A Generic Approach to Coalition Formation

Krzysztof R. Apt and Andreas Witzel

Institute for Logic, Language and Computation
University of Amsterdam
CWI, Amsterdam

Introduction and Motivation

Coalitions are an important notion in cooperative game theory. Many stability concepts exist, but how do stable coalitions come about?

To study coalition formation generically and from an algorithmic point of view, we introduce

- an abstract preference relation over coalition structures
- operators to merge and split coalitions
- a stability notion for coalition structures

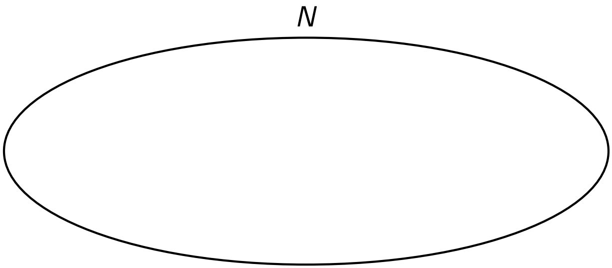
and identify conditions under which

- stable coalition structures exist
- merge and split sequences terminate
- merge and split sequences reach a unique stable outcome

Outline

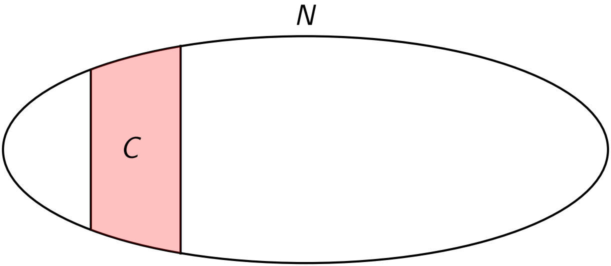
Basic Definitions

- Grand coalition: The complete set of players $N = \{1, \ldots, n\}$



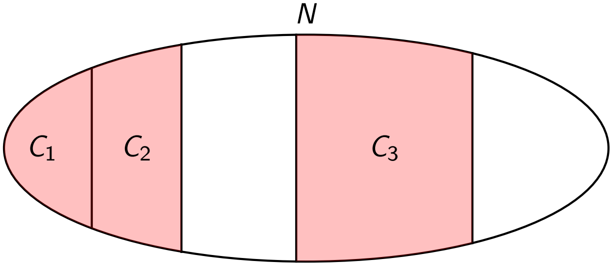
Basic Definitions

- Grand coalition: The complete set of players $N = \{1, \ldots, n\}$
- Coalition: Non-empty subset of N



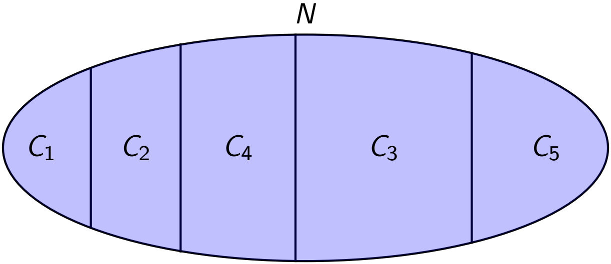
Basic Definitions

- Grand coalition: The complete set of players $N = \{1, \ldots, n\}$
- Coalition: Non-empty subset of N
- Collection: Family $\mathcal{C} = \{C_1, \ldots, C_l\}$ of mutually disjoint coalitions. We abbreviate $\bigcup \mathcal{C} := \bigcup_{i=1}^{l} C_i$



Basic Definitions

- Grand coalition: The complete set of players $N = \{1, \ldots, n\}$
- Coalition: Non-empty subset of N
- Collection: Family $\mathcal{C} = \{C_1, \ldots, C_l\}$ of mutually disjoint coalitions. We abbreviate $\bigcup \mathcal{C} := \bigcup_{i=1}^{l} C_i$
- Partition: Collection $\mathcal{C}$ with $\bigcup \mathcal{C} = N$



Abstract Preference Relation

We need a way to express the social preference over alternative partitions of the same subset of players.

We assume a preference relation $\triangleright$ defined for collections $\mathcal{A}$ and $\mathcal{B}$ with $\bigcup \mathcal{A} = \bigcup \mathcal{B} = M \subseteq N$.

Intuitively, $\mathcal{A} \triangleright \mathcal{B}$ means that the way $\mathcal{A}$ partitions M is preferable to the way $\mathcal{B}$ does.

E.g. $\{\{2,3\},\{5\}\} \triangleright \{\{2\},\{3,5\}\}$ means that, independent of the remaining players, grouping players 2 and 3 together and leaving 5 alone is preferred to leaving 2 alone and grouping 3 and 5 together.

Abstract Preference Relation Properties

For all collections $\mathcal{A}, \mathcal{B}, \mathcal{C}$ with $\bigcup \mathcal{A} = \bigcup \mathcal{B} = \bigcup \mathcal{C}$
and all collections $\mathcal{D}, \mathcal{E}$ with $\bigcup \mathcal{D} = \bigcup \mathcal{E}$ and $\bigcup \mathcal{A} \cap \bigcup \mathcal{D} = \emptyset$,
we assume:

$$\mathcal{A} \rhd \mathcal{B} \rhd \mathcal{C} \text{ imply } \mathcal{A} \rhd \mathcal{C} \qquad \text{(transitive)}$$

$$\mathcal{A} \rhd \mathcal{B} \text{ and } \mathcal{D} \rhd \mathcal{E} \text{ imply } \mathcal{A} \cup \mathcal{D} \rhd \mathcal{B} \cup \mathcal{E} \qquad (\text{monotonic}_1)$$

$$\mathcal{A} \rhd \mathcal{B} \text{ implies } \mathcal{A} \cup \mathcal{D} \rhd \mathcal{B} \cup \mathcal{D} \qquad (\text{monotonic}_2)$$

In certain cases a distinction will be made if additionally
for all collections $\mathcal{A}$ and $\mathcal{B} \neq \mathcal{A}$:

$$\mathcal{A} \not\rhd \mathcal{A} \qquad \text{(irreflexive)}$$

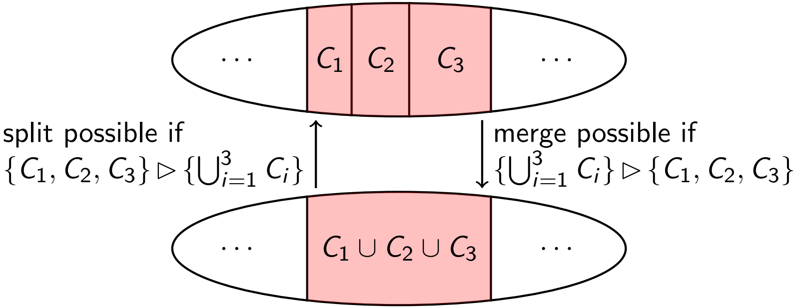
$$\mathcal{A} \rhd \mathcal{B} \text{ or } \mathcal{B} \rhd \mathcal{A} \qquad \text{(total)}$$

Merge and Split Operators

We model coalition formation by two rules describing possible transformations of any given partition of the grand coalition:

merge: $\mathcal{P} \to \mathcal{P} \setminus \mathcal{C} \cup \bigcup \mathcal{C}$, if $\mathcal{C} \subseteq \mathcal{P}$ and $\{\bigcup \mathcal{C}\} \triangleright \mathcal{C}$

split: $\mathcal{P} \to \mathcal{P} \setminus \bigcup \mathcal{C} \cup \mathcal{C}$, if $\bigcup \mathcal{C} \subseteq \mathcal{P}$ and $\mathcal{C} \triangleright \{\bigcup \mathcal{C}\}$



If $\triangleright$ is irreflexive, any sequence of merge and split operations terminates.

Outline

TU-games

As examples, we consider several instantiations of the preference relation in the context of coalitional TU-games.

Reminder: A TU-game is a pair (v, N), where

- N: set of players, as before
- $v : 2^N \rightarrow \mathbb{R}$: value function assigning a value to each coalition

Instantiations of the Preference Relation

For $\mathcal{A} = \{A_1, \ldots, A_m\}$ and $\mathcal{B} = \{B_1, \ldots, B_n\}$ with $\bigcup \mathcal{A} = \bigcup \mathcal{B}$, the following preference relations satisfy the required properties, i.e. transitivity and monotonicity:

- $\mathcal{A} \rhd \mathcal{B}$ iff $\sum_{i=1}^{m} v(A_i) > \sum_{i=1}^{n} v(B_i)$ (utilitarian order)
- $\mathcal{A} \rhd \mathcal{B}$ iff $\prod_{i=1}^{m} v(A_i) > \prod_{i=1}^{n} v(B_i)$ (Nash order)
- $\mathcal{A} \rhd \mathcal{B}$ iff $\max_{A \in \mathcal{A}} v(A) > \max_{B \in \mathcal{B}} v(B)$ (elitist order)
- $\mathcal{A} \rhd \mathcal{B}$ iff $\min_{A \in \mathcal{A}} v(A) > \min_{B \in \mathcal{B}} v(B)$ (egalitarian order)
- $\mathcal{A} \rhd \mathcal{B}$ iff $v^*(\mathcal{A}) >_{lex} v^*(\mathcal{B})$ (leximin order)

$v^*(\mathcal{A})$ denotes the sequence of all $v(A_i)$ in decreasing order and $>_{lex}$ the usual lexicographic order.

One example for a preference relation which is *not* monotonic:

- $\mathcal{A} \rhd \mathcal{B}$ iff $\frac{\sum_{i=1}^{m} v(A_i)}{m} > \frac{\sum_{i=1}^{n} v(B_i)}{n}$ (average order)

Outline

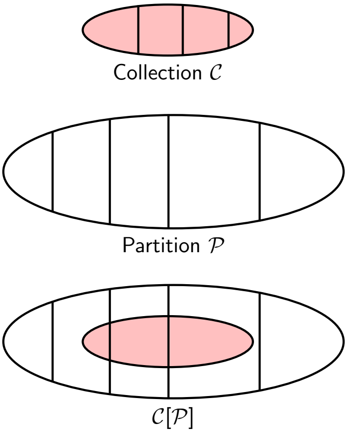
$\mathcal{C}$ in the Frame of $\mathcal{P}$

Given a collection $\mathcal{C}$ and a partition $\mathcal{P} = \{P_1, \ldots, P_k\}$,

$$\mathcal{C}[\mathcal{P}] := \{P_1 \cap \bigcup \mathcal{C}, \ldots, P_k \cap \bigcup \mathcal{C}\} \setminus \{\emptyset\}$$

is called $\mathcal{C}$ in the frame of $\mathcal{P}$.

This notion is used to compare a possible defecting coalition $\mathcal{C}$ with the involved players' current configuration in $\mathcal{P}$.



Stability Notion

A defection function defines what defection are considered possible given some partition $\mathcal{P}$. It yields a family of collections of coalitions.

$\mathcal{C} \in \mathbb{D}(\mathcal{P})$ means that the players in $\mathcal{C} = \{C_1, \ldots, C_l\}$ could leave $\mathcal{P}$ and form l new coalitions according to $\mathcal{C}$.

A partition $\mathcal{P}$ is $\mathbb{D}$-stable iff

$$\mathcal{C}[\mathcal{P}] \rhd \mathcal{C} \quad \text{for all } \mathcal{C} \in \mathbb{D}(\mathcal{P}) \text{ with } \mathcal{C}[\mathcal{P}] \neq \mathcal{C}$$

That is, all possible defecting collections $\mathcal{C}$ prefer their current configuration in $\mathcal{P}$.

Two Defection Functions

Two natural defection functions (both independent of $\mathcal{P}$) are

- $\mathbb{D}_p$, which allows all partitions of the grand coalition:

  $\mathbb{D}_p(\mathcal{P}) = \{\text{all partitions of } N\}$

- $\mathbb{D}_c$, which allows all collections in the grand coalition:

  $\mathbb{D}_c(\mathcal{P}) = \{\text{all partitions of all subsets of } N\}$

Stability Results

Theorem

A partition $\mathcal{P}$ is $\mathbb{D}_p$-stable iff for all partitions $\mathcal{P}' \neq \mathcal{P}$, we have $\mathcal{P} \rhd \mathcal{P}'$.

Corollary

If $\rhd$ is irreflexive and total, then a $\mathbb{D}_p$-stable partition exists.

Theorem

A partition $\mathcal{P} = \{P_1, \ldots, P_k\}$ is $\mathbb{D}_c$-stable iff

- $\{A \cup B\} \rhd \{A, B\}$ *for all disjoint subsets* A, B *of some* P_i*, and*
- $\{T\}[\mathcal{P}] \rhd \{T\}$ *for all* $T \subseteq N$ *which are not subsets of any* P_i*.*

Unique Stable Outcomes of Merge and Split Operations

Theorem

Suppose that $\triangleright$ is irreflexive and $\mathcal{P}$ is a $\mathbb{D}_c$-stable partition. Then

- *$\mathcal{P}$ is the outcome of every exhaustive sequence of merge and split operations starting from any initial partition;*
- *$\mathcal{P}$ is the unique $\mathbb{D}_p$-stable partition; and*
- *$\mathcal{P}$ is the unique $\mathbb{D}_c$-stable partition.*

This result shows the relation between merge and split operations and the stability notion.

A Remark and Future Work

When instantiating the preference relation, it is necessary to check whether the resulting stability notion correctly reflects intuitively stable situations. In the paper, we give an example where this is not the case in the setting of hedonic games (but also one where it does work).

We are currently studying some extensions and their relations to the existing results.
An underlying network structure between the players (representing e.g. friendship relations) might

- determine which coalitions are feasible (e.g. only connected players), or
- induce preferences over coalitions (e.g. distance in friendship network)

Alternatively, in TU-games preferences could be induced by comparison of player values, e.g. the Shapley value.