Team Structure and Quality Improvement in Collaborative Environments

Narine Manukyan
Department of Computer Science
University of Vermont
Burlington, Vermont 05405
Narine.Manukyan@uvm.edu

Margaret J. Eppstein
Department of Computer Science
University of Vermont
Burlington, Vermont 05405
Maggie.Eppstein@uvm.edu

Jeffrey D. Horbar
Vermont Oxford Network &
Department of Pediatrics, UVM
Burlington, Vermont 05401
horbar@vtoxford.org

***Abstract*—∞ Teams comprising diverse individuals have been shown to increase the collective creativity in jointly solving problems. However, in contexts where the purpose of collaboration is knowledge diffusion in complex environments, it is not clear whether team diversity will help or hinder effective learning. For example, in organized quality improvement collaboratives (QICs), healthcare institutions exchange information on clinical practices and outcomes with the aim of improving health outcomes at their own institutions. However, what works in one hospital may not work in others with different local contexts, due to non-linear interactions among various treatments and practices. While there is limited evidence that some QICs have resulted in improved care, it is not yet clear what factors contribute to the effectiveness of these team collaborations. In this study, we use an agent-based model to study how different strategies of team formation, including team diversity and size, affect quality improvement in simulated collaborative environments. We show that, in this context, teams comprising similar individuals outperform those with more diverse teams, and that this advantage increases with the complexity of the landscape and level of noise in assessing fitness. Furthermore, we show that larger teams of relatively homogeneous agents perform better than smaller teams, and that effective learning through team collaborations is dependent on the level of knowledge of team members' performance levels. Thus, our results suggest that groups of similar hospitals should collaborate as a single team and openly share detailed information regarding their clinical practices and outcomes. To facilitate this, we propose a virtual collaboration framework that would allow hospitals to efficiently identify potentially better practices in use at other institutions similar to theirs, without any institutions having to sacrifice the privacy of their own data. Our results may also have implications for other types of data-driven diffusive learning, such as in personalized medicine.**

***Keywords*—Collaborative learning; knowledge diffusion; quality improvement; complex environments; agent-based modeling; team diversity; team learning.**

I. Introduction

Knowledge sharing has the potential to benefit all parties involved. Much recent research has focused on studying knowledge sharing among teams of individuals collaborating to jointly solve problems [1]–[8]. In this context, diverse teams have been shown to offer some advantage. For example, in [9] the authors show that groups of diverse problem solvers can outperform more homogeneous groups of higher-ability problem solvers, because diverse individuals bring different perspectives and heuristics that aid in the creativity of the collective intelligence. Similarly, in [10] the authors show that teams with higher numbers of newcomers perform better because newcomers add to the diversity of the team. However, when the purpose of collaboration is knowledge diffusion in complex environments rather than knowledge creation, it is not clear whether diverse teams help or hinder performance.

For example, many clinicians are now participating in organized quality improvement collaboratives (QICs), in which teams from different healthcare organizations exchange information on current practices and outcomes. Nonprofit institutions such as the Vermont Oxford Network [11], [12] act as facilitators for these QICs. Team members identify potentially better practices in use at teammates' institutions and then try them out in the local context of their home institutions [13], [14]. In this type of collaborative environment, the goal is for all hospitals to improve their own performance by learning from the experiences of others in their teams. However, what works in one hospital might not work in others with different local contexts, due to non-linear interactions among various treatments and practices. Indeed, it is becoming increasingly recognized that such complex interactions are not uncommon in healthcare [15]–[19]. While there is positive but limited evidence that QICs can result in improved quality of care [20], it is not clear which factors contribute to the effectiveness of teamwork in QICs [21]–[23].

The primary goal of this contribution is to study how different strategies of team formation affect quality improvement in healthcare through information sharing and learning. Wright [24] introduced the concept of visualizing biological evolution as search of a "fitness landscape", where an individual's position in the landscape is determined by its N heritable characteristics ("features") and the height of the landscape at any given location corresponds to the reproductive success ("fitness") of the individual. The distance between individuals on the landscape corresponds to the dissimilarity in their features. In [25], the authors adopted this landscape search analogy for modeling quality improvement in healthcare. In this context, features represent clinical practices and treatments and fitness represents the probability of positive patient

This work was funded in part by the NIH Eunice Kennedy Shriver National Institute of Child Health & Human Development award 1R21HD068296.

outcomes at healthcare organizations searching the landscape. They used an agent-based model (ABM) to show that multi-institutional QICs often perform better than traditional randomized controlled trials, due to a combination of greater statistical power and more context-dependent evaluation of treatments, especially in complex environments with multiple interactions between features [15]. However, in [25], team members were randomly selected for each set of trials and team sizes were held constant.

Here, we use a similar ABM to study the impact of team characteristics on performance improvement in team members. In support of this goal we developed two methods: (a) A clustered initialization method that generates synthetic agents, such that the distribution of inter-agent attribute similarities resembles an observed distribution in 51 real hospitals working together in QICs; (b) A method for team formation that groups similar agents into the same teams. We then assess the sensitivity of performance improvement to team size, random *vs.* homophilous team formation, uniform random *vs.* clustered initialization of agent attributes, complexity of the landscape, and noise in the fitness evaluation. Based on the results of our simulations, we propose a new virtual collaboration framework for quality improvement in hospitals.

II. Methods

A. Modeling the Problem

We use the same clinical fitness landscape model as used in [25], where hospitals are modeled as agents searching for better health outcomes for their patients. The probability of patient survival $Pr(s_x)$ (or some other desired outcome) at a given healthcare institution is simulated with a high dimensional logistic function as follows:

$$Pr(s_x) = \left(1 + exp\left(-\left(\beta_0 + \sum_{i=1}^{N} \beta_i x_i + \sum_{i=1}^{(N-1)} \sum_{j=i+1}^{N} \gamma_{ij} x_i x_j + H\right)\right)\right)^{-1} \quad (1)$$

where x is a vector of N binary features ($x_i \in \{-1, 1\}$), each representing the presence or absence of the use of a specific practice, intervention, or other modifiable characteristic of the institution. Note that normalized Hamming distances between these vectors can be used to measure similarity between simulated institutions. Coefficients β_i and γ_{ij} are randomly drawn from a normal distribution with a mean of 0 and standard deviation of $L^{-0.5}$, where L is the total nuber of non-zero terms in the model. As in [25] we restrict our landscapes to those with an average fitness of 0.5 ($\beta_0 = 0$), include non-zero coefficients (β_i) for all main effects, and only model up to two-feature interactions (γ_{ij}); i.e., potential higher order interactions (H) are always set to zero. Heterogeneity in patient-level responses is modeled using Bernoulli trials with survival probability given by Eq. (1). Thus, trials with fewer numbers of patients have higher levels of noise in the fitness function, due to stochastic effects. In the remainder of this manuscript, we use the terms "agent" and "individual" to mean an abstraction of a healthcare institution.

B. Population Initialization

In [25] the authors used scattered and clustered initial populations of agents on the landscape. In the first case, uniformly scattered populations of M agents were created with N randomly generated binary features. The resulting median of pairwise normalized Hamming distances (nHD) in each scattered population was approximately 0.5. For the clustered populations the authors generated agents with median nHD of 0.1 by starting with a population of identical copies of a random individual and perturbing random features until the desired median nHD was achieved. They argue that a population of real hospitals is more likely to be clustered than scattered, due to a long history of shared learning. To begin to investigate this assumption, we assessed the actual distribution of pairwise nHDs of 51 hospitals, on a landscape of 93 binarized practices, that participated in 7 QICs through the Vermont Oxford Network (www.vtoxford.org) in September 2003. These hospitals had a median pairwise nHD of 0.34, ranging from 0 to 0.73 (Fig. 1a). Although these observations are limited, they do support the notion that hospitals are clustered rather than scattered in the feature space, although not to the degree modeled in [25], where the clustered populations exhibited a more compact distribution of pairwise nHDs. Thus, for this study, we developed a new algorithm for population initialization that we refer to as $MakeSnakingCluster$, which can generate distributions more similar to that of the observed hospitals.

The $MakeSnakingCluster$ algorithm for binary-featured landscapes works as follows. There are two tunable parameters that control the resulting distribution; d is a specified HD, and K is an integer between 1 and $M-1$. First we create a random individual with N binary features as the core individual. Then we create K individuals that are d HD away from the core individual by flipping d randomly selected bits of each of K copies of the core individual. We next pick one of these K generated individuals as the core individual for the next step and repeat the process. The algorithm terminates when M individuals have been generated (if M is not evenly divisible by K, then the last iteration is terminated early).

Although we define the algorithm above for binary-featured landscapes, it is easily generalized to landscapes with real-valued features by replacing the HDs with Euclidean distances. In Fig. 2 we illustrate an example population generated by the $MakeSnakingCluster$ algorithm in a 2-dimensional real-valued feature space, since it is easier to visualize than an N-dimensional binary space. Notice that the $MakeSnakingCluster$ algorithm generates an elongated cluster of individuals that tend to snake through the landscape, hence the name.

We compare the distribution of the pairwise HDs of the observed hospital data (Fig. 1a) with one representative distribution generated by the $MakeSnakingCluster$ algorithm using $M = 51$ individuals, $N = 93$ features, $K = 5$ steps and $d = 13$ HD (Fig. 1b). The resulting distributions of the pairwise nHDs of the clustered populations are similar to the distribution of the observed data, in contrast to distributions of scattered populations with the same N and M (e.g., Fig. 1c).

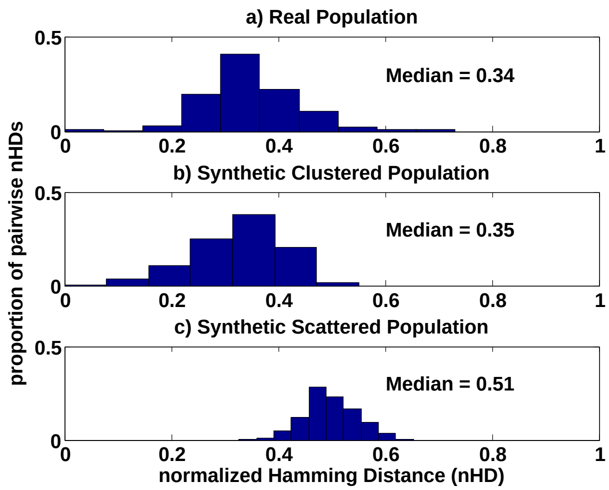

Figure 1. Representative histograms of pairwise nHDs with $N = 93$ features, $M = 51$ agents for a) a dataset of real hospitals, b) clustered synthetic random agents generated by $MakeSnakingCluster$ with $K = 5$ steps and $d = 13$ HD, c) scattered synthetic random agents.

C. Team Structure

One potentially important influence on team learning is the team construction mechanism; i.e., deciding which agents should be in the same teams. In our ABM we compare randomly formed teams (as used in [25]) to teams formed by the principal of homophily, in which similar agents are grouped together. We generate T teams of M_T homophilous agents (where $M_T = \lceil \frac{M}{T} \rceil$), using an algorithm we call $PickSimilarTeams$, as follows. First we calculate all the pairwise HDs in the population. Then for each agent we calculate the mean of the HDs between the agent and its most similar $M_T - 1$ neighbors in the population. The first team is selected to be the agent with the smallest calculated mean HD to its $M_T - 1$ closest neighbors. We then remove the individuals that were assigned to this team from the available population and repeat the process for the remaining population until we have T teams.

A visual illustration of the $PickSimilarTeams$ algorithm is shown in Fig. 2b, where the $PickSimilarTeams$ algorithm has divided the population of $M = 50$ individuals shown in Fig. 2a into T = 5 teams, with $M_T = 10$ individuals in each team. Notice that agents that are close to each other in feature space are in the same teams, hence we refer to these teams as homophilous.

D. Team Learning

We use the team learning algorithm described in [25], with minor modifications. In each generation, every agent selects one feature that has the highest difference between the majority consensus of the teammates that have higher and lower fitnesses than the agent, such that the selected feature

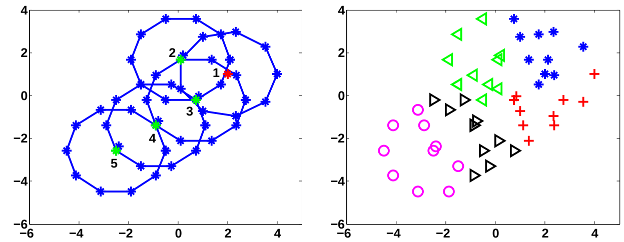

Figure 2. a) Illustration of $MakeSnakingCluster$ algorithm ($M = 50$, $N = 100$, $K = 5$ and $d = 13$), where red triangular dot represents the first core individual, and green rectangular dots represent subsequent core individuals. Although this illustration shows equal spacing of individuals on each radius around the core individuals, this constraint is not present in the actual algorithm. b) Visual illustration of teams picked by $PickSimilarTeams$ algorithm. Members of the same teams are shown with the same color and shape combinations. There are $T = 5$ teams with N_{indv} = 10 individuals per each team.

of the agent is different from the majority consensus feature value of the fitter teammates. The agent then flips the bit for this feature and tries this new state (calculates the fitness) in its local context and, if it is better than the previous state, it adopts the new feature value. Unlike in [25], where the most fit member of each team does no exploration, in this study the fittest individual in each team selects the feature that has the smallest difference between the agent's feature value and the majority consensus of all other teammates' feature values, tries its complement value, and adopts it if better. Agents are not allowed to retry the same features within 5 trial steps.

In [26] the authors show that, although imitating the best agent can be effective in noise free environments, in noisy environments this can lead to decreased performance. This is due to the fact that extreme success of the best agent in a noisy environment is often due to random effects (e.g., luck) rather than skill. The authors suggest that imitating the average members in a population might be a better strategy. The feature selection strategy we describe above can similarly mitigate the effects of noise, while also providing agent-specific customized recommendations for change based on where each agent's fitness lies relative to the others in its team.

E. Simulations

We assessed the impact of different team characteristics for $M = 100$ agents and $N = 100$ features by varying the following factors: (i) the initial population was generated to be either scattered or clustered (using $MakeSnakingCluster$ algorithm with $K = 10$ and $d = 13$); (ii) teams were selected either randomly (with resulting average median pairwise nHDs within teams of size 10 of 0.49 for scattered and 0.32 for clustered initial populations) or using the $PickSimilarTeams$ algorithm, resulting in more homophilous teams (with average median pairwise nHDs within teams of size 10 of 0.31 for scattered and 0.16 for clustered initial populations); (iii) the number of hospitals per team was varied as $M_T \in \{2, 4, 5, 10, 20, 50, 100\}$; (iv) the complexity of the fitness landscape was varied by setting the number of random two-feature interactions to be one of 0, 495 or 2475, corresponding to 0%, 10% or 50% of all possible two-feature interactions; (v) the noise for each trial was varied by setting the number of patients in each trial to be one of 10, 20, 40 or 320. In most of the experiments the agents had full knowledge of their teammates' performance levels. However, we also tried some experiments where agents had no knowledge and simply guessed their teammates' performance levels, in order to see how important this knowledge is during the feature selection process. We generated 200 random landscapes for each specified number of two-feature interactions using Eq. (1). All experiments with a given combination of parameter settings were averaged over the performances on these 200 landscapes. Each population was allowed to search each landscape for 100 trial steps.

F. Statistical Comparisions

Pairs of experiments that differed in only one parameter were compared as follows. We integrated each fitness curve over all 100 trial steps, for each of the 200 random landscapes with the specified number of 2-feature interactions. We compared these integrated values using 2-tailed paired t-tests.

III. Results

As in [25], team search consistently outperformed random search ($p < 0.01$, Fig. 3). When averaged over 200 random landscapes, the performance of individual random searchers was statistically indistinguishable whether started from initially scattered or initially clustered populations, so we only show one of these curves in Fig. 3. This finding also indicates that there is no inherent fitness advantage conferred by either of these two types of population initializations. In addition, homophilous teams (Fig. 3, black lines) significantly outperformed random teams (Fig. 3, red lines), both for scattered initial populations (Fig. 3, dashed lines) and clustered initial populations (Fig. 3, solid lines)($p < 0.01$). Our results show that agents are more effectively learning from teammates that have similar local contexts, and homophilous teams perform increasingly better than random teams as the complexity of the landscape increases (compare Fig. 3a,b,c). As expected, when noise in the fitness function increases due to smaller numbers of patients, learning becomes more difficult (Fig. 4). Interestingly, the advantage of homophilous teams over random teams becomes increasingly pronounced as the noise level increases, both for scattered and clustered initial populations (Fig.4). Other important influences on team learning are the team size and the accuracy with which agents can judge each others' performance levels. To investigate these influences, we performed two sets of experiments with clustered populations, in which we varied the sizes of the teams. In the first case, the agents had perfect knowledge of their teammates' fitnesses (as in most of the other simulations in this study). In the second case, agents had no information about the performance levels of their teammates, and therefore made random assumptions about their fitnesses. These two cases capture two extreme learning environments (completely secretive or completely open agents), and thus bracket the potential range of learning

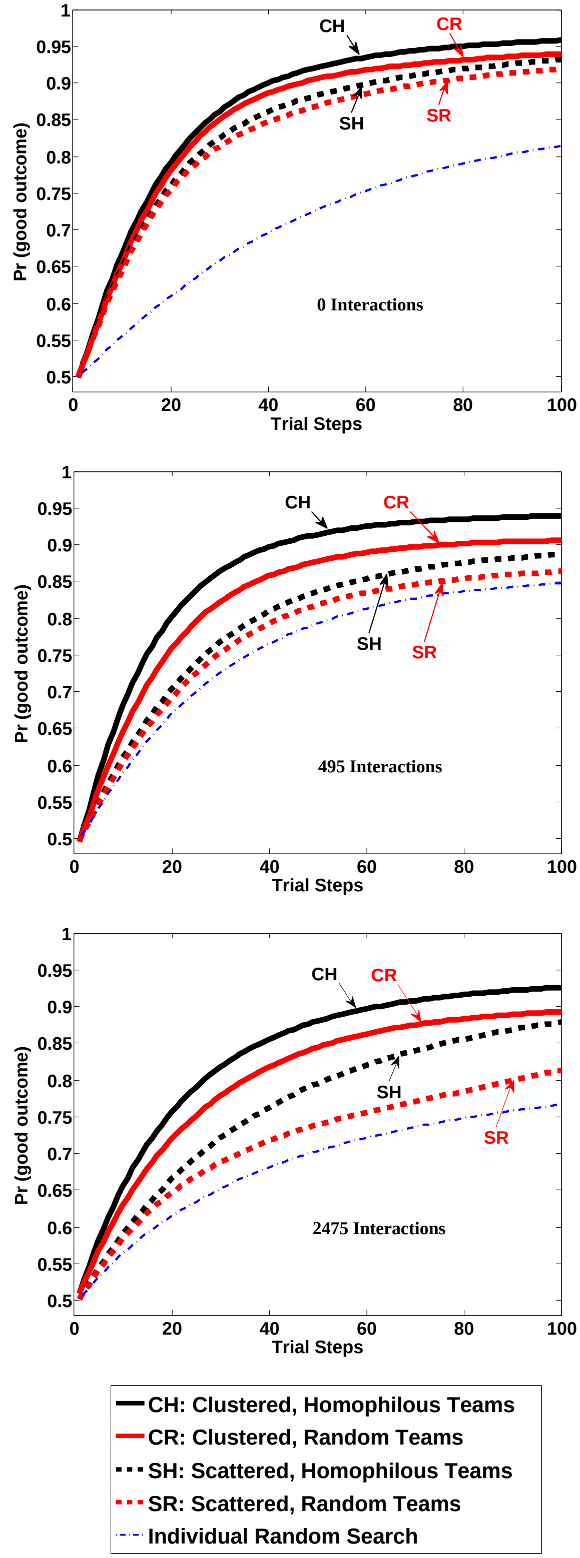

Figure 3. Mean fitnesses on 200 random landscapes at each of 100 trial steps, for landscapes with a) 0, b) 495, and c) 2475 two-feature interactions.

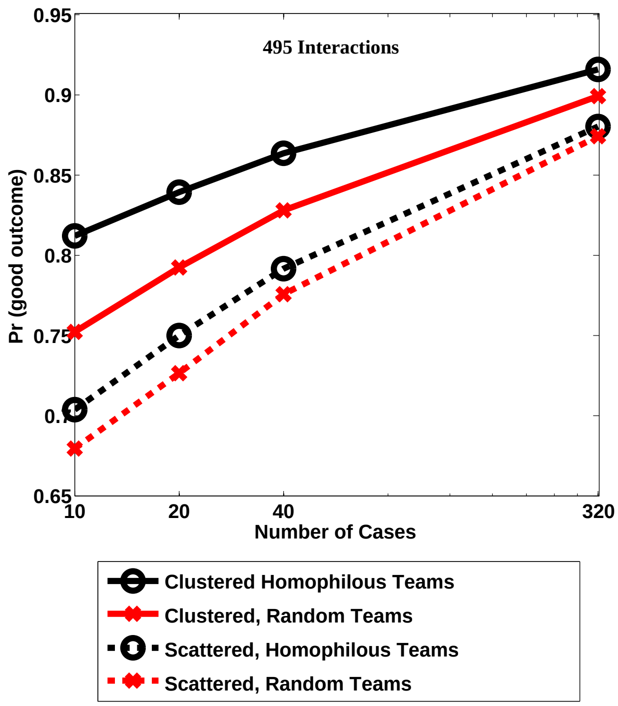

Figure 4. Mean fitnesses on 200 random landscapes with 495 two-feature interactions and clustered initial populations averaged over 100 trial steps, shown as a function of the number of patients in each trial. Note that increasing the number of cases decreases the noise in the fitness function.

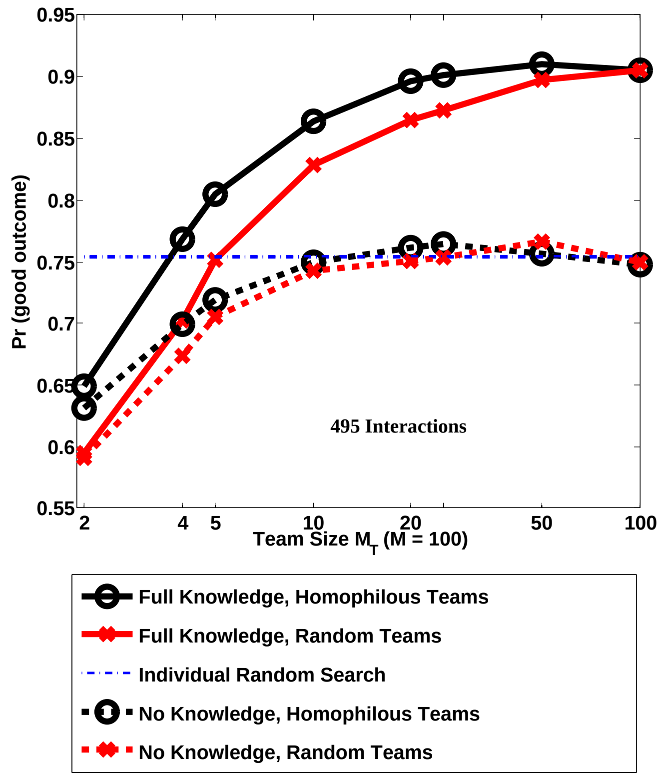

Figure 5. Mean fitnesses on 200 random landscapes with 495 two-feature interactions and clustered initial populations, over different team sizes and the amount of information team members have regarding their teammates' fitnesses. The horizontal line denotes the performance of random searchers.

environments (in reality, learning environments are typically somewhere in between these two extremes). In both cases, the performance of agents increased as the number of individuals in the teams increased in these clustered populations (Fig. 5). However, when agents had no knowledge about their teammates' performance levels, they never performed much better (and for small teams actually performed worse) than individual random search (the latter shown by the horizontal line in Fig. 5). On the other hand, in the experiments where agents had full knowledge about each others' performance levels, agents performed much better than random searchers, except for very small team sizes (Fig. 5). Note that, in these simulations a single large team of clustered agents with perfect knowledge performed the best (Fig. 5). In the next section we discuss some implications of these results.

IV. Discussion

Team learning has an important role in knowledge diffusion; we found that individuals in teams always had higher individual performance than those searching individually, as long as teams were not too small and agents were able to accurately access each others' performance levels. There is much interest in ascertaining how to best promote quality improvement through team learning [1]–[8]. In this work we have used an ABM to examine the sensitivity of quality improvement at individual simulated hospitals to different team collaboration scenarios, to try to gain insight into which factors can improve team learning. We found that teams that are formed with individuals that have more similar contexts (treatments and clinical practices) performed consistently better than randomly formed (and hence more diverse) teams, across a wide range of landscape complexity and noise levels in trial results. The consistent nature of this finding suggests that homophilous teams may be beneficial in real world collaborative learning environments, like QICs, where the emphasis is on knowledge diffusion (rather than knowledge creation). Typically, in QICs hospitals self-organize into teams without much direction from facilitators. Our results suggest that if organizations like the Vermont Oxford Network encourage teaming of similar hospitals, the hospitals might exhibit higher rates of performance improvement.

Our examination of one data set of 51 real hospitals showed that they exhibited significant clustering on a fitness landscape of 93 clinical practices. This supports the argument that hospitals are more likely to be clustered than scattered on the real clinical fitness landscape, as suggested in [25]. Our simulations using agents that were clustered in a similar way showed that these populations of individuals always performed better than randomly scattered agents. While one cannot control the distribution of actual hospitals who elect to participate in QICs, this finding suggests that clustered initial populations could potentially improve performance in evolutionary algorithms trying to optimize complex fitness functions.

Privacy of information can hinder knowledge diffusion. For example, in the healthcare domain, detailed data of patient practices and outcomes is collected and maintained securely by organizations such as the Vermont Oxford Network, but this information is not shared publicly. In our simulations, when hospitals had no information about their teammates' performance levels, they were not able to learn effectively from teammates and agents did just as well or better using random search. On the other hand, performance gains were high when they had access to perfect information about teammates.

Another interesting factor that affected individual learning was the number of individuals in a team. Our results showed that, in clustered populations where all agents were already relatively homogenous, larger teams performed better, since more information was available to learn from.

These results suggest that, in an ideal world, one would have similar hospitals collaborate as one large team and have open access to all data about each other, in order to derive optimal benefits from the collaboration. However in the real world, the maximum number of individuals in QIC teams is limited both by organizational costs related to team assembly into a collaborative environment (whether real or virtual) and by the number of individuals that can effectively work together in that environment, and real hospitals have significant privacy concerns regarding sharing detailed data on practices and outcomes.

Thus, our inferred optimal learning strategy is in conflict with the realities of team learning in QICs, both in terms of realistic team sizes and data privacy concerns. To mitigate these conflicts, we propose a new virtual collaboration framework that would allow hospitals to efficiently identify potentially better practices in use at other institutions similar to theirs without any hospitals having to sacrifice the privacy of their own institutional data. The system would compare the practices and other characteristics of a hospital to those in the knowledge base to identify similar institutions, and compare outcomes to identify which of those are better or worse performers. The system would then make intelligent customized recommendations to each hospital using an algorithm similar to the feature selection algorithm described in Section II.*D*. Hospitals would be required to share detailed information on their practices and outcomes to be able to use the system, but would be incentivized to do so by being able to benefit from the collective knowledge. In fact, many healthcare organizations are already providing similar confidential data to organizations like the Vermont Oxford Network for internal analysis, so extending this with a recommendation system would not be onerous. This framework would not require any kind of contact between the members (online or in person). Hence it would minimize time and other costs associated with the collaborative learning.

Our findings may also prove useful in other application domains, such as in collaborations designed to share best practices within franchises of a business, each with slightly different local contexts. In addition, with the growing availability of genomic data and electronic medical records, there has been increasing interest in the potential for personalized medicine [27]–[29]. It is conceivable that large databases of human DNA sequences and other relevant patient-specific attributes, health conditions, treatments, and outcomes, could be queried using an approach similar to that proposed here for virtual QICs, to suggest promising personalized treatments.

V. Summary and Conclusions

Healthcare institutions are increasingly participating in quality improvement collaboratives (QICs). In these collaborations multi-institutional teams share information to identify potentially better practices that might improve patient outcomes. Different selected practices are subsequently evaluated in the local contexts of these institutions. In this paper we modeled this collaborative learning approach using an agent-based model to study how different team characteristics affect quality improvement of agents (hospitals). We analyzed the practices of a group of real hospitals that participated in QICs and found that these hospitals are clustered in feature space. Hence, we introduced a new method for generating synthetic agents that are similarly clustered. We also introduced a new method for selecting teams of homophilous agents.

Our simulations show that, in this type of learning environments, (a) more homophilous teams show an advantage over more diverse teams, (b) this advantage increases with the complexity of the fitness landscape, (c) the advantage also increases with the noise level of the fitness evaluation, and (d) larger teams generally perform better.

Based on these results, we propose a new virtual collaboration framework that could allow hospitals to efficiently improve quality by learning from a secure and confidential knowledge base using an intelligent recommendation system to select which features to test next in their own institutions. Such a system would not require the time and resources necessary for direct collaborations, and participating hospitals would not have to publicly divulge potentially sensitive information regarding their own level of performance.

While this work was specifically motivated to inform quality improvement in healthcare institutions, our results may also have bearing on other types of learning environments where the aim is the diffusion of contextually relevant knowledge in complex environments.

Acknowledgment

The authors would like to thank Mike Kenny, Kathy Leahy, and the Vermont Oxford Network for their time and effort in providing the hospital data. We also thank Stuart A. Kauffman, Jeffrey S. Buzas, and Donna M. Rizzo for stimulating discussions that benefitted this work.

References

[1] P. Cohen and H. Levesque, "Teamwork," *Nous*, pp. 487–512, 1991.

[2] M. West, *Effective teamwork: Practical lessons from organizational research*. Oxford: Blackwell Publishing, 2012.

[3] B. Poulton and M. West, "Effective multidisciplinary teamwork in primary health care," *Journal of Advanced Nursing*, vol. 18, no. 6, pp. 918–925, 2008.

[4] M. Turner, *Groups at work: Theory and research*. Hillsdale, NJ: Lawrence Erlbaum, 2000.

[5] J. LePine, R. Piccolo, C. Jackson, J. Mathieu, and J. Saul, "A meta-analysis of teamwork processes: tests of a multidimensional model and relationships with team effectiveness criteria," *Personnel Psychology*, vol. 61, no. 2, pp. 273–307, 2008.

[6] M. Marks, J. Mathieu, and S. Zaccaro, "A temporally based framework and taxonomy of team processes," *Academy of Management Review*, pp. 356–376, 2001.

[7] E. Salas, N. Cooke, and M. Rosen, "On teams, teamwork, and team performance: Discoveries and developments," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 50, no. 3, pp. 540–547, 2008.

[8] T. Manser, "Teamwork and patient safety in dynamic domains of healthcare: a review of the literature," *Acta Anaesthesiologica Scandinavica*, vol. 53, no. 2, pp. 143–151, 2008.

[9] L. Hong and S. Page, "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*, vol. 101, no. 46, pp. 16 385–16 389, 2004.

[10] R. Guimera, B. Uzzi, J. Spiro, and L. Amaral, "Team assembly mechanisms determine collaboration network structure and team performance," *Science*, vol. 308, no. 5722, pp. 697–702, 2005.

[11] J. Horbar, "The Vremount Oxford Network: evidence-based quality improvement for neonatology," *Pediatrics*, vol. 103, no. Supplement E1, pp. 350–359, 1999.

[12] J. Horbar, R. Soll, and W. Edwards, "The Vremount Oxford Network: a community of practice," *Clinics in perinatology*, vol. 37, no. 1, p. 29, 2010.

[13] D. Berwick, "Broadening the view of evidence-based medicine," *Quality and Safety in Health Care*, vol. 14, no. 5, pp. 315–316, 2005.

[14] J. Horbar, J. Rogowski, P. Plsek *et al.*, "Collaborative quality improvement for neonatal intensive care," *Pediatrics*, vol. 107, no. 1, pp. 14–22, 2001.

[15] A. Rodrigues, *Drug-drug interactions*. New York, NY: Informa Healthcare, 2008.

[16] K. Johnell and I. Klarin, "The relationship between number of drugs and potential drug-drug interactions in the elderly: A study of over 600000 elderly patients from the swedish prescribed drug register," *Drug Safety*, vol. 30, no. 10, pp. 911–918, 2007.

[17] A. Dechartres, I. Boutron, L. Trinquart *et al.*, "Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study," *Annals of internal medicine*, vol. 155, no. 1, p. 39, 2011.

[18] R. Gray, "A bayesian analysis of institutional effects in a multicenter cancer clinical trial," *Biometrics*, pp. 244–253, 1994.

[19] R. Horwitz, B. Singer, R. Makuch, and C. Viscoli, "Can treatment that is helpful on average be harmful to some patients? A study of the conflicting information needs of clinical inquiry and drug regulation," *Journal of Clinical Epidemiology*, vol. 49, no. 4, pp. 395–400, 1996.

[20] L. Schouten, M. Hulscher, J. Everdingen, R. Huijsman, and R. Grol, "Evidence for the impact of quality improvement collaboratives: systematic review," *Bmj*, vol. 336, no. 7659, pp. 1491–1494, 2008.

[21] L. Schouten, R. Grol, and M. Hulscher, "Factors influencing success in quality-improvement collaboratives: development and psychometric testing of an instrument," *Implementation Science*, vol. 5, no. 1, pp. 1–9, 2010.

[22] S. Taylor, S. Dy, R. Foy *et al.*, "What context features might be important determinants of the effectiveness of patient safety practice interventions?" *BMJ Quality & Safety*, vol. 20, no. 7, pp. 611–617, 2011.

[23] P. Newton, E. Halcomb, P. Davidson, and A. Denniss, "Barriers and facilitators to the implementation of the collaborative method: reflections from a single site," *Quality and Safety in Health Care*, vol. 16, no. 6, pp. 409–414, 2007.

[24] S. Wright, "The roles of mutation, inbreeding, crossbreeding and selection in evolution," in *Proceedings of the sixth international congress on genetics*, vol. 1, no. 6, 1932, pp. 356–366.

[25] M. Eppstein, J. Horbar, J. Buzas, and S. Kauffman, "Searching the clinical fitness landscape," *PLoS ONE*, vol. 7, no. 11, p. e49901, 2012.

[26] J. Denrell and C. Liu, "Top performers are not the most impressive when extreme performance indicates unreliability," *Proceedings of the National Academy of Sciences*, vol. 109, no. 24, pp. 9331–9336, 2012.

[27] T. Sandmann and M. Boutros, "Screens, maps & networks: from genome sequences to personalized medicine," *Current Opinion in Genetics & Development*, vol. 22, pp. 36–44, 2012.

[28] G. Vaidyanathan, "Redefining clinical trials: the age of personalized medicine," *Cell*, vol. 148, no. 6, pp. 1079–1080, 2012.

[29] E. Juengst, R. Settersten, J. Fishman, and M. McGowan, "After the revolution? Ethical and social challenges in personalized genomic medicine," *Personalized Medicine*, vol. 9, no. 4, pp. 429–439, 2012.