

MATHEMATICAL TERRORISM

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Alexander Gutfraind

February 2010

© 2010 Alexander Gutfraind
ALL RIGHTS RESERVED

MATHEMATICAL TERRORISM

Alexander Gutfraind, Ph.D.

Cornell University 2010

Mathematical research of terrorism has the potential to inform both scholars and policy-makers. This thesis presents several projects in this emerging area: (1) an ordinary-differential equations model of a terrorist organization focused on evaluating various counter-terrorism measures and predicting the evolution of terrorist conflicts; (2) a model of nuclear smuggling where the adversary is described as a Markov process on a transportation network and algorithms for positioning sensor arrays on the network; (3) a new formulation of nuclear smuggling that allows fast computation using approximation algorithms with performance guarantees; (4) a model for constructing cascade-resilient networks, with implications for analyzing the structure of terrorist networks, specifically their susceptibility to betrayal.

BIOGRAPHICAL SKETCH

The journey to the thesis took many unlikely turns. As a proud citizen of the Byelorussian Soviet Socialistic Republic Sasha was admitted to October Youth, and was working towards becoming a Pioneer. However, he was soon studying Hebrew in Israel and then English in Canada. He came to the University of Waterloo to become an Computer Engineer, but soon switched to Applied Math, since General Relativity couldn't fit the schedule. Achim Kempf agreed to supervise a thesis in Theoretic Physics, but first suggested a small problem in the evolution of the genetic code which became a Master's thesis. He came to Cornell to study Mathematical Biology and Networks, but taking Jon Kleinberg's seminar on networks found a gap - no papers in mathematical terrorism. Filling this gap was fun.

ACKNOWLEDGEMENTS

Without Gino Tenti's substitute calculus lectures I would have never heard of Applied Mathematics. Without Stan Lipshitz's time and Achim Kempf's investment and encouragement I would have never developed the confidence to bring ideas and defend them. Without Rich Durrett's selfless generosity I would have never had a chance to write this thesis. Without my parents and friends I would not have had anybody to brag about itto.

TABLE OF CONTENTS

Biographical Sketch	iii
Acknowledgements	iv
Table of Contents	v
List of Figures	vii
1 Introduction	1
References	10
2 Understanding Terrorist Organizations with a Dynamic Model	11
2.1 Introduction	12
2.2 A Mathematical Model	14
2.3 Analysis of the Model	18
2.4 Discussion	22
2.4.1 Nascent terrorist organizations	22
2.4.2 Conditions for Victory	24
2.4.3 Stable Equilibria	26
2.5 Counter-Terrorism Strategies	27
2.5.1 Targeting the leaders	27
2.5.2 Encouraging desertion	29
2.5.3 Minimization of Strength S	30
2.6 Conclusions	31
2.A Appendix	32
2.A.1 The Dynamical System	32
2.A.2 Proof of the theorem	33
2.A.3 Concrete Example of Strength Minimization	35
References	37
3 Interdiction of a Markovian Evader	40
3.1 Introduction	40
3.2 The interdiction model	43
3.2.1 Markovian evaders	45
3.3 Evader models	48
3.3.1 Least-cost-guided evader	49
3.3.2 Non-retreating evader	52
3.4 Solving the Markovian Evader Interdiction Problem (MENI)	54
3.4.1 Global heuristics	55
3.4.2 Betweenness centrality heuristic	58
3.4.3 Performance Results	60
3.5 Conclusions and Outlook	61
3.A The Motion Heuristic	65

References	68
4 Optimal Interdiction of Unreactive Markovian Evaders	70
4.1 Introduction	71
4.2 Unreactive Markovian Evader	72
4.2.1 Evaders	73
4.2.2 Interdictor	74
4.2.3 Objective function	75
4.3 Complexity	77
4.4 An Efficient Interdiction Algorithm	81
4.4.1 Submodularity of the interdiction problem	82
4.4.2 Greedy algorithm	84
4.4.3 Solution quality	85
4.4.4 Exploiting submodularity with Priority Evaluation	86
4.5 Outlook	90
4.A Mixed integer program for UME	92
References	95
5 Designing Networks for Cascade Resilience	97
5.1 Introduction	98
5.2 Formal Model	103
5.2.1 Measuring Resilience	104
5.2.2 Measuring Efficiency	105
5.3 Network Designs	107
5.3.1 Solution Approach	107
5.3.2 Network Designs	109
5.4 Results	112
5.4.1 Optimal Network	112
5.4.2 Effect of attenuation	116
5.5 Discussion	118
5.6 Conclusions and Future Work	119
5.A Continuity of Fitness	122
5.B Extent and Contagion Risk	123
5.C Simulation Methodology	125
5.D Analytic Results	126
5.E Configuring the Optimal Design	128
5.F Sensitivity Analysis	128
References	136

LIST OF FIGURES

2.1	Phase plane of the model	19
2.2	Evolution of strength, leaders and foot soldiers (S, L, F , respectively) in two terrorist organizations as a function of time.	20
2.3	Plot of the sink (thick red) and trend lines (thin dashed black)	20
2.4	The effects of the parameters b and k on the dynamical system	21
2.5	The effects of the parameters p , r and d on the dynamical system	22
2.6	The effects of d_L and d_F on the dynamical system	30
2.7	Illustration for the proof	34
2.8	Graphical calculation of optimal budget allocation	36
3.1	Example network where least-cost path interdiction is suboptimal	43
3.2	Computation of the transition probabilities M_{ij}	51
3.3	The cost of reaching the target as a function of λ	52
3.4	Comparison on the grid networks of several algorithms	62
3.5	Comparison on the Washington, DC network of several algorithms	63
3.6	Example network where motion likelihood heuristic outperforms the betweenness heuristic.	66
4.1	Illustration of the reduction of Set Cover to UME-Decision.	79
4.2	Comparison between the basic greedy and the priority greedy algorithms	90
5.1	WWII underground network “Francs-tireurs Partisans”	100
5.2	Illustration of the 6 designs.	111
5.3	Fitness of the optimal configuration for various designs	113
5.4	Resilience of the optimal design.	115
5.5	Efficiency of the optimal design.	115
5.6	Average degree in the optimal configuration of each design.	116
5.7	Fitness of the optimal configuration in the stars design	116
5.8	Fitness of the optimal configuration for each design when $g = 0.1$	117
5.9	Cell size k in the optimal configuration of each design.	129
5.10	Connectivity p in the optimal configuration of each design.	129
5.11	Standard deviation in resilience, within the top 5% of solutions.	130
5.12	Standard deviation in efficiency, within the top 5% of solutions.	131
5.13	Standard deviation in average degree, within the top 5% of solutions.	132
5.14	Standard deviation in cell size k of the top 5% of solutions.	133
5.15	Standard deviation in connectivity p of the top 5% of solutions.	134

CHAPTER 1

INTRODUCTION

The application of mathematical methods to counter-terrorism is not typically found in the sophomore applied math curriculum and certainly our undergraduate students can find more benign applications of mathematics. However, this thesis argues that terrorism is an area where mathematical methods can make an impact in a variety of targets/research problems.

State of the Field Judging by the practitioners, mathematical terrorism is an area lying at the clash of disciplines such as operations research, economics and machine learning (in that order). To date much of the research and policy interest in mathematical models has been as a tool for helping in intelligence gathering - using some sort of an algorithms for finding the bad guys in a big pile of heterogeneous data. This is a vast problem. Unfortunately there are indications that little real progress has been made to develop such tools (unless the terrorists commit credit card fraud, in which case terrorism reduces to a known problem and machine learning can help.) The difficulty is not surprising because (1) our adversaries are smart, (2) training data is in short supply (mercifully).

Slow progress did not stop people from trying. The earliest quantitative papers specifically in terrorism have appeared in the 1980s as a consequence of a wave of hijackings [11]. In recent years some of the most interesting findings has been in identifying regularities in terrorist attacks and levels of violence. Namely, it has been found that casualties from terrorist events tend to scale as a power law with an exponent of 2.5. Moreover, conflicts across the globe and across regions appear to be converging towards a single type of violence as measured by the distribution of casualties [7]. It was also found that terrorist violence come in waves of period 13 ± 1 years [3], but not the waves proposed on qualitative grounds [4]. It is notable that these patterns have been

entirely missed by the political science community and it remains to be seen whether good models can be developed to account for such trends.

Philosophy Unlike perhaps fluid dynamics (or e.g. climate change), terrorism as a whole is far too complex a phenomenon to be described by a general mathematical model. The same is of course true of disciplines like biology where no single model can describe the phenomena in all scales simultaneously (e.g. both DNA replication and the evolution of ecosystems). Yet in both biology and terrorism there are niches where models can help evolve a more powerful understanding. At the very least the attempt to formulate a model helps raise questions that would not occur to botanists (or counter-terrorism experts) who tend to be qualitatively-minded.

It is worth quoting Rapoport, who in his critique of Richardson's model of war observed that models help in ways not typically appreciated [5]:

Contrary to a prevalent meaning of "model" in many theoretical formulations, the main function of a mathematical model is not an "explanatory" one. A mathematical model is more characteristically a point of departure rather than a point of arrival in the construction of a theory. In this way it is akin to the null hypothesis, which, incidentally, also involves the construct of a mathematical model. In most cases, null hypotheses are made so that they can be refuted. As a by-product of the refutation of the null hypothesis, biases are usually discovered which point to the direction of search for "causes." It is much the same with mathematical models. These models are often deliberately made simple-minded with full knowledge that they do not represent reality. Their chief value is that they lead to *compelling* consequences. These consequences are then compared with observations. As

often as not, the derived consequences do *not* agree with the observations. But then the direction and magnitude of the departures may indicate the direction of further search. Richardson fully realizes both the usefulness and the limitations of mathematical models and repeatedly emphasizes both.

Rapoport's ideal of mathematical modeling is not always followed in this thesis. First, most of the models are based on qualitative notions rather than on comparison with data. More serious still is the dangerous tendency found in all of the chapters in this thesis to advance specific policy recommendations - these must be qualified for obvious methodological reasons. For instance some of the models make precise claims even though the models are founded on poorly-understood simplifications (at one point the author claims to have a method for finding "at least 63% of the optimal solution").

These defects are all signs of the immaturity of the field. As it would grow one expects to see new models becoming increasingly realistic and validated. They would also begin receiving more attention from policy-makers and non-mathematicians. For an example of this trend consider a recent econometric paper on the interaction between Improvised Explosive Devices (IED) and countermeasures in Iraq [6]. The authors showed that the countermeasures are effective despite the paradoxical increase in both the number of casualties and attacks (reason: other targets became so hardened that IED attacks became the "inferior good" that insurgencies started "buying" more.) That paper was cited by the U.S. Department of Defense in its congressional testimony, helping continue billions of dollars of funding for Joint IED Defeat Organization (\$13B by April 2008).

Overview of Contents The current thesis is a collection of papers, about half of which have already passed peer review (Ch. 2,4) while others, some almost as mature, are

just entering it (Ch. 3,5). The problems come from different areas of terrorism, and the models involve very different methodologies, including ordinary differential equations, Markov processes on networks, approximation algorithms for discrete optimization problems, as well as network analysis.

Chapter 2 This chapter was motivated by a basic problem in counter-terrorism [19]: when fighting Al-Qaida, is the best strategy to attack the leadership of the group or to attack its rank-and-file i.e. its foot soldiers? The former seem to be more valuable targets, in some ways easier to stop, but is attacking them more effective in the long term? How serious is the oft-suggested risk of “blowback” (counter-terrorism measures helping the terrorists)? This problem is not unlike the problems found in biology where sometimes the objective is to develop a strategy to control a species which might be invading an ecosystem. Does it better to attack the reproducing adults or the larvae?

To address those issues Chapter 2 introduces a dynamic model of terrorist organizations (the original model was published in a terrorism journal [8], and is to appear in a collection about mathematical models in counter-terrorism [9]). The model is a very simplified compartmental model consisting of two coupled linear ordinary differential equations (ODEs). For comparison, there are surprisingly few ODE-based models of terrorism, and the existing ones look at different questions.

The original version of the model was non-linear and therefore hard to analyze (Steve Strogatz suggested that the author linearize it.) Interestingly, the biggest problem with non-linearities in this area is not analytical (we thankfully have numerical integration to fall back to, unlike Gauss). Rather it is the number of possible non-linearities and their dependence on specifics of the conflict. Whereas in the fight against Al-Qaida in Western Europe it might be worthwhile to consider non-linearity due to resentment of

counter-terrorism operations, in the fight against the same group in Iraq it is perhaps more important to consider non-linearity due to self-catalytic recruitment of security forces, and so forth - very different forces appear more important in different terrorist groups. It is likely that the linear model captures a more general aspect of terrorism, in ways that the extensions cannot. Also, the simplicity should hopefully make it an attractive target for extensions by enterprising undergraduate students in dynamical systems (indeed the author has already received two e-mails about re-implementing it.)

Despite the radical oversimplification, the resulting model makes some interesting predictions: theoretically it is possible to guarantee the demise of a terrorist group based on certain derivative conditions that should be possible to check in real life. The model shows that if the leadership is being reduced and the organization is weakened then it can yet recover, but it cannot recover from a decline in foot soldiers and a weakening. This is essentially due to the compartmental structure - the system has a delay in restoring the leaders pool but no delay with recruiting foot soldiers. This finding is suggestive of what happened in Al-Qaida in 2001, and surely contains some truth, but of course the real situation is not quite so simple.

Chapter 3 A major terrorism threat comes from non-conventional weapons. Indeed it has already happened at least twice beside the famous Anthrax attacks (Tokyo Sarin gas attacks and the much lesser known salmonella poisonings in The Dalles, OR [22]). Studies indicate that the threat is real:

1. there is considerable interest by Islamist terrorists in acquiring attack capabilities (other violent groups are typically more interested in doing a good show, and large casualty figures are seen as counterproductive),
2. a team of 12 technically-trained terrorists should be able to build a working gun-

type uranium weapon as soon as they acquire enriched Uranium and

3. cost-benefit studies show that a single nuclear weapon attack prevented every 400 years is sufficient to justify astronomical expenditures in prevention.

Thus far considerable resources have been put towards (a) securing the borders and (b) stopping nuclear smuggling from ex-soviet nuclear sites (whose precise locations are apparently unknown even to the Russians). For (b) a network of detectors for radioactive materials has been deployed in parts of the ex-Soviet Union. The mathematical problems arising from this effort are in operations research. Namely, where should those detectors be placed subject to various constraints such as money? This set of problems is known as “network interdiction”, and has research going back to the origins of operations research, namely, min-cut problems (notice though that in this problem edges have costs rather than capacities).

The work in chapter 3 (also in chapter 4) was started as a summer internship in Los Alamos National Lab with Aric Hagberg (Theoretical Division), Feng Pan and David Izraelevitz (Risk Analysis and Decision Support Systems Division.) The project’s first contribution is to develop a Markovian model of the nuclear smuggler (termed “evader”) and to show how to compute the evader’s expected cost to reaching his smuggling objective. Based on this model it is possible to define a stochastic version of network interdiction where the objective is to maximize the expected cost of reaching the target. Stochasticity in those models is not from deliberate evasion, but rather represents the evader’s inability to find the best path to follow, since his computation would be confounded by the difficulty of estimating risk on different paths, even if the least-time or least-distance path was easy to find. This model significantly advances the field in which the main model is of an omniscient minimizer who sees all the detectors and finds the least-cost path around them (these are large oversimplifications). The new model has

a parameter that adjusts the level of stochasticity from complete randomness to complete determinism (in that latter limit it becomes the omniscient model.) The author's contribution has been to

1. develop a mechanistic explanation of the model,
2. develop a software implementation of the algorithms,
3. show that the omniscient model is bad because it not only leads to suboptimal solutions, but paradoxically, it can also lead to solutions that make matters worse, and
4. most importantly, to develop two solution heuristics that can find good solutions even in very large problems.

Those heuristics select edges to be interdicted based on global information about the transportation network (edge betweenness centrality) and information about evader motion (which edges are likely to be transited). They are shown to perform much better than standard methods on both synthetic and real transportation networks.

Unfortunately, the problem of finding the optimal detector placement on the transportation network cannot be approximated arbitrarily well in polynomial time. Indeed the heuristics do not come with guaranteed performance and synthetic instances exist where they both are arbitrarily bad. On the whole though, the stochastic model and the heuristic represent a significant increase in the quality of the solutions one can find and also a significant increase in modeling realism.

Chapter 4 The complexity of the evader problem in chapter 3 suggested to the author to try to solve a simpler version of the network interdiction problem, as follows

(originally published in [11]).¹ Specifically, suppose detectors could be hidden so that the evader is unable to determine their location on the transportation network. This is not unrealistic: placing decoy detectors would make it hard to find the real ones. Also in large transportation networks finding the shortest path (or specifically the least-risk path) becomes very hard.

An evader satisfying this condition could be termed “unreactive”. Unlike with the deterministic model discussed earlier, this evader is moving stochastically on the network and hence, there is no guarantee that it would be possible to catch him despite his unreactivity. However, the optimization problem itself becomes much easier (not surprisingly.) The new objective function - the probability of capture as a function of the detection set - is submodular in the detection set. This is due to the fact that there is no synergy when detectors isolate the target and instead there is an interference between different detectors: the sum of their gain alone \geq their gain when working together.

Submodularity implies that a greedy algorithm finds solutions that are $> 63\%$ of the optimal solution. The author has further improved performance using a “fast initialization” technique that exploits unique properties of this problem. As a result it was experimentally found that the computation time seems to be independent of the number of edges in the network, and it becomes possible to solve network interdiction on very large transportation networks (up to 30,000 nodes have been tried, but it should scale to millions of nodes and is highly parallelizable.) It is possible that the fast initialization technique could be exploited in other important submodular optimization problems.

Chapter 5 An interesting and hard problem in terrorism research is to understand the network structure of terrorist groups. Since terrorist groups are often operating in a

¹A professor of the author once quoted Polya as saying that if you see a problem that you cannot solve, then try finding a simpler problem (which you might not be able to solve either.)

hostile social and legal environment, it is paramount for their secret social network to be able to resist captures or betrayals. A classic design of secret societies is based on cells so that the capture of one node typically leads to only the capture of the other members of the cell.

This problem motivates chapter 5, which describes an investigation into the optimal structure of terrorist networks (originally appeared as [12] and now revised in response to two peer reviews.) What can be the outcome of such an investigation, beyond satisfying academic curiosity and helping Al-Qaida? Actually, terrorist networks are not the only type of secret societies. Indeed, dissident groups and guerrillas such as those who sprung in Nazi-occupied Europe are also organized in this way. More importantly, cascading failures such as those found in terrorist networks are also found in a variety of other problems (including computer virus infections, electric power networks, supply chains and others too numerous to list).

This cascade risk motivates the search for mechanisms or topological properties that would increase the networks' cascade resilience. The chapter introduces a simple model under which networks are optimized for resilience and also performance, and describes the optimal networks under this model. The results are perhaps surprising and in particular, it was found that when the exogenous risk of cascades is low the optimal networks are densely-connected, but they are also densely-connected when the risk of cascades is very high (depending on the parameter values). Moreover, one of the practical insights is that it was found that certain types of networks are not suitable to attempts to improve cascade resilience, and it is better to spend resource to improve their performance ignoring the risk of cascades.

REFERENCES

- [1] T. Sandler, J. T. Tschirhart, and J. Cauley, “A theoretical analysis of transnational terrorism,” *American Political Science Rev.*, vol. 77, pp. 36–54, Mar 1983.
- [2] N. F. Johnson, M. Spagat, J. A. Restrepo, O. Becerra, J. C. Bohrquez, N. Surez, E. M. Restrepo, and R. Zarama, “Universal patterns underlying ongoing wars and terrorism,” *arxiv.org*, vol. physics/0506213, 2006.
- [3] A. Clauset, M. Young, and K. S. Gleditsch, “On the frequency of severe terrorist events,” *Journal Of Conflict Resolution*, vol. 51, p. 58, 2007.
- [4] D. C. Rapoport, “The four waves of terrorism,” in *Attacking Terrorism: Elements of a Grand Strategy* (A. K. Cronin and J. M. Ludes, eds.), pp. 46–73, Georgetown University Press, Washington, DC, USA, 2004.
- [5] A. Rapoport, “Lewis F. Richardson’s mathematical theory of war,” *Journal of Conflict Resolution*, vol. 1, pp. 249–, Sep 1957.
- [6] M. A. Hanson, “The economics of roadside bombs,” *SSRN eLibrary*, 2008. <http://ideas.repec.org/p/cwm/wpaper/68.html>.
- [7] B. Ganor, *The Counter-Terrorism Puzzle: A Guide for Decision Makers*. Piscataway, NJ: Transaction Publishers, 2005.
- [8] A. Gutfraind, “Understanding terrorist organizations with a dynamic model,” *Studies in Conflict and Terrorism*, vol. 32, pp. 45–59, Jan 2009.
- [9] A. Gutfraind, *Mathematical Methods in Counterterrorism*, ch. Understanding Terrorist Organizations with a Dynamic Model. Springer-Verlag, 2009.
- [10] B. Hoffman, *Inside Terrorism*. Columbia University Press, USA, 2006.
- [11] A. Gutfraind, A. A. Hagberg, and F. Pan, “Optimal interdiction of unreactive Markovian evaders,” in *CPAIOR 2009* (J. Hooker and W.-J. van Hoeve, eds.), Lecture Notes in Computer Science, Springer, May 2009. <http://arxiv.org/abs/0903.0173>.
- [12] A. Gutfraind, “Constructing networks for cascade resilience,” 2009. <http://arxiv.org/abs/0906.0786>.

CHAPTER 2

UNDERSTANDING TERRORIST ORGANIZATIONS WITH A DYNAMIC MODEL

Chapter Abstract Terrorist organizations change over time because of processes such as recruitment and training as well as counter-terrorism (CT) measures, but the effects of these processes are typically studied qualitatively and in separation from each other. Seeking a more quantitative and integrated understanding, we constructed a simple dynamic model where equations describe how these processes change an organization's membership. Analysis of the model yields a number of intuitive as well as novel findings. Most importantly it becomes possible to predict whether counter-terrorism measures would be sufficient to defeat the organization. Furthermore, we can prove in general that an organization would collapse if its strength and its pool of foot soldiers decline simultaneously. In contrast, a simultaneous decline in its strength and its pool of leaders is often insufficient and short-termed. These results and other like them demonstrate the great potential of dynamic models for informing terrorism scholarship and counter-terrorism policy making.

2.1 Introduction

Our goal is to study terrorist organizations using a dynamic model. Generally speaking, in such a model a phenomenon is represented as a set of equations which describe it in simplified terms. The equations represent how the phenomenon changes in time or space, and cast our empirically-based knowledge in precise mathematical language. Once the model is constructed, it can be studied using powerful mathematical techniques to yield predictions, observations and insights that are difficult or impossible to collect empirically [1, 2]. For example, a dynamic model could be constructed for the various militant groups operating in Iraq and then used to predict their strength a year in the future. Moreover, given the model, it would be possible to evaluate the efficacy of various counter-insurgency policies.

Mathematical models can help fill a large methodological void in terrorism research: the lack of model systems. Whereas biologists studying pathogens can do experiments *in vitro*, there are no such model systems in terrorism research, except for mathematical models. In this sense, the method developed below provides an *in vitro* form of terrorism, which can be investigated in ways not possible in its *in vivo* kind. Like all model systems, mathematical models are imperfect because they rely on large simplifications of the underlying political phenomena, and one can rightfully ask whether their predictions would be sufficiently accurate. Fortunately, complex phenomena in fields like biology have been studied very successfully with this mathematical technique [2]. Therefore, even phenomena as complex as found in terrorism research may, in some cases, be productively studied using mathematical models and indeed, previous models have brought considerable insights¹.

¹E. g. dynamic models: [3, 4, 5, 6, 7, 8, 9], rational choice models: [10, 11, 12, 13], agent-based models: [14, 15].

In the rest of the paper we describe a simple model of a terrorist organization. The model is new in its focus, methodology and audience: We focus on a single terrorist organization and model its processes of recruitment, its internal dynamics as well as the impact of counter-terrorism measures on it. As to methodology, with a few exceptions [4, 6, 8, 9] and perhaps a few others the powerful mathematical technique of differential equations has not been applied in terrorism research. Finally, the paper is specifically written to an audience of non-mathematicians: the main body of the paper uses non-technical language to explain the terminology and to describe the equations and assumptions used in the model, while the technical analysis is exposed in the appendix.

The model described below was built following two design principles. First, it was desired to have a model of broad applicability across organizations and conflicts. Indeed, the model is so general that it can be applied to insurgencies or even to some non-terrorist organizations. As we shall see, despite this generality it makes non-trivial observations and more importantly it specifies sufficient conditions for victory over the organization (see subsection 2.4.2). Second, it was desired to build a simple model so as to facilitate interpretation, analysis and further development. It was hoped that the model would establish a methodological prototype that could be easily extended and modified to fit specific cases.

The organization of the paper is as follows. Section 2 describes the model - its variables, parameters and relations between them. Section 3 graphically illustrates the model's predictions about terrorist organizations. Sections 4 and 5 discuss the insights gleaned from the model, and the implications to counter-terrorism policies. The conclusions are in Section 6. Finally, all of the technical arguments are gathered in the appendix.

2.2 A Mathematical Model

There are many ways of describing a terrorist organization, such as its ideology or political platform, its operational patterns, or its methods of recruitment. Here we consider it from the “human resources” point of view. Namely, we are interested in examining how the numbers of “leaders” and “foot soldiers” in the organization change with time. The former includes experienced managers, weapon experts, financiers and even politicians and religious leaders who help the organization, while the latter are the more numerous rank-and-file. These two quantities arguably give the most important information about the strength of the organization. The precise characteristics of the two groups and their relative sizes would depend on the organization under consideration. Nevertheless, this distinction remains relevant even in the very decentralized organizations like the post-Afghanistan al-Qaeda movement, because we can identify the “leaders” as the experienced terrorists, as compared to the new recruits (see discussions in [16, 17]). The division between those two groups is also important in practice because decision makers often need to choose which of the groups to target [18, 19, Ch.5]: while the leaders represent more valuable targets, they are also harder to reach. Later on in section 5 we actually compare the policy alternatives.

Therefore, let us represent a terrorist organization as two time-varying quantities, L and F , corresponding to the number of leaders and foot soldiers, respectively. Also, L and F determine the overall “strength” S of the organization. Because leaders possess valuable skills and experience, they contribute more to the strength than an equivalent number of foot soldiers. Hence, strength S is taken to be a weighted sum of the two variables, with more weight ($m > 1$) given to leaders:

$$S = mL + F$$

We now identify a set of processes that are fundamental in changing the numbers of leaders and foot soldiers. These processes constitute the mathematical model. While some of them are self-evident, others could benefit from quantitative comparison with data. The latter task is non-trivial given the scarcity of time-series data on membership in terrorist organizations and hence we leave it out for future work.

The histories of al-Qaeda and other terrorist organizations e. g., [20, 21, 22] suggest that the pool of terrorist leaders and experts grows primarily when foot soldiers acquire battle experience or receive training (internally, or in terrorist-supporting states, [23]). Consequently, the pool of leaders (L) is provisioned with new leaders at a rate proportional to the number of foot soldiers (F). We call this process “promotion” and label the parameter of proportionality p . This growth is opposed by internal personnel loss due to demotivation, fatigue, desertion as well as in-fighting and splintering [24, Ch.6]. This phenomenon is modeled as a loss of a fraction (d) of the pool of leaders per unit time. An additional important influence on the organization are the counter-terrorism (CT) measures targeted specifically at the leadership, including arrests, assassinations as well as efforts to disrupt communications and to force the leaders into long-term inactivity. Such measures could be modeled as the removal of a certain number (b) of people per unit time from the pool of leaders. CT is modeled as a constant rate of removal rather than as a quantity that depends on the size of the organization because the goal is to see how a fixed resource allocation towards CT would impact the organization. Presumably, the human resources and funds available to fight the given terrorist organization lead, on average, to the capture or elimination of a fixed number of operatives. In sum, we assume that on average, at every interval of time the pool of leaders is nourished through promotion, and drained because of internal losses and CT (see appendix, equation (2.1).)

The dynamics of the pool of foot soldiers (F) are somewhat similar to the dynamics

of leaders. Like in the case of leaders, some internal losses are expected. This is modeled as the removal of a fraction (d) of the pool of operatives per unit time where for simplicity the rate d is the same as the rate for leaders (the complex case is discussed in subsection 2.5.2.) Much like in the case of leaders above, counter-terrorism measures are assumed to remove a fixed number (k) of foot soldiers per unit time. Finally and most importantly, we need to consider how and why new recruits join a terrorist organization. Arguably, in many organizations growth in the ranks is in proportion to the strength of the organization, for multiple reasons: Strength determines the ability to carry out successful operations, which increase interest in the organization and its mission. Moreover, strength gives the organization the manpower to publicize its attacks, as well as to man and fund recruitment activities. By assuming that recruitment is proportional to strength, we capture the often-seen cycle where successful attacks lead to greater recruitment, which leads to greater strength and more attacks. Overall, the pool of foot soldiers is nourished through recruitment, and drained because of internal losses and CT (see appendix, equation (2.2))^{2,3}.

The numerical values of all of the above parameters (p, d, b, r, m, k) are dependent on the particular organization under consideration, and likely vary somewhat with time⁴. Fortunately, it is possible to draw many general conclusions from the model without knowing the parameter values, and we shall do so shortly. Finally, it should be noted that

²A minor assumption in our model is that once a foot soldier is promoted a new foot soldier is recruited as a replacement. It is shown in the appendix that if in some organizations such recruitment is not automatic, then the current model is still valid for these organizations as long as $p < r$. In any case the drain due to promotion is marginal because foot soldiers are far more numerous than leaders even in relatively “top heavy” organizations.

³This model is similar to structured population models in biology, where the foot soldiers are the “larvae” and the leaders are the “adults”. However, an interesting difference is that whereas larvae growth is a function of the adult population alone, in a terrorist organization the pool of foot soldiers contributes to its own growth.

⁴The simplest approach to estimating them would be to estimate the number and leaders and foot soldiers at some point in time, and then find the parameter values by doing least-squares fitting of the model to the data on the terrorist attacks, where we consider the terrorist attacks to be a proxy of strength. However, this approach has some limitations.

counter-terrorism need not be restricted to the parameters b, k (removal of leaders and foot soldiers, respectively), and measures such as public advocacy, attacks on terrorist bases, disruption of communication and others can weaken the organization by reducing its capabilities as expressed through the other parameters.

In the above description, we assumed that counter-terrorism measures are parameters that can be changed without affecting recruitment. This is a significant simplification because in practice it may be difficult to respond to terrorist attacks without engendering a backlash that actually promotes recruitment e. g., [25, 26]. Nevertheless, the advantages of this simplification outweigh the disadvantages: Firstly, it is clear that any model that would consider such an effect would be much more complicated than the current model and consequently much harder to analyze or use. Secondly, the current model can be easily extended to incorporate such an effect if desired. Thirdly, the strength of this effect is difficult to describe in general because it depends extensively on factors such as the specific CT measures being used, the terrorist actions and the political environment. Indeed, [9] who incorporated this effect, constructed their model based on observations of a specific context within the current conflict in Iraq.

The model includes additional implicit assumptions. First, it assumes a state of stable gradual change, such that the effect of one terrorist or counter-terrorism operation is smoothed. This should be acceptable in all cases where the terrorist organization is not very small and thus changes are not very stochastic. Second, the model assumes that an organization's growth is constrained only by the available manpower, and factors such as money or weapons do not impose an independent constraint. Third, it is assumed that the growth in foot soldiers is not constrained by the availability of potential recruits - and it is probably true in the case of al-Qaeda because willing recruits are plentiful (for the case of England, see [27]). We discuss this point further in subsection 2.4.3.

2.3 Analysis of the Model

Having written down the governing equations, the task of studying a terrorist organization is reduced to the standard problem of studying solutions to the equations. Because the equations indicate rates of change in time, the solutions would be two functions, $L(t)$ and $F(t)$, giving the number of leaders and foot soldiers, respectively, at each time. Let us suppose that currently (time 0) the organization has a certain number of leaders and foot soldiers, L_0 and F_0 , respectively and is subject to certain CT measures, quantified by b and k . We want to see whether the CT measures are adequate to defeat the organization. Mathematically, this corresponds to the question of whether at some future time both L and F would reach zero. Intuitively, we expect that the organization would be eliminated if it is incapable of recovering from the losses inflicted on it by CT. In turn, this would depend on its current capabilities as well as the parameters p, d, r, m which characterize the organization.

Mathematical analysis of the model (see the appendix) shows that most terrorist organizations⁵ evolve in time like the organizations whose “orbits” are displayed in Fig. 2.1(a,b). In Fig. 2.1(a) we plotted eight different organizations with different starting conditions. Another perspective can be seen in Fig. 2.1(b) which graphically illustrates the dynamical equations via arrows: the direction of each arrow and its length indicate how an organization at the tail of the arrow would be changing and at what rate. By picking any starting location (L_0, F_0) and connecting the arrows, it is possible to visually predict the evolution into the future. Another illustration is found in Fig. 2.2, which shows how two example organizations change with time.

⁵That is, those with realistically low rates of desertion: $d < \frac{1}{2}(r + r\sqrt{1 + \frac{mp}{r}})$. A higher rate of desertion d always causes the organization to collapse and is not as interesting from a policy perspective (see subsection 2.5.2 for a discussion of desertion).

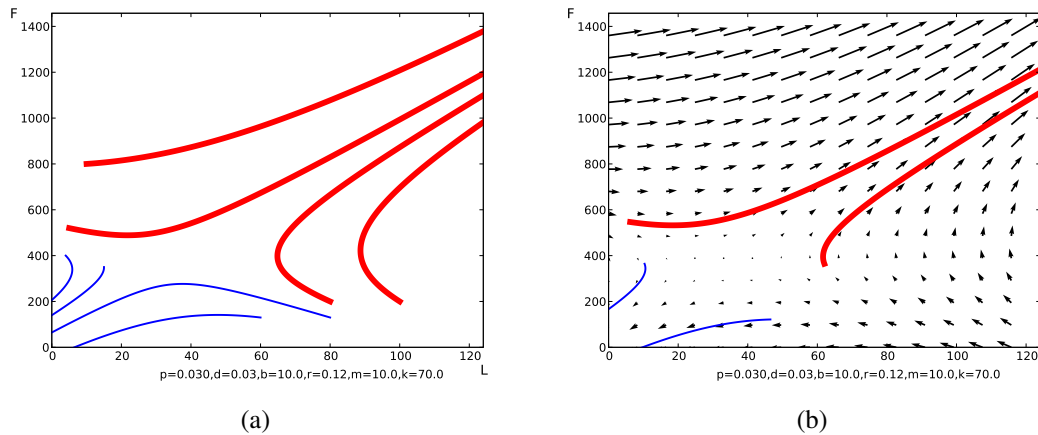


Figure 2.1: (a) Typical solution curves of the equations coded by ultimate fate: thin blue for successfully neutralized organizations and thick red for those remaining operational and growing. The parameters were set to representative values, but as was said earlier, all realistic organizations are qualitatively similar and resemble these. (b) "Vector field" of L and F . At each value of L, F the direction and length of the arrow give the rate of change in L and F .

In general, it is found that the dynamics of the organization is dependent upon the position of the organization with respect to a threshold line, which can be termed the "sink line": an organization will be neutralized if and only if its capabilities are below the sink line. In other words, the current CT measures are sufficient if and only if the organization lies below that threshold (thick red line on Fig. 2.3). The threshold is impassable: an organization above it will grow, and one below it is sure to collapse. This threshold is also very sharp: two organizations may lie close to the line, but the one above it would grow, while the one below it would shrink even if the differences in initial capabilities are small. In addition to the sink line, the model also predicts that all successful organizations would tend towards a particular trajectory. This "trend line" (a dashed black line on Fig. 2.3) is discussed further in subsection 2.4.1.

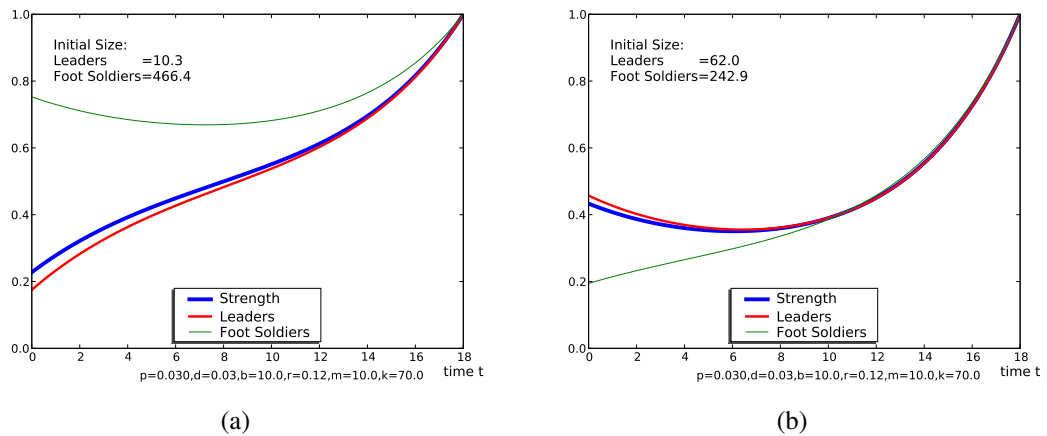


Figure 2.2: Evolution of strength, leaders and foot soldiers (S, L, F , respectively) in two terrorist organizations as a function of time. In (a), due to CT, F falls initially but eventually the organization recovers through promotion. In (b), L and S fall initially but eventually the organization recovers through recruitment. The vertical axis has been rescaled by dividing each quantity by the maximum it attains during the time evolution. This makes it possible to represent all three quantities on the same plot. The units of time are unspecified since they do not affect the analysis. Of course, in a more complex model it would be desirable to consider periodic events like election cycles or generational changes.

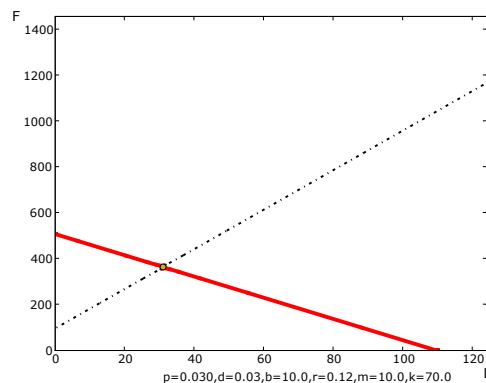


Figure 2.3: Plot of the sink (thick red) and trend lines (thin dashed black). The two lines intersect at a “saddle point”.

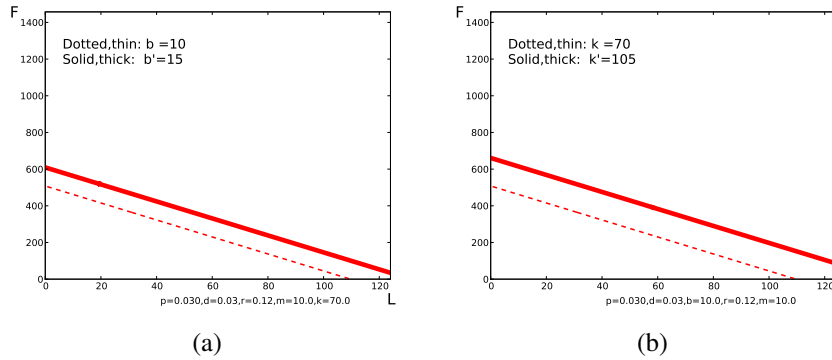


Figure 2.4: The effects of the parameters b and k on the dynamical system, (a) and (b) respectively, as seen through the effect on the sink line. In each case, as the CT measures are increased, the sink line moves up confining below it additional terrorist organizations.

Suppose now that the model predicts that the given organization is expected to grow further despite the current CT measures, and therefore increased CT measures would be needed to defeat it. To see the effect of additional CT measures, we need to examine how the dynamical system changes in response to increases in the values of the parameters, in particular, the parameters b and k which express the CT measures directed at leaders and foot soldiers, respectively (Fig. 2.4).

It is also possible to affect the fate of the organization by influencing the values of other parameters affecting its evolution, such as recruitment and promotion (Fig. 2.5). In general, to bring the terrorist organization under control it is necessary to change the parameters individually or simultaneously so that the organization's current state, (L, F) , is trapped under the sink line. An interesting finding in this domain is that both b and k are equivalent in the sense that both shift the sink link up in parallel (Fig. 2.4).

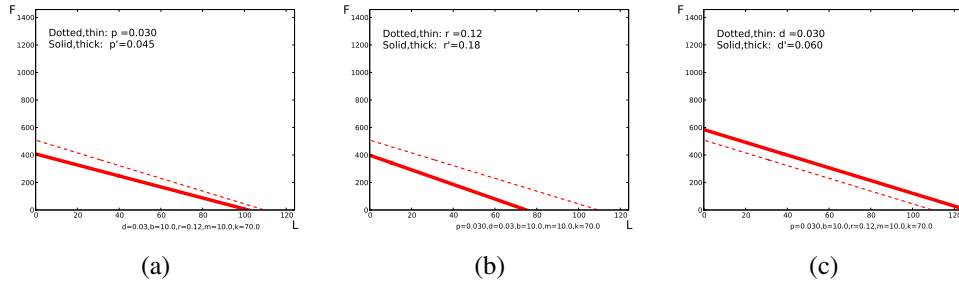


Figure 2.5: The effects of the parameters p (a), r (b) and d (c) on the dynamical system as seen through the effect on the sink line. When p or r are increased the organizations are able to grow faster, causing the sink line to move down, making the existing CT measures no longer sufficient to neutralize some terrorist organizations. In contrast, when d is increased, the sink line moves up because the organization is forced to replace more internal losses to survive.

2.4 Discussion

2.4.1 Nascent terrorist organizations

Recall that the sink line (Fig. 2.3) distinguishes two classes of terrorist organizations - those destined to be neutralized and those that will continue growing indefinitely. Within the latter group, another distinction is introduced by the trend line - a distinction with significance to counter-terrorism efforts: organizations lying to the left of it have different initial growth patterns compared to those lying to the right (Fig. 2.1). The former start with a large base of foot soldiers and a relatively small core of leaders. In these organizations, F may initially decline because of CT, but the emergence of competent leaders would then start organizational growth (e. g. Fig. 2.2(a)). In contrast, the latter type of organizations start with a large pool of leaders but comparatively few recruits. CT could decimate their leadership, but they would develop a wide pool of foot soldiers,

recover and grow (e. g. Fig. 2.2(b)). Thus, all successful terrorist organizations may be classified as either “p-types” (to the left of the trend line) or “r-types” (to the right of the trend line) in reference to the parameters p of promotion and r of recruitment. In p-type organizations early growth occurs mainly through promotion of their foot soldiers to leaders, while in the r-types mainly through recruitment of new foot soldiers.

This classification could be applied to many actual organizations. For example, popular insurgencies are clearly p-type, while al-Qaeda’s history since the late 1990s closely follows the profile of an r-type: Al-Qaeda may be said to have evolved through three stages: First, a core of followers moved with bin Laden to Afghanistan. They were well-trained but the organization had few followers in the wider world (for a history see [28]). Then the attacks and counter-attacks in the Fall of 2001 reduced the organization’s presence in Afghanistan leaving its operatives outside the country with few leaders or skills. Finally the organization cultivated a wide international network of foot soldiers but they were ill-trained as compared to their predecessors. This description closely matches the profiles in Fig. 2.1 where r-type organizations start from a small well-trained core, move toward a smaller ratio of leaders to foot soldiers and then grow through recruitment.

As was noted, nascent organizations tend towards the trend line, regardless of how they started (Fig. 2.1). The slope of this line is $\frac{r + \sqrt{r^2 + 4rmp}}{2p}$, and this number is the long-term ratio between the number of foot soldiers and the number of leaders. Notice that this formula implies that ratio is dependent on just the parameters of growth - r, m, p - and does not depend on either d or the CT measures k, b . This ratio is generally not found in failing organizations, but is predicted to be ubiquitous in successful organizations. It may be possible to estimate it by capturing a division of an organization and it can help calculate the model’s parameters. However, it is important to note that L includes

not just commanding officers, but also any individuals with substantially superior skills and experience. The existence of the ratio is a prediction of the model, and if the other parameters are known, it could be compared to empirical findings.

2.4.2 Conditions for Victory

Recall, that the model indicates that all terrorist organizations belong to one of three classes: r-types, p-types and organizations that will be defeated. Each class exhibits characteristic changes in its leaders, foot soldiers and strength (L , F and S resp.) over time. This makes it possible to determine whether any given organization belongs to the third class, i. e., to predict whether it would be defeated.

One finding is that if a terrorist organization weakens, i. e. shows a decline in its strength S , it does not follow that it would be defeated. Indeed, in some r-type organizations it is possible to observe a temporary weakening of the organization and yet unless counter-terrorism (CT) measures are increased, the organization would recover and grow out of control (see Fig. 2.2(b)). Even a decline in the leadership is not by itself sufficient to guarantee victory. The underlying reason for this effect is out-of-control growth in F , which would ultimately create a new generation of terrorist leaders. Similarly, it is possible for an organization to experience a decline in its pool of foot soldiers and yet recover. These cases indicate that it is easy during a CT campaign to incorrectly perceive apparent progress in reducing the organization as a sign of imminent victory.

Fortunately, under the model it is possible to identify reliable conditions for victory over the organization (see the appendix for the proof):

1. *For a p-type organization, it is impossible to have a decline in strength S . If such*

a decline is made to happen, the organization would be defeated.

2. *For an r-type organization, it is impossible to have a decline in foot soldiers F . If such a decline is made to happen, the organization would be defeated.*

Consequently:

A terrorist organization would collapse if counter-terrorism measures produce both: (1) a decline in its strength S and (2) a decline in its foot soldiers F .

In a notable contrast, declines in strength and the *leaders* are *not* sufficient in all cases (see Fig. 2.2(b)). To apply the theorem to an organization of an unknown type, one needs merely to estimate whether the organization's pool of foot soldiers and strength are declining. The latter could be found indirectly by looking at the quantity and quality of terrorist operations. It is not necessary to know the model's parameters or changes in the pool of leaders - the latter could even be increasing. Furthermore, while it may take some time to determine whether S and F are indeed declining, this time could be much shorter compared to the lifetime of the organization. Therefore, the theorem suggests the following two-step approach:

1. Estimate the scale of CT measures believed to be necessary to defeat the organization.
2. Measure the effect on S and F . If they both declined, then sustain the scale of operations (i. e. do not reduce b, k); Otherwise an increase in CT measures would be necessary.

The theorem and findings above give sufficient conditions for victory but they do not characterize the only possible victory scenario. For example, it is possible for an organization to see an increase in its pool of foot soldiers F yet ultimately collapse: these are organizations that lie to the right of the trend line and just slightly under the sink line. More generally, it should be remembered that to prove the theorem it was necessary to use a simplified model of a terrorist organization, as described in section 2.2. Nevertheless, it is likely that some form of the theorem would remain valid in complicated models because the model is built on fundamental forces that are likely to be retained in these models.

2.4.3 Stable Equilibria

Recall that the model does not have a stable equilibrium (Fig. 2.3). Yet, in many practical cases, terrorist organization seem to reach a stable equilibrium in terms of their structure and capabilities. It is plausible that such stability is the result of a dynamic balance between the growing terrorist organization and increasing CT measures directed against it. Indeed, rather than staying constant numbers like b, k , CT may actually grow when the organization presents more of a threat⁶. Aside from CT, stability may be the result of organizations reaching an external limit on their growth - a limit imposed by constraints such as funding, training facilities or availability of recruits. The case of funding could be modeled by assuming that the growth of the organization slows as the organization approaches a maximum point, (L_{max}, F_{max}) . Alternatively, it is quite possible and consistent with the model that there would be a perception of stasis because the organization is changing only slowly.

⁶It would be a straightforward task to modify the model to incorporate such a control-theoretic interaction, but the task is more properly the subject of a follow-up study.

2.5 Counter-Terrorism Strategies

Recall that the general counter-terrorism (CT) strategy in this model is based on the location of the sink line, which we want to place above the terrorist organization (in Fig. 2.1). To implement this strategy, it is necessary first to calculate the model's parameters for a given organization (p, r, m, d) , and second, to determine the efficacy of the current counter-terrorism measures (b, k) . Then, it remains "just" to find the most efficient way of changing those parameters so as to move the sink line into the desired location. Let us now consider several strategic options.

2.5.1 Targeting the leaders

An important "counter terrorist dilemma" [19] is whom to target primarily - the leaders or the foot soldiers. Foot soldiers are an inviting target: not only do they do the vital grunt work of terrorism, they also form the pool of potential leaders, and thus their elimination does quiet but important damage to the future of the organization. Moreover, in subsection 2.4.1 we saw that while an organization can recover from a decline in both its strength and leadership pool, it cannot recover from declines in both its strength and its foot soldiers pool. That finding does not say that attacking leaders is unlikely to bring victory - indeed, they form an important part of the organization's overall strength, but it does suggest that a sustained campaign against an organization is more likely to be successful when it includes an offensive against its low-level personnel. Yet, it seems that the neutralization of a terrorist leader would be more profitable since the leader is more valuable to the organization than a foot soldier, and his or her loss would inevitably result in command and control difficulties that may even disrupt terrorist attacks.

When we use the model to address the problem quantitatively, we find that the optimal strategy is actually dependent upon the organization, that is to say the parameters p, d, r, m (but not on b, k). For example, for the parameter values used in the figures above, an increase in b gives a greater rise in the sink line than an equal increase in k . Specifically, for those parameter values every two units of b are equivalent to about ten units of k . In general, when m, r, d are high but p is low then attack on the leadership is favored, while attack on the foot soldiers is best when p is high but m, r, d are low - in agreement with intuition⁷. In the first parameter range, foot soldiers are recruited so rapidly that attacking them is futile, while in the second set leaders are produced quickly so the only strategy is to fight the foot soldiers to prevent them from becoming leaders. In any case, policy prescriptions of this kind must be applied with consideration of counter-terrorism capabilities and policy costs. Thus, while on paper a particular strategy is better, the other strategy could be more feasible.

It is often argued that counter-terrorism policies have considerable side effects. For instance, there is evidence that targeted assassinations of leaders have led terrorist organizations to escalate, in what has been called the “boomerang effect” [29, p.125]. Fortunately, the model suggests that the policy maker has useful substitutes, with possibly fewer policy side effects. As Fig. 2.5 shows, making recruitment (r) lower has an effect similar to increasing k . Likewise, decreasing the rate of promotion to leadership (p) can substitute for increasing b . This agrees with intuition: for example, in the case of the foot soldiers, growth can be contained either actively through e. g. arrests or proactively by slowing the recruitment of new operatives (through e. g. attacks on recruitment facilities or advocacy).

⁷Mathematically to obtain this result we first compute the derivatives of the fixed point with respect to both b and k , then project them to the orthogonal to the sink link and then use an optimization solver to find the parameter values of the model which maximize (and minimize) the ratio between the lengths of the projections.

2.5.2 Encouraging desertion

Fatigue and attrition of personnel have been empirically found to be an important effect in the evolution of terrorist organizations. In interviews with captured or retired terrorists, they often complained about the psychological stress of their past work, its moral contradictions, and the isolation from relatives and friends [24, Ch.6]. This is part of the reason why terrorist organizations cannot remain inactive (as in a cease fire) for very long without experiencing irreplaceable loss of personnel due to loss of motivation, and many organizations even resort to coercion against desertion. Therefore, encouraging operatives to leave through advocacy or amnesties may be an effective counter-terrorism strategy.

The model introduced here brings theoretical insight into this phenomenon. One prediction of the model is that even if such desertion exceeds recruitment (i. e. $d > r$) the organization would still sustain itself as long as it has a sufficiently large rate of promotion (p) or leaders of sufficiently high caliber (m). However, if d is even greater, namely, exceeds $d = \frac{1}{2}(r + r \sqrt{1 + \frac{mp}{r}})$, then the model predicts that the organization would be destroyed regardless of starting conditions, or counter-terrorism efforts (b, k).

Organizations with lower d are, of course, also effected by desertion. Earlier, in Fig. 2.5 we saw how increasing d raises up the sink line. To see the phenomenon in more detail, we replaced d by two (not necessarily equal) parameters d_L and d_F for the desertion of L and F , respectively. The two parameters change the slope of the sink line: increasing d_L flattens it, while increasing d_F makes it more steep (Fig. 2.6). Therefore, increasing d_L could be a particularly effective strategy against nascent r-type organizations, while increasing d_F could be effective against the nascent p-types.

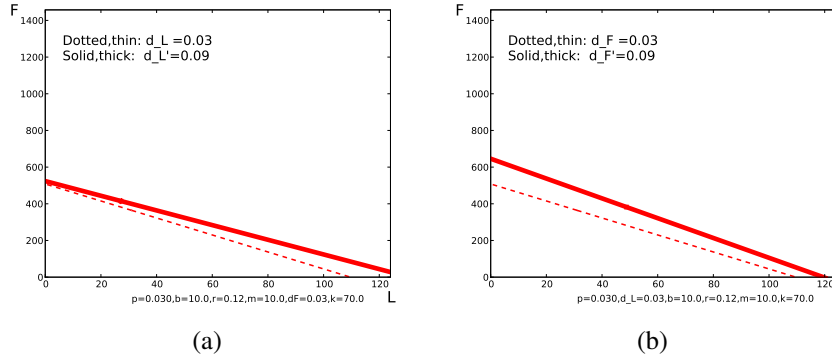


Figure 2.6: The effects of d_L (a) and d_F (b) on the dynamical system, as seen through the effect on the sink line. As the desertion rates increase, the sink line moves up and its slope changes, thus trapping additional terrorist organizations.

2.5.3 Minimization of Strength S

Counter terrorism (CT) is often the problem of resource allocation among competing strategies. Therefore, suppose that resources have become available towards a CT operation against the terrorist organization. Namely, suppose we can remove leaders and operatives in a single blow (unlike the parameters b, k in the model which take a gradual toll). A reasonable approach to allocating these resources efficiently would be to divide them between operations targeting the leadership and those targeting the foot soldiers in such a way that the terrorist organization's strength S is minimized⁸. However, by some simplified economic analysis, it is possible to show (see appendix) that this counter-terrorism strategy is in general *suboptimal*. Instead, for a truly effective resource allocation, it is necessary to consider the dynamics of the organization being targeted and the true optimum may be considerably different. For example, when the ratio of promotion to recruitment is relatively large (i. e. $\frac{p}{r} \gg 0$), then the optimum shifts increasingly towards attacking the foot soldiers since they become much harder to

⁸Mathematically, this would be two variable minimization of S constrained by a budget.

replace than leaders.

On an intuitive level, the reason why the strategy is suboptimal is because often, the losses we can inflict most effectively on the organization are precisely those losses that the organization can restore most easily. Hence, in the long-term a strategy targeting strength S would be ineffective. Instead, when making a CT strategy it would be valuable to understand the target organization's dynamics, and in particular, to build a dynamical model. Such a model would help because it can identify an organization's unique set of vulnerabilities due to the composition of its human capital and its properties as a dynamical system.

2.6 Conclusions

Much of the benefit of mathematical models is due to their ability to elucidate the logical implications of empirical knowledge that was used to construct the model. Thus, whereas the empirical facts used to construct the models should be uncontroversial, their conclusions should offer new insights. The model proposed here is a very simplified description of real terrorist organizations. Despite its simplicity, it leads to many plausible predictions and policy recommendations. Indeed, the simplicity of the model is crucial to making the model useful. More detailed models of this kind could provide unparalleled insights into counter-terrorism policies and the dynamics of terrorism.

APPENDIX

2.A Appendix

2.A.1 The Dynamical System

The original differential equations are:

$$\frac{dL}{dt} = pF - dL - b \quad (2.1)$$

$$\frac{dF}{dt} = r(mL + F) - dF - k \quad (2.2)$$

If we wished to incorporate the drain of the foot soldiers due to promotion ($-pF$) in Eqn.2.2, then we could adjust the original parameters by the transformation $r \rightarrow r - p$ and $m \rightarrow \frac{rm}{r-p}$. However, this would affect some of the analysis below, because for $r < p$ it would not longer be the case that $r > 0$, even though $rm > 0$ would still hold true. Alternatively, we could change the internal losses parameter for foot soldiers : $d_F \rightarrow d_F + p$ and break the condition $d_F = d_L$.

The linearity of the system of differential equations makes it possible to analyze the solutions in great detail by purely analytic means. The fixed point is at:

$$L_* = \frac{kp - b(r-d)}{d(r-d) + rmp} \quad F_* = \frac{kd + rmb}{d(r-d) + rmp} \quad (2.3)$$

The eigenvalues at the fixed point are

$$\lambda_{1,2} = \frac{r - 2d \pm \sqrt{(r - 2d)^2 + 4(rmp + d(r - d))}}{2} \quad (2.4)$$

From Eqn.(2.4), the fixed point is a saddle when $rmp + d(r - d) > 0$, i. e. $\frac{r - \sqrt{r^2 + 4rmp}}{2} < d < \frac{r + \sqrt{r^2 + 4rmp}}{2}$ (physically, the lower bound on d is 0). The saddle becomes a sink if

$r < 2d$ and $rpm + d(r - d) < 0$. By Eqn.(2.3), this automatically gives $F_* < 0$, i. e. the organization is destroyed⁹. It is impossible to obtain either a source because it requires $r - 2d > 0$ and $rpm + d(r - d) < 0$, but the latter implies $d > r$, and so $r - 2d > 0$ is impossible; or any type of spiral because $(r - 2d)^2 + 4(rpm + d(r - d)) < 0$ is algebraically impossible¹⁰. It is also interesting to find the eigenvectors because they give the directions of the sink and trend lines:

$$e_{1,2} = \begin{pmatrix} 2p \\ r \pm \sqrt{r^2 + 4rpm} \end{pmatrix} \quad (2.5)$$

We see that the slope of e_2 , which is also the slope of the sink line - the stable manifold - is negative. Therefore, we conclude that the stable manifold encloses, together with the axes, the region of neutralized organizations. Concurrently, the slope of e_1 - the trend line i. e. the unstable manifold - is positive. Thus, the top half of the stable separatrix would point away from the axes, and gives the growth trend of all non-neutralized organizations ($\frac{\Delta F}{\Delta L} = \frac{r + \sqrt{r^2 + 4rpm}}{2p}$).

2.A.2 Proof of the theorem

Recall, we wish to show that a terrorist organization that experiences both a decline in its strength and a decline in the number of its foot soldiers will be destroyed. The proof rests on two claims: First, a p-type organization cannot experience a decline in strength, and second, an r-type organization cannot experience a decrease in F (for a graphic illustration see Fig. 2.7). Thus, both a decline in strength and a decline in the number of foot soldiers cannot both occur in an r-type organization nor can they both occur in

⁹Of course, the dynamical system is unrealistic once either F or L fall through zero. However, by the logic of the model, once F reaches zero, the organization is doomed because it lacks a pool of foot soldiers from which to rebuild inevitable losses in its leaders.

¹⁰The degenerate case of $\lambda = 0$ has probability zero, and is not discussed.

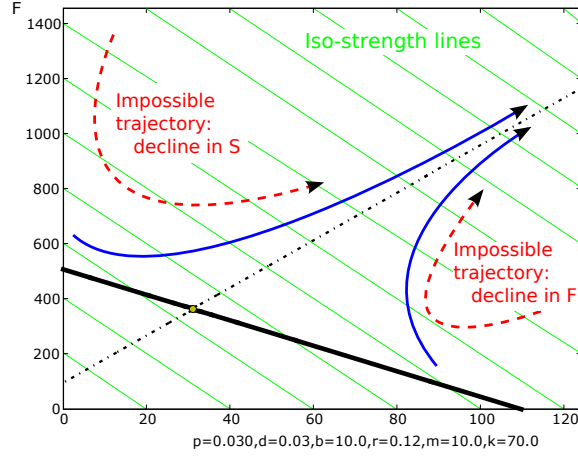


Figure 2.7: The phase plane with possible (solid blue) and impossible (dashed red) trajectories, and lines of equal organization strength (green). Because orbits of the p-type must experience an increase in strength S , the left red line cannot be an orbit. Also, r-type orbits must experience an increase in F , and so the right red line cannot be an orbit either.

a p-type organization. Hence, such a situation can only occur in the region of defeated organizations.

As to the first claim, we begin by showing that the slope of the sink line is always greater than the slope of the iso-strength lines ($= -m$). By Eqn. (2.5) the slope is $\frac{r - \sqrt{r^2 + 4rmp}}{2p} = -m \frac{2}{1 + \sqrt{1 + \frac{mp}{r}}} > -m$. Therefore, the flow *down* the sink line has $\frac{dS}{dt} > 0$ (Down is the left-to-right flow in the figure). Now, we will show that in a p-type organization, the flow must experience an even greater increase in strength. Let A be the matrix of the dynamical system about the equilibrium point and let the state of the terrorist organization be $(L, F) = d_1 e_1 + d_2 e_2$ where e_1, e_2 are the distinct eigenvectors corresponding to the eigenvalues λ_1, λ_2 . Consideration of the directions of the vectors (Eqn.(2.5)) shows that for a p-type organization, $d_1 > 0$ and $d_2 < 0$. The direction of flow is therefore $d_1 \lambda_1 e_1 + d_2 \lambda_2 e_2$. Notice that $\lambda_1 > 0, \lambda_2 < 0$, and so the flow has a positive component ($= d_1 \lambda_1$) in the e_1 direction (i. e. up the trend line). Since the flow

along e_1 experiences an increase in both L and F , it must experience an increase in strength. Consequently, a p-type organization must have $\frac{dS}{dt}$ which is even more positive than the flow along the sink line (where $d_1 = 0$). Thus, $\frac{dS}{dt} > 0$ for p-types.

As to the second claim, note that r-type organizations have $d_1 > 0$ and $d_2 > 0$. Moreover, in an r-type organization, the flow $d_1\lambda_1e_1 + d_2\lambda_2e_2$ has $\frac{dF}{dt}$ greater than for the flow *up* the right side of the sink line (right-to-left in the figure): the reason is that e_1 points in the direction of increasing F and while in an r-type $d_1 > 0$, along the sink line $d_1 = 0$. The flow up the sink line has $\frac{dF}{dt} > 0$, and so $\frac{dF}{dt} > 0$ in an r-type organization. In sum, $\frac{dS}{dt} < 0$ simultaneously with $\frac{dF}{dt} < 0$ can only occur in the region $d_1 < 0$ - the region of defeated organizations. QED.

2.A.3 Concrete Example of Strength Minimization

In subsection 2.5.3 we claimed that the task of minimizing S is different from the optimal counter-terrorism strategy. Here is a concrete example that quantitatively illustrates this point. Suppose a resource budget B is to be allocated between fighting the leadership and fighting the foot soldiers, and furthermore, that the cost of removing l leaders and f foot soldiers, respectively, is a typical convex function: $c_1l^\sigma + c_2f^\sigma$ (c_1 and c_2 are some positive constants and $\sigma > 1$)¹¹. Notice that whereas uppercase letters L, F indicate the number of leaders and foot soldiers, respectively, we use lowercase l, f to indicate the number to be *removed*. The optimal values of l and f can be easily found graphically using the standard procedure in constrained optimization: the optimum is the point of tangency between the curve $B = c_1l^\sigma + c_2f^\sigma$ and the lines of constant difference in strength: $\Delta S = ml + f = \text{constant}$ (Fig.2.8(a)). However, as illustrated in Fig. 2.8(b), if

¹¹ $\sigma > 1$ because e. g. once the first say 20 easy targets are neutralized, it becomes harder to find and neutralize the next 20 (the law of diminishing returns.) In any case the discussion makes clear that for most cost functions the suggested optimum would be different from the true optimum.

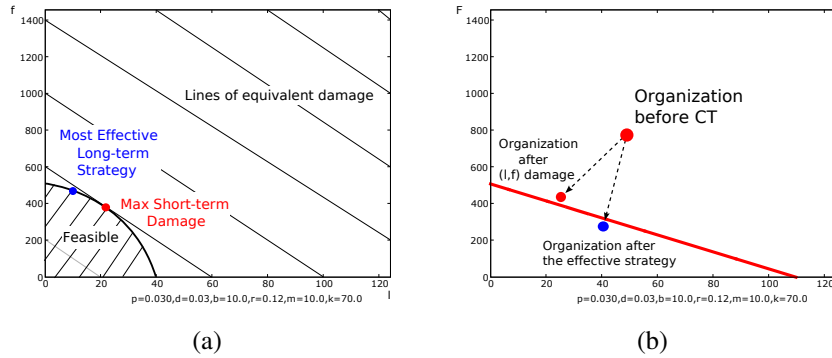


Figure 2.8: Graphical calculation of optimal budget allocation (a) and contrast between minimization of S and the actual optimum (b). In (a), the optimal choice of (l, f) is given by the point of tangency between the feasible region and the lines of constant S . In (b), the red line is the sink line. The minimization of S through the removal of about 20 leaders and 400 foot soldiers would not bring the organization under the sink line, but a different (still feasible) strategy would.

such a strategy is followed, the terrorist organization may still remain out of control. It is preferable to choose a different strategy - in the example it is the strategy that focuses more on attacking the foot soldiers and thus brings the organization under the sink line (red line), even though the ΔS is not as large. In general, the difference between the strategies is represented by the difference between the slope of the sink line and the slopes of the lines of equivalent damage to strength. The latter always have slope $-m$ while the former becomes arbitrarily flat as $\frac{p}{r} \rightarrow \infty$.

REFERENCES

- [1] R. Aris, *Mathematical Modelling Techniques*. Mineola, NY: Dover Publications, 1995.
- [2] S. P. Ellner and J. Guckenheimer, *Dynamic Models in Biology*. Princeton, NJ: Princeton University Press, 2006.
- [3] J. Allanach, H. Tu, S. Singh, P. Willett, and K. Pattipati, “Detecting, tracking, and counteracting terrorist networks via Hidden Markov Models,” in *IEEE Aerospace Conference Proceedings* (R. Profet, D. Woerner, and R. Wright, eds.), pp. 3246–3257, Mar 2004.
- [4] T. Chamberlain, “Systems dynamics model of Al-Qa’ida and United States ”Competition”,” *J. Homeland Security and Emergency Management*, vol. 4, no. 3:14, pp. 1–23, 2007.
- [5] J. D. Farley, “Evolutionary dynamics of the insurgency in Iraq: A mathematical model of the battle for hearts and minds,” *Studies in Conflict and Terrorism*, vol. 30, pp. 947–962, 2007.
- [6] G. Feichtinger, R. F. Hartl, P. M. Kort, and A. J. Novak, “Terrorism control in the tourism industry,” *J. Optimization Theory and Applications*, vol. 108, pp. 283–296, Feb 2001.
- [7] N. F. Johnson, M. Spagat, J. A. Restrepo, O. Becerra, J. C. Bohrquez, N. Surez, E. M. Restrepo, and R. Zarama, “Universal patterns underlying ongoing wars and terrorism,” *arxiv.org*, vol. physics/0506213, 2006.
- [8] D. Stauffer and M. Sahimi, “Discrete simulation of the dynamics of spread of extreme opinions in a society,” *Physica A*, vol. 364, pp. 537–543, Mar 2006.
- [9] F. Udwardia, G. Leitmann, and L. Lambertini, “A dynamical model of terrorism,” *Discrete Dynamics in Nature and Society*, vol. 2006, May 2006. 32 pages.
- [10] C. H. Anderton and J. R. Carter, “On rational choice theory and the study of terrorism,” *Defence and Peace Economics*, vol. 16, pp. 275–282, Aug 2005.
- [11] T. Sandler, J. T. Tschirhart, and J. Cauley, “A theoretical analysis of transnational terrorism,” *American Political Science Rev.*, vol. 77, pp. 36–54, Mar 1983.

- [12] T. Sandler, "Collective action and transnational terrorism," *The World Economy*, vol. 26, pp. 779–802, Jun 2003.
- [13] R. Wintrobe, *Rational Extremism: The Political Economy of Radicalism*. Cambridge, UK: Cambridge University Press, 2006.
- [14] E. P. MacKerrow, "Understanding why - dissecting radical Islamic terrorism with agent-based simulation," *Los Alamos Science*, vol. 28, pp. 184–191, 2003.
- [15] M. Tsvetovat and K. M. Carley, "On effectiveness of wiretap programs in mapping social networks," *Computational Mathematical Organization Theory*, vol. 13, pp. 63–87, Mar 2007.
- [16] B. Hoffman, "Al-Qaeda, trends in terrorism, and future potentialities: An assessment," *Studies in Conflict and Terrorism*, vol. 26, pp. 429–442, Nov-Dec 2003.
- [17] M. Sageman, *Understanding Terror Networks*. University of Pennsylvania Press, PA, Apr 2004.
- [18] J. B. Wolf, *Antiterrorist Initiatives*. New York: Plenum Press, 1989.
- [19] B. Ganor, *The Counter-Terrorism Puzzle: A Guide for Decision Makers*. Piscataway, NJ: Transaction Publishers, 2005.
- [20] W. Laqueur, *A History of Terrorism*. Piscataway, NJ: Transaction Publishers, Aug 2001.
- [21] C. C. Harmon, *Terrorism Today*. Oxford, UK: Routledge, 1 ed., 2000.
- [22] B. Hoffman, *Inside Terrorism*. Columbia University Press, USA, 2006.
- [23] A. J. Jongman and A. P. Schmid, *Political Terrorism*. Piscataway, NJ: Transaction Publishers, 2005.
- [24] J. Horgan, *The Psychology of Terrorism*. New York, NY: Routledge, Sep 2005.
- [25] B. Ganor, "Terrorist organization typologies and the probability of a boomerang effect," *Studies in Conflict and Terrorism*, vol. 31, pp. 269–283, Mar 2008.
- [26] M. A. Hanson and M. B. Schmidt, "The Impact of Coalition Offensive Operations on the Iraqi Insurgency," *SSRN eLibrary*, 2007. <http://ideas.repec.org/p/cwm/wpaper/56.html>.

- [27] E. Manningham-Buller, "The international terrorist threat to the UK," *MI5 website*, November 2006. <http://www.mi5.gov.uk/output/the-international-terrorist-threat-to-the-uk-1.html>.
- [28] L. Wright, *The Looming Tower*. London: Allen Lane, 2006.
- [29] M. Crenshaw, "The counter-terrorism and terrorism dynamic," in *Terrorism and the 2000 Olympics* (A. Thompson, ed.), p. 125, Defence Studies Center, Canberra, Australia, 1996.

CHAPTER 3

INTERDICTION OF A MARKOVIAN EVADER

Chapter Abstract Network interdiction is a combinatorial optimization problem on an activity network arising in a number of important security-related applications. It is classically formulated as a bilevel maximin problem representing an “interdictor” and an “evader”. The evader tries to move from a source node to the target node along the shortest or safest path while the interdictor attempts to frustrate this motion by cutting edges or nodes. The interdiction objective is to find the optimal set of edges to cut given that there is a finite interdiction budget and the interdictor must move first. We reformulate the interdiction problem for stochastic evaders by introducing a model in which the evader follows a Markovian random walk guided by the least-cost path to the target. This model can represent incomplete knowledge about the evader and the graph as well as partial interdiction. We formulate the optimization problem for this model and show how, by exploiting topological ordering of the nodes, one can achieve an order-of-magnitude speedup in computing the objective function over a naive algorithm. We also introduce optimization heuristics based on betweenness centrality. These can rapidly find high-quality interdiction solutions by providing a global view of the network.^{1 2}

3.1 Introduction

Mathematical modeling of network interdiction originated in the study of military supply chains and interdiction of transportation networks [1, 2]. The problem is currently studied in different classes of networks and in a variety of contexts, and finds applications in countering of nuclear proliferation programs [3], control of infectious dis-

¹*This chapter is released under Los Alamos National Laboratory LA-UR-08-06551*

²Joint work with Aric Hagberg, David Izraelevitz and Feng Pan - Los Alamos National Laboratory

eases [27], and disruption of terrorist networks [5]. The underlying networks may represent transportation networks, but more generally may be social or activity networks. Recent interest in the problem has been in part due to the threat of smuggling of nuclear materials and devices [6]. In the case of nuclear smuggling, interdiction might correspond to the installation of special radiation-sensitive detectors along the selected transportation edges.

The problem is often posed in terms of two agents called “interdictor” and “evader” where the evader attempts to minimize some objective function in the network, *e.g.* distance, cost, or risk when traveling from network location s to location t , while the interdictor attempts to limit success by removing network nodes or edges. The interdictor has limited resources and can thus only remove a finite set of nodes or edges. In the simplest formulation, the interdictor seeks to identify a set of edges (or nodes) on the network whose removal maximizes the cost of the least-cost path from a source to a destination node, while the evader seeks to find and traverse the best unimpeded path. This interdiction problem is known as the “most vital edges” (or “most vital nodes”) problem [7] and it has been shown to be NP-hard [8] and NP-hard to approximate to better than a factor of 2 [9]. Methods for solving network interdiction problems have included exact algorithms for solving integer programs, such as branch-and-bound, as well as decomposition methods to rebuild the network by iteratively adding relevant paths to reduce the size of both the underlying network and the number of binary decision variables. A more recent approach, based on structure-dependent cutting planes, exploits the relationship between the ordered set of evading paths and binary interdiction variables [10].

A common assumption in many studies is that there is perfect knowledge of hard-to-compute network parameters, such as the cost to the evader to traverse an edge in terms

of resource consumption or probability of detection. However, it is clear that the evader, and, to a lesser extent, the interdictor, have unreliable and incomplete information about the network. These uncertainties place the interdiction problem within stochastic optimization, where one seeks to find those edges that are vital *on average*. Indeed, under uncertainty the evader must be described in probabilistic terms. By constructing such probabilistic evader models one can expect to develop more robust interdiction solutions. This problem of stochastic interdiction has been the focus of a number of recent studies [3, 11, 12, 13, 14].

Failure to account for evader uncertainty can lead to suboptimal decisions, namely, solutions that do not maximize (and even decrease) the evader’s expected cost to reach the target. Consider for instance the network in Fig. 3.1. There are four paths from the source to the target: one each through nodes 1, 2, 3 and the one direct path (0, 5) with costs 9, 8, 8 and 8.01, respectively. If only one edge can be removed, the solution in the least-path-cost formulation is to remove edge (4, 5) which increases the path cost from 8.0 to 8.01. However if the evader is unable to determine which path has the least cost and takes any path with equal (or nearly equal) probability, then this solution is not optimal. Interdiction at (4, 5) actually *decreases* the expected cost from ≈ 8.25 to 8.01, because it removes the costly path through node 1. The optimal choice is interdiction of any one of the edges (0, 2), (2, 4), (0, 3), or (3, 4), which increases the expected cost from ≈ 8.25 to ≈ 8.33 .

In this paper we propose a Markovian network interdiction framework which can capture a wide range of network evader behavior (Sec. 3.2). We then demonstrate the general framework with a simple model based on low-level evader decision-making processes (Sec. 3.3). Finally we develop efficient heuristic algorithms for the interdiction problem based on edge betweenness centrality and predicted evader motion, and then

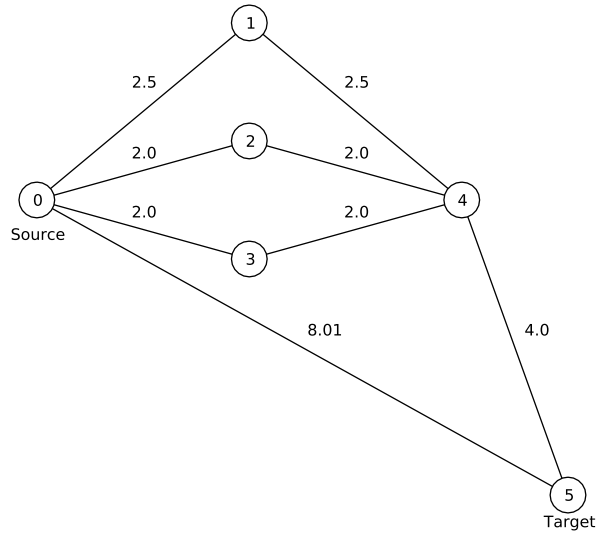


Figure 3.1: Example network where the least-cost path interdiction formulation produces a suboptimal solution $((4, 5))$. Interdicting any one of $(0, 2)$, $(2, 4)$, $(0, 3)$, or $(3, 4)$ is the true optimum.

present performance results comparing various heuristic methods (Sec. 3.4).

3.2 The interdiction model

Our interdiction formulation is a stochastic generalization of the max-min shortest path interdiction problem [1, 2]. In the max-min formulation an evader attempts to traverse a network on a path from an origin s to a destination t . Let p be some path between s and t in a graph $G(N, E)$ with the set of nodes N and the set of weighted edges E . Let $c(p)$ be the path cost computed by summing the cost C_{ij} over the edges (i, j) of p , where C_{ij}

may include costs due to interdiction.³ The edge costs are assumed to be given in the problem and may depend on direction (in the case that $G(N, E)$ is a directed graph). In the following “edge cost” is used here interchangeably with “edge weight”.

The network interdiction strategy is represented by choosing edges to interdict from the feasible interdiction set R which is typically a subset of the edge set E with a limited size (budget) B . We set the value of $r_{ij} = 1$ if edge (i, j) is interdicted, and $r_{ij} = 0$ otherwise. Let $D_{ij} \geq 0$ be the added cost of traversing (i, j) when it is interdicted. When the value of D_{ij} is very large all paths avoid the interdicted edge (i, j) (assuming that there is an alternative path) which effectively removes the edge (i, j) from the graph. One may write $C'_{ij} = C_{ij} + r_{ij}D_{ij}$ but it is more convenient to “drop the primes” that is, to use C_{ij} at all times to denote cost that includes possible interdiction. This makes the matrix \mathbf{C} a function of r .

In the shortest-path model, the evader only travels on least-cost paths, and is fully aware of interdiction decisions. This gives the optimization problem

$$\max_{r \in R} \min_{p \in PT} c(p), \quad (3.0)$$

where $c(p)$ is implicitly a function of r . The above formulation is for interdiction of edges but of course, a similar problem could be considered for node interdiction by introducing for all $i \in N$ node costs D_i and decision variables on nodes r_i etc.

A stochastic version of the interdiction problem can be constructed by supposing that an evader may take any path from s to t , according to some probability distribution, rather than always choosing the least-cost path. Randomness in the evader path decision is caused by uncertainty about interdiction decisions r or network costs, mistaken

³The additivity of costs is natural for resources such as money or time, but it also holds for simple models of risk. The probability of evasion, $q(p)$ on p could be represented through the cost $c(p)$ by setting $c(p) = -\log q(p)$ and similarly for each edge in the path. If the probabilities of evasion on all edges of p are independent events, then $q(p)$ is just the product of their probabilities, or equivalently the exponential of the sum of the edge costs: $q(p) = \exp(-\sum_{(i,j) \in p} C_{ij})$.

cost computations, or possibly even by intent to increase unpredictability. The path p becomes a random variable distributed as $P(p)$, and the expected cost of traveling from s to t is then

$$E[c] = \sum_{p \in PT} P(p)c(p). \quad (3.0)$$

The interdiction problem becomes

$$\max_{r \in R} \sum_{p \in PT} P(p|r)c_r(p), \quad (3.0)$$

where $P(p|r)$ is now the probability of traversing a path given the interdiction set r . The conditional probability $P(p|r)$ implicitly contains the evader's strategy. The shortest-path optimization problem (3.2) is clearly just a special instance of (3.2) when the expectation is conditioned on traversal of only least-cost paths.

3.2.1 Markovian evaders

In order to compute $P(p|r)$ values and $E[c]$ it is necessary to develop stochastic evader models. Consider for simplicity a Markov random walk for the evader model where the evader chooses the next step without memory of the previous steps. In general an evader might follow a non-Markovian random walk but a Markovian model achieves a good balance between realism and computational efficiency. In the Markovian model the expected cost $E[c]$ can be found simply for all possible starting locations (as described below) while for the non-Markovian evader such a computation would likely involve computing $c(p)$ and $P(p|r)$ for each possible path p . Since the number of possible paths between two nodes can be enormous, feasible computations of $E[c]$ would require simplifying the set of allowed paths, which would reduce solution realism.

Complete information about a Markovian evader is encoded in a distribution of starting nodes, \mathbf{a} , and a Markovian transition probability matrix, \mathbf{M} . An element M_{ij} of this

matrix is the probability that an evader at node i will move along edge (i, j) . The distribution of starting nodes is assumed to be given and independent of the interdiction strategy r , while the \mathbf{M} matrix is assumed to be determined as soon as the graph and r are known. It is convenient to set the single entry $M_{tt} = 0$. Then \mathbf{M} can be interpreted as an evader model where the evader is removed from the network when reaching the target t . We also assume that the cost of traversing edges is given by a matrix \mathbf{C} .

Armed with $\mathbf{a}, \mathbf{M}, \mathbf{C}$ the objective $E[c]$ can be computed by writing and solving a recurrence relation for the the expected cost of paths that reach t as a function of path length.⁴ Suppose a path p from s to t is specified by the edge sequence $(s, x_1), (x_1, x_2), \dots, (x_l, t)$. The conditional probability that the evader will traverse this path is $M_{sx_1}M_{x_1x_2}\dots M_{x_l t}$. The cost of this path is

$$c(p) = C_{sx_1} + C_{x_1x_2} + \dots + C_{x_l t}, \quad (3.0)$$

where C_{ij} includes the cost of passing an interdicted edge if (i, j) is interdicted. Let $\pi^{(n)}$ be the probability vector whose j coordinate is the probability that a path of length n begins at s and ends at j . Thus, $\pi^{(n)}$ is the sum of the probabilities of all paths of length n that end at j . Since \mathbf{M} defines a Markov chain, $\pi^{(n)} = \pi^{(n-1)}\mathbf{M} = \pi^{(0)}\mathbf{M}^n = \mathbf{a}\mathbf{M}^n$, where \mathbf{a} is the distribution over the starting nodes. If all paths must begin at s , then \mathbf{a} is just the unit vector in the s direction.

Let the vector $\mathbf{h}^{(n)}$ represent the expected cost for paths of length n that terminate at each node. Namely the j coordinate $h_j^{(n)}$ is the expected cost of a path with length n that terminates at j . The vector satisfies the recurrence

$$h_j^{(n)} = \sum_i h_i^{(n-1)}M_{ij} + \pi_i^{(n-1)}M_{ij}C_{ij}. \quad (3.0)$$

⁴The length of a path p is the number of edges in the path, while the cost $c(p)$ of a path is the sum of the costs of the edges. These are equal only if the cost of each edge is unity.

The entire vector can be written as

$$\mathbf{h}^{(n)} = \mathbf{h}^{(n-1)}\mathbf{M} + \boldsymbol{\pi}^{(n-1)}(\mathbf{M} \odot \mathbf{C}), \quad (3.0)$$

where $\mathbf{M} \odot \mathbf{C}$ is the matrix formed by element-wise multiplication of \mathbf{M} and \mathbf{C} . The expected cost of paths of any length is given by

$$\mathbf{h} := \sum_{n=0}^{\infty} \mathbf{h}^{(n)}. \quad (3.0)$$

Using the relation

$$\sum_{n=0}^{\infty} \boldsymbol{\pi}^{(n)} = \sum_{n=0}^{\infty} \boldsymbol{\pi}^{(0)}\mathbf{M}^n = \mathbf{a}(\mathbf{I} - \mathbf{M})^{-1}, \quad (3.0)$$

and summing Eq. (3.2.1) over all n gives

$$\mathbf{h} = \mathbf{a}(\mathbf{I} - \mathbf{M})^{-1}(\mathbf{M} \odot \mathbf{C})(\mathbf{I} - \mathbf{M})^{-1}, \quad (3.0)$$

where \mathbf{I} is the identity matrix.

Equation (3.2.1) is key to our approach: the vector element, h_t , expresses in closed form the expected cost of paths starting at s and ending at t . Each part of Eq. (3.2.1) has an intuitive meaning. The vector $\mathbf{a}(\mathbf{I} - \mathbf{M})^{-1}$ is the expected number of times that each of the nodes is visited by the evader when starting at a distribution \mathbf{a} [11, p.419]; the vector $\mathbf{a}(\mathbf{I} - \mathbf{M})^{-1}(\mathbf{M} \odot \mathbf{C})$ is the expected cost of reaching each of the nodes from their immediate predecessor nodes; and \mathbf{h} gives the expected cost of reaching each of the nodes from the starting distribution \mathbf{a} .

The interdiction objective is to maximize h_t . Because the interdiction variable r affects the costs and then the matrix \mathbf{M} , this results in the nonlinear optimization problem

$$\max_{r \in R} \left[\mathbf{a}(\mathbf{I} - \mathbf{M})^{-1}(\mathbf{M} \odot \mathbf{C})(\mathbf{I} - \mathbf{M})^{-1} \right]_t. \quad (3.0)$$

This optimization problem could be termed the *Markovian Evader Network Interdiction* (MENI) problem.

This expression can be generalized for the case of multiple evaders where each evader represents a threat scenario or an adversarial group. Each evader k then has certain probability $w^{(k)}$ of occurring ($\sum_k w^{(k)} = 1$), as well as a distinctive source distribution $\mathbf{a}^{(k)}$, target node $t^{(k)}$ and transition matrix $\mathbf{M}^{(k)}$. The generalized objective is a weighted sum of Eq. (3.2.1) over all evaders.

The specifics of the Markovian evader model would effect whether the sum in Eq. (3.2.1) converges. In general it is sufficient to know that $(\mathbf{I} - \mathbf{M})^{-1}$ exists because Eq. (3.2.1) implies

$$\|\mathbf{h}\|_1 \leq \|\mathbf{a}(\mathbf{I} - \mathbf{M})^{-1}\mathbf{M}(\mathbf{I} - \mathbf{M})^{-1}\|_1 \max_{(i,j) \in E} C_{ij}, \quad (3.0)$$

where the right-hand side is the average path length before reaching the target times the maximal edge cost. Letting $\mathbf{v}_1 = \mathbf{a}(\mathbf{I} - \mathbf{M})^{-1}$ and $\mathbf{v}_2 = \mathbf{v}_1\mathbf{M}$, observe that $\|\mathbf{h}\|_1 \leq k \|\mathbf{v}_2(\mathbf{I} - \mathbf{M})^{-1}\|_1$ for some constant $k > 0$. In turn existence is guaranteed if the target node is an absorbing state of the Markov chain defined by \mathbf{M} , that is, a chain where $M_{tj} = 0$ if $j \neq t$ and t is reached from any node with non-zero probability after finitely many steps [11, Sec.11.2].⁵

3.3 Evader models

We now develop a concrete Markovian evader model and introduce and analyze algorithms for finding interdiction sets. Recall that the advantage of stochastic models is not only the promise of better interdiction solutions but also the additional information they provide about evader motion. Specifically, with stochastic models and in particular with Markovian models it is possible to compute probabilities that each node and edge in the

⁵The \mathbf{M} matrix here is the \mathbf{Q} matrix in Ref. [11] except for a small detail: \mathbf{Q} includes only the transitions between non-absorbing states, but here \mathbf{M} does include the absorbing state - the target node t . As a compensation, we impose $\mathbf{M}_{tt} = 0$ implying that the evader is removed from the graph upon reaching t .

network will be traversed - a fact that could be exploited for designing algorithms (see Sec. 3.A).

3.3.1 Least-cost-guided evader

As was noted in the introduction the evader may often be unable to determine correctly the least-cost or least-risk path to the target because of limited information about the network topology, interdiction decisions, or the costs and risks along alternative paths. Let us suppose then that the errors the evader makes in finding the best path are random in the sense that the evader does not systematically overestimate or underestimate the costs of certain types of paths. The random error assumption implies that at every node the transition along least-risk path would still be the most likely. To make the discussion more concrete, recall that the Markovian evader model is specified through the probability M_{ij} that evader at node i would traverse $i \rightarrow j$. Therefore, suppose that in general M_{ij} increases with the probability of (successful) evasion on path q_{ij} - defined as the path consisting of the edge (i, j) and then of the least-risk path from j to the target. One choice is to assume that an evader would traverse edge (i, j) with probability proportional to q_{ij} , or more generally, proportional to a positive power of q_{ij}

$$M_{ij} \propto \left(\frac{q_{ij}}{q_{i*}} \right)^\lambda, \quad (3.0)$$

where $\lambda > 0$ is a parameter, $q_{i*} = \max_j q_{ij}$ is the probability of evasion if the least-risk path from i to the target is followed. (The constant of proportionality is found from $\sum_j M_{ij} = 1$.) When $\lambda \rightarrow \infty$ the evader moves deterministically along the least-risk path and when $\lambda \rightarrow 0$ the motion is perfectly random. The least-risk path has the highest probability, but the difference with other paths vanishes as $\lambda \rightarrow 0$. Hence, the model can be called the “least-risk-guided evader”

If the values of the probabilities q_{ij} and q_{i^*} are not known directly they can be found by relating them to edge costs. One approach is to find the cost $c(p_{ij})$ of the least-cost path p from i through j , and the cost $c(p_{i^*})$ of the least-cost path p_{i^*} from i , (see Eq. (3.2.1) and Fig. 3.2.) Then the probabilities of evasion may be computed from the cost by the relation $q_{ij} = e^{-c(p_{ij})}$, and $q_{i^*} = e^{-c(p_{i^*})}$. Substitution to Eq. (3.3.1) gives

$$M_{ij} \propto e^{-\lambda(c(p_{ij})-c(p_{i^*}))}, \quad (3.0)$$

where $c(p_{ij})$ is the cost of the least cost from i through j to the target.

This model, termed the “least-cost-guided evader”, is similar to one developed for routing in ad-hoc wireless networks. In that application \mathbf{M} is used to determine where to transmit a message when the final destination cannot be reached directly [16].

In some applications edge costs are easy to obtain but evasion probabilities q_{ij} on paths are known only poorly or not at all. In those problems it is more natural to assume Eq. (3.3.1) rather than Eq. (3.3.1) (the costs of all the relevant paths could be found using Dijkstra’s algorithm.) Eq. (3.3.1) could be independently motivated by supposing that the evader attempts to minimize costs (as in the least-cost-path formulation) but uncertainty causes stochastic deviations from the least-cost path. The adherence to the least-cost path is determined by the parameter λ . This is somewhat similar to the stochastic motion of a quantum particle.

In the model the parameter λ represents the precision of the information the evader has about the graph and interdiction decisions. The effect of λ on the motion of the evader is not linear. Rather for all networks we have studied there is a continuous phase transition as λ is increased from effectively stochastic regime to deterministic motion, with corresponding decrease in the average cost of reaching the target node (Fig. 3.3).

Notice that although M_{ij} values in Eq. (3.3.1) depend on the cost of least-cost path,

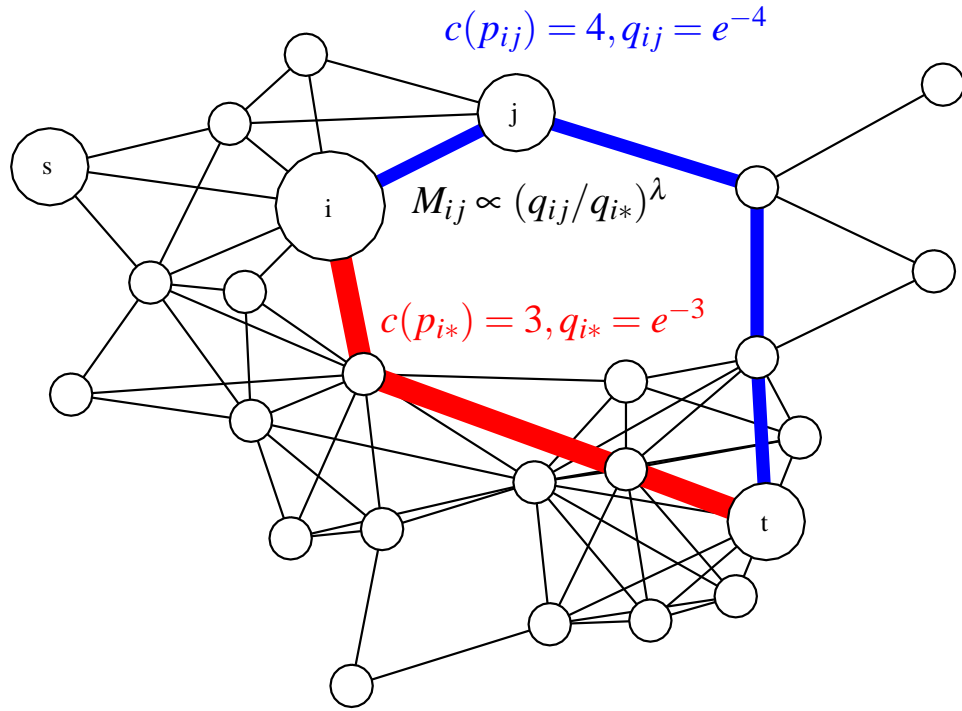


Figure 3.2: Computation of the transition probabilities M_{ij} . The least-cost path from node i to the target t is the path p_{i^*} (thick red) with cost $c(p_{i^*}) = 3$. Through node j the shortest path to t is (thin blue) path p_{ij} with cost $c(p_{ij}) = 4$. If the probability of evasion on p_{ij} is assumed to be $q_{ij} := q(p_{ij}) = e^{-c(p_{ij})}$ then the edge transition probability to j is $M_{ij} \propto (q_{ij}/q_{i^*})^\lambda = e^{-\lambda}$.

when $\lambda < \infty$ this dependence is a smooth function of path costs. Thus the new formulation provides a more desirable description of evader motion because it avoids the sensitivity to path costs seen in the shortest-path evader model.

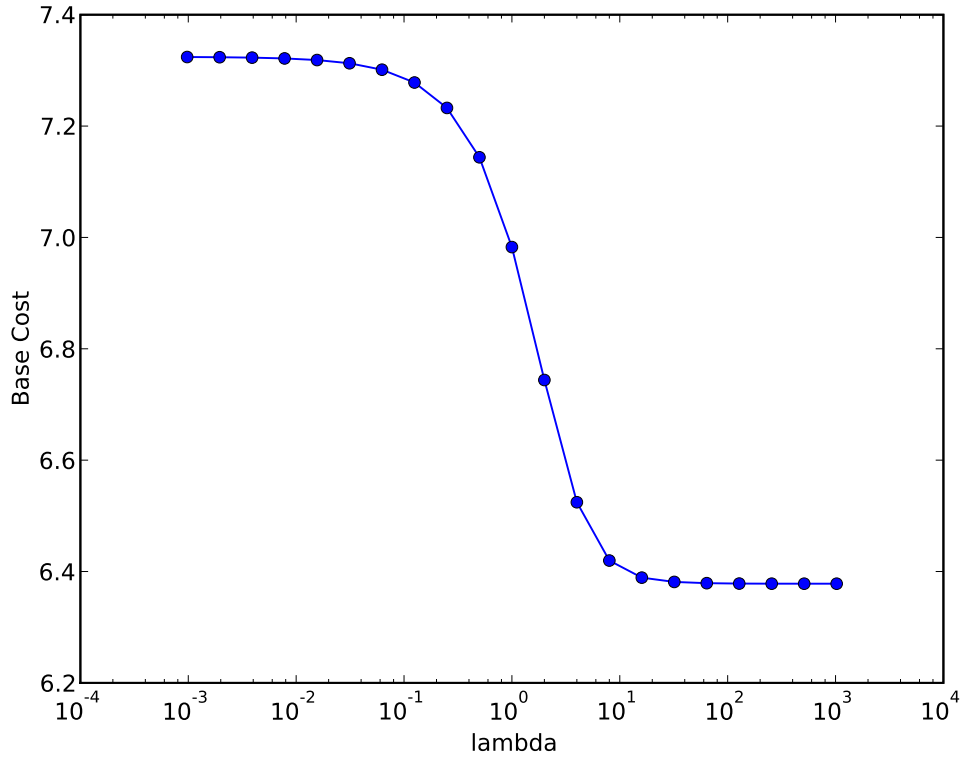


Figure 3.3: The cost of reaching the target as a function of λ . No edges have been interdicted, and the effect is solely due to deviations from the least-cost path. The evader is the non-treating evader discussed in the next section, and the underlying network is a 20×20 grid graph with 20 shortcut edges.

3.3.2 Non-retreating evader

A useful variant the least-cost-guided model is the non-retreating evader. In this model it is assumed that an evader always moves to nodes that are closer to the target node t than the current node. To represent this model assume that there is zero probability of motion through (i, j) if node i is at least as close to the target as node j , namely, $c(p_{i*}) \leq c(p_{j*})$, where $c(p_{i*})$ and $c(p_{j*})$ are the smallest costs of paths to the target from nodes i and j ,

respectively, computed using Eq. (3.2.1). An interesting effect of this assumption is that the evader would never cross a node or an edge twice.

Consequently the set of nodes becomes a partially ordered set and as a result, there exists a relabeling σ of the nodes such that if $c(p_{i^*}) > c(p_{j^*})$ then $\sigma(i) > \sigma(j)$. A simple (non-unique) procedure is to label the target node t as 0 ($\sigma(t) = 0$) and then rank the nodes in the order of their distance (cost) along least-cost path to t , breaking ties arbitrarily. Computationally, this is the same as the order the nodes are reached by a shortest path (Dijkstra's) algorithm starting at t . The transition probability becomes

$$\hat{M}_{ij} = \begin{cases} M_{ij}, & c(p_{i^*}) > c(p_{j^*}), \\ 0, & c(p_{i^*}) \leq c(p_{j^*}). \end{cases}$$

In this case all paths must reach the target after at most $|N| - 1$ steps, where $|N|$ is the number nodes in G , and hence $\hat{\mathbf{M}}$ becomes nilpotent of power $|N| - 1$. Moreover, by labeling the nodes up in order of increasing cost, $\hat{\mathbf{M}}$ can be written as a lower-triangular matrix with zero diagonal. For example, if the evader traverses a 2×3 grid with the target at one corner then one possible σ gives the matrix

$$\hat{\mathbf{M}} = \begin{pmatrix} 0 & & & & & & \\ 1 & 0 & & & & & \\ 1 & 0 & 0 & & & & \\ 0 & 1 & 0 & 0 & & & \\ 0 & 0.5 & 0.5 & 0 & 0 & & \\ 0 & 0 & 0 & 0.5 & 0.5 & 0 & \end{pmatrix}.$$

The special matrix structure facilitates an order-of magnitude speedup in the computation of Eq. 3.2.1. For a general \mathbf{M} , computing $\mathbf{a}(\mathbf{I} - \mathbf{M})^{-1}$ involves Gaussian elimination at a cost of $2|N|^3/3$ operations. For a nilpotent lower-triangular $\hat{\mathbf{M}}$ the cost falls

to $O(|N|^2)$ since we can use backward-forward substitutions instead of Gaussian elimination. The cost of computing the objective function Eq. (3.2.1) is also expected to drop to $O(|N|^2)$ despite the need to reorder the matrix \mathbf{C} when the nodes are relabeled.

3.4 Solving the Markovian Evader Interdiction Problem (MENI)

The challenge of network interdiction consists of developing both realistic models and tractable algorithms. The Markovian evader model adds realism but does not reduce the computational complexity of finding good interdiction solutions. Indeed it is clear that the Markovian model is computationally hard because in the limit of $\lambda \rightarrow \infty$, the model becomes the least-cost interdiction problem which is NP-Hard [7, 8] and also hard to approximate [9]. Therefore, this section discusses solution heuristics based on network structure.

A common approach to solving many combinatorial optimization problems is based on local, or neighborhood, search algorithms. In general, such local search methods such as simulated annealing (SA), genetic algorithms (GA), and Tabu search (TS) [18], take a random solution (or a population thereof) and improve it through a series of incremental changes. These class of algorithms can often provide good solutions but usually do not guarantee optimality.

We implemented the above local search algorithms and tested them on the Markovian evader network interdiction problem for some sample networks. Even after tuning parameters in the algorithms we found that all the algorithms gave comparably poor performance. It was clear that the solutions were highly suboptimal. For instance in many cases it was easy to markedly improve the solution by simply interdicting edges incident to the target nodes.

It is not surprising that general local search algorithms do not work well for large instances of Markovian evader network interdiction: the interdiction solution space is exponential in the budget so any iterative improvement process of local search can only explore a very small fraction of solutions in a reasonable number of steps. In addition the solution space is very rugged as there are synergies (non-linear gains) when multiple edges are interdicted on paths going to a single target because the evader is prevented from easily bypassing the interdicted edges.

3.4.1 Global heuristics

The weak performance of the local search algorithms suggests that fast high-quality heuristics can only come from more specialized solvers that exploit the structure of the interdiction problem. Particularly promising algorithms are those that take advantage of ranking functions (called here “heuristics”) that rank edges according to global information about graph structure. (The specifics of the function are explained in the next subsection.)

The ranking functions are to be incorporated into two greedy procedures for constructing the interdiction set: Direct Greedy Heuristic (DGH) and “Randomized Greedy Heuristic” (RGH). DGH simply performs greedy sequential selection of edges based on their rank. The ranking function H_S is computed for all edges e based on the structure of the graph and the current interdiction set S (Alg. 1).

This algorithm is fast since it merely has to compute the ranking heuristic B times. A concern with such an approach is that it may be hard to find high-accuracy global heuristics that also run fast. It will be shown later in this section that graph betweenness centrality performs surprisingly well.

Algorithm 1: Direct Greedy algorithm using global heuristic H (DGH) for budget

B

$S \leftarrow \emptyset$

while $B > 0$ **do**

$S \leftarrow S \cup \{\operatorname{argmax}_{e \in E \setminus S} H_S(e)\}$, resolving ties arbitrarily.

$B \leftarrow B - 1$

Output(S)

A more serious issue is that it is not difficult to find instances of MENI in which betweenness centrality performs poorly because it finds edges that are easy to bypass. This prompted the development of a more reliable algorithm, RGH. The algorithm is in turn based on a very simple local search method, a “randomized greedy” (RG) algorithm show in Alg. 2. In RG the interdiction set S is constructed incrementally by taking the best from a sample of edges. Namely, at each step a random sample $E_L \subset E \setminus S$ of size L is generated and then the edge e that gives the greatest increase in the evader cost is added to S . The sample size L is typically much smaller than the number of edges in the graph because it is computationally expensive to calculate the change $\Delta_S(e)$ in the objective function due to a single edge.

The RG algorithm performs similarly to the other local search methods we considered above. However, by itself it is not expected to have good performance (performance data is discussed in the next subsection). The RGH algorithm makes the RG algorithm more efficient using global information. Specifically, rather than randomly sampling from $E \setminus S$ it chooses some edges that were ranked highly by a heuristic H_S (Alg. 3).

Note that the RGH algorithm has two selection steps - first is the selection of the sample which is followed by the selection of the best edge in this sample. Only the edge with the greatest gain is selected, where gain refers to increase in evader cost. The

Algorithm 2: Randomized greedy (RG) construction of the interdiction set S with budget B and sample size L

$S \leftarrow \emptyset$

while $B > 0$ **do**

$E_L \leftarrow \{L \text{ random elements from } E \setminus S\}$

for all $e \in E_L$ **do**

$\Delta_S(e) := h(S \cup \{e\}) - h(S)$

$S \leftarrow S \cup \{\operatorname{argmax}_{e \in E_L} \Delta_S(e)\}$, resolving ties arbitrarily.

$B \leftarrow B - 1$

Output(S)

Algorithm 3: Randomized greedy with heuristic (RGH) for budget B and sample size L

$S \leftarrow \emptyset$

while $B > 0$ **do**

$E_L \leftarrow \{\lfloor \frac{L}{2} \rfloor \text{ random elements from } E \setminus S\}$

$E_L \leftarrow E_L \cup \{e \mid e \in \text{top } \lfloor \frac{L}{2} \rfloor \text{ elements ranked by } H_S(e)\}$

for all $e \in E_L$ **do**

$\Delta_S(e) := h(S \cup \{e\}) - h(S)$

$S \leftarrow S \cup \{\operatorname{argmax}_{e \in E_L} \Delta_S(e)\}$, resolving ties arbitrarily.

$B \leftarrow B - 1$

Output(S)

running-time complexity of this algorithm is $O(BL)$, assuming that the cost of computing the heuristic is dominated by the cost of computing the gain.

Note that in Algorithm 3 only some of the edges are selected using the heuristic and the rest are selected randomly. Allowing some edges to be selected randomly ensures

that the algorithm is not deterministic. Determinism would prevent exploring the entire set of feasible solutions even in principle. It may be useful for some problems to further increase exploration by replacing argmax with a stochastic selection biased towards the best edges.

3.4.2 Betweenness centrality heuristic

Our next task is to introduce ranking of edges based on global information about the problem. One such ranking is to compute edge “betweenness centrality” metric - the fraction of least-cost paths between pairs of nodes in a network that cross a given edge [19]. This metric identifies those edges that are critical to connectivity within a network because they participate in a large number of least-cost paths linking nodes on a network, such as bridge edges that joins two graph components.

An effective heuristic based on this notion could be constructed as long as we also consider the distribution of source sites, \mathbf{a} and the target node of the evader t . Recall that a_s is the probability that the evader would start at node s . Let σ_{st} be the number of least-cost paths between nodes s and the target node t in the graph, and let $\sigma_{st}(e)$ be the number of those paths that pass through edge e . Therefore, we define the source-weighted centrality of edge e with respect to t as the sum:

$$C_{\mathbf{a},t}(e) = \sum_{s:t \neq s \in V} a_s \frac{\sigma_{st}(e)}{\sigma_{st}}. \quad (3.0)$$

Finally, we let $H_S(e) = C_{\mathbf{a},t}(e)$ in algorithms 1 and 3 and this defines the algorithms “DGH-betweenness” and “RGH-betweenness”, respectively. Notice that this quantity needs to be re-computed during execution: as the interdiction set S is increased, the costs of the edges change and so are the least-cost paths.

An algorithm for calculating a metric of this kind for all $e \in E$ in $O(|E| + |N| \log |N|)$

time is found in Ref. [20]. For the RGH algorithm, the running time of this algorithm equals asymptotically the complexity of the original RG algorithm because the time for computing change in the evader cost due to the addition of a single edge (Eq. 3.2.1) dominates the cost of the heuristic (assuming a constant number of evaders and $\Theta(|E|) = |N|$).

This heuristic is of course similar to the least-cost evader model (Subsec. 3.3.1) which is also based on the least-cost paths to the target. A concern in both cases is high sensitivity to edge costs. Moreover, when used in a heuristic it would seem unsuitable to interdiction of stochastically-moving evaders. Fortunately, the least-cost path from each node tends to be the most probable evader path in that model. For any value of $\lambda > 0$ interdiction of the least-cost path should shift the evader to taking alternative paths, which are by definition more costly. Interdicting of least-cost path is provably optimal in simple types of interdiction problems where a single pair of source-target nodes is connected by several non-intersecting paths. It is interesting that when the paths to the target do intersect, the least-cost path is not always the best location for interdiction. Indeed, it is possible to construct graphs where as a result of attracting evaders from multiple source nodes, relatively high cost paths are more likely to be traversed than the least-cost paths. As a result one finds examples where betweenness-based heuristics (even RGH-betweenness) consistently find poor interdiction locations (for sufficiently small λ). There an alternative heuristic, termed “motion likelihood” is superior. Briefly, the idea is to use for the ranking function the probability of the evader transitioning each edge. This is calculated from the model for evader motion. While promising, algorithms based on evader motion were found to be inferior in solution quality and running time compared to the current betweenness centrality-based heuristic (see the Appendix for a detailed discussion).

3.4.3 Performance Results

Theoretical consideration suggest that in interdiction problems DGH would have faster running time but RGH would find higher quality solutions. Unexpectedly simulations showed that DGH is the best algorithm in both running time and solution quality. For the simulations we constructed 50 network interdiction problems and also run the algorithms on real transport networks. Each of the simulated networks was a 400-node 20x20 grid graph with 20 randomly-added “jump” edges representing air links. The weights on all edges were sampled randomly from 0.5..1.5.

The results of the comparison show that both DGH- and RGH-betweenness outperform the basic algorithm (RG) Fig. 3.4. This advantage was maintained in both $\lambda = 0.001$ and $\lambda = 1000$ cases. The performance of the heuristics on empirical transportation network was qualitatively similar Fig. 3.5.⁶ Indeed, in those larger networks the RG algorithm, being a local search algorithm, cannot efficiently find any solutions that materially increase the evader’s cost. All other models of transportation networks that were tried gave qualitatively identical performance (including GTG-based problems [17], as well as networks generated from Waxman’s model RG2 [22].)

The gain from the heuristics is significant in a more absolute sense: The solutions found by the algorithms were benchmarked to the cost found by a variant of RG that exhaustively considers all the edges in the graph. It was found that the global heuristic found solutions which were always $> 95\%$ of the exhaustive solution and sometimes even superior to it. This implies that it finds solutions almost indistinguishable from solutions found by an algorithm using about 80 times as many cost evaluations.

⁶(1) 9559-node 29682-edge Washington, DC network with weights indicating time of travel. (2) 3353-node 8859-edge Rome, Italy network with costs indicating distance. The networks are available at the DIMACS 9th Implementation Challenge website <http://www.dis.uniroma1.it/~challenge9/download.shtml>

When observing the simulations for the DC networks we saw that sometimes a larger interdiction set actually decreased evader cost, for reasons that were illustrated in Fig. 3.1. The data of Fig. 3.4 shows the best solution found for interdiction S within budget B ($|S| \leq B$) rather than in budget B ($|S| = B$).

Another interesting finding is that when B is much smaller than the size of the solution space, $|E|$, there was almost no difference in solution quality between RGH-betweenness and the much faster DGH-betweenness. Thus, employing the heuristics directly (without computing the gain in cost) causes no apparent loss of solution quality. For the larger budgets relative performance is even better for DGH-betweenness (one-tailed t statistic has probability < 0.001 at 0.05 level.) This is probably due to the excessive greediness of the RGH-betweenness as it selects not only the highest betweenness edges but also the edge with the greatest gain in evader cost, as compared to just one greedy step in DGH-betweenness. In many other discrete optimization problems it was noted that greedy algorithms tend to find weak local optima rather than better or even global optima.

When using DGH the running time is improved by factor of $> L$ over RGH (where L is the RGH sample size, in the above $L = 20$). Thus the fast DGH-betweenness algorithm offers the best running time and solution quality.

3.5 Conclusions and Outlook

The main contribution of this work are:

- a stochastic model of the evader motion based on Markovian guided random walk
- demonstration the importance of stochastic models as opposed to simple least-cost

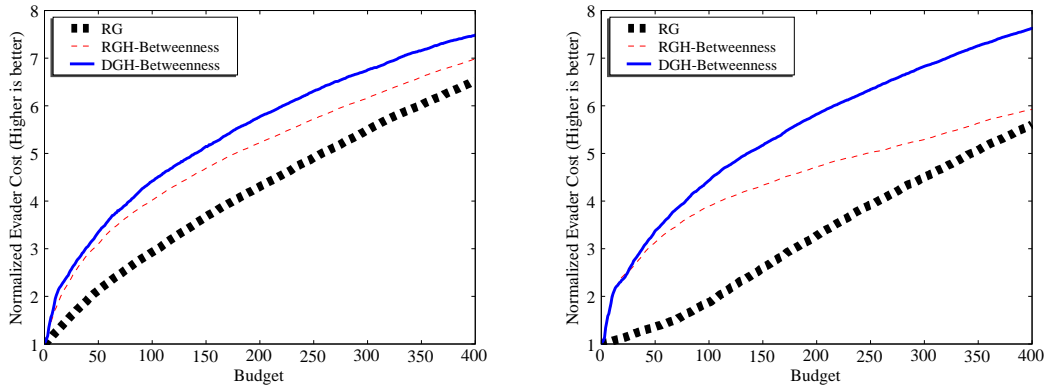


Figure 3.4: Comparison on the grid networks between a basic RG algorithm and DGH and RGH which use the betweenness heuristic. The data shows the average of 50 simulated problems per value of λ : (top) $\lambda = 0.001$, (bottom) $\lambda = 1000$. The convergence of the algorithms around budget of 400 for $\lambda = 0.001$ occurs because most of the paths to the targets become interdicted many times. To compare solutions across several different networks with different interdiction complexities, the solution on each network was normalized by the evasion cost on this network without any interdiction. There were two equally likely evaders (corresponding to two targets) and 10 source nodes selected at uniformly at random (note that the number of source nodes does not effect either solution quality or running time). The sample size in RG and RGH-betweenness was $L = 20$.

path models

- demonstration the unsuitability of local search heuristics to network interdiction
- construction a pair of global search heuristics based on between centrality

In future work it would be very useful for network interdiction and other applications to construct additional heuristics for finding edges that cannot be cheaply bypassed by the evader.

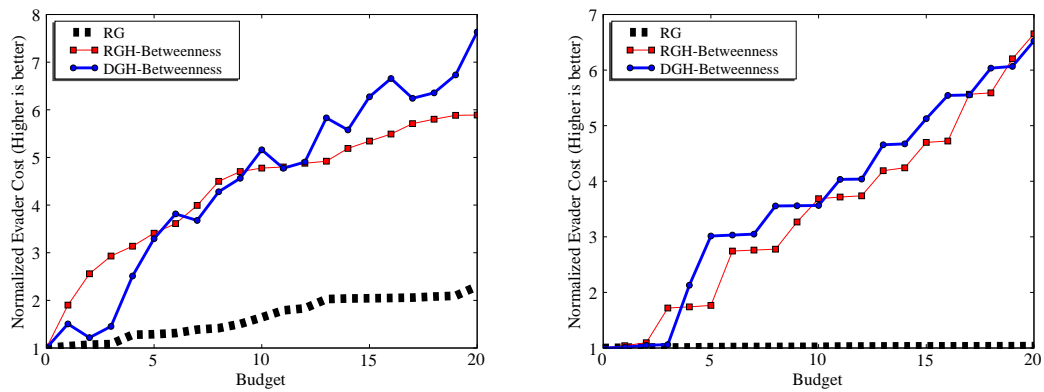


Figure 3.5: Comparison on the Washington, DC network between a basic RG algorithm and algorithms which use global heuristics. The results on the Rome network were qualitatively similar. Note the much smaller maximum budget compared to the previous set of simulations. The data shows the average of 5 simulated problems per value of λ : (top) $\lambda = 0.001$, (bottom) $\lambda = 1000$. The interdiction added cost, D_{ij} , was half the diameter of the network (specifically, its largest connected component.) As above, the solution on the network was normalized by the evasion cost on this network without any interdiction. For this much larger graph, the number of sources was increased while maintaining that about 2.5% of the nodes are source nodes. The differences between DGH and RGH are not statistically significant and both beat RG (one-tailed t statistic has probability < 0.01 at 0.05 level.)

The evader model developed here is a step toward a more refined model that more closely ties evader motion with its computational and informational constraints. Research into more refined models promises further gains in computational performance and realism.

Acknowledgments AG would like to thank David Shmoys for suggesting the use of randomized greedy algorithm for interdiction, and Vadas Gintautas for countless fruitful discussions. Part of this work was funded by the Department of Energy at Los Alamos

National Laboratory under contract DE-AC52-06NA25396 through the Laboratory Directed Research and Development Program.

APPENDIX

3.A The Motion Heuristic

One promising heuristic is the evader motion itself: Edges that are likely to be traversed by the evader are also likely to be good interdiction locations. Intuitively, interdiction of such edges would compel the evader to take alternative paths which are often considerably more costly.

This heuristic is also appealing because it should be able to avoid ineffective interdiction in cases where there are several paths to the target of approximately equal length. Such a situation requires interdiction of all paths (which might be too expensive) if interdiction is to have any effect. The heuristic would sense this because the likelihood of motion would be split equally among alternative paths, which would decrease their ranking compared to interdiction elsewhere on the graph.

The technical details of the implementation are as follows. For any given Markovian evader model \mathbf{M} and starting vector \mathbf{a} the expected number of times the evader crosses an edge $e = (i, j)$ is

$$E[e] = \mathbf{a}(\mathbf{I} - \mathbf{M})^{-1}M_{ij}. \quad (3.0)$$

The motion likelihood algorithms (termed “DGH-likelihood” and “RGH-likelihood”) are constructed by setting $H_S(e) := E[e]$ in Alg. 1 and 3, respectively. Notice that \mathbf{M} is an implicit function of the interdiction set S because S affects edge costs and hence evader motion as expressed through \mathbf{M} . Therefore, it is necessary to recompute $E[e]$ at each step of the algorithm. Fortunately, this heuristic is relatively inexpensive to compute and can be evaluated for all edges in the network simultaneously at asymptotic time of $O(|N|^2)$ for the non-retreating evader model (for a general \mathbf{M} it takes $O(|N|^3)$ because of

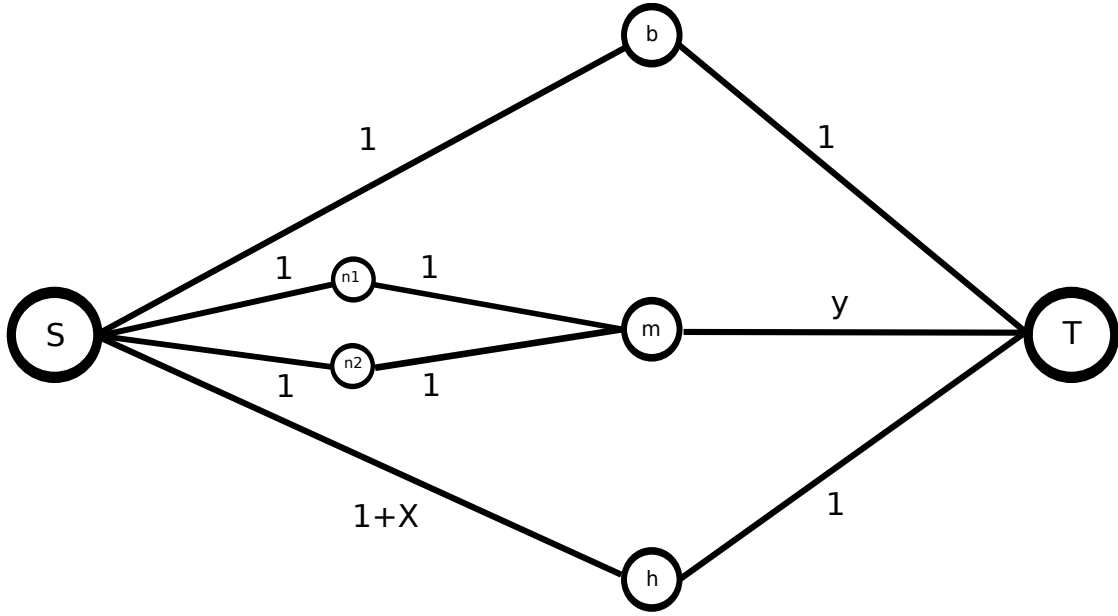


Figure 3.6: Example network where motion likelihood heuristic outperforms the betweenness heuristic.

Gaussian elimination.) This makes it more expensive than betweenness centrality ranking ($O(|E| + |N| \log |N|)$) but cheaper than computing evader cost in Eq. (3.2.1), which would cost $\Theta(|N|^2)$ per edge in the non-retreating model.

It is easy to find examples where the motion heuristics outperform the betweenness heuristics, as in Fig. 3.A. Suppose an evader must move in this graph from S to T , and the interdiction problem has budget 1. There are four paths through nodes b , $n1$, $n2$ and h with costs 2 , $2 + y$, $2 + y$ and $2 + x$, respectively and consider the case where $x > y$. The edges of highest betweenness are (S, b) and (b, T) , and for sufficiently large value of λ they are the optimal interdiction locations since they lie on the least-cost path. However, for a range of values $[0, \lambda_c]$ (where λ_c is a constant that depends on x and y) the edge with the highest likelihood of motion is (m, T) because it is the confluence of two distinct paths whose cost is close to the cost of the least-cost path (assuming $y \ll 1$).

Interdiction of this edge would redirect the evader towards both the least-cost path, but also towards the high cost path through h (cost $2 + x$). Indeed, it is easy to see that in the limit of $\lambda \rightarrow 0$, interdiction of (m, T) gives cost $E_m := 2 + \frac{x}{2}$. Depending on the values of x and y , E_m is greater than the cost after interdiction along the least-cost path, $E_b := 2 + \frac{2y+x}{3}$ (say at edge (b, T)). The difference, $\frac{x}{6} - \frac{2y}{3}$ can be arbitrarily large when $x - y \rightarrow \infty$. The precise threshold value of λ where the phenomenon appears is not easy to express analytically.⁷

By generalizing this example it is possible to construct problem instances in which motion likelihood is superior to betweenness when the interdiction budget > 1 . Notice also that networks containing this graph motif can defeat not only DGH-betweenness but also the RGH-betweenness algorithm as long as the number of least-paths is large, since RGH will not choose the high likelihood edge into its sample.

A wide range of simulations were used to assess those heuristics but the results were inconclusive. In some simulations such as on 20x20 grids above (Fig. 3.4), motion-based heuristics outperforms in solution quality equivalent algorithms based on betweenness. Yet in GTG graphs [17] of 100 nodes betweenness performs better. The recommendation of betweenness (previous section) stems from its faster running time in both DGH and RGH implementations. As well in the worst simulations betweenness achieved evader cost equal to 90% cost found by motion likelihood. The inconclusive findings highlight the need for additional research into those heuristics but also into novel approximation algorithms that come with guaranteed performance bounds.

⁷The limit of $\lambda \rightarrow 0$ appears somewhat special because in this case the evader ignores the presence of interdiction - the added cost on the edge, and so interdiction appears to directly increase cost rather than redirecting the evader. Actually redirection occurs even in this limit: in the non-retreating evader model, the probability of motion through (b, T) is 0 when the total cost on that edge is $C_{bT} \geq 2 + y$. Similarly, there is no motion through (m, T) if its total cost $C_{mT} \geq 2$. Redirection still occurs under arbitrarily small λ with the more general least-cost-path guided model because it is possible to redirect while contributing an arbitrarily small amount to expected cost: this contribution consists of the cost of the path times a term with a negative exponent in the edge cost, so it vanishes.

REFERENCES

- [1] P. M. Ghare, D. C. Montgomery, and W. C. Turner, "Optimal interdiction policy for a flow network," *Naval Research Logistics Quarterly*, vol. 18, no. 1, pp. 37–45, 1971.
- [2] A. W. McMasters and T. M. Mustin, "Optimal interdiction of a supply network," *Naval Research Logistics Quarterly*, vol. 17, no. 3, pp. 261–268, 1970.
- [3] D. P. Morton, F. Pan, and K. J. Saeger, "Models for nuclear smuggling interdiction," *IIE Transactions*, vol. 39, no. 1, pp. 3–14, 2007.
- [4] B. Pourbohloul, L. Meyers, D. Skowronski, M. Krajdén, D. Patrick, and R. Brunham, "Modeling control strategies of respiratory pathogens," *Emerg. Infect. Dis.*, vol. 11, no. 8, pp. 1246–56, 2005.
- [5] N. Memon and H. Larsen, "Practical algorithms for destabilizing terrorist networks," in *Intelligence and Security Informatics*, pp. 389–400, 2006.
- [6] F. Pan, W. Charlton, and D. P. Morton, "Interdicting smuggled nuclear material," in *Network Interdiction and Stochastic Integer Programming* (D. Woodruff, ed.), pp. 1–19, Kluwer Academic Publishers, Boston, 2003.
- [7] H. W. Corley and D. Y. Sha, "Most vital links and nodes in weighted networks.," *Oper. Res. Lett.*, vol. 1, pp. 157 – 160, Sep 1982.
- [8] A. Bar-Noy, S. Khuller, and B. Schieber, "The complexity of finding most vital arcs and nodes," tech. rep., University of Maryland, College Park, MD, USA, 1995.
- [9] E. Boros, K. Borys, and V. Gurevich, "Inapproximability bounds for shortest-path network interdiction problems," tech. rep., Rutgers University, Piscataway, NJ, USA, 2006.
- [10] F. Pan and D. P. Morton, "Minimizing a stochastic maximum-reliability path," *Networks*, vol. 52, pp. 111–119, 2008.
- [11] M. P. Atkinson, Z. Cao, and L. M. Wein, "Optimal stopping analysis of a radiation detection system to protect cities from a nuclear terrorist," *Risk Analysis*, vol. 28, pp. 353–371, Apr 2008.
- [12] H. Bayrak and M. Bailey, "Shortest path network interdiction with asymmetric information," *Networks*, vol. 52, pp. 133–140, 2008.

- [13] U. Janjarassuk and J. Linderoth, “Reformulation and sampling to solve a stochastic network interdiction problem,” *Networks*, 2008. to appear.
- [14] D. Reich and L. Lopes, “The most likely path.” preprint, 2008.
- [15] C. M. Grinstead and J. L. Snell, *Introduction to Probability*. USA: American Mathematical Society, second revised ed., Jul 1997.
- [16] C. L. Barrett, S. J. Eidenbenz, L. Kroc, M. Marathe, and J. P. Smith, “Parametric probabilistic routing in sensor networks,” *Mobile Networks and Applications*, vol. 10, pp. 529–544, 2005.
- [17] M. O. Ball, B. L. Golden, and R. V. Vohra, “Finding the most vital arcs in a network,” *Oper. Res. Lett.*, vol. 8, no. 2, pp. 73–76, 1989.
- [18] I. H. Osman and J. P. Kelly, *Meta-Heuristics: Theory and Applications*. Norwell, MA, USA: Kluwer Academic Publishers, 1996.
- [19] L. C. Freeman, “A set of measures of centrality based on betweenness,” *Sociometry*, vol. 40, no. 1, pp. 35 – 41, 1977.
- [20] U. Brandes, “A faster algorithm for betweenness centrality,” *Journal of Mathematical Sociology*, vol. 25, no. 2, pp. 163–177, 2001.
- [21] M. Bradonjić and J. S. Kong, “Wireless ad hoc networks with tunable topology,” in *Forty-Fifth Annual Allerton Conference*, (UIUC, Illinois, USA), pp. 1170–1177, 2007.
- [22] B. Waxman, “Routing of multipoint connections,” *IEEE J. on Selected Areas in Communications*, vol. 6, pp. 1617–1622, Dec 1988.

CHAPTER 4

OPTIMAL INTERDICTION OF UNREACTIVE MARKOVIAN EVADERS

Chapter Abstract The interdiction problem arises in a variety of areas including military logistics, infectious disease control, and counter-terrorism. In the typical formulation of *network* interdiction, the task of the interdictor is to find a set of edges in a weighted network such that the removal of those edges would maximally increase the cost to an evader of traveling on a path through the network.

Our work is motivated by cases in which the evader has incomplete information about the network or lacks planning time or computational power, *e.g.* when authorities set up roadblocks to catch bank robbers, the criminals do not know all the roadblock locations or the best path to use for their escape.

We introduce a model of network interdiction in which the motion of one or more evaders is described by Markov processes and the evaders are assumed not to react to interdiction decisions. The interdiction objective is to find an edge set of size B , that maximizes the probability of capturing the evaders.

We prove that similar to the standard least-cost formulation for deterministic motion this interdiction problem is also NP-hard. But unlike that problem our interdiction problem is submodular and the optimal solution can be approximated within $1 - 1/e$ using a greedy algorithm. Additionally, we exploit submodularity through a priority evaluation strategy that eliminates the linear complexity scaling in the number of network edges and speeds up the solution by orders of magnitude. Taken together the results bring closer the goal of finding realistic solutions to the interdiction problem on global-scale networks.^{1 2}

¹*This chapter is released under Los Alamos National Laboratory LA-UR-09-00560*

²Joint work with Aric Hagberg and Feng Pan - Los Alamos National Laboratory

4.1 Introduction

Network interdiction problems have two opposing actors: an “evader” (*e.g.* smuggler) and an “interdictor” (*e.g.* border agent.) The evader attempts to minimize some objective function in the network, *e.g.* the probability of capture while traveling from network location s to location t , while the interdictor attempts to limit the evader’s success by removing network nodes or edges. Most often the interdictor has limited resources and can thus only remove a very small fraction of the nodes or edges. The standard formulation is the max-min problem where the interdictor plays first and chooses at most B edges to remove, while the evader finds the least-cost path on the remaining network. This is known as the B most vital arcs problem [1].

This least-cost-path formulation is not suitable for some interesting interdiction scenarios. Specifically in many practical problems there is a fog of uncertainty about the underlying properties of the network such as the cost to the evader in traversing an edge (arc, or link) in terms of either resource consumption or detection probability. In addition there are mismatches in the cost and risk computations between the interdictor and the evaders (as well as between different evaders), and all agents have an interest in hiding their actions. For evaders, most least-cost-path interdiction models make optimal assumptions about the evader’s knowledge of the interdictor’s strategy, namely, the choice of interdiction set. In many real-world situations evaders likely fall far short of the optimum. This paper, therefore, considers the other limit case, which for many problems is more applicable, when the evaders do not respond to interdictor’s decisions. This case is particularly useful for problems where the evader is a process on the network rather than a rational agent.

Various formulations of the network interdiction problem have existed for many

decades now. The problem likely originated in the study of military supply chains and interdiction of transportation networks [2, 3]. But in general, the network interdiction problem applies to wide variety of areas including control of infectious disease [27], and disruption of terrorist networks [5]. Recent interest in the problem has been revived due to the threat of smuggling of nuclear materials [6]. In this context interdiction of edges might consist of the placement of special radiation-sensitive detectors across transportation links. For the most-studied formulation, that of max-min interdiction described above [1], it is known that the problem is NP-hard [7, 8] and hard to approximate [9].

4.2 Unreactive Markovian Evader

The formulation of a stochastic model where the evader has limited or no information about interdiction can be motivated by the following interdiction situation. Suppose bank robbers (evaders) want to escape from the bank at node s to their safe haven at node t_1 or node t_2 . The authorities (interdictors) are able to position roadblocks at a few of the roads on the network between s , t_1 and t_2 . The robbers might not be aware of the interdiction efforts, or believe that they will be able to move faster than the authorities can set up roadblocks. They certainly do not have the time or the computational resources to identify the global minimum of the least-cost-path problem.

Similar examples are found in cases where the interdictor is able to clandestinely remove edges or nodes (*e.g.* place hidden electronic detectors), or the evader has bounded rationality or is constrained in strategic choices. An evader may even have no intelligence of any kind and represent a process such as Internet packet traffic that the interdictor wants to monitor. Therefore, our fundamental assumption is that the evader does not respond to interdiction decisions. This transforms the interdiction problem

from the problem of increasing the evader’s cost or distance of travel, as in the standard formulation, into a problem of directly capturing the evader as explicitly defined below. Additionally, the objective function acquires certain useful computational properties discussed later.

4.2.1 Evaders

In examples discussed above, much of the challenge in interdiction stems from the unpredictability of evader motion. Our approach is to use a stochastic evader model to capture this unpredictability [6, 10]. We assume that an evader is traveling from a source node s to a target node t on a graph $G(N, E)$ according to a guided random walk defined by the Markovian transition matrix \mathbf{M} ; from node i the evader travels on edge (i, j) with probability M_{ij} . The transition probabilities can be derived, for example, from the cost and risk of traversing an edge [10].

Uncertainty in the evader’s source location s is captured through a probability vector \mathbf{a} . For the simplest case of an evader starting known location s , $a_s = 1$ and the rest of the a_i ’s are 0. In general the probabilities can be distributed arbitrarily to all of the nodes as long as $\sum_{i \in N} a_i = 1$. Given \mathbf{a} , the probability that the evader is at location i after n steps is the i ’th entry in the vector $\boldsymbol{\pi}^{(n)} = \mathbf{aM}^n$

When the target is reached the evader exits the network and therefore, $M_{tj} = 0$ for all outgoing edges from t and also $M_{tt} = 0$. The matrix \mathbf{M} is assumed to satisfy the following condition: for every node i in the network either there is a positive probability of reaching the target after a sufficiently large number of transitions, or the node is a dead end, namely $M_{ij} = 0$ for all j . With these assumptions the Markov chain is absorbing and the probability that the evader will eventually reach the target is ≤ 1 . For equality

to hold it is sufficient to have the extra conditions that the network is connected and that for all nodes $i \neq t$, $\sum_j M_{ij} = 1$ (see [11].)

A more general formulation allows multiple evaders to traverse the network, where each evader represents a threat scenario or a particular adversarial group. Each evader k is realized with probability $w^{(k)}$ ($\sum_k w^{(k)} = 1$) and is described by a possibly distinct source distribution $\mathbf{a}^{(k)}$, transition matrix $\mathbf{M}^{(k)}$, and target node $t^{(k)}$. This generalization makes it possible to represent any joint probability distribution $f(s, t)$ of source-target pairs, where each evader is a slice of f at a specific value of t : $\mathbf{a}^{(k)}|_s = f(s, t^{(k)}) / \sum_s f(s, t^{(k)})$ and $w^{(k)} = \sum_s f(s, t^{(k)})$. In this high-level view, the evaders collectively represent a stochastic process connecting pairs of nodes on the network. This generalization has practical applications to problems of monitoring traffic between any set of nodes when there is a limit on the number of “sensors”. The underlying network could be *e.g.* a transportation system, the Internet, or water distribution pipelines.

4.2.2 Interdictor

The interdictor, similar to the typical formulation, possesses complete knowledge about the network and evader parameters \mathbf{a} and \mathbf{M} . Interdiction of an edge at index i, j is represented by setting $r_{ij} = 1$ and $r_{ij} = 0$ if the edge is not interdicted. In general some edges are more suitable for interdiction than others. To represent this, we let d_{ij} be the interdiction efficiency, which is the probability that interdiction of the edge would remove an evader who traverses it.

So far we have focused on the interdiction of edges, but interdiction of nodes can be treated similarly as a special case of edge interdiction in which all the edges leading to an interdicted node are interdicted simultaneously. For brevity, we will not discuss node

interdiction further except in the proofs of Sec. 4.3 where we consider both cases.

4.2.3 Objective function

Interdiction of an unreactive evader is the problem of maximizing the probability of stopping the evader before it reaches the target. Note that the fundamental matrix for \mathbf{M} , using \mathbf{I} to denote the identity matrix is

$$\mathbf{N} = \mathbf{I} + \mathbf{M} + \mathbf{M}^2 + \dots = (\mathbf{I} - \mathbf{M})^{-1}, \quad (4.0)$$

and \mathbf{N} gives all of the possible transition sequences between pairs of nodes before the target is reached. Therefore given the starting probability \mathbf{a} , the expected number of times the evader reaches each node is (using (4.2.3) and linearity of expectation)

$$\mathbf{aN} = \mathbf{a}(\mathbf{I} - \mathbf{M})^{-1}. \quad (4.0)$$

If edge (i, j) has been interdicted ($r_{ij} = 1$) and the evader traverses it then the evader will not reach j with probability d_{ij} . The probability of the evader reaching j from i becomes

$$\hat{M}_{ij} = M_{ij} - M_{ij}r_{ij}d_{ij}. \quad (4.0)$$

This defines an interdicted version of the \mathbf{M} matrix, the matrix $\hat{\mathbf{M}}$.

The probability that a single evader does not reach the target is found by considering the t 'th entry in the vector E after substituting $\hat{\mathbf{M}}$ for \mathbf{M} in Eq. (4.2.3),

$$J(\mathbf{a}, \mathbf{M}, \mathbf{r}, \mathbf{d}) = 1 - \left(\mathbf{a} [\mathbf{I} - (\mathbf{M} - \mathbf{M} \odot \mathbf{r} \odot \mathbf{d})]^{-1} \right)_t, \quad (4.0)$$

where the symbol \odot means element-wise (Hadamard) multiplication. In the case of multiple evaders, the objective J is a weighted sum,

$$J = \sum_k w^{(k)} J^{(k)}, \quad (4.0)$$

where, for evader k ,

$$J^{(k)}(\mathbf{a}^{(k)}, \mathbf{M}^{(k)}, \mathbf{r}, \mathbf{d}) = 1 - \left(\mathbf{a}^{(k)} \left[\mathbf{I} - \left(\mathbf{M}^{(k)} - \mathbf{M}^{(k)} \odot \mathbf{r} \odot \mathbf{d} \right) \right]^{-1} \right)_{t^{(k)}}. \quad (4.0)$$

Equations (4.2.3) and (4.2.3) define the *interdiction probability*. Hence the *Unreactive Markovian Evader* interdiction problem (UME) is

$$\operatorname{argmax}_{\mathbf{r} \in F} J(\mathbf{a}, \mathbf{M}, \mathbf{r}, \mathbf{d}), \quad (4.0)$$

where r_{ij} represents an interdicted edge chosen from a set $F \subseteq 2^E$ of feasible interdiction strategies. The simplest formulation is the case when interdicting an edge has a unit cost with a fixed budget B and F are all subsets of the edge set E of size at most B . This problem can also be written as a mixed integer program as shown in the Appendix.

Computation of the objective function can be achieved with $\sim \frac{2}{3} |N|^3$ operations for each evader, where $|N|$ is the number of nodes, because it is dominated by the cost of Gaussian elimination solve in Eq. (4.2.3). If the matrix \mathbf{M} has special structure then it could be reduced to $O(|N|^2)$ [10] or even faster. We will use this evader model in the simulations, but in general the methods of Secs. 4.3 and 4.4 would work for any model that satisfies the hypotheses on \mathbf{M} and even for non-Markovian evaders as long as it is possible to compute the equivalent of the objective function in Eq. (4.2.3).

Thus far interdiction was described as the removal of the evader from the network, and the creation of a sub-stochastic process $\hat{\mathbf{M}}$. However, the mathematical formalism is open to several alternative interpretations. For example interdiction could be viewed as redirection of the evader into a special absorbing state - a “jail node”. In this larger state space the evader even remains Markovian. Since $\hat{\mathbf{M}}$ is just a mathematical device it is not even necessary for “interdiction” to change the physical traffic on the network. In particular, in monitoring problems “interdiction” corresponds to labeling of intercepted traffic as “inspected” - a process that involves no removal or redirection.

4.3 Complexity

This section proves technical results about the interdiction problem (4.2.3) including the equivalence in complexity of node and edge interdiction and the NP-hardness of node interdiction (and therefore of edge interdiction). Practical algorithms are found in the next section.

We first state the decision problem for (4.2.3).

Definition 1. UME-Decision.

Instance: A graph $G(N, E)$, interdiction efficiencies $0 \leq d_i \leq 1$ for each $i \in N$, budget $B \geq 0$, and real $\rho \geq 0$; a set K of evaders, such that for each $k \in K$ there is a matrix $\mathbf{M}^{(k)}$ on G , a sources-target pair $(\mathbf{a}^{(k)}, t^{(k)})$ and a weight $w^{(k)}$.

Question: Is there a set of (interdicted) nodes Y of size B such that

$$\sum_{k \in K} w^{(k)} \left(\mathbf{a}^{(k)} \left(\mathbf{I} - \hat{\mathbf{M}}^{(k)} \right)^{-1} \right)_{t^{(k)}} \leq \rho? \quad (4.0)$$

The matrix $\hat{\mathbf{M}}^{(k)}$ is constructed from $\mathbf{M}^{(k)}$ by replacing element $M_{ij}^{(k)}$ by $M_{ij}^{(k)}(1 - d_i)$ for $i \in Y$ and each (i, j) corresponding to edges $\in E$ leaving i . This sum is the weighted probability of the evaders reaching their targets. \square

The decision problem is stated for node interdiction but the complexity is the same for edge interdiction, as proved next.

Definition 2. Edge interdiction is polynomially equivalent to node interdiction.

Proof. To reduce edge interdiction to node interdiction, take the graph $G(N, E)$ and construct G' by splitting the edges. On each edge $(i, j) \in E$ insert a node v to create the

edges $(i, v), (v, j)$ and set the node interdiction efficiency $d_v = d_{ij}, d_i = d_j = 0$, where d_{ij} is the interdiction efficiency of (i, j) in E .

Conversely, to reduce node interdiction to edge interdiction, construct from $G(N, E)$ a graph G' by representing each node v with interdiction efficiency d_v by nodes i, j , joining them with an edge (i, j) , and setting $d_{ij} = d_v$. Next, change the transition matrix \mathbf{M} of each evader such that all transitions into v now move into i while all departures from v now occur from j , and $M_{ij} = 1$. In particular, if v was an evader's target node in G , then j is its target node in G' . □ □

Consider now the complexity of node interdiction. One source of hardness in the UME problem stems from the difficulty of avoiding the case where multiple edges or nodes are interdicted on the same evader path - a source of inefficiency. This resembles the *Set Cover* problem [12], where including an element in two sets is redundant in a similar way, and this insight motivates the proof.

First we give the definition of the set cover decision problem.

Definition 3. Set Cover. *For a collection C of subsets of a finite set X , and a positive integer β , does C contain a cover of size $\leq \beta$ for X ?* □

Since *Set Cover* is NP-complete, the idea of the proof is to construct a network $G(N, E)$ where each subset $c \in C$ is represented by a node of G , and each element $x_i \in X$ is represented by an evader. The evader x_i is then made to traverse all nodes $\{c \in C | x_i \in c\}$. The set cover problem is exactly problem of finding B nodes that would interdict all of the evaders (see Fig. 4.1.)

Theorem 4. *The UME problem is NP-hard even if $d_i = h$ (constant) \forall nodes $i \in N$.*

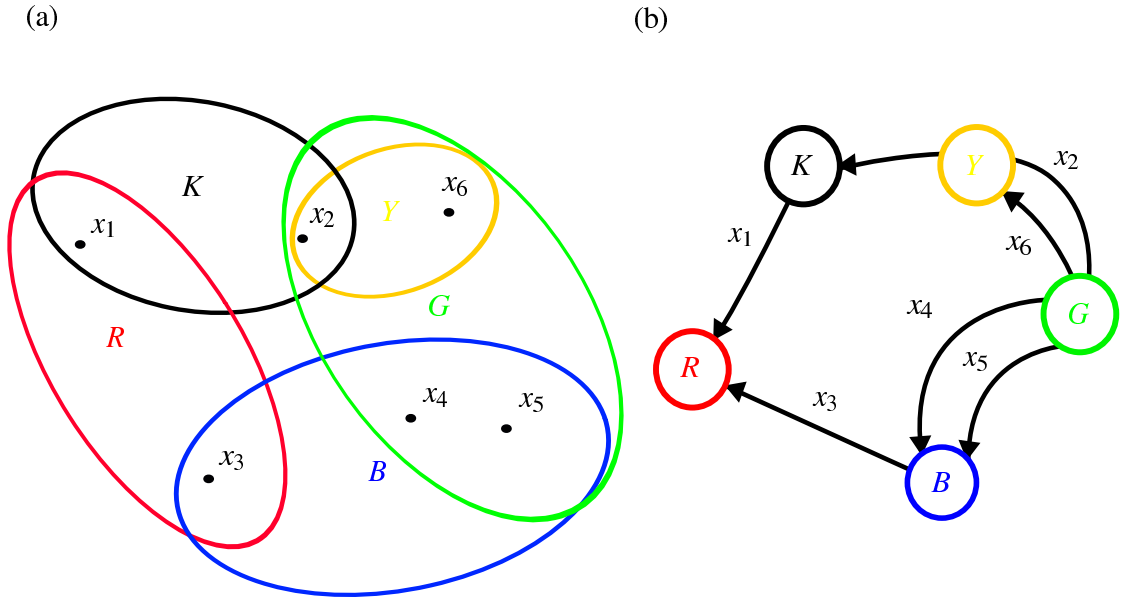


Figure 4.1: Illustration of the reduction of Set Cover to UME-Decision. (a) A set cover problem on elements $x_1 \dots x_6 \in X$ with subsets $K = \{x_1, x_2\}$, $R = \{x_1, x_3\}$, $B = \{x_3, x_4, x_5\}$, $G = \{x_2, x_4, x_5, x_6\}$, $Y = \{x_2, x_6\}$ contained in X . (b) The induced interdiction problem with each subset represented by a node and each element by an evader. Each arrow indicates the path of a single evader.

Proof. First we note that for a given a subset $Y \subseteq N$ with $|Y| \leq B$, we can update $\mathbf{M}^{(k)}$ and compute (1) to verify *UME-Decision* as a yes-instance. The number of steps is bounded by $O(|K||N|^3)$. Therefore, *UME-Decision* is in NP.

To show *UME-Decision* is NP-complete, reduce *Set Cover* with X, C to *UME-Decision* on a suitable graph $G(N, E)$. It is sufficient to consider just the special case where all interdiction efficiencies are equal, $d_i = 1$. For each $c \in C$, create a node c in N . We consider three cases for elements $x \in X$; elements that have no covering sets, elements that have one covering set, and elements that have at least two covering sets.

Consider first all $x \in X$ which have at least two covering sets. For each such x create

an evader as follows. Let O be any ordering of the collection of subsets covering x . Create in E a Hamiltonian path of $|O| - 1$ edges to join sequentially all the elements of O , assigning the start, a and end t nodes in agreement with the ordering of O . Construct an evader transition matrix of size $|C| \times |C|$ and give the evader transitions probability $M_{ij} = 1$ iff $i, j \in C$ and $i < j$, and $= 0$ otherwise.

For the case of zero covering sets, that is, where $\exists x \in X$ such that $x \notin S$ for all $S \in C$, represent x by an evader whose source and target are identical: no edges are added to E and the transition matrix is $\mathbf{M} = 0$. Thus, J in Eq. (4.2.3) is non-zero regardless of interdiction strategy.

For the case when x has just one covering set, that is, when $\exists x \in X$ such that there is a unique $c \in C$ with $x \in c$, represent c as two nodes i and j connected by an edge exactly as in the case of more than one cover above. After introducing j , add it to the middle of the path of each evader x if i is in the path of x , that is, if $c \in C$. It is equivalent to supposing that C contains another subset exactly like c . This supposition does not change the answer or the polynomial complexity of the given instance of *Set Cover*. To complete the reduction, set $B = \beta$, $\rho = 0$, $X = K$, $w^{(k)} = 1/|X|$ and $d_i = 1, \forall i \in N$.

Now assume *Set Cover* is a yes-instance with a cover $\hat{C} \subseteq C$. We set the interdicted transition matrix $\hat{M}_{ij}^{(k)} = 0$ for all $(i, j) \in E$ corresponding to $c \in \hat{C}$, and all $k \in K$. Since \hat{C} is a cover for X , all the created paths are disconnected, $\sum_{k \in K} (\mathbf{a}^{(k)} (\mathbf{I} - \hat{\mathbf{M}}^{(k)})^{-1})_{t^{(k)}} = 0$ and *UME-Decision* is an yes-instance.

Conversely, assume that *UME-Decision* is a yes-instance. Let Y be the set of interdicted nodes. For $y \in Y$, there is element y of C . Since all the evaders are disconnected from their target and each evader represents a element in X , $Y \subseteq C$ covers X and $|Y| \leq \beta$. Hence, *Set Cover* is a yes-instance. Therefore, *UME-Decision* is NP-complete. \square

□

This proof relies on multiple evaders and it remains an open problem to show that UME is NP-hard with just a single evader. We conjecture that the answer is positive because the more general problem of interdicting a single unreactive evader having an arbitrary (non-Markovian) path is NP-hard. This could be proved by creating from a single such evader several Markovian evaders such that the evader has an equal probability of following the path of each of the Markovian evaders in the proof above.

Thus far no consideration was given to the problem where the cost c_{ij} of interdicting an edge (i, j) is not fixed but rather is a function of the edge. This could be termed the “budgeted” case as opposed to the “unit cost” case discussed so far. However, the budgeted case is NP-hard as could be proved through reduction from the knapsack problem to a star network with “spokes” corresponding to items.

4.4 An Efficient Interdiction Algorithm

The solution to the UME problem can be efficiently approximated using a greedy algorithm by exploiting submodularity. In this section we prove that the UME problem is submodular, construct a greedy algorithm, and examine the algorithm’s performance. We then show how to improve the algorithm’s speed by further exploiting the submodular structure using a “priority” evaluation scheme and “fast initialization”.

4.4.1 Submodularity of the interdiction problem

In general, a function is called submodular if the rate of increase decreases monotonically, which is akin to concavity.

Definition 5. A real-valued function on a space S , $f : S \rightarrow \mathbb{R}$ is submodular [13, Prop. 2.1iii] if for any subsets $S_1 \subseteq S_2 \subset S$ and any $x \in S \setminus S_2$ it satisfies

$$f(S_1 \cup \{x\}) - f(S_1) \geq f(S_2 \cup \{x\}) - f(S_2). \quad (4.0)$$

Definition 6. $J(\mathbf{r})$ is submodular on the set of interdicted edges.

Proof. First, note that it is sufficient to consider a single evader because in Eq. (4.2.3), $J(\mathbf{r})$ is a convex combination of k evaders [13, Prop. 2.7]. For simplicity of notation, we drop the superscript k in the rest of the proof.

Let $S = \{(i, j) \in E \mid r_{ij} = 1\}$ be the interdiction set and let $J(S)$ be the probability of interdicting the evader using S , and let $Q(p)$ be the probability of the evader taking a path p to the target. On path p , the probability of interdicting the evader with an interdiction set S is

$$P(p|S) = Q(p) \left(1 - \prod_{(i,j) \in p \cap S} (1 - d_{ij}) \right). \quad (4.0)$$

Moreover,

$$J(S) = \sum_p P(p|S). \quad (4.0)$$

If an edge $(u, v) \notin S$ is added to the interdiction set S (assuming $(u, v) \in p$), the probability of interdicting the evader in path p increases by

$$P(p|S \cup \{(u, v)\}) - P(p|S) = Q(p) d_{uv} \prod_{(i,j) \in p \cap S} (1 - d_{ij}),$$

which can be viewed as the probability of taking the path p times the probability of being interdicted at (u, v) but not being interdicted elsewhere along p . If $(u, v) \in S$ or $(u, v) \notin p$ then adding (u, v) has, of course, no effect: $P(p|S \cup \{(u, v)\}) - P(p|S) = 0$.

Consider now two interdiction sets S_1 and S_2 such that $S_1 \subset S_2$. In the case where $(u, v) \notin S_1$ and $(u, v) \in p$, we have

$$P(p|S_1 \cup \{(u, v)\}) - P(p|S_1) = Q(p)d_{uv} \prod_{(i,j) \in p \cap S_1} (1 - d_{ij}), \quad (4.1)$$

$$\geq Q(p)d_{uv} \prod_{(i,j) \in p \cap S_2} (1 - d_{ij}), \quad (4.2)$$

$$\geq P(p|S_2 \cup \{(u, v)\}) - P(p|S_2). \quad (4.3)$$

In the above (4.2) holds because an edge $(u', v') \in (S_2 \setminus S_1) \cap p$ would contribute a factor of $(1 - d_{u'v'}) \leq 1$. The inequality (4.3) becomes an equality iff $(u, v) \notin S_2$. Overall (4.3) holds true for any path and becomes an equality when $(u, v) \in S_1$. Applying the sum of Eq. (4.4.1) gives

$$J(p|S_1 \cup \{(u, v)\}) - J(p|S_1) \geq J(p|S_2 \cup \{(u, v)\}) - J(p|S_2), \quad (4.3)$$

and therefore $J(S)$ is submodular. □

Note that the proof relies on the fact that the evader does not react to interdiction. If the evader did react then it would no longer be true in general that $P(p|S) = Q(p) \left(1 - \prod_{(i,j) \in p \cap S} (1 - d_{ij})\right)$ above. Instead, the product may show explicit dependence on paths other than p , or interdicted edges that are not on p . Also, when the evaders are not Markovian the proof is still valid because specifics of evader motion are contained in the function $Q(p)$.

4.4.2 Greedy algorithm

Submodularity has a number of important theoretical and algorithmic consequences. Suppose (as is likely in practice) that the edges are interdicted incrementally such that the interdiction set $S_l \supseteq S_{l-1}$ at every step l . Moreover, suppose at each step, the interdiction set S_l is grown by adding the one edge that gives the greatest increase in J . This defines a greedy algorithm, Alg. 4.

Algorithm 4: Greedy construction of the interdiction set S with budget B for a graph $G(N, E)$.

```
 $S \leftarrow \emptyset$   
while  $B > 0$  do  
   $x^* \leftarrow \emptyset$   
   $\delta^* \leftarrow -1$   
  for all  $x \in E \setminus S$  do  
     $\Delta(S, x) := J(S \cup \{x\}) - J(S)$   
    if  $\Delta(S, x) > \delta^*$  then  
       $x^* \leftarrow \{x\}$   
       $\delta^* \leftarrow \Delta(S, x)$   
   $S \leftarrow S \cup x^*$   
   $B \leftarrow B - 1$   
Output( $S$ )
```

The computational time is $O(B|N|^3|E|)$ for each evader, which is strongly polynomial since $|B| \leq |E|$. The linear growth in this bound as a function of the number of evaders could sometimes be significantly reduced. Suppose one is interested in interdicting flow $f(s, t)$ that has a small number of sources but a larger number of targets. In the current formulation the cost grows linearly in the number of targets (evaders) but

is independent of the number of sources. Therefore for this $f(s, t)$ it is advantageous to reformulate UME by inverting the source-target relationship by deriving a Markov process which describes how an evader moves from a given source s to each of the targets. In this formulation the cost would be independent of the number of targets and grow linearly in the number of sources.

4.4.3 Solution quality

The quality of the approximation can be bounded as a fraction of the optimal solution by exploiting the submodularity property [13]. In submodular set functions such as $J(S)$ there is an interference between the elements of S in the sense that sum of the individual contributions is greater than the contribution when part of S . Let S_B^* be the optimal interdiction set with a budget B and let S_B^g be the solution with a greedy algorithm. Consider just the first edge x_1 found by the greedy algorithm. By the design of the greedy algorithm the gain from x_1 is greater than the gain for all other edges y , including any of the edges in the optimal set S^* . It follows that

$$\Delta(\emptyset, x_1)B \geq \sum_{y \in S_B^*} \Delta(\emptyset, y) \geq J(S_B^*). \quad (4.3)$$

Thus x_1 provides a gain greater than the average gain for all the edges in S_B^* ,

$$\Delta(\emptyset, x_1) \geq \frac{J(S_B^*)}{B}. \quad (4.3)$$

A similar argument for the rest of the edges in S_B^g gives the bound,

$$J(S_B^g) \geq \left(1 - \frac{1}{e}\right) J(S_B^*), \quad (4.3)$$

where e is Euler's constant [13, p.268]. Hence, the greedy algorithm achieves at least 63% of the optimal solution.

This performance bound depends on the assumption that the cost of an edge is a constant. Fortunately, good discrete optimization algorithms for submodular functions are known even for the case where the cost of an element (here, an edge) is variable. These algorithms are generalizations of the simple greedy algorithm and provide a constant-factor approximation to the optimum [14, 15]. Moreover, for any particular instance of the problem one can bound the approximation ratio, and such an “online” bound is often better than the “offline” *a priori* bound [16].

4.4.4 Exploiting submodularity with Priority Evaluation

In addition to its theoretical utility, submodularity can be exploited to compute the same solution much faster using a priority evaluation scheme. The basic greedy algorithm recomputes the objective function change $\Delta(S_l, x)$ for each edge $x \in E \setminus S_l$ at each step l . Submodularity, however, implies that the gain $\Delta(S_l, x)$ from adding any edge x would be less than or equal to the gain $\Delta(S_k, x)$ computed at any earlier step $k < l$. Therefore, if at step l for some edge x' , we find that $\Delta(S_l, x') \geq \Delta(S_k, x)$ for all x and any past step $k \leq l$, then x' is the optimal edge at step l ; there is no need for further computation (as was suggested in a different context [16].) In other words, one can use stale values of $\Delta(S_k, x)$ to prove that x' is optimal at step l .

As a result, it may not be necessary to compute $\Delta(S_l, x)$ for all edges $x \in E \setminus S$ at every iteration. Rather, the computation should prioritize the edges in descending order of $\Delta(S_l, x)$. This “lazy” evaluation algorithm is easily implemented with a priority queue which stores the gain $\Delta(S_k, x)$ and k for each edge where k is the step at which it was last calculated. (The step information k determines whether the value is stale.)

The priority algorithm (Alg. 5) combines lazy evaluation with the following fast

initialization step. Unlike in other submodular problems, in UME one can compute $\Delta(\emptyset, x)$ simultaneously for all edges $x \in E$ because in this initial step, $\Delta(\emptyset, x)$ is just the probability of transition through edge x multiplied by the interdiction efficiency d_x , and the former could be found for all edges in just one operation. For the “non-retreating” model of Ref. [10] the probability of transition through $x = (i, j)$ is just the expected number of transitions through x because in that model an evader moves through x at most once. This expectation is given by the i, j element in $\mathbf{a}(\mathbf{I} - \mathbf{M})^{-1} \odot \mathbf{M}$ (derived from Eq. (4.2.3)). The probability is multiplied by the weight of the evader and then by d_x : $\Delta(\emptyset, x) = \sum_k \left(\mathbf{a}^{(k)}(\mathbf{I} - \mathbf{M}^{(k)})^{-1} \right)_i M_{ij}^{(k)} w^{(k)} d_x$. In addition to these increments, for disconnected graphs the objective $J(S)$ also contains the constant term $\sum_k w^{(k)} \left(\sum_{i \in Z^{(k)}} a_i \right)$, where $Z^{(k)} \subset N$ are nodes from which evader k cannot reach his target $t^{(k)}$.

In subsequent steps this formula is no longer valid because interdiction of x may reduce the probability of motion through other interdicted edges. Fortunately, in many instances of the problem the initialization is the most expensive step since it involves computing the cost for all edges in the graph. As a result of the two speedups the number of cost evaluations could theoretically be linear in the budget and the number of evaders and independent of the size of the solution space (the number of edges).

The performance gain from priority evaluation can be very significant. In many computational experiments, the second best edge from the previous step was the best in the current step, and frequently only a small fraction of the edges had to be recomputed at each iteration. In order to systematically gauge the improvement in performance, the algorithm was tested on 50 synthetic interdiction problems. In each case, the underlying graph was a 100-node Geographical Threshold Graph (GTG), a possible model of sensor or transportation networks [17], with approximately 1600 directed edges (the threshold parameter was set at $\theta = 30$). Most of the networks were connected. We set the cost of

Algorithm 5: Priority greedy construction of the interdiction set S with budget B

```
 $S \leftarrow \emptyset$   
 $PQ \leftarrow \emptyset$  {Priority Queue:  $(value, data, data)$ }  
for all  $x = (i, j) \in E$  do  
     $\Delta(x) \leftarrow$  {The cost found using fast initialization}  
     $PUSH(PQ, (\Delta(x), x, 0))$   
 $s \leftarrow 0$   
while  $B > 0$  do  
     $s \leftarrow s + 1$   
    loop  
         $(\Delta(x), x, n) \leftarrow POP(PQ)$   
        if  $n = s$  then  
             $S \leftarrow S \cup \{x\}$   
            break  
        else  
             $\Delta(x) \leftarrow J(S \cup \{x\}) - J(S)$   
             $PUSH(PQ, (\Delta(x), x, s))$   
     $B \leftarrow B - 1$   
Output( $S$ )
```

traversing an edge to 1, the interdiction efficiency d_x to 0.5, $\forall x \in E$, and the budget to 10. We used two evaders with uniformly distributed source nodes based on the model of [10] with an equal mixture of $\lambda = 0.1$ and $\lambda = 1000$. For this instance of the problem the priority algorithm required an average of 29.9 evaluations of the objective as compared to 31885.2 in the basic greedy algorithm - a factor of 1067.1 speedup.

The two algorithms find the same solution, but the basic greedy algorithm needs to

recompute the gain for all edges uninterdicted edges at every iteration, while the priority algorithm can exploit fast initialization and stale computational values. Consequently, the former algorithm uses approximately $B|E|$ cost computations, while the latter typically uses much fewer (Fig. 4.2a).

Simulations show that for the priority algorithm the number of edges did not seem to affect the number of cost computations (Fig. 4.2b), in agreement with the theoretical limit. Indeed, the only lower bound for the number of cost computations is B and this bound is tight (consider a graph with B evaders each of which has a distinct target separated from each evader’s source by exactly one edge of sufficiently small cost). The priority algorithm performance gains were also observed in other example networks.³

The priority algorithm surpasses a benchmark solution of the corresponding mixed integer program (See Appendix) using a MIP solver running CPLEX (version 10.1) in consistency, time, and space. For example, in runs on 100-node GTG networks with 4 evaders and a budget of 10, the priority algorithm terminates in 1 to 20 seconds, while CPLEX terminated in times ranging from under 1 second to 9.75 hours (the high variance in CPLEX run times, even on small problems, made systematic comparison difficult.) The difference in solution optimality was zero in the majority of runs. In the hardest problem we found (in terms of its CPLEX computational time - 9.75 hours), the priority algorithm found a solution at 75% of the optimum in less than 10 seconds.

For our implementation, memory usage in the priority algorithm never exceeded 300MiB. Further improvement could be made by re-implementing the priority algorithm so that it would require only order $O(|E|)$ to store both the priority queue and the vectors of Eq. (4.2.3). In contrast, the implementation in CPLEX repeatedly used over 1GiB for

³Specifically, the simulations were a two evader problem on a grid-like networks consisting of a lattice (whose dimensions were grown from 8-by-8 to 16-by-16) with random edges added at every node. The number of edges in the networks grew from approximately 380 to 1530 but there was no increasing trend in the number of cost evaluations.

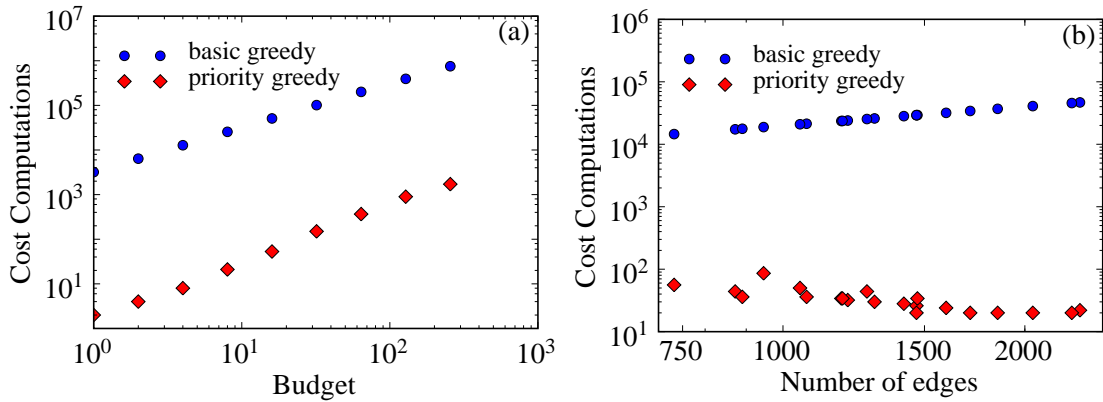


Figure 4.2: Comparison between the basic greedy (blue circles) and the priority greedy algorithms (red diamonds) for the number of cost evaluations as a function of (a) budget, and (b) number of edges. In (a) each point is the average of 50 network interdiction problems. The average coefficient of variation (the ratio of the standard deviation to the mean) is 0.10 for basic greedy and 0.15 for the priority greedy. Notice the almost perfectly linear trends as a function of budget (shown here on a log-log scale, the power ≈ 1.0 in both.) In (b), the budget was fixed at 10 and the number of edges was increased by decreasing the connectivity threshold parameter from $\theta = 50$ to $\theta = 20$ to represent, e.g., increasingly dense transportation networks.

the search tree. As was suggested from the complexity proof, in runs where the number of evaders was increased from 2 to 4 the computational time for an exact solution grew rapidly.

4.5 Outlook

The submodularity property of the UME problem provides a rich source for algorithmic improvement. In particular, there is room for more efficient approximation schemes and practical value in their invention. Simultaneously, it would be interesting to classify

the UME problem into a known approximability class. It would also be valuable to investigate various trade-offs in the interdiction problem, such as the trade-off between quality and quantity of interdiction devices.

As well, to our knowledge little is known about the accuracy of the assumptions of the unreactive Markovian model or of the standard max-min model in various applications. The detailed nature of any real instance of network interdiction would determine which of the two formulations is more appropriate.

Acknowledgments

AG would like to thank Jon Kleinberg for inspiring lectures, David Shmoys for a helpful discussion and assistance with software, and Vadas Gintautas for support. Part of this work was funded by the Department of Energy at Los Alamos National Laboratory under contract DE-AC52-06NA25396 through the Laboratory Directed Research and Development Program.

Postscript

After the original paper appeared in print the authors became aware of a prior proposal to exploit submodularity to monitor Markovian network traffic (Berman et al. [18]). Even with that, the current work represents a clear advance in terms of the underlying model and the priority algorithm, that far outperforms the running time of the basic greedy algorithm used in [18].

APPENDIX

4.A Mixed integer program for UME

In the unreactive Markovian evader interdiction (UME) problem an evader $k \in K$ is sampled from a source distribution $\mathbf{a}^{(k)}$, and moves to a sink $t^{(k)}$ with a path specified by the matrix $\mathbf{M}^{(k)}$. This matrix is the Markov transition matrix with zeros in the row of the absorbing state (sink). The probability that the evader arrives at $t^{(k)}$ is $(\mathbf{a}^{(k)}(\mathbf{I} - \mathbf{M}^{(k)})^{-1})_{t^{(k)}}$ and is 1 without any interdiction (removal of edges).

Notation summary

$G(N, E)$: simple graph with node and edge sets N and E , respectively.

K : the set of evaders.

$w^{(k)}$: probability that the evader k occurs.

$a_i^{(k)}$: probability that node i is the source node of evader k .

$t^{(k)}$: the sink of evader k .

$\mathbf{M}^{(k)}$: the modified transition matrix for the evader k .

d_{ij} : the conditional probability that interdiction of edge (i, j) would remove an evader who traverses it.

B : the interdiction budget.

$\pi_i^{(k)}$: decision variable on conditional probability of node evader k traversing node i .

r_{ij} : interdiction decision variable, 1 if edge (i, j) is interdicted and 0 otherwise.

Definition 7. Unreactive Markovian Evader interdiction (UME) problem

$$\begin{aligned} \min_{\mathbf{r}} \quad & H(\mathbf{r}) = \sum_{k \in K} w^{(k)} h^{(k)}(\mathbf{r}), \\ \text{s.t.} \quad & \sum_{(i,j) \in E} r_{ij} = B, \\ & r_{ij} \in \{0, 1\}, \quad \forall (i, j) \in E, \end{aligned}$$

where

$$h^{(k)}(\mathbf{r}) = \min_{\pi} \pi_i^{(k)}, \quad (4.0)$$

$$\text{s.t.} \quad \pi_i^{(k)} - \sum_{(j,i) \in E} (M_{ji}^{(k)} - M_{ji}^{(k)} d_{ji} r_{ji}) \pi_j^{(k)} = a_i^{(k)}, \quad \forall i \in N, \quad (4.0)$$

$$\pi_i^{(k)} \geq 0, \quad \forall i \in N. \quad (4.1)$$

The constraint (4.0) is nonlinear. We can replace this with a set of linear constraints, and the evader problem becomes

$$h^{(k)}(\mathbf{r}) = \min_{\pi, \theta} \pi_i^{(k)}, \quad (4.2a)$$

$$\text{s.t.} \quad \pi_i^{(k)} - \sum_{(j,i) \in E} \theta_{ji}^{(k)} = a_i^{(k)}, \quad \forall i \in N, \quad (4.2b)$$

$$\theta_{ji}^{(k)} \geq M_{ji}^{(k)} \pi_j^{(k)} - M_{ji}^{(k)} d_{ji} r_{ji}, \quad \forall (j,i) \in E, \quad (4.2c)$$

$$\theta_{ji}^{(k)} \geq M_{ji}^{(k)} (1 - d_{ji}) \pi_j^{(k)}, \quad \forall (j,i) \in E, \quad (4.2d)$$

$$\theta_{ij}^{(k)} \geq 0, \quad \forall (i,j) \in E, \quad (4.2e)$$

$$\pi_i^{(k)} \geq 0, \quad \forall i \in N. \quad (4.2f)$$

$$(4.2g)$$

If we set $r_{ij} = 0$, the constraint (4.2c) is dominating (4.2d), and θ_{ij} will take value $M_{ij}^{(k)} \pi_i^{(k)}$ at optimal because of the minimization. If we set $r_{ij} = 1$, the constraint (4.2d) is dominating since $\pi_j^{(k)} \leq 1$. Although formulation (4.2) has an additional variable θ , at the optimum the two formulations are equivalent because π and \mathbf{r} have the same values.

REFERENCES

- [1] H. W. Corley and D. Y. Sha, “Most vital links and nodes in weighted networks.,” *Oper. Res. Lett.*, vol. 1, pp. 157 – 160, Sep 1982.
- [2] A. W. McMasters and T. M. Mustin, “Optimal interdiction of a supply network,” *Naval Research Logistics Quarterly*, vol. 17, no. 3, pp. 261–268, 1970.
- [3] P. M. Ghare, D. C. Montgomery, and W. C. Turner, “Optimal interdiction policy for a flow network,” *Naval Research Logistics Quarterly*, vol. 18, no. 1, pp. 37–45, 1971.
- [4] B. Pourbohloul, L. Meyers, D. Skowronski, M. Krajdén, D. Patrick, and R. Brunham, “Modeling control strategies of respiratory pathogens,” *Emerg. Infect. Dis.*, vol. 11, no. 8, pp. 1246–56, 2005.
- [5] J. D. Farley, “Breaking Al Qaeda cells: A mathematical analysis of counterterrorism operations (a guide for risk assessment and decision making),” *Studies in Conflict and Terrorism*, vol. 26, pp. 399–411, 2003.
- [6] F. Pan, W. Charlton, and D. P. Morton, “Interdicting smuggled nuclear material,” in *Network Interdiction and Stochastic Integer Programming* (D. Woodruff, ed.), pp. 1–19, Kluwer Academic Publishers, Boston, 2003.
- [7] M. O. Ball, B. L. Golden, and R. V. Vohra, “Finding the most vital arcs in a network,” *Oper. Res. Lett.*, vol. 8, no. 2, pp. 73–76, 1989.
- [8] A. Bar-Noy, S. Khuller, and B. Schieber, “The complexity of finding most vital arcs and nodes,” tech. rep., University of Maryland, College Park, MD, USA, 1995.
- [9] E. Boros, K. Borys, and V. Gurevich, “Inapproximability bounds for shortest-path network interdiction problems,” tech. rep., Rutgers University, Piscataway, NJ, USA, 2006.
- [10] A. Gutfraind, A. Hagberg, D. Izraelevitz, and F. Pan, “Interdicting a Markovian evader.” Preprint, 2009.
- [11] C. M. Grinstead and J. L. Snell, *Introduction to Probability*. USA: American Mathematical Society, second revised ed., Jul 1997.

- [12] R. M. Karp, “Reducibility among combinatorial problems,” in *Complexity of Computer Computations* (R. E. Miller and J. W. Thatcher, eds.), pp. 85–103, New York: Plenum, 1972.
- [13] G. Nemhauser, L. Wolsey, and M. Fisher, “An analysis of the approximations for maximizing submodular set functions-I,” *Mathematical Programming*, vol. 14, pp. 265–294, 1978.
- [14] S. Khuller, A. Moss, and J. S. Naor, “The budgeted maximum coverage problem,” *Information Processing Letters*, vol. 70, no. 1, pp. 39–45, 1999.
- [15] A. Krause and C. Guestrin, “A note on the budgeted maximization on submodular functions,” tech. rep., Carnegie Mellon University, 2005. CMU-CALD-05-103.
- [16] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, “Cost-effective outbreak detection in networks,” in *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, (New York, NY, USA), pp. 420–429, ACM, 2007.
- [17] M. Bradonjić and J. S. Kong, “Wireless ad hoc networks with tunable topology,” in *Forty-Fifth Annual Allerton Conference*, (UIUC, Illinois, USA), pp. 1170–1177, 2007.
- [18] D. K. Oded Berman and C. W. Xu, “Locating flow-intercepting facilities: New approaches and results,” *Annals of Operations Research*, vol. 60, pp. 121–143, 1995.

CHAPTER 5

DESIGNING NETWORKS FOR CASCADE RESILIENCE

Chapter Abstract Complex socio-economic networks such as transportation networks, information systems and even underground organizations are often designed for resilience - to be able to function even if some of the nodes (depots, routers, operatives etc.) are compromised by a human or natural adversary. In many cases the adversary threatens to cause a cascade where the failure of a single node leads to far-reaching domino effect where some of the adjacent nodes are lost as well, followed by their neighbors and so on. Such cascades motivate the search for mechanisms and network designs that would increase the networks' cascade resilience while maintaining efficiency. This work introduces a mathematical model in which networks are designed by solving an optimization problem. The results indicate that a network consisting of multiple star-like cells maximizes a combination of cascade resilience and efficiency. Also, perhaps surprisingly it was found that in many network designs and parameter values, edge density of the optimal network topology does not monotonically decrease when cascade risk increases, but may increase again when cascade risk is high. This implies that certain networks ought not to be modified for cascade resilience, since the cost in efficiency is too high. Understanding cascade resilience and its structural phase transitions will help identify vulnerabilities in violent underground groups, but also design more resilient networks in many diverse application areas.

Keywords: networks, resilience, cascade, contagion, epidemics on networks, terrorism, terrorist networks

5.1 Introduction

Cascades on networks have been a major theme in the field of network research. For example, in a power grid the loss of a single transmission or generator node may cause nearby nodes to be overloaded, becoming disconnected or damaged. This failures might propagate widely through the network, leading to widespread blackouts and large economic damage. Rebuilding the network might involve both repair of hardware and complex start-up procedures that take considerable time. These cascade phenomena are also found in other domains such as disease control, computer networks, financial markets and social systems. A particularly interesting case is clandestine social networks, such as terrorist networks or guerrillas operating in a hostile environment. If one of the nodes (i.e. operatives) is captured by law enforcement agencies, it may betray all the nodes connected to it leading to their likely capture.

The focus of this work is to investigate a novel problem in network research - how to build cascade-resilient networks. To date much research has explored the extent of cascades and the nature of their propagation, specifically looking at important classes of networks [1, 2, 3, 4, 5]. Here the focus is different in that the topology is not fixed because in many networks cascade resilience is a design criterion. Here “resilience” refers to properties of the network that reduce the damage caused by cascades (also called “contagions” or “epidemics”). The objective here is to identify topological features that can endow networks with high contagion resilience. It is hoped that ultimately it would be possible to identify a general prescription for building cascade-resilient networks for many different domains. Finding such features would be useful not only in networks that are being designed *de novo*, but also in a much broader class of networks where some changes could be made to the topology even if complete redesign is infeasible. This include power grids (which could be locally upgraded), social networks facing epidemics

(through isolation of certain nodes) and others.

Because the relative importance of efficiency and resilience depends on operating conditions, the optimal design is expected to be not a single pattern, but multiple different patterns, with possible sharp transitions between them. In research of terrorism and guerrilla movements, a classic pattern is the tree-of-cliques cellular hierarchy (Fig.5.1) [6, 7] (see also research on crime networks in industry [8], and drugs [9]). The advantages of networks structured around independent cells are well-recognized by the terrorist groups themselves, and have already been explored by researchers [10, 11]. In this application area, the current research promises to provide better understanding of the structure of terrorist networks and in particular, it will help predict which networks would be very difficult to interdict.

However, it is clear that the “optimal” topology depends on conditions such as the risk of cascades and the purpose of the network (e.g. organize violence vs. transport electric power), whose effect is poorly-understood.

The model introduced in this paper addresses this problem in a simplified context of graph theory, as follows. It is clear that in many networks it is possible to increase cascade resilience by many means, not only through their topology. For instance, in controlling respiratory diseases, it is possible to ask people to wear face mask reducing the spread of contagions. Also, in clandestine social networks one sees additional types of nodes (“dead drops” or “couriers”) whose purpose is to prevent cascades. Those types of defenses are interesting both in practice and in theory, but they involve heterogeneous graphs (with multiple types of nodes and edges) whose models are both more complex and application-specific. Therefore here the focus is on simple graphs as models of networks in the view that the conclusions derived from such models would also be applicable to more complex situations. In the remainder, the words “network” and “graph”

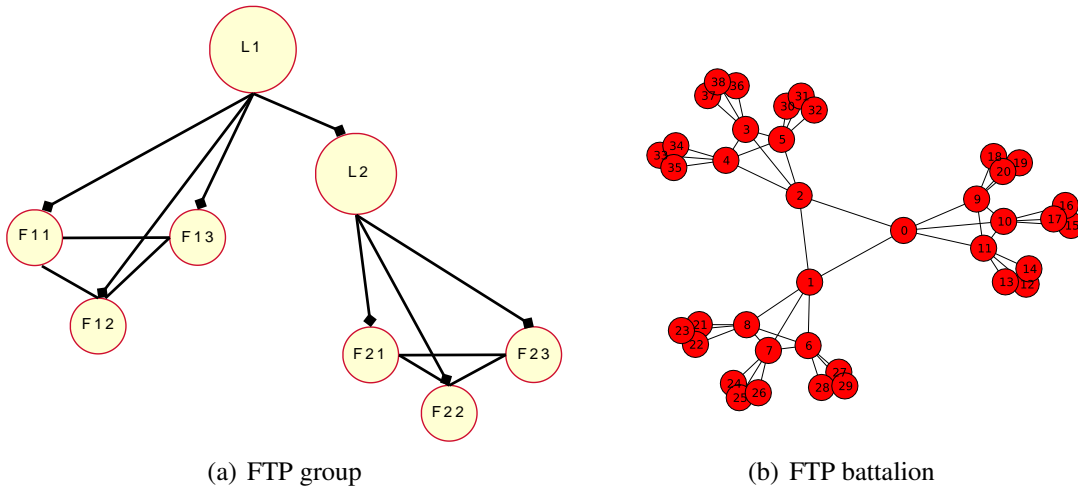


Figure 5.1: The French World-War II underground network *Francs-tireurs Partisans* (FTP). Its basic organizational unit was the combat group (a). This was divided into two “teams” of three fighters, where leader $L1$ was in overall command and in command of team 1. His lieutenant, $L2$, led team 2 and assumed overall command if $L1$ was captured. The small degree of the nodes ensured that the capture of any one node did not cause a significant fraction of the organization becoming captured as well. Such groups were organized into a hierarchy (b) where 3 groups made a “section”, 3 of which made a “company”, and finally 3 companies made a “battalion”. In the battalion figure, a leaf node corresponds to the leader of a group (subordinates not shown).

G will be used to mean the same object: a tuple (V, E) , where V is a set of “nodes” and E is a set of “edges”, where each edge is an unordered pair of nodes.

Even on simple graphs, designing resilience to cascades is not a simple problem to formulate because in practical network designs it is necessary to balance resilience with suitably-defined performance/efficiency. Indeed, intuitively the most cascade-resilient network is the network with no edges (no cascades can propagate), but it is also the least useful kind of network. Therefore, searching for the most resilient design is not the right objective, neither in practice nor in theory. Rather the true objective is to maximize a certain combination of resilience and efficiency, which is termed “fitness”. It is expected

that typically maximizing resilience and maximizing efficiency will be in opposition, requiring tradeoffs. Just as disconnected networks are resilient and inefficient, highly-efficient networks such as densely-connected graphs are likely to have low resilience [12].

Such a trade-off suggests formulating the question as one of mixed-objective optimization, where the solution space is a space of simple graphs possibly with a fixed number nodes. In this formulation one must overcome two keys problems. First, this is a very large search space even if the graph has just a few dozen nodes or edges. Secondly, it is important to be able to smoothly measure efficiency in both connected and disconnected topologies but many familiar graph functions are only suitable for only connected topologies. Both of those issues are addressed here: to reduce the search space the graphs are constructed using parametrized generating programs, and efficiency is measured by a metric termed “distance-attenuated reach”.

Research on social networks indicates that resilience and efficiency might be just two of several design criteria that also include e.g. “information-processing requirements”, that impose additional constraints on network designs [8]. In the original context “information-processing” refers to the need to have ties between individuals involved in a particular task, when the task has high complexity. Each individual might have a unique set of expertise into which all the other agents must tap directly. Generalizing from sociology, such “functional constraints” might considerably limit the flexibility in constructing resilient and efficient networks. For example, in the context of terrorism, this constraint significantly decreased the quality of attacks that could be successfully carried out in the post 9/11 security environment [13]. Such functional constraints could be addressed by looking at a narrow set of models of networks which already incorporate such constraints. The specific models to be examined are motivated

by analytical expediency, but in general one may want to consider a particular design palette dictated by the application at hand. The current objective is to identify which elements might be useful for increasing the fitness of networks, and thus suggest elements that are useful to have in the palette.

There is a very extensive literature on both cascades and resilience. For instance, a number of investigations considered resilience to removal of nodes or edges. It was realized that certain types of scale-free networks (but not others) are sensitive to targeted node removal [14, 15, 16, 17, 18, 19]. Researchers also looked at different models of contagion [1, 20], as well as non-topological mechanisms for increasing resilience [21, 22]. The general area of resilience has attracted a lot of research in the area of secret societies such as terrorist networks [10, 12, 23, 24, 9, 25, 26, 11]. In fact many secret societies are benign, including non-governmental organizations and dissident movements operating in hostile political environments. Related problems have also been studied in epidemiology, where the question focused on immunization strategies (e.g. [27]) but apparently not as a question of optimal network design. Game-theoretic methods have recently been applied to the resilience-efficiency trade-off in terrorist network design [26, 11].

The main contribution of this work is to systematically attack the cascade resilience problem of networks. Also novel is a metric for defining efficiency of networks, which enables studying disconnected topologies in detail.

The paper is organized in the following way: section 2 formalizes the problem mathematically and section 3 introduces various classes of networks. The results are exhibited in section 4, and discussed in section 5. Mathematical details are found in the Appendix.

5.2 Formal Model

The novel approach here is to represent the resilience-efficiency trade-off above as an optimization problem where the decision variable is the topology G of a simple graph and the objective function is a metric overall performance termed “fitness”, $F(G)$. The graph G is chosen from a set \mathbb{G} , and fitness combines resilience $R(G)$ and efficiency $W(G)$, through a weight parameter r :

$$\max_{G \in \mathbb{G}} \left[\underbrace{rR(G) + (1-r)W(G)}_{F(G)} \right]. \quad (5.0)$$

The set \mathbb{G} is here called the “design” of the network. This design could be quite narrow (e.g. an Erdos-Renyi random graph on n nodes [28]) or very broad (e.g. any n -node graph). Both of the functions $R(G)$ and $W(G)$ (W stands for “work”) will have range $\subseteq [0, 1]$, to make them independent of network size. The parameter $r \in [0, 1]$ depends on the problem being studied. Much of the discussion below will be about the point $r = 0.5$ in which the two are weighed equally. On a practical level, r could represent the cost of restoring the network after a cascade - is it light or catastrophic. In practice r is determined by considering the strategic purpose of the network being designed, and making an estimate of the best balance point between the two objectives.

The next subsection introduces two simple measures of $R(G)$ and $W(G)$ that should be applicable to a range of different scenarios. These measures are necessary to make progress, but of course the general multi-objective formulation (5.2) has much broader applicability. Another issue to be addressed shortly is defining a suitable search space of graphs \mathbb{G} (see subsec. 5.3).

5.2.1 Measuring Resilience

In order to define $R(G)$ one needs to provide (a) a realistic model of contagions, and (b) a metric of contagion resilience relevant to that model. While research on graph theory has led to the development of a variety of metrics of robustness or resilience [7], here, unlike in many other studies the interest is in resilience to cascades not to e.g. disconnection.

One particularly important class of cascades are those that start at a single node then spread probabilistically to neighboring nodes possibly reaching a large fraction of the network. Under this model, a very natural definition of resilience is the expected size of the surviving network. To make the definition neutral with respect to the size of the original network, the quantity is normalized:

$$R(G) = 1 - \frac{1}{n-1} E[\text{extent of a cascade}]. \quad (5.0)$$

For simplicity, assume that cascades start at all nodes with uniform probability. It would be easy to extend this to cases where factors such as graph topology, node degree and even node type (for heterogeneous graphs) play a role. Note that the definition considers the case where cascades spread beyond the immediate neighbors of the starting node, since this possibility is important in practice.

A simple model of a contagion is the following “SIR” model [29]: each node in the graph can be in one of three states “susceptible”, “infected” and “removed” designated S, I and R respectively (these names are borrowed from epidemiology). Time is assumed to move in uniform discrete steps. A node in S state at time t stays in this state, unless a neighbor “infects” the node, causing it to move to state I at time $t + 1$. Specifically, a node in state S at time t has a node-independent probability τ of turning to I state at time $t + 1$ if an adjacent node is at state I at time t . Finally, a node in I state at time t would

always become R at time $t + 1$. Once in state R , a node would remain there for all future times. In general the rate of transition $I \rightarrow R$ could take more than one time step but adding this effect would mostly serve to increase the probability of transmission [29], which is already parametrized by τ .

5.2.2 Measuring Efficiency

The efficiency of a network is of course completely dependent on its ultimate function. Ideally, the efficiency metric would (a) be general enough for a variety problems, (b) be suitable for both connected, weakly-connected and disconnected networks and (c) computationally and analytically simple. Arguably for many applications the distance between pairs of nodes in the network is the most important determinant of the network's efficiency. This idea motivates the following "distance-attenuated reach" metric of efficiency, which gives the average neighborhood size of each node, corrected by the distance to the nodes in this neighborhood. Namely, for all nodes $u \in V$, weigh each $v \in V \setminus \{u\}$ by the inverse of its distance to u :

$$W(G) = \frac{1}{n(n-1)} \sum_{u \in V} \sum_{v \in V \setminus \{u\}} \frac{1}{d(u,v)^g}, \quad (5.0)$$

where $g > 0$ is a parameter and normalization by $n(n-1)$ ensures that $0 \leq W(G) \leq 1$. As usual, for any node v disconnected from u , set $\frac{1}{d(u,v)^g} = 0$. An equivalent formula is the following: if $V_{u,d}$ is the set of nodes around u at distance d from u (d goes from 1 to ∞), then $W(G) = \frac{1}{n(n-1)} \sum_{u \in V} \sum_{d=1}^{\infty} \frac{|V_{u,d}|}{d^g}$. Thus, if the network has short paths, then the metric would approach 1.0 because $V_{u,d}$ would be large for small d , and if it has long paths or is disconnected, then $V_{u,d}$ would be small for small d . (One may even generalize this metric to replace d^g with $\frac{d^g}{d^g + D}$ for some $D > 0$ to take into account the possibility of connecting to nodes through a costly alternative route that bypasses the given network.)

The parameter g could be termed “connectivity attenuation“ of the network because it represents the rate at which distance decreases the connectivity between nodes. In some problems (such as the Internet) the presence of v on the same connected component as u is completely sufficient for providing the services of v to u (such as serving documents), implying that $g \rightarrow 0$. In other problems (such as trust networks) one can trust only one’s friends and much less their friends, corresponding to attenuation $g \gg 1$. Attenuation is expected to have a significant effect on the optimization problem because it is hard to build resilient networks when the attenuation is rapid because to decrease cascade risk one cannot reduce the density of edges, as such a reduction would radically reduce efficiency $W(G)$.

The above definitions of efficiency and resilience provide the following intuition about the optimal design of a network. Since edges increase efficiency of the graph, as the probability of cascades decreases ($\tau \rightarrow 0$) the optimal network would grow more dense in order to reduce the average distance on the network. However, since cascades propagate through edges, as $\tau \rightarrow 1$, the optimal networks would become more sparse in order to maximize resilience at the expense of efficiency. As would be shown in the results section, this intuition is often incorrect.

In general, the current modeling approach has close parallels in the work of Lindelauf et al. on terrorist networks [26, 11]. Like here, they consider two optimization criteria: “secrecy” and “information”, corresponding to resilience and efficiency. Similar to this paper, their secrecy metric is based on the principle that the capture of a node will lead to the loss of their immediate neighbors, with some probability. The issue of cascades is not considered, presumably to achieve analytic tractability. Similar to here, the metric of information is defined through distances between pairs of nodes, but they take the inverse of the average distance (giving 0 when the network is disconnected).

Most different is the interesting application of game theory to find the optimal network. Their network is the Nash bargaining solution involving a “secrecy player” and an “information player”. This could be viewed as a type of multi-objective optimization, and is conceptually similar to the current approach if one sets r to 0.5. Their work is distinguished in that they propose a rich set of metrics for secrecy corresponding to different adversarial scenarios and derive a rich set of analytic results, whereas here just one metric is used and the main tool is simulation (with some analytic results). The current work is also related to [30], who explored the interesting question of cascades due to overload of node capacities, but the results are suspect since local search heuristics were applied to search for the optimum.

5.3 Network Designs

The optimization problem above faces the difficult obstacle of a large solution space. For general graphs on say n nodes, the set of possible solutions to the above problem is exponential in n as $n \rightarrow \infty$. Therefore, any practical approach must restrict the space to a small subset. Moreover, given a set of highly-rated networks (eg. top 1%), it may be challenging to characterize them - to identify what features give those particular networks their desirable properties.

5.3.1 Solution Approach

In order to solve those two problems, this work considers what is arguably the most important subset of the search space. Namely, the focus is on the set of all graphs on n nodes which can be constructed using a number of simple models termed “network

designs”. Each model contains parameters which specify how the network is to be generated, where each setting of the parameters is termed “configuration”. Thus, instead of searching through the space of graphs, the search is through the set of programs generating graphs from a model, or more concretely, the search is through the set of parameters that control those programs. A graph-generating program is an analog of instructions or protocols by which real networks are constructed. Practical limitations prevent those instructions from being complex, and hence the set of graphs constructed by such a program is the more relevant search space for practical applications, than e.g. the set of all graphs on n nodes.

The optimization process is similar to evolution where selection occurs on the phenotype of organisms - here: graphs - but the organisms are largely specified by their genotype - here: the designs and their configuration parameters. In other words, each design D has configurations C_1^D, C_2^D, \dots . In experiments each configuration C_i^D is inputted to a program that generates a sample of networks, whose average performance provides an estimate of the fitness of C_i^D (see Appendix, sec. 5.C for details).

As to the characterization problem, given that one would know how the networks were generated, one can more easily characterize the optimal configuration simply by looking at the parameter values of the programs that generated them. Implicitly, this procedure assumes that for a given set of parameter values all of the models/programs produce similar networks, as far as efficiency and resilience are concerned. Hence, the words “network” and “configuration” will be used interchangeably, even though the former refers to a single graph, while the latter to a class graphs generated using the same process.

5.3.2 Network Designs

A variety of graph generating models have been proposed to describe empirical networks: Erdos-Renyi (ER) [31], small world [32], preferential attachment [33], and many others. However, the models recreate networks whose construction principles are quite different from what is needed in cascade-resilient networks. In particular, many models produce graphs with a relatively large number of high degree nodes - a feature that strongly promotes cascades [2] and under certain conditions even facilitates epidemics that sweep most of the nodes in the network [5]. While in some applications, such as scientific collaboration networks, cascades are desirable because they spread knowledge, the current investigation explores cascade-retarding networks.

Fortunately, a good source of suitable networks is found in studies of clandestine networks. It is known that terrorist networks are often partitioned into cells which operate largely independently of each other. Moreover, the leadership of the terrorist group is often not even in direct contact with the cells, instead providing strategic guidance through public forums (see e.g.[34, 35]). These ideas of decentralization are of course found in other applications such as the electronic communication networks and organizational design theory.

Let us then consider networks on n nodes constructed through 6 simple designs, chosen both based on empirical findings as well as the possibility of rigorous analysis in some cases (see Appendix, sec. 5.D). These designs provide both a mathematical simplification of the network optimization problem and also structural motifs that could be incorporated in networks, where it is desired to achieve higher cascade resilience.

Three of the designs are based on identical “cells” where each cell is either a clique (a complete graph), a star (with a central node called “leader”) or a cycle (nodes connected

in a ring). Each of these have a single parameter, k - the number of nodes in the cell. Let us also consider n -node graphs consisting of randomly-connected cliques (sometimes termed “cavemen”), randomly-connected stars, in both cases according to probability p . Consider also the simpler and well-studied Erdos-Renyi (ER) random graph with probability p (see Fig. 5.2). Of course, the six designs hardly exhaust the range of structures found in real networks, and future research ought to consider other designs in more detail.

Armed with these designs, the computation below will determine which of them is optimal (i.e. how to build the cascade-resilience network), as well as how the optimal design should be configured (e.g. how large should each cell be). The networks produced by different configurations of a single design could be quite different from each other. For instance, the cellular designs could be configured to create networks of multiple disconnected components¹, as well as the more extreme networks without any edges or with all nodes in the same component. Some of the designs are expected to perform better since they generalize simpler designs: the connected stars design includes both the ER design (by setting $k = 1$) as well as the stars design (by setting $p = 0$). Similarly, connected cliques includes both the ER graph (by setting $k = 1$) as well as the cliques graph (by setting $p = 0$). Note also that some designs have structural limitation so that possibly none of their configurations can achieve $R(G) = 1$ or $W(G) = 1$. For example, the stars design cannot achieve $W(G) = 1$ for any positive attenuation exponent. As we shall see, this will effect its optimal configuration for extreme values of the parameters.

It is clear that the above palette of designs is far too short to provide immediate value to all of the application areas where cascade resilience is desired. The objective

¹It has been argued that disconnected graphs are not realistic as models for many applications, but several reasons suggest otherwise. First, a variety of networks are only connected in the topological sense, and in fact, a very small number of edges act as bridges between parts of the network. Second, in many networks certain edges are highly resilient to cascade propagation, or could be made to be. Thus the disconnected network provides a simplified model of networks containing regular and resilient edges.

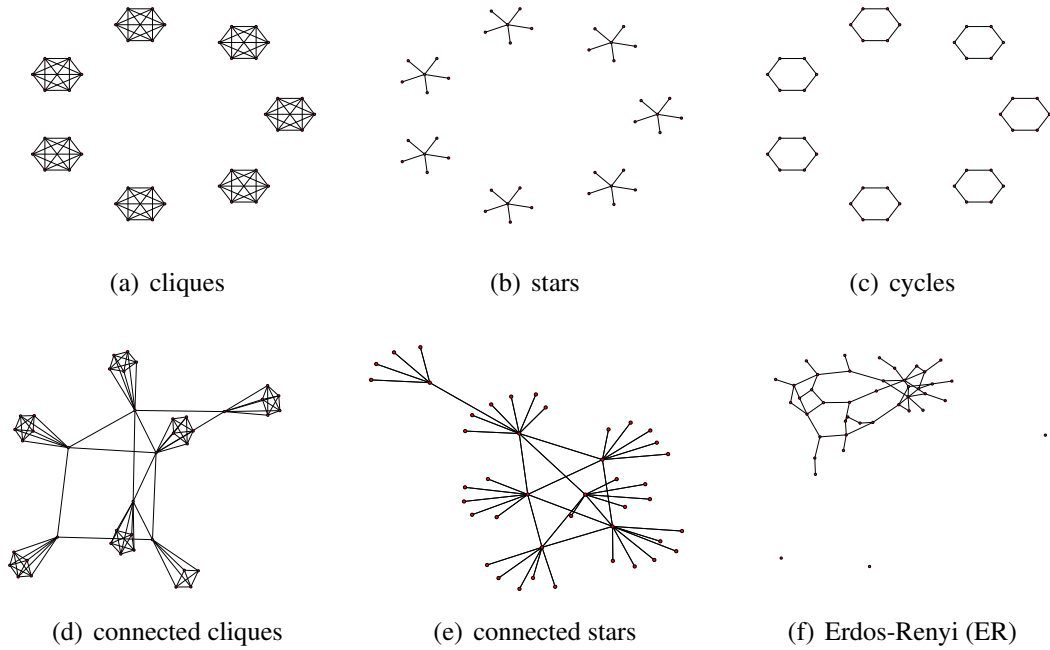


Figure 5.2: Illustration of the 6 designs.

here is to propose an approach which could be applied to different domains, as well as begin constructing a theory to address cascade resilience. It should also be noted that the “optimal design” is likely to be a function of exogenous parameters such as cascade risk, τ , weight of resilience, r , as well the application area: functional constraints which may make some of the designs unsuitable.

A particularly important problem is understanding the structure of terrorist networks. These networks are prototypical examples of networks that are maximized to resilience - their adversary is various government agencies, and efficiency - to be able to recruit and to carry out violence. Unfortunately, it is hard to obtain detailed data on their structure, with the notable exception of the 9/11 network [36] and some historic underground groups, such as FTP illustrated in the introduction. For the FTP network, a “battalion” (228 nodes, 462 edges) was constructed based on the account in [6]. Although both the FTP fighters and the 9/11 terrorists are secret societies, the author does not propose any

moral equivalence between their methods or objectives.

5.4 Results

Having completely specified the optimization problem in the last section, the current shows the optimal solutions. The optimum depends of course on the parameters τ , r and g representing contagion risk, the importance of resilience as opposed to efficiency, and the attenuation of the network, respectively. Together they create what is perhaps a surprisingly rich picture of optimal networks, but one that can be completely understood with intuition.

5.4.1 Optimal Network

The first set of experiments explores the effect of contagion risk (τ) while keeping $g = 1$. In each setting of τ , the optimal configuration of each design was determined by varying parameters such as cell size. Consider first the fitness of the designs at their optimal configuration(Fig. 5.3). A basic qualitative observation is that within each design, as τ increases, the fitness decreases - one cannot win when fighting cascades, only delay. This monotonicity could be proved in general (see Appendix, sec. 5.B). It can be seen that the connected stars design is superior to all others in fitness(Fig. 5.3) but the simpler stars design is almost as fit. This suggests that the connections between cells brings additional fitness only at extreme ranges of τ .

Consider now the empirical networks: 9/11 and FTP. It is interesting that the 9/11 network is quite successful for low values of τ (< 0.2), but then it rapidly deteriorates. This is due to a rapid increase in the extent of cascades - rapid decline in resilience. This

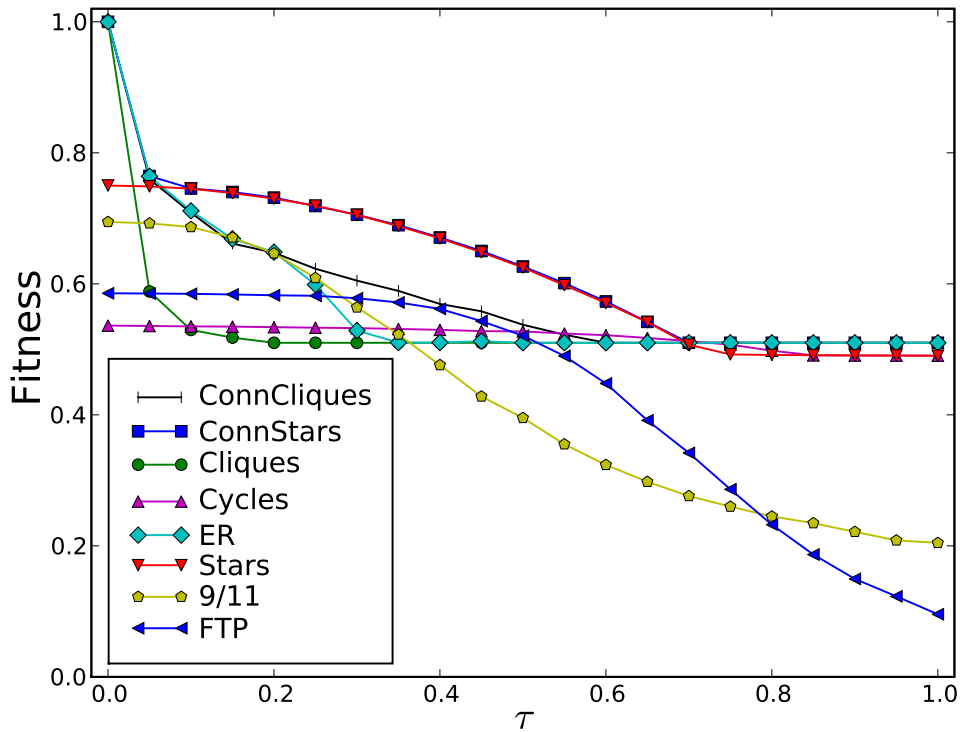


Figure 5.3: Fitness of the optimal configuration for various designs, and for the 9/11 and FTP networks. $r = 0.49$, $g = 1$.

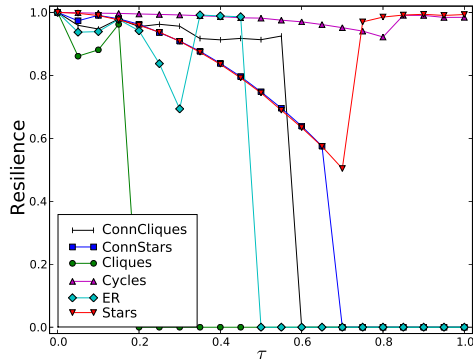
onset of rapid decline suggests that in some types of networks, the network might be initially hard to defeat, but there is a point after which efforts against it start to pay off. If τ is representative of the security environment, then one can say that the 9/11 is relatively ill-adapted to more rigorous security measures implemented after the attacks. Indeed, it is likely that the 9/11 attacks would have been thwarted under the current regime since one of the nodes was captured before 9/11. In contrast, the cellular tree hierarchy of the FTP network is more suitable for intermediate range of cascade risk τ , but the average distances in it are too long to provide high efficiency. Therefore, its performance is comparatively poor in very low and very high ranges of τ . Other networks, such as the ubiquitous scale-free networks found in e.g. scientific collaborations and friendship

relationships [28] are not discussed here, but it is clear that because high-degree hubs are found in these networks, such networks would be extremely vulnerable to cascades once τ is slightly larger than 0. Of course, their organizing principle is not based on retarding cascades.

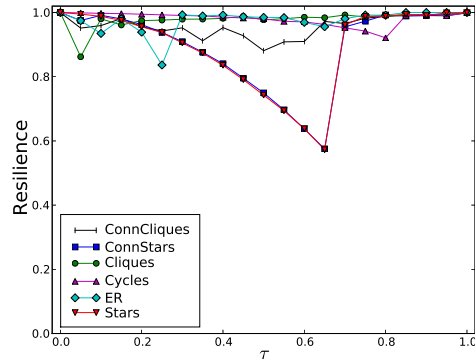
In certain applications it is possible to invest in reducing the cascade propagation probability, τ (e.g. using *nomes de guerre* in a secret society). Then the curves in Fig. 5.3 could also be viewed as expressing the value of efforts to reduce cascades by reducing τ . If the slope is steep then the gains are large. It is important to remember that the fitness curves indicate the fitness of the optimal configuration of each design, rather than a static network (except for FTP and 9/11 series). If however the configuration was made fixed then the fitness decrease would be even steeper as τ increased since changing the configuration can mitigate some of the decrease.

The fitness of an optimal network is a continuous function of the parameter r , and so the counterpart of Fig. 5.3 but for $r = 0.51$ is almost indistinguishable (see Appendix, sec. 5.A for justification). In contrast resilience, efficiency and other properties of the optimum may experience discontinuous “phase transitions” as r is changed. For example, in the cliques design a transition occurs at $r = 0.5$ and cascade risk ($\tau \gg 0$). Below $r < 0.5$ the optimal configuration maximizes efficiency, whereas for $r > 0.5$ it maximizes resilience (Fig. 5.4,5.5, 5.13).

Intuition suggested that the networks grow more sparse as contagion risk grows. Instead, the results were surprising because the trend was non-monotonic (Fig. 5.6). Unexpectedly, for $\tau \gg 0$ and $r < 0.5$ some network designs (e.g. cliques, connected cliques, connected stars) became denser, instead of sparser, and for them the most sparse networks were formed in the intermediate values of τ . At intermediate τ values the optimum achieves both relatively high resilience and high efficiency. At higher τ values,

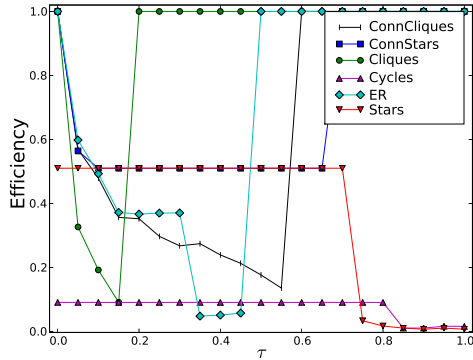


(a) $r = 0.49$

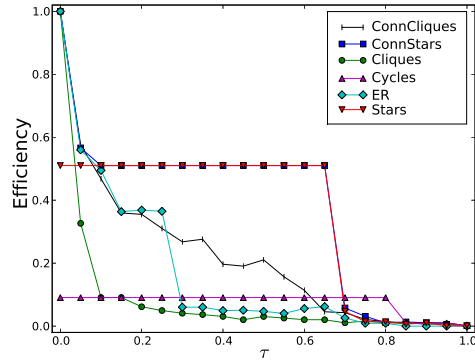


(b) $r = 0.51$

Figure 5.4: Resilience of the optimal design.



(a) $r = 0.49$



(b) $r = 0.51$

Figure 5.5: Efficiency of the optimal design.

when $r < 0.5$ it pays to sacrifice resilience because fitness is increased when efficiency is made larger through an equal or lesser sacrifice in resilience. The transition does not occur in the stars design at $r = 0.5$ because with stars it is relatively harder to increase efficiency.

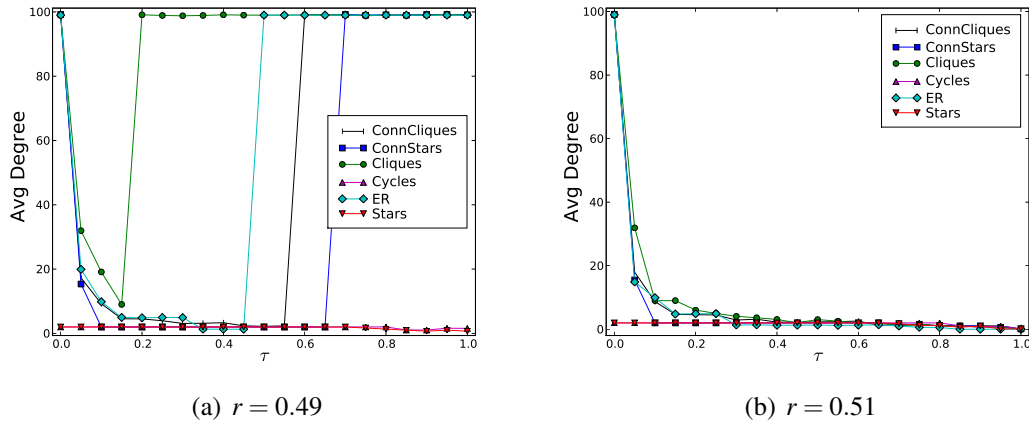


Figure 5.6: Average degree in the optimal configuration of each design.

5.4.2 Effect of attenuation

Various combinations of the parameters have interesting effects on the fitness, as shown on Fig. 5.7. Comparing fitness for attenuation $g = 0.1$ against $g = 10$ notice that decreas-

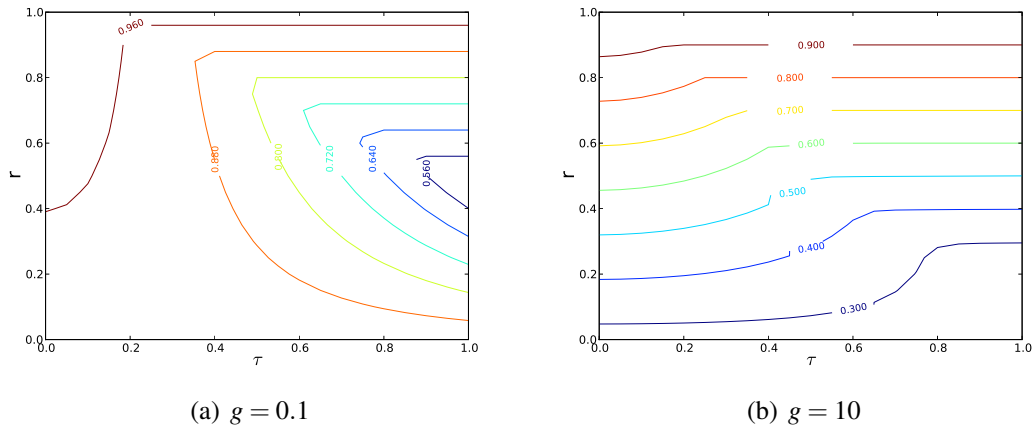


Figure 5.7: Fitness of the optimal configuration in the stars design for various values of g , τ and r .

ing g improves the fitness of the optimal configurations, as expected. Furthermore when g is small, it is easy to find highly-optimal configuration because the networks could be made sparse - improving cascade-resilience without significantly reducing efficiency.

In contrast, when attenuation g is large, efficiency cannot be achieved and the optimal configuration of the stars design is to have cells of size 1 maximizing resilience. It is perhaps surprising that for smaller g , high fitness is most difficult to achieve when efficiency and resilience are approximately equally weighted (i.e. r is near 0.5), especially when τ is near 1.0.

The attenuation g has also interesting effects on the relative merits of various designs (Fig. 5.8). For example, the cycle design becomes competitive with stars when $g = 0.1$: the relatively large distance in cycles help stop cascades and do not decrease efficiency very much since g is low. It becomes the best design for $\tau = 0.55$ and larger.

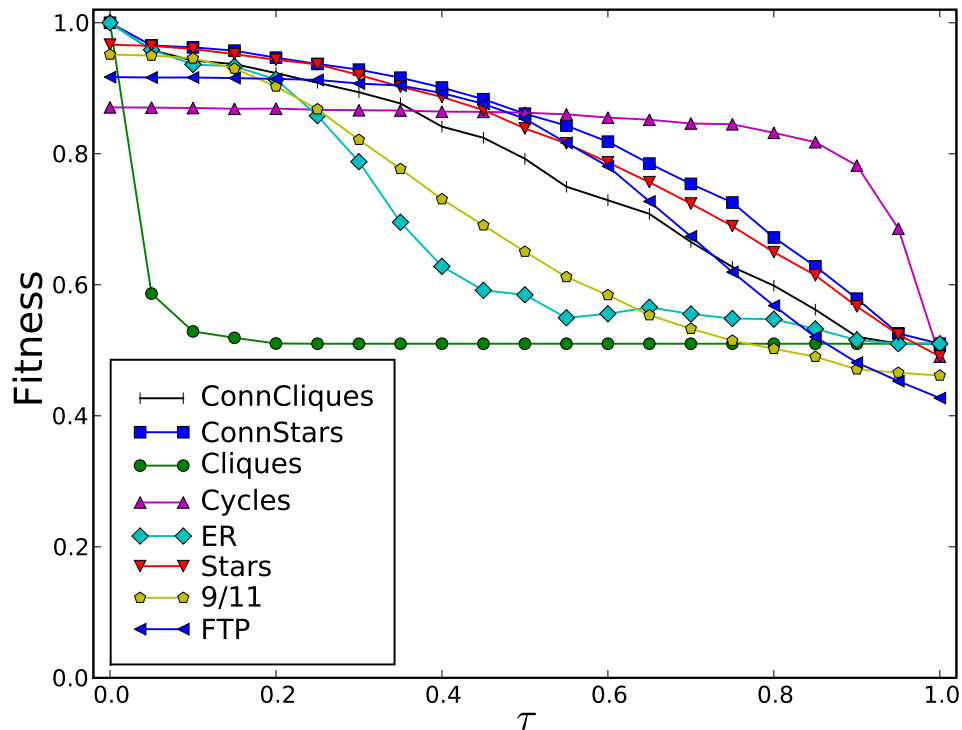


Figure 5.8: Fitness of the optimal configuration for each design when $g = 0.1$.
Data is for $r = 0.49$.

5.5 Discussion

The success of the stars design could be analyzed more qualitatively. The fitness function combines resilience $R(G)$ which decreases when the graph becomes more strongly connected, and efficiency $W(G)$ which decreases when the graph becomes more sparse. The optimality of the star-based designs is due to a good trade-off between R and W : the central node in each cell (its “leader”) provides a good firewall against cascades because in each cell most pairs are separated by distance of 2, but this separation reduces efficiency only modestly. In the cliques design the separation is 1 (too short for resilience), and in the cycles design it is too long (\sim a quarter of cell size, which is too long for efficiency).

Mathematically, the existence of a non-trivial solution is due to the different functional relationships. To a first-order approximation, efficiency decreases inversely with average distance ($\sim \frac{1}{\text{avg distance}^g}$) while cascade propagation probability decreases exponentially ($\sim \tau^{\text{avg distance}}$, for $\tau < 1$ assuming a bounded number of alternative paths). For example, for the star design $R = 1 - \tau^2$ and $W = 2^{-g}$ as $n = k \rightarrow \infty$. Therefore, the optimal network’s structure exploits the exponential decrease in cascades without sacrificing too much efficiency. In the range $\tau \in [0.2, 0.7]$ and $r \approx 0.5$, an average distance of ≈ 2 , as in the star graph, might be optimal.

Structuring the network into communities might help increase in fitness. Notice that the connected stars design mostly outperforms the random graph $G(n, p)$ even though it includes it (by setting $k = 1$). A similar effect occurs in connected cliques vs. cliques - why? Perhaps to achieve high performance it is helpful to build the graph around “communities” - sets of densely-connected nodes. Indeed the optimal configuration away from $\tau \rightarrow 0$ is precisely based on $k > 1$ (see Appendix, sec. 5.E). The effect of com-

munity structure on cascades has been explored extensively, and the new result is that that communities achieve resilience while maintaining efficiency. For more details on configuration parameters see Appendix, sec. 5.E. For sensitivity analysis see Appendix, sec. 5.F.

The finding that under high cascade risk the optimal network is dense is interesting, because our expectation was that the optimal network would be sparse and tree-like. Instead, it was found that at high τ values the optimal networks have low fitness values (Fig. 5.3) and for $r < 0.5$ are not optimized for resilience at all. This may have interesting parallels in a variety of application areas. Consider for instance non-violent social movements, like the movement that brought about the independence of colonial India, or those involved in the electoral revolutions in Serbia, Ukraine and Georgia. The organizers of those movements intentionally chose to organize openly rather than to form an underground. This openness greatly facilitated the movement's growth, although it put at risk the individuals who participated. The parallel in the model is to the sacrifice of resilience to cascades in order to gain higher efficiency. This work suggests that such a sacrifice is worth making even when cascade risk is high as long as efficiency is more valued or replacing nodes is easy ($r < 0.5$).

5.6 Conclusions and Future Work

This work has explored the problem of designing networks for cascade resilience. The main contributions are:

- A general definition of the problem as a multi-objective optimization problem
- Metrics for efficiency and resilience that work well in various networks, including

disconnected topologies

- Evidence for optimality of star-like topologies
- Evidence for non-monotonocities in the edge density as a function of cascade risk
- Evidence that the cellular hierarchical network, such as the FTP is suitable for intermediate ranges of risk, and that the 9/11 network would have been easily defeated under a more rigorous security environment.

Much further work remains to be done in this area. The multi-objective optimization problem for networks could be modeled in other interesting ways, such as by examining the efficient frontier, or by maximizing just efficiency $W(G)$ under a constraint on resilience: $R(G) \geq r$. Another interesting area is optimal design in heterogeneous rather than simple graphs. The former are important in practice and are potentially more involved. For example, there could be two or more classes of nodes, with different effects on efficiency and resilience. One of them could be “immune” to contagions. As well, the current contagion model could be usefully generalized to other models (multiplexed contagions [1], the *SIS* model or threshold contagions [28]). It should also be worthwhile to explore questions about dynamics, such as, how to grow networks while maintaining both their efficiency and cascade resilience. More theoretically, the discussion of designs suggests consideration of Kolmogorov complexity as applied to networks. It is possible that the number of parameters in a designs constraints the optimality of the network. The work could also be extended to consider novel application areas such as the improvement of financial credit networks, whose structure may make them vulnerable to bankruptcies [37, 38].

Acknowledgments

This work has benefited from discussions with Michael Genkin, Roy Lindelauf, Richard Durrett and Michael Macy. Consultations with Shane Henderson greatly helped in designing the simulations. Aaron Clauset generously provided the data on the 9/11 network.

APPENDIX

5.A Continuity of Fitness

It was claimed in sec. 5.4.1 that fitness is continuous. Notice that the claim is not about the continuity of fitness of a single configuration as a function of r but rather that:

Claim: $f(r) = \max_{G \in \mathbb{G}} F(G, r)$ is continuous for $r \in [0, 1]$.

Proof: The argument constructs a bound on the change in f in terms of the change in r . Consider an optimal configuration C_1 of a design for $r = r_1$ and let its fitness be $f_1 = F(C_1, r_1)$ (there is slight abuse of notation since C is a configuration, which is usually an ensemble of graphs).

Observation 1: consider the fitness of C_1 at $r = r_2$. Because C_1 is fixed and the metrics are bounded ($0 \leq R \leq 1$ and $0 \leq W \leq 1$), the fitness change is bounded by the change in r :

$$\begin{aligned} |f_1 - F(C_1, r_2)| &= |r_1 R(C_1) + (1 - r_1) W(C_1) - r_2 R(C_1) - (1 - r_2) W(C_1)| \\ &= |(r_1 - r_2) R(C_1) - (r_1 - r_2) W(C_1)| \\ &\leq |r_1 - r_2|. \end{aligned}$$

Observation 2: let C_2 be the optimal configuration for $r = r_2$ and let $f_2 = F(C_2, r_2)$. Since C_2 is optimal for $r = r_2$ it satisfies: $f_2 \geq f(C_1, r_2)$, and so $-f_2 \leq -F(C_1, r_2)$.

It follows that $f_1 - f_2 \leq f_1 - F(C_1, r_2)$. Together with Observation 1 get the bound:

$$f_1 - f_2 \leq |r_1 - r_2|.$$

Observation 3: applying the argument of Observations 1&2 but reversing the roles of C_1 and C_2 implies that $f_2 - f_1 \leq |r_1 - r_2|$.

Observations 2&3 give $|f_1 - f_2| \leq |r_1 - r_2|$. Continuity is then proved by recalling

that f_2 is a function of r and taking the limit: $\lim_{r_2 \rightarrow r_1} |f_1 - f(r_2)| \leq \lim_{r_2 \rightarrow r_1} |r_1 - r_2| = 0$.

5.B Extent and Contagion Risk

Proposition: Let

$$f(\tau) = \max_{G \in \mathbb{G}} \left[\underbrace{rR(G, \tau) + (1-r)W(G)}_{F(G, \tau)} \right]$$

be the fitness of the optimal graph G^* for a fixed network design \mathbb{G} , for cascade probability τ . Then $f(\tau)$ is a non-increasing function of τ .

Proof of Proposition: The proof relies on a simple claim that resilience of networks does not increase when τ increases. Namely:

Claim: $\forall G$, a simple graph, if $\tau_+ > \tau$ then $R(G, \tau) \geq R(G, \tau_+)$, that is, increasing cascade probability does not increase resilience.

If the claim is proved the remainder is almost trivial, because we need to show that when the fitness of all the points on the space has been made smaller or kept the same (by increasing τ), the new maximum value would not be greater than the old. Rigorously, assume by contradiction that $\tau_+ > \tau$ and fitness *increased*, namely:

$$f(\tau_+) > f(\tau). \tag{5.-3}$$

Let $G_{\tau_+}^*, G_{\tau}^*$ by any two optimal networks for τ_+ and τ , respectively, namely: $G_{\tau_+}^* \in \operatorname{argmax}_{G \in \mathbb{G}} [rR(G, \tau_+) + (1-r)W(G)]$ and $G_{\tau}^* \in \operatorname{argmax}_{G \in \mathbb{G}} [rR(G, \tau) + (1-r)W(G)]$.

By optimality of G_{τ}^* , get that at τ

$$\underbrace{F(G_{\tau}^*, \tau) - F(G_{\tau_+}^*, \tau)}_{\equiv \Delta} \geq 0. \tag{5.-3}$$

Expanding Δ :

$$\begin{aligned}
\Delta &= F(G_{\tau}^*, \tau) - [rR(G_{\tau_+}^*, \tau) + (1-r)W(G_{\tau_+}^*)] \\
&\leq F(G_{\tau}^*, \tau) - rR(G_{\tau_+}^*, \tau_+) - (1-r)W(G_{\tau_+}^*) \text{ by the Claim.} \\
&= f(\tau) - F(G_{\tau_+}^*, \tau_+) \\
&< 0 \text{ by the assumption (5.B).}
\end{aligned}$$

This implies that $F(G_{\tau}^*, \tau) - F(G_{\tau_+}^*, \tau) < 0$ and therefore contradicts that G_{τ}^* is an optimal network for τ (Eq. 5.B).

The argument is more general. One could apply this method to the parameter g of attenuation, showing that fitness is non-increasing when attenuation is increased.

Proof of Claim: Consider the first step of the cascade, i.e. as it expands from a single node in condition I . There are no removed nodes, and hence the probability the contagion would expand from the initial is at least as large under τ_+ as it is under τ . Next, consider any state \mathcal{H} of the graph containing nodes in S, I , and R conditions (susceptible, infected and removed, resp.). Consider also any state \mathcal{H}' which can be produced in one cascade event from \mathcal{H} (note that the time steps of the simulation above may each contain several such events, but any node can only change its condition once per time step). The possible differences between \mathcal{H} and \mathcal{H}' are a single node (1) changing $I \rightarrow R$ or a node (2) changing $S \rightarrow I$. Observe that for (1) and (2) the probability of the event happening is at least as large under τ_+ as compared to τ . Since any cascade can be decomposed into a finite number of cascade steps, the expected extent of the contagion is at least as large under τ_+ as compared to τ . More formally, let $X_{i,t}$ be the event that the contagion is in state i at time t , where a “state” contains information on which nodes are in each of the S, I, R conditions. Consider two network states $X_{j,t+1}$ and $X_{k,t+1}$, reachable from $X_{i,t}$ where $X_{k,t+1}$ has not less nodes in condition I . Under the measure induced by τ_+ the probability of transitions to state $X_{k,t+1}$ is at least as large under τ , for all such

states. By induction on t from 1 to t (t is time where the epidemic has no more infected nodes, i.e. nodes in condition I), the mean extent must be at least as large under τ_+ .

5.C Simulation Methodology

The resilience metric is most easily computed by simulation where a node is selected at random to be “infected”, and the simulation is run until all nodes are in states S or R , and none is in state I . A cascade/contagion that starts at a single node would run for up to n steps, but usually much fewer since typically $\tau < 1$ and/or the graph is not connected. To achieve good estimate of the average extent, the procedure was replicated 40 times, and then continued as long as necessary to achieve an error of under ± 1 node with a 95% confidence interval.

For each design and configuration, the program generated 1 – 10 sample networks (depending on the variability characteristic of the design) of 100 nodes each, and computed the average objective function value. The coefficient of variation in the fitness of the sample networks was monitored to ensure that the average is a reliable measure of performance. Typically variation was < 0.2 except near phase transitions of connectivity and percolation. In designs consisting of cells of size k , in order to consider a spectrum of k values some of which might not divide 100, the number of nodes was chosen to be either the largest multiple of k less than 100 or the smallest multiple larger than 100. If k was fractional, cell sizes were sampled from a normal distribution with mean k and standard deviation 0.3, rounded to the nearest integer. In general, normalization in the definitions of resilience and efficiency ensures that even when the number of nodes is tripled the effect of network size on fitness is very small for the above designs (around

± 0.05 in numerical experiments).

Optimization was performed using simple grid search without grid refinement. Alternative methods (e.g. Nelder-Mead) were considered but grid search was chosen despite its computational cost because it suffers no convergence problems even in the presence of noise (present due to variations in topology and contagion extent), and collects data useful for sensitivity analysis.

An analytic computation of the cascade extent metric was investigated. It is possible in theory because the contagion is a Markov process with states in the superset of the set of nodes, 2^n . Unfortunately, such a state space is impractically large. When G is a tree, then an analytic expression exists², and it might be feasible when the treewidth is small [39, 29]. However, in many graphs below the tree approximation is not suitable. A fruitful approximate approach is to represent the contagion approximately as a system of differential equations which can be integrated numerically [4]. These possibilities were not pursued since the simulation approach was sufficient and could be applied to all graphs.

5.D Analytic Results

The information provided by simulations is valuable but limited, as simulations cannot be run for the entire infinity of parameter values and design configurations. Fortunately, it is easy to analytically derive the values of the resilience, efficiency (and hence fitness) functions for certain simple designs: the cycles and the stars designs. Recall that n is the number of nodes and k is the number of nodes per cell. For $k = 1$, in both designs

²Specifically, the mean contagion size is $1 + \frac{pG_0'(1)}{1-pG_1'(1)}$, where $G_0(x)$ generates the degree distribution and $G_1(x) = \frac{G_0'(x)}{G_0'(1)}$ generates the probability of arrival to a node [29].

$R = 1$ and $W = 0$. When $k \geq 2$, for the cycle design:

$$R(n, k, \tau) = 1 - \frac{1}{n-1} \left[2\tau \frac{1 - \tau^{k-1}}{1 - \tau} - (k-1)\tau^k \right]$$

$$W(n, k, g) = \frac{1}{n-1} \begin{cases} \frac{1}{\left(\frac{k}{2}\right)^g} + 2\sum_{j=1}^{\frac{k}{2}-1} \frac{1}{j^g} & \text{k even} \\ 2\sum_{j=1}^{\frac{k-1}{2}} \frac{1}{j^g} & \text{k odd} \end{cases}$$

and for the stars design:

$$R(n, k, \tau) = 1 - \frac{1 - \frac{1}{k}}{n-1} [2 + \tau(k-2)] \tau$$

$$W(n, k, g) = \frac{1 - \frac{1}{k}}{n-1} [2 + 2^{-g}(k-2)] .$$

These expressions are not readily useful for continuous optimization since k is discrete, but they can be used to obtain a plot of the fitness function, and identify phase transitions. Thus, they help inform optimization for designs where no analytic expression is available.

In the stars design, when R and W are weighted equally ($r = \frac{1}{2}$), fitness takes a relatively simple form: $F = \frac{1}{2} + \frac{1}{2} \frac{1 - \frac{1}{k}}{n-1} [2(1 - \tau) + (2^{-g} - \tau^2)(k-2)]$. This implies that increasing cell size k , for k large, improves fitness iff $2^{-g} - \tau^2 > 0$. Hence the optimal configuration has one cell ($k = n$), until a threshold near $\tau = 2^{-g/2}$ (for $g = 1$, approximately 0.71). This agrees with the findings in Fig. 5.9. Also, the rate of change in fitness with respect to τ , $\frac{dF}{d\tau} = \frac{1 - \frac{1}{k}}{n-1} [-2 - 2\tau(k-2)]$, is always negative, as expected on more general grounds (see Appendix, sec. 5.B). It is linear in τ (because it is a tree graph) but superlinear in k (because of the mutual hazard induced by adding nodes to cells.)

5.E Configuring the Optimal Design

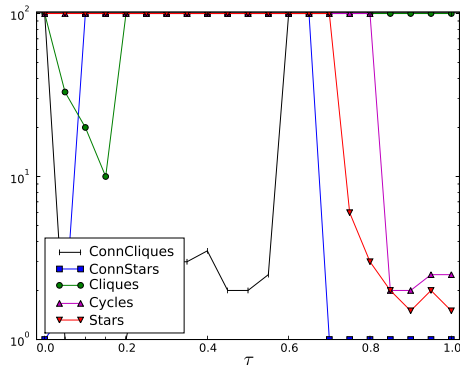
As τ is varied, the optimal configuration changes. This section shows those changes in the values of the parameters k (cell size) and p (connectivity). In other words, it indicates how each of the designs ought to be configured to attain optimal fitness, as a function of resilience weighting, r , and cascade probability, τ .

The cell size parameter k is non-monotonic for various designs under $r < 0.5$ (Fig. 5.9). For example, for the connected cliques design, at low contagion risk ($\tau < 0.1$), k is high (comparable to the size of the network, i.e. $k \rightarrow n$), then it falls to a small number. At high contagion risk ($\tau > 0.6$) the network is again highly connected again with $k \rightarrow n$. Thus for $\tau \rightarrow 1$, the optimal network is the fully-connected graph.

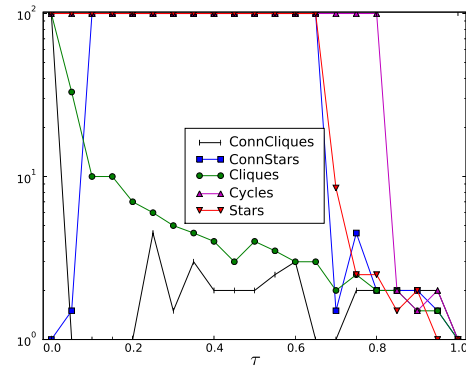
In general, designs involving both the p and k parameters show an intricate interplay between the two (Fig. 5.10). For example, in the connected stars design under $r < 0.5$ there are two phase-transitions in connectivity p : as τ increases at $\tau \rightarrow \tau_l^* \approx 0.1$ it transitions from a connected graph to disconnected cells, and at $\tau \rightarrow \tau_u^* \approx 0.7$ back to full connectivity. If $r > 0.5$ the second transition is extinguished. The data requires care to interpret. For example, in the connected stars design $\tau \in [0.1, 0.65]$, when $r > 0.5$ the fluctuations in the p are noise because there is a single cell and a single cell leader ($k = n$), and so the parameter p has no effect. For sensitivity analysis see Appendix, section 5.F.

5.F Sensitivity Analysis

It is desirable to determine how much variability exists within the optimal values. One possible approach is to consider all configurations whose fitness ≥ 0.95 of the fitness

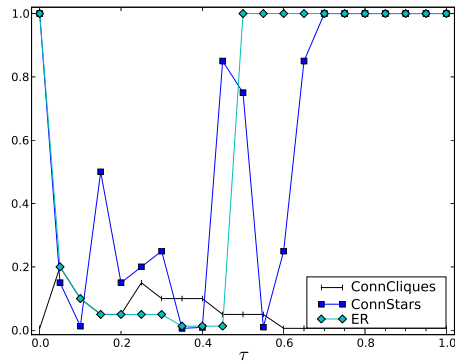


(a) $r = 0.49$

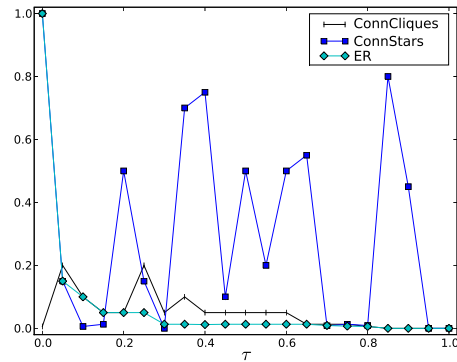


(b) $r = 0.51$

Figure 5.9: Cell size k in the optimal configuration of each design.



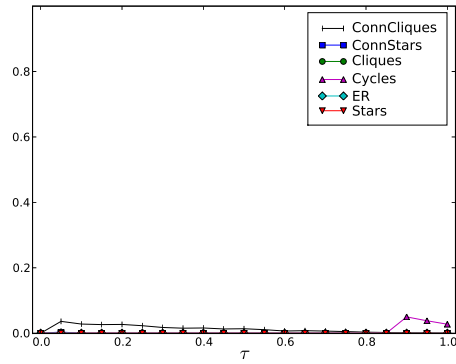
(a) $r = 0.49$



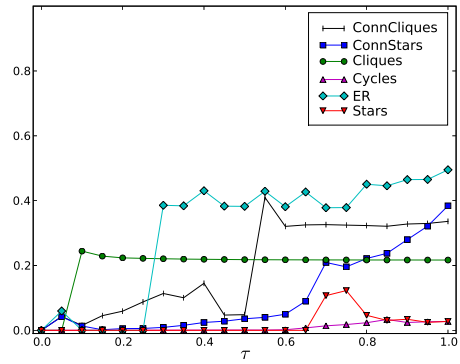
(b) $r = 0.51$

Figure 5.10: Connectivity p in the optimal configuration of each design.

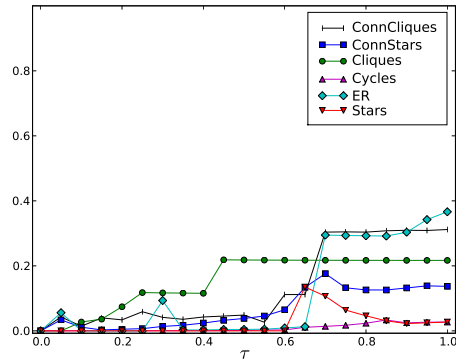
of the optimal solution, and describe the variability in this space. Since in practice the space is infinite, sampling is necessary. The plots of standard deviation within various properties of those configurations are shown in Figs. 5.11,5.12,5.13,5.15,5.14.



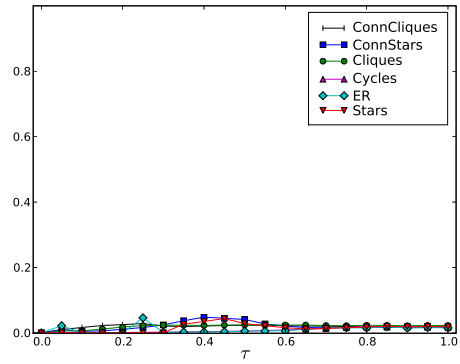
(a) $r = 0.25$



(b) $r = 0.49$

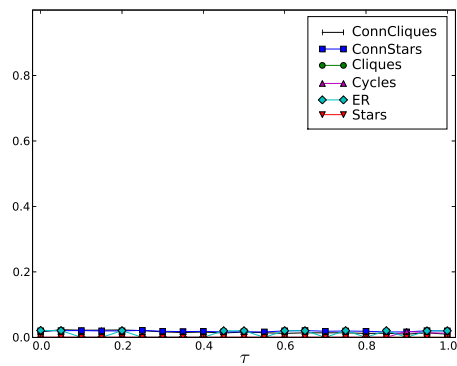


(c) $r = 0.51$

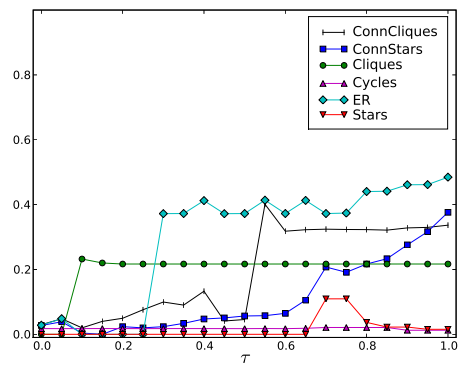


(d) $r = 0.75$

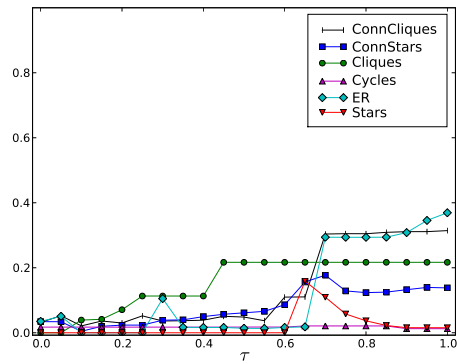
Figure 5.11: Standard deviation in resilience, within the top 5% of solutions.



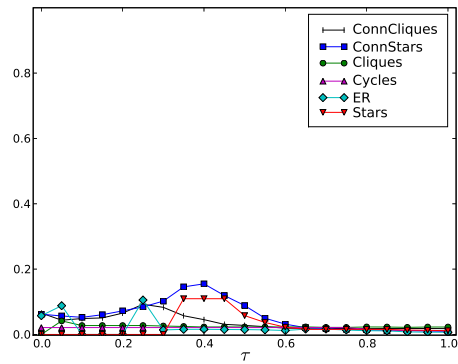
(a) $r = 0.25$



(b) $r = 0.49$

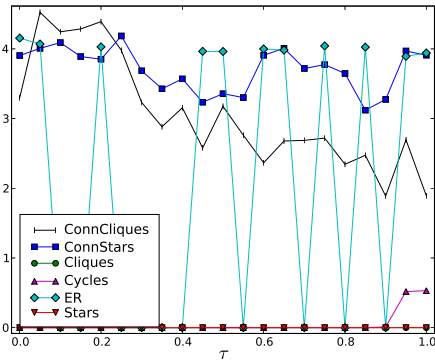


(c) $r = 0.51$

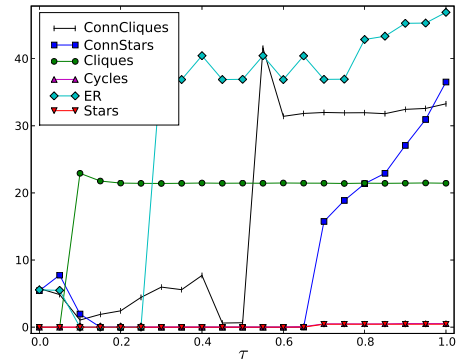


(d) $r = 0.75$

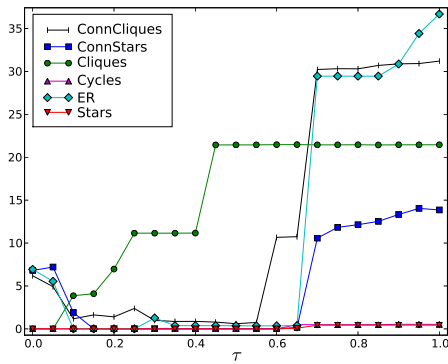
Figure 5.12: Standard deviation in efficiency, within the top 5% of solutions.



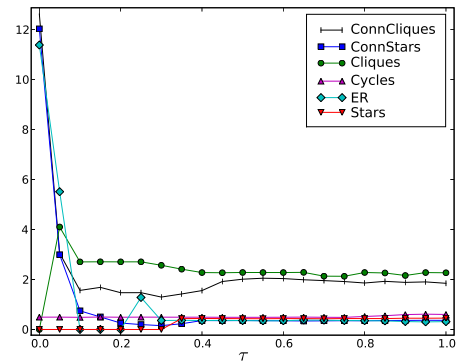
(a) $r = 0.25$



(b) $r = 0.49$

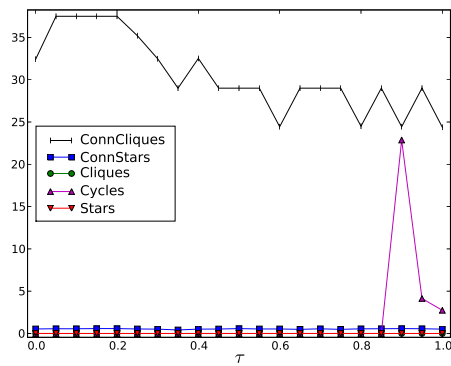


(c) $r = 0.51$

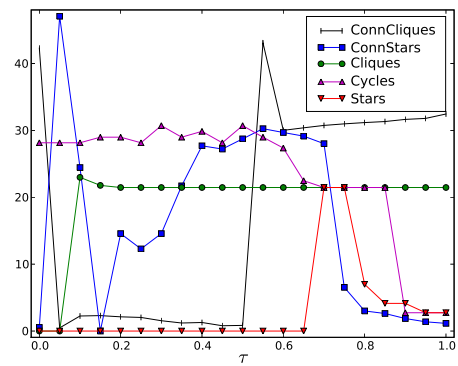


(d) $r = 0.75$

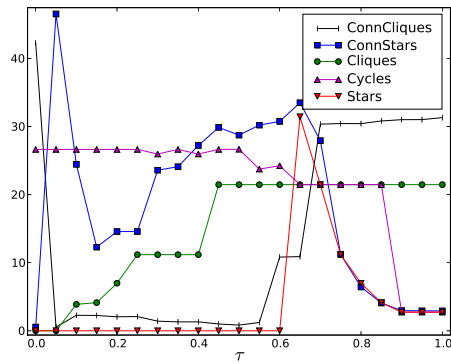
Figure 5.13: Standard deviation in average degree, within the top 5% of solutions.



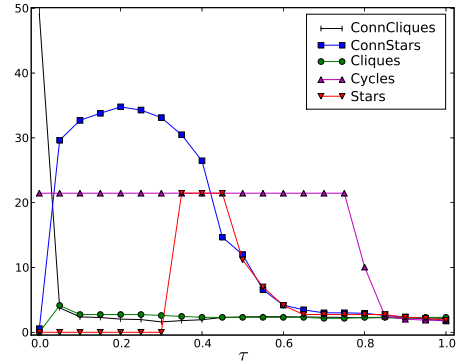
(a) $r = 0.25$



(b) $r = 0.49$

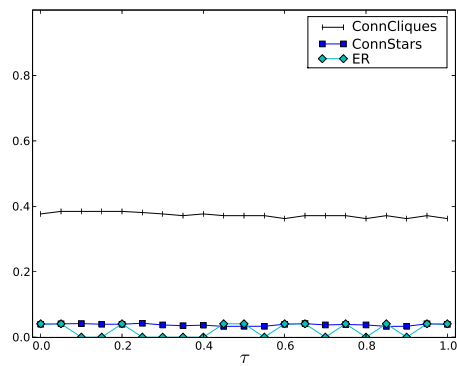


(c) $r = 0.51$

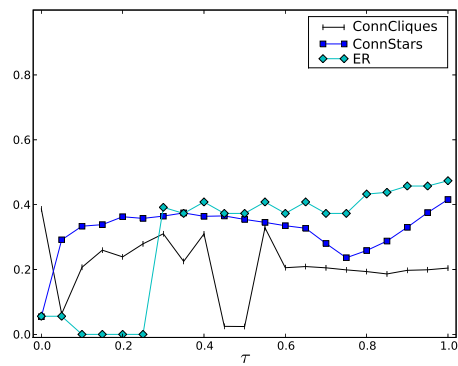


(d) $r = 0.75$

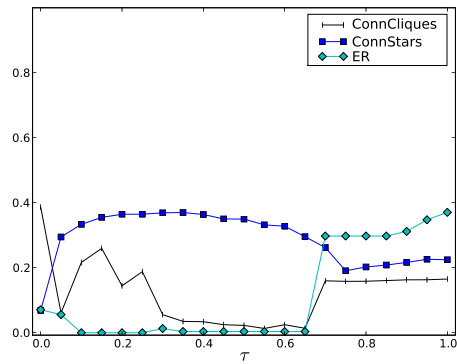
Figure 5.14: Standard deviation in cell size k of the top 5% of solutions.



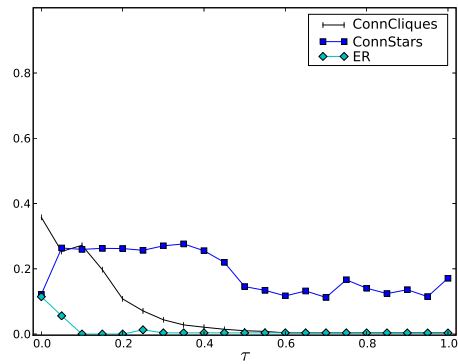
(a) $r = 0.25$



(b) $r = 0.49$



(c) $r = 0.51$



(d) $r = 0.75$

Figure 5.15: Standard deviation in connectivity p of the top 5% of solutions.

Overall, as one would expect, the properties are more variable near the transition point $r = 0.5$ as compared to r values away from $r = 0.5$. Moreover, variability is high within each design whenever the design undergoes a phase transition, since multiple different phases have nearly equal fitness. Designs with two parameters are more variable than those with a single parameter because the latter can sometimes reproduce the same graphs with many different parameter settings - the parameters have “non-orthogonal” effects.

REFERENCES

- [1] D. Centola and M. Macy, “Complex contagions and the weakness of long ties,” *American J. Sociology*, vol. 113, pp. 702–734, Nov 2007.
- [2] P. Crepey, F. P. Alvarez, and M. Barthelemy, “Epidemic variability in complex networks,” *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, vol. 73, no. 4, p. 046131, 2006.
- [3] W. Huang and C. Li, “Epidemic spreading in scale-free networks with community structure,” *J Stat Mech*, vol. P01014, 2007.
- [4] M. J. Keeling, “The effects of local spatial structure on epidemiological invasions,” *Proc R Soc Lond B*, vol. 266, pp. 859–867, Jan 1999.
- [5] R. Pastor-Sarorras and A. Vespignani, “Epidemic spreading in scale-free networks,” *Phys Rev Lett*, vol. 86, pp. 3200–3203, Apr 2001.
- [6] F. O. Miksche, *Secret Forces*. London, UK: Faber and Faber, 1st ed., 1950.
- [7] G. W. Klau and R. Weiskircher, “Robustness and resilience,” in *Network Analysis*, Lecture Notes in Computer Science 3418, pp. 417–437, Springer-Verlag, 2005.
- [8] W. E. Baker and R. R. Faulkner, “The social organization of conspiracy: Illegal networks in the heavy electrical equipment industry,” *American Sociological Review*, vol. 58, no. 6, pp. 837–860, 1993.
- [9] C. Morselli, K. Petit, and C. Giguere, “The Efficiency/Security Trade-off in Criminal Networks,” *Social Networks*, vol. 29, no. 1, pp. 143–153, 2007.
- [10] G. Gunther and B. L. Hartnell, “On minimizing the effects of betrayals in resistance movements,” in *Proceedings of the Eighth Manitoba conference on Numerical Mathematics and Computing*, pp. 285–306, 1978.
- [11] R. H. Lindelauf, P. E. Borm, and H. Hamers, “On Heterogeneous Covert Networks,” *SSRN eLibrary*, 2008.
- [12] J. Zawodny, “Internal organization problems and the sources of tensions of terrorist movements as catalysts of violence,” *Terrorism: An International Journal (continued as Studies in Conflict and Terrorism)*, vol. 1, no. 3/4, pp. 277–285, 1978.

- [13] G. Woo, *Mathematical Methods in Counterterrorism*, ch. Intelligence Constraints on Terrorist Network Plots. Springer-Verlag, 2009. to appear.
- [14] R. Albert, H. Jeong, and A.-L. Barabasi, “Error and attack tolerance of complex networks,” *Nature (London)*, vol. 406, pp. 378–381, 2001.
- [15] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, “Network robustness and fragility: Percolation on random graphs,” *Phys. Rev. Lett.*, vol. 85, pp. 5468–5471, Dec 2000.
- [16] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, “Resilience of the internet to random breakdowns,” *Phys. Rev. Lett.*, vol. 85, pp. 4626–4628, Nov 2000.
- [17] J. C. Doyle, D. L. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka, and W. Willinger, “The ”robust yet fragile” nature of the Internet,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 41, pp. 14497–14502, 2005.
- [18] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, “Attack vulnerability of complex networks,” *Phys. Rev. E*, vol. 65, p. 056109, May 2002.
- [19] P. Holme, “Efficient local strategies for vaccination and network attack,” *Europhys. Lett.*, vol. 68, pp. 908–914, Dec 2004.
- [20] P. Crucitti, V. Latora, and M. Marchiori, “Model for cascading failures in complex networks,” *Phys. Rev. E*, vol. 69, p. 045104, Apr 2004.
- [21] Y.-C. Lai, A. Motter, and T. Nishikawa, “Attacks and cascades in complex networks,” in *Complex Networks: Lecture Notes in Physics 650*, pp. 299–310, Springer-Verlag, 2004.
- [22] A. E. Motter, “Cascade control and defense in complex networks,” *Phys. Rev. Lett.*, vol. 93, p. 098701, Aug 2004.
- [23] B. L. Hartnell, “The optimum defense against random subversions in a network,” in *Proceedings of the Tenth Southeast conference on Combinatorics Graph Theory and Computing*, pp. 494–499, 1979.
- [24] A. S. Finbow and B. L. Hartnell, “On designing a network to defend against random attacks of radius two,” *Networks*, vol. 19, no. 7, pp. 771–792, 1989.
- [25] M. Baccara and H. Bar-Isaac, “How to organize crime,” *Review of Economic Studies*, vol. 75, pp. 1039–1067, October 2008.

- [26] R. H. Lindelauf, P. E. Borm, and H. Hamers, “The Influence of Secrecy on the Communication Structure of Covert Networks,” *Social Networks*, 2009.
- [27] B. Pourbohloul, L. Meyers, D. Skowronski, M. Krajdén, D. Patrick, and R. Brunham, “Modeling control strategies of respiratory pathogens,” *Emerg. Infect. Dis.*, vol. 11, no. 8, pp. 1246–56, 2005.
- [28] M. E. J. Newman, “The structure and function of complex networks,” *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
- [29] M. E. J. Newman, “Spread of epidemic disease on networks,” *Phys. Rev. E*, vol. 66, p. 016128, Jul 2002.
- [30] D. Newth and J. Ash, “Evolving cascading failure resilience in complex networks,” *Complexity International*, vol. 11, pp. 125–136, Jan 2005.
- [31] P. Erdos and A. Renyi, “On the evolution of random graphs,” *Publication of the Mathematical Institute of the Hungarian Academy of Science*, vol. 5, pp. 17–67, 1960.
- [32] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *Nature (London)*, vol. 393, pp. 440–442, Jun 1998.
- [33] R. Albert and A.-L. Barabasi, “Emergence of scaling in random networks,” *Science*, vol. 286, pp. 509–512, Oct 1999.
- [34] J. Arquilla and D. Ronfeld, *Networks and Netwars: The Future of Terror, Crime, and Militancy*. Santa Monica, CA: RAND Corporation, 2001.
- [35] K. M. Carley, “Destabilization of covert networks,” *Comput Math Organiz Theor*, vol. 12, pp. 51–66, 2006.
- [36] V. E. Krebs, “Mapping networks of terrorist cells,” *Connections*, vol. 24, no. 3, pp. 43–52, 2002.
- [37] S. Battiston, D. D. Gatti, M. Gallegati, B. Greenwald, and J. E. Stiglitz, “Credit chains and bankruptcy propagation in production networks,” *Journal of Economic Dynamics and Control*, vol. 31, pp. 2061–2084, 2007.
- [38] G. Iori, G. D. Masi, O. V. Precup, G. Gabbi, and G. Caldarelli, “A network analysis of the italian overnight money market,” *Journal of Economic Dynamics and Control*, vol. 32, pp. 259–278, 2008.

- [39] T. Colcombet, “On families of graphs having a decidable first order theory with reachability,” in *Automata, Languages and Programming: 29th International Colloquium, ICALP* (J. Diaz *et al.*, eds.), 2002.