The Role of Novelty in Securing Investors for Equity Crowdfunding Campaigns

Emőke-Ágnes Horvát
School of Communication
Northwestern University
Evanston, USA

Johannes Wachs
Center for Network Science
Central European University
Budapest, Hungary

Rong Wang
School of Communication
Northwestern University
Evanston, USA

Anikó Hannák
Center for Network Science
Central European University
Budapest, Hungary

Abstract

In recent years crowdfunding has diversified and grown beyond most experts' projections. Originally aiming to serve venture ideas and entrepreneurs outside the focus of traditional capital markets, the crowdfunding marketplace has developed a complicated relationship with novel ideas. Yet, there is little to no research on the relationship between project novelty and success in crowdfunding. This paper measures the novelty of crowdfunding campaigns using the content and language of their pitches, capturing their tendency to combine different venture sectors and topics in distinctive ways. Using a unique data set that covers four years of activity on a leading equity crowdfunding platform, we investigate the link between novelty and success, as well as how novelty appeals to different kinds of investors. We find that novelty derived from campaign pitches is negatively related with fundraising success even when controlling for quality and style of writing. We also find that novel campaigns are more likely to attract less-frequent, large-sum investors. Our findings contribute to the long-standing debate related to the trade-offs between innovativeness and conventionality in maximizing chances of startup survival. Our results also have important implications for entrepreneurs writing fundraising pitches and for platform providers who wish to facilitate successful innovation.

Introduction

Novelty lies at the center of entrepreneurship (Davidsson and Wiklund 2001; Tan, Shao, and Li 2013). Existing literature on the topic has examined how novelty as a form of competitive differentiation and distinctiveness impacts startup survival and productivity (Hyytinen, Pajarinen, and Rouvinen 2015; Jennings, Jennings, and Greenwood 2009). On the one hand, novelty brings innovative products and services to the market and faces thus less competition. On the other hand, novelty increases uncertainty for potential financiers, which makes fundraising more challenging (Hyytinen, Pajarinen, and Rouvinen 2015; Cunningham 2017). Due to these conflicting forces, the value of novelty in entrepreneurship has been debated for decades. Studies attempting to uncover its effect on venture success rely on observational, experimental, and survey data. These studies, however, each have limitations as they: *i)* either focus on only the entrepreneurial or investor side, *ii)* lack detailed information about the business proposals, or *iii)* solely consider successful ventures, leading to survival bias in the analysis. Because of these limitations, previous studies are at risk of finding spurious correlations between innovation and business outcomes or missing the most relevant dimension of novelty.

The recent rise of Web-enabled crowdfunding systems offers abundant data to overcome these limitations: they provide information about entrepreneurs and investors alike for both successful and failed crowdfunding campaigns. We use a unique set of data from a leading equity crowdfunding platform that provides startup funding to mostly British ventures from European investors. Using comprehensive data from 2012 to 2016 of over 77,000 investments into 698 campaigns representing a wide range of economic sectors, this paper examines empirically *how an early-stage startup's novelty, measured using the text of its pitch to investors, affects campaign success*.

Equity crowdfunding allows entrepreneurs and startups to attract funds from a large group of investors (i.e., the crowd) in return for a stake in company ownership through an online platform (Mollick 2014b; Cholakova and Clarysse 2015; Vulkan, Astebro, and Sierra 2016; Drover et al. 2017). What makes equity crowdfunding different from other methods of raising early capital for startups is the reliance on the crowd reached via the Internet: entrepreneurs make an open call over an online platform where potential funders evaluate the information provided in the campaign pitch (Ahlers et al. 2015). This method scales up from the so-called *family-and-friends* funding model to professional financial markets, which are full of investors motivated by the prospect of excess returns (Cholakova and Clarysse 2015). Given the impressive growth and current maturity of equity crowdfunding (Drover et al. 2017), not only that the phenomenon provides an excellent framework for examining important open questions in entrepreneurship, but online capital markets have become important in their own right as part of a wealth of transformative decentralized markets. Other examples include online labor markets such as Airbnb, TaskRabbit and Uber, where research emphasizes the complexity of success. The combination of algorithms, website design, and the human biases brought into the system have crucial impact on outcomes (Fradkin et al. 2015; Thebault-Spieker, Terveen,

and Hecht 2017; Wachs et al. 2017; Hannák et al. 2017). Here too, it is unclear whether the platforms truly spur innovation.

Existing literature on equity crowdfunding investigates the link between funding success and *a.)* the human and social capital of entrepreneurs (Ahlers et al. 2015), *b.)* their previous experience and gender (Marom and Sade 2013; Horvát and Papamarkou 2017), *c.)* the quality of the venture along with associated risks (Ahlers et al. 2015; Mohammadi and Shafi 2018), as well as *d.)* the early performance of a campaign (Vulkan, Astebro, and Sierra 2016). However, to the best of our knowledge, no research has looked into aspects of novelty in association with equity campaign success. There are also relatively few works on the outcomes of crowdfunding which use text data (Mitra and Gilbert 2014). Likely these two gaps are linked: without mining the information contained in the text of the pitch, ventures can only be differentiated in general ways such as their economic sector, size, and team composition. As equity crowdfunding requires convincing investors from a broad, digital audience, we claim that the textual data encodes important information about the potential success or failure of a venture. Moreover, the style and presentation of the pitch gives investors valuable signals (Ahlers et al. 2015).

To address these gaps, we extract a variety of text-based features from the campaign pitches. These features include linguistic style, quality of writing, and the relative emphasis of the pitch on past accomplishments of the team behind the venture. Then, to tackle the difficulties in quantifying a multifaceted concept like novelty, we draw on language expectancy theory (Burgoon, Denning, and Roberts 2002; Parhankangas and Renko 2017) and entrepreneurship literature (Hyytinen, Pajarinen, and Rouvinen 2015). Ingrained in the idea that innovation is facilitated by the recombination of previous ideas (Weitzman 1998; Fleming 2001), our approach consists in extracting a novelty measure from campaign text that quantifies the tendency to combine a diversity of topics into the campaign pitch. The so-called *topic entropy* (Hall, Jurafsky, and Manning 2008) measures the extent to which campaigns defy categorization relative to their peers. We show that this specific quantification of the complex notion of novelty is negatively associated with success, supporting the theory of innovation maximization fallacy (Hyytinen, Pajarinen, and Rouvinen 2015; Chan and Parhankangas 2017). We also find interesting relationships between novelty and the kinds of investors campaigns attract. Our results contribute to the long-standing debate related to the theoretical link between innovativeness and startup proliferation. Beyond theory, our findings also have useful implications for practicing entrepreneurs and platform providers.

Related Work

Developments in Web-based systems have impacted business organization dramatically. Combining the key innovations of crowdsourcing and open marketplaces (Howe 2008; Adjei, Noble, and Noble 2010), crowdfunding has emerged as an attractive platform for entrepreneurial initiatives by individuals and firms seeking to obtain capital from a large pool of investors (Belleflamme, Lambert, and Schwienbacher 2014). Some suggested that crowdfunding supports innovation because it facilitates unorthodox ventures underserved by traditional capital markets (Riedl 2013). Anecdotal evidence about the motivation of investors aside, little is known about how projects are evaluated (Cholakova and Clarysse 2015) and whether investors are actually attracted to certain kinds of unconventionality. We argue that the question is even more interesting in the case of equity crowdfunding, where investors buying equity have an ownership stake in the company, instead of simply enjoying novel consumer experiences (Lukkarinen et al. 2016). To theorize how equity crowdfunding might enable the funding of innovative ventures, we consider: *1.)* entrepreneurship literature on the different types of novelty and their debated impact on firm survival, *2.)* research on other forms of crowdfunding that introduce investigations of novelty in relation to fundraising success, and *3.)* communication literature on persuasion and language expectation.

Entrepreneurship as a broad research domain is concerned with novelty and value creation in the economy. Its literature talks in length about the tension between the necessity and inherent uncertainty of innovativeness. Several theoretical arguments and empirical evidences suggest that novelty offers survival-enhancing capabilities for instance by increasing a firm's market power and ability to escape competition (Hyytinen, Pajarinen, and Rouvinen 2015). Countertheories and example cases indicate an innovation maximization fallacy: innovativeness leads to uncertain payback times, which limits access to external funding and results in greater likelihood of failure (Chan and Parhankangas 2017). As it is unclear how novelty impacts success, scholars are developing the appropriate data and methods to uncover context-dependent links between specific types of novelty and firm survival that apply to young ventures. For instance, Jennings et al. studied new companies established within a single regional district of the legal profession and found a U-shaped relationship between employment-system novelty and organizational productivity (Jennings, Jennings, and Greenwood 2009). Cunningham investigated a set of medical device startups from the U.S. and showed that technological innovativeness had no impact on competitive advantage (Cunningham 2017). Dutta and Folta analyzed the Angel Investors Performance Project survey data and found that the number of patents and patent citations, used as a proxy for novelty, correlate with receiving funding (Dutta and Folta 2016). Inspired by the ambitions of these and similar studies, we look at a new context (that of equity crowdfunding) rich in details, containing both successes and failures, which allows us to test for a great variety of novelty-related factors investigated in entrepreneurship literature.

Studies of the effects of project novelty on crowdfunding performance are scarce and have focused on post-production performance (Xu et al. 2016). Most relevant to our paper, is the recent work of Chen and Parhankangas who show in the context of the rewards-based crowdfunding site Kickstarter that radical innovativeness performs poorly, while more moderate, incremental novelties balancing familiarity, expected benefits, and feasibility result in favorable funding

outcomes (Chan and Parhankangas 2017). This study measured innovativeness through online surveys where Amazon Turkers evaluated campaign videos. Here, we are aiming for a more scalable approach based on the automatic analysis of pitch text. Evidence from a donation-based crowdfunding platform suggests that similarity to other projects on the same platform has negative effects on funding success (Meer 2014), in line with the idea that highly conventional projects face higher competition.

The most straightforward way in which new ventures on crowdfunding platforms can appeal to funders is through effective communication about their legitimacy and potential (Parhankangas and Renko 2017). Signaling theory has been used before to identify communication-based attributes that influence fundraising success (Ahlers et al. 2015). Mentions of the human and social capital of entrepreneurs and the quality of the product or service (Agrawal, Catalini, and Goldfarb 2011; Mollick 2014b) as well as past experience and prior success along external certification like awards, government grants and patents (Ahlers et al. 2015; Marom and Sade 2013) have been shown to facilitate fundraising. Besides the credentials and embedding narratives, recent evidence suggests that pitch language itself trough the linguistic style and word usage can shape the perception and funding chances of a new venture (Allison et al. 2015; Mitra and Gilbert 2014; Parhankangas and Renko 2017; Xu et al. 2014). Our work investigates both the content and the style of language to provide a holistic picture of the important factors impacting fundraising. Moreover, we build our novelty measure around the pitch text. In doing so we use key insights from language expectancy theory which states that in order to succeed, entrepreneurs need to meet the expectations of their target investors (Burgoon, Denning, and Roberts 2002). Accordingly, a viable measure of novelty in crowdfunding should occupy the space between investor expectations and the strategic positioning of the firm among established venture types and sectors.

Data, Methods, and Measures

In this section we describe the data and the methods we use to generate additional measures that are associated with crowdfunding success.

Crowdfunding Data

The site we study is a leading UK-based equity crowdfunding platform that provides capital to entrepreneurs to launch or expand their businesses. Our sample consists of 698 campaigns launched between May 22, 2012 and January 24, 2016. In this sample, 21,979 unique investors from the greater European area made 77,628 pledges. The platform works according to an all-or-nothing model, meaning that a campaign is successful if it becomes fully funded by reaching or exceeding its funding goal. In line with existing literature (Mollick 2014b), we use this criteria to define our dependent variable. 31.9% of the campaigns managed to raise their targets, the median of which was £80,000 (interquantile distance: £100,000). In exchange for the amount it seeks to raise, the median campaign offered 10% of its equity (interquantile distance: 8.1%). This percentage is stable regardless of the type of venture proposed. Campaigns are labeled by the entrepreneurs' self-reported high-level venture sectors including *Health*, *Entertainment*, and *High Technology*.

Campaign Pitch Text as Data

This paper focuses primarily on the text data associated with each campaign. Specifically, each campaign has a detailed pitch to potential investors in which entrepreneurs attempt to make their case. Just like previous work (Ceyhan, Shi, and Leskovec 2011), we control for the length of the campaign pitch as this may indirectly influence the text-based variables we introduce next.

Relative Section Lengths Uniquely to this platform, campaign pitch text is consistently split up into nine sections. These are: *Introduction*, *Intended impact*, *Substantial accomplishments to date*, *Monetisation strategy*, *Use of proceeds*, *Target market*, *Characteristics of target market*, *Marketing strategy*, and *Competition strategy*. We compare the relative character lengths of each section between campaigns by dividing by their total lengths. Given the importance of signaling in crowdfunding markets (Ahlers et al. 2015), we focus on the relative length of *Substantial accomplishments to date*. We examined a random sample of campaigns and found that this section frequently contained information both about past performance of the venture (revenue, number of customers, patents or copyrights in hand) and signals about the human capital of the founders (degrees, awards, previous ventures). We show later that it is the only section whose excess length consistently predicts success of campaigns.

Writing Quality Previous work indicates that spelling errors are negatively correlated with success (Mollick 2014b). Inspired by this idea, we quantify writing quality using *Proselint*,[1] an automated tool that suggests improvements to written text. In contrast with spelling and grammar checkers, Proselint suggests more qualitative improvements to English language text. For example, it checks for consistency in spelling and spacing, for use of archaic forms, cliches, or jargon, and for redundancy or excessive use of "weasel-words". We run Proselint on each campaign's text and count the number of suggestions made by the program. Scaling for text length, we hypothesize that campaigns with fewer Proselint suggestions have higher writing quality.

Verbal Content and Style The Linguistic Inquiry and Word Count (LIWC) dictionary and associated software are the most commonly used language analysis tools for investigating stylistic aspects of language (Pennebaker, Mehl, and Niederhoffer 2003; Tausczik and Pennebaker 2010). LIWC matches stemmed words with a comprehensive dictionary whose entries have been assigned by human experts into over 90 dimensions that range from the basic linguistic (e.g., adjectives, prepositions) to the psychological level (e.g., negative emotions). The tool returns the percentage of words that fall into each category. LIWC has been used in

[1] https://github.com/amperser/proselint

Topic Number	Keywords
1	art, luxuri*, tower, cowork*
2	properti*, agent, estat*, landlord
3	idea, startup, player, signag*
4	merchant, loyalti*, vyke, call
5	music, bar, studio, care
6	properti*, investor, loan, lend
7	devic*, softwar*, manufactur*, switch
8	film, dog, collect, pet
9	sport, footbal*, fan, ticket
10	car, wed, learn, bike
11	wine, travel, hotel, card
12	book, guest, film, properti*
13	photo, video, chees*, print
14	golf, news, vape, player
15	fashion, magazin*, boat, owner
16	student, school, educ*, univers*
17	food, game, restaur*, driver
18	project, job, recruit, investor
19	film, bank, stove, sport
20	chariti*, children, donat*, devic*

Table 1: Keywords from each topic. Stars denote stemmed words.

the crowdfunding literature to show the importance of effective communication in case of campaigns delivering social good (Parhankangas and Renko 2017). Our paper analyzes the verbal content and style of pitches to look for elements of language that are associated with success, regardless of the type or sector of a venture. As an example, we visualize the distribution of the use of psychologically negative words in the first plot of Figure 1. The difference between successful (S) and failed (F) campaigns is statistically significant (Kolmogorov-Smirnov test, $p < 0.01$) with successful campaigns containing more negative emotion ($M_S = 0.41$ and $IQR_S = 0.39$ vs $M_F = 0.31$ and $IQR_F = 0.32$).

Campaign Topics We model the topics occurring in our corpus of campaign pitches using Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003). Topic modeling algorithms are machine learning methods for discovering systematically the thematic structure and linguistic context of large text data. Within this framework, LDA is a widely accepted unsupervised approach. We apply standard NLP preprocessing techniques like stop-word removal and stemming and use LDA with 20 topics thereby associating each campaign with a 20-dimensional topic vector[2]. We show the keywords associated with each topic in Table 1. Similar approaches have been used before to analyze crowdfunding outcomes: An et al. found that topic modeling of campaign text can provide valuable information for recommending campaigns to investors (An, Quercia, and Crowcroft 2014). We use the topics as a basis for a novelty measure that captures the mixture of topics in a pitch as follows.

[2]We replicated our subsequent findings with 10, 20, and 30 topics and across a range of the tunable LDA parameters.

Topic Entropy We measure the extent to which campaigns are spread across topics using entropy. Low topic entropy campaigns are highly concentrated in fewer topics, while high topic entropy campaigns draw from many more topics. Recall that n-dimensional topic modeling generates an n-dimensional vector for each campaign in which the j-th entry of a campaign's vector measures the extent to which the campaign fits into the j-th topic. The entries in the vector sum to unity, facilitating a probabilistic interpretation of the vectors. Given the n-dimensional topic vector T_i of campaign i, its *topic entropy* is defined as:

$$TE_i = -\sum_{j}^{n} T_i(j) \log(T_i(j)).$$

We normalize topic entropy by the log of the number of topics in our topic model so that the measure is bounded between 0 and 1. As shown in the middle plot of Figure 1, successful campaigns have a significantly lower topic entropy than failed ones ($M_S = 0.41$ and $IQR_S = 0.51$ vs $M_F = 0.51$ and $IQR_F = 0.32$; Kolmogorov-Smirnov test, $p < 0.01$).

Topic entropy has been used to study the evolution of academic fields (Hall, Jurafsky, and Manning 2008) by quantifying the diversity of topics appearing in academic conferences. A more recent study uses topic entropy to measure diversity of theme in political discourse on Twitter during protests (Munger et al. 2018). Similarly, we suggest that the topic entropy of a campaign measures the extent to which it combines different ideas. Excess topic entropy can be interpreted as the unpredictability of sector combinations or a lack of focus. Since we train the topic model on the entire set of campaigns, we claim that campaigns that occupy multiple topics are highly distinctive and hence novel. The LDA algorithm makes it highly unlikely that there are significant repeated pairings of topics across campaigns[3]. As investors on the platform make decisions under high levels of uncertainty, we hypothesize that the novel, high topic entropy campaigns are less likely to succeed. Our methodology is also similar to approaches taken by research on measuring novelty in creative fields. In one example, short artistic videos are clustered by their features and novelty is measured by each video's distance from centroids of the clusters (Redi et al. 2014). In another, the feature space of graphic designs is approximated by a Gaussian mixture model, and the novelty of a new image is define by a likelihood function (Wachs et al. 2018). In both cases, the novelty of a product is measured by its deviation from some classification of items into groups. More novel items are the ones which are most difficult to classify. In our case, the groups are topics and the measure of the difficulty of the classification of a campaign is topic entropy. Finally, observe that novelty as measured through topic entropy does not have a temporal component. Given the time scale and coverage of our data set, temporal extensions of the topic entropy were not meaningful.

[3]Indeed, the robustness of our findings to a range of topic numbers supports this interpretation.

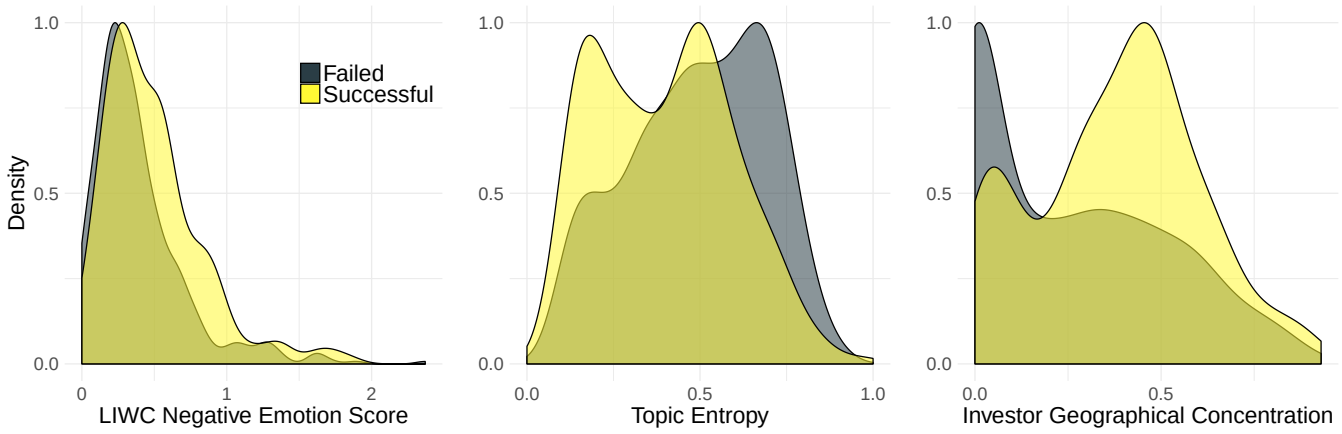

Figure 1: Distributions of negative emotion score, which is quantified by the percentage of words that match negative psychological processes; topic entropy, which measures the novelty of the campaign pitch; and investor geographical concentration, which measures the country-level concentration of investors contributing to the campaign. The difference between successful and failed campaigns is statistically significant in all three cases.

Investor Diversity

A crucial aspect of startup success in crowdfunding is the investor base mobilized by the campaign. Previous work on modeling fundraising success has recognized the importance of the sheer number of contributing funders (Ceyhan, Shi, and Leskovec 2011), especially in the early stages of the campaign (Colombo, Franzoni, and Rossi-Lamastra 2015). There has been less work on differentiating investors—for exceptions see (Agrawal, Catalini, and Goldfarb 2011; An, Quercia, and Crowcroft 2014). The idea is that different campaign strategies will inherently attract different investors, as it has been argued most importantly, in terms of their location and investment behavior.

Geographical Concentration Geographic considerations are particularly interesting due to crowdfunding's ability and goal to eliminate the spatial constraints that separate funders and entrepreneurs (Agrawal, Catalini, and Goldfarb 2011; Lin and Viswanathan 2016). To measure the geographical concentration of a campaign i's investors, we compute the Herfindahl index over the share of countries S where investments come from as:

$$GC_i = \sum_{j}^{c} S(j)^2,$$

where c is the number of countries. The index ranges from $1/c$ to 1, where low values indicate high geographic diversity and high values indicate high geographic concentration. The distribution of investor geographic concentration for successful and failed campaigns in the third plot of Figure 1 shows that on average, successful campaigns are associated with more concentrated investors ($M_S = 0.4$ and $IQR_S = 0.29$ vs $M_F = 0.21$ and $IQR_F = 0.45$; Kolmogorov-Smirnov test, $p < 0.01$).

Investment Frequency and Amount Another aspect of diversity, specific to crowdfunding, lies in the range of investor resources and strategies (Mollick 2014a). Here, we focus on the two most straightforward dimensions of investment behavior that result from these differences. For each investor we determine whether the number of investments they made on the platform, and the value of those investments lie in the low, medium, or high categories. In line with previous characterizations of investors (An, Quercia, and Crowcroft 2014), we say that an investor is a "low/medium/high frequency investor" if they pledged two times or less, between 3 and 8 times, or more than 8 times. Similarly, an investor is considered "low/medium/high amount investor" if their overall contribution on the site was less than £100, between £100 and £1,000, or greater than £1,000. To aggregate these investor-level variables to individual campaigns, we determine the percentage of contributions coming from each of the investment activity categories.

Results

We run various logistic regression models to predict the binary outcome of campaigns: venture has raised at least the target amount (1) or not (0). Our fully specified model contains adapted and extended variables that have been hypothesized to influence fundraising success in equity crowdfunding (including campaign target, percentage equity offered, and proxies for human and intellectual capital); in other types of crowdfunding; and in offline entrepreneurship literature (entrepreneurs' past record, venture sector, and novelty of idea). We report both log-likelihood and McFadden's pseudo-R^2 as measures of model fit.

We show our primary findings in Table 2. All continuous features are standardized to have mean 0 and a standard deviation of 1. Our models indicate that features derived from pitch text and investor behavior significantly predict success even when controlling for a suite of different factors that probe the generalizability of our findings. To assess model fit, we also report a likelihood ratio test value for each model compared to the previous one. Here, significance indicates that we reject the null hypothesis that the focal model does

not significantly improve on the previous model.

The Role of Novelty in Predicting Fundraising

Model 1 represents our baseline for the prediction problem in the current context. In contrast to most studies on crowdfunding (Mollick 2014b), the size of the campaign target is not a significant predictor of success. We suggest that this may be due to the richness of the sector labelling in the data. We also see that the amount of equity offered is only significant before we control for the text features, although high values have been broadly considered to be a negative signal (Ahlers et al. 2015; Vulkan, Astebro, and Sierra 2016; Mohammadi and Shafi 2018).

Model 2 expands the first one by introducing topic entropy, our primary measure of the novelty and distinctiveness of a campaign's pitch, and indicates that topic entropy is a negative predictor of success. According to the likelihood ratio test for the two models, Model 2 improves significantly on the base model. Recall our interpretation of topic entropy as distinctiveness: if any specific combinations of topics were extremely prevalent among the campaigns, the topic model would have shifted to make the underlying combination of words a topic itself. Our findings are robust for a larger number of topics. We also note that other measures of the uniformity of the distribution of the topics like the Herfindahl index yield similar results. Specifically, a higher Herfindahl index, indicating high concentration in few topics, is positively related with the success of the campaign.

Novelty remains a significant predictor of fundraising success even when accounting for a whole range of additional variables in Models 3–5. We find that text variables are significant and that their inclusion generally improves model fit. The relative amount of text spent on sharing past accomplishments is a consistently positive predictor of success. Poor writing quality, proxied by the count of Proselint suggestions the text generates is a significant predictor of failure. Interestingly, simple measures of human capital, like mentioning MBA or PhD degrees, are not significantly predictive of success, c.f. (Ahlers et al. 2015) and their inclusion does not significantly improve our model. Similarly, intellectual capital, as measured by patents and awards, have no significant impact on funding success either. Finally, the selected LIWC features perform well in the models. The heavy use of punctuation (e.g., dashes and exclamation marks), prepositions, and adjectives are positively associated with success. Negative emotions expressed through the pitch text have the largest effect among the LIWC variables: going from one standard deviation below the mean (.09%) to one above it (.85%) in the frequency of negative words increases chance of success by a factor of 2.

To better understand the impact of topic entropy on success we rerun Model 5 as a generalized additive model (Hastie and Tibshirani 1990), allowing a non-linear relationship between topic entropy and success. We plot the model's prediction of campaign success as a function of topic entropy in Figure 2. Accordingly, moving from one standard deviation below average topic entropy to one standard deviation above average translates to roughly a 10% decrease in success.

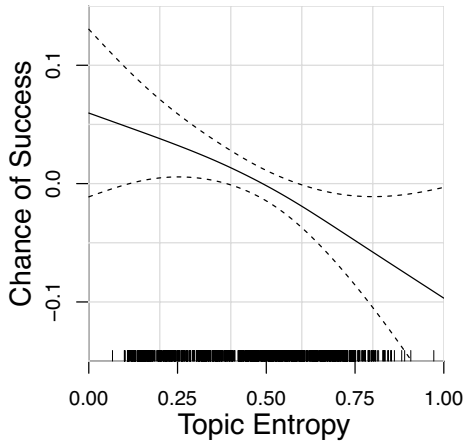

Figure 2: A plot of Model 5 rerun as a generalized additive model, with topic entropy allowed to vary smoothly. The plot shows the relationship between topic entropy and campaign success.

Effects of Investor Diversity on Success

Model 6 improves significantly on Model 5 and achieves a pseudo-R^2 of 0.68. It indicates that campaigns are more or less successful depending on the composition of their investors. We see that the overrepresentation of high amount investors is actually a negative for campaigns. In other words, campaigns on this platform are succeeding because of the mid and low-quantity investors, i.e., those investing less than £1,000. Attracting the most active investors, on the other hand, is a significant positive signal. Finally, successful campaigns tend to attract a geographically concentrated group of investors.

Given the apparent importance of mobilizing different kinds of investors we now investigate whether novel campaigns, measured by their topic entropy, are likely to attract different kinds of investors. Such a relationship may explain how novelty as we define it hinders success. For each investor feature we calculated, we split the campaigns into two groups. We consider campaigns with above and below the median of high amount investors, high frequency investors, and geographic concentration. We plot the distribution of topic entropy across these groups in Figure 3. In two cases we find a significant statistical difference: campaigns drawing more high amount investors typically have higher topic entropy (Kolmogorov-Smirnov test, $p < 0.05$). Campaigns attracting investments from frequent donors have lower entropy (KS test, $p < 0.05$). On the other hand, there is no significant difference in topic entropy between campaigns attracting more or less geographically concentrated investors (KS test, $p = 0.72$).

All in all, given the importance of small and mid-sized investors in campaign success, equity crowdfunding is not an ideal platform for highly novel ideas, provided that novelty is measured at the level of campaign pitches. Apparently, simple and straightforward campaign pitches do bet-

	Dependent Variable: Campaign Success					
	(1)	(2)	(3)	(4)	(5)	(6)
Campaign Target (log)	−0.19	−0.13	−0.12	−0.13	−0.12	0.004
	(0.16)	(0.16)	(0.17)	(0.17)	(0.18)	(0.21)
Equity Offered (%)	−0.35**	−0.34**	−0.34**	−0.33**	−0.21	−0.21
	(0.15)	(0.15)	(0.15)	(0.15)	(0.17)	(0.18)
Text Length (log)	−0.14	−0.32*	−0.21	−0.23	−0.34*	−0.44**
	(0.17)	(0.19)	(0.19)	(0.19)	(0.21)	(0.22)
Topic Entropy		−0.45**	−0.47***	−0.46**	−0.48**	−0.39**
		(0.18)	(0.18)	(0.18)	(0.19)	(0.20)
Rel. Length "Past Accomplishments"			0.42**	0.43**	0.39*	0.32
			(0.20)	(0.20)	(0.21)	(0.21)
Proselint Suggestions (log)			−0.23	−0.21	−0.31**	−0.39**
			(0.14)	(0.15)	(0.15)	(0.17)
Mentions MBA				−0.66	−0.78	0.02
				(2.21)	(3.14)	(3.88)
Mentions Patents				0.62	0.59	0.59
				(0.53)	(0.56)	(0.58)
LIWC: All Punctuation					0.53***	0.47**
					(0.19)	(0.21)
LIWC: Negative Emotion					0.34**	0.38**
					(0.15)	(0.16)
Prepositions					0.55***	0.54***
					(0.18)	(0.20)
Adjectives					0.53***	0.62***
					(0.18)	(0.19)
High Amount Investors (%)						−0.68**
						(0.30)
High Frequency Investors (%)						0.63***
						(0.24)
Investor Geographic Concentration						0.46**
						(0.20)
Constant	−2.89***	−2.85***	−2.89***	−2.95***	−3.32***	−3.68***
	(0.57)	(0.57)	(0.58)	(0.58)	(0.62)	(0.65)
McFadden's Pseudo-R^2	0.59	0.6	0.61	0.61	0.64	0.68
Observations	697	697	697	697	697	697
Log-Likelihood	−178.61	−175.32	−171.56	−170.86	−156.35	−140.87
Likelihood Ratio Test vs Previous	-	6.58**	7.52**	1.41	29.01***	30.96***
Note:	Logistic regressions. Sector dummies included. *p <0.1, **p <0.05, ***p <0.01					

Table 2: Logistic regression results

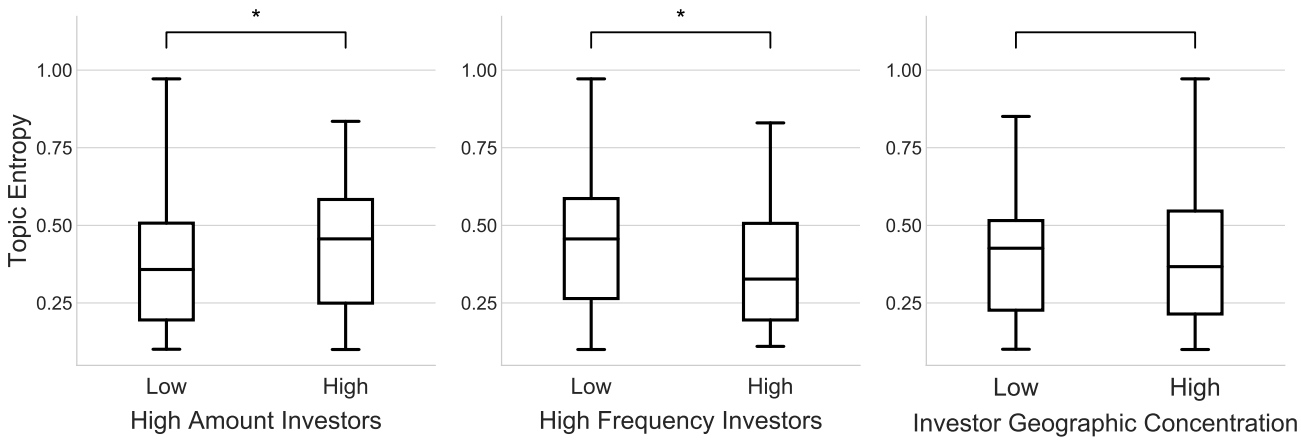

Figure 3: Differences in topic entropy for campaigns with above and below median amounts of high amount investors, high frequency investors, and investor geographic concentration. The first two comparisons yield significant differences: high topic entropy campaigns have a greater proportion of high amount investors and a lower proportion of high frequency investors.

ter. Comparing the standardized regression coefficients in Table 2, we note that variation in topic entropy induces a larger change in the chance of a campaign's success than the same variation in the campaign's target and the share of equity offered by the campaign, two key predictors of crowdfunding campaign success in the literature (Mollick 2014b).

Conclusion and Discussion

Our findings indicate that the pitch of an equity crowdfunding campaign can significantly affect fundraising success. A campaign's novelty (as measured by the topic entropy of the pitch's text) is negatively correlated with success: campaigns which are easily categorized into a few coherent topics are significantly more successful than their more textually diverse counterparts. This result holds even after controlling for writing quality and style as well as a suite of variables that have been found by previous studies to impact success. Novelty predicts success with effect size larger than these previously studied variables including share of equity offered and amount sought. It also induces a small but significant improvement in our model fit. When exploring the appeal of novelty to investors, we see that novel projects tend to draw attention from investors who are less active, but provide higher contributions. Our findings relating novelty to success are robust to several topic model specifications and a different measure of campaign dispersion across topics. In contrast, simpler measures of campaign novelty such as counts of stemmed tokens like "new*", "novel*", or "innovat*" are not significantly related with fundraising success.

Our work has several implications that include theoretical considerations about equity crowdfunding and design recommendations for crowdfunders and platforms.

Theoretical Considerations This paper makes three theoretical contributions. First, it is the first empirical study to examine the relationship between venture novelty and funding success in the context of equity crowdfunding. Using one potential framing of novelty, it adds new evidence to a long debate in the literature. In agreement with the innovation maximization fallacy, we find that high levels of novelty are unfavorable. As opposed to existing models (Zuckerman 2016) as well as empirical results from science and popular culture (Uzzi et al. 2013; Boudreau et al. 2016; Askin and Mauskapf 2017), however, we do not find an inverted U-shaped relationship between novelty and success. Second, it provides strong evidence that equity crowdfunding success not only depends on entrepreneurs' venture pitch, but also on the geographical concentration of investors and the campaign's ability to attract high frequency investors with small to medium contributions. With this, we contribute to a growing body of literature on investor dynamics and the collective decision-making aspect behind crowdfunding (Zhang and Liu 2012). Third, this study describes an easy and scalable way of measuring venture novelty by considering the mix of topic information in the project narrative and effectively establishing how topically succinct a pitch is. The proposed measure represents thus an indirect quantification of novelty from textual patterns that is entangled with the perception of distinctiveness.

Design Recommendations Our study leads to several recommendations for companies thinking about launching equity crowdfunding campaigns. The first is that language and presentation matter. Quality of writing, emphasis on past accomplishments, and stylistic dimensions of the pitch signal venture quality and are related with success. We also show that good writing contributes to convincing investors.

Our results suggest a more comprehensive picture of how to better market a campaign. Entrepreneurs would profit from grounding their novel ideas in familiar settings. Though cues which distinguish their campaign from others may isolate them from direct competition, excessive distinctiveness can quickly alienate investors. Information asymmetry between entrepreneurs and investors is one of the

greatest challenges in equity crowdfunding, one that becomes more pronounced with excess novelty. Campaigns must strive to consider both expectations of familiarity and innovation.

From the perspective of crowdfunding platform managers, our findings suggest potential value for a categorization system for campaigns. Categories can help entrepreneurs clarify their position, while using information cues to convince investors of the value of their new ideas. A glossary of sector information provided by the platform could help reduce information asymmetry between entrepreneurs and investors and help novel ventures gain more credibility. Platforms could also play an information-mediating role to help novel ventures attract more distant investors.

Finally, our study has important limitations as well. To begin with, the results pertain to the users of one equity crowdfunding platform. The novelty-preferences of this particular set of users remain inherently elusive. Moreover, novelty is a multidimensional concept which is difficult to validate. The subjective nature of what is new and fresh is no doubt one reason why the concept is so difficult to quantify. We suggest that more work is needed to understand these dimensions. One potential direction which we believe is especially promising is to map networks of investors and entrepreneurs. We also see further potential for the analysis of campaign texts, including messages from entrepreneurs to prospective investors. As crowdfunding continues to grow and the field generates more data, a better understanding of the important phenomena of novelty and innovativeness is within reach.

Acknowledgments

The authors are grateful to the company for sharing their unique data set. The Data Science team has provided indispensable information at various stages of this work. The authors thank Martina Iori for helpful conversations. This work was funded in part by the U.S. National Science Foundation (IIS-1755873).

References

Adjei, M. T.; Noble, S. M.; and Noble, C. H. 2010. The influence of c2c communications in online brand communities on customer purchase behavior. *Journal of the Academy of Marketing Science* 38(5):634–653.

Agrawal, A. K.; Catalini, C.; and Goldfarb, A. 2011. The geography of crowdfunding. Working Paper 16820, National Bureau of Economic Research.

Ahlers, G. K.; Cumming, D.; Günther, C.; and Schweizer, D. 2015. Signaling in equity crowdfunding. *Entrepreneurship Theory and Practice* 39(4):955–980.

Allison, T. H.; Davis, B. C.; Short, J. C.; and Webb, J. W. 2015. Crowdfunding in a prosocial microlending environment: Examining the role of intrinsic versus extrinsic cues. *Entrepreneurship Theory and Practice* 39(1):53–73.

An, J.; Quercia, D.; and Crowcroft, J. 2014. Recommending investors for crowdfunding projects. In *Proceedings of the 23rd ACM International Conference on World Wide Web*, 261–270.

Askin, N., and Mauskapf, M. 2017. What makes popular culture popular? Product features and optimal differentiation in music. *American Sociological Review* 82(5):910–944.

Belleflamme, P.; Lambert, T.; and Schwienbacher, A. 2014. Crowdfunding: Tapping the right crowd. *Journal of business venturing* 29(5):585–609.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3(Jan):993–1022.

Boudreau, K. J.; Guinan, E. C.; Lakhani, K. R.; and Riedl, C. 2016. Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Science* 62(10):2765–2783.

Burgoon, M.; Denning, V.; and Roberts, L. 2002. Language expectancy theory. In *The persuasion handbook: Developments in theory and practice*. 117–136.

Ceyhan, S.; Shi, X.; and Leskovec, J. 2011. Dynamics of bidding in a P2P lending service: Effects of herding and predicting loan success. In *Proceedings of the 20th ACM International Conference on World Wide Web*, 547–556.

Chan, C. S. R., and Parhankangas, A. 2017. Crowdfunding innovative ideas: How incremental and radical innovativeness influence funding outcomes. *Entrepreneurship Theory and Practice* 41(2):237–263.

Cholakova, M., and Clarysse, B. 2015. Does the possibility to make equity investments in crowdfunding projects crowd out reward-based investments? *Entrepreneurship Theory and Practice* 39(1):145–172.

Colombo, M. G.; Franzoni, C.; and Rossi-Lamastra, C. 2015. Internal social capital and the attraction of early contributions in crowdfunding. *Entrepreneurship Theory and Practice* 39(1):75–100.

Cunningham, C. 2017. When does novelty pay? In *Academy of Management Proceedings*, volume 2017, 16957.

Davidsson, P., and Wiklund, J. 2001. Levels of analysis in entrepreneurship research: Current research practice and suggestions for the future. *Entrepreneurship Theory and Practice* 25(4):81–100.

Drover, W.; Busenitz, L.; Matusik, S.; Townsend, D.; Anglin, A.; and Dushnitsky, G. 2017. A review and road map of entrepreneurial equity financing research: Venture capital, corporate venture capital, angel investment, crowdfunding, and accelerators. *Journal of Management* 43(6):1820–1853.

Dutta, S., and Folta, T. B. 2016. A comparison of the effect of angels and venture capitalists on innovation and value creation. *Journal of Business Venturing* 31(1):39–54.

Fleming, L. 2001. Recombinant uncertainty in technological search. *Management Science* 47(1):117–132.

Fradkin, A.; Grewal, E.; Holtz, D.; and Pearson, M. 2015. Bias and reciprocity in online reviews: Evidence from field experiments on Airbnb. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, 641–641.

Hall, D.; Jurafsky, D.; and Manning, C. D. 2008. Studying the history of ideas using topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 363–371.

Hannák, A.; Wagner, C.; Garcia, D.; Mislove, A.; Strohmaier, M.; and Wilson, C. 2017. Bias in online freelance marketplaces: Evidence from taskrabbit and fiverr. In *CSCW*, 1914–1933.

Hastie, T., and Tibshirani, R. 1990. *Generalized additive models*. Wiley Online Library.

Horvát, E.-Á., and Papamarkou, T. 2017. Gender differences in equity crowdfunding. In *Proceedings of the Fifth AAAI Conference on Human Computation and Crowdsourcing*, 51–60.

Howe, J. 2008. *Crowdsourcing: How the power of the crowd is driving the future of business*. Random House.

Hyytinen, A.; Pajarinen, M.; and Rouvinen, P. 2015. Does innovativeness reduce startup survival rates? *Journal of Business Venturing* 30(4):564–581.

Jennings, J. E.; Jennings, P. D.; and Greenwood, R. 2009. Novelty and new firm performance: The case of employment systems in knowledge-intensive service organizations. *Journal of Business Venturing* 24(4):338–359.

Lin, M., and Viswanathan, S. 2016. Home bias in online investments: An empirical study of an online crowdfunding market. *Management Science* 62(5):1393–1414.

Lukkarinen, A.; Teich, J.; Wallenius, H.; and Wallenius, J. 2016. Success drivers of online equity crowdfunding campaigns. *Decision Support Systems* 87:26–38.

Marom, D., and Sade, O. 2013. Are the life and death of an early stage venture indeed in the power of the tongue? Lessons from online crowdfunding pitches.

Meer, J. 2014. Effects of the price of charitable giving: Evidence from an online crowdfunding platform. *Journal of Economic Behavior and Organization* 103(C):113–124.

Mitra, T., and Gilbert, E. 2014. The language that gets people to give: Phrases that predict success on Kickstarter. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 49–61.

Mohammadi, A., and Shafi, K. 2018. Gender differences in the contribution patterns of equity-crowdfunding investors. *Small Business Economics* 50(2):275–287.

Mollick, E. 2014a. *The Danger of Crowding Out the Crowd in Equity Crowdfunding*. Penn Wharton Public Policy Initiative. Book 29.

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

Munger, K.; Bonneau, R.; Nagler, J.; and Tucker, J. A. 2018. Elites tweet to get feet off the streets: Measuring regime social media strategies during protest. *Political Science Research and Methods* 1–20.

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

Pennebaker, J. W.; Mehl, M. R.; and Niederhoffer, K. G. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology* 54(1):547–577.

Redi, M.; O'Hare, N.; Schifanella, R.; Trevisiol, M.; and Jaimes, A. 2014. 6 seconds of sound and vision: Creativity in micro-videos. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 4272–4279.

Riedl, J. 2013. Crowdfunding technology innovation. *IEEE Computer* 46:100–103.

Tan, J.; Shao, Y.; and Li, W. 2013. To be different, or to be the same? An exploratory study of isomorphism in the cluster. *Journal of Business Venturing* 28(1):83–97. Special Issue: Institutions, Entrepreneurs, Community.

Tausczik, Y. R., and Pennebaker, J. W. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29(1):24–54.

Thebault-Spieker, J.; Terveen, L.; and Hecht, B. 2017. Toward a geographic understanding of the sharing economy: Systemic biases in UberX and TaskRabbit. *ACM Transactions on Computer-Human Interaction (TOCHI)* 24(3):21.

Uzzi, B.; Mukherjee, S.; Stringer, M.; and Jones, B. 2013. Atypical combinations and scientific impact. *Science* 342(6157):468–472.

Vulkan, N.; Astebro, T. B.; and Sierra, M. F. 2016. Equity crowdfunding: A new phenomena. *Journal of Business Venturing Insights* 5:37–49.

Wachs, J.; Hannák, A.; Vörös, A.; and Daróczy, B. 2017. Why do men get more attention? Exploring factors behind success in an online design community. In *Eleventh International AAAI Conference on Web and Social Media*.

Wachs, J.; Daróczy, B.; Hannák, A.; Páll, K.; and Riedl, C. 2018. And now for something completely different: Visual novelty in an online network of designers. In *Proceedings of the 10th ACM Conference on Web Science (WebSci)*. ACM.

Weitzman, M. 1998. Recombinant growth. *Quarterly Journal of Economics* 113(2):331–360.

Xu, A.; Yang, X.; Rao, H.; Fu, W.-T.; Huang, S.-W.; and Bailey, B. P. 2014. Show me the money! An analysis of project updates during crowdfunding campaigns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 591–600. ACM.

Xu, B.; Zheng, H.; Xu, Y.; and Wang, T. 2016. Configurational paths to sponsor satisfaction in crowdfunding. *Journal of Business Research* 69(2):915–927.

Zhang, J., and Liu, P. 2012. Rational herding in microloan markets. *Management science* 58(5):892–912.

Zuckerman, E. 2016. Optimal distinctiveness revisited. *The Oxford handbook of organizational identity* 183.