

# Optimality and Proportionality in Online Friends Partitioning under Uncertainty

Paper ID # 1884

## Abstract

We study the *friendship-based online coalition formation* problem, in which agents that appear one at a time should be partitioned into coalitions, and an agent's utility for a coalition is the number of her neighbors (i.e., *friends*) within the coalition. Unlike prior work, agents' friendships may be *uncertain*. We analyze the desirability of the resulting partition in the common term of *optimality*, aiming to maximize the social welfare, and the novel *proportionality* criterion, aiming to proportionally reflect agents' friendships in the resulting partition. For optimizing social welfare, we first design an online algorithm termed *Maximum Predicted Coalitional Friends* (MPCF), which is enhanced with predictions of each agent's number of friends within any possible coalition. For common classes of random graphs, we prove that MPCF is optimal, and, for certain graphs, provides the *same* guarantee as the best known competitive algorithm for settings *without* uncertainty. We then introduce the notion of *proportionality* from multi-winner elections into hedonic games by devising *coalitional* variants of *Proportional Justified Representation* and *Extended Justified Representation* called *CPJR* and *CEJR* (resp.). We prove a strong positive result for the case of uncertainty: *For a natural random graph model, MPCF is optimal in terms of both social welfare as well as guaranteeing CPJR and CEJR*. When friendships are *certain* this is not necessarily the case, but we show that CPJR can be easily attained under mild assumptions, while CEJR can be approximated.

## Introduction

Your employers have reserved the best banquet venue of the city for a company event. When you arrive, you see multiple tables of various capacities that have been set to accommodate the participants, yet some seats are already taken. You would like to share your table with some friends from work, while avoiding colleagues you do not like that much. As such, your managers would like to ensure that each employee is assigned to a table with as many of his friends as possible, so he would have a good experience during the event. In each table, the organizers also desire that every sufficiently large group of participants with similar preferences would have common friends in a manner proportional to their size. A similar scenario was considered by Bilò, Monaco, and Moscardell (?). Additional real-life examples

include academic research (?) and international unions (?), where *agents* perform activities in *coalitions* rather than on their own. Such situations fall within the phenomenon of coalition formation, which is noticed in our social, economic, and politic life.

In this paper, we introduce and study a model capturing such real-life scenarios. A popular framework for studying coalition formation is that of *hedonic games* (?), which disregards externalities, i.e., agents utilities solely depend on the coalition they are part of. The outcome of such games is a set of disjoint coalitions (hereafter, *partition*). In our model, agents' friendships may be encoded by an unweighted and undirected graph and an agent's utility for a coalition is the number of her neighbors (i.e., *friends*) within the coalition, that is, the agent's *degree* in the graph induced by the coalition. As in our example, there are many contexts where it is more realistic to assume that coalitions may have a limited *physical* space and agents arrive over time. Therefore, we assume an upper bound on the size of each coalition. Then, a central authority (i.e., an *online* algorithm) has to *immediately* and *irrevocably* decide whether to add an arriving agent into an existing coalition or to create a new one containing, at this moment, only her.

Unlike prior work which assumes that an arriving agent reveals her exact friendships with previously disclosed agents (?), we consider scenarios where friendships are *uncertain*, and design an online algorithm augmented with a *coalitional friends predictor*, i.e., an oracle that predicts an agent's number of friends within a given coalition. Not only that such a predictor is simple and easy to interpret, it is also useful: Vertices' degree information has been previously employed in heuristic and approximation algorithms for other graph problems (e.g., maximum independent set (?)). Since an agent's degree is simply her frequency in the union of all edges, a predictor can also be readily attained via estimation methods of elements' frequencies in a dataset, as used for certain data analysis problems (?).

The quality of a partition has been measured by various solution concepts, e.g., stability and optimality (?). In this paper, we first explore the objective of *maximizing the (utilitarian) social welfare*. We study a simple algorithm called *Maximum Predicted Coalitional Friends* (MPCF), which assigns agents to coalitions greedily with respect to the predictor. For general graphs, MPCF returns a partition with a

high social welfare even when the predictions are slightly off. For illustrating MPCFs good performance, we follow a vast trend of research on average-case analysis (see, e.g., (??)), and analyze MPCF also under a very natural and common random graph model: the Chung-Lu-Vu (CLV) model (?). When the expected degree of each agent in the subgraph induced by any coalition is used as a prediction, we prove that MPCF stochastically dominates *any* other algorithm for graphs drawn from the CLV model and analyze its expected social welfare. For *deterministic* graphs, we show that it has the *same* competitive ratio as the best known algorithm by Flammini et al. (?).

Then, we explore *proportionality*, which states that agents friendships should be reflected proportionately in the resulting partition. That is, as existing proportionality axioms in elections cannot be directly applied, we initiate the study of lifting them to ASHG encoded by *unweighted* and *undirected* graphs. We devise *coalitional* variants of *Proportional Justified Representation* (?) and *Extended Justified Representation* (?) called *CPJR* and *CEJR* (resp.). Uncertainty about friendships allows us to supply a strong positive result: *Under the CLV model, MPCF is optimal in terms of both social welfare as well as guaranteeing CPJR and CEJR*. When friendships are *certain* this may not be the case, but we show that CPJR can be easily satisfied in polynomial-time, while CEJR is unsatisfied in online settings, yet it can be approximated. After showing that the best possible approximation is computationally hard to attain, we devise a polynomial-time scheme satisfying a slightly worse CEJR guarantee.

## Related Work

Our work can be viewed as *additively separable hedonic games* (ASHGs) (?) with symmetric and binary preferences under the restriction that the size of each coalition is bounded. In such classes, agents' preferences are encoded by means of a weighted and possibly directed graph from which cardinal utilities are extracted, depicting each agent's valuation of any other agent. A coalition is then evaluated via the *summation* of its members' valuations of others within the coalition. Hedonic games have been introduced by Dreze and Greenberg (?), and later expanded to the study of various solutions concepts such as stability, fairness, and optimality (see, e.g., (??)). One major concern is designing computationally manageable classes of hedonic games, which led to an abundance of game representations. Some are *ordinal* and can *fully* express every preference over coalitions (??), yet they may require exponential space. In contrast, *cardinal* hedonic games, based on weighted graphs (??), are *not* fully expressive but only require *polynomial* space for reasonable weights. Apart from ASHG, averaging the weights within each coalition yields the classes of fractional hedonic games (?) and modified fractional hedonic games (?). ASHG also contain friend-oriented hedonic games (?) that distinguish between friends and enemies by allowing for only two possible weights.

Partitions in hedonic games are typically measured in terms of stability and optimality. While in (??) properties guaranteeing the existence of stable partitions in ASHG

were supplied, their computational aspects were studied in (??). Yet, in our work we focus on maximizing social welfare, as studied in (??). (??) also explore welfare-optimal partitions, as well as Pareto-optimality. In *online* settings, Flammini et al. (?) study scenarios similar to standard *symmetric* ASHG, where agents arrive along with their incident edges. In contrast, we aim at maximizing social welfare in *online* hedonic games when agents' friendships (and thus their utilities) are *uncertain* (i.e., edges are *not* revealed). Unlike Flammini et al. (?), we focus not only on maximizing welfare, but also introduce *proportionality* into hedonic games, which is a well-known topic in elections (?).

Our work is also closely related to a recently popular trend of augmenting online algorithms with predictions, with the aim of bypassing the worst-case lower bounds of online problems caused by the uncertainty of the future. Though the idea of using advice to obtain semi-online algorithms is not new (?), Munoz and Vassilvitskii (?) propose to use a predictor oracle for improving revenue optimization in auctions by setting a good reserve (or minimum) price, and Lykouris and Vassilvitskii (?) consider the online caching problem with predictions. These works led to a series of learning augmented results in various fields (e.g., clustering (?), ski-rental (?)). To the best of our knowledge, we are the first to introduce predictions to online coalition formation problems, including a scheme that is stochastically optimal for realistic and natural random graphs.

Similarly to a recent work on online bipartite matching by Aamand et al. (?), we also use predictions derived from agents' expected degrees. However, in their work agents arrive along with their incident edges, while in ours edges may not be revealed. Further, while they use an oracle that predicts an agent's degree *in the full graph*, the oracle used by our MPCF algorithm predicts an agent's degree *within the subgraph induced by a given coalition*. Their algorithm greedily matches an arriving agent with a *minimum* predicted degree neighbor that is yet to be covered. In contrast, MPCF assigns an arriving agent to a coalition with *maximum* predicted number of friends that has the *minimum* total expected degree among such coalitions. In other graph problems, other kinds of predictions have also been used (See, e.g., (??)).

We also investigate *proportional representation* in hedonic games, which is a central desiderata in elections (?). Most proportionality axioms in elections usually require that each sufficiently large group of agents is represented in the committee. For instance, extended justified representation (?) enforces that there exists at least one voter approving a certain number of committee members, whereas proportional justified representation (?) demands that there are sufficiently many committee members that are each approved by at least one voter. Stronger notions were later proposed, e.g., (?). Though proportionality axioms have been also studied in online committee elections (??), as far as we know, prior research did not regard such axioms as a fairness criteria in ASHG (See, e.g., (?)), let alone in online settings.

## Online Partitioning of Friends

We consider the problem of partitioning a finite set  $N = \{1, \dots, n\}$  of  $n$  agents within an undirected social network  $G = (N, E)$  into coalitions, where an agent benefits from the arrival of another agent into her coalition so long as they are neighbors (i.e., friends). Agent  $i$ 's preferences can thus be succinctly represented by a (binary) cardinal utility function  $v_i : N \rightarrow \{0, 1\}$  with  $v_i(i) = 0$  that satisfies  $v_i(j) = 1$  if agent  $j \neq i$  is a friend of agent  $i$  (i.e.,  $(i, j) \in E$ ); otherwise,  $v_i(j) = 0$ . We denote by  $\mathcal{N}_i$  the set of coalitions agent  $i$  belongs to, i.e.,  $\mathcal{N}_i = \{C \subseteq N : i \in C\}$ . Agent  $i$ 's utility can be *additively* aggregated to preferences over each coalition  $C \in \mathcal{N}_i$  via  $v_i(C) = \sum_{j \in C} v_i(j)$ . Such representation allows us to compare agents utilities such that a certain cardinal value expresses the same intensity of a preference for all agents. In fact, our model corresponds to *additively separable hedonic games* (ASHGs) with symmetric and binary preferences (?), and thus our model will be referred to as an ASHG for brevity.

An outcome is thus a *partition*  $\mathcal{C}$  of  $N$  into disjoint coalitions, where  $|\mathcal{C}|$  denotes the number of its coalitions. Let  $\mathcal{C}(i)$  be the coalition  $C \in \mathcal{C}$  such that  $i \in C$ . Hence, for a partition  $\mathcal{C}$ ,  $\mathcal{C}(i) \succeq_i \mathcal{C}'(i)$  iff  $v_i(\mathcal{C}) \geq v_i(\mathcal{C}')$ , where  $v_i(\mathcal{C}) = v_i(\mathcal{C}(i))$  is the utility  $i$  receives from a partition  $\mathcal{C}$ . We focus on real-life scenarios where the size of each coalition is bounded. For instance, regarding our company event example, each table can accommodate a limited number of employees. Hence, for a positive integer  $\alpha$ , we consider partitions  $\mathcal{C}$  that are  $\alpha$ -bounded, i.e.,  $|C| \leq \alpha$  for every coalition  $C \in \mathcal{C}$ . We assume that  $\alpha \geq 2$  as the case  $\alpha = 1$  is trivial. We denote by  $\mathcal{C}_\alpha$  the collection of all  $\alpha$ -bounded outcomes. For an integer  $n > 0$ , for brevity, we henceforth denote  $[n] := \{1, \dots, n\}$  where  $[0] = \{\emptyset\}$ .

In our *online* model, agents appear one at a time in the order  $1, \dots, n$ . At each time  $t$ , an online algorithm  $\mathcal{A}$  shall produce a partial partition  $\mathcal{C}^t$  of the agents *who arrived until time  $t$* , without any knowledge regarding future agents. Upon the arrival of agent  $t$ ,  $\mathcal{A}$  should *immediately* and *irrevocably* decide whether to insert her to an existing coalition in  $\mathcal{C}^{t-1}$  or create a new coalition  $\{t\}$ . As the number of agents is not known upfront, from the perspective of an online algorithm, it may be possibly infinite. Prior research considered scenarios where friendships are revealed upon arrival (?). See Appendix A for a sample instance in such settings. Unlike prior studies, and unless stated otherwise, we consider that agents do *not* reveal upon arrival their edges to previously arrived agents. For such environments, we consider online algorithms augmented with a "coalitional friends predictor"  $\varphi_i : 2^N \rightarrow \mathbb{R}_{\geq 0}$  for each agent  $i \geq 2$ , which are possibly stochastic, inferred from additional knowledge about the graph or machine-learned from past data. We use such notation to stress that the predictors can operate on any coalition, but in practice they only operate on subsets of previously disclosed agents. Intuitively, the predictor captures agent  $i$ 's *uncertainty* on her friendships: it is an oracle that, given any coalition  $C \subseteq N$ , predicts the number of agent  $i$ 's friends within  $C$ . We also assume that we are given  $\varphi_i(N)$ , which predicts agent  $i$ 's degree in the full graph. For the predictor

---

## Algorithm 1: Maximum Predicted Coalitional Friends

---

```

1: Initialize an empty partition  $\mathcal{C} \leftarrow \emptyset$ .
2: while an online agent  $t \in N$  arrives do
3:   Set  $\mathcal{C}' = \{C \in \mathcal{C} : |C| < \alpha\}$ .
4:   if  $|\mathcal{C}'| > 0$  then
5:     Set  $\mathcal{S} = \arg \max_{C \in \mathcal{C}'} \varphi_t(C)$ .
6:      $C = \arg \min_{C \in \mathcal{S}} \sum_{i \in C} \varphi_i(N)$  (ties broken randomly)
7:     Add agent  $t$  to the coalition  $C$ .
8:   else
9:     Create a new coalition  $\{t\}$  and add it to  $\mathcal{C}$ .
10:  end if
11: end while
Output: The partition  $\mathcal{C}$ .

```

---

to be well-defined, we assume that  $\varphi_i(C) = \sum_{j \in C} \varphi_i(\{j\}) \forall C \subseteq N$ .

## Maximizing Social Welfare under Uncertainty

In this section, we evaluate the quality of partitions by measures of optimality. For a coalition  $C \subseteq N$ , let  $SW(C) = \sum_{i \in C} v_i(C)$  be the *social welfare* of  $C$ . The *social welfare* of a partition  $\mathcal{C}$  is then  $SW(\mathcal{C}) = \sum_{C \in \mathcal{C}} SW(C) = \sum_{i \in N} v_i(C)$ . Hence, a partition  $\mathcal{C}$  is *welfare-optimal* if it maximizes the social welfare amongst all possible partitions, i.e.,  $\mathcal{C} \in \arg \max_{\mathcal{C}' \in \mathcal{C}_\alpha} SW(\mathcal{C}')$ . For an integer  $\alpha \geq 2$ , the central goal of an online algorithm  $\mathcal{A}$  is computing a welfare-optimal  $\alpha$ -bounded partition  $\mathcal{C}^n$ .

Our *Maximum Predicted Coalitional Friends* (MPCF) algorithm (Algorithm 1) uses the predictors to greedily assign an arriving agent  $t$  to a coalition  $C \in \mathcal{C}^t$  of size less than  $\alpha$  that contains the maximum predicted number of friends, which yields the maximum increase in the social welfare of the current partition. If multiple such coalitions exist, the algorithm assigns agent  $t$  to the coalition  $C$  whose total predicted expected degree within the entire graph is *minimal*, i.e.,  $C$  minimizes  $\sum_{i \in C} \varphi_i(N)$ . Intuitively, coalitions with *low* total degrees should be filled as early as possible since we will have more chances to assign agents to coalitions with *higher* total degrees. If no such coalition exists, then MPCF creates a new coalition  $\{t\}$ .

As an algorithm, MPCF is simple, but our main novelty is the analysis of its behaviour with respect to the quality and choice of the predictors. For general graphs, we show that MPCF maintains a high social welfare even when the predictors are *noisy*. For depicting its good performance, we prove that MPCF is *optimal* for a very natural random graph model and analyze its expected social welfare. In the sequel, we denote the coalition to which agent  $i$  is assigned by an online algorithm  $\mathcal{A}$  as  $\mathcal{A}_i(G)$ . For brevity, we hereafter denote MPCF by  $\mathcal{A}^*$ .

## Robustness of MPCF to Noisy Predictors

Generally, we would like the algorithm to perform well when the predictions are decent (or even accurate), yet maintain reasonable performance even when the predictions are

slightly *noisy*. For *general* graphs, we thus show that MPCF will return a partition with a high social welfare even when the predictions are slightly off. Formally, given another predictor  $\varphi' = (\varphi'_i)_{i \in N}$ ,  $\varphi_i$  and  $\varphi'_i$  may disagree upon their predictions of agent  $i$ 's friendship with any other agent  $j$  (i.e.,  $\varphi_i(\{j\}) \neq \varphi'_i(\{j\})$  may hold) and thus induce different orderings on the agents who arrived *before* agent  $i$ . Thus, let  $\Delta(\varphi_i, \varphi'_i)$  be the *minimal* set of agents that should be removed such that the two predictors  $\varphi_i$  and  $\varphi'_i$  will induce the same ordering over the remaining agents in  $[i - 1]$ .

Let  $\Delta(\varphi, \varphi') := \cup_{i \in N} \Delta(\varphi_i, \varphi'_i)$ . We give an upper bound on the difference between the social welfares incurred by executing MPCF with the predictors  $\varphi_i$  and  $\varphi'_i$  (Theorem 1). Given any other predictor  $\varphi'$ , we infer that MPCF will still be near-optimal for a small enough value of  $|\Delta(\varphi, \varphi')|$ . In particular,  $|\Delta(\varphi, \varphi')|$  is upper bounded by the number of mispredicted neighbors of each agent.

**Theorem 1.** *Let  $\mathcal{A}_\varphi^*(G)$  be the partition generated by MPCF with a predictor  $\varphi$  on a graph  $G$ . Then,  $\text{SW}(\mathcal{A}_\varphi^*(G)) \geq \text{SW}(\mathcal{A}_{\varphi'}^*(G)) - |\Delta(\varphi, \varphi')|(2\alpha - 3)$  for any pair of predictors  $\varphi$  and  $\varphi'$ .*

*Proof. (Sketch)* In Appendix B, we first prove by induction that the removal of a single agent in the graph  $G$  cannot yield a better social welfare. Then, we invoke this result inductively on the graph  $G'$  resulting from the removal of the agents in  $\Delta(\varphi, \varphi')$  to obtain  $\text{SW}(\mathcal{A}_{\varphi'}^*(G)) \geq \text{SW}(\mathcal{A}_{\varphi'}^*(G'))$ . An agent  $j \in \Delta(\varphi, \varphi')$  can receive at most a utility of at most  $\alpha - 1$ . Upon the removal of agent  $j$ , the utility of agent  $j$ 's neighbors reduces by at most 1, and thus the partition's social welfare can decrease by at most  $2\alpha - 3$ . By invoking this reasoning recursively:  $\text{SW}(\mathcal{A}_\varphi^*(G)) \leq \text{SW}(\mathcal{A}_\varphi^*(G')) + |\Delta(\varphi, \varphi')|(2\alpha - 3)$ . Our result follows from  $\text{SW}(\mathcal{A}_\varphi^*(G')) = \text{SW}(\mathcal{A}_{\varphi'}^*(G'))$  as removing the agents in  $\Delta(\varphi, \varphi')$  yields that the predictors  $\varphi$  and  $\varphi'$  induce the same set of predicted friends for each agent.  $\square$

Namely, so long as  $\varphi$ 's number of mispredicted neighbors for each agent  $i$  is *small*, MPCF is still *near-optimal*.

## Optimality of MPCF for the CLV Model

For depicting MPCFs good performance and effectively analyzing it, we consider the natural Chung-Lu-Vu (CLV) random graph model (?) that generates graphs with arbitrary expected degree sequences. Our analysis stresses the applicability of our algorithm to practical graphs since the CLV model can generate graphs with power law distributed degrees, exhibited by many real-world graphs, e.g., Internet topology (?). By Newman, Watts, and Strogatz (?), one can obtain a fairly accurate model of many social networks by using the Molloy-Reed method (?), which samples a graph from a family of random graphs with degrees distributed following a power law with exponential cutoff.

Formally, for a sequence  $\mathbf{p} = (p_1, \dots, p_n) \in [0, 1]^n$ , we consider the random graph  $G_{\mathbf{p}}$  in which each pair of agents  $i \neq j$  are friends with probability  $p_i p_j$  and these events are mutually independent. This model corresponds

to the setting where agents pick their edges with probabilities proportional to  $\mathbf{p}$  which describes the relative distribution over the agents. Within this model, for each agent  $i$  we use the coalitional friends predictor that returns the expected number of agent  $i$ 's friends within a coalition  $C \subseteq [i]$ , i.e.,  $\varphi_i(C) = p_i \sum_{j \in C} p_j$ . Under this choice of predictors, we next show that the social welfare of the partition returned by MPCF *stochastically dominates* the social welfare incurred by *any* other algorithm  $\mathcal{A}$ , i.e., MPCF is *optimal* for graphs within the CLV model. First, we treat the case that an algorithm  $\mathcal{A}$  may leave an agent  $i$  in a singleton coalition even when her neighborhood is non-empty. Specifically, we prove in Lemma 1 that an agent having no friends hinders the social welfare of the partition generated by MPCF compared to the case where his neighborhood is non-empty.

**Lemma 1.** *Under the CLV model, let  $\mathbf{p} \in [0, 1]^n$  be a weight vector. For any agent  $i$ , let  $\mathbf{p}_{-i} \in [0, 1]^{n-1}$  be obtained from  $\mathbf{p}$  by removing its  $i^{\text{th}}$  entry. Then,  $\mathbb{P}[\text{SW}(\mathcal{A}^*(G_{\mathbf{p}})) \geq k] \leq \mathbb{P}[\text{SW}(\mathcal{A}^*(G_{\mathbf{p}_{-i}})) \geq k - v_i(\mathcal{A}_i^*(G_{\mathbf{p}}))]$  and  $\mathbb{P}[v_j(\mathcal{A}_j^*(G_{\mathbf{p}})) \geq k] \leq \mathbb{P}[v_j(\mathcal{A}_j^*(G_{\mathbf{p}_{-i}})) \geq k - 1]$  for each agent  $j \neq i$  and any  $k \geq 0$ .*

*Proof. (Sketch)* The proof in Appendix C shows that for any instance graphs  $G$  and  $G_{-i}$  of  $G_{\mathbf{p}}$  and  $G_{\mathbf{p}_{-i}}$  (resp.),  $\text{SW}(\mathcal{A}^*(G_{-i})) \geq \text{SW}(\mathcal{A}^*(G)) - v_i(\mathcal{A}_i^*(G))$  holds.  $\square$

Next, we show that another appealing property of MPCF is its capability of leveraging the density of a graph for attaining a higher social welfare. Given two weight vectors  $\mathbf{p}, \mathbf{p}' \in [0, 1]^n$ , we say that  $\mathbf{p}'$  *dominates*  $\mathbf{p}$  if  $p_i \leq p'_i$  for every agent  $i$ . This indicates that a graph distributed as  $G_{\mathbf{p}'}$  is *denser* than a graph distributed as  $G_{\mathbf{p}}$ . In Appendix D, we prove that the social welfare of the partition generated by MPCF for a given graph is stochastically dominated by the social welfare obtained for a *denser* one:

**Lemma 2.** *Under the CLV model, let  $\mathbf{p}, \mathbf{p}' \in [0, 1]^n$  be s.t.  $\mathbf{p}'$  dominates  $\mathbf{p}$ . Then,  $\mathbb{P}[\text{SW}(\mathcal{A}^*(G_{\mathbf{p}})) \geq k] \leq \mathbb{P}[\text{SW}(\mathcal{A}^*(G_{\mathbf{p}'})) \geq k] \forall k \geq 0$ .*

We now give our main result about MPCF's optimality:

**Theorem 2.** *Under the CLV model, let  $\mathbf{p} \in [0, 1]^n$  be a weight vector and let  $\mathcal{A}$  be an online algorithm for our problem. Then,  $\mathbb{P}[\text{SW}(\mathcal{A}(G_{\mathbf{p}})) \geq k] \leq \mathbb{P}[\text{SW}(\mathcal{A}^*(G_{\mathbf{p}})) \geq k]$  for any  $k \geq 0$ . In particular, for each agent  $i$ ,  $\mathbb{P}[v_i(\mathcal{A}_i(G_{\mathbf{p}})) \geq k] \leq \mathbb{P}[v_i(\mathcal{A}_i^*(G_{\mathbf{p}})) \geq k]$  for any  $k \geq 0$ .*

*Proof. (Sketch)* The proof in Appendix E stems from Lemmas 1 by induction on the number of agents  $n$ .  $\square$

Theorem 2 proves that MPCF is *optimal* for the CLV model. It also indicates that even if some online algorithm  $\mathcal{A}$  satisfies Lemma 2, MPCF still exploits the graph's density better than  $\mathcal{A}$  by Theorem 2. Though optimality only holds when the predictions are each agent's expected number of friends within a given coalition, MPCF is still near-optimal if the predictions are noisy due to Theorem 1.

**The Expected Social Welfare of MPCF.** We herein analyze the expected social welfare incurred by MPCF under the CLV model for  $n \geq 3$ , in both the asymptotic and the non-asymptotic case. Our first main result within this model is a set of equations that describe the social welfare incurred by MPCF. First, we consider the set  $I$  of the agents in  $G_{\mathbf{p}}$  that are assigned to singleton coalitions by MPCF. Let  $G_{\mathbf{p}}^+ = (N, E_{\mathbf{p}}^+)$  be the graph obtained from  $G_{\mathbf{p}}$  as follows: If  $p_i p_j > 0$  for a pair of agents  $i \neq j$ , then  $(i, j) \in E_{\mathbf{p}}^+$ . Let  $c_\ell$  be the number of coalitions in  $\mathcal{A}^*(G_{\mathbf{p}})$  with exactly  $\ell$  agents. We prove the following relation between the partition returned by MPCF and the graph  $G_{\mathbf{p}}^+$ :

**Lemma 3.** *Under the CLV model, let  $\mathbf{p} \in [0, 1]^n$  and let  $I$  be the agents in the graph  $G_{\mathbf{p}}$  that are assigned to singleton coalitions by MPCF. Then,  $I$  is an independent set of  $G_{\mathbf{p}}^+$ . Further, each coalition  $C \in \mathcal{A}^*(G_{\mathbf{p}})$  is connected in  $G_{\mathbf{p}}^+$  and  $SW(\mathcal{A}^*(G_{\mathbf{p}})) \leq \sum_{\ell=1}^{\alpha} \ell(\ell-1)c_\ell$ . Since  $n - |I| = \sum_{\ell=2}^{\alpha} \ell c_\ell$ , then  $SW(\mathcal{A}^*(G_{\mathbf{p}})) \leq (\alpha-1)(n - |I|)$ .*

*Proof.* Assume, towards contradiction, that the partition generated by  $G_{\mathbf{p}}$  contains two isolated agents  $i < j$  with  $p_i p_j > 0$ . Then, when  $j$  appears, MPCF adds agent  $j$  to the coalition  $\{i\}$ , which contradicts the fact that  $j$  is isolated. Thus,  $I$  is an independent set of  $G_{\mathbf{p}}^+$ . Next, note that when an agent  $i$  is inserted to an existing coalition  $C$ , there exists at least one agent  $j \in C$  such that  $p_i p_j > 0$  by Algorithm 1. Hence,  $C$ 's social welfare is at most  $|C|(|C| - 1)$ , yielding that  $SW(\mathcal{A}^*(G_{\mathbf{p}})) \leq \sum_{\ell=1}^{\alpha} \ell(\ell-1)c_\ell$ .  $\square$

We next show how to compute  $|I|$  for a symmetric variant of the CLV model for  $n \geq 3$ , where we consider the random graph  $G_d$  parameterized by the number of agents  $n$  and a vector  $\mathbf{d} = \{d_i\}_{i \in [n]}$ , while  $d_i$  is the expected degree of agent  $i$ . Any pair of agents  $i \neq j$  with the same expected degree  $d$  are friends with probability  $d/n$ . As in the previous section, we can analyze MPCF when the predictions for each agent  $i$  are given by the expected number of agent  $i$ 's friends within a coalition  $C \subseteq N$ , i.e.,  $\varphi_i(C) = |\{j \in C : d_j = d_i\}| \cdot d_i/n$ . Let  $Y_t^d$  be the number of agents with expected degree  $d$  who are in singleton coalitions by MPCF after agent  $t$  arrives.  $\{Y_t^d\}_{t \in [n]}$  is a Markov chain whose expected evolution is:

$$\mathbb{E}[Y_{t+1}^d - Y_t^d] = -\left(1 - \left(1 - \frac{d}{n}\right)^{Y_t^d}\right) \Pi_{d' \leq d} \left(1 - \frac{d'}{n}\right)^{Y_t^{d'}} \quad (1)$$

The first term is the probability that at least one isolated agent with expected degree  $d$  is a friend of agent  $t+1$ , and the second term is the probability that agent  $t+1$  has no isolated friend with lower expected degree (which would have been prioritized). Letting  $k_t^d = -\log(1 - d/n)$  and  $Z_t^d = -k_t^d \cdot Y_t^d$ , (1) can be simplified as:

$$\mathbb{E}[Z_{t+1}^d - Z_t^d] = k_t^d (1 - e^{Z_t^d}) \Pi_{d' < d} e^{Z_t^{d'}} \quad (2)$$

Following Kurtz (?) and many subsequent works (See, e.g., (?)), these Markov chains can be approximated by the solution of the following system of differential equations:

$$\frac{dz^d(t)}{dt} = k_t^d (1 - e^{z^d(t)}) \Pi_{d' < d} e^{z^{d'}(t)} \quad (3)$$

As  $k_t^d$  is independent of time  $t$ , there is a constant  $k^d$  s.t.  $k_t^d \equiv k^d$ . Similarly to (? , Theorem 6.1), we obtain that the solution  $z^d(t)$  approximates the number of isolated agents with expected degree  $d$  at time  $t$  via  $-z^d(t)/k^d$ . In fact, by letting  $\{\delta_f\}_{f=1}^\ell$  be the unique expected degrees, we obtain that  $|I|$  can be approximated by  $\sum_{f=1}^\ell -z^{\delta_f}(t)/k^{\delta_f}$ . We now conclude that the solution to (3) thus provides an *approximate* upper bound in terms of  $\{\delta_f\}_{f=1}^\ell$  on  $SW(\mathcal{A}^*(G_{\mathbf{p}}))$  for the non-asymptotic case, which is exact in the asymptotic case where  $n \rightarrow \infty$ :

**Lemma 4.** *As  $SW(\mathcal{A}^*(G_{\mathbf{p}})) \leq (\alpha-1)(n - |I|)$  (Lemma 3), we infer an expected upper bound of  $(\alpha-1)(n + \sum_{f=1}^\ell z^{\delta_f}(n)/k^{\delta_f})$  on the social welfare of the partition generated by MPCF when executed on  $SW(\mathcal{A}^*(G_{\mathbf{p}}))$ .*

In Appendix G, we provide a similar system for the general CLV model, whose solution can be similarly derived for cases such as the symmetric CLV model, and discuss its solvability. Now, we analyze MPCF's expected social welfare for Erdős-Rényi random graphs, where all edges in a random graph  $G_p$  independently appear with the same probability  $p \in [0, 1]$ . Each agent's expected degree is the same and equals to  $d := np$ . For a wide range of the parameters  $n, \alpha, p$ , we give in Appendix H an *exact* expression for MPCF's expected social welfare (up to a small additive error):

**Theorem 3.** *Let  $p \in [0, 1]$  and  $n, \alpha \in \mathbb{N}$ . Assume that  $p = o(\log n)/n$  and  $p \geq 1/n^{1+o(1)}$ . Then,  $\mathbb{E}[SW(\mathcal{A}^*(G_p))] = n^2 + \alpha n - \frac{n \ln(e^{p\alpha} + e^{pn} - 1)}{p} \pm n^{1.5+o(1)}$ .*

**Remark 1.** (Deterministic Graphs) *By Flammini et al. (? , Theorem 3.8), for undirected and unweighted graphs, no deterministic online algorithm has a competitive ratio better than  $\alpha - 1$ . For a restricted CLV model where weights are binary (i.e.,  $p_i \in \{0, 1\}$  for any agent  $i$ ), the resulting graph is deterministic and is exactly  $G_{\mathbf{p}}^+$ . In that case, MPCF is equivalent to the strictly  $\alpha$ -competitive greedy algorithm devised by Flammini et al. (? , Theorem 3.9), and we infer MPCF on deterministic graphs is almost optimal and achieves a strict competitive ratio of  $\alpha$ .*

## Proportionality in Online Hedonic Games

The principle of *proportional representation* in the context of HGs states that agents friendships should be reflected proportionately in the resulting partition. Informally, most agents shall be somewhat satisfied by that partition, which should also best represent the spectrum of different friendships within the society. Yet, existing proportionality axioms do not directly apply to HGs: capturing proportionality in HGs requires to reason about *multiple* coalitions instead of a *single* committee (see, e.g., (?)). Within each coalition, any subgroup of agents with similar sets of friends should have a number of friends that is proportional to the size of the subgroup. For instance, in our company event example, consider 100 employees and that each table can accommodate at most 10 participants. If there is a table where 50% of the participants have at least 5 common friends, then those participants should have at least 5 friends in their table. In this

section, we thus lift proportionality axioms from the multi-winner voting realm to hedonic games.

For evaluating and formulating new variants of proportionality that are suitable for *simple* and *symmetric* ASHG, we assume the number of agents  $n$  is *known*. First, we define the notion of *coalitional cohesiveness*, which extends the notion of cohesiveness in elections (?) to hedonic games. Cohesiveness dictates that the preferences within each subgroup of any coalition are sufficiently aligned. Recall that  $F_i$  is agent  $i$ 's neighborhood in the social network at time  $t$ . Given an integer  $\alpha \geq 2$ , for any  $C \in \mathcal{C}^t$  and each  $m \in [\alpha]$ , we say that a subgroup  $S \subseteq C$  is *m-coalitionally cohesive* (or *m-cohesive* for short) if  $|S| \geq m \cdot \frac{n}{\alpha}$  (i.e.,  $S$  is large enough) and  $|\cap_{i \in S} F_i| \geq m$  (i.e.,  $S$ 's members have at least  $m$  common neighbors). We also adapt the *approximate* variant of cohesiveness explored by (?). Given  $\eta \geq 1$ , the above subgroup of agents  $S \subseteq C$  is  $(\eta, m)$ -cohesive if  $|S| \geq \eta \cdot m \cdot \frac{n}{\alpha}$  and  $|\cap_{i \in S} F_i| \geq m$ .

Next, we redefine and study notions of proportionality in the context of hedonic games, that are commonly considered in the committee elections literature, which include proportional justified representation (PJR) (?) and extended justified representation (EJR) (?), as well as their approximate versions. The extension of proportional justified representation (PJR) (?) to hedonic games demands that the preferences of sufficiently large and cohesive groups of agents shall not be disregarded by the resulting partition, but shall be adequately represented in each coalition. Formally, coalitional PJR is defined as follows:

**Definition 1.** Given  $\eta \geq 1$ , the partition  $\mathcal{C}^t$  at time  $t$  satisfies  $\eta$ -coalitional PJR ( $\eta$ -CPJR) if for each coalition  $C \in \mathcal{C}^t$ , for all  $m \in [\alpha]$  and for each  $(\eta, m)$ -cohesive subgroup of agents  $S \subseteq C$ , it holds that the agents within  $S$  are jointly friends of at least  $m$  agents within  $C$ , i.e.,  $\sum_{i \in S} v_i(C) \geq m$ .

The stronger notion of coalitional EJR (?) demands that there exists at least one agent in each large enough and cohesive coalition with not just one, but several friends within the coalition. Formally:

**Definition 2.** Given  $\eta \geq 1$ , the partition  $\mathcal{C}^t$  at time  $t$  satisfies  $\eta$ -coalitional EJR ( $\eta$ -CEJR) if for each coalition  $C \in \mathcal{C}^t$  it holds that for each  $m \in [\alpha]$  and each  $(\eta, m)$ -cohesive subgroup of agents  $S \subseteq C$  there exists an agent  $i \in S$  that is friends with at least  $m$  agents within  $C$ , i.e.,  $v_i(C) \geq m$ .

If a partition satisfies 1-CPJR, then we say that it satisfies CPJR for brevity. CEJR is defined similarly. By (?), CPJR is implied by CEJR in the committee elections setting, and thus CEJR provides a stronger axiomatic guarantee of proportionality. Next, we analyze CPJR and CEJR when friendships are revealed, where CPJR can be easily attained under mild assumptions, while CEJR can be approximated. When friendships are *uncertain*, we show that our MPCF algorithm is *optimal* under the CLV model, yet it does not always satisfy CPJR and CEJR under full certainty.

## Proportionality when Friendships are Revealed

We first study CPJR and CEJR in scenarios *without uncertainty* (i.e., friendships are revealed). In such settings, at each

time  $t$  and for each agent  $i \in [t]$ , let  $\mathcal{N}_i^t = \{C \subseteq [t] : i \in C\}$  be all possible coalitions containing agent  $i$  and let  $F_i$  be agent  $i$ 's friends in the social network. Thus, agent  $t$  arrives along with  $v_t^t$ , where  $v_i^t : \mathcal{N}_i^t \rightarrow \{0, 1\}$  is the valuations of agent  $i \in [t]$  to the agents *who arrived until time  $t$* . Her utility for a coalition  $C \subseteq [t]$  is the number of her friends within  $C$ , i.e.,  $v_i(C) = |F_i \cap C|$ .  $\mathbf{v}^t = (v_i^t)_{i \in [t]}$  is the joint valuation of the agents *who arrived until time  $t$* .

**Achieving CPJR.** CPJR can be easily satisfied by an adaption of the *method of equal shares* (MES) (??), which we term *coalitional MES* (CMES) and it executes as follows. Each agent  $i$  has an initial budget of *one* dollar. Agent  $i$  spends her money across the execution by buying any agent  $j$  she approves of (i.e.,  $v_i(j) = 1$ ). Namely, at time  $t$ , agent  $t$  is assigned to a coalition whose members that approve  $t$  have at least  $n/\alpha$  dollars in total, where those members are then asked to jointly pay  $n/\alpha$  dollars. If multiple coalitions satisfying this property exist, CMES uniformly assigns  $t$  to some coalition with the *least* total budget. If such a coalition does not exist, then a new coalition  $\{t\}$  is created. Each agent's payment can be determined similarly to the classical MES, though the algorithm remains unaffected by the spread of  $n/\alpha$  among the agents. As agents pay  $n/\alpha$  dollars in total for each agent assigned to an already existing coalition, CMES assigns at most  $\alpha$  agents to each coalition. We now prove that:

**Theorem 4.** The partition returned by CMES is CPJR.

*Proof.* By contradiction, we assume that there are a coalition  $C \in \mathcal{C}^n$ ,  $m_C \in [\alpha]$  and an  $m_C$ -cohesive subgroup of agents  $S_C \subseteq C$  such that  $|\cup_{i \in S_C} F_i \cap C| \leq m_C - 1$ . At each time  $t$  where agent  $t$  is assigned to  $C$ , the agents within  $S_C$  that approve  $t$  pay exactly  $n/\alpha$  dollars, and since  $|\cup_{i \in S_C} F_i \cap C| \leq m_C - 1$ , they pay at most  $(m_C - 1) \cdot n/\alpha$  in total. Due to  $|S_C| \geq m_C \cdot n/\alpha$ , the agents within  $S_C$  have at least  $n/\alpha$  dollars at each time instant. Hence, at each time  $t$  where  $t \in \cap_{i \in S_C} F_i$ , those agents have enough money for buying  $t$ . Thereby, each candidate in  $\cap_{i \in S_C} F_i$  will be assigned to some existing coalition. There are at least  $m_C$  such agents. However, as  $|\cup_{i \in S_C} F_i \cap C| \leq m_C - 1$  there are at most  $m_C - 1$  such agents, which therefore constitutes a contradiction completing the proof.  $\square$

**Achieving CEJR.** We begin with an impossibility result which depicts that gaining  $\eta$ -CEJR is indeed much harder. Even when  $\alpha$ -bounded partitions can only contain a *single* coalition, our problem can be viewed as a special case of online approval committee elections (?), where the single coalition generated by an online algorithm constitutes the winning committee. Letting  $H(\alpha)$  be the  $\alpha$ -th harmonic number (i.e.,  $H(\alpha) = \sum_{j=1}^{\alpha} 1/j$ ), by Do et al. (?), Theorem 5.3) we thus infer the following lower bound:

**Corollary 1.** There exists no deterministic online algorithm that generates a  $(1 - \varepsilon)H(\alpha)$ -CEJR partition for any  $\varepsilon > 0$ .

Adapting the algorithm in (?), Section 5.2) for committee elections, we next provide the optimal *Greedy Coalitionally Cohesive* (GCC) scheme, i.e., GCC satisfies  $H(\alpha)$ -CEJR.

Informally, GCC assigns each agent  $i$  to a coalition containing a sufficiently large number of agents that are friends of  $i$ . At time  $t$ , if there exists a coalition  $C \in \mathcal{C}^t$  and  $S \subseteq C$  with  $S \subseteq N_t^t$  and  $|S| \geq H(\alpha) \cdot m \cdot n/\alpha$  s.t. each agent in  $S$  is a neighbor of less than  $m$  agents assigned to  $C$  thus far, then GCC assigns  $t$  to  $C$ . If multiple such coalitions exist, GCC uniformly assigns  $t$  to the coalition of *smallest* cardinality. If such a coalition does not exist, then a new coalition  $\{t\}$  is created. We now prove that:

**Theorem 5.** *The partition returned by GCC is  $H(\alpha)$ -CEJR and each coalition contains at most  $\alpha$  agents.*

*Proof. (Sketch)* By construction, the partition is  $H(\alpha)$ -CEJR. Using a budgeting argument, we prove in Appendix J that each coalition contains at most  $\alpha$  agents.  $\square$

Next, we further illuminate on the complexity of obtaining a partition satisfying  $\eta$ -CEJR. We remark that GCC requires checking whether a coalition is  $(\eta, m)$ -cohesive, where verifying its existence is generally NP-hard. The proof in Appendix K is by reduction from Maximum  $k$ -Subset Intersection (?).

**Theorem 6.** *Checking whether there exists an  $m$ -cohesive coalition is NP-hard.*

Theorem 6 dictates that GCC cannot be executed in polynomial time, and thus achieving an *optimal* partition as stated by Corollary 1 is challenging. Hence, we supply *Sub-Coalitions by Greedy Budgeting (SCGB)*, a polynomial-time algorithm that yields a slightly worse CEJR guarantee than GCC. Our scheme adapts the algorithm in (?, Section 5.3) for committee elections to hedonic games. First, let  $w(\cdot)$  be the inverse function of  $x \mapsto x^x$ , i.e.,  $w(\alpha) = x$  if  $\alpha = x^x$ . Note that  $w(\alpha) = O(\log \alpha)$  and  $\log \alpha = O(w(\alpha)^2)$ . Let  $\beta = \lceil w(\alpha) \rceil$ . SCGB independently creates  $\beta$  sub-coalitions generated similarly to CMES, each of size  $\lfloor \alpha/\beta \rfloor$ . Formally, each agent is given an initial budget of  $(1, \dots, 1) \in [0, 1]^\beta$ , i.e., there are  $\beta$  independent dollars where each one is associated with a specific possible sub-coalition. The  $j^{\text{th}}$  coin can be used for buying agents who are approved by at least  $n\beta^j/\alpha$ . Each agent costs  $n\beta/\alpha$  dollars. At time  $t$ , we find the largest triple  $j \in [\beta]$  and a coalition  $C \in \mathcal{C}^t$  with  $S \subseteq C$  satisfying  $S \subseteq N_t^t$  (we first maximize over  $j$  and then over  $|S|$ ), s.t.  $|S| \geq n\beta^j/\alpha$  and each agent in  $S$  has at least  $n\beta/(\alpha|S|)$  dollars of type  $j$  left. That is, those agents can afford to buy agent  $t$  assuming that each of them pays the same amount of money using the coins of type  $j$ . If such a triple  $(j, C, S)$  exists, then SCGB assigns  $t$  to  $C$ . If multiple such triples exist, SCGB uniformly assigns  $t$  to some coalition with the *least* total budget of type  $j$ . In both cases, each agent in  $S$  pays  $n\beta/(\alpha|S|)$  dollars for  $t$ . If such a triple does not exist, then a new coalition  $\{t\}$  is created. As each agent has  $\beta$  dollars in total, then buying each agent costs  $n\beta/\alpha$  dollars and thus SCGB assigns at most  $\alpha$  agents to each coalition. In Appendix L, we show that:

**Theorem 7.** *SCGB returns a  $\lceil w(\alpha) \rceil^2$ -CEJR partition.*

### Proportionality under Uncertainty

Surprisingly, when friendships are *uncertain*, our MPCF algorithm is also *optimal* for guaranteeing CEJR and CPJR in

the CLV model. That is, the probability that the partition produced by MPCF satisfies CEJR (CPJR) *dominates* the probability that the partition generated by *any* other algorithm  $\mathcal{A}$  satisfies CEJR (CPJR). The proof in Appendix I stems from minor modifications of the proof for Theorem 2.

**Theorem 8.** *Under the CLV model, let  $\mathbf{p} \in [0, 1]^n$  be a weight vector and let  $\mathcal{A}$  be an online algorithm for our problem. Then,  $\mathbb{P}[\mathcal{A}(G_{\mathbf{p}}) \text{ is CEJR}] \leq \mathbb{P}[\mathcal{A}^*(G_{\mathbf{p}}) \text{ is CEJR}]$ . In fact,  $\mathbb{P}[\mathcal{A}(G_{\mathbf{p}}) \text{ is CPJR}] \leq \mathbb{P}[\mathcal{A}^*(G_{\mathbf{p}}) \text{ is CPJR}]$ .*

**Corollary 2.** *Under the CLV model, MPCF is optimal in terms of both social welfare and proportionality.*

Our strong positive result does not generalize to cases where friendships are revealed. In such settings, though MPCF is almost optimal in terms of social welfare by Remark 1, the following example shows that the partition produced by MPCF does *not* necessarily satisfy CPJR or CEJR:

**Example 1.** *Our example is inspired by (?, Theorem 3). For  $\alpha \geq 3$ , consider that  $n = \alpha + 1$ , agent  $\alpha + 1$  is the only friend of agent 1 and the agents  $2, \dots, \alpha$  are all friends with each other. Then, MPCF will return the partition  $\mathcal{C} = (\lfloor \alpha \rfloor)$  consisting of a single coalition. Note that  $\{1\}$  is a 1-cohesive coalition, yet  $v_i(C) = 0$ , and thus yielding that the partition  $\mathcal{C}$  is neither CPJR nor CEJR.*

### Conclusions and Future Work

We have explored an online variant of partitioning agents in an undirected social network into coalitions of a bounded size. Initially, we gave the first results for maximizing social welfare in online hedonic games where algorithms have access to (possibly machine-learned) predictions, capturing uncertainty. Our work also initiated the study of lifting proportionality axioms from elections to hedonic games. We first analyzed the notions of CPJR and CEJR in scenarios where friendships are revealed. When friendships are *uncertain*, our MPCF algorithm is *optimal* in terms of *both* social welfare and proportionality for a vast family of natural random graphs. Our results can be seen as evidence that predictions are a promising tool for improving algorithms in online hedonic games, even if predictions are slightly noisy.

Our work opens the way for many future studies. Immediate directions are exploring other classes of hedonic games in online settings and studying proportionality in general uncertain domains. It is also appealing to examine scenarios where assignments may be postponed, agents may be reassigned after each arrival, or both.

Possimus sunt quae, nihil vel quas impedit dignissimos exercitationem?Quaerat eveniet veritatis doloremque numquam rerum reprehenderit magnam ad repudiandae maxime voluptates, numquam facilis vitae id non, voluptate veritatis impedit error dolorum officiis atque similique molitia sed facilis?Error consequuntur temporibus deleniti distinctio ut, ad sint ipsum aliquam nam consectetur officiis, qui repellat repellendus distinctio enim recusandae reiciendis eum veritatis quaerat, voluptates magnam voluptate cumque suscipit quod, dicta amet labore nisi veniam dolores explicabo quam eveniet ipsum illum?Quam ipsa consectetur similique officia quasi, ea praesentium eveniet explicabo labore

nemo, quasi saepe a dolores neque laboriosam rem obcaecati, necessitatibus laboriosam nesciunt illo eaque pariatur, error quis voluptatum maxime?