Random Acts of Pizza: Success Factors of Online Requests

Tim Althoff and **Niloufar Salehi** and **Tuan Nguyen**
{althoff, niloufar, tylernht}@stanford.edu
Project Group Number: 25

Abstract

Many online communities such as Q&A sites, crowd-funding platforms, and online charities are based on users requesting from one another and receiving responses from within the community; the success of these requests is critical to the success of the community. In this paper, we explore the various factors that influence success. One of the most interesting aspects in this regard is how the formulation of the requests affects it's chances of succeeding. We argue that previous attempts at unraveling factors of success were complicated by their diverse nature. We introduce a dataset of several thousand of requests over the course of more than two years where every requests asks for the very same, a pizza. This allows us to analyze the question of *how* successful request were made, in a way that has not been approached before. Our findings include that putting in some effort, creating a sense of trust, and the constructing the right kind of narrative are significantly correlated with success.

Introduction

We live in a time where we increasingly turn to the web for help. However, our needs often go far beyond what a public domain can offer, we need help from real people. A large number of platforms allow users to post requests through which they can ask others for help. Understanding the dynamics of successful requests is critical in many domains such as crowdfunding projects for startups on crowdfunding platforms like Kickstarter [4], asking for answers to specific questions on Quora [7] or StackOverflow [9], disadvantaged people asking for loans on social peer to peer lending sites such as Kiva [5] or donations to non-profits and charities such as GlobalGiving [3] or Donors Choose [2], and people facing hardship and asking for help on online communities such as Reddit [8]. Not all of these requests are successful however, which raises the question of what differentiates the fortunate. Prior research suggests aspects such as what is asked for, how it is asked, and who is asking whom as largely influential in the process.

We explore these factors by simplifying some of the complexities faced in previous work. In this paper we present a novel case study of factors of online requests that naturally controls for the subject of the request (try asking the internet for a Ferrari!). Further, we eliminate group dynamics by looking at requests that a single person could satisfy. Moreover, knowledge gathering platforms discourage duplicate requests (questions) and thereby introducing complex biases (particularly, when one is trying to control for content to try to understand how people are asking for it). In this project, we study "Random Acts of Pizza", an online community devoted to giving away free pizzas to strangers that ask for one. We argue that this is a unique platform for our study since all requests ask for the same thing, a pizza. Compared to many previously studied settings, the structure of textual requests in this particular community is relatively simple. For instance, prior interaction between the requester and the giver before the request or during the decision making process is rare and there are no video messages or communication (as on Kickstarter) that have been shown to have a strong influence on success [28]. Furthermore, it is not discouraged to post a request with similar content to an existing request. This is not the case in the study of online popularity, Q&A sites or Kickstarter-like crowdfunding platforms.

Our contribution in this paper is three-fold: (1) We present a new dataset of online requests that all ask for the same thing and that has, to the best of our knowledge, not been subject to scientific analysis before. (2) We present an analysis of various factors of success for online requests through matching and statistical hypothesis testing. (3) We translate our findings into concrete guidelines for the requester to maximize her chances of success.

We find that several factors are significantly correlated with success of online requests: It helps to put in effort to write a longer narrative, to create a sense of trust by telling a personal story and including pictures, to signal to give back to the community in the future, and to be an active and well-regarded member of the community (see Results section for more details and a discussion)

The remainder of this paper is structured as follows. We first present related work on success of online requests. Then the problem statement and the corresponding dataset is introduced. The next section elaborates on several factors that could have an impact on success. We explain our methodology in Methods section before we summarize our findings in the Results section. Lastly, we conclude this paper with a short discussion and summary and describe avenues for future work.

Related Work

In the following, we review related work in the domain of online requests and different factors of success such as content, temporal dynamics, the narrative, user similarity and status.

The Cost Spectrum of Online Interaction

Online communities and online social networks allow users to interact with each other as well as each other's content. Most will allow you to like or up-vote posts of other people, to (re)share them, or to comment on them. These modes of interactions have been studied extensively in the context of online popularity [23, 35]. Often the goal of studying online popularity is to understand what drivers user consumption, what content to display on websites, or how to design "viral" marketing campaigns.

One dimension that has (to the best of our knowledge) not been studied explicitly yet is the cost of these interactions. Votes, re-shares and comments are usually free of charge and also cost very little time. Thus, the threshold to interact in these ways is fairly low and they happen relatively often. However, other modes of interaction have higher costs. For instance, answering somebody else's question might take considerable effort and time, and help funding projects on crowdsourcing platforms or donating to certain non-profits even comes at a financial cost. In contrast to the mentioned low cost interactions people specifically ask for a nontrivial amount of help. The act of asking for something is commonly called a request (see our problem definition below) and have adopted this terminology in this paper.

However, determining the factors that influence popularity poses similar challenges in that the content either needs to be modeled explicitely [35] or controlled for [23]. Lakkaraju et al. [23] use resubmissions of the same content (in this case a picture) to control for content and study how the title, the community, and the time matter for online popularity. When analyzing factors of success of online requests this "content" is very similar to "what is being asked for". In this paper, we attempt to understand *how* one should ask for it by controlling what is being asked for.

Temporal Dynamics of Successful Requests

Previous research has largely focused on crowdfunding platforms such as Kickstarter [4] or peer to peer lending sites such as Prosper [6] and more often studied their temporal dynamics rather than the content of the requests. For instance, Ceyhan et al. [16] find that loans with some early funding are more likely to get fully funded (coined herding effect). Similar effects were found on Kickstarter when modeling funding success as a time series prediction task that included previous donations as well as social features such as the number of tweets about the project [18]. Particularly, the last donation to fully fund a project was found to exhibit special characteristics such as being significantly larger than average [34].

Studies of Kickstarter projects [28] further revealed that the objective to be funded matters. For instance, technology and design projects are much more likely to receive funding than film and arts projects. The same study also found that the size of the requester's friends network was predictive of funding success and that including a video with your project proposal increased your chances by 26%.

How You Ask Matters

Danescu-Niculescu-Mizil et al. [17] built a computational model of politeness and investigated the relationship between politeness and the power in requests. The theory of politeness predicts that those in high power tend to be less polite in their requests or responses. This was shown empirically using data on Wikipedia admin elections: Editors that are eventually successful in the admin election tend to be more polite than their unsuccessful counterparts. Their work shows that politeness is often negatively correlated with power which suggests that people in need would have to be more polite when posting a request.

Herzenstein et al. [20] consider the significance of narratives and claimed identities (trustworthy, economic hardship, hardworking, successful, moral, and religious) for the success of obtaining loans on the peer to peer lending platform Prosper [6]. In these narratives the borrowers try to present themselves favorably to gain the trust of their lenders. The authors investigate how these narratives impact loan funding as well as loan performance (do lenders get paid back after two years?) on top of more traditional measures such as demographics and credit scores. They show that the textual part of a request, the narrative, can significantly influence the outcome of the request. However, they mainly looked at requests to support the decision making of lenders rather than teaching the borrowers how to successfully post requests ("[...] we believe lenders swayed by multiple identities are more likely to fall prey to borrowers that underperform or fail."). Furthermore, they admit that narratives are more complex than the six identified identities. Particularly emotion (sympathy, anger, etc.) and other linguistic features could inform similar models.

Who You Are Matters

Social media platforms often have a system for evaluating or up-voting other users or content they have published. These systems use user-to-user evaluation as a mechanism to identify a range of matters such as the level of authority of a certain user or how useful an answer or contribution is. Recent studies have shown that these evaluations are largely influenced by the relative status of the users ([19], [25]) and also by the similarity in their characteristics [14]. These studies suggest that who you are matters in how you and your contributions are evaluated in the community.

In their study [14], Anderson et. al. measure similarity as a combination of two factors: similarity of interests (mostly covering the types of content produced), and similarity of social ties (computing mutual connections to members of the community). They measure status as a function of the total number of actions (say number of edits on a Wikipedia article on the domain). Their results show that users with higher levels of similarity are generally more likely to evaluate the other positively. In fact, between similar users, difference in status matters less in evaluations.

Success of Online Requests

In the context of the literature review above we focus on analyzing the success of online requests.

Problem Statement

Dictionaries typically define a request as "an act of asking politely or formally for something [30]". We consider a request successful if the person acting receives what she asked for. Presumably, the success of a request depends on (1) what is being asked for, (2) how the person asks for it (the form of the request), (3) the person asking, and (4) who this request is directed at. Obviously, what you are asking for (1) is important: If you ask the internet for a $300k Ferrari your chances are smaller than if you ask for a small favor such as a retweet about your most recent article. We seek to eradicate this effect by looking at requests that are all aimed at obtaining the same thing. In particular, we are interested in the effect and interaction of how one should ask for something (2) and how this depends on the person asking (3) and the person or group the request is directed at (4).

Note that this problem statement differs from understanding content popularity in that satisfying a request usually has a higher cost than e.g. clicking on an article or upvoting a post. It further differs from crowdfunding in that a request can often be satisfied by a single person as opposed to a group of people. In that regard analyzing success of requests bears similarities with studies of question and answer sites. However, these typically feature a wide range of questions (requests) and therefore do not naturally control for what is being asked for (1). Furthermore, they often employ mechanisms that discourage asking the same question multiple times which complicates the analysis. If how people are asking for it is to be studied then having multiple requests with the the same is actually advantageous.

Dataset

Introduction to Reddit Reddit [8] is a social news and entertainment website where registered users submit content in the form of either a link or a text post. Other users then vote the submission "up" or "down", which is used to rank the post and determine its position on the site's pages and front page. Reddit users (called "Redditors") can also post comments about the submission, and respond back and forth in a conversation tree of comments; the comments themselves can also be up-voted and down-voted.

Reddit encourages links over text submissions, by allowing redditors to accumulate points ("karma") for highly rated links, but not for text-posts. Redditors also accumulate karma for highly rated comments, on posts of both kinds [1].

All content entries are organized by areas of interest called "subreddits". In the following we introduce one particular subreddit that has recently gained much attention and allows us to study the success of online requests in a very natural setting.

Random Acts of Pizza Random Acts of Pizza [12] is a community on the website Reddit.com that facilitates the sending and receiving of pizzas between strangers, "because... who doesn't like helping out a stranger? The purpose is to have fun, eat pizza and help each other out. Together, we aim to restore faith in humanity, one slice at a time [11]."

Random Acts of Pizza gives people the opportunity to help those in need of a delicious dinner, or to be the recipient of one. Redditors can submit requests for free pizza, and if their story is compelling enough a fellow "RAOPer" might decide to send them one. One of the Subreddit's founders, Daniel Rogers of San Antonio, Texas, said he got the idea during a period of unemployment, which led him to feel "disconnected from society." One ABC News article describes the stories as "moving, even heartbreaking [15]." A typical post might sound something like this, from Gabriel in New York: "It's been a long time since my mother and I have had proper food. I've been struggling to find any kind of work so I can supplement my mom's social security... A real pizza would certainly lift our spirits [15]."

This subreddit community provides an almost ideal environment to study the success of online request since every participant asks for the very same, a pizza. Therefore, these requests provide a well controlled dataset in which only the form of the request or the person requesting have an effect on the success of the request. Related work has often focussed on question & answer sites such as StackOverflow [13, 14]. However, the range or complexity of what people ask for varies a lot on such sites. Others focused on Wikipedia and the admin application and voting process [14, 17]. While in this case all applications seek to reach the same status of an admin the process is more complicated as your application essentially involves your lifetime achievements as a Wikipedia editor. Furthermore, this request can only be satisfied by a group of people (current admins voting on your case). This allows for interesting analyses of collective behavior and mechanisms but potentially complicates effects on the success of the request itself.

The Dataset In accordance with subreddit moderators and reddit employees we crawled the entire history of the Random Acts of Pizza subreddit [12]. The first post in this community dates back to December 8, 2010. For more information on the dataset please refer to Table 1. Most users in this community post or comment very few times while a small number of them contribute dozens of posts and thousands of comments (see Figure 1 and 2). To compute user and community features (see Section) we further crawled the entire lifetime history of posts and comments across all subreddits for all users involved in the Random Acts of Pizza subreddit.

We split the annotated pizza requests into development (4077 observations) and test set (1651 observations) such that both sets mirror the average success rate in our dataset of about 25%. All features were developed on the development test only while the test set was used only once to evaluate the prediction accuracy of our proposed model.

Potential Factors of Success in Online Requests

The main question we seek to answer in this project is what makes online requests successful, i.e. identifying characteristics that distinguish successful requests from unsuccessful ones. It is an unsolved question to what degree it matters

total number of submissions	1,870,902
total number of posts	231,869
total number of comments	1,639,033
total number of votes	43,268,390
total number of subreddits	13,723
time span RAOP	Dec 2010 – Sep 2013
posts in RAOP	21,577
users that posted or commented in RAOP	11,551
users that have received a pizza	3,439
users that have given a pizza	1,657
total number of requests success annotation	5,728
identifiably successful requests	1,410
identifiably successful requests for which we know the giver	379
identifiably unsuccessful requests	4,318
total size of community network	5,743

Table 1: Basic statistics of the Reddit and RAOP dataset

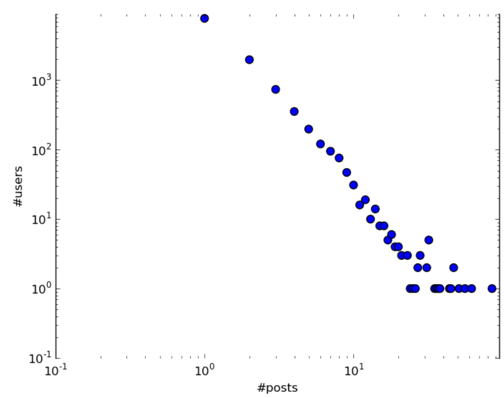

Figure 1: Log-log plot of number of RAOP users with a specific number of posts.

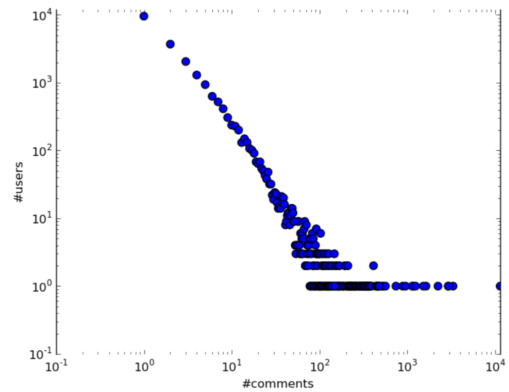

Figure 2: Log-log plot of number of RAOP users with a specific number of comments.

who the poster is (in addition to how she asks for it). Given that we control for what is being requested (pizza) we differentiate between factors depending on the (1) textual request and on the (2) user in the community to shine some light on this question.

Basic Factors

It is possible that we have temporal and seasonal effects in the data, i.e. that requests on specific months, weekdays, or hour of the day are more successful on average than others. We extract these three quantities and further the month of the request since the beginning of the community (this can be thought of the interaction between year and month or simply the age of the community). We convert the hour of the day to Eastern time (EST) for all requests (unfortunately, the time zone information on Reddit is seems broken so that we could not use time zones as a proxy for location). Note that the RAOP community consists mostly of individuals in the United States such that the time is off at most three hours for users from the west coast. We have the hypothesis that the day within the month matters as well as many people receive their paychecks in the end or beginning of the month. Because of this the financial situation might get tighter until the end of the months so that more people should be in need of a pizza at the end of the month.

We used the Stanford Named Entity Recognizer to extract locations from the request as some times the requester states her location. However, many requesters do not publish their location and the coverage was too small to allow for significant results so we dropped this feature from our analysis.

Textual Factors

Length The length of the request (measure in number of words) is a measure of how much effort the requester put into the request and likely to be correlated with the success of the request. In contrast, the length of the request title was not significant and was therefore omitted from the following analysis.

Sentiment It is possible that reddit users are more likely to give a pizza if the request is very positive or very negative, i.e. if it has a clear sentiment. We use two different sentiment classifiers [29, 32] to annotate request as positive, neutral, or negative. The language on Reddit differs from more traditional and edited sources such as professional magazines and is probably much closer to shortenend language such as Twitter. In these communities, emoticons are often used to convey emotion or sentiment. Therefore we also extract emotions from the text using regular expressions [31].

Politeness Previous research suggests that users that ask more politely might be more successful [17]. We measure politeness as the fraction of words that are “please”, “thanks”, “thank you”, or hedges. Hedges are “words whose job is to make things fuzzier or less fuzzy”, i.e. the exact meaning of of some qualities or quantities is blurred by them

[24]. Danescu-Niculesu-Mizil et al. found hedges to be an important linguistic concept to be in politeness (see Table 3 in [17]). We use a lexicon of hedges defined in [17] for our analysis.

Content What do successful requests talk about? The ways that individuals talk and write provide windows into their emotional and cognitive worlds. Previous research suggests that people's physical and mental health can be predicted by the words they use which lead to the definition of the Linguistic Inquiry and Word Count (LIWC) corpus [33]. This corpus enables us to study the effect of function words (pronouns, articles, etc.), social processes (family, friends), affective processes (positive or negative emotion), as well as personal concerns (work, leisure, money, death, etc.). We use counting statistics of words associated with these categories as features. For example, the category for friends includes words like friend, girlfriend, boyfriend, roommate, fiance, fellow, neighbor, buddy, partner, lover, etc.

Language Complexity The readability of a request could have an impact on the reader's perception of the educational, cultural, and socioeconomic background of the requester. We extract the Flesch-Kincaid Grade Level [21] which is frequently used to judge the readability level of various books and texts (based on number of words per sentence and number of syllables per word). For example, a score of 8.2 would indicate that the text is expected to be understandable by an average 8th grade student in the United States.

We further use a spell-checker to identify misspelled words in the request text [10]. In other (albeit more professional) contexts (e.g. Kickstarter), spelling errors have been found to have a negative impact on funding success [28].

Trust Some users specifically state their ability and willingness to prove that their story or claims are valid. Often the proof is a photo which shows their empty bank accounts, empty fridges, location, or identity. We count the occurrences of http links, image links, and "proof"/"prove". Other users signal their ambition to help other people to a pizza after they do better (often after their next paycheck). In RAOP this is referred to as "pay it forward" which we also extract from the text.

Novelty and Humor We decided not to use these attributes since they did not seem significant from manual inspection of the development set and very hard to classify automatically.

User and Community Factors

It is reasonable to assume that not only what you say and how you say it will matter in getting you a pizza. In other words, before paying the price of a pizza for somebody they don't know, potential givers will look into that persons profile and their past activities. We hypothesize that the user community around RAOP, and where the requester stands in this community will have an effect on their chances of success. In this regard, we studied this community from a number of perspectives and tried to answer questions such as:

- Will activity on Reddit/ROAP significantly effect success?
- Who is giving pizza's? Who is receiving them?
- Will your performance in other parts of Reddit effect your chances of success on RAOP?
- Will user similarity effect the chances of giving/receiving pizza?

In order to answer these questions, we have extracted a number of community features from Reddit and RAOP. Note that we compute each of these features with conscious of time. For instance, we only take posts/activity into account that happened before the particular request was posted. Since reddit does only offer aggregate statistics "of the present" we crawled full user histories to be able to recompute these measures at arbitrary points in time.

Age This factor looks into the relevancy of how long a person has been in the Reddit community and their success in getting a pizza. In this regard we look at both the persons "age" in all of Reddit, as well as in the ROAP subreddit. Here age is computed as the amount of time passed since their first activity.

Activity Do people care about how active you are on Reddit? This is a very important factor as it hints to two key concepts. The first concept is trust, if a user has posted frequently, other users have a sense of familiarity with that person, they feel this person is less likely to be a fraud, and thus trust their story better. The second concept is that of encouraging contribution. Within many communities members are likely to reward those who have made large impacts. Thus, we believe activity on Reddit/RAOP will increase the chances of success. In the activity category we also look at the total number of subreddits the user has posted to.

Score and Karma While activity looks into how many posts or comments you have made, this factor looks into how you were treated by the community. Accordingly we have extracted two features for each RAOP user: score and karma. On the reddit platform, users have the option to "up-vote" or "down-vote" other user's posts or comments. We compute the score of a user as the sum of all up-votes ever received minus all down-votes on content the user has published (posts or comments). Karma, is computed as the sum of all votes (up or down) ever received [14].

User Similarity Here we ask the question of whether people are more likely to offer others more simliar to themselves pizza? It intuitively seems that people would choose the former if asked directly. But the main question here is do givers care to check the requesters background and activities? And if they do, will it matter to them whether the person is active on the same subreddits that they are? To compute user similarity we have extracted all subreddits each user has participated on prior to requesting/giving pizza by posting or commenting. The similarity between two users is the number of common subreddits they have contributed to weighted by their number of posts/comments in that particular subreddit. So if two users have one subreddit in common and one has X posts/comments, while the other has Y, their similarity is increased by X+Y. We seek to learn to what extent being similar to someone will effect a users chance of receiving a pizza from them.

Subreddit Communities From our analysis we came to the conclusion that users do in fact go back to check a users activities before giving them pizza. This raises the question of whether participating in certain communities raises success rates? Is the network clustered? Are certain clusters better for getting pizza?

We further consider how helpful these user and community features are in predicting success (on top of the previously defined temporal and text-based features.

Methods

This section outlines our methodology used to analyze which factors impact the success of requests in the RAOP community.

Matching

On RAOP, successful requests tend to be longer than their unsuccessful counterparts (see Figure 3). We compute the success rate with respect to a given feature as the fraction of successful requests among all requests that share this feature. Figure 4 shows such a success rate plot for request length (error bars represent bootstrapped 95% confidence intervals [22]). The plot shows that long requests are almost twice as likely to be successful as short requests.

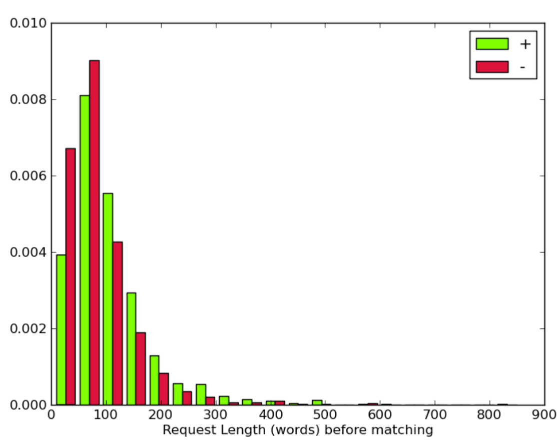

Figure 3: Request length in words before matching

We need to be very aware of this strong effect when analyzing the other textual factors. Some of the factors are almost certainly correlated with the request length. For example, the number of occurrences of function words should increase proportionally with the request length. We therefore control for request length by forming a pairs of successful and unsuccessful requests that are very similar in length. In this process, called matching, we form pairs for which the relative difference in lengths is minimal. We compute a globally optimal match through the Hungarian matching algorithm [26] and drop outliers where no reasonable match can be found. This gives us a set of 843 pairs of successful and unsuccessful requests. As Figure 5 shows, the length distributions of successful and unsuccessful requests are well matched now and has no effect on success rate any more.

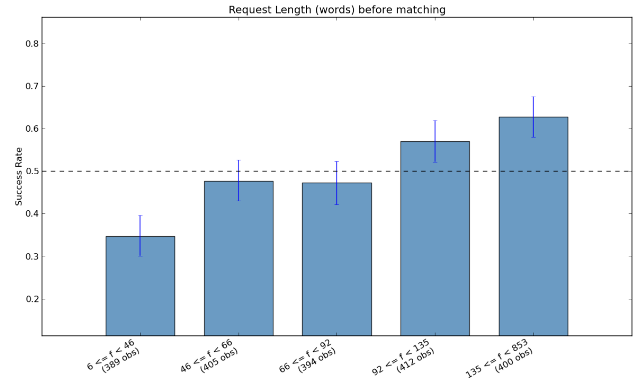

Figure 4: Success rate for different request lengths before matching

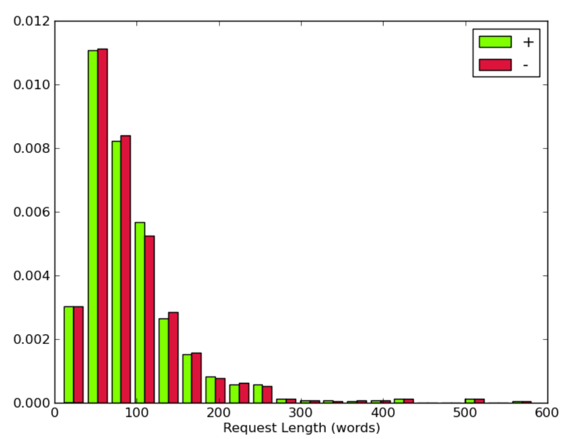

Figure 5: Request length in words after matching

Feature Selection through Significance Testing of Logistic Regression Models

The goal of statistical analysis here is to detect which factors have impact on the success rate and which do not. We collect a set of 70 features, including linguistic, LIWC (content), user-community and temporal dynamics. Then we perform hypothesis testing and evaluate testing error to give the final model. Since we try to understand the relationships of factors having on the success rate rather than to make precise prediction, we use simple logistic regression and its hypothesis testing rather than a more complex machine learning method (that might give better predictions but often fail to explain the underlying dynamics in the data well). The true model can be nonlinear or very complicated; however, as long as the logistic model has good correlation with the empirical success rate, we propose those features have significant impact on the outcome. Here, we explain how we fit and evaluate the model.

The first step of analysis is to do pre-selection on the original set of features. We fit a linear logistic regression on each individual feature along with the bar plot to give candidates that have good correlation with the success rate. Here is the result of the pre-selection process:

- Length
- Linguistic: hedges, kincaid (readability), thanks-rate, proof-rate, images, links, payitforward
- LIWC: feeling, friends, leisure, money, time, work
- Community-age
- Time: month, hour, weekday, community-age, 4h-before, first-half-month
- User: age, pizza-age, karma, score, posts count, posts-pizza count, comments count, comments-pizza count, subreddit count

Our analysis will be on several models, and then select the features from each to offer the best final model. Here is the list of proposed models:

- Random Baseline
- Length
- Length + Linguistic
- Length + LIWC
- Length + Community-age
- Length + User
- Length + Time
- Full model: include all features

Our approach is to compare all model against the base model which is length-only. This will be done via ANOVA. We can see which group of variable are significant after this is done. Hence, for each significant group of variables, we seek for significant variables within them. Then, we arrive a set of candidates for the final model. At the end, we will compare the fit of these condidates with the full model to obtain the final set of variables.

Results

This section presents the empirical results for the individual factors described in Section before we analyze the performance of logistic regression model based on all factors.

Individual Factors

All features are summarized in Table 2 and will be described in more detail in the following.

Feature	Effect	P-value
length	positive	$3.35 * 10^{-17}$
timestamp	negative	$1.25 * 10^{-4}$
day in month	negative	$3.47 * 10^{-2}$
LIWC friends	negative	$2.23 * 10^{-4}$
LIWC leisure	negative	$1.22 * 10^{-2}$
LIWC money	positive	$4.30 * 10^{-2}$
LIWC work	positive	$1.61 * 10^{-2}$
LIWC feel	positive	$3.67 * 10^{-2}$
LIWC time	positive	$3.27 * 10^{-3}$
number of images	positive	$1.06 * 10^{-5}$
pay it forward claim	positive	$1.44 * 10^{-2}$
votes on users' posts until request	positive	$1.70 * 10^{-4}$
number of posts by user until request	positive	$1.15 * 10^{-10}$
age of user in RAOP subcommunity at request	positive	$4.16 * 10^{-7}$
age of user on Reddit at request	positive	$4.35 * 10^{-3}$
number of requests in 4h window before	-	$1.49 * 10^{-1}$
fraction of positive sentiment sentences	-	$2.20 * 10^{-1}$
fraction of neutral sentiment sentences	-	$8.55 * 10^{-1}$
fraction of negative sentiment sentences	-	$4.83 * 10^{-1}$
number of emoticons	positive trend	$6.68 * 10^{-2}$
number of spelling errors	-	$6.41 * 10^{-1}$
Flesch-Kincaid Readability Score	-	$1.36 * 10^{-1}$
Number of thanks/thank you	positive trend	$7.82 * 10^{-2}$
Number of hedges	-	$7.39 * 10^{-1}$

Table 2: List of features with their effect (when increasing the feature value) on success rate as well as their statistical significance (two-tailed Wilcoxon Signed Rank Test on length-matched pairs except for the length feature itself which uses random pairs). The top part only shows features significant at a level of $p = 0.05$.

Temporal Features We find that month and weekday, while exposing some seasonal pattern, were not significant in our matched dataset. The hour of the day also showed a clear seasonal pattern where requests in the morning and around lunchtime were more often successful than requests at night. However, the effect was not significant, possibly because we are averaging over three timezones in the United States (all timestamps are reported in UTC).

Requests on RAOP became less likely to succeed over time (see Figure 6). Each bar represent 20% of requests in our matched sample in their temporal order. Requests in the first couple of months were significantly more successful than in the last couple of months. This can be explained by the increasing number of requests over the lifetime of the community (see Figure 7), i.e. there is more "competition" for a roughly constant number of pizzas given out. Possibly, the amount of abuse increased with the growing popularity of the subreddit and the publication of a questionable Wired article that encouraged people to lie to obtain free pizza [27] so that givers became less willing or more cautious but it is interesting to see that the number of pizzas given stayed roughly the same.

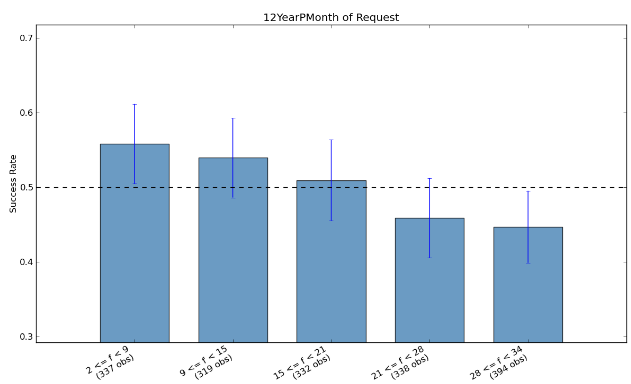

Figure 6: Success rate over time

We furthermore find a significant different between requests early in the month and requests late in the month. Figure 8 reveals that this effect is mostly due to an increasing number of requests at the end of the month (though there is a slight decrease in the number of pizzas given as well). This tells us that more reddit users request pizza at the end of the month, very possibly out of need and in tight financial situations. However, the number of pizzas given out stays roughly the same leading to a decreasing success rate. This led us to consider a feature capturing the competition between requests, the number of requests posted in a four hour window before this request. While there seemed to be a negative trend showing that more competition makes you less likely to succeed this effect was not found statistically significant.

Sentiment Our sentiment feature turned out not to besignificant in our analysis (even though the tool we used handles negation and performed well on other datasets). We then

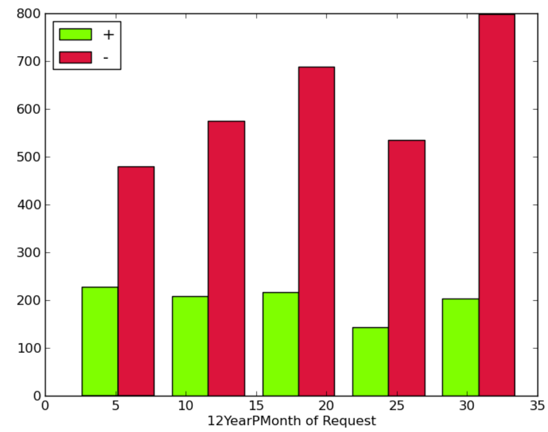

Figure 7: Number of successful (green) and unsuccessful (red) requests over time

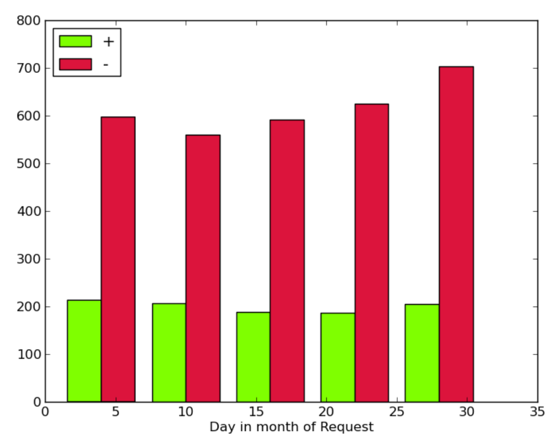

Figure 8: Number of successful (green) and unsuccessful (red) requests in the first and second half of the month.

used two other sentiment models to verify these results. We used the Stanford CoreNLP tool to annotate individual sentences with sentiment scores and measured the fraction of positive, neutral, or negative sentences a request had. We furthermore analyzed word counts for very positive and negative words using the LIWC dataset (see below in the Content paragraph). Neither of these two approaches turned out to produce significant predictors of success. However, we observed a positive trend for the number of emoticons used in the text (between 0 and 13 in our dataset). However, this trend was not significant at a 5% level using a Wilcoxon Signed Rank test ($p = 0.067$).

Politeness We found that neither the number of hedges nor the number of "please" to be significant in our analysis. Through looking at "thanks" and "thank you" we found out that some users edit their post after they received pizza to thank their generous giver. Not accounting for this accidental conditioning on the success variable this feature was very significant at $p = 10^{-5}$. After ignoring edits that noted success the feature stopped to be significant at a 5% level ($p = 0.078$).

Content From the LIWC categories we analyzed the following turned out to be significant factors to determine success. It is interesting to see that talking about friends, roommates and partners seems to hurt your chance of success (the category includes words such as friend, roommate, girlfriend, boyfriend, fiance, ...) . It also hurts to talk about Leisure (family reading, weekend, cook, apartment, celebrate, game, ...), i.e. free time and fun activities.

In contrast, it seems like a good idea to talk about Money (pay, money, paid, account, buy, rent , ban, bills, check, ...). Terms related to Work (pay, job, student, college, working, school, unemploy, financial, ...) also had a positive impact on the success rate. Two more categories, that are positively correlated with success rate, though less significantly, are Feel (feel, hard, hot, feeling, rough, tight, cool and Time (day, now, time, week, until, last, month).

Further note that categories containing positive words or negative words did not have a significant impact on success. This underlines the sentiment results above in that a strong overall positive or negative sentiment does not have a consistent effect of success.

Language Complexity We did not find a significant correlation between Flesch-Kincaid Grade Level scores and success as well as the presence or number of spelling errors and success.

Trust While the count statistic for "proof"/"prove" did not turn out to be significant but it matters a lot whether the requests contains an image or not (even after filtering out images that show the received pizza after success).

The willingness to forward the pizza to another reddit user also increased one's chances to receive pizza It is interesting to note that in our dataset only 25.0% of the the users that claimed to forward the pizza actually did after they were successful. While this might seem disappointing it is significantly larger than the fraction of users that never claimed to forward the pizza but did anyway (13.2%).

Popularity Successful requests tend to have a lot more votes (mostly up-votes) and a higher number of comments (see Figure 9 and 10). However, as we do not have the exact timestamp of when a request became fulfilled we cannot use this information that is not immediately available at the time of request.

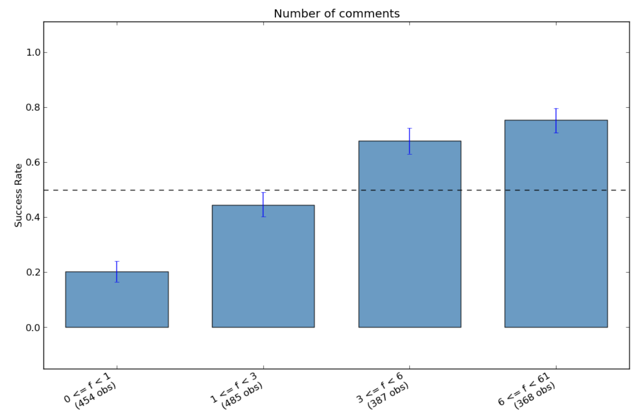

Figure 9: Success rate for different number of comments. Note that the number of comments might change after becoming successful so that this feature cannot be used for prediction.

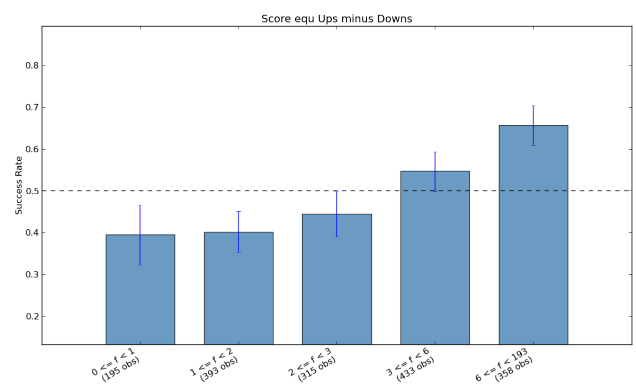

Figure 10: Success rate for different voting scores (difference between up-votes and down-votes). Note that the number of up-votes and down-votes might change after becoming successful so that this feature cannot be used for prediction.

User and Community Our results show significant effect of the community on success rates. Here we talk about a few of these factors and their effects:

Age: Our results show that age has a significant impact on rate of success. This factor has fairly the same effect if computed for all of Reddit, or on the RAOP subreddit. The longer the user has been a member of the community the higher their chance of being rewarded by a pizza.

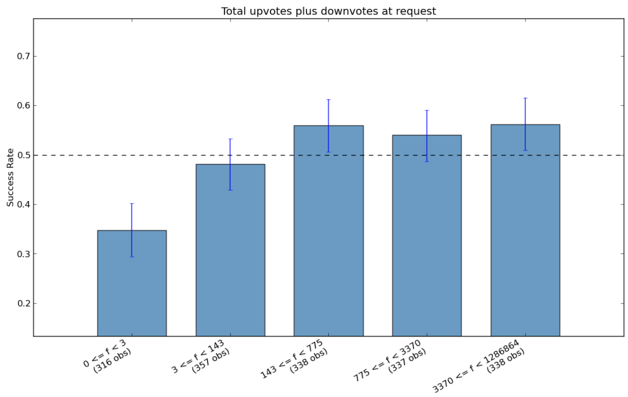

Figure 11: Success rate for different values of karma (upvotes plus downvotes)

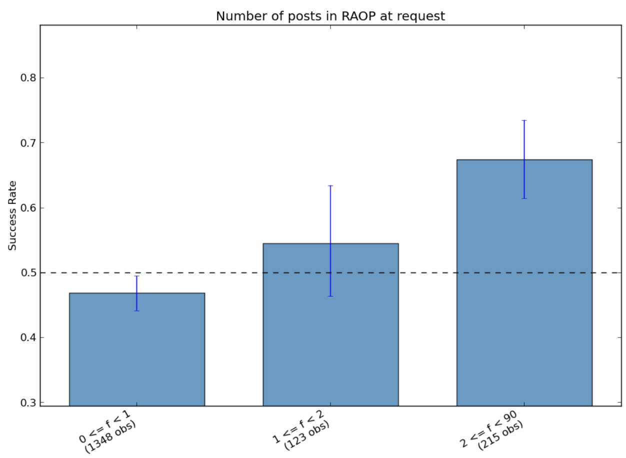

Figure 12: Success rate for different numbers of posts on RAOP before requesting

Karma: We looked into how karma (and score), effect success rate. As mentioned before, karma is a valuable way to see how the community responded to this user's activity. The plot for how karma is related to success rate is shown in Figure 11. Interestingly the plot for score (up-votes minus down-votes) follows the same general trend. This suggests that most requesters either do not accumulate a lot of downvotes or that these tend to play a minor role in the evaluation.

Activity: Figure 12 shows how the number of posts the user has had on the RAOP subreddit (at time of request) effects their chance of success. This figure shows that users who have been active on the community before requesting have a higher chance of success.

User Similarity: To compute user similarity we looked at all pairs of pizza receiver and pizza giver that we had extracted from the data (395 pairs). For each pair we computed the weighted similarity on common subreddits. We aimed to

cluster	success rate
cluster 0	0.32
cluster 1	0.28
cluster 2	0.27
cluster 3	0.26
cluster 4	0.32
cluster 5	0.26
cluster 6	0.32
cluster 7	0.29
cluster 8	0.29
cluster 9	0.27

Table 3: Success rates for different subreddit clusters.

test the hypothesis that the similarity of giver and receiver pairs that were actually observed are higher than random pairs. For this random null model we pair the receiver with all other givers and measure the similarity of these pairs. To do so we compute the sum of all posts/comments the two have had on common subreddits. Empirically, we find that these null model similarities for a given receiver have a large variance (some reddit users are active in many, many subreddits). Therefore we take the median of these null model similarities and compare against them in a paired test. We find that in 54.2% (214 out of 395) of cases the similarity of the actual giver/receiver pair is larger than the median null model similarity. While this is much lower than we anticipated it is significant as tested through a Wilcoxon Signed Rank Test ($T = 20049$, p=2.73e-04). We conclude that while significant the similarity between giver and receiver in terms of shared subreddits is not a large factor.

Subreddits: We also looked at other Subreddits requesters were active on to see if certain communities could raise a persons chances of success. We extracted a network of all subreddits, were two subreddits have a link of weight w, if w people are active on both these subreddits. Then we clustered these subreddits into 10 meaningful clusters based on the network, each cluster had at least 450 subreddits (see Figure 13). Our results show that there is not a big difference in success rates based on which other subreddits users have participated in (see Table 3), but certain communities of subreddits can increase the chance of success by around 18%.

Statistical Model Evaluation

Early analysis showed that request length was strongly correlated with success. The Chi-square ANOVA test between the model of length versus the pure random model gives the p-value of 3.74e-14 indicating very high significance of this feature. Therefore, we use length as our baseline model for testing the impact of other groups. However, there needs to be a note of caution here. It does not necessarily follow that the more you write, the more you are likely to get pizza (correlation versus causation). The more you write the more you well get the opportunity to craft a compelling narrative that shows appreciation, respect, and your needs. There are also unobserved factors associated with the length of writing including grammatical formalism or clear narration that might

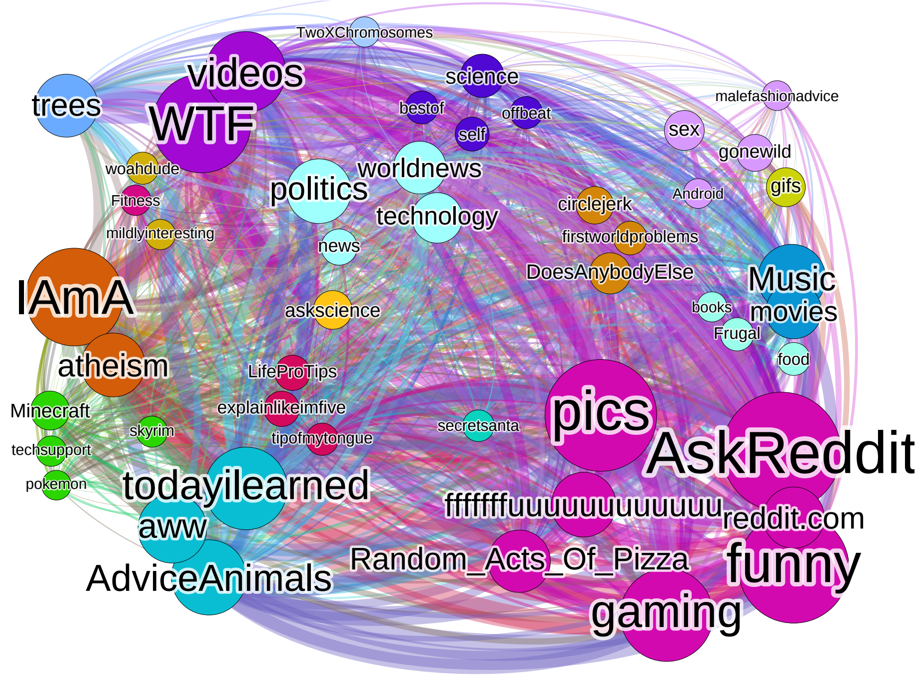

Figure 13: The graph of subreddit clusters. Note that we only show the most important subreddits in this visualization.

cause a boost in success rate as well. In other words, just writing long requests does not necessarily increase your success rate. Most likely, one needs to make good use of this opportunity as well (see our results from LIWC categories).

Now, we move on to do ANOVA Chi-square test on all other models against the length model. Here are the p-values reported:

Features	P-value	Significance
Linguistic vs Length	4.505e-05	*
LIWC vs Length	4.112e-12	*
Users vs Length	2.2e-16	*
Time vs Length	1.154e-08	*
Community Age vs Length	3.642e-08	*

This suggests that all those models would give significant boost in the fit. We then turn to analyze the statistical significance of each parameter in each model using z-score test:

Linguistic Model:

Features	P-value	Significance
length	7.13e-13	*
emoticons	0.93806	
hedges	0.25410	
kincaid	0.10447	
thank	0.01550	*
image	0.00377	*
links	0.56543	
payitforward	0.00171	*

LIWC Model:

Features	P-value	Significance
length	3.32e-14	*
Feel	0.082	
Friends	2.83e-06	*
Leisure	0.409	
Money	0.199	
Time	9.01e-07	*
Work	0.536	

Time Model:

Features	P-value	Significance
length	6.50e-12	*
Month	0.0270	*
Hour	0.1091	
Weekday	0.9942	
Community Age	9.45e-09	*
4h-before	0.0537	
First Half of Month	0.0120	*

User Model:

Features	P-value	Significance
length	7.46e-16	*
Age	0.0791	
Age-Pizza	0.0388	*
Karma	0.7893	
Score	0.7369	
Posts count	0.6494	
Posts-Pizza count	4.79e-06	*
Comments count	0.4354	
Comments-Pizza count	7.55e-05	*
Subreddits count	0.5090	

Community age Model:

Features	P-value	Significance
Length	1.50e-11	*
Community age	4.38e-08	*

After this steps, we obtain the two initial sets of candidates for the final models: length, community age, Friends, Time, Posts-Pizza count, Comments-Pizza count, thanks, image, payitforward, month, First-Half-Month, Age-Pizza, Age

Doing the ANOVA test of these candidates against the full model (all variables), we obtain p-value 0.3146, which suggest that these candidates capture all the variability. Hence, we use them as our final model. Here is the final fit and their p-value:

Final Model:

Features	P-value	Significance
length	1.22e-11	*
Community Age	$< 2e\text{-}16$	*
Friends	4.46e-06	*
Time	1.20e-07	*
Posts-Pizza count	5.10e-07	*
Comments-Pizza count	0.000323	*
Thank	0.017766	*
Image	0.003056	*
payitforward	0.037217	*
month	0.011218	*
First-Half-Month	0.002724	*
Age-Pizza	0.003163	*
Age	4.12e-05	*

As predicted, they are all statistically significant. Hence we have derived our final model.

Predicting Success

The previous section focused on statistical significance tests and model fit mostly on training data. In this subsection we show that the studied dynamics indeed have predictive value on unseen data as well. We use Logistic Regression models as before and measure the area under the curve (AUC) in the Receiver Operating Characteristic (ROC) that plots the true positive rate against the true negative rate (a very common measure for predictive accuracy in the presence of noise). We perform ten-fold cross-validation on our dataset to reduce random effects. We further explore unigram (bag-of-words) and bigram models to compare our performance to a standard model for text-based prediction. Since these models have a larger number of parameters we employ Support Vector Machines with L1 penalty on the parameters to enforce sparsity in the parameters (found to decrease overfitting).

We define a simple baseline model using just the length feature and the timestamp of the requests. Count stats include image links, the normalized number of thank you's, and whether the requester signals to "pay it forward". LIWC includes all the significant LIWC categories (Friends, Money, Work, Time, Feel, and Leisure). The community model includes the user account age and the number of accumulated up- and down-votes at the time of requests and represent the user and community features. The "full model" simply represents the combination of all of these features.

Model	**ROC AUC**
random baseline	0.500
count stats	0.564
LIWC	0.575
community	0.578
count stats + liwc	0.589
unigram	0.594
count stats + unigram	0.595
bigram	0.596
length and time	0.602
length only	0.603
liwc + unigram	0.604
count stats + liwc + unigram	0.604
length and time + unigram	0.614
time + community	0.622
community + unigram	0.624
length and time + count stats + liwc + unigram	0.625
length and time + count stats + liwc	0.635
time + community + unigram	0.643
full model + unigram	0.659
full model	**0.667**

Table 4: List of features with their effect (when increasing the feature value) on success rate as well as their statistical significance (two-tailed Wilcoxon Signed Rank Test on length-matched pairs except for the length feature itself which uses random pairs).

Table 4 shows several model combinations with their ROC AUC scores. All models perform significantly better than chance and the proposed features improve significantly upon simple baselines such as length or unigram models. Unsurprisingly, unigram (text) models provide the biggest gain when adding them to text-less features such as the community model. They do not add predictive accuracy to the full model (which has significantly less features / parameters). It is interesting to note that the significant features in the unigram model include many terms related to previously specified features such as "friend", "pay", or "jpg".

Discussion

Given prior research it might seem surprising that sentiment and politeness (other than gratitude) do not play a big role in success of online requests for pizza. It is possible that the tools used to capture sentiment and politeness, while demonstrated to work well on other datasets, fail to capture the same in reddit posts that tend to be short and more colloquial. Furthermore, the language complexity as measured by readability scores did not have a significant impact either.

With respect to temporal dynamics we found that the community as a whole shows a trend to fulfill smaller fractions of pizza requests over time. Obviously, this is nothing a new user in the community can influence to maximize her chance of success (unless she is able to travel back in time).

However, our findings suggests that there are several factors that the user can control that are significantly correlated with success. First, the request should be fairly long allowing the user to introduce herself and her situation. It also helps to put in additional effort to upload a photo. This is often used to increase the level of trust, e.g. by attempting to verify certain claims through the photo such as identity, location, financial situation, or simply an empty fridge. Our findings also suggest that pizza givers value the requesters ambition to give back to the community by forwarding a pizza later (even though some never do). With respect to the request content we found that talking about your friends and partners as well as your leisure activities can have a negative impact on your success rate. Instead it seems advisable to talk more about money, most likely a bad financial situation, and work. It also seems to help to express gratitude and appreciation in your request.

On the community side of matters, our findings show that user communities formed around social platforms are a key factor in understanding those platforms and the underlying mechanisms in them. We have so far found that the longer a person has contributed to the system (age) and the more active they have been (activity), the higher their chances of being rewarded by the system. We have also seen the effect of Karma (or how well received the user has been) on success. As we move forward with the project we will look deeply into network metrics and properties around the Reddit/RAOP community to understand and discover new features of this online social platform. We would like to emphasize here that due to the nature of the platform we have studied, "being rewarded" by the system has much more weight than receiving an up-vote, or good review. Here each reward costs around 15 dollars for the person giving away pizza. Thus we believe this paper brings a novel perspective and studying the RAOP online community could have high value in understanding similar platforms.

Conclusion

We explored what factors differentiate successful requests from unsuccessful requests. We introduced a dataset of several thousand of requests over the course of more than two years where every requests asks for the very same. This allowed us to analyze the question of how to ask for something and what factors matter for the success of an online request. Our findings include that putting in some effort, expressing gratitude, creating a sense of trust, the constructing the right kind of narrative, as well as being an active member of the community are significantly correlated with success.

Future Work

We believe that these findings could be very relevant to other online communities as well. For future work it will be interesting to see to what degree the concepts introduced in this paper generalize to other online communities. This further calls for a more general framework or models that explains the observed dynamics across communities. The importance of individual concepts might still very much depend on the community. For instance, spelling errors decreased funding success on professional platforms such as Kickstarter but did not do so significantly on the more colloquial Reddit community. Further, trust-related features such as including a picture is valued very much on Random Acts of Pizza where user accounts are not subject to a verification process whereas that might be different for communities that have trusted/verified accounts.

We are currently working on deeper linguistic analyses of the requests including modeling the narratives as sequences of verbs and measuring the consistency and content of the narrative through patterns of verb tenses.

Lastly, we are interested in factors of success across the spectrum of interaction cost. How do the dynamics change from getting people to click on or like your post to getting them to donate money to your project? Or more concretely on Reddit, what types of posts get up-voted as opposed to commented or gilded (bearing a small financial cost to gift the poster with premium access for a month).

References

[1] http://en.wikipedia.org/wiki/reddit. Last accessed on October 16, 2013.

[2] http://www.donorschoose.org/. Last accessed on October 16, 2013.

[3] http://www.globalgiving.org/. Last accessed on October 16, 2013.

[4] http://www.kickstarter.com/. Last accessed on October 16, 2013.

[5] http://www.kiva.org/. Last accessed on October 16, 2013.

[6] http://www.prosper.com/. Last accessed on October 16, 2013.

[7] http://www.quora.com/. Last accessed on October 16, 2013.

[8] http://www.reddit.com/. Last accessed on October 16, 2013.

[9] http://www.stackoverflow.com/. Last accessed on October 16, 2013.

[10] Pyenchant: A spellchecking library for python. http://pythonhosted.org/pyenchant/. Last accessed on October 16, 2013.

[11] Random acts of pizza. http://randomactsofpizza.com/. Last accessed on October 16, 2013.

[12] Random acts of pizza subreddit. http://www.reddit.com/r/Random_Acts_Of_Pizza/. Last accessed on October 16, 2013.

[13] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec. Discovering value from community activity on focused question answering sites: a case study of stack overflow. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 850–858. ACM, 2012.

[14] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec. Effects of user similarity in social media. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 703–712. ACM, 2012.

[15] J. Berman. Random acts of pizza: How to donate. http://abcnews.go.com/US/random-acts-pizza-donate/story?id=13950694#.UMmm6aWo9jY. Last accessed on October 16, 2013.

[16] S. Ceyhan, X. Shi, and J. Leskovec. Dynamics of bidding in a p2p lending service: effects of herding and predicting loan success. In *Proceedings of the 20th international conference on World wide web*, pages 547–556. ACM, 2011.

[17] C. Danescu-Niculescu-Mizil, M. Sudhof, D. Jurafsky, J. Leskovec, and C. Potts. A computational approach to politeness with application to social factors. In *Proceedings of ACL*, 2013.

[18] V. Etter, M. Grossglauser, and P. Thiran. Launch hard or go home!: predicting the success of kickstarter campaigns. In *Proceedings of the first ACM conference on Online social networks*, pages 177–182. ACM, 2013.

[19] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th international conference on World Wide Web*, pages 403–412. ACM, 2004.

[20] M. Herzenstein, S. Sonenshein, and U. M. Dholakia. Tell me a good story and i may lend you money: The role of narratives in peer-to-peer lending decisions. *Journal of Marketing Research*, 48(SPL):S138–S149, 2011.

[21] J. P. Kincaid, R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, DTIC Document, 1975.

[22] P. Koehn. Statistical significance tests for machine translation evaluation. In *Proceedings of EMNLP*, pages 388–395, 2004.

[23] H. Lakkaraju, J. McAuley, and J. Leskovec. Whats in a name? understanding the interplay between titles, content, and communities in social media. In *Seventh International AAAI Conference on Weblogs and Social Media*, 2013.

[24] G. Lakoff. *Hedges: A study in meaning criteria and the logic of fuzzy concepts*. Springer, 1975.

[25] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1361–1370. ACM, 2010.

[26] L. Lovász and M. D. Plummer. Matching theory, volume 29 of annals of discrete mathematics. *North-Holland*, 5:42–46, 1986.

[27] J. McGirk. The cheat code to life. http://www.wired.com/underwire/2013/07/how-to-cheat/all/. Last accessed on October 16, 2013.

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

[29] V. Narayanan, I. Arora, and A. Bhatia. Fast and accurate sentiment classification using an enhanced naive bayes model. In H. Yin, K. Tang, Y. Gao, F. Klawonn, M. Lee, T. Weise, B. Li, and X. Yao, editors, *Intelligent Data Engineering and Automated Learning IDEAL 2013*, volume 8206 of *Lecture Notes in Computer Science*, pages 194–201. Springer Berlin Heidelberg, 2013.

[30] Oxford Dictionaries. http://oxforddictionaries.com/-us/definition/american_english/request. Last accessed on October 16, 2013.

[31] C. Potts. Sentiment symposium tutorial: Tokenizing. http://sentiment.christopherpotts.net/tokenizing.html. Last accessed on October 16, 2013.

[32] Stanford NLP Group. Stanford CoreNLP. http://nlp.stanford.edu/software/corenlp.shtml. Last accessed on October 16, 2013.

[33] Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54, 2010.

[34] R. Wash. The value of completing crowdfunding projects. 2013.

[35] T. Yano and N. A. Smith. What's worthy of comment? content and comment volume in political blogs. In *ICWSM*, 2010.