Simulation: The basics

Russell J. Funk

Carlson School of Management
University of Minnesota

April 20, 2020

Roadmap

- Motivation
- Concepts
- Computation
- Organizational research
- Frontiers

Motivation

What is simulation?

For our purposes. . .

- Simulation is a research approach that uses computers to imitate some complex social or organizational system.
- Researchers have been using simulation to model social behavior for many years (and not always using computers).
  - residential segregation (Schelling, 1978)
  - organizational decision making (Cohen et al., 1972)
  - cooperation (Axelrod, 1984)

We'll focus on two uses. . .

- **Understanding the real world**—using simulation to better understand the nature and structure of observed data.
- **Creating your own world**—using simulation to better understand whether and how small sets of rules can generate complex phenomena.

Concepts

Using simulation to understand real world data

- Researchers often use simulation to understand the nature and structure of real data.
- When we observe some complex phenomenon in the social world , our inclination is to look for some social process as an explanation.
- However, even randomly generated data can exhibit complex structures.
- We may attribute the distribution of word frequencies to the economy of word use (c.f., Zipf) but random text also follows such distributions (c.f., Manning & Schütze, 1999).
- Within this context, simulation is useful for generating some "null" model against which we can compare our observed data.
- We saw an example of this in the Bearman et al. (2004) paper.

Broadly, work that uses this approach often has a "theory testing" orientation.

Using simulation to understand real world data

Challenges (and some solutions)

Getting the null model right

- You don't need to search long to find a random model from which your real data deviate.
- But that doesn't mean that your data were generated by a meaningful social process.
- There are all sorts of constraints in the real world that contribute to observed outcomes.
- Ideally, you want your null model to incorporate the constraints that are not of interest.

Using simulation to create your own worlds

- Researchers also use simulation to generate their own worlds.
- Within this literature, agent-based modeling is probably the most well-known stream.
- The basic idea is to define a small set of simple micro-level rules governing the behavior of "agents" and seeing whether these lead to complex macro level phenomenon.
- For example, Schelling (1978) shows that small in-group preferences can lead to significant residential segregation.
- We saw an example of this in the Hernandez and Menon (2017) paper.
- This approach is particularly valuable when complex interdependencies or adaptive behaviors are of interest that would be difficult to model empirically.

Broadly, work that uses this approach often has a "theory development" orientation.

Using simulation to create your own worlds

Challenges (and some solutions)

Keeping things simple

- When you are creating your own world, there are no constraints on what you can do.
- There is a tendency to make models match too many aspects of the "real world," thereby making the model too complex.
- Complex models are difficult to evaluate because it is unclear what features are driving observed outcomes.
- Also, if your model is complex enough, you can more or less get any result you want.
- One common solution to this challenge is to "tie your hands" as a researcher a bit.
- By limiting the complexity of the model, it's probably easier for you to be wrong (and more compelling when you're right).

Using simulation to create your own worlds

Challenges (and some solutions)

Connecting to real world data

- Another nice solution is to tie your simulation to real data.
- You might show some interesting emergent phenomenon that results from some micro mechanisms, but does that happen at realistic values of your parameters?
- If you can use empirical data to inform your model (e.g., values of your "hyper-parameters") your results are likely to be more compelling.
- We saw an example of this in the Guimera et al. (2005) paper.

Computation

What's the connection to computational social science?

At a surface level...

- Computation is pretty much the de facto (though not the only) approach to simulation.
- Because they typically entail many interactions and interdependencies, simulations can be very computationally intensive.

At a deeper level...

- Some people would argue that simulation IS computational social science.
- Simulation really embodies the inductive ethos of computational social science.
- Given simple rules, what kinds of complex phenomena can emerge?

Organizational research

Simulation and organizational research

- Researchers have been using simulation to study organizations for a long time.
- However, it has been more widely used in some subfields than others.

Learning

- search processes (Levinthal, 1997)
- exploration and exploitation (March, 1991)
- decision making (Cohen et al., 2972)

Networks

- diffusion processes (Lazer & Freidman, 2007)
- evolution (Tatarynowicz et al., 2016)

Structure

- formal organization (Clement & Puranam, 2018)
- emergence of culture (Centola & Baronchelli, 2015)

Frontiers

Where is simulation going next?

My $0.02

Understanding "big data"

- The massive increase in availability of "big data" has created a lot of opportunities for research.
- But there's also a lot of noise in big data sets—patterns that seem meaningful but that may not be.
- Simulation offers a way to compare these patterns to null models.

Multi-method research

- Along similar lines, as data has become more available, there is also increasing pressure to measure things directly (rather than purely simulating) when we can.
- This creates new opportunities for multi method research designs, that couple empirical data analysis with simulation (e.g., for unpacking mechanisms or exploring implications).

Appendix