Network Science of Teams: Current State and Future Trends

Liangyue Li and Hanghang Tong

***Abstract*—Teams are increasingly indispensable to achievements in any organization. Despite the organizations substantial dependency on teams, fundamental knowledge about the conduct of team-enabled operations is lacking, especially at the social, cognitive and information level in relation to team performance and network dynamics. Generally speaking, the team performance can be viewed as the composite of its users, the tasks that the team performs and the networks that the team is embedded in or operates on. The goal of this article is to (1) provide a comprehensive review of the recent advances in optimizing teams performance in the context of networks; and (2) identify the open challenges and future trends. We believe this is an emerging and high-impact topic in computational social science, which will attract both researchers and practitioners in the data mining as well as social science research communities. Our emphasis will be on (1) the recent emerging techniques on addressing team performance optimization problem; and (2) the open challenges/future trends, with a careful balance between the theories, algorithms and applications.**

***Index Terms*—Network science of teams, team performance characterization, performance prediction, team optimization.**

I. Introduction

In defining the essence of professional teamwork, Hackman and Katz [1] stated that teams function as 'purposive social systems', defined as people who are readily identifiable to each other by role and position and who work interdependently to accomplish one or more collective objectives. Teams are increasingly indispensable to achievements in any organization. This is perhaps most evident in multinational organizations where communication technology has transformed the geographically dispersed teams and networks. Business operations in the large organizations now involve large, interactive, and layered networks of teams and personnel communicating across hierarchies and countries during the execution of complex and multifaceted international businesses. Despite the organizations' substantial dependency on teams, fundamental knowledge about the conduct of team-enabled operations is lacking, especially at the *social, cognitive* and *information* level in relation to team performance and network dynamics. What do high-performing engineering/design/sale teams share in common with respect to their communication patterns? How to predict a team's performance before it starts to work on the assigned project? How to foster productive behavioral changes of team members and leaders in order to optimize performance?

The authors are from the School of Computing, Informatics, Decision Systems Engineering, Arizona State University, Tempe, AZ 85281, USA. Corresponding Author e-mail: ({liangyue, htong6}@asu.edu).

Generally speaking, the **team performance** can be viewed as the composite of the following three aspects, including (1) its users, (2) tasks that the team performs and (3) the networks that the team is embedded in or operates on. In this article, we will provide a comprehensive review of the recent advances in *characterizing*, *predicting* and *optimizing* teams' performance in the context of composite networks (i.e., social-cognitive-information networks).

Research in sociology and psychology has long been trying to characterize the high-performing teams in organizations. The basics of team effectiveness were identified by J. Richard Hackman, who uncovered a groundbreaking insight: what matter most to collaboration are certain enabling conditions. Recent studies find that three of Hackman's conditions – a compelling direction, a strong structure, and a supportive context – continue to be particularly critical to team success [2]. We would comprehensively survey related literatures in sociology, psychology and computer science.

Understanding the dynamic mechanisms that drive the success of high-performing teams can provide the key insights into building the best teams and hence lifting the productivity and profitability of the organizations. For this purpose, we introduce some of the recent work on developing novel predictive models to forecast the long-term performance of teams (*point prediction*) as well as the pathway to impact (*trajectory prediction*). It is also worthwhile to quantitatively examine the relationship between the team level and individual level performances to build a joint predictive model.

From the practical perspective, it is important to form a good team in the context of networks for a given tasks. For an existing team, it is often desirable to optimize its performance through expanding the team by bringing a new team member with certain expertise, finding a new candidate to replace a current under-performing team member or downsizing the team for the purpose of cost reduction. We would introduce recent advances in team performance optimization.

II. Team Performance Characterization

A. Collective Intelligence

The notion of individual intelligence was first proposed by Charles Spearman when he noticed that school kids who did well in one school subject tend to do well in many other school subjects [3]. The observations that the average correlation among individual's performance on a variety of cognitive tasks is positive and the first factor extracted using a factor analysis accounts for about 30-50% of the variance indicate the existence of general intelligence. The first factor is usually referred to as general intelligence. We can give people a relatively limited set of items and the scores of these items can predict how they perform across a variety of domains and over a long period of time. Such intelligence test can predict not only how kids do in school in multiple subjects, but also the probability that they would be successful in their future career. This is perhaps the most empirically replicated facts in most of the psychology.

A group of researchers at CMU set out to test whether a similar notion of collective intelligence exists in a team of people, i.e., whether a single factor exists from the team's performance on a variety of tasks [4]. They enlisted 40 and 152 teams of size two to five for their two studies. They assigned a diverse set of group tasks to these teams. The tasks can be categorized into four types, namely, 'generate', 'choose', 'negotiate', and 'execute'. The results support their initial hypothesis that the average correlation among the teams' scores on the diverse set of tasks is positive and the factor analysis reveals that one single factor can account for more than 43% of the variance. Additionally, the collective intelligence score calculated using the first factor can strongly predict the team's performance on a future criterion tasks (e.g., video game and architectural design). Surprisingly, the average team member intelligence and the maximum team member intelligence are not that predictive of the future performance, which tells us that simply assigning a team of smart people does not promise a smart team. But what are the ingredients that are important to an intelligent team? Surprisingly, the team processes, e.g., group cohesion, motivation, and satisfacation, traditionally regarded as important to team performance, are not predictive of collective intelligence. The collective intelligence is found to be positively correlated with the average social perceptiveness of the team members and negatively correlated with the variance in the number of speaking turns by team embers.

B. Virtual Teams in Online Games

The above research about collective intelligence (CI) is mainly on traditional teams where team members have face to face interactions. It would be interesting to examine whether the collective intelligence also exists in virtual teams. Virtual teams are diverse, dispersed, digital and dynamic, e.g. the Multiplayer Online Battle Arena (MOBA) teams. Considering that such teams perform tasks at a fast pace without explicit face-to-face or verbal communitication, other means of coordinations might player a more critical role here, e.g., tacit coordiation, or coordinations that happen without explicit verbal communication [5]. Studying how collective intelligence works in such MOBA teams could also inform the operations of other virtual teams commonly seen in business world, where teams are dipsersed across geographical boundaries and making decisions at a fast pace.

One recent study [5] examines collective intelligence in *League of Legends (League)* teams, a popular game with worldwide monthly active user base of 67 million. In *League*, a match is between two teams of five members and teams can be formed either through the game's matchmaking algorithms or by recruiting other players in the game community. One team's goal is to destroy the opponent team's base. The authors hypothesize that (1) CI will predict team performance in *League*, (2) social perceptiveness and proportion of woman will be positively associated with CI in *League*, and (3) CI will not be associated with equality of contribution to conversation or decision making in *League* teams. In order to know the CI, game performance, and team characteristics, the authors collect data from three sources: (1) all team members completed a questionnaire on their own about information on their demographics, psychological variables, cognition, affect, etc; (2) the teams took the Test of Collective Intelligence (TCI), an online test battery, as a group to measure the collective intelligence of each team; and (3) the play statistics including the team performances are provided by Riot Games. There were 248 teams that completed all components of the study and 85% of the teams are all males. The authors find that CI also exists in *League* teams from factor analysis and it is positively correlated with the performance measure of the teams controlling for individual and team play time. Besides, CI is positively correlated with the number of woman in the team and is positively correlated with social perceptiveness, but the proportion of woman and social perceptiveness are not correlated. What's interesting is that the equality of communication measured by standard deviation of chat lines and chat word count is not significantly correlated with CI. In addition, CI is negatively correlated with some group process, e.g., perceived equality in decision making, frequency of game-specific communication. These suggest that highly dispersed and dynamic virtual teams tend to adopt a tacit coordination method.

C. Networks in Sports Teams

Recently, a number of works start to examine the network structure in sports teams in relation to their performances [6], [7]. Using Euro Cup 2008 tournament data, researchers construct a directed network of "ball flow" among players in the team [6], where nodes represent players and edge weights indicate the number of successful passes between two players. They use the betweenness centrality of the player with regard to the opponent's goal as the performance measure of a player and the team level performance is defined as the average performance of the top-k players. They find that the difference between two teams' defined performance measure is indicative of their winning probability. In a similar study, researchers use English Premier League soccer team data to find that increased network density among team members lead to increased team performance and increased centralization of team play decreases the performance [7].

D. Networks in GitHub Teams

Social coding platforms such as GitHub offer a unique experience to developers as they can subscribe to activities of other developers. Using GitHub data, researchers construct two types of networks [8]: a project-project network, where nodes represent projects and two nodes are connected if they share at least one common developer; and developer-developer network, were nodes represent developers and two nodes are connected if they have collaborated in the same project. They find that in the project-project network, the diameter of the largest connected component is 9 with the average shortet path 3.7, which is more interconnected than human networks; and in the developer-developer network, the average shortest path is 2.47. Compared with the average shortest path of Facebook 4.7, we see social coding enables substantially more collaborations among developers.

III. Team Performance Prediction

A. Long-term Performance Prediction

For the discussion in this section, we mainly use research teams since their performance can be measured by the impact of their team products (e.g., research papers, patents). Understanding the dynamic mechanisms that drive those high-impact scientific work is a long-debated research topic and has many important implications, ranging from personal career development and recruitment search, to the jurisdiction of research resources. Scholars, especially junior scholars, who could master the key to producing high-impact work would attract more attentions as well as research resources; and thus put themselves in a better position in their career developments. High-impact work remains as one of the most important criteria for various organization (e.g. companies, universities and governments) to identify the best talents, especially at their early stages. It is highly desirable for researchers to judiciously search the right literature that can best benefit their research.

Recent advances in characterizing and modeling scientific success have made it possible to forecast the long-term impact of scientific work. Wuchty et al. [9] observe that papers with multiple authors receive more citations than solo-authored ones. Uzzi et al. [10] find that the highest-impact science work is primarily grounded in atypical combinations of prior ideas while embedding them in conventional knowledge frames. Recently, Wang et al. [11] develop a mechanistic model for the citation dynamics of individual papers. In particular, they identify three fundamental drives underlying the citation histories of individual papers, namely, preferential attachment, temporal citation trend, and fitness. They combine these three factors into a mechanistic model, which fits well on the Physical Review corpus and is able to predict future citations with good accuracy. In data mining community, efforts have also been made to predict the long-term success. Carlos et al. [12] estimate the number of citations of a paper based on the information of past articles written by the same author(s). Yan et al. [13] design effective content (e.g., topic diversity) and contextual (e.g., author's h-index, venue's centrality) features for the prediction of future citation counts.

To collectively address a number of key algorithmic challenges, namely, scholarly feature design (C1), non-linearity (C2), domain heterogeneity (C3), and dynamics (C4), in relation to predicting long-term scientific impact, a joint predictive model *iBall* is proposed [14]. First (for C1), they find that the citation history of a scholarly entity (e.g., paper, researcher, venue) in the first three years (e.g., since its publication date) is a strong indicator of its long-term impact (e.g., the accumulated citation count in ten years); and adding additional

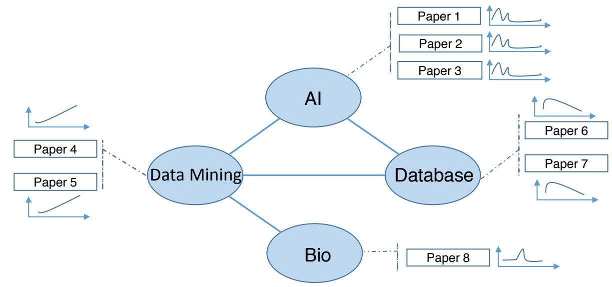

Fig. 1. An illustrative example of the joint predictive model iBall [14]. Papers from the same domain (e.g., AI, Databases, Data Mining and Bio) share similar patterns in terms of attracting citations over time. Certain domains (e.g., AI and Data Mining) are more related with each other than other domains (e.g., AI and Bio). The authors want to jointly learn four predictive models (one for each domain), with the goal of encouraging the predictive models from more related domains (e.g., AI and Data Mining) to be 'similar' with each other.

contextual or content features brings little marginal benefits in terms of prediction performance. This not only largely simplifies the feature design, but also enables them to forecast the long-term scientific impact at its early stage. Second (for C2), their joint predictive model is flexible, being able to characterize both the linear and non-linear relationship between the features and the impact score. Third (for C3), they propose to jointly learn a predictive model to differentiate distinctive domains, while taking into consideration the commonalities among these similar domains (see an illustration in Figure 1). Fourth (for C4), they further propose a fast on-line update algorithm to adapt our joint predictive model efficiently over time to accommodate newly arrived training examples (e.g., newly published papers).

B. Performance Trajectory Forecasting

From the prediction perspective, more often than not, it is of key importance to forecast the pathway to impact for scholarly entities (e.g., how many citations a research paper will attract in each of several consecutive years in the future). The impact pathway often provides a good indicator of the shift of the research frontier. For instance, the rapid citation count increase of the deep learning papers reveals an emerging surge of this topic. The impact pathway can also help trigger an early intervention should the impact trajectory step down in the near future.

The state of the art has mainly focused on modeling the long-term scientific impact for the early prediction, as we have discussed in the previous subsection. They are not directly applicable to forecasting the impact pathway, e.g., citation counts in each of the next 10 years. One baseline solution is to treat the impacts across different years independently and to train a separate model for each of the impacts. This treatment might ignore the inherent relationship among different impacts across different years, and thus might lead to sub-optimal performance. Having this in mind, a better way could be to apply the existing multi-label/multi-task learning methods to exploit the relation among impacts across different years. Nonetheless, these general-purpose multi-label/multi-task learning approaches might overlook some unique characteristics of the impact pathway prediction.

A new predictive model (*iPath*) is proposed to simultaneously fulfill two design objectives with the unique properties of impact pathway prediction [15]. First, *prediction consistency.* Intuitively, the scholarly impacts at certain years might be correlated with each other, which, if vetted carefully, could boost the prediction performance (i.e., multi-label or multi-task learning). Here, one difficulty for impact pathway prediction is that such a relation structure is often not accurately known a priori. The *iPath* model is capable of simultaneously inferring the impact relation structure from the training data and leveraging such (inferred) relation to improve the prediction performance. Second, *parameter smoothness.* For a given feature of the predictive model, one do not expect its effect on the impacts of adjacent years would change dramatically. The *iPath* model is able to capture such temporal smoothness.

C. Team Performance vs. Individual Performance

The great Greek philosopher Aristotle articulated more than 2,000 years ago that "*the whole is greater than the sum of its parts*". This is probably most evident in *teams*, which, through appropriate synergy, promise a collective outcome (i.e., team performance) that is superior than the simple addition of what each individual team member could achieve (i.e., individual productivity). For example, in professional sports (e.g., NBA), the peak performance of a grass-root team is often attributed to the harmonic teamwork between the team players rather than the individual player's capability. Beyond teams, the *part-whole* relationship also routinely finds itself in other disciplines, ranging from crowdsourcing (e.g., Community-based Question Answering (CQA) sites [16]), to reliability assessment of a networked system of components [17].

From the algorithmic perspective, an interesting problem is to predict the outcome of the whole and/or parts [18]. In organizational teams, it is critical to appraise the individual performance, its contribution to the team outcome as well as the team's overall performance [19]. Despite much progress has been made, the existing work either develop separate models for predicting the outcome of whole and parts without explicitly utilizing the part-whole relationship [14], [15], or implicitly assume the outcome of the whole is a *linear* sum of the outcome of the parts [16], which might oversimplify the complicated part-whole relationships (e.g., non-linearity). The key to address these limitations largely lies in the answers to the following questions, i.e., to what extent does the outcome of parts (e.g., individual productivity) and that of the whole (e.g., team performance) correlated, beyond the existing linear, independency assumption? How can we leverage such potentially non-linear and interdependent 'coupling' effect to mutually improve the prediction of the outcome of the whole and parts collectively? The challenges come as two-folds. First (*Modeling Challenge*), the relationship between the parts outcome and whole outcome might be complicated, beyond the simple addition or linear combination. Moreover, the composing parts of the whole might not be independent with each other. In a networked system, the composing parts are connected with each other via an underlying network. Such part-part interdependency could have a profound impact on both the part outcome correlation as well as each part's contribution to the whole outcome. Second (*Algorithmic Challenge*), the complicated part-whole relationship (i.e., non-linearity and interdependency) also poses an algorithmic challenge, as it will inevitably increase the complexity of the corresponding optimization problem.

To address these challenges, a joint predictive model named *PAROLE* is proposed to simultaneously and mutually predict the part and whole outcomes [20]. First, *model generality*, the proposed model is flexible in admitting a variety of linear as well as non-linear relationships between the parts and whole outcomes, including *maximum aggregation, linear aggregation, sparse aggregation, ordered sparse aggregation* and *robust aggregation*. Moreover, it is able to characterize part-part interdependency via a graph-based regularization, which encourages the tightly connected parts to share similar outcomes as well as have similar effect on the whole outcome. Second, *algorithm efficacy*, the authors propose an effective and efficient block coordinate descent optimization algorithm, which converges to the coordinate-wise optimum with a linear complexity.

IV. Team Performance Optimization

A. Team Formation

Team formation studies the problem of assembling a team of people to work on a project. The first work that studies team formation in the context of social networks finds a team of experts who possess the desired skills and have strong team cohesion to ensure the team success [21]. In particular, they define two communication cost based on the diameter as well as the minimum spanning tree of the induced team subgraph. Since the corresponding optimization problems are NP-complete, they devise approximation algorithms by exploiting the relationship to Multiple-Choice Cover and Group Steiner Tree problems. As follow-up work, Anagnostopoulos et al [22] study forming teams to accommodate a sequence of tasks arriving in an online fashion. Rangapuram et al [23] allow incorporating many realistic requirements (e.g., inclusion of a designated team leader) into team formation based on a generalization of the densest subgraph problem. Beyond that, minimizing the tensions among the team members is considered [24]. With the presence of the underlying social network, the set cover problem is complicated by the goal of lowering the communication cost at the same time. Cao et al [25] develop an interactive group mining system that allows users to efficiently explore the network data and from which to progressively select and replace candidate members to form a team. Bogdanov et al [26] study how to extract a diversified group pulled from strong cliques given a network; this ensures that the group is both comprehensive and representative of the whole network. Cummings and Kiesler [27] find that prior working experience is the best predictor of collaborative tie strength. To provide insights into designs of online communities and organizations, the systematic differences in appropriating social softwares among different online enterprise communities is analyzed in [28]. The patterns of informal networks and communication in distributed global software teams using social network analysis is also investigated in [29]. Specific communication structures are proven critical to new product development delivery performance and quality [30]. To assess the skills of players and teams in online multi-player games and team-based sports, "team chemistry" is also accounted for in [31], [32].

B. Team Member Replacement

The churn of team members is a common problem across many application domains. For example, an employee in a software or sales team might decide to leave the organization and/or be assigned to a new tasks. The loss of the key member (i.e., the irreplaceable) might bring the catastrophic consequence to the team performance. *How can we find the best alternate (e.g., from the other members within the organization), when a team member becomes unavailable?* Despite the frequency with which people leave a team before a project/task is complete and the resulting disruption [33], replacements are often found opportunistically and are not necessarily optimal.

It is conjectured that there will be less disruption when the team member who leaves is replaced with someone with similar relationships with the other team members. This conjecture is inspired by some recent research which shows that team members prefer to work with people they have worked with before [34] and that distributed teams perform better when members know each other [27]. Furthermore, research has shown that specific communication patterns amongst team members are critical for performance [30]. Thus, in addition to factors such as skill level, maintaining the same or better level of familiarity and communication amongst team members before and after someone leaves should reduce the impact of the departure. In other words, for team member replacement, the similarity between individuals should be measured in the context of the team itself. More specifically, a good team member replacement should meet the following two requirements. First (*skill matching*), the new member should bring a similar skill set as the current team member to be replaced. Second (*structure matching*), the new member should have a similar network structure as the current team member in connecting the rest of the team members.

Armed with this conjecture, *Team Member Replacement* problem is formally defined by the notation of graph similarity/kernel [35], [36]. By modeling the team as a labeled graph, the graph kernel provides a natural way to capture both the skill and structure match as well as the interaction of both. However, for a network with n individuals, a straightforward method would require $O(n)$ graph kernel computations for one team member replacement, which is computationally intractable. For example, for the *DBLP* dataset with almost 1M users (i.e., $n \approx 1,000,000$), the authors find that it would take 6,388s to find one replacement for a team of size 10. To address the computational challenges, they propose a family of fast algorithms by carefully designing the pruning strategies and exploring the smoothness between the existing and the new teams. From their extensive experimental evaluations, they find that (1) by encoding both the skill and structural matching, it leads to a much better replacement result. Compared with the best alternative choices, they achieve 27% and 24% *net increase* in average recall and precision, respectively; (2) the fast algorithms are orders of magnitude faster and scale *sub-linearly*. For example, their pruning strategy alone leads up to 1,709× speed-up, without sacrificing any accuracy.

C. Team Enhancement

Different from *Team Member Replacement*, *Team Refinement* considers refining a team by replacing one member with another with the desired skill sets and communication connections. In the above two problems, the team size remains the same. In *Team Expansion*, we want to expand the team by adding a member with certain skill sets and communication structure. For instance, a software project team wants to develop a new feature of natural language search and a new member with Natural Language Processing (NLP) skill will be recruited. On the contrary, in *Team Shrinkage*, the size of a team needs to be reduced in response to new challenge such as a shortage of the available resource (e.g., a budget cut). In all cases, the resulting disruption [33] should be minimized.

By careful inspection, Li et al. [36] identify the problem similarity between *Team Refinement*, *Team Expansion* and *Team Replacement* and propose these problems can be formulated in a way to share common technical solutions. In *Team Refinement*, one team member is edited to a desired skill and network structure configuration. Since such edited member might not exist in the rest of the network, they call it a 'virtual member'. By replacing this 'virtual member' as in *Team Replacement*, they can solve *Team Refinement*. Similarly, in *Team Expansion*, the desired new member might also be a 'virtual member'. After adding this 'virtual member' to the current team and then replacing the 'virtual member', they can solve *Team Expansion*. They propose to reduce the disruption induced by the team alteration by maintaining the team-level similarity (between the original and the new teams), which includes skill similarity as well as structural similarity.

D. Interactive Visualization System

A system called *TeamOPT* (http://team-net-work.org/) is developed to assist users in optimizing the team performance interactively to support the changes to a team [37] (See Fig. 2 for an example). *TeamOPT* takes as input a large network of individuals (e.g., co-author network of researchers) and is able to assist users in assembling a team with specific requirements and optimizing the team in response to the changes made to the team. To the best of our knowledge, this is the first system specializing in forming and optimizing teams with the following key features. First (*effectiveness*), they carefully identify the design objectives and develop effective algorithms with the key technique of graph kernels. Compared with other competitors, their algorithm can achieve the highest precision and recall in finding the best team member candidate. Second (*interaction*), they design fast solutions to their algorithms, enabling an interactive user experience with users' feedback

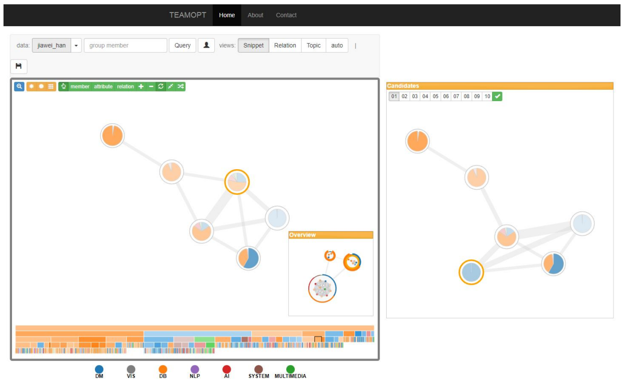

Fig. 2. A snapshot of the *TeamOPT* interactive visualization system.

in the loop. Third (*deployment*), they build the system with HTML5, Javascript, D3.js (front-end) and Python CGI (back-end).

V. Future Directions

As an emerging field, the network science of teams is still in its early stage and remains an active area of exploration. Future directions include modeling the hierarchical structure within organizations by extending the *PAROLE* model and modeling the heterogeneous goals among the team members. In the team optimization work, one implicit assumption is that the original team is performing well and maintaining the similarity with the original team can promise a similar high performance. We want to point out that when the assumption does not hold, one can leverage the actual or predicted future performance as feedback to guide the team optimization process, using advanced reinforcement learning techniques. Since team operations often involve important staffing decisions, it is critical to have team performance prediction and optimization to be explainable to the end users [38], [39].

References

[1] J. R. Hackman and N. Katz, *Group behavior and performance*. New York: Wiley, 2010, pp. 1208–1251.

[2] M. Haas and M. Mortensen, "The secrets of great teamwork." *Harvard business review*, vol. 94, no. 6, pp. 70–6, 2016.

[3] C. Spearman, "" general intelligence," objectively determined and measured," *The American Journal of Psychology*, vol. 15, no. 2, pp. 201–292, 1904.

[4] A. W. Woolley, C. F. Chabris, A. Pentland, N. Hashmi, and T. W. Malone, "Evidence for a collective intelligence factor in the performance of human groups," *science*, vol. 330, no. 6004, pp. 686–688, 2010.

[5] Y. J. Kim, D. Engel, A. W. Woolley, J. Y.-T. Lin, N. McArthur, and T. W. Malone, "What makes a strong team?: Using collective intelligence to predict team performance in league of legends," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 2017, pp. 2316–2329.

[6] J. Duch, J. S. Waitzman, and L. A. N. Amaral, "Quantifying the performance of individual players in a team activity," *PloS one*, vol. 5, no. 6, p. e10937, 2010.

[7] T. U. Grund, "Network structure and team performance: The case of english premier league soccer teams," *Social Networks*, vol. 34, no. 4, pp. 682–690, 2012.

[8] F. Thung, T. F. Bissyande, D. Lo, and L. Jiang, "Network structure of social coding in github," in *Software maintenance and reengineering (csmr), 2013 17th european conference on*. IEEE, 2013, pp. 323–326.

[9] S. Wuchty, B. F. Jones, and B. Uzzi, "The increasing dominance of teams in production of knowledge," *Science*, vol. 316, no. 5827, pp. 1036–1039, 2007.

[10] B. Uzzi, S. Mukherjee, M. Stringer, and B. Jones, "Atypical combinations and scientific impact," *Science*, vol. 342, no. 6157, pp. 468–472, 2013.

[11] D. Wang, C. Song, and A.-L. Barabási, "Quantifying long-term scientific impact," *Science*, vol. 342, no. 6154, pp. 127–132, 2013.

[12] C. Castillo, D. Donato, and A. Gionis, "Estimating number of citations using author reputation," in *String processing and information retrieval*. Springer, 2007, pp. 107–117.

[13] R. Yan, J. Tang, X. Liu, D. Shan, and X. Li, "Citation count prediction: learning to estimate future citations for literature," in *CIKM*, 2011, pp. 1247–1252.

[14] L. Li and H. Tong, "The child is father of the man: Foresee the success at the early stage," in *KDD*. ACM, 2015, pp. 655–664.

[15] L. Li, H. Tong, J. Tang, and W. Fan, "*iPath:* forecasting the pathway to impact," in *SDM*. SIAM, 2016, pp. 468–476.

[16] Y. Yao, H. Tong, F. Xu, and J. Lu, "Predicting long-term impact of cqa posts: a comprehensive viewpoint," in *KDD*. ACM, 2014, pp. 1496–1505.

[17] C. Chen, H. Tong, L. Xie, L. Ying, and Q. He, "Fascinate: Fast cross-layer dependency inference on multi-layered networks," in *KDD*, ser. KDD '16, 2016, pp. 765–774.

[18] B. R. Jasny and R. Stone, "Prediction and its limits," *Science*, vol. 355, no. 6324, pp. 468–469, 2017.

[19] L. Liu and E. Zhao, "Team performance and individual performance: Example from engineering consultancy company in china," in *2011 International Conference on Management and Service Science*, Aug 2011, pp. 1–4.

[20] L. Li, H. Tong, Y. Wang, C. Shi, N. Cao, and N. Buchler, "Is the whole greater than the sum of its parts?" in *KDD*, 2017, pp. 295–304.

[21] T. Lappas, K. Liu, and E. Terzi, "Finding a team of experts in social networks," in *KDD*, 2009, pp. 467–476.

[22] A. Anagnostopoulos, L. Becchetti, C. Castillo, A. Gionis, and S. Leonardi, "Online team formation in social networks," in *WWW*, 2012, pp. 839–848.

[23] S. S. Rangapuram, T. Bühler, and M. Hein, "Towards realistic team formation in social networks based on densest subgraphs," in *WWW*, 2013, pp. 1077–1088.

[24] B. Golshan and E. Terzi, "Minimizing tension in teams," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, ser. CIKM '17. New York, NY, USA: ACM, 2017, pp. 1707–1715.

[25] N. Cao, Y.-R. Lin, L. Li, and H. Tong, "g-Miner: Interactive visual group mining on multivariate graphs," in *CHI*, 2015.

[26] P. Bogdanov, B. Baumer, P. Basu, A. Bar-Noy, and A. K. Singh, "As strong as the weakest link: Mining diverse cliques in weighted graphs," in *ECML/PKDD (1)*, 2013, pp. 525–540.

[27] J. N. Cummings and S. B. Kiesler, "Who collaborates successfully?: prior experience reduces collaboration barriers in distributed interdisciplinary research," in *CSCW*, 2008, pp. 437–446.

[28] M. Muller, K. Ehrlich, T. Matthews, A. Perer, I. Ronen, and I. Guy, "Diversity among enterprise online communities: collaborating, teaming, and innovating through social media," in *CHI*, 2012, pp. 2815–2824.

[29] K. Chang and K. Ehrlich, "Out of sight but not out of mind?: Informal networks, communication and media use in global software teams," in *CASCON*, 2007, pp. 86–97.

[30] M. Cataldo and K. Ehrlich, "The impact of communication structure on new product development outcomes," in *CHI*, 2012, pp. 3081–3090.

[31] C. DeLong, L. G. Terveen, and J. Srivastava, "Teamskill and the nba: applying lessons from virtual worlds to the real-world," in *ASONAM*, 2013, pp. 156–161.

[32] C. DeLong and J. Srivastava, "Teamskill evolved: Mixed classification schemes for team-based multi-player games," in *PAKDD (1)*, 2012, pp. 26–37.

[33] R. Zadeh, A. D. Balakrishnan, S. B. Kiesler, and J. N. Cummings, "What's in a move?: normal disruption and a design challenge," in *CHI*, 2011, pp. 2897–2906.

[34] P. J. Hinds, K. M. Carley, D. Krackhardt, and D. Wholey, "Choosing work group members: Balancing similarity, competence, and familiarity," in *Organizational Behavior and Human Decision Processes*, 2000, pp. 226–251.

[35] L. Li, H. Tong, N. Cao, K. Ehrlich, Y. Lin, and N. Buchler, "Replacing the irreplaceable: Fast algorithms for team member recommendation," in *WWW*. ACM, 2015, pp. 636–646.

[36] ——, "Enhancing team composition in professional networks: Problem definitions and fast solutions," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 3, pp. 613–626, 2017.

[37] ——, "TEAMOPT: interactive team optimization in big networks," in *CIKM*, 2016, pp. 2485–2487.
[38] L. Li, H. Tong, and H. Liu, "Towards explainable networked prediction," in *CIKM*, 2018, pp. 1819–1822.
[39] Q. Zhou, L. Li, N. Cao, N. Buchler, and H. Tong, "Extra: explaining team recommendation in networks," in *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018*, 2018, pp. 492–493.