Size Does Matter: Why Knowledge Markets Can Fail at Scale!

Himel Dev[1], Chase Geigle[1], Qingtao Hu[1], Jiahui Zheng[2] and Hari Sundaram[1]
[1]University of Illinois at Urbana-Champaign
[2]Peking University
hdev3@illinois.edu

ABSTRACT

In this paper, we model the community question answering (CQA) websites on Stack Exchange platform as knowledge markets, and analyze how and why these markets can fail at scale. Analyzing CQA websites as markets allows site operators to reason about the failures in knowledge markets, and design policies to prevent these failures. Our main contribution is to provide insight on knowledge market failures. We explore a set of interpretable economic production models to capture content generation dynamics in knowledge markets. The best performing of these, well-known in economic literature as Cobb-Douglas equation, provides an intuitive explanation for content generation in the knowledge markets. Specifically, it shows that (1) factors of content generation such as user participation and content dependency have *constant elasticity*—a percentage increase in any of the inputs leads to a constant percentage increase in the output, (2) in many markets, factors exhibit *diminishing returns*—the incremental, marginal output decreases as the input is incrementally increased, (3) markets vary according to their *returns to scale*—the increase in output resulting from a proportionate increase in all inputs, and finally (4) many markets exhibit *diseconomies of scale*—measures of market health decrease as a function of overall system size (number of participants).

ACM Reference format:
Himel Dev[1], Chase Geigle[1], Qingtao Hu[1], Jiahui Zheng[2] and Hari Sundaram[1]. 2017. Size Does Matter: Why Knowledge Markets Can Fail at Scale!. In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17),* 10 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In this paper, we analyze a large group of community question answering (CQA) websites on Stack Exchange network through the Economic lens of a market. Framing Stack Exchange websites as knowledge markets has intuitive appeal: in a hypothetical knowledge market, if no one wants to answer questions, but only ask, or conversely, there are individuals who want to only answer but not ask questions, the "market" will collapse. What then, is the relationship among actions (say between questions and answers) in such knowledge markets, for us to deem it healthy? Are larger markets with more participants healthier since there will be more people to ask and answer questions?

Studying CQA websites through an Economic lens allows site operators to reason about whether they should grow the user base. Since most of the popular CQA websites (e.g. Quora, Stack Exchange) do not charge participants, but instead depend on site advertisements for revenue, there is a natural temptation for operators of these websites to grow the user base, so that there is increase in revenue. As we show in this paper, for most Stack Exchange websites, growth in the user base is counter-productive in the sense that they turn unhealthy—specifically, more questions remain unanswered.

Explaining the macroscopic behavior of knowledge markets is important, yet challenging. One can regress some variable of interest (say number of questions) on variables including number of users, time spent in the website among others. However, explaining why the regression curve looks like the way it does is hard. As we show in this work, using an Economic lens of a market allows us to model dependencies between number of participants, and the amount of content, and to predict the production of content.

Our main contribution is to model CQA websites as knowledge markets, and to provide insight on the relationship between size and health of these markets. To this end, we develop models to capture content generation dynamics in knowledge markets. We analyze a set of basis functions (the function form of how an input contributes to output) and interaction mechanisms (how the inputs interact with each other), and identify the optimal *power basis* function and the *interactive essential* interaction form using a prediction task on the outputs (questions, answers and comments). This form, is the well-known Cobb-Douglas form that connects production inputs with output. Using the best model fits for each Stack Exchange, we show that the Cobb-Douglas model predicts the production of content with high accuracy.

The Cobb-Douglas function provides intuitive explanation for content generation in Stack Exchange markets. It demonstrates that in Stack Exchange markets— (1) factors such as user participation and content dependency have *constant elasticity*—percentage increase in any of these inputs will have constant percentage increase in output; (2) in many markets, factors exhibit *diminishing returns*—decrease in the marginal (incremental) output (e.g., answer production) as an input (e.g. number of people who answer) is incrementally increased, keeping the other inputs constant; (3) markets vary according to their *returns to scale*—the increase in output resulting from a proportionate increase in all inputs; and (4) many markets exhibit *diseconomies of scale*—measures of health decrease as a function of overall system size (number of participants).

There are two reasons why we see diminishing returns in the Stack Exchange markets. First, the total activity of participants for any Stack Exchange, unsurprisingly follows a power-law pattern.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

What is interesting is that the power law exponent falls with increase in size for most Stack Exchanges, implying that new users do not participate in the same manner as earlier users. Second, we can identify a stable core of users, who continue to actively participate for long periods of time, contributing to the network health.

Finally, we show diseconomies of scale through experiments on system size, analysis of health metrics and user exchangeability. For most Stack Exchanges, we see that as system size grows, the ratio of answers to questions falls below critical point, when some questions go unanswered. Furthermore, using health metrics of number of questions with an accepted answer, and number of questions with at least one answer, we observe that most Stack Exchanges decline in health with increase in size. Finally, we compare the top contributors with the bottom contributors to see if they are "exchangeable." Most Stack Exchanges are not exchangeable in the sense the contributions of the top and the bottom contributors are qualitatively different and differ in absolute terms. These experiments on diseconomies of scale are consistent with the insight from Cobb-Douglas model of production that predicts diminishing returns.

2 RELATED WORK

Our work draws from, and improves upon, several research threads.

Sustainability. There is a relatively recent body of work studying sustainability in knowledge markets. Notably, Srba and Bielikova [16] conducted a case study on why StackOverflow, the largest and oldest of the networks on StackExchange, is failing. They reveal some insights into market failure such as novice and negligent users generating low quality content perpetuating the decline of the network. However, they do not present a systematic way to understand and prevent failures in knowledge markets. Wu et al. [20] introduced a framework for understanding the user strategies in a knowledge market—revealing the importance of diverse user strategies for sustainable markets.

Activity Dynamics. There have been a number of papers on modeling the activity dynamics of online platforms [2, 19, 21]. Walk et al. [19] modeled user-level (micro-scale) activity dynamics in StackExchange using two factors: intrinsic activity decay, and positive peer influence. However, the model proposed there can not be used to study content-driven success or failure of online platforms because (1) it does not reveal insights on the collective platform dynamics, and (2) it does not concentrate on the eventual success or failure of a platform. Wu and Zhang [21] proposed a discrete generalized beta distribution (DGBD) model that reveals several insights into the collective dynamics, notably the concept of a size dependent distribution. In this paper, we improve upon the concept of a size dependent distribution.

Scale Study. Lin et al. [12] examined Reddit communities to characterize the effect of user growth in voting patterns, linguistic patterns, and community network patterns. The study reveals that these patterns do not change much after a massive growth in the size of the user community. In this paper, we examine the consequence of scale on knowledge markets from a different perspective by using a set of health and stability metrics.

Stability. Successes and failures of networks have been studied from the perspective of stability. Patil et al. [14] studied the dynamics of group stability in social networks. They define stability based on the average increase or decrease in member growth. Our paper examines stability in a different manner—namely, by considering the relative exchangability of users as a function of scale.

User Growth. Successes and failures of networks have also been widely studied from the perspective of user growth [3, 9, 11, 23]. Notably, Backstrom et al. [3] studied the mechanisms that underpin of how users join communities in a social network; Kairam et al. [9] examined diffusion (growth via social ties) and non-diffusion (growth without social ties) processes to design models that predict the longevity of social groups. These works, however, do not model active user growth. To capture active user growth, Ribeiro [15] proposed a daily active user (DAU) prediction model for membership based websites; the model classifies membership based websites as sustainable and unsustainable. While this perspective is important, we argue that the success and failure of networks based on their *content production* can perhaps be more meaningful.

Content Generation. Analyzing patterns of user content generation is crucial for developing principled content generation models. Guo et al. [7] analyzed three social networks and revealed the stretched exponential distribution of user contribution. In this paper, we argue that the distribution of user contribution is size-dependent in CQA networks.

Modeling CQA Websites. Furtado et al. [5] explore user behavior profiles and their dynamics in five StackExchange networks by performing an agglomerative clustering on manually specified user attributes. This can be viewed as a providing an understanding of behavior dynamics at a "micro" level (at the level of individual users). A major difference of our work from theirs is that in this paper we take a "macro" view of the behavior dynamics of CQA networks by looking at the behavior of an entire network as a function of its user population. Kumar et al. [10] proposes taking an economic view of CQA networks, like we do, but under their formulation users must be on one-and-only-one side of the two-sided market. Our model differs in that it does not make any assumption about the presence or absence of overlap in the group of users that provide answers and the group of users that provide questions. Furthermore, their model does not provide a systematic way of understanding the amount of *content* the market generates. Rather, they focus on the growth of the two kinds of users. Yang et al. [22] identifies the scalability problem of CQA networks that we study here—namely, the volume of question content eventually subsumes the capacity of the answerers within the community. Understanding and modeling this phenomenon is one of the goals of this paper.

3 PROBLEM FORMULATION

The goal of this paper is to develop a model for content generation in knowledge markets. Content is integral to the success and failure of a knowledge market. Therefore, we aim to better understand the content generation dynamics.

A model for content dynamics should have the following properties: macro-scale, explanatory, predictive, minimalistic, comprehensive.

Macro-scale: The model should capture content generation dynamics via aggregate measures. Aggregate measures help us understand the collective market by summarizing a complex array of information about individuals, which is especially important for policy-making.

Explanatory: The model should be insightful about the behavior of a knowledge market. Understanding market behavior is a crucial first step in designing policies to maintain a resilient, sustainable market.

Predictive: The model should allow us to make predictions about future content generation and resultant success or failure. These market predictions are integral to the prevention and mitigation of market failures.

Minimalistic: The model should have as few parameters as necessary, and still closely reflect the observed reality.

Comprehensive: The model should encompass content generation dynamics for different content types (e.g., question, answer, comment) in varieties of knowledge markets. This is important for developing a systematic way to understand the successes and failures of knowledge market.

In remaining sections we propose models that meet the aforementioned requirements, and show that our best-fit model accurately reflects the content generation dynamics and resultant successes and failures of real-world knowledge markets.

4 MODELING KNOWLEDGE MARKETS

In this section we introduce economic production models to capture content generation dynamics in real-world knowledge markets. We first draw an analogy between economic production and content generation, and report the content generation factors in knowledge markets (Section 4.1). Then, we concentrate on the knowledge markets in Stack Exchange—presenting production models for different content types (Section 4.2).

4.1 Preliminaries

Economic production mechanisms well describe content generation in knowledge markets. In Economics, *production* is defined as the process by which human labor is applied, usually with the help of tools and other forms of capital, to produce useful goods or services—the *output* [17]. We assert that participants of a knowledge market function as labor to generate content such as questions and answers. Analogous to economic output, content contributes to participant utility.

Motivated by the production analogy, we design macroeconomic production models to capture content generation dynamics in knowledge markets. In these models, instead of directly modeling content generation as a dynamic process (function of time), we model it in terms of associated factors which are dynamic.

There are two key factors that affect content generation in knowledge markets, namely user participation and content dependency. User participation is the most important factor in deciding the quantity of generated content. The participation of more users induce more questions, answers, and other contents in a knowledge market. Content dependency also affects the quantity of generated content for different types. Content dependency refers to the dependency of one type of content (e.g., answers) on other type of content (e.g., questions). In absence of questions, there will be no answers in a knowledge market, even in the presence of many potential participants who are willing to answer.

4.2 Modeling Stack Exchange

Stack Exchange is a network of community question answering websites where each site is based on a focused topic. Each user of Stack Exchange network participates in one or more of these sites based on their interest. Stack Exchange sites are free knowledge markets where participants generate content for non-monetary reputation-based incentives. These markets are diverse, varying in theme (subject matter), size (number of users and amount of activity), and age (number of days in existence).

We design production models for three primary content types in Stack Exchange: questions (the root content), answers (which nest below questions), and comments (which can nest either beneath questions or answers). Based on the content dependency and user roles in content generation, we propose the following relationships for question, answer and comment generation in Stack Exchange (See Table 1 for notation).

Table 1: Notations used in the model

Symbol	Definition
$U_q(t)$	# of users who asked questions at time t
$U_a(t)$	# of users who answered questions at time t
$U_c(t)$	# of users who made comments at time t
$N_q(t)$	# of active questions at time t
$N_a(t)$	# of answers to active questions at time t
$N_c^q(t)$	# of comments to active questions at time t
$N_c^a(t)$	# of comments to active answers at time t
$N_c(t)$	# of comments to active questions/answers at time t
f_x	The functional relationship for content type x

There is a single factor in generating N_q questions: the number of users U_q who ask questions (askers).

$$N_q = f_q(U_q)$$

There are two key factors in generating N_a answers: the number of questions N_q, and the number of users U_a who answer questions (answerers).

$$N_a = f_a(N_q, U_a)$$

There are two types of comments: comments on questions, and comments on answers. Accordingly, there are three key factors in generating N_c comments: the number of questions N_q, the number of answers N_a, and the number of users U_c who make comments (commenters).

$$N_c^q = f_{c^q}(N_q, U_c)$$
$$N_c^a = f_{c^a}(N_a, U_c)$$
$$N_c = N_c^q + N_c^a$$

The aforementioned relationships imply that the amount of generated content of each type depends on the function describing its factor dependent growth, and the availability of factor(s). These relationships make three assumptions. First, different content types interact only through their use of factors. Second, the functional relationships depend on the consumption or usage of each factor.

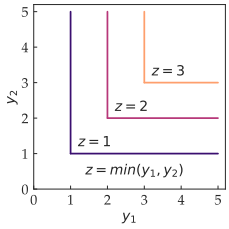
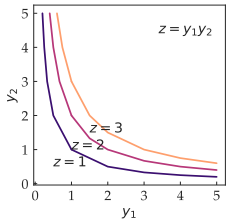
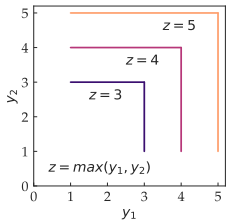
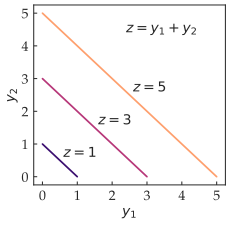
Third, the functional relationships depend on the interaction among the factors—how the factors of a particular content type interact.

Now, we transform the functional relationships into production models by first choosing a basis function to capture how a content type consumes its factor(s), and then choosing an interaction type to capture the interaction among factors.

Basis Function. We use a basis function to capture the effect of a given factor on a particular content type. While there is a variety of basis functions available for regression, we consider three basis functions widely used in economics and growth modeling [4]: power– $g(x) = ax^{\lambda}$; exponential– $g(x) = ab^{x}$; and sigmoid– $g(x) = \frac{L}{1+e^{k(x-x_0)}}$.

Interaction among the Factors. We use an aggregate function to capture the interaction among multiple factors of a given content type. Specifically, we consider the pairwise interaction functions listed in Table 2.

Table 2: Pairwise interaction between factors with contour

Essential: Essential factors are both required for content generation, with zero marginal return for a single factor. For a pair of essential factors, content generation is determined by the more limiting factor: $z = min(y_1, y_2)$ [18]. This is known as Liebig's law of the minimum.	
Interactive Essential: In interactive essential interaction, we get diminishing return (instead of zero return) for a single factor: $z = y_1 y_2$ [18]. If factors are consumed using power basis function, i.e., $y_i = ax_i^{\lambda}$, it captures Cobb-Douglas production function.	
Antagonistic: For antagonistic factors, content generation is determined solely by the availability of the factor which yields the largest return: $z = max(y_1, y_2)$ [18]. This interaction implies that the production process has maximum possible efficiency.	
Substitutable: Factors that can each support production on their own are substitutable relative to each other: $z = w_1 y_1 + w_2 y_2$ [18]. This implies that there exists some equivalence between the two factors. This is analogous to the general additive models.	

We combine basis function and interaction type to design production models for different content types. For example, answer generation can be modeled using power basis and essential interaction as $N_a = min(A_1 N_q^{\lambda_1}, A_2 U_a^{\lambda_2})$. We consider twelve such possible models (combination of three basis and four interaction type) for answer and comment generation in Stack Exchange.

User Role Distribution. A fundamental assumption of our model is the awareness of user roles (e.g., asker, answerer, and commenter) and their distribution (e.g., how many users are askers?). We empirically observe that all Stack Exchange markets have a stable distribution of user roles. In fact, given the number of users, we can accurately predict the number of participants of a particular role.

We apply linear regression to determine the number of participants U_x of a particular role $x \in \{q, a, c\}$, from the number of users U in a Stack Exchange market. For each market, we compute three distinct coefficient of determination (R^2) for predicting three roles (asker, answerer, and commenter) using linear regression. In Figure 1 we show the distribution of R^2 for regressing user roles across 156 Stack Exchange markets. We use letter value plots[1] to present these distributions—showing precise estimates of their tail behavior. We observe that, in most markets, the R^2 values are close to 1. Further, the tail capturing low R^2 values consists of markets with relatively small number of monthly users (<200) .

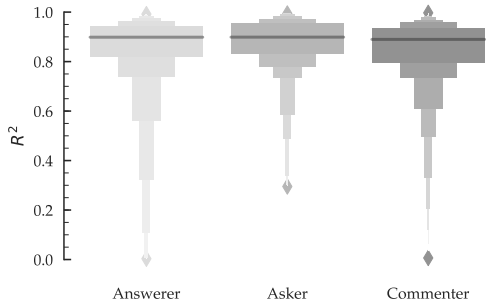


Figure 1: The R^2 distribution for regressing user roles (asker, answerer, and commenter) across 156 Stack Exchange. In most Stack Exchange, the role distribution is stable—as manifested by the high R^2 in letter value plots.

Number of Users. The number of users is the only free input to our content generation models; the remaining inputs are functions of number of users. In all these models, the growth or decline of number of users is exogenous—determined outside the model, by non-economic forces. There is a growing body of literature that concentrate on modeling active user growth [15, 23]. In this paper, we build upon these works to concentrate on a subsequent problem—the relation between user growth and content generation.

5 DATASET

We collected the latest release (September, 2017) of Stack Exchange dataset. This snapshot is a complete archive of all activities in all Stack Exchange. There are 169 sites in our collected dataset. For the purpose of empirical analysis, we only consider the sites that have been active for at least 12 months beyond the ramp up period (site created, but few or no activity). There are 156 such sites. The age of these sites vary from 14 months to 111 months, number of user from 1072 to 547175, number of posts (questions and answers) from 1600 to 1985869. Further, the sites have small overlaps in user base;

[1]The letter-value plot display information about the distribution of a variable [8]. It conveys precise estimates of tail behavior using letter values; boxplots lack such precise estimation.

therefore, we can reasonably argue that the underlying markets are independent.

In Figure 2 we present letter value plots (in log-scale) to show the distribution of number of users, number of posts, and age for the Stack Exchange markets.

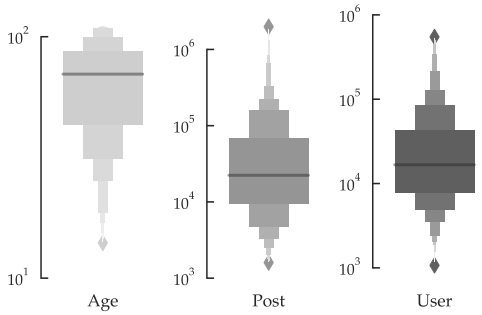


Figure 2: The distribution of number of users, number of posts, and age for Stack Exchange markets (in log-scale). The markets vary in all three dimensions.

6 EVALUATING OUR PROPOSED MODELS

In this section we identify optimal models (basis and interaction) from three different perspectives: the accuracy of fitting content generation time series observed in our dataset (Section 6.1), the performance of predicting content volume in long run (Section 6.2), and the perplexity of characterizing content generation dynamics at early stage (Section 6.3).

6.1 Model Fitting

We fit each variant of production model (basis and interaction) for each content type to the observed content generation time series (monthly granularity) in each Stack Exchange. Notice that among the different variants of production models, the models using power or exponent basis have a parsimonious set of parameters. For example, answer generation model using power basis function requires only three parameters for interactive essential interaction (See Section 4.2), and four parameters for remaining interaction types. In contrast, answer generation model using sigmoid basis function requires five parameters for interactive essential interaction, and six parameters for remaining interaction types.

Parameter Estimation. Our parameter learning process has three sequential steps enforced by the content dependency among question, answer and comment; we first learn the best-fit parameters for modeling question, followed by answer, followed by comment. At each step, we use the parameters learnt in earlier steps to generate input.

We restrict some parameters of our production models to be non-negative, e.g., non-negative exponents in power basis. These restrictions are important because the underlying factors positively affect the output. We use the trust-region reflective algorithm to solve our constrained least square optimization problem. The algorithm is appropriate for solving non-linear least squares problems with constraints.

Evaluation Method. We evaluate the overall fitting accuracy using four metrics: root mean square error (RMSE), normalized root mean square Error (NRMSE), explained variance score (EVS), and Akaike information criterion (AIC). Given two series, the observed series for content w, $N_w(t)$, and the prediction $\hat{N_w}(t)$ of the series by a model with k parameters, we compute these metrics as follows: $\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(N_w(t) - \hat{N_w}(t))^2}$; $\text{NRMSE} = \frac{RMSE}{max(N_w(t)) - min(N_w(t))}$; $\text{EVS} = 1 - \frac{Var(N_w(t) - \hat{N_w}(t))}{Var(N_w(t))}$; $\text{AIC} = T * ln(\frac{1}{T}\sum_{t=1}^{T}(N_w(t) - \hat{N_w}(t))^2) + 2k$. Among the four metrics, RMSE and NRMSE are error metrics (low value implies good fit), AIC is an information theoretic metric to capture the trade-off between model complexity and goodness-of-fit (low value implies good model), and EVS refers to a model's ability to capture variance in data (high value implies good model). All four metrics are consistent with the non-linear least squares problem.

Fitting Results. Now, we compare the fitting accuracy of production models for all markets in Stack Exchange, using the four metrics, each summarized via mean across all Stack Exchange, for each content type. We use content generation time series with monthly granularity as observed data. We found that the models with exponential and sigmoid basis do not fit the data for many Stack Exchange. Accordingly, in Table 3, we only present the results for production models with power basis and different interaction types. Notice that the models with interactive essential interaction outperform the remaining models for all metrics and content types. We performed paired t-tests to determine if the improvements for interactive essential interaction are statistically significant; the results are positive with $p < 0.01$.

Table 3: The comparison of fitting accuracy of production models (with power basis and different interaction types) for all Stack Exchange. The models with interactive essential interaction outperform the remaining models for all metrics and content types. The improvements for interactive essential interaction are statistically significant—validated via paired t-tests ($p < 0.01$).

Content	Interaction Type	Avg. RMSE	Avg. NRMSE	Avg. EVS	Avg. AIC
Question	Single Factor	25.742	0.086	0.791	104.473
Answer	Essential	70.307	0.092	0.789	208.820
	I. Essential	**64.624**	**0.083**	**0.825**	**196.395**
	Antagonistic	72.765	0.094	0.778	210.958
	Substitutable	68.900	0.089	0.805	207.609
Comment	Essential	146.644	0.084	0.833	328.245
	I. Essential	**137.228**	**0.081**	**0.845**	**318.243**
	Antagonistic	155.969	0.088	0.818	334.118
	Substitutable	155.433	0.089	0.820	335.102

Thus we use production models with power basis and interactive essential interaction for prediction tasks.

6.2 Forecasting Content Generation

We apply production models with power basis and interactive essential interaction to forecast content volume in long run—one year ahead in future. Specifically, we train each model using the content generation data from first 12 months (beyond the ramp period), and then examine how well the model forecasts content

dynamics in next 12 months. We validate the forecasting capability by examining the overall prediction error (NRMSE) into the future.

We compute the prediction NRMSE across all Stack Exchange, and summarize the results using mean (μ) and variance (σ)— (i) question: $\mu = 0.11$, $\sigma = 0.09$; (ii) answer: $\mu = 0.12$, $\sigma = 0.09$; (iii) comments: $\mu = 0.11$, $\sigma = 0.19$. Notice that our models can forecast future content dynamics with high accuracy. We performed these experiments for different time granularity, e.g., week, month, quarter, and found consistent conclusion. We do not report these results for brevity.

6.3 Parameter Estimation for New Markets

To better predict the success or failure of young markets (6-12 months old), we use model parameters learnt from mature markets (at least 36 months old) as informative priors for new markets. We incorporate these informative priors using Bayesian parameter estimation technique. We validate the parameter estimation capability by first training models using the first 6 month's data from young market, and the informative priors based on the nearest (based on user size) mature market, and then examining how well the model forecasts content dynamics in remaining months.

We compute the prediction NRMSE across all young Stack Exchange, and summarize the results using mean (μ) and variance (σ)— (i) question: $\mu = 0.10$, $\sigma = 0.12$; (ii) answer: $\mu = 0.09$, $\sigma = 0.08$; (iii) comments: $\mu = 0.14$, $\sigma = 0.13$.

7 CHARACTERIZING KNOWLEDGE MARKETS

In this section we characterize the knowledge markets in Stack Exchange—explaining the best-fit models and their foundations (Section 7.1), revealing two key distributions that control the markets (Section 7.2), and uncovering the stable core that maintains market equilibrium (Section 7.3).

7.1 Model Interpretation

First, we explain the best-fit models found in Section 6.1. We observe that content generation in Stack Exchange markets are best modeled through the combination of power basis and interactive essential interaction. In addition, we found that the best-fit exponents (λ parameter in basis $g(x) = ax^{\lambda}$, where x is a factor) of these models lie between 0 and 1 (inclusive), for all factors of all content types, for all Stack Exchange.

A model that uses power basis (where exponents lie between 0 and 1) and interactive essential interaction is known as the Cobb-Douglas production function [6]. In its most standard form for production of a single output z with two inputs x_1 and x_2, the function is:

$$z = ax_1^{\lambda_1} x_2^{\lambda_2}.$$

Here, coefficient a represents the *total factor productivity*—the portion of output not explained by the amount of inputs used in production [6]. As such, its level is determined by how efficiently the inputs are utilized in production. The exponents λs represent the *output elasticity* of the inputs—the percentage change in output that results from the percentage change in a particular input [6].

The Cobb-Douglas function provides intuitive explanation for content generation in Stack Exchange markets. In particular, the explanation stands on three phenomena or principles: constant elasticity, diminishing returns, and returns to scale.

Constant Elasticity. In Stack Exchange markets, factors such as user participation and content dependency have *constant elasticity*—percentage increase in any of these inputs will have constant percentage increase in output [6], as claimed by the corresponding exponents in the model. For example, in `academia` ($N_A = 6.93 N_q^{0.18} U_a^{0.65}$), 1% increase in number of answerers (U_a) leads to 0.65% increase in number of answers (N_a).

Diminishing Returns. For a particular factor, when the exponent is less than 1, we observe *diminishing returns*—decrease in the marginal (incremental) output as an input is incrementally increased, while the other inputs are kept constant [6]. This 'law of diminishing returns' has many interesting implications for the Stack Exchange markets, including the diminishing benefit of having a new participant in a market. For example, in `academia`, if the number of answerers is 100, then the marginal contribution of a new answerer is $c(101^{0.65} - 100^{0.65}) = 0.129c$, where c is a constant; in contrast, if the number of answerers is 110, then the marginal contribution of a new answerer is $c(111^{0.65} - 110^{0.65}) = 0.125c$. Thus, for answer generation in `academia`, including a participant when the number of participants (system size) is 110 is likely to be less beneficial compared to including a participant when the system size is 100.

Returns to scale. The knowledge markets in Stack Exchange vary in terms of scale efficiency, as manifested by their *returns to scale*—the increase in output resulting from a proportionate increase in all inputs [6]. If a market has high returns to scale, then greater efficiency is obtained as the market moves from small- to large-scale operations. For example, in `academia`, for answer generation, the returns to scale is $0.18 + 0.65 = 0.83 < 1$. The market becomes less efficient as answer generation is expanded, requiring more questions and answerers to increase the number of answers by same amount.

7.2 Two Key Distributions

Next, we discuss two key distributions that control content generation in knowledge markets, namely participant activity and subject POV (perspective). These two distributions induce the three phenomena reported in section 7.1.

Participant Activity. The distribution of participant activities implicitly drives a market's return in terms of user participation, as manifested by the corresponding exponent. For example, in a hypothetical knowledge market where each answerer contributes equally, the answer generation model should be $N_a = A N_q^{\lambda_1} U_a^{1.0}$. In reality, the distribution of participant activities is a size dependent distribution controlled by the number of participants (system size). As the system size increases, most participants contribute to the head of the distribution (few activities), whereas very few join the tail (many activities).

We systematically reveal the size dependent distribution for participant activities in three steps. First, we empirically fit power-law distribution to the activities of participants in a month, for each month, for each Stack Exchange. We follow the standard procedure to fit a power-law distribution [1]. We observe that power-law well describe the monthly activity distributions. Second, we plot the

exponents of power-law against the number of participants for all observed months in a Stack Exchange, for each Stack Exchange. We observe that for most Stack Exchange power-law exponent decreases as the system size increases. Third, we apply linear regression to reveal the relationship between power-law exponents and system size. We observe that in general power-law exponents are negatively correlated with system size. This negative correlation is strongly visible in big knowledge markets that have at least 500 monthly participants in each month.

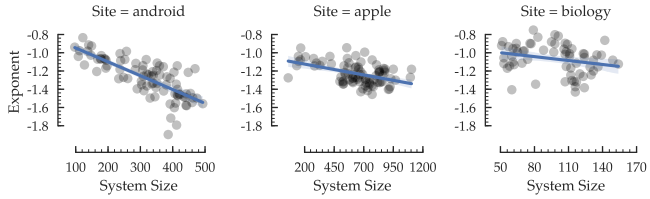


Figure 3: The visibility of size dependent distribution in `android` (strong), `apple` (moderate), and `biology` (weak). In most markets, the power-law exponent decreases with system size (similar to academia). In other markets, there exists a non-zero correlation between system size and power-law exponent.

In Figure 3 we present empirical evidence of size dependent distribution for answer generation in three markets: `android`, `apple`, and `biology`. We choose these examples to cover three possible visibility of size dependent distribution, as manifested by the correlation between power law exponent and system size—strong correlation ($|r^2| > 0.5$), moderate correlation($0.3 < |r^2| < 0.5$), weak correlation ($|r^2| < 0.3$).

Subject POV. The distribution of subject POV implicitly drives a market's return in terms of content dependency, as manifested by the corresponding exponent. Subject POV refers to the number of distinct perspectives on a particular content (primarily question) that imposes a conceptual limit to the number of dependent contents (answers). For example, an open-ended question such as 'What's your favorite book?' has many possible answers, whereas a close-ended question such as 'What's the solution for 3x+5 = 2?' has a single correct answer. In reality, most questions are neither completely open-ended nor completely closed; however, from an answerer's perspective, there's a diminishing utility in answering a question that already has an answer. This diminishing utility varies from question to question—questions asking for recommendations attract many answers, whereas questions seeking factual information attract few answers.

7.3 Uncovering the Stable Core

Now, we discuss about the core user community that assist maintaining the Cobb-Douglas models in a knowledge market. The Cobb-Douglas models indicate the presence of dynamic equilibria where the increase or decrease in user community does not affect the models. We assert that there is a stable user community in each knowledge markets who contribute a large fraction of contents; whereas the remaining users are unstable and contribute a small fraction. This is particularly applicable for high-threshold contents that require more effort, e.g, answers and comments.

We reveal the presence of core user community by summarizing the age of users with different levels of answer contribution across all Stack Exchange. First, we create 10 contribution categories, where each category covers an interval of length 5; the intervals are 1-6, 6-11, ..., 46-51. Then, we assign each user to one of these categories based on his/her average answer contribution per month. Note that all users who on average contribute more than 50 answers per month are capped at 50. Next, for each category, we compute the average age (the number of contributing months) of users within the category. We present this average age vs contribution curve in Figure 4. We observe that average age is an increasing function of contribution level—the users who contribute a lot to any market on a monthly basis also contribute for many months.

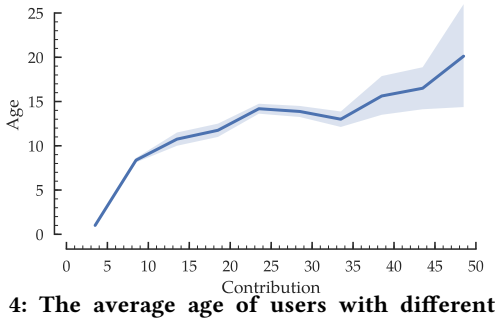


Figure 4: The average age of users with different levels of monthly answer contribution across all Stack Exchange. The users who contribute a lot to any market on a monthly basis also contribute for many months.

8 FAILURES AT SCALE

In this section we discuss how and why knowledge markets may fail at scale. We first empirically examine diseconomies of scale (Section 8.1), then analyze the effects of scale on market health (Section 8.2), and finally study user exchangeability under scale changes (Section 8.3).

8.1 Diseconomies of Scale

First, we examine disceconomies of scale—the ratio of answers to questions declining with the increase in number of users. The opposite of diseconomies is economies, when the ratio increases with the increase in number of users. The concept of diseconomies is important because the decrease in answer to question ratio implies the increase in gap between market supply (answer) and demand (question). In fact, if the ratio falls below 1.0, the gap becomes critical—guaranteeing there will be some questions with no answers.

In Figure 5 we present the economies and diseconomies of scale in three Stack Exchange markets: `cstheory`, `puzzling` and `superuser`. We choose these examples to cover three cases: strong diseconomies, strong economies, and weak economies. Among the three markets, `superuser` shows strong diseconomies of scale: if the number of users increases by 1%, then answer to question ratio declines by 0.95%. The other two markets show economies of scale, where `cstheory` shows strong economies: if the number of users

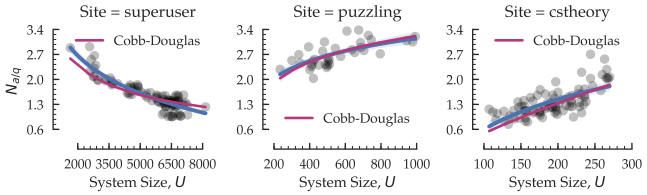


Figure 5: Diseconomies and economies of scale: the ratio of answers to questions decreasing (diseconomies) or increasing (economies) with the increase in number of users. 124 (out of 156) Stack Exchange markets exhibit diseconomies of scale. Examples: `superuser` (strong diseconomies), `puzzling` (weak economies), and `cstheory` (strong economies).

increases by 1%, then answer to question ratio increases by 0.8%; and `puzzling` shows weak economies: if the number of users increases by 1%, then answer to question ratio increases by 0.2%. Note that most markets, especially the ones with more than 500 monthly active participants, exhibit diseconomies of scale similar to `superuser`. Only five markets exhibit strong economies of scale in Stack Exchange: `cstheory`, `expressionengine`, `puzzling`, `ja_stackoverflow`, and `softwareengineering`.

The Cobb-Douglas curves well fit the empirical trends of economies and diseconomies (as shown in Figure 5). We derive these curves by dividing the answer models by the corresponding question models, and subsequently developing curves that capture economies and diseconomies ($N_{a/q}$) as a function of number of users (system size). We get similar curves via log regression. Between the two model types, the Cobb-Douglas models provide better explanation.

The Cobb-Douglas models well explain the economies and diseconomies of scale. As per the models, the primary cause of diseconomies is the difference between the diminishing returns of questions and answers for user participation. In other words, in most market, for user input, the marginal question output is higher compared to marginal answer output, i.e., an average user is likely to ask more questions and provide few answers. This causes the ratio of answers to questions to decline with the increase in number of users.

8.2 Analyzing Health

Next, we examine the disadvantage of scale through two health metrics— H_1 : fraction of answered questions (questions with at least one answer); and H_2 : fraction of questions with accepted answer (questions for which asker marked an answer as accepted). H_1 and H_2 capture the true gap between market supply (answer) and demand (question). The increase in number of users may cause decline in H_1 and H_2, as both metrics are related to the ratio of answer to questions. In fact, if the ratio falls below 1.0, it guarantees the decline of both metrics.

In Figure 6 we present the advantage and disadvantage of scale (through H_1 and H_2) for three Stack Exchange markets: `cstheory`, `puzzling` and `superuser`. We observe that the results are consistent with our analysis of economies and diseconomies—`cstheory` exhibit health advantage at scale, `puzzling` remains stable, whereas `superuser` exhibit disadvantage at scale. These three examples cover the possible health effects of scale in knowledge markets.

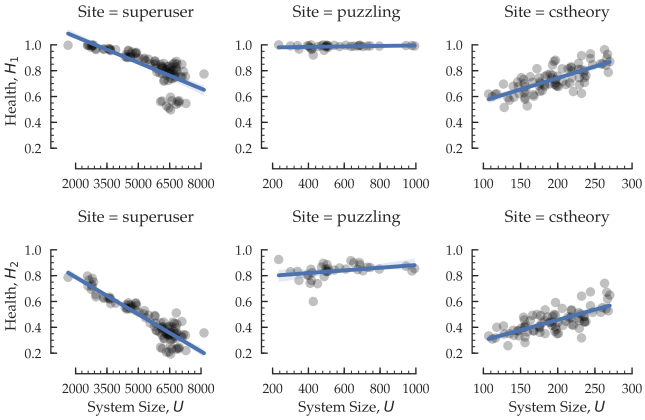


Figure 6: Health disadvantage and advantage of scale: the fraction of answered questions (H_1) and the fraction of questions with accepted answer (H_2) decreasing (disadvantage) or increasing (advantage) with the increase in number of users.141 (out of 156) markets exhibit disadvantage at scale. Examples: `superuser` (disadvantage), `puzzling` (neutral), and `cstheory` (advantage).

8.3 Effects on Exchangeability

Now, we empirically study the effects of scale on user exchangeability. By exchangeability, we specifically mean the gap between the top contributors and other participants in a knowledge market. Studying this gap is important because it can reveal if a market's success or failure depends on a small group of users.

To empirically study user exchangeability, we define two metrics that reflect the gap between the top contributors and other participants in a knowledge market. The first metric I_1 is defined as the ratio of average contribution for top k% users and bottom k% users. For computing I_1, we measure the contribution of each user as the ratio ($n_{a/q}$) of number of answers provided by the user to the number of questions asked by the user. Notice that I_1 is a ratio based metric and we define user contribution to be consistent with this metric. The second metric I_2 is defined as the sum of two distance: (i) the distance between the contribution of top k% users and average k% users, and(ii) the distance between the contribution of average k% users and bottom k% users. For computing I_2, we measure the contribution of each user as a tuple (n_a, n_q), consisting of the number of answers (n_a) provided by the user and the number of questions asked by the user (n_q). Notice that I_2 is an interval based metric and we define user contribution to be consistent with this metric. While both metrics have certain limitations (e.g., does not account for other content types), these metrics allow us to comprehend user exchegeability in Stack Exchange.

In Figure 7 we present the exchangeability of users under scale changes (through I_1 and I_2) for three Stack Exchange markets: `cstheory`, `puzzling` and `superuser`. Among the three markets, `superuser` exhibit the highest gap between top contributors and the other participants. However, as the number of participants increase, this gap decreases, i.e., the users become more exchangeable. In contrast, `cstheory` exhibit the lowest gap between top contributors and the other participants. However, as the number of participants increase, this gap increases, i.e., the users become less exchangeable.

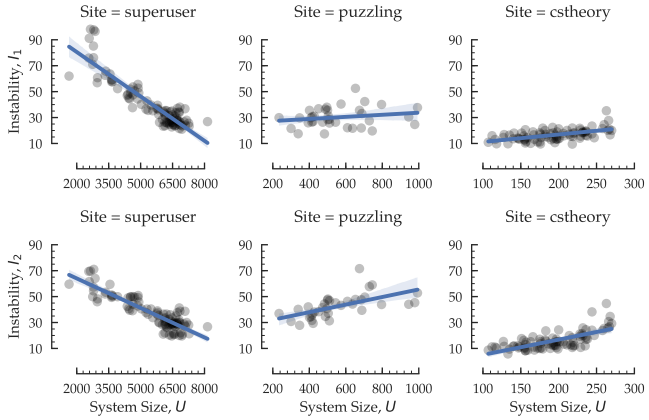


Figure 7: User exchangeability under scale: the gap (I_1 or I_2) between the top contributors and other participants in a knowledge market decreasing or increasing with the increase in number of users. Most markets exhibit a large gap between the top contributors and other participants, i.e., participants are dissimilar. Examples: `superuser` (high dissimilarity), `puzzling` (moderate dissimilarity), and `cstheory` (low dissimilarity).

9 DISCUSSION

In this section we discuss the implications of our research (Section 9.1) and the limitations of our work (Section 9.2).

9.1 Implications

Our work has implications for several research threads.

Power Law of Participation. Ross Mayfield coined the term 'Power Law of Participation'–when a small number of community members participate in high engagement activities, while the larger community participate in low threshold activities [13]. We observe power-law of participation in Stack Exchange markets. Stack Exchange supports varieties of activities ranging from low-threshold (e.g., voting) to high-engagement (e.g., collaborative editing, and linking similar questions); with a small fraction of users participating in high-engagement activities.

We assert that both low-threshold and high-engagement activities are required for a knowledge market's survival, and should proportionately increase with the increase in number of participants. However, in reality, for most knowledge markets, the user community contributing high-engagement activities does not scale with the system size—which creates a gap between market supply and demand for high-engagement activities, and consequently affects market health.

Microfoundations. The size-dependent distribution of user contribution implies that users who join a community later in its lifecycle exhibit different behavior than those who were present from the beginning. This very well may imply that the distribution of individual user behaviors (not just their overall production) is *also* a function of the system size. We should expect to see more a stable user behavior distribution over time for networks that appear to be more scale-insensitive; preliminary results suggest that this may indeed be the case.

9.2 Limitations

Now, we discuss several limitations of our work. First, the economic production models do not account for user growth. While there exists several user growth models for two-sided markets [10], membership based websites [15], and online social networks [23], it would be useful to introduce an economic user growth model that properly complements the production models. A potential direction in this research is to apply Malthusian growth model. Second, the production models inherit the fundamental assumptions of macroeconomics such as an aggregate is homogeneous (without looking into its internal composition), and aggregates are functionally related etc. It would be useful to empirically study these assumptions for knowledge markets.

10 CONCLUSION

In this paper, we examined CQA websites on the Stack Exchange platform through an economic lens by modeling them as knowledge markets. We analyzed a set of basis functions (the functional form of how an input factor, such as the number of available answerers, contributes to the overall output of the network) and interaction mechanisms (how the input factors interact with each other) to capture the content generation dynamics in knowledge markets. The resulting best-fit model, Cobb-Douglas, predicts the production of content with high accuracy. In addition, we showed that the model provides intuitive explanations for content generation in Stack Exchange markets. Namely, (1) factors such as user participation and content dependency have *constant elasticity*, meaning that a percentage increase in any of these inputs will result in a constant percentage increase in output; (2) input factors exhibit *diminishing returns* in that there is a decrease in the marginal (incremental) output (content production) as an input (e.g. number of people who answer) is incrementally increased while holding the other inputs constant; and (3) the efficiency of markets varies as manifest by their *returns to scale*—the increase in output resulting from a proportionate increase in all inputs. Finally, we explore the implications of the Cobb-Douglas economic model by showing the presence of diseconomies of scale in terms of the production of content, several measures of network health, and finally the exchangability of users in the network. We conclude that there is more to a successful CQA network than merely the total number of active users. Blindly increasing network sizes and ignoring these potential diseconomies of scale is unwise if we wish to sustain healthy knowledge markets.

REFERENCES

[1] Lada A Adamic. 2000. Zipf, power-laws, and pareto-a ranking tutorial. *Xerox Palo Alto Research Center, Palo Alto, CA, http://ginger. hpl. hp. com/shl/papers/ranking/ranking. html* (2000).

[2] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2012. 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 (KDD '12)*. ACM, New York, NY, USA, 850–858. https://doi.org/10.1145/2339530.2339665

[3] Lars Backstrom, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. 2006. Group Formation in Large Social Networks: Membership, Growth, and Evolution. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06)*. ACM, New York, NY, USA, 44–54. https://doi.org/10.1145/1150402.1150412

[4] Desta Fekedulegn, Siútáin Mártín Pádraig Mac, and Jim J. Colbert. 1999. Parameter Estimation of Nonlinear Models in Forestry. *Finnish Society of Forest Science and the Finnish Forest Research Institute* (1999).

[5] Adabriand Furtado, Nazareno Andrade, Nigini Oliveira, and Francisco Brasileiro. 2013. Contributor Profiles, Their Dynamics, and Their Importance in Five Q&A Sites. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, 1237–1252. https://doi.org/10.1145/2441776.2441916

[6] Glossary of economics. 2017. Glossary of economics — Wikipedia, The Free Encyclopedia. (2017). https://en.wikipedia.org/wiki/Glossary_of_economics [Online; accessed 30-October-2017].

[7] Lei Guo, Enhua Tan, Songqing Chen, Xiaodong Zhang, and Yihong (Eric) Zhao. 2009. Analyzing Patterns of User Content Generation in Online Social Networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*. ACM, New York, NY, USA, 369–378. https://doi.org/10.1145/1557019.1557064

[8] Heike Hofmann, Hadley Wickham, and Karen Kafadar. 2017. Letter-Value Plots: Boxplots for Large Data. *Journal of Computational and Graphical Statistics* 26, 3 (2017), 469–477. https://doi.org/10.1080/10618600.2017.1305277

[9] Sanjay Ram Kairam, Dan J. Wang, and Jure Leskovec. 2012. The Life and Death of Online Groups: Predicting Group Growth and Longevity. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining (WSDM '12)*. ACM, New York, NY, USA, 673–682. https://doi.org/10.1145/2124295.2124374

[10] Ravi Kumar, Yury Lifshits, and Andrew Tomkins. 2010. Evolution of Two-sided Markets. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining (WSDM '10)*. ACM, New York, NY, USA, 311–320. https://doi.org/10.1145/1718487.1718526

[11] Ravi Kumar, Jasmine Novak, and Andrew Tomkins. 2006. Structure and Evolution of Online Social Networks. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06)*. ACM, New York, NY, USA, 611–617. https://doi.org/10.1145/1150402.1150476

[12] Zhiyuan Lin, Niloufar Salehi, Bowen Yao, Yiqi Chen, and Michael Bernstein. 2017. Better When It Was Smaller? Community Content and Behavior After Massive Growth. In *Proceedings of the 11th International AAAI Conference on Web and Social Media*.

[13] Ross Mayfield. 2006. Power Law of Participation. (2006). http://ross.typepad.com/blog/2006/04/power_law_of_pa.html [Online; accessed 30-October-2017].

[14] Akshay Patil, Juan Liu, and Jie Gao. 2013. Predicting Group Stability in Online Social Networks. In *Proceedings of the 22Nd International Conference on World Wide Web (WWW '13)*. ACM, 1021–1030. https://doi.org/10.1145/2488388.2488477

[15] Bruno Ribeiro. 2014. Modeling and Predicting the Growth and Death of Membership-based Websites. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14)*. ACM, New York, NY, USA, 653–664. https://doi.org/10.1145/2566486.2567984

[16] I. Srba and M. Bielikova. 2016. Why is Stack Overflow Failing? Preserving Sustainability in Community Question Answering. *IEEE Software* 33, 4 (July 2016), 80–89. https://doi.org/10.1109/MS.2016.34

[17] Jim Stanford. 2008. Economics for Everyone: On-Line Glossary of Terms & Concepts. (2008).

[18] David Tilman. 1980. Resources: A Graphical-Mechanistic Approach to Competition and Predation. *The American Naturalist* 116, 3 (1980), 362–393. https://doi.org/10.1086/283633 arXiv:https://doi.org/10.1086/283633

[19] Simon Walk, Denis Helic, Florian Geigl, and Markus Strohmaier. 2016. Activity Dynamics in Collaboration Networks. *ACM Trans. Web* 10, 2, Article 11 (May 2016), 32 pages. https://doi.org/10.1145/2873060

[20] Lingfei Wu, Jacopo A. Baggio, and Marco A. Janssen. 2016. The Role of Diverse Strategies in Sustainable Knowledge Production. *PLoS ONE* (2016).

[21] Lingfei Wu and Jiang Zhang. 2011. Accelerating growth and size-dependent distribution of human online activities. *Phys. Rev. E* 84 (Aug 2011), 026113. Issue 2. https://doi.org/10.1103/PhysRevE.84.026113

[22] Jie Yang, Alessandro Bozzon, and Geert-Jan Houben. 2015. "Harnessing Engagement for Knowledge Creation Acceleration in Collaborative Q&A Systems". In *Proceedings of the 23rd International Conference on user Modeling, Adaptation, and Personalization (UMAP)*, Francesco Ricci, Kalina Bontcheva, Owen Conlan, and Séamus Lawless (Eds.). 315–327.

[23] Chengxi Zang, Peng Cui, and Christos Faloutsos. 2016. Beyond Sigmoids: The NetTide Model for Social Network Growth, and Its Applications. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. ACM, New York, NY, USA, 2015–2024. https://doi.org/10.1145/2939672.2939825