Network-Centric Simulation and Virtual Experimentation

Geoffrey P. Morgan
Kathleen M. Carley
Carnegie Mellon University
School of Computer Science
Institute for Software Research
5000 Forbes Ave
Pittsburgh, PA 15213
gmorgan@cs.cmu.edu, kathleen.carley@cs.cmu.edu

Keywords:
Socio-Cultural Modeling, Model Validation

ABSTRACT: Based on a segment of an annual course offered at Carnegie Mellon, we will discuss virtual experimentation and model consumption. We will introduce Construct, a network-centric simulation tool that may be of interest to modelers focused on larger group behavior. Construct incorporates key insights from sociological literature. We will teach attendants how to critically consume model results and how to design experiments involving models. Attendees will be taught how to use Construct and will execute simple virtual experiments for Construct.

1. Introduction

Simulation is a very powerful method useful to people in many research areas. This tutorial is designed to provide attendants with skills in critically consuming simulation results and in the design of simulation experiments. We use Construct (Carley, 1991), a network-centric simulation of information diffusion, as an example to discuss these concepts. We will discuss Construct and place the simulation within a taxonomy of modeling methods. We will provide instruction in how to use Construct through a UI and also provide material for the advanced user on how to fully explore the range of options Construct offers.

1.1 Why is this material important?

As simulation methods become more common-place, it becomes more and more important for researchers in many fields, as well as policy makers, to be able to usefully consume simulation results. Being able to evaluate experimental design and to critically evaluate model assumptions will help researchers from diverse fields work together and improve the quality of simulations that are developed.

Construct, as a network-centric simulation tool, is an interesting perspective on organizational modeling. Like many of the modeling methodologies presented at BRiMS, Construct has individual actors with unique characteristics. Yet Construct actors exist within a larger context defined by their access to knowledge and resources and their connections with other actors. Modelers of many disciplines may be interested in examining the possibilities of integrating network-centric techniques into their work, especially as network data becomes easier to access and more understandable to model consumers.

1.2 Tutorial Purpose

The tutorial intends to educate attendants in how to critically evaluate simulation work and to educate them in the construction of simulation experiments. It will present material refined yearly for a Summer Institute in network methodologies held at Carnegie Mellon since 2001.

1.3 Intended Audience

The intended audience of this tutorial includes people first learning about simulations and simulation techniques. It provides a larger context on simulation that may be useful to these attendees. People also interested in learning about network-centric simulation, either to use existing simulations or to incorporate elements into their own tools, will also be well served.

2. Contents

The tutorial has three sections, these are 1) Introduction to Construct, 2) Virtual Experimentation, and 3) Practical Construct Use through ORA.

The first section, *Introduction to Construct*, defines and describes Construct in a larger modeling context, describing the unique groups of features that Construct possesses and key concepts relevant to Construct from the simulation literature. It also describes the mechanics of Construct at a high level.

The second section, *Virtual Experimentation*, takes advantage of the first section by using Construct as an object for practical exercises. Following this exercise, we define virtual experimentation and discuss when such a technique is useful. We discuss, at a high-level, methods for parameter assignment.

The third section, *Practical Construct Use through ORA*, returns to Construct from an operational perspective. ORA, the Organizational Risk Analyzer, is an application which includes both best-of-breed network analytics as well as a simulation UI to create construct experiments. We will discuss how to use this interface and perform some simple experiments.

The remaining sub-sections describe each section in more detail.

2.1 An Introduction to Construct

The tutorial will begin by introducing and describing Construct. Construct is a turn and agent-based simulation useful for modeling information and belief diffusion. Each aspect of this previous definition will be defined and attendees will be encouraged to identify elements from their own experience that share these characteristics.

We approach modeling as a method of representation that removes those details not pertinent to the reason to model. A complete and accurate representation of an object would no longer be a model, but instead be the thing itself. As such, we present the attendee with the idea that models are common-place and everywhere, from recipes to toy cars to military simulations.

We describe Construct's turn mechanics, which are similar to Newell's decision cycle (Newell, 1990) and to Boyd's OODA loop (Boyd, 1987). We focus on the importance on perception as opposed to perfect apprehension of reality and the importance of bounded rationality to realistic modeling of human effort. We describe the importance of agency and also reference the importance of randomness in turn-based simulations to avoid accidental primacy among agents.

We also discuss two important human drives for socialization, based on the sociological literature: homophily (McPherson and Smith-Lovin, 1987) and expertise-seeking. Homophily is the preference for same, while expertise-seeking is the desire to interact with actors with rare knowledge. These two drives are more or less in action in many interesting modeling contexts.

From here, the tutorial will take a very brief break and then move on to the next section, Virtual Experimentation.

2.2 Virtual Experimentation

We begin this section by defining the term "Virtual Experiments". They are experiments operated upon a model. We remind ourselves that unlike experiments on reality, we are not discovering universal truths but instead what the model would predict. Models may be misapplied to problems that their assumptions poorly reflect. We keep the modifier "virtual" to remind us of this possibility. Finally, we remind the attendant of the useful maxim: "All models are wrong, but some are useful."

We suggest that virtual experiments are best used when testing reality would be unethical, infeasible, or incredibly expensive. Many real-world problems, particularly those involving people, exhibit these characteristics.

Next, the attendant is asked to think critically about Construct and its assumptions, based on the previous section of the tutorial. We introduce the term "model assumption" and note that such assumptions are integral to the model, as opposed to how the model was used in a particular instance. After the exercise, we identify these key assumptions for Construct.

Next, each attendant is asked to identify phenomena where these assumptions may not hold (or be useful). This exercise is used to promote a group discussion. We then remind the attendant that all models they produce or

consume have their own set of assumptions, and that considering whether those assumptions are appropriate for the problem at hand is nearly always a useful exercise.

We then return to the larger issue of Virtual Experiments, identifying some key issues for discussion as they apply to simulation work, these are:

- Independent Variables
- Dependent Variables
- Method
- Control Conditions
- Generality
- Power

Independent variables, for simulation purposes, that are simulation parameters that we intend to manipulate over the course of the experiment. We encourage the attendant, for each independent variable, to ask themselves several questions, such as "Why am I changing this variable?" and "Why these values?". We also remind the attendant of the potential problem of combinatoric explosion.

Dependent variables, for simulation purposes, are the things that are being measured. We introduce the idea of construct validity in connection with dependent variables, suggesting that it is important that what they're measuring in the simulation should, in some way, relate to what they would want to measure in the real world.

Method is one of the largest differences between standard experiments in reality versus experiments involving simulation. In simulation, this is largely confined to setting other variables that must be set for simulation operation to an appropriate constant setting. We offer three strategies for setting these constants:

- No Impact – Set the variable so that the process it informs is non-operational, this is appropriate if the process is orthogonal to your particular purposes
- A reasonable base-line – This is an appropriate strategy when the process is clearly important to your experiment but not central. Ideally, previous work has found that this is a useful operational setting.
- Monte Carlo - a random distribution – If the relevant process is important to your phenomena but a single setting is difficult to defend, it may be all the user can do to define an appropriate range and distribution that the variable should fall into. As such, this variable will be randomly set for every instance of the simulation and provide useful noise.

Control conditions, in simulation work, are model settings where the process those conditions inform have very little effect on the phenomena of interest. For network-centric simulation, Erdos-Renyi networks (Erdös & Rényi, 1960) are often used as a control condition because these networks have very little inherent structure. Yet, Erdos-Renyi networks are very unlikely in human phenomena!

We discuss the issue of generality in the sense that all model parameters should be informed from literature, and should not be described merely in the context of their particular situation. A particular network-centric example is provided.

We also discuss the issue of power. Simulation, because it can be repeated as many times as the modeler has computer horsepower, will often find statistical significance in the comparison of even the most minute differences. We encourage the attendant to focus on trends in their simulation results, rather than specific percentage differences. The use of explicit percentages often encourages the model consumer to believe that those percentages are likely to be found as-is in the real-world as well.

The model section concludes with a discussion of validity. Face validity, simple validity, and docking are described.

A brief break separates this section from the next. The break will be primarily used to make sure that those interested in the final section have the software installed on their machine.

2.3 Practical Construct Use through ORA

This section is hands-on exercise using ORA to create simple Construct experiments. ORA is a Windows-only tool. Attendees will be provided the software at the start of the tutorial (if possible, asked to install it before the conference).

The tutorial will focus on how to start ORA and on how to load networks into ORA. These networks will be used for the Construct exercises.

Once the attendees have loaded the networks, we will move to the Simulation UI, which allows ORA users to access Construct.

ORA's construct UI will be discussed and an example will be demonstrated. Users will define a base case and alternatives for comparison. They will run Construct and view the results through ORA.

Advanced materials (Hirshman, Carley, & Kowalchuk, 2007a; Hirshman, Carley, & Kowalchuck, 2007b) on how to use XML to initialize ORA will be offered, but not discussed in depth in this section

3. Conclusion

As a powerful method, Simulation is useful to people in many research areas. We have designed this tutorial to provide attendants with skill in the critical evaluation of simulations and simulation experiments. We feel that this will be valuable both to simulation designers and simulation consumers.

Construct offers an interesting and different perspective on agent-oriented simulation with insights that may be useful to others, either as a stand-alone tool or technologies they may want to consider applying to their own tools.

4. References

Boyd, J (1987) A discourse on winning and losing. Air University, Maxwell Air Force Base

Carley, K. M. (1991). A theory of group stability. *American Sociological Review*. 56, 331-354.

Erdös, P., and Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*. 5, 17-61.

McPherson, J. M., Smith-Lovin, L. (1987). Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups. *American Sociological Review*. 52, 370-379.

Newell, A (1990) Unified theories of cognition. Harvard University Press, Cambridge

Hirshman, B., Carley, K.M., and Kowalchuck, M.J. (2007a). Loading networks in Construct. *Technical Report CMU-ISRI-07-116*. Pittsburgh: Carnegie Mellon University.

Hirshman, B., Carley, K.M., and Kowalchuck, M.J. (2007b). Specifying agents in Construct. *Technical Report CMU-ISRI-07-117*. Pittsburgh: Carnegie Mellon University.

Author Biographies

GEOFFREY P MORGAN is a PhD student in Carnegie Mellon's Computation, Organization, and Society program. Morgan has experience in developing agent-based models, cognitive models, and network-centric simulation. He is interested in organizational impacts on individual decision-making and on optimizing organizational performance.

KATHLEEN M CARLEY is the director of the Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University. Her research combines cognitive science, social networks and computer science to address complex social and organizational problems.