

Threshold Phenomena in Epistemic Networks

Patrick Grim

Group for Logic & Formal Semantics, Department of Philosophy, SUNY Stony Brook
Visiting Scholar, Center for Study of Complex Systems, University of Michigan, Ann Arbor
pgrim@notes.cc.sunysb.edu

Abstract

A small consortium of philosophers has begun work on the implications of epistemic networks (Zollman 2008 and forthcoming; Grim 2006, 2007; Weisberg and Muldoon forthcoming), building on theoretical work in economics, computer science, and engineering (Bala and Goyal 1998, Kleinberg 2001; Amaral et. al., 2004) and on some experimental work in social psychology (Mason, Jones, and Goldstone, 2008). This paper outlines core philosophical results and extends those results to the specific question of thresholds. Epistemic maximization of certain types does show clear threshold effects. Intriguingly, however, those effects appear to be importantly independent from more familiar threshold effects in networks.

1. Epistemology and Scientific Networks

Epistemology is defined as the study of knowledge. The traditional focus in the field, however, has long been limited to the study of the individual epistemic agent. Traditional epistemology treats knowledge acquisition as an individual endeavor. In Hume, Descartes, and Kant, epistemology is told as the story of a single individual trying to figure out what the world is like—an attempt to answer the question of how an individual agent figures out the structure of the world.

A small consortium of contemporary philosophers has begun work on a different approach (Zollman 2008 and forthcoming; Grim 2006, 2007; Weisberg and Muldoon forthcoming). In this recent work the essential emphasis is not on communities of epistemic agents rather than on the individual.

How does an individual figure out the structure of the world? The truth is that no individual does. It is cultures and communities that plumb the structure of reality; individuals figure out the structure of the world only as they participate in the epistemic networks in which they are embedded.

Science is undoubtedly our pre-eminent example of knowledge acquisition. But what characterizes contemporary scientific research is not a catalog of isolated investigators but coordinated interactive *networks* of investigation. To understand knowledge acquisition in science one must understand more than the work of

individual participants. One must understand the structure and dynamics of the enterprise as whole.

Here questions are importantly different than those in traditional epistemology. A scientific community can be envisaged as a network of interactive agents attempting to limn reality on the basis of uneven, conflicting, and sometimes ambiguous data. How does the network structure of collaboration and competition, of data sharing and information transfer, affect knowledge acquisition in the community at large? What kinds of network structures, of what kinds of agents, will best achieve scientific goals—scientific goals of accuracy, for example? In what ways will those structures be sensitive to the specific form of the problem, or to the distribution or uncertainty of data? Those are questions central to this new approach, and questions for which the work outlined below offers some early and partial clues.

Given what we know of networks in general, it is to be expected that the dynamics of information acquisition and exchange across an epistemic network will not be reducible to any simple aggregate measure across individuals. The modeling results offered here substantiate that expectation in full. One of the implications of epistemic networks, tracked here in terms of thresholds, is the robust and surprising finding that a scientific community may learn *more* when its individual scientists learn *less*. In terms of central scientific goals such as accuracy, increased informational linkages between scientists may not always be a good thing.

17th century science was characterized by distributed informational networks with limited linkages between investigators. 21st century science is characterized by totally connected networks across the internet. One way of phrasing a central result in what follows is that for some central scientific goals, including accuracy, and for some topics of investigation, the network structure of 17th century science appears to be *superior* to our own.

Section 2 outlines the notion of epistemic landscape crucial to the model, with details in section 3 of initial networks surveyed. The core result that a scientific community can learn more when individual scientists learn less is presented in section 4. Sections 5 and 6 further explore the question of precisely what properties of

networks are important for that result. Here results show clear thresholds for epistemic maximization of certain types with increasing number of links in random networks.

Epistemic maximization on networks of the type at issue, it turns out, exhibits clear threshold phenomena. But it also turns out that the epistemic thresholds at issue are surprisingly independent from other network; they do not correlate cleanly with thresholds in any of the other network properties one might expect.

Results here are intended as an introduction, with first hints regarding some of the surprises and subtleties of informational dynamics across epistemic networks. These are offered as a first word on the topic, rather than the last word; it quickly becomes clear how much we do not yet understand, and how much more work remains to be done.

2. Epistemic Landscapes

We can envisage an epistemic landscape as a topology of ideal data—data regarding alternative medical treatments, for example (Fig. 1). In such a graph points in the xz plane represent particular combinations of radiation and chemotherapy, or particular hypotheses regarding the best combination. Graph height on the y axis represents some measure of success: the proportion of patients in fact cured with combinations of radiation and chemotherapy at that rate. If you use radiation therapy at rate x , and chemotherapy at rate z , you will get the proportion of cures represented on the y axis hovering over that point.

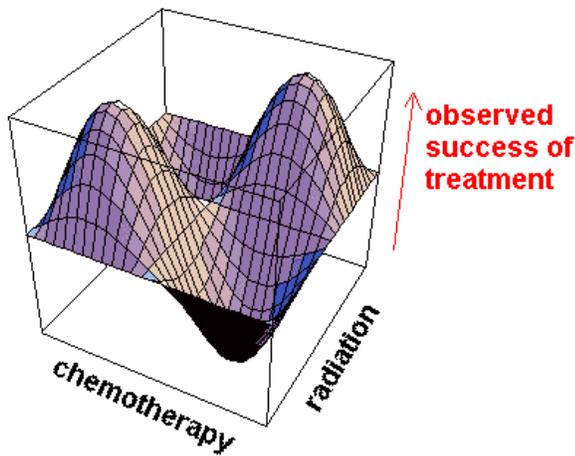


Fig. 1 A three-dimensional epistemic landscape. Points on the xz plane represent hypotheses regarding optimal combination of radiation and chemotherapy; graph height on the y axis represents some measure of success.

This first epistemic landscape is a medical one, but the specific topic of investigation is unimportant for our broader epistemic concerns. One might have an epistemic landscape of magnetology readings for different hypothetical locations of a shipwreck, or of irridium

stratigraphy world-wide as feedback regarding different hypotheses regarding the timing of the K-T asteroid collision, or any measurable variable y that confirms some hypotheses rather than others regarding the interplay of variables x and z .

It is important to emphasize, however, that the concept of an epistemic landscape represents *ideal* data across a full range of possible hypotheses. Different investigators will test different hypotheses and will get differential feedback regarding those hypotheses. As an individual investigator, however, one will not be able to see the epistemic landscape as a whole. One will see results only at a point in the graph, in a small area or in a scattering of points.

Despite those limitations, our job description as epistemic agents is to find the theory that is best supported by data. The goal of investigation is to find the highest points in the epistemic landscape—the best confirmed hypotheses, or the most warranted predictions, the most reliable medical treatments. Fortunately, we do not work alone: we are linked to other investigators as part of a larger network.

The model at issue here employs simpler two-dimensional epistemic landscapes (Fig. 2).

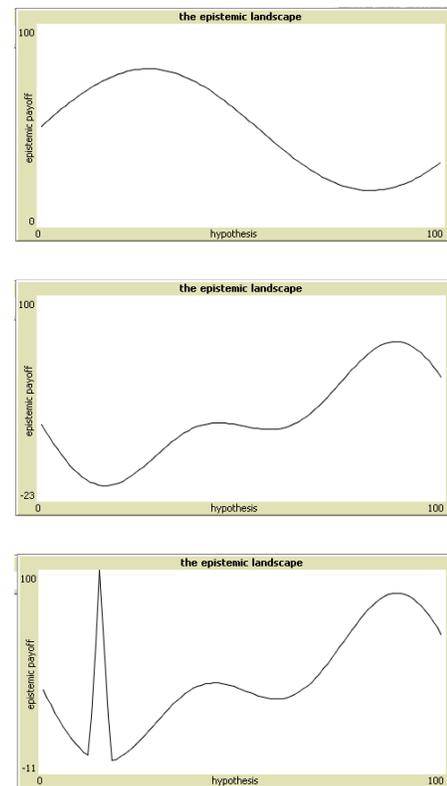


Fig. 2 Two-dimensional epistemic landscapes. Values on the x axis represent alternative hypotheses. Values on the y axis represent the ideal epistemic payoff for particular hypotheses.

In the first landscape data converges smoothly to a single best hypothesis or medical treatment. The second represents a slightly more complex landscape, in which particular combinations of drugs do well, perhaps, but combinations in between do worse. The third is a still more complex landscape, in which some peaks are smooth and easily climbed, but represent inferior outcomes. The hypotheses or medical treatments they lead to are not the best. The best outcome, however—that hypothesis that would be best confirmed, or that medical treatment that would be most effective—is hidden in a spike with a narrow base, and is thus harder to find.

3. Modeling Epistemic Networks

Suppose we have a population of agents, each of whom starts with a hypothesis. Here that hypothesis is represented by a single point on the x-axis of an epistemic landscape. In testing their hypotheses, our agents accumulate data as feedback—a percentage of patient cures, for example. But our agents are also networked to others; they can see not only the success rate of their own hypothesis but the success rate for the hypotheses of those to whom they are linked.

Agents change their hypotheses based on the success rates of those to whom they are linked. As an agent in this model, you can see how well the hypotheses of some other agents are doing; if their hypotheses are better supported by the data than yours, you shift your hypothesis in their direction. If your hypothesis is the best of those visible to you, on the other hand, you stick with it.

With even a network model this simple there are a number of intriguing parameters. One of the parameters built into this model is a ‘shaking hand’: when you aim to duplicate another’s hypothesis, you may be slightly off. Your lab conditions may be slightly different from that of the other agent, or your chemicals impure, or your sample slightly biased. You therefore end up with a hypothesis that is not a precise match of that you are imitating but is merely close by. One result, of course, is that you therefore explore more of the epistemic landscape. The model used here builds in a ‘shaking hand’ that puts one in random region within 4 points either side of a target hypothesis.

The model also incorporates elements of ‘speed’ and ‘inertia’. In pursuing a more successful hypothesis, does one jump to that conclusion or approximate it halfway each time? This model employs the latter assumption, with a ‘speed’ of 50%. It also builds in an ‘inertia’ factor of 50%, representing agents’ stubborn investment in their current hypothesis. In each of 100 steps each agent has only a 50% probability of shifting in the direction of a superior hypothesis.

The crucial parameter the model is designed to investigate, of course, is network structure (Fig. 3).

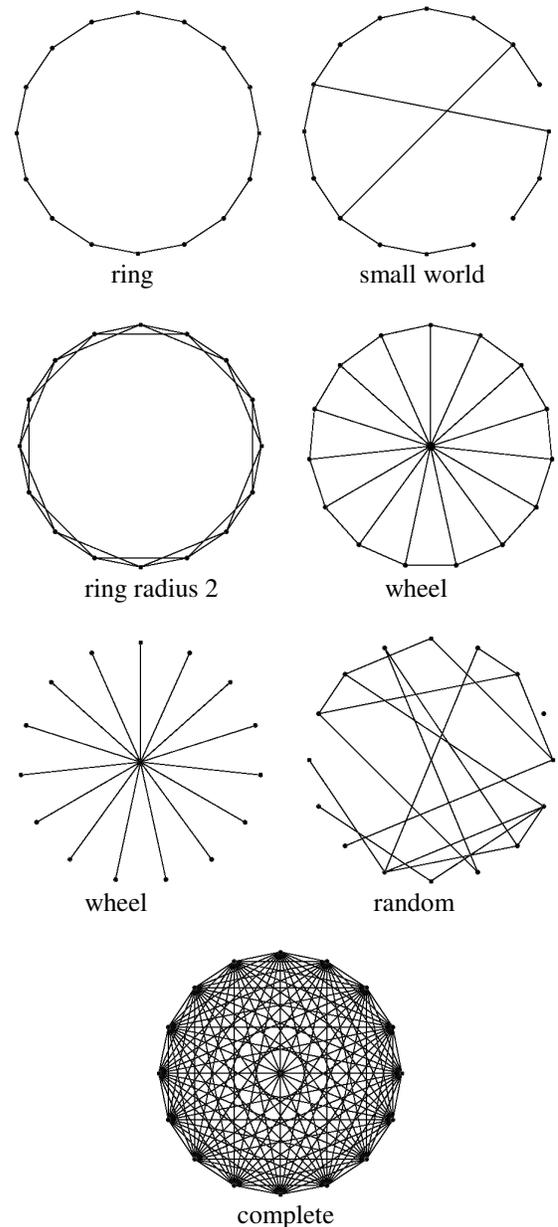


Fig. 3 Structures of epistemic networks used in the model. Shown here with 20 nodes for visibility, networks used in the model linked a population of 50 nodes.

Working with 50 agents, we studied networks in which the network structure is:

- (1) a simple ring, with contacts to a single agent on each side;
- (2) a small world network, here a simple ring with a 9% probability of rewiring;
- (3) a ring with double radius, in which each agent has contacts with two agents on each side;
- (4) a wheel, also known as a ‘royal family’ in the economics literature, in which each agent on the ring also has contact with a central agent;

(5) a hub, in which agents have contact only through that central agent;

(6) a random network, here with a 10% probability of combination between any two nodes;

(7) A total, connected, or complete network, in which all nodes are linked to all others.

Though illustrated in terms of 20 agents in Fig. 3, the model employed the network types above with 50 agents. With the noted parameters for shaking hand, speed, and inertia outlined above, each agent in the network updated on those two which it was linked through a series of 100 steps. 100 runs were performed for each network with re-randomization of agent hypotheses and network structure in the case of small world and random networks.

4. Network Effects in Epistemic Maximization

In which networks, exploring which epistemic landscapes, will agents succeed in finding the optimal hypothesis? Here our measure was whether any agent in the network found the optimal hypothesis; assuming a process of convergence to the highest all in a connected cluster would eventually follow suit.

For the simpler two epistemic landscapes in Fig. 2, with smooth climbs to their peaks, all of the networks surveyed found the highest point in all 100 cases. Those networks which systematically found it most quickly were those with the highest connection and degree. In these studies it was the total or connected network which most quickly found the peak.

Results were intriguingly different for the more complex network represented by the epistemic landscape in Fig. 4, in which the optimal hypothesis is hidden in a narrow peak.

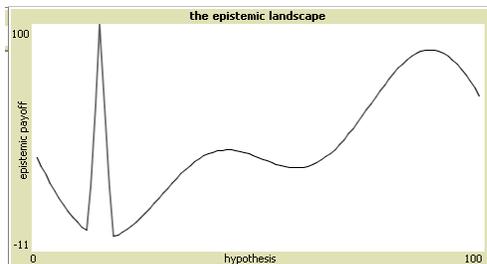


Fig. 4. The more complex epistemic landscape, in which the optimal hypothesis is hidden in a narrow peak.

In this case none of the networks surveyed found the optimal hypothesis in all 100 runs. The percentages of successful runs, moreover, show a wide variance with different network structure (Fig. 5).

For this epistemic landscape, a regular ring of networked agents connected to a single neighbor on each side, with a 'shaking hand' of 4 points and using speed and inertia

factor of 50%, converges on the highest point in 66% of the runs.

For a 'small world' variation, with a 9% probability of rewired connections, the success rate drops immediately to 55%.

Connect each node not with a single neighbor on each side, as in the single ring, but with two neighbors on each side and the success rate drops immediately to 40%.

Networks configured as wheels and hubs give a 42% and 37% success rate in finding the optimal hypothesis.

Random networks with a 10% probability of connection between any two nodes give a success rate of 47%.

Worst of all, operating on this epistemic landscape with the background assumptions noted, are total or connected networks. Here the chance of finding the epistemic optimum is a mere 32%--approximately half the success rate of a simple ring.

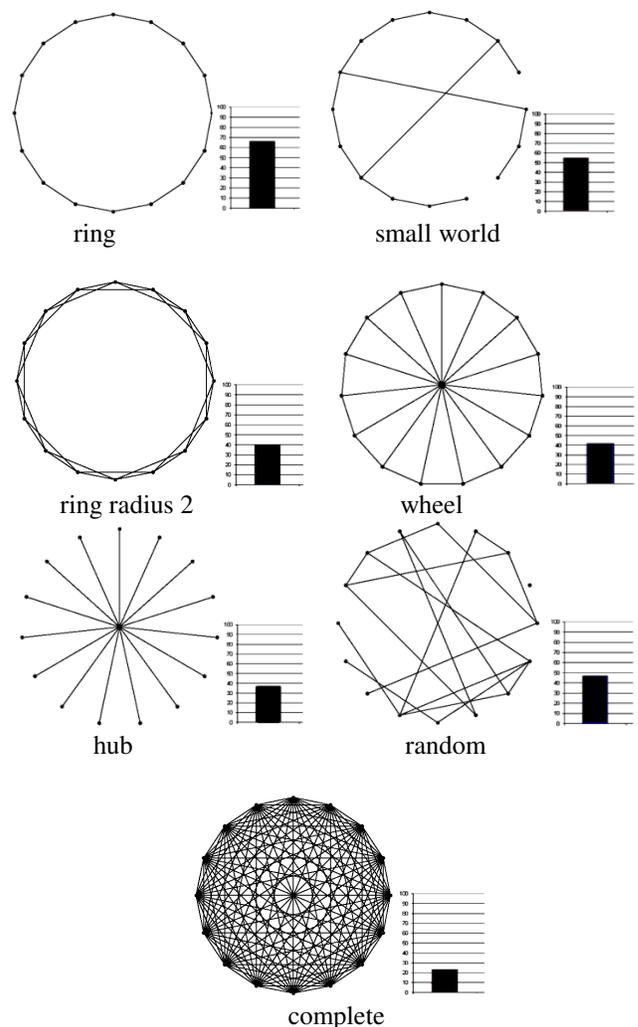


Fig. 5 Percentages of 100 runs in which networks of different structures found the optimal hypothesis in the epistemic landscape of Fig. 4.

Scientific exploration has multiple desiderata. For the central scientific goal of accuracy—here represented by finding the optimal hypothesis in an epistemic landscape—it is clear that informational linkages between investigators is thus not an unadulterated good. Given a tendency to follow the lead of a more successful hypothesis, even with an ‘inertia’ parameter of 50% and a 50% restriction on ‘speed,’ increased linkages between nodes can result in a convergence on a suboptimal hypothesis. The problem with increased linkages is that there is convergence to that hypothesis that is ‘best in show’—the best presently occupied by anyone in the network—but which is not the highest in the landscape or the best in fact. Epistemic networks with increased linkages have a tendency to let exploitation trump exploration, resulting in quick convergence to a sub-optimal epistemic outcome.

A scientific community may therefore learn less when its individual scientists learn more—more, that is, regarding others’ immediate results. The scientific networks of the 21st century, massively connected through the internet, approach the character of a total or connected graph. The scientific networks of the 17th century were significantly more distributed: investigators communicated in large part on an individual basis and by letter, perhaps in a structure as distributed as that of a ring network. On the agent updating assumptions used here, and in cases in which the epistemic landscape has something of the character of Fig. 4, it is science on a 17th century network rather than a 21st century network that could be expected to prove more successful.

5. Increased Linkages in Random Networks: An Epistemic Threshold

Precisely what property is it of the networks above that facilitates epistemic success in such a case?

To this point we have been working with a small sample of assorted networks. The results are suggestive, but further exploration is needed. What is required for a deeper understanding is observed variability in epistemic success with variation in a single network parameter or small combination of parameters. In pursuit of that understanding, the results that follow leave behind the wide variety of networks surveyed above, concentrating instead on a single form. The goal within that focus is to track the effect on epistemic outcome of varying the number of links between nodes.

A number of assumptions are retained from the studies above. We again use the epistemic landscape of Fig. 4, with networks of 50 agents, a ‘shaking hand’ of 4 points variability, and using an updating ‘speed’ and ‘inertia’ of 50%. Here our networks will be random networks throughout, however—perhaps the most studied of all network structures. None of the networks used, therefore, have the clear symmetry of a simple ring, and none will be

as connected as a total or complete network. Results can therefore be expected to lie between the extremes found above. What this closer focus allows us is the possibility of systematically varying the single parameter of number of links between agents to track implications for epistemic network success on the given landscape.

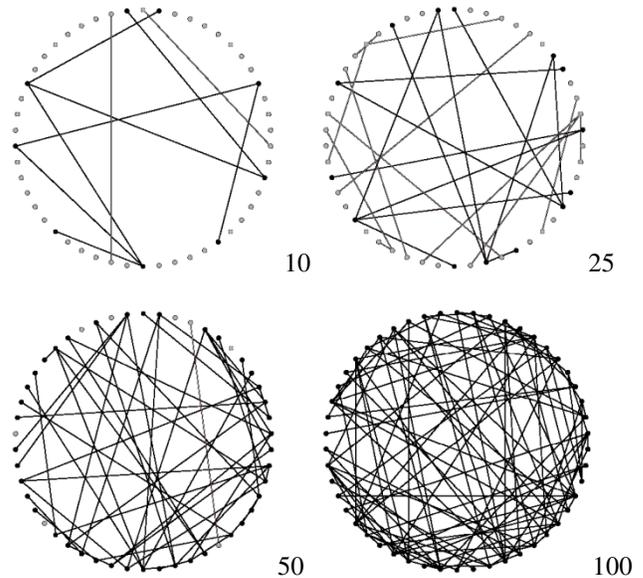


Fig. 6 Sample random networks with increasing numbers of links. Giant cluster shown in black in each case, with other links in gray.

How does increased connectivity in random networks affect epistemic outcome? The results appear in Fig. 7, showing the result of increasing number of random links between nodes from 5 to 300. Shown for each number of links at intervals between 5 and 300 is the percentage of 1000 runs in which random networks with that number of links succeeded in finding the highest point on the epistemic landscape.

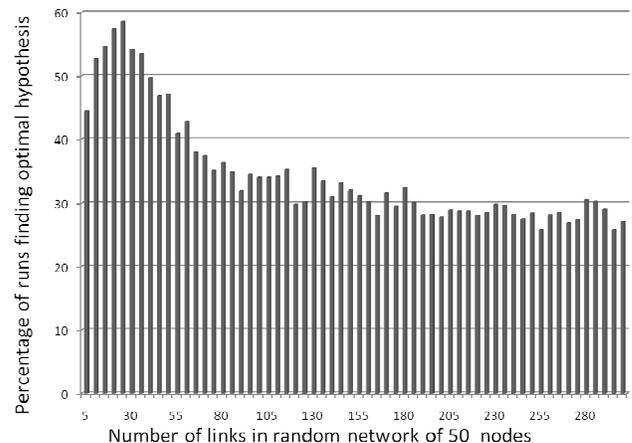


Fig. 7 Percentage of 1000 runs in which random networks of 50 nodes with the given number of links succeeded in finding the optimal hypothesis in the epistemic landscape of Fig. 4.

The results show a clear epistemic threshold at approximately 25 nodes. Although this is in line with the results outlined above, it is nonetheless surprising that those random epistemic networks did best in which the number of links was so low. Within the assumptions noted, those random epistemic networks did best in which the number of links was approximately half the number of nodes. From that peak at approximately 25 links the success of networks decays as links increase to approximately 100, the point at which there are twice as many links as nodes. From that point on decay in epistemic success continues, though at a much slower rate.

6. Comparing Threshold Phenomena

One reason for using random networks in this study is that they are among the best understood. There are, in particular, well established threshold phenomena regarding increased linkages in random networks. A primary question, therefore, is this: Does the epistemic threshold of increased linkages documented above merely track some more familiar threshold in network phenomena?

What is it about a network that accounts for epistemic success of the sort detailed here? The usual suggestion I have heard in response to that question is ‘clustering coefficient’: the mean probability that two nodes linked to a common node will also be linked to each other (Watts & Strogatz 1998; for clarification see Barrat and Weigt 2000 and Newman et. al. 2001).

The hypothesis that it is variability in clustering that is responsible for variation in epistemic success fits perfectly the extremes of results outlined in earlier sections. A ring network has a clustering coefficient of precisely 0: none of an agent’s neighbors are also neighbors of each other. Fully connected networks have a clustering coefficient of 1: all of each agent’s neighbors are also agents of each other.

If it were clustering coefficient that were the deciding factor regarding epistemic success of the sort outlined, however, we would expect the graph of epistemic success in Fig. 7 to mirror the graph of changes in clustering coefficient. Figure 8 shows mean clustering coefficient across our 1000 runs at each number of links between nodes.

Whatever network property is responsible for threshold phenomena in epistemic nets, Fig. 8 makes it clear that it cannot be clustering coefficient. Clustering traces a clear linear ascent with increased number of nodes, with no reflection of the epistemic thresholds evident in the graph of epistemic success.

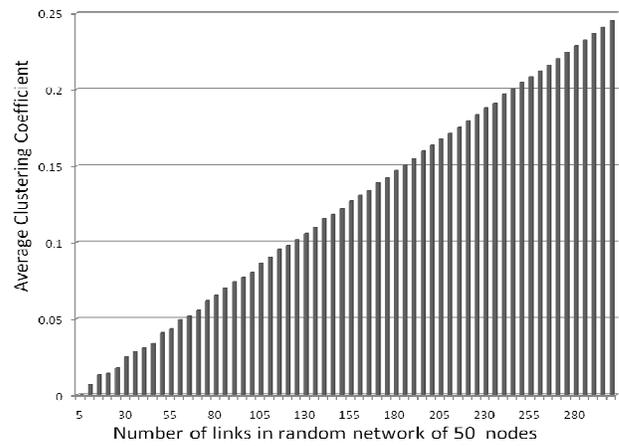
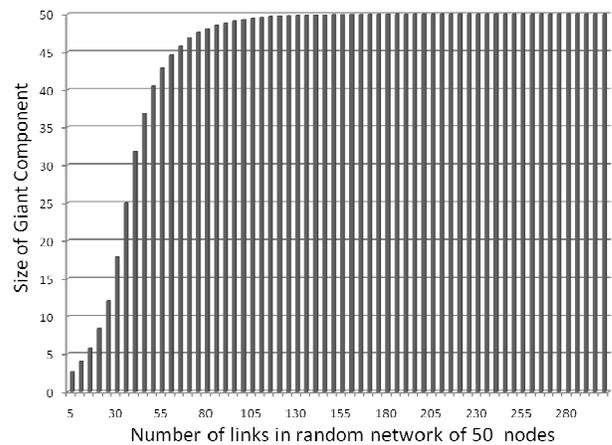


Fig. 8 Mean clustering coefficient across 1000 runs at each number of links in random networks of 50 nodes.

There is a well-known threshold in random nets with increasing number of links: the size of the giant component, or largest number of connected links. As established quite early (Erdos and Renyi 1959, 1960), the increase in size of the giant component with increased numbers of links in random networks is far from linear: it shows a clear threshold at approximately the number of nodes. At that point there is a dramatic increase in the size of the giant component.

For the networks at issue, that threshold appears in Fig. 9. The chart shows the average giant component size in our 1000 runs across each number of added links.



In attempting to explain results regarding epistemic success, however, the threshold of giant component size seems to be in the wrong place. It begins at approximately half the number of nodes, tapering off at slightly more than the number of nodes. This threshold, of course, is also unidirectional. Though it may be a contributing factor to the epistemic success threshold noted, it cannot offer a complete explanation.

Here we offer a final candidate for a network property that might correlate with epistemic threshold: the average path length within the giant or largest cluster. Simple average path length is not a useful measure here, because it is defined only for connected graphs that contain no isolated nodes or sub-graphs. That requirement does not hold uniformly for networks at issue below approximately 150 links, and so cannot be used to measure across the spread. Average path length within the giant component, however, can be used as a measure. Graphed across our runs that measure generates the threshold shown in Fig. 10.

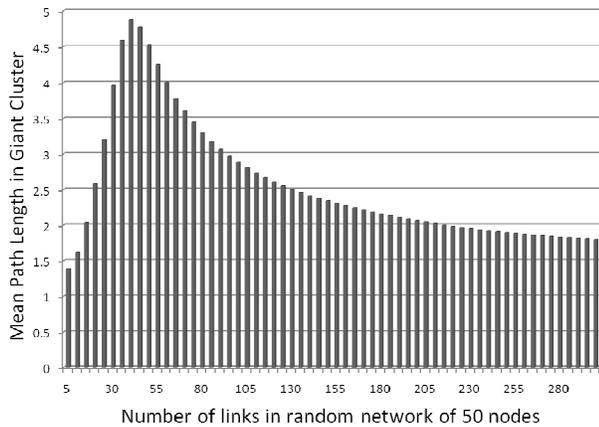


Fig. 10 Average path length within the largest cluster across 1000 runs at each number of links in random networks of 50 nodes.

Here finally we have a network property with a threshold similar in shape to that we are tracking. On that basis mean path length in the giant cluster of a network seems a prime candidate for a factor that favors epistemic success on the particular epistemic landscape at issue here.

Questions remain, however. Although the graph of mean path length has much the shape of the graph of epistemic success, its peaks are at the wrong points. Mean path length peaks at that point that the number of links equals the number of nodes. Epistemic success, in contrast, peaks at the point that the number of links is half that. Epistemic success seems to level when links equal twice the number of nodes, whereas the decay in mean path length is more gradual.

Although qualitative comparison of effects indicagtes mean path length in giant cluster as a primary candidate among contributing causes, therefore, it does not appear to give us the whole story. Even if the general shape of success qualitatively tracks that of mean path length in the largest cluster, we do not yet understand why it tracks it at a different point. Although the similarities in thresholds is striking, we don't yet understand the differences.

6. Further work

The results above are confined to particular assumptions regarding updating dynamics within networks, using a specific epistemic landscape for comparison throughout. Further work is required to explore the effect of changing some of those updating assumptions, perhaps mirroring different incentive structures for investigators. Further work is also required in order to explore the relation of thresholds noted to particular characteristics of epistemic landscape: to track what characteristics of networks optimize epistemic success on specific landscapes.

In the long run, such an exploration promises a new epistemology, offering a better understanding of the dynamics of knowledge acquisition in science. In the long run, goals may be normative as well as descriptive. Such an exploraiton might also offer the possibility of optimizing ongoing scientific exploration. Given first indications of what an area of investigation is like—a first glimpse of an epistemic landscape—we might be able to tell whether such a landscape would most effectively be investigated by big science or small, by a few groups of closely linked investigators, a scattered set of independents, or some combination of the two.

As indicated in introduction, these results are intended as a brief introduction to some of the surprises and subtleties of informational dynamics across epistemic networks—a first word on the topic, but far from the last.

Acknowledgments

Important components of this work were developed collaboratively with students and colleagues at the University of Michigan and Stony Brook. I am grateful to Aaron Bramson for early and important programming, and to Jennifer Carter, Andrew Donaruma, Aklima Khondoker, Ryan O'Shea, and Adam Rosenfeld for discussion and further development.

References

- L. A. N. Amaral, A. Diaz-Guilera, A. A. Moreira, A. L. Goldberger, and L. A. Lipsitz, 2004. Emergence of complex dynamics in a simple model of signaling networks. *Proceedings of the National Academy of Sciences USA* 101, 15551-15555.
- V. Bala and S. Goyal, 1998. Learning from Neighbors. *Review of Economic Studies* 65 (1998), 565-621.
- A. Barrat and M. Weigt, 2000. On the properties of small-world networks. *European Physical Journal B* 13, 547-560.

P. Erdos and A. Renyi, 1959. On random graphs. *Publicationes Mathematicae* 6, 290-297.

P. Erdos and A. Renyi, 1960. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5, 17-61.

P. Grim, 2006. Tangled Webs: The Philosophical Importance of Networks. The Marshall Weinberg Lecture, University of Michigan.

P. Grim, forthcoming. Network Simulations and their Philosophical Implications: Models for Semantics, Pragmatics, and Epistemology. In Paul Humphreys, ed., *Models and Simulations III*.

J. Kleinberg, 2001. Small-world phenomena and the dynamics of information. *Advances in Neural Information Processing Systems* 14, 431-438.

W. A. Mason, A. Jones, and R. L. Goldstone, 2008. Propagation of Innovation in Networked Groups. *Journal of Experimental Psychology: General* 137, 422-433.

M. E. J. Newman, 2001. Clustering and preferential attachment in growing networks. *Physical Review E* 64, 025102.

D. J. Watts and S. H. Strogatz, 1998. Collective dynamics of 'small-world' networks. *Nature* 393, 440-442.

M. Weisberg and R. Muldoon, forthcoming. Epistemic Landscapes and the Division of Cognitive Labor. *Philosophy of Science*.

K. Zollman, 2008. The Communication Structures of Epistemic Communities. *Philosophy of Science* 74: 574-587.

K. Zollman, forthcoming. Social Structure and the Effects of Conformity. *Synthese*.

K. Zollman, forthcoming. The Epistemic Benefit of Transient Diversity. *Erkenntnis*.