

99designs: An Analysis of Creative Competition in Crowdsourced Design

Ricardo Matsumura Araujo

Federal University of Pelotas, Brazil

ricardo@inf.ufpel.edu.br

Abstract

We provide a study of crowdsourced design based on the popular service 99designs, a website where users create design contests for other users to submit their ideas for evaluation, competing for a monetary amount offered. By quantitatively analyzing a large number of contests, we report on the dynamics of the interactions between contests and designers. We show results on the effects of different financial incentives and contests' properties on the number and quality of submissions, as well as on how rewards are distributed across designers. We find that higher financial incentives do not translate to more effort by individual designers, but nonetheless have an impact on the quality outcome of contests by attracting a larger pool of designers. We also show that a majority of contests are dominated by a disproportionate few designers, which are both more active and effective in the service.

Introduction

Crowdsourcing is an approach to solve problems by posing these problems to an open crowd, where a reward is offered for solutions meeting some quality threshold. Several services now exist that facilitates, through the Internet, the interaction between those seeking to solve problems and those capable of solving them. These services range from the very general (e.g. Amazon's Mechanical Turk) to the very specific, which propose to tackle only a particular kind of problem (e.g. t-shirts, logo design), and thrive by allowing anyone to have access to a diverse and potentially talented pool of workers. On the other end of the transaction, workers have access to an influx of problems to solve, with different levels of difficulty and reward.

There are many ways to structure a crowdsourcing service. In one typical setting, those seeking to solve problems create a contest specifying the problem, quality thresholds and other information that may help to solve the problem, and specify a reward to be given only to the best solution presented. This reward is often monetary.

Even in such (rather simple) setting, there are several parameters that may impact the outcome. A major challenge in crowdsourcing systems is to be able to attract many good

workers. For that to happen, the right set of incentives must be put in place. For instance, one must choose how much the winner will be paid. Intuitively, the more one pays, the more workers will be interested and work hard to earn the reward. However, evidences have been shown that quality may not change with the size of the reward (Mason and Watts 2010; Huang, Singh, and Mukhopadhyay 2013).

We report on an exploratory analysis of data extracted from the popular crowdsourcing service 99designs¹, which focus on design tasks such as logo, business card and web design. This service allows users to create contests where other users may submit design ideas for evaluation; at the end, the contest holder chooses a design and the winning designer is rewarded a pre-specified monetary amount. Several services exist that are similar to 99designs (e.g. Crowdspring, Crowdsite), making it a valuable resource to understand a wider range of crowdsourcing services.

Our goal is to better understand the mechanics and dynamics of contests from the perspectives of both contest holders and designers. In order to do so, we collected data from over 38,000 contests. One central contribution of this work is the focus on the quality of creative contests, which are evaluated in a more qualitative manner than more mechanical tasks.

We show results that corroborate the service's marketing pitch, showing that quality of a contest does correlate with the number of designers participating. We further detail this phenomenon, providing evidence of the role of rewards in attracting more designers and with the amount of effort put by them. In addition, we provide results and discussions on the distributions of designers per contest and of awards won per designer, showing that there are evidences that there are consistently more effective designers and that a large portion of designers are unable to win a single contest.

This work is organized as follows. We first briefly discuss relevant previous work related to ours. We then provide a description of 99designs, followed by a description of the data set, including the methodology used to capture it. Results on the relationship between quality of a contest outcome and the quantity of designers participating are then shown and we further detail how different attributes of the contests affect these variables. Finally, we provide results on how de-

¹<http://www.99designs.com>

signers are distributed across contests and how effective they are in winning contests.

Related Work

This paper adds to a growing set of studies covering different crowdsourcing systems. The first attempt to define crowdsourcing is attributed to (Howe 2006), where it is stated that “crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call”.

Amazon’s Mechanical Turk is arguably the most popular general-purpose crowdsourcing service. It has been used not only for solving problems from a range of business domains (Howe 2006), but also as a tool for social (Buhmester, Kwang, and Gosling 2011; Mason and Suri 2012) and user (Kittur, Chi, and Suh 2008) studies. Mechanical Turk itself has been studied intensively, aiming at better understanding and characterizing its dynamics, including its limitations and capabilities (Ipeirotis 2010; Mason and Watts 2010).

In addition to Mechanical Turk, several systems focus on tasks from more specific domains, aiming at building a stronger community and making interaction easier for the composing parts. These systems often provide unique features that fit their contexts and several studies have tackled the problem of understanding these services’ communities (Harper et al. 2008; Brabham 2010; Archak 2010; Harris 2011; Yang, Adamic, and Ackerman 2008). In (Yuen, King, and Kwong-Sak 2011), a survey of the many mechanisms in place is made. As far as we know, no previous work focused on 99designs, although it is often mentioned in the literature and is one of the most successful cases.

The effect of incentives in crowdsourcing systems is an important aspect of such systems and a major research area for both experimental and theoretical researchers. In (Archak and Sundararajan 2009; Chawla, Hartline, and Sivan 2012) a theoretical perspective was applied to the question of how to design an optimal contest for different types of participants. In (Mason and Watts 2010) several experiments using Mechanical Turk were conducted, concluding that workers’ accuracy is essentially unchanged with changes in the amount offered for the completion of the same task, even though higher incentives attract more workers. In (Huang, Singh, and Mukhopadhyay 2013) it was argued that quality can even *decrease* with higher incentives as they can drive away better-than-average workers repelled by the increased competition. On the other hand, there are evidences that well-placed (intrinsic and extrinsic) incentives do improve the quality of solutions in Mechanical Turk (Rogstadius et al. 2011). A similar result is reported by (Horton and Chilton 2010). Improvement in quality with higher incentives was also observed outside Mechanical Turk - e.g. (Harper et al. 2008).

99designs

99designs follows much of the mechanisms found in a typical crowdsourcing service. The main concept behind the service is that of a *contest*, which can be created by any user.

Logo design contests

Our team of design professionals compete to produce a unique design for you. Post a contest, review the designs and pick a winner, it's that easy.

| Contest Title | Contest Holder | Ends | Entries | Package |
|--|----------------|---------------------|---------|-------------------------|
| New Logo For - CHARGER CONSTRUCTION & DESIGN We are a general contracting and construction company | MATMAD | selecting finalists | 113 | silver \$499 |
| KokoonMe needs a new logo We are the LinkedIn of Health and Wellbeing | Jayshortt | selecting winner | 41 | silver \$495 |
| Create the next logo for COBB CORP Cobb Corp is an investment company that specialises in the Finance and Energy industry sectors | Dcobb | selecting winner | 200 | \$400 |
| logo for CherryPlayer.com Cherryplayer is a free Windows multimedia player that needs a new logo! | Johan8511 | won | 619 | bronze \$299 guaranteed |
| Help Pacific Mortgage Consultants Inc with a new logo Mortgage Company | Plorentin | won | 65 | bronze \$299 |
| Help Fuff & Puff (GAD) under name very small print with a new logo New Company that will strive to excel in business of establishing a chain of net | Tlahop | won | 35 | bronze \$299 |
| Create the next logo for AGI Solutions, Inc. or AGIS AGI Solutions Inc. is a business consulting or professional services firm targeted for medium to large size... | Aginwin | selecting finalists | 1584 | \$899 |

Figure 1: Screenshot of a public browsable list of open contests in 99designs.

Contests are created within several design categories - logo, website, business card, book etc. A contest is mostly composed of a short title, a design brief and a package. The *title* is a short description of what is needed, while the *design brief* is a longer version fully detailing the contest. Several *packages* are available at different prices. Along other minor information, a package specifies how much will be paid to the winning designer (the prize) and whether payment is *guaranteed* or not.

A contest by default will run for at least 7 days. In the first 4 days (qualifying round), submissions are open and the contest holder may evaluate submissions and give feedback through comments. At the end of this round, the contest holder chooses up to 6 designers to move to the next round. If no designers are chosen and the contest is not guaranteed, the contest ends without a winner. If at least one designer is chosen, the contest becomes guaranteed - i.e. a designer is guaranteed to be paid in the end of the contest. At this last round, new submissions may only be made by these chosen designers. Finally, after 3 more days, a winning design is chosen and its designer is paid. Hence, 99designs uses a winner-takes-all mechanism, which may not be the most efficient for the participants (Cavallo and Jain 2012).

At any round, the contest holder will evaluate designs by giving them ratings from 1 to 5 stars, which have the following description, in order: “Has Potential”, “Right Direction”, “Good Design”, “Great Design” and “Leading Contender”. Additionally, the contest holder may eliminate any design from the competition and leave individual feedback on each design. The order of submission is always shown, so one can track the evolution of entries.

A designer, on the other side of the transaction, is given a browsable list (Figure 1) of open contests and may choose to enter at any time during the first round by submitting a

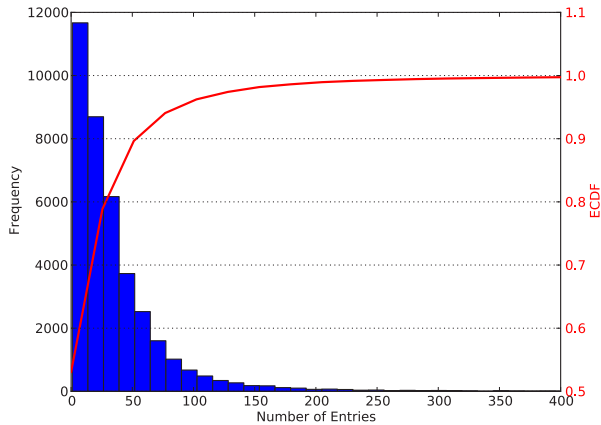


Figure 2: Number of entries per contest.

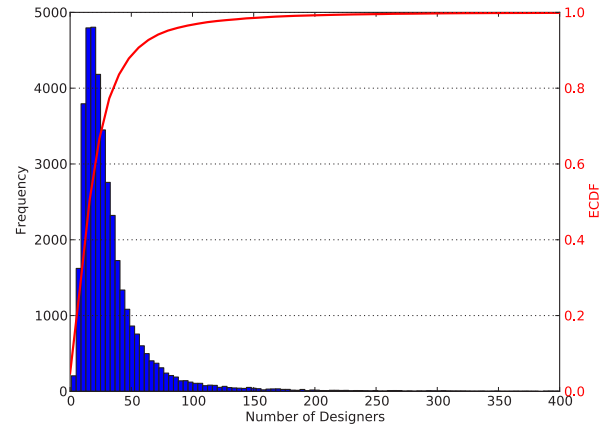


Figure 3: Number of designers per contest.

design. Designers may withdraw their designs at any time, which causes the image associated to the design to become hidden while all other information is kept at the contest.

Data Description

99designs partitions its contests in several categories, covering a range of design tasks (e.g. book cover design, web page design). The most popular of these categories is *logo design*, with over 60% of open contests fitting into this category. In this paper we focus in this category due to its popularity. The data set is available at <http://ricardoaraujo.net/datasets/99designs/>.

We collected a total of 38,361 logo design contests from 99designs website, between 2010 and early 2012, with most (52%) contests taking place in 2011, along with associated 1,453,870 designs from 63,049 unique designers.

The median number of designs per contest is 24 and the median number of designers per contest is 9. The distributions are skewed as shown in Figure 2 and 3. Most contests (90%) have at most 58 submissions, but a very few contests attract many more - the most popular contest, aiming at creating a logo for popular Occupy Movement, gathered 5097 submissions from 1591 designers. Likewise, 90% of the contests have at most 57 designers and the distribution’s mode is 17 designers.

Quantity versus Quality

An implicit assumption behind this type of crowdsourcing model is that by receiving designs from many different designers, one is bound to like some of the submissions - i.e. quality follows quantity. In this section we tackle the hypothesis that contests with a higher number of designers yield higher quality outcomes.

While quantity is easily measurable by taking the number of designers that submitted ideas, care must be taken when attempting to measure quality. There are many different ways to measure the quality of a contest. A straightforward way would be to consider all contests that finished with

a winner as a success. Unfortunately, we did not have access to contests that did not finish, as they are not publicly available. Moreover, this definition is very coarse since not all contests with winners may have a satisfactory outcome. This is because after the first round of submissions, the contest becomes guaranteed - there will be a winner no matter what happens in the next rounds. Furthermore, several contests are guaranteed from the start.

A more informative way to measure quality is to consider the ratings given by the contest holder for individual submissions. We consider a contest a success if at least one high quality submission was made. This is a reasonable definition, since the goal of a contest is to generate one useful design. Given the semantics embedded in the rating system, we define a high quality submission as one rated with at least 4 stars (“Great Design”) by the contest holder.

One could argue for averaging submissions’ ratings to measure a contest quality. However, a contest that is able to attract good designers may also attract many more low quality submissions (Figure 4). These submissions do not affect the end result as it only takes a single acceptable design for a contest to be successful, but they do introduce a bias, rendering the average meaningless.

In order to characterize the relationship between quantity and quality, we grouped contests by the number of participating designers (e.g. all contests with 10 designers, all contests with 20 designers and so on) and for each contest in each group we checked to see if there was at least one acceptable entry. Only groups with at least 50 contests were considered in order to reduce noise (essentially cutting out the few contests with an unusually high number of designers).

Figure 5 plots the resulting fraction of contests with a successful submission, which we interpret as the empirical probability of success of a contest, against the number of designers. There is a strong Spearman’s correlation between quantity and quality ($\rho = 0.79$) and the plot is almost linear but for a slight observable damping when increasing the number of designers. Hence, it is the case that more design-

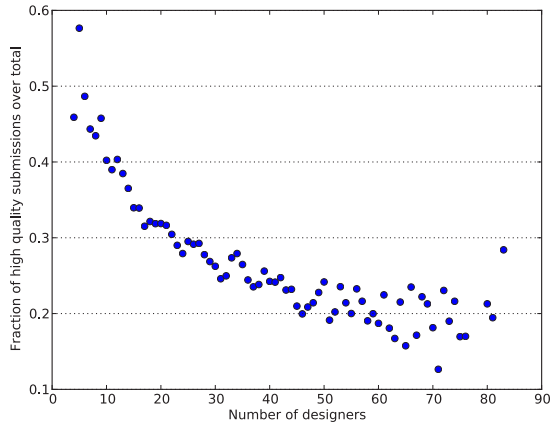


Figure 4: Fraction of designs with rating at least 4 over total number of submissions for contests with different number of designers.

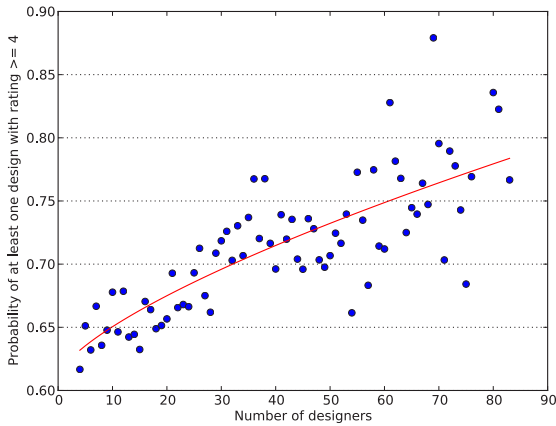


Figure 5: Fraction of contests receiving a submission of rating at least 4 for different number of participating designers. The solid line represents fitted power function $y(x) = -0.0079 * x^{0.6985} + 0.6111$

ers lead to a better chance of finding a satisfactory design. A linear relationship (for the observed range of values) is expected if one considers the hypothesis that each designer has the same probability of producing an acceptable submission; more designers lead to a linear increase in the probability of success. Of course, a dampening is expected as the probability of success approaches 1.0 or 0.0 and a slight dampening can already be observed for the shown range of values.

Effects of Financial Incentives

A key decision when creating a contest is how much will be paid to the winner. The reasoning goes that the more you pay, the higher the chances of attracting more and better designers and have an acceptable final product. However,

in (Mason and Watts 2010) evidences were shown that although more workers are indeed attracted by higher financial incentives in a crowdsourcing system (based on Mechanical Turk), an increase in quality do not typically follow. On the other hand, there are cases where financial incentive does play a role in the quality of the outcomes (Harris 2011). In this section, we study the effects of financial incentives in 99designs. We compare the number of submissions, number of designers and the quality of submissions for different price points.

Reward Values

For logo designs and US customers in 99designs, there are three default price points that can be chosen: U\$299, U\$499 and U\$699. However, one can create a custom package and offer a different amount. We found that most prizes are around these values. We chose to consider amounts grouped around these default ones: (0, 200], (200, 400], (400, 600], (600, 800], (800, ∞]. In this analysis, we only consider contests priced in US Dollars (over 90% of the data set).

We start by testing the hypothesis that higher rewards attract more designers. Figure 6 shows how the number of designers are distributed across these price ranges. A Tukey HSD test shows no statistically significant differences between the two first groups ($p = 0.90$), but a statistically significant difference for all other group combinations ($p < 0.001$). As expected, higher financial incentives lead to an increased number of designers submitting to contests. Considering only the three center price ranges (which all have equal ranges), the largest increase in number of designers happen when transitioning from (400, 600] to (600, 800] - a 72% increase in median number of designers with an average cost only 40% higher. Going from (200, 400] to (400, 600], on the other hand, increase the median number of designer by 38% but with a 67% higher average cost. The middle range has the worse cost per designer, with both lower and upper ranges with about the same average cost per designer.

A higher financial incentive also has a positive impact in a contest's probability of success. Figure 7 plots the fraction of successful contests for each price range. While the lower range sees only a 70% chance of success, the most generous contests reaches a 83% rate.

An important question is whether the observed higher probability of success is solely due to the increase in number of designers or also due to better designs being submitted. Figure 8 shows evidence that controlling for the number of designers, the differences in the outcomes are quite small, with both intercepts and slopes of the best fit lines having very similar values and no discernible pattern among them. This implies that more designers are attracted to higher prizes, but these designers are not incentivized to produce *better* designs.

One way to measure how much effort designers put into contests is by the number of submissions they make. Designers can submit different, independent, designs or submit tweaks to their previously submitted designs, based on feedback given by the contest holder. In any case, more submis-

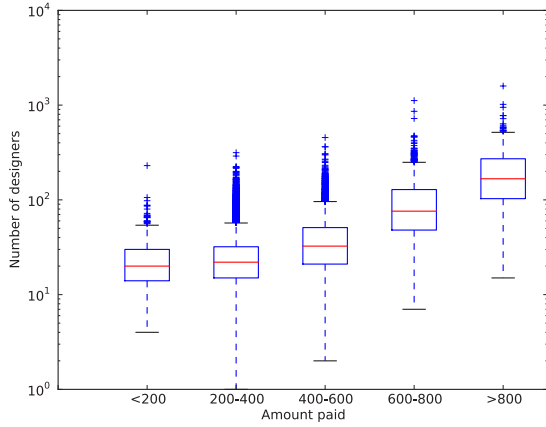


Figure 6: Number of designers per contest for different price ranges.

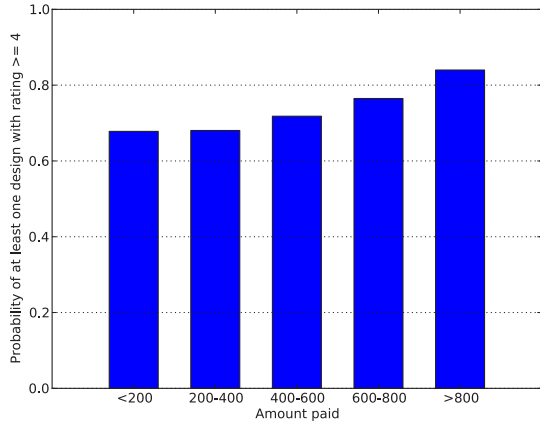


Figure 7: Empirical probability of success for different price ranges.

sions typically mean the designer cares enough to listen to feedback and provide improvements to the initial proposal. Indeed, no contest in our data set was won by a designer that submitted a single entry. Having designers submit several entries is an essential part of contests (as can be seen in Figure 2).

Figure 9 shows that the number of submissions per designer is largely unchanged for each price range. Even though an ANOVA rejects the hypothesis that all the means are equal, $F(4, 35364) = 13.73, p < 0.001$, the measured differences are very small and do not show any trend. This corroborates the evidence that higher financial incentives do not make designers try consistently harder.

Guaranteed Contests

Choosing a prize value is not the only important decision when creating a contest. A potential issue with the way the

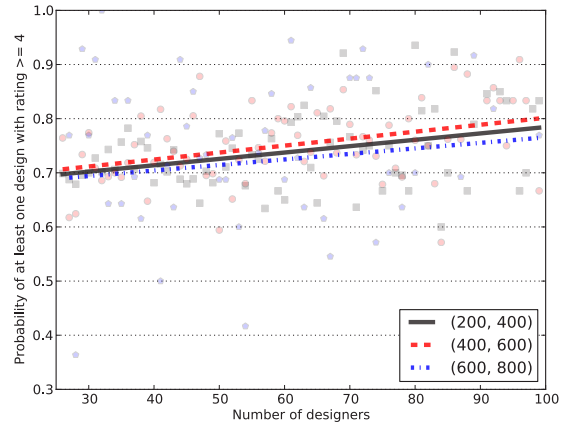


Figure 8: Probability of success for different number of participating designers and price ranges with best fit lines for each range. Upper and lower price ranges are not shown due to the small number of contests in them.

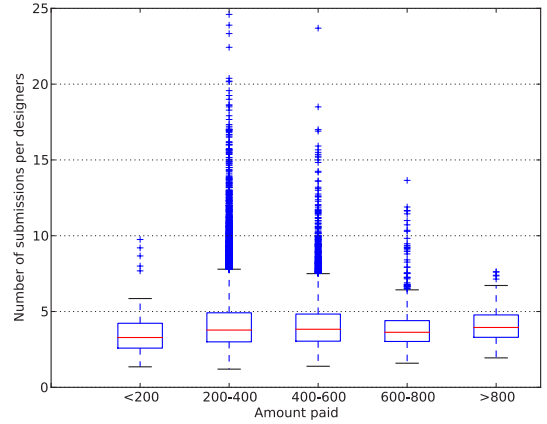


Figure 9: Number of submissions per designer for different price ranges. From left to right, the means are 3.7, 4.2, 4.1, 3.8 and 4.1.

service works is that contest holders can cancel a contest after the initial round of submissions. Since designers must send their ideas for evaluation, a deceitful contest holder could wait for the first round to end and walk away with the best submission so far, without paying anything to the designers. The contest holder would probably not get the best possible design and would not have the copyrights for the design, but he could be paying nothing for a possibly usable design.

In addition to deciding on a price, a contest holder can choose to reassure potential designers by guaranteeing payment from the start of the contest. In this case, he or she is charged at the moment of the contest's creation, with 99designs intermediating and guaranteeing the transfer to the winner at the end of the contest. This provides further incen-

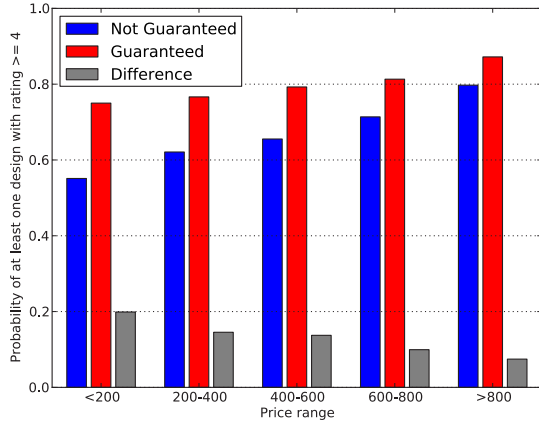


Figure 10: Fraction of contests with rating at least 4 for different price ranges and whether the contest is guaranteed or not.

tive for designers that could otherwise not enter the contest out of fear of being deceived. The distribution between guaranteed and non-guaranteed contests is fairly balanced. Those guaranteed compose 42% ($N = 14,991$) of the contests in our data set.

Guaranteed contests see a statistically significant increase of 25% in the mean number of designers when compared to non-guaranteed contests; $t(23214) = 24.65, p < 0.0001$. As a consequence, the probability of success increases by 22%, from 0.63 for non-guaranteed contests to 0.77; $t(37420) = 29.86, p < 0.0001$. Interestingly, guaranteeing a contest has a higher effect for intermediate reward values (Figure 10). One possible reason for this is that designers could believe that deceitful contests would choose lower prizes to reduce their risks - if the prize happens to be enforced for some reason, the cost would be lower. In other words, designers seem to evaluate the risk of being deceived against the size of the reward.

Designers and Success

From the 63,049 designers in the data set, 25,384 (40.2%) only participated in one contest. The most active designer participated in 1071 contests and the mean is 7.8 contests per designer. The most successful designer won 150 contests (0.4% of the total number of finished contests). On the other hand, 84% of designers did not win a contest and 8% won a single contest. The distribution closely follows a power-law (Figure 11).

Considering only designers that won at least one contest, the number of contests a designer participates in, naturally, correlated with the number of contests he or she wins (Figure 12). More than a few designers submit to thousands of contests and are able to show a high number of wins. However, most designers do not have a high success rate. Figure 13 shows the distribution of success rate, as defined by the ratio of numbers of wins to number of contests, for designers that participated in at least 10 contests. The distribution

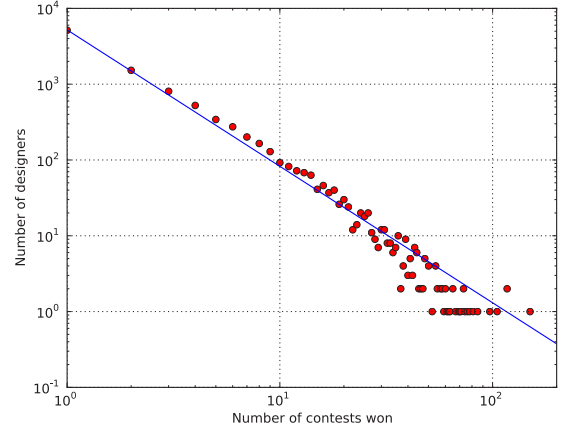


Figure 11: Distribution of wins among designers. The solid line represent a power-law distribution with exponent 1.8.

is log-normal, without evidences of extreme outliers - e.g. exceptional designers that win most of the contests they participate in.

Table 1 compares designers below the first quartile and above the third quartile when ranked by success rate. Rather than making a statement about competence, these differences hint at two different designer classes, or different strategies applied by designers. In one case, designers submit to many different contests but make few iterations in each one; in the other case, they enter into much fewer contests and make more submissions to each. Hence, higher effectiveness is obtained by focusing on fewer contests.

The less effective designers, on average, enter into higher valued contests when compared to their more effective counterparts, but both end up gaining exactly the same amount per contest they win. However, when comparing how much is earned per *submission* (entries) the difference is striking: the more effective designers are able to earn a lot more for each submission. If we take submissions as a proxy measure for effort, this shows that while effective designers are able to win more contests by increasing effort put into each one, they end up making much less total effort across all contests they participate in.

Conclusions

In this paper we reported on the results of quantitative studies conducted over data extracted from the popular crowdsourcing service 99designs. Our goal was to understand the dynamics of key features of the service, which are common to several crowdsourcing services. In particular, we focused on the effects of financial incentives on the quality of outcomes and the relationship between quantity and quality.

We analyzed over 38,000 contests spanning three years of activities. We found that there is an almost direct linear relation between the number of designers participating in contests and the quality of these contests. This corroborates the service's concept that quality follows quantity - given a large

Table 1: Comparison of different attributes for designers below the first quartile and above the third quartile for designers ranked by success rate. Each set is composed of 2400 designers and p-values are the results of t-tests.

| Attribute | Lower quartile Mean (σ) | Upper quartile - Mean (σ) | t-statistic | df | p-value |
|-------------------|----------------------------------|------------------------------------|-------------|-------|----------|
| Contests/designer | 99.61 (124.30) | 14.45 (40.91) | 31.65 | 2867 | < 0.0001 |
| Entries/contest | 3.60 (1.64) | 5.04 (4.46) | 14.68 | 2991 | < 0.0001 |
| Wins/contest | 2.34 (2.94) | 3.14 (7.99) | 4.49 | 2998 | < 0.0001 |
| U\$/contest | 402.76 (167.76) | 378.98 (154.95) | 17.97 | 67997 | < 0.0001 |
| U\$/win | 369.03 (125.73) | 371.65 (141.73) | 0.69 | 4661 | 0.715 |
| U\$/submission | 3.03 (1.94) | 56.28 (84.93) | 30.49 | 2366 | < 0.0001 |

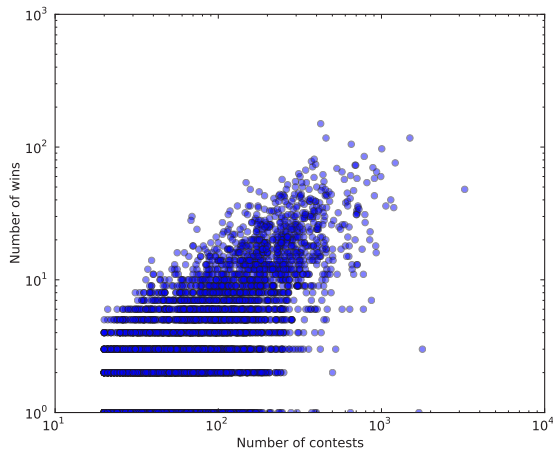


Figure 12: Log-log plot of number of contests versus number of contests won for designers that won at least one contest. The Pearson correlation coefficient is $\rho = 0.69$.

enough number of designers submitting, there is a high probability of obtaining an acceptable design as the outcome.

We then examined the effects of financial incentives on the quality of designs. We showed that offering a higher incentive does attract more designers and increase the probability of receiving a high-quality design. We found that the increase in number of designers is the major explanation for the increase in quality; controlling for the number of designers leads to little to no change in quality by increasing the amount offered. Indeed, we presented evidences that designers in higher paying contests do not work harder compared to lower paying contests. This last result is in line with the results presented by (Mason and Watts 2010), which was applied to non-creative setting, showing that similar mechanics seem to be in place for both types of crowdsourced tasks.

Finally, we studied the service from the point of view of the designers. We found that the distribution of wins follows a power-law, with some designers winning a disproportionate number of contests while a majority winning very few. We ranked designers by their effectiveness (wins per contest entered) and extracted features of highly effective designers, showing that effectiveness comes from focusing efforts into fewer contests, a strategy that was shown to pay off.

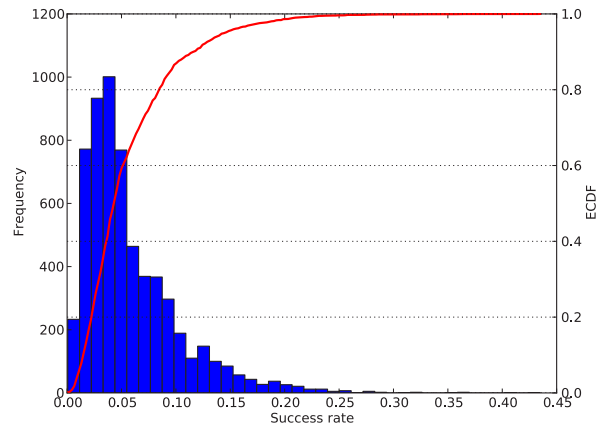


Figure 13: Histogram and ECDF for designers' ratio of wins to number of contests entered.

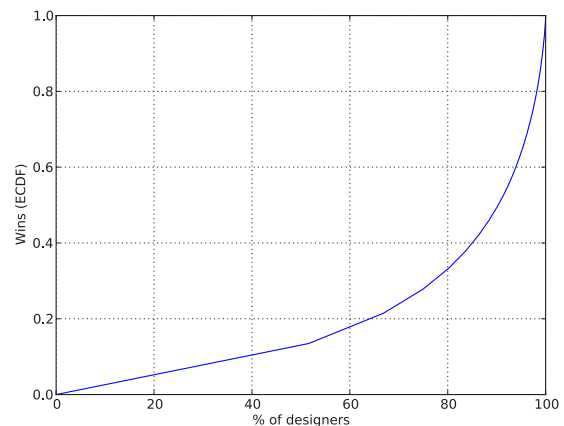


Figure 14: Empirical CDF for number of wins per designer for all designers with at least one contest won.

These results together are evidence that the probability of success of a contest is more related to the probability of attracting at least one of the few better designers in the system, rather than the raw number of designers participating. In this sense, having more designers is just a way to increase

the odds of having those good designers, but it may not be the most effective way to do so. In the limit, if one knew beforehand who the best designers are, one could do better by approaching them directly. Of course, in creative business “best” cannot be measured objectively and a creative match between a designer and the contest holder is expected to also play an important role. In this sense, 99designs provides value by allowing a better matching - not only attracting many designers increase the odds of attracting good designers, but also increase the odds of having a diverse pool of good designers, hence increasing the chances of finding a match.

This paper provided two main contributions. First, it provided a first quantitative study on a very popular, community-driven, crowdsourcing system. We hope that it opens doors for future research in this system, as it provides an interesting venue for research on creative crowdsourcing in a more structured and specialized place than e.g. Mechanical Turk. Second, this work provides a different look at crowdsourced product quality, measuring it from the contest level instead of the individual workers’ level. By doing so, we believe we were able to better capture the main goal of similar crowdsourcing systems - i.e. to obtain one single usable solution - which can prove useful for the design and use of similar systems.

Acknowledgments

This work is supported by CNPq (Brazilian National Research Council) through grant number 477937/2012-8.

References

Archak, N., and Sundararajan, A. 2009. Optimal design of crowdsourcing contests. *Proceedings of the International Conference on Information Systems 2009*.

Archak, N. 2010. Money, glory and cheap talk: analyzing strategic behavior of contestants in simultaneous crowdsourcing contests on topcoder.com. In *Proceedings of the 19th international conference on World wide web, WWW '10*, 21–30. New York, NY, USA: ACM.

Brabham, D. C. 2010. Moving the Crowd at Threadless. *Information, Communication & Society* 13(8):1122–1145.

Buhrmester, M.; Kwang, T.; and Gosling, S. D. 2011. Amazon’s Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 6(1):3–5.

Cavallo, R., and Jain, S. 2012. Efficient crowdsourcing contests. In *AAMAS '12: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems.

Chawla, S.; Hartline, J. D.; and Sivan, B. 2012. Optimal crowdsourcing contests. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '12*, 856–868. SIAM.

Harper, F. M.; Raban, D.; Rafaei, S.; and Konstan, J. A. 2008. Predictors of answer quality in online Q&A sites.

Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems 865–874.

Harris, C. 2011. You’re hired! an examination of crowdsourcing incentive models in human resource tasks. *Proceedings of the WSDM Workshop on Crowdsourcing for Search and Data Mining (CSDM)* 15–18.

Horton, J. J., and Chilton, L. B. 2010. The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM conference on Electronic commerce, EC '10*, 209–218. New York, NY, USA: ACM.

Howe, J. 2006. The rise of crowdsourcing. *Wired magazine* 14(6):1–4.

Huang, Y.; Singh, P.; and Mukhopadhyay, T. 2013. How to Design Crowdsourcing Contest: A Structural Empirical Analysis. *Proceedings of the Conference on Information Systems and Technology (CIST 2012)*.

Ipeirotis, P. G. 2010. Analyzing the Amazon Mechanical Turk marketplace. *XRDS: Crossroads, The ACM Magazine for Students* 17(2):16.

Kittur, A.; Chi, E. H.; and Suh, B. 2008. Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08*, 453–456. New York, NY, USA: ACM.

Mason, W., and Suri, S. 2012. Conducting behavioral research on Amazon’s Mechanical Turk. *Behavior Research Methods* 44(1):1–23.

Mason, W., and Watts, D. J. 2010. Financial incentives and the “performance of crowds”. *SIGKDD Explorations Newsletter* 11(2).

Rogstadius, J.; Kostakos, V.; Kittur, A.; Smus, B.; Laredo, J.; and Vukovic, M. 2011. An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* 11.

Yang, J.; Adamic, L. A.; and Ackerman, M. S. 2008. Crowdsourcing and knowledge sharing: strategic user behavior on taskcn. *Proceedings of the 9th ACM conference on Electronic commerce* 246–255.

Yuen, M.-C.; King, I.; and Kwong-Sak, L. 2011. A survey of crowdsourcing systems. In *Proceedings of the 2011 Privacy, security, Risk and Trust (PASSAT)*, 766–773.