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

Konstantin Avrachenkov*, Lucas Lopes Felipe§, Daniel Sadoc Menasché§
*INRIA, Sophia Antipolis (France), §Federal University of Rio de Janeiro (Brazil)

***Abstract*—Community detection is one of the most fundamental problems in network science, machine learning and social network analysis. In this paper, we leverage a connection between community detection algorithms and a subclass of hedonic games which admits a potential. Then, we (i) show that the considered community detection algorithm converges in polynomial time to an equilibrium and (ii) analyze convergence towards accurate solutions. Among our findings, we discover that hedonic games are efficient to track communities under synthetic and real networks, being superior to state of the art methods, 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.**

***Index Terms*—Community detection, hedonic potential games**

I. Introduction

Community detection is one of the most fundamental problems in network science, machine learning and social network analysis [12], [14]. In essence, it consists of finding groups of related entities, possibly without previous knowledge about the ground-truth concerning the groups that each element is part of [1], [2], [8]. In this paper, we focus on the problem of dividing a network into two disjoint sets of nodes, also referred to as *graph bisection*, corresponding to two communities or 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 network as input, *how long* will it take, in the worst case, for the algorithm to converge to a candidate community?

Accuracy: given a candidate solution, how *close* is it to the ground truth?

Although the questions above are very basic, their answers are still broadly open. Most of the work on community detection algorithms is either heuristic, focusing both on efficiency and accuracy [3], [11], or amenable to asymptotic analysis with respect to accuracy [15]. The latter typically accounts for stylized network models such as the stochastic block model (SBM) [1], [16]. In a series of recent papers, Hajek *et al.* investigate polynomial time algorithms which also fall in the latter category (see [8], [9] 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], [10]. 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. From the social network perspective, the hedonic game approach is attractive as it is essentially local and represents well the limitation on acquiring and maintaining social connections [14].

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. In particular, we derive a lower bound on the potential gain per step which, together with the fact that the absolute value of the potential is bounded, implies that convergence occurs in polynomial time.

Second, we analyze convergence towards *good* solutions, i.e., that are close to 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. When a node decides which community to join, a parameter denoted by α captures the relevance of neighbors and non-neighbors in that choice. Smaller (resp., larger) values of α correspond to increased preference (resp., rejection) towards a community with more neighbors (resp., non-neighbors). Unfortunately, searching for the best value of α is non-trivial. Fortunately, we discovered in numerical investigations that certain communities correspond to equilibria of the potential game for a broad range of values of α and that those robust solutions are competitive against the ones found by state of the art methods [1], [3], [4], [11].

Outline. The following section presents the considered community detection algorithm, and Sections III to V evaluate it with respect to its efficiency and accuracy. Section VI contains related work, and Section VII concludes.

II. 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 accuracy of the considered community detection algorithm in Sections III and IV, respectively. Theorem 1 states that the considered algorithm converges in polynomial time to an equilibrium, and Theorems 2 and 3 are instrumental to find good equilibria.

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 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 has incentive to leave the community. A community is externally stable if no node outside the community has incentive to enter the community. A partition is stable if all its communities are internally and externally stable, and is referred to as an equilibrium. To determine if a node has incentive to deviate from its current community, we define the value of a node pair and the value of a node.

Given a network partitioning $\{\mathcal{A}, \mathcal{B}\}$, assume that nodes i and j are in same community, i.e., $i, j \in \mathcal{A}$ or $i, j \in \mathcal{B}$. Then, the value of node pair (i, j) 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 weighs 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 $\alpha = 0$ nodes incur no penalty for having non-neighbors in their communities. In that case, they join the community wherein their number of neighbors is maximized. When $\alpha = 1$, in contrast, nodes are penalized for the absence of links, favoring fine grained small communities. In that case, nodes join the community wherein their number of non-neighbors, or 'strangers', is minimized.

Choice of α**.** The choice of α is critical for the outcome of community detection based on hedonic games. In previous works, authors have considered different strategies to select *the best* value for α [2], [10]. In this paper, in contrast, we take a different perspective on the problem, and search for equilibria that are *robust* against changes in α, e.g., that hold for a large interval of values of α or even for all $\alpha \in [0, 1]$.

Hedonic value. Consider a node j in community $\mathcal{B}$. The hedonic value of node j is the sum of the values of the node pairs (i, j), for all $i \in \mathcal{B}$, and is denoted by v_j. Then, from (1),

$$v_j = \sum_{(i,j): i \in \mathcal{B}} v_{ij} = d_{Bj} - (n_B - 1)\alpha \tag{2}$$

where d_{Bj} is the degree of node j only accounting for its connections at community $\mathcal{B}$ and n_B is the number of nodes in the community of node j, $n_B = |\mathcal{B}|$. Let $\overline{d}_{Bj}$ be the number of 'strangers' to j, i.e., the number of nodes in $\mathcal{B}$, apart from j, minus its degree at that community, $\overline{d}_{Bj} = n_B - 1 - d_{Bj}$. Note that $v_j = d_{Bj}$ (resp., $v_j = -\overline{d}_{Bj}$) if $\alpha = 0$ (resp., $\alpha = 1$).

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 $\varphi_k = \frac{1}{2} \sum_{i: i \in A_k} v_i$. 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 (Equilibrium)**.** *A solution to the community detection problem is a partitioning wherein no node has incentive to move in (or out) to (or from) any community, i.e., the hedonic value of every node is larger than or equal to its hedonic value if it deviates 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.

III. POLYNOMIAL TIME CONVERGENCE

Next, we state our main result concerning the algorithm convergence time. To that aim, we consider two assumptions: (i) α is rational, i.e., $\alpha = b/c$, for $b, c \in \mathbb{N}$ and (ii) at each step of an iterative best-response algorithm, at least one node accumulates a strictly positive gain in potential. The algorithm converges when no node can increase the potential.

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

The proof of the above theorem is provided in Appendix A. 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 $\Phi = U/c$, where $U, c \in \mathbb{N}$. 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 $1/c$. This lower bound together with the maximum and minimum values attained by Φ imply that the algorithm converges in at most $2cV^2$ 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.

IV. Robustness and accuracy

Let $\mathcal{C}(P,(G=(\mathcal{V},\mathcal{E}),\sigma))$ 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,

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

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 communities, 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 for all $\alpha\in[\alpha_0,\alpha_1]$. 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 1(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 community. 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. 1(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 1(c)-(e): the equilibrium that minimizes the number of edges across communities, different from the grand coalition (Figure 1(d)), is the most robust (Figure 1(c)).

The following theorem, whose proof is found at Appendix B, 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 communities have same size is an equilibrium for any $\alpha\in[0,1]$.*

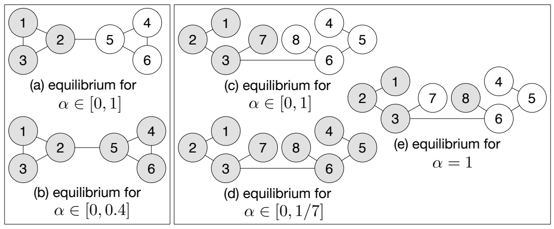


Fig. 1. Robustness of hedonic equilibria.

Let the *imbalance* of a partitioning be the difference between the sizes of its communities. In a *perfectly balanced* partitioning, communities have the same size. The proof of the above theorem (Appendix C) involves showing that perfectly balanced equilibria for $\alpha=0$, e.g., Figs. 1(a) and 1(c), are also equilibria for $\alpha=1$. Then, the result follows from Theorem 2.

V. Evaluation

To evaluate community detection based on hedonic games, we consider a setup motivated by the tracking of a community in a dynamic graph. Our goals are to (i) numerically indicate the efficiency of hedonic games for community tracking and detection; (ii) establish the relationship between robustness and accuracy and (iii) illustrate how the dependency of the hedonic game best-response algorithm with respect to its initialization can be leveraged for tracking purposes.

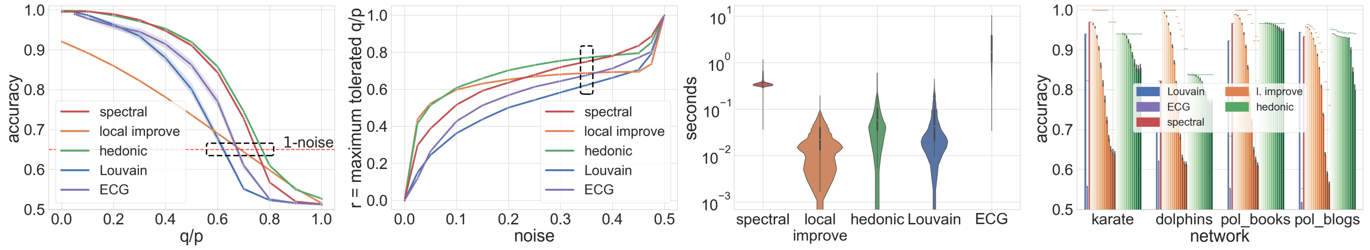
Motivated by [7], we consider the problem of tracking a community that evolves over time in a dynamic social network. At each time slot, each node switches community with probability f. We refer to f as the *noise level.*

In our evaluation of the hedonic strategy, we stick to *'practically' robust equilibria*, i.e., to partitions wherein (i) nodes are at equilibrium for $\alpha=1$ and (ii) at least 90% of the nodes have no incentive to deviate from their communities, for all $\alpha\in[0,1]$.

We compare the hedonic approach against four strategies:

1) the *spectral* method, which is known to be asymptotically optimal for a broad class of networks [15];
2) the *Louvain* algorithm [3], which is a modularity-based approach empirically known to perform well in practice;
3) a state of the art ensemble method (*ECG*) [11], which (i) generates an ensemble of k partitions, (ii) produces a re-weighted graph based on the ensemble, and (iii) performs a last partition on the re-weighted graph. In our experiments, we let $k=32$, which is the default value considered in [11];
4) a *local improvement* strategy [1] wherein, given an input partitioning, (i) in one pass mark all nodes that would have more neighbors if they switch communities and then, (ii) move all such nodes.

The three first strategies are insensitive to the initial network configuration, whereas local improvement and the hedonic game approaches can leverage the initial configuration for community tracking purposes. In addition, whereas the spectral, Louvain and local improvement algorithms are deterministic, ECG and hedonic game are randomized algorithms,



(a) accuracy vs q/p (noise=0.350) (b) max q/p such that accuracy ≥ 1-noise (c) time efficiency (d) real networks

Fig. 2. Community tracking: (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 $\approx 40\%$; (c) hedonic games are time-efficient; (d) results extend to real networks. The efficiency of hedonic games is close to local improvement, and accuracy is competitive against state of the art approaches by leveraging knowledge about the tracked community for initialization purposes.

TABLE I
NETWORKS ATTRIBUTES

Network	V	E	edges density	nodes per community	
Karate	34	78	0.1390	17	17
Dolphins	62	159	0.0840	20	42
Political Blogs	1222	16714	0.0224	586	636
Political Books	92	374	0.0893	43	49

involving random choices in the generation of ensembles and in the ordering of nodes for best-response, respectively.

A. Community tracking in synthetic networks

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 $p \geq q$. 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 0.1 in all considered cases) from ten (p, q) pairs, where each pair is evaluated in 25 network instances of size $n = 500$, each instance being evaluated under 25 random initial conditions.

Given ground-truth about the state of the community at a given slot, our goal is to estimate its following state at the upcoming slot. Let $f = 0.35$. Figure 2(a) shows how the accuracy of robust hedonic equilibria compares against the other four strategies. 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 its counterparts. Nonetheless, in the hardest settings ($q/p \geq 0.78$) a naive strategy which simply replicates the input into the output outperforms all the 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 2(a), $r = 0.78$ under the hedonic strategy, and Figure 2(b) shows how r varies as function of f.[1] In particular, Figure 2(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.

Efficiency. Figure 2(c) shows violin plots of the time to convergence of the considered strategies. The hedonic game method convergence time accounts for the time to find robust equilibria. Figure 2(c) indicates that the Louvain, local improvement and hedonic methods are time-efficient, being able to find robust equilibria in almost linear time. In particular, the low computational complexity of the hedonic method is in agreement with the results in Section III.

B. Community tracking in real networks

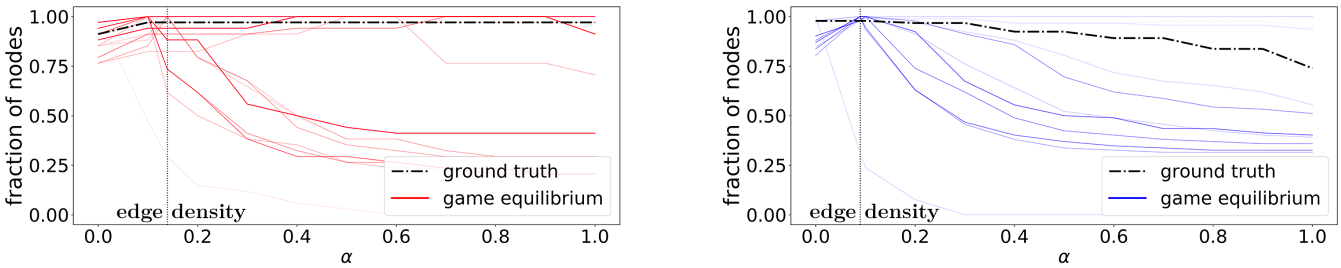
Next, we evaluate the considered community tracking approaches under four real networks.[2] Political blogs and political books are derived from human social networks, accounting for the polarization in the realm of politics, karate club is a small network of friendships, and dolphins is a biological social network (see Table I). Reference ground truth about communities is available for the four considered networks, all of them comprising two major communities each.

Evaluation results for real networks are reported in Figure 2(d). For each network, hedonic robust and local improvement strategies correspond to 11 bins each, accounting for noise level f varying between 0 and 0.5 in increments of 0.05. The spectral method, Louvain algorithm and ECG correspond to a single bin each, as they are insensitive to f. Each bar shows the average accuracy over 100 executions of the corresponding method, with 95% confidence intervals, and the maximum accuracy achieved by each of the considered methods is marked with horizontal dashes above the bars.

Together, Figures 2(c) and 2(d) indicate that the hedonic game approach presents a good compromise between efficiency and accuracy. As shown in Figure 2(d), in real networks the robust equilibria produced by the hedonic game approach are competitive against the four considered alternatives. Except for the dolphins network, the average accuracy of the hedonic game approach is within 0.05 of the highest average accuracy among the considered approaches. As the noise f increases, the accuracy of the hedonic game approach decreases, but such a decrease is smaller than that of the local improvement

[1] Source code and datasets to reproduce results together with animations and additional results available at https://github.com/lucaslopes/hedonic/

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



(a) karate (b) political books

Fig. 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, $\alpha = 0, 0.1, \ldots, 1.0$).

approach. Indeed, the hedonic game approach remains competitive in the setup with maximum noise level of $f = 0.5$, which is further investigated in the sequel.

C. Community detection in real networks

Next, we consider the problem of community detection without any prior knowledge of ground truth. In particular, we consider the same real networks as in the previous section, with two communities each, and assume that the hedonic game algorithm is initialized with each node having a 50% chance of being part of each of the two communities ($f = 0.5$).

The edge density of a network is the ratio of the number of existing edges over the possible number of edges, and was proposed as a reference for α in [2], [10]. Indeed, as indicated in the analysis that follows, setting α as the edge density typically produces accurate results. This observation, combined with the evidence that robust equilibria also tend to be accurate (Figure 2), motivates the search for robust equilibria starting from α around the edge density.

Role of robustness. Next, we further investigate the relationship between robustness of equilibrium with respect to α and accuracy. To that aim, for each network we find a set of equilibria for various values of α. For each 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 for the karate and political books networks. 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. For the dolphins and political blogs, the ground truth configurations were less robust with respect to changes in α than some of their counterparts. In particular, the ground truth in those networks is more unbalanced than under the karate and political books networks (see Table I and Section IV). In summary, taking edge density as a reference value for α, one can produce a set of equilibria and robustness is instrumental to further select among those, specially in scenarios with balanced partitionings.

VI. 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. Then, we compared its accuracy against state of the art methods, including modularity-based approaches [6], ensemble methods [11] and spectral methods which are known to be asymptotically optimal for a broad class of networks [15].

The Louvain method [3], [6], for instance, is an efficient approach for community detection. Our method differs from the Louvain approach in at least four aspects, as it (i) allows to pre-determine the number of communities in the network, (ii) is based on a multi-agent perspective towards the problem, (iii) accounts for equilibrium robustness with respect to a resolution parameter and (iv) can leverage prior knowledge about communities through its initialization, which is instrumental for tracking purposes.

Spectral methods produce high-quality communities, but their applicability to large-scale problems is hampered by the computational complexity of $O(V^3)$. In [17] the authors propose heuristics to circumvent the complexity of spectral methods. Hedonic games are complementary to [17]. 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. When such knowledge becomes outdated to the extent of becoming irrelevant, e.g., $f = 0.5$, spectral methods [17] or ECG [11] can sporadically be used to fully recompute the network state.

We have shown that the proposed game-based method can find a candidate community in $O(V^2)$ 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 (i) equilibrium complexity [5] and (ii) equilibrium selection [13]. Equilibrium complexity refers to the problem of determining the complexity to find *a* Nash equilibrium. Equilibrium selection is the problem of determining, among all equilibria, *the* best one. By establishing a connection between those 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.

VII. Conclusion

Community detection is a basic building block in the modern data science pipeline. In this paper, we investigated the efficiency and accuracy of a community detection algorithm based on hedonic games. First, we have shown that the considered algorithm is provably efficient: its polynomial computational time corresponds to the complexity of finding a Nash equilibrium. Then, we indicated its ability to accurately track and detect communities. In particular, the outcome produced by the proposed method can leverage previous knowledge about communities, through its initialization, which is instrumental for tracking purposes. By introducing the novel concept of equilibrium robustness with respect to a resolution parameter, we have shown that community tracking based on hedonic games outperforms state of the art solutions [2], [10]. We envision that this work opens up a number of interesting avenues for future investigation, in the frontier between game theory and community detection, including extensions to detect multiple communities, each of which corresponding to a local maximum of the potential function.

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] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment 2008*, 10 (2008), P10008.

[4] Chan, T. F., Ciarlet Jr, P., and Szeto, W. On the optimality of the median cut spectral bisection graph partitioning method. *SIAM Journal on Scientific Computing 18*, 3 (1997), 943–948.

[5] 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.

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

[7] Greene, D., Doyle, D., and Cunningham, P. Tracking the evolution of communities in dynamic social networks. In *Intl. conference on advances in social networks analysis and mining* (2010), IEEE.

[8] 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.

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

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

[11] Poulin, V., and Théberge, F. Ensemble clustering for graphs: comparisons and applications. *Applied Network Science 4*, 1 (2019).

[12] Prokhorenkova, L., and Tikhonov, A. Community detection through likelihood optimization. In *The Web Conference* (2019).

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

[14] Sheikholeslami, F., et al. Egonet tensor decomposition for community identification. In *GlobalSIP* (2016), IEEE, pp. 341–345.

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

[16] 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.

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

Appendix A
Proof of Theorem 1

Proof. Next we complete the proof outlined in the main body of the paper. To that aim, we consider, without loss of generality, a tagged node j that can gain from moving from community $\mathcal{B}$ to community $\mathcal{A}$. Our goal is to show that the instantaneous gain derived from such a move is at least $1/c$. This result, together with the fact that the potential is lower bounded (resp., upper bounded) by $-V^2$ (resp., V^2) implies that the algorithms converges in at most $2cV^2 = O(V^2)$ steps.

Recall that n_B is the number of nodes in the community $\mathcal{B}$ of the tagged node j, before moving, whereas n_A is the number of nodes in $\mathcal{A}$, with $n_B + n_A = V$. Recall also that d_{Bj} and d_{Aj} are the degrees of j in $\mathcal{B}$ and in $\mathcal{A}$, before and after a move, and $\overline{d}_{ij}$ is the number of nodes in i not connected to j, with $\overline{d}_{Aj} = n_A - d_{Aj}$ and $\overline{d}_{Bj} = n_B - d_{Bj} - 1$.

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

$$\begin{aligned}\Delta v_{Bj} &= (1-\alpha)d_{Aj} - \alpha\overline{d}_{Aj} - (-\alpha\overline{d}_{Bj} + (1-\alpha)d_{Bj}) \\ &= -\alpha(n_A - n_B + 1) + d_{Aj} - d_{Bj} \end{aligned} \tag{4}$$

It follows from (4) that $c\Delta v_{Bj} \in \mathbb{Z}$,

$$c\Delta v_{Bj} = -b(n_A - n_B + 1) + c(d_{Aj} - d_{Bj}). \tag{5}$$

Then,

$$\Delta v_{Bj} > 0 \Rightarrow c\Delta v_{Bj} > 0 \Rightarrow c\Delta v_{Bj} \geq 1 \Rightarrow \Delta v_{Bj} \geq \frac{1}{c}.$$

where the first inequality follows from the assumption that node j has incentive to deviate (assumption (ii)) and the second implication follows from the fact that $c\Delta v_{Bj} \in \mathbb{Z}$ (eq. (5) and assumption (i)). □

Appendix B
Proof of Theorem 2

Proof. We consider a given node j of interest, that is part of community $\mathcal{B}$. Then, we show that node j has no incentive to change from $\mathcal{B}$ to $\mathcal{A}$ for any $\alpha \in [\alpha_0, \alpha_1]$, as far as it does not have incentive to change for $\alpha = \alpha_0$ and $\alpha = \alpha_1$. We use the same terminology as in Theorem 1 and leverage (4). Indeed, $\Delta v_{Bj}(\alpha)$ is a linear function of α (see eq. (4)), so $\Delta v_{Bj}(\alpha_0) \leq 0$ and $\Delta v_{Bj}(\alpha_1) \leq 0$ imply $\Delta v_{Bj}(\alpha^\star) \leq 0$ for $\alpha^\star \in [\alpha_0, \alpha_1]$.

We have just shown that the result holds for any node in community $\mathcal{B}$. By symmetry, the same arguments apply for nodes in community $\mathcal{A}$, which concludes the proof. □

Appendix C
Proof of Theorem 3

Proof. We consider a given tagged node j, that is part of community $\mathcal{B}$, and use the same terminology as in Theorem 1. Let B_α and A_α be the hedonic value of the tagged node before and after its move. It follows from (2) and (4) that

- $\alpha = 0$: equilibrium if $B_0 \geq A_0$, i.e., $d_{Bj} \geq d_{Aj}$
- $\alpha = 1$: equilibrium if $B_1 \geq A_1$, i.e.,

$$d_{Bj} - (n_B - 1) \geq d_{Aj} - (n_A + 1 - 1). \tag{6}$$

When assuming communities of same size, $n_A = n_B = n$, the last condition translates to $d_{Bj} \geq d_{Aj} - 1$, which is implied by the condition $d_{Bj} \geq d_{Aj}$ which holds for $\alpha = 0$. Therefore, if two communities have same size, and comprise an equilibrium for $\alpha = 0$, they also comprise an equilibrium for $\alpha = 1$. □