



Publicly Accessible Penn Dissertations

---

2017

# The Network Structure Of Collective Innovation

Devon Brackbill

University of Pennsylvania, [dbrackbill@asc.upenn.edu](mailto:dbrackbill@asc.upenn.edu)

Follow this and additional works at: <https://repository.upenn.edu/edissertations>



Part of the [Communication Commons](#)

---

## Recommended Citation

Brackbill, Devon, "The Network Structure Of Collective Innovation" (2017). *Publicly Accessible Penn Dissertations*. 2195.  
<https://repository.upenn.edu/edissertations/2195>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/edissertations/2195>  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

# The Network Structure Of Collective Innovation

## **Abstract**

Prior research on how to design collaboration networks among scientists, engineers, and strategists surprisingly predicts that inefficient networks that slow down the rate of collaboration will lead to better performance on complex problems. However, empirical research has provided mixed evidence for these ideas. Here, we test this theory using an online Data Science Competition that experimentally manipulates the network efficiency of teams working on a complex problem. The results support the idea that less efficient collaboration networks increase collective performance on complex problems. The results have important implications for designing problem-solving teams in numerous domains.

## **Degree Type**

Dissertation

## **Degree Name**

Doctor of Philosophy (PhD)

## **Graduate Group**

Communication

## **First Advisor**

Damon Centola

## **Keywords**

computational social science, data science, experiment, network structure

## **Subject Categories**

Communication

THE NETWORK STRUCTURE OF COLLECTIVE INNOVATION

Devon Brackbill

A DISSERTATION

in

Communication

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

---

Damon Centola  
Associate Professor, Annenberg School  
Associate Professor, School of Engineering  
Faculty, Warren Center for Network and Data Sciences  
University of Pennsylvania

Graduate Group Chairperson

---

Joseph Turow  
Robert Lewis Shayon Professor of Communication  
Associate Dean for Graduate Students  
University of Pennsylvania

Dissertation Committee  
Damon Centola, Associate Professor of Communication, Associate Professor School of Engineering  
Michael X. Delli Carpini, Walter H. Annenberg and Professor of Communication  
Joseph N. Cappella, Gerald R. Miller Professor of Communication

THE NETWORK STRUCTURE OF COLLECTIVE INNOVATION

COPYRIGHT

2017

Devon Brackbill

This work is licensed under the  
Creative Commons Attribution-  
NonCommercial-ShareAlike 3.0  
License

To view a copy of this license, visit

<https://creativecommons.org/licenses/by-nc-sa/3.0/us/>

*To my family.*

## **ACKNOWLEDGMENTS**

This dissertation was made possible by grants from the National Science Foundation Doctoral Dissertation Improvement Grant (Sociology #1519026), by the Wharton Risk Center Russell Ackoff Fellowship at the University of Pennsylvania, and by the Annenberg School for Communication at the University of Pennsylvania.

## ABSTRACT

### THE NETWORK STRUCTURE OF COLLECTIVE INNOVATION

Devon Brackbill

Damon Centola

*Prior research on how to design collaboration networks among scientists, engineers, and strategists surprisingly predicts that inefficient networks that slow down the rate of collaboration will lead to better performance on complex problems. However, empirical research has provided mixed evidence for these ideas. Here, we test this theory using an online Data Science Competition that experimentally manipulates the network efficiency of teams working on a complex problem. The results support the idea that less efficient collaboration networks increase collective performance on complex problems. The results have important implications for designing problem-solving teams in numerous domains.*

# TABLE OF CONTENTS

<b>ACKNOWLEDGMENTS .....</b>	<b>IV</b>
<b>ABSTRACT.....</b>	<b>V</b>
<b>TABLE OF CONTENTS.....</b>	<b>VI</b>
<b>LIST OF TABLES .....</b>	<b>VIII</b>
<b>LIST OF ILLUSTRATIONS .....</b>	<b>IX</b>
Figures in Chapter 1 .....	ix
Figures in Chapter 2 .....	x
<b>PREFACE .....</b>	<b>XI</b>
<b>CHAPTER 1: THE NETWORK STRUCTURE OF COLLECTIVE INNOVATION: AN EXPERIMENTAL STUDY .....</b>	<b>1</b>
Abstract.....	1
Introduction .....	1
Methods .....	2
Results .....	5
Discussion .....	9
<b>Supplementary 1: Design of the Experimental Study .....</b>	<b>10</b>
Experimental Design .....	10
Subject Recruitment.....	11
Subject Pool .....	12
Subject Experience During the Experiment .....	13
Data Science Problems.....	19
Network Metrics .....	23
Data Analysis .....	23
Subject Retention.....	25
Ensuring Data Quality .....	25
<b>Robustness Analysis.....</b>	<b>26</b>
Individual Rate of Exploration .....	26
Population Scaling Effects .....	27
Robustness to Design Choices: Allowing Non-Incremental Search .....	31



Attempts to Reduce Within-Network Variability .....	32
Power Analysis.....	33
<b>CHAPTER 2: APPLICATIONS OF NETWORK ENGINEERING TO TEAM PROBLEM SOLVING .....</b>	<b>36</b>
<b>Introduction .....</b>	<b>36</b>
<b>Basic Theory and Model.....</b>	<b>40</b>
Exploration-Exploitation Tradeoff .....	40
Efficiency and Performance .....	41
Model.....	41
Previous Results .....	42
<b>Experimental Design.....</b>	<b>48</b>
Task.....	48
Recruitment.....	51
Experimental Treatments.....	51
<b>Measures.....</b>	<b>53</b>
Diversity.....	53
Diffusion and Convergence Speed .....	53
Rate of Exploration.....	53
Performance.....	54
<b>Statistical Analysis .....</b>	<b>55</b>
Analysis .....	55
Attrition .....	56
<b>Experimental Results.....</b>	<b>56</b>
Solution Diversity .....	56
Speed of Diffusion.....	57
Speed of Convergence .....	58
Rate of Exploration.....	59
Performance.....	60
Average Group Performance .....	61
Cumulative Performance of the Average.....	62
Robustness to Non-Local Exploration.....	63
Discussion of Experimental Results.....	65
<b>Revised Model.....</b>	<b>66</b>
Simulation Results .....	67
Discussion of Simulation Results .....	69
<b>Discussion .....</b>	<b>71</b>
<b>BIBLIOGRAPHY .....</b>	<b>73</b>

## LIST OF TABLES

Table S1. Descriptive statistics of the data problems used in the study.

## LIST OF ILLUSTRATIONS

### **Figures in Chapter 1**

Fig. 1. Structure of the experiment.

Fig. 2. Inefficient networks found better solutions than efficient networks.

Fig. 3. Average performance in efficient and inefficient networks across time.

Fig. 4. Rate of solution diffusion.

Fig. 5. Diversity of solutions.

Fig. 6. Fraction adopting most popular solution.

Fig. S1. Schema of the experiment.

Fig. S2. Screenshot of the experimental interface when subjects explored their model.

Fig. S3. Screenshot of the experimental interface when a neighbor had a better solution.

Fig. S4. Screenshot of the experimental interface when a subject finished a round and adopted a better solution.

Fig. S5. Screenshot of the experimental interface when a subject finished a round and adopted a worse solution.

Fig. S6. Screenshot of the experimental interface when a subject finished a round and submitted the same solution.

Fig. S7. Screenshot of the experimental interface when a subject ran out of time on a round.

Fig. S8. Schema of simple (A) and complex (B) data science problems.

Fig. S9. Attempts to explore new solutions.

Fig. S10. Scaling effects of the best solution with population size.

Fig. S11. Scaling effects of the average (mean) solution with population size.

Fig. S12. Robustness to allowing non-incremental search.

Fig. S13. Power tests for "failure to copy."

## **Figures in Chapter 2**

Fig. 1. Simple and complex landscapes in one dimension.

Fig. 2. Screenshot of the experimental interface.

Fig. 3. Network structures in the experiment.

Fig. 4. Diversity of solutions.

Fig. 5. Rate of solution diffusion.

Fig. 6. Kaplan-Meier survival plot of the time until adoption of the top solution.

Fig. 7. Attempts to explore new solutions.

Fig. 8. Difference in performance between inefficient and efficient networks.

Fig. 9. Average performance in efficient and inefficient networks across time.

Fig. 10. Payoff to investing in inefficient networks relative to efficient networks.

Fig. 11. Performance differences between conditions was robust to allowing non-incremental search.

Fig. 12. The effects of network efficiency on collective performance.

## PREFACE

Do efficient communication networks increase collective intelligence? Scientists, engineers, and strategists all work within highly connected environments where each person's solutions are used to inspire and inform the work of others. The communication networks between researchers can determine the rate at which new ideas and innovations reach the rest of the community, giving rise to better solutions to difficult problems. As the complexity of the problem increases, so does the putative need for more efficient collaboration networks. Firms, research organizations, and universities have all invested in developing network technology to improve communications between researchers trying to solve complex problems. However, recent theoretical evidence suggests that these efforts may be counterproductive. These theories suggest that when teams are solving the most complex problems, increasing network efficiency can actually reduce the overall progress of members of communication networks. Current empirical studies testing these theories have provided mixed support, with one initial study providing supportive evidence, and a second larger study providing contrary evidence. These studies have relied upon simplified games that are meant to capture the central elements of group problem-solving, and they have recruited subjects using convenience samples, such as Amazon Mechanical Turk or undergraduate populations.

In order to address some of the concerns about external validity with previous research on this topic, this dissertation draws on research using “computational social science,” where the goal is to capture real world behavior while also maintaining precise causal control (Centola, 2010; Lazer et al., 2009; Salganik, Dodds, & Watts, 2006; van de Rijt, Kang, Restivo, & Patil, 2014). This dissertation tests these theories by gathering original data from an Internet-based experiment called the Annenberg Data Science Competition (<https://www.datascience.upenn.edu>). In order to recruit the subjects who solve these complex problems in the real world – statisticians and data scientists – I decided to build an online platform so that individuals could participate through their web browser from any computer with an Internet connection. I invited statisticians from across the country to solve complex statistical problems on

teams, and exogenously manipulated their collaboration network to see how it affected their ability to solve complex problems. The Annenberg Data Science Competition was modeled after crowd-sourced data science competitions, such as the Netflix Prize, the KDD Cup Challenge, and Kaggle.com, where data scientists from around the world compete to build the most accurate forecasting models from a data set.

By building a platform that attracts the people in the real world who normally solve such complex problems and by situating them in a realistic environment, this design allows me to capture the actual problem-solving behavior of individuals working on highly complex problems. Additionally, the experimental design allows me to causally identify the effect of network efficiency on collective performance.

The results support the idea that less efficient collaboration networks increase collective intelligence on complex problems. These results have important implications for the design of teams that are working on complex problems in design, engineering, and science.

The dissertation consists of two sections. **Chapter 1** (“The Network Structure of Collective Innovation: An Experimental Study”) is a concise presentation of the project for a general science audience who is interested in network theory and collective intelligence. It presents the main findings from the experiment, and defers further discussion of Design of the Experimental Study and Robustness Analyses to the end of this section. **Chapter 2** (“Applications of Network Engineering to Team Problem-Solving”) reports the results for a more specialized audience in management, business, and organizational theory who would be interested in finding practical ways to make use of the study’s findings. I first present the original model, and then show experimental results supporting that model. Then, I revise the model with novel simulations to show that collaborative efficiency can be changed in other ways beyond rewiring networks, as the original theory presupposed. I conclude by discussing practical ways to slow down collaboration among problem-solving teams using the theoretical results from this section.



# CHAPTER 1: THE NETWORK STRUCTURE OF COLLECTIVE INNOVATION: AN EXPERIMENTAL STUDY

## **Abstract**

Prior research on how to design collaboration networks among scientists, engineers, and strategists surprisingly predicts that inefficient networks that slow down the rate of collaboration will lead to better performance on complex problems. However, empirical research has provided mixed evidence for these ideas. Here, we test this theory using an online Data Science Competition that experimentally manipulates the network efficiency of teams working on a complex problem. The results support the idea that less efficient collaboration networks increase collective performance on complex problems. The results have important implications for designing problem-solving teams in numerous domains.

## **Introduction**

Do efficient communication networks increase the rate of innovation? In many complex problems, researchers, engineers, scientists, and designers face a tradeoff between exploring new possibilities by creating new solutions, or exploiting existing solutions by collaborating with others (Gupta, Smith, & Shalley, 2006; March, 1991). Theories from diffusion research suggest that when teams work to solve problems, more efficient collaboration networks would improve performance (Rogers, 2003; Strang & Soule, 1998). Such efficient networks would rapidly disseminate the most novel and high-quality solutions, which would improve group performance. However, recent theoretical work has indicated that excessive connectivity can undermine collective performance on complex problems (Fang, Lee, & Schilling, 2010; Lazer & Friedman, 2007). These theories predict that inefficient networks promote collective innovation by preserving the group's solution diversity, which prevents them from prematurely adopting a suboptimal solution. As a result, embedding teams in inefficient collaboration networks allows them to more



effectively navigate the large non-convex solution spaces that characterize many important problems in technology, science, and public policy (Kauffman, Lobo, & Macready, 2000; Kauffman & Macready, 1995; Lazer & Friedman, 2007).

Despite the importance of understanding how to build teams to solve complex problems, existing empirical tests have not been able to consistently support one hypothesis over the other. An initial empirical study supported the idea that inefficient collaboration would increase collective performance (Mason, Jones, & Goldstone, 2008), but a second larger experiment found the opposite (Mason & Watts, 2012). Both studies involved recruiting convenience samples to participate in a stylized online game that had some similarities to real-world complex problems. Here, we provide novel evidence to address the empirical disagreement by examining how collaboration affects domain experts in a real-world complex problem.

## **Methods**

To test the hypothesis that network efficiency decreases collective performance, we conducted a controlled experiment. Using a web-based platform, we recruited statisticians and data scientists from the World Wide Web to participate in a Data Science Competition (see Design of the Experimental Study). Individuals were given a regression problem where they needed to find the best predictive model, such as predicting the sales volume for a popular retailer. Subjects interacted with these data sets using a custom-built platform via their web browser that displayed their model performance on each round and allowed them to adjust their model. Models were scored based on a predictive accuracy metric (see Design of the Experimental Study).

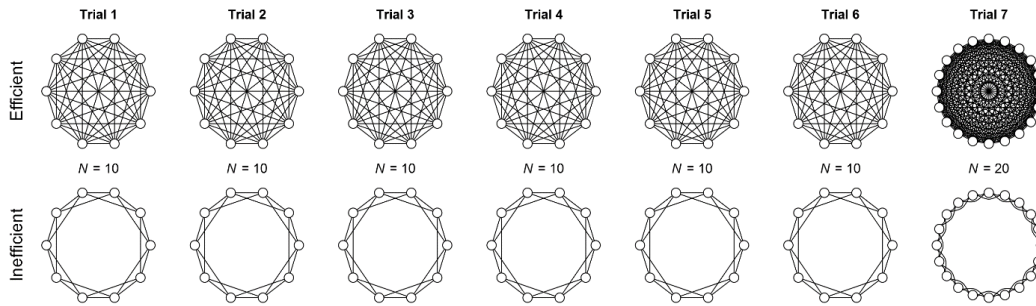
The competition lasted for 15 rounds, where each round required individuals to make a decision to either explore a new solution or exploit a neighbor's solution. Exploration meant deciding to change their solution by adding or removing a single variable from their statistical model, which was meant to capture how individuals incrementally search from their current solutions (Cohen & Levinthal, 1990; March, 1991; March & Simon, 1958; Nelson & Winter, 1982; Stuart & Podolny, 1996) (see Design of the Experimental Study). In contrast, exploitation meant

copying an existing solution from one of their network neighbors. To copy a better solution, individuals saw the performance scores of their best network neighbors, and could choose to adopt a better solution. Additionally, individuals could maintain their current solution. Once a participant submitted a decision for the round, they received immediate feedback on the performance of their new solution. If the solution was better, they adopted it, and proceeded to the next round, and if the solution was worse, they were returned to their previous solution. Each round lasted for 1 minute. Individuals never knew if they had the best possible solution, and they were incentivized to find better solutions each round, and received financial rewards based on the quality of their final solution.

Individuals solved a complex combinatorial optimization problem that shares essential features with complex problems in design, engineering, and complex problem-solving (Kauffman, 1993, 1995; Kauffman et al., 2000; Kauffman & Macready, 1995). Each solution was represented as a sequence of decisions to either include or exclude a variable in a statistical model. Each decision affected the overall fitness of the entire solution, and the problems were sufficiently complex so that there was a high degree of interdependency among the components of the solution (see Design of the Experimental Study), which is a hallmark of complex problems (Kauffman, 1993). This complex interdependency gave rise to solution “fitness landscapes” where teams could get caught on many locally optimal solutions (Kauffman, 1993; Wright, 1932). To ensure that our findings were general across numerous data science problems, we used several different problems across the experimental trials (see Design of the Experimental Study).

Participants in the study were randomly assigned to one of two collaboration network conditions – an efficient network with minimum possible average path length ( $L = 1$ ), which was a fully connected network, or an inefficient network with higher average path length ( $L = 1.67$  for  $N = 10$ , and  $L = 2.89$  for  $N = 20$ ), which was a ring lattice with average degree  $Z=4$  (Fig. 1). To make each group in a trial as similar as possible, each participant was given a random starting solution that was matched with someone in the other condition (see Design of the Experimental Study). This allowed us to see how differences in network structure could affect two initially

similar populations. The design resulted in 14 independent networks, and seven matched pair trials. Population sizes were fixed within a given trial, and we ran six pairs with  $N=10$  subjects in each network, and one pair with  $N = 20$  subjects, comprising 160 participants in the study. Subjects were recruited based on their statistical ability (see Design of the Experimental Study).

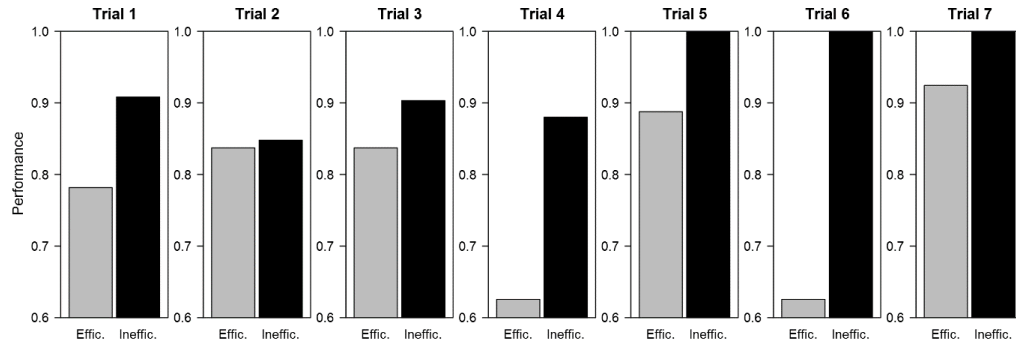


**Fig. 1. Structure of the experiment.** All experimental trials consisted of two networks, one efficient ( $L = 1$ ), and one inefficient ( $L=1.67$  or  $2.89$ ). In each of the seven trials, subjects were randomly assigned to one network condition, and then randomly assigned to a single node in the network. On the initial round of each trial, subjects received a random starting solution. On subsequent rounds, subjects saw the performance of their immediate network contacts and could copy these solutions. In a single trial, random initial solutions were matched across conditions, and groups faced the same data science problem.

Participants in the study were shown an identical user interface in both experimental conditions. Features of the social network, such as the average path length and the size of the population, were unobservable to participants (see Design of the Experimental Study). More generally, every aspect of the participants' experience was equivalent across experimental conditions. The only difference was the structure of the social networks. Thus, any differences in collective performance may be attributed to the effects of network efficiency on the process of collective innovation (see Design of the Experimental Study).

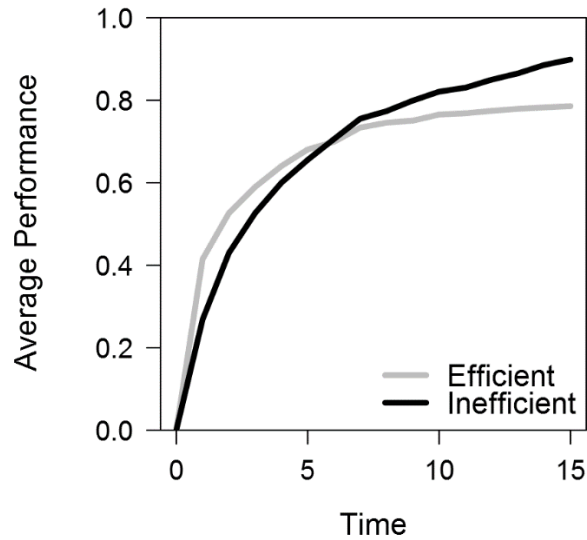
## Results

The results show that the efficiency of the collaboration network had a significant effect on the quality of collective innovation. We find that inefficient networks discovered better solutions than efficient networks ( $P = 0.02$ , Wilcoxon signed-rank test). This effect was consistent across each of the seven trials (Fig. 2). On average, inefficient networks found solutions that were 21% better than those discovered by efficient networks.



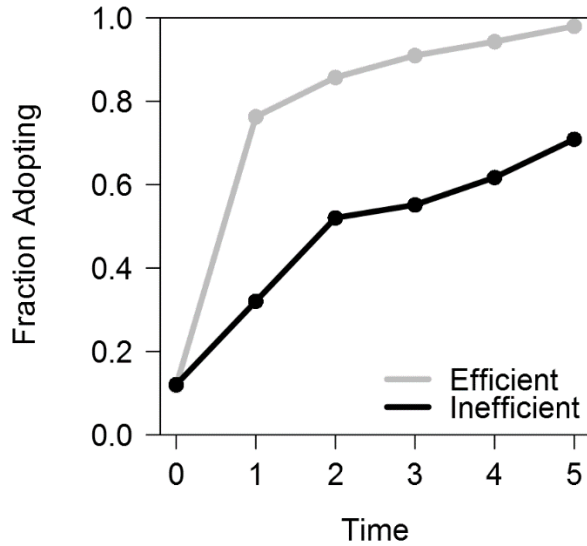
**Fig. 2. Inefficient networks found better solutions than efficient networks.** The maximum solution found by an individual in efficient (light) and inefficient (dark) networks is plotted across all seven trials. The performance of each solution is scaled based on the best possible solution on a given data set (=1) compared to the group's average starting performance (=0).

Initially, all inefficient networks had worse average solutions than their efficient network pairs, and on average their mean solutions were 30% worse than efficient networks ( $P < 0.05$ , Wilcoxon signed-rank). However, as the theory predicts (Lazer & Friedman, 2007), this suboptimal performance did not persist throughout the experiment (Fig. 3). By round 14, inefficient networks had significantly reversed the trend and were generating better average solutions than efficient networks ( $P < 0.05$ , Wilcoxon signed-rank). By the study's conclusion, every inefficient network had a better mean solution than its efficient network pair. On average, efficient networks generated mean solutions that were 17% higher than efficient networks on the final round ( $P = 0.02$ , Wilcoxon signed-rank).



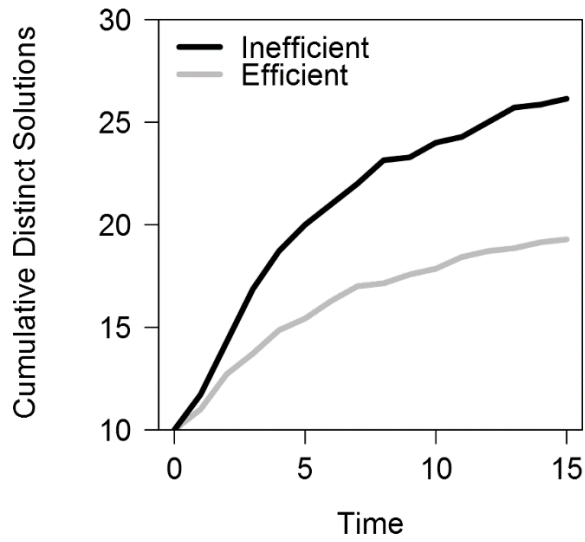
**Fig. 3. Average performance in efficient and inefficient networks across time.** Lines represent the average performance across all seven trials for efficient (light) and inefficient (dark) networks. Average performance was initially higher within efficient networks as compared to inefficient networks. However, by the study's conclusion inefficient networks had a better average performance than efficient networks within each trial.

The performance of inefficient networks was heavily influenced by the speed of solution diffusion, both depressing the initial average performance, but also preserving diversity and allowing for better solutions to arise. Diffusion rates were significantly lower in inefficient networks than in efficient networks ( $P < 0.01$ , Wilcoxon rank sum test). When the top solution was found in efficient networks, 76% of individuals adopted it on the following round on average (Fig. 4). In contrast, only 32% of individuals adopted the best solution on the next round after its discovery in an inefficient network.

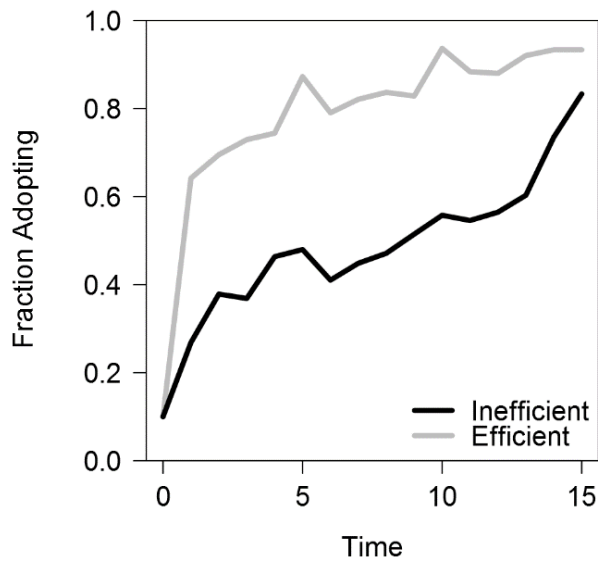


**Fig. 4. Rate of solution diffusion.** Proportion of individuals adopting the best solution after it was found at  $t=0$  over the following 5 rounds. Efficient networks took on average 5 rounds to diffuse the solution to the entire population. In contrast, inefficient networks did not see universal diffusion of the solution on average across all trials in the observation window of the study. The figure shows the mean fraction who adopt a group's best solution across seven trials.

The lower rate of solution diffusion in inefficient networks led to a greater diversity of solutions in these networks (Fig. 5). Inefficient networks discovered a larger portion of the solution space, on average successfully adopting 36% more distinct solutions across all time compared to efficient networks ( $P = 0.03$ , Wilcoxon signed rank test). Additionally, inefficient networks were significantly less likely to herd onto the most popular solution on each round ( $P < 0.05$ , Wilcoxon signed-rank test, for rounds 1 through 14) (Fig. 6). By maintaining more solution diversity, inefficient networks had a higher likelihood that additional explorations would find better solutions.



**Fig. 5. Diversity of solutions.** The cumulative number of distinct solutions discovered in efficient (light) and inefficient (dark) networks throughout time.



**Fig. 6. Fraction adopting most popular solution.** The average fraction of the population who adopt the most popular solution on each round is plotted across 15 rounds. There were fewer cases of duplicated solutions in inefficient (dark) than in efficient (light) networks.

## **Discussion**

As with all experiments, design choices that aided the control of the study also put constraints on the behaviors that we could test. A limitation of the design is that the subjects were experts in statistics who received a single complex problem and had to solve it in a limited time. However, increasing the length of time would make no difference for the results because individuals in efficient networks converged on a single solution and were unable to move to a better solution even with more time. Many real world complex problems have a similar high level of interdependency where individuals can get stuck on local optima, including problems in engineering, technology, and public policy (Kauffman et al., 2000; Kauffman & Macready, 1995). Additionally, the experts used in this study have many similarities with problem-solvers in other real-world domains, including in terms of their experience, approaches to collaboration, and search strategies. While the design restricted individuals to incremental search, research has shown that when faced with complex problems, animals, individuals, and organizations search incrementally and do not abandon solutions, so our design captures the essential elements of collective innovation (Cohen & Levinthal, 1990; March, 1991; March & Simon, 1958; Nelson & Winter, 1982; Stuart & Podolny, 1996). Further, the results were similar when we allowed individuals to search in a non-incremental manner (see Robustness Analyses).

Additionally, this study focused on one aspect of collective problem-solving, namely solution discovery among problem-solvers working in parallel (Lazer & Friedman, 2007). There are many additional aspects to collective problem-solving, such as efforts to coordinate and motivate members of large teams. When the logistics of coordinating large groups is the most pressing issue, then efficient networks should be used because they are known to promote faster and more universal adoption of a norm or technological standard (Centola & Baronchelli, 2015). Many of the large gains in productivity that have resulted from investments in communication technology have helped organizations find better ways to coordinate their behavior. Consistent with theory, more efficient collaboration would allow groups to perform better on these simple



problems. However, the large gains that have resulted from increases in efficiency on these simple problems will be at odds with group performance when they face the most complex problems, as the findings from this study suggest. Engineers, designers, and scientists face both simple and complex problems, and as a result, organizations may need to find ways to flexibly change their network structures depending on the complexity of the problem they face (Shore, Bernstein, & Lazer, 2015).

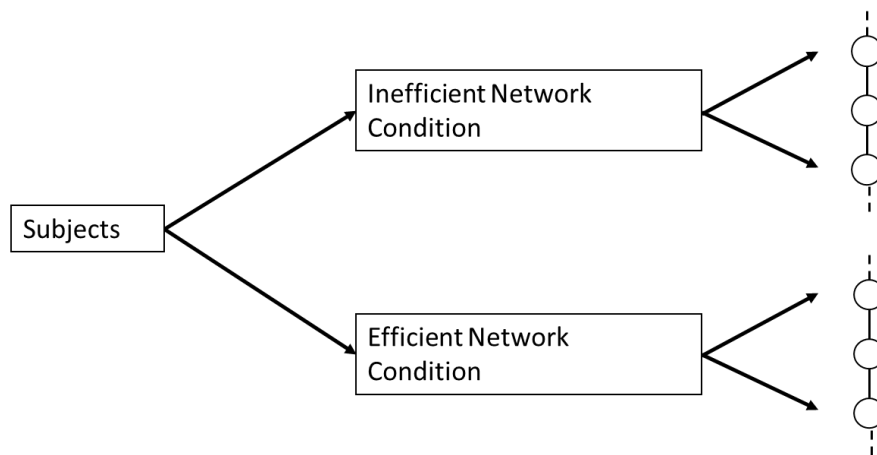
In contrast to previous research, our results show no benefits to efficient networks over inefficient ones when groups face complex problems (Mason & Watts, 2012). Instead, our findings suggest that inefficient networks may be an important part of collective innovation on complex problems. This finding agrees with theoretical (Fang et al., 2010; Kim & Park, 2009; Lazer & Friedman, 2007) and empirical (Mason et al., 2008; Mason, 2014) research on complex problems that finds inefficient networks promote collective solution diversity, which in turn improves the best solution and the average group solution in the long run. Concerns about groupthink, production blocking, and the common knowledge effect appear to be well-founded, and have the potential to prevent groups from finding the best solutions (Diehl & Stroebe, 1987, 1991; Janis, 1972, 1982). Surprisingly, finding ways to break ties, restrict information flow, and slow down collaboration may be an important way to increase the rate of discovery on the most complex problems.

## **Supplementary 1: Design of the Experimental Study**

### ***Experimental Design***

Each trial of the study consisted of a matched pair of networks, one efficient and one inefficient network. As subjects came into the study, they were randomized to one of the network conditions. The schema for this design is shown in Figure S1. Once subjects were randomized to a network condition, they were randomly assigned to one node in the network, and they maintained this position throughout the experiment. In each trial, both networks had the same size (either  $N=10$

or  $N=20$ ), but they differed in terms of their average path length. The networks also differed in terms of their degree, density, and diameter, but average path length is the central independent variable (Lazer & Friedman, 2007). Seven independent trials of the study were run. Across all seven trials, half of the subjects were randomly assigned to the efficient network condition and the remaining half were enrolled in the inefficient network condition. By holding all variables constant except network structure, we can identify how network efficiency affects collective problem-solving.



**Fig. S1. Schema of the experiment.** Each subject is randomly assigned to a network condition, and then randomly assigned to a single node within the network.

### ***Subject Recruitment***

Participants in our study were recruited via online advertisements posted on the World Wide Web to participate in the “Annenberg Data Science Competition.” When subjects arrived to the study website, they registered to participate by completing a form that required them to submit their email address, and choose an avatar and a username. All participants were required to provide informed consent in order to complete the registration process. Advertisements were placed online and direct emails were sent to several thousand interested participants. This recruitment

campaign generated 1,182 unique registrations in the system. From this pool, we invited subjects back to participate in the competition on a specific date. By following a link that only became active shortly before the competition began, participants could access the online platform at the specified date and time. When participants arrived at a live competition, they viewed instructions on how the competition worked, and waited while other subjects arrived. When a sufficient number of subjects arrived to conduct a single trial of the study (i.e. 20 individuals), all 20 subjects were then randomized to experimental conditions as described above, at which point the trial would begin in both conditions. The study was run for a 113-day period, December 10, 2015 through March 31, 2016 over which time online advertisements were posted to attract subjects to participate in the study. In total, 160 unique subjects were recruited to participate in the main study. Of the 160 individuals in the main study, 80 participated in inefficient networks, and 80 participated in efficient networks. An additional 20 subjects were recruited for a robustness trial, which is reported below.

### ***Subject Pool***

Due to the complex nature of the problem, we recruited subjects who were specialists in statistics. We wanted individuals who actually work on these complex problems, so that we could capture the behavior of real world teams of problem-solvers. The subjects in the study were skilled in statistics and quantitative methods. In order to participate, subjects had to understand how to run a linear regression model, and how to interpret coefficients, p-values, and model performance. While subjects did receive an introductory video, this video only described how the platform interface worked, and it did not include instructions about statistics. As a result, subjects were informed that the competition would be demanding on their skills, and that they should only participate if they had the skills to understand the statistics problems used in the study. All recruitment efforts were directed toward forums where individuals with quantitative skills would visit. To assess the statistical skills of the sample, we provided a voluntary survey question that

asked how many statistics courses they had taken. Of the individuals who responded to this question, the participants took a mean of 3 (sd=3.6) college-level statistics courses.

### ***Subject Experience During the Experiment***

To isolate the causal effect of network structure, the interface in each condition was identical. Individuals began a trial with a randomized initial solution. Then, individuals decided whether they wanted to revise their current model by adding or removing a single variable, or copy a solution from one of their neighbors. For example, one data set required individuals to find the variables (such as age, pH, and acidity) that predict taste ratings for wines (Fig. S2). Subjects made the decision to explore or copy by clicking on a radio button on the right side of the interface that allowed them to select their own solution, or select another player's solution (Fig. S2). The interface when subjects selected to revise their own solution is shown in Figure S2. Subjects could add or remove one variable by clicking the button with the variable's name. When they had made their choice for the round, they had to press the red "Submit" button on the right side of the interface. When a better option was available to copy, subjects saw the interface in Figure S3, which included a pop-up to indicate the better model. To copy a neighbor's model, individuals had to click on the radio button next to their model and then press the "Submit" button to end the round. The interface displayed the option to copy only when one of the individual's neighbors had a better solution. On rounds where the individual had the best solution in their local neighborhood, the interface defaulted to showing the interface to explore their model in Figure S2.

After submitting their decision to either copy or explore for the round, individuals received feedback based on the quality of their solution, and they waited while other players finished the round. When individuals found a better solution, they received a pop-up tracking their improvement as shown in Figure S4. When individuals tried a new solution, but it was not better than their previous solution, they received a notification that they would be returned to their previous solution as shown in Figure S5. If individuals decided to submit the same solution, they

received a notification indicating their choice as shown in Figure S6. Finally, if individuals made no choice before the timer ran out, they would remain at their previous model, and they received a notification as shown in Figure S7.

This sequence was repeated for 15 rounds in total, and each round lasted for 1 minute. The entire experiment lasted 15 minutes with an additional instructional video prior to the competition. We registered every click on each round—either decisions to explore or copy—so we had complete records of individual decisions. To motivate subjects, rewards were based on the quality of their final solution with a maximum payout of \$10. This design allowed us to examine the effect of network efficiency on the quality of group solutions.

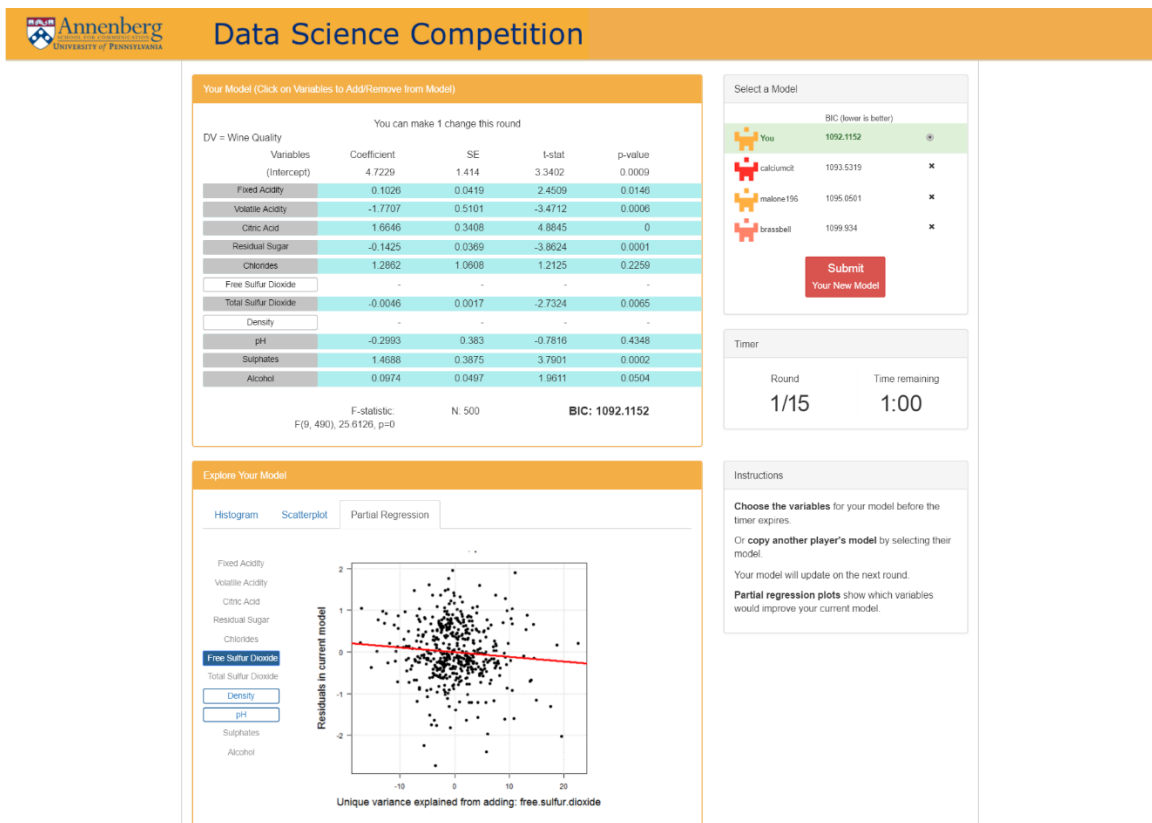


Fig. S2. Screenshot of the experimental interface when subjects explored their model.

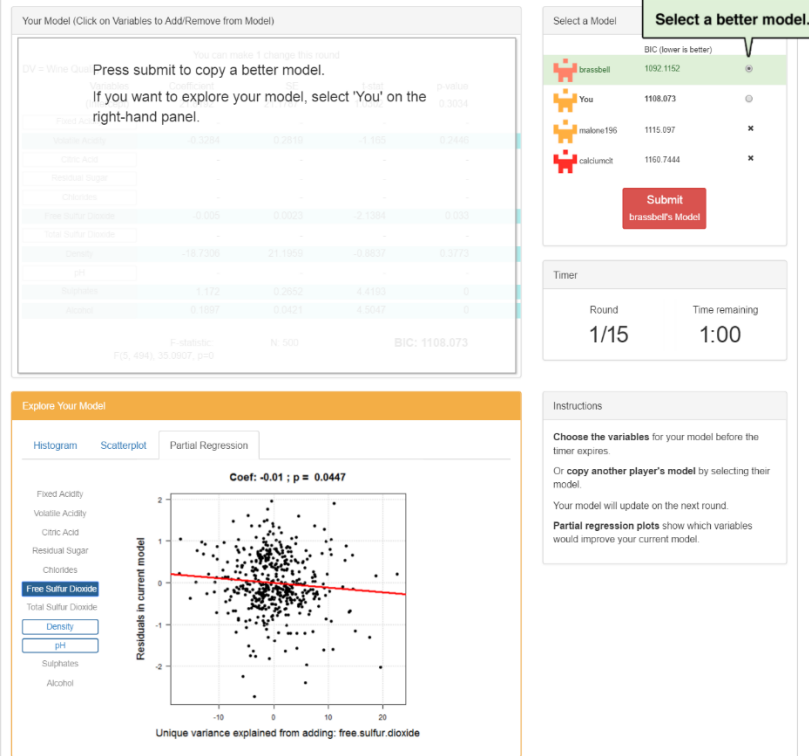
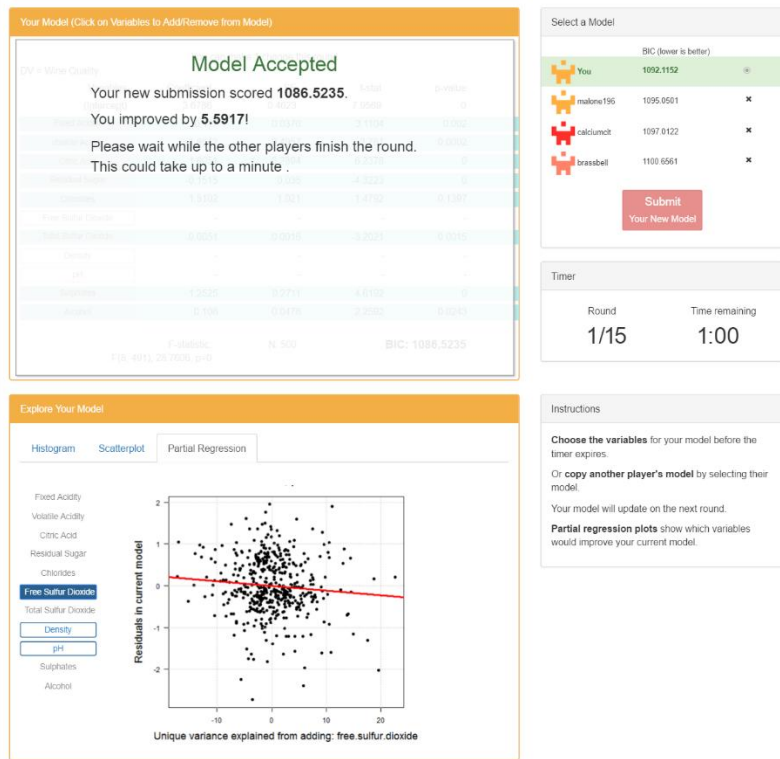


Fig. S3. Screenshot of the experimental interface when a neighbor had a better solution.



**Fig. S4. Screenshot of the experimental interface when a subject finished a round and adopted a better solution.**

Your Model (Click on Variables to Add/Remove from Model)

### Model Rejected

Your new submission scored **1118.4411**, which is not an improvement.  
 We've returned you to your previous model, which scored **1086.5235**.  
 Please wait while the other players finish the round.  
 This could take up to a minute .

Variable	beta	std. error	t	p-value
Free Sulfur Dioxide	0.117	0.4023	0.2910	0.7705
Density	0.0076	0.0010	7.1094	0.0000
pH	-0.0004	0.0002	-1.9964	0.0462
Sulphates	0.0001	0.0010	-0.2021	0.8438
Alcohol	1.2625	0.2711	4.6569	0.0000

F(8, 491), 29.7008, p=0      T-statistic:      N: 502      BIC: 1086.5235

Explore Your Model

Fixed Acidity

Volatile Acidity

Citric Acid

Residual Sugar

Chlorides

Free Sulfur Dioxide

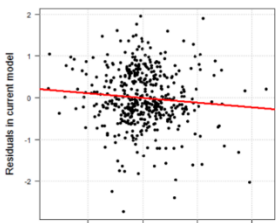
Total Sulfur Dioxide

Density

pH




Sulphates

Alcohol



Unique variance explained from adding: free.sulfur.dioxide

Select a Model

	BIC (lower is better)	
 You	1086.5235	
 malone195	1095.6501	✖
 calclumc01	1095.6501	✖
 brassbell	1095.6501	✖

Submit Your New Model

Timer

Round	Time remaining
1/15	1:00

Instructions

**Choose the variables** for your model before the timer expires.  
 Or **copy another player's model** by selecting their model.  
 Your model will update on the next round.  
**Partial regression plots** show which variables would improve your current model.

**Fig. S5. Screenshot of the experimental interface when a subject finished a round and adopted a worse solution.**



Your Model (Click on Variables to Add/Remove from Model)

## Model Unchanged

You submitted the same model, which scored **1092.1152**.

Please wait while the other players finish the round.  
This could take up to a minute

Variable	beta	std. error	t-statistic	p-value
Free Sulfur Dioxide	-0.1025	0.0410	-2.4999	0.0146
Volatile Acidity	4.7229	1.4114	3.3452	0.0009
Density	-0.0000	0.0000	0.0000	0.0000
Total Sulfur Dioxide	0.0000	0.0000	0.0000	0.0000
Residual Sugar	-0.1425	0.0386	-3.6624	0.0001
Chlorides	1.2882	1.0608	1.2122	0.2259
Alcohol	0.0000	0.0000	0.0000	0.0000
pH	-0.2093	0.3463	-0.6046	0.5446
Sulphates	1.4636	0.3675	3.9811	0.0000
Alcohol	0.0014	0.0467	0.0291	0.9764

F-statistic: 1092.1152    N: 502    BIC: 1092.1152  
 F(9, 490), 29.9726, p=0

Explore Your Model

Fixed Acidity

Volatile Acidity

Citric Acid

Residual Sugar

Chlorides

**Free Sulfur Dioxide**

Total Sulfur Dioxide

Density

pH

Sulphates

Alcohol

Select a Model

Player	BIC (lower is better)	Action
You	1092.1152	✖
caliumcit	1093.5319	✖
malone195	1095.6501	✖
brassbell	1099.934	✖

Submit Your New Model

Timer

Round	Time remaining
1/15	1:00

Instructions

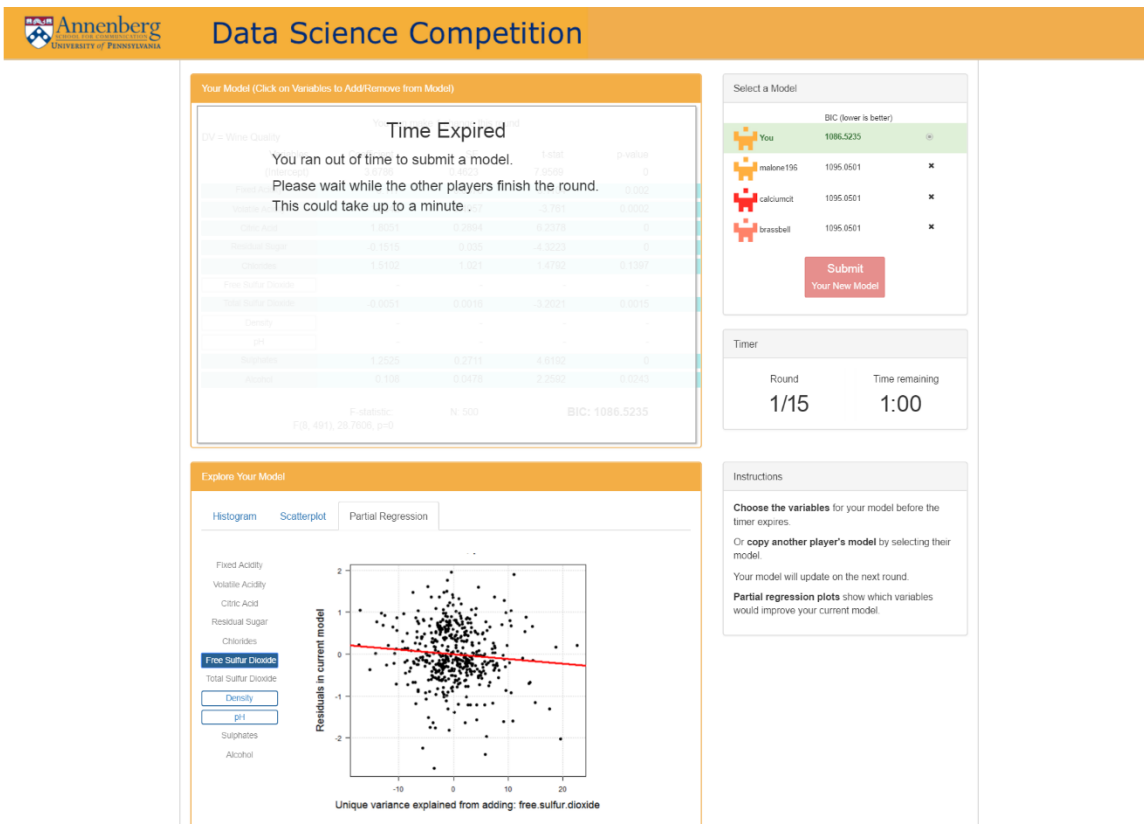
**Choose the variables** for your model before the timer expires.

Or **copy another player's model** by selecting their model.

Your model will update on the next round.

**Partial regression plots** show which variables would improve your current model.

**Fig. S6. Screenshot of the experimental interface when a subject finished a round and submitted the same solution.**



**Fig. S7.** Screenshot of the experimental interface when a subject ran out of time on a round.

### **Data Science Problems**

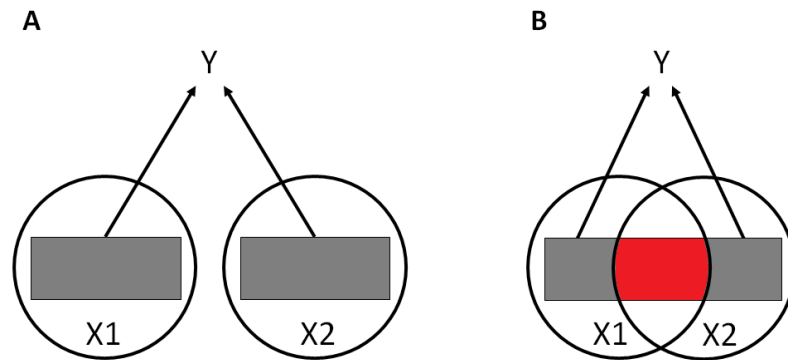
Each experimental trial involved a network of individuals who were invited to compete on a platform designed specifically to host a data science competition and to study this research question. Solutions were evaluated on each round based on the Bayesian Information Criterion (BIC) of their model, which provides continual feedback on their performance. The BIC is chosen because it is a good asymptotic measure of out-of-sample model performance and performs similarly to cross-validation predictive accuracy (Shao, 1997). The BIC rewards constructing sparse models that explain sufficient variance in the dependent variable. Individuals begin with a randomly assigned model and can explore from that point.

In order to ensure that the problem contains sufficient complexity, we used a method of creating data sets that draws directly on the *NK* model (Kauffman, 1993). The crucial feature of complex problems is that changing one dimension affects the fitness contribution from another dimension. Such synergies among the components produce many local maxima in the problem space so that incremental, local search can miss the global maximum. In contrast to  $R^2$ , which will always improve with the addition of more variables, performance metrics such as BIC penalize variables that do not provide additional information. In a simple landscape, where variables do not interact, each predictor variable contributes statistically independent information that improves the quality of the model. When a variable explains no variance in the predictor variable, its contribution is clear because the BIC will decline. In contrast, in a complex landscape, there are correlations among the variables, which is common in real world data sets. Variables often contain redundant information that is already captured by another variables. As a result, adding redundant information to the model will result in a worse BIC score and worse out-of-sample performance because the model is effectively being fit to noise that is idiosyncratic to the training data set.

To ensure that the data science problems in this study had this complex structure, we adopted existing data sets and increased the correlations among the variables. The procedure works by holding fixed the amount of variation in the dependent variable that is explained by all the predictor variables, but then shuffling that predictive variation among the predictor variables. By altering the correlation among the  $X$ 's, we can shift from a simple to a complex problem. In a complex problem, the contribution to the fitness of one variable depends on whether another variable is already included in the model or not. This interdependency among the components produces a complex fitness landscape.

An example is shown in Figure S8. For example, imagine  $X_1$  and  $X_2$  are predictive of the dependent variable,  $Y$ , but both are correlated with each other. If  $X_1$  is already included in the model, the addition of  $X_2$  might result in a worse BIC, particularly if the portion of each variable

that is predictive of  $Y$  is shared among the two variables. However, if  $X_1$  had not previously been in the model, the addition of  $X_2$  might explain even more variation and result in a better BIC. In contrast, if  $X_1$  and  $X_2$  are independent, then their joint inclusion in a model will always improve the BIC (assuming both predict the dependent variable). When faced with a complex landscape with many correlated variables, as is common in many real world data problems, greedy optimization can result in settling on local peaks rather than finding the best model, which is a well-known problem for step-wise regression techniques (Friedman, Hastie, & Tibshirani, 2001; Sribney, 2011).



**Fig. S8. Schema of simple (A) and complex (B) data science problems.** Each circle represents the variance in a variable,  $X_1$  and  $X_2$ . The gray components within each variable are the portions of each variable that are predictive of the dependent variable,  $Y$ . Panel A: In simple problems, the “fitness” of a predictor variable in terms of explaining the dependent variable depends only upon that variable. This means that each variable is independent of all other variables. Panel B: In complex problems, the “fitness” of a predictor variable depends upon the presence of other variables that are correlated with it. Even if  $X_2$  predicts the dependent variable, it will only improve the BIC score if it provides sufficient independent variation that is not accounted for by another variable in the model,  $X_1$ . Even though the entire variation predicted by  $X_1$  and  $X_2$  is the same as the simple problem, the “fitness” of adding new variables depends

upon how much additional variance explained beyond what is already explained by the existing variables. The red box represents covariation between X1 and X2 that is predictive of the target variable. Because this information is redundant, adding X2 to the model might not improve it any further. If the unique variance explained by adding X2 is too small (gray box within X2), then the BIC model performance will actually get worse by the addition of X2.

We applied this procedure to three data sets to generate complex data science problems. To ensure that the problems were sufficiently complex, we ran all linear regression models, and then counted the number of local optima in the solution space. A solution was a local optima when adding or removing any single variable would result in a worse solution. A simple problem should have one local optima, whereas a complex problem has several. The data problems were large and complex, with 2,048 to 16,384 possible solutions and 9 to 16 local optima, as shown in Table S1.

Trial	Data Set	Variables	Solutions	Local Optima
1	Wine	11	2,048	11
2	Viral News	14	16,384	9
3	Viral News	14	16,384	9
4	Sales Forecast	14	16,384	16
5	Sales Forecast	14	16,384	16
6	Sales Forecast	14	16,384	16
7	Sales Forecast	14	16,384	16

**Table S1. Descriptive statistics of the data problems used in the study.**

### **Network Metrics**

Average path length (or characteristic path length) is the mean geodesic or shortest path connecting two pairs of vertices (Wasserman & Faust, 1994). It is defined as the following for undirected graphs:

$$L = \frac{1}{n} \sum_{i \in N} L_i = \frac{1}{n} \sum_{i \in N} \frac{\sum_{j \in N, j \neq i} d_{i,j}}{n-1}$$

where  $L_i$  is the average distance between node  $i$  and all other nodes;  $d_{i,j}$  is the shortest path connecting nodes  $i$  and  $j$ ; and  $n$  is the population size. It is a measure of the efficiency with which information can flow through a network. Higher path lengths indicate less efficient communication networks, and lower path lengths indicate more efficient information spread. We use the undirected, unweighted version of this metric because of the experiment's design.

### **Data Analysis**

The performance of each model was measured in terms of its Bayesian Information Criterion (BIC), which is also known as the Schwarz Criterion (Schwarz, 1978). The BIC is a measure of a model's out of sample performance on a new data set that it has not been trained on. The BIC is a function both of the likelihood function and a regularization term that penalizes the addition of more parameters. The BIC is defined as:

$$BIC = -2 \bullet \ln \hat{L} + k \bullet \ln(n)$$

where  $\hat{L}$  is the maximum of the likelihood function of the model,  $k$  is the number of free parameters to be estimated, and  $n$  is the number of observations in the data set. In the case of a linear regression used in the experiment,  $k$  is the number of regressors including the intercept in the model.

To create a measure of group performance, we rescaled the BIC metric onto the range [0,1]. Since lower BIC indicates that a model is a better fit to the data, we created a measure of group top performance by transforming the BIC as follows:

$$Best_i = \frac{mean_i(BIC_{t=0}) - \max_i(BIC_t)}{mean_i(BIC_{t=0}) - \max^*(BIC)}$$

where the numerator is the difference between group  $i$ 's average starting solution on the initial round ( $t=0$ ) and the maximum solution on the current round  $t$ , and the denominator is the difference between the group's average starting solution and the best possible solution. To capture average group performance, we use the following formula:

$$Average_i = \frac{mean_i(BIC_{t=0}) - mean_i(BIC_t)}{mean_i(BIC_{t=0}) - \max^*(BIC)}$$

where the numerator is the difference between the group's average solution at  $t=0$  and the group's average at time  $t$ , and the denominator is the same. Both metrics range from 0 (the group's initial solution) to 1 (the best possible solution), and indicate how much groups have improved from their initial starting solution. The average initial starting solution was fixed between conditions within a given trial because we matched the same starting solutions between conditions, and the best possible solution was fixed for each data set. As a result, this metric is directly comparable between conditions so long as individuals started from the same solutions in both conditions and they used the same data problem.

To assess the performance of the best solution each group found, we compared  $Best_i$  for each network structure in the seven trials using a Wilcoxon signed-rank test. This test is a non-parametric test for matched pairs comparing the probability that observations from one condition will be greater than those from another condition. In essence, it tests whether it is more likely than chance within each matched pair that one group will consistently have a larger value than the other. It is very similar to the paired  $t$ -test, but it provides a more conservative estimate of significance because it does not assume a normal distribution. We found that the null hypothesis that there was no difference in the top solution across conditions could be accepted with a probability of  $P = 0.02$ . All statistical tests used a two-sided test of significance.

To examine the average performance of groups, we compared  $Average_i$  in each condition using the Wilcoxon signed-rank test after the initial round. We also conducted this test

on the final round. To construct aggregate statistics across all seven trials, we also averaged across all efficient and inefficient networks.

To examine the rate of diffusion of top solutions, we counted the fraction of individuals who adopted the best solution on every round following its discovery. Because the experiment only lasted for 15 rounds, there is some missing data because groups may have found their top solution very close to the end of the experiment.

To examine the difference in diversity between conditions, we compared the set of unique solutions that were adopted in all fifteen rounds using a Wilcoxon signed-rank test. We tested differences between conditions using the Wilcoxon signed-rank test. To quantify herding, we calculated the proportion of individuals who adopted the most popular solution on each round, and then calculated the average across the seven trials. We tested differences in the herding rate using the Wilcoxon signed-rank test.

### ***Subject Retention***

The experiment had a high retention rate, with 86% of all subjects completing the final round. There was no significant difference in retention rates across conditions, with 85% of subjects finishing the study in inefficient networks, and 87% finishing in efficient networks ( $P = 0.75$ , Wilcoxon signed-rank test). The most common reasons for attrition were due to network connectivity issues, where the platform would disconnect a user if they closed their browser tab. We used the data from an individual until they left the study, or completed the final round.

### ***Ensuring Data Quality***

We took several precautions in order to ensure that subjects did not violate the design of the experiment. Such precautions can be more difficult in online experiments because researchers may have less control over the behavior of the subjects than in traditional laboratory settings. We took several steps to ensure that the data was sound. In order to prevent individuals from participating in the study multiple times, we designed the system so that if a user tried to use a second browser tab to simultaneously participate, the system would produce an error, and only

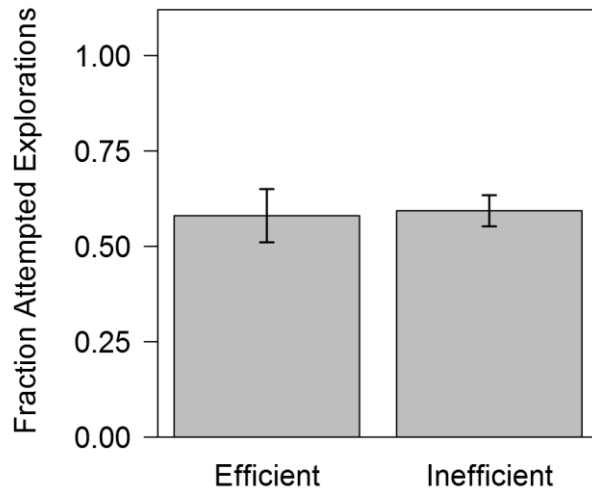


allow one active browser tab to communicate on the same computer. Additionally, we required users to enter their email address before playing the game, and all payments were sent to these addresses, which made it more difficult for users to gain access to the system multiple times. To do so, a user would have had to enroll with multiple email addresses. Even if users were able to bypass these measures, each trial of the study used a new data problem so that repeat users would not have any advantage over new players. The interface was explained with a video instruction as users waited for the game to start, so there was very little reason to believe that there was any skill or learning that could occur from having played the game before.

## **Robustness Analysis**

### ***Individual Rate of Exploration***

Previous research has hypothesized that the mechanism through which inefficient networks promote better solutions is that individuals are incentivized to explore more in inefficient networks (Mason & Watts, 2012). This increased exploration is expected to decrease the likelihood that the collective will converge on a premature suboptimal solution. Our results do not support this hypothesis (Fig. S9). Instead, network structure did not affect the rate of exploration between condition ( $P = 0.93$ , Wilcoxon signed rank test). The rate of individual exploration was measured as the number of times an individual made a revision to their model and pressed the “submit” button.



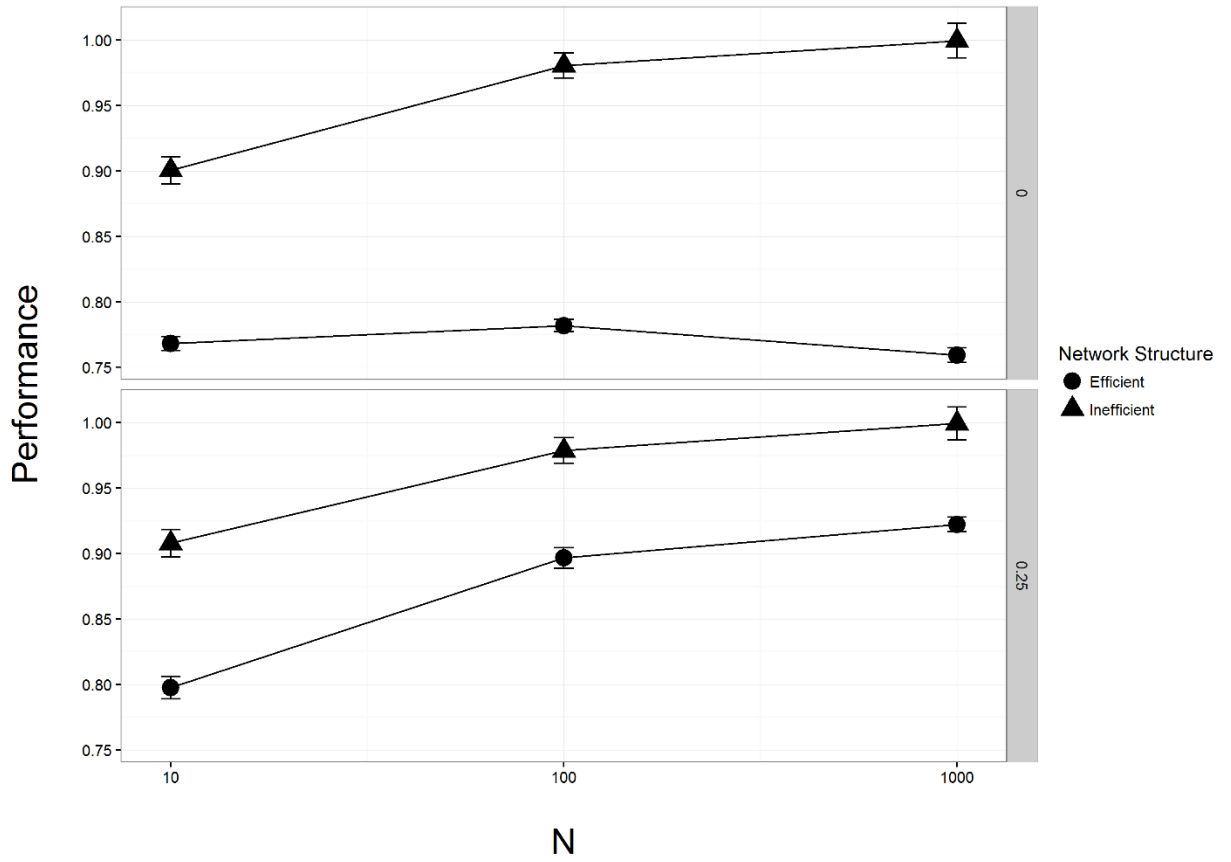
**Fig. S9. Attempts to explore new solutions.** The average fraction of plays that were attempts to explore are plotted for each network across all trials. Error bars are two standard errors of the mean.

***Population Scaling Effects***

To examine how the results scale to large populations, we conducted simulations with increasing population size. The simulations use a data science problem that was used in the experiment, and allow agents to interact for 15 rounds. We used the exact starting solutions that were used in the experiment for this data science problem. We then test the differences between fully connected networks (Efficient) and lattices where every node has degree = 4 (Inefficient) for population sizes  $n = 10, 100, \text{ and } 1,000$ . The average path lengths in the inefficient network increase from  $L = 1.7, \text{ to } 12.9, \text{ to } 125.2$ .

The results for the best solution scale to larger populations (Fig. S10). As population size increases, inefficient networks perform better because it is more likely that the exploration in inefficient networks will find a better top solution. In contrast, in efficient networks performance does not increase because early copying has locked in efficient networks on suboptimal solutions.

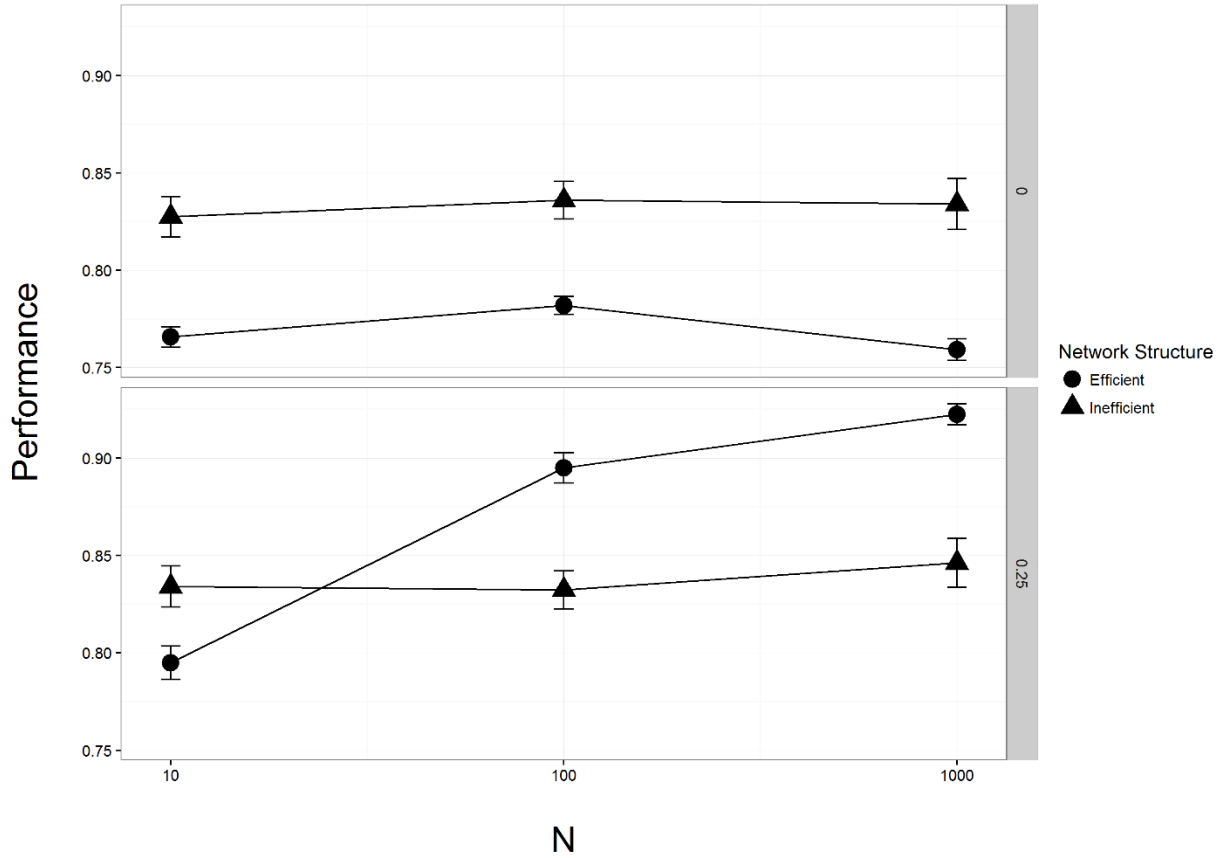
To examine how the results vary with different assumptions about how frequently individuals prefer to explore even when a better solution is available to copy, we then vary how often individuals make this “error.” The bottom row of Fig. S10 shows the results when individuals see a better solution, but choose to explore instead of copying 25% of the time, which was the rate of exploration observed from the empirical data. The results are very similar for inefficient networks. In contrast, for efficient networks, performance increases because the greater diversity that results from additional exploration helps prevent groups from getting stuck on a local optimum. This improvement from individual exploration is not sufficient to equal the performance of inefficient networks, which indicates that the effect of network structure on group performance in this experiment is larger than the effect of individual preferences to explore.



**Fig. S10. Scaling effects of the best solution with population size.** Performance of the best solution is plotted for 3 population sizes (10, 100, and 1000) with two levels of error (0, and 0.25). An error of 0 indicates that when an agent sees a better solution, it fails to copy that solution 0% of the time, and 0.25 indicates a 25% failure to copy a better solution. An error of 0.25 is similar to the effects observed in the empirical data. 100 simulations for each point. Simulations were run on a single data science problem from the experiment for 15 rounds.

The average group performance is not affected by scaling to larger population sizes when individuals always copy the best solution (Fig. S11, top). However, when individuals have a 25% preference to explore even when a better option exists to copy, the efficient network performs better in terms of its average (Fig. S11, bottom). At larger population sizes, the efficient

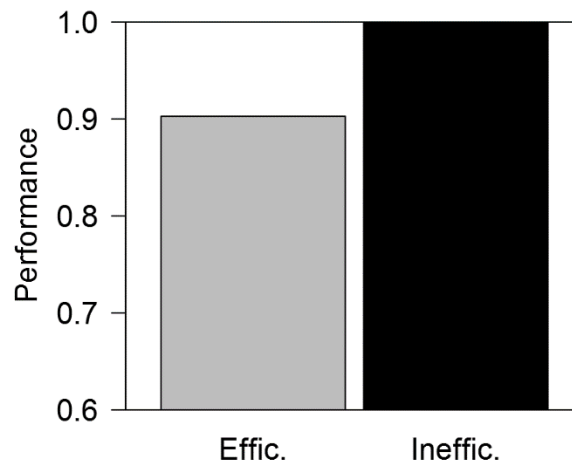
network is able to spread a good solution to more individuals in the population, thus lifting up the group's average performance.



**Fig. S11. Scaling effects of the average (mean) solution with population size.** Performance of the average solution is plotted for 3 population sizes (10, 100, and 1000) with two levels of error (0, and 0.25). An error of 0 indicates that when an agent sees a better solution, it fails to copy that solution 0% of the time; an error of 0.25 indicates the agent would fail to copy a better solution 25% of the time. 100 simulations for each point. Simulations were run on a single data science problem from the experiment for 15 rounds.

### ***Robustness to Design Choices: Allowing Non-Incremental Search***

To examine the sensitivity of the results to design choices, we conducted a robustness experiment. Our original experiment allowed individuals to make a single revision each round (i.e., incremental search). We chose this design because we wanted to capture realistic search processes by individuals and organizations in high-risk situations, where there are strong incentives to add slowly to a solution that has received heavy investment (Cohen & Levinthal, 1990; March, 1991; March & Simon, 1958; Nelson & Winter, 1982; Stuart & Podolny, 1996). In the robustness experiment, we permitted individuals to make as many changes to their model on each round before submitting their new solution and receiving feedback (i.e., non-incremental search). We ran a single trial, comparing an inefficient network to an efficient one. The results show that allowing individuals to search non-incrementally does not change the differences between conditions (Fig. S12). The inefficient network still performed better than the efficient network even when both were allowed to make non-incremental searches.



**Fig. S12. Robustness to allowing non-incremental search.** The maximum solution found by an individual in an efficient (light) and inefficient (dark) network where subjects were allowed to make non-incremental searches on all rounds. The performance of each solution is scaled based

on the best possible solution on a given data set (=1) compared to the group's average starting performance (=0).

At the individual level, individuals did attempt to explore more widely when given the option. Across all attempts to explore, 38% of attempts involved non-incremental search. However, a majority of these attempts were unsuccessful, and individuals were more successful when they explored incrementally. When incremental exploration was used, individuals successfully found a better solution 19.4% of the time, in contrast to a success rate of only 9.8% for attempts to change more than one element of their solution. This result confirms the intuition that incremental exploration is both preferred by individuals, and also represents a more reasonable choice because the quality of the solution will likely be more similar to the current solution, and will likely be better.

#### ***Attempts to Reduce Within-Network Variability***

Within a complex landscape like the ones used in this study, there is considerable variability in group performance throughout the search process. A decision by a single individual to revise one component of their solution can directly affect the diversity of the entire group and the direction that the group can explore. As a result, it is possible that the within-network variation might mask any between-network variability.

We took several precautions to minimize the variation within each network. The experiment used a matched pair design, where each individual in the inefficient network was given the same starting solution as an individual in the efficient network. Additionally, we provided individuals with suggestive information about which solutions might be better. This information came in the form of added variable plots at the bottom of the interface, where individuals could see if adding another variable would likely improve or worsen their solution. This information allows search to be much more efficient than simple random changes to the solution string, which is how the theoretical model operationalizes search (Lazer & Friedman, 2007). Additionally, the

variable names were real, and had actual correlation with the dependent variable, so any intuitions that individuals had about the causal relationships would help them explore more effectively. These features are expected only to speed up the dynamics so that the differences between conditions become more quickly apparent.

Individual preferences to explore even when a better solution is available could diminish the differences between the network conditions. At the extreme, if individuals always explore and never copy, there will be no difference between the efficient and inefficient collaboration structures. While this “failure to copy” is a function of individual preferences, designing an interface that clearly demarcates the options that a user faces can help decrease instances of failing to copy. We designed the system defaults in the interface so that the best choice on any given round was made immediately apparent to users via pop-up boxes and prompts so that decisions to explore their model or exploit an existing solution could be made efficiently without cognitive interference (Fig. S3).

### ***Power Analysis***

To examine the likelihood that the experiment would detect the effects of network efficiency on collective performance, we conducted numerous power tests using simulations. Traditional power analysis tests in individual-level experiments begin by specifying an assumption about the effect size of the experimental manipulation. This assumption is either based on past empirical studies in the same research topic, or from estimations and intuitions about the effect size from with related studies. This model of power analysis is not appropriate for collective-level experiments. Instead, we constructed agent-based simulations using the exact data science problems in the experiment to test if collective performance would emerge from different assumptions about individual-level behavior.

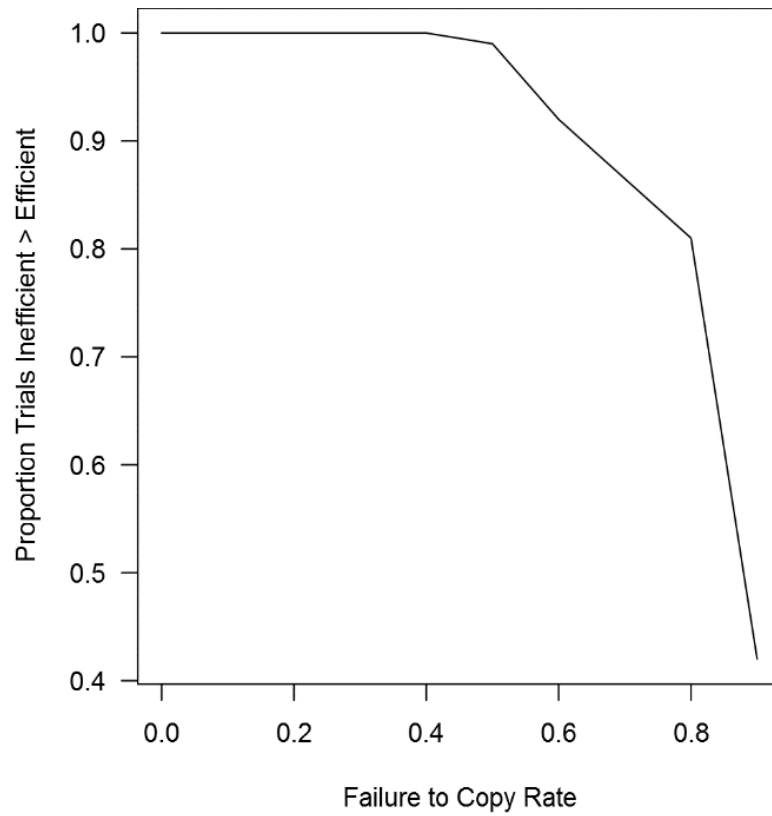
There are two individual-level parameters that affect the likelihood of detecting a difference between the network conditions at the collective level. First, individuals differ in terms of their willingness to explore even when a better solution exists. This “failure to copy” parameter



could affect the ability to detect network differences. At the extreme, if individuals always explore and never copy, there will be no difference between the efficient and inefficient collaboration networks. Second, individuals may differ in how skillful they are when they explore. Skill in exploration means that an individual is more likely than chance to choose a variable that will improve their model, either from intuition about what variables will be effective, or from an understanding of the statistical information presented to the user. If individuals are better than chance at exploring, then the timescale for the effects will be increased.

To measure the sensitivity of the results to these unknown parameters, we conducted agent-based simulations that varied the degree to which individuals “failed to copy.” In all of the simulations, we assumed that individuals were no better than chance when they explored, so these tests provided a conservative estimate of the timescales. The results reported here are for  $N=10$ , using the final data science problem, where agents begin on the exact starting locations used in the experiment. We ran 100 simulations for 15 rounds at each value of the Failure to Copy Rate. Similar results were found for all the other data science problems.

The results show that as the rate of failing to copy increases, the differences between networks become smaller (Fig. S13). The effect is gradual, however, which indicates that for a large range of individual preferences to explore new solutions rather than exploit existing options we will be able to detect a significant difference between conditions. In the study, about 25% of individuals failed to copy the best solution in the efficient network on the round immediately following its discovery. Figure S13 indicates that even when individuals ignore better solutions 25% of the time, we can still expect the inefficient network to outperform the efficient network on a majority of the trials. This suggests that we have a high probability of observing the predicted effects in our empirical setting.



**Fig. S13. Power tests for “failure to copy.”** As rates of failing to copy increase, the ability of the experiment to detect differences between the network conditions diminishes.

## CHAPTER 2: APPLICATIONS OF NETWORK ENGINEERING TO TEAM

### PROBLEM SOLVING

#### Introduction

Solving complex problems requires teams of researchers. Examples include physicians discovering a correct diagnosis and prescription (Coleman, Katz, & Menzel, 1957), pharmaceutical engineers refining drugs to slow cancer growth, software developers moving through cycles of design and testing to refine the efficiency of a new program (Graham & Sichelman, 2010), state governments crafting public policy based on information about past effectiveness (Walker, 1969), or financial analysts updating portfolio allocations based on previous performance (Pan, Altshuler, & Pentland, 2012). Successful performance on a complex problem depends crucially on a team achieving the correct balance between innovation and collaboration, or what the theoretical literature has called exploration and exploitation (Gupta et al., 2006; March, 1991). For each individual, the decision to pursue innovation, i.e., to independently explore the solution space, is risky and entails the costs of time and effort. However, it offers the potential rewards of discovery, where a new innovation could dramatically improve the group. Alternatively, individuals can choose to exploit the knowledge already existing in their networks. This strategy will not reveal any new solutions, but it could improve their relative performance on the problem and help spread known solutions to others. Individuals' decisions to innovate or collaborate translate into group-level performance – either accelerating the process of collective discovery or hastening the diffusion of previous solutions.

In order for an organization to survive, it must find a way to balance the tradeoff between innovation and collaboration (March, 1991). Numerous studies have suggested that an organization's collaboration network can be used to balance the forces of innovation and collaboration (Benner & Tushman, 2003; Ethiraj & Levinthal, 2004; Fang et al., 2010; Mihm, Loch, Wilkinson, & Huberman, 2010; K. D. Miller, Zhao, & Calantone, 2006; O'Reilly & Tushman, 2004;

Raisch, Birkinshaw, Probst, & Tushman, 2009; Rivkin & Siggelkow, 2007; Siggelkow & Rivkin, 2005, 2006; Taylor & Greve, 2006). In particular, theoretical findings suggest that well-designed research networks can improve the rate of scientific and technological discovery. However, existing empirical research provides conflicting evidence regarding what the ideal structure for collective problem-solving actually is.

Research on diffusion provides one theory as to how to structure collaboration networks (Rogers, 2003; Strang & Soule, 1998). Fast and efficient communication networks have the potential to provide problem-solvers with the most recent information in their networks. A single breakthrough by one person will spread rapidly throughout the network, benefitting all. Studies on research and development networks have provided some evidence to support this view (Kim & Park, 2009).

However, theoretical research has shown that excessive communication might inhibit collective problem-solving in the long run (Fang et al., 2010; Lazer & Friedman, 2007). If groups collaborate at the expense of searching, their information diversity declines. They will begin to focus on a sub-set of the solutions, ignoring other potentially beneficial options, which can lead to groupthink, production blocking, and the common knowledge effect (Diehl & Stroebe, 1987, 1991; Janis, 1982; Stasser & Titus, 1985). Since diversity is crucial to group performance (Hong & Page, 2004; Page, 2007), slowing down collaboration might be better. In particular, decentralizing communication into separate sub-units, such as separating research and development teams from the central organization, has been shown to be beneficial (Benner & Tushman, 2003; O'Reilly & Tushman, 2004).

Experimental studies that have attempted to test these theoretical ideas have provided mixed support because they relied on small, highly stylized networks and used subjects from convenience samples to solve simple problems that do not capture the richness of experts solving real-world complex problems (Mason et al., 2008; Mason & Watts, 2012; Roberts & Goldstone, 2006; Wisdom, Song, & Goldstone, 2013). Existing theoretical results about the network structure of collective innovation suggest that managers and policymakers should take

the profoundly counter-intuitive action of reducing the efficiency of their collaboration networks. In a highly competitive market, such drastic institutional redesign requires clear evidence of the effect of collaboration networks on important problems. The existing empirical literature does not provide such clear guidelines. This paper examines the thesis that network efficiency has an inverse relationship to problem complexity using an online experiment where individuals must solve a complex problem, and it then applies these insights to management and organizational design to provide practical strategies to adjust collaborative efficiency to maximize collective innovation.

Due to the difficulty of separating the causal effects of network structure from individual characteristics (Shalizi & Thomas, 2011), this study uses a controlled experiment to manipulate a group's collaboration network, which identifies the causal effect of network structure on collective problem-solving (Centola, 2010, 2011; Centola & Baronchelli, 2015). This design allows me to test how network efficiency affects the ability of groups to solve complex problems.

To study this question, we examine the problem of statisticians and data scientists solving complex statistical problems. Participants are invited to a data science competition where they must solve a statistical problem by choosing which predictor variables among a large number of options should be included in a statistical model, and by collaborating with other competitors. In order to recruit the subjects who solve these complex problems in the real world, we built an online platform that participants could access through their web browser from any computer with an Internet connection. This design allows us to capture the real-world problem-solving behavior of individuals working on highly complex problems. This design builds on much research in computational social science that emphasizes precise causal control while also allowing generalizability in terms of the situation and the individuals performing the behavior (Centola, 2010; Lazer et al., 2009; Salganik et al., 2006; van de Rijt et al., 2014). This study is an attempt to find the people in the real world who normally solve such complex problems and test theoretical ideas about network efficiency.

To do this, we built upon a movement around crowd-sourced data science, which is a well-established implementation of collective intelligence that frequently delivers on the promise that the crowd can beat the experts (Aldhous, 2012). In one competition, crowdsourced users improved upon an insurance company's risk model by 270% ("Kaggle Winners Tapped As Data Analytics Consultants," n.d.), and one machine learning competition platform reports that in every competition with an industry benchmark, the crowd has produced a better model (Byrne, n.d.). The data science problem that participants face involves issues of variable selection and feature engineering, which are perhaps the most important aspects of data analysis, especially in the era of "big data" that are "wide," in the sense of having a large number of predictor variables (Kuhn & Johnson, 2013). As a result, selecting an appropriate subset of predictor variables is an increasingly challenging problem for data scientists and researchers, and detecting feature importance and generating new features are processes at the frontiers of data science because they cannot always be automated and require creative human input.

The goal of this study is to evaluate the thesis that network efficiency has an inverse relationship to problem complexity. As the problem complexity increases, reductions in network efficiency promote more effective problem-solving. This study examines these dynamics among human problem-solvers who are solving a realistic complex problem. The assumptions about agents embedded in current theoretical models that make these counter-intuitive predictions about network efficiency and collective performance might depart significantly from the behavior of humans, so testing how humans solve complex problems as part of a larger collaboration network is important. Additionally, due to the uncertainty arising from previous experimental studies, this study seeks to provide clear causal evidence about whether communication structure is an effective tool for managing information diversity and group performance in technological and scientific problems. I rely on a model of technological and scientific innovation rather than search across one- and two-dimensional problem spaces. As a result, this study seeks to generalize to a wider branch of complex, high-dimensional problems, such as biomedical research, technological innovation, and software design, as opposed to only a domain of problems involving search

across low-dimensional spaces. In order for significant resources to be invested in restructuring collaboration networks, it must first be made clear that the effects hold up consistently in a representative problem in science and technology and second that these effects are substantively large and consistent.

This paper is divided into three sections. The first section examines the **Basic Theory and Model** and presents the theoretical prediction that inefficient networks are better for complex problems. The second section examines results from an **Experimental Study** to test these ideas. And the third section expands the theoretical scope of these ideas by considering a **Revised Model** that examines ways managers can manipulate collaborative efficiency beyond network structure. The original theory considered only the possibility of rewiring network ties in an undirected, unweighted network. But I extend the theory by considering how managers might be able to reduce collaborative efficiency in other ways by considering weighted and dynamically evolving networks. This expanded theoretical model offers more practical ways that managers can adjust the efficiency of their collaboration networks in order to increase collective performance.

## **Basic Theory and Model**

### ***Exploration-Exploitation Tradeoff***

The complex problems studied here are members of a well-defined class of problems in which agents receive a *reward signal* based on their current state and attempt to adjust their *policy*, or behavior, in an attempt to maximize their utility (Sutton & Barto, 1998). In these *reinforcement learning* problems, individuals can repeatedly submit solutions and receive ongoing feedback from the environment informing them of their performance. These problems often require individuals to navigate through a large, high-dimensional space to find globally optimal solutions in the face of uncertain payoffs. Additionally, individuals typically operate in parallel, meaning all the agents are working on the same problem and their payoffs do not affect each other (Lazer &

Friedman, 2007). Examples include physicians discovering a correct diagnosis and prescription (Coleman et al., 1957), pharmaceutical engineers refining drugs to slow cancer growth, software developers moving through cycles of design and testing to refine the efficiency of a new program (Graham & Sichelman, 2010), state governments crafting public policy based on information about past effectiveness (Walker, 1969), financial analysts updating portfolio allocations based on previous performance (Pan et al., 2012), or data scientists revising their predictive models in a machine learning contest. In such problems, individuals face a tradeoff between exploiting already existing solutions in the network and exploring new options, which are risky and uncertain, but offer the possibility for breakthroughs that improve the entire group (Gupta et al., 2006; March, 1991; Sutton & Barto, 1998). Maintaining the correct balance between exploration and exploitation is crucial to group performance, and changing the network structure is a central theorized way to achieve an optimal balance (Lazer & Friedman, 2007; March, 1991).

### ***Efficiency and Performance***

Theoretical models of the effect of collaboration structure on group performance on complex problems have been developed using simulations from agent-based models (ABMs). For example, theoretical work has examined how individual decision rules for innovation versus imitation affect the group's performance (Rendell et al., 2010; Roberts & Goldstone, 2006). More relevant to this study, two ABMs have examined the importance of collaboration structure (Fang et al., 2010; Lazer & Friedman, 2007), and this study works from the model of Lazer and Friedman (L&F), detailed below (Lazer & Friedman, 2007).

### ***Model***

In the L&F model, agents search across a complex, high-dimension problem space. Each solution is a bit string (a sequence of 0's and 1's) that indicates a binary decision in each dimension. On each round, individuals behave according to the following rules. First, they check if any of their network neighbors have a better solution than their current solution. If there is a better one, they copy the max solution from their neighbors. If none of their neighbors have a better



solution, they change their current solution by randomly altering one element of their solution bit string. If that change is better than their current solution, they adopt it, and if it is no better, then they return to their previous solution. Agents behave deterministically in this fashion, but due to the randomness inherent in the search process, different collective outcomes can emerge for the identical starting positions.

L&F examine the effect of efficient versus inefficient networks on a group's performance when solving such a complex problem. Efficiency was measured as the average path length in each network.<sup>1</sup> The higher the average path length, the less efficient the network was. On each round, agents searched across a complex, high-dimensional problem space, either by exploring a new option or by copying a solution from one of their neighbors in the network. They find that on complex problems, efficient networks allowed groups to perform better in the short-term because information about initially promising solutions was disseminated rapidly. However, efficient networks performed worse in the long-run compared with inefficient networks because early copying prematurely restricted the efficient network's search to a smaller portion of the landscape. This reduction in information diversity made it less likely that they would find the globally optimal solution. These findings reversed for simple problems: efficient networks performed better in both the short- and long-term because there was no possibility of getting stuck on suboptimal solutions in a simple problem.

### ***Previous Results***

While theoretical models provide a clear series of results on the effect of network structure on collective problem-solving (Fang et al., 2010; Lazer & Friedman, 2007), experimental studies have provided conflicting evidence in support of the hypotheses. One initial study found supporting evidence for the theory, while a second larger follow-up study found disconfirming evidence (Mason et al., 2008; Mason & Watts, 2012; Wisdom et al., 2013). The conflicting

---

<sup>1</sup> Average path length is the average number of steps along the shortest paths for all possible pair of nodes.

experimental evidence can be resolved by designing an experiment that draws from the theoretical research. In particular, experimental studies need to use models of complex problems that capture the interdependencies among high-dimensional options in technological and scientific problems.

The theoretical studies provide agents with a complex problem that draws deeply from models of technological and scientific innovation. In contrast, experimental tests have tended to simplify the problem to searching across one- or two-dimensional landscapes. This abstraction and simplification does not capture the richness of scientific innovation. The theoretical work conceives of complex problems as requiring optimization along numerous dimensions (Kauffman, 1995; Lazer & Friedman, 2007; Levinthal, 1997; Levinthal & March, 1981; J. Miller & Page, 2009; Siggelkow & Levinthal, 2003). Complex problems have a high level of interdependency among their parts. One dimension may interact with others in surprising and unpredictable ways, producing non-monotonic relationships among the dimensions and making the direction of future innovation uncertain (Fleming & Sorenson, 2001; Kauffman, 1993; Kauffman et al., 2000; Kauffman & Macready, 1995).

Most real technological innovations have such complex interdependencies. For example, research on semiconductors has shown the crucial interdependence between temperature and the amount of impurity in silicon. Small changes in impurity levels have drastic effects on a semiconductor's resistance to electrical current at certain temperatures, such that at many levels the semiconductor fails. However, at other levels of interaction, the system provides valuable electrical properties (Millman, 1979).

These complex problems in science and technology are often conceptualized as movement across a "fitness landscape," which is a term originating in biology to refer to the distribution of fitness values for all combinations of a genotype (Wright, 1932). In technology, a fitness landscape refers to the performance of a solution on some dimension as a function of the solution, where each solution is a decision across numerous dimensions. I use the *NK* model to capture the interdependency among each component, following theoretical studies of complexity

(Kauffman, 1993, 1995; Lazer & Friedman, 2007; Valente, 2008). The primary benefit of the *NK* model as a representation of complex problems is that it offers a parameter to directly manipulate complexity. Additionally, its statistical properties are well understood (Kauffman, 1993). The model is often used to understand technological and scientific innovation (Fleming & Sorenson, 2001; Kauffman et al., 2000; Kauffman & Macready, 1995; Lazer & Friedman, 2007; Levinthal, 1997; March, 1991).

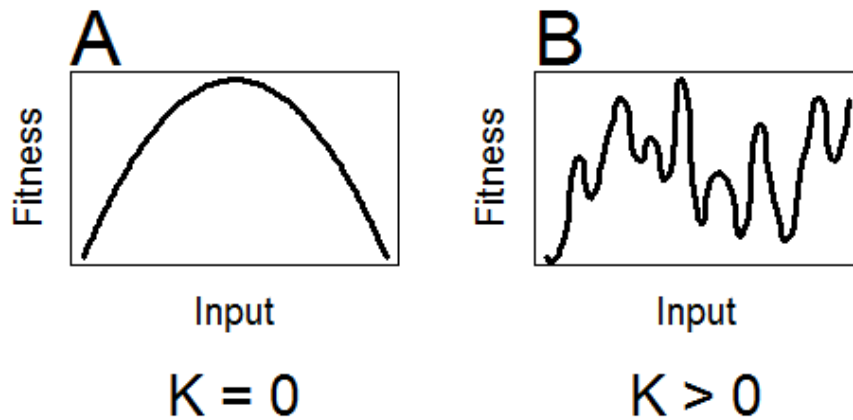
Two parameters control the *NK* model. *N* indicates the dimensionality of the problem, where a problem space is modeled as a bit string of length *N* with a 1 indicating that a component has been activated, and a 0 indicating a component's deactivation.<sup>2</sup> From the perspective of the agent, the goal is to determine which combination of 0's and 1's will make the best solution. Calculating the "fitness" of a solution depends upon the second parameter, *K*, which determines the level of interdependence between each dimension. When  $K = 0$ , each bit contributes to the fitness independently. But when  $K > 0$ , the components interact, and the fitness contribution of each element depends on *K* other components.

While it is impossible to plot high-dimensional problem spaces, Figure 1 shows a one-dimensional stylization of the difference between a simple problem space when  $K = 0$  (panel A) and a complex one when  $K > 0$  (panel B). Simple problems have an easily identifiable global maximum. When agents adjust their input along the x-axis by continually searching across and updating their position along their local environment, they will eventually reach the peak. In contrast, complex landscapes have many local maxima. Since individuals and organizations typically search incrementally (Cohen & Levinthal, 1990; March, 1991; March & Simon, 1958; Nelson & Winter, 1982; Stuart & Podolny, 1996), they tend to get caught on local maxima in complex landscapes. Complex problems are thus rugged, multi-peaked fitness landscapes in

---

<sup>2</sup> A third parameter, *A*, allows each dimension to have more than two options, and numerous models have expanded upon this idea (Li et al., 2006; Valente, 2008), but the central statistical properties of *NK* landscapes are invariant to these changes (Kauffman, 1993).

numerous dimensions, where “rugged” refers to the number of local maxima where agents can get stuck.



**Fig. 1. Simple and complex landscapes in one dimension.**

In contrast to the rich models of complexity used in the theoretical studies, the experiments that have attempted to test this theory have tended to simplify the task to search across one- (Mason et al., 2008) and two-dimensional spaces (Mason & Watts, 2012). For example, in Mason and Watts (M&W) (Mason & Watts, 2012), subjects played an online game called “Wildcat Wells,” where they searched across a desert landscape to find the best locations to drill an oil well. The underlying landscape was a hilly space with multiple peaks, or locally optimal drilling locations. Such spaces allow generalization to an important domain of problems, including search and rescue operations, but they do not capture the multi-dimensionality of technological and scientific innovation.

Such simplification is concerning because the theory stresses an interaction between problem complexity and network structure. As problem complexity increases, *reductions* in network efficiency will promote more effective problem solving. But when problems are very simple, *increases* in network efficiency will lead to the best results (Lazer & Friedman, 2007). These theoretical ideas help explain why M&W found that efficient networks were always better on a two-dimensional problem space—a finding that runs contrary to the theoretical predictions.

But it is doubtful that the findings from M&W generalize to technological and scientific search on more complex problems. Having the proper model of a complex problem is thus crucial. The primary benefit of the *NK* model as a representation of complex problems is that it offers a parameter to directly manipulate complexity so that the modeler can rapidly tune a multi-dimensional problem space between simple and complex (Kauffman, 1993). The model of complex problems used in this study (the *NK* Model) allows me to more precisely test hypotheses about the interaction between problem difficulty and network structure than previous experiments have allowed.

Additionally, the individuals in M&W likely did not explore incrementally, as individuals were free to jump throughout the problem space to very distant regions, and most of the users in the study appeared to explore in a random, global manner. In contrast, most research on scientific, technological, and organizational innovation stresses that groups explore myopically by making incremental changes to existing practices and methods (Cohen & Levinthal, 1990; March, 1991; March & Simon, 1958; Nelson & Winter, 1982; Stuart & Podolny, 1996). Such incremental movements are a rational response when exploring is costly or when organizations are in a highly competitive environment where the risk of moving far away from an existing solution could result in institutional failure. The ability and incentives to explore non-incrementally meant that getting caught on locally optimal solutions was unlikely because subjects could simply look at other solutions and search close to them.

Further, M&W provided the visual position of other players and then allowed users to either directly copy that location or explore. However, users could use information about the position of others when forming their decision to explore. Users could explore *close to* other users, meaning that they were receiving a social signal and combining it with their exploration. For example, participants rarely directly copied each other, but instead explored similar locations based on social information. This popularity effect results in a mixture between pure copying and pure exploration that is outside of the theoretical model. It likely reduced the amount of pure copying that players used in the experiment. Such a strong reduction in direct, measured copying

would weaken the effects of the collaboration network manipulations and would make it more likely for the efficient network to perform well. I have conducted simulations that show that reductions in the amount of actual copying behavior could eliminate the differences between conditions.

A more recent paper explores the effect of network structure on the ability of groups to solve a whodunit problem, such as playing the game of Clue (Shore et al., 2015). The problem was meant to have a higher degree of verisimilitude and involved gathering information about a terrorist plot and formulating theories based on how the plot occurred. The study found supportive evidence for the theorized inverse relationship between network efficiency and collective performance. This problem offers substantial improvements upon search across a two-dimensional map. The multiple dimensions of the problem (Who would carry out the attack? What would be the target? Where would the attack take place? And when would it occur?) generalize to a broader space of realistic problems. Additionally, they produce dynamics where individuals are resistant to random global search and instead search incrementally, which is more realistic. However, this problem departs significantly from the theoretical *NK* model, as it offers no way to directly manipulate the problem complexity and instead only produces a large problem space without any guarantees about complexity. In fact, complexity in this area simply means the size of the problem space, which is a product of the size of the four dimensions (Who? What? Where? When?). This paper supports the theorized inverse relationship between network efficiency and group performance, but it is unclear whether whodunit problems can map onto an important domain of problems involving technological and scientific innovations.

Taken together, the existing empirical results provide conflicting evidence about whether managers and policymakers should reduce the efficiency of their collaboration networks in order to increase collective performance. We currently do not know if taking such profound actions would be beneficial for structuring scientific or corporate collaboration networks. Before funding agencies incentivize different collaboration structures or corporate research divisions restructure

their organizations, it is important to know whether these effects exist and how large the effect sizes are when groups solve complex problems.

## **Experimental Design**

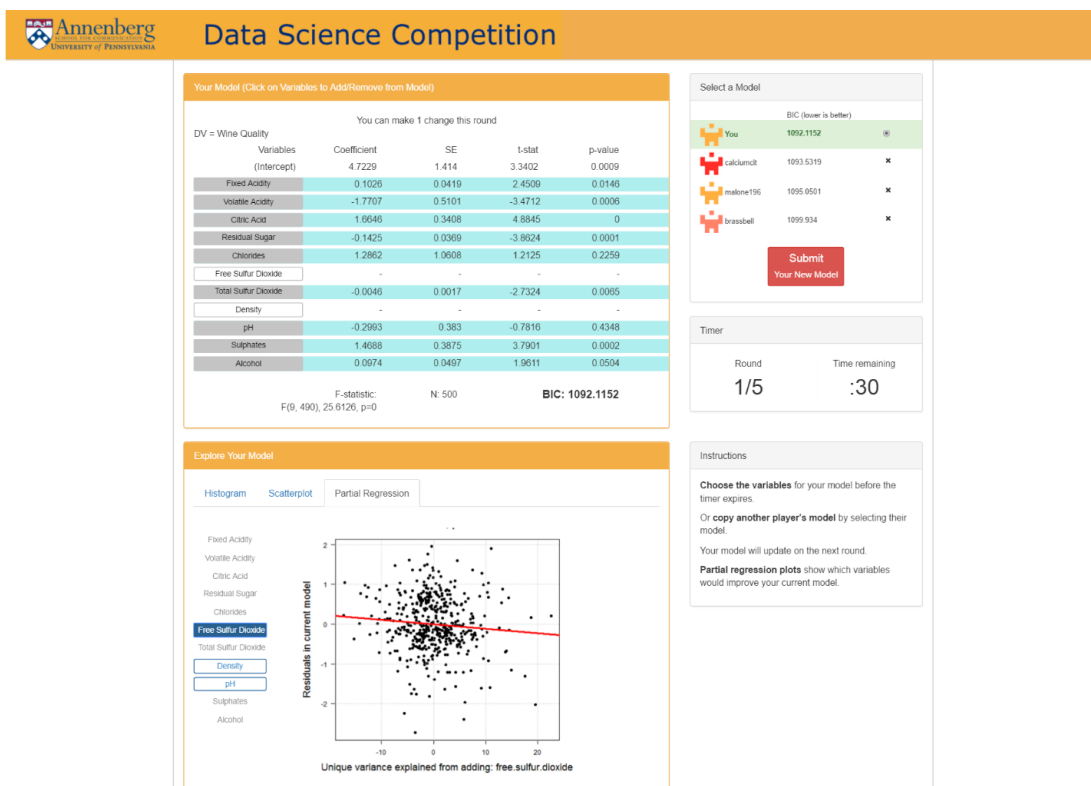
### ***Task***

To causally identify the relationship between network efficiency and collective performance, I developed a complex task that had several important properties: (1) maximum problem realism, which means that the task was a particular real-world complex problem that shared similar features to other complex problems; (2) maximum subject realism, which means that the subjects we recruited should be the people who actually work on these complex problems in the real world, not just convenience samples; and (3) maximum accuracy of data collection, which means that we could capture as much behavioral data as possible without relying on any self-report measures.

To satisfy these three criteria, the experiment consisted of a Data Science Competition where experts in statistics were recruited from the World Wide Web to solve a complex problem. Statistics and data science competitions have become increasingly popular online, where individuals compete to build the best statistical forecast model that makes the most accurate predictions (criteria 1). For example, one of the data problems in our study required subjects to predict sales volumes at a retailer as a function of economic variables. I built an online platform that allowed us to recruit interested and motivated individuals online who have quantitative skills (criteria 2). The structure of the competition was round-based, which allowed us to unobtrusively measure subjects' decisions to either explore new solutions or exploit existing ones over 15 distinct rounds in the competition (criteria 3).

Subjects interacted with the competition using a custom-built platform that displayed their model performance on each round and allowed them to revise their solution (Fig. 2). The competition lasted for 15 rounds. In each round, individuals decided either to *explore* or *exploit*.

Exploration meant changing their solution by adding or removing a single variable from their statistical model. To search incrementally for a better solution, individuals could select a single variable to add or remove from their current model (Cohen & Levinthal, 1990; March, 1991; March & Simon, 1958; Nelson & Winter, 1982; Stuart & Podolny, 1996). Exploiting meant coping an existing solution from a network neighbor. Additionally, individuals could maintain their current solution. To copy a better solution, individuals saw the performance scores of their best network neighbors, and could choose to adopt a better solution.



**Fig. 2. Screenshot of the experimental interface.** To explore their model, individuals could interact with their current solution by changing a variable in the top left panel. To exploit neighboring solutions, participants could select a better model from their network neighbor in the top right panel. When individuals had a made a selection for that round, they pressed the “Submit” button, and then proceeded to the next round. The bottom panel contained plotting information about their current model solution.



Once a participant submitted a decision for the round, they received immediate feedback on the performance of their new solution. If the solution was better, they adopted it, and proceeded to the next round, and if the solution was worse, they were returned to their previous solution. Each round lasted for 1 minute. Individuals never knew if they had the best possible solution, and they were incentivized to find better solutions each round, and received financial rewards based on the quality of their final solution.

Models were scored based on the Bayesian Information Criterion (BIC). The BIC is a good asymptotic measure of out-of-sample model performance and performs similarly to cross-validation predictive accuracy (Shao, 1997), so it captures the ability of the model to predict unseen data. The BIC rewards constructing sparse models that explain sufficient variance in the dependent variable.

The forecasting problems were sufficiently complex, and involved subjects solving a high-dimensional combinatorial optimization problem. Complex problems have high interdependency among the components of the solutions such that changing one dimension affects the fitness contribution from another dimension (Kauffman, 1993; Kauffman et al., 2000; Kauffman & Macready, 1995). Such synergies among the components produce many local maxima in the problem space so that incremental, local search can miss the global maximum. Performance metrics such as BIC penalize variables that do not provide additional information. In a simple landscape, where variables do not interact, each predictor variable contributes statistically independent information that improves the quality of the model. When a variable explains no variance in the predictor variable, its contribution is clear because the BIC will decline. In contrast, in a complex landscape, there are correlations among the variables, which is common in real world data sets. Variables often contain redundant information that is already captured by another variables. As a result, adding redundant information to the model will result in a worse BIC score and worse out-of-sample performance because the model is effectively being fit to noise that is idiosyncratic to the training data set.

To ensure that the data science problems in this study had this complex structure, I adopted existing data sets and increased the correlations among the variables. The procedure works by holding fixed the amount of variation in the dependent variable that is explained by all the predictor variables, but then shuffling that predictive variation among the predictor variables. By altering the correlation among the  $X$ 's, we can shift from a simple to a complex problem. In a complex problem, the contribution to the fitness of one variable depends on whether another variable is already included in the model or not. This interdependency among the components produces a complex fitness landscape.

To ensure that the results did not depend upon a single problem, I used three data sets to generate complex data science problems. The data problems contained numerous locally optimal but globally suboptimal solutions, ranging from 9 to 16 local optima, and the number of possible solutions ranged from 2,048 to 16,384.

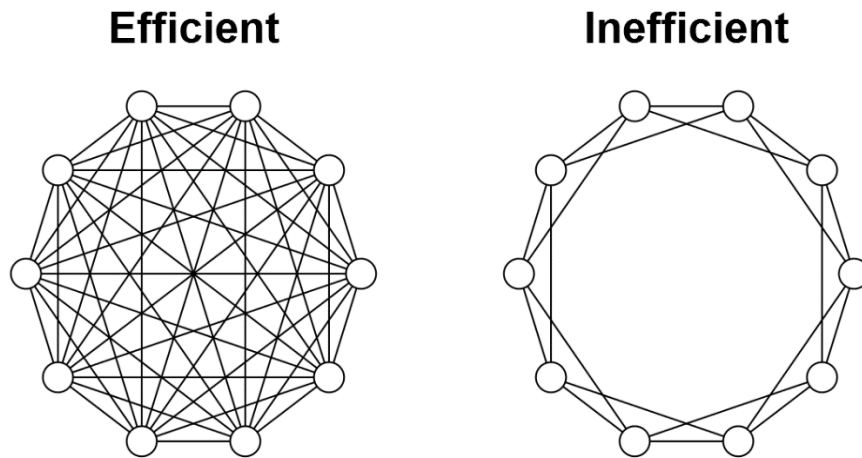
### ***Recruitment***

In order to recruit subject who were specialized in statistics and quantitative methods, I contacted individuals in quantitative departments of several universities, and advertised in several online forums that are frequented by individuals with interest in statistics and data science. Individuals had taken a mean of 3 (sd=3.6) upper-level statistics courses.

### ***Experimental Treatments***

The structure of the collaboration network determined which solutions an individual could see and choose to copy. Subjects were randomly assigned to one of two collaboration networks (Fig. 3). In the efficient network, the structure was a fully connected network, which has the minimum possible average path length ( $L = 1$ ). In the inefficient network, the structure was a ring lattice, where each individual was connected to their nearest four neighbors, which has a larger average path length ( $L = 1.67$  for  $N = 10$ , and  $L = 2.89$  for  $N = 20$ ). In total, 14 independent networks were run, with seven matched pair trials. In the first six trials, populations of size  $N = 10$  were used in each condition, and in the final trial, the population sizes were increased to  $N = 20$ .

In total, there were 160 unique participants in the study. To make each group in a trial as similar as possible, each participant was given a random solution that was matched with an identical initial solution in the other network within a given trial.



**Fig. 3. Network structures in the experiment.** Each experimental trial consisted of two networks, one efficient ( $PL = 1$ ), and one inefficient ( $PL=1.67$  or  $2.89$ ). In each of the seven trials, subjects were randomly assigned to one network condition, and then randomly assigned to a single node in the network. On the initial round of each trial, subjects received a random starting solution. On subsequent rounds, subjects saw the performance of their immediate network contacts and could copy these solutions. In a single trial, random initial solutions were matched across conditions, and faced the same data science problem.

Participants in the study were shown an identical user interface in both experimental conditions. Features of the social network, such as the average path length and the size of the population, were unobservable to participants. More generally, every aspect of the participants' experience was equivalent across experimental conditions. The only difference was the structure of the social networks. Thus, any differences in solution quality may be attributed to the effects of network efficiency on the process of collective innovation.

## **Measures**

### ***Diversity***

Group-level solution diversity was measured in two ways. The first considered the number of unique solutions. Several different units of time were considered, including each round as well as cumulatively across the entire experiment. The second measure attempted to quantify the magnitude of diversity beyond a simple count of the number of unique solutions. The Hamming distance is defined as the number of differences in the elements of two strings. We took the average Hamming distance of a group's solutions, comparing each individual to every other individual and then averaging across the entire group.

### ***Diffusion and Convergence Speed***

Diffusion speed was measured as the fraction of individuals who adopted the best solution on the rounds following its discovery by a member of that group. The convergence speed measured the number of rounds until an individual adopted the solution that was best in that network.

### ***Rate of Exploration***

On each round, individuals could either choose to explore by submitting a new solution, copy by selecting another participant's solution, or maintain their current solution. This mutually exclusive division between exploration and exploitation reflects the way that these problems have been conceptualized in theoretical and empirical research (Lazer & Friedman, 2007; March, 1991; Mason & Watts, 2012; Shore et al., 2015). Explorations were registered within the database when individuals selected to add or remove a model component and pressed the submit button. The rate of exploration simply measured the proportion of all submissions that involved a decision to revise an existing solution.

## **Performance**

Performance was measured in two ways to capture different utility functions that groups might have. The maximum performance captured the best solution that was found in each group. This metric is of interested in group-problem solving situations where an organization will only implement a single solution, such as among a team of engineers designing a new product. In contrast, the average performance took the arithmetic mean across all the individual solutions within a group. This metric is of interest when the utility function of a group depends upon all members, such as among a team of salespeople who have different strategies for pursuing leads and who ultimately generate the total sales revenue for a company.

The performance of each model was measured in terms of its Bayesian Information Criterion (BIC), which is also known as the Schwarz Criterion (Schwarz, 1978). The BIC is a measure of a model's out of sample performance on a new data set that the model has not seen before. The BIC is a function both of the likelihood function and a regularization term that penalizes the addition of more parameters. The BIC is defined as:

$$BIC = -2 \bullet \ln \hat{L} + k \bullet \ln(n)$$

where  $\hat{L}$  is the maximum of the likelihood function of the model,  $k$  is the number of free parameters to be estimated, and  $n$  is the number of observations in the data set. In the case of a linear regression used in the experiment,  $k$  is the number of regressors including the intercept in the model. In short, the BIC metric rewards models that fit the data closely, while penalizing models that are overly complex and rely on many parameters.

To create a measure of group performance, I rescaled the BIC metric onto the range [0,1]. Since a lower BIC indicates that a model is a better fit to the data, I created the following measure of group performance by transforming the BIC as follows:

$$Best_i = \frac{mean_i(BIC_{t=0}) - \max_i(BIC_t)}{mean_i(BIC_{t=0}) - \max^*(BIC)},$$

where the numerator is the difference between group  $i$ 's average starting solution on the initial round ( $t=0$ ) and the maximum solution on the current round  $t$ , and the denominator is the difference between the group's average starting solution and the best possible solution. To capture average group performance, I use the following formula:

$$Average_i = \frac{mean_i(BIC_{t=0}) - mean_i(BIC_t)}{mean_i(BIC_{t=0}) - \max^*(BIC)}$$

where the numerator is the difference between the group's average solution at  $t=0$  and the group's average at time  $t$ , and the denominator is the same. Both metrics range from 0 (the group's initial solution) to 1 (the best possible solution), and indicate how much groups have improved from their initial starting solution. The average initial starting solution was fixed between conditions within a given trial because I matched the same starting solutions between conditions, and the best possible solution was fixed for each data set. As a result, this metric is directly comparable between conditions only if individuals started from the same solutions in both conditions and they used the same data problem.

## **Statistical Analysis**

### ***Analysis***

The analysis is conducted at the group level because the network structure treatment is applied to the group as a whole. As a result, an individual-level analysis would violate assumptions about independence. At the individual-level, I had 160 unique subjects, but at the group level, I had only 14 groups. To ensure that the analysis was conservative with such a small sample size of groups, I used the nonparametric Wilcoxon signed rank test to make paired comparisons between trials, and the nonparametric Wilcoxon rank sum test to make comparisons between the distributions of both conditions. These tests were used to evaluate the hypotheses about solution diversity, speed of diffusion, rates of exploration, and performance. Additionally, a

Cox proportional hazards model was constructed to test how the time to convergence differed between conditions.

### ***Attrition***

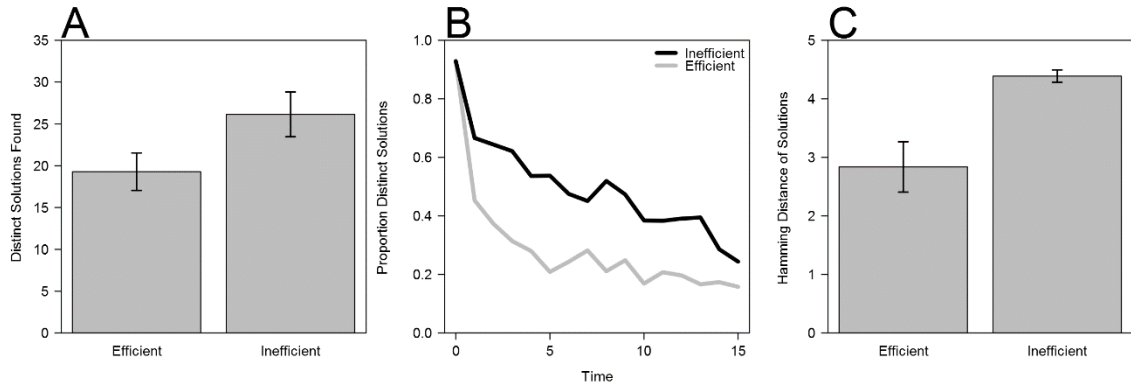
There was a low attrition rate for the entire experiment: 14% of subjects did not complete all rounds of the competition. There was no significant difference in attrition rates between conditions ( $P = 0.75$ , Wilcoxon signed-rank test). Data was used from individuals until they left the study, or finished the final round.

## **Experimental Results**

Our results provide support for the proposition that slow collaboration as manipulated through inefficiency in the network structure promotes collective diversity, which improves collective performance.

### ***Solution Diversity***

Inefficient networks had more group-level diversity of solutions than efficient networks (Fig. 4). Inefficient networks explored a larger portion of the solution space, on average exploring 36% more distinct solutions throughout the experiment compared to efficient networks ( $P = 0.03$ , Wilcoxon signed rank test) (Fig. 4A). On every round, inefficient networks had more unique solutions (Fig. 4B), averaging 72% more distinct solutions at each time point in inefficient networks compared to efficient networks ( $P < 0.001$ , Wilcoxon rank sum test). On average, inefficient networks were 55% more diverse in terms of the Hamming distance of the solutions offered as compared to efficient networks ( $p = 0.03$  Wilcoxon signed rank test) (Fig. 4C). Individuals in inefficient networks had solutions that were on average 4.4 steps away from the other members of the group, as compared to only 2.8 changes in efficient networks.

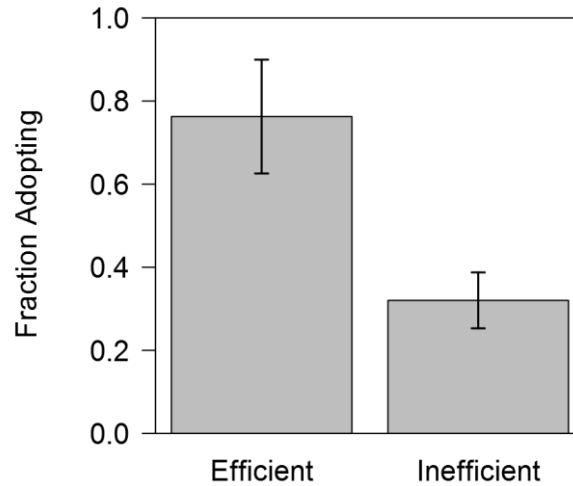


**Fig. 4. Diversity of solutions.** Panel A: The average number of distinct solutions in efficient and inefficient networks across all trials. Panel B: Average number of distinct solutions on each round. Panel C: Average Hamming distance of all solutions in efficient and inefficient networks. Error bars show 1 standard error of the mean.

### ***Speed of Diffusion***

Networks with higher path lengths were less efficient in terms of the speed of diffusing good solutions. When an individual found the solution that would be the best in the efficient network, an average of 76% of individuals adopted it on the following round (Fig. 5). In contrast, only an average of 32% of individuals adopted the best solution on the next round after its discovery in an inefficient network, which was significantly lower than efficient networks ( $P < 0.01$ , Wilcoxon rank sum test).

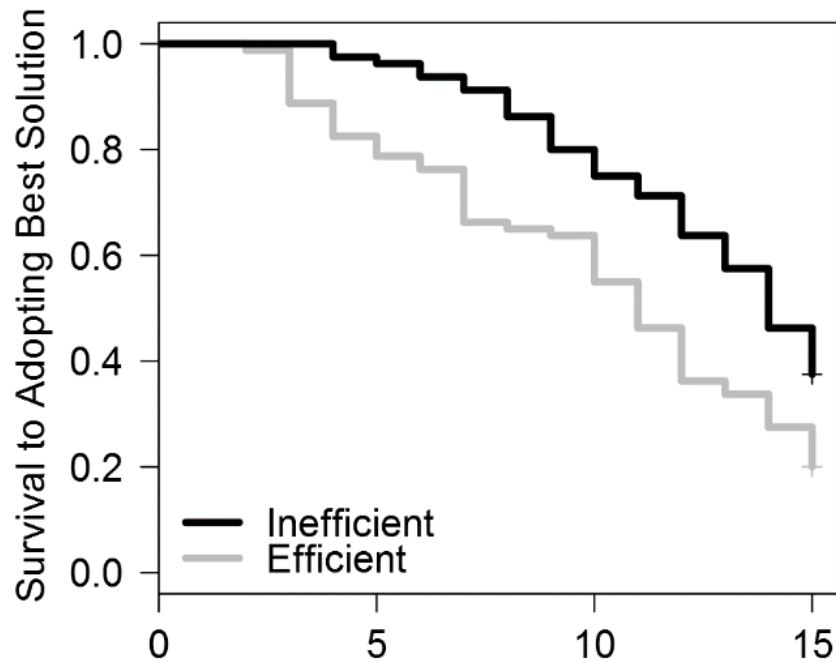




**Fig. 5. Rate of solution diffusion.** Mean proportion of individuals adopting the top solution on the round after it was found ( $t+1$ ). Error bars show two standard errors of the mean.

### ***Speed of Convergence***

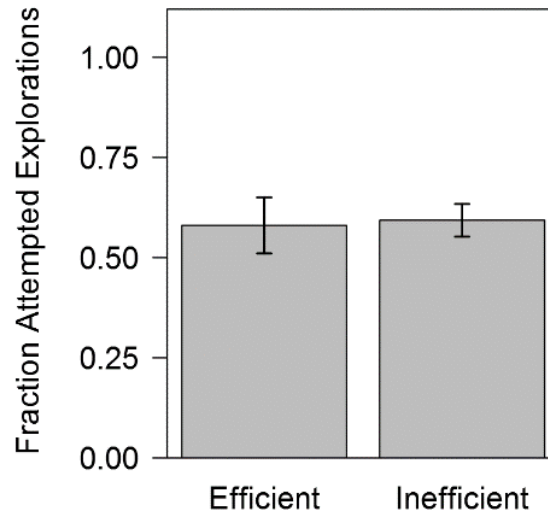
Efficient networks were more likely to converge prematurely on an early solution than inefficient networks. Figure 6 shows the Kaplan-Meier plot of the survival time until individuals adopted the solution that was best in their network. On average across all trials, the hazard of adopting the best solution was 83% higher in efficient networks as compared to inefficient networks ( $P < 0.01$ , Cox proportional hazards model), which indicates that individuals were quicker to adopt in efficient networks.



**Fig. 6. Kaplan-Meier survival plot of the time until adoption of the top solution.** Survival indicates an individual has not adopted the best solution. Efficient networks adopted their best solution earlier and more quickly than inefficient networks.

### ***Rate of Exploration***

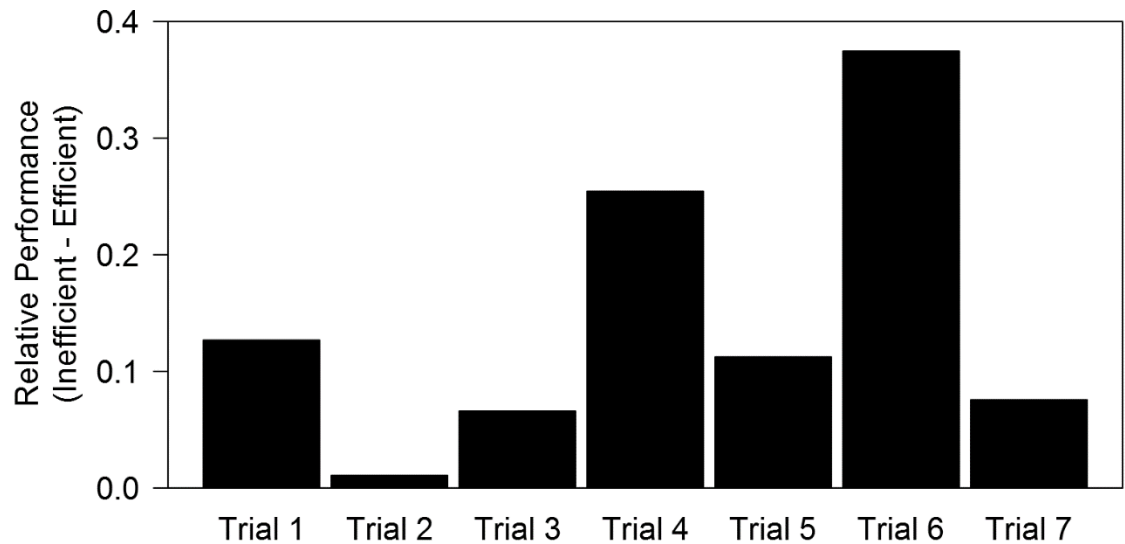
In contrast to the significant differences in diffusion rates and speed of convergence, there were no difference in decisions to explore new solutions between conditions. Network structure did not affect the rate of exploration among subjects ( $P = 0.93$ , Wilcoxon signed rank test) (Fig. 7). This finding differs from previous studies that have found significant increases in exploration when teams were embedded in more efficient networks (Mason & Watts, 2012).



**Fig. 7. Attempts to explore new solutions.** The average fraction of plays that were attempts to explore a new solution are plotted for each network across all trials. Error bars are two standard errors of the mean.

### ***Performance***

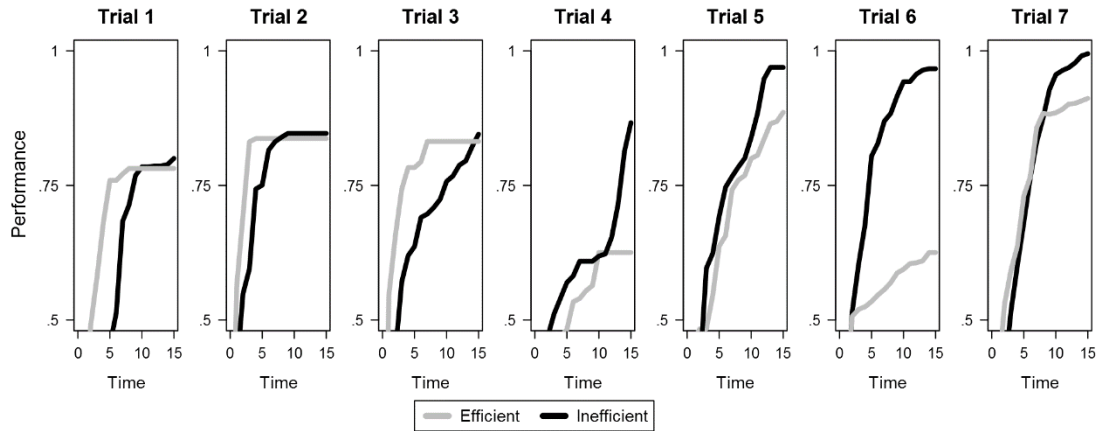
Inefficient networks led to better performance at the collective level. Inefficient networks discovered better top solutions than efficient networks in every trial ( $P = 0.02$ , Wilcoxon signed-rank test) (Fig. 8). On average, inefficient networks found solutions that were 21% better than those discovered by efficient networks. Additionally, in three out of the seven trials, inefficient networks discovered the best possible solution, which was verified after an exhaustive search of the entire solution space.



**Fig. 8. Difference in performance between inefficient and efficient networks.** The difference between the maximum solution found by an individual in inefficient and efficient networks is plotted across all seven trials. Inefficient networks found better top solutions than efficient networks in all trials.

### ***Average Group Performance***

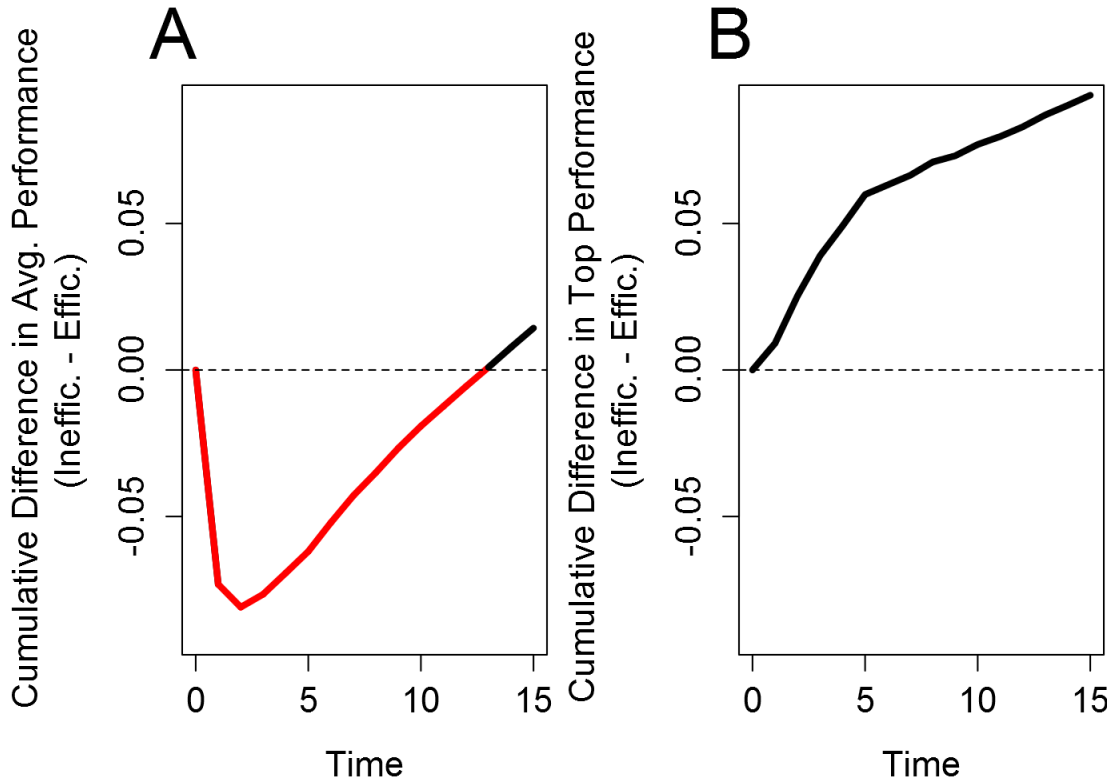
The results for the average group performance confirm the theoretical prediction that in the short-term efficient networks will diffuse better solutions and perform better, but in the long-term, inefficient networks will find better solutions and improve the group's average (Fig. 9). After the first round of the competition, inefficient networks had average solutions that were 30% worse than efficient networks ( $P < 0.05$ , Wilcoxon signed-rank). However, by the final round, every inefficient network had a better average solution than its efficient network pair. On average, inefficient networks generated mean solutions that were 17% higher than efficient networks by the end of the study ( $P = 0.02$ , Wilcoxon signed-rank).



**Fig. 9. Average performance in efficient and inefficient networks across time.** Lines represent the average performance across all seven trials for efficient (light) and inefficient (dark) networks. Average performance was initially higher within every efficient networks as compared to inefficient networks. However, by the study's conclusion inefficient networks had a better average performance than efficient networks within each trial.

### ***Cumulative Performance of the Average***

Investing in inefficient networks requires a tradeoff between the short-term, where average solutions are worse than efficient networks, and the long-term, where eventually better solutions are found and diffused to the entire population. To understand these tradeoffs, the cumulative performance of each group's performance across all trials is plotted in Figure 10. In short time frames, inefficient networks have a large cost in comparison to efficient networks (Fig. 10A). However, after round 13, inefficient networks have recovered this cost, and cumulatively outperform efficient networks, a trend that will continue indefinitely because the efficient networks have gotten stuck on a suboptimal solution. More striking is the results for top performance, where inefficient networks consistently offer better solutions for all time (Fig. 10B).



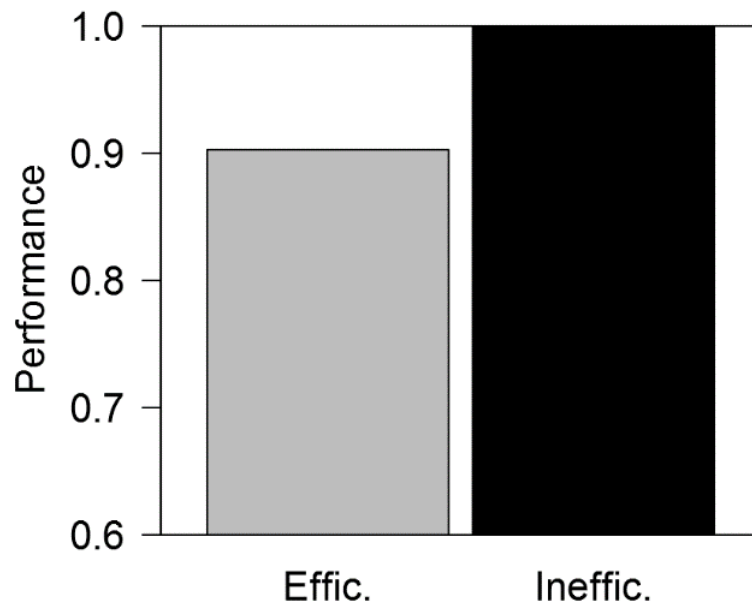
**Fig. 10. Payoff to investing in inefficient networks relative to efficient networks.** Panel A: Difference in cumulative average performance over time (inefficient minus efficient). Red line indicate times where it is better to invest in efficient networks, whereas black line indicates where the cumulative return is higher to inefficient networks. Panel B: Difference in top performance over time (inefficient minus efficient).

### ***Robustness to Non-Local Exploration***

To examine the sensitivity of the results to design choices, I conducted a robustness experiment with  $N=10$  individuals in an inefficient and efficient network. The main experiment allowed individuals to make a single revision each round (i.e., incremental search). This design captured realistic search processes by individuals and organizations in high-risk situations, where there are strong incentives to add incrementally to a solution (Cohen & Levinthal, 1990; March, 1991; March & Simon, 1958; Nelson & Winter, 1982; Stuart & Podolny, 1996). To examine if the results

are robust to this design constraint, I permitted individuals to make as many changes to their model on each round before submitting their new solution and receiving feedback. Allowing this non-incremental search might affect the results by preventing efficient networks from getting stuck on suboptimal solutions because an individual could get lucky and find another peak even though the group had converged on an initial solution.

I ran a single efficient network and allowed individual to make non-incremental searches, and compared the performance of this efficient network to the inefficient networks that were restricted to only incremental search. I expect that allowing non-incremental search will increase performance. The results show that allowing individuals to search non-incrementally did not substantively change the results in any way. The inefficient network still performed 11% better than the efficient network, and there were very little differences between the efficient network with incremental search and prior trials that allowed only non-incremental search.



**Fig. 11. Performance differences between conditions was robust to allowing non-incremental search.**

At the individual level, participants did attempt to explore more widely when given the option to conduct non-incremental search. Across all attempts to explore, 38% of attempts involved non-incremental search. However, a majority of these attempts were unsuccessful, and individuals were more successful when they explored incrementally. When incremental exploration was used, individuals successfully found a better solution 19.4% of the time, in contrast to a success rate of only 9.8% for attempts to change more than one element of their solution. This result suggests that incremental exploration is both preferred by individuals, and also represents a more informed choice because the quality of the solution will likely be more similar to the current solution.

### ***Discussion of Experimental Results***

Slowing down the rate of collaboration increased the performance of groups of statisticians and data scientists when they solved a complex problems. Inefficient networks with higher average path length restricted the diffusion of solutions, which led to more diversity. In contrast to previous empirical results, I find no evidence that efficient networks provide a benefit to collective performance (Mason & Watts, 2012). Instead, the results agree with previous theoretical (Lazer & Friedman, 2007) and empirical research (Mason et al., 2008) showing that slowing down collaboration can prevent groupthink and premature adoption of suboptimal solutions.

While previous research found higher rates of exploration in efficient networks as compared to inefficient networks (Mason & Watts, 2012), we find no differences between conditions in the rate of exploration. This difference can be explained because their design allowed for partial copying where individuals could explore close to the previous solutions that others had offered. Such social influence and popularity effects can profoundly alter the collective dynamics when searching across 2-dimensional landscapes, which can create particle swarm dynamics (Eberhart & Kennedy, 1995; Eberhart, Shi, & Kennedy, 2001). In contrast, this study assumed that individuals were either exploring or copying, and that these were categorically



distinct from one another, which is in line with previous research on the exploration–exploitation tradeoff (Lazer & Friedman, 2007; Shore et al., 2015). By not confounding the results with social influence and popularity, this study showed no difference in individual exploration across conditions. This is because the original theory actually predicts no difference in individual exploration on average across conditions for rates of exploration.

### **Revised Model**

The experiment was able to confirm the theoretical model, which posits that increasing inefficiency in terms of network average path length increases a group’s solution quality when they are solving a complex problem. This has important implications for how managers establish ties among the members of an organization, such as the connections among members of a research and development arm of an organization (Kim & Park, 2009). However, there are many situations where managers cannot add or remove the ties in a network, such as when individuals have the ability to choose their own ties. Additionally, even if managers have the ability to rewire a network to maintain a desired collaborative inefficiency, these rewired networks can be fragile. The addition of a single long tie that spans the network can dramatically increase the collaborative efficiency of the network and undermine attempts to maintain inefficient collaboration as determined by high average path lengths (Watts & Strogatz, 1998). As a result, even though the effects in this study are substantively important, they may be difficult to practically implement as currently theorized.

The theoretical literature has developed the idea that network slowness can be manipulated by the average path length of the network. Research has focused on undirected, unweighted graphs. Here, I develop the idea that “collaborative slowness” can be manipulated in other ways, namely weighted path length (e.g., communication costs) and temporal networks that delay communication and prevent premature convergence.

The central idea is that longer average path lengths slow down collaboration. The insight that inefficiency or slowness might promote collective problem-solving immediately suggests two

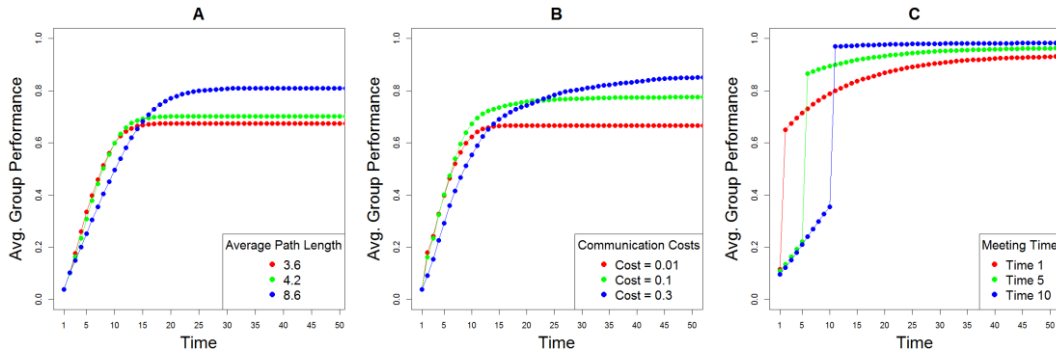
alternative ways to improve collective problem-solving that draw on recent work in network science. First, communication costs, or the costs to form a tie or use an existing tie, can slow down collaboration (Goyal & Vega-Redondo, 2005; Jackson, 2005; Slikker & van den Nouweland, 2000). These communication costs can be conceptualized as placing weights on each tie so that individuals pay different attention to each of their neighbors based on the strength of the network tie. By adjusting the weights on these ties, it is possible for a manager to tune the weighted average path length of a network.

Second temporal networks can manipulate the speed and frequency of collaboration. Temporal or dynamic networks involve changes in the network structure over time (Gloor, 2005; Holme & Saramäki, 2012; Juszczyszyn, Musial, Kazienko, & Gabrys, 2012). Networks that dynamically rewire over time, perhaps from influence by managers, offer the possibility of fine-tuning the collaboration structure to account for the unfolding process of collective exploration over time. In initial discovery phases, groups might collaborate less frequently to increase the diversity of solutions. However, later in the process, the network structure might be made more efficient so that the best solutions are rapidly diffused to the population.

In this section, I explore theoretically how these two manipulations of collective efficiency affect the dynamics of group problem-solving.

### ***Simulation Results***

To examine the effect of these additional manipulations of collaborative slowness, I conduct a series of agent-based models using the same basic model as L&F (2007). The simulations then examine how variations in collaboration structure affect the group's performance. Figure 12, Panel A replicates the basic finding from L&F: making a network more inefficient by increasing its path length leads to better collective performance in the long run. The most efficient network (red) performs worse in the long-run in comparison to the least efficient network (blue).



**Fig. 12. The effects of network efficiency on collective performance.** Panel A replicates L&F by considering the probability of rewiring ring lattices using the small worlds model (Watts & Strogatz, 1998). Probabilities include  $p = 0.01, 0.1,$  and  $0.3$ , which are equivalent to average path lengths of 8.6, 4.2, and 3.6. As path length increases, the average performance increases. Panel B examines communication costs for copying a neighbor's solution in a fully connected network. As costs increase, the network performance increases. Panel C examines the timing of meetings in a fully connected network. As the meeting time is delayed from round 1, to 5, and to 10, performance increases. Within a single replication of each simulation, agents begin at the same locations, thus allowing for within-subjects confidence intervals, which are very small in the figures. All panels use 1,000 simulations with 100 agents in each network.

To examine how performance is affected by communication costs, I revise the basic model in the following way. Instead of always copying a neighbor's solution if it is better than their own, agents must pay a tax for copying another solution. Agents will only copy if the neighbor's solution minus a tax is better than their current solution. This manipulation can also be understood as imposing switching costs for adopting a radically new solution. In contrast to the intuition that communication costs might harm a collective's performance, the simulations in Figure 12, Panel B show that as communication costs increase, groups perform better in the long run. Networks with the most efficient communication (red) perform worse than networks with higher communication costs (blue). The intuition is that these costs will encourage agents to

search more widely, but when a far better solution is found, it will still be able to diffuse across the network.

The simulations for temporal networks alter the baseline model by considering the effect of meeting time on performance (Figure 12, Panel C). Meetings are defined as a fully connected network. On times where there are no meetings, agents search in isolation. The effect of meetings is consistently large for all times, as it improves average performance. However, the largest absolute performance comes from delaying the meeting time (blue) as opposed to meeting early in the design process (red). Each additional round that a meeting is delayed increases the likelihood that a superior solution is found.

### ***Discussion of Simulation Results***

Collectively, these theoretical models expand the scope of the theory involving collective problem solving by examining alternative mechanisms that slow down a group's exploitation of known solutions. In addition to the theoretical results, these findings have important policy benefits. For example, maintaining a network structure with a high path length might be difficult for a manager or policymaker, especially with decentralized agents. When a network with long path length, such as a ring lattice, adds only a few random ties, its path length rapidly diminishes. In other words, if even a few individuals within an organization build long ties, the benefit from high path length can be rapidly erased. Maintaining this high path length can be a difficult problem that requires repeated, costly intervention on the part of managers and policymakers. While managers and policymakers may have little control over the network structure, they often do have control over communication costs or the timing of meetings and conferences. In what follows, I expand on the policy actions that might arise from this research.

Research into communication costs has revealed several ways that they can be used to increase performance in organizations. One proposal to address communication overload has been to increase the cost of email, which would reduce collaborative efficiency and encourage individuals to only share information that is extremely important (Newport, 2016; "To Make Email

Easier We Must Make it Harder - Study Hacks - Cal Newport," n.d.). Another way that communication costs might be harnessed is through the geographic costs associated with physical distance. Managers might physically position research divisions at geographically distant regions to make it more costly in terms of time and effort for individuals to communicate, while still allowing these ties to exist (Benner & Tushman, 2003; O'Reilly & Tushman, 2004). Additionally, corporate trends toward creating virtual teams that are geographically dispersed, which has traditionally been seen as a problem, may actually have the benefit of increasing the costs of collaboration and thus lowering collaborative efficiency (Maznevski & Chudoba, 2000). Finally, organizations might use desk and office layouts to structure more effective collaborate environments so that it would be more difficult for some individuals to communicate, using physical distance as a cost on tie maintenance (Malone, 1983).

Insights into temporal networks could be used by organizations to decrease collaborative efficiency and thus increase collective innovation. Managers often face an important dilemma in structuring meetings. Should teams of engineers meet early and discuss initial solutions to their problems, or should they schedule meetings later after each individual has had a chance to explore the problem independently? Initial brainstorming sessions might lead to early convergence on suboptimal solutions, whereas delaying collaboration might result in wider individual search. Research on groupthink, production blocking, and the common knowledge effect has indicated that slowing down collaboration in problem-solving groups might prevent a premature consensus on a suboptimal solution (Diehl & Stroebe, 1987, 1991; Janis, 1972, 1982). Additionally, corporate boards need to make incredibly complex decisions about an organization's future, and reducing the frequency of board meetings can have a strong effect on increasing firm value (Vafeas, 1999). Finally, increasing research has focused on the declines in job satisfaction that result from numerous meetings (Rogelberg, Leach, Warr, & Burnfield, 2006). If such research is incorporated into management decisions, it could have the effect of reducing the frequency of meetings, which could slow down collaboration, and thus increase collective performance on complex problems.

## **Discussion**

Experimental and theoretical results confirm the central idea that network slowness increases collective performance by increasing a group's solution diversity. Simulations on varying the group's initial diversity show a clear increase in performance to having more diversity because such groups will be more likely to avoid settling on premature suboptimal solutions, and they will be more likely to search a wider portion of the solution space. Groups that were more diverse were able to find better average solutions and better top solutions.

Diversity is frequently recommended as an important method for building competitive and creative teams (Page, 2007), so the results of this study have a familiar resonance. However, the mechanism as to why diversity is important here is different. In Hong and Page (Hong & Page, 2004), diversity is useful because different problem-solving approaches can allow groups to remain robust to different problems. In this study, diversity has a temporal element, where premature convergence of the group onto a small subset of solutions prevents them from searching more widely in the solution space.

While the experimental control in this study allowed us to test the effects of network structure on how groups solve a complex problem, we had to restrict our study to a single problem in statistics, namely variable selection in a complex regression. There might be elements of this problem that do not generalize to other complex problems. For example, our interface gave individuals suggestive variables that might improve their model if they decided to explore. This feature captures some elements of real-world problem solving where experts have an indication about which way to improve next. However, many complex problems are highly uncertain and individuals may have no prior information about which direction would be best to explore. The problem was chosen in part because its structure (i.e., a high-dimensional optimization problem with high levels of interdependency) matches many real-world problems in science, technology, and design (Kauffman et al., 2000; Kauffman & Macready, 1995). In order to ensure the generalizability of these findings, it may be important to test these results in other problem domains.

A central assumption of this project was that individuals were working in parallel, which means that each individual was maximizing their own private utility function by submitting a solution, receiving feedback from the environment, and deciding to collaborate with others. There are many collective problems where individuals are not working in parallel, but rather in tandem. Examples include engineering problems where each individual works on a separate aspect of the problem, or web site design where each team is responsible for a different part of the page. Understanding how to structure collaboration for large, distributed teams when there is a significant amount of division of labor is an important area for future research.

The results in this study reveal the dangers inherent in over-connectedness. While there have been many benefits from increasing connectivity historically (Gertner, 2012), the theory presented here indicates that efficient networks will improve performance only on simple problems. Instead, on the most complex problems that we face, this trend toward increased connectivity will harm our performance. Finding ways to break down collaboration, either by altering the communication network, making communication more costly, or delaying the time of collaboration, may increase the performance of teams in engineering, design, business, and research.

## BIBLIOGRAPHY

- Aldhous, P. (2012, December 8). Specialist Knowledge Is Useless and Unhelpful. Slate. Retrieved from [http://www.slate.com/articles/health\\_and\\_science/new\\_scientist/2012/12/kaggle\\_president\\_jeremy\\_howard\\_amateurs\\_beat\\_specialists\\_in\\_data\\_prediction.html](http://www.slate.com/articles/health_and_science/new_scientist/2012/12/kaggle_president_jeremy_howard_amateurs_beat_specialists_in_data_prediction.html)
- Benner, M. J., & Tushman, M. L. (2003). Exploitation, Exploration, and Process Management: The Productivity Dilemma Revisited. *Academy of Management Review*, 28(2), 238–256. <http://doi.org/10.5465/AMR.2003.9416096>
- Byrne, C. (n.d.). The Rise Of The DIY Data Scientist. Retrieved September 18, 2015, from <http://www.fastcolabs.com/3014018/the-rise-of-the-diy-data-scientist>
- Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, 329(5996), 1194–1197.
- Centola, D. (2011). An Experimental Study of Homophily in the Adoption of Health Behavior. *Science*, 334(6060), 1269–1272. <http://doi.org/10.1126/science.1207055>
- Centola, D., & Baronchelli, A. (2015). The spontaneous emergence of conventions: An experimental study of cultural evolution. *Proceedings of the National Academy of Sciences*, 112(7), 1989–1994.
- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, 35(1), 128–152. <http://doi.org/10.2307/2393553>
- Coleman, J., Katz, E., & Menzel, H. (1957). The diffusion of an innovation among physicians. *Sociometry*, 20(4), 253–270.
- Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, 53(3), 497–509. <http://doi.org/10.1037/0022-3514.53.3.497>
- Diehl, M., & Stroebe, W. (1991). Productivity loss in idea-generating groups: Tracking down the blocking effect. *Journal of Personality and Social Psychology*, 61(3), 392–403. <http://doi.org/10.1037/0022-3514.61.3.392>
- Eberhart, R. C., & Kennedy, J. (1995). A new optimizer using particle swarm theory. In *Proceedings of the sixth international symposium on micro machine and human science* (Vol. 1, pp. 39–43). New York, NY. Retrieved from [http://www.ppgia.pucpr.br/~alceu/mestrado/aula3/PSO\\_2.pdf](http://www.ppgia.pucpr.br/~alceu/mestrado/aula3/PSO_2.pdf)
- Eberhart, R. C., Shi, Y., & Kennedy, J. (2001). *Swarm intelligence*. Elsevier. Retrieved from <https://books.google.com/books?hl=en&lr=&id=Z1z3byh1JAsC&oi=fnd&pg=PR13&dq=kennedy+eberhart+2001&ots=rzFWDs20vi&sig=UQ8y9CTKTstGdOdfb7XWewjw0OoQ>
- Ethiraj, S. K., & Levinthal, D. (2004). Bounded Rationality and the Search for Organizational Architecture: An Evolutionary Perspective on the Design of Organizations and Their Evolvability. *Administrative Science Quarterly*, 49(3), 404–437. <http://doi.org/10.2307/4131441>



- Fang, C., Lee, J., & Schilling, M. A. (2010). Balancing Exploration and Exploitation Through Structural Design: The Isolation of Subgroups and Organizational Learning. *Organization Science*, 21(3), 625–642. <http://doi.org/10.1287/orsc.1090.0468>
- Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. *Research Policy*, 30(7), 1019–1039. [http://doi.org/10.1016/S0048-7333\(00\)00135-9](http://doi.org/10.1016/S0048-7333(00)00135-9)
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer series in statistics Springer, Berlin. Retrieved from <http://statweb.stanford.edu/~tibs/book/preface.ps>
- Gertner, J. (2012). *The idea factory: Bell Labs and the great age of American innovation*. Penguin. Retrieved from [https://books.google.com/books?hl=en&lr=&id=uOMt\\_XCo81QC&oi=fnd&pg=PT9&dq=the+idea+factory&ots=\\_Sh-4tYpzi&sig=qlwwLze0SEffml\\_jW4VuF5zyKa8](https://books.google.com/books?hl=en&lr=&id=uOMt_XCo81QC&oi=fnd&pg=PT9&dq=the+idea+factory&ots=_Sh-4tYpzi&sig=qlwwLze0SEffml_jW4VuF5zyKa8)
- Gloor, P. A. (2005). Capturing team dynamics through temporal social surfaces. In *Information Visualisation, 2005. Proceedings. Ninth International Conference on* (pp. 939–944). IEEE. Retrieved from [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=1509185](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1509185)
- Goyal, S., & Vega-Redondo, F. (2005). Network formation and social coordination. *Games and Economic Behavior*, 50(2), 178–207.
- Graham, S. J. H., & Sichelman, T. M. (2010). Patenting by Entrepreneurs: An Empirical Study. *Michigan Telecommunications and Technology Law Review*, 17, 111–180.
- Gupta, A. K., Smith, K. G., & Shalley, C. E. (2006). The Interplay Between Exploration and Exploitation. *Academy of Management Journal*, 49(4), 693–706. <http://doi.org/10.5465/AMJ.2006.22083026>
- Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics Reports*, 519(3), 97–125.
- Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America*, 101(46), 16385–16389. <http://doi.org/10.1073/pnas.0403723101>
- Jackson, M. O. (2005). A survey of network formation models: stability and efficiency. *Group Formation in Economics: Networks, Clubs, and Coalitions*, 11–49.
- Janis, I. L. (1972). *Victims of Groupthink*. New York: Houghton Mifflin.
- Janis, I. L. (1982). *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. New York: Houghton Mifflin.
- Juszczyszyn, K., Musial, K., Kazienko, P., & Gabrys, B. (2012). Temporal changes in local topology of an email-based social network. *Computing and Informatics*, 28(6), 763–779.
- Kaggle Winners Tapped As Data Analytics Consultants. (n.d.). Retrieved September 18, 2015, from <http://www.informationweek.com/software/business-intelligence/kaggle-winners-tapped-as-data-analytics/240150254>
- Kauffman, S. (1993). *The Origins of Order: Self-organization and Selection in Evolution*. Oxford University Press.

- Kauffman, S. (1995). *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*. Oxford University Press.
- Kauffman, S., Lobo, J., & Macready, W. G. (2000). Optimal search on a technology landscape. *Journal of Economic Behavior & Organization*, 43(2), 141–166. [http://doi.org/10.1016/S0167-2681\(00\)00114-1](http://doi.org/10.1016/S0167-2681(00)00114-1)
- Kauffman, S., & Macready, W. (1995). Technological evolution and adaptive organizations: Ideas from biology may find applications in economics. *Complexity*, 1(2), 26–43. <http://doi.org/10.1002/cplx.6130010208>
- Kim, H., & Park, Y. (2009). Structural effects of R&D collaboration network on knowledge diffusion performance. *Expert Systems with Applications*, 36(5), 8986–8992. <http://doi.org/10.1016/j.eswa.2008.11.039>
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer. Retrieved from <http://link.springer.com/content/pdf/10.1007/978-1-4614-6849-3.pdf>
- Lazer, D., & Friedman, A. (2007). The Network Structure of Exploration and Exploitation. *Administrative Science Quarterly*, 52(4), 667–694. <http://doi.org/10.2189/asqu.52.4.667>
- Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... others. (2009). Life in the network: the coming age of computational social science. *Science (New York, NY)*, 323(5915), 721.
- Levinthal, D. A. (1997). Adaptation on Rugged Landscapes. *Management Science*, 43(7), 934–950. <http://doi.org/10.1287/mnsc.43.7.934>
- Levinthal, D. A., & March, J. G. (1981). A model of adaptive organizational search. *Journal of Economic Behavior & Organization*, 2(4), 307–333. [http://doi.org/10.1016/0167-2681\(81\)90012-3](http://doi.org/10.1016/0167-2681(81)90012-3)
- Li, R., Emmerich, M. T. M., Eggermont, J., Bovenkamp, E. G. P., Bäck, T., Dijkstra, J., & Reiber, J. H. C. (2006). Mixed-Integer NK Landscapes. In T. P. Runarsson, H.-G. Beyer, E. Burke, J. J. Merelo-Guervós, L. D. Whitley, & X. Yao (Eds.), *Parallel Problem Solving from Nature - PPSN IX* (pp. 42–51). Springer Berlin Heidelberg. Retrieved from [http://link.springer.com/chapter/10.1007/11844297\\_5](http://link.springer.com/chapter/10.1007/11844297_5)
- Malone, T. W. (1983). How Do People Organize Their Desks?: Implications for the Design of Office Information Systems. *ACM Trans. Inf. Syst.*, 1(1), 99–112. <http://doi.org/10.1145/357423.357430>
- March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, 2(1), 71–87. <http://doi.org/10.1287/orsc.2.1.71>
- March, J. G., & Simon, H. (1958). *Organizations*. Cambridge, MA: Blackwell.
- Mason, Jones, A., & Goldstone, R. L. (2008). Propagation of innovations in networked groups. *Journal of Experimental Psychology: General*, 137(3), 422–433. <http://doi.org/10.1037/a0012798>
- Mason, W. A. (2014). Collective search as human computation. In P. Michelucci (Ed.), *Handbook of Human Computation*. Springer Press.

- Mason, & Watts, D. J. (2012). Collaborative learning in networks. *Proceedings of the National Academy of Sciences*, 109(3), 764–769. <http://doi.org/10.1073/pnas.1110069108>
- Maznevski, M. L., & Chudoba, K. M. (2000). Bridging space over time: Global virtual team dynamics and effectiveness. *Organization Science*, 11(5), 473–492.
- Mihm, J., Loch, C. H., Wilkinson, D., & Huberman, B. A. (2010). Hierarchical Structure and Search in Complex Organizations. *Management Science*, 56(5), 831–848. <http://doi.org/10.1287/mnsc.1100.1148>
- Miller, J., & Page, S. (2009). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life: An Introduction to Computational Models of Social Life*. Princeton University Press.
- Miller, K. D., Zhao, M., & Calantone, R. J. (2006). Adding Interpersonal Learning and Tacit Knowledge to March's Exploration-Exploitation Model. *Academy of Management Journal*, 49(4), 709–722. <http://doi.org/10.5465/AMJ.2006.22083027>
- Millman, J. (1979). *Microelectronics*. New York, NY: McGraw-Hill. Retrieved from <http://cds.cern.ch/record/106895>
- Nelson, R., & Winter, S. (1982). *An Evolutionary Theory of Economic Change*. Cambridge, MA: Belknap Press.
- Newport, C. (2016, February 18). A Modest Proposal: Eliminate Email. Retrieved April 27, 2016, from <https://hbr.org/2016/02/a-modest-proposal-eliminate-email>
- O'Reilly, C. A., & Tushman, M. L. (2004). The ambidextrous organization. *Harvard Business Review*, 82(4), 74–81.
- Page, S. E. (2007). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press.
- Pan, W., Altshuler, Y., & Pentland, A. (2012). Decoding social influence and the wisdom of the crowd in financial trading network. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom)* (pp. 203–209). IEEE. Retrieved from [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6406285](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6406285)
- Raisch, S., Birkinshaw, J., Probst, G., & Tushman, M. L. (2009). Organizational Ambidexterity: Balancing Exploitation and Exploration for Sustained Performance. *Organization Science*, 20(4), 685–695. <http://doi.org/10.1287/orsc.1090.0428>
- Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., ... Laland, K. N. (2010). Why Copy Others? Insights from the Social Learning Strategies Tournament. *Science*, 328(5975), 208–213. <http://doi.org/10.1126/science.1184719>
- Rivkin, J. W., & Siggelkow, N. (2007). Patterned Interactions in Complex Systems: Implications for Exploration. *Management Science*, 53(7), 1068–1085. <http://doi.org/10.1287/mnsc.1060.0626>
- Roberts, M. E., & Goldstone, R. L. (2006). EPICURE: Spatial and Knowledge Limitations in Group Foraging. *Adaptive Behavior*, 14(4), 291–313. <http://doi.org/10.1177/1059712306072336>

- Rogelberg, S. G., Leach, D. J., Warr, P. B., & Burnfield, J. L. (2006). "Not Another Meeting!" Are Meeting Time Demands Related to Employee Well-Being? *Journal of Applied Psychology*, 91(1), 83–96. <http://doi.org/10.1037/0021-9010.91.1.83>
- Rogers, E. M. (2003). *Diffusion of Innovations*. New York: Free Press.
- Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762), 854–856.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.
- Shalizi, C. R., & Thomas, A. C. (2011). Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *Sociological Methods & Research*, 40(2), 211–239. <http://doi.org/10.1177/0049124111404820>
- Shao, J. (1997). An asymptotic theory for linear model selection. *Statistica Sinica*, 7(2), 221–242.
- Shore, J., Bernstein, E., & Lazer, D. (2015). Facts and Figuring: An Experimental Investigation of Network Structure and Performance in Information and Solution Spaces. *Organization Science*, 26(5), 1432–1446. <http://doi.org/10.1287/orsc.2015.0980>
- Siggelkow, N., & Levinthal, D. A. (2003). Temporarily Divide to Conquer: Centralized, Decentralized, and Reintegrated Organizational Approaches to Exploration and Adaptation. *Organization Science*, 14(6), 650–669. <http://doi.org/10.1287/orsc.14.6.650.24840>
- Siggelkow, N., & Rivkin, J. W. (2005). Speed and Search: Designing Organizations for Turbulence and Complexity. *Organization Science*, 16(2), 101–122. <http://doi.org/10.1287/orsc.1050.0116>
- Siggelkow, N., & Rivkin, J. W. (2006). When Exploration Backfires: Unintended Consequences of Multilevel Organizational Search. *Academy of Management Journal*, 49(4), 779–795. <http://doi.org/10.5465/AMJ.2006.22083053>
- Slikker, M., & van den Nouweland, A. (2000). Network formation models with costs for establishing links. *Review of Economic Design*, 5(3), 333–362.
- Sribney, B. (2011, July). What are some of the problems with stepwise regression? Retrieved from <http://www.stata.com/support/faqs/statistics/stepwise-regression-problems/>
- Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, 48(6), 1467–1478. <http://doi.org/10.1037/0022-3514.48.6.1467>
- Strang, D., & Soule, S. A. (1998). Diffusion in Organizations and Social Movements: From Hybrid Corn to Poison Pills. *Annual Review of Sociology*, 24, 265–290.
- Stuart, T. E., & Podolny, J. M. (1996). Local search and the evolution of technological capabilities. *Strategic Management Journal*, 17(S1), 21–38. <http://doi.org/10.1002/smj.4250171004>
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1). MIT press Cambridge. Retrieved from [http://www.cell.com/trends/cognitive-sciences/pdf/S1364-6613\(99\)01331-5.pdf](http://www.cell.com/trends/cognitive-sciences/pdf/S1364-6613(99)01331-5.pdf)

- Taylor, A., & Greve, H. R. (2006). Superman or the Fantastic Four? knowledge combination And experience in Innovative Teams. *Academy of Management Journal*, 49(4), 723–740. <http://doi.org/10.5465/AMJ.2006.22083029>
- To Make Email Easier We Must Make it Harder - Study Hacks - Cal Newport. (n.d.). Retrieved from <http://calnewport.com/blog/2016/04/12/to-make-email-easier-we-must-make-it-harder/>
- Vafeas, N. (1999). Board meeting frequency and firm performance. *Journal of Financial Economics*, 53(1), 113–142. [http://doi.org/10.1016/S0304-405X\(99\)00018-5](http://doi.org/10.1016/S0304-405X(99)00018-5)
- Valente, M. (2008). Pseudo-NK: An enhanced model of complexity (No. 2008/26). LEM Working Paper Series. Retrieved from <http://www.econstor.eu/handle/10419/89504>
- van de Rijt, A., Kang, S. M., Restivo, M., & Patil, A. (2014). Field experiments of success-breeds-success dynamics. *Proceedings of the National Academy of Sciences*, 111(19), 6934–6939.
- Walker, J. L. (1969). The diffusion of innovations among the American states. *American Political Science Review*, 63(03), 880–899.
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and applications* (Vol. 8). Cambridge university press.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393(6684), 440–442.
- Wisdom, T. N., Song, X., & Goldstone, R. L. (2013). Social Learning Strategies in Networked Groups. *Cognitive Science*, 37(8), 1383–1425. <http://doi.org/10.1111/cogs.12052>
- Wright, S. (1932). The roles of mutation, inbreeding, cross-breeding and selection in evolution. *Proceedings XI International Congress of Genetics*, 1, 356–366.