SHOULD WE SIMULATE MENTAL MODELS TO ASSESS WHETHER THEY AGREE?

Eric A. Lavin

Department of Computer Science
Northern Illinois University
1425 W. Lincoln Hwy
DeKalb, IL 60115, USA
ealavin@gmail.com

Philippe J. Giabbanelli

Department of Computer Science
Northern Illinois University
1425 W. Lincoln Hwy
DeKalb, IL 60115, USA
giabba@cs.niu.edu

Andrew T. Stefanik

Department of Computer Science
Northern Illinois University
1425 W. Lincoln Hwy
DeKalb, IL 60115, USA
stefanik35@gmail.com

Steven A. Gray

Department of Community Sustainability
Michigan State University
480 Wilson Road Room 131
East Lansing, MI 48824, USA
grayste1@msu.edu

Robert Arlinghaus

Division of Integrative Fisheries Management, Humboldt-Universität zu Berlin, Unter den Linden 6
and
Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310
Berlin, Germany
arlinghaus@igb-berlin.de

ABSTRACT

Modeling approaches can support policy coherence by capturing the logistics of an intervention involving multiple individuals, or by identifying goals and preferences of each individual. An important intermediate step is to identify agreement among individuals. This may be achieved through intensive qualitative methods such as interviews, or by automatically comparing models. Current comparisons are limited as they either assess whether individuals think of the same factors, or see the same causal connections between factors. Systems science suggests that, to test whether individuals really share a paradigm, we should mobilize their whole models. Instead of comparing their whole models through multiple simulation scenarios, we suggested using network centrality. We performed experiments on mental models from 264 participants in the context of fishery management. Our results suggest that if stakeholder groups agree on the central factors (per Katz centrality), they also tend to agree on simulation outcomes and thus share a paradigm.

Keywords: Fuzzy Cognitive Maps, Ecological Management, Mental Models, Participatory Modeling.

SpringSim-ANSS, 2018 April 15-18, Baltimore, MD, USA; ©2018 Society for Modeling & Simulation International (SCS)

1 INTRODUCTION

Interventions on complex systems may require coordination among individuals to identify a common purpose, establish suitable and efficient scenarios, and finally implement them. This is grounded in the theoretical concept of *policy coherence*, emphasizing the need to share outcomes across sectors and ensure consistency (Dube et al. 2014). A coherent policy is articulated around a specific issue, and requires that individuals work together through forms ranging from cooperation (sharing a mission but keeping resources separate) to collaboration (sharing a mission and resources). Modeling approaches have been proposed to capture how individuals can work together, given their different capacities and associated time scales (e.g. organizational and political cycles) (Giabbanelli and Crutzen 2017). This assumes that a shared mission has already been identified, and focuses on the logistics of accomplishing it. At the other end of the spectrum, various modeling approaches are routinely used to elicit individual goals and preferences (Gray et al. 2015b). For instance, Integrated Assessment (IA) is a popular approach for socio-environmental problems, which seeks stakeholder participation and involves modeling approaches such as Systems Dynamics, Bayesian Networks, Agent-Based Models, and Fuzzy Cognitive Maps (Giabbanelli, Gray, and Aminpour 2017, (Letcher) et al. 2013). Between eliciting individual preferences and managing the logistics when pursuing a common mission, there is one essential step: identifying shared preferences such that we can agree on the mission. Qualitative methods are a popular approach to identify shared preferences (e.g., using depth interviews or focus groups), but they can be resource intensive (e.g., training staff to run and analyze sessions, compensating participants for their time) and involve non-trivial logistics given the need for some synchronicity (e.g., staff and one or multiple participants interact at the same time). In contrast, we are interested in fully automatic, scalable quantitative methods. The objective is to identify agreements using little resources (hence making it scalable), and allowing for asynchronous participation. Achieving scalability and supporting asynchronous participation are essential to support the participation of many stakeholders, residing in any place. That is, it supports the uptake of citizen science in a more inclusive form of decision-making process, which has known benefits in areas such as conservation planning (Gray et al. 2017).

To automatize the full process leading to the identification of shared preferences, we need the first part of the process to already be automatic. That is, among all possible processes to elicit individual preferences, we are limited to those which are scalable and asynchronous. Our contribution consists of finding similarities among stakeholders by analyzing individual preferences. Finding similarities among individuals is related to the concept of a *cultural model*. Cultures are about the social behavior and norms of societies: they tell us what is shared across individuals. In Axelrod's landmark model for "The Dissemination of Culture", individuals have a list of features which can take different values (e.g., what hat they like to wear). To eliciting such models automatically, participants can simply email their features and associated values. Then, commonalities are straightforwardly defined as having the same value on the same feature. However, this approach only considers isolated facts: it does not tell us *why* individuals have a given value, or the consequences of having it. For instance, one individual may have a black hat because it is a religious requirement, and another may have a black hat because it is a fashion trend. The same observation is thus rooted in different causal antecedents, and the 'commonality' is superficial. We can go one level higher by (i) eliciting a network from each individual, and (ii) comparing whether the same nodes also share antecedents and consequences. This is known as the problem of comparing causal maps (Giabbanelli and Tawfik 2017). However, research on systems thinking shows that this is still not a complete comparison. Indeed, Donella Meadows and others (Meadows 2008, Malhi et al. 2009) have posited that isolated nodes are around the lowest level, while links (i.e. antecedents and consequences) are slightly higher, but the highest level is the *paradigm* which mobilizes the entire system. Assessing links is thus insufficient to really conclude that individuals will agree on a course of action. For instance, given a cycle with four links, individuals could agree on three links but disagree on the fourth one, and that is enough to create complete disagreement (e.g., by turning a reinforcing loop into a balancing one). To mobilize a whole system, we can extract mental models from individuals, and then simulate a variety of 'what-if' questions on the models. If they generally provide the same simulation output, then we can conclude that the individuals themselves are in broad agreement.

Several studies have shown that mental models suitable for simulations could be extracted in the form of Fuzzy Cognitive Maps (FCMs) (Amirkhani et al. 2017, Jetter and Kok 2014), which offers some scalability (Giabbanelli 2014) and fully allows for asynchronous participation using platforms such as `MentalModeler` (Gray et al. 2014). However, running several 'what-if' scenarios over hundreds or more FCMs is computationally intensive (Lavin and Giabbanelli 2017) and requires to identify the 'right' scenarios. In this paper, we ask the fundamental question: can we tell that simulation models (in the form of FCMs) are going to agree only based on their structure, without actually running simulations? This would provide a more accurate assessment of similarity between individuals than current structural approaches (only comparing edges instead of using the whole model), and would present performance advantages over conducting many simulations. We propose to use centrality metrics, as they mobilize the whole network structure. To identify which centrality metric(s) are suitable, we examine whether agreeing on centrality is correlated with agreeing on which simulation outcomes are important across different scenarios. Our experiments are performed using real-world data, collected for the socio-ecological problem of fishery management. The dataset provides a large sample of 264 mental models from different types of stakeholders.

In short, our main contributions are as follows:

1. We propose a new approach to investigate agreements between mental models, by using network centrality instead of individual edges (known to be potentially uninformative) and without resorting to a large number of simulations (causing a computation burden).
2. We demonstrate our approach on a large real-world dataset with different types of stakeholders.
3. We show that one centrality metric correlates with simulation outcomes, thus a structural analysis may suffice to comprehensively assess agreements between models instead of running simulations.

The remainder of this paper is organized as follows. In section 2, we provide a technical background on Fuzzy Cognitive Maps and network centrality. In section 3, we explain our simulation set-up by summarizing the context of our dataset, testing its quality using extreme settings, and identifying four what-if scenarios. Our simulation scripts are hosted on the Open Science Framework at https://osf.io/qyujt/, within 'Comparing FCMs'. In section 4, we report our results on the correlation between simulation outcomes and selected centrality metrics. The implication of these results for participatory modeling and policy coherence are discussed in section 5. We conclude by summarizing our achievements in identifying agreements across participants more comprehensively than previous structural approaches, without extensive simulations.

2 BACKGROUND

2.1 Fuzzy Cognitive Maps

2.1.1 Why Are Fuzzy Cognitive Maps Used As Mental Models?

A 'mental model' broadly captures an individual's perspective on a specific topic. We thus need to refine what goes into a 'perspective'. On the one hand, individuals could provide statements backed by abundant evidence (Dalkey 1969), such as using meta-reviews to justify why they think that a system works in a certain way. On the other hand, a perspective may involve information of any type or quality, which would include sensorial imprints (e.g., "do you like the smell of this fish?"). The perspective that we seek to capture in a mental model lies in between: we are interested in meanings and understandings rather than sensorial imprints, but we do not require that they are supported by evidence (Giabbanelli and Crutzen 2014). For instance, we can ask a fisherman how he thinks that fishing is going to impact a fish population, and the answer neither requires evidence nor should it focus on senses.

Research in cognitive sciences has long been concerned with how individuals form and organize knowledge in their memory, and how it can be elicited. To capture an individual's mental model, we thus need to tap into the specific part of their memory where the knowledge of interest is held. The intention is often

to tap into the *semantic* memory, which holds the individual's conceptualization of the world (Binder and Desai 2011), rather than in the episodic memory, which is about specific events. This results in a mental model where individuals generalize and abstract from their experiences, rather than focusing on a specific instance. The practical question is thus: how do we elicit a perspective through an individual's semantic memory? And, as a corollary: to which extent is the resulting mental model an artifact of the elicitation process? McNeese and Ayoub suggested that, if the elicitation method closely aligns with how human knowledge is internally represented, then (i) its elicitation would be easier, (ii) the representations would tie into and open up spontaneous access to associated knowledge within memory, and (iii) the knowledge of one person can be compared with another to search for invariance (McNeese and Ayoub 2011). Since semantic memory provides functional *relationships* between objects, some elicitation methods (e.g., mind mapping, and semantic networks) focus on capturing interrelatedness. They thus produce *networks* or *maps*. In other words, if mental models are published and shared in the form of maps, it owns to the fact that we seek to capture semantic memory whose structure is network-based.

A relationship from A to B implies that changes in A *may* have effects on B. This may be characterized with parameters such as intensity of the change, time, and previous history. For instance, merely suggesting to someone not to fish a juvenile pike may have limited to no effect. In contrast, strongly insisting on the illegal or damaging effects of fishing non-harvestable fish may have a larger effect. This effect may not be immediate (i.e., a lag may occur), as one may finish fishing but change strategy the next time. Such dynamic relationships can be *represented* using System Dynamics (SD). However, eliciting SD models can be challenging, as individuals may not be readily able to provide a clear number or to precisely estimate the duration of a time lag. This may require in depth interviews or focus groups to create graphical functions (Verigin et al. 2016). As we are interested in a more scalable process (see Introduction), we may thus focus on only acquiring some parameters characterizing a relationship. A Fuzzy Cognitive Map (FCM) represents neither time nor history, as only the current value of a concept can influence the next one. The focus is on capturing the intensity of the change, in a way that is intuitive to participants, and can easily be done through online questionnaires. Rather than being forced to provide numbers that they may not have thought of, participants evaluate the strength of a relationship using linguistic variables such as 'Low', 'Medium', 'Strong'. Then, fuzzy logic is used to associate a membership function to these variables, and the defuzzification process eventually results in a number used by the model (Amirkhani et al. 2017, Giabbanelli 2014).

2.1.2 How Do Fuzzy Cognitive Maps Work?

Definition 1. *At a given iteration t, a Fuzzy Cognitive Map $F^t = (V^t, E, f)$ is formed of:*

1. *A set V^t of n nodes, representing concepts of the system such as states or entities. The values can change through the simulation, and V_i^t indicates the value of node i at iteration t. Values are bounded in the range [0, 1] where 0 indicates the absence of a concept and 1 indicates that it is maximal.*
2. *A set of weighted, directed edges E representing causal relationships. Weights and directionality are typically represented using an adjacency matrix A, where $A_{i,j}$ is the weight of the link from i to j. If $A_{i,j}$ is negative then an increase in i causes a decrease in j. Conversely, if $A_{i,j}$ is positive then an increase in i causes a decrease in j.*
3. *A clipping function f, also known as transfer function. As nodes' values are updated, the clipping function ensures that the result remains in the range [0, 1].*

Rather than representing units, an FCM models the extent to which concepts are present. For example, in an ecological model, a node could stand for the density of fish in a given space, where 0 means no fish and 1 means a maximal density. That value cannot go beyond 1 since it is maximal, and the density cannot be negative either. Concept values are updated over discrete iterations, which do not intend to correspond to

real-world time steps. The update uses an inference function, where the next value of each concept V_i^{t+1} is computed in Equation 1 based on (i) its current value V_i^t, (ii) the value of connected concepts V_j^t, (iii) the strength of these connections $A_{j,i}$, and (iv) the clipping function f.

$$V_i^{t+1} = f\left(V_i^t + \sum_{j=1, j\neq i} V_j^t \times A_{j,i}\right), \tag{1}$$

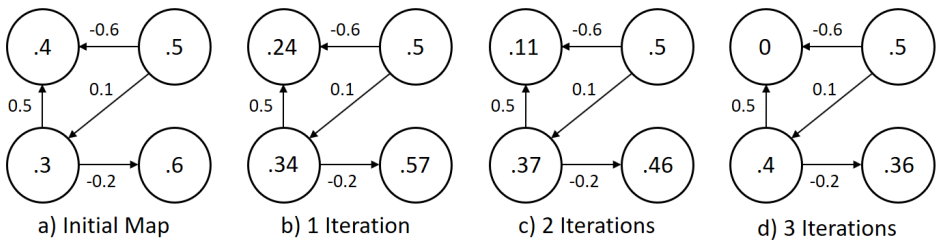

Figure 1: An FCM with four nodes, whose values are updated using equation 1 over 3 iterations. Nodes are usually labeled with concept names: we instead display their values to illustrate the operations.

The update is repeated (Figure 1) until a subset $S \subseteq V$ stabilizes. For instance, if the intention of the mental model is to summarize how various factors impact the long-term presence of fish in a lake, then we update the model until the 'fish' concept changes by less than a very small ε. Formally, the simulation ends when:

$$|V_i(t+1) - V_i(t)| \leq \varepsilon, \forall i \in S, \tag{2}$$

Since there is no randomness in equations (1) and (2), an FCM does not need repeated simulation runs. The choice for f and ε is normally left to the modelers. The function f has to be monotonic (to preserve the order of nodes' values) and a sigmoidal function is recommended for planning scenarios (Tsadiras 2008). The hyperbolic tangent *tanh* is widely used (Groumpos and Stylios 2000, Mago et al. 2013), and will be employed here as well. The value of ε is rarely reported, as it is assumed to be a small constant. We use $\varepsilon = .001$. The objective of the elicitation process is to obtain the set of nodes and edges from participants. Initial values for the nodes are set when applying the model to a specific scenario/what-if question.

FCMs are implemented by several libraries and software such as `MentalModeler` (Gray et al. 2014). For intensive simulations, we will use the Python library introduced in (Lavin and Giabbanelli 2017).

2.2 Network Centrality

From a network perspective, an FCM is composed of nodes and edges. Since we do not look at changes in nodes' values, we use V rather than V^t to refer to the set of nodes. Intuitively, the centrality of a node represents its 'importance'. The convention is that the higher the centrality, the more important the node.

Definition 2. *Node centrality c induces at least a semi-order on the set of nodes, allowing to conclude that $x \in V$ is at least as central as $y \in V$ with respect to centrality c if $c(x) \geq c(y)$. Generally, the difference or ratio of two centrality values cannot be interpreted as a quantification of how much more central one node is than the other (Koschützki et al. 2005).*

Different centrality indices can produce very different ranges of values: they can be between 0 and 1 when counting what *fraction* of some network dynamics involves a node, or significantly larger than 1 when counting the *total* contribution of a node. In addition, Definition 2 emphasizes that the relative difference between the values may not be meaningful: values show whether a node is more or less central than another one, but not *how much* more or less. To address these two potential issues when dealing with multiple centrality indices, we normalize all of them into rankings. That is, the node with the highest centrality value now takes the value of 1, the next highest takes the value of 2, and so on.

Centrality indices can evaluate different aspects to decide that a node is 'important'. Reachability indices focus on the cost that it takes for a node to reach others, while flow indices assess how much traffic may pass through a node, and feedback indices compute a node's importance based on its neighbor's importance (Koschützki et al. 2005). As we cannot presume that one type of centrality will best correlate with simulation outcomes, we use different types, all implemented using the `NetworkX` library in Python:

1. For reachability indices, we use *closeness centrality*, which is the inverse of the sum of the shortest distances between a node and all others.
2. For flow indices, we use *betweenness centrality* and its variant *load centrality* (Goh, Kahng, and Kim 2001). Betweenness is the fraction of shortest paths between all pairs that go through a given node. Load is different from betweenness as shown in (Brandes 2008): it considers that traffic splits equally between neighbors, thus leading to a different approximation of how traffic accumulates.
3. For feedback indices, we use *Katz centrality*. A node accumulates influence through incoming neighbors, and to a lesser extent through further away nodes. A damping factor α adjusts the influence of nodes based on the distance. We use the default value $\alpha = 0.1$ in `NetworkX`.
4. For a local index which depends only on a node's neighbor, we use the *degree* defined as the total number of edges incidents to a node.

3 SIMULATION SET-UP

3.1 Dataset

The northern pike is an important recreational fishery in Germany (Arlinghaus et al. 2008). Recreational fishermen using a rod and a line are referred to as *anglers*. The dynamics of the northern pike population depend on several socio-ecological factors, and anglers are one of the agents of ecological change. Appreciating this role through an ecological understanding helps to identify preferred management policies. To examine the anglers' ecological understanding, we recently collected their FCMs (Gray et al. 2015a). The process started by mailing a solicitation to all 461 angling clubs in the German state of Lower Saxony, of which 41 agreed to participate, and 17 were retained after criteria such as interest to improve their knowledge of pike fisheries management. Up to 25 anglers were accepted as volunteers from each club. Within a club, we report separately on club managers, water managers, and 'regular' anglers (i.e. without a management role for the water or the club) as their role may come with a different level of familiarity regarding socio-ecological dynamics. After excluding participants who provided missing data, we had a total of 136 anglers, 79 club managers, and 32 water managers. To compare their maps with those produced by biologically trained academic experts, we also collected the maps of 17 "volunteers comprised of researchers, post-docs and PhD students employed [at] a research institute specializing in fish ecology and biology, and an inland fisheries institute" (Gray et al. 2015a). We thus had a total of 264 FCMs coming from four categories of participants. Note that the previous study reported a higher sample size as it also counted participants who were excluded from the analysis. While the *cognitive* map (i.e. the network without its fuzzy quantification) of our problem was not included here due to space limitations, it can be found in (Gray et al. 2015a).

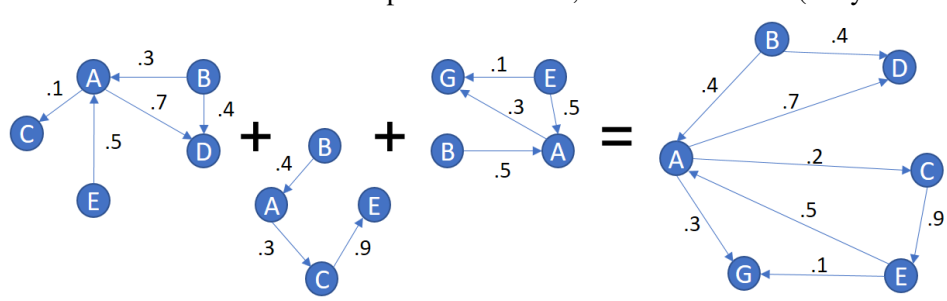

Figure 2: Aggregation of three individual FCMs into one, using a weighted average.

If participants created FCMs without constraints, they may use different terms to refer to the same concept (e.g., number of pikes, pike population, amount of pikes), (Giabbanelli and Tawfik 2017). We set a model boundary and limited linguistic variability using a list of standardized terms, which allowed to assess whether a concept in one map is the same as in another map. Our standardization involved "independent focus groups with anglers and fishery experts about the biology and fisheries ecology of pike" (Gray et al. 2015a), resulting in a list of 19 concepts (Table 1). We created *aggregate* FCMs representing the mental model held by a group of stakeholders, in addition to the 264 *individual* FCMs. A node or edge appears in the aggregate if it was expressed by any stakeholder. An edge's value is obtained using a weighted average for the value of the edge in all maps that contained it (Figure 2). The FCM of a group thus uses the FCMs of all of its stakeholders. This leads to 5 aggregate maps: all anglers, all water managers, all club managers, all experts, and all non-experts (i.e., anglers, water managers, and club managers) (Gray et al. 2015a).

Table 1: The values of our 19 concepts were set depending on the what-if scenario or validation test.

	Scenarios				**Validation**	
Concepts	1	2	3	4	1	2
Spawning Grounds	.6	.35			.6	.1
Angling Pressure	.75			.5	.2	.9
Refuges	.25	.5	.25		.6	.1
Pike Population (adults over legal size limit)	.31				.9	
Stocked Pike (adults over legal size limit)	.21					.4
Stocked Pike (under legal size limit)	.25		.5	.25		
Baitfish	.15					
Other Predatory Fish	.15					.8
Algae	.18				.5	.1
Depth of Water	.3					
Wild Pike (under legal size limit)	.5					
Emergent Riparian Plants	.45				.5	.1
Benthic Invertebrates	.25					.1
Zooplankton	.27				.5	.1
Submerged Aquatic Plants	.45				.5	.1
Cormorant	.1					.8
Plant Nutrients	.18					
Turbidity of Water	.2				.1	.8
Surface Area of a Body of Water	.75					

The focus of these FCMs is the relationship between the ecosystem (including anglers) and the population of pike over the legal size-limit (meaning anglers can legally keep the fish if caught). We briefly detail the main concepts of the system. Pike maintain their population by reproducing in a spawning habitat, which requires vegetative mats to keeps eggs in a better oxygenated area (Casselman and Lewis 1996). Thus, by keeping highly vegetative areas, we expect an increase in juvenile pike (that cannot be fished for legally), which ultimately increases the amount of harvestable pike. This vegetation further increases the amount of zooplankton and feed for other (small) fish, which become part of the food supply for pike. However, pikes are not the top of the food chain in their ecosystem. Other fish threaten pike and their offspring, and these predators also benefit from having more small fish. Furthermore, pikes of all sizes are also threatened by cormorants, a group of aquatic birds. On the human side, strong fishing habits can greatly reduce the amount of fish in a lake. To compensate, the pike population is boosted through a practice called *stocking* which consists of placing pike of legal size and below into the environment.

3.2 Validation

Our dataset only serves to test our proposed methods to correlate simulation outcomes and centrality. Nonetheless, it is important to test its validity, as an improper data collection process may introduce bias in our results. A broad test used for complex models (Verigin et al. 2016), and particularly for FCMs (Giabbanelli, Torsney-Weir, and Mago 2012), consists of establishing extreme scenarios where there is no doubt in what the correct outcome should be. Starting with a very large population for pike (set to 0.9) and an abundance of other species in the system (other predatory fish = cormorant = 0.8), our two selected scenarios should clearly increase (validation 1) or decrease (validation 2) the total pike population of legal size.

In the validation expected to *increase* the pike population, we use conditions optimal for the breeding of pike as studied by Casselman (Casselman and Lewis 1996). We want high amounts of refuge and spawning areas (spawning grounds = refuge = .6). We want the water to be clean and not deep, thus we want lower values for the turbidity (turbidity = .1). Along with that, we need a sufficient amount of plants for food and cover (algae = emergent plants = submergent plants = zooplankton = .5). We keep plant nutrients low (plant nutrients = .2) and avoid overfishing through a low angling pressure (angling pressure = .2).

In the validation expected to *decrease* the pike population, we set poor conditions for pike to thrive by following the opposite logic of the previous validation. We make it difficult for pike to reproduce (spawning grounds = refuge = .1). The water will be dirty and deep with a large amount of predators (cormorant = predatory fish = turbidity =.8). Food and cover are scarce (algae = emergent plants = submergent plants = zooplankton = invertebrates =.1). Finally, the lake will be significantly used for fishing, thus it has a large amount of anglers fishing (angling pressure = .9) and involves stocking (stocked adult pike = .6).

As both validation scenarios use an expert understanding of the system, we apply them to the expert maps. That is, we expect that the expert FCMs resulting from our data collection process will make the same conclusions as derived from the scientific literature. Two-third of the expert maps (64.70%) produced the expected increase in scenario 1, and the expected decrease in scenario 2. This validates our data collection process, which was applied to all stakeholders. We note that, while the aggregate FCM of all experts is obtained through a standard process, it predicts an increase in pike in both scenarios (although less so in scenario 2). We will thus report our results on the aggregate FCMs as well as across all individual FCMs.

3.3 What-if Scenarios

We implemented four what-if scenarios, representing possible actions to manage the pike population: increase spawning habitat (1), increase refuge (2), increase juvenile stocking (3), and decrease angling pressure (4). Scenarios 1–2 have been proposed in different studies as possible actions to manage a stocked lake for pike fishing (Pierce et al. 1995, Casselman and Lewis 1996) while scenario 3 was discussed in our previous work (Gray et al. 2015a). A study found that a high angling pressure could reduce the fish population by half, thus scenario 4 considers alleviating this pressure (Mosindy et al. 1987).

All scenarios are applied in the context of an 'average' initial situation. In short, we consider a lake with a medium amount of fish per hectare (which does not attract many predators), a supportive environment for growth of the pike population, and a heavy fishing presence. Specifically, we first examined the characteristics of lakes that are stocked with pike for fishing. The average density of pike in a stocked lake is 31 per hectare, thus we set our value to .31 (Pierce et al. 1995). Research further shows the average stocking proportionality to be 21 per hectare, therefore we set it to .21 as well. We assume that there are more pike below the legal fishing size (value of .5), since larger pike are exploited at 2 to 9 times greater rates than smaller pike (Pierce et al. 1995). To create a supportive environment, we use clear water (low turbidity value of .2), a medium depth of .4 and a non-dense amount of submergent and emergent vegetation (both at .45) (Casselman and Lewis 1996). Studies mention the spawning areas for pike to be in shallower water (Casselman and Lewis 1996) which competes with human desire for shallow areas near the shore. These

areas provide refuge from predators and a stable breeding ground. This competition leads us to assume the pike only have a moderate amount of space for spawning and refuge. Thus, we will set these values to .25 and .35 respectively. For surface area, we place the value as .75. Since the dynamics of food supply lead to having less predators than prey in an ecosystem, we set other predatory fish to .15 and cormorant to .1. Angling pressure was set to .75 to represent the large amount of fishing in the areas.

4 RESULTS

Our goal is to assess whether there is a correlation between some centrality metric(s) and simulation outcomes, across all four scenarios and stakeholder groups. That is, nodes are ranked by both centrality and simulation outcomes, and we compute the correlation between these two rankings. To avoid being sensitive to outliers, we use the robust Theil–Sen estimator to fit a line (also known as Kendall robust line-fit method). Instead of the typical [-1, 1] range for correlations, the estimator returns 1 when there is a perfect fit *regardless* of whether it is a positive or negative correlation. Negative or null values indicate a poor fit. Figure 3 exemplifies the Theil–Sen estimator. All results are available on https://osf.io/qyujt/.

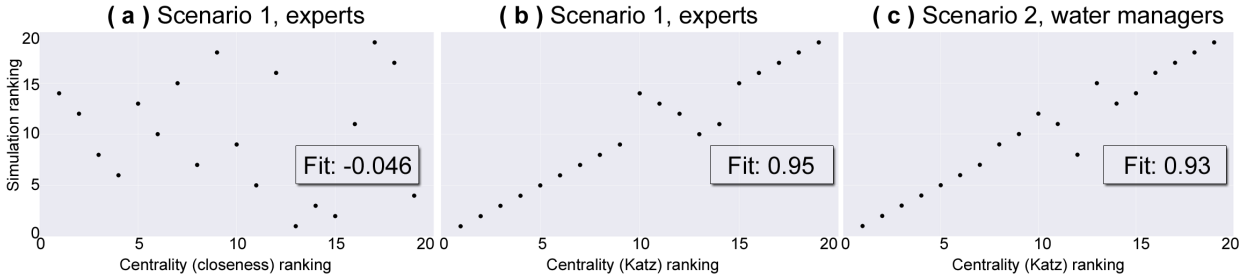

Figure 3: Correlation between centrality and simulation ranking for different scenarios and groups.

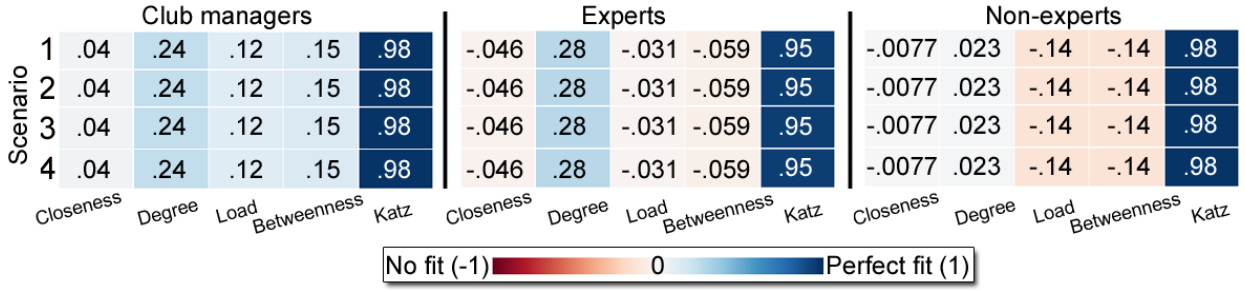

Figure 4: Across all scenarios and groups of stakeholders, only Katz centrality had a very high fit.

Sample results on the aggregate maps are shown in Figure 4. The results show that Katz centrality has a very high fit across all scenarios and aggregate maps (from .93 for water managers to 1 for anglers), whereas all other centrality metrics have a very poor fit. As we emphasized in section 3.2 that the aggregate map can behave differently from most of its underlying FCMs, we also assessed the fit across individual maps for each scenario. Results for scenario 2 are shown in Figure 5, and results for other scenarios (available online) support the same two observations. First, the four centrality metrics that showed no correlation at the *aggregate* level also show no correlation at the *individual* level. Second, in contrast to the aggregate level, Katz centrality only exhibits a (small) correlation when applied to the expert group, and behaves the same as other metrics for other groups. The implication is that Katz centrality is suitable to compare aggregate maps, but may only be used to compare individual maps in very specific cases. A possible explanation is as follows. Aggregate maps are highly connected, since an edge exists if a single participant used it. These many connections provide abundant feedback, which the Katz centrality seeks to measure. In contrast, individual maps (and particularly maps of non-experts) have much less feedback.

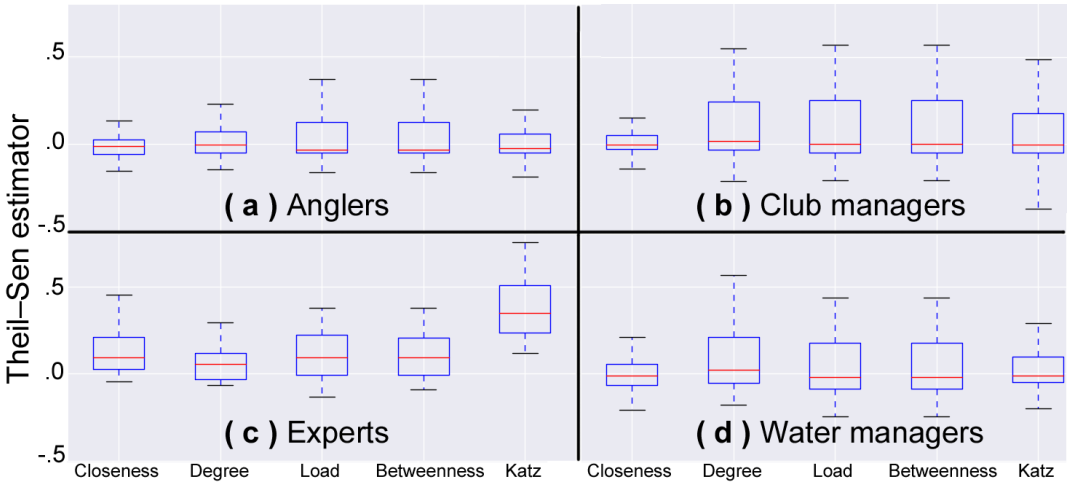

Figure 5: Fit across all stakeholders in each group, for scenario 2.

5 DISCUSSION

To know whether two simulation models agree, we can check whether they yield similar outputs across an extensive and computationally demanding set of simulations. We examined whether analyzing the *structure* of the models was sufficient, instead of simulating them. Our application context is participatory modeling, where the mental models of stakeholders are represented as Fuzzy Cognitive Maps (FCMs) and we need to identify whether stakeholders agree before pursuing a course of action. As systems science emphasizes that contrasting systems in terms of their independent nodes or edges does not fully capture their dynamics, we instead searched for an analysis that involves the whole model structure. Since FCMs are built on networks, we used different measures of network centrality to take into account reachability, flow, or feedback. We then assessed whether the importance of nodes as judged by the centrality tended to agree with their importance in terms of simulation outputs. Using the mental models of 264 stakeholders from four groups, we found an almost perfect agreement when applying Katz centrality at the level of a group of stakeholders, and a moderate agreement when applying it to expert stakeholders only. Only 19 set terms could be used in individual models, which may create more agreement than if we allowed for complete linguistic variability.

The implication is that if two groups of stakeholders view nodes as being as central (in terms of Katz), then that they would make the same conclusion as to what happens to these nodes across a broad range of scenarios. In practice, we can thus consider that agreeing on centrality is enough to conclude that two groups share a paradigm. The implication to compare individual models is less clear, as centrality was only moderately useful, and for a single category of participants. Future research would thus need to assess whether other feedback-based measures of centrality provide better results at the individual level. Alternatively, one may allow to perform a few simulations (instead of none), and use the time series of node values to supplement the purely topological information used by a centrality measure. This is becoming possible as dynamic centrality measures on temporal network data are increasingly available.

6 CONCLUSION

We demonstrated that we do not need to simulate the mental models of groups to know whether they agree, as network centrality can sufficiently characterize these models. However, further research is needed to adequately characterize the mental models of individuals, which still require simulations to be compared.

ACKNOWLEDGMENTS

The authors received financial support from the College of Liberal Arts & Sciences and the Department of Computer Science at Northern Illinois University, and through the Besatzfisch project funded by the

German Federal Ministry of Education and Research. We thank Devin Cole, Jinwuk Lee, Aditya Sabbineni, Eric Warner, Johanna Hilsberg, and Andrew McFall for help on collecting and analyzing the mental maps.

REFERENCES

Amirkhani, A. et al. 2017. "A review of fuzzy cognitive maps in medicine: Taxonomy, methods, and applications". *Computer Methods and Programs in Biomedicine* vol. 142 (Sup. C), pp. 129 – 145.

Arlinghaus, R. et al. 2008. "Understanding the heterogeneity of recreational anglers across an urban–rural gradient in a metropolitan area (Berlin, Germany), with implications for fisheries management". *Fisheries Research* vol. 92 (1), pp. 53 – 62.

Binder, J. R., and R. H. Desai. 2011. "The neurobiology of semantic memory". *Trends in Cognitive Sciences* vol. 15 (11), pp. 527–536.

Brandes, U. 2008. "On variants of shortest-path betweenness centrality and their generic computation". *Social Networks* vol. 30 (2), pp. 136 – 145.

Casselman, J. M., and C. A. Lewis. 1996. "Habitat requirements of northern pike (Essoxlucius)". *Canadian Journal of Fisheries and Aquatic Sciences* vol. 53 (S1), pp. 161–174.

Norman C Dalkey 1969. "The Delphi method: an experimental study of group opinion".

Dube, L. et al. 2014. "From policy coherence to 21st century convergence: a whole-of-society paradigm of human and economic development". *Ann. N. Y. Acad. Sci.* vol. 1331 (1), pp. 201–215.

Giabbanelli, P. J. 2014, May. "Modelling the spatial and social dynamics of insurgency". *Security Informatics* vol. 3 (1), pp. 2.

Giabbanelli, P. J., and R. Crutzen. 2014. "Creating groups with similar expected behavioural response in randomized controlled trials: a fuzzy cognitive map approach". *BMC Medical Research Methodology* vol. 14 (1), pp. 130.

Giabbanelli, P. J., and R. Crutzen. 2017. "Using Agent-Based Models to Develop Public Policy about Food Behaviours". *Comput Math Methods Med* vol. 2017, pp. 5742629.

Giabbanelli, P. J., S. A. Gray, and P. Aminpour. 2017. "Combining fuzzy cognitive maps with agent-based modeling: Frameworks and pitfalls of a powerful hybrid modeling approach to understand human-environment interactions". *Environ Modell Softw* vol. 95 (Sup. C), pp. 320 – 325.

Giabbanelli, P. J., and A. A. Tawfik. 2017, Sep. "Overcoming the PBL Assessment Challenge: Design and Development of the Incremental Thesaurus for Assessing Causal Maps (ITACM)". *Technology, Knowledge and Learning*.

Giabbanelli, P. J., T. Torsney-Weir, and V. K. Mago. 2012. "A fuzzy cognitive map of the psychosocial determinants of obesity". *Applied Soft Computing* vol. 12 (12), pp. 3711 – 3724.

Goh, K.-I., B. Kahng, and D. Kim. 2001. "Universal Behavior of Load Distribution in Scale-Free Networks". *Phys. Rev. Lett.* vol. 87, pp. 278701.

Gray, S. et al. 2014. "Modeling with citizen scientists: Using community- based modeling tools to develop citizen science projects". *Proc Int Cong Env Modell Softw* vol. Stream F.

Gray, S. et al. 2015a. "The structure and function of angler mental models about fish population ecology". *Journal of Outdoor Recreation and Tourism* vol. 12 (Sup. C), pp. 1 – 13.

Gray, S. et al. 2015b. "Using fuzzy cognitive mapping as a participatory approach to analyze change, preferred states, and perceived resilience of social-ecological systems". *Ecol Soc* vol. 20 (2), pp. 11.

Gray, S. et al. 2017. "Combining participatory modelling and citizen science to support volunteer conservation action". *Biological Conservation* vol. 208 (Sup. C), pp. 76 – 86.

Groumpos, P. P., and C. D. Stylios. 2000. "Modelling supervisory control systems using fuzzy cognitive maps". *Chaos, Solitons & Fractals* vol. 11 (1–3), pp. 329 – 336.

Jetter, A. J., and K. Kok. 2014. "Fuzzy Cognitive Maps for futures studies—A methodological assessment of concepts and methods". *Futures* vol. 61 (Sup. C), pp. 45 – 57.

Koschützki, D. et al. 2005. *Centrality Indices*, pp. 16–61. Springer Berlin Heidelberg.

Lavin, E. A., and P. J. Giabbanelli. 2017. "Analyzing and Simplifying Model Uncertainty in Fuzzy Cognitive Maps". *Proceedings of the 2017 Winter Simulation Conference*.

(Letcher), R. A. K. et al. 2013. "Selecting among five common modelling approaches for integrated environmental assessment and management". *Environ Modell Softw* vol. 47 (Sup. C), pp. 159 – 181.

Mago, V. K. et al. 2013. "Analyzing the impact of social factors on homelessness: a Fuzzy Cognitive Map approach". *BMC Medical Informatics and Decision Making* vol. 13 (1), pp. 94.

Malhi, L. et al. 2009. "Places to Intervene to Make Complex Food Systems More Healthy, Green, Fair, and Affordable". *Journal of Hunger & Environmental Nutrition* vol. 4 (3-4), pp. 466–476.

McNeese, M. D., and P. J. Ayoub. 2011. "Concept mapping in the analysis and design of cognitive systems: a historical review". *Applied Concept Mapping: capturing, analyzing and organizing knowledge*.

Meadows, D. H. 2008. *Thinking in Systems: A Primer*. Chealsea Green Publishing.

Mosindy, T. E. et al. 1987. "Impact of angling on the production and yield of mature walleyes and northern pike in a small boreal lake in Ontario". *North Am J Fish Mana* vol. 7 (4), pp. 493–501.

Pierce, R. B. et al. 1995. "Exploitation of northern pike in seven small north-central Minnesota lakes". *North Am J Fish Mana* vol. 15 (3), pp. 601–609.

Tsadiras, A. K. 2008. "Comparing the inference capabilities of binary, trivalent and sigmoid fuzzy cognitive maps". *Information Sciences* vol. 178 (20), pp. 3880–3894.

Verigin, T. et al. 2016. "Supporting a Systems Approach to Healthy Weight Interventions in British Columbia by Modeling Weight and Well-being". In *Proceedings of the 49th Annual Simulation Symposium*, ANSS '16, pp. 9:1–9:10, Society for Computer Simulation International.

AUTHOR BIOGRAPHIES

ERIC A. LAVIN is a Graduate Research Assistant in the Department of Computer Science at Northern Illinois University (USA). His thesis focuses on Fuzzy Cognitive Maps. His work has appeared in the Winter Simulation Conference. His email address is ealavin@gmail.com.

PHILIPPE J. GIABBANELLI is an Assistant Professor in the Department of Computer Science at Northern Illinois University (USA). He received his Ph.D. from Simon Fraser University (Canada). His research interests include data mining and simulation models. His email address is giabba@cs.niu.edu.

ANDREW T. STEFANIK received his MS in Computer Science from Northern Illinois University (USA), where he worked on simulation and network science projects with Dr. Giabbanelli. He is currently a professional developer. His email address is stefanik35@gmail.com.

STEVEN A. GRAY is an Assistant Professor in the Department of Community Sustainability at Michigan State University (USA). He received his Ph.D. from Rutgers University-New Brunswick. His research focuses on socio-environmental modeling. His email is grayste1@msu.edu.

ROBERT ARLINGHAUS received his Ph.D. from Humboldt-Universität zu Berlin (Germany), where he is now a professor. His research focuses on the human dimensions and bio-economics of recreational fisheries, fish and fisher behavioural ecology. His email address is arlinghaus@igb-berlin.de.