# Exploring Social Influence on the Usage of Resources in an Online Learning Community


Ogheneovo Dibie
Institute of Cognitive Science
Dept. of Computer Science
University of Colorado
Boulder, Colorado USA
ogdi0204@colorado.edu

Tamara Sumner
Institute of Cognitive Science
Dept. of Computer Science
University of Colorado
Boulder, Colorado USA
sumner@colorado.edu

Keith Maull
University Corporation for
Atmospheric Research
Boulder, Colorado USA
kmaull@ucar.edu

David Quigley
Institute of Cognitive Science
Dept. of Computer Science
University of Colorado
Boulder, Colorado USA
david.quigley@colorado.edu



## ABSTRACT

This research investigates the usage distribution of instructional resources shared among educators in an online learning community. The usage of a resource is defined by the number of unique educators who use (click on) it. We explored what the usage distribution of these resources looks like and we investigated what underlying mechanisms may have generated the observed distribution. Our results indicate that the usage distribution of resources follows a power law. Furthermore, our results also suggest that an educator's decision to use a resource may be influenced by the prior decisions of others. 82.6% of 2500 simulations of an information cascade model developed to model the resource selection process of educators resulted in a power law distribution as observed in our data. Information cascades provide a natural way of understanding how individuals may imitate the decisions of others even when such decisions do not align with their perssonal preferences.


## 1. INTRODUCTION

Research consistently indicates that online learning communities can improve the instructional practices of educators and produce increases in student learning outcomes by providing educators with access to learning resources and best practices shared by their peers [5]. Given the importance of these community-contributed resources to educator instruction, understanding the factors that encourage their usage is an intriguing question with important implications for educator instruction, student learning and agencies that support these communities.

We explored this question in the context of a community of Earth Science educators that used an online curriculum planning tool called the Curriculum Customization Service (CCS). The CCS provides educators with access to digital versions of their class textbook, digital library resources and community-contributed resources. This study is based on 6th-9th grade Earth Science educators that shared and used community-contributed resources in the CCS over a period of four academic years.

We began by exploring the observed usage distribution of community-contributed resources in the CCS, and then turned our attention to the influence of three mechanisms on the observed usage distribution. First, we investigated how resource visibility influences resource selection—postulating that the position or rank of a resource in the list it is displayed in *may* impact selection behavior. Second, we investigated how the quality (or perceived quality) of a resource might have influence on selection. Finally, we examine how social factors, specifically how the decisions of others in the community, provide insights into the observed resource usage distribution.

## 2. METHODS AND RESULTS

We discovered that the usage distribution of community-contributed resources follows a power law. Also known as Zipf, Pareto-Levy or scale-free distributions [4], a quantity $x$ obeys a power law if it is drawn from a probability distribution $p(x) \propto x^{-\alpha}$ where $\alpha$ is known as the exponent or scaling parameter. Power laws appear in a wide array of man made and natural phenomena [3] such as the distribution of calls to telephone numbers, scientific paper citations and the frequency of use of English words [4]. We determined that the usage distribution of resources followed a power law using software implementations[1] [2] of the rigorous statistical approach of Clauset et al. [3] for detecting power laws in empirical data. [3]. Our empirical data was found to follow a power law with an $\alpha$ of 4.44 and an $x_{min}$ value of 15. Figure 1 illustrates that a power law provides a closer fit to the complimentary cumulative distribution function (CCDF) [3]

[1]plfit: https://pypi.python.org/pypi/plfit
[2]powerlaw: https://pypi.python.org/pypi/powerlaw
[3]The CCDF is defined by $\Pr(X \geq x)$

of the empirical data in comparison to the lognormal and exponential distribution.

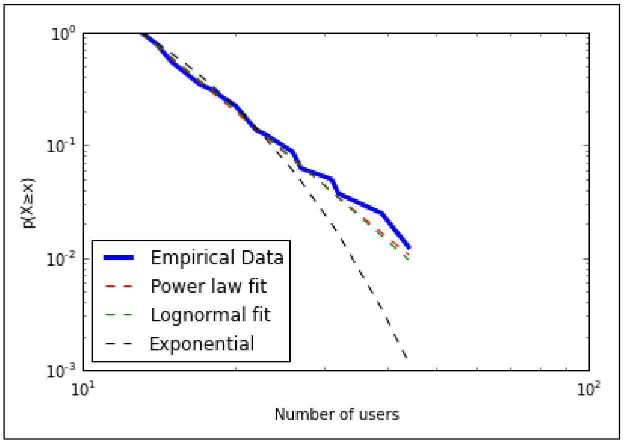


**Figure 1: Comparisons of the complimentary cummulative distribution function (CCDF) of the empirical data, the power law, lognormal and exponential distribution fits to the data.**

### 2.1 Mechanisms behind the power law distribution of resources

**Resource position**: Correlation tests between the mode, median and last click position of resources and their usage show only a very weak correlation between the usage of a resource and its position during the 2012-2013 school year. This suggests that a resource position had little to no influence on usage.

**Resource quality**: We then investigated the relationship between resource quality (inferred before a user clicks on it) and its usage in two steps. First, we used the presence of a description in the listing of a resource as a marker of its quality. Thus, resources with a description were deemed as having high quality and those without a description were regarded as low quality. We then investigated if there was a statistically significant difference in usage between resources of high quality and those of low quality, and consequently discovered no statistically significant difference in usage between resources of both groups. Our next investigation into the impact of a resource's quality on usage investigated whether there was any correlation between the number of quality signals of a resource and its usage. To do this, we developed a composite resource quality score that incorporated all signals of a resource's quality that can be inferred by a user before clicking. These signals were mapped to the resource quality indicators developed by Bethard et al. [1]. Our results show only a weak correlation of 0.124 between resource quality and usage ($t = 2.8343, df = 516, p = 0.002387$)

**Social influence**: Finally, we looked at the impact of aggregate social influence on the usage of community-contributed resources. We found a statistically significant positive correlation of 0.634 at a p-value of $2.2e^{-16}$ between saves and usage. Unlike our earlier tests on position and quality, this indicates that the social influence conveyed through the saving of resources may be in part responsible for driving usage.

We then explored if an information cascade model simulating the decision making processes of educators can generate a power law usage distribution as observed in our data. Our model extends the informational cascade model of Bikchandani, Hirshleifer, and Welch (BHW) [2] in three ways. First, instead of the binary decision model of BHW, a decision will be made between $1..r$ resources at any time. Second, in contrast to the BHW model, the decision of an individual is not always visible to others as a public signal. In our context, the only public signal available is whether or not a user saves a resource. After clicking on a resource, users will leave public signals with a uniform random probably $p$. This probability is exogenously fixed at 0.41—determined from computing the ratio of saves to unique clicks on all resources across all school years. Finally, a user's private signal $p_s$ for a resource $r$ is drawn from a discrete uniform probability distribution such that $p_s \in [0, 1]$. 2500 simulations of the information cascade model described above were processed with each simulation evaluated to see if they follow a power law using the procedure of Clauset et al. [3]. Consequently, 82.6% of these simulations were determined to follow a power law distribution. The outcomes of this experiment strongly suggests that an information cascade model simulating the decision making process of educators can lead to a power law usage distribution as observed in our data. This provides strong support for the social influence hypothesis as a generative mechanism for the observed usage distribution.

## 3. DISCUSSION & CONCLUSION

For agencies that support online learning communities, this research has important implications for resource presentation and recommendation. In presenting resources, social influence signals can be de-emphasized to limit the chances that they will detract users from evaluating a resource's inherent quality. For example, in the CCS, the number of educators that have saved a resource can be hidden and require active effort from users to be revealed. In recommending resources, high quality but barely used resources can be recommended to educators in ways that give them high precedence. This could include personalized recommendations while active on the platform or email recommendations. This paper is based upon research supported by National Science Foundation awards #1043638 and #1147590.

## 4. REFERENCES


[1] Bethard, S., Wetzer, P., Butcher, K., Martin, J. H., and Sumner, T. Automatically characterizing resource quality for educational digital libraries. In *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries* (2009), ACM, pp. 221–230.

[2] Bikhchandani, S., Hirshleifer, D., and Welch, I. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy* (1992), 992–1026.

[3] Clauset, A., Shalizi, C. R., and Newman, M. E. Power-law distributions in empirical data. *SIAM review 51*, 4 (2009), 661–703.

[4] Easley, D., and Kleinberg, J. *Networks, crowds, and markets: Reasoning about a highly connected world.* Cambridge University Press, 2010.

[5] Louis, K. S., and Marks, H. M. Does professional community affect the classroom? teachers' work and student experiences in restructuring schools. *American journal of education* (1998), 532–575.