Expert Systems With Applications 221 (2023) 119781

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A coalition formation framework of smallholder farmers in an agricultural cooperative

Samriddhi Sarkar [a,c,*], Tuhin Biswas [a], Mariana Curado Malta [b,c], Deolinda Meira [b], Animesh Dutta [a]

[a] National Institute of Technology Durgapur, India
[b] CEOS.PP, Polytechnic of Porto, Portugal
[c] ALGORITMI Research Centre, LASI, University of Minho, Guimarães, Portugal

ARTICLE INFO

Keywords:
Coalition formation
Multi-agent systems
Cooperative of smallholder farmers
Similarity

ABSTRACT

Agricultural cooperatives remain a significant component of the food and agriculture industry to help the stakeholders to provide services and have opportunities for themselves. One of the aims of an agricultural cooperative is to answer to the needs within the communities of the farmers. Agricultural cooperatives enable individual farmers to increase productivity and maximise their social welfare. Together the farmer members of an agricultural cooperative can buy input supplies cheaper and sell more of their products in larger markets at higher prices, which is not possible for an individual smallholder farmer otherwise. Some studies have shown that farmers who were members of cooperatives have gained higher revenue for their products and spent less on input. However, organising the hundreds of farmers into smaller groups to perform collective farming and marketing is crucial to strengthening their position in the food and agriculture industry. Thereby, in our work, we consider an agricultural cooperative of smallholder farmers as a multi-agent based coalitional model, where coalitions are formed based on the similarity among the smallholder farmers. In this paper, we propose a model and implement a heuristic-based algorithm to find the disjoint partition of the agents set. We evaluate the model and the algorithm based on the following criteria: (i) individual gain, (ii) runtime analysis, (iii) solution quality, and (iv) scalability. We theoretically prove that our coalitional model of an agricultural cooperative has conciseness, expressiveness and efficiency properties. Experimental results confirm that our algorithm is time efficient and scalable. We show, both empirically and theoretically, that our algorithm generates a solution within a bound of the optimal solution. We also show that our coalition model generates positive revenue for the smallholder farmers and the payoff division rule is individual rational. In addition, we generate a new dataset in the context of an agricultural cooperative to show the effectiveness and efficiency of the proposed coalitional model of the cooperative.

1. Introduction

The Food and Agriculture Organisation of the United Nations (FAO) estimates that over 2.5 billion people depend on agriculture for their livelihood (Rapsomanikis, 2015). Sustainable growth and profitability are long-standing challenges in the agriculture sector in many countries (Byerlee et al., 2008). According to Byerlee et al. (2008), almost 85% (as of 2008) of the farmers in the world are smallholder farmers, owning land plots smaller than 2 ha (Rapsomanikis, 2015). These small farmers do not produce enough to dictate the price of their produce in the market at an individual level. They have more difficulties buying inputs at bulk rates, increasing the volume to open new markets, lowering the cost of equipment, to access required services or facilities.

A recent survey by Abdul-Rahaman and Abdulai (2020a) of 447 rice farmers in rural Ghana examine the role of farmer groups and collective marketing in improving smallholder farm performance. Their study finds that group formation has a positive impact on the earnings of an individual farmer as well as on the farm's net revenues. Abdul-Rahaman and Abdulai (2020a) have shown that the farmers who were members of farmer's groups that collectively accessed the market gained higher income for their production and incurred lower input costs. Organising smallholder farmers effectively into groups to perform collective farming and marketing is crucial to strengthening their position in the food value chain. Despite having small land, the aggregated productivity of the smallholder farmers is higher than the medium

* Corresponding author at: National Institute of Technology Durgapur, India.
E-mail addresses: samriddhi2010@gmail.com (S. Sarkar), tuhin11biswas@gmail.com (T. Biswas), mariana@iscap.ipp.pt (M. Curado Malta), meira@iscap.ipp.pt (D. Meira), animesh@cse.nitdgp.ac.in (A. Dutta).

https://doi.org/10.1016/j.eswa.2023.119781
Received 9 July 2022; Received in revised form 28 January 2023; Accepted 28 February 2023
Available online 4 March 2023
0957-4174/© 2023 Elsevier Ltd. All rights reserved.

and large-size farmers (Rapsomanikis, 2015). Agricultural cooperatives remain a significant component of the food and agriculture industry to help the stakeholders provide services-the members of an agricultural cooperative pool their lands, human resources, and other resources to cultivate jointly. Working together facilitates a more rational use of resources and the adoption of scientific production methods. In general, developing agricultural cooperatives in developing countries is expected to help smallholder farmers increase their market share, agricultural income, and crop productivity and lower production costs (Ma & Abdulai, 2017). It allows the smallholder farmers to do what big farms can do, generating more income. Each member of an agricultural cooperative is rewarded according to their participation - in exchange for participating in the cooperative activity. The cooperator receives economic advantages, which the doctrine calls "mutualistic advantages" or "surplus". Since the surplus is the result of the work of the cooperators, it is distributed among the cooperators in proportion to their transactions with the cooperative.

Agricultural cooperatives fulfil an important social function. This social function justifies the relevant role of cooperatives in the social development of people, communities, and nations. According to the World Cooperative Monitor,[1] there are 3 million cooperatives in the world, and at least 12% of the world population are cooperators of a cooperative. On average, one in six people is a member or a customer of a cooperative, and cooperatives employ 10% of the world's population. The 300 largest cooperatives and mutuals generate 2146 billion USD in turnover while providing the services and the infrastructure society needs to thrive.

The cooperatives' social function is particularly relevant for smallholder farmers since the aim of a cooperative is closely linked to the promotion of the interests of the cooperative members, i.e., the objective is to meet the needs of the communities where the cooperative acts. The International Labour Organisation (ILO) highlights that "Cooperatives are sustainable enterprises that are owned and run by their membership, and are built on values that encourage cooperation, empowerment, and solidarity, rather than just profits."[2] Similarly, the International Cooperative Alliance (ICA) states, "Collectively members own their cooperative, and through democratic arrangements, they participate in its governance. Individually they have a right to information, a voice, and representation."[3] Cooperatives do not have an autonomous purpose concerning their members but are an instrument to satisfy the individual needs of every cooperative member, who work, consume, sell and provide services in and through the cooperative (Fajardo et al., 2017). The objective of a cooperative is to maximise the benefits that members can derive from the operations they carry out through/or with the cooperative; it is never to make a profit to share. Although focused on the needs of their members, cooperatives work to achieve sustainable development of their communities according to criteria approved by the members.

The incorporation of members from the territorial area where the cooperative mainly carries out its activity has been a constant in this type of organisation, whose ultimate goal is to meet the needs felt by the community. Thus cooperatives appear as entities that generate stable jobs - mainly because cooperatives have strong local roots, develop activities that cannot be relocated due to their nature - and foster an entrepreneurial spirit (Meira, 2022).

Agricultural cooperatives have a substantial presence in the agricultural sector worldwide (Candemir et al., 2021). Members of an agricultural cooperative are less prone to economic, environmental, and social threats. Cooperative values such as democratic decision-making, equality, and solidarity within the cooperatives help farmers cope with market risks and other unforeseen events.

The crisis of COVID-19 has demonstrated the resilience and creativity of cooperatives since they are the organisations that put social and environmental objectives first, reinvesting most of their profits back into the organisation itself. Given their strong roots in the community and their participatory, bottom-up governance model, the cooperatives demonstrated their ability to build and offer innovative solutions to many of the challenges posed by the pandemic. This resilience and innovation during the pandemic have ensured that cooperatives have maintained their contribution in crucial sectors such as access to food, home help, and health (Álvarez et al., 2022).

Cooperatives of the 21st century introduced an "intensive use of information technologies in the different stages" of its value chain (Ciruela-Lorenzo et al., 2020, pg.1); alongside this introduction of information technologies, there is an increased use of smart tools to support cooperative activities (Ciruela-Lorenzo et al., 2020). Our work is framed on this trending topic of smart tools, where we aim to provide the top organisations of cooperatives, i.e. the International Cooperatives Alliance,[4] and also market-stakeholders such as hardware-software houses, with a tool in the form of an algorithm that can be integrated into, e.g. the software of virtual cooperatives (See Bialoskorski Neto, 2001 to know about the definition of virtual cooperatives), or in the management software-tools of the cooperatives, among other possibilities.

Previous work (see Sarkar et al., 2020) explores the possibility of multi-agent-based coalition formation of smallholder farmers in an agricultural cooperative. They aim to provide algorithms that create virtual groups of smallholder farmers to increase their social welfare. While agent technology is used as a metaphor to design and operate virtual enterprises (Fischer et al., 1996), coalition formation enables the farmers to purchase the resources at volume discount prices. Researchers of Distributed Artificial Intelligence (DAI) have been working on coalition formation and its applications for the last three decades to coordinate the agents as in "coalition of buyers in e-commerce" - where buyers form coalitions to take advantage of volume discount (Chen et al., 2008; Li et al., 2003; Li & Sycara, 2002; Li et al., 2010), "coalition in electricity marketplaces" - where energy consumers form coalitions with complementary or similar energy profiles avail discounts on the purchase (Alam et al., 2013; Chalkiadakis et al., 2011; Contreras & Wu, 1999), "coalition in virtual organisations" - where individuals, departments or organisations form coalitions to combine their resources and capabilities to attain particular market niches (Jennings & Dang, 2004), "coalition of sensors" - where sensors form coalitions to track the target of interest precisely or to save cost (Dang et al., 2006), "coalition formation in social networks" - where individuals in a social network are modelled as the agents to improve the group's utility using coalition formation game theory-based approach to identify the best communities (Zhou et al., 2015), "coalitions of robots" - where a large set of robots join together and perform complicated tasks (Agarwal et al., 2015, 2014). The coalition formation process generally focuses on coordinating the stakeholders by partitioning them into smaller groups to achieve individual or collective goals whenever they cannot do so on their own. The entire coalition formation process typically involves three main activities: (i) generating the coalitions and calculating their values using a characteristic function, (ii) finding a partition of the set of agents into disjoint coalitions that maximise the sum of the coalition values (also called as Coalition Structure Generation (CSG) problem), and (iii) distribution of the payoff among the agents for their contribution to the coalitions (Rahwan et al., 2015). To date, several

[1] See https://monitor.coop/en/media/library/research-and-reviews/world-cooperative-monitor-2021 - accessed on August 26, 2022.

[2] Available at https://www.ilo.org/global/topics/cooperatives/areas-of-work/WCMS_445131 - accessed on June 26, 2022.

[3] "Blue Print for a Co-operative Decade" - available at https://www.ica.coop/sites/default/files/2021-11/ICA%20Blueprint%20-%20Final%20-%20Feb%2013%20EN.pdf - accessed in June 27, 2022.

[4] See https://www.ica.coop/ - accessed on June 26, 2022.

approaches are proposed to solve this problem, and the approaches proposed for the applications of coalition formation are heavily dependent on the requirements of any application. Furthermore, the traditional approaches can work on relatively small sets of agents as this is an NP-complete problem. However, they cannot be applied to solve the farmers' coalition formation (where hundreds of farmers are involved) in a reasonable amount of time.

This paper proposes a multi-agent-based cooperation model of an agricultural cooperative consisting of hundreds of smallholder farmers in the form of a virtual enterprise. We consider the smallholder farmers as the agents. A virtual enterprise is defined as a temporary network made up of independent and autonomous entities to take advantage of specific market opportunities to maximise the profits of the individuals and the virtual enterprise (Petersen et al., 2001). Participants of a virtual enterprise mainly contribute with their competencies, and they act as a single corporation to the outside world (Fischer et al., 1996; Petersen et al., 2001). Thus, in this paper, we propose an extension of Sarkar et al. (2020) by investigating how the similarity among the agents' resource requirements can be used to improve the formation of coalitions.

The main contributions of the paper are :

1. A formal model of an agricultural cooperative of smallholder farmers in the sense of coalitional games. In particular, a model of an individual smallholder farmer as an agent and a coalitional model of the agricultural cooperative as a network of heterogeneous agents. We theoretically prove that our model possesses the conciseness, expressiveness and efficiency properties (Tran-Thanh et al., 2013).
2. A framework to connect the agents in an agricultural cooperative and a method to calculate the similarity of the agents.
3. A characteristic function to evaluate the worth of a coalition in the context of an agricultural cooperative of smallholder farmers, and a set of constraints that demonstrate the essence of an agricultural cooperative.
4. A novel algorithm to find the partition of the agents into disjoint coalitions based on the similarity of the agents for a large-scale system consisting of hundreds of agents: (i) we theoretically establish the approximation ratio of the solution generated by our proposed algorithm; (ii) the proposed algorithm does not need to store $2^n - 1$ coalition values in memory. Instead, it calculates only the values of coalitions in the coalition structure that the algorithm finally generates.
5. A method of dividing the payoff among the agents of the final coalition structure in the context of an agricultural cooperative. The members of the agricultural cooperative profit in proportion to their contribution to the cooperative. We show that the proposed payoff division rule is individual rational.
6. A new dataset in the context of an agricultural cooperative to evaluate the effectiveness and efficiency of the proposed model and algorithm. We detail the steps to create this dataset. This dataset can be used in the future to implement improved algorithms for coalition formation of smallholder farmers.

This article is organised as follows: Section 2 provides the related work. Section 3 presents the research problem of coalition formation in a virtual agricultural cooperative. Section 4 presents a multi-agent based modelling of an individual farmer and an agricultural cooperative of smallholder farmers. Section 5 presents our coalitional game based solution approach. Section 6 presents the methodology and experimental evaluation of the model, detailing the experimental setup, data generation, and metrics for the evaluation. Section 7 presents the results and analysis of the results. Section 8 presents the limitations of the work and future work. Finally, Section 9 concludes the article.

2. Related work

2.1. Cooperative games perspective

The literature on coalition formation has been growing in the last three decades. In this section, we present the literature on coalition formation into two strands. The first strand focuses on the theoretical aspects of coalition formation, and the second focuses on the real-life applications of coalition formation. The second strand is the most relevant to our work on forming coalitions of smallholder farmers. In the second strand, we discuss some of the works that focus on developing algorithms for general coalition formation issue (for details of theoretical aspects of coalition formation, see Rahwan et al., 2015 and for details of real-life applications of coalitions formation see, Sarkar et al., 2022).

2.1.1. Coalition structure generation algorithms

A wide range of algorithms has been developed for the theoretical strand of coalition formation. Some of these algorithms focus on enumerating all possible coalitions. Most of the algorithms use a classical characteristic function to calculate the values of the coalitions. The algorithms that make no assumptions about value functions are the most straightforward ones. However, the memory requirement to store the coalition values grows exponentially with the number of agents, and also the representation is not scalable (Tran-Thanh et al., 2013). On the other hand, some algorithms impose restrictions on forming certain infeasible coalitions. A few algorithms consider infeasible coalitions as feasible but with a value ∞ (Voice et al., 2012). However, the latter is not a feasible approach. The following sections summarise the (i) Classic algorithms and (ii) Graph-based algorithms.

- *Classic algorithms:* The classic algorithms for coalition formation can be clustered into three categories: (i) *Exact algorithm:* given a finite number of agents, the algorithm guarantees to find an optimal solution. The time complexity of such an algorithm is $O(3^n)$ (Rahwan & Jennings, 2008), (ii) *Anytime algorithm:* given a short execution time and strict deadlines, an anytime algorithm can return a valid solution. In a situation with such time constraints, it becomes pointless to use an exact algorithm since it needs to be run until the end to generate a solution. The worst-case time complexity of an anytime is $O(n^n)$ (Service & Adams, 2010), even if it is interrupted before it ends, (iii) *Heuristics algorithm:* this type of algorithm produces a good enough solution for a given problem within a reasonable amount of time (Shehory & Kraus, 1998). Farinelli et al. (2013) propose a heuristic approach called C-link algorithm based on the hierarchical clustering method, which merges two coalitions if the resultant coalition brings more gain to the system. The C-link starts with singleton coalitions, and at each iteration, it finds the most suitable pair of coalitions. The merging process is stopped if the grand coalition is formed or merging two coalitions results in a negative value.
- *Graph-based algorithms:* The classic algorithms of CSG enumerate every possible subset of agents as a potential coalition. However, there are many real-life scenarios where the formation of some coalitions is forbidden. A graph can express these types of constraints. In a graph-based setting, vertices represent the agents, and the edges represent the synergies or relationships among the agents (Bistaffa et al., 2017). Here, only the feasible coalitions are enumerated. A coalition is feasible if the agents in a coalition are connected to each other, i.e., if they form a connected sub-graph (Bistaffa et al., 2017) or a complete sub-graph (clique) (Tošić & Agha, 2004). The idea is that the agents in a coalition are similar in terms of the level of skills, capabilities, or requirements. Voice et al. (2012) propose three algorithms for (i) coalition enumeration (SlyCE), (ii) coalition

value evaluation (D-SlyCE), and for (iii) coalition structure generation (DyCE). SlyCE performs the preprocessing and generates feasible coalitions. These resultant coalitions are the input for DyCE. DyCE depends on the working principle of dynamic programming, which makes the algorithm inefficient in terms of time and memory requirements. Bistaffa et al. (2014, 2017) propose an anytime algorithm for the graph-based settings, where the working principle depends on the concept of edge contraction. They validate their method using three benchmark (*m+a*) characteristic functions[5]: (i) collective energy purchasing, (ii) edge sum with coordination, and (iii) coalition size with distance cost.

2.1.2. Applications

The theoretical strand of coalition formation mainly focuses on finding coalition structures while minimising the time and space complexity. For real-life applications of coalition formation, the authors address the issue of modelling the application according to the requirements of coalition formation and designing the characteristic function according to the need of that application. We can find some work in group buying in electronic marketplaces and electricity marketplace similar to forming coalitions in an agricultural cooperative. In a market-like setting, sellers and customers trade among themselves, where the sellers wish to sell large quantities of products to the customers, while the customers seek partnerships with other customers to get discounts on product prices. Coalition formation is beneficial to model such cooperation and negotiation of rational entities in many market-like settings (Sarkar et al., 2022). The following sub-sections detail the state-of-the-art approaches on electronic marketplaces and electricity marketplace.

- *Group buying in an electronic marketplace:* In an electronic marketplace (e-marketplace), buyers take advantage of group buying through coalitions (Yamamoto & Sycara, 2001). Usually, these buyers have less bargaining power in the market, and together they can obtain a product without paying more than their valuation. He and Ioerger (2004) address the problem of small buyers who want to buy items in bundles at a lower price without buying more than their real demand. The customers are likely to join coalitions that will maximise their social welfare. If joining any coalition does not fetch them the highest profit, they will try to leave that coalition. Contreras and Wu (1999) search for the best coalition to join based on Bilateral Shapley values (Shapley, 1953). The Shapley value is a widely used mechanism in game theory to distribute the surplus earned by the members among them in a stable way (Li et al., 2003). In some situations, buyers want to buy a combination of heterogeneous items in a marketplace with different preferences and reservation prices. To address this issue, Li and Sycara (2002) and Li et al. (2010) propose a polynomial time approximation algorithm called combinatorial coalition formation.
- *Group buying in an electricity marketplace:* Energy resources sometimes share their unused excess energy with others who need energy at the moment. This domain of energy trading, where matching demand and supply is crucial, has gained much popularity lately. Agent-based modelling of energy trading enables autonomy among the participants (Yasir et al., 2018). In order to get discounts on electricity through joint purchases, Vinyals et al. (2012) use a complementary based method to organise consumers in a virtual power consumer group. This group acts as a single coalition and takes advantage of the discount. They focus on forming coalitions among electricity consumers, where the energy requirement of the consumers complement each other. They have also utilised the properties of social network to establish connections among consumers. Chalkiadakis et al. (2011) propose a virtual cooperative of distributed energy resources, where the energy resources form coalitions to sell their energy profitably to the main electricity grid. However, they discuss the formation of a virtual energy cooperative from a game-theoretic standpoint, as they use mechanism design and core-stable payment allocation. Alam et al. (2013) address the issue of energy exchange between households in a community of remote villages in the developing world, where the households heavily rely on renewable generation units and electric batteries. They propose an agent-based approach to reduce the overall battery usage, thus prolonging the life of batteries and reducing energy losses. They propose a linear programming model with a wide range of constraints. The constraints perfectly present the essence of the whole system.
- *Agricultural cooperatives* Some researchers of artificial intelligence use multi-agent based simulation as a valuable tool to validate their proposals for the agriculture sector, e.g., farm-land auction, food-defence training and analysis, and agricultural supply chain management (Borodin et al., 2016). Sarkar et al. (2020) present a multi-agent based modelling of an agricultural cooperative and propose a core-stable payoff allocation method.

2.2. Other perspectives

In this section, we present studies related to an agricultural cooperative from the perspective of agricultural science and economics.

An agricultural cooperative involves a network of farmer members who cultivate as a unit. Through cooperation, these cooperatives enable independent farmers to enter a bigger market and buy input supplies at lower prices. Compared to individual farmers of investor-owned farms, the members of an agricultural cooperative are more economically protected and face lower risks.[6] A large number of work can be found on the advantages and benefits of an agricultural cooperative. Traditionally, an agricultural cooperative is defined as a form of an economic organisation of a set of farmers (Candemir et al., 2021). However, Tortia et al. (2013) mention that there is no clear economic definition of agricultural cooperatives. Researchers have presented different governance structures and economic behaviour to model an agricultural cooperative. Hueth and Marcoul (2015) define cooperatives as a form of coalition among farmers with similar objectives. However, their work focuses on differentiating "member-owned" farm and "investor-owned" farms based on the organisational environment. One of the seminal works in cooperatives' economic behaviour is by Staatz (1983), where the author models an agricultural cooperative as an *n*-person cooperative game. Here, the author focuses on a game theory based modelling to allocate the costs and benefits to heterogeneous farmers in an agricultural cooperative. Ma and Abdulai (2017) examine and find the statistically significant impacts of agricultural cooperative membership on smallholder farmers using data from a survey of 481 apple-producing households in rural China. Abdul-Rahaman and Abdulai (2020a) observe that effective organisation of smallholder farmers into groups to undertake production and marketing activities help them achieve higher prices for the product and reduce input costs.

3. Research problem identification

3.1. Gaps in the state-of-the-art

We identify the following research gaps:

[5] It is the sum of a monotonic function $v(C)^+$ and an anti-monotonic function $v(C)^-$.

[6] https://eos.com/blog/agricultural-cooperatives/ - accessed on 28 August, 2022.

- **Impact of cooperative membership:** Ma and Abdulai (2017) empirically examine a vector of outcome variables, including output price, gross income, farm profit, and return on investment (RoI). However, their evaluation is limited to analysing if the cooperative membership has significant impact on the farmers' income and profit.
- **Organising the farmers in groups:** Abdul-Rahaman and Abdulai (2020a) observe the effectiveness of organising the smallholder farmers into groups. However, they only aim to investigate the interrelationship between group membership and collective marketing decisions. Bonroy et al. (2019) observe that farmers tend to cooperate in a cooperative with fewer members. The authors observe that the likelihood of cooperating decreases as the number of farmers increases. Bonroy et al. further suggested that a solution to this problem is to create sub-divisions within the cooperative.
- **Combinatorial explosion:** Sarkar et al. (2020) propose an exact algorithm based on an exhaustive search to find the partition of the agents. This work formally defines the benefits of coalition formation of smallholder farmers and proposes a core-stable allocation method of the payoff. Sarkar et al. then evaluate the model using synthetic data generated by four statistical data distributions and show the results for a small set of agents (only 12 agents). This is a combinatorial optimisation problem, which can be solved optimally by an exhaustive search (Rahwan et al., 2015). However, this is infeasible for large-scale systems, as the number of possible coalition structures over n-agents - known as the *Bell number* $O(n^n)$. For an NP-hard problem, an exact algorithm cannot handle a large set of agents (Farinelli et al., 2013). On the other hand, an agricultural cooperative usually has hundreds of members (Ma & Abdulai, 2017 have shown results for 481 farmers). Hence, this approach becomes infeasible for a large set of agents.
- **Heterogeneity:** Moreover, Sarkar et al. (2020) considered the agents as homogeneous in the sense that every agent can form coalitions with every agent. However, one of the essential issues of concern related to agricultural cooperatives is the farmers' heterogeneity (Candemir et al., 2021). This heterogeneity may come from the farm size, the type of product a farmer produces, or other personal characteristics. Unlike Sarkar et al. (2020), we consider that the smallholder farmers have a heterogeneous demand for resources, and the farmers with similar production capability should earn the same revenue.
- **Similarity of the farmers:** Additionally, Bonroy et al. (2019) suggest that creating groups with similar characteristics can encourage the farmers to cooperate. In this context, in this work, we coordinate the smallholder farmers based on their similarity in resource requirements. To the best of our knowledge, the idea of forming farmers' coalitions based on similarity has remained unexplored so far. In this regard, we find CSG the best fit to address this issue as CSG finds the best way of partitioning an agent set into subsets.
- **Characteristic Function design:** Bistaffa et al. (2014, 2017), Farinelli et al. (2013) use an approach to the collective purchasing domain similar to our problem. The algorithm proposed by Farinelli et al. (2013) provides solutions for problems involving thousands of agents (more than 2500) in collective energy purchasing settings. However, this algorithm is not applicable to our model as the stopping criteria in our case do not guarantee termination. The algorithm by Farinelli et al. (2013) stops if there is no advantage in joining together the most suitable pair of coalitions, i.e., the algorithm stops if a pair of coalitions has a negative utility or the grand coalition is formed. In our case, merging two coalitions will not result in a negative value instead, the sum of the coalition values will gradually increase. Bistaffa et al. (2014, 2017) propose a characteristic function that is applied to graph-based models, and the solution approach is applicable to scenarios where the coalition values are calculated with an (m+ a)-like characteristic function. However, our model requires a different characteristic function that resembles $v(C) = v(C^+) - v(C^-)$, where *v(C)* increases as $v(C^-)$ decreases.

Consequently, we need a solution that defines a specific model by addressing all the issues mentioned above for an agricultural cooperative of smallholder farmers. The next section presents the problem statement for this problem.

3.2. Problem statement

We define the coordination problem of a set of farmers in an agricultural cooperative, i.e., a farmers' coalition formation problem *FCF*, as a cooperative game. Let F be the set of farmers $F = \{f_1, f_2, \dots, f_n\}$ in an agricultural cooperative, where n is the number of farmers in *F*. We denote a coalition $C = \{f_1, f_2, \dots, f_m\}$ as a non-empty subset of *F*, where $m \leq n$. A characteristic function $v : 2^F \Rightarrow \mathbb{R}$ maps each coalition C to a real value *v(C)*. A coalition structure *(CS)* over *F* is a partitioning of F into a set of disjoint coalitions $\{C_1, C_2, \dots, C_s\}$, where $s = \mid CS \mid$. In other words, *FCF* satisfies the following constraints:

- $C_i, C_j \neq \emptyset, i, j \in \{1, 2, \dots, s\}$
- two coalitions are mutually disjoint, i.e., $C_i \cap C_j = \emptyset$, for all $i \neq j$
- $\bigcup_{i=1}^{s}(C_i) = F$, where $C_i \in CS$

Definition 1. A characteristic function *v* is used to measure the utility of a coalition. When the value of a coalition does not depend on the actions of non-members, such settings are known as Characteristic Function Games (*CFGs*). The value of any coalition structure *CS* is defined by $V(CS) = \sum_{C_i \in CS}(v(C_i))$.

- $CS \in \prod^F \mid C \subseteq CS$
- $V(CS) = \sum_{C_i \in CS}(v(C_i))$
- the objective is to find the $CS^\star \in \prod^F$, where $CS^\star = argmax_{CS \in \prod^F} V(CS)$

Definition 2. An imputation is a payoff vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$ that is both group rational and individual rational. A payoff vector $\boldsymbol{x}$ is individual rational if no player needs to agree to receive less than what the player could obtain by working alone, $x_i \geq v(\{f_i\})$, for all $i = [1, n]$. A payoff vector is group rational if the total amount received by the players is the same as the amount that a grand coalition could obtain, $\sum_{i=1}^{n} x_i = v(C)$, where $v(C)$ is the amount obtained by the grand coalition.

4. System modelling

Modelling is one of the critical activities in understanding, designing, implementing, and operating any computational system. In this section, we present an abstract representation of the virtual enterprise model of an agricultural cooperative. Our work focuses on the horizontal grouping of smallholder farmers in an agricultural supply chain (see Fig. 1) to improve overall farm performance and on an individual level. Smallholder farmers are usually organised horizontally (see Fig. 2) into farmer groups, and through a solid and cohesive social network, they are incorporated into an agricultural supply chain (Abdul-Rahaman & Abdulai, 2020b). Farmer groups are then incorporated into local cooperatives organised in a multi-tiered fashion, known as a federated (i.e., centralised) structure. The local cooperative, in turn, may become a member of a regional cooperative, and a regional cooperative may also decide to be a member of an inter-regional cooperative. However, in this work, we only focus on the local cooperative (see the bottom layer of Fig. 2), which is centralised in nature.

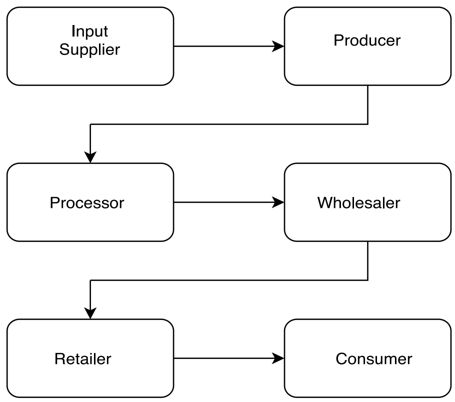


Fig. 1. Supply chain management in agriculture (Lazzarini et al., 2001).

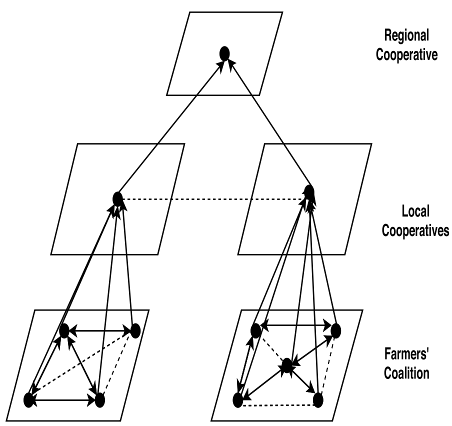


Fig. 2. Federated agricultural cooperative (Lazzarini et al., 2001).

An agricultural cooperative is composed of some farmers that produce crops. The main purpose of initiating agricultural cooperation is to generate more revenue for the farmers. For such a cooperation problem of heterogeneous and self-interested agents, an agent-based approach is a natural way of addressing the problem (Vasirani et al., 2011). Automation is more efficient for finding beneficial coalitions than people in complex organisations (Jennings & Dang, 2004). Against this background, we model the same issue as a multi-agent based approach where each agent acts on behalf of a farmer to maximise utility. Agent-based models potentially present a way to model real-life organisations as a complex system (Farmer & Foley, 2009). The coalition formation problem of smallholder farmers can be viewed at two levels: (i) individual farmers and (ii) cooperative (the farm level). We first present a model of an individual farmer and describe the underlying components of an agricultural cooperative; then, we describe the steps of forming a cooperative by connecting farmers according to their needs — these connections are not made randomly but involve a series of phases. Finally, we present a coalition model of the cooperative to find the best way of partitioning the farmers into groups in a fair way.

4.1. Model of an individual farmer

This section provides a model of an individual farmer in a cooperative. We assume that every farmer requires resources to produce crops and positively contributes to the cooperative in terms of crop production. Each farmer member of the cooperative is represented by an autonomous agent f_i in our model. Our model assumes that the agents are heterogeneous in terms of production capability and resource (e.g., seeds, fertilisers, water, electricity supply, transportation) requirements. The crop production capability depends on the landholding l_{f_i} of an agent f_i. Let there be r resource types required to yield a crop and the types of resources required are $Res = \{re^1, re^2, \dots, re^r\}$. Specifically, let an agent f_i expects to have outcome P_i. The production capacity can be expressed as a function of land and of resources: P_i : $l_{f_i} \times \sum_{k=1}^{r}(re^k) \in \mathbb{R} \geq 0$. The utility of an agent is the amount of revenue that an agent earns alone. However, in this work, we do not consider Z due to the unavailability of this data.

$v(\{f_i\}) = [PQ(\omega, Z) - W\omega]$(Abdul-Rahaman & Abdulai, 2020a)

$v(\{f_i\})$ is the value that an agent obtains when (s)he works alone, i.e., the value of a singleton coalition, P is the price of output, Q is the expected output level, ω is the input quantities, W is the vector of input prices, and Z is the combination of farm factors $Z_{farm} = \langle$farm size, access to credit, the average price per kg, gross farm revenue, yield, fertiliser, and chemical$\rangle$ and the household factors $Z_{farm} = \langle$age, education, gender, market perception$\rangle$ (Abdul-Rahaman & Abdulai, 2020a). The product of P and Q together is a function of (ω, Z), i.e., the yield depends on how much inputs are required and how the farm factors influence the yield. Abdul-Rahaman and Abdulai (2020a) show that the group members obtain higher yields, receive higher prices and generate higher gross farm revenues, resulting in significantly higher farm net revenues than non-members.

4.2. Model of a cooperative

In this section, we describe how the agents can be connected, given the model in Section 4.1, to form a cooperative of agents. We then transform this cooperative model into a coalitional game model.

An agricultural cooperative acts as a business enterprise, where the members cooperate for a common understanding of business as a single corporation (Fischer et al., 1996). When it comes to achieving a common goal, an appropriate structure is central to building an efficient team in an organisation (Gaston & DesJardins, 2008). According to Gaston and DesJardins, the key to effective team building is the underlying agent interaction topology (Gaston & DesJardins, 2008). According to Newman (2003), the type of connection among the people is important to understand the functioning of any human society. In agent organisations, agent-to-agent interactions may be limited by a social network. The social network may result from various factors, such as physical proximity, communication limitations, limited knowledge of other agents and their capabilities, trust relationships, and organisational structure (Bulka et al., 2007).

Cooperative game-theoretical scenarios in economics also address the issue of coalition formation in synergistic graphs (Myerson, 1977). In coalition formation theory, the coalitions are goal-directed and short-lived, i.e., they are formed with a purpose in mind and get dissolved when that purpose does not exist anymore (Rahwan et al., 2015). An agricultural cooperative could change over time; a member could choose to leave the cooperative. This situation will result in disbanding of two agents or the formation of new connections among the agents. As a result, the interconnection topology of the underlying graph considerably changes. Hence, in this section, we construct a graph $G = (F, E)$ based on the peer-to-peer connections between individual agents. In the context of an agricultural cooperative, peer-to-peer connections can be established based on two constraints mentioned by Voice et al. (2012): (i) the physical constraints and (ii) the social constraints. The next section presents two different formation of peer-to-peer connections based on these constraints. However, in this work, we consider the social constraints in constructing the graph of the agents. In the existing graph-based model for coalition formation, the coalitions are formed based on the membership restricted to coalitions composed of friends of friends (Vinyals et al., 2012). In order to adopt this approach to our problem, we need geo-spatial data of the land, which is presently unavailable.

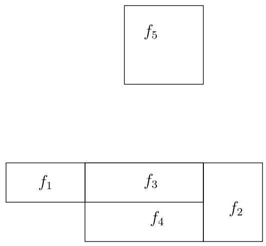


Fig. 3. Aerial view of farm land of smallholder farmers.

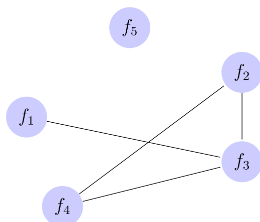


Fig. 4. Graphical representation of the land of the farmers. The lines show that the different land are adjacent to each other. The disconnected vertex f_5 has no connection with the rest of the vertices.

4.2.1. The physical constraints

In the context of an agricultural cooperative, a "physical constraint" is the proximity of the land of the cooperative members. The steps for constructing the graph are based on physical constraints. There are two steps:

- *Step 1 — Enrolment:* agents enrol themselves in the cooperative and announce the location of their land;
- *Step 2 — Graph Construction:* an undirected graph $G = (F, E)$ (the peer-to-peer connection between the agents) is constructed. In this graph, agents are considered as the vertices of the graph, and there exists an edge between two vertices if the land of those corresponding agents are adjacent (see Fig. 3).

Constructing a graph based on the physical constraints can result in a disconnected graph, for example, f_5 is a disconnected vertex in Fig. 4 (derived from Fig. 3). In the graph constrained coalition formation approach, a coalition C is feasible if a set of vertices form a connected subgraph. Therefore, a disconnected vertex in a disconnected graph will be treated as a singleton coalition by the graph constrained coalition formation approach. However, in an agricultural cooperative, farmers join the cooperative intending to work together with others, not of working alone. In the next section, we describe the steps of graph formation based on social constraints to avoid single coalitions.

4.2.2. The social constraints

In an agricultural cooperative, farmers can be grouped based on social constraints regarding various criteria, e.g., on the basis of: common needs, common problems, a specific category, shared interests, similarity in their need for agricultural resources, social affinity, homogeneous in social and economic status, among others (Rani & HRD, 0000). In this section, we present the steps for constructing a graph that is based on the social constraints (see Fig. 5). They are three steps:

- *Step 1 — Enrolment:* Agents enrol themselves in the cooperative and announce the area of the land l_{f_i} they own. The cooperative then determines the amount of each resource $re_i{}^r$ an agent needs and allocates them to the agent. The input quantities (here, the resource requirements) of an agent f_i is $\omega_i = Res_i = (re_i{}^1, re_i{}^2, \dots, re_i{}^r) \in \mathbb{R} \geq 0$.
We assume that the relationship between the landholding l_{f_i} and the amount of resources Res_i required by an agent follows a simple linear regression model. The independent variable in our model is l_{f_i}, and the dependent variables are the amount $re_i{}^r$ of each resource that an agent requires. The relationship between the dependent and independent variables is positive because both variables grow together. The value of each resource is determined by the following linear equation:

$$re_i{}^r = m^r.(l_{f_i}) + b^r$$

where, $m = \frac{change\ in\ re_i{}^r}{change\ in\ l_{f_i}}$, l_{f_i} is the amount of landholding of agent, b^r is the intercept.
Lets take an example of a cooperative with 5 agents $\{f_1, f_2, f_3, f_4, f_5\}$ and 3 resources $\{re^1, re^2, re^3\}$ seeds, fertiliser, and farm machinery. The amount of landholdings of the agents are $\{0.99, 0.38, 0.89, 1.25, 0.08\}$. The entries of Table 1 are calculated on the basis of requirement of the resources of each agent.
- *Step 2 — Normalisation:* The normalisation technique is used when the data has different scales and helps to compare the data with each other. We normalise the amount of resources (see Table 2). The agent with the highest amount of capability is assigned the highest weightage. To calculate the rest of the values, we use the statistical min–max normalisation technique (Yasir et al., 2018).

$$re^r{}_{norm} = \frac{re^r - re^r{}_{min}}{re^r_{max} - re^r{}_{min}}$$

- *Step 3 — Link Establishment:* The rationale behind grouping the agents in terms of similarity in the resource requirement ensures equal revenue earned by the agents. In this sense, given two agents $f_i, f_j \in F$, the similarity between their resource requirement Res_i and Res_j is measured by the absolute difference between them. Thereafter, the problem is represented by an undirected and unweighted graph $G = (F, E)$, where the agents are expressed by vertices, where the similar agents are connected by edges (de Oliveira Ramos et al., 2015). In this step, we first evaluate the similarity criteria between two agents (Step 3.1) and then decide which two vertices of graph G will be connected (Step 3.2). The following lines explain each sub-step:
  - **Step 3.1 — Similarity calculation:** We calculate the absolute difference between f_i and f_j. In order to do so, we extract the Manhattan distance between f_i and f_j from the normalised vectors of their resource needs. The lower the value of the Manhattan distance, the greater the similarity between the two agents. The minimum Manhattan distance $d_{f_{ij}}$ is just the absolute value of the differences:

$$\begin{aligned} d_{f_{ij}} &= |\text{ difference in } re^1 \,| + |\text{ difference in } re^2 \,| \\ &\quad + \cdots + |\text{ difference in } re^r \,| \\ &= |\, re_i{}^1 - re_j{}^1 \,| + |\, re_i{}^2 - re_j{}^2 \,| + \cdots + |\, re_i{}^r - re_j{}^r \,| \\ &= \sum_{x=1}^{r} |\, re_i{}^x - re_j{}^x \,| \end{aligned}$$

  - **Step 3.2 — Connection:** We connect two agents f_i and f_j with an edge e_{ij} if the agents are similar based on their

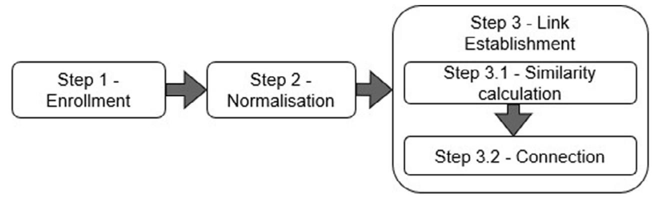


Fig. 5. Steps for modelling an agricultural cooperative based on the social constraints.

Manhattan distance $d_{f_{ij}}$.

$$e_{ij} = \begin{cases} 1 & \text{if } \omega_{f_{ij}} \leq k \\ 0 & \text{otherwise} \end{cases}$$

For our example, the Manhattan distances are (see Table 3): $d_{f_{12}} = 4.27$, $d_{f_{13}} = 0.69$, $d_{f_{14}} = 1.86$, $d_{f_{15}} = 6.38$ $d_{f_{23}} = 3.58$, $d_{f_{24}} = 6.13$, $d_{f_{25}} = 2.11$, $d_{f_{34}} = 2.55$, $d_{f_{35}} = 5.69$ and $d_{f_{45}} = 8.25$. The greater the Manhattan distance, the less similar the two agents are. Hence the most similar agent of f_1 is f_3 and the least similar agent is f_5. For agent f_2, the most similar agent is f_5 and the least similar agent is f_4. For agent f_3, the most similar agent is f_1 and the least similar agent is f_5. For agent f_4, the most similar agent is f_1 and the least similar agent is f_5. For agent f_5, the most similar agent is f_2 and the least similar agent is f_4.

5. Our solution: A coalitional model for a virtual agricultural cooperative

5.1. Introduction

We follow the process of coalition structure generation in order to find the best way to organise the agents into groups. The coalition formation process involves three steps: (i) coalition value calculation, (ii) coalition structure generation, and (iii) payoff division (Sandholm et al., 1999). In the following sub-sections, we specify how to execute each step in the context of a virtual cooperative formation of small-holder farmers. Also, we demonstrate that the described model has the following properties (Tran-Thanh et al., 2013):

- **Conciseness:** It is concise, i.e., it computes less number of coalitions than the classical representation of the CSG problem (see Claim 1);
- **Expressiveness:** It is fully expressive, i.e., it can represent any *CFGs* (see Claim 3);
- **Efficiency:** It is efficient, i.e., ii has the ability to develop efficient algorithms for solving the optimisation problems (see Claim 4).

5.2. Coalition value calculation

We evaluate the coalitional value in two steps, (i) generation and (ii) evaluation of the coalitions. In particular, we generate coalitions from the Table 3 in Section 4.2.2. Then we evaluate the coalition value, i.e., the utility of a coalition with the help of a characteristic function.

5.2.1. Coalition generation

Within a cooperative, an agent may require multiple resources. For example, an agent needs seeds, fertiliser, and a cultivator. We consider that each agent looks for potential partners for the coalition who are the most similar to them in Table 3. In graph theory, for a connected graph, a coalition is feasible if every two agents are adjacent in the induced subgraph, i.e., if the agents represented by the vertices form a connected subgraph (Bistaffa et al., 2017). However, our model only generates the coalitions of a few particular sizes from κ to $\kappa\prime$.

We take the following approach to enumerate the feasible coalitions:

- every individual agent forms a singleton coalition.
- there will be a feasible coalition of two or more agents if the agents are similar (see Eq. (1)). For our model, we only generate $\binom{F}{\kappa}$ to $\binom{F}{\kappa\prime}$, where κ is the size of a coalition, and each coalition of a size is the union of two coalitions of previous sizes, i.e., $\binom{F}{\kappa} = \binom{F}{\kappa-i} \cup \binom{F}{\kappa-j}$, for all $i, j = [1, m-1]$.

Claim 1. *Here we analyse the conciseness of our model since we do not need to store all* $2^{|n|}$ *values.*

Table 1
Resource requirement for 5 agents.

		Resource		
		re^1	re^2	re^3
Agents	f_1	7792.58	8126.38	10 799.74
	f_2	3006.70	3135.49	4166.98
	f_3	7016.13	7316.67	9723.66
	f_4	9877.05	10 300.13	13 688.60
	f_5	638.45	665.80	884.84

Table 2
Resource requirement for 5 agents after normalisation.

		Resource		
		re^1	re^2	re^3
Agents	f_1	0.63	0.63	0.63
	f_2	–0.79	–0.79	–0.79
	f_3	0.40	0.40	0.40
	f_4	1.25	1.25	1.25
	f_5	–1.49	–1.49	–1.49

Table 3
Similarity of 5 agents based on their resource requirement.

		Agents				
		f_1	f_2	f_3	f_4	f_5
Agents	f_1	–	4.27	0.69	1.86	6.38
	f_2	4.27	–	3.58	6.13	2.11
	f_3	0.69	3.58	–	2.55	5.69
	f_4	1.86	6.13	2.55	–	8.25
	f_5	6.38	2.11	5.69	8.25	–

In the traditional approach, for 5 agents there will be a total of $2^5 = 32$ possible coalitions. However, in our example, the number of possible coalitions is less than $2^n - 1$ (see Proposition 1). Moreover, our approach does not generate singleton coalitions, grand coalitions and coalitions larger than a particular size ($\kappa\prime$). Due to the superadditivity nature of the problem, the larger sized coalitions will always be beneficial; however, that will make the system unstable. Instead, we choose the coalition size with a trial and error method to maintain the system's balance. Table 4 shows the feasible coalitions of sizes 2 to 4 for 5 agents. When we create coalitions of size κ, for each agent, our approach creates coalitions with κ number of closest agents. For example, for lower bound $\kappa = 2$ and upper bound $\kappa\prime = 4$, our algorithm generated the 2-sized coalitions $\{f_1, f_3\}, \{f_1, f_4\}, \{f_2, f_5\}$. While generating a κ-sized coalition with agent f_i in it, the algorithm finds other $\kappa - 1$ agents closest to agent f_i. Hence, the 3-sized coalitions are $\{f_1, f_3, f_4\}, \{f_2, f_3, f_5\}$. The forbidden coalitions of size 2 in our algorithm are $\{f_1, f_2\}$ and $\{f_2, f_4\}$. In this approach, since not all coalitions are feasible to be formed, we do not enumerate all coalitions. Though, this constraint lowers the number of coalitions from exponential (i.e. 2^n), to polynomial (i.e. $O(n^k)$), the problem remains NP-hard (Bistaffa et al., 2017).

Proposition 1. *The number of possible coalitions in our model is polynomial and not exponential.*

Proof. Our model finds coalitions of a few given sizes for each agents from κ to $\kappa\prime$, i.e., for each agent, $(\kappa\prime - \kappa)$ number of coalitions are created. Hence, in the worst case, possible number coalitions generated in our case is: $n(\kappa\prime - \kappa) = O(n)$, where $(\kappa\prime - \kappa)$is a constant.

5.2.2. Coalition evaluation

In this section, we define the characteristic function to evaluate the potential of the coalitions. The characteristic function estimates the total revenue a group of agents can earn for their joint venture (the

Table 4
All feasible coalitions for 5 agents — traditional approach vs. proposed approach. For our approach we check coalitions for different κ and $\kappa\prime$ values (due to lack of space in this table we denote an agent with an integer i instead of f_i).

Coalition size	Coalitions - traditional approach	Coalition - our model with $\kappa = 2$ and $\kappa\prime = 3$	Coalition - our model with $\kappa = 2$ and $\kappa\prime = 4$
1	{1}, {2}, {3}, {4}, {5}	∅	∅
2	{1, 2}, {1, 3}, {1, 4}, {1, 5}, {2, 3} {2, 4}, {2, 5}, {3, 4}, {3, 5}, {4, 5}	{1, 3}, {1, 4}, {2, 5}	{1, 3}, {1, 4}, {2, 5}
3	{1, 2, 3}, {1, 2, 4}, {1, 2, 5}, {1, 3, 4}, {1, 3, 5}, {1, 4, 5}, {2, 3, 4}, {2, 3, 5} {2, 4, 5}, {3, 4, 5}	{1, 3, 4}, {2, 3, 5}	{1, 3, 4}, {2, 3, 5}
4	{1, 2, 3, 4}, {1, 2, 3, 5}, {1, 2, 4, 5} {1, 3, 4, 5}, {2, 3, 4, 5}	∅	{1, 2, 3, 4}, {1, 2, 3, 5}
5	{1, 2, 3, 4, 5}	∅	∅

coalition value $v(C)$). We present the expected revenue of any coalition C:

$$v(C_j) = [PQ_{C_j} - W\omega_{C_j}] \tag{1}$$

In our model, the coalition value depends on two factors (Abdul-Rahaman & Abdulai, 2020a): (i) the joint expenditure for the inputs of a coalition C_j and (ii) the selling price of the crops, where P is the market valuation of the crop, Q is the expected collective output of each member of coalition C_j, W is the input price and ω is the input quantity. Furthermore, we consider the following constraints to find the coalition value with Eq. (1).

Almost every farmer suffers from post-harvest loss. Inadequate transport infrastructure due to bad roads, contamination from repeated loading and unloading, and lack of refrigerators are some of the reasons for the loss of farm production. This is modelled as:

Constraint 1. *The output of each agent Q_{f_i} and each coalition Q_{C_j} should be positive:*

$$Q_{f_i} > 0,$$

$$Q_{C_i} > 0$$

Boundaries within the land of smallholder farmers lead to a decrease in cultivable land. Members of the agricultural cooperative may agree to pool their land for joint cultivation on a cooperative basis, which increases the size of the operational landholding (Mahendra Dev, 2014). This is modelled as:

Constraint 2. *The sum of individual landholdings is less than or equal to the total landholding of a coalition. The sum of individual landholdings is less than that of a coalition C_j if the lands of the coalition members are adjacent to each other.*

$$l_{f_i} + l_{f_j} \le l_{C_j}$$

In our model, we avoid forming singleton coalitions as it violates the basic idea of cooperative formation. Also, we restrict larger coalitions. Although larger coalitions may perform their tasks more efficiently, communication and coordination overhead grow beyond the size of a particular coalition (Sarkar et al., 2022). This is modelled as:

Constraint 3. *No agent should work alone or the coalition size should not exceed a certain limit.:*

$1 < |C_{f_i}| < \kappa\prime$, *where $\kappa\prime$ is the upper limit of coalition size*

From a buyer's point of view, quantity-based discounts provide a considerable incentive that makes them form coalitions and take advantage of lower prices without ordering more than their actual demand (Li et al., 2003). Group buying decreases the unit price of a product as the number of buyers increases.

Constraint 4. *The unit price of each resource for an agent when (s)he is working alone is greater than the unit price of resources (ρ^r) for an agent when (s)he is working in a coalition.*

$$\forall \rho^r \in W, W_{C_j} < \sum_{f_i \in C_j} W_{f_i}$$

Proposition 2 (*Price Reduction*). *Purchasing the resources in a group is more beneficial than buying them individually.*

Let us consider the example again: a group of agents is now purchasing the resources. The unit price schedule of the resources is: $\rho^r : N \to N$, or $\rho^r : N \to R$, presented through a descending step function. The amount of discount depends on the number of units sold together (Yamamoto & Sycara, 2001). This function is on either natural or real numbers based on the character of the resource. For example, seeds or fertiliser can be ordered in fractions. However, a farmer cannot lend a fraction amount of a tractor. Section 5.2.2 presents the step function for the price reduction scheme.

$$\rho^r = \begin{cases} \rho^r(|C_{f_i}| = 1) & \text{for } |C_{f_i}| = 1 \\ \rho^r(|C_{f_i}| = a) & \text{for } |C_{f_i}| = a \\ \rho^r(b \le |C_{f_i}| \le c) & \text{for } b \le |C_{f_i}| \le c \\ \rho^r(d \le |C_{f_i}|) & \text{for } d \le |C_{f_i}| \end{cases}$$

We describe our characteristic function with a simple example. Assume there are 3 agents and only 1 input item with a price schedule described in Section 5.2.2. For instance, if 1 item is sold out, the unit price is 100. If 2 items are sold out then the unit price is 95, and if 3 items are sold out then the unit price is 85.

Now the value of the singleton coalitions will be: $v(\{f_1\}) = [PQ_{f_1} - 100 \times \omega_{f_1}]$, $v(\{f_2\}) = [PQ_{f_2} - 100 \times \omega_{f_2}]$ and $v(\{f_3\}) = [PQ_{f_3} - 100 \times \omega_{f_3}]$. For size-2 coalitions, the values will be: $v(\{f_1, f_2\}) = [PQ_{\{f_1,f_2\}}(l_{\{f_1,f_2\}}) - 95 \times \omega_{\{f_1,f_2\}}]$, $v(\{f_1, f_3\}) = [PQ_{\{f_1,f_3\}}(l_{\{f_1,f_3\}}) - 95 \times \omega_{\{f_1,f_3\}}]$ and $v(\{f_2, f_3\}) = [PQ_{\{f_2,f_3\}}(l_{\{f_2,f_3\}}) - 95 \times \omega_{\{f_2,f_3\}}]$. Finally, for the size-3 coalition, the value will be: $v(\{f_1, f_2, f_3\}) = [PQ_{\{f_1,f_2,f_3\}}(l_{\{f_1,f_2,f_3\}}) - 85 \times \omega_{\{f_1,f_2,f_3\}}]$.

Claim 2. *If the number of agents in a coalition increases, the possibility of profit increases.*

The characteristic function has two parts $x = (P \times Q)$ and $y = (W \times \omega)$.

$$\frac{x}{y} = \frac{(P \times Q)}{(W \times \omega)} = \frac{\textit{market price of crop} \times \textit{Expected amount of crop}}{\textit{market price of the inputs} \times \textit{Expected amount of inputs}}$$

Now, the term "market price of the inputs" in the denominator decreases as the number of agents n increases (see Section 5.2.2), while the "market price of the crop" remains constant. For the other two terms, "Expected amount of crop" is greater than "Expected amount of inputs". Hence, the numerator is always greater than the denominator as the number of agents increases.

Farmers join the coalitional model to earn more. Purchasing resource helps them lower their expenditure. In our model, we only allow beneficial coalitions to form.

Constraint 5. *The value of each coalition is always positive.*

$$C_i > 0, \textit{where } V(CS) = \sum_{C_j \in CS} v(C_j)$$

Claim 3. *Here we analyse the expressiveness of our model, i.e., for any instance of CFG (see Definition 1), there exists an equivalent coalitional game of an agricultural cooperative FCF on the same set of agents F, where for a feasible coalition C, $v_{FCF}(C)$ is equal to the value of gain within CFG, i.e., $v_{CFG}(C)$.*

Consider any two agents f_i and f_j, where they have different capabilities to produce a crop. In our model, coalition value is decided by the net revenue that a coalition can earn, and the coalition values of two different coalitions are always unique. Because Q_{f_i} is unique for each agent, so does the ω_{C_j}. Hence, $v(C) = [PQ - W\omega]$ is unique for each coalition and assigns a real value to each coalition.

5.3. Coalition structure generation

5.3.1. Introduction

A coalition structure $CS = \{C_1, C_2, \ldots, C_s\}$ is formed so that each coalition C_i in the *CS* as well as the *CS* draw positive utility. The CSG problem can be formulated as:

$$Maximise \sum_{C \in \{\text{feasible coalition}\}} v(C) \cdot x_C \tag{2}$$

such that each agent joins at most one coalition:

$$\sum_{C \in \{\text{feasible coalition}\}} x_C = 1 \tag{3}$$

x_C is a binary decision variable, where $x_C \in \{0, 1\}$ and the *feasible coalition* is the set of possible coalitions we obtained from the *Coalition Generation* step (see Section 5.2.1).

The next section presents the proposed algorithm to solve the CSG problem in an agricultural cooperative.

5.3.2. Algorithm details

This section first details how our proposed algorithm works, in particular, how the similarity criteria is used to compute the feasible coalitions, and how the heuristic based approach works to find the final solution.

Generating the Feasible Coalitions

The algorithm (see Algorithm 1) starts by calculating the similarity ($sim_{F\times F}$) among the agents using the resource requirement of each agent (lines 1–3 of Algorithm 1). For the given lower and upper bound of coalition size, κ and $\kappa\prime$ respectively, the algorithm creates the coalitions based on the similarity matrix Sim (lines 4–8). E.g., Table 3 presents the similarity among the agents using the Manhattan distance. The algorithm calculates the feasible coalitions using this Sim matrix (see Table 4).

Proposed Heuristic Approach

In order to find a coalition structure $CS_{SetCover()}$, the algorithm uses the *SetCover()* function with the previously generated coalitions (line 9). However, the coalition structure $CS_{SetCover()}$ generated here does not have disjoint coalitions. After this, the algorithm detects all the agents $f_i{}^{intersect\_agent}$ that are present in two or more than two coalitions. Each of these agents $f_i{}^{intersect\_agent}$ is then added to a coalition where the rest of the agents are the most similar to agent $f_i{}^{intersect\_agent}$ (lines 10–12). At this stage, the algorithm gets a coalition structure that has disjoint coalitions and calculates the *V(CS)* value (lines 13–14). If the *V(CS)* has any coalition with a negative value, the algorithm performs one final step to remove any negative valued coalition. Each agent of a negative valued coalition is added to the coalition with least *v(C)* value (lines 15–17) and the algorithm returns the final *CS* with updated *V(CS)* value (line 19).

Algorithm 1: Pseudo Code for the Proposed Algorithm

Input: n=number of agents, l_f=land of the agents, r= number of resource, W_r= price of each resource, ω= resource requirement, κ=lower bound of coalition size, $\kappa\prime$=upper bound of coalition size

Output: Disjoint coalition structure with approx $O(\frac{n}{\kappa})$ number of coalitions

1 Create set F with n no. of agents.
2 Take land value l_{f_i} for each agent and calculate their input quantity vector ω_{f_i} (resource requirement).
3 Create $Sim = sim_{F\times F}$ a $(n \times n)$ matrix for n no. of agents.
4 **for** $i \leftarrow \kappa$ **to** $\kappa\prime$ **do**
5   **for** $j \leftarrow 0$ **to** *length(Sim)* **do**
6   Compute the coalition C based on the values of $sim_{F\times F}$
7   **end**
8 **end**
/* for each f_i and for each coalition size κ, there will be (κ-1) no. of agents that are similar to f_i. */
9 Find a coalition structure $CS_{SetCover()}$ using *SetCover()* function with the coalitions C obtained in step 6.
/* the coalitions in the $CS_{SetCover()}$ are not disjoint, there are some agents ($f^{intersect\_agent}{}_i$) who are present in more than one coalitions. */
10 **for** *all coalitions in* $CS_{SetCover()}$ **do**
11   First remove all $f_i{}^{intersect\_agent}$ from all coalitions Add $f_i{}^{intersect\_agent}$ to the coalition where other agents f_i are most similar to $f_i{}^{intersect\_agent}$; // duplicate removal
12 **end**
13 **for** *each C in* $CS_{SetCover()}$ *after duplicate removal in step 10-11* **do**
14   use equation (1) to compute the *v(C)*
15   **for** *each C with -ve v(C)* **do**
16   assign the agents to the coalitions with least *v(C)* values.
17   **end**
18 **end**
19 Return the final *CS* and *V(CS)* using equation (1).

5.4. Payoff division

To share the surplus within a coalition, a method is needed to distribute the surplus among the agents in a fair way that is consistent with the characteristic function (Conitzer & Sandholm, 2004). In an agricultural cooperative the surplus is distributed among the farmer members in the ratio of the members' contribution to the cooperative (International labour office, 1984). In this model, we consider that a member's contribution is equivalent to the land the member owns. Moreover, another type of payment is made to the cooperative members: payment for the labour contributed toward cultivation and other activities. However, in this work, we do not consider the distribution of payments for labour. A coalition C_j is said to have a positive surplus if $v(C_j) > 0$, and the payoff vector of a coalition is $x(C_j) = (x_1{}^j, x_2{}^j, \ldots, x_n{}^j)$, where $\sum_{i=1}^{n} x_i{}^j = v(C_j)$. The surplus sharing rule is:

$$x_i(C_j) = V(CS) \times \frac{l_{f_i}}{\sum_{f_i \in C_j} l_{f_i}}$$

Proposition 3. *This imputation is Individual rational, i.e.,* $x_i(C_j) \geq v(\{f_i\})$.

Proof. We proof this by the method of contradiction, lets assume: $v(\{f_i\}) > x_i(C_j)$

$$v(\{f_i\}) > x_i(C_j)$$
$$v(\{f_i\}) - x_i(C_j) > 0$$

$$(PQ_{f_i} - W_{f_i}\omega) - (PQ_{f_i} - W_{f_{ij}}\omega) > 0$$

where PQ_{f_i} is common in both the cases and $W_{f_{ij}}$ is the vector of price applicable for agent f_i when that agent is a participant of coalition C_j.

$$\begin{aligned} -W_{f_i}\omega + W_{f_{ij}}\omega &> 0 \\ W_{f_{ij}}\omega - W_{f_i}\omega &> 0 \\ W_{f_{ij}}\omega &> W_{f_i}\omega \end{aligned}$$

If the input quantity vector remains unchanged, then $W_{f_{ij}}\omega > W_{f_i}\omega$ is false. Because according to our model, the unit price applicable for an agent decreases when the resources are purchased collectively. Hence, $W_{f_{ij}}\omega$ is lesser than $W_{f_i}\omega$. So, we can say that our initial assumption was false, i.e., $x_i(C_j) > v(\{f_i\})$.

Claim 4. *Here we analyse the efficiency of our algorithm, i.e., our proposed algorithm can efficiently solve the optimisation problem.*

Proposition 4. *Approximation ratio of our approach is $\frac{n}{\lfloor \frac{n}{\kappa} \rfloor}$, i.e., the solution of our approach $V(CS')$ is within a bound from the optimal solution $V(CS^\star)$.*

Proof. Suppose our approach has found a coalition structure CS' and let $CS^\star$ be the optimal coalition structure. Given that the lower size limit of the coalitions are κ, at most $\lfloor \frac{n}{\kappa} \rfloor$ coalitions will be formed. Suppose each coalition has an average coalition value v_c. In this case, our $V(CS')$ will be $\lfloor \frac{n}{\kappa} \rfloor \times v_c$. Let us assume that $C^\star$ is the highest valued coalition with coalition value $v(C^\star)$ that is formed by the Dynamic Programming (DP), i.e, the grand coalition, where $V(CS^\star) = v(C^\star)$. Moreover, according to Proposition 3, $v(C^\star) > n \times v_s$, where v_s is the value of singleton coalitions.

$$\begin{aligned} \frac{V(CS')}{V(CS^\star)} &= \frac{V(CS')}{v(C^\star)} \\ &\approx \frac{\lfloor \frac{n}{\kappa} \rfloor \times v_c}{n \times v_s} && [v(C^\star) > n \times v_s] \\ & && [V(CS') > \lfloor \frac{n}{\kappa} \rfloor \times v_c] \\ &= \frac{\lfloor \frac{n}{\kappa} \rfloor}{n} \times \frac{v_c}{v_s} && [\frac{v_c}{v_s} \gg 1] \\ &> \frac{\lfloor \frac{n}{\kappa} \rfloor}{n} \end{aligned}$$

6. Methodology of the experimental evaluation

In this section, we describe the goals of the evaluation in Section 6.2, the experimental setup in Section 6.3, the procedures of the evaluation in Section 6.5, the data characteristics that we use to evaluate the model under general and specific settings in Section 6.6, the data generation in Section 6.7 and finally, the description of the metrics used for the evaluation in Section 6.4.

6.1. Simulation model

We describe the simulation model in Section 6.5. The simulation model is designed to mimic the formation of a real-life agricultural cooperative of smallholder farmers. Here, the smallholder farmers have less than 2 ha landholding (Rapsomanikis, 2015). They join a cooperative to reduce operational cost and increase their profit. To achieve this goal, these smallholder farmers are divided into coalitions based on their similarity in resource requirements.

6.2. Goals of the evaluation

The main goals of the empirical evaluation are:

- to compare the quality of our solution with the exact algorithm DP — for a total of agents from 15 to 23;
- to evaluate the speed-up achievement of our system compared to the exact algorithm DP — for a total of agents from 15 to 23;
- to evaluate the run-time performance of our algorithm for larger agents sets, to show that our algorithm can be scaled up;
- to evaluate the effectiveness and impact of coalition formation in an agricultural cooperative setup — for a total of agents from 15 to 23.

6.3. Experimental setup

We used an Intel (R) Xeon (R) CPU E7-4830 v3, 2.10 GHz environment, Linux operating system (64-bit) with 160 GB of RAM.

All the algorithms were implemented in *PYTHON 3.8 (IDLE)*. The basic libraries to compute coalitions and combinations are imported from *PYTHON 3* standard libraries.

6.4. Metrics

This section presents the metrics to evaluate the proposed model (Section 4) and algorithm (Algorithm 1). The evaluation of the model will validate the proposed multi-agent based modelling of a virtual agricultural cooperative of smallholder farmers. The evaluation of the proposed algorithm aims to investigate its efficiency.

This section has two parts, the first presents the metrics used for the evaluation of the model; the second the rationale for the choice of metrics used.

6.4.1. Explanation of the metrics used

We use the metrics runtime comparison, solution quality, and scalability to evaluate the performance of the algorithm:

- **Runtime comparison:** It shows the time efficiency of the proposed algorithm, i.e., how faster the algorithm can generate a solution compared to the state-of-the-art algorithm (Bistaffa et al., 2017; Hussin & Fatima, 2016). To evaluate the runtime performance, we compare the runtime of the proposed algorithm with the runtime of the exact algorithm DP.
- **Solution quality:** Comparing the solution generated by our approach with the state-of-the-art approach shows that the proposed algorithm can return a solution closer to the best solution (Bistaffa et al., 2017; Hussin & Fatima, 2016). To evaluate the solution quality we measure how close the *V(CS)* value generated by our algorithm is to the optimal *V(CS)* value (Hussin & Fatima, 2016). Let CS_{FCF} be the coalition structure returned by our algorithm, and CS_{Opt} be the exact optimal solution. The solution quality for our algorithm is measured as follows:

$$\frac{V(CS_{FCF})}{V(CS_{Opt})} \times 100 \quad (4)$$

- **Scalability:** The scalability metric checks if the proposed algorithm can create a solution within a reasonable time as the number of agents increases. To examine the scalability of the proposed algorithm we evaluate it with $n \in \{25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500\}$ to demonstrate the scalability of the solution approach. We then measure the time that our algorithm takes to return a solution. For this evaluation, we provided a lower limit $\kappa = 3$; and an upper limit $\kappa\prime = 5$.

We use the metric individual gain to evaluate the correctness of the model.

- **individual gain:** The individual gain metric shows if the proposed multi-agent based coalitional model benefits the stakeholders and draws extra revenue to them compared to the revenue earned in a non-coalitional model. To evaluate the individual gain of the agents, we show the increase in revenue when the agents join the coalitional model described in Section 4.2.2 instead of working alone. We calculate the revenue of each agent using Section 5.4.

6.4.2. Rationale for the choice of metrics

The decision to choose the metrics "runtime comparison" and "solution quality" is based on the work of Bistaffa et al. (2017) and Hussin and Fatima (2016). The decision to choose the metric "scalability" is based on the work of Alam et al. (2013) and Bistaffa et al. (2017). The decision to choose the metric "individual gain" is based on the work of Alam et al. (2013). Alam et al. show the usefulness of energy exchange between homes in a community to reduce battery usage. In this work, we show the usefulness of cooperation via coalition formation in an agricultural cooperative to increase social welfare.

Bistaffa et al. (2017) also use the metrics: (i) bounding technique effectiveness — which measures how much search space is explored; (ii) Anytime performance — which shows whether the algorithm can return a valid solution if it is stopped before completion; and (iii) Parallelism — which shows the ability of the algorithm to perform more than one computation simultaneously. We do not use any of these metrics to evaluate our model. In fact, the metric:

- "Bounding technique effectiveness" is not used as our model does not need to explore the search space to find the coalition structure.
- "Anytime performance" is not used as our model is not a real-time system with deadlines.
- "Parallelism" is not used for now. We may use it in future work.

6.5. Procedures

We programmed an algorithm to implement the model of a virtual agricultural cooperative, a coalitional model that makes it possible to increase the social welfare of the cooperative members. We ran simulations to evaluate the effectiveness of the algorithm. Our algorithm[7] takes a heuristic based approach to find the coalition structure. However, our data preprocessing steps (described in Section 4.2.2) use an exact method to find coalitions of similar agents.

We conducted the tests according to the metrics defined in Section 6.4.

- For a small-scale simulation (from a total of agents of 15 to a total of 23 agents[8]), we use the metrics "runtime comparison", "solution quality", and "individual gain". We compare our algorithm against the state-of-the-art exact algorithm DP in terms of "runtime comparison" and "solution quality". For "individual gain" we compare the results we get from the payoff division method against the payoff division method of a non-coalitional model.
- For a large-scale simulation (from a total of agents of 25 to a total of 500 agents[9]), we use the metrics "scalability". Here we show that our algorithm can create a solution within a reasonable time frame for hundreds of agents. Since DP cannot generate results for more than 30 agents (see Farinelli et al., 2013), it is not possible to compare the $V(CS)$ values generated by our algorithm and DP.

[7] Implementation of our algorithm is publicly available at https://github.com/samriddhi2010/FCF.git.

[8] We started the simulations with a total number of agents as 15 since this is the minimum requirement for forming any cooperative society. (https://megcooperation.gov.in/faqs.html - accessed on 30 March, 2022).

[9] Ma and Abdulai (2017) show the impacts of agricultural cooperative membership on 481 smallholder farmers, so we have chosen the maximum number of agents as 500.

6.6. Parameter description

In this section, we describe the parameters of the data that we use to evaluate the coalition values using the characteristic function (Eq. (1)).

- **Number of Agents (n):** We initially run the experiments for the exact algorithm DP and the proposed algorithm using a community of 15 to 23 agents. We then run the experiment only for our algorithm on a community of 500 agents.
- **Size of landholding** (l_{f_i}): Each agent of the agricultural cooperative holds land with less than 2 ha of size.
- **Market value of the crop (P):** We assume that our farmers produce rice. We take the price schedule of paddy in the state of West Bengal in India. The average price of the paddy in that region is Rs. 14.30/kg.[10]
- **Expected outcome of an agent (Q):** Production of any crop is proportional to the amount of land one has. In this experiment, we consider that our farmers produce rice. In India, the average productivity of rice is 5400 kg/ha.[11]
- **Input quantities** (ω): According to the data provided on The Tamil Nadu Agricultural University's portal,[12] we consider that the amount of the input quantities are directly proportional to the size of the land.
- **Input price (W):** We take the data of Paddy cultivation in the year 2014–15. Seed 7844 INR/ha, Fertilisers and Manures 8180 INR/ha, Machine Power 10 871 INR/ha.[13]
- **Discounts:** We use the price schedule shown in Fig. 6. The horizontal axis shows the amount of land vs. the price reduction we have considered in this work. We have considered the price reduction on the sum of three resources together.
- **Coalition size** (κ): We create coalitions of agents with similar needs of sizes ranging from κ to $\kappa\prime$. For the total of agents 15 to 23 we use $\kappa = 3$ and $\lfloor \sqrt{n} \rfloor$ and $\kappa\prime = 5$ and $\lfloor \frac{n}{2} \rfloor$. For agents from 25 to 500 we use $\kappa = 3$ and $\kappa\prime = 5$.

6.7. Data generation

The raw data is provided by the Ministry of Agriculture and Farmers Welfare, Govt. of India.[14] We take the data of the agriculture census (the year 2015–2016) and only use the data of a "Tehsil" of the state "West Bengal", India.[15] We extract the number of male farmers[16] under each category of size of operational landholdings (see Table 5) of those farmers of "Tehsil" Habra-I in the district North 24 Parganas, West Bengal, India. The dataset is then created with only the data that concerns the farmers whose land is less than 2 ha.

We created the landholding data of the agents using the Uniform Distribution for each range (smaller than 0.5, from 0.5 to 1.0, and from 1.0 to 2.0), and then we shuffle the sample data to make the data randomly distributed. We use Python's Random uniform method to create the data as this method only requires the lowest and highest sample value. The samples created by the Uniform method are uniformly

[10] https://enam.gov.in/web/dashboard/trade-data accessed on 12 January, 2022.

[11] https://agritech.tnau.ac.in/agriculture/agri_costofcultivation_rice.html - accessed on 2 January, 2022.

[12] https://agritech.tnau.ac.in/agriculture/agri_costofcultivation_rice.html - accessed on 30 March, 2022.

[13] https://agritech.tnau.ac.in/agriculture/agri_costofcultivation_rice.html - accessed on 30 March, 2021.

[14] https://agcensus.nic.in - accessed on 30 March, 2022.

[15] https://agcensus.dacnet.nic.in/TL/TehsilT1table1.aspx - accessed on 30 March, 2022.

[16] Here, considering the statistics of male farmers do not introduce any bias in the evaluation of our algorithm and the statistics available for the male farmers was enough to conduct our study.

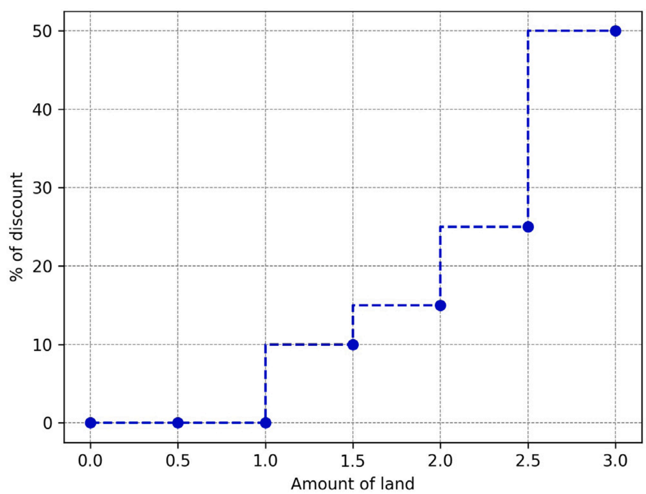


Fig. 6. A sample price schedule for the bulk purchasing.

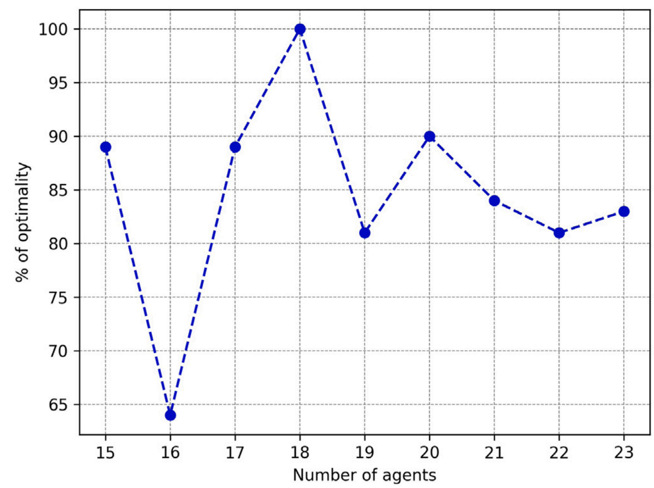


Fig. 7. A comparison of $V(CS)$ values of the proposed approach with the optimal one using equation (4).

Table 5
Landholding vs. size (less than equal to 2 ha of a "Tehsil" of state West Bengal, India).

Size of land (in ha)	Number of farmers	% total of farmers
<0.5	2752	36.36%
0.5–1.0	3220	42.54%
1.0–2.0	1597	21.09%
Total	7569	100%

Table 6
Time taken by DP vs. the proposed approach for a total of agents from 15 to 23.

No. of agents	% of optimality	Time taken by DP (in sec)	Time taken by our approach (in sec)
15	89%	31.107	0.006
16	64%	97.845	0.006
17	89%	304.091	0.005
18	100%	969.127	0.008
19	81%	3008.845	0.007
20	90%	9472.370	0.009
21	84%	29 484.043	0.010
22	81%	91 522.574	0.009
23	83%	269 376.294	0.010

distributed over the half-open interval, where each sample includes the lowest value of the interval but excludes the highest one.

The size of landholding and number of farmers are presented in Table 5: 36.36% of the agent set is of landholdings below 0.5 ha, 42.54% of the agent set is of landholdings from 0.5 to 1.0, and 21.09% of the agent set is of landholdings from 1.0 to 2.0 ha.

6.8. Benchmark

To the best of our knowledge, Sarkar et al. (2020) first propose the concept of coalition formation for smallholder farmers in an agricultural cooperative (Sarkar et al., 2022). Sarkar et al. use an exhaustive search method to find the solution and show the results using four statistical data distributions for a small set of agents. The exhaustive search method cannot find a solution for more than 12 agents. Therefore, in this work, we do not compare our solution with the solution of exhaustive search. Instead, we choose DP to compare the performance of our algorithm. We test DP and our algorithm using the same dataset, as comparing the algorithms with the same dataset is necessary (Bistaffa et al., 2017). With our experimental setup, DP can generate results up to 23 agents. DP is not dependent on the distribution of coalition values (de Oliveira Ramos et al., 2015) and capable of generating an optimal solution if $2^n - 1$ number of coalition values are given. Hence, we generate $2^n - 1$ number of coalitions to test the runtime and solution quality of DP using the dataset described in Sections 6.6 and 6.7. We further justify the rationale of not comparing our algorithm with other state-of-the-art algorithms: (i) *CFSS* by Bistaffa et al. (2017) use energy consumption profile that represents energy consumption of the consumers, (ii) de Oliveira Ramos et al. (2015) use the data of plug-in electric vehicle.

Therefore, such a comparison between the two approaches would not be justified given that the modelling, characteristic function, dataset are different from each other (de Oliveira Ramos et al., 2015; Tran-Thanh et al., 2013).

7. Experimental evaluation

7.1. Results

7.1.1. Runtime

Table 6 shows the time taken by the DP algorithm vs. our approach. For 23 agents, DP takes more than 74 h (269 376.294 s), while our approach takes 0.010 s.

7.1.2. Solution quality

Fig. 7 shows the comparison between the *V(CS)* value of the proposed approach and the *V(CS)* value of the optimal algorithm for agents from 15 to 23. We can see that our solution generates the worst result for 16 agents (64% of the optimal), and the best result for 18 agents (100% of the optimal).

7.1.3. Scalability

Fig. 8 shows the comparison of the computational time required to compute *V(CS)*. For an instance with 500 agents, our algorithm takes 95 s to provide a solution.

7.1.4. Individual gain

Fig. 9 shows the gain of the agents when they work individually vs. when they join the coalitional model. The x-axis represents the total value of land owned by the agents, and the y-axis represents the agent's revenue. The top most point of a line indicates the amount received by the agent (using the payoff division rule - Section 5.4) while working in a coalitional model. The bottom-most point of the line indicates the amount received by the agent while working alone.

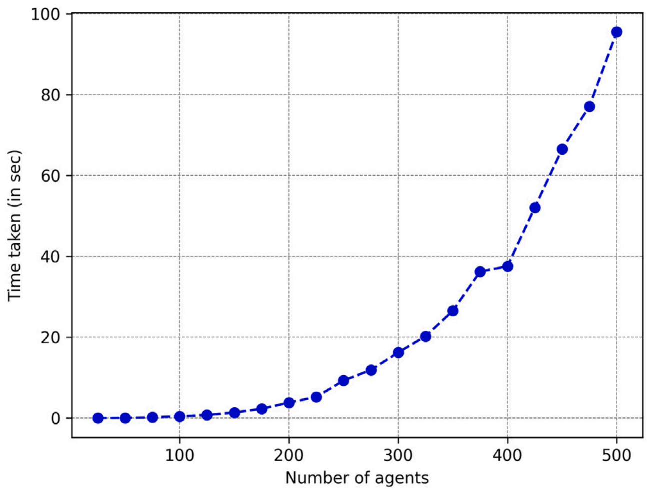


Fig. 8. Scalability of the approach: Computation time vs. number of agents,.

7.2. Analysis of the results

In this section, we analyse the results obtained in the evaluation of our model for coalition formation of smallholder farmers in an agriculture cooperative. To do so, we present the analysis of the results according to the metrics in Section 6.4.

7.2.1. Runtime

We stated that our algorithm takes much less time than the exact algorithm DP (see Table 6). This is due to the fact that we prune significant portions of the search space by only forming the coalitions of similar agents using the lower and upper limits. Our approach does not need to store the coalition values in memory. Instead, we only calculate the *v(C)* values of the coalitions that make the set cover (line 9). In the traditional approaches, the operations of generating and evaluating the coalition structures consume about 99% of the execution time (Cruz et al., 2017). However, our algorithm takes a heuristic approach and makes a set cover with feasible coalitions (line 9–16) to find the coalition structure.

7.2.2. Solution quality

During the experiment, we observe that the DP algorithm is always producing the grand coalition as the optimal solution. This phenomenon happens because, according to our characteristic function, the grand coalition generates the highest amount of expected outcome *Q*, while the grand coalition receives the highest amount of discount. As a result, the gain of the grand coalition is higher than any other smaller sized coalition. Thus, the coalition structures (see Fig. 7) closer to the grand coalition represent good solutions.

Table 7 presents two different coalition structures *CS* and *V(CS)* values, where $n \in \{15, 16, 17, 18, 19, 20, 21, 22, 23\}$. Furthermore, we observe that the *V(CS)* values returned by algorithm 1 for different values of κ and $\kappa\prime$, vary. For the values presented in Table 7 we consider the lower limit of the coalition size as 3 and $\lfloor\sqrt{n}\rfloor$. For the upper limit of the coalition size we consider the either of (5/6/7/8) and $\lfloor\frac{n}{2}\rfloor$.

7.2.3. Scalability

Our algorithm can work for larger agents sets (see Fig. 8). First, we adopt an exact approach to find feasible coalitions of similar agents. As a result, infeasible coalitions are not generated and evaluated. We then adopt a heuristic approach to find the solution. Unlike the other traditional approaches, our algorithm does not need to store $2^n - 1$ number of coalition values in the memory. Thus this allows our algorithm to scale up, unlike other classical CSG approaches. The classical CSG approaches store $2^n - 1$ coalitions and search an exponentially large search space consisting of coalition structures to find the solution. In contrast, we find only one coalition structure that has coalitions of similar agents.

7.2.4. Individual gain

Here we discuss the patterns we observe in the results presented in Fig. 9. We observe that the length of the lines is proportional to the amount of landholding, i.e., the absolute difference between the gains in a coalitional model and a non-coalitional model increases if the amount of landholding increases. However, the lengths of the lines may vary depending on the discount function — the greater the discount, the greater the length of the line. Nevertheless, our experiments show a clear benefit in adopting our proposed coalitional model, as the agents always gain more than in the scenarios where the agents work alone.

7.3. Discussion

This section presents a discussion on the findings that aim to examine the impact of the creation of virtual agricultural cooperatives in the context of smallholder farmers. The main aim is to investigate the model and algorithm qualitative behaviour via experiments.

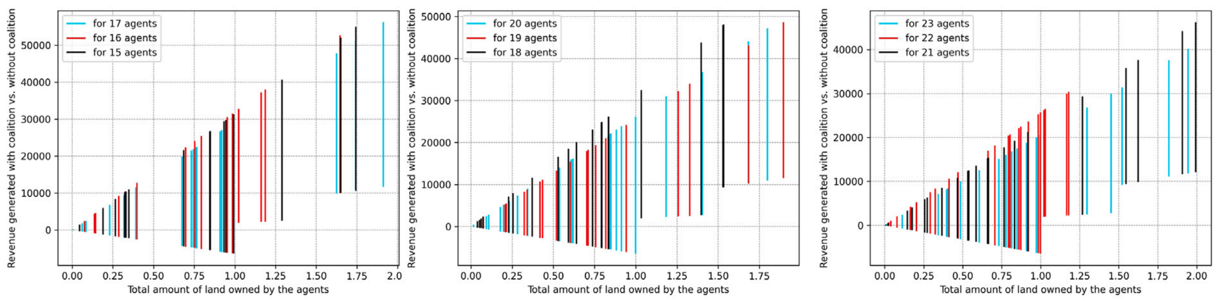
7.3.1. Coalition value vs. coalition size

We further investigate the coalition values in the final coalition structure (line 19) returned by Algorithm 1. Figs. 10(a)–10(b) show the coalition value vs. coalition size for the small-scale agents sets - total number of agents from 15 to 23, and Figs. 11(a)–11(e) show the coalition value vs. coalition size for the large-scale agent sets - total number of agents from 25 to 500. In both small-scale and large-scale agents sets, the upper limit of *v(C)* values generated by our algorithm lies in the range of 250 000–350 000. Furthermore, we observe that the *v(C)* values of the large sized coalitions are less than the smaller sized coalitions. For example, in Fig. 10(a), the largest sized coalition is 11 (for agent set 18). A similar phenomenon we can observe in Fig. 11 as well. This behaviour can be explained by reasoning about the amount of landholdings of the agents. Larger coalitions have agents with less amount of landholdings. For example, for agent set 18, the landholdings of the agents in the largest sized coalition are: [0.52, 0.73, 0.59, 0.63, 0.25, 0.37, 0.039, 0.22, 0.078, 0.05, 0.06] (see Table 8). Thus, in our algorithm, we do not set any strict limit on the coalition size; rather we allow the algorithm to adjust the size of the coalition based on the *v(C)* value.

7.3.2. Similarity vs. Gain

Figs. 12(a)–12(c) show the similarity in receiving revenue of the agents. We use different coloured bars to show the revenue received by the agents of each dataset. We observe that the lengths of the bars are increasing almost monotonically along the x-axis. This behaviour suggests that the agents with more land will receive more revenue in our proposed model. Although the bar lengths for each dataset increase almost linearly, the cases where the bars do not gradually increase are: in agent set 16, agent set 18, agent set 19, agent set 20, agent set 22 and agent set 23 (see Section 7.3.3 for further analysis).

We also observe that, in Figs. 12(a)–12(c) most of the bars lie in the range of 0 ha to 1 ha. This behaviour can be explained by the fact that almost 80% of the data lie in the first two ranges, i.e., 0 to 0.5 and 0.5 to 1.0 (see Table 5). In Table 8 we show the coalition-wise amount of landholding of the agents. Here, first, we observe that the agents with similar landholding are always in the same coalition. Therefore, we can claim that our algorithm is successfully generating coalitions of similar agents and the agents with a similar amount of landholding are earning similar revenue in our proposed model. Second, we observe that in Fig. 12(a) for the result of agent set 16, the agents near 1 ha (on x-axis), i.e., the agents of coalition $\{f_5, f_{11}, f_0, f_2, f_{13}\}$ have an irregular amount of revenue. In other words, an agent having landholding more



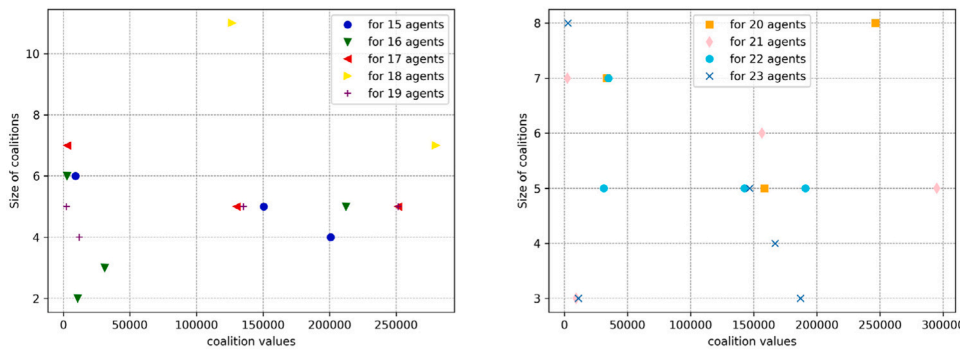
(a) Total number of agents: 15, 16 and 17 (b) Total number of agents: 18, 19 and 20 (c) Total number of agents: 21, 22 and 23

Fig. 9. Comparative analysis of revenue received by the agents while working alone vs. working in coalitions — for a total amount of agents from 15 to 23.

Table 7
Coalition structures *CS*, coalition values of $C \in CS$ and *V(CS)* for agents from 15 to 23. We have varied the κ and $\kappa\prime$ values and obtained different results. In the first row for each no. of agent, the lower limit is $\kappa = 3$ and the upper limit is $\kappa\prime = 5/6/7/8$. In the next row for each no. of agent, the lower limit is $\kappa = \lfloor\sqrt{n}\rfloor$ and the upper limit is $\kappa\prime = \lfloor\frac{n}{2}\rfloor$.

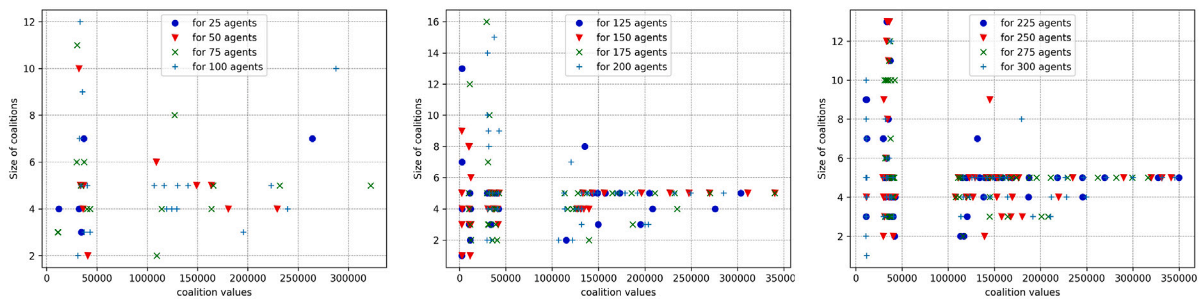
No. of agents	Coalition structure *CS*	Coalition value *v(C)* where ($C \in CS$)	Coalition structure value *V(CS)*	κ value	$\kappa\prime$ value
15	{{14, 1, 3, 6, 13, 5}, {9, 10, 4, 2}, {12, 0, 8, 11, 7}}	[9140.83, 200 789.26, 150 573.54]	360 503.64	3	5
	{{0, 7, 8, 11, 12}, {1, 3, 5, 6, 13, 14}, {9, 10, 4, 2}}	[150 573.54, 9140.83, 200 789.26]	360 503.64	$\lfloor\sqrt{15}\rfloor$	$\lfloor\frac{15}{2}\rfloor$
16	{{1, 3}, {4, 6, 7, 10, 8, 14}, {9, 12, 15}, {5, 11, 0, 2, 13}}	[10 637.74, 2606.64, 31 034.81, 212 295.67]	256 574.88	3	5
	{{9, 15}, {0, 2, 5, 11}, {10, 4, 6, 7, 8, 14}, {12, 1, 3, 13}}	[2656.78, 177 466.59, 2606.64, 123 242.85]	305 972.88	$\lfloor\sqrt{16}\rfloor$	$\lfloor\frac{16}{2}\rfloor$
17	{{0, 1, 13, 15, 8, 6, 11}, {2, 4, 7, 14, 16}, {3, 5, 9, 10, 12}}	[2835.78, 130 096.77, 251 440.14]	384 372.70	3	5
	{{0, 1, 6, 8, 11, 13, 15}, {4, 7, 14, 16, 2, 10, 12}, {3, 5, 9}}	[2835.78, 194 756.76, 186 780.15]	384 372.70	$\lfloor\sqrt{17}\rfloor$	$\lfloor\frac{17}{2}\rfloor$
18	{{3, 5, 11, 13, 2, 7, 0}, {4, 14, 15, 17, 1, 12, 9, 6, 8, 10, 16}}	[279 911.99, 126 622.10]	406 534.09	3	6
	{{0, 4, 14, 15, 17}, {1, 6, 8, 9, 10, 16, 12}, {3, 5, 7, 11, 13, 2}}	[116 316.12, 2061.88, 251 932.02]	370 310.02	4	7
19	{{3, 8, 10, 14, 18}, {5, 6, 11, 13, 15}, {1, 4, 7, 12, 16}, {0, 9, 2, 17}}	[2204.99, 135 241.93, 251 411.41, 12 025.80]	400 884.14	3	5
	{{5, 6, 4, 7, 12, 13, 15, 16}, {11, 0, 2, 9, 17}, {1}, {14, 3, 8, 10, 18}}	[294 717.38, 38 704.85, 11 548.01, 2204.99]	347 175.24	$\lfloor\sqrt{19}\rfloor$	$\lfloor\frac{19}{2}\rfloor$
20	{{1, 3, 8, 13, 19}, {17, 11, 15, 12, 2, 6, 10, 14}, {18, 7, 9, 4, 0, 16, 5}}	[158 463.65, 246 224.74, 33 578.54]	438 266.94	3	6
	{{4, 7, 9, 18, 3, 8, 10, 13, 19, 0, 2, 5, 6, 14, 16}, {1, 11, 12, 15, 17}}	[236 797.52, 250 058.45]	486 855.98	$\lfloor\sqrt{20}\rfloor$	$\lfloor\frac{20}{2}\rfloor$
21	{{0, 1, 5, 7, 12, 15, 17}, {9, 16, 20}, {4, 6, 8, 3, 11}, {10, 13, 2, 14, 18, 19}}	[2370.12, 32 702.34, 40 225.34, 327 499.50]	402 797.31	3	8
	{{3, 9, 16, 20}, {0, 4, 5, 6, 8, 11, 12, 15, 1, 7, 17}, {13, 2, 10, 14, 18, 19}}	[42 229.63, 119 128.46, 327 499.50]	488 857.60	$\lfloor\sqrt{21}\rfloor$	$\lfloor\frac{21}{2}\rfloor$
22	{{1, 10, 11, 14, 15}, {4, 12, 13, 17, 18}, {2, 16, 19, 9, 20}, {0, 7, 3, 5, 6, 8, 21}}	[190 796.42, 31 227.30, 142 729.81, 34 948.80]	399 702.35	3	5
	{{9, 16, 20}, {5, 6, 8, 3, 4, 12, 13, 17, 18, 21}, {19, 2}, {15, 0, 1, 7, 10, 11, 14}}	[36 656.60, 38 659.19, 2855.92, 258 131.06]	336 302.78	$\lfloor\sqrt{22}\rfloor$	$\lfloor\frac{22}{2}\rfloor$
23	{{0, 3, 6, 9, 21}, {11, 15, 16, 22}, {8, 17, 20, 1, 2, 10, 13, 18}, {12, 7, 19}, {14, 4, 5}}	[146 830.66, 166 836.09, 2800.92, 187 001.02, 11 104.32]	514 573.03	3	5
	{{7, 12, 15, 16, 19, 3, 9, 11, 21, 22}, {1, 2, 4, 5, 8, 10, 13, 17, 18, 20}, {0, 6, 14}}	[444 756.44, 37 052.70, 33 414.10]	515 223.26	$\lfloor\sqrt{23}\rfloor$	$\lfloor\frac{23}{2}\rfloor$

than 1 ha is receiving lesser revenue than an agent who has landholding less than 1 ha. Table 8 shows that the amount of landholdings of the agents in that coalition: [1.18, 1.64, 1.16, 1.02, 0.98]. In this example, agent f_2 having landholding 1.02 ha is getting less revenue than the agent having landholding 0.98 ha (f_{13}). This phenomenon again supports the fact that putting an agent having larger landholding with the agents having lesser landholding will incur less revenue for the former. Maybe putting f_{13} in the coalition $\{f_1, f_3\}$ could have avoided the phenomenon, as f_1 and f_3 are having 0.79 ha and 0.95 ha, respectively. However, the payoff distribution rule ensures that the revenue is always proportional to the amount of landholding (see Section 7.3.3). For agent set 16 we get 2 coalition structures, $CS_{16} = \{\{f_1, f_3\}, \{f_4, f_6, f_7, f_{10}, f_8, f_{14}\}, \{f_9, f_{12}, f_{15}\}, \{f_5, f_{11}, f_0, f_2, f_{13}\}\}$ with *V(CS)* = 256 574.88 and $CS_{16}{}' = \{\{f_9, f_{15}\}, \{f_0, f_2, f_5, f_{11}\}, \{f_{10}, f_4, f_6, f_7, f_8, f_{14}\}, \{f_{12}, f_1, f_3, f_{13}\}\}$ with greater *V(CS)* =

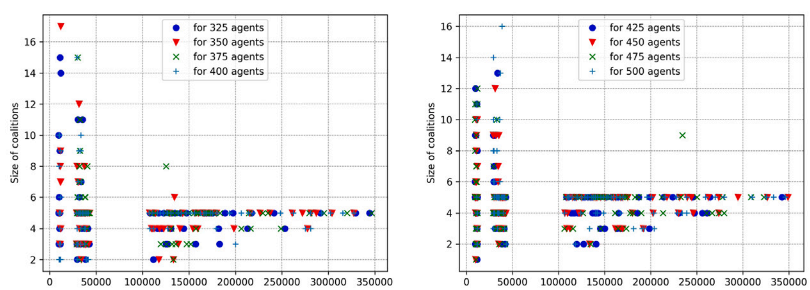


(a) Total number of agents: 15, 16, 17, 18 and 19 (b) Total number of agents: 20, 21, 22 and 23

Fig. 10. Distribution of coalition values of agents from 15 to 23.



(a) Total number of agents: 25, 50, 75 and 100 (b) Total number of agents: 125, 150, 175 and 200 (c) Total number of agents: 225, 250, 275 and 300



(d) Total number of agents: 325, 350, 375 and 400 (e) Total number of agents: 425, 450, 475 and 500

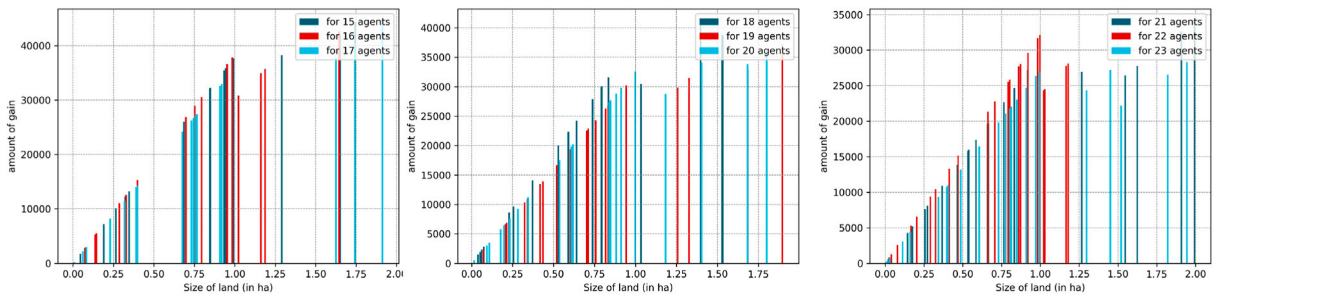
Fig. 11. Distribution of coalition values.

305 972.88 (see Table 7). The second coalition structure has the following amount of landholdings: [[0.69, 0.69], [1.16, 1.02, 1.18, 1.64], [0.07, 0.28, 0.32, 0.39, 0.13, 0.14], [0.75, 0.79, 0.95, 0.98]]. Comparing the two *CS*, we can see that the coalition $\{f_4, f_6, f_7, f_{10}, f_8, f_{14}\}$ is common in both the *CS*s. However, we observe that compared to CS_16, in CS_{16}' the coalition $\{f_0, f_2, f_5, f_{11}\}$ has excluded agent f_{13} and as we discussed earlier in this section, the *V(CS)* may have increased due to the exclusion of f_{13} from that coalition.

7.3.3. Total landholding vs. profit

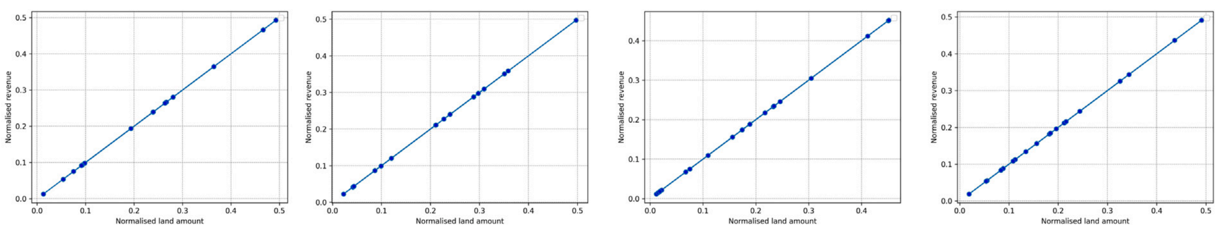
Figs. 13(a)–13(g) present the amount of revenue received by the agents according to the size of their landholding. These figures are derived from Figs. 12(a), 12(b), 12(c). In this section, we do not investigate all the cases of revenue received by each agent of every dataset separately; instead, we investigate only a few cases where the bars in Figs. 12(a), 12(b), 12(c) do not increase evenly (as we mentioned in the previous section).

In this work, we consider three ranges for the landholdings (see Table 5). Figs. 13(a)–13(g) present the normalised values of landholding along the x-axis and the normalised value of the revenue received along the y-axis. We normalise the values to check if the revenue received by an agent is proportional to their landholding, as discussed in Section 5.4, and both the values are always equal. Therefore, we can claim that the equation (see Section 5.4) to divide the payoff among the agents is capable of dividing the payoff according to an agent's contribution toward a coalition.

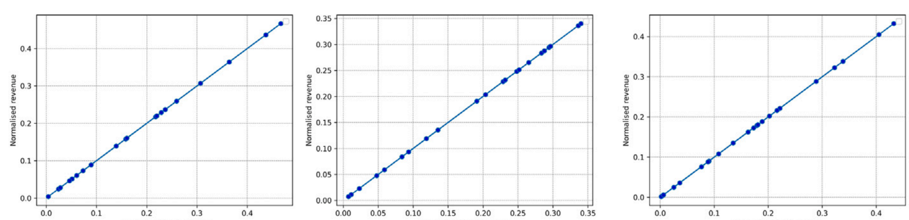


(a) Total number of agents: 15, 16 and 17 (b) Total number of agents: 18, 19 and 20 (c) Total number of agents: 18, 19 and 20

Fig. 12. Similarity in the revenue according to the land size.



(a) For 15 agents (b) For 16 agents (c) For 18 agents (d) For 19 agents

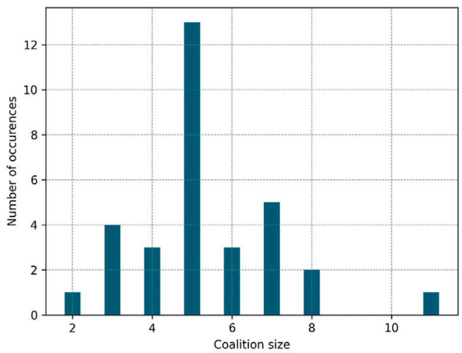


(e) For 20 agents (f) For 22 agents (g) For 23 agents

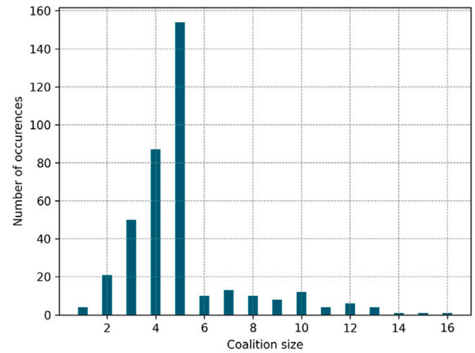
Fig. 13. Normalised value of the land vs. revenue received of agents.

Table 8
Coalition structure and the amount of landholding of the agents.

No. of agents	Coalition structure	Amount of landholding (in ha)
15.	{{14, 1, 3, 6, 13, 5}, {9, 10, 4, 2}, {12, 0, 8, 11, 7}}	[[0.045, 0.18, 0.26, 0.32, 0.32, 0.34], [1.64, 1.74, 1.29, 0.99], [0.68, 0.84, 0.93, 0.84, 0.94]]
16.	{{1, 3}, {4, 6, 7, 10, 8, 14}, {9, 12, 15}, {5, 11, 0, 2, 13}}	[[0.79, 0.95], [0.28, 0.32, 0.39, 0.07, 0.13, 0.14], [0.69, 0.75, 0.69], [1.18, 1.64, 1.16, 1.02, 0.98]]
17.	{{0, 1, 13, 15, 8, 6, 11}, {2, 4, 7, 14, 16}, {3, 5, 9, 10, 12}}	[[0.31, 0.39, 0.22, 0.39, 0.006, 0.06, 0.08], [0.76, 0.67, 0.74, 0.75, 0.73], [1.62, 1.74, 1.91, 0.91, 0.90]]
18.	{{3, 5, 11, 13, 2, 7, 0}, {4, 14, 15, 17, 1, 12, 9, 6, 8, 10, 16}}	[[1.52, 1.39, 1.03, 1.53, 0.79, 0.83, 0.79], [0.52, 0.73, 0.59, 0.63, 0.25, 0.37, 0.039, 0.22, 0.078, 0.05, 0.06]]
19.	{{3, 8, 10, 14, 18}, {5, 6, 11, 13, 15}, {1, 4, 7, 12, 16}, {0, 9, 2, 17}}	[[0.34, 0.21, 0.20, 0.07, 0.32], [0.75, 0.81, 0.70, 0.71, 0.83], [1.89, 1.68, 1.25, 0.94, 1.32], [0.43, 0.41, 0.60, 0.51]]
20.	{{1, 3, 8, 13, 19}, {17, 11, 15, 12, 2, 6, 10, 14}, {18, 7, 9, 4, 0, 16, 5}}	[[0.99, 0.83, 0.88, 0.84, 0.91], [1.79, 1.68, 1.40, 1.18, 0.19, 0.17, 0.28, 0.23], [0.61, 0.60, 0.53, 0.34, 0.09, 0.10, 0.01]]
21.	{{0, 1, 5, 7, 12, 15, 17}, {9, 16, 20}, {4, 6, 8, 3, 11}, {10, 13, 2, 14, 18, 19}}	[[0.36, 0.14, 0.27, 0.25, 0.021, 0.006, 0.17], [0.76, 0.66, 0.83], [0.53, 0.53, 0.58, 0.65, 0.46], [1.90, 1.99, 1.26, 0.91, 1.54, 1.62]]
22.	{{1, 10, 11, 14, 15}, {4, 12, 13, 17, 18}, {2, 16, 19, 9, 20}, {0, 7, 3, 5, 6, 8, 21}}	[[1.02, 1.16, 1.02, 0.99, 1.17], [0.41, 0.46, 0.326, 0.66, 0.29], [0.79, 0.80, 0.70, 0.86, 0.87], [0.92, 0.98, 0.20, 0.03, 0.02, 0.07, 0.16]]
23.	{{0, 3, 6, 9, 21}, {11, 15, 16, 22}, {8, 17, 20, 1, 2, 10, 13, 18}, {12, 7, 19}, {14, 4, 5}}	[[0.77, 0.90, 0.80, 0.84, 0.81], [0.99, 1.45, 1.29, 0.97], [0.34, 0.40, 0.39, 0.01, 0.007, 0.16, 0.02, 0.11], [1.82, 1.94, 1.52], [0.73, 0.60, 0.48]]



(a) Distribution of coalition values of agents from 15 to 23



(b) Distribution of coalition values of agents from 25 to 100

Fig. 14. Visualisation of *V(C)* value.

7.3.4. Price discount vs. coalition size

The size of the coalitions varies depending on the discount function. We initially take the lower limit and upper limit in our proposed algorithm 1 (line 4). However, coalitions of the given size sometimes generate negative valued coalitions. We remove the agents from the negative valued coalitions to the least valued ones (lines 15–16) to generate a coalition structure without any negative valued coalition. For the amount of discount lesser than that of Fig. 6 will require larger coalitions to form in order to avoid any negative valued coalition. Our experimental setup shows the most commonly formed coalition sizes in Figs. 14(a) and 14(b). For the smaller agent sets, i.e., agents from 15 to 23, the most commonly formed coalition size is 5. In some cases, our algorithm failed to return any valid solution with some of the lower and upper bounds (especially the upper bounds). However, for the larger agent set, the algorithm produced valid results with lower and upper limits 3 and 5, respectively. Here, we observe that some large size coalitions are generated but are few in number. When we investigated the larger datasets, we found that for the agent set 175 a coalition of 16 agents has been formed. Our intuition is, such a coalition is formed to avoid generating a coalition with negative value as we can see from Fig. 11 the coalition with 16 agents has less *v(C)* value. Same for the coalition with 12 and 10 agents. Moreover, some coalitions with greater *v(C)* value has less number of agents, e.g., a 2-sized coalition having closer to 150 000 or a 4-sized coalition having *v(C)* value closer to 250 000. We further investigated the amount of landholdings of the same agent set and found that each agent in that coalition has landholding less than 0.22 ha. With this background, we can say that, although the final coalition structure has larger sized coalitions than the given limits (line 4), finally, our algorithm is generating coalitions with agents having similar profiles (line 15–16).

7.3.5. Performance of the algorithm

On the basis of the experiments carried out, the proposed algorithm can be considered an efficient approach to address the coalition formation problem in a scenario where similar agents are purchasing goods collectively and making a profit. In this section, we present a discussion based on the overall research findings.

Based on the working principle of an algorithm, the proposed algorithm is: (i) centralised — where the final decision is taken collectively and (ii) static — where the coalitions are formed prior to the commencement of the agricultural cooperative. Our algorithm starts with all agents and does not allow dynamism, i.e., agents are not allowed to join or leave in an instant. Furthermore, our algorithm is suitable for complex, and large-scale systems as the algorithm is capable of generating a good enough approximate solution within the permissible time (Sarkar et al., 2022).

We adopt a similarity matching approach to generate the coalitions (see Section 5.2.1). This method is advantageous in several ways: (i) the algorithm takes a heuristic approach to find the coalition structure. However, the process of coalition generation takes an exact approach to finding the coalitions of similar agents. Most probably, this approach helps the algorithm to find a good quality solution, (ii) generating coalitions among the similar agents works as constraints that specify which groups of agents should/should not work together. In other words, the heterogeneity of the agents is restraining them from working with everyone. This approach makes our algorithm desirable as it focuses on a few feasible coalitions and efficiently generates a valid coalition structure (see Table 6), (iii) we assume that the similar agents should always be in the same coalition and earn the in the same proportion. The similarity matching approach enables our algorithm to achieve this objective and find a good quality solution without storing any coalition value in the memory, (iv) we use the set cover approach to find the final coalition structure (see line 13). Despite being an *NP*-complete problem, our algorithm returns a solution in a reasonable amount of time compared to the state-of-the-art exact approaches (it takes 95 s for 500 agents). This phenomenon can be explained by the number of coalitions our algorithm explores to find a coalition structure. Adopting the similarity matching criteria allows the algorithm to generate only the feasible coalitions.

Using the set cover approach to find the coalition structure is beneficial in another way as well, as it makes our algorithm non-sensitive to the coalition values. The runtime of our algorithm depends on the number of coalitions that our model generates, and the number of coalitions generated is polynomial and not exponential, unlike the state-of-the-art approaches where the number of coalitions is $2^n - 1$ (see Proposition 1).

7.4. Summary

Based on the experiments carried out in this section, our proposed coalition model is an efficient approach to address the coalition formation of smallholder farmers based on similarity among the farmers. We model an agricultural cooperative as a multi-agent-based coalitional model. The run-time of our similarity-based heuristic approach is only a few seconds for hundreds of agents (95 s for 500 agents), while the worst-case solution quality is 64% (for 16 agents). Moreover, agents always gain more when working in a coalition model than when working alone; this gain is proportional to an agent's contribution toward a coalition. So, we can say that our proposed model successfully divides farmers into small groups to improve their social welfare.

8. Limitations of our work and future work

Based on the evaluation and discussion in the previous section, our approach is an efficient way to form coalitions in an agricultural cooperative or where group buying is involved. However, despite promising results, it is essential to highlight that our approach has some drawbacks:

- Unlike traditional approaches like DP, our approach may produce multiple near-optimal solutions as shown in Table 7 by varying the lower and upper limit of constraint 3 (although the solutions are within a certain bound from the optimal solution). In future work, we need to identify κ and $\kappa\prime$ values to get the best outcome.
- Our approach provides an approximation of the solution. Nevertheless, it does not provide any guarantee of the stability of the solution.
- Forming coalitions provides the stakeholders to have the power to bargain. However, our approach does not generate coalitions with the same *v(C)* values in this work. On the other hand, as we discussed in Section 7.3.4, larger-sized coalitions fail to draw a greater *v(C)* value. Moreover, forming larger coalitions may fetch communication overheads to the system. Hence, we need to investigate this issue further and find the best way of forming coalitions in the context of coalition formation in an agricultural cooperative.
- Although no singleton coalition is getting formed for the smaller size agent set, where $n \in \{15, 16, 17, 18, 19, 20, 21, 22, 23\}$ (see Fig. 10), one singleton coalitions was formed for agent set 125, two for agent set 150 and one for agent 300 (see Fig. 11). We need to find a general rule to integrate these agents into the proper coalition so the system remains stable.
- For this work, we do not use any actual discount function since, to our knowledge, no such discount function is presently available. Instead, the information related to a discount function in the context of an agricultural cooperative can be retrieved by surveying the physical marketplace.
- Our work focuses on determining the best combination (the coalition) of similar smallholder farmers to minimise the production costs within a local agricultural cooperative. Therefore, farmers within the cooperative will be chosen to achieve the best coalition. The selling price is an interesting perspective for future research. However, this work does not consider different selling prices in different market segments and other variables.
- Due to the unavailability of the parameters ω and Z of Section 4.1, we do not use these parameters in this work.

In future work, we want to evaluate our model where coalitions will be formed based on the constraints described in 4.2.1. Adopting this method will satisfy Constraint 2, i.e., forming coalitions among agents with adjacent land will increase the total amount of operational land. However, it requires the land's geo-spatial data and the extra amount of land. Future work will evaluate our model with a real discount function and with individual entries with all the parameters of Section 4.1. The selling price in different market segments may be added in future work, where each market segment will generate a different output price that determines revenue. We will also develop a mobile application or website for further dissemination and validation of our proposed coalitional model of an agricultural cooperative. The steps described in Section 4.2.1 will be implemented to automate the agricultural cooperative of smallholder farmers. We will then conduct a case study among smallholder farmers who cultivate alone and compare the results with our proposed coalitional model. Furthermore, we will explore better heuristic approaches and machine learning-based techniques to form coalitions of smallholder farmers.

9. Conclusions

Smallholder farmers have great difficulty entering the market because their production is small, and they cannot face the price pressure imposed by large producers. Moreover, they face great difficulties in lowering the production price and increasing their productivity, among other difficulties. One solution is to get together in cooperation. This cooperation materialises in agricultural cooperatives.

Smallholder farmers in an agricultural cooperative can build groups to have more product market share, among other advantages. Our work presents a coalitional game-based modelling of an agricultural cooperative of smallholder farmers to increase their revenue through joint cultivation. We adopt a coalition formation approach, as it provides an efficient way to coordinate a group of agents who aim to work jointly to optimise their social welfare as a result of their cooperation.

An agricultural cooperative usually consists of hundreds of heterogeneous farmers, and organising these farmers in a structured way is crucial in understanding, designing, implementing, and operating the cooperative. Our model considers an individual farmer as an agent, and we present agent-based modelling of an agricultural cooperative of those agents. We use social constraints to transform this agent-based model of an agricultural cooperative into a coalitional formation model. Finally, we present the three steps of coalition formation in the context of forming virtual cooperatives of smallholder farmers:

- coalition value calculation — where we design a characteristic function that resembles $v(C) = v(C^{+}) - v(C^{-})$, where *v(C)* increases as $v(C^{-})$ decreases;
- coalition structure generation — where we implement a heuristic-based algorithm to find the disjoint partition of the agents set;
- payoff division — where we divide the payoff among the agents according to their contributions to the cooperative.

We evaluated the algorithm and the model with four metrics: runtime comparison, solution quality, scalability and individual gain. In order to perform the evaluation, we created a new dataset that contains the landholding values and the resource requirement values of the agents. With the later help, the algorithm calculates the similarity among the agents and finally finds a disjoint coalition structure. Although the general problem of coalition structure generation is a *NP*-hard problem, our algorithm provides a solution within an acceptable time for hundreds of agents.

We proved that:

1. the algorithm is time efficient since for 500 agents, it takes 95 s;
2. the algorithm is effective in providing good solutions that are within a bound of the optimal solution; in the worst case for 16 agents, the solution is 64% of the optimal;
3. the algorithm has enough scalability to handle a large number of agents (in this work, we show the results for up to 500 agents);
4. the coalitional model of the agricultural cooperative always draws positive revenue for the agents based on an agent's contribution to the cooperative.

Additionally, we study:

1. how the coalition sizes regulate the coalition values,
2. how the similarity among the agents of a coalition affects the revenue of these agents,
3. how the amount of landholding regulates the profit,
4. how the price discount regulates the coalition size, and
5. the rationale behind the performance of the algorithm.

Our work shows how the coalition formation technique can be applied to real-world scenarios by relaxing some restrictive assumptions while modifying some.

This work contributes to providing tools to enhance the world's social economy/informal economy — particularly in developing countries. It is a fact that smallholder farmers need to come together to survive — the question was: how can we be efficient in bringing them together? Our contribution will address this need. Empowering smallholder farmers is a pressing need. In fact, by empowering these farmers, we are contributing to the following UN's sustainable development goals: #1 — No Poverty, #2 — Zero Hunger; #8 — Decent Work and

Economic Growth; #10 — Reduced Inequalities; #16 — Peace and Justice.

Future work will focus on the graph-based modelling of an agricultural cooperative of smallholder farmers and on finding approaches that can find better solutions in lesser time than the present one. Here, we will use parallelism by having different threads to search the solutions simultaneously. In order to use graph-based modelling, we also need to focus on collecting geospatial data on the land. In addition, we will focus on finding approaches to guarantee the stability of the solution. Future work will further investigate the model with the values of actual parameters that we could not be used in the present work.

CRediT authorship contribution statement

Samriddhi Sarkar: Investigation, Conceptualization, Methodology, Data curation, Software, Validation, Formal analysis, Writing – original draft. **Tuhin Biswas:** Software, Validation. **Mariana Curado Malta:** Supervision, Investigation, Conceptualization, Methodology, Writing - Original Draft, Writing – review & editing. **Deolinda Meira:** Resources, Writing. **Animesh Dutta:** Supervision, Investigation, Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data is publicly available and the link is provided in the manuscript.

Acknowledgements

This work has been supported by Portuguese national funds through FCT - within the Project Scope UID/CEC/00319/2019.

This work has been supported by Portuguese national funds through FCT - within the Project Scope UIDB/05422/2020.

References

Abdul-Rahaman, A., & Abdulai, A. (2020a). Farmer groups, collective marketing and smallholder farm performance in rural Ghana. *Journal of Agribusiness in Developing and Emerging Economies*, *10*(5), 511–527.

Abdul-Rahaman, A., & Abdulai, A. (2020b). Social networks, rice value chain participation and market performance of smallholder farmers in Ghana. *African Development Review*, *32*(2), 216–227.

Agarwal, M., Agrawal, N., Sharma, S., Vig, L., & Kumar, N. (2015). Parallel multi-objective multi-robot coalition formation. *Expert Systems with Applications*, *42*(21), 7797–7811.

Agarwal, M., Kumar, N., & Vig, L. (2014). Non-additive multi-objective robot coalition formation. *Expert Systems with Applications*, *41*(8), 3736–3747.

Alam, M., Ramchurn, S., & Rogers, A. (2013). Cooperative energy exchange for the efficient use of energy and resources in remote communities. In *12th international conference on autonomous agents and multi-agent systems, (AAMAS)* (pp. 731–738). IFAAMAS.

Álvarez, J., Bouchard, M., & Marcuello, C. (2022). Social economy and covid-19: An international approach. *Revista de Economía Pública, Social y Cooperativa*, (104), 205–207.

Bialoskorski Neto, S. (2001). Virtual cooperatives in Brazil and the globalization process. *Journal of Rural Cooperation*, *29*(886-2016-64598), 153–165.

Bistaffa, F., Farinelli, A., Cerquides, J., Rodriguez-Aguilar, J. A., & Ramchurn, S. D. (2014). Anytime coalition structure generation on synergy graphs. In *13th international conference on autonomous agents and multi-agent systems (AAMAS)*. IFAAMAS/ACM.

Bistaffa, F., Farinelli, A., Cerquides, J., Rodríguez-Aguilar, J., & Ramchurn, S. D. (2017). Algorithms for graph-constrained coalition formation in the real world. *ACM Transactions on Intelligent Systems and Technology*, *8*(4), 1–24.

Bonroy, O., Garapin, A., Hamilton, S. F., & Souza Monteiro, D. M. (2019). Free-riding on product quality in cooperatives: Lessons from an experiment. *American Journal of Agricultural Economics*, *101*(1), 89–108.

Borodin, V., Bourtembourg, J., Hnaien, F., & Labadie, N. (2016). Handling uncertainty in agricultural supply chain management: A state of the art. *European Journal of Operational Research*, *254*(2), 348–359.

Bulka, B., Gaston, M., & desJardins, M. (2007). Local strategy learning in networked multi-agent team formation. *Autonomous Agents and Multi-Agent Systems*, *15*(1), 29–45.

Byerlee, D., De Janvry, A., Sadoulet, E., Townsend, R., & Klytchnikova, I. (2008). *World development report 2008: agriculture for development*: *Technical Report*, Washington, DC: The World Bank.

Candemir, A., Duvaleix, S., & Latruffe, L. (2021). Agricultural cooperatives and farm sustainability–A literature review. *Journal of Economic Surveys*, *35*(4), 1118–1144.

Chalkiadakis, G., Robu, V., Kota, R., Rogers, A., & Jennings, N. (2011). Cooperatives of distributed energy resources for efficient virtual power plants. In *10th international conference on autonomous agents and multiagent systems (AAMAS)* (pp. 787–794).

Chen, D.-N., Jeng, B., Lee, W.-P., & Chuang, C.-H. (2008). An agent-based model for consumer-to-business electronic commerce. *Expert Systems with Applications*, *34*(1), 469–481.

Ciruela-Lorenzo, A. M., Del-Aguila-Obra, A. R., Padilla-Meléndez, A., & Plaza-Angulo, J. J. (2020). Digitalization of agri-cooperatives in the smart agriculture context. proposal of a digital diagnosis tool. *Sustainability*, *12*(4), 1325.

Conitzer, V., & Sandholm, T. (2004). Computing Shapley values, manipulating value division schemes, and checking core membership in multi-issue domains. In *American association for artificial intelligence, (AAAI), Vol. 4* (pp. 219–225).

Contreras, J., & Wu, F. F. (1999). Coalition formation in transmission expansion planning. *IEEE Transactions on Power Systems*, *14*(3), 1144–1152.

Cruz, F., Espinosa, A., Moure, J. C., Cerquides, J., Rodriguez-Aguilar, J. A., Svensson, K., & Ramchurn, S. D. (2017). Coalition structure generation problems: optimization and parallelization of the IDP algorithm in multicore systems. *Concurrency Computations: Practice and Experience*, *29*(5), Article e3969.

Dang, V. D., Dash, R. K., Rogers, A., & Jennings, N. R. (2006). Overlapping coalition formation for efficient data fusion in multi-sensor networks. In *American association for artificial intelligence (AAAI), Vol. 6* (pp. 635–640).

Fajardo, G., Fici, A., Henrÿ, H., Hiez, D., Meira, D., Münkner, H.-H., & Snaith, I. (2017). *Principles of European cooperative law*.

Farinelli, A., Bicego, M., Ramchurn, S., & Zuchelli, M. (2013). C-link: A hierarchical clustering approach to large-scale near-optimal coalition formation. In *3rd international joint conference on artificial intelligence*.

Farmer, J. D., & Foley, D. (2009). The economy needs agent-based modelling. *Nature*, *460*(7256), 685–686.

Fischer, K., Muller, J. P., Heimig, I., & Scheer, A.-W. (1996). Intelligent agents in virtual enterprises. In *1st international conference on the practical application of intelligent agents and multi-agent technology (PAAM)* (pp. 205–223).

Gaston, M. E., & DesJardins, M. (2008). The effect of network structure on dynamic team formation in multi-agent systems. *Computational Intelligence*, *24*(2), 122–157.

He, L., & Ioerger, T. R. (2004). Combining bundle search with buyer coalition formation in electronic markets: A distributed approach through explicit negotiation. In *6th international conference on electronic commerce* (pp. 95–104). ACM.

Hueth, B., & Marcoul, P. (2015). Agents monitoring their manager: A hard-times theory of producer cooperation. *Journal of Economics & Management Strategy*, *24*(1), 92–109.

Hussin, A., & Fatima, S. (2016). Heuristic methods for optimal coalition structure generation. In *Multi-agent systems and agreement technologies* (pp. 124–139). Springer.

Jennings, N., & Dang, V. (2004). Coalition formation algorithms for virtual organizations. *Intellingence, Agents, Multimedia Group, Electronics and Computer Sceince–University of Southampton*.

Lazzarini, S., Chaddad, F., & Cook, M. (2001). Integrating supply chain and network analyses: the study of netchains. *Journal on Chain and Network Science*, *1*(1), 7–22.

Li, C., Rajan, U., Chawla, S., & Sycara, K. (2003). Mechanisms for coalition formation and cost sharing in an electronic marketplace. In *5th international conference on electronic commerce* (pp. 68–77).

Li, C., & Sycara, K. (2002). Algorithm for combinatorial coalition formation and payoff division in an electronic marketplace. In *1st international joint conference on autonomous agents & multiagent systems (AAMAS)* (pp. 120–127). ACM.

Li, C., Sycara, K., & Scheller-Wolf, A. (2010). Combinatorial coalition formation for multi-item group-buying with heterogeneous customers. *Decision Support Systems*, *49*(1), 1–13.

Ma, W., & Abdulai, A. (2017). The economic impacts of agricultural cooperatives on smallholder farmers in rural China. *Agribusiness*, *33*(4), 537–551.

Mahendra Dev, S. (2014). Small farmers in India: Challenges and opportunities.

Meira, D. (2022). Cooperative governance and sustainability: An analysis according to new trends in European cooperative law. In *Perspectives on cooperative law: Festschrift in honour of professor Hagen Henrÿ* (p. 223). Springer Nature.

Myerson, R. B. (1977). Graphs and cooperation in games. *Mathematics of Operations Research*, *2*(3), 225–229.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM Review*, *45*(2), 167–256.

International labour office, G. (1984). Basic economics of an agricultural co-operative: a learning element for staff of agricultural cooperatives.

de Oliveira Ramos, G., Burguillo, J. C., & Bazzan, A. L. C. (2015). A self-adapting similarity-based coalition formation approach for plug-in electric vehicles in smart grids. *Multiagent and Grid Systems*, *11*(3), 167–187.

Petersen, S. A., Divitini, M., & Matskin, M. (2001). An agent-based approach to modelling virtual enterprises. *Production Planning and Control*, *12*(3), 224–233.

Rahwan, T., & Jennings, N. R. (2008). An improved dynamic programming algorithm for coalition structure generation. In *7th international joint conference on autonomous agents and multiagent systems (AAMAS), Vol. 3* (pp. 1417–1420).

Rahwan, T., Michalak, T. P., Wooldridge, M., & Jennings, N. R. (2015). Coalition structure generation: A survey. *Artificial Intelligence*, *229*, 139–174.

Rani, R., & HRD, A. D. Training programme on formation and management of producers groups and federations. National Institute of Agricultural Extension Management, Hyderabad.

Rapsomanikis, G. (2015). *The economic lives of smallholder farmers: An analysis based on household data from nine countries*. Rome: Food and Agriculture Organization of the United Nations.

Sandholm, T., Larson, K., Andersson, M., Shehory, O., & Tohmé, F. (1999). Coalition structure generation with worst case guarantees. *Artificial Intelligence*, *111*(1–2), 209–238.

Sarkar, S., Curado Malta, M., & Dutta, A. (2022). A survey on applications of coalition formation in multi-agent systems. *Concurrency Computations: Practice and Experience*, Article e6876.

Sarkar, S., Malta, M. C., & Dutta, A. (2020). A coalition formation framework for platform cooperatives of smallholder farmers. In *IEEE/WIC/ACM international joint conference on web intelligence and intelligent agent technology (WI-IAT)* (pp. 803–809). IEEE.

Service, T. C., & Adams, J. A. (2010). Anytime dynamic programming for coalition structure generation. In *9th international conference on autonomous agents and multiagent systems (AAMAS), Vol. 1* (pp. 1411–1412).

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, *2*(28), 307–317.

Shehory, O., & Kraus, S. (1998). Methods for task allocation via agent coalition formation. *Artificial Intelligence*, *101*(1–2), 165–200.

Staatz, J. M. (1983). The cooperative as a coalition: a game-theoretic approach. *American Journal of Agricultural Economics*, *65*(5), 1084–1089.

Tortia, E. C., Valentinov, V. L., & Iliopoulos, C. (2013). Agricultural cooperatives. *Journal of Entrepreneurial and Organizational Diversity*, *2*(1), 23–36.

Tošić, P. T., & Agha, G. A. (2004). Maximal clique based distributed coalition formation for task allocation in large-scale multi-agent systems. In *International workshop on massively multiagent systems* (pp. 104–120). Springer.

Tran-Thanh, L., Nguyen, T.-D., Rahwan, T., Rogers, A., & Jennings, N. R. (2013). An efficient vector-based representation for coalitional games.

Vasirani, M., Ossowski, S., Kota, R., Cavalcante, R. L. G., & Jennings, N. R. (2011). Using coalitions of wind generators and electric vehicles for effective energy market participation. In *10th international conference on autonomous agents and multiagent systems (AAMAS)* (pp. 1099–1100). IFAAMAS.

Vinyals, M., Bistaffa, F., Farinelli, A., & Rogers, A. (2012). Coalitional energy purchasing in the smart grid. In *IEEE international energy conference and exhibition (ENERGYCON)* (pp. 848–853). IEEE.

Voice, T., Ramchurn, S., & Jennings, N. (2012). On coalition formation with sparse synergies. In *11th international conference on autonomous agents and multiagent systems (AAMAS)*. IFAAMAS.

Yamamoto, J., & Sycara, K. (2001). A stable and efficient buyer coalition formation scheme for e-marketplaces. In *5th international conference on autonomous agents (AGENTS)* (pp. 576–583). ACM.

Yasir, M., Purvis, M., Purvis, M., & Savarimuthu, B. T. R. (2018). Complementary-based coalition formation for energy microgrids. *Computational Intelligence*, *34*(2), 679–712.

Zhou, L., Lü, K., Yang, P., Wang, L., & Kong, B. (2015). An approach for overlapping and hierarchical community detection in social networks based on coalition formation game theory. *Expert Systems with Applications*, *42*(24), 9634–9646.