

An Analysis of Crowdfunding Data

Abstract

The development and implementation of crowdfunding as an institution is reviewed, and a large data set from one of the largest crowdfunding platforms, Kickstarter, is examined using both principal component analysis (PCA) and a projection to latent structures (PLS). From this analysis it becomes apparent that projects may be divided into “failed” “successful” and “wildly successful”, and that while a variety of factors correspond to project success or failure, wild successes are only indicated reliably by the number of comments on the project. A tendency for Kickstarter to produce a consumerist mindset is proposed to explain this, despite a conscious effort on their part to discourage that.

What is crowdfunding?

Crowdfunding is the collection of money from many contributors either in expectation of a completed product (*ex post facto*) or to assist in the achievement of a desired result (*ex ante*) [25]. It is common for the contributors to receive some token reward for their support, even in the case of *ex ante* crowdfunding, although in some cases this can be seen as the work itself if it is a public good [9][11][17][19][25]. In the *ex post facto* case, crowdfunding can be likened to a pre-order service, although this may imply a level of security that is not present. In other cases, it can be likened to gambling, loans, or charity [14][21][23][26][29]. Because of this similarity to extant financial structures, especially within investment banking, it is in danger of over or under regulation (depending who you talk to), as the government tries to figure out what to do with it. Currently, the JOBS Act contains provisions exempting crowdfunding from some regulation, and these are either overreaching or insufficient [3][7][13][16][18][20][22][32][34]. Almost everyone however, agrees that as an emerging phenomenon, crowdfunding has the potential to fundamentally change the economic and social environment of the world [4].

Influences

Crowdfunding comes from the intersection of two major social traditions. The first and most direct is “crowdsourcing”, a term coined by Jeff Howe to describe the act of letting the internet (or more generally, the public) do your work for you. This comes about either because the public is the end consumer and it is most effective to let them decide what they want, or because the scope of the project is so large that it would be impractical to do alone, or even hire an expert. Perhaps the earliest instance of crowdsourcing was the first edition of the Oxford English Dictionary, in which the British public helped record their language as it was used. The largest and most successful example however, is probably Wikipedia [10][24].

The other large influence on crowdfunding is the tradition of investment. Many ventures have a high barrier to entry. Traditionally, those attempting these ventures, if lacking the capital, would seek an investor to fund their activities, and the investor would back them after assessing the risk of the venture. However, for especially large ventures, it is desirable to spread the risk among many investors. Crowdfunding can be seen as an extension of this tradition then, with the large social networks of “Web 2.0” and hyper-connectivity of the internet in general allowing people to seek investment without regards for borders and people to back at entirely minimal risk [1][14][21][29].

Issues

As noted above, there are several legal hurdles that crowdfunding currently faces, mostly with regards to disclosure of ability and spending. These stem from the tradition of investment banking, which is strictly regulated. However, crowdfunding is different, and crafting such laws is delicate work: it is important that crowdfunding be viable as well as regulated, and its robustness, as a new phenomenon, is not yet established [3][7][13][16][18][20][22][32][34].

Several other inherent problems with crowdfunding exist. One is that as any specific crowdfunding effort becomes larger, the average level of insider knowledge of the contributors decreases, resulting in “herding behavior”, where people join in because they assume it is safe because of the large amount of support it has [8]. Also, while “the vast majority” of project founders make efforts to fulfill their obligations, 75% of them deliver later than initially estimated [27]. Possibly, this is because with the barriers to entry so low, it is common for creators to get in over their heads, especially in fields in which they are not expert.

Advantages

Despite these issues, crowdfunding has several advantages over traditional venture capitalism. If nothing else, it offers a potentially large sum of money in areas where funding is notoriously difficult, including music [1][25], journalism [9], and scientific research [15][33]. Many of these areas can be defined as public goods, those in which everyone benefits equally, whether they invested or not. The psychology of investment in these areas is an interesting topic, because any reward is indirect [9][17][19].

Because of its close ties with social networking and “Web 2.0”, a successful crowdfunding effort can also be a useful tool for raising awareness [9].

The Crowdfunding Process

Crowdfunding Sites

In theory, crowdfunding can be an organic process, and early examples often are. President Obama’s campaign fund-raising is cited as an example of this [25]. Currently however, the most common form of crowdfunding is via a crowdfunding platform or website. Project creators make a page on the website, and members can pledge different amounts of money over a set period of time. At the end of the period, the project creator either keeps it all (Keep it All, KiA) or, if a predetermined goal is not met, receives nothing (All-or-Nothing, AoN) [25].

There are many crowdfunding sites, and given the volatile environment, an exhaustive list would be quickly outdated. However, some more notable sites, with a focus on those funding scientific research, are given below.

Kickstarter

www.kickstarter.com

Notable for having the highest-funded projects of any platform. Occasionally gains notoriety by wildly popular or outlandish projects. AoN funding.

IndieGoGo

An Analysis of Crowdfunding Data

www.indiegogo.com

One of the largest crowdfunding sites, it initially focused on film. Unlike Kickstarter, it allows charitable projects and gives creators the option of AoN or KiA funding.

RocketHub

rockethub.com

Another large crowdfunding website, RocketHub is notable for its testimony before congress regarding crowdfunding. AoN funding.

SellaBand

www.sellaband.com

An early crowdfunding website that specializes in funding music. In principle, it takes the role of a record label. AoN funding.

Petridish.org

www.petridish.org

A site for funding scientific research, it focuses on projects that may be too risky to qualify for government funding and costs that are not normally within the scope of a grant. AoN funding.

Microryza

www.microryza.com

Also focuses on scientific funding, no direct rewards are given to backers. AoN funding.

IamScientist

www.iamscientist.com

Seeks to be not only a scientific crowdfunding platform, but a social network for related purposes. AoN funding.

#Scifund Challenge

www.scifundchallenge.com

A scientific research crowdfunding platform. KiA or AoN funding.

Project Creation

There is a contention that the structure and presentation of a crowdfunding project can have a significant effect on its funding. In equity crowdfunding, it was found that financial road-maps and identified risk factors are important to a successful project, while external certification of the project has little impact [2]. The social network of the project creator is an important factor in crowdfunding, and along with the underlying quality of the project is the major determinants of funding success [1] [27].

On platforms like Kickstarter, where specific pledge amounts are suggested, the anchoring effect and the paradox of choice combine to decide optimum pledge levels to set. The anchoring effect is where the first number seen becomes the context for all the other numbers (this can be seen on restaurant menus, where the first item is never the cheapest). The paradox of choice holds that given too many options, the consumer is ultimately paralyzed by choice. In this context, it means that people will not consider choices past a certain number. Thus, the optimum amount for the second-lowest reward is slightly more than \$20 [11].

Given the choice of offering either pre-orders for the product or sharing future profits as a reward

An Analysis of Crowdfunding Data

structure, entrepreneurs prefer the pre-order option for projects with low initial capital requirements, and profit sharing otherwise [5].

If the goal is awareness as much as successful funding, duration and exposure of the project during the funding period are the most important factors [9].

Demographics

The final factor to consider in a crowdfunding effort is the investors themselves. With the advent of social networking, there are fewer geographical considerations, except that closer backers probably form a core of family and friends who are more constant and early investors [1]. Males and adolescents are generally more likely to take risks than females and adults, so it has been hypothesized that they would be more likely to participate in crowdfunding. If anything however, a contrary trend exists, although it remains to be seen if this is borne out in larger samples [6].

Problem Description

A large amount of data is available from Kickstarter and other crowdfunding websites about both successful and unsuccessful crowdfunding efforts. Although these data are all readily available, little analysis of them has been done. It is desirable both to develop a model of success in these projects so that it can be replicated where possible and to develop a model of project delivery and backer satisfaction so that a potential backer can know if it is a good idea to contribute. However, only the former is likely to arise from readily available data.

Data Overview

First, the data was cleaned for duplicate and missing values, and then from the variables available, the following data were assembled for each project:

name_launch—project name concatenated with launch date. A primary ID.

url—the url of the project, kept so that outliers can be investigated in greater detail. A secondary ID.

city—the city in which the project is based. Excluded from analysis.

state—the country or US state in which the project is based. A secondary ID.

nearby—the number of other projects in the same location. Intended as an indication of KS-using population in the area. An X-variable.

parent_category—any of: Art, Comics, Dance, Design, Fashion, Film & Video, Food, Games, Music, Photography, Publishing, Technology, or Theater. A secondary ID.

category—a subdivision of parent_category. For example, “Games” is divided into “Tabletop Games” and “Video Games”. A qualitative X-variable.

launched—the date and time on which the funding period began. An X-variable.

deadline—the date and time on which the funding period ended. Excluded from analysis.

duration—the duration of the funding period in days. An X-variable.

end_day—the day of the week on which the project ended. Because Kickstarter can send reminders to potential backers 48 hours before a project ends, the end day’s position in the week could have an influence on the final “push” of funding. A qualitative X-variable.

goal—the amount sought in funding. An X-variable.

pledged—the amount raised in funding. A Y-variable.

An Analysis of Crowdfunding Data

currency—the currency used for funding. Either of: USD or GBP. A qualitative X-variable.

percent—the amount raised in funding as a percent of the amount sought. A successful project will have at least 100% funding. A Y-variable.

success—“True” if the project succeeded and “False” if it did not. A qualitative Y-variable.

state—the final project state. Any of: successful, failed, suspended, or canceled. A class ID.

backers—the final number of backers of the project. A Y-variable.

avg_pledge—the average amount pledged by each backer. Although the creator has no direct control over this, it is considered as an X-variable because it does not directly correspond to success.

updates—the number of times before the project’s end that the creator posted updates. This may indicate creator presence in the project and communication with backers. An X-variable.

comments—the number of user comments on the project. This may indicate community engagement or the size of the audience. An X-variable.

has_video—a Boolean variable corresponding to the presence of a video. An X-variable.

authored—the number of projects previously authored by the same creator. This was necessarily inferred from available data and may be incomplete. An X-variable.

has_TWT—a Boolean variable corresponding to a connected Twitter account. An X-variable.

has_FB—a Boolean variable corresponding to a connected Facebook account. An X-variable.

TWT_followers—the number of followers on Twitter. Excluded from analysis because sparse.

FB_friends—the number of friends on Facebook. Excluded from analysis because sparse.

lvls—the number of levels available to pledge at. An X-variable.

min_lvl—the minimum level available to pledge at (This is not the minimum pledge allowed, but the minimum amount with a reward attached). An X-variable.

2nd_lvl—the second level available to pledge at. An optimal value might be \$20 [11]. An X-variable.

max_lvl—the maximum level available to pledge at. An X-variable.

lvl_ratio—the ratio of the 2nd to 1st pledge levels. There may be an optimum somewhere [11]. An X-variable.

lvl_mean—the mean (average) of the pledge levels. An X-variable.

lvl_med—the median of the pledge levels. An X-variable.

lvl_repeats—a Boolean variable corresponding to if any reward level appears more than once. If it does, this may indicate rewards in limited amounts, which may in turn indicate higher-value rewards. An X-variable.

In total, there are 29700 projects in the data set, beginning in April 2012 and ending in March 2013. Because I may refer to them by name later, the top five highest-funded projects in the data set are shown in Table 1.

An Analysis of Crowdfunding Data

Table 1: Top-funded Kickstarter projects.

Name	Category	Launch	Goal	Pledged	Percent	Backers	Comments
OUYA: A New Kind of Video Game Console	Video Games	7/10/12	\$950,000	\$8,600,000	905%	63,416	22,985
The Veronica Mars Movie Project	Narrative Film	3/13/13	\$2,000,000	\$5,700,000	285%	91,584	18,170
Torment: Tides of Numenera	Video Games	3/6/13	\$900,000	\$4,200,000	465%	74,405	46,127
Project Eternity	Video Games	9/14/12	\$1,100,000	\$4,000,000	362%	73,986	59,463
Reaper Miniatures Bones: An Evolution Of Gaming Miniatures	Games	7/23/12	\$30,000	\$3,400,000	11395%	17,681	45,099

After the data are cleaned, they are moved into SIMCA, where a principal component analysis can be used to get a feel for the data and outliers can be examined individually and potentially removed. Metrics identified with project success can be identified and a partial least-squares regression used to find the most significant predictors of success.

Results

Five models were generated from the data set, each a refinement of the last. The first two models, PCA models both, served to clarify the structure of the data and to identify important variables. The second two models, PLS models both, served to explore the nature of success and to find its indicators. The fifth model, a PLS-DA model attempted to classify a project into one of four outcomes (successful, failed, canceled, or suspended) based only on its other characteristics.

An Analysis of Crowdfunding Data

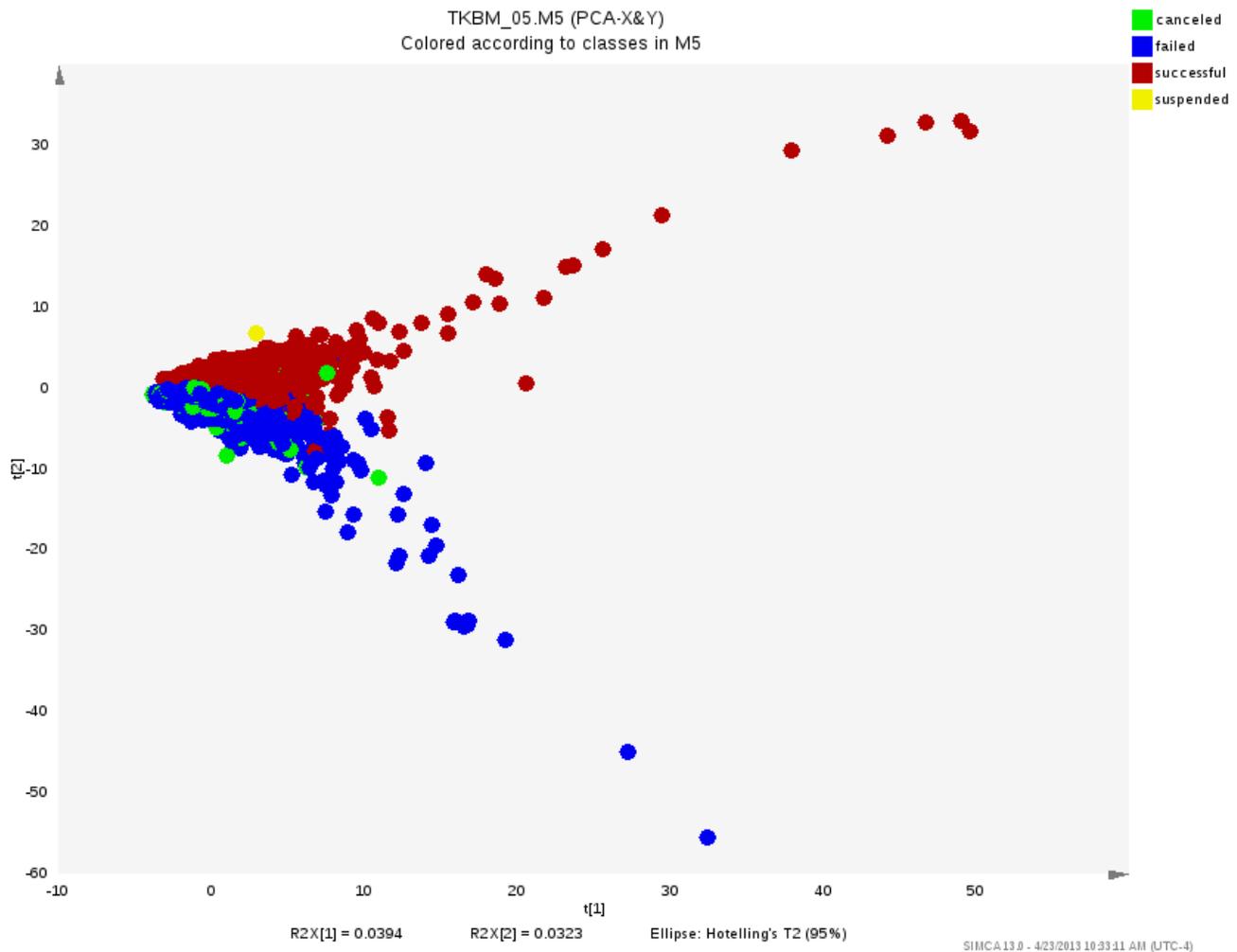


Figure 1: Model 1 score plot.

The first model immediately illustrates the structure of the data set (Figure 1). While the second component (the Y-axis) segregates projects by success, the first component spreads the projects out into two “tails”. The tail on the successful half is projects that are wildly successful, such as Ouya and the Veronica Mars movie. Likely, the corresponding tail of failed projects is an artifact: while success can have varying degrees, failure really cannot. Selective coloring reveals no obvious patterns in project location or category.

An Analysis of Crowdfunding Data

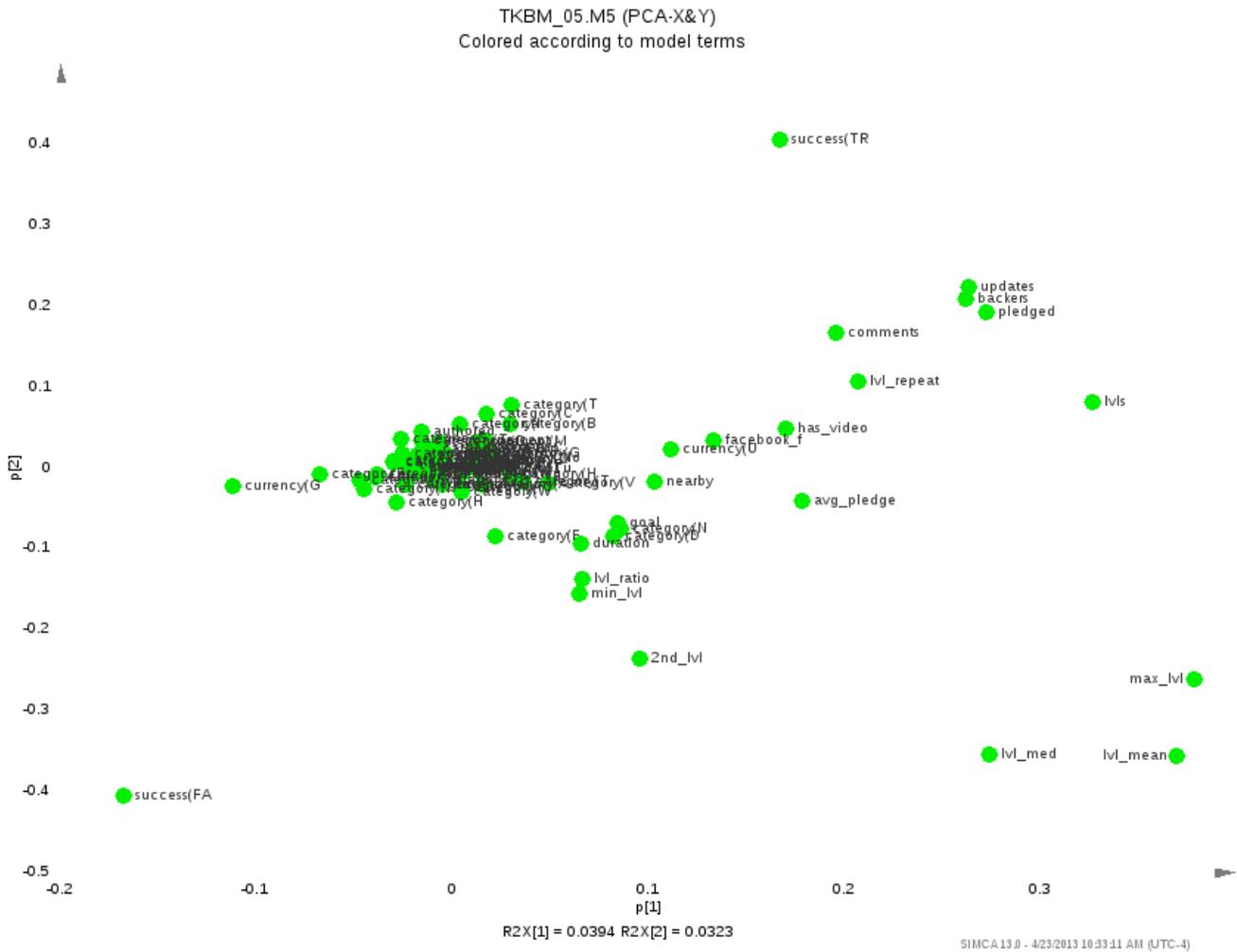


Figure 2: Model 1 loadings plot. Note the mess of variables near the origin.

The first model also shows that some variables don't matter as much. The number of canceled and suspended projects is low enough that these classes can be safely ignored in favor of a binary "True" or "False". Additionally, a large cluster of variables near the origin of the loadings plot (Figure 4) can be safely ignored. Most of these loadings are from two qualitative variables: project category and deadline day of the week. It is possible that SIMCA does not process qualitative variables in a way that is as meaningful as quantitative variables, and that this explains their apparent irrelevance. Variables related to the size, although not the presence, of social networks were excluded because of their sparse information. Finally, the date/time variable "launched" was included as an X-variable. The resulting model R2 improved from worse than less than 0.1 to better than 0.2 for two components each, but did not have significantly different structure.

An Analysis of Crowdfunding Data

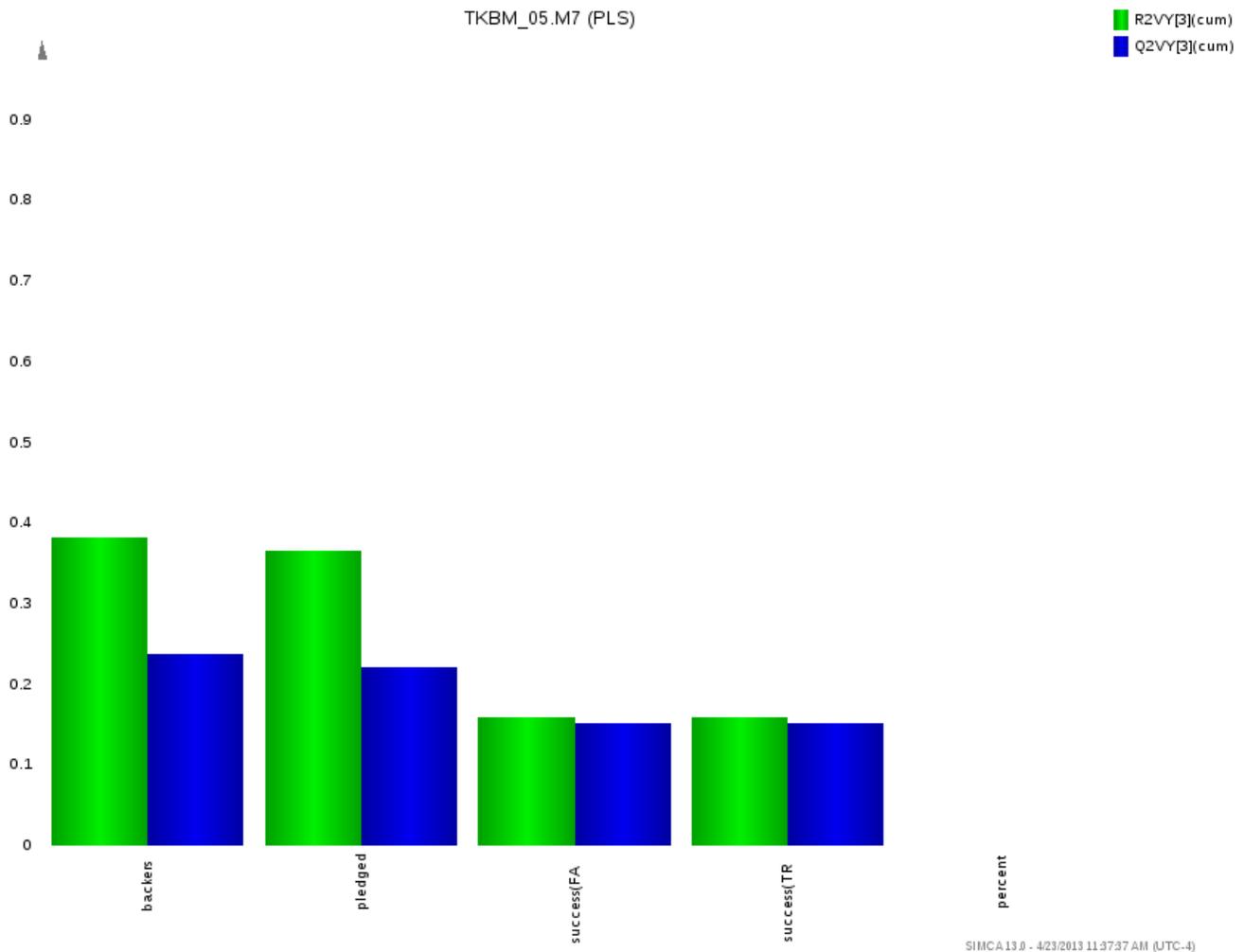


Figure 3: Model 3 X/Y overview plot.

The third model was a PLS model based on the second model, using amount pledged, percent of goal pledged, success (a boolean value), and backers as Y-variables. It is immediately apparent that the percent of goal pledged is entirely not predictable (Figure 2). Furthermore the tail of failed projects has collapsed into normal projects, while the tail of successful projects has become exaggerated (Figure 1). By examining the contributions to the tail, as opposed to the normal projects, it is seen that number of comments is by far the greatest predictor of this “runaway” success, while number of updates follows far behind, and no other factors contribute significantly.

An Analysis of Crowdfunding Data

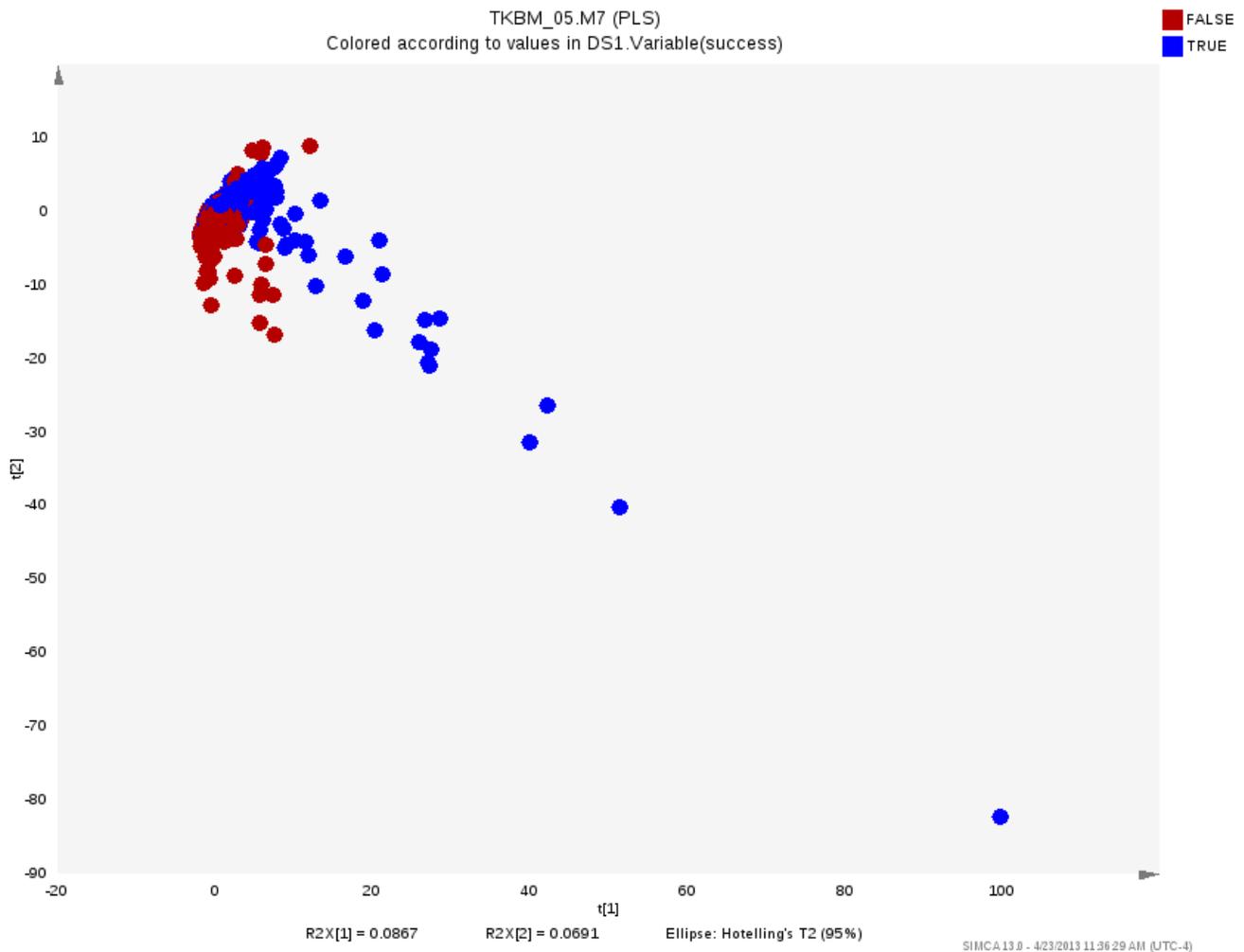


Figure 4: Model 3 score plot.

A fourth model is substantially similar to the third, but excluding the percent variable. The immediate effect of this was to flip the orientation of second component, but little else. An examination of the loadings plot of this model reveals a telling underlying structure (Figure 5). It can be seen that the Boolean succeed or fail, and most of the other factors lie roughly in a line, while pledged and backers lie nearly on top of one another, forming an orthogonal line with comments. At one end and almost all other factors at the other.

The fifth model, a PLS-DA model, does not immediately reveal anything notable about the data.

An Analysis of Crowdfunding Data

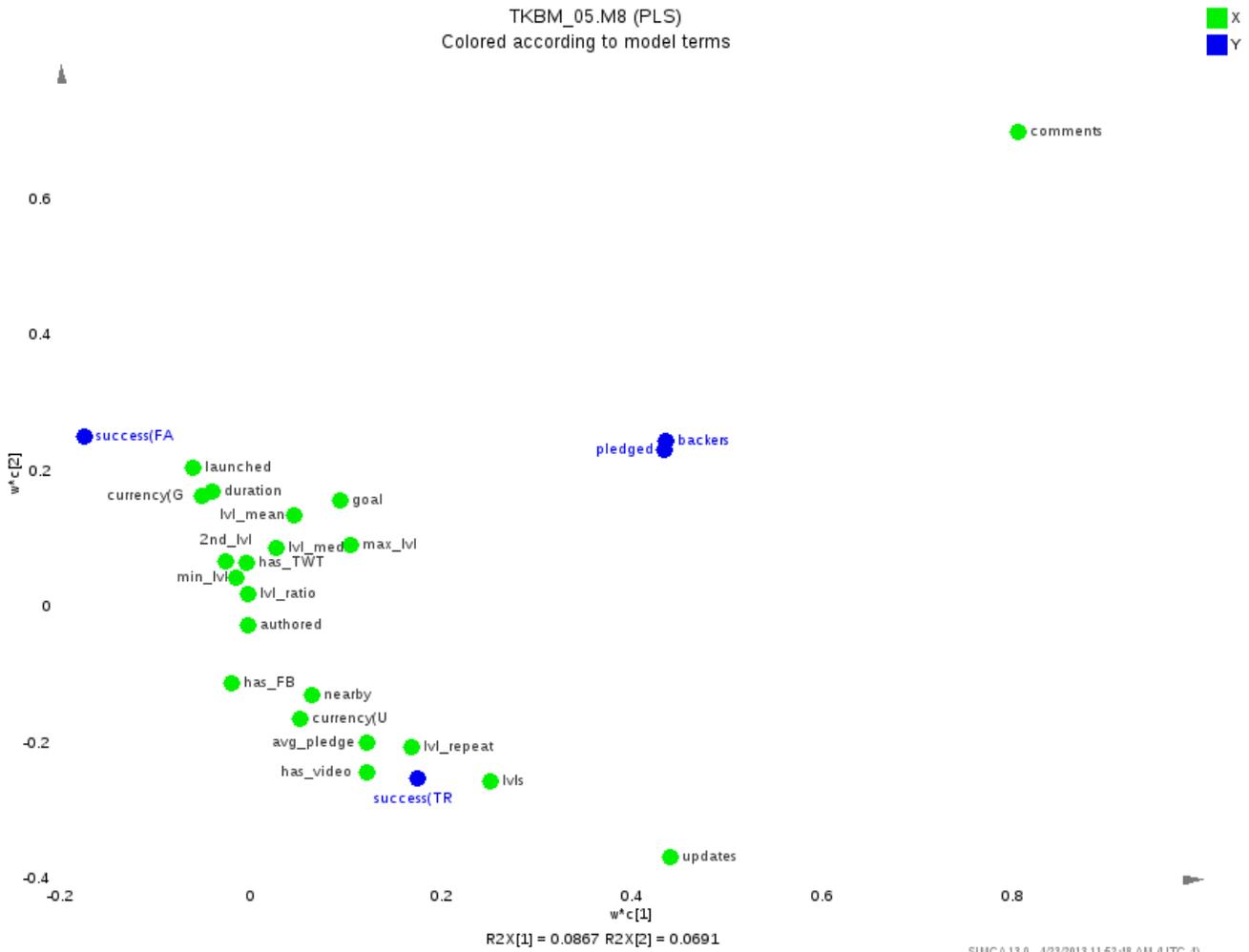


Figure 5: Model 4 loadings plot. Note the Y-axis is reversed from Figure 4.

Conclusions

The irrelevance of percent is telling, as it indicates that the goal of a crowdfunding project does not significantly alter the amount given. Instead, because of Kickstarter’s rewards structure, the number of backers and the amount pledged depends mainly on the audience of the project, something quantified in this data set only as the number of comments on its page. However, the line from failure to success runs perpendicular to this, and a project with a smaller audience can be successfully funded by setting a lower goal or updating frequently.

This data also implies a strong tendency towards an *ex post facto* funding situation, in which backers contribute for their own perceived gain more than for the sake of the project. This is distinct from what may be expected in other crowdfunding situations, where the strength of the social network is a major determinant of project success. However, Kickstarter in particular has been attempting to distance itself from this mindset [12][31]. This may also be due in part to the herbing behavior posited by Burtch [8].

An Analysis of Crowdfunding Data

While drawing these conclusions however, one should be careful to remember that the final PLS model still only has an R^2 value slightly under 0.3, so while these correlations may hold, there is a lot of room for variance in any individual project.

Limitations

The methodology employed here is only somewhat scalable: for every observation added, the “nearby” column will need to be entirely recalculated, but the authored column only grows higher. Also, the impact of prior authorship may be unknowably limited by the extent of data, reaching back only to June 2012. For similar reasons, authorship may be inextricably correlated with the passage of time.

Additionally, this data can tell us nothing about several factors, including the quality of the project, the quality of rewards offered, the written description accompanying the project, promised “stretch goals”, the length of the video, project follow-through, backer satisfaction, and extent and timing of media coverage. Collecting information for any of these would take varying degrees of effort, from wide-ranging surveys to external tools to manual extraction.

Future Studies

There are several potential future studies that are ripe for investigation. The simplest would be to split the data into two sets: a validation and a training set. Because the data is constantly accumulating with time, it would be important that launch date be an X-variable in this model. It may be possible to interpret the entire reward structure as a spectrum, rather than through summary variables like minimum and maximum. If this is possible, the advantage would be that the whole data set would be accounted for, but potential difficulties arise from the degree to which reward levels are arbitrary and vary from project to project. Finally, an individual project can be seen as a batch, but this data set uses only a landmark-based approach, not a dynamic one. If data were to be collected over the course of each project instead of at the end, it may be possible to better model the process, and such a model would be of value to project creators as well as backers.

Given the difference posited here between Kickstarter and other crowdfunding platforms, it would be of interest to collect data from other crowdfunding platforms and compare. For example, Microryza does not offer direct rewards to backers, so it may be expected that a project’s goal closely corresponds to the amount raised.

Acknowledgments

I would like to thank Dan Misener (www.thekickbackmachine.com) for the use of his data set.

References

- [1] Agrawal, Ajay; Catalini, Christian; Goldfarb, Avi. The Geography of Crowdfunding. **2010**, NET Institute Working Paper No 10-08.
- [2] Ahlers, Gerrit K.C.; Cumming, Douglas J.; Guenther, Christina; Schweizer, Denis. Signaling in Equity Crowdfunding. **2012**, Working Paper.
- [3] Barabas, Ronald L. Crowdfunding: Trends and Developments Impacting Entertainment

An Analysis of Crowdfunding Data

- Entrepreneurs. *Ent., Arts and Sports Law J.* **2012**, 23, 38-40.
- [4] Bechter, Clemens; Jentsch, Stefan; Frey, Michael. From Wisdom to Wisdom of the Crowd and Crowdfunding. *J. of Comm. and Comp.* **2011**, forthcoming.
- [5] Belleflamme, Paul; Lambert, Thomas; Schwienbacher, Armin. Crowdfunding: Tapping the Right Crowd. **2012**, *CORE Discussion Paper* No. 2011/32.
- [6] Boman, Mimmi. The effect age and gender have on our will to crowd fund a festival: A survey study. Student thesis, Umeå University, Sweden, 2012.
- [7] Bradford, C. Steven. Crowdfunding and the Federal Securities Laws. *Columbia Bus. Law Rev.* **2012**, 1, 1-149.
- [8] Burtch, Gordon. Herding Behavior as a Network Externality. In *ICIS 2011 Proceedings*, Thirty Second International Conference on Information Systems, Shanghai, China, Dec. 6, 2011. Paper 28.
- [9] Burtch, Gordon; Ghose, Anindya; Wattal, Sunil. An Empirical Examination of the Antecedents and Consequences of Contribution Patterns in Crowd-Funded Markets. **2012**, Working Paper.
- [10] Castillo, M. The Wisdom of Crowds. *Am. J. Neuroradiol.* **2013**, ?, 1-3.
- [11] Chen, Edwin. Kickstarter Data Analysis: Success and Pricing. *Edwin Chen's Blog*, 25 Apr. 2011 [Online], <http://blog.echen.me/2011/04/25/kickstarter-data-analysis-success-and-pricing/>
- [12] Chen, Perry; Strickler, Yancey; Adler, Charles. Kickstarter is not a store. *Kickstarter Blog*, 20 Sept. 2012 [Online], <http://www.kickstarter.com/blog/kickstarter-is-not-a-store>
- [13] Cohn, Stuart R. The New Crowdfunding Registration Exemption: Good Idea, Bad Execution. *Florida Law Rev.* **2012**, forthcoming.
- [14] Dushnitsky, Gary; Klueter, Thomas. Is there an *eBay* for ideas? Insights from online knowledge marketplaces. Forthcoming in *Eur. Management Rev.* **2010**.
- [15] Editorial. Citizen Science. *Nat. Neurosci.* **2010**, 13, 1151.
- [16] Fink, Andrew. Protecting the crowd and raising capital through the JOBS act. **2012**, Working Paper.
- [17] Gambardella, Massimiliano. How to (Crowd-)Fund and Manage the (User-)Innovation: The Case of Big Buck Bunny. In *OSDOC '12*, Proceedings of the Workshop on Open Source and Design of Communication, p 51-56.
- [18] Griffin, Zachary J. Crowdfunding: Fleecing the American Masses. *Case Western Reserve J. of Law, Tech., & the Internet* **2012**, forthcoming.
- [19] Grossman, Lisa. People Power. *New Scientist*, Dec. 2012, p 22.
- [20] Hazen, Thomas Lee. Crowdfunding or Fraudfunding? Social Networks and the Securities Laws —Why the Specially Tailored Exemption must be Conditioned on Meaningful Disclosure. *N. Carolina Law Rev.* **2012**, forthcoming.
- [21] Heminway, Joan MacLeod. What is security in the crowdfunding era? *Ohio St. Entrepren. Bus. L. J.* **2012**, 7, 335.

An Analysis of Crowdfunding Data

- [22] Heminway, Joan MacLeod; Hoffman, Sheldon Ryan. Proceed with your Peril: Crowdfunding and the Securities Act of 1933. *Tennessee Law Rev.* **2011**, 78, 879.
- [23] Hildebrand, Thomas; Puri, Manju; Rocholl, Jörg. Skin in the Game: Incentives in Crowdfunding. **2010**, Working Paper.
- [24] Howe, Jeff. The Rise of Crowdsourcing. *Wired*, Jun. 2006 [**Online**], <http://www.wired.com/wired/archive/14.06/crowds.html>
- [25] Kappel, Tim. Ex Ante Crowdfunding and the Recording Industry: A Model for the U.S. *Loy. L. A. Ent. L. Rev.* **2009**, 29, 375-385.
- [26] Lehner, Dr. Othmar M. Crowdfunding Social Ventures: A Model and Research Agenda. *Routledge Venture Capital J.* **2012**, forthcoming.
- [27] Mollick, Ethan. The Dynamics of Crowdfunding: Determinants of Success and Failure. **2012**, Working Paper.
- [28] Ordanini, Andrea; Miceli, Lucia; Pizzetti, Marta; Parasraman, A.; Crowdfunding: Transforming Customers into Investors through Innovative Service Platforms. *J. of Service & Management* **2011**, 22, 443-470.
- [29] Rubinton, Brian J. Crowdfunding: Disintermediated Investment Banking. **2011**, Working Paper.
- [30] Schwienbacher, Armin; Larralde Benjamin. Crowdfunding of small entrepreneurial ventures. In *Handbook of Entrepreneurial Finance*; forthcoming. Oxford University Press.
- [31] Strickler, Yancey. Is lateness failure? *Kickstarter Blog*, 17 Dec. 2012 [**Online**], <http://www.kickstarter.com/blog/is-lateness-failure>
- [32] Verstein, Andrew. The Misregulation of Person-to-Person Lending. *The UC Davis Law Rev.* **2011**, 45, 445-530.
- [33] Wheat, Rachel E.; Wang, Yiwei; Byrnes, Jarrett E.; Ranganathan, Jai. Raising money for scientific research through crowdfunding. *Trends in Ecology & Evolution* **2013**, 28, 71-72.
- [34] Wroldsen, John S. The Social Network and the Crowdfund Act: Zuckerberg, Saverin, and the Venture Capitalists' Dilution of the Crowd. *Vand. J. Ent. & Tech. L.* **2013**, forthcoming.