

Probability Bounds for Overlapping Coalition Formation

Michail Mamakos

Electrical and Computer Engineering
 Technical University of Crete
 Chania, Greece
 mamakos@intelligence.tuc.gr

Georgios Chalkiadakis

Electrical and Computer Engineering
 Technical University of Crete
 Chania, Greece
 gehalk@intelligence.tuc.gr

Abstract

In this work, we provide novel methods which benefit from obtained probability bounds for assessing the ability of teams of agents to accomplish coalitional tasks. To this end, our first method is based on an improvement of the Paley-Zygmund inequality, while the second and the third ones are devised based on manipulations of the two-sided Chebyshev's inequality and the Hoeffding's inequality, respectively. Agents have no knowledge of the amount of resources others possess; and hold private Bayesian beliefs regarding the potential resource investment of every other agent. Our methods allow agents to demand that certain *confidence levels* are reached, regarding the resource contributions of the various coalitions. In order to tackle real-world scenarios, we allow agents to form *overlapping coalitions*, so that one can simultaneously be part of a number of coalitions. We thus present a protocol for iterated overlapping coalition formation (OCF), through which agents can complete tasks that grant them utility. Agents lie on a social network and their distance affects their likelihood of cooperation towards the completion of a task. We confirm our methods' effectiveness by testing them on both a random graph of 300 nodes and a real-world social network of 4039 nodes.

1 Introduction

Cooperative transferable utility (*TU*) games [Chalkiadakis *et al.*, 2011] have been widely studied as they provide a rich framework for cooperation and coordination among rational agents. In such games, agents make binding agreements for jointly tackling tasks that result to some *utility (coalitional value)* awarded to the team (*coalition*). The importance of their study arises from the fact that in many real-world settings it may be impossible for agents to achieve their goals on their own, or that it is simply more profitable for them to form coalitions. Moreover, interconnected electronic societies and social networks offer a natural environment for the completion of goals of individuals, most of the times by completing tasks through common actions. In this setting, the

formation of a coalition is feasible only if its members are interconnected [Myerson, 1977].

Even though the vast majority of the literature assumes that coalitions have to be disjoint—i.e., an agent can belong in only one coalition at a time—this assumption does not hold in many realistic scenarios. In our work, agents can decide to join a number of *overlapping coalitions*, also referred to as *partial coalitions* [Chalkiadakis *et al.*, 2010].

Moreover, in realistic settings it is natural that agents contribute some amount of their (divisible) resource in order to complete the coalitional task. It is conceivable that an agent is certain only of the quantity of her own resource, while she has private beliefs about others' potential contribution. Thus, acting under incomplete information, she can only probabilistically reason about the success of a potential coalition [Chalkiadakis *et al.*, 2007; Kraus *et al.*, 2003]. Against this background, we provide three novel methods that allow agents to establish probability bounds on the corresponding uncertainty. The bounds are derived given private agent beliefs about others' potential resource investment. These beliefs correspond to *Beta* and *Dirichlet* distributions which can be easily manipulated and updated in a principled manner. Our first method exploits an improvement of the *Paley-Zygmund inequality* [Paley and Zygmund, 1932], while the second and the third proceed by appropriately handling the *two-sided Chebyshev's inequality* [Mitzenmacher and Upfal, 2005] and the *Hoeffding's inequality* [Hoeffding, 1963], respectively. Agents using any of them can demand an arbitrary *confidence level* for the resource contribution of a coalition.

In order for agents to complete tasks by forming overlapping coalitions, an appropriate protocol is required. Thus, we provide a generic protocol for iterated overlapping coalition formation. Under this protocol, in each iteration (round) a number of tasks arrives, and a proposer gets assigned to each task, who is responsible for forming a coalition so that the task gets completed; naturally, the underlying graph structure influences the plausibility of emergence of the coalitions.

To the best of our knowledge, this is the first paper to compute probability bounds for coalition formation under uncertainty. Moreover, this is the first work to tackle *overlapping coalition formation (OCF)* under uncertainty. We evaluate our methods by conducting experiments over both a 300 nodes Erdős-Renyi random graph and a real social network which is a 4039 nodes snapshot from Facebook [Leskovec and Krevl,

2014]. Our results show that our methods consistently outperform, in terms of effectiveness in task completion, a baseline method that selects coalitions based only on expected resource quantity. Moreover, they are also time-efficient, and their behaviour is robust against increases in demanded confidence level. As such, we can conclude that they are indeed suitable for probabilistic reasoning in order to decide which coalitions to form in real-world networks.

2 Background and Related Work

The study of domains with overlapping coalitions was initiated by [Shehory and Kraus, 1998], who provided an approximate solution to the corresponding optimal *coalition structure generation (CSG)* problem [Chalkiadakis *et al.*, 2011]. The subsequent work of [Dang *et al.*, 2006] presented two domain-oriented algorithms for (overlapping) CSG in sensor networks. The formal definition of cooperative games with overlapping coalitions, however, was not until [Chalkiadakis *et al.*, 2010]. In that model, an agent can participate in more than one coalition, by contributing to each a portion of her resources; thus, a partial coalition is given by a vector $r_C = (r_C^1, \dots, r_C^n)$, where $r_C^i \in [0, 1]$ denotes the resource fraction which agent i contributes to coalition C , so $r_C^i = 0$ means that agent i is not part of coalition C . In this way an *overlapping coalition formation game* is given by (i) a set of players $N = \{1, \dots, n\}$ and (ii) a function $v : [0, 1]^n \mapsto \mathbb{R}$.

In [Chalkiadakis *et al.*, 2010], an expressive class of OCF games, *threshold task games (TTGs)* is also put forward. A threshold task game $G = (N; w; t)$ is given by (i) a set of agents $N = \{1, \dots, n\}$, (ii) a vector $w = (w_1, \dots, w_n) \in \mathbb{R}^+$ denoting the quantity of resources the agents possess and (iii) a list $t = (t_1, \dots, t_l)$ of task types where each t_h is described by a threshold value $T_h \geq 0$ and a utility $v_h \geq 0$, so a task type is denoted as $t_h = (T_h, v_h)$. In TTGs, agents form coalitions to complete tasks, in order to gain utility v_h by fulfilling the requirement T_h for t_h .

Coalition formation (focusing on coalitional stability) under uncertainty was studied in [Chalkiadakis and Boutilier, 2004; Jeong and Shoham, 2008; Suijs *et al.*, 1999], while [Kraus *et al.*, 2003] dealt with forming coalitions via auctions. Finally, cooperative games on graphs have been widely studied, both in non-overlapping and in overlapping settings (see, e.g., [Myerson, 1977; Deng and Papadimitriou, 1994; Chalkiadakis *et al.*, 2012; Zick *et al.*, 2012]). However, no previous work has explicitly dealt with the *decision problem* agents face in uncertain overlapping coalitional settings.

3 Modeling Uncertainty

In [Ueda *et al.*, 2011] a concise representation scheme based on the idea of *agent types* is proposed, where agents $i, j \in N$ are recognizably equivalent if for any coalition C , $i, j \notin C$: $v(C \cup i) = v(C \cup j)$. Such a representation cannot apply as a scheme for conciseness in threshold task games, as a task is characterized by a scalar, which is the sum of the resources required for its completion, and not by type vectors (as in [Ueda *et al.*, 2011]). However, we achieve representational conciseness by restricting the quantities of the agent

resources and the threshold values of the task types to be integer values. Without these restrictions, the number of possible partial coalitions would be infinite. We assume that resources are integers, and $q_{max} \in \mathbb{N}^+$ is such that an agent $i \in N$ possesses a resource amount $w_i \in \{1, \dots, q_{max}\}$.

Agent beliefs. Each agent i knows only the quantity of her own resource, while she has a private belief X_j^i about the resource investment of any other agent j . Belief X_j^i is composed of random variables D_j^i and B_j^i , distributed as follows:

- $D_j^i \sim Multinomial(p^{ij} = (p_1^{ij}, \dots, p_{q_{max}}^{ij}))$, where p_r^{ij} is the probability that j offers quantity r of her resource in a coalition.
- $p^{ij} \sim Dirichlet(\alpha^{ij})$, $\alpha^{ij} = (\alpha_1^{ij}, \dots, \alpha_{q_{max}}^{ij})$
- $B_j^i \sim Binomial(q^{ij})$, where q^{ij} is the probability that j accepts an offer by i for participation in a coalition.
- $q^{ij} \sim Beta(a^{ij}, b^{ij})$, where a^{ij} (b^{ij}) corresponds to the number of i 's proposals that j has accepted (declined).

Note that α^{ij} , a^{ij} and b^{ij} above are hyperparameters corresponding to easily updated counters. We estimate p_r^{ij} as $\alpha_r^{ij} / \sum \alpha^{ij}$, and q^{ij} as $a^{ij} / (a^{ij} + b^{ij})$. Thus, $E[X_j^i] = q^{ij} \sum_{r=1}^{q_{max}} p_r^{ij} \cdot r$. Therefore, the expected, according to i 's beliefs, quantity of the sum of resource contribution of a group of agents $X_C^i = \sum_{j \in C} X_j^i$, $C \subseteq N \setminus i$, can be computed in $O(n \cdot q_{max})$ time, through linearity of expectations. We assume that $X_1^i, \dots, X_{i-1}^i, X_{i+1}^i, \dots, X_n^i$ are independent. Due to independence, the variance of X_C^i is the sum of the variances of each X_j^i , $j \in C$, and thus it can be computed in $O(n \cdot q_{max})$ time as well.

4 Obtaining and Exploiting Bounds

In real-world TTG settings, agents act under uncertainty regarding the amount of resources that others may contribute to a coalition. They can thus obtain probability bounds over the resource contribution of a group, and exploit them to select with confidence groups that can carry out their tasks (Alg.1).

We henceforth refer to X_C^i as X , for fixed $i \in N$ and $C \subseteq N \setminus i$, when this causes no confusion. The threshold value T of a task must be exceeded, so that utility u is granted to the members of the successful coalition, and thus agent i is interested in computing $P(X \geq T)$. Since the cost of this computation can be extremely high,¹ the agents resort to computing bounds over this quantity.

4.1 The IPZY Method

Our first method builds on an improvement of the Paley-Zygmund inequality (IPZY). Firstly, we present the Paley-Zygmund inequality [Paley and Zygmund, 1932]:

¹The distribution of the sum of independent integer-valued random variables can be computed via the *Convolution Theorem*, which exploits *Fast Fourier Transform* [Proakis and Manolakis, 1996], in $O(n^2 \cdot q_{max} \cdot \log(n \cdot q_{max}))$ time. However, for large n , this computation time can be prohibitive. By contrast, we obtain probability bounds in $O(n \cdot q_{max})$ time, since the inequalities we utilize depend, at worst, on the expected value and variance of X .

Theorem 1 (Paley-Zygmund inequality). *Let a random variable $X \geq 0$ with finite variance and $0 \leq \theta \leq 1$. It holds:*

$$P(X > \theta E[X]) \geq \frac{(1 - \theta)^2 E[X]^2}{E[X^2]}$$

The Paley-Zygmund inequality can be written as:

$$P(X > \theta E[X]) \geq \frac{(1 - \theta)^2 E[X]^2}{\text{Var}[X] + E[X]^2} \quad (1)$$

We now state the following inequality:

Theorem 2. *Let a random variable $X \geq 0$ with finite variance, and $0 \leq \theta \leq 1$. It holds that:*

$$P(X > \theta E[X]) \geq \frac{(1 - \theta)^2 E[X]^2}{\text{Var}[X] + (1 - \theta)^2 E[X]^2} \quad (2)$$

Proof. Our proof is based on the one-sided Chebyshev's inequality. We do not present it due to space restrictions.² \square

Clearly, Eq. (2) provides a better bound than Eq. (1), since for lower bounds greater values are preferred, and the denominator of Eq. (2) is smaller than the one of Eq. (1). Hence, lower probability bounds can be obtained using Eq. (2).

Agent i wants to assess the ability of a group of agents C to complete task $t = (T, u)$. Thus, a lower bound for $P(X > T)$ can be obtained using Eq. (2), by setting:

$$\theta = T/E[X] \quad (3)$$

Notice that it must hold that $E[X] > T$, or else $\theta \geq 1$.

Furthermore, let i demand a certain *confidence level* c , $0 < c < 1$, so that the bound obtained for a group, using Eq. (2), with θ set as in Eq. (3), is at least c . To select a group of agents, either to directly join or invite its members to join forces together, i would select the group that has the *smallest size* among those that offer her confidence level of at least c (cf. Alg. 1). In this way, i would have her confidence requirement satisfied; and by selecting an appropriate group with the smallest possible size, a greater portion of u could be distributed to each individual of the coalition.

4.2 The CH2S Method

Our second method builds on a manipulation of the two-sided Chebyshev's inequality, and it is thus termed *CH2S*.

Theorem 3 (Two-sided Chebyshev's inequality). *Let a random variable X with finite variance, for any $k > 0$:*

$$P(|X - E[X]| \geq k) \leq \frac{\text{Var}[X]}{k^2} \quad (4)$$

Now, we can manipulate Eq. (4), in order to obtain a *lower probability bound*, in the following way:

$$P(|X - E[X]| < k) \geq 1 - \frac{\text{Var}[X]}{k^2} \quad (5)$$

The probability bound obtained by Eq. (5) can be exploited in a similar way as in the IPZY method, so we adopt the same notation as before. Here again, agent i wants to assess the

ability of a group C to complete task $t = (T, u)$, and the lower bound of Eq. (5) can be obtained, by setting:

$$k = E[X] - T \quad (6)$$

Thus, i can obtain a lower bound, equal to $1 - \text{Var}[X]/k^2$, for the probability of the event that $X \in (T, E[X] + k)$. Agent i can select a group of agents, out of potentially many, as follows. She chooses the one that has the smallest size among those which provide her with a bound, computed by Eq. (5), and with k as in Eq. (6), that exceeds her required confidence level c , $0 < c < 1$.

4.3 The HF Method

The third method (*HF*) is based on Hoeffding's inequality.

Theorem 4 (Hoeffding's inequality). *Let X_1, \dots, X_n be independent random variables, where all X_i are bounded so that $X_i \in [l_i, u_i]$, and let $X = \sum_{i=1}^n X_i$. Then it holds that:*

$$P(|X - E[X]| \geq k) \leq 2 \exp\left(-\frac{2k^2}{\sum_{i=1}^n (u_i - l_i)^2}\right) \quad (7)$$

Hoeffding's inequality can be applied since random variables X_j^i are independent and X_j^i takes integer values in $[0, q_{max}]$. *HF* exploits Eq. (7) identically to the manner that *CH2S* exploits Eq.(4), since both Eq.(4) and Eq.(7) provide upper bounds on both tails, and thus their difference is only on the value of the obtained bound.

Algorithm 1: Selecting a coalition that meets an agent's required confidence level

Data: Agent's confidence level $c \in (0, 1)$

Input: A set of coalitions $\mathbf{C} = \{C_1, C_2, \dots\}$;

Task Threshold T

Output: Coalition $Best \in \mathbf{C}$

```

1  $min \leftarrow +\infty, Best \leftarrow \emptyset, bound \leftarrow 0$ 
2 foreach  $C \in \mathbf{C}$  do
3   if  $method = IPZY$  then
4      $\theta \leftarrow T/E[X]$ 
5      $bound \leftarrow \frac{(1-\theta)^2 E[X]^2}{\text{Var}[X] + (1-\theta)^2 E[X]^2}$ 
6   else
7      $k \leftarrow E[X] - T$ 
8     if  $method = CH2S$  then
9        $bound \leftarrow 1 - \frac{\text{Var}[X]}{k^2}$ 
10    else //  $method = HF$ 
11       $bound \leftarrow 1 - 2 \exp(-2k^2 / (|C| q_{max}^2))$ 
12    if  $bound > c \wedge |C| < min$  then
13       $Best \leftarrow C$ 
14       $min \leftarrow |Best|$ 
15 return  $Best$ 
    
```

4.4 Discussion

The power of our methods lies in the fact that they do not need to compute the exact distribution of X , since they obtain probability bounds. We have assumed variable independence, and thus an agent's beliefs can be represented by

²Note also that an alternative proof appears in Wikipedia.

easy-to-maintain conjugate priors (Betas and Dirichlets) over these variables [Grimmett and Stirzaker, 2001]. However, our methods do not depend in any way on the exact priors used.

Note also that *IPZY* and *CH2S* do apply if *dependence* among the variables holds, as well. Computing the distribution of X would then require the computation of the joint probability distribution of the dependent variables, leading to the following computation for $P(X_C^i \geq T)$ and $|C| = m$:

$$P(X_C^i \geq T) = \sum_{\kappa=T}^{m \cdot q_{max}} P(X_C^i = \kappa) = \sum_{\kappa=T}^{m \cdot q_{max}} P\left(\sum_{j=1}^m X_j^i = \kappa\right)$$

where the first equality holds since the events are mutually exclusive. The number of solutions to the equation $X_1^i + \dots + X_m^i = \kappa$, where every X_j^i is a non-negative integer, is known [Komatsu, 2003] to be $\binom{\kappa+m-1}{m-1} = O((\kappa+m)^m)$. Therefore, the computation of $P(X \geq T)$ is exponential in the number of the agents when $X_j^i, j \in C$, are dependent. However, *IPZY* and *CH2S* would still work though requiring more time, since computing the variance of X would now take $O((n \cdot q_{max})^2)$, due to the computation of the covariance of each pair of variables. Moreover, a different modeling of beliefs as Dirichlets over the Cartesian product of all variables, or as Dirichlet mixture models, would also be required. Experimentation with dependent variables is future work.

5 An OCF Protocol

We now propose a generic protocol for iterated overlapping coalition formation. It models real-world situations where S tasks per iteration arrive, over a period of I iterations (or rounds), and the resources of the agents are *replenished* at the end of each round (as such, there is no need for long-term strategic planning on the part of the agents). Note that the protocol can be applied whether there exists an underlying graph in the environment or not. The agents of a coalition C must be connected by some path, however not all agents connected by a given path have to be members of C . The protocol, summarized in Alg. 2, is kept simple, so that it does not interfere with the agents' deliberation over bounds. To the best of our knowledge, this is the first protocol for decentralized overlapping coalition formation under uncertainty.

At each round, tasks are exogenously created and a distinct proposer is exogenously associated with each task. The assignment of proposers to tasks takes place uniformly at random. Following that, the proposers are concurrently asked to form proposals. Thus, each proposer has to select a group of agents, offering a portion of u (the utility granted by the completion of task $t = (T, u)$) to each of its members, and asking in return for a resource quantity so that the threshold T is met. A proposer can receive proposals from other proposers.

Group selection. The number of possible groups of agents is $O(2^n)$, and thus taking all of them under consideration would be inefficient for a proposer. Instead, we let a proposer i sample K groups of agents, where each agent $j \in N \setminus i$ is included in a group with probability $2^{-D(i,j)}$, where $D(i, j)$ is the geodesic distance (shortest path) between nodes i and j . Such a distance between every pair of nodes can be computed in an offline step. Hence, the closer i is to j , with respect to

their distance in the (undirected) graph, the more likely it is that j will be included in a group sampled by i , while if i and j are not connected, that is $D(i, j) = \infty$, then they cannot cooperate. In this way, the position of an agent in the social network affects the likelihood of making a proposal to another agent, for cooperation towards completing a common task. Thereafter, the proposer uses one of our methods in order to select one of the K groups, and then submits individual $\langle q, \pi \rangle$ proposals to its members. The *requested quantity* $q \in \mathbb{N}^+$ is the rounded average of samples taken from α^{ij} multiplied by q^{ij} (the belief that j accepts a proposal), and *agent payoff* $\pi \in \mathbb{R}^+$ is proportional to q , and is distributed to j upon the completion of the task.

Only the proposer and the proposed-to agent have knowledge of the proposal submitted. Furthermore, all proposals to an agent are revealed to her simultaneously.

Response to proposals. Each agent, who has received at least one proposal, responds by either accepting or rejecting each of the proposals. Therefore, agent j has to select which of the, at most S , proposals of the form $\langle q, \pi \rangle$ to accept, where $q \in \mathbb{N}^+$ and $\pi \in \mathbb{R}^+$, in order to *maximize* her gained utility (at the completion of the tasks), given her resource quantity $w_j \in \mathbb{N}^+$. Thus, the optimal response of an agent is the one that maximizes $\sum_{\tau} x_{\tau} \cdot \pi_{\tau}$, subject to $\sum_{\tau} x_{\tau} \cdot q_{\tau} \leq w_j$, where x is a binary vector of size equal to the number of the proposals that agent j has received. Proposition 1 below follows from a straightforward reduction from KNAPSACK.

Proposition 1. *The optimal response of an agent is NP-Hard.*

Notice that a responder has no knowledge of the other members that a proposer has proposed to, and hence she cannot infer the probability of task completion. Since an agent's response does not depend on the responses of others, the independence assumption regarding agents' beliefs is reasonable.

Beliefs update. If the response of j to proposal $\langle q, \pi \rangle$, submitted by proposer i , is positive then i increases both α^{ij} (Beta update) and α_q^{ij} (Dirichlet update) by 1. If the response is negative then b^{ij} (only Beta update) is increased by 1. Furthermore, if coalition C succeeds in completing a task, then every member of C gets to learn the contribution of every other member, and has her Dirichlet distribution updated in an identical way to that of the proposer.

Notice this allows the modeling of agent behaviour and preferences, e.g. an agent could observe that another agent accepts proposals from the rest of the agents but not from her. We aim to study such phenomena in future work. Finally, after receiving the responses, if the proposer's task has not been accomplished, she covers its remaining needs (if possible).

6 Experimental Evaluation

In this section we provide results on the effectiveness of our methods. We conducted experiments on both an Erdős-Renyi random graph [Bollobás, 2001] of 300 nodes-agents and a real social network—a snapshot of a part of Facebook with 4039 agents [Leskovec and Krevl, 2014]. Our methods are tested for different values of confidence level c demanded by the agents. We compare our methods to a baseline method, in which the group that an agent chooses to make proposals

Algorithm 2: Overlapping coalition formation protocol

```

1  $I \leftarrow$  number of rounds
2  $S \leftarrow$  number of tasks per round
3 for  $i = 1$  to  $I$  do
4    $tasks \leftarrow createTasks(S)$ 
5    $proposers \leftarrow createProposers(tasks)$ 
6   proposers form proposals concurrently, each by
     sampling  $K$  groups and selecting one using {IPZY,
     CH2S, or HF} and Algorithm 1
7   inform proposed members
8   agents accept or reject the proposals they received
9   inform proposers about responses
10  foreach  $uncompleted\ t \in tasks$  do
11    Let proposer of  $t$  complete it if she can
12  foreach  $completed\ t \in tasks$  do
13    inform members about the participation of the
     rest and distribute utility
14  replenish agents' resources and update beliefs
    
```

to is the *smallest* among those whose *expected value* $E[X]$ exceeds the task threshold value T .³ Thus, we refer to this method as *EV*. Selecting *EV* as the baseline is natural: *EV* essentially describes the simplest course of action that an agent acting under uncertainty could take. Moreover, there is no pool of alternative decision-making methods we can use as a baseline in such a setting.⁴

Game parameters. In the experiments on both graphs, q_{max} was set to 30 and the resource weight w_i of each agent was a (rounded to integer) sample from $\mathcal{N}(15, 5^2)$. The hyperparameters of each agent's Beta were initialized to $a^{ij} = 1$ and $b^{ij} = 1$; and the α_r^{ij} of the Dirichlets to $\alpha_r^{ij} = w_j / (D(i, j) \cdot (|r - w_j| + 1))$. In this way α^{ij} is bell-shaped, with its mode being at w_j , the actual (maximum offerable) resource quantity of j . Furthermore, the smaller $D(i, j)$ is, the more sharply peaked the prior is, so the belief updates on α^{ij} have greater impact. The value of K , the number of groups that a proposer i samples, was set to 30. The number of samples taken from α^{ij} , for defining the requested quantity q of i 's proposal $\langle q, \pi \rangle$ to j , was set to 20. The number of rounds I and the number of tasks per round S were set to 200 and 16, respectively,⁵ for each run on both graphs. Before making proposals, a proposer invests 75% of her resource to a task.

All results are average values across 30 runs for each experimental setting. The same set of proposers and tasks were generated at the x^{th} round of the y^{th} run in each setting ($x \in \{1, \dots, I = 200\}$ and $y \in \{1, \dots, 30\}$). We report on the to-

³The smaller the group, the greater the portion of utility u its members receive upon task completion.

⁴The computation of a centralized solution is intractable, and thus it cannot be used as a means of comparison. Even the positive results in [Zick *et al.*, 2012] hold for very restricted cases.

⁵In our experiments we used the standard pseudopolynomial dynamic programming algorithm for KNAPSACK [Kellerer *et al.*, 2004], which, due to the limited number of tasks (and hence proposals), performed in a fraction of time (< 1 ms), as expected (despite Proposition 1); so, agent responses were in fact optimal.

tal number of completed tasks by the agents, where for each setting the used method and the value of c were the same for all agents. We tested values of $c \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$. (In the case of *EV* there is no c .) We also present the (average) *size of the group* that the proposers, using each method, choose as the *best* one to make proposals to. Furthermore, since there is no guarantee that the agents will have their demand for confidence level c fulfilled, we report the *confidence fails*, the number of times that a proposer failed to achieve that c . In those cases, we let the proposer make proposals to the group which provided the confidence level closest to c . Henceforth, we denote the *best group size* by bgs , and the *confidence fails* by cf . The total number of tasks in a run was $I \cdot S = 200 \cdot 16 = 3200$ (thus cf has a max value of 3200). The implementation was in Python 3 and experiments were run on a PC with an i3 3.3GHz processor and 4GB of RAM.

6.1 Erdős-Renyi Graph Model

In this random graph model, edge $\{i, j\}$ is added on the (undirected) graph with probability p . In our setting, we set $n = 300$ and $p = 0.03$ (average node distance = 2.83). The threshold values T of the generated tasks were sampled from $\mathcal{N}(200, 10^2)$, and the utility u of each task was $10 \cdot T$.

As shown in Table 1, our methods consistently outperform *EV*. Even for *IPZY* with $c = 0.5$, the number of completed tasks is much larger compared to that of *EV*, with a very small increase in bgs . The standard deviation of the number of completed tasks is small for every method. Furthermore, when $c \leq 0.8$, proposers using *IPZY* or *CH2S* have their required confidence satisfied almost always. For $c = 0.9$ however, the number of cf increases steeply: for instance, for *CH2S* with $c = 0.8$, only 1.33% = 42.7/3200 of the proposals did not meet that requirement, while for *CH2S* with $c = 0.9$ that percentage increased to 33.33% = 1066.8/3200. Now, lack of confidence, caused by uncertainty, is a major source of downturns in economic environments [Caldara *et al.*, 2016]. Ideally, we would like to have a metric defining the quality of a method, depending on the number of completed tasks and confidence fails, but it is not obvious which one that should be. Thus, *HF* can be deemed unsuitable for this setting, since the number of cf for every value of c is at least 1022.4, and so are *IPZY* and *CH2S* for $c = 0.9$ —notwithstanding that for $c = 0.9$ more tasks were completed, since their completion came with an abrupt increase in cf . On the other hand, both *IPZY* and *CH2S* achieve a confidence level of $c = 0.8$ while maintaining a reasonably low cf number. Then, for *IPZY* and *CH2S*, results presented in Table 3 suggest that the more demanding the agents are (with respect to their confidence level), the better the ratio of the number of completed tasks to best group size is. Overall, Tables 1 and 3 suggest that *CH2S* performs slightly better than *IPZY*.

Finally, we report that the average time for an entire round to be completed, including coalitional evaluations for all 16 tasks and belief updates for all agents—for some multiple times, since they can participate in multiple tasks—was 2.7 sec for *IPZY*, 2.9 sec for *CH2S* and 1.8 sec for *HF*.

Table 1: Results for the Erdős-Renyi graph with 300 agents.

	#completed tasks (%) {std}	bgs	cf
EV	637.46 (19.92) {22.78}	35.16	-
IPZY, $c = 0.5$	876.70 (27.39) {18.50}	36.27	0.1
IPZY, $c = 0.6$	990.76 (30.96) {24.97}	36.87	0.2
IPZY, $c = 0.7$	1202.16 (37.56) {24.83}	38.13	1.6
IPZY, $c = 0.8$	1615.30 (50.47) {22.64}	41.30	14.9
IPZY, $c = 0.9$	2352.51 (73.52) {18.43}	52.95	692.9
CH2S, $c = 0.5$	1115.33 (34.85) {17.29}	37.60	1.1
CH2S, $c = 0.6$	1235.26 (38.60) {22.46}	38.40	2.6
CH2S, $c = 0.7$	1451.73 (45.36) {19.82}	39.94	6.9
CH2S, $c = 0.8$	1843.96 (57.62) {15.46}	43.52	42.7
CH2S, $c = 0.9$	2396.81 (74.91) {16.23}	54.85	1066.8
HF, $c = 0.5$	2316.83 (72.40) {21.63}	52.86	1022.4
HF, $c = 0.6$	2356.80 (73.65) {18.42}	54.95	1616.1
HF, $c = 0.7$	2368.80 (74.02) {11.66}	56.48	2213.9
HF, $c = 0.8$	2370.86 (74.09) {17.45}	57.46	2727.7
HF, $c = 0.9$	2372.03 (74.13) {18.29}	57.88	3095.6

Table 2: Results for the Facebook graph with 4039 agents.

	#completed tasks (%) {std}	bgs	cf
EV	462.26 (14.45) {31.16}	396.12	-
IPZY, $c = 0.5$	562.76 (17.58) {29.62}	398.68	241.1
IPZY, $c = 0.6$	608.31 (19.01) {28.16}	399.97	295.7
IPZY, $c = 0.7$	676.30 (21.13) {25.99}	401.68	397.5
IPZY, $c = 0.8$	803.50 (25.11) {26.95}	404.45	525.4
IPZY, $c = 0.9$	1080.33 (33.76) {24.07}	412.39	742.5
CH2S, $c = 0.5$	648.60 (20.26) {26.29}	400.48	356.8
CH2S, $c = 0.6$	684.61 (21.40) {28.92}	401.27	411.8
CH2S, $c = 0.7$	747.10 (23.35) {27.92}	402.61	482.8
CH2S, $c = 0.8$	860.93 (26.90) {25.56}	405.19	565.4
CH2S, $c = 0.9$	1108.21 (34.63) {24.57}	412.69	799.1
HF, $c = 0.5$	1147.66 (35.86) {11.71}	408.35	772.4
HF, $c = 0.6$	1171.20 (36.60) {15.72}	410.17	852.4
HF, $c = 0.7$	1201.93 (37.56) {16.42}	412.46	946.1
HF, $c = 0.8$	1242.40 (38.82) {12.31}	415.42	1071.1
HF, $c = 0.9$	1285.46 (40.17) {11.92}	420.02	1272.0

6.2 Facebook Snapshot Graph

The graph of the snapshot of Facebook we experimented on consists of 4039 connected nodes (agents). The threshold values T of the generated tasks were sampled from $\mathcal{N}(2250, 40^2)$, and the utility u of each task was $10 \cdot T$.

As observed in Table 2, task completion was more challenging in this sparser network (with an average node distance of 3.69). However, the ordering of the methods, for *IPZY* and *CH2S*, based on the ratio of the number of completed tasks to *bgs*, was virtually the same with that of the Erdős-Renyi graph, as observed in Table 3. Moreover, we observe that proposers using *HF* completed more tasks compared to the other two methods, for every value of c , with the value of *cf* not being tremendously larger. The values in Table 3 corresponding to *HF* were greater compared to those of *IPZY* and *CH2S* for any value of c , and monotonically increase with c , even exceeding 3 for $c = 0.9$. Thus, our methods outperform the baseline in this environment as well; and *HF* is a winner.

The difficulty in achieving the completion of tasks can be attested by the fact that even when proposers used the *EV*

Table 3: # of completed tasks to best group size ratio.

	Erdős-Renyi graph	Facebook graph
EV	18.13	1.17
IPZY, $c = 0.5$	24.17	1.41
IPZY, $c = 0.6$	26.87	1.52
IPZY, $c = 0.7$	31.52	1.68
IPZY, $c = 0.8$	39.11	1.99
IPZY, $c = 0.9$	44.42	2.62
CH2S, $c = 0.5$	29.66	1.62
CH2S, $c = 0.6$	32.16	1.71
CH2S, $c = 0.7$	36.34	1.84
CH2S, $c = 0.8$	42.37	2.12
CH2S, $c = 0.9$	43.69	2.69
HF, $c = 0.5$	43.82	2.81
HF, $c = 0.6$	42.89	2.85
HF, $c = 0.7$	41.94	2.91
HF, $c = 0.8$	41.26	2.99
HF, $c = 0.9$	40.98	3.06

method, about $6.5\% = 207.2/3200$ of them could not find a (sampled) group whose expected value exceeded the task threshold value. In that case they selected the one with the highest expected value. Thus, it is not surprising that for all our three methods, and for every value of confidence level c , proposers frequently could not fulfill their requirement for exceeding c . The fact that *cf* increases smoothly with c , as seen in Table 2, combined with the significant increase of the ratio of completed tasks to best group size for $c = 0.9$, observed in Table 3, lets us conclude that the best value of c in this setting would be 0.9. The average time for the completion of an entire round was 26.5 sec for *IPZY*, 27.7 sec for *CH2S*, and 21.3 sec for *HF*.

7 Conclusions and Future Work

We presented, for the first time in the literature, three methods that derive probability bounds for effective overlapping coalition formation, where the agents have incomplete information of the value of the resources that the other agents can invest. This setting extends straightforwardly to environments with *multiple* (rather than one) resource types. All three of them allow agents to demand *confidence levels*; and significantly outperformed the baseline (which picked coalitions based solely on their expected resource quantity) in terms of the number of tasks completed, and the ratio of this quantity to the size of the group the proposers selected.

Future work includes testing our methods using alternative formation protocols. Moreover, we intend to experiment with settings that allow resources to be depleted over time. This would necessitate long-term strategic planning, which would naturally influence the agents' attempts to derive bounds.

Acknowledgments

Michail Mamakos gratefully acknowledges financial support from the Hellenic Artificial Intelligence Society (EETN) for attending this conference.

References

- [Bollobás, 2001] Béla Bollobás. Random graphs. *Cambridge Studies in Advanced Mathematics*, 2001.
- [Caldara *et al.*, 2016] Dario Caldara, Cristina Fuentes-Albero, Simon Gilchrist, and Egon Zakrajšek. The macroeconomic impact of financial and uncertainty shocks. *European Economic Review*, 2016.
- [Chalkiadakis and Boutilier, 2004] Georgios Chalkiadakis and Craig Boutilier. Bayesian reinforcement learning for coalition formation under uncertainty. In *Proc. of the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1090–1097, 2004.
- [Chalkiadakis *et al.*, 2007] Georgios Chalkiadakis, Evangelos Markakis, and Craig Boutilier. Coalition formation under uncertainty: Bargaining equilibria and the bayesian core stability concept. In *Proc. of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 400–407, 2007.
- [Chalkiadakis *et al.*, 2010] Georgios Chalkiadakis, Edith Elkind, Evangelos Markakis, Maria Polukarov, and Nick R Jennings. Cooperative games with overlapping coalitions. *Journal of Artificial Intelligence Research*, 39(1):179–216, 2010.
- [Chalkiadakis *et al.*, 2011] Georgios Chalkiadakis, Edith Elkind, and Michael Wooldridge. *Computational aspects of cooperative game theory*. Morgan & Claypool Publishers, 2011.
- [Chalkiadakis *et al.*, 2012] Georgios Chalkiadakis, Evangelos Markakis, and Nicholas R Jennings. Coalitional stability in structured environments. In *Proc. of the 11th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 779–786, 2012.
- [Dang *et al.*, 2006] Viet Dung Dang, Rajdeep K. Dash, Alex Rogers, and Nicholas R. Jennings. Overlapping coalition formation for efficient data fusion in multi-sensor networks. In *Proc. of the 21st National Conference on Artificial Intelligence (AAAI)*, pages 635–640, 2006.
- [Deng and Papadimitriou, 1994] Xiaotie Deng and Christos H Papadimitriou. On the complexity of cooperative solution concepts. *Mathematics of Operations Research*, 19(2):257–266, 1994.
- [Grimmett and Stirzaker, 2001] Geoffrey Grimmett and David Stirzaker. *Probability and random processes*. Oxford University Press, 2001.
- [Hoeffding, 1963] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.
- [Jeong and Shoham, 2008] Samuel Jeong and Yoav Shoham. Bayesian coalitional games. In *Proc. of the 23rd National Conference on Artificial Intelligence (AAAI)*, pages 95–100, 2008.
- [Kellerer *et al.*, 2004] Hans Kellerer, Ulrich Pferschy, and David Pisinger. *Knapsack problems*. Springer, Berlin, 2004.
- [Komatsu, 2003] Takao Komatsu. On the number of solutions of the diophantine equation of frobenius-general case. *Mathematical Communications*, 8(2):195–206, 2003.
- [Kraus *et al.*, 2003] Sarit Kraus, Onn Shehory, and Gilad Taase. Coalition formation with uncertain heterogeneous information. In *Proc. of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1–8, 2003.
- [Leskovec and Krevl, 2014] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data>, June 2014.
- [Mitzenmacher and Upfal, 2005] Michael Mitzenmacher and Eli Upfal. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge University Press, 2005.
- [Myerson, 1977] Roger B Myerson. Graphs and cooperation in games. *Mathematics of Operations Research*, 2(3):225–229, 1977.
- [Paley and Zygmund, 1932] REAC Paley and A Zygmund. A note on analytic functions in the unit circle. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 28, pages 266–272. Cambridge Univ Press, 1932.
- [Proakis and Manolakis, 1996] John G Proakis and Dimitris G Manolakis. *Digital Signal Processing: Principles, Algorithms, and Applications*. Prentice Hall, 1996.
- [Shehory and Kraus, 1998] Onn Shehory and Sarit Kraus. Methods for task allocation via agent coalition formation. *Artificial Intelligence*, 101(1):165–200, 1998.
- [Suijs *et al.*, 1999] Jeroen Suijs, Peter Borm, Anja De Waegenare, and Stef Tijs. Cooperative games with stochastic payoffs. *European Journal of Operational Research*, 113(1):193–205, 1999.
- [Ueda *et al.*, 2011] Suguru Ueda, Makoto Kitaki, Atsushi Iwasaki, and Makoto Yokoo. Concise characteristic function representations in coalitional games based on agent types. In *Proc. of the 22nd International Joint Conference on Artificial Intelligence*, pages 393–399, 2011.
- [Zick *et al.*, 2012] Yair Zick, Georgios Chalkiadakis, and Edith Elkind. Overlapping coalition formation games: Charting the tractability frontier. In *Proc. of the 11th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 787–794, 2012.