B.Sc. Dissertation

Crowdfunding Recommendation Network:
Analysis of the Antecedents and Consequents of Peer Recommendations

By

Yang Xiya

A0143972Y

Department of Information System and Analytics

School of Computing

National University of Singapore

2018/2019

Project No: H175340
Advisor: Assoc Prof Hahn Jungpil

Deliverables:
Report: 1 Volume

Abstract

Crowdfunding is playing an increasingly important role in the world of entrepreneurial financing. Past literature has mainly focused the various success factors as well as the motivations of the platform participants but ignored the phenomenon of peer recommendations. However, peer recommendations in the context of crowdfunding network could have significant impact on the fundraising outcome and the overall platform dynamics due to the word-of-mouth effect and the reinforcement effect. In this study, a systematic methodology is adopted to analyse the various antecedent characteristics of project recommendations. Furthermore, the study also examines the impact of peer recommendations on the crowdfunding performance of the project and also seeks to verify the validity of the reinforcement model.

Keywords: crowdfunding, peer recommendation, reinforcement model, clustering, panel regression

Acknowledgement

I would like to express my deepest gratitude to my supervisor Associate Professor Hahn Jungpil for his valuable guidance throughout the whole project, without which the project would not come to fruition. I would like to thank him for the critical advice raised as well as the unwavering support throughout the project.

I would also like to thank Ms Yang Lusi for the valuable advice, which helped to shed light on the paths that the project could take. I would also like to thank Mr Wang Zhiyi greatly for his indispensable advice as well as the stimulating discussions we had, without which the project would not come to be what it is right now.

I would like to express my gratitude to my friends, Zhong Mengjia and Zhu Ziwei, for the numerous discussion sessions together, without whom the journey to complete the project would no doubt be less interesting and intellectually stimulating.

I would like to thank my main evaluator Associate Professor Huang Ke-Wei, whose valuable advice and feedback during and after the midterm presentation helped to shed light on the various ways to increase the depth and the value of the project, making the experience a more fulfilling one.

Table of Contents

1. Introduction

Crowdfunding is playing an increasingly important role in the world of entrepreneurial financing (Giudici et al., 2012). It disseminates the risks the traditional venture capitalists face to a distributed crowd of early adopters and leverages the web technologies of the rich-media-laden Internet 2.0 era to forge a virtual community that nurture and support a project throughout its birth.

Crowdfunding has been recognised to have significant potentials for changing the world (Best, 2013). It hands the power of financing to the crowd and let them decide which ideas are worth funding and to bring alive. This is a form of empowerment not only for the crowd but also for the project creators themselves. By tapping funds from a wide audience across the globe, many entrepreneurs are likely to have access to a financing market that they would otherwise not have had access to.

Moreover, the importance of crowdfunding platforms is also shown in the fact that they allow the investors to have greater control over their own project (Gerber & Hui, 2013), in contrast to the traditional financing avenue where a small group of investors on an idea have high bargaining power. Furthermore, the value of the crowdfunding platforms is also imbued in the connections that are built between the project creators and the backers. These connections have motivated the project creators and the backers alike to stay focused and committed to the projects they pledge their time or money to through communications and relationship building (Gerber & Hui, 2013).

Since crowdfunding brings unique advantages and significant value to various groups of project creators and backers, it has become critical to determine the success factors that influence the fundraising outcome. Many studies have been carried out, with a number coming to the conclusion that project-level attributes that signal project quality are of great importance in determining fundraising success (Koch & Siering, 2015; Bi et al., 2017). Even though project-level factors are studied the earliest, the social aspect of fundraising has also received more and more attention due to the equally important role it plays in determining fundraising success (Hong et al., 2018; Koch & Siering, 2015; Kromidha & Robson, 2016; Thies et al., 2016).

The highly interactive nature of the crowdfunding platforms has not only fostered creator-to-backer interaction but also creator-to-creator interaction. In particular, project creators sometimes mention and recommend other projects to their own backers during project updates. This project aims to understand the effect of peer project recommendations on the fundraising outcome of crowdfunding projects. The fundraising process is usually based on an all-or-nothing criterion, where the project is fully funded if the pledged goal is achieved. Otherwise, the project will not receive any of the funds pledged by the contributors.

Another research question the study aims to address is the characterisation of the recommendation behaviour of the project creators in order to interpret the motivations of their recommendations. There are two classes of competing models, namely the reinforcement model and the substitution model, that attempt to explain the rationale behind which individuals contribute to a public good (Shang & Croson, 2009). In this study, these models are extended to explain the peer recommendation behaviour of the project creators. The reinforcement model specifies that the project creators are encouraged to make more recommendations after they have derived sufficient utility following recommendations of their own project. On the other hand, the substitution model theorises that the project creators are likely to decrease their recommendation of others' projects after receiving more recommendations from others because their marginal utility of making another recommendation decreases. This could be explained by their motivation of obtaining mutual recommendations to boost the project's fundraising level.

Understanding the motivations behind fundraising on a crowdfunding platform has important policy implications on the incentives that the platform could design to encourage contribution and community forming. Moreover, for the individuals, understanding the effect of peer recommendation and other project characteristics on fundraising success and the reasons why others make recommendations enable the individuals to improve his or her strategy in improving funding levels and reaching targets.

This report will go into details about the current literature related to crowdfunding and identifies the gap addressed by this study. Then, the methodology adopted is explained before the presentation of results. This is followed by the discussion of implications and limitations.

2. Literature Review

Previous works have studied user motivations and behaviour on crowdfunding platforms from various angles. One group of studies have examined the reasons behind which project creators and backers are drawn to crowdfunding platforms instead of traditional financing avenues (Section 2.1). Another group of studies have analysed the factors that contribute to the funding success of the projects listed on the platform, with the more recent ones highlighting the importance of social factors on fundraising success (Section 2.2).

On the other hand, there has also been abundant literature on the design and impact of recommendation engines in general e-commerce context, which demonstrates the importance of recommendations on consumer behaviour (Section 2.3). In the context of crowdfunding, the project creators have the choice to recommend projects to their own backers. This creates a word-of-mouth effect that has been studied in previous works to influence consumer decisions in various ways (Section 2.4).

Despite the importance of the social effect of peer recommendations among project creators, it has not been examined in previous works. This study proposes a method to address this gap in existing literature. One of the objectives is to determine the impact of peer recommendations on fundraising outcome. Another objective of the study is to analyse the social effect of peer recommendations based on two competing classes of economic models, the reinforcement model and the substitution model (Section 2.5).

The following sections will illustrate the findings of existing studies in detail wherever relevant to the current study, after which the gaps in the existing literature and the contributions that this project aims to make will be highlighted (Section 2.6).

2.1 Motivations of Platform Participants

To study user behaviour on crowdfunding platforms, it is important to understand the value that crowdfunding platforms provide to its participants. Previous works have pointed out that some of the motivations for project creators to use crowdfunding platforms include a desire to expand awareness of work, connect with others, gain approval, maintain control over work and learn (Gerber & Hui, 2013). It has also been highlighted that online crowdfunding platforms facilitate interaction between users, which is essential for project creators to reach out to networks of investors or consumers (Belleflamme et al., 2010).

On the other hand, the motivations for backers to fund the projects have also been analysed. These include a desire to collect rewards, help others, support causes, and be part of a community (Gerber & Hui, 2013). It has also been found that the connections and interactions reinforce the commitment of the project creators and backers, motivating them to participate on crowdfunding platforms (Gerber et al., 2012).

2.2 Contributory Factors of Fundraising Success

Due to the value that crowdfunding platforms provide to both project creators and backers, there have also been various studies that focus on the factors that impact fundraising success on the platforms. These include project-related attributes such as project description, related images and videos (Koch & Siering, 2015), presence of spelling errors (Mollick, 2014), specific uses of project updates (Xu et al., 2014) and reward scheme design (Thürridl & Kamleitner, 2016; Wang, Yang, Kang & Hahn, 2016). A common attribute in most factors leading to fundraising success is that they effectively signal the high quality of the project (Koch & Siering, 2015; Bi et al., 2017).

Recent studies have started to examine the social aspect of fundraising. It was found that project creators are more successful in having their projects funded if they have backed other projects before (Koch & Siering, 2015). Moreover, bidirectional social interactions between project owners and backers also have a stronger positive impact on fundraising performance than unidirectional communication of project updates (Kromidha & Robson, 2016). Moreover, the effects of different social media platforms on fundraising success has been studied (Hong et al., 2015). It has also been found that backers who identify with the projects in their own social network are likely to pledge more, contributing to fundraising success (Kromidha & Robson, 2016). In addition, the personality signals that the project creator sends such as agreeableness and openness can have a positive impact on the project's fundraising performance (Thies et al., 2016). Moreover, some of the literature has also studied the impact of social network structures on crowdfunding success (Hong et al., 2018).

Besides these factors, previous works have also examined how overwhelmingly successful projects can affect the fundraising success of other projects. It has been found that these projects tend to have spillover effects on other projects within the same category but can have cannibalisation effects on projects in different categories (Liu et al., 2015).

2.3 Recommendation Engines and Recommendation Networks

Recommendation engines have become ubiquitous on today's online platforms. Various studies have proposed different recommendation engine design paradigms including collaborative filtering (Sarwar et al., 2001), Bayesian learning (Schwaighofer et al., 2005), supervised learning (Benchettara et al., 2010) and reinforcement learning (Shani et al., 2005).

The proliferation of recommendation engines has prompted studies to understand the value they bring and analyse the effects they have on human behaviour. Prior studies have mainly focused on the recommendation networks on e-commerce platforms. One study highlights that the strength and recency of received recommendations have a positive effect on sales (Pathak et al., 2010), while another study found that recommendations tend to have a stronger positive impact on less popular products (Oestreicher-Singer & Sundararajan, 2012). It has also been shown that the diversity of products in recommendations have a positive impact on demand in the co-purchase network (Lin et al., 2015).

2.4 Impact of Word-of-Mouth on Purchase Decisions

Previous research has investigated the impact of online word-of-mouth on sales performance. It has been found that there is a positive reinforcement effect between product performance and word-of-mouth (Duan et al., 2008). That is, better-performing items tend to attract a higher volume and greater sources of word-of-mouth, which contributes to even higher revenues. This highlights the importance of understanding the effect of product recommendations on crowdfunding performance in this project. Moreover, research has also shown that product categories with higher volumes of word-of-mouth tend to attract more purchases than product categories with lower volumes of word-of-mouth (Davis & Khazanchi, 2008), which suggests the need to control product category as well as other confounding variables in the current study. Moreover, word-of-mouth has also been shown to have an important positive impact on project success in the crowdfunding context, along with signals of project quality (Bi et al., 2017).

2.5 Reinforcement Models and Substitution Models

One of the important factors that mediate crowdfunding success is the social information about others' contributions. It is valuable to examine how the availability of information of others' contributions might affect the contributions of a particular individual in order to understand the underlying fundraising process. In the current study, it is also of value to break down how the presence of others' recommendations to a project impacts an individual project creator's decision to recommend the project. Two classes of models have been proposed to address the impact of social information in the public goods literature. The first class of models posit the presence of the reinforcement effect, where contributions of others are positively correlated with individual contributions, while the second class of models claim the presence of the substitution effect, where the knowledge of the others' contributions leads to lower individual contributions (Shang & Croson, 2009).

2.5.1 Reinforcement Models

The reinforcement models have been put forth based on different hypotheses about the underlying consumer behaviour in the context of public goods. One study viewed the phenomenon from the perspective of reciprocity, where the contributors are morally obliged to make at least the least amount of contributions that others are making (Sugden, 1984). When the information of the least generous contributor is unknown, social information may affect an individual's belief about this amount and increase their contributions (Shang & Croson, 2009). However, if the principle of reciprocity holds and the contribution of the least generous person is known, social information about others' contributions will always decrease the lower bound on the individual's contribution, which may lead to a negative impact on the individual's contribution amount instead.

Others have claimed that in a social context, individuals care about not only the utility they derive from consumption but also others' perceptions of their dispositions and their social status (Bernheim, 1994). It is argued that any slight deviations from social norm can impair an individual's status. In the context of crowdfunding, this serves as an incentive for individuals to contribute disproportionately more in comparison to the actual utility they obtain. Other conformity models have also theorised that deviations from social norms can directly impact an individual's utility (Akerlof, 1982; Jones, 1984)

Another study has also found that others' contributions in the general fundraising context signal the quality of the project, thereby attracting more contributions and accelerating fundraising process (Vesterlund, 2003). These studies provide the theoretical foundations to assess the reinforcement effect of peer recommendations in this study should it exist.

2.5.2 Substitution Models

The substitution models have been based on the individual's need to maximise utility by adjusting contributions to public goods and private goods (Roberts, 1984; Warr, 1982). Earlier studies have claimed that individuals reduce their own contributions as others' contributions to a public good increase on a dollar-by-dollar basis in order to maximise their own utility (Roberts, 1984; Warr, 1982). Later research has put forth a modified model which states that individuals' contributions to a public good decrease by less than the amount that others' contributions increase due to the positive psychological effect that individuals receive from giving (Andreoni, 1989; Andreoni, 1990).

Another perspective shared by researchers is that the provision of public goods is hinged upon a certain amount of fixed costs, above which individuals' contributions are no longer crucial in securing the provision of the public goods (Andreoni, 1998). Moreover, others' contributions to the minimum required fixed costs can be substitutes for one's own, leading to a negative correlation between others' contributions and individual contributions (Cornelli, 1996; Romano, 1991).

With regards to crowdfunding platforms, research has shown that substitution effects dominate for primarily donation-based crowdfunding markets (Burtch et al., 2013), where individuals experience a decrease in their marginal utility as contributions from others increase. However, more recent research has found that there are significant reinforcement effects in reward-based crowdfunding markets, especially when taking into account the social buzz that the projects receive (Thies et al., 2014). In addition, it has been shown that an increase in the dollar amount contributed per day contribute to fundraising success, corroborating the validity of the reinforcement model (Cordova et al., 2015).

2.6 Gaps in Literature and Contributions

2.6.1 Formulation of the First Research Objective

As established in the existing literature, the crowdfunding platforms provide significant value to both project creators and backers who have various motivations to participate in such platforms. Therefore, it is important to understand the various factors contributing to crowdfunding success. While project-level factors have been studied in detail, more recent research has highlighted the importance of social connections and social network structure on crowdfunding performance. One focal area of social relationships is the phenomenon of peer recommendations among project creators. However, existing literature has only focused on recommendation algorithms and their effects on customer behaviour in the context of co-purchase or co-view networks. The recommendations coming from project creators can potentially have a more positive impact on the performance of the recommended project than a recommendation produced by an algorithm. This is because the project creators have established various degrees of social relationships with their backers, which can have a positive stimulus on recommendation adoption through the well-studied word-of-mouth effect, and word-of-mouth has been shown to have a significant impact on product performance that is comparable to other factors that signal product quality. Therefore, the first objective of the study is to determine the impact of peer recommendations on fundraising success.

2.6.2 Formulation of the Second Research Objective

In addition, while research has studied the application of the reinforcement models and the substitution models to crowdfunding from the backers' perspective. It remains unclear what the motivations the project creators have to recommend others' projects from the lens of the two classes of models. There are two competing hypotheses with regards to the project creators' peer recommendation behaviour.

<u>Hypothesis 1: the project creators reduce their voluntary recommendations of others' projects as they receive more recommendations due to the decrease in marginal utility.</u>

This hypothesis is supported by the substitution model, where the contributions to the crowdfunding community through giving recommendations reaches a point of diminishing returns for the project creators. Giving more recommendations improves the project creators' private utility to a smaller and smaller extent as they have already garnered sufficient attention from the peers.

<u>Hypothesis 2: the project creators tend to make more recommendations of others' projects as their own project is featured in more recommendations.</u>

This is supported by the reinforcement model and indicates the prevailing existence of community spirit, in that receiving more recommendations stimulates the project creators to give back in return. Therefore, the second objective of the study is to empirically examine which hypothesis is valid in the context of crowdfunding.

3. Research Methodology

3.1 Research Approach

As discussed in 2.6, this study aims to achieve two research objectives. The first concerns the impact of peer project recommendations and the network influence of the recommenders on a project's fundraising performance. The second is about assessing the validity of the reinforcement model or the substitution model in explaining the recommendation behaviour of the project creators. Therefore, this study will provide empirical evidence to support one of the two formulated hypotheses and to refute the other.

After defining the research objectives, the data requirements are studied. Then, data extraction is performed, which is followed by data preprocessing, including critical tasks such as converting the currency of all projects to US Dollars and removing projects that are still ongoing for project-level regression. After the data is consolidated and ready, various models are used to answer the two proposed research questions.

In order to systematically code the motivations for peer project recommendations, feature extraction and transformation are performed based on the relevance of the various features to motivate project recommendations. K-means clustering is then used to determine the characteristics prevalent in the dominant groups of project recommendations in the context of recommendation motivations. The coded cluster groups are then further transformed to be used in subsequent analysis.

To determine the impact of peer recommendations and network influence of recommender on fundraising success, the outcome variable needs to first be operationalised as the logarithmically transformed pledged amount. The regression forms are determined and specified before the data is transformed into the features to be used in the various regression models. These regression models include linear regression at the project level, panel regression with fixed effects. Furthermore, propensity score matching is also conducted to confirm the results obtained (Figure 1).

In order to assess the validity of either the reinforcement model and the substitution model in explaining the recommendation behaviour of the project creators, the panel regression model is used to determine the impact of prior received recommendations for each project on the amounts of ensuing recommendations given to other projects (Figure 1).

After the models are executed, the results from running the model are interpreted in order to identify potential issues in processing the data and to identify the effects of the independent variables of interest on the respective outcome variables to be specified in the following sections.

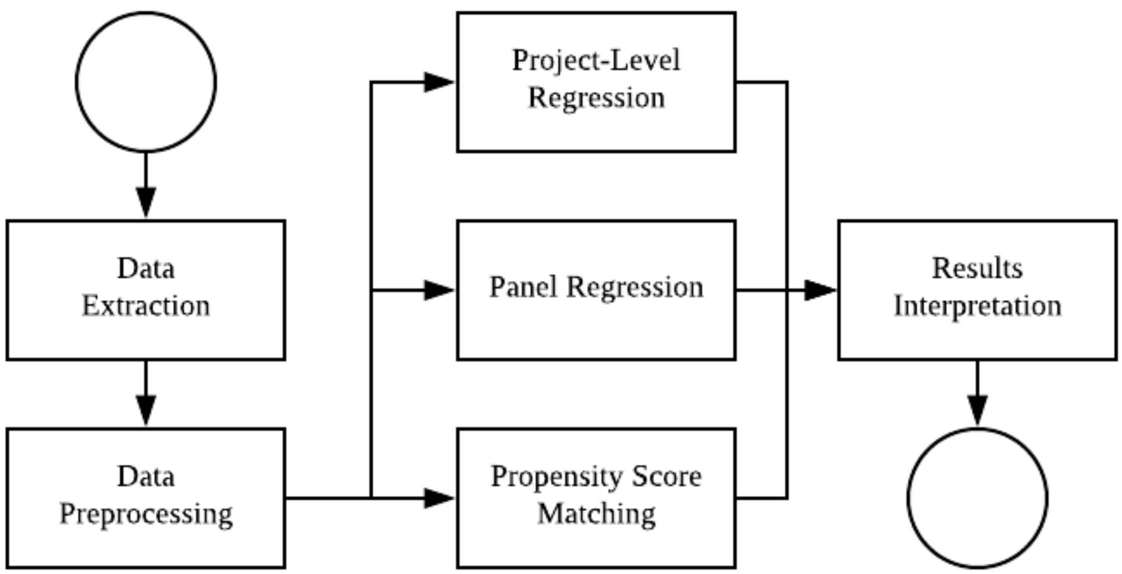

Figure 1 Overview of Research Approach

3.2 Data Specifications

3.2.1 Data Collection

The collected data comprises comprehensive project information from June 2015 to September 2018. First and foremost, project updates within this period for all Kickstarter projects are collected to identify the occurrence of peer recommendations. The dataset includes information on project ID, update URL, update time, update title, update description, image count, video count and embedded URLs. A group of project updates are only for backers and their content is not available for collection. Therefore, these updates are excluded and the study is only focused on the publicly available project updates.

In addition, the metadata of all projects with updates in this period is also collected. The metadata provides information on the project category, country of origin, funding goal, amount pledged, number of backers, duration, project description, image count and video count. Moreover, to control for the dynamic pattern of fundraising performance through panel data analysis, the data on the amount of funds raised for each project on each date within the stated period is also collected.

3.2.2 Data Description

The projects are distributed unevenly across the major categories. As can be seen in Fig. 2, the majority of the projects (13667) are from the Games category, followed by Design (10067).

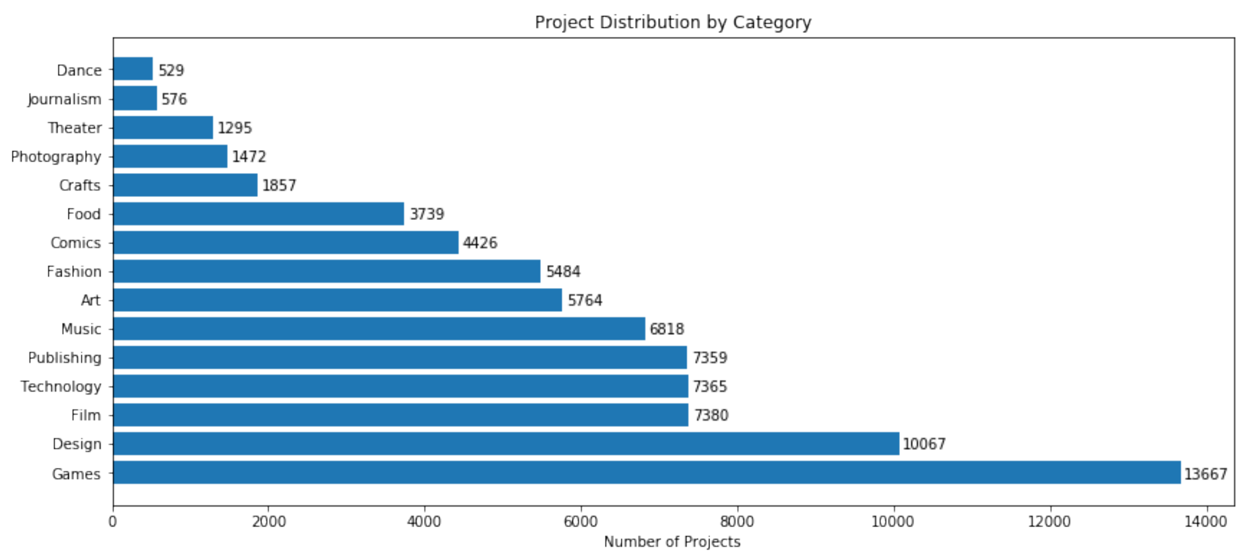

Figure 2 Project Distribution by Category

In terms of geographical distribution, the majority of the projects are based in the United States (63.44%), followed by the United Kingdom (11.53%) and Canada (5.05%), as seen in Fig. 2.

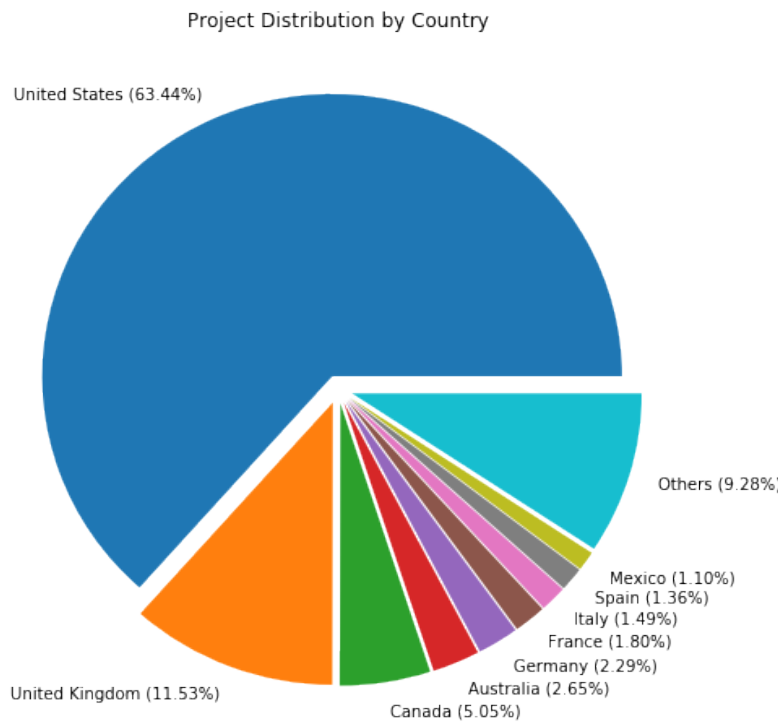

Figure 3 Project Distribution by Country

Furthermore, the average values for the set goal, the pledged amount and the number of backers are also explored across the different project categories (Fig. 5). It can be seen that the Technology category has the highest average set goal, while the Design category has the highest average pledged amount and the Games category has the highest average number of backers. These categories are also among the top 4 categories in the number of projects (Fig. 5).

3.2.3 Recommendation Network and Centrality Measures

The datasets described above are used to construct a peer recommendation network for network characteristics extraction. Each project creator is a node in the network, while the acts of project recommendations are directed edges in the network. Project recommendations occur in the project updates when project creators promote others' projects and provide a hyperlink to the recommended project.

The structure of the recommendation network changes over time as more recommendations occur, adding edges to previously unconnected nodes. Therefore, the dynamic nature of the recommendation network favours the use of a panel regression model, where network-related measures are computed for each time step to serve as independent time-variant effects in the regression model. This approach ensures that the network centrality measures accurately reflect the network structure within the fundraising period for each project. However, as will be discussed later, a general linear regression model without time-variant effects will also be fitted. In this case, the network-related measures at the last time step will be used as independent variables.

The study focuses on the network centrality measures of the project creators initiating peer recommendations and assesses their impact on fundraising performance. Since the recommendation network is a directed graph, each of the centrality measures has two different versions. One is computed with regards to the incoming edges while the other is computed with regards to the outgoing edges. Network centrality measures computed based on incoming edges measure the influence of the project based on the attention they receive, while network centrality measures computed based on outgoing edges measure the importance of the project in terms of the recommendations it endows upon other projects. In particular, degree centrality, harmonic centrality, eigenvector centrality and betweenness centrality measures are used.

Degree centrality is the simplest network centrality measure among the above-stated measures. It is the normalised counts of immediate neighbours attached to the node to be measured. Therefore, it is a local centrality measure. The formulae are given below, where the degree centrality of the i^{th} node is equal to the number of incoming or outgoing edges divided by the maximum number of degree possible (n-1) where n is the number of nodes in the graph. In the context of recommendation network, the in-degree centrality measures the centrality of each project based on how many other projects have recommended it while the out-degree centrality measures the centrality of each project based on how many other projects have been recommended by it.

$$DegreeCentrality_{in,i} = \frac{d_{in,i}}{n-1} \qquad DegreeCentrality_{out,i} = \frac{d_{out,i}}{n-1}$$

Harmonic centrality measures how far a node is to all other nodes in the graph. It is a modified version of the closeness centrality adapted to graphs with disconnected components (Rochat, 2009). The closeness centrality measure computes the reciprocal of the sum of shortest distances from a node to all other nodes. In graphs with disconnected components, the computed value approaches zero for all nodes in the graph. On the other hand, harmonic centrality takes the reciprocal of the shortest distances from one node to all other nodes first and then sum up the values. As such, only summands for distances to nodes in disconnected components will be zero, so the computed harmonic centrality measure will be non-zero for different nodes and can be used for comparison.

In the case of recommendation network, two versions of harmonic centrality can also be formulated. In the first version, the directions of the edges are reversed before computing harmonic centrality. In this version, nodes with high centrality scores have shorter recommendation chains coming from other nodes in the network. On the other hand, a second version can also be formulated, where harmonic centrality is computed based on the recommendation network with edges going from the recommendation provider to the recommendation recipient. In this version, nodes with high centrality scores have shorter recommendation chains going to other nodes in the network. The formulae for computing harmonic centrality measure are given below. $dist'(i,j)$ represents the shortest distance between the i^{th} node and the j^{th} node in the recommendation network with reversed edge directions, while $dist(i,j)$ represents the shortest distance between the i^{th} node and the j_{th} node in the unmodified recommendation network.

$$HarmonicCentrality_{in,i} = \frac{n-1}{\sum_j \frac{1}{dist'(i,j)}} \qquad HarmonicCentrality_{out,i} = \frac{n-1}{\sum_j \frac{1}{dist(i,j)}}$$

In the context of recommendation network, projects with high in-harmonic centrality values are able to be reached by the backers from other projects with less steps on an aggregate level. On the other hand, projects with high out-harmonic centrality values allow their own backers to reach other projects in the network with less steps on an aggregate level.

Eigenvector centrality is a measure where a node's centrality is its summed connections to others, weighted by their centralities (Bonacich, 1987). Similarly, there are also two formulations of eigenvector centrality. The first formulation dictates that the centrality of a node is the sum of the centralities of its downstream neighbours connected via outgoing edges, while the second formulation states that the centrality of a node is the sum of the centralities of its upstream neighbours connected via incoming edges. Let e denote the centrality vector of all nodes and let A denote the adjacency matrix. e and A satisfy the respective conditions specified below.

Formulation 1: $\lambda e = Ae$ Formulation 2: $\lambda e = A^T e$

λ is introduced such that the equations above have a non-zero solution if not all eigenvalues of A are zero. The largest eigenvalue is usually the preferred one (Bonacich, 1987). These equations can be solved numerically via power iteration. In the context of recommendation network, a greater proportion of backers tend to visit projects with high in-eigenvector centrality values. On the other hand, projects with high out-eigenvector centrality values distribute a greater proportion of backers to other projects in the network.

Furthermore, the pagerank centrality is also adopted. It is a variant of the eigenvector centrality by creating weakly connected edges across all pairs of nodes (Page, Brin, Motwani & Winograd, 1999). This creates connections across the disconnected components and results in a finite, irreducible and aperiodic Markov chain with a unique stationary distribution. In particular, the modified transition probability matrix, $\bar{P}$, is obtained from the original transition probability matrix, P, through the following equation. Note that the transition probability matrix is simply the normalised version of the adjacency matrix.

$$\bar{P} = \alpha P + \frac{1-\alpha}{n}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

The parameter, α, determines the extent to which the original transition probabilities are preserved. Similarly, two versions of pagerank centrality can be formulated as shown below to represent the centrality of the node in terms of its downstream neighbours (Formulation 1) or its upstream neighbours (Formulation 2) respectively.

Formulation 1: $\lambda e = \bar{P} e$ Formulation 2: $\lambda e = \bar{P}^T e$

The interpretation of the pagerank centrality in the context of recommendation network is the same as the interpretation of the eigenvector centrality.

3.3 Detailed Methodologies

3.3.1 Feature Selection

Feature selection will be performed to address the possible challenge of multicollinearity. For example, multicollinearity could occur as the above-stated network centrality measures are all concerned with the influence of the node in the recommendation network. Therefore, there can be high correlations among the network centrality measures. Moreover, multicollinearity could also exist among other independent variables. Multicollinearity hinders the correct identification of the true effects of the independent variables on the outcome variable. Moreover, multicollinearity can also lead to greater uncertainties in the regression coefficients estimated. In order to overcome the challenge, feature selection is required to filter out independent variables that are less important in determining fundraising success.

Two approaches will be adopted. The first is Least Absolute Shrinkage and Selection Operator (LASSO) regression. LASSO regression can be applied each of the specified regression forms where relevant through the introduction of an L1 regularisation term. The regularisation term constrains the feasible region of the regression parameters and enables the regression coefficients to shrink to zero, keeping only the most important coefficients. Therefore, LASSO regression serves as a feature selection strategy if independent variables introducing high collinearity can be ruled out in the feature selection process.

However, in practice, LASSO regression could be biased to be durable with multicollinearity. Therefore, Principal Component Analysis (PCA) will also be carried out if LASSO regression is proven to be ineffective in handling multicollinearity. PCA is a method of linear transformation such that the variance of the data points along each of the principal components is sequentially maximised. In particular, the principal components are along the directions of the eigenvectors of the covariance matrix sorted in the order of the magnitude of the eigenvalues that they are associated with. Specifically, in order to perform feature selection with PCA, a required proportion of variance to be preserved such in the range of 70-90% will first be selected (Rea & Rea, 2016). The minimum number of principal components such that the threshold proportion of variance can be preserved will be selected. Identify the independent variables with the greatest proportions of variance projected onto these principal components and select them as the features to be used in the regression models.

In particular, for each regression model to be specified later, we start with the full model. Adjust the strength of the regularisation strength parameter to allow the sequential exit of the less important variables in determining the outcome variable for LASSO regression. For PCA, identify the variable with the least contribution in proportion of variance to the top selected principal components as the exit variable. Check the Variance Inflation Factor (VIF) after the exit of each variable until all variables have acceptable VIF values not exceeding 10 (O'brien, 2007). The threshold may be adjusted in model building if it is deemed to be too large to exclude variables introducing high collinearity. Then, the algorithm stops and the remaining features are used in the reduced model for identification of the coefficients of interest. Last but not least, the various groups of independent variables with high collinearity may also be grouped together in order to specifically exclude the redundant variables to handle the issue of multicollinearity.

3.3.2 Clustering Analysis

The objective of clustering analysis is to systematically determine the different categories of recommendation motivations. K-means clustering is adopted for this task to group recommendations into similar clusters based on extracted motivation-related features (Algorithm 1).

```
Select a number of clusters (k)
Initialise k centroids randomly
repeat
    for every data point do
        for every centroid do
            calculate the distance between the data point and the centroid
        assign the data point to the cluster with the nearest centroid
    for every centroid do
        recalculate the centroid of each cluster based on reassigned data points
until no data point is reassigned in the current iteration
```

Algorithm 1 K-means Clustering

The internal clustering validation metric of silhouette score is used to select the optimal value of the hyper-parameter, K. The computation of silhouette score starts with calculating the average distance of each point, *i*, to every other point, *j*, in the same cluster.

$$a(i) = \frac{1}{|C_i| - 1} \sum_{j \in C_i, j \neq i} d(i, j)$$

In the formula, $|C_i|$ denotes the number of points in the cluster which *i* belongs to. *d(i,j)* denotes the distance between point *i* and *j*. Afterwards, the value of *b(i)* is also calculated.

$$b(i) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} d(i, j)$$

In the computation of *b(i)*, the average distance of point *i* to all points within another cluster is calculated. The computation is repeated for every cluster different from the one that point *i* is in and the minimum of the average distance is then taken as the value of *b(i)*. Following the computation of *a(i)* and *b(i)*, the silhouette score associated with the point I is the computed in the formula below.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

The score is the scaled value that incorporates information on the cohesion and separation of the clusters based on a single point, *i*. The score will be high if the point, *i*, is close to other points within the same cluster and distant from other points in every other cluster and vice versa.

To assess the clustering performance based on different values of K, the silhouette score is averaged across all points in the clusters obtained after running K-means based on a chosen K. Then, the value of K that produces the highest silhouette score is then selected as the number of clusters that best characterise the resultant clustered points.

3.3.2 Project-Level Regression

Linear regression is applied to identify the impact of peer recommendations on fundraising performance. The dependent variable is the percentage funds raised of set goal achieved. Three different sets of independent variables will be used in three different regression models. The first choice of independent variable is a binary indicator variable denoting whether the project has been recommended by others in the network. The quantity to be estimated is the average increase in the percentage of funds raised over set goal across all projects that have received peer recommendations. The second choice of independent variable is the number of recommendations received by the project, where the interest of study is the average increase in the percentage of funds raised over set goal for each recommendation received by the project. The third choice of the independent variables is the set of network centrality measures. The quantity of interest is the average increase in the percentage of funds raised over set goal per unit increase in each of the aggregate network centrality measures for all other projects that have given recommendations to the project under study.

In order to identify the effect of respective independent variable(s) on the outcome variable, the fundraising goal, the duration, the reward-related variables including the minimum reward amount, the maximum reward amount and the number of reward levels, the length of description in words, the length of title in words, the year and month of project creation, the category, the country of origin and the various network centrality scores of the project are controlled to eliminate their confounding effects on the outcome variable. The regression form is as follows.

$$y_i = x_i^T \beta + \epsilon_i$$

y_i is the percentage of funds raised over funding goal for project i. x_i is the vector containing both the independent variables of interest and the control variables. β is the coefficient vector for the independent and control variables. ϵ_i is a normally distributed random variable centred at zero for y_i denoting the noise in the data. ϵ_i is distributed independently across $y_i's$ with constant variance under Ordinary Least Squares (OLS) assumptions. OLS estimators, $\hat{\beta}$, for the coefficients will be obtained, where the coefficients for the respective independent variables will be analysed and the t-test for statistical significance of the coefficients will also be interpreted to identify the respective effects of interest stated above.

Nevertheless, project-level regression models may suffer from the problem of endogeneity. This is because, in the context of recommendation network, the successful projects may attract more recommendations, making it difficult to identify the true effect of peer recommendations on fundraising success. To address this issue, panel regression is adopted.

3.3.3 Panel Regression

The availability of panel data in terms of the amount of funds contributed to each project on each day allows a more fine-grained time series regression model with fixed effects to be developed. This allows the performance of the project to be controlled, which helps to overcome the problem of endogeneity described in the section above. Four different sets of independent variables will be used in four different regression models. The first three are similar to the ones stated in 3.3.1. The first independent variable is the occurrence of any peer recommendation for project i at time steps, $t-k, t-k+1, \dots, t-1, t$, where k is an arbitrary constant controlling how many lagged variables to be included. The second is the total number of project recommendations received by project i at time steps, $t-k, t-k+1, \dots, t-1, t$. The third is the set of aggregate network centrality measures for all projects recommending project i at time steps, $t-k, t-k+1, \dots, t-1, t$. For these three, the outcome variable is the amount of funds raised at time $t+1$ in US Dollars.

For these models that study the effect of peer recommendations on fundraising performance, it should be noted that the effect of recent peer recommendations on fundraising performance is likely to be much stronger. This is because visitor counts are usually highest when a new project update containing recommendation is posted, directing the attention of the backers to the recommended project. The visitor counts and the effect of peer recommendation are likely to taper off further into the future after the project update is posted. However, to account for the lasting effect of project recommendations after the second day, lagged variables are also included in model fitting with a relatively small k in the range of 3 to 5.

On the other hand, the fourth set of independent variables have a different interest of study, which is to determine the validity of reinforcement model and substitution model in the recommendation behaviour of the project creators. The independent variables are the lagged variables indicating the number of recommendations that project i received at time steps, $t-k, t-k+1, \dots, t-1, t$. The outcome variable is the number of recommendations that project i made on other projects at time step $t+1$. A positive and statistically significant coefficient of any of the lagged variables supports the reinforcement model, whereas a negative and statistically significant coefficient of any of the lagged variables supports the substitution model.

The following variables will be controlled.

$$X_{i,t} = \left(F_{i,t}, B_{i,t}, DD_{i,t}, U_{i,t}, RR_{i,t}, RG_{i,t}, R_{i,t}, DL_{i,t}, CM_{i,t}\right) \ \alpha_i = \left(C_i, TL_i, CO_i, G_i\right)$$

$F_{i,t}$ denotes the amount of funds already raised for project *i* at time *t*. $B_{i,t}$ denotes the number of backers for project *i* at time *t*. $DD_{i,t}$ denotes the remaining number of days to deadline for project i at time t. $U_{i,t}$ denotes the number of updates posted for project *i* at time *t*. $RR_{i,t}$ denotes the number of recommendations that project *i* has already received at time *t*. $RG_{i,t}$ denotes the number of recommendations that have already been given to other projects in project *i*'s updates up to time *t*. $R_{i,t}$ denotes the reward-related project characteristics for project *i* at time *t*, including the minimum reward amount, the maximum reward amount and the number of reward levels. $R_{i,t}$ represents three different variables instead of a single variable. The notation of $R_{i,t}$ is used here for brevity. $DL_{i,t}$ denotes the length of project *i*'s description measured in number of words at time *t*. $CM_{i,t}$ denotes the set of recommendation network centrality measures for project *i*. C_i denotes the category that project *i* belongs to. TL_i denotes the length of project *i*'s title measured in number of words. CO_i denotes the country of origin for project *i*. G_i denotes the funding goal of project *i*.

The regression form is as follows.

$$y_{i,t+1} = \sum_{j=0}^{k} \widetilde{X_{\iota,t-J}}^{T} \widetilde{\beta_J} + X_{i,t}^{T}\beta + \alpha_i^{T}\gamma + \epsilon_i$$

Here, $y_{i,t+1}$ is the outcome variable denoting the amount of funds raised or the number of recommendations given by project i at time step $t+1$. $\widetilde{X_{\iota,t-J}}$ represents the independent variables of interest for project i at time $t-j$. $X_{i,t}$ represents the independent variables of interest for project i at time t. α_i represents the vector of fixed effects variables that are time-invariant and project-specific. ϵ_i is a normally distributed random variable centred at zero that satisfy the OLS assumptions of constant variance and independence.

3.3.4 Propensity Score Matching

Propensity Score Matching (PSM) is a method that explicitly controls for the confounding variables. The interest of PSM is to study the effect of a treatment that the subjects are observed to go through on the outcome variable. In the context of this study, the subjects are the crowdfunding projects being studied with their features recorded at a specific time step. The treatment is the occurrence of recommendation on a given project at a specific time step and the outcome variable is the amount of funds that backers pledged at the next time step. The way PSM controls for the confounding variables is through the introduction of a propensity score, which is the probability of an individual receiving a treatment given a set of chosen observable pre-treatment characteristics.

$$e(X_{i,t}) = \Pr\left(W_{i,t} = 1 | X_{i,t} = x_{i,t}\right)$$

In the equation above, $e(X_{i,t})$ denotes the propensity score assigned to project i at time t represented as a vector of chosen pre-treatment characteristics. $W_{i,t}$ is an indicator variable denoting the occurrence of recommendation on project i at time t. The key property that is satisfied by $e(X_{i,t})$ is

$$X_{i,t} \perp W_{i,t} | e(X_{i,t})$$

That is, given the propensity score assigned to project *i* at time *t*, the pre-treatment characteristics of the project at time t are independent of the occurrence of recommendation on the project at time t (Rosenbaum & Rubin, 1983). This implies that the exposure of the projects with similar propensity scores to recommendation is random. If $0 < e\left(X_{i,t}\right) < 1$, then the unbiased estimates of average effects of recommendation can be obtained through paired matching (Rosenbaum & Rubin, 1983).

In particular, in order to perform PSM, a series of rigorous steps are advised to be followed (Caliendo & Kopeinig, 2008). The first step is to determine the functional form for propensity score estimation. The logistic regression model is adopted.

$$e\left(X_{i,t}\right) = \frac{\exp\left(h(X_{i,t})\right)}{1 + \exp\left(h(X_{i,t})\right)}$$

Here, $h(X_{i,t})$ is a linear function of the project characteristics, the parameters of which will be determined through maximum likelihood estimation. In terms of feature selection, only variables that simultaneously influence project recommendation and the next-day amount of pledged funds should be included in the functional form (Leuven & Sianesi, 2018; Smith & Todd, 2005). In particular, these characteristics should not be affected by the occurrence of project recommendation. Of the available project characteristics, the following are chosen to meet the stated criteria. These variables have been explained in detail in section 3.3.3.

$$\left(F_{i,t}, B_{i,t}, DS_{i,t}, DD_{i,t}, U_{i,t}, RR_{i,t}, RG_{i,t}, C_i, R_i, DL_i, TL_i, CO_i, CM_i, G_i\right)$$

Before proceeding to the next step, it will be checked whether the property $X_{i,t} \perp W_{i,t} | e(X_{i,t})$ holds approximately. The propensity score is first discretised into multiple bins. Within each bin, it will be tested if the means of each variable used in propensity score estimation for projects at various time steps with and without receiving recommendations have a statistically significant difference using two-sample t-test. In the case of failure of the test, higher order terms will be incorporated into $h(X_{i,t})$ to allow the propensity score to be estimated with a greater degree of flexibility in order for the property $X_{i,t} \perp W_{i,t} | e(X_{i,t})$ to be fulfilled.

After performing the necessary check, for each project i that received recommendation at time t_i, a project j that did not receive recommendation at time t_j will be found, where $e(X_{j,t_j})$ is closest in distance to $e(X_{i,t_i})$ among those in the control group without the occurrence of recommendation. Since there is likely to be strong serial correlation in X_{i,t_i}, it will be required that $j \neq i$. For each pair found, the difference between the amounts of funds raised at time $t_i + 1$ and $t_j + 1$ for projects i and j respectively is computed. Afterwards, a paired two-sample t-test can be carried out to determine if peer recommendation has a statistically significant impact on the amount of funds pledged on the next day.

4. Results and Discussions

4.1 Recommendation Motivations

In order to extract recommendation motivations, feature engineering is first carried out. A word vectors model is trained on the project titles and then the feature vectors are extracted. A low-dimensional embedding of the word vectors is visualised in a 2-dimensional space using t-distributed nearest stochastic neighbour embedding. Furthermore, the cluster points are color-coded according to project category, showing that projects from the same category are distributed together in the reduced dimensions.

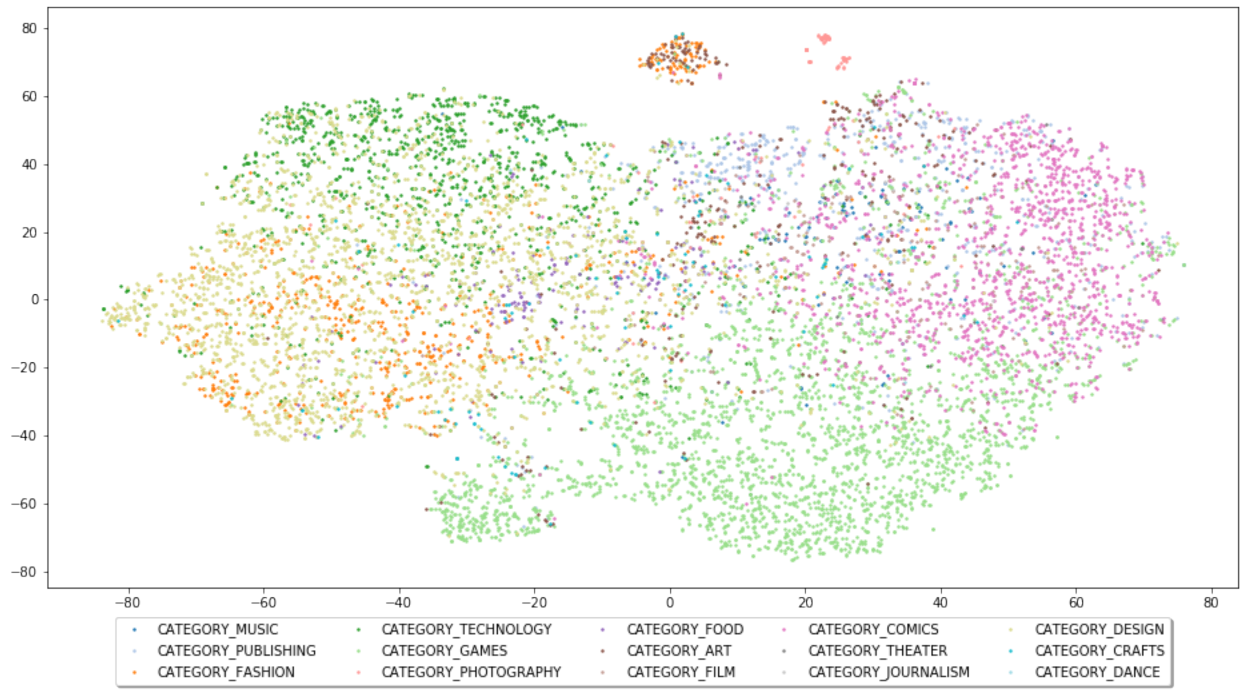

Figure 4 Scatter Plot of Projects in the TSNE Visualisation of Generated Word Vectors

Four groups of features are used in coding recommendation motivations These include reciprocity features, project performance features, location feature and project similarity feature.

Feature Group	Feature Names
Reciprocity	transformed_days_since_rec transformed_days_until_rec
Performance	recommender_pledged_percentage

	recipient_pledged_percentage recommender_backers recipient_backers
Location	is_same_city
Similarity	description_distance

Table 2 Groups of Motivation-related Features

The feature, transformed_days_since_rec, is computed based on the following below:

$$transformed\_days\_since\_rec = \frac{1}{1 + number\ of\ days\ since\ being\ recommended}$$

The reciprocal transformation is taken such that if the recommender has never been recommended by the recipient before, the reciprocal transformation will produce a zero value. Similarly, the feature, transformed_days_until_rec, is computed according to the formula below.

$$transformed\_days\_until\_rec = \frac{1}{1 + number\ of\ days\ until\ being\ recommended}$$

The reciprocal transformation will produce a zero value if the reciprocal recommendation never occurred. On the other hand, an immediate reciprocal recommendation will produce a value of one. Furthermore, the number of days may not take on integer values.

The recommender_pledged_percentage and recommender_backers measure the performance levels of the recommender project based on the percentage amount pledged of goal and the number of backers at the time of recommendation respectively. On the other hand, the recipient_pledged_percentage and recipient_backers measure the performance levels of the recommendation recipient project based on the percentage amount pledged of goal and the number of backers at the time of recommendation as well.

is_same_city is a binary variable that equals 1 if the recommender and the recipient are from the same city, and 0 otherwise. A city-level indicator is chosen over country-level indicators due to its greater specificity and its greater suitability in capturing the recommender-recipient geographical proximity.

Furthermore, description_distance – the cosine distance between the project descriptions in the low-dimensional embedding of the extracted word vectors – captures the dissimilarity between the recommender and the recipient projects.

After the features are computed, standardisation and logarithmic transformation are applied on the non-binary features due to their highly skewed distributions. A pairwise plot is produced below for the transformed features to examine the suitability of the transformed features for clustering.

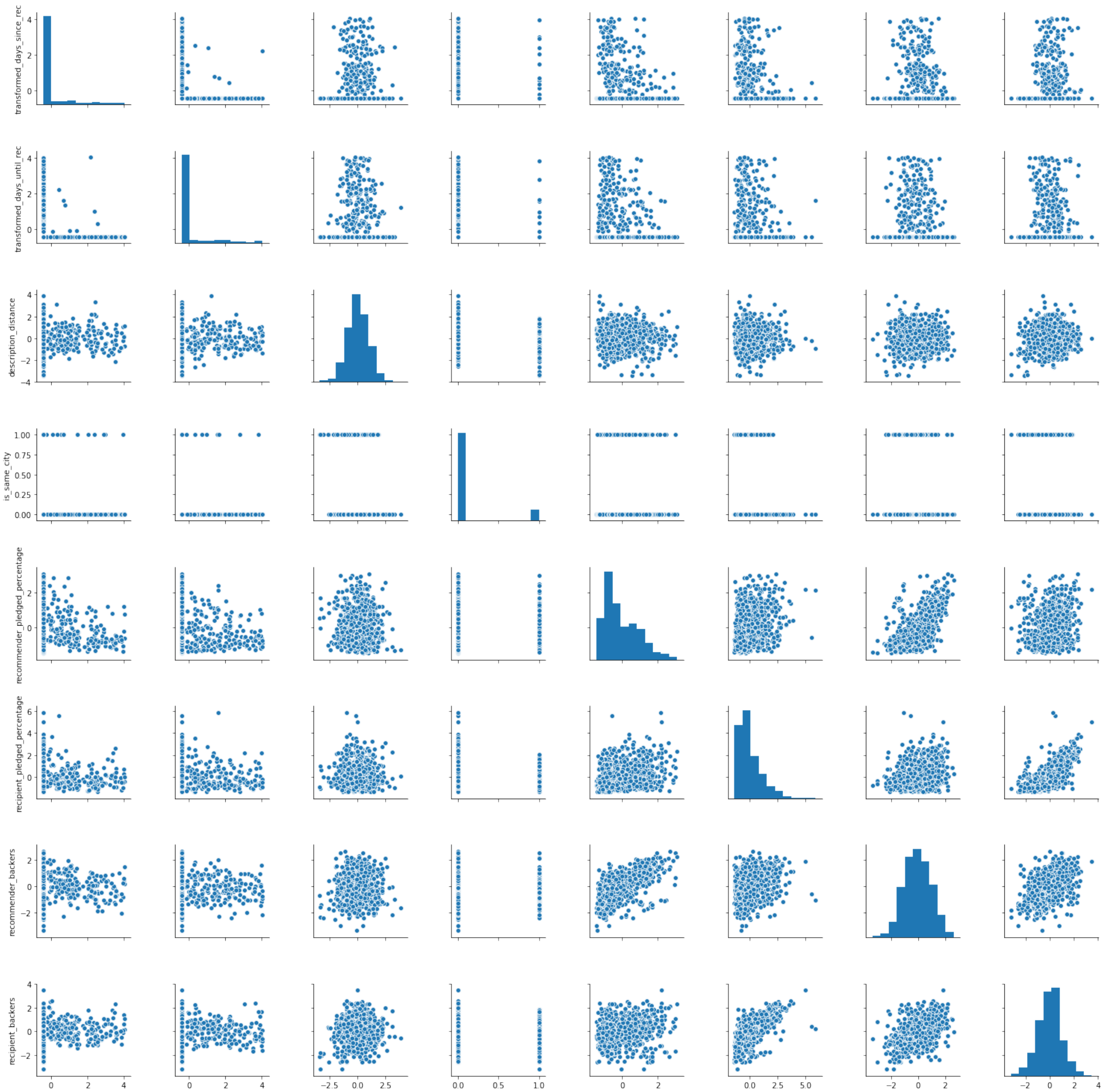

Figure 5 Pairwise Plot of Transformed Features

K-means models based on a range of K values from 2 to 9 are run on the entire set of recommendation data. To determine the most suitable K, silhouette scores are computed and displayed in the line chart below.

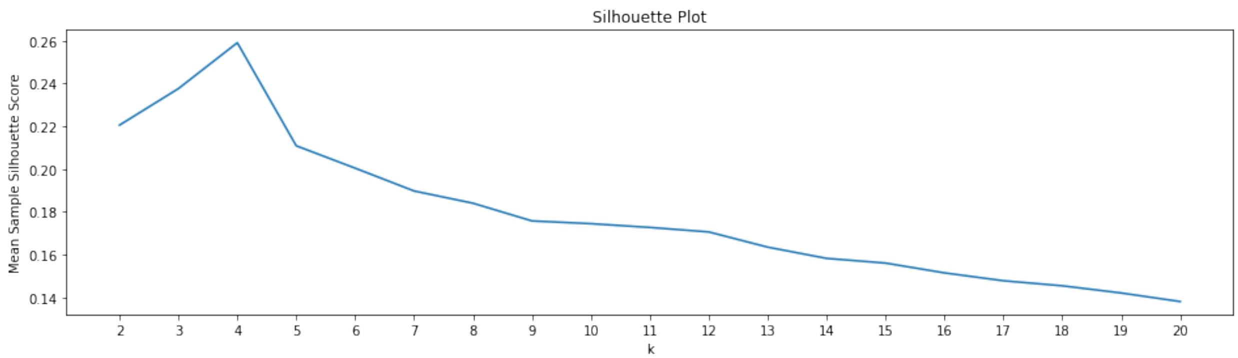

Figure 6 Average Silhouette Scores by K Value

As can be seen, the K value of 4 produces the most optimal cluster formations. The labels for the four clusters are extracted. To interpret the characteristics of the clusters with respect to the recommendation motivations, the centroids are computed and shown in Table 3. Furthermore, the sizes of the clusters are also displayed in Table 4 below.

Feature Columns	Centroid Values for Each Cluster				Overall Mean	Overall Standard Deviation
	0	1	2	3		
Transformed Days Since Being Recommended	0.03	0.02	0.60	0.00	0.08	0.20
Transformed Days Until Being Recommended	0.03	0.02	0.01	0.60	0.08	0.20
Description Cosine Distance	0.32	0.30	0.31	0.31	0.31	0.09
Same City Indicator Variable	0.07	0.18	0.05	0.05	0.12	0.32
Recommender Pledged Percentage	17.10	3.03	2.73	2.76	7.51	51.90
Recipient Pledged Percentage	12.05	1.59	2.94	2.93	5.25	18.69
Recommender Backers	2248.11	338.42	518.17	524.24	993.52	2078.58
Recipient Backers	1873.79	238.38	556.39	534.11	831.84	1989.02

Table 3 Cluster Centroid Feature Values with Overall Feature Means and Standard Deviations

Cluster	0	1	2	3
Size	8237	11798	2752	2764
Coding	Status	Proximity	Reciprocation	Anticipated Reciprocation

Table 4 Cluster Size and Coded Motivations

Cluster 0 is characterised by high mean recommender backers (2248.11) and pledged percentage (17.10) as well as high mean recipient backers (1873.79) and pledged percentage (11.6648). Therefore, the recommendations in Cluster 0 mainly occur across pairs of projects that are both highly successful. This suggests that the recommenders could have been drawn to the status of the recipient projects.

Cluster 1 has the highest same city indicator variable mean value (0.18). The other feature means are lower than their respective overall means. As such, Cluster 1 is characterised by recommendations that occur across projects of high proximity. Conceptually, the proximity factor could indicate the existence of a past relationship between the recommender and the recipient.

Cluster 2 and 3 are smaller in size as compared to Cluster 0 and 1. Cluster 2 and 3 have significantly higher mean transformed days since being recommended (0.60) and mean transformed days until being recommended (0.60) respectively while other mean feature values are very close or below their respective overall means. This indicates that Cluster 2 contains recommendations that could be motivated by the perceived obligation to reciprocate project recommendation and Cluster 3 contains recommendations that could be motivated by the anticipation of recipient reciprocating project recommendation.

The generated cluster labels are aggregated over the days on which each project is live. These labels allow the identified recommendation motivations to be coded and used in the regression models in sections 4.2 and 4.3. In addition, two other motivations have been identified which are not captured in the clustering analysis. They are coded based on matching indicative regular expression patterns respectively. These two motivations are collaborative promotion and community spirit respectively. Table 7 shows the expressions adopted to identify the two motivations.

Motivation	Expression	Number of Recommendations
Collaborative Promotion	Brand cooperation Back both Backers of both	66
Community Spirit	Community spirit	22

Table 7 Additional Motivation Matching

The recommendations containing expressions and their variants are then coded accordingly. These serve as additional independent variables in the motivation category for the regression models used in the subsequent sections.

4.2 Project-level Regression

In order to determine the effect of recommendations and recommender centrality on the project fundraising performance, a project-level regression model is adopted. Due to the existence of high multicollinearity among the centrality scores, LASSO regression is applied to the groups of in- and out- centrality measures separately to predict the logarithmically transformed pledged amount. The regularisation strength is adjusted so that only one variable from each group of centrality measures remains. The harmonic centrality measure is selected for the in-centrality group while the pagerank centrality is selected for the out-centrality group. The VIF values of all variables after LASSO feature selection fall below 5, suggesting that there is no multicollinearity issues. In the project-level regression model, the independent variables of interest are the number of recommendations given and received as well as the average in- and out-centrality of the recommenders. The regression results are displayed in Table 8.

In terms of notation, the coefficients for the variables are listed in the two rightmost columns and the robust standard errors are shown in parentheses. The coefficients with p-values less than 0.01 are marked with triple asterisks. The coefficients with p-values between 0.01 and 0.05 are marked with double asterisks. The coefficients with p-values between 0.05 and 0.10 are marked with a single asterisk.

The regression model using the logarithmically transformed pledged amount as the dependent variable has a 26.3% R-squared value while the model using the logarithmically transformed number of backers as the dependent variable has a 30.7% R-squared value. The control variables of project category and country of origin are omitted for brevity.

Variable Groups	**Variable Names**	**Pledged Amount**	**Number of Backers**
Recommendation Number	rec_received_cum	0.163*** (0.0116)	0.146*** (0.00654)
	rec_given_cum	0.109*** (0.0127)	0.0972*** (0.00717)
Recommender Centrality	in_harmonic	0.0787*** (0.00355)	0.0561*** (0.00200)
	out_pagerank	0.0724*** (0.00331)	0.0621*** (0.00187)
Control Variable sand Constant	desc_words	-0.0177*** (0.00132)	-0.0123*** (0.000747)
	duration_days	-0.0143*** (0.000651)	-0.0102*** (0.000367)
	goal_usd	-4.48e-08*** (5.84e-09)	-2.10e-08*** (3.29e-09)
	launched_month	0.0826*** (0.00248)	0.0414*** (0.00140)
	launched_year	1.042*** (0.00928)	0.577*** (0.00523)
	mean_reward	-7.74e-05*** (9.17e-06)	-6.23e-05*** (5.17e-06)
	name_words	0.202*** (0.00278)	0.111*** (0.00157)
	project_reward_number	0.232*** (0.00155)	0.132*** (0.000872)
	Constant	-2,101*** (18.83)	-1,164*** (10.62)
Summary Statistics	Observations	160,045	160,045
	R-squared	0.263	0.307

Table 8 Project-level Regression Results

The number of recommendations received has a positive and statistically significant relationship with the fundraising performance at 95% confidence level. In particular, one additional received recommendation leads to a 16.3% increase in the pledged amount (p-value=0.000) and a 14.6% increase in the number of backers (p-value=0.000).

Moreover, at 95% confidence level, the number of recommendations given also has a positive and statistically significant effect on the fundraising performance. Each additional recommendation given results in a 10.9% increase in the pledged amount (p-value=0.000) and a 9.72% increase in the number of backers gained (p-value=0.000). This shows that while giving recommendations has a weaker impact on the fundraising performance than receiving recommendations, the effect is statistically significant and it is within the power of the project creators to make recommendations to other projects in order to improve their own fundraising outcome.

Furthermore, it is observed that each unit increase in the in-harmonic centrality of the recommenders is associated with a 7.87% increase in the pledged amount (p-value=0.000) and a 5.61% increase in the number of backers gained (p-value=0.000). On the other hand, each unit increase in the average out-pagerank centrality of the recommenders is associated with a 7.24% increase in the pledged amount and a 6.21% increase in the number of backers gained (p-value=0.000). Therefore, both in- and out-centrality measures of the recommenders are observed to have a positive and statistically significant effect on fundraising performance at 95% confidence level.

However, the project-level regression suffers from endogeneity and the lack of determination of more fine-grained results. The issue of endogeneity arises from the fact that the dependent variable and the independent variable may simultaneously influence each other. For example, the successful projects may influence recommenders with high out-centrality scores, leading to misinterpretation of the causal effect of out-centrality on project fundraising performance. To deal with the challenge, panel regression models are adopted and explained in the following sections.

4.3 Panel Regression

4.3.1 Day-level Project Performance

To determine the day-level impact of independent variables on crowdfunding performance, a panel regression model with fixed effects is adopted. In all the regression models, the VIF values are checked for the independent variables and the control variables. All of the VIF scores are below 10 and most of the VIF scores are below 5, indicating that multicollinearity is an issue (O'brien, 2007).

In the fixed effect model, the time variable is the number of days since funding campaign launch and the group variable is project ID. Table 6 shows the results for two separate regression models using the logarithmically transformed next-day pledged amount and the logarithmically transformed next-day backer number as dependent variables respectively.

The model using the next-day pledged amount as the dependent variable achieves 5.3% R-squared value while the model using the next-day backers gained as the dependent variable achieves 9.1% R-squared value. Furthermore, the following control variables are removed from the regression results due to the presence of multicollinearity. They are the mean reward amount, the total number of reward categories and the project goal in USD. In addition, the following control variables are removed from the regression results for brevity. They are the day of funding period, the country code, the campaign launch year and the campaign launch month.

Variable Groups	Variable Names	Next-day Pledged Amount	Next-day Backer Number
Recommendation Number	rec_given	0.0130	-0.0110
		(0.0304)	(0.0132)
	rec_received	0.312***	0.216***
		(0.0289)	(0.0138)
Recommender Centrality	recommender_in_harmonic_centrality	0.00584***	0.00213***
		(0.00139)	(0.000704)
	recommender_out_harmonic_centrality	-0.00733***	-0.00423***
		(0.00154)	(0.000754)
Recommendation Motivation Clusters	in_cluster_0	0.0303	0.0337**
		(0.0334)	(0.0168)
	in_cluster_1	0.120***	0.0342**
		(0.0337)	(0.0159)
	in_cluster_2	-0.0280	-0.0351**
		(0.0406)	(0.0178)
	out_cluster_0	0.0385	0.0308**
		(0.0338)	(0.0156)
	out_cluster_1	0.0879**	0.0463***
		(0.0365)	(0.0154)
	out_cluster_2	0.0594	0.0593***
		(0.0443)	(0.0191)
Additional Recommendation Motivations	recommender_cs_bin	-0.0357	0.0697
		(0.211)	(0.135)
	recipient_cs_bin	0.114	0.138
		(0.270)	(0.186)
	collaborative_promotion	0.582***	0.214**
		(0.149)	(0.0868)
Control Variable sand Constant	rec_given_cum	-0.00891	-0.00806**
		(0.00696)	(0.00394)
	rec_received_cum	-0.0216***	-0.0191***
		(0.00638)	(0.00361)
	project_name_words	0.00158	0.00298**
		(0.00364)	(0.00145)
	log_pledged_usd	-0.121***	
		(0.00191)	
	log_backers		-0.111***
			(0.00237)
	Constant	3.593***	1.266***
		(0.0241)	(0.00975)
Summary Statistics	Observations	4,984,429	4,984,429
	R-squared	0.053	0.091
	Number of entry_project_id	160,036	160,036

Table 9 Fixed Effect Models with the Logarithmically Transformed Next-day Pledged Amount and the Next-day Backer Number as the Dependent Variables

The number of recommendations received (rec_received) has a positive and statistically significant effect on the next-day funding performance at 95% confidence level. In the two regression models, each recommendation received is associated with a 31.2% and a 21.6% increase in the next-day amount of funds raised (p-value=0.000) and the next-day number of backers gained (p-value=0.000) respectively. On the other hand, there is no statistically significant relationship between the number of recommendations given and the next-day funds raised (p-value=0.668) or the next-day number of backers gained (p-value=0.341).

The results show a positive and statistically significant relationship between the average in-harmonic centrality of the recommenders and the next-day pledged amount (p-value=0.000) as well as the next-day backers gained (p-value=0.000) at 95% confidence level, where each unit increase in the average in-harmonic centrality of recommenders results in a minor increase of 0.584% in the next-day pledged amount and 0.213% in the next-day backers gained. This shows that the recommenders' centrality in terms of whether the recommenders themselves have received recommendations from multiple sources has a minor albeit statistically significant effect on the funding performance.

On the other hand, the out-harmonic centrality of the recommenders have a negative and statistically significant effect on the fundraising performance. In particular, each unit increase in the average out-harmonic centrality results in a 0.733% decrease in the next-day pledged amount (p-value=0.000) and a 0.423% decrease in the next-day backers gained (p-value=0.000) at 95% confidence level. This suggests that all else equal, the more central the recommenders are in terms of how many recommendations they have given out, the more diluted the impact of their recommendations is. Specifically, given the same number of recommendations, it is expected that less traffic would follow from the recommenders with high out-centrality values and pledge to the recommended project.

From the perspective of the coded recommendation motivations, Cluster 0 shows a positive and statistically significant difference from Cluster 3 in determining the next-day number of backers for both incoming (p-value=0.042) and outgoing recommendations (p-value=0.049). In particular, all else equal, an increase in the number of incoming recommendations based on status results in a 3.37% increase of next-day backers gained and an increase in the number of outgoing recommendations based on status results in a 3.08% increase of next-day backers gained. However, an increase in the number of incoming recommendations and outgoing recommendations based on status has no statistically significant effect on the next-day pledged amount with p-values of 0.364 and 0.255 respectively.

In addition, Cluster 1 shows a positive and statistically significant difference from Cluster 3 in determining both the next-day number of backers and the next-day pledged amount at 95% confidence level. In particular, all else equal, each additional incoming recommendation coded as proximity leads to a 12.0% increase in the next-day pledged amount (p-value=0.000) and a 3.42% in the next-day backers gained (p-value=0.031) while each additional outgoing recommendation coded as proximity results in a 8.79% increase in the amount of next-day pledged funds (p-value=0.016) and a 4.63% increase in the next-day backers gained (p-value=0.003). This shows that proximity-based recommendations tend to have a stronger positive effect on the funding performance as compared to the anticipation of reciprocation.

On the other hand, all else equal, there exists a statistically significant relationship at 95% confidence level between Cluster 2 and Cluster 3 in determining the number of backers gained on the next day. In particular, given that the total number of recommendations given remains unchanged, each additional recommendation given based on reciprocation in place of a recommendation given based on anticipated reciprocation increases the next-day backers gained by 5.93% (p-value=0.002). In addition, each additional recommendation received based on reciprocation in place of a recommendation received based on anticipated reciprocation decreases the next-day backers gained by 3.51% (p-value=0.049). This suggests that recommendations given for the purpose of reciprocation may improve the image of the project in the eyes of the potential backers. This could act as a quality signal of the project. On the other hand, all else being equal, recommendations received for the purpose of reciprocation tend to degrade project performance as compared to recommendations received for the purpose of anticipating reciprocation. This suggests that when the recommendation received is based on reciprocation, the backers of the recommender may experience a weaker quality signal of the recommended project, resulting in poorer performance with the other conditions being held equal.

Furthermore, the identified community spirit recommendation motivation, recommender_cs_bin and recipient_cs_bin, did not show a statistically significant effect on the next-day pledged amount (p-value=0.866 for recommender_cs_bin and p-value=0.674 for recipient_cs_bin) or the next-day backers gained (p-value=0.605 for recommender_cs_bin and p-value=0.459 for recipient_cs_bin). Nevertheless, the collaborative promotion motivation has a strongly positive and statistically significant effect on the project's fundraising performance, leading to a 58.2% increase in the next-day pledged amount (p-value=0.000) and a 21.4% increase in the next-day backers gained (p-value=0.013). This indicates the strong effectiveness of collaborative promotion as a motivation as compared to ordinary recommendations.

Moreover, the cumulative numbers of recommendations received and given both show a statistically significant and negative correlation with the fundraising performance at 95% confidence level. In particular, all else being equal, each unit increase in the cumulative number of recommendations given results in a 0.806% decrease in the next-day backers gained (p-value=0.041). Each unit increase in the cumulative number of recommendations received results in a 2.16% decrease in the next-day pledged amount (p-value=0.001) and a 1.91% decrease in the next-day backers gained (p-value=0.000). This shows that the accumulation of received or given recommendations reduces the efficacy of additional recommendations being received or given.

To add on, the negative and statistically significant correlation is also observed between the cumulative pledged amount (log_pledged_usd) or cumulative backers number (log_backers) and the fundraising performance at 95% confidence level. In particular, each percentage increase in the cumulative pledged amount is associated with 12.1% decrease in the next-day amount pledged (p-value=0.000) and each percentage increase in the cumulative number of backers is associated with 11.1% decrease in the next-day number of backers gained (p-value=0.000).

4.3.2 Determination of the Reinforcement Effect

In order to determine whether the reinforcement effect model or the substitution effect model prevails in the context of crowdfunding recommendation, the next-day number of recommendations given is used as the dependent variable (Table 7).

Variable Groups	Variable Names	Number of Next-day Recommendations Given
Cumulative Number of Recommendations	rec_given_cum	-0.0462***
		(0.00312)
	rec_received_cum	0.0295***
		(0.00258)
Recommender Centrality	recommender_in_harmonic_centrali	-0.000711
		(0.000728)
	recommender_out_harmonic_central	0.00112*
		(0.000677)
Recommendation Motivation Clusters	in_cluster_0	0.00500
		(0.0130)
	in_cluster_1	-0.0367***
		(0.0110)
	in_cluster_2	-0.0139***
		(0.00468)
	in_cluster_3	0.717***
		(0.0278)
	out_cluster_0	-0.0662***
		(0.0123)
	out_cluster_1	-0.293***
		(0.0143)
	out_cluster_2	-0.0343***
		(0.00467)
	out_cluster_3	-0.0306***
		(0.00877)
Additional Recommendation Motivations	recommender_cs_bin	0.0679
		(0.138)
	recipient_cs_bin	-0.0449
		(0.0393)
	collaborative_promotion	0.0308
		(0.0702)
Control Variable sand Constant	log_pledged_usd	-0.00123***
		(8.65e-05)
	log_backers	0.00443***
		(0.000364)
	project_name_words	0.000452**
		(0.000185)
	Constant	-0.00221*
		(0.00118)
Summary Statistics	Observations	4,824,393
	Number of entry_project_id	159,744
	R-squared	0.059

Table 7 Fixed Effect Model with Number of Next-day Recommendations Given as the Dependent Variable

In the second fixed effect model built, the next-day recommendations given is used as the dependent variable. The main variable of interest is the cumulative number of recommendations received.

From Table 7, the cumulative number of recommendations received has a positive and statistically significant effect on the number of recommendations given on the next day at 95% confidence level. In particular, each additional received recommendation results in an increase of 0.0295 in the next-day number of recommendations given for each project (p-value=0.000). The results agree with the reinforcement model, in that an increase in the number of recommendations made by others stimulates the recommender to make more recommendations to the other projects, thereby contributing to the development of the recommendation network by creating new edges and increasing inter-connectedness.

Furthermore, it can also be seen from Table 7 that the cumulative number of recommendations given by a project has a negative and statistically significant relationship with the number of next-day recommendations given at 95% confidence level. In particular, each additional recommendation given results in on average a 0.0462 decrease in the number of next-day recommendations given (p-value=0.000). This indicates the existence of the phenomenon of decaying momentum as the number of recommendations given gradually decrease as the already given recommendations increase in number.

4.4 Propensity Score Matching

Propensity score matching is also conducted to corroborate the previous results. The logit function is adopted where the pre-treatment characteristics include the cumulative number of recommendations given and received, the logarithmically transformed goal in USD, the funding campaign start year and month, mean rewards, number of reward levels as well as the number of words in the name of the project. The pre-treatment characteristics are fitted into the logit model to generate a propensity score that represents the treatment probability, where the treatment is the occurrence of recommendation.

Then, propensity score matching is carried out to match each treated subject with a control subject with similar propensity scores. A t-test is then conducted to determine if a statistically significant difference exists between the treatment group and the control group. The variables of interest, the t-statistic are and the p-values under the t-test are recorded in the table below.

Variable	Mean Treated	Mean Control	t-statistic	p-value
log_day_pledged_usd	6.2472	6.1807	2.46	0.014
log_day_backers	2.5606	2.5152	2.85	0.005
rec_given_cum	2.6137	2.0088	13.25	0.000
rec_received_cum	4.0373	2.8933	20.57	0.000
long_entry_goal_usd	9.1881	9.3733	-13.24	0.000
log_pledged_usd	9.7826	10.359	-30.28	0.000

Table xx Propensity Score Matching Results

Propensity score matching also shows a statistically significant difference in the amount pledged next day as well as the number of backers gained next day at 95% confidence level based on the T-test, confirming the results obtained in the panel regression model.

5. Summary and Further Discussions

Earlier literature on crowdfunding has focused on the motivations and success factors of the crowdfunding projects. With the rising popularity of recommendation engines, recommender systems have also become a focal point of studies. However, literature is lacking in assessing the impact of peer recommendations in the context of crowdfunding. In particular, the effect of peer recommendation could be potentially stronger than machine-generated predictions due to the word-of-mouth effect. Furthermore, the two classes of models, namely the reinforcement model and the substitution model, may also be applied to understand the peer recommendation behaviour through the lens of an established theory.

5.1 Summary of Findings

The study begins by exploring the various antecedents of peer recommendations in the crowdfunding context. The four dominant motivations have been identified as status, proximity, reciprocation and anticipated reciprocation. Furthermore, collaborative promotion and community spirit have also been separately coded as antecedents of project recommendations.

The regression models show that an increase in the number of recommendations received have a positive effect on the fundraising performance. High in-centrality measures of the recommenders could improve the fundraising performance while high out-centrality measures of the recommenders may dilute the backer traffic and lead to worse fundraising performance with all else being held equal. The various motivations have different effects on fundraising performance but the proximity and the collaborative promotion motivations have been identified to cause the greatest improvement in fundraising performance.

In addition, it is found that the increase in the number of recommendations received stimulates the project creator to give more recommendations to other projects in the community. This supports the validity of the reinforcement model in the context of crowdfunding peer recommendations. In addition, there is a decaying momentum in the number of recommendations being made as the project creators tend to decrease the number of recommendations they contribute as they make more contributions to recommend other projects.

5.2 Implications of Study

The study has determined the antecedents and of peer recommendations in the crowdfunding context, an area that has not been investigated by current research. The study shows that the reinforcement model may not be only interpreted in the context of public goods. For the project creators, contributing to the common good of the crowdfunding platform by making recommendations to other projects and increasing overall project visibility across the platform fulfils a similar function as contributing to a public good. As such, the reinforcement model shows that the existence of project recommendations tends to promote further recommendations that catalyse a chain reaction to knit the crowdfunding community together.

Furthermore, the study has also investigated the effect of peer recommendations as well as the related motivations on the crowdfunding performance. The results can be applied in practice to guide project recommendation. For example, recommending too many projects may be unfavourable for the recommended project. Therefore, the crowdfunding platform may monitor the out-centrality values of the projects and send a notification to projects with high out-centrality values to let them understand the impact of making too many recommendations. In particular, it is found that receiving project recommendations improve the fundraising outcome while giving recommendations does not improve fundraising performance at the day level. All else being equal, recommendations coming from creators with high in-centrality values have a more positive effect on the fundraising performance as compared to recommendations coming from creators with high out-centrality values due to the dilution effect.

5.3 Limitations

The current methodology adopts clustering as the coding mechanism for recommendation motivations. However, the insufficient data and noise in the data prevent the clustering methodology to fully capture the existing motivations. This results in a partial view of the recommendation motivations without getting the full picture. Future studies can focus on alternative data sources in order to devise new methods for motivation extraction.

6. References

Akerlof, G. A. (1982). Labor contracts as partial gift exchange. The quarterly journal of economics, 97(4), 543-569.

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and Ricardian equivalence. Journal of political Economy, 97(6), 1447-1458.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. The economic journal, 100(401), 464-477.

Andreoni, J. (1998). Toward a theory of charitable fund-raising. Journal of Political Economy, 106(6), 1186-1213.

Belleflamme, P., Lambert, T., & Schwienbacher, A. (2010, June). Crowdfunding: An industrial organization perspective. In Prepared for the workshop Digital Business Models: Understanding Strategies', held in Paris on June (pp. 25-26).

Benchettara, N., Kanawati, R., & Rouveirol, C. (2010, August). Supervised machine learning applied to link prediction in bipartite social networks. In Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on (pp. 326-330). IEEE.

Bernheim, B. D. (1994). A theory of conformity. Journal of political Economy, 102(5), 841-877.

Best, J. (2013). Crowdfunding's Potential for the Developing World, infoDev, Finance and Private Sector Development Department. Washington, DC: World Bank.

Bi, S., Liu, Z., & Usman, K. (2017). The influence of online information on investing decisions of reward-based crowdfunding. Journal of Business Research, 71, 10-18.

Bonacich, P. (1987). Power and centrality: A family of measures. American journal of sociology, 92(5), 1170-1182.

Burtch, G., Ghose, A., & Wattal, S. (2013). An empirical examination of the antecedents and consequences of contribution patterns in crowd-funded markets. Information Systems Research, 24(3), 499-519.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. Journal of economic surveys, 22(1), 31-72.

Cornelli, F. (1996). Optimal selling procedures with fixed costs. *Journal of Economic Theory*, *71*(1), 1-30.

Cordova, A., Dolci, J., & Gianfrate, G. (2015). The determinants of crowdfunding success: evidence from technology projects. Procedia-Social and Behavioral Sciences, 181, 115-124.

Davis, A., & Khazanchi, D. (2008). An empirical study of online word of mouth as a predictor for multi-product category e-commerce sales. Electronic markets, 18(2), 130-141.

Duan, W., Gu, B., & Whinston, A. B. (2008). The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry. Journal of retailing, 84(2), 233-242.

Gerber, E. M., Hui, J. S., & Kuo, P. Y. (2012, February). Crowdfunding: Why people are motivated to post and fund projects on crowdfunding platforms. In Proceedings of the International Workshop on Design, Influence, and Social Technologies: Techniques, Impacts and Ethics (Vol. 2, No. 11). New York, NY: ACM.

Gerber, E. M., & Hui, J. (2013). Crowdfunding: Motivations and deterrents for participation. ACM Transactions on Computer-Human Interaction (TOCHI), 20(6), 34.

Giudici, G., Nava, R., Rossi Lamastra, C., & Verecondo, C. (2012). Crowdfunding: The new frontier for financing entrepreneurship?.

Hong, Y., Hu, Y., & Burtch, G. (2015). How does social media affect contribution to public versus private goods in crowdfunding campaigns?.

Hong, Y., Hu, Y., & Burtch, G. (2018). Embeddedness, Pro-Sociality, and Social Influence: Evidence from Online Crowdfunding.

Jones, S. R. (1984). The economics of conformism. Blackwell.

Kromidha, E., & Robson, P. (2016). Social identity and signalling success factors in online crowdfunding. Entrepreneurship & Regional Development, 28(9-10), 605-629.

Koch, J. A., & Siering, M. (2015). Crowdfunding success factors: the characteristics of successfully funded projects on crowdfunding platforms.

Leuven, E., & Sianesi, B. (2018). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing.

Liu, J., Yang, L., Wang, Z., & Hahn, J. (2015). Winner takes all? The “blockbuster effect” in crowdfunding platforms.

Lin, Z., Goh, K. Y., & Heng, C. S. (2015). The demand effects of product recommendation networks: an empirical analysis of network diversity and stability.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. Journal of business venturing, 29(1), 1-16.

Oestreicher-Singer, G., & Sundararajan, A. (2012). Recommendation networks and the long tail of electronic commerce. Mis quarterly, 65-83.

O’brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. Quality & quantity, 41(5), 673-690.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. Stanford InfoLab.

Pathak, B., Garfinkel, R., Gopal, R. D., Venkatesan, R., & Yin, F. (2010). Empirical analysis of the impact of recommender systems on sales. Journal of Management Information Systems, 27(2), 159-188.

Rea, A., & Rea, W. (2016). How Many Components should be Retained from a Multivariate Time Series PCA?. arXiv preprint arXiv:1610.03588.

Roberts, R. D. (1984). A positive model of private charity and public transfers. Journal of Political Economy, 92(1), 136-148.

Rochat, Y. (2009). Closeness centrality extended to unconnected graphs: The harmonic centrality index (No. CONF).

Romano, R. E. (1991). When excessive consumption is rational. The American Economic Review, 553-564.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001, April). Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international conference on World Wide Web (pp. 285-295). ACM.

Schwaighofer, A., Tresp, V., & Yu, K. (2005). Learning Gaussian process kernels via hierarchical Bayes. In Advances in neural information processing systems (pp. 1209-1216).

Shang, J., & Croson, R. (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. The Economic Journal, 119(540), 1422-1439.

Shani, G., Heckerman, D., & Brafman, R. I. (2005). An MDP-based recommender system. Journal of Machine Learning Research, 6(Sep), 1265-1295.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators?. Journal of econometrics, 125(1-2), 305-353.

Sugden, R. (1984). Reciprocity: the supply of public goods through voluntary contributions. The Economic Journal, 94(376), 772-787.

Thies, F., Wessel, M., & Benlian, A. (2014). Understanding the dynamic interplay of social buzz and contribution behavior within and between online platforms–evidence from crowdfunding.

Thies, F., Wessel, M., Rudolph, J., & Benlian, A. (2016, June). Personality Matters: How Signaling Personality Traits can Influence the Adoption and diffusion of Crowdfunding Campaigns. In ECIS (p. ResearchPaper36).

Thürridl, C., & Kamleitner, B. (2016). What goes around comes around? Rewards as strategic assets in crowdfunding. California management review, 58(2), 88-110.

Vesterlund, L. (2003). The informational value of sequential fundraising. Journal of public Economics, 87(3-4), 627-657.

Wang, Z., Yang, L., Kang, Y., & Hahn, J. (2016). Strategies of Effective Reward Scheme Design on Crowdfunding Platforms. In Academy of Management Proceedings (Vol. 2016, No. 1, p. 15229). Briarcliff Manor, NY 10510: Academy of Management.

Warr, P. G. (1982). Pareto optimal redistribution and private charity. Journal of Public Economics, 19(1), 131-138.

Xu, A., Yang, X., Rao, H., Fu, W. T., Huang, S. W., & Bailey, B. P. (2014, April). Show me the money!: An analysis of project updates during crowdfunding campaigns. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 591-600). ACM.