Basics of Game Theory for Bioinformatics

Fioravante Patrone

DIPTEM, University of Genoa, patrone@diptem.unige.it

Abstract. In this "tutorial" it is offered a quick introduction to game theory and to some suggested readings on the subject. It is also considered a small set of game theoretical applications in the bioinformatics field.

1 Introduction

It could be considered strange the fact that game theory (we shall abbreviate it as GT) is used in the field of bioinformatics. After all, not only GT was created to model economic problems, but also its foundational assumptions are very close relatives of those made in neoclassical economics: basically, the assumption that "players" are "rational" and intelligent decision makers (usually assumed to be human beings). If we deal with viruses, or genes, it is not so clear whether these basic assumptions retain any meaning, and we could proceed further, wondering whether there are "rational decision makers" around.

We shall see that there are some grounds for such an extension in the scope of game theory, but at the same time we acknowledge that there is another reason (almost opposite) to use game theory in the field of bioinformatics. This second reason can be found in the fact that game theory can be seen as "math + intended interpretation". Of course, if we discard the "intended interpretation", we are left only with mathematics: by its very nature, math is "context free", so that we are authorized to use all of the mathematics that has been developed in and for game theory, having in mind whatever "intended meaning" we would like to focus on (a relevant example of this "de-contextualization" is offered in Moretti and Patrone (2008), about the so-called Shapley value). This switch in the interpretation of the mathematical tools needs sound justifications, if it has to be considered a serious scientific contribution, but this can be done, as it has been done also in other very different fields of mathematics.

So, we shall try to emphasize a bit these two approaches to the subject of this tutorial. Since we do not assume any previous knowledge of game theory, we shall start with a very quick sketch of the basics of game theory (section 3), at least to set the possibility of using its language. Since game theory is a subject quite extended in width and depth, for the reader interested to go further (maybe considering to apply GT on its own) we cannot do anything better than providing suggestions for further readings. For this reason, in section 3 we shall provide a concise guide to the main relevant literature, especially to GT books.

We shall then move to illustrate *some* of the GT applications in the field of bioinformatics. Due to our personal contributions to the field, we shall mainly stress the applications of a basic concept and tool for “cooperative games”: the so-called Shapley value. This will be the subject of the first subsection of section 4. The second subsection will be a very quick tour touching some of the diverse ways in which GT has been used in bioinformatics.

2 Game theory: its basics

There is no doubt about the date of birth of GT: it is 1944, year in which *Theory of games and economic behavior*, by John von Neumann and Oskar Morgenstern, appeared. That book built the language of the discipline, its basic models (still used today, with the improvement given by Kuhn (1953) to the definition of games in extensive form), and strongly suggested that GT should be the adequate mathematical tool to model economic phenomena. Of course, there was some GT before 1944 (extremely important is the so-called “minimax theorem” of von Neumann, appeared in 1928), but that book set the stage on which many actors have been playing since then.

From the “birth” of GT it has been accumulated a lot of models, results, applications, foundational deepening. After more than 60 years, GT is really a *non disposable* tool in many areas of economic theory (consumers’ behavior, theory of the firm, industrial organization, auctions, public goods, etc.), and has spread its scope to other social sciences (politics, sociology, law, anthropology, etc.), and much beyond.

The core situation that GT models is a situation in which individuals “strategically interact”. The way in which interaction is meant is the following: there is a set of individuals, each of whom has to make a choice from a set of available actions; it will be the *full set of choices* made by *each* individual to determine the result (or outcome). An important point is that individuals generally have different preferences on the final outcomes. Moreover, interaction is said to be “strategic”: it is important that individuals are aware of this interactive situation, so that they are urged to analyze it, and for that it will be important to know the structure of the interactive situation, the information available, and also the amount of intelligence of the players.

The form which is most used to model such a situation is the so-called “strategic form”. It is quite easy to describe: it is a tuple: $G = (N, (X_i)_{i \in N}, (f_i)_{i \in N})$, where:
- N is the (finite) set of players
- X_i is the set of actions (usually said “strategies”) available to player $i \in N$
- $f_i : X \to \mathbb{R}$, where $X = \prod_{i \in N} X_i$, are the “payoffs”.

Many things should be said about the payoffs, but the main point is the following: given a so-called “strategy profile”, i.e. $x \in X$, it will determine an outcome $h(x)$; $f_i(x)$ is the personal evaluation of the outcome by player i, measured on some scale. Using a terminology which is standard in neoclassical economics, one can see f_i as the composition of the function h with the “utility function” u_i of player i (u_i is defined on the set of outcomes).

We introduce here a couple of very simple (but quite relevant) examples of games in strategic form. The first one is the so-called "prisoner's dilemma": $N = \{I, II\}$, $X_I = X_{II} = \{C, NC\}$, and f_I, f_{II} are described in table 1 (where, e.g., $f_I(C, NC) = 4, f_{II}(C, NC) = 1$.

Table 1. Prisoner's dilemma

$I \backslash II$	C	NC
C	$(2, 2)$	$(4, 1)$
NC	$(1, 4)$	$(3, 3)$

The description of a game by means of a table like table 1 is quite common in the case in which there are two players, with finite strategy sets (the table is often referred as a "bimatrix", since cells contain two numbers, the payoffs for each of the players).

A second example is the "battle of the sexes", which is given in table 2.

Table 2. Battle of the sexes

$I \backslash II$	L	R
T	$(2, 1)$	$(0, 0)$
B	$(0, 0)$	$(1, 2)$

It is important to stress the fact that the standard assumption behind a game modeled in strategic form is that each player chooses his strategy without knowing the choices of the other players (just think of a sealed bid auction).

The relevant question is: what is a "solution" for a game in strategic form? Before answering this question, it is important to notice that game theory can be seen both from a prescriptive (or normative) and a descriptive (or positive) point of view, these "dual" points of view are often found in models for the social sciences. That is: a "solution" can be seen either as:
- what players *should* do (prescriptive point of view)
- what we *expect* that players (will) do (descriptive point of view)

The basic idea of solution for strategic games is the Nash equilibrium, which can find justifications based both on normative and descriptive bases. The formal definition is as follows.

Given a game G, a strategy profile $\bar{x} \in X$ is a Nash equilibrium if:
- for every $i \in N$, $f_i(\bar{x}) \geq f_i(x_i, \bar{x}_{-i})$ for all $x_i \in X_i$ (here x_{-i} denotes the set of the remaining "coordinates" of x, after having deleted the $i - th$).

An essential condition behind the use of the Nash equilibrium as a solution for games in strategic form brings us to consider the so-called "institutional setting" under which the game is played: we refer here to the fact that players *cannot sign* *binding agreements*. This assumption is fundamental in considering the game as a *non cooperative* game: if players can sign binding agreements, we say that we have a *cooperative game*. Namely, if we assume that agreements are not binding, one can see the definition of Nash equilibrium as some kind of stability condition for the agreement $\bar{x}$: the condition that defines a Nash equilibrium amounts to say that no player has an incentive to unilaterally deviate from the agreement.

What are the Nash equilibria for the two examples introduced before? For the prisoner's dilemma, the unique Nash equilibrium is (C, C), which yields a result that *is not efficient*. This surprising fact (players are rational and intelligent!) has been the reason for a lot of interest about this game, which is by far the "most famous" one in GT.

For the battle of the sexes there are *two* Nash equilibria: (T, L) and (B, R). Not only the battle of the sexes provides an example in which the Nash equilibrium is not unique, but it has the additional feature that players have opposite preferences on these two equilibria: player I prefers (T, L), while player II prefers (B, R). But the most interesting point is that the battle of the sexes shows quite clearly that a Nash equilibrium is a *profile* of strategies, and that it is (generally) nonsense to speak of *equilibrium strategies* for the players: which would be the equilibrium strategies for player I?

It has to be stressed that the model of a game in strategic form can be used in principle to represent both cooperative and non-cooperative games, but *de facto* it is essentially used only for non-cooperative games. The same thing can be said of the second basic model used in GT: games in extensive form. We shall not discuss this model, which is useful to describe interactive situations in which the "timing" of the choices made by players is relevant (chess, poker, English auctions): we shall just mention the fact that it can be considered as a natural extension of a "decision tree". It is also important the fact that a game in extensive form can be transformed in a canonical way into a game in strategic form: this fact (already emphasized in von Neumann (1928)) is relevant in extending the scope of games in strategic form, going beyond the immediate and trivial interpretation (players decide "contamporarily"). This is obtained by means of the idea of "strategy" as an action plan, allowing to use games in strategic form o describe also situations in which players are called to make choices that can be based on the observation of previous choices made by other players.

So far, we have almost left out the cooperative games. Even if the strategic and extensive form models can be used, the most common model for cooperative games is given by the so-called games in "characteristic form" (also said "in coalitional form"). Such a model provides a description of the interactive situation which is much less detailed than those provided by the other two models. Apart from this, it is a model which is "outcome oriented", since it displays the "best" results that can be achieved (possibly using binding agreements) by coalitions, i.e., group of players.

We give here the formal definition of a game in characteristic form. To be more precise, we shall confine ourselves to the (simpler) version that is identified as the "transferable utility" case.

A game in characteristic form with transferable utility (quickly: a TU-game) is a couple (N, v), where:
- N is a finite set (the players)
- $v : \mathcal{P}(N) \to \mathbb{R}$, with $v(\emptyset) = 0$ is the so-called "characteristic function".

Of course, the key ingredient is v, the characteristic function, which assigns to each subset of N (usually referred as a "coalition" of players), a real number which has the meaning of describing (in some common scale) the best result that can be achieved by them.

As it can be clearly seen, there is no room for "strategies" in this model: the basic info is the (collective) outcome that can be achieved by each coalition.

Due to the quite different approach w.r.t. the model used for non-cooperative games, it should not come as a surprise that the "solutions" available for TU-games obey different types of conditions, compared to the Nash equilibrium. Before going to describe the two main solution concepts for TU-games (the core and the Shapley value), it is worth emphasizing a relevant difference between cooperative and non-cooperative games for what concerns the approach to the idea of a solution for them. Games in strategic and extensive form, used in the non-cooperative setting, try to offer an adequate description of the possibilities of choice available to the players: this fact allows us to use essentially a basic principle (that of Nash equilibrium) as the ground for solution concepts[1]. On the other hand, the astounding simplicity of a TU-game has a price to be paid: the huge amount of details which is missing in the model has to be recovered somehow when we pass to the "solution" issue. As we shall see, the two main solutions that we shall describe obey different lines of thought. The overall result is that we find significantly different groups of solution concepts: to give just an example, the "bargaining set", the "kernel", the "nucleolus" are three different solutions that all try to incorporate conditions[2] different from those incorporated in the core or the Shapley value.

Having said this, let us consider one of the two anticipated solution concepts for a TU-game: the *core*. In the reminder of this section we shall assume that the TU-games that we shall consider satisfy the *superadditivity* condition[3]:

[1] It must be noticed that there are other relevant solutions, apart the Nash equilibrium. To quote the most relevant, we can mention: subgame perfect equilibrium, perfect equilibrium, proper equilibrium, sequential equilibrium. These different solution concepts try to cope with deficiencies of the Nash equilibrium, but are no more that (relevant) variants of the basic idea behind the Nash equilibrium.

[2] The "bargaining set" tries to take care of the possibilities of "bargaining" among the players, through proposals for an allocation, and the objections and counter objections that can be made w.r.t. the proposed allocation. The "kernel" and the "nucleolus" were introduced having in mind approaches to find allocations belonging to the bargaining set (which is difficult to determine).

[3] It is not an unavoidable condition, but it allows for more "natural" interpretations of formal conditions that we shall use.

- a TU-game is said to be superadditive if $v(S \cup T) \geq v(S) + v(T)$ for any $S, T \subseteq N$, s.t. $S \cap T = \emptyset$.

The interpretation of this condition is quite natural: if two groups of players join together, they should be able to achieve at least the sum of what they are able to achieve acting separately.

For a superadditive game, the best collective results can be achieved when all of the players act together: so, a critical point to identify a solution is to look at $v(N)$ and to reasonable methods to split it among the players.

Some terminology is needed:
- $x = (x_i)_{i \in N} \in \mathbb{R}^N$ is said to be an *allocation*
- an allocation x is said to be an imputation if:
- - $x_i \geq v(\{i\})$ for each $i \in N$ (*individual rationality* condition).
- - $\sum_{i \in N} x_i = v(N)$ (*efficiency* condition).

It is clear that being an imputation is essentially a necessary condition for a reasonable way of dividing $v(N)$ among the players: if the individual rationality condition is violated for some i, this player will prefer to "play alone", instead of joining the "grand coalition" N.

The "core" of a TU-game can be seen as an almost straightforward extension of the individual rationality condition to some sort of "group rationality". The definition is actually quite simple.

An allocation $x = (x_i)_{i \in N} \in \mathbb{R}^N$ belongs to the *core* of the game (N, v) if:
- $\sum_{i \in S} x_i \geq v(S)$ for all $S \subseteq N$
- $\sum_{i \in N} x_i = v(N)$.

It could be surprising, maybe, but the core of a superadditive TU-game can be empty. A standard example is given by the so-called simple majority game ($N = \{1, 2, 3\}$): $v(N) = v(\{1, 2\}) = v(\{1, 3\}) = v(\{2, 3\}) = 1$, $v(\{1\}) = v(\{2\}) = v(\{3\}) = 0$. If an allocation satisfies the conditions for being in the core, it must be:
- $x_1 + x_2 \geq 1$
- $x_1 + x_3 \geq 1$
- $x_2 + x_3 \geq 1$

From this we get that $x_1 + x_2 + x_3 \geq 3/2$, which contradicts the efficiency condition.

Another interesting example is given by the so-called "glove game". Members of N belong to two subsets that partition N, "L" and "R", whose intended meaning is that each player is owner of exactly one "left" or "right" glove, respectively. Unpaired gloves are worthless, pairs of glove have value 1: this is enough to get a TU-game (clearly $v(S) = \min\{S \cap L, S \cap R\}$).

If #R $=$ #L $+1$, there is a *unique* allocation in the core: "left" owners get 1, "right" owners get 0. This is true for all values of #L, even if #L $= 10^6$. If one can imagine that left owners have some kind of advantage, due to the relative scarcity of left gloves w.r.t. right gloves, this advantage intuitively seems to be not so strong (especially for "big" values of #R) to justify the fact that left owners would get all, leaving nothing to right gloves owners.

Despite of these examples, the core is an important solution concept, and has been proved to be especially useful in economic applications, in particular for the class of the so-called market games, for which it can be proved (under reasonable assumptions) that it is always non empty. It is interesting to notice that, despite its simplicity, it was not proposed as a solution in the seminal book of von Neumann and Morgenstern. It appeared later, in 1953, proposed by Gillies in his PhD thesis (Gillies (1953)) that was devoted to the analysis of "stable sets": the idea of a "stable set" was the proposal by von Neumann and Morgenstern as a solution for TU-games. This kind of solution, quite interesting from a theoretical point of view, proved to be unpractical, except for games with few players, offering moreover in some cases results that do not offer any kind of sensible interpretation.

In the same year in which the idea of the "core" appeared, another, quite different, proposal was made by Shapley (1953). There are many reasons of interest for the solution proposed by Shapley (now named *Shapley value*), and we shall list here quickly the most prominent:
- contrary to the core, which is a set of imputations, it is a *single valued* solution
- it was conceived as a way to "predict" the gain that a player could expect (in average), playing a given game
- it was introduced in an "axiomatic" way[4]: that is, Shapley introduced a set of conditions that a solution should satisfy, and proved that there is a unique way to associate to any TU-game an allocation that satisfies these conditions.

Let us see a set of axioms that determine the Shapley value (it is a set of conditions a bit different from that used by Shapley, but the differences are more formal than substantial and offer us a quicker way to approach it). Given a set of players, N, consider the set of all superadditive TU-games having N as player set: $\mathcal{SG}(N)$. We want to find a map $\Phi : \mathcal{SG}(N) \to \mathbb{R}^N$ that satisfies the following conditions:

- EFF $\sum_{i \in N} \Phi((N,v))_i = v(N)$.
- SYM Symmetric players get the same. Two players i and j are symmetric if $(v(S) \cup \{i\}) \backslash \{j\} = (v(S) \cup \{j\}) \backslash \{i\}$ for all $S \subseteq N$.
- NPP If i is a "null player", i.e. $v(S \cup \{i\}) = v(S)$ for all $S \subseteq N$, then i gets zero.
- ADD $\Phi(v+w) = \Phi(v) + \Phi(w)$ for all $v, w \in \mathcal{SG}(N)$. The game $v+w$ is defined "pointwise" as follows: $(v+w)(S) = v(S) + w(S)$ for any coalition S.

The interesting result is that *there is a unique* Φ satisfying these four conditions, and this map is called "Shapley value". It can be shown that the Shapley value is some appropriate average of the marginal contributions of the players ("marginal contribution" of player i to a coalition S is $v(S \cup \{i\}) - v(S)$), from where one can get the following formula that expresses the Shapley value:

---

[4] The axiomatic approach had been applied in a very successful way by Arrow (1951) and Nash (1950) few years before.

$\Phi_i(v) = \frac{1}{n!} \sum_{\sigma} m_i^{\sigma}(v)$.

Here σ is a permutation of N; j is the unique element in N s.t. $i = \sigma(j)$, and $m_i^{\sigma}(v) = v(\{\sigma(1), \sigma(2), \ldots, \sigma(j)\}) - v(\{\sigma(1), \sigma(2), \ldots, \sigma(j-1)\})$.

As could be expected, in the simple majority game the Shapley value assigns to each player 1/3 (this is immediately seen using SYM and EFF). For the glove game, if $\#\mathrm{L} = 10^6$, we have that "L" players get approximately 0.500443 and "R" players 0.499557. A much more "reasonable" result than those offered by the core.

It is interesting to notice that *significantly different* axiomatic characterizations are available: Young (1985), Hart and Mas-Colell (1987).

Another aspect worth being mentioned is that the axiomatic characterization given by Shapley not necessarily "works" if we restrict the attention to subsets of $\mathcal{SG}(N)$. A particularly relevant example is provided by the class of *simple games*: a TU-game is simple if v assumes just the values 0 or 1, and $v(N) = 1$. It is obvious that summing two simple games we do not get a simple game (at least, $v(N)$ would be equal to 2), so the ADD condition becomes completely irrelevant, and the three remaining axioms are not enough for a complete characterization of the Shapley value for this class of games. Dubey (1975) has shown that it is enough to substitute ADD with TRANSF:

- TRANSF For any two simple games v and w, $\Phi(v \vee w) + \Phi(v \wedge w) = \Phi(v) + \Phi(w)$.

A similar problem is found for the class of microarray games, which we shall extensively discuss in section 4.

3 Suggested readings

As said in the introduction, the room allowed for introducing basic facts about GT is obviously too small to consider even quite relevant questions: to quote just one, a due discussion about the meaning of the payoffs and of the values of the characteristic form. So, we shall provide a relatively small set of bibliographical items that could be useful for readers willing to go further these small notes.

First of all, books that are good as a "preliminary step", almost at a divulgative level: Davis (1970) and Straffin (1996).

A more complete introduction to GT is offered by the books of Osborne (2003) , Binmore (1992), Dutta (1999), Luce and Raiffa (1957).

At a more advanced level, I would suggest three books that are essentially texts for graduate students: Owen (1968), a very good source for cooperative games; Myerson (1991), whose highlights are (in my humble opinion) the chapters on utility theory, on communication and on cooperation under uncertainty; Osborne and Rubinstein (1994), that I would recommend for repeated games and implementation theory. A text that is worth to consider as a general reference source is the Handbook of Game Theory, in three volumes, edited by Aumann and Hart (1992).

It is worth mentioning, having in mind the bio*informatics* reader, a recent volume entitled "Algorithmic Game Theory", edited by Nisam *et al.*, (2007).

Interesting sources to discuss foundational and interpretational issues are the already quoted book by Luce and Raiffa, Kreps (1990) and Aumann (1987).

From the point of view of web-based sources, it is quite rich the web page maintained by Shor: `http://www.gametheory.net`. A lot of material (most of which in Italian, but it can be found also a video course on GT, available for download and in English) can be found also in my web page:
`http://dri.diptem.unige.it`. To conclude, I mention a video lecture, by Tommi Jaakkola, MIT: "Game theoretic models in molecular biology",
`http://videolectures.net/pmsb06_jaakkola_gtmmb/`.

4 Applications of Game Theory

We shall stress in this section applications that use the Shapley value for cooperative games offering, for other applications, just a quick description and references.

4.1 Using the Shapley value

A first application that we shall briefly describe is Kaufman *et al.*, (2004). Focus of this contribution is the RAD6 DNA repair pathway for *S. Cerevisiae* (yeast). In such a pathway are involved 12 genes, and the question is to measure the relevance of each gene for this pathway. The laboratory technique used is somehow classical, i.e. the knock-out of genes. Critical factor is that *groups* of genes are knocked-out: in such a way, it is identified a "coalition" S, composed of the remaining genes. The value $v(S)$ assigned to S is the survival rate to UV radiation exposure. In such a way one gets a TU-game (N, v).

There is a non-trivial problem, however. Even if the knock-out technique is classical, one has to consider that the number of subsets of $N = \{1, 2, \dots, 12\}$ is $2^{12} = 4096$, so that it is not sensible to conduct all of the experiments needed to get "the whole game" (N, v). This is a quite common issue in "practical" applications of GT, and has been somehow analyzed in Moretti and Patrone (2004), in which has been taken into account the fact that getting the data for knowing (N, v) is a costly operation so that one has to balance costs with the main aim of the application (in that paper, the focus was on the fairness of the imputation of costs).

Getting back to the contribution of Kaufman *et al.* (2004), they propose the Shapley value to evaluate the relevance of the different genes in the RAD6 pathway. Of course, having not all of the data for (N, v), they need some approximation/estimation techniques, for which we refer the reader to the paper.

We shall switch to another application of GT, to which we shall devote some more space to it, since it presents at least a couple of non-trivial questions related with the application of the GT tools.

In Moretti *et al.* (2007) it has been introduced a subclass of TU-games, named "microarray games", aiming at answering the following research question: find a *relevance index* for genes (using the data obtained by a microarray gene experiment), taking into account the way in which groups of genes "move together" when switching from healthy to ill condition.

We shall focus here on microarray data from mRNA expression. Having selected a set of genes to analyze, each sample from an experiment will provide an array of data (expression levels for the selected genes), so that the outcome of an experiment will be a matrix collecting the data from all samples. The starting point for applying GT is to obtain a TU-game (a microarray game) from such a matrix. This is achieved in two steps: the first one is a *discretization* procedure, which converts the expression matrix into a boolean matrix. This is done fixing thresholds[5] that distinguish, e.g., "normally expressed" genes (to which it will be attributed the value 0 in the boolean matrix) and "abnormally expressed" genes (value 1).

Given a boolean matrix and a set ("coalition") of genes S, $v(S)$ is defined in the following way: it is the number of samples for which S contains all of the genes to which is assigned value 1, divided by the total number of samples.

Having the microarray game (N, v), one can apply all of the mathematical tools developed for TU-games, and refer to any kind of solution, Shapley value included. Of course, it remains the problem of justifying the use of a particular "solution", given the context. The path followed in the paper is somehow reversed, w.r.t. what has been just noticed. The idea followed is to propose a set of conditions that a "relevance index" should satisfy and then see what one gets. A critical role is reserved to *partnerships* of genes. Partnerships in the context of TU-games were introduced by Kalai and Samet (1988) to analyze weighted Shapley value. The definition is the following:
- given (N, v), $S \in \mathcal{P}(N) \setminus \{\emptyset\}$ is a partnership if: for each $T \subsetneq S$ and each $R \subseteq N \setminus S$, $v(R \cup T) = v(R)$.

So, a partnership, in the setting of microarray games, represents a set of genes that are relevant when they "move together", and so it is natural to attribute an adequate "weight" to them. The assumptions made on the "relevance index" (a map F defined on the set of microarray games with gene set N, $\mathcal{M}(N)$, to $\mathbb{R}^N$) that is sought for, are given by the following three conditions ("axioms") which involve partnerships:

- PR The solution F has the *Partnership Rationality* property, if $\sum_{i \in S} F_i(N, v) \geq v(S)$ for each $S \in \mathcal{P}(N) \setminus \{\emptyset\}$ such that S is a partnership of genes in the game v.

The PR property determines a lower bound of the power of a partnership, i.e., the total relevance of a partnership of genes in determining the onset of the tumor in the individuals should not be lower than the average number of cases of tumor enforced by the partnership itself.

- PF The solution F has the *Partnership Feasibility* property, if $\sum_{i \in S} F_i(N, v) \leq$

[5] The choice of the threshold is clearly delicate: we refer to the original paper of Moretti *et al.* (2007) for some discussion about it.

$v(N)$ for each $S \in \mathcal{P}(N) \setminus \{\emptyset\}$ such that S is a partnership of genes in the game v.
Contrary to PR, the PF properties fixes a natural upper bound of the power of a partnership: the total relevance of a partnership of genes in determining the tumor onset in the individuals should not be greater than the average number of cases of tumor enforced by the grand coalition.

- PM The solution F has the *Partnership Monotonicity* property, if $F_i(N, v) \geq F_j(N, v)$ for each $i \in S$ and each $j \in T$, where $S, T \in \mathcal{P}(N) \setminus \{\emptyset\}$ are partnerships of genes in v such that $S \cap T = \emptyset$, $v(S) = v(T)$, $v(S \cup T) = v(N)$, $|S| \leq |T|$.

The PM property is very intuitive: genes in the smaller partnership should receive a higher relevance index than genes in the bigger one, where the likelihood that some genes are redundant is higher.

To these crucial axioms it has been added a couple of "almost obvious" conditions:

- NG For a null gene (i.e., null "player") $i \in N$, $F_i(N, v) = 0$
- ES We refer to Moretti *et al.* (2007) for the somehow technical definition. The intended meaning is that all of the samples deserve the same importance.

It turns out that there is a unique "relevance index" that satisfies these five conditions, and that it coincides with the Shapley value.

This theoretical result poses a "practical" computational question: is it possible to compute the Shapley value for a given microarray game, taking into account that the "players" are of the order of thousands, or tenths of thousands? The answer, luckily, is positive. Not only it is possible, but the computation is also very easy and very quick: the reason for that has to be found in the fact that the Shapley value can be computed directly from the discretized (boolean) matrix, *without having to consider* the microarray game. This is a fact not new in GT: for example, in the field of the so-called "Operation Research Games" it often happens that a GT solution can be computed directly form the data of the original OR problem, without the need of knowing the game that has been built on it, the "Bird rule" being perhaps the most well know example of this kind[6]. Going back to "microarray games" and Shapley value, the meaning of what we have noticed is that the role of GT is mainly connected with the *justification* of the computations that will rank genes in order of relevance.

Applications of this approach have been made in Moretti *et al.* (2007) to colon cancer data, provided in Alon *et al.* (1999); to data about neuroblastoma in Albino *et al.* (2008); to the influence of pollution, in Moretti *et al.* (2008), where the GT analysis is combined with classical statistical tools; to autism, in Esteban and Wall (2009). It has also been made a comparison between the use of

[6] It provides an element of the core of a "minimum cost spanning tree" game, just looking directly at the information on the connection costs that are associated to the connection graph, see Bird (1976).

the Shapley value and of the Banzhaf value[7] for microarray games, in Lucchetti *et al.* (2009).

4.2 A quick look at other GT applications

We shall give here a very concise description of few GT applications that offer an idea of the different GT tools and different applied fields, falling under the heading of "Game theoretic applications in bioinformatics".

In Kovács *et al.* (2005), in the context of energy and topological networks of proteins, it is discussed the possibility of introducing *protein games* to reduce the number of energy landscapes occurring when a protein binds to another macromolecule.

Bellomo and Delitala (2008) analyze the role of mathematical kinetic theory of active particles to model the early stages of cancer phenomena, including in the available tools also *stochastic games*.

Wolf and Arkin (2003) stress that regulatory network dynamics in bacteria can be analyzed using motifs, modules and games. In particular, it is suggested that the dynamics of modules determines the strategies used by the organism to survive in the *game* in which cells and the environment are involved.

Frick and Schuster (2003) suggest that, for some range of the relevant parameters, microorganisms that can choose between two different pathways for ATP production can be engaged in a *prisoner's dilemma game*. The effect is that of an "inefficient" outcome, meaning in this setting a waste of resources.

In Nowik (2009), it is described the competition that takes place among motoneurons that innervate a muscle. The problem is modeled as a *game*, where the "strategies" for the motoneurons are their "activity level". This model is able to justify the emergence of the so-called "size principle", and to provide predictions which can be subjected to experimental testing.

The last paper that we mention is Schuster *et al.* (2008), which is a survey on the use of *evolutionary game theory* to analyze biochemical and biophysical systems.

References

Albino, D., Scaruffi, P., Moretti, S., Coco, S., Truini, M., Di Cristofano, C., Cavazzana, A., Stigliani, S., Bonassi, S., Tonini, G.P.: Identification of low intratumoral gene expression heterogeneity in Neuroblastic Tumors by wide-genome expression analysis and game theory. Cancer, **113**, 1412Ũ-1422, (2008).

Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack D., Levine, A.J.: Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissue probed by oligonucleotide arrays. Proceedings of the National Academy of Sciences, **96**, 6745–6750, (1999).

Arrow, K.J.: Social Choice and Individual Values. Wiley, New York (1951); second edition: 1963.

[7] The Banzhaf value was originally introduced as a "power index" for simple games by Banzhaf (1965), and then it has been extended to a value defined on al of $\mathcal{SG}(N)$.

Aumann, R.J.: Game Theory, in “The new Palgrave Dictionary of Economics”, (editors: J. Eatwell, M. Milgate e P. Newman), Macmillan, London, 460-482 (1987).

Aumann, R.J., Hart, S. (editors): Handbook of Game Theory. 3 vols, North-Holland, Amsterdam (1992).

Banzhaf, J.F. III: Weighted voting doesn’t work: A game theoretic approach. Rutgers Law Review, **19**, 317–343, (1965).

Bellomo, N., Delitala, M.: From the mathematical kinetic, and stochastic game theory to modelling mutations, onset, progression and immune competition of cancer cells. Physics of Life Reviews, **5**, 183Ũ-206, (2008).

Binmore, K.: Fun and games. Heath and Company, Lexington (MA) (1992).

Bird, C.G.: On cost allocation for a spanning tree: A game theoretic approach. Networks, **6**, 335–350, (1976).

Davis, M.D.: Game Theory: A Nontechnical Introduction. Basic Books, New York (NY, USA) (1970); second edition: 1983, reprinted 1997 by Dover, Mineola (NY, USA).

Dubey, P.: On the Uniqueness of the Shapley Value. International Journal of Game Theory, **4**, 131–139 (1975).

Dutta, P.K.: Strategies and Games: Theory and Practice. MIT Press, (1999).

Esteban, F.J., Wall, D.P.: Using game theory to detect genes involved in Autism Spectrum Disorder. Top, in press, DOI - 10.1007/s11750-009-0111-6, (2009).

Frick, T., Schuster, S.: An example of the prisoner’s dilemma in biochemistry . Journal Naturwissenschaften, **90**, 327–331, (2003).

Gillies, D.B.: Some Theorems on n-Person Games. PhD thesis, Department of Mathematics, Princeton University, Princeton (1953).

Hart, S., Mas-Colell, A.: Potential, value and consistency. Econometrica, **57**, 589–614 (1987).

Kalai, E., Samet, D.: Weighted Shapley Values. In: “The Shapley Value, Essays in Honor of Lloyd S. Shapley”, A. Roth (ed.), Cambridge University Press, Cambridge, 83–100 (1988).

Kaufman, A., Kupiec, M., Ruppin, E.: Multi-knockout genetic network analysis: the Rad6 example. In: “Proceedings of the 2004 IEEE computational systems bioinformatics conference” (CSB 04), (2004).

Kovács, I., Szalay, M., Csermely, P.: Water and molecular chaperones act as weak links of protein folding networks: Energy landscape and punctuated equilibrium changes point towards a game theory of proteins. FEBS Letters, **579**, Issue 11, Pages 2254–2260, (2005).

Kreps, D.M.: Game Theory and Economic Modeling. Oxford University Press, Oxford (1990).

Kuhn, H.W.: Extensive Games and Problems of Information. in “Contributions to the Theory of Games”, II, (editors: H.W. Kuhn e A.W. Tucker), Annals of Math. Studies, **28** (1953).

Lucchetti, R., Moretti, S., Patrone, F., Radrizzani, P.: The Shapley and Banzhaf values in microarray games. Computers & Operations Research, in press, DOI: 10.1016/j.cor.2009.02.020, (2009).

Luce, R.D., Raiffa, H.: Games and Decisions. Wiley, New York (1957).

Moretti, S., van Leeuwen, D., Gmuender, H., Bonassi, S., van Delft, J., Kleinjans, J., Patrone, F., Merlo, F.: Combining Shapley value and statistics to the analysis of gene expression data in children exposed to air pollution. BMC Bioinformatics, **9**, 361, (2008).

Moretti, S., Patrone, F.: Cost Allocation Games with Information Costs. Mathematical Methods of Operations Research, **59**, 419–434, (2004).

Moretti, S., Patrone, F.: Transversality of the Shapley value. Top, **16**, 1-41, (2008). Invited paper: from pages 42 to 59, comments by Vito Fragnelli, Michel Grabisch, Claus-Jochen Haake, Ignacio García-Jurado, Joaquín Sánchez-Soriano, Stef Tijs; rejoinder at pages 60-61

Moretti, S., Patrone, F., Bonassi, S.: The class of Microarray games and the relevance index for genes. Top, **15**, 256–280, (2007)

Myerson, R.B.: Game Theory: Analysis of Conflict. Harvard University Press, Cambridge (MA) (1991).

Nash, J.F. Jr.: The Bargaining Problem. Econometrica, **18**, 155–162 (1950).

von Neumann, J.: Zur Theorie der Gesellschaftsspiele. Matematische Annalen, **100**, 295–320 (1928).

von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press, Princeton (1944); second edition: 1947; third edition: 1953.

Nisam, N., Roughgarden, T., Tardos, É., Vazirani, V.V. (eds.): Algorithmic Game Theory. Cambridge University Press, Cambridge, (2007). Freely downloadable a non-printable pdf from: `http://www.cambridge.org/journals/nisan/downloads/nisan_non-printable.pdf`.

Nowik,I.: The game motoneurons play. Games and Economic Behavior, **66**, 426–461, (2009).

Osborne, M.J.: An introduction to game theory. Oxford University Press, Oxford, (2003).

Osborne, M., Rubinstein A.: A course in Game Theory. MIT Press, Cambridge (MA), (1994). Freely downloadable (upon registration) from: `http://theory.economics.utoronto.ca/books/`.

Owen, G.: Game Theory. Academic Press, New York, (1968); second edition: 1982; third edition: 1995.

Schuster, S., Kreft, J.-U., Schroeter, A., Pfeiffer, T.: Use of Game-Theoretical Methods in Biochemistry and Biophysics. Journal of Biological Physics, **34**, 1–17, (2008).

Shapley, L.S.: A Value for n-Person Games. In “Contributions to the Theory of Games”, II, (editors: H.W. Kuhn e A.W. Tucker), Annals of Math. Studies, **28**, Princeton University Press, Princeton (NJ), 307–317 (1953).

Straffin, P.: Game Theory and Strategy. The Mathematical Association of America, Washington, DC. 1995.

M Wolf, D.M., Arkin, A.P.: Motifs, modules and games in bacteria. Current Opinion in Microbiology, **6**, 125–134, (2003).

Young, H.P.: Monotonic Solutions of Cooperative Games. International Journal of Game Theory **14**, 65–72 (1985).