Models and Simulations: Final Project Proposal

Instructions: Your research proposal should contain three parts: (1) a statement of the question you wish to address, (2) a plan of how you will address the question, and (3) a conjecture about what you expect to find and why your findings are philosophically important. I will describe each part of the proposal a bit more in depth. After describing each of three parts, I provide an example proposal. Before you start writing your final paper, you should set up a meeting with me to discuss your proposal in person.

In the first section of your proposal, you should state a research question and briefly summarize existing background research that addresses questions similar to yours. What should you include? As a general rule of thumb, you should imagine that your proposal will be read by an educated person who is unfamiliar with your area of research. For instance, imagine you are describing your final project to a professor in a different course.

With that rule of thumb in mind, the background section ought to contain three parts, which may occur in any order. First, you should motivate the importance of your research question. For example, suppose your (very complicated) research question is, "Does cooperative behavior emerge in repeated prisoners' dilemmas if agents are embedded in a network that evolves according to a preferential attachment model?" To motivate the importance of that question, you might want to describe why prisoners' dilemmas are important and where they occur in the real world. You might also wish to clarify why academics have had difficulty explaining the emergence of cooperative behavior so that you reader knows the your question is difficult; you might also wish to summarize how preferential attachment models have been used to explain interesting features of social networks.

Next, although your reader is educated, he or she may not know the terminology used in your research question. So you should define all terms that would be unfamiliar to the average reader. For example, if you asked the very complicated research question above, then you ought to explain briefly what network models are, clarify what "preferential attachment" means, and describe prisoners' dilemmas.

Finally, you should summarize similar existing research and explain how your project is different. Doing so lets the reader know that you are attacking a problem that is really novel. For example, you might say, "In [Citation X], the authors develop a dynamic network model in which agents interact with others if previous interactions have been profitable. However, their model differs substantially from preferential attachment models because Y. I will use preferential attachment models for reason Z."

In the second part of your proposal, you ought to explain how you plan to answer your question. In particular, you ought to describe the model that you will implement and how it can be used to address your question. If you plan to modify an existing model, describe what features of an existing model you will change. If you plan to implement an algorithm (e.g., a particular network formation algorithm) that was written by someone else, you should say so. Finally, after describing your model, you should explain how you plan to analyze and visualize the results (e.g., using particular types of graphs).

In the third and final section of your proposal, you ought to state what you expect to learn from your proposal. For example, you might say, "Result X in [author's name] model seem to depend critically on assumption Y. When assumption X is dropped, as it is in my model, I expect the following behavior to emerge, which differs significantly from previous findings in way Z." Do not worry: your conjecture need no be right! However, it is important to explain what you expect your results to be before you have completely implemented your model and run simulations. Think of your final project as an experiment. You may not know exactly what the outcome of your experiment will be, but if you cannot explain what it might tell you about your research question, then why are you running the experiment at all?

Your entire proposal ought to be no longer than four pages in length.

1 Example Proposal

1.1 Background

Pressing questions in science and policy-making require interdisciplinary collaboration. Economists collaborate with physicists to construct financial models, which dictate economic policy and the behavior of investment bankers. Biologists collaborate with statisticians to develop methods for high-dimensional data analysis that are then used to answer crucial questions in genetics. Chemists collaborate with medical researchers to develop prescription drugs. And the examples could be multiplied many-fold.

Despite the importance of collaboration, scientific institutions often implicitly encourage working *within* an established discipline. For example, physicists are typically trained by other physicists. Hence, they collaborate most frequently with physicists and publish most frequently in physics journals. In general, the organization of scientific disciplines strongly influences the (i) the training of researchers, (ii) the collaborative relationships that scientists form, and (iii) the dissemination of research. These factors, in turn, affect the speed and significance of scientific discoveries.

Recently, philosophers have developed several formal models to investigate the effects of collective scientific practices on the likelihood, speed, and significance of discovery.[1] However, none of these models investigates the way in which interdisciplinary collaboration affects scientific research. In this project, I will address this question by constructing a new model of scientific inquiry. Specifically, I will try to answer the following two questions:

Central Question: How should scientists choose collaborators so as to maximize the speed, reliability, and significance of their findings? Do scientists choose collaborators in a way that approximates the ideal?

1.2 The Model

To model collaboration among scientists, I combine two types of models: *network* models[2] and epistemic landscapes (EL) models[3] In my network model, scientific communities are represented by graphs like those in **Figure**

[1]See Kitcher [1990] Kitcher [1995], Strevens [2003], Weisberg and Muldoon [2009], and Zollman [2010].

[2]For general network models of learning and information sharing, see Goyal [2005] and Chapters 7 and 8 in Jackson [2008]. For models of scientific communities, in particular, see Newman [2001] and Newman [2004].

[3]See [Weisberg and Muldoon, 2009].

1, where nodes represent individual scientists, edges represent which pairs of scientists are collaborators, and colors represent the scientist's current research approaches. What is a research approach?

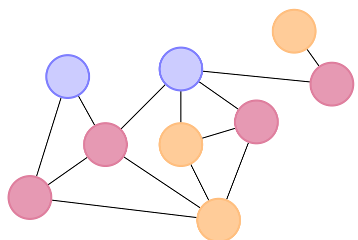

Figure 1: A Collaboration Network

The term "approach" is borrowed from EL models. At any given time in an EL model, each scientist has some finite number of available actions called "approaches"; these approaches represent which research questions a scientist deems important, which experimental techniques she employs, how she analyzes her data, and so on. EL models assume that different approaches to research generate results of differing *significance* (i.e. utility), and hence, scientists explore various research approaches to generate significant findings. The relationship between approaches and their significance is called an *epistemic landscape*. An example of an epistemic landscape is pictured below; the height (i.e. z-coordinate) of the surface represents epistemic significance, and the different pairs of xy-coordinates represent different possible approaches.

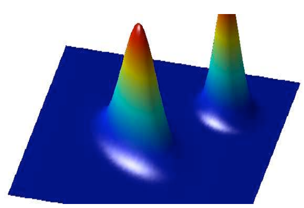

Figure 2: An Epistemic Landscape

Modeling collaboration requires making one small adjustment to EL models. The small adjustment is to assign significance not to a single approach, but rather to every set of approaches. For example, both game-theoretic modeling and mathematical techniques from evolutionary biology (e.g. replicator dynamics) are significant research approaches, but the two may combine to produce research of greater significance than other pairs of research approaches (e.g., techniques from game theory and quantum mechanics may not yield useful findings when combined).

I will model a network of scientists investigating a common epistemic landscape. Different academic disciplines will be represented by different regions of the landscape. Overlapping regions, therefore, represent sets of approaches that might be applicable in several disciplines. On each stage of inquiry, agents choose some limited number of collaborators and move in the epistemic landscape. Upon moving and choosing collaborators, researchers learn the significance of their findings and those of other scientists in their field. For example, two biologists learn the significance of each others' collaborations, but a biologist cannot learn the significance of a collaboration between an economist and physicist.

Given such a model, one can then investigate a series of questions that make precise the informal questions concerning collaboration with which this proposal began. For example, which strategies for exploring the epistemic landscape produce significant collaborations across disciplines? Are the networks produced by such optimal methods for pursuing collaborators similar in structure to the types of co-authorship networks that are observed in scientific practice?

To answer the first question, I will model agents as employing boundedly rational strategies (e.g. "imitate the best" researcher in one's field) when choosing collaborators and compare which of the strategies leads agents to find the most significant results on the epistemic landscape. To answer the second, I will compare the graph statistics (e.g. mean path length between agents or clustering coefficient) of the networks produced by various boundedly rational strategies with the corresponding graph statistics of real world co-authorship networks.

1.3 Expected Results

I conjecture that networks produced by the optimal learning rules will resemble co-authorship networks for the following reason. As agents learn, they will form collaborations with researchers who have been successful in the past. The result is that certain collaborators will become very popular, whereas agents who are known to produce insignificant work will be unpopular collaborators. If the number of successful agents is sufficiently small, then most agents in the collaboration network will have only a few neighbors (i.e. collaborators), whereas a small number (the best researchers) will have an enormous number of neighbors. So one should observe a degree distribution that decreases steeply in degree; this resembles the power law degree distributions in practice. I conjecture that a similar argument can be given to show that the diameter of the resulting collaboration networks (i.e. the longest path in the network) will be short, which is another common property of co-authorship networks. These results confirm similar findings obtained by [Anderson, 2012].

These results would be important because many researchers have conjectured that structural features of scientific networks are best explained by researchers' desires to obtain fame, grant money, or power by collaborating with the better-known scientists. My results would indicate that scientists motivated solely by the production of significant research might form similar collaborative networks. Perhaps most importantly, the results would show that one need not be skeptical of scientific research produced by researchers with epistemically "impure" motives, as scientists who are motivated purely by the pursuit of truth might produce exactly the same results.

References

Katharine Anderson. Skill specialization and the formation of collaboration networks. *Unpublished Manuscript*, 2012.

S. Goyal. Learning in networks. *Group Formation in Economics: Networks, Clubs and Coalitions*, pages 122–170, 2005.

Matthew O. Jackson. *Social and economic networks*. Princeton Univ Press, 2008.

Philip Kitcher. The division of cognitive labor. *The Journal of Philosophy*, 87(1):5—22, 1990.

Philip Kitcher. *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. Oxford University Press, USA, 1995.

Mark E.J. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404—409, 2001.

Mark E.J. Newman. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl 1):5200—5205, 2004.

Michael Strevens. The role of the priority rule in science. *The Journal of philosophy*, 100(2):55—79, 2003.

Michael Weisberg and Ryan Muldoon. Epistemic landscapes and the division of cognitive labor. *Philosophy of Science*, 76(2):225—252, April 2009.

Kevin J. Zollman. The epistemic benefit of transient diversity. *Erkenntnis*, 72(1):17—35, 2010.