7-2012

Skill Specialization and the Formation of Collaboration Networks

Katharine A. Anderson
Carnegie Mellon University, andersok@andrew.cmu.edu

SKILL SPECIALIZATION AND THE FORMATION OF COLLABORATION NETWORKS

KATHARINE A. ANDERSON

Abstract. In recent years, there has been increased interest among funding organizations and administrators in supporting the acquisition of interdisciplinary skills in collaborative communities, such as universities, national labs, and knowledge-based firms. However, there has been relatively little work exploring the effects of interdisciplinarity on the structure and function of these communities. In this paper, I use a collaboration network–in which the nodes are individuals, and two nodes are connected if they've collaborated on a problem–to formalize the structure of collaborative relationships. Using a formal model of the collaborative process, I examine the effects of increased interdisciplinarity on the structure of this collaboration network. I show that when collaborative communities become more interdisciplinary, the links in the network become more concentrated among a few, high-degree individuals, and superstars emerge. These individuals, who are so productive that their contributions dominate the overall community, are a potential unintended consequence of policies intended to increase interdisciplinarity. I then define a specialist to be an individual whose skills cluster in a single area, and a generalist to be an individual whose skills span several areas, and I examine the roles that specialists and generalist play in the network. I show that while specialists have more links in the collaboration network, generalists are more likely to bridge between different communities. Given that individuals in these communities tend to benefit from being highly connected, while the community as a whole benefits from bridging activities, this result suggests that generalists may be undersupplied, which lends support to policies that fund the acquisition of interdisciplinary skills in situations where bridging activities are valued by the community at large but are not individually rational.

Collaborative problem-solving is vital to the generation of new knowledge in universities, national labs, and within industry. The organization of these problem-solving communities is important because it governs the flow of information and ideas, and can ultimately help determine the pace of innovation. A collaboration network captures the organization of a collaborative community, highlighting both overall structure and the role of the individual.

Nodes in these networks are individual problem solvers, and two individuals are connected if they have worked together on a problem.

Collaboration is particularly important in knowledge-generation, because it allows individuals with different sets of skills to work together and solve problems that no single individual in the community could solve alone. This suggests that individuals with different skill sets should have different roles in the collaborative community. In particular, suppose we have two individuals: a specialist, whose skills are tightly clustered in a single area, and a generalist, whose skills span several different areas. What role will those two individuals play in the collaborative community? Will they occupy different positions in the collaboration network? And how does the mixture of these two types in the problem-solving population affect the overall structure of the collaboration network? Will the collaboration network of a population with more generalists differ substantially from the collaboration network of a population with more specialists?

In this paper, I use a model of collaboration network formation to make predictions about the role of skill specialization in individual network position and the structure of collaboration networks. I show that the collaborative process exaggerates small differences in skill level–individuals with a small number of additional skills have dramatically more links, leading to an empirically-familiar "hub and spoke" network structure. I show that when collaborative communities become more interdisciplinary, the links in the network become more concentrated among a few, high-degree individuals, and superstars emerge. These individuals are so productive that their contributions dominate the overall community. On an individual level, a person's position in the collaboration network differs depending on whether they are a specialist or generalist. Specialists are much more likely to be the high-degree hubs in the network, while generalists are much more likely to be bridges between different communities.

This work has some interesting implications for the funding of interdisciplinary researchers. In the past few decades, there has been considerable interest, both within funding organizations and among the ranks of university, hospital, and national lab administrators, in increasing the number of interdisciplinary researchers in academic institutions. Implicitly, it is assumed that 1) interdisciplinary researchers have positive affects on the productivity of the field as a whole, and 2) the number of interdisciplinary researchers is lower than is optimal. However, there is relatively little research to back up either of these assumptions. This work provides some support for both of these claims. There is some research which indicates the importance of bridging activity in the flow of information through organizations.[1] This research suggests that generalists are more likely to fill those bridging roles, making them valuable to the field as a whole. Moreover, considering that high degree is correlated with high productivity in collaborative communities, while bridging activity is not, this work suggests that there is not much individual incentive to be an interdisciplinary researcher, and thus absent the intervention of a central authority, there may be an under-supply of interdisciplinary researchers in the population.

1. Models of Collaboration Network Formation

Collaboration networks provide a useful way of looking at the structure and function of collaborative communities. One striking feature of these networks is their great heterogeneity. Some individuals have very few links, while others have a great many. Some individuals are firmly ensconced within their communities, while others create bridges between communities. Moreover, while there are similarities in the overall structure of collaboration networks, the details of network structure clearly vary from one context to the next. As social scientists, we would like to believe that these variations in network position and network structure are due to some underlying heterogeneity between individuals and heterogeneity between collaborative communities. Indeed, the models used by empirical networks researchers implicitly assume such a connection exists. However, most theoretical models of network formation do not make the connection between individual heterogeneity and network heterogeneity. The model I use in this paper is explicitly designed to draw this connection, and thus before moving forward, it is worth taking a step back to look at the variety of models that have been used to generate social network structures.

These models can roughly be divided into two categories: statistical models and decision-based models. In statistical models, individual nodes are connected via some kind of stochastic process. The most famous of these models is "preferential attachment", in which new nodes connect to existing nodes with a probability that is proportional to the number of connections a node already has. These models are very good at replicating the large-scale

[1]Burt (2001)

structures common to most social networks, including the skewed degree distribution mentioned above.[2] The disadvantage of these models is that because individuals are not making decisions about their connections, they do not respond to incentives. Moreover, the primary factors that distinguish low-degree nodes from high-degree nodes are age and "luck".[3] Although this undoubtedly captures some of the variation in node degree in collaboration networks, there is surely some additional variation due to individual skill heterogeneity.

A second class of models, called decision-based models, allow individuals to make their linking decisions based on optimizing some kind of objective function.[4] The advantage of these models is that because individuals make decisions, they can be made to respond to incentives. However, most decision-based models of network formation assume that individuals are heterogeneous.[5] This means that the networks formed from these models are symmetric, and bear little resemblance to empirically-observable collaboration networks. In particular, there are no high-degree nodes. Thus, these models cannot be used to answer questions about who will end up in what position in the network.

The model in this paper is a member of a class of decision-based models in which heterogeneity in individual skill sets is translated into heterogeneity in network structure. The following section elaborates both on this general class of models, and on the specific model that I will be using in this paper.

2. Model

2.1. **The General Model.** The model of network formation that I will use here is based on the general model of collaboration network formation with heterogeneous skills, presented in Anderson (2011). In that model, there is a population of heterogeneous problem solvers, each of whom faces a complex problem. Each individual is endowed with a set of skills useful for solving the problems faced by the population as a whole. These skills are pieces of knowledge, abilities, and tools that useful for solving problems, and are not easily passed

[2]Barabási and Albert (1999) is the seminal paper. A number of follow up papers, including Jackson and Rogers (2007) make changes to the original model, which create a better match between the degree distribution obtained from the model and empirical reality.
[3]That is to say, a stochastic element.
[4]See, for example, Jackson and Wolinsky (1996) and Goyal and Moraga-González (2001)
[5]An exception is a literature springing from Galeotti et al. (2006), which includes Goeree et al. (2009). This literature is based on Jackson and Wolinsky (1996) and allows individuals to be heterogeneous in the cost of making links. However, this heterogeneity is obviously considerably different than heterogeneity over skills.

from one individual to the next.[6] Although it is possible that an individual will have all of the skills required to solve her problem alone, complex problems are unlikely to yield to the abilities of a single individual. When an individual does not have all of the skills required to solve her problem by herself, she can collaborate with others. Together, all of these collaborations form a collaboration network.

More formally, let $I = \{1, 2, ...N\}$ denote the set of problem solvers, and$S = \{a_1...a_M\}$ denote the (finite) set of all skills. Individuals are endowed with two things: a problem, $\omega_i \subseteq S$, and a skill set, $A_i \subseteq S$. The problem that a person has to solve is defined by the set of skills that must be applied in solving it. Her skill set is the human capital (skills) that she can apply towards solving it. The distribution of problems across individuals is Ω and the distribution of skill sets is Ψ.

A collaboration is any subset of the problem solvers, $C \subseteq I$. A problem can be solved by a given collaboration if that set of problem solvers has all of the required skills–that is, a collaboration, C, can solve a problem, ω, if $\omega \subseteq \bigcup_{k \in C} A_k$. An individual chooses a set of collaborators, C_i, where $i \in C_i$ by default.

For a given set of collaborations, $C = \{C_1...C_N\}$, a collaboration network, $g(C)$, is given by the set of all individuals linked via those collaborations. That is, $ij \in g$ if $j \in C_i$. [7]

In equilibrium, an individual chooses her set of collaborators to maximize her payoffs. A problem yields a payoff of 1 if solved, and 0 if not solved: $\pi_i(C_i) = \begin{cases} 1 \text{ if } \omega_i \subseteq \bigcup_{k \in C_i} A_k & 0 \text{ otherwise} \end{cases}$. The individual splits the payoff from the problem with her collaborators–each gets $\frac{\pi_i(C_i)}{|C_i|}$. An individual will potentially be asked to join other collaborations and solve other individuals' problems, and therefore her total payoff is given by $u_i(C) = \frac{\pi_i(C_i)}{|C_i|} + \sum_{j \neq i \, st \, i \in C_j} \frac{\pi_i(C_j)}{|C_j|}$. Given this payoff structure, it is incentive compatible for the individual to choose the smallest set of collaborators that will let her solve the problem, and the individual will be indifferent between any two such minimum sets. [8]

[6]Note that this latter characterization distinguishes skills from information. Whereas information is easily passed from one person to another and aggregated across multiple individuals, skills are not.

[7]This network is technically directed. However, Anderson (2011) notes that links are pairwise stable, and thus it is reasonable to think of the network as being undirected. Here, I will do the same.

[8]Anderson (2011) notes that any payoff structure that incentivizes minimizing the set of collaborators will yield the same results. So, for example, if links are costly, the result is the same.

Let C^* denote a payoff-maximizing set of collaborations, $\{C_1^*...C_N^*\}$. The network that results from such a payoff-maximizing set of collaborations is called a "complementary skills network".

Thus, the inputs to the general model are a population-wide distribution of skills, Ψ, and a population-wide distribution of problems, Ω. The output is a set of payoff-maximizing sets of collaborations, denoted $\mathcal{C}(\Psi, \Omega)$, and a set of equilibrium networks, $\Gamma(\Psi, \Omega)$, where $g(C^*) \in \Gamma(\Psi, \Omega)$ if $C^* \in \mathcal{C}(\Psi, \Omega)$. In other words, the inputs to the model are a set of population characteristics, and the output is a set of networks that the problem solvers might form, given those population characteristics. This direct mapping between individual heterogeneity (Ψ and Ω) and network structure ($\Gamma(\Psi, \Omega)$) is the strength of this model, because it allows one to link the characteristics of a population to the structure of the network, and an individual's skill set to her position in the network.

2.2. **A population with specialists and generalists.** Here, I am considering the link between skill specialization and network structure, so I will look at a particular specification for Ψ and Ω. The existence of "generalists"–people whose skills span more than one area–requires that skills be divided into at least two subcategories. Thus, I will assume that the set of M skills is divided into two "disciplines": A and B. Similarly, I will divide the population in two, with half of the population being part of discipline A and half being part of discipline B.

The distribution of skills in the population (Ψ) is defined as follows. Skills are distributed independently.[9] An individual has probability p_{own} of having each skill in her own discipline, and a probability p_{other} of having each skill in the other discipline. I assume that an individual is more likely to have skills in her own discipline, meaning that $p_{own} \geq p_{other}$. Thus, the probability of an individual having k skills in her own discipline and j skills in the other discipline is $p_{own}^k (1 - p_{own})^{\frac{M}{2}-k} p_{other}^j (1 - p_{other})^{\frac{M}{2}-j}$. In this framework, specialists and generalists emerge naturally. A specialist is any individual whose skills all fall within their own discipline, while a generalist is an individual with skills in both disciplines.

For the moment, I will assume that the distribution of problems in the population (Ω) is such that individuals face discipline-specific problems. That is, individuals in discipline A

[9]This is similar to the assumption of the Bernoulli Skills Model, which is outlined in Anderson (2011). In the Bernoulli Skills Model, all M skills are distributed independently with probability p.

all face the same problem, requiring all of the discipline A skills: $\omega^A = \left\{a_1 ... a_{\frac{M}{2}}\right\}$, while individuals in discipline B all face the same problem requiring all of the discipline B skills: $\omega^B = \left\{a_{\frac{M}{2}} ... a_M\right\}$. Note that this strict division of problems into "discipline A problems" and "discipline B problems" will be relaxed later.

In this model of skill specialization, the distribution of skills and problems in the population is defined by three parameters: M, p_{own}, and p_{other}. Together, p_{own} and p_{other} define the total amount of interdisciplinarity within the skill population. On one extreme, if $p_{other} = 0$, then nobody obtains skills in another discipline, and the two populations are entirely separate. As $p_{other} \to p_{own}$, there are more people with skills in multiple disciplines, and the two populations merge. Thus, these two parameters adjust the mixture of specialists and generalists in the population.

2.3. **Model Implementation.** The heterogeneity of collaboration networks makes them a great tool for examining the function of collaborative problem solving communities. However, this heterogeneity also makes proving analytical results extremely difficult. This is especially true for "global" network measures (those that depend on the structure of the network as a whole, rather than just the local neighborhood of a node). Fortunately, in this case there is little detail lost by simulating the model.[10] For each set of parameters, I run 1000 simulations, generating 1000 possible networks. All results that follow are obtained by averaging over those 1000 runs.[11] Although I present results for only a handful of representative parameter combinations, experiments indicate that the results hold across the relevant range of parameter values.

2.4. **An Example.** Figure 2.1 illustrates the outcome of a single run with $N = 100$, $M = 5$, $p_{own} = 0.25$, and $p_{other} = 0.15$.

3. Results

3.1. **Network Structure and Specialization.** First, consider the effect of specialization on the overall structure of the collaboration network. There has been considerable interest

[10]As a check on this, I have simulated the Bernoulli Skills Model, laid out in Anderson (2011). For reasonably large populations (N = 100), the results are similar to those obtained via analytical methods

[11]This differs only slightly from the procedure that would be required to obtain closed form solutions. In an analytical environment, one would still average over the ensemble of networks that results from a particular distribution of skills. In this case, I am also averaging over a number of draws from the same family of skill distributions, represented by the parameter values.

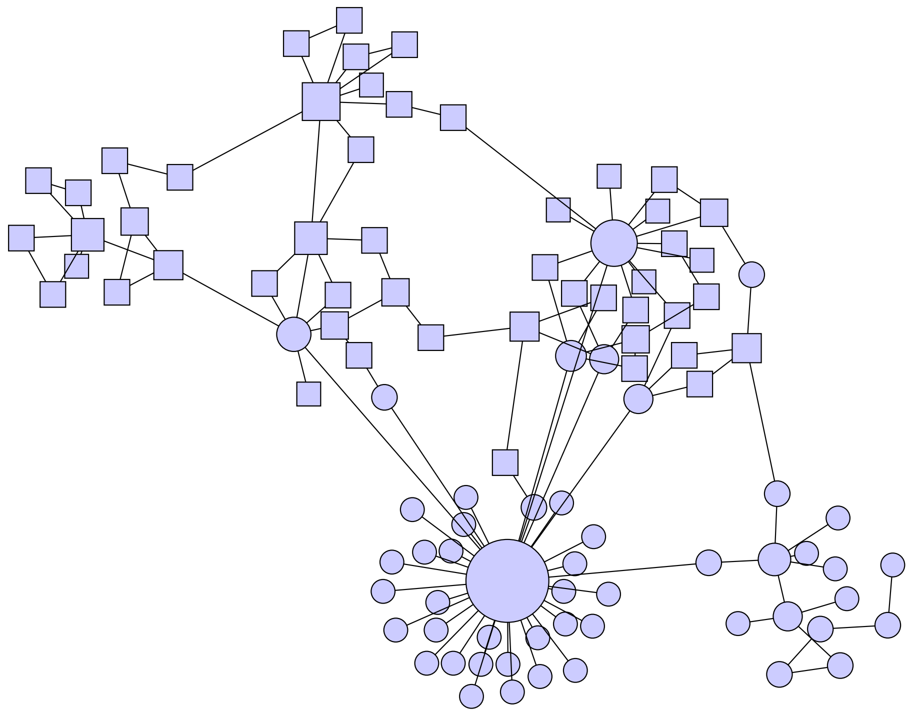

FIGURE 2.1. An example of a collaboration network with $N = 100$, $M = 5$, $p_{own} = 0.25$, and $p_{other} = 0.15$.

among funding organizations in supporting the acquisition of interdisciplinary skills in academic communities. There has also been considerable money and institutional power devoted to this goal among the administrators of research labs and academic institutions. However, there is little understanding of how encouraging the acquisition of interdisciplinary skills in individuals would impact the collaborative community as a whole. Thus, one question with important implications is how increasing the prevalence of people with interdisciplinary skills (generalists) affects the structure of the collaborative community. The amount of interdisciplinarity in the problem-solving population is controlled by two parameters, pown and pother. As $p_{other} \to p_{own}$, there is more and more overlap in the skill sets of the two disciplines, and thus more interdisciplinary workers. As $p_{other} \to 0$, there is less overlap, and there are fewer interdisciplinary workers in the population. (See Figure 3.1 for an illustration).

One measure of network structure that is particularly interesting is the degree distribution. Empirically, the degree distribution of collaboration networks (and, indeed, most social networks of all types) is skewed–the majority of the links are concentrated among a few,

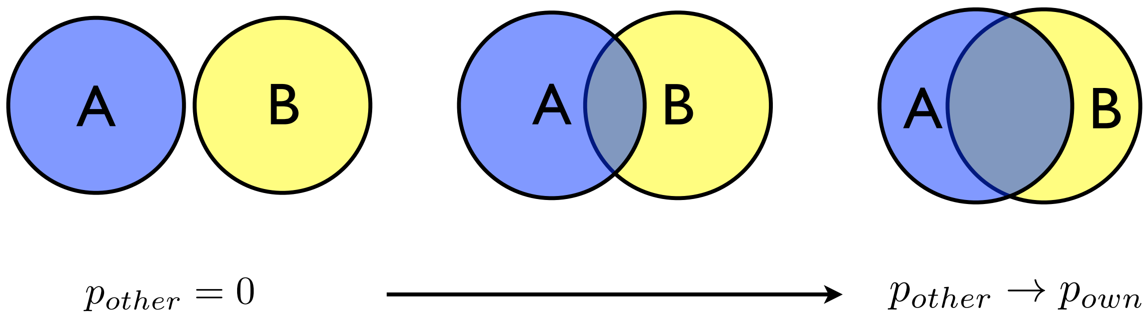

FIGURE 3.1. An illustration of the effects of adjusting p_{own} and p_{other}.

high-degree individuals, while most individuals have very few links.[12] The resulting network has a distinctive "hub-and-spoke" structure. However, the amount of skew in the degree distribution differs from one population to another. For example, the degree distribution of coauthorship networks in the medical community is much less skewed than the degree distribution of coauthorship networks in the physics community.[13] This has implications for the way that these two communities function–in the physics community, the average characteristics of the field is determined by a few, high-degree "superstars", whereas in the biomedical community, the average characteristics are determined by the masses. Thus, it would be valuable to know how the underlying characteristics of the community impact the resulting degree distribution.

Overall, the degree distribution of a network that results from a population with independent skills tends to be skewed towards a few high-degree nodes, much as empirical collaboration networks are. The source of this skew is explored in greater detail below.

Figure 3.2 illustrates the relationship between the shape of the degree distribution, and the amount of interdisciplinarity in the population (the inset shows just the tail of the degree distribution, better illustrating the behavior in this region). In all of the curves shown, the number of skills required to solve the problem (M) is fixed. Similarly, the total fraction of the skills that the average individual has ($p_{own} + p_{other}$) is fixed. What varies is degree to which the collaborative communities overlap ($p_{own} - p_{other}$). When there aren't many

[12]For example, we can see this in networks of interfirm collaboration (Powell et al. (1996), Iyer et al. (2006)), creative artists in broadway plays (Uzzi and Spiro (2005)), film actors (Barabási and Albert (1999)), jazz musicians (Gleiser and Danon (2003)), and coauthorship networks in a variety of fields: see Newman (2001) (physics), Moody (2004) (sociology), Goyal et al. (2006) (economics), and Acedo et al. (2006) (management science).
[13]Newman (2000)

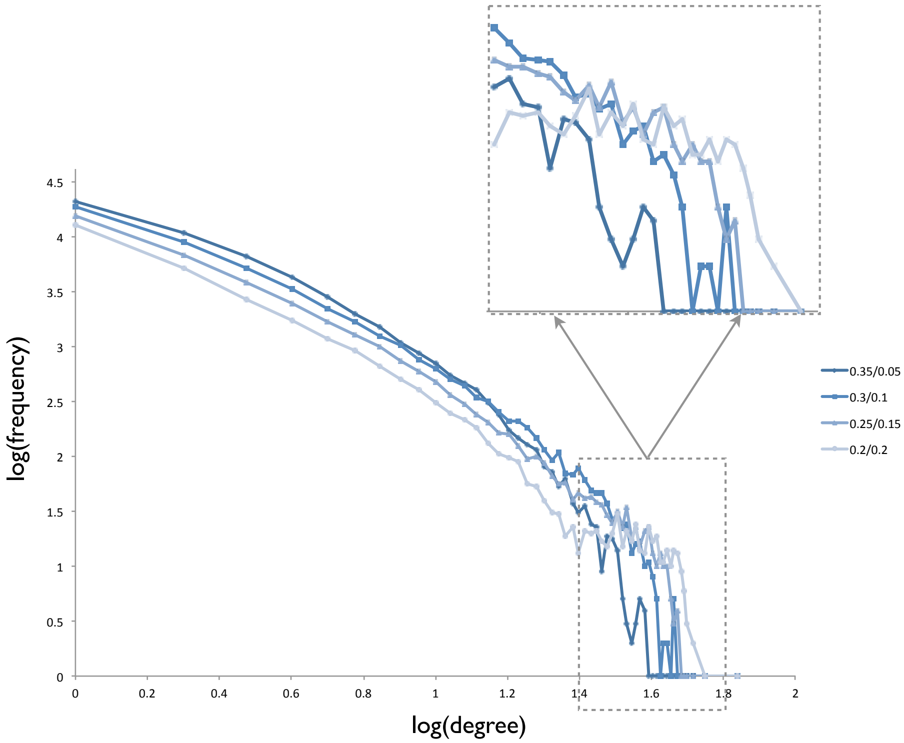

FIGURE 3.2. The degree distribution of networks with different levels of interdisciplinarity. As $p_{own} \to p_{other}$, the population contains a larger proportion of generalists, and the degree distribution becomes increasingly skewed towards a few, high-degree nodes.

generalists (the dark blue line in Figure 3.2), there are more individuals with low degree, and fewer individuals with high degree. In contrast, when generalists are more prevalent (the light blue line), the tail of the distribution extends, and there are more high-degree individuals.

The Gini coefficients for these distributions provide a way of quantifying the shape of the degree distribution. A higher Gini coefficient indicates a more uneven distribution of links across individuals. When $p_{own} = 0.35$ and $p_{other} = 0.05$, the Gini coefficient of the degree distribution is 0.72. The Gini coefficient increases across the four distributions, and when $p_{own} = p_{other} = 0.20$, the Gini coefficient is 0.87. This indicates that as the level of interdisciplinarity increases ($(p_{own} - p_{other}) \to 0$), the degree distribution of the network becomes more skewed towards a few, high-degree nodes. This suggests that if funding organizations put money into supporting the acquisition of interdisciplinary skills, they may

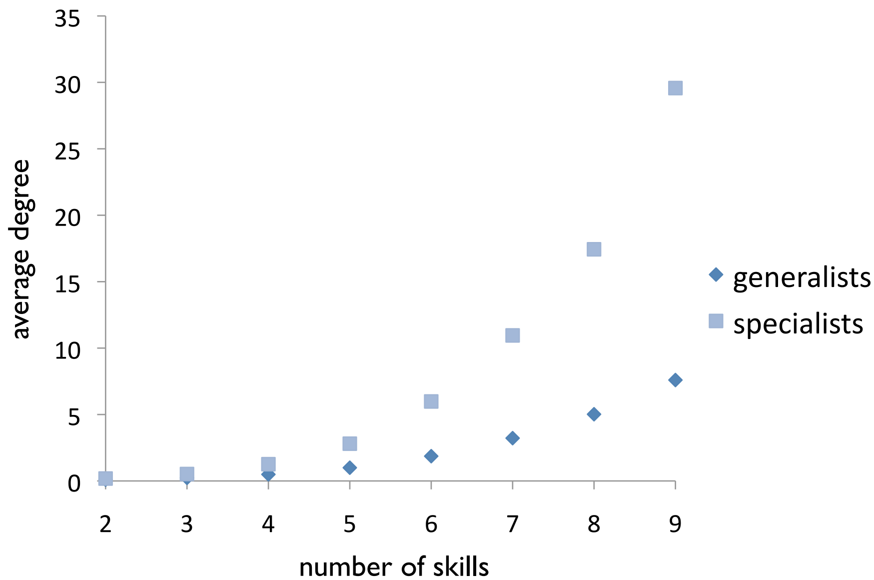

FIGURE 3.3. Specialists have higher degree than generalists with the same number of skills. Here, $M = 20$, $p_{own} = 0.35$, and $p_{other} = 0.05$. The average is taken over 1000 networks.

also inadvertently concentrate the productivity of the community in a smaller number of hands.

3.2. **The Network Position of Specialists and Generalists.** One of the most striking features of collaboration networks is their skewed degree distribution–most of the links are concentrated in a small number of hands. The natural question is: who are these high-degree individuals going to be? Are specialists or generalists more likely to have a large number of links? A comparison between specialists and generalists reveals that specialists have more links, on average, than generalists with the same number of skills. This gap is wider for highly-skilled individuals (see Figure 3.3).

In order to understand why specialists have more links, consider how an individual's set of skills is translated into her expected degree on a network. An individual's expected degree is a function of two things: 1) the number of people who need the various subsets of her skills and 2) the number of other people who have that subset. Thus, the expression for an individual's expected degree on the collaboration network is $E\left[d\left(A\right)\right] = \sum_{C\subseteq A} \frac{\Psi(C)}{\sum_{D\subseteq S\setminus C}\Psi(C\cup D)}$.

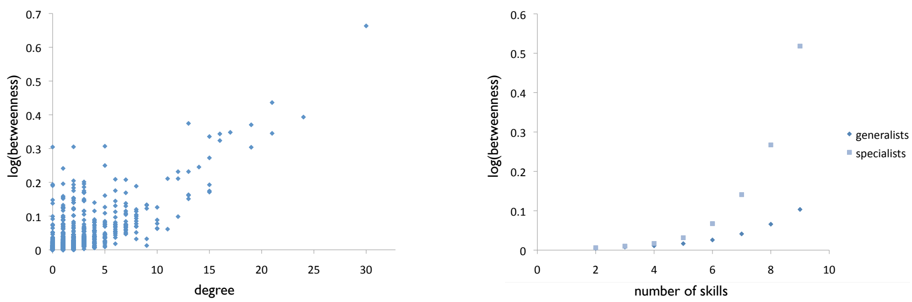

FIGURE 3.4. Due to the strong correlation between betweenness and degree (first panel), specialists also have higher betweenness than generalists with the same number of skills (second panel). Here, $M = 20$, $p_{own} = 0.35$, and $p_{other} = 0.05$. The average is taken over 1000 networks.

[14] Because of the summation over all possible subsets of the skill set, this quantity is super-additive in skills. However, only combinations of skills that are in demand by a particular population contribute to an individual's degree. Because problems are disciplinary, only some combinations of skills will be in demand. All subsets of a specialist's skills will be demanded by the people in her discipline, whereas only subsets of a generalist's skills will be in demand.

This suggests that if problems are also made interdisciplinary–that is, if some problems require a mixture of the skills in both disciplines–then the gap between specialists and generalists should disappear. To illustrate this, consider a modification to the model. Instead of every individual in a discipline getting the same problem, distribute problems in a way that allows some problems to be interdisciplinary. For every individual, a problem has $\frac{M}{2}$ skills. Skills $s = 1, 2, ...\frac{M}{2}$ are drawn from the set of skills from the person's own discipline with probability $q \leq \frac{1}{2}$ and from the other discipline with probability $1 - q$. When $q = 0$, problems are disciplinary, and when $q = \frac{1}{2}$, problems are entirely a-disciplinary. Varying q between 0 and $1/2$ reveals exactly what one would expect: as the community faces more interdisciplinary problems, the skills of generalists become more useful, and the gap in Figure 3.3 narrows. In the extreme of $q = \frac{1}{2}$, where all problems are interdisciplinary, the gap closes completely and generalists have equal degree to specialists, on average.

[14]Anderson (2011)

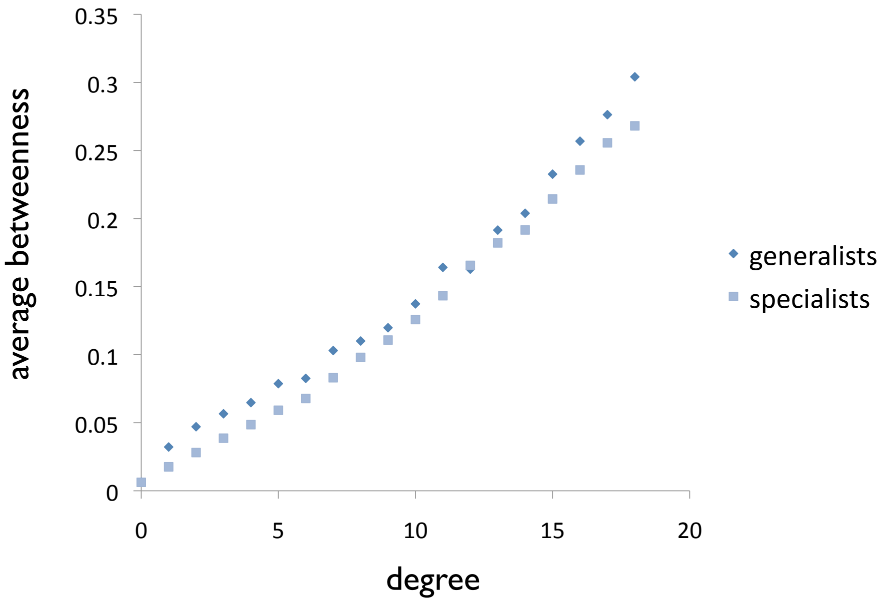

FIGURE 3.5. Generalists have higher betweenness than specialists of the same degree. Here, $M = 20$, $p_{own} = 0.35$, and $p_{other} = 0.05$. The average is taken over 1000 networks.

Clearly, degree is not the only network measure of importance to outcomes. Individuals who form "bridges" between communities have an important role as the brokers of information. So a natural question is whether specialists or generalists are more likely to bridge between communities? One natural measure of how well an individual connects communities is betweenness–a measure of the number of shortest paths that pass through a particular individual. However, as can be seen in Figure 3.4, betweenness is highly correlated with degree. Thus, if we simply look at the betweenness of individuals with the same number of skills, the fact that specialists have higher degree swamps the results. By looking at the betweenness of individuals with the same degree, we cancel out this effect. When we do this, generalists have significantly higher betweenness than specialists in the lower tail of the degree distribution.[15] This effect is illustrated in Figure 3.5.

Together, these two results suggest that while specialists will tend to have more collaborators (and thus, perhaps, be more productive), generalists will tend to play bridging roles between communities. This is interesting, because it suggests a possible disconnect between

[15]In the upper tail of the degree distribution, the data is much sparser, so it will always be difficult to draw conclusions in this region.

individual incentives and the needs of the community at large. There is a wide-spread belief that bridges between disciplines play a vital role in the health of the research community as a whole. Given that generalists are more likely to bridge communities, this means that the community as a whole benefits from having generalists around. However, if individual success is measured via productivity, then it is to the advantage of the individual to specialize. Together, these two facts mean that fewer people than is optimal may become generalists. Thus, to the extend that administers and funding organizations value bridges between communities, these results may support policies that financially support the acquisition of interdisciplinary skills.[16]

4. Conclusion

In this paper, I have used a theoretical model of the formation of collaboration networks to explore the connection between individual skill specialization and network structure. On a network-wide level, an increase in the number of generalists in a population will tend to skew the degree distribution of the network towards a few, high-degree superstars. On an individual level, specialists are more likely to occupy those high-degree positions in the network, while generalists are more likely to occupy bridging positions between different communities. While this model assumes a fixed, exogenous skill population, the results presented here have implications for the acquisition of skills. In particular, to the extent that we value the spillover effects of bridging between communities, this work suggests that the external support of interdisciplinary work might be needed.

[16]While this model assumes a fixed skill population, it is also interesting to consider what these results suggest about who would find it optimal to become a specialist or a generalist. For a moment, think of the number of skills that an individual has as their "ability" level. Individuals with high ability get a large bump in degree from specializing their skills. In contrast, the increase in their betweenness is relatively modest. On the other hand, low ability individuals get a relatively modest boost in degree from diversifying their skills, relative to the boost in betweenness from becoming generalists. This suggests that low ability individuals may wish to generalize, while high ability individuals may wish to specialize. The crossover point will depend on the relative weight placed by the individual on higher degree and higher betweenness. If we consider that the highest-ability individuals in a population might choose to become generalists for reasons that are outside the model (personality, boredom, natural spillover), then it is plausible that both very high and very low ability individuals would become generalists. This might explain why generalists tend to have higher variance in outcomes than specialists.

References

Acedo, Francisco Jose, Carmen Barroso, Cristobal Casanueva, and Jose Luis Galan, "Co-authorship in management and organizational studies: an empirical and network analysis," *Journal of Management Studies*, 2006, *43* (5), 957–983.

Anderson, Katharine, "Collaboration Network Formation and the Demand for Problem Solvers with Heterogenous Skills," December 2011.

Barabási, Albert-László and Réka Albert, "Emergence of Scaling in Random Networks," *Science*, October 1999, *286* (5439), 509–512.

Burt, Ronald S, "Structural Holes versus Network Closure as Social Capital," *Network*, 2001, *1* (May 2000), 31–56.

Galeotti, A, S Goyal, and J Kamphorst, "Network formation with heterogeneous players," *Games and Economic Behavior*, 2006, *54* (2), 353–372.

Gleiser, Pablo and Leon Danon, "Community Structure in Jazz," *Advances in Complex Systems*, 2003, *6* (4), 565–573.

Goeree, Jacob K, Arno Riedl, and Aljaž Ule, "In search of stars: Network formation among heterogeneous agents," *Games and Economic Behavior*, 2009, *67* (2), 445–466.

Goyal, Sanjeev and José Luis Moraga-González, "R&D Networks," *The RAND Journal of Economics*, 2001, *32* (4), 686–707.

———, Marco J Van Der Leij, and Jose Luis Moraga-Gonzalez, "Economics: An Emerging Small World," *Journal of Political Economy*, 2006, *114* (2), 403–412.

Iyer, Bala, Chi-Hyon Lee, and N Venkatraman, "Managing in a Small World Ecosystem : Some Lessons from the Software Sector," *California Management Review*, 2006, *48* (3), 28–47.

Jackson, Matthew O and Asher Wolinsky, "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 1996, *71* (1), 44–74.

——— and Brian W Rogers, "Meeting Strangers and Friends of Friends: How Random Are Social Networks?," *American Economic Review*, 2007, *97* (3), 890–915.

Moody, James, "The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999," *American Sociological Review*, 2004, *69* (2), 213–238.

Newman, M E J, "The structure of scientific collaboration networks," *Proceedings of the National Academy of Sciences of the United States of America*, 2000, *98* (2), 7.

Newman, M.E.J., "Scientific collaboration networks. I. Network construction and fundamental results," *Physical Review E*, 2001, *64* (1), 1–8.

Powell, Walter W, Kenneth W Koput, and Laurel Smith-Doerr, "Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology," *Administrative Science Quarterly*, 1996, *41* (1), 116.

Uzzi, Brian and Jarrett Spiro, "Collaboration and Creativity: The Small World Problem," *American Journal of Sociology*, 2005, *111* (2), 447–504.