Stepping Back from the Trees to see the Forest: Network Approaches to Valuing Intelligence

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

in partial fulfillment

of the requirements for the degree

Doctor of Philosophy

by

Christopher M. Smith

May

2013

APPROVAL SHEET

The dissertation

is submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

AUTHOR

The dissertation has been read and approved by the examining committee:

Prof. William T. Scherer

Advisor

Prof. Peter A. Beling

Prof. Michael Smith

Prof. Jeremy Marcel

Prof. Duncan E. McGill

Accepted for the School of Engineering and Applied Science:

Dean, School of Engineering and Applied Science

May

2013

Abstract

Determining value of intelligence can be a difficult problem. One way to value intelligence is to judge a document's worth by its location within a structure of a given corpus of documents. Network applications are a natural extension of this logic. I introduce a methodology for value of information (VOI) for networks, comparable to VOI for influence diagrams. Additionally, citation networks and Google's PageRank algorithm are examples of valuing information based on its location within a structure. Dynamic network analysis (DNA) has been used to allow social network analysis (SNA) involving multi-nodal networks by creating inferences across networks with common nodes. I introduce the application of the DNA layered approach to information networks in an attempt to determine value of intelligence. These applications demonstrate supplemental, and objective ways of measuring intelligence.

For my wife, daughters and son. With hard work, all things are possible.

Notation and Acronyms

ACM	Association for Computing Machinery
CASOS	Center for Computational Analysis of Social and Organizational Systems
CIA	Central Intelligence Agency
CRS	Congressional Research Service
DARPA	Defense Advanced Research Projects Agency
DIA	Defense Information Agency
DNA	Dynamic Network Analysis
DNI	Director of National Intelligence
EVII	Expected value of imperfect information
EVPI	Expected value of perfect information
G2	Intelligence staff section
IC	Intelligence Community
IED	Improvised Explosive Device
IEW/TA	Intelligence, Electronic Warfare, and Target Acquisition
INSCOM	Intelligence and Security Command
IR	Intelligence Report
ISR	Intelligence, Surveillance, and Reconnaissance
JHU/APL	Johns Hopkins University/Applied Physics Laboratory
JT-COIC	Joint Training Counter IED Operations Integration Center
JWICS	Joint Worldwide Intelligence Communications System
MC	Main Component
MORS	Military Operations Research Society
NDIC	National Defense Intelligence College
NGIC	National Ground Intelligence Center
NPV	Net Present Value
ODNI	Office of the Director of National Intelligence
OPTVIEW	Operational Value of Intelligence, Electronic Warfare, and Target Acquisition
OR	Operations Research
PPBES	Planning, Programming, Budgeting, and Execution System
SNA	Social Network Analysis
VOI	Value of Information

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Motivation

Since the events of 9-11, the United States has waged a war against terrorists. This war is unlike most previous wars, in that there are few battle lines, the enemy does not necessarily wear a uniform, and leadership and location of the enemy has shifted through different countries and continents throughout the course of the war. In order for civilian and military leadership to make effective plans to win this war, they need effective strategic intelligence. Intelligence, defined by Lowenthal, "refers to information that meets the stated or understood needs of policy makers and has been collected, processed, and narrowed to meet those needs."[1] This intelligence helps the leadership understand the enemy they're facing, promotes development of counter measures to help defeat the enemy's strategy, lending support towards a winning U.S. strategy. Intelligence is a key aspect of this war, and for that matter, any war.

The U.S. Intelligence Community (IC) produces over a billion pieces of information a day.[2] Current efforts to improve intelligence focus on collecting more information rather than analysis of existing information/intelligence. Collection management, within the IC, is a practice that focuses on knowing the needs of the intelligence users and determining how best to allocate limited collection resources.[3] Collection managers also review existing allocation strategies by determining if intelligence collected is valuable and relevant. Providing an objective, supplemental measure to collection managers, that is calculated while intelligence is still in the intelligence cycle would be useful to collection managers to appropriately allocate the national intelligence resources.

1.2 Purpose and Scope

The idea of a quantitative measure for intelligence is a huge topic. Ideally, a true measure of the value of intelligence might incorporate a number of different factors both subjective and objective. The focus of this dissertation will be on a supplemental and objective quantitative measure of intelligence. Specifically, a network based approach to demonstrating value of intelligence within a given corpus of documents. The overall intent of this research effort is to examine whether or not a layered network methodology indeed offers a quantitative approach towards measuring the various aspects of intrinsic value of intelligence dealing with the connectivity within existing knowledge

1.3 Organization of Document

The introduction is contained within chapter 1. In chapter two, I explore the literature behind valuing intelligence and how it might apply to this methodology. I recommend applying a supplemental valuing methodologies within the intelligence cycle, or before decisions or actions are made using the intelligence. In the third chapter I offer a compliment to value of information (VOI) in influence diagrams, but for networks. Additionally, I explore possible applications for VOI based valuations of intelligence. In chapter four, I outline my methodology for applying layered networks to an intelligence network in order to demonstrate value within the corpus. The fifth chapter outlines some results from applying the layered network approach. Finally, the sixth chapter is the conclusion and suggestions for future work over the entirety of my research.

2 Background

2.1 Introduction

The U.S. Intelligence Community (IC) harvests over a billion pieces of data a day, but often lacks the ability to analyze that data and produce valuable intelligence.[4] Efforts to improve intelligence often focus on collection rather than analysis. According to the 9-11 Commission's report, the intelligence community had information that may have either helped decision makers mitigate or even stop the 9-11 attack. Even though the IC had enough information to determine that "the system was blinking red," according to a Central Intelligence Agency (CIA) supervisor, "no one looked at the bigger picture; no analytic work foresaw the lightning that could connect the thundercloud to the ground."[5]

Collection management within the IC is a practice that focuses on knowing the needs of the intelligence users and determining how best to allocate limited collection resources.[6] Collection managers also review existing allocation strategies by determining if intelligence collected is valuable and relevant. Existing methods for evaluating intelligence, however, have focused on their value to the decision maker. Since decisions, especially at the national level, are subject to conflicting priorities and politics, it is difficult to determine what role intelligence plays in decisions. Current value of information methods, however, rely crucially on whether or not information is used in a decision and since collection managers can never truly determine how intelligence is used, existing methods are often inapplicable (see Figure 1).

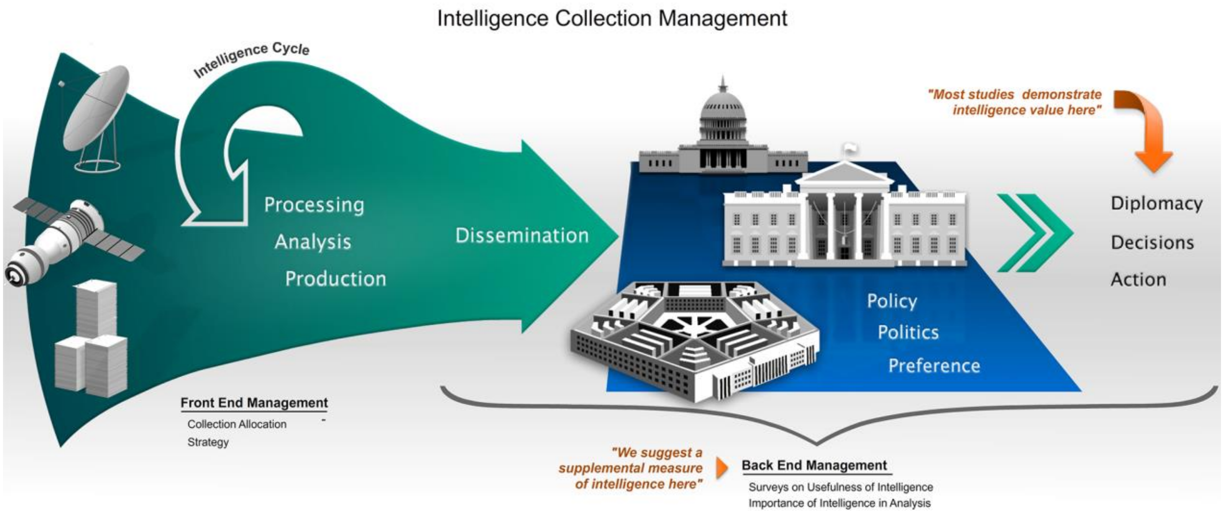

Figure 1 Intelligence collection management within the intelligence cycle

The IC, and collection managers specifically, will benefit from an assessment methodology that allows a quantitative measure of the impact certain intelligence has on an analysis. A robust method of this type will allow managers to objectively measure “good” or effective intelligence vs. “bad” or ineffective intelligence. A quantitative measure applied across the different classes of analyses would provide a relative scoring of the intelligence that different sources produce thereby allowing collection managers to more cost effectively manage expensive collection resources.

The structure of the paper is as follows. The first section introduces the concept of valuing intelligence within the intelligence cycle. The second section defines national security intelligence for the purposes of this paper. The third section describes the variety of methods or theories currently used to “value” information both in the government and in private sector and also outlines why these methods fall short when applied to national security intelligence. The fourth section discusses research on how the valuation of intelligence has been attempted to date and their limitations. The fifth section makes the suggestion of two possibilities for measuring the supplemental value of intelligence within the intelligence cycle. Finally, in the sixth section we offer a conclusion and suggestions for future work.

2.2 Definition of Intelligence

A useful definition for intelligence given by Jennifer Sims is “the collection, analysis, and dissemination of information for decision-makers engaged in a competitive enterprise.”[7] Specifically for the purposes of this paper, we will be discussing strategic level intelligence, but we feel that this work will be effective at all levels of decision making using intelligence. Understanding the definition of intelligence is the first step in measuring of it. Although various authors have put forth a number of definitions of intelligence, one of the main distinctions of the various definitions, and most important to the measurement of intelligence, is that some authors suggest that covert actions taken by the government should be included in the definition of intelligence.[8,9,10] This would turn the term intelligence into something that not only described information, but also described actions taken by the government. The definition of intelligence for this paper will not include covert action.

A useful distinction, when defining intelligence, is the difference between secrets and mysteries. According to John C. Gannon, then Deputy Director for Intelligence in the CIA, in a speech to the World Affairs Council in 1996, the intelligence business distinguishes between secrets and mysteries. Specifically, “secrets, at least theoretically, can be obtained one way or another… mysteries on the other hand, are unknown or unexplained phenomena.” An example he gives for a secret is “the specifications for a new weapon system being developed by a foreign government.”[11] In a talk delivered to a Military Operations Research Symposium (MORS) working group at the Office of the director of National Intelligence (ODNI) in December of

2011, Director of National Intelligence (DNI) James Clapper, referring to the unrest in Egypt at the beginning of 2011, suggested an example of a mystery might be exactly when President Mubarak would choose to step down.[12] Clearly the decision on when President Mubarak would step down was not a secret. It is likely that President Mubarak did not know when he would step down until the very morning that he did.

At a very general level, data is harvested and systematically turned into intelligence using a process called the Intelligence Cycle, seen below in Figure 2. The first stage is planning and direction where the decision makers or intelligence managers direct that some target should be collected on and what different methods should be used to collect the information. The second stage is collection, where the various methods assigned to collect the information actually collect the specified information. The third step is processing, where the vast amount of information is translated, if needed and reduced down to the needed information to give to the analysts. Analysis and production is the fourth step where analysts actually convert the information into intelligence by "integrating, evaluating and analyzing all available data."[13] The final step is dissemination where the intelligence is distributed to the decision makers who may have initiated the process in the planning and direction step. Though it is usually displayed as a cycle where information goes from one step to the other, in reality the progression tends towards the dissemination stage, but frequently information/intelligence is moved back steps based on what the information contains.[14] For example, if while processing (stage 3) a satellite photo, an analyst notices that a location of interest is only partially captured by the photo, due to new construction, they might send the photo back to the planning and direction stage (stage 1) with direction to include more land mass in the photo. For a more detailed discussion on the intelligence cycle with a specific focus towards operations research in the intelligence cycle, see Kaplan.[15]

Some have stated that intelligence should be defined by information that supports decision-making by reducing uncertainty.[16] While this may be true in some cases, it does not sufficiently

Figure 2 Intelligence Cycle (Adapted from The Work of the Nation)

describe the entirety of the process. This definition fails to capture the useful work by intelligence analysts in their efforts to gain a general understanding of competitors. Prior to actions by the Japanese on Pearl Harbor, US intelligence analysts collected a vast amount of information on the Japanese. Much of this information was instrumental in understanding the way the Japanese would prosecute the war.[17] This information was not a prelude to a stated decision, but yet was still considered useful intelligence.

2.3 Approaches to Valuing Information

The first approach that applied a quantitative value to information was Shannon's Information Theory model.[18] In it, Shannon introduces the five components to communication: the source, the transmitter, the channel, the receiver and the destination. A message, primarily binary in nature, is introduced into this system and measured on how accurately the message is transmitted through the system. Although he discusses ways to measure the communication of non-binary messages, his model largely focuses on the syntactic level of communication. When looking at many of the developments of information theory since Shannon's work, the work revolves around innovations in data compression efficiency and efficiencies in the rate information can travel through a noisy channel.[19] Although this is the first research modeling of information through a system, the research focuses on describing the syntax of information moving from the source to the receiver, not necessarily the value of the information flowing between the two nodes.

Another application of valuing information is on the open market. There is a market for information to be bought and sold on the economy. In 1961, Stigler wrote about determining the market price of information.[20] He develops a mathematical representation for the price one might pay to know more information about a market or product, specifically through advertising. In addition, information economics also describes certain specific properties of information that make information different from other goods sold in markets. According to Varian, these properties include: experience good, returns to scale, and public goods.[21] "Experience good" is the property of information that you can only tell whether the information is valuable to you once know it, but once you know it, you will most likely have already paid for it. Returns–to-scale refers to the fact that information can be costly to create or produce, but relatively cheap to reproduce. Pure public goods means that information is both nonrival and non-excludable. Nonrival means that one person consuming information does not diminish the amount of information available to others. Non-excludable means that one cannot keep another from consuming the information. Though the many of the properties of information are still applicable to intelligence, much of the information economics is currently not applicable because by definition, the practices within information economics reside in a commoditized environment. In the intelligence world, a government might pay money for information, or for the option to view information, but will certainly not sell the information on the other side.

2.4 Decision Focused Methods

Some methods for valuing information are decision focused. In other words, the quality or value of the information is based upon the decisions made or recommended using that information. One such approach within decision analysis is the value of information (VOI). According to Clemen et al. we gather information in order to reduce uncertainty so that we can make decisions that give us a better chance of having a good outcome.[20] We will review the basic principles of VOI in this paper as in might be used in the intelligence cycle (for details see Clemen et al.[22]).

When there is a decision to be made that determines an irrevocable allocation of resources, a decision analysis framework is an applicable approach.[23] There are numerous classes of models one can use to model a decision process: decision trees with probabilities, network models, and multiple objective decision analysis to name a few. A decision problem is represented with more structure and detail by using state variables and modeled using probabilities, often elicited from experts and stakeholders.

An illustrative example of how we might use a decision tree to help decision makers solve a problem is where decision makers are faced with the problem of whether to spend money to harden, or better defend, a site that terrorists may attack. In an actual application there may be other, more important variables than the cost in dollars, but for this purpose we will use the cost in dollars. Suppose the following given conditions in Table 1.

Full cost of an attack	-$200,000
Full cost to harden a site	-$25,000
Reduction in cost of an attack after hardening	70%
Probability of an attack	90%

Table 1 Initial conditions on illustrative example (all amounts in $ thousands)

An attack on the site would cost the defenders $200 million, yet it would take $25 million to harden the site. If the site is hardened, the site would only experience 70% of the damage in cost of an attack. Also, we assume the probability of an attack is 90%. In this very simplistic illustrative example, shown on a decision tree in Figure 3, the optimal decision would be to harden the site because the expected value of loss would be $151 million, rather than the $180 million loss if decision makers did not harden the site. Again, this is an illustrative example, so we shall assume that analysts have done due diligence in determining the true probability of an attack (against the possibility of a deception by foreign intelligence) and the magnitude of an attack which would affect the cost of an attack as well as the cost to harden a site.

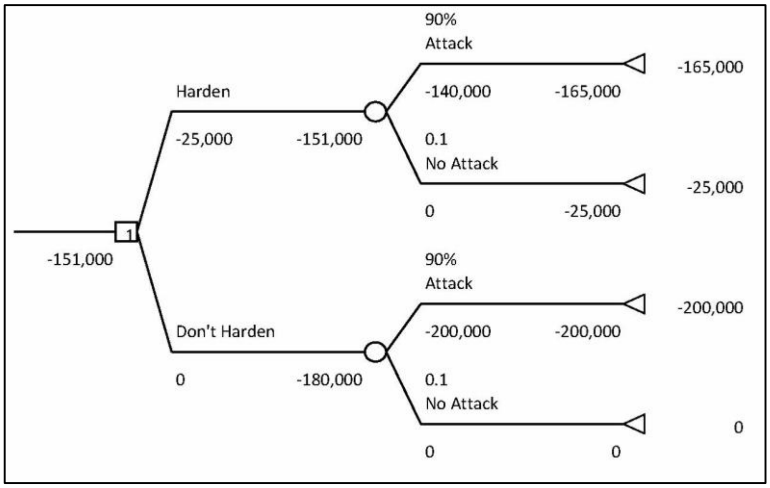

Figure 3 Decision tree of illustrative example

In the above example, information is not modeled, but the approach can be adapted to include information. VOI explores the impact of specific pieces of information on the decision recommended. Ideally, we would want the information given from an expert to be perfect, in other words, no errors. In reality, we understand that this rarely happens, but in theory if we find the expected value of perfect information (EVPI), then we can find the upper value of the worth of that information. In order to mathematically describe perfect information, we will assume that our expert is making predictions on a terrorist attack. Our expert always gives perfect information; mathematically this is a conditional probability expressed as:

$$P(\text{Expert says "attack"} \mid \text{There is an attack}) = 1 \qquad (1)$$

In other words, the probability given there was an attack, that our expert said there would be an attack is equal to 1. In addition perfect information assumes:

$$P(\text{Expert says "no attack"} \mid \text{there is an attack}) = 0 \qquad (2)$$

The probability that our expert predicts no attack and there actually is an attack is equal to zero. And finally,

$$P(\text{Expert says "attack"} \mid \text{there is no attack}) = 0 \qquad (3)$$

The probability that our expert predicts an attack and there is no attack is equal to zero.

The information that our perfect expert gives us must follow all of these rules.[24] Once our expert makes a determination about a terrorist attack, there is no more uncertainty about the occurrence of that attack, though other probabilities may remain, such as the amount of loss of life, or the damage in dollars. Knowing the equations 1-3, will allow us to get to the information that we really want to know,

$$P(\text{There is an attack} \mid \text{Expert says "attack"}) \qquad (4)$$

Or given the expert says there will be an attack, there actually will be an attack.

We use Bayes' Theorem to "flip" the probabilities of the equations 1-3, in order to derive the probability of equation 4 (for details, see Devore[25]). Admittedly, this problem is relatively simple and allows us to quickly understand that the probability of equation 4 is equal to 1, or in other words, there is 100% probability that given our expert says there will be an attack, an attack will occur. Bayes' Theorem is a useful tool for intelligence analysis in its own right. There have been a number of declassified intelligence analyses using Bayes' theorem to update probability of a hypothesis.[26,27,28]

Going back to our notional expert with perfect information, we can use a decision tree to show us the decision on whether to use the perfect information, or take our chances with the laid out in the decision tree without information. In Figure 4, we see that if we choose to take the perfect information, our expected loss would be $126 million, instead of $151 million. In this

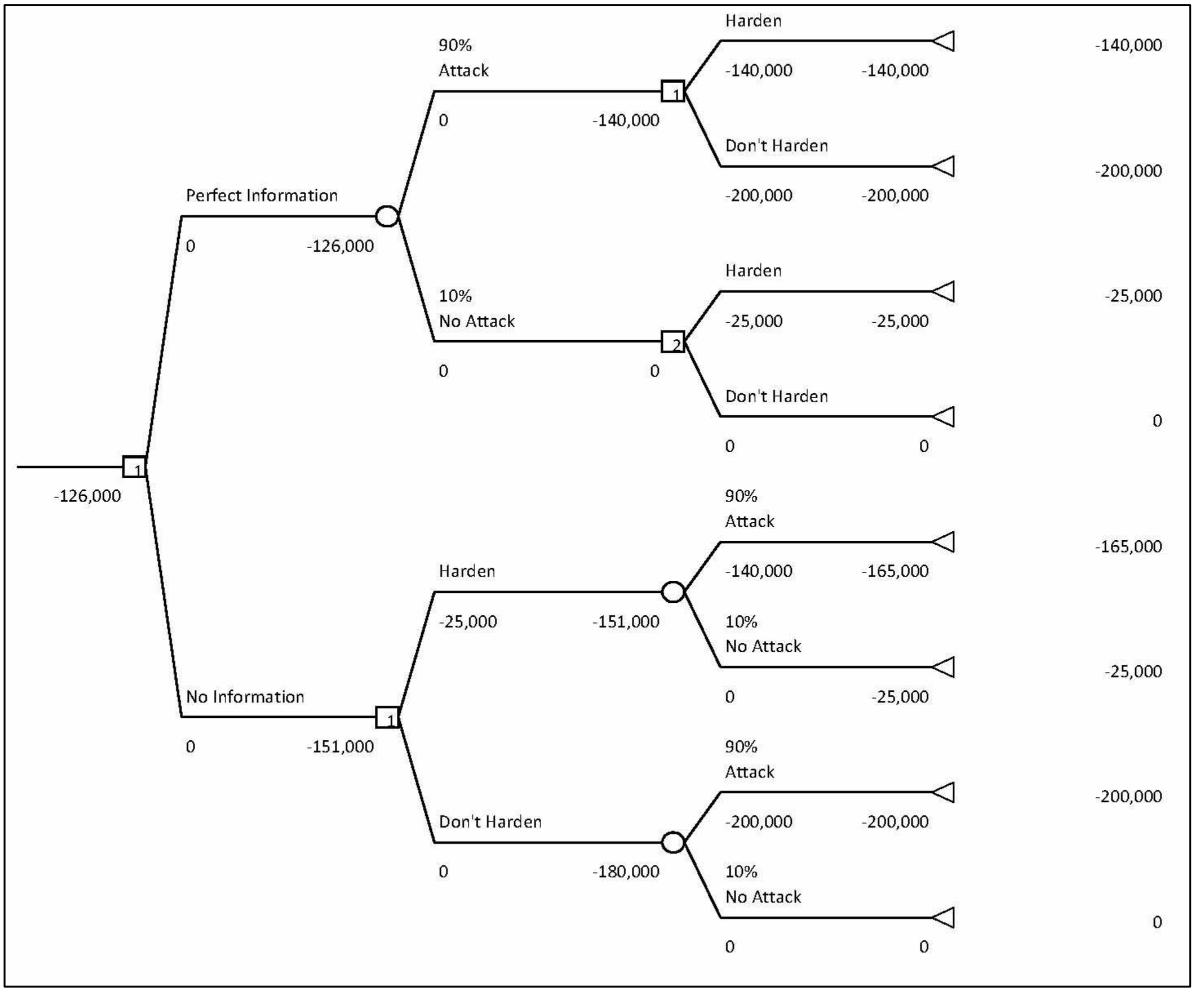

Figure 4 Decision tree of illustrative example with perfect information

notional scenario, the decision maker can expect to save $25 million by using the perfect information. This means that this decision maker should be willing to pay up to $25 million in order to get access to this perfect information. EVPI, therefore, can be interpreted as the upper bound on how much one might be willing to pay for information that helps with a decision.

In reality, we can see that in most situations we will never have perfect information, but sometimes we have enough proof on the accuracy of the information to be able to arrive at the expected value of imperfect information (EVII). Continuing this scenario, consider that we know the information in Table 2 about the accuracy of a source of information. In the source's history, 95% of the time when they've predicted an attack in the past, an attack has occurred. Additionally, when the source has predicted there will be no attack, 70% of the time no attack occurs.

$P(PA \mid Attack)$	95%
$P(PNA \mid No\ Attack)$	70%
PA = Predict Attack	
PNA = Predict No Attack	
Attack = Enemy Attacks	
No Attack = Enemy Does Not Attack	

Table 2 Given information on accuracy of imperfect information source

Using this information, we can modify the decision tree to display how valuable having this source's information might be to the decision makers in Figure 5. Using imperfect information we see that that although the decision maker should still use the source's information, the amount they might pay for that information would be significantly less. In this situation, the imperfect information allows us to expect to lose $150,825,000, while not using the information we would expect to lose $151 million. In this case, the EVII is only $175,000, or using the source's information, we expect to save $175,000, so decision makers should not pay more than this amount for the information that the EVII. In this simple example, the information reduces the expected loss by less than .2% and would likely not be worth considering, especially if there was uncertainty on the probability estimates.

In this illustrative and very simplistic example we demonstrate the application of VOI to a decision problem. Keisler studies more in depth analysis of EVPI within two act linear loss decision problems,[29,30] reengineering of automobile manufacturers decisions[31] and portfolio decision analysis[32]. Each of these applications illustrate in detail the application of EVPI in different applications.

Sometimes the assumption that a decision maker will always choose the least costing or in the case of a multi objective decision problem, most value option is a false assumption. There is an interesting paradox within cognitive psychology that people sometimes make choices opposite of those choices they would agree were common sense.[33] Risk analysts explore these paradoxes using risk attitudes. An example of which is that some people are risk averse, or would pay money to avoid a situation involving possible loss, even if there was a corresponding possibility of gain. On the other hand, some people are risk seeking, or would pay money to be involved in a situation described above. Risk analysis uses utility functions to quantitatively demonstrate a decision maker's risk attitude.[34] The use of risk analysis in conjunction with decision analysis can be a powerful tool. Delquié shows that the value of information can be highly dependent upon the decision maker's predilections prior to the analysis being conducted.[35]

One of the prime applications for risk analysis is risk management. While risk is defined as the probability of a negative outcome, in risk analysis, analysts try to answer the following questions:

1. What can happen? (i.e., What can go wrong?) – the scenario or vulnerability
2. How likely is it that that will happen? – likelihood / probability

3. If it does happen, what are the consequences? – the damage[36]

Applications of risk management attempt to determine variables with high variability in a situation or scenario, attempt to reduce the variability and reduce the probability or impact of a

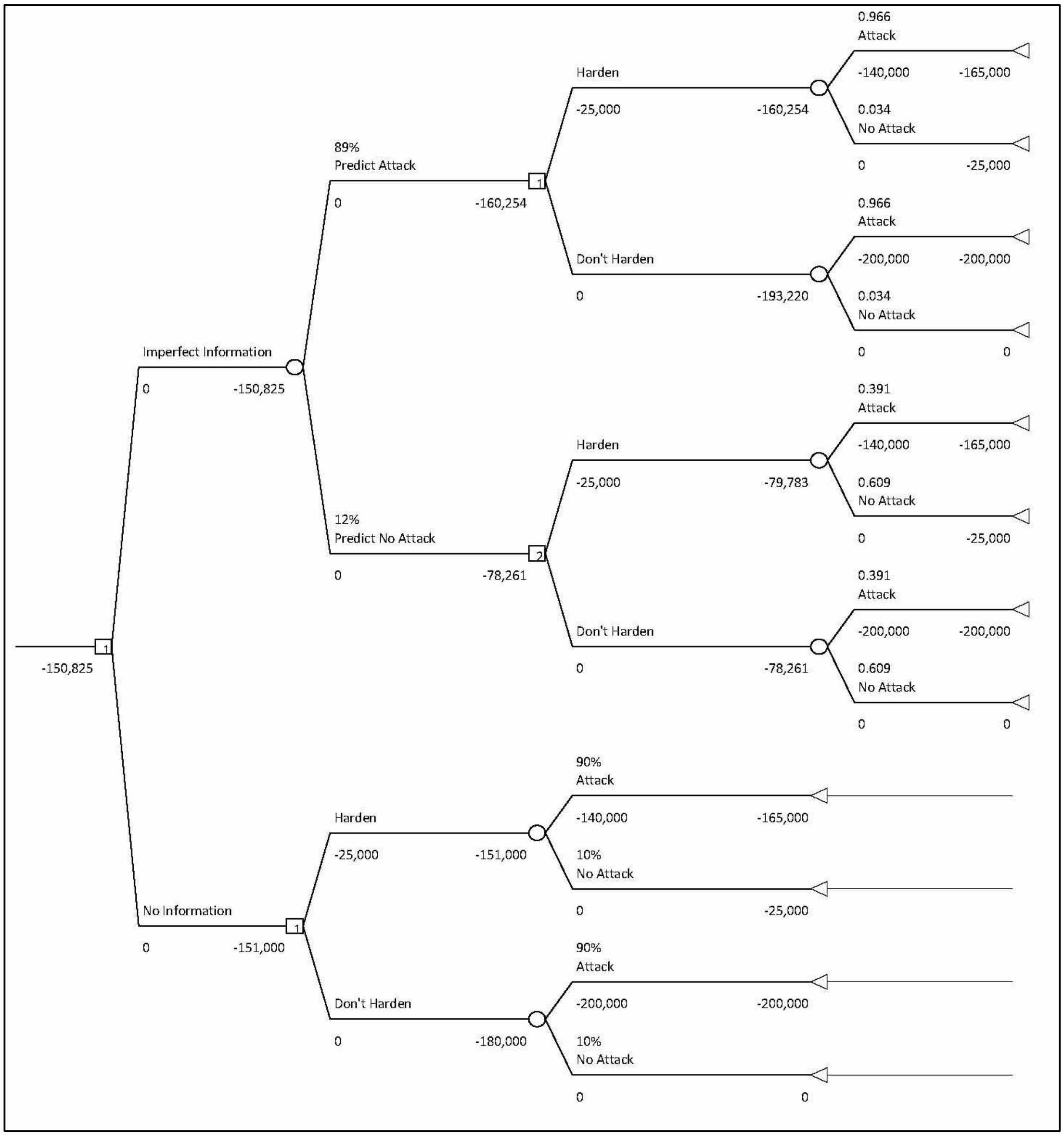

Figure 5 Decision tree of illustrative example with imperfect information

negative outcome as defined by the decision maker. This usually involves trade-offs between risky options or activities.[37] Risk analysis and risk management are used in a variety of different applications within the government.[38,39,40,41,42] Information is usually a key component to determining risk.[43] Risk management modelers use sensitivity analysis to explore the model or risk they've developed to identify which variables offer the most variability to their objective function. Clemen et al. describes sensitivity analysis as answering the question, "What makes a difference in this decision?"[44] These variables with a high degree of variability are then explored and effort is taken to reduce the variability. In the view of risk management, high variability is not preferred because an outcome with a high probability of prediction is the eventual goal.

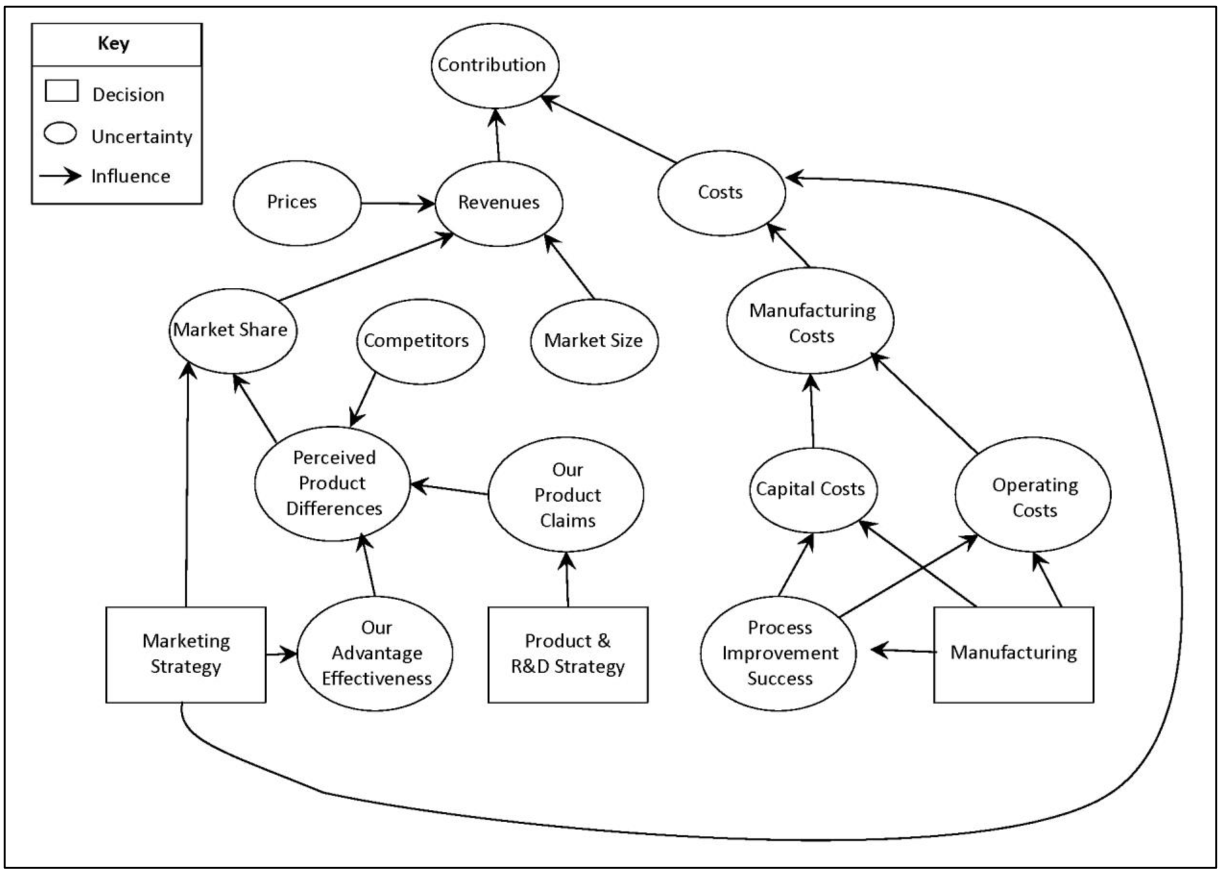

Figure 6 Example of an influence diagram of a financial alternative for a company (Adapted from Bodily et al.)

One method of sensitivity analysis is a visualization tool called a tornado diagram. A tornado diagram shows the magnitude of the effect a variable has on the objective when all other uncertainties are held constant, sometimes called a single factor sensitivity analysis. If one creates a model, such as a decision tree, or an influence diagram, they can use a one way sensitivity analysis tool called a tornado diagram to show the impact of the variable on the overall decision recommended by the model. The tornado diagram allows the modeler to vary one part of the influence diagram while holding all other variables constant. Bodily et al. describe a six step process of value creating within a company. [45] One of the steps involves evaluating the risk and

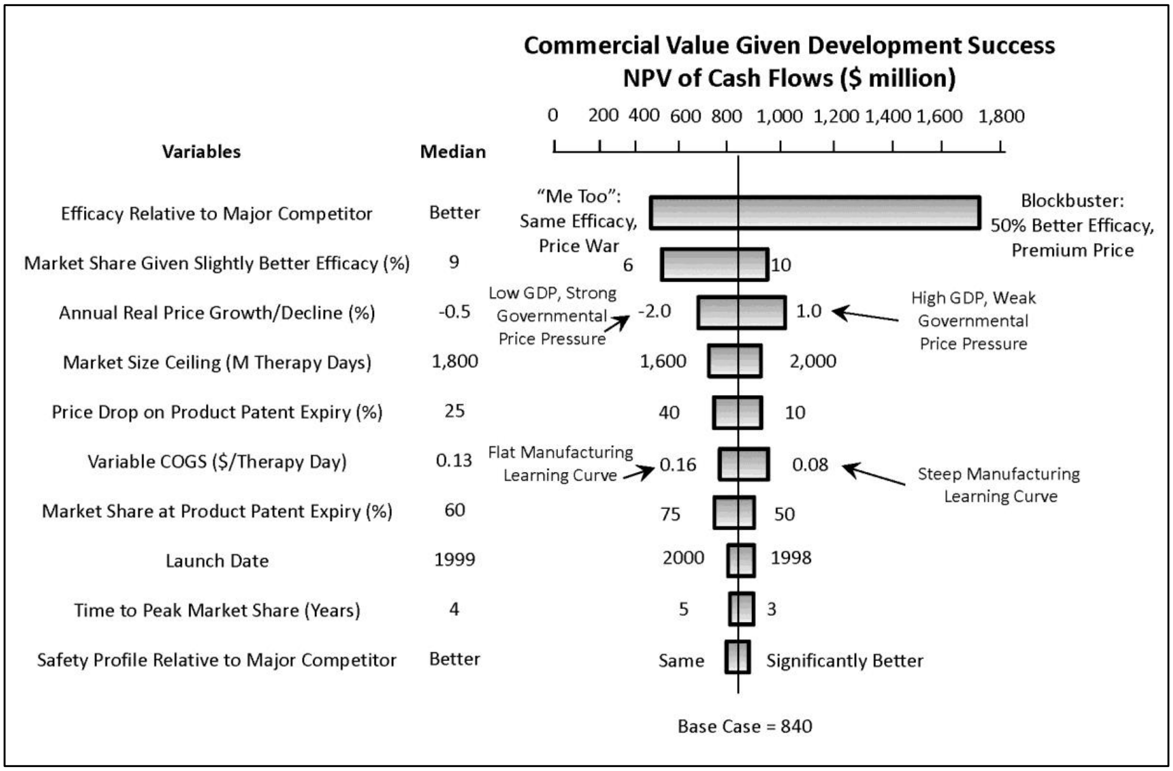

Figure 7 Tornado diagram showing sensitivity analysis of variables within financial alternative (Adapted from Bodily et al.)

potential return of the alternatives, where Bodily et al. demonstrate using an influence diagram and a corresponding tornado diagram showing the sensitivity of the net present value (NPV) to the variables within that influence diagram in Figure 6 and Figure 7. The purpose is to not only communicate to decision makers the variability within their model, but also to identify the variables with the greatest impact, both positive and negative on the NPV of the alterative.

2.5 Review of current valuation of intelligence

Current methods for "valuing" intelligence tend to be subjective at best, but focused primarily on the decisions made based on the intelligence. People who collect the intelligence also provide a numerical "value" of the intelligence with respect to credibility and likelihood; the issue with this is that the people who collect the intelligence have no incentive to downgrade the intelligence, but some incentive to upgrade the "value."[46] These admiralty rankings are seen in Table 3 and Table 4. In other words, the people who collect the information, also grade the information.

Code	Descriptors	Estimated Probability of Truth
1	Confirmed	100%
2	Probably true	80%
3	Possibly true	60%
4	Doubtful	40%
5	Improbable	20%
6	Misinformation	
7	Deception	
8	Cannot be judged	50%

Table 3 Information accuracy codes[47]

Code	Descriptors	Estimated Probability of Truth
A	Completely reliable	100%
B	Usually reliable	80%
C	Fairly reliable	60%
D	Not usually reliable	40%
E	Unreliable	20%
F	Cannot be judged	50%

Table 4 Information reliability codes[48]

Information might have a B-2 rating, which would mean the source is usually reliable, and that the information offered by the source is probably true. In this way, intelligence analysts use a somewhat crude tool to apply probability to the information as a way of assessing its quality. The intent for this method is to capture the subjectivity in some form of a ranking, where the analyst derives the reliability and accuracy. The implication is clear that even assuming the trustworthiness and integrity needed for most intelligence analysts, if an intelligence analyst wants his report read at the higher levels, the data on which his intelligence rests should be judged as reliable or at least potentially higher than the information's actual reliability. This need not be a case of low integrity, but rather an impassioned analyst who believes in an estimate and is willing to give more credibility or to overlook inconsistencies in sources who give information that supports his view.

"I'll know it when I see it," is a quote from Justice Stewart remarking about being able to identify videos that were sexually explicit and of questionable moral value, but wouldn't be able to describe it. [49] Some decision makers and senior intelligence analysts have the false belief that "good" intelligence can be recognized in a similar manner. There are many examples throughout history of senior decision makers or even analysts receiving intelligence predicting an outcome, and simply ignoring the information, this paper will provide one such example.

Author Robert Young outlines the capabilities of the French Intelligence during the time just prior to the start of World War II. They had almost unprecedented access and intelligence into future attacks as well as knowledge about new doctrine the Nazis would use.[50] Though the senior military and political leaders undoubtedly had the information, as well as the historically proven, good estimates, they either ignored the information or chose to do nothing about it. In fact just after the successful German invasion into Poland, the senior French military attaché to Germany Colonel Rivet remarked that the attack had "come as no surprise to anyone but the Poles "and our General Staff."[51]

There can be a number of reasons why decision makers and senior intelligence analysts might ignore or not recognize "good" intelligence. Tversky and Kahneman produced a number of works studying cognitive biases that affect all humans, most notably winning a Nobel Prize for work contained in Judgment Under Uncertainty: Heuristics and Biases.[52] Additionally, Kahneman has continued their research and reviews the state of the art in a recent book, Thinking, Fast and Slow.[53] Some of the types of biases that all humans can suffer from are representativeness, availability, and adjustment and anchoring. In representativeness bias, we sometimes incorrectly assume that one person or thing who might have some attributes similar to a stereotype actually represents and contains all the attributes that stereotype. One example of this is when seeing a meek looking individual with glasses, and considering what job the person might hold, librarian might spring to mind since that individual might fit some common stereotypes of a librarian. When a person suffers from availability bias, they mistake the actual frequency or probability of an event with the ease on might bring it to mind. For example, after just having seen a plane crash on the news, one might mistakenly believe that they happen more frequently than they actually do. Finally adjustment and anchoring is a bias where "people make estimates starting from an initial view that is adjusted to yield the final answer."[54] The final estimate tends to be highly impacted by where they started from. These are only reflective of a broad range of heuristics and biases and while it has been shown that humans can train themselves to avoid these biases, they typically do not. In fact, many people are unaware when they are suffering from these biases. Additionally, Vertzberger outlines cognitive biases at work specifically in political decision makers.[55] One of the cognitive issues he addresses is what he calls the cognitive style. Vertzberger remarks that people, especially political decision makers, create a cognitive system of the world. This system could be very complex in nature or not, and also the individual might believe their system covers the entirety of reality or not. Based on the cognitive style of the individual decision maker(s), they could easily discount otherwise "good" intelligence because the information did not fit or ran contrary to their cognitive style.

Classification can tend to provide some *defacto* "value"; more highly classified intelligence tends to be more highly "valued." According to Lowenthal, "some intelligence processionals have mistakenly equated the degree of difficulty involved in obtaining information with its ultimate value to analysts and policy makers."[56] Although by definition, classification is a supposed to be a measure of the amount of damage the information could do to national security; this is not always the practice with classification.[57] It seems easy to equate the level of effort and amount that we are willing to spend to protect data as a representation of how valuable it may be. Anyone using classification, even in an unofficial or unconscious way of demonstrating the value of intelligence, would be using a flawed system. The problem with this is that the classification system within the IC is inconsistent at best, with different agencies sometimes valuing the same information at different classification levels.[58] Additionally, any e-mail or content published on classified sites, no matter how mundane or personal it might be, is automatically classified at the level of the system which transmits it. Not only would valuing intelligence based on classification be inconsistent, but in many cases the value would be less based on the actual value of the content or how much it helps to complete a picture, but more on the method it was transmitted.

In a recently published article, Stephen Marrin discusses the difficulty of measuring intelligence.[59] He states that there is an inability to decide what metrics intelligence should be judged by. According to him, some feel that intelligence is supposed to be accurate; others feel it should be serve as a warning to decision makers, and finally others feel intelligence must be "influential or useful."[60] Without being able to define what "good" intelligence is, it is hard to measure it.

A number of different methodologies have been applied to valuing intelligence. In 1960, Caldwell et al attempted to model the value of intelligence largely by surveying field grade officers with World War II experience. Researchers would offer a military scenario and offer 100 sets of two "stimuli" consisting of examples of enemy disposition and told to choose their preference of importance between the two "stimuli." Additional information was given to the officer such as the source of the intelligence. Using this information, Caldwell et al developed a linear model that one could use to determine the "value" of a piece of intelligence based on its origin and subject matter.[61]

Over the course of the War on Terror, from 2001 to present, Whitlock et al. led a series of assessments in both the military sector and the national intelligence arena to more empirically establish the value of various intelligence capabilities. On the military intelligence side, they sought to explicitly associate intelligence to operational outcomes—providing a strong indication to leadership on which capabilities were contributing to the counter-terrorism and counter-insurgency operations. In this approach, intelligence reporting was associated to actual military operations through three lenses: temporal, locational, and relational. The purpose was to gain greater insight on the relative contributions to impact programmatic decisions in the Department of Defense.[62] On the national intelligence side, they applied natural language processing capabilities to establish a connection between reporting across the intelligence community against the national intelligence priority structure for that mission area. The priority scheme was broken down into 5-8 major objectives, each with seven dimensions of intelligence targets. The emphasis on both reports were less on establishing the value of an individual report, but rather, to bring clarity to where intelligence collectors and analysts are focusing their effort relative to the priority scheme—thus serving as a resource management tool.[63]

The RAND Institute has studied a way of measuring the operational value of intelligence for military operations. Their approach follows a similar methodology to these earlier approaches; they depend upon the decisions or results based on the intelligence. Their Operational Value of Intelligence, Electronic Warfare, and Target Acquisition (OPTVIEW) related the collection of intelligence to commanders information requirement, assessed the performance capability of different technical collection systems and also could be used to assess performance over a given operation scenario.[64,65] This model and simulation combined was developed primarily to assess the operation of technical Intelligence, Electronic Warfare and Target Acquisition (IEW/TA) alternatives to support the Army Planning, Programming, Budgeting and Execution system (PPBES) rather than evaluate across a variety of intelligence collection methods. The OPTVIEW would consist of a simulation environment using a possible future engagement, and a model that described the predicted amount and quality of intelligence prospective IEW/TA assets might provide to battlefield commanders, allowed simulated commanders to action on the information given, and then measured the ramifications. The tools within OPTVIEW "depended fundamentally on subjective-judgment data" rather than objective analysis.[66] While OPTVIEW might be considered useful in what it was designed to do, the authors have no information as to the validation of this tool, it measures the effects of having battlefield information based on the decisions made within the simulation, rather than the estimating the overall help each system brings to helping the commanders create a picture of what is happening on the battlefield.

The basic themes that these studies have in common are that they attempted to address the difficulty of measuring intelligence. These methods were primarily results based methods which depended heavily or entirely on feedback from users of the intelligence and the results of decisions based on these decisions. Each of these methodologies have their benefits and limitations but arguably, none provide an objective measure of the value of intelligence across a variety of disciplines while avoiding the problematic view of assessing value based on decisions made with the intelligence or results of those decisions.

2.6 Problems applying existing methodologies to Intelligence

The Defense Science Board Advisory committee addresses why Operations Research (OR) tools, like decision analysis have not been adopted within the IC.[67] The study states that OR is applied inconsistently throughout the defense and Intelligence, Surveillance and Reconnaissance (ISR) communities, and they recommend that OR use should be improved. As Marrin noted, the board notes that one of the main difficulties is the inability of decision makers at the various levels of government to agree on priorities or valuation, which makes creating an objective function difficult.[68] Decisions at the higher levels are often made for reasons besides supporting intelligence. Specifically, resources, control, turf, constituent issues, intuition, policies priorities and politics often play into the decision maker process at higher levels. Additionally, Lowenthal states that measuring the value added "is difficult because intelligence officers and their policy clients do not agree on what adds value."[69] When stakeholders cannot arrive at an agreement on valuing something, determining an objective value to use in decision analysis methods mentioned above can be extremely difficult. Most attempts, as seen above attempt to apply decision analysis methods using objectives focused on actions after the intelligence has been (or sometimes hasn't been) used.

There are a number of reasons some decision makers may not use the intelligence from the IC. According to Lowenthal, decision makers are free to ignore intelligence and estimates from the IC and often do so.[70] Some decision makers believe that the analysis the IC produces leaves too much that is "uncertain and ambiguous" and that with their multibillion dollar budget, the IC should be able to know more than what they do in current practice. Also, policy makers have been known to be suspicious of intelligence that supports their political rivals instead of their own position. Finally, as the Defense Science Board addressed, decision makers sometimes make decisions based on politics rather than intelligence.[71] In applications of decision theory, if a decision is not made based on the information or rather is made regardless of the information, the information is valueless.

For these two reasons, we argue that intelligence is a specific subset of information that current valuation methods are inherently problematic. Not only can different levels within the IC not agree on what makes intelligence valuable, if the decisions made by policy makers are not made based on information (and arguably it is difficult to ever determine which decisions were made using intelligence and which were not) then the information is valueless. Since we agree that the collection, analysis and production of intelligence is a valuable process and should not stop, we argue that instead of attempting to apply these methods to valuing intelligence, a new method should be applied.

2.7 Value within the Intelligence Cycle

Decision analysis is a powerful tool, but must be used appropriately. We suggest a supplemental method of valuing intelligence by choosing objectives of valuation while the intelligence is still in the intelligence cycle. We suggest two possible methods for providing a supplemental and quantitative value of intelligence within the intelligence cycle. The first method is based on measuring a piece of intelligence's impact on analysis with respect to the entire picture. How might the analyst's recommendation change based on not having that information? The second method would be to apply existing dynamic social network methodologies to a corpus of intelligence documents to demonstrate a measure of importance within the structure of that corpus of documents.

We argue that valuing intelligence based on its impact on analysis addresses the difficulties above. In many of the proposed measures of information within the IC, the information tends to be valued based on decisions made, either with a battlefield commander or by a political entity. While this view can be valuable, it can also have some theoretical and practical drawbacks. In decision theory, if it can be shown or perhaps even posited that decisions made were made irrespective of the information given, as sometimes happens in strategic decision makers, then by definition, the information has no value to the decision. Additionally, while information or intelligence can be extremely accurate, there are many human reasons why decision makers make decisions seemly inconsistent with the information given. For this reason, we suggest a supplemental measure of the value of intelligence. Something that can be measured while the intelligence is still in the intelligence cycle, before or even after it is disseminated, but based on the picture that it helps to create of what might be going on with respect to the subject of the analysis. One demonstrated way of determining the effect information can have on analysis is sensitivity analysis. Sensitivity analysis identifies those variables that cause the end result of analysis to vary, and determine an amount of variability. Methods such as tornado diagrams can quantify the impact of a variable on the analysis. In conventional analysis, once risk managers or the IC equivalent have attempted to mitigate the variability by vetting sources and content, not much attention is paid to sensitivity analysis. We argue, when the variable is information/intelligence, that this sensitivity analysis identifies what intelligence would impact the analysis most if it did not exist. Looking at this in a different way, this intelligence has a strong impact on the analysis, and, therefore, has value to its end result.

Some in the IC hold the belief that no piece of intelligence has value, but rather its value comes in the overall picture that they help the analysts determine of what is going on in the picture. If one looks at analyzing intelligence as assembling a jigsaw puzzle where analysts attempt to fill in pieces of multiple puzzles with unknown number of pieces, sometimes with deceiving pieces, not all the pieces and the pieces change in time. Another way that one might demonstrate the objective value of information, within the intelligence cycle is to measure the impact it has on the overall picture that is created using this intelligence. We propose another way of looking at the valuation of intelligence and using this analogy, i.e., we should value the pieces of the puzzle that contribute greatly to our understanding of the overall picture. This means valuing either the largest pieces or the key pieces that contribute to overall understanding. Networks are a common analysis tool and can be a pictorial representation of the overall structure of the subject of analysis. Networks are already being used to explore the structure of information and show quantitative value of pieces of information, most notably the PageRank algorithm which is the driving algorithm behind the Google search engine.[72] For more examples reference Radicchi et al.[73] Dynamic network analysis (DNA), introduced by Kathleen Carley, takes social networks to a new level of understanding the structure and impact of individuals within organizations.[74] We propose that DNA could be adapted to analyze the structure of a corpus of documents as well. In this way, we could identify, not only key structural components of a network of information, but also identify structurally critical documents within that corpus of documents. In this way, the authors propose that depending upon the question interesting to the analysts, they could quickly find documents which are important or relevant to that question. Since often times the value of information depends upon the context of the situation or question, this would help researchers quickly determine value of information or intelligence within a number of different contexts.

2.8 Future work and Conclusion

The authors believe that methods of valuation that involve objectively measuring the value of information or intelligence to analysis, within the intelligence cycle is an important supplemental technique. The authors will conduct further research in applying sensitivity analysis and network structure of the information based methods of valuation. There are already sensitivity analysis methods for many decision analysis techniques. These should be explored to determine adaptability to this valuation method. Additionally, the authors have argued there is inherent value within the structure of the intelligence, and that using a network approach, we can demonstrate value with respect to context. Using the analogy from above, if there is a picture that might explain a situation, in analytical terms a picture is often referred to as a graph. We shall develop a method of determining the impact of analysis on a network using graph theoretic approaches. Preliminary results show significant promise, in each of these approaches, to valuing information within the intelligence cycle and the associated sensitivity analysis. Once appropriate methods have been found and applied to individual pieces of intelligence within separate analyses, the big picture of intelligence valuation can be investigated. If an intelligence source is consistently found to produce objectively high “valuing” intelligence, we can determine the relative “value” of intelligence that sources offer. These can be used to assist collection managers with their cost benefit analysis in determining appropriate asset allocation and provide objective, quantitative support to varied decision makers as evidence of the benefit of sources.

3 Value of Information Applied to Networks

3.1 Introduction

Intelligence collection and analysis have been compared to solving jigsaw puzzles. The complexity lies in that the puzzle set may have 1000 pieces, but the Intelligence Community (IC) is only able to get a small subset of the pieces and must try to determine the overall picture. This example demonstrates the problem facing many intelligence professionals.

In this analogy, each puzzle piece represents a piece of intelligence, and inherent on each puzzle piece is a pattern that might help the analyst determine where the piece goes in the overall picture. Occasionally, there is a puzzle piece, analogous to actionable intelligence, something like the location of Ayman al-Zawahiri for the next two days, but except for these rare bits of intelligence, each piece has value in the picture that it helps to create. Intelligence professionals and academicians have long been searching for a way to identify "good" or "valuable" intelligence. Using this representation of intelligence, we shall demonstrate a methodology to show "value" to the overall picture each piece of intelligence has by quantifying each piece's impact on the picture.

Intelligence analysts use network graphs to show pictures of everything from transnational criminal organizations, flow of illicit money through banks, to social networks of terrorist organizations. Each of these networks are pictures created using pieces of intelligence in a manner similar to the puzzle described above. Each piece of intelligence can add unique nodes or links to the graph. The methodology we use shall demonstrate the impact of each piece of intelligence by quantifying the impact of those unique links and nodes that it adds to the overall network.

3.2 Background

Networks have been used to model a number of different phenomena in the real world from modeling business relations between companies, neural networks to networks of citations between papers.[75] A network is comprised of a set of items called nodes, or sometimes vertices, and connections between these nodes called links or edges. A collection of these nodes and links is called a component, sometimes networks contain a number of components that are not connected to each other. A network is used to understand the structure and relationship of nodes by viewing them in relation to each other and their links. An example of a network that one might use every day is the World Wide Web (WWW). Each of the separate pages on the WWW can be thought of as nodes, while buttons or hyperlinks embedded on the page that allow the user to go from one page to another can be thought of as links. An example of a network is shown in

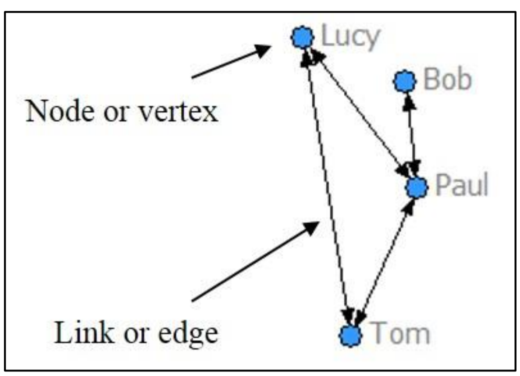

Figure 8 Example of nodes and links

Figure 8. Viewing nodes and relationships together allows researchers to explore their relationship to one another and identify patters and groups. Occasionally, networks have two different classes of nodes, called bipartite networks. Typically networks are limited to displaying only two types of nodes, especially since networks are also commonly represented by

a matrix, where nodes are the titles of the columns and rows.[76] Borgatti et al. provides a good discussion on the principle theories behind network science.[77] In recent years, the availability of relatively inexpensive computing power has allowed researchers to view and analyze networks on a much larger scale than in the past.[78,79,80]

Networks are used for a variety of different reasons within the IC. One of the applications is social network analysis (SNA), though there are a number of other examples where networks can be useful. Network charting, a process of identifying people, groups, things and drawing connecting lines between them showing various types of association, and network analysis, the process of taking a network chart and trying to make sense of the data represented by looking for patterns among the data are two examples of other uses of networks.[81] SNA tends to be the most mathematical and analytical of the network methods, though many of the techniques and methods can be applied to different types of networks. A good example of an SNA intelligence application might be from Kennedy et al using a social network to determine the most effective forms modeling the disruption or destruction of the network.[82] The authors explore the use of a network to interrupt maximum flow of information through a network by using interdiction. Additionally, Overbey et al looked at a network of twitter users specifically to determine influence of users during the 2011 Egyptian revolution.[83] Some have had concerns with using intelligence driven networks to apply network theory, useable results can still be drawn. Most notably, Malcolm Sparrow listed three main challenges to using networks in intelligence work:

- Incompleteness – analysts will not necessarily have access to all information allowing them to include all nodes and links that exist in reality
- Fuzzy boundaries – difficulty in determining agents or nodes to include or exclude in the network

- Dynamism – these networks, like many in real life, change over time and are dynamic, not static

Even though these complications exist, Sparrow argues that analysis of networks using network theory is still useful to the intelligence community.[84]

The methodology in this research uses a technique of deleting links and sometimes nodes to gauge impact on the overall network. Others have used deletion within a network for a number of purposes. Borgatti et al. use deletion (and addition) of random nodes within a number of random networks to simulate random error within a network. They then gained a confidence interval around some centrality scores to measure robustness of the scores.[85] Glacet et al. used deletion of edges to measure the effect of topology on the number of liars one might expect to encounter in a network of advice givers on directions.[86] These deletion techniques tend to be focused on the purposes of the graph or testing the resiliency of different measures. To our knowledge deletion techniques have not been used to evaluate information.

Though the intelligence profession is arguably the second oldest profession, there has been little work on the field of measuring the importance of intelligence.[87] Most recently, the work from Gill et al. has attempted to define intelligence and begin to construct an intelligence theory.[88] An important aspect of the theory of intelligence is how one measures intelligence, or "values" intelligence. Though this methodology is clearly not general enough to be used in this fashion, it might serve as an example for quantitatively assessing the "value" of intelligence rather than assessing it qualitatively.

One method of sensitivity analysis is a one way sensitivity analysis method called value of information (VOI). It is applied to an influence diagram, such as the one in Figure 6.[89] A method of visualizing the VOI is called a tornado diagram, such as the one seen in Figure 7. A tornado diagram shows the magnitude of the effect a variable has on the objective when all other uncertainties are held constant, sometimes called a single factor sensitivity analysis. If one creates a model, such as a decision tree, or an influence diagram, they can use a one way sensitivity analysis tool called a tornado diagram to show the impact of the variable on the overall decision recommended by the model. The tornado diagram allows the modeler to vary one part of the influence diagram while holding all other variables constant. Bodily et al. describe a six step process of value creating within a company. [90] One of the steps involves evaluating the risk and potential return of the alternatives, where Bodily et al. demonstrate using an influence diagram and a corresponding tornado diagram showing the sensitivity of the net present value (NPV) to the variables within that influence diagram in Figure 6 and Figure 7. The purpose is to not only communicate to decision makers the variability within their model, but also to identify the variables with the greatest impact, both positive and negative on the NPV of the alterative.

3.3 Value of Information (VOI) Applied to Networks

3.3.1 Deterministic Sensitivity Analysis Approach

We propose to apply the one-way sensitivity analysis, value of information (VOI) concept to networks. Once a network has been analyzed using a network theory metric, like degree centrality, or betweenness centrality, etc. Our methodology, put simply, is a deterministic sensitivity analysis of each link, where we run the relevant measure(s) for the original network, delete one link, run the relevant network measure(s) again, develop an "average" value for that network, then replace the deleted link, choose another link and start the process again. This would yield a collection of "average" values one could compare to the original "average" value.

The term average is in quotes because we do not believe that the actual average is applicable in this sense. Many of the naturally occurring networks, especially citation networks, in recent years been found to have a scale-free property which means that the number of links k, originating from a given node exhibits a power law distribution $P(k) \sim k^{-\gamma}$ where γ is a parameter generally in the range $2 < \gamma < 3$.[91,92,93] In other words the distribution of highly connected nodes follows a power distribution, rather than a normal distribution where the average might be appropriate. In order to have an "average" value to compare for these networks, we propose the use of the 95th percentile as an average measure.

When we compare the "average" values, the direction of the change is often times irrelevant, in other words only the absolute difference from the original "average" value is important to the amount of impact the link has on the original network. Additionally, since normalized metric values are numbers between 1 and zero, the percent difference can be more informative than the actual difference.

Therefore, we find the percent of absolute difference in the 95th percentile between the original network and the networks created by iteratively deleting a link, then replacing the link and deleting a new link. This will allow analysts to determine how impactful each link is to the overall structure of the network with respect to the metric used.

3.3.1.1 95th Percentile as an "average" measure

The measure of percentile is defined by Devore as the 99th percentile separates the highest 1% from the bottom 99%, or in this case, the 95th percentile separates the highest 5% from the bottom 95%.[94] We propose that the 95th percentile can capture the "average" value of many metrics and can be used to determine overall change across the network.

3.3.1.2 Network measures' resiliency to deletion of links

Reasonably, one might wonder how resilient network measures are to deletion of links. There has been a number of works exploring this question.[95,96,97,98] The end results are that for most traditional centrality metrics, networks are largely resilient to minimal node deletion. Of course we would like to explore this effect in other networks with additional metrics.

3.4 Notional Example

We decided to explore this methodology on a data set using a notional example. In the late summer of 2011, riots occurred in London, UK. Throughout the next month, it's estimated that 5 people died and over 300 police officers were injured.[99] There were over 234 recorded riot related incidents centered on London but spread over the UK.[100] Understandably, government intelligence analysts might want to have current situational awareness of riot related occurrences as well as the temper of the populace. There are only limited resources to be able to monitor twitter users, so analysts need to determine a subset of the twitter users tweeting about London riots to monitor in order to keep their finger on the pulse of the populace.

3.4.1 Data set

The researchers at Johns Hopkins University's Applied Physics Lab (JHU/APL) mined public twitter reports, geographically located around the London riot locations, and those using hashtags that identified the tweet as something related to the riots (e.g. #londonisburning, londonriots, etc). From this, they created a data set of twitter users tweeting other twitter users about the London riots. Using this data set, the authors created a network where the nodes are twitter users and the links connecting them exist if one twitter user either sent or received a tweet from another twitter user. Since each tweet either received from or sent to another twitter user is evidence of a connection that exists between two users, this network was symmetrized and

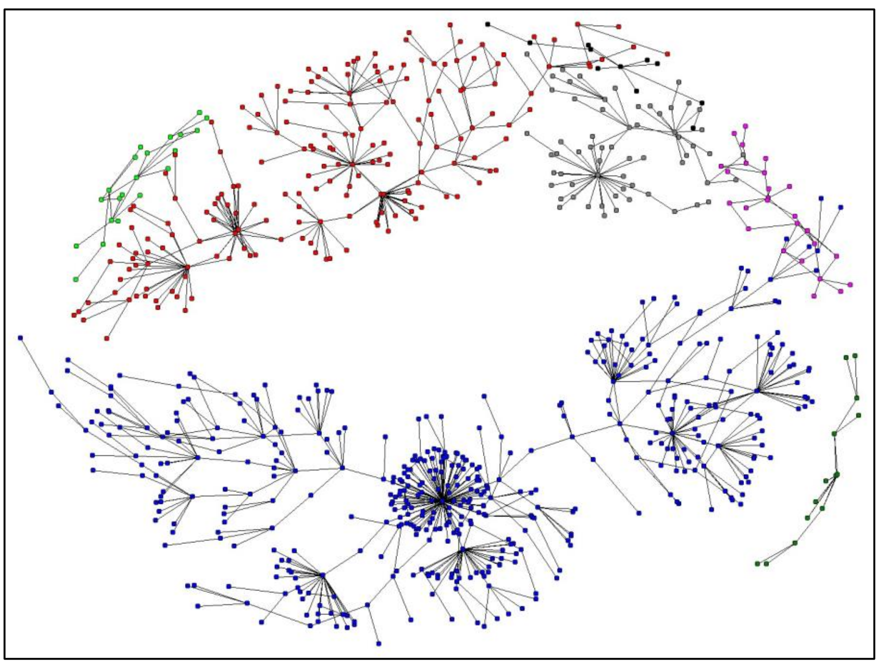

Figure 9 Graph of entire network

directionality was deleted from this graph. Additionally, the largest seven components were retained in the network, removing all isolates (nodes not connected to any other nodes), pendants (two nodes with only a connection between the two nodes) or triads (three nodes connected to one another, but to no other node).The network created has 770 nodes, 850 links between the nodes and contains seven different components (see Figure 9).

3.4.2 Problem

Using this graph, notional analysts can apply the network metric of betweenness to identify the nodes with great influence on the information network. Betweenness is a measure of the number of times a node lies on the shortest path between two other nodes. This value can be normalized so that nodes betweenness values on different components can be compared to one another. Betweenness centrality has been used to explore a number of different networks such as identifying the key actors in terrorist networks[101] as well as other twitter networks.[102,103,104] Calculating the normalized betweenness values for the nodes can help analysts determine which twitter users to dedicate limited monitoring resources.

3.4.3 Notional Solution

A notional analyst can run the betweenness values on the data set we had created, and capture the collection of 5% (39 users) of the twitter users with the highest betweenness values (see Appendix A). Using this list, the analyst can apply monitoring technology to follow the twitter traffic flowing across the twitter accounts of these users.

3.5 Application of methodology

Since the notional solution determined in 3.4.3 is dependent upon the structure of the network created, they might want to determine how sensitive their solution is to the links given in the data set. If we assume that each link represents an intelligence report of the existence of a twitter link between two twitter users, we might be able to find which of these intelligence reports would cause the notional solution list to change the most if analysts did not have that intelligence report.

We coded in python a program to iteratively delete each link, run betweenness values for all remaining nodes, and then replace the link and repeat the process with another link. We then found the 95th percentile value for all of the resultant networks and compared them to find the percent of absolute difference from the original network. In this way, we mapped out the impact of each link on the overall network.

Additionally, to validate our metrics we found the number of Twitter users that would change in the list of the top 5% of betweenness nodes within the network. The results can be found in Table 5 below. We then tested our 95th percentile of percent absolute difference for statistical significance compared to the statistical average of percent absolute difference.

3.6 Results

In the Table 5 below we can see that the link between kenworthy39 and PCStanleyWMP is

From	To	95th Percentile Norm. Betweenness	Abs % Diff from Original	# of Twitter users changed from to original list
kenworthy39	PCStanleyWMP	0.0071	37.8%	12
PCStanleyWMP	MasherMiles	0.0071	37.7%	12
rioferdy5	bignarstie	0.0071	37.6%	10
rioferdy5	MasherMiles	0.0071	37.5%	12
youtube	bignarstie	0.0072	37.3%	16
youtube	MyrtleTakesTea	0.0072	36.9%	16
iamwill	100Monkeys_Fan	0.0072	36.8%	16
iamwill	haydenmead	0.0082	27.9%	16
bignarstie	EmTalib	0.0090	20.8%	6
FFMonst	EmTalib	0.0090	20.7%	6
YoungRv	FFMonst	0.0091	20.6%	6
hannahchw	PolicingToday	0.0094	17.9%	6
…	…	…	…	…
MissBeRavin85	AleshaOfficial	0.0103	9.8%	4
…	…	…	…	…
nickkeane	kenworthy39	0.0120	5.3%	2
…	…	…	…	…
thefadotcom	craigcampbell_	0.0113	1.0%	0
…	…	…	…	…
bignarstie	KushKwame	0.0114	0.0%	0

Table 5 Partial listing of Absolute % difference in betweenness from original network and change to original list

the most impactful on the network as determined by our metric described above. In other words, if the network did not contain the link between these two users, the end list of the top 5% of betweenness users would change by 12 people, or by over 30%. Each of the 850 links were measured for 95th percentile of normalized betweenness and compared to the original network 95th percentile. From this, the authors selected the top twelve links and intermittently along the rest of the list to determine the number of Twitter users that would change in the top 5%

betweenness list. Although there are four links where a higher number on the list changed than those with a higher impact value we propose, we measured the statistical significance of the result. We found that our measure had a .91 correlation with 3.3E-7 significance, compared to a comparison of the statistical average percent absolute difference, which had a .61 correlation with 0.01 significance. This demonstrates that our metric overall was more highly correlated with picking the number of users changing from the top 5% list.

The result of this analysis is that the analyst now can review the network they've created and perhaps verify that the link between the top eight most impactful links are actually legitimate. They can actually view the tweets between these users and verify that the tweets are informational and not spam floating between Twitter users. Upon verifying those links, the analyst can be relatively sure that their limited Twitter monitoring assets are being used appropriately.

3.7 Conclusion / Future Work

For this example, we demonstrated the effect of deleting the information given by links and treating them as individual intelligence reports of a twitter connection between two users. Upon the completion of the analysis based on the notional example, we applied our methodology to determine the impact that each link has on the overall network, based on the metric of betweenness. We believe that this methodology could be applied to a whole host of naturally occurring networks to show impact of the intelligence to the network.

How much value a piece of intelligence has can be a very subjective thing, with people bringing different definitions to what makes something valuable. On the other hand, we can quantify how much impact a piece of intelligence has on our overall picture of understanding on a topic of interest using this methodology. The goal would be to try to measure the value of information that each puzzle piece brings to the overall picture.

We would like to explore the effects of scaling to our methodology by applying our methodology to more networks, perhaps even non-Twitter networks. Ideally we would like to explore actual intelligence data sets for applications of our methodology. In real life intelligence, one document might imply the existence of multiple nodes and links. We would like to explore our methodology on the deletion (and resiliency of our methodology) on valuing intelligence reports where to determine impact, multiple links or nodes are deleted signifying that the intelligence report did not exist.

Finally, since most real-life networks change over time, we would like to explore our methodology's application in a temporal nature, showing how a document's impact could change over time as the structure of the resultant network changes.

3.8 Introducing intelligence Potential

Building upon information theory and applications of VOI in decision analysis and network impact, we introduce the application of this paper. In a similar manner to Shannon's development of the entropy of a message, we propose to name this proposed measure potential. Once the intelligence has been used to create a model of our understanding of some situation (a decision, a network or other situations), the analyst uses the model to answer some analyst question. As described in VOI and Network impact above, once the question has been answered, analysts can look back to determine which pieces of intelligence had the most impact and quantitatively assess that impact. One can determine the quantitative impact or "value" on the result of a question, once the intelligence is placed into the perspective of the model.

If we look at the numerical impact of that intelligence for that application, we might apply something similar to the coefficient of determination or R^2. R^2 in linear regression is defined as the "proportion of observed y variation that can be explained by the linear regression model."[105] In this case, we propose that the potential of a piece of intelligence is defined as the average proportion of impact over all applicable analysis techniques. The metric would involve dividing the impact for a particular model the total amount of impact for all intelligence within that model, or $I_i = \frac{impact_i}{\sum impact_{total}}$. That I_i term would be the impact for that specific piece of intelligence for that model; the larger the term, the larger the impact. The dimensionless term ranges between 0 and 1, with 1 meaning the intelligence was the only thing that impacted the model, and 0 meaning the intelligence had no impact on the model. Taking the average of all I_i would give us the average impact, or potential of the intelligence.

This average would be a measure of the impact for the piece of intelligence over all applicable analysis, thus a quantitative measure for intelligence with respect to the questions/analysis applied.

3.8.1 A notional application of finding Potential

Suppose there are two notional pieces of intelligence, each used in a number of analytical studies. Also assume that each of the methodologies of the studies had a way to measure the value of information or in this case, intelligence. One piece of intelligence, called Intelligence A, was used three analytical studies, one decision analysis and two network studies. Applying the potential methodology suggested in this paper, the fraction of value of information in dollars, over the sum of all the value of information, in dollars, over the study, for that study, the value of that intelligence was .33 ($10,000/$30,000). In a similar manner, we could find the value of the intelligence in the other two network studies (.44 and .16). If we take the average value of intelligence over all the studies, we find that overall the intelligence is valued at .31. Using the other piece of notional intelligence, Intelligence B, and the same methodology, we find that the value of intelligence for that piece of intelligence is .18 (average of .01, .47, .10, and .15). See Table 6.

In this notional example, the potential of quantitative analysis of Intelligence A is greater

Potential of Intelligence A			
	Analysis 1	Analysis 2	Analysis 3
$impact_i$	$ 10,000	3500	1208
Sum of total impact	$ 30,000	8003	7546
I_i	0.33	0.44	0.16
Potential	0.31		

Potential of Intelligence B				
	Analysis 4	Analysis 5	Analysis 6	Analysis 7
$impact_i$	54	$ 56,000	$ 102,000	12
Sum of total impact	6033	$ 120,000	$ 1,000,000	80
I_i	0.01	0.47	0.10	0.15
Potential	0.18			

Table 6 Notional example of comparing potential of intelligence applications

than the potential of Intelligence B. Even though Intelligence B has been used in more studies, Intelligence A has been more impactful in the studies in which it's been used. In this way, analysts can use a quantitative and objective measure of potential to supplement a subjective assessment of the value of a piece of intelligence.

3.9 Future Research

The next step for this research is validation. We shall apply our methodology, to a number of analyses that use the same pieces of intelligence and demonstrate our valuing methodology on the intelligence. We might then compare our results to analyst feelings on the importance of different pieces of intelligence.

In future work we shall want to develop additional ways of finding the impact within methodologies, similar to VOI in decision analysis and within impact in networks described in this paper. This methodology would become more robust if there were additional ways of calculating the impact.

The measure of impact proposed in this paper is meant to act in a way to supplement, not replace current and more subjective means of "valuing" intelligence. Potential would provide an objective measure of the impact of a piece of intelligence on quantitative analysis, and measure the value within the intelligence cycle. Additionally, this measure does not represent the "truthfulness" or accuracy of the intelligence. If in the future, the intelligence is found to be incorrect, analysts can remove the intelligence from the analysis, and reassess other pieces of intelligence's impact.

4 Layered Network Approach

4.1 Introduction

During wartime operations or in successful intelligence gathering operations frequently the result is a large collection of documents. Often time sifting through these documents to determine valuable information is a long laborious process made even more difficult by the fact that often times the documents are in a different language than the gaining community. Translation resources are scarce and already overworked, being able to prioritize documents to focus translation effort, or simply being able to determine which documents to read first might be helpful. The value of a document within a collection or corpus of documents might change based on what question the analyst wants to answer. If the corpus consisted of academic journals, the question of "what are the foundational authors working on?" vs. "what are the foundational works within this corpus?" might have completely different answers, which would change the value of each document. In order to assist with this problem, assume that translation efforts can be made to determine the author(s), title of the paper, the journal it was published in, the year of publication, and key words associated with that paper. We introduce DNA applied to a network of information documents consisting of this information.

4.2 Background

Networks have been used to model a number of different phenomena in the real world from modeling business relations between companies, neural networks to networks of citations between papers.[106] A network is comprised of a set of items called nodes, or sometimes vertices, and connections between these nodes called links or edges. A network is used to understand the structure and relationship of nodes by viewing them in relation to each other and their links. An example of a network that one might use every day is the World Wide Web (WWW). Each of the separate pages on the WWW can be thought of as nodes, while buttons or hyperlinks embedded on the page that allow the user to go from one page to another can be thought of as links. An example of a network is shown in Figure 8. Viewing nodes and relationships together allows researchers to explore their relationship to one another and identify patters and groups. Occasionally, networks have two different classes of nodes, called bipartite networks. Typically networks are limited to displaying only two types of nodes, especially since networks are also commonly represented by a matrix, where nodes are the titles of the columns and rows.[107] Borgatti et al. provides a good discussion on the principle theories behind network science.[108] In recent years, the availability of relatively inexpensive computing power has allowed researchers to view and analyze networks on a much larger scale than in the past.[109,110,111]

Two of the larger areas where networks have been applied are in citation networks and in social network analysis (SNA).[112,113] Citation networks are structures where bibliometric datasets are used to create structure from academic journals. There are a variety of different citation networks that have been created, for example a citation network between academic papers can be created by treating the papers as nodes and references of the papers as links to other papers. This basic graph would be directed (one paper refers to an earlier published paper, so the link only goes in one direction) and unweighted, since one can assume that each referenced paper carries the same weight. In this same way, networks can and are created to explore the structure between papers, authors and other meta-data developed from a journal article. [114] Although these networks can be used for a variety of things, one purpose they are used for is to rank journal articles using a variety of metrics. Since 1927, citation networks have been used to rank journal articles using different metrics.[115] Many of the more modern rankings take advantage of metrics from network science, the most popular being the number of other authors or papers that cite a paper or author, but new metrics have been developed as well. The *h* index is a popular metric to measure "the cumulative impact and relevance of an individual's scientific research output."[116] Another example of a network metric used for ranking academic journals is PageRank, originally designed for helping to display the "importance" of a website in Google's algorithm.[117]

Another area where networks have been applied in recent years is SNA. A social network is comprised of "a set of people or groups of people with some pattern of contacts or interactions between them."[118] Networks are used to study the structure of these relationships for a number of different reasons such as exploring patterns of friendships,[119] intermarriages between families[120] and terrorist organizations[121,122] to name a few. Social network applications have also been applied to information.[123,124] In SNA, analysts study the links between people and sometimes events or locations independently. They use network metrics to determine critical individuals or to understand how the network operates.

A relatively recent method has been applied to SNA to take better advantage of the wealth of information surrounding social interactions, called dynamic network analysis (DNA).[125] In DNA, meta-matrices are created to explore different interactions. For example, an agent x agent (AxA) meta matrix would have interactions between agents, where a non-zero element in an *i*,*j* location means that agent *i* is connected to agent *j*, while a zero denotes no interaction. The strength of the interaction could be captured in the magnitude of the non-zero element, or there could be just a value of 1 in a non-weighted meta-matrix. In this same way, meta-matrices are developed amongst different permutations of agents, knowledge, resources, tasks/events, organizations and locations.[126] These meta-matrices are like lenses. In DNA, the analyst can choose different meta-matrices depending upon the information they are interested in, and view them together on one network. The network then becomes a multi-nodal network that includes interactions between multiple nodes, not just interactions between one or two classes of nodes. Additionally, probability is applied to the links because often times there are some questions as to either the existence of the link or its strength. Many of the current applications are to terrorist networks.[127,128,129,130]

4.3 DNA concepts applied to information networks in the intelligence field

We argue that DNA can be applied to information networks to gain greater flexibility and understanding out of the structures of the information. We introduce applying the meta-matrix concept to information networks. In this example, we introduce an application of meta-matrices using the meta-data from academic journals, though this could easily be applied to other information networks. The meta-data we chose to model in our information meta-matrices are the title of the paper, the author(s), key word(s), publication date, the journal in which the paper appeared, and the author(s) institution. Given this collection of meta-data, in a similar way of displaying meta-matrices as Diesner et al, we offer Table 7 as a suggestion of meta-matrices developed from academic journal meta-data.[131]

	Paper	Author(s)	Key Word(s)	Publication Date	Journal	Author(s) Institution
Paper	P - Matrix Citation Network (*Which papers cite which papers*)	PA - Matrix Author Network (*Which papers were written by which authors*)	PK - Matrix Contextual Network (*Which papers contain which key word(s)*)	PD - Matrix Publishing Precedence Network (*Which papers have been published when*)	PJ - Matrix Journal Content Network (*Which papers are from which journals*)	PI - Matrix Institutional Capability (*What titles are where*)
Author(s)		A - Matrix Co-Author Network (*Which authors wrote papers with which other authors*)	AK - Matrix Knowledge Network (*Who is writing about what topics*)	AD - Matrix Author Timeline Network (*Which authors have published when*)	AJ - Matrix Literature Network (*What authors write for which journals*)	AI - Matrix Employment Network (*What authors work where*)
Key Word(s)			K - Matrix Information Network (*Which key words inform on key words*)	KD - Matrix Key Word Timeline Network (*What key words were published about when*)	KJ - Matrix Journal Context Network (*What key words describe which journals*)	KI - Matrix Competency Capability (*What knowledge is where*)
Publication Date				D - Matrix Timeline Network (*What publication dates interact with publication dates*)	JD - Matrix Journal Timeline Network (*What publication dates are journals published during*)	ID - Matrix Intitutional Timeline Network (*What publication dates are organizations publishing during*)
Journal					J - Matrix Citation Network (*Which journals cite which journals*)	JI - Matrix Institutional References (*What journals get their papers from where*)
Author(s) Institution						I - Matrix Inter-organizational Network (*Which organizations work with which organizations*)

Table 7 Meta-matrices for academic journals

These interaction matrices listed here are mathematically defined in Appendix B. They show interactions between the two node classes of the matrix. For example, the **A** matrix, is an interaction matrix with authors on both the columns and rows of the matrix. Additionally, if there are *R* number of authors, then the matrix is of the order of *R*x*R*. We have defined the **A** matrix as A*ij* is equal to *n* if and only if author *i* is a co-author with author *j* in *n* particular published works; otherwise A*ij* = 0. This matrix is a symmetric matrix with weights. This is an example of a co-author citation network that answers the question: *which authors wrote papers with which other authors*?

The benefit to the structure is that given the variety of different meta-matrices, depending upon what an analyst chooses to explore, they have a number of matrices or lenses to put together in a network. Additionally, these matrix constructs allow the generation of other information like reachability, and inference based on existing information. Reachability matrices (some are mathematically defined in Appendix C) allow analysts to determine how closely two nodes (of the same time or different) are aligned. For example, in an $\mathbf{A}^1$ matrix, one could determine which authors were connected by a paper that they wrote together by looking at the column and row where they intersect and there would be a "1." An $\mathbf{A}^2$ matrix would show people connected by two steps or less, or if an author *i* wrote an article with author *j*, and author *j* wrote an article with author *k*, then the row and column for author *i* and author *k* would have a "1" showing that they are two or less steps connected to one another. Used with authors, analysts can determine closely aligned authors vs. ones not closely aligned.

An inference matrix is a matrix based on a mathematical multiplication of a meta-matrix in order to infer information based on existing information (some mathematical definitions for inference matrices are in Appendix D). While the **A** matrix would show if two authors have published papers together, and how many, one might ask if authors can be similar enough to infer that two authors know one another, but creating an inference matrix. If we took the author/knowledge matrix (**AK** that answers the question: *who is writing about what topics?*), and folded it or multiplied the matrix by its own transpose ($\mathbf{AK} \bullet \mathbf{AK}^T$), we would get an $\mathbf{A}_K$ matrix, which is the author interaction matrix inferred on knowledge, or in this case, key words. In other words, if authors are writing using the same key words, even if they have not published together, the matrix would display a number n, based on our definition, which represents the number of times each has published using the same key words. The inferred information within this interaction matrix is that if the two are writing about the same concept, using the same key words, that they are connected. These two concepts have already applied with success DNA to organizations like terrorist cells.[132]

4.4 Intelligence questions applied to information networks

One might wonder why we demonstrate applying these DNA concepts to information networks like academic journals in order to demonstrate value within intelligence fields. Value can be a subjective term, what is valuable to one may not be valuable to another. Especially in the intelligence world, determining the value of intelligence can be difficult.[133] Smith et al. argue that defining "value" of intelligence can be difficult for a number of reasons.[134] They argue that other existing methods have focused evaluating value after intelligence has left the intelligence cycle and become "tainted" by the effects of politics, preference and sometimes just probability. They argue that a supplemental measure of the "value" of intelligence can be found within the intelligence cycle, before the intelligence is disseminated.

The intelligence cycle is the cycle within the Intelligence Community which turns harvested information into intelligence ready for dissemination seen in Figure 2. The first stage is planning and direction where the decision makers or intelligence managers direct that some target should be collected on and what different methods should be used to collect the information. The second stage is collection, where the various methods assigned to collect the information actually collect the specified information. The third step is processing, where the vast amount of information is translated, if needed and reduced down to the needed information to give to the analysts. Analysis and production is the fourth step where analysts actually convert the information into intelligence by "integrating, evaluating and analyzing all available data."[135] The final step is dissemination where the intelligence is distributed to the decision makers who may have initiated the process in the planning and direction step. Though it is usually displayed as a cycle where information goes from one step to the other, in reality the progression tends towards the dissemination stage, but frequently information/intelligence is moved back steps based on what the information contains.[136] For example, if while processing (stage 3) a satellite photo, an analyst notices that a location of interest is only partially captured by the photo, due to new construction, they might send the photo back to the planning and direction stage (stage 1) with direction to include more land mass in the photo. For a more detailed discussion on the intelligence cycle with a specific focus towards operations research in the intelligence cycle, see Kaplan.[137]

Determining value within the intelligence cycle, without some measureable results after using the intelligence can seem somewhat daunting. One example of an application of value of intelligence within the intelligence cycle is offered by Smith et al. in their work Intelligence Impact on Network Analysis and Quantitative Intelligence Analysis.[138,139] In these works, the introduce the idea of measuring a piece of intelligence's impact on a network developed from intelligence. Additionally, taking the method further, they showed the possibility of applying that type of method across all analysis using particular pieces of intelligence and introduced the measure of potential, where the potential of a piece of intelligence is the fraction between 0 and 1 of its average impact on quantitative analyses done using that intelligence. In this way, Smith et al. have demonstrated an application for demonstrating quantitative "value" to analysis.

The method proposed in this paper would allow analysts to develop their own metrics for determining importance. Sometimes there are different questions that might determine worth or

			Metrics				
			Betweenness Centrality	Eigenvector Centrality	Simmelian Ties	Total Degree Centrality	Appropriate MetaMatrices
Questions	1	Who are the most important nodes (authors, papers, etc.) within this network?	1	1		1	Any
	2	Which nodes (authors, papers, etc.) are the most popular?		1		1	Any
	3	Which nodes (authors, papers, etc.) are popular and connected to popular nodes?		1			Any
	4	Which nodes (authors, papers, etc.) are most likely to have information pass through them?	1				Any
	5	Which nodes (authors, papers, etc.) have the most ties to different cliques or subgroups?			1		Any
	6	Which authors are most likely to be writers of foundational knowledge within this corpus of documents?		1			$\mathbf{A}$, $\mathbf{A}_J$, $\mathbf{A}_K$, $\mathbf{A}_{J,K}$, etc.
	7	Which papers are most likely to have foundational knowledge within this corpus?		1			$\mathbf{P}$, $\mathbf{P}_J$, $\mathbf{P}_K$, $\mathbf{P}_A$, $\mathbf{P}_{J,K}$, etc.
	8	Are there any emerging or fading foundational papers within this corpus?				1	$\mathbf{P}$
	9	Are there any emerging or fading key words within this corpus?				1	$\mathbf{K}_P$
	10	Are there any emerging or fading authors within a research discipline?				1	$\mathbf{A}_P$

Table 8 Notional questions about a corpus with corresponding network metrics to measure them

value depending upon the use of the information. Table 8 suggests some questions researchers may ask about a corpus of documents, like a collection of academic journals or a collection of intelligence. Within the matrix of questions are appropriate network measures that can answer those questions. One can note that there may be more than just one metric that can provide a measure for the question. Some of the metrics are measures from network theory, but some of the measures are metrics developed by DNA analysts. This list is not exhaustive, but illustrative and a good start.

Using these questions, one can explore a corpus of documents using a network framework to start answering the question of value. Definitions for the metrics found in Table 8 are laid out in Appendix E.

4.5 Applied to data set

The data we applied our methodology to is from CiteSeerx (http://citeseerx.ist.psu.edu), a digital library primarily focused on the computer and information science. We downloaded the July 26, 2011 database in MySQL format. The data base consists of a number of linked databases with meta-data from over one million articles from the mid-1940s to 2011.[140] This consisted of information on over a million different papers. We reduced this data set to be more manageable by only including those papers which were cited over 100 times. This reduced our range of dates to papers from 1945 to 2007. From this collection of metadata, we queried the database to create edge lists, or spreadsheets containing a columns of "From," "To," and "Weight" which defines each edge connection. These edge lists followed the mathematical definitions found in Appendix B. We used ORA v.2.3.6, a dynamic meta-network assessment and analysis tool created and maintained by CASOS at Carnegie Mellon University to create our meta-matrices from the edge lists. This tool has been used extensively to explore meta-matrices in the social network environment. Our meta-matrices now contained the numbers of nodes shown below in Table 9.

Title	Number
Authors	5317
Journals	538
Key Words	3312
Papers	4793

Table 9 Numbers of nodes within network

The data within the CiteSeerX database was not very clean or organized well. The data had clearly been text mined from the various articles, which caused some parity errors. For example, some of the key words had the word "and" attached to the front of them. For this reason, the edge lists needed to be cleaned to ensure that the key words: "andclustering" and "clustering" counted as one node. We had to do this for each separate node within each of the different edge lists we created. For this reason, we did not create edge lists with author institution information, after some attempt at cleaning this data, we determined that it was just too unrefined to be cleaned within any reasonable time. In addition, we theorized that we might create meta-matrices including the publication date, but finally determined that we could add the date as a property of the paper and still accomplish our analysis. Future work may involve creating these matrices to see if there are ways to use them to explore further questions within the network.

Using a notional situation with the following parameters: if there were a collection of journal articles (4793 in this case) from a country of interest in their native language, we have certain translated meta-data from those articles but not the full translations of those journal articles, can we determine a priority of translation or at least a list of those articles that would be worth translating? In other words, which articles are valuable enough to dedicate sparse resources to translate and have government researchers read?

As discussed earlier, value questions can largely be subjective. Some questions come to mind when trying to determine the value of each of these journal articles. What would determine value? For this, we may look towards the chart in Table 8 for different questions that might help determine value. Current methods might involve using an author or paper citation network to determine, perhaps which papers are most central, or which authors are most central, and perhaps translate those papers.

5 Analysis/Results

We used our data set to explore four hypotheses.

Hypothesis 1. There is a difference in statistical significance between Turing Award winners and non-winners from 2007 until 2011 relevant metrics among an author citation network, and our $\mathbf{A}$, $\mathbf{A}_J$, $\mathbf{A}_K$, and $\mathbf{A}_{J,K}$.

Hypothesis 2. We can use our meta-matrices to determine the top recurring ranked authors within the corpus.

Hypothesis 3. There is a visible difference in some nodes when comparing author citation network, and our $\mathbf{A}$, $\mathbf{A}_J$, $\mathbf{A}_K$, and $\mathbf{A}_{J,K}$.

Hypothesis 4. Using our meta-matrices we can demonstrate the migration in importance of key words over a period of years.

Hypothesis 5. We can determine a number of papers from our original list, for translation and attention efforts to be focused on

5.1 Turing award winners vs. non-winners

We used an author citation network to explore Hypothesis 1. The A.M. Turing award is the Association for Computing Machinery's (ACM) most prestigious technical award, similar in the computer science world to the Nobel Prize. It is given for "major contributions of lasting importance to computing."[141] We identified the winners from year 2007 through 2011 within the data set, and used a pivot table to determine averages and standard deviations of the metrics given in Table 8. We then performed an *f*-test 2 sample for variance to determine if there was a statistical difference in variance of the metrics between winners and non-winners. We then performed a two sample *t* test assuming equal or unequal variance based on the results of the f test (p value ≤ 0.05). This let us know if there was a statistical difference between the average metric value for award winners and non-award winners. Since there are so few award winners

(7) to non-winners (5310) greater statistical analysis would be problematic, but this analysis could let us know if this network might be helpful in predicting award winners.

We also performed similar analysis and tests on an **A**, $\mathbf{A}_J$, $\mathbf{A}_K$ and $\mathbf{A}_{J,K}$ matrices to determine if there is a difference in statistical significance or perhaps an improvement in significance by selecting one of the meta-matrices. The results are in Appendix F. Highlighting some of the results, the author citation network, or AxA(Directed) does a reasonably good job in predicting award winners across most of the metrics chosen, though some of the meta-matrices improve on the statistical significance in some of the metrics. For example, the $\mathbf{A}_J$ matrix is significant to the 0.1 level in the betweenness centrality metric and is significant to the 0.003 level in Simmelian ties. A possibly reason for this might be that rather than looking at the selection of authors that cite other authors, looking at authors that write for similar journals might demonstrate a better structure in understanding the communication paths between authors and what different cliques they might have ties within. Additionally, the **A** network improves the significance in Eigenvector Centrality dramatically (0.017 to 7.0E-7). This might be because the co-author network might capture those foundational authors who publish together. Depending upon the question asked, one might have a better result in predicting award winners by using a different matrix.

It is important to note the metrics in the statistically significant measures for all matrices are higher for Turing award winners. The only exception is in the **A** network in Eigenvector Centrality. This is because all the award winners had a value of zero for the metric, as opposed to a non-zero result for non-winners in the same metrics. From this we can reasonably say that winners tend to be higher in these metrics for all the matrices.

This similar method applied to intelligence data sets can help analysts identify key agents or people within the subject of the data set. Key people within organizations can be found even if they are not the current head of an organization, or if there is no official head of the organization. Additionally, this method could be used to determine sources or authors of intelligence who are influential within the corpus.

5.2 Top recurring ranked authors

From exploring Hypothesis 1, we know that when looking at the metrics and Turing award winners as an indication of importance of the author, Turing award winners tend to be higher in each of the metrics explored. Using this logic we can explore the authors across many metrics and choose the top 10 ranking authors across all of the metrics in each matrix. This would allow us to get a good understanding of important authors within the corpus of data. The

Author	Networks in Top 10
MichaelBurrows	A-Dir, A, A_K
MichaelIJordan	A-Dir, A
MartnAbadi	A-Dir, A_J, A_K
AadCJDuinmaijer	A, A_J
AndantonPGWelbers	A, A_J
MarcelJMPelgrom	A, A_J
BarbaraHLiskov	A_K, $A_{J,K}$
MurdochJGabbay	A_K, $A_{J,K}$
AapoHyvrinen	A_K, $A_{J,K}$

Table 10 Authors in Top 10 of standard metrics

results for this analysis are in Appendix H. It is interesting to note that 9 authors showed up as the top ranking authors in multiple networks, see Table 10.

These authors have ranked within the top ten for a variety of network metrics including the ones used earlier in this paper. It is worth noting that the average H-Index (according to the ISI Web of Science calculation done at the time of writing of this article), a standard measure of evaluation of a scientific author's work, for this group is 34.2. The H-Index is a measure that calculates the *h* number of journal articles cited *h* number of times. The measure of an H-index of 12 is a calculation that might earn a prospective candidate tenure at a major research university, while at the time of the writing of the reference for this article, 2005, the highest H-index was 110 by Edward Witten.[142] The average H Index scores reflected in this sample reflect that the authors are well cited, but not extremely well cited. The authors believe that this relatively low H-index score is reflective of the inference made through the various levels. Although the authors might not have a large H-Index, they are very well connected within the various networks explored.

5.3 The visible effect of using different matrices

If you look at the standard author citation network, and were looking at metrics for

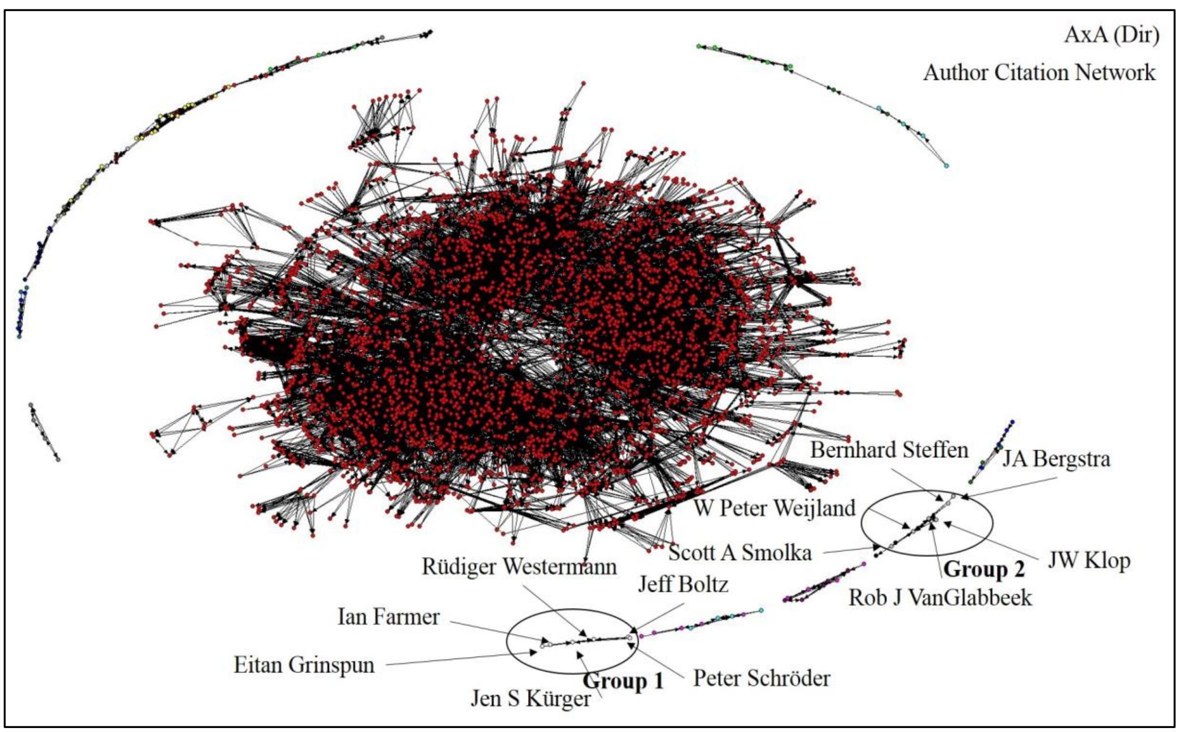

Figure 10 Diagram of Author Citation Network with two groups identified

possible interesting authors, you would probably overlook the authors of these two groups. In order to prove the hypothesis in Hypothesis 3, we chose some authors in the fringe of an author citation network and followed their migration towards the main component when looking at other meta-matrices. The picture in Figure 10 is the author citation network with all isolates and pendants removed (nodes with no connections or two nodes whose only connection is between themselves). The picture was made with UCINet v6.374.[143]

Each group is in the periphery of the standard author citation network, and are components to themselves. Although the papers they wrote (Group 1: Krüger et al. [144] and Boltz

Paper	Cited
Krüger et al.	688
Boltz et al.	697
VanGlabbeek et al. (1990)	490
VanGlabbeek et al. (1996)	749
Bergstra at al.	1007

Table 11 Number of citation for papers

et al.[145], Group 2: VanGlabbeek et al. (1990)[146], VanGlabbeek et al. (1996)[147], Bergstra at al.[148]) were pretty highly cited according to Google Scholar (see Table 11.)

Group 1 authors wrote about an innovation connected with Graphical Processing Units (GPUs) which were developed in 1999, but have only recently entered the mainstream market due to innovations similar what the authors wrote about, which made GPUs as versatile as Central Processing Units (CPUs). This allowed for robust 3D visualization within computers around 2003. Group 2 wrote about specific applications of process algebra, a topic which was one of the reasons why Robin Milner won the Turing Award in 1991. Process algebra helps to algebraically model behavior of any system, something very useful to computer scientists seeking to develop realistic models through programming. These authors wrote on very specific innovations, relevant to those interested in GPUs or process algebra, but not as well cited by those in the mainstream computer science world. For this reason, they reside on the periphery of the author citation network, and therefore would be marginalized in network metrics.

As noted earlier, Turing award winners ranked higher in the metrics, so higher metric scores equate to stronger authors. If we just looked at the metrics for these authors in the standard author citation network analysts might not recognize the importance of these authors. On the other hand, if you look at a number of the different inference matrices developed

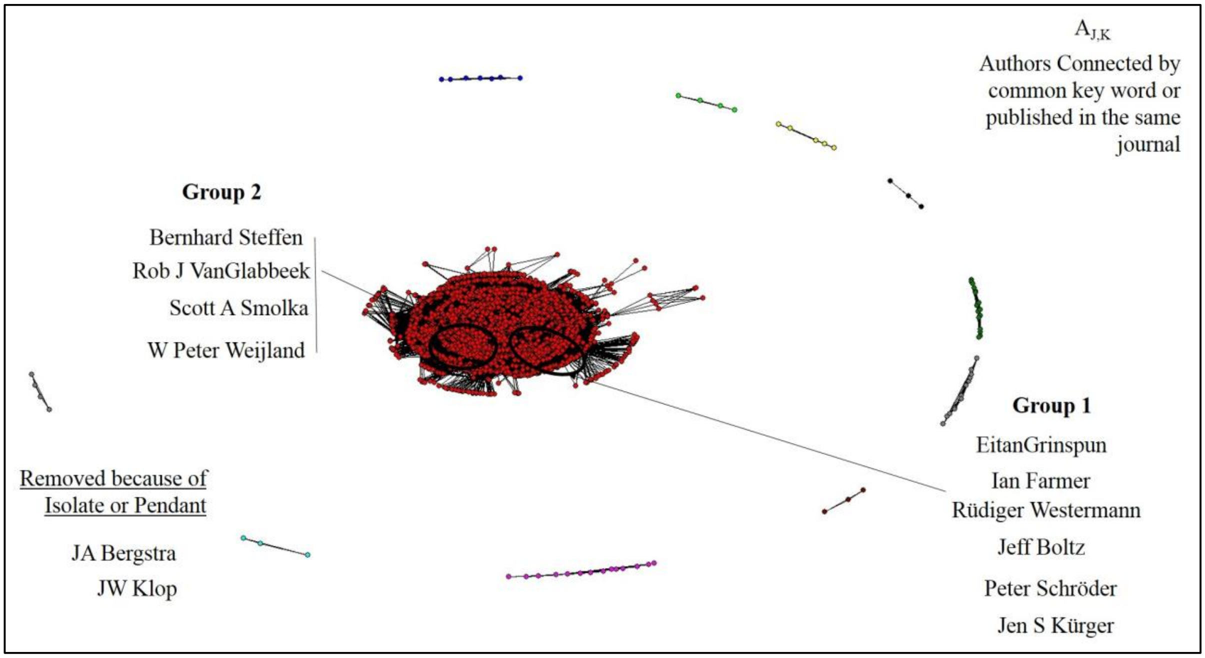

Figure 11 Diagram of $A_{J,K}$ Network with two groups identified

using this methodology, you see that they are, in fact, more connected than the average in a number of metrics. Because of inference connections based on the key words they used, and the journal they published in, we see that they are in fact more important than one might have previously thought. See Appendix G for the actual values of the various metrics explored. Specifically if we look at the metrics for the $\mathbf{A}_{J,K}$ network, we can see that if we use the fact that these authors are writing for popular journals as well as writing about popular key words, we see that their metric values rise. The reason for this is that they are now captured in the main component within the network as seen in Figure 11.

In typical intelligence focused social analysis networks, sometimes people exist on the fringe of the network, because they may only have been seen contacting a few people, and none in the main component. This methodology could use things like key words or intelligence sources in common to view people or locations in a different structure where they may become more important than at first thought in classic social network analysis.

5.4 Analysis of key word popularity

In order to prove Hypothesis 4, that we can use this layered network structure, which is currently not available under existing methods, in order to analyze key word popularity. An example of this is demonstrated when we take the **KP** network which contains the property of year published of the paper. We can create different **KP** network containing only those papers published in a specific year. In this case, we chose to explore the key words between the years 2000 and 2004. If we fold these matrices (multiply them by their transpose) we get $\mathbf{K}_P$ or a K matrix where key words are connected if they occur on the same paper. We used this matrix to link the key word with the paper, and more specifically the date the paper was published, which is an attribute of the paper. This matrix tells us which key words connected by a paper within that year.

Using these matrices, we can explore the popularity of the key words by year, using the metric Total Degree Centrality, which counts the total number of links going into the node. In order to compare apples to apples, this number has been unitized since there are a number of components within the total network, some components larger than others. Using this metric, we

Rank	Key Word	Ranking Year
41	clustering	2000
41	clustering	2004
189	collaborativefiltering	2000
34	collaborativefiltering	2004
51	evaluation	2000
2	evaluation	2004
43	experimentation	2000
35	experimentation	2004
58	featureselection	2000
7	featureselection	2004
147	keywords	2000
129	keywords	2004
183	measurement	2000
33	measurement	2004
57	textclassification	2000
21	textclassification	2004

Table 12 Key word by year analysis

looked at the 300 most popular key words for papers published in the year 2000, as well as those published in 2004 in order to see change in popularity amongst the key words in the 5 years. In Table 12, we see all key words that were in the top 300 key words in both 2000 and 2004.

We see some interesting results in the green highlighted sections. We notice that the key word "clustering" stayed pretty constant as a key word throughout the years. On the other hand, "collaborativefiltering," "evaluation," and "textclassification" all rose in popularity over the five years. It also looks like "collaborativefiltering" and "textclassification" were actually more popular in the years between 2000 and 2004. This analysis can help researchers identify and map research trends over time. Applied to intelligence, this can help analysts determine trends among a variety of topics such as terrorist activities, tribal names and specific types of attacks.

This type of analysis will help researchers realize the emergence (or decline) of key words related to specific terrorist attacks (i.e. Specific type of IED, etc), locations, resources or leaders. This will help researchers identify changes in trends and allow warning for possible dangerous directions of trends.

5.5 Papers to translate

In the end, this notional example was developed to determine the answer to Hypothesis 5, whether we can use our meta-matrices to determine a priority of which papers to translate. Depending upon what analysts think is important, or which questions they ask, they'll use different metrics and therefore find different answers. Additionally, if we explore papers based on different structures, we find that we'd translate different papers.

In this scenario, we will assume that analysts are interested in what the foundational authors within this corpus are working on currently. Essentially, we are asking question 6 in Table 8, which authors are most likely to be writers of foundational work and then we find their works within the corpus. The metric that we will use to explore this question is eigenvector centrality. There are a number of works using eigenvector centrality (or Google's PageRank, which is a derivation of eigenvector centrality) as a measure of importance of an author.[149,150,151] In fact, it's been shown that eigenvector centrality correlates well with H index scores.[152] For more detailed information on eigenvector centrality, see Appendix E. We rank ordered the authors in an author citation network (AxA-Dir), as well as the $\mathbf{A}$, $\mathbf{A}_J$, $\mathbf{A}_K$, $\mathbf{A}_{J,K}$ matrices by eigenvector centrality.

Based on these results, and our results from viewing Turing winner results, we would recommend translating the lists from the author citation network (AxA-Dir), the **A** matrix and the $\mathbf{A}_{J,K}$ matrix. Then we identified the top 0.5% authors (or 26 authors) in each of the above meta-matrices with respect to eigenvector centrality, creating a total list of 65 foundational authors (removing repeats). For a list of these authors see Appendix I. We then identified what papers these authors wrote within our corpus. See Appendix I. From these 65 foundational authors we found that there are 137 papers that they wrote, which we would recommend translation. Our use of meta-matrices allowed us to quickly identify the 2.9% of the papers that have value to the analyst question. These papers should give analysts a reasonable idea of what the foundational authors have worked on within the corpus. This might allow researchers to know what the foremost thinkers, within this corpus of papers, are thinking and researching.

This analysis can allow researchers to quickly assess or prioritize a large corpus of information in a short amount of time. By choosing different questions, and different metrics and meta-matrices, analysts can quickly identify intelligence reports with high valuing according to the question asked.

5.6 Discussion and Analysis of Results

Validation for this kind of applied analysis can be difficult but must be attempted. In Nanda et al studied almost a million dissertations and journal articles covering 13 subject fields from the mid-1950s to 2000, and found that many fields of discipline demonstrated "delayed or poor emphasis on validated knowledge."[153] When talking about national security intelligence, although one might never know the actual truth of a situation, as sparrow suggests we must do the best we can with the information we have.[154] In general, results are valid to the extent that they actually measure what is intended to be measured.[155] In our research, we have attempted to ensure that we have measured what we've attempted to measure. We are attempting to apply the meta-matrix methodology from Carley et al. to a corpus of documents in order to arrive at a multi-dimensional view of the information contained therein.[156] We use a combination of (1) statistical analysis, (2) comparison to other measures of importance from previous work (i.e. H Index, number of citations, etc.), (3) expert elicitation, and (3) qualitative analysis of the results. In the future, if we had more resources like availability of computer science experts, time and probably research money we would attempt a deeper exploration of validation of our results.

Ideally, we might have unlimited resources to validate our analysis and results. If we did, we would perhaps follow the work of Sargent who discusses the process for validating simulation models.[157] He outlines four different ways that a simulation could be validated, 1) the developing team, if large enough could internally validate the model; 2) the model users could attempt to validate the model based on their use and understanding of what they need in a model; 3) an independent team could validate the model or 4) (and he does not recommend this approach for a number of reasons) is to develop a scoring model to measure the validation of the model. Although we are not attempting to validate a model, rather a process, these methods could still be useful. In this respect, we are currently choosing the first way to validate the model. The developing team has applied statistical, quantitative and qualitative approaches to validating this work. We could give this approach to a team of intelligence analysts and collection managers to see if they thought there was value in this approach. This might be a good way, but at this time, the authors cannot mandate the adoption of our methodology for intelligence professionals. An independent team of validators might be preferred as well. Finally, we could apply some scoring methodology comparing our results to existing measures of author quality (H Index) or paper quality (citation count).

For an independent validation effort we might collect a number of recognized experts in the field of computer science and computer science literature, and first give them the actual corpus of documents represented within our data set. We could ask them to familiarize themselves with the corpus and return with their recommendations of the foundational authors within this corpus. We could then ask them to view the list our method created and compare and contrast their list to ours. After this, we might again ask them for a revised list of authors. At the end of this exercise, we could compare the two lists from the experts to our list and see how many of the authors from our list made it into either of the two expert lists. We might also see if there was a difference in inclusions between the two lists by the experts.

Another way we might independently validate our findings might be if we had compare results in each of our hypotheses to previously peer reviewed results using other methodologies. We might demonstrate that we could come up with similar results faster than previous methods. In this case, though there are a number of applications of networks and network measures in demonstrating "value" of documents, or journals in this case, our research, using meta-matrices, brings a more 3 dimensional approach to viewing a corpus of documents. To the knowledge of the authors this has not been done before quantitatively. Current methods of evaluating authors (H-Index) or journals (citations), although currently regarded as the standard for academic evaluation, tend to be a one-dimensional metric that might not capture a more holistic view of the author or journal. Though we use H-Index and citation counts in demonstrating the high quality of node selection, we also recognize that node selection does not always demonstrate the highest values of either of these measures. Additionally, H Index score distributions tend to vary widely between different disciplines. The underlying cause is that different disciplines have different average citations counts.[158] What might be a high H score for a paper in the Mathematics field, would be very low in the Biology & Biochemistry field.

Finally, we could attempt to apply some scoring methodology to validate our methodology. Given the above note about H index values and citation counts, if we had a reliable means of calculating either (although the ISI Web of Science tends to be the gold standard in calculating H Index scores and citation counts, there is work noting it's deficiencies,[159] additionally each author must be measured individually, which can take a lot of time for our 5317 authors and 4793 papers), we might compare our results in a bootstrapped distribution. We could show the average H index scores or citation counts of the collection of authors or papers we've selected, and run a bootstrap collecting a similar sample a large number of times, and calculate the average H index for those randomly selected samples. We might then compare our results to the distribution and calculate the probability of selecting our average value of metrics randomly or calculate the p-value. This could give us some measure of the statistical significance of our results. Again, though, this would be a comparison to a one dimensional statistic. We would be comparing apples to oranges. While H Index scores relate *n* number of papers with *n* citation, the implied connotation is that those authors with high H Index scores are high quality researchers. This may be the case, but there are certainly other aspects of good researchers that are not addressed in this H Index score. These are the things we are attempting to quantify using our meta-matrix approach to viewing a corpus of data. Perhaps there are important nodes or links when you view authors or papers in terms of the connections to key words or to journals. Surely an argument can be made that a good quality of a researcher is that they research and advance the state of the art on popular topics. Another possible quality might be the connections gained through consistently publishing one's work in high quality, and the possible connections to other authors or other research one can make through knowing the same people who habitually write for high quality journals or present at the high quality conferences.

5.7 Discussion of results

Given the above discussion on possible methods of validation if we had unlimited resources, we did explore methods of validation. The validation methods for our analysis are given in the details below.

5.7.1 Hypothesis 1 Discussion

In Hypothesis 1, we demonstrate the models ability to use the four network measures to see if there is a statistical difference between Turning Award winners and non-winners. We use

	Centrality Betweenness	Centrality Eigenvector	Simmelian Ties	Centrality Total Degree
AxA(Dir)	0.125	0.017	0.722	0.0004
AxA	0.392	0.0000007	0.276	0.066
AxA(K)	0.39	0.192	0.109	0.177
AxA(J)	0.1	0.96	0.003	0.002
AxA(J,K)	0.177	0.043	0.197	0.048
		Stat Sig p value <= 0.05		
		Stat Sig p value <= 0.1		

Table 13 P values of average differences between network metrics of Turing Award winners and non-winners

a p-values found in Table 13 below to demonstrate statistical significance between averages of winners and non-winners. This hypothesis uses statistical analysis to validate the differences between Turing winners and non-winners. These p values determine that some of the meta-matrices besides the standard author citation network do a statistically significant job of separating the averages of the metrics shown.

5.7.2 Hypothesis 2 Discussion

In Hypothesis 2, we explored the top ranked authors within our corpus. This is different from foundational authors. When we mention top ranked, this is because they rose to the top 10 in terms of a number of different metrics which take into account more than just importance. For a total list of the metrics involved see Appendix H. In an effort to prove that our methodology was better than randomly selecting authors, we used the hypergeometric distribution to estimate the probability we could take a random sample and capture at least one of the Turing award winners. We used this distribution because it met each of the three assumptions from Devore, 1) set to be sampled consists of N individuals (finite population), 2) each individual can be characterized as a success or a failure and there are M successes in the population, and 3) a sample of n individuals is selected without replacement I such a way that each subset of size n is equally likely to be chosen. [160] In this distribution there N=5317 total authors, there are M=20 successes or Turning Award winners and the sample of n=10 selections from each matrix completes the parameters for this distribution. This distribution models the selection of one sample of 10 authors. The probability of randomly picking 1 or more Turing award winners from one matrix is 0.037 or 3.7%. Since we draw from these n samples from 5 matrices, there is more to the overall probability. In this case, Liskov was found by sampling from 2 matrices ($\mathbf{A}_k$ and $\mathbf{A}_{J,K}$). The probability of randomly picking 1 or more Turing award winners from exactly two matrices is .00137 or 0.13%. Being more conservative, the probability of randomly picking a Turing award winner from 2 or more matrices is .0014 or 0.14%. Bottom line, the probability that a similar to the one above randomly picks at least one turning award winner by chance is close to zero. Our method is much better than a random selection of authors from a list a picking at least Turing award winners. Given that Turing award winners are a good indicator of quality of authors.

Additionally, we found the average H Index of the selected authors are 34.2. When you peel back the onion on the H Index scores, the story is a little more interesting. Below in Table 14, the H Index scores for each of the 9 authors is shown. While the average score of 34.2 as an H index score is impressive, an average professor successfully gaining tenure is around 12, you see that there is a disparity in H Index scoring. Some of the authors, predictably, have a high H Index, but some have pretty low scores. As we've mentioned before, the H Index is, by definition, a very one dimensional metric. We might hope that our methodology would not just return only those with a high H Index score. In this case, we see that authors Duinmaijer, Welbers and Pelgrom are all listed and have all been chosen by being in the top of the metrics in the $\mathbf{A}$ and $\mathbf{A}_J$ matrices. While none of the three by themselves are uniquely distinguishing, together they co-authored a paper, Matching Properties of MOS Transistors in the IEEE Journal of Solid State Circuits back in 1989. This paper has been cited over 2170 times according to Google Scholar. These authors were likely chosen in the A network because they co-wrote a very well cited paper together, and were likely chosen in the AJ network because this highly cited paper was published in a very respected journal. This paper is among the highest cited papers within our network, but was not chosen in the author citation network because there were other authors, whose work over multiple papers set them farther apart than these authors.

The other outlier in terms of H Index score seems to be Murdoch Gabbay. He was among the top ten in metrics over the $\mathbf{A}_K$ and the $\mathbf{A}_{J,K}$ matrices. Although, he has published over 61 journal articles since 1999 and had over 20 co-authors, it seems that he stands out for a different reason. He co-authored a paper, "A new approach to abstract syntax involving binders variable binding" and "A new approach to abstract syntax with variable binding," published in the IEEE Computer Society Press in 1999, and in the Formal Aspects of Computing journal in 2002, respectively. These papers combined were cited 683 times according to Google Scholar.

Author	H Index
MichaelBurrows	29*
MichaelIJordan	99
MartnAbadi	75
AadCJDuinmaijer	1**
AndantonPGWelbers	1**
MarcelJMPelgrom	6**
BarbaraHLiskov	43
MurdochJGabbay	9**
AapoHyvrinen	45
Avg H-Index	34.2

H Index References
Palsberg, Jens , Dept Chair UCLA Computer Science Dept, *The H Index for Computer Science* website. Accessed 17 April 2013. http://www.cs.ucla.edu/~palsberg/h-number.html
*Arnetminer.org, Search and Mining of Academic Social Networks website. Accessed 17 April 2013. http://arnetminer.org/person-ranklist/hindex/151-0.html
***ISI Web of Knowledge* website- Manual Caluclation by author. Accessed 17 April 2013. http://apps.webofknowledge.com/UA_GeneralSearch_input.do?product=UA&search_mode=GeneralSearch&SID=1D55CobbdG3eMK45MgH&preferencesSaved=

Table 14 Listing of H Index scores for selected authors

This paper describes a "semantic basis of meta-logics for specifying and reasoning about formal systems involving name binding, α-conversion, capture avoiding substitution, and so on," specifically applying to Fraenkel-Mostowski (FM) sets within Set Theory.[161] It's basically a popular naming convention used within these FM-sets, which have been adopted and employed by future researchers of FM-sets. In essence, Gabbay is an above average researcher and author, but also developed a popular naming structure within the computer science field. As far as getting into the top of the charts of the AK matrix, we're not entirely sure why this happened. We looked at the total importance of the key words used in the 2002 paper (there were no key words used in the 1999 paper) and found nothing notable. Exploring both eigenvector centrality and total degree centrality (E.C; D.C) the ranks for each of the given key words are shown: abstract syntax (1392/1837; 1308/3092), Alpha-conversion (1837/1837; 2267/3092), permutation actions (1837/1837; 2267/3092), Set theory (1837/1837; 2267/3092), and structural induction (1837/1837; 2267/3092). Each of the key words ranks near the end of the list of key words in terms of either eigenvector centrality (popular words connected to other popular words) or degree centrality (simply popularity of the node).

A hypothesis on why the methodology selected Gabbay is that he might be a combination Connector and Maven from Gladwell's Tipping Point.[162] In Tipping Point, Gladwell points out three types of people who help to create tipping points, or "the moment of critical mass, the threshold, the boiling point:" the Connector, the Maven and the Salesman.[163] The Connector is a person who collects acquaintances and habitually connects people based on some shared interest or drive. Mavens are people who collect information about various things and are hyperaware of what the information means and are sensitive to any changes in the information. They are problem solvers by digesting information and sharing this information with others. Salesman are, as the name states, charismatic persuaders with powerful negotiation skills. Together these personalities tend to be at the nexus of tipping points. We suggest that Gadday might be some combination of Connector, based on his many journal articles and large number of co-authors, as well as possibly a Maven, as shown in the variety of topics his papers have spanned as well as the success his one paper about FM-sets has enjoyed.

5.7.3 Hypothesis 3 Discussion

In our analysis of Hypothesis 3, we show the visual effects of using a meta-matrix approach as opposed to a standard one dimensional analysis of an author citation network. We demonstrate the visual effects of analyzing authors within each of the matrices and show visually why nodes might have different network metric scores based on the difference in structure of the overall networks. Although the two groups we identified had scores of almost zero for network measures in the author centrality network, the scores for some of the authors within the groups shot up to above average values in the meta-networks. For validation we show that these authors which were marginalized in the author citation network, but shown to have more connectivity in other meta-networks, were actually writing on things interesting to the overall community, perhaps just not with the core community of those with over 100 cites for their papers. The average citation count is 726.2 for the 6 papers written by the authors of group one and group two.

5.7.4 Hypothesis 4 Discussion

Hypothesis 4 was focused on key word popularity. In order to validate our findings, we looked up our key words in the ISI Web of Knowledge. We used the search on Topic or Title for each of the searches, selected only the Science Citation Index Expanded (SCI-Expanded), and limited the date to the time ranges given. When the key word is two words, we put the key words in quotes when running the search. Below, in Table 16, is a chart of our findings. Shown in this chart, we see that all of the hits of papers increase over the given time frame. This is an independent verification that the key words were written about more during the given time frame. We've calculated the slope over the time frame and each is a positive slope. On the other hand, some of the raw numbers are not what we'd expect from a popular key words according to our methodology. On the other hand, our data set is comprised of computer science papers having been cited over 100 times, and the ISI database is much larger, containing more disciplines. What we wanted to see is that the key words had a high quantity of papers aligned with them throughout the time period, and also that the slope reflected what our analysis reflected. We wanted to see that the slope of collaborative filtering, evaluation feature selection, measurement, and text classification had positive slopes. . In every case, the slope is positive, so the key word frequency was rising, but many of the total number of papers using the keywords, specifically collaborative filtering, and text classification, are relatively low in quantity and not a strong slope.

Number of papers with same topic or title in ISI Web of Science							
	clustering	collaborative filtering	evaluation	feature selection	keywords	measurement	text classification
2000	17040	10	29900	108	189	55873	12
2001	17244	14	29912	85	176	56743	10
2002	18573	27	32152	135	218	57842	15
2003	20241	42	35089	205	307	60543	39
2004	21908	73	38163	285	417	64552	54

Slope of papers with same topic or title in ISI Web of Science							
	clustering	collaborative filtering	evaluation	feature selection	keywords	measurement	text classification
Slope	1273.3	15.4	2170.3	47.4	58.7	2115.8	11.3

Table 16 Listing of numbers of papers and slope with the given high ranking key words in the given time

In order to do an evaluation of this check, we chose some of the lowest ranked key words

Number of papers with same topic or title in ISI Web of Science					
	imagerepresentation	machinelearning	tapestry	wirelessnetwork	sparseness
2000	68	535	82	121	73
2001	68	628	76	149	79
2002	76	843	89	219	94
2003	80	1149	95	355	110
2004	91	1501	108	478	151

Slope of papers with same topic or title in ISI Web of Science					
	imagerepresentation	machinelearning	tapestry	wirelessnetwork	sparseness
Slope	5.8	245.3	7.1	92	18.7

Table 15 Listing of numbers of papers and slope with the given low ranking key words in the given time

from our methodology shown in Table 15. We entered these key words in to the ISI search query in a similar manner to compare to see if there are indeed differences in the slope. The results were somewhat disappointing. What we might have liked to see is that the key words had a much less popularity, which is true when compared to some of the key words above. Additionally, it seems as though the positive slope exists for these key words as well. If we could check the number of papers written in the SCI-Expanded data base, and do an analysis that took into account the total number of papers written per year, we might be able to verify that indeed the number of these key words was increasing.

5.7.5 Hypothesis 5 Discussion

In Hypothesis 5 we answered an analyst questions: what are the foundation authors in our corpus writing? We developed a list of foundational authors from our corpus of documents using eigenvector centrality and in the list of 65 authors, our list contained 5 of the 20 Turing award winners in our data, a good indication that we captured foundational authors. In order to evaluate our results, we decided to compare our results to a random draw of authors in a way similar to our directed draw.

We sampled from three networks (author citation network, **A**, and $\mathbf{A}_{J,K}$) based on our results from our analysis of Hypothesis 1, that these matrices did a statistically significant job sorting out the Turing award winners from non-winners. We chose the top 0.5% authors in terms of eigenvector centrality from each of these networks. In order to evaluate the same random selection, we determined the probabilities using combinations based on the hypergeometric distribution. The actual hypergeometric distribution was not applicable in this sense because we took authors from all three lists. The upper bound of the probability of randomly drawing 5 or more Turning award winners from the three meta-matrices is shown in the probability statement below:

$$P(X \geq 5) = 1 - \sum_{4}^{k=0} \sum_{4-k}^{i=0} \sum_{4-k-i}^{j=0} p(k)p(i)p(j)$$

Where the p(*k*), p(*i*), p(*j*) is the probability of picking a winner from matrix *k*, *i*, or *j*, and the distribution of picking a winner is:

$$p(x) = \binom{5297}{26-x} * \frac{\binom{20}{x}}{\binom{5317}{26}}$$

The overall probability of randomly selecting 5 or more Turing award winners from 3 separate matrices is 0.0000109 or 0.001%. We believe this means that our methodology would be better than selecting authors at random at least in terms of picking Turing award winners. We believe it is a safe assumption that selecting Turing award winners are a good indicator of picking foundational authors.

In determining validation of our methodology, we also should look at results from existing methodologies. The common practice is to use author citation networks and their metrics to analyze a corpus of documents like this. If the standard method had been used to analyze this corpus, we might have taken the top 1.5% of author nodes in eigenvector centrality from the author citation network. Instead we got about 2/3 of the list different which gives us the opportunity to view a different perspective of importance. We did a comparison of the overall list to see what the differences may have been. In Table 17 we see a list of all the author nodes we got from each of the different matrices. These author nodes were further refined to arrive at our list of 65, since some of the author nodes are different spelling of the same author. The Turing award winners are listed in bold. Additionally, there is a column of the rank of eigenvector centrality score in the author citation network. Finally, there is a color coding to represent which names came from which lists. The top of the list with no color coding represents the top 26 authors we pulled from the author citation network. We see that there is an overlap of 5 authors on the author citation network list and the $\mathbf{A}_{J,K}$ matrix. One of those overlapped authors is the Turing award winner Butler Lampson. Additionally, many of the authors in the $\mathbf{A}_{J,K}$ matrix also have a relatively high eigenvector centrality rank within the author citation network. While the author citation network performs well and gives us four of our five Turing award winners, the $\mathbf{A}_{J,K}$ matrix gives us the fifth award winner (the 39th ranked author node on

#	Node Title	AxA-Dir E.C Rank		#	Node Title	AxA-Dir E.C Rank		
1	MichaelBurrows	1		37	EdwardDLazowska	96		
2	MartnAbadi	2		38	MSatyanarayanan	130		
3	EdwardWobber	3		39	MiguelCastro	173		
4	RogerNeedham	4		40	DavidAPatterson	176		
5	**RonaldLRivest**	5		41	WillyZwaenepoel	181		
6	SallyFloyd	6		42	MauriceHerlihy	248		
7	**AdiShamir**	7		43	MichaelJWest	312		
8	**LAdleman**	8		44	RobertNSidebotham	312		
9	DahliaMalkhi	9		45	AndreSchiper	366		AxA-Dir
10	**ButlerLampson**	10		46	GregNelson	412		AxA(J,K)
11	MichaelReiter	11		47	GeorgGottlob	708		AxA
12	MartinAbadi	12		48	ChrisHanson	1395		
13	AndrewDGordon	13		49	CHanson	2114		
14	ScottShenker	14		50	ChristopherTHaynes	2114		
15	Stephen	15		51	CTHaynes	2114		
16	TKent	15		52	DanielPFriedman	2114		
17	VictorLVoydock	15		53	DHBartley	2114		
18	LorenzoAlvisi	18		54	DOxley	2114		
19	JosephYHalpern	19		55	ed	2114		
20	CARHoare	20		56	EugeneKohlbecker	2114		
21	MichaelDahlin	21		57	Gbrooks	2114		
22	JeanPhilippeMartin	22		58	GeraldJSussman	2114		
23	VanJacobson	23		59	GJRozas	2114		
24	YoramMoses	24		60	GJSussman	2114		
25	MikeBurrows	25		61	GuillermoJRozas	2114		
26	JoanFeigenbaum	26		62	GuyLSteeleJr	2114		
27	FredBSchneider	35		63	HAbelson	2114		
28	RajJain	36		64	HalAbelson	2114		
29	BoltBeranek	37		65	JonathanRees	2114		
30	**BarbaraHLiskov**	39		66	JonathanReeseditors	2114		
31	RobbertVanRenesse	41		67	KentMPitman	2114		
32	LeslieLamport	44		68	MitchellWand	2114		
33	HectorGarciaMolina	48		69	NIAdamsIV	2114		
34	AndrewSTanenbaum	52		70	RHalstead	2114		
35	KennethPBirman	73		71	RichardKelsey	2114		
36	BrianNBershad	81		72	RKentDybvig	2114		
				73	WilliamClinger	2114		

Table 17 List of 73 author foundational author nodes with corresponding E.C. rank in AxA-Dir matrix

the author citation network is the same author node as the fifth Turing award winner, we only pulled the top 0.5% or 26 from each network). The $\mathbf{A}_{J,K}$ matrix shows those people connected with the inference of journal and key words, so we might expect that those popular nodes connected to other popular nodes within this composite matrix would also perform well in the standard author citation network.

A demonstration of the worth of our methodology comes from the list of the **A** matrix, or the co-author network. The author nodes within this list rank towards the end of the list of eigenvector centrality in the author citation network. The final rank in the author citation network is 2114, so a majority of our list comes from the very bottom of eigenvector centrality in terms of the author citation network. Each of these authors came from a paper titled Revised[5] Report on the Algorithmic Language.[164] It's a paper, cited according to Google Scholar 304 times since it was published in 1998. It's a paper evaluating the progress of a computer language called Scheme, created by Steele and Sussman, two of the co-authors of this paper. This paper is one of the largest collections of co-authors on one paper in our **A** matrix. Although individually, these people may not be topping the list in the author citation network, the fact that there is such a large collection of co-authors might be useful to analysis. Considering how many authors went into this comprehensive work, it might be useful to know that there is this huge collaboration effort ongoing within the ranks of the authors. Bringing in a national security intelligence scenario, if the subject of this data was terrorists, terrorist knowledge, terrorist locations, etc., while a collection of people with these characteristics might not be at the top of a terrorist popularity list, the fact that this is one of the largest single groups working together on a single event would probably be of interest to intelligence analysts seeking to understand the capabilities of a terrorist organization. These are the types of things that our methodology can bring to the surface if applied to a corpus of documents, which might remain opaque if standard methods are used.

Another possible way of understanding why this methodology selected these authors, beyond the list of the author citation network, is that these authors are again, some mixture of Connectors, Mavens and Salesman from Gladwell's Tipping Point.[165] The authors selected by eigenvector centrality, or popular authors connected to other popular authors, with respect to the **A** matrix, or the co-author matrix are a collection of co-authors who are co-author on a lot of papers, but also are well cited (the data set only contains those papers cited over 100 times). These people might be thought of as Connectors in Gladwell's terminology. Additionally, the list of high eigenvector centrality authors from the author citation network might be thought of as some combination of Salesmen and Mavens. Clearly they have a mastery of the information enough to have written papers that a lot of others cite, but also there must be some combination of Salesmen in the papers to have persuaded so many others to cite the paper in their own work. Finally, the list in the middle, coming from the $\mathbf{A}_{J,K}$ matrix, or authors connected by journals and key words, might be thought of as the collection of authors who are writing in the popular journals and about the popular key words, or perhaps Mavens of the popular topics.

In the spirit of the discussion above about validation if we had unlimited resources, we sent our list of foundational authors to a former professor of computer science at West Point, now a government employee still working in computer and network science. We asked him his opinion on the list of authors and whether in his expert opinion thought we had captured the foundational authors. He responded that he thought we had captured some good authors. Unfortunately he did not have the time to go into detail about his answer, just that he thought that we had done a good job of identifying the foundation authors of computer science literature.

In order to test our methodology for robustness, we randomly selected 5% of the authors and deleted them and re-ran our results network methodology results. We did this on three

	Run 1	Run 2	Run 3
# missing from original list of 74	11	9	7
% Selected correctly	85%	88%	91%

Table 18 Accuracy of list of foundational authors in additional random runs (before repeated names deleted)

different runs and found the results listed in Table 18. Based on these results, we believe that the outcome is robust against the size of the network. In each of the three runs, the list of foundational authors retained over 85% of its original content. Though this methodology is of course sensitive to change in the size of the network (network methodology is essentially a measure of the structure – reducing or adding to that structure changes the overall structure and therefore, the metrics associated with that structure), we found that a reduction in 5% of the structure, still resulted in the list of foundational author nodes being over 85% accurate on average. Of course the $\mathbf{A}_{J,K}$ matrix was the most affected, since this matrix is actually a result of multiplying the $\mathbf{A}_J$ and $\mathbf{A}_K$ matrices together. We should expect that any changes to these two matrices are heightened when the two matrices are multiplied together.

5.8 Conclusion of Discussion

Validation is an important aspect of any research methodology. Being able to definitively show that a proposed methodology is better than existing models or at least is better than random selection can be very important. Following Sargent's approach to validating a simulation model, we attempted to validate our methodology within our research group, though we have discussed alternative methods for attempting validation in future research. In our research we've, when applicable, demonstrated that our methodology works better than drawing random samples. Since we're demonstrating a more rounded approach and methodology to selecting interesting nodes, comparison to existing methods of evaluation like H Index for authors and citation counts for papers, although are useful for comparison, we expect our results to be different and not replicating the information contained within the H Index or citation counts. After we demonstrated that our methodology beats random selection, we discussed how many of our results compared to some traditional analysis and also discussed why results we found might deviate from the existing methods and metrics of evaluation. Additionally we discussed how we would like to validate this work if we had unlimited resources. Future work should focus on this validation process.

5.9 Conclusion/Future Work

It's been said that analyzing intelligence is like assembling a jigsaw puzzle where analysts attempt to fill in pieces of multiple puzzles each with unknown number of pieces, sometimes containing deceiving pieces, also sometimes not containing the pieces, and what's on the face of the piece as well as its shape change in time. It's also been said that no single piece of intelligence has value, with the exception of the rare immediately actionable piece of intelligence, but rather its value comes in the overall picture that they help the analysts create. This methodology offers not only a way to value pieces of information within a corpus based on different questions, but also allows a way to further understand the picture that the corpus helps to create.

Future work might involve furthering DNA use on information networks. As in real life, some intelligence assets are more reliable than others. For this reason, DNA allows the use of probability in their links. In a similar manner, it could be argued that some sources of documents, in this case journals or conferences, are more reliable than others. It would be interesting to see an application of probability to information networks. Additionally finding a cleaner data set that might allow us to create author institution networks. Finally, this methodology has applications with all kinds of data sets, not just academic journals. For obvious reasons we applied this methodology to an open source collection of documents, but we would like to apply this methodology to a classified data set as well.

6 Conclusion/Future Work

Collection managers have a difficult task of assigning the appropriate mix of collection assets to answer the needs of our national decision maker's questions. A methodology that might assist them in real time, or while the intelligence is still in the intelligence cycle might help them better allocate these assets. Additionally, these methodologies might assist by creating objective "value" of intelligence metrics for cost/benefit analysis of collection methods in times of dwindling budgets.

Intelligence helps our nation win its wars. Civilian and military leadership depend upon the intelligence collected, processed and narrowed to meet their needs. In the non-traditional war against terrorists, there are no front lines, so knowing more about our enemies than they know about us could be one of the key elements to winning the war. Useful measures that might improve our nation's intelligence capabilities are important and useful to our nation.

Appendix A. Results of Notional Analysis

	Top 5% of Betweenness
1	100Monkeys_Fan
2	alttiii1
3	bignarstie
4	Blair_Gibbs
5	EmTalib
6	FFMonst
7	hannahchw
8	HantsPolice
9	haydenmead
10	Hmilligan1
11	hugefoodlover
12	Hulkhogan
13	iamwill
14	kenworthy39
15	laurevans311
16	MasherMiles
17	missgucci13
18	MrCliveC
19	multizonecraig
20	MyrtleTakesTea
21	nickkeane
22	OliverPhelps
23	PCStanleyWMP
24	PoliceFedChair
25	PolicingToday
26	ravensrod
27	RBKC_Markets
28	rioferdy5
29	Riotcleanup
30	rob2d2
31	sampepper
32	SmithySmite
33	spookybizzle
34	stevegarfield
35	teanamu
36	thefadotcom
37	VC_UEL
38	WestYorksPolice
39	YoungRv

Table 19 List of highest betweenness users

This list is a list of the top 5% of users in betweenness value within the tweet network. This is the list that, depending upon the existence of tweets within the network might change from 16 (41%) to no change.

Appendix B. Interaction meta-matrix mathematical definitions

Notations

A single matrix is notated with capitalized and bold letters. We use the following abbreviations to refer to different node classes:

A	Author
P	Paper
K	Key Word
J	Journal Title
I	Author Institution
D	Publication Date

R	# of researcher
T	# of titles/papers
W	# of key words
JO	# of Journal Titles/IDs
I	# of author institutions/locations
Y	# of dates/years

Table 20 Legend of notations

Additionally, we use the following matrix notations when describing a measure:

A	Capitalized letters represent on mode (square) network matrices, e.g. the interconnections between *Authors*
A^T	The transpose of a matrix, i.e. rows and columns are swapped
A_{ij}	The entry in the *i*th row and the *j*th column of the matrix A
AK	Two mode network matirx, e.g. the connections between *Author* and *Key Words*
A_K	Inference network, e.g. the resultant matrix when $AK \bullet AK^T$

Table 21 Matrix calculation notation

One Node Graphs

A (RxR): Co-Author Network (*which authors wrote papers with which other authors?*) where $A_{ij} = n$ iff author *i* is a co-author with author *j* in *n* published works; else $A_{ij} = 0$. Nondirected, symmetric matrix with weights.

P (TxT): Citation Network (*which papers cite which papers?*) where $P_{ij} = 1$ iff paper *i* cites paper *j*; else $P_{ij} = 0$. Directed matrix.

K (WxW): Information Network (*which key words inform on which key words?*) where $K_{ij} = 1$ iff key word i is on same paper as key word j; else $K_{ij} = 0$. Nondirected, symmetric matrix.

J (JOxJO): Citation Network (*which journals cite which journals?*) where $J_{ij} = n$ iff paper in journal *i* cites paper in journal *j*, *n* times; else $J_{ij} = 0$. Directed matrix with weights.

I (LxL): Inter-organizational Network (*which organizations work with which organizations?*) where $I_{ij} = n$ iff author at institution *i* publishes a paper with author at institution *j* , *n* times; else $I_{ij} = 0$. Nondirected, symmetric matrix with weights.

D (YxY): Timeline Network (*what publication dates interact with publication dates?*) where $D_{ij} = 1$ iff date i cites date j; else $D_{ij} = 0$. Nondirected, symmetric matrix.

Two Node Graphs

PA (TxR): Author Network (*Which papers were written by which authors authors cite which papers?*) where $PA_{ij} = 1$ iff paper *i* was written by author *j*; else $PA_{ij} = 0$. Nondirected matrix.

PK (TxW): Contextual Network (*Which papers contain which key word(s)?*) where $PK_{ij} = 1$ iff paper *i* contains key word *j*; else $PK_{ij} = 0$. Nondirected matrix.

PJ (TxJO): Journal Content Network (*Which papers are from which journals?*) where $PJ_{ij} = 1$ iff paper *i* is published in journal *j*; else $PJ_{ij} = 0$. Nondirected matrix.

PI (TxL): Institutional Capabilities Network (*What titles are where?*) where $PI_{ij} = 1$ iff paper *i* is written from an author from institution *j*; else $PI_{ij} = 0$. Nondirected matrix.

PD (TxY): Publishing Precedence Network (*Which papers have been published when?*) where $PD_{ij} = 1$ iff paper *i* was published in year *j*; else $PD_{ij} = 0$. Nondirected matrix.

AK (RxW): Knowledge Network (*Who is writing about what topics?*) where $AK_{ij} = n$ iff author *i* publishes using key word *j*, *n* times; else $AK_{ij} = 0$. Nondirected matrix with weights.

AJ (RxJO): Literature Network (*What authors write for which journals?*) where $AJ_{ij} = 1$ iff author *i* publishes with journal *j*; else $AJ_{ij} = 0$. Nondirected matrix.

AI (RxL): Employment Network (*what authors work where?*) where $AI_{ij} = n$ iff author *i* publishes *n* paper(s) while at institution *j*; else $AI_{ij} = 0$. Nondirected matrix with weights.

AD (RxY): Author Timeline Network (*Which authors have published when?*) where $AD_{ij} = n$ iff author *i* publishes *n* papers during year *j*; else $AD_{ij} = 0$. Nondirected matrix with weights.

KJ (WxJO): Journal Context Network (*What key words describe which journals?*) where $KJ_{ij} = n$ iff key word *i* is used on *n* published papers from journal *j*; else $KJ_{ij} = 0$. Nondirected matrix with weights.

KI (WxL): Competency Capabilities Network (*What knowledge is where?*) where $KI_{ij} = n$ iff key word *i* is used on *n* published papers from institution *j*; else $KI_{ij} = 0$. Nondirected matrix with weights.

KD (WxY): Key Word Timeline Network (*What key words were published about when?*) where $KD_{ij} = n$ iff key word *i* is used on *n* published papers during year *j*; else $KD_{ij} = 0$. Nondirected matrix with weights.

JI (JOxL): Institutional Reference Network (*What journals get their papers from where?*) where $JI_{ij} = n$ iff *n* paper(s) in journal *i* is written by an author from institution *j*; else $JI_{ij} = 0$. Nondirected matrix with weights.

JD (JOxY): Journal Timeline Network (*What publication dates are journals published during?*) where $JD_{ij} = n$ iff *n* paper(s) *i* are published during year *j*; else $JD_{ij} = 0$. Nondirected matrix with weights.

ID (LxY): Institutional Timeline Network (*What publication dates are organizations publishing during?*) where $ID_{ij} = n$ iff *n* author(s) at institution *i* publishes during year *j*; else $ID_{ij} = 0$. Nondirected matrix with weights.

Appendix C. Reachability matrices mathematical definitions

$\mathbf{A}^n$: $(\mathbf{A}+\mathbf{I})^n$, where **I** is the identity matrix, matrix (order RxR) – Citation Network (*which authors co-author with which other authors?*) Given a graph of **A**, using Boolean operators {0 representing no edge, 1 representing an edge} instead of weights, create matrix $\mathbf{A}^n$ where $A_{ij} = 1$ iff author *i* is *n* or fewer steps away from author *j* in a citation network; else $A_{ij} = 0$. $\mathbf{A}^R$ is the fully connected matrix, also known as the universal matrix, matrix known to be strongly connected. Nondirected matrix.

$\mathbf{P}^n$: $(\mathbf{P}+\mathbf{I})^n$, where **I** is the identity matrix, matrix (order TxT) – Citation Network (*which papers cite which papers?*) Given a graph of **P**, using Boolean operators {0 representing no edge, 1 representing an edge} instead of weights, create matrix $\mathbf{P}^n$ where $P_{ij} = 1$ iff paper *i* is *n* or fewer steps away from paper *j* in a citation network; else $P_{ij} = 0$. Symmetrize the result. $\mathbf{P}^T$ (order T, not transpose) is the fully connected matrix, also known as the universal matrix, matrix known to be strongly connected. Nondirected matrix.

$\mathbf{K}^n$: $(\mathbf{K}+\mathbf{I})^n$, where **I** is the identity matrix, matrix (order WxW) – Information Network (*which key words inform on which key words?*) Given a graph of $\mathbf{K}^n$, using Boolean operators {0 representing no edge, 1 representing an edge} instead of weights, create matrix $\mathbf{K}^n$ where $K_{ij} = 1$ iff key word *i* is *n* or fewer steps away from key word *j* in a key word network; else $K_{ij} = 0$. $\mathbf{K}^W$ is the fully connected matrix, also known as the universal matrix, matrix known to be strongly connected. Nondirected matrix.

$\mathbf{J}^n$: $(\mathbf{J}+\mathbf{I})^n$, where **I** is the identity matrix, matrix (order JOxJO) – Citation Network (*which journals cite which journals?*) Given a graph of **J**, using Boolean operators {0 representing no edge, 1 representing an edge} instead of weights, create matrix $\mathbf{J}^n$ where $J_{ij} = 1$ iff journal *i* is *n* or fewer steps away from journal *j* in a journal citation network; else $J_{ij} = 0$. $\mathbf{J}^{JO}$ is the fully connected matrix, also known as the universal matrix, matrix known to be strongly connected. Nondirected matrix.

$\mathbf{I}^n$: $(\mathbf{I}+\mathbf{I'})^n$, where **I'** is the identity matrix, matrix (order LxL) – Inter-organizational Network (*which organizations work with which organizations?*) Given a graph of **I**, using Boolean operators {0 representing no edge, 1 representing an edge} instead of weights, create matrix $\mathbf{I}^n$ where $I_{ij} = 1$ iff institution *i* is *n* or fewer steps away from institution *j* in a institution reference network; else $I_{ij} = 0$. $\mathbf{I}^L$ is the fully connected matrix, also known as the universal matrix, matrix known to be strongly connected. Nondirected matrix.

$\mathbf{D}^n$: $(\mathbf{D}+\mathbf{I})^n$, where **I** is the identity matrix, matrix (order YxY) – Timeline Network (*what publication dates interact with publication dates?*) Given a graph of **D**, using Boolean operators {0 representing no edge, 1 representing an edge} instead of weights, create matrix $\mathbf{D}^n$ where $D_{ij} = 1$ iff date *i* is *n* or fewer steps away from date *j* in a matrix timeline network; else $D_{ij} = 0$. $\mathbf{D}^Y$ is the fully connected matrix, also known as the universal matrix, matrix known to be strongly connected. Nondirected matrix.

Appendix D. Inference matrices mathematical definitions

A$_K$ (RxR): Citation Network inference with key words (*which authors publish using the same key words as which other authors?)* (**AK**) x (**AK**)^T matrix; Where $A_K = \sum_{k=1}^{R} A_{K_{ik}} A^T_{K_{kj}}$. Nondirected, symmetric matrix with weights.

A$_J$ (RxR): Citation Network inference with journals (*which authors publish in the same journals as which other authors?*) (**AJ**) x (**AJ**)^T matrix; Where $A_J = \sum_{k=1}^{R} A_{J_{ik}} A^T_{J_{kj}}$ Nondirected, symmetric matrix with weights.

A$_P$ (RxR): Citation Network inference with papers (*which authors cite which authors?*) (**AP**) x ((**P**) x (**PA**)); Where $A_P = \sum_{k=1}^{R} A_{P_{ik}} A^T_{P_{kj}}$. Directed matrix with weights.

A$_{J,K}$ (RxR): Citation Network inference with journals (*which authors publish in the same journals as which other authors?*) ((**AJ**) x (**AJ**)^T) x ((**AK**) x (**AK**)^T); Where $A_{J,K} = \sum_{k=1}^{R} A_{J_{ik}} A_{K_{kj}}$. Nondirected, symmetric matrix with weights.

K$_P$ (WxW): Key Word Network inference with papers (*which key words are noted on the same paper as which other key words?*) (**KP**) x (**KP**)^T; Where $K_P = \sum_{k=1}^{W} K_{P_{ik}} K^T_{P_{kj}}$. Nondirected, symmetric matrix with weights.

Appendix E. Definitions for network metrics used

Betweenness Centrality[166]

Unscaled Betweenness Centrality in symmetric networks

$$C_B(x) = \sum_{i=1}^{|J|} \sum_{j=i+1}^{|J|} \frac{g_{i,j}(x)}{g_{i,j}}$$

Unscaled Betweenness Centrality in asymmetric networks

$$C_B(x) = \sum_{i=1}^{|J|} \sum_{j \neq i}^{|J|} \frac{g_{i,j}(x)}{g_{i,j}}$$

Scaled Betweenness Centrality

$$C_B'(x) = \frac{C_B(x)}{C_B^{max}}$$

Where $g_{i,j}$ is the number of shortest paths between two nodes *i* and *j*, while $g_{i,j}(x)$ is the number of shortest paths including node *x*.

$$C_B^{max} = \begin{cases} \dfrac{|J|^2 - 3|J| + 2}{2} & \text{symmetric networks} \\ |J|^2 - 3|J| + 2 & \text{asymmetric networks} \end{cases}$$

Unscaled betweenness is the summation of the number of times a node lies on the shortest path between two nodes. Nodes high in betweenness have the potential to act as a gatekeeper between groups or control the flow of information over a network.

Eigenvector Centrality[167]

The Eigenvector centrality of a node *x*, $C_E(x)$ is defined as the linear combination of the eigenvector centrality of its neighbors *i* and *j*:

$$C_E(x) = \frac{1}{\lambda} \sum_{j=1}^{|J|} x_{ij} C_E(X)$$

Where λ is a constant. We can rewrite this equation as:

$$\lambda C_E = W \cdot C_E$$

In equation #, C_E is an eigenvector and W is the network matrix. For calculating eigenvector centrality lambda is the largest eigenvalue of the adjacency matrix W and CE is the corresponding eigenvector. Note that W is always symmetrized before computing the measure which guarantees real (rather than complex) valued eigenvalues.

Scaled Eigenvector Centrality, $C'_E(x)$

$C'_E(x) = \frac{C_E(x)}{C_E^{max}}$ with $C_E^{max} = \sqrt{0.5}$

Independently from the network size, the maximum value is always $\sqrt{0.5}$. Consequently, we use this value to scale the unscaled values of eigenvector centrality.

Nodes high in eigenvector centrality are considered popular nodes connected to other popular nodes.

Simmelian Ties[168]
Given an adjacency matrix **G**, within this matrix **G** the number of cliques of size ≥ 3 are recorded and membership assigned. S_i is the number of entities that are in a clique with entity *i*. This measure is then normalized over the network.
A measure that reflects the number of informal ties within cliques. The more simmelian ties, the more social norms one has to conform to, and theoretically, the more constrained their actions are because of being constrained by the large number of norms. Additionally, this measure is an indication of the number of cliques this node has access to which can indicate importance. This measure is then normalized by the group size to make the measure one between zero and one.

Total Degree Centrality[169]

Unscaled Degree Centrality in symmetric networks $C_D(x)$

$$C_D(x) = \begin{cases} \sum_{j=1, j\geq i}^{|N|} w_{j,i} & \text{allow self loops} \\ \sum_{j=1, j>i}^{|N|} w_{j,\mathrm{i}} & \text{ignore self loops} \end{cases}$$

Unscaled Degree Centrality in asymmetric networks $C_{D\_total}(x)$

$$C_{D\_total}(x) = \begin{cases} \sum_{j=1, j\neq i}^{|N|} w_{j,i} + \sum_{j=1, j\neq i}^{|N|} w_{i,j} & \text{allow self loops} \\ \sum_{j=1, j\neq i}^{|N|} w_{j,i} + \sum_{j=1, j\neq i}^{|N|} w_{i,j} & \text{ignore self loops} \end{cases}$$

Scaled Degree Centrality C'_D

$$C'_D(x) = \frac{C_D(x)}{C_D^{max}}$$

The unscaled centrality of node *x* counts the absolute number of neighbors in the unweighted case, or sums up the line weights connected to every node, where $|N|$ is the network size and *w* is the weight of links between nodes i and j. This is one of the most basic measures of centrality. This is a measure of the popularity of the node; the more links or weighted links the node has, the more popular they are.

Appendix F. Results for statistical analysis of Turing award winners vs. non-winners

			Centrality Betweenness	Centrality Eigenvector	Simmelian Ties	Centrality Total Degree
AxA(Dir)	Award Winners	Average	0.00058	0.06886	0.00008	0.00015
		Std Dev	0.00133	0.11665	0.00020	0.00013
	Non-winners	Average	0.00018	0.00253	0.00006	0.00003
		Std Dev	0.00119	0.01735	0.00024	0.00007
AxA	Award Winners	Average	0.00010	0.00000	0.00110	0.00014
		Std Dev	0.00026	0.00000	0.00101	0.00015
	Non-winners	Average	0.00001	0.00132	0.00073	0.00007
		Std Dev	0.00007	0.01936	0.00090	0.00011
AxA(K)	Award Winners	Average	0.00023	0.04404	0.02747	0.00079
		Std Dev	0.00056	0.07056	0.03158	0.00116
	Non-winners	Average	0.00003	0.00481	0.00503	0.00012
		Std Dev	0.00015	0.01859	0.01370	0.00034
AxA(J)	Award Winners	Average	0.00213	0.00559	0.03331	0.00381
		Std Dev	0.00245	0.01245	0.01338	0.00151
	Non-winners	Average	0.00033	0.00594	0.01589	0.00180
		Std Dev	0.00164	0.01847	0.01555	0.00176
AxA(J,K)	Award Winners	Average	0.00064	0.05334	0.10740	0.02646
		Std Dev	0.00080	0.04592	0.12011	0.02271
	Non-winners	Average	0.00018	0.00898	0.04154	0.00525
		Std Dev	0.00052	0.01702	0.07825	0.00915

Table 23 Turing award winner averages and standard deviations

	Centrality Betweenness	Centrality Eigenvector	Simmelian Ties	Centrality Total Degree
AxA(Dir)	0.125	0.017	0.722	0.0004
AxA	0.392	0.0000007	0.276	0.066
AxA(K)	0.39	0.192	0.109	0.177
AxA(J)	0.1	0.96	0.003	0.002
AxA(J,K)	0.177	0.043	0.197	0.048

(green)	Stat Sig p value <= 0.05
(yellow)	Stat Sig p value <= 0.1

Table 22 P values of differences in means between Turning award winners and non-winners

In Table 23, we see the averages of Turing award winners across the 4 selected network metrics. Interesting to note, that in each of these metrics, award winners have a higher average score than non-award winners, with the exception, AxA –Eigenvector Centrality, being when the average for award winners is zero. This suggests that larger values in these metrics indicate a tendency towards Turing award winners, and while this may not necessarily help us predict award winners, it is certainly suggestive of a direction that indicates more respected authors.

Table 22 displays the p-value significance in the differences between the means of award winners and non-winners within the various metrics. Of note, analysts interested in the Betweenness and Simmelian Ties metrics, we might suggest using the A_J matrix to determine worth for authors because of significant p values between award winners and non-award winners in those two matrices. Additionally in Eigenvector Centrality, we would recommend using AxA matrix due to the reduction in p value of significance.

Appendix G. Comparison of group 1 and group 2 network metrics

AxA(Dir)	Node Title	Centrality Betweenness	Centrality Eigenvector	Simmelian Ties	Centrality Total Degree
	JensKrger	0	0	0	0
	RdigerWestermann	0	0	0	0
	EitanGrinspun	0	0	0	0
	IanFarmer	0	0	0	0
	JeffBolz	0	0	0	0
	PeterSchrder	0	0	0	0
Network	Average	0.00017	0.0028	0.00006	0.000029
	Std Dev	0.001	0.0192	0.0002	0.000066

AxA	Node Title	Centrality Betweenness	Centrality Eigenvector	Simmelian Ties	Centrality Total Degree
	JensKrger	0	0	0	0
	RdigerWestermann	0	0	0	0
	EitanGrinspun	0	0	0.0006	0.0001
	IanFarmer	0	0	0.0006	0.0001
	JeffBolz	0	0	0.0006	0.0001
	PeterSchrder	0	0	0.0006	0.0001
Network	Average	0.000008	0.0013	0.0007	0.000066
	Std Dev	0.00007	0.019	0.0009	0.00011

AxA(K)	Node Title	Centrality Betweenness	Centrality Eigenvector	Simmelian Ties	Centrality Total Degree
	JensKrger	0	0	0.0009	0
	RdigerWestermann	0	0	0.0009	0
	EitanGrinspun	0.0002	0.0043	0.0278	0.0006
	IanFarmer	0.0002	0.0043	0.0278	0.0006
	JeffBolz	0.0002	0.0043	0.0278	0.0006
	PeterSchrder	0.0002	0.0043	0.0278	0.0006
Network	Average	0.0000321	0.0049	0.005	0.0001
	Std Dev	0.00015	0.019	0.014	0.00034

AxA(J)	Node Title	Centrality Betweenness	Centrality Eigenvector	Simmelian Ties	Centrality Total Degree
	JensKrger	0	0.0012	0.0124	0.0014
	RdigerWestermann	0	0.0012	0.0124	0.0014
	EitanGrinspun	0	0.0012	0.0124	0.0014
	IanFarmer	0	0.0012	0.0124	0.0014
	JeffBolz	0	0.0012	0.0124	0.0014
	PeterSchrder	0	0.0012	0.0124	0.0014
Network	Average	0.003	0.006	0.016	0.0018
	Std Dev	0.0016	0.019	0.02	0.0017

AxA(J,K)	Node Title	Centrality Betweenness	Centrality Eigenvector	Simmelian Ties	Centrality Total Degree
	JensKrger	0	0.016	0	0.0071
	RdigerWestermann	0	0.0156	0	0.0069
	EitanGrinspun	0.0007	0.0193	0.2682	0.0175
	IanFarmer	0.0007	0.0155	0.2682	0.0164
	JeffBolz	0.0007	0.0194	0.2682	0.0175
	PeterSchrder	0.0007	0.0162	0.2682	0.0165
Network	Average	0.00018	0.009	0.042	0.0053
	Std Dev	0.0005	0.017	0.078	0.0092

Denotes metric scores above the average

Table 24 Comparison of network metrics for Group 1 among different networks

AxA(Dir)	Node Title	Centrality Betweenness	Centrality Eigenvector	Simmelian Ties	Centrality Total Degree
	BernhardSteffen	0	0	0.0004	0
	JABergstra	0	0	0	0
	JWKlop	0	0	0	0
	RobJVanGlabbeek	0	0	0.0004	0
	ScottASmolka	0	0	0.0004	0
	WPeterWeijland	0	0	0	0
Network	Average	0.00017	0.0028	0.00006	0.000029
	Std Dev	0.001	0.0192	0.0002	0.000066

AxA	Node Title	Centrality Betweenness	Centrality Eigenvector	Simmelian Ties	Centrality Total Degree
	BernhardSteffen	0	0	0.0004	0
	JABergstra	0	0	0	0
	JWKlop	0	0	0	0
	RobJVanGlabbeek	0	0	0.0004	0
	ScottASmolka	0	0	0.0004	0
	WPeterWeijland	0	0	0	0
Network	Average	0.000008	0.0013	0.0007	0.000066
	Std Dev	0.00007	0.019	0.0009	0.00011

AxA(K)	Node Title	Centrality Betweenness	Centrality Eigenvector	Simmelian Ties	Centrality Total Degree
	BernhardSteffen	0	0	0	0
	JABergstra	0	0	0	0
	JWKlop	0	0	0	0
	RobJVanGlabbeek	0	0.0028	0.0036	0.0001
	ScottASmolka	0	0	0	0
	WPeterWeijland	0	0.0028	0.0036	0.0001
Network	Average	0.0000321	0.0049	0.005	0.0001
	Std Dev	0.00015	0.019	0.014	0.00034

AxA(J)	Node Title	Centrality Betweenness	Centrality Eigenvector	Simmelian Ties	Centrality Total Degree
	BernhardSteffen	0	0.0003	0.0094	0.0011
	JABergstra	0	0	0	0
	JWKlop	0	0	0	0
	RobJVanGlabbeek	0.0007	0.0022	0.0273	0.0032
	ScottASmolka	0	0.0003	0.0094	0.0011
	WPeterWeijland	0	0.0019	0.0186	0.0021
Network	Average	0.003	0.006	0.016	0.0018
	Std Dev	0.0016	0.019	0.02	0.0017

AxA(J,K)	Node Title	Centrality Betweenness	Centrality Eigenvector	Simmelian Ties	Centrality Total Degree
	BernhardSteffen	0	0.0031	0	0.0017
	JABergstra	0	0	0	0
	JWKlop	0	0	0	0
	RobJVanGlabbeek	0.001	0.0269	0	0.0129
	ScottASmolka	0	0.0031	0	0.0017
	WPeterWeijland	0.0011	0.0286	0	0.0129
Network	Average	0.00018	0.009	0.042	0.0053
	Std Dev	0.0005	0.017	0.078	0.0092

Denotes metric scores above the average

Table 25 Comparison of network metrics for Group 2 among different networks

Appendix H. Ten important authors across the various matrices

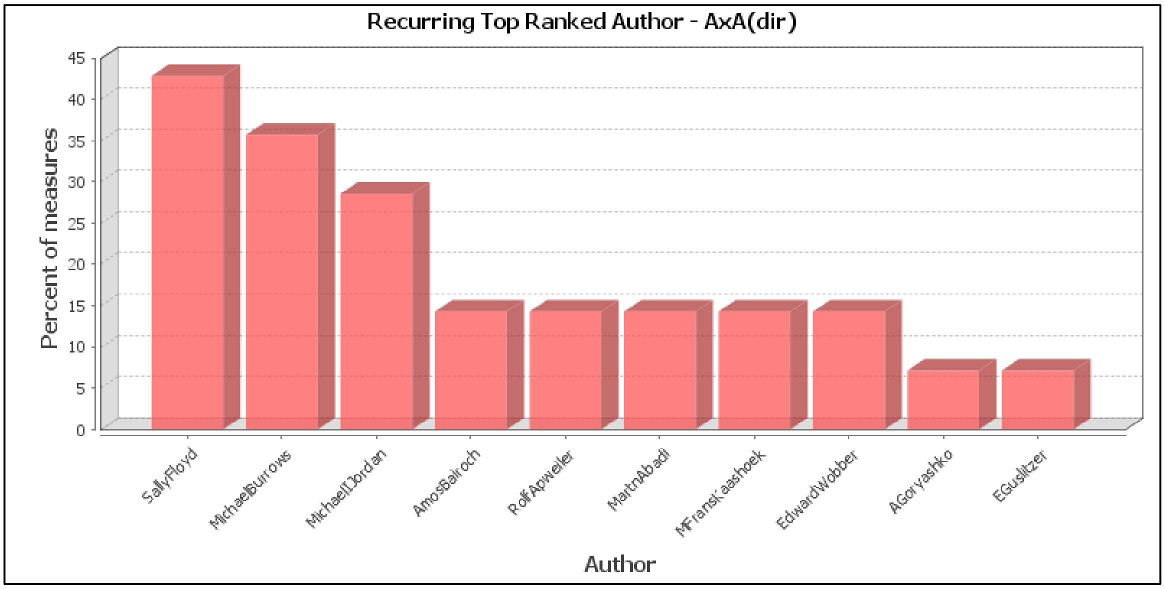

Figure 12 Ranking of top 10 authors in author citation network (AxA-Dir)

Figure 13 Ranking of top 10 authors in the A network

Figure 14 Ranking of top 10 authors in the A_J network

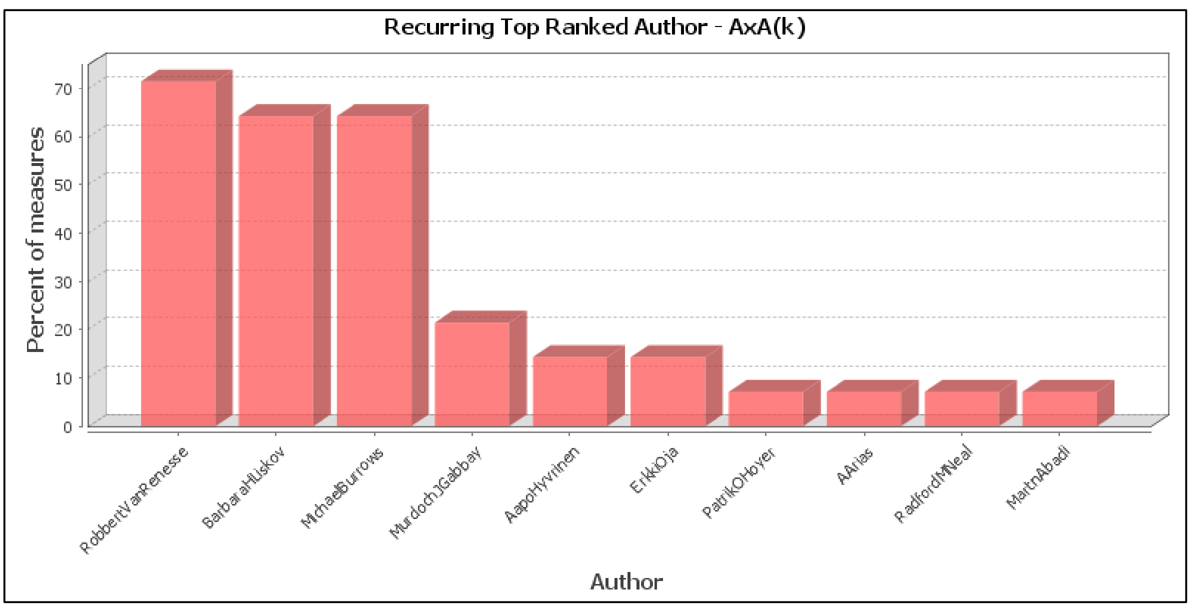

Figure 15 Ranking of top 10 authors in the A_K matrix

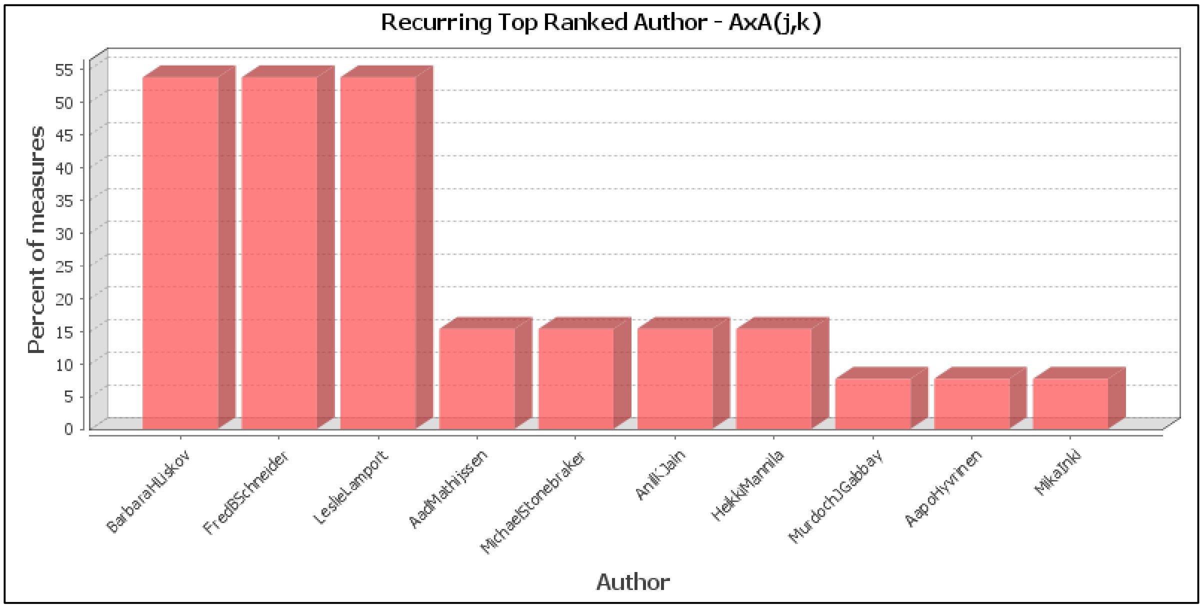

Figure 16 Ranking of top 10 authors in the $A_{J,K}$ matrix

Appendix I. Results for selection of journal articles to translate by network

#	Node Title		#	Node Title
1	AdiShamir		34	KennethPBirman
2	AndreSchiper		35	KentMPitman
3	AndrewDGordon		36	LAdleman
4	AndrewSTanenbaum		37	LeslieLamport
5	BarbaraHLiskov		38	LorenzoAlvisi
6	BoltBeranek		39	MartinAbadi
7	BrianNBershad		40	MauriceHerlihy
8	ButlerLampson		41	MichaelBurrows
9	CARHoare		42	MichaelDahlin
10	ChrisHanson		43	MichaelJWest
11	ChristopherTHaynes		44	MichaelReiter
12	DahliaMalkhi		45	MiguelCastro
13	DanielPFriedman		46	MitchellWand
14	DavidAPatterson		47	MSatyanarayanan
15	DHBartley		48	NIAdamsIV
16	DOxley		49	RajJain
17	ed		50	RHalstead
18	EdwardDLazowska		51	RichardKelsey
19	EdwardWobber		52	RKentDybvig
20	EugeneKohlbecker		53	RobbertVanRenesse
21	FredBSchneider		54	RobertNSidebotham
22	GBrooks		55	RogerNeedham
23	GeorgGottlob		56	RonaldLRivest
24	GeraldJSussman		57	SallyFloyd
25	GregNelson		58	ScottShenker
26	GuillermoJRozas		59	Stephen
27	GuyLSteeleJr		60	TKent
28	HalAbelson		61	VanJacobson
29	HectorGarciaMolina		62	VictorLVoydock
30	JeanPhilippeMartin		63	WilliamClinger
31	JoanFeigenbaum		64	WillyZwaenepoel
32	JonathanRees		65	YoramMoses
33	JosephYHalpern			

Table 26 List of 65 Foundational Authors from author citation network (AxA-Dir), A, and $A_{J,K}$ Networks

#	Author	Name	Date
1	BarbaraHLiskov	A Behavioral Notion of Subtyping	1994
2	RajJain	A binary feedback scheme for congestion avoidance in computer networks	1990
3	ButlerLampson	A Calculus for Access Control in Distributed Systems	1991
3	MartnAbadi	A Calculus for Access Control in Distributed Systems	1991
3	MichaelBurrows	A Calculus for Access Control in Distributed Systems	1991
4	AndrewDGordon	A Calculus for Cryptographic Protocols - The Spi Calculus	1998
4	MartnAbadi	A Calculus for Cryptographic Protocols - The Spi Calculus	1998
5	GeorgGottlob	A Comparison of Structural CSP Decomposition Methods	1999
6	RajJain	A Delay-Based Approach for Congestion Avoidance in Interconnected Heterogeneous Computer Networks	1989
7	RonaldLRivest	A digital signature scheme secure against adaptive chosen-message attacks	1988
8	JosephYHalpern	A Logic for Reasoning about Probabilities	1990
9	MartAbadi	A logic of authentication	1990
9	MichaelBurrows	A logic of authentication	1990
9	RogerNeedham	A logic of authentication	1990
10	BoltBeranek	A majority consensus approach to concurrency control for multiple copy databases	1979
11	AdiShamir	A Method for Obtaining Digital Signatures and Public-Key Cryptosystems	1978
11	LAdleman	A Method for Obtaining Digital Signatures and Public-Key Cryptosystems	1978
11	RonaldLRivest	A Method for Obtaining Digital Signatures and Public-Key Cryptosystems	1978
12	MauriceHerlihy	A Methodology for Implementing Highly Concurrent Data Objects	1993
13	MauriceHerlihy	A quorum-consensus replication method for abstract data types	1986
14	ScottShenker	A Simple Algorithm For Finding Frequent Elements In Streams And Bags	2003
14	LeslieLamport	A Temporal Logic of Actions	1990
15	AndrewSTanenbaum	Amoeba - A Distributed Operating System for the 1990s	1990
15	RobbertVanRenesse	Amoeba - A Distributed Operating System for the 1990s	1990
16	ChrisHanson	Amorphous Computing	1995
16	GeraldJaySussman	Amorphous Computing	1995
16	HaroldAbelson	Amorphous Computing	1995
17	JosephYHalpern	An Analysis of First-Order Logics of Probability	1990
18	VanJacobson	An Analysis of TCP Processing Overhead	1989
19	VanJacobson	An Architecture for Wide-Area Multicast Routing	1996
20	CARHoare	An axiomatic basis for computer programming	1969
21	AndrewSTanenbaum	An Efficient Reliable Broadcast Protocol	1989
22	BoltBeranek	An overview of the KL-ONE knowledge representation system	1985
23	ScottShenker	Analysis and Simulation of a Fair Queueing Algorithm	1989
24	KennethPBirman	Astrolabe: A Robust and Scalable Technology For Distributed . . .	2003
24	RobbertVanRenesse	Astrolabe: A Robust and Scalable Technology For Distributed . . .	2003
25	ButlerLampson	Authentication in distributed systems: Theory and practice	1992
25	EdwardWobber	Authentication in distributed systems: Theory and practice	1992
25	MartinAbadi	Authentication in distributed systems: Theory and practice	1992
25	MichaelBurrows	Authentication in distributed systems: Theory and practice	1992

Table 27 Papers by Foundational Authors (1-25)

#	Author	Name	Date
26	ButlerLampson	Authentication in the Taos Operating System	1994
26	EdwardWobber	Authentication in the Taos Operating System	1994
26	MartinAbadi	Authentication in the Taos Operating System	1994
27	MichaelBurrows	Autonet: A high-speed	1991
27	RogerMNeedham	Autonet: A high-speed	1991
28	MikeWest	Bayesian Density Estimation and Inference Using Mixtures	1994
29	KennethPBirman	Bimodal Multicast	1998
30	DahliaMalkhi	Byzantine Quorum System	1998
30	MichaelReiter	Byzantine Quorum System	1998
30	LorenzoAlvisi	Byzantine quorum systems	1998
30	MichaelDahlin	Byzantine quorum systems	1998
31	FredBSchneider	COCA: A Secure Distributed Online Certification Authority	2000
31	RobbertVanRenesse	COCA: A Secure Distributed Online Certification Authority	2000
32	MahadevSatyanaraya	Coda: A Highly available File System for a Distributed Workstation Environment	1990
33	CARHoare	Communicating Sequential Processes	1985
34	LeslieLamport	Composing Specifications	1993
34	MartnAbadi	Composing Specifications	1993
35	FredBSchneider	Concepts and notations for concurrent programming	1983
36	RajJain	Congestion control and traffic management in atm networks: Recent advances and a survey	1995
37	SallyFloyd	Connections with Multiple Congested Gateways in Packet-Switched Networks Part 1: One-way Traffic	1991
38	SallyFloyd	Controlling High Bandwidth Aggregates in the Network	2001
38	ScottShenker	Controlling High Bandwidth Aggregates in the Network	2001
39	MichaelKReiter	Crowds: Anonymity for Web Transactions	1997
40	HectorGarciamolina	Data Caching Issues in an Information Retrieval System	1990
41	JoanFeigenbaum	Delegation Logic: A Logic-based Approach to Distributed Authorization	2000
42	SallyFloyd	Difficulties in Simulating the Internet	2001
43	MSatyanarayanan	Disconnected operation in the Coda file system	1989
44	GeorgGottlob	Disjunctive Datalog	1997
45	WillyZwaenepoel	Distributed Process Groups in the V Kernel	1985
46	DahliaMalkhi	Dynamic Byzantine Quorum Systems	2000
46	LorenzoAlvisi	Dynamic Byzantine Quorum Systems	2000
46	MichaelKReiter	Dynamic Byzantine Quorum Systems	2000
47	SallyFloyd	Dynamics of TCP Traffic over ATM Networks	1994
48	MSatyanarayanan	Energy-aware adaptation for mobile applications	1999
49	FredBSchneider	Enforceable Security Policies	1998
50	GregNelson	Eraser: a dynamic data race detector for multithreaded programs	1997
50	MichaelBurrows	Eraser: a dynamic data race detector for multithreaded programs	1997
51	AndrewSTanenbaum	Experiences with the amoeba distributed operating system	1990
51	RobbertVanRenesse	Experiences with the amoeba distributed operating system	1990
52	SallyFloyd	Explicit Congestion Notification	1994

Table 28 Papers by Foundational Authors (26-52)

#	Author	Name	Date
53	MAbadi	Explicit substitutions	1991
53	GeorgGottlob	Extending Object-Oriented Systems with Roles	1994
54	FredBSchneider	Fail-Stop Processors: An Approach to Designing Fault-Tolerant Computing Systems	1983
55	GregNelson	Fast decision procedures based on congruence closure	1980
56	MauriceHerlihy	Fast Randomized Consensus using Shared Memory	1988
57	JosephYHalpern	From Statistical Knowledge Bases to Degrees of Belief	1996
58	ScottShenker	Fundamental Design Issues for the Future Internet	1995
59	AndrewSTanenbaum	Globe: A wide-area distributed system	1999
60	HectorGarciaMolina	GlOSS: Text-Source Discovery over the Internet	1999
61	KennethPBirman	Horus: A flexible group communication system	1996
61	RobbertVanRenesse	Horus: A flexible group communication system	1996
62	HectorGarciamolina	How to assign votes in a distributed system	1985
63	AdiShamir	How to Leak a Secret	2001
63	RonaldLRivest	How to Leak a Secret	2001
64	GeorgGottlob	Identifying the minimal transversals of a hypergraph and related problems	1995
65	FredBSchneider	Implementing Fault-Tolerant Services Using the State Machine Approach: A Tutorial	1990
66	BoltBeranek	Improving round-trip time estimates in reliable transport protocols	1987
67	JosephYHalpern	Knowledge and Common Knowledge in a Distributed Environment	1984
67	YoramMoses	Knowledge and Common Knowledge in a Distributed Environment	1984
68	RobertHHalstead	Lazy Task Creation: A Technique for Increasing the Granularity of Parallel Programs	1990
69	AndreSchiper	Light weight causal and atomic group multicast	1991
69	KennethBirman	Lightweight causal and atomic group multicast	1991
70	BrianNBershad	Lightweight remote procedure call	1990
70	EdwardDLazowska	Lightweight remote procedure call	1990
71	MauricePHerlihy	Linearizability: a correctness condition for concurrent objects	1990
72	SallyFloyd	Link-sharing and Resource Management Models for Packet Networks	1993
72	VanJacobson	Link-sharing and Resource Management Models for Packet Networks	1995
73	VanJacobson	Low-Complexity Video Coding for Receiver-Driven Layered Multicast	1997
74	HectorGarciamolina	Main memory database systems: An overview	1992
75	WillyZwaenepoel	Manetho: Transparent Rollback-Recovery with Low Overhead	1992
76	AndrewDGordon	Mobile ambients	2000
77	CARHoare	Monitors: An Operating System Structuring Concept	1974
78	RobertHHalstead	MULTILISP: a language for concurrent symbolic computation	1985
79	GeorgGottlob	On the Complexity of Propositional Knowledge Base Revision	1992
80	SallyFloyd	On Traffic Phase Effects in Packet-Switched Gateways	1992
80	VanJacobson	On Traffic Phase Effects in Packet-Switched Gateways	1992
81	AndrewSTanenbaum	Orca: A language for parallel programming of distributed systems	1992
82	RajJain	Packet Trains: Measurements and a New Model for Computer Network Traffic	1986
83	MichaelBurrows	Performance of Firefly RPC	1989
84	MSatyanarayanan	Pervasive Computing: Vision and Challenges	2001

Table 29 Papers by Foundational Authors (53-84)

#	Author	Name	Date
85	BarbaraHLiskov	Practical Byzantine fault tolerance and proactive recovery	2002
85	MiguelCastro	Practical Byzantine fault tolerance and proactive recovery	2002
86	ScottShenker	Pricing in computer networks: Motivation	1993
87	SShenker	Pricing in Computer Networks: Reshaping the Research Agenda	1995
88	MartnAbadi	Private Authentication	2002
89	DahliaMalkhi	Probabilistic Quorum Systems	1997
89	MichaelReiter	Probabilistic Quorum Systems	1997
90	AndrewSTanenbaum	Programming languages for distributed computing systems	1989
91	SallyFloyd	Promoting the Use of End-to-End Congestion Control in the Internet	1998
92	RogerMNeedham	Protecting Poorly Chosen Secrets from Guessing Attacks	1993
93	BarbaraHLiskov	Protecting privacy using the decentralized label model	2000
94	MartNAbadi	Prudent engineering practice for cryptographic protocols	1996
94	RogerNeedham	Prudent engineering practice for cryptographic protocols	1996
95	DahliaMalkhi	Quorum Systems	1999
96	DavidAPatterson	RAID: High-Performance	1994
97	SallyFloyd	Random Early Detection Gateways for Congestion Avoidance	1993
97	VanJacobson	Random Early Detection Gateways for Congestion Avoidance	1993
98	LLamport	Reaching Agreement in the Presence of Faults	1980
99	JosephYHalpern	Reasoning about Knowledge and Probability	1994
100	FredBSchneider	Recognizing Safety and Liveness	1986
101	WillyZwaenepoel	Recovery in Distributed Systems Using Optimistic Message Logging and Checkpointing	1988
102	KennethPBirman	Reliable Communication in the Presence of Failures	1987
103	FredBSchneider	Renesse. COCA: A Secure Distributed On-line Certification Authority	2002
103	RobbertVanRenesse	Renesse. COCA: A Secure Distributed On-line Certification Authority	2002
104	MauricePeterHerlihy	Replication Methods for Abstract Data Types	1984
105	CHanson	Revised 5 report on the algorithmic language Scheme	1998
105	CTHaynes	Revised 5 report on the algorithmic language Scheme	1998
105	DHBartley	Revised 5 report on the algorithmic language Scheme	1998
105	DOxley	Revised 5 report on the algorithmic language Scheme	1998
105	GBrooks	Revised 5 report on the algorithmic language Scheme	1998
105	GJRozas	Revised 5 report on the algorithmic language Scheme	1998
105	GJSussman	Revised 5 report on the algorithmic language Scheme	1998
105	HAbelson	Revised 5 report on the algorithmic language Scheme	1998
105	JonathanReeseditors	Revised 5 report on the algorithmic language Scheme	1998
105	KMPitman	Revised 5 report on the algorithmic language Scheme	1998
105	MWand	Revised 5 report on the algorithmic language Scheme	1998
105	RHalstead	Revised 5 report on the algorithmic language Scheme	1998
105	RichardKelsey	Revised 5 report on the algorithmic language Scheme	1998
105	RKDybvig	Revised 5 report on the algorithmic language Scheme	1998
105	WilliamClinger	Revised 5 report on the algorithmic language Scheme	1998
105	DanielPFriedman	Revised^5 Report on the Algorithmic Language Scheme	1998
105	ed	Revised^5 Report on the Algorithmic Language Scheme	1998
105	ed	Revised^5 Report on the Algorithmic Language Scheme	1998
105	ed	Revised^5 Report on the Algorithmic Language Scheme	1998
105	EugeneKohlbecker	Revised^5 Report on the Algorithmic Language Scheme	1998
105	GuyLSteeleJr	Revised^5 Report on the Algorithmic Language Scheme	1998
105	NIAdamsIV	Revised^5 Report on the Algorithmic Language Scheme	1998

Table 30 Papers by Foundational Authors (85-105)

#	Author	Name	Date
106	ScottShenker	RSVP: A New Resource Reservation Protocol	1993
107	FredBSchneider	SASI Enforcement of Security Policies: A Retrospective	2000
108	MichaelJWest	Scale and Performance in a Distributed File System	1988
108	MSatyanarayanan	Scale and Performance in a Distributed File System	1988
108	RobertNSidebotham	Scale and Performance in a Distributed File System	1988
109	BrianNBershad	Scheduler Activations: Effective Kernel Support for the User-Level Management of Parallelism	1992
109	EdwardDLazowska	Scheduler activations: Effective kernel support for the user-level management of parallelism	1992
110	ScottShenker	Scheduling for reduced CPU energy	1994
111	MiguelCastro	SCRIBE: A large-scale and decentralized application-level multicast infrastructure	2002
112	MartnAbadi	Secrecy by Typing in Security Protocols	1998
113	Stephen	Security mechanisms in high-level network protocols	1983
113	TKent	Security mechanisms in high-level network protocols	1983
113	VictorLVoydock	Security mechanisms in high-level network protocols	1983
114	DavidAPatterson	Serverless Network File Systems	1996
114	MichaelDDahlin	Serverless Network File Systems	1996
115	JoanFeigenbaum	Sharing the Cost of Multicast Transmissions	2001
115	ScottShenker	Sharing the Cost of Multicast Transmissions	2001
116	GregNelson	Simplification by cooperating decision procedures	1979
117	JeanPhilippeMartin	Small Byzantine Quorum Systems	2001
117	LorenzoAlvisi	Small Byzantine Quorum Systems	2001
117	MichaelDahlin	Small Byzantine Quorum Systems	2001
118	JoanFeigenbaum	Strategyproof Sharing of Submodular Costs: budget balance versus efficiency	1999
118	ScottShenker	Strategyproof Sharing of Submodular Costs: budget balance versus efficiency	1999
119	SallyFloyd	TCP and Explicit Congestion Notification	1994
120	AndrewSTanenbaum	The Amoeba Distributed Operating System	1992
121	LeslieLamport	The Byzantine Generals Problem	1982
122	GeorgGottlob	The complexity of logic-based abduction	1995
123	GeorgGottlob	The DLV System for Knowledge Representation and Reasoning	2002
124	LeslieLamport	The Existence of Refinement Mappings	1988
124	MartnAbadi	The existence of refinement mappings	1991
125	LeslieLamport	The part-time parliament	1998
126	EdwardDLazowska	The Performance Implications of Thread Management Alternatives for Shared-Memory Multiprocessors	1989
127	VanJacobson	The PIM Architecture for Wide-Area Multicast Routing	1996
128	KennethPBirman	The process group approach to reliable distributed computing	1993
129	SallyFloyd	The Synchronization of Periodic Routing Messages	1994
129	VanJacobson	The Synchronization of Periodic Routing Messages	1994
130	HectorGarciamolina	The TSIMMIS Approach to Mediation: Data Models and Languages	1997
131	AndrSchiper	Total Order Broadcast and Multicast Algorithms: Taxonomy And Survey	2003
132	SallyFloyd	Traffic Phase Effects in Packet-Switched Gateways	1992
132	VanJacobson	Traffic Phase Effects in Packet-Switched Gateways	1992
133	WillyZwaenepoel	TreadMarks: Shared Memory Computing on Networks of Workstations	1996
134	RonaldLRivest	Untraceable electronic mail	1981
135	SallyFloyd	V.: Random Early Detection gateways for Congestion Avoidance	1993
135	VanJacobson	V.: Random Early Detection gateways for Congestion Avoidance	1993
136	MauriceHerlihy	Wait-free synchronization	1991
137	SallyFloyd	Wide-Area Traffic: The Failure of Poisson Modeling	1995

Table 31 Papers by Foundational Authors (106-137)

References

[1] M. M. Lowenthal, Intelligence*: From Secrets to Policy*. 4th Ed. (Washington, DC: CQ press, 2009), 53.

[2] Mike McConnell, “Overhauling Intelligence,” *Foreign Affairs*, 86, no. 2 (2007): 53.

[3] Clark, Robert M. The Technical Collection of Intelligence. (Washington, DC.:CQ Press, 2011). Pg 261-292.

[4] Mike McConnell, “Overhauling Intelligence,” *Foreign Affairs*, 86, no. 2 (2007): 53.

[5] National Commission on Terrorist Attacks upon the United States, (Philip Zelikow, Executive Director; Bonnie D. Jenkins, Counsel; Ernest R. May, Senior Advisor), *The 9/11 Commission report*, (New York: W.W. Norton & Company, 2004), 277.

[6] Clark, Robert M. The Technical Collection of Intelligence. (Washington, DC.:CQ Press, 2011). Pg 261-292.

[7] Sims, J. “Defending adaptive realism: Intelligence theory comes of age.” In P. Gill, S. Marrin & M. Phythian (Eds.) *Intelligence Theory: Key questions and debates* (151-165). New York, NY:Routledge, 2008.

[8] Godson, Roy. Dirty Tricks or Trump Cards: U.S. Covert Action and Counterintelligence. (New Brunswick, NY: Transaction Publishers, 2001), 1-18.

[9] Warner, Michael. “Intelligence as risk shifting.” In: P. Gill, S. Marrin & M. Phythian (Eds.) *Intelligence Theory: Key questions and debates*. New York, NY:Routledge, 2008. Pg 19.

[10] Betts, Richard K. “Analysis, war, and decision: Why intelligence failures are inevitable.” In. P. Gill, S. Marrin & M. Phythian (Eds.) *Intelligence Theory: Key questions and debates*. New York, NY:Routledge, 2008. Pg 106.

[11] Gannon, J.C. Speech before World Affairs Council, 20 Mar 1996. <https://www.cia.gov/news-information/speeches-testimony/1996/ddi_speech_032096.html> last accessed 29 Sept 2011.

[12] Clapper, J. Speech before Military Operations Research Society meeting, 12 Dec 2011. MAJ Chris Smith’s personal recollection of speech.

[13] The Work of a Nation. Central Intelligence Agency. < https://www.cia.gov/library/publications/additional-publications/the-work-of-a-nation/ >, last accessed 10 Dec 2012. Pg. 27

[14]Treverton, Gregory F. and C Bryan Gabbard. Technical Report: Assessing the Tradecraft of Intelligence Analysis. Rand Corporation. Santa Monica, CA. 2008. Pg 3-4

[15] Kaplan, Edward. OR Forum – Intelligence Operations Research: The 2010 Philip McCord Morse Lecture. Operations Research Articles in Advance. Pg 1-13. <http://or.journal.informs.org/content/early/2012/07/03/opre.1120.1059.full.pdf> last accessed 10 Dec 2012.

[16] Wheaton, Kristan J., and Michael T. Beerbower. "Towards a New Definition of Intelligence." *Stan. L. & Pol'y Rev.* 17 (2006): 319.

[17] Mahnken, T. *Uncovering Ways of War: U.S. Intelligence and Foreign Military Innovation 1918-1941*. Ithaca, NY: Cornell University Press, 2002.

[18] Shannon, C.E. “A Mathematical Theory of Communication.” *Bell System Technical Journal* (University of Illinois Press, 1948) 379-423 and 623-656.

[19] Verdú, Segio. Fifty Years of Shannon Theory. IEEE Transactions on Information Theory. 44(6) 1998. Pg 2057-2078.

[20] Stigler, George J. The Economics of Information. The Journal of Political Economy Vol. 69. Vol. 3. Pg 213- 225.

[21] H.R. Varian, “Markets for Information Goods,” in conference proceedings: *Monetary Policy in a World of Knowledge-Based Growth, Quality Change, and Uncertain Measurement*, (New York, NY: Palgrave, 2000):85-99.

[22] Clemen, Robert T., Terence Reilly. Making Hard Decisions with Decision Tools 2nd Ed. (Duxbury, Pacific Grove, CA). 2001. pg. 496.

[23] Howard, Ronald A. *Decision analysis: Applied decision theory*. Stanford Research Institute, 1966.

[24] Ibid. pg. 497-498.

[25] Devore, Jay L. Probability and Statistics for Engineering and the Sciences. 7th Ed. Thomson, Belmont, CA) 2008. pg. 72 – 74.

[26] Zlotnick, Jack. “Bayes’ Theorem for Intelligence Analysis.” Studies in Intelligence Online. Paper presented at the Conference on The Diagnostic Process, Ann Arbor, Michigan. 1970. <https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/vol16no2/html/v16i2a03p_0001.htm> Last Accessed, 13 December 2012.

[27] Fisk, Charles E. “The Sino-Soviet Border Dispute: A Comparison of the Conventional and Bayesian Methods for Intelligence Warning.” Studies in Intelligence Online. Posted Online 2007. < https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/vol16no2/html/v16i2a04p_0001.htm > Last Accessed, 13 December 2012.

[28] Schweitzer, Nicholas. “Bayesian Analysis for Intelligence: Some focus on the Middle East.” Studies in Intelligence Online. Posted 2007. < https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/vol20no2/html/v20i2a03p_0001.htm > Last Accessed, 13 December 2012.

[29] Keisler, Jeffrey M. “Technical Note: Comparative static analysis of information value in a canonical decision problem.” The Engineering Economist. Vol. 49 2004. 339-349.

[30] Keisler, Jeffrey M. “Additivity of Information Value in Two-Act Linear Loss Decisions with Normal Priors.” Risk Analysis 25(2) 2005. 351-359.

[31] Keisler, Jeffrey M. and Mark Brodfuehrer. “An application of value of information to decision process reengineering.” The Engineering Economist Vol 54 2009. 197-221.

[32] Keisler, Jeffrey M. “Value of Information in Portfolio Decision Analysis.” Decision Analysis 1(3) 2004. 177-189.

[33] Clemen, Robert T., Terence Reilly. Making Hard Decisions with Decision Tools 2nd Ed. (Duxbury, Pacific Grove, CA). 2001. pg. 527.

[34] Ibid. pg. 531.

[35] Delquié, Phillippe. “The value of information and intensity of preference.” Decision Analysis. Vol 5. No.3. Sept. 2008. Pp. 129 – 139.

[36] Kaplan, Stanley and B. John Garrick. On the Quantitative Definition of Risk. Risk Analysis. Vol. 1, No. 1. 1981. Pg 12.

[37] Sage, Andrew P. *Systems Engineering*. (John Wiley and Sons, Inc. New York, NY: 1992) pg. 246.

[38] Keeney, Ralph L. Modeling values for anti-terrorism analysis. Risk Analysis. Vol 27, No. 3. 2007. Pg 585-596.

[39] Merrick, Jason and Gregory S. Parnell. A comparative analysis of PRA and Intelligent Adversary Methods for Counterterrorism Risk Management. Risk Analysis. Vol. 31, No. 9. 2011.

[40] Parnell, Gregory S., Michael Frimpon, John Barnes, Jack M. Kloeber, Jr., Richard F. Deckro, and Jack A. Jackson. Safety Risk Analysis of an Innovative Environmental Technology. Risk Analysis Vol 21. No.1. 2001.

[41] Willis, Henry H. Using Risk Analysis to Inform Intelligence Analysis – Working Paper. Rand Corporation, Santa Monica, CA. February 2007.

[42] Parnell, G. Smith, C. and Moxley, F. "Intelligent adversary risk analysis: A Bioterrorism Risk Management Model." Risk Analysis, 2010; 30(1): 32-48.

[43] Sage, Andrew P. Systems Engineering. (John Wiley and Sons, Inc. New York, NY: 1992) pg. 246.

[44] Ibid. pg. 175.

[45] Bodily, Samuel E. and Michael S. Allen. "A Dialogue Process for Choosing Value-Creating Strategies." Interfaces. Vol. 29, No. 6. Pg. 16-28.

[46] Prunckun, H. 2010. *Handbook of Scientific Methods of Inquiry for Intelligence Analysis*. Lantham, MD: The Scarecrow Press. Pg 30-31.

[47] Ibid. Pg. 29-30.

[48] Ibid. Pg. 29-30.

[49] Justice Potter Stewart, Concurring Opinion in Jacobellis v. Ohio 378 U.S. 184(1964).

[50] Robert Young, "French Military Intelligence and Nazi Germany 1938–1939" in *Knowing One's Enemies: Intelligence Assessment Before the Two World* Wars, eds. Ernest May (Princeton: Princeton University Press, 1986), 271–309

[51] Ibid. p.308.

[52] Amos Tversky and Daniel Kahneman, "Judgment under uncertainty: Heuristics and Biases." *Science,* New Series, Vol. 185, No. 4157. Sep 27, 1974. Pg. 1124-1131.

[53] Daniel Kahneman, *Thinking, Fast and Slow*. (New York, NY: Farrar, Straus and Giroux) 2011.

[54] Amos Tversky and Daniel Kahneman, "Judgment under uncertainty: Heuristics and Biases." *Science*, New Series, Vol. 185, No. 4157. Sep 27, 1974. Pg. 1128.

[55] Yaacov Vertzberger, *The World in Their Minds: Information Processing, Cognition, and Perception in Foreign Policy Decisionmaking* (Stanford: Stanford University Press, 1990), 133–144, 172–176, 190–191.

[56] M. M. Lowenthal, Intelligence*: From Secrets to Policy*. 4th Ed. (Washington, DC: CQ press, 2009), 105.

[57] Presidential Executive Order, "U.S. President. Executive Order 13526 - Classified National Security Information," *Federal Register* 75, no. 2 (2010): 707-731.

[58] Director of National Intelligence. 2008. *Intelligence Community Classification Guidance Findings and Recommendations Report*. Washington, DC: The Office of the Director of National Intelligence.

[59] Stephen Marrin, "Evaluating the Quality of Intelligence Analysis: By what (Mis) Measure?," *Intelligence and National Security* (London, UK: Routledge, 2012).pg 1-17.

[60] Ibid. pg.17.

[61] Caldwell, W.V., C.H. Coombs, R.M. Thrall, and M.S. Schoeffler. "A Model for evaluating the output of intelligence systems." Naval Research Logistics Quarterly 8(1): 25-40.

[62] Frank B. Strickland, Chris Whitlock. "Empirically Based Intelligence Management: Using Operations Research to Improve Programmatic Decision Making." *IBM Center for The Business of Government*. 2012. Accessed online : http://www.businessofgovernment.org/report/empirically-based-intelligence-management-using-operations-research-improve-programmatic-deci. Last accessed: 4 October 2012.

[63] Frank B. Strickland, Chris Whitlock. "Assessing the Value of Intelligence: Lessons for Leaders." *IBM Center for The Business of Government*. 2012. Accessed online : http://www.businessofgovernment.org/report/assessing-value-intelligence-lessons-leaders. Last accessed: 4 October 2012.

[64] Cesar, Edison, Patrick Allen, Rick Eden. "Finding a new approach to measure the operational value of intelligence for military operations." Rand Corporation, Santa Monica, CA. December 1992.

[65] Cesar, Edison, Patrick Allen, Steven Bankes, John Bondanella, Rick Eden, H. Edward Hall, Clairice Veit, Loretta Verma, Robert Weissler, Barry Wilson. "A new approach for measuring the operational value of intelligence for military operations." Rand Corporation, Santa Monica, CA. December 1994.

[66] Cesar, Edison, Patrick Allen, Rick Eden. "Finding a new approach to measure the operational value of intelligence for military operations." Rand Corporation, Santa Monica, CA. December 1992. Pg. v.

[67] Defense Science Board Advisory Group on Defense Intelligence. Operations Research for Intelligence, Surveillance and Reconnaissance (ISR). Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics. Washington, DC. January 2009.

[68] Ibid. Pg 23.

[69] Lowenthal, M. M. 2009. *Intelligence: From Secrets to Policy*. Washington, DC: CQ press. Pg. 112

[70] Ibid. pg. 183.

[71] Defense Science Board Advisory Group on Defense Intelligence. Operations Research for Intelligence, Surveillance and Reconnaissance (ISR). Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics. Washington, DC. January 2009.

[72] Brin, S., and Page, L. 1998. Anatomy of a large scale hypertextual web search engine. In Proceedings of the 7th International World Wide Web conference (Brisbane, Australia, Apr 14-18). Pg. 107-117.

[73] Radicchi F., Fortunato S., and Vespignani A. (2011) Chapter 7: Citation Networks. In: Scharnhorst A. Borner K, Van den Besselaar P (Eds) Models of science dynamics. Springer, Berlin, pp 233-257.

[74] Carley, Kathleen M. "Dynamic Network Analysis" In: Breiger, R., Carley, K., Pattison, P. (Eds), Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers. Committee on Human Factors, National Research Council, pp. 133-145.

[75] Newman, Mark EJ. "The structure and function of complex networks." *SIAM review* 45.2 (2003): 167-256. Pg 168.

[76] Newman, Mark. *Networks: an introduction*. Oxford University Press, Inc., 2010. Pg 110, 122.

[77] Borgatti, Stephen P., and Daniel S. Halgin. "On network theory." *Organization Science* 22.5 (2011): 1168-1181.

[78] Newman, Mark EJ. "The structure and function of complex networks." *SIAM review* 45.2 (2003): 167-256. Pg 168.

[79] Dorogovtsev, Sergey N., and Jose FF Mendes. "Evolution of networks." *Advances in physics* 51.4 (2002): 1079-1187. Pg 1080.

[80] Boccaletti, Stefano, et al. "Complex networks: Structure and dynamics." *Physics reports* 424.4 (2006): 175-308.

[81] Heuer, Richards J. and Randolph H. Pherson, *Structured Analytical Techniques for Intelligence Analysts* (Washington, DC: CQ Press, 2011), 68.

[82] Kevin T. Kennedy, Richard F. Deckro, James T. Moore, and Kenneth M. Hopkinson, "Nodal Interdiction," *Mathematical and Computer Modeling* 54(2011): 3116-3125.

[83] Overbey, Lucas A., Christopher Paribello, and Terresa Jackson. "Identifying Influential Twitter Users in the 2011 Egyptian Revolution." *Social Computing, Behavioral-Cultural Modeling and Prediction.* Springer Berlin Heidelberg, 2013. 377-385.

[84] Malcolm K. Sparrow, "The application of network analysis to criminal intelligence; an assessment of the prospects," *Social Networks* 13(1991): 261-263.

[85] Stephen P. Borgatti, Kathleen M. Carley, and David Krackhardt,"On the robustness of centrality measures under conditions of imperfect data," *Social Networks* 28(2006).

[86] Christian Glacet, Nicholas Hanusse, and David Ilcinkas, "The impact of edge deletion on the number of errors in networks," Lecture Notes in Computer Science: Principles of Distributed Systems 7109 (2011), 378.

[87] Knightley, Phillip, *The Second Oldest Profession: Spies and Spying in the Twentieth Century* (New York: Penguin, 1988).

[88] Gill, Peter, Stephen Marrin, and Mark Phythian, eds., *Intelligence Theory: Key Questions and Debates* (New York: Routledge, 2008).

[89] Bodily, Samuel E. and Michael S. Allen. "A Dialogue Process for Choosing Value-Creating Strategies." Interfaces. Vol. 29, No. 6. Pg 24.

[90]Ibid. Pg 16-28.

[91] Price, Derek de Solla. "A general theory of bibliometric and other cumulative advantage processes." *Journal of the American Society for Information Science* 27.5 (2007): 292-306.

[92] Albert, Réka, and Albert-László Barabási. "Statistical mechanics of complex networks." *Reviews of modern physics* 74.1 (2002): 47.

[93] Barabási, Albert-László, and Eric Bonabeau. "Scale-Free." *Scientific American* (2003).

[94] Devore, Jay L. *Probability and Statistics for Engineering and the Sciences*. 7th Ed. Duxbury Press, 2011.

[95] Kennedy, Kevin T., et al. "Nodal interdiction." *Mathematical and Computer Modelling* 54.11 (2011): 3116-3125.

[96] Glacet, Christian, Nicolas Hanusse, and David Ilcinkas. "The impact of edge deletions on the number of errors in networks." *Principles of Distributed Systems* (2011): 378-391.

[97] Borgatti, Stephen P., Kathleen M. Carley, and David Krackhardt. "On the robustness of centrality measures under conditions of imperfect data." *Social Networks* 28.2 (2006): 124-136.

[98] Sterling, Sara E. *Aggregation techniques to characterize social networks*. No. AFIT/GOR/ENS/04-12. Air Force Inst of Tech, Wright-Patterson AFB OH, School of Engineering and Management, 2004.

[99]Travis, A. National Archives reveals historic parallels between 2011 and 1981 riots. The Guardian. 29 December 2011. <http://www.guardian.co.uk/uk/2011/dec/30/national-archives-1981-riots-parallels>, last accessed: 9 February 2012.

[100]Rogers, S., Sedghi, A. and Evans, L. UK riots: every verified incident – interactive map. The Guardian. 29 December 2011. <http://www.guardian.co.uk/news/datablog/interactive/2011/aug/09/uk-riots-incident-map>, Last accessed: 9 February 2012.

[101] T. Coffman, S. Greenblatt, and S. Marcus, “Graph-based technologies for intelligence analysis,” *Communications of the ACM*, vol. 47, no. 3, pp. 45–47, 2004.

[102] Cheong, F and Cheong, C 2011, 'Social Media Data Mining: A Social Network Analysis of Tweets during the Australian 2010-2011 Floods', in Peter B. Seddon and Shirley Gregor. (ed.) *PACIS 2011 - 15th Pacific Asia Conference on Information Systems: Quality Research in Pacific*, Brisbane, Australia, 7-11 Jul 2011, pp. 1-16.

[103] Ediger, David, et al. "Massive social network analysis: Mining twitter for social good." *Parallel Processing (ICPP), 2010 39th International Conference on*. IEEE, 2010.

[104] Gruzd, Anatoliy, Barry Wellman, and Yuri Takhteyev. "Imagining Twitter as an imagined community." *American Behavioral Scientist* 55.10 (2011): 1294-1318.

[105] Jay L. Devore. *Probability and Statistics for Engineers and the Sciences*. 7th Ed. (Thomson: Belmont, CA) 2008. Pg. 463.

[106] Newman, Mark EJ. "The structure and function of complex networks." *SIAM review* 45.2 (2003): 167-256. Pg 168.

[107] Newman, Mark. *Networks: an introduction*. Oxford University Press, Inc., 2010. Pg 110, 122.

[108] Borgatti, Stephen P., and Daniel S. Halgin. "On network theory." *Organization Science* 22.5 (2011): 1168-1181.

[109] Newman, Mark EJ. "The structure and function of complex networks." *SIAM review* 45.2 (2003): 167-256. Pg 168.

[110] Dorogovtsev, Sergey N., and Jose FF Mendes. "Evolution of networks." *Advances in physics* 51.4 (2002): 1079-1187. Pg 1080.

[111] Boccaletti, Stefano, et al. "Complex networks: Structure and dynamics." *Physics reports* 424.4 (2006): 175-308.

[112] Radicchi, Filippo, Santo Fortunato, and Alessandro Vespignani. "Citation networks." *Models of Science Dynamics* (2012): 233-257. Pg. 234.

[113] Borgatti, Stephen P., and Daniel S. Halgin. "On network theory." *Organization Science* 22.5 (2011): 1168-1181. Pg. 1168

[114] Radicchi, Filippo, et al. "Diffusion of scientific credits and the ranking of scientists." *Physical Review E* 80.5 (2009): 056103.

[115] Nisonger, Thomas E. "JASIS and library and information science journal rankings: A review and analysis of the last half-century." *Journal of the American Society for Information Science* 50.11 (1999): 1004-1019. Pg. 1004.

[116] Hirsch, Jorge E. "An index to quantify an individual's scientific research output." *Proceedings of the National Academy of Sciences of the United States of America* 102.46 (2005): 16569-16572. Pg. 16569.

[117] Ding, Ying, et al. "PageRank for ranking authors in co-citation networks." *Journal of the American Society for Information Science and Technology* 60.11 (2009): 2229-2243.

[118] Newman, Mark EJ. "The structure and function of complex networks." *SIAM review* 45.2 (2003): 167-256. Pg 172.

[119] Rapoport, Anatol, and William J. Horvath. "A study of a large sociogram." *Behavioral Science* 6.4 (2007): 279-291.

[120] Padgett, John F., and Christopher K. Ansell. "Robust Action and the Rise of the Medici, 1400-1434." *American journal of sociology* (1993): 1259-1319.

[121] Sparrow, Malcolm K. "The application of network analysis to criminal intelligence: An assessment of the prospects." *Social networks* 13.3 (1991): 251-274.

[122] Ressler, Steve. "Social network analysis as an approach to combat terrorism: past, present, and future research." *Homeland Security Affairs* 2.2 (2006): 1-10.

[123] Diesner, Jana, and Kathleen M. Carley. "Revealing social structure from texts: meta-matrix text analysis as a novel method for network text analysis." *Causal mapping for information systems and technology research: Approaches, advances, and illustrations* (2005): 81-108.

[124] Otte, Evelien, and Ronald Rousseau. "Social network analysis: a powerful strategy, also for the information sciences." *Journal of information Science* 28.6 (2002): 441-453.

[125] Carley, Kathleen M. "Dynamic network analysis." *Dynamic social network modeling and analysis: Workshop summary and papers*. Committee on Human Factors, National Research Council, 2003.

[126] Diesner, Jana, and Kathleen M. Carley. "Revealing social structure from texts: meta-matrix text analysis as a novel method for network text analysis." *Causal mapping for information systems and technology research: Approaches, advances, and illustrations* (2005): 81-108. Pg. 11.

[127] Carley, Kathleen M. *A dynamic network approach to the assessment of terrorist groups and the impact of alternative courses of action*. CARNEGIE-MELLON UNIV PITTSBURGH PA INST OF SOFTWARE RESEARCH INTERNAT, 2006.

[128] Carley, Kathleen M., Ju-Sung Lee, and David Krackhardt. "Destabilizing Networks1." *Connections* 24.3 (2002): 79-92.

[129] Carley, Kathleen M., et al. "Destabilizing dynamic covert networks." *Proceedings of the 8th international Command and Control Research and Technology Symposium*. 2003.

[130] Carley, Kathleen M., Jeffrey Reminga, and Natasha Kamneva. "Destabilizing terrorist networks." *Institute for Software Research* (1998): 45.

[131] Diesner, Jana, and Kathleen M. Carley. "Revealing social structure from texts: meta-matrix text analysis as a novel method for network text analysis." *Causal mapping for information systems and technology research: Approaches, advances, and illustrations* (2005): 81-108. Pg. 11.

[132] Krackhardt, David, and Kathleen M. Carley. *PCANS model of structure in organizations*. Carnegie Mellon University, Institute for Complex Engineered Systems, 1998.

[133] United States. Defense Science Board. Task Force on Operations Research Applications for Intelligence, Surveillance, and Technology United States. Office of the Under Secretary of Defense for Acquisition. *Report of the Defense Science Board Advisory Group On Defense Intelligence, Operations Research Applications for Intelligence, Surveillance and Reconnaissance (ISR)*. Washington, D.C.: Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics, 2009.

[134] Smith, CM, WT Scherer, A Todd, and DT Maxwell. Quantitative approaches to representing the value of information within the intelligence cycle. Working paper.

[135] The Work of a Nation. Central Intelligence Agency. < https://www.cia.gov/library/publications/additional-publications/the-work-of-a-nation/ >, last accessed 10 Dec 2012. Pg. 27

[136]Treverton, Gregory F. and C Bryan Gabbard. Technical Report: Assessing the Tradecraft of Intelligence Analysis. Rand Corporation. Santa Monica, CA. 2008. Pg 3-4

[137] Kaplan, Edward. OR Forum – Intelligence Operations Research: The 2010 Philip McCord Morse Lecture. Operations Research Articles in Advance. Pg 1-13. <http://or.journal.informs.org/content/early/2012/07/03/opre.1120.1059.full.pdf> last accessed 10 Dec 2012.

[138] Smith, CM, WT Scherer, and S Carr. Intelligence Impact on Network Analysis. Working paper.

[139] Smith, CM, WT Scherer, and A Todd. Quantitative Intelligence Analysis. Working paper.

[140] Li, Huajing, et al. "CiteSeerx: an architecture and web service design for an academic document search engine." *Proceedings of the 15th international conference on World Wide Web*. ACM, 2006.

[141] *A.M. Turing Award*. ACM. Web. 01 Mar 2013.

[142] Peterson, Ivars. Rating Researchers. Science News. December 2, 2005. < http://www.sciencenews.org/view/generic/id/6824/description/Rating_Researchers > Last Accessed 22 March 2013.

[143] Borgatti, S.P., M.G. Everett, and L.C. Freeman. 2002. *UCINET 6 for Windows: Software for Social Network Analysis*. Harvard: Analytic Technologies.

[144] Krüger, Jens, and Rüdiger Westermann. "Linear algebra operators for GPU implementation of numerical algorithms." *ACM Transactions on Graphics (TOG)*. Vol. 22. No. 3. ACM, 2003

[145] Bolz, Jeff, et al. "Sparse matrix solvers on the GPU: conjugate gradients and multigrid." *ACM Transactions on Graphics (TOG)*. Vol. 22. No. 3. ACM, 2003.

[146] Van Glabbeek, Rob J., Scott A Smolka, and Bernhard Steffen. "Reactive, generative, and stratified models of probabilistic processes." *Logic in Computer Science, 1990. LICS'90, Proceedings, Fifth Annual IEEE Symposium on e*. IEEE, 1990.

[147] Van Glabbeek, Rob J., and W. Peter Weijland. "Branching time and abstraction in bisimulation semantics." *Journal of the ACM (JACM)* 43.3 (1996): 555-600

[148] Bergstra, Jan A., and Jan Willem Klop. "Process algebra for synchronous communication." *Information and control* 60.1 (1984): 109-137.

[149] Pinski, Gabriel, and Francis Narin. "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics." *Information Processing & Management* 12.5 (1976): 297-312.

[150] Bollen, Johan, Marko A. Rodriquez, and Herbert Van de Sompel. "Journal status." *Scientometrics* 69.3 (2006): 669-687.

[151] Dellavalle, Robert P., et al. "Refining dermatology journal impact factors using PageRank." *Journal of the American Academy of Dermatology* 57.1 (2007): 116-119.

[152] Campiteli, Monica G., et al. "Hirsch index as a network centrality measure." *arXiv preprint arXiv:1005.4803* (2010).

[153] Nanda, Santosh K., et al. "Emphasis on validation in research: A meta-analysis." *Scientometrics* 48.1 (2000): 45-64.

[154] Malcolm K. Sparrow, "The application of network analysis to criminal intelligence; an assessment of the prospects," *Social Networks* 13(1991): 261-263.

[155] Wasserman, Stanley, and Katherine Faust. *Social network analysis: Methods and applications*. Vol. 8. Cambridge university press, 1994.

[156] Carley, Kathleen M. "Dynamic Network Analysis" In: Breiger, R., Carley, K., Pattison, P. (Eds), Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers. Committee on Human Factors, National Research Council, pp. 133-145.

[157] Sargent, Robert G. "Verification and validation of simulation models." *Proceedings of the 37th conference on Winter simulation*. Winter Simulation Conference, 2005.

[158] Vieira, E. S., and J. A. N. F. Gomes. "Citations to scientific articles: Its distribution and dependence on the article features." *Journal of Informetrics* 4.1 (2010): 1-13.

[159] Harzing, Anne-Wil. "Google Scholar-a new data source for citation analysis." *University of Melbourne* (2008).

[160] Devore, Jay L. Probability and Statistics for Engineering and the Sciences. 7th Ed. Thomson, Belmont, CA) 2008. pg. 116.

[161] Gabbay, Murdoch, and Andrew Pitts. "A new approach to abstract syntax involving binders." *Logic in Computer Science, 1999. Proceedings. 14th Symposium on*. IEEE, 1999.

[162] Gladwell, Malcolm. *The tipping point: How little things can make a big difference*. Little, Brown, 2006.

[163] Ibid. Pg 12.

[164] Adams IV, N. I., et al. "Revised 5 report on the algorithmic language scheme." *ACM Sigplan Notices* 33.9 (1998): 26-76.

[165] Gladwell, Malcolm. *The tipping point: How little things can make a big difference*. Little, Brown, 2006.

[166] Wei, Wei, Jürgen Pfeffer, Jeffrey Reminga, and Kathleen M. Karley. *Handling Weighted, Asymmetric, Self-Looped, and Disconnected Networks in ORA*. No. CMU-ISR-11-113. Carnegie-Mellon Univ Pittsburgh, PA, Inst of Software Research Internat, 2011. Pg. 13-14.

[167] Ibid. Pg. 18-19.

[168] ORA software. *ORA Help: Measures*. Center for Computational Analysis of Social and Organizational Systems (CASOS). Version 3.0.0.2.

[169] Wei, Wei, Jürgen Pfeffer, Jeffrey Reminga, and Kathleen M. Karley. *Handling Weighted, Asymmetric, Self-Looped, and Disconnected Networks in ORA*. No. CMU-ISR-11-113. Carnegie-Mellon Univ Pittsburgh, PA, Inst of Software Research Internat, 2011. Pg. 9-10.