Probabilistic Argument Maps for Intelligence Analysis: Completed Capabilities

Robert Schrag[*], Joan McIntyre[+], Melonie Richey[$], Kathryn Laskey[%], Edward Wright[*], Robert Kerr[*], Robert Johnson[*], Bryan Ware[*], Robert Hoffman[&]
[*]Haystax Technology, McLean, VA and Las Vegas, NV (Schrag)
[$]Camber Corporation, Fairfax, VA
[+]Innovative Analytics and Training, Washington, D.C.
[%]George Mason University, Fairfax, VA
[&]Florida Institute for Human-Machine Cognition, Pensacola, FL

Abstract

Intelligence analysts are tasked to produce well-reasoned, transparent arguments with justified likelihood assessments for plausible outcomes regarding past, present, or future situations. Traditional argument maps help to structure reasoning but afford no computational support for probabilistic judgments. We automatically generate Bayesian networks from argument map specifications to compute probabilities for every argument map node. Resulting analytical products are operational, in that (e.g.) analysts or their decision making customers can interactively explore different combinations of analytical assumptions.

1   Introduction

In intelligence analysis, argument mapping [1] presents a problem-solving framework built around key elements of the intelligence issue being addressed, makes analytic reasoning shortfalls and information gaps more visible, prompts consideration of both supporting and refuting evidence mitigating confirmation bias [5], allows for comparison of multiple hypotheses, and translates easily into standard written formats with bottom line up front and supporting reasoning organized logically.

Haystax has developed a probabilistic argument mapping framework called FUSION[1]. Faced with the challenge of operationalizing subject matter experts' (SMEs') policy-guided reasoning about person trustworthiness in a comprehensive risk model [10], we first developed CARBON, now one of many models supported by the FUSION framework. The CARBON domain's high volume (hundreds) of policy statements and need for SMEs both to understand the model and to author its elements inspired us to develop and apply a technical approach that enhances argument maps with SME-accessible probabilistic reasoning.

We developed the FUSION framework having recognized the general need for and latent power of a probabilistic argument mapping approach—across many application areas, including our own software product and service line. In the last three years of building FUSION, we have identified and resolved subtle representation and reasoning issues in a coherent, integrated computational framework with APIs and UIs at multiple levels, including a top-level GUI. We recently began addressing the specific requirements of argumentation for intelligence analysis, appealing initially as a driving use case to the CIA's Iraq retaliation scenario [4], where Iraq might respond to US forces' bombing of its intelligence headquarters by conducting major, minor, or no terror attacks, given limited evidence about Saddam Hussein's disposition and public statements, Iraq's historical responses, and the status of Iraq's national security apparatus.

Intelligence analysts traditionally develop their judgments about the likelihood of a given situation's outcome using ad hoc methods that consider probabilistic notions but do not necessarily implement mathematically sound probabilistic reasoning. Bayesian network inference propagates beliefs in all directions—not just up from leaf nodes towards root hypotheses, but also back down[2], in a process that is generally too complex for any human to follow, completely, beyond small pedagogical examples. For a very large class of intelligence analysis problems [3], this belief propagation is very fast—much faster than needed to support graphical user interface (GUI) interaction. Once propagation has settled, observed probabilities are all consistent. Clicking around an argument map FUSION model in the GUI, analysts can observe which of their input likelihood assessments have what effects on computed beliefs, for all nodes.

FUSION models[3] are intuitively simple yet technically sophisticated. We have developed software [13] to convert probabilistic argument maps into corresponding Bayesian networks (BNs). The conversion software recognizes a pattern of types of argument map links that are incident on a given statement and constructs a conditional probability table (CPT) for the corresponding BN node (a random variable representing the statement's truth or falsity) to implement appropriate reasoning. The SME—here, the

---

[1] SMALL-CAPS typeface distinguishes tools and frameworks.

[2] Note the finding set in Figure 2, e.g.

[3] A FUSION model is a probabilistic argument map (a computational model of an analytical argument). We use the terms "argument" and "model" interchangeably.

analyst—thus works with argument maps (as if on a dashboard), and BN mechanics and minutiae[4] all remain conveniently "under the hood."

2 Analyst's structured FUSION model

Given the Iraq retaliation scenario description from [4], our intelligence analyst followed a structured argumentation process drawing loosely on analysis of competing hypotheses (ACH)[5], to draft a purely textual argument. A fragment appears as Table 1. The process Note the analyst's grouping of evidence statements into five categories—past reactions, capability, initial responses, and political and sychological motivations—which we take to be exemplary of predictive intelligence questions as an analytical problem class. Under the structured process, an analyst asserts first hypotheses, then evidence statements (formulated as hypothesis-neutral), then rates each evidence statement for consistency with and relevance to each hypothesis.

Table 1. FUSION upgrades the practice of intelligence analysis from textual arguments with ad hoc (analyst-intuited) likelihood reasoning to graphical argument maps with principled probabilistic reasoning. Shown is a fragment of our analyst's specification for the Iraq retaliation scenario, addressing just one outcome hypothesis.

Hypothesis 2 --Iraq will sponsor some minor terrorist actions in the Middle East—*Refuting with High Uncertainty*
Past reaction to similar events—*Refuting with High uncertainty*
- Absence of terrorist offensive during the 1991 Gulf War—*Refuting, Credibility High, Relevance Low*
- Iraq responded with low scale response to "provocations" by Iran—*Supporting, Credibility High, Relevance Low*
Capability to respond – military and intelligence capabilities—*Supporting with High Uncertainty*
- Small network of agents which could be used to attack US interests in the Middle East and Europe—*Supporting, Credibility Medium, Relevance Low*
- Network has only been used to go after Iraqi dissidents—*Refuting, Credibility Medium, Relevance Low*
Initial responses to the bombing—*Supporting with High Uncertainty*
- Saddam public statement of intent not to retaliate—*Refuting, Credibility Low; Relevance Low*
- Increase in frequency/length of monitored Iraqi agent radio broadcasts—*Supporting, Credibility Medium, Relevance Low*
- Iraqi embassies instructed to take increased security precautions—*Supporting, Credibility Medium, Relevance Low*

Political motivations driving response decision—*Refuting with High Uncertainty*
- Assumption that Saddam would not want to provoke another US attack—*Refuting, Validity Medium, Relevance Medium*
Implication of Saddam's psychological makeup for a decision on responding—*Supporting with High Uncertainty*
- Assumption: Failure to retaliate would be a loss of face for Saddam—*Supporting, Validity Low, Relevance Low*

In Table 1, likelihood reasoning is captured as follows.

- *Relevance* captures the degree to which a posited statement supports or refutes the (sub-)hypothesis statement to which it is connected. Our analyst includes explicit headings for the evidence categories.
- *Credibility* captures the degree to which an evidence statement is considered believable, based on attributes of an associated source report.
- *Validity* captures the analyst's assessment of the legitimacy of a posited assumption statement—one for which sourced evidence is unavailable or unexpected.
- *Uncertainty* captures the analyst's (presumably, ad hoc) roll-up accounting for the other three likelihood notions above, respecting argument structure.

Figure 1 is a screenshot[6] of our encoding of the analyst's argument as a Fusion model, which includes outcome hypothesis nodes (circled yellow), evidence category nodes (circled green), and evidence nodes (right of category nodes), plus additional nodes for the sake of logic (IraqRetailiatesWithTerror) and organization (IraqChoosesTerror). The former is true if either "TerrorAttacks" statement is true. The latter collects support from the four category nodes that in her model are the same for the two terror hypotheses, using indication strengths per her specification. For brevity, we've hidden all the evidence credibility and assumption validity nodes. We've set appropriate findings on all evidence, assumption credibility, and validity nodes. Hypothesis 2 (minor terror) has a computed belief of 17%, hypothesis 3 (major terror) 2%. By comparison, our analyst estimated a belief range of 20–45% for hypothesis 2. The traditional process rolls up likelihoods from evidence to hypotheses, normalizing to 1.0 across hypotheses. In contrast, Bayesian belief propagation is multi-directional, updating beliefs over an entire model. A version of this model addressing only Hypothesis 2 computes 23% belief—within the analyst's bounds.

---

[4] The conversion software creates auxiliary BN nodes for some link type patterns (e.g., MitigatedBy in [13]).

[5] ACH (see [4], chapter 8) is intended to induce a workflow enhancing the elicitation of hypotheses and evidence and to reduce biases towards any particular lines of reasoning. It elicits informal likelihoods but falls short of eliciting the conditional probabilities that are essential to true Bayesian reasoning. While some ACH tools do implement ad hoc likelihood combination methods, ACH itself has no integral probabilistic framework.

[6] With the GUI, a user can edit a model to add, delete, or change nodes or links, navigate to show or hide a displayed node's upstream, downstream, or neighbor nodes, find (per text search) and display a hidden node, select either bottom-to-top or right-to-left argument stream orientation, and explore different situations by entering (or clearing) BN "hard findings" that arbitrarily (often temporarily) state unequivocally that a given statement should be taken either as true or false. Upon a finding entry, FUSION performs BN belief propagation and updates the display.

For each node, the modeler specifies a full-sentence statement and chooses a short label for display on the node's GUI icon. The GUI will display the full statement on mouse-over or drill-down.
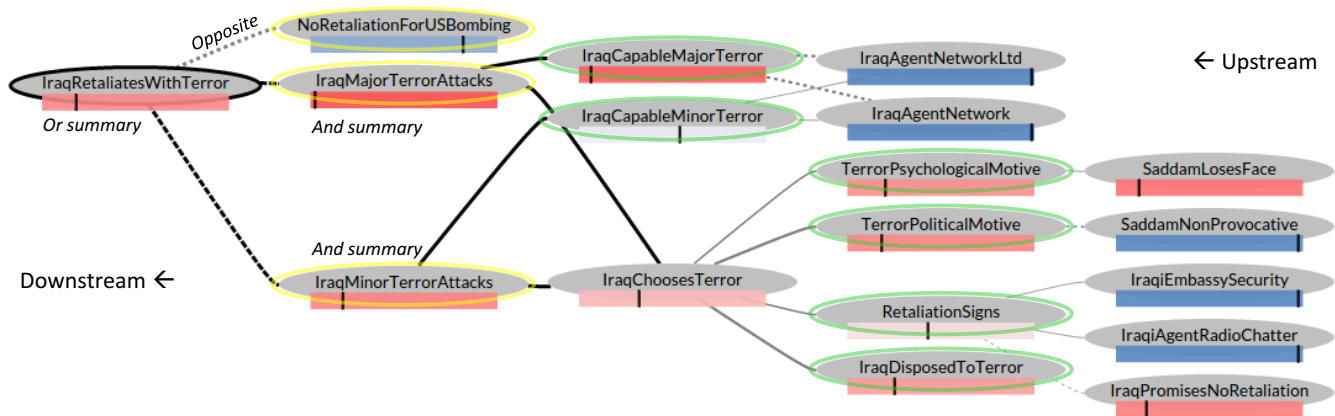
Figure 1. This screenshot from FUSION's model GUI exhibits our probabilistic argument mapping capability. Statement nodes are connected by positive (solid grey line) and negative (dashed grey line) indication links of various strengths (per line thicknesses). Argument flow (from evidence to outcomes) is from right to left—e.g., IraqChoosesTerror is strongly indicated by RetaliationSigns. Black lines inform logic summary nodes (annotated post-screenshot). Belief bars (bottom, each node) reflect belief level using both tick marks and colors (explained at Figure 2).

Note that the FUSION model is more compact than the full textual specification. The model mentions each statement only once. Besides uncluttering the modeling canvas, this convention helps enforce consistency. Consider that given assumption statements should carry the same truth values in a fair comparison of different hypotheses. So, we shouldn't assess NoRetaliationForUSBombing with IraqiAgentRadioChatter turned on and IraqMinorTerrorAttacks with IraqiAgentRadioChatter turned off. The GUI shows model state under one given set of assumption values at one time. When a user changes assumption values, computed beliefs displayed for all statements (including outcomes) are updated together.

Table 2. FUSION supports probabilistic argument map model link types (center column).

	IndicatedBy	
	CounterIndicatedBy	
Downstream[7] statement	MitigatedBy	Upstream statement(s)
	RelevantIf	
	OppositeOf	
	ImpliedByConjunction	
	ImpliedByDisjunction	

FUSION implements distinct varieties of relevance different from our analyst's usage that we have taken as FUSION indication strength.

- Indication relevance. Saddam's promise not to retaliate matters to our assessment of Iraq's choice/outcome not to retaliate only if he is not lying. Connecting the statement (not shown) impugning Saddam's veracity and his statement promising non-retaliation is a FUSION

RelevantIf link (not a standard IndicatedBy link), which serves to discount consideration of the promise when we believe Saddam to be dishonest in making it.[8]

- Temporal relevance, reflecting decay either in importance of a past event or in continuing reliability of a past state observation. In FUSION, an event's/observation's relevance decays per a user-specified half life [10].

3 Scientist's incentive-oriented FUSION model

Figure 2 is a screenshot of a model by one of our scientists emphasizing Saddam's incentives to act, considering the issues of maintaining diplomatic status, maintaining a face of strength with his public, and whether US might not expect retaliation (so harden defenses, likely foiling any attack) if he promises none. By setting a hard finding of false or true on the incentive-collecting node SaddamWins, we can examine computed beliefs (plotted in Figure 3) under Saddam's worst- and best-case scenarios. We see that Saddam is much more likely to have engaged in terror in a situation in which he loses than one in which he wins—so terror is not in his best interests. Figure 3 also plots beliefs for the situation in which there is no finding and the 50% prior probability on SaddamWins prevails.

---

[7] Per argument map convention, "downstream" is left, "upstream" right in the left-flowing argument maps of Figure 1 and Figure 2.

[8] A FUSION MitigatedBy link works symmetrically, discounting an influence when the mitigator is true.
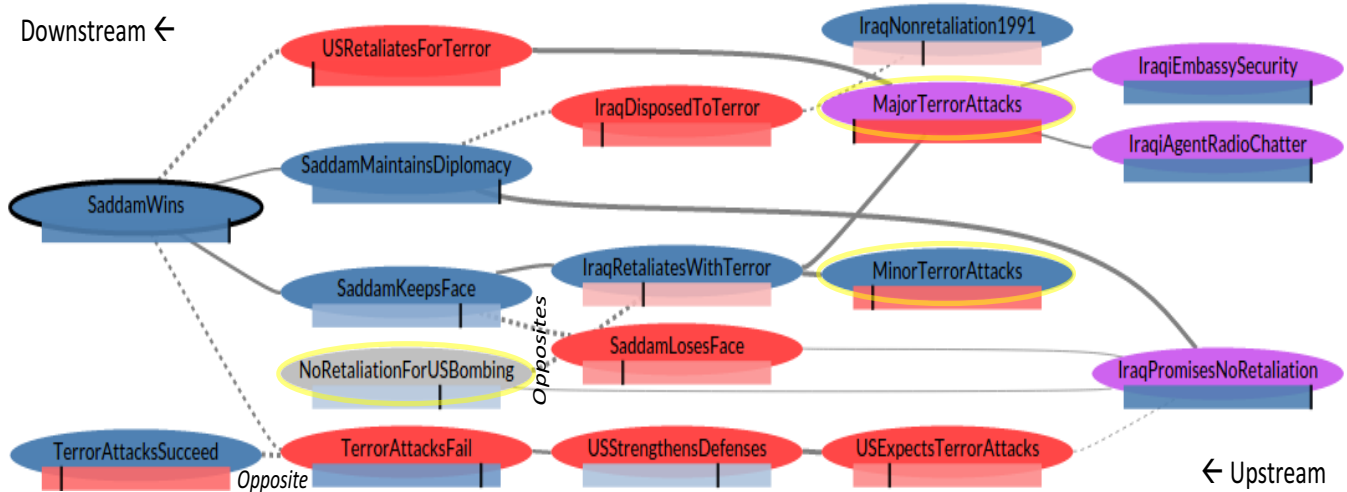
*Figure 2. Our scientist's model of the argument excerpted in Table 1 includes (with different names) many of the same statements as its counterpart in Figure 1, with many similar links. **SaddamWins** (hard finding false) captures Saddam's incentives to act or not. Node colors capture whether Saddam's disposition (or attitude) regarding a given statement is favorable (blue) or unfavorable (red). Fusion computes these dispositions from a single directive—to propagate the favorability of **SaddamWins** upstream (only), respecting link polarities. One node (gray, bottom left) is not touched by this propagation. Four nodes (purple, top right) have ambiguous status with respect to Saddam's disposition. Belief bars augment spatial tick marks with colors chosen to reflect the degree of concern a given statement would pose given a node's disposition. Lower belief (redder bar) poses less concern regarding a statement viewed unfavorably, more regarding one viewed favorably, higher belief (bluer bar) the reverse. Red/blue contrast thus draws attention to a statement that should be of concern to Saddam.*
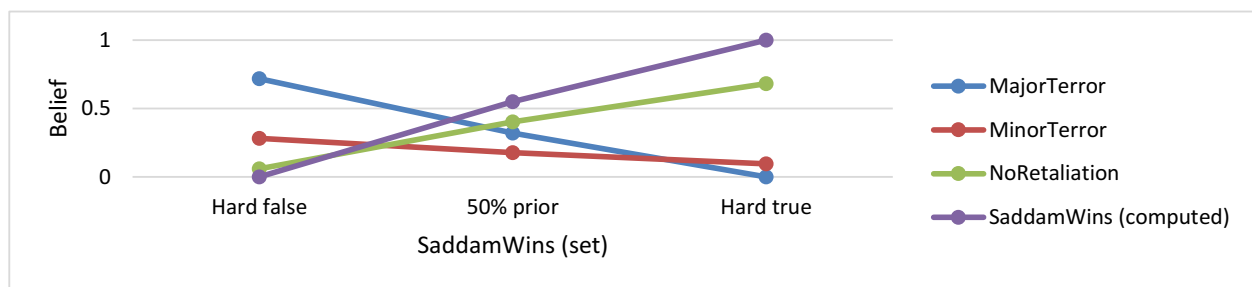

*Figure 3. Given a node (like **SaddamWins**) collecting an actor's incentives, FUSION can compute beliefs for outcomes under the actor's worst-case, best-case, and other scenarios.*

4 Related work

We believe FUSION's combination of argument maps and BNs to be unique.

Karvetski et al. [7] propose BN expert-facilitated BN development following an ACH-based protocol. The BN adaptation is intended to overcome ACH weaknesses associated with informal treatment of uncertainty. Modeling the 1984 Rajneeshee bioterror attack themselves, the authors envision how practicing analysts might productively collaborate in an ACH style. Appealing to standard elicitation techniques, they elicit (from each supposed analyst) 118 coarse-grained probability assessments to complete the CPTs for 14 nodes (one ternary[9], 13 binary) with a combined total of 19 parents.[10] Like us, they eschew duplicate nodes (which unnecessarily complicate probability reasoning). A corresponding FUSION model would require no more than 27 indication polarity-and-strength assessments. We have designed FUSION to eliminate the need for a knowledge representation and reasoning specialist (a BN expert) to facilitate knowledge acquisition, so that analysts can build argument models themselves.

---

[9] A FUSION model would factor the ternary node into three binary ones, over which it would apply an xOr Logic constraint.

[10] They limit model size by factoring 12 outcome hypotheses into three outcome aspects—who, where, and why. FUSION can support this approach.

We agree with these authors' statements below regarding ACH uninformed by mathematically sound probability reasoning. These statements also apply to argument maps so uninformed.

The measures of consistency, relevance, and credibility are poorly defined and elicited unreliably. This allows for highly subjective and unique interpretations among analysts. For example, the consistency measure should answer a well-defined question such as, "Given hypothesis H, how likely are we to see evidence e?" rather than the question "How consistent are hypothesis H and evidence e?" Emphasizing the direction of the question can clear up confusion between interpretations. (p213)

The Senate Select Committee on Intelligence [12] criticized the pre-war assessments of weapons of mass destruction (WMDs) in Iraq for the tendency of analysts to consider uncertainty only at each separate stage of reasoning rather than over the whole chain of reasoning. Heuer [4] was not unaware of this problem, but he offered limited advice on the subject for ACH users. (p215)

Karvetski et al. acknowledge the value of argument mapping as an elicitation tool, but do not go so far as to integrate it with their product BN in an argument mapping tool, as FUSION does.

Probabilistic abstract argumentation frameworks [8] assign probabilities to nodes and links locally and use these to compute probabilities globally. These frameworks generally assume conditional independence among all nodes, so do not accommodate conditional probabilities and cannot meaningfully capture causality or other rich relationships.[11]

Markov logic networks, similarly, do not naturally accommodate conditional probabilities, so representing causality is cumbersome [6]. They are notoriously hard to build directly. More often, they are applied in a machine learning setting. They are attractive in that the only parameters to be specified are weights on logical formulas. We implemented a propositional Markov logic interpreter to experiment with the Iraq Retaliation scenario but were unable to engineer the necessary fundamental conditional dependence relationships (without going all the way to implement BNs, less efficiently, in this framework).

---

[11] We take the recent dissertation of Li [8] to be representative of the state of the art. Li proposes framework extensions to accommodate conditional independence—after having briefly mentioned BNs, chooses Nilsson's probabilistic logic [9] as a foil, and dismisses the lot: *The standard uncertainty management approaches as mentioned are unable to propagate uncertainty through argument evaluation; i.e., given uncertainty associated with arguments, these approaches cannot propagate the uncertainty to uncertainty about which arguments are justified.* (p12) FUSION does this now.

6 Conclusion

Probabilistic argument maps are applicable wherever traditional argument maps are. By choosing Logic statement nodes and/or by applying hard findings to upstream-most non-Logic nodes, a probabilistic argument map can be rendered entirely deterministic. Thus, FUSION models are a superset of standard argument maps. Probabilistic reasoning offers a powerful alternative to crisp logical reasoning, accounting naturally for uncertainty about evidence or influences. FUSION also probabilistically enhances nonmonotonic defeat and relevance reasoning— via its MitigatedBy and RelevantIf link types. We continue to develop and apply the FUSION framework [11].

References

[1] CIA Directorate of Intelligence, "A Tradecraft Primer: The Use of Argument Mapping," *Tradecraft Review* 3(1), Kent Center for Analytic Tradecraft, Sherman Kent School, 2006.

[2] Kevin Burns, "Bayesian HELP: Assisting Inferences in All-Source Intelligence," *Cognitive Assistance in Government,* Papers from the AAAI 2015 Fall Symposium, 7–13.

[3] Kevin Burns, "Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking (ICArUS): Phase 2 Challenge Problem Design and Test Specification," MITRE Technical Report, MTR 149412, McLean, VA, 2014.

[4] Richards J. Heuer, Jr., *Psychology of Intelligence Analysis,* Central Intelligence Agency Historical Document. https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/psychology-of-intelligence-analysis (Posted: Mar 16, 2007 01:52 PM. Last Updated: Jun 26, 2013 08:05 AM.)

[5] P. E. Lehner, L. Adelman, B. A. Cheikes, and M. J. Brown, "Confirmation bias in complex analyses," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans,* 38(3), 584–592, 2008.

[6] Dominik Jain, "Knowledge Engineering with Markov Logic Networks: A Review," in *DKB 2011: Proceedings of the Third Workshop on Dynamics of Knowledge and Belief,* 2011.

[7] Christopher W. Karvetski, Kenneth C. Olson, Donald T. Gantz, and Glenn A. Cross, "Structuring and analyzing competing hypotheses with Bayesian networks for intelligence analysis," *EURO J Decis Process* (2013) 1:205–231.

[8] Hengfei Li, *Probabilistic Argumentation,* A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of the University of Aberdeen Department of Computing Science, 2015.

[9] Nils J. Nilsson, "Probabilistic Logic," *Artificial intelligence,* 28(1):71–87, 1986.

[10] Robert Schrag, Edward Wright, Robert Kerr, and Bryan Ware, "Processing Events in Probabilistic Risk Assessment," *9th International Conference on Semantic Technologies for Intelligence, Defense, and Security (STIDS),* 2014.

[11] Robert Schrag, Edward Wright, Robert Kerr, Robert Johnson, Bryan Ware, Joan McIntyre, Melonie Richey, Kathryn Laskey, and Robert Hoffman, "Probabilistic Argument Maps for Intelligence Analysis: Capabilities Underway," *16th Workshop on Computational Models of Natural Argument,* 2016.

[12] United States Senate Select Committee on Intelligence, "Report on the U.S. Intelligence, Community's Prewar Intelligence Assessments on Iraq," One Hundred Eighth Congress,

Second Session. U.S. Government Printing Office, Washington, DC, 2004.

[13] Edward Wright, Robert Schrag, Robert Kerr, and Bryan Ware, "Automating the Construction of Indicator-Hypothesis Bayesian Networks from Qualitative Specifications," Haystax Technology technical report, 2015, https://labs.haystax.com/wp-content/uploads/2016/06/BMAW15-160303-update.pdf.