Replication Robust Payoff Allocation in Submodular Cooperative Games

Dongge Han, Michael Wooldridge, Alex Rogers, Olga Ohrimenko, Sebastian Tschiatschek

***Abstract*—Submodular functions have been a powerful mathematical model for a wide range of real-world applications. Recently, submodular functions are becoming increasingly important in machine learning (ML) for modelling notions such as information and redundancy among entities such as data and features. Among these applications, a key question is payoff allocation, i.e., how to evaluate the importance of each entity towards a collective objective? To this end, classic solution concepts from cooperative game theory offer principled approaches to payoff allocation. However, despite the extensive body of game-theoretic literature, payoff allocation in submodular games is relatively under-researched. In particular, an important notion that arises in the emerging submodular applications is redundancy, which may occur from various sources such as abundant data or malicious manipulations where a player replicates its resource and acts under multiple identities. Though many game-theoretic solution concepts can be directly used in submodular games, naively applying them for payoff allocation in these settings may incur robustness issues against replication. In this paper, we systematically study the replication manipulation in submodular games and investigate *replication robustness*, a metric that quantitatively measures the robustness of solution concepts against replication. Using this metric, we present conditions which theoretically characterise the robustness of semivalues, a wide family of solution concepts including the Shapley and Banzhaf value. Moreover, we empirically validate our theoretical results on an emerging submodular ML application—ML data markets.**

***Impact Statement*—With the increasing take-up of ML techniques in real-world settings, payoff allocation has significant impact towards fairness, trustworthiness, safety, and knowledge discovery in ML applications, e.g., performing analysis or debugging of ML systems by finding the key contributors or bottleneck entities. Many emerging ML applications exhibit submodular characteristics, while properties of classic game-theoretic payoff allocation on submodular games are under-researched. This paper investigats an important issue of redundancy arising from replication in submodular ML applications. Using a replication robustness metric, we provide theoretical guarantees for the robustness of common game-theoretic payoff allocation methods against replication. Our findings can guide the use of game-theoretic payoff allocation in submodular ML applications, and impact real-world applications and future research on payoff allocation in ML systems in general, such as fair compensation in multi-party ML systems and feature importance interpretation in the medical domains.**

Manuscript received Jan 15, 2022.

Dongge Han, Michael Wooldridge and Alex Rogers are with Department of Computer Science, the University of Oxford, OX1 3QD, Oxford, U.K. (e-mail: dongge.han.oxford@gmail.com, michael.wooldridge@cs.ox.ac.uk, alex.rogers@cs.ox.ac.uk).

Olga Ohrimenko is with School of Computing and Information Systems, the University of Melbourne, Victoria 3010, Australia (e-mail: oohrimenko@unimelb.edu.au).

Sebastian Tschiatschek is with the University of Vienna, Faculty of Computer Science, Währinger Straße 29, 1090 Vienna, Austria (e-mail: sebastian.tschiatschek@univie.ac.at).

***Index Terms*—Cooperative Game Theory, Submodularity, Semivalue, Shapley value, Banzhaf value**

I. Introduction

Submodularity has long been an important topic in mathematics, operations research, economics and optimisation. Submodular functions [35] exhibit the natural property of *diminishing returns*. Informally, given a ground set of elements (e.g., physical entities such as sensors, goods, or digital entities such as data, features), the marginal contribution of a single element when added to a set of elements diminishes with the increasing size of the set. This property frequently occurs in real-world settings, making submodular functions a powerful mathematical model for a wide range of applications, such as cooperative cost allocations [12], sensor placement [20] and facility location problems (FLP) [7]. Recently in the field of machine learning (ML), submodular functions are becoming increasingly important as they naturally model notions of *information, diversity and redundancy* [3]. In these classic and emerging applications, a key question is *how to evaluate the importance of each entity towards the collective objective, i.e., payoff allocation?* On the one hand, in cooperative settings, importance evaluations can enable fair allocation of the collective reward towards each member. On the other hand, evaluating the importance of each entity helps to identify crucial insights into the system such as the key contributors or redundant entities. An example use case is ML model interpretation [33, 26] – typically a trained blackbox ML model cannot be interpreted by humans. To interpret the model and ensure it is trustworthy, we can evaluate the importance it gives to each input feature when making a prediction.

A principled approach to payoff allocation is provided by cooperative game theory [4], which models the entities as players and their interactions (typically) in the form of a characteristic function game $G = (N, v)$, where a characteristic function v evaluates each possible set of players. Under this formulation, the most popular game-theoretic solution concept is the Shapley value [36], which allocates the payoff to each player as a weighted average of its' marginal contributions towards all possible sets of other players, and has been widely applied in network centrality [1], ML interpretation [26], data valuation [16, 2], etc. Despite the extensive body of game-theoretic literature, submodular games (i.e., games with submodular characteristic functions) are relatively under-explored, a setting where players may not be incentivised to cooperate and form a grand coalition. Nevertheless, with the ever-grown interest in ML applications, the above setting becomes

increasingly common than ever and leads to an urgent need to study payoff allocation in submodular games. In fact, many problems in ML are submodular by nature, and players form a grand coalition inherently (e.g., among passive entities such as data and features) or according to rules which require the cooperation among players. For example, consider multiple hospitals collaboratively training an ML model by pooling their medical records: if accuracy of the trained model is the central objective, the hospitals will agree to cooperate and form the grand coalition in order to train a better prediction model, even though a player may be less useful in terms of marginal contributions with increasing data.

Closely related to the submodular games is the notion of *redundancy* [3]. On the one hand, redundancy may come from a benign source, e.g., abundant data typically carry partially redundant information and yields diminishing returns. This motivates important problems such as data selection [41, 18], feature selection [8] and data summarisation [25]. On the other hand, redundancy may arise as a result of malicious manipulations, e.g., a replication manipulation, where a malicious player may replicate its resource and act under multiple false identities. In both the malicious and benign cases, redundancy often does not bring significant additional value to the collective objective, but may have substantial impact on the payoff allocation. Though many common game-theoretic solution concepts can be directly applied in the emerging submodular ML applications, there are no theoretical guarantees for these solution concepts regarding redundancy. Consequently, naively applying them for payoff allocation in these settings may incur robustness issues and thereby have the adverse effect of incentivizing the aforementioned replication manipulation.

In this paper, we systematically study the replication manipulation in submodular games and investigate *replication robustness*, a metric which quantitatively measures the robustness of solution concepts against replication manipulations. Using this metric, we present conditions which theoretically characterise the robustness of semivalues [10], a wide family of Shapley-like solution concepts including the Shapley value and the Banzhaf value [22]. Though we model the redundancy from the perspective of malicious manipulations, the theoretical framework can also be extended to study redundancy that occur under the benign cases, for example, for promoting diversity among features in ML feature subset selections, or encourage diverse behaviours among robotic agents in multiagent reinforcement learning.

The outline of our paper is as follows: In Section III we first define submodular games, the replication manipulation and replication robustness. To illustrate the effect of redundancy, we look at a classic submodular problem – the facility location problem. In Section IV, we compare the replication robustness of the Shapley value and the Banzhaf value when a malicious player replicates its resource and acts as two identities. In Section V, we extend our theoretical results to general semivalues and an arbitrary number of replications, and we present a necessary and sufficient condition which characterises the replication robustness of general semivalues. Finally in Section VI, we apply our theoretical results to an emerging ML application – the ML data market [30, 2], and empirically validate our theoretical results of replication robustness across various solution concepts.

II. Background

In this section, we introduce our notation and concepts from cooperative game theory [4].

Cooperative Games. Formally, a cooperative game with transferable utility (hereafter simply a *cooperative game*) is given by a tuple $G = (N, v)$, where $N = \{1, \dots, n\}$ is the set of players of the game and $v\colon 2^N \to \mathbb{R}$ is a *characteristic function*, which assigns a real value $v(\mathcal{C})$ to every subset of players $\mathcal{C} \subseteq N$, referred to as *coalitions*. The *grand coalition* is the set N of all players. For clarity, we will introduce the general definition of semivalues [10] in Section V-A. Before this, we introduce here the concept of marginal contribution and some common semivalues. Intuitively, the *marginal contribution* of a player to coalition $\mathcal{C}$ is the difference that this player makes towards $\mathcal{C}$ before and after joining it, i.e., $MC_i(\mathcal{C}) \coloneqq v(\mathcal{C} \cup \{i\}) - v(\mathcal{C})$.

Solution Concepts. A solution concept [4] describes the outcome of a cooperative game, i.e., the partition of players into coalitions, and a payoff function which assigns a payoff $\varphi_i(N, v) \in \mathbb{R}$ to each player i. As discussed in the introduction, we focus on payoff allocations in the emerging ML settings where the players form the grand coalition inherently. Therefore, we will refer to the solution concepts as the payoff allocation with respect to the grand coalition.

The following is a collection of properties which are commonly used to axiomatize solution concepts [4].

(A1) *Symmetry*: Two players i and j who have the same marginal contribution in any coalition have the same payoff, i.e., $(\forall \mathcal{C} \subseteq N \setminus \{i, j\}\colon v(\mathcal{C} \cup \{i\}) = v(\mathcal{C} \cup \{j\})) \to \varphi_i(N, v) = \varphi_j(N, v)$.

(A2) *Efficiency*: The payoff values of all players sum to $v(N)$, i.e., $v(N) = \sum_{i \in N} \varphi_i(N, v)$.

(A3) *Null-player*: a player whose marginal contribution is zero in any coalition has zero payoff, i.e., $(\forall \mathcal{C} \subseteq N\colon v(\mathcal{C} \cup \{i\}) = v(\mathcal{C})) \to \varphi_i(N, v) = 0$.

(A4) *Linearity*: Given two cooperative games $G^1 = (N, v^1)$ and $G^2 = (N, v^2)$, then for any player $i \in N$, $\varphi_i(N, v^1 + v^2) = \varphi_i(N, v^1) + \varphi_i(N, v^2)$.

(A5) *2-Efficiency* [22]: $\varphi_i(N, v) + \varphi_j(N, v) = \varphi_{p_{ij}}(N', v')$ characterises neutrality of collusion, where $\varphi_{p_{ij}}(N', v')$ is player p_{ij}'s payoff in a game in which players i and j are merged into a single player p_{ij}, i.e., $N' = N \setminus \{i, j\} \cup \{p_{ij}\}$.

Next, we review three common semivalues.

- The *Shapley Value* [36] is a common solution concept, defined as the weighted average marginal contributions of a player towards coalitions of other players, and the unique value that satisfies **(A1)-(A4)**:

$$\varphi_i^{\text{Shapley}} = \sum_{\mathcal{C} \subseteq N \setminus \{i\}} \frac{|\mathcal{C}|!(|N| - |\mathcal{C}| - 1)!}{|N|!} MC_i(\mathcal{C})$$

- The *Banzhaf Value* [22] is commonly used as a measure for voting power, which is defined by the average marginal

contribution of a player towards all coalitions of other players, uniquely characterized by axioms **(A1)**, **(A3)**-**(A5)**:

$$\varphi_i^{\text{Banzhaf}} = \frac{1}{2^{|N|-1}} \sum_{\mathcal{C} \subseteq N \setminus \{i\}} MC_i(\mathcal{C})$$

- *Leave-one-out (LOO)* assigns to each player its marginal contribution towards the coalition of all other players:

$$\varphi_i^{\text{LOO}} = MC_i(N \setminus \{i\}).$$

III. Submodular Games and Replication

We now introduce submodular functions and how they can be used as characteristic functions in cooperative games. For illustration, we show an example class of submodular games defined by a classic submodular function – the facility location function. We will also use this example to validate our theoretical findings in Section V-H. Following the definition of submodular games, we will show how replication manipulations can be performed, and define criteria for evaluating the robustness of solution concepts against replication.

A. Submodular Games

The following property lists three equivalent definitions of submodular set functions (aka submodular functions) [35]:

Definition III-A.1 (Submodular Set Functions)**.** *Let N be a finite set. A submodular function is a set function $f\colon 2^N \to \mathbb{R}$, where 2^N denotes the power set of N, which satisfies one of the following equivalent conditions:*

1) *$\forall X, Y \subseteq N$ with $X \subseteq Y$ and $\forall x \in N \setminus Y$, we have $f(X \cup \{x\}) - f(X) \geq f(Y \cup \{x\}) - f(Y)$.*
2) *$\forall S, T \subseteq N$, we have $f(S)+f(T) \geq f(S \cup T)+f(S \cap T)$.*
3) *$\forall X \subseteq N$ and $x_1, x_2 \in N \setminus X$ such that $x_1 \neq x_2$, we have $f(X \cup \{x_1\})+f(X \cup \{x_2\}) \geq f(X \cup \{x_1, x_2\})+f(X)$.*

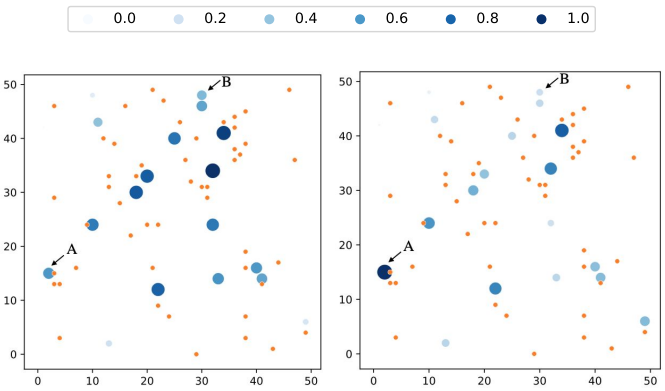
The first one of the equivalent conditions demonstrates diminishing returns, i.e., the marginal value of an entity towards a set decreases as the set grows. Due to its natural relation to the marginal contributions of cooperative games, we next define submodular games using the first condition.

Definition III-A.2 (Submodular Game)**.** *A characteristic function game $G = (N, v)$ with a finite non-empty set of players $N = \{1, \ldots, n\}$, is a* submodular game *if the characteristic function v is submodular, i.e, $\forall \mathcal{C} \subseteq \mathcal{C}' \subseteq N \setminus \{i\}\colon v(\mathcal{C} \cup \{i\}) - v(\mathcal{C}) \geq v(\mathcal{C}' \cup \{i\}) - v(\mathcal{C}')$.*

Recall that the difference in value made by a player i by joining a coalition $\mathcal{C}$ is denoted as the *marginal contribution* of player i towards coalition $\mathcal{C}$, i.e., $MC_i(\mathcal{C}) \coloneqq v(\mathcal{C} \cup \{i\}) - v(\mathcal{C})$. Therefore in a submodular game, the marginal contribution of a player towards a coalition $\mathcal{C}$ is no less than its contribution towards a superset $\mathcal{C}'$, as summarised in the next assumption.

Assumption 1. *In a submodular game $G = (N, v)$, the marginal contributions of each player $i \in N$ satisfy*

$$\forall \mathcal{C} \subseteq \mathcal{C}' \subseteq N \setminus \{i\}\colon MC_i(\mathcal{C}) \geq MC_i(\mathcal{C}'). \tag{1}$$



(a) The Shapley Value (b) The Banzhaf value

Fig. 1: The Shapley value and Banzhaf value for the Facility Location Game. The figures show a 50x50 map, where the orange dots represent 50 customers. Blue dots refer to 20 facility locations, with larger and darker dots representing larger Shapley/Banzhaf values, which are normalised between [0,1] for a clear comparison.

B. Motivating Example: Facility Location Problem

A classic example in submodular optimisation is the facility location problem (FLP). As an important topic in operations research, an FLP considers the question of how to select a cost-effective subset from a ground set of potential locations for placing new facilities [3, 34, 7, 11]. Here, the facilities can refer to hospitals, plants, docking stations, etc. There exist several different formulations of the FLP [29, 7], and we will adopt the formulation following Nemhauser et al. [29], which consists a set of potential facility sites $\mathcal{L}$ where new facilities can be opened, a set of customers D to be serviced, and a matrix U which represents utilities of each customer from each facility location (e.g., proximity). The FLP is uncapacitated, i.e, it is always optimal to satisfy the demand of a customer from the open facility which provides them with the highest utility. By modelling the FLP as a submodular game, we can evaluate the importance of each facility location by computing their payoff allocations using the common solution concepts. To do this, we can consider the players as the set of facility locations $\mathcal{L}$, and the characteristic function as the facility location function $Fac(\mathcal{C}) = \sum_{d \in D} \max_{i \in \mathcal{C}} u_{id}$, i.e., the value of each coalition $\mathcal{C} \subseteq \mathcal{L}$ is the sum of utilities of all customers from the open facilities $i \in \mathcal{C}$.

Example 1 (Facility Location Game)**.** *Let D be a set of customers and $\mathcal{L}$ a set of facility locations. Define a utility function $u\colon \mathcal{L} \times D \to \mathbb{R}_+$, represented by a matrix $U \in \mathbb{R}_+^{|\mathcal{L}| \times d}$, where each entry $u_{id} \in U$ is the utility of customer d for facility location i. A facility location game is defined as $G = (\mathcal{L}, v)$, where the players $\mathcal{L}$ are facility locations and the characteristic function is the facility location function, i.e., $\forall \mathcal{C} \subseteq \mathcal{L}, v(\mathcal{C}) = Fac(\mathcal{C}) = \sum_{d \in D} \max_{i \in \mathcal{C}} u_{id}$.*

Fig. 1 illustrates the Shapley and Banzhaf value on an example facility location game with $|\mathcal{L}| = 20$ facility locations (blue), and $|D| = 50$ customers (orange) randomly placed in a 50×50 map. The utility of customer d from facility

i, $u_{id} = 100 - (|x_i - x_d| + |y_i - y_d|)$, decreases with the Manhattan distance to the facility. In comparison, the locations with higher Shapley value typically have a larger number of nearby customers, while locations with higher Banzhaf values often have a larger number of nearby customers and fewer nearby facility locations. In the figure, for example, A is distant from nearby facilities, and ranks higher in terms of the Banzhaf value than the Shapley value, and conversely for B, which has multiple nearby facilities. This example provides a illustrative comparison between the Shapley value and the Banzhaf value against redundancy. Compared with the Shapley value, the Banzhaf value assigns higher weights to mid-sized coalitions than the Shapley value, hence emphasizes a player's complementary value towards other players. In comparison, the Shapley value puts higher weights on smaller coalitions, hence emphasizes a player's individual value. Therefore, in the FLP example, a player's (i.e., a facility) Banzhaf value is able to better detect redundant players (nearby facilities) compared with the player's Shapley value, making the Banzhaf value more robust to redundancy. In the rest of the paper, we will go beyond the FLP example to further investigate the cause of their distinct behaviours and formally show the intuition in general submodular cooperative games.

C. Replication Manipulation

As illustrated in the facility game in Fig. 1, an important notion that commonly arises in submodular settings is *redundancy*, which may occur naturally from abundant resources or from malicious manipulations such as replication. For example, a standard submodular ML problem is data summarisation [3, 25], which aims to find a concise subset to best represent the ground set of data.

In the following definition, we introduce the replication manipulation, where a malicious player replicates its resource (e.g., digital entities such as online identities, data, features) and acts under multiple false identities. Here we model redundancy from the point of view of malicious manipulations, nevertheless, the theoretical framework can also be extended to study redundancy that occur under the benign cases.

Definition III-C.1 (Replication Manipulation)**.** *In a submodular game $G = (N, v)$, a (malicious) player i executes a replication action k times on its resources D_i and acts as $k+1$ players $\mathcal{C}^R = \{i_0, i_1, \ldots, i_k\}$ each holding one replica of D_i. Denote the induced game as $G^R = (N^R, v^R)$, where the induced set of players are $N^R = N \setminus \{i\} \cup \mathcal{C}^R$, and the induced characteristic function v^R satisfies $\forall \mathcal{C} \subseteq N \setminus \{i\}, \forall i_k \in \mathcal{C}^R$: $v^R(\{i_k\} \cup \mathcal{C}) = v(\{i\} \cup \mathcal{C})$ and $v^R(\mathcal{C}) = v(\mathcal{C})$. By replicating, player i receives a* total payoff *which is the sum of the payoff of all its $k+1$ replicas, i.e., $\varphi_i^{\text{tot}}(k) = \sum_{\kappa=0}^{k} \varphi_{i_\kappa}(N^R, v^R)$.*

The next assumption captures the fact that adding redundant resources to a coalition typically does not change the value of the coalition (e.g., redundant feature or replicated data). We refer to this property as *replication redundancy* and formalize it in the following assumption:

Assumption 2 (Replication Redundancy)**.** *A replica does not contribute additional value to coalitions which already contain another replica or the original resource:*

$$\forall i, j \in \mathcal{C}^R \colon (i \in \mathcal{C}) \rightarrow MC_j(\mathcal{C}) = 0.$$

Despite the fact that redundant resources do not bring significant additional value to the collective objective, it may have substantial impact on the payoff allocation. For example, a malicious player may be able to gain a higher total payoff by performing the replication manipulation described in Definition III-C.1. The next definition formalizes the notion of replication robustness of solution concepts, i.e., a property that ensures that a player through replication gains a total payoff no more than its original payoff.

Definition III-C.2 (Replication Robustness)**.** *A solution concept φ is* replication robust *if the payoff of the replicating player i in the original game G is no less than the total payoff of the player's replicas $\mathcal{C}^R$ in the induced game G^R after replication, i.e.,*

$$\varphi_i(N, v) \geq \sum_{i_\kappa \in \mathcal{C}^R} \varphi_{i_\kappa}(N^R, v^R).$$

To illustrate the condition, consider a malicious player who aims to increase its payoff by performing replication manipulation. A solution concept that is replication robust could then be used to counteract such malicious behaviours as the replication manipulation does not result in an increased payoff for the malicious player. For an example in the benign case, consider the problem of feature importance interpretation: adding to a set of features $\mathcal{C}$ a feature f' that is redundant to feature $f \in \mathcal{C}$ in the set can be considered as a replication manipulation, and a replication robust solution concept will allocate the two redundant features a total value no greater than the value of the feature f on its own.

Having defined the replication manipulation and robustness criteria, we next study the behaviours of the semivalues under replication and their robustness properties.

IV. Replication Robustness of Common Semivalues with $k = 1$ Replications

To start with, we take a look at the two most common semivalues, the Shapley value and the Banzhaf value, and study how the total payoff of the malicious player changes when the player replicates its resources and splits into two identities.

A. Robustness of the Shapley Value

The following theorem shows that the Shapley value is not replication robust in submodular games. Specifically, under payoff allocation according to the Shapley value, the malicious player can always obtain a non-negative gain in total payoff by replicating its resource and splitting into two identities.

Theorem IV-A.1. *Let G= (N, v) be a submodular game with replication redundant characteristic function v. Assume that player $i \in N$ replicates and obtains the total payoff of the*

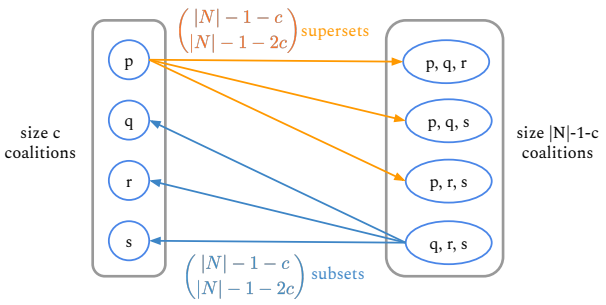


Fig. 2: Illustration of the proof for Theorem IV-A.1. Given an example game with 5 players $N = \{i, p, q, r, s\}$, we match the coalitions (excluding the target player i) of size-c and size-$(|N|-1-c)$. (here $c = 1$ and $|N|-1-c = 3$). Specifically, each size-c coalition in L (left) has $\binom{|N|-c-1}{|N|-2c-1}$ supersets in R (right), each size-$|N|-c-1$ coalition in (R) has $\binom{|N|-c-1}{|N|-2c-1}$ subsets of size-c. Arrows indicate the set inclusion relations.

two identities $\mathcal{C}^R = \{i_1, i_2\}$ in the new game $G^R = (N^R, v^R)$. The change in total payoff of player i because of relication is:

$$\delta\varphi_i^{\text{Shapley}} = \sum_{\mathcal{C}\subseteq N\setminus\{i\}} \frac{|\mathcal{C}|!(|N|-|\mathcal{C}|-1)!}{(|N|+1)!}(|N|-2|\mathcal{C}|-1)MC_i(\mathcal{C}).$$

Moreover, the total payoff of player i after replication is no less than its payoff in the original game, i.e, $\delta\varphi_i^{\text{Shapley}} \geq 0$.

Proof: The derivation of the change in total payoff is provided in Appendix A. Here we focus on showing that the change in total payoff is non-negative, i.e., $\delta\varphi_i^{\text{Shapley}} \geq 0$. To prove this, we make use of the submodularity property, which compares the marginal contributions of player i towards pairs of coalitions of other players $\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq N \setminus \{i\}$. To pair the coalitions, we make two observations on $\delta\varphi_i^{\text{Shapley}}$: Given two coalitions $\mathcal{C}_1, \mathcal{C}_2 \subseteq N \setminus \{i\}$ with complementary sizes, i.e., $|\mathcal{C}_1| + |\mathcal{C}_2| = |N \setminus \{i\}| = |N| - 1$, (1) their weights in $\delta\varphi_i^{\text{Shapley}}$ are opposite and add up to zero, (2) there are equal number of size c and $|N|-1-c$ coalitions, i.e., $\binom{|N|-1}{c} = \binom{|N|-1}{|N|-1-c}$. These observations suggest that we may find a bijective mapping between the size c coalitions and their size $|N|-1-c$ supersets. Formally, for any coalition size $c < (|N|-1)/2$, we look for a bijective mapping f between coalitions with inclusion relations and of complementary sizes, that is, $f : \{\mathcal{C}_1 \subseteq N \setminus \{i\} \mid |\mathcal{C}_1| = c\} \mapsto \{\mathcal{C}_2 \subseteq N \setminus \{i\} \mid |\mathcal{C}_2| = |N|-1-c\}$ such that $\mathcal{C}_1 \subseteq f(\mathcal{C}_1)$. The corner case where $c = (|N|-1)/2$ can be omitted as they have zero weight in $\delta\varphi_i^{\text{Shapley}}$, i.e., $\frac{c!(|N|-c-1)!}{(|N|+1)!}(|N|-2c-1) = 0$. To show the existence of the bijective mapping, we model the coalitions and their inclusion relations ($\subseteq$ and $\supseteq$) by a bipartite graph (An example is shown in Figure 2). For any coalition size $c < (|N|-1)/2$, define bipartite graph $B_c = (L, R, E)$ where each vertex corresponds to a coalition, i.e., vertices $L = \{\mathcal{C}_1 \subseteq N \setminus \{i\} \mid |\mathcal{C}_1| = c\}$ are the size c coalitions, and vertices $R = \{\mathcal{C}_2 \subseteq N \setminus \{i\} \mid |\mathcal{C}_2| = |N|-1-c\}$ are the size $|N|-1-c$ coalitions, and $|L| = |R|$ from observation (2). Denote edges E as the set inclusion relations, that is, $E = \{\{\mathcal{C}_1, \mathcal{C}_2\} \mid \mathcal{C}_1 \in L, \mathcal{C}_2 \in R, \mathcal{C}_1 \subseteq \mathcal{C}_2\}$. The graph is k-regular where every vertex has the same degree $k = \binom{|N|-1-c}{|N|-1-2c}$. To see this, we first show that each coalition $\mathcal{C}_1 \in L$ has $\binom{|N|-1-c}{|N|-1-2c}$ supersets in R. To find a size $|N|-1-c$ coalition $\mathcal{C}_2 \in R$ that is a superset of $\mathcal{C}_1$, we can add $|N|-1-2c$ players by choosing from the remaining $|N|-1-c$ players, i.e., $N \setminus \{i\} \setminus \{\mathcal{C}_1\}$. Therefore, there are $\binom{|N|-1-c}{|N|-1-2c}$ choices and hence the same number of supersets. Similarly, we can show that each coalition $\mathcal{C}_2 \in R$ has $\binom{|N|-1-c}{|N|-1-2c}$ subsets in L, by removing $|N|-1-2c$ members. Having shown that the B_c is k-regular, by Hall's Marriage Theorem for regular graphs, there exists a perfect matching on B_c and hence a bijective mapping f. Finally, we pair the terms according to f:

Let $\mathcal{C}^c = \{\mathcal{C} \subseteq N \setminus \{i\} \mid |\mathcal{C}| = c\}$ denote all size c coalitions excluding player i,

$$\begin{aligned}
\delta\varphi_i^{\text{Shapley}} &= \sum_{\mathcal{C}\subseteq N\setminus\{i\}} \frac{|\mathcal{C}|!(|N|-|\mathcal{C}|-1)!}{(|N|+1)!}(|N|-2|\mathcal{C}|-1)MC_i(\mathcal{C}) \\
&= \sum_{0\leq c<\frac{|N|-1}{2}} \frac{c!(|N|-c-1)!(|N|-2c-1)}{(|N|+1)!}\left(\sum_{\mathcal{C}_1\in\mathcal{C}^c} MC_i(\mathcal{C}_1) - \sum_{\mathcal{C}_2\in\mathcal{C}^{|N|-1-c}} MC_i(\mathcal{C}_2)\right) \\
&= \sum_{0\leq c<\frac{|N|-1}{2}} \frac{c!(|N|-c-1)!(|N|-2c-1)}{(|N|+1)!} \sum_{\mathcal{C}_1\in\mathcal{C}^c} \underbrace{\left(MC_i(\mathcal{C}_1) - MC_i(f(\mathcal{C}_1))\right)}_{\geq 0 \text{ due to submodularity}} \geq 0.
\end{aligned}$$

This concludes our proof that under the Shapley value, a player can only gain a higher total payoff by replication. ■

B. Robustness of the Banzhaf Value

Theorem IV-A.1 shows that the Shapley value is not robust against replication if the player replicates its resource and acts as two players. In what follows we will show that under the same replication manipulation, the Banzhaf value is neutral.

Theorem IV-B.1. *Let G= (N, v) be a submodular game with replication redundant characteristic function v. Assume that player $i \in N$ replicates and obtains the total payoff of the two identities $\mathcal{C}^R = \{i_1, i_2\}$ in the new game $G^R = (N^R, v^R)$. Under payoff allocation using the Banzhaf value, the change in total payoff of player i because of replication is zero, i.e., $\delta\varphi_i^{\text{Banzhaf}} = 0$*

Proof: The neutrality of the Banzhaf value under the replication is a natural consequence of the weights defined on the coalitions. It is also closely related to the 2-efficiency axiom, where the Banzhaf value is neutral to the merging or splitting of two players. The complete proof is provided in Appendix B. ■

In comparison, when the player replicates and acts as two identities, the Shapley value is not replication robust, while the Banzhaf is neutral. This raises a few interesting questions: (1) What governs the robustness of the solution concepts which lead to the different behaviours between the Shapley and Banzhaf values? (2) Can we draw the same conclusion for more than one replications, for example, is the Banzhaf value neutral to an arbitrary number of replications? To answer these questions, we next examine a wider class of solution concepts, i.e., semivalues [10], which include both the Shapley value and Banzhaf value. More importantly, we extend our results to the more general case where the player performs an arbitrary number ($k \geq 1$) of replications.

V. Replication-robustness of General Semivalues with $k \geq 1$ Replications

In many real-world applications, the details of replication are only private to the malicious player due to anonymity. Take the online social networks for an example, the digital identity of a player is typically private and only accessible to the player itself, and a single player can create multiple false identities. Therefore, it is important to account for the case of an arbitrary number of replications where k is unknown. However, with an arbitrary number of replications, the change in total payoff no longer exhibits the structured form which allows for coalition pairing and was exploited in the previous proofs. Therefore, to analyse the robustness of the semivalues for $k \geq 1$ replications, we take the following steps (in the following *V-A* refers to Section V-A, etc.):

V-A. represent semivalues as an importance weighted sum of average marginal contributions across coalition sizes,

V-B. transform the submodularity into an inequality on the average marginal contributions across coalition sizes,

V-C. express the total payoff of the malicious player after replication as (new) importance weighted sum on the (original) average marginal contributions,

V-D. present conditions on the importance weights which lead to replication robustness,

V-E. use the above robustness conditions to evaluate a given semivalue such as the Shapley value.

A. Semivalues as Weighted Average Marginal Contributions

As the first step, we introduce the semivalues [10], a wide class of Shapley-like solution concepts including both the Shapley and Banzhaf value. The semivalue of a player can be defined as a weighted sum over its marginal contributions towards coalitions of other players. The weights of player i's marginal contributions towards coalitions $\mathcal{C}$ are denoted by $w_{\mathcal{C},N}$. In particular, $w_{\mathcal{C},N}$ only depends on the size of the coalition $\mathcal{C}$ but not on the players' identities inside the coalition, i.e.,

$$\varphi_i(N,v) = \sum_{\mathcal{C}\subseteq N\setminus\{i\}} w_{|\mathcal{C}|,N} MC_i(\mathcal{C}). \tag{2}$$

Therefore, by grouping together equal-sized coalitions, a semivalue assigns to each player $i \in N$ a real-valued payoff, expressed as a weighted sum of player i's *average marginal contributions towards size-c coalitions* $z_i(c)$:

$$\varphi_i(N,v) = \sum_{c=0}^{N-1} \alpha_c z_i(c), \quad \text{where} \tag{3}$$

$$z_i(c) = \binom{|N|-1}{c}^{-1} \sum_{\mathcal{C}\subseteq N\setminus\{i\},|\mathcal{C}|=c} MC_i(\mathcal{C})$$

$$\alpha_c = \binom{|N|-1}{c} w_{c,N} \quad \text{(Importance Weights)}$$

Proof Sketch: The derivation from Equation (2) to (3) is straightforward and can be obtained by grouping the marginal contributions of player i towards equal-sized coalitions. The normalisation factor $\binom{|N|-1}{c}$ is the number of size-c coalitions of players excluding i. The proof is in Appendix C. ∎

We will refer to α_c as *importance weights*, as they quantify the importance of a player's marginal contributions towards different coalition sizes. In addition, the importance weights in a semivalue form a probability distribution, that is, $\sum_{c=0}^{|N|-1} \alpha_c = 1$. The next corollary presents the importance weights of some common semivalues, namely, the Shapley value, Banzhaf value, and Leave-one-out value.

Corollary V-A.1 (Importance Weights for Common Semivalues)**.** *The Shapley value is defined by the weights* $w_{c,N} = \frac{c!(|N|-1-c)!}{|N|!} = \frac{1}{|N|}\binom{|N|-1}{c}^{-1}$*, hence the importance weights are uniform across all coalition sizes, i.e.,* $\alpha_c^{\text{Shapley}} = \binom{|N|-1}{c} w_{c,N} = \frac{1}{|N|}$*. In contrast, the Banzhaf value is defined by the weights* $w_{c,|N|} = \frac{1}{2^{|N|-1}}$*, hence the importance weights form a bell shape* $\alpha_c^{\text{Banzhaf}} = \frac{1}{2^{|N|-1}}\binom{|N|-1}{c}$ *which favours mid-sized coalitions. Finally, for the Leave-one-out value,* $\alpha_c^{\text{LOO}} = \mathbb{1}_{c=|N|-1}$.

The wide class of semivalues applies to many different real-world applications, such as voting [28], interpretable machine learning [26, 27], reinforcement learning [24, 14], text summarisation [32], etc. The equality of weights over coalition sizes is not a constraint for the application, but a choice of design that allows the semivalues to take into account the influence of a player on groups of different sizes while preserving anonymity and symmetry. By adjusting the importance weights α_c, a semivalue balances a player's *individual value* and *complementary value*. In particular, putting higher importance on smaller coalitions (larger α_c for smaller c) favours the individual value and vice-versa. So far the representation of semivalues via importance weights has provided some insights for differentiating common solution concepts. In the following sections, we will show that this representation has significant implications for understanding the difference in robustness of solution concepts against replication in submodular games.

B. Average Marginal Contributions vs. Coalition Sizes

Intuitively, in a submodular game with diminishing returns, a player tends to be less useful in terms of marginal contribution when contributing towards a larger coalition. Can we formally show this intuition? Unfortunately, this does not always hold true for arbitrary pairs of coalitions: given coalitions $\mathcal{C}_1$ and $\mathcal{C}_2$ where $|\mathcal{C}_1| \leq |\mathcal{C}_2|$, there is no direct comparison between a player's marginal contributions towards these two coalitions, only except for when $\mathcal{C}_1$ is a subset of $\mathcal{C}_2$. Nevertheless, we can formalise this intuition under *average marginal contributions*. We now present in the following a useful property of submodular games that the average marginal contributions $z_i(c)$ decrease with coalition size under the submodularity assumption.

Lemma V-B.1. *Given a submodular game, the average marginal contribution* $z_i(c)$ *of a player* i *monotonic decreases with coalition size* c*, i.e.,*

$$\forall 0 \leq c < |N|-1, \quad z_i(c) \geq z_i(c+1). \tag{4}$$

Proof: Given player $i \in N$, we show for any coalition size c, $z_i(c) \geq z_i(c+1)$, by taking the following steps:

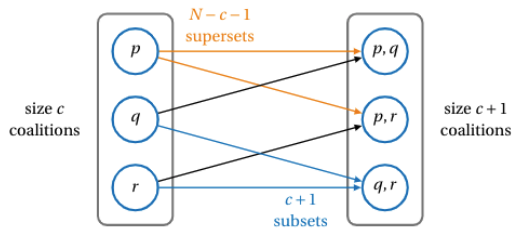


Fig. 3: Illustration of the Proof for Lemma V-B.1: Given an example game with 4 players $N = \{i, p, q, r\}$, we compare the average marginal contribution of player i towards size-c and size-$(c+1)$ coalitions by matching the coalitions. Specifically, each size-c (here $c = 1$) coalition $\mathcal{C}_1$ (Left) has $|N| - c - 1$ supersets of size-$(c + 1)$. This can be shown by adding any one of the remaining $|N| - c - 1$ players ($-c$ refers to the c players already in the coalition and -1 refers to the player i). Conversely, each size-$(c + 1)$ coalition $\mathcal{C}_2$ (Right) has $c + 1$ subsets of size-c. This can be shown by removing any one of its $c + 1$ members. Arrows indicate the "$\subseteq$" relation.

(1) Map the size-c coalitions (excluding i) to their size-$(c + 1)$ supersets (excluding i), and vice-versa: each size-c coalition $\mathcal{C}_1$ can be mapped to $(|N| - 1 - c)$ number of size-$(c + 1)$ supersets $\mathcal{C}_2$ where $\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq N \setminus \{i\}$. This can be achieved by adding one of the remaining $(|N|-1-c)$ elements $j \in N \setminus (\{i\} \cup \mathcal{C}_1)$. Conversely, each $\mathcal{C}_2$ can be mapped to $(c+1)$ subsets $\mathcal{C}_1$ of size-c. This can be achieved by removing any one of the member elements $j \in \mathcal{C}_2$. An example is shown in Figure 3 for an illustration.

(2) With the mappings between size c and $c+1$ coalitions, we show that $z_i(c) \geq z_i(c+1)$ by the submodularity property: $\forall \mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq N \setminus \{i\} \implies MC_i(\mathcal{C}_1) \geq MC_i(\mathcal{C}_2)$. The detailed derivations are as follows: $\forall c \in [0, 1, \ldots, |N| - 2]$, denote $\mathcal{C}^c := \{\mathcal{C} \subseteq N \setminus \{i\} \mid |\mathcal{C}| = c\}$ as all possible coalitions of size c excluding player i, then

$$
\begin{aligned}
z_i(c+1) - z_i(c) &= \sum_{\mathcal{C}_2 \in \mathcal{C}^{c+1}} \binom{|N|-1}{c+1}^{-1} MC_i(\mathcal{C}_2) - \sum_{\mathcal{C}_1 \in \mathcal{C}^c} \binom{|N|-1}{c}^{-1} MC_i(\mathcal{C}_1) \\
&= \sum_{\mathcal{C}_2 \in \mathcal{C}^{c+1}} \Big(\binom{|N|-1}{c+1}^{-1} MC_i(\mathcal{C}_2) - \sum_{\mathcal{C}_1 \in \mathcal{C}^c, \mathcal{C}_1 \subseteq \mathcal{C}_2} \underbrace{\tfrac{1}{|N|-1-c}}_{(1)} \binom{|N|-1}{c}^{-1} \underbrace{MC_i(\mathcal{C}_1)}_{\geq MC_i(\mathcal{C}_2)} \Big) \\
&\leq \sum_{\mathcal{C}_2 \in \mathcal{C}^{c+1}} \Big(\binom{|N|-1}{c+1}^{-1} MC_i(\mathcal{C}_2) - \sum_{\mathcal{C}_1 \in \mathcal{C}^c, \mathcal{C}_1 \subseteq \mathcal{C}_2} \tfrac{1}{|N|-1-c} \binom{|N|-1}{c}^{-1} MC_i(\mathcal{C}_2) \Big) \\
&= \sum_{\mathcal{C}_2 \in \mathcal{C}^{c+1}} \Big(\binom{|N|-1}{c+1}^{-1} MC_i(\mathcal{C}_2) - \underbrace{\tfrac{c+1}{|N|-1-c}}_{(2)} \binom{|N|-1}{c}^{-1} MC_i(\mathcal{C}_2) \Big) \\
&= \sum_{\mathcal{C}_2 \in \mathcal{C}^{c+1}} \Big(\binom{|N|-1}{c+1}^{-1} MC_i(\mathcal{C}_2) - \binom{|N|-1}{c+1}^{-1} MC_i(\mathcal{C}_2) \Big) \\
&= 0
\end{aligned}
$$

(1) $\mathcal{C}_1$ is counted once in each of its $(|N|-1-c)$ supersets $\mathcal{C}_2$ of size-$(c+1)$, and (2) is because each $\mathcal{C}_2$ has $c+1$ subsets $\mathcal{C}_1$ of size-c. And this concludes our proof for $z_i(c) \geq z_i(c+1)$. ∎

We have shown that in a submodular game, a player is more useful *on average* when contributing towards a smaller coalition, i.e., the player's average marginal contribution towards a smaller coalition $z_i(c)$ is no less than its average marginal contribution to a bigger coalition $z_i(c+1)$. With this property, we are ready to extend the replication robustness results to the general class of semivalues and an arbitrary number of replications $k \geq 1$.

C. Payoff Changes under Replication with $k \geq 1$

To study the replication robustness of the semivalues, we first derive the total payoff of the replicating player according to the solution concepts after replication. Interestingly, we observe that under the replication redundancy assumption, the replicating player's total payoff can be expressed as a weighted sum of the player's average marginal contributions $z_i(c)$ *from the original game*, as detailed in the following lemma.

Lemma V-C.1. *Let $G = (N, v)$ be a submodular game with replication redundant characteristic function v. By replicating k times and acting as $k + 1$ players $\mathcal{C}^R = \{i_0, \ldots, i_k\}$ in the induced game $G^R = (N^R, v^R)$, the malicious player i receives a total payoff of*

$$
\varphi_i^{\text{tot}}(k) = \sum_{c=0}^{|N|-1} \alpha_c^k z_i(c), \text{ where} \tag{5}
$$

$$
z_i(c) = \binom{|N|-1}{c}^{-1} \sum_{\mathcal{C} \subseteq N \setminus \{i\}, |\mathcal{C}|=c} MC_i(\mathcal{C}),
$$

$$
\alpha_c^k = (k+1)\binom{|N|-1}{c} w_{c,N^R} \quad \text{(new importance weights)}.
$$

Proof Sketch: By symmetry the replicas yield equal payoff, i.e., $\varphi_i^{\text{tot}}(k) = (k+1)\varphi_{i_k}(N^R, v^R)$. Due to replication redundancy (Assumption 2), a replica player makes a nonzero marginal contribution only towards coalitions with no other replicas $\mathcal{C} \subseteq N^R \setminus \mathcal{C}^R$, which correspond to the same set of coalitions of the other players in the original game $\mathcal{C} \subseteq N \setminus \{i\}$ because $N^R \setminus \mathcal{C}^R = N \setminus \{i\}$. Following this insight, we can compute the new importance weights α_c^k over the player's original average marginal contributions $z_i(c)$. The complete proof is included in Appendix D. ∎

Equation (5) reduces to Equation (3) for no replications, i.e., $\alpha_c^k = \alpha_c$ when $k = 0$. Note the property that the importance weights sum to 1 only applies to $k = 0$ and does not apply to the new importance weights after replication. Importantly, $z_i(c)$ are the average marginal contributions defined on the original game $G = (N, v)$ as in Equation (3), instead of on the induced game, thus they *are invariant under replication*. As stated in Equation (5), the total payoff of the replicating player is a weighted sum over $z_i(c)$ with the new importance weights α_c^k. Since the average marginal contributions $z_i(c)$ in the original game stay invariant after replication, the change in the total payoff of the replicating player φ_i^{tot} is reflected in the change in α_c^k across different number of replications k. This makes α_c^k a key factor for characterising replication robustness. The next corollary demonstrates the importance weights after replication for the common semivalues.

Corollary V-C.1 (New Importance Weights for Common Semivalues after Replication)**.** *After k replications, the new importance weights for the total payoff of the malicious player are: for the Shapley value $\alpha_c^k = \frac{(k+1)\binom{|N|-1}{c}}{(|N|+k)\binom{|N|+k-1}{c}}$, for the Banzhaf value $\alpha_c^k = \frac{(k+1)}{2^{|N|+k-1}}\binom{|N|-1}{c}$, and for the Leave-one-out value $\alpha_c^k = \mathbb{1}_{c=|N|-1,k=0}$.*

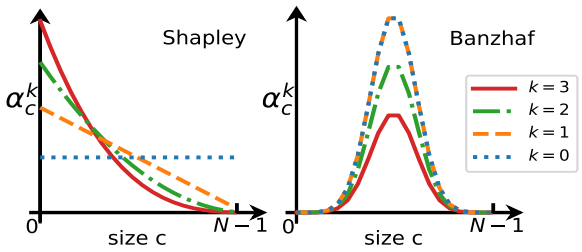


Fig. 4: Changes of α_c^k under different number of replications k, plot using Equation (5) with $|N| = 20$. The x-axis represents the sizes c of coalitions of the other players $N \setminus \{i\}$, and the y-axis shows the new importance weights α_c^k assigned to each coalition size. Each curve represents a different number of replications k. (Left) Across the different curves, the importance weights α_c^k of the Shapley value shift towards smaller coalitions as k increases (Lemma V-E.1). (Right) In contrast, the Banzhaf importance weights α_c^k are unchanged with the first replication, afterwards, α_c^k decreases across all coalition sizes as k increases. Since $z_i(c)$ decreases over coalition size c due to the submodular characteristic function (Lemma V-B.1), the weight shift of Shapley value causes φ_i^{tot} to be increasing, and non-increasing for the Banzhaf value.

Proof: The new importance weights can be obtained by plugging in the weights of the solution concepts in the induced game to Equation (5). ∎

In Example 2 and Fig. 4, we compare the Shapley value and the Banzhaf value using Equation 5, and illustrate the difference between these two solution concepts in terms of their new importance weights after replication.

Example 2 (Payoff Changes of the Malicious Player)**.** *Let $G = (N, v)$ be a submodular game with 3 players $N = \{i, p, q\}$. The marginal contributions of player i towards coalitions of other players are $MC_i(\emptyset) = 3$, $MC_i(\{p\}) = MC_i(\{q\}) = 2$, $MC_i(\{p, q\}) = 1$. Player i replicates once and acts under two identities $\mathcal{C}^R = \{i_1, i_2\}$. The induced game is then $G^R = (N^R, v^R)$ where $N^R = \{i_1, i_2, p, q\}$. To see the changes in i's total payoff, we first compute the average marginal contributions of i in the original game: $z_i(0) = MC_i(\emptyset) = 3$; $z_i(1) = \frac{1}{2}(MC_i(p) + MC_i(q)) = 2$; $z_i(2) = MC_i(p, q) = 1$. Then we compute the total payoffs $\varphi_i^{\text{tot}}(k)$ of player i according to the Shapley and Banzhaf value using Equation (5), where $k = 0$ refers to no replication, and $k > 0$ represents replicating k times:*

(Shapley) $\varphi_i^{\text{tot}}(0) = \sum_{c=0}^{2} \alpha_c^0 z_i(c) = 3 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 2$;
$\varphi_i^{\text{tot}}(1) = \sum_{c=0}^{2} \alpha_c^1 z_i(c) = 3 \cdot \frac{1}{2} + 2 \cdot \frac{1}{3} + 1 \cdot \frac{1}{6} = \frac{7}{3}$.

(Banzhaf) $\varphi_i^{\text{tot}}(0) = \sum_{c=0}^{2} \alpha_c^0 z_i(c) = 3 \cdot \frac{1}{4} + 2 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} = 2$;
$\varphi_i^{\text{tot}}(1) = \sum_{c=0}^{2} \alpha_c^1 z_i(c) = 3 \cdot \frac{1}{4} + 2 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} = 2$.

The above example demonstrates our observations from Fig. 4: the importance weights α_c^k of Shapley value shifts from uniform in the coalition sizes $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, to having larger weights towards smaller coalition sizes $(\frac{1}{2}, \frac{1}{3}, \frac{1}{6})$ after replication. Due to submodularity, the average marginal contributions are larger for smaller coalition sizes, hence the total payoff increases as a result of replication. Whereas for the Banzhaf value, the importance weights $\alpha_c^k = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ are invariant under the first replication, hence the total payoff is unchanged. In the following section we provide a formal characterisation of these observations.

D. Replication Robustness Condition

In this section, we will present the condition which characterises the replication robustness for semivalues. Specifically, this condition provides a sufficient and necessary condition on the importance weights α_c^k for guaranteeing replication robustness against any arbitrary number of replications k. By using a replication robust solution concept, a player should have no incentive to perform any number of replications in order to increase its payoff.

Theorem V-D.1 (Replication Robustness Condition)**.** *Given a submodular game with replication redundant characteristic function, a solution concept of the form $\varphi_i = \sum_{c=0}^{|N|-1} \alpha_c z_i(c)$ is replication robust* if and only if *for any number of replications k,*

$$\forall 0 \leq p \leq |N| - 1, \sum_{c=0}^{p} \alpha_c^0 \geq \sum_{c=0}^{p} \alpha_c^k, \tag{6}$$

where α_c^k are the importance weights as defined in Equation (5).

Proof: **Sufficiency.** We will first show the sufficient condition, that is, Equation (6) implies replication robustness, i.e., $\varphi_i^{\text{tot}}(0) - \varphi_i^{\text{tot}}(k) \geq 0$. Due to submodularity, the average marginal contributions of a player decrease as growing coalition sizes, according to Lemma V-B.1. Together with replication redundancy, we have the average marginal contributions satisfy the following condition $z_i(0) \geq \ldots \geq z_i(|N|-1) \geq 0$.

Let $\delta_c^k = \alpha_c^0 - \alpha_c^k$ denote the difference in importance weight over coalition size c before and after replication, then by Equation (6), $\forall p \in \{0, 1, \ldots, |N| - 1\}, \sum_{c=0}^{p} \delta_c^k \geq 0$. Therefore, we proceed to show the inequality recursively:

$$\begin{aligned}
\varphi_i^{\text{tot}}(0) - \varphi_i^{\text{tot}}(k) &= \textstyle\sum_{c=0}^{|N|-1} \delta_c^k z_i(c) \\
&= z_i(0) \textstyle\sum_{c=0}^{0} \delta_c^k + \sum_{c=1}^{|N|-1} \delta_c^k z_i(c) \\
&\overset{(1)}{\geq} z_i(1) \textstyle\sum_{c=0}^{0} \delta_c^k + \sum_{c=1}^{|N|-1} \delta_c^k z_i(c) \\
&= z_i(1) \textstyle\sum_{c=0}^{1} \delta_c^k + \sum_{c=2}^{|N|-1} \delta_c^k z_i(c) \\
&\overset{(2)}{\geq} z_i(2) \textstyle\sum_{c=0}^{1} \delta_c^k + \sum_{c=2}^{|N|-1} \delta_c^k z_i(c) \geq \ldots \\
&= z_i(|N|-2) \textstyle\sum_{c=0}^{|N|-2} \delta_c^k + \sum_{c=|N|-1}^{|N|-1} \delta_c^k z_i(|N|-1) \\
&\geq z_i(|N|-1) \textstyle\sum_{c=0}^{|N|-2} \delta_c^k + \sum_{c=|N|-1}^{|N|-1} \delta_c^k z_i(|N|-1) \\
&= z_i(|N|-1) \textstyle\sum_{c=0}^{|N|-1} \delta_c^k \geq 0.
\end{aligned}$$

where (1) is because $z_i(0) \geq z_i(1)$ and $\sum_{c=0}^{0} \delta_c \geq 0$, and (2) is because $z_i(1) \geq z_i(2)$ and $\sum_{c=0}^{1} \delta_c \geq 0$. With this, we have shown the sufficient condition, and we will next show the necessary condition.

Necessity. We now show that Equation (6) is also a necessary condition. To do this, we will prove by contradiction:

Let $\tilde{\delta}_c^k := \tilde{\alpha}_c^0 - \tilde{\alpha}_c^k$. Recall in Equation (6) for any coalition size $0 \leq p \leq |N| - 1, \sum_{c=0}^{q} \tilde{\delta}_c^k \geq 0$. Assume the contrary that there exists a set of coalition sizes (index) q_m where the

condition does not hold:

$$\exists Q_m = \{q_0, q_1, \ldots, q_m\}, \text{ such that } \forall q \in Q_m, \sum_{c=0}^{q} \tilde{\delta}_c^k < 0,$$

Without loss of generality, we assume the coalition sizes are ordered and that $q_0 < q_1 < \ldots < q_m \leq |N| - 1$. We now show that there exist average marginal contributions $z_i(c)$'s which violates replication robustness, and we construct them as follows: Looking at the smallest index that causes the contrary assumption $q_0 = \min Q_m$, all indices below q_0 satisfy the original condition, i.e.,

$$\sum_{c=0}^{p} \tilde{\delta}_c^k \begin{cases} < 0 & \text{if } p = q_0 \\ \geq 0 & \text{if } p < q_0 \end{cases}$$

Therefore, the sum of $\tilde{\delta}_c^k$ over all indices under q_0 is less than the absolute value of that at q_0, i.e., $0 \leq \sum_{c=0}^{q_0-1} \tilde{\delta}_c^k < -\tilde{\delta}_{q_0}^k = |\tilde{\delta}_{q_0}^k|$. Therefore, we denote $\gamma < 1$ as the ratio, such that $\sum_{c=0}^{q_0-1} \tilde{\delta}_c^k = \gamma |\tilde{\delta}_{q_0}^k|$. To construct $z_i(c)$, we let $\forall c > q_0, z_i(c) = 0$ for all indices above q_0, and let $z_i(q_0) = \gamma z_i(0) + \epsilon$ where $0 < \epsilon \leq (1-\gamma) z_i(0)$. Note that the $\epsilon > 0$ is for $z_i(q_0)$ to be strictly greater than $\gamma z_i(0)$, and $\epsilon \leq (1-\gamma) z_i(0)$ guarantees submodularity where $z_i(q_0) \leq z_i(0)$. In fact, a trivial choice would be a constant function for all indices no greater than q_0, i.e., $\forall q \in \{0, \ldots, q_0\}, z_i(q) = Const.$, but we will adopt the former option which also accounts for strictly submodular cases. Then, we have

$$\begin{aligned} \tilde{\varphi}_i^{\text{tot}}(0) - \tilde{\varphi}_i^{\text{tot}}(k) &= \textstyle\sum_{c=0}^{|N|-1} \tilde{\delta}_c^k z_i(c) \\ &\overset{(1)}{=} \textstyle\sum_{c=0}^{q_0} \tilde{\delta}_c^k z_i(c) = (\sum_{c=0}^{q_0-1} \tilde{\delta}_c^k z_i(c)) + \tilde{\delta}_{q_0}^k z_i(q_0) \\ &\overset{(2)}{\leq} (\textstyle\sum_{c=0}^{q_0-1} \tilde{\delta}_c^k) z_i(0) + \tilde{\delta}_{q_0}^k z_i(q_0) = |\tilde{\delta}_{q_0}^k| (\gamma z_i(0) - z_i(q_0)) \\ &= |\tilde{\delta}_{q_0}^k| (\gamma z_i(0) - \gamma z_i(0) - \epsilon) = -\epsilon |\tilde{\delta}_{q_0}^k| < 0, \end{aligned}$$

where (1) is due to $\forall q_0 < q \leq |N| - 1, z_i(q) = 0$, and (2) is due to submodularity.

This forms a contradiction. Thus we have shown that Theorem V-D.1 is both a necessary and sufficient condition for replication robustness, and this concludes our proof. ∎

The left hand side of inequality in Eq. (6) refers to a sum of importance weights in the case of no replications while the right hand refers to after replication. Intuitively, the condition ensures that after replication, there is not significant increase in importance weight on the small coalition sizes, towards which a player has larger average marginal contributions $z_i(c)$ due to submodularity. This effect was illustrated in Fig. 4 and the theorem is a formal characterisation. The significance of this theorem is that it provides a necessary and sufficient condition for guaranteeing replication robustness for all semivalues and for any number of replications. Therefore, a solution concept that satisfies the condition is replication robust without the need of knowing the number of replications k or the replicated false identities. Note that to guarantee robustness, the necessary and sufficient condition does not require the sum to be monotonic increasing/decreasing as the number of replications k. Extending from Theorem V-D.1, the following two corollaries provide sufficient conditions for monotonic decreasing (Corollary V-D.1) and monotonic increasing (Corollary V-D.2) total payoff of the replicating player with respect to the number of replications. These conditions will help us characterise the robustness of the common semivalues.

Corollary V-D.1 (Monotonic Decreasing Total Payoff)**.** *Given a submodular game with a replication redundant characteristic function, a semivalue is replication robust and the total payoff of the malicious player decreases monotonically i.e.,*$\varphi_i^{\text{tot}}(k) \geq \varphi_i^{\text{tot}}(k+1)$, if *for any number of replications k,*

$$\forall 0 \leq p \leq N-1, \sum_{c=0}^{p} \alpha_c^k \geq \sum_{c=0}^{p} \alpha_c^{k+1} \tag{7}$$

Note that the condition stated in Corollary V-D.1 is stricter than that in Theorem V-D.1 which implied $\varphi_i^{\text{tot}}(0) \geq \varphi_i^{\text{tot}}(k)$, but additionally ensures that the total payoff of a replicating player monotonic decreases with the number of replications, i.e., $\forall k \geq 0, \varphi_i^{\text{tot}}(k) \geq \varphi_i^{\text{tot}}(k+1)$.

Proof Sketch: We need to show that Equation (7) implies the total payoff decreases with k:

$$\forall k, \varphi_i^{\text{tot}}(k) - \varphi_i^{\text{tot}}(k+1) = \sum_{c=0}^{|N|-1} (\alpha_c^k - \alpha_c^{k+1}) z_i(c) \geq 0,$$

Denote $\delta_c^k := \alpha_c^k - \alpha_c^{k+1}$, then we can substitute δ_c^k in the recursive proof for the sufficient condition of Theorem V-D.1, by doing so we will reach the above conclusion. ∎

Corollary V-D.2 (Monotonic Increasing Total Payoff)**.** *Given a submodular game with a replication redundant characteristic function, a semivalue is not replication robust, and the total payoff of the malicious player increases monotonically i.e.,*$\varphi_i^{\text{tot}}(k) \leq \varphi_i^{\text{tot}}(k+1)$, if *for any number of replications k,*

$$\forall 0 \leq p \leq N-1, \sum_{c=0}^{p} \alpha_c^{k+1} \geq \sum_{c=0}^{p} \alpha_c^{k} \tag{8}$$

Proof Sketch: We need to show that Equation (8) implies the total payoff increases with k:

$$\forall k, \varphi_i^{\text{tot}}(k+1) - \varphi_i^{\text{tot}}(k) = \sum_{c=0}^{|N|-1} (\alpha_c^{k+1} - \alpha_c^{k}) z_i(c) \geq 0,$$

The proof is similar to the monotonic increasing case above. Denote $\delta_c^k := \alpha_c^{k+1} - \alpha_c^k$, and we can reuse the proof for Theorem V-D.1 to reach the above conclusion. ∎

E. Robustness of Common Solution Concepts

Now using the robustness conditions presented in the previous section, we can revisit the Shapley value and the Banzhaf value, as well as Leave-one-out with $k \geq 1$ replications. The following theorem is a consequence of our robustness condition for the three common semivalues.

Theorem V-E.1. *Let $G = (N, v)$ be a submodular game where v is replication redundant, the Shapley value is not replication robust, whereas the Banzhaf value and Leave-one-out are replication robust. For the Shapley value, the total payoff of the replicating player i monotonic increases over the number of replicas, and converges to i's characteristic value, i.e., $\lim_{k\to\infty} \varphi_i^{\text{tot}}(k) = v(\{i\})$. For the Banzhaf and Leave-one-out values, $\lim_{k\to\infty} \varphi_i^{\text{tot}}(k) = 0$.*

Proof: **For the Shapley Value.** We prove that the Shapley value is not replication robust, and the total payoff of the replicating player monotonic increases with growing k in the following three steps:

1. Express the Shapley value after replication according to Lemma V-C.1 in terms of average marginal contributions and importance weights, i.e., $\alpha_c^k = \frac{(k+1)}{(|N|+k)}\binom{|N|-1}{c}\binom{|N|+k-1}{c}^{-1}$.

2. In Lemma V-E.1 (presented following this theorem), we show that the Shapley value satisfies Equation (9b):

$$\forall 0 \leq p \leq N-1, \quad \sum_{c=0}^{p} \alpha_c^k \leq \sum_{c=0}^{p} \alpha_c^{k+1}.$$

3. By Theorem V-D.1, the Shapley value is not replication robust. In addition, Corollary V-D.1 shows that for the Shapley value, the total payoff of the replicating player monotonic increases with respect to increasing number of replications k.

Finally, the limit is computed as follows:

$$\begin{aligned}\lim_{k\to\infty} \varphi_i^{\text{tot}}(k) &= \lim_{k\to\infty} \sum_{c=0}^{|N|-1} \tfrac{k+1}{|N|+k}\binom{|N|-1}{c}\binom{|N|+k-1}{c}^{-1} z_i(c) \\ &= \sum_{c=0}^{|N|-1} \binom{|N|-1}{c} z_i(c) \underbrace{\lim_{k\to\infty} \tfrac{k+1}{|N|+k}}_{=1} \underbrace{\lim_{k\to\infty} \binom{|N|+k-1}{c}^{-1}}_{=\mathbb{1}_{c=0}} \\ &= z_i(0) = MC_i(\emptyset) = v(\{i\})\end{aligned}$$

Proofs of the Banzhaf value and Leave-one-out value are in Appendix E. ∎

We observe that due to replication redundancy and submodularity, the solution concepts which emphasize the complementary value tend to be more replication robust.

The robustness property of the Shapley value generalises our findings in Section IV-A for the $k = 1$ case. With an increasing number of replications, the player's total payoff monotonic increases and converges to its own characteristic value. This is due to the following properties shown in the next lemma.

Lemma V-E.1. *For the Shapley value, the importance weights α_c^k of the total payoff of a replicating player satisfy the following properties: $\forall k \geq 0, \forall 0 \leq p \leq |N|-1$,*

$$\sum_{c=0}^{|N|-1} \alpha_c^k = 1 \tag{9a}$$

$$\sum_{c=0}^{p} \alpha_c^k \leq \sum_{c=0}^{p} \alpha_c^{k+1} \tag{9b}$$

$$\sum_{c=0}^{p} \alpha_c^{k+1} - \alpha_c^k \geq \sum_{c=0}^{p} \alpha_c^{k+2} - \alpha_c^{k+1} \tag{9c}$$

Proof: The complete proof is included in Appendix F. ∎

Equation (9a) describes an interesting phenomenon that the new importance weights of the Shapley value after replication (i.e., α_c^k) always sum to 1. This is a special property of the Shapley value which is not shared by all semivalues. In contrast to the property of semivalues, i.e., $\sum_{c=0}^{|N|-1} \alpha_c = 1$, Equation (9a) states that the sum of the new importance weights after replication α_c^k over the coalitions of honest players in the original game always sum to 1. Moreover, Equation (9b) shows that these importance weights gradually *shifts* towards the smaller coalitions with each added replication, which results in the monotonic increasing total payoff. Collectively, Equation (9a) and Equation (9b) results in the convergence of the malicious player's total payoff to its characteristic value. Additionally, Equation (9c) implies that the gain of adding one replica decreases with replication, hence the first replication yields the highest unit gain.

To summarise, we have analysed and compared the replication robustness of the common semivalues, namely, the Shapley value, the Banzhaf value and the Leave-one-out value. In the following section, we discuss the design of other replication robust solution concepts.

F. Other Replication Robust Payoff Allocations

In this section, we describe how to apply the replication robustness condition to find other robust solution concepts and illustrate this with an example robust solution concept derived from the Shapley value.

Observation 1: To satisfy the robustness conditions in Theorem V-D.1, it suffices to satisfy one of the following conditions for each summand of coalition size c:

$$\alpha_c^0 \geq \alpha_c^k, \text{ or monotonicity: } \alpha_c^k \geq \alpha_c^{k+1} \tag{10}$$

Observation 2: The identity of the replicating player and the number of replicas k are private information that is often not accessible. Therefore, we should make sure that k does not appear in the solution concept.

We now derive a robust solution by down-weighing the Shapley value using these two observations. Our solution will take the following form, where the factor $\gamma_{|N|}^{|\mathcal{C}|}$ is a function of the total number of players $|N|$ and coalition size $|\mathcal{C}|$:

$$\tilde{\varphi}_i(N,v) := \sum_{\mathcal{C}\subseteq N\setminus\{i\}} \gamma_{|N|}^{|\mathcal{C}|} w_{|\mathcal{C}|,N} MC_i(\mathcal{C}) \tag{11}$$

where $w_{|\mathcal{C}|,N} = \frac{|\mathcal{C}|!(|N|-|\mathcal{C}|-1)!}{|N|!}$ are the Shapley coefficients.

Definition V-F.1. *(Robust Shapley value) Equation (11) with*

$$\gamma_{|N|}^{|\mathcal{C}|} = \begin{cases} \frac{\lceil \frac{|N|-1}{2} \rceil ! \lfloor \frac{|N|-1}{2} \rfloor !}{|\mathcal{C}|!(|N|-|\mathcal{C}|-1)!} & \text{if } |\mathcal{C}| < \lfloor \frac{|N|-1}{2} \rfloor, \\ 1 & \text{otherwise.} \end{cases}$$

defines the Robust Shapley value.

Corollary V-F.1. *The Robust Shapley value is replication robust. Moreover, in a submodular game $G = (N, v)$, the loss for a replicating player i by replicating k times $\varphi_i^{\text{tot}}(0) - \varphi_i^{\text{tot}}(k) \geq \frac{1}{|N|}\sum_{c=0}^{|N|-1}(1 - \frac{k+1}{2^k})\gamma_{|N|}^c z_i(c)$.*

Proof: The complete proof is included in Appendix G ∎

The Robust Shapley value satisfies axioms symmetry **(A1)**, null-player **(A3)**, linearity **(A4)**. Additionally, the total allocated payoff does not exceed the value of the grand coalition.

Like the Banzhaf value and Robust Shapley value, there are many other possible solution concepts which are replication robust. These solution concepts can be crafted by designing the importance weights. Recall that semivalues balance between a player's individual value and complementary value through the

TABLE I: Computation Time of the Shapley and Banzhaf value in the Facility Location Game (in seconds) for n players using the naive approach and our algorithm (Algorithm 1).

		n=10	n=15	n=20	n=50	n=100
Shapley value	*naive*	0.283	14.182	558.525	-	-
	ours	0.004	0.007	0.011	0.099	0.225
Banzhaf value	*naive*	0.245	12.623	482.803	-	-
	ours	0.002	0.003	0.006	0.056	0.170

importance weights. As a rule of thumb, the solution concepts which emphasize the complementary value and put larger importance weights on the mid-sized and larger coalitions tend to be more replication robust.

G. Perturbed Replication

Sometimes, the manipulations may not be an exact replication. We now consider a related scenario where the malicious player replicates its resources and splits into multiple identities, then performs a small perturbation on each of its replicated resources to avoid being detected as a replica, such as adding noise or performing transformations on an image. A perturbed replication can be formulated as follows: In the submodular game $G = (N, v)$, a malicious player i replicates its resources D_i k times and acts as $k+1$ players $\mathcal{C}^R = \{i_0, \ldots, i_k\}$ where $D_{i_k} = D_i$. The player further perturb its replicas as $\mathcal{C}^P = \{p_0, \ldots, p_k\}$, where $D_{p_k} = f_k(D_i)$ for some perturbation function f_k. The malicious player receives a total payoff as a sum of all its perturbed replicas, i.e., $\varphi_i^{\text{replicate}} = \sum_{i_k \in \mathcal{C}^R} \varphi_{i_k}$ and $\varphi_i^{\text{perturb}} = \sum_{p_k \in \mathcal{C}^P} \varphi_{p_k}$. Assume the effect of perturbations are small such that (1) the marginal contribution of the perturbed replicas towards the other players remain unchanged, that is:

$$\forall \mathcal{C} \subseteq N \setminus \{i\}, MC_{p_k}(\mathcal{C}) = MC_{i_k}(\mathcal{C}),$$

and (2) the marginal contributions of each perturbed replica towards coalitions containing other perturbed replicas are small, that is, there exists a small quantity $\exists \epsilon > 0$ s.t.,

$$\forall p_k \in \mathcal{C}^P, \emptyset \neq \mathcal{C}^p \subseteq \mathcal{C}^P \setminus \{p_k\}, \mathcal{C} \subseteq N \setminus \{i\}, MC_{p_k}(\mathcal{C}^p \cup \mathcal{C}) \leq \epsilon$$

Lemma V-G.1. *Compared with replication, the additional gain in total payoff of the malicious player due to the perturbation when replicating k times is given by:*

$$\varphi_i^{\text{perturb}} - \varphi_i^{\text{replicate}} \leq (k+1)\epsilon.$$

Proof: The proof is included in Appendix H. ∎

In this way, perturbations which yield negligible marginal values towards other players and the other perturbed replicas will yield negligible benefit compared with the non-perturbed replicas. Therefore, the solution concepts which are replication robust against (pure) replications are also guaranteed to be $\epsilon-$robust against the perturbed replication manipulation.

H. Robustness Results on the Facility Location Game

Having presented the theoretical results on replication robustness, we now demonstrate these findings on the facility location game as defined in Example 1. Before that, we present the following theorem, which efficiently computes the Shapley and Banzhaf value in the facility location game, allowing us efficiently visualise their convergence properties.

Theorem V-H.1. *The Shapley and Banzhaf value of a facility location i in a facility location game can be computed as*

$$\varphi_i^{\text{Shapley}} = \sum_{d \in D} \Big[u_{id} \frac{1}{|\mathcal{L}| - |\mathcal{L}_{id}|} - \sum_{t=1}^{|\mathcal{L}_{id}|} \frac{1}{\lambda(t) + \lambda(t)^2} u_{e_{it}^d} \Big],$$

$$\varphi_i^{\text{Banzhaf}} = \frac{1}{2^{|\mathcal{L}|-1}} \sum_{d \in D} \Big[2^{|\mathcal{L}_{id}|} u_{id} - \sum_{t=1}^{|\mathcal{L}_{id}|} 2^{|\mathcal{L}_{id}|-t} u_{e_{it}^d} \Big],$$

where $\lambda(t) \coloneqq (|\mathcal{L}| - |\mathcal{L}_{id}| + t - 1)$, $\mathcal{L}_{id} \coloneqq \{j \in \mathcal{L} \mid u_{jd} \leq u_{id}\}$ and $u_{e_{it}^d}$ is the utility value of the t-th largest element after i along the dimension (customer) d, D is the set of customers, u_{id} is the utility of a customer d from facility location i.

Proof Sketch: Denote $w_{\mathcal{C}}$ as the weight assigned by the Shapley (Banzhaf) value to coalition $\mathcal{C}$. We observe that $\varphi_i = \sum_{\mathcal{C} \subseteq \mathcal{L} \setminus \{i\}} w_{\mathcal{C}} MC_i(\mathcal{C}) \overset{(*)}{=} \sum_{d \in D} \big[\underbrace{\sum_{\mathcal{C} \subseteq \mathcal{L}_{id}} w_{\mathcal{C}} u_{id}}_{(\#1)} - \underbrace{\sum_{\mathcal{C} \subseteq \mathcal{L}_{id}} w_{\mathcal{C}} \max_{j \in \mathcal{C}} u_{jd}}_{(\#2)} \big]$, where $(*)$ is because the marginal contribution of i to coalition $\mathcal{C}$ in dimension (customer) d is zero *unless* i is the largest element in $\mathcal{C}$ in the d-th dimension, i.e., subsets of $\mathcal{L}_{id} = \{j \in \mathcal{L} \mid u_{jd} \leq u_{id}\}$. Along each dimension d, $(\#1)$ is a weighted sum of i's marginal contributions towards coalitions $\mathcal{C}$ where i is the largest element; $(\#2)$ sums up for each $j \in \mathcal{L}_{id}$ over coalitions $\mathcal{C} \subseteq \mathcal{L}_{id}$ where j is the largest element. The proof is included in Appendix J. ∎

The Shapley and Banzhaf value of the facility location game can be computed using Lemma V-H.1, which can be implemented efficiently by sorting the facility location utility matrix along each dimension (customer), as summarised in Algorithm 1 in Appendix I. Table I demonstrates that our algorithm (ours) significantly improves the computation efficiency compared with the naive algorithm (naive) which enumerates all possible coalitions. The output values computed by both algorithms are verified to be equal. We observed that the naive algorithm struggle in games with large number of players (e.g., $n \geq 50$) while ours scales up easily. With the help of Algorithm 1, we can efficiently visualise the Shapley and Banzhaf value in Fig. 1 and validate their robustness and convergence properties under replication in Fig. 5.

Fig. 5 shows the replication robustness of the Shapley value and the Banzhaf value in the facility location game. Specifically, the original facility location game includes $|\mathcal{L}| = 10$ players (facility locations) and $|D| = 10$ customers, each utility value u_{ij} is an integer uniformly sampled from $[0, 20]$. Among the players, a malicious player i replicates itself k times and acts under $k+1$ identities. The curves show the total payoff of the player i with respect to growing number of replications k from 0 to 50. The graph validates some of our findings. Specifically, the Shapley value is not replication robust: the total payoff of the player monotonically increases, and converges to the characteristic value of the player, i.e., $v(\{i\})$. Moreover, the unit gain of the player for adding each

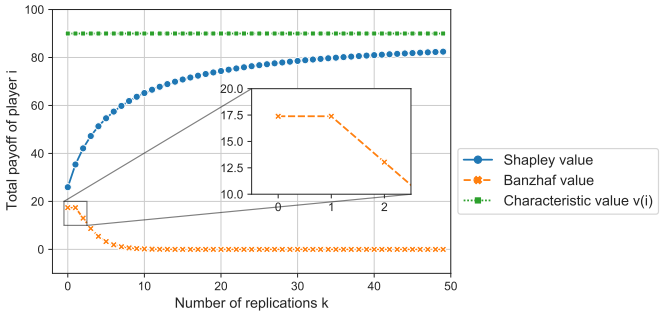


Fig. 5: Replication robustness of the Shapley and Banzhaf value for the Facility Location Game

player monotonically decreases. This can be seen from the decreasing height between pairs of adjacent points. In contrast, the Banzhaf value is replication robust: the total payoff of the replicating player monotonic decreases, and converges to 0. Moreover, by comparing $k = 0$ and $k = 1$ (zoomed), we can see that the Banzhaf value is neutral to the first replication.

VI. Case Study: ML Data Markets

Our theoretical results can be applied to study redundancy and payoff allocation in many submodular real-world ML applications such as multiagent sensing, feature importance evaluation and multi-party ML. In this section, we investigate payoff allocation for ML data markets [2, 30] – an emerging application which readily connects data buyers (i.e., ML practitioners) with data sellers, providing a nice alternative to addressing the challenge of data acquisition in real-world ML applications. A naive implementation of such a market in the form of a direct data exchange is likely to fail in practice as data can be freely replicated, and hence may be easily be resold by a buyer. Moreover, acquiring ownership of a large dataset may exceed the budget of the buyer. These issues can be alleviated by modelling the market as an integral part of a cloud ML platform: At each round of interaction, the buyer provides a classification task, specified by a validation dataset D_{val}. The data from multiple data sellers will be pooled securely to jointly train a model $\mathcal{M}(\cup_{i\in N} D_i)$ towards the classification task. The buyer will then pay a fee according to the performance of the model, and the sellers are allocated a payoff according to their data's contributions.

In the following, we model the market as a submodular game, and apply our theoretical insights to study robust payoff allocation against replication, i.e., data replication attacks.

A. Data Market as a Submodular Game

We model each round of interaction as a cooperative game $G = (N, v)$, where the players i are the data sellers $N = \{1, \ldots, n\}$, each holding a dataset D_i. A natural characteristic function is given by the accuracy $\mathcal{G}(\mathcal{M}, D_{\text{val}})$ achieved by the model $\mathcal{M}$ trained on the data held by players in the coalition:

$$v(\mathcal{C}) \coloneqq \mathcal{G}(\mathcal{M}(\cup_{i\in\mathcal{C}} D_i), D_{\text{val}})$$

Submodularity is often a good model for approximating properties of this accuracy—the value of additional training datasets typically diminishes with growing data size [18].

B. Data Replication Attack and Replication Robustness

In the context of a data market game, a replication manipulation can be implemented by a malicious player through replicating its data and acting under multiple false identities. For many ML models, redundant data do not significantly change the model's performance and hence satisfies the replication redundancy assumption. As the market game is an instance of submodular game with replication redundant characteristic function, we can directly apply our replication robustness results shown for submodular games to evaluate the common solution concepts. That is, in a market game G, the Shapley value is not robust against the data replication attack, and the malicious player is incentivised to replicate its data and act under multiple identities in order to increase its total payoff. Whereas the Banzhaf value, Leave-one-out value, and Robust Shapley value are robust against the data replication attack. Similarly, the conditions for replication robustness presented in Section V-D also hold for the ML data market game.

C. Experiments

In this section, we provide empirical validations to our theoretical results on the ML data market. Specifically, we will present experiments which justify our assumptions on submodularity and replication redundancy. We then compare the replication robustness of the discussed solution concepts.

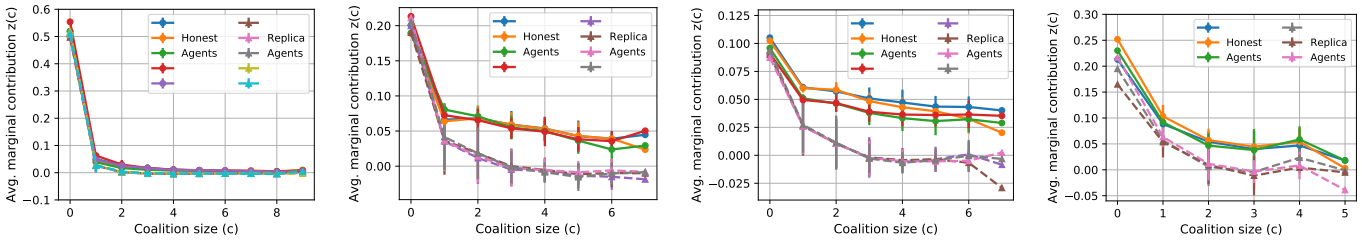
1) Experiment Setup: We test our results on three standard ML tasks (datasets) of varied sizes. On each task, we assign to each player a subset of the data, and a malicious player replicates its data and we gradually increase the number of replications k. The datasets and assignments are as follows:

(a) *Covertype* [9]: Each input consists of 10 continuous features (e.g., elevation, slope, hillshade 9am, etc.), and the output is a prediction of the forest cover type out of 7 classes. We use the dataset provided by Kaggle which consists of ∼15000 training datapoints uniformly distributed in the 7 output classes. 5 honest players each holds 1000 datapoints, 5 replicas share 1000 datapoints.

(b) *CIFAR-100* [21]: 32x32x3 images of 20 superclasses and 100 subclasses C_{sub}. We carried out 4 sets of experiments with varied data assignments as follows:
- *Uniform:* Players $0-4$ assigned data from 100 C_{sub} uniformly, players $5-7$ (replicas) same as Player 0.
- *Disjoint:* Players $0-4$ each assigned 20 C_{sub}, players $5-7$ (replicas) assigned the same data as Player 0.
- *Mixed:* Players $0-4$ assigned varied portions of each C_{sub}, players $5-7$ (replicas) same as Player 0.

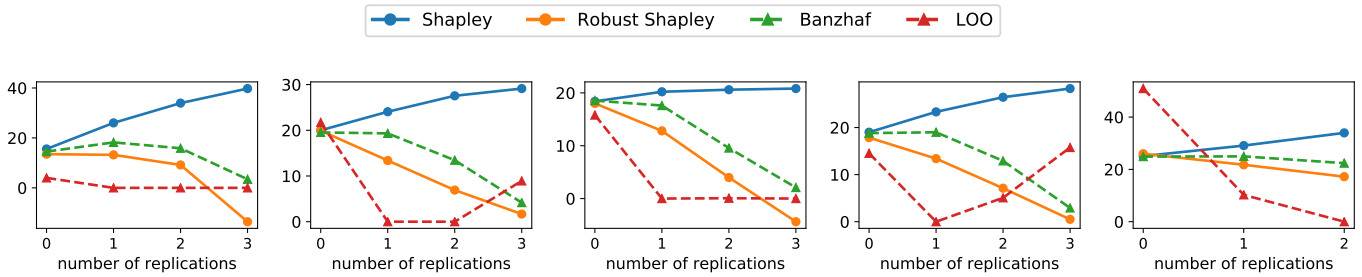
(c) *Tiny ImageNet* [23]: 64x64x3 images of 20 random classes. 3 honest players each holds 2000 datapoints and 3 replicas hold the same 2000 datapoints.

To construct the ML models, we used a 4-layer (512 units per layer) fully-connected neural network for Covertype prediction. For CIFAR-100, we used the VGG-16 architecture [37] with 10 convolutional layers (kernel size 3), max-pooling, and 2 fully-connected layers (1024 units per layer). For Tiny-ImageNet, we used the VGG-16 with 2 fully-connected layers (4096 units per layer). Adam optimizer [17] is used to train



(a) Covertype (b) CIFAR-100 (uniform) (c) CIFAR-100 (disjoint) (d) Tiny ImageNet

Fig. 6: Average marginal contributions $z_i(c)$ across various datasets. Solid lines are non-replicating players while dashed lines are replicas which belong to the malicious player. Error bars show standard deviations of the marginal contributions of each coalition size. Observe that $z_i(c)$ monotonic decreases with coalition size as a result of submodularity.



(a) Covertype (b) CIFAR(uniform) (c) CIFAR (disjoint) (d) CIFAR (mixed) (e) Tiny ImageNet

Fig. 7: Percentage of total replica values in the total allocated payoffs w.r.t number of replications. x-axis represents increasing number of replications by the malicious player, e.g. x=3 refers to an induced game where the malicious player holds 4 replicas.

the models. For the Covertype classification, we use learning rate of 0.0001, minibatch size 128. For CIFAR-100, we use learning rate of 0.001, minibatch size 64. For Tiny ImageNet we use learning rate of 0.001, minibatch size 64.

2) Validations on properties of the ML Data Market: We empirically validate Assumption 1 (submodularity) and Assumption 2 (replication redundancy). Fig. 6 shows the average marginal contributions $z_i(c)$ for each player over coalition sizes c. Observe that $z_i(c)$ is monotonic decreasing, which according to Lemma V-B.1, is a result of the submodularity of the characteristic function. The curves further validate replication redundancy with $z_i(c) \approx 0$ for the replica players when c exceeds the number of honest players.

3) Replication Robustness: Fig. 7 compares the replication robustness of various solution concepts. The curves show the changes in total payoffs of the malicious player as a percentage of the total allocated payoffs, over growing number of replicas. On the Covertype, CIFAR-100, and Tiny ImageNet tasks, we start with 5,4,3 honest players respectively and 1 malicious player, and along the x-axis, we gradually increase the number of replicas. In all settings, the Shapley value is vulnerable to replication, and the total share of value gained by the replica player increases. Both the Banzhaf value and Robust Shapley value are replication robust. The Leave-one-out value is sensitive to the randomness during training, because it only includes a player's marginal contribution towards all other players. We plot the percentage for easy comparison, which also preserves the trend of the actual value.

VII. Related Work

Our theoretical results relate closely to the seminal game-theoretic literature on merging/splitting proofness, collusion, and false name manipulations. Lehrer [22] is the first to present an axiomatization of the Banzhaf value with the 2-efficiency axiom, which characterized the neutrality of the Banzhaf value on merging two players as one. Similarly, Haller [13] studied the collusion of two players where they both keep their identities thus the total number of players are unchanged: under a proxy agreement, one player acts as a proxy while the other a null player, whereas under an association agreement, two players act on each other's behalf. van den Brink [40] studied the interplay between efficiency and collusion neutrality of two players. Knudsen and Østerdal [19] studied the merging and splitting-proofness on convex games, and introduced some possibility/impossibility results. Ohta et al. [31] studied false name manipulation in an open environment and proposed anonymity-proof Shapley value against malicious players who split their *skills* and act as multiple identities, where skills are assumed to be unique. Related to the splitting manipulations, our present work looks at the replication manipulation arising in submodular games. Such cases have not been adequately addressed previously. Moreover, our results extends from bilateral amalgamation to an arbitrary number of replica players. Related work on emerging ML applications include (1) ML data markets, e.g., Agarwal et al. [2] first introduced an algorithmic framework for data marketplaces. Ohrimenko et al. [30] studied collaborative ML data markets where each player must participate both as seller and buyer. (2) ML model interpretation [38, 15],

which explains ML models through the feature importance. Many have adopted game-theoretic solution concepts such as the Shapley Value [26, 39, 6, 5], and (3) Submodular data and feature selection [41, 18, 8, 25].

VIII. Conclusions

In this work, we studied the robustness of solution concepts against redundancy as a result of replication in submodular games. In summary, we showed a necessary and sufficient condition which characterises the robustness of semivalues in general. Using this condition, we showed that the Shapley value is not replication robust, i.e., the total payoff of the malicious player monotonic increases with growing number of replications. Whereas the Banzhaf value, Robust Shapley and Leave-one-out value are replication robust. We demonstrate the distinct robustness and convergence properties of the Shapley and Banzhaf value on a submodular facility location game. Moreover, we applied our theoretical results to an emerging application of ML data markets, and empirically validated our theoretical results across three standard ML datasets. Interesting future directions include extending our theoretical framework for submodular games with partial redundancy; and applying our theoretical findings to submodular ML applications such as feature evaluation and multiagent learning.

Acknowledgment

Wooldridge was supported by the Turing AI Fellowship entitled "The Large Agent Collider: Robust agent-based modelling as scale" (Grant number EP/W002949/1). The authors would like to thank the editors and the reviewers for their time and valuable comments that helped improve the manuscript.

References

[1] K. V. Aadithya, B. Ravindran, T. P. Michalak, and N. R. Jennings. Efficient computation of the shapley value for centrality in networks. In *International workshop on internet and network economics*, pages 1–13. Springer, 2010.

[2] A. Agarwal, M. Dahleh, and T. Sarkar. A marketplace for data: An algorithmic solution. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 701–726, 2019.

[3] J. Bilmes and W. Bai. Deep submodular functions. *arXiv preprint arXiv:1701.08939*, 2017.

[4] G. Chalkiadakis, E. Elkind, and M. Wooldridge. *Computational Aspects of Cooperative Game Theory*. Morgan & Claypool Publishers, 2011.

[5] J. Chen, L. Song, M. J. Wainwright, and M. I. Jordan. L-shapley and c-shapley: Efficient model interpretation for structured data. In *International Conference on Learning Representations (ICLR)*, 2019.

[6] S. B. Cohen, E. Ruppin, and G. Dror. Feature selection based on the shapley value. In *International Joint Conference on Artificial Intelligence (IJCAI)*, volume 5, pages 665–670, 2005.

[7] G. Cornuejols, M. Fisher, and G. L. Nemhauser. On the uncapacitated location problem. In *Annals of Discrete Mathematics*, volume 1, pages 163–177. Elsevier, 1977.

[8] A. Das, A. Dasgupta, and R. Kumar. Selecting diverse features via spectral regularization. *Advances in neural information processing systems*, 25:1583–1591, 2012.

[9] D. Dua and C. Graff. UCI machine learning repository, 2017. URL http://archive.ics.uci.edu/ml.

[10] P. Dubey, A. Neyman, and R. J. Weber. Value theory without efficiency. *Mathematics of Operations Research*, 6(1):122–128, 1981.

[11] M. L. Fisher, G. L. Nemhauser, and L. A. Wolsey. An analysis of approximations for maximizing submodular set functions—ii. In *Polyhedral combinatorics*, pages 73–87. Springer, 1978.

[12] M. X. Goemans and M. Skutella. Cooperative facility location games. *Journal of Algorithms*, 50(2):194–214, 2004.

[13] H. Haller. Collusion properties of values. *International Journal of Game Theory*, 23(3):261–281, 1994.

[14] D. Han, C. X. Lu, T. Michalak, and M. Wooldridge. Multiagent model-based credit assignment for continuous control. *arXiv preprint arXiv:2112.13937*, 2021.

[15] D. Janzing, L. Minorics, and P. Blöbaum. Feature relevance quantification in explainable ai: A causal problem. In *International Conference on Artificial Intelligence and Statistics*, pages 2907–2916. PMLR, 2020.

[16] R. Jia, D. Dao, B. Wang, F. A. Hubis, N. Hynes, N. M. Gürel, B. Li, C. Zhang, D. Song, and C. J. Spanos. Towards efficient data valuation based on the shapley value. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1167–1176, 2019.

[17] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[18] K. Kirchhoff and J. Bilmes. Submodularity for data selection in machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 131–141, 2014.

[19] P. H. Knudsen and L. P. Østerdal. Merging and splitting in cooperative games: some (im) possibility results. *International Journal of Game Theory*, 41(4):763–774, 2012.

[20] A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9(2), 2008.

[21] A. Krizhevsky. Learning multiple layers of features from tiny images. Technical report, 2009.

[22] E. Lehrer. An axiomatization of the banzhaf value. *International Journal of Game Theory*, 17(2):89–99, 1988.

[23] F. Li, A. Karpathy, and J. Johnson. Tiny imagenet visual recognition challenge. URL https://tiny-imagenet.herokuapp.com/.

[24] J. Li, K. Kuang, B. Wang, F. Liu, L. Chen, F. Wu, and J. Xiao. Shapley counterfactual credits for multi-agent reinforcement learning. *arXiv preprint arXiv:2106.00285*, 2021.

[25] H. Lin and J. Bilmes. A class of submodular functions for document summarization. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, pages 510–520, 2011.

[26] S. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*, 2017.

[27] S. M. Lundberg, G. G. Erion, and S.-I. Lee. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*, 2018.

[28] T. Matsui and Y. Matsui. A survey of algorithms for calculating power indices of weighted majority games. *Journal of the Operations Research Society of Japan*, 43 (1):71–86, 2000.

[29] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical programming*, 14(1):265–294, 1978.

[30] O. Ohrimenko, S. Tople, and S. Tschiatschek. Collaborative machine learning markets with data-replication-robust payments. *arXiv preprint arXiv:1911.09052*, 2019.

[31] N. Ohta, V. Conitzer, Y. Satoh, A. Iwasaki, and M. Yokoo. Anonymity-proof shapley value: extending shapley value for coalitional games in open environments. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 2*, pages 927–934, 2008.

[32] R. Patel, M. Garnelo, I. Gemp, C. Dyer, and Y. Bachrach. Game-theoretic vocabulary selection via the shapley value and banzhaf index. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2789–2798, 2021.

[33] M. T. Ribeiro, S. Singh, and C. Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[34] S. Salhi. Discrete location theory. *Journal of the Operational Research Society*, 42:1124–1125, 1991.

[35] A. Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer Science & Business Media, 2003.

[36] L. S. Shapley. *17. A Value for n-Person Games*, pages 307–318. Princeton University Press, 2016. doi: doi:10.1515/9781400881970-018. URL https://doi.org/10.1515/9781400881970-018.

[37] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[38] E. Štrumbelj and I. Kononenko. Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems*, 41(3):647–665, 2014.

[39] M. Sundararajan and A. Najmi. The many shapley values for model explanation. In *International Conference on Machine Learning*, pages 9269–9278. PMLR, 2020.

[40] R. van den Brink. Efficiency and collusion neutrality in cooperative games and networks. *Games and Economic Behavior*, 76(1):344–348, 2012.

[41] K. Wei, R. Iyer, and J. Bilmes. Submodularity in data subset selection and active learning. In *International Conference on Machine Learning*, pages 1954–1963. PMLR, 2015.

APPENDIX A
PROOF FOR THEOREM IV-A.1

Theorem IV-A.1. *Let G= (N, v) be a submodular game with replication redundant characteristic function v. Assume that player $i \in N$ replicates and obtains the total payoff of the two identities $\mathcal{C}^R = \{i_1, i_2\}$ in the new game $G^R = (N^R, v^R)$. The change in total payoff of player i because of relication is:*

$$\delta\varphi_i^{\text{Shapley}} = \sum_{\mathcal{C}\subseteq N\setminus\{i\}} \frac{|\mathcal{C}|!(|N|-|\mathcal{C}|-1)!}{(|N|+1)!}(|N|-2|\mathcal{C}|-1)MC_i(\mathcal{C}).$$

Moreover, the total payoff of player i after replication is no less than its payoff in the original game, i.e, $\delta\varphi_i^{\text{Shapley}} \geq 0$.

Proof: The second half of the theorem is provided in the main text, here we provide the derivation for $\delta_i^{\text{Shapley}}$. In the induced game $G^R = (N^R, v^R)$ where player i replicates into two players $\{i_1, i_2\}$, the total number of players increases by one, i.e., $|N^R| = |N| + 1$ and $v^R(\mathcal{C} \cup \{i_1, i_2\}) = v^R(\mathcal{C} \cup \{i\})$ as i_1 and i_2 are replicas of i and as a result of replication redundancy. We next write out the sum of the Shapley values $\varphi_{i_2}^R$ and $\varphi_{i_2}^R$ of i_1, i_2 in G^R.

$$\begin{aligned}
\varphi_{i_1}^R(N^R, v^R) &\coloneqq \sum_{\mathcal{C}\subseteq N^R\setminus\{i\}} \frac{|\mathcal{C}|!(|N^R|-|\mathcal{C}|-1)!}{|N^R|!} MC_{i_1}(\mathcal{C}) \\
&= \sum_{\mathcal{C}\subseteq N^R\setminus\{i\}} \frac{|\mathcal{C}|!(|N|+1-|\mathcal{C}|-1)!}{(|N|+1)!} MC_{i_1}(\mathcal{C}) \\
&\overset{(1)}{=} \sum_{\mathcal{C}\subseteq N^R\setminus\{i_1,i_2\}} \frac{|\mathcal{C}|!(|N|-|\mathcal{C}|)!}{(|N|+1)!} MC_{i_1}(\mathcal{C}) + \sum_{\mathcal{C}\subseteq N^R\setminus\{i_1,i_2\}} \frac{(|\mathcal{C}|+1)!(|N|-|\mathcal{C}|-1)!}{(|N|+1)!} \underbrace{MC_{i_1}(\mathcal{C}\cup\{i_2\})}_{=0,\text{ replication redundancy}} \\
&= \underbrace{\sum_{\mathcal{C}\subseteq N^R\setminus\{i_1,i_2\}}}_{=N\setminus\{i\}} \frac{|\mathcal{C}|!(|N|-|\mathcal{C}|)!}{(|N|+1)!} \underbrace{MC_{i_1}(\mathcal{C})}_{=MC_i(\mathcal{C})} = \sum_{\mathcal{C}\subseteq N\setminus\{i\}} \frac{|\mathcal{C}|!(|N|-|\mathcal{C}|)!}{(|N|+1)!} MC_i(\mathcal{C}),
\end{aligned}$$

where (1) is by grouping the coalitions of players (excluding i_1) into two groups, one group containing all coalitions without i_2 and the other with i_2 in.

By symmetry, $\varphi_{i_2}^R(N^R, v^R) = \varphi_{i_1}^R(N^R, v^R)$, and the total payoff of i in the induced game is:

$$\varphi_i^R = 2\varphi_{i_1}^R(N^R, v^R) = \sum_{\mathcal{C}\subseteq N\setminus\{i\}} \frac{2|\mathcal{C}|!(|N|-|\mathcal{C}|)!}{(|N|+1)!} MC_i(\mathcal{C})$$

On the other hand, the total payoff of player i in the original game is:

$$\varphi_i = \sum_{\mathcal{C}\subseteq N\setminus\{i\}} \frac{|\mathcal{C}|!(|N|-|\mathcal{C}|-1)!}{|N|!} MC_i(\mathcal{C})$$

Therefore, the change in total payoff of player i is :

$$\begin{aligned}
\delta\varphi_i^{\text{Shapley}} &= \varphi_i^R - \varphi_i \\
&= \sum_{\mathcal{C}\subseteq N\setminus\{i\}} \frac{|\mathcal{C}|!(|N|-|\mathcal{C}|-1)!}{|N|+1!}(2(|N|-|\mathcal{C}|)-(|N|+1))MC_i(\mathcal{C}) \\
&= \sum_{\mathcal{C}\subseteq N\setminus\{i\}} \frac{|\mathcal{C}|!(|N|-|\mathcal{C}|-1)!}{|N|+1!}(|N|-2|\mathcal{C}|-1)MC_i(\mathcal{C})
\end{aligned}$$

■

APPENDIX B
PROOF FOR THEOREM IV-B.1

Theorem IV-B.1. *Let G= (N, v) be a submodular game with replication redundant characteristic function v. Assume that player $i \in N$ replicates and obtains the total payoff of the two identities $\mathcal{C}^R = \{i_1, i_2\}$ in the new game $G^R = (N^R, v^R)$. Under payoff allocation using the Banzhaf value, the change in total payoff of player i because of replication is zero, i.e., $\delta\varphi_i^{\text{Banzhaf}} = 0$*

Proof: In the induced game $G^R = (N^R, v^R)$ where player i replicates into two players $\{i_1, i_2\}$, the total number of players increases by one, i.e., $|N^R| = |N| + 1$ and $v^R(\mathcal{C} \cup \{i_1, i_2\}) = v^R(\mathcal{C} \cup \{i\})$ as i_1 and i_2 are replicas of i and as a result of replication redundancy. We next write out the sum of the Banzhaf values $\varphi^R_{i_2}$ and $\varphi^R_{i_2}$ of i_1, i_2 in G^R.

$$\begin{aligned}
\varphi^R_{i_1}(N^R, v^R) &:= \sum_{\mathcal{C}\subseteq N^R\setminus\{i\}} \frac{1}{2^{|N^R|}} MC_{i_1}(\mathcal{C}) \\
&= \sum_{\mathcal{C}\subseteq N^R\setminus\{i\}} \frac{1}{2^{|N|+1}} MC_{i_1}(\mathcal{C}) \\
&\overset{(1)}{=} \sum_{\mathcal{C}\subseteq N^R\setminus\{i_1,i_2\}} \frac{1}{2^{|N|+1}} MC_{i_1}(\mathcal{C}) + \sum_{\mathcal{C}\subseteq N^R\setminus\{i_1,i_2\}} \frac{1}{2^{|N|+1}} \underbrace{MC_{i_1}(\mathcal{C}\cup\{i_2\})}_{=0,\text{ replication redundancy}} \\
&= \underbrace{\sum_{\mathcal{C}\subseteq N^R\setminus\{i_1,i_2\}}}_{=N\setminus\{i\}} \frac{1}{2^{|N|+1}} \underbrace{MC_{i_1}(\mathcal{C})}_{=MC_i(\mathcal{C})} = \sum_{\mathcal{C}\subseteq N\setminus\{i\}} \frac{1}{2^{|N|+1}} MC_i(\mathcal{C}),
\end{aligned}$$

where (1) is by grouping the coalitions of players (excluding i_1) into two groups, one group containing all coalitions without i_2 and the other with i_2 in.

By symmetry, $\varphi^R_{i_2}(N^R, v^R) = \varphi^R_{i_1}(N^R, v^R)$, and the total payoff of i in the induced game is:

$$\varphi^R_i = 2\varphi^R_{i_1}(N^R, v^R) = \sum_{\mathcal{C}\subseteq N\setminus\{i\}} \frac{1}{2^{|N|}} MC_i(\mathcal{C}) = \varphi_i(N, v)$$

Therefore, the change in total payoff of player i is zero, i.e.,:

$$\delta\varphi_i^{\text{Banzhaf}} = \varphi^R_i - \varphi_i = 0$$

■

APPENDIX C
PROOF FOR EQUATION (3)

Proof: The derivation from Equation (2) to (3) can be obtained by grouping the marginal contributions of player i towards equal-sized coalitions. The payoff of the player φ_i can be computed as a weighted sum of its average marginal contributions $z_i(c)$ to each coalition size c, where $\binom{|N|-1}{c}$ is the number of size-c coalitions of players excluding i. The detailed steps are as follows:

$$\begin{aligned}
\varphi_i(N, v) &= \sum_{\mathcal{C}\subseteq N\setminus\{i\}} w_{|\mathcal{C}|,N} MC_i(\mathcal{C}) \\
&= \sum_{c=0}^{|N|-1} \sum_{|\mathcal{C}|=c, \mathcal{C}\subseteq N\setminus\{i\}} w_{|\mathcal{C}|,N} MC_i(\mathcal{C}) = \sum_{c=0}^{|N|-1} w_{|\mathcal{C}|,N} \sum_{|\mathcal{C}|=c, \mathcal{C}\subseteq N\setminus\{i\}} MC_i(\mathcal{C}) \\
&= \sum_{c=0}^{|N|-1} \underbrace{\binom{|N|-1}{c} w_{|\mathcal{C}|,N}}_{\alpha_c} \underbrace{\binom{|N|-1}{c}^{-1} \sum_{|\mathcal{C}|=c, \mathcal{C}\subseteq N\setminus\{i\}} MC_i(\mathcal{C})}_{z_i(c)} = \sum_{c=0}^{|N|-1} \alpha_c z_i(c).
\end{aligned}$$

■

APPENDIX D
PROOF FOR LEMMA V-C.1

Lemma V-C.1. *Let $G = (N, v)$ be a submodular game with replication redundant characteristic function v. By replicating k times and acting as $k+1$ players $\mathcal{C}^R = \{i_0, \ldots, i_k\}$ in the induced game $G^R = (N^R, v^R)$, the malicious player i receives a total payoff of*

$$\varphi_i^{\text{tot}}(k) = \sum_{c=0}^{|N|-1} \alpha_c^k z_i(c), \text{ where} \tag{5}$$

$$z_i(c) = \binom{|N|-1}{c}^{-1} \sum_{\mathcal{C}\subseteq N\setminus\{i\}, |\mathcal{C}|=c} MC_i(\mathcal{C}),$$

$$\alpha_c^k = (k+1)\binom{|N|-1}{c} w_{c,N^R} \quad \text{(new importance weights)}.$$

Proof: In the induced game G^R, let $w_{|\mathcal{C}|,N+k}$ be the weights of the solution concept by definition, i.e., $\varphi_i(N^R, v^R) \coloneqq \sum_{\mathcal{C}\subseteq N^R\setminus\{i\}} w_{|\mathcal{C}|,N+k} MC_i(\mathcal{C})$. Then the total payoff of the malicious player after k replications is:

$$\begin{aligned}
\varphi_i^{\text{tot}}(k) &\overset{(1)}{=} (k+1)\varphi_{i_k}(N^R, v^R) \\
&= (k+1) \sum_{\mathcal{C}\subseteq N^R\setminus\{i_k\}} w_{|\mathcal{C}|,|N|+k} MC_{i_k}(\mathcal{C}) \\
&= (k+1) \sum_{\mathcal{C}\subseteq N^R\setminus\mathcal{C}_R} w_{|\mathcal{C}|,|N|+k} MC_{i_k}(\mathcal{C}) + (k+1) \sum_{\mathcal{C}\subseteq N^R\setminus\{i_k\}, \mathcal{C}\cap\mathcal{C}_R\neq\phi} w_{|\mathcal{C}|,|N|+k} \underbrace{MC_{i_k}(\mathcal{C})}_{\overset{(2)}{=}0} \\
&= (k+1) \sum_{S\subseteq N\setminus\{i\}} w_{|\mathcal{C}|,|N|+k} MC_i(\mathcal{C}) \\
&= \sum_{c=0}^{N-1} \underbrace{(k+1)\binom{N-1}{c} w_{|\mathcal{C}|,|N|+k}}_{\alpha_c^k} z_i(c)
\end{aligned}$$

where (1) is due to symmetry and (2) is due to replication-redundancy. ■

APPENDIX E
PROOF FOR THEOREM V-E.1 FOR THE BANZHAF VALUE AND LOO

Theorem V-E.1. *Let $G = (N, v)$ be a submodular game where v is replication redundant, the Shapley value is not replication robust, whereas the Banzhaf value and Leave-one-out are replication robust. For the Shapley value, the total payoff of the replicating player i monotonic increases over the number of replicas, and converges to i's characteristic value, i.e., $\lim_{k\to\infty} \varphi_i^{\text{tot}}(k) = v(\{i\})$. For the Banzhaf and Leave-one-out values, $\lim_{k\to\infty} \varphi_i^{\text{tot}}(k) = 0$.*

Proof: We have proven the robustness properties for the Shapley value in the main text. Here we will provide the proofs for the Banzhaf and LOO values.

(1) For the Banzhaf value. We prove that the importance weights α_c^k satisfy $\forall k \geq 0, \alpha_c^k \geq \alpha_c^{k+1}$ for both the Banzhaf value and Leave-one-out, which is a sufficient condition for $\forall p, \sum_{c=0}^{p} \alpha_c^0 \geq \sum_{c=0}^{p} \alpha_c^k$. Therefore, according to our robustness condition in Theorem V-D.1, both the Banzhaf value and Leave-one-out value are replication robust. In addition, both values monotonic decrease as the number of replications k according to Corollary V-D.1. In particular, for the Banzhaf value,

$$\begin{cases} \alpha_c^k = \frac{(k+1)}{2^{|N|+k-1}}\binom{|N|-1}{c} \\ \alpha_c^{k+1} = \frac{(k+2)}{2^{|N|+k}}\binom{|N|-1}{c} \end{cases} \implies \frac{\alpha_c^k}{\alpha_c^{k+1}} = 2\frac{(k+1)}{(k+2)} \geq 1$$

As we can see, $\alpha_c^0 = \alpha_c^1$ for the Banzhaf value. This implies that the total payoff of the malicious player is unchanged when it replicates for the first time. Therefore, the Banzhaf value is neutral for $k = 1$, which conforms with our finding in Theorem IV-B.1. The limit of the total payoff is:

$$\lim_{k\to\infty} \varphi_i^{\text{tot}}(k) = \lim_{k\to\infty} \sum_{c=0}^{|N|-1} \alpha_c^k z_i(c) = \sum_{c=0}^{|N|-1} \binom{|N|-1}{c} z_i(c) \lim_{k\to\infty} \frac{k+1}{2^{|N|+k-1}} = 0$$

(2) For the Leave-one-out value. $\forall k \geq 0, \frac{\alpha_c^{k+1}}{\alpha_c^k} = 0$, and the total payoff is zero with any positive number of replications, i.e., $\forall k > 0, \varphi_i^{\text{tot}}(k) = 0$. ■

APPENDIX F
PROOFS FOR LEMMA V-E.1

Lemma V-E.1. *For the Shapley value, the importance weights α_c^k of the total payoff of a replicating player satisfy the following properties: $\forall k \geq 0, \forall 0 \leq p \leq |N| - 1$,*

$$\sum_{c=0}^{|N|-1} \alpha_c^k = 1 \tag{9a}$$

$$\sum_{c=0}^{p} \alpha_c^k \leq \sum_{c=0}^{p} \alpha_c^{k+1} \tag{9b}$$

$$\sum_{c=0}^{p} \alpha_c^{k+1} - \alpha_c^k \geq \sum_{c=0}^{p} \alpha_c^{k+2} - \alpha_c^{k+1} \tag{9c}$$

Proof: **Proof for Equation (9a):** Equation (9a) shows that α_c^k always sums to 1 under changing k for the Shapley value.

$$\begin{aligned}
\sum_{c=0}^{|N|-1}\alpha_c^k &= \sum_{c=0}^{|N|-1}\frac{k+1}{|N|+k}\binom{|N|-1}{c}\binom{|N|+k-1}{c}^{-1}, \text{ due to Corollary V-C.1}\\
&= (k+1)\sum_{c=0}^{|N|-1}\frac{(|N|-1)!(|N|+k-1-c)!}{(|N|-1-c)!(|N|+k)!}\\
&= \frac{(k+1)!(|N|-1)!}{(|N|+k)!}\sum_{c=0}^{|N|-1}\frac{(|N|+k-1-c)!}{(|N|-1-c)!k!}\\
&= \frac{1}{\binom{|N|+k}{k+1}}\sum_{c=0}^{|N|-1}\binom{|N|+k-1-c}{k}\\
&\overset{(1)}{=} \frac{1}{\binom{|N|+k}{k+1}}\sum_{i=k}^{|N|-k-1}\binom{i}{k}\\
&\overset{(2)}{=} \frac{1}{\binom{|N|+k}{k+1}}\binom{|N|+k}{k+1}\\
&= 1,
\end{aligned}$$

where (1) is by substituting $i = |N|+k-1-c$ and (2) by the Hockey-Stick identity.

Proof for Equation (9b): This shows that the importance weights α_c^k *shift* to the smaller coalitions under growing k.

$$\begin{aligned}
\sum_{c=0}^{p}\alpha_c^k &= \sum_{c=0}^{p}\frac{(k+1)(|N|-1)!(|N|+k-1-c)!}{(|N|-1-c)!(|N|+k)!}\\
&= \frac{(k+1)!(|N|-1)!}{(|N|+k)!}\sum_{c=0}^{p}\frac{(|N|+k-1-c)!}{(|N|-1-c)!k!}\\
&= \frac{1}{\binom{|N|+k}{k+1}}\sum_{c=0}^{p}\binom{|N|+k-1-c}{k}\\
&= \frac{1}{\binom{|N|+k}{k+1}}\Big(\sum_{c=0}^{|N|-1}-\sum_{c=p+1}^{|N|-1}\binom{|N|+k-1-c}{k}\Big)\\
&\overset{(1)}{=} 1-\frac{1}{\binom{|N|+k}{k+1}}\sum_{c=p+1}^{|N|-1}\binom{|N|+k-1-c}{k}\\
&\overset{(2)}{=} 1-\frac{1}{\binom{|N|+k}{k+1}}\binom{|N|+k-p-1}{k+1}\\
&= 1-\frac{(|N|-1)!}{(|N|-p-2)!}\frac{1}{(|N|+k)...(|N|+k-p)},
\end{aligned}$$

where (1) is by Equation (9a) and (2) is by the Hockey-Stick identity. Similarly,

$$\sum_{c=0}^{p}\alpha_c^{k+1} = 1-\frac{(|N|-1)!}{(|N|-p-2)!}\frac{1}{(|N|+k+1)...(|N|+k+1-p)}$$

Therefore,

$$\begin{aligned}
\sum_{c=0}^{p}\alpha_c^{k+1}-\sum_{c=0}^{p}\alpha_c^{k} &= \frac{(|N|-1)!}{(|N|-p-2)!}\frac{(|N|+k+1)-(|N|+k-p)}{(|N|+k+1)...(|N|+k-p)}\\
&= \frac{(|N|-1)!}{(|N|-p-2)!}\frac{p+1}{(|N|+k+1)...(|N|+k-p)} \geq 0
\end{aligned}$$

Proof for Equation (9c): With this additional condition, the *unit gain* of the total payoff for adding a replica decreases monotonically for each replication. This means that the player obtains the most unit gain for the first replication. To show this property, we denote for any k, denote $\delta^k := \sum_{c=0}^{p}\alpha_c^{k+1}-\sum_{c=0}^{p}\alpha_c^{k}$. From the proof of Equation (9b): $\delta^k = \frac{(|N|-1)!}{(|N|-p-2)!}\frac{p+1}{(|N|+k+1)...(|N|+k-p)}$, therefore,

$$
\begin{aligned}
RHS - LHS \text{ of (9c)} &= \delta^{k+1} - \delta^k \\
&= \frac{(|N|-1)!(p+1)}{(|N|-p-2)!}\left(\frac{1}{(|N|+k+2)...(|N|+k+1-p)} - \frac{1}{(|N|+k+1)...(|N|+k-p)}\right) \\
&= \frac{(|N|-1)!(p+1)}{(|N|-p-2)!}\left(\frac{(|N|+k-p)-(|N|+k+2)}{(|N|+k+2)...(|N|+k-p)}\right) \\
&= \frac{(|N|-1)!(p+1)}{(|N|-p-2)!}\frac{-(p+2)}{(|N|+k+2)...(|N|+k-p)} \leq 0
\end{aligned}
$$

■

APPENDIX G
PROOF FOR COROLLARY V-F.1

Corollary V-F.1. *The Robust Shapley value is replication robust. Moreover, in a submodular game $G = (N, v)$, the loss for a replicating player i by replicating k times $\varphi_i^{\text{tot}}(0) - \varphi_i^{\text{tot}}(k) \geq \frac{1}{|N|}\sum_{c=0}^{|N|-1}(1 - \frac{k+1}{2^k})\gamma_{|N|}^c z_i(c)$.*

Proof: **Replication-robustness** We prove that similar to the Banzhaf value, the Robust Shapley value satisfies Equation (10) in Observation 1: $\forall k \geq 0, \frac{\alpha_c^k}{\alpha_c^{k+1}} \geq 1$. Hence it satisfies Theorem V-D.1, and therefore sufficient for replication robustness. There are 3 possible cases:

Case 1: $c < \lfloor\frac{|N|+k-1}{2}\rfloor \leq \lfloor\frac{|N|+k}{2}\rfloor$

In this case, both $\tilde{\alpha}_c^k$ and $\tilde{\alpha}_c^{k+1}$ will be down-weighed from the Shapley coefficients where $\gamma_{|N|+k}^c = \frac{\lceil\frac{|N|+k-1}{2}\rceil!\lfloor\frac{|N|+k-1}{2}\rfloor!}{c!(|N|+k-c-1)!}$:

$$
\begin{aligned}
\tilde{\alpha}_c^k &= \gamma_{|N|+k}^c \alpha_c^k = (k+1)\binom{|N|-1}{c}\frac{\lfloor\frac{|N|+k-1}{2}\rfloor!\lceil\frac{|N|+k-1}{2}\rceil!}{(|N|+k)!} \\
\tilde{\alpha}_c^{k+1} &= \gamma_{|N|+k+1}^c \alpha_c^{k+1} = (k+2)\binom{|N|-1}{c}\frac{\lfloor\frac{|N|+k}{2}\rfloor!\lceil\frac{|N|+k}{2}\rceil!}{(|N|+k+1)!} \\
\text{Hence } \frac{\tilde{\alpha}_c^k}{\tilde{\alpha}_c^{k+1}} &= \frac{k+1}{k+2}\frac{|N|+k+1}{\lceil\frac{|N|+k}{2}\rceil} \geq \frac{1}{2} * 2 = 1.
\end{aligned}
$$

Case 2: $c \geq \lfloor\frac{|N|+k}{2}\rfloor \geq \lfloor\frac{|N|+k-1}{2}\rfloor$

Both $\tilde{\alpha}_c^k$, $\tilde{\alpha}_c^{k+1}$ take the original form of Shapley coefficients after replication, i.e, $\gamma_{|N|}^c = 1$:

$$
\begin{aligned}
\tilde{\alpha}_c^k &= \alpha_c^k = (k+1)\binom{|N|-1}{c}\frac{c!(|N|+k-1-c)!}{(|N|+k)!} \\
\tilde{\alpha}_c^{k+1} &= \alpha_c^{k+1} = (k+2)\binom{|N|-1}{c}\frac{c!(|N|+k-c)!}{(|N|+k+1)!} \\
\frac{\tilde{\alpha}_c^k}{\tilde{\alpha}_c^{k+1}} &= \frac{k+1}{k+2}\frac{|N|+k+1}{|N|+k-c} \overset{(1)}{\geq} 2\frac{k+1}{k+2} \geq 1, \qquad \text{where (1) is due to } c \geq \lfloor\tfrac{|N|+k}{2}\rfloor.
\end{aligned}
$$

Case 3: $\lfloor\frac{|N|+k-1}{2}\rfloor \leq c < \lfloor\frac{|N|+k}{2}\rfloor$

In this case, $\tilde{\alpha}_c^k$ will take the original form, while $\tilde{\alpha}_c^{k+1}$ will take the down-weighed form. Moreover, $|N|+k$ must be even, hence $c = \lfloor\frac{|N|+k-1}{2}\rfloor$.

$$
\begin{aligned}
\tilde{\alpha}_c^k &= \alpha_c^k = (k+1)\binom{|N|-1}{c}\frac{c!(|N|+k-1-c)!}{(|N|+k)!} = (k+1)\binom{|N|-1}{c}\frac{\lfloor\frac{|N|+k-1}{2}\rfloor!\lceil\frac{|N|+k-1}{2}\rceil!}{(|N|+k)!} \\
\tilde{\alpha}_c^{k+1} &= \gamma_{|N|+k+1}^c \alpha_c^{k+1} = (k+2)\binom{|N|-1}{c}\frac{\lfloor\frac{|N|+k}{2}\rfloor!\lceil\frac{|N|+k}{2}\rceil!}{(|N|+k+1)!} \\
\text{Hence } \frac{\tilde{\alpha}_c^k}{\tilde{\alpha}_c^{k+1}} &= \frac{k+1}{k+2}\frac{|N|+k+1}{\lceil\frac{|N|+k}{2}\rceil} \geq 2\frac{k+1}{k+2} \geq 1
\end{aligned}
$$

We have shown that $\forall k \geq 0, \frac{\alpha_c^k}{\alpha_c^{k+1}} \geq 1$, and hence the Robust Shapley value is replication-robust.

Payoff loss Note that from the above derivations, in all 3 cases, $\forall k \geq 0, \frac{k+2}{k+1}\frac{\tilde{\alpha}_c^k}{\tilde{\alpha}_c^{k+1}} \geq 2$:

$$
\begin{aligned}
\varphi_i^{\text{tot}}(0) &= \sum_{c=0}^{|N|-1} \tilde{\alpha}_c^0 z_i(c) \coloneqq \frac{1}{|N|} \sum_{c=0}^{|N|-1} \gamma_{|}N|^c z_i(c) \\
\varphi_i^{\text{tot}}(k) &= \sum_{c=0}^{|N|-1} \tilde{\alpha}_c^k z_i(c) \\
&= (k+1) \sum_{c=0}^{|N|-1} \frac{\tilde{\alpha}_c^k}{k+1} z_i(c) = (k+1) \sum_{c=0}^{|N|-1} (\frac{\tilde{\alpha}_c^{k-1}}{k} \frac{k}{\tilde{\alpha}_c^{k-1}}) \frac{\tilde{\alpha}_c^k}{k+1} z_i(c) \\
&= (k+1) \sum_{c=0}^{|N|-1} \frac{\tilde{\alpha}_c^{k-1}}{k} \underbrace{\frac{k}{\tilde{\alpha}_c^{k-1}} \frac{\tilde{\alpha}_c^k}{k+1}}_{\leq \frac{1}{2} \text{ as } \forall k>0, \frac{\tilde{\alpha}_c^k/(k+1)}{\tilde{\alpha}_c^{k-1}/(k)} \leq \frac{1}{2}} z_i(c), \\
&\leq (k+1) \sum_{c=0}^{|N|-1} \frac{1}{2} \frac{\tilde{\alpha}_c^{k-1}}{k} z_i(c) \leq \ldots \\
&\leq (k+1) \sum_{c=0}^{|N|-1} \frac{1}{2^k} \tilde{\alpha}_c^0 z_i(c) \\
&= (\frac{k+1}{2^k}) \frac{1}{|N|} \sum_{c=0}^{|N|-1} \gamma_{|N|}^c z_i(c) \\
\textit{Hence } \varphi_i^{\text{tot}}(0) - \varphi_i^{\text{tot}}(k) &\geq \frac{1}{|N|} \sum_{c=0}^{|N|-1} (1 - \tfrac{k+1}{2^k}) \gamma_{|N|}^c z_i(c).
\end{aligned}
$$

This concludes our proof for Corollary V-F.1. ∎

APPENDIX H
PROOFS FOR LEMMA V-G.1

Lemma V-G.1. *Compared with replication, the additional gain in total payoff of the malicious player due to the perturbation when replicating k times is given by:*

$$\varphi_i^{\text{perturb}} - \varphi_i^{\text{replicate}} \leq (k+1)\epsilon.$$

Proof: Compared with replication, the additional gain in payoff due to the perturbation is

$$
\begin{aligned}
\varphi_i^{\text{perturb}} - \varphi_i^{\text{replicate}} &= \sum_{p_k \in \mathcal{C}^P} \sum_{\mathcal{C} \subseteq N \setminus \{i\}, \mathcal{C}^p \subseteq \mathcal{C}^P \setminus \{p_k\}} w_{|\mathcal{C} \cup \mathcal{C}^p|, |N|+k} MC_{p_k}(\mathcal{C} \cup \mathcal{C}^p) - \\
&\qquad \sum_{i_k \in \mathcal{C}^R} \sum_{\mathcal{C} \subseteq N \setminus \{i\}, \mathcal{C}^r \subseteq \mathcal{C}^R \setminus \{i_k\}} w_{|\mathcal{C} \cup \mathcal{C}^r|, |N|+k} MC_{i_k}(\mathcal{C} \cup \mathcal{C}^r) \\
&= \sum_{k=0}^{|\mathcal{C}^R|-1} \sum_{\mathcal{C} \subseteq N \setminus \{i\}} w_{|\mathcal{C}|, |N|+k} (MC_{p_k}(\mathcal{C}) - MC_{i_k}(\mathcal{C})) + \\
&\qquad \sum_{p_k \in \mathcal{C}^P} \sum_{\mathcal{C} \subseteq N \setminus \{i\}, \mathcal{C}^p \subseteq_{\neq \emptyset} \mathcal{C}^P \setminus \{p_k\}} w_{|\mathcal{C} \cup \mathcal{C}^p|, |N|+k} MC_{p_k}(\mathcal{C} \cup \mathcal{C}^p) \\
&= \sum_{p_k \in \mathcal{C}^P} \sum_{\mathcal{C} \subseteq N \setminus \{i\}, \mathcal{C}^p \subseteq_{\neq \emptyset} \mathcal{C}^P \setminus \{p_k\}} w_{|\mathcal{C} \cup \mathcal{C}^p|, |N|+k} MC_{p_k}(\mathcal{C} \cup \mathcal{C}^p) \\
&\overset{(1)}{\leq} (k+1) \sum_{\mathcal{C} \subseteq N \setminus \{i\}, \mathcal{C}^p \subseteq_{\neq \emptyset} \mathcal{C}^P \setminus \{p_k\}} w_{|\mathcal{C} \cup \mathcal{C}^p|, |N|+k} \epsilon \\
&\overset{(2)}{\leq} (k+1)\epsilon,
\end{aligned}
$$

where (1) is due to the assumption on ϵ, where $\forall \emptyset \neq \mathcal{C}^p \subseteq \mathcal{C}^p \setminus \{p_k\}, \mathcal{C} \subseteq N \setminus \{i\}, MC_{p_k}(\mathcal{C} \cup \mathcal{C}^p) \leq \epsilon$. (2) is due to the definition of semivalues where the weights of coalitions sum to 1. ∎

APPENDIX I
ALGORITHM 1

Algorithm 1 Efficient Shapley and Banzhaf value Computation for the Facility Location Game

Input: Locations $\mathcal{L}$, customers D, utility matrix U
Output: Shapley and Banzhaf value of all locations

1: Sort the facility locations by (ascending) utility for each customer d, where $U^{d\uparrow}$ is the sorted utility vector of customer d and $\mathcal{L}^{d\uparrow}$ are the sorted facility locations.
2: **for** each location $i \in \mathcal{L}$ **do**
3: $\quad l_i^d \leftarrow$ index of i in $\mathcal{L}^{d\uparrow}$, i.e., $l_i^d = |\mathcal{L}_{id}| - 1$
4: $\quad \varphi_i^{\text{Shapley}} \leftarrow \sum_{d\in D}[\frac{U_{id}}{n-l_i^d+1} - \sum_{t=0}^{l_{id}} \frac{U^{d\uparrow}_{l_i^d-t-1}}{(n-l_i^d+t)+(n-l_i^d+t)^2}]$
5: $\quad \varphi_i^{\text{Banzhaf}} \leftarrow \frac{1}{2^{|\mathcal{L}|-1}} \sum_{d\in D}[2^{|l_i^d|+1}U_{id} - \sum_{t=0}^{l_i^d} U^{d\uparrow}(l_i^d - t + 1)]$
6: **end for**

APPENDIX J
PROOFS FOR THEOREM V-H.1

Before deriving the Shapley and Banzhaf value for the facility location game, we first need to show the following mathematical identity which will be used for the derivation.

Lemma J-.1.

$$\sum_{k=0}^{m} \frac{\binom{m}{k}}{\binom{n}{k}} = \frac{n+1}{n+1-m} \tag{12}$$

Proof Sketch.: The identity can be shown in two steps: First, we show the identity $\frac{\binom{m}{k}}{\binom{n}{k}} = \frac{\binom{n-k}{m-k}}{\binom{n}{m}}$ by expansion of the terms. Then, we can take the denominator $\binom{n}{m}$ out of the summation over k, and as a common mathematical identity, the sum reduces to $\sum_{k=0}^{m}\binom{n-k}{m-k} = \binom{n+1}{m}$. Finally, by expanding the terms we arrive at $\frac{\binom{n+1}{m}}{\binom{n}{m}} = \frac{n+1}{n+1-m}$. ■

Theorem V-H.1. *The Shapley and Banzhaf value of a facility location i in a facility location game can be computed as*

$$\varphi_i^{\text{Shapley}} = \sum_{d\in D}\big[u_{id}\frac{1}{|\mathcal{L}|-|\mathcal{L}_{id}|} - \sum_{t=1}^{|\mathcal{L}_{id}|}\frac{1}{\lambda(t)+\lambda(t)^2}u_{e_{it}^d}\big],$$
$$\varphi_i^{\text{Banzhaf}} = \frac{1}{2^{|\mathcal{L}|-1}}\sum_{d\in D}\big[2^{|\mathcal{L}_{id}|}u_{id} - \sum_{t=1}^{|\mathcal{L}_{id}|}2^{|\mathcal{L}_{id}|-t}u_{e_{it}^d}\big],$$

where $\lambda(t) \coloneqq (|\mathcal{L}| - |\mathcal{L}_{id}| + t - 1)$, $\mathcal{L}_{id} \coloneqq \{j \in \mathcal{L} \mid u_{jd} \le u_{id}\}$ and $u_{e_{it}^d}$ is the utility value of the t-th largest element after i along the dimension (customer) d, D is the set of customers, u_{id} is the utility of a customer d from facility location i.

Proof: **(1) Proof for the Shapley value.**

Let $v(\mathcal{C}) \coloneqq Fac(\mathcal{C})$ and $n \coloneqq |\mathcal{L}|$ as the number of players. Denote $w_\mathcal{C}$ as the weights of the Shapley value, i.e., $\varphi_i^{\text{Shapley}} = \sum_{\mathcal{C}\subseteq\mathcal{L}\setminus\{i\}} w_\mathcal{C} MC_i(\mathcal{C})$, where $w_\mathcal{C} \coloneqq \frac{1}{n}\binom{n-1}{|\mathcal{C}|}^{-1}$. Observe that

$$\varphi_i = \sum_{\mathcal{C}\subseteq\mathcal{L}\setminus\{i\}} w_\mathcal{C} MC_i(\mathcal{C}) \overset{(*)}{=} \sum_{d\in D}\big[\underbrace{\sum_{\mathcal{C}\subseteq\mathcal{L}_{id}} w_\mathcal{C} u_{id}}_{(\#1)} - \underbrace{\sum_{\mathcal{C}\subseteq\mathcal{L}_{id}} w_\mathcal{C} \max_{j\in\mathcal{C}} u_{jd}}_{(\#2)}\big],$$

where $(*)$ is because the marginal contribution of i for dimension d is zero *unless* i is the largest element for that dimension and $\mathcal{L}_{id} = \{j \in \mathcal{L} \mid u_{jd} \le u_{id}\}$ is the coalition of all elements which have smaller values in the d-th dimension than element i.

Along each dimension d, $(\#1)$ is a weighted sum over coalitions $\mathcal{C}$ where i is the largest element; and $(\#2)$ sums up for each $j \in \mathcal{L}_{id}$ over all coalitions $\mathcal{C} \subseteq \mathcal{L}_{id}$ where j is the largest element. We next compute $(\#1)$ and $(\#2)$ separately for each

dimension $d \in D$:

$$\begin{aligned}
(\#1) &= \sum_{\mathcal{C}\subseteq\mathcal{L}_{id}} w_{\mathcal{C}} u_{id} = u_{id} \sum_{\mathcal{C}\subseteq\mathcal{L}_{id}} w_{\mathcal{C}} \\
&= u_{id} \sum_{c=0}^{|\mathcal{L}_{id}|} \sum_{\mathcal{C}\subseteq\mathcal{L}_{id}, |\mathcal{C}|=c} w_{\mathcal{C}} && \text{, where } w_{\mathcal{C}} := \frac{1}{n}\binom{n-1}{c}^{-1} \\
&= u_{id} \frac{1}{n} \sum_{c=0}^{|\mathcal{L}_{id}|} \binom{n-1}{c}^{-1} \binom{|\mathcal{L}_{id}|}{c} \\
&\overset{(1)}{=} u_{id} \frac{1}{n} \frac{n}{n-|\mathcal{L}_{id}|} && \text{, (1) by Lemma J-.1 } \sum_{k=0}^{m} \frac{\binom{m}{k}}{\binom{n}{k}} = \frac{n+1}{n+1-m} \\
&= u_{id} \frac{1}{n-|\mathcal{L}_{id}|},
\end{aligned}$$

Next we compute $(\#2)$. For simplicity, let $+, -$ denote set operations $\mathcal{C} \cup \{e\}, \mathcal{C} \setminus \{e\}$, and denote e^d_{it} is t-th largest element (after element i) in the d-th dimension.

$$\begin{aligned}
(\#2) &= \sum_{\mathcal{C}\subseteq\mathcal{L}_{id}} w_{\mathcal{C}} \max_{j\in\mathcal{C}} u_{jd} \\
&= \sum_{\mathcal{C}\subseteq\mathcal{L}_{id}-(e^d_{i1})} w_{\mathcal{C}+e^d_{i1}} u_{e^d_{i1}d} + \sum_{\mathcal{C}\subseteq\mathcal{L}_{id}-(e^d_{i1}+e^d_{i2})} w_{\mathcal{C}+e^d_{i2}} u_{e^d_{i2}d} + \cdots \\
&= u_{e^d_{i1}d} \sum_{\mathcal{C}\subseteq\mathcal{L}_{id}-(e^d_{i1})} w_{\mathcal{C}+e^d_{i1}} + u_{e^d_{i2}d} \sum_{\mathcal{C}\subseteq\mathcal{L}_{id}-(e^d_{i1}+e^d_{i2})} w_{\mathcal{C}+e^d_{i2}} + \cdots \\
&= u_{e^d_{i1}d} \underbrace{\sum_{\mathcal{C}\subseteq\mathcal{L}_{id}-(e^d_{i1})} w_{\mathcal{C}+e^d_{i1}}}_{=:\beta_1} + u_{e^d_{i2}d} \underbrace{\sum_{\mathcal{C}\subseteq\mathcal{L}_{id}-(e^d_{i1}+e^d_{i2})} w_{\mathcal{C}+e^d_{i2}}}_{=:\beta_2} + \cdots
\end{aligned}$$

$$\begin{aligned}
\text{In particular, } \beta_t &= \sum_{\mathcal{C}\subseteq\mathcal{L}_{id}-(e^d_{i1}+\ldots+e^d_{it})} w_{\mathcal{C}+e^d_{it}} \\
&= \sum_{c=0}^{|\mathcal{L}_{id}|-t} \sum_{\mathcal{C}\subseteq\mathcal{L}_{id}-(e^d_{i1}+\ldots+e^d_{it}), |\mathcal{C}|=c} w_{\mathcal{C}+e^d_{it}} \\
&= \frac{1}{n} \sum_{c=0}^{|\mathcal{L}_{id}|-t} \binom{n-1}{c+1}^{-1} \binom{|\mathcal{L}_{id}|-t}{c} \\
&\overset{(1)}{=} \frac{1}{n} \sum_{c=0}^{|\mathcal{L}_{id}|-t} \binom{n-1}{c+1}^{-1} \Big[\binom{|\mathcal{L}_{id}|-t+1}{c+1} - \binom{|\mathcal{L}_{id}|-t}{c+1}\Big] \\
&\overset{(2)}{=} \frac{1}{n} \sum_{x=1}^{|\mathcal{L}_{id}|-t+1} \binom{n-1}{x}^{-1} \Big[\binom{|\mathcal{L}_{id}|-t+1}{x} - \binom{|\mathcal{L}_{id}|-t}{x}\Big] \\
&\overset{(3)}{=} \frac{1}{n} \Big[\frac{n}{n-|\mathcal{L}_{id}|+t-1} - 1 - \frac{n}{n-|\mathcal{L}_{id}|+t} + 1\Big] \\
&= \frac{1}{\lambda(t)+\lambda(t)^2},
\end{aligned}$$

where (1) is by Pascal's identity, (2) by substituting $x = c+1$, (3) by Lemma J-.1 and observing that $\binom{n}{k}$ is zero for $k > n$, and where $\lambda(t) = n - |\mathcal{L}_{id}| + t - 1$.

$$\text{Hence, } \varphi_i = \sum_{d\in D} \Big[u_{id} \frac{1}{n-|\mathcal{L}_{id}|} - \sum_{t=1}^{|\mathcal{L}_{id}|} \frac{1}{\lambda(t)+\lambda(t)^2} u_{e^d_{it}}\Big].$$

(2) Proof for the Banzhaf value.

Similar to the proof for the Shapley value, we expand the Banzhaf value as follows:

$$\varphi(i) = \sum_{\mathcal{C}\subseteq\mathcal{L}-i} w_{\mathcal{C}} MC_i(\mathcal{C}) = \sum_{d\in D} \big[\underbrace{\sum_{\mathcal{C}\subseteq\mathcal{L}_{id}} w_{\mathcal{C}} u_{id}}_{(\#1)} - \underbrace{\sum_{\mathcal{C}\subseteq\mathcal{L}_{id}} w_{\mathcal{C}} \max_{j\in\mathcal{C}} u_{jd}}_{(\#2)} \big].$$

By definition of the Banzhaf value $w_{\mathcal{C}} \coloneqq \frac{1}{2^{n-1}}$, next we compute $\#1$ and $\#2$.

$$\begin{aligned}
(\#1) &= \sum_{d\in D} \sum_{\mathcal{C}\subseteq\mathcal{L}_{id}} w_{\mathcal{C}} u_{id} \\
&= \sum_{d\in D} u_{id} \sum_{\mathcal{C}\subseteq\mathcal{L}_{id}} w_{\mathcal{C}} \\
&= \sum_{d\in D} u_{id} \sum_{c=0}^{|\mathcal{L}_{id}|} \sum_{\mathcal{C}\subseteq\mathcal{L}_{id},|\mathcal{C}|=c} w_{\mathcal{C}} \\
&= \sum_{d\in D} u_{id} \frac{1}{2^{n-1}} \sum_{c=0}^{|\mathcal{L}_{id}|} \binom{|\mathcal{L}_{id}|}{c} \\
&= \frac{1}{2^{n-1}} \sum_{d\in D} 2^{|\mathcal{L}_{id}|} u_{id}
\end{aligned}$$

We then expand $(\#2)$ in a similar approach to the Shapley value (for notations c.f. above theorem),

$$(\#2) = \sum_{\mathcal{C}\subseteq\mathcal{L}_{id}} w_{\mathcal{C}} \max_{j\in\mathcal{C}} u_{jd} = u_{e^d_{i1}d} \underbrace{\sum_{\mathcal{C}\subseteq\mathcal{L}_{id}-(e^d_{i1})} w_{\mathcal{C}+e^d_{i1}}}_{=:\beta_1} + u_{e^d_{i2}d} \underbrace{\sum_{\mathcal{C}\subseteq\mathcal{L}_{id}-(e^d_{i1}+e^d_{i2})} w_{\mathcal{C}+e^d_{i2}}}_{=:\beta_2} + \cdots$$

$$\begin{aligned}
\text{where } \beta_t &= \sum_{\mathcal{C}\subseteq\mathcal{L}_{id}-(e^d_{i1}+\ldots+e^d_{it})} w_{\mathcal{C}+e^d_{it}} \\
&= \sum_{c=0}^{|\mathcal{L}_{id}|-t} \sum_{\mathcal{C}\subseteq\mathcal{L}_{id}-(e^d_{i1}+\ldots+e^d_{it}),|\mathcal{C}|=c} w_{\mathcal{C}+e^d_{it}} \\
&= \frac{1}{2^{n-1}} \sum_{c=0}^{|\mathcal{L}_{id}|-t} \binom{|\mathcal{L}_{id}|-t}{c} \\
&= \frac{1}{2^{n-1}} 2^{|\mathcal{L}_{id}|-t}
\end{aligned}$$

$$\text{Hence, } \varphi_i = \frac{1}{2^{n-1}} \sum_{d\in D} \big[2^{|\mathcal{L}_{id}|} u_{id} - \sum_{t=1}^{|\mathcal{L}_{id}|} 2^{|\mathcal{L}_{id}|-t} u_{e^d_{it}} \big].$$

■