Harnessing the Wisdom of the Crowd in the Evaluation of Innovative Ideas

Jingshu Du
Vrije Universiteit Amsterdam
Management & Organization
j.du@vu.nl

Abstract

In this paper, I aim to explore which criterion, or a set of criteria, contribute to the popularity of an innovative idea in a crowd. So far there is no consensus as to what constitute a “good” innovative idea (Poetz and Schreier, 2012; Criscuolo et al., 2017). Therefore, a set of five criteria regarding the quality of innovative ideas are proposed and subsequently tested on their effect on the choice made by the crowd: clarity, originality, creativity, feasibility, and (visual) attractiveness. Second, I test whether democratize the decision-making power to the vast majority (e.g.: the crowd) is indeed a reliable strategy, when it comes to evaluate the potential of early-stage projects that are yet fuzzy and uncertain (e.g.: innovative ideas). More specifically, I investigate whether and to what extent the social factors are at play in crowd-based evaluation. Based on a unique dataset collected over a period of three years, most of the hypotheses are confirmed.
Harnessing the Wisdom of the Crowd in the Evaluation of Innovative Ideas

- “Crowds are moving into the mainstream; even if you don’t take advantage of them, your competitors surely will” (Boudreau and Lakhani, 2012)

1. Introduction

An important and recurring theme in the innovation literature has been the generation and selection of innovative ideas (Schumpeter, 1934; Chesbrough, 2003; Van de Ven, 2000, 1993; Burgelman, 1983). While increasingly more research is devoted to the “ideation” phase (e.g.: Cooper and Edgett, 2008; Vandenbosch et al., 2006), namely how to improve the creativity and novelty of ideas (Fleming et al., 2007; Wheelwright and Clark, 1992), we know rather little about the subsequent selection of those innovative ideas once they are generated. Traditionally, idea selection has been largely restricted to a handful of key decision makers in the organization, known as “gate-keepers” or innovation managers (Cooper, 1990; Cooper, 2008). However, relying on few decision makers can be highly risky, as the decisions are inevitably influenced by the individual’s bounded rationality (Cyert and March, 1963; March and Simon, 1958); path-dependency of prior knowledge and experiences (Nelson and Winter, 1982); limited attention span (Afuah and Tucci, 2012), and information processing capabilities (Simon, 1978; Tushman and Nadler, 1978). These limitations of individuals have resulted in “local search” of organizations (Martin and Mitchell, 1998; Stuart and Podolny, 2010; Piezunka and Dahlander, 2015) and more alarmingly, lead to numerous “false positives” and “false negatives” (Kaplan, 2011; Gassman, 2014) in the evaluation and selection process of innovative ideas. The situation is further exacerbated by the highly uncertain and risky nature of the “fuzzy front end” of innovation (Burgleman, 1983; Khurana and Rosenthal, 1998; Schemmann et al., 2016), which adds an additional layer of complexity to the selection process — being new also implies there is no readily available routine to follow when making decisions, nor it would be easy to disentangle any possible human factor surrounding an innovative idea and the inherently risky nature of the idea itself, thus makes it unclear to attribute the subsequent success or failure of an innovative idea to few decision maker/gatekeeper involved.

To overcome the cognitive limitation of few individual decision makers in the idea selection process, and to empower a broader circle of stakeholders— the “collective intelligence”, a number of new approaches have been proposed in the open innovation paradigm (Chesbrough, 2003). Among which, leveraging the “wisdom of the crowd” (Boudreau et al 2011; Jeppesen et al 2010; Afuah and Tucci, 2012) has gained increasing popularity in recent years. As it is stated, “crowds are moving into the mainstream; even if you don’t take advantage of them, your competitors surely will” (Boudreau and Lakhani, 2012, p 53).

The shift in focus from merely depending on a few key decision makers to an increasing reliance on the vast majority in the crowd has been guided by the expectation that the designation of decision making power to the crowd will generate fairer, more accurate, and thus more reliable evalua-
tions, compared to the (likely) biased decisions made by only few people in/related to the organization. Examples such as the Chicago-based T-shirt company Threadless (Lakhani and Kanji, 2008) and the various communities in the open source software (West and Lakhani, 2008; von Hippel and von Krogh, 2003) all seem to support this assumption. However, research has also shown that managers remain understandably cautious in delegating tasks to the crowd, and a main reason is that managers don't clearly understand how to manage the process (Boudreau and Lakhani, 2012; Birkinshaw et al., 2011). In fact, despite its potential, not every crowd-based evaluation works (Birkinshaw et al., 2011). Therefore, whether these potential benefits of the “wisdom of the crowd” can really be realized and what are the underlying mechanisms through which crowd-based evaluation functions, remain an under-researched puzzle in the innovation literature.

The goal of this paper is mainly two-fold: First, I aim to explore which criterion, or a set of criteria, contribute to the popularity of an innovative idea in a crowd. So far there is no consensus as to what constitute a “good” innovative idea (Poetz and Schreier, 2012; Criscuolo et al., 2017). In this research, a set of five criteria regarding the quality of innovative ideas are proposed and subsequently tested on their effect on the choice made by the crowd: clarity, originality, creativity, feasibility, and (visual) attractiveness, and how these are compared to the decisions made by experts/professionals. Second, I test whether democratize the decision-making power to the vast majority (e.g.: the crowd) is indeed a reliable strategy, when it comes to evaluate the potential of early-stage projects that are yet fuzzy and uncertain (e.g.: innovative ideas). In other words, whether the crowd-based evaluation is indeed neutral and trustworthy as it is assumed to be. More specifically, I investigate whether and to what extent the social factors are at play in crowd-based evaluation. Although mostly being online and anonymous to one another, crowd is composites of real people in the real life (offline) setting. As such, they may carry and be influenced by the characteristics in the “offline” social environment (such as homophily and in-group favoritism), which may likely affect their (seemingly neutral) evaluation of early-stage innovative ideas. Based on 416 crowd-sourced innovative ideas and 1679 evaluations that are uniquely collected over a three-year time period, my findings corroborate the hypotheses that crowd makes different evaluations compared to that of experts, whereas homophily and in-group favouritism indeed (still) play an important role even in the seemingly neutralized, online setting. Different from what we know in offline social settings, crowd in-group dynamism tends to decrease along with the group tenure. Moreover, participants who received initially disadvantageous evaluations tend to lobby for more in-group support, and allocating more decision making power to the crowd will affect crowd in-group activities.

2. Theoretical Background and Hypothesis Development

2.1 Crowdsourcing and Crowd-based Evaluation

The “wisdom of the crowd” has gained increasing popularity in both academia and the business world since its inception in 2006 (Howe, 2006). The most popular form of leveraging the wisdom of the crowd has been outsourcing challenging innovation tasks which were initially assigned to internal employees, to an undefined network of outside participants, usually in the form of an open call (Howe, 2006). This is guided by the assumption that virtually everyone in the crowd has some potential to contribute valuable information (Greengard 2011). By broadcasting the task to the public and solicit-
ing ideas from the crowd, companies can reach out to many previously untouched brains and territo-
ries, which helps in overcoming the barriers of “local search” (Afauh and Tucci, 2012) and unveiling
potentially better ideas. Prior research has shown that this loosely coupled organizational form (Bru-
soni et al., 2001; Dahlander and Wallin, 2006; Jeppesen and Laursen, 2009; Orton and Weick, 1990)
allows firms to obtain surprisingly innovative solutions in a cost-efficient manner (Bullinger et al.,
2010; Harhoff and Mayrhofer, 2010; Nambisan and Baron, 2009, 2010; Poetz and Schreier, 2012;
Terwiesch and Ulrich, 2009; Franke et al., 2013), and as such, it is becoming an increasingly popular
phenomenon.

Existing research on crowd-based innovations has been mainly focusing on the benefits of the
crowd in the idea generation process, showing that for certain types of problems, crowds outperform
the company (Boudreau and Lakhani, 2012), and the ideas the crowd generates score significantly
higher in terms of novelty and customer benefit compared to professionals (Poetz and Schreier, 2012).
Moreover, those firms which use crowdsourcing to capture the knowledge of the crowd can develop
distinctive innovation competences, which help them to achieve better performance (Xu et al., 2015).

Despite the numerous benefits of idea generation from crowdsourcing, serious problems arise.
Among which, a key challenge is the large number of ideas to be processed and selected. Management
typically underestimates the amount of work that is needed after ideas are collected from the crowd
(Birkinshaw et al., 2011; Van Dijk and Van den Ende, 2002), and sometimes a crowd can return a vast
amount of noise that may be of little relevance (Keen 2007). In that case, having many innovative
ideas does not really help sort through nor synthesize information. Instead, it may lead to information
and cognitive overload, which can cause problems on evaluating submitted feedbacks and identifying
qualified ones (Zhao and Zhu, 2014). Especially in nowadays information-rich contexts that are char-
acterized by a large supply of ideas and (digital) information, the scarce resource is typically not the
ideas per se, but the amount of attention that individuals can allocate to searching for, sorting through,
and evaluating the available ideas (Ocasio, 1997; Hansen and Haas, 2001). As Simon (1997: 40) ar-
gued, “a wealth of information creates a poverty of attention”, which can be very challenging to han-
dle by only few decision makers in the organization.

Another problem with idea selection by a limited number of key decision makers is the high
likelihood of biased decisions. Due to cognitive limitation and bounded rationality of individuals
(Cyert and March, 1963; March and Simon, 1958), most decision makers fail to pick new product
winners correctly— only one out of every five new product selected to launch is successful (Ho and
Chen, 2007). Crowd-sourcing ideas although allows for increased opportunity to tap into distant ideas,
the approach itself however cannot solve the root cause problem of “local search” (Martin and
Mitchell, 1998; Stuart and Podolny, 2010) and “learning myopia” (Levinthal and March, 1993) if the
ideas sourced are still processed through few individuals. In fact, research shows that instead of help-
ing firms to branch out to distant ideas, crowding actually narrows the attention of organizations: as
the management is more likely to pay attention to suggestions that are familiar, not distant (Piezunka
and Dahlander, 2015). Consequently, Riedl and colleagues (2010) indicate that due to the inherent bi-
ases of decisions made by few people, there is a strong need for an evaluation mechanism to identify
the best ideas (Zhao and Zhu, 2014). To this end, some organizations start to employ popular voting/
rating mechanisms by the public to evaluate the quality of ideas solicited from the crowd. This approach leverages the “wisdom of the crowd” in its subsequent idea evaluation process, as a great diversity of viewpoints and input from the crowd can deter self-serving bias and belief perseverance (Bonabeau, 2009) and can therefore help combating pattern obsession and negative framing effects (Bonabeau, 2009), which, in turn, may help mitigate biases of (few) individual decision makers, and could generate potentially more fair and reliable evaluations.

Despite its potential, so far, little research has been conducted to investigate the effect and mechanism of the crowd in the evaluation of innovative ideas. Understanding crowd-based evaluation is important, not only because there is lots of (yet untouched) “wisdom of the crowd” (Howe, 2006; Afuah and Tucci, 2012), but also the need for involving more stakeholders in the key stage-gates of the innovation process (Cooper, 2008) for better innovation decisions.

2.2 Crowd-based vs. Expert Evaluation

One main reason for relying on crowd-based evaluation is the (assumed) objectivity and reliability of the collective intelligence, as compared to depending on only a few key decision makers in (or related to) the organization. Further, outsourcing evaluation tasks to the crowd may also lower the managerial burdens of the management, and make the decision making process more efficient. But a pressing question remains unanswered: Does collective intelligence really correct for decision biases? (Bonabeau, 2009). Despite the growing role of crowds in making decisions once left to experts, little is known about how crowds and experts may differ in their judgement of innovative ideas. Indeed, for crowds, there is even considerable debate over whether their decisions are actually based on rational criteria, and if they are, what differences might exist between the judgment of crowds and that of experts (Mollick and Nanda, 2015).

Crowd-based evaluation may differ from expert evaluation for a number of reasons. First, there may exist significant differences in the disciplinary, social, and educational background of the crowds and the experts, which make their knowledge repository rather different. Crowd-based evaluation is valued mainly because of its diversity of opinions. When gathered together, these opinions may complement each other and thus show more objective patterns compared to those of (few independent) experts. Second, compared to the experts, many members in the crowd may have a “peripheral” position to the specific idea/problem at hand (Dahlander and Frederiksen, 2012). As such, most participants in the crowd may tend to take a different angle compared to the one of the experts, who have their expertise centered on the idea. As centrality in a community (e.g.: in this context, the position of experts) can cause individuals to become inward-looking and be blind to distant novelty (Katz and Allen, 1982; Dahlander and Frederiksen, 2012), the peripheral vision of the crowd may help develop its own set(s) of criteria, which can be different from those of the experts'. Third, crowds are also different from experts in terms of their focus interest points (von Hipple, 1979). While experts may be more attracted to the technological/complex knowledge domain of an innovative idea, crowds may on

---

1 Examining funding decisions for proposed theatre projects, Mollick and Nanda (2015) found significant agreement between the funding decisions of crowds and experts, which suggests that leveraging crowd-based evaluation may boost the efficiency of the organization.
the other hand pay more attention to the functional and usability of an idea, which lead to different evaluations about the same idea. Taken together, I hypothesize:

**Hypothesis 1 (Baseline hypothesis): Crowd-based evaluation significantly differ from expert evaluation.**

### 2.3 Social Dynamics in Crowd-based Evaluation

In general, the evaluation of innovative ideas has been (briefly) mentioned in two literature streams. The innovation literature typically attributes the success or failure of an idea to its technical characteristics (Utterback, 1971; Cooper, 2008; Bower and Christensen, 1995; Christensen and Overdorf, 2000), with the underlying assumption that technically more advanced ideas are more easily selected (Cooper, 2008). While the management literature, on the other hand, has suggested that once an initial idea is proposed, it is the social interaction that is conducted surrounding the idea that determines its evaluation (Eisenhardt and Bourgeois, 1988; Kijkuit and van den Ende, 2007). I argue that, even in online settings which feature anonymous actors (e.g.: in the crowd context), homophily and group effects may still play an important role affecting crowd-based evaluations, for the following reasons:

First, social contacts may help overcome information asymmetry (Stigler, 1961) of an innovative idea by connecting the evaluator to, and explicate the potential of, a specific idea. This, in turn, may increase the likelihood of an idea being noticed and subsequently being picked up by the evaluator. As a large number of ideas are collected from the crowd, it becomes technically impossible for the evaluator to go through every single idea that is submitted. Therefore, social links may help channeling the idea to the evaluator and gain (potential) support from him/her. For example, based on data from a crowdfunding platform that connects artist-entrepreneurs with investors over the internet for financing musical projects, Agrawal and colleagues (2011) show that most early backers of the project are located in the same geographical area of the submitter. These are likely to be people with whom the submitter has a personal connection, established through real-life interactions. In a similar vein, using evidence from case studies on projects posted on three large crowdfunding platforms, Ordanini et al. (2011) documented that in the initial phase of a crowdfunding project (which the authors call friend-funding phase), contributions are primarily made by the close friends of the submitter, who join the platform only to support their friends’ initiative.

Second, for the potential of an innovative idea to be recognized, it needs mutual understanding of the idea submitter and the evaluator (Miura and Hida, 2004; Mumford and Gustafson, 1988; West and Anderson, 1996) especially in non-redundant and heterogeneous networks. Hence, despite its assumed objectivity, homophily may still dominate decision making in crowd-based evaluations.

Third, reciprocity may also play an important role in crowd-based evaluations. In the context of evaluating complex, uncertain, and yet fuzzy innovative ideas, there is typically no clear-cut criteria been put forward upfront, allowing for sufficient leeway for crowd members to express each other’s opinions, which can be based on and influenced by their social networks. As a consequence, the network structure of crowd-based evaluation may function by means of reputation and sub-group norms (Krackhardt, 1999; Reagans and McEvily, 2003). These mechanisms ensure that members are more likely to demonstrate cooperative behavior and facilitate the development of group rules or shared val-
ues (Kijkuit and van den Ende, 2007; Tsai & Ghoshal, 1998). As it is proposed, the network surrounding an idea may evolve from a non-redundant structure to a cohesive structure creating social integration, clear expectations and a common frame of reference (Kijkuit and van den Ende, 2007). The idea submitter’s social contacts within a platform are useful because they spread information about an idea and more importantly, they may trigger reciprocity through a feeling of perceived obligation (Coleman, 1988). Considering the above-mentioned arguments, I hypothesize:

Hypothesis 2: (even in neutralized online settings) Homophily and In-group favoritism play important roles in crowd-based evaluation.

The decision-making literature on group processes has found a positive correlation between social integration and consensus formation (Lott and Lott, 1961; Smith et al., 1994). Coleman (1988) stresses the importance of social cohesion. In this network structure, control or action comes from trust, norms of cooperation and reputation (Coleman, 1988; Obstfeld, 2005). A cohesive network structure creates benefits such as support, coordinated action and clear expectations (Coleman, 1988; Obstfeld, 2005; Reagans and Zuckerman, 2001). As group members interact with each other for a longer period of time, it is likely for them to develop strong ties which are characterized by frequent two-way interactions (Granovetter, 1973; Hansen, 1999). The formation of these ties take time and require more effort, but create trust (Reagans and McEvily, 2003) and mutual understanding (Gilsing and Nooteboom, 2005), which facilitate the transfer and construction of knowledge, especially more complex knowledge (Handley, 2006; Hansen, 1999; Roberts, 2006; Uzzi, 1999). Therefore, the longer the crowd members spend together, the more likely the homophily and in-group favoritism play a role in crowd-based evaluation.

Hypothesis 3: Crowd member’s social activity (in-group favoritism) tends to evolve with group tenure, in such a way that the longer the crowd members are in the same task, the more in-group favoritism is displayed.

However, it is unlikely that such effect will appear equally to every crowd member, in particular in the context of anonymous online setting which is designed to diminish the (likely biased offline) social interactions. One differentiating factor is the (dis)advantageous position of the crowd member. According to social comparison theory (Festinger, 1954), the existence of a discrepancy in a group with respect to opinions or abilities will lead to action on the part of members of that group to reduce the discrepancy (Festinger, 1954). When a discrepancy exists with respect to opinions or abilities there will be tendencies to change one’s own position so as to move closer to others in the group (Festinger, 1954); The stronger the attraction to the group, the stronger will be the pressure toward uniformity concerning abilities and opinions within that group (Festinger, 1954). I suppose that, the crowd members who are in a disadvantageous position, may feel stronger the need to catch up and to “lobby” for more support from his/ her immediate contacts for the innovative idea, hence, may display stronger resort to group support.

Two reasons may explain this motivation for in-group “lobbying”: First, the crowd member with poorly-received (lowly-rated) innovative idea may feel injustice in the evaluation that he/ she has received and the innovative idea is under-valued. This may provoke him/ her to reach out for support from his/ her immediate circle— in many cases the group that he/ she is in; Second, the crowd mem-
ber with poorly-received (lowly-rated) innovative idea may see the need to get more support just to avoid being lagging behind, which will also likely to stimulate him/her to engage in more in-group lobbying. Taken together, I hypothesize:

Hypothesis 4: More importantly, in-group favoritism is stronger for those who have a disadvantageous position in the crowd-based evaluation.

Apart from (anonymous) interactions in the online context, crowd members may at the same time also engage in offline activities with one another. For instance, people who are in the same dance club may at the same time also go to the same (or different) golf club(s). For dance lessons, being in the same dance club implies members are in the “immediate” network, while being in the same golf club meaning although members share contacts, these contacts are in a “distant” network. Prior research tends not to distinguish the different natures of the networks, and focus primarily on immediate networks. In this paper, I argue, group members may be more inclined to resort to group support from the contacts in their “distant” network compared to their “immediate” network, for the following reasons:

First, group members collaborate but at the same time also compete with each other. In such co-opetition relationships (Nalebuff et al., 1996), one may feel pressure when exposing too much information with his/her immediate in-group members, while more at ease if lobbying with non-competing contacts in other, distant networks.

Second, *paribus ceteris*, and assuming strict reciprocity among group members (everybody votes for everybody else in the group), gaining support from immediate in-group members can only bring perfectly balanced results (e.g.: the number of votes received per idea), which is insufficient for any group member to stand out. Therefore, one way to gain more support and to possibly outperform the others in the immediate network, is to resort to support in distant networks. Taken together, I hypothesize:

Hypothesis 5: More importantly, in-group favoritism is stronger in distant networks than in immediate networks in the crowd-based evaluation.

2.4 Organizational Design in Crowd-based Evaluation

The design of crowd-based evaluation is crucial to fully leverage the wisdom of the crowd. This is mainly based on two related but opposing factors: control and motivation.

On the one hand, common to all forms of collective intelligence is a loss of control (Bonabeau, 2009). Hence, although delegating a large part of the decision-making power to the crowd, the organizer may (still) want to retain a high level of control in crowd-based evaluation. Giving away (part of) its control over the final result may bring issues of distrust, misfit, and unreliability of the decisions made. Hence, one of the biggest issues with respect to control is whether to include outsiders—not to mention the crowds—in the process (Bonabeau, 2009). The mechanism design is always difficult when relying on collective intelligence (e.g.: the crowd) in making decisions (Bonabeau, 2009). Closely pertaining to control is quality assurance. As the application of crowd-based evaluation often lack any explicit refereeing nor tracing process that might provide some degree
of quality assurance (Mollick and Nanda, 2015), making sure the crowd does not just make random decisions nor abuse its power then becomes very important in crowd-based evaluation. A reasonable approach is then limiting the decision-making power of the crowd. As a result, a certain level of control may be needed at the top of the organization, in order to ensure both the control and quality of crowd-based evaluations.

On the other hand, keeping too much control at the top may diminish the engagement and motivation of the participants in the crowd. A key element to crowd-based activities is to ensure many (engaged) participants in the process (Birkinshaw et al., 2011). Therefore, organizations must keep a continuous flow of enthusiastic participants to keep engagement high, or they need to provide incentives to sustain people’s motivation over time (Bonabeau, 2009). One way to achieve this is to give sufficient power to the crowd in making decisions. According to the dual-factor theory (Herzberg et al., 1959; Herzberg, 1965), an increased level of responsibility would motivate the participants to be more engaged in their activities. In other words, distributing and decentralizing the decision-making power to the crowd may motivate participants to be (more) engaged in crowd evaluation.

Considering both sides of the spectrum: control and motivation, an increased level of decision-making power to the crowd may also stimulate in-group (lobbying) activities, as the crowd now has a bigger say on its part in making/ influencing decisions, therefore the expectancy (Porter and Lawler, 1968; Vroom, 1964) to obtain a better result (e.g.: better evaluation of the idea) through in-group support may now be higher than if the crowd has little power in making/ influencing decisions. This, in turn, may motivate the participants to make more efforts in lobbying his/ her group contacts as the decision-making power assigned to the crowd increases. Taken together, I hypothesize the following:

Hypothesis 6: The more decision making power is distributed to the crowd, the more in-group favoritism occurs in crowd-based evaluation.

3. Data and Sample

For this research, a unique dataset has been employed, which is collected over a period of three years from September 2015 to October 2017. The unit of analysis in this research is the social networks of individual participants in a crowdsourcing environment, and the influence of social dynamics on the crowd-based evaluation. For the purpose of this research, the research context is created in the form of multiple rounds of contests that were integrated in an inter-university elective programme hosted by one of the largest research universities in The Netherlands. The programme (and the courses therein) is open to all third-year students (from diverse disciplines) in all Dutch research universities across the country. This creates an ideal research setting that all students applied for this programme are intrinsically interested in the topic (otherwise they may have chosen a different programme), and because students are from different disciplines and different universities, chances are very low that they know each other already before the course starts. This provides an ideal setting to study how social dynamics play out and gradually evolve during the course of three consecutive contests each year. The crowdsourcing environment follows the tournament-based design that was proposed by Afuah and Tucci.

---

2 In total, there are 13 research universities in The Netherlands. These universities are: University of Amsterdam, Vrije Universiteit Amsterdam, Erasmus University (Rotterdam), Leiden University, Tilburg University, Maastricht University, Utrecht University, Delft University of Technology, Edinhoven University of Technology, University of Groningen, Twente University, Radboud University Nijmegen, Wageningen University.
In each year, three innovation challenges are sequentially posted to participants in the crowd, and innovative ideas to address these challenges are solicited. This mirrors the behavior of businesses that outsource their problem solving needs by crowd-sourcing. Each participant is anonymized with a random ID in the crowd-sourcing system, therefore the crowd members can only see a random number associated with the innovative idea, but not the name nor any identifier of the author of the idea. This is intended to bring neutrality to the contest. After the innovative ideas to the first innovation challenge are collected and the results are announced, the second innovation challenge is posted, so on and so forth. For each innovation challenge, each participant has the opportunity to select/vote for the innovative ideas that they like the most. As the voting is completely voluntary, most but not every participant voted (which is very much the case in crowd-based evaluation). In total, 416 innovative ideas were submitted, and 1679 votes were casted.

The evaluation of the quality of these innovative ideas are based on a weighted combination of two parts: 1) a centralized evaluation, namely expert evaluation. In this case, the evaluation of the course coordinator; and 2) a decentralized evaluation, crowd-based evaluation—in this case, the number of votes received per innovative idea from the crowd. For the first part, expert evaluation, each innovative idea is evaluated by the coordinator on a set of four criteria: information clarity, originality, creativity, and feasibility. For the second part, crowd evaluation, following tournament-based crowd-sourcing (Afuah & Tucci, 2012; Wooten & Ulrich, 2017), in each round of challenge, each participant is given the opportunity to vote for the five best innovative ideas of their favorite. All innovative ideas are then ranked by their popularity—the number of votes the idea has received from fellow participants. For the purpose of this research experiment, the weight that each part counts in the final decision of the evaluation has changed from year 2015 to year 2016/2017: in year 2015, both the expert evaluation and crowd evaluation count for 50% of the final evaluation (grade) of each innovative idea; while in 2016 and 2017, the weight assigned to expert evaluation increased to 70%, and the weight assigned to the crowd evaluation decreased to 30%.

Further, in 2017, more guidelines have been added to the voting system, and clear criteria have been provided (information clarity, originality, creativity, and feasibility, same as what are used in the expert evaluation). Instead of simply casting a vote, participants who wish to support their favorite innovative ideas now also have to rate clearly on each individual criterion before a final vote can be casted.

Considering the potential bias of peer voting due to the social reasons elaborated earlier, we asked three independent raters (all are at the Master degree level and are familiar with the subject of study) to rate the innovative ideas separately on the same criteria: information clarity, originality, creativity, and feasibility. Moreover, considering the possible influence of visual appearance (attractiveness) of an idea, the visual attractiveness is also rated by the independent raters (which are later on used as a control variable). Next, Fleiss’ Kappa is calculated in order to analyze inter-rater reliability—the reliability of agreement among these different raters (Fleiss, 1971). The results from Fleiss’ Kappa show that the agreement among three independent raters can be considered as good (between 0.5-0.7), thus the three independent assessments are integrated into one score per criterion by taking the average score of the three raters on each criterion.

---

3 This change was based on popular request from the students in the course in 2015, that in-group lobbying is going on despite the intention was to diminish such activities by anonymizing voters and submitters in online voting.
4. Variable Definition

This part of the research will provide an overview of the variables in this research, including their operational definitions.

4.1.1 Dependent Variable

The crowdsourcing platform is specifically designed for this research purpose, which keeps track of the voting history: e.g.: who voted for whom and how many votes each of the innovative ideas received. The number of votes is corrected by deducting the number of ‘self-vote’ from the total votes received per idea (as students are allowed to cast one of their votes for their own idea). While some students used all their votes in support of others’ ideas, some other students did vote for themselves. Through this approach, the count variable ‘Total_Votes’ is constructed, which functions as the dependent variable in this study. More specifically, as each year there are three rounds of contests, the number of votes a participant received for his/ her innovative idea is calculated per contest. This enables me to study the evolution of participants voting behavior. Appendix A provides the descriptives for the total number of votes individuals received for their innovative ideas.

4.1.2 Independent Variables

In this study, two sets of independent variables are used. The first set of independent variables are concerned with the evaluation criteria of the quality of the innovative ideas, and the second set of independent variables are related to the social dynamics (in-group favoritism) in the crowd-based evaluation.

4.1.2.1 Evaluation Criteria of the Quality of Innovative Ideas

Prior research has shown that the quality of innovation serves as a determinant for the variance of the preference of an innovation (Jones and Stevens, 1999; Maute and Locander 1994). Based on prior research (e.g.: Poetz and Schreier, 2012) and experience, a set of four evaluation criteria are incorporated in this research, each of them is measured on a scale from 1 to 10:

**Information Clarity**: How novel, accurate, and relevant the information is. Prior research has shown that because the information is voluntarily offered, ideas from the crowd often show a low degree of elaboration and thus can sometimes be vague and immature (Magnusson 2009; Di Gangi and Wasko 2009; Di Gangi et al. 2010; Bayus, 2013). Therefore, in this research, clarity of the innovative idea is considered as one of the key dimensions of evaluation.

**Originality**: How original the innovative idea is. In addition to information clarity, originality of the idea constitutes another important dimension of evaluation. As a large number of ideas are proposed, it should not be surprising if some of the ideas are already known to the focal organization, which, in turn, brings little added-value to the organization. For example, based on a research of the crowd-sourced ideas for the company Dell, Bayus (2013) showed that several consumer ideas that are proposed by the crowd have actually already been implemented by the company, some have even been used for a long time already. Therefore, innovative ideas which have low levels of originality may be of little use to the organization.
Creativity: How creative/innovative the idea is. Prior research suggests that several ideas from the crowd enables distant search and can be very creative. Consequently, creative ideas are novel (relatively new compared to other available ideas) (Amabile 1996, Shalley et al. 2004, George 2007, Burroughs et al. 2008).

Feasibility: Whether and how feasible the innovative idea is. Prior research shows that a brilliant but unfeasible idea can actually be disempowering if the organization lacks the capacity to act on the idea (Birkinshaw et al., 2011). The vast majority of really creative ideas have no commercial value (Levitt 1963, Silverberg and Verspagen 2007) and that many patented ideas ended up being dusted on the shelf (Czarnecki et al. 2011; Chesbrough, 2003). Achieving the organization’s innovation goals therefore requires that some ideas are actually valuable enough to be implemented (Mumford and Gustafson 1988, West 2002, Franke et al. 2006). In the words of Levitt (1963, p. 79), “Ideas are useless unless used. The proof of their value is in their implementation.” Consequently, in this study, feasibility is a fourth evaluation criterion of the ideas solicited from the crowd.

1) Creating a normative foundation for quality

In this research, the innovative ideas are evaluated based on a set of four criteria as discussed above. In order to assess the quality of each innovative idea, three independent raters are recruited to assess the quality of the submitted ideas on each of these four criteria. To gain neutral assessments, each innovative idea is evaluated per item by the three independent raters, and the Fleiss’ Kappa is consequently calculated based on their rating. As good Fleiss’ Kappa scores imply mutual agreement among the raters, the average scores among the three raters per item per innovative idea were then generated. These are used as a set of quality criteria in the regression.

First, the information clarity criterion is based on whether the submitted idea makes use of true assumptions, elaborated on the information used, and to what extent the quality and use of these external sources is sufficient. Second, the value that captures the level of originality is based on to what extent an idea already exists. This value ranges between ideas being new to the world, new to the industry, new to the firm, and known to the firm. Third, to determine whether an idea is creative, it is based on whether an idea is novel and new. Finally, the quality of the idea is also associated with whether the it is feasible. In this context, feasibility is interpreted as whether challenges in implementing the idea are identified and if the submitter provides a clear strategy to deal with these challenges.

A Fleiss’ Kappa is calculated to analyze the reliability of agreement and create one reliable assessment for each of the qualities per innovative idea. This method assesses the reliability of agreement between a fixed numbers of raters for each of the ratings per categories per subject (Fleiss, 1971: 378). The Fleiss’s Kappa analysis provides an indication about to what extent the overall evaluation of the raters are internally consistent, it also provides an indication about whether and to what extent the raters disagree. The following notation of Fleiss’ Kappa captures the internal consistency between the independent ratings of individuals:

\[ k = \frac{\bar{p} - \bar{pe}}{1 - \bar{pe}} \]
The following formula captures the calculation with which Fleiss’ Kappa can be generated:

\[ P_j = \frac{1}{n_n} \sum_{i=1}^{n} n_{ij}, \quad 1 = \sum_{j=1}^{k} P_j \]

Where \( N \) represents the total number of innovative ideas, the \( n \) represents the number of ratings per subject, \( k \) represents all the possible rating categories, and \( n_{ij} \) represents the number of raters who assigned the \( i \)-th subject to the \( j \)-th category (Fleiss, 1971: 379). The results that are generated by this equation are interpreted with the use of the interpretation table proposed by Landis and Koch (1977). The \( k \) value that is generated by the Fleiss’ Kappa analysis can be interpreted as: .01< a poor agreement; .01-.20 a slight agreement; .21-.40 a fair agreement; .41-.60 a moderate agreement; .61-.80 a substantial agreement; and .81-1.00 an almost perfect, to a perfect agreement (Landis & Koch, 1977).

2) Transforming capabilities into quality variables

A pilot exercise was conducted to independently evaluate the submitted ideas of year 2015 and 2016, which led to a first representation of the agreement among the raters. Based on results of the pilot exercise, a grading “scheme” was created on how each of the four criteria should be assessed. After this, an independent evaluation was conducted based on the agreed measurement scale. The measurement scale was also adjusted to a five-point Likert scale. Fleiss (1971) has argued that fewer categories are more likely to result in agreement. The re-evaluations have resulted in relatively high levels of internal agreement amongst the ratings of the three independent raters.

Appendix A provides the combined Fleiss’ Kappa scores of the ratings of three raters per evaluation criterion per contest. More specifically, the Fleiss’ Kappa scores show values ranging between .483 and .908, indicating that there is moderate to almost perfect agreement amongst the three independent raters. These scores from independent raters are then averaged to generate the combined Fleiss’ Kappa scores that are used in the analyses. The Fleiss’ Kappa calculations are available upon request.

After having confirmed the internal consistency between the independent ratings of the raters, the ratings are composed into a weighted average for each of the four evaluation criteria. These weighted averages are captured by four variables with an interval measurement scale with values ranging from ‘1’ to ‘5’. First, information clarity is captured by the variable ‘AVG_INFO_Value’. Second, the variable ‘AVG_ORIG_Value’ captures the originality of an idea. Third, the variable ‘AVG_CREA_Value’ captures the level of creativity of an idea. Finally, the feasibility of an idea is captured by the variable ‘AVG_CRIT_Value’.

4.1.2.2 Social Dynamics in Crowd-based Evaluations

Another set of independent variables is concerned with the social dynamics in the crowd-based evaluation. Prior research shows that social networks can exert strong influence on human behavior and consequently, on the evaluation outcome. In this study, I empirically examine it in an online setting.
The voting system in this study provides the foundation on which the evolution of social networks and social influences in crowd-based evaluation can be examined. As the innovative ideas are submitted to an online platform, the participants can read the submitted ideas of their fellow students and cast their votes accordingly. Being anonymized to the students, the background information of each participant at the ‘back office’ enables me to capture the personal characteristics of each voter and the idea submitter. In this context, it is possible to examine to what extent social networks influence the voting behavior of individuals in the online setting.

Research on social networks, has shown that individuals who share the same characteristics are more likely to connect to one another (Feld, 1981; Marsden, 1988; Louch 2000, Jeppesen & Fredericksen, 2006). These individuals are also expected to develop strong ties that are characterized by a high level of interaction intensity and knowledge sharing. For this reason, shared group membership, shared educational background, shared gender, shared ethnicity between the voter and the idea submitter are included as main independent variables.

4.1.2.3 Social Network Variables

Homophily is an important topic in sociology. Feld (1981) and Louch (2000) have stated that individuals connect through central points of focus. These points of focus create shared characteristics that make it more likely for individuals to connect with one another. Feld (1971) provides shared places and activities as examples of these points of focus. Based on this research, it is assumed that people who share the same characteristics are more likely to develop a preference for one another’s ideas and are therefore more likely to vote for each other. In order to control for the influence of these shared characteristics, this research incorporates a variable that measures whether individuals share the same educational background as university of origin. Moreover, the sample shows that students come from the same university may or may not follow the same subject of study. Considering the mutually exclusive and collectively exhaustive (MECE) principle, four variables are therefore constructed that take into account both the educational background as well as the university of origin.

To control for the educational background, this research incorporates the count variable ‘Shared_Background’. This variable captures the number of votes individuals received from students with the same educational background. The student’s educational background varies between fourteen studies, namely: (1) Biomedical, (2) Business Analytics, (3) Communication and Information Sciences, (4) Economics, (5) History, (6) International Studies, (7) Law, (8) Mathematics, (9) Motion Sciences, (10) Political Sciences, (11) Psychology, (12) Public Administration and Organizational Sciences, (13) Health Science, and (14) Science, Business and Innovation. If the voter and the idea submitter share the same study background, a ‘1’ is designated; otherwise, it is noted with a ‘0’. The total count of designated ‘1’ then captures the total number of votes that are received from voters with a shared educational background.

On the other hand, students are also found to come from different universities. Therefore, some students might share the same university as point of focus (Louch, 2000). To control for the effect of shared university, this research incorporates the count variable ‘Shared_Place’. This variable contains the total number of votes an idea received from voters who share the same University of Origin. The students in the sample are from six different universities. If the voter and the idea submitter
are from the same university, a ‘1’ is designated, otherwise, a ‘0’ is noted. The total count of designated ‘1’ then captures the total number of votes which are received from voters who are with a shared university of origin.

In addition, Marsden (1988) suggests that not only the external characteristics but also some inherited characteristics constitute the homogeneity between individuals. Two of these characteristics that have been proposed by Marsden (1988) are gender and ethnical background. This implies that individuals are more likely to connect to someone of their own gender and their own ethnicity. To control for gender this research incorporated the count variable ‘\textit{Shared\_Gender}’. This variable captures the number of votes that are received from voters who share the same gender as the submitter of the idea. It is coded a ‘1’ if both the voter and the submitter share the same gender, and a ‘0’ if otherwise. The ethnical similarities are taken into consideration by implementing the count variable ‘\textit{Shared\_Ethnicity}’. This variable captures the total number of votes received from voters who are with the same ethnical background. If the voter and the individual share the same ethnicity, a ‘1’ is designated, otherwise a ‘0’ is coded. The total count of designated ‘1’s in each category then captures the total number of votes that are casted by a voter with a shared ethnicity with the idea submitter.

4.1.2.4 Group favoritism variable

Besides the above-mentioned individual level constructs, another important determinant of social networks is whether individuals share the same group membership (Brewer 1979; Postmes et al., 2000). Being part of a group is expected to increase the homogeneity between group members, as a group is found to create and share a social identity, which, in turn, may lead to group favoritism and even cliques. In the experiments, students were required to deliver a group assignment at the end of the course period. For this group assignment, the students were randomly distributed over 10 groups in 2015, 14 groups in 2016, and 14 groups in 2017. This construction provides the basis for comparing who has voted for whom and if they are related to the same group membership. To this end, a dummy variable ‘\textit{Shared\_Group}’ is added, capturing whether the votes received by an innovative idea is actually casted by one of the group members of the submitter. The value ‘1’ is designated to votes that are from members of the same group, and a value of ‘0’ otherwise. The total count of designated ‘1’s then captures the total number of votes received from members of the same group.

During the course period, most students are at the same time also taking another course in the programme. In which they are also assigned to groups. This provides an ideal research setting for studying the immediate social networks (groups the students are assigned to in the focal course), and distant social networks (groups in the other course). To study the effect of both immediate and distant networks, student groups in the other course are also coded.

In addition, the prior literature has suggested that group favoritism is something that develops over time (Brewer 1979; Postmes et al., 2000). This argument suggests that time may moderate the relation between shared group membership and the number of votes received by group members. Therefore, a variable indicating the sequential number of each challenge is created, which captures the time element. This makes it possible to examine whether voting behavior changes during the course period. In order to capture the time element, the ‘\textit{Number\_Challenge}’ variable is added to the analy-
sis. As each year we run three challenges, this variable has an ordinal measurement scale ranging from ‘1’ to ‘3’.

### 4.2 Control Variables

Apart from the above-mentioned variables, in this study I used a number of control variables.

**Activity Year**: The year in which the contests were conducted. A series of three year dummy variables are created.

**Visual Attractiveness of the Idea**: In addition to the four evaluation criteria (which are focusing on the content of an idea), the level of quality of the presentation of a submission is also incorporated as a control variable. The need to control for the effect of visual attractiveness is derived from research towards the importance of aesthetics for product attractiveness, as the level of visual appealiness may potentially contributed to the number of votes received by an idea and is therefore used as a control variable. In order to construct this variable, it is also integrated in the evaluation process together with the other four evaluation criteria. Appendix B presents positive Fleiss’ Kappa’s for the assessment of visual appealiness of the submitted ideas for each of the contests and years, which imply that there is almost perfect agreement among the raters. Consequently, the ratings of the raters were composed into a weighted average variable ‘AVG_PRES_Value’. The presentation value ranges between ‘1’ solutions of individuals that have done nothing to make the solution document look nice, and ‘5’ for visually very appealing ideas.

**Participant Group**: The final set of control variables that is added to this research is a set of dummy variables that capture the (random) group number of each participant. Different from the social dynamics variable which is created at the individual level (the voter who is in the same group with the idea submitter is coded as ‘1’, and ‘0’ if otherwise), this variable is created at the group level. This group level control variable is potentially important, as it may represent the heterogeneity among different groups, e.g.: symbolize the resources a group has (or is able to access).

**VoterNotInDistantNetwork**: as each student (participant) in the focal course can also follow a parallel course that is running during the same period as the focal course, students may develop their social contacts there as well, which may potentially influence our findings. This variable captures whether the voter to the innovative idea is not in the other course (if not, then coded as ‘1’, and ‘0’ otherwise).

**CandidateNotInDistantNetwork**: This variable captures whether the candidate/ author of the innovative idea is not in the other course (if not, then coded as ‘1’, and ‘0’ otherwise).

### Analysis Results

#### 5.1 Descriptive Statistics

Appendix A provides an overview of the variables that are included in this research, including a concise description of the variable codes, what is intended to measure by the variable, and the type of measurement that is used to capture the effect. On average, the students (N=244) received 4.070 votes for each of the solutions they submitted for the challenges. The maximum amount of votes received by an individual was 11. The composition of these votes follows the following distribution: 1.57 votes were received from voters who share the same group; 2.06 votes were received from voters with the same gender; 3.10 votes were received from voters who share the same ethnicity; 0.91 votes were re-
ceived from voters who did not share the same university of origin nor the same educational background; 0.08 votes were received from voters who had the same educational background but not the same university of origin; 1.93 votes were received from voters who only share the same university of origin; and 1.30 votes were from voters who share both the same educational background as well as the same university of origin.

Moreover, a Pearson correlation analysis was conducted to estimate whether there are exceptionally strong correlations between the variables included in this research. Taken together, these results indicate that social dynamics and social networks are positively related to the total number of votes that are received by an idea. This assumption is based on the positive correlations between these variables and the independent variable. The correlation analysis also shows that compared to the social dynamics factors, quality factors are less determining for whether an innovation is supported.

More specifically, the results indicate that individuals that share similar group membership are more likely to vote for one another. The number of votes is also positively related to similar gender and ethnicities. Subsequently, the results also indicate that the total number of votes is related to whether the voter did not share the same university of origin or same educational background; individuals who only share educational background; individuals who only share the same university of origin; and individuals who shared both the same educational background as the same university of origin. What is noticeable about these specific results is that individuals who do not share their educational background nor university of origin are also more likely to vote. This result indicates that some individuals do vote objectively. In addition, some correlation coefficients also show that some groups also influence the number of votes received by individuals.

5.2 (Exploratory) Social Network Analysis

I did some exploratory social network analysis to explore the data (in 2017). The analyses are conducted with the software NetDraw of Ucinet, and the results are shown below. Each node represents a participant in the crowd evaluation, a link (tie) between the two nodes means one voted for the other (following the direction of the tie). The participants who are in the same group are marked with the same color. It is clear to see the evolution of the voting networks and groups become more cohesive over the process, and it is noticeable that in the third round, a group is completely isolated from the crowd and the group members only voted for each other, which provides strong evidence for in-group activity even in the online setting. The visualization is shown on the next page, in Figure 1: Social Network Analysis of the Evolution of Voting Ties in a Crowd-sourcing Environment.
Figure 1: Social Network Analysis of the Evolution of Voting Ties in a Crowd-sourcing Environment
5.3 Analysis and Findings

The analysis results are in appendix. As the dependent variable is count variable (number of votes received per innovative idea in each round of the innovation contest in each year) which is overly dispersed, I use the less-efficient-but-more-accurate Poisson quasi maximum likelihood regression model in the data analyses. The PQML method is a more general family than the negative binomial regressions for overly-dispersed count data. In the robustness checks, we also used OLS and negative binomial regressions which give similar results.

The first hypothesis is about the difference between crowd-based evaluation and expert evaluation. The results are listed in Table 3. Surprisingly, and contradictory to what was initially supposed, the results show that none of the quality criteria (information clarity, originality, creativity, feasibility as well as visual attractiveness) seem to affect the number of votes received per innovative idea (Model 1), which indicates some other factors may at play in the process. Also as a result, Hypothesis 1 is supported. In Model 2, I tested the impact of the social factors on the number of votes per idea, the results show that, indeed there are strong relationships between Homophily and the popularity of an idea in a crowd. Among the factors, shared gender, shared ethnicity, as well as shared place (university) all seem to strongly influence the number of votes received per idea, but not shared background. Also, in general, if the candidate is not in the other course (distant network), the popularity of his/her innovative idea may be impaired. This is reasonable as visibility in a crowd may help with gaining popularity of an idea. In Model 3, I test the effect of in-group favoritism on the popularity (number of votes received) of an innovative idea. Interestingly, both immediate networks (in the same group in the focal course) and distant networks (in the same group in the parallel course) show strong relationship with the number of votes received per innovative idea, but not when they are combined (in the same group in both focal course and the parallel course). Hence, the co-opetition arguments may be helpful in understanding this situation. In Model 4, both the homophily variables as well as the in-group favoritism variables are added to the regression, the results stay very similar. Finally, in the last model (Model 5), all variables are added into the model. Interestingly, the effect of the quality criteria of an innovative idea now gets to be (partially) realized. Among the five quality criteria, the attractiveness of how an innovative idea is presented to the crowd does not play any role in predicting the number of votes received from the crowd. From the rest four major quality criteria, information clarity (how clear the information is described/conveyed to the crowd) of the innovative idea is the most significant in positively explaining the popularity of the innovative idea, followed by its originality. While whether the innovative idea is creative, and whether it is critically reflected upon, both do not seem to be important, which show little relevance to their influence on the preference of the crowd. The effect and significance level of both the social variables (e.g.: homophily) and the in-group favoritism variables, stay almost unchanged. Therefore, the results show that social networks and in-group favoritism indeed play important roles in determining the number of votes received per idea, rather than the intrinsic quality of the idea itself.
More specifically, Hypothesis 2 – 5 are about social dynamics in the crowd-based evaluation process. The results (Table 3) show that, gender, ethnical background, and university of origin all play important roles in determining the number of votes received by an innovative idea. Hence, the results support Hypothesis 2 that homophily indeed exists in online setting, albeit to varying extents. Also, whether the submitter is in the other course (“distant network”) or not plays a positive role in determining the number of votes the submitter received. However, whether the voter and the submitter share the same study background (thus, have more in-depth knowledge about the field) does not seem to affect the popularity of an idea. Regarding Hypothesis 3, results (Table 4) show that homophily and in-group favoritism indeed evolves with group tenure, however not as the hypothesized uni-directional, but more complex: Model 1-3 are with the variables denoting the intrinsic quality of an innovative idea, while Model 4-6 are purely focusing on the group dynamism in the crowd-sourcing context. It seems the longer the participants stay in the same group, the less influential the role social factor plays (decrease for both same gender and same ethnicity, but increase for same university). Therefore, contradictory to what is supposed, Hypothesis 3 is disconfirmed.

Regarding in-group favoritism, while I do not find strong support for any evolution of such patterns from contest round 1 to contest round 3, I do find participants’ resort to distant network (“In-Group Voting DistantNetwork”) becomes increasingly more salient over time along with group tenure. When differentiating between “immediate network” groups (groups in the course where the innovation contests are taking place, named as “FocalNetwork”) and “distant network” groups (groups in the other course, named as “DistantNetwork” which students attend during the same study period as the focal course), interestingly, it seems although there is in-group favoritism in both types of groups, it is groups that are formed in the distant networks (the parallel course), rather than in the immediate networks (the focal course) that show stronger in-group support. Therefore, Hypothesis 5 is supported.

Regarding the possible differentiating effect of in-group favoritism between different types of participants (here we focus on the “stars”—students who outperformed the others in the previous round of contest and the “laggards”—students who underperform the others in the previous round of contest, performance is compared with the medium grade in class, Hypothesis 4, results are shown in Table 5 and Table 6), interestingly, for the “stars”, regarding the quality criteria, it seems creativity plays a more important role in determining the popularity of their innovative idea than the others, while the criticality (feasibility) of the idea seem to even negatively impact the popularity of the idea. Students who outperform the average in the class, seem to resort more to the distant networks for support of their innovative ideas. While for the “laggards”, it is the more “superficial” criteria (information clarity and visual attractiveness) that affect how well their innovative ideas are received. Also, for this group of students, they appear to resort more to the support of students of the same ethnical background, as well as students who come from the same university. Interestingly, voters who have the same study background as the laggard submitters tend to rate the innovative idea of the laggards more negatively. Moreover, compared to their better-performing counterparts who mostly draw support from the distant networks (groups in the parallel course), the “laggards” tend to predominantly rely on in-group support from their immediate contacts (groups in the focal course where the innovation contests took place). Hence, Hypothesis 4 is partially supported.

Regarding organizational design and the distribution of decision making power to the crowd, Table 7 shows preliminary results by separating and comparing the results of year 2015 from year
In year 2015, 50% decision is centralized and 50% decision is distributed to the crowd; while in year 2016, 70% decision is centralized and 30% decision is distributed to the crowd. By changing the allocation of the weight of decision making power between the “top” (central management, in this case, the contests coordinator) and the crowd, I explore whether and to what extent the allocation of the weight of decision making power affects the social factors and in-group favoritism of crowd-based evaluation. The results suggest that the weight given to the crowd may indeed influence the extent of crowd members’ social (lobbying) activities. More specifically, the more weight is given to the crowd (in year 2015, 50% decision making power is distributed to the crowd, compared to only 30% in year 2016), in general the more likely the crowd members resort to support via the social network activities (e.g.: based on homophily, such as same ethnicity, same background, same university, except same-gender) to gain support for their ideas, instead of resorting to in-group activities (e.g.: based on in-group favoritism).

6. Conclusion and Discussion

This research is aimed to answer the research questions regarding the mechanisms in crowd-based evaluations. Overall, this research has contributed to the field of research concerning crowd-sourcing as a new mechanism for innovation, by investigating the social networks and social dynamism in the evaluation of the crowd-sourcing process. This social dynamics dimension is derived from social-network theory that has stated that individual behavior can be explained through network ties. The analysis shows that social dynamics factors do influence the outcome the crowd-based evaluation. More specifically, the findings of this research show that similar characteristics (homophily) between individuals makes them more likely to favor the innovations from one another. In addition, the results also show that group members tend to favor the innovations generated within their groups, albeit to varying extents.

In general, within the crowdsourcing context, the results show that individuals who share the same group membership are more likely to favor the solutions of one another. Thus, individuals within the same group are more likely to vote for an idea of group member. However, no statistically significant relation has been found for the positive development of the influence of shared group membership votes over time. Furthermore, the results show that the commonly received quality criteria (creativity, clarity, originality, and feasibility) of innovative ideas do not play significant roles in determining the popularity of the idea, but it is the social dynamics that are at play in the final outcome. More specifically, the results pointed out that individuals who share the same gender, the same ethnicity, as well as the same university of origin, are more likely to vote for one another’s solutions, which is not the case if they share the same study background (major). Finally, the results also show that individuals who are lagging behind in previous round of innovation contest display different social activities compared to their counterparts (the winners). Also, organizational design (the distribution and allocation of the decision making power between the central/top management and the crowd) is also likely to influence the social behavior in a crowd.

In addition, Feld (1981) argued that the connections between people not always arise from chance, but often arise from a central point of focus. These central points of focus actively connect people or passively constrain them to interact with one another, creating social ties (Feld, 1981;
Louch, 2000). The connection between loosely tied individuals even has the potential to generate new foci through the organization of new activities around the newly constituted network. Consequently, I hypothesize that individuals who share the same point of focus are more likely to create a preference for each other’s innovations. The findings of this research show consistency with these assumptions, as homophily (personal characteristics) is found to positively affect the preference of individuals. Individuals who share the same characteristics are found to prefer on another’s solutions. Furthermore, Marsden (1988) has found that the homogeneity in interpersonal networks is also triggered by personal characteristics of the involved individuals. It has been argued that individuals who share the same personal characteristics are more likely to connect with each other during social interaction. Overall, the findings of this research indicate that shared characteristics between individuals, makes it more likely to favor each other’s work, which still holds in online settings.

In conclusion, this research has proven that social dynamics factors are important determinants of the outcome of crowdsourcing. With this result the research contributes to prior research that has suggested that technological superiority alone does not provide a sufficient determinant for the outcome of the innovation process (Frost and Egri, 1989; Dougherty & Hardy, 1996). It has been argued that social dynamics processes within organizations also play a significant role in the outcome of innovations (Eisenhardt & Bourgeois, 1988; Frost & Egri, 1989; Maute & Locander, 1994; Jones & Stevens, 1999; Howard-Grenville, 2007; Kaplan 2008; Monin et al., 2013). Thus, these results are consistent with the findings of prior research and extend the current field of knowledge by researching it in an online setting, as well as identifying what micro-processes play in the crowdsourcing process. These results have implications for both theory and practices. This research adds to the social network theory by extending its boundaries to the virtual, online, and “blind-review” settings, showing that although being anonymous, social factors such as homogeneity still plays a crucial role in deciding the popularity of a submission, which, in turn, weakens the assumption that the seemingly neutral and trustworthy crowd may be better evaluators than the centralized decision makers in innovation. Organizers of crowd evaluation should take this into account when designing the crowding/evaluation event. On the other hand, for idea submitters, it is the originality and clarity of information that help the idea to gain more popularity among their potential voters, instead of its creativity and how appealing the idea may appear.

Furthermore, prior research has also pointed out that these social dynamics factors might constrain innovation, arguing that innovation needs a certain organizational power basis to get supported (Dougherty, 1996; Jones & Stevens 1999). The results provided by this research are consistent with this assumption, as the quality criteria that are determined by businesses themselves is sometimes found neglected in a tournament-based crowdsourcing environment. In this context, the assessment of voters for determining what solution to vote for is based on personal and group characteristics rather than the quality criteria. It has also been suggested that different parties try to alter the outcome of the innovation process by engaging in a political process, by building coalitions and through negotiating (Dougherty, 1996; Jones & Stevens 1999). This research shows that individuals who share the same group membership are more likely to favor one another’s innovations. This group favoritism can be seen as a synonym for the coalitions that are formed within the organizational context. These situations could cause high quality innovations to be neglected over the ones that have more political support.
6.1 Avenues for Future Research

The deployment of crowdsourcing mechanism is becoming increasingly important as more businesses use these mechanisms as a source of innovation (Dougherty, 2003; Howe, 2006; Afuah and Tucci, 2012; Poetz and Schreier, 2012). However, there is still need to further elaborate on how social constructs and politics within businesses and between individuals determine the outcome of the crowdsourcing process.

Future studies are invited to address the problem of external validity that is encountered by this research. By repeating this experiment with a more diverse crowd in which individuals are not yet socially tied biases can be avoided. One way to attract a more diverse crowd is by extending the crowdsourcing environment in a company or to the World Wide Web. This has become increasingly possible, as the Internet enables us to reach a more distant and diverse crowd without enduring major costs (Fuchs and Schreier, 2011; Poets and Schreier, 2012; Afuah and Tucci, 2012). Wooten and Ulrich (2017) further elaborate on this approach by arguing that future research should actively seek out involvement of different industries.

Alternatively, future research could also further elaborate on the social dynamics that are involved in the crowd-based evaluation process. Jeppesen and Federiksen (2006) have argued that innovative users are often found to be hobbyists that are eager to share information. Prior research has suggested that potential customer benefits and customer experience are key determinants of whether crowds are willing to engage in the crowd-based evaluation process (Scheier & Prügl, 2008; Franke et al., 2012; Barry, 2013). Furthermore, future research could also examine how the crowd engages in a political battle in order to influence each other to gain more support for their innovation. Due to data limitations this research was only able to sketch part of the social dynamics. Hence, future research should look more specifically into what personal characteristics play a role in the crowd-based evaluation process.

On the other hand, future research might also explore to what extent the biases that are created by shared characteristics cause the neglecting of high quality solutions. In other words, is the problem really solved by the solution with the highest quality or do other dynamics in play a role. Alternatively to the research of Poetz and Schreier (2012), another direction for future research might be to investigate whether the business or the crowd is more effective to determine what ideas are best. Blohm, Leismeister and Helmut (2013) have provided some research in this direction. Their research focuses on how businesses can cope with the large amount of ideas that are generated by crowds. The research concludes with a set of challenge and coping strategies to deal with the increasing amount of data. However, more research on this topic is needed in order to gain a better understanding of the trade-offs between internal sourcing and outsourcing the selection of ideas.
## Appendix A: Variables Descriptions

### Table 1. Variables Definitions, Measurements and Categories

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Measure</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. TotalVotes</td>
<td>The total number of votes received per innovative idea in a certain round of contest, minus the self-vote.</td>
<td>Count</td>
</tr>
<tr>
<td>2. AVG_INFO</td>
<td>Average value for information clarity of the innovative idea designated by the three independent raters.</td>
<td>Continuous</td>
</tr>
<tr>
<td>3. AVGORIGINAL</td>
<td>Average value for originality of the innovative idea designated by the three independent raters.</td>
<td>Continuous</td>
</tr>
<tr>
<td>4. AVG_CREATIVE</td>
<td>Average value for creativity of the innovative idea designated by the three independent raters.</td>
<td>Continuous</td>
</tr>
<tr>
<td>5. AVG_CRITICAL</td>
<td>Average value for criticality/ feasibility of the innovative idea designated by the three independent raters.</td>
<td>Continuous</td>
</tr>
<tr>
<td>6. AVG_PRESENTATION</td>
<td>Average value for visual attractiveness of the innovative idea designated by the three independent raters.</td>
<td>Continuous</td>
</tr>
<tr>
<td>7. SameGenderVoting</td>
<td>The total number of votes received from someone who has the same gender as the individual, minus the self-vote.</td>
<td>Count</td>
</tr>
<tr>
<td>8. SameEthnicityVoting</td>
<td>The total number of votes received from someone who has the same ethnicity as the individual, minus the self-vote.</td>
<td>Count</td>
</tr>
<tr>
<td>9. SameBackgroundVoting</td>
<td>The total number of votes received from someone who has the same study major as the individual, minus the self-vote.</td>
<td>Count</td>
</tr>
<tr>
<td>10. SameUniversityVoting</td>
<td>The total number of votes received from someone who comes from the same university as the individual, minus the self-vote.</td>
<td>Count</td>
</tr>
<tr>
<td>11. InGroupVoting_FocalNetwork</td>
<td>The total number of votes received from someone who is in the same group in the focal course (where the innovation contests took place) as the individual, minus the self-vote.</td>
<td>Count</td>
</tr>
<tr>
<td>12. InGroupVoting_DistantNetwork</td>
<td>The total number of votes received from someone who is in the same group in the parallel course (which runs in parallel with the focal course) as the individual, minus the self-vote.</td>
<td>Count</td>
</tr>
<tr>
<td>13. InGroupVoting_Both_Focal_Distant</td>
<td>The total number of votes received from someone who is in the same group both in the focal course as well as in the parallel course as the individual, minus the self-vote.</td>
<td>Count</td>
</tr>
<tr>
<td>14. VoterNotInDistantNetwork</td>
<td>The voter is not in the other course that is running in parallel with the focal course (where the innovation contests took place).</td>
<td>Dummy (0/1)</td>
</tr>
<tr>
<td>15. CandidateNotInDistantNetwork</td>
<td>The candidate is not in the other course that is running in parallel with the focal course (where the innovation contests took place).</td>
<td>Dummy (0/1)</td>
</tr>
<tr>
<td>16. Year_Dummies</td>
<td>Year dummies, denoting in which year the contests took place (2015, 2016, 2017)</td>
<td>Dummies</td>
</tr>
<tr>
<td>18. # of Innovation Challenge</td>
<td>The number of the round of innovation challenge (contest), each year three rounds of innovation contests took place</td>
<td>Dummies</td>
</tr>
</tbody>
</table>
### Appendix B: Fleiss’ Kappa score per year for each category

**Table 2. Fleiss’ Kappa score per year for each category**

<table>
<thead>
<tr>
<th>Year</th>
<th>Rating category</th>
<th>Challenge 1 $k$</th>
<th>Challenge 2 $k$</th>
<th>Challenge 3 $k$</th>
<th>Total $k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>Information clarity</td>
<td>.642</td>
<td>.483</td>
<td>.784</td>
<td>.647</td>
</tr>
<tr>
<td></td>
<td>Originality quality</td>
<td>.592</td>
<td>.549</td>
<td>.592</td>
<td>.592</td>
</tr>
<tr>
<td></td>
<td>Creativity quality</td>
<td>.702</td>
<td>.624</td>
<td>.629</td>
<td>.668</td>
</tr>
<tr>
<td></td>
<td>Critical thinking quality</td>
<td>.713</td>
<td>.656</td>
<td>.535</td>
<td>.654</td>
</tr>
<tr>
<td></td>
<td>Presentation quality</td>
<td>.511</td>
<td>.843</td>
<td>.685</td>
<td>.689</td>
</tr>
<tr>
<td>2016</td>
<td>Information clarity</td>
<td>.732</td>
<td>.687</td>
<td>.767</td>
<td>.744</td>
</tr>
<tr>
<td></td>
<td>Originality quality</td>
<td>.492</td>
<td>.660</td>
<td>.552</td>
<td>.570</td>
</tr>
<tr>
<td></td>
<td>Creativity quality</td>
<td>.760</td>
<td>.509</td>
<td>.569</td>
<td>.624</td>
</tr>
<tr>
<td></td>
<td>Critical thinking quality</td>
<td>.466</td>
<td>.628</td>
<td>.489</td>
<td>.532</td>
</tr>
<tr>
<td></td>
<td>Presentation quality</td>
<td>.753</td>
<td>.765</td>
<td>.908</td>
<td>.802</td>
</tr>
<tr>
<td>VARIABLES</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>avg_info</td>
<td>0.0250</td>
<td></td>
<td></td>
<td>0.0689**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0405)</td>
<td></td>
<td></td>
<td>(-0.0309)</td>
<td></td>
</tr>
<tr>
<td>avg_origial</td>
<td>0.0545</td>
<td></td>
<td>0.0563**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0447)</td>
<td></td>
<td>(-0.0253)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>avg_creative</td>
<td>0.0595</td>
<td></td>
<td>-0.00754</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0476)</td>
<td></td>
<td>(-0.0288)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>avg_critical</td>
<td>0.0508</td>
<td></td>
<td>-0.00114</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0378)</td>
<td></td>
<td>(-0.0243)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>avg_presentation</td>
<td>-0.00880</td>
<td></td>
<td></td>
<td>0.00433</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0332)</td>
<td></td>
<td></td>
<td>(-0.0202)</td>
<td></td>
</tr>
<tr>
<td>SameGenderVoting</td>
<td></td>
<td>0.108***</td>
<td></td>
<td>0.105***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.0173)</td>
<td></td>
<td>(-0.0178)</td>
<td></td>
</tr>
<tr>
<td>SameEthnicityVoting</td>
<td>0.0906***</td>
<td></td>
<td>0.0675***</td>
<td></td>
<td>0.0715***</td>
</tr>
<tr>
<td></td>
<td>(-0.0171)</td>
<td></td>
<td>(-0.0163)</td>
<td></td>
<td>(-0.0164)</td>
</tr>
<tr>
<td>SameBackgroundVoting</td>
<td>-0.0169</td>
<td>-0.0176</td>
<td>-0.00176</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.0151)</td>
<td></td>
<td>(-0.0137)</td>
<td></td>
<td>(-0.0164)</td>
</tr>
<tr>
<td>SameUniversityVoting</td>
<td>0.0564***</td>
<td></td>
<td>0.0535***</td>
<td></td>
<td>0.0484***</td>
</tr>
<tr>
<td></td>
<td>(-0.0172)</td>
<td></td>
<td>(-0.0148)</td>
<td></td>
<td>(-0.0172)</td>
</tr>
<tr>
<td>InGroupVoting_FocalNetwork</td>
<td></td>
<td>0.123***</td>
<td>0.0411**</td>
<td>0.0468**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.0187)</td>
<td></td>
<td>(-0.0174)</td>
<td></td>
</tr>
<tr>
<td>InGroupVoting_DistantNetwork</td>
<td></td>
<td>0.115***</td>
<td>0.0581***</td>
<td></td>
<td>0.0542***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.017)</td>
<td></td>
<td>(-0.016)</td>
<td></td>
</tr>
<tr>
<td>InGroupVoting_Both_Focal_Distant</td>
<td></td>
<td>-0.199***</td>
<td>-0.144***</td>
<td></td>
<td>-0.0562</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-0.0471)</td>
<td></td>
</tr>
<tr>
<td>VoterNotInDistantNetwork</td>
<td>0.0561</td>
<td>0.00758</td>
<td>0.0372</td>
<td>0.012</td>
<td>-0.00415</td>
</tr>
<tr>
<td></td>
<td>(0.0343)</td>
<td>(-0.0236)</td>
<td>(-0.024)</td>
<td>(-0.022)</td>
<td>(-0.025)</td>
</tr>
<tr>
<td>CandidateNotInDistantNetwork</td>
<td>-0.145</td>
<td>-0.145***</td>
<td>-0.219***</td>
<td>-0.192***</td>
<td>-0.190**</td>
</tr>
<tr>
<td></td>
<td>(0.0901)</td>
<td>(-0.0545)</td>
<td>(-0.0529)</td>
<td>(-0.0598)</td>
<td>(-0.0923)</td>
</tr>
<tr>
<td>Year_dummies</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>InnovationChallenge_1</td>
<td>0.0751</td>
<td>-0.123***</td>
<td>0.110**</td>
<td>-0.0665**</td>
<td>-0.0608</td>
</tr>
<tr>
<td></td>
<td>(0.0569)</td>
<td>(-0.0333)</td>
<td>(-0.0478)</td>
<td>(-0.0341)</td>
<td>(-0.0437)</td>
</tr>
<tr>
<td>InnovationChallenge_2</td>
<td>0.0515</td>
<td>-0.0303</td>
<td>0.0219</td>
<td>-0.029</td>
<td>-0.0127</td>
</tr>
<tr>
<td></td>
<td>(0.0555)</td>
<td>(-0.0254)</td>
<td>(-0.0309)</td>
<td>(-0.0248)</td>
<td>(-0.0334)</td>
</tr>
<tr>
<td></td>
<td>1.169***</td>
<td>0.991***</td>
<td>1.065***</td>
<td>0.921***</td>
<td>0.739***</td>
</tr>
<tr>
<td>----------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>(0.222)</td>
<td>(-0.0439)</td>
<td>(-0.0551)</td>
<td>(-0.0491)</td>
<td>(-0.136)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,099</td>
<td>1,677</td>
<td>1,677</td>
<td>1,677</td>
<td>1,099</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-2262</td>
<td>-3106</td>
<td>-3232</td>
<td>-3083</td>
<td>-2053</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses:

*** p<0.01, ** p<0.05, * p<0.1
<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TotalVotes</td>
<td>TotalVotes</td>
<td>TotalVotes</td>
<td>TotalVotes</td>
<td>TotalVotes</td>
<td>TotalVotes</td>
</tr>
<tr>
<td></td>
<td>Round 1</td>
<td>Round 2</td>
<td>Round 3</td>
<td>Round 1</td>
<td>Round 2</td>
<td>Round 3</td>
</tr>
<tr>
<td>avg_info</td>
<td>0.135***</td>
<td>0.0480</td>
<td>0.0694**</td>
<td>0.108***</td>
<td>0.105***</td>
<td>0.0809***</td>
</tr>
<tr>
<td></td>
<td>(0.0519)</td>
<td>(0.0409)</td>
<td>(0.0392)</td>
<td>(0.0264)</td>
<td>(0.0210)</td>
<td>(0.0166)</td>
</tr>
<tr>
<td>avg_origial</td>
<td>0.0896**</td>
<td>0.0373</td>
<td>0.00562</td>
<td>0.109***</td>
<td>0.0594***</td>
<td>0.0246</td>
</tr>
<tr>
<td></td>
<td>(0.0442)</td>
<td>(0.0322)</td>
<td>(0.0526)</td>
<td>(0.0323)</td>
<td>(0.0195)</td>
<td>(0.0202)</td>
</tr>
<tr>
<td>avg_creative</td>
<td>-0.0465</td>
<td>-0.0350</td>
<td>0.0153</td>
<td>0.00555</td>
<td>-0.0243</td>
<td>-0.0265</td>
</tr>
<tr>
<td></td>
<td>(0.0577)</td>
<td>(0.0523)</td>
<td>(0.0382)</td>
<td>(0.0283)</td>
<td>(0.0160)</td>
<td>(0.0201)</td>
</tr>
<tr>
<td>avg_critical</td>
<td>0.0355</td>
<td>-0.0208</td>
<td>-0.0303</td>
<td>0.0281</td>
<td>0.0416**</td>
<td>0.0475*</td>
</tr>
<tr>
<td></td>
<td>(0.0550)</td>
<td>(0.0329)</td>
<td>(0.0363)</td>
<td>(0.0325)</td>
<td>(0.0199)</td>
<td>(0.0268)</td>
</tr>
<tr>
<td>avg_presentation</td>
<td>0.00845</td>
<td>0.0594*</td>
<td>0.000964</td>
<td>0.0237</td>
<td>0.0574***</td>
<td>-0.116**</td>
</tr>
<tr>
<td></td>
<td>(0.0377)</td>
<td>(0.0325)</td>
<td>(0.0241)</td>
<td>(0.0325)</td>
<td>(0.0199)</td>
<td>(0.0268)</td>
</tr>
<tr>
<td>SameGenderVoting</td>
<td>0.0865***</td>
<td>0.124***</td>
<td>0.0766***</td>
<td>0.108***</td>
<td>0.105***</td>
<td>0.0809***</td>
</tr>
<tr>
<td></td>
<td>(0.0267)</td>
<td>(0.0224)</td>
<td>(0.0182)</td>
<td>(0.0264)</td>
<td>(0.0210)</td>
<td>(0.0166)</td>
</tr>
<tr>
<td>SameEthnicityVoting</td>
<td>0.140***</td>
<td>0.0577***</td>
<td>0.0137</td>
<td>0.109***</td>
<td>0.0594***</td>
<td>0.0246</td>
</tr>
<tr>
<td></td>
<td>(0.0290)</td>
<td>(0.0215)</td>
<td>(0.0184)</td>
<td>(0.0323)</td>
<td>(0.0195)</td>
<td>(0.0202)</td>
</tr>
<tr>
<td>SameBackgroundVoting</td>
<td>0.00771</td>
<td>0.00797</td>
<td>-0.00889</td>
<td>0.00555</td>
<td>-0.0243</td>
<td>-0.0265</td>
</tr>
<tr>
<td></td>
<td>(0.0351)</td>
<td>(0.0231)</td>
<td>(0.0193)</td>
<td>(0.0283)</td>
<td>(0.0160)</td>
<td>(0.0201)</td>
</tr>
<tr>
<td>SameUniversityVoting</td>
<td>0.0209</td>
<td>0.0387</td>
<td>0.0803***</td>
<td>0.0281</td>
<td>0.0416**</td>
<td>0.0475*</td>
</tr>
<tr>
<td></td>
<td>(0.0373)</td>
<td>(0.0276)</td>
<td>(0.0269)</td>
<td>(0.0325)</td>
<td>(0.0199)</td>
<td>(0.0268)</td>
</tr>
<tr>
<td>InGroupVoting_FocalNetwork</td>
<td>0.0261</td>
<td>0.0415</td>
<td>0.0824**</td>
<td>0.0237</td>
<td>0.0574***</td>
<td>0.0401</td>
</tr>
<tr>
<td></td>
<td>(0.0309)</td>
<td>(0.0276)</td>
<td>(0.0365)</td>
<td>(0.0297)</td>
<td>(0.0206)</td>
<td>(0.0274)</td>
</tr>
<tr>
<td>InGroupVoting_DistantNetwork</td>
<td>-0.0201</td>
<td>0.0888***</td>
<td>0.106***</td>
<td>0.00334</td>
<td>0.0716***</td>
<td>0.117***</td>
</tr>
<tr>
<td></td>
<td>(0.0313)</td>
<td>(0.0296)</td>
<td>(0.0287)</td>
<td>(0.0276)</td>
<td>(0.0229)</td>
<td>(0.0259)</td>
</tr>
<tr>
<td>In_GroupVoting_Both_Focal_Distant</td>
<td>-0.0783</td>
<td>0.00926</td>
<td>0.0233</td>
<td>-0.163*</td>
<td>-0.116**</td>
<td>-0.112*</td>
</tr>
<tr>
<td></td>
<td>(0.103)</td>
<td>(0.0795)</td>
<td>(0.0704)</td>
<td>(0.0884)</td>
<td>(0.0563)</td>
<td>(0.0616)</td>
</tr>
<tr>
<td>VoterNotInDistantNetwork</td>
<td>0.0275</td>
<td>-0.00384</td>
<td>-0.0136</td>
<td>0.0266</td>
<td>0.0144</td>
<td>-0.0167</td>
</tr>
<tr>
<td></td>
<td>(0.0272)</td>
<td>(0.0337)</td>
<td>(0.0372)</td>
<td>(0.0312)</td>
<td>(0.0281)</td>
<td>(0.0319)</td>
</tr>
<tr>
<td>CandidateNotInDistantNetwork</td>
<td>-0.160</td>
<td>-0.125</td>
<td>-0.136</td>
<td>-0.213**</td>
<td>-0.131*</td>
<td>-0.187***</td>
</tr>
<tr>
<td></td>
<td>(0.145)</td>
<td>(0.104)</td>
<td>(0.0920)</td>
<td>(0.108)</td>
<td>(0.0686)</td>
<td>(0.0710)</td>
</tr>
<tr>
<td>Year_dummies</td>
<td>-0.138</td>
<td>0.0849</td>
<td>0.0857</td>
<td>-0.0407</td>
<td>0.163**</td>
<td>0.232***</td>
</tr>
<tr>
<td></td>
<td>(0.0947)</td>
<td>(0.0915)</td>
<td>(0.0861)</td>
<td>(0.0894)</td>
<td>(0.0732)</td>
<td>(0.0682)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.471**</td>
<td>0.662***</td>
<td>0.841***</td>
<td>0.890***</td>
<td>0.852***</td>
<td>0.966***</td>
</tr>
</tbody>
</table>

Table 4. Poisson Quasi Maximum Likelihood Regression – Evolution with Group Tenure.
<table>
<thead>
<tr>
<th></th>
<th>(0.225)</th>
<th>(0.159)</th>
<th>(0.198)</th>
<th>(0.0688)</th>
<th>(0.0751)</th>
<th>(0.0818)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>356</td>
<td>378</td>
<td>365</td>
<td>531</td>
<td>592</td>
<td>554</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-673.2</td>
<td>-691.7</td>
<td>-662.8</td>
<td>-982</td>
<td>-1082</td>
<td>-1001</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses:

*** p<0.01, ** p<0.05, * p<0.1
Table 5. Poisson Quasi Maximum Likelihood Regression  
– Split Sample (Better Performers- Grades higher than Medium)

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg_info</td>
<td>-0.000418</td>
<td>0.000693</td>
<td>0.00732</td>
<td>0.00410</td>
<td>0.00636</td>
<td>0.00441</td>
</tr>
<tr>
<td></td>
<td>(0.0418)</td>
<td>(0.0452)</td>
<td>(0.0423)</td>
<td>(0.0336)</td>
<td>(0.0423)</td>
<td>(0.0323)</td>
</tr>
<tr>
<td>avg_origial</td>
<td>0.0120</td>
<td>0.0446</td>
<td>0.0490</td>
<td>0.0476</td>
<td>0.0343</td>
<td>0.0634*</td>
</tr>
<tr>
<td></td>
<td>(0.0456)</td>
<td>(0.0406)</td>
<td>(0.0367)</td>
<td>(0.0343)</td>
<td>(0.0388)</td>
<td>(0.0330)</td>
</tr>
<tr>
<td>avg_creative</td>
<td>0.146***</td>
<td>0.0615</td>
<td>0.0609</td>
<td>0.0502</td>
<td>0.0615</td>
<td>0.0465</td>
</tr>
<tr>
<td></td>
<td>(0.0543)</td>
<td>(0.0420)</td>
<td>(0.0431)</td>
<td>(0.0325)</td>
<td>(0.0417)</td>
<td>(0.0322)</td>
</tr>
<tr>
<td>avg_critical</td>
<td>-0.0958**</td>
<td>-0.0317</td>
<td>-0.0608**</td>
<td>0.00313</td>
<td>-0.0255</td>
<td>-0.0274</td>
</tr>
<tr>
<td></td>
<td>(0.0372)</td>
<td>(0.0264)</td>
<td>(0.0300)</td>
<td>(0.0252)</td>
<td>(0.0302)</td>
<td>(0.0323)</td>
</tr>
<tr>
<td>avg_presentation</td>
<td>0.0128</td>
<td>-0.0155</td>
<td>-0.0159</td>
<td>-0.00214</td>
<td>-0.00528</td>
<td>-0.0141</td>
</tr>
<tr>
<td></td>
<td>(0.0316)</td>
<td>(0.0303)</td>
<td>(0.0248)</td>
<td>(0.0225)</td>
<td>(0.0308)</td>
<td>(0.0195)</td>
</tr>
<tr>
<td>SameGenderVoting</td>
<td>0.106***</td>
<td>0.0636**</td>
<td>0.119***</td>
<td>0.105***</td>
<td>0.0863***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0243)</td>
<td>(0.0253)</td>
<td>(0.0166)</td>
<td>(0.0247)</td>
<td>(0.0198)</td>
<td></td>
</tr>
<tr>
<td>SameEthnicityVoting</td>
<td>0.0613***</td>
<td>0.0429**</td>
<td>0.0302</td>
<td>0.0597***</td>
<td>0.0189</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0188)</td>
<td>(0.0181)</td>
<td>(0.0231)</td>
<td>(0.0196)</td>
<td>(0.0206)</td>
<td></td>
</tr>
<tr>
<td>SameBackgroundVoting</td>
<td>0.0425</td>
<td>0.0401</td>
<td>0.0144</td>
<td>0.0474</td>
<td>0.00654</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0335)</td>
<td>(0.0373)</td>
<td>(0.0281)</td>
<td>(0.0347)</td>
<td>(0.0306)</td>
<td></td>
</tr>
<tr>
<td>SameUniversityVoting</td>
<td>0.0158</td>
<td>0.0273</td>
<td>-0.000211</td>
<td>0.0162</td>
<td>0.00680</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0502)</td>
<td>(0.0511)</td>
<td>(0.0374)</td>
<td>(0.0506)</td>
<td>(0.0361)</td>
<td></td>
</tr>
<tr>
<td>InGroupVoting_FocalNetwork</td>
<td>0.108***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0330)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>InGroupVoting_DistantNetwork</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.145***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0281)</td>
</tr>
<tr>
<td>InGroupVoting_Both_Focal_Distant</td>
<td>0.0666</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0788</td>
</tr>
<tr>
<td></td>
<td>(0.0664)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0543)</td>
</tr>
<tr>
<td>VoterNotInDistantNetwork</td>
<td>-0.0437</td>
<td>-0.0177</td>
<td>-0.0353</td>
<td>-0.0158</td>
<td>-0.0182</td>
<td>-0.0270</td>
</tr>
<tr>
<td></td>
<td>(0.0407)</td>
<td>(0.0356)</td>
<td>(0.0348)</td>
<td>(0.0263)</td>
<td>(0.0360)</td>
<td>(0.0239)</td>
</tr>
<tr>
<td>CandidateNotInDistantNetwork</td>
<td>-0.0165</td>
<td>-0.229***</td>
<td>-0.180**</td>
<td>-0.133*</td>
<td>-0.207**</td>
<td>-0.124</td>
</tr>
<tr>
<td></td>
<td>(0.112)</td>
<td>(0.0740)</td>
<td>(0.0786)</td>
<td>(0.0699)</td>
<td>(0.0817)</td>
<td>(0.0826)</td>
</tr>
<tr>
<td>Year_dummies</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>InnovationChallenge_dummies</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.426</td>
<td>0.102</td>
<td>0.121</td>
<td>0.177</td>
<td>0.0840</td>
<td>0.212</td>
</tr>
<tr>
<td></td>
<td>(0.739)</td>
<td>(0.720)</td>
<td>(0.610)</td>
<td>(0.544)</td>
<td>(0.725)</td>
<td>(0.484)</td>
</tr>
<tr>
<td>Observations</td>
<td>387</td>
<td>387</td>
<td>387</td>
<td>387</td>
<td>387</td>
<td>387</td>
</tr>
<tr>
<td>--------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-752.5</td>
<td>-724</td>
<td>-718.7</td>
<td>-708.8</td>
<td>-723.5</td>
<td>-705.2</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses:
*** p<0.01, ** p<0.05, * p<0.1
<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TotalVotes</td>
<td>0.0492</td>
<td>0.132***</td>
<td>0.141***</td>
<td>0.130***</td>
<td>0.132***</td>
<td>0.137***</td>
</tr>
<tr>
<td></td>
<td>(0.0468)</td>
<td>(0.0383 )</td>
<td>(0.0363 )</td>
<td>(0.0387 )</td>
<td>(0.0383 )</td>
<td>(0.0365 )</td>
</tr>
<tr>
<td>TotalVotes</td>
<td>-0.00599</td>
<td>-0.0229</td>
<td>-0.0465</td>
<td>-0.0386</td>
<td>-0.0509</td>
<td>-0.0648</td>
</tr>
<tr>
<td></td>
<td>(0.0822)</td>
<td>(0.0571 )</td>
<td>(0.0562 )</td>
<td>(0.0562 )</td>
<td>(0.0539 )</td>
<td>(0.0525 )</td>
</tr>
<tr>
<td>TotalVotes</td>
<td>0.0164</td>
<td>0.0511</td>
<td>0.0275</td>
<td>0.0471</td>
<td>0.0516</td>
<td>0.0343</td>
</tr>
<tr>
<td></td>
<td>(0.0675)</td>
<td>(0.0409 )</td>
<td>(0.0444 )</td>
<td>(0.0414 )</td>
<td>(0.0422 )</td>
<td>(0.0449 )</td>
</tr>
<tr>
<td>SameGenderVoting</td>
<td>0.118***</td>
<td>0.105***</td>
<td>0.111***</td>
<td>0.114***</td>
<td>0.104***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0261)</td>
<td>(0.0254 )</td>
<td>(0.0251 )</td>
<td>(0.0243 )</td>
<td>(0.0240 )</td>
<td></td>
</tr>
<tr>
<td>SameEthnicityVoting</td>
<td>0.0975***</td>
<td>0.0877***</td>
<td>0.0834***</td>
<td>0.0960***</td>
<td>0.0851***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0258)</td>
<td>(0.0234 )</td>
<td>(0.0274 )</td>
<td>(0.0240 )</td>
<td>(0.0264 )</td>
<td></td>
</tr>
<tr>
<td>SameBackgroundVoting</td>
<td>-0.0464*</td>
<td>-0.0484**</td>
<td>-0.0440**</td>
<td>-0.0369</td>
<td>-0.0397*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0237)</td>
<td>(0.0238 )</td>
<td>(0.0221 )</td>
<td>(0.0240 )</td>
<td>(0.0238 )</td>
<td></td>
</tr>
<tr>
<td>SameUniversityVoting</td>
<td>0.0987***</td>
<td>0.0998***</td>
<td>0.102***</td>
<td>0.0968***</td>
<td>0.0988***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0345)</td>
<td>(0.0352 )</td>
<td>(0.0352 )</td>
<td>(0.0340 )</td>
<td>(0.0353 )</td>
<td></td>
</tr>
<tr>
<td>InGroupVoting_FocalNetwork</td>
<td>0.0671**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0443</td>
</tr>
<tr>
<td></td>
<td>(0.0287)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0293 )</td>
</tr>
<tr>
<td>InGroupVoting_DistantNetwork</td>
<td></td>
<td>0.0405*</td>
<td></td>
<td></td>
<td></td>
<td>0.0148</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0238)</td>
<td></td>
<td></td>
<td></td>
<td>(0.0250 )</td>
</tr>
<tr>
<td>InGroupVoting_Both_Focal_Dista nt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.187***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0721)</td>
</tr>
<tr>
<td>VoterNotInDistantNetwork</td>
<td>0.103*</td>
<td>-0.0115</td>
<td>-0.0121</td>
<td>-0.0118</td>
<td>-0.00141</td>
<td>-0.00479</td>
</tr>
<tr>
<td></td>
<td>(0.0571)</td>
<td>(0.0421 )</td>
<td>(0.0380 )</td>
<td>(0.0412 )</td>
<td>(0.0358 )</td>
<td>(0.0339 )</td>
</tr>
<tr>
<td>CandidateNotInDistantNetwork</td>
<td>-0.140</td>
<td>-0.133</td>
<td>-0.166</td>
<td>-0.0986</td>
<td>-0.0672</td>
<td>-0.0946</td>
</tr>
<tr>
<td></td>
<td>(0.120)</td>
<td>(0.124 )</td>
<td>(0.115 )</td>
<td>(0.125 )</td>
<td>(0.116 )</td>
<td>(0.117 )</td>
</tr>
<tr>
<td>Year_dummies</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td></td>
</tr>
<tr>
<td>InnovationChallenge_dummies</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.484</td>
<td>1.116*</td>
<td>1.270*</td>
<td>1.254*</td>
<td>1.324**</td>
<td>1.424**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.721)</td>
<td></td>
<td>(0.765 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6. Poisson Quasi Maximum Likelihood Regression – Split Sample (Laggards - Grades lower than Medium)

<table>
<thead>
<tr>
<th></th>
<th>Observations</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>336</td>
<td>336</td>
<td>336</td>
<td>336</td>
<td>336</td>
<td>336</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-657.6</td>
<td>-609.3</td>
<td>-607.1</td>
<td>-608.2</td>
<td>-606.3</td>
<td>-605.3</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses:

*** p<0.01, ** p<0.05, * p<0.1
Table 7. Poisson Quasi Maximum Likelihood Regression – Org. Design

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SameGenderVoting</td>
<td>0.108***</td>
<td>0.0992***</td>
<td>0.0989***</td>
<td>0.137***</td>
<td>0.133***</td>
<td>0.128***</td>
</tr>
<tr>
<td></td>
<td>(0.0206)</td>
<td>(0.0194)</td>
<td>(0.0191)</td>
<td>(0.0285)</td>
<td>(0.0260)</td>
<td>(0.0250)</td>
</tr>
<tr>
<td>SameEthnicityVoting</td>
<td>0.0881***</td>
<td>0.0787***</td>
<td>0.0779***</td>
<td>0.0773***</td>
<td>0.0483*</td>
<td>0.0578**</td>
</tr>
<tr>
<td></td>
<td>(0.0252)</td>
<td>(0.0244)</td>
<td>(0.0250)</td>
<td>(0.0243)</td>
<td>(0.0271)</td>
<td>(0.0245)</td>
</tr>
<tr>
<td>SameBackgroundVoting</td>
<td>0.0307**</td>
<td>0.0291**</td>
<td>0.0285**</td>
<td>-0.0399</td>
<td>-0.0564</td>
<td>-0.0332</td>
</tr>
<tr>
<td></td>
<td>(0.0138)</td>
<td>(0.0121)</td>
<td>(0.0123)</td>
<td>(0.0332)</td>
<td>(0.0381)</td>
<td>(0.0391)</td>
</tr>
<tr>
<td>SameUniversityVoting</td>
<td>0.0228</td>
<td>0.0270*</td>
<td>0.0280*</td>
<td>-0.0793</td>
<td>0.0635</td>
<td>0.0465</td>
</tr>
<tr>
<td></td>
<td>(0.0179)</td>
<td>(0.0140)</td>
<td>(0.0152)</td>
<td>(0.0496)</td>
<td>(0.0506)</td>
<td>(0.0522)</td>
</tr>
<tr>
<td>VoterNotInFBA</td>
<td>0.00124</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0243)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CandidateNotInFBA</td>
<td>-0.0369</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.223*</td>
</tr>
<tr>
<td></td>
<td>(0.0804)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.118)</td>
</tr>
<tr>
<td>InGroupVoting_FocalNetwork</td>
<td>0.0405</td>
<td>0.0409</td>
<td></td>
<td>0.0242</td>
<td></td>
<td>0.0353</td>
</tr>
<tr>
<td></td>
<td>(0.0308)</td>
<td>(0.0309)</td>
<td></td>
<td>(0.0350)</td>
<td></td>
<td>(0.0326)</td>
</tr>
<tr>
<td>InGroupVoting_DistantNetwork</td>
<td>0.0641</td>
<td>0.0623</td>
<td></td>
<td>0.0690**</td>
<td></td>
<td>0.0617**</td>
</tr>
<tr>
<td></td>
<td>(0.0394)</td>
<td>(0.0404)</td>
<td></td>
<td>(0.0303)</td>
<td></td>
<td>(0.0266)</td>
</tr>
<tr>
<td>In_GroupVoting_Both_Focal_Distant</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-0.0203</td>
<td></td>
<td>-0.0604</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0720)</td>
<td></td>
<td>(0.0610)</td>
</tr>
<tr>
<td>Innovation Challenge_dummies</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>Constant</td>
<td>1.175***</td>
<td>1.071***</td>
<td>1.075***</td>
<td>1.098***</td>
<td>1.003***</td>
<td>1.023***</td>
</tr>
<tr>
<td></td>
<td>(0.0630)</td>
<td>(0.0791)</td>
<td>(0.0815)</td>
<td>(0.0685)</td>
<td>(0.0750)</td>
<td>(0.0765)</td>
</tr>
<tr>
<td>Observations</td>
<td>382</td>
<td>382</td>
<td>382</td>
<td>717</td>
<td>717</td>
<td>717</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-704.1</td>
<td>-700.7</td>
<td>-700.6</td>
<td>-1374</td>
<td>-1364</td>
<td>-1354</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1
References


