SUCCESS PREDICTION OF CROWDFUNDING CAMPAIGNS WITH PROJECT NETWORK: A MACHINE LEARNING APPROACH

Chao Zhong
School of Information
Renmin University of China
Beijing, China, 100872
zhognchao@ruc.edu.cn

Wei Xu
School of Information
Renmin University of China
Beijing, China, 100872
weixu@ruc.edu.cn

Wei Du*
*Corresponding author
School of Information
Renmin University of China
Beijing, China, 100872
ahduwei@ruc.edu.cn

ABSTRACT

In the last decade, crowdfunding has emerged as a new form of Internet finance, providing founders with a channel through which they can raise funds from the public. Prior studies have mainly investigated two types of crowdfunding success predictors: conventional numerical features (e.g., project goal, duration, number of rewards, number of comments and the presence of a video) and features extracted from textual description and project images. In comparison, few studies have examined the effect of interrelations among projects on crowdfunding performance. For example, a founder can learn from historically invested projects when launching one's own project. In this study, we extend the previous understanding by introducing the concept of "project network," which can be constructed by extracting founders' activities on crowdfunding platforms. Network-based features are extracted from the project network through the Node2vec method. Experimental results show that models with network-based features outperform those without network-based features. Furthermore, the dense dataset with densely connected projects achieves better prediction performance than the original one, further validating the role of a project network in success prediction. Another implication is that a small proportion of connected projects could help predict project success to avoid a high calculation cost.

Keywords: Crowdfunding; Success prediction; Project network; Machine learning; Node embedding

1. Introduction

In recent years, crowdfunding has emerged as a new form of Internet finance on a global scale. Crowdfunding platforms provide founders with a channel to raise funds from public by displaying their own projects in platforms (Yang et al., 2016). Specifically, crowdfunding can be divided into crowdlending as well as donation-, reward-, and equity-based crowdfunding according to the forms of rewards (Leimeister, 2012). This paper focuses on reward-based crowdfunding which requires fundraisers to reward investors with products or services (Zheng et al., 2017). Reward-based crowdfunding is becoming the dominant type of crowdfunding considering the funds raised and the number of projects completed (Kraus et al., 2016). In many countries, crowdfunding has already become an important way for entrepreneurs or small enterprises to raise funds. Kickstarter, the leading crowdfunding platform in the United States, has released over 424,000 projects until November 2018 (Ryoba et al., 2020). In comparison, although crowdfunding started later in China, it has also received full public attention. Demohour[1] is the first crowdfunding platform

[1] http://www.demohour.com/

established in 2011. Subsequently, crowdfunding has developed rapidly in China, and JD Finance[2] has become one of the largest reward-based crowdfunding platforms. As of 2020, over 21,000 projects launched in JD Finance have raised a total amount of US$1.03 billion.

A crowdfunding campaign is considered successful if it reaches the funding goal within a stipulated time period. In reality, the proportion of successful projects is relatively low on most platforms. For example, successful projects account for only 36.52% of all projects on Kickstarter (Ryoba et al., 2020). Founders, backers, and platforms can benefit from the prediction of a project's success probability. For founders, they can plan the follow-up work early if they know the probability of project success is relatively high (Etter et al., 2013). They can also apply corresponding methods to identify the influential features that can help increase the probability of project success (Yuan et al., 2016). For backers, project success prediction can help them invest in projects that are more likely to succeed. Moreover, backers can help raise funds for interesting projects by increasing their pledge or introducing those projects to their friends to increase the success probability of such projects (Wash, 2013). For platforms, predicting project success in advance can help them improve project success rate by implementing specific functions and providing strategies in the project guide (Yang et al., 2020).

Many studies have investigated the crowdfunding success predictors of crowdfunding campaigns. Early studies have explored the impact of numerical features (e.g., raising goal, campaign duration, etc.) on project success (Mollick, 2014; Li et al., 2016; Ryoba et al., 2020). Mollick (2014) found that goal amount, project category, the number of comments and updates, and the number of social network ties affect project success. Further, a few studies have considered the role of text or image information. For example, Yuan et al. (2016) proposed an improved Latent Dirichlet Allocation (LDA) model to extract textual features from project description and reward description to enhance models that only use numerical features. Yang et al. (2020) introduced image features to improve the performance of project success prediction. However, so far, few studies have examined the interrelations among crowdfunding projects and their impact on crowdfunding performance.

Our research is motivated by the assumption that the performance of crowdfunding projects is interrelated. On the one hand, project founders' previous experiences gained in historical projects can help them operate a new one (Cappa et al., 2020). In other words, the outcomes of their past projects affect those of their future projects. On the other hand, project founders' investing and following behaviors indicate their interests or expertise in corresponding projects. The investing and following behaviors can be interpreted as a signal of their endorsement. In turn, the project founders who perform the investing and following behaviors can learn from the corresponding crowdfunding projects when launching their own projects. According to the aforementioned two mechanisms, we assume that the hidden project interrelations parsed from project founders' online activities can be conducive to predict project success.

Therefore, we construct a project network by identifying project founders' online activities (e.g., launching, investing, and following) and extracting network-based features to improve the performance of project success prediction. First, we propose to construct a project network to profile the implicit relations among crowdfunding projects by analyzing project founders' online activities. Second, we extract network-based features from the project network and combine them with basic numerical features to solve the project success prediction problem. Finally, using a real-world crowdfunding dataset, we perform an empirical analysis to confirm whether these network-based features indeed boost project performance compared to the baselines. To the best of our knowledge, this paper is the first attempt to apply the project network analysis to the success prediction of crowdfunding projects.

The remainder of this paper is organized as follows. Section 2 introduces previous studies related to crowdfunding and success prediction. A survey of node embedding algorithms is also provided. Section 3 focuses on the construction of a project network and describes the proposed model. Section 4 elaborates on experimental procedures and empirical results. Finally, the conclusions of our work and discussions of future research directions are provided in Section 5.

2. Related Works

2.1. An Overview of Crowdfunding

The term "crowdfunding" comes from the concept of crowdsourcing and is originally defined as "an open call, essentially through the Internet, for the provision of financial resources either in form of donation or in exchange for some forms of rewards and/or voting rights in order to support initiatives for specific purposes" (Belleflamme et al., 2013). In addition, many other researchers have proposed similar definitions of crowdfunding (Lehner, 2013; Colgren 2014; Belleflamme et al., 2015). There are two main crowdfunding models: "keep-it-all" and "all-or-nothing". In the first model, the founders can keep all pledges even if their campaigns fail. In the second model, the founders can only receive the funds after their projects succeed (Cumming et al., 2019). Most crowdfunding platforms currently apply the all-or-nothing model. Normally, crowdfunding involves three main stakeholders: the founders who need to raise

[2] http://z.jd.com/bigger/search.html

funds by launching projects, the backers who fund the projects, and the crowdfunding platforms (Kaur and Gera, 2013).

2.2. Determinants of Project Success

Predicting the success of a crowdfunding project is a much-discussed topic in the literature. Related to this, previous studies have explored many factors that influence the performance of crowdfunding projects.

In the early stage, many numerical features have been studied for crowdfunding project success prediction. Li et al. (2016) grouped numerical features into three subsets: static features (can be obtained when the project is launched, such as project goal), dynamic features (can change over time, such as the number of backers), and social media features (can be obtained from social media, such as the number of Facebook friends). Greenberg et al. (2013) analyzed the effects of static features (i.e., project goal, project category, project duration, the number of rewards, the presence of a video, the connection of Twitter or Facebook, and the number of sentences in project description) and social media features (i.e., the number of Facebook friends, and the number of Twitter followers) on project success. Mollick (2014) and An et al. (2014) proved that dynamic features (i.e., the number of backers, the number of updates, and the number of comments) that provide signals of backers' enthusiasm for or interest in a project can help prediction. Chen (2015) evaluated a dynamic feature (i.e., the number of Facebooks shared) and validated that sharing crowdfunding campaigns on social media helps predict success. In addition, Kaur and Gera (2017) counted the number of tweets posted on Facebook and Twitter by backers, founders, or communicators and used the number to improve the performance of success prediction. With the goal of enhancing the interpretability and simplicity of the results, Ryoba et al. (2020) proposed a feature subset selection tool to select features from a whole set of static, dynamic and social media features. Several existing studies have attempted to leverage the social networks of founders and backers for success prediction. Guided by social identity theory, Kromidha and Robson (2016) reported that the social identities of founders or backers within larger social networks can help raise more money. The empirical evidence of this research showed that the number of founders' Facebook friends and the number of sharing by backers can affect the performance of a crowdfunding project.

Meanwhile, the textual description of crowdfunding projects discloses detailed project information to potential backers. Many studies have shown how textual features, such as description length as well as sentiment and topical features of project description, can be used to predict crowdfunding success. For example, Zhou et al. (2018) conceptualized the process by which a founder obtains funding from backers using project description as part of the persuasion process. They identified the length, readability (the ease of understanding), and tone (the ratio of positive and negative words) of textual description in determining the persuasion effect as a potential antecedent of funding success. Wang et al. (2017) argued that sentiment in project description represents a founder's attitude and can impact backers' investment intention. In addition, the topic (Yuan et al., 2016), linguistic styles (i.e., concrete language, precise language, interactive style and psychological distancing) (Parhankangas and Renko, 2017), message framing (positive or negative framing) (Moradi and Dass, 2019), and narrative style (e.g., results in progress and ongoing journey) (Cappa et al., 2020) in project descriptions are also considered as important factors of project success.

Furthermore, rich visual images and videos in the project description increase a project's attractiveness to potential backers. The number of images, as a simple feature, has been introduced to help predict crowdfunding success (Yang et al., 2020; Beier and Wagner, 2015). Chen et al. (2019) also provided some examples to demonstrate the differences in visual styles (e.g., delight vs. horror and bright vs. dark,) between successful and failed projects and validated their role in success prediction. Meanwhile, Kim and Park (2017) analyzed founders' facial expressions in their profiles and found that smiling faces can serve as a significant signal of project quality that helps establish trust. Shi et al. (2021) assumed that multimedia information, such as the combination of textual and audio features, is helpful for project success prediction.

Some studies have also investigated the interactions between founders and backers in predicting project success. For example, Wang et al. (2018) measured interactions, namely, comments from backers and replies from founders. Troise (2020) built an interaction network and extracted topological metrics (e.g., degree centrality and density) to help predict crowdfunding performance. However, the interrelations among crowdfunding projects were rarely investigated. In this study, the interrelations between project p_i and project p_j can be: 1) p_i *has historical experience on* p_j, meaning p_i and p_j are launched by the same founder and p_i finished before p_j, 2) p_i *endorses* p_j, meaning p_j is followed or supported by a founder who have launched p_i, 3) p_i *learns from* p_j, meaning p_j is followed or supported by a founder who will launch p_i.

Other fileds have examined the role of entity interrelations on entity performance. For example, Zhang et al. (2015) believed that stock price movements are influenced by multiple competitors and collaborators, therefore they used the company interrelationships (i.e., cooperation and competition) to predict stock performance. Gitinabard et al. (2019) noticed that students interact with each other when discussing questions on blended courses. He proposed a model to extract student interrelations and used it to predict students' course performance. Similarly, our study holds the view that the performance of crowdfunding projects is interrelated and that the influence will propagate along the network. For example, a founder' success experience in one's past projects can benefit follow-up projects. Moreover, projects launched by different founders can influence one another through the founders' activities (e.g., supporting or following projects). From this point of view, we propose to construct a project network by analyzing founders' activities and extracting the network-based features to enhance the crowdfunding performance prediction.

2.3. Prediction Algorithms on Crowdfunding Success

Previous studies generally adopted machine learning (ML) methods to deal with the crowdfunding success prediction problem. Logistic regression (LR) is widely used in this task due to its high interpretability (Mollick, 2014; Zhou et al., 2018; Kaur and Gera, 2017; An et al., 2014). LR can sometimes work well. For example, Kaur and Gera (2017) compared the performance of LR with the performance of naive Bayes, decision tree (DT) and random forest RF), and found that LR performed best on their dataset. However, LR failed in most cases because the data usually cannot be linearly learned. An et al. (2014) used both LR and support vector machines (SVM) with three different kernels and found that all three SVMs outperformed LR. Many other empirical studies have applied various ML techniques other than LR, including SVM, DT, RF, etc. For example, Greenberg et al. (2013) trained an SVM classifier and a DT classifier on their static and social media features. SVMs were also trained on sentiment features extracted from text through sentiment analysis (Wang et al., 2017). Moreover, the RF is an ensemble algorithm that integrates multiple DTs and can be used to overcome the shortcomings of overfitting and low stability in a single decision tree (Rokach, 2016; Chen et al., 2015; Yuan et al., 2016; Ahmad et al., 2017). Etter et al. (2013) used k-nearest neighbors (KNN) and Markov chain to solve prediction problem. Meanwhile, Li et al. (2016) aimed to estimate the time of success and applied the survival analysis and censored regression approach. Ryoba et al. (2020) used whale optimization algorithm to select features and used KNN to predict project success.

Table 1 lists the main determinants and prediction algorithms used for crowdfunding success. Basic numerical features, especially the static ones, are commonly used features that are often selected as the baseline feature set. LR, SVM, DT and RF are the widely used algorithms for project success prediction. In our work, we aim to examine whether the project network can provide a better solution to the project success prediction problem. In particular, we focused on feature extraction of the project network and tested the performance of the network-based features on these widely used algorithms. Given the superior performance of boosting method and the popularity of deep learning, we also evaluated adaptive boosting (AdaBoost), extreme gradient boosting (XGBoost), light gradient boosting (LightGBM) and deep neural networks (DNN) in our work.

2.4. Node Embedding Algorithms

Graph analysis has attracted increasing attention. In the early years, traditional local network features, such as degree and clustering coefficients, were widely used in network analysis. Node embedding algorithms that extract global structural information of nodes can precisely capture the network features compared to the traditional features. There are three kinds of algorithms for node embedding: factorization, deep learning and random walk.

Factorization-based algorithms use variable matrices (e.g., adjacency matrix, Laplacian matrix, and transition probability matrix) to represent connections in network and embed nodes by matrix factorization, including LLE (Roweis and Saul, 2000), graph factorization (Ahmed et al., 2013), GraRep (Cao et al., 2015), Hope (Qu et al., 2016), etc. Most factorization-based algorithms require the computation of eigenvalues and eigenvectors, which can be time-consuming on large real-world networks.

Deep learning methods, e.g., graph convolutional networks (GCN), solve the node embedding problem by designing a convolution operator on graph (Kipf and Welling, 2016). Similar to the traditional CNN, the new embedding of one node aims to iteratively aggregate its neighbors' embeddings. Velikovi et al. (2017) proposed graph attention network (GAT) as an improved version of GCN by using attention machinism to assign different weights on different neighbors.

As extensions of word2vec in graph anlysis, the basic idea of random walk-based algorithms is that one node affects its neighbor nodes, just like two consecutive words in a sentence. Deepwalk and Node2vec are the two most popular algorithms based on random walks. DeepWalk samples nodes through random walks, and a sample of nodes is regarded as one sentence. Then, word embedding (e.g., word2vec (Mikolov et al., 2013)) can be used to convert those nodes into vectors (Perozzi et al., 2014). However, the random walk strategy in DeepWalk does not consider the edge weights and only explores neighbors in depth-first search (DFS) style. Grover and Leskovec (2016) proposed the Node2vec algorithm with a new sampling strategy that combine DFS and breadth-first search (BFS) style. In addition, the edge weights are involved in the calculation of transition probability between nodes. Therefore, Node2vec has high flexibility and can be used in various contexts. For example, Zhou et al. (2021) used the Node2vec method to detect Internet financial fraud. Meanwhile, Kazemi and Abhari (2020) proposed a model based on Node2vec to learn the representations of papers in the scientific literature network. According to a survey of node embedding algorithms, Node2vec performed best in classification tasks in various datasets compared to factorization-based algorithms and deep learning methods (Goyal and Ferrara, 2018). In summary, Node2vec can learn node embeddings by preserving neighbor nodes in network. Moreover, it can scale to large real-world networks efficiently, making it a famous node embedding technique. Therefore, we chose Node2vec to extract network features in the current study.

Table 1: Determinants and Prediction Algorithms of Project Success

Studies	Algorithms used	Features used
Mollick (2014)	LR	Static and dynamic features
Zhou et al. (2018)	LR	Static features, social media features, and the length, readability and tone of project description
Kaur and Gera (2017)	DT, naive Bayes, RF, and LR	Static and social media features
An et al. (2014)	LR and SVM	Static and dynamic features
Greenberg et al. (2013)	DT and SVM	Static and social media features
Wang et al. (2017)	SVM and sentiment analysis	Statis, dynamic features, and sentiment features of project description
Chen et al. (2015)	RF	Static and dynamic features
Yuan et al. (2016)	Domain-constraint latent Dirichlet allocation and RF	Static and topical features of the project and reward description
Ahmad et al. (2017)	DT and weighted RF	Static and social media features
Etter et al. (2013)	KNN and Markov chain	Static, dynamic, and social media features
Li et al. (2016)	Survival analysis and censored regression	Static, dynamic, and social media features
Ryoba et al. (2020)	Whale optimization algorithm and KNN	Static, dynamic, and social media features
Kromidha and Robson (2016)	Ordinary least squares regression (OLSR)	Static, dynamic, and social media features
Parhankangas and Renko (2017)	LR	Static features and linguistic styles of project description
Cappa et al. (2020)	OLSR	Static and dynamic features and narrative style of project description
Chen et al. (2019)	Glove and convolutional neural networks (CNN)	Static and textual features of project description and visual features of project images
Yang et al. (2020)	OLSR	Static and dynamic features, the text length of project description, and the numbers of images and videos.

3. Proposed Methodology

The proposed framework for crowdfunding success prediction with a project network is shown in Figure 1. This framework includes three main components: network construction, feature extraction, and prediction. In this paper, we propose the construction of a project network by using interrelations among crowdfunding projects. Node embedding method is applied to extract features of the project network. Finally, these features are combined with basic numerical features in the current research to train the widely-used ML models (e.g., LR, SVM, DT, RF, AdaBoost, XGBoost, LightGBM and DNN) to predict project success.

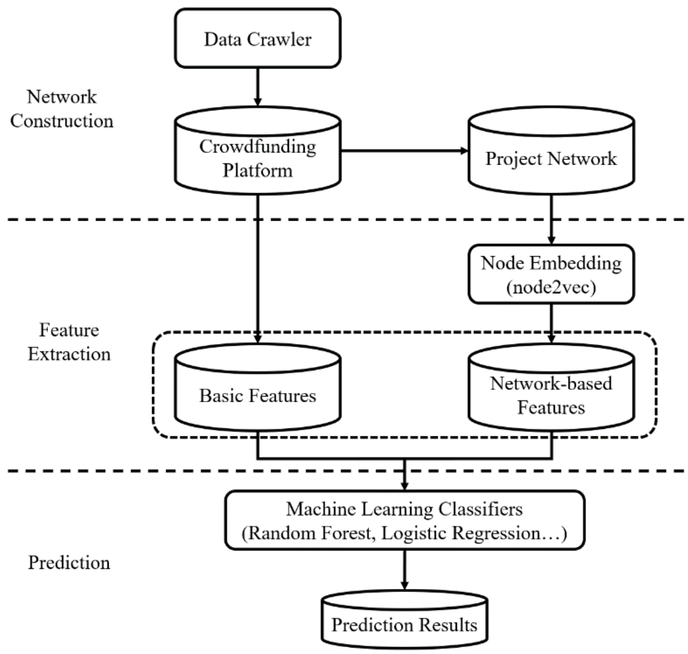

Figure 1: The Framework for Crowdfunding Success Prediction with Project Network

3.1. Network Construction

In this work, we construct a project network by analyzing the project founders' activities. Figure 2 provides an illustration of founders' activities relating to projects, from which interrelations among projects can be extracted. In this example, founder A launched projects p1, p2, and p3 successively, and founder B launched projects p4. The closing times of campaigns p1-p4 are also provided in Figure 2. We assume that the performance of these crowdfunding projects is interrelated given the following mechanisms. Mechanism 1 explains how projects launched by the same founder are interrelated, and mechanism 2 explains how projects launched by different founders are interrelated. Note that founders' activities are not limited to the listed three (i.e., launching, following and supporting), but the three representative activities, compared to only "viewing projects", indicates the strong interrelations among projects. Figure 2 illustrates how we construct a project network for p5 by its launching date (i.e., 2020/12/01) to predict its success probability.

Mechanism 1. Each founder's crowdfunding experience in past projects may affect the performance of subsequent projects (Cappa et al., 2020). We define this type of link as "historical experience". In Figure 2, we use solid arrows to indicate the impact of p4's experience on p5 and the impact of p1's experience on p2 and p3.

Mechanism 2. If a founder follows or invests in another founder's projects, their projects are interrelated by the following way.

First, a founder's following and investing behaviors indicate one's interests and confidence in corresponding projects, which can be interpreted as a signal of endorsements; we define this type of link as "endorse." In Figure 2, founder B followed p1 and invested in p3. Hence, p1 and p3 may be interesting or valuable enough to gain B's attention. If B is an experienced founder, then B's endorsements in p1 and p3 indicates the projects' quality. The experience of B can be interpreted from B's historical experience in operating projects (i.e., p4). Therefore, we build "endorse" links from p4 to p1 and p3.

Second, a founder can improve one's projects by learning from others' crowdfunding projects; we define this type of link as "learn". Founders learn crowdfunding strategies, such as attracting potential backers, from their supporting experiences (Zheng et al., 2016). Learning can be divided into two types: direct and indirect. The indirect experience is learned by observing others' activities (Argote and Todorova, 2007; Darr et al., 1995). In the context of crowdfunding, Yang and Hahn (2015) pointed that founders obtain direct experience by learning from their own previous projects (i.e., what we call "historical experience" in this paper) and indirect experience by backing other founders' initiatives. In the case of Figure 2, founder B followed p1 and invested in p3. Hence, the learning experience from p1 and p3 can affect the performance of p5, i.e., the founder of p5 learns from the founders of p1 and p3.

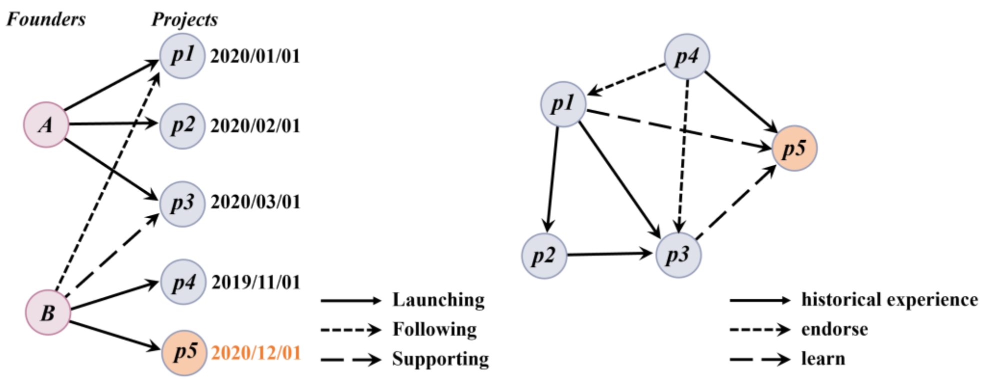

Figure 2: An Illustration of Project Network Construction

3.2. Feature Extraction

We chose the widely-used numerical features as basic features listed in Table 2 by following previous studies (Wang et al., 2020; Yuan et al., 2016; Zhou et al., 2018). The basic features are then combined with network-based features to predict project crowdfunding performance. This research aims to early predict the success probability of a project when the project is just launched, therefore we ignore the dynamic features during the fundraising lifecycle.

Table 2: The Descriptions of Basic Features

Basic features	Description
goal	A founder-defined goal amount when the project was launched
Weibo	Whether a Weibo[3] account was provided on a founder's home page
video	Whether a video was provided on the project page
photo_num	The number of photos used on the project page
reward_num	The number of rewards showed on the project page
project_num	The number of historical projects

We assume that projects within the project network influence mutually and that the interrelations can help enhance the success prediction task. For example, the more historical experience in-edges linked to one project, the more operational crowdfunding experience the founder of this project has, which can contribute to the success of this project. Moreover, a project with many successful neighbor projects is likely to gain guidance or endorsements from its neighbors. Thus, the structural information of the project network is valuable for project success prediction.

We used Node2vec algorithm to extract network-based features from the project network. Node2vec treats each node and a random walk as a word and a sentence, respectively. A random walk can be described as a random process. Performing a random walk on a graph results in a path that consists of a sequence of nodes. For example, in Figure 3, given a starting node t, a random walk visits the next node v with a certain probability. The probability that the current node v will visit the next node x is calculated as follows:

$$P(c_i = x | c_{i-1} = v) = \begin{cases} \frac{\pi_{vx}}{Z} & \text{if } x \text{ is the neighbor of } v \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where Z is the normalized constant, and π_{vx} is the unnormalized transition probability from node v to node x. Here, π_{vx} is determined by the previous node t in the walk and is equal to the product of $\alpha_{pq}(t,x)$ and w_{vx}, i.e., $\pi_{vx} = \alpha_{pq}(t,x)w_{vx}$. In addition, w_{vx} is the edge weight from node v to node x, and $\alpha_{pq}(t,x)$ is a function that can adjust the importance of DFS and BFS. This can be calculated by following equation:

$$\alpha_{pq}(t,x) = \begin{cases} \frac{1}{p} & \text{if } d_{tx} = 0 \\ 1 & \text{if } d_{tx} = 1 \\ \frac{1}{q} & \text{if } d_{tx} = 2 \end{cases}, \tag{2}$$

where d_{tx} is the shortest path distance between node t and node x.

[3] https://weibo.com/

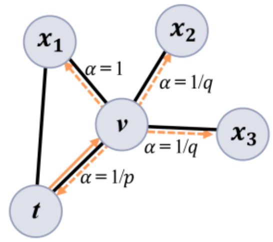

Figure 3: An Example of a Random Walk

Looking at Figure 3, given a starting node t, the random walk has sampled node v, and the next node will be generated in (x_1, x_2, x_3, t). The probability value of visiting the next node can be calculated through Equations (1) and (2). Figure 3 provides the α values calculated by Equation (2) on each link. If $p > max(q, 1)$, the random walk tends to visit the node far away from t, e.g., x_2 and x_3, which is called DFS style. If $p < max(q, 1)$, the random walk tends to visit the node close to t, e.g., x_1 and t,which is called BFS style. Hence, the Node2vec method allows us to smoothly walk with both BFS and DFS through the two parameters p and q.

Node2vec uses Skip-gram model from Word2vec. Skip-gram considers a window of surrounding words $\{w_{m-L}, \dots, w_{m-1},\ w_{m+1}, \dots, w_{m+L}\}$ for a given word w_m, and the goal is to train a model that predicts the surrounding words. Figure 4 shows the architecture of Node2vec. In the Node2vec algotithm, Skip-gram is applied to train a model that predicts the surrounding nodes in random walks based on the given target node. After random walking and training Skip-gram, we can obtain the low-dimensional (dimension size=d) vector representations of nodes in network. In addition, the project network is not static when predicting the performance of projects because the project network is constructed by the launching date for each tagat project. To deal with the dynamics of the project network, we use a dynamic embedding strategy in Node2vec (Mahdavi et al., 2018). When new projects are added to the network, we retrain Node2vec and update the current embeddings of all nodes.

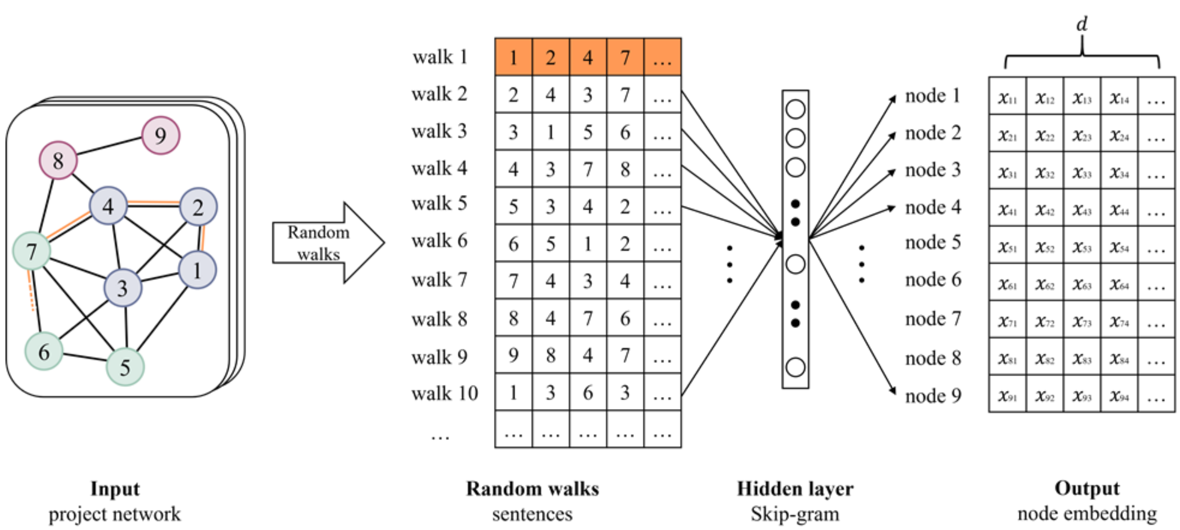

Figure 4: The Architecture of Node2vec

We use the Node2vec algorithm to learn project embeddings in Figure 2. We set the dimension size as 128. Then we use t-SNE (t-distributed stochastic neighbor embedding) (Maaten and Hinton, 2008), an unsupervised machine learning algorithm for the visualization of high-dimensional datasets, to visualize the project embeddings in Figure 5. We can observe that p1, p4 and p3 are close to p5 while p2 is far from p5. The visualization indicates that p1, p4 and p3 are more useful in predicting the performance of p5 compared to p2.

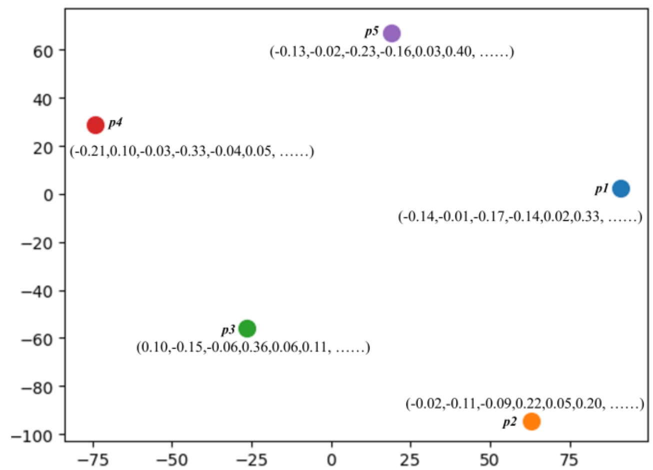

Figure 5: Visualization of Project Embeddings in Figure 2 with t-SNE

3.3. Prediction

We employ the commonly used ML techniques in the field of crowdfunding project success prediction (Greenberg et al., 2013; Mollick, 2014; Kaur and Gera, 2017; Zhou et al., 2018) and state-of-the-art classifiers, including LR, SVM, DT, RF, AdaBoost, XGBoost, LightGBM and DNN, as baselines. A brief overview of the ML methods used in this article is given below.

LR is a model developed from linear regression to deal with classification problems. One advantage of LR is its high interpretability that provides the predictive significance values of features (Kaur and Gera, 2017).

SVM is a classic algorithm proposed by Vapnik to deal with the binary classification problem (Vapnik, 1995). Although SVM can solve both linear and non-linear problems by using kernel functions (e.g., radial basis function (RBF), sigmoid function (SF), etc.). In particular, RBF is the most effective kernel function in most cases. Thus, this paper uses RBF as the kernel function in SVM.

DT is a tree-structured model with conditional control statements. DTs are viewed as easily interpretable models when compared with other classifiers such as black-box deep learning models. This paper applies CART algorithm to prevent overfitting.

Ensemble methods including **RF**, **AdaBoost**, **XGBoost** and **LightGBM** are introduced to combine the base DT classifiers for better prediction performance. RF is an ensemble of decision trees trained with bagging method. AdaBoost, XGBoost and LightGBM are popular boosting algorithms to boost the performance of DTs (Freund, 1997; Chen and Guestrin, 2016; Meng, 2018).

DNN, also known as multilayer perceptron, is a multilayer feedforward artificial neural network model. DNN can improve the performance of conventional neural networks by increasing the number of layers. To overcome the "vanishing gradient" problem due to the large number of layers, activation functions such as ReLU and MaxOut have been proposed to replace Sigmoid.

4. Experiments and Results

4.1. The Dataset

We collected a real-world crowdfunding dataset from JD Finance, one of the most popular reward-based crowdfunding websites in China. Specifically, we retrieved all the projects displayed on the JD Finance website since its establishment up until April 7, 2020. Data from a total of 15,384 projects (12,483 successful projects and 2,901 failed projects) were collected. We divided the training set and the test set by time, using the projects before January 1, 2020 as the training set and the remaining data as the test set. Using the network construction method in Section 3.1, we constructed a project network with 23,934 historical experience links, 8,103 endorse links, and 4,966 learn links. The network density is 2.85e-4. Oversampling was used to deal with the imbalanced dataset in the training set, given that the imbalance can lead to a learning bias (Cang and Yu, 2012).

4.2. The Performance Measures

Like most crowdfunding platforms, JD Finance adopts an all-or-nothing model. In this scenario, each finished project is either a success or a failure. Thus, the project success prediction is a classification task. We used confusion matrix, a widely used tool in classification tasks, to evaluate the models. Table 3 is an example of a confusion matrix, which contains the number of samples correctly classified and incorrectly classified. Based on Table 3, four common performance metrics are defined, namely, Accuracy, Precision, Recall, and F1.

Table 3: Sample Confusion Matrix

Confusion Matrix		Actual Class	
		Success	Failure
Predicted Class	Success	TP	FP
	Failure	FN	TN

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{3}$$

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

4.3. Experimental Results

We compared the performance of selected models to evaluate the role of project network in project success prediction. Eight popular classifiers, namely, LR, SVM, DT, RF, AdaBoost, XGBoost, LightGBM and DNN, were compared in this section. Table 4 provides the parameters used in various classifiers and Node2vec. Details on the performance of the eight models with all features (basic features and network-based features) are provided in Table 5.

Table 4: Parameters Used in Various Classifiers and Node2vec

Algorithm	**Key parameters**	**Parameter interpretation**
LR	penalty = l2	The norm used in the penalization
SVM	kernel = rbf, gamma = 1, C = 10	The type of kernel, the values of γ and C
DT	criterion = gini	The function to measure the quality of a split
RF	n_estimators = 30, criterion = gini	The number of trees, the function to measure the quality of a split
AdaBoost	base_estimator = DT, lr = 0.1, n_estimators=100, ba = samme	The base classifier, the learning rate, the maximum number of classifiers, boosting algorithm
XGBoost	lr = 0.1, n_estimators = 100, objective = binary:logistic	The learning rate, the maximum number of classifiers, the cost fuction
LightGBM	boosting_type = gbdt, n_leaves = 31, objective = binary	The boosting type, the number of leaves, the cost function
DNN	n_layers = 5, activation = Relu, N = [256,512,1024,1024,1024]	The number of hidden layers, the activation function, the number of neurons in each layer
Node2vec	d=128, r=5, k=10, p=1, q=0.5	Embedding size, the number of walks per node, walk length, the values of p and q

Table 5: Prediction Performance (%) of Various Classifiers with All Features

	LR	SVM	DT	RF	XGBoost	LightGBM	AdaBoost	DNN
F1	70.1	76.1	82.7	89.1	88.3	87.9	82.7	84.8
Precision	88.7	87.0	82.8	82.5	82.7	83.6	82.9	83.9
Recall	58.1	67.1	82.6	96.7	94.6	92.6	82.5	85.6
Accuracy	60.0	65.5	72.0	80.5	79.8	79.5	72.1	74.9

As shown in Table 5, RF has the largest F1, Recall, and Accuracy, while LR has the best performance on Precision. Given our focus on the overall performance of classifiers, RF thus shows the best performance with the largest F1 and Accuracy values, outperforming the boosting models and the advanced DNN.

4.4. Ablation Study

Here, we built various feature sets, i.e., E1, E2, and E3, for training models in order to evaluate the role of network-based features in predicting project success. E1 contains only basic features as listed in Table 2. E2 contains only network-based features (embedding via Node2vec). E3 concatenates basic features with network-based features. These three feature sets were then fed to LR, SVM, DT, RF, AdaBoost, XGBoost, LightGBM and DNN, respectively. The experimental results are provided in Table 6.

Table 6: Prediction Performance (%) of Various Classifiers with Various Feature Sets

Feature set	Measure	LR	SVM	DT	RF	XGBoost	LightGBM	AdaBoost	DNN
E1 (only basic features)	F1	69.0	72.9	77.7	79.5	76.9	76.6	77.3	76.2
	Precision	88.9	88.5	85.1	85.1	86.6	87.5	83.9	86.1
	Recall	56.4	62.1	71.7	74.7	69.1	68.1	71.7	68.4
	Accuracy	58.8	62.7	68.8	68.7	66.2	66.2	65.8	65.3
E2 (only network-based features)	F1	69.5	72.1	81.6	88.2	88.4	87.9	80.7	82.9
	Precision	80.8	81.4	81.5	81.5	82.7	82.9	82.3	82.9
	Recall	61.0	64.7	81.8	96.1	95.1	93.6	79.1	82.9
	Accuracy	56.6	59.4	70.2	79.2	79.6	78.9	68.9	72.3
E3 (all features)	F1	70.1	76.1	82.7	89.1	88.3	87.9	82.7	84.8
	Precision	88.7	87.0	82.8	82.5	82.7	83.6	82.9	83.9
	Recall	58.1	67.1	82.6	96.7	94.6	92.6	82.5	85.6
	Accuracy	60.0	65.5	72.0	80.5	79.8	79.5	72.1	74.9

As can be seen from Table 6, RF achieves relatively good performance in all feature sets. Moreover, E3 has the best performance of F1 score over E1 and E2 except for XGBoost. Taking RF as an example, the baseline RF with E1 has F1=79.5%, Precision=85.1%, Recall=74.7%, and Accuracy=68.7%. RF with network-based features (i.e., E2) has F1=88.2%, Precision=81.5%, Recall=96.1%, and Accuracy=79.2%; it also outperforms the baseline RF with E1. The results indicate the good performance of using network-based features to predict project success. In addition, RF with the combined feature set (i.e., E3) achieves the best performance among others. Another observation is that the performance gaps between models with E2 and models with E3 are small. This indicates that the project network can better predict project success even without conventional basic features.

4.5. Analysis of Feature Importance

Given that the RF classifier had the best performance on both F1 and Accuracy, we used RF as the prediction model in the following analysis of feature importance. One advantage of RF is that the model can show importance of features based on Gini impurity or out-of-bag error (Cutler et al., 2014). In this paper, we assessed feature importance based on Gini impurity. Figure 6 shows the ranking of top 12 most important features. Among all features, the number of rewards is the most important one, followed by the number of historical projects and photos. More rewards can attract more backers, because they provide a variety of options to backers. The other nine important features are network-based features. Specifically, the sum of the importance of network-based features (128 dimension) is 92.0%, thus proving useful and valuable role of the project network in project success prediction.

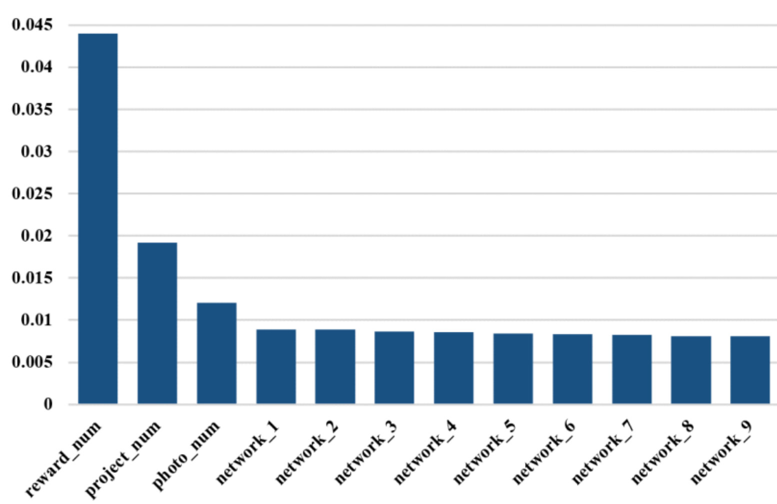

Figure 6: The Importance of Features in RF

4.6. Experiments on the Dense Dataset

To further validate the role of project network, we removed the isolated nodes that were not connected with any other nodes in the original dataset. We labelled the resulting dataset as the "dense dataset". Compared to the original dataset, the dense dataset suffers less from network sparsity problem. The dense dataset consists of 9,141 projects. The network density is $8.08e^{-4}$ (three to four times higher than the density of original dataset). Figure 7 gives an example of the comparison between an original dataset and a dense dataset. The gray nodes at the periphery of the network in the original dataset are isolated with other nodes. In the dense data set, we removed these isolated nodes.

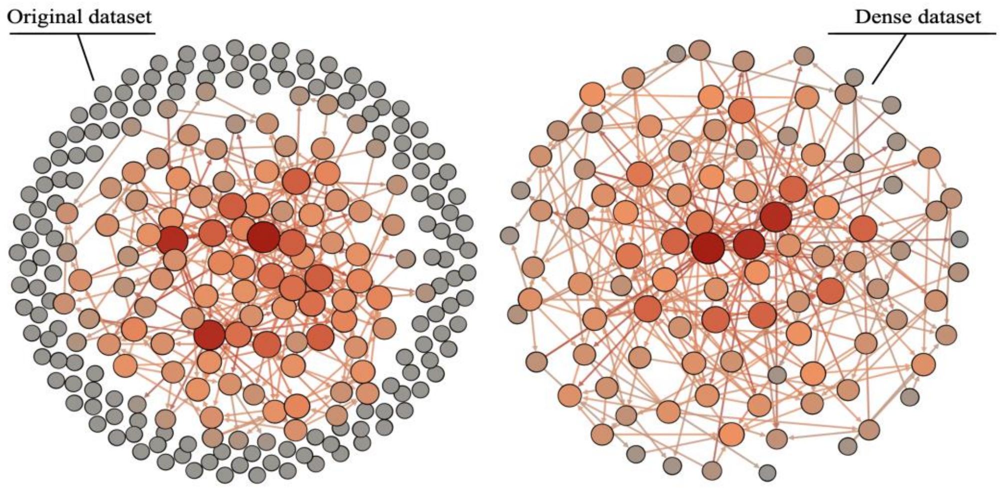

Figure 7: Example of a Comparison between an Original and a Dense Dataset

Results show that classifiers, including LR, SVM, DT, RF, AdaBoost, XGBoost, LightGBM and DNN, perform better with the dense dataset than those with the original dataset. Table 7 shows the experimental results of various classifiers with the dense dataset after using all features. The performance is improved compared to the results in Table 5. The classifier RF can reach F1=93.1% on the dense dataset but can only reach F1=89.1% on the original dataset. The values of Precision, Recall, and Accuracy also increase for RF with the dense dataset compared to those with the original dataset.

Table 7: Prediction Performance (%) on the Dense Dataset

	LR	SVM	DT	RF	XGBoost	LightGBM	AdaBoost	DNN
F1	71.8	76.0	87.3	93.1	92.6	92.7	88.0	89.4
Precision	93.9	92.1	88.2	87.5	88.2	88.1	88.5	88.6
Recall	58.1	64.7	86.4	99.1	97.5	97.8	87.6	90.1
Accuracy	60.1	64.2	78.1	87.2	86.4	86.5	79.2	81.3

The high accuracy on the dense dataset indicates the role of the project network, that is, a densely connected network can better predict project success than a sparsely connected network. This shows that Node2vec can extract more valuable features from a dense dataset. The results also provide clues for future project success prediction, that is, a small proportion of a connected project network can effectively predict project success given that large network calculation involves high calculation cost.

4.7. Experiments for Various Edges

To validate the performance of three types of edges in project network, we trained Node2vec with different edges and made predictions with corresponding embeddings. The results are shown in Table 8. As can be seen, the model with only historical experience edges has the worst prediction performance. The performance of the other two models with only endorse edges or only learn edges is similar, outperforming the model with only historical experience edges. Historical experience edges only exist within the same founders' projects, while endorse edges and learn edges connect projects launched by different founders. Therefore, historical experience edges constitute complete graphs which contain relatively monotonic information. The results indicate that Node2vec extract more valuable features from the network with endorse or learn edges than that with historical experience edges. Table 8 also shows that the model using all edges has the best performance. Also, we used the paired t-test to compare the prediction results of the model with all edges and model with a single type of edges, and the results indicate the significant difference. The significance levels are labelled in the first column of Table 8.

Table 8: Prediction Performance (%) of Various Edges with RF

	Accuracy	Precision	Recall	F1
Only historical experience edges*	79.5	80.9	97.5	88.3
Only endorse edges***	80.3	81.7	97.5	88.9
Only learn edges**	80.3	81.7	97.6	88.9
All edges	80.5	82.5	96.7	89.1

Note: The significance levels indicate the difference between the models with a single type of edges and the model with all edges by using the paired t-test. ***: $p<0.001$; **: $p<0.01$; *: $p<0.05$.

4.8. Discussions

The information hidden in the project network is helpful in achieving successful prediction. On the basis of this intuition, we extracted network-based features from the project network and combined them with basic features to enhance the performance of project success prediction. The experimental results of our ablation study show that the classifiers using all features have good prediction performance, thus proving the value of network-based features extracted from the project network. In addition, the dense dataset with better predictive performance can be extracted from the entire project network. Finally, the results of our empirical tests reveal that the prediction performance varies on different ML classifiers. By using a real-world dataset, we found that RF has the best performance among all classifiers. The possible reason is that RF, as an ensemble learning method, can overcome the shortcomings of overfitting and low stability of a single classifier (Rokach, 2016).

5. Conclusions and Future Works

In the field of crowdfunding project success prediction, previous studies mainly focused on basic numerical features, while recent works have explored the impacts of project descriptions, reward descriptions, and photos used on crowdfunding websites. However, the interrelations among projects have not been fully explored. Thus, the current work aims to fill the aforementioned research gap. In this paper, we propose the concept of “project network” which builds interrelations among crowdfunding projects by analyzing founders’ activities on crowdfunding platforms. Network-based features extracted from the project network are used for success prediction. Based on a real-world crowdfunding dataset crawled from JD Finance website, RF with all features achieved the best performance. A dense dataset with densely connected projects can obtain superior performance than the original one.

This study provides several practical implications. First, founders can benefit from both their historical crowdfunding experiences and those of other founders. For a founder, reviewing one’s past projects, studying the successful experiences of others’ projects, and obtaining support or endorsement from expert founders could help increase the success probability of one’s own follow-up crowdfunding campaigns. Second, backers could evaluate the future performance of a crowdfunding project by analyzing its interrelations with other projects. Third, crowdfunding platforms can implement functions to encourage further interactions among founders or establish intimate interrelations among projects, e.g., recommending successful projects to a new founder or building a tutorial system among founders.

Our future work will enhance the model by mining more valuable information from the project network. In terms of limitations, first, our current study only considers network features achieved with Node2vec. Adding some traditional static network features, such as node degree, clustering coefficient, closeness centrality, and betweenness centrality, may help improve the performance to some extent. Second, future works can construct a heterogeneous network that includes different types of entities, such as projects, founders, and backers. Furthermore, performing heterogeneous network analysis can help capture more information to guarantee a successful prediction. Third, future studies could extract features from dynamic changes of the project network. Last, real-time construction of project network and online deployment of algorithms can further validate the network effect.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (No. 20XNQ038).

REFERENCES

Ahmad, F. S., Tyagi, D., & Kaur, S. (2017). Predicting crowdfunding success with optimally weighted random forests. In *Proceedings of IEEE 2017 International Conference on Infocom Technologies and Unmanned Systems* (pp. 770–775). IEEE.

Ahmed. A., Shervashidze, N., Narayanamurthy, S., Josifovski, V., & Smola, A. J. (2013). Distributed large-scale natural graph factorization. In *Proceedings of International Conference on World Wide Web* (pp. 37–48). ACM.

An, J., Quercia, D., & Crowcroft, J. (2014). Recommending investors for crowdfunding projects. In *Proceedings of International Conference on World Wide Web* (pp. 261–270). ACM.

Argote, L., & Todorova, G. (2007). Organizational learning: Review and future directions. *International Review of Industrial and Organizational Psychology*, 193–234.

Beier, M., & Wagner, K. (2015, December). Crowdfunding success: A perspective from social media and e-commerce. In *Proceedings of International Conference on Information Systems (ICIS)*.

Belleflamme, P., Lambert, T., & Schwienbacher, A. (2013). Crowdfunding: Tapping the right crowd. *Journal of Business Venturing*, *29*(5), 585–609.

Belleflamme, P., Lambert, T., & Schwienbacher, A. (2013). Individual crowdfunding practices. *Venture Capital*, *15*(4), 313–333.

Belleflamme, P., Omrani, N., & Peitz, M. (2015). The economics of crowdfunding platforms. *Information Economics and Policy*, *33*(1), 11–28.

Cang, S., & Yu, H. (2012). Mutual information based input feature selection for classification problems. *Decision Support Systems*, *54*(1), 691–698.

Cao, S., Lu, W., & Xu, Q. (2015). Grarep: Learning graph representations with global structural information. In *Proceedings of International on Conference on Information and Knowledge Management* (pp. 891–900). ACM.

Camacho. D., Lledot, A.P., Orgaz, B. G., & Pardo, A. G. (2020). The four dimensions of social network analysis: An overview of research methods, applications, and software tools. *Information Fusion*, *63*(1), 88–120.

Cappa, F., Pinelli, M., Maiolini, R., & Leone, M. I. (2020). “Pledge” me your ears! The role of narratives and narrator experience in explaining crowdfunding success. *Small Business Economics*, *57*(2), 953-973.

Chen, S. Y., Chen, C. N., Chen, Y. R., Yang, C. W., Lin, W. C., & Wei, C. P. (2015, July). Will your project get the green light? Predicting the success of crowdfunding campaigns. In *Proceedings of Pacific Asia Conference on Information Systems (PACIS)*.

Chen, J., Wu, Y., Fan, L., Lin, X., Zheng, H., Yu, S., & Xuan, Q. (2019). N2vscdnnr: A local recommender system based on node2vec and rich information network. *IEEE Transactions on Computational Social Systems, 6*(3), 456-466.

Cheng, C., Tan, F., Hou, X., & Wei, Z. (2019, August). Success prediction on crowdfunding with multimodal deep learning. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 2158-2164).

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (SIGKDD)* (pp. 785-794). ACM.

Colgren, D. (2014). Tech practices: The rise of crowdfunding social media, big data, cloud technologies. *Strategic Finance*, *95*(10), 56-58.

Cumming, D. J., Leboeuf, G., & Schwienbacher, A. (2020). Crowdfunding models: Keep-it-all vs. all-or-nothing. *Financial Management*, *49*(2), 331-360.

Cutler, A., Cutler, D. R., & Stevens, J. R. (2004). Random Forests. *Machine Learning*, *45*(1), 157-176.

Darr, E. D., Argote, L., & Epple, D. (1995). The acquisition, transfer, and depreciation of knowledge in service organizations: Productivity in franchises. *Management Science*, *41*(11), 1750-1762.

Deepika, S. S., & Geetha, T. V. (2018). A meta-learning framework using representation learning to predict drug-drug interaction. *Journal of Biomedical Informatics*, *84*, 136-147.

Etter, V., Grossglauser, M., & Thiran, P. (2013, October). Launch hard or go home! Predicting the success of Kickstarter campaigns. In *Proceedings of the first ACM Conference on Online Social Networks* (pp. 177-182).

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*(1), 119-139.

Gitinabard, N., Xu, Y., Heckman, S., Barnes, T., & Lynch, C. F. (2019). How widely can prediction models be generalized? Performance prediction in blended courses. *IEEE Transactions on Learning Technologies*, *12*(2), 184-197.

Goyal, P., & Ferrara, E. (2018). Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems*, *151*, 78-94.

Greenberg, M. D., Pardo, B., Hariharan, K., & Gerber, E. (2013). Crowdfunding support tools: Predicting success & failure. In *Proceedings of Extended Abstracts on Human Factors in Computing Systems* (pp. 1815-1820).

Grover, A., & Leskovec, J. (2016, August). node2vec: Scalable feature learning for networks. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (SIGKDD)* (pp. 855-864). ACM.

Herrero, A., Hernandez-Ortega, B., & San Martín, H. (2020). Potential funders’ motivations in reward-based crowdfunding. The influence of project attachment and business viability. *Computers in Human Behavior*, 106.

Kaur, H., & Gera, J. (2017). Effect of social media connectivity on success of crowdfunding campaigns. *Procedia Computer Science*, *122*, 767-774.

Kazemi, B., & Abhari, A. (2020). Content-based Node2Vec for representation of papers in the scientific literature. *Data & Knowledge Engineering*, *127*.

Kim, J., & Park, J. (2017). Does facial expression matter even online? An empirical analysis of facial expression of emotion and crowdfunding success, In *Proceedings of International Conference on Information Systems (ICIS)*.

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, *160*(1), 3-24.

Kraus, S., Richter, C., Brem, A., Cheng, C. F., & Chang, M. L. (2016). Strategies for reward-based crowdfunding campaigns. *Journal of Innovation & Knowledge*, *1*(1), 13-23.

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

Kuppuswamy, V., & Bayus, B. L. (2018). Crowdfunding creative ideas: The dynamics of project backers. In *Proceedings of the Economics of Crowdfunding* (pp. 151-182). Palgrave Macmillan, Cham.

Leimeister, J. M. (2012). Crowdsourcing, Crowdfunding, Crowdvoting, Crowdcreation. *Controlling &Management*. *56*(6), 388–392.

Lehner, O. M. (2013). Crowdfunding social ventures: a model and research agenda. *Venture Capital*, *15*(4), 289-311.

Li, Y., Rakesh, V., & Reddy, C. K. (2016, February). Project success prediction in crowdfunding environments. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining* (pp. 247-256).

Li, Z., Duan, J. A., & Ransbotham, S. (2020). Coordination and dynamic promotion strategies in crowdfunding with network externalities. *Production and Operations Management*, *29*(4), 1032-1049.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(11).

Mahdavi, S., Khoshraftar, S., & An, A. (2018, December). dynnode2vec: Scalable dynamic network embedding. In *Proceedings of International Conference on Big Data* (pp. 3762-3765). IEEE.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *30*, 3146-3154.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

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

Moradi, M., & Dass, M. (2019). An investigation into the effects of message framing on crowdfunding funding level. *Journal of Electronic Commerce Research*, *20*(4), 238-254.

Ou, M., Cui, P., Pei, J., Zhang, Z., & Zhu, W. (2016, August). Asymmetric transitivity preserving graph embedding. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (SIGKDD)* (pp. 1105-1114).

Parhankangas, A., & Renko, M. (2017). Linguistic style and crowdfunding success among social and commercial entrepreneurs. *Journal of Business Venturing*, *32*(2), 215-236.

Perozzi, B., Al-Rfou, R., & Skiena, S. (2014, August). Deepwalk: Online learning of social representations. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (SIGKDD)* (pp. 701-710).

Ribeiro, L. F., Saverese, P. H., & Figueiredo, D. R. (2017, August). struc2vec: Learning node representations from structural identity. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (SIGKDD)* (pp. 385-394).

Rokach, L. (2016). Decision forest: Twenty years of research. *Information Fusion*, *27*, 111-125.

Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, *290*(5500), 2323-2326.

Ryoba, M. J., Qu, S., & Zhou, Y. (2020). Feature subset selection for predicting the success of crowdfunding project campaigns. *Electronic Markets*, 1-14.

Shen, Z., Chen, F., Yang, L., & Wu, J. (2019). Node2vec representation for clustering journals and as a possible measure of diversity. *Journal of Data and Information Science*, *4*(2), 79.

Shi, J., Yang, K., Xu, W., & Wang, M. (2021). Leveraging deep learning with audio analytics to predict the success of crowdfunding projects. *The Journal of Supercomputing*, *77*(7), 1-21.

Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., & Mei, Q. (2015, May). Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 1067-1077).

Troise, C. (2020). Discovering the underlying dynamics of crowdfunding networks: Entrepreneurs' ties, crowdfunders' connections and community spin-offs. *Journal of Enterprising Communities: People and Places in the Global Economy*, *14*(2), 2020.

Vapnik, V., (1995). *The nature of statistical learning theory*. New York: Springer.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2015). APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems*, *75*, 38-48.

Wang, W., Zhu, K., Wang, H., & Wu, Y. C. J. (2017). The impact of sentiment orientations on successful crowdfunding campaigns through text analytics. *IET Software*, *11*(5), 229-238.

Wang, N., Li, Q., Liang, H., Ye, T., & Ge, S. (2018). Understanding the importance of interaction between creators and backers in crowdfunding success. *Electronic Commerce Research and Applications*, *27*, 106-117.

Wang, W., Zheng, H., & Wu, Y. J. (2020). Prediction of fundraising outcomes for crowdfunding projects based on deep learning: A multimodel comparative study. *Soft Computing*, *24*(11), 8323-8341.

Wash, R. (2013, June). The value of completing crowdfunding projects. In *Proceedings of Seventh International AAAI Conference on Weblogs and Social Media*.

Yang, L., & Hahn, J. (2015). Learning from prior experience: an empirical study of serial entrepreneurs in IT-enabled crowdfunding. In *Proceedings of International Conference on Information Systems (ICIS).*

Yang, Y., Wang, H. J., & Wang, G. (2016). Understanding crowdfunding processes: A dynamic evaluation and simulation approach. *Journal of Electronic Commerce Research*, *17*(1), 47–64.

Yang, J., Li, Y., Calic, G., & Shevchenko, A. (2020). How multimedia shape crowdfunding outcomes: The overshadowing effect of images and videos on text in campaign information. *Journal of Business Research*, *117*, 6-18.

Yuan, H., Lau, R. Y., & Xu, W. (2016). The determinants of crowdfunding success: A semantic text analytics approach. *Decision Support Systems*, *91*, 67-76.

Zhang, W., Li, C., Ye, Y., Li, W., & Ngai, E. W. (2015). Dynamic business network analysis for correlated stock price movement prediction. *IEEE Intelligent Systems*, *30*(2), 26-33.

Zhang, Z., Zhou, X., Zhang, X., Wang, L., & Wang, P. (2018). A model based on convolutional neural network for online transaction fraud detection. *Security and Communication Networks*, *2018*.

Zheng, H., Hung, J. L., Qi, Z., & Xu, B. (2016). The role of trust management in reward-based crowdfunding. *Online Information Review*, *40*(1).

Zheng, H., Xu, B., Wang, T., & Xu, Y. (2017). An empirical study of sponsor satisfaction in reward-based crowdfunding. *Journal of Electronic Commerce Research*, *18*(3), 269.

Zhou, M. J., Lu, B., Fan, W. P., & Wang, G. A. (2018). Project description and crowdfunding success: An exploratory study. *Information Systems Frontiers*, *20*(2), 259-274.

Zhou, H., Sun, G., Fu, S., Wang, L., Hu, J., & Gao, Y. (2021). Internet financial fraud detection based on a distributed big data approach with node2vec. *IEEE Access*, *9*, 43378-43386.

Zia, K., Shafi, M., & Farooq, U. (2020). Improving recommendation accuracy using social network of owners in social internet of vehicles. *Future Internet*, *12*(4), 69.