On the Efficiency and Accuracy of a Community Detection Algorithm based on Hedonic Games

Konstantin Avrachenkov
INRIA, Sophia Antipolis, France

Lucas Lopes Felipe
UFRJ, Rio de Janeiro, Brazil

Daniel Sadoc Menasché
UFRJ, Rio de Janeiro, Brazil

ABSTRACT

Community detection is one of the most fundamental problems in network science and machine learning. In this paper, we leverage a connection between community detection algorithms and a subclass of hedonic games which admits a potential. Then, we (a) show that the considered community detection algorithm converges in polynomial time to an equilibrium and (b) analyze convergence towards accurate solutions. Among our findings, we discover that hedonic games are efficient to track communities under synthetic networks, being superior to spectral clustering, in terms of efficiency and accuracy, when considering the problem of community tracking subject to up to 25% of noise. To achieve high accuracy, we introduce the novel concept of equilibrium robustness with respect to a resolution parameter.

1. INTRODUCTION

Community detection is one of the most fundamental problems in network science and machine learning. In essence, it consists of finding groups of related elements, possibly without previous knowledge about the ground-truth concerning the groups that each element is part of [1, 2, 6]. In this paper, we focus on the problem of finding a single community in a network, sometimes referred to as *graph bisection*, which consists of partitioning the vertices of a given network into two disjoint components of initially unknown sizes, comprising a community and the remainder of the network.

Among the questions pertaining community detection, two of the most basic ones are related to efficiency and accuracy:

Efficiency: given a community detection algorithm, and a graph as input, *how long* will it take, in the worst case, for the algorithm to converge to a candidate community?

Effectiveness: in face of ground truth, given a candidate solution, how *accurately* does it capture a community?

Although the questions above are very basic, their answers are still broadly open. Most of the work on community detection algorithms is either heuristic or amenable to asymptotic analysis with respect to accuracy. Heuristic solutions involve algorithms which work well in practice, but for which no guarantees about accuracy or time to convergence are provided. Algorithms whose asymptotic accuracy is guaranteed when the number of nodes grows large typically account for stylized models such as the stochastic block model (SBM) [12]. In series of recent papers, Hajek *et al.* investigate polynomial time algorithms which fall in the latter category (see [7, 6] and references therein). In this paper, our aim is to analyze a community detection algorithm that is analytically tractable with respect to its time to converge and empirically accurate. In addition, we aim at a methodology that is based on first principles, i.e., for which the rationale behind the dynamics towards convergence can be interpreted as a game and whose solutions are selected based on their robustness with respect to the model parameters.

Our key insight to tackle the aforementioned goal is to leverage a connection between community detection algorithms and potential games, as established in [2]. In particular, we restrict ourselves to a subclass of hedonic games which admits a potential, henceforth referred to as hedonic potential [2, 8]. Hedonic games are games wherein each node has a corresponding utility for being part of a given community, known as its hedonic value, which represents how well it fits inside the community. From a computational perspective, the hedonic value of hedonic potential games is a function that can be computed in polynomial time, based solely on information about the number of neighbors of a node and the number of nodes in each community.

Contributions. First, we show that the considered community detection algorithm based on hedonic potential games converges in polynomial time to an *equilibrium*, i.e., to a configuration wherein no agent has incentives to change from its community. The proof leverages properties of the potential function. By deriving a lower bound on the potential gain per step, noting that the potential is bounded, we conclude that convergence occurs in polynomial time.

Second, we analyze convergence towards *good* solutions. A good solution is one that resembles a community in the ground truth. Although hedonic games have already been considered for community detection purposes [2], one of the key challenges for their adoption involves the parametrization of the algorithm. One of the parameters, denoted by α, captures the preference of nodes to join a community wherein they have more neighbors. Larger (resp., smaller) values of α correspond to increased preference (resp., rejection) towards a community with more (resp., less) neighbors. Unfortunately, searching for the best value of α is non-trivial. Fortunately, we discovered in our numerical investigations that certain communities correspond to equilibria of the potential game for a broad range of values of α, and that those robust communities are competitive against the ones found by spectral methods.

Outline. In the following section we present the considered community detection algorithm, and the subsequent

Workshop on AI in Networks and Distributed Systems 2020 Milan, Italy

sections follow the above outline. Section 6 contains related work, and Section 7 concludes.

2. HEDONIC GAMES

In this section we begin by introducing basic terminology and assumptions considered throughout the work. Then, we report our results related to efficiency and effectiveness of the considered community detection algorithm in Sections 3 and 4, respectively. Theorem 1 states that the considered algorithm converges in polynomial time to an equilibrium, and Theorem 2 will be instrumental to find good equilibria in the following section.

Given graph $G = (\mathcal{V}, \mathcal{E})$, with $V = |\mathcal{V}|$ vertices and $E = |\mathcal{E}|$ edges, a community is intuitively described as a set of nodes that have many connections across them, and few connections with other nodes in the system. In what follows, we make this definition precise. To that aim, we begin by defining the notion of value of a node in a partition.

A community (or coalition) is a set $\mathcal{S}$ of nodes, $\mathcal{S} \subseteq \mathcal{V}$. A community is connected if any pair of nodes in $\mathcal{S}$ is connected through a path. A community detection algorithm partitions the network into subsets $\{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_K\}$ such that $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset$ for $i \neq j$ and $\cup_{i=1}^{K} \mathcal{S}_i = \mathcal{V}$. In this paper, except otherwise noted we assume $K = 2$. In that case, we may also refer to the partitioning as $\{\mathcal{A}, \mathcal{B}\}$ where $\mathcal{A}$ and $\mathcal{B}$ are the two communities, with $\mathcal{S}_1 = \mathcal{A}$ and $\mathcal{S}_2 = \mathcal{B} = \mathcal{V} \setminus \mathcal{A}$.

A community is internally stable if no node in the community can benefit from leaving the community. A community is externally stable if no node outside the community can benefit by entering the community. A partition is stable if no node has incentive to change from its community. To determine if a node has incentive to deviate from its current coalition, we define the value of a node.

Assume that i and j are in partition $\mathcal{A}$, i.e., $i, j \in \mathcal{A}$. The value of each node pair (i, j), where $i, j \in \mathcal{A}$ is given by [2]:

$$v_{ij} = \begin{cases} 1 - \alpha, & (i,j) \in \mathcal{E} \\ -\alpha, & (i,j) \notin \mathcal{E} \\ 0, & i = j \end{cases} \tag{1}$$

where $0 \leq \alpha \leq 1$.

Role of α***.*** Parameter α controls the *resolution* at which communities are detected. It weights the relevance of the presence of links against the absence of those when determining the value of a node pair. Small values of α favor nodes to form coarse grained communities with more links, possibly reaching a grand coalition, which comprises all nodes in a single community. This occurs because when $\alpha = 0$ there is no incentive for nodes to isolate themselves, or for nodes to connect to isolated nodes, as an isolated node has zero value as opposed to a positive value if connected to its neighbors. Large values of α, in contrast, weight a penalty for the absence of links, favoring the creation of fine grained small communities, as isolated nodes may have incentives to remain isolated given that they have zero value, as opposed to negative value if they join a community with few neighbors.

Choice of α***.*** The choice of α is critical for the performance of community detection based on hedonic games. In previous works, authors have considered different strategies to select *the best* value for α [8]. In this paper, in contrast, we take a different perspective on the problem, and search for equilibria that hold *for all* α.

Hedonic value. Consider a node i in community $\mathcal{A}$. The hedonic value of node i is given by

$$v_i = \sum_{(i,j):j \in \mathcal{A}} v_{ij} = d_i' - (N_i - 1)\alpha \tag{2}$$

where d_i' is the degree of node i only accounting for its connections at community $\mathcal{A}$ and N_i is the number of nodes in the cluster of node i, $N_i = |\mathcal{A}|$. $N_i - 1 - d_i'$ is the number of "strangers", i.e., the number of nodes in $\mathcal{A}$, apart from i, minus its degree at that cluster.

Note that the considered games admit additively separable preferences, i.e., the value of the community is the sum of the values of node pairs. The value (or potential) of a community is given by $\varphi_k = \frac{1}{2} \sum_{i:i \in k} v_i$. The value of a configuration (or partitioning) is given by $\Phi = \sum_k \varphi_k = \frac{1}{2} \sum_{k \in \mathcal{K}} \sum_{i,j \in \mathcal{S}_k} v_{ij}$. When considering two communities $\mathcal{A}$ and $\mathcal{B} = \mathcal{V} \setminus \mathcal{A}$, $2\Phi = \sum_{i,j \in \mathcal{A}} v_{ij} + \sum_{i,j \in \mathcal{B}} v_{ij}$.

Definition 1. *A solution to the community detection problem is a partitioning wherein no node has incentive to move in (or out) to (or from) any cluster, i.e., the value of every node in each community is larger than its value if it transitions to another community.*

Given the above comments regarding the additive nature of the potential, a solution to the problem may also be characterized based on the potential.

Proposition 1. *A solution to the community detection problem is a partitioning wherein no node can unilaterally transition to another community and increase the potential.*

In what follows, we show that a best-response algorithm is guaranteed to converge in polynomial time to a solution. Then, we consider the problem of equilibrium selection.

3. POLYNOMIAL TIME CONVERGENCE

Next, we state our main result concerning the algorithm convergence time. To that aim, we consider two assumptions: (a) α is rational, i.e., $\alpha = c/d$, for $c, d \in \mathbb{N}^+$ and (b) at each step of an iterative algorithm, each node accumulates a strictly positive gain in potential. The algorithm converges when no node has incentive to deviate.

Theorem 1. *Community detection based on hedonic games detects a community in* $O(V^2)$.

The proof of the above theorem is provided in [3]. In what follows, we briefly outline the basic insights involved. Note that the potential Φ satisfies $-V^2 \leq \Phi \leq V^2$. In addition, note also that the potential can be written as $\Phi = U/d$, where $U, d \in \mathbb{N}^+$. Now, consider a node j that can gain from moving between communities. The key step in the proof consists in showing that such move yields an instantaneous gain lower bounded by $1/d$. This lower bound together with the maximum and minimum values attained by Φ imply that the algorithm converges in at most $2dV^2$ steps.

From a theoretical point of view, it remains an open problem whether the above theorem still holds if α is not a rational number. From a practical standpoint, the above theorem states that the algorithm converges to a solution in polynomial time. In what follows, we consider the problem of equilibrium selection, i.e., the problem of finding an equilibrium that approximates well the ground truth.

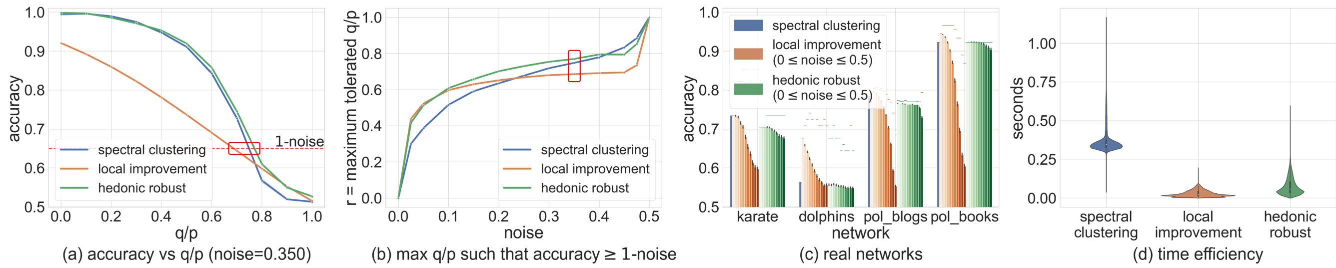


Figure 1: (a) In synthetic networks, under noise level of 0.35, hedonic games are beneficial for $q/p \leq 0.78$; (b) tolerating noise levels up to $\approx 35\%$; (c) results extend to real networks; (d) hedonic games are time-efficient.

4. ROBUSTNESS AND ACCURACY

Let $\mathcal{A}(P, (G = (\mathcal{V}, \mathcal{E}), \sigma))$ be the accuracy of a solution P to the community-detection problem (G, σ), where G is the graph and σ is the ground truth. Then,

$$\mathcal{A}(P, (G, \sigma)) = \sum_{i \in \mathcal{V}} \mathbb{1}(\sigma_i = P_i)/V \tag{3}$$

where $\mathbb{1}(c)$ is an indicator function which equals 1 if condition c holds and 0 otherwise. Note that as any solution to the problem still holds if we swap the 0's and 1's corresponding to the labels of the two clusters, it is assumed that the above equation is evaluated considering the most favorable alternative to maximize (3).

We refer to an equilibrium as $(\alpha_0, \alpha_1)-$*robust* if it remains an equilibrium after perturbing α over the $[\alpha_0, \alpha_1]$ range. We leverage the intuition that if an equilibrium is robust over a broader range it is more likely to accurately capture ground truth.

We illustrate the relationship between equilibrium robustness and accuracy through two simple toy examples. Figure 2(a) shows a 6 node network, comprising two cliques of 3 nodes connected through a single edge. Communities $\mathcal{A}$ and $\mathcal{B}$ correspond to nodes $\{1, 2, 3\}$ and $\{4, 5, 6\}$, respectively. If $\alpha = 0$, the hedonic value of each node equals 2, which is the number of neighbors of each node in its cluster. If one of the nodes 1, 3, 4 or 6 deviates to the other community, its hedonic value decreases to 0. If nodes 2 or 5 deviate, their hedonic value decreases to 1. It is worth noting that for $\alpha = 0$ the configuration wherein all nodes are in the same community (grand coalition) is also an equilibrium (Fig. 2(b)). However, the latter equilibrium is less robust than the former, holding for $\alpha \in [0, 0.4]$. As another example, consider the network in Figures 2(c)-(e). The equilibrium that minimizes the number of edges across communities, different from the grand coalition, is the most robust.

The following theorem indicates that to show that a given equilibrium holds for a range of values of α, it suffices to consider the two extremes of the range.

THEOREM 2. *Given an equilibrium Q which holds for $\alpha = \alpha_0$ and $\alpha = \alpha_1 > \alpha_0$, Q is an (α_0, α_1)-robust equilibrium.*

While searching for robust equilibria, the following theorem suggests that $\alpha = 0$ is instrumental to find candidates.

THEOREM 3. *An equilibrium for $\alpha = 0$ wherein the two partitions have same size is an equilibrium for any $\alpha \in [0, 1]$.*

The proof of the above theorem involves showing that equlbria for $\alpha = 0$ wherein the two partitions are balanced,

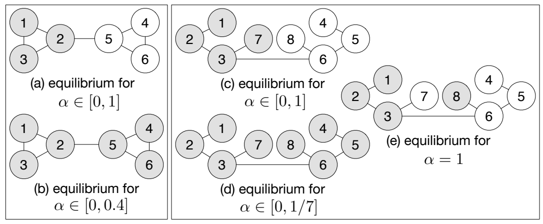


Figure 2: Robustness of hedonic equilibria.

e.g., Figs. 2(a) and 2(c), are also equilibria for $\alpha = 1$. The result then follows from Theorem 2. In what follows, we evaluate robust equilibria in synthetic and real networks.

5. EVALUATION

To evaluate hedonic community detection, we consider a setup motivated by the tracking of a community in a dynamic graph. Our goals are to (a) numerically indicate the efficiency of hedonic clustering; (b) establish the relationship between robustness and accuracy and (c) illustrate how the dependency of hedonic clustering with respect to its initialization can be leveraged for tracking purposes, assuming that a community evolves over time and that, at each time slot, each node switches community with probability f. We refer to f as the *noise level.*

In our evaluation of the hedonic strategy, we stick to *'practically' robust equilibria*, i.e., to partitions wherein (a) nodes are at equilibrium for $\alpha = 1$ and (b) at least 90% of the nodes have no incentive to deviate from their communities, for all $\alpha \in [0, 1]$. We also consider the *local improvement* strategy [1]: given an input partitioning, in one pass mark all nodes that would have more neighbors if they switch communities. Then, move all such nodes. For a description of the *spectral method* we refer the reader to [11].

We begin with synthetic networks. To determine ground truth, we consider the planted bisection model (PBM), a member of the SBM class. A graph $G = (\mathcal{V}, \mathcal{E})$ in the family of PBM comprises two communities of size n each, $|\mathcal{V}| = 2n$. Intra (resp., inter) edge communities occur with probability p (resp., q), with $p > q$. Note that q/p captures relative inter-community strength. The larger the ratio, the harder it is to identify communities. We vary p between 0.01 and 0.1 in increments of 0.01, and adjust q accordingly to obtain a pair (p, q) corresponding to one of the reported q/p ratios. For each q/p ratio, results are obtained with 95% confidence intervals (smaller than 0.01 in all considered cases) from 25 runs, each run comprising 25 network instances of size $n = 500$, for each of the ten considered (p, q) pairs.

Given ground-truth about the state of the community at a given slot, our goal is to estimate its following state. Let $f = 0.35$. Figure 1(a) shows how the accuracy of hedonic and spectral methods compare, noting the latter is insensitive to its initialization. As q/p increases, accuracy decreases as the problem becomes harder and exact recovery may be infeasible [1]. For all values of q/p hedonic community detection remains more accurate than the spectral method. Nonetheless, in the hardest settings ($q/p \geq 0.78$) a naive strategy which simply replicates the input into the output outperforms the three considered methods. Such a naive strategy yields an accuracy of $1 - f = 0.75$. Let r be the maximum tolerated ratio q/p such that for $q/p \geq r$ the naive strategy outperforms its alternatives. In Figure 1(a), $r = 0.78$ under the hedonic strategy, and Figure 1(b) shows how r varies as function of f.[1]In particular, Figure 1(b) suggests that the hedonic and spectral methods are complementary: whereas the first is more efficient and allows to track communities up to a noise level of $\approx 40\%$, one must rely on the latter when the noise surpasses such a threshold.

Additional evaluation results are reported in Figure 1(c), accounting for real networks (detailed in [3]). In such networks, we identified that robust equilibria are competitive against alternatives. In particular, in all the considered networks we were able to find robust equilibria whose accuracy outperforms spectral methods (marked as dots above the bars), and we are currently investigating approaches to select those equilibria as preferred choices. Figure 1(d) shows that hedonic methods are time-efficient, in agreement with our results in Section 3, being able to find robust equilibria in almost linear time.

6. RELATED WORK

Game theory provides a concrete multi-agent interpretation to the problem of community detection and its solution [2]. In this paper, we have indicated that a certain class of potential games admits an efficient algorithm to find equilibria, and compared its accuracy against spectral methods [11]. Modularity-based approaches [5] constitute an additional class of community detection methods.

The Louvain method [5], for instance, is an efficient approach for community detection. Our method differs from the Louvain approach in at least two aspects, as it (a) allows to pre-determine the number of communities in the network, and (b) is based on a multi-agent perspective towards the problem. In this paper we considered the problem of exact recovery of a community [1]. Alternatively, previous works also considered the problem of approximated recovery [12].

Spectral methods produce high-quality communities, but their applicability to large-scale problems is hampered by the computational complexity of $O(V^3)$. In [13] the authors propose heuristics to circumvent the complexity of spectral methods. Hedonic games are complementary to [13]. In particular, we indicate that for the problem of community tracking hedonic games may be the preferred choice as they allow us to leverage existing knowledge about the ground truth, and spectral methods [13] can be used to fully recompute the network state once prior knowledge is outdated.

We have shown that using the proposed game-based method one can find a candidate community in polynomial time. In addition, we have considered criteria for selecting "good" equilibria. Indeed, two of the main classical problems in the realm of computational game theory are determining (a) equilibrium complexity [4] and (b) equilibrium selection [10]. Equilibrium complexity generally refers to the problem of determining the complexity to find *a* Nash equilibrium. Equilibrium selection, in turn, is the problem of determining, among all equilibria, *the* best one. By establishing a connection between those two fundamental game theoretic problems and community detection, we envision further adopting tools from the former, e.g., evolutionary strategies of evolutionary games, to tackle the latter.

7. CONCLUSION

Community detection is a basic building block in the modern data science pipeline. In this paper, we have indicated that a simple algorithm inspired by hedonic-games, whose polynomial computational time corresponds to the complexity of finding a Nash equilibrium, is able to accurately track communities. We envision that this work opens up a number of interesting avenues for future investigation, in the frontier between game theory and community detection. In particular, we envision the extension of our work to detect multiple communities, each of which corresponding to a local maximum of the potential function. The comparison of robust equilibria as considered in this work against maximum likelihood solutions [1, 8, 9] is also as exciting topic for future work.

8. REFERENCES

[1] Abbe, E., Bandeira, A. S., and Hall, G. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory 62*, 1 (2015), 471–487.

[2] Avrachenkov, K., Kondratev, A., Mazalov, V., and Rubanov, D. Network partitioning algorithms as cooperative games. *Comp. Social Networks 5* (2018).

[3] Avrachenkov, K., Lopes, L., and Menasché, D. On the efficiency and accuracy of a community detection algorithm, 2020. `https://tinyurl.com/hedonicgit`.

[4] Daskalakis, C., Goldberg, P. W., and Papadimitriou, C. H. The complexity of computing a Nash equilibrium. *SIAM Journal on Computing 39*, 1 (2009), 195–259.

[5] Gach, O., and Hao, J.-K. Improving the Louvain algorithm for community detection with modularity maximization. In *Artificial Evolution* (2013), Springer.

[6] Hajek, B., et al. Recovering a hidden community beyond the Kesten–Stigum threshold in $O(|E| \log |V|)$ time. *Journal of Applied Probability 55*, 2 (2018), 325–352.

[7] Hajek, B., and Sankagiri, S. Community recovery in a preferential attachment graph. *IEEE Transactions on Information Theory 65*, 11 (2019), 6853–6874.

[8] Mazalov, V. V. Comparing game-theoretic and maximum likelihood approaches for network partitioning. In *Trans. Comp. Intelligence*. Springer, 2018, pp. 37–46.

[9] Prokhorenkova, L., and Tikhonov, A. Community detection through likelihood optimization. In *The Web Conference* (2019), pp. 1498–1508.

[10] Samuelson, L. *Evolutionary games and equilibrium selection*. MIT, 1997.

[11] Von Luxburg, U. A tutorial on spectral clustering. *Statistics and computing 17*, 4 (2007), 395–416.

[12] Xu, M., Jog, V., Loh, P.-L., et al. Optimal rates for community estimation in the weighted stochastic block model. *The Annals of Statistics 48*, 1 (2020), 183–204.

[13] Yan, D., Huang, L., and Jordan, M. I. Fast approximate spectral clustering. In *ACM SIGKDD* (2009), pp. 907–916.

[1]Source code to reproduce results together with animations and additional results available at [3].

APPENDIX

A. PROOF OF THEOREM 1

THEOREM 4. *Given a graph with V vertices, the proposed clustering algorithm based on hedonic games finds a solution in $O(V^2)$ steps.*

PROOF. Assumptions:

1. α is rational, i.e.,
$$\alpha = \frac{c}{d} \tag{4}$$
2. at each step, each node accumulates a strictly positive gain in potential.

The proof follows in three steps.

1. note that the potential Φ is upper bounded by $M = V^2$, i.e.,
$$\Phi \leq M = V^2 \tag{5}$$
In addition, note also that the potential can be written as
$$\Phi = \frac{U}{d}, \tag{6}$$
where U and d are integer numbers;
2. consider without loss of generality a node j that can gain from moving from one cluster B to cluster A and show that the instantaneous gain is at least $1/d$;
3. the two points above, together with (5) and (6), imply that the algorithm converges in at most $2dV^2$ steps. This is because in the worst case the initial value of U at the first iteration is $-V^2d$ and at the last iteration it is V^2d.

Next we complete the proof by showing that step (2) holds. To that aim, suppose that j is a candidate to move from cluster B to cluster A, with positive gain from such a move.

Let n_i be the number of nodes in cluster i, $i \in \{A, B\}$ and n_{1i} be the degree of j in i. Let n_{2i} be the number of nodes in i not connected to j.

$$n_{2A} = n_A - n_{1A}, \quad n_{2B} = n_B - n_{1B} - 1 \tag{7}$$
$$n = n_{2A} + n_{2B} + n_{1A} + n_{1B} + 1 = n_A + n_B \tag{8}$$

Let Δv_{ij} be the gain in value due to a move of node j from i,

$$\Delta v_{Bj} = (1-\alpha)n_{1A} - \alpha n_{2A} - (-\alpha n_{2B} + (1-\alpha)n_{1B}) \tag{9}$$
$$= -\alpha((n_{1A} + n_{2A}) - n_B + 1) + n_{1A} - n_{1B} \tag{10}$$
$$= -\alpha(n_A - n_B + 1) + n_{1A} - n_{1B} \tag{11}$$

In particular, note that if $\alpha \approx 1$ then $\Delta v_{Bj} = n_B - n_A - 1 + n_{1A} - n_{1B} = n_{2B} - n_{2A}$.

Then, if follows from (11) that

$$d\Delta v_{Bj} = -c(n_A - n_B + 1) + d(n_{1A} - n_{1B}). \tag{12}$$

Note that as we assume that node j has incentive to move from one cluster to the other,

$$\Delta v_{Bj} > 0 \Rightarrow d\Delta v_{Bj} \geq 1 \Rightarrow \Delta v_{Bj} \geq \frac{1}{d} \tag{13}$$

as desired. The first passage follows from the fact that $d\Delta v_{Bj}$ is an integer by assumption. □

B. PROOF OF THEOREM 2

THEOREM 5. *Given a configuration c that solves the clustering problem for α_0 and α_1, with $\alpha_0 < \alpha_1$, configuration c is also a solution for any $\alpha \in [\alpha_0, \alpha_1]$.*

PROOF. We consider a given node j of interest, that is part of cluster B. We will show that node B has no incentive to change from B to A for any $\alpha \in [\alpha_0, \alpha_1]$, as far as it does not have incentive to change for $\alpha = \alpha_0$ and $\alpha = \alpha_1$. In what follows, we use the same terminology as in Theorem 1 and leverage (11).

We consider two cases. First, if $n_A \geq n_B$, Δv_{Bj} decreases as α increases. If $\Delta v_{Bj}(\alpha_0) \leq 0$, then $\Delta v_{Bj}(\alpha^\star) \leq 0$, for $\alpha^\star > \alpha_0$. Similarly, if $n_A < n_B$, Δv_{Bj} increases as α decreases. If $\Delta v_{Bj}(\alpha_1) \leq 0$, then $\Delta v_{Bj}(\alpha^\star) \leq 0$, for $\alpha^\star < \alpha_1$.

We have just shown that the result holds for any node in cluster B. By symmetry, the same arguments apply for nodes in cluster A, which concludes the proof. □

C. PROOF OF THEOREM 3

Next, we establish a condition wherein an equilibrium for $\alpha = 0$ implies an equilibrium for $\alpha = 1$.

THEOREM 6. *An equilibrium for $\alpha = 0$ wherein the two partitions have same size is an equilibrium for any $\alpha \in [0, 1]$.*

PROOF. Let d_B and d_A be the degrees before and after a move, and let V_B be the number of nodes in the partition of the tagged node, again, before moving, and counting the tagged node. Let $_A$ be the number of nodes in the partition without the tagged node, i.e., the partition where the node will be found after its move, but without counting the tagged node. Then, $V_B + V_A = V$ where V is the number of nodes in the network.

Note that the degrees of the nodes are random variables. However, for the purposes of the analysis that follows, we can assume that the degrees before and after the move are constants.

Let B_α and A_α be the utility (hedonic value) of the tagged node before and after its move.

It follows from (2) that

- $\alpha = 0$: equilibrium if $B_0 \geq A_0$, i.e., $d_B \geq d_A$
- $\alpha = 1$: equilibrium if $B_1 \geq A_1$, i.e.,
$$d_B - (V_B - 1) \geq d_A - (V_A + 1 - 1). \tag{14}$$
When assuming partitions of same size in ground truth, $V_A = V_B = n$ and the condition translates to $d_B \geq d_A - 1$

Therefore, if partitions in ground truth have same size, and if partition is equilibrium for $\alpha = 0$, it implies it is also equilibrium for $\alpha = 1$

$$d_B \geq d_A \tag{15}$$
$$d_B \geq d_A - 1 \tag{16}$$

which concludes the proof. □

D. ADDITIONAL RESULTS ON REAL NETWORKS

We consider four real networks to evaluate hedonic-based community detection (see Table 1).[2] Political blogs and political books were two natural choices to consider, given the polarization in the realm of politics. In addition, we also consider the karate network comprising friendships between the 34 members of a karate club at a US university, and the dolphins network, an undirected social network of frequent associations between 62 dolphins. Reference ground truth about communities is available for the four considered networks, all comprising two communities each.

Accuracy. Table 2 reports the accuracy of hedonic-based community detection against spectral clustering. The hedonic clustering results correspond to the best accuracy obtained after 1,000 runs of Algorithm 1. In essence, it shows that for three out of the four considered networks there exists an hedonic-game equilibrium which outperforms the solution of spectral methods.

Equilibrium selection. Next, we consider the problem of equilibrium selection under real network. Table 2 suggests that the karate network and the political books network are at two extremes of the spectrum. For the first, we were not able to find any equilibrium which outperforms spectral clustering. For the latter, there exists an equilibrium which holds across all values of α. Indeed, this equilibrium outperforms the accuracy of spectral clustering, in agreement with the results obtained using synthetic networks and the PBM model as reported in the previous section.

Next, we further investigate the relationship between robustness of equilibrium with respect to α and accuracy, with a special focus on the dolphins and political blogs networks. To that aim, for each network, and for various values of α we find an equilibrium. For each of those equilibrium, we vary the value of α and assess how it impacts the fraction of nodes that have no incentive to deviate. Figure 3 reports our results. Each line in Figure 3 corresponds to an equilibrium. The thicker the line, the higher the accuracy. The ground truth is represented through dotted lines. For the karate and political books networks, we observe that indeed ground truth tends to be robust, in agreement with our previous observations. Nonetheless, for the dolphins and political blogs the ground truth configurations were less robust with respect to changes in α than some of their counterparts. We are currently investigating additional criteria for equilibrium selection to account for those cases, noting that for all the four considered networks at least one equilibrium competitive against spectral clustering exists (Table 2).

Algorithm 1: Community detection: best response to find local equilibrium

Input : $G = (V, E)$, α and initial partitioning $\{A, B\}$
Default values: $\alpha = 2|\mathcal{E}|/(|\mathcal{V}|(|\mathcal{V}| - 1))$, i.e., edge density, and
A and B of equal size, drawn uniformly at random
Output : Partitions $\mathcal{A}$ and $\mathcal{B} = \mathcal{V} \setminus \mathcal{A}$

changed $\leftarrow$ *true*
while *changed* **do**
    changed $\leftarrow$ $false$;
    for $i = 1 \to |\mathcal{V}|$ **do**
        let v be the i-th node from $\mathcal{V}$;
        let $\mathcal{P}$ be the partition wherein v is found, $\mathcal{P} \subseteq \{\mathcal{A}, \mathcal{B}\}$ and $\overline{\mathcal{P}} \leftarrow \mathcal{V} \setminus \mathcal{P}$;
        if *v has incentive to deviate* **then**
            $\mathcal{A} \leftarrow \mathcal{P} \setminus v$ and $\mathcal{B} \leftarrow \overline{\mathcal{P}} \cup \{v\}$;
            changed $\leftarrow$ *true*;
        end
    end
end
return $\mathcal{A}$ and $\mathcal{B}$

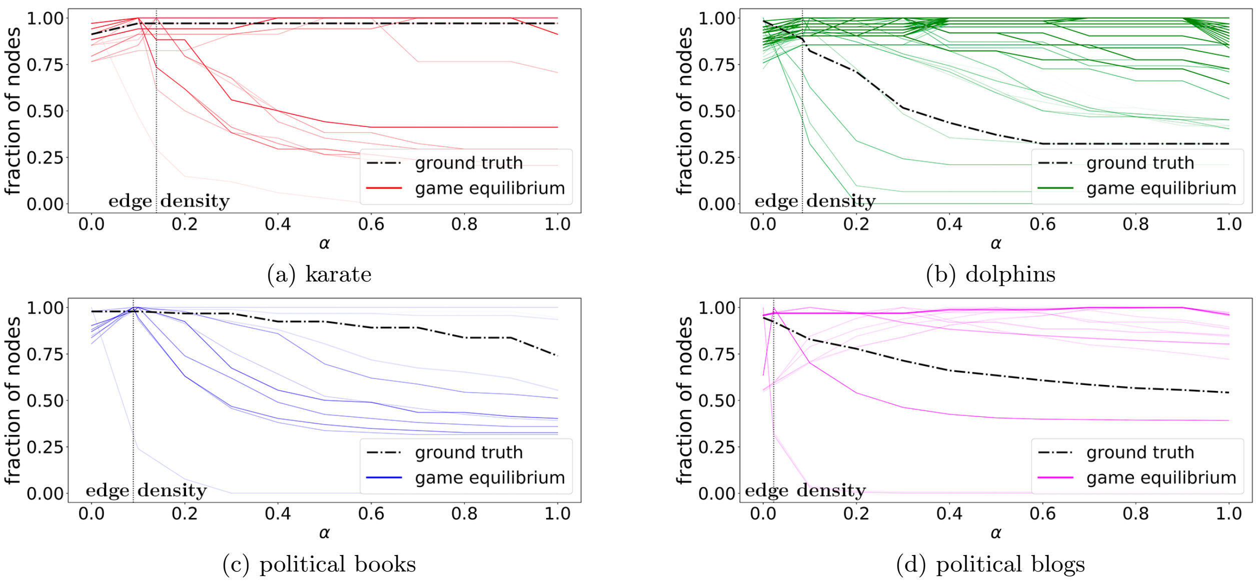
[2] Available at `http://www-personal.umich.edu/~mejn/netdata/`

Table 1: **Summary of real network properties**

Network	V	E	Edge Density $2E/(V(V-1))$	Vertex Proportion $\max(\lvert\mathcal{A}\rvert,\lvert\mathcal{B}\rvert)/V$
Karate	34	78	0.1390	50.00
Dolphins	62	159	0.0840	68.33
Political Books	92	374	0.0893	53.33
Political Blogs	1222	16714	0.0224	52.04

Table 2: **Summary of accuracy for real networks**

Network	Spectral	$\alpha=0$	$\alpha=$ Edge Density	$\alpha=0.5$	$\alpha=1$
Karate	97.05	94.11	94.11	94.11	94.11
Dolphins	82.25	98.38	91.93	91.93	91.93
Political Books	96.73	96.73	96.73	96.73	96.73
Political Blogs	93.37	95.66	94.84	94.92	95.17



(a) karate (b) dolphins (c) political books (d) political blogs

Figure 3: Fraction of nodes that have no incentive to deviate, for various values of α, in real networks: under ground truth (dotted black line) and under equilibrium (each of the other lines corresponds to an equilibrium for a given α, $\alpha = 0, 0.1, \ldots, 1.0$).