Modeling of Crowdsourcing Platforms and Granularity of Work Organization in Future Internet

Tobias Hoßfeld, Matthias Hirth, Phuoc Tran-Gia
University of Würzburg,
Institute of Computer Science,
Chair of Communication Networks
D-97074 Würzburg, Germany
{tobias.hossfeld|matthias.hirth|phuoc.trangia}@uni-wuerzburg.de

***Abstract*—Beside of social media networks, crowdsourcing is one of the emerging new applications and business models in the Future Internet, which can dramatically change the future of work and work organization in the on-line world. The crowdsourcing technology can be viewed as "Human Cloud" technique, in contrast to "Machine Cloud Computing". Using a crowd with a large number of internationally widespread workers and the flexibility of micro-payment services, crowdsourcing platforms like Amazon's MTurk and Microworkers can outsource traditional forms of work organization on a microscopic level of granularity to a large, anonymous crowd of workers, the human cloud. In such platforms work or tasks are organized at a finer granularity and jobs are split into micro-tasks that need to be performed by a human cloud.**

It is a need of analysis to understand the anatomy of such a platform and of models to describe the time-dependent growth of the human cloud, in order to predict the traffic impact of such novel applications and to forecast the growth dynamics. The purpose of this paper is a measurement-based statistical analysis of a crowdsourcing platform, using the Microworkers.com platform as example. The obtained results are then used to model the growth of such fast-changing environments in the Internet using well-known models from biology. Based on the findings from the population growth, we develop a deterministic fluid model which is an extension of the SIR model of epidemics, in order to investigate the platform dynamics.

I. Introduction

In the last decades, the Internet evolved from a simple collection of websites providing pure information towards a service and application platform by implementing new paradigms. The rise of the Peer-to-Peer paradigm led to new applications and services which allowed Internet users sharing files and user generated content among each other. Later on, the application of the Web 2.0 paradigm empowered Internet users to become application and service developer themselves. Examples of this new generation of websites are blogs, wikis or media-sharing platforms. Thereby, the users are connected to each other by means of social networks creating new path to communicate and share information. Such social media networks like Facebook or YouTube as prevailing Internet service platforms are technologically realized with machine cloud computing.

This work was conducted within the Internet Research Center (IRC) at the University of Würzburg.

Nowadays, a newly emerging service platform and business model in the Internet is established by the crowdsourcing paradigm. In contrast to outsourcing, where a job is performed by a designated worker or employee, crowdsourcing means to outsource a job to a large, anonymous crowd of workers, the so-called human cloud, in the form of an open call. This human cloud is abstracted by a crowdsourcing platform, which distributes the work submitted by an employer among the human worker resources and acts as mediator between worker and employer. The crowdsourcing paradigm is changing dramatically the future of work and work organization in the Internet. The work is organized at a finer granularity and jobs are split into cheap micro-tasks that can be fast performed by the human cloud. With a huge number of international workers and the flexibility of micro-payment services like PayPal, crowdsourcing platforms like Amazon Mechanical Turk or Microworkers are already serious business.

The importance of crowdsourcing will increase even more in the future. Several companies like IBM or BMW used and use already the concept of crowdsourcing to improve their products and to foster new innovations. Due to the increasing interest in crowdsourcing, there is a lot of ongoing research in this area[1]. However, most of the research focuses on the demographics of the crowdsourcing users and the business aspects, but the impact of crowdsourcing platforms on future Internet traffic is still unknown. Because of the size of the human cloud, these crowdsourcing platforms will change significantly the Internet traffic in a similar manner as Facebook or other social media networks today. Thus, it is an important telecommunications issue to model – beside the growth of social media network – the growth of crowdsourcing platforms. An appropriate model is required in order to estimate the traffic growth in the emerging Future Internet and to forecast the growth dynamics, in order to cope with the particular traffic demands of such crowdsourcing platforms.

The main contributions of this paper are 1) a measurement-based statistical analysis of a crowdsourcing platform and the

[1]During the last few year specialized conferences and workshops where initiated like the Conference on the Future of Distributed Work (CrowdConf), the Workshop on Crowdsourcing for Search and Data Mining (CSDM), the Workshop on Crowdsourcing and Human Computation (CHI), and the International Conference on Computational Aspects of Social Networks (CASoN).

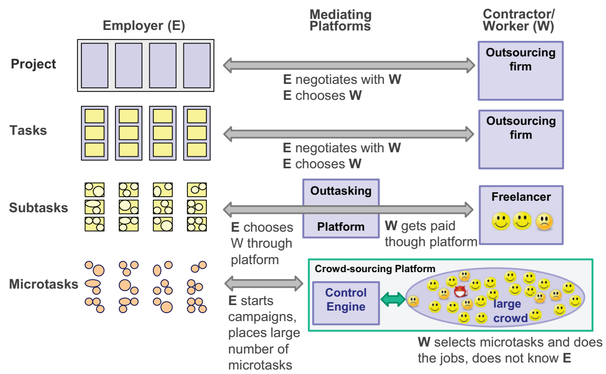

Fig. 1. Evolution of the granularity of work.

granularity of work, as well as 2) models for the population growth of the human cloud as well as for the platform dynamics. We use anonymized data traces from Microworkers in order to investigate which models are appropriate. In a first step, the growth of the Microworkers population is analyzed and well-known growth models applied in a range of fields including biology and sociology are used. In a second step, we use the findings obtained from the population growth model in order to investigate the platform dynamics. We develop a fluid model which is an extension of the SIR (Susceptible-Infected-Recovered) model of epidemics.

The remainder of the paper is structured as follows. At first, we provide a comprehensive background on crowdsourcing and discuss related work in Section II. The anatomy of the Microworkers platform with respect to the workers and the granularity of work are analyzed in Section III. The growth model of the human cloud is derived in Section IV, while the fluid model of the platform dynamics is developed in Section V. Finally, we conclude our work in Section VI.

II. Background and Related Work

As background of the modeling approaches, we shortly outline the evolution of work granularity from outsourcing to crowdsourcing and introduce the crowdsourcing scheme as well as common terminology in this context. In addition, related work is reviewed briefly.

A. Granularity of Work

Crowdsourcing platforms can be viewed as a further development of the outsourcing principle, by reducing the administrative overhead and the size of the outsourced tasks. We illustrate this evolution in the granularity of work in Figure 1.

The largest work package that can be completed by a foreign conductor is a *project*. In this case, the *employer* chooses directly which firm should accomplish the project and negotiates with this firm the terms of the contract.

Every project consists of a number of *tasks* which represent the milestones of the project. Similar to a project, these tasks can also be accomplished by a foreign firm. The procedure is similar, the employers chooses the outsourcing firm and negotiates the contract details.

The next smaller granularity is a *subtask*. This can be viewed as a small part of the whole project, where no knowledge about the whole project is required. In contrast to tasks, subtasks can be accomplished by individuals e.g. freelancers. In order to access the freelancers, the employer uses an *out-tasking platform*. On out-tasking platforms, freelancers can upload a personal profile with their references, skills and salary expectations. An employer chooses a freelancer according to his demands. In contrast to outsourcing a project or a task, the employer does neither directly communicate with the freelancer nor pays him directly. The out-tasking platform acts as a mediator between them.

The finest granularity of work is the *microtask*. A microtask can be accomplished within a few minutes to a few hours and thus does not need a long term employment. Further, it is irrelevant to the employer who accomplishes the task and usually the task has to be repeated several times. The repetitive tasks are combined in a *campaign*, which the employer submits to the crowdsourcing platform. Similar to the out-tasking platform, the *crowdsourcing platform* acts as a mediator between the employer and the anonymous *workers*. However, the workforce in the crowdsourcing approach is not a *designated* worker but a human cloud.

B. Crowdsourcing Scheme

Applying the crowdsourcing scheme, an employer submits a campaign to the crowdsourcing platform which mediates the tasks to the human cloud. Thus, an employer does not selectively choose a worker for a certain task, but offers the task to a large crowd of workers who can freely choose to work on this task or not. In the remainder of this paper, we use the term microtask and task in the context of crowdsourcing, equivalently. A *user* of a crowdsourcing platform has registered and may act as worker as well as employer at the same time. In the case of the Microworkers platforms, 90.52 % of the users are pure workers, 3.59 % are pure employers, and 5.89 % are workers and employers.

The life cycle of a campaign in the Microworkers platform comprises the following steps. (1) First, the employer submits a campaign to the crowdsourcing platform. This includes a description of the task, the payment per task, how the workers have to proof a completed task, and how many tasks n are needed. (2) Then an employer of Microworkers reviews the campaign and approves it, if it corresponds to their guidelines. (3) Afterwards, the workers start working on the campaign and submit their finished tasks. (4) As soon as the desired n tasks are completed, the campaign is paused. The employer has to review the submitted tasks within 7 days, if they are valid. If m tasks are not valid, the campaign resumes until m new tasks are submitted. (5) If the employer rated n tasks valid, the campaign is completed.

We define the *completion time* of a campaign as the time difference between the submission of the campaign (1) by the employer and the submission of the nth valid task (5).

C. Related Work

During the last few years numerous commercial crowdsourcing platforms like Amazon Mechanical Turk (MTurk) [1], Microworkers [2], Shorttask, and Clickworker were launched. However, the majority of the crowdsourcing research is based on the analysis of and tests run on MTurk.

In [3] Ipeirotis analyzed the characteristics of the MTurk marketplace. According to his findings, there are a few very active users who account for most of the reward on MTurk and a huge number of users investing small amounts. The top users, are companies with business models based on crowdsourcing or companies which offer services to crowd source specific tasks or type of tasks. This shows that crowdsourcing enabled new business concepts and offers cost-effective possibilities to perform existing tasks.

Even if 90 % of the tasks are paid 0.10 USD or less, the hourly wages for a worker are about 5 USD. This demonstrates that crowdsourcing is a real alternative for people in low-wage jobs. The payment and the time required to complete the task groups also indicate, that the jobs on MTurk belong to the above mentioned microtasks. Based on these measurements Wang et al. [4] showed the effect of various parameters, like e.g. payment on the completion time of tasks.

Besides economic aspects, the demographics of MTurk have been addressed in previous work, too. Using user based surveys, Ross et al. and Ipeirotis [5]–[9] analyzed the demographic development of MTurk. They showed that most of the users working on MTurk are from the USA and that the percentage of Indian workers is dramatically increasing since MTurk added the possibility to receive payments in India. Opening crowdsourcing platforms to tradition outsourcing countries, with huge human resources offers tremendous growth potentials. This makes crowdsourcing platforms, like other Social Media Networks e.g. Facebook or YouTube, to the traffic hot-spots of the Future Internet.

To the best of our knowledge, this is the first paper which investigates the growth and the dynamics of crowdsourcing platforms which help to predict the traffic in the emerging Future Internet.

III. Anatomy of the Microworkers Crowdsourcing Platform

For analyzing the anatomy of a crowdsourcing platform and in particular the granularity of work organization, we got the necessary data from the Microworkers operator. All information which could identify a user had been removed or anonymized by the platform operator. The database snapshot covers the time from May 2009 when the platform was launched, to October 2010. At this time the platform had about 80,000 registered users who submitted over 15,000 campaigns and completed 860,000 tasks.

As our data is not based on user surveys, but directly retrieved from the platform operator, we have exact and correct information. The only exception is the home country of the users, which is set by the user himself during the registration process and thus might not be correct in all cases. However, this is not relevant for the models developed in this paper. A detailed analysis of the Microworkers platform can be found in a previous work [10].

A. Demographics of Microworkers Users

The users of Microworkers are from over 190 countries world wide. The largest countries according to their number of workers are Indonesia, Bangladesh, India, the United States, the Philippines and Romania. Most of the employers are from the United States, followed by Indonesia, India, the United Kingdom, the Philippines and Romania. These countries account for 3/4 of all workers and employers, respectively.

The number of workers per country is almost proportional to the number of users from this country. This is caused by the fact, that most of the users are workers and only small percentage employers. Most of the workers are located in Asia, while there is also a significant amount of workers from the USA and Europe. Most of the employers are from the USA even if this country only accounts for 13 % of all users. This indicates that in crowdsourcing, work from high-wage countries is delegated to workers in low-wage countries, similar to outsourcing.

In order to verify this, we analyzed the United Nations' Human Development Index (HDI) [11] of the home countries. The HDI is an indicator of a country's development and categorized them in low, medium, high and very high developed countries. Our analysis showed that workers are more likely to be from low or medium developed countries than employers. However, there are still about 30 % of the workers from high and very high developed countries. In our dataset, the HDI and the number of workers or employers are statistically uncorrelated.

The distribution of the worker and the employers shows that crowdsourcing is neither used only in high wage countries in order to spend some free time, nor it is used only to exploit workers from low wage countries. In contrast, both the workers and the employers are international. These results indicate the potential world-wide use of crowdsourcing in the Future Internet.

B. Modeling the Granularity of Work Organization

In this section, we analyze the granularity of work on crowdsourcing platforms which is quantified in terms of the payment per task, the number of tasks per campaign indicating its repetitiveness, and the completion time of campaigns.

In order to distinguish different campaign types, a cluster analysis was performed by using the Expectation Maximization algorithm as implemented in the data-mining tool *Weka* [12]. The following input variables for clustering the campaigns were used, a) completion time of a campaign, b) number of tasks per campaign, c) payment per task, d) estimated duration for completing the task as announced in the job description by the employer, e) length of the job description in number of characters as indicator for the complexity of a job, f) task ratio which is the relative number of tasks rated

TABLE I
GRANULARITY OF WORK AS MEASURED FOR MICROWORKERS.

cluster id	avg. completion time of campaigns	avg. number of tasks per campaign	avg. payment per task	avg. task ratio per campaign
0	0.48 days	47	0.11 USD	99 %
1	2.69 days	44	0.15 USD	92 %
2	5.58 days	256	0.11 USD	99 %
3	19.58 days	32	0.45 USD	95 %
4	25.99 days	53	0.59 USD	69 %

TABLE II
GOODNESS-OF-FIT FOR MICROWORKERS GROWTH MODELS.

model	function	parameter	gof R^2
square	$N_{\mathrm{squ}}(t) = at^2 + bt$	a=0.216173, b=45.87	0.9988
logistic	$N_{\log}(t) = \frac{N_0 \cdot K}{N_0+(K-N_0)e^{-r_0 t}}$	N_0=4507.4, K=133162, r_0=0.00734	0.9959
exponential	$N_{\exp}(t) = N_0 e^{rt}$	N_0=3434.76, r =0.00678	0.9572
hyperbolic	$N_{\mathrm{hyp}}(t) = \frac{K}{t_c-t}$	K=10099646, t_c=619.95	0.9302

to be valid by the employer. The task ratio is an indicator for the difficulty of tasks.

As a result, five different clusters were identified which are summarized in Table I. The campaigns in cluster 0 contain simple jobs, which are completed very quickly, but are very low paid. Accordingly, the task ratio is very high. The campaigns in cluster 1 have little higher requirements which are reflected by larger completion times, slightly higher payments, but lower task ratios compared to cluster 0. The tasks of the campaigns in cluster 2 are comparable to the tasks of campaigns in cluster 0, however, these campaigns are much larger in terms of number of tasks which results in a corresponding larger campaign completion time. The tasks of the campaigns in cluster 3 and cluster 4 show a higher complexity which is indicated by the larger completion times and the larger payment. Especially the campaigns in cluster 4 seem to contain the most complex tasks, since the task ratio is quite low with about 70 %. A closer look at the database snapshot reveals that the employers of the campaigns in cluster 4 are running several campaigns. Thus, the employers are used to crowdsourcing and the characteristics of the granularity of work and consequently know what they can request from the workers in terms of payment and quality of the submitted tasks.

In general, we see that although some tasks are paid up to 8.00 USD, the payment for 98 % of the tasks is less than one dollar. This indicates that crowdsourcing tasks on Microworkers are simple tasks which can be completed quickly. Therefore, we analyze next the completion times of campaigns in more detail. In our analysis, we only consider completed campaigns where all n valid tasks are already submitted. We exclude campaigns that have been stopped by the employer, blocked by Microworkers or campaigns which are still running.

Figure 2 shows the cumulative distribution function of the completion times of campaigns for the five different clusters. The percentage value in Figure 2 shows the ratio of campaigns belonging to a certain cluster. The campaigns in cluster 0 have a very short completion time below 2.5 days. About 90 % of the campaigns in cluster 1 are completed within one week. Although the number of tasks is five to six times larger for campaigns in cluster 2 than in cluster 0 or cluster 1, 90 % of the campaigns are already completed within two weeks. The completion time of campaigns in cluster 3 and cluster 4 is about 50 days and 60 days, respectively, due to the more complex tasks.

Our analysis shows, that crowdsourcing tasks are mostly less paid, require only a small amount of time and are highly repetitively in contrast to traditional forms of work organization.

IV. POPULATION GROWTH MODELS

The scope of this section is to model the population growth of a crowdsourcing platform. The growth of the Microworkers population is analyzed based on the available measurement traces and well-known growth models applied in a range of fields including biology and sociology are investigated for their applicability. In particular, four different population growth models are reviewed in order to describe the number of users over time who registered at the Microworkers platform.

As growth models, we consider the following (1) unbounded models for exponential, hyperbolic, and square growth, as well as (2) the bounded logistic growth model which leads to a maximum number of users in the system. The resulting population curves according to these models as well as the measured number of registered users over time are depicted in Figure 3. The parameters of the different models are obtained by minimizing the least-square errors between the measurement and the model data using the Levenberg-Marquardt algorithm as implemented in Matlab. The goodness-of-fit in terms of coefficient of determination R^2 as well as the parameters of the different models are given in Table II.

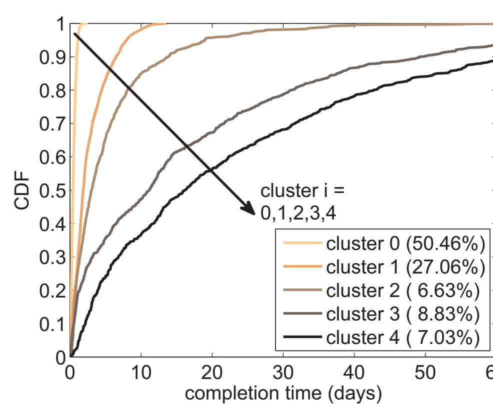

Fig. 2. Completion times of campaigns for the different clusters.

A. Unbounded Growth Model

The *exponential growth model* is associated with Thomas Robert Malthus (1766-1834) and is for example used to describe the growth of bacteria. The growth of the population $\frac{d}{dt}N_{\text{exp}}(t)$ depends on the actual number $N_{\text{exp}}(t)$ of bacteria or users in the case of crowdsourcing which are already existing at time t, i.e. $\frac{d}{dt}N_{\text{exp}}(t) \sim N_{\text{exp}}(t)$. It is defined by

$$N_{\text{exp}}(t) = N_0 e^{rt} \,. \tag{1}$$

The parameter N_0 describes the initial population at time $t = 0$. The model parameter r is called *Malthusian parameter* or *population growth rate* which determines the outcome of the model. For $r = 0$, the population does not change. For $r < 0$, the population exponentially declines, while for $r > 0$ the population exponentially increases which is the case for Microworkers. However, the exponential growth model overestimates the number of users after $Q3-10$, showing that this model is not appropriate for Microworkers, cf. Figure 3.

The *hyperbolic growth model* has a singularity in finite time which grows to infinity at a finite time t_c. This model was suggested to describe the world population until the early 1970s. It is defined by

$$N_{\text{hyp}}(t) = \frac{K}{t_c - t} \,, \tag{2}$$

where K is a scale factor. The absolute population growth rate in the moment t is proportional to the square of the number of users $N_{\text{hyp}}(t)$ at time t, i.e. $\frac{d}{dt}N_{\text{hyp}}(t) \sim N_{\text{hyp}}(t)^2$. However, the hyperbolic model also does not apply to the Microworkers population, as we can see from Figure 3 and Table II. In particular, according to the measurements, the singularity would already appear in February 2011, but the number of Microworkers users is still far away from infinity.

The *square growth model* is not often observed in nature. The interpretation of this model is that the growth rate $\frac{d}{dt}N_{\text{squ}}(t)$ of the population increases linearly over time, i.e. $\frac{d}{dt}N_{\text{squ}}(t) \sim t$. The model is described by

$$N_{\text{squ}}(t) = at^2 + bt \tag{3}$$

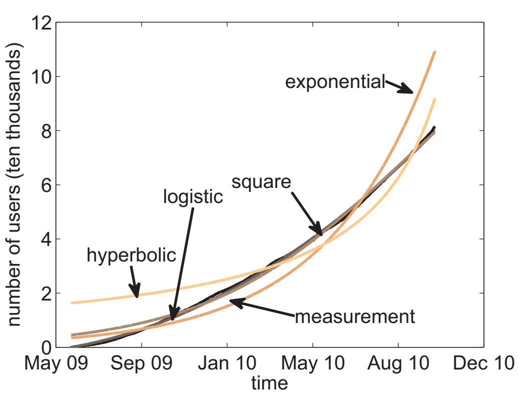

Fig. 3. Growth models for the number of users signed to Microworkers.

with the square growth factor a and the linear growth factor b. This model fits very well with the growth of the Microworkers population and no difference between the measurements and the square growth model can be seen in Figure 3. The coefficient of determination R^2 is very close to 1 which means a perfect match between model and measurements.

From a mathematical point of view, square growth seems to be valid for Microworkers, however, the model is unbounded which means that the number of users is not limited. Possible explanations for unbounded growth are the exponential growth of world population or the usage of multiple Microworkers accounts for same workers. A look into the future reveals that – according to the square growth model – there will be more than one million Microworkers users in January 2015, which seems to be quite realistic.

B. Logistic Growth Model

The logistic function is applied in various fields like biology, sociology or economics, and especially in demographics for describing population growth. The logistic growth model was developed by Pierre Verhulst (1804-1849) who suggested that the rate of population increase may be limited, i.e., it may depend on population density,

$$r(t) = r_0 \left(1 - \frac{N(t)}{K}\right) \,. \tag{4}$$

In the beginning, growth is approximately exponential which slows down when saturation begins, while finally growth stops. The parameter r_0 is referred to as the *intrinsic growth rate* and is the maximum possible rate of population growth. The parameter K is the maximum number of users in the system. The dynamics of the population is described by the differential equation

$$\frac{d}{dt}N_{\log}(t) = rN(t) = r_0 N(t) \left(1 - \frac{N(t)}{K}\right) \tag{5}$$

which has the solution

$$N_{\log}(t) = \frac{N_0 \cdot K}{N_0 + (K - N_0)e^{-r_0 t}} \tag{6}$$

with the initial population size N_0.

The logistic curve converges towards $\lim_{t\to\infty} N_{\log}(t) = K$. In the case of Microworkers, it is $N_0 < K$. Thus, the population increases until it reaches the maximum capacity K. According to the least squares fitting of the measurement data to the logistic model, the number of users is limited to $K = 133162$ users, as can be seen Table II, and will already been reached in 2012, see Figure 4. Since the prediction of the future is impossible, we continue to investigate how the dynamics of the system can be influenced in the following section and use the results from the logistic growth model, since it also shows a nearly perfect match with the measurement with $R^2 = 0.9959$ and is widely accepted for modeling population growth.

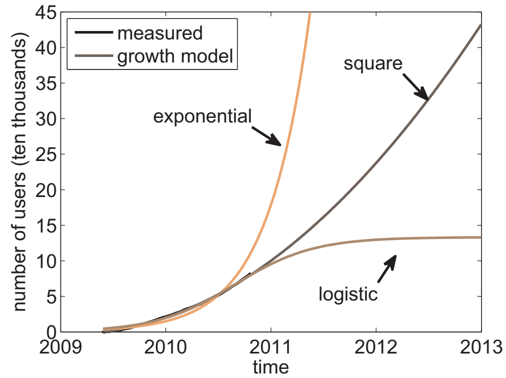

Fig. 4. Looking into the future: population growth of Microworkers according to different models.

V. Platform Dynamics using SIR Model

In this section, we describe the platform dynamics and develop a deterministic fluid model which is an extension of the SIR (Susceptible-Infected-Recovered) model [13] of epidemics. This model is used to research different types of popularity which trigger and influence the decision of users being active or inactive within the platform. These types of popularity lead to different dynamic models which are referred to as *globally influenced dynamics (GID) model* and *local user decision (LUD) model*. In addition, we investigate how the dynamics of the platform can be influenced, e.g. by advertisement campaigns, such that the critical mass for platform success is reached faster.

The scope of this section is the investigation of the platform dynamics of users becoming active and inactive, respectively. In particular, we are interested to evaluate whether the crowdsourcing platform gets successful and after which time. This is achieved by considering the steady state of the system. In addition, we are interested how the dynamics can be influenced by means of advertisement campaigns and what is the impact of this advertisement.

The platform dynamics in terms of number of active and inactive users can be described by using a deterministic fluid model. As a result of the logistic growth model in Section IV, we assume that there is a fixed maximum number K of users in the system.[2] This means that we consider a fixed population consisting of the number N of non-Microworkers users, which haven't signed up so far, the number A of active users, and the number I of inactive users.

$$K = N + A + I \tag{7}$$

The active users have signed up to Microworkers and are actively using the platform, either as employer or employee, while the inactive users are also registered but did not use their accounts for several month.

[2]It has to be noted that the SIR model can also be adapted to the results of the square growth model, such that the number of the users is not limited.

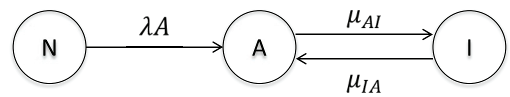

Fig. 5. Extension of the SIR model for platform dynamics.

Figure 5 illustrates the population dynamics of the crowdsourcing platform. Non-Microworkers users sign up and register at the crowdsourcing platform with a rate of λA. This is analogous to the basic SIR model by W. O. Kermack and A. G. McKendrick (1927), where susceptible, infected, and recovered individuals are considered in order to describe the transmission of disease through individuals. Each individual in the population has an equal probability of contracting the disease with a rate of λ. Then, all infected individual are able to transmit the disease to susceptible individuals with rate λA, with A being the number of infected people. The application of the SIR model to the crowdsourcing platform can be interpreted as follows. The active users A of the crowdsourcing platform infect non-Microworkers users N with crowdsourcing at rate λA. Hence, the rate of new infections, i.e. users subscribing to Microworkers, is λAN. Thus, the dynamics of the crowdsourcing population can be derived with the following differential equations. It has to be noted that all variables are time-dependent, however, for the sake of readability we use a shorter notion, e.g. A instead of $A(t)$ or λ instead of $\lambda(t)$.

$$\begin{aligned} \frac{dN}{dt} &= -\lambda AN \\ \frac{dA}{dt} &= -\mu_{AI}A + \mu_{IA}I + \lambda AN \\ \frac{dI}{dt} &= -\mu_{IA}I + \mu_{AI}A \end{aligned} \tag{8}$$

Active users get inactive with rate μ_{AI}, while inactive users get active again with rate μ_{IA}. There are two different reasons why an active user may get inactive. This is reflected by the GID and the LUD model which will be explained in the following.

A. Globally Influenced Dynamics (GID) Model

The GID model assumes that the usage of the crowdsourcing platform is influenced by the global opinion and popularity of the platform. Thus, the more people are active, the more people will be attracted. This includes both, the non-Microworkers users as well as the inactive users. However, this will also result in the opposite effect. If there are many inactive users in the platform, which are registered, but do not actively participate in crowdsourcing, this may disappoint active users which consequently get inactive. Accordingly, active users get inactive with rate γI and inactive users get active again with rate δA, respectively.

$$\mu_{AI} = \gamma I \quad \text{and} \quad \mu_{IA} = \delta A \tag{9}$$

In the steady state $\lim_{t\to\infty}$, all K users are either active or inactive, i.e.

$$\lim_{t\to\infty} N(t) = 0\,. \tag{10}$$

In addition, the population sizes do not change anymore in the steady state, i.e. $\frac{dN}{dt} = \frac{dA}{dt} = \frac{dI}{dt} = 0$. Thus, the differential equation system in Eq. (8) gets a linear equation system in the steady state

$$0 \;=\; -\gamma IA + \delta AI\,, \tag{11}$$

which can be easily solved by

$$A = 0 \vee I = 0 \vee \gamma = \delta\,. \tag{12}$$

Using

$$K \;=\; A + I\,, \tag{13}$$

a case differentiation yields to the following results for the steady state.

$$\gamma > \delta \;\Rightarrow\; I = K, A = 0, N = 0 \tag{14}$$
$$\gamma < \delta \;\Rightarrow\; I = 0, A = K, N = 0 \tag{15}$$
$$\gamma = \delta \;\Rightarrow\; I = I_0, A = K - I_0, N = 0 \tag{16}$$

Thus, if the crowdsourcing platform operator is able to ensure that more users are active than inactive, i.e. $\delta > \gamma$, then all users will actively use the platform. In practice, this can be achieved by different forms of advertisement campaigns, by a well operated platform, by solving problems between employers and employees easily, etc.

B. Local User Decision (LUD) Model

The LUD model assumes that users individually decide to use the crowdsourcing platform. A user is dissatisfied individually, e.g. the completion time for a campaign is too large for an employer or a job completed by an employee was not accepted by the employer such that the employee did not receive any reward. In that case, the users get inactive independent of the overall platform popularity – in contrast to the GID model. The same is true for inactive users getting active again. Independent of the global opinion, inactive users will take new jobs or launch campaigns, i.e. getting active.

$$\mu_{AI} = \gamma \quad \text{and} \quad \mu_{IA} = \delta \tag{17}$$

For the steady state, we arrive at the following equations

$$0 \;=\; -\gamma A + \delta I\,, \tag{18}$$
$$K \;=\; A + I\,, \tag{19}$$

which can be easily solved by

$$I \;=\; \frac{\gamma}{\gamma+\delta} K\,, \tag{20}$$

$$A \;=\; \frac{\delta}{\gamma+\delta} K\,. \tag{21}$$

The user decisions about the usage of the platform are independent from other users' opinions according to the LUD model. As a consequence for the platform operator, he constantly has to give incentives to its users for being active to increase the rate δ.

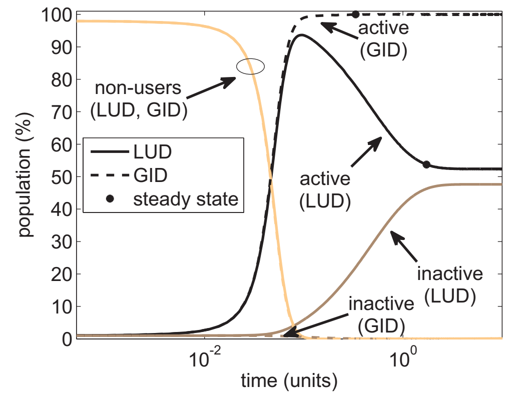

Fig. 6. Evolution of populations for LUD and GID model.

C. Influencing the Platform Dynamics

So far, we have investigated the steady state how many users will finally actively participate in the platform. However, an important factor for the platform operator from an economic point of view is the time when the critical mass is reached. According to the one-third hypothesis (OTH) by Hugo O. Engelmann, a group's prominence increases as it approaches one-third of the population and diminishes when it exceeds or falls below one-third of the population. This OTH can be applied to a crowdsourcing platform, accordingly. However, since we assume different kind of popularities and opinion forming in the GID and LUD model, we decided to consider the point in time t_0 when the steady state is reached in order to compare numerical results of the GID and LUD model.

Figure 6 exemplary shows the evolution of populations for the LUD and GID model. The dots in the figure represent t_0 for the LUD and GID model, respectively. The x-axis scaled logarithmically denotes the time which we normalized by the arrival rate λ, while the y-axis shows the relative number of non-, active, and inactive users. It can be seen that the curves for the non-users N are identical for the LUD and GID model, since Equation 8 is identical for both models. However, the number A of active users grows faster in the GID model, since the global opinion triggers the growth of A. Therefore, the steady state is reached faster for GID than for LUD. The curve for the active users for LUD is first dominated by the arrival of non-users λ until it converges to $\frac{\delta}{\gamma+\delta}$.

A comparison of the time until the steady state is reached for both models is shown in Figure 7. Different platform registration rates λ are considered, that are $\lambda = [0.1; 0.2; 0.3; 1]$. On the x-axis, the rate δ for getting active is varied from 1 to 10, while the rate γ is fixed with $\gamma = 1$. Hence, it is $\delta >= \gamma$ which means that in the GID model all users will finally be active. As we can see, the steady state is always reached faster for GID for the same λ than for LUD. In addition, the step from $\lambda = 0.1$ to $\lambda = 0.2$ leads to a significant improvement for the platform operator. Thus, for reaching fast the critical mass, a platform operator should try to motivate enough people to join and try the platform, especially in the beginning.

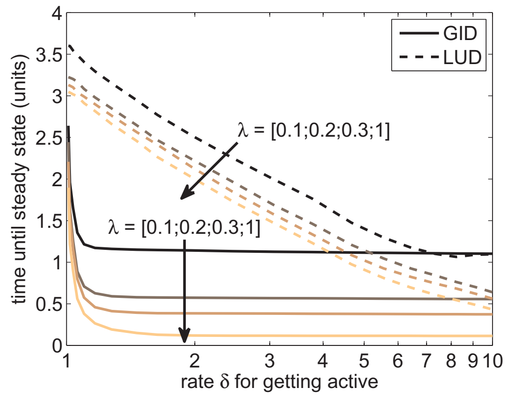

Fig. 7. Time until steady state is reached for different platform registration rates λ. The larger the rate λ is, the earlier the steady state is reached.

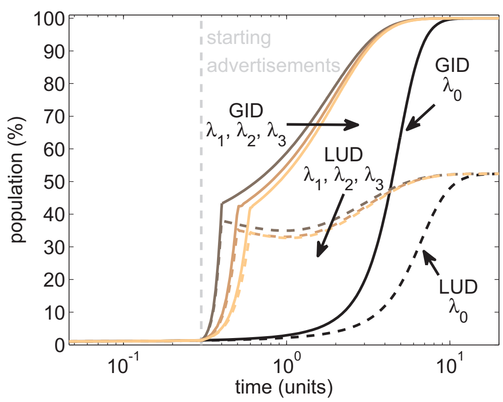

Fig. 8. Influence of different advertisement campaigns λ_i starting at $t = 0.2$. The original curves without advertisement campaign are labeled with λ_0.

Later on, the platform owner can influence the dynamics of the system by advertisement campaigns or other incentives, such that the users sign in. We consider now different registration rates $\lambda_i(t)$ reflecting different advertisement campaigns which vary in length and intensity. Nevertheless for different campaigns i and j, the same number of people is motivated, i.e. $\int_0^\infty \lambda_i(t) = \int_0^\infty \lambda_j(t)$. The evolution of the system without advertisement campaign is referred to as λ_0.

$$\lambda_i(t) = \begin{cases} 0.01, & t < 0.3 \\ {}^{0.4}/_{i} & 0.3 \leq t \leq 0.3 + i \cdot 0.1 \\ 0.01, & t > 0.3 + i \cdot 0.1 \end{cases} \tag{22}$$

The influence of the different advertisement campaigns is illustrated in Figure 8 for $i = 1, 2, 3$ when the campaign starts at $t = 0.2$. It can be seen again that the advertisement campaign has a significant impact on the system dynamics. Furthermore, we see that the larger the peak of the advertisement campaign is, the faster the users join the platform.

In summary, the platform operator has different options to influence the population dynamics. New users should be given incentives or motivated by advertisements etc., especially in the beginning, in order to increase $\lambda(t)$. However, short, but intensive campaigns are more successful. In addition, the platform operator should foster users keeping active, i.e. increasing δ or reducing γ by appropriate means.

VI. Conclusions

Crowdsourcing is one of the emerging new applications in the Future Internet based on the Human Cloud technique. It will dramatically change the future of work and work organization and significantly impact the traffic in the Future Internet. Based on measurements of the Microworkers platform, the anatomy of a crowdsourcing platform and the granularity of work was analyzed and shown to be much finer than in traditional forms of work organization. The population growth of the platform follows either a square or a logistic growth model. We developed a fluid model based on an epidemic model from biology to model the dynamics of the platform and the activity of its users. The growth model and the dynamics model allow forecasting the traffic caused by crowdsourcing platforms in the Future Internet.

Acknowledgements

The authors thank Nhatvi Nguyen (Weblabcenter Texas) for his valuable comments and Matjaz Vrecar for his technical support.

References

[1] Amazon.com, Inc., "Mechanical Turk," www.mturk.com.
[2] Weblabcenter, Inc., "Microworkers," www.microworkers.com.
[3] P. Ipeirotis, "Analyzing the Amazon Mechanical Turk Marketplace," *CeDER Working Papers*, Sep. 2010.
[4] J. Wang, S. Faridani, and P. G. Ipeirotis, "Estimating the Completion Time of Crowdsourced Tasks Using Survival Analysis Models," in *Proceedings of the Workshop on Crowdsourcing for Search and Data Mining (CSDM)*, Hong Kong, China, Feb. 2011.
[5] P. Ipeirotis. (2008, Mar.) Mechanical Turk: The demographics. Online. [Online]. Available: http://behind-the-enemy-lines.blogspot.com/2008/03/mechanical-turk-demographics.html
[6] J. Ross, A. Zaldivar, L. Irani, and B. Tomlinson, "Who are the Turkers? Worker Demographics in Amazon Mechanical Turk," Department of Informatics, University of California, Irvine, USA, Tech. Rep., Jan. 2009.
[7] P. Ipeirotis. (2009, Mar.) Turker Demographics vs Internet Demographics. Online. [Online]. Available: http://behind-the-enemy-lines.blogspot.com/2009/03/turker-demographics-vs-internet.html
[8] ——. (2010, Mar.) The New Demographics of Mechanical Turk. Online. [Online]. Available: http://behind-the-enemy-lines.blogspot.com/2010/03/new-demographics-of-mechanical-turk.html
[9] J. Ross, L. Irani, M. S. Silberman, A. Zaldivar, and B. Tomlinson, "Who are the crowdworkers? Shifting demographics in mechanical turk," in *Proceedings of the 28th of the International Conference on Human Factors in Computing Systems (CHI 2010)*, Atlanta, Georgia, USA, Apr. 2010.
[10] M. Hirth, T. Hoßfeld, and P. Tran-Gia, "Anatomy of a Crowdsourcing Platform - Using the Example of Microworkers.com," in *Proceedings of the Workshop on Future Internet and Next Generation Networks (FINGNet)*, Seoul, Korea, Jun. 2011.
[11] United Nations Development Programme, *Human Development Report 2010*. Palgrave Macmillan, 2010.
[12] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) Explorations Newsletter*, vol. 11, pp. 10–18, Nov. 2009.
[13] J. D. Murray, *Mathematical Biology*. Springer, 2002, vol. 3.