

RESEARCH ARTICLE

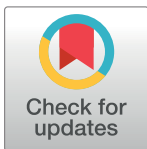
# Approximation of Nash equilibria and the network community structure detection problem

Suciu Mihai-Alexandru<sup>☉</sup>, Gaskó Noémi<sup>☉</sup>, Lung Rodica Ioana<sup>☉\*</sup>

Centre for the Study of Complexity, Babeş-Bolyai University, Cluj Napoca, Romania

☉ These authors contributed equally to this work.

\* [rodica.lung@econ.ubbcluj.ro](mailto:rodica.lung@econ.ubbcluj.ro)



## Abstract

Game theory based methods designed to solve the problem of community structure detection in complex networks have emerged in recent years as an alternative to classical and optimization based approaches. The Mixed Nash Extremal Optimization uses a generative relation for the characterization of Nash equilibria to identify the community structure of a network by converting the problem into a non-cooperative game. This paper proposes a method to enhance this algorithm by reducing the number of payoff function evaluations. Numerical experiments performed on synthetic and real-world networks show that this approach is efficient, with results better or just as good as other state-of-the-art methods.

## OPEN ACCESS

**Citation:** Mihai-Alexandru S, Noémi G, Ioana LR (2017) Approximation of Nash equilibria and the network community structure detection problem. PLoS ONE 12(5): e0174963. <https://doi.org/10.1371/journal.pone.0174963>

**Editor:** Wen-Bo Du, Beihang University, CHINA

**Received:** September 14, 2016

**Accepted:** March 19, 2017

**Published:** May 3, 2017

**Copyright:** © 2017 Mihai-Alexandru et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All network files are available at <https://data.mendeley.com/datasets/hcbzbhfdhm/draft?a=3cd86897-b3d9-4761-a84f-d6c0eedc6905> <http://dx.doi.org/10.17632/hcbzbhfdhm.1>.

**Funding:** This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI ([www.uefiscdi.gov.ro](http://www.uefiscdi.gov.ro)), project number PN-II-RU-TE-2014-4-2332.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Game theory plays an important role in modeling and solving real-world conflicting situations. In recent years game theory concepts have been used to address the network community structure detection problem as a game [1, 2]. This problem has multiple applications in economics, politics, sociology, biology, physics, and chemistry. Its importance arises from the fact that the community structure offers structural and functional information about the network that cannot be derived from other indicators. Intuitively, a community is described as a group of nodes that are highly connected to each other and sparsely connected to the outside. One of the main challenges related to this problem comes from the lack of a universally accepted formal definition for the community structure encompassing all aspects that emerge from the intuitive description above. A comprehensive survey of the problem, including possible definitions for the community structure, can be found in [3].

Game theory models strategic and conflicting situations offering a variety of solution concepts that are known as *game equilibria*. The most popular example is the Nash equilibrium, which models a situation of a game in which no player has any incentive for unilateral deviation [4]. The Nash equilibrium has been used to characterize the network community structure by considering a game in which nodes are players that strategically choose their community to maximize their payoffs. But this approach leads to multi-player games that

are difficult to solve when the payoff functions do not have nice mathematical properties. A heuristic model that computes Nash equilibria for such games has been proposed and adapted for the community detection problem in [1, 5] for static unweighted networks, and in [2] for dynamic networks. The model is based on extremal optimization and a generative relation for strategy profiles that enables their comparison [6], guiding the search towards game equilibria.

This paper proposes a reduced version of the generative relation—requiring less payoff function evaluations—for Nash equilibria characterization. We first provide a formal proof that this relation, which only takes into account a fraction of players when comparing two strategy profiles, can also characterize the Nash equilibria of the game. Numerical experiments are performed to evaluate this approach; results are compared with those provided by other state-of-art methods from the literature.

## Methods

The community structure detection problem is usually described as searching for a partition over the set of network nodes such that nodes within each set are *more* connected to each other than to nodes in other sets of the partition. While the concept of community as a social term is straight forward, its formalization is not trivial. Several attempts to define it can be found in the literature [7–14], but none of them captures all aspects embedded in the intuitive definition above.

A popular attempt defines the community structure as the optimum value of a fitness function defined to capture the network structure; the advantage of considering such a function is that heuristic methods can further be used to compute the optimal structure. Examples of such functions are: the *modularity* [15], the *community score* [16], the *community fitness* [17], and the *modularity density* [18]. However, none of these functions represent the real structure for all types of networks as their optimal values do not always correspond to the known community structure [19, 20].

A recent approach considers the community structure as the equilibrium of a mathematical game. Game theoretic approaches usually transform the community detection problem into a game in which nodes have the roles of players that have to choose a community to maximize their payoffs. The first game-based models use the modularity function [15] to derive payoffs for players [21, 22]. Game theoretic models have been proposed mostly for networks with overlapping communities [21, 23, 24] and for dynamic networks [2, 25]; models based on cooperative games theory have been proposed in [26–28]; other game based approaches can be found in [1, 5, 29].

In [29] the game is defined also over the set of nodes, i.e. nodes are players that maximize a payoff computed as the number of neighbors the node has in a community and a fraction of its neighbors that are connected among themselves. The authors propose a method to compute the Nash equilibria of this game, called Nash Stability based Community Detection (Nash-CoDe). Numerical results are reported in terms of modularity and coverage for a set of real-world networks.

Another formal game theoretic approach for unweighted networks constructs a hedonic game [22] in which nodes are the players and payoffs are computed using a local modularity function. The advantage of this approach is that the properties of the payoffs guarantee the existence of a Nash equilibrium for this game. The authors also design a community detection algorithm (CDG) that alternates three mechanisms to find the equilibrium. Numerical results are reported on a set of real-world networks and synthetic networks.

### Reduced nash ascendancy

Consider a game  $\Gamma = (N, S, U)$ , where  $N$  is the set of players,  $N = \{1, \dots, n\}$ ,  $S = \prod_{i=1}^n S_i$  is the set of strategy profiles of the game with  $S_i$  the set of strategies of player  $i$ , and  $U = \{u_i\}_{i \in N}$ ,  $u_i : S \rightarrow \mathbb{R}$  the player's payoff functions. A Nash equilibrium (NE) is a strategy profile such that there is no player that can improve his payoff by changing his strategy while all other players maintain theirs unchanged. Currently there are a plethora of methods that compute the Nash equilibria of a game. Among them an interesting class is formed of the methods based on evolutionary computation as they are supposed to be highly adaptable to various payoff functions and, more important, do not require nice mathematical properties for them [6, 30–32]. In [6] the basis for the direct search for Nash equilibria by evolutionary algorithms is designed by the proposal of a generative relation called the Nash ascendancy relation that can be used to characterize and furthermore compare strategy profiles of the game with respect to NEs. A quality operator  $k : S \times S \rightarrow \mathbb{N}$  computed as the number of players that would improve their payoffs by unilaterally switching their strategies from one strategy profile to the other is used:

$$k(s, q) = \text{card}\{i \in N \mid u_i(s) \leq u_i(q_i, s_{-i}), s_i \neq q_i\}, \tag{1}$$

where  $\text{card}\{\}$  denotes the cardinality of a set, and

$$(q_i, s_{-i}) = (s_1, s_2, \dots, s_{i-1}, q_i, s_{i+1}, \dots, s_n).$$

Operator  $k$  is used to define the Nash generative relation that enables the comparison of two strategy profiles with respect to the NE:

- if  $k(s, q) < k(q, s)$  we say that  $s$  Nash *ascends* (or Nash dominates)  $q$ , and that  $s$  is considered *better* than  $q$  in Nash sense;
- if  $k(s, q) > k(q, s)$  we say that  $s$  is *ascended* (or Nash dominated) by  $q$  in Nash sense;
- if  $k(s, q) = k(q, s)$  then  $s$  and  $q$  are indifferent to each other in Nash sense.

A strategy profile that is not ascended/dominated by any other is called in this context Nash non-ascended/non-dominated (NND). In [6] it is shown that if a game has at least one NE, then the NND set coincides with the set of NEs of the game. In the absence of Nash equilibria, the NND set can offer useful information about practical solutions that have properties similar to Nash equilibria to problems that were not previously approached by game theoretic tools.

However, the ascendancy relation requires up to  $2 \cdot n$  payoff function evaluations for each pair  $(s, q)$ ; this can become computationally expensive when dealing with large games. In this paper we are exploring the possibility of replacing the Nash ascendancy relation described above with a reduced version [33], which requires less payoff function evaluations when comparing two strategy profiles. The main result of this section is the formal proof that the reduced version also characterizes the Nash equilibria of the game, i.e. in the presence of the Nash equilibrium the set of non-dominated solutions with respect to this relation is equal to the set of equilibria.

The reduced relation considers only a fraction  $p$  of the nodes when computing  $k$ , where  $p \in (0, 1]$  such that  $n_p = [p \cdot n] > 0$ , where  $[\cdot]$  denotes the integer part. We denote by  $I_p \subset N$ ,  $I_p = \{i_1, i_2, \dots, i_{n_p}\}$  any set of  $n_p$  players, and  $\mathcal{I}_p \subset \mathcal{P}(N)$  the set of all subsets  $I_p$  of  $N$  having size  $n_p$ . Operator  $k_p : S \times S \times \mathcal{I}_p \rightarrow \mathbb{R}$  is defined as:

$$k_p(s, q, I_p) = \text{card}\{i \in I_p \mid u_i(s) \leq u_i(q_i, s_{-i}), s_i \neq q_i\}. \tag{2}$$

**Definition 1.** We say that strategy profile  $s$   $p$ -Nash ascends  $q$  with respect to  $I_p \in \mathcal{I}_p$  if

$$k_p(s, q, I_p) < k_p(q, s, I_p). \tag{3}$$

If  $p = 1$ , the  $p$ -Nash ascendancy relation is the same as the Nash ascendancy relation.

**Definition 2.** Strategy  $s \in S$  is  $p$ -Nash non-ascended ( $p$ -Nash non-dominated) if there is no  $q \in S$  and no  $I_p \in \mathcal{I}_p$  such that

$$k_p(q, s, I_p) < k_p(s, q, I_p).$$

We denote the set of  $p$ -Nash non-dominated solutions of  $\Gamma$  as  $p$ NND.

**Proposition 1.** For any  $s, q \in S$  we have:

$$\sum_{I_p \in \mathcal{I}_p} k_p(s, q, I_p) = C_{n-1}^{n_p-1} k(s, q), \tag{4}$$

where  $C_{n-1}^{n_p-1}$  denotes the binomial coefficient.

*proof.* Let  $k(s, q) = l$  and  $I = \{i_1, \dots, i_l\}$  the set of  $l$  players that can improve their payoffs by unilaterally deviating from  $s$  to  $q$ . Then each player  $i_j$  belongs to  $C_{n-1}^{n_p-1}$  sets  $I_p \in \mathcal{I}_p$  and we can write:

$$\sum_{I_p \in \mathcal{I}_p} k_p(s, q, I_p) = C_{n-1}^{n_p-1} \cdot l = C_{n-1}^{n_p-1} k(s, q).$$

**Proposition 2.** Any  $p$ -Nash non-dominated solution is a Nash non-dominated solution.

*proof.* Let  $s \in p$ NND, and  $q \in S$ . Then, based on Prop. 1, we have

$$k(s, q) = \frac{1}{C_{n-1}^{n_p-1}} \sum_{I_p \in \mathcal{I}_p} k_p(s, q, I_p) \leq \frac{1}{C_{n-1}^{n_p-1}} \sum_{I_p \subset N} k_p(q, s, I_p) = k(q, s) \tag{5}$$

meaning that  $s \in NND$ .

This result can be extended in the following manner:

**Proposition 3.** Let  $p, r \in [0, 1]$  with  $r > p$ . Then  $\forall I_r \in \mathcal{I}_r$ :

$$k_r(s, q, I_r) = \frac{1}{C_{n_r-1}^{n_p-1}} \sum_{I_p \subset I_r} k_p(s, q, I_p) \tag{6}$$

*proof.* Let  $k_r(s, q, I_r) = l, l < n_r$  and  $I = \{i_1, \dots, i_l\}$  the set of  $l$  players from  $I_r$  that can improve their payoffs by unilaterally deviating from  $s$  to  $q$ . Then each player  $i_j$  belongs to  $C_{n_r-1}^{n_p-1}$  subsets  $I_p \subset I_r$  and we can write:

$$\sum_{I_p \subset I_r} k_p(s, q, I_p) = C_{n_r-1}^{n_p-1} \cdot l = C_{n_r-1}^{n_p-1} k_r(s, q, I_r).$$

By considering  $r = 1$  in Proposition 3 we obtain the result in Proposition 1. Moreover, based on proposition 3 we can also formulate the following result:

**Proposition 4.** Let  $p, r \in [0, 1]$  with  $r > p$ . Then any  $p$ -Nash non-dominated solution is also a  $r$ -Nash non-dominated solution.

*proof.* The proof is straight-forward from Proposition 3 and similar to that of Proposition 2.

**Proposition 5.** Any Nash equilibrium of  $\Gamma$  is a  $p$ -Nash non-ascended solution,  $p \in (0, 1]$ .

*proof.* The proof is direct from the definition of the NE.

From Propositions 2, 5, and based on [6] it follows that  $p$ -Nash non-dominated solutions are also Nash equilibria for the game, if such an equilibrium exists. Thus, approximating  $p$ -

Nash non-dominated solutions leads to the same results as approximating Nash non-dominated solutions by using the Nash ascendancy relation.

### Community structure detection game

In this paper we approach the game proposed in [5] with the reduced variant of the generative relation used to characterize Nash equilibria. Given a network  $G = (V, E)$ , where  $V$  is the set of vertices and  $E$  is the set of edges, consider the following game  $\tilde{\Gamma} = (N, S, U)$  [5], where:

- $N$  is the set of players represented by network nodes,  $N = V = \{1, 2, \dots, n\}$ ; where  $n$  is the network size;
- $S = S_1 \times S_2 \times \dots \times S_n$  the set of strategy profiles, and  $S_i$  the set of strategies of player  $i$ ; in this case  $S_i$  represents the set of possible communities node  $i$  may choose from; an element  $s \in S$  is a possible community structure; we write  $s = (C_1, C_2, \dots, C_n)$ ,  $C_i$  represents the community of node  $i$ ,  $i \in \{1, 2, \dots, n\}$ ; we denote strategies  $C_i$  by using integers,  $C_i \in \{1, \dots, C_{max}\}$  where  $C_{max}$  is a maximum value expected for the number of communities in the network, and with  $C_w = \{i \in N | C_i = w\}$ ,  $w = 1, \dots, C_{max}$  the community containing all nodes having  $C_i = w$ ,  $i \in \{1, \dots, n\}$ ; for example,  $C_i = 2$  indicates that player  $i$  belongs to community 2 with all other nodes  $j \in N$  with  $C_j = 2$ , and we will denote by  $C_2$  the community containing all nodes  $i \in N$  with  $C_i = 2$ .
- $U$  is the payoff function,  $U = (u_1, u_2, \dots, u_n)$  where  $u_i : S \rightarrow \mathbb{R}$  is the payoff function of player  $i \in N$ , computed as the node's contribution to its community  $C_w$ , where  $C_i = w$  [1, 17]. We compute the payoff  $u_i(C_1, C_2, \dots, C_i, \dots, C_n)$  of node  $i$  relative to community  $C_w$  as the difference between the community fitness of  $C_w$  with node  $i$  included in it and its fitness when  $i$  is excluded:

$$u_i(C_1, C_2, \dots, C_i, \dots, C_n) = f(C_w \cup \{i\}) - f(C_w \setminus \{i\}), \text{ with } C_i = w. \tag{7}$$

The fitness  $f(C)$  of community  $C$  in Eq (7) [17] is:

$$f(C) = \frac{\sum_{j \in C} k_j^{in}}{(\sum_{j \in C} (k_j^{in} + k_j^{out}))^\alpha},$$

where  $k_j^{in}$  is the internal degree of node  $j$  in community  $C$  (the number of links connecting node  $j$  to other nodes in  $C$ ),  $k_j^{out}$  is the external degree of node  $j$  with respect to community  $C$  (the number of links connecting node  $j$  to nodes outside of  $C$ ), and  $\alpha$  is a parameter controlling the size of the community. In what follows we will consider  $\alpha = 1$ .

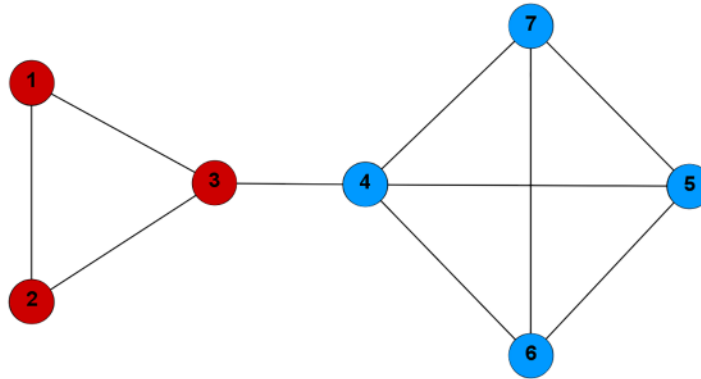
**Example 1.** Consider the network in Fig 1 and two different community structures (Figs 1 and 2).

In the first situation (Fig 1) we have:

$$f(C_1) = \frac{2 + 2 + 2}{(2 + 0) + (2 + 0) + (2 + 1)} = \frac{6}{7}, \quad f(C_2) = \frac{12}{13}$$

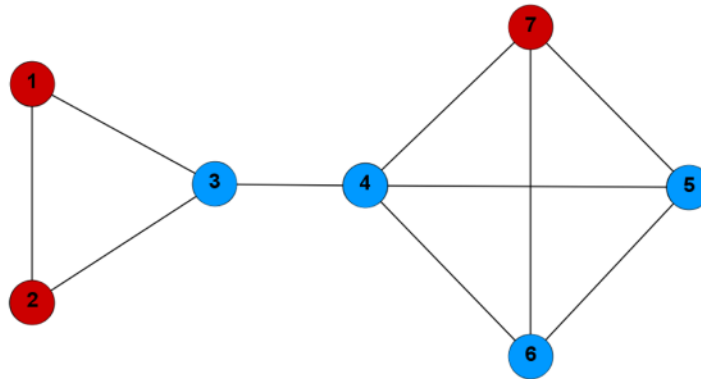
and the payoff of the third node (player) is the following:

$$u_3(1, 1, 1, 2, 2, 2, 2) = \frac{6}{7} - \frac{1 + 1}{(1 + 1) + (1 + 1)} = \frac{5}{14}.$$



**Fig 1.** The communities  $C_1 = \{1, 2, 3\}$  (red),  $C_2 = \{4, 5, 6, 7\}$  (blue) form the strategy profile  $s = (1, 1, 1, 2, 2, 2, 2)$ .

<https://doi.org/10.1371/journal.pone.0174963.g001>



**Fig 2.** The communities  $C'_1 = \{1, 2, 7\}$  (red),  $C'_2 = \{3, 4, 5, 6\}$  (blue) form the strategy profile  $q = (1, 1, 2, 2, 2, 2, 1)$ .

<https://doi.org/10.1371/journal.pone.0174963.g002>

In the second case (Fig 2):

$$f(C'_1) = \frac{2}{7}, \quad f(C'_2) = \frac{8}{13}.$$

The payoff for the third player is:

$$u_3(1, 1, 2, 2, 2, 2, 1) = \frac{8}{13} - \frac{2 + 2 + 2}{(2 + 2) + (2 + 1) + (2 + 1)} = \frac{1}{65}.$$

We have  $k(s, q) = 0$ , because  $u_3(s) \not\leq u_3(1, 1, 2, 2, 2, 2, 2)$  and  $u_7(s) \not\leq u_7(1, 1, 1, 2, 2, 2, 1)$  (we need to verify only these two inequalities, because  $s_i = q_i$  for  $i = 1, 2, 4, 5, 6$ ). On the other hand  $k(q, s) = 2$ , because  $u_3(q) \leq u_3(1, 1, 1, 2, 2, 2, 1)$  and  $u_7(q) \leq u_7(1, 1, 2, 2, 2, 2, 2)$ . So we can say that  $s$  Nash ascends  $q$ , or that  $s$  is better than  $q$  in Nash sense.

If we consider  $p = 0.50$  and  $I_p = \{3, 1, 5\}$  we obtain  $k_p(s, q, I_p) = 0$  and  $k_p(q, s, I_p) = 1$ , which means that  $s$  0.50-Nash ascends strategy profile  $q$  with respect to  $I_p$ .

While we cannot guarantee the existence of the Nash equilibrium for this game [21, 23], we can attempt to approximate it by using a search heuristic with the generative relation for characterization of Nash equilibria [6]. Particularly, the assumption that the NNDs of game  $\tilde{\Gamma}$  represent

the community structure of the network was empirically tested in [1] by means of numerical experiments performed on synthetic and real-world networks by using the Mixed Nash Extremal Optimization (MNEO) algorithm. MNEO showed superior performance compared to other methods in identifying the community structure for some networks considered difficult.

### Computation method: $p$ -Mixed Nash Extremal Optimization

In order to test the  $p$ -ascendancy relation, we adapt the Mixed Nash Extremal Optimization [1] to use of the  $p$ -ascendancy relation instead of the Nash ascendancy. Thus, the only modification made to MNEO, which is described in what follows, is the replacement, in line 4 of Algorithm 1, of the Nash ascendancy relation with the  $p$ -Nash ascendancy relation (Definition 1) with respect to a randomly generated set  $I_p$ . The particular set  $I_p \in \mathcal{I}_p$  of nodes that are tested whenever two individuals are compared by using the reduced Nash ascendancy relation is randomly generated each time a comparison is performed.  $p$  is a parameter of the method and we denote by  $p$ MNEO the reduced version of MNEO.

**Encoding.** Each individual  $P_j$  encodes a network structure as a vector of integers of size  $n$ ,  $P_j = (p_{j1}, \dots, p_{jn})$ , for node  $i$ ,  $p_{ji}$  represents the community of node  $i$  for individual  $j$ . Each individual searches for a fixed number of communities, which is set within a minimum possible value,  $C_{min}$ , and a maximum one,  $C_{max}$ .

**Populations.**  $p$ MNEO evolves a population  $P$  and an archive  $A$ ; individuals in  $A$  preserve the best solutions found so far by each individual in  $P$ . Paired individuals  $(P_j, A_j)$  from the two populations follow the rules of extremal optimization: individual  $P_j$  explores the search space; the best value found by  $P_j$  is preserved in  $A$ ,  $j = \overline{1, M}$ , where  $M$  is the size of the two populations.

**Extremal optimization.** Individuals in population  $P$  explore the search space following the rules of an EO-based iteration: for each individual  $P_j$ , the  $\ell$  nodes having the worst fitnesses are randomly assigned to different communities. If the newly obtained partition  $P'_j$   $p$ -Nash ascends the corresponding archive member  $A_j$  relative to a randomly generated  $I_p \in N$ , it will replace it. If not, the search continues next iteration from  $P'_j$  (Algorithm 1).

#### Algorithm 1 The Extremal Optimization iteration.

**Input:** Current population  $P$ , archive  $A$ ;  
**Parameters:** number of nodes to be changed  $\ell$ ;  
 Archive member  $A_j$ ,  $j = \overline{1, M}$ , preserves the best solution found so far by  $P_j$ .  
 1: **for** each individual  $P_j$  in  $P$  **do**  
 2:   Select the  $\ell$  nodes having the lowest payoffs from  $P_j$ ;  
 3:   Randomly assign another community to each selected node—create offspring  $P'_j$ ;  
 4:   **if**  $P'_j$   $p$ -Nash ascends  $A_j$  (Alg. 2) **then**  
 5:     set  $A_j := P'_j$ ;  
 6:   **end if**  
 7:   set  $P_j = P'_j$   
 8: **end for**

The  $p$ -Nash ascendancy relation is implemented as described in Algorithm 2. An important aspect of this relation is that at line 4 it tests if node  $i$  is in the same community in both  $P$  and  $A$ : if it is so, we do not attempt to compare payoff values, even though they might be different because of the rest of the structures, i.e. equal strategies in  $P$  and in  $A$  do not imply equal payoffs, but they are not compared as they are not included in  $k_p(P, A, \cdot)$  Eq (3).

**Algorithm 2**  $p$ -Nash ascendancy relation

**Input:** Individuals  $P$  and  $A$ ; Probability  $p$   
**Output:** 1 if  $P$   $p$ -Nash ascends  $A$ , -1 if  $A$  Nash ascends  $P$ , 0 if they are indifferent to each other.

```

1: Set  $k_p(P, A) = 0$ ;  $k_p(A, P) = 0$ ;
2: for  $h = 0$ ;  $h < n_p = np$ ;  $h++$  do
3:   Randomly select node  $i$ ;
4:   if  $P_i < > A_i$  then
5:     if  $u_i(P) < u_i(A_i, P_{-i})$  then
6:        $k_p(P, A)++$ ;
7:     end if
8:     if  $u_i(A) < u_i(P_i, A_{-i})$  then
9:        $k_p(A, P)++$ ;
10:    end if
11:  end if
12: end for
13: if  $k_p(P, A) < k_p(A, P)$  then
14:   return 1; ( $P$   $p$ -Nash ascends  $A$ )
15: else
16:   if  $k_p(A, P) < k_p(P, A)$  then
17:    return -1; ( $A$   $p$ -Nash ascends  $P$ )
18:   else
19:    return 0; ( $P$  and  $A$  are indifferent to each other)
20:   end if
21: end if

```

**Diversity preserving mechanism.**  $p$ MNEO uses the mixing mechanism used by MNEO designed to preserve the node degrees in the network and proposed in [34]. Two links are randomly chosen, deleted, and the nodes are re-connected in a different manner. Links are selected randomly with a probability  $\rho$  controlling the magnitude of change in the network.  $p$ MNEO alternates the search on the original network for  $\Lambda$  iterations with the search on the modified network for  $\lambda$  iterations.

**Outline of  $p$ MNEO.** Algorithm 3 presents an outline of  $p$ MNEO.

**Algorithm 3** Outline of  $p$ - Mixed Network Extremal Optimization.

```

1: Randomly initialize all individuals in  $P$  and  $A$ ;
2: Evaluate all individuals in  $P$  and  $A$ ;
3: for  $NrGen$  from 0 to  $MaxGen$  do
4:   Set  $\ell_{ngen} = \max\left\{1, \left[\frac{1}{10} \cdot n \cdot (n - 2)^{\frac{NrGen}{MaxGen}}\right]\right\}$ 
5:   Run an EO iteration (alg. 1);
6:   if  $\Lambda$  iterations were performed on the original network then
7:     Mix network with probability  $\rho$ ;
8:     Randomly re-initialize population  $A$ ;
9:     Evaluate all individuals in  $P$  and  $A$ ;
10:  end if
11:  if Network is mixed and  $\lambda$  iterations were performed then
12:    Restore network to original structure;
13:    Randomly re-initialize population  $A$ ;
14:    Evaluate all individuals in  $P$  and  $A$ ;
15:  end if
16: end for
17: Output: individual from  $A$  having the best modularity;

```



**Output.** The individual with best modularity  $Q$  [15]. The modularity function is not used to guide the search, but only to select the best individual in the final population and is computed as:

$$Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{d_i d_j}{2m} \right) \delta(C_i, C_j), \tag{8}$$

where the sum runs over all pairs of vertices  $i$  and  $j$ ,  $A$  is the adjacency matrix,  $m$  the total number of links in the network,  $d_i$  the degree of node  $i$ ,  $C_i$  the community of node  $i$  and  $\delta(C_i, C_j)$  equals 1 if nodes  $i$  and  $j$  belong to the same community and 0 otherwise. When two community structures are compared, a higher modularity value presumably indicates a better solution.

**Parameters.**  $p$ MNEO uses the following parameters:

- fraction  $p$  of nodes that are evaluated during the  $p$ -Nash ascendancy relation test;
- Population size  $M$ ;
- Maximum number of iterations  $MaxGen$ ;
- $\rho$ —network mixing probability;
- $\Lambda$  and  $\lambda$ —number of iterations MNEO performs the search on the original network and the modified one, respectively;
- $C_{min}$  and  $C_{max}$ —minimum and maximum number of communities searched for.

## Results

We illustrate the effect of using the reduced generative relation within  $p$ MNEO by means of numerical experiments performed on a Cournot oligopoly and on synthetic and real-world networks with known community structures.

### Experimental set-up

**Cournot oligopoly.** The Cournot oligopoly is a well known competitive market game, suitable as a benchmark for testing equilibria detection methods as it presents one known equilibrium, and it is scalable.

Let  $q_i, i = 1, \dots, n$  quantities of an homogeneous product—produced by  $n$  companies respectively. We consider the market clearing price as

$$P(Q) = a - Q,$$

where  $Q$  is the aggregate quantity on the market. We have

$$P(Q) = \begin{cases} a - Q, & \text{for } Q < a, \\ 0, & \text{for } Q \geq a. \end{cases}$$

If the total cost for the company  $i$  of producing quantity  $q_i$  is  $C(q_i) = cq_i$ , i.e. there are no fixed costs and the marginal cost  $c$  is constant,  $c < a$ , and that companies choose their quantities simultaneously, the payoff for the company  $i$  is its profit:

$$u_i(q_1, q_2, \dots, q_n) = q_i P(Q) - C(q_i).$$

If  $Q = \sum_{i=1}^n q_i$ , the Cournot oligopoly has one Nash equilibria that can be computed by

$$q_i = \frac{a - c}{n + 1}, \forall i \in \{1, \dots, n\}.$$

We will use the Cournot oligopoly to illustrate the behavior of pMNEO when considering different probabilities for the generative relation. However, since convergence to local equilibria is not an issue in this case, we do not need the mixing mechanism to ensure diversity. We denote by pMNEO<sup>-</sup> the algorithm used to perform numerical experiments on the Cournot oligopoly.

**Parameter settings for the Cournot oligopoly** We tested pMNEO<sup>-</sup> for 10, 50, 100, 500, and 1000 players, by using 10 individuals encoding a strategy profile of the game; in our experiments we set  $a = 24$ ,  $c = 9$ ,  $MaxGen = 10^4$ ,  $S = [0, 10]^n$ . Tested values for probability  $p$  are 0.25, 0.50, 0.75, and 1. Each generation we compute and report the distance to the Nash equilibrium of the game.

**Community structure detection. Benchmarks** pMNEO is tested on the GN and LFR synthetic benchmarks [35, 36] (generated using the code available at <https://sites.google.com/site/andrealancichinetti/software>) and on four real world datasets with known community structure: the bottle-nose *dolphin* network [37], the *football* network [36], the Zachary *karate* club network [38], and the *books* about US politics network (<http://www.orgnet.com>, last accessed 9/3/2015). A set of three networks with unknown community structure is also tested: *risk map* [39], *jazz* [40], and *contiguous USA* [41].

The GN benchmark consists of eight network sets having 128 nodes grouped in 4 communities of 32 nodes; each node has a degree of 16. Each set is characterized by the number of links a node has outside its community,  $z_{out} \in \{1, 2, \dots, 8\}$ , and contains 30 networks.

Three sets of LFR benchmarks were generated: one set having 128 nodes, and 2 with 1000 nodes, each of them consisting of 5 sets containing 30 networks with different mixing parameters  $\mu \in \{0.1, 0.2, \dots, 0.5\}$ . The mixing parameter  $\mu$  is the ratio between the number of links a node has outside its community and its degree. The LFR parameters are: average vertex degree 20, maximum vertex degree 50, community size [10, 50] for the S (*small*) set with 128 and 1000 nodes, and [20, 100] for the B (*big*) set with 1000 nodes.

**Parameters** pMNEO has only one specific parameter apart from MNEO, the fraction  $p$  of nodes used within the  $p$ -Nash ascendancy relation. Four values of  $p$  were tested: 25%, 50%, 75%, and 100%. When  $p = 100\%$ , pMNEO is the same with MNEO. The termination condition was set as the maximum number of generations for each benchmark, indicated with the numerical results for each set. In order to assess if there are differences between results obtained by using these values, all other parameters are maintained constant (with values indicated in [1]), i.e.: population size 50, probability  $\rho$  for the network mixing mechanism 0.02,  $\Lambda = 30$ , and  $\lambda = 10$ . The parameters  $C_{min}$  and  $C_{max}$  were set to include the real number of communities and to allow 25% of the population to use it as a parameter.

**Comparison with other methods** The results obtained by pMNEO are compared to the following state-of-the-art algorithms: OSLOM [42], Infomap [43], Modularity optimization (ModOpt) [44], and the Louvain method [45] (for these algorithms we use the code and parameter settings available at <https://sites.google.com/site/andrealancichinetti/software>, downloaded on March 2014). Each of these algorithms are known to be fast and efficient.

**Performance measure** A reliable performance measure for assessing the quality of the results offered by different community detection methods is the normalized mutual information (NMI)<sup>1</sup> [17]. The NMI can be used when the community structure is available; the higher the value of the NMI, the better the solution; the maximum value of 1 indicates that the

compared structures are identical. Thus, if a result is compared to the known structure, a NMI value of 1 indicates that the correct structure has been found.

NMI values obtained by different methods are compared by using the Wilcoxon signed-rank non-parametric test. Thus we consider the NMIs reported in 30 independent runs for each real network and for the 30 networks of each GN or LFR set, by each method. The Wilcoxon sign rank assesses if there is a significant difference between two sample means; the null hypothesis that two samples come from the same population can be rejected with a level of significance  $\alpha = 0.05$  if the computed  $p$ -value is smaller than 0.05.

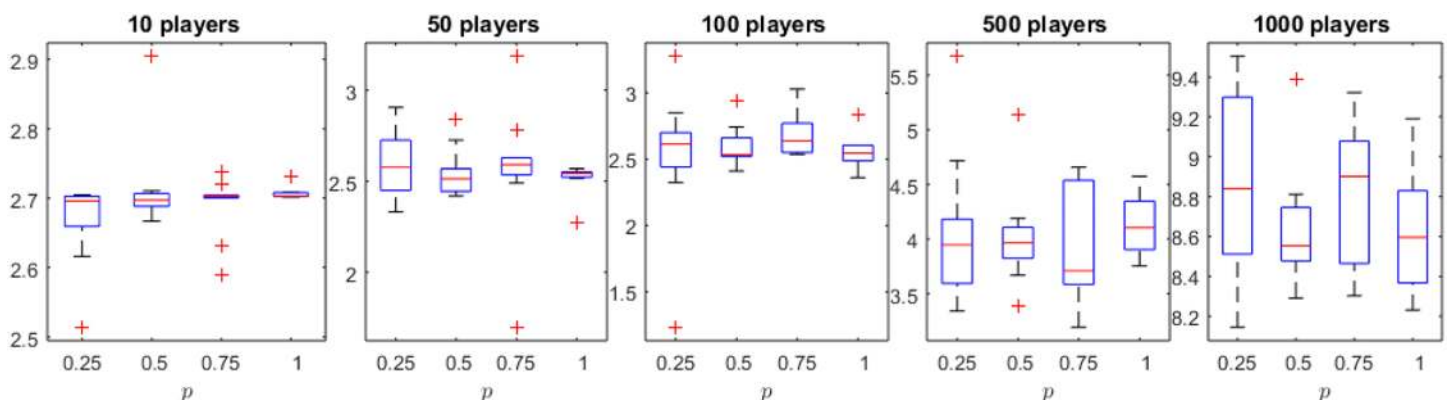
Also in order to evaluate the quality of the reported solution from the equilibria point of view, we verify for each one if there are any players that would improve their payoffs by unilateral deviation, and report the fraction of this number to the number of nodes in the network. In this manner we can assess if  $p$ MNEO actually computes Nash equilibria of the community structure detection game.

### Results and discussion

**Cournot oligopoly.** The results reported by  $p$ MNEO<sup>-</sup> on the Cournot oligopoly are presented in Figs 3–5. Fig 3 presents boxplots of minimum distance to NE for all the player settings. There are no significant differences between results reported when using different  $p$  values. Fig 4 complements this information by showing the evolution of the distance to NE with similar behavior for all  $p$  values. Fig 5 however, shows significant differences between the duration of runs (in seconds) for all  $p$  values, supporting the hypothesis that the use of the reduced generative relation yields faster, but just as good, results for the Cournot oligopoly.

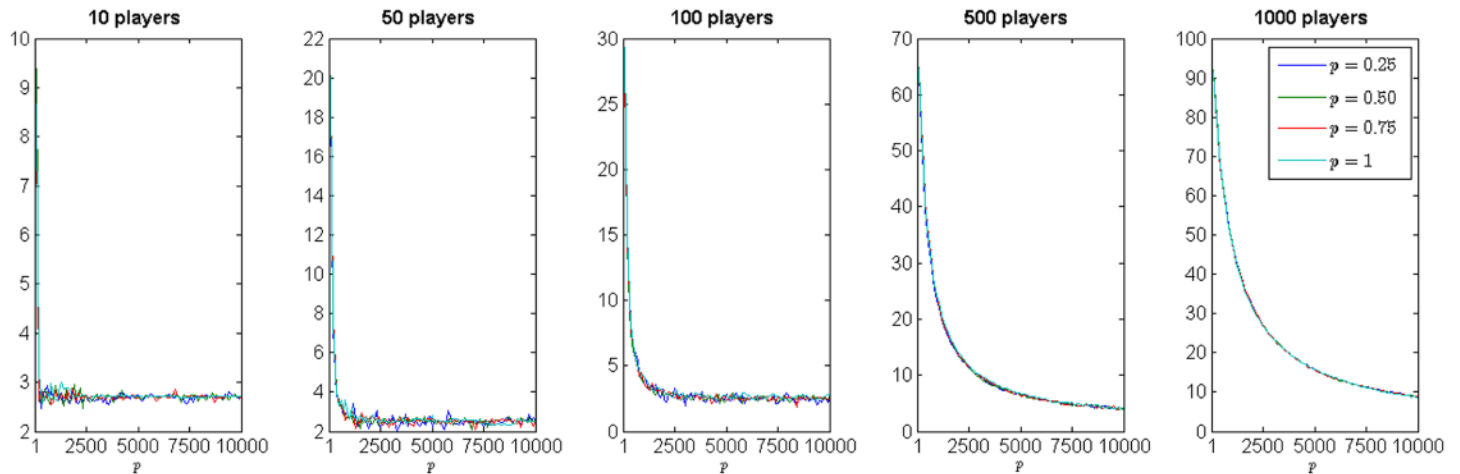
**Community structure detection.** NMI values computed for the results obtained by  $p$ MNEO by using the four different values of  $p$  as well as with the other methods for all considered networks are presented as boxplots accompanied by black-white matrices representing Wilcoxon  $h$ -values for each pair of methods tested in order to make the assessment about the statistical significance of possible differences straightforward (a white square in the matrix indicate no statistical differences in results while a black one indicates significant differences). The numbers 1 to 8 indicate the methods as they are ordered in the box-plots in the left.

Results obtained on networks with clear community structures are presented in the Supporting information. Thus, for the GN benchmarks with  $z_{out} \leq 5$  all methods compute the



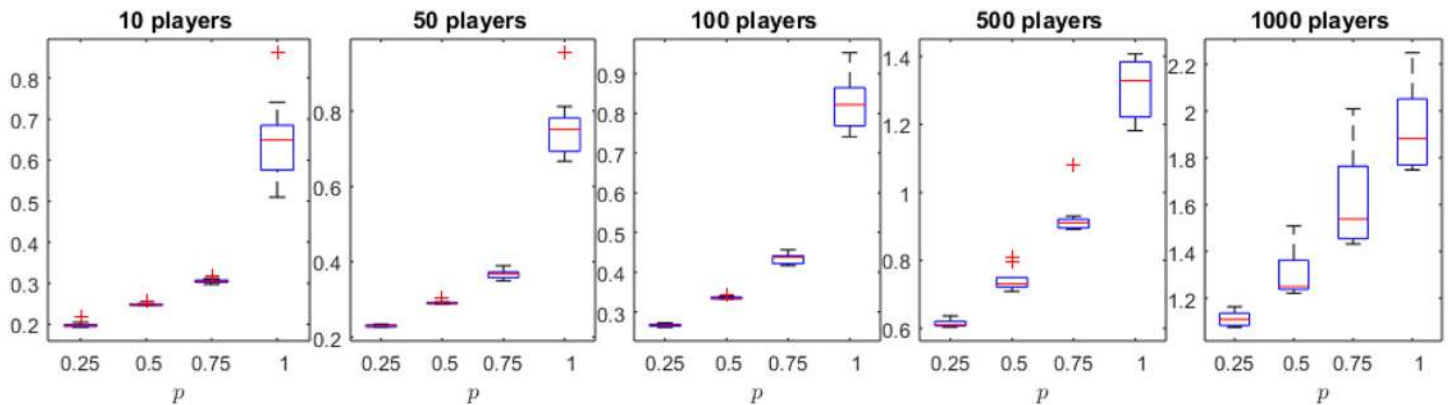
**Fig 3. Distance to Nash equilibrium for the Cournot oligopoly and different  $p$  values.** The null hypothesis that there is no statistical difference between means of results could not be rejected by using a Wilcoxon sum-rank test (0.05 significance level).

<https://doi.org/10.1371/journal.pone.0174963.g003>



**Fig 4. Evolution of the distance to NE in all cases: again no significant difference between the four  $p$  values is observed.**

<https://doi.org/10.1371/journal.pone.0174963.g004>



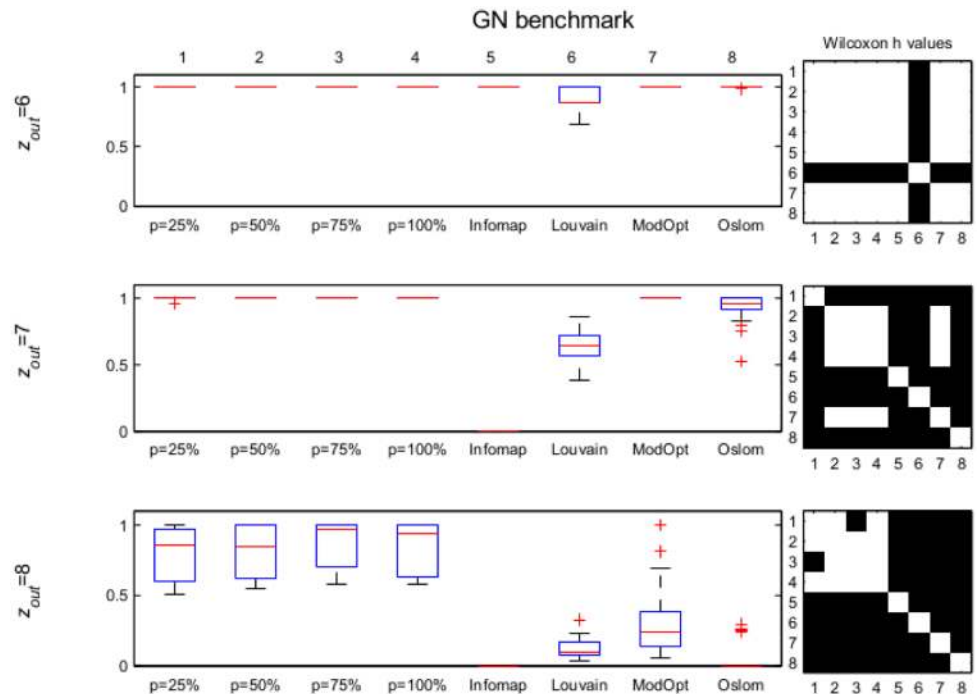
**Fig 5. Duration of the runs (in sec.).** The null hypothesis that differences between mean values are not significant was rejected by using a Wilcoxon sum-rank test with a significance level of 0.05.

<https://doi.org/10.1371/journal.pone.0174963.g005>

correct structure (S1 Fig). For  $z_{out} = 6$  and 7 (Fig 6),  $p$ MNEO correctly identifies the structure for all values of  $p$ , with the exception of one network for  $p = 25\%$  and  $z_{out} = 7$ . For  $z_{out} = 8$ , which is the most difficult set in the GN benchmark, all results obtained by  $p$ MNEO are significantly better than those obtained by the other methods. Regarding other results in literature obtained with an EO based method [46], we find that they report the fraction of nodes correctly classified as 1 for  $z_{out}$  values from 1 to 6 and decreasing to approx. 0.8 to  $z_{out} = 8$ .

For the LFR set with 128 nodes and  $\mu \leq 0.3$ , results are similar with those for the GN (S2 Fig). For the most difficult set, having  $\mu = 0.5$ ,  $p$ MNEO results are significantly better than those reported by Infomap and OSLOM. The results obtained for the LFR sets with 1000 nodes (Fig 7 and S3 Fig) show that all eight methods are competitive as they are all capable to identify the community structure with NMI values close to 1 (equal to 1 for Infomap).

All results obtained on the synthetic networks show no significant difference among the values of the  $p$  parameter. Thus, we may conclude that, for networks presenting similar



**Fig 6. GN  $z_{out} = 6, 7, 8$ ; boxplots of NMI values for all methods considered (left).** On the right, the black-white matrix represents Wilcoxon  $h$  values for each pair of methods considered, numbered in the order they appear in the boxplot. A black square indicates a statistical difference between results. For  $z_{out} = 8$  results obtained with all values of  $p$  are significantly better than all other methods considered.

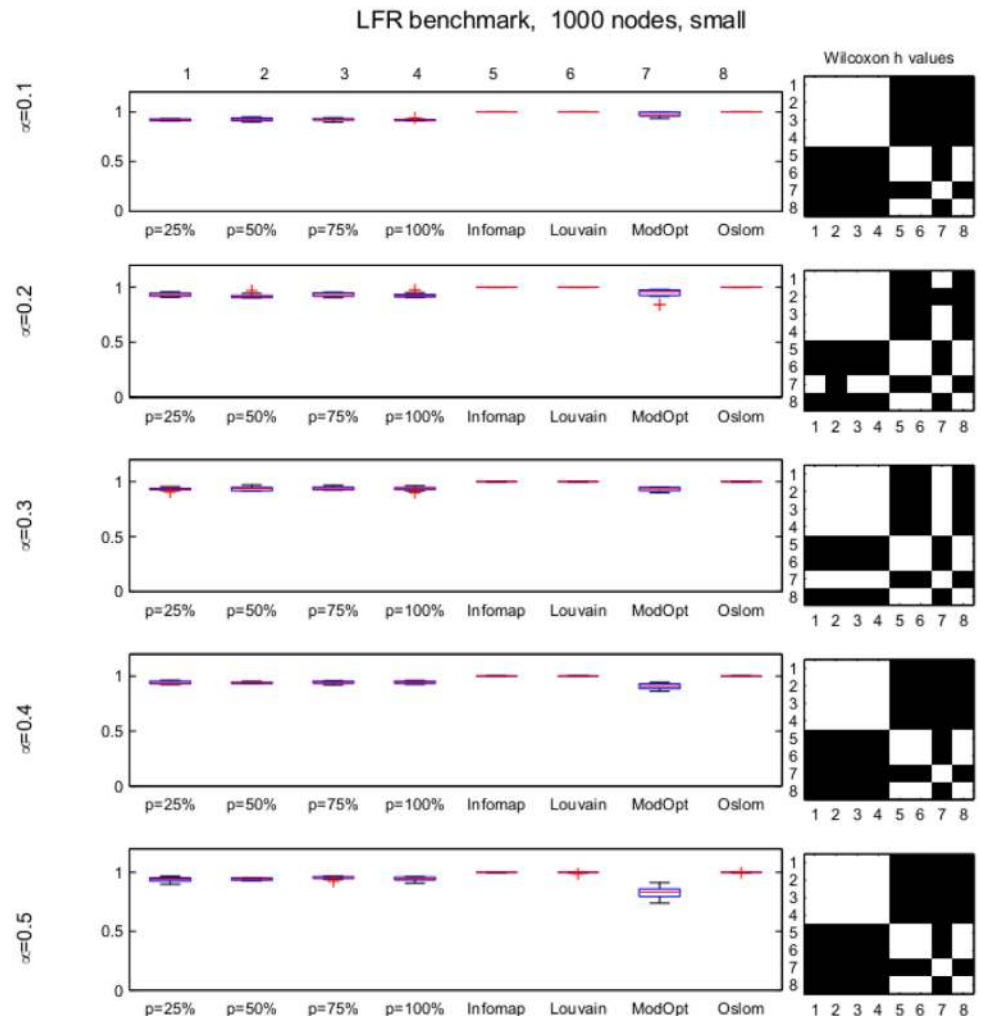
<https://doi.org/10.1371/journal.pone.0174963.g006>

structures with the tested ones, the Nash ascendancy relation can be replaced with the reduced variant considering even only 25% of the nodes when comparing two strategy profiles. The evolution of NMI values over time for the GN  $z_{out} = 8$  set, represented in Fig 8, shows that for this network set the evolution is similar, with overlapping trends, for all  $p$  values.

Fig 9 presents the results obtained for the four real-world networks tested. They show that for all  $p$  settings,  $p$ MNEO results are the same or better than those obtained by the other methods. Again, there is no statistical difference among results obtained for different  $p$  values in most cases.

Most other methods based on game theory in literature do not present extensive numerical results to allow comparisons. In [28] experiments are performed on the GN and LFR 1000 S sets, using their method, SNS-CD, and also the Game algorithm in [21] and NashCoDe in [29]. On the GN set, SNS-CD reports a NMI equal to 1 for  $z_{out} \leq 7$  and decreases below 0.5 for  $z_{out} = 8$ . According to their results, Game and NashCoDe reported an NMI of 0 for all GN sets. In the case of LFR 1000 S, SNS-CD reports best results among game theoretic based approaches, with NMI values of approx. 0.95 for  $\mu = 0.1, 0.2$ , and 0.85 for  $\mu = 0.3, 0.4$  and decreasing below 0.5 for  $\mu = 0.5$ .  $p$ MNEO results are better than all these (Figs 6, 7 and S1 Fig) for all tested  $p$  values.

**Is it Nash equilibrium?** While computing the Nash equilibrium is recognized to be challenging task, verifying that a strategy profile is a Nash equilibrium is a computationally expensive, yet trivial: we need to check for each player if unilateral deviations increase their payoff. When the strategy set is discrete, we can verify this by exhaustive search. To do so, for all results



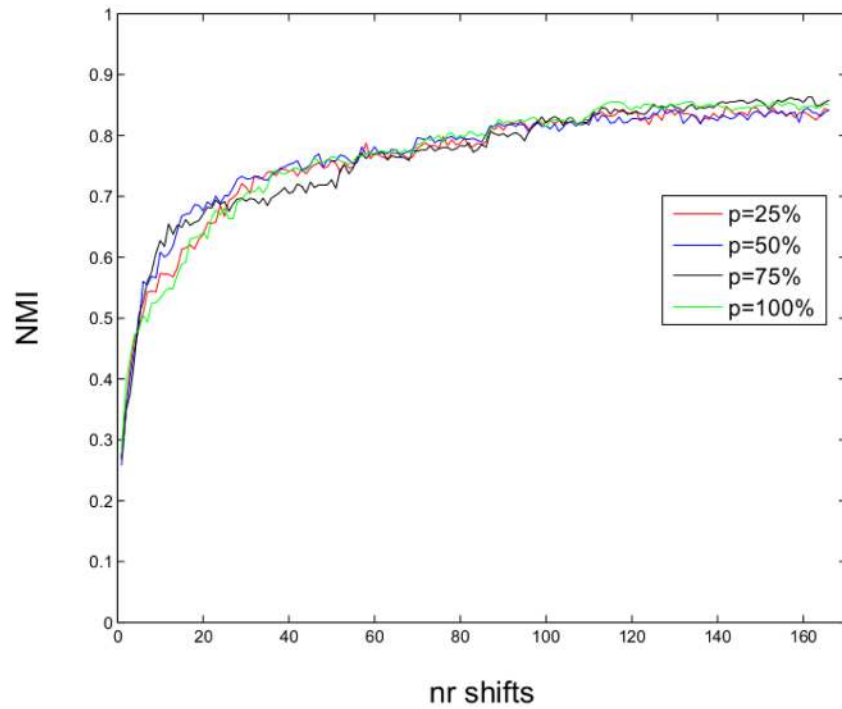
**Fig 7. LFR small, 1000 nodes; boxplots of NMI values for all methods considered (left).** On the right, the black-white matrix represents Wilcoxon  $h$  values for each pair of methods considered, numbered in the order they appear in the boxplot. A black square indicates a statistical difference between results.

<https://doi.org/10.1371/journal.pone.0174963.g007>

reported by  $p$ MNEO we counted the number of nodes (players) that would improve their pay-offs by changing community. In order to compare results among networks with different sizes we divided this number to the network size. Results are presented in Tables 1 and 2.

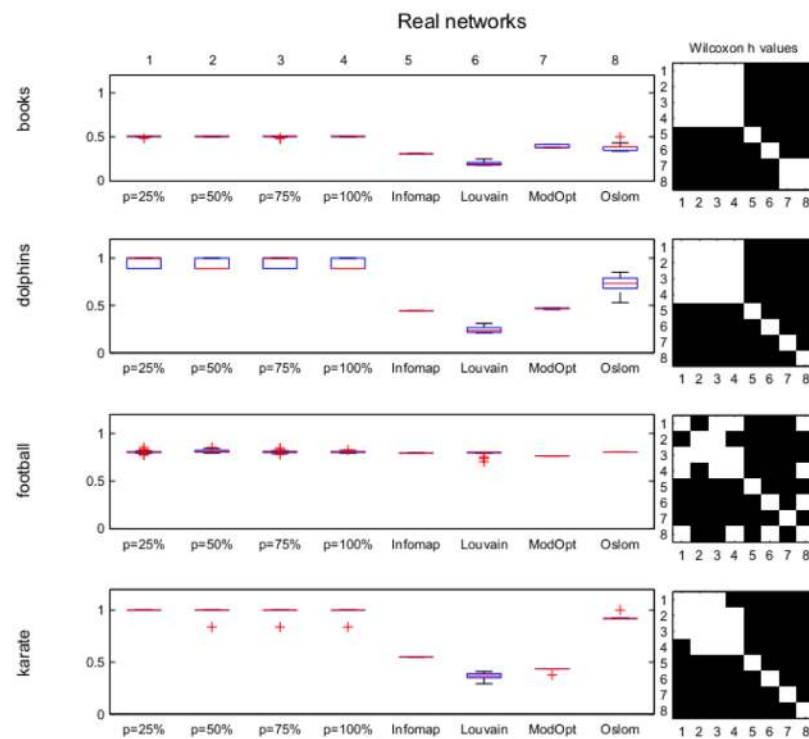
Results in Table 1 show that for all small networks the reported solutions are Nash equilibria for all  $p$  values. For the LFR sets with 1000 nodes, where the algorithm did not reach a NMI value of 1, there are nodes that can improve their payoffs by changing communities (more for LFR small). For the real world networks, we find that the solutions are Nash equilibria. Looking at Table 2 we find that even when the real community structure is not available, the provided solutions (with a high modularity value) are good approximations of Nash equilibria.

Furthermore, for real-world networks where the concept of ‘real community structure’ is debatable [47], these results that have the Nash equilibrium property of stability against unilateral deviations may actually provide a more realistic approach to this concept. For example, the *books* dataset is constructed from online data by considering books as nodes and edges are



**Fig 8. Evolution of the mean NMI values over time for the GN  $z_{out} = 8$  set and different  $p$  values.**

<https://doi.org/10.1371/journal.pone.0174963.g008>



**Fig 9. Real-world networks; boxplots of NMI values for all methods considered (left). On the right, the black-white matrix represents Wilcoxon  $h$  values for each pair of methods considered, numbered in the order they appear in the boxplot. A black square indicates a statistical difference between results.**

<https://doi.org/10.1371/journal.pone.0174963.g009>

**Table 1. Fraction of nodes that can improve their payoffs by unilateral deviation in solutions reported by  $\rho$ MNEO for networks with known community structure.** A zero value indicates that reported solutions are indeed Nash equilibria of game  $\tilde{\Gamma}$ . We report average values  $\pm$  standard deviations and confidence limits of the mean.

$\mu$		$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 1$
<b>GN</b>					
$z_{out}:1-8$	mean $\pm$ stdev	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00
	min	0.00	0.00	0.00	0.00
	CL mean	0.00 0.00	0.00 0.01	0.00 0.00	0.00 0.00
<b>LFR 128</b>					
0.1-0.5	mean $\pm$ stdev	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00
	min	0.00	0.00	0.00	0.00
	CL mean	0.00 0.00	0.00 0.01	0.00 0.00	0.00 0.00
<b>LFR 1000 big</b>					
0.1	mean $\pm$ stdev	0.01 $\pm$ 0.02	0.00 $\pm$ 0.01	0.00 $\pm$ 0.01	0.01 $\pm$ 0.01
	min	0.00	0.00	0.00	0.00
	CL mean	0.01 0.02	0.00 0.01	0.00 0.01	0.00 0.01
0.2	mean $\pm$ stdev	0.00 $\pm$ 0.00	0.01 $\pm$ 0.02	0.00 $\pm$ 0.01	0.00 $\pm$ 0.01
	min	0.00	0.00	0.00	0.00
	CL mean	0.00 0.00	0.01 0.02	0.00 0.01	0.00 0.01
0.3	mean $\pm$ stdev	0.00 $\pm$ 0.01	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.01
	min	0.00	0.00	0.00	0.00
	CL mean	0.00 0.01	0.00 0.00	0.00 0.00	0.00 0.01
0.4-0.5	mean $\pm$ stdev	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00
	min	0.00	0.00	0.00	0.00
	CL mean	0.00 0.00	0.00 0.01	0.00 0.00	0.00 0.00
<b>LFR 1000 small</b>					
0.1	mean $\pm$ stdev	0.10 $\pm$ 0.05	0.11 $\pm$ 0.04	0.12 $\pm$ 0.04	0.17 $\pm$ 0.11
	min	0.05	0.05	0.07	0.04
	CL mean	0.07 0.14	0.08 0.14	0.09 0.15	0.09 0.24
0.2	mean $\pm$ stdev	0.14 $\pm$ 0.06	0.12 $\pm$ 0.06	0.11 $\pm$ 0.08	0.09 $\pm$ 0.05
	min	0.07	0.03	0.00	0.01
	CL mean	0.10 0.18	0.08 0.17	0.05 0.16	0.06 0.13
0.3	mean $\pm$ stdev	0.11 $\pm$ 0.07	0.10 $\pm$ 0.08	0.12 $\pm$ 0.09	0.08 $\pm$ 0.06
	min	0.03	0.01	0.05	0.00
	CL mean	0.06 0.16	0.04 0.15	0.05 0.18	0.04 0.12
0.4	mean $\pm$ stdev	0.07 $\pm$ 0.03	0.07 $\pm$ 0.04	0.05 $\pm$ 0.03	0.06 $\pm$ 0.03
	min	0.03	0.00	0.01	0.01
	CL mean	0.05 0.10	0.04 0.10	0.03 0.07	0.03 0.08
0.5	mean $\pm$ stdev	0.06 $\pm$ 0.04	0.04 $\pm$ 0.03	0.04 $\pm$ 0.03	0.05 $\pm$ 0.04
	min	0.02	0.01	0.01	0.01
	CL mean	0.03 0.08	0.02 0.06	0.02 0.07	0.02 0.08
<b>Real networks</b>					
books	mean $\pm$ stdev	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00
	min	0.00	0.00	0.00	0.00
	CL mean	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00
dolphins	mean $\pm$ stdev	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00
	min	0.00	0.00	0.00	0.00
	CL mean	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00

(Continued)



Table 1. (Continued)

$\mu$		$p = 0.25$	$p = 0.5$	$p = 0.75$	$p = 1$
fotball	mean $\pm$ stdev	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00
	min	0.00	0.00	0.00	0.00
	CL mean	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00
karate	mean $\pm$ stdev	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00
	min	0.00	0.00	0.00	0.00
	CL mean	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00

<https://doi.org/10.1371/journal.pone.0174963.t001>

Table 2. Fraction of nodes that can improve their payoffs by unilateral deviation in solutions reported by  $p$ MNEO for real world networks with unknown community structure. A zero value indicates that reported solutions are indeed Nash equilibria of game  $\tilde{\Gamma}$ . We report average values  $\pm$  standard deviations and confidence limits of the mean also for modularity  $Q$ .

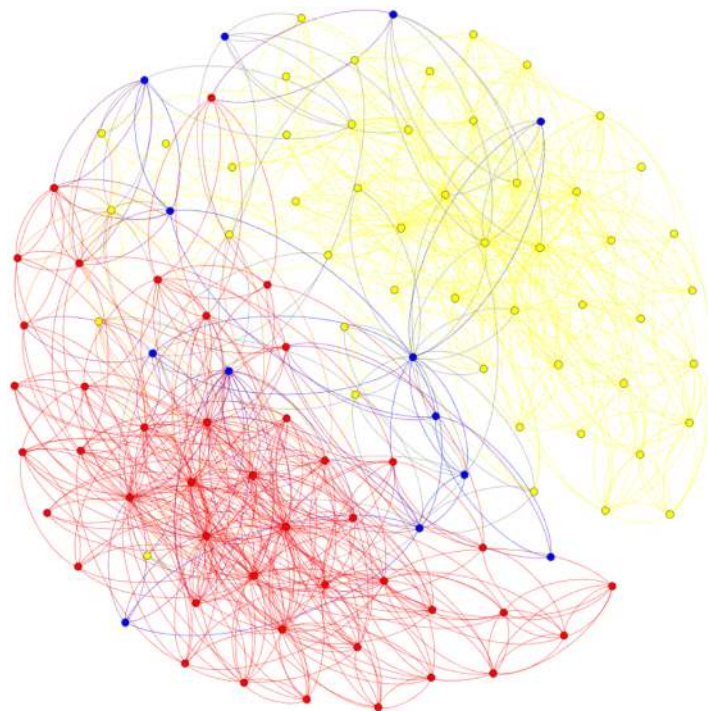
Network		$p = 0.25$	$p = 0.5$	$p = 0.75$	$p = 1$
riskmap	$Q$	0.62 $\pm$ 0.00	0.61 $\pm$ 0.00	0.61 $\pm$ 0.01	0.61 $\pm$ 0.00
	mean $\pm$ stdev	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00
	min	0.00	0.00	0.00	0.00
	CL mean	0.00 0.00	0.00 0.00	0.00 0.00	0.00 0.00
jazz	$Q$	0.44 $\pm$ 0.00	0.44 $\pm$ 0.00	0.44 $\pm$ 0.00	0.44 $\pm$ 0.00
	mean $\pm$ stdev	0.00 $\pm$ 0.01	0.01 $\pm$ 0.01	0.01 $\pm$ 0.01	0.00 $\pm$ 0.01
	min	0.00	0.00	0.00	0.00
	CL mean	0.00 0.01	0.00 0.01	0.00 0.01	0.00 0.01
contiguous USA	$Q$	0.58 $\pm$ 0.00	0.58 $\pm$ 0.01	0.58 $\pm$ 0.01	0.58 $\pm$ 0.01
	mean $\pm$ stdev	0.01 $\pm$ 0.01	0.00 $\pm$ 0.01	0.01 $\pm$ 0.01	0.00 $\pm$ 0.01
	min	0.00	0.00	0.00	0.00
	CL mean	0.00 0.02	0.00 0.01	0.00 0.02	0.00 0.01

<https://doi.org/10.1371/journal.pone.0174963.t002>

linking books that have been co-purchased by same buyers. The ‘real’ structure is considered to be set by the book’s political content. However, by studying the ‘real’ structure we can find nodes that have no link with other nodes in the community they are assigned to, which makes their ‘correct’ identification impossible (Figs 10 and 11). In this situation the equilibrium approach provides an alternative solution that may offer information about the real and not the expected structure of the network.

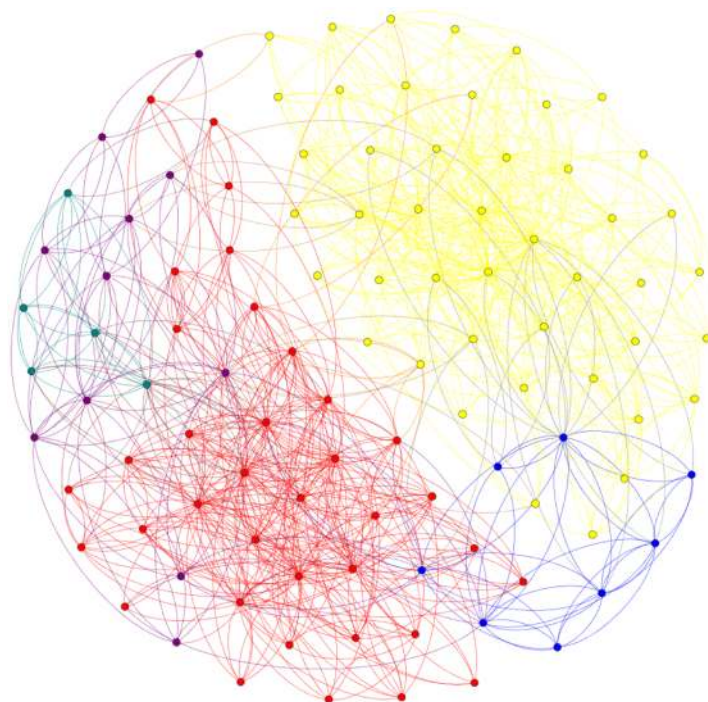
**Efficiency.** One of the reasons the reduced Nash ascendancy relation was considered was to decrease the number of payoff function evaluations. Reducing this number up to only 25% should also reduce the running time of  $p$ MNEO. Fig 12 illustrates for each set the average duration necessary to run the experiments for different  $p$  values as percentage of the average duration for the  $p = 100\%$  case. Results are similar for all the sets: reducing to 25% the percent of nodes that are used in the reduced ascendancy relation leads to an approximatively 10% decrease in running time. We may conclude that using the Nash ascendancy relation influences the running time less than other components of the extremal optimization algorithm.

In order to assess if there is indeed an economy in payoff function evaluations, we counted the number of payoff function calls that were actually performed from the  $p$ Nash ascendancy procedure (Alg. 2). Results are presented in Fig 13. The first thing to notice is that values



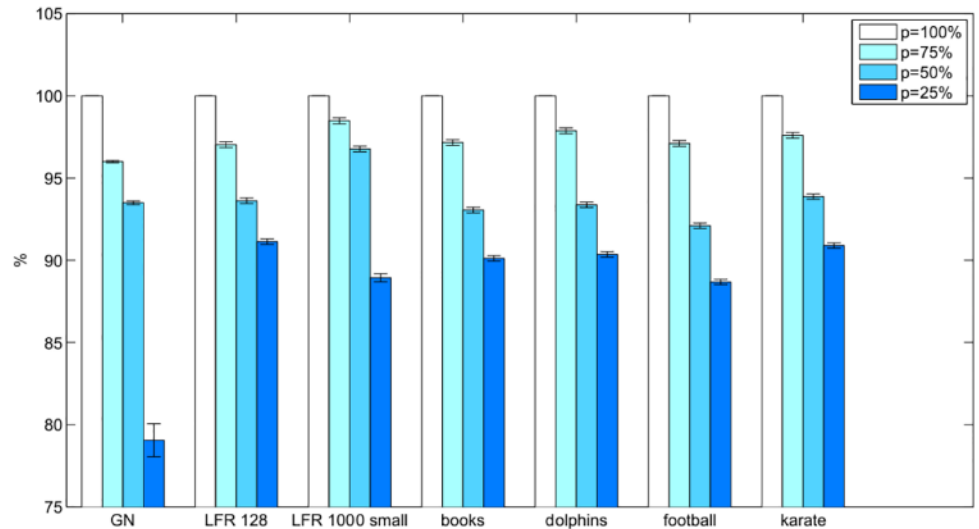
**Fig 10. Political books dataset: The ‘real’ community structure is represented.** We can see yellow nodes that have more colored links than yellow, indicating more links in the other two communities.

<https://doi.org/10.1371/journal.pone.0174963.g010>



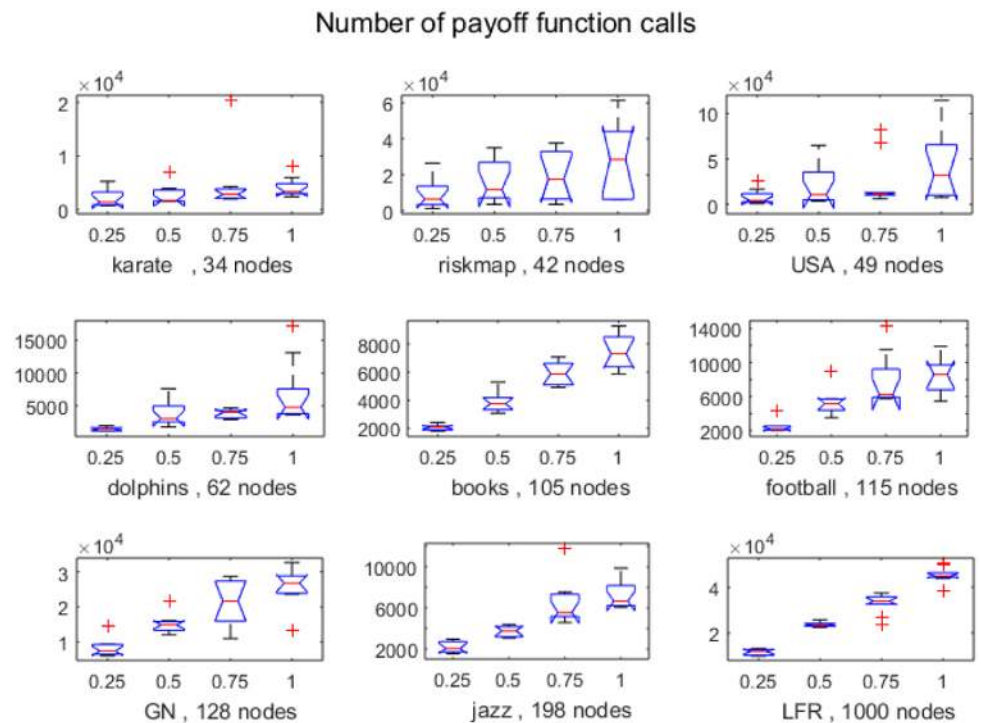
**Fig 11. Political books dataset: A solution reported by  $pMNEO$  with  $NMI = 0.52$  and verified to be a Nash equilibrium; some of the ‘difficult’ nodes are placed in a different community.**

<https://doi.org/10.1371/journal.pone.0174963.g011>



**Fig 12. Differences in running times for the four tested  $p$  values as percentage of the running time for the  $p = 100\%$  case for each network set.**

<https://doi.org/10.1371/journal.pone.0174963.g012>



**Fig 13. Number of payoff function calls in Fig. 2, 10 individuals. Notches indicate confidence limits of the mean.**

<https://doi.org/10.1371/journal.pone.0174963.g013>

**Table 3. Wilcoxon sum-rank results regarding differences in payoff function calls for the tested  $p$  values. An \* indicates significant difference.**

Network	0.25 → 0.5	0.25 → 0.75	0.25 → 1	0.5 → 0.75	0.5 → 1	0.75 → 1
karate	–	–	*	–	*	–
riskmap	–	–	*	–	–	–
USA	–	*	*	–	–	–
dolphins	*	*	*	–	*	*
books	*	*	*	*	*	*
football	*	*	*	*	*	–
GN	*	*	*	*	*	–
jazz	*	*	*	*	*	–
LFR 1000	*	*	*	*	*	*

<https://doi.org/10.1371/journal.pone.0174963.t003>

reported for different  $p$  values on the same network do not have the same ratio between them as the  $p$  values themselves. The explanation for that is that the payoff function call in Alg. 2 is performed only if the selected player has different strategies in the two tested profiles (line 4). A significant difference would indicate the utility of considering the reduced generative relation for the community structure detection problem.

The Wilcoxon sum-rank test shows that the differences between  $p = 0.25$  and  $p = 1$  are significant for all networks (Table 3), while looking at NMI results reported for these values do not differ significantly (only in few cases). Table 3 also shows that most significant differences appear when comparing 0.25 with the others, and least differences between 0.75 and 1. These results show that even if the gain in running time -which also depends on the implementation—may not seem remarkable, there are benefits in using the reduced generative relation.

Compared with the other methods, however,  $p$ MNEO is slower, like many other evolutionary techniques. A typical run for a network with 1000 nodes, population of 50 individuals and 6000 generations takes about one hour on a Intel(R) Core(TM) Quad CPU Q9400 @2.66GHz, while for Osloom and Infomap it takes minutes. However, the results obtained on difficult network show that  $p$ MNEO is a more refined method with potential to reveal structures that are not discoverable by other methods.

## Conclusion

We explore the possibility of using a reduced version of the Nash ascendancy relation within an extremal optimization method designed to identify the community structure of a network. The reduced ascendancy relation takes into account only a fraction of the network nodes when comparing two strategy profiles. We show that  $p$ -Nash non-ascended solutions are also Nash equilibria of the game. We address the practical concern that using the reduced relation may not yield results as good as the Nash ascendancy relation when maintaining the same number of iterations by using numerical experiments. Results show that there are very few significant differences in results between different values for the fraction of nodes and that the performance of  $p$ MNEO is significantly better than that of other state-of-the-art methods on the tested benchmarks.

Regarding the running time, the differences between different probability levels are low: an average decrease of 10% in running time when switching from  $p = 100\%$  to  $p = 25\%$ . This result indicates that the Nash ascendancy relation, while indeed computationally expensive, does not influence the running time of  $p$ MNEO as much as other components of the method.

However, the reduced version does offer a less computationally expensive alternative to the Nash ascendancy relation, that can be used to enhance other heuristics that attempt to solve the community structure detection problem using game theoretic approaches.

## Supporting information

**S1 Fig. GN  $z_{out} = \{1, \dots, 5\}$ ; boxplots of NMI values for all methods considered (left).** On the right, the black-white matrix represents Wilcoxon  $h$  values for each pair of methods considered, numbered in the order they appear in the boxplot. A black square indicates a statistical difference between results. As these networks have very well defined community structures, all methods are capable to identify the correct structure.  
(EPS)

**S2 Fig. LFR small, 128 nodes; boxplots of NMI values for all methods considered (left).** On the right, the black matrix represents Wilcoxon  $h$  values for each pair of methods considered, numbered in the order they appear in the boxplot. A black square indicates a statistical difference between results.  
(EPS)

**S3 Fig. LFR big, 1000 nodes; boxplots of NMI values for all methods considered (left).** On the right, the black-white matrix represents Wilcoxon  $h$  values for each pair of methods considered, numbered in the order they appear in the boxplot. A black square indicates a statistical difference between results.  
(EPS)

## Acknowledgments

This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number PN-II-RU-TE-2014-4-2332.

## Author Contributions

**Conceptualization:** LRI GN.

**Data curation:** GN SMA.

**Formal analysis:** GN SMA LRI.

**Funding acquisition:** LRI.

**Investigation:** GN SMA LRI.

**Methodology:** LRI GN.

**Project administration:** LRI.

**Resources:** GN SMA.

**Software:** LRI GN SMA.

**Supervision:** LRI.

**Validation:** LRI GN SMA.

**Visualization:** GN SMA LRI.

**Writing – original draft:** LRI GN.

**Writing – review & editing:** LRI.

## References

1. Suciú M, Lung R, Gasko N. Mixing Network Extremal Optimization for Community Structure Detection. In: Ochoa G, Chicano F, editors. *Evolutionary Computation in Combinatorial Optimization*. vol. 9026 of *Lecture Notes in Computer Science*. Springer International Publishing; 2015. p. 126–137. Available from: [http://dx.doi.org/10.1007/978-3-319-16468-7\\_11](http://dx.doi.org/10.1007/978-3-319-16468-7_11)
2. Lung RI, Chira C, Andreica A. Game Theory and Extremal Optimization for Community Detection in Complex Dynamic Networks. *PLoS ONE*. 2014; 9(2):e86891. <https://doi.org/10.1371/journal.pone.0086891> PMID: 24586257
3. Fortunato S. Community detection in graphs. *Physics Reports*. 2010; 486:75–174. <https://doi.org/10.1016/j.physrep.2009.11.002>
4. Nash JF. Non-cooperative games. *Annals of Mathematics*. 1951; 54:286–295. <https://doi.org/10.2307/1969529>
5. Lung RI, Gog A, Chira C. A Game Theoretic Approach to Community Detection in Social Networks. In: *Nature Inspired Cooperative Strategies for Optimization, NICSO 2011, Cluj-Napoca, Romania, October 20–22, 2011*. p. 121–131. Available from: [http://dx.doi.org/10.1007/978-3-642-24094-2\\_8](http://dx.doi.org/10.1007/978-3-642-24094-2_8)
6. Lung RI, Dumitrescu D. Computing Nash Equilibria by Means of Evolutionary Computation. *Int J of Computers, Communications & Control*. 2008; III(suppl.issue):364–368.
7. Barabási AL. *Linked—the new science of networks*; 2002.
8. Radicchi F, Castellano C, Cecconi F, Loreto V, Parisi D. Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101(9):2658–2663. <https://doi.org/10.1073/pnas.0400054101> PMID: 14981240
9. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. *Physical review E*. 2004; 69(2):26113. <https://doi.org/10.1103/PhysRevE.69.026113>
10. Molecular KK, Molecular KK. Letters To Nature. *Nature*. 2005; 433(February):895–900.
11. Mokken RJ. Cliques, clubs and clans. *Quality & Quantity*. 1979; 13(2):161–173. <https://doi.org/10.1007/BF00139635>
12. Hu Y, Chen H, Zhang P, Li M, Di Z, Fan Y. Comparative definition of community and corresponding identifying algorithm. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*. 2008; 78(2):1–7.
13. Gao Z, Lu Z, Chen W, Zhong J, Bi Y. The Maximum Community Partition Problem in Networks. In: *2013 International Conference on Social Intelligence and Technology*. IEEE; 2013. p. 28–36. Available from: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84881123765&partnerID=tZOtx3y1>
14. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. Complex networks: Structure and dynamics. *Physics Reports*. 2006; 424(4-5):175–308. <https://doi.org/10.1016/j.physrep.2005.10.009>
15. Newman MEJ. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*. 2006; 103(23):8577–8582. <https://doi.org/10.1073/pnas.0601602103>
16. Pizzuti C. GA-Net: A genetic algorithm for community detection in social networks. In: *Parallel Problem Solving from Nature X*. Springer; 2008. p. 1081–1090.
17. Lancichinetti A, Fortunato S, Kertész J. Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics*. 2009; 11(3):033015. <https://doi.org/10.1088/1367-2630/11/3/033015>
18. Li Z, Zhang S, Wang RS, Zhang XS, Chen L. Quantitative function for community detection. *Phys Rev E*. 2008; 77:036109. <https://doi.org/10.1103/PhysRevE.77.036109>
19. Fortunato S, Barthélemy M. Resolution limit in community detection. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104(1):36–41. <https://doi.org/10.1073/pnas.0605965104> PMID: 17190818
20. Chira C, Gog A. Fitness evaluation for overlapping community detection in complex networks. In: *Evolutionary Computation (CEC), 2011 IEEE Congress on*; 2011. p. 2200–2206.
21. Chen W, Liu Z, Sun X, Wang Y. A game-theoretic framework to identify overlapping communities in social networks. *Data Mining and Knowledge Discovery*. 2010; 21(2):224–240. <https://doi.org/10.1007/s10618-010-0186-6>
22. McSweeney PJ, Mehrotra K, Oh JC. A Game Theoretic Framework for Community Detection. In: *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE; 2012.

- p. 227–234. Available from: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84874230938&partnerID=tZOtx3y1>
23. Mandala S, Kumara S, Chatterjee K. A Game-Theoretic Approach to Graph Clustering. *INFORMS Journal on Computing*. 2014; 26(3):629–643. <https://doi.org/10.1287/ijoc.2013.0588>
  24. Sung YS, Kumara S. Detecting overlapping communities in online social networks using game theoretic approach. In: *IIE Annual Conference and Expo 2013*. Institute of Industrial Engineers; 2013. p. 77–86. Available from: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84900314301&partnerID=tZOtx3y1>
  25. Alvari H, Hajibagheri A, Sukthankar G. Community detection in dynamic social networks: A game-theoretic approach. In: *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*. IEEE; 2014. p. 101–107. Available from: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84911087193&partnerID=tZOtx3y1>
  26. Zhou L, Cheng C, Lü K, Chen H. Using coalitional games to detect communities in social networks. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. vol. 7923 LNCS; 2013. p. 326–331. Available from: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84880008779&partnerID=tZOtx3y1>
  27. Zhou L, Lü K, Liu W. An approach for community detection in social networks based on cooperative games theory. *Expert Systems*. 2016; 33(2):176–188. <https://doi.org/10.1111/exsyt.12141>
  28. Basu S, Maulik U. Community detection based on strong Nash stable graph partition. *Social Network Analysis and Mining*. 2015; 5(1):61. <https://doi.org/10.1007/s13278-015-0299-4>
  29. Narayanam R, Narahari Y. A game theory inspired, decentralized, local information based algorithm for community detection in social graphs. In: *Proceedings—International Conference on Pattern Recognition*; 2012. p. 1072–1075. Available from: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84874571413&partnerID=tZOtx3y1>
  30. Li J, Kendall G, John R. Computing Nash Equilibria and Evolutionarily Stable States of Evolutionary Games. *IEEE Transactions on Evolutionary Computation*. 2016; 20(3):460–469. <https://doi.org/10.1109/TEVC.2015.2490076>
  31. Li C, Cao H, Du M. A novel method to compute Nash equilibrium in non-cooperative n-person games based on differential evolutionary algorithm. *Intelligent Decision Technologies*. 2014; 8(3):207–213. <https://doi.org/10.3233/IDT-140189>
  32. Pavlidis NG, Parsopoulos KE, Vrahatis MN. Computing Nash equilibria through computational intelligence methods. *Journal of Computational and Applied Mathematics*. 2005; 175(1 SPEC. ISS.):113–136. <https://doi.org/10.1016/j.cam.2004.06.005>
  33. Lung RI, Mihoc TD, Dumitrescu D. Nash equilibria detection for multi-player games. In: *IEEE Congress on Evolutionary Computation*; 2010. p. 1–5.
  34. Maslov S, Sneppen K. Specificity and Stability in Topology of Protein Networks. *Science*. 2002; 296:910–913. <https://doi.org/10.1126/science.1065103> PMID: 11988575
  35. Lancichinetti A, Fortunato S. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Phys Rev E*. 2009; 80:016118. <https://doi.org/10.1103/PhysRevE.80.016118>
  36. Girvan M, Newman MEJ. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*. 2002; 99(12):7821–7826. <https://doi.org/10.1073/pnas.122653799>
  37. Lusseau D, Schneider K, Boisseau O, Haase P, Slooten E, Dawson S. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*. 2003; 54(4):396–405. <https://doi.org/10.1007/s00265-003-0651-y>
  38. Zachary WW. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*. 1977; 33(4):452–473. <https://doi.org/10.1086/jar.33.4.3629752>
  39. Steinhäuser K, Chawla NV. Identifying and evaluating community structure in complex networks. *Pattern Recognition Letters*. 2010; 31(5):413–421. <https://doi.org/10.1016/j.patrec.2009.11.001>
  40. Gleiser PM, Danon L. Community structure in jazz. *Advances in complex systems*. 2003; 6(04):565–573. <https://doi.org/10.1142/S0219525903001067>
  41. Knuth DE. *The Art of Computer Programming, Volume 4, Fascicle 0: Introduction to Combinatorial Algorithms and Boolean Functions (Art of Computer Programming)*. 1st ed. Addison-Wesley Professional; 2008.
  42. Lancichinetti A, Radicchi F, Ramasco JJ, Fortunato S. Finding statistically significant communities in networks. *PloS one*. 2011; 6(4):e18961. <https://doi.org/10.1371/journal.pone.0018961> PMID: 21559480

43. Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*. 2008; 105(4):1118–1123. <https://doi.org/10.1073/pnas.0706851105>
44. Sales-Pardo M, Guimera R, Moreira AA, Nunes Amaral LA. Extracting the hierarchical organization of complex systems. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104(39):15224–15229. <https://doi.org/10.1073/pnas.0703740104> PMID: 17881571
45. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008; 2008(10):P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
46. Duch J, Arenas A. Community detection in complex networks using extremal optimization. *Physical Review E*. 2005; 72(2):027104. <https://doi.org/10.1103/PhysRevE.72.027104>
47. Palchykov V, Gemmetto V, Boyarsky A, Garlaschelli D. Ground truth? Concept-based communities versus the external classification of physics manuscripts. *EPJ Data Science*. 2016; 5(1):28. <https://doi.org/10.1140/epjds/s13688-016-0090-4>