Hedonic Clustering Games

[Extended Abstract]

Moran Feldman
CS Dept., Technion
Haifa, Israel
moranfe@cs.technion.ac.il

Liane Lewin-Eytan
IBM Haifa Research Lab.
Haifa, Israel
lianel@il.ibm.com

Joseph (Seffi) Naor*
CS Dept., Technion
Haifa, Israel
naor@cs.technion.ac.il

ABSTRACT

Clustering, the partitioning of objects with respect to a similarity measure, has been extensively studied as a global optimization problem. We investigate clustering from a game theoretic approach, and consider the class of *hedonic clustering games*. Here, a *self organized* clustering is obtained via decisions made by independent players, corresponding to the elements clustered. Being a hedonic setting, the utility of each player is determined by the identity of the other members of her cluster. This class of games seems to be quite robust, as it fits with rather different, yet commonly used, clustering criteria. Specifically, we investigate hedonic clustering games in two different models: *fixed clustering*, which subdivides into k-median and k-center, and *correlation clustering*. We provide a thorough and non-trivial analysis of these games, characterizing Nash equilibria, and proving upper and lower bounds on the price of anarchy and price of stability. For fixed clustering we focus on the existence of a Nash equilibrium, as it is a rather non-trivial issue in this setting. We study it both for general metrics and special cases, such as line and tree metrics. In the correlation clustering model, we study both minimization and maximization variants, and provide almost tight bounds on both price of anarchy and price of stability.

Categories and Subject Descriptors

C.2.4 [**Computer Systems Organization**]: Computer-Communication Networks—*Distributed Systems*

General Terms

Theory, Performance

Keywords

Clustering Games, Hedonic Games, Price of Anarchy, Price of Stability

*Work supported in part by the Technion-Micorsoft Electronic Commerce Research Center and by ISF grant 954/11.

SPAA'12, June 25–27, 2012, Pittsburgh, Pennsylvania, USA.

1. INTRODUCTION

Clustering is the partitioning of objects or elements with respect to a similarity measure. The greater is the similarity of elements belonging to a cluster, or the distance between elements belonging to different clusters, the "better" is the clustering. Clustering has been extensively treated as a global optimization problem, employing a variety of optimization methods. We adopt here a novel game theoretic approach, and consider a setting in which a *self organized* clustering is obtained from decisions taken by independent players. We assume that the players correspond to the elements clustered, and their goal is to maximize their own utility functions. From the perspective of a single player, the quality, or the utility of a clustering, depends on the player's similarity to elements in her own cluster and perhaps on dissimilarity to elements in other clusters.

Our clustering games belong to the well known class of *hedonic* games, introduced in the Economics literature as a model of coalition formation. In a hedonic game, the utility of a player is solely determined by the identity of the players belonging to her coalition, and is independent of the partition of the other players into coalitions. Hedonic games were first introduced and analyzed by [11] in the context of cooperative games, and were motivated by situations in which individuals carry out joint activities as coalitions. Examples of such situations are individuals organizing themselves in groups for consumption or production purposes, or individuals relying upon local communities for the provisioning of public goods. Thus, hedonic games can be used to model settings arising in a wide variety of social, economic, and political problems, ranging from communication and trade to legislative voting. See [7] for a discussion of several real-life situations fitting the hedonic model. The notion of stability in hedonic games has been investigated both from cooperative, as well as non-cooperative, aspects [3, 7, 5]. The non-cooperative framework makes sense in environments lacking a social planner, or if the cost of coordinating movements is high. We note that most work on hedonic games has mainly focused on the existence of stable coalition partitions, whether core stable or individually stable, and on the complexity of finding such outcomes.

We investigate non-cooperative hedonic clustering games, in which elements are independent selfish players. Each player joins a group maximizing her utility, and the resulting clustering is the outcome of the choices of all players. We present a case study of two different well-known clustering models, with commonly used utility functions. The first model is *fixed clustering*, in which the number of clusters is fixed, and each cluster has a *centroid* whose position is determined by the identity of the cluster members. A player's utility depends on the location of the *centroid* of her cluster. The second model is *correlation clustering*, in which a player's utility depends on her similarity to other elements in her cluster as well as on her dissimilarity to elements in other clusters. In general, various settings in which players form clusters, and then each cluster provides a public good, or a service from a set of available alternatives, is captured by hedonic clustering games. Following are two motivating examples coming from different application areas.

In an ad-hoc (or sensor) network there is a large number of *autonomous devices* which are spread over a geographic area and wish to communicate with each other. In order to establish communication, devices invest transmission power which depends on the physical distance between them. Power is a critical resource for battery-limited devices, and thus the goal of each device is to minimize its transmission power and save on battery time. Fixed clustering is a proven method for enhancing energy efficiency and lifetime of large ad-hoc networks, and has been extensively studied in this context [1, 2]. Proposed clustering protocols organize the devices in data aggregation clusters to reduce network traffic. Each cluster has a center that receives data from other devices in the cluster, and sends it beyond the cluster limits, possibly after aggregating the received data and reducing its volume. A device will then join a cluster having the closest center to minimize the power needed for transmission, thus leading to a game-theoretic setting. We note that clustering in ad-hoc networks has been studied from a game-theoretic perspective in [19], yet their game definition is completely different from ours.

In online web advertising, publishers wish to join advertising services. Publishers are partitioned into clusters, and each cluster provides a different type of advertising service to its members. The type of service a cluster offers is derived from the attributes of its members, where possible attributes are, *e.g.*, fields of specialization, geographical area, organization size, types of product and annual budget. Publishers join or leave a cluster depending on the advertising service that the cluster offers and on the attributes of its members. For example, a new and relatively small business would prefer not to be coupled with a well-known large company specializing in a similar field. Thus, the utility of each player (publisher) depends on her similarity to the cluster, *i.e.*, how close are her advertising needs to the service provided by the cluster, and on her dissimilarity to players in other clusters (in the latter case, small and large businesses would be considered dissimilar). This is precisely the type of utility captured by correlation clustering.

Despite extensive work on clustering, not much work has been done from a game theoretic perspective, and we believe that this work contributes in that direction. We emphasize that the focus of our paper is not on a specific setting, but rather on the study of the general game theoretic framework in the context of hedonic clustering games.

1.1 Our Model

Our clustering problems are defined on a set of n points lying in a metric space with a distance function $d(\cdot,\cdot)$. The points correspond to selfish, non-cooperative, players (or users) moving between clusters at will. Players within a cluster are provided a service depending on the set of players belonging to it. A player achieves a utility from being a member of a cluster, and will naturally join the one maximizing her utility (or minimizing her cost). The notion of *social welfare* (or *social cost*) corresponds to the overall utility achieved by the system (or overall cost). The strategies of a player in a clustering game correspond to the set of clusters to which she can belong. Every choice of strategies by the players partitions them into clusters, and is called a *clustering configuration*. A Nash equilibrium[1] of the clustering game corresponds to a clustering configuration in which no user can unilaterally increase her utility (reduce her cost) by changing clusters. We investigate the clustering game in two different models: *fixed clustering* and *correlation clustering*.

Fixed Clustering.

In a fixed clustering game, the number of clusters is known beforehand and is denoted by k. Each cluster C has a *centroid*, $c(C)$, defined to be the element minimizing the cost of the cluster. We consider two well-known definitions in the clustering literature, known as k*-median* and k*-center*. In the k-median clustering problem, the cost of a cluster is defined as the sum of distances between all members of the cluster and its centroid. The centroid is thus defined as $c(C) = \arg\min_{u\in C}\left\{\sum_{v\in C} d(u,v)\right\}$. In the k-center clustering problem, the cost of a cluster is defined by its radius, *i.e.*, the maximum distance between an element in the cluster and its center. Hence, the centroid is $c(C) = \arg\min_{u\in C}\left\{\max_{v\in C} d(u,v)\right\}$. We note that in both models the choice of a centroid might not be unique, and therefore a tie-breaking rule is needed. We elaborate on this issue later on.

In both models, the strategy space of a player is defined by the k clusters. Each player v chooses the cluster C that minimizes her distance to the centroid $c(C^{+v})$, where C^{+v} denotes the cluster C with the addition of player v. Note that following v's addition to C, the centroid of C might change, *i.e.*, it might be that $c(C) \neq c(C^{+v})$. For k-median, the social cost is defined to be the sum of costs of all the clusters, whereas for k-center the social cost is defined to be the maximum cost of a cluster.

In fixed clustering, the service offered by a cluster is represented by its centroid. The example of ad-hoc networks fits this model, since the centroid is the node to which transmissions within a cluster are sent, and transmission costs depend on the distances to the centroid.

Correlation Clustering.

In settings where only the relationship among objects is known, correlation clustering is a natural approach. Unlike most clustering formulations, specifying the number of clusters as a separate parameter is not necessary. We assume that the similarity metric is captured by a distance metric $d(\cdot,\cdot) \in [0,1]$. If $d(u,v) \approx 0$, then u and v are very similar, and if $d(u,v) = 1$, then they are highly distinctive, unrelated elements. Each element v has a weight w_v denoting its "measure of influence" on other elements. Elements wish to be clustered with similar elements of high weight, and to be partitioned away from unrelated elements of high weight. Since the number of clusters is not fixed, the possible strate-

[1] We consider in this paper only pure Nash equilibria.

gies of a player are either to join an existing cluster, or to create a new cluster and become its sole member.

Given an element v, denote by C_v its cluster in a given configuration. Typically, two variants are studied in correlation clustering. In the *minimization variant*, the objective of each element v is to minimize its cost $\sum_{u \in C_v} w_u \cdot d(u, v) + \sum_{u \notin C_v} w_u \cdot (1 - d(u, v))$, *i.e.*, an element pays for being in a cluster with unrelated nodes of high weight, and for being partitioned away from similar elements of high weight. The *social cost* is defined as the sum of the costs paid by all elements. In the *maximization* variant, the objective of each node is to maximize its utility $\sum_{u \in C_v} w_u \cdot (1 - d(u, v)) + \sum_{u \notin C_v} w_u \cdot d(u, v)$, *i.e.*, an element achieves utility from being in a cluster with similar nodes of high weight and from being partitioned away from unrelated nodes of high weight. Again, the *social welfare* is the sum of the utilities achieved by all elements.

Correlation clustering essentially models scenarios in which the objective is to either minimize the difference or maximize the similarity among objects within clusters. This type of clustering depends on the relationship among elements. The advertising example fits this model if distances between objects represent willingness to be clients of the same advertising service, and weights represent market influence.

1.2 Our Contribution

We provide a thorough and non-trivial analysis of hedonic clustering games under several models, characterizing Nash equilibria, and proving upper and lower bounds on price of anarchy and price of stability[2]. Our study covers a broad subclass of hedonic games which seems to lack previous investigation from a game theoretic perspective. This subclass captures clustering as a self organizing process governed by game theoretic considerations. We note that it is important to study clustering games in several models, since it is a diverse subject, and cannot be captured by a single framework [6]. Our models seem to have quite a robust definition, as they fit well with rather different, yet commonly used, clustering criteria.

The first clustering model we consider is fixed clustering. For a general metric, we show that Nash equilibrium does not necessarily exist. Clearly, imposing high enough penalties on players for changing the location of a centroid (when moving to a different cluster) would guarantee the existence of Nash equilibrium for both k-median and k-center models. We prove that setting the penalty to be equal to the distance traveled by the centroid suffices. This choice of penalty is very natural, and can be thought of as a fee imposed by a cluster on nodes joining it, in order to cover the incurred expenses. This choice of penalty also resembles the way costs are determined by the VCG mechanism [16].

Since Nash equilibrium does not necessarily exist in general, we study the fixed clustering game in specific metrics, *i.e.*, tree and line metrics. The strict definition of hedonic games requires that the members of a cluster uniquely determine the location of its centroid. To achieve that one needs to specify some (possibly arbitrary) static tie-breaking rules. However, there exist instances for which no Nash equilibrium exists under any choice of a static tie-breaking rule, even for line metrics. We circumvent this issue by using tie-breaking rules that are *history dependent*; when the choice of a centroid is not unique, it depends on the previous states of the system.[3] In this respect, our work on fixed clustering deviates from the class of hedonic games. However, we emphasize that even if fixed clustering games were strictly hedonic, our results would not have followed from existing literature.

The proof of existence of Nash equilibrium in a tree metric for the k-median model is rather involved and non-trivial. It is based on a characterization of a centroid, definition of a potential function, and a judiciously chosen schedule of moves of players resulting in equilibrium. We note that for the k-center model, the proof of existence of Nash equilibrium under tree metrics requires allowing centroids to be located in an arbitrary location between the two end points of an edge. We believe that this relaxation is not necessary (as is the case for line metric), but we have not managed to prove that. For both the k-median and k-center models, we show that the price of stability is 1, while the price of anarchy is unbounded.

Going back to the example of ad-hoc networks, from a designer's perspective, our work implies that simple greedy-like algorithms are sufficient for reaching Nash equilibrium, and thus the devices do not need to run more sophisticated protocols.

The second clustering model we consider is correlation clustering, for which we obtain results for both minimization and maximization variants. This model is closely related to additively-separable hedonic games [13], and techniques used for this class of games can be extended to show that Nash equilibria always exist for correlation clustering games, but finding them is PLS-complete.

Our main results for correlation clustering games are lower and upper bounds on both the price of anarchy and price of stability. The bounds are proved by characterizing the distance between nodes belonging to the same cluster vs. the distance between nodes belonging to different clusters. The specific bounds are:

- In the special case of equal node weight, the price of stability is 1 for both the minimization and maximization variants. For arbitrary weights, the price of stability is strictly larger than 1.
- For the minimization variant, an upper bound on the price of anarchy is $O(n^2)$ and the corresponding lower bound is $n-1$. For the special case of equal nodes weight, the lower bound still holds, and there is an improved upper bound of $n-1$ on the price of anarchy.
- For the maximization variant, the price of anarchy is $\Theta(\sqrt{n})$. This bound holds even if all nodes have equal weight and the metric is a tree metric. In case of a line metric, the price of anarchy is $\Theta(n^{1/3})$, even for equal nodes weight.

For both the fixed clustering and correlation clustering models, an intriguing question is what kind of mechanisms can be used to reduce the price of anarchy. We make a first

[2]Price of stability is defined as the ratio between the social value of the *best* Nash equilibrium and the social optimal solution, while price of anarchy is defined as the ratio between the social value of the *worst* Nash equilibrium and the social optimal solution.

[3]The history of the system is one of the parameters for our tie-breaking rules. For a one-shot game, these rules reduce to static tie-breaking rules.

step in this direction by showing that the price of anarchy of the maximization variant of correlation clustering can be bounded by k (for $k \geq 2$) by limiting the game to k clusters. Moreover, if all node weights are equal, this constraint can be removed after the game reaches a Nash equilibrium, and the price of anarchy still remains at most k.

Previous work.

Clustering is a vast area of research with abundant results, and therefore we mention only a few, directly related, results. Fixed clustering is the classic approach to clustering data and the goal of the optimization problem is to find a partitioning of the nodes to k clusters, such that the cost is minimized. The k-center problem was considered by Gonzalez [14], who gave the first 2-approximation algorithm (see also the 2-approximation algorithm of [17].) For the k-median problem, the first constant factor algorithm was given by [8]. The approximation factor has since been improved in a sequence of papers (see, *e.g.*, [23]). We note that in case of a tree metric, k-median can be solved optimally in polynomial time [22].

Correlation clustering was first defined by [4]. They considered the version where the edges of a complete graph are labelled as either "+" (similar) or "-" (different), and the goal is to find a partition of the nodes into clusters that agrees as much as possible with the edge labels. They considered both maximizing agreements and minimizing disagreements, obtaining a constant factor approximation for the former and a PTAS for the latter. These results were generalized for real-valued edge weights; [10, 12, 9] obtained a logarithmic approximation algorithm for the minimization version and [9] obtained a constant (greater than $1/2$) factor approximation for the maximization version.

From game-theoretic perspectives, in addition to being a hedonic game, our correlation clustering game falls into a class known as *polymatrix games*, introduced by Yanovskaya [24]. Few other games that fall into both classes were also considered. We mention only those that are most closely related to our model. Hoefer [18] considered a game called "MaxAgree" which is equivalent to the maximization variant of our correlation clustering game with equal node weights and a limit ℓ on the number of possible clusters. For this game [18] shows that best response dynamics converge in polynomial time to Nash equilibrium (which does not seem likely in our model), and gives bounds on the price of anarchy. Another example is the game version of Max-Cut, considered by multiple works, *e.g.*, [18, 15], and, in a sense, is the inverse of our model.

A different game theoretic representation of clustering is given by [20, 21], however, their approach is completely different from ours.

2. FIXED CLUSTERING

In this section we consider the fixed clustering game in which the number of clusters is known beforehand, and is denoted by k. We investigate both the k-median and k-center variants of this model. In both variants, given the set of nodes in a cluster, the choice of a centroid may not be unique. As already mentioned, in order to fit perfectly into the hedonic model, a static tie-breaking rule for choosing a unique centroid is required. However, such tie breaking rules may have a negative impact on our results, as demonstrated by Theorem 2.1.

THEOREM 2.1. *With (static) tie breaking rules, there may not exist Nash equilibrium for both the k-median and k-center variants, even for line metrics with three nodes.*

We circumvent this issue by analyzing game dynamics that allow tie-breaking rules which are *history dependent*. Initially, an arbitrary static tie breaking rule R is used. (Rule R also applies to a one shot game.) During the dynamics, whenever a player performs a move and changes her strategy, each centroid remains at the same node if it is still a valid location for a centroid; otherwise, the static tie breaking rule R is used to relocate it. The use of a history dependent rule implies that the cost observed by a player does not solely depend on the identity of the other players in her cluster, but also on the history of the game. We can thus consider the fixed clustering model as a hedonic game with an additional attribute.

Due to space limitations, most proofs of this section are omitted.

2.1 The k-Median Model

In the k-median model, the cost of a cluster C is defined as the sum of the distances between all members of the cluster and its centroid, and the centroid $c(C)$ is defined as the node minimizing the cost of the cluster, that is, $c(C) = \min_{u\in C}\left\{\sum_{v\in C} d(u,v)\right\}$. We denote by $D(u,C)$ the sum of distances between u and the other nodes in C. Thus, $c(C) = \min_{u\in C} D(u,C)$. For ease of notation, we denote by $D(C)$ the cost of a cluster C, that is, $D(C) = D(c(C), C)$.

The cost of a node v in the k-median clustering game is defined as its distance from the centroid of its cluster, $d(v, c(C_v))$. A clustering configuration of the k-median clustering game is a Nash equilibrium if no player can reduce its cost by choosing a different cluster, assuming the other players stay in their cluster. We assume a node v changes its strategy from cluster C_1 to C_2 only in case it strictly decreases its distance to the center, that is $d(v, c(C_1)) > d(v, c(C_2^{+v}))$. As mentioned, we assume that following a move performed by a player, the centroid of a cluster will not change its location unless forced (*i.e.*, only in case the sum of distances of the points from the new location of the centroid is strictly lower than from its previous location).

We first consider the general metric case, and prove that Nash equilibrium does not necessarily exist. Moreover, we show that this is the case even if centroids are allowed to be located at any location along edges, rather than only at a node of their own cluster. Then, we notice that by imposing a high enough penalty on a node whose move to a cluster changes the location of its centroid, Nash equilibrium is guaranteed to exist. Moreover, we show that it is enough to set the penalty to be equal to the distance traveled by the centroid. Motivated by these results, we study the game in line and tree metrics, and show that in these cases Nash equilibrium always exists, with no further assumptions such as penalties. Note that the existence of Nash equilibrium in the line case is implied by our result for the tree case. Still, we consider line metrics separately, since we can show that best response dynamics converge to equilibrium.

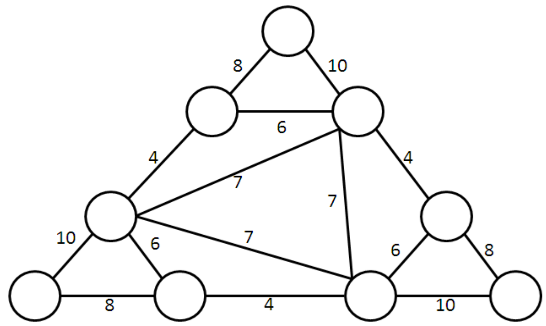

Figure 1: Graph with no Nash equilibrium (assuming two clusters).

2.1.1 The General Metric Case

For a general metric, there is no guarantee that a Nash equilibrium exists. Figure 1 displays a graph representing a metric having no Nash equilibria, assuming there are two clusters. The numbers on the edges of the graph represent the distances. The proof that no Nash equilibrium exists for this graph is done via a case analysis, showing that any selection of two nodes to be the centroids of the clusters results in a configuration where either a cluster has a non optimal centroid choice, or there is a node wishing to deviate.

A natural way to try to guarantee the existence of a Nash equilibrium is by allowing the centroid of a cluster to be located at any location along an edge of the graph. Formally, every edge (u, v) and $\lambda \in [0, 1]$ represents a possible location for a centroid. The distance of this centroid from a node w is $\min\{d(u, w) + \lambda \cdot d(u, v), d(v, w) + (1 - \lambda) \cdot d(u, v)\}$. The centroid of the cluster is placed at the location minimizing the sum of distances from the nodes of the cluster. Unfortunately, this generalization fails to guarantee the existence of a Nash equilibrium. The graph in Figure 1 represents a counter example for this case as well. In order to establish the counter example based on this graph, we use the notation of weak Nash equilibria.

Our original definition of Nash equilibrium states that a centroid will not change location unless forced. If the centroid is forced to change location, it might have multiple locations to which it can move. The choice among these possible new locations is based on some arbitrary tie breaking rule. On the other hand, a configuration is said to be in *weak Nash equilibrium* if there is no node u wishing to deviate to any cluster $C \neq C_u$, given that following u's deviation, C's centroid will move to the location farthest away from u among its possible locations. In other words, if there are multiple possible locations for the centroid of C after u's move, then the centroid will move to the worst location from the point of view of u. Moreover, this move occurs even if the original location of the centroid is still a possible location for it.

Clearly, every Nash equilibrium is also a weak Nash equilibrium, though the reverse is not necessarily true. We can show that the graph in Figure 1 allows no weak Nash equilibria with two clusters; and therefore, it also does not allow any Nash equilibria (assuming two clusters). The proof is established by case analysis, and is based on Theorem 2.2, where we show that it is sufficient to consider only half integral locations on edges as optional placements for centroids.

THEOREM 2.2. *An instance of the k-median clustering game with integral edge distances has a weak Nash equilibrium if and only if it has a weak Nash equilibrium configuration in which centroids are placed in half integral locations on edges.*

In order to guarantee the existence of Nash equilibrium, we add a rule penalizing a node whose move to a cluster C changes the location of the centroid of C. That is, the cost of a node u performing an improvement move will consist of two values:

- The distance from the updated centroid of the new cluster, $d(u, c(C^{+u}))$.
- A cost equal to the distance between the original and new centroids of C, $d(c(C), c(C^{+u}))$.

Intuitively, one can think of this choice of a penalty as a fee imposed by a cluster on nodes joining it, in order to cover the incurred expenses. Note that the total cost of the nodes in the cluster deserted by u can only decrease.

LEMMA 2.3. *Under the above penalties, a node u will only deviate from cluster C_1 to cluster C_2 if its distance from the original location of the centroid $c(C_2)$ is shorter than its distance from $c(C_1)$.*

Lemma 2.3 implies that the above penalizing rule prevents a node from deviating unless the deviation results in an improved social value. Since the social value cannot improve forever, the game must converge to Nash equilibrium. The following theorem formalizes this argument.

THEOREM 2.4. *Under the above penalties, every best response move strictly decreases the social cost. Hence, there always exists Nash equilibrium, and it is guaranteed to be reached by best response dynamics.*

COROLLARY 2.5. *Under the above penalties, the price of stability is* 1.

PROOF. By Theorem 2.4, the social cost is guaranteed to decrease with every best response move. Thus, the optimal solution must be a Nash equilibrium. □

Note that Nash equilibrium in the setting without penalties is also valid with penalties (as penalties can only decrease the benefit of a move). The price of anarchy for this case is unbounded, as shown next.

THEOREM 2.6. *The price of anarchy of the k-median clustering game is unbounded, even for a line metric.*

PROOF. Consider a line with three nodes at locations 0, 1 and M, for some large M, and assume $k = 2$. There are two clusters $\{0\}$ and $\{1, M\}$, and the centroid of the second cluster is the node 1. Clearly this is a Nash equilibrium, since the only node which is not a centroid will gain nothing by deviating. However, the cost of this equilibrium is M. On the other hand, the optimal solution has the following two clusters: $\{0, 1\}$ and $\{M\}$, yielding a cost of 1. The price of anarchy of this instance is thus M. □

2.1.2 The Line Metric Case

We prove that if players are nodes on a line, Nash equilibrium always exists (without penalties, and allowing a centroid to be placed only at a node). We begin by characterizing the centroid of a cluster on a line.

LEMMA 2.7. *The centroid of a cluster on a line is a median*[4] *of all nodes in the cluster.*

In case the number of nodes in a cluster is odd, there exists a single median node. Thus, this node is the single optional centroid node, and divides the other nodes in the cluster to right and left sets where each set contains $(m-1)/2$ nodes. In case the number of nodes in a cluster is even, there are two optional centroid nodes. These are two consecutive nodes on the line, each dividing the nodes to right and left sets such that one of the sets contains $\lceil \frac{m-1}{2} \rceil$ nodes and the other set contains $\lfloor \frac{m-1}{2} \rfloor$ nodes.

Given a clustering configuration $\mathcal{F}$, we define a potential function Φ, and show that it strictly decreases with each strategy change performed by a player. We assume a player changes her strategy only if it strictly decreases her distance to the center. The potential function is equal to the social cost function, that is

$$\Phi(\mathcal{F}) = \sum_{i=1}^{k} \sum_{v \in C_i} d(v, c(C_i)) \ . \qquad (1)$$

The following lemma is proved using the characterization of the centroid given by Lemma 2.7.

LEMMA 2.8. *The potential function Φ strictly decreases during the clustering game's natural dynamics.*

As the strategy space of all nodes is finite, the potential function will decrease following each move performed by a node until reaching a local (or global) minimum, corresponding to Nash equilibrium.

COROLLARY 2.9. *Nash equilibrium always exists for the k-median clustering game on a line.*

2.1.3 The Tree Metric Case

We assume here that the distances are defined by a tree metric, and the players are exactly the nodes of the tree. We prove that in this case, a Nash equilibrium always exists. First, we characterize the centroid node of a cluster for the tree case. To this end we define the *median node of a tree* as follows.

DEFINITION 2.10. *Given a tree T with $|V|$ nodes, a node $v^* \in T$ is called a* median node *if its removal partitions T into connected components of size at most $\left\lceil \frac{|V|-1}{2} \right\rceil$ each.*

The following lemma is well known.

LEMMA 2.11. *There are at most 2 median nodes in a tree.*

In order to define the relation between a median and the centroid of a cluster, we use the next definition.

DEFINITION 2.12. *A cluster C has the* closure property *if all nodes on each path between two nodes of C belong to C as well.*

LEMMA 2.13. *The centroid of a cluster with the closure property in a tree is a median node of the cluster.*

[4]A median node of a cluster is a cluster node for which the number of cluster nodes to its left and to its right differ by at most 1.

In order to prove that the k-median clustering game always has a Nash equilibrium, we use the potential function of Equation (1), and describe a schedule that converges to equilibrium. Unlike the line metric case, we cannot simply use the potential function for each improvement move performed by a node, since the closure property can easily be violated by a best response move. Instead, we need to define a set of moves that keep the closure property, allowing us to use the median characterization of centroids, and that are guaranteed to strictly decrease the potential function.

We describe the convergence schedule. It consists of iterations, where each iteration starts and ends with a clustering configuration in which all clusters have the closure property. We call such a configuration a *closed configuration*. Starting from a closed configuration, consider a node v in cluster C_1 wishing to move to cluster C_2. All nodes on the path between v and $c(C_2)$, denoted by $\delta(v, c(C_2))$, are either already in C_2 or wish to move to C_2 as well (if there are other clusters with equal distance to centroid, we choose C_2). This follows since nodes in $\delta(v, c(C_2))$ are closer than v is to $c(C_2^{+v})$. Note that the center of C_2 will be in the same way given that any node of $\delta(v, c(C_2))$ moves to C_2. In addition, if there is a node $w \in \delta(v, c(C_2))$ having a better cluster C_m such that $d(w, c(C_m^{+w})) < d(w, c(C_2^{+w}))$, then the same holds for v as well since: $d(v, c(C_m^{+v})) \leq d(w, c(C_m^{+w})) + d(w, v) < d(w, c(C_2^{+w})) + d(w, v) = d(v, C_2^{+v})$.

Thus, when starting a new iteration of the schedule, we choose a node u adjacent to C_2, and make it perform a best response move from C_1 to C_2 (we call this move "first best response"). Following this move, there are two options. In case C_1^{-u} remains connected, then it still has the closure property, and we can select a new first best response (assuming we did not reach Nash equilibrium). Otherwise, C_1^{-u} contains multiple connected components, and there is a component C_1' containing the centroid of C_1^{-u}. Let C_1'' denote $C_1^{-u} - C_1'$, *i.e.*, C_1'' contains all nodes of C_1 except for the node that made the first best response (u) and the nodes of the connected component containing C_1^{-u}'s centroid (that is, C_1'). Next, we move all nodes of C_1'' to cluster C_2, and finish the iteration. Note that beside of the "first best response" move, all other moves in the iteration need not be best response moves.

Both C_1' and $C_2 \cup \{u\} \cup C_1''$ are closed clusters. No other cluster is affected by the iteration, hence, by the end of the iteration, we get back to a closed configuration. We are now ready to use our potential function, and show that it strictly decreases with each iteration of the above schedule. We conceptually divide the iteration into three steps, and show that none of them can increase the potential function, and at least one of them strictly decreases it. The three steps are as follows.

1. The first best response move made by u, including the possible relocation of the clusters' centroids.
2. The move of the nodes of C_1'' to C_2, assuming the clusters' centroids are not allowed to move.
3. A possible move of the clusters' centroids to new locations (following the move of the nodes of C_1'').

LEMMA 2.14. *In step 1 of the iteration, the potential function P strictly decreases.*

PROOF. We consider a node u moving from C_1 to C_2, and show that the sum of distances from all nodes in C_1^{-u} and

C_2^{+u} to their respective centers decreases. This is clearly true for u (the distance strictly decreases since u had an incentive to move) as well as for C_1^{-u} (as the centroid of C_1^{-u} is relocated in order to minimize its total distance from the cluster's nodes). As for C_2, observe that it has the closure property by the above schedule, hence Lemma 2.13 applies. We consider two cases. Let $Q_1, \ldots, Q_m$ denote the connected components in C_2 formed after the removal of $c(C_2)$.

1. Node u joins a connected component Q_j of size at most $\left\lceil \frac{|C_2|}{2} \right\rceil - 1$ in C_2. Then, after it joins C_2, $c(C_2)$ remains a median node of C_2^{+u} and thus the centroid remains at the same location.
2. Node u joins a connected component Q_j of size $\left\lceil \frac{|C_2|}{2} \right\rceil$ in C_2. This can occur if $|C_2|$ is even, and then $\left\lceil \frac{|C_2|}{2} \right\rceil = \left\lceil \frac{|C_2|-1}{2} \right\rceil$ and Q_j is a maximal connected component. In this case, there are two optional centroids (median nodes) in C_2, that is $c(C_2)$ and $v_{in}(Q_j)$ (the entry point to the connected component Q_j). Clearly, $D(c(C_2), C_2) = D(v_{in}(Q_j), C_2)$. Now, after v joins Q_j, the centroid is forced to move to $v_{in}(Q_j)$, but the total distance of the vertices in C_2 from their centroid remains unchanged. □

LEMMA 2.15. *In steps 2 and 3 of the iteration, the potential function P cannot increase.*

COROLLARY 2.16. *The potential function P strictly decreases in each iteration of the above schedule.*

As the strategy space of all players is finite, the potential function decreases following each iteration performed by the schedule, until reaching a local minimum (or global), corresponding to Nash equilibrium.

COROLLARY 2.17. *Nash equilibrium always exists for the k-median clustering game in a tree.*

We consider only dynamics of closed configurations. The next lemma states that it is not restrictive.

LEMMA 2.18. *For any configuration of the clustering game on a tree, there always exists a closed configuration which is at least as good with respect to the social cost.*

As the potential function of the game is equal to the objective function, an optimal k-clustering closed configuration for the k-median model is also in Nash equilibrium (no move can further decrease the global minimum point of the objective function). Moreover, following Lemma 2.18, the cost of such a configuration is equal to the cost of an optimal configuration. We thus get the following corollary.

COROLLARY 2.19. *The price of stability of the clustering game in a tree is 1.*

Note that since a constant-approximation for the k-median problem (or even an optimal one for tree metrics) can be computed in polynomial time, a Nash equilibrium with a price of anarchy $O(1)$ can be guaranteed by setting the initial configuration to be such an approximate solution, converting it to a closed configuration, and then scheduling the moves as described above until reaching equilibrium. Since the cost of the clustering configuration strictly decreases along the process, it will reach Nash equilibrium with price of anarchy $O(1)$.

2.2 The k-Center Model

In the k-center model, the cost of a cluster C is defined by its radius, which is the maximal distance between its centroid and a node of the cluster. The centroid $c(C)$ is defined as the node minimizing the cost of the cluster, that is, $c(C) = \min_{u \in C} \left\{ \max_{v \in C} d(u,v) \right\}$. The cost of a node v in the k-center clustering game is defined as its distance from the centroid of its cluster, $d(v, c(C_v))$. A clustering configuration of the k-center clustering game is in Nash equilibrium if no user can reduce its cost by choosing a different cluster, assuming the other users stay in their individual clusters.

We first consider the case of a line metric, and prove that Nash equilibrium always exists. Then, we turn to tree metrics, and guarantee the existence of Nash equilibrium in case centroids can be placed anywhere along edges. Finally, we consider general metrics, and prove that Nash equilibrium does not necessarily exist, even if centroids are allowed to be placed at any location along edges. We note that by imposing a high enough penalty on a node whose move affects the location of the target cluster's centroid, existence of Nash equilibrium is guaranteed. We further show that setting the penalty to be equal to the distance traveled by the centroid is enough. As for the price of anarchy, we show it is unbounded in all settings considered.

2.2.1 The Line and Tree Metric Cases

THEOREM 2.20. *Nash equilibrium always exists for the k-center clustering game on a line.*

We suspect that Nash equilibrium also always exists in tree metrics, and best response dynamics are guaranteed to reach it. However, we manage to prove this only in case the centroid of a cluster is allowed to be placed at any location along an edge (rather than only at a node). Formally, every edge (u, v) and $\lambda \in [0, 1]$ represent a possible location for a centroid. The distance of this centroid from a node w is $\min\{d(u,w) + \lambda \cdot d(u,v), d(v,w) + (1-\lambda) \cdot d(u,v)\}$. The centroid of the cluster is placed at the location minimizing the maximum distance from a node of the cluster. Note that the centroid could be located at a node which does not belong to its cluster, however, the configuration may be stable only if this node is also the centroid of its own cluster.

LEMMA 2.21. *In case centroids are allowed to be placed on edges, the centroid of a cluster is always located in the middle of its diameter (since we are dealing with a tree metric, all diameters share their middle point).*

The following theorem is proved using the characterization of the centroid given by Lemma 2.21.

THEOREM 2.22. *In case centroids are allowed to be on edges, Nash equilibrium always exists for tree metrics and it is guaranteed to be reached by best response dynamics.*

THEOREM 2.23. *In case centroids are allowed to be on edges, the price of anarchy of the k-center clustering game remains unbounded, even for the case of a line metric.*

2.2.2 The General Metric Case

Similarly to the k-median model, Nash equilibrium might not exist for the case of a general metric. Figure 2 represents

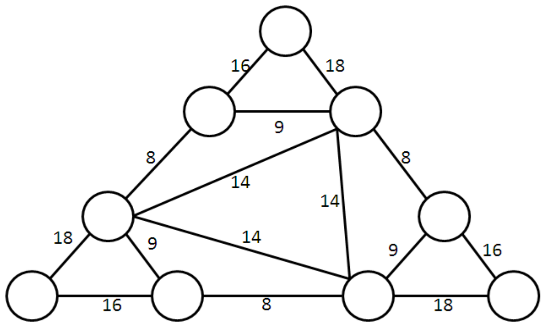

Figure 2: Graph with no Nash equilibria (assuming two clusters).

an instance that does not have a Nash equilibrium, assuming there are two clusters. The numbers on the edges of the graph represent the distances. The proof that no Nash equilibrium exists for this graph is established similarly as for the graph in Figure 1.

One can try, again, the trick that worked for tree metrics, *i.e.*, allow the centroid of a cluster to be located at any location along an edge of the graph. Unfortunately, this generalization fails to guarantee the existence of Nash equilibrium in this case. The graph in Figure 2 represents a counter example for this case as well. The proof is established by case analysis and is based on Theorem 2.24, where we show that in Nash equilibrium centroids can be placed only in half integral locations on edges.

THEOREM 2.24. *In any Nash Equilibrium configuration of the k-center clustering game with integral edge distances, centroids must be placed in half integral locations on edges.*

In order to guarantee the existence of Nash equilibrium, we add to the game a rule penalizing a node whose move to a cluster C changes the location of the centroid of C. That is, the cost of a node u performing an improvement move will consist of two values:

- The distance from the updated centroid of the new cluster, $d(u, c(C^{+u}))$.
- A cost equal to the distance between the original and new centroids of C, $d(c(C), c(C^{+u}))$.

THEOREM 2.25. *Under the above penalties, there is always a Nash equilibrium solution, and it is guaranteed to be reached by best response dynamics.*

3. CORRELATION CLUSTERING

In this section we consider the clustering game in the correlation clustering model. We investigate both minimization and maximization variants. In the minimization variant, a clustering configuration of the correlation clustering game is in Nash equilibrium if no user can unilaterally reduce its cost $\sum_{u\in C_v} w_u \cdot d(u,v) + \sum_{u\notin C_v} w_u \cdot (1-d(u,v))$ by choosing a different cluster (respectively, in the maximization variant, a user cannot increase its profit $\sum_{u\in C_v} w_u \cdot (1-d(u,v)) + \sum_{u\notin C_v} w_u \cdot d(u,v)$). For ease of notation, given a cluster C, we denote the total weight of its nodes by $w(C)$. We denote by V the set of elements, and by E the set of all pairs of elements (E is the set of edges in a complete graph having V as the set of nodes). Due to space limitations, most proofs of this section are omitted from this extended abstract.

The following lemma shows that the two variants are closely related.

LEMMA 3.1. *A configuration of the game is in Nash equilibrium for the minimization variant if and only if it is in Nash equilibrium for the maximization variant.*

Correlation clustering is a hedonic game and it is closely related to the class of additively-separable hedonic games [7]. It can be shown that for every correlation clustering game, there exists an additively-separable hedonic game with the same set of Nash equilibria. However, the reverse is not true, i.e., additively-separable hedonic games strictly generalize correlation clustering games. We omit the details.

Another interesting subclass of additively-separable hedonic games are symmetric additively-sparable hedonic games [7]. A potential function argument shows that every symmetric additively-sparable hedonic game has a Nash equilibrium, but finding it is PLS-hard [13]. Although these results do not extend immediately to correlation clustering games, their techniques can be used for this type of games as well. The next two theorems follow from the use of these techniques.

THEOREM 3.2. *There always exists Nash equilibrium for the correlation clustering game. Moreover, best response dynamics of this game always converge to a Nash equilibrium.*

THEOREM 3.3. *Computing Nash equilibrium in the correlation clustering game is PLS-Complete.*

In the rest of this section we focus on the price of stability and price of anarchy of correlation clustering games. Unlike the set of Nash equilibria, these values are different for each of the game classes mentioned above, and therefore, we cannot use the relations between our model and these classes to derive results in this context. We begin with several observations on the price of stability.

LEMMA 3.4. *In the special case where all elements have an identical weight w, the price of stability is* 1.

PROOF. In the case of identical weights, the potential function Φ from the proof of Theorem 3.2 can be rewritten as

$$\Phi(\mathcal{F}) = w^2 \sum_{v\in V} \left(\sum_{u\in C_v} d(u,v) + \sum_{u\notin C_v} (1-d(v,u)) \right) .$$

Note that given a clustering configuration $\mathcal{F}$, the value of the potential function is identical to the value of the social objective function up to a multiplicative factor of w. This implies that any best response move performed by a player will strictly decrease the social objective value. Thus, an optimal clustering configuration is also a Nash equilibrium, as no player can further reduce its cost. The proof for the maximization variant follows directly from Lemma 3.1. □

In the case of arbitrary weights of elements, the price of stability differs between the minimization and maximization variants and can be strictly larger than 1, as shown in the following example. Consider the graph depicted in Figure 3.

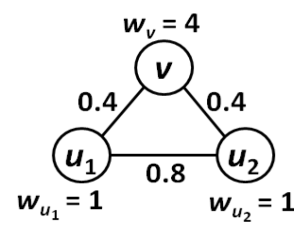

Figure 3: Graph with price on stability greater than 1.

Table 1: Price of stability

Variant	Nash Equilibrium	Optimal Configuration	Price of Stability
Minimum	5.6	5.4	≈ 1.037
Maximum	6.4	6.6	1.03125

The similarity between any two nodes appears on the edge between them, and the weight of a node appears next to it.

It is easy to verify that Nash equilibrium in this graph only occurs when all nodes share a single cluster. However, the optimal configuration is when v shares a cluster with either u_1 or u_2, and the third node is in a different cluster. Table 1 presents the price of stability provided by this example for both variants.

We now turn to analyze the price of anarchy. Despite the close relationship between the minimization and maximization variants of the game, the results obtained for their price of anarchy are quite different, as shown in the next sections. Before turning to analyzing each variant, we present general properties of Nash equilibrium which are later used for establishing the bounds on the price of anarchy for both game variants.

Properties of Nash Equilibria.

We prove two lemmata bounding distances in Nash equilibria.

LEMMA 3.5. *Consider two nodes u and v. If there exists a Nash equilibrium where u and v share a common cluster C, then $d(u, v) \leq 1 - \frac{w_u + w_v}{2w(C)}$.*

LEMMA 3.6. *Consider two nodes u and v. If there exists a Nash equilibrium where u and v belong to two different clusters C_u and C_v, then $d(u, v) \geq \frac{w_u + w_v}{2(w(C_u) + w(C_v))}$.*

3.1 Price of Anarchy - Minimization Variant

We present an upper bound of $O(n^2)$ and a lower bound of $(n-1)$ on the price of anarchy of the minimization variant of the correlation clustering game. We begin with the special case of equal weights and then proceed to arbitrary weights. The next definition is used in the sequel.

DEFINITION 3.7. *Given a clustering configuration, an edge e is of one of two types: either it is an internal edge within a single cluster, or it is an external edge between two different clusters.*

THEOREM 3.8. *If all nodes have the same weight w, price of anarchy is at most $n - 1$.*

PROOF. Without loss of generality, assume $w = 1$. Let $\mathcal{E}$ be a Nash equilibrium of the game, and let $\mathcal{O}$ be an optimal solution. In order to evaluate the contribution of an edge $e = (u, v)$ (of distance $d(u, v)$) to the total cost of $\mathcal{E}$ and $\mathcal{O}$, there are four cases to be considered.

- e is an internal edge in $\mathcal{E}$ and $\mathcal{O}$. Then, e contributes a cost of $2d(u, v)$ to both $\mathcal{E}$ and $\mathcal{O}$.
- e is an external edge in $\mathcal{E}$ and $\mathcal{O}$. Then, e contributes a cost of $2(1 - d(u, v))$ to both $\mathcal{E}$ and $\mathcal{O}$.
- e is an internal edge within cluster C in $\mathcal{E}$, but is an external edge in $\mathcal{O}$. Since e is an edge within C in a Nash equilibrium, Lemma 3.5 implies that $d(u, v) \leq 1 - 2/2w(C) \leq 1 - 1/n$. Therefore, the cost contribution of e to $\mathcal{E}$ is $2d(u, v) \leq 2(1 - 1/n)$. On the other hand, the cost contribution of e to $\mathcal{O}$ is $2(1 - d(u, v)) \geq 2/n$. Thus, the ratio between the cost contribution of e to $\mathcal{E}$ and its contribution to $\mathcal{O}$ is at most $\frac{2(1-1/n)}{2/n} = n - 1$.
- e is an external edge between two clusters C_1 and C_2 in $\mathcal{E}$, but is an internal edge in $\mathcal{O}$. Since e is an edge between clusters in a Nash equilibrium, Lemma 3.6 implies that $d(u, v) \geq 2/(2w(C_1) + 2w(C_2)) \geq 1/n$. Therefore the cost contribution of e to $\mathcal{E}$ is $2(1 - d(u, v)) \leq 2(1 - 1/n)$. On the other hand, the cost contribution of e to $\mathcal{O}$ is $2d(u, v) \geq 2/n$. Thus, the ratio between the cost contribution of e to $\mathcal{E}$ and its contribution to $\mathcal{O}$ is again at most $n - 1$.

It follows that the cost contribution of all edges to the cost of $\mathcal{E}$ is at most $n - 1$ times their contribution to the cost of $\mathcal{O}$, completing the proof. ☐

For general weights we are only able to prove a weaker result.

THEOREM 3.9. *The price of anarchy of the minimization variant of the correlation clustering game is $O(n^2)$.*

The following theorem establishes a lower bound of $(n-1)$ on the price of anarchy of the minimization variant. Note that it matches the upper bound for the special case of equal node weights.

THEOREM 3.10. *The price of anarchy of the minimization variant is at least $n-1$. This holds even when all nodes have the same weight and the metric is a line metric.*

PROOF. We show an instance in which the price of anarchy is at least $n-1$. Assume the weight of all nodes is 1. Let A and B be two disjoint sets of $n/2$ nodes. The distances between the nodes are defined as follows:

$$d(u, v) = \begin{cases} 0 & u, v \in A \text{ or } u, v \in B \\ 1/n & \text{otherwise} \end{cases}$$

This distance function is a special case of a line metric. Consider a clustering in which the nodes of A form one cluster and the nodes of B form another one. We show that this clustering is in Nash equilibrium. Due to symmetry, it suffices to show that a node $u \in A$ does not have an incentive to deviate. The cost of u under the current clustering is $(n/2 - 1) \cdot 0 + n/2 \cdot (1 - 1/n) = (n - 1)/2$. Node u has two deviation options. The first one is to move to the cluster of B which yields the same cost of $(n/2-1)\cdot 1 + n/2 \cdot 1/n = n/2 - 1 + 1/2 = (n-1)/2$. The other option is to form a new cluster, which increases the cost to $(n/2-1)\cdot 1 + n/2 \cdot (1-1/n) = n/2 - 1 + n/2 - 0.5 = n - 1.5$. Hence, the clustering is in Nash equilibrium. The social cost of the clustering solution is $n(n-1)/2$. On the other hand, the cost of the configuration in which all nodes belong to a single cluster is: $n \cdot [(n/2-1)\cdot 0 + n/2 \cdot 1/n] = n/2$. Thus, the price of anarchy of this game instance is at least $\frac{n(n-1)/2}{n/2} = n-1$. □

3.2 Price of Anarchy - Maximization Variant

In this section we provide tight bounds on the price of anarchy of the maximization variant of correlation clustering for general metrics and line metrics.

THEOREM 3.11. *The price of anarchy of the maximization variant is* $O(\sqrt{n})$.

The proof of Theorem 3.11 proceeds as follows. We first note that an upper bound on the maximum social welfare (total profit) is $(n-1)\cdot w(V)$. Then, we establish a lower bound of $\Omega(\sqrt{n})\cdot w(V)$ on the total profit of any Nash equilibrium solution.

THEOREM 3.12. *The price of anarchy of the maximization variant is* $\Omega(\sqrt{n})$. *This bound holds even if all nodes have the same weight and the metric is a tree metric.*

THEOREM 3.13. *The price of anarchy of the maximization variant in the case of a line metric is* $\Theta(n^{1/3})$. *Moreover, this bound is tight even if all nodes have unit weight.*

3.2.1 Bounding the Price of Anarchy

We suggest a method for bounding the price of anarchy (at the cost of making slight modifications to the rules).

LEMMA 3.14. *If only k clusters are allowed (for $k \geq 2$), then the price of anarchy is at most* k^{-1}.

COROLLARY 3.15. *Consider the case in which all nodes have equal weight, and best response dynamics is executed by first allowing it to reach Nash equilibrium in which nodes are limited to $k \geq 2$ clusters, and then it is allowed to continue till it finds a true Nash equilibrium. Then, the resulting Nash equilibrium has value of at least k^{-1} times the optimal social value.*

PROOF. By Lemma 3.14, the Nash equilibrium reached while the nodes are limited to k clusters has social value of at least OPT/k, where OPT is the optimal social value. Moreover, the proof of Lemma 3.14 actully shows that the social value of this Nash equilibrium is also at least W/k, where W is the maximal social value possible for any configuration (with any number of clusters). Using ideas similar to the proof of Lemma 3.4, we get that the social value cannot decrease by best response dynamics. Hence, the social value of the final Nash equilibrium is at least as large as the social value of the initial Nash equilibrium. □

Acknowledgements

We thank Ari Freund, Yossi Richter, and Elad Yom-Tov for helpful discussions. Special thanks go to Adam Kalai, Ronny Lempel, and Nir Ailon for fruitful comments on our work.

4. REFERENCES

[1] A. Abbasi and M. Younis. A survey on clustering algorithms for wireless sensor networks. *Computer Communications*, 30(14-15):2826–2841, June 2007.

[2] S. Bandyopadhyay and E.J. Coyle. An energy efficient hierarchical clustering algorithm for wireless sensor networks. In *Proceedings of INFOCOM*, pages 1713–1723. March 2003.

[3] S. Banerjee, H. Konishi and T. Sonmez. Core in a Simple Coalition Formation Game. *Social Choice and Welfare*, 18(1):135–153, 2001.

[4] N. Bansal, A. Blum, S. and Chawla. Correlation clustering. *Machine Learning J.*, 56(1-3):86–113, 2004.

[5] F. Bloch and E. Diamantoudi. Noncooperative formation of coalitions in hedonic games. *International Journal of Game Theory*, 40(2):263–280, 2010.

[6] A. Blum. Thoughts on clustering. In *NIPS Workshop on Clustering Theory*, 2009.

[7] A. Bogomolnaia and M.O. Jackson. The stability of hedonic coalition structures. *Games and Economic Behavior*, 38(2):201–230, 2002.

[8] M. Charikar, S. Guha, E. Tardos and D. Shmoys. A constant factor approximation algorithm for the k-median problem. In *Proceedings of STOC*, pages 1–10. ACM, 1999.

[9] M. Charikar, V. Guruswami and A. Wirth. Clustering with qualitative information. In *Proceedings of FOCS*, pages 524–533. IEEE Computer Society, 2003.

[10] E. Demaine and N. Immorlica. Correlation clustering with partial information. In *Proceedings of APPROX*, pages 71–80. Springer, 2003.

[11] J.H. Dreze and J. Greenberg. Hedonic coalitions: Optimality and stability. *Econometrica*, 48(4):987–1003, 1980.

[12] D. Emanuel and A. Fiat. Correlation clustering – Minimizing disagreements on arbitrary weighted graphs. In *Proceedings of ESA*, pages 208–220. Springer, 2003.

[13] M. Gairing and R. Savani. Computing Stable Outcomes in Hedonic Games. *Lecture Notes in Computer Science*, 10:174–185, 2010.

[14] T. F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985.

[15] L. Gourvès and J. Monnot. On Strong Equilibria in the Max Cut Game. In *Proceedings of WINE*, pages 608–615. 2009.

[16] T. Groves. Incentives in teams. *Econometrica*, 41(4):617–631, 1973.

[17] D. S. Hochbaum and D. B. Shmoys. A best possible heuristic for the k-center problem. *Mathematics of Operations Research*, 10:180–184, 1985.

[18] M. Hoefer. Cost Sharing and Clustering under Distributed Competition. Ph.D. Thesis, Universität Konstanz, 2007.

[19] G. Koltsidas and F. N. Pavlidou. A game theoretic approach to clustering of ad-hoc and sensor networks. *Telecommunication Systems*, 47(1-2):81–93, June 2011.

[20] M. Pelillo. What is a cluster? perspectives from game theory. In *NIPS Workshop on Clustering Theory*, 2009.

[21] S. R. Bulò. A game-theoretic framework for similarity-based data clustering. Ph.D. Thesis, Università Ca' Foscari di Venezia, 2009.

[22] A. Tamir. An $O(pn^2)$ algorithm for the p-median and related problems on tree graphs. *Operations Research Letters*, 19(2):59-64, August 1996.

[23] V. Vazirani. Approximation Algorithms. Springer Verlag, 2001.

[24] E. B. Yanovskaya. Equilibrium points in polymatrix games. (in Russion) *Litovskii Matematicheskii Sbornik*, 8:381–384, 1968.