Reputation and Attention Wealth

Chunyan Wang,Yiwei Yao,Ye Jin

In this paper, we investigate the distribution and growth dynamic of individual producer's aggregated attention wealth in crowd-sourcing phenomenon by mining data from digg. com. We observe that, the distribution of attention wealth deviates from power-law with a peak near the tail. We interpret the deviation with a stochastic model for attention wealth growth incorporating both promotion to front page and reputation effect in the social network. We further confirm our results by simulating the growth process. This model gives quantitative measure of the reputation's impact to wealth distribution in attention economy. It also provides new elements for the understanding of collective attention world.

PACS numbers:

INTRODUCTION

With the advent of information age, attention as a mental capitalism with scarcity has been studied by researchers for a better understanding of various phenomenon such as commercial trends and donation dynamics under many contexts including circulation sizes and TV ratings [1, 2]. Among these many media, world wide web provides an unprecedented empirical laboratory for the study of attention economy, where millions of people could publish their productions in the form of articles, music and videos to gain other people's attention [3, 4].

In the past few years, extensive studies have been conducted to find strong regularities and interesting patterns of world wide web [5–7]. The study of crowd-sourcing phenomenon, which is exemplified by web sites such as YouTube, digg and Wikipedia, offers insights on Internet and attention economy [8]. Great efforts have been made on investigating how the contributions or rewards in the form of attention distribute among peers: heavy-tailed distribution is found to be ubiquitous [9]; the submission count of each producer in Youtube or Digg is found to follow power-law distribution [10]; the digg count of a promoted story to front page in digg web site obeys log-normal distribution, which is well interpreted by a dynamic model characterized by a novelty factor [11]. However, few efforts have been focused on treating each published item's attention as an aggregative wealth of producers, like capital in economic world. It is thus interesting to study such attention wealth to explore new findings. In the meanwhile, what also attracts people's great interest, is the role played by attention in feedback loops of peer production. It was found that producers with more followers tend to attract more attention [12]. Thus, questions like how individual producers' reputation grows and what role of reputation is in social network for attention wealth accumulation, is of value and interest.

In this paper, we focus on the growth dynamic study of individual producer's attention wealth accumulation under the impact of reputation in the network. We realize this by viewing single producer's existing attention of all products as a wealth signifying one's reputation in the community, which would in return boom wealth accumulation in future publications. We propose a model of promotion probability for producers with different attention wealth level. Based on this model, we give predictions and interpretations of income and wealth distribution. The model fits well with observations, and is further backed up by simulations.

METHODS

The web site in our study, digg. com, is an interactive media with transparent attention information which allows its users to share news or stories they create or discovered in the form of a URL from the Internet [11]. A digg number, which is counting how many viewers like the story in the past, is shown next to each story's headline, reflecting how much attention has been earned by the story [13]. Viewers of the web site have to select and direct attention to a few items from a very large pool of submissions. The web site provides functions like 'promotion' and 'subscription' to help viewers decide what to read. If the digg count of a new submission meets a specific number, it is promoted to the front page as a popular story of digg web site†. This kind of promotion mechanism is found to be influential on single product's attention increment [11]. The promoted stories receive significant more attention than those not promoted. On the other side, there are also functions like 'subscription' facilitating the formation of networks and reputation in the community. If one reads and likes a submission, one can subscribe the submission's producer and receive a notice, whenever there is a new submission from the pro-

[1]There was a period that promotion started at 15 digg count. Now, the actual promotion algorithm is claimed to be more complicated, taking story source and category into account. But our data indicates that only 21 out of $1,862,470$ stories were promoted with a digg count less than 15. So we still use 15 as a criterion in this research.

ducer.

Many fruitful progress has been made relating to the impacts of promotion and subscription in social network [11, 12, 14]. However, questions regarding to how a single story gets promoted, how a producer's attention wealth accumulates, as well as how the producer's reputation comes to impact promotion probability is still unclear. To resolve these problems, we gather our data from digg.com by tracking all the submissions' information including producer ID from January 2007 to November 2007. To study attention wealth distribution, we select out producers who made their first submission after January 2007. Data of 318,131 producers and their 1,862,470 submissions are all fully recorded. Time gap between submission date and final digg count record is long enough to ensure that the digg count of submissions becomes saturated. Each submission is labeled as promoted or non-promoted to distinguish whether it had ever met the specific digg count and entered the front page.

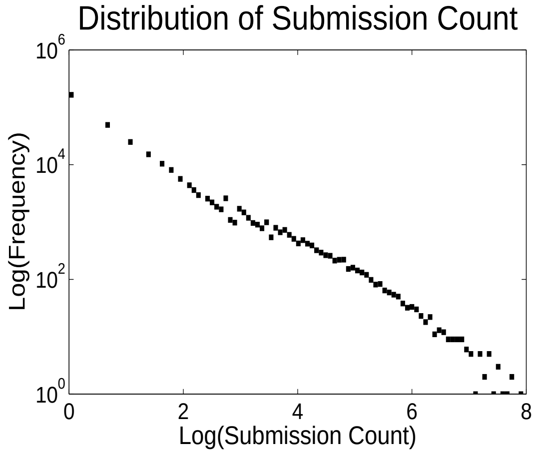

FIG. 1: Distribution of individual producer's submission count in log-log scale. The straight line indicates a power-law distribution of submission count.

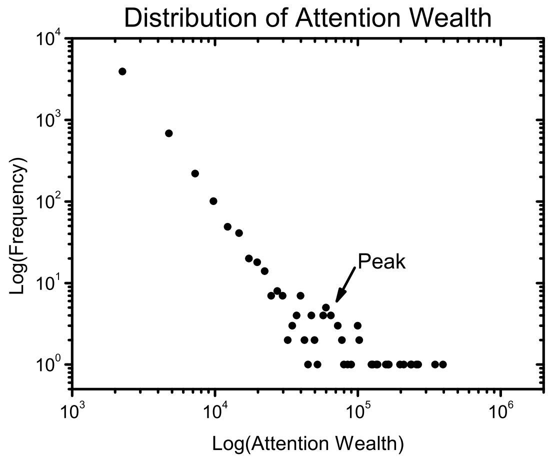

FIG. 2: Distribution of attention wealth in log-log scale. The peak near the tail indicates an aggregation of wealthy class.

RESULTS

In our first series of study, we look at the individual producer's contribution to the community. We first examine each producer's number of submissions. In accordance with previous observations and theories [10], the submission count follows well with power-law distribution [Fig.1]. While submission count reveals one aspect of single producer's contribution in the community, the distribution of aggregated total attention received by single producer from all submissions has not yet been studied. We look at individual's attention wealth distribution by summing up each producer's all submissions' digg count from the first upload to the latest by the end of year 2007. In doing so, we obtain information about all the producers' attention wealth, which is similar to the definition of wealth in real economic world. The plot of the distribution is as shown in Fig.2. It could be seen that the plot follows power-law distribution in the low wealth region and deviates from power-law with peak near the tail, where wealth level is around 10^5. It is natural to ask why the attention wealth deviates from submission count distribution if wealth is just the count of submissions multiplied by the expectation value of each product's digg count. We answer this question by proposing a model presented below, incorporating the promotion mechanism and the impact of reputation in social network.

On the digg web site, once a new submission's digg count comes to a threshold up to 15, the submission is promoted to the front page of the web site. The promotion mechanism directs viewers' attention onto those few submissions. Our data shows that, while the promoted submission takes only up to 1.16 of total submissions, the overall attention accounted for promoted submissions is 2.5 times that of non-promoted ones. The mean of promoted submission is 955.9, while the mean of non-promoted submission is only 4.4. Hence, the growth of attention wealth is largely depending on the producer's chance to get promoted. It has been shown that the digg count of promoted story builds up quickly at first and eventually comes to saturate following a log-normal distribution [11]. However, it is still unclear about dominating factors rendering a submission getting promoted. One may list many factors that have influence on the chance for promotion, including quality of submissions, topics of interest, or the producer's network in the community. While it is true that all of these factors may have their impact on promotion, we argue that the producer's reputation or network is the dominating factor in determining each submission's chance of getting promoted. As a matter of fact, the information provided to the viewer, before the submission actually being viewed, is just the submission's title, digg count and producer's reputation. Sometimes, titles might be a leading factor for success, but all the producers are trying to cater viewers to gain more attention, making it very hard to outperform by a unique title. The variance of digg count before it actually getting promoted is also rarely noticeable. Based on these facts, we assume that the producer's reputation is the only factor influencing the chance for promotion, and that submissions are identical to each other, except for its producer's information. We also assume a uniform distribution of satisfaction level on viewers' side for simplicity.

In our model, we categorize viewers' attention into three parts: those directed by front page to promoted stories(promotion-generated attention); those generated by subscription in the form of producer's reputation(reputation-generated attention), and also those randomly paid to the pool of mass non-promoted submissions who later sink into the information sea(randomly-generated attention). Each promoted submission's digg count is constituted by these three parts. For those non-promoted stories, the digg count is constituted by the later two parts. As we are neglecting factors such as quality or topic of submissions in our model, a community like this without promotion and subscription functions would be no more than a system of identical publications, leaving viewers randomly to choose. We could view this system as constituted by identical, independent and distinguishable submissions engaging in different digg count or attention level. We define the total number of submissions in the system as N, the attention level or digg count in this case occupied by submissions as ε_i, number of submissions on attention level i as n_i, the total amount of attention in our third category as E. The system is then composed of N identical, independent, distinguishable submissions. Despite the many differences, the system here under our study is very similar to the free particle system being extensively studied. At any certain time point, we have:

$$N = \sum_{i=0}^{m} n_i \tag{1a}$$

$$E = \sum_{i=0}^{m} \varepsilon_i \cdot n_i \tag{1b}$$

, where $i = 0, 1, 2 \cdots m$, $\varepsilon_i = i$. If the system in our study is closed and isolated in quasi-static equilibrium state, the probability of total attention E distribute among N items in way n_i is proportional to:

$$\Omega = \frac{N!}{n_1! n_2! \cdots n_m!}$$

The entropy of the system could be defined as $S = Ln(\Omega)$. Since the system will behave in the most probable state, maximizing entropy condition combined with restrictions in Eqs. (1) give:

$$n_i = \exp(-\beta \cdot \varepsilon_i). \tag{2}$$

The partition could then be written as:

$$Z = \sum_{i=1}^{m} \exp(-\beta \cdot \varepsilon_i) \tag{3}$$

When m is large,

$$Z \to \frac{1}{1 - \exp(-\beta)} \tag{4}$$

Thus the probability of a submission's digg count equaling to i is,

$$\Pr(\varepsilon = \varepsilon_i) = \frac{\exp(-\beta \cdot \varepsilon_i)}{Z} = (1 - \exp(-\beta)) \times \exp(-\beta \cdot \varepsilon_i). \tag{5}$$

The probability for digg count larger than a specific integer X is,

$$\Pr(\varepsilon \geq X) = \frac{\sum_{i=X}^{m} \Pr(\varepsilon_i)}{Z} \approx \exp(-\beta \cdot X). \tag{6}$$

If X is set as the boundary criteria for promotion, in empirical digg web site where X equals to 15, then Eqs. (6) gives promotion probability of the identical submission system without taking social network into consideration.

A direct examine of above derivation is to select each user's first submission to see their distribution. As producer has no reputation for their first arrival, the digg count of first submissions would follow distribution predicted by Eqs. (5), if our model is correct. By choosing out all the first time submissions in our data, we get Fig. 3. The probability of promotion in ln scale varies linearly with view count, indicating $Ln(\Pr(x = x_i)) = \alpha - \beta \cdot x_i$. Linear fit gives $\alpha_1 = -1.005, \beta_1 = 0.384$. From Eqs. (5), we have $\exp(\alpha) = 1 - \exp(-\beta)$, and by taking β_1 into the equation, we could give another value of α, $\alpha_2 \approx -1.14$. α_2 and α_1 from the fitting result are approximately the same, which confirms our hypothesis and the model.

Now, we take into a producer's reputation effect into account. Given the uniform distribution of viewers' satisfaction level and identical submission assumptions, the probability of a view triggering subscription is constant, named as k_1. If a producer's attention wealth of digg count constituted by the first category and the third category is t_1^p and t_3^p respectively, with p denoting index of p_{th} submission, then one's number of subscribers according to our model would be $k_1 \cdot (T_1^p + T_3^p)$, where $T_1^p = \sum_{j=1}^{p} t_1^j$, $T_3^p = \sum_{j=1}^{p} t_3^j$. A simplified deduction would

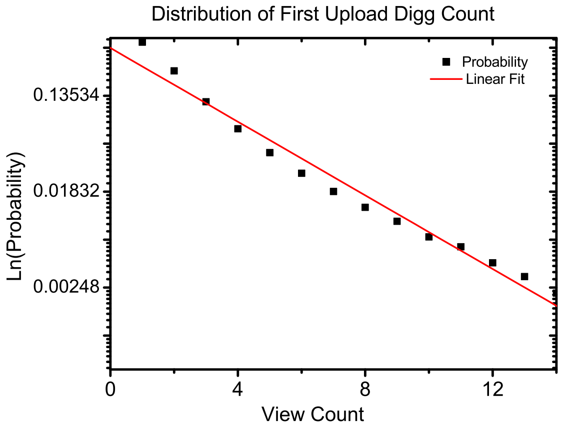

FIG. 3: Distribution of first upload follows the prediction of Eqs. (5). Linear fit gives $\beta_1 = 0.384$. α_1 obtained from linear fit and α_2 from calculation coherent with each other.

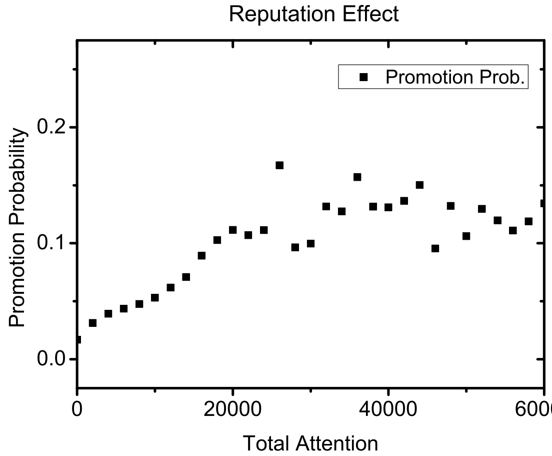

FIG. 4: Promotion probability increases monotonously with producer's reputation measured by attention wealth. The promotion probability comes to saturated at 0.15.

assume those subscribers' react to subscribed new publications with a constant chance of visiting, defined as k_2. Thus a producer with $k_1 \cdot (T_1^p + T_3^p)$ subscribers would have a $t_2^{p+1} = k_1 \cdot k_2 \cdot (T_1^p + T_3^p)$ increase of digg count in following upload due to previous achievements in reputation or network. As k_1 and k_2 are both small($k_1 \cdot k_2 \sim 10^{-4}$) as later shown, we could find that: t_2^{p+1}

$$= k_1 \cdot k_2 \cdot (T_1^p + T_3^p) = k_1 \cdot k_2 \cdot (T^p - t_2^p - t_2^{p-1} - \cdots - t_2^2)$$
$$= k_1 \cdot k_2 \cdot T^p - k_1 \cdot k_2 \cdot k_1 \cdot k_2 \cdot \sum_{j=1}^{p-1} (t_1^j + t_3^j) \approx k_1 \cdot k_2 \cdot T^p, \tag{7}$$

where T^p is the total attention wealth after p_{th} upload. So a new submission from a producer with total attention wealth T will have a digg count $x + k_1 \cdot k_2 \cdot T$, x still following distribution in Eqs. (5). And a new submission's promotion probability becomes:

$$\begin{aligned}\Pr((x + k_1 \cdot k_2 \cdot T) > X) &= \Pr(x > (X - k_1 \cdot k_2 \cdot n)) \\ &= \exp(-\beta \cdot (X - k_1 \cdot k_2 \cdot n))\end{aligned} \tag{8}$$

We can gather out producers on different attention wealth level's promotion probability against their attention wealth level as shown in Fig. 4. The promotion probability increases monotonously with attention wealth, and tends to saturate when attention wealth comes to above 20,000. The saturation can be explained by the fact that the number of viewers in the digg community is finite, and also the number of viewers using the subscription function.

Since over 90% of producers' within the community is in the region of low level, which is in accordance with previous findings [15], we zoom in our scale to see how the promotion probability increases with attention wealth, as

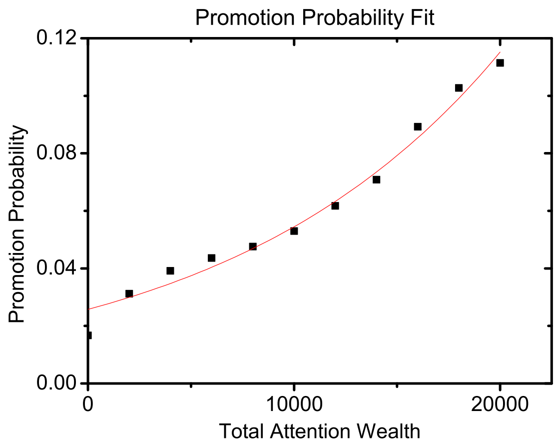

FIG. 5: Exponential fit of the promotion probability against reputation in the increasing region. The fit gives $a = 0.02576$, $b = 0.00007$, from $Pr. = a \cdot exp^{b \cdot T}$.

shown in Fig. 5. Data fits into exponential growth, giving $\beta_3 = 0.2439$, $k_1 \cdot k_2 = 2.87 \times 10^{-4}$, when $X = 15$. β_3 is approximately equal to the β_1 and β_2 obtained from first time submissions. We will use β_3 for future analysis.

With the above model, we could now predict the digg count distribution of a producer's following submission, given that the producer is now on attention wealth level T. We have,

$$\begin{aligned}\Pr(\Delta T = \varepsilon_i | T) = (1 - \lambda)(1 - \exp(-\beta)) \times \exp(-\beta \cdot \varepsilon_i) \\ + \lambda \cdot \frac{1}{\varepsilon_i \cdot \sqrt{2\pi\sigma^2}} \exp(-\frac{(\ln \varepsilon_i - r_0)^2}{2\sigma^2})\end{aligned} \tag{9a}$$

$$\lambda = \exp(-\beta(X - k_1 \cdot k_2 \cdot T)), \tag{9b}$$

where λ is the promotion probability, r_0 and σ are the mean and standard deviation of the promoted story's natural logarithm [11]. Individual producer's attention wealth is then given as:

$$T = \sum \Delta T. \tag{10}$$

This dynamic model gives the dynamic growth process of producer's attention wealth. It can also interpret the existence of a peak at the tail of attention wealth distribution. Noticing that the peak is in the region of 10^5, which is also where the reputation starts to make a difference on promotion probability, we believe that, the peak is caused by social network of the digg community, or the reputation of those established producers. Their reputation in the network tends to help them draw more attention in the following uploads, as well as wealth accumulation.

A direct proof of above proposition's correctness is to simulate attention wealth growth process based on the dynamic model. We simulate the growth of 318,131 producers by generating each publication's digg count obeys Eqs. 9. The submission count of each producer is chosen to be of the same distribution as in real digg community. The promotion probability λ of each producer's next upload is set according to one's current attention wealth level. The simulated distribution of attention wealth is shown in Fig. 6. It has the same scale and characteristics as we observed in real data with a peak near the tail. One benefit for such simulation is that, we can turn off the impact of reputation by setting $\lambda = const.$. Fig. 7 shows the simulated result of attention wealth distribution, when λ is fixed as 0.015. It is obvious that the peak disappears when we are turning off the network, which confirms our hypothesis that the aggregation of wealthy people is resulted from reputation in social network.

Besides simulating the growth process of attention wealth, our result could also predict income distribution for certain attention wealth level with the same uploading frequency. To present this, we pick out the month income information of 96 producers in the attention wealth level between 18000 and 20000 in April. According to our model when n=19000, the income distribution would follow a bimodal distribution, with weighting factor $\lambda_1 = 0.8906$, and $\lambda_2 = 0.1094$.

$$\begin{cases} \Pr(\varepsilon = \varepsilon_i | T) = \lambda_1 \cdot \Pr_1(\varepsilon_i) + \lambda_2 \cdot \Pr_2(\varepsilon_i) \\ \Pr_1(\varepsilon_i) = (1 - \exp(-\beta)) \times \exp(-\beta \cdot \varepsilon_i) \\ \Pr_2(\varepsilon_i) = \frac{1}{\varepsilon_i \sqrt{2\pi\sigma^2}} \exp(-\frac{(\ln \varepsilon_i - r_0)^2}{2\sigma^2}) \\ \lambda_2 = \exp(-\beta(X - k_1 \cdot k_2 \cdot T)) \\ \lambda_1 + \lambda_2 = 1 \end{cases} \tag{11}$$

Again, comparing with real income distribution gathered from digg, we use Expectation-Maximization(EM) algorithm for a bimodal fit to the 96 producer's income. The

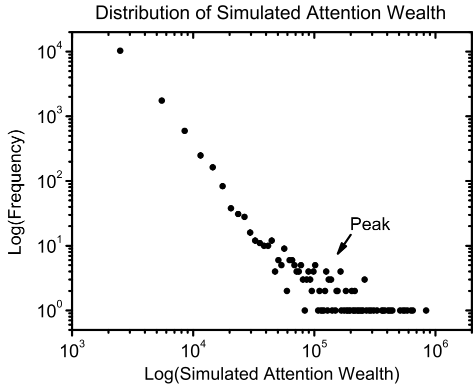

FIG. 6: Simulated attention wealth distribution using our dynamic growth model. The scale of the plot is similar to observations from real digg web site. Peak near the tail indicates aggregation of wealthy people in attention economy.

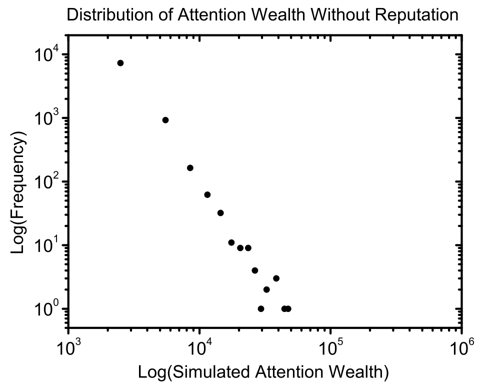

FIG. 7: Simulated attention wealth distribution without reputation effect by setting promotion probability $\lambda = 0.015$. Peak near the tail disappears, which confirms our proposition that the aggregation of wealthy people is due to reputation effect.

EM results give $\lambda_1 = 0.9039$, and $\lambda_2 = 0.0961$. Other wealth level's income data also fits well with predictions.

Another potential application of our model is to look at the persistence paradox proposed by Fang et, al. The study shows that, while the average quality of submissions increases with the index number of uploads, the more frequently an individual uploads content, the less likely the producer makes a success [16]. The definition of success could vary, but similar results appear. We repeat the work by defining success as getting promoted to front page, which is of importance in gaining attention. The conditional probability of success $h(k)$ is defined as the number of promoted stories, whose previous $k-1$ submissions are non-promoted, divided by number of submissions whose previous $k-1$ submissions are all

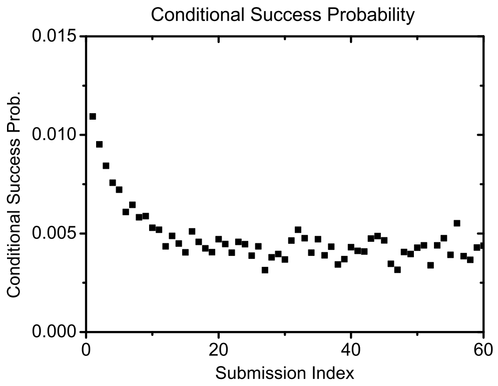

FIG. 8: Conditional success probability decreases with index number.

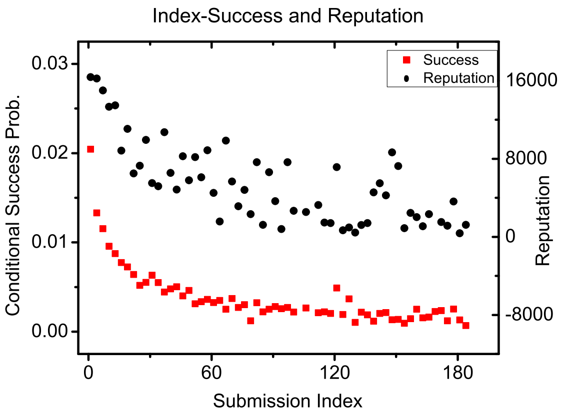

FIG. 9: The conditional success probability decreases in the same trend with reputation.

non-promoted. As shown in Fig.8, the conditional success probability decreases, despite of the increasing index number, along with the possible improvements in quality. We try to give one possible explanation based on reputation effect. The plot in Fig. 9 demonstrates that, while it is true the conditional success probability decreases with index number, the reputation, which is quantified by attention wealth, is also decreasing along with index number. So one possible explanation for the persistence paradox might be that people are losing not because they don't make enough efforts or improvements, but for that they don't have a good social network. The frequent successors are more likely to outperform, as they have a higher reputation level. If this is true, it is also the condition when quality grows with index number, the reputation level decreases in opposite way, and the chance of success is following the trend of reputation, rather than quality.

V. DISCUSSION AND CONCLUSION

In this paper, we treat attention paid to each item as an aggregative wealth for each producer. By summing up all items' attention owned by certain producer, we demonstrate that the attention wealth distribution deviates from power-law with a peak near the tail, indicating an aggregation of wealthy people. To interpret this, we further investigate the impact of reputation in the social network to promotion probability. A model for attention wealth growth process is proposed and simulated to compare with observations. The disappearing peak after we turn off network in simulation suggests that the aggregation in wealthy region is due to reputation. Useful predictions like income distribution are also examined based on our model. We believe that, many emerging topics of interest can be explored if applying this reputation model in a broader category of attention economy.

ACKNOWLEDGEMENTS

[1] T.H. Davenport, and J.C. Beck, *The Attention Economy: Understanding the New Currency of Business*, (Harvard Business School Press, Cambridge, 2001).
[2] C. Shapiro, H. Varian, *Information Rules: A Strategic Guide to the Network Economy*, (Harvard Business School Press, Boston, 1999).
[3] G. Franck, Science Communication,(53-55), 286 (1999).
[4] M.H. Goldhaber, First Monday, 2 (1998).
[5] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan and R. Stata et al., Computer Networks, 33 (2000).
[6] B.A. Huberman, P. Pirolli, J.E. Pitkow, and R.M. Lukose, Science, (95), 280 (1998).
[7] B.A. Huberman and L.A. Adamic., Nature, (130), 399 (1999).
[8] B.A. Huberman, Computer, (11), 41 (2008).
[9] K.M. Rezaul and V.A. Grout, Proceedings of the second International Joint e-Conference on Computer, Information, and Systems Sciences, and Engineering (2006).
[10] D.M. Wilkinson, ACM Conference on Electronic Commerce (2008).
[11] Fang Wu and B.A. Huberman, Proc. Natl. Acad. Sci., /textbf(105), 17599 (2007).
[12] Fang Wu and B.A. Huberman, IEEE SocialCom '09 (2009).
[13] How Digg Works. http://www.digg.com/how
[14] Fang Wu and B.A. Huberman, ACM E-Commerce '08 (2008).
[15] A. Klamer and H.P. Dalen, Journal of Economic Methodology, /textbf(9), 289 (2002).
[16] Fang Wu, B.A. Huberman, First Monday, 15 (2010).