Tilburg University

A New Approximation Method for the Shapley Value Applied to the WTC 9/11 Terrorist Attack

Hamers, Herbert; Husslage, Bart; Lindelauf, R.; Campen, Tjeerd

A NEW APPROXIMATION METHOD FOR THE SHAPLEY VALUE APPLIED TO THE WTC 9/11 TERRORIST ATTACK

By

Tjeerd van Campen, Herbert Hamers, Bart Husslage,
Roy Lindelauf

21 November 2016

TILBURG ◆ UNIVERSITY

A new approximation method for the Shapley value applied to the WTC 9/11 terrorist attack

Tjeerd van Campen[1]   Herbert Hamers[1]   Bart Husslage[2]
Roy Lindelauf[3]

November 21, 2016

Abstract

The Shapley value (Shapley (1953)) is one of the most prominent one-point solution concepts in cooperative game theory that divides revenues (or cost, power) that can be obtained by cooperation of players in the game. The Shapley value is mathematically characterized by properties that have appealing real-world interpretations and hence its use in practical settings is easily justified.

The down part is that its computational complexity increases exponentially with the number of players in the game. Therefore, in practical problems that consist of more that 25 players the calculation of the Shapley value is usually too time expensive. Among others the Shapley value is applied in the analysis of terrorist networks (cf. Lindelauf et al. (2013)) which generally extend beyond the size of 25 players.

In this paper we therefore present a new method to approximate the Shapley value by refining the random sampling method introduced by Castro et al. (2009). We show that our method outperforms the random sampling method, reducing the average error in the Shapley value approximation by almost 30%. Moreover, our new method enables us to analyze the extended WTC 9/11 network of Krebs (2002) that consists of 69 members. This in contrast to the restricted WTC 9/11 network considered in Lindelauf et al. (2013), that only considered the operational cells consisting of the 19 hijackers that conducted the attack.

1 Introduction

Cooperative game theory with transferable utilities studies situations in which players can work together and create additional revenues (or costs reductions) instead of the situation in which

---

[1]CentER and Department of Econometrics and Operations Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.

[2]Department of Mathematics, Fontys University of Applied Sciences, P.O. Box 90900, 5000 GA Tilburg, The Netherlands. Corresponding author: b.husslage@fontys.nl

[3]Military Operational Art & Science, Netherlands Defense Academy, P.O. Box 90002, 4800 PA Breda, The Netherlands.

each player acts on its own. When all players involved in the game work together the so-called *grand coalition* forms. The objective is to find an allocation of the total revenue (or cost) of the grand coalition among the players that is considered 'fair' by all (sets of) players.

There are many solution concepts in cooperative game theory, each satisfying its own set of properties. One of the most prominent one-point solution concepts satisfying intuitive properties is the Shapley value (Shapley (1953)). It is defined as the average of all $n!$ marginal vectors in a cooperative game consisting of n players. Each marginal vector is determined by a specific order of the players, say σ, in which the player in the i-th position in σ, say player k, is allocated the difference between the value of the coalition consisting of k and its predecessors and the value of the coalition consisting only of the predecessors of k. Although the concept of marginal contributions is intuitive clear, computing the Shapley value as the average marginal contribution of each player is considerably harder. The Shapley value has both theoretical and practical merits. From a theoretical perspective, for instance, the Shapley value can be characterized using properties that are in general considered as appealing and fair (cf. Shapley (1953), Young (1985)) or it can be regarded as a special case of the theory of power indexes (Shapley and Shubik (1954)). An overview and further generalizations of the Shapley value are provided in the book edited by Roth (1988).

The Shapley value has been applied to many real-world cases. Consider the airport landing fee problem described in Littlechild and Owen (1973). Littlechild and Owen proposed a division of the landing fees using the Shapley value. More recent, Deidda et al. (2009) use the Shapley value to allocate the cost of transporting water from reservoirs to the users in the Turia river basin. In genetics Moretti et al. (2007) use the Shapley value of the microarray game to determine the most important genes for specific body functions. Lindelauf et al. (2013) model terrorist networks by connectivity games and use the Shapley value to identify key players in such networks. More applications of the Shapley value in different fields can be found in the survey of Moretti and Patrone (2008).

The biggest challenge in applying the Shapley value to real-world applications is its computational time which generally increases exponentially with the number of players. More precisely, it can be shown that computing the Shapley value is an NP-complete problem (cf. Deng and Papadimitriou (1994), Faigle and Kern (1992)).

A notable exception is the class of games that are mathematically defined in such a way that it becomes possible to compute the Shapley value in polynomial time (e.g. airport games (Littlechild and Owen (1973)) and sequencing games (Curiel et al. (1989))). In the context of terrorist networks Michalak et al. (2013) developed an algorithm to compute the Shapley value for connectivity games on networks that is polynomial in the number of connected subgraphs when the underlying network structure is known. Another exception is when properties of the Shapley value can be used to determine it in polynomial time (e.g. unanimity games (Shapley (1953))). Unfortunately the majority of games lacks both a convenient mathematical structure and Shapley value properties that can be determined in polynomial time. Then even in a game consisting of 25 players the computation of the Shapley value is too time expensive. Real-world applications more often than not deal with larger number of players, hence underlining the importance of heuristics in determining the Shapley value.

Two kinds of heuristics have been developed up to now. The first one focusses on the special class of simple games, i.e., games that only attain a value of 0 or 1. Within this class voting games receive most attention. See for instance Fatima et al. (2008) who present an algorithm

that has time complexity linear in the number of players. They show that their algorithm out-performs the following approximation methods: Monte Carlo simulation (Mann and Shapley (1960)), multi-linear extension (MLE) (Owen (1972)), modified MLE (Leech (2003)) and random permutation (Zlotkin and Rosenschein (1994)). Moreover, they provide some error bound analysis and show that their algorithm is also applicable to k-majority games.

The second type of heuristic uses the average of a random subset of marginal vectors, see Castro et al. (2009). This approximation method for the Shapley value has been used more often lately in the literature. The majority of these papers focusses on specific classes of games. In Bachrach et al. (2010), for example, the focus is again on simple games. Liben-Nowell et al. (2012) consider convex games. Furthermore, Narayanam and Narahari (2008) and Narayanam and Narahari (2011) use this method in social networks.

One area where better approximation methods for the Shapley value are needed is in the ranking procedure of individuals in networks of a terrorist, insurgent or criminal nature. For instance in Lindelauf et al. (2013) the operational WTC 9/11 network is analyzed, consisting of the 19 hijackers of the four planes involved in the attack. Using a cooperative game that takes both compositional and structural variables into account a ranking is obtained using the Shapley value. However the extended WTC 9/11 network consists of 69 members (see Krebs (2002)) which is too large to analyze with the Shapley value in order to obtain a ranking of the players in a reasonable amount of time. The algorithm of Michalak et al. (2013) that considers connected subgraphs is also too time expensive as the number of such subgraphs becomes too large. Furthermore there is a need for a general approximation method to the Shapley value that is also applicable to games without network structure.

In this paper we therefore introduce a refinement of the random sampling method introduced by Castro et al. (2009) and establish rankings for all the 69 members in the extended WTC 9/11 network. We do this by first selecting a random set of permutations of all players in the game. Second we modify these permutations in such a way that each player attains each position in the permutation the same number of times. This slight modification results in a better performance than the heuristic of Castro et al. (2009). On average the error in the Shapley value approximation is reduced by almost 30%.

This paper is organized as follows. In Section 2 some well-known aspects of the Shapley value are recalled. Section 3 discusses the heuristic of Castro et al. (2009) and our new heuristic which we will refer to as *structured random sampling*. Section 4 provides the performance analysis of the structured random sampling method. Section 5 applies our heuristic to the network of hijackers and accomplices involved in the 9/11 attack. Section 6 ends with conclusions.

2   The Shapley value

A cooperative game is a pair (N, v), where $N = \{1, 2, \ldots, n\}$ denotes the set of players. These players can cooperate and form different coalitions. A map v assigns a value $v(S)$ to each possible coalition $S \subseteq N$, which reflects the potential benefit or power of coalition S. By definition $v(\emptyset) = 0$. The Shapley value allocates the value of the grand coalition, i.e., $v(N)$, over all individual players. The most commonly used definition of the Shapley value is based on marginal vectors. Let Π contain all possible orderings of the players in the grand coalition N and let $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N) \in \Pi$ be one such ordering. If player i is at position k, i.e., $\sigma_k = i$, then its marginal contribution $m_v^\sigma(i)$ is defined as $m_v^\sigma(i) = v(\{\sigma_1, \ldots, \sigma_k\}) - v(\{\sigma_1, \ldots, \sigma_{k-1}\})$. Hence,

the marginal contribution of player i is the extra value (or benefit) that player i contributes to the already established coalition $\{\sigma_1, \ldots, \sigma_{k-1}\}$. Clearly, the marginal contribution of a player depends on the chosen ordering σ. Considering all $n!$ possible orderings in Π, we define the Shapley value $\varphi_i(v)$ of player i as its average marginal contribution:

$$\varphi_i(v) = \frac{1}{n!} \sum_{\sigma \in \Pi} m_v^\sigma(i). \tag{1}$$

The following example illustrates how to compute the Shapley value via the marginal contributions of players.

Example 2.1 (*Computing the Shapley value via marginal contributions*). Consider the 3-person game (N, v) with v as in Table 1.

S	$\emptyset$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
$v(S)$	0	1	3	0	5	7	4	10

Table 1: An example of a 3-person game.

There are $3! = 6$ orderings possible in this 3-person game. The corresponding marginal contributions are depicted in Table 2.

σ	$m_v^\sigma(1)$	$m_v^\sigma(2)$	$m_v^\sigma(3)$
$(1,2,3)$	1	4	5
$(1,3,2)$	1	3	6
$(2,1,3)$	2	3	5
$(2,3,1)$	6	3	1
$(3,1,2)$	7	3	0
$(3,2,1)$	6	4	0

Table 2: The marginal contributions of the 3-person game in Table 1.

Averaging the marginal contributions of each player results in the Shapley value: $\varphi(v) = (3\frac{5}{6}, 3\frac{1}{3}, 2\frac{5}{6})$. $\square$

Computing the Shapley value via marginal contributions is time-expensive since all $n!$ possible orderings of the players need to be considered. In several special cases, however, a time-efficient closed formula to compute the Shapley value exists. This is the case, for example, when the game (N, v) is represented as a linear combination of unanimity games. Given $N = \{1, 2, \ldots, n\}$ and a non-empty subset $T \subseteq N$ of players, a unanimity game u_T is defined as

$$u_T(S) = \begin{cases} 1 & \text{if } T \subseteq S, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

for each possible coalition $S \subseteq N$. Interpretation: a benefit of 1 can only be attained when all players in T are involved in the coalition. Consider a game (N, v) that is represented as a

4

linear combination of several unanimity games, i.e., $v(S) = \sum_{T \subseteq N, T \neq \emptyset} c_T \cdot u_T(S)$, where c_T is the coefficient of the corresponding unanimity game. For this sum-of-unanimity-games game (or, shorter, SOUG game) the Shapley value of player i can readily be computed (cf. Shapley (1953)):

$$\varphi_i(v) = \sum_{T:T \ni i} \frac{c_T}{|T|}. \tag{3}$$

Any cooperative game can be represented as a SOUG game (c.f. Shapley (1953)) and this representation provides an efficient way to compute the Shapley value by means of (3). Unfortunately, to reformulate a random cooperative game as a SOUG game all $2^n - 1$ coalition values $v(S)$ of the game (N, v) need to be considered. Hence, this procedure is only useful when the SOUG game is already provided.

The following example illustrates how to efficiently compute the Shapley value when the underlying SOUG game is provided.

Example 2.2 (*Computing the Shapley value for a SOUG game*). Consider the following 3-person unanimity games $u_{\{1\}}$, $u_{\{1,2\}}$ and $u_{\{2,3\}}$. Table 3 presents the coalition values for each unanimity game.

S	$\emptyset$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
$u_{\{1\}}$	0	1	0	0	1	1	0	1
$u_{\{1,2\}}$	0	0	0	0	1	0	0	1
$u_{\{2,3\}}$	0	0	0	0	0	0	1	1

Table 3: An example of three unanimity games.

Let $c_{\{1\}} = 2$, $c_{\{1,2\}} = -3$, $c_{\{2,3\}} = 5$ and let all other coefficients equal 0. Then, for the corresponding 3-person SOUG game (N, v) it follows that $v(S) = 2 \cdot u_{\{1\}}(S) - 3 \cdot u_{\{1,2\}}(S) + 5 \cdot u_{\{2,3\}}(S)$, see Table 4.

S	$\emptyset$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
$v(S)$	0	2	0	0	-1	2	5	4

Table 4: The 3-person SOUG game corresponding to Table 3.

Applying the procedure of Example 2.1 we could compute the average marginal contribution of each player, and, hence, the Shapley value. Note, however, that (3) provides a straightforward way to compute the Shapley value by considering the explicit formulation of the SOUG game. For example, the Shapley value of player 2 is $\varphi_2(v) = \frac{c_{\{2\}}}{|\{2\}|} + \frac{c_{\{1,2\}}}{|\{1,2\}|} + \frac{c_{\{2,3\}}}{|\{2,3\}|} + \frac{c_{\{1,2,3\}}}{|\{1,2,3\}|} = \frac{0}{1} + \frac{-3}{2} + \frac{5}{2} + \frac{0}{3} = 1$. In a similar way the Shapley values of the other two players follow and $\varphi(v) = (\frac{1}{2}, 1, 2\frac{1}{2})$. $\quad\rule{0.3em}{0.6em}$

3 Approximation methods

In cases where the Shapley value of a cooperative game is too time expensive to compute we resort to approximation methods. Such methods are either based on the class of the underlying game or, for more general classes of games, are based on sampling methods. In this section we first review the nowadays popular random sampling method. Then we introduce a new approximation method, the so-called structured random sampling method. We illustrate each approximation method by an example.

3.1 Random sampling

In the last decade there has been a focus on approximating the Shapley value of arbitrary games with many players using a random sample of the marginal vectors. This random sampling method was first introduced by Castro et al. (2009). They applied the following three-step procedure.

Procedure random sampling

Input: n-person cooperative game (N, v).

1. Select a subset Π_r of r orderings from all $n!$ possible orderings, i.e., $\Pi_r \subset \Pi$.

2. Compute the marginal contributions $m_v^\sigma(i)$ for all players $i \in N$ and for all orderings $\sigma \in \Pi_r$.

3. Approximate the Shapley value for each player i by averaging the marginal contributions obtained at step 2, i.e., $\hat{\varphi}_i(v) = \frac{1}{r} \sum_{\sigma \in \Pi_r} m_v^\sigma(i)$.

Notice that in step 1 the modeler can choose how many orderings he wants to use, i.e., he selects the value of r. We will use the random sampling method as a benchmark for our new structured random sampling method. In the following example we apply the random sampling method to the 3-person game introduced in Section 2 (Table 1) with $r = 3$. In Table 5 one possible choice of orderings is given.

Example 3.1 (*Approximating the Shapley value via random sampling*)**.** Consider the 3-person game (N, v) of Example 2.1. Say we sample half of the possible orderings, i.e., $r = 3$. Then compute the marginal contributions for each of the three players in the sampled orderings. Table 5 depicts a possible scenario. Finally, average the marginal contributions to obtain an approximated Shapley value for each player. Hence, $\hat{\varphi}(v) = (3, 3\frac{1}{3}, 3\frac{2}{3})$.

σ	$m_v^\sigma(1)$	$m_v^\sigma(2)$	$m_v^\sigma(3)$
$(1, 2, 3)$	1	4	5
$(1, 3, 2)$	1	3	6
$(3, 1, 2)$	7	3	0

Table 5: The marginal contributions of randomly selected orderings of the 3-person game in Table 1.

In Example 3.1 the approximated Shapley values for players 1 and 3 differ substantially from the real Shapley values $\varphi_1(v) = 3\frac{5}{6}$ and $\varphi_3(v) = 2\frac{5}{6}$. In this example the difference is mainly due to the lack of orderings in the sample in which player 1 is the last player since these orderings largely influence the Shapley value of both players 1 and 3. Hence, influential orderings like these can easily be missed when sampling at random. This observation led to the idea to ensure that each player attains each position in the ordering the same number of times. This is why we propose to use a new approximation method that adds structure to the orderings that are sampled.

3.2 Structured random sampling

Structured random sampling method is inspired by the random sampling method. The idea is to ensure that each player attains each position in the ordering the same number of times. To realize this the randomly selected orderings are tweaked by swapping players to their preferred positions in the orderings. The marginal contributions of the players in these new orderings are then used to approximate the Shapley value.

The swapping method is illustrated for a 4-person game in Table 6. In this case $r = 8$ orderings are randomly selected and divided into $n = 4$ groups of size $t = 2$. In the two orderings in the first group player 1 is swapped with the player at the first position. In the two orderings in the second group player 1 is swapped with the player at the second position et cetera. These new orderings are used to compute the marginal contributions of player 1 and they in turn are averaged to approximate the Shapley value of player 1. This procedure is then repeated for players 2 to n, each time using the original randomly selected r orderings as a starting point for the swapping method. In Table 6 player 1 attains each position in the ordering exactly $t = 2$ times. The remaining positions in the orderings however remain random. The same holds for players 2 to n when the swapping method is applied to construct the orderings for these players.

Group	σ_1	σ_2	σ_3	σ_4		Group	σ_1	σ_2	σ_3	σ_4
1	.	.	.	1		1	1	.	.	σ_1
1	.	1	.	.		1	1	σ_1	.	.
2	1	.	.	.		2	σ_2	1	.	.
2	.	1	.	.	$\rightarrow$	2	.	1	.	.
3	.	.	.	1		3	.	.	1	σ_3
3	.	1	.	.		3	.	σ_3	1	.
4	1	.	.	.		4	σ_4	.	.	1
4	.	1	.	.		4	.	σ_4	.	1

Table 6: Swapping player 1 to his preferred positions in the orderings.

The procedure to approximate the Shapley value of an arbitrary game using the structured random sampling method is as follows.

Procedure structured random sampling

Input: n-person cooperative game (N, v).

1. Select a subset Π_r of r orderings from all $n!$ possible orderings, i.e., $\Pi_r \subset \Pi$, with $r = t \cdot n$ and $t \in \mathbb{N}$.

2. Divide the subset Π_r in n groups of size t.

3. For each player i:

   (a) Swap player i with the player at position j for each of the t orderings in group j, where $j \in \{1, \ldots, n\}$, resulting in a set Π'_r of r new orderings.

   (b) Compute the marginal contributions $m_v^\sigma(i)$ of player i for all new orderings $\sigma \in \Pi'_r$.

   (c) Approximate the Shapley value of player i by averaging the marginal contributions obtained at step 3b, i.e., $\hat{\varphi}_i(v) = \frac{1}{r} \sum_{\sigma \in \Pi'_r} m_v^\sigma(i)$.

The following example illustrates how the structured random sampling procedure could be applied to the cooperative game of Example 2.1.

Example 3.2 (*Approximating the Shapley value via structured random sampling*). Consider again the 3-person game (N, v) of Example 2.1. Assume that we sample the same subset Π_r of $r = 3$ orderings as in the random sampling example, see the second column of Table 7. Divide this subset into $n = 3$ groups of size $t = 1$. Now consider player 1. Swapping this player with the player at the first, second and third position results in the new orderings depicted in the third column of Table 7. The fourth column in this table depicts the corresponding marginal contributions $m_v^\sigma(1)$ of player 1 in the new orderings. Averaging these marginal contributions yields an approximation of the Shapley value for player 1, i.e., $\hat{\varphi}_1(v) = (1 + 7 + 6)/3 = 4\frac{2}{3}$.

Group	Ordering	Swap 1	$m_v^\sigma(1)$	Swap 2	$m_v^\sigma(2)$	Swap 3	$m_v^\sigma(3)$
1	$(1,2,3)$	$(1,2,3)$	1	$(2,1,3)$	3	$(3,2,1)$	0
2	$(1,3,2)$	$(3,1,2)$	7	$(1,2,3)$	4	$(1,3,2)$	6
3	$(3,1,2)$	$(3,2,1)$	6	$(3,1,2)$	3	$(2,1,3)$	5

Table 7: The marginal contributions of randomly selected orderings of the 3-person game in Table 1.

Starting again from the original subset Π_r in the second column and swapping player 2 with the player at the first, second and third position results in a new subset of orderings for player 2, see the fifth column of Table 7. The sixth column depicts the corresponding marginal contributions, resulting in $\hat{\varphi}_2(v) = 3\frac{1}{3}$. Repeating this swapping method for the third player results in the orderings and marginal contributions depicted in the seventh and eight column of Table 7, which in turn lead to $\hat{\varphi}_3(v) = 3\frac{2}{3}$. Hence, $\hat{\varphi}(v) = (4\frac{2}{3}, 3\frac{1}{3}, 3\frac{2}{3})$. ⬜

Two important observations follow from Examples 3.1 and 3.2. First, both methods calculate the same number of marginal contributions per player in order to approximate the Shapley value. The structured random sampling method however also needs to swap players. This extra operation results in a slightly larger computation time, see Section 4.3. Second, random sampling is efficient, i.e., $v(N)$ is distributed over the n players, whereas structured random sampling is not efficient. This lack of efficiency is due to the fact that structured random

sampling only considers marginal contributions of a single player for each sampled ordering. In spite of these two drawbacks the structured random sampling method outperforms the random sampling method when it comes to approximating the Shapley value as will be shown in the next section.

4   Performance analysis

In this section the performance of our proposed structured random sampling method is compared to the performance of the random sampling method. The idea is to apply both sampling methods on arbitrary games with many players and compare the results with one another as well as with the exact Shapley values of these games. First we describe how the errors between the approximated and exact Shapley values are measured. Then we run two types of performance analyses: one on the number of orderings used in the sampling methods and one on the number of players in the cooperative game.

We test the performance of both sampling methods for the class of SOUG games. In order to generate random SOUG games the number of players, the number of unanimity games and the corresponding coefficients of these unanimity games have to be chosen randomly. Using this class of games has two main advantages. First, since each cooperative game can be represented as a SOUG game, and the SOUG games are generated randomly, the cooperative games used in the analysis are also random. Second, since the generated SOUG games are known, the exact Shapley values of the players in the game can be computed efficiently, see Example 2.2. Hence, we will be able to compute error measures for both approximation methods by comparing the approximated Shapley values with the exact Shapley values.

The error measures are computed as follows. For a randomly generated SOUG game v_j the exact Shapley values as well as the approximated Shapley values are computed for all players in the game. Then both the absolute error and the percentage error are computed for each player. Averaging these errors leads to the average absolute error and the average percentage error of the sampling method for the corresponding game v_j. This provides insight in the absolute as well as the relative size of the errors incurred by our approximation method. To improve the validity of the error measures not just one but 50 random SOUG games v_j are generated, hence, $j = 1, \ldots, 50$. Averaging the average absolute error and the average percentage error results in the error measures 'average average absolute error' (AAAE) and 'average average percentage error' (AAPE) for these 50 random SOUG games v_j, i.e.,

$$\text{AAAE} = \frac{1}{50} \sum_{j=1}^{50} \left(\frac{1}{n} \sum_{i=1}^{n} |\hat{\varphi}_i(v_j) - \varphi_i(v_j)| \right) \tag{4}$$

$$\text{AAPE} = \frac{1}{50} \sum_{j=1}^{50} \left(\frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{\varphi}_i(v_j) - \varphi_i(v_j)|}{|\varphi_i(v_j)|} \right) \tag{5}$$

These average error measures provide for better benchmarking of the approximation methods to the exact method. To facilitate averaging over multiple games the value of the grand coalition, i.e., $v_j(N)$, in each random SOUG game is normalized. The general procedure to compute the error measures is summarized in the following four-step approach.

Procedure error measures

1. Randomly generate 50 SOUG games and normalize the value of the grand coalition in each game.

2. Compute the exact Shapley values for all players in all 50 games.

3. Use random sampling to approximate the Shapley values for all players in all 50 games and compute the error measures AAAE en AAPE.

4. Use structured random sampling to approximate the Shapley values for all players in all 50 games and compute the error measures AAAE en AAPE.

A note on the computation of the error measure AAPE. Consider a player with a Shapley value close to zero. This Shapley value may be approximated very well by (structured) random sampling, but may still result in a relatively large percentage error. Therefore, in the computation of the error measure AAPE only players with a Shapley value larger than 0.1% of $v(N)$ are taken into account.

4.1   Number of orderings

The idea behind both sampling methods is to consider only a fraction of the total number of orderings of the players. It is expected that the approximations will be better, i.e., closer to the exact Shapley values, when the number of orderings used in the sampling methods increases. A simulation consists of 50 randomly generated SOUG games. Each of these SOUG games consists of 100 players and is constructed from 100 unanimity games, with a randomly selected subset $T \subset N$. The coefficients, which are multiplied with the unanimity games, are random integers taken from the interval $[-10, 10]$. In the simulations the number of orderings vary from only 500 orderings up to 5000 orderings. Note however that for a game with 100 players an astoundingly $100! \approx 9.33 \cdot 10^{157}$ different orderings of the players exist. Even when sampling 5000 orderings only a tiny fraction of the total number of orderings is used in the approximation, i.e., less than 0.01%.

Figure 1 shows the results of the performance analysis on the number of orderings. In this figure the 'average average absolute error' (AAAE) and 'average average percentage error' (AAPE) are provided for both sampling methods when the number of orderings is varied from 500 up to 5000 orderings.

From this figure it can be seen that average absolute errors (the AAAEs) for both sampling methods are very small. The average percentage errors (the AAPEs) however are quite large. For example, using 5000 orderings the AAPE is 16% for random sampling and 11% for structured random sampling. Hence, more orderings should be sampled to obtain more accurate approximations. Still, considering the tiny fraction of orderings used in the approximation of the Shapley values the results of both sampling methods are promising. Furthermore, it can be seen that structured random sampling outperforms random sampling. That is, the errors in the approximations from structured random sampling are significantly smaller than the errors of random sampling. This effect is even more profound when the number of orderings is small, i.e., close to 500.
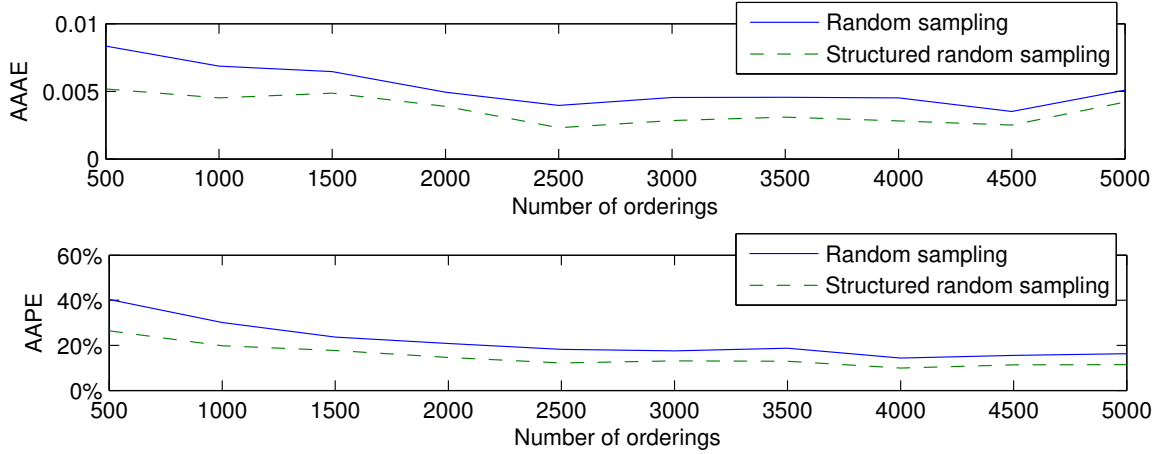
Figure 1: Performance analysis on the number of orderings.

4.2 Number of players

The second analysis tests the performance of both approximation methods for different numbers of players. Again arbitrary games are considered by randomly generating 50 SOUG games in each simulation. The parameter settings of the SOUG games construction are quite similar to the settings in the performance analysis of the number of orderings. The only differences with the previous analysis are that the number of orderings is fixed to 5000 orderings and the number of players in each game varies from 10 to 100 players. Furthermore, the normalized value of the grand coalition in each game of n players is factored by $0.01n$ thereby enabling the comparison of the absolute errors in the approximation methods for games with different numbers of players.

Figure 2 shows the results of the performance analysis on the number of orderings. In this figure the 'average average absolute error' (AAAE) and 'average average percentage error' (AAPE) are provided for both samplings methods while the number of players is varied from 10 up to 100 players.

Both sampling methods are seen to have small AAAEs en AAPEs, but these errors seem to increase as the number of players increases. This is most likely due to the fact that the number of orderings used in all the approximations is fixed to 5000 orderings. When the number of players increases the number of possible orderings increases as well. Hence, a smaller fraction of the total number of orderings is sampled and the approximations are expected to be less accurate. Nevertheless, the approximation errors increase only slowly when the number of players increases. For example, the AAPEs of random sampling and structured random sampling are respectively 4% and 3% in a game of 10 players. These numbers are respectively 13% and 9% in a game of 100 players. Hence, a factor 10 increase in the number of players results in only a factor 3 increase in the average errors. Again, structured random sampling outperforms random sampling. This time the effect is even more profound when the number of players in the game increases.
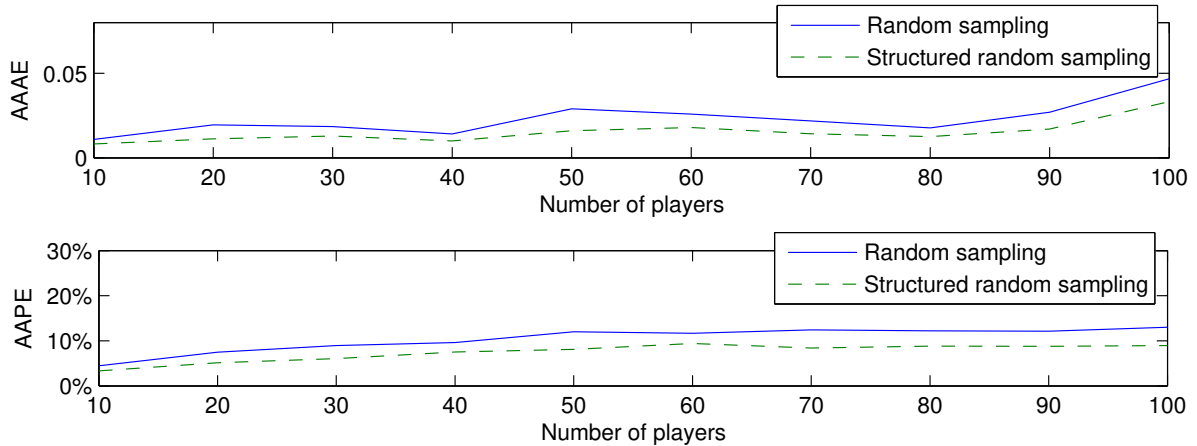
Figure 2: Performance analysis on the number of players.

4.3 A note on computation times

Each performance analysis implies running several simulations. In each simulation 50 random SOUG games are considered. All simulations and computations were performed with MATLAB, release R2014, on a computer with four Intel X3430 processors of 2.4Ghz and with 4.00Gb RAM.

The computation time of the structured random sampling method is slightly larger than the computation time of random sampling. This extra time is due to swapping players to their preferred positions in the orderings in the structured random sampling procedure. Overall, this time difference is, however, negligible. Average computation times are about 500s to 900s with the number of orderings ranging from 2500 to 5000. These times are about 300s to 850s with the number of players ranging from 50 to 100.

5 Application: Al Qaeda's network of the WTC 9/11 attack

In this section the structured random sampling method is applied to a real-world case, i.e., the WTC 9/11 terrorist attack of Al Qaeda. On September 11th, 2001, Al Qaeda hijacked four planes in the United States of America. Two of the planes flew into the Twin Towers of the World Trade Center of New York. A third plane flew into the Pentagon and a fourth plane crashed in Pennsylvania. The attack resulted in the immediate death of 2977 people and left many more injured.

Based on publicly available resources Krebs (2002) mapped Al Qaeda's terrorist network responsible for the WTC attack. This includes a network of 19 pilots and hijackers, which were directly responsible for the 9/11 attack, and an extended network of 69 members that includes accomplices, see Figure 3. In this figure the size of each node is proportional to the number of connections of each member. Lindelauf et al. (2013) computed the Shapley value of the connectivity game based on the network of the 19 pilots and hijackers. Computing the Shapley value for the extended network of 69 members can however not be accomplished using their

method due to the extremely large number of $69! \approx 1.7 \cdot 10^{98}$ orderings to be considered in the computations.
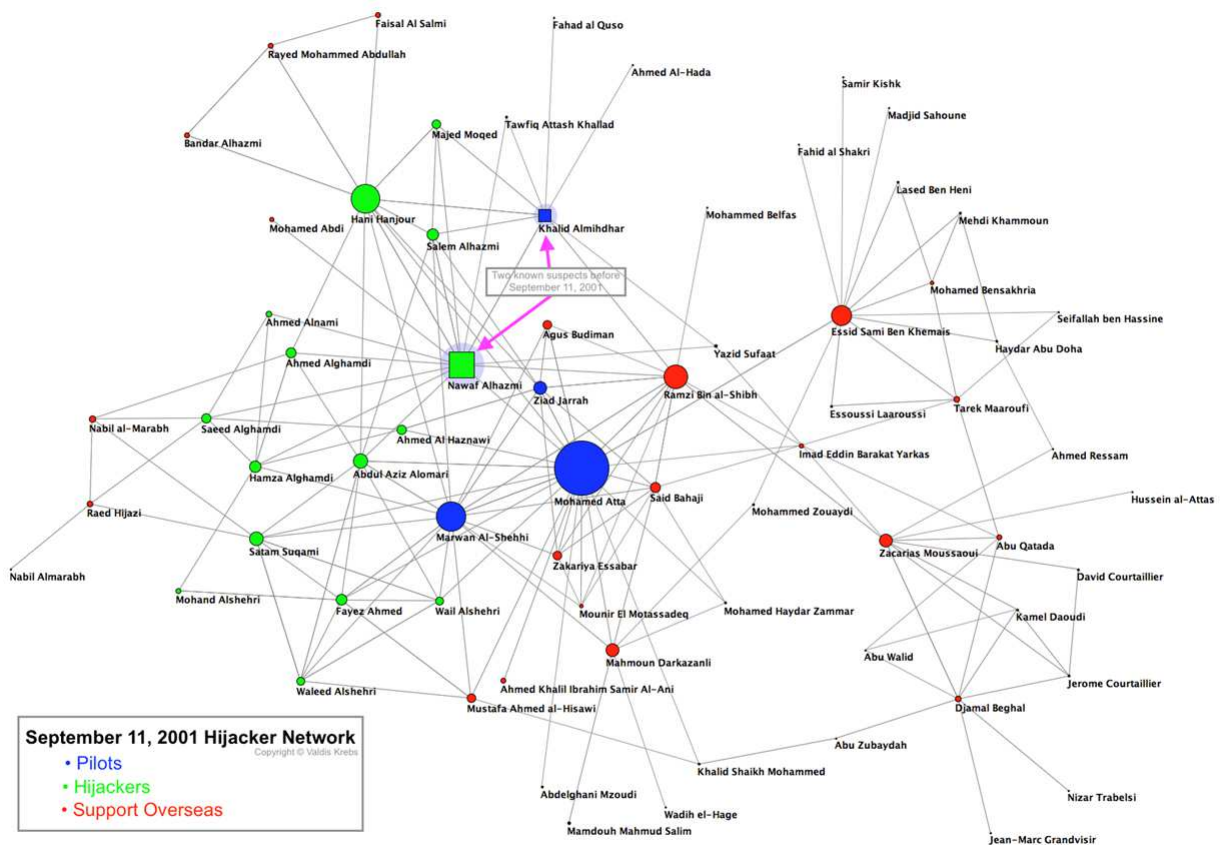

Figure 3: The network of pilots, hijackers and accomplices in the WTC 9/11 attack. The size of each node is proportional to the number of connections of each member. *Source: Krebs (2002). Used with permission.*

Fortunately, our structured random sampling method is able to approximate the Shapley value for each of the 69 members in the network. These approximated Shapley values can be used to construct a ranking of the members in the network. In this ranking the members with the highest Shapley values attain the top positions in the ranking. Table 8 depicts the first 15 members in the ranking based on the approximated Shapley value.

In the connectivity game for this network the value of the grand coalition, i.e., $v(N) = 1$, is to be divided over all 69 members. The Shapley values of the top-10 members add up to 1.05, with individual Shapley values ranging between 0.0949 and 0.1137, see Table 8. The Shapley values of the remaining 59 members in the network are at least a factor ten smaller, see, for example, Hamza Alghamdi at the 11-th position. More than half of the members even obtain a negative Shapley value. Hence, the Shapley value identifies the first ten members in the ranking in Table 8 as the key players in the WTC network.

13

Ranking	Name	Appr. Shapley value
1	Mohamed Atta	0.1137
2	Essid Sami Ben Khemais	0.1111
3	Hani Hanjour	0.1107
4	Djamal Beghal	0.1070
5	Khalid Almihdhar	0.1069
6	Mahmoun Darkazanli	0.1067
7	Zacarias Moussaoui	0.1009
8	Nawaf Alhazmi	0.0995
9	Ramzi Bin al-Shibh	0.0985
10	Raed Hijazi	0.0949
11	Hamza Alghamdi	0.0090
12	Fayez Ahmed	0.0088
13	Marwan Al-Shehhi	0.0046
14	Satam Suqami	0.0038
15	Saeed Alghamdi	0.0037

Table 8: First 15 members in WTC network of Figure 3 according to the approximated Shapley value.

The ranking in Table 8 was obtained using a sample of only 10 000 orderings (with a computation time of 5 minutes). Increasing the number of orderings up to 500 000 orderings (taking up 4 hours of computation time) had no effect on the members present in the top-10. Note that even in the case of 500 000 orderings still only a tiny fraction of the total number of 69! orderings is sampled. Hence, only a tiny fraction of all possible orderings is needed to obtain consistent results.

The top-10 contains members like Mohamed Atta (pilot of flight 11, WTC North), Hani Hanjour (pilot of flight 77, Pentagon), Khalid Almihdhar and Nawaf Alhazmi (both hijackers of flight 77). Zacarias Moussaoui, who was arrested before the WTC attacks took place and became known as 'the 20-th hijacker', is also present in the top-10. Besides these pilots and hijackers, members like Essid Sami Ben Khemais, a former head of operations for Al-Qaeda in Italy, and Djamal Beghal can be found in the top-10 of the ranking. Both have been identified by US government of plotting attacks on US embassies world wide. Figure 3 shows their connections to the members involved in the WTC 9/11 attack.

In this application we have only considered the structure of the WTC network. A big advantage of game theoretic measures is that they, in contrast to standard centrality measures, can also be applied when additional information on nodes and relationships is present (see Lindelauf et al. (2013)). In this application we could for example model the network as a weighted connectivity game. Our structured random sampling method could then approximate the Shapley value of each member resulting in a ranking of the members in the WTC network taking into account both compositional and structural variables. Comparing this ranking to rankings obtained by standard centrality measures (that only consider network structure) may lead to valuable insights in dismantling such networks.

6 Conclusions

The Shapley value is one of the most prominent one-point solution concepts in cooperative game theory. In general however its computational complexity is exponential in the number of players. There are numerous real world applications in which the number of players in the game is too large to calculate the Shapley value, warranting methodology to approximate the Shapley value. Next to the already existing heuristic of random sampling we presented a structured random sampling approach to approximate the Shapley value and compared it to the random sampling procedure. Our simulations showed that the errors in the approximations using structured random sampling are substantially smaller than the errors using random sampling, becoming more prominent when the number of orderings is small or the number of players increases. Average computation times of the structured random sampling method were only slightly larger than the random sampling method, making structured random sampling a preferred choice when approximating the Shapley value. We applied our structured random sampling method on a connectivity game for Al Qaeda's extended network of the WTC 9/11 attack to identify the key players in this network, which previously was impossible due to the size of the WTC network.

Acknowledgements

References

Bachrach, Y., E. Markakis, E. Resnick, A. D. Procaccia, J. S. Rosenschein, and A. Saberi (2010, March). Approximating Power Indices: Theoretical and Empirical Analysis. *Autonomous Agents and Multi-Agent Systems 20*(2), 105–122.

Castro, J., D. Gómez, and J. Tejada (2009, May). Polynomial Calculation of the Shapley Value Based on Sampling. *Comput. Oper. Res. 36*(5), 1726–1730.

Curiel, Imma, Giorgio Pederzoli, and Stef Tijs (1989). Sequencing games. *European Journal of Operational Research 40*(3), 344 – 351.

Deidda, D., J. Andreu, M. A. Perez, G. M. Sechi, R. Zucca, and P. Zuddas (2009). A cooperative game theory approach to water pricing in a complex water resource system. In Proceedings of the 18th World IMACS/MODSIM Congress, Cairnes, Australia.

Deng, Xiaotie and Christos H. Papadimitriou (1994, May). On the complexity of cooperative solution concepts. *Math. Oper. Res.*, **19**, 257–266.

Faigle, U. and W. Kern (1992). The Shapley value for cooperative games under precedence constraints. *International Journal of Game Theory 21*(3), 249–266.

Fatima, Shaheen S., Michael Wooldridge, and Nicholas R. Jennings (2008). A linear approximation method for the Shapley value. *Artificial Intelligence 172*(14), 1673 – 1699.

Krebs, V. E. (2002). Mapping networks of terrorist cells. *Connections 24*(3), 43–52.

Leech, Dennis (2003). Computing Power Indices for Large Voting Games. *Management Science 49*(6), 831–837.

Liben-Nowell, D., A. Sharp, T. Wexler, and K. Woods (2012). Computing Shapley Value in Supermodular Coalitional Games. In Joachim Gudmundsson, Julin Mestre, and Taso Viglas (Eds.), *Computing and Combinatorics*, Volume 7434 of *Lecture Notes in Computer Science*, pp. 568–579. Springer Berlin Heidelberg.

Lindelauf, R., H. Hamers, and B. Husslage (2013). Cooperative game theoretic centrality analysis of terrorist networks: The cases of Jemaah Islamiyah and Al Qaeda. *European Journal of Operational Research 229*(1), 230–238.

Littlechild, S. C. and G. Owen (1973). A Simple Expression for the Shapley Value in a Special Case. *Management Science 20*(3), 370–372.

Mann, Irwin and Lloyd S. Shapley (1960). Values of Large Games, IV: Evaluating the Electoral College by Montecarlo Techniques. Technical report, RAND Corporation.

Michalak, T., T. Rahwan, P.L. Szczepanski, O. Skibski, R. Narayanam, M. Wooldridge, and N.R. Jennings (2013). Computational Analysis of Connectivity Games with Applications to the Investigation of Terrorist Networks. In *Proc. 23rd International Joint Conference on AI (IJCAI)*, pp. 293–301. AAAI Press / International Joint Conferences on Artificial Intelligence.

Moretti, S. and F. Patrone (2008). Transversality of the Shapley value. *TOP 16*(1), 1–41.

Moretti, S., F. Patrone, and S. Bonassi (2007). The class of microarray games and the relevance index for genes. *TOP 15*(2), 256–280.

Narayanam, R. and Y. Narahari (2008). Determining the top-k nodes in social networks using the Shapley value. In Lin Padgham, David C. Parkes, Jrg P. Mller, and Simon Parsons (Eds.), *AAMAS (3)*, pp. 1509–1512. IFAAMAS.

Narayanam, R. and Y. Narahari (2011, Jan). A Shapley Value-Based Approach to Discover Influential Nodes in Social Networks. *Automation Science and Engineering, IEEE Transactions on 8*(1), 130–147.

Owen, G. (1972). Multilinear Extensions of Games. *Management Science 18*(5), pp. P64–P79.

Roth, Alvin E. (Eds) (1988). *The Shapley Value: Essays in Honor of Lloyd S. Shapley*. Cambridge University Press.

Shapley, L. S. (1953). A value for n-person games. Contribution to the Theory of Games. *Annals of Mathematics Studies*, **2**, 28.

Shapley, L. S. and M. Shubik (1954, 9). A Method for Evaluating the Distribution of Power in a Committee System. *American Political Science Review*, **48**, 787–792.

Young, H.P. (1985). Monotonic solutions of cooperative games. *International Journal of Game Theory 14*(2), 65–72.

Zlotkin, Gilad and Jeffrey S. Rosenschein (1994). Coalition, Cryptography, and Stability: Mechanisms for Coalition Formation in Task Oriented Domains. In *Proceedings of the Twelfth National Conference on Artificial Intelligence (Vol. 1)*, AAAI '94, Menlo Park, CA, USA, pp. 432–437. American Association for Artificial Intelligence.