B.Sc. Dissertation

Evaluating the Novelty of Crowdfunding Projects

By

Chen Yanran

A0119404M

Department of Information System

School of Computing

National University of Singapore

2016/2017

Project No: H175190
Advisor: Assoc Prof Hahn Jungpil

Deliverables:
Report: 1 Volume

Abstract

With the advancement of Information Technology, online crowdfunding has gained its popularity as an untraditional fundraising method. Although major crowdfunding platforms strongly advocate and place emphasis on creative or novel projects, little is known about how they assess project novelty. Previous literatures in crowdfunding have largely focused on factors influencing the project success rate, but have paid much less attention to the measurement and implication of novelty in crowdfunding. In this study, we define and categorize novelty in the context of crowdfunding, examine and verify findings from previous researches. Our aim is to derive an effective method to evaluate the novelty of crowdfunding projects.

Keywords: crowdfunding, radical novelty, incremental novelty, dissimilarity, novelty perception

Acknowledgement

I would like to express my deepest gratitude to my supervisor Associate Professor Hahn Jungpil for his patient guidance, valuable advice, and enthusiastic encouragement through the whole project. I am grateful for his generosity of spending much time during our consultation sessions and support in many ways.

I would also like to thank Ms Yang Lusi, and Mr Wang Zhiyi for their encouragement and valuable help with all the resources and guidance.

Table of Contents

1. Introduction

Crowdfunding is the process to fund a project by raising money from a large group of individuals. Starting from the launch of ArtistShare, a website where musicians look for donations to produce their albums, in 2003, crowdfunding has begun to gain its popularity gradually (Freedman & Nutting, 2015). The world has witnessed a rapid increase in the volume raised by crowdfunding – from $880 million in 2010, to a considerable $16 billion in 2014 (Massolution, 2015). According to an estimation of the World Bank, the market for crowdfunding will further grow to $90 billion by 2025 (World Bank, 2015).

Based on the type of return that the project creators promise to their funders, crowdfunding projects can be classified into the following four categories: 1) donation-based, where the project is largely philanthropic and no direct return will be provided; 2) lending-based, also known as P2P lending, where funders lend out money and expect to get a certain rate of return; 3) equity-based, where equity shares or equivalents are given to funders in return for their investment; 4) reward-based, where funders receive some rewards for their funding, varying from a token of appreciation to a product prototype (Mollick, 2014).

Today, there exist over 1,000 crowdfunding platforms online (Drake, 2015), many of which aim to turn creative ideas into economically viable products (Ordanini, Miceli, Pizzetti, & Parasuraman, 2011). As large companies, under pressure of retaining existing market share, often take a conservative stance on disruptive innovations, it is not uncommon to see innovators working in smaller firms or even on their own. If not personally wealthy, such innovators would need to resort to external parties to get their ideas funded, which makes crowdfunding inherently a popular way to fund novel projects (Riedl, 2013). Furthermore, studies also discover that, among prominent crowdfunding platforms, there is a "strong record of success" for creative projects (Barnett, 2014).

In most crowdfunding platforms, a project usually needs to be inspected and approved by the platform before it is permitted to launch. However, we are not able to learn how the innovativeness of a project is evaluated: is it assessed by certain metrics, or purely by subjective impression? Some further questions will surface, yet unanswered: are the approved projects really “creative” or “novel”? If they have some extent of novelty, how can we measure their creativeness? Besides, what kinds of projects are more likely to exhibit novelty in crowdfunding platforms? Although previous research has examined various factors in crowdfunding projects, little is known about the nature of innovation on crowdfunding platforms.

The significance of project novelty, as well as our limited knowledge on it, has motivated us to look deeper into this topic. Therefore, this research aims to propose a method that can evaluate the novelty of crowdfunding projects. As the research process rolls out, we will first look for proper ways to define and categorize project novelty, and then attempt to derive a method to evaluate the novelty of a given project, and further examine the effectiveness of the proposed method.

2. Literature Review

2.1 Crowdfunding

Most of the previous studies on crowdfunding have an ultimate goal of promoting success rate of projects. With this emphasis, some researchers have identified factors that directly impact the success rate or performance of a crowdfunding project, such as the project’s presentation (Marom & Sade, 2013), the creator’s background (Zvilichovsky et al., 2013), the project updates (Xu et al., 2014), reward schemes of the project (Xiao et al., 2014) and stretch goals (Li & Jarvenpaa, 2015). Some papers focus on the backers instead, and discuss factors that influence their backing behavior, such as herding effect (Burtch, 2011), perceived risks (Gierczak et al., 2014), geographical distances (Agrawal et al., 2010), cultural differences (Burtch et al., 2014a), and heterogeneity of backers (Lin et al., 2014).

Compared to the amount of existing studies devoted into the exploration of project success rate and fundraising performance, the literature available in studying the novelty of crowdfunding projects appear to be relatively limited. Prior to the emergence of crowdfunding, studies have shown that venture capital firms tend to seek and fund seed capital investments that are novel (Dimov & Murray, 2008). As one of the substitutes to venture capital, crowdfunding is likely to share this proactive attitude towards creative projects. However, it should also be noted that crowdfunding backers may have different preferences and behaviors, based on types of projects they focus on. For example, people who frequently back equity-based projects share more similarities with traditional investors. Hence, they will prefer extremely novel projects, as these projects have the potential to open new market and thus render considerable return on their investment (Metrick & Yasuda, 2010). On the contrary, the motive for reward-based backers is usually that the project itself looks interesting or attractive to them (Cholakova & Clarysse, 2015). Based on these understandings, it is thus worthwhile to investigate how the novelty of a crowdfunding project is related to its attraction to potential backers, which may further influence the amount of funds raised.

Different opinions exist regarding this topic. Previous studies in traditional investments have already made it clear that greater novelty is associated with more risk and larger profit at the same time. Given this seemingly self-contradictory ground theory, some scholars believe that traditional investors not only favor but also encourage innovations (Kortum & Lerner, 2000). Nevertheless, other researches have shown that, since the risks in relation to a ground-breaking innovation is strong, venture capitalists would rather avoid such investments unless its potential of success has been proven to some extent (Kleinschmidt, 2007). Considering all these discussions, we cannot readily conclude whether equity-based backers, whose behaviors are supposed to follow the pattern of traditional investors, will act positively or negatively towards a novel crowdfunding project.

On the other hand, backers in reward-based crowdfunding are more like ordinary consumers, the difference being that they pre-order, instead of directly purchasing, products (Chan & Parhankangas, 2017). Technically, they often do not possess the financial expertise and business acumen as professional equity investors do; psychologically, they provide funds usually in exchange for products or services they are interested in, without ambitions of receiving money or shares as long-term returns (Gerber, Hui, & Kuo, 2012). As a result, literatures from traditional investment are unlikely to be applicable to them. If, instead, analyzed from consumer perspective, backers in reward-based crowdfunding may not always find novel projects favorable to their tastes. According to innovation resistance theories, consumers tend to reject products that are perceived as extremely innovative, as these products entail high learning costs, may disruptively change their routine behaviors and hence cause mental distress (Heidenreich & Kraemer, 2015). However, there are also studies pointing out that, despite their potential resistance towards highly novel products, reward-based backers are fond of incrementally innovative projects (Chan & Parhankangas, 2017).

As revealed by all the literatures above, our knowledge about novelty of crowdfunding projects is still very limited: many of the concepts and theories are directly extended from other areas such as venture capital and consumer psychology, without sufficiently justifying why they are still applicable in a crowdfunding context, not to mention that some research results even appear to contradict each other.

Among the existing discussions about the relation between the project novelty and backer responses, a prerequisite question has been overlooked: how can we know about the novelty of a crowdfunding project in the first place? Nowadays, a widely-adopted procedure is asking individuals to evaluate a project, and using their opinions to represent the novelty of the given project. Methods as such can be inaccurate, as human impressions are highly subjective. Different individuals may give answers that greatly deviate from each other, causing the result to vary every time as the sampled respondents change. One possible solution to the inaccuracy would be to distribute the same project evaluation among many individuals – as the sample size increases, the average response would be close to the true value. This solution, however, will not be practical if it is to be applied in the industry, since the time and monetary costs needed to acquire one novelty score is too large.

Regarding this situation, we are motivated to derive an accurate and efficient way to measure the novelty for crowdfunding projects, as it is the very base for other novelty-related studies. We hope this research can shed more light on the research in crowdfunding about project novelty and can inspire further studies in this area.

In the following sections, we will discuss the existing theories on novelty. In addition to considering the limited materials in crowdfunding itself, we will also make inferences from other relevant research areas.

2.2 Definition of Novelty

2.2.1 One-dimensional

In the context of project management in corporations, project novelty is often considered an essential factor, as it is associated with the riskiness, and hence the success probability, of the project. As a result, companies usually place great emphasis on analyzing project novelty before decision-making. Under such circumstances, project novelty is defined as the newness of project characteristics, relative to the firm's past experience (Voelker, Niu & Miles, 2011). This definition only involves a one-dimension criterion, which is the "dissimilarity" compared to the previous finished projects of this company.

A similar definition is the one often used in novelty detection. The purpose of novelty detection is to differentiate a given test data when it is distinct from all the existing training data. Since its main task is to separate the outlier from the normal class, novelty detection is also known as one-class classification (Pimentel et al., 2014). As suggested by its name, this novelty measurement is one-dimensional, same as the previous measurement.

Despite the difference in contexts, the above two definitions both rely on dissimilarity to conceptualize novelty, which is also consistent with the case of telecom services. According to Magnusson, Matthing and Kristensson (2003), the core dimension of novelty, or originality, is the uniqueness that differentiates this idea from others.

With dissimilarity being the one and single dimension, this definition of novelty will be largely objective and can be generalized to measure novelty for an arbitrary item, provided that there exists a clearly defined and reasonably sized reference set and an effective way to calculate dissimilarity between items in that set and the target.

2.2.2 Two-dimensional

When developing a recommendation system, it is often desirable to recommend to users projects that are tailored to their preferences and yet new to them. Therefore, in such circumstance, determining whether a project is novel or not is of great significance. In the research of Castells, Vargas and Wang (2011), they identified two possible directions to build an Item Novelty Model. The first direction is popularity-based; that is, if a project is popular, which means it must have been seen by many people, then it is less likely to be "new" to a given user. The corresponding formula is as follows:

$$\text{novelty (i)} = -log_2 p(i)$$

where p(i) is the probability that item i is observed.

The second direction is distance-based, where inter-item distance reflects the dissimilarity between items to some extent:

$$\text{novelty(i|S)} = \sum_{j \in S} p(j|S) d(i,j)$$

where d is a distance measure, and S is the reference set, which is the user's previously-picked-up item list. Similarly, the novelty defined in this way is also calculated by an exhaustive way.

Following prior studies, this research provides us with a two-dimensional definition for novelty. In other words, a novel item is expected to satisfy the following two criteria:

1) It is rarely seen by people.
2) It is largely different from existing items.

It is interesting enough to notice that, in our common way of thinking, the novelty of an item is about its existence, not the frequency it is seen. For instance, iPhone 4 was considered novel when it was released, because there never existed such smartphone prior to its time. If it is yet another non-smartphone in the given example, then although the model just came out of R&D department and had never been seen by any consumer, it still can hardly be defined as "novel", not in the way iPhone 4 was perceived. This first criterion, which seems to be counter-intuitive, would be more comprehensible in the context of recommendation system design. The aim of a recommender is to predict blanks or potentials in the user utility (Leskovec, Rajaraman & Ullman, 2011), which can only be found in unpopular products. That being said, the case of crowdfunding differs from that of recommender designing, and popularity is thus no longer a significant concern.

Besides, novelty is sometimes defined using another set of two dimensions – radical and incremental (Madjar, Greenberg & Cheng, 2011). As explained by the authors, this analogy was originally adopted in innovation literatures, from where they borrowed and then applied to creativity or novelty. It is worth mentioning that novelty is not equivalent to innovation: novelty is usually the evaluating criterion for an idea solely in its generation stage, while innovation also take further implementation stage into account (Anderson, Potočnik & Zhou, 2014). Being novel is merely a necessary condition for any innovation, but not a sufficient one. An innovation also has to be useful and appropriate (Amabile, 1996), which, depending on variety of its context, can be further translated into technical feasibility, market potential, or customer acceptance (Hart et al., 2003). The problem about this definition is that the two dimensions are not mutually exclusive. Unlike the vertical and horizontal axes in the two-dimensional space, radical and incremental novelty can be correlated and intertwined. In other words, they are not completely independent from each other, which makes them a less fit choice as a set of "dimensions".

2.2.3 Three-dimensional

In addition to the previous study on recommendation system, Zhang (2013) has proposed more comprehensive criteria for novelty. He first traced the original definition of "novelty" back to Wordnet dictionary, which appeared to be "original and of a kind not seen before" and "pleasantly new or different". Then he summarized that novelty was equivalent to three characteristics:

1) The item is not known to the user.
2) The item is dissimilar to items in profile of the user.
3) The item is satisfying for the user.

By combining the three criteria, he obtained the following formula to calculate the novelty score for a given item:

$$\mathrm{novelty(i, u)} = p(i|unknown, u) \times dis(i, pref_u) \times p(i|like, u)$$

where the three factors are respective measurements for the three characteristics.

It should be noted that a successful recommendation system is usually required to be customized one, able to recommend "novel" items that are tailored for each user's preference. As a result, profile for a specific user is taken into consideration, which renders highly subjective novel item list for that particular user. However, this is not exactly the same as our case of crowdfunding projects, where project "novelty" is meant to be an objective concept, regardless the audience viewing it. Therefore, the first and last criteria become less relevant in crowdfunding context, while the second emphasizes on "difference" or "dissimilarity" and largely aligns with the previous one-dimensional definition.

2.3 Types of Novelty

Continuing from the dissimilarity-based definition, product novelty can be further classified into two types: P-Creativity and H-Creativity. "P" stands for psychological. A product is of P-Creativity for an individual if this product is dissimilar from previous thoughts of this particular individual, which is a relative concept. "H", on the contrary, stands for historical; H-Creativity implies that the product is absolutely novel in the human history (Sarkar & Chakrabarti, 2006), which is also what we intend to study in this research.

Based on the degree of dissimilarity, there are different types of novelty from management perspective (Lopez-Mesa & Vidal, 2006). A widely-accepted theory is the five models proposed by Slaughter (1998), where she classified innovation into: 1) incremental innovation, where only minor changes are made; 2) modular innovation, which focuses on the change within a component itself; 3) architectural innovation, which changes the connection between components; 4) system innovation, where several innovations are integrated together in order to achieve a new function; 5) radical innovation, which is a breakthrough that changes the nature of an industry. These categories can be extended into crowdfunding contexts. For example, an accessory that allows a user to plug ear pods into his iPhone 7 while charging is an architectural innovation, while Pebble, the pioneer smartwatch, can be considered as a radical innovation. That being said, it is sometimes difficult, if not impossible, to go down to such details while analyzing novelty, so 5 models can be reasonably simplified into 2 – radical and incremental, as mentioned earlier in Section 2.2.2.

In summary, there exist various ways to define and categorize novelty. The one-dimensional definition states that an item dissimilar from other existing items is novel, while the two dimensional definition include another criterion "unpopularity" in addition to dissimilarity. The three-dimensional definition consists of dissimilarity, unpopularity, and pleasantness, but all the three criteria are subjective concepts specific to the individual in concern. Depending on the viewing angles, novelty can be categorized into two types – psychological and historical, or five types – incremental, modular, architectural, system, and radical. The five categories are sometimes simplified into two: incremental and radical.

After considering the concepts discussed above and assessing their appropriateness in the crowdfunding context, we decide to adopt the one-dimensional definition and the simplified two categories for novelty. Hence, project novelty is defined as the extent to which the project is dissimilar to preceding items, and will be further categorized into radical novelty and incremental novelty.

3. Data and Method

3.1 Data Description

Empirical data was retrieved from Kickstarter, one of the leading crowdfunding platforms. The dataset collected consists of 48,862 projects in total, whose launch timestamps span from October 2009 to January 2016. For each project, all the observable attributes are available, such as project name, category, state, backers, pledged amount, project description, project reward, and images embedded in description. Based on the objective of this study, we are particularly interested in the content that can help us to infer the novelty of a project, namely the project description. Other information serves as additional reference for measuring the innovativeness of a project.

Among descriptions of all projects, we find out the average length, in terms of number of letters, is around 3,231, with a minimum of 3 and a maximum of 41,837.

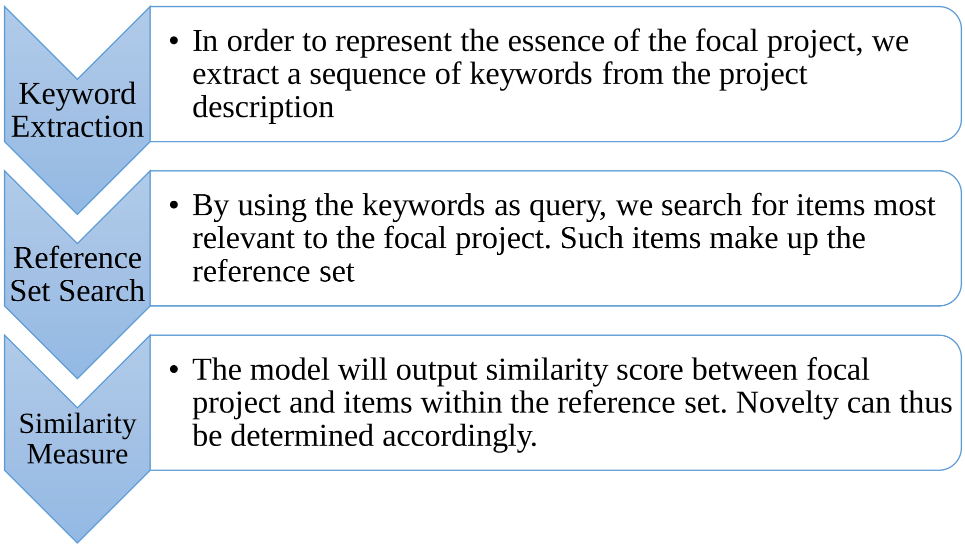
3.2 Method

As confirmed by existing literatures, the core criterion to analyze an item’s novelty, also the most appropriate in the context of crowdfunding, is its dissimilarity from existing items. The difficult part here is to find this set of “existing items”. One naïve thought might be that this set should comprise of all the items that have ever appeared on earth from the very beginning of human history. Apparently, it is not possible to construct such a set, nor is it meaningful. The reason for "not meaningful" is simple: it makes sense to compare a recently-developed smartphone with previous models to see whether it adds any novel feature, but it makes no sense to compare a smartphone to an apple – even though they are completely dissimilar, it does not imply the smartphone is novel. Therefore, as Sarkar and Chakrabarti pointed out (2006), instead of comparing the given item with everything else in the universe, it suffices to just compare it with items sharing common or relevant characteristics.

Thus, in this study, we will assess project novelty with the focus of measuring inter-item dissimilarity.

3.3 Model Description

For each given Kickstarter project, we first extract its key features from project description, in the form of a sequence of keywords. Then by using keywords matching and context extraction, we can find the existing items that are most relevant to the given project. Using these relevant items as the reference set, we will compute the similarity between the given project and the set. Overall, low similarity would indicate greater novelty. The process is demonstrated in the flow chart below.



3.3.1 Keyword Extraction

Since each crowdfunding project is assumed to create something new and be distinct from others, there will be no training set that is suitable for every focal project under supervised methods, so we adopt an unsupervised keyword extraction method called TextRank. The detailed process is described as follows.

Pre-processing. Before starting, the project description needs to be pre-processed, such as removing stop words and cleansing. In order to achieve this, we make use of Natural Language Toolkit (NLTK), a package available in Python. Besides removing all the stop words listed in English dictionary, other words that occur frequently in Kickstarter projects are also manually added into the stop word list, such as “product”, “project” and “reward”, etc.. Considering that stemming may result in words in their incomplete forms – for example, the third-person singular verb “senses” will become “sens” after stemming, which is not even an English word – we replace it by lemmatizing. WordNet Lemmatizer within NLTK allows us to lemmatize based on word type: after recognizing “senses” as a verb, its lemmatized form will become “sense”, the correct result as we expect. Also, Since TextRank, the keyword selection step after pre-processing, requires a complete paragraph as input, we will perform no tokenization but leave the project description in the form of sentences.

Keyword/Keyphrase Selection. With the project pre-processed and cleansed, we then apply TextRank, a widely-adopted graph-based ranking method, to identify keywords or keyphrases from project descriptions. TextRank derives its main idea from networking theories: in any directed network made of edges and vertices, score of a vertex V_i is defined as:

$$S(V_i) = (1 - d) + d \times \sum_{j \in In(V_i)} \frac{1}{|Out(V_j)|} S(V_j)$$

Where d is a factor between 0 and 1, often set to 0.85, $In(V_i)$ is the set of points who direct into V_i, and Out (V_j) is the set of points to whom V_j is directed. Iterate until the scores converge, and the vertices with higher scores – usually those connected to more points – are deemed as more significant (Brin & Page, 1998).

To transform text into graph, each word will be treated as a vertex, and the co-occurrence between each pair of vertices will become the edge bridging them. Two words are co-occurring, and thus connected, if their distance in document is within N ($2 < N < 10$) words. During the initial run of the algorithm, only words, but not phrases, will be considered as vertices. After the iterations are completed, post-processing will be conducted for words with relatively high scores to identify multi-word phrases (Mihalcea & Tarau, 2004). The selected words and phrases are hence the keywords and keyphrases for the given project.

3.3.2 Reference Set Search

In this step, we aim to find a proper reference set (i.e., a set that contains items that are most relevant to the given project). If the given project and the reference set share some characteristics in common, yet some differences exist between them, then the project is likely to have a certain extent of novelty. If even the most relevant items have little in common with the project, then we can conclude with confidence that the project must be novel to a great extent.

Given our objective to measure dissimilarity between focal project and its most relevant items, without constraining the range to any specific platform, we need to obtain the reference set from external sources on the Internet. In order to select the items most relevant to the given project, we intend to make use of keyword matching of search engines. For each given project, we create a query using the keywords or phrases identified in the keywords extraction process. By using these keywords as the input and restricting the time to be one day prior to the launched time of the focal project, the search engine can return us a list of results relevant to our keywords input. Since the search results are sorted by relevancy, we are able to select the top 10 results as the reference items that are most relevant to the given project. We then extract the textual contents from these 10 webpages, which will be compared with the given project for dissimilarity later.

Currently, there are various search engine APIs available on the Internet, among which we choose Google Custom Search API for its sophistication and accuracy. The web crawling followed is done by Beautiful Soup, a package available in Python, if the link to be crawled is a normal HTML page. If it happens to be a PDF file, we then adopt another package PyPDF2 to download it and convert its contents to plain-text documents.

3.3.3 Measuring Text Similarity

With the reference set ready, we then proceed to measure the text dissimilarity between documents in the reference set and the description of given project. According to our design, the reason that a document can be selected into the reference set is because it matches some keywords of the given project description. Therefore, it might be ineffective to represent dissimilarity using cosine similarity metrics between the Term Frequency–Inverse Document Frequency (TF-IDF) vectors, which is calculated on the basis of exact word matching.

As a result, we propose to use Similarity Queries, which constructs a Latent Semantic Indexing (LSI) space and transforms documents into a stream of vectors in the defined LSI space. Unlike TF-IDF based algorithms that derive similarity from exact words shared by two documents in common, LSI can reveal the deeper similarity in "topic" by performing semantic generalization.

To execute the procedure, we first train a LSI model on English Wikipedia, which is considered an unbiased general corpus. The dimensions of the LSI space are the topics modeled from English Wikipedia by performing semantic analysis. Following convention, we set the number of topics to 100. Next, we convert documents in the reference set to the trained LSI space and index them according to each dimension. Meanwhile, the same is carried out for the focal project description. Now that each of them is represented by a set of coordinates, we are able to perform similarity queries for the given project description against the reference set, using cosine distance between two vectors to represent their difference. The queries will return, for each document in the set, its coordinate and its similarity score with the given project description. We then compute the average similarity score for each given project with its top 10 most relevant items; low average similarity score implies high novelty.

Gensim, which is a python library, will be used to implement this approach for measuring inter-platform dissimilarity.

3.4 Model Validation

After obtaining the measure of project novelty through algorithms above, additional analysis needs to be conducted to validate our approach.

In order to test the consistency between machine and human judgement, we conduct surveys on Amazon Turk (MTurk) to evaluate how novelty of crowdfunding projects is perceived by potential backers, and then proceed to compare with the result of our model. It is unlikely to carry out surveys for all projects in the dataset, given the budget restriction. Thus, we construct a random sample of projects for validation.

First of all, we understand that an individual's judgement of novelty is made based on his knowledge and experience at the particular moment. In other words, time is a crucial factor affecting one's perception of novelty. It is not uncommon to see many once ground-breaking technologies become so popular and widely-used that, several years later, nobody would consider them to be novel any more. By restricting search results to precede the project launch date, the model assesses project novelty at its debut. Then, ideally, we would expect the survey responses to be based at the same time point, so the two results will be comparable. While the actual human perception at launch time is not retrievable, we need to find the next-to-best measure as alternative. We assume that, if the time passed since the project launch is shorter, then the individual's current perception of project novelty will be a more appropriate proxy for perception at its launch time. Based on this assumption, we first narrow the sampling range down to the latest projects in the dataset, that is, projects whose launch times are between October 2015 and January 2016.

Within this sample range, there are 157 distinct categories and subcategories in total, such as Design, Fashion, Food, Music, and Technology. To ensure our sample is unbiased, we intend to select from categories that are typical and representative. Drawing inspirations from previous literatures (Mollick & Kuppuswamy, 2014), we decide to focus on Technology, Design, Games in particular, as projects under these categories are perceived to have larger variation of novelty across categories (Chan & Parhankangas, 2017). Considering that products from both Technology and Game are usually in tangible forms while Design may be not, we add one additional category "Film & Video" to make even between the tangible and intangible.

We randomly sample 30 projects in each selected category, which gives us 120 projects in total. Projects under each category are sorted according to the model output of similarity score and then divided into three groups – low similarity (i.e., high novelty), medium similarity (i.e., medium novelty), and high similarity (i.e., low novelty). Combining the low similarity groups for all four categories gives us the low similarity set; same is done for the medium and high similarity groups.

With the three sets available, the subsequent sampling is carried out in multiple rounds. Each round we draw one project from a set without replacement, and construct a survey using the three projects drawn out in the same round. In this way, we obtain 40 copies of survey, each containing three projects, with low, medium, and high similarity respectively based on the results of our model. We then repeat the same process and obtain another 40 copies, which has the same overall contents as the first batch, but different in terms of combinations of projects for surveys.

In each survey, the respondent is first asked to provide basic demographic information, including age, gender, education level, employment status. Afterwards, the respondent is given a project description and needs to answer several questions based on his understanding of the given description; the settings are the same for the other two projects. The design of survey items are adapted from a previous study (Chan & Parhankangas, 2017). For each project, the survey consists of one question about the respondent's first impression of the focal project, one set of questions regarding the focal project's radical novelty, and one set of questions regarding the incremental novelty[1]. In each set, every respondent is required to answer 6 questions. Following prior research convention, a 7-point Likert scale is used in the survey, with degree 1 indicating strongly disagree and degree 7 indicating strongly agree. Each participant is compensated $5 for their participation. A sample survey can be found in Appendix 1.

4. Result Analysis

4.1 Data Cleansing

[1] Question set 1 contains a statement "There are no other project/product like this on the market right now", while question set 2 contains a statement "The project/product is quite similar to existing project/products on the market". The two questions are supposed to have opposite ranks (i.e., the extent of agreement). Responses that do not satisfy this criterion will be rejected and not documented into the results.

Among descriptions of all the 120 projects, 8 are considered invalid descriptions. These descriptions are either written in languages other than English (e.g., French and German), or contains only links to images and videos but no meaningful textual contents. After excluding such projects from consideration, we are left with 112 valid projects to evaluate. Since each project is evaluated twice, that gives us 224 valid data points.

By our survey design, for each project, all 6 questions in question set 1 revolves around the main idea "how radically innovative you think the focal project is", only asked from different angles. Similar settings also holds for the design of question set 2. Due to the deliberately repetitive design, answers within the same question set are supposed to be consistent. Depending on this rule, we are able to detect and remove self-contradictory responses, which indicates that the respondents may have randomly selected answers for questions, instead of basing his responses on the exact project description.

We normalize all the data, and then, for each project, we group responses according to the question set it belongs to, and calculate the sample variance within each question set. If the variance for either group is larger than the standard deviation (i.e., 1), this response is discarded.

After filtering and cleansing, we are left with 190 individual responses.

4.2 Descriptive Analysis

Information available in each response consists of three parts: demographical information (i.e., age, gender, education level, and employment status), crowdfunding familiarity information (i.e.., knowledge of crowdfunding, whether visit Kickstarter, and whether invest in new ventures), and project specific information (i.e., ranks for the two sets of questions). Before proceeding any further, we conduct the descriptive analysis to grab general understanding of the data structure.

Among all the 190 responses, 174 claim that the corresponding respondents have visited Kickstarter website (91.6%), 108 show that their respondents have been involved in new venture or entrepreneurial investments (56.8%). This has again verified people's passion about investment as well as the popularity of online crowdfunding. In terms of demographics, 132 are from the younger generation (aged from 18 to 35), which are more tech-savvy. 170 are employed or entrepreneurs, indicating that a dominant majority has relatively stable income. Such demographical profiles match those of typical online crowdfunding backers (Kimbia, 2014), suggesting that our survey sample is representative.

In order to investigate whether demographical characteristics have any impact on an individual's perception of project novelty, we take the following procedures. For each individual response, we create new variable "Average Radical" and "Average Incremental" by calculating the mean rank for question sets 1 and 2 respectively. Then we study the two new variables against various demographical characteristics.

4.2.1 Age V.S. Novelty Perception

In the survey, we group ages into the several intervals: Below 18, 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75 and above, but it turns out that all our respondents fall into the age range from 18 to 64. We calculate the means of "Average Radical" and "Average Incremental" for every age group that we have, and the results are displayed in Figure 1. Since there is only one data point for age group 55-64, we cannot verify that its behaviors are representative for the whole group. Therefore, this data point is treated as an outlier and the corresponding age group is excluded from our analysis. From the remaining four groups, we see no obvious patterns for Average Radical and Average Incremental as the age increases. Hence we suppose that an individual's age has little impact on his perception of novelty and our survey responses will not be biased because of age.

4.2.2 Gender V.S. Novelty Perception

We plot a similar Figure 2 to study the impact of gender on novelty perception. The means of Average Incremental are around the same across genders; however, the mean of Average Radical for male is lower than that of female by 8%, which indicates that males may have stricter criteria for radical novel product than females. We further divide the ranks into 3 groups: 1-3, 3-5, and 5-7. By calculating the number of responses within different rank groups, we break down the structure into more details and obtain Figure 3 and 4. It seems that, compared to females, males are more likely to give moderate ranks in terms of both radical and incremental novelty. Moreover, both genders tend to give higher ranks in Average Incremental than in Average Radical, which suggests that incrementally innovative projects may be more common than radically innovative projects.

4.2.3 Income V.S. Novelty Perception

Besides age and gender, we are also curious to study the impact of financial factors, such as income. Since income is a sensitive question that should be preferably avoided in survey design, we ask about the respondent's employment status and education level instead: The employed and entrepreneurs usually have more stable and higher income than the unemployed and students, and studies have already shown that there exists a strong positive correlation between education and earnings (Blaug, 1972). Therefore, we could reasonably use these two characteristics as indicators of personal income. The results are plotted in Figure 5 and 6. For both Average Radical and Average Incremental, an increase of education level seems to have positive impact on their ranks. However, we cannot conclude from the data whether or not the employed and entrepreneurs will have a higher rank relative to the unemployed and students. As a result, the impact of income on novelty perception is still unclear and needs further investigation.

4.2.4 Crowdfunding Familiarity V.S. Novelty Perception

In crowdfunding websites like Kickstarter, it is not uncommon to see two projects offering similar products. For instance, ilDock and Auxillite are both accessories that enable the user to charge and listen to his iPhone 7 at the same time. When asked to evaluate the novelty of Auxillite, if an individual actively browses the website and frequently participates in backing events, then he stands for a higher chance of having encountered ilDock before. As a result, his novelty score for Auxillite may be lower than a new comer in crowdfunding. Inspired by such scenarios, we intend to assess whether the respondent's familiarity with crowdfunding will indeed affect his perception of project novelty.

In our survey, two questions are directed to test the respondent's familiarity with crowdfunding. The first question asks the respondent to rank the extent of his knowledge about the industry, using a 7-point Likert scale. The second question asks if the respondent has personally participated in such investments. Again, we calculate the means of Average Radical and Average Incremental for different familiarity groups, and the results are displayed below in Figure 7 and 8.

It is interesting to note that, people only invest in crowdfunding projects when they think their background knowledge is above average, and backers who know crowdfunding extremely well will always make investments. Despite the different behavior patterns, there is no obvious trend of novelty perception change as the familiarity increases, indicating that the impact of crowdfunding familiarity on novelty perception is little.

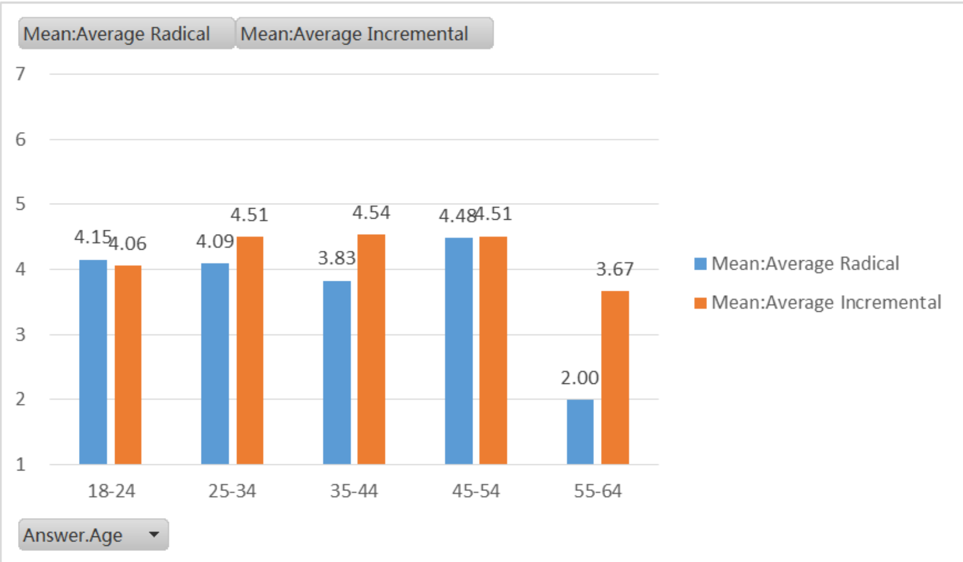


Figure 1. *Bar chart of the means of Average Radical and Average Incremental for age groups.*

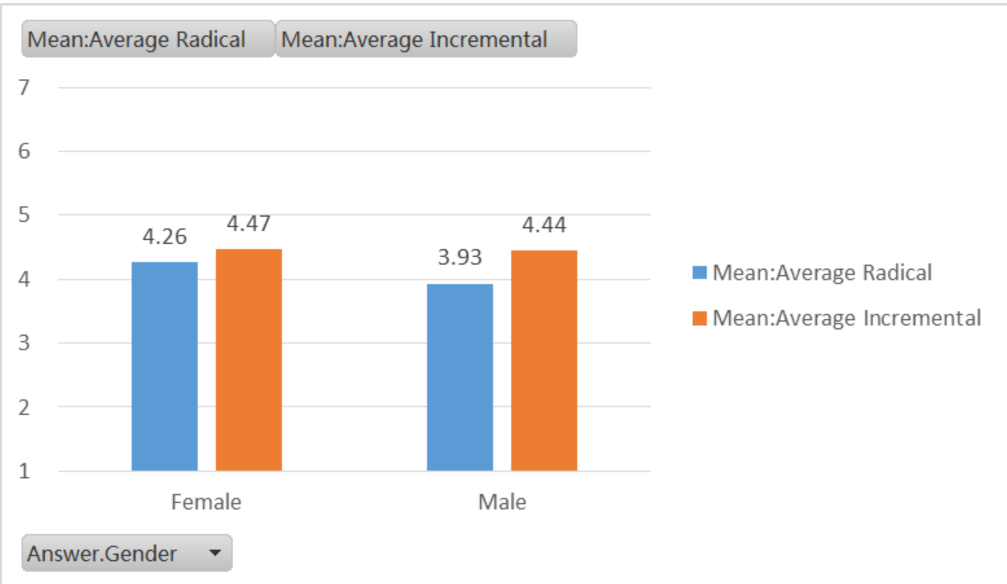


Figure 2. *Bar chart of the means of Average Radical and Average Incremental for genders.*

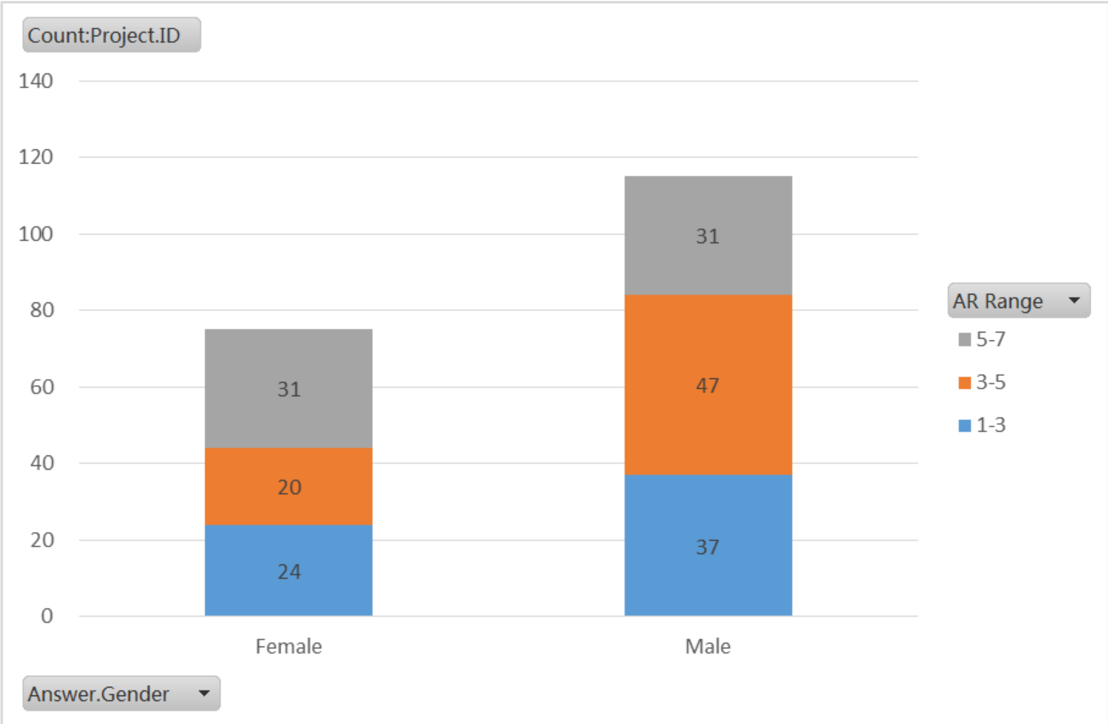


Figure 3. Stacked bar chart for the number of responses within different Average Radical rank groups against genders.

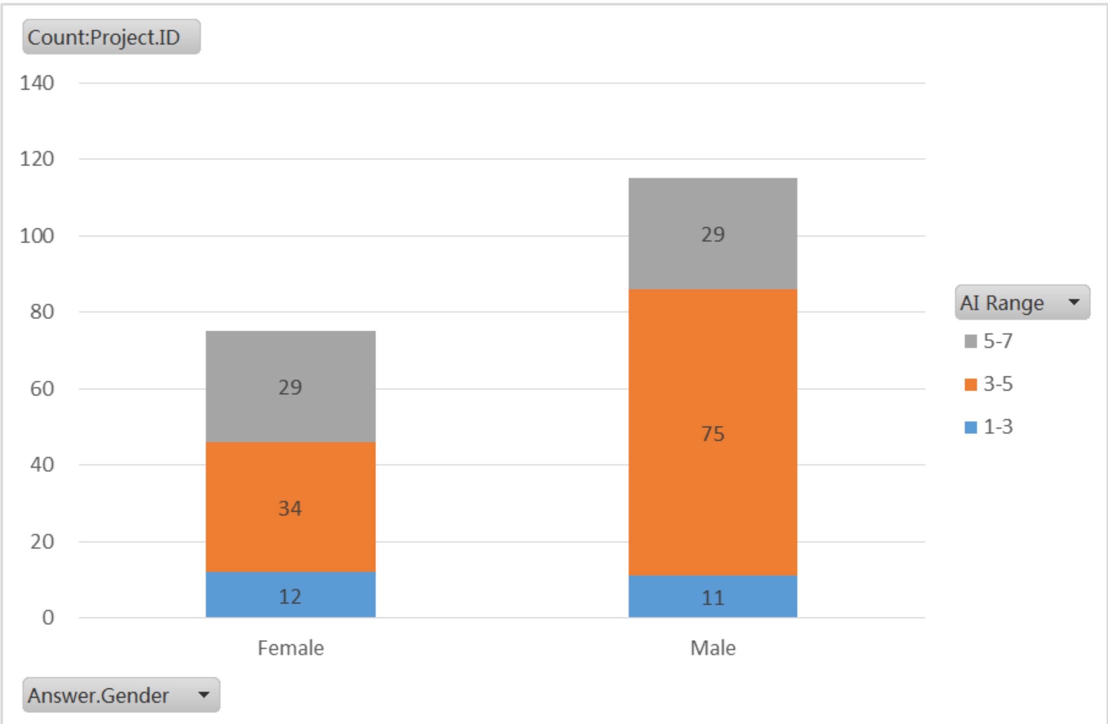


Figure 4. Stacked bar chart for the number of responses within different Average Incremental rank groups against genders.

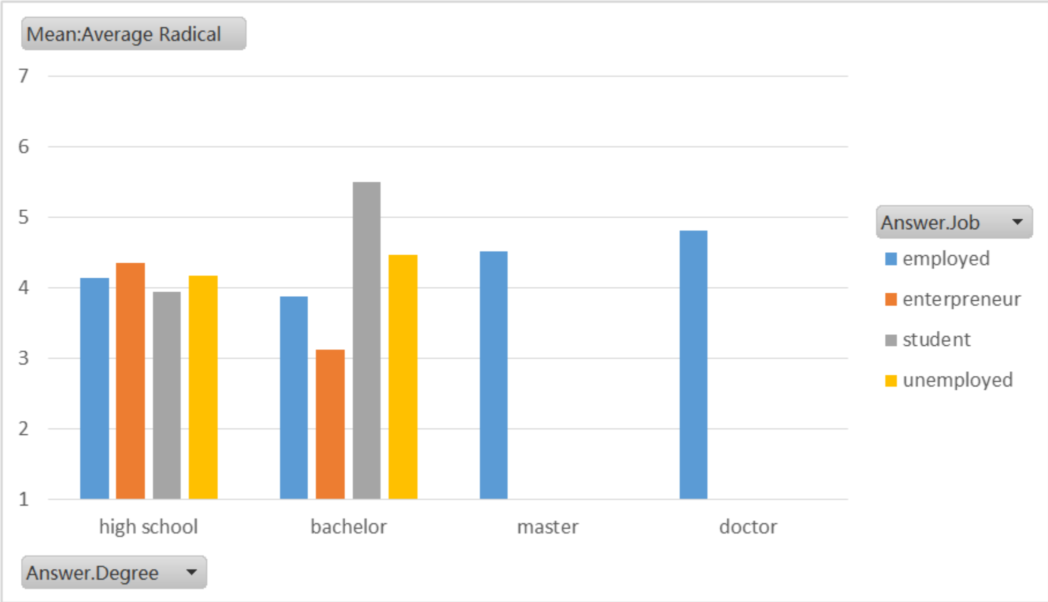


Figure 5. Bar chart of the mean of Average Radical for different education levels and employment statuses.

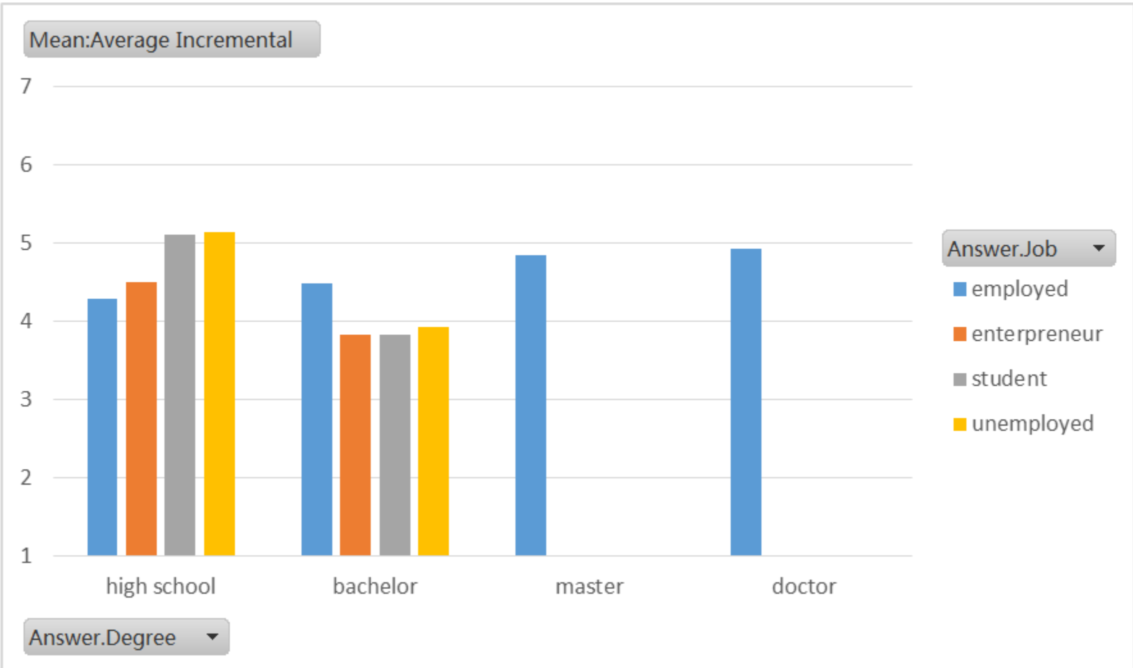


Figure 6. Bar chart of the mean of Average Incremental for different education levels and employment statuses.

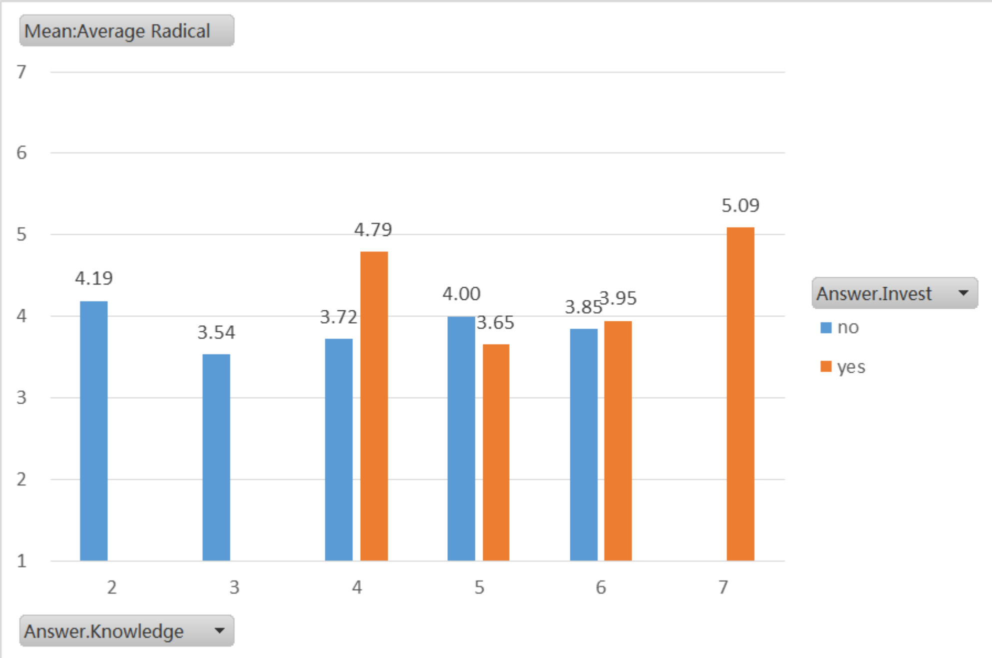


Figure 7. Bar chart of the mean of Average Radical for different familiarity groups.

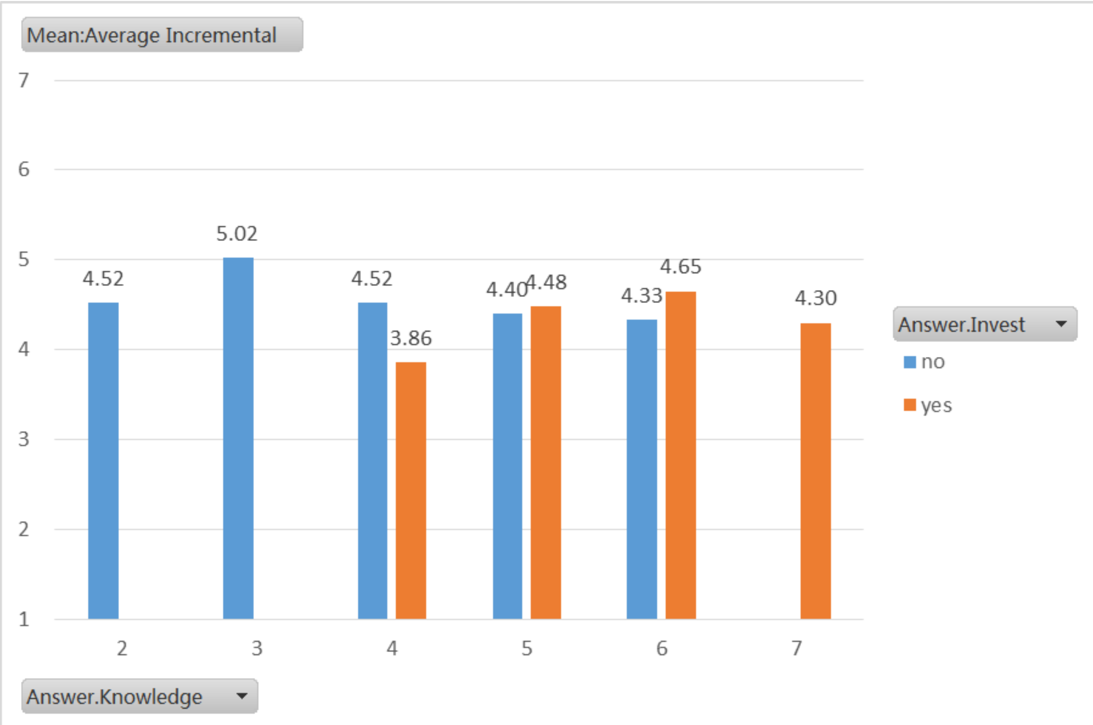


Figure 8. Bar chart of the mean of Average Incremental for different familiarity groups.

4.2.5 Novelty Perception V.S. Preference

Last but not least, we intend to verify people's attitudes towards radical and incremental novelty, which has been argued and discussed in previous literatures. Respondents are asked to select their first impression for each given project, with five options: very positive (denoted by positive+), somewhat positive (denoted by positive), neutral, somewhat negative (denoted by negative), very negative (denoted by negative-). For responses in all the three rank groups (i.e., 1-3, 3-5, and 5-7), we calculate the percentage of each option and present the results in Figure 9 and Figure 10 below.

For Average Radical, the result is straight forward: the more radically innovative people think one project is, the better impression they will have on that project. This supports opinions of Metrick and Yasuda (2010) that extremely innovative projects are more favorable to backers.

Incremental novelty is yet another story. Interestingly, people seem to prefer projects that have either extremely high incremental novelty or extremely low incremental novelty. The high side coincides with the study of Chan and Parhankangas (2017) that reward-based backers favor projects that are more incrementally innovative. One possible explanation for the low side might be that, since low radical novelty suggests the product is modified little based on existing items, learning cost for such products would be low. Such products can be easily comprehended, and would very likely look familiar to the backer, thus appealing to his positive emotions.

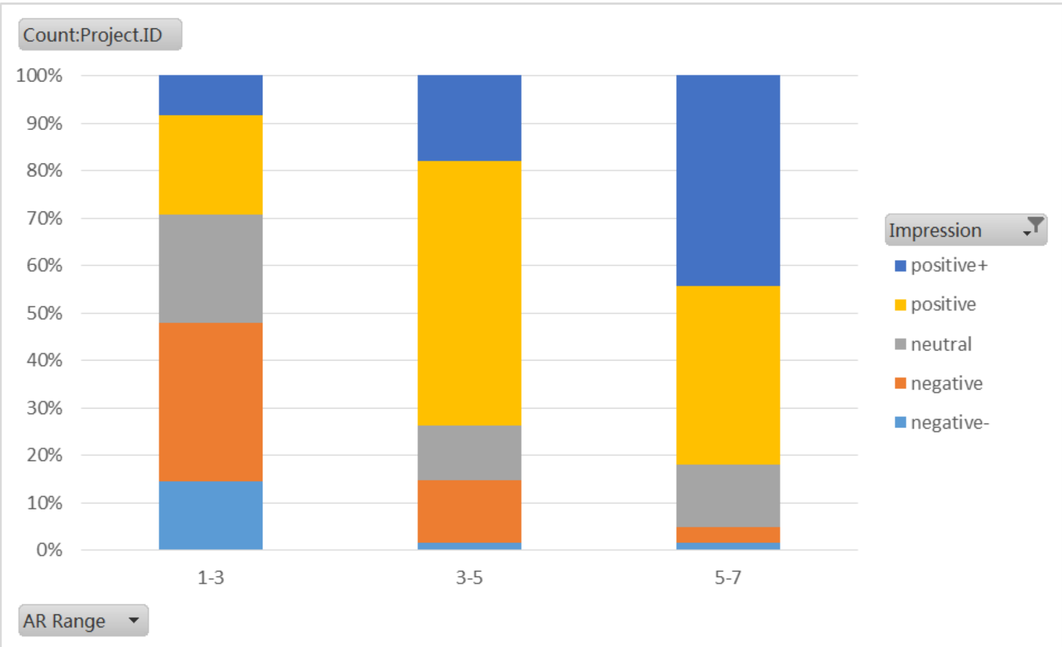


Figure 9. Stacked bar chart of impression structure for Average Radical rank groups.

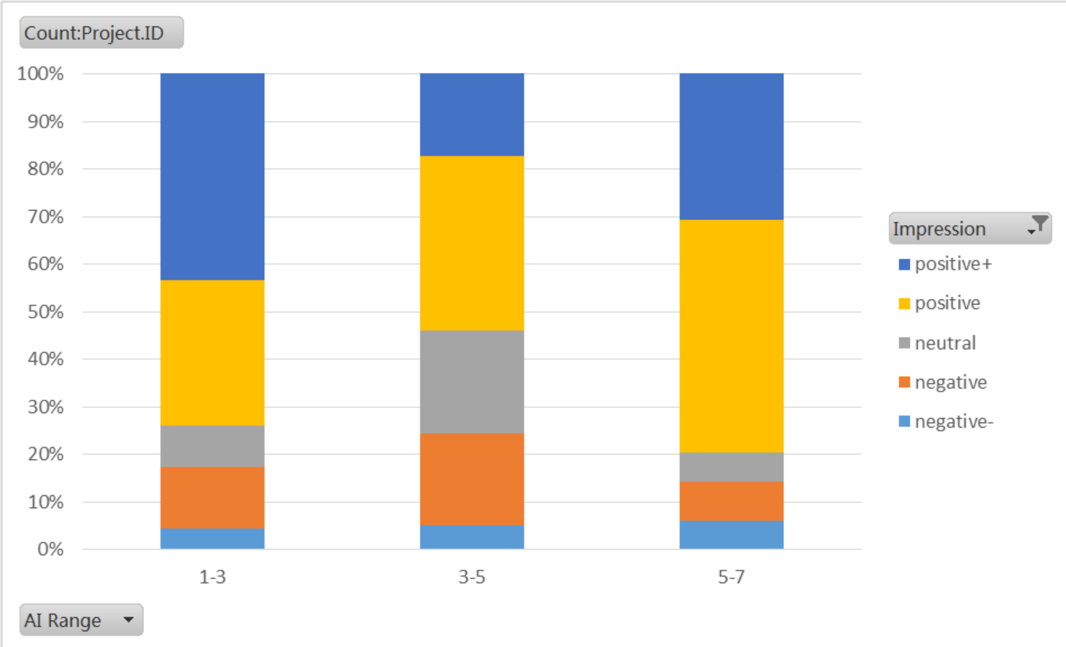


Figure 10. Stacked bar chart of impression structure for Average Incremental rank groups.

4.3 Factor Analysis

In our survey, question set 1 (consisting of 6 items) is designed to reflect radical novelty, while question set 2 (also consisting of 6 items) is designed to reflect incremental novelty. The two groups can be easily seen on the correlation plot (Figure 1). The alpha reliability coefficients for radical and incremental are 0.95 and 0.85 respectively; the high values suggest that items within the same question set have high covariance and possibly measure the same underlying subject. From both correlation plot and the alpha coefficients, the 2-factor latent structure seems obvious.

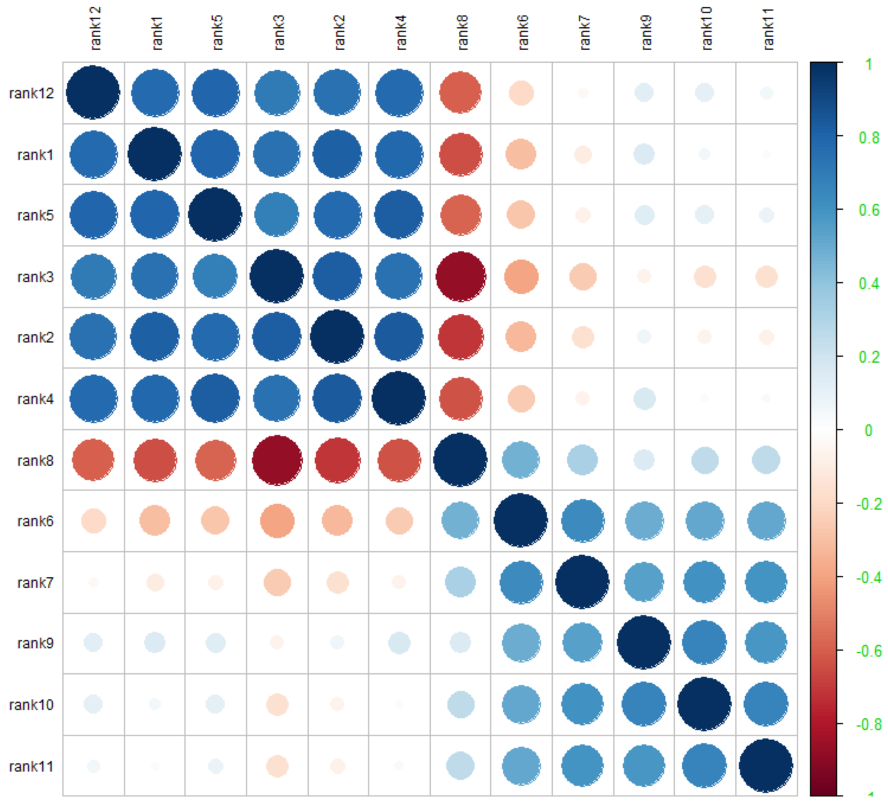


Figure 11. Correlation Plot

Some items are designed to be mixed in order to avoid response bias. Rank1-5 & rank12 belong to question set 1, while rank6-11 belong to question set 2.

We perform Exploratory Factor Analysis (EFA) to determine the appropriate number of factors and further verify the factor structure of the items. The Scree Plot is as shown in Figure 2.

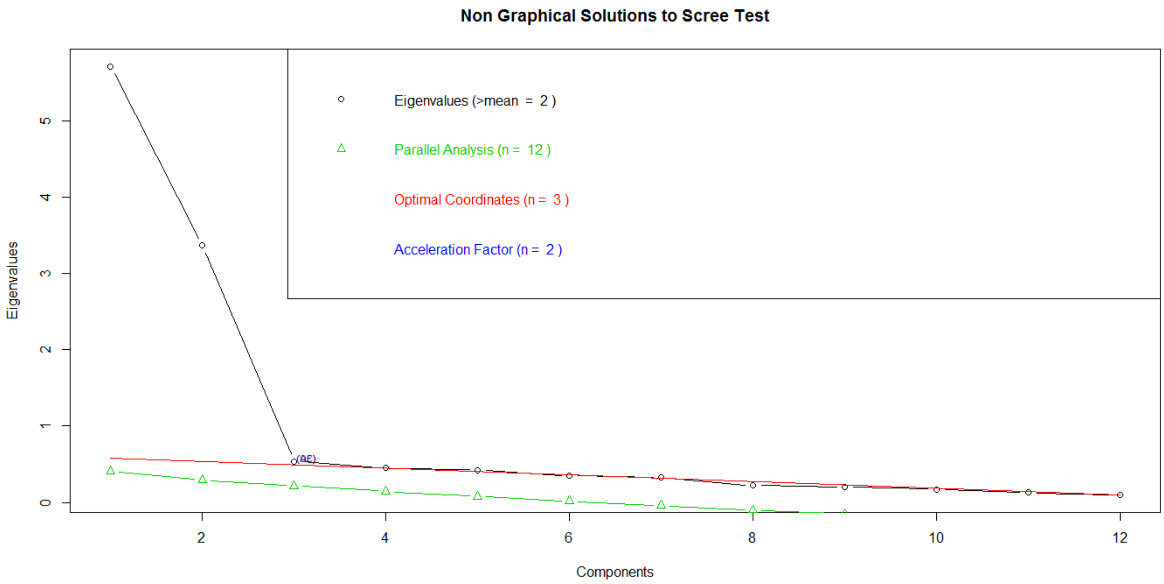


Figure 12. Scree Plot

As demonstrated in Figure 2, the scree, or the drastic change of the slope of the eigenvalue curve, occurs at the third eigenvalue. According to Cattell's scree test, two components need to be retained (Raîche, Riopel & Blais, 2006), which aligns with our speculation above.

With the number of factor confirmed to be 2, we proceed with a principal component factor analysis with varimax rotation. The result shows that Factor 1, representing the radical novelty, explains 44.7% of item variances, and Factor 2, representing the incremental novelty, explains another 25.9% of variances. Items and their respective loadings are displayed in Table 1 below.

	Factor 1	**Factor 2**
rank1: This project/product is original.	0.888	
rank2: This project/product is radically new.	0.91	
rank3: There are no other project/product like this on the market right now.	0.863	-0.241
rank4: The individual or team behind this project/product suggests radically new ways for doing things.	0.9	
rank5: The individual or team behind this project/product is a good source of highly creative ideas.	0.873	
rank12: The individual or team behind this project/product demonstrates originality in his/her/their work.	0.853	
rank6: The project/product is a modification of an existing project/product.	-0.317	0.677
rank7: The individual or team behind this project/product easily modifies previously existing work processes to suit current needs.	-0.117	0.751
rank9: The project/product is an incremental improvement over an existing project/product.	0.13	0.759
rank10: The individual or team behind this project/product is very good at adapting already existing ideas.		0.83
rank11: The individual or team behind this project/product uses previously existing ideas or work in an appropriate new way.		0.782
rank8: The project/product is quite similar to existing project/products on the market.	-0.758	0.359

Table 1. Factor Loadings

Given the sample size of 190, only loadings whose absolute values are larger than 0.45 will be deemed meaningful and have practical significance (Hair, Tatham, Anderson & Black, 1998). Hence, we drop the loadings that do not meet this threshold and produce the adjusted factor loadings in Table 2.

	Factor 1	Factor 2
rank1: This project/product is original.	0.888	
rank2: This project/product is radically new.	0.91	
rank3: There are no other project/product like this on the market right now.	0.863	
rank4: The individual or team behind this project/product suggests radically new ways for doing things.	0.9	
rank5: The individual or team behind this project/product is a good source of highly creative ideas.	0.873	
rank12: The individual or team behind this project/product demonstrates originality in his/her/their work.	0.853	
rank6: The project/product is a modification of an existing project/product.		0.677
rank7: The individual or team behind this project/product easily modifies previously existing work processes to suit current needs.		0.751
rank9: The project/product is an incremental improvement over an existing project/product.		0.759
rank10: The individual or team behind this project/product is very good at adapting already existing ideas.		0.83
rank11: The individual or team behind this project/product uses previously existing ideas or work in an appropriate new way.		0.782
rank8: The project/product is quite similar to existing project/products on the market.	-0.758	

Table 2. Adjusted Factor Loadings

With the original data and the factor loading matrix, we make use of the factor.score function within R Library "psych" to obtain the 2 factor scores for each response.

4.4 Correlation Analysis

After obtaining the 2 factor scores, which are indicators for human judgement of project novelty, we are to calculate the correlation between human judgement and the model output.

As mentioned in previous sections, human perception of novelty can be highly subjective; two individuals may reach completely different conclusions, even if they are asked to evaluate the same project. For the correlation result to be more fair and reasonable, we wish to rule out the individual subjective impact and preserve only the novelty perception shared by the general. In order to achieve this goal, we inspect the remaining data again.

Originally, each project is evaluated twice, and hence should correspond to a pair of responses. However, it should be noted that, since we removed some unqualified responses in earlier steps, only part of the remaining responses will come in pair while the other part exists on its own. Calculations show that 81 projects still have a pair of corresponding responses; 28 projects now only have a single response, with 12 of them from the first batch survey and 16 of them from the second batch survey.

After normalize the data frame that consists of Factor 1 score, Factor 2 score and model similarity score for all the projects left, we focus on those still with a pair of responses. We calculate the difference of its factor scores between the two responses. If the difference is smaller than the standard deviation (i.e., 1), it means that the two respondents reach a consensus regarding novelty of the focal project, so we can assume their consistent opinions is a proper representative of the "true" project novelty. Otherwise, scores of the two responses are considered deviated largely from each other. It suggests the subjective impact is large and, since there is no available reference for the "true" project novelty, we will exclude the particular project from consideration.

This filter process leaves us with the last 47 projects, which we use for correlation analysis. We calculate the correlation among Factor1 score, Factor2 score, and the model similarity score. The results are displayed in Table 3 below.

It can be seen that the radical novelty and model similarity are negatively correlated, which is consistent with our expectation of the model. A slightly positive correlation exists for incremental novelty and model similarity. This may be due to the dissimilarity-based definition of the novelty that we adopt for the model: the more radical innovative a project is, the less similarity it will share with existing items. Nevertheless, since the incremental novelty is based on modifying existing items, there is no decisive conclusion that the larger modification proportion leads to a higher incremental innovative score – an immediate counter example would be 100% modification proportion, which makes the project radically innovative and deviated far from incremental novelty.

By further segmenting projects according to their category, we obtain the correlations for projects in the four categories, respectively. The results are shown in the Table 4 to 7.

Correlation Table of the 47 Projects			
	avg.radical	avg.incremental	similarity
avg.radical	1	0.17784	-0.56255
avg.incremental	0.17784	1	0.020891
similarity	-0.56255	0.020891	1

Table 3. Correlation table of the 47 projects

Correlation Table of Design Projects			
	avg.radical	avg.incremental	similarity
avg.radical	1	0.034376	-0.44841
avg.incremental	0.034376	1	0.33495
similarity	-0.44841	0.33495	1

Table 4. Correlation table of projects under category Design

Correlation Table of Film Projects			
	avg.radical	avg.incremental	similarity
avg.radical	1	0.243345	-0.35116
avg.incremental	0.243345	1	-0.08035
similarity	-0.35116	-0.08035	1

Table 5. Correlation table of projects under category Film

Correlation Table of Game Projects			
	avg.radical	avg.incremental	similarity
avg.radical	1	0.445395	-0.82522
avg.incremental	0.445395	1	-0.34657
similarity	-0.82522	-0.34657	1

Table 6. Correlation table of projects under category Game

Correlation Table of Tech Projects			
	avg.radical	avg.incremental	similarity
avg.radical	1	-0.03774	-0.72398
avg.incremental	-0.03774	1	0.159554
similarity	-0.72398	0.159554	1

Table 7. Correlation table of projects under category Tech

We find that for tangible products such as games and technology gadgets, the negative correlation between radical novelty and model similarity score tends to be larger. The reason may be that such products are less abstract and easier to describe, so the keywords extracted can give a better summary of their features in the first place. Moreover, matching tangible items is usually more accurate than intangible ones. For instance, if a pair of smartphones are matched by same core, screen size, and storage size, then the similarity degree between them is likely to be high. However, if two movies are matched by the same theme "adventures, teenagers, heroes", it is possible that one of them is Harry Potter and the other is Hunger Games, the two of which are, to many, not similar in nature.

4.5 Regression Analysis

Combining the factor scores and model similarity score with project category and length of project description, we perform linear regression on the data in R. in order to examine factors that influence human perception of project novelty, we regress Factor 1 score, i.e., the radical novelty score, on model similarity, project category and the length of description. The model summary is shown below:

Coefficients for Independent Variables					
	Estimate	Std. Error	t value	Pr(>|t|)	significance level
(Intercept)	-0.24	1.80E-01	-1.343	0.1866	
similarity	-0.61	1.19E-01	-5.096	8.24E-06	***[2]
categoryFilm	-0.01	2.17E-01	-0.065	0.9487	
categoryGame	-0.00	2.42E-01	-0.018	0.9859	
categoryTech	0.27	2.30E-01	1.166	0.2503	
lenDescription	0.00	2.36E-05	2.474	0.0176	*[3]

Table 8. Coefficients for the independent variables.

It can be seen that the model similarity indeed has a significant negative impact on the radical novelty score. In other words, whether an individual will perceive a project as radically novel largely depends on contents of the project itself, rather than the category it belongs to, or the way it is presented. Length of project description is also significant at 95% confidence level. The positive coefficient sign indicates that the longer project

[2] *** indicates that the variable is significant at 99.9% confidence level.
[3] * indicates that the variable is significant at 95% confidence level.

description is, the chance that people perceive this project as novel is higher. There are two possible explanations. Firstly, a radical innovative project may have many ground-breaking features, some of which may be even beyond current imagination of the audience. In order to facilitate the backers' understanding of his project, the creator may need to put more efforts into elaboration, compared to describing a more commonly-seen, less innovative product. Secondly, a longer description enables backers to obtain a more nuanced comprehension of the project and hence are more likely to appreciate innovative parts of the project. In addition, the novel features can also be reinforced and emphasized, so that they are less likely to go unrecognized.

The adjusted R-squared is 35.3%, meaning that these independent variables explain 35.3% of the variation in radical novelty score. Obviously, there are still factors that are not accounted for in this model, which may require future studies to investigate.

5. Discussion and Conclusion

As an emerging phenomenon in the fundraising market, crowdfunding has attracted attention of scholars. While much efforts have been devoted into the exploration of project success rate, only limited researches focus on the role of project novelty in crowdfunding. In this study, we aim to derive an effective way of evaluating the novelty of crowdfunding projects. First of all, we define novelty as the dissimilarity from existing items. Our model is to extract keywords from the project description, and then find items that are most relevant to the focal project via keyword matching. We then apply Latent Semantic Indexing to calculate the similarity score between the focal crowdfunding project and its relevant items. Low similarity would imply high novelty. The model result is then validated against human responses on random selected sample of projects. We analyze the correlation between model output and human judgement in both general sense and context of specific categories. We further perform regression to verify the impact of project dissimilarity on human perception of project novelty, as well as to identify other factors that may influence backers' novelty perception.

5.1 Result Summary

The general correlation between model similarity score and human novelty perception is -56.2%, indicating that our model result is consistent with human judgement to a large extent, and that low similarity indeed implies high novelty. Moreover, the model appears to be more accurate for tangible projects (e.g., technology, games) than intangible ones (e.g., design, videos). The regression result suggests that similarity score and description length are two significant factors that influence individual novelty perception of crowdfunding projects. Similarity score is, as shown in the correlation analysis, negatively related to novelty perception, while the description length is positively related to novelty perception.

Besides, we use our data of human responses to test individuals' attitudes towards novelty, since some conclusions from previous literatures contradict each other. It turns out that backers will become more positive towards a crowdfunding project as its radical novelty increases. In terms of incremental novelty, backers favor those projects with either extremely low incremental novelty, or extremely high incremental novelty.

5.2 Implications

Although major online crowdfunding platforms usually place emphasis on project novelty, there is no known effective method to assess it other than manual evaluation. This is where the practical value of our research lies in. We propose a way to measure project novelty, one that is automated and with acceptable accuracy. Using algorithms to generate novelty scores for projects, this method requires much less time and costs than manual inspections. Due to its cost-effectiveness, the method has high scalability. Once adapted in online crowdfunding platforms, it can help platforms to identify the novelty degree of each project, and platforms can thus recommend projects to backers according to their novelty. Since backers express a positive attitude towards projects with higher radical novelty, recommending such projects will increase their chance of backing. Platforms, which charge fees for fully funded projects, will also benefit from the increase of success rate.

Moreover, backers with preferences on novelty can then filter projects according to their novelty scores and save time for untargeted searching.

5.3 Limitations

Several limitations exist in the research. Firstly, in the reference set construction phase, we have encountered situations where some webpages in the top 10 search results have already expired or prohibit crawler visits. Our solution is to replace such webpages by subsequent results, which may affect the similarity score for the focal projects. Secondly, although we have restricted the sample projects to be the latest ones, there is still time discrepancy between model result and human judgement: the model assesses the novelty at the time point it was launched, while the respondents evaluate the novelty based on his knowledge and perception “now”. The results will make more sense if two measurements at the same time point are compared. Thirdly, during model validation, since each project is evaluated by only two respondents, we cannot be sure that the average response of two persons is definitely representative for the general. If one project is evaluated by a large number of respondents, say, 100, then the average response will be closer to the true general human perception, and hence the comparison result will be more reliable and convincing.

Reference

Agrawal, A., Catalini, C., & Goldfarb, A. (2010). Entrepreneurial Finance and the Flat-World Hypothesis: Evidence from Crowd-Funding Entrepreneurs in the Arts. NET Institute Working Paper.

Amabile, T. M. (1996), Creativity In Context: Update To The Social Psychology Of Creativity. New Edition. Westview Press, Boulder, 1996.

Anderson, N., Potočnik, K. and Zhou, J. (2014). Innovation and Creativity in Organizations: A State-of-the-Science Review, Prospective Commentary, and Guiding Framework. Journal of Management. Vol.40, No.5, July 2014, pp. 1297-1333.

Barnett, C. (2014). 7 Crowdfunding Tips Proven To Raise Funding.

Berry, M. W., & Kogan, J. (2010). Text Mining: Applications and Theory.

Blaug, M. (2007). The Correlation between Education And Earnings: What Does it Signify?. Higher Education. Vol.1, No.1, February 1972, pp. 53-76.

Brin, S., & Page, L. (1998) The Anatomy of a Large-Scale Hypertextual Web Search Engine. In: Seventh International World-Wide Web Conference (WWW 1998), April 14-18, 1998, Brisbane, Australia.

Burtch, G. (2011). Herding Behavior as a Network Externality. In: Thirty Second International Conference on Information Systems, Shanghai, China.

Castells, P., Vargas, S., & Wang, J. (2011). Novelty and Diversity Metrics for Recommender Systems: Choice, Discovery and Relevance.

Chan, C. S. R. and Parhankangas, A. (2017), Crowdfunding Innovative Ideas: How Incremental and Radical Innovativeness Influence Funding Outcomes. Entrepreneurship Theory and Practice, Vol.41, No.2, March 2017, pp. 237–263.

Cholakova, M. and Clarysse, B. (2015), Does the Possibility to Make Equity Investments in Crowdfunding Projects Crowd Out Reward-Based Investments?. Entrepreneurship Theory and Practice, Vol.39, No.1, January 2015, pp. 145–172.

Dimov, D. and Murray, G.(2008). Determinants of the Incidence and Scale of Seed Capital Investments by Venture Capital Firms. Small Business Economics, Vol.30, No.2, February 2008, pp. 127-152.

Drake, D. (2015). 2,000 Global Crowdfunding Sites to Choose from by 2016: Top 5 Growth Indicators.

Freedman, D. M., & Nutting, M. R. (2015). A Brief History of Crowdfunding Including Rewards, Donation, Debt, and Equity Platforms in the USA.

Gensim. (n.d.). Similarity Queries.

Gerber, E.M., Hui, J.S., & Kuo, P.Y. (2012). Crowdfunding: Why people are motivated to post and fund projects on crowdfunding platforms. ACM Conference on Computer Supported Cooperative Work, Seattle, WA.

Gierczak, M., Bretschneider, U., & Leimeister, J.M. (2014). Is All That Glitters Gold? Exploring the Effects of Perceived Risk on Backing Behavior in Reward-Based Crowdfunding. In: Thirty Fifth International Conference on Information Systems, Auckland, New Zealand.

Hair, J.F., Tatham, R.L., Anderson, R.E. and Black, W. (2009). Multivariate data analysis. 7th Edition. Prentice-Hall, London, 2009.

Hart, S., Jan Hultink, E., Tzokas, N. and Commandeur, H. R. (2003), Industrial Companies' Evaluation Criteria in New Product Development Gates. Journal of Product Innovation Management. Vol. 20, No.1, pp. 22–36.

Heidenreich, S. and Kraemer, T. (2015). Passive innovation resistance: The curse of innovation? Investigating consequences for innovative consumer behavior. Journal of Economic Psychology. Vol.51, December 2015, pp. 134–151.

Hoffman, M. D., Blei, D. M., & Bach, F. (2010). Online Learning for Latent Dirichlet Allocation. Proceedings of the 23rd International Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, pp. 856-864.

Jacob. (2010). Part of Speech Tagging with NLTK Part 4 – Brill Tagger Vs Classifier Taggers.

Kickstarter. (2016). Stats. Retrieved from https://www.kickstarter.com/help/stats?ref=footer

Kimbia. (2014). Crowdfunding Donor Demographics Report.

Kleinschmidt, M. (2007). Venture Capital, Corporate Governance, and Firm Value. 2007 edition, Deutscher Universitätsverlag, 2007.

Kortum, S. and Lerner, J. (2000). Assessing the impact of venture capital on innovation. Rand Journal of Economics, Vol.31, No.4, winter 2000, pp. 674–692.

Leskovec, J., Rajaraman, A. and Ullman, J. Mining of Massive Datasets. 1st Edition. Cambridge University Press, Cambridge, 2011.

Lin, Y., Boh, W.F., & Goh, K.H. (2014). How Different Are Crowdfunders? Examining Archetypes of Crowdfunders and Their Choice of Projects.

Lopez-Mesa, B., & Vidal, R. (2006). Novelty Metrics in Engineering Design Experiments. In: International Design Conference - Design 2006, Dubrovnik - Croatia, May 15 - 18, 2006.

Madjar, N., Greenberg, E., & Cheng, Z. (2011). Factors for radical creativity incremental creativity and routine noncreative performance. Journal of Applied Psychology. Vol.96, No.4, Jul 2011, pp. 730-743.

Magnusson, P. R., Matthing, J. and Kristensson, P. (2003). Managing User Involvement in Service Innovation Experiments With Innovating End Users. Journal of Service Research. Vol.6, No.2, November, 2003, pp. 111-124

Marom, D., & Sade, O. (2013). Are the Life and Death of a Young Start-up Indeed in the Power of the Tongue? Lessons from Online Crowdfunding Pitches.

Massolution. (2015). 2015CF Crowdfunding Industry Report.

Metrick, A. and Yasuda, A. (2010). Venture capital and the finance of innovation. Second Edition, Wiley, New York, 2010.

Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing Order into Texts.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. Journal of

Business Venturing. Vol.29, No.1, January 2014, pp. 1-16.

Mollick, E. R. and Kuppuswamy, V. (2014). After the Campaign: Outcomes of Crowdfunding. UNC Kenan-Flagler Research Paper No. 2376997.

Pimentel, M. A. F., Clifton, D. A., Clifton, L., & Tarassenko, L. (2014). A review of novelty detection. Signal Processing. Vol. 99, June 2014, pp. 215–249.

Raîche, G., Walls, T. A., Magis, D., Riopel, M. and Blais, J. (2013). Non-Graphical Solutions for Cattell's Scree Test. Methodology. Vol.9, January 2013, pp. 23-29.

Riedl, J. (2013). Crowdfunding Technology Innovation. Computer. Vol.46, No.3, March 2013, pp. 100-103.

Sarkar, P., & Chakrabarti, A. (2006). Determining and measuring product novelty to ascertain innovation in product development cycle. In: National Conference on Design for Product Life Cycle-2006 Birla Institute of Technology and Science, Feb 17-18 2006, Pilani, Rajasthan, India.

Slaughter, E. S. (1998). Models of Construction Innovation. Journal of Construction Engineering and Management. Vol.124, No.3, May 1998, pp. 226.

Voelker, T. A., Niu, K., & Miles, G. E. (2011). Finding the Best Fit: Considering the Role of Novelty in Project Selection for Small Businesses. Small Business Institute Journal, Vol.7, No.2, October 2011, pp. 46.

World Bank. (2013). Crowdfunding's Potential for the Developing World.

Xiao, S., Tan, X., Dong, M., & Qi, J. (2014). How to Design Your Project in the Online Crowdfunding Market? Evidence from Kickstarter. In: Thirty Fifth

International Conference on Information Systems, Auckland, New Zealand.

Xu, A., Yang, X., Rao, H., Fu, W.-T., Huang, S.-W., & Bailey, B.P. (2014). Show Me the Money: An Analysis of Project Updates During Crowdfunding Campaigns. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems: ACM, pp. 591-600.

Zhang, L. (2013). The Definition of Novelty in Recommendation System. Journal of Engineering Science and Technology Review. Vol.6 No.3, January 2013, pp. 141-145.

Zvilichovsky, D., Inbar, Y., & Barzilay, O. (2013). Playing Both Sides of the Market: Success and Reciprocity on Crowdfunding Platforms. In: Thirty Fourth International Conference on Information Systems, Milan, Italy.

Appendix: Sample Survey

Survey Instructions:
Please complete the following 2 sections according to your own situation:

Section I: Demographic Information
Section II: Project Evaluation

Section I: Demographic Information

1. What is your gender?
   - Male
   - Female
2. What is your age?
   - Below 18
   - 18-24
   - 25-34
   - 35-44
   - 45-54
   - 55-64
   - 65-74
   - 75 and above
3. What is the highest degree or level of school you have completed? If currently enrolled, then the highest degree you received.
   - Below high school
   - High school
   - Bachelor degree
   - Master degree
   - Doctoral degree
4. What is your current employment status?
   - Employed
   - Self-employed / Entrepreneurs
   - Not employed
   - Student
   - Retired
5. To what extent do you know about crowdfunding? (1 = not familiar at all; 7 = know very well)
6. Have you ever visited Kickstarter website?
   - Yes
   - No
7. Have you ever invested in any new venture / entrepreneurial activities?
   - Yes
   - No

Section II: Project-based Evaluation

In this section, you will be given descriptions of 3 crowdfunding projects. Please read each description and answer the questions below.

Project 1555813533

["Well, I'm putting this description up rather than a video as literally my voice is so quiet you won't be able to hear me. Trust me, I tried making a video, I could barely hear myself. :DAnyway, onto what this project is about. Leaves Of Dreams is a creative hub based in Watton, run by myself and my husband. For the purposes of funding campaigns such as these I am the Social Media Officer. (Yep, we're only a start up at the moment but we have big ambitions, hence the fancy title) ;) :)We aim to support and promote local independent creatives of all stripes Some indeed not that local as we have one lady who is keen to get on board with us and she lives in Yorkshire but hopes to be sending work down in the near future.We know that in the vicinity local to us, which is Norwich, there are a number of galleries, bookshops etc but nothing quite like what we're doing and we think that's our selling point.Many shops will sell goods in a themed way, i.e books are sold as one thing, jewellery and clothes as another and so on. However, because the 'theme' of what we do is supporting creatives, we sell all of those things together. (Obviously in the main on behalf of the creatives we support) Even though it might seem odd, the people who have come in so far for whatever reason, have commented that it really seems to work well together.This is something we are really passionate about and we really want to make it work but in order to do that we need your help. We both feel awful to be asking for money on the one hand, but on the other hand part of why we're doing what we're doing is to create a 'creative community'. We both feel that 'professional image' (i.e. being slick, on top of everything all the time and completely independent.) only serves to leave people miserable. In contrast, by engaging within a 'creative community' model we hope to help people realise the importance of being interconnected and not isolated. i.e. being interdependent but stronger for it, which is really the far more truthful version of how everyone is pretty much all the time, rather than the 'professional' concept that has become ingrained in society through our current business models and practices.Fortunately, we are already in a good position with regard to the location of our shop. As a rural community, a lot of Watton's shops are independent and we have been warmly welcomed by other shopkeepers there and are even starting now to really get noticed by the public bit by bit.In short, we know we can make this happen but to see it through we need your help.Overall this campaign is something we hope/expect to function as an ongoing 'fund raising' effort. In the end we hope we may be generating enough revenue on our own that we don't necessarily need it but in the early stages it seems like a good start. So, you've read about what we do, now let's see what we can all do together! :)"]

1. What is your first impression of this project/product?
   - Very positive
   - Somewhat positive
   - Neutral
   - Somewhat negative

- Very negative

2. Rank the extent to which you agree with the following statements (1 = strongly disagree, 7 = strongly agree)
    - This project/product is original.
    - The project/product is radically new.
    - There are no other project/product like this on the market right now.
    - The individual or team behind this project/product suggests radically new ways for doing things.
    - The individual or team behind this project/product is a good source of highly creative ideas.
    - The project/product is a modification of an existing project/product.
    - The individual or team behind this project/product easily modifies previously existing work processes to suit current needs.
    - The project/product is quite similar to existing project/products on the market.
    - The project/product is an incremental improvement over an existing project/product.
    - The individual or team behind this project/product is very good at adapting already existing ideas.
    - The individual or team behind this project/product uses previously existing ideas or work in an appropriate new way.
    - The individual or team behind this project/product demonstrates originality in his/her/their work.

Project 945311721
["Introducing WooBots - a high-quality wood robot collection designed for action-packed fun, just in time for Christmas. Play, transform and enjoy!About WooBots They're new, they're full of energy and they're taking over the streets. The WooBots have hit the ground to change the game for everyone. WooBot is a collection of premium wooden robots specially crafted to transform into different states in just a few seconds. Each character is totally unique in design, and the limits to creativity are endless. This is a toy for people of all ages, and one that will make the most amazing gift for Christmas this year. Meet WooBots Family Beetle: The youngest member of the team, Beetle is full of youthful enthusiasm, and can move at incredible speeds! Truck: This guy always knows where he's headed and is the one to lead the crew into battle. He becomes a sturdy truck in just a few seconds. Warship: He's tough and unstoppable, and contains the most awesome weapon of the whole team. Watch out when he's angry AutoBus: Big, robust and reliable, AutoBus is one of the bedrocks of the team who you'll always want on your side. He's always there with backup for the team. Jet Fighter: They'll never stop this guy when he's in jet mode! He's aerodynamic in flight, and always ready for action on the ground. Play and Transformation All of the WooBots can turn into two different pre-determined forms, but what you do with them is really up to your imagination, making them not only a fun toy or a collector's item, but also a tool for developing creativity. And we're not just talking about little ones. the WooBots are just as much fun for grown ups too! Why Did We Use Wood?

It all started from a piece of wood, why? Because gone are the days of cheap and nasty plastic toys. The wood we use is of the highest quality, making it durable for years to come.Concept and Design We've spent a huge amount of time perfecting these toys, and though it wasn't an easy feat, we've created something we're very proud of. Modified Blueprint Each toy is 13-15cm tall and made of 15 to 20 wooden blocks, cut by a laser machine for precision. We've assembled the toys by hand to ensure that each of them moves and transforms perfectly."]

1. What is your first impression of this project/product?
   - Very positive
   - Somewhat positive
   - Neutral
   - Somewhat negative
   - Very negative

2. Rank the extent to which you agree with the following statements (1 = strongly disagree, 7 = strongly agree)
   - This project/product is original.
   - The project/product is radically new.
   - There are no other project/product like this on the market right now.
   - The individual or team behind this project/product suggests radically new ways for doing things.
   - The individual or team behind this project/product is a good source of highly creative ideas.
   - The project/product is a modification of an existing project/product.
   - The individual or team behind this project/product easily modifies previously existing work processes to suit current needs.
   - The project/product is quite similar to existing project/products on the market.
   - The project/product is an incremental improvement over an existing project/product.
   - The individual or team behind this project/product is very good at adapting already existing ideas.
   - The individual or team behind this project/product uses previously existing ideas or work in an appropriate new way.
   - The individual or team behind this project/product demonstrates originality in his/her/their work.

Project 489384530

["uprice.ca is the first website that allows customers to pick the price they're willing to pay for products. By eliminating costs and grouping customer orders, we make it possible to get top quality products for the lowest possible prices. What Makes uprice.ca Different? We do not mark up the price of products on our site. If we get it for $10, you get it for $10. We are able to do this by buying as a community and placing one large order rather than hundreds of small orders. This eliminates all of the unnecessary costs associated with getting product to you. What

Type Of Products will uprice.ca carry?Upon launch, we will carry products in the following categories: Electronics Apparel Books/Music/Video Toys and Hobby Furniture and Home Furnishings MembershipsAll offers on uprice.ca will be exclusive to members only. You can become a regular member or support our Kickstarter campaign to become a founding member. TimelinesBy working closely with our web development partner, we are scheduled to launch on April 1st 2016."]

1. What is your first impression of this project/product?
    - Very positive
    - Somewhat positive
    - Neutral
    - Somewhat negative
    - Very negative

2. Rank the extent to which you agree with the following statements (1 = strongly disagree, 7 = strongly agree)
    - This project/product is original.
    - The project/product is radically new.
    - There are no other project/product like this on the market right now.
    - The individual or team behind this project/product suggests radically new ways for doing things.
    - The individual or team behind this project/product is a good source of highly creative ideas.
    - The project/product is a modification of an existing project/product.
    - The individual or team behind this project/product easily modifies previously existing work processes to suit current needs.
    - The project/product is quite similar to existing project/products on the market.
    - The project/product is an incremental improvement over an existing project/product.
    - The individual or team behind this project/product is very good at adapting already existing ideas.
    - The individual or team behind this project/product uses previously existing ideas or work in an appropriate new way.
    - The individual or team behind this project/product demonstrates originality in his/her/their work.