

**Original citation:**

Grandi, Umberto and Turrini, Paolo (2016) *A network-based rating system and its resistance to bribery*. In: Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, USA, 9-15 Jul 2016. Published in: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence pp. 301-307.

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/97682>

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

© ACM, 2016. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence pp. 301-307. (2016) <https://dl.acm.org/citation.cfm?id=3060664>

**A note on versions:**

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)

# A Network-Based Rating System and its Resistance to Bribery

**Umberto Grandi**

University of Toulouse  
France  
umberto.grandi@irit.fr

**Paolo Turrini**

Imperial College London  
United Kingdom  
p.turrini@imperial.ac.uk

## Abstract

We study a rating system in which a set of individuals (e.g., the customers of a restaurant) evaluate a given service (e.g. the restaurant), with their aggregated opinion determining the probability of all individuals to use the service and thus its generated revenue. We explicitly model the influence relation by a social network, with individuals being influenced by the evaluation of their trusted peers. On top of that we allow a malicious service provider (e.g., the restaurant owner) to bribe some individuals, i.e., to invest a part of his or her expected income to modify their opinion, therefore influencing his or her final gain. We analyse the effect of bribing strategies under various constraints, and we show under what conditions the system is bribery-proof, i.e., no bribing strategy yields a strictly positive expected gain to the service provider.

## 1 Introduction

Imagine to be the owner of a new and still relatively unknown restaurant. The quality of food is not spectacular and the customers you have seen so far are only limited to a tiny number of friends of yours. Your account on Tripadvisor® has received no review and your financial prospects look grim at best. There is one easy solution to your problems: you ask your friends to write an enthusiastic review for you, in exchange for a free meal. After this, Tripadvisor® lists your restaurant as excellent and the number of customers, together with your profit, suddenly flourishes.

Systems such as Tripadvisor®, where a small proportion of customers writes reviews and influences a large number of potential customers, are not *bribery-proof*: each restaurant owner - or the owner of whichever service - is able to offer a compensation - monetary or not - in exchange for positive evaluation, having an impact on the whole set of potential customers. Tripadvisor® is based on what we call “Objective Rating”, or  $\mathbb{O}$ -RATING: individual evaluations are aggregated into a single figure, which is seen by, and thus influences, every potential customer.

What we study in this paper is a system in which each individual *only* receives the evaluation given by the set of trusted peers, his or her friends, and only this aggregated opinion

influences his or her decision. This is what we call “Personalised Rating”, or  $\mathbb{P}$ -RATING, which can be seen a generalisation of  $\mathbb{O}$ -RATING in which influence has a complex network-structure. So, while in the case of  $\mathbb{O}$ -RATING the restaurant owner knows exactly how influence flows among the customers, this might not be the case with  $\mathbb{P}$ -RATING.

**Our contribution** We analyse the effect of bribing strategies in the case of  $\mathbb{O}$ -RATING and  $\mathbb{P}$ -RATING under various constraints, depending on the presence of customers who do not express any opinion and the knowledge of the network by the service provider: the exact network is known, the network is known but not the customers’ exact position, the network is completely unknown. We show under what conditions the system is bribery-proof, i.e., there is no bribe yielding a strictly positive expected gain to the service provider, and we provide algorithms for the computation of (all) optimal bribing strategies when they exist.

Intuitively, being able to know and bribe influential customers is crucial for guaranteeing a positive expected reward of a bribing strategy. However, while with large populations of non-voters “random” bribes can still be profitable, the effect of  $\mathbb{P}$ -RATING is largely different from that of  $\mathbb{O}$ -RATING and, as we show, the expected profit in the former can be severely limited and drops below zero in all networks, under certain (mild) conditions on the cost of bribes.

Our study can be applied to all situations in which individuals influence one another in the opinion they give and bribery can have a disruptive role in determining collective decisions.

**Related research lines** Our approach relates to several research lines in artificial intelligence, game theory and (computational) social choice [Brandt *et al.*, 2015].

**Network-based voting and mechanism design** We study social networks in which individuals’ local decisions can be manipulated to modify the resulting global properties. A similar approach is taken by Apt and Markakis [2014] and Simon and Apt [2015], which study the changes on a social network needed to make a certain product adopted among users. Further contributions include rational secret sharing and multi-party computation [Abraham *et al.*, 2006], the strategic manipulation of peer reviews [Kurokawa *et al.*, 2015], and the growing literature on voting in social networks [Conitzer, 2012; Salehi-Abari and Boutilier, 2014; Elkind, 2014; Tsang *et al.*, 2015; Procaccia *et al.*, 2015].

**Lobbying and Bribery** Our framework features an external agent trying to influence individual decisions to reach his or her private objectives. Lobbying in decision-making is an important problem in the area of social choice, from the seminal contribution of [Helpman and Persson, 1998] to more recent studies in multi-issue voting [Christian *et al.*, 2007]. Lobbying and bribery are also established concepts in computational social choice, with their computational complexity being analysed extensively [Faliszewski *et al.*, 2009; Baumeister *et al.*, 2011; Bredereck *et al.*, 2014; 2016].

**Reputation-based systems** We study the aggregation of possibly insincere individual evaluations by agents that can influence one another through trust relations. In this sense ours can be seen as a study of reputation in Multi Agent Systems, which has been an important concern of MAS for the past decades [Conte and Paolucci, 2002; Sabater and Sierra, 2005; Garcin *et al.*, 2009]. In particular, our framework treats reputation as a manipulable piece of information, not just a static aggregate of individual opinions, coherently with the work of Conte *et al.* [2008] and Pinyol and Sabater-Mir [2013].

**Paper structure** Section 2 presents the basic setup, introducing  $\mathbb{O}$ -RATING,  $\mathbb{P}$ -RATING and bribing strategies. Section 3 focusses on  $\mathbb{O}$ -RATING, studying its bribery-proofness under various knowledge conditions. Section 4 evaluates  $\mathbb{P}$ -RATING against the same knowledge conditions. In Section 5 we compare the two systems, taking the cost of bribery into account. We conclude by summarising the main findings and pointing at future research directions (Section 6).

## 2 Basic setup

In this section we provide the basic formal definitions.

### 2.1 Restaurant and customers

Our framework features an object  $r$ , called *restaurant*, being evaluated by a finite non-empty set of individuals  $C = \{c_1, \dots, c_n\}$ , called *customers*. Customers are connected by an undirected graph  $E \subseteq C \times C$ , called the *customers network*. Given  $c \in C$  we call  $N(c) = \{x \in C \mid (c, x) \in E\}$  the *neighbourhood* of  $c$ , always including  $c$  itself.

Customers concurrently submit an *evaluation* of the restaurant, drawn from a set of values  $Val \subseteq [0, 1]$ , together with a distinguished element  $\{*\}$ , symbolising no opinion. Examples of values are the set  $[0, 1]$  itself, or a discrete assignment of 1 to 5 stars, as common in online rating systems. We make the assumption that  $\{0, 1\} \subseteq Val$  and that  $Val$  is closed under the operation  $\min\{1, x + y\}$  for all  $x, y \in Val$ . The vast majority of known rating methods can be mapped onto the  $[0, 1]$  interval and analysed within our framework.

We represent the evaluation of the customers as a function  $eval : C \rightarrow Val \cup \{*\}$  and define  $V \subseteq C$  as the subset of customers that expresses an evaluation over the restaurant, i.e.,  $V = \{c \in C \mid eval(c) \neq *\}$ . We refer to this set as the set of *voters* and we assume it to be always non-empty, i.e., there is at least one customer that expresses an evaluation.

### 2.2 Two rating systems

In online rating systems such as Tripadvisor® every interested customer can see - and is therefore influenced by - (the average of) what the other customers have written. We call this method  $\mathbb{O}$ -RATING, which stands for *objective rating*.

Given an evaluation function  $eval$  of a restaurant, the associated  $\mathbb{O}$ -RATING is defined as follows:

$$\mathbb{O}\text{-RATING}(eval) = \text{avg}_{c \in V} eval(c)$$

Where  $\text{avg}$  is the average function across real-valued  $eval(c)$ , disregarding  $*$ . We omit  $eval$  when clear from the context.

$\mathbb{O}$ -RATING flattens individual evaluations into a unique objective aggregate, the rating that a certain restaurant is given. What we propose is a refinement of  $\mathbb{O}$ -RATING, which takes the network of influence into account. In this system customers are *only* interested in the evaluation of other customers they can trust, e.g., their friends. We call our method  $\mathbb{P}$ -RATING, which stands for *personalised rating*. It is defined for a pair customer-evaluation  $(c, eval)$  as follows:

$$\mathbb{P}\text{-RATING}(c, eval) = \text{avg}_{k \in N(c) \cap V} eval(k)$$

So the  $\mathbb{P}$ -RATING( $c, eval$ ) calculates what customer  $c$  comes to think of the restaurant, taking the average of the opinions of the customers  $c$  is connected to. Again we omit  $eval$  whenever clear from the context.

Observe that in case a customer has no connection with a voter, then  $\mathbb{P}$ -RATING is not defined. To facilitate the analysis we make the technical assumption that *each customer is connected to at least one voter*. Also observe that when  $E = C \times C$ , i.e., in case the network is complete and each individual is influenced by each other individual, then for all  $c \in C$  and  $eval$  we have that  $\mathbb{P}\text{-RATING}(c, eval) = \mathbb{O}\text{-RATING}(eval)$ .

### 2.3 Utilities and strategies

We interpret a customer evaluation as a measure of his or her *propensity* to go to the restaurant. We therefore assume that the utility that a restaurant gets is proportional to its rating. To simplify the analysis we assume a factor 1 proportionality.

**The case of  $\mathbb{O}$ -RATING.** For the  $\mathbb{O}$ -RATING, we assume that the initial utility  $u^0$  of the restaurant is defined as:

$$u_{\mathbb{O}}^0 = |C| \mathbb{O}\text{-RATING}(eval).$$

Intuitively, the initial utility amounts to the number of customers that actually go to the restaurant, weighted with their (average) predisposition.

At the initial stage of the game, the restaurant owner receives  $u^0$ , and can then decide to invest a part of it to influence a subset of customers and improve upon the initial gain. We assume utility to be fully transferrable and, to facilitate the analysis, that such transfers translate directly into changes of customers' predispositions.

**Definition 1.** A strategy is a function  $\sigma : C \rightarrow Val$  such that  $\sum_{c \in C} \sigma(c) \leq u^0$ .

Definition 1 imposes that strategies are *budget balanced*, i.e., restaurants can only pay with resources they have.

Let  $\Sigma$  be the set of all strategies. We denote  $\sigma^0$  the strategy that assigns 0 to all customers and we call *bribing strategy* any strategy that is different from  $\sigma^0$ . After the execution of a bribing strategy, the evaluation is updated as follows:

**Definition 2.** *The evaluation  $eval^\sigma(c)$  after execution of  $\sigma$  is  $eval^\sigma(c) = \min\{1, eval(c) + \sigma(c)\}$ , where  $* + \sigma(c) = \sigma(c)$ , if  $\sigma(c) \neq 0$ , and  $* + \sigma(c) = *$ , if  $\sigma(c) = 0$ .*

In this definition we are making the assumption that the effect of bribing a non-voter to vote is equivalent to that of bribing a voter that had a 0-level review, as, intuitively, the individual has no associated predisposition to go to the restaurant.

A strategy is called *efficient* if  $\sigma(c) + eval(c) \leq 1$  for all  $c \in C$ . Let  $B(\sigma) = \{c \in C \mid \sigma(c) \neq 0\}$  be the set of bribed customers. Let  $V^\sigma$  be the set of voters after the execution of  $\sigma$ . Executing  $\sigma$  induces the following change in utility:

$$u_{\mathbb{O}}^\sigma = |C| \mathbb{O}\text{-RATING}(eval^\sigma) - \sum_{c \in C} \sigma(c).$$

Intuitively,  $u_{\mathbb{O}}^\sigma$  is obtained by adding to the initial utility of the restaurant the rating obtained as an effect of the money invested on each individual minus the amount of money spent.

We define the revenue of a strategy  $\sigma$  as the marginal utility obtained by executing it:

**Definition 3.** *Let  $\sigma$  be a strategy. The revenue of  $\sigma$  is defined as  $\mathbf{r}_{\mathbb{O}}(\sigma) = u_{\mathbb{O}}^\sigma - u_{\mathbb{O}}^0$ . We say that  $\sigma$  is profitable if  $\mathbf{r}_{\mathbb{O}}(\sigma) > 0$ .*

Finally, we recall the standard notion of dominance:

**Definition 4.** *A strategy  $\sigma$  is weakly dominant if  $u_{\mathbb{O}}^\sigma \geq u_{\mathbb{O}}^{\sigma'}$  for all  $\sigma' \in \Sigma$ . It is strictly dominant if  $u_{\mathbb{O}}^\sigma > u_{\mathbb{O}}^{\sigma'}$  for all  $\sigma' \in \Sigma$ .*

Hence a non-profitable strategy is never strictly dominant.

**The case of  $\mathbb{P}$ -RATING.** The previous definitions can be adapted to the case of  $\mathbb{P}$ -RATING as follows:

$$u_{\mathbb{P}}^0 = \sum_{c \in C} \mathbb{P}\text{-RATING}(c, eval)$$

which encodes the initial utility of each restaurant, and

$$u_{\mathbb{P}}^\sigma = \sum_{c \in C} \mathbb{P}\text{-RATING}(c, eval^\sigma) - \sum_{c \in C} \sigma(c)$$

which encodes the utility change after the execution of a  $\sigma$ . Finally, let the revenue of  $\sigma$  be  $\mathbf{r}_{\mathbb{P}}(\sigma) = u_{\mathbb{P}}^\sigma - u_{\mathbb{P}}^0$ . If clear from the context, we use  $\mathbb{P}\text{-RATING}^\sigma(c)$  for  $\mathbb{P}\text{-RATING}(eval^\sigma, c)$ .

In order to determine the dominant strategies, we need to establish how the customers vote, how they are connected, and what the restaurant owner knows. In this paper we assume that the restaurant knows *eval*, leaving the interesting case when *eval* is unknown to future work. We focus instead on the following cases: the restaurant knows the network, the restaurant knows the shape of the network but not the individuals' position, and the network is unknown. We analyse the effect of bribing strategies on  $\mathbb{P}$ -RATING in each such case. Notice how for the case of  $\mathbb{O}$ -RATING the cases collapse to the first. We also look at the special situation in which every customer is a voter.

Given a set of such assumptions, we say that  $\mathbb{O}$ -RATING (or  $\mathbb{P}$ -RATING) are *bribery-proof* under those assumptions if  $\sigma^0$  is weakly dominant.

**Discussion** Our model is built upon a number of simplifying assumptions which do not play a significant role in the results and could therefore be dispensed with: (i) customers' ratings correspond to their propensity to go to the restaurant. (ii) the restaurant utility equals the sum of all such propensities (iii) bribe  $\sigma(c)$  affects evaluation  $eval(c)$  linearly. All these assumptions could be generalised by multiplicative factors, such as an average price  $R$  paid at the restaurant, and a "customer price"  $D_c$ , such that  $eval^\sigma(c) = eval(c) + \frac{\sigma(c)}{D_c}$ .

### 3 Bribes under $\mathbb{O}$ -RATING

In this section we look at bribing strategies under  $\mathbb{O}$ -RATING, first focussing on the case where everyone expresses an opinion, then moving on to the more general case.

#### 3.1 All vote

Let us now consider the case in which  $V = C$ . Recall that  $B(\sigma)$  is the set of customers bribed by  $\sigma$ . We say that two strategies  $\sigma_1$  and  $\sigma_2$  are *disjoint* if  $B(\sigma_1) \cup B(\sigma_2) = \emptyset$ . By direct calculation it follows that the revenue of disjoint strategies exhibits the following property:

**Lemma 1.** *If  $V = C$  and  $\sigma_1$  and  $\sigma_2$  are two disjoint strategies, then  $\mathbf{r}_{\mathbb{O}}(\sigma_1 \circ \sigma_2) = \mathbf{r}_{\mathbb{O}}(\sigma_1) + \mathbf{r}_{\mathbb{O}}(\sigma_2)$ .*

We now show that bribing a single individual is not profitable.

**Lemma 2.** *Let  $\sigma$  be a bribing strategy,  $V = C$  and  $|B(\sigma)| = 1$ . Then,  $\mathbf{r}_{\mathbb{O}}(\sigma) \leq 0$ , i.e.,  $\sigma$  is not profitable.*

*Proof sketch.* Let  $\bar{c}$  be the only individual such that  $\sigma(\bar{c}) \neq 0$ . By calculation,  $\mathbf{r}(\sigma) = u_{\mathbb{O}}^\sigma - u_{\mathbb{O}}^0 = \mathbb{O}\text{-RATING}^\sigma - \mathbb{O}\text{-RATING} - \sum_{c \in C} \sigma(c) = \min\{1, eval(\bar{c}) + \sigma(\bar{c})\} - eval(\bar{c}) - \sigma(\bar{c}) \leq 0$ .  $\square$

By combining the two lemmas above we are able to show that no strategy is profitable for bribing the  $\mathbb{O}$ -RATING.

**Proposition 3.** *If  $V = C$ , then no strategy is profitable.*

*Proof sketch.* Any bribing strategy  $\sigma$  can be decomposed into  $n$  pairwise disjoint strategies such that  $\sigma = \sigma_{c_1} \circ \dots \circ \sigma_{c_n}$  and  $|B(\sigma_{c_j})| = 1$  for all  $1 \leq j \leq n$ . By applying Lemma 1 and Lemma 2 we then obtain that  $\mathbf{r}_{\mathbb{O}}(\sigma) \leq 0$ .  $\square$

From this it follows that  $\sigma^0$  is weakly dominant and thus  $\mathbb{O}$ -RATING bribery-proof when all customers voted.

#### 3.2 Non-voters

Let us now consider the case of  $V \subset C$ , i.e., when there is at least one customer who is not a voter. In this case Lemma 1 no longer holds, as shown in the following example.

**Example 1.** *Let  $C = \{A, B, C\}$ , and let  $eval(A) = 0.5$ ,  $eval(B) = 0.5$ , and  $eval(C) = *$ . The initial resources are  $u^0 = \mathbb{O}\text{-RATING} \times 3 = 1.5$ . Let now  $\sigma_1(A) = 0.5$  and  $\sigma_1(B) = \sigma_1(C) = 0$ , and let  $\sigma_2(C) = 0.5$  and  $\sigma_2(A) = \sigma_2(B) = 0$ . Now  $u_{\mathbb{O}}^{\sigma_1} = 0.75 \times 3 - 0.5 = 1.75$  and  $u_{\mathbb{O}}^{\sigma_2} = 0.5 \times 3 - 0.5 = 1$ , but  $u_{\mathbb{O}}^{\sigma_1 \circ \sigma_2} = 0.6 \times 3 - 1 = 1$ .*

The example (in particular  $\sigma_1$ ) also shows that  $\mathbb{O}$ -RATING in this case is not bribery-proof.

We now turn to characterise the set of undominated bribing strategies. We begin by showing that bribing a non-voter is

always dominated. Let first  $\sigma$  be a strategy such that  $\sigma(\bar{c}) \neq 0$  for some  $\bar{c} \in C \setminus V$  and recall that  $V^\sigma$  is the set of voters after execution of  $\sigma$ . Let us define the  $\bar{c}$ -greedy restriction of  $\sigma$  to be any strategy  $\sigma^{-\bar{c}}$  such that:

- $V^{\sigma^{-\bar{c}}} = V^\sigma \setminus \{\bar{c}\}$ , i.e., the greedy restriction eliminates  $\bar{c}$  from the set of voters.
- For each  $c \in V^\sigma \setminus \bar{c}$ ,  $\max(1, eval(c) + \sigma(c)) = \max(1, eval(c) + \sigma^{-\bar{c}}(c))$ , i.e., the greedy restriction does not waste further resources.
- If there exists  $c \in V^\sigma \setminus \bar{c}$  such that  $eval(c) + \sigma^{-\bar{c}}(c) < 1$  then  $\sum_{c \in C} \sigma^{-\bar{c}}(c) = \sum_{c \in C} \sigma(c)$ , i.e., the  $\sigma^{-\bar{c}}$  redistributes  $\sigma(\bar{c})$  among the remaining voters.

We now show that each strategy bribing a non-voter is strictly dominated by any of its greedy restrictions.

**Proposition 4.** *Let  $V \neq C$ , and  $\bar{c} \in C \setminus V$ . Then each strategy  $\sigma$  with  $\sigma(\bar{c}) \neq 0$  is strictly dominated by  $\sigma^{-\bar{c}}$ .*

*Proof.* Let  $\sigma$  be a strategy with  $\sigma(\bar{c}) \neq 0$  for some non-voter  $\bar{c}$ , and let  $\sigma^{-\bar{c}}$  be one of its greedy restriction defined above.

$$\begin{aligned} & u_{\mathbb{O}}^{\sigma^{-\bar{c}}} - u_{\mathbb{O}}^{\sigma} = \\ & |C|(|\mathbb{O}\text{-RATING}^{\sigma^{-\bar{c}}} - \mathbb{O}\text{-RATING}^{\sigma}|) + \sum_{c \in C} \sigma(c) - \sum_{c \in C} \sigma^{-\bar{c}}(c) = \\ & |C| \left( \frac{\sum_{c \in C} eval^{\sigma^{-\bar{c}}}(c)}{|V|} - \frac{\sum_{c \in C} eval^{\sigma}(c)}{|V \cup \bar{c}|} \right) + \\ & \quad + \left( \sum_{c \in C} \sigma(c) - \sum_{c \in C} \sigma^{-\bar{c}}(c) \right) \end{aligned}$$

Observe first that  $\sigma^{-\bar{c}}$  is a redistribution, hence  $\sum_c \sigma(c) - \sum_c \sigma^{-\bar{c}}(c) \geq 0$ , i.e., the second addendum in the above equation is positive. Consider now the case where there exists  $c \in V^\sigma \setminus \bar{c}$  such that  $eval(c) + \sigma^{-\bar{c}}(c) < 1$ . Then by the definition of  $\sigma^{-\bar{c}}$  we have that  $\sum_{c \in V^\sigma} eval^{\sigma}(c) = \sum_{c \in V^{\sigma^{-\bar{c}}}} eval^{\sigma^{-\bar{c}}}(c)$ , i.e., the greedy restriction preserves the overall evaluation. By straightforward calculation this entails that  $u_{\mathbb{O}}^{\sigma^{-\bar{c}}} - u_{\mathbb{O}}^{\sigma} > 0$ . If no such  $c$  exists, and therefore  $\mathbb{O}\text{-RATING}^{\sigma^{-\bar{c}}} = 1$  we have that either  $\mathbb{O}\text{-RATING}^{\sigma} < 1$  or, by the efficiency requirement and the fact that  $\sigma(\bar{c}) \neq 0$ , we have that  $\sum_{c \in C} \sigma(c) > \sum_{c \in C} \sigma^{-\bar{c}}(c)$ . In either cases we have that  $u_{\mathbb{O}}^{\sigma^{-\bar{c}}} - u_{\mathbb{O}}^{\sigma} > 0$ .  $\square$

Let an  $\mathbb{O}$ -greedy strategy be any efficient strategy that redistributes all the initial resources  $u_{\mathbb{O}}^0$  among voters. Making use of the previous result, we are able to characterise the set of all dominant strategies for  $\mathbb{O}$ -RATING.

**Proposition 5.** *Let  $V \neq C$ . A strategy is weakly dominant for  $\mathbb{O}$ -RATING if and only if it is an  $\mathbb{O}$ -greedy strategy.*

*Proof sketch.* For the right-to-left direction, first observe that all  $\mathbb{O}$ -greedy strategies are payoff-equivalent, and that a non-efficient strategy is always dominated by its efficient counterpart. By Proposition 4 we know that strategies bribing non-voters are dominated, and by straightforward calculations we obtain that in presence of non-voters it is always profitable to bribe as much as possible. For the left-to-right direction,

observe that a non-greedy strategy is either inefficient, or it bribes a non-voter, or does not bribe as much as possible. In either circumstance it is strictly dominated.  $\square$

While there may be cases in which the number of weakly dominant strategies under  $\mathbb{O}$ -RATING is exponential, all such strategies are revenue equivalent, and Proposition 5 gives us a polynomial algorithm to find one of them: starting from an evaluation vector  $eval$ , distribute all available resources  $u_{\mathbb{O}}^0$  to the voters, without exceeding the maximal evaluation of 1. By either exhausting the available budget or distributing it all, we are guaranteed the maximum gain by Proposition 5.

## 4 Bribes under $\mathbb{P}$ -RATING

In this section we look at bribing strategies under  $\mathbb{P}$ -RATING, against various knowledge conditions on the social network. As for Section 3 we start by looking at the case where everyone votes and later on allowing non-voters. Before doing that, we introduce a useful graph-theoretic measure of influence.

**Definition 5.** *The influence weight of a customer  $c \in C$  in a network  $E$  and under a strategy  $\sigma$  is defined as follows:*

$$w_c = \sum_{k \in N(c)} \frac{1}{|N(k) \cap V^\sigma|}$$

Intuitively, each individual's rating influences the rating of each of its connections, with a factor that is inversely proportional to the number of second-level connections. We formalise this statement in the following lemma, which is restricted to strategies that only bribe voters:

**Lemma 6.** *The utility obtained by playing  $\sigma$  with  $\mathbb{P}$ -RATING is  $u_{\mathbb{P}}^{\sigma} = \sum_{c \in V^\sigma} w_c \times eval^{\sigma}(c) - \sum_{c \in C} \sigma(c)$ .*

*Proof.* By calculation:

$$\begin{aligned} u_{\mathbb{P}}^{\sigma} + \sum_{c \in C} \sigma(c) &= \sum_{c \in C} \mathbb{P}\text{-RATING}(c) = \sum_{c \in C} \text{avg}_{k \in N(c) \cap V^\sigma} eval^{\sigma}(k) = \\ &= \sum_{c \in C} \left[ \frac{1}{|N(c) \cap V^\sigma|} \sum_{k \in N(c) \cap V^\sigma} eval^{\sigma}(k) \right] = \\ &= \sum_{k \in V^\sigma} \left[ eval^{\sigma}(k) \times \sum_{k' \in N(k)} \frac{1}{|N(k') \cap V^\sigma|} \right] = \\ &= \sum_{c \in V^\sigma} w_c \times eval^{\sigma}(c) \quad \square \end{aligned}$$

In particular, when  $V = C$ , we obtain  $w_c = \sum_{k \in N(c)} \frac{1}{deg(k)}$ , where  $deg(c) = |N(c)|$  is the degree of  $c$  in  $E$ .

### 4.1 All vote, known network

We begin by studying the simplest case in which the restaurant knows the evaluation  $eval$ , the network  $E$  as well as the position of each customer on the network. The following corollary is a straightforward consequence of Lemma 6:

**Corollary 7.** *Let  $V = C$  and let  $\sigma_1$  and  $\sigma_2$  be two disjoint strategies, then  $\mathbf{r}_{\mathbb{P}}(\sigma_1 \circ \sigma_2) = \mathbf{r}_{\mathbb{P}}(\sigma_1) + \mathbf{r}_{\mathbb{P}}(\sigma_2)$ .*

We are now able to show a precise characterisation of the revenue obtained by any efficient strategy  $\sigma$ :

**Proposition 8.** Let  $V = C$ , let  $E$  be a known network, and let  $\sigma$  be an efficient strategy. Then  $\mathbf{r}_{\mathbb{P}}(\sigma) = \sum_{c \in C} (w_c - 1)\sigma(c)$ .

*Proof.* By calculation, where Step (2) uses Lemma 6, and Step (4) uses the fact that  $\sigma$  is efficient:

$$\begin{aligned} \mathbf{r}_{\mathbb{P}}(\sigma) &= u_{\mathbb{P}}^{\sigma} - u_{\mathbb{P}}^0 = & (1) \\ &= \left[ \sum_{c \in C} w_c \text{eval}_c^{\sigma} - \sum_{c \in C} \sigma(c) - \sum_{c \in C} w_c \text{eval}(c) \right] = & (2) \\ &= \sum_{c \in C} [w_c [\min\{1, \text{eval}(c) + \sigma(c)\} - \text{eval}(c)]] - & (3) \\ &\quad - \sum_{c \in C} \sigma(c) = \sum_{c \in C} (w_c - 1)\sigma(c). & (4) \end{aligned}$$

□

Proposition 8 tells us that the factors  $w_c$  are crucial in determining the revenue of a given bribing strategy. Bribing a customer  $c$  is profitable whenever  $w_c > 1$  (provided its evaluation was not 1 already), while bribing a customer  $c$  with  $w_c \leq 1$  is at most as profitable as doing nothing, as can be seen in the example below. Most importantly, it shows that  $\mathbb{P}$ -RATING is not bribery-proof when the restaurant knows both the network and the customers' evaluations.

**Example 2.** Let  $E$  be a four arms stars, and let  $A$  be the individual in the centre. Assume each individual values the restaurant 0.5. We have that  $w_A = 2.2$  and  $w_c = 0.7$  for all  $c$  different from  $A$ . Consider now two bribing strategies:  $\sigma^A$  which bribes  $A$  with 0.5, and  $\sigma^B$  which bribes a single individual  $B \neq A$  with the same amount. What we obtain is that  $\mathbf{r}_{\mathbb{P}}(\sigma^A) = 0.6$ , while  $\mathbf{r}_{\mathbb{P}}(\sigma^B) = -0.15$ .

Given a network  $E$  and an evaluation vector  $\text{eval}$ , let Algorithm 1 define the  $\mathbb{P}$ -greedy bribing strategy.

**Input:** Evaluation function  $\text{eval}$  and network  $E$

**Output:** A bribing strategy  $\sigma_{\mathbb{P}}^G : C \rightarrow \text{Val}$

Budget =  $u_{\mathbb{P}}^0$

$\sigma_{\mathbb{P}}^G(c) = 0$  for all  $c \in C$

Compute  $w_c$  for all  $c \in C$

Sort  $c \in C$  in descending order  $c_0, \dots, c_m$  based on  $w_c$

for  $i=0, \dots, m$  do

    if Budget  $\neq 0$  then

        if  $w_{c_i} > 1$  then

$\sigma_{\mathbb{P}}^G(c_i) = \min\{1 - \text{eval}(c_i), \text{Budget}\}$

            Budget = Budget -  $\sigma_{\mathbb{P}}^G(c_i)$

        end

    end

    return  $\sigma_{\mathbb{P}}^G$

end

**Algorithm 1:** The  $\mathbb{P}$ -greedy bribing strategy  $\sigma_{\mathbb{P}}^G$

As a consequence of Proposition 8 we obtain:

**Corollary 9.** The  $\mathbb{P}$ -greedy bribing strategy defined in Algorithm 1 is weakly dominant.

As in the case of  $\mathbb{O}$ -RATING, Corollary 9 has repercussions on the computational complexity of bribery: it shows that

computing a weakly dominant strategy can be done in polynomial time. Notice how the most costly operation lies in the computation of the influence weights  $w_c$ , which can be performed only once, assuming the network is static. Similar problems, such as recognising whether bribing a certain individual is profitable, or estimating whether individuals on a network can be bribed above a certain threshold, are also computable in polynomial time.

## 4.2 All vote, unknown network

We now move to study the more complex case of an unknown network. Surprisingly, we are able to show that no bribing strategy is profitable (in expectation), and hence  $\mathbb{P}$ -RATING is bribery-proof in this case. Recall that we are still assuming that the restaurant knows  $\text{eval}$  and everybody voted.

We begin by assuming that the restaurant knows the structure of the network, but not the position of each participant. Formally, the restaurant knows  $E$ , but considers any permutation of the customers in  $C$  over  $E$  as possible. Let us thus define the expected revenue of a strategy  $\sigma$  over a given network  $E$  as the average over all possible permutations of customers:  $\mathbb{E}[\mathbf{r}_{\mathbb{P}}(\sigma)] = \sum \frac{1}{n!} [u_{\rho}^{\sigma} - u_{\rho}^0]$ , where we abuse notation by writing  $u_{\rho}^{\sigma}$  as  $u_{\mathbb{P}}^{\sigma}$  under permutation  $\rho$  over the network  $E$ . What we are able to show is that all strategies are at most as profitable as  $\sigma^0$  in expected return:

**Proposition 10.** Let  $V = C$ , let the network structure of  $E$  be known but not the relative positions of customers on  $E$ . Then  $\mathbb{E}[\mathbf{r}_{\mathbb{P}}(\sigma)] = 0$  for all strategies  $\sigma$ .

*Proof sketch.* Let  $|C| = n$ . We show the result for any strategy  $\sigma$  that bribes a single customer  $\bar{c}$ . The general statement follows from the linearity of  $\mathbb{E}[\mathbf{r}(\sigma)]$ . Equation (5) uses Proposition 8 to compute the revenue for each permutation  $\rho$  of customers  $C$  on the network:

$$\begin{aligned} \mathbb{E}[\sigma] &= \sum_{\rho} \frac{1}{n!} (u_{\rho}^{\sigma} - u_{\rho}^0) = \sum_{\rho} \frac{1}{n!} (w_{\rho(\bar{c})} - 1)\sigma(\bar{c}) = & (5) \\ &= \sum_{c \in C} \frac{(n-1)!}{n!} (w_c - 1)\sigma(\bar{c}) = \frac{(n-1)!}{n!} \sum_{c \in C} (w_c - 1) = 0 & (6) \end{aligned}$$

The last line follows from the observation that  $\sum_c w_c = |C|$  and hence  $\sum_c (w_c - 1) = 0$ , by a consequence of Definition 5 when everybody votes. □

Hence, if we assume a uniform probability over all permutations of customers on the network, a straightforward consequence of Proposition 10 concludes that it is not profitable (in expectation) to bribe customers.

**Corollary 11.** If  $V = C$  and the network is unknown, then no bribing strategy for  $\mathbb{P}$ -RATING is profitable in expected return.

## 4.3 Non-voters, known network

With  $\mathbb{P}$ -RATING it is possible to find a network where bribing a non-voter is profitable:

**Example 3.** Consider 4 individuals  $\{B, C, D, E\}$  connected only to a non-voter in the middle. Let  $\text{eval}(j) = 0.2$  for all  $j$  but the center. We have  $u_{\mathbb{P}}^0 = 1$ . Let  $A$  be the non-voter, and let  $\sigma_1(A) = 1$  and 0 otherwise. The utility of  $\sigma_1$  is:

$$\mathbb{P}\text{-RATING}^{\sigma_1}(A) + 4\mathbb{P}\text{-RATING}^{\sigma_1}(j) - 1 = 1.76$$

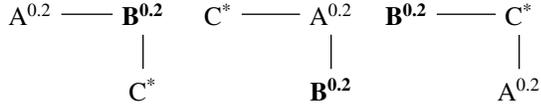


Figure 1: Customers permutations in Example 4.

All other strategies can be shown to be dominated by  $\sigma_1$ . Take for instance a strategy  $\sigma_2$  such that  $\sigma_2(B) = 0.8$ ,  $\sigma_2(C) = 0.2$  and 0 otherwise. The utility of  $\sigma_2$  is  $u_{\mathbb{P}}^{\sigma_2} = 1.25$ .

It is quite hard to obtain analytical results for strategies bribing non-voters, due to the non-linearity of the  $\mathbb{P}$ -RATING in this setting. We can however provide results in line with those of the previous section if we restrict to *voter-only strategies*, i.e., strategies  $\sigma$  such that  $\sigma(c) = 0$  for all  $c \notin V$ . In this case, a similar proof to Proposition 8 shows the following:

**Proposition 12.** *Let  $V \neq C$ ,  $E$  be a known network, and  $\sigma$  be an efficient bribing strategy such that  $B(\sigma) \subseteq V$ . Then,  $r_{\mathbb{P}}(\sigma) = \sum_{c \in C} (w_c - 1)\sigma(c)$ .*

The difference with the case of  $V = C$  is that  $w_c$  can be arbitrarily large in the presence of non-voters, such as in our Example 3.

#### 4.4 Non-voters, unknown positions

Unlike the case of  $V = C$ , in this case it is possible to define bribing strategies that are profitable (in expected return).

**Example 4.** *Let  $C = \{A, B, C\}$ , and the initial evaluation  $eval(A) = eval(B) = 0.2$  and  $eval(C) = *$ . Assume that the structure of the network is known, but the position of the individuals is not. Let the three possible network positions (without counting the symmetries) be depicted in Figure 1. Let  $\sigma(B) = 0.2$  and  $\sigma(A) = \sigma(C) = 0$ . In the first case:*

$$\begin{aligned} r_{\mathbb{P}}^1(\sigma) &= \mathbb{P}\text{-RATING}(A) + \dots + \mathbb{P}\text{-RATING}(C) - 0.2 - u_{\mathbb{P}}^0 = \\ &= 0.3 + 0.3 + 0.4 - 0.2 - 0.6 = 0.2 \end{aligned}$$

In the second case  $r_{\mathbb{P}}^2(\sigma) = 0$  while in the third:

$$r_{\mathbb{P}}^3(\sigma) = 0.4 + 0.3 + 0.2 - 0.2 - 0.6 = 0.1$$

Therefore,  $\mathbb{P}$ -RATING is not bribery-proof (in expectation) in the presence of non-voters when the network is unknown. Interesting computational problems open up in this setting, such as identifying the networks that allow for profitable bribing strategies, and their expected revenue.

## 5 Boundaries of bribery-proofness

The previous sections have shown that having a network-based rating systems, where individuals are influenced by their peers, is not bribery-proof, even when the position of individuals in a given network is not known. However bribing strategies have a different effect in the overall score. While the utility of  $\mathbb{O}$ -RATING is a sum of the *global* average of voters' evaluation, the utility of  $\mathbb{P}$ -RATING is a sum of *local* averages of voters' evaluation against the one of their peers.

Therefore a strategy bribing one voter affects everyone in the case of  $\mathbb{O}$ -RATING, but it can be shown to have a limited effect in the case of  $\mathbb{P}$ -RATING.

**Proposition 13.** *Let  $\sigma$  be an efficient strategy s.t.  $|B(\sigma)| = 1$ , and let  $\bar{c}$  be such that  $\sigma(\bar{c}) \neq 0$ . Then  $r_{\mathbb{P}}(\sigma) < N(\bar{c})$ .*

*Proof.* By calculation, we have that:

$$\begin{aligned} r_{\mathbb{P}}(\sigma) &= \sum_{c \in C} \mathbb{P}\text{-RATING}^{\sigma}(c) - \sigma(\bar{c}) - \sum_{c \in C} \mathbb{P}\text{-RATING}(c) = \\ &= \sum_{c' \in N(\bar{c})} \mathbb{P}\text{-RATING}^{\sigma}(c) - \sigma(\bar{c}) - \sum_{c' \in N(\bar{c})} \mathbb{P}\text{-RATING}(c) \leq \\ &\leq 1 \times N(\bar{c}) - \sigma(\bar{c}) - \sum_{c' \in N(\bar{c})} \mathbb{P}\text{-RATING}(c) < N(\bar{c}) \quad \square \end{aligned}$$

The previous result shows that increasing the number of individuals that are not connected to an agent that is bribed, even if these are non-voters, does not increase the revenue of the bribing strategy. This is not true when we use  $\mathbb{O}$ -RATING.

**Proposition 14.** *Let  $\sigma$  be an efficient strategy. The revenue  $r_{\mathbb{O}}(\sigma)$  of  $\sigma$  is monotonically increasing with the number of non-voters, and is unbounded.*

*Proof.* It follows from our definitions that:

$$r_{\mathbb{O}}(\sigma) = \left( \frac{|C|}{|V^{\sigma}|} - 1 \right) \left[ \sum_{c \in C} eval(c) + \sigma(c) \right]$$

The above figure is unbounded and monotonically increasing in the number of non-voters, which can be obtained by increasing  $C$  keeping  $V^{\sigma}$  fixed.  $\square$

So while  $\mathbb{P}$ -RATING and  $\mathbb{O}$ -RATING are not bribery-proof in general, it turns out that the impact of the two in the overall network are significantly different. In particular, under realistic assumptions such as a very large proportion of non-voters and with participants having a few connections, bribing under  $\mathbb{O}$ -RATING is increasingly rewarding, while under  $\mathbb{P}$ -RATING this is no longer the case.

## 6 Conclusive remarks

We introduced  $\mathbb{P}$ -RATING, a network-based rating system which generalises the commonly used  $\mathbb{O}$ -RATING, and analysed their resistance to external bribery under various conditions. The main take-home message of our contribution can be summarised in one point, deriving from our main results:

$\mathbb{P}$ -RATING and  $\mathbb{O}$ -RATING are not bribery-proof in general. However, if we assume that a service provider has a cost for bribing an individual, there are situations in which  $\mathbb{P}$ -RATING is fully bribery proof, while  $\mathbb{O}$ -RATING is not. For instance, if the cost of bribing an individual  $c$  is at least  $N(c)$  then  $\mathbb{P}$ -RATING is bribery-proof. As observed previously, this is not necessarily true for  $\mathbb{O}$ -RATING. In particular, if we assume the presence of unreachable individuals the difference is more significant. As shown, for  $\mathbb{P}$ -RATING we need to bribe individuals with  $w_c > 1$ . With  $\mathbb{O}$ -RATING is sufficient to find one voter who accepts a bribe.

There is a number of avenues open to future research investigation. The most important ones include the case of partially known customers' evaluation, and the study of ratings of multiple restaurants, where the probability of a customer choosing a restaurant determines his or her probability not to choose the others.

## Acknowledgments

Umberto Grandi acknowledges the support of the Labex CIMI project “Social Choice on Networks” (ANR-11-LABX-0040-CIMI). Paolo Turrini the support of Imperial College London for the Junior Research Fellowship Designing negotiation spaces for collective decision-making (DoC-AI1048). This work has also been partly supported by COST Action IC1205 on Computational Social Choice.

## References

- [Abraham *et al.*, 2006] I. Abraham, D. Dolev, R. Gonen, and J. Halpern. Distributed computing meets game theory: robust mechanisms for rational secret sharing and multiparty computation. In *Proceedings of the 25th ACM symposium on Principles of distributed computing (PODC’06)*, 2006.
- [Apt and Markakis, 2014] Krzysztof R. Apt and Evangelos Markakis. Social networks with competing products. *Fundamenta Informaticae*, 129(3):225–250, 2014.
- [Baumeister *et al.*, 2011] Dorothea Baumeister, Gábor Erdélyi, and Jörg Rothe. How hard is it to bribe the judges? A study of the complexity of bribery in judgment aggregation. In *Proceedings of the Second International Conference on Algorithmic Decision Theory (ADT-2011)*, 2011.
- [Brandt *et al.*, 2015] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2015.
- [Bredereck *et al.*, 2014] R. Bredereck, J. Chen, S. Hartung, S. Kratsch, R. Niedermeier, O. Suchý, and G. J. Woeginger. A multivariate complexity analysis of lobbying in multiple referenda. *Journal of Artificial Intelligence Research*, 50:409–446, 2014.
- [Bredereck *et al.*, 2016] Robert Bredereck, Piotr Faliszewski, Rolf Niedermeier, and Nimrod Talmon. Large-scale election campaigns: Combinatorial shift bribery. *Journal of Artificial Intelligence Research (JAIR)*, 55:603–652, 2016.
- [Christian *et al.*, 2007] R. Christian, M. Fellows, F. Rosamond, and A. Slinko. On complexity of lobbying in multiple referenda. *Review of Economic Design*, 11(3):217–224, 2007.
- [Conitzer, 2012] Vincent Conitzer. Should social network structure be taken into account in elections? *Mathematical Social Sciences*, 64(1):100–102, 2012.
- [Conte and Paolucci, 2002] Rosaria Conte and Mario Paolucci. *Reputation in Artificial Societies: Social Beliefs for Social Order*. Kluwer Academic Publishers, 2002.
- [Conte *et al.*, 2008] Rosaria Conte, Mario Paolucci, and Jordi Sabater-Mir. Reputation for innovating social networks. *Advances in Complex Systems*, 11(2):303–320, 2008.
- [Elkind, 2014] Edith Elkind. Coalitional games on sparse social networks. In *10th International Conference on Web and Internet Economics (WINE-2014)*, 2014.
- [Faliszewski *et al.*, 2009] Piotr Faliszewski, Edith Hemaspaandra, and Lane A. Hemaspaandra. How hard is bribery in elections? *Journal of Artificial Intelligence Research (JAIR)*, 35:485–532, 2009.
- [Garcin *et al.*, 2009] F. Garcin, B. Faltings, and R. Jurca. Aggregating reputation feedback. In *Proceedings of the International Conference on Reputation*, 2009.
- [Helpman and Persson, 1998] Elhanan Helpman and Torsten Persson. Lobbying and legislative bargaining. Working Paper 6589, National Bureau of Economic Research, June 1998.
- [Kurokawa *et al.*, 2015] David Kurokawa, Omer Lev, Jamie Morgenstern, and Ariel D. Procaccia. Impartial peer review. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-2015)*, pages 582–588, 2015.
- [Pinyol and Sabater-Mir, 2013] Isaac Pinyol and Jordi Sabater-Mir. Computational trust and reputation models for open multi-agent systems: a review. *Artificial Intelligence Review*, 40(1):1–25, 2013.
- [Procaccia *et al.*, 2015] Ariel D. Procaccia, Nisarg Shah, and Eric Sodomka. Ranked voting on social networks. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-2015)*, 2015.
- [Sabater and Sierra, 2005] Jordi Sabater and Carles Sierra. Review on computational trust and reputation models. *Artificial Intelligence Review*, 24(1):33–60, 2005.
- [Salehi-Abari and Boutilier, 2014] Amirali Salehi-Abari and Craig Boutilier. Empathetic social choice on social networks. In *Proceedings of the 13th International conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2014)*, 2014.
- [Simon and Apt, 2015] Sunil Simon and Krzysztof R. Apt. Social network games. *Journal of Logic and Computation*, 25(1):207–242, 2015.
- [Tsang *et al.*, 2015] Alan Tsang, John A. Doucette, and Hadi Hosseini. Voting with social influence: Using arguments to uncover ground truth. In *Proceedings of the 15th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2015)*, 2015.