Testing a Power Law Model of Knowledge Propagation: Case Study of the Out of Eden Walk Project

Octavio Lerma, Leobardo Valera,
Deana Pennington, and Vladik Kreinovich
Cyber-ShARE Center
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
lolerma@episd.edu, leobardovalera@gmail.com
ddpennington@utep.edu, vladik@utep.edu

Abstract

To improve teaching and learning, it is important to understand how knowledge propagates. In general, when a new piece of knowledge is introduced, people start learning about it. Since the potential audience is limited, after some time, the number of new learners starts to decrease. Traditional models of knowledge propagation are based on differential equations; in these models, the number of new learners decreases exponentially with time. Recently, a new power law model for knowledge propagation was proposed. In this model, the number of learners decreases much slower, as a negative power of time. In this paper, we compare the two models on the example of readers' comments on the Out of Eden Walk, a journalistic and educational project in which informative messages ("dispatches") from different parts of the world are regularly posted on the web. Readers who learned the new interesting information from these dispatches are encouraged to post comments. Usually, a certain proportion of readers post comments, so the number of comments posted at different times can be viewed as a measure characterizing the number of new learners.

So, we check whether the number of comments is consistent with the power law or with the exponential law. To make a statistically reliable conclusion on which model is more adequate, we need to have a sufficient number of comments. It turns out that for the vast majority of dispatches with sufficiently many comments, the observed decrease is consistent with the power law (and none of them is consistent with the exponential law).

1 Formulation of the Problem

Traditional models of knowledge propagation. To improve teaching and learning, it is important to understand how knowledge propagates. Traditional models of knowledge propagation are based on differential equations, similar to epidemics propagation; see, e.g., [1, 3, 9].

Power law models of knowledge propagation. Some empirical data suggests that the spread follows the power law [1, 5], when the number of people who learned the new material is proportional to some power t^α of time t.

Power laws in general and power law as a description of knowledge propagation: successes and challenges. Power laws are ubiquitous in real life; these laws underlie the *fractal* techniques which has been successful in many application areas; see, e.g., [2, 8, 12].

In many applications, the success of power laws is not just an empirical fact, there are convincing theoretical explanations for their ubiquity. Such explanations exist for knowledge propagation as well: for example, in [7], we used the general motivations for power laws to provide a theoretical explanations for the success of power law in describing knowledge propagation.

However, while the theoretical foundations for the power law are reasonably well-developed, the empirical foundations remains rather thin, since the power law was tested only on a few cases. Thus, more testing is needed.

What we do in this paper. In this paper, we empirically check whether power law is indeed a better description of knowledge propagation than models based on differential equations. For that purpose, we consider an example – comments on the messages ("dispatches") posted as a part of the Out of Eden Walk project [10]; see also [6, 11, 14].

The structure of this paper is as follows: In Section 2, we give a brief description of the case study and explain how it is related to knowledge propagation. In Section 3, we provide technical details related to the traditional and power law models of knowledge propagation. In Section 4, we describe our methodology for comparing the two models. Section 5 contains the results of this comparison. Finally, Section 6 contains conclusions and future work.

2 Description of the Case Study

Out of Eden Walk project: a description. Commenced on January 10th, 2013 in Ethiopia, the Out of Eden Walk is a 7-year, 21,000 mile long, storytelling journey created by two-time Pulitzer Prize winning journalist Paul Salopek. This project is sponsored by the National Geographic Society. Reports from this journey regularly appear in the National Geographic magazine, in leading newspapers such as New York Times, Washington Post, Chicago Tribune, Los Angeles Times, and on the US National Public Radio (NPR).

This project has important educational and knowledge propagation goals. The Out of Eden Walk is a very ambitious project, its main objective is to enhance education and knowledge propagation as main features of journalism, to reinvent digital reporting in the age of nano-headlines by embracing the concept of *slow journalism*: revealing human stories and world events from the ground, at a walking pace. The slow journalism of the Out of Eden Walk is immersive and sustained reporting, yet conveyed through the state-of-the-art digital platforms, with presence on the web, on Facebook, on Twitter, and in the traditional media.

The project has largely succeeded in these goals: now in its third year, the website has thousands of followers worldwide, not counting Facebook, Twitter, and other followers. Over 200 schools worldwide regularly use Salopek's reports as an education tool, to enrich the students' understanding of different worldwide cultures.

Out of Eden Walk project: technical details. After visiting a new geographic area, Paul Salopek selects an important topic related to this area and publishes a *dispatch* describing his impressions, experiences, and thoughts. As of January 2015, there are close to 100 dispatches.

Followers are welcome to add comments after each dispatch. After two weeks, each dispatch gathers from 15 to more than 250 comments. Many of these comments are made by teachers and students who use these dispatches as part of their learning experience.

Sometimes, Paul Salopek replies to some of these comments, often by providing additional details about the story. These replies, in their turn, elicit more comments, etc. All these comments are part of the knowledge propagation process.

What we do. In this project, we trace, for each dispatch, how the number of comments made by the readers changes with time. This number reflects how the knowledge contained in a dispatch propagates with time.

Specifically, as we mentioned earlier, we check whether this propagation is better described by a traditional model based on differential equations or by a power law. To make this comparison, let us first recall the formulas describing these two approaches to quantifying knowledge propagation.

3 Power Law Model vs. Traditional Approach: Technical Details

The power law model. The power law formulas predict that the number of comments $r(t)$ decreases with t as $r(t) = A \cdot t^{-\alpha}$. This model has two parameters A and $\alpha > 0$.

In practice, after a large period of time t, the number of new comments decreases to practically 0 – and this is exactly what the power law predicts, since $A \cdot t^{-\alpha} \to 0$ as $t \to \infty$.

The traditional model. The traditional description of knowledge propagation is based on first order differential equations. In particular, a general way to describe how the number of comments $r(t)$ changes with time is to use a differential equation $\dfrac{dr}{dt} = -f(r)$, for an appropriate function $f(r)$.

As we have mentioned, in practice, after a large period of time t, the number of new comments decreases to practically 0. In this case, we have $r(t) \approx 0$ and $\dfrac{dr}{dt} \approx 0$. So, we should have $f(0) = 0$.

In principle, we can have models of different complexity. We can have models with a linear function $f(r)$, we can have models with a more general quadratic dependence $f(r)$, etc. To make a fair comparison, we should select a class of models which is characterized by the same number of parameters as the power law model – otherwise, the traditional model will be more accurate just because we allow it to use more parameters to adjust to the data. Let us start with the simplest case of a linear function $f(r)$. A general linear function with $f(0) = 0$ has the form $f(r) = \alpha \cdot r$ for some α. For this function $f(r)$, the above differential equation has a 2-parametric family of solutions $r(t) = A \cdot \exp(-\alpha \cdot t)$.

For quadratic functions $f(r) = \alpha \cdot r + c \cdot r^2$, we already have a 3-parametric family of solutions. So, to keep our comparison fair, in this paper, we use the exponential model $r(t) = A \cdot \exp(-\alpha \cdot t)$ as a traditional model of knowledge propagation.

4 How We Compare the Two Models

Selecting a time period. Our observation is that once a dispatch is posted, there is a short period with practically no comments, then the bulk of the comments start, first with a big burst in comments, and then usually gradually decreasing.

Both power law and exponential law describe only how the number of comments decreases with time. Thus, to compare the observations with the model, we start with the day on which the most comments were posted, and considers this day and several days after that. In most cases, the vast majority of comments are posted within the first month, so we limited our data to 30 consecutive days (starting with the day in which the largest number of comments were posted).

Finding parameters of the model: Least Squares approach. In our analysis, for each of the two models, we first find the values of the parameters leading to the best fit, and then check how good is the resulting fit.

Let us start with the power law model. The number of responses fluctuates, so clearly responding is a random process. Let $p(t)$ denote the probability with which a person responds at moment t. The overall numbers of responses $r(t)$ can be viewed as a the sum $r(t)_1 + \ldots + r(t)_N$, where $r(t)_i$ is the number of responses coming from the i-th reader at moment t. Since the probability of a person responding is equal to $p(t)$, this means that each value r_i is equal to 1 with probability $p(t)$ and to 0 with the remaining probability $1 - p(t)$. One can easily see that the expected value $E[r_i]$ is then equal to $p(t)$, and the variance $V[r_i]$ is equal to $p(t) \cdot (1 - p(t))$. Since we assume that all the respondents are independent, the expected value of the sum is equal to the sum of the expected values, and the variance of the sum is equal to the sum of variances. Thus, the expected value is equal to $N \cdot p(t)$, and the variance is equal to $N \cdot p(t) \cdot (1 - p(t))$.

It is reasonable to estimate the probability $p(t)$ by the corresponding frequency, i.e., as the ratio $\dfrac{r(t)}{N}$. In this case, the variance can be estimated as $V(t) = N \cdot \dfrac{r(t)}{N} \cdot \left(1 - \dfrac{r(t)}{N}\right)$. Here, $r(t) \ll N$, so $1 - \dfrac{r(t)}{N} \approx 1$, and $V(t) \approx N \cdot \dfrac{r(t)}{N} = r(t)$. Since the standard deviation is equal to the square root of the variance, we conclude that

$$r(t) \approx r_0(t), \text{ with accuracy } \sigma(t) = \sqrt{r(t)},$$

where $r_0(t)$ denotes the corresponding model.

Finding parameters of the model: case of power model. For the power model $r_0(t) = A \cdot t^{-\alpha}$, the above formula takes the form

$$r(t) \approx A \cdot t^{-\alpha}, \text{ with accuracy } \sigma(t) = \sqrt{r(t)}. \tag{1}$$

For large N, due to the Central Limit theorem, the distribution of an approximation error is close to Gaussian, so, in principle, we can find the parameters A and α by using the Maximum Likelihood method, which in this case takes the form

$$\sum_t \frac{(r(t) - A \cdot t^{-\alpha})^2}{\sigma^2(t)} = \sum_t \frac{(r(t) - A \cdot t^{-\alpha})^2}{r(t)} \to \min .$$

From the computational viewpoint, however, this approach has a limitation. Such a Gaussian-based Maximum Likelihood (= Least Squares) approach is usually applied to situations when the model linearly depends on the parameters. In our case, however, the objective function is a quadratic function of these parameters, and so, by differentiating this objective function with respect to all these parameters and equating the resulting partial derivatives to 0, we get an easy-to-solve system of linear equations. In our case, the dependence of the model on the parameters A and α is strongly non-linear, and, as a result, we end up with a mode-difficult-to-solve system of nonlinear equations. To simplify computations, it is therefore desirable to reduce our problem to the case when the model linearly depends on parameters.

It is known that in log-log scale, the power law becomes a linear dependence. Specifically, if we take logarithms of both sides of the formula $r(t) = A \cdot t^{-\alpha}$, then we get $\ln r(t) = \ln(A) - \alpha \cdot \ln(t)$.

To use this fact, we need to find out how the inaccuracy in $r(t)$ is transformed into the inaccuracy with which we know $\ln(r(t))$. We know that the inaccuracy $\Delta r(t)$ has standard deviation $\sqrt{r(t)}$. When this inaccuracy is small, for $\ln(r(t))$, we have $\dfrac{\Delta(\ln(r(t))}{\Delta r(t)} \approx \dfrac{d(\ln(r(t))}{dr(t)} = \dfrac{1}{r(t)}$. Thus, in this approximation, $\Delta(\ln(r(t)) = \dfrac{\Delta r(t)}{r(t)}$. In general, when we multiply a random variable ξ by a positive constant c, its standard deviation $\sigma[\xi]$ is multiplied by the same constant: $\sigma[c \cdot \xi] = c \cdot \sigma[\xi]$. Thus, we conclude that $\sigma[\ln(r(t))] = \dfrac{\sigma[r(t)]}{r(t)}$. Since we already know that $\sigma[r(t)] = \sqrt{r(t)}$, we thus conclude that $\sigma[\ln(r(t))] = \dfrac{1}{\sqrt{r(t)}}$. For the power model, we have $\ln(r(t)) \approx \ln(A) - \alpha \cdot \ln(t)$, so we have

$$\ln(r(t)) \approx \ln(A) - \alpha \cdot \ln(t) \ \text{ with accuracy } \ \sigma = \frac{1}{\sqrt{r(t)}}.$$

If we multiply both sides of this approximate equality by $\sqrt{r(t)}$, we conclude that

$$\sqrt{r(t)} \cdot \ln(r(t)) \approx a \cdot \sqrt{r(t)} + b \cdot \ln(t) \cdot \sqrt{r(t)} \ \text{ with accuracy } \ \sigma = 1,$$

where we denoted $a \stackrel{\text{def}}{=} \ln(A)$ and $b \stackrel{\text{def}}{=} -\alpha$. For this problem, the Maximum Likelihood (= Least Squares) method means minimizing the sum

$$\sum_t \left(\sqrt{r(t)} \cdot \ln(r(t)) - a \cdot \sqrt{r(t)} - b \cdot \ln(t) \cdot \sqrt{r(t)}\right)^2.$$

Once we find the corresponding values of a and b, we can then find $A = \exp(a)$ and $\alpha = -b$.

Finding parameters of the model: case of exponential model. For the exponential model $r_0(t) = A \cdot \exp(-\alpha \cdot t)$, the formula for finding the parameters takes the form

$$r(t) \approx A \cdot \exp(-\alpha \cdot t), \ \text{with accuracy } \sigma(t) = \sqrt{r(t)}. \tag{2}$$

Similarly to the case of the power model, we can conclude that the distribution of an approximation error is close to Gaussian, so, we can find the parameters A and α by using the Maximum Likelihood method, which in this case takes the form

$$\sum_t \frac{(r(t) - A \cdot \exp(-\alpha \cdot t))^2}{\sigma^2(t)} = \sum_t \frac{(r(t) - A \cdot \exp(-\alpha \cdot t))^2}{r(t)} \to \min .$$

The exponential law $r(t) = A \cdot \exp(-\alpha \cdot t)$ becomes linear if we consider the dependence of $\ln(r(t))$ on time t: $\ln(r(t)) = \ln(A) - \alpha \cdot t$.

We already know, from our analysis of the power model, that $\sigma[\ln(r(t))] = \dfrac{\sigma[r(t)]}{r(t)}$. Thus, we have

$$\ln(r(t)) \approx \ln(A) - \alpha \cdot t \ \text{ with accuracy } \ \sigma = \frac{1}{\sqrt{r(t)}}.$$

If we multiply both sides of this approximate equality by $\sqrt{r(t)}$, we conclude that

$$\sqrt{r(t)} \cdot \ln(r(t)) \approx a \cdot \sqrt{r(t)} + b \cdot t \cdot \sqrt{r(t)} \ \text{ with accuracy } \ \sigma = 1,$$

where we denoted $a \stackrel{\text{def}}{=} \ln(A)$ and $b \stackrel{\text{def}}{=} -\alpha$. For this problem, the Maximum Likelihood (= Least Squares) method means minimizing the sum

$$\sum_t \left(\sqrt{r(t)} \cdot \ln(r(t)) - a \cdot \sqrt{r(t)} - b \cdot t \cdot \sqrt{r(t)} \right)^2 .$$

Once we find the corresponding values of a and b, we can then find $A = \exp(a)$ and $\alpha = -b$.

From the traditional Least Squares to the robust ℓ^1 method for parameter estimation. When a model predicts the values E_i and observations are O_i, the traditional Least Squares estimate selects the parameters for which the sum of the squares of the differences $\sum_i (O_i - E_i)^2$ is the smallest possible. As we have mentioned, this works well if the difference $O_i - E_i$ is normally distributed.

In practice, in addition to a normally distributed differences $O_i - E_i$, we also have outliers. For example, a measuring instrument may malfunction, generating a number which is far away from the actual value of the measured quantity. Such outliers can drastically change the Least Squares estimate.

For example, if the model is a constant $E_i = \text{const}$, the Least Square estimate for this constant is simply the arithmetic mean of all observed values $\frac{1}{n} \cdot \sum_{i=1}^{n} O_i$. If there are no outliers, this works well. For example, if the actual value is 0, and standard deviation is $\sigma = 1$, then after 100 observations, we get 0 with an accuracy of $\frac{\sigma}{\sqrt{n}} \approx 0.1$. However, if due to a malfunction, one of the recordings is 10000, we get the average 100.

In situations when outliers are possible, it is therefore reasonable to use methods which are less sensitive to outliers. Such methods are known as *robust*; see, e.g., [4]. One of the most widely used robust methods is the ℓ^1-*method*, when we select parameters for which the sum of absolute values $\sum_i |O_i - E_i|$ is the smallest possible.

For example, in the above example when the model is a constant, the ℓ^1 methods results in selecting a median of all the observation instead of the arithmetic mean, and one can easily check that the median is much less sensitive to outliers.

To take into account the possibility of outliers, in this paper, in addition to using the Least Squares method to find the parameters of the model, we also use an ℓ^1 method.

How to check whether the data fits a model. Once we have found the parameters of a model, we can apply the Pearson's chi-squared test to check whether the data fits the corresponding formula. This test checks whether the number of events $O_1, \ldots, O_n$ that occurred in n different situations is consistent with the model that predicts that, one average, E_i events will occur in situation i. To apply this test, we compute the sum $\mu = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$. This distribution of this value is close to the chi-squared distribution with $n - n_{\text{param}}$ parameters, where n_{param} is the number of parameters that was determined based on this distribution (in our case, $n_{\text{param}} = 2$). Then, we compare a p-value by comparing the value of the statistic to a chi-squared distribution, and use this p-value to decide whether the observations fit the model.

In our case, different situations i correspond to different moments of time t, the expected values come from the corresponding model $E_i = r_0(t)$, and the observed values are $O_i = r(t)$. Thus, for this situation, Pearson's chi-squared test means computing the expression

$$\chi^2 \stackrel{\text{def}}{=} \sum_t \frac{(r(t) - r_0(t))^2}{r_0(t)}.$$

In particular, for the power model, we compute the value

$$\chi_p^2 \stackrel{\text{def}}{=} \sum_t \frac{(r(t) - A \cdot t^{-\alpha})^2}{A \cdot t^{-\alpha}}.$$

For exponential model, we compute the value

$$\chi_e^2 \stackrel{\text{def}}{=} \sum_t \frac{(r(t) - A \cdot \exp(-\alpha \cdot t))^2}{A \cdot \exp(-\alpha \cdot t)}.$$

Based on this value of χ^2 and on the number $n-2$ of degrees of freedom, we compute the corresponding p-value p_p or p_e.

The values of χ^2 and p based on ℓ^1-estimates will be denoted, correspondingly, by $\chi^2_{p,1}$, $\chi^2_{e,1}$, $p_{p,1}$, and $p_{e,1}$.

If this p-value is smaller than some threshold p_0 (usually, $p_0 = 0.05$), then we can conclude that the data is inconsistent with the corresponding model, and the model is rejected. Otherwise, if the p-value is greater than or equal to the threshold p_0, we conclude that the data is consistent with the model.

5 Comparison Results

Selecting dispatches. When the sample size is small, both models fit. Empirically, we have found that the models can be separated if we have at least 50 comments. Some dispatched are shorter than others; these dispatches are marked as “trail notes”. None of these trail notes has 50 or more comments; so, we only considered “proper” dispatches, i.e., dispatches which are not trail notes.

Our interest is in analyzing knowledge propagation. New dispatches appear all the time, and new comments are added all the time. Thus, the more recent the dispatch, the more probable it is that new comments will be added and therefore, that the available comments do not yet present a final description of how the corresponding knowledge propagates. So, in our analysis, we concentrated on the earliest dispatches, for which the picture of knowledge propagation is (most probably) complete.

Specifically, to compare the two models, out of the 25 earlier "proper" dispatches, we selected all the dispatches which by February 1, 2015, had at least 50 comments. There were seven such dispatches; these dispatches are listed in the following table.

For each dispatch, there is usually a few days delay until the bulk of the comments appear. At first, there is a burst of comments, then the number of comments gradually decreases, and after a month, very few new comments appear. So, for our analysis, we limited our data to a time period that:

- starts on the first day when the largest number of comments appear, and
- ends 30 days after the posting of the original dispatch.

Comparison results. The result of analyzing the selected ten dispatches are given in the following table; here, N_c is the number of comments, and all the p-values are rounded to two digits. P-values exceeding 0.05 – that indicate that the model is consistent with the data – are underlined.

The ℓ^1 values corresponding to the exponential model are not given, but the resulting p-values are similar to the values corresponding to the Least Squares estimates.

Dispatch Title	N_c	χ^2_p	$\chi^2_{p,1}$	χ^2_e	p_p	$p_{p,1}$	p_e
Let's Walk	271	30.6	30.0	31,360	<u>0.33</u>	<u>0.37</u>	0.00
Sole Brothers	61	22.1	22.8	83	<u>0.76</u>	<u>0.74</u>	0.00
The Glorious Boneyard	59	16.3	18.6	262	<u>0.96</u>	<u>0.91</u>	0.00
The Self-Love Boat	67	63.1	60.0	124	0.00	0.00	0.00
Go Slowly–Work Slowly	91	33.0	31.5	821	<u>0.24</u>	<u>0.29</u>	0.00
The Camel and the Gyrocopter	52	28.4	24.6	72	<u>0.45</u>	<u>0.65</u>	0.00
Lines in Sand	69	21.4	18.3	89	<u>0.81</u>	<u>0.92</u>	0.00

Conclusions:

- None of the dispatches is consistent with the exponential law.
- For both the Least Squares method and a more robust ℓ^1-method to estimate the parameters, the vast majority of the dispatches is consistent with the power law.

So, this data supports the power law model in comparison with the more traditional exponential model.

Discussion. A possible reason why comments on some dispatches fit neither the power law nor the exponential law is that Salopek sometimes replies to the comments, and these replies trigger another wave of comments. As a result, some observed distributions of comments over time are bimodal or close to bimodal.

6 Conclusions and Future Work

Conclusions. To improve teaching and learning, it is important to know how knowledge propagates. Traditional models of knowledge propagation are similar to differential-equations-based models of propagation in physics. Recently, an alternative fractal-motivated power-law model of knowledge propagation was proposed, that, in several cases, provides a more adequate description of knowledge propagation. In this paper, we compare this model with the traditional model on the example of the comments to the Out of Eden Walk project, an ambitious journalistic and educational project aimed at educating the general audience about different societies around the world, their culture, their history, etc.

It turns out that for the related data, the power law is indeed a more adequate description:

- for the vast majority of dispatches, the dependence of number of comments on time is consistent with the power law, while
- the differential equations-motivated exponential law is not consistent with any of this data sets.

This shows that the fractal-motivated power law is indeed a more adequate description of knowledge propagation – to be more precise, a more adequate first approximation to describing knowledge propagation.

Possible future work. There are many possibilities to expand our analysis:

- we can further analyze our power law models,
- we can go beyond the power law models and try to get models which are even more adequate for describing knowledge propagation,

- instead of simply counting comments corresponding to different dispatches, we can also look into the substance of these comments and dispatches, and finally,
- we can use information beyond web-posted comments.

Let us describe these possibilities one by one.

Possible further analysis of the power law models corresponding to different dispatches. First, it would be interesting to *further analyze* our power law results. In our analysis, we simply checked whether the dependence is described by the power law or not. However, the power laws come with different parameters α. The larger α, the faster the number of comments decreases with time. It would be interesting to check whether the values α corresponding to different dispatches depend on the overall number of comments. Based on the few dispatches that we analyzed, it appears that when we have more initial comments, then the decrease is slower, but so far, we do not have enough data to rigorously confirm this observation.

Beyond power law models. As we have mentioned earlier, the power law explains the time distribution of comments only for a little more than a half of dispatches. It is therefore desirable to design models that would explain the time dependence for all the dispatches – or at least for a large portion of dispatches. There are two possible ways to get a better fit.

One possibility is to consider *more complex* models. A natural way to do it is to take into account that a power law is linear in log-log scale, it leads to a linear dependence of $\ln(r(t))$ on $\ln(t)$:

$$\ln(r(t)) \approx \ln(A) - \alpha \cdot \ln(t);$$

thus, a natural idea is to see if a quadratic dependence

$$\ln(r(t)) \approx \ln(A) - \alpha \cdot \ln(t) + \beta \cdot (\ln(t))^2$$

leads to a better fit.

Another possibility is to take into account some *external events* that affect the number of comments: for example, the number of comments spikes when Salopek replies to comments, when he appears on National Public Radio, or when an article of his is published in the *National Geographic* magazine.

Probably, both ideas need to be implemented to get a better fit with the observed number of comments to different dispatches.

Beyond counting the number of comments. In our analysis, we simply counted the number of comments. It is desirable to also take into account the substance of these comments – and also the substance of the dispatch.

Our main interest is in the educational applications. From the educational viewpoints, *not all comments are equal*:

- some comments simply praise the dispatch, without any explicit indication that the replier learned something new from it;
- other comments explicitly indicate that the replier learned some new information from the dispatch;
- a few comments go even further, indicating that the replier plans to inform their friends and colleagues about some material from this dispatch – in particular, some repliers who are teachers plan to use this material in their classes.

It would be interesting to analyze how the numbers of such "learning" and "teaching" comments changes with time.

Similarly, *not all dispatches are equal*. Some topics – e.g., the dispatches about the beasts of burden such as camels or mules – caused many comments, and these comments come a few days after the dispatch is posted. On the other hand, other topics cause fewer replies. It would be interesting to analyze which types of dispatches elicit more comments. This may be useful in teaching, since eliciting comments from students is a known way to improve their learning.

It is also desirable to use *information about the repliers*:

- The majority of repliers sign their comments with their *names*. It would be interesting to trace how many repliers for each dispatch are new and how many also replied to one or more of the previous dispatches; this will show how the audience changes in time.
- Some repliers indicate their *geographic location*. It would be interesting to analyze the geographic distribution of comments, and to see how the geographic distribution of comments to a given dispatch depends on the geographic area that is the subject of this dispatch.

Beyond web-based replies. Finally, we should take into consideration that, in addition to comments posted on the website, repliers also post comments on Twitter, Instagram, and Facebook. It to desirable to analyze these comments as well.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. OCI-1135525 for the CI-Team Diffusion project: The Virtual Learning Commons and HRD-1242122 for the Cyber-ShARE Center of Excellence renewal.

We are greatly thankful to Paul Salopek for his encouragement and advice.

References

[1] L. M. A. Bettencourt, A. Cintron-Arias, D. I. Kaiser, and C. Castillo-Chavez, "The power of a good idea: quantitative modeling of the spread of ideas from epidemiological models", *Physica A*, 2006, Vol. 364, pp. 513–536.

[2] B. K. Chakrabarti, A. Chakraborti, and A. Chatterjee, *Econophysics and Sociophysics: Trends and Perspectives*, Wiley-VCH, Berlin, 2006.

[3] P. A. Geroski, "Models of technology diffusion", *Research Policy*, 2000, Vol. 29, pp. 603–625.

[4] P. J. Huber and E. M. Ronchetti, *Robust Statistics*, Wiley, Hoboken, New Jersey, 2009

[5] J. Jee, L. C. Klippel, M. S. Hossain, N. Ramakrishnan, and B. Mishra, "Discovering the ebb and flow of ideas from text corpora", *IEEE Computer*, February 2012, pp. 77–81.

[6] M. Le Masurier, "What is Slow Journalism?", *Journalism Practice*, 2014.

[7] L. O. Lerma, D. Pennington, and V. Kreinovich, "On Early Stages of Idea Propagation, the Number of Adopters Grows as $n(t) \sim c \cdot t^a$: Theoretical Explanation of the Empirical Observation", *Journal of Uncertain Systems*, 2014, Vol. 8, No. 3, pp. 180–185.

[8] B. B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman Publ., San Francisco, California, 1983.

[9] P. Morone and R. Taylor, *Knowledge Diffusion and Innovation: Modeling Complex Entrepreneurial Behaviors*, Edward Elgar Publ., Cheltenham, UK, 2010.

[10] Out of Eden Walk http://www.outofedenwalk.com/

[11] Pulitzer Centre, *Out of Eden: Paul Salopeks Walk from Ethiopia to Patagonia*, http://pulitzercenter.org/projects/out-of-eden

[12] S. I. Resnick, *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer-Verlag, New York, 2007.

[13] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC Press, Boca Raton, Florida, 2011.

[14] H. S. Tavangar and B. Mladic-Morales, *The Global Education Toolkit for Elementary Learners*, Corwin, Thousand Oaks, California, 2014.