Network Formation with Heterogeneous Nodes and Stochasticity*

Remco Heesen†

June 17, 2016

1 Introduction

Academic superstars are a familiar phenomenon. These scientists write the papers that everyone reads and talks about, they make media appearances, give presidential addresses, and they win grants and awards. The work of an academic superstar generally attracts more attention than that of the average scientist.

Quantifying attention as the number of citations to their papers, sociologists found an easy way to identify academic superstars. They also noted that superstars are rare: the vast majority of scientists receives no more than a handful of citations, while a rare few get extremely many (Price 1965, Cole 1970). In particular, the distribution of citations follows a "power law": the number of papers that gets cited n times is proportional to $n^{-\alpha}$ for some α. Redner (1998) estimates α to be around three.

*I would like to thank Kevin Zollman, Teddy Seidenfeld, and Katharine Anderson for valuable comments and Dominik Klein for suggesting corollary 15 and outlining a proof.

†Department of Philosophy, Baker Hall 161, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA. Email: rheesen@cmu.edu.

The standard explanation for this characteristic power law pattern in the literature relies on so-called preferential attachment models. Preferential attachment models are network formation models in which new nodes link to older nodes with probabilities proportional to the number of links that older node already has. It can be shown that this generates a power law distribution of links with an exponent equal to three (Barabási and Albert 1999). Interpreting nodes as papers and links as citations creates a close match between empirically observed citation patterns and those predicted by preferential attachment.

A strength of this explanation is its ability to match the observed distribution of citations very closely. A weakness is that it gives no insight as to why scientists would choose to cite papers according to this pattern.

Here I offer an alternative model in which the content of papers plays an explicit role in scientists' decision whether to read or cite it. I set aside various biases that scientists might have in choosing what to read or cite and instead view these decisions purely as an information exchange. My goal is to show that this factor by itself is sufficient to produce the citation patterns that may be seen in real life. In particular, my aim is to establish something like the following claim.

Claim 1. *If, in choosing whose work to read, scientists are motivated only by gathering as much information as possible given their means, then the patterns of interaction that emerge are highly imbalanced: some scientists get a lot of attention, while most get very little.*

Insofar as I succeed in establishing this claim, it shows that scientists' desire for information (which is surely one of scientists' many motivations) can work as a mechanism that leads to the existence of academic superstars. As such it can be an alternative to preferential attachment models or supplement them. A key strength of the present model is that it offers some limited insight into the question which papers become highly cited and which do not, a question on which preferential attachment models are completely silent (the initial differences in citations among papers are generated completely randomly in such models).

From a technical perspective, the model I present is a strategic network formation model in the style of Jackson and Wolinsky (1996) and Bala and Goyal (2000). In those early network formation models, nodes are distinguished only by the characteristics they have relative to the network (i.e., their links and their neighbors' links). The "information" being transmitted between the nodes takes the form of deterministic, additive "bits", of which each node has exactly one. Following Anderson (2016) I allow some heterogeneity between nodes. In Anderson's model each node possesses some subset of the finite set of bits of information (called "skills" by Anderson) available in the model. I generalize this further by allowing information to take the form of sets of random variables, interpreted as experiments that scientists would like to learn the outcome of. This constitutes the main technical innovation achieved here.

The model is described in section 2.1. The next two sections explore some specific features of the model. Section 2.2 defines a notion that allows one to compare the information each scientist has access to, and proves that this notion works for its intended purpose. Section 2.3 builds on this to define the relative frequencies of information in a scientific community. Two assumptions concerning the behavior of scientists are stated in section 2.4, along with some considerations of why these assumptions might be acceptable as a form of bounded rationality. The main result, which substantiates claim 1, is then proved in section 2.5.

As a further defense of the two assumptions made in section 2.4, I prove that Bayesian scientists will satisfy them under certain conditions. Section 3.1 introduces the extra machinery I need. Section 3.2 proves a theorem related to the first assumption, and section 3.3 does the same for the second assumption.

2 A Network Formation Model with Heterogeneous Nodes and Stochasticity

2.1 The Model: Scientists, Worlds, and Information

The goal of this model is to capture important aspects of the way scientists exchange information. This goal has guided the assumptions to be stated in the following.

Let I be a set of scientists. Let Ω be a set of possible worlds the scientists may find themselves in. Ω may be of arbitrary cardinality. Think of I as being either finite or countably infinite. I will make more specific assumptions on the size of I along the way.

Suppose that there are m types of experiments that might be performed. Whenever an experiment is performed this generates some data. Suppose that these data can be summarized by some number or vector X_j. The possible values of X_j are elements of the set $\mathcal{X}_j$ $(j = 1, \ldots, m)$.

Each scientist $i \in I$ has an information set A_i which contains the information i has gathered through experimentation. i has done experiment j some (finite) number of times, call this number $n(i, j)$. Thus A_i contains $n(i, j)$ realizations of X_j for each j. I will write this as follows:

$$\begin{aligned} A_i &:= \{X_{j,i,k} \mid 1 \leq j \leq m, 1 \leq k \leq n(i,j)\} \\ &= \{X_{1,i,1}, \ldots, X_{1,i,n(i,1)}, X_{2,i,1}, \ldots, X_{2,i,n(i,2)}, \ldots, X_{m,i,1}, \ldots, X_{m,i,n(i,m)}\}, \end{aligned}$$

where $X_{j,i,k}$ is scientist i's k-th realization of experiment j. The scientists are assumed to view the realizations associated with a given experiment j as identically distributed random variables: $X_{j,i,k} \sim X_j$ for all i and k. The idea is that scientists are interested in finding out which possible world they are in, and observing the random variables in information sets provides information that is relevant to this goal. It is clear that the function $n(i, \cdot)$ characterizes the amount of information in the set A_i by specifying how many of each type of random variable it contains (of course $n(i, \cdot)$ does not specify the content of the information, only the amount).

Call the set of all information sets for a given set of scientists I A_I:

$$A_I := \{A_i \mid i \in I\}.$$

Definition 2 (scientific community)**.** A scientific community $\mathcal{C} = (I, A_I)$ is an ordered pair consisting of a set of scientists I and the set of information sets associated with the scientists in I.

Finally, assume that scientists are able to form one-way connections such that if a scientist connects to another scientist, she obtains the information in that other scientist's information set (but no other information, and the other scientist does not receive her information). There may be a cost associated with forming a connection (this will be made explicit in section 3.1).

A connection reflects one scientist learning the results of another scientist's experiments. The paradigm case I have in mind is reading those results in a paper published by the other scientist (but it may also reflect, e.g., an oral transfer of information).

I will assume that every scientist knows who has done what experiment how many times (i.e., every scientist knows $n(i, j)$ for all i and j). So when a scientist is considering to connect to another scientist she knows in advance which random variables she will learn the value of by making this connection.

2.2 Comparing Information Sets: a Simple Abstraction

I mentioned that an information set A_i is characterized by the number of realizations $n(i, j)$ of each experiment j that it contains. This section explores that idea. In particular, I want to have a way of talking about larger and smaller information sets.

When talking about finite sets, the notions of larger and smaller are captured by the subset relation. Unfortunately these notions cannot be applied directly in the framework I have set up. For example, consider scientist i, whose information set contains one realization each of experiments 1 and 2:

$$A_i = \{X_{1,i,1}, X_{2,i,1}\},$$

and scientist i', whose information set contains two realizations each of experiments 1 and 2:

$$A_{i'} = \{X_{1,i',1}, X_{1,i',2}, X_{2,i',1}, X_{2,i',2}\}.$$

I want to say that A_i is a subset of $A_{i'}$, but by the definition of subset this is not true, because the elements of A_i, $X_{1,i,1}$ and $X_{2,i,1}$, do not occur in $A_{i'}$. However, the elements of A_i have natural equivalents in $A_{i'}$: $X_{1,i',1}$ and $X_{2,i',1}$.

Thus, from the standpoint of a third scientist considering to connect to i or i', the two random variables in A_i are equivalent to the two singled-out random variables from $A_{i'}$, in the sense that *a priori* the information they will provide is indistinguishable for the third scientist. Since $A_{i'}$ contains additional information in the form of two more random variables, the third scientist should prefer to connect to i' rather than i.

To make these ideas more precise, it would be useful to have a notion of information set that abstracts away from particular instances of random variables, but focuses instead on the amount of information contained in them. For this purpose, I propose to call two random variables equivalent if they only differ in their second index (which specifies which scientist did the experiment and is as such irrelevant to the amount of information contained in the random variable). This allows me to define a more abstract notion of information set containing equivalence classes of random variables rather than the random variables themselves.

Definition 3 (equivalent random variables)**.** Two random variables $X_{j,i,k}$ and $X_{j',i',k'}$ are equivalent if $j = j'$ and $k = k'$. Define the associated equivalence classes as follows:

$$X_{j,k} := \{X_{j,i,k} \mid i \in I\}.$$

Definition 4 (abstracted information sets)**.** An abstracted information set $[A_i]$ contains the equivalence classes of the random variables contained in the information set A_i. That is, for any $i \in I$:

$$[A_i] := \{X_{j,k} \mid X_{j,i,k} \in A_i\} = \{X_{1,1}, \ldots, X_{1,n(i,1)}, \ldots, X_{m,1}, \ldots, X_{m,n(i,m)}\}.$$

Thus by looking at $[A_i]$ one can tell how many realizations of each of the m experiments are in A_i, without referring to their specific realizations. The following proposition and its corollary show that the normal notions of set containment and set equality, applied to abstracted information sets, accurately capture the amount of information in an information set.

Proposition 5. *For all* $i, i' \in I$*:*

$$[A_i] \subseteq [A_{i'}] \Leftrightarrow n(i,j) \leq n(i',j), \forall j \in \{1, \ldots, m\}.$$

Proof. ($\Rightarrow$) Assume $[A_i] \subseteq [A_{i'}]$. Let $j \in \{1, \ldots, m\}$ be any of the types of experiments. The set $[A_i]$ contains $X_{j,1}, \ldots, X_{j,n(i,j)}$. Therefore the set $[A_{i'}]$ contains $X_{j,1}, \ldots, X_{j,n(i,j)}$. By definition $[A_{i'}]$ contains $X_{j,1}, \ldots, X_{j,n(i',j)}$. But these facts are consistent only if $n(i,j) \leq n(i',j)$.

($\Leftarrow$) Assume $n(i,j) \leq n(i',j)$ for all $j \in \{1, \ldots, m\}$. Let $X_{j,k} \in [A_i]$. Then $k \leq n(i,j) \leq n(i',j)$. Therefore $X_{j,k} \in [A_{i'}]$. So $[A_i] \subseteq [A_{i'}]$. $\square$

Corollary 6. *For all* $i, i' \in I$*:*

$$[A_i] = [A_{i'}] \Leftrightarrow n(i,j) = n(i',j), \forall j \in \{1, \ldots, m\}.$$

This corollary, which follows immediately from proposition 5, expresses the idea that two information sets A_i and $A_{i'}$ contain essentially the same information (at least from the viewpoint of a third scientist considering to connect to i or i') if they contain the same number of realizations of each experiment.

The following two propositions show that the other standard set-theoretic relations (set union and intersection) also have sensible interpretations when applied to abstracted information sets.

Proposition 7. *For all* $i, i', i'' \in I$*:*

$$[A_{i''}] = [A_i] \cup [A_{i'}] \Leftrightarrow n(i'', j) = \max\{n(i,j), n(i',j)\}, \forall j \in \{1, \ldots, m\}.$$

Proof. ($\Rightarrow$) Assume $[A_{i''}] = [A_i] \cup [A_{i'}]$. Let $j \in \{1, \ldots, m\}$. Since $[A_{i''}]$ contains anything that is contained in either $[A_i]$ or $[A_{i'}]$, it must contain $X_{j,n(i,j)}$ and $X_{j,n(i',j)}$. So $[A_{i''}]$ contains $X_{j,\max\{n(i,j),n(i',j)\}}$. But $[A_{i''}]$ does not contain $X_{j,\max\{n(i,j),n(i',j)\}+1}$ because that is not contained in either $[A_i]$ or $[A_{i'}]$. So $n(i'', j) = \max\{n(i,j), n(i',j)\}$.

($\Leftarrow$) Assume $n(i'', j) = \max\{n(i,j), n(i',j)\}$ for all $j \in \{1, \ldots, m\}$. Let $X_{j,k} \in [A_{i''}]$. Then $k \leq n(i'', j) = \max\{n(i,j), n(i',j)\}$. So either $k \leq n(i,j)$ or $k \leq n(i', j)$ (or both). In the former case $X_{j,k} \in [A_i]$, while in the latter $X_{j,k} \in [A_{i'}]$. But then certainly $X_{j,k} \in [A_i] \cup [A_{i'}]$. So $[A_{i''}] \subseteq [A_i] \cup [A_{i'}]$.

Now let $X_{j,k} \in [A_i] \cup [A_{i'}]$. Then either $X_{j,k} \in [A_i]$ or $X_{j,k} \in [A_{i'}]$ (or both). In the former case $k \leq n(i,j) \leq \max\{n(i,j), n(i',j)\} = n(i'', j)$, while in the latter $k \leq n(i', j) \leq \max\{n(i,j), n(i',j)\} = n(i'', j)$. So $X_{j,k} \in [A_{i''}]$. So $[A_i] \cup [A_{i'}] \subseteq [A_{i''}]$. Together with $[A_{i''}] \subseteq [A_i] \cup [A_{i'}]$ this yields $[A_{i''}] = [A_i] \cup [A_{i'}]$. $\square$

Proposition 8. *For all* $i, i', i'' \in I$*:*

$$[A_{i''}] = [A_i] \cap [A_{i'}] \Leftrightarrow n(i'', j) = \min\{n(i,j), n(i',j)\}, \forall j \in \{1, \ldots, m\}.$$

The proof is essentially the same as for the previous proposition.

2.3 Relative Frequency of Information

Assuming that I is finite, the relative frequency of an abstracted information set can be straightforwardly defined.

Definition 9 (relative frequency of information)**.** A function q describes the relative frequency of information in a scientific community $\mathcal{C} = (I, A_I)$ (I will abbreviate this as "$\mathcal{C}$ satisfies q") if for any abstracted information set $[A]$

$$q([A]) = \frac{|\{i \in I \mid [A_i] = [A]\}|}{|I|}.$$

For a given $[A]$, $q([A])$ clearly expresses the proportion of scientists in I whose abstracted information set is equal to $[A]$. Let $[A_I]$ be the set of abstracted information sets occurring in the set of scientists I:

$$[A_I] := \{[A_i] \mid i \in I\}.$$

Then

$$\begin{aligned}
q([A]) &\geq 0 && \text{for any } [A], \\
q([A]) &> 0 && \text{if and only if } [A] \in [A_I], \\
\sum_{[A] \in [A_I]} q([A]) &= 1.
\end{aligned}$$

So any q that satisfies the above definition for some scientific community $\mathcal{C}$ is indeed a frequency distribution.

Note that if only finite sets of scientists are considered, the relative frequency of any abstracted information set must be a rational number ($q([A]) \in \mathbb{Q}$ for all $[A]$). Also, only finitely many abstracted information sets can have positive relative frequency ($[A_I]$ has finite size). When in subsequent sections I prove results "for all relative frequencies q" this should be understood as referring to any function q that satisfies the above definition for at least one scientific community $\mathcal{C}$ consisting of finitely many scientists.

Note also that if at least one scientific community satisfies q, infinitely many will do so (since any integer number of copies of the same community may together form a larger community that also satisfies q). Moreover, communities satisfying q may then have arbitrarily large sets of scientists I associated with them (by the same argument). This is important for the argument in subsequent sections.

2.4 Assumptions Concerning the Behavior of Scientists

Assume that the distribution of information q is given. I am interested in the connections that will form between the scientists. In order to prove anything about these, I will need some assumptions. I could do this by specifying the decision problem the scientists face and making assumptions about how they solve it. That is the approach I will take in chapter 3.

However, in this section I will only assume that each scientist i has somehow obtained a (sequential) decision procedure $\delta_i^{\mathcal{C}}$. $\delta_i^{\mathcal{C}}$ specifies which scientists i connects to as a function of $\mathcal{C}$, i.e., as a function of the set of available scientists I and their information sets A_I. As decisions are allowed to be made sequentially, the $n+1$-st connection made under $\delta_i^{\mathcal{C}}$ is also a function of the information obtained from the information sets of the first n scientists i connects to.

Taking the decision procedure $\delta_i^{\mathcal{C}}$ to be fixed, a given decision to connect is a function of the information gained through previous connections. Since that information takes the form of random variables, decisions to connect may themselves be viewed as random. This is the significance of statements I will make in the rest of this chapter about the probability of connecting to a given scientist, the probability of connecting to a certain number of scientists, or the expected number of connections.

These probabilistic statements take as fixed the set of scientists and the abstracted information set of each scientist (i.e., information about how many of each type of experiment each scientist has done), while they take as random the actual realizations of these random variables. So these probabilities reflect the assessment that a third-party observer would make before any connections are formed, knowing the experiments that each scientist has done, but not the outcomes of these experiments.

The decision procedures scientists follow may be arbitrarily good or bad (relative to the unspecified decision problem). In this section I will only give some minimal constraints on $\delta_i^{\mathcal{C}}$. These minimal behavioral assumptions turn out to be sufficient for the main result, theorem 13.

In a slight abuse of notation, I use $\delta_i^{\mathcal{C}}$ to denote both i's decision procedure and the total number of connections that i makes. The latter is a random variable, taking values in the non-negative integers, because it depends (in general) on the information obtained through the process.

Let $\delta_{i,[A]}^{\mathcal{C}}$ denote the number of scientists with abstracted information set $[A]$ that i connects to. Obviously, $\delta_{i,[A]}^{\mathcal{C}}$ is also a random variable and $\delta_{i,[A]}^{\mathcal{C}}$ is related to $\delta_i^{\mathcal{C}}$:

$$
\begin{aligned}
\delta_i^{\mathcal{C}} &= \sum_{[A]\in[A_I]} \delta_{i,[A]}^{\mathcal{C}}, && \text{and therefore} \\
\delta_{i,[A]}^{\mathcal{C}} &\leq \delta_i^{\mathcal{C}} && \text{for all } [A].
\end{aligned}
$$

In the following I state two assumptions that are used in the main result in the next section. The first assumption says that scientists are unlikely to connect to a very large number of scientists with the same abstracted information set. Moreover, the rate at which the probability for a large number of connections drops off is similar for each scientist, and similar regardless of how many scientists are available to connect to (similar enough that the rate can be uniformly bounded).

Assumption 10 (Uniformly Bounded Connection Probabilities)**.** *For any* *relative frequency q, abstracted information set $[A]$, and $\varepsilon > 0$, there exists $N_{[A],\varepsilon}$ such that for all $n > N_{[A],\varepsilon}$, for all $\mathcal{C}$ that satisfy q and for all $i \in I$, if $\Pr\left(\delta^{\mathcal{C}}_{i,[A]} \geq 1\right) > 0$ then*

$$n \Pr\left(\delta^{\mathcal{C}}_{i,[A]} \geq n \mid \delta^{\mathcal{C}}_{i,[A]} \geq 1\right) \leq \varepsilon.$$

A simple scenario where this assumption holds is when the number of connections is uniformly bounded almost surely (i.e., there exists an N such that for all $\mathcal{C}$ satisfying q and for all $i \in I$ $\Pr\left(\delta^{\mathcal{C}}_{i} \geq N\right) = 0$). This condition is sufficient for assumption 10. So if there is reason to believe that there is a number such that no scientist could benefit from making more connections than that number then assumption 10 is justified.

A slightly less simple scenario where this assumption is justified is when scientists behave like Bayesian statisticians and they all have the same loss function: in this case their optimal decision procedures will be identical. Section 3.3 will makes this more precise.

The following lemma can now be proved. It says that as the size of the set of scientists I gets large, the probability that a scientist i connects to all $q([A]) \cdot |I|$ scientists with abstracted information set $[A]$ gets small relative to the expected number of connections to scientists with abstracted information set $[A]$. This result is important for the proof of theorem 13, in fact it is only via lemma 11 that assumption 10 is used in the proof.

Lemma 11. *Assume that scientists behave according to assumption 10. For any relative frequency q and abstracted information set $[A]$ there exists $N_{[A]}$ such that for all $\mathcal{C}$ satisfying q with $|I| > N_{[A]}$, for all $i \in I$:*

$$q([A]) \cdot |I| \Pr\left(\delta^{\mathcal{C}}_{i,[A]} \geq q([A]) \cdot |I|\right) \leq \frac{1}{2}\mathbb{E}\left[\delta^{\mathcal{C}}_{i,[A]}\right].$$

Moreover, the inequality is strict for any $i \in I$ satisfying $\mathbb{E}\left[\delta^{\mathcal{C}}_{i,[A]}\right] > 0$.

Proof. Let $[A]$ be an abstracted information set. If $q([A]) = 0$ the desired inequality holds because both sides are zero. Assume $q([A]) > 0$. For $N_{[A]}$

choose the value $N_{[A],1/4}/q([A])$ (where $N_{[A],1/4}$ is $N_{[A],\varepsilon}$ as defined in assumption 10 for the value $\varepsilon = 1/4$). Let I be a set of scientists such that $I \subset_q I_\infty$ and $|I| > N_{[A]}$. Let $i \in I$ be a scientist. Distinguish two cases:

1. $\mathbb{E}\left[\delta^{\mathcal{C}}_{i,[A]}\right] > 0$. Since $\delta^{\mathcal{C}}_{i,[A]}$ is a nonnegative, integer-valued random variable, this is equivalent to $\Pr\left(\delta^{\mathcal{C}}_{i,[A]} \geq 1\right) > 0$. Since $q([A]) \cdot |I| > N_{[A],1/4}$, it follows from assumption 10 that

$$q([A]) \cdot |I| \Pr\left(\delta^{\mathcal{C}}_{i,[A]} \geq q([A]) \cdot |I| \mid \delta^{\mathcal{C}}_{i,[A]} \geq 1\right) \leq \frac{1}{4} < \frac{1}{2}.$$

By the definition of conditional probability this can also be written as

$$q([A]) \cdot |I| \Pr\left(\delta^{\mathcal{C}}_{i,[A]} \geq q([A]) \cdot |I|\right) < \frac{1}{2} \Pr\left(\delta^{\mathcal{C}}_{i,[A]} \geq 1\right).$$

Markov's inequality yields

$$\Pr\left(\delta^{\mathcal{C}}_{i,[A]} \geq 1\right) \leq \mathbb{E}\left[\delta^{\mathcal{C}}_{i,[A]}\right].$$

Combining the last two inequalities gives the desired (strict) inequality.

2. $\mathbb{E}\left[\delta^{\mathcal{C}}_{i,[A]}\right] = 0$. Since $\delta^{\mathcal{C}}_{i,[A]}$ is a nonnegative random variable, it follows that $\Pr\left(\delta^{\mathcal{C}}_{i,[A]} \geq n\right) = 0$ for all $n > 0$. In this case the desired inequality must hold because both sides are equal to zero. □

The second assumption says that scientists will not connect to someone with strictly less information than someone else.

Assumption 12 (Never Consider Subsets)**.** *A scientist will not connect to a second scientist i whenever a third scientist i' is available to connect to and $[A_i] \subset [A_{i'}]$ (i.e., $[A_i] \subseteq [A_{i'}]$ and $[A_i] \neq [A_{i'}]$).*

This assumption is plausible if each random variable in an information set gives independently valuable information. In such cases scientists behaving optimally will satisfy this assumption. Section 3.2 will make this more precise by giving conditions under which there is an optimal procedure for each scientist satisfying this assumption and giving (only slightly stronger) conditions under which all optimal procedures satisfy this assumption.

2.5 The Main Result: Supermodular In-Degrees

Consider the network formed by viewing the set of scientists as a set of nodes and drawing a (directed) edge between i and i' if i connects to i'. The in-degree of a scientist i' in a given network, denoted $d(i')$, is the number of scientists connecting to i':

$$d(i') := |\{i \in I \mid i \text{ connects to } i'\}|.$$

This number can be viewed as a measure of the prominence of an individual scientist in the community.

Since the scientists' decision procedures may depend on the information they gather as they go along, in general it does not follow with certainty from $\delta_i^{\mathcal{C}}$ whether or not i will connect to a given i'. But $\delta_i^{\mathcal{C}}$ specifies how likely it is that i will connect to i' (as a function of the probability distributions on the information sets). So one can define the notion of the expected in-degree of a scientist (where the expectation is relative to the probabilities of connecting induced by the decision procedures $\delta_i^{\mathcal{C}}$):

$$\mathbb{E}\left[d(i')\right] := \sum_{i \in I} \Pr\left(i \text{ connects to } i' \mid \delta_i^{\mathcal{C}}\right).$$

The use of the sum over i in this definition is unproblematic because each of the probabilities in the sum are independent. This is because i's decision procedure $\delta_i^{\mathcal{C}}$ is not allowed to depend on other scientists' connections: it depends only on the scientists that are available to connect to and their information sets.

Let $[A] \in [A_I]$ be an abstracted information set. The number of connections formed to scientists with abstracted information set $[A]$ is of course

$$\sum_{i' \in I : [A_{i'}] = [A]} d(i')$$

If $\mathcal{C}$ satisfies the relative frequency q there are $q([A]) \cdot |I|$ scientists with abstracted information set $[A]$. Therefore the average in-degree among such scientists is simply the above sum divided by $q([A]) \cdot |I|$. I will use this as a definition of the in-degree of an arbitrary scientist with abstracted information set $[A]$:

$$d([A]) := \frac{1}{q([A]) \cdot |I|} \sum_{i' \in I : [A_{i'}] = [A]} d(i').$$

But then the expected in-degree of an arbitrary scientist with abstracted information set $[A]$ is

$$\begin{aligned}
\mathbb{E}\left[d([A])\right] &:= \frac{1}{q([A]) \cdot |I|} \sum_{i' \in I : [A_{i'}] = [A]} \mathbb{E}\left[d(i')\right] \\
&= \frac{1}{q([A]) \cdot |I|} \sum_{i' \in I : [A_{i'}] = [A]} \sum_{i \in I} \Pr\left(i \text{ connects to } i' \mid \delta_i^{\mathcal{C}}\right) \\
&= \frac{1}{q([A]) \cdot |I|} \sum_{i \in I} \sum_{i' \in I : [A_{i'}] = [A]} \Pr\left(i \text{ connects to } i' \mid \delta_i^{\mathcal{C}}\right) \\
&= \frac{1}{q([A]) \cdot |I|} \sum_{i \in I} \mathbb{E}\left[\delta_{i,[A]}^{\mathcal{C}}\right], \qquad (1)
\end{aligned}$$

where $\delta_{i,[A]}^{\mathcal{C}}$ is the number of scientists with abstracted information set $[A]$ that i connects to, as discussed in section 2.4. Note that the latter two definitions are only meaningful if $q([A]) > 0$. Define $d([A])$ and $\mathbb{E}[d([A])]$ to be zero whenever $q([A]) = 0$.

The sense in which the scientist discussed above is arbitrary is that if you collected all the scientists with abstracted information set $[A]$ together and drew one of them at random the above is the expected in-degree of that randomly drawn scientist. Moreover if scientists are indifferent whether they connect to one scientist or another whenever they have the same abstracted information set (as Bayesian scientists minimizing their risk function would be, see lemma 21) and break such ties by choosing randomly which scientist with the same abstracted information set to connect to, then $\mathbb{E}[d([A])] = \mathbb{E}[d(i')]$ for all scientists i' with abstracted information set $[A]$. In this case the expected in-degree is the same for all scientists with the same abstracted information set and $\mathbb{E}[d([A])]$ measures the prominence in the community of scientists with abstracted information set $[A]$. Even if this is not the case, however, $\mathbb{E}[d([A])]$ is useful as a measure of the average prominence of scientists as a function of their abstracted information set $[A]$.

With all of this in place, I can now state the main theorem. It states that if the set of scientists is sufficiently large, the expected prominence of a given scientist increases rapidly (faster than linearly) in the size of her information set.

Theorem 13. *For any relative frequency q there exists a number N such that for all communities $\mathcal{C}$ satisfying q and assumptions 10 and 12, if $|I| > N$ then for all abstracted information sets $[A]$ and $[B]$ with $q([A] \cup [B]) > 0$*

$$\mathbb{E}[d([A] \cup [B])] + \mathbb{E}[d([A] \cap [B])] \geq \mathbb{E}[d([A])] + \mathbb{E}[d([B])].$$

If the sets involved are distinct (that is, neither $[A]$ nor $[B]$ is equal to $[A] \cup [B]$) then this can be strengthened to

$$\mathbb{E}[d([A] \cup [B])] \geq \mathbb{E}[d([A])] + \mathbb{E}[d([B])].$$

And if in addition $\mathbb{E}[d([A] \cup [B])] > 0$ then this can be further strengthened to

$$\mathbb{E}[d([A] \cup [B])] > \mathbb{E}[d([A])] + \mathbb{E}[d([B])].$$

Proof. Let q be a relative frequency. Recall from section 2.3 that only finitely many abstracted information sets can have positive relative frequency under q. Choose

$$N = \max_{[A]:q([A])>0} N_{[A]},$$

where $N_{[A]}$ is as defined in lemma 11. Let $\mathcal{C}$ be a scientific community satisfying q and assumptions 10 and 12. Assume that $|I| > N$ and let $[A]$ and $[B]$ be any two abstracted information sets such that $q([A] \cup [B]) > 0$.

As in the statement of the theorem, I distinguish two cases: either (1) at least one of the sets $[A]$ or $[B]$ is equal to $[A] \cup [B]$, or (2) both $[A]$ and $[B]$ are proper subsets of $[A] \cup [B]$. I will prove the theorem for the two cases separately.

1. Assume without loss of generality that $[A] = [A] \cup [B]$. It follows that $[A] \cap [B] = ([A] \cup [B]) \cap [B] = [B]$. So $\mathbb{E}[d([A])] = \mathbb{E}[d([A] \cup [B])]$ and $\mathbb{E}[d([B])] = \mathbb{E}[d([A] \cap [B])]$. The result follows immediately (in this case there is equality between the two sides).

2. For the second case it is assumed that $[A] \subset [A] \cup [B]$ and $[B] \subset [A] \cup [B]$. If $q([A]) > 0$ I can use equation (1) to get a formula for $\mathbb{E}[d([A])]$:

$$\mathbb{E}[d([A])] = \frac{1}{q([A]) \cdot |I|} \sum_{i \in I} \sum_{i' \in I : [A_{i'}] = [A]} \Pr\left(i \text{ connects to } i' \mid \delta_i^{\mathcal{C}}\right).$$

Since $[A] \subset [A] \cup [B]$, assumption 12 applies. This says scientist $i \in I$ will not connect to scientist i' (with $[A_{i'}] = [A]$) as long as there are scientists with abstracted information set $[A] \cup [B]$ available to connect to. So i may connect to i' only if she has connected to all $q([A] \cup [B]) \cdot |I|$ scientists with abstracted information set $[A] \cup [B]$:

$$\Pr\left(i \text{ connects to } i' \mid \delta_i^{\mathcal{C}}\right) \leq \Pr\left(\delta_{i,[A]\cup[B]}^{\mathcal{C}} \geq q([A] \cup [B]) \cdot |I|\right),$$

where $\delta_{i,[A]\cup[B]}^{\mathcal{C}}$ represents the number of scientists with abstracted information set $[A] \cup [B]$ that i connects to. Applying this to $\mathbb{E}\left[d([A])\right]$ yields

$$\begin{aligned}\mathbb{E}\left[d([A])\right] &= \frac{1}{q([A]) \cdot |I|} \sum_{i \in I} \sum_{i' \in I : [A_{i'}] = [A]} \Pr\left(i \text{ connects to } i' \mid \delta_i^{\mathcal{C}}\right) \\ &\leq \frac{1}{q([A]) \cdot |I|} \sum_{i \in I} \sum_{i' \in I : [A_{i'}] = [A]} \Pr\left(\delta_{i,[A] \cup [B]}^{\mathcal{C}} \geq q([A] \cup [B]) \cdot |I|\right) \\ &= \sum_{i \in I} \Pr\left(\delta_{i,[A] \cup [B]}^{\mathcal{C}} \geq q([A] \cup [B]) \cdot |I|\right), \end{aligned} \tag{2}$$

where the final equality follows because the probabilities in the sum do not depend on i' and the inner sum contains $q([A]) \cdot |I|$ terms.

Inequality (2) has been obtained under the assumption that $q([A]) > 0$, but the bound on $\mathbb{E}[d([A])]$ is true when $q([A]) = 0$ as well: in that case $\mathbb{E}[d([A])] = 0$, which is clearly no bigger than a sum of probabilities.

Because $[B] \subset [A] \cup [B]$, an analogous argument (which, for the reason just given, works regardless of whether $q([B]) > 0$ or $q([B]) = 0$) allows me to bound $\mathbb{E}[d([B])]$:

$$\mathbb{E}\left[d([B])\right] \leq \sum_{i \in I} \Pr\left(\delta_{i,[A] \cup [B]}^{\mathcal{C}} \geq q([A] \cup [B]) \cdot |I|\right). \tag{3}$$

Since assumption 10 is satisfied and $|I| > N \geq N_{[A] \cup [B]}$ all of the conditions of lemma 11 are satisfied. Hence, for all $i \in I$,

$$q([A] \cup [B]) \cdot |I| \Pr\left(\delta_{i,[A] \cup [B]}^{\mathcal{C}} \geq q([A] \cup [B]) \cdot |I|\right) \leq \frac{1}{2} \mathbb{E}\left[\delta_{i,[A] \cup [B]}^{\mathcal{C}}\right], \tag{4}$$

where the inequality is strict for any $i \in I$ satisfying $\mathbb{E}\left[\delta_{i,[A] \cup [B]}^{\mathcal{C}}\right] > 0$. By the definition of $\mathbb{E}\left[d([A] \cup [B])\right]$ (per equation (1)):

$$\mathbb{E}\left[d([A] \cup [B])\right] = \frac{1}{q([A] \cup [B]) \cdot |I|} \sum_{i \in I} \mathbb{E}\left[\delta_{i,[A] \cup [B]}^{\mathcal{C}}\right].$$

If $\mathbb{E}\left[d([A] \cup [B])\right] > 0$ it follows that $\mathbb{E}\left[\delta^{\mathcal{C}}_{i,[A]\cup[B]}\right] > 0$ for at least one $i \in I$. Together with inequality 4 this yields

$$\mathbb{E}\left[d([A] \cup [B])\right] \geq 2 \sum_{i \in I} \Pr\left(\delta^{\mathcal{C}}_{i,[A]\cup[B]} \geq q([A] \cup [B]) \cdot |I|\right),$$

where the inequality is strict if $\mathbb{E}\left[d([A] \cup [B])\right] > 0$. Combining this result with inequalities (2) and (3) obtained above gives:

$$\mathbb{E}\left[d([A] \cup [B])\right] \geq \mathbb{E}\left[d([A])\right] + \mathbb{E}\left[d([B])\right].$$

The inequality is strict whenever $\mathbb{E}\left[d([A] \cup [B])\right] > 0$, as required. □

There are some limitations to this result. Scientists have to behave in accordance with assumptions 10 and 12 (although assumption 10 is only used through lemma 11). And the distribution of information q needs to be kept constant as I increases in size: it is easy to construct an example where the result fails for any finite I if abstracted information sets can become increasingly rare as I gets bigger.

2.6 A Corollary: Exponential In-Degrees

Theorem 13 shows that expected in-degree is a supermodular function of the size of a scientist's information set. As I alluded to, supermodularity can be thought of as "faster than linear". This section makes this more precise. I prove a corollary to theorem 13 that shows that, in "dense" scientific communities, expected in-degree increases exponentially as the size of a scientist's information set increases.

Recall that scientist i's abstracted information set $[A_i]$ contains $n(i, j)$ equivalence classes of random variables for each experiment j. Let E be some subset of the experiments and let $\hat{n}^{[A_i]}_E = \min\{n(i, j) \mid j \in E\}$, i.e., $\hat{n}^{[A_i]}_E$ is the lowest number of times any of the experiments in E was performed by scientist i. Equivalently, $\hat{n}^{[A_i]}_E$ is the largest number such that each experiment in E was performed at least $\hat{n}_E^{[A_i]}$ times by scientist i.

Let q be a relative frequency. As before, let $[A_I]$ be the set of abstracted information sets that have positive relative frequency under q. Define

$$\hat{n}_E = \max_{[A] \in [A_I]} \hat{n}_E^{[A]}.$$

So $\hat{n}_E$ denotes the largest number such that a scientific community satisfying q contains at least one scientist who has performed each experiment in E at least $\hat{n}_E$ times.

Suppose scientist i has performed each experiment in E at least once and at most $\hat{n}_E$ times, but has performed no other experiments, i.e., $1 \leq n(i,j) \leq \hat{n}_E$ if $j \in E$ and $n(i,j) = 0$ if $j \notin E$. As a result, there are $\hat{n}_E^{|E|}$ possible abstracted information sets this scientist might have. I will say a relative frequency is "dense" if all of these combinations occur in it with positive frequency.

Assumption 14 (dense relative frequency)**.** *A relative frequency q (or a scientific community $\mathcal{C}$ satisfying q) is dense relative to a subset of the experiments E if for any combination of numbers $n_1, \ldots, n_m$ (where $1 \leq n_j \leq \hat{n}_E$ if $j \in E$ and $n_j = 0$ if $j \notin E$) there is an abstracted information set $[A_i]$ such that $q([A_i]) > 0$ and $n(i,j) = n_j$ for all j.*

Let $[A_E(n)]$ denote the abstracted information set containing exactly n realizations of each experiment in E, and no other experiments. The following result, a corollary of theorem 13, shows that the expected prominence of a scientist with abstracted information set $[A_E(n)]$ increases exponentially with n.

Corollary 15. *Let q be any relative frequency and E any subset of experiments with $|E| > 1$. There exists a number N such that for all communities $\mathcal{C}$ satisfying q and assumptions 10, 12 and 14, if $|I| > N$ then for all n (with $1 \leq n \leq \hat{n}_E$)*

$$\mathbb{E}\left[d([A_E(\hat{n}_E)])\right] \geq 2^{(\hat{n}_E-n)(|E|-1)}\mathbb{E}\left[d([A_E(n)])\right] \geq 2^{(\hat{n}_E-1)(|E|-1)}\mathbb{E}\left[d([A_E(1)])\right].$$

Proof. It suffices to show that under the conditions stated,

$$\mathbb{E}\left[d([A_E(n)])\right] \geq 2^{|E|-1}\mathbb{E}\left[d([A_E(n-1)])\right]$$

for all $2 \leq n \leq \hat{n}_E$: the result then follows by iteratively applying this inequality.

Label the experiments in E as $1, 2, \ldots, |E|$. Write $[A_E]$ for $[A_E(n)]$, i.e., the abstracted information set where each experiment in E is performed n times. For any $F \subset E$, write $[A_F]$ for the abstracted information set where each experiment in F is performed n times and each experiment in $E \setminus F$ is performed $n-1$ times. Note that with this notation $[A_\emptyset] = [A_E(n-1)]$ and for any nonempty F, $[A_\emptyset] \subset [A_F] \subset [A_E]$ by proposition 5. Consider $[A_{E\setminus\{1\}}]$ and $[A_{E\setminus\{2\}}]$. By proposition 7,

$$[A_E] = [A_{E\setminus\{1\}}] \cup [A_{E\setminus\{2\}}].$$

By assumption 14, $q([A_E]) > 0$. Hence, theorem 13 applies, yielding

$$\mathbb{E}\left[d([A_E])\right] \geq \mathbb{E}\left[d([A_{E\setminus\{1\}}])\right] + \mathbb{E}\left[d([A_{E\setminus\{2\}}])\right].$$

If $|E| = 2$ this finishes the proof because $\mathbb{E}[d([A_{E\setminus\{1\}}])] \geq \mathbb{E}[d([A_\emptyset])]$ and $\mathbb{E}[d([A_{E\setminus\{2\}}])] \geq \mathbb{E}[d([A_\emptyset])]$, and hence

$$\mathbb{E}\left[d([A_E])\right] \geq 2\mathbb{E}\left[d([A_\emptyset])\right].$$

If $|E| > 2$ the same strategy can be applied again. By proposition 7,

$$
\begin{aligned}
[A_{E\setminus\{1\}}] &= [A_{E\setminus\{1,2\}}] \cup [A_{E\setminus\{1,3\}}], \\
[A_{E\setminus\{2\}}] &= [A_{E\setminus\{1,2\}}] \cup [A_{E\setminus\{2,3\}}].
\end{aligned}
$$

By assumption 14, $q([A_{E\setminus\{1\}}]) > 0$ and $q([A_{E\setminus\{2\}}]) > 0$, and hence by theorem 13

$$
\begin{aligned}
\mathbb{E}\left[d([A_E])\right] &\geq \mathbb{E}\left[d([A_{E\setminus\{1\}}])\right] + \mathbb{E}\left[d([A_{E\setminus\{2\}}])\right] \\
&\geq 2\mathbb{E}\left[d([A_{E\setminus\{1,2\}}])\right] + \mathbb{E}\left[d([A_{E\setminus\{1,3\}}])\right] + \mathbb{E}\left[d([A_{E\setminus\{2,3\}}])\right] \\
&\geq 4\mathbb{E}\left[d([A_\emptyset])\right].
\end{aligned}
$$

In general, this strategy can be repeated $|E| - 1$ times, taking out one performance of an experiment each time and doubling the number of terms in the sum, yielding

$$
\mathbb{E}\left[d([A_E])\right] \geq 2^{|E|-1}\mathbb{E}\left[d([A_\emptyset])\right]. \qquad \square
$$

3 When Bayesian Scientists Satisfy the Assumptions

3.1 A Sequential Decision Problem

So far I have showed that under two assumptions concerning the behavior of scientists, the links formed by scientists in my model display a certain asymmetry: a few scientists get contacted many times, while most get contacted zero or few times. I argued that the two behavioral assumptions are plausible if scientists are at least somewhat rational (see section 2.4).

Now I will instead assume that scientists are fully rational Bayesian statisticians. By showing that such scientists satisfy the two behavioral assumptions of section 2.4 (given some conditions), I prove that the links formed by these fully rational scientists display the same asymmetry.

Let $\mathcal{C} = (I, A_I)$ be a scientific community and let Ω be the set of possible worlds the scientists may find themselves in. Let Ξ denote the set of all possible probability distributions defined on some suitable σ-field $\mathcal{W}$ of subsets of Ω. So for each $\xi \in \Xi$, $\xi : \mathcal{W} \to [0, 1]$ is a probability function and the triple $(\Omega, \mathcal{W}, \xi)$ is a probability space.

I will use $\xi \in \Xi$ to denote an individual scientist's subjective assessment at a given time of how likely it is that she is in a given (set of) world(s). If the scientist learns some information she updates her subjective beliefs by Bayes conditioning. Let $\xi(A_1, \ldots, A_n)$ denote the probability measure obtained by Bayes conditioning ξ with respect to the random variables in the information sets $A_1, \ldots, A_n$. Note that $\xi(A_1, \ldots, A_n) \in \Xi$.

Recall that $A_I = \{A_i \mid i \in I\}$ is the set of information sets, with

$$A_i = \{X_{j,i,k} \mid 1 \leq j \leq m, 1 \leq k \leq n(i, j)\}.$$

Each random variable $X_{j,i,k}$ represents a realization of experiment j, so it follows some given distribution X_j with possible outcomes $\mathcal{X}_j$. At some points, I will make an additional assumption to the effect that any collection of realizations of experiment j forms an i.i.d. dataset given any relevant set of possible worlds, and independent of the realizations of the other experiments.

Definition 16 (simple scientific community)**.** Call the scientific community $\mathcal{C} = (I, \{A_i \mid i \in I\})$ simple relative to a probability space $\mathcal{P} = (\Omega, \mathcal{W}, \xi)$ if for all $W \in \mathcal{W}$ $X_{j,i,k}$ is independent of $X_{j',i',k'}$ given W unless $i = i'$, $j = j'$, and $k = k'$ (conditional independence) and $X_{j,i,k}|W \sim X_j|W$ (conditional identical distributions).

This does not and should not deny that the distributions X_j may depend on the possible world the scientist is in: presumably the purpose of gathering data is to learn about what type of world she is in.

The scientists face the following sequential decision problem: they have to choose some decision from a set of options D. In an epistemological context a simple example would be "announce what possible world you are in" but in general the relation between D and Ω may be more complicated. D may be different for each scientist, but I will not emphasize this fact by putting an index on D.

Before scientists make their decision they are allowed to connect to other scientists in I. Connecting to a scientist i costs c for each connection, but yields all of the information in i's information set A_i. This is a so-called "one way, one pays" link: the scientist who pays c receives the data obtained from any experiments the other scientist has done, but the scientist who does not pay receives no new information.

Definition 17 (decision problems and procedures)**.** A decision problem $\mathcal{D} = (\mathcal{C}, D)$ is an ordered pair consisting of a scientific community $\mathcal{C}$ and a set D of terminal decisions. A (sequential) decision procedure δ for $\mathcal{D}$ is a function that outputs a sequence $i_1, i_2, \ldots$ of members of I and a terminal decision $d_\delta \in D$. i_{n+1} may depend on the information in the information sets $A_{i_1}, \ldots, A_{i_n}$. d_δ may depend on $A_{i_1}, A_{i_2}, \ldots$. Let $\Delta_{\mathcal{D}}$ denote the set of all decision procedures for $\mathcal{D}$.

Before moving on I give two side remarks on the nature of the sequential decision problem. First, because scientists receive only one information set by connecting (not the data the other scientist has obtained by connecting to yet other scientists) and are not affected by who connects to them, an individual scientist's performance is not affected by other scientists' strategy. So I do not need to use full-blown game theory; individual decision theory suffices.

Second, this is a sequential decision problem because I assume a scientist can connect to a scientist and see the information in her information set before deciding whether to connect to another scientist.

Next, I need to introduce a way of evaluating the outcome of the sequential decision problem. This will depend on the loss associated with the terminal decision and the cost of connecting.

Definition 18 (sequential risk function)**.** Let $\mathcal{P}$ be a probability space and $\mathcal{D}$ a decision problem. Let $\ell(w, d)$ be the loss associated with terminal decision $d \in D$ in world $w \in \Omega$. The risk of an immediate decision is defined (relative to $\mathcal{P}$) as

$$\rho_0(\xi, d) = \int_\Omega \ell(w, d)\, d\xi(w)$$

for all $d \in D$.

Let $c > 0$. Then the risk under loss ℓ and cost of connecting c is defined (relative to $\mathcal{P}$) as

$$\rho(\xi, \delta) = \mathbb{E}\left[\rho_0(\xi(A_{i_1}, A_{i_2}, \dots, A_{i_\delta}), d_\delta) + c\delta\right] \tag{5}$$

for all $\delta \in \Delta_{\mathcal{D}}$ (where, as before, δ is used to denote the number of connections made under procedure δ).

So the prior risk associated with a procedure δ is the expected risk of an immediate decision relative to the subjective beliefs the scientist will have after she finished connecting, plus c times the number of connections made. ξ, ℓ, ρ_0, and ρ may be different for each scientist, but as with D I will suppress this fact notationally.

3.2 Existence and Properties of the Optimal Procedure

Let $\Delta_{\mathcal{D}}$ denote the set of all possible sequential decision procedures that a given scientist could follow when faced with the type of problem I described in section 3.1. Chow and Robbins (1963) prove that this set contains an optimal procedure.

Theorem 19 (Chow and Robbins (1963))**.** *Let $\mathcal{P} = (\Omega, \mathcal{W}, \xi)$ be a probability space and $\mathcal{D} = (\mathcal{C}, D)$ a decision problem. Let ρ be a risk function (specified relative to $\mathcal{P}$ for a given loss function ℓ and cost of connecting $c > 0$). Then there exists a $\delta^* \in \Delta_{\mathcal{D}}$ such that*

$$\rho(\xi, \delta^*) = \inf_{\delta \in \Delta_{\mathcal{D}}} \rho(\xi, \delta).$$

The proof is rather long and can be found in Chow and Robbins (1963, section 2). A proof for the specific case that I am considering, and using the same notation that I use, is in DeGroot (2004, section 12.9).

Let B be an information set. If a scientist with prior ξ learns the information in B she updates to $\xi(B)$. As noted above, $\xi(B) \in \Xi$, so $(\Omega, \mathcal{W}, \xi(B))$ can be used as the probability space in the theorem above. So by the theorem there exists an optimal procedure in $\Delta_{\mathcal{D}}$ for a scientist with prior $\xi(B)$. Let $\delta_B^{\mathcal{D}} \in \Delta_{\mathcal{D}}$ denote such an optimal procedure (this is a random variable because B is random). This procedure is optimal for decision problem $\mathcal{D}$ relative to prior $\xi(B)$. An optimal procedure relative to ξ can then be written $\delta_\emptyset^{\mathcal{D}}$, since ξ conditioned on an empty information set is equal to ξ.

Now I will prove an elementary result about the value of information due to Good (1967). The result guarantees that getting free information never makes a scientist worse off, and gives a necessary and sufficient condition under which the information actually makes her better off.

Theorem 20 (Good (1967))**.** *Let $\mathcal{P} = (\Omega, \mathcal{W}, \xi)$ be a probability space and $\mathcal{D} = (\mathcal{C}, D)$ a decision problem. Let ρ be a risk function (specified relative to $\mathcal{P}$ for a given loss function ℓ and cost of connecting $c > 0$). Then the optimal risk is at least as high as the expected optimal risk after conditioning on some arbitrary information set B:*

$$\rho\left(\xi, \delta_\emptyset^{\mathcal{D}}\right) \geq \mathbb{E}\left[\rho\left(\xi(B), \delta_B^{\mathcal{D}}\right)\right].$$

The inequality is strict if and only if there is a set of possible outcomes for

B with positive probability for which

$$\rho\left(\xi(B), \delta_\emptyset^{\mathcal{D}}\right) > \rho\left(\xi(B), \delta_B^{\mathcal{D}}\right),$$

i.e., if conditional on B $\delta_\emptyset^{\mathcal{D}}$ is no longer optimal.

Proof. Note first that the existence of $\delta_\emptyset^{\mathcal{D}}$ and $\delta_B^{\mathcal{D}}$ (for any realization of B) is guaranteed by theorem 19.

Now observe that $\rho(\xi, \delta_\emptyset^{\mathcal{D}}) =$

$$\begin{aligned}
&= \mathbb{E}\left[\rho_0\left(\xi\left(A_1, \ldots, A_{\delta_\emptyset^{\mathcal{D}}}\right), d_{\delta_\emptyset^{\mathcal{D}}}\right) + c\delta_\emptyset^{\mathcal{D}}\right] && \text{(definition of } \rho(\xi, \delta_\emptyset^{\mathcal{D}})\text{)}\\
&= \mathbb{E}\left[\mathbb{E}\left[\rho_0\left(\xi\left(A_1, \ldots, A_{\delta_\emptyset^{\mathcal{D}}}\right), d_{\delta_\emptyset^{\mathcal{D}}}\right) + c\delta_\emptyset^{\mathcal{D}} \mid B\right]\right] && \text{(iterated expectation)}\\
&= \mathbb{E}\left[\mathbb{E}\left[\rho_0\left(\xi\left(A_1, \ldots, A_{\delta_\emptyset^{\mathcal{D}}}, B\right), d_{\delta_\emptyset^{\mathcal{D}}}\right) + c\delta_\emptyset^{\mathcal{D}}\right]\right] && \text{(Bayesian updating)}\\
&= \mathbb{E}\left[\mathbb{E}\left[\rho_0\left(\xi\left(B, A_1, \ldots, A_{\delta_\emptyset^{\mathcal{D}}}\right), d_{\delta_\emptyset^{\mathcal{D}}}\right) + c\delta_\emptyset^{\mathcal{D}}\right]\right] && \text{(conditioning commutes)}\\
&= \mathbb{E}\left[\rho\left(\xi(B), \delta_\emptyset^{\mathcal{D}}\right)\right]. && \text{(definition of } \rho(\xi(B), \delta_\emptyset^{\mathcal{D}})\text{)}
\end{aligned}$$

For a fixed realization of B, by definition of the infimum,

$$\rho\left(\xi(B), \delta_\emptyset^{\mathcal{D}}\right) \geq \inf_{\delta \in \Delta} \rho(\xi(B), \delta) = \rho\left(\xi(B), \delta_B^{\mathcal{D}}\right).$$

Putting these two together yields

$$\rho\left(\xi, \delta_\emptyset^{\mathcal{D}}\right) = \mathbb{E}\left[\rho\left(\xi(B), \delta_\emptyset^{\mathcal{D}}\right)\right] \geq \mathbb{E}\left[\rho\left(\xi(B), \delta_B^{\mathcal{D}}\right)\right]$$

as desired. It is now also obvious that the inequality above reduces to an equality if and only if with probability one $\delta_\emptyset^{\mathcal{D}}$ is still an optimal procedure after observing B. Otherwise the inequality is strict, yielding the condition stated in the theorem. □

Recall that section 2.2 defined two random variables to be equivalent if they differ only in their second index (which specifies which scientist it belongs to). This yields equivalence classes of random variables defined as follows: $X_{j,k} := \{X_{j,i,k} \mid i \in I\}$. The abstracted information set of a scientist i, denoted $[A_i]$, specifies i's information through the associated equivalence classes: $[A_i] := \{X_{j,k} \mid X_{j,i,k} \in A_i\}$.

Corollary 6 showed that two information sets contain the same number of realizations of each experiment if and only if their associated abstracted information sets are identical. Intuitively, this means that a scientist considering to connect to one of two such information sets should be indifferent: the information she will obtain is *a priori* indistinguishable. The following lemma makes this precise.

Lemma 21. *Let $\mathcal{P} = (\Omega, \mathcal{W}, \xi)$ be a probability space and $\mathcal{D} = (\mathcal{C}, D)$ a decision problem, with $\mathcal{C}$ simple relative to $\mathcal{P}$. Let ρ be a risk function (specified relative to $\mathcal{P}$ for a given loss function ℓ and cost of connecting $c > 0$). If two information sets A_i and $A_{i'}$ are such that $[A_i] = [A_{i'}]$ then the risk relative to $\xi(A_i)$ and the risk relative to $\xi(A_{i'})$ are the same in expectation for any decision procedure. That is, for any $\delta \in \Delta_{\mathcal{D}}$:*

$$\mathbb{E}[\rho(\xi(A_i), \delta)] + c = \mathbb{E}[\rho(\xi(A_{i'}), \delta)] + c.$$

Proof. Since $[A_i] = [A_{i'}]$, by corollary 6, $n(i, j) = n(i', j)$ for all j. So A_i and $A_{i'}$ contain the same number of random variables of each type j.

Let $S := \{S_{j,k} \mid S_{j,k} \subseteq \mathcal{X}_j, 1 \leq j \leq m, 1 \leq k \leq n(i, j)\}$ denote a set of possible outcomes for A_i (by specifying a set of possible outcomes $S_{j,k}$ for each $X_{j,i,k} \in A_i$). Then S is also a set of possible outcomes for $A_{i'}$ (because $A_{i'}$ contains the same number of each type of random variable).

Let $W \in \mathcal{W}$. Because $X_{j,i,k} \mid W \sim X_{j,i',k} \mid W$ for all relevant j and k it follows that

$$\Pr(X_{j,i,k} \in S_{j,k} \mid W) = \Pr(X_{j,i',k} \in S_{j,k} \mid W)$$

for all relevant j and k. Therefore (using the conditional independence that follows from $\mathcal{C}$ being simple):

$$\Pr(A_i \in S \mid W) = \Pr(A_{i'} \in S \mid W).$$

So A_i and $A_{i'}$ are identically distributed in all relevant sets of worlds. It follows that $\xi(A_i) = \xi(A_{i'})$ whenever their realizations match ($X_{j,i,k} = X_{j,i',k}$ for all relevant j and k). So the scientist learns from the information set A_i in exactly the same way as she would from $A_{i'}$. Moreover, by conditional independence, what she learns from information sets she connects to in the future is not affected either, i.e., for all δ,

$$\xi(A_i, A_1, \ldots, A_\delta) = \xi(A_{i'}, A_1, \ldots, A_\delta)$$

whenever $X_{j,i,k} = X_{j,i',k}$ for all relevant j and k. The result follows. $\square$

It may be objected that the lemma assumes that the decision problem faced by the scientist after connecting to i or i' is the same, while in fact the set of available information sets should be different: $A_{I\setminus\{i\}}$ after connecting to i as opposed to $A_{I\setminus\{i'\}}$ after connecting to i'.

I argue that this does not actually matter: after connecting to i or i' the only difference in the resulting decision problems is whether i' or i (respectively) is still available to connect to, and the lemma guarantees that the scientists views those two options as equivalent. So the lemma really does say that a scientist should be indifferent between connecting to i and i' when $[A_i] = [A_{i'}]$.

The following result is obtained using lemma 21 and theorem 20. It says that connecting to an information set with more information instead of one with less information never makes a scientist worse off, and gives a necessary and sufficient condition such that it makes the scientist better off.

Theorem 22. *Let $\mathcal{P} = (\Omega, \mathcal{W}, \xi)$ be a probability space and $\mathcal{D} = (\mathcal{C}, D)$ a decision problem, with $\mathcal{C}$ simple relative to $\mathcal{P}$. Let ρ be a risk function (specified relative to $\mathcal{P}$ for a given loss function ℓ and cost of connecting $c >$*

0*). Suppose that i's information set contains at least as much information as i''s: $A_i \subseteq A_{i'}$. Then*

$$\mathbb{E}\left[\rho\left(\xi(A_i), \delta^{\mathcal{D}}_{A_i}\right)\right] + c \geq \mathbb{E}\left[\rho\left(\xi(A_{i'}), \delta^{\mathcal{D}}_{A_{i'}}\right)\right] + c.$$

The inequality is strict if and only if there is a set of possible outcomes of $A_{i'}$ with positive probability such that if $X_{j,i,k} = X_{j,i',k}$ for all $1 \leq j \leq m$ and for all $1 \leq k \leq n(i,j)$ (i.e., A_i matches $A_{i'}$ in their common part), $\delta^{\mathcal{D}}_{A_i}$ is not optimal when conditioning on $A_{i'}$:

$$\rho\left(\xi(A_{i'}), \delta^{\mathcal{D}}_{A_i}\right) > \rho\left(\xi(A_{i'}), \delta^{\mathcal{D}}_{A_{i'}}\right).$$

Proof. First off, note that the existence of the optimal procedures $\delta^{\mathcal{D}}_{A_i} \in \Delta_{\mathcal{D}}$ and $\delta^{\mathcal{D}}_{A_{i'}} \in \Delta_{\mathcal{D}}$ is guaranteed by theorem 19 (for any possible realizations of A_i and $A_{i'}$).

Next, define B as follows:

$$\begin{aligned} B &:= \{X_{j,i,k} \mid 1 \leq j \leq m, n(i,j) < k \leq n(i',j)\} \\ &= \{X_{1,i,n(i,1)+1}, \ldots, X_{1,i,n(i',1)}, \ldots, X_{m,i,n(i,m)+1}, \ldots, X_{m,i,n(i',m)}\}. \end{aligned}$$

So for each j, B contains $n(i',j) - n(i,j)$ (which is nonnegative by proposition 5) realizations of X_j and they are indexed such that A_i is disjoint from B. In a sense, B contains the information that A_i is missing relative to $A_{i'}$: $[A_i \cup B] = [A_{i'}]$. So by lemma 21, for all $\delta \in \Delta_{\mathcal{D}}$:

$$\mathbb{E}\left[\rho\left(\xi\left(A_{i'}\right), \delta\right)\right] + c = \mathbb{E}\left[\rho\left(\xi\left(A_i \cup B\right), \delta\right)\right] + c.$$

It follows that the optimal risk should also be the same. So let $\delta^{\mathcal{D}}_{A_i \cup B} \in \Delta_{\mathcal{D}}$ denote the optimal procedure (which exists by theorem 19) relative to $\xi(A_i \cup B)$ given a realization of $A_i \cup B$, to get

$$\mathbb{E}\left[\rho\left(\xi\left(A_{i'}\right), \delta^{\mathcal{D}}_{A_{i'}}\right)\right] + c = \mathbb{E}\left[\rho\left(\xi\left(A_i \cup B\right), \delta^{\mathcal{D}}_{A_i \cup B}\right)\right] + c,$$

But of course conditioning on $A_i \cup B$ is the same as conditioning first on A_i, then on B: $\xi(A_i \cup B) = \xi(A_i, B)$. Therefore

$$\mathbb{E}\left[\rho\left(\xi\left(A_i \cup B\right), \delta^{\mathcal{D}}_{A_i \cup B}\right)\right] + c = \mathbb{E}\left[\rho\left(\xi\left(A_i, B\right), \delta^{\mathcal{D}}_{A_i \cup B}\right)\right] + c.$$

Since $\delta^{\mathcal{D}}_{A_i \cup B}$ is optimal relative to the subjective probability $\xi(A_i, B)$ and $\delta^{\mathcal{D}}_{A_i}$ is optimal relative to the subjective probability $\xi(A_i)$, it follows from theorem 20 that for a given realization of A_i

$$\rho\left(\xi\left(A_i\right), \delta^{\mathcal{D}}_{A_i}\right) + c \geq \mathbb{E}\left[\rho\left(\xi\left(A_i, B\right), \delta^{\mathcal{D}}_{A_i \cup B}\right) \mid A_i\right] + c.$$

Here the expectation on the right-hand side is only with respect to B. The inequality is strict whenever there is a set of possible outcomes for B with positive probability for which $\delta^{\mathcal{D}}_{A_i}$ is not an optimal procedure, given A_i. Taking the expectation with respect to A_i yields:

$$\begin{aligned}
\mathbb{E}\left[\rho\left(\xi\left(A_i\right), \delta^{\mathcal{D}}_{A_i}\right)\right] + c &\geq \mathbb{E}\left[\mathbb{E}\left[\rho\left(\xi\left(A_i, B\right), \delta^{\mathcal{D}}_{A_i \cup B}\right) \mid A_i\right]\right] + c \\
&= \mathbb{E}\left[\rho\left(\xi\left(A_i, B\right), \delta^{\mathcal{D}}_{A_i \cup B}\right)\right] + c \\
&= \mathbb{E}\left[\rho\left(\xi\left(A_{i'}\right), \delta^{\mathcal{D}}_{A_{i'}}\right)\right] + c,
\end{aligned}$$

which is the desired inequality. The inequality is strict if and only if there is a set of outcomes of A_i with positive probability for which there is a set of outcomes of B with positive probability such that $\delta^{\mathcal{D}}_{A_i}$ is not an optimal procedure relative to $\xi(A_i, B)$, i.e.,

$$\rho\left(\xi\left(A_i, B\right), \delta^{\mathcal{D}}_{A_i}\right) > \rho\left(\xi\left(A_i, B\right), \delta^{\mathcal{D}}_{A_i \cup B}\right),$$

which by lemma 21 is equivalent to there being a set of outcomes of $A_{i'}$ with positive probability such that if A_i matches $A_{i'}$ in their common part,

$$\rho\left(\xi\left(A_{i'}\right), \delta^{\mathcal{D}}_{A_i}\right) > \rho\left(\xi\left(A_{i'}\right), \delta^{\mathcal{D}}_{A_{i'}}\right). \qquad \square$$

Again it may be objected that the theorem assumes that the decision problem the scientist faces after the first connection is the same whether she connects to i or i', while in fact the problems differ: after connecting to i, i' is still available to connect to and i is not, whereas after connecting to i' the situation is reversed.

But I claim that this is not a substantive assumption of the theorem. Consider the following. If the scientist ends up connecting to both i and i', she ends up conditioning on $A_i \cup A_{i'}$ and she is in the same situation whether she connected to i or i' first. If she connects to only one of them, the theorem shows that connecting to i' is at least as good as connecting to i, and the omitted options are irrelevant because they are unused.

So the theorem shows that in any decision problem, if $[A_i] \subseteq [A_{i'}]$, connecting to i' is always at least as good as connecting to i, and strictly better whenever $[A_i] \subset [A_{i'}]$ and the extra information in $A_{i'}$ has any chance of making a difference to future decisions. I take this to be a kind of justification, within the Bayesian framework, of assumption 12, which says that scientists will choose to connect to i' rather than i whenever $[A_i] \subset [A_{i'}]$.

3.3 Uniformly Bounded Connection Probabilities

In this section I compare the optimal procedures for different scientists in different contexts (where the context is the set of scientists that are available to connect to). I assume that all scientists are trying to solve the same problem. In particular, each scientist has to choose a decision from the same set D, they consider the same set of possible worlds Ω, they have the same loss function ℓ, before observing the information in their own information set they have the same prior $\xi \in \Xi$, and they have the same cost of connecting

$c > 0$.

Fix a relative frequency of information q. Then for any scientific community $\mathcal{C} = (I, A_I)$ satisfying q, the scientists that are available for any scientist $i \in I$ to connect to are given by the set $I \setminus \{i\}$. As before, $\Delta_\mathcal{D}$ denotes the set of all sequential decision procedures available to i given decision problem $\mathcal{D} = (\mathcal{C}, D)$. So $\Delta_\mathcal{D}$ contains all procedures that specify which scientists in $I \setminus \{i\}$ to connect to (possibly as a function of the information gained from previous connections) and which decision from D to choose after it stops connecting.

As before, a Bayesian scientist should evaluate the risk of a sequential decision procedure using the formula in equation (5). In particular, scientist $i \in I$ has initial information set A_i, and so her subjective beliefs are represented by $\xi(A_i)$. Thus, she should evaluate the risk of any procedure $\delta \in \Delta_\mathcal{D}$ as $\rho(\xi(A_i), \delta)$.

By theorem 19, there exists an optimal procedure for scientist i. Following the notation I used above, the optimal procedure for scientist i (i.e., the optimal procedure relative to prior $\xi(A_i)$) is denoted by $\delta^\mathcal{D}_{A_i}$.

Now suppose that for every abstracted information set $[A]$ that occurs in the community (i.e., every $[A]$ such that $q([A]) > 0$) there were an infinite number of scientists with abstracted information set $[A]$. In particular, define $C_\infty = (I_\infty, A_{I_\infty})$ such that for all $[A]$:

$$|\{i \in I_\infty \mid [A_i] = [A]\}| = \begin{cases} \infty & \text{if } q([A]) > 0, \\ 0 & \text{if } q([A]) = 0. \end{cases}$$

If i could connect to any scientist in the infinite set I_∞ (other than herself), the decision problem she faces is $\mathcal{D}_\infty = (\mathcal{C}_\infty, D)$. The set of all sequential decision procedures is $\Delta_{\mathcal{D}_\infty}$, and the optimal procedure from that set is $\delta^{\mathcal{D}_\infty}_{A_i}$.

As before, I will use $\delta^\mathcal{D}_{A_i}$ to stand both for the sequential decision procedure and the number of connections made by that procedure. For any abstracted information set $[A]$, $\delta^\mathcal{D}_{A_i,[A]}$ denotes the number of connections made to scientists with abstracted information set $[A]$ by procedure $\delta^{\mathcal{D}}_{A_i}$. In this section I am interested in the probability that $\delta^{\mathcal{D}}_{A_i,[A]} \geq n$ for large n.

Assuming that the optimal risk is finite, the following result follows immediately.

Lemma 23. *Let $\mathcal{P} = (\Omega, \mathcal{W}, \xi)$ be a probability space, D a set, and ρ a risk function (specified relative to $\mathcal{P}$ for a given loss function ℓ and cost of connecting $c > 0$). Let q be a relative frequency, $[A]$ an abstracted information set, $\mathcal{C} = (I, A_I)$ a scientific community such that either $\mathcal{C}$ satisfies q or $\mathcal{C} = \mathcal{C}_\infty$, and $\mathcal{D} = (\mathcal{C}, D)$ the associated decision problem. For all $i \in I$, if $\Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right) > 0$ then*

$$\lim_{n\to\infty} n \Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq n \mid \delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right) = 0.$$

In other words, for all $\varepsilon > 0$ there is an $N^{\mathcal{D}}_{i,[A],\varepsilon}$ such that for all $n > N^{\mathcal{D}}_{i,[A],\varepsilon}$

$$n \Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq n \mid \delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right) \leq \varepsilon.$$

Proof. It follows immediately from equation (5) that

$$\mathbb{E}\left[\delta^{\mathcal{D}}_{A_i}\right] \leq \frac{1}{c} \cdot \rho\left(\xi(A_i), \delta^{\mathcal{D}}_{A_i}\right) < \infty.$$

So the expected number of connections is finite. Since the expected number of connections to scientists with abstracted information set $[A]$ cannot be higher than the expected total number of connections, it follows that it must be finite as well: $\mathbb{E}\left[\delta^{\mathcal{D}}_{A_i,[A]}\right] < \infty$. Then it is easy to prove that

$$\lim_{n\to\infty} n \Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq n\right) = 0.$$

So for all $\varepsilon > 0$ there is an $M(\varepsilon)$ such that for all $n > M(\varepsilon)$

$$n \Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq n\right) \leq \varepsilon.$$

From the assumption $\Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right) > 0$ it follows that for all $\varepsilon > 0$

$$\varepsilon \cdot \Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right) > 0.$$

The rest of the proof is straightforward. Let $\varepsilon > 0$. Choose $N^{\mathcal{D}}_{i,[A],\varepsilon} := M\left(\varepsilon \cdot \Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right)\right)$. Let $n > N^{\mathcal{D}}_{i,[A],\varepsilon}$. It follows that

$$n \Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq n\right) \leq \varepsilon \Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right).$$

And thus

$$n \Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq n \mid \delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right) = \frac{n \Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq n\right)}{\Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right)} \leq \varepsilon. \qquad \square$$

The above result is completely trivial for finite sets of scientists I, as then there always exist n such that $\Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq n\right) = 0$ (namely any n such that $n > q([A]) \cdot |I|$). However, this lemma proves that the probability of a large number of connections goes to zero relatively fast even when there are infinitely many opportunities to connect.

This bound on the probability of a large number of connections may be wildly different depending on the scientist i and on the community $\mathcal{C}$ that i is a part of. The probability is bounded by ε only if n is larger than $N^{\mathcal{D}}_{i,[A],\varepsilon}$. What I would like is a uniform bound such that the probability is bounded by ε for all i and $\mathcal{C}$ if n is larger than $N_{[A],\varepsilon}$. Since there are infinitely many possible scientists and sets of scientists it is not clear that I can simply take the maximum over all i and $\mathcal{C}$ of all the $N^{\mathcal{D}}_{i,[A],\varepsilon}$ and get a finite number. The rest of this section seeks to establish that this can be done.

Lemma 24. *Let $\mathcal{P} = (\Omega, \mathcal{W}, \xi)$ be a probability space, D a set, and ρ a risk function (specified relative to $\mathcal{P}$ for a given loss function ℓ and cost of connecting $c > 0$). Let q be a relative frequency, i a scientist, $[A]$ an abstracted information set, and $\varepsilon > 0$. There exists an $N_{i,[A],\varepsilon}$ satisfying the following. Let $n > N_{i,[A],\varepsilon}$, let $\mathcal{C} = (I, A_I)$ be a scientific community satisfying q, and let $\mathcal{D} = (\mathcal{C}, D)$ be the associated decision problem. If $i \in I$*

and $\Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right) > 0$ *then*

$$n \Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq n \mid \delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right) \leq \varepsilon.$$

Proof. It follows from lemma 23 that for all decision problems $\mathcal{D} = (\mathcal{C}, D)$ with $\mathcal{C}$ satisfying q there exists an $N^{\mathcal{D}}_{i,[A],\varepsilon}$ such that the desired inequality holds. This lemma shows that there is a finite $N_{i,[A],\varepsilon}$ such that $N^{\mathcal{D}}_{i,[A],\varepsilon} \leq N_{i,[A],\varepsilon}$ for all such $\mathcal{D}$.

Let $[A]$ be an abstracted information set and let i be a scientist. Consider the community $\mathcal{C}_{[A]} = (I_{[A]} \cup \{i\}, A_{I_{[A]} \cup \{i\}})$, where $I_{[A]}$ satisfies

$$\left|\left\{i' \in I_{[A]} \mid [A_{i'}] = [B]\right\}\right| = \begin{cases} \infty & \text{if } [B] = [A], \\ 0 & \text{if } [B] \neq [A]. \end{cases}$$

So $\mathcal{C}_{[A]}$ is a community consisting, in addition to i, of infinitely many scientists with abstracted information set $[A]$.

Consider the decision problem $\mathcal{D}_{[A]} = (\mathcal{C}_{[A]}, D)$ faced by scientist i. By theorem 19, there exists an optimal procedure $\delta^{\mathcal{D}_{[A]}}_{A_i} \in \Delta_{\mathcal{D}_{[A]}}$. By lemma 23, for all $\varepsilon > 0$ there is an $N^{\mathcal{D}_{[A]}}_{i,[A],\varepsilon}$ such that for all $n > N^{\mathcal{D}_{[A]}}_{i,[A],\varepsilon}$,

$$n \Pr\left(\delta^{\mathcal{D}_{[A]}}_{A_i,[A]} \geq n \mid \delta^{\mathcal{D}_{[A]}}_{A_i,[A]} \geq 1\right) \leq \varepsilon.$$

I claim that $N^{\mathcal{D}_{[A]}}_{i,[A],\varepsilon}$, as just defined, serves as the desired finite bound.

Let $n > N^{\mathcal{D}_{[A]}}_{i,[A],\varepsilon}$. Let $\mathcal{C} = (I, A_I)$ satisfy q and let $\mathcal{D} = (\mathcal{C}, D)$ be the associated decision problem. Assume that $i \in I$ and $\Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right) > 0$. Compare $\delta^{\mathcal{D}}_{A_i,[A]}$ and $\delta^{\mathcal{D}_{[A]}}_{A_i,[A]}$. These are each the optimal solution to the same sequential decision problem, except that the sets of scientists available to connect to are different. There are two differences between the sets I and $I_{[A]}$: (1) I contains a finite number of scientists with abstracted information set $[A]$, while $I_{[A]}$ contains an infinite number, and (2) I (potentially) contains some scientists with abstracted information sets other than $[A]$, while $I_{[A]}$ does not. It follows that

$$n\,\Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq n \mid \delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right) \leq n\,\Pr\left(\delta^{\mathcal{D}[A]}_{A_i,[A]} \geq n \mid \delta^{\mathcal{D}[A]}_{A_i,[A]} \geq 1\right) \leq \varepsilon.$$

The first inequality holds for all $n \in \mathbb{N}$, so certainly for all $n > N^{\mathcal{D}[A]}_{i,[A],\varepsilon}$. This inequality holds because whenever it is optimal to connect to more scientists with abstracted information set $[A]$ in the set of scientists I it must also be optimal to do so in the set of scientists $I_{[A]}$: the presence of extra scientists with abstracted information set $[A]$ cannot make it worse to connect to more such scientists, and the absence of scientists with other abstracted information sets similarly cannot make it worse to connect to more scientists with abstracted information set $[A]$. □

With this lemma in hand I can prove the theorem of this section.

Theorem 25. *Let $\mathcal{P} = (\Omega, \mathcal{W}, \xi)$ be a probability space, D a set, and ρ a risk function (specified relative to $\mathcal{P}$ for a given loss function ℓ and cost of connecting $c > 0$). For any relative frequency q, abstracted information set $[A]$, and $\varepsilon > 0$, there exists $N_{[A],\varepsilon}$ such that for all $n > N_{[A],\varepsilon}$, if $\mathcal{C} = (I, A_I)$ is a scientific community satisfying q with $\mathcal{C}$ simple relative to $\mathcal{P}$, $\mathcal{D} = (\mathcal{C}, D)$ the associated decision problem, $i \in I$ a scientist, and $\Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right) > 0$, then*

$$n\,\Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq n \mid \delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right) \leq \varepsilon.$$

Proof. It follows from lemma 24 that for all i there exists an $N_{i,[A],\varepsilon}$ with the desired properties. It remains to show that there exists an $N_{[A],\varepsilon}$ such that $N_{i,[A],\varepsilon} \leq N_{[A],\varepsilon}$ for any i that could potentially be part of a scientific community satisfying q (i.e., any i such that $q([A_i]) > 0$).

Note first that for any i and i' such that $[A_i] = [A_{i'}]$ the optimal procedures are closely related. If the realizations of the random variables in A_i and $A_{i'}$ are the same ($X_{j,i,k} = X_{j,i',k}$ for all relevant j and k), then $\xi(A_i) = \xi(A_{i'})$

(compare the argument in the proof of lemma 21). In that case the two scientists are facing exactly the same problem, so their optimal procedures $\delta^{\mathcal{D}}_{A_i}$ and $\delta^{\mathcal{D}}_{A_{i'}}$ are simply identical. But since the random variables in A_i and $A_{i'}$ are identically distributed, the probability of any given realization is the same. Thus by the law of total probability the overall probability of any number of connections must be the same (since that probability equals the conditional probability given some realization of A_i or $A_{i'}$ integrated over the distribution of A_i or $A_{i'}$), i.e., for all $[A]$ and for all $n \in \mathbb{N}$

$$n \Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq n \mid \delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right) = n \Pr\left(\delta^{\mathcal{D}}_{A_{i'},[A]} \geq n \mid \delta^{\mathcal{D}}_{A_{i'},[A]} \geq 1\right).$$

So if i's probabilities are bounded, i''s must be bounded the exact same way: $N_{i,[A],\varepsilon}$ or $N_{i',[A],\varepsilon}$ can serve as a bound for either i or i''s connection probabilities. So $N_{i,[A],\varepsilon}$ bounds the connection probabilities for all i' such that $[A_i] = [A_{i'}]$.

Recall that $q([A]) > 0$ for only finitely many abstracted information sets $[A]$. Let $\mathcal{C}' = (I', A_{I'})$ be a community that satisfies

$$|\{i \in I' \mid [A_{i'}] = [A]\}| = \begin{cases} 1 & \text{if } q([A]) > 0, \\ 0 & \text{if } q([A]) = 0. \end{cases}$$

So $\mathcal{C}'$ contains exactly one scientist with abstracted information set $[A]$ for each $[A]$ such that $q([A]) > 0$. Thus

$$N_{[A],\varepsilon} := \max_{i \in I'} N_{i,[A],\varepsilon}$$

is well-defined (because I' is a finite set). I claim that $N_{[A],\varepsilon}$ serves as the desired bound.

Let $n > N_{[A],\varepsilon}$, and let $\mathcal{C} = (I, A_I)$ be a scientific community satisfying q with $\mathcal{C}$ simple relative to $\mathcal{P}$, $\mathcal{D} = (\mathcal{C}, D)$ the associated decision problem, $i \in I$ a scientist, and $\Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right) > 0$. Since i is a member of $\mathcal{C}$, which satisfies q, $q([A_i]) > 0$. So there exists $i' \in I'$ such that $[A_i] = [A_{i'}]$. Since $n > N_{i',[A],\varepsilon}$, by the foregoing argument

$$n \Pr\left(\delta^{\mathcal{D}}_{A_i,[A]} \geq n \mid \delta^{\mathcal{D}}_{A_i,[A]} \geq 1\right) = n \Pr\left(\delta^{\mathcal{D}}_{A_{i'},[A]} \geq n \mid \delta^{\mathcal{D}}_{A_{i'},[A]} \geq 1\right) \leq \varepsilon. \quad \square$$

It is easy to further generalize this theorem by following a similar strategy to the last step in the above proof. One generalization uses the fact that there are only finitely many abstracted information sets $[A]$ such that $q([A]) > 0$ to establish the existence of an N_ε such that $N_{[A],\varepsilon} \leq N_\varepsilon$ for all $[A]$. Another generalization notes that if there are finitely many different problems the scientists are solving (finitely many different $\mathcal{P}$, D, and ρ, instead of all these things being the same for all scientists) the above theorem still holds.

The above theorem, however, is sufficient for my goal in this section. That goal is to show that assumption 10 is satisfied by a group of scientists if they are all solving the same problem and behave like Bayesian statisticians. Theorem 25 shows exactly that. So assumption 10 is justified, within the Bayesian framework, at least whenever all scientists are solving the same problem.

Combining theorems 22 and 25 with theorem 13 yields the following result. If (1) scientists behave like Bayesian statisticians that are all solving the same problem, (2) every piece of information is valuable (in the sense of having a non-zero chance of changing future optimal decisions), and (3) the population is large enough, then scientists' expected in-degrees are a supermodular function of their information set. So one should expect large differences in in-degree, and one should expect these differences to track the amount of information a scientist has (such that more information means a higher in-degree).

References

Katharine A. Anderson. A model of collaboration network formation with heterogeneous skills. *Network Science*, 4(2):188–215, Jun 2016. ISSN 2050-1250. doi: 10.1017/nws.2015.33. URL `http://journals.cambridge.org/article_S2050124215000338`.

Venkatesh Bala and Sanjeev Goyal. A noncooperative model of network formation. *Econometrica*, 68(5):1181–1229, 2000. ISSN 1468-0262. doi: 10.1111/1468-0262.00155. URL `http://dx.doi.org/10.1111/1468-0262.00155`.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999. ISSN 00368075. URL `http://www.jstor.org/stable/2899318`.

Y. S. Chow and Herbert Robbins. On optimal stopping rules. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 2(1):33–49, 1963. ISSN 0044-3719. doi: 10.1007/BF00535296. URL `http://dx.doi.org/10.1007/BF00535296`.

Jonathan R. Cole. Patterns of intellectual influence in scientific research. *Sociology of Education*, 43(4):377–403, 1970. ISSN 00380407. URL `http://www.jstor.org/stable/2111839`.

Morris H. DeGroot. *Optimal Statistical Decisions*. John Wiley & Sons, New Jersey, 2004.

I. J. Good. On the principle of total evidence. *The British Journal for the Philosophy of Science*, 17(4):319–321, 1967. ISSN 00070882. URL `http://www.jstor.org/stable/686773`.

Matthew O. Jackson and Asher Wolinsky. A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74, 1996. ISSN

0022-0531. doi: http://dx.doi.org/10.1006/jeth.1996.0108. URL http://www.sciencedirect.com/science/article/pii/S0022053196901088.

Derek J. de Solla Price. Networks of scientific papers. *Science*, 149(3683): 510–515, 1965. ISSN 00368075. URL http://www.jstor.org/stable/1716232.

S. Redner. How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B - Condensed Matter and Complex Systems*, 4(2):131–134, 1998. ISSN 1434-6028. doi: 10.1007/s100510050359. URL http://dx.doi.org/10.1007/s100510050359.