Developing a Language for Applied Causal Analysis: The Assessment of Causal Networks in Interdisciplinary Research

by

Kyle Turpin

Department of Civil and Environmental Engineering
Duke University

Date:_______________________
Approved:

____
Mark Borsuk, Advisor

____
David Carlson

____
Marc Jeuland

Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science in the Department of
Civil and Environmental Engineering in the Graduate School
of Duke University

2018

ABSTRACT

Developing a Language for Applied Causal Theory—the Assessment of Causal Networks in Interdisciplinary Research

by

Kyle Turpin

Department of Civil and Environmental Engineering
Duke University

Date:_______________________
Approved:

____
Mark Borsuk, Advisor

____
David Carlson

____
Marc Jeuland

An abstract of a thesis submitted in partial
fulfillment of the requirements for the degree
of Master of Science in the Department of
Civil and Environmental Engineering in the Graduate School of
Duke University

2018

Abstract

Integration of disparate research fields has become a major concern in recent years due to the increasing complexity of the issues that face policy makers and researchers. Concerted efforts have therefore been initiated to remove the traditional barriers between research fields to allow for greater cooperation between policy makers and researchers, particularly in the fields of health, the environment, and development. The Bridge Collaborative is one such organization dedicated to facilitating this process through the use of results chains. However, because of a lack of experimental data or observational datasets traditionally endemic to interdisciplinary policy research, they lack an effective mechanism for analyzing causal dependence among network variables. The purpose of this thesis is therefore to create a method of analyzing causal relationships using expert knowledge that can still pass the rigorous tests necessary to assert causality in the traditional experimental and observational data approaches. Building upon previous work of statisticians, philosophers, and computer scientists, I create a question template that will allow a researcher to easily check and refine a causal network and explore alternatives to that network based on experience and elicited expert judgement alone. I then perform a case study using this template based on the work of the Food-Energy-Water (FEW) Catalyst project, a group initiative within the Bridge Collaborative, to review a causal network based on a systematic literature search.

I conclude that a causal network can indeed be constructed, explored, and adjusted using logical reasoning and expert judgement—a finding that has implications for researchers seeking to create reliable models using causal networks as their base.

Contents

List of Tables

List of Figures

Acknowledgements

This thesis would not have been possible without the patience and guidance of my advisor, Dr. Mark Borsuk. Additional thanks are due to Dr. Ryan Calder, Rebecca Marx, Andrea Alatorre Troncoso, and the other researchers responsible for the creation of the FEW Catalyst network from which I drew my case study. Special acknowledgement should also be given to Dr. Ryan Calder who provided me the baseline code for my case study and helped me think through analyzing the causal network. A final thanks to my parents, sister, and friends who listened to me agonize about my thesis for the past year—your support and patience will not soon be forgotten!

1. Introduction

The nature of knowledge connectivity in the Information Era is complex and institutional in nature. Links between the environment, health, and human development have been known for decades (Haines et. al., 2006), however these fields' current research goals remain narrowly defined (Maxwell et. al. 2016), funding for new research projects is very community specific, and researcher expertise and networks are largely isolated (Petruney, 2016). Additionally, leaders often only consider a narrow range of policy options based off the research to which they have access and understand, professional incentives maintain the research funding status quo, and training and field procedures remain highly segregated across different fields (Brandt et. al., 2015). This engenders a research culture that is insular in nature where cross-field collaboration is the exception rather than the norm. Enacting effective change will require a paradigm shift from a research field composed of a patchwork of different evidence bases and procedures to a centralized model and evidence base. It is from this database that related fields can then draw upon resources, thereby allowing a more holistic approach to solving research and policy questions (Whitmee et. al., 2015).

To illustrate this challenge, consider an epidemiologist and a water resources engineer working on projects in the same small village in India: they could both be experts in their respective fields, but it would be unrealistic to expect the epidemiologist to also be an expert on water resource engineering and vice-versa. That being said, both doubtlessly have information to contribute to each other's projects—the epidemiologist's project tracking cholera outbreaks is fundamentally tied to the village's water resources; the engineer's project of providing the village with clean drinking water depends largely on the villagers' current consumption habits. The two are well-versed in the literature of their respective fields but do not necessarily follow the research being done in other fields—if these researchers had some means of communication allowing them to correct their work then each would likely walk away with a more holistic perspective that better enables them to solve their respective project problems. However, their work institutions are not geographically close, so how does one go about connecting these two minds? This is the paradigm shift that the Bridge Collaborative seeks to achieve.

The Bridge Collaborative was initiated in 2016 by The Nature Conservancy, Duke University, the Program for Appropriate Technology in Health (PATH), and the Food Policy Research Institute (IFPRI) (The Bridge Collaborative, 2018). The group's driving mission is to incite a fundamental shift in the way researchers and policy-makers approach planning, funding, and executing collaborative initiatives aimed at combatting our world's most pressing challenges. The group draws its inspiration from the Human Genome Project (1990-2003), an international collaborative spanning numerous research fields to create a functioning encyclopedia of the nature and location of all 20,500 genes making up the human genome (National Human Genome Research Institute, 2018). Similarly, the Bridge Collaborative seeks to tear down disciplinary walls preventing environmental, health, and development professionals from collaborating in the solving of world issues. To achieve this, the Collaborative has three long term goals: to better align the disparate agendas of the health, environment, and development research communities toward common goals; to standardize and unify research community project planning and evidence assessment methods thereby making it transferrable across disciplines; and to catalyze the synthesis of evidence across sectors to accelerate how the research community responds to problems (Tallis et. al., 2017). The first year of its existence saw over 150 members spanning 110 organizations come together to create a Practitioner's Guide on cross-sector planning and evidence evaluation to operate as a tool for researchers to better able access topical information from other research sectors. The objectives of this thesis build on the Practitioner's Guide, so I will begin by describing its basic framework—we will therefore start by exploring the power of this tool and examining the areas within that my thesis will make the greatest contributions.

1.1 Understanding the Practitioner's Guide

The Practitioner's Guide was created to catalyze problem solving across the health, environment, and development research sectors by codifying guidelines adherent to a set of six principles that the Collaborative identified as the basis for cross-sectional research transfer (Tallis et. al., 2017). These six principles are as follows:

1.) Use evidence to inform decisions about which systems function and which do not.
2.) Act now and learn by doing by acknowledging that progress can be made despite knowledge gaps and encouraging flexibility and learning in how we act.
3.) Seek and respect other perspectives by constantly believing that the goals of one research sector can be better achieved by embracing ideas from other sectors.
4.) Be intentional about inclusion to foster an environment of collaboration between sectors and for under-represented groups.
5.) Strive to do no harm by constantly seeking out novel solutions that will not provide success to one sector to the detriment of another.
6.) Share information openly and transparently while respecting individual's desire for privacy or anonymity to facilitate a community of trust among researchers.

The Collaborative, building off previous work on integrative research strategies (OECD, 2001; Petruney, 2016; PAGE, 2014), found that following these principles was critical toward framing their work across sectors to promote mindsets of trust among researchers (a group composed, essentially, of professional skeptics). However, whereas previous work only called for greater intersectional collaboration, the Practitioner's Guide sought to create operational guidelines for researchers to conduct evidence gathering and project planning. The crux of these operation guidelines lies in results chain analysis and evidence evaluation— as it currently stands, all health, environment, and development work more-or-less comply with the following iterative guidelines:

1.) “Create the team
2.) Define and analyze the situation and set goals
3.) Analyze possible interventions or develop research hypotheses using results chains
4.) Choose interventions (or research questions) and select impacts, metrics and a monitoring plan
5.) Implement
6.) Monitor, evaluate, and adapt” (Tallis et. al., 2017)

The Collaborative focuses its attention on that third step— creating results chains that incorporate multiple sector viewpoints to reveal interactions and connections across sectors. A results chain, as defined by the Practitioner’s Guide, is the visual representation of the logic and theory by which an intervention leads to positive and negative consequences (Tallis et. al., 2017). Results chains are usually composed of nodes, which represent causal drivers or consequences, and links, which represent hypotheses about how a change in one node will impact those further downstream. Each node of the chain is therefore a random variable and the arrows connecting them indicate both a proposed relation and the directionality of that relationship between variables—the direction of the error indicates which node is the cause and which node is the effect. I will further discuss the relationships that can exist between nodes in a later section.

Consider the following scenario based off an example from the Practitioner’s Guide (Tallis et. al., 2017): a local town council is working with a timber company to design a sustainable method of thinning forests to decrease wildfire intensity. Fire intensity has increased in recent years leading to difficulty in conserving endangered tree species and causing some to consider the use of machinery for forest thinning. Simultaneously, the region is seeking to combat high rates of respiratory illness by introducing more readily available inhalers to the local population. To further complicate the situation, a renewable energy group is introducing micro-energy subsidies to encourage private adoption of solar energy panels. Each of these, in the status quo, is trying to achieve their goals using only their group's point of view (Fig, 1), but in reality, the groups' goals are all interconnected (Fig. 2). Mechanical thinning and sustainable energy could be used in tandem to encourage environmental conservation while simultaneously decreasing the intensity of forest fires, thereby improving respiratory health.

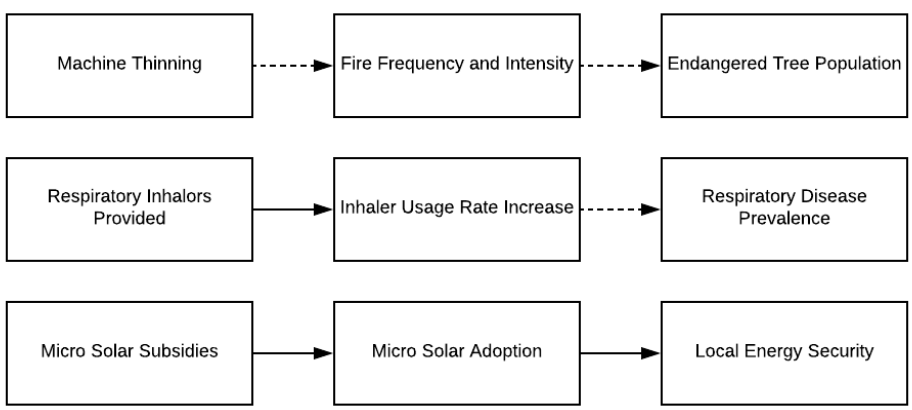


Figure 1. A simplified results chain showing single sector causal logic where the nodes (squares) are actions, solid lines are positive hypothesized links, and

dotted lines are negative hypothesized links. There is no connection between any of the three organizations' agendas. (Adapted from Tallis et. al., 2017).

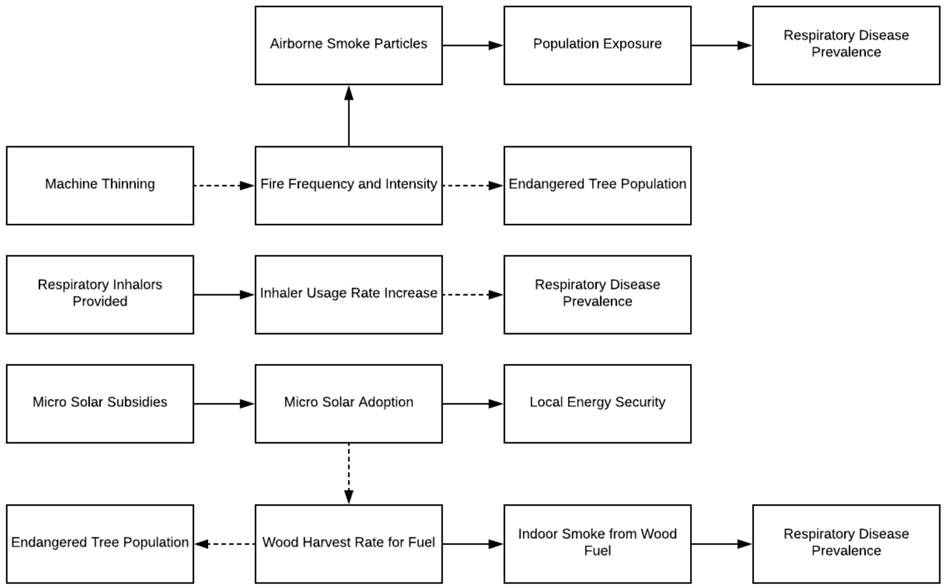


Figure 2. A simplified, albeit more comprehensive, results chain showing a holistic view of how the three organizations' agendas can work in tandem to better achieve everyone's goals. (Adapted from Tallis et. al., 2017)

This is only a simplified version of a results chain—actual chains can include hundreds of nodes and links—this is not surprising considering the complexity of many modern-day research and policy problems. Each link in the chain is a hypothesis that opens the door for evidence analysis where one can identify the logic and evidence of links that are logically weak or evidence insufficient. To provide a template for link assessment, the Collaborative created an evidence grading schematic that provides an easy-to-use interface for researchers to evaluate what methods and data analysis researchers in other fields are using. This makes it easier for intersectional link assessment of evidence strength. The Guide goes on to give strong guidelines on team creation, factors to consider during the research process, and tips on increasing the intersectionality of research questions. These aspects, although helpful to researchers, are not helpful to the main focus of this thesis; I will therefore now shift my attention from the guide and begin the discussion of the underlying logical principles therein that I believe require further refinement.

1.2 Thesis Objective

A central driver of the Bridge Collaborative's goal to synchronize the missions of separate sectors is the use of results chains—a strong tool when used correctly. However, I argue that the Collaborative's current approach to creating result chains asserts links between causal nodes without adequately addressing the nuances of the links themselves or questioning whether other links exist that are not currently drawn to that node. My research goal is therefore to create an analysis tool that assesses causal chains built off logic and expert opinion (henceforth called experience-based models) with the same level of rigor applied to assessing causal chains constructed from experimental or observational data. An experience-based model, for clarification, relies on reasonable observations and assumptions—this diverges from statistical and observational models insofar as it does not require empirical observational data but can instead be built upon previous research and logic. This technique is useful to researchers because the vast majority of interdisciplinary and policy research projects are unable to be conducted in a controlled setting— if I achieve my research goal, we can use expert elicitation to create and test the links within a causal network in the areas that statistical data is non-existent or non-pertinent. I will spend the rest of this thesis arguing the necessity of recognizing link nuances, explaining a system through which an analyst can elicit the beliefs of experts, converting those beliefs into logical assumptions, and then using those logical assumptions to create a results chain. The goal of this text is to create a template from which researchers can assess whether they are logically able to draw a causal link between nodes and how they can explore alternative causal models to the ones originally hypothesized. I will begin by laying the groundwork of causal systems and explaining why they are so important to this question.

2. History of Causal Analysis

The history of research has been driven by the researcher's desire to establish causation. In its most primitive state, research is a project driven by the observation that A has some sort of relationship to B—if I notice that the leaves of deciduous trees change color at the end of every September then I know there is some type of relationship between the time of year and the color of the tree leaves. The research project in question would then seek to establish a causal relationship between time of year and tree color through observation and data collecting. But how does our primitive researcher go about establishing this connection? Current accumulated knowledge tells us that the time of year determines the amount of sunlight energy a leaf receives and that the color change is due to a breakdown of chlorophyll and a stalling of photosynthetic processes in response to decreased light exposure (Leopold, 2005). However, our primitive researcher does not know this and is faced with how to test for a causal relationship between the two observed variables (time of year and leaf color). It is from such questions that the controlled experimental research method was born.

2.1 The Controlled Experiment

The controlled experiment is a method that determines that control of the variables, and not how they were controlled, is critical to experimental design (Shipley, 2016). This method consists of proposing a hypothetical cause-effect relationship, manipulating one variable while holding all others constant, and then comparing the actual outcome with the predicted outcome. For example, if I hypothesized that leaving my garden hose on in the backyard is the root cause of mud, then I would leave the garden hose off and control all other variables in the expectation that no mud would form. The weakness of this method is that one is never sure that all relevant variables have been verified or that all variables can indeed be controlled. I might leave my garden hose off and then return the next day to still find mud—if I had not considered the afternoon storms the day before then I might have been tempted to erroneously state that leaving my garden hose on is not a cause of mud. Another weakness is that there are so many confounding factors in a natural setting that it is next to impossible to account for them all, as seen by the variables in our researcher's deciduous tree example (time of year, weather, soil type, tree parasites, etc.). While the controlled experiment is a powerful tool and has been responsible for a great many discoveries in science, its chief weakness makes it of limited use in natural or overly-complicated systems.

2.2 The Randomized Experiment

To combat this weakness, Sir Ronald Fisher proposed the idea of experimental randomization in the early 20th century. To illustrate this idea, let us consider one of the agricultural applications that Fisher would have examined in his research as an agricultural scientist on crop yields (Box, 1980; Shipley, 2016). Suppose we want to determine whether the addition of a certain kind of fertilizer will significantly increase crop output—to achieve this we divide our field into thirty plots and sow seeds across each plot uniformly. Our treatment variable is our fertilizer, which is either added to a section or not added— for each plot, we will write down a number on a scrap of paper that decides whether the plot will receive fertilizer or not and then throw that scrap into a proverbial hat. Half of the plots will then be randomly selected to receive fertilizer and half will not—in this way the fertilizer will be distributed randomly across the plots. If we find that the fertilized plots, on average, have higher yields then, barring a very rare random corollary, we have good evidence that there is a positive association between fertilizer and crop yield.

The genius of this approach is that it greatly reduces the chance of confounding variables making a significant impact on the experiment's results. Even if soil moisture or sunlight exposure has impacts on crop yield, the randomized scattering of fertilizer, if it does indeed increase crop yield, will show higher results on average across all the plots on which it is applied. In the traditional controlled experiment where half the field is fertilized and the other half is held constant, local soil moisture or sunlight exposure discrepancies between the two halves would impact the results despite the researcher's best efforts to control all variables. Randomization, while being unable to categorically remove unaccounted or uncontrollable variables, can at least decrease the likelihood of such factors affecting the experimental design.

By randomizing our experiment, we can create a sampling distribution that allows us to calculate the probability of our treatment variable's effect being due to chance rather than causal relation. Statisticians, including Fisher, built further on this idea by asserting that randomization differentiates between causal associations and associations due to an unaccounted common variable shared by the treatment and response variables (Fisher, 1935; Kempthorpe, 1979; Kendall & Stuart, 1983). When a systematic relationship is observed that cannot be attributed to sampling fluctuations then the natural conclusion is that some type of causal mechanism is at work—randomization allows us to attach a statistical value to the likelihood that this causal mechanism is the one that we are asserting. We can then quantitatively assert that a causal relationship exists beyond a reasonable doubt even if we cannot assert that it is a logical certainty. Fisher states this best himself when he wrote that randomization "relieves the experimenter from the anxiety of considering and estimating the magnitude of the innumerable causes by which his data may be disturbed" (Fisher, 1935).

2.3 Uncontrolled Observation

Fisher's work made great improvements in how researchers of the 20th century onward approached experimental design, however Fisher's applications of this approach remained dependent on experimental data. However, many of the issues that policy makers, researchers, and the Bridge Collaborative face are large, complicated problems that span many fields and are difficult to study in a controlled setting. A more flexible system of establishing causality is needed in this instance that allows for causal relationships to be established from natural observations rather than controlled experiments. Fisher was able to mathematize causality because he was still ultimately conducting a controlled experiment manipulating plant soil nutrients that gave him data with which he could create and analyze distributions. This is not the case in many of the interdisciplinary problems tackled by the Bridge Collaborative—Fisher's approach therefore falls short.

It is at this point that the work of Judea Pearl becomes relevant—his writings contributed greatly to the field through his assertions that causal relationships could be derived from uncontrolled observations in addition to controlled experiments. Probability theory, the ultimate product of Fisher's work and of other statisticians like him, provides both the principles and ability to draw inferences from scientific observations which is why it is such an attractive tool for many disciplines concerned with causal relationships. As Judea Pearl, professor at UCLA and author of "Causality", states in his preface, "In the last decade, owing partly to advances in graphical models, causality has undergone a major transformation: from a concept shrouded in mystery into a mathematical object with well-defined semantics and well-founded logic. […] Put simply, causality has been mathematized" (Pearl, 2000). However, Pearl observes that relying purely upon probability theory to explain causality has its limits: "A joint distribution tells us how probable events are and how probabilities would change with subsequent observations, but a causal model also tells us how these probabilities would change as a result of external interventions—such as those encountered in policy analysis, treatment management, or planning everyday activity" (Pearl, 2000). He then astutely remarks that construction of a causal network capable of being used for research requires that a strong set of assumptions, based on causal knowledge, be made so that interventions in the system would allow for the identification of causal mechanics between variables. Building off our hose example from earlier, we can format our scenario into a system that allows for interventions to be made to assess mechanistic relationships between variables.

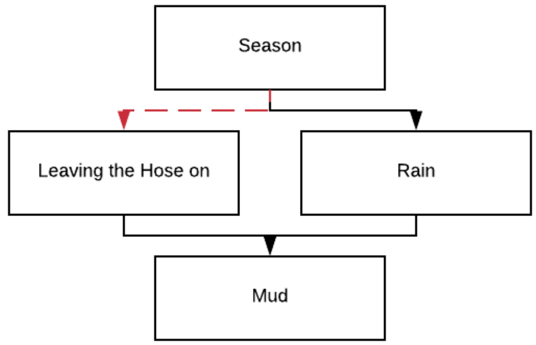


Figure 3. A simple causal network using Pearl's intervention oriented causal network.

Turning our hose on to water the grass is closely aligned with the season—one will not water the grass during a time of year that sees heavy rains. Similarly, the season determines the amount of rain that is seen. Both turning the hose on and rain will cause mud and both have an intimate relationship with season. Pearl states that we can test these causal mechanisms by an intervention—our intervention in this instance will be watering our lawn every day despite the time of the year. As our dotted red line in Fig. 3 represents, the causal relationship from season to mud through watering your lawn is broken; however, there still exists a causal relationship between season and mud through rain. I will go deeper into the mechanics and rules of causal networks in the next section, but this simple example illustrates what Pearl believed could be done through causal networks: asserting causal relationships through assumptions (empirical or otherwise) and then assessing causality through interventions in the system. As Pearl states, “causal relationships are ontological, describing objective physical constraints in our world, whereas probabilistic relationships are epistemic, reflecting what we know or believe about the world. Therefore, causal relationships should remain unaltered as long as no change has taken place in the environment, even when our knowledge of the environment undergoes changes” (Pearl, 2000). Probability theory will oftentimes distill causal relationships out of existence through the reduction of real world events to probabilistic language: knowing that the hose is on will lead the probability theory researcher to believe that it is not raining (after all, who would leave a hose on while it is raining) and incorporate that belief into his/her mathematical representation of the likelihood of mud. Pearl’s point is that a causal network will help you quickly identify that no relationship exists between whether the hose was on and the rain after one removed the link between season and the hose—only after consulting the causal network can one then convert into probabilistic language.

This now brings us to the question of how to create a causal network based off observations and then how to determine whether causality exists between individual parts of the network. Pearl reasoned through much of the logic of creating a translational device between causal models and observation-based statistical models, but it is Shipley who explicitly shows how this logic can be applied in a research scenario to determine causal relationships among variables. I will dedicate this next section to explaining this translational device to provide background for the work of his that will be the main basis of my contribution: a template providing the logical framework for causal relationship assessment based on the informal observation and experience of experts.

3. Causal Relations

Traditionally, significant causal research has been advanced with two types of models: the statistical and the experimental. Shipley, Pearl, and I argue that a causality model is the underlying logical framework of both statistical and experimental models and can be used independently of the other two models for research purposes—it is this approach that will prove most effective to the Bridge Collaborative as an inter-disciplinary communication device. To illustrate the differences between the systems, I will re-examine the mud example I used in the last section. Within an observational model, we would make the following hypothesis: “Observing rain will provide information about whether we will also observe mud.” This logic is symmetric insofar as it posits that observing rain gives us information about whether we will see mud and seeing mud gives us information about whether it has recently rained. The statistical model is similar to the observational model but goes one step further by assigning mathematical relationships between the variables. We can assign a relationship between the depth of mud and the amount of rain, allowing us to attach numbers to the logic we used in our observational model. An experience-based model relies upon assumptions based off research and logic—for example, we experience mud in our yard on a Tuesday and know that the sprinkler system only goes off on Thursdays, so we can deduce that it was likely rain that caused the mud. This section will discuss the logical relationships that can exist between network parts—the overarching question that this approach seeks to answer is how to assert a causal network based on logic and expert opinion as rigorously as one would assess a network created from experimental or observational data.

The problem before us is how to take informal observation or experience and translate it into causal representation that is reliable. However, before we can do this, it will first be useful to delve deeper into how causality has been translated into the language of statistics—this seemingly easy task of developing a translational link between causal language and statistical language took scientists the better part of the 20th century to decode (Shipley, 2016). The difficulty lies in having an apparatus that translates meaning perfectly between the two languages—although there exist rough correlations between causal and statistical languages, a translational system robust enough to perfectly convert causal meaning to statistical significance must be established to accommodate researcher needs.

Causal chains provide a graphical representation of relationships that a researcher believes exist between variables. For a controlled experiment, a variable is held constant and others are changed to see if there is any effect on a variable or variables downstream of the one held constant. In statistical analysis, probability distributions are determined for each node of the causal chain and full conditional probabilities can be determined for terminal nodes by running the entire system through a numeric simulator. However, for the causal network connections that I propose, reliant on expert opinions rather than on computable data or controlled results, it is necessary to use logical mechanisms to verify the existence of links between nodes. To this end, I will be using a concept known as directed separation (known more colloquially as "d-separation") which allows for the assessing of causal network connections without necessarily having a deep understanding of statistical knowledge.

3.1 d-Separation

d-Separation is effective when using a directed, acyclic graph such as the one represented in Figure 4; the baseline parts of the causal chain were introduced in section 1.2, but it is necessary to understand the finer vocabulary of causal chains when introducing d-separation. As mentioned earlier, the direction of the arrowhead connecting nodes indicates whether a node is a cause or an effect: A is a cause of B; B is both an effect of A and a cause of D. When one is able to follow the direction of the arrows across numerous nodes, one calls this path a directed path (A→B→D is a directed path); when one is able to draw a path irrespective to the direction of the arrows, one denotes this as an undirected path (A→B→D→E→F is an undirected path). A and E are causally independent variables, meaning that changes in either variable will not affect the other. B, C, and D are all causally dependent on A so that a change in A will produce a change in B, C, and D. A is a direct cause of B and C because a change in A will provoke a change in B and C regardless of the change of other variables. However, A is an indirect cause of D as A only affects D through the intermediary nodes of B and C, making D a causal descendant of A (one can reverse this relation and call A a causal ancestor of D).

With these terms in mind, an important concept of a causal graph is the idea of collider and non-collider nodes. A collider node is one that is caused by other nodes: it is a node with two or more arrows pointing at it, as we see for D. A fork node is the opposite, where a node receives no arrows but is the source of two or more, as we see for node E. An unshielded collider node is a set of three vertices where the central node is a collider and the extreme nodes share no connecting lines, such as B→D←E.

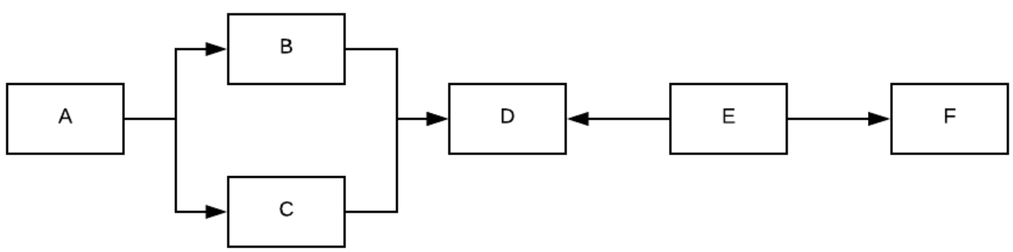


Figure 4. A simple causal chain where each letter represents a node.

With terms defined, we can begin describing the less intuitive idea of causal conditioning and its resulting importance in the concept of d-separation. As stated before, each node in the directed graph is a random variable; causal conditioning is the concept that determines whether any of these random variables are active (ON) or inactive (OFF). A node that is both a cause and an effect (like node B in A→B→D) is naturally active because it allows the causal influence of A to be transferred to D. Conversely, a collider node (like D in B→D←E) is considered naturally inactive as it blocks the causal influence of B from being transferred down a directed path to E. Conditioning on a variable changes its state: if a node was active and we condition upon it then it becomes inactive; the same can be said about the contrary. Logically, this makes sense: if we condition that B is true, then it no longer matters for D whether A has occurred or not, effectively blocking the indirect causal influence that A had on D through B. Similarly, if we condition D to be true, then we activate the node since we know that for D to be true then one or more of E, B, and C must be occurring too.

The concept of d-separation builds off this logical base: it is the "necessary and sufficient condition for two nodes in a directed, acyclic graph to be observationally independent upon conditioning on some other set of vertices" (Shipley, 2016). It specifies how observational or statistical causal information flows along the nodes and lines of a causal graph. Put simply, it is the logical base for how we can assess how causality is relayed from node to node when we condition upon nodal information; it is the backbone of causal relationships and gives us a logical tool to deduce causal graph relationships. It is a translational mechanism between the language of causal logic and the language of probability and will provide us a stronger quantitative analysis tool later. To determine d-separation between two vertices, one must complete the following steps:

A.) Identify every unique undirected path (ignoring the arrow direction) along the causal graph network that links the two vertices.

B.) Designate the set of vertices for which one is interested in determining d-separation. We will call this set D.

C.) If any undirected path between the two vertices contains a non-collider node, then there is no causal influence between the two vertices. This is because conditioning on a non-collider node makes the pathway inactive, as explained earlier.

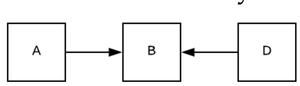
D.) For every undirected path between the vertices aside from set D, if any colliders exist along these paths then no causal influence exists along that path between the vertices. This is because conditioning upon a collider turns it active, so any nodes outside of our conditioned set D would therefore be inactive and incapable of transferring causal influence between the two nodes.

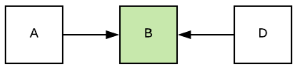
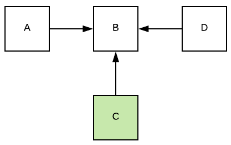
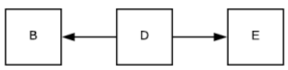
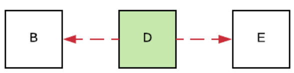
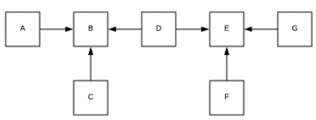
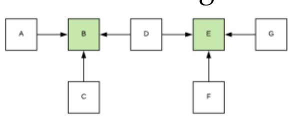
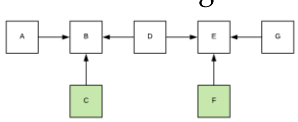
Logically, wrapping one's head around d-separation is difficult, but it is powerful tool once one has a full understanding of the concept. In Fig 4 we would say that A and F are d-separated, because the non-collider node D becomes inactive upon conditioning on the set, however we can also say that D and F are not d-separated when conditioning on E makes the set active and capable of passing along causal influence.

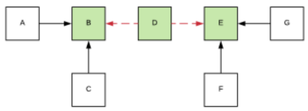
3.2 Introducing Shipley's Logic Chart

Now that we have a working grasp of d-separation, we can introduce the chart Shipley compiled of the various probabilistic independence relationships that can be deduced using d-separation in a directed graph (Shipley, 2016). For sake of clarity I have created a chart modeled off Shipley's original work; all variables will be in reference to Fig. 5 below the table:

Table 1. A re-creation of Shipley's logic chart. In the left column I provide the logical statement as written by Shipley along with a graphical representation; in the right column I provide an explanation of what is occurring logically. A dark square indicates that variable is being conditioned upon; dotted lines indicate that causal inference can no longer pass through that arrow.

Independence Relation	Explanation
A and D are unconditionally independent. 	No directed path exists between A and D due to B being inactive, so they have unconditional independence.

When we condition on B, A and D are not independent. 	Due to B being a collider between A and D, conditioning B makes it active thereby allowing causal transfer between A and D, making them no longer unconditionally independent.
When we condition on C, A and D are not independent. 	C is a causal ancestor of B and conditioning on C consequently activates B. B allows causal inference to transfer between A and D, making them no longer unconditionally independent.
B and E are not unconditionally independent. 	Because we are not conditioning upon anything, our forked node D, by definition, is active and allows causal transfer between B and E.
B and E are independent if we condition on D. 	By conditioning on D, we inactivate the node and close the only causal pathway linking B and E; this causes independence of the two variables in relation to each other.
A and G are unconditionally independent. 	The only causal pathway between A and G will be blocked by the inactive collider nodes B and E, allowing for unconditional independence.
When we condition upon both B and E then A and G are no longer independent. 	By conditioning upon B and E, we activate both nodes and create an undirected path of causal transfer between the nodes so that A and G are no longer independent.
When we condition upon both C and F then A and G are no longer independent. 	C and D are causal ancestors of B and E, so conditioning upon them activates B and E allowing for the same undirected causal transfer mentioned in the previous section.

When we condition upon B, D, and E, then A and G will be independent of each other. 	Although conditioning upon B and E activates both collider nodes, conditioning on D inactivates the forked node and removes any undirected causal path connecting A and G.

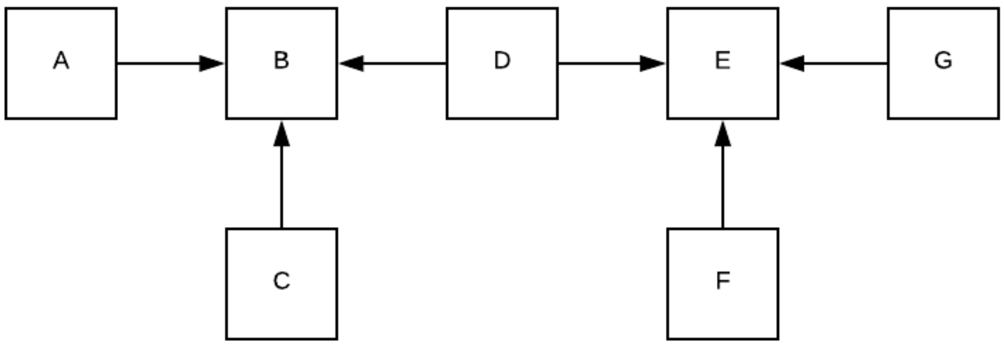


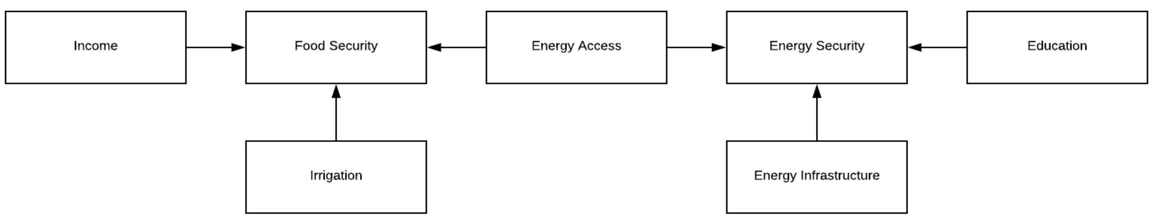
Figure 5. Reference causal network for explicatory chart.

This provides a base network of all the causal relationships that can exist between variables and provides us with the tools to explore what this logical framework looks like in a real-world setting.

4. Causal Analysis in the Real World

This section will create a qualitative understanding of Shipley and Pearl's logic for the use of question creation and application to real world causal networks. The structure of the network is the same as the one created by Shipley but is constructed of analogous variables to outline what Shipley's logic arguments look like in a real-world setting. To create this analogous real-world network, I chose for my nodes several research categories identified by the Food-Energy-Water (FEW) Catalyst team, a group associated with the Bridge Collaborative testing the effectiveness of network-based literary searches. The group created a search algorithm that they entered into Web of Science, an online scientific indexing service produced and maintained by the Institute for Scientific Information, for the purposes of creating a baseline literary database for scientific papers pertaining to energy, food, water, and other development issues that the Bridge Collaborative investigates. After this first wave creation of the database, the group then went through the abstracts of each of the thousands of papers that the algorithm collected and placed those papers into umbrella categories that summarized the topics of the papers therein. After the creation of these "nodes", the group then did a closer re-examination of the papers in the umbrella categories to determine the relationships that could be drawn between the individual categories, effectively creating a causal network system. It is from this database that I drew the inspiration for my analogous causal network.

For the below analogous causal network (Fig. 6), I drew the arrows loosely based off the arrows that the FEW Catalyst group drew in their work, however I modified a few for pedagogical purposes. In a later section I will then assess an actual network created by the FEW Catalyst search algorithm to test whether the framework assessment tools I develop in the next section are robust enough to address real world research issues addressed by Bridge Collaborative members.



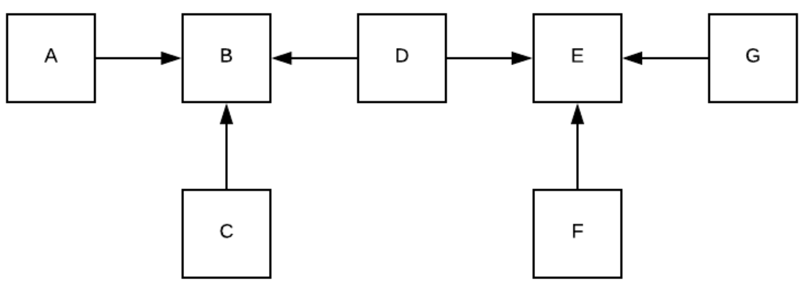


Figure 6. The analogous real-world causal network created using the FEW Catalyst search results and modified to match the structure of the network created by Shipley. The figure below it, Fig. 5, is included for easier reference.

For ease of writing, I will refer to the categories in my analogous network by the alphabetic categorizations that Shipley used to denote his network and then provide a description of how the logic would look in the analogous network. Additionally, my analysis here and in the next sections will be assuming binary variables. A reference table, Table 2, is included below to illustrate the nodal categories and the binary variable options for each node.

Table 2 A table of the alphabetical nodal names Shipley used in his causal network (Fig. 5), the analogous real-world nodal names that I used in my causal network (Fig. 6), and the binary variable categories available for each node.

Shipley Nodal Name	Real World Nodal Name	Binary Nodal Categories
A	Income	High/Low
B	Food Security	Secure/Insecure
C	Irrigation	Sufficient/Insufficient
D	Energy Access	Present/Absent
E	Energy Security	Secure/Insecure
F	Energy Infrastructure	Present/Absent
G	Education	High/Low

Variables in the real world are obviously more complicated than the binary ones I am using for my analysis, however I posit that multinomial or continuous nodes will follow the same causal logic as a binary system. I am using a binary system for pedagogical reasons so that the underlying logic of causal relationships can best be understood—I will discuss multinomial and continuous nodes in section 5.3. Now, let us begin parsing through this analogous network to get a better idea of which questions we will need to ask an expert to determine causal independence. For clarification, "conditioning" a node for the rest of this paper will mean that we have information on the variable in question; if we do not condition upon a node then that means we have no information on it.

4.1 Category 1: A and D are unconditionally independent when B is not conditioned upon.

The first of Shipley's categories examines nodes A, B, and D in which B is not conditioned upon and A and D have unconditional independence. In the real-world network, we would not condition upon Food Security such that Income and Energy Access are independent. That is to say, if we do not know anything about whether the population has food security, then learning about the population's income will tell us nothing about that population's energy access and vice versa—but does this make sense? Before continuing with this idea, I believe it is important to first pause and consider if node B has an "and" relationship with A and D or an "or" relationship and whether this relationship matters in d-separation. Consider our mud example in subsection 2.3—this is an example of an "or" relationship: if we see mud in our front yard, then we know it either rained <u>or</u> the hose was left running (both may very well have occurred, but only one is sufficient and necessary to cause mud). If we condition that we saw mud, then learning that it did not rain the day before would give us information that the hose was probably the cause of the mud—contrarily if we do not know if there is mud in the yard, then knowing that the hose was left on would give us no information on the likelihood that it rained and vice versa. If Income and Energy Access have an "or" relationship, then only one of these nodes is needed to cause Food Security—so long as we do not know about Food Security, then Income and Energy Access will be d-separated and the two will be causally independent of each other. If, however, Income and Energy Access have an "and" relationship, then Food Security will require both Income and Energy Access. Provided we do not have any information on Food Security then we can still posit that Income and Energy Access are d-separated—to illustrate this, let us consider if our mud example had an "and" relationship, that is, both the hose and rain are necessary for mud formation. If we do not know if there is mud then, like in our "or" relationship, learning that the hose was left on will give us no information on the likelihood that it rained the day beforehand and vice versa. The two are only linked by the fact that both are needed for mud—if we do not know if there is mud, then learning about one will not give us information about the other (thereby making the hose and rain d-separated). For an unconditioned collider node such as B, it therefore is not important to d-separation whether the parental nodes have an "and" or an "or" relationship—we will further explore whether this logic remains true in the next sub-section when we condition that we do know if there is mud or not.

Before continuing, however, it should be noted that it makes little intuitive sense that a population's energy access and its income are not linked—in general, higher income countries will have better energy access. We must therefore ask our expert whether the two share any causal parents that may allow for co-dependence when Food Security is de-activated. As a reminder, the main drive of this thesis is to provide researchers with tools for assessing links that are not drawn in the existent causal chain—this is a classic example of where one might use such tools. Here, one would first ask the expert whether Energy Access and Income have some level of significant dependence on each other or on other variables when Food Security is not known (put simply, will the nodes continue to vary in a similar or linked fashion when the Food Security link is removed). If the researcher affirms that this is the case, then the next step would be to ask whether the researcher could hypothesize the causal node that may link the two outside of Food Security. After finding this linking node (for our real-world example we will say that this linking node is the parent node, Economic Stability) we could then explain that lack of d-separation despite Food Security not being known. We will explore this idea in later sections.

4.2 Category 2: B is conditioned upon such that A and D are not independent.

We will next examine the relationship between Food Security, Income, and Energy Access, however now we have knowledge about the population's food security.

Because Food Security is a collider node, conditioning upon it activates it and allows for non-directed causal relation to connect Income and Energy Access. This is where we will finish our analysis of the "and-or" relationships that the nodes connected to a collider node can have. In the last section, we determined that "and-or" relationships had no impact on the d-separation of an unconditioned collider node, however I hold that this is not the case for a conditioned collider node. Let us first examine the "or" relationship by continuing with our mud example—in this instance, we condition that we see mud and then learn that it did not rain yesterday. We are therefore correct in saying that knowing that there is mud and that it did not rain yesterday gives us information that the hose was on, thereby making rain and the hose dependent. Is this correct in all cases, though?

Remember that in our mud example we initially linked the hose and rain with a parental node, Season. So long as Season was the common cause for the two descendant nodes then knowing that there was mud made information on the hose and rain dependent. However, after our intervention (turning on the hose every day of the year) severed the connection between Season and the hose, our information about the hose being on and whether it rained became d-separated again—despite seeing mud we could no longer infer knowledge about rain due to the hose being on every day of the year. In our real-world example, we hypothesized via our expert advice that the parental node, Economic Stability, connected Income and Energy Access. So long as that connection remains in place, information about the population's Food Security will create causal dependence between Energy Access and Income. However, should the connection between Economic Stability and one of its daughter nodes be severed, then having knowledge about the population's Food Security is not enough for causal inference to transfer between Income and Energy Access. "Or" relationships will therefore only follow Shipley's logic when the collider node's two parental nodes also share a parental node.

The "and" relationship is much more straightforward and follows Shipley's logic—if rain and the hose are necessary for mud, then seeing mud will tell us that it rained and the hose was left running. Even if we claim that the hose and rain share no parental node (removing Season in this case), seeing mud and knowing that the hose was left running yesterday will allow us to infer that it also rained. If Income and Energy Access are both required for Food Security, then knowing a population has food security will mean that information on its income or energy access is no longer causally independent and is consequently no longer d-spaced (this is the case with or without the Economic Stability node). In our conditioned collider node case we therefore must ask our expert, "In your experience, can a population have food security while either having only high income or only high energy access or are both required?" This will answer whether Income and Energy Access have an "and" or an "or" relationship and will allow for subsequent questions to be asked on whether there is causal dependence between Income and Energy Access (as we did at the end of subsection 4.1). If the expert affirms that the two nodes have an "and" relationship and are co-dependent, then we can conclude that the current depiction of the causal chain is correct. There may be intermediary nodes between the collider node and its current parents—for example, it may be more accurate to say that Income leads to the node, Greater Purchasing Power, which leads to Greater Food Access, which leads to Food Security—but this will not affect the causal dependence between Income and Energy Access.

If the expert states that we have an "or" relationship and the two nodes are dependent then we can conclude that they have a common parental node—Economic Stability, for example. If the expert states we have an "or" relationship and the two nodes are independent, then we can conclude that the collider node is correct but its parental nodes do not share a parental node. Finally, if the expert states that Income and Energy Access have an "and" relationship but operate independently when we know a population's Food Security, then we will have to re-examine the arrows drawn from Food Security to Energy Access and Income. We will explore this relationship when we discuss forked nodes. I will conclude this sub-section by stating that "and-or" analysis should be the first line of questioning for determining nodal relations as it has such a profound impact on d-separation. However, going forward I will assume that each of the relationships in the following sub-sections has an "and" relationship as this is the relationship that Shipley appeared to consider in Table 1. I will make note if any of the relationships will be affected by an "or" relationship but believe that constant un-necessary reference to it will distract from the causal relation analysis and undermine the purpose of this section.

4.3 Category 3: C is conditioned upon such that A and D are not independent.

Now, instead of conditioning upon Food Security, we will condition upon an ancestral node, Irrigation—this means we have information on the population's irrigation habits. Causal chains work in a way such that conditioning upon an ancestral node in a unidirectional chain will result in the activation of all descendant nodes. Here, Irrigation will activate our collider node Food Security and allow for non-directed causal transfer between Income and Energy Access. At first glance, using the same logic presented in the previous category, this makes sense; however, this category is nuanced by the degree of the relationship between irrigation and food security. For example, it has been proven that irrigation can oftentimes lead to huge improvements in daily caloric consumption of farmers using irrigation methods (Nkhata, 2014) which would most certainly contribute to food security. However, researchers have simultaneously observed that certain irrigation techniques can lead to the salinization and inundation of soils over long periods of time, leading to lower crop yields and further environmental degradation in the region (Lal, 2008). Additionally, there is great regional variability in regard to the effectiveness of irrigation on crop production—areas like Southeast Asia, where irrigation is a cultural aspect of farming, obtain far great yields from irrigation techniques than do regions of Africa, where irrigation techniques are not as established a part of regional farming cultures (Godray et. al., 2010). This is to say that the relationship between the Irrigation and Food Security nodes is much more nuanced than initially represented. An expert on the subject should therefore be questioned about the regional specificities of the nodes as well as whether the ancestral node has any links to other parts of the chain. For example, Energy Access and Irrigation often will have a causal relationship due to electric pumps leading to better irrigation techniques. A connection between Irrigation and Energy Access does not change the co-dependence of Income and Energy Access when Food Security is activated (this co-dependence will depend on whether the nodal "and-or" relationship, as elucidated in the final paragraph of the last sub-section), but the arrow between Irrigation and Food Security would need to be refined to express both the positive and negative impacts that irrigation has on food security. This could easily be done by adding two intermediary nodes—Greater Food Productivity and Environmental Degradation—between Irrigation and Food Security, thereby better elucidating the relationship as seen below. Receiving information on this relationship from the expert does not change the resulting co-dependence of Income and Energy Access through the activated Food Security node, but it is a clear example of how expert elicitation can create arrows and nodes that were not previously included in the causal chain.

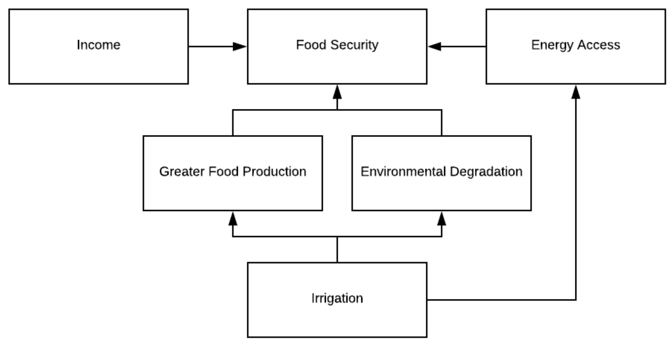


Figure 7 Updated left-hand portion of Fig 6 showing the nuanced linked between Irrigation and Food Security as well as the link between Irrigation and Energy Access.

4.4 Category 4: B and E are not unconditionally independent when D is not conditioned upon.

We now move our analysis to the forked node D which, when not conditioned upon, is naturally active, consequently allowing for non-directional causal transference between B and E. In our analogous example, we examine a population whose energy access we do not know. If we observe energy security or food security in the region then we can causally relate the two through this activated Energy Access node—even if we do not have information on the node itself, the fact that it exists as a causal bridge between the two daughter nodes is enough to causally related the two. By saying that

Energy Security and Food Security are not causally independent, we are merely saying that learning about one will give us information about the other. If my expert tells me that two nodes are causally not independent (that is, I can make an intervention to energy security and see a proportional change to food security), then that would naturally lead me to conclude that there is a parent node connecting the two about which I have no knowledge. By looking to the literature (or asking our expert), I will quickly find that improved energy access will often lead to greater energy and food security (Bazillian et. al., 2011) allowing me to connect Energy Security and Food Security through the forked node of Energy Access. I could then test this new network by making an intervention to Energy Access—if our hypothesis is correct then we will see changes to Food Security and Energy Security. This category is a tricky one to assess with expert advice as it is difficult determining when to treat nodes as connected by an unconditioned forked node rather than as a unidirectional causal chain. The question we first need to address is whether it is truly a cause-effect relationship between the two nodes or whether it is a parent node that connects the two. In this instance, having energy security is not always enough to have food security—a region that has suffered an intense summer will be resplendent in solar energy but will likely lack the water resources necessary for abundant food production. However, on average, a region with consistent energy access will have the infrastructure to continue electrical pumping of irrigation water. We can therefore determine whether we have a forked node by asking the logical questions, “Are Energy Security and Food Security independent given Energy Access?” If the answer is, “No,” then we can begin asking the expert questions to assert the relationship between the three variables—although, if Energy Security is codependent with Food Security given no information on Energy Access then we likely have a forked node (we would need to verify with our expert that the forked node is indeed Energy Access and not some other variable). If Energy Security is independent of Food Security without knowing anything about Energy Access, then we will either have a unidirectional chain (Food Security→Energy Access→Energy Security or the opposite direction) or a have a collider node (Food Security→Energy Access←Energy Security). To determine which option it is, we would ask our expert, “In your experience, does Energy Access require the presence of both Food Security and Energy Security?” If the answer is “Yes,” then we have an “and” relationship collider node where Food Security and Energy Security are d-separated. If the answer is “No,” then we either have an “or” relationship collider node or a uni-directional chain. We could verify this by asking the follow-up question “Can a population have Energy Security without Energy Access?” Intuitively, the answer will likely be “No,” implying a unidirectional causal chain rather than an “or” collider.

4.5 Category 5: When D is conditioned upon, B and E are independent.

Causal networks adhere to the Markov Chain rule, which states that it is only the previous node that determines our assessment of the probability of the next step in the chain occurring. If we have four ancestral nodes and only condition upon our last node, then it no longer has any impact on our decision-making process that the other three nodes occurred or not—it is only that last ancestral node that will impact our next step in our chain. This is the logic that allows for causal independence between Energy Security and Food Security when we have knowledge about our population's energy access. We know, looking at our network, that if our population has energy access then it will have food security and energy security—finding out that the population has energy security will tell us nothing new about the population's food security, since we already knew this information from our conditioning node. The only thing that impacts Food Security and Energy Security is the Energy Access node—all other nodes before that (in any direction) become irrelevant after we make our conditioning statement. We therefore would ask our expert, "Given our knowledge of our population's energy access, is there any co-variability of our population's food security or energy security outside of its energy access?" If the answer is "No," then we can conclude that all information regarding Energy Security and Food Security is dependent only on Energy Access and therefore independent from each other. If the answer is "Yes," then that is an indicator that there is another node connecting Energy Security and Food Security outside of Energy Access about which we know nothing. This would require asking our expert or re-assessing the literature for information on what this connection might be.

4.6 Category 6: A and G are unconditionally independent.

The logic behind this category is similar to that behind the logic of category one, only there is a great degree of separation between Income and Education. Since none of our collider nodes are being conditioned upon, Income and Education have no causal relation to each other—however, just as with category one, this does not make intuitive sense. It would make sense that higher levels of education and income would go hand-in-hand, meaning that if one knows information about the one then they could deduce information about the other. As in category one, this would naturally attract the attention of a researcher questioning an expert—if the expert asserts that there is causal dependence between Income and Education then there must be a link between the two nodes preventing d-separation. Similar to the ending of the last sub-section, the researcher would then question the expert or turn to the literature to determine the common node(s) connecting the two and then question the expert on the relationship between the nodes using the tools I outlined in the last sections. For example, if the expert posits that Income and Education are linked by Societal Stability, then my follow up question would be, "Is it possible to have high quality education and high income without societal stability?" If the answer is "No," then the following question would be "Does income and education co-vary when we have no information on a population's energy access or societal stability?" If the answer is "Yes," then we have identified all the missing links between the two; if the answer is "No," then there are still links connecting the two that must be identified.

4.7 Category 7: When B and E are conditioned upon, A and G are no longer independent.

Conditioning upon a collider node, as previously mentioned, activates that node and allows for causal information to pass through that node—so long as all the collider nodes between two nodes are conditioned upon and have "and" relationships, then there will always be a non-directed path for causal information to travel. In this instance, we are conditioning that our studied population has energy security and food security—finding out information about the population's education levels will therefore, due to our conditioning, give us information on the population's income levels. The only instance in which this will not be true is if there is an "or" relationship collider node whose parent nodes do not share a causal parent, as outlined at the end of subsection 4.2.

4.8 Category 8: When C and F are conditioned upon, A and G are no longer independent.

Conditioning upon ancestral nodes, as previously stated, will activate collider nodes. In this instance, we condition that our communities have energy infrastructure and irrigation, activating our collider nodes, Energy Security and Food Security, and allowing for a non-directed causal path to be drawn between Income and Education. This follows the logic laid out in subsections 4.3 and 4.7.

4.9 Category 9: When B, D, and E are conditioned upon, A and G are independent.

Our final category examines how causal relation translates across two collider nodes and a forked node. Although conditioning on our collider nodes activates them, conditioning upon our forked node deactivates it and effectively cuts the causal path linking Income and Education. Conditioning upon Energy Access asserts that learning about the studied population's income will tell us nothing new about the population's education and vice versa. This category and category five are the nodes that require the most tact in addressing since there are numerous variables that need to be explored before any assertion about causal relationship can be made. Conditioning on our forked node will inactivate it, as conditioning upon that node will make any information on both sides of it irrelevant to the other side. In this situation, we already have information on the collider nodes and the forked node—we would then want to ask our expert about whether there would be any codependence between Income and Education such that knowing about the one will give us previously unknown information on the other. Barring examples from previous examples where undrawn links existed between Income and Education, if there is no causal dependence then the causal chain is drawn correctly. If there is causal dependence, then that would require the chain to be reexamined to see if Energy Access is truly a forked node.

4.10 Summary

The takeaway of this section is that missing links between nodes can oftentimes be identified by simply asking if the two nodes are dependent on each other under particular conditions concerning other nodes—will gaining information on the one give us previously unknown information about the other? This section went through several different causal scenarios where dependence or lack thereof could be explored to re-write current nodes and arrows or add new ones entirely. The next section will explore how to take the tools elicited in this section and translate them into a working language that an expert can understand and from which the researcher can get impactful answers regarding their causal network design.

5. Questioning Causal Networks

We will now create a template of questions that researchers can use to assess causal relation and independence in causal networks. The power of this technique is that researchers conducting interdisciplinary projects that are unable to be executed in a controlled setting or that lack extensive observational data can test the feasibility of their logically created causal networks through the elicitation of expert opinion. Although such informal elicitation would never meet the stringent requirements necessary for a controlled experiment, these carefully worded questions can still allow researchers to mentally test valid hypotheses in situations where they cannot produce the data that typically needs to be used for a Fisher style statistical causal chain. The following questions will be created in such a way so as to allow researchers to identify and respond to areas of their causal network that can no longer be claimed after expert elicitation. This logic therein will be based on the Pearl-Shipley analysis that was described in an earlier section and adapted to the scenarios that arose from observation of the analogous chain in the last section. I will create the questions based off the types of relationships that can exist in a causal network and then create a master list at the end of this section of the questions that should be posed to an expert to logically assert causal relationships.

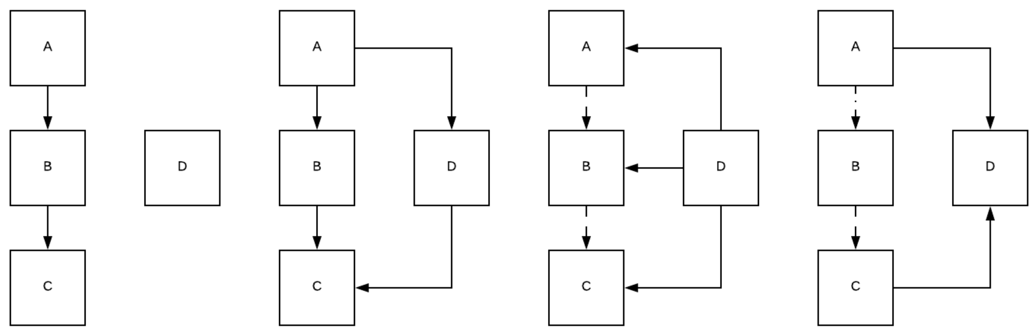
5.1 Initial Network Assessment

As was mentioned in the last section, the first line of questioning that a researcher should pose to an expert is on whether any relationships exist among the nodes that are not already shown by arrows in the graph. These situations are made clear when causal dependence exists between nodes that do not have an established causal path following Shipley's rules. The prudent researcher would therefore begin by going through the network one node at a time with the expert and asking how that node is conditionally related to the nodes nearest to it—this will effectively involve examining the nodes as I did in subsections 4.1-4.5 in the last section. However, the researcher must simultaneously put conditions on the question: an expert could draw a causal relationship between any two nodes if asked (this is just the nature of our inter-connected world), however we are only interested in a connection if it provides a previously undrawn directed connection between the nodes. I will discuss how to assess this later in this section.

In questioning causal relationships between nodes, we will likely add new nodes and arrows as we did in several examples shown in section four. We can naturally intuit when this will occur by questioning our experts about the nature of nodal conditional relationships and about the absence of potential nodes. The former line of questioning was covered in sub-sections 4.1-4.3 and can be elicited from an expert by asking questions meant to define parental nodal relationships (as was seen in 4.1-4.2 where defining the "and-or" relationship was instrumental in determining d-separation) or that helped separate complex relationships into simpler ones (like the kind that exists between Irrigation and Food Security). The latter line of questioning was covered in sections 4.4-4.6 and occurs in situations where causal dependence exists between nodes without the causal network supporting such co-dependence via Shipley's chart (Table 1). This situation requires asking the expert about potential ancestors and descendants the two nodes might share—from here, the researcher can then ask the expert about the nature of those relationships in a similar manner as how was done in sub-section 4.4 & 4.6. Our causal network will ultimately be used to test hypotheses in our research community, so having as comprehensive a causal network as possible is necessary for a well-executed project design.

This initial analysis will help assess the overall sensibility of the proposed network and will identify areas where causal relationships do not follow the rules established in Table 1. Relationships denoted in the causal chain should be reflective of the associated real-world phenomena— incorrect causal relationships drawn between nodes will lead to faulty conclusions that may hinder progress in a field rather than develop it. This next sub-section will therefore show how to analyze multiple causal chains and use the tools we developed in the last section to ask our expert pertinent questions that allow us to make correct assertions about nodal relationships.

5.2 Questioning Nodal Relations



I II III IV

Figure 8. Representation of different causal relationships that can exist between nodes. Figure I is our initial hypothesized causal relationship between the nodes; II, III, and IV are all alternate ways in which pathways can exist between the variables. Solid lines indicate a relationship, dotted lines represent relationships that are not considered in the analysis because the connecting node is not conditioned upon.

Using the Bridge Collaborative's approach, a researcher will create a causal network based off literature and then use that as a basis for an experimental design to test the impacts that individual nodes have on the entire chain. The initial network analysis exposes any sections that an expert's response raises concerns—these are the sections where an expert either asserts there is causal dependence where causal dependence should not exist or else that the network is not indicative of a real-world scenario. Figure 9 illustrates five different ways that several nodes can be inter-related. The researcher therefore needs to ask the correct questions to the expert to categorically rule out all but one of these possible relations. We will do this step by step.

Let us assert that the causal relationship we first propose to the expert is option I from Fig. 9—we ask if there is causal independence between nodes A and C when we know nothing about B, to which she replies that there is causal dependence. Given her non-statistics background, the researcher will have to ask a series of questions to the expert to determine which of the other relationships might be most fitting for the project. Given the co-dependence between A & C despite our lack of knowledge of B, we ask our expert whether there are any other direct relationships between A & C—she replies that, given her experience, there is a strong possibility that node D may relate the two. After the addition of node D, our updated chain looks like option II in Fig. 9—our next question is to determine whether A and C are causally dependent if both B & D are unknown. If our researcher answers they are, then that will mean one of two things: either we will need to add another node to connect A & C or else node D is a forked node as seen in option III in Fig. 9. Our expert is confident that B & D are the only two direct relationships linking A & C, so we will need verify whether our chain is some variant of option III. To do this, we ask our expert whether A, B, and C (or just A & C depending on the nature of the forked node) will be causally independent when we have information on D. If this is the case, D will be a forked node and we can assert that option III is the correct causal chain.

Now, let us backtrack to when we asked our expert if A & C are co-dependent when we have no information about B & D—let us now posit that they are independent of each other. We will now be deciding between scenarios II and IV. We can easily differentiate between the two by asking the expert whether, in his experience, D can be observed when only A or C is present. If the expert answers that both are required, then we can immediately deduce that our causal chain will look like option IV with an "and" relationship collider. If, however, the expert answers that A is known to cause D but C is either not required or he does not know if it is required, then we will need to decide between the unidirectional chain shown in option II and the "or" relationship collider node of option IV. We can differentiate between the two by asking our expert, "If we have information on D, will A & C be causally independent?" If the expert replies, "Yes," then we can conclude that we have the "or" relationship collider node in scenario IV. If the expert replies, "No," then we either have the causal chain shown in option II or an "or" collider node where A & C share a causal parent. We can make this distinction by simply asking the expert if A & C have common parent node—assuming that they do not, then we can conclude that option II is the best one.

The analysis provided above shows how one can use the tools I explained in section four to walk through a causal chain with an expert to determine its viability. I only discussed the four potential combinations that will cover the bulk of potential network types a researcher may encounter—there are other relationship combinations that could be explored, but for sake of brevity and understanding I decided to not focus on these. For clarification of my steps, I am including Figure 9 to graphically explain my question process.

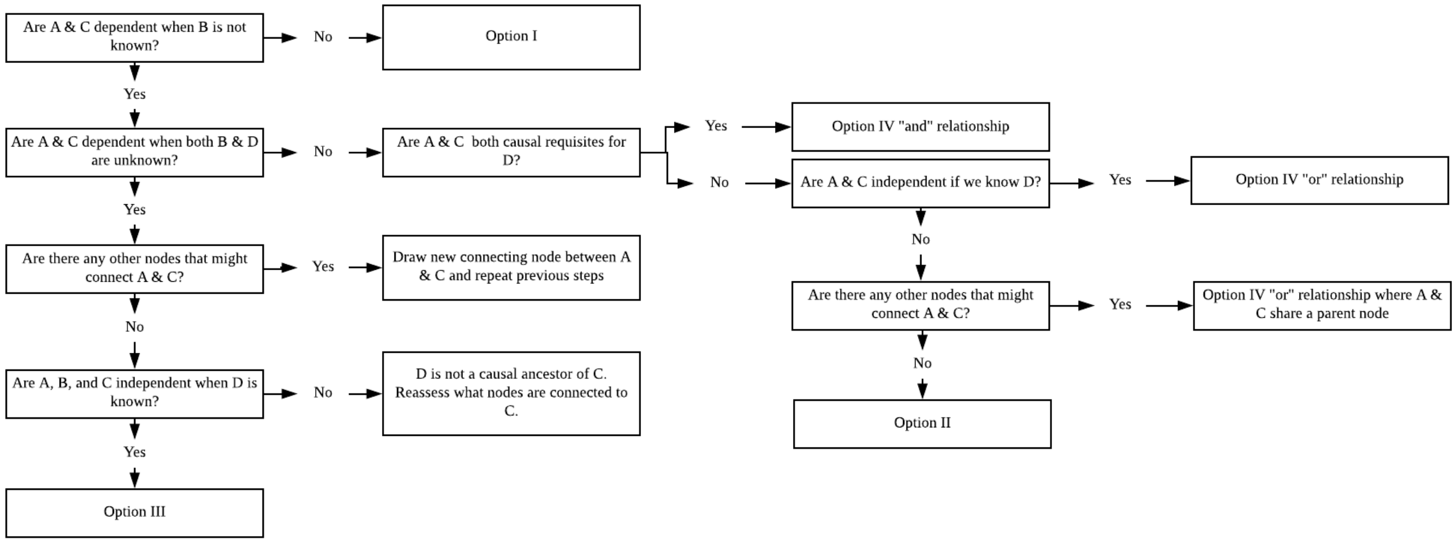


Figure 9 A graphical yes-no chart showing how one might structure the questioning of an expert for causal chain analysis. All steps are outlined in the text of sub-section 5.2

5.3 Introducing Multinomial Variables

As I mentioned earlier, all of my analysis has focused on binary variables for simplification, however the complex nature of the real world means that I would be remiss in not including a section on how to assess multinomial or continuous variables. For clarification, I have included a comparison of binomial, multinomial, and continuous systems in Table 3 to illustrate the differences between the systems. At its heart, the causal relationships between nodes will remain the same—Shipley's rules will hold as firmly for a non-binary system as a binary system. However, the types of questions asked to an expert will change—for instance, instead of asking an expert, "Can a population have food security while only having either high income or high energy access?" a researcher assessing a multinomial network would have to ask "Will the probability of a population having food stores exceeding 40 bushels of grain/person change if the population either only has a mean annual household income of $2,000 or only 20% of homes are electrified?" As you can see, number of degrees of relation increase, but the underlying causal logic between nodes remains the same—if the answer to the multinomial question is "yes" then our relationship will likely be a collider node with an "or" relationship between the parental nodes. The complexity derives itself from the researcher having to ask many iterations of this same question to determine the causal relationships for all the multinomial degrees that can exist between nodes. The continuous scale builds off the logic of the multinomial system, except, unlike the multinomial system which continues to have discrete values, the continuous system, as the name suggests, runs on a continuous scale. Again, the addition of multinomial and continuous systems does not affect the underlying causal relationships described in this paper—it only increases the precision with which we can examine nodal relationships and assign statistical values.

Table 3 Example nodes from previously used causal chains denoting the differences between binary, multinomial, and continuous systems.

System Type	System Designation	Energy Access	Education Availability	Income
Binary System	Yes/No	• Present • Absent	• Sufficient • Insufficient	• High • Low
Multinomial System	Multiple States	• No Access • Unreliable Access • Reliable Access	• Low availability • Decent availability • Widespread availability	• Below national avg. • Near national avg. • Above national avg.
Continuous System	Continuous	% reliability	% literacy	$/day

56

6. Case Study

This final section of my thesis will be applying my question tree to a real-world example created using the FEW Catalyst database. To create my causal chain, I used R code that had been created for the FEW Catalyst project and then personalized it to my project to create the comprehensive causal relationship web shown below (Fig. 10). This is a representation of the links asserted by the literature between different nodal topics—the thicker the arrow, the greater the body of evidence to support it. In the creation of my causal chain I will be using nodes that have abundant evidence linking them to each other as I believe that will provide a good proxy for an expert—instead of interviewing an expert, I will be "interviewing" the papers. These papers will be drawn from the database from which this figure was created.

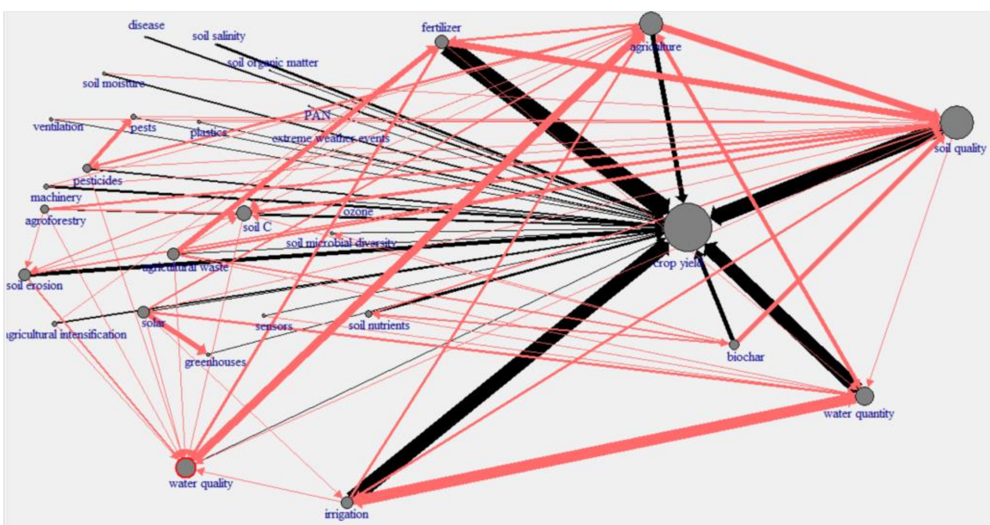


Figure 10. Causal network created using R code adapted from that of Dr. Ryan Calder. The direction of the arrow also indicates causal direction.

Just by looking over the network, one can quickly see there are only six to eight nodes with thick lines connecting them despite there being many smaller nodes. It is the purview of the FEW catalyst project to characterize all of these connections—I am only concerned with using my system to analyze a real-world literature-based model to see if the question tree is a workable schematic. I will therefore spend this section using my analysis technique to examine the links of five nodes—my hypothesized causal chain, drawn using the arrows shown in Fig. 10, is shown in Fig. 11 below.

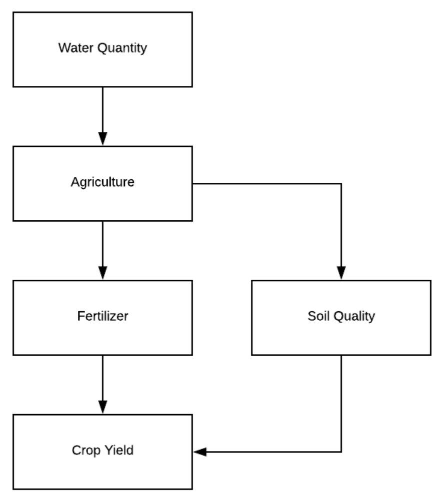


Figure 11. A simplified causal chain created using nodes from the R-programmed causal chart show in Fig. 10.

6.1 Creating the Causal Network

The above network's nodes were labeled exactly as they were in the FEW diagram from Fig. 10, however these nodal categories are too broad for us to conduct any meaningful causal analysis with the tools I've developed. For example, Agriculture is a very broad category with many different facets—it would be difficult assessing causal relationships between Agriculture and other nodes when it is unclear what exactly is meant by "Agriculture." I will therefore begin my case study by better defining the nodes in Figure 11 according to the literature used in the database to create the diagram. To do this, I first accessed the FEW database (in this case, it is a Mendeley database compiled by the researchers) and looked at all the papers under each of the relevant tags—for example, the papers linking Water Quantity with Agriculture were under the tag "water quantity → agriculture". I did this for each of the causal relationships to refine Fig. 11 into Fig. 12—I will now provide a short explanation of the changes that were made and why they were made.

Water Quantity was changed to Water Efficiency because the literature reflected that this is the factor that has the largest impact on most aspects of agricultural success (Fereres et. al., 2014). Agriculture was refined into Agricultural Intensity, as this matched well with the literature concerned with Water Efficiency and was mentioned multiple times in papers under the "Agriculture → Fertilizer" and "Agriculture → Soil Quality" tags. Integrated soil fertility management systems are a popular technique for increasing agricultural intensity while simultaneously maintaining soil health for future harvests (Agegnehu & Amede, 2017)—this requires the large usage of fertilizer, which naturally refines the Fertilizer node into Fertilizer Usage. However, heavily farmed soils, if not well maintained, experience nutrient exhaustion and soil erosion after years of constant use; additional pollution of heavy metals and pesticides often accompany many modern farming practices and further contribute to soil degradation (Kirchmann & Thorvaldsson, 2000). Such treatment of the soil can lead to substantial loss of soil organic matter—upwards of 60% in the top soils of regions farmed for over fifty years—and the eventual loss of the land as an effective grower of produce (Dalal & Chan, 2001). This seems to indicate that Soil Quality could be specified to Soil Degradation as this is one of the main effects of Agricultural Intensity and an impacting factor in determining Crop Yield. Our final causal chain therefore is the one shown in Fig. 13 below:

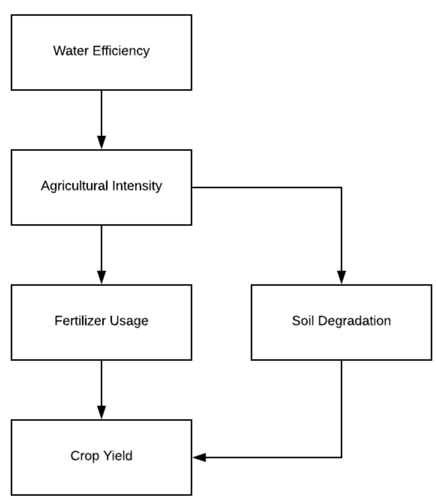


Figure 12 The updated Fig. 11 causal chain based off of our updated information provided from the database papers.

6.2 Examining Causal Relationships

Now that we have a refined causal network to work with, we can begin using the tools introduced in section four and the template introduced in section five to question our expert as we did in section five. Our goal is to test for the absence of arrows as well as question the arrows that already exist using the concept of d-separation. Our main areas of interest will be our Agricultural Intensity node, which is a forked node, and our Crop Yield node, which is a collider node; however, all the causal relationships will be examined. To clarify, we are trying to determine, via the advice of our expert, whether information in one node is co-dependent with the information in another node. This is not a question of making interventions, which posits that changing one node will have an effect on the other node—we are mentally testing the logical relationships among our nodes that may lead to intervention analysis at a later state of the experiment.

Water Efficiency's relationship to Agricultural Intensity is fairly straightforward—as was stated in the literature used to create the initial network, knowledge of a region's agricultural intensity often reflects how the region's water resources are used (Agegnehu & Amede, 2017). Knowledge of the latter gives us information of the former. We then will question the relationship between Water Efficiency and its descendant nodes of Fertilizer Usage and Soil Degradation. We would frame our questions to our expert as follows, "If we do not know anything about the population’s agricultural intensity, will knowledge of that population’s fertilizer usage or the region’s soil degradation be independent from information about the water efficiency of that population?” (this is similar to our first question in the template in Fig. 9, but the structure of this network is different). Starting with fertilizer usage, the FEW database does not provide any compelling papers showing that a population’s fertilizer usage gives any information on the water efficiency of their practices outside of its Agricultural Intensity link—an expert in a real scenario might disagree, but for our purposes we can conclude that the only causal pathway from Water Efficiency to Fertilizer Usage is through Agricultural Intensity. Our relationship between Water Efficiency and Soil Degradation is more interesting because knowledge about the region’s soil degradation does give us information about the efficiency of the local population’s water practices. Regions of high desertification (loss of topsoil to aerial and fluvial erosion) are often the result of water mismanagement leading to loss of soil matter to runoff (Lal, 2001). If we knew a region’s soil had recently predominantly been degraded by fluvial processes then, barring an errant climatic event, that would give us information on the efficiency of the local population at managing its water resources. We can therefore draw a line from Water Efficiency to Soil Degradation as there is a cause-effect relationship that exists here outside of the Agricultural Intensity node that also connects them. We finish our analysis of the Water Efficiency node by examining whether it information about Crop Yield is co-dependent with Water Efficiency when Agricultural Intensity, Fertilizer Usage, and Soil Degradation are unknown. Turning to the FEW database, the evidence we see in the papers is that Water Efficiency and Crop Yield are only related through the links already drawn—we therefore will assert that the two will be independent provided the intermediary nodes are unknown, so no new nodes or links will be drawn between the two.

We next examine our Agricultural Intensity node—the aim of our analysis, again, is to determine whether there are any missing links between this node and others to which it is not already connected. The drawn causal network, at this stage, is backed up by the literature and we are now utilizing expert opinion to answer questions that the literature cannot fully address. The only node that Agricultural Intensity does not share a link with is Crop Yield, which is shielded by the Fertilizer Usage and Soil Degradation nodes. We must therefore ask, if Fertilizer Usage and Soil Degradation are not known, are Crop Yield and Agricultural Intensity independent of each other? By turning to the literature, we see that this appears to not be the case—the two nodes are co-dependent in regard to technology and environmental well-being. Technological advancements have increased our crop productivity two and three-fold, depending on the crop, during the 20th century; at the same time, increased homogenization of crops, eutrophication of water resources, and exposure to animal waste damages environmental conditions and leaves crop yields more vulnerable to disease and contamination (Miller, 2008). This would draw two new nodes connecting Agricultural Intensity with Crop Yield—if we had an expert to pose questions to, we would continue following our Fig. 9 template and ask questions to determine the nature of the relationship between the new nodes, Agricultural Intensity, and Crop Yield (i.e. Are Agricultural Intensity and Crop Yield causal parents of Agricultural Technology? Are Agricultural Intensity and Crop Yield independent when we know our population's Agricultural Technology? Etc.). However, since we do not have an expert and the literature is pretty clear about the relationship the two new nodes have to Agricultural Intensity and Crop Yield, we will add them in as unidirectional chains without pursuing this line of questioning. This updated causal network can be seen in Figure 13 below:

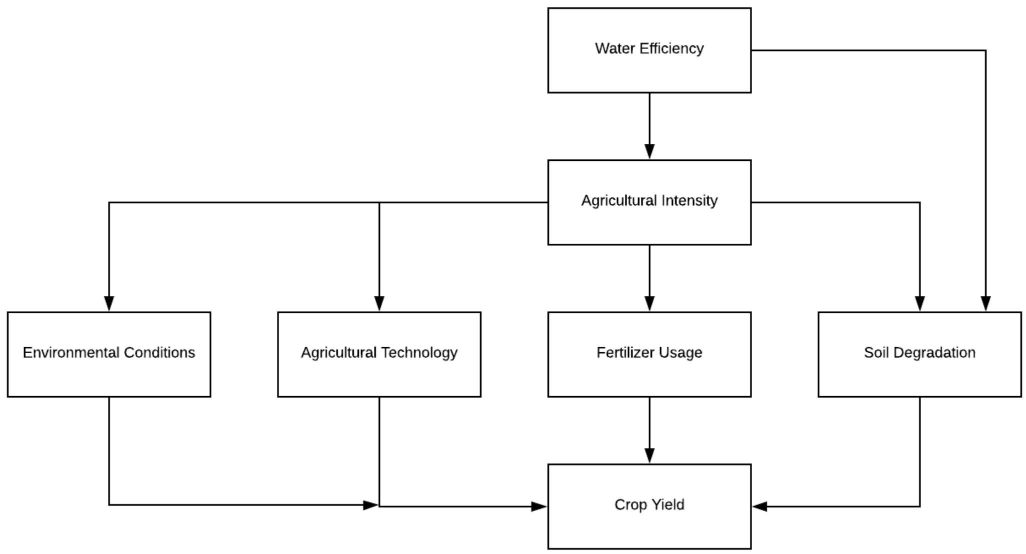


Figure 13 An updated Fig. 11 causal network after examining the causal relationships of Water Efficiency and Agricultural Intensity.

Our only collider node in this network is Crop Yield—we would normally question whether its parent nodes have an "and" relationship or an "or," however because Fertilizer Usage and Soil Degradation share a parent node, Agricultural Intensity, the type of collider node will have no impact on the d-separation when Crop Yield is conditioned upon (remember back to sub-section 4.2 for an explanation of this logic). The final part of our analysis is assessing the relationships of our newly added nodes to the rest of the network. Firstly, improved agricultural water harvesting technology leads to improved crop yields in dry seasons, so we can draw a link from Water Efficiency to Agricultural Technology creating another causal relation between

Water Efficiency and Crop Yield (Bouma et. al., 2015). Technological advancement of fertilizer over the past decades has increased dramatically (Yan et. al., 2008), so Agricultural Technology and Fertilizer Usage would be causally dependent even if we had no information on Agricultural Intensity—a link must therefore be drawn between these two nodes. Similarly, Fertilizer Usage can prevent nutrient depletion in soils (Spiers and McGill, 1979; Andreev et. al., 2018; Wortman et. al., 2017)), so a link should be drawn between Fertilizer Usage and Soil Degradation. Finally, soil erosion and degradation are a large component of the environmental degradation associated with agriculture (Miller, 2008), so a causal relationship will exist between Environmental

Conditions and Soil Degradation. Our final causal network, after updating in response to literature responses to our questions, will look like the one shown in Figure 14 below.

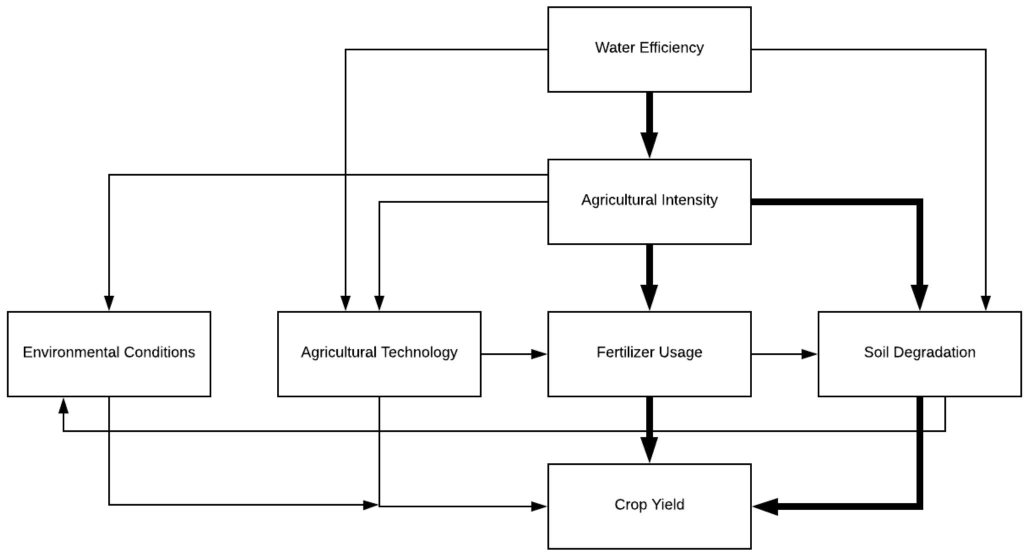


Figure 14 The final causal network after being updated with the responses from the FEW Catalyst literature search database. The original network links are shown in bold

The above analysis was done based off of the question template shown in Figure 9, but it shows the power of asking an expert simple questions such as, "Do Agricultural Intensity and Crop Yield vary independently when we have no knowledge on Fertilizer Usage or Soil Degradation?" Returning again and again to whether nodes within a network vary independently with unconnected nodes given a lack of information on intermediary variables will fill in the bulk of undrawn nodes. By following the question template, the researcher can then better define the nodal relations in the network—this part was difficult to express in this section due to the lack of an expert, thereby preventing us to follow each branch of the question template to completion. However, even without an expert we were able to drawn eight new links and two new nodes using our literature search alone—I am confident that the experience based opinions of an expert could be used to even better refine this network.

7. Conclusions

Logical analysis of causal networks can be a powerful tool for cross-field analysis of real-world research problems; however, this thesis was only able to explore these implications so far without applying the question template to a real expert. The entire purpose of expert elicitation is to fill knowledge gaps that have not made it through the rigorous controlled experiments that dominate literature. Although the case study in the last section was adequate for pedagogical purposes, I think it did not fully test whether the questioning template in Figure 9 will be an effective tool in framing questions posed to experts. For example, the language of the questions might be too technical (for example, there may be misunderstanding in asking an expert whether two variables vary if their connecting node is unknown). I would therefore like to see this template used in an actual research causal network to see if it is effective.

The second issue is the covering of multinomial and continuous variables, which I only mentioned briefly in subsection 5.3. I believe that this is the area that prevents more researchers from delving deeper into the causal relations between their network nodes—the task of questioning an expert about the multinomial or continuous causal nodes is an arduous and nuanced task that requires the iterative asking of many precise questions to an expert. Even if I provide a template for asserting causal relationships to researchers, I am skeptical whether it will be used for pragmatic reasons. I believe future work can be done on attaching statistical relationships to binomial, multinomial, and continuous variables to incorporate the tools I created into decision making models. This mathematization would require further research into the mechanics that Shipley and Pearl discuss in later chapters of their books.

Finally, this thesis's analysis only addresses monotonic causal relationships—that is, the relationship between variables is either positive or negative. Further work should be done for non-monotonic relationships—for example, the output of a kitchen could increase with the addition of 3 chefs to help the initial one fill orders, however the infamous "too many cooks in the kitchen" threshold could be passed with the addition of a 5^{th} chef causing the output of the kitchen to plummet. The way the causal relationship questions are currently structured is not sufficient to adequately characterize a non-monotonic causal relationship. Further logical work could be done to add questions to the existing template on how best to elicit and characterize experience-based opinions on non-monotonic relationships from experts.

The logical network assessment mechanisms I developed in this thesis have already been incorporated into an NSF proposal for a project seeking to conceptualize food-energy-water (FEW) interactions (Mark Borsuk, personal communication, October 25, 2018). As has been mentioned several times in this paper, although there is good characterization of individual interactions (i.e. water-food, water-energy etc.) the scope of such relationships is generally myopic and not indicative of the complexity of connections that exist between all three nodes. If approved, the project would have three goals:

1.) Integration of known or presumed FEW relations into a Bayesian network and use causal network analysis tools to explore alternate causal relationships for subsequent model building

2.) Advancement of cyber-infrastructure to extract information of model variables from remote sensing data.

3.) Testing of alternate models creating in goal 1 with data using AI and econometric causal tools to create an effective tool in decision making.

My work is being incorporated into goal one where the causal analysis tools that I have developed will be used to explore potential alternatives to conventionally accepted FEW models. This will then provide the basis for those alternative models to be explored so as to create a more robust decision-making model.

The next logical step after creating a causal relationship identification tool based on expert elicitation would to create guidelines for expert screening. Although experts are highly familiar with their respective fields and likely have the experience necessary to answer the questions I develop in this thesis, the risk of bias or exaggerated knowledge will always be a central weakness of this tool. Even if not intentional, an expert's bias may express itself in the information elicited due to the narrow context under which the expert worked within the field. For example, the field of human development research is a wide one and may require more than one expert to answer the questions outlined in my template fully. Future work should then focus on creating field specific expert screening guidelines and on adapting my template to being asked across a range of experts rather than one "master" expert. This would then provide a stronger basis for the implementation of my question template and facilitate the translation of my template into a real world research setting.

In summary, the work that I have done is but the first step in using logic to analyze the functionality of causal networks. There is still much work that can be done to quantize and refine this analytical process to benefit researchers interested in bridging the divides between their respective fields. However, I am confident that the Bridge Collaborative and organizations with similar end-goals will continue striving to integrate the rich databases of the world to answer the increasingly complex problems that the upcoming decades hold in store. To this end, I hope my work proves beneficial.

References

Agegnehu, G., & Amede, T. (2017). Integrated Soil Fertility and Plant Nutrient Management in Tropical Agro-Ecosystems: A Review. PEDOSPHERE, 27(4), 662–680. https://doi.org/10.1016/S1002-0160(17)60382-5

Andreev, N., Ronteltap, M., Boincean, B., & Lens, P. N. L. (2018). Lactic acid Fermentation of Human Excreta for Agricultural Application. JOURNAL OF ENVIRONMENTAL MANAGEMENT, 206, 890–900. https://doi.org/10.1016/j.jenvman.2017.11.072

Bazilian, M., Rogner, H., Howells, M., Hermann, S., Arent, D., Gielen, D., Steduto, P., Mueller, A., Komor, P., Tol, R., Yumkella, K. (2011). Considering the Energy, Water, and Food Nexus: Towards an Integrated Modelling Approach. ENERGY POLICY Vol. 39, Issue 12, pp. 7896-7906.

Bouma, J., Hegde, S., Lasage, R. (2015). Assessing the Returns to Water Harvesting: a Meta-Analysis. AGRICULTURAL WATER MANAGEMENT, 164(2016), pp. 100-109.

Borsuk, Mark. Personal Interview. October 25, 2018.

Box, J. F. (1980). R. A. Fisher and the Design of Experiments, 1922-1926. THE AMERICAN STATISTICIAN Vol. 34, No. 1, pp. 1-7.

Brandt, P., Ernst, A., Gralla, F., Luederitz, C., Lang, DJ., Newig, J., Reinert, F., Abson, DJ., von Wehrden, H. (2015) A Review of Transdisciplinary Research in Sustainability Science. ECOLOGICAL ECONOMICS, 92:1-5.

Bridge Collaborative. "Who We Are: Our Story." Available at: http://bridgecollaborativeglobal.org/who-we-are/our-story/. Accessed on: July 15, 2018.

Calder, Ryan. Personal Email. September 9, 2018.

Cassman, K. G., Dobermann, A., Walters, D. T., & Yang, H. (2003). Meeting Cereal Demand while Protecting atural Resources and Improving Environmental Quality. ANNUAL REVIEW OF ENVIRONMENT AND RESOURCES, 28, 315–358.

https://doi.org/10.1146/annurev.energy.28.040202.122858

Dalal, R. C., & Chan, K. Y. (2001). Soil Organic Matter in Rainfed Cropping Systems of the Australian Cereal Belt. AUSTRALIAN JOURNAL OF SOIL RESEARCH, 39(3), 435–464. https://doi.org/10.1071/SR99042

Fereres, E., Orgaz, F., Gonzalez-Dugo, V., Testi, L., & Villalobos, F. J. (2014). Balancing Crop Yield and Water Productivity Tradeoffs in Herbaceous and Woody Crops. FUNCTIONAL PLANT BIOLOGY, 41(10–11, SI), 1009–1018. https://doi.org/10.1071/FP14042

Fisher, R. A. (1935). The Design of Experiments. Edinburgh, Oliver & Boyd.

Godfray, H., Beddington, J., Crute, I., Haddad, L, Lawrence, D., Muir, J., Pretty, J., Robinson, S., Thomas, S., Toulmn, C. (2010). Food Security: The Challenge of Feeding 9 Billion People. SCIENCE EXPRESS, 10.1126/science.1185383, pp. 2-3.

Haines A, Kovats RS, Campbell-Lendrum D, Corvalán C. (2006). Climate Change and Human Health: Impacts, Vulnerability and Public Health. PUBLIC HEALTH, 120(7), 585-96.

Horrigan, L., Lawrence, R.S., Walker, P. (2002). How Sustainable Agriculture can Address the Environmental and Human Health Harms of Industrial Agriculture. ENVIRONMENTAL HEALTH PERSPECTIVE, 110, 445–456.

Kirchmann, H., & Thorvaldsson, G. (2000). Challenging Targets for Future Agriculture. EUROPEAN JOURNAL OF AGRONOMY, 12(3–4), 145–161. https://doi.org/10.1016/S1161-0301(99)00053-2

Kempthorpe, O. (1979). The Design and Analysis of Experiments. Huntington, NY, Robert E. Krieger.

Kendall, M. G., Stuart, A. (1983). The Advanced Theory of Statistics. London, Charles Griffin.

Lal, R. (2008). Soils and Sustainable Agriculture. A Review. AGRONOMIC SUSTAINABILITY DEVELOPMENT, 28, 57–64, pp 63.

Leopold, D. (2005). Guide to Fall Colors in Upstate New York. SUNY-ESF Syracuse, Department of Environmental and Forest Biology.

Maxwell, SL, Fuller, RA, Brooks, TM, Watson, JE. (2016). Biodiversity: The Ravages of Guns, Nets and Bulldozers. NATURE, 536(7615), 143-5.

Miller, F. (2008). After 10,000 years of Agriculture, whither Agronomy? AGRONOMY JOURNAL, 100(1), 22-34.

National Human Genome Research Institute. (2016). An Overview of the Human Genome Project. Available at: https://www.genome.gov/12011238/an-overview-of-the-human-genome-project/. Accessed on: July 15, 2018.

Nkhata, R. (2014). Does Irrigation have an Impact on Food Security and Poverty? Evidence from Bwanje Valley Irrigation Scheme in Malawi. International Food Policy Research Institute, Working Paper 04, 9-10.

OECD. (2001). The DAC Guidelines: Strategies for Sustainable Development. Paris: OECD, p. 26-28.

Pearl, J. (2000). Causality. Cambridge University Press, Cambridge, UK.

PAGE (Partnership on Action for Green Economy). (2016). Integrated Planning and Sustainable Development: Challenges and Opportunities. New York: United Nations Development Program, 18-19.

Petruney, T. (2016). Resource Package for Integrated Development. Washington, DC: FHI 360, pp.5.

Shipley, B. (2016) Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference with R. Cambridge: University Printing House. Second edition.

Stavi, I., Bel, G., & Zaady, E. (2016). Soil Functions and Ecosystem Services in Conventional, Conservation, and Integrated Agricultural Systems. A review. AGRONOMY FOR SUSTAINABLE DEVELOPMENT, 36(2). https://doi.org/10.1007/s13593-016-0368-8

Tallis H, Kreis K, Olander L, Ringler C et al. 2017. Bridge Collaborative Practitioner's Guide: Principles and Guidance for Cross-sector Action Planning and Evidence Evaluation. Washington DC: The Nature Conservancy. Available at: http://bridgecollaborativeglobal.org/wp-content/uploads/2018/02/Practitioners_Guide_Final_2.pdf Accessed on: July 29, 2018.

Vaneeckhaute, C., Ghekiere, G., Michels, E., Vanrolleghem, P. A., Tack, F. M. G., & Meers, E. (2014). Assessing Nutrient Use Efficiency and Environmental Pressure of Macronutrients in Biobased Mineral Fertilizers: a Review of Recent Advances and Best Practices at Field Scale. In D. Sparks (Ed.), ADVANCES IN AGRONOMY, 128, 137–180. https://doi.org/10.1016/B978-0-12-802139-2.00004-4

Whitmee, S.,Haines, A., Beyrer, C., Boltz, F., Capon, A., de Souza Dias, B., Ezeh, A. Frumkin, H., Gong, P., Head, P., Horton, R., Mace, G., Marten, R. Myers, S., Nishtar, S., Osofsky, S., Pattanayak, S., Pongsiri, M., Romanelli, C., Soucat, A., Vega, J., Yach, D. (2015). Safeguarding Human Health in the Anthropocene Epoch: Report of The Rockefeller Foundation–Lancet Commission on Planetary Health." THE LANCET, 386: 1973-1974.

Wortman, S. E., Holmes, A. A., Miernicki, E., Knoche, K., & Pittelkow, C. M. (2017). First-Season Crop Yield Response to Organic Soil Amendments: A Meta-Analysis. AGRONOMY JOURNAL, 109(4), 1210–1217. https://doi.org/10.2134/agronj2016.10.0627

Yan, X., Jin, J., He, P., Liang, M. (2008). Recent Advances on the Technologies to Increase Fertilizer Use Efficiency. AGRICULTURAL SCIENCES IN CHINA, 7(4), 469-479.