Manipulation of explicit reputation in innovation and knowledge exchange communities: The example of referencing in science

Michael A. Zaggl

TUM School of Management, Technische Universität München, Arcisstr. 21, 80333 Munich, Germany

ARTICLE INFO

ABSTRACT

This paper investigates the manipulation of reputation in the context of innovation and knowledge exchange communities. Reputation is crucial for overcoming the free-riding problem and enables community members to be rewarded because their contributions to the common good can be measured. However, the concept of reputation can include the notion of manipulation, which we define as the attempt to change one's reputation without contributing to the community. To investigate the topic of reputation manipulation, we build on the concept of reputation-based reward systems and extend it by distinguishing between *implicit reputation*, which is uncodified, and *explicit reputation*, which is codified and centrally counted. We argue that the possibilities for manipulation differ between these two distinctions. We investigate reputation manipulation empirically in the context of science, which is built on an explicit reputation-based reward system, and we use the received citations as an indicator for reputation. We distinguish two forms of manipulation—unjustified self-citing and unjustified reciprocal citing—and find evidence of both within a bibliometric dataset. This paper contributes to the design of knowledge exchange communities by highlighting the opportunities and challenges arising from explicit reputation-based reward systems, specifically the opportunities for manipulation. It also contributes to the work on misconduct in science.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Reputation plays a crucial role in innovation and knowledge exchange communities because it can account for the contributions of the communities' members. Reputation has been researched in contexts other than communities, such as organizations and their stakeholders (Lange et al., 2011) and online markets (Bolton et al., 2013; Lanzolla and Frankort, 2016; Moreno and Terwiesch, 2014; Yoganarasimhan, 2013), where a positive reputation increases the likelihood of mutually beneficial transactions. In the context of organizations and their stakeholders, reputation enables the transfer of decision-relevant information to recipients such as potential employees, customers, and suppliers. Similarly, reputation helps to overcome the problem of information asymmetries between buyers and sellers in markets. However, the role of reputation differs in the context of communities of innovation and knowledge exchange such as open source software development (Henkel, 2006; Osterloh and Rota, 2007; von Krogh et al., 2012) and science (i.e., basic research) (Merton, 1988; Stephan, 1996; Stephan, 2010). Here, reputation is based on community members' contributions. Consequently, reputation represents a kind of deposit that enables selective incentives and directs benefits exclusively to the contributors (Milinski et al., 2002; Oliver, 2013; Olson, 1965; von Hippel and

von Krogh, 2003). Reputation enables others to reciprocate or the contributor to signal hidden qualities in the expectation of future benefits (Lerner and Tirole, 2002).

The value represented by reputation inevitably incites manipulation attempts (see Charness et al., 2014). We define manipulation as the attempt to change one's reputation without actually contributing to the community. Manipulation is distinct from free riding. Both avoid the cost of contributing, but while free riders are excluded from selective benefits (i.e., benefits exclusive to the contributors, see Olson, 1965), manipulators profit from selective benefits because they pretend they have contributed. The phenomenon of reputation manipulation is almost as old as that of reputation itself. Now, online communities are dictating new conditions, including the possibility to codify reputation information and translate it into an *explicit* representation. This explicit representation of reputation changes the possibilities for manipulation by undermining traditional protection measures such as social punishment for self-praise.

This paper addresses the topic of reputation manipulation in innovation and knowledge exchange communities. It enriches our understanding of such communities (Faraj and Johnson, 2011; Faraj et al., 2011; West and Lakhani, 2008) and the role of explicit reputation. Currently, there is a lack of understanding of the possibilities

in communities for the manipulation of reputation, resulting in the misuse of selective incentives. Manipulation is a relevant phenomenon as shown, for example, by Sojer et al. (2014) in a recent study on unethical code reuse in open source software development and by Hutter et al. (2015) in innovation contests. Based on the extant literature, we define knowledge exchange and innovation communities as members interacting voluntarily with no common affiliation but with similar goals such as innovation or knowledge accumulation (West and Lakhani, 2008). Community members have a shared language and common rules, and they rely to a large degree on self-governance (Rullani and Haefliger, 2013), and they have the inherited characteristics of collective action systems (see Allen, 1983; Olson, 1965; Ostrom, 2000, 2007).

We chose the scientific system to examine the manipulation of explicit reputation. This system underlies the conditions of collective action and references or citations are a meaningful representation of reputation (Dasgupta and David, 1994; Stephan, 1996). We extract data in the form of a citation network that consists of several thousand publications from the Web of Science. Building on the free-rider hypothesis (Olson, 1965; Hardin, 1968), we expect manipulation to be a significant phenomenon and publications involved in this manipulation to be of inferior quality. Based on these expectations, we develop hypotheses and apply a deductive research design. We find significant evidence for author self-citing and author reciprocal citing in the scientific system. Self-citing is not a new phenomenon (Fowler and Aksnes, 2007; Glänzel et al., 2004; Hyland, 2003), but we extend the concept to include manipulative self-citing. Despite strong interest of the conceptual literature (Phelan, 1999; Posner, 2000), reciprocal citing has rarely been studied empirically at the individual level (i.e., author or publication level), the exceptions being studies using small samples (Paisley, 1990; White et al., 2004). Since self-citation and reciprocal citation are not generally attempts to manipulate reputation, we develop a simple and effective strategy to distinguish alleged manipulation from self- and reciprocal citations that occur without the intention to manipulate (we refer to these as justified self- and reciprocal citations). By controlling for shared content between cited and citing publications, we construct a proxy criterion to distinguish between manipulated and justified citations. Thus, we account for alternative motivations than manipulation. In addition to shedding light on the phenomena of self- and reciprocal citing, our empirical investigation shows that manipulation is associated with inferior quality. This implies that the ability of the scientific system to self-organize (Dasgupta and David, 1994; Martin, 2012) can (partly) correct for manipulation activity.

This paper contributes to the literature in three ways. First, it extends the literature on innovation and knowledge exchange communities (Faraj and Johnson, 2011; Faraj et al., 2011). It provides a new perspective on communities by introducing the dichotomy of implicit versus explicit reputation and by approaching the topic of reputation manipulation. Online communities provide a unique opportunity for implementing governance mechanisms to reduce or even resolve a wide range of cooperation and coordination problems. Exploiting explicit reputation is promising because it increases visibility and efficacy compared to implicit reputation. If reputation is explicit, there is no need for all participants to keep track of all other members' reputations. Instead, reputation is calculated and stored centrally.

We show that manipulation is an issue in such systems, and that governance mechanisms should be designed to minimize the possibilities to manipulate. More generally, this paper provides insights into collective innovation and selective incentives (Oliver, 2013; Olson, 1965; Osterloh and Rota, 2007; von Hippel and von Krogh, 2003). It highlights the downside to selective incentives by relating them to the concept of manipulation. Thus, we challenge the current one-sided, positive characterization of selective incentives by arguing that reputation-based reward systems are inherently accompanied by misleading incentives.

Second, we contribute to the development of a more integrated theory of innovation and knowledge exchange communities. The extant literature includes multiple empirical studies on communities that are isolated (e.g., Franke and Shah, 2003; Lüthje et al., 2005; Jeppesen and Frederiksen, 2006; Antorini et al., 2012; Hienerth et al., 2014). This paper makes a crucial step towards a more holistic and integrated picture of the organizational form of communities by providing a general understanding of reputation and its functioning as one of the main building blocks of communities.

Third, we contribute to the literature on the economics of science (Aghion et al., 2009; Stephan, 1996; Stephan, 2012), specifically the growing body of work on misconduct in science (see Lacetera and Zirulia, 2011; Martin, 2012; Martin, 2013) including the research on citation behavior by bibliometric methods (Fowler and Aksnes, 2007; Hyland, 2003; Kostoff, 1998). Manipulation based on authors' self- and reciprocal citing is more subtle and seemingly less harmful than the currently discussed forms of misconduct such as plagiarism or data fabrication. However, this subtlety might lead to lower concern over these practices and their more widespread use. Thus, the topic of manipulation in the form of self- and reciprocal citing might be underestimated at first glance.

This paper provides some practical implications for the governance of innovation and knowledge exchange communities. First, managers of innovation communities should be aware of the advantages and disadvantages of explicit compared to implicit reputation-based reward systems. Second, community designers and managers should consider the possibilities of reputation manipulation as this could affect the design and governance of existing platforms (e.g., for managing software projects such as GitHub). Third, we recommend policymakers in the scientific systems identify self-citations and distinguish them from other citations.

The remainder of this paper is structured as follows. Section 2 reviews the literature on reputation-based reward systems, elaborates the concept of implicit versus explicit reputation, and develops our hypotheses on manipulation in the form of self- and reciprocal citing in the scientific system. Section 3 describes the dataset and the research design. Section 4 presents the results. Section 5 discusses the findings and provides implications for theory and for practitioners.

2. Theoretical background and hypotheses

This section sets the theoretical background to the paper. Based on the notion of reputation-based reward systems, we distinguish between implicit and explicit reputation to study manipulation in the latter case (Section 2.1). We present reasons for our choice of the scientific system as the context for an empirical investigation (Section 2.2) and develop our hypotheses (Section 2.3).

2.1. Implicit and explicit reputation-based reward systems

Reputation-based reward systems are elegant solutions for public good and collective action problems. Reputation accounts for contributions and enables mechanisms that compensate individuals for innovating and revealing their private knowledge (Dasgupta and David, 1994; Stephan, 1996; von Hippel and von Krogh, 2003). Reputation-based reward systems can differ in terms of their representation of reputation. We introduce a distinction between *implicit* and *explicit* reputation.

We introduce the term *implicit reputation-based reward system* to refer to systems that rely on reputation but lack central coordination of reputation-related information with the result that