Design of Coalition Resistant Credit Score Functions for Online Discussion Forums

Ganesh Ghalme
Indian Institute of Science
Bangalore, India
ganeshg@iisc.ac.in

Sujit Gujar
International Institute of Information Technology
Hyderabad, India
sujit.gujar@iiit.ac.in

Amleshwar Kumar
Indian Institute of Science
Bangalore, India
amleshwarkumar10717@gmail.com

Shweta Jain
Indian Institute of Science
Bangalore, India
shwetajains20@gmail.com

Y. Narahari
Indian Institute of Science
Bangalore, India
narahari@iisc.ac.in

ABSTRACT

We consider the problem of designing a robust credit score function in the context of online discussion forums. Credit score function assigns a real-valued credit score to each participant based on activities on the forum. A credit score of a participant quantifies the usefulness of contribution made by her. However, participants can manipulate a credit score function by forming coalitions, i.e., by strategically awarding upvotes, likes, etc. among a subset of agents to maximize their credit scores. We propose a coalition resistant credit score function which discourages such strategic endorsements. We use community detection algorithms to identify close-knit communities in the graph of interactions and characterize *coalition identifying* community detection metric. In particular, we show that *modularity* is coalition identifying and provide theoretical guarantees on modularity based credit score function. Finally, we validate our theoretical findings with simulations on illustrative datasets.

ACM Reference Format:
Ganesh Ghalme, Sujit Gujar, Amleshwar Kumar, Shweta Jain, and Y. Narahari. 2018. Design of Coalition Resistant Credit Score Functions for Online Discussion Forums . In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 9 pages.

1 INTRODUCTION

The internet has transformed the way we communicate ideas and share knowledge. Online discussion forums (ODFs) play an important role by providing a platform for the internet users to interact. There exist numerous ODFs where participants (i.e. the agents) can post the reviews of events, restaurants, movies, books, services, and so on. Such online forums are also commonly used as a web-based discussion platform to discuss academic topics, personal experiences, views, etc. We are motivated by forums such as Quora, Stack Exchange, Reddit, WikiAnswers, and Yahoo! Answers, which provide a platform for open-ended discussions, and by platforms such as Piazza and iClicker which supplement traditional classroom teaching. These forums are also used by massive online open courses (MOOCs) to provide the students with a platform to actively discuss subject related topics.

The usefulness (or quality) of a posted content is generally gauged by its popularity, which is often measured by the number of upvotes, likes, comments, shares, cross-references etc. received for the post. We refer to these parameters as *popularity indicators.* It is a common practice to incentivize participants to post high-quality content by providing gift vouchers, discounts, reward points, expert ranks, badges, etc. In this paper, we abstract and call such rewards as *credit scores* awarded to each agent. These forums, however, are susceptible to strategic manipulation by participating agents. For example, a subset of agents may favor the posts by within group agents by awarding upvotes, likes, or shares to posts by within-group agents. We refer to this as manipulation by coalition formation. The goal of this paper is to design credit score functions that prevent coalition formation. We call such functions *coalition resistant* credit score functions.

We consider an online discussion forum setup with heterogeneous agents. Each agent is characterized by a quality parameter. The quality of an agent reflects the usefulness/value of her contribution to the forum, which depends on her expertise, understanding, commitment, or skill level in the field. Our objective is to design a coalition resistant *credit score function* which awards a credit score to each agent by taking into account the strategic nature of interactions. We analyze the network structure arising from agents' interaction graphs such as follow network graph, upvote graph, like-network, etc. In the context of a follow graph, a coalition means that the within-group agents follow each other, whereas, in an upvote network, the within coalition agents upvote the posts by each other. In the rest of the paper, we call such a graph *popularity indicators graph.*

If a subset of agents form a coalition, the realized popularity indicators graph will most likely contain a dense subgraph reflecting the coalition structure emerged from strategic interactions. To detect such dense subgraphs, we use community detection algorithms which are well studied in the social networks literature. We hypothesize that under few reasonable conditions, a strategic coalition can be detected by community detection algorithms. Coalition resistant credit score functions ensure, for every agent, that the expected credit score is maximized when the agent does not join a coalition. The idea is to impose penalties on the agents found in a community

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

detected using a community detection metric. The proposed credit score suitably penalizes the agents found in dense communities and thus renders "not forming a coalition" as a best response strategy when the rest of the agents do not form a coalition. In summary, following are our specific contributions.

Contributions. We first characterize the class of community detection metrics which can be used to effectively detect strategic coalitions. We show that any community detection metric that satisfies a certain coalition identifying property (Definition 4.3) can be used to define a credit score function to be coalition resistant (Theorem 4.4). In particular, we propose a coalition resistant credit score function that uses the modularity metric (Section 5). We next show that the expected credit score of an agent that is not part of a coalition is positive even if the agent falsely gets detected in a community by a community detection algorithm (Lemma 5.5).

Our proposed credit score function depends on the true quality parameters and bias parameter. However, in practical scenarios these parameters are not known. To address this, we propose a learning algorithm to design credit score functions that preserve the coalition resistant property (Section 6). Finally, we validate our results through extensive simulations on representative data (Section 7).

2 RELATED WORK

Online discussions have grown in both volume and content in the recent times. From academic content to almost everything warranting a debate, people post their opinions online. Pendry and Salvatore [20] provide a detailed account of benefits of such online discussion forums. However, these forums are susceptible to strategic manipulations; Dellarocas [6] discuss strategic manipulations possible in the context of expressed opinions on forums while endorsing the products by other agents. Most of the work in this space focus on predicting and/or detecting a collusive manipulation on online forums such as online shopping systems [14, 15, 19], community question answer systems [7, 13, 23], social networking platforms [16] and recommender systems [22], [9, Chapter 27]. For a recommender systems setup, Resnick and Sami [22] and Friedman et al. [9, Chapter 27] provide algorithms which render a recommender system manipulation resistant i.e. truthful reporting of ratings is a reputation score maximizing strategy. In this paper, we explore the problem of strategic endorsement on ODFs and design a credit score robust to strategic coalition formation. For an online discussion forum setting, we believe our work in this paper is the first to exploit the community structure identified in the interactions graph to design a coalition resistant credit score function.

Coalition Formation. Coalition formation in a social network may be desirable in some situations but undesirable in others. Designing incentives for participants to generate the best answer by aggregating the information in a question answer forum is discussed in [11]. Sless et al. [24] propose a mechanism to facilitate the formation of a coalition for completing a mission critical task. In contrast to these works, our goal in this paper is to discourage coalition formation by designing appropriate credit score functions. The work by Niu et al. [19] is closest to ours; they propose an algorithm to detect collusive cheating in online shopping platforms by analyzing the underlying social network. Our work substantially differs from [19] as we consider question answer forums where each user is assigned a credit score and we provide a credit score function that prevents a collusive behaviour.

Community Detection with Modularity. Typically, a community detection problem is formulated as that of maximizing a certain community metric which acts as a quality measure of the detected communities. We choose modularity [8, 10] as a metric to detect the community structure in the network. The undirected version of the modularity [17, 18] based community detection problem is very well studied. In this paper, we consider modularity for weighted directed graphs as defined by Chen et al. [3]. For a detailed discussion on community detection, we refer the reader to [4].

3 THE MODEL

Let $N = \{1, 2, \ldots, n\}$ denote the set of all participating agents in a given forum and $p_i > 0$ represent the quality of agent i, i.e. the probability that a post by an agent i receives a popularity indicator from a random agent. Let the directed graph $G = (V, E)$ denote the popularity indicators graph with $V = N$ and an edge $(i, j) \in E$ denotes that an j receives a popularity indicator from agent i. We denote the total number of popularity indicators to an agent i in a graph G by $u_i(G)$ i.e. the total in-degree of node i.

Let Π denote the set of all partitions of N and $C = \{C_1, \ldots, C_l\} \in \Pi$ denote a *coalition structure* among the agents. If C_k is such that $C_k = \{j\}$, we say that the agent j has not formed a coalition with any agent. We overload the notation, $\emptyset := \{C_1 = \{1\}, \ldots, C_n = \{n\}\}$ to denote non-coalition. The coalition structure emerges as a result of a strategic coalition among utility maximizing agents. We next formalize this coalition formation model.

Coalition Formation Model

When agent i and agent j act independently, the directed edge from i to j is a realization of a Bernoulli trial with success probability p_j. However, when agents i and j form a coalition, the bias parameter $p >> p_j$ determines the edge probability. To study the effect of such coalitions on realized popularity indicators graph, we define a *Model Graph of a Coalition Structure.*

Definition 3.1. **Model Graph of a Given Coalition Structure** A *Model Graph of given a coalition structure* $C \in \Pi$, denoted by $\mathcal{N}_C$, is a weighted directed complete graph where nodes are the agents and the directed edge (i, j) has a weight

$$w_{i,j} = \begin{cases} p & \text{if } i \text{ and } j \text{ are in the same coalition} \\ p_j & \text{otherwise.} \end{cases}$$

Model Graph captures the expected popularity indicators recieved by each agents under given coalition structure. Note that the popularity indicators graph is a realization of the model graph. We denote by $k_i^{in}(\mathcal{N}_C) := \sum_{j \neq i} w_{j,i}$ and $k_i^{out}(\mathcal{N}_C) := \sum_{j \neq i} w_{i,j}$ the sum of weights of incoming and outgoing edges respectively from node i in model graph $\mathcal{N}_C$. Observe that $k_i^{in}(\mathcal{N}_C)$ is the expected number of popularity indicators an agent i receives. If $C = \emptyset$, $k_i^{in}(\mathcal{N}_\emptyset) = (n-1)p_i$ and if $C = N$, i.e. the agents form a grand coalition, then $k_i^{in}(\mathcal{N}_C) = (n-1)p$. Note that since $p >> p_i$, the value of expected popularity indicators to agent i is strictly increasing with

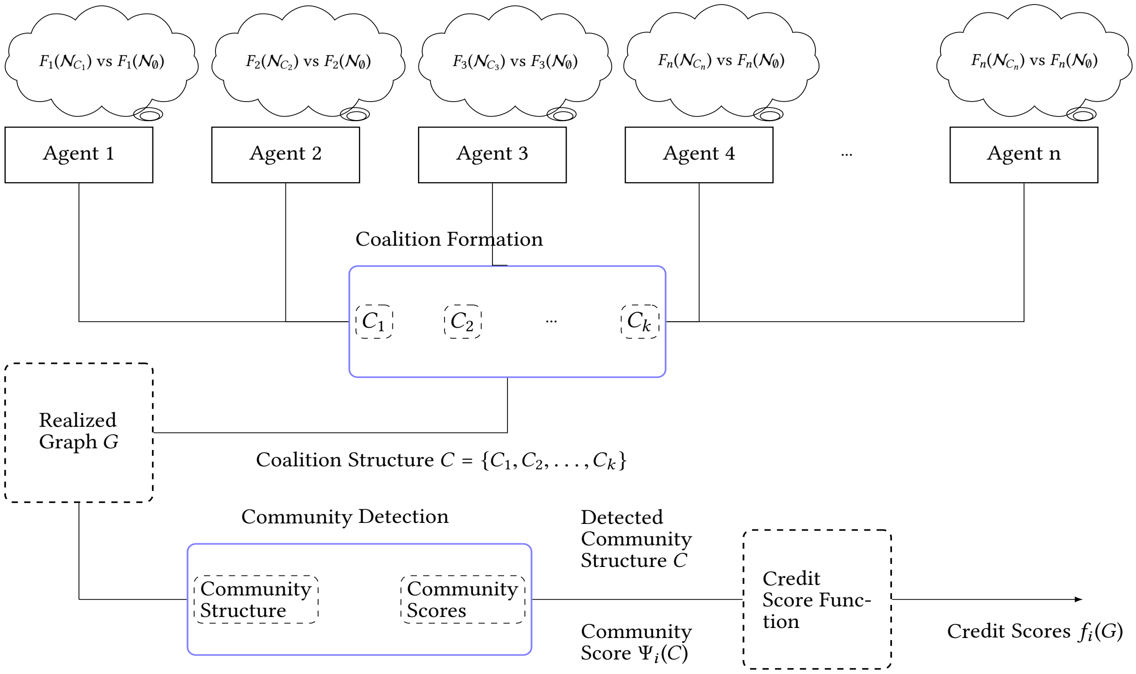

$F_1(\mathcal{N}_{C_1})$ vs $F_1(\mathcal{N}_\emptyset)$
$F_2(\mathcal{N}_{C_2})$ vs $F_2(\mathcal{N}_\emptyset)$
$F_3(\mathcal{N}_{C_3})$ vs $F_3(\mathcal{N}_\emptyset)$
$F_n(\mathcal{N}_{C_n})$ vs $F_n(\mathcal{N}_\emptyset)$
$F_n(\mathcal{N}_{C_n})$ vs $F_n(\mathcal{N}_\emptyset)$
Agent 1
Agent 2
Agent 3
Agent 4
...
Agent n
Coalition Formation
C_1
C_2
...
C_k
Realized Graph G
Coalition Structure $C = \{C_1, C_2, \ldots, C_k\}$
Community Detection
Community Structure
Community Scores
Detected Community Structure C
Community Score $\Psi_i(C)$
Credit Score Function
Credit Scores $f_i(G)$

Figure 1: Conceptual building blocks of the paper

the size of the coalition agent i is part of. Thus, the grand coalition will always give agents maximum number of popularity indicators. A naive score function which depends only on the number of popularity indicators received by agents is not strategy-proof as it incentivizes agents to form larger coalitions. Thus, it is required to design credit score function that identify and penalize the agents forming strategic coalitions.

The popularity indicators graph $G = (V, E)$ is realized as follows. The nodes are agents and the directed edge (i, j) is realized according to the independent Bernoulli trial with parameter $w_{i,j}$ (i.e. the edge weight $w_{i,j}$ in the Model Graph). Note that for a given underlying coalition structure C, there can be multiple graph realizations based on the results of Bernoulli trials. The agents are not fully aware of the realized graph which is formed based on the coalition structure. However, the agents have a partial knowledge about the realized graph as they know the coalitions they form. Thus, an agent makes decision to form a coalition based on the credit score received on model graph. We run a community detection algorithm on realized graph G to detect these dense substructures (i.e. community structure) in the graph. Based on the detected community structure and the community measure, we reward credit scores to each agent. Figure 1 summarizes the flow considered in this paper.

Coalition detection problem closely resembles the community detection problem as we look for dense subgraphs in both the cases. However, finding a community structure is not same as detecting a strategic coalition. The graph with a close-knit community structure may not have any coalition in it; it is natural to expect that a group of high-quality agents is detected as a community without being a strategic coalition. In this case, the detected community structure do not provide any useful information about the underlying coalition structure. In this paper, we make a simplifying but a practical assumption that the agents' qualities $p_i < 0.5$. We base our assumption on the observation that the popularity indicators are rare to obtain on a large forum. The popularity indicator received for the post depends on how many agents are interested to view the post concerning a particular topic and further on the popularity indicator to views ratio. In most of the large social networks such as Quora and StackExchange, the average upvotes to views ratio is very small e.g. as of 2017, Reddit saw 82.54 billion pageviews, 725.85 million comments, and 6.89 billion upvotes from its users [21]. In this paper we focus on designing a credit score function to disincentivize strategic coalition formation considering that the formed coalitions can be detected reasonably well. This leads to our first assumption.

A1 The community detection algorithm effectively detects the coalitions.

We leave the quantitative study of the effectiveness of several community detection algorithms to identify coalitions as a future work. Note that there can be other ways of manipulation: for example, an agent may not participate in any coalition formation, but can still maximize relative credit score by not awarding popularity indicator to other agents. We leave the study of such manipulations as an interesting future direction. In this paper we make the following assumption.

A2 The agents can manipulate only by forming coalitions.

4 COALITION RESISTANT CREDIT SCORE FUNCTIONS

In this section, we characterize a class of credit score functions which are coalition resistant and can be used to incentivize the agents to avoid coalition formation. Let G be the realized popularity indicators graph from $\mathcal{N}_C$. We run a community detection algorithm on G to identify the community structure C. Typically, a community detection algorithm finds a node partition which maximizes a community detection measure $\Psi : G \longrightarrow \mathbb{R}$. Community detection measure quantifies the *goodness* of a given subset $C_k \in C$ to be called a well-knit community. Let $\Psi^{C_k}(G)$ denote the community measure for a community $C_k \in C$ on G. Write $\Psi_i(G) = \Psi^{C_k}(G)$ for all i in C_k to represent that each agent in a community gets the same community score. We now define our credit score function as follows:

Definition 4.1. **Credit score function:** Given a popularity indicators graph G with the detected community structure $C = \{C_1, \ldots, C_l\}$, the credit score function $f := (f_1(G), f_2(G), \ldots f_n(G)) : G \longrightarrow \mathbb{R}^n$ is defined as

$$f_i(G) = u_i(G) - \beta_i \Psi_i(G). \tag{1}$$

Where, $\beta_i \geq 0$ and $u_i(G)$ represent the penalty parameter and the number of popularity indicators to agent i respectively.

If a coalition structure C is formed then the *expected credit score function* defined on model graph of a given coalition structure C is denoted by $F(\mathcal{N}_C) := (F_1(\mathcal{N}_C), F_2(\mathcal{N}_C), \ldots F_n(\mathcal{N}_C))$ and is given as: $F_i(\mathcal{N}_C) = k_i^{in}(\mathcal{N}_C) - \beta_i \Psi_i(\mathcal{N}_C)$.

A credit score function is said to be *Coalition resistant* if it assigns an agent the highest expected credit score when the agent does not join any coalition. We formalize this in the following definition.

Definition 4.2. **Coalition resistant credit score function:** Let $\mathcal{N}_C$ be the model grapf of a coalition structure C. We say a credit score function is coalition resistant if for all agents i

$$F_i(\mathcal{N}_C) \leq F_i(\mathcal{N}_\emptyset)$$

We first define a coalition identifying community metric and then show that any community detection algorithm which maximizes coalition identifying community metric can be used to design a coalition resistant credit score function.

Definition 4.3. **Coalition identifying community metric:** Let $\mathcal{N}_C$ be the model graph of a coalition structure C and $i \in C_k$, we say a community detection metric Ψ is *coalition identifying* if

$$\Psi_i(\mathcal{N}_C) \geq \Psi_i(\mathcal{N}_\emptyset) \; \forall \, C \neq \emptyset \; \in \Pi, \; \forall i \in N. \tag{2}$$

We now have the following theorem.

Theorem 4.4. *For a credit score function F to be coalition resistant, the community metric must be coalition identifying .*

Proof. Given F is coalition resistant then for a given value of β_i, we have:

$$F_i(C) \leq F_i(\emptyset)$$
$$\Rightarrow k_i^{in}(\mathcal{N}_C) - k_i^{in}(\mathcal{N}_\emptyset) \leq \beta_i [\Psi_i(\mathcal{N}_C) - \Psi_i(\mathcal{N}_\emptyset)]$$

Note that $k_i^{in}(\mathcal{N}_C) = (a-1)p + (n-a)p_i$, where $a = |C_k|$, and $k_i^{in}(\mathcal{N}_\emptyset) = (n-1)p_i$. Since, $p >> p_i$, we have $k_i^{in}(\mathcal{N}_C) - k_i^{in}(\mathcal{N}_\emptyset) > 0 \; \forall C \neq \emptyset$. Thus, for a given penalty parameter $\beta_i > 0$, we need $\Psi_i(\mathcal{N}_C) - \Psi_i(\mathcal{N}_\emptyset) \geq 0$ □

We now investigate modularity; a popular coalition detection metric in the next section and show that it is coalition identifying metric.

5 COALITION DETECTION USING MODULARITY

In this section, we first define modularity for weighted directed networks. We then consider a special case with $p = 1$ and show that under this assumption one can find a suitable penalty parameter β which renders credit score function coalition resistant. We then relax the assumption $p = 1$ and obtain a trade-off between the coalition size and the credit score optimality.

5.1 Community Detection and Modularity Metric

Let $C \in \Pi$ be a vertex partition of given graph G and $C_\ell \in C$ be a community. Modularity of C_ℓ is defined as the difference between intra-community edges observed and the intra-community edges expected in community C_ℓ in the model graph. More formally, the modularity of a directed graph given by [12] is defined as

$$\Psi^{C_\ell}(G) = \frac{1}{m} \sum_{i,j \in C_\ell} \left(A_{i,j} - \frac{k_i^{in} k_j^{out}}{m} \right), \tag{3}$$

Where, m is total number of edges, $A_{i,j} = \mathbb{1}_{(i,j) \in E}$ and $k_i^{in}(k_i^{out})$ is a number of incoming (outgoing) edges to (from) node i. The modularity score for a community structure $C = \{C_1, C_2, \ldots, C_\ell\}$ is defined as $\sum_{i=1}^{\ell} \Psi^{C_i}(G)$.

In this paper, we model agent behavior using the underlying complete and weighted model graph. An agent, in the absence of a prior knowledge of the realizations of popularity indicators (i.e. realized social graph) maximizes the credit score on corrosponding model graph. The above definition of modularity for directed graph can be easily extended to directed weighted graph as follows. As earlier, consider a partition $C = \{C_1, C_2, \ldots, C_\ell\}$ of a complete weighted graph G. The modularity of community C_ℓ is defined as

$$\Psi^{C_\ell}(G) = \frac{1}{m} \sum_{i,j \in C_\ell} \left(a_{i,j} - \frac{k_i^{in} k_j^{out}}{m} \right) \tag{4}$$

Where, $m = \sum_{i,j \in G} a_{i,j})$ is the total edge weight, $a_{i,j} = w_{i,j}$ and $k_i^{in}(k_i^{out})$ is the total incoming (outgoing) edge weight of node i. A modularity maximization based community detection algorithm returns the community structure that maximizes the modularity score. We now present our theoretical analysis with respect to the modularity measure.

5.2 Theoretical Analysis

We start with the following observation: if an agent does not form any coalition, the agent's expected modularity is non-positive under assumption A1.

OBSERVATION 1. *Let C be the coalition structure where agent i does not form a coalition then* $\Psi_i(\mathcal{N}_C) < 0$

This can be easily shown with the assumption A1:

$$\Psi_i(\mathcal{N}_C) = \frac{1}{m}\sum_{i,j\in\{i\}} A_{ij} - \frac{k_i^{in}k_j^{out}}{m} = -\frac{k_i^{in}k_i^{out}}{m^2} < 0 \quad (A_{ii} = 0)$$

When agents do not form any coalition the assumption that agents are detected in a singleton community is too restrictive from a practical standpoint. We relax assumption in section 5.3.

We begin our analysis by considering that observation 1 holds. In order to prove that modularity is a coalition identifying community metric, it is enough to prove that if an agent forms a coalition then the expected modularity score is positive. We first present a special case where every within coalition agent assigns popularity indicators to each other (i.e. $p = 1$). In the next part, we show the trade-off between the coalition size to be detected and the bias probability range for which the credit scores poses coalition identifying property. Let the agent i gets detected in a community $C_\ell \subset C$. We denote $a = |C_\ell|, \gamma = \sum_{s\in N} p_s$ and $\alpha = \sum_{s\notin C_\ell} p_s$.

CASE 1 : $p = 1$

In this case, we assume that an agent awards a popularity indicator to all in-coalition agents i.e. $p = 1$. For instance, in the follow network each agent in coalition C_k follow $j \in C_k$ with probability 1 and $j \notin C_k$ with probability p_j. In upvote network all the in-coalition agents upvote posts by each other and remain indifferent while upvoting the posts by agents outside the coalition.

THEOREM 5.1. *Let* $p_i \leq \frac{1}{2}$ *for all i and* $p = 1$ *then modularity is a coalition identifying community metric for all* $n \geq 7$ *and* $3 \leq a \leq n - 4$.

PROOF. It is enough to show that $\Psi_i(\mathcal{N}_C) \geq 0$ for all $C \in \Pi$ (eq. (2)). Let $i \in C_\ell$ for some C_ℓ. Recall the notation, $a = |C_\ell|, \gamma = \sum_{s\in N} p_s$ and $\alpha = \sum_{s\notin C_\ell} p_s$. For any C s.t. $i \in C_\ell \subset C$ we have

$$\begin{aligned} m &\geq a(a-1) + a\alpha + \sum_{j\notin C_l}(\gamma - p_j) \\ &\qquad \text{(there can be other coalitions other than } C_\ell\text{)} \\ &\geq a(a-1) + a\alpha + (n-a)\gamma - \alpha \\ &\geq a(a-1) + \alpha(a-1) + (n-a)\gamma. \end{aligned}$$

Further, $k_i^{in} = (a-1) + (n-a)p_i$ and $k_i^{out} = (a-1) + \alpha \ \forall i \in C_\ell$. Putting these values in, Equation 3

$$\begin{aligned} \Psi_i(\mathcal{N}_C) &= \frac{1}{m^2}\Big(m \times a(a-1) - \sum_{i,j\in C_\ell} k_i^{in}k_j^{out}\Big) \\ &= \frac{1}{m^2}\Big(m \times a(a-1) - (a(a-1) \\ &\qquad + (n-a)(\gamma-\alpha))(a(a-1) + a\alpha)\Big) \\ &\geq \frac{1}{m^2}\Big(a^2(a-1)^2 + a(a-1)^2\alpha + a(a-1)(n-a)\gamma - \\ &\qquad [a^2(a-1)^2 + a^2(a-1)\alpha + a(a+\alpha-1)(n-a)(\gamma-\alpha)]\Big) \\ &\geq \frac{a}{m^2}\Big(-(a-1)\alpha + (n-a)\alpha(-\gamma + a - 1 + \alpha)\Big) \\ &\geq \frac{\alpha a}{m^2}(-(a-1) - (n-a)(\gamma-\alpha) + (n-a)(a-1)) \\ &\geq \frac{\alpha a}{m^2}(-(a-1) + (n-a)(a-1-(\gamma-\alpha))) \\ &\qquad\qquad (\ a - (\gamma-\alpha) \geq a/2 \text{ as } p_i < 0.5) \\ \Psi_i(\mathcal{N}_C) &\geq \frac{\alpha a}{2m^2}(n(a-1) - a(a+1) + 2) \end{aligned}$$

It is easy to check that $\Psi_i(\mathcal{N}_C) > 0, \forall n \geq 7$ and $3 \leq a \leq n - 4$ □

We now prove that with the appropriate value of $\beta_i' s$, the proposed credit score function is coalition resistant with respect to the modularity as a community detection metric.

THEOREM 5.2. *Let* $i^* = \min_i p_i > 0$, *the proposed credit score function is coalition resistant for* $\beta_i \geq \frac{2n^2(n-1)^2}{n-3}\frac{(1-p_i)}{p_{i^*}}$.

PROOF. Note that $\Delta k_i^{in}(\mathcal{N}_C) \leq (a-1)(1-p_{i^*})$, to see this observe that

$$\begin{aligned} \Delta k_i^{in}(\mathcal{N}_C) &= k_i^{in}(\mathcal{N}_C) - k_i^{in}(\mathcal{N}_\emptyset) \\ &= a - 1 + (n-a)p_i - (n-1)p_i \\ &= (a-1)(1-p_i) \end{aligned}$$

From ɵObservation 1, we have $\Delta\Psi_i(\mathcal{N}_C) = \Psi_i(\mathcal{N}_C) - \Psi_i(\mathcal{N}_\emptyset) > \Psi_i(\mathcal{N}_C)$. Now, from Theorem 5.1 we have

$$\Psi_i(\mathcal{N}_C) \geq \frac{p_{i^*}a}{2m^2}(n(a-1) - a(a+1) + 2)$$

For $F_i(.)$ to be coalition resistant, we need

$$\beta_i \geq \max_a \frac{(1-p_i)}{p_{i^*}}\frac{2m^2}{a(n-a-2)}$$

as, $m \leq n(n-1)$ and $a(n-a-2) \geq n - 3 \ \forall\ 1 \leq a \leq n - 3$, it is enough to have

$$\beta_i \geq \frac{2n^2(n-1)^2}{n-3}\frac{(1-p_i)}{p_{i^*}}$$

□

CASE 2 : $p < 1$

Let us now consider that the agents in a coalition upvote each other based on a common agreed upon bias parameter $p < 1$. We show that even in this case one can detect the communities under some relaxed conditions on the size of a coalition and the total number of agents. We begin with the following simple observation,

OBSERVATION 2. $\frac{x(y-x)}{(x-1)(y-x-1)} \leq \frac{4}{3} \forall y \geq 100$ *and* $x \in [4, y-4]$

We next prove that modularity is a coalition identifying metric for $n \geq 100$ and $a \in [4, n-4]$.

THEOREM 5.3. *Modularity is a coalition identifying community metric if* $p > \frac{2}{3}, p_i \leq \frac{1}{2} \forall i \in N, n \geq 100$ *and* $a \in [4, n-4]$.

Proof. Using the same notation as in Theorem 5.1 we have,

$$
\begin{aligned}
m &\geq a(a-1)p - \alpha(a-1) - (n-a)\gamma \\
k_i^{in} &= (a-1)p + (n-a)p_i \\
k_i^{out} &= (a-1)p\alpha \\
\Psi_i(\mathcal{N}_C) &= \frac{a}{m^2}\Big(a^2(a-1)^2 - \alpha a(a-1)^2 p + a(a-1)(n-a)\gamma p - \\
&\quad \Big(a(a-1)p - (n-a)(\gamma-\alpha)\Big)\Big(a(a-1)p - a\alpha\Big)\Big) \\
&\geq \frac{a}{m^2}\Big((a-1)^2\alpha p - a(a-1)\alpha p + \\
&\quad (n-a)(a-1)\alpha p - (n-a)(\gamma-\alpha)\Big) \\
&= \frac{a\alpha}{m^2}\Big(-p(a-1) + (n-a)(a-1)p - (n-a)(\gamma-\alpha)\Big) \\
&= \frac{a\alpha}{m^2}\Big((a-1)p(n-a-1) - (n-a)(\gamma-\alpha)\Big) \\
&\geq \frac{a\alpha}{m^2}\Big((a-1)(n-a-1)p - (n-a)\frac{a}{2}\Big) \quad \Big(\text{as } \gamma-\alpha \leq \frac{a}{2}\Big)
\end{aligned}
$$

$$
\implies \Psi_i(\mathcal{N}_C) \geq 0 \text{ iff } p \geq \frac{a(n-a)}{2(a-1)(n-a-1)}.
$$

Thus for all values of coalition size a such that $\forall 4 \leq a \leq n-4$ and for all ODF's with size $n \geq 100$ modularity is a coalition identifying metric for $p \geq \frac{2}{3} \geq \frac{a(n-a)}{2(a-1)(n-a-1)}$ (by Observation 2). □

For a typical social network where the number of nodes is large ($n \gg 10^3$) compared to the coalition size a (i.e. $\frac{n-a}{n-a-1} \approx 1$), we observe that the modularity identifies a the coalitions correctly for a reasonable coalition sizes ($log(n) \leq a \leq n - log(n)$) and bias parameter range $1 \geq p \geq \frac{1}{2}$.

Theorem 5.4. *For every $n \geq 100$ and $a \in [4, n-4]$ with $p > \frac{2}{3}$, the proposed credit score function is coalition resistant for $\beta_i \geq \frac{(p-p_i)}{p_{i^*}} \frac{n^2(n-1)^2}{(n-5)(p-\frac{2}{3})}$*

Proof. Following similar arguments as in Theorem 5.2 we have,

$$
\begin{aligned}
\Delta k_i^{in}(\mathcal{N}_C) &= (a-1)p + (n-a)p_i - (n-1)p_i \\
&= (a-1)(p-p_i)
\end{aligned}
$$

Using Theorem 5.3,

$$
\begin{aligned}
\Delta\Psi_i(\mathcal{N}_C) &\geq \frac{a\alpha}{m^2}\Big((a-1)(n-a-1)p - (n-a)\frac{a}{2}\Big) \\
&\geq \frac{ap_{i^*}}{m^2}\Big((a-1)(n-a-1)p - (n-a)\frac{a}{2}\Big)
\end{aligned}
$$

We need,

$$
\begin{aligned}
\beta_i &\geq \frac{(p-p_{i^*})(a-1)m^2}{ap_i(a-1)(n-a-1)p - (n-a)\frac{a}{2}} \\
&= \frac{(p-p_i)}{p_i}\frac{m^2}{(n-a-1)\Big(p - \frac{a(n-a)}{2(n-a-1)(a-1)}\Big)} \\
&\geq \frac{(p-p_i)}{p_{i^*}}\frac{n^2(n-1)^2}{(n-5)(p-\frac{2}{3})} \qquad (\text{ Observation 2})
\end{aligned}
$$

□

Note that agents can strategize over the bias parameter p. We proved in Theorem 5.3 modularity metric is coalition identifying for $p > \frac{2}{3}$, however, we would like to report here that the coalitions with $p \leq \frac{2}{3}$ may not get correctly identified with modularity maximization algorithms. This renders the proposed credit score function coalition resistant instead of coalition-proof. We leave the design of community detection metric which renders the credit score function coalition resistant for any value of a bias parameters as an interesting future work.

5.3 Relaxing Assumption A1

In the next theorem, we prove that even if agents do not form any coalition but gets detected in some non-singleton community, the expected modularity score is still negative thus resulting in non-negative credit score for the agent.

Lemma 5.5. *Let $\{C_1, C_2, \ldots, C, \ldots, C_\ell\}$ be the community structure detected by modularity maximization algorithm and $i \in C$ with $\emptyset$ as a coalition structure then, $\Psi_i(\mathcal{N}_\emptyset) < 0$.*

Proof. In a model graph of $\emptyset$, the expected number of edges is given as $m = \sum_i (n-1)p_i = (n-1)\gamma$. If a community detection algorithm detects i in community C then,

$$
\begin{aligned}
\Psi_i(\mathcal{N}_\emptyset) &= \frac{1}{m}\sum_{s,t\in C}\left(A_{s,t} - \frac{k_s^{in}(\mathcal{N}_\emptyset)k_t^{out}(\mathcal{N}_\emptyset)}{m}\right) \\
\sum_{s,t\in C} A_{s,t} &= \sum_{s\in C_\ell} p_s(a-1) = (\gamma-\alpha)(a-1) \\
\sum_{s,t\in C} k_s^{in}(\mathcal{N}_\emptyset)k_t^{out}(\mathcal{N}_\emptyset) &= \sum_{s\in C}\left(k_s^{in}(\mathcal{N}_\emptyset)\sum_{t\in C}k_t^{out}(\mathcal{N}_\emptyset)\right) \\
&= \sum_{s\in C}(n-1)p_s \sum_{t\in C}(\gamma - p_t) \\
&= (n-1)(\gamma-\alpha)(a\gamma - (\gamma-\alpha)) \\
&= (n-1)(\gamma-\alpha)(\gamma(a-1)+\alpha)
\end{aligned}
$$

Putting everything together,

$$
\Psi_i(\mathcal{N}_\emptyset) = \frac{(\gamma-\alpha)}{m}\Big(\frac{m \times (a-1) - (n-1)(\gamma(a-1)+\alpha)}{m}\Big)
$$

using $m = (n-1)\gamma$ we get,

$$
\Psi_i(\mathcal{N}_\emptyset) = \frac{-\alpha(\gamma-\alpha)}{(n-1)\gamma^2} < 0
$$

□

5.4 Selection of Penalty Parameters (β_is)

Setting $\beta_i = \infty$ trivially satisfies the *coalition resistant* property as no agent can improve their scores by forming a coalition. We want the lowest possible value of $\beta_i > 0$ (the lower bound) such that coalition property still holds. Note that, the penalty parameter β_i depends on the quality p_i of the agent. It is natural to expect that the low-quality agents get penalized heavily when found in a close-knit community. Further, observe that the value of β_i depends on the total number of agents which again can be justified by the fact that $\Psi_i(.)$ is bounded and do not depend on the size of the graph whereas k_i^{in} does. As a result of high value of β_i, the agents may end up getting large negative scores leading discouraging agents to participate in the forum.

Another possibility of computing the value of β_i is to learn p_i's by deploying a suitable machine learning algorithm. A naive learning algorithm may be easily manipulated by the agents. The learning algorithm designed without game theoretic considerations may encourage all the agents to form a grand coalition leading estimate of p_i to 1 for all i. This lead to setting $\beta_i = 0$ for all the agents and thus coalition resistant property is lost. Though learning in presence of strategic agents has been addressed in the literature [1, 5], none of both the techniques are appropriate in our context.

6 COALITION RESISTANT LEARNING

The lower bound on the penalty parameter given in Theorem 5.2 depend on the true quality of the agent. However, in a typical practical situation, the true quality of agents may not be known a priori. We now show that a learning algorithm, with a pessimistic estimate of the agent quality based on the popularity indicators received from the outside community agents, can achieve a coalition resistant property. Note here that the within coalition agents are not honestly upvoting based on the quality and hence should be excluded when we estimate the true quality of an agent. We next formalize the above argument. We begin by stating a well-known Hoeffding's inequality result for our ready reference.

Lemma 6.1. ***Hoeffding's Inequality*** *Let $X_1, X_2, \ldots, X_n$ be i.i.d random variables with mean μ such that $X_i \in [a, b] \forall i$ where $-\infty < a < b < +\infty$ then*

$$\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \geq t\Big) \leq \exp\Big(-\frac{2nt^2}{(b-a)^2}\Big)$$

and

$$\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \leq -t\Big) \leq \exp\Big(-\frac{2nt^2}{(b-a)^2}\Big)$$

Learning Agent Qualities from Detected Community Structure

Let $\{C_1, C_2, \ldots C_l\}$ be a coalition structure and let $X_{i,j}$ be a random variable given by

$$X_{i,j} = \begin{cases} \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{otherwise} \end{cases} & \text{if } i, j \text{ are in the same coalition} \\ \begin{cases} 1 & \text{w.p. } p_i \\ 0 & \text{otherwise} \end{cases} & \text{otherwise} \end{cases} \tag{5}$$

Let agent $i \in C$ and denote by $\hat{p_i} := \frac{1}{n-|C|} \sum_{j=1, j \notin C}^{n-|C|} X_{i,j}$, the empirical mean of the quality estimate from outside group agents. Further let $\hat{p} = \frac{1}{|C|} \sum_{j=1, j \in C}^{|C|} X_{i,j}$ be the empirical estimate of the bias parameter of coalition C. Then from Lemma 6.1, with probability at least $1 - \frac{2}{(n-|C|)^4}$ the following is true,

$$\underbrace{\hat{p}_i - \sqrt{\frac{2\log(n-|C|)}{n-|C|}}}_{\hat{p}_i^-} \leq p_i \leq \underbrace{\hat{p}_i + \sqrt{\frac{2\log(n-|C|)}{n-|C|}}}_{\hat{p}_i^+}, \tag{6}$$

and with probability at least $1 - \frac{2}{|C|^4}$ we have,

$$\underbrace{\hat{p} - \sqrt{\frac{2\log(|C|)}{|C|}}}_{\hat{p}^-} \leq p \leq \underbrace{\hat{p} + \sqrt{\frac{2\log(|C|)}{|C|}}}_{\hat{p}^+}, \tag{7}$$

Since, $2/3 + \epsilon \leq p \leq 1$, we use the following bounds:

$$\hat{p}^- = max\left(\frac{2}{3} + \epsilon, \hat{p}^-\right) \text{ and } \hat{p}^+ = min\left(1, \hat{p}^+\right)$$

The next lemma shows that using the appropriate bounds on the quality parameter and the bias parameter one can obtain a lower bound on penalty parameter which, with high probability, renders credit score function "coalition resistant".

Lemma 6.2. *The proposed credit score function is coalition resistant with penalty parameter $\hat{\beta}_i = \frac{n^2(n-1)^2}{n-5} \frac{(\hat{p}^+ - \hat{p}_i^-)}{\hat{p}_i^-(\hat{p}^- - 2/3)}$.*

Proof. From Theorem 5.4, we have if $\beta_i \geq \frac{n^2(n-1)^2}{n-5} \frac{(p-p_i)}{p_i(p-2/3)}$ then the proposed credit score function is coalition resistant for all coalitions C. Thus, it is enough to prove that $\hat{\beta}_i \geq \frac{n^2(n-1)^2}{n-5} \frac{(p-p_i)}{p_i(p-2/3)}$. It is easy to verify that this holds with high probability i.e. with probability $min\left(1 - \frac{2}{(n-|C|)^4}, 1 - \frac{2}{|C|^4}\right)$ using Equations (6) and (7). □

In the next section, we evaluate the proposed credit score function by simulations with known ground truth.

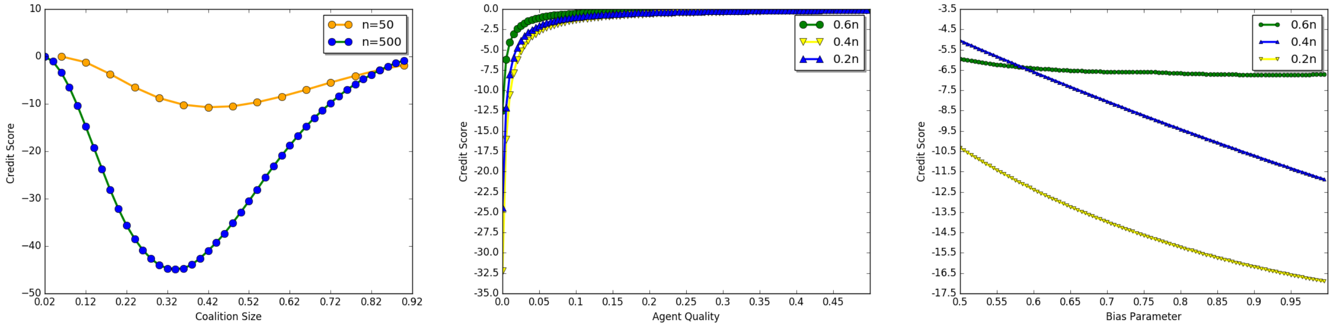
7 SIMULATION ANALYSIS

In this section, we evaluate and validate the proposed credit score function by performing simulations on representative, synthetic datasets. We run experiments with $n = 20$, 50 and 500 agents with randomly generated quality parameter for each agent. The objectives are (1) to investigate the effect of strategic (i.e. coalition size) as well as intrinsic (quality, bias) parameters on the scores awarded to the agents by the algorithm and (2) validate the assumption A1. We compute an optimal modularity maximizing community structure for $n = 20$ and $n = 50$. As modularity maximization is NP-Hard [2] we use Walktrap algorithm to test our hypothesis for $n = 500$.

Effect of Input Parameters on Credit Score Function

We fix an agent i with quality parameter p_i, generated randomly from a uniform distribution with support $[0, 0.5]$. The credit scores are scaled down by n^2.

Coalition Size: We vary the size of the coalition containing agent i from 1 to $(n-4)$ by randomly adding non-coalition agent into the coalition, one at a time. We run the experiment for 100 iterations with different quality parameters. In each iteration, we randomly add agents to the coalition having i and take the average credit score of agent i. We consider the bias parameter $p = 1$ for this experiment; also we use penalty parameter $\beta_i = \frac{2n^2(n-1)^2}{n-3} \frac{(1-p_i)}{p_{i^*}}$ as proposed in Theorem 5.2. Figure 2 shows the average credit score of the agent i vs. coalition size (fraction of agents forming coalition). As we increase the coalition size, the modularity value of the detected community increases, leading to a dip in credit score values around



credit score vs coalition size **credit score vs bias parameter** **credit score vs agent quality**

Figure 2: Effect of input parameters on credit score function

a coalition size of $\sqrt{n}$. If we further increase the coalition size, the number of extra popularity indicators received from in-coalition agents dominates the change in modularity score.

Agent Quality: We expect that a low quality agent, if found in a well knit community should be penalized more. We sample the quality of agents other than i uniformly from $[0, 0.5]$. We vary agent i's quality from 10^{-4} to 0.5. Figure 2 shows the relation between the awarded credit score to agent i and agents i's quality. The figures show the results with $n = 50$ and for different values of coalition sizes (i.e. $0.2n, 0.4n, 0.6n$). A similar trend is observed for higher values of n. We use the penalty parameter β_i (same as given in Theorem 5.2).

Bias Parameter: We increase the bias parameter in the range $0.5 \leq p \leq 1$ with which the agent awards popularity indicators to other within-coalition agents. We run this experiment for different coalition sizes ($0.2n$, $0.4n$ and $0.6n$), by randomly forming a coalition with specified coalition size. It is seen from Figure 2 that as we increase the bias parameter, the awarded credit score to an agent decreases.

Validating Assumption A1

In this experiment, we validate the assumption A1 by calculating for different coalition sizes, the average number of agents which are not part of a given coalition but falsely detected in a community (false positives) as well as the average number of agents who are part of coalition and are not detected in a community (false negatives). The results in Table 1 are averaged over 100 iterations with quality parameters randomly generated from uniform distribution over $[0, 0.5]$. We use $n = 50$, $p = 0.7$ and communities are obtained by optimally computing the modularity maximizing vertex partition. The results improve for higher values of bias parameter p as the coalitions become increasingly close knit. It can be seen from this table that for coalition of size 5 or more there are hardly false positives. In addition, for coalition of size ≤30, there are not many true negatives. Thus, the assumption A1 that community detection algorithm detects coalitions correctly is valid for most of the practical sizes of coalitions. In addition, in an unlikely event of very large coalition getting formed, the honest agents are unlikely to get penalized, though some coalition forming agents may escape penalties.

coalition	false	false
size	positives	negatives
3	1.41	0
5	0	0
10	0	0
15	0	0
20	0	0.02
30	1.4	4.69
45	0.76	12.38

Table 1: Errors in detecting coalitions

8 DISCUSSION AND FUTURE WORK

In this work, we designed coalition resistant credit score functions for online discussion forums by considering the popularity indicators as proxies for the agents' activities on the forum. We used modularity maximization based community detection algorithms to detect coalitions and adjust scores appropriately. An immediate future direction of work is to analyze other community detection metrics for robustness against coalition formation. We also leave as a future work, the design of a credit function which effectively aggregate the credit scores obtained from popularity indicator graphs such as follow network, like network, etc. The proposed credit function is not equipped to handle negative votes (downvotes, dislikes, unfollows etc) and design of functions for such cases forms an interesting future direction. One can also consider settings where agents can manipulate the credit scores by means other than coalition formation.

REFERENCES

[1] Moshe Babaioff, Yogeshwer Sharma, and Aleksandrs Slivkins. Characterizing truthful multi-armed bandit mechanisms. *SIAM Journal on Computing*, 43(1): 194–230, 2014.

[2] Ulrik Brandes, Daniel Delling, Marco Gaertler, Robert Görke, Martin Hoefer, Zoran Nikoloski, and Dorothea Wagner. Maximizing modularity is hard. *arXiv preprint physics/0608255*, 2006.

[3] M. Chen, K. Kuzmin, and B. K. Szymanski. Community detection via maximization of modularity and its variants. *IEEE Transactions on Computational Social Systems*, 1(1):46–65, 2014.

[4] Michele Coscia, Fosca Giannotti, and Dino Pedreschi. A classification for community discovery methods in complex networks. *Statistical Analysis and Data Mining*, 4(5):512–546, 2011.

[5] Ofer Dekel, Felix Fischer, and Ariel D Procaccia. Incentive compatible regression learning. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 884–893, 2008.

[6] Chrysanthos Dellarocas. Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management Science*, 52(10):1577–1593, 2006.

[7] Amir Fayazi, Kyumin Lee, James Caverlee, and Anna Squicciarini. Uncovering crowdsourced manipulation of online reviews. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 233–242, 2015.

[8] Santo Fortunato and Marc Barthélemy. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36–41, 2007.

[9] Eric Friedman, Paul Resnick, and Rahul Sami. Manipulation-resistant reputation systems. In *Algorithmic Game Theory*, chapter 27, pages 677–697. Cambridge University Press Cambridge, UK, 2007.

[10] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[11] Shaili Jain, Yiling Chen, and David C. Parkes. Designing incentives for online question and answer forums. In *Proceedings of the 10th ACM Conference on Electronic Commerce*, EC '09, pages 129–138, 2009.

[12] Elizabeth A Leicht and Mark EJ Newman. Community structure in directed networks. *Physical review letters*, 100(11):118703, 2008.

[13] Baichuan Li, Tan Jin, Michael R Lyu, Irwin King, and Barley Mak. Analyzing and predicting question quality in community question answering services. In *Proceedings of the 21st International Conference on World Wide Web*, pages 775–782, 2012.

[14] Yuli Liu, Yiqun Liu, Ke Zhou, Min Zhang, and Shaoping Ma. Detecting collusive spamming activities in community question answering. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1073–1082, 2017.

[15] Arjun Mukherjee, Bing Liu, and Natalie Glance. Spotting fake reviewer groups in consumer reviews. In *Proceedings of the 21st international conference on World Wide Web*, pages 191–200, 2012.

[16] Atif Nazir, Saqib Raza, and Chen-Nee Chuah. Unveiling facebook: a measurement study of social network based applications. In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, pages 43–56, 2008.

[17] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[18] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74:036104, 2006.

[19] J. Niu, L. Wang, Y. Chen, and W. He. Detecting collusive cheating in online shopping systems through characteristics of social networks. In *2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 311–316, 2014.

[20] Louise F. Pendry and Jessica Salvatore. Individual and social benefits of online discussion forums. *Computers in Human Behavior*, 50:211 – 220, 2015.

[21] Reddit. Reddit — Wikipedia, the free encyclopedia, 2017. URL https://en.wikipedia.org/wiki/Reddit. [Online; accessed 11-August-2017].

[22] Paul Resnick and Rahul Sami. The influence limiter: provably manipulation-resistant recommender systems. In *Proceedings of the 2007 ACM conference on Recommender systems*, pages 25–32, 2007.

[23] Chirag Shah and Jefferey Pomerantz. Evaluating and predicting answer quality in community QA. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 411–418, 2010.

[24] Liat Sless, Noam Hazon, Sarit Kraus, and Michael Wooldridge. Forming coalitions and facilitating relationships for completing tasks in social networks. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS '14, pages 261–268, 2014.