Ideal Composition of a Group for Maximal Knowledge Building in Crowdsourced Environments

Anamika Chhabra[1]**, S. R. S. Iyengar and Jaspal S. Saini**

anamika.chhabra@iitrpr.ac.in, sudarshan@iitrpr.ac.in, jaspal.singh@iitrpr.ac.in
Department of Computer Science and Engineering
Indian Institute of Technology Ropar, India

Abstract

Crowdsourcing has revolutionized the process of knowledge building on the web. Wikipedia and StackOverflow are witness to this uprising development. However, the dynamics behind the process of crowdsourcing in the domain of knowledge building is an area relatively unexplored. It has been observed that an ecosystem exists in the collaborative knowledge building environments (KBE) (Chhabra, Iyengar, Saini, & Bhat, 2015), which puts users of a KBE into various categories based on their expertise. Classical cognitive theories indicate triggering among the knowledge units to be one of the most important reasons behind accelerated knowledge building in collaborative KBEs. We use the concept of ecosystem and the triggering phenomenon to highlight the necessity for the right mix of users in a KBE. We provide a hill climbing based algorithm which gives the ideal mixture of users in a KBE, given the amount of triggering that takes place among the users. The study will help the portal designers to accordingly build suitable crowdsourced environments.

Keywords: Crowdsourcing; knowledge building; ecosystem; triggering

Introduction

Due to the introduction of Web 2.0 and Web 3.0 (Lassila & Hendler, 2007; Hendler, 2009), a collection of tools for collaboration, integration and interaction have become available on the internet (Bryant, 2006). Accessing knowledge on any given topic is not a difficult task nowadays. The online knowledge building systems have become successful largely because of crowdsourcing. It is a technique provided by Web 2.0 which is used to gather information from the people in the crowd. In a crowdsourced knowledge building system, the knowledge-generation process is outsourced to a community of users (Doan, Ramakrishnan, & Halevy, 2011). Some of the knowledge building systems effectively exploiting the benefits of crowdsourcing are Wikipedia, StackOverflow and Quora. These systems aggregate the human knowledge on various topics and have been successful in making use of the immense potential of the masses. Many of these systems also make use of various incentivizing mechanisms to motivate users to participate more in the knowledge building process (Anderson, Huttenlocher, Kleinberg, & Leskovec, 2013). Due to the ease in accessing internet, the process of knowledge building has further evolved (Saxton, Oh, & Kishore, 2013). Even knowledge seekers are participating in the knowledge building process by, for example, asking questions (Wang, Gill, Mohanlal, Zheng, & Zhao, 2013; Hanrahan, Convertino, & Nelson, 2012). These questions then trigger other users to add their knowledge to the system (Dierkes,

Child, & Nonaka, 2003; Anderson, Huttenlocher, Kleinberg, & Leskovec, 2014). Prime examples include StackOverflow and Quora. Group knowledge building by a mixture of experts and non-experts has mostly replaced individual knowledge building by experts (Kittur & Kraut, 2008; Galton, 1907; Surowiecki, Silverman, et al., 2007). All this has inspired us to look into the dynamics of group knowledge building. We begin by addressing the following set of questions: Why is it that groups perform better than individuals in knowledge building? In order to have the best knowledge building experience, should we consider the type and number of participating users? In general, does the composition of the group matter?
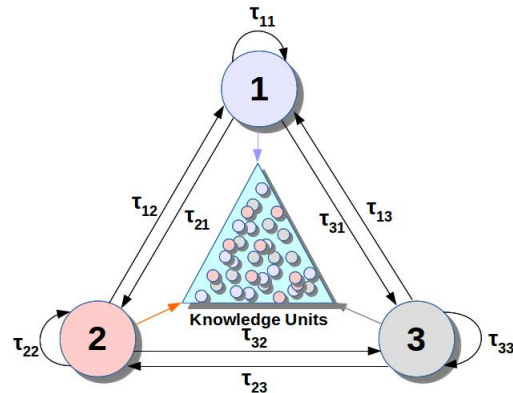

Figure 1: The Triggering Phenomenon (A knowledge building system consisting of three categories getting triggered by each other)

To investigate these questions, our analysis is based on the existence of ecosystem (Chhabra, Iyengar, Saini, Bhat, & Kumar, 2015) in crowdsourced KBEs, Luhmann's theory of autopoietic systems (Seidl, 2004; Luhmann, 1995) and Piaget's theory of equilibration (Piaget, 1977, 1976). We also follow on classical literature on the triggering phenomenon occurring among the knowledge frames. In (Chhabra, Iyengar, Saini, & Bhat, 2015), the authors observed the existence of ecosystem among the users of a knowledge building system. This ecosystem divides the users into various categories based on their expertise. Each of these categories participates in the knowledge building process in a different manner. For example, some users have a trait of asking questions, some take pride in answering others' questions, while some others like articulating the given text in a better way. There is

yet another group of users, who contribute to the knowledge building process by just reading and voting others' posts. Following Luhmann's theory, we propose that groups in a knowledge building system perform better, because these categories create an enormous amount of perturbation among each other. This perturbation triggers more ideas into the KBE. The theory when applied to modern knowledge building systems gives an insight into why people participate in these crowdsourced systems. The irritation that is created on coming across some content on these systems inspires them to contribute more. According to Piaget's theory, this phenomenon of triggering continues until the state of equilibration, after which the triggering reduces. This theory explains the reasons behind the slowing growth of content on Wikipedia with time (Suh, Convertino, Chi, & Pirolli, 2009; Lam & Riedl, 2011).

The paper thoroughly investigates the crowdsourced knowledge building process in terms of ecosystem and triggering. We differentiate between the knowledge added by users without looking at each others' ideas and the knowledge added when they come across the content posted by others. We see that the latter constitutes a considerable part of the knowledge built by a crowdsourced system. The proposed model helps in understanding the reason behind the benefits provided by a group, as compared to individual knowledge building. We further show that it is important to have the right mix of users in a KBE for effective knowledge building. For example, users possessing good subject matter knowledge may not add it to the system until instigated. This instigation is provided by the users asking meaningful questions. In this way, less knowledgeable users also contribute to the knowledge building process.

Outline. The remainder of the paper is organized as follows: It first discusses related work on knowledge building. It then explains certain classical theories which have inspired the development of the current model followed by the triggering model. The paper further explains an algorithm to find the ideal distribution of users given the triggering matrix. In the end, it concludes and discusses directions for future work.

Related Work

In the past, most of the models on knowledge building have focused on various characteristics that a KBE should possess. However, the characteristics of users that are participating in the knowledge building process have not been well understood.

Nonaka (1994) explains that the knowledge creation happens through a constant dialogue between tacit knowledge[2] and explicit knowledge[3]. The author developed a theoretical framework which provides an analytical perspective on various dimensions of knowledge creation. Gerry Stahl (2000) provides a conceptual framework for collaborative KBEs

which consists of important phases that should be supported by any computer-based KBE. The model given by the author explains the relationship of collaborative group processes to individual cognitive processes. Scardamlia and Brieter (1994) in their work consider learners also as important members of knowledge building community. They emphasized the importance of a community effort rather than individual effort for accelerated knowledge building. Cress et al. (2008) provide a theoretical framework for describing the process of learning and knowledge building and the way these two processes influence each other. They make use of Luhmann theory and Piaget's theory of equilibration and propose two processes externalization[4] and internalization[5] as the basis of interaction between the social and the cognitive system. Minsky (1977) and Rumelhart (1991) consider the knowledge units as frames. Norman (1981) and Just et. al (1980) assert that the frames are linked together and one frame may trigger other frames. When a frame is triggered, all the frames linked to it are triggered. Fisher and Lipson (1985) state that for a given frame to come into the system, suitable triggering conditions are required.

A study which is conducted in the context of a right mix of users in the domain of problem solving has been performed by Scott (2004; 2008). The author states that a group of randomly selected people outperforms a group of best performing people. This is due to the fact that these random people bring diverse knowledge into the system and hence are able to perform better while solving a problem. On contrary, the best performing agents bring in similar type of knowledge to the system and hence might not be able to solve the problem that well. However, Thompson (2014) came up with a counter paper to Scott's work recently claiming that his paper does not provide any foundation for the argument that diversity actually trumps ability. Krause et al. (2011) argue that adding diversity to a group can be more advantageous than adding expertise to the group. Erickson et. al (2012) provide a framework to select the crowd matching organizational needs. The authors state that different tasks require different crowds with different skills and knowledge. Kobern et. al (Kobren, Tan, Ipeirotis, & Gabrilovich, 2015) observe in their recent work that intelligently assigning tasks to the users significantly increases the value of a crowdsourced system.

The Triggering Model

This section first introduces the motivation behind the model by presenting classical theories which explain the phenomenon of triggering among the knowledge units and an overview of the concept of existence of ecosystem in crowdsourced environments. It then describes the types of KUs followed by the problem statement. The Triggering Model and the algorithm are explained afterwards.

---

[2]It is codified knowledge which is transmittable in some formal language

[3]It is the knowledge that human beings possess and acquire through their experience over time

[4]Externalization is a process by which users add their knowledge to the system

[5]Internalization is the process of taking the information from the system

Motivation

This subsection introduces Luhmann theory, Piaget's theory of equilibration and the previous work on ecosystem.

Elementary Theories on Cognition and Triggering Luhmanns theory describes a knowledge building system as an autopoietic system [6]. It states that once started, further ideas (or cognitions) are produced by existing ideas of the same system. The existing ideas create perturbations in the cognitive system of users, which then trigger more ideas. The theory clearly distinguishes cognitive systems from the KBE. We apply the same concept to the knowledge building process, where users are the cognitive systems who build knowledge in the KBE. The changes in the environment lead to irritations in the cognitive systems. The theory further talks about the concept of *structural coupling*, which is the relation between cognitive systems and the environment. It states that although environment can create irritations in all the cognitive systems (users here), it might not be able to trigger all of them. The actual systems that can get triggered due to these perturbations are determined by the structural coupling of these systems to the environment. Also, different cognitive systems may have different structural coupling.

The question arises whether this phenomenon of triggering goes on indefinitely, or does it reach some threshold point. Piagets Model of equilibration states that people contribute to the knowledge building process because of cognitive conflicts (perturbations as per Luhmann Model), which means that when they see some information that is incongruent to their existing knowledge, it creates a disturbance in their mind. This disturbance leads them to add their knowledge to the system. This knowledge addition then leads to the equilibration between the systems knowledge and the users knowledge. We further build on this theory explaining that initially, there is a huge need for equilibration, hence enormous amount of knowledge gets added to the system. However, as time passes, users knowledge starts matching with the knowledge of the system, and hence, there is less disturbance. This provides some support for slowing growth of knowledge building in the KBEs with time.

Lateral thinking and triggering process (Rumelhart, 1991) are the main reasons for generation of ideas in a collaborative KBE. Triggering is a procedure by which an idea or a comment spearheads the generation of another idea or a thought (Just & Carpenter, 1980). It is associated with how an idea or a comment becomes an impetus for the generation of other ideas. The generation of knowledge units in terms of triggering has also been explained in the information processing theory (Norman, 1981; Seidl, 2004). This theory characterizes human knowledge as a series of Knowledge Frames. The frames are related to each other by various conditions. When a frame is triggered by stimuli, other frames, which are linked to this frame, may also be triggered. The frames play an important role in guiding the way for creating and retrieving more knowledge.

Ecosystem and Formation of Categories In (Chhabra, Iyengar, Saini, & Bhat, 2015), the authors conducted experiments on a custom-made annotation system CAS and observed that people showed expertise in performing mainly one kind of activity while participating in the knowledge building process. The existence of ecosystem was observed in Wikipedia and StackOverflow as well (Chhabra, Iyengar, Saini, Bhat, & Kumar, 2015). This phenomenon divides all the users into various categories based on their expertise. For example, the categories observed in CAS (Crowdsourced Annotation System) were Articulators, Probers, Solvers and Explorers. There were different categories observed in the case of Wikipedia and StackOverflow. These observations led to the conclusion that ecosystem is a characteristic of crowdsourced KBEs. The categories formed in the ecosystem trigger each other with different triggering factors. Figure 1 shows the triggering phenomenon occurring in a three category system. τ_{ij} is the triggering factor, which quantizes the number of knowledge units of category i that will be generated per user of category i due to one knowledge unit of category j. This gives rise to a triggering matrix storing the triggering factors for all the categories. We intuitively believe that some values in this matrix will be low, high or some will even be zero, depending on whether one category is a prerequisite for the generation of KUs of another category or not. For example, in CAS, a question triggers users in 'Solver' category more than it may trigger the users in 'Pointer' category. We call triggering within categories *'Intra-triggering'* and across categories as *'Inter-triggering'*. The values of intra-triggering are perceived to be less than those of inter-triggering due to the similarity of traits among the users of same category. The self loops of categories in the diagram represent intra-triggering.

A collaborative knowledge building system, due to the presence of ecosystem, provides an ideal environment for the triggering process to take place. The users in one category trigger the users in all other categories. Due to this triggering process only, these collaborative environments are able to realize 'the whole is greater than sum of its parts' phenomenon.

Sources of Knowledge Units in a Knowledge Building System

We can classify all the knowledge generated in the system on the basis of whether it is an outcome of group dynamics or not. The total generated knowledge can be divided into two categories: (i) Internal Knowledge and (ii) Triggered Knowledge.

Internal Knowledge Internal knowledge is a subset of the user's knowledge which is added to the system independent of the effect of group dynamics. This is precisely the knowledge that the user would have added to the system if she had been participating in the knowledge building process individ-

---

[6]Autopoietic system refers to a system which once started, keeps recreating and maintaining itself.

ually (and not in a group). Hence, addition of internal knowledge to the system consists of only the process of externalization. As an example, consider the following experiment: If a user is asked to name all the countries in the world (which are more than 190 in total), assume, she is able to come up with 40-50 of these countries. These generated knowledge units are what we consider as her internal knowledge. Please note that in this case, it may so happen that the user knows some more countries' names, but currently she does not recall them. These names are not a part of internal knowledge since they never got added to the system.

Note that different users may have overlapping internal knowledge. This hints us towards the fact that there exits an upper bound on the internal knowledge contribution of a particular category, i.e. after a certain threshold number of users of category i, adding more users to this category will not lead to more internal knowledge contribution of the category. This shows that the 'law of diminishing benefits' holds here rather than the 'law of large numbers'. This information may help us to find an upper bound on the maximum number of users in a particular category.

Triggered Knowledge This is the kind of knowledge that gets added to the system as a result of the group dynamics. When people participate in the knowledge building process as a group, they get triggered on seeing each others' ideas and hence, generate more knowledge. This knowledge is called triggered knowledge. Addition of triggered knowledge to the system is a process, in which first internalization and then externalization takes place. The users first internalize the knowledge from the system, and in turn externalize more knowledge to thee system. This externalized knowledge can further be of two types:

1. First type consists of the knowledge that the user had in her cognitive system already, but she gets reminded of it only after getting triggered by some other user's idea. In our previous example on countries' names, for example, it may so happen that on hearing about some country of Europe, she gets reminded of some other countries' names from Europe. This type of knowledge is called *'Recollected Knowledge'*.

2. The other type consists of the new knowledge which was not in users' cognitive system already. This type of knowledge has also been referred to as *Emergent Knowledge* (Cress & Kimmerle, 2008). For example, consider the scenario where a user knows these two facts: (a) There are infinite Fermat Numbers. (b) Every number can be uniquely decomposed into a product of prime numbers (Fundamental Theorem of Arithmetic). Now consider that she comes across another fact which was added to the system by some other user, which is, (c) Every two Fermat Numbers are coprime to each other. The user combines her existing knowledge ((a) and (b)) and the knowledge internalized from the system ((c)) and comes up with a new knowledge, for example, that there are infinite prime numbers. This new

deduced knowledge when added to the system, also falls under the category of triggered knowledge.

Problem Statement

Problem: *Given the number of users, the number of categories, the total internal knowledge of the users and the amount of triggering taking place among the categories, can we output an ideal number of users for each category that leads to the maximum knowledge building and hence accelerates the knowledge building process?*

To motivate the discussion of the problem, let us assume a simplistic two-category scenario as shown in the figure 2. As we can see that when the users of category B add a KU to the system, the users of category A are triggered to a very less extent (i.e. 0.1). So does this mean that we should keep as less as possible number of users in category A However, we also see that whenever users of category A add a KU to the system, it triggers the users of category B to a great extent (i.e. 0.9). Now, if it is given that we can take only 100 total users in the system, how many users should be assigned to both the categories so that the total knowledge generated is maximum. Surprisingly, in this case, 50 users in each of the categories will lead to the maximum knowledge building if it is given that all the users have same amount of internal knowledge. In general considering all the parameters and the amount by which various categories trigger each other, can we find out an ideal number of users that should be assigned to each category for maximum knowledge building?
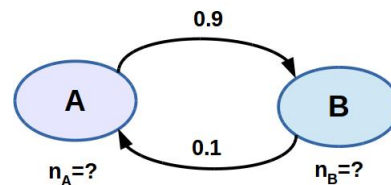

Figure 2: A Simplistic two-category Case

To get an answer to the problem explained above, we will first build a model of crowdsourced knowledge building which makes use of the triggering phenomenon, ecosystem and the two types of knowledge units (internal and triggered). With the help of this model, we will show that different distributions of users across categories gives rise to varying amount of total knowledge produced. We will then develop an algorithm that takes the triggering matrix along with the internal knowledge as input and gives the ideal distribution of users leading to the maximal knowledge building.

The Model

Consider a knowledge building system (KBS) consisting of n users. We assume that ecosystem exists in the KBS i.e. each users tends to be an expert *mainly* in one particular type of activity. Further we consider a stricter version of the above scenario where each user belongs to *only* one category i.e. a user will generate only one type of knowledge units, although

it can trigger users to generate varied types of knowledge units. This assumption has been made to reduce the complexity of the model. In practice, it may happen that the users contribute to rest of the categories to a small extent. Given that the current model mainly concentrates on the triggering phenomenon in KBS, incorporation of this small amount of contribution of knowledge to the other categories will make the model extremely complex, and may be considered in the next version of the model. We assume that all the n users enter the knowledge building system at time $t = 0$. For example, in settings like NB (Zyto, Karger, Ackerman, & Mahajan, 2012) and CAS (Chhabra, Iyengar, Saini, & Bhat, 2015), all the users are usually present in the system at the start of the knowledge building process. Throughout the text, we will be using the words 'type' and 'category' interchangeably.

Let n_i be the number of users in category i and m be the number of categories such that,

$$\sum_{i=1}^{m} n_i = n$$

Let r_i be the internal knowledge contribution of a user of category i. For simplicity, we assume that all the users of a category add same amount of internal knowledge to the system. This assumption has been taken so that the model can concentrate on explaining the most important feature which is triggering among the categories. This assumption, however, can be easily relaxed without significant changes to the model. The users of a category will add their internal knowledge to the system in discrete time stamps depending on the type of knowledge building system under consideration. In some cases, they may add it in the beginning. In other cases, they may keep adding it gradually to the system as some function of time. Let $r_i(t)$ be the internal knowledge contribution of a user of category i at time t such that,

$$\sum_{t=0}^{\infty} r_i(t) = r_i$$

Let τ_{ij} be the number of knowledge units (KUs) of type i that get triggered due to one KU of type j. Now, our aim is to calculate the total number of KUs of type i that get added to the KBS at time t. Let $k_i(t)$ represent that number. We consider that the number of KUs of a category that get added at time t are directly dependent on the following parameters:

1. The number of KUs of all the categories[7] that get added to the system at time $t - 1$. This is because of the fact that all the categories' KUs will trigger new KUs of the considered category at the next time step.

2. The triggering factors from all other categories to the considered category.

3. The number of users in the considered category. More the number of users, more the KUs generated[8].

4. The internal knowledge of the users of the considered category.

These parameters give rise to the following expression for $k_i(t)$,

$$k_i(t) = n_i(\tau_{i1}.k_1(t-1) + \tau_{i2}.k_2(t-1) + \cdots + \tau_{im}.k_m(t-1)) + n_i.r_i(t)$$

i.e.,

$$k_i(t) = n_i\left(\sum_{j=1}^{m} \tau_{ij}.k_j(t-1) + r_i(t)\right) \quad (1)$$

$\forall 0 \leqslant i \leqslant m$

The above expression gives us the number of KUs of category i that get added to the system. We are now interested to similarly compute the KUs of all the categories that get added to the system at time t.

Let $K(t)$ represent the column vector consisting of the knowledge generated by various categories at time t as its elements.

$$K(t) = \begin{bmatrix} k_1(t) \\ k_2(t) \\ \vdots \\ k_m(t) \end{bmatrix}$$

Let T be the triggering matrix, N be a diagonal matrix storing the number of users in each category and $R(t)$ be the column matrix storing the function by which each category users add their internal knowledge to the system. The matrices T, N and $R(t)$ are shown as below:

$$T = \begin{bmatrix} \tau_{11} & \tau_{12} & \cdots & \tau_{1m} \\ \tau_{21} & \tau_{22} & \cdots & \tau_{2m} \\ \vdots & & & \vdots \\ \tau_{m1} & \cdots & \cdots & \tau_{mm} \end{bmatrix}$$

$$N = \begin{pmatrix} n_1 & & & & \\ & n_2 & & \mathbf{0} & \\ & & \ddots & & \\ & \mathbf{0} & & n_{m-1} & \\ & & & & n_m \end{pmatrix}$$

$$R(t) = \begin{bmatrix} r_1(t) \\ r_2(t) \\ \vdots \\ r_m(t) \end{bmatrix}$$

Given the above notations, we can write equation 1 as:

$$K(t) = N(T.K(t-1) + R(t)) \quad (2)$$

---

[7]Including the category under consideration, due to some amount of intra-triggering

[8]Upto a certain threshold, after which diminishing benefits appear due to overlap of knowledge of users of the same category

where $K(t)$, N, T and $R(t)$ are the matrices as defined above. The above equation gives us a recursive formula for finding the total number of KUs added to the system at time t. Our aim now is to calculate the total KUs of each category added to the system, when N, T and R are given, where R is defined as below:

$$R = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix} \quad (3)$$

where r_is represent total internal knowledge of category i. We will soon show that the total knowledge generated in the system is independent of the function $r(t)$ for any category, i.e. in whichever way the users may add their internal knowledge to the system, the total knowledge generated in the system over a sufficiently long period of time in all cases comes out to be equal.

Lemma 1 *The total knowledge generated in the system at time t is given by:*

$$K(t) = \sum_{i=0}^{t} (N.T)^i.N.R(t-i)$$

Proof. Substituting the value of $K(t-1)$ in equation 2, we get,

$$K(t) = N(T.(N.T.K(t-2) + N.R(t-1)) + R(t))$$
$$K(t) = (N.T)^2.K(t-2) + (N.T).N.R(t-1) + N.R(t)$$

Continuing like this, we get,

$$K(t) = (N.T)^t.K(0) + (N.T)^{t-1}.N.R(1) +$$
$$(N.T)^{t-2}.N.R(2) + \cdots + (N.T)^0.N.R(t)$$

Since at time $t = 0$, no triggering happens, and the total knowledge of the system is given by the internal knowledge of all the users, we take $K(0) = N.R(0)$, which gives us,

$$K(t) = (N.T)^t.N.R(0) + \cdots + (N.T)^0.N.R(t)$$
$$K(t) = \sum_{i=0}^{t} (N.T).N.R(t-i)$$

$\square$

In order to get the total KUs ever generated in the KBS, we consider another matrix K as follows:

$$K = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_m \end{bmatrix}$$

where k_is are scalar values storing the total number of KUs of category t ever generated in the KBS.

Theorem 1 *Given the matrices N, T and R, the net knowledge in the system at the end of the Knowledge building process is given by,*

$$K = (I - N.T)^{-1}.N.R$$

assuming

$$\rho(N.T) < 1$$

where I is an Identity Matrix of the order $m \times m$ and ρ is teh spectral radius.

Proof. Lemma2 gives the KUs generated at time t for all the categories. Summation of $K(t)$ over the time period from 0 to ∞ gives,

$$K = \sum_{t=0}^{\infty} \left(\sum_{i=0}^{t} (N.T)^i.N.R(t-i) \right)$$
$$K = \sum_{i=0}^{t} \sum_{t=0}^{\infty} (N.T)^i.N.R(t-i)$$
$$K = \sum_{i=0}^{t} (N.T)^i.N.R, \text{ where } R \text{ is as defined in (3)}$$

If the spectral radius of $(N.T)$ is less than 1, $\sum_{i=0}^{t}(N.T)^i$ converges to $(I - N.T)^{-1}$, i.e.

$$K = (I - N.T)^{-1}.N.R,$$

where I is an Identity matrix. $\square$

In the above proof, $\rho(N.T)$ has been taken to be less than 1. This is because we are assuming a bounded system. If we take ρ to be greater than or equal to 1, then the term $\sum_{i=0}^{t}(N.T)^i$ increases exponentially, which is not a practical scenario.

The above theorem shows that the total knowledge in the system is independent of the distribution of $R(t)$ i.e. whether the entire internal knowledge of users enters the system at $t = 0$ or it enters in discrete time, the net knowledge in the system remains unchanged. The theorem also helps computing the triggered knowledge contribution in the knowledge building, as shown by the corollary that follows.

Corollary 1 *The contribution of triggered knowledge to the total knowledge of the system is given by N.T.K*

Proof. From Theorem 1,

$$K = (I - N.T)^{-1}.N.R$$
$$\implies (I - N.T).K = N.R$$

The contribution of triggered knowledge in K is equal to $K - N.R$, where $N.R$ is the total internal knowledge added to the system.

$$\implies K - N.T.K = N.R$$
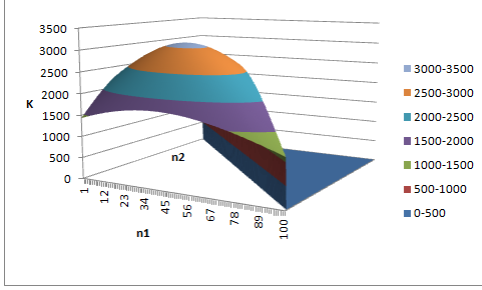$$\implies K - N.R = N.T.K$$

$\square$

Figure 3: The convex plot of net knowledge (K) generated across all the possible distributions

Intuitively, the corollary indicates that the ratio of triggered knowledge to the total knowledge is directly proportional to the triggering matrix T. This means that the total knowledge that a system is able to produce depends on how much the users of the system are able to trigger each other.

In order to know the shape of the surface curve when we change the number of users across the categories, we simulated the results of Theorem 1 by taking the parameters: $n = 100$, $m = 3$ and

$$T = \begin{bmatrix} 0.001 & 0.003 & 0.0004 \\ 0.03 & 0.001 & 0.005 \\ 0.03 & 0.008 & 0.003 \end{bmatrix}$$

For simplicity, the internal knowledge (R) for all the categories was taken to be equal. It was observed that the surface plot comes out to be a convex plot which exhibits a peak corresponding to a particular distribution of users. Figure 3 shows the plot where the axes for the represent n_1, n_2 and K (n_3 is taken as $(n - n_1 - n_2)$). The maxima of the curve was found at $K = 3081.40849516$ with the distribution $[39, 30, 31]$.

Hill Climbing Algorithm for Ideal Skillset Distribution

Since we obtained a convex surface plot with respect to the distributions, we can develop a hill climbing algorithm which starts from any random distribution and gradually moves towards the ideal distribution of users which leads to an accelerated knowledge building. We now provide a 'Hill Climbing Algorithm' (See Algorithm 1) for finding the ideal skill set distribution across categories.

For verification, we simulated the proposed algorithm on the same specifications as defined above. Figure 4 shows how the algorithm moves from an initial distribution (which is input to the algorithm) to ideal distribution i.e. the distribution which produces maximum knowledge in the given KBE. The final output of the algorithm ($[39, 30, 31]$) is coherent with results obtained from the previous simulation, supporting the correctness of the proposed algorithm.

Conclusion and Future Work

This paper is a step towards understanding the dynamics of knowledge building in a crowdsourced environment. We con-

---

Algorithm 1 Hill Climbing Algorithm For Ideal Distribution of Users across Categories

Input: Number of users n, Number of categories m, Triggering Matrix T and Internal Knowledge matrix R.
Output: Ideal Distribution Matrix N

1. Pick a random distribution $N = [n_1, n_2, \ldots, n_m]$ of users across different categories, such that $\sum_{i=1}^{m} n_i = n$

2. Consider all the $2 * \binom{m}{2} = m(m-1)$ neighbouring distributions (Nb) of D, defined as:

$$Nb = [\,]$$
$$for\ i = 1 : m$$
$$\quad for\ j = 1 : m$$
$$\quad\quad If\ i \neq j :$$
$$\quad\quad\quad D = (n_1, n_2, \ldots, n_i - 1, \ldots, n_j + 1, \ldots, n_m)$$
$$\quad\quad\quad Nb.append(D)$$

3. Calculate $K_{Nb[j]} \ \forall\ 1 \leqslant j \leqslant m(m-1)$ based on the following formula:

$$K = (I - N.T)^{-1}.N.R$$

where $K_{Nb[j]}$ represents the net knowledge in the system with the distribution $Nb[j]$

Choose i such that $K_{Nb[i]} = \max_{j} K_{Nb[j]}$

4. If $K_{Nb[i]} > K_N$ then $N = Nb[i]$ and we repeat steps 2 to 4 else:
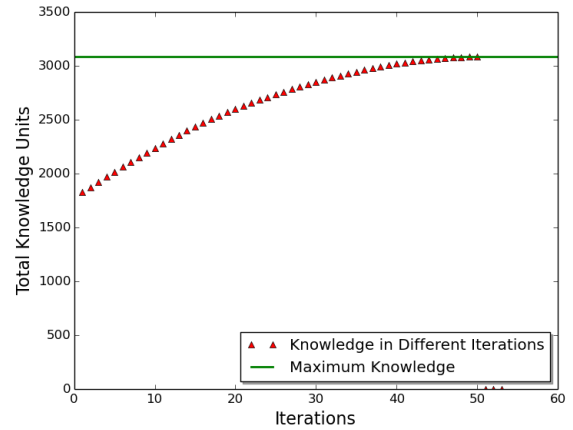   return N //The ideal skill set distribution

---

Figure 4: The execution of Hill Climbing Algorithm: moving from one distribution to another neighboring distribution, leading to the perfect one

sider the variance of expertise present in the crowd which divides the users of crowd into categories. In line with the previous research in cognitive sciences, we consider triggering among the knowledge frames to be the reason behind better performance of the groups than individuals. Through a model, we explain the interplay of these categories taking into account the varied amount of triggering that takes place among them. We then compare this triggered knowledge with the knowledge that would have been added to the system in the absence of interaction among the users. We find that triggered knowledge is a significant part of the total knowledge produced.

We emphasize the importance of a right mix of users across the categories in order to maximize the total knowledge. We observe that as we keep changing the distribution of users across categories, we get different amount of knowledge produced with respect to time. In that context, we try to find the ideal distribution which accelerates the knowledge building process to the best possible extent. We computationally observe that the knowledge curve with respect to the distribution of users is a convex curve, the peak of which gives the ideal distribution. We further develop a Hill Climbing Algorithm which starts from any random distribution and outputs the ideal distribution for the best knowledge building experience. We believe that this study will inform the portal designers to take better decisions while developing a KBE. For example, knowing the number of categories in the system and the triggering among them, they can use various incentivizing mechanisms to encourage the participation of users in different categories. This theory can also help explain why some KBEs like Wikipedia have flourished while some others like could not proliferate.

In future, we plan to verify the proposed model on various KBEs like Wikipedia, Quora, StackOverflow etc. The model could also be tested by changing the existing distribution of users in a KBE, and observing the impact of this change towards the acceleration/deceleration in the knowledge building process. Knowledge building is a complex phenomenon having a number of parameters. It is not possible to build a single model considering all the parameters. The current version considers the triggering among the users as the most important parameter while making certain assumptions for the others. For example, all the users are assumed to be present in the system at the start of the knowledge building process, which may not always be the case. We also assume that every user belongs to precisely one category. Although the users possess traits making them part of mainly one category, they may contribute to a small extent in rest of the categories as well. The current model also does not consider the overlap among the knowledge of various users, which will give an upper bound on the users of a particular category. As a next step, we plan to consider some of these assumptions in the model.

References

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2013). Steering user behavior with badges. In *Proceedings of the 22nd international conference on world wide web* (pp. 95–106).

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with massive online courses. In *Proceedings of the 23rd international conference on world wide web* (pp. 687–698).

Bryant, T. (2006). Social software in academia. *Educause quarterly*, *29*(2), 61.

Chhabra, A., Iyengar, S., Saini, P., & Bhat, R. S. (2015). Presence of an ecosystem: a catalyst in the knowledge building process in crowdsourced annotation environments. *Proceedings of the 2015 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2015)*.

Chhabra, A., Iyengar, S., Saini, P., Bhat, R. S., & Kumar, V. (2015). Ecosystem: A characteristic of crowdsourced environments. *arXiv preprint arXiv:1502.06719*.

Cress, U., & Kimmerle, J. (2008). A systemic and cognitive view on collaborative knowledge building with wikis. *International Journal of Computer-Supported Collaborative Learning*, *3*(2), 105–122.

Dierkes, M., Child, J., & Nonaka, I. (2003). *Handbook of organizational learning and knowledge*. Oxford University Press.

Doan, A., Ramakrishnan, R., & Halevy, A. Y. (2011). Crowdsourcing systems on the world-wide web. *Communications of the ACM*, *54*(4), 86–96.

Erickson, L., Petrick, I., & Trauth, E. (2012). Hanging with the right crowd: Matching crowdsourcing need to crowd characteristics.

Fisher, K., & Lipson, J. I. (1985). Information processing interpretation of errors in college science learning. *Instructional Science*, *14*(1), 49–74.

Galton, F. (1907). Vox populi (the wisdom of crowds). *Nature*, *75*, 450–451.

Hanrahan, B. V., Convertino, G., & Nelson, L. (2012). Modeling problem difficulty and expertise in stackoverflow. In *Proceedings of the acm 2012 conference on computer supported cooperative work companion* (pp. 91–94).

Hendler, J. (2009). Web 3.0 emerging. *Computer*, *42*(1), 111–113.

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(46), 16385–16389.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological review*, *87*(4), 329.

Kittur, A., & Kraut, R. E. (2008). Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proceedings of the 2008 acm conference on computer supported cooperative work* (pp. 37–46).

Kobren, A., Tan, C. H., Ipeirotis, P., & Gabrilovich, E. (2015). Getting more for less: Optimized crowdsourcing with dynamic tasks and goals. In *Proceedings of the 24th international conference on world wide web* (pp. 592–602).

Krause, S., James, R., Faria, J. J., Ruxton, G. D., & Krause, J. (2011). Swarm intelligence in humans: diversity can trump ability. *Animal Behaviour*, *81*(5), 941–948.

Lam, S. K., & Riedl, J. (2011). The past, present, and future of wikipedia. *Computer*, *44*(3), 87–90.

Lassila, O., & Hendler, J. (2007). Embracing" web 3.0". *Internet Computing, IEEE*, *11*(3), 90–93.

Luhmann, N. (1995). *Social systems*. Stanford University Press.

Minsky, M. (1977). Frame-system theory. *Thinking: Readings in cognitive science*, 355–376.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization science*, *5*(1), 14–37.

Norman, D. A. (1981). Categorization of action slips. *Psychological review*, *88*(1), 1.

Page, S. E. (2008). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton University Press.

Piaget, J. (1976). *Piagets theory*. Springer.

Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures.(trans a. rosin)*. Viking.

Rumelhart, D. E. (1991). Understanding understanding. *Memories, thoughts and emotions: Essays in honor of George Mandler*, 257–275.

Saxton, G. D., Oh, O., & Kishore, R. (2013). Rules of crowdsourcing: Models, issues, and systems of control. *Information Systems Management*, *30*(1), 2–20.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The journal of the learning sciences*, *3*(3), 265–283.

Seidl, D. (2004). Luhmanns theory of autopoietic social systems.

Stahl, G. (2000). A model of collaborative knowledge-building. In *Fourth international conference of the learning sciences* (Vol. 10, pp. 70–77).

Suh, B., Convertino, G., Chi, E. H., & Pirolli, P. (2009). The singularity is not near: slowing growth of wikipedia. In *Proceedings of the 5th international symposium on wikis and open collaboration* (p. 8).

Surowiecki, J., Silverman, M. P., et al. (2007). The wisdom of crowds. *American Journal of Physics*, *75*(2), 190–192.

Thompson, A. (2014). Does diversity trump ability? *Notices of the AMS*, *61*(9).

Wang, G., Gill, K., Mohanlal, M., Zheng, H., & Zhao, B. Y. (2013). Wisdom in the social crowd: an analysis of quora. In *Proceedings of the 22nd international conference on world wide web* (pp. 1341–1352).

Zyto, S., Karger, D., Ackerman, M., & Mahajan, S. (2012). Successful classroom deployment of a social document annotation system. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1883–1892).