Stability of Overlapping Coalitions

GEORGIOS CHALKIADAKIS
University of Southampton, UK
and
EDITH ELKIND
Nanyang Technological University, Singapore
and
EVANGELOS MARKAKIS
Athens University of Economics and Business, Greece
and
MARIA POLUKAROV
University of Southampton, UK
and
NICHOLAS R. JENNINGS
University of Southampton, UK

Categories and Subject Descriptors: F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems

General Terms: Economics, Theory

Additional Key Words and Phrases: Coalition Formation

In multiagent systems, it is often the case that agents need to form teams in order to achieve their goals. This may happen even if the agents are controlled by different parties: while in this case the agents do not necessarily share the same set of goals and priorities, they still need to cooperate, as individual agents may not possess sufficient resources to perform their tasks independently. In such scenarios, a natural tool for modeling agent collaboration is *coalitional game theory* [Branzei et al. 2008], a research area that studies collaboration and team formation among rational entities.

The two main questions studied by coalitional game theory are: (1) What teams, or *coalitions*, of agents are likely to arise? and (2) How are the agents going to share the benefits of cooperation? To answer these questions, it is usually assumed that the game ends with the agents forming the *grand coalition*, i.e., the coalition that includes all agents, or a *coalition structure*, i.e., a partition of the set of agents into pairwise disjoint coalitions. The resulting gains are then shared so as to promote *fairness* (i.e., each agent is paid in proportion to his contribution) or *stability* (i.e., the gains are distributed so as to minimize the agents' incentives to deviate from

Authors' addressses: gc2@ecs.soton.ac.uk, eelkind@ntu.edu.sg, markakis@gmail.com, mp3@ecs.soton.ac.uk, nrj@ecs.soton.ac.uk

the current outcome). The set of stable outcomes is usually referred to as the *core*.

Now, when agents are physical entities, such as robots, and the coalitions perform tasks requiring physical movement, it is natural to stipulate that all coalitions in a coalition structure are pairwise disjoint: simply put, no robot can be in two places at the same time. On the other hand, in the case of software agents—or, in fact, humans—this requirement may be too restrictive: it is not unusual for a server to perform several computations in parallel, for a software engineer to work on several assignments at once, or for an investor to provide funds to several start-ups. In such settings, an agent splits his resources (such as processing power, time or money) among several tasks. These tasks, in turn, may require participation of more than one agent: a computation may run on several servers, a software project usually involves more than one engineer, and a start-up may rely on several investors. Thus, each task corresponds to a coalition of agents, but agents' contributions to those coalitions may be fractional, and, moreover, agents can participate in several tasks at once, resulting in *overlapping* coalitions.

In our recent work [Chalkiadakis et al. 2008; Elkind et al. 2009], we proposed a model for overlapping coalition formation, which allows us to handle overlapping coalitions and reason about stability of overlapping coalition structures. In our model, each agent has a unit of a (unique) resource, and the agents can form coalitions by committing a fraction of their resources to a particular task. Thus, the resulting coalitions are *partial*, in the sense that agents are not required to commit all of their resources to any single coalition. Any such coalition can be described by a vector $(r_1, \dots, r_n) \in [0, 1]^n$, where r_i is the fraction of the ith agent's resource contributed to this coalition. An *overlapping coalition formation game (an OCF game)* is then given by its *characteristic function*, which for each partial coalition outputs the value that this coalition can obtain. This value has to be divided among the agents that make a non-zero contribution to this partial coalition. Thus, the outcome of an OCF game is an *overlapping coalition structure*, i.e., a collection of partial coalitions such that no agent allocates more than one unit of resource altogether, and an *imputation*, i.e., a collection of payoff vectors, each of which indicates how the value of the corresponding partial coalition is distributed among its members. Note that we do not allow players to receive payments from partial coalitions they are not contributing to—a natural assumption for most real-life scenarios. However, within a partial coalition its value may be split arbitrarily.

Just as in the non-overlapping setting, a desirable property of an overlapping coalition structure is stability. That is, we are interested in identifying overlapping coalition structures and imputations such that no group of agents can profitably deviate from them. However, defining stability in the overlapping setting is not a straightforward task. Indeed, whereas in the non-overlapping scenario a player deviates by simply withdrawing all of his resources from the coalition in which he participates, in the overlapping setting a player has a choice of withdrawing his contribution from a given partial coalition completely, reducing it, or keeping it unchanged; moreover, he has to make this decision for each partial coalition he is involved in. The definition of stability therefore depends on whether a deviator expects to get any payoff from the coalitions that he did not abandon completely: indeed, it may be profitable for a player to deviate from one of his partial coalitions only if he can keep his payoffs from another partial coalition.

In our work, we consider three different definitions of stability, which vary in terms of what the deviators expect to earn from their coalitions with non-deviators. Our first definition is the most conservative of these three: it assumes that as soon as a player deviates, i.e., changes his contribution to one of his partial coalitions, he does not expect to get any payoffs from his coalitions with non-deviators. Thus, we term the corresponding notion of the core the *conservative* core, or the *c-core*. Among the three definitions of the core that we consider, the conservative core is the closest to the traditional definition of the core. This motivates us to study it in more detail. Specifically, in [Chalkiadakis et al. 2008], we characterize the elements of the c-core by proving that under some technical conditions on the characterisitic function, a pair (overlapping coalition structure, imputation) is in the c-core if and only if the total payoff received by each group of players S is no less than what the players in S could jointly earn by forming partial coalitions among the members of S only. While seemingly simple, this result requires a rather involved proof. The reason for this is that when players in S deviate, the resulting payoff cannot be distributed among them arbitrarily, but needs to respect the boundaries of the partial coalitions formed by the members of S. An important implication of this result is that any outcome in the c-core maximizes the social welfare. We then generalize the definition of balancedness to games with overlapping coalitions, and use it to characterize the overlapping coalition structures that can be stabilized by an appropriate choice of an imputation. Finally, we extend the notion of convexity to OCF games and prove than any convex OCF game has a non-empty c-core.

Now, the notion of the c-core assumes that the deviators are pessimistic, in the sense that they expect the non-deviators to stop collaborating with them. While this model can be appropriate when one deals with human agents, who may treat a deviation as a breach of trust, in case of software agents the situation is less clear: it may be the case that as long as a particular coalition is not affected, its members are not going to punish the deviators, who can therefore count on their share of payoffs from partial coalitions not hurt by the deviation. Thus, in [Elkind et al. 2009] we introduce two alternative notions of stability. The first one assumes that if the deviators do not change their contribution to a particular partial coalition, they can still get their share of that coalition's profits. This implies that the deviators' decision which partial coalition to abandon has to be made on a per coalition basis, rather than for all coalitions simultaneously, Therefore, we call the corresponding notion of the core the *refined core*, or the *r-core*. Note that under this definition it is easier for the deviators to succeed, and, consequently, fewer outcomes are stable, so the r-core is contained in the c-core. The second notion of stability introduced in [Elkind et al. 2009] assumes that the deviators make even more fine-grained choices than in the previous case, while being rather optimistic about the non-deviators' behavior. Namely, they believe that if they withdraw some of their contribution from a partial coalition C so that the value of C goes down, but still exceeds the non-deviators' share of C's original value, then the deviators can claim the rest of the C's profit. We refer to the corresponding notion of the core as the *optimistic core*, or the *o-core*. Clearly, the o-core is contained in the r-core.

We then explore the relationships among these definitions. We have argued above that there is a natural precedence ordering over the three cores: the c-core of a game always contains its r-core, which, in turn, always contains its o-core. We prove that these containments can be strict, and, moreover, there exist games in which the o-core is empty while the r-core is not; the same is true for the r-core vs. the c-core. Thus, the three definitions of the core introduced in our work capture genuinely different notions of OCF-stability.

To develop a better intuition for stability in OCF games, we instantiate our three notions of the core for a simple, but expressive class of coalitional games, which we call *threshold task games (TTGs).* In these games, there is a single type of resource (e.g., time or money), each agent has a certain amount of this resource (which we will refer to as his *weight*), and there is a number of tasks, each of which has a resource requirement and a utility. A (partial) coalition can embark on a task as long as its total weight meets this task's resource requirement; it then obtains the corresponding utility, which can be divided among the members of the coalition in an arbitrary way. Besides providing a natural model for a range of team formation scenarios, these games serve as a convenient testing ground for our notions of stability. Indeed, we use these games to illustrate the differences among our three notions of the overlapping core. Further, given a set of agents' weights and the list of tasks, we can define two variants of the corresponding TTG: one in which overlapping coalition structures are not permitted, and one where they are allowed. This enables us to compare our three notions of the core to the classic notion of the core. Finally, TTGs are compactly representable, which allows us to study computational aspects of stability in these games. In particular, it turns out that our three notions of the core can differ in this respect as well: for example, when players' weights are integers given in unary, the problem of checking whether an outcome is in the c-core of a TTG is polynomial-time solvable, while the corresponding problems for the r-core and the o-core are coNP-hard. These results, besides being interesting in their own right, provide us with a better understanding of the differences among the concepts of stability that we have introduced.

There are obvious similarities between our definition of the OCF games and Aubin's definition of fuzzy games [Aubin 1981]. Indeed, in fuzzy games a player can participate in a coalition at various *levels*, and the value of a coalition S depends on participation levels of the agents in S. Thus, at a first glance, the definition of a fuzzy game seems to be identical to the definition of an OCF game, as both are given by characteristic functions defined on $[0, 1]^n$. However, there are several crucial differences between fuzzy games and OCF games. First, fuzzy games and OCF games differ in their definition of an outcome. Indeed, while in OCF games an outcome is an (overlapping) coalition structure (together with an imputation), in fuzzy games the only allowable outcome is the formation of the grand coalition. Similarly, to be in the c-core (r-core, o-core) of an OCF game, an outcome needs to be stable against any deviation to a (possibly overlapping) coalition structure. In contrast, to be in the Aubin core, an outcome needs to be stable against a deviation to a partial ("fuzzy") coalition, but not necessarily against a deviation to a coalition structure. Indeed, to the best of our knowledge, the formation of coalition structures (overlapping or not) is not addressed in the fuzzy games literature. To overcome this difficulty, one could try to extend the definition of the fuzzy games to allow for coalition structures. However, in [Elkind et al. 2009] we argue that this approach fails to capture several delicate aspects of overlapping coalition formation.

Even more importantly, the definition of the fuzzy core given in [Aubin 1981] is very different from all three notions of the core proposed in our work. Specifically, this definition assumes that when a group of players deviates from the grand coalition, each deviating player i receives both her payoff from the deviation, and her original payoff from the grand coalition, scaled down by a factor of $(1 - r_i)$, where r_i is the amount of resources that she withdraws when deviating. Thus, the fuzzy core is even more "optimistic" from the deviators' perspective than the o-core. Indeed, the deviators do not worry what the grand coalition will be able to do once they leave. They simply assume that if they withdraw, say, 40% of their resources, they will get 60% of what they used to get. However, in many games—and, in particular, in TTGs—if some players (even partially) abandon the grand coalition, the latter may not have sufficient resources to complete any task. Clearly, in this case the deviators could not possibly get any payoff from what remains of the grand coalition. Thus, the fuzzy core may be empty, even if in practice the game is stable. In [Elkind et al. 2009], we provide a specific example of such a game.

To summarize, in [Chalkiadakis et al. 2008] and [Elkind et al. 2009], we have initiated the formal study of games with overlapping coalitions. There is a number of open questions suggested by this work. First, while we have made some progress on understanding the complexity-theoretic aspects of stability in OCF games, so far we have focused on the complexity of checking whether a given outcome is stable. Another natural problem in this domain is checking whether a game has a stable solution, i.e., whether its c-core (r-core, o-core) is non-empty. From a somewhat different angle, it would be interesting to see if the characterization results proved in [Chalkiadakis et al. 2008] for the c-core can be extended to the r-core and the o-core. Finally, an important research direction is to develop a better understanding of scenarios where overlapping coalitions can naturally arise, and to identify the appropriate stability concepts for these scenarios.

REFERENCES

Aubin, J.-P. 1981. Cooperative fuzzy games. *Mathematics of Operations Research 6,* 1, 1–13.

Branzei, R., Dimitrov, D., and Tijs, S. 2008. *Models in Cooperative Game Theory.* Springer.

Chalkiadakis, G., Elkind, E., Markakis, E., and Jennings, N. R. 2008. Overlapping coalition formation. In *Proceedings of the 4th International Workshop on Internet and Network Economics (WINE'08).* 307–321.

Elkind, E., Polukarov, M., Chalkiadakis, G., and Jennings, N. R. 2009. Exploring the stability of overlapping coalitions. Submitted.