Skill Specialization and the Formation of Collaboration Networks

Katharine Anderson
Tepper School of Business
Carnegie Mellon University
Pittsburgh, PA 15213
andersok@andrew.cmu.edu

Abstract

In this paper, I examine the role of skill specialization in collaboration network structure, and individual position in the collaborative community. Using a model of skill specialization and collaboration network formation, I show that as disciplines become less insular, the collaboration network becomes increasingly dominated by a small number of individuals. I compare specialists and generalists with the same number of skills and show that specialists will tend to have more links in the network than generalists with the same number of skills. However, I the show that generalists are more likely than specialists to occupy key central positions in the network.

1 Introduction

Problem solving and innovation are important in many contexts, including academic research, policy-making, product development, and entrepreneurship. Broadly speaking, innovators come in two varieties: specialists and generalists. Specialists are people who have a deep knowledge of a very narrow subject area. Generalists are people whose knowledge ranges across multiple subject areas.[1] In this paper, I examine the different roles that specialists and generalists play in collaborative problem-solving communities. Using a model of skills and collaborative connections, I show that while specialists tend to have more collaborators than generalists, generalists are much more likely to play a central role, connecting otherwise disparate communities.

Collaboration is a vital part of innovation and problem solving. By collaborating, individuals can pool their skills and solve problems that none of them could hope to solve alone. This set of collaborative interactions bind individual innovators into a collaborative community. One way of studying this community is by examining a collaboration network, in which any two individuals are connected if they have collaborated on a project.

Consider the collaboration network pictured in Figure 1.1. It is an academic co-authorship network, where two authors are connected if they have co-authored a paper together. The nodes are sized according to the number of connections they have. For a collaboration network, this reflects the number of collaborations an individual has. This network exhibits a characteristic that is typical of most collaboration networks, across contexts—the degree distribution of the network is skewed (see Figure 1.1), meaning that a small number of individuals participate in a large number of projects, while most individuals participate in relatively few.[4] The result is what is called a "hub and spoke"

[1]Isaiah Berlin would call these two types of people "foxes" and "hedgehogs". Foxes are people who know many things. Hedgehogs are people who know one thing, but know it well.

[4]This skewed degree distribution has been observed in a wide range of collaboration networks. Most prominantly, it can be seen in coauthorship networks in a variety of fields. Newman (2001) examines coauthorship networks for several subdisciplines of physics, biomedical fields, and computer science. Moody (2004) does

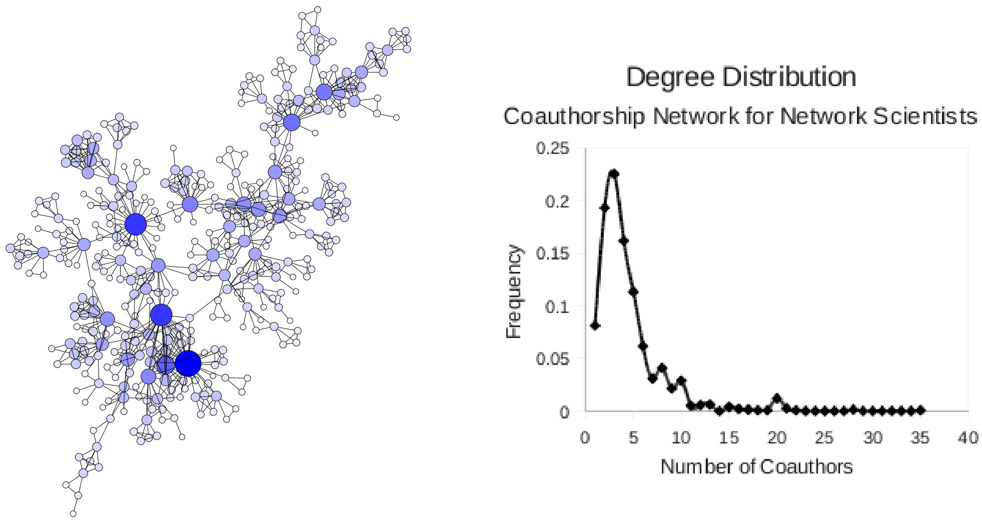

Figure 1.1: Panel A is an example of a collaboration network. This network is an academic coauthorship network, in which two indivdiuals are connected if they've coauthored a paper together.[3] The nodes are sized and colored according to their degree (in this case, the number of collaborators they have). Note that a few individuals have very high degree in the network, while most individuals have low degree. Panel B illustrates this degree distribution. This skewed degree distribution can be seen in a wide variety of collaboration networks. The high degree nodes will tend to have more power and influence in the community, and will have better access to information.

network, in which a large number of low-degree "spokes" are arrayed around a small number of high-degree "hubs". Although we can easily observe this structure empirically, it is an open question which factors determine how skewed that degree distribution is–in other words, can we identify characteristics of a collaborative community that will make a collaboration network more centralized around a few, high-degree individuals? This is an important question, because the structure of the social network governs many of the other behaviors that we care about–it shapes the nature of discourse by governing the flow of information and ideas, it can help or hinder communication, and because social connections guide the process of opinion formation, the social network can affect the time it take the community to reach a consensus.

Moreover, an individual innovator's position in the collaboration network is an important determinant of her outcomes. The hubs in the network in Figure 1.1 occupy a central position in the collaborative community. This gives them particularly good access to new information. They are also more likely to be the prestigious members of the community, wielding a disproportionate amount of influence. Individuals who form bridges between otherwise unconnected communities have a similarly special role in collaborative communities. Consider the network pictured in Figure 1.2. This network pictures two different collaborative communities, which are linked by a few individuals. Ronald Burt (2004) calls the gaps between different communities "structural holes". The individuals who bridge these structural holes have access to information and ideas from both communities. Burt also argues that because they control the flow of information between the communities, bridges have greater power in the community than those who do not bridge structural holes.

---

the same for sociology. Goyal et al (2006) looks at economists. Acedo et al (2006) present data on researchers in management and organizational studies, and while they do not directly address the degree distribution, their data includes more high-degree nodes than would be expected in a random network, suggesting a fat-tailed distribution. A skewed degree distribution has also been observed in interfirm collaboration (Powell et al (1996), Iyer et al (2006)), creative artists in broadway plays (Uzzi and Spiro (2005)), film actors (Barabasi and Albert (1999)), and jazz musicians (Gleiser and Danon (2003)).

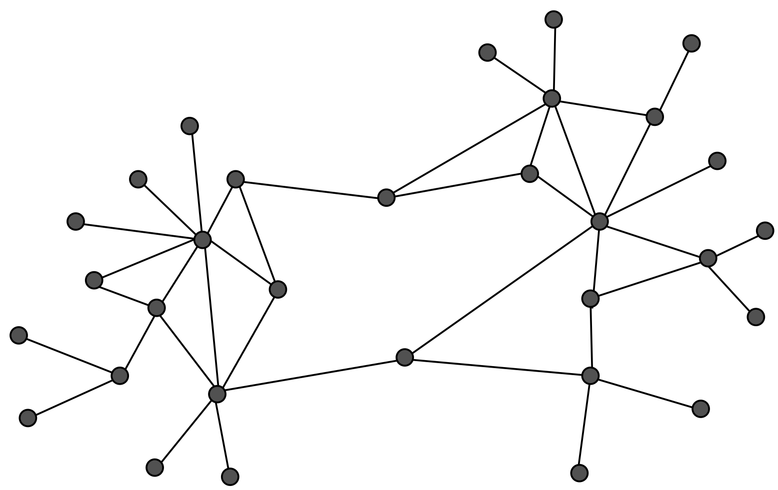

Figure 1.2: An example of a collaboration network with two seperate communities. The two nodes in the middle bridge a structural hole that exists between the two communities. These nodes benefit from being able to control the flow of information between the two communities.

This heterogeneity of individuals is one of the great strengths of using social networks to model collaborative communities. However, we do not yet have good models of what determines either the overall structure of a collaboration network, or an individual's place in a social network, and thus their position in the collaborative community. What factors distinguish hubs from spokes? What determines which innovators will bridge structural holes? In this paper, I consider one of the factors that contributes to different individuals taking different positions in a collaboration network—the specialization of that individual's skills.

Recently, there has been considerable attention given to interdisciplinary and multidisciplinary work. But is there any reason for an individual to choose an interdisciplinary set of skills? There is evidence that generalists pay a penalty for diversifying their skills—Adamic et al show that the contributions of generalists make less of an impact than the contributions of their more focused peers. Given that penalty, it is difficult to justify the decision to become a generalist in a non-collaborative environment. This can lead to an under-supply of generalists in the collaborative community. [5]

In this paper, I consider whether becoming a generalist confers some kind of network-based benefit in a collaborative community. In the following, I use a model of collaboration network formation to look at the positions of specialists and generalists in an endogenously-generated collaboration network. I show that when problems are uni-disciplinary, specialists will have more links than generalists with a similar number of skills. However, generalists will tend to occupy more central positions in the collaboration network. Thus, even when there are no interdisciplinary problems, being a generalist can provide network-based advantages for an individual problem solver.

2 Models of Collaboration Network Formation

The goal of this paper is to compare the positions of specialists and generalists on an endogenously-generated collaboration network. Thus, 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

[5] Anderson (2010)

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 structures common to most social networks, including the skewed degree distribution mentioned above. 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 factor that distinguishes low-degree nodes from high-degree nodes is age. Although this undoubtedly captures some of the variation in node degree in collaboration networks, there is surely some additional variation due to individual skills.

A second class of models, called decision-based models, allow individuals to make their linking decisions based on optimizing some kind of objective function.[6] 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. 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. (See Anderson (2010). In these models, individual problem solvers have skills, which are useful for solving a problem. These skills are pieces of knowledge, abilities, and tools that useful for solving problems, and are not easily passed from one individual to the next.[7] Individuals pool their skills by collaborating on problems, and the result is a collaboration network. Suppose, for example, we are interested in the network of interactions between entrepreneurs. In this case, the problem being solved is the formation of a new entrepreneurial firm. If this particular entrepreneurial venture is concerned with the development of a web application, then the required skills might be computer programming, user interface design, and marketing.[8] Although some, gifted individuals might have all of the skills required to start the venture, conventional wisdom is that most people will have only a small fraction of the required skills. Individual entrepreneurs collaborate with others who have complementary skills. The resulting collaboration is represented by a link between those individuals on the collaboration network. Thus, this class of models takes as an input a population of problem solvers, with a distribution of skills, and produces a collaboration network. The number of links that an individual has (her degree on the collaboration network) is the number of collaborations that she participates in—in this case, the number of ventures that she has a hand in founding.

3 The Bernoulli Skills Model

The Bernoulli Skills Model Anderson (2010) introduces a special case of this class of models, which she calls "The Bernoulli Skills Model", in which the probability of having any given skill is independent of the probability of having any other skill. The model in this paper generalizes the Bernoulli skills model, and thus it will be useful to start with this model before moving on. See Anderson (2010) for more details of this model.

Suppose there are N problem solvers. Each has a copy of a problem, ω, which requires M skills: $S = \{s_1 ... s_M\}$. Each individual will have a subset of those skills, $A_i \subseteq S$. In the Bernoulli Skills Model, the skills are distributed independently–an individual has any given skill with probability p. This means that the skill sets are the result of M Bernoulli random trials. Because skills are uncorrelated, it makes sense to define an individual's "ability" to be the number of skills she has, $a_i = |A_i|$. Note that in the limit of large N, the agents' abilities are distributed binomially, $a \sim b\,(M, p)$.

An individual may have all of the skills required to solve the problem, but she may not. If she lacks some of the required skills, she can work with another agent or agents on her problem. A *collaboration* is a subset of the problem solvers, C. An individual and her collaborators can *solve* a problem if together they possess all of the required skills–that is, if the problem is solved if $\omega \subseteq$

[6]See, for example, Jackson and Wolinsky (1996) and Goyal and Moranga-Gonzalez (2001)

[7]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.

[8]This is obviously an over-simplification, but it does illustrate the issues involved.

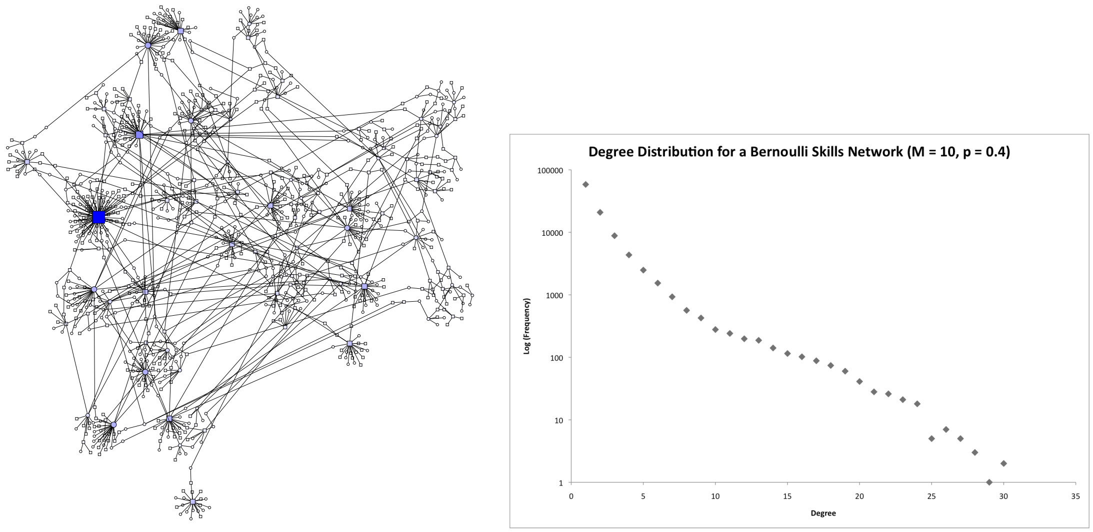

Figure 3.1: Panel A is a Bernoulli Random Skills network with 100 agents solving a problem requiring $M = 10$ skills. The probability of having each skill is independant, with $p = .4$ of having any individual skill. Panel B is the degree distribution of that network. Note that the number of individuals with a given degree is plotted on a log scale for ease of reading.

$\bigcup_{j \in C_i} A_j$. If the problem is solved, the collaborators get a payoff of 1. If they cannot solve the problem, the payoff is π.[9] Assume that she chooses a minimal set of collaborators[10] that allows her to solve the problem.[11] If there are multiple minimal sets of collaborators, she chooses one at random. By linking individuals who work together on a problem, we form a collaboration network.

Figure 3.1 illustrates both a network and the degree distribution for a simulation of a Bernoulli Random Skills network with 100 individuals, $M = 10$ skills, and $p = .4$ probability of having each skill.

4 Modifying the Bernoulli Skills Model–2 disciplines, 1 problem

In the Bernoulli skills model, skills are completely uncorrelated among problem solvers. In order to answer questions about specialists and generalists, we need to modify the distribution of skills in the community so that skills cluster together into disciplines. To model this kind of collaborative community, I divide the skills into two disciplines: A and B. The agents are divided into type A and type B according to whether they belong to discipline A or discipline B. The agents in discipline A face a problem requiring only type A skills and those in discipline B face a problem requiring only type B skills. The agents in each discipline have an easier time acquiring skills within their discipline than outside of it. In particular, agents in discipline A have the skills in discipline A with probability p_{own} and skills in discipline B with probability $p_{other} \leq p_{own}$. Similarly, agents in discipline B have skills in discipline B with probability p_{own} and those in discipline A with probability $p_{other} \leq p_{own}$.

[9] This particular production function could be written as

$$f(A_i, \omega, C_i) = \begin{cases} \text{payoff} & \text{if } \omega \subseteq \bigcup_{j \in C_i} A_j \\ 0 & \text{otherwise} \end{cases}$$

[10] There are many reasons that a problem solver might want to choose a minimal set of collaborators. Suppose, for example, there is a payoff to solving a problem which is split evenly between the problem solver and her collaborators. Then a payoff-maximizing problem solver will have an incentive to minimize the number of people she works with on the problem. A problem solver might also want to minimize the number of collaborators she has for interpersonal reasons (eg: because it simplifies communication or minimizes the possibility for conflict).

[11] Because the payoff to solving the problem is non-zero, this is incentive compatible.

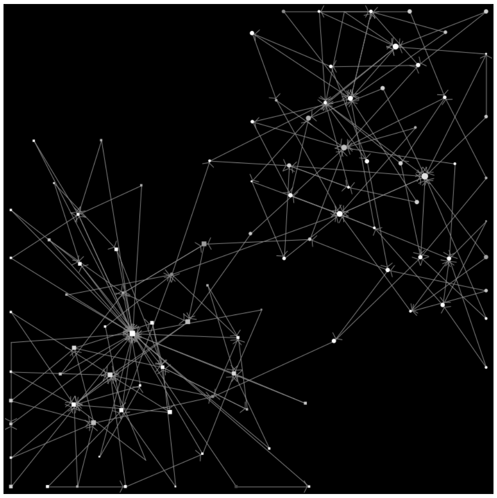

Figure 4.1: An example of a simulated collaboration network with two disciplines. In this case, there are 20 skills total–10 in each discipline. The probability of having a skill is $p_{own} = .35$ for skills in the agent's own discipline and $p_{other} = .05$ for skills in the other discipline.

The links are formed as in the Bernoulli Skills model–individuals in discipline A choose a minimal set of collaborators to solve their discipline A problem, and individuals in discipline B choose a minimal set of collaborators to solve their discipline B problem.

When p_{own} is strictly greater than p_{self} the resulting network has two distinct collaborative communities, as illustrated in Figure 4.1

The level of specialization in the network is parameterized by the pair $(p_{own,}p_{other})$. When the probability of having skills in the other discipline is low, compared with the probability of having skills in one's own discipline ($p_{other} << p_{own}$), the disciplines are highly codified, and skills are highly specialized. In the extreme case, if $p_{other} = 0$, then no individual can have skills in another discipline, and the collaborative community is divided completely in two. On the other hand, if $p_{own} = p_{other}$, then skills are uncorrelated, as in the Bernoulli Skills Model–individuals in the two disciplines still face different problems, but they have similar sets of skills.

The overall structure of the network depends on the level of specialization in the collaborative community. Figure 4.2 illustrates how the degree distribution changes for different values of p_{own} and p_{other}. In each of these networks, individuals have the same number of skills on average (that is, $\frac{M}{2}p_{own} + \frac{M}{2}p_{other}$ is the same in all cases). However, as p_{own} and p_{other} become more dissimilar, the degree distribution of the network becomes less skewed–that is, as skills become less specialized, a larger and larger fraction of the problems are solved by a smaller and smaller fraction of the problem solving community.[12]

This raises the question of who those individuals with the exceptionally high degree will be. Are they the specialists or the generalists? In the Bernoulli Skills Model, expected degree is strictly increasing in the ability of an individual. However, in this model, the relationship between degree and ability is a bit more complex. Even for a given ability level, specialists tend to have more links than generalists. For example, for the case with $M = 10$, $p_{own} = .35$, and $p_{other} = .05$, the following is the average degree of specialists and generalists in the network, by ability.

[12]Note that in the extreme case, where $p_{own} = p_{other}$, the two communities have the same distribution of skills, but half face problem A and half face problem B. In that case, the degree distribution of the network is the same as in the Bernoulli Skills Model.

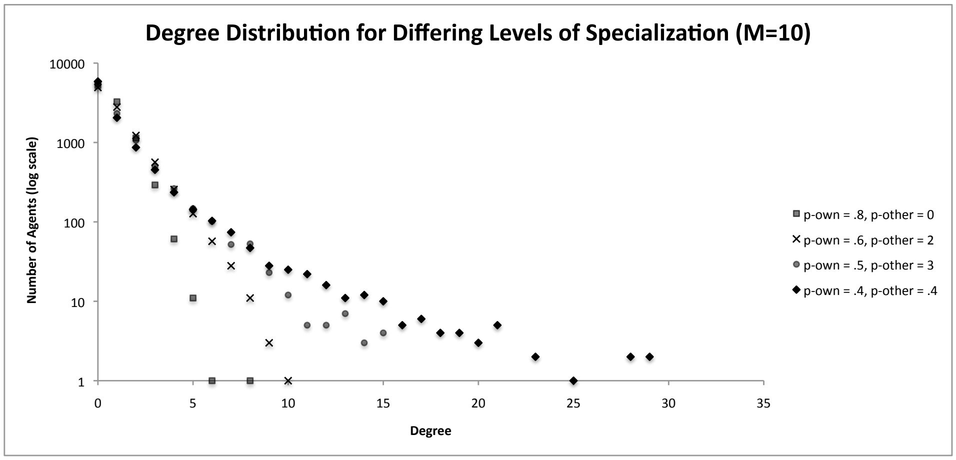

Figure 4.2: Degree distribution for collaborative communities with different levels of skill specialization. As p_{other} and p_{own} get further apart, the disciplines become more specialized, and the degree distribution of the network becomes less skewed. Note that the number of individuals with a given degree is plotted on a log scale for ease of reading.

Ability (# skills)	Average Degree, Specialist	Average Degree, Generalist
1	0.2	N/A[13]
2	0.7	0.3
3	2.4	0.8
4	8.6	2.3
5	24.9	6.8

Generalists have lower degree than specialists of the same ability because an individual's desirability as a collaborator within a community rises superadditively with the number of skills she can offer that community (an individual with skills a and b can help anyone who needs skill a, anyone who needs skill b, *and anyone who needs both*). By splitting their skills between the two communities, generalists don't benefit from the synergies between skills, and end up less attractive to both communities.

Looking at degree alone, one might wonder why any agent would ever choose to have skills in more than one discipline. However, degree is only one measure of an individual's role in the network–we might also consider a whether a node is central to the collaborative community. One particularly relevant measure of centrality in this context is a node's "constraint". The constraint of a node measures the extent to which a node's neighbors are similar to each other. Individuals with low constraint tend to bridge disparate communities. This is a powerful position in the network, because they benefit from being able to control the flow of information between communities. Although not all generalists in this model have low constraint, many of the individuals with the lowest constraint in these networks are generalists. For example, in the network in Figure 4.1, 5 of the 7 agents who have low constraint are generalists. This is an intuitive result–generalists have skills to offer to both collaborative communities, and thus are more likely to bridge those communities.

5 Conclusion

The results of this paper suggest that specialists and generalists will play very different roles in collaborative communities. When problems are focused within a single discipline, specialists will tend to participate in a larger number of collaborations, because their skills are all useful in combination with one another. Generalists tend to solve fewer problems, because only a subset of their skills are relevant in any given context. However, generalists are more likely to be the bridging members of the community, connecting otherwise disparate groups. This central role may provide incentive for generalists to diversify their skills, despite the apparent penalties they pay for doing so. However, since not all generalists play this bridging role, diversifying ones skills is still potentially a risky strategy. Given the extremely important role that these bridges seem to play in the collaborative process, this indicates that generalists may be providing a benefit to society that they are not being compensated for, and thus generalists may be undersupplied from a societal perspective.

References

[1] Adamic, Lada A., Xiao Wei, Jiang Yang, Sean Gerrish, Kevin Nam, Gavin Clarkson (2010) "Individual focus and knowledge contribution".

[2] Anderson, Katharine A. (2010) "Foxes and Hedgehogs: Equilibrium Skill Acquisition Decisions in Problem-solving Populations".

[3] Anderson, Katharine A. (2010) "Collaboration Network Formation and the Demand for Problem Solvers with Heterogenous Skills".

[4] Burt, Ronald S. (2004) "Structural Holes and Good Ideas," *American Journal of Sociology 110 (2)*, pp. 349-399.

[5] Hong, Lu and Scott E. Page (2001). "Problem Solving by Heterogeneous Agents," *Journal of Economic Theory, 97*, pp 123-163.

[6] Jackson, Matthew O. (2008). *Social and Economic Networks.* Princeton University Press: Princeton, NJ.

[7] Newman, M.E.J. (2006). *Phys. Rev. E,* 74, 036104.

[8] Phillips, Katherine W., Elizabeth A. Mannix, Margaret A. Neale, and Deborah H. Gruenfeld (2004). "Diverse groups and information sharing: The effects of congruent ties," *Journal of Experimental Social Psychology, 40,* pp. 497-510.

[9] Polzer, Jeffrey T., Laurie P. Milton, and William B. Swann, Jr. (2002). "Capitalizing on Diversity: Interpersonal Congruence in Small Work Groups," *Administrative Science Quarterly, 47 (2)*, pp. 296-324.

[10] Thomas-Hunt, Melissa C., Tonya Y. Ogden, and Margaret A. Neale (2003). "Who's Really Sharing? Effects of Social and Expert Status on Knowledge Exchange Within Groups," *Management Science, 49 (4),* pp. 464-477.