PRADA-TF: Privacy-Diversity-Aware Online Team Formation

Yash Mahajan

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Science

Jin-Hee Cho, Chair
Chang-Tien Lu
Terrence J. Moore

May 3rd, 2021
Blacksburg, Virginia

Keywords: Online Team Formation, Privacy, Diversity, Mechanism Design, Online Social Networks

PRADA-TF: Privacy-Diversity-Aware Online Team Formation

Yash Mahajan

(ABSTRACT)

In this work, we propose a PRivAcy-Diversity-Aware Team Formation framework, namely PRADA-TF, that can be deployed based on the trust relationships between users in online social networks (OSNs). Our proposed PRADA-TF is mainly designed to reflect team members' domain expertise and privacy preserving preferences when a task requires a wide range of diverse domain expertise for its successful completion. The proposed PRADA-TF aims to form a team for maximizing its productivity based on members' characteristics in their diversity, privacy preserving, and information sharing. We leveraged a game theory called *Mechanism Design* in order for a mechanism designer as a team leader to select team members that can maximize the team's social welfare, which is the sum of all team members' utilities considering team productivity, members' privacy preserving, and potential privacy loss caused by information sharing. To screen a set of candidate teams in the OSN, we built an expert social network based on real co-authorship datasets (i.e., Netscience) with 1,590 scientists, used the semi-synthetic datasets to construct a trust network based on a belief model called *Subjective Logic*, and identified trustworthy users as candidate team members. Via our extensive simulation experiments, we compared the seven different TF schemes, including our proposed and existing TF algorithms, and analyzed the key factors that can significantly impact the expected and actual social welfare, expected and actual potential privacy leakout, and team diversity of a selected team.

PRADA-TF: Privacy-Diversity-Aware Online Team Formation

Yash Mahajan

(GENERAL AUDIENCE ABSTRACT)

In this work, we propose a PRivAcy-Diversity-Aware Team Formation framework, namely PRADA-TF, that can be deployed based on the trust relationships between users in online social networks (OSNs). In professional work settings, often times we need to collaborate with other people to solve or complete a fairly complex problem or task. The task may commonly require creativity and/or high intelligence; but it is often given with a deadline. Our proposed PRADA-TF is mainly designed to reflect team members' domain expertise and privacy preserving preferences when a task requires a wide range of diverse domain expertise for its successful completion. The proposed PRADA-TF aims to form a team based on members' characteristics in their diversity, privacy preserving, and information sharing so as to maximize the performance of the team. We leveraged a game theory called *Mechanism Design* in order for a mechanism designer as a team leader to select team members that can maximize the sum of all team members' reward considering team productivity, members' privacy preserving, and potential privacy loss caused by information sharing. To screen a set of candidate teams in the OSN, we built an expert social network based on real co-authorship datasets with 1,590 scientists, used the semi-synthetic datasets to construct a trust network representing the trust relationship between the users in OSNs, and identified trustworthy users as candidate team members. Via our extensive simulation experiments, we compared the seven different team formation (TF) schemes, including our proposed and existing TF algorithms, and analyzed the key factors that can significantly impact the expected and

actual task rewards (utilities), expected and actual potential privacy leakout, and team diversity of a selected team.

Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisor, Dr. Jin-Hee Cho, for her guidance throughout this research. She has been really helpful all the way, guiding me patiently through each and every problem I faced, and nudging me in the right direction. I could not have made the progress I made, without her support and guidance, and I have nothing but exceptionally good things to say about her. I would also like to extend my thanks to Dr. Chang-Tien Lu for his time and encouragement for going through the crucial part of my thesis and providing some valuable inputs. I express my honest appreciation to Dr. Terrence J. Moore for his dedication and important comments for the revision of this document as well as inputs for future work directions.

Finally, I would like to thank my family and friends for all the moral support that they have provided over the course of this research.

Contents

List of Figures

List of Tables

List of Abbreviations

ED	Expertise Diversity
IF	Information Sharing
MD	Mechanism Design
OSN	Online Social Network
PPL	Potential Privacy Loss
SL	Subjective Logic
SW	Social Welfare
TD	Team Diversity
TF	Team Formation
VoI	Value of Information

Chapter 1

Introduction

These days we can ordinarily observe many examples of online team formation, such as in crowdsourcing systems, aiming to form a networked community of team members with relevant and diverse skill sets. Depending on the characteristics of a task, the criteria to select team members can differ aiming to maximize team productivity. This chapter discusses the motivation behind the research, followed by the research goal and questions, example scenario, problem statement, key contribution and the outline of this thesis.

1.1 Motivation

In professional work settings, often times we need to collaborate with other people to solve or complete a fairly complex problem or task. The task may commonly require creativity and/or high intelligence; but it is often given with a deadline. Much social science research has shown that high productivity or successful task completion is closely related to team composition in terms of the levels of relevant and diverse skills, team coherence, trust among members, information sharing, and/or shared mental model [1, 2, 3]. In the computer science domain, a team formation problem is known as NP-Hard [4]. How to form the team itself is another concern to consider particularly when we are pressured to form a team to complete dynamic tasks. Considering all these critical components of forming an optimal team composition, how to choose our collaborators or team members is critical to determining the successful completion of a given complicated task. Although there are many critical components to best form a team to complete a complicated task, we are interested in studying the effect of diversity and influence on online team composition and team performance where each team member's privacy is well maintained through out the process of completing the complicated task. The complicated task often requires highly novel, innovative, and solid ideas, understanding/organizing different kinds of knowledge in a coherent manner, and/or effective and efficient communication skills while maintaining a required level of work and communication integrity. In this work, we consider 'privacy' as the dimension of work and communication integrity.

Diversity is known as one of the key elements we need to consider to derive high-quality solutions particularly in the process of solving fairly complicated problems requiring critical thinking or domain expertise. Diverse thoughts or multidisciplinary approaches are known to be very creative and novel as well as to bring more productive and beneficial outcome, compared to the counterparts relying on homogeneous thoughts or single-discipline based approaches [5]. Scientifically and empirically many researchers have proven the positive effect of high diversity on team or organization productivity [6]. However, in our real world, most tasks are given with a deadline to complete a task. The time constraint may not fully utilize the maximum benefit of diversity because reaching a consensus in building a coherent idea among people with diverse background requires a sufficient amount of time, which is not allowed under the situation of completing a time-sensitive task. However, little work has explored the issues and/or tradeoffs between diversity and consensus for decision making under time pressure [7, 8].

An individual's privacy issue is a serious concern these days. Although a selected team may communicate via a secure channel, as more team members share information, there would be greater chances for some shared information to be leaked out to the outside world. Privacy loss minimization is studied in distributed constraint satisfaction (DCS) problems [9, 10], aiming to preserve perfect privacy without trusting anyone in interactions while minimizing interactions. However, in the context of team formation with a task requiring high expertise and diversity to achieve maximum team productivity, the privacy loss minimization techniques are not applicable. The relationship between privacy loss and team performance was studied in [11]. However, the authors examined what types of shared information can minimize privacy loss in a train traffic control task, which does not necessarily require different domain expertise among team members. However, preserving privacy in a context of team formation with a task requiring different types of domain expertise has not been studied in the literature.

1.2 Research Goal & Questions

In this work, we aim to develop a PRivAcy-Diversity-Aware Team Formation framework, namely PRADA-TF, that can be deployed based on trust relationships between users in online social networks. The proposed PRADA-TF is designed to (i) reflect team members' domain expertise and privacy preservation preferences when a task requires a wide range of diverse domain expertise to be successfully completed; and (ii) maximize team productivity based on team diversity, privacy, and information sharing with acceptable computational overhead.

To achieve this goal, we will answer the following **key research question** in this study:

RQ1 *What is the relationship between team performance and team members' privacy preserving preferences?*

RQ2 *What are the effects of team diversity on the team performance?*

RQ3 *How do the trust relationships in users of a given online social network affect team performance?*

RQ4 *How does the team behave in an actual real world scenario?*

RQ5 *How does the team size affect the social welfare of the team?*

1.3 Example Scenario

We assume a scenario such that a team needs to be formed in online platforms in order to perform fairly complicated tasks. An example can be like an online crowdsourcing system, such as Amazon Mechanical Turk, where a requestor wants to form a team to execute a given task. We assume that the task requires fairly diverse skills and knowledge and collaboration across team members for a successful completion of the task. Thus, selecting qualified team members is the key to the team performance and the success of the given task.

A team will be formed based on team members who have diverse domain expertise (or knowledge) where each member has a certain level of preference in preserving his or her privacy. A member's privacy preference will affect communication patterns among team members in which an amount of information naturally leads to high team productivity. In an online social network (OSN), a team leader is a user aiming to form a team to achieve a certain task and can reach out to his or her friends or friends of the friends to gather promising candidates of team members. We assume that each user as a trustor can estimate trust in his/her friends, trustees, based on domain expertise or willingness to share information, which is available to a trustor through direct or indirect experiences. Through the chain of trust relationships between users, the team lead can gather a set of promising candidates of team members and select a set of team members based on certain criteria. We described the details of how to calculate a user's trust in another user, and how the team leader collects a set of promising candidate team members and accordingly selects the final team members in Section 4.4.

1.4 Problem Statement

In this work, we leverage the *mechanism design* [12] as a game theoretic solution where the MD, as a team leader, aims to identify an effective team composition based on diverse domain expertise of team members and effective communications allowing to share quality information. On the other hand, each player, as a potential team member, has a certain level of his/her privacy preference when working with other members because exposing privacy to some extent may be inevitable for collaborative teamwork. However, at the same time, sharing more information with lower privacy preserving preferences may lead to higher potential privacy leakout.

The tradeoff issue between information sharing and privacy preservation is well-known in the TF problem [13, 14]. The MD wants all players to reveal their truthful types in the expertise and privacy preference to make a best decision to select qualified team members that can maximize team performance while maximally preserving their privacy preferences. Given a player i's utility, u_i formulated to maximize his/her contribution to the team performance and privacy preservation while minimizing potential privacy loss, the MD will aim to achieve:

$$\arg\max_{x \in X} \sum_{i \in \mathcal{T}} u_i(x, \hat{\theta}_i, \hat{\theta}_{-i} | \theta_i), \quad \forall \theta_i \in \Theta_i \tag{1.1}$$

where θ_i is player i's true preferences in privacy preservation and expertise level where Θ_i refers to a set of preferences. The x is a particular team composition decision belonging to a set of candidate team composition decisions, X. $\hat{\theta}_i$ is player i's revealed preference to the MD and $\hat{\theta}_{-i}$ refers to revealed preferences of all other players, except player i. The $\mathcal{T}$ is a set of team members i's chosen by decision x. The $u_i(x, \hat{\theta}_i, \hat{\theta}_{-i}|\theta_i)$ is detailed in Eq. (4.2).

1.5 Key Contributions

In this work, we made the following **key contributions**:

1. We developed the PRADA-TF to identify a set of team members forming a team that can maximize social welfare of the selected team based on the concept of *mechanism design* [15]. To the best of our knowledge, this work is the first that considered both diversity and privacy of team members to solve a TF problem, given a task requiring diverse domain expertise for its successful completion.

2. We selected a set of candidate team members based on the trust relationships between users which are estimated by a belief model called *Subjective Logic*. Few studies have considered the prior trust relationships between team members in the team formation process on the OSN environment along with applying the mechanism design in game theory for team formation.

3. Unlike the existing team formation studies mainly focusing on the 'expected' team performance, we investigated the social welfare of a selected team in both expected and actual social welfare where candidate team members' behaviors are modeled based on their privacy preferences revealed in the team formation stage and their actual privacy preferences used in the task execution stage.

4. To reflect a realistic scenario of the online expert social network used in this work, we created a semi-synthetic dataset in order to build an expert social network estab-

lished based on the *Netscience* [16]. The *Netscience* contains a coauthorship network of 1,590 scientists working on Network Theory and Experiments compiled from the bibliographies of two review articles [17, 18] on networks.

5. We conducted extensive experiments to evaluate the performance of the proposed PRADA-TF in terms of expected and actual social welfare, expected and actual potential privacy leakout, and team diversity of a selected team. We conducted the performance analysis of seven different schemes that determine the selection of candidate teams where four schemes are the variants of the proposed PRADA-TF, two are the state-of-the-art counterparts, and one is a baseline model. From this study, we showed that our proposed PRADA-TF outperformed the exiting counterpart and baseline TF algorithms overall. In particular, overall selecting members based on the utility, estimated based on team performance, privacy preserving, and potential privacy loss, outperformed other schemes. In addition, compromising more privacy showed less or no performance improvement in social welfare while there exists a certain level of team diversity (i.e., not too high or too low) that can lead the team to high performance (i.e., high social welfare).

1.6 Outline

This thesis is structured as follows:

- The next chapter discusses research papers and case studies related to collaborative team formation and the key factors that impact the team performance.

- The Preliminary chapter details all the prerequisites for the proposed privacy-diversity-aware team formation framework.

- The Design chapter describes the mechanism design based approach for team formation. It details on how the trust network is formed and how candidates are selected from this network using the 7 different candidate team selection (CTS) methods.

- The Experimental Setup chapter, introduces all the metrics used for comparing the performance of the 7 different CTS methods. Along with it, it also describes the dataset used and its preprocessing along with the Parameterization.

- The Results section discusses the extensive simulations done to compare the performance of the proposed framework, along with the different effects observed.

- The conclusions chapter briefly mentions the overview of the entire process and discusses the results obtained. It also summarizes the rest of this document.

Chapter 2

Background

Privacy-Diversity-Aware Online Team Formation focuses on two components. First one being the team formation process, where given N individuals, the goal to form the most effective and efficient team ($\leq N$), given the task. The second one is exploring and analysing the impact of key factors such as privacy and diversity, on the performance of the team. This chapter discusses the related literature on the components mentioned above.

2.1 Collaborative Team Formation

Lappas et al. [4] first created a team formation (TF) problem to identify a subset of individuals in a social network based on their expertise and communication cost incurred among team members. They proved that the TF problem is NP-Hard and validated their proposed algorithm based on the DBLP dataset. Li et al. [19] also solved the same TF problem similar to [4]. However, the authors added additional design features to consider required skill sets and accordingly a required number of experts for each skill set to develop a general team formation framework. Kargar and An [20] solved the same TF problem [4], but considered additional cost factors, including the communication cost between team members and between team members and a leader. Anagnostopoulos et al. [21] proposed a solution for the TF problem by considering three factors: a sufficient level of skill sets, low communication cost, and fair workload among team members.

Bhowmik et al. [22] formulated the TF problem as an unconstrained submodular function maximization problem in which a team can be formed by leveraging skill cover softening, efficient team communication, and relaxation of connectivity. Gajewar and Sarma [23] solved a TF problem based on the density of a selected subgraph of a social network consisting of experts. Since the TF problem assumes that low communication cost will incur over a densely connected network, high network density is treated as desirable to increase the team compatibility. However, this may not be true because a highly connected individual node can be overloaded due to a large volume of requests from its neighbors and may fail delivering an assigned work. Datta et al. [24] proposed a cost-effective TF algorithm which meets requirements of skill sets and whose team members are socially close in order to reduce communication cost. However, being socially close between members can be effective when a given task does not require highly novel ideas. Although high homophily can make team communications easier, it may not necessarily contribute to deriving novel, innovative ideas due to the nature of similarity in the ideas/thoughts. Basiri et al. [25] tackled the same TF problem but used a meta-heuristic algorithm, called BRADO (BRAin Drain Optimization) [26], which is a type of swarm algorithms. Wang et al. [27] conducted a comprehensive performance comparison of the major TF algorithms based on the proposed benchmark for fair comparison. Wang et al. [28] took a game theoretic approach and modeled each worker in crowdsourcing as a selfish entity which does not necessarily cooperate to the request to join a social crowdsourcing team.

Based on the literature review above, we found that the diversity of team composition and its impact (i.e., both positive and negative aspects) on team performance has not been sufficiently addressed, given a fairly complicated, time-sensitive task requiring highly novel, creative, innovative, and solid ideas. In addition, while a node's high connectivity with other nodes are treated as a desirable aspect to reduce communication cost, there is lack of understanding of the adverse effect of working with highly influential people who may not be able to contribute to the team due to their limited capabilities in time or effort.

For our research problem, we aim to devise a mechanism that can maximise the team performance while preserving the agents' privacy preferences, in a team formation problem. Not much prior research has considered team formation as a mechanism design problem. Wright and Vorobeychik [29] present the first formal mechanism design framework for team formation and present four mechanisms, of which two are novel and two are extensions of known mechanisms. They define the team formation problem as a hedonic game with N sets of players and a tuple $\succ$ which defines each players preference over the set of players, in forming a team.

2.2 Key Factors Impacting Team Performance

2.2.1 Diversity vs. Consensus

Diversity refers to identity-based differences amongst two or more people and more than one objective characteristic of a group. Diversity is a subjective phenomenon, created by the group members based on the dissimilarity (or similarity) of social identities [30]. Three levels of diversities exists in workplace: surface-level, deep-level and hidden. *Surface-level diversity* refers to an individual's visible characteristics, such as age, race, sex and visible disability. *Deep-level diversity* refers to non-observable and cognitive traits of an individual, such as attitude, knowledge, and abilities. *Hidden diversity* includes traits disclosed or concealed at the individual's discretion (deep-level), such as sexual orientation or hidden disability.

Ancona and Caldwell [31] examined the impact of variations in functional diversity (deep-level) and organizational tenure on team performance and process in product development teams. They found that greater functional diversity is associated with more external communication but both functional diversity and tenure have a negative impact on the team performance due to multiple reasons, such as conflicts arising due to differences in perspectives. Bechtoldt et al. [30] considered diversity in team members' personality traits and how they affect the team creativity and performance. The findings are: Based on the Big-Five Personality Inventory which includes extraversion, agreeableness, openness, conscientiousness, and neuroticism, heterogeneity in extraversion and agreeableness and homogeneity in conscientiousness are preferred to enhance team performance.

After reviewing 40 years of literature on demography and diversity, Phillips and O'Reilly [32] concluded that there isn't any consistent relationship between diversity and organisational performance and instead proposed that mediating variables might exist between diversity and performance. In the meta-analysis to provide a relationship between team diversity and team outcomes, Horwitz and Horwitz [33] found that there exists a positive impact of task-related diversity with the team performance, in both quantity and quality. On the other hand, there exists no relationship between bio-demographic diversity with the team outcome. Pieterse et al. [34] explored the impact of cultural diversity on team performance and found: cultural diversity has a positive effect on team productivity with teams focused on developing knowledge and increasing competence (high learning approach orientation) and teams who are focused on avoiding loss in terms of knowledge (low performance avoidance orientation). One important factor observed from the experiment is that every team members' goal orientation has a direct impact on how the diverse teams can profit from their diversity, by elaborating on their enhanced pool of knowledge.

Using conflict as a mediating variable, Liang et al. [35] explored the effect of team diversity on software project performance by considering the three group composition types, which are *knowledge diversity* (KD), - differences among team members in education, technical knowledge and perspectives, *value diversity* (VD), - differences in how members perceive what the task, goal and outcomes should be and *social category diversity* (SD), - differences in demo-graphic characteristic, against task conflict and relationship conflict. The authors found that KD has positive effect on the team performance because high KD increases the task conflict (disagreement regarding the content of the task) among team members while decreasing the relationship conflict (interpersonal incompatibilities) which makes it more likely to produce better team performances. On the other hand, VD negatively affects the team performance as it increases relationship conflict. SD has a mixed effect on the team performance because it positively influences both task conflict and relationship conflict, thereby affecting them both in opposing ways.

Cohen and Yashinski [36] showed that finding an optimal diverse team of people is an NP-Complete problem. They proposed two algorithms: *Fixed Parameter Algorithm (FPT)* which iterates over instantiations of an abstract template in increasing order until it finds a concrete template that can be satisfied by the candidates for the team; and *GreedyDiverse* which uses the greedy technique to iteratively select a skill and candidate that improves the group the most. Their experiments showed that FPT performs better than GreedyDiverse when there is a skew in data.

Cognitive consensus refers to similarity of team members' in perceiving, defining and conceptualizing key issues. Extremely high diversity and consensus in collective representations are viewed as unfit and dysfunctional for a lot of situations and therefore a balance between is consensus and dissensus is required [8]. However this optimal level of consensus that can positively influence the team performance, depends on a lot of factors, including the types of individuals involved and the nature of the task. Knight et al. [7] studied the impact of demographic diversity in Top Management Teams (TMT) of 83 high-technology firms and found that demographic diversity has negatively related to the consensus of the team.

2.2.2 Information Sharing vs. Privacy Preserving

The relationship between the extent of information and team performance has been significantly investigated. Many studies proved that information sharing is a clear driving force leading to high team performance and success [13, 14, 37]. Although information sharing has been studied as the key positive determinant affecting team performance, the related adverse impact on information sharing via close interactions among team members has not been sufficiently explored.

Privacy loss or minimization issues have been studied in distributed constraint satisfaction problems [9, 10]. In DCS problems, the distributed negotiation or cooperation is studied while preserving perfect privacy by not trusting anyone in interactions. Both works [9, 10] showed a tradeoff between privacy and efficiency. Unlike the DCS problems, team work requires more continuous and close interactions that can directly maximize team performance. Privacy loss is inevitable. Harbers et al. [11] studied the trade-off between privacy loss and team performance in the train traffic control domain. They investigated what type of and how often information should be shared among team members. Based on their empirical experiments, sharing affective load information is the most favourable for better performance and minimal privacy loss in most conditions. On the other hand, sharing information about total load led to the most privacy loss. Based on the literature review above, team formation problems with privacy preserving have not been sufficiently studied.

ϵ-differential privacy was introduced by Dwork [38] with an intuition that a person's privacy cannot be compromised with the release of their data if the data is not present in the database. It basically provides each individual with the same privacy that would result from having their data removed from the database. Differential privacy offers strong guarantee against adversaries due to its composability, robustness to post processing, and graceful degradation when there is correlated data present. Kasiviswanathan and Smith [39] provides the formulation of the guarantees in terms of the inferences drawn by a Bayesian adversary which is satisfied by ϵ-differential privacy and even by its relaxation. Differential privacy works on the notion that the result of any function on a database, is not overly dependent on one individual's data. McSherry and Talwar [40] extended differential privacy and gave game theoretic guarantees, including approximate truthfulness, collusion resistance, and repeatable play. Using the mechanism, they ensure that each participant has very selected effect on the outcome of the mechanism, which in-turn provide very limited incentive to lie. Nissim et al. [41] and Xiao et al. [42] argue that external incentives are necessary for individuals to participate and report truthfully. Xiao et al. [42] introduced a transformation that transforms a truthful mechanism into a deferentially private mechanism that remains truthful based on the ideas for privately releasing histogram data. They advocated directly incorporating privacy into the player's utility and developing a mechanism that takes into account the combined utility of a player.

Chapter 3

Preliminaries

All the preliminaries for the proposed PRADA-TF are defined, in this chapter. This chapter covers the formal definition of our Task model, Information model and the Adversarial model in detail along with its different components.

3.1 Task Model

We consider the following key components of a complex task to be given to a prospective team:

- *Number of team members (m)*: A given team consists of the m number of members.

- *A set of required domain expertise (E)*: The successful completion of a given task requires that a given team has expertise in a set of domain knowledge to perform the task, denoted by $E = \{e_1, e_2, \ldots, \ldots, e_l\}$, where $l \leq m$. For the ease of referring to what expertise domain is required, we maintain a vector $\mathbf{L}$ to access an element of each expertise domain, such as $\mathbf{L}(e_i)$ for domain i where $\mathbf{L}(e_i)$ returns the extent of knowledge required in expertise domain e_i in E as a nonnegative real number where the sum of expert is set to a given constant, ϵ (i.e., $\sum_{e_i \in E} L(e_i) = \epsilon$).

3.2 Information Model

When a player shares information in a team, the information may have a different extent of contributing to the successful task completion. We call it *Value-of-Information (VoI)* in our work where VoI refers to how valuable given information is to support a given task based on the following criteria:

- *Credibility* (crd_{ih}) represents the extent of credibility for given information provided based on player i's expertise in domain h, measured as a real number in $[0, 1]$.
- *Usefulness* (uf_{ih}) refers to the extent of usefulness (or relevance) for given information based on player i's expertise in domain h, measured as a real number in $[0, 1]$.
- *Novelty* (nov_{ih}) indicates the extent of the novelty in given information based on player i's expertise in domain h, measured as a real number in $[0, 1]$.

We simply formulate VoI of given information by player i as:

$$\mathrm{VoI}_{ih} = w_{\mathrm{crd}} \cdot \mathrm{crd}_{ih} + w_{\mathrm{uf}} \cdot \mathrm{uf}_{ih} + w_{\mathrm{nov}} \cdot \mathrm{nov}_{ih}, \tag{3.1}$$

where each weight is ranged in $[0, 1]$ as a real number with $w_{\mathrm{crd}}+w_{\mathrm{uf}}+w_{\mathrm{nov}} = 1$ and represents how much each component of VoI_{ih} is weighed. If player i executes a given task requiring expertise $\mathbf{L}$, then crd_{ih}, uf_{ih}, and nov_{ih} are computed based on player i's expertise in domain h by:

$$\mathrm{crd}_{ih} = \theta^e_{ih}, \quad \mathrm{uf}_{ih} = \min\left[1, \frac{\theta^e_{ih}}{\mathbf{L}(e_h)}\right], \tag{3.2}$$

$$\mathrm{nov}_{ih} = \frac{\sum_{j \in \mathcal{T}, j \neq i} \max\left[0, \left(\sqrt{\theta^e_{ih}} - \sqrt{\theta^e_{jh}}\right)\right]}{|\mathcal{T}|},$$

where θ^e_{ih} is player i's truthful revelation of expertise level in domain h ranged in $[0, 1]$ as a real number and $\mathcal{T}$ is a set of other players j's in a given team composition with player i. Each dimension of VoI implies as follows: crd_{ih} refers to $\theta^e_i(h)$ representing player i's actual expertise quality in domain h; uf_{ih} means how much player i's expertise in domain h contributes to the teamwork based on the required extent of expertise in domain h, $\mathbf{L}(e_h)$; and nov_{ih} indicates the extent of player i's contribution to the required expertise in domain h compared to other team members j's contribution.

3.3 Adversarial Model

We consider possible private information leakout by team members based on their level of distrust. In this work, we consider an expert social network consisting of experts with various backgrounds based on *Netscience* [16], which is described in Section 5.2.1. We also used a belief model called *Subjective Logic* in order to derive the MD's trust in each user j (i.e., expert) in the expert social network, denoted by P^{MD}_j (see Eq. (4.4)). We simply use $1 - P^{MD}_j$ as the MD's distrust in each player j, which is discussed in detail in Section 4.4.1. Given a candidate team chosen by the MD, we formulate the extent of member i's private information that can be potentially leaked out to outside of the team (i.e., unauthorized parties) by other team members j's by:

$$pl_i = \exp\left(-\frac{\lambda}{\left(\sum_{h \in E}(1-\hat{\theta}^p_{ih})\right)\left(1-\prod_{j \in M, i \neq j} P^{MD}_j\right)}\right), \tag{3.3}$$

where pl_i refers to the extent of possible *privacy loss* for team member i. The λ is a constant to adjust the scale depending on the number of domain expertise used (i.e., $\lambda = |E|$). The $\sum_{h \in E}(1-\hat{\theta}^p_{ih})$ is the sum of member i's shared information with the team based on the revealed privacy preference of player i (since the MD only knows the revealed privacy preferences by players) in the given domains (i.e., E). The $\left(1-\prod_{j \in M, i \neq j} P^{MD}_j\right)$ refers to the probability that any one of other team members j's leak out i's shared information with the team to outside of the team (i.e., unauthorized parties). Note that when $(\sum_{h\in E}(1-\hat{\theta}^{p}_{ih}))(1-\prod_{j\in M, i\neq j} P^{MD}_{j}))$ is zero (i.e., i fully shares information without any privacy preference or all other members are fully trusted with zero distrust), pl_i returns zero, representing no chance of leaking out private information. Note that it is assumed the MD is trustworthy and does not leak any private information since the MD's objective is to maximize the team performance which is estimated by the social welfare.

Chapter 4

Key Design Features of the Proposed PRADA-TF

In this chapter, we describe our proposed PRADA-TF that uses a game theoretic approach using the mechanism design [15]. We describe a player's type, payoff computation, preference revelation, team selection process, and actual behavior modeled in this work.

4.1 Mechanism Design for Team Formation

We use the mechanism design to solve a TF problem. A set of players, $N = \{1, 2, \cdots, n\}$, participate in a team formation where a set of team choice $x \in X$ is given with $X = \{x_1, x_2, \ldots, x_n\}$. Each player i has a truthful private signal (i.e., type) $\theta_i \in \Theta_i$, representing preferences over outcomes. A set of truthful private signals by all players is denoted by $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$, which describes the profile of all truthful types for the n players. The state θ is selected randomly from the state space $\Theta \equiv \Theta_1 \times \Theta_2, \ldots, \Theta_n$, representing the set of all possible profiles of types $\theta \in \Theta$. The MD aims to select x, a set of team members, to form a team based on the members' preferences, θ's (i.e., the decision rule by $x(\theta)$) to maximize the sum of the payoffs of all players, $\sum_{i \in N} u_i(x, \hat{\theta}_i, \hat{\theta}_{-i} | \theta_i)$, where $u_i(x, \hat{\theta}_i, \hat{\theta}_{-i} | \theta_i)$ refers to player i's utility when the MD selects x when player i's revealed (or announced) preference type is $\hat{\theta}_i$, other players j's revealed preference types are denoted by $\hat{\theta}_{-i}$, and θ_i

is player i's truthful preference. $u_i(x, \hat{\theta}_i, \hat{\theta}_{-i}|\theta_i)$ is given in Eq. (4.2).

4.1.1 Player's Types

Each player i has its own type θ_i with two private signals, the degrees of domain expertise and privacy preserving preferences (i.e., $\theta_i = \{\theta_i^e, \theta_i^p\}$) by:

- *Expertise profile* (θ_i^e): We assume that each player has a set of values representing expertise in $|M|$ knowledge domains where player i's expert domains and their strengths are indicated in a vector θ_i^e with o elements. θ_{ih}^e, $h = 1, \ldots, o$ is a real number in $[0, 1]$, such that $\sum_{h \in M} \theta_{ih}^e \leq |M|$ where M is a set of domain expertise whose subset is E as $E \subseteq M$. Recall that E is a set of expertise domains considered in a given task, implying $\sum_{h \in E} \theta_{ih}^e \leq |E|$.
- *Privacy preference* (θ_i^p): Each player i has a different level of privacy preserving preference when sharing information, denoted by θ_i^p. Higher θ_i^p means player i is less willing to share information to minimize privacy exposure. Note that θ_i^p refers to a vector of player i's truthful privacy preference in sharing information in domain h, denoted by θ_{ih}^p, and is set as a real number in $[0, 1]$ with $\theta_i^p = \langle \theta_{i1}^p, \theta_{i2}^p, \ldots, \theta_{io}^p \rangle$ where player i has expertise in o number of domains.

We assume that a player cannot lie about expertise type, which will be given based on objective criteria (e.g., publications, degrees, years of experience). However, the player may lie about his/her privacy type because information sharing behavior is constrained by one's privacy preference, which is not known without direct experience. Hence, when $\theta_{ih}^e = \hat{\theta}_{ih}^e$, it implies that players reveal truthful privacy preferences while they reveal truthful expertise preferences as default.

4.2 Player's Payoff

Player i's payoff is estimated by:

$$u_i(x, \hat{\theta}_i, \hat{\theta}_{-i} | \theta_i) = u_i^{\text{team}}(x, \hat{\theta}_i, \hat{\theta}_{-i} | \theta_i) + u_i^{\text{priv}}(x, \hat{\theta}_i, \hat{\theta}_{-i} | \theta_i) - pl_i,$$

where $u_i^{\text{team}}(x, \hat{\theta}_i, \hat{\theta}_{-i} | \theta_i)$ is the expected team performance when the MD decides to choose team x where $\hat{\theta}_i$ is player i's revealed type, $\hat{\theta}_{-i}$ is other players $-i$'s revealed types, and θ_i is player i's truthful type. The pl_i refers to the loss caused by user i's private information leakout as shown in Eq. (3.3).

We measure $u_i^{\text{team}}(x, \hat{\theta}_i, \hat{\theta}_{-i} | \theta_i)$ based on how much credible, useful, and novel information (i.e., VoI) can be influenced by player i's expertise and privacy preferences and is obtained by:

$$u_i^{\text{team}}(x, \hat{\theta}_i, \hat{\theta}_{-i} | \theta_i) = \sum_{h \in E} VoI_{ih} \cdot (1 - \hat{\theta}_{ih}^p)^2 \tag{4.1}$$

This models the decrease of novelty when more information is shared as discussed in [43, 44]. We estimate $u_i^{\text{priv}}(x, \hat{\theta}_i, \hat{\theta}_{-i} | \theta_i)$ to reflect how much an individual's privacy preference is preserved. This privacy-related utility is obtained by

$$u_i^{\text{priv}}(x, \hat{\theta}_i, \hat{\theta}_{-i} | \theta_i) = \sum_{h \in E} (1 - VoI_{ih}) \cdot (\hat{\theta}_{ih}^p)^2, \tag{4.2}$$

reflecting that less sharing in less valuable information preserves player i's privacy as well as introduces little adverse impact on team performance.

4.3 A Player's Preference Revelation

Unlike prior TF research, we additionally validate the quality of TF algorithms when a task is actually executed by the selected team. Players do not have to reveal their truthful preferences, $\theta_i = (\theta_i^e, \theta_i^p)$, where $\theta_i^e = \{\theta_{i1}^e, \theta_{i2}^e, \ldots, \theta_{in}^e\}$ and $\theta_i^p = \{\theta_{i1}^p, \theta_{i2}^p, \ldots, \theta_{in}^p\}$. We denote player i's revealed preferences by $\hat{\theta}_i = (\hat{\theta}_i^e, \hat{\theta}_i^p)$. A player's actual behavior is modeled based on whether the player reveals his/her truthful privacy type considering the following two cases:

- *Case 1:* $\theta_i^p == \hat{\theta}_i^p$ where player i's revealed privacy type is the same as his/her truthful type. In this case, the MD can make an accurate decision based on truthful information.

- *Case 2:* $\theta_i^p \neq \hat{\theta}_i^p$ where player i's revealed type is not the same as his/her truthful type. In this case, player i can consider whether to compromise the truthful privacy type based on the estimated utility. Player i's exhibited privacy preference level in an actual task execution, denoted by $\theta_i^{p'}$, is determined by:

$$\theta_i^{p'} = \begin{cases} \hat{\theta}_i^p \text{ if } u_i(x^*, \hat{\theta}_i, \hat{\theta}_{-i} | \theta_i) \ > \ u_i(x, \theta_i, \hat{\theta}_{-i} | \theta_i), \\ \theta_i^p \ \text{ otherwise.} \end{cases} \tag{4.3}$$

  Here $u_i(x^*, \hat{\theta}_i, \hat{\theta}_{-i} | \theta_i)$ is the payoff when x^* decision is taken by the MD where $\hat{\theta}_i$ is the revealed type of the player i, $\hat{\theta}_{-i}$ is other players' revealed types, and θ_i is player i's truthful type. Thus, a player will compromise the privacy if and only if the payoff by compromising the truthful privacy using the revealed privacy brings a better payoff than using the truthful privacy preference.

We consider pc_i as a real number in $[0, 1]$ to indicate how much player i can compromise its revealed type, $\hat{\theta}_i^p$, which can be selected as a real number in $[pc_i \cdot \theta_i^p, \theta_i^p]$. Since pc_i determines the lower bound of the range a player compromises his/her privacy preference, higher pc_i means a player can compromise less while lower pc_i means the player can compromise more. Note that when player i compromises his/her privacy preference, it implies that the player announces a lower privacy preference than the truthful privacy preference in order to increase its benefit from the contribution to the team performance.

4.4 Team Selection Process

An MD is a team leader aiming to form a team in an OSN. The MD will recruit candidate team members from the MD's ego network, which is defined as a social network consisting of all users (or players) within k-hop distances from the MD. Higher k will result in considering a larger number of member candidates, and vice-versa. Among all users within the k-hop distances from the MD, the MD will select a set of team members via two rounds: (1) Select a set of member candidates from the users within the k-hop distances based on the trust relationships between users. We call this k-hop network *the MD's k-hop trust network*. From the users in the MD's k-hop trust network, the MD will select the top ϕ number of member candidates; and (2) From the top ϕ number of member candidates selected based on the MD's k-hop trust network, the MD runs the social welfare function in Eq. (5.5) to select a final set of m team members. For the process of (1), the MD needs to build the k-hop trust network and select top ζ member candidates from the k-hop trust work based on their trust values, as described below.

4.4.1 Building k-Hop Trust Network

We build a k-hop trust network (e.g., an expert social network) in order for the MD to select a set of team member candidates through the trust relationships between users in the network. We model the MD's trust in users A's who are directly connected to the MD, called *direct trust* in A, based on truthful expertise and privacy preferences in A. This implies that if two users have direct experiences, they know each other's truthful types in expertise and privacy preferences. However, if A is not directly connected to the MD, then the MD needs to estimate A's trust through the trust relationships with other users who directly know A. There should be trust decay as the chain of the trust path becomes longer. In addition, since the MD may reach A through multiple paths within the k-hop trust network, there will be multiple trust values obtained from the multiple paths. The MD needs a method to combine these multiple trust values into a single trust value. In order to discount trust and combine multiple trust values into the single trust value, we leverage the *discounting* operator and the *consensus* operator in Subjective Logic (SL) [45].

In this work, we estimate the MD's trust in each user in a given trust network (e.g., an expert social network) based on a binomial opinion (i.e., trust or distrust) offered by SL [45]. The binomial opinion $\omega_B^A = (b_B^A, d_B^A, u_B^A, a_B^A)$ where parameters b_B^A, d_B^A and u_B^A denote the degree to which A trusts, distrusts or is uncertain about the trustworthiness of B in the current instance, respectively, where $b_B^A + d_B^A + u_B^A = 1$. Additionally, a_B^A is a base rate probability A would assign to B *a priori*. This base rate can be interpreted as A's prior belief or preference to B. For simplicity, we consider the base rate to be equal for each belief mass, i.e., $a_B^A = 1/2$ for b_B^A and $1 - a_B^A = 1/2$ for d_B^A. The projected probability of A's trust in B is given by:

$$P_B^A = b_B^A + a_B^A u_B^A. \tag{4.4}$$

Similarly, A's distrust in B is obtained by $d_B^A + a_B^A u_B^A$, which is the same as $1 - P_B^A$, as $b_B^A + d_B^A + u_B^A = 1$ and $a_B^A = 1/2$. Since we need to obtain the MD's trust in each user in the given social network, the MD's trust can be obtained through a chain of trust relationships where the target user (i.e., a user the MD wants to obtain trust) is distant from the MD. For example, when A trusts B, B trusts C, and C trusts D, then we want to obtain A's trust in D where A is the MD. We obtain the so called referral trust via the *discounting operator* below.

The *discounting operator* is used to obtain indirect trust by increasing the uncertainty in the expectation value (see Eq. (4.4)). Assume three agents A, B and C, where A has referral trust in B by $\omega_B^A = (b_B^A, d_B^A, u_B^A, a_B^A)$ and B has functional trust in C represented by $\omega_C^B = (b_C^B, d_C^B, u_C^B, a_C^B)$. The indirect functional trust of A in C can be obtained by discounting B's trust in C by A's trust in B. This is given by $\omega_C^{A:B} = \omega_B^A \otimes \omega_C^B$ where $\otimes$ is the discounting operator and $\omega_c^{A:B} = (b_C^{A:B}, d_C^{A:B}, u_C^{A:B}, a_C^{A:B})$ with

$$\begin{aligned} b_C^{A:B} = b_B^A b_C^B, \;\; d_C^{A:B} = b_B^A d_C^B \\ u_C^{A:B} = d_B^A + u_B^A + b_B^A u_C^B, \;\; a_C^{A:B} = a_C^B. \end{aligned} \tag{4.5}$$

When the MD obtains multiple trust values from users who directly interact with the target user via direct interactions, the MD needs to derive the agreed trust based on the multiple trust values. In that case, we use the below *consensus operator*.

The *consensus operator* is used to obtain trust by combining two beliefs into one with reduced uncertainty in the expectation value. Assuming A's trust in C to be $\omega_C^A = (b_C^A, d_C^A, u_C^A, a_C^A)$ and B's trust in C to be $\omega_C^B = (b_C^B, d_C^B, u_C^B, a_C^B)$, the consensus between ω_C^A and ω_C^B is denoted by $\omega_C^{A\oplus B} = (b_C^{A\oplus B}, d_C^{A\oplus B}, u_C^{A\oplus B}, a_C^{A\oplus B})$ with

$$b_C^{A\oplus B} = \frac{b_C^A u_c^B + b_C^B u_C^A}{\beta}, \;\; d_C^{A\oplus B} = \frac{d_C^A u_c^B + d_C^B u_C^A}{\beta} \qquad (4.6)$$
$$u_C^{A\oplus B} = \frac{u_C^A u_c^B}{\beta}, \;\; a_C^{A\oplus B} = a_C^A$$

where $\beta = u_C^A + u_C^B - u_C^A u_B^A$. With the help of the discounting and consensus operators above, the MD's opinion in A, ω_A^{MD}, can be obtained to derive the MD's trust based on Eq. (4.4). We ensure using an independent path from the MD to the target user where no users appear in multiple paths. Finally, the MD can rank the trust of all users in the k-hop trust network by using the expected trust based on Eq. (4.4) and select top ϕ candidate team members.

For simplicity, we initialize each user's trust value based on both expertise preference and willingness to share information based on privacy preference. For example, the MD's belief in trusting B via direct experience in terms of whether B will be relevant for the given task requiring E set of expertise domains is formulated by:

$$b_B^{MD} = \frac{\sum_{h\in E}\Big(w_e\theta_h^e + w_s(1-\theta_h^p)\Big)}{|E|}, \qquad (4.7)$$

where $w_e + w_s = 1$. Assuming with a fairly small uncertainty u_B^{MD} (e.g., $K/(N+K)$ where $K = 2$ which is commonly assumed for a bionomial opinion and N is sufficiently large), we simply derive $d_B^{MD} = 1 - (b_B^{MD} + u_B^{MD})$ based on the requirement of additivity with $b_B^{MD} + d_B^{MD} + u_B^{MD} = 1$.

4.5 Construction of Candidates Network using Real Datasets

Using a densely connected social network, where the MD can reach most of the users using *k-hops*, trust values are calculated for all the members in the OSN platform, as described in Section 4.4. From these trust values, the top ϕ players are selected and then used in the candidate team selection schemes, described as below.

4.5.1 Selection of Candidate Teams

After the top ϕ players are selected based on the MD's k-hop trust network, the MD further cuts down prospective team members by applying the different heuristic candidate team selection (CTS) methods to avoid high complexity. Hence, our proposed PRADA-TF scheme can have the following variants:

- *Utility-based Serial Dictatorship (USD)* selects top ζ number of candidates out of ϕ number of players using a player's utility function in Eq. (4.2).

- *Expertise diversity-based CTS (ED-CTS)* cuts down top ζ number of candidates out of ϕ number of players using Eq. (5.2), similar to [46], selecting the candidates based on the diversity of a player's expertise contributing to the required expertise of a given task.

- *VoI-based CTS (VoI-CTS)* selects top ζ number of candidates out of ϕ number of players using VoI in Eq. (3.1).

- *Information Sharing (IF-CTS)* selects top ζ number of candidates out of ϕ number of players using a player's revealed privacy type, $\hat{\theta}_i^p$.

Using ζ number of candidate members, the MD selects the top m players to form the final team which maximizes the social welfare among all possible subsets of a team with m out of ζ candidates. That is, given m is sufficiently small, the MD considers all possible team composition x's and select a team that maximizes the social welfare based on Eq. (5.5). Note that social welfare calculated in the team selection process is the expected social welfare based on the reported preference type and is not indicative of the result of actual task execution. In this work, we also demonstrate the actual social welfare based on the actual preferences used by team members in a given team composition x.

Chapter 5

Experimental Setup

In this chapter, we describe the performance metrics, datasets, environmental setup, and comparing schemes used for the comparative performance analysis of our proposed PRADA-TF with the existing counterparts.

5.1 Metrics

We use the following metrics to evaluate the performance of the TF algorithms considered in this work:

- *Team Diversity ($\mathcal{TD}$)*: This metric refers to the extent of diversity in team members' expertise background. $\mathcal{TD}$ is measured by:

$$\mathcal{TD} = \frac{\sum_{i \in \mathcal{T}} \mathcal{H}_i}{|\mathcal{T}|}, \tag{5.1}$$

where $\mathcal{T}$ refers to a set of members in a selected team and $\mathcal{H}_i$ refers to the extent of team member i's uniqueness compared to other members' expertise types. $\mathcal{H}_i$ is estimated based on the Hellinger distance [47] by:

$$\mathcal{H}_i = \frac{\sum_{j \in \mathcal{T}, j \neq i} \mathcal{H}(\theta_i^e, \theta_j^e)}{|\mathcal{T}| - 1}, \tag{5.2}$$

where the difference between agent i's background and agent j's background, $\mathcal{H}(\theta_i^e, \theta_j^e)$, for given team $\mathcal{T}$ and $i, j \in \mathcal{T}$, is computed by:

$$\mathcal{H}(\theta_i^e, \theta_j^e) = \sqrt{\sum_{h \in \mathcal{E}} D_{ij}^e}, \tag{5.3}$$

$$\text{where} \;\; D_{ij}^e = \frac{\sum_{h \in E} \max\left[0, \theta_{ih}^e - \theta_{jh}^e\right]}{|E|}. \tag{5.4}$$

The θ_{ih}^e and θ_{jh}^e are the vectors of truthful expertise types of players i and j in domain h.

- *Social Welfare* ($\mathcal{SW}_{\mathcal{T}}$): This refers to a team's expected social welfare estimated based on Eq. (4.2) and is given by:

$$\mathcal{SW}_{\mathcal{T}} = \sum_{i \in \mathcal{T}} u_i(x, \hat{\theta}_i, \hat{\theta}_{-i} | \theta_i). \tag{5.5}$$

Note that the actual SW is estimated by replacing revealed preferences, $\hat{\theta}_i$, with exhibited preferences, θ_i', at the execution time. Therefore, in the experimental results, we show both expected and actual SW.

- *Potential Privacy Leakout ($\mathcal{PPL}$)*: This metric refers to the amount of penalty a player may have because of potential privacy leakout by other players. We use pl_i in Eq. (3.3) to measure this metric. We demonstrate the expected and actual PPL where the expected PPL is estimated by a player's revealed privacy preference, $\hat{\theta}_i^p$, while the actual PPL uses the privacy level actually used by a player at task execution.

5.2 Experimental Setup

5.2.1 Datasets

Although deriving the expertise of a participant is relatively straightforward based on objective verifiable criteria, it is highly challenging to obtain a user's privacy preference in OSN platforms. Therefore, we developed a semi-synthetic dataset by leveraging the *Netscience* [16] dataset which contains a coauthorship network of 1,590 scientists working on Network Theory and Experiments compiled from the bibliographies of two review articles [17, 18] on networks. To derive the expertise of each author in the network, we used the Scopus API to extract publication records and metrics for the top 3 subject areas (according to All Science Journal Classification (ASJC)) each author has publication in. Subject areas from the ASJC are then consolidated further into 5 broader fields: 'Biology and Biochemistry', 'Sciences', 'Arts and Social Sciences', 'Engineering' and 'Multidisciplinary.' From these subject areas, using the corresponding publication record and citations, a weighted expertise level is calculated for all the authors in the network. Now to convert this sparse network into a small world network, and make authors more reachable from the MD, additional edges are added based on the cosine similarity (> 0.9) in expertise level and the subject areas. The processed *Netscience* data generated a network with $1,269$ nodes, $28,072$ edges, and 5 subject-area specific communities. Finally, each author's privacy preference, θ_i^p is drawn from a Gaussian distribution with mean $\mu = 0.5$ and standard deviation $\sigma = 0.3$, and is in the range $[0, 1]$

5.2.2 Parameterization

We consider all the 1,269 authors participating in a given TF problem. We selected an author with a highest betweenness to be an MD playing as a team leader. With the MD as

Table 5.1: Key Parameters, Meanings, and Default Values

Param.	**Meaning**	**Value**
θ^{e}_{ik}	Strength in expertise in domain k	[0, 1]
θ^{p}_{ik}	Privacy preserving preference in domain k	[0, 1]
$\hat{\theta}^{p}_{ik}$	Revealed privacy preserving preference in domain k	[0, 1]
pc_i	Extent to which a player can lie about its privacy preserving preference	[0, 1]
w_e, w_s	Weights for expertise and privacy privacy preserving respectively	[0, 1]
ϕ	Number of candidates selected from the trust network	200
ζ	Number of participants selected using CTS schemes	40
m	Number of team members	20
$\lvert E\rvert$	Number of expertise domains	5
$w_{\mathrm{crd}}, w_{\mathrm{uf}}, w_{\mathrm{nov}}$	Weights for the three components of VoI	$[0, 1]$
ϵ	Sum of domain expertise levels required by a given task (i.e., $\sum_{e_i \in E} e_i = \epsilon$)	5
$\mathbf{L}$	A vector of domain expertise levels in a given task, $\{e_1, e_2, \ldots, e_l\}$	$[1, 1, 1, 1, 1]$
λ	A constant to scale pl_i in Eq. (3.3)	$\lvert E\rvert$

the center of the network, the MD's ego network is created with users who are reachable via 5-hop ($k = 5$) distances from the MD. The MD's trust values in the users of the MD's trust network are calculated by constructing a trust network, as described in Section 4.4. The MD's trust values in users within the MD's trust network are recursively calculated. This allows to consider all independent paths from the MD to each user within the MD's trust network. This implies that high-degree users tend to have higher trust values because using a consensus operator more to combine trust values from multiple paths can increase estimated trust, as discussed earlier. The weights for VoI_{ih} (i.e., $w_{\text{crd}}, w_{\text{uf}}$, and w_{nov} in Eq. (3.1)) are set to 0.2, 0.3, and 0.5, respectively. w_e and w_s in Eq. (4.7) are equally weighted with $w_e = w_s = 0.5$. Top 200 ($= \phi$) players with the highest trust values are selected from which 40 ($= \zeta$) players are selected based on a given candidate team selection method described in 'Selection of Candidate Teams' of Section 4.4. Finally from the 40 players shortlisted, 20 ($= m$) players that maximize the social welfare are selected to form the final team to execute the task. We summarized the default values of the key design parameters for our experiments in Table 5.1. All results are collected based on the mean values from 1,000 simulation runs and shown with the standard deviation at each data point.

5.3 Comparing Schemes

The variants of the proposed PRADA-TF scheme (i.e., USD, ED-CTS, VoI-CTS, and IF-CTS; see Section 4.5.1) are compared against the following two existing counterparts and one baseline model in terms of the metrics in Section 5.1:

- *Homophily-based CTS (H-CTS)* selects top ζ number of candidates based on the degree of players' homophily in terms of their expertise required in a given task, similar to [48], based on a cosine-similarity metric.
- *Centrality-based CTS (C-CTS)* [49] selects top ζ number of candidates based on players' betweenness to identify the influential members for the team.
- *Random* selects ζ number of candidates at random.

Chapter 6

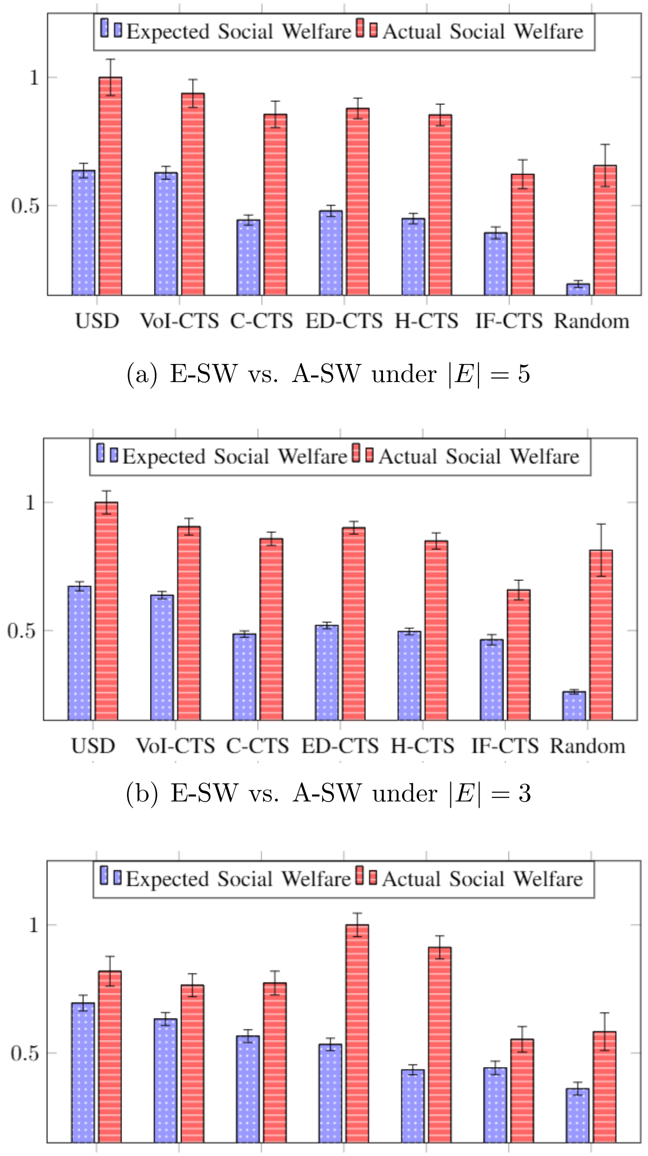
Numerical Results & Analysis

Having described the team formation process, and the environmental setup for the simulations, in this chapter we analyse the results and discuss the key trends observed and inferred from these results. Five metrics are considered, including Expected Social Welfare (E-SW), Actual Social Welfare (A-SW), Expected Potential Privacy Loss (E-PPL), Actual Potential Privacy Loss (A-PPL) and Team Diversity ($\mathcal{TD}$), and the (1) effect of different task types, (2) effect of compromising privacy, (3) effect of different team size and (4) effect of varying k-hops, is compared against these five metrics.

6.1 Effect of Different Task Types

6.1.1 Effect of Different Task Types on Social Welfare

Figs. 6.1(a)-6.1(c) show the performance comparison of 7 different candidate team selection (CTS) methods based on Expected Social Welfare (E-SW) and Actual Social Welfare (A-SW). In those figures, we observed that A-SW is likely to be higher than E-SW because team members are more likely to compromise their privacy preferences, aiming to increasing information sharing and accordingly higher utility. Fig. 6.1(a) demonstrates the performance of the above mentioned CTS methods for a task requiring fairly diverse skill-sets (i.e., $L(e_i) = [1, 1, 1, 1, 1]$ with $|E| = 5$). From the figure, we observe the performance order in terms of



(a) E-SW vs. A-SW under $|E| = 5$

(b) E-SW vs. A-SW under $|E| = 3$

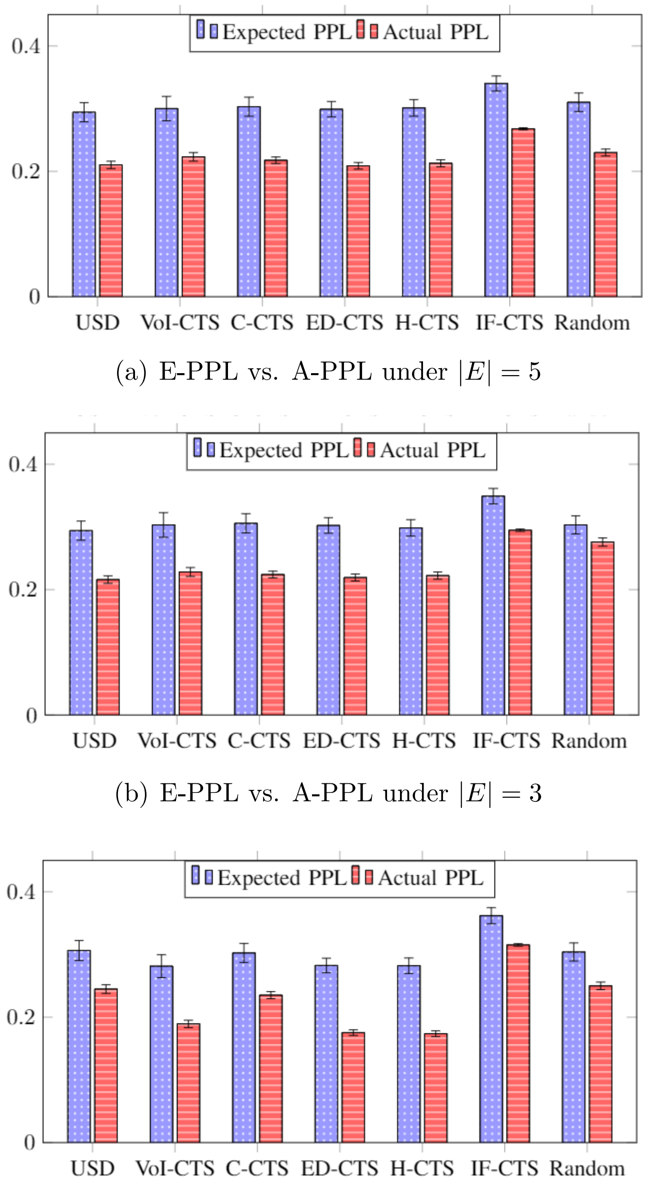
(c) E-SW vs. A-SW under $|E| = 1$

Figure 6.1: Performance comparison of different candidate team selection (CTS) methods based on expected social welfare (ESW) and actual social welfare (ASW), when the number of domains varies with $|E| = 5, 3$ or 1 and the corresponding task composition, $\mathbf{L}(e_i) = [1, 1, 1, 1, 1], [5/3, 5/3, 5/3]$, or [5], respectively. (a) is under $|E| = 5$, (b) is under $|E| = 3$, and (c) is under $|E| = 1$. Note that the lower bound weight of a revealed privacy preference (pc_i) is set to $= 0.8$, the number of hops in an online trust network (k-hop) is set to 5, and the error bar represents the standard deviation.

A-SW as: USD > VoI-CTS > C-CTS $\simeq$ ED-CTS $\simeq$ H-CTS > Random > IF-CTS. This is in accordance to the trend observed in E-SW. However, when C-CTS, ED-CTS, and H-CTS are compared against Random, we notice that players selected using the Random scheme tend to sacrifice their privacy less to balance the privacy loss suffered. In Fig. 6.1(b) using $|E| = 3$ with $L(e_i) = [5/3, 5/3, 5/3]$, we observe a similar trend to Fig. 6.1(a). USD performs the best while IF-CTS does not perform well. The performance of the other CTS methods is comparable. The performance of ED-CTS is as good as VoI-CTS, indicating that as the required task for completion becomes more subject-area specific while a team formed based on expertise diversity tends to perform better and contributed more. This can be further confirmed in Fig. 6.1(c), which delineates a task requiring just one subject-area expertise $|E| = 1$ and the strength in expertise required is represented by $L(e_i) = [5]$. Interestingly, contrary to the results from the previous task compositions (i.e., $|E| = 5$ and 3), ED-CTS tends to perform the best followed by Random and then H-CTS. Although E-SW observed for USD and VoI-CTS is the highest, in an actual task execution, the compromise in privacy exhibited by the team formed using ED-CTS, H-CTS and even Random is higher than that of the former. This is because PPL is estimated based on other team members' distrust which may be higher only when a set of candidate team members is selected based on the single expertise where users with less adjacent users due to lack of similarity with other users are less likely to be linked with other users and accordingly this can reduce the player's trust lower.

6.1.2 Effect of Different Task Types on Potential Privacy Leakout

Figs. 6.2(a)-6.2(c) demonstrate the effect of different task types on Potential Privacy Leakout (PPL). We observe that the actual PPL (A-PPL) is lower than the expected PPL (E-PPL) because E-PPL is calculated using the revealed type of the team members whereas in an



(a) E-PPL vs. A-PPL under $|E| = 5$

(b) E-PPL vs. A-PPL under $|E| = 3$

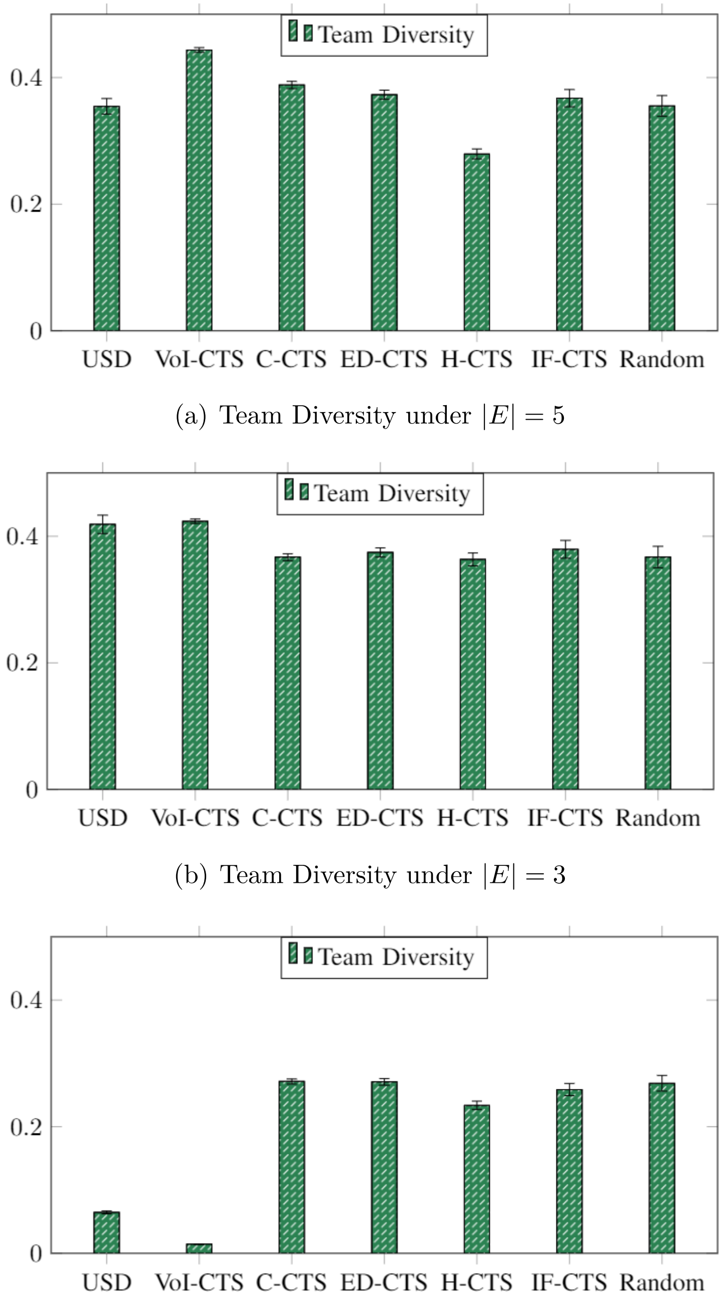
(c) E-PPL vs. A-PPL under $|E| = 1$

Figure 6.2: Performance comparison of different candidate team selection (CTS) methods based on expected potential privacy leakout (E-PPL) and actual potential privacy leakout (A-PPL) when the number of domains varies with $|E| = 5, 3$ or 1 and the corresponding task composition, $\mathbf{L}(e_i) = [1, 1, 1, 1, 1], [5/3, 5/3, 5/3]$, or [5], respectively. (a) is under $|E| = 5$, (b) is under $|E| = 3$, and (c) is under $|E| = 1$. Note that the lower bound weight of a revealed privacy preference (pc_i) is set to $= 0.8$, the number of hops in an online trust network (k-hop) is set to 5, and the error bar represents the standard deviation.

actual task execution, a player might either stick to its revealed type or revert back to its true type, which would be higher, leading to less information shared and consequently less privacy loss. From Fig. 6.2(a), we observe that A-PPL is the least for USD whereas it is the highest for IF-CTS. That is, there clearly exists a trade-off between information sharing and privacy. Sharing more information regardless of its value (VoI) will naturally lead to higher privacy loss. As observed in Figs. 6.1(a)-6.1(c), information sharing (as shown in IF-CTS) is not a driving factor to maximize SW because of its high PPL. On the other hand, all the C-CTS, ED-CTS, H-CTS and Random show similar levels of PPL indicating that all players act similarly in a selfish manner to protect their privacy irrespective of the scheme used. In Fig. 6.2(b), we observe a similar trend wherein IF-CTS has the highest A-PPL whereas all the rest of the CTS methods have comparable A-PPL values. Notice that this directly affects the SW (see Fig. 6.1(b)). In addition, in Fig. 6.2(c), the performance order in terms of A-PPL is observed as: ED-CTS $\simeq$ Random $>$ H-CTS $>$ C-CTS $>$ IF-CTS $>$ USD $>$ VoI-CTS. This is particularly interesting because from Fig. 6.1(c), we can see that ED-CTS and Random followed by H-CTS perform the best. From these comparison, we can conclude that there exists an inverse relationship between privacy loss and team performance estimated by SW.

6.1.3 Effect of Different Task Types on Team Diversity

Figs. 6.3(a)-6.3(c) show the diversity of the team formed using the 7 different candidate team selections methods under three different types of tasks. Under tasks requiring a fairly diverse expertise (e.g., $|E| = 5$ or 3), we can clearly observe a fairly high diversity under highly performing schemes (e.g., see USD or VoI-CTS in Figs. 6.3(a) and 6.3(b)). For example, in Fig 6.3(b), team diversity is a standout factor that leads to higher utility, which can be confirmed by looking at USD and VoI-CTS. Additionally, analyzing Fig. 6.3(c), we can



(a) Team Diversity under $|E| = 5$

(b) Team Diversity under $|E| = 3$

(c) Team Diversity under $|E| = 1$

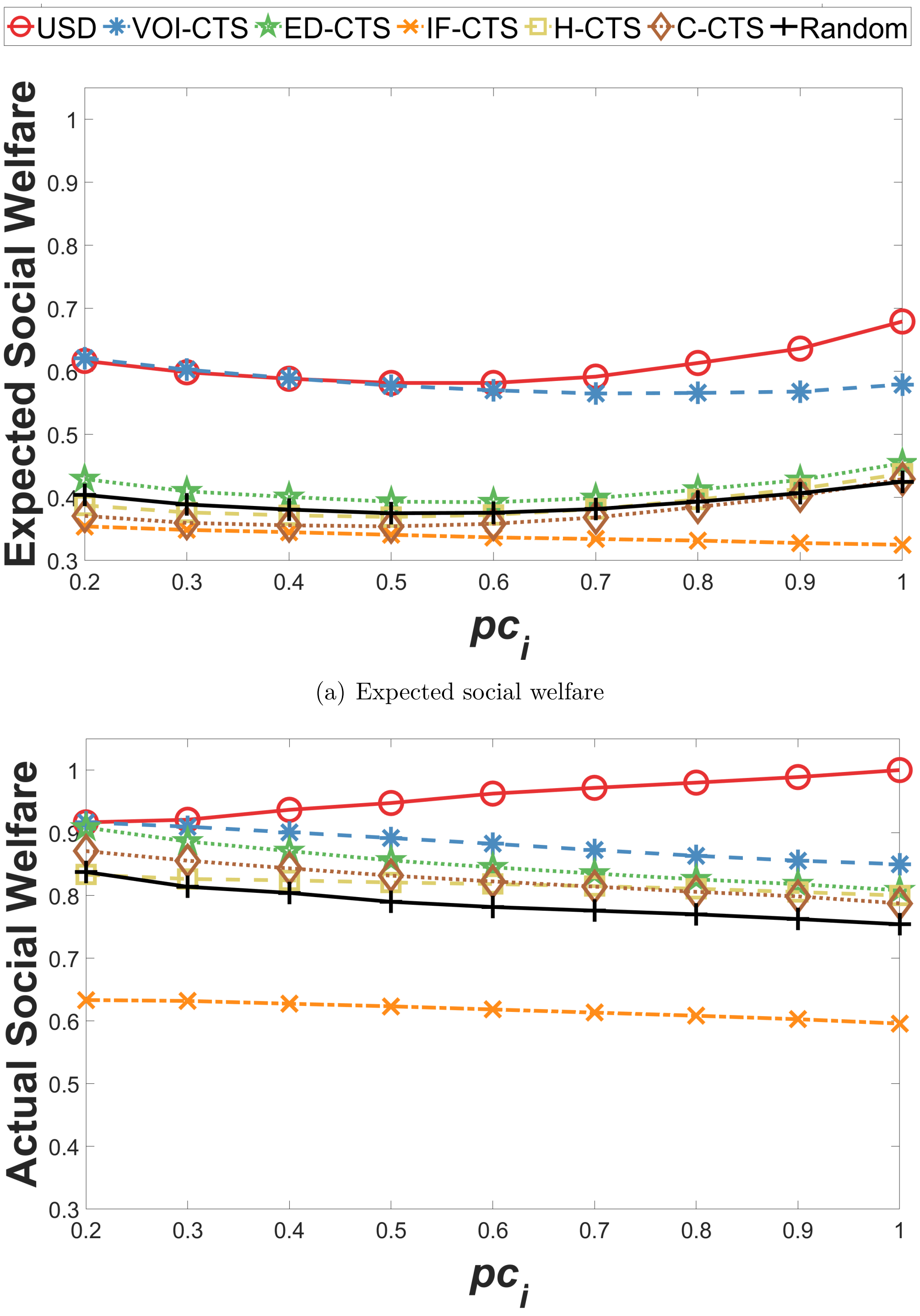
Figure 6.3: Performance comparison of different candidate team selection (CTS) methods based on team diversity, when the number of domains varies with $|E| = 5, 3$ or 1 and the corresponding task composition, $\mathbf{L}(e_i) = [1, 1, 1, 1, 1], [5/3, 5/3, 5/3]$, or $[5]$, respectively. (a) is under $|E| = 5$, (b) is under $|E| = 3$, and (c) is under $|E| = 1$. Note that the lower bound weight of a revealed privacy preference (pc_i) is set to $= 0.8$, the number of hops in an online trust network (k-hop) is set to 5, and the error bar represents the standard deviation.

notice that high diversity is aligned with high SW. From this observation, we can say that team diversity is important when the task requires subject area specific expertise. However, under a task requiring a single domain expertise, it is unclear that having a certain level of diversity is closely related to high team performance (i.e., high SW), as shown in Fig. 6.3(c).

6.2 Effect of Compromising Privacy

6.2.1 Effect of Different Task Types on Social Welfare

Fig. 6.4 shows the effect of varying the lower bound of compromising a team member's privacy preference (pc_i) under the 7 different CTS methods in terms of the five metrics. Similar to Fig. 6.3, USD performs the best followed by VoI-CTS amongst all the E-SW and A-SW, as shown in Figs. 6.4(a) and 6.4(b). The overall trend observed from Fig. 6.4(a) is that as pc_i increases (revealing more truthful privacy preferences), E-SW decreases when pc_i ranges from 0.2 to 0.5 whereas it increases when pc_i ranges from 0.5 to 1. That is, when $pc_i = [0.2, 0.5]$, the utility achieved by compromising the privacy is less than the privacy loss suffered as E-SW decreases for $pc_i = [0.5, 1]$. Contrary to the expected trend, overall A-SW decreases as pc_i increases except for USD which increases as pc_i increases. Although A-SW is always higher than E-SW, fewer players choose to compromise their privacy because the utility gained by compromising privacy in comparison to the privacy lost is less when pc_i increases and the players' revealed privacy types become closer to their true types. Since, in USD, players are selected based on the utility function, as when the extent of dishonesty decreases with higher pc_i, E-SW and A-SW increase.



(a) Expected social welfare

(b) Actual social welfare

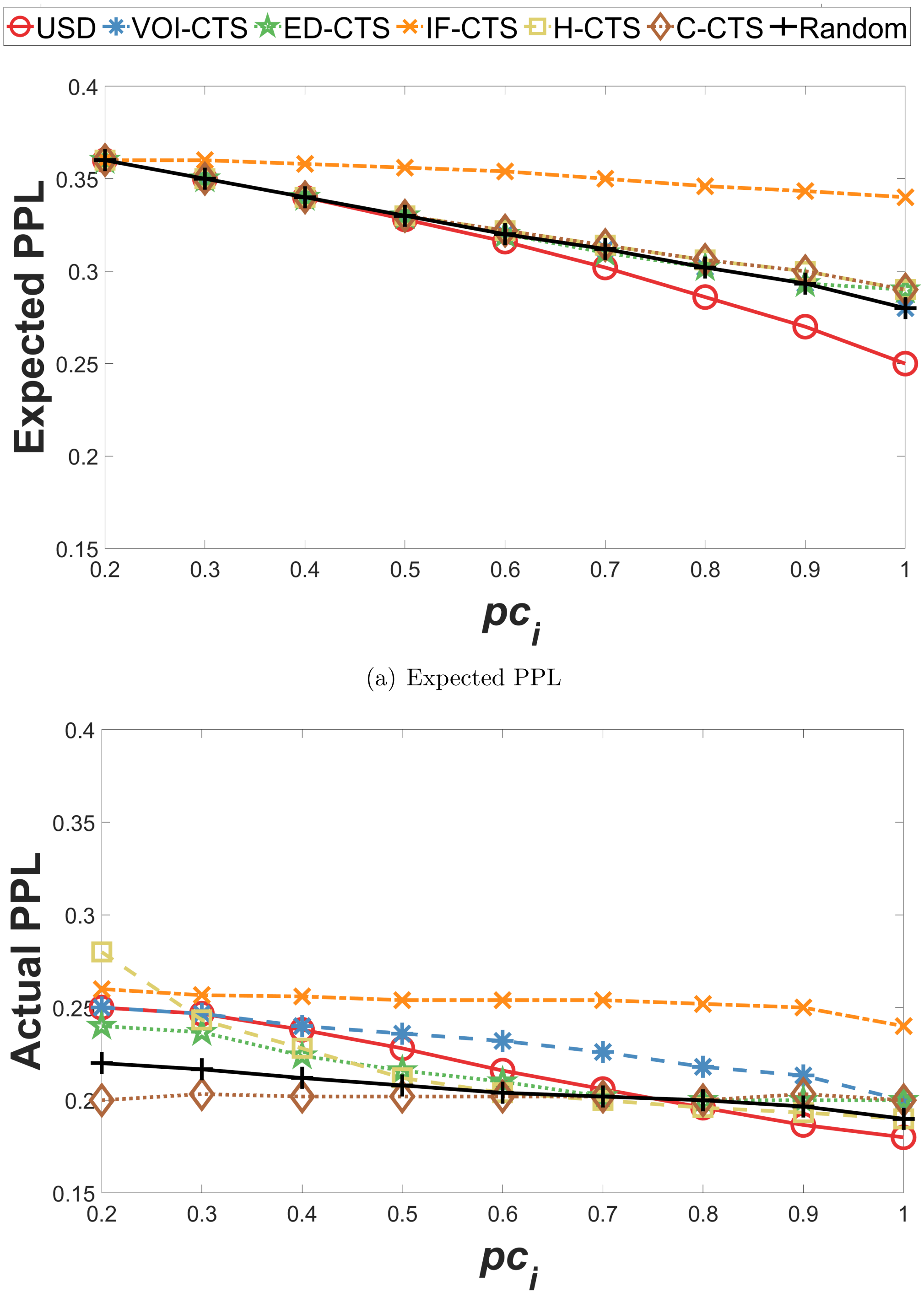
Figure 6.4: Comparison of different candidate team selection (CTS) methods based on the expected social welfare and actual social welfare under varying the lower bound weight of a revealed privacy preference (pc_i), where the environment is set as the number of domains ($|E|$) = 5 and task composition ($\mathbf{L}(e_i) = [1, 1, 1, 1, 1]$) under a 5-hop trust network (i.e., $k = 5$).

6.2.2 Effect of Different Task Types on Potential Privacy Leakout and Team Diversity

In Fig. 6.5(a), we can see that the E-PPL decreases relatively sharply for USD whereas IF-CTS has the highest value across all values of pc_i. Overall we observed that the E-PPL decreases as pc_i increases. In Fig. 6.5(b), we notice that except for USD, which follows the general trend, the rest of all the CTS methods remain unchanged (e.g., C-CTS) or have a slight decrease in their value of A-PPL as pc_i increases. The decreased A-PPL is because as pc_i increases, players start reporting higher privacy preference (less information sharing but revealing more truthful information) which leads to less potential privacy loss. In addition, in an actual task execution, a player might revert to his/her true privacy preference, which would further increase privacy preservation. Fig. 6.6 shows the trend in diversity as pc_i varies. We can view that VoI-CTS has the highest increase in the team diversity as pc_i increases whereas for USD the team diversity decreases. For all the rest of the method, the team diversity follows an almost zero incline. Additionally, compared to Fig. 6.4(a), we can see that A-SW increases as the team diversity increases for USD whereas the A-SW decreases when the team diversity increases for VoI-CTS. This confirms that the team diversity has an inverse relationship with ASW for this particular task type.

6.3 Effect of Different Team Sizes

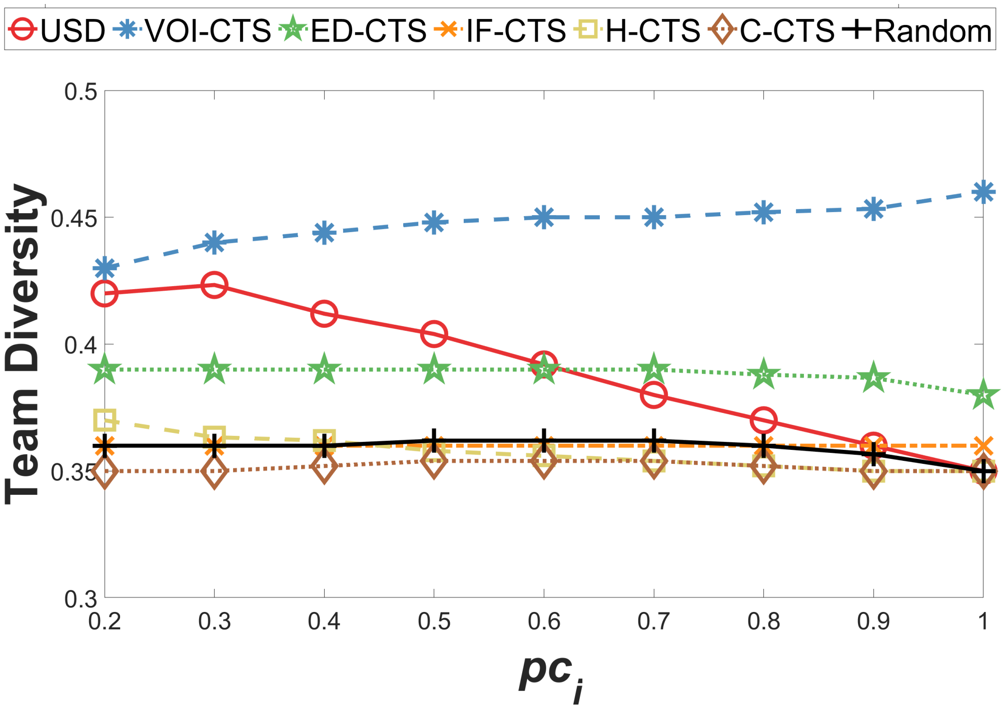
Figs. 6.7, 6.8 & 6.9, show how different team sizes affect the performance of different CTS methods in terms of the five metrics used in this work. Comparing Figs. 6.7(a) & 6.7(b) against Figs. 6.8(a) & 6.8(b), we can infer three key observations. First, as the team size increases, PPL increases, which consequently decreases the utility received by the player.



(a) Expected PPL

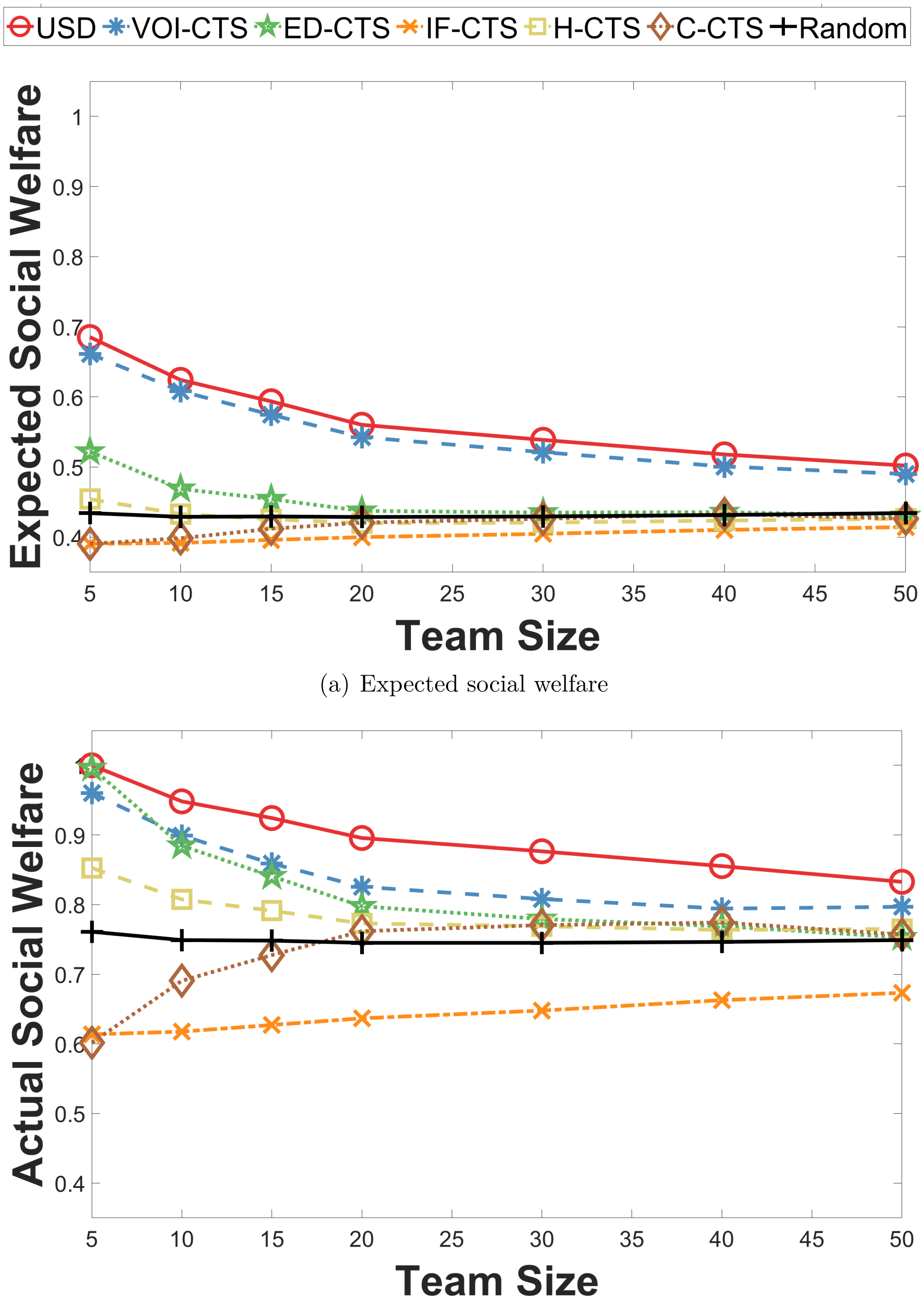
(b) Actual PPL

Figure 6.5: Comparison of different candidate team selection (CTS) methods based on expected potential privacy loss and actual potential privacy loss under varying the lower bound weight of a revealed privacy preference (pc_i), where the environment is set as the number of domains ($|E|$) = 5 and task composition ($\mathbf{L}(e_i) = [1, 1, 1, 1, 1]$) under a 5-hop trust network (i.e., $k = 5$).



(a) Team diversity

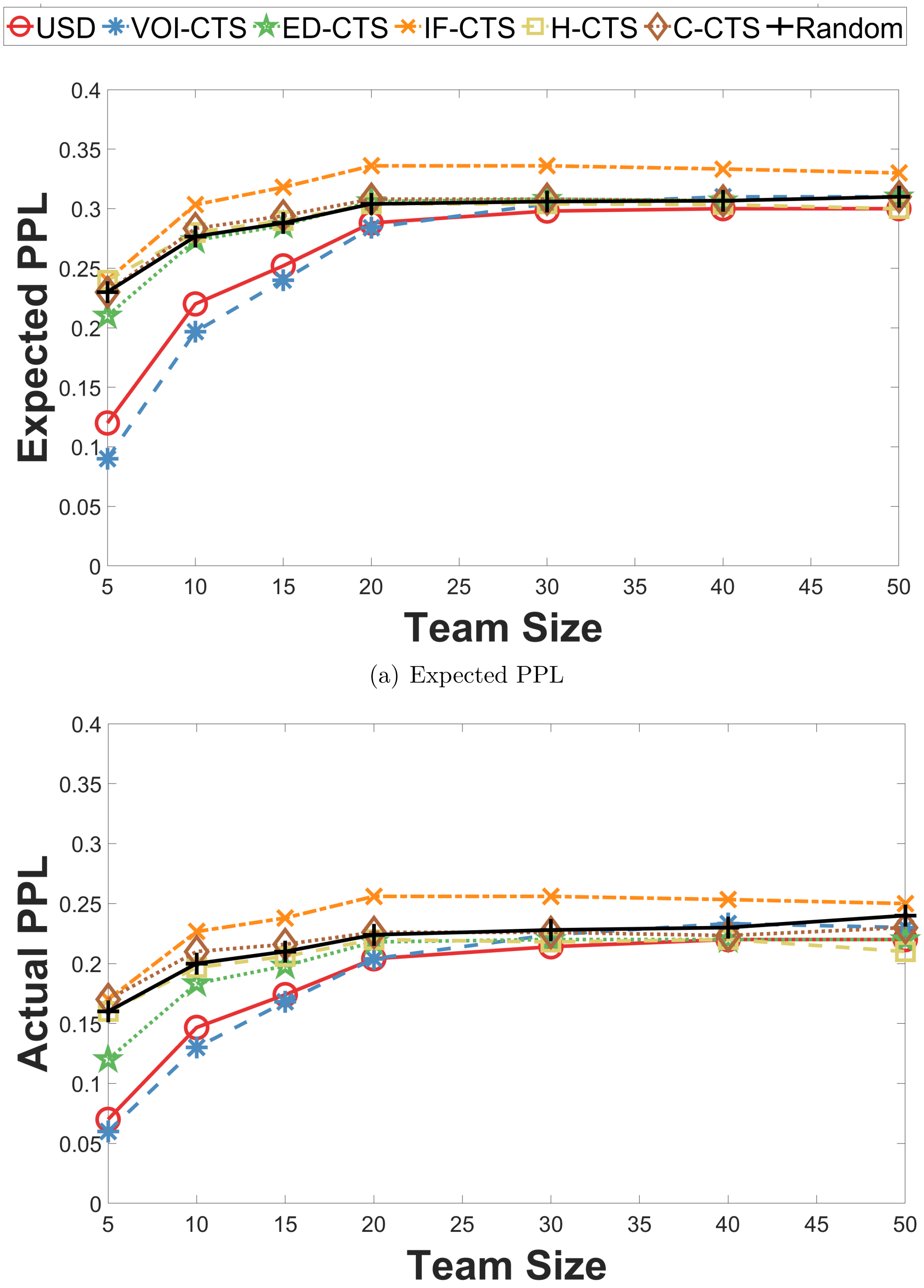
Figure 6.6: Comparison of different candidate team selection (CTS) methods based on team diversity under varying the lower bound weight of a revealed privacy preference (pc_i), where the environment is set as the number of domains ($|E|$) = 5 and task composition ($\mathbf{L}(e_i) = [1, 1, 1, 1, 1]$) under a 5-hop trust network (i.e., $k = 5$).



(a) Expected social welfare

(b) Actual social welfare

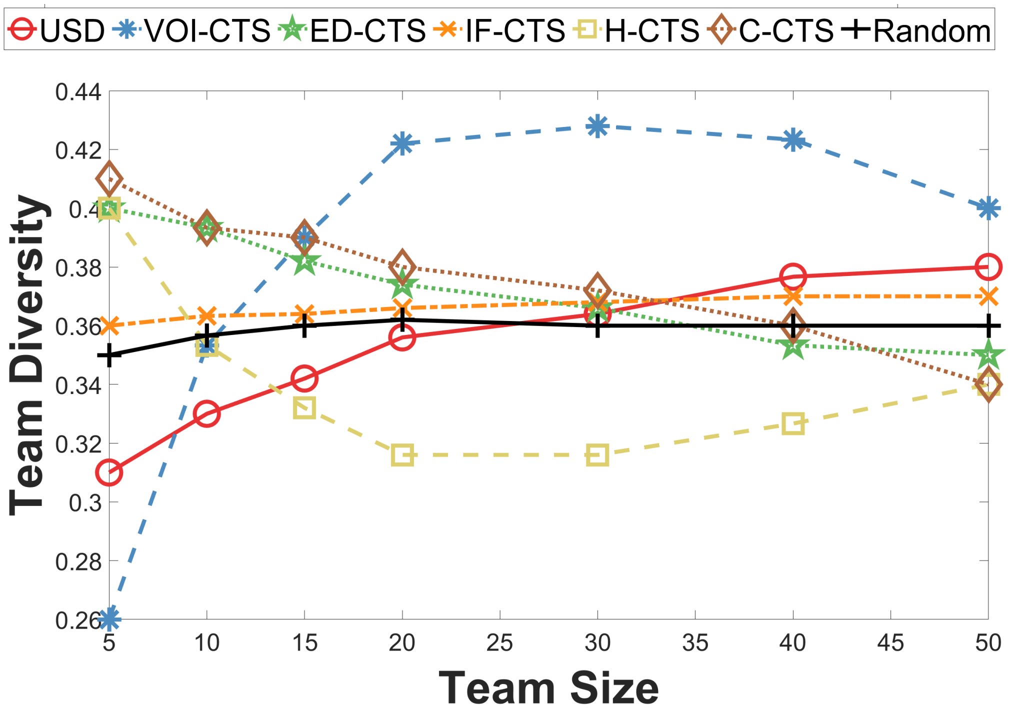
Figure 6.7: Comparison of different candidate team selection (CTS) methods based on expected social welfare and actual social welfare under varying team size, where the environment is set as the number of domains ($|E|$) = 5, and task composition ($\mathbf{L}(e_i) = [1, 1, 1, 1, 1]$) under a 5-hop trust network.



(a) Expected PPL

(b) Actual PPL

Figure 6.8: Comparison of different candidate team selection (CTS) methods based on expected ppl and actual ppl under varying team size, where the environment is set as the number of domains ($|E|$) = 5, and task composition ($\mathbf{L}(e_i) = [1, 1, 1, 1, 1]$) under a 5-hop trust network.



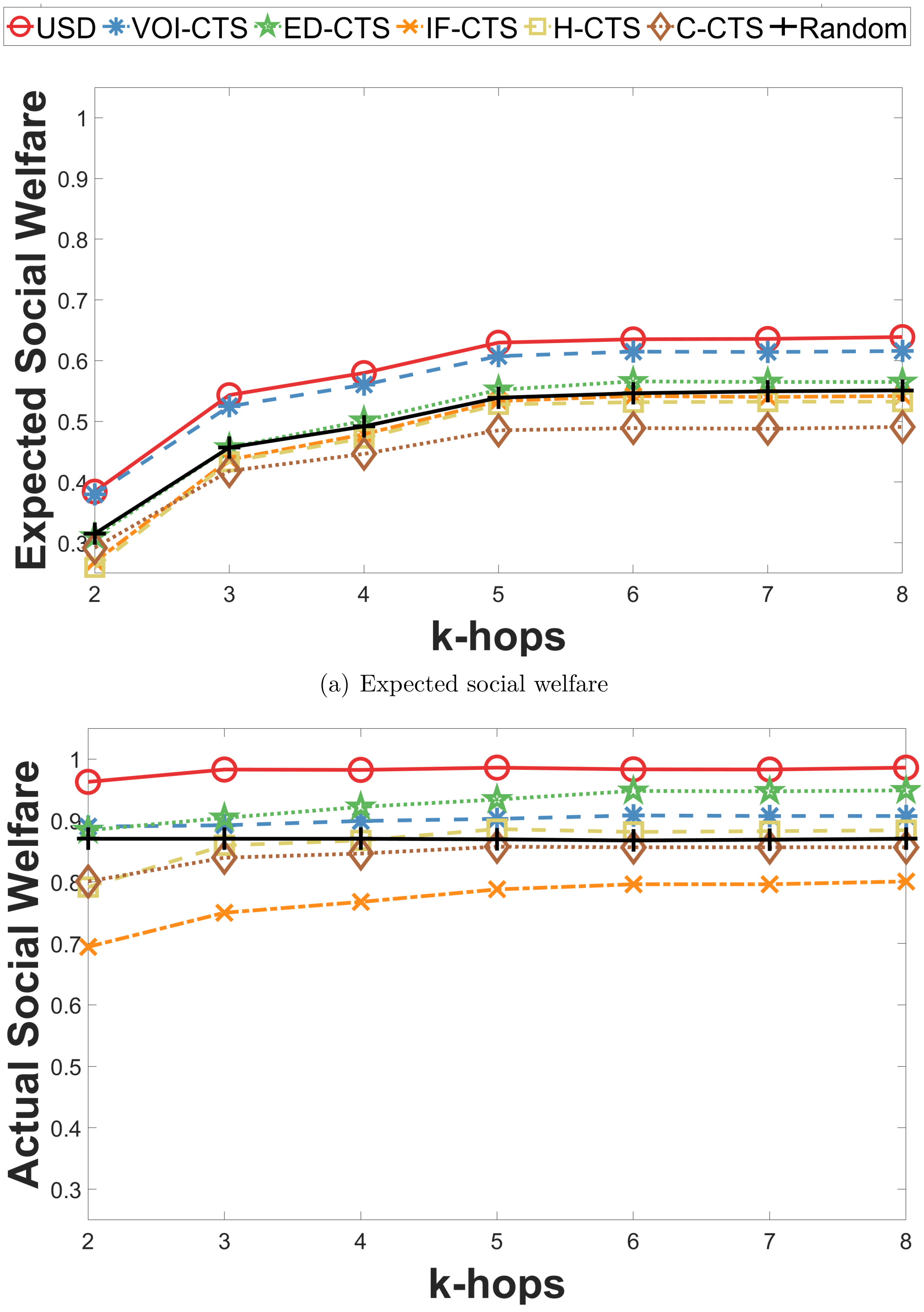
(a) Team diversity

Figure 6.9: Comparison of different candidate team selection (CTS) methods based on team diversity under varying team size, where the environment is set as the number of domains $(|E|) = 5$, and task composition $(\mathbf{L}(e_i) = [1, 1, 1, 1, 1])$ under a 5-hop trust network.

Looking at Figs. 6.7(b) and 6.8(b), we can see that as the team size increases, A-PPL increases while A-SW decreases. Second, as discussed in Section 6.1, there exists a trade-off between information sharing and privacy, which directly affects the utility received by the team player. We already discussed poor performance of IF-CTS. Unlike the general trend observed, as the team size increases, the utility from collective information sharing predominates the risk of privacy loss and consequently utility increases steadily. Lastly, C-CTS selects influential candidates in the network. From Fig. 6.8(b), it is observed that when the team size is small, a team with influential players have high A-PPL, which increases as the team size increases. However, after examining Fig. 6.7(b), we notice that although A-PPL increases, A-SW also increases. This implies that the valuable information shared by these influential members outweighs A-PPL. Fig. 6.9(a) shows the effect of different team size on the team diversity under the seven CTS schemes. The team diversity for USD and VoI-CTS increases as the team size increases because more task specific candidates are selected, increasing the expertise diversity. However, the team diversity decreases for ED-CTS, H-CTS, and C-CTS as the team size increases because although players are chosen based on diversity (for ED-CTS and H-CTS) and influence (C-CTS), the team consisting of high diversity individual cancel each other out with regards to diversity. Overall as the team sizes increase, all metrics converge to certain points. This implies that under a smaller size of the team, what CTS method to use is more important in promoting team performance.

6.4 Effect of Varying k-hop

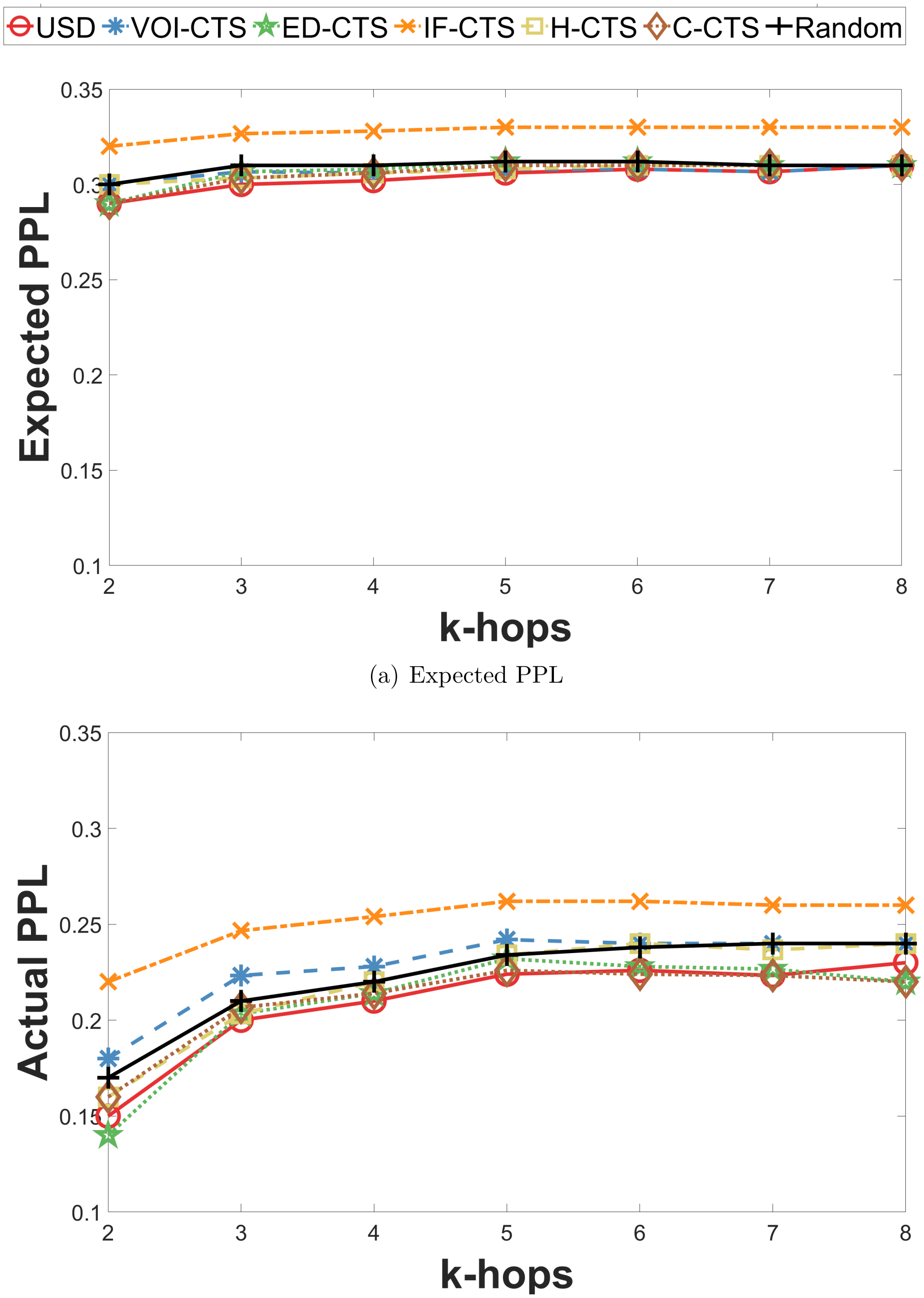
Figs. 6.10 & 6.11, demonstrate the effects of varying k-hops for E-SW and A-SW, E-PPL and A-PPL and Team Diversity respectively. The general trend inferred by observing Figs. 6.10 & 6.11 is that the values for all the metrics - E-SW, A-SW, E-PPL, A-PPL, team diversity,



(a) Expected social welfare

(b) Actual social welfare

Figure 6.10: Comparison of different candidate team selection (CTS) methods based on expected social welfare and actual social welfare under varying k-hops, where the environment is set as the number of domains ($|E|$) = 5, and task composition ($\mathbf{L}(e_i) = [1, 1, 1, 1, 1]$).



(a) Expected PPL

(b) Actual PPL

Figure 6.11: Comparison of different candidate team selection (CTS) methods based on expected ppl and actual ppl under varying k-hops, where the environment is set as the number of domains ($|E|$) = 5, and task composition ($\mathbf{L}(e_i) = [1, 1, 1, 1, 1]$).

stabilizes after 5-hops. The reason for this observed trend is that, in our network which consisting of 1,269 nodes, all the nodes and all the independent paths from the MD are discovered using 5-hops. Thereby, the trust values stabilizes and similar teams are formed when $k > 7$. Consequently for all the other results $k = 5$ is used, which covers 99% of the total nodes in the network.

From Fig. 6.10, we can notice that as more participants are considered for candidate team selection, and as the trust values increase (due to multiple independent paths being considered for *consensus*), the social welfare increases. However as the value of k increases, and participants further away from MD are considered, the resulting trust values for these participants is relatively low due to discounting the trust. Therefore, from Fig. 6.11, it is seen that the PPL value increases due to increase in distrust, for both Expected and Actual scenario.

6.5 Algorithmic Complexity

Table 6.1 shows the algorithmic complexity of all the CTS methods. From the table, you can see that ED-CTS has the highest algorithmic complexity whereas the H-CTS, and IF-CTS have the lowest. ED-CTS needs to calculate the expertise diversity by comparing each player with all other players in the network whereas on the other hand IF-CTS simply selects candidates based on their reported privacy preserving preferences. Overall the complexities are quite comparable for all the methods.

Additionally, Fig. 6.12, shows the actual runtime of all the 7 CTS methods. We can see that the runtimes for all 7 methods are pretty close and in the range $10^{-1.5}$. Contrary to the previous discussion, USD and H-CTS have the highest running time because it takes into account the additional operation costs too. For example, USD requires the VoI of each participants to calculate the resultant utility. However the figure shows that the runtimes are favourable and can be used in a real world situation. In the future, these CTS methods, can be further optimized.

Table 6.1: Algorithmic complexity of the 7 different CTS methods, where N is the number of player, $\mathcal{E}$ is the number of expertise domains, and E is the number of edges.

CTS Methods	**Time Complexity**
USD	$\mathcal{O}(\|N\|\|\mathcal{E}\|)$
VoI-CTS	$\mathcal{O}(\|N\|\|\mathcal{E}\|)$
C-CTS	$\mathcal{O}(\|N\|\|E\|)$
ED-CTS	$\mathcal{O}(\|N^2\|\|\mathcal{E}\|)$
H-CTS	$\mathcal{O}(\|N\|)$
IF-CTS	$\mathcal{O}(\|N\|)$

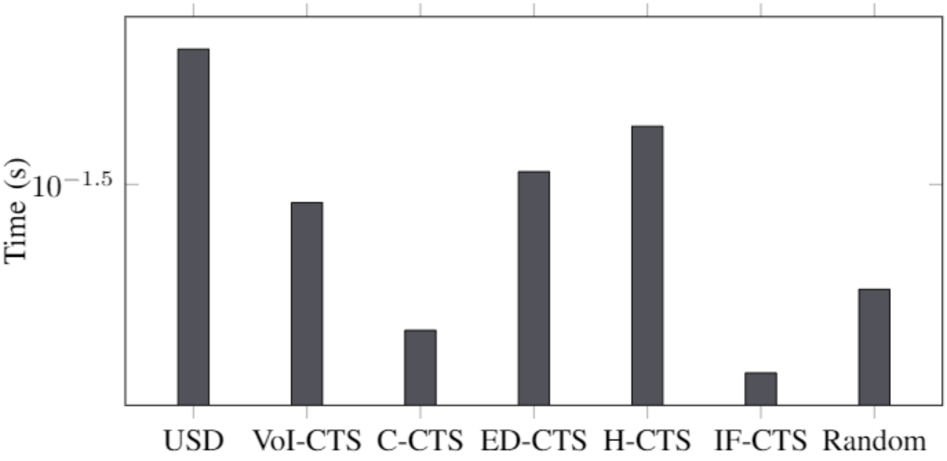


Figure 6.12: Comparison of different candidate team selection (CTS) methods based on running time, where the environment is set as the number of domains ($|E|$) = 5, pc_i=0.8 and task composition ($\mathbf{L}(e_i) = [1, 1, 1, 1, 1]$).

Chapter 7

Conclusions & Future Work Directions

In this work, we approached the Team Formation problem by considering both, diversity and privacy preserving preferences of a prospective team member. We first described the motivation for this research problem followed by the research goals and questions that we aim to answer through this study. A problem statement is then defined followed by the key contributions and outline of this thesis document. We then discussed the existing literature on collaborative team formation and the key factors impacting the team performance. This is followed by the preliminaries including the task model, information model and adversarial model. The key design features of the proposed PRADA-TF include the player's type, player's payoff, player's privacy preference revelation followed by the team selection process. We then described the experimental setup including the semi-synthetic dataset used, parameterization and comparing schemes. Finally we conducted detailed experimentation and discussed the findings in the Numerical Result and Analysis chapter. Below we summarize the key findings of this research.

7.1 Key Findings

We now address the answers for the five research questions raised in Chapter 1:

- **RQ1** *What is the relationship between team performance and team members' privacy preserving preferences?*

  Team members are more likely to compromise their privacy preference only when keeping their truthful preference significantly hurts their utilities in team performance.

- **RQ2** *What are the effects of team diversity on the team performance?*

  Although team diversity has a positive relationship with the team's social welfare consisting of team performance, individual privacy preserving, and adversarial behaviour only when the task requires subject-area specific expertise, information sharing has a strictly inverse relationship with the team's social welfare.

  Team diversity is important when the task requires subject area specific expertise. However, under a task requiring a single domain expertise, it is unclear that having a certain level of diversity is closely related to high team performance (i.e., high SW).

- **RQ3** *How do the trust relationships in users of a given online social network affect team performance?*

  We varied the k-hop value to investigate the effect of trust derivation on the expected/actual social welfare, team diversity, and privacy sacrifice. We found there exists a minimum k value that allows the team's performance to start converging to a certain point. This is well aligned with a real world scenario that working with a candidate introduced based on a very long trust chain won't work due to too shallow trust between a team leader and the team member.

- **RQ4** *How does the team behave in an actual real world scenario?*

  We observed that A-SW is likely to be higher than E-SW because team

members are more likely to compromise their privacy preferences, aiming to increasing information sharing and accordingly higher utility. Additionally, the actual PPL (A-PPL) is lower than the expected PPL (E-PPL) because E-PPL is calculated using the revealed type of the team members whereas in an actual task execution, a player might either stick to its revealed type or revert back to its true type, which would be higher, leading to less information shared and consequently less privacy loss.

- **RQ5** *How does the team size affect the social welfare of the team?*

  As the team size increases, information shared collectively, tends to outweigh the privacy loss suffered by individual members of the team.

7.2 Future Work Direction

Future work would consider:

- A more heuristic process for Trust Network Analysis (TNA), to consider only those paths which lead to high trust values, instead of considering all independent paths. This would make the team formation process more optimized.
- Coming up with an incentive-compatible-direct mechanism where it a dominant strategy for the player to reveal truthful information, which can then be solved for Bayesian Nash Equilibrium (BNE).
- Collecting the privacy preference of individuals anonymously and using it for the detailed analysis would help to see more interesting real-world trends.
- More realistic adversarial behavior wherein specific malicious users perform targeted

attacks against specific set of users.

- Communication cost between team members can be considered, which can vary based on the differences between the team member's expertise background or simply how far apart the two members are in the network.

- Competing teams where recruitment needs to be considered and where there is a possibility of schedule and resource constraint, if a player joins multiple teams.

- Dynamic feedback system wherein a set of prospective team members provided by PRADA-TF can be re-valuated based on the feedback provided by the team leader (MD).

- More expertise domains as currently in this work only 5 expertise domains are considered. Although the proposed system can handle dynamically changing expertise domains, more expertise domains can be considered for realising some interesting trends.

- Running a simple regression or any statistical technique to measure the relationship between privacy and diversity, privacy and social welfare and diversity and social welfare.

This work has been submitted to International Conference on Web Services (ICWS) 2021.

Bibliography

[1] B.-C. Lim and K. J. Klein, “Team mental models and team performance: a field study of the effects of team mental model similarity and accuracy,” *Journal of Organizational Behavior*, vol. 27, no. 4, pp. 403–418, 2006.

[2] S. Mohammed and B. C. Dumville, “Team mental models in a team knowledge framework: expanding theory and measurement across disciplinary boundaries,” *Journal of Organizational Behavior*, vol. 22, no. 2, pp. 89–106, 2001.

[3] B. Edwards, E. Day, W. Arthur, Jr, and S. Bell, “Relationships among team ability composition, team mental models, and team performance.” *The Journal of Applied Psychology*, vol. 91, pp. 727–36, 2006.

[4] T. Lappas, K. Liu, and E. Terzi, “Finding a team of experts in social networks,” in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’09, New York, NY, USA, 2009, pp. 467–476.

[5] S. E. Page, *Diversity and Complexity*, 1st ed. USA: Princeton University Press, 2010.

[6] ——, *The Difference: How The Power of Diversity Creates Better Groups, Firms, Schools, and Societies.* Princeton, NJ: Princeton University Press, 2007.

[7] D. Knight, C. L. Pearce, K. G. Smith, J. D. Olian, H. P. Sims, K. A. Smith, and P. Flood, “Top management team diversity, group process, and strategic consensus,” *Strategic Management Journal*, vol. 20, no. 5, pp. 445–465, 1999.

[8] S. Mohammed and E. Ringseis, “Cognitive diversity and consensus in group decision

making: The role of inputs, processes, and outcomes," *Organizational behavior and human decision processes*, vol. 85, pp. 310–335, 08 2001.

[9] M. C. Silaghi and D. Mitra, "Distributed constraint satisfaction and optimization with privacy enforcement," in *Proceedings. IEEE/WIC/ACM International Conference on Intelligent Agent Technology, 2004. (IAT 2004).*, 2004, pp. 531–535.

[10] R. Greenstadt, J. P. Pearce, and M. Tambe, "Analysis of privacy loss in distributed constraint optimization," in *Proceedings of the 21st National Conference on Artificial Intelligence*, ser. AAAI'06, vol. 1. AAAI Press, 2006, pp. 647–653.

[11] M. Harbers, R. Aydogan, C. M. Jonker, and M. A. Neerincx, "Sharing information in teams: Giving up privacy or compromising on team performance?" in *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*, ser. AAMAS '14. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 413–420.

[12] A. G. Erdman and G. N. Sandor, *Mechanism Design (3rd Ed.): Analysis and Synthesis (Vol. 1).* USA: Prentice-Hall, Inc., 1997.

[13] J. Mesmer-Magnus and L. DeChurch, "Information sharing and team performance: A meta-analysis," *Journal of Applied Psychology*, vol. 94, no. 2, pp. 535–546, Mar. 2009.

[14] A. R. Wellens, "Effects of telecommunication media upon information sharing and team performance: some theoretical and empirical observations," in *Proceedings of the IEEE National Aerospace and Electronics Conference*, vol. 2, 1989, pp. 726–733.

[15] Y. Narahari, R. Narayanam, D. Garg, and H. Prakash, *Foundations of Mechanism Design.* London: Springer London, 2009, pp. 1–131.

[16] M. E. J. Newman, “Finding community structure in networks using the eigenvectors of matrices,” *Phys. Rev. E*, vol. 74, p. 036104, Sep 2006.

[17] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, “Complex networks: Structure and dynamics,” *Physics Reports*, vol. 424, no. 4, pp. 175–308, 2006.

[18] M. Newman, “Newman mej.. the structure and function of complex networks. siam rev 45: 167-256,” *SIAM Review*, vol. 45, 08 2003.

[19] C.-T. Li, M.-K. Shan, and S.-D. Lin, “On team formation with expertise query in collaborative social networks,” *Knowledge and Information Systems*, vol. 42, no. 2, pp. 441–463, Feb 2015.

[20] M. Kargar and A. An, “TeamExp: Top-k team formation in social networks,” in *IEEE 11th International Conference on Data Mining Workshops*, Dec. 2011, pp. 1231–1234.

[21] A. Anagnostopoulos, L. Becchetti, C. Castillo, A. Gionis, and S. Leonardi, “Online team formation in social networks,” in *Proceedings of the 21st International Conference on World Wide Web (WWW’12).* New York, NY, USA: ACM, 2012, pp. 839–848.

[22] A. Bhowmik, V. Borkar, D. Garg, and M. Pallan, “Submodularity in team formation problem,” in *Proceedings of the 2014 SIAM International Conference on Data Mining*, 2014, pp. 893–901.

[23] A. Gajewar and A. D. Sarma, “Multi-skill collaborative teams based on densest subgraphs,” in *Proceedings of the 2012 SIAM International Conference on Data Mining*, 2011, pp. 165–176.

[24] S. Datta, A. Majumder, and K. Naidu, “Capacitated team formation problem on social networks,” in *Proceedings of the 18th ACM SIGKDD International Conference on*

Knowledge Discovery and Data Mining (KDD '12), New York, NY, USA, 2012, pp. 1005–1013.

[25] J. Basiri, F. Taghiyareh, and A. Ghorbani, "Collaborative team formation using brain drain optimization: a practical and effective solution," *World Wide Web*, vol. 20, no. 6, pp. 1385–1407, Nov. 2017.

[26] J. Basiri and F. Taghiyareh, "Introducing a socio-inspired swarm intelligence algorithm for numerical function optimization," in *4th International Conference on Computer and Knowledge Engineering (ICCKE)*, Oct. 2014, pp. 462–467.

[27] X. Wang, Z. Zhao, and W. Ng, "USTF: A unified system of team formation," *IEEE Transactions on Big Data*, vol. 2, no. 1, pp. 70–84, Mar. 2016.

[28] W. Wang, J. Jiang, B. An, Y. Jiang, and B. Chen, "Toward efficient team formation for crowdsourcing in noncooperative social networks," *IEEE Transactions on Cybernetics*, vol. 47, no. 12, pp. 4208–4222, Dec 2017.

[29] M. Wright and Y. Vorobeychik, "Mechanism design for team formation," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, ser. AAAI'15. AAAI Press, 2015, pp. 1050–1056.

[30] M. Bechtoldt, C. De Dreu, and B. Nijstad, "Team personality diversity, group creativity, and innovativeness in organizational teams," 2007 (Accessed on: 04/04/2021). [Online]. Available: http://www.susdiv.org/uploadfiles/RT3.2_PP_Carsten.pdf/

[31] D. G. Ancona and D. F. Caldwell, "Demography and design: Predictors of new product team performance," *Organization Science*, vol. 3, no. 3, pp. 321–341, 1992.

[32] K. Phillips and C. O'Reilly, *Demography and Diversity in Organizations: A Review of 40 Years of Research*, 01 1998, vol. 20, pp. 77–140.

[33] S. Horwitz and I. Horwitz, “The effects of team diversity on team outcomes: A meta-analytic review of team demography,” *Journal of Management - J MANAGE*, vol. 33, 2007.

[34] A. Pieterse, D. Knippenberg, and D. Van Dierendonck, “Cultural diversity and team performance: The role of team member goal orientation,” *Academy of Management Journal*, vol. 56, pp. 782–804, 2012.

[35] T.-P. Liang, C.-C. Liu, T.-M. Lin, and B. Lin, “Effect of team diversity on software project performance,” *Industrial Management and Data Systems*, vol. 107, pp. 636–653, 2007.

[36] S. Cohen and M. Yashinski, “Crowdsourcing with diverse groups of users,” in *Proceedings of the 20th International Workshop on the Web and Databases*, ser. WebDB’17. New York, NY, USA: Association for Computing Machinery, 2017, p. 7–12.

[37] J. S. Olson, J. Grudin, and E. Horvitz, “A study of preferences for sharing and privacy,” in *CHI ’05 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA ’05. New York, NY, USA: Association for Computing Machinery, 2005, p. 1985–1988.

[38] C. Dwork, “Differential privacy,” in *Automata, Languages and Programming*, M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 1–12.

[39] S. Kasiviswanathan and A. Smith, “On the ‘semantics’ of differential privacy: A bayesian formulation,” *Journal of Privacy and Confidentiality*, vol. 6, 03 2008.

[40] F. McSherry and K. Talwar, “Mechanism design via differential privacy,” in *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS’07)*, 2007, pp. 94–103.

[41] K. Nissim, C. Orlandi, and R. Smorodinsky, “Privacy-aware mechanism design,” in *Proceedings of the 13th ACM Conference on Electronic Commerce*, ser. EC ’12. New York, NY, USA: Association for Computing Machinery, 2012, p. 774–789.

[42] X. Xiao, G. Wang, and J. Gehrke, “Differential privacy via wavelet transforms,” *IEEE Trans. on Knowl. and Data Eng.*, vol. 23, no. 8, p. 1200–1214, Aug. 2011.

[43] D. Trapido, “How novelty in knowledge earns recognition: The role of consistent identities,” *Research Policy*, vol. 44, no. 8, pp. 1488 – 1500, 2015.

[44] A. Briggs, “Novelty and appropriability: The role of entrepreneurial knowledge in sharing information,” 2009.

[45] A. Jøsang, *Subjective Logic: A Formalism for Reasoning Under Uncertainty*, 1st ed. Springer Publishing Company, Incorporated, 2016.

[46] L. S. Marcolinon, A. X. Jiang, and M. Tambe, “Multi-agent team formation: Diversity beats strength?” Aug. 2013, pp. 279–285.

[47] D. Hand, “Statistical decision theory: Estimation, testing, and selection by friedrich liese, klaus-j. miescke,” *International Statistical Review*, vol. 76, pp. 450–450, Feb. 2008.

[48] I. Kamel, Z. Al Aghbari, and K. Kamel, “SmartRecruiter: A similarity-based team formation algorithm,” *International Journal of Big Data Intelligence*, vol. 3, pp. 228–238, 01 2016.

[49] P. Dey, M. Ganguly, and S. Roy, “Network centrality based team formation: A case study on t-20 cricket,” *Applied Computing and Informatics*, vol. 13, no. 2, pp. 161 – 168, 2017.