International Journal of Communication 7 (2013), 653–680 1932–8036/20130005

The Functionality of Social Tagging as a Communication System

POONG OH[1]
PETER MONGE
University of Southern California

This study examines changes over time in the functionality of social tagging systems. Social tagging systems are conceptualized as a special kind of communication system that enables users to organize and discover information resources using tags as signals. Functionality is defined as the extent to which the system minimizes both encoding and decoding efforts, following Zipf's principle of least effort. The Yule-Simon model is adopted as the growth mechanism of the system. Two hypotheses are proposed: The first predicts decline, increase, and stabilization of the tag-frequency distribution; the second predicts an oscillation in the functionality leading to a dampened stability. The empirical data support the first hypothesis and partially support the second hypothesis. The article discusses the implications for collective intelligence.

Keywords: social tagging system, power-law distribution, encoding and decoding efforts, Zipf-Mandelbrot model, Yule-Simon model, preferential attachment

The rapid development of information and communication technologies over recent decades has led to an explosive increase in electronically stored information. According to Hilbert and López (2011), the world's technological capacity for information storage roughly doubled every 40 months between 1987 and 2007, and its growth rate has been accelerated by the recent development of compression techniques. However, the proliferation of information does not immediately imply its utility. For example, a library would be little more than a huge collection of paper without a well-functioning classification system. Therefore, the increasing information in a system requires a reliable and efficient *meta*-information system, one that generates and manages the "information *about* the information resources,"

[1] This research was supported by the National Science Foundation (IIS-0838548) and by a grant from the Annenberg School for Communication and Journalism. The authors express their appreciation to Amanda Beacom, Manuel Castells, Edward Fink, Janet Fulk, David Kempe, Kristina Lerman, Drew Margolin, Seungahn Nah, and Malcom Parks for helpful comments on earlier drafts of this article.

Poong Oh: poongoh@usc.edu
Peter Monge: monge@usc.edu
Date submitted: 2012–02–18

(Michener, 2006, p. 3; emphasis added) including their content, quality, structure, and accessibility. Recently, however, scholars have begun to question whether conventional meta-information systems are adequate for organizing large amounts of information—in particular, online information resources (Golder & Huberman, 2006).

One potential solution that has emerged is social tagging systems (also called collective tagging or social bookmarking systems), which some scholars claim are functionally equivalent or even superior to conventional systems (Boulos & Wheeler, 2007). Others, however, have questioned the reliability and efficiency of social tagging systems. Begelman, Keller, and Smadja (2006) argue that differences in individuals' perceptions reduce the utility of each person's tags for others. For example, some tags, such as "to-read" and "cool," are too personal to be informative for others. Also, it has been pointed out that the decentralized structure of social tagging systems cannot regulate the increasing *noise* (i.e., unwanted disturbances, Pierce, 1980) as effectively as the conventional information management systems (Chi & Mytkowicz, 2008). On the other hand, the proponents of social computing and Web 2.0–based applications maintain that, even though highly personalized tags may not be useful for others, the larger totality of tags can produce a reliable system as a form of collective intelligence that emerges from the dynamic interactions among many individuals, through which undesirable noise can be canceled out (Boulos & Wheeler, 2007). Further, Zauder, Lazic, and Zorica (2007) argue that social tagging systems facilitate knowledge discovery, functioning as a voting or recommendation system.

With this ongoing debate in mind, this study provides a new approach to examining the functionality of social tagging systems. First, the study examines how the functionality of social tagging systems *changes over time* rather than whether they are functional at a certain point in time. Social tagging systems evolve, which most previous studies ignore. As new information resources enter these systems, new tags should be created and assigned to the new resources. The creation and assignment of new tags necessarily alter the existing tag-resource association structure and, thereby, the functionality of the entire social tagging system as well. These new tags may either enhance system functionality—for example, by canceling out the existing errors (Boulos & Wheeler, 2007)—or decrease its functionality—for instance, as new sources of noise (Chi & Mytkowicz, 2008). Therefore, it is important to assess the change in the functionality of social tagging systems to better understand their dynamism and complexity.

Second, the present study conceptualizes the process by which individuals organize and discover information resources via tags as a special form of *communication* between taggers and tag users. Specifically, resource organization and discovery are conceptually equivalent to encoding and decoding processes, respectively. Tag-resource association, however, pertains to the coding system in Shannon's (1948) general communication model. The functionality of a social tagging system is defined as the degree to which it facilitates both processes by keeping the balance between taggers' efforts to organize information resources and tag users' efforts to discover their target information based on the principle of least effort (Zipf, 1949). Previous studies have assessed functionality either from tag users' viewpoints as resource discovery tools (Razikin, Goh, Chua, & Lee, 2011) or from taggers' viewpoints as resource organization tools (Lipczak & Milios, 2011) rather than both.

Finally, the study identifies the growth mechanisms by which social tagging systems arise from the dynamic interactions among multiple users and their mutual influences. People tend to consult with others and to imitate their choices, especially when they need to select tags out of many possible alternatives (Cattuto, Loreto, & Pietronero, 2007). We formalize such a tendency as "preferential attachment" and adopt the Yule-Simon model (Simon, 1955) as the underlying mechanism by which particular structural properties of tag-resource associations emerge.

The article is organized as follows. The next section focuses on the two key characteristics of social tagging systems—flexible and decentralized structures—and conceptualizes social tagging systems as a communication system between taggers and tag users. The following section reviews the Zipf-Mandelbrot and Yule-Simon models, which provide competing explanations for the emergence of particular structural properties of word-meaning association in natural languages as applied to social tagging systems. This discussion provides the basis for the development of hypotheses on the change in the functionality of social tagging systems over time. The method and results sections describe the data set and analysis procedures and present the results of the data analysis. The article closes with a discussion of the implications of the findings for collective intelligence and for the evolution of complex and dynamic systems in general.

Conceptualization of Social Tagging Systems

Marking information with descriptive terms and classifying documents into separate groups for later retrieval has been a common way of organizing information since long before the digital revolution. Classification ideas can be dated to as early as the 3rd century B.C.E., when a huge collection of books was organized by Demetrius, a student of Aristotle, at the Royal Library of Alexandria in Egypt (MacLeod, 2004). In the 16th century, the library at Leiden University in the Netherlands used a complete alphabetical cataloging system (Berkvens-Stevelinck, 2004). Information management systems can be easily found in our everyday lives as well (Spink, 2010). For example, filing hundreds of documents and retrieving specific ones out of the collection is one of the most important skills for office workers. However, there are fundamental differences between conventional information organization/retrieval systems and social tagging systems.

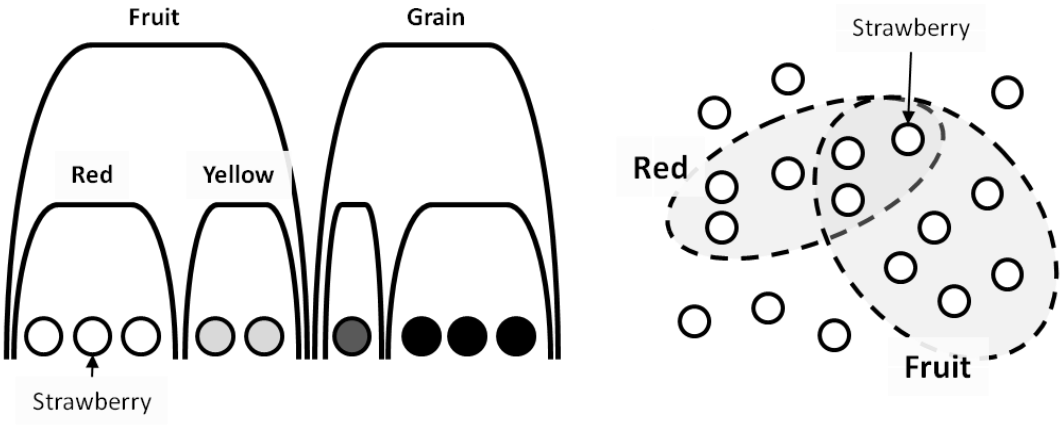
Tags as Meta-Information

According to information foraging theory (Pirolli, 2007), information resources are not evenly distributed in a given environment; rather, they are clustered in "information patches." Each patch has its own cues, "information scents," that enable information foragers to infer the content of the patch without looking directly into it and thereby facilitate information discovery. Resource organizing is a tagging process in which individual users divide a set of information resources into several subsets by placing one or more tags on each subset. Resource discovery occurs when people select a subset of information by referring to the tags on it and searching within the subset. In this sense, tags function as meta-information, the information about information resources (Sinclair & Cardew-Hall, 2007).

Tagging Systems: Structural Flexibility

Jacob (2004) defines tagging systems as a form of categorization system distinct from the classification form. Classification involves the orderly and systematic assignment of each object to one and only one class within a system of mutually exclusive and exhaustive classes (Figure 1a). On the other hand, categorization divides the set of objects into categories whose members share some perceptible similarity within a given context (Figure 1b). An object with complex characteristics would simultaneously belong to multiple categories in a categorization system that allows overlaps among categories. For example, *strawberry* can be categorized as both *red* in terms of color and *fruit* as a kind of crop. At the system level, the *red* and the *fruit* categories share a common area that includes *strawberry*. This inclusive nature of categorization systems allows a great degree of flexibility in organizing information resources compared to classification systems.

However, the structural flexibility of tagging systems occurs as a result of a trade-off with precision. Resource discovery in tagging systems necessarily involves considerable uncertainty, even though it would be less than the level of uncertainty if no tags were used. Typically, a single tag is attached to more than one object. In the above example, the tag *red* can be attached not only to *strawberry* but to *rose*, *fire truck*, and any other red objects. In that case, the search boundary has been narrowed down from all objects to the red objects, but further search is required within the boundary. In fact, Razikin et al. (2011) reported that each of the top 100 tags on a popular social tagging website, *del.icio.us*, are attached to 1,300 documents on average, meaning that a tag user has to continue to search among the 1,300 documents even after narrowing down the search boundary provided by that tag. In contrast, in classification systems such as libraries, where every information resource has a unique call number, patrons can precisely locate target information without uncertainty.



(a) Classification (b) Categorization

Figure 1. Comparison between classification and categorization.

Social Tagging Systems: Structural Decentralization

Social tagging systems were originally designed to enable individual users to share their personalized meta-information systems with others (Golder & Huberman, 2006). The underlying mechanism is simple. Individual users of a Web service add freely determined descriptive keywords or tags to commonly available information resources. In most cases, users place tags for personal purposes according to their own judgments and preferences, such as a bookmark for future revisiting. In online environments, however, personalized tagging systems can be easily made available to others, functioning as "navigational advice" for others, regardless of the taggers' intentions (Kammerer, Nairn, Pirolli, & Chi, 2009). The aggregate of personalized tagging systems constitutes a social tagging system at a collective level.

A *social* tagging system, as its name implies, is generated from the dynamic interactions among multiple users. The decentralized structure of social tagging systems ensures such dynamics (Golder & Huberman, 2006). In traditional meta-information systems, classifying and indexing information resources is exclusively performed by either authors who create the resources or librarians trained to strictly follow conventional classification rules (e.g., the Dewey Decimal Classification system). In addition, the descriptive terms are already determined by centralized authorities (i.e., controlled vocabulary by experts, Macgregor & McCulloch, 2006). In contrast, social tagging systems allow *anyone*—usually consumers rather than authors or librarians—to freely attach tags to the information resources. For this reason, the tag sets in social tagging systems are often called "folksonomies" (Hendry, Jenkins, & McCarthy, 2006). In short, social tagging systems are not governed by predetermined rules or standards; they are generated from and continually changed by decentralized local interactions among the users.

However, the decentralized structure—more specifically, the lack of predetermined rules and standards—does not necessarily imply that the systems will remain unstructured over time. Rather, a highly centralized structure will emerge from dynamic interactions among users, which are, paradoxically, facilitated by the decentralized structure of the systems. To demonstrate, the lack of centralized control over tag selection inevitably entails considerable uncertainty—that is, which of several possible tags to choose for either resource organization or resource discovery. In that case, individual users may consult what other users usually do and imitate the others' choices, considering others' choices as a substitute for predetermined rules and standards. In fact, most social tagging sites make it easy to imitate others by providing the information about tag popularity (e.g., a tag cloud, which is a visual presentation of tag frequency). In this sense, individuals' tag selection involves nontrivial interdependence, and, therefore, the entire social tagging system is not reducible to a mere aggregation of independently developed personalized systems.

This phenomenon is called *emergence*, where the collective behavior at the global level arises from the interactions among the local parts of the system, those which are following simple rules (Wolf & Holvoet, 2005). Well-known examples include ant pheromone trails, aggregations of cockroaches, termite mounds, and collective behavior of human crowds, such as the applause of opera audiences and the formation of traffic jams (Sumpter, 2010). Similarly, the decentralized structure of social tagging systems facilitates the interactions among individual users, which, in turn, gives rise to global structures, which are

irreducible to the aggregation of individual tagging systems. The possibility that individual users in social tagging systems follow simple rules (e.g., imitating others) and its impact on the emergence of global patterns will be discussed in a later section.

Social Tagging Systems as a Communication System

Tags are meta-information rather than the information resources that users intend to organize or discover. Instead, tags help users to organize and discover information resources by signaling the contents and locations of the information resources. In this sense, tag-resource association is conceptually equivalent to the signal-message association in Shannon's (1948) general communication model.

Shannon's model includes three components—source, receiver, and communication channel—and the two processes—encoding and decoding.[2] Sources who wish to send messages to receivers have to encode their messages into sets of signals following coding rules. Receivers, on the other hand, receive the signals instead of the messages. Therefore, receivers need to infer the messages that the sources originally attempted to deliver based on the received signals (i.e., decoding). The communication process described by Shannon can be applied to a social tagging system (Figure 2). In a social tagging system, users organize information resources by placing tags and, thereby, make it easy for other users to discover the resources. In this case, the information resources can be seen to be communicated from those who organize the resources (i.e., taggers) to those who discover the resources (i.e., tag users). However, it is important to note that the information resources are not actually transported from taggers to tag users. The information resources are already available to all the users. What taggers select and send is a set of tags to indicate the contents and/or locations of information resources. On the other hand, what tag users receive is the set of tags, from which they need to infer the contents and/or locations of the information resources. In Shannon's terms, taggers *encode* the contents and/or locations of information resources into tags; tag users *decode* the contents and/or locations of information resources from the tag. The communication between a tagger and a tag user will be successful if the resource that the tagger indicates by a tag is *precisely* one that the tag user infers from the tag. The communication will have a high probability of being successful if the resource that the tagger is *likely* to indicate by a tag is

---

[2] In Shannon's model, *communication channel* is defined as "the medium used to transmit the signal from transmitter to receiver" (1948, p. 398), which determines the number of signals that can be transmitted per unit time (i.e., channel capacity). However, when social tagging systems are conceptualized as a communication system among users, it is an asynchronous communication that does not involve a time constraint. Therefore, the channel component was not considered in this article.

one that the tag user is *likely* to infer from the tag. More generally, the probability structure of the tag-resource association in a social tagging system determines the degree to which users successfully communicate information resources in the system.

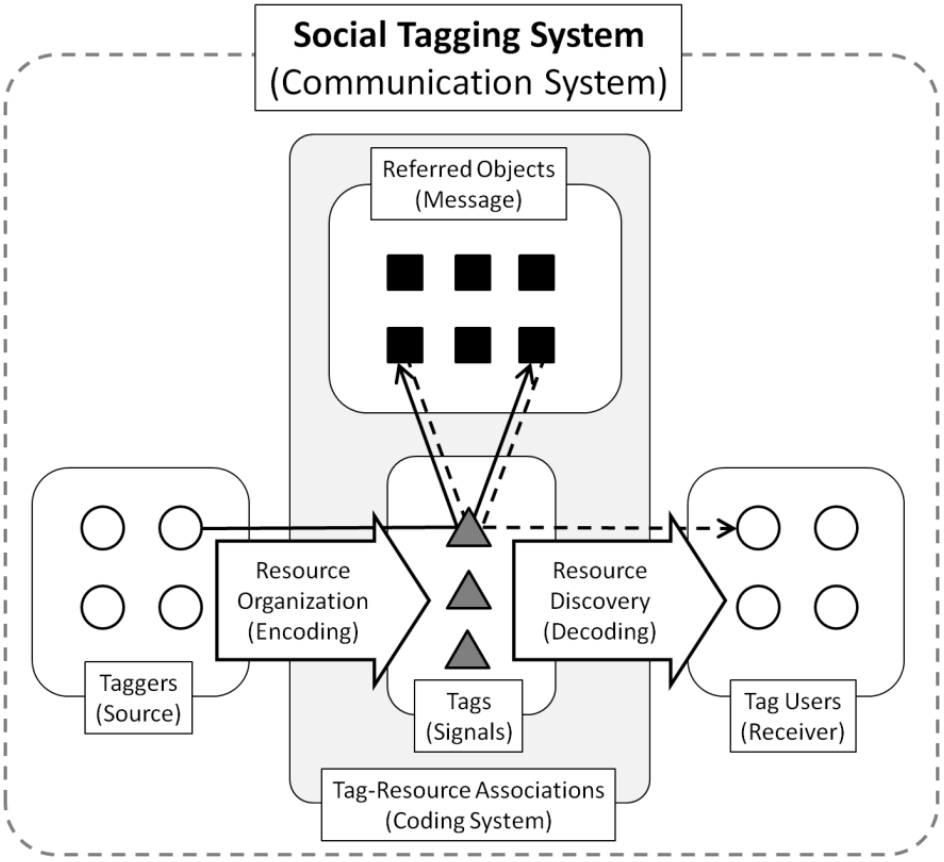


Figure 2. Schematic diagram of a social tagging system as a communication system.

Note. *The terms in parentheses are the corresponding components of Shannon's (1948) general communication model.*

Functionality of Social Tagging Systems

The tag-resource association in a social tagging system is conceptually equivalent to the signal-message association, which Shannon called "coding systems" (1948, p. 389). His original goal was to develop optimal coding systems under limited channel capacities, and he formulated a set of mathematical theorems for this purpose. One of them states that the variety of signals should be equal to or greater than that of messages for unambiguous communication (Theorem 9, p. 421; see also the law of requisite variety, Ashby, 1957). Applying Shannon's theorem to social tagging systems, the variety of tag set should be equal to or greater than that of information resource set for unambiguous communication. Put differently, a distinctive tag should be attached to every resource in the same way that a unique call number is attached to every book in a library system. However, social tagging systems entail a certain amount of uncertainty as a result of the trade-off between precision and flexibility.

Zipf-Mandelbrot Model: The Principle of Least Effort

The trade-offs between the precision and flexibility of communication systems are easily found in natural human languages, which were first theorized by Zipf (1949). He considered the optimal coding system as one that minimizes both encoding and decoding costs. The decoding cost is minimized if every signal or word is assigned to one and only one object in a way similar to Shannon's theorem. Therefore, from the receivers' viewpoint, coding systems that are maximally diversified (i.e., precise) are desired, which is called the "Force of Diversification" (p. 21). On the other hand, the encoding cost is minimized if it is possible to communicate every meaning with a single word. Thus, from the sources' viewpoint, a coding system that is maximally unified (i.e., flexible) is preferred, which is called the "Force of Unification" (p. 21). Putting these together, Zipf hypothesized that natural languages have evolved into a state of "vocabulary balance," where the two opposing forces are perfectly balanced.

To support his hypothesis, Zipf (1949) examined a massive amount of empirical data from varied sources such as magazines, newspapers, novels, speech scripts, and other texts, written in diverse languages, such as English, French, German, and so on. He found an approximate hyperbolic relationship between the frequency of occurrence of each word (f) and the rank (r) of the word in the corpora (Figure 3). Rank is the order of the words according to their frequencies, where rank one is the most frequently used word. Thus,

$$r \times f = C, \tag{1}$$

where C is a constant determined by the frequency of the most frequent word ($r = 1$). Equation (1) can be solved for f by dividing both sides by r:

$$f = C \cdot r^{-1}. \tag{2}$$

Further, it can be rewritten by taking the logarithms of both sides:

$$\log f = -\log r + \log C. \tag{3}$$

By Equations (2) and (3), the relationship between *f* and *r* is presented as a power-law distribution with an exponent = -1 or as a straight line with a slope = -1 on a log-log scale. Zipf (1949) viewed such a word-frequency distribution as the state at which the vocabulary balance is achieved, calling it the "ideal" distribution (p. 26). Later, Zipf's hypothesis was further substantiated by Mandelbrot's (1953) mathematical proof that shows the encoding and decoding efforts are optimally balanced when the slope of word-frequency distribution equals -1.

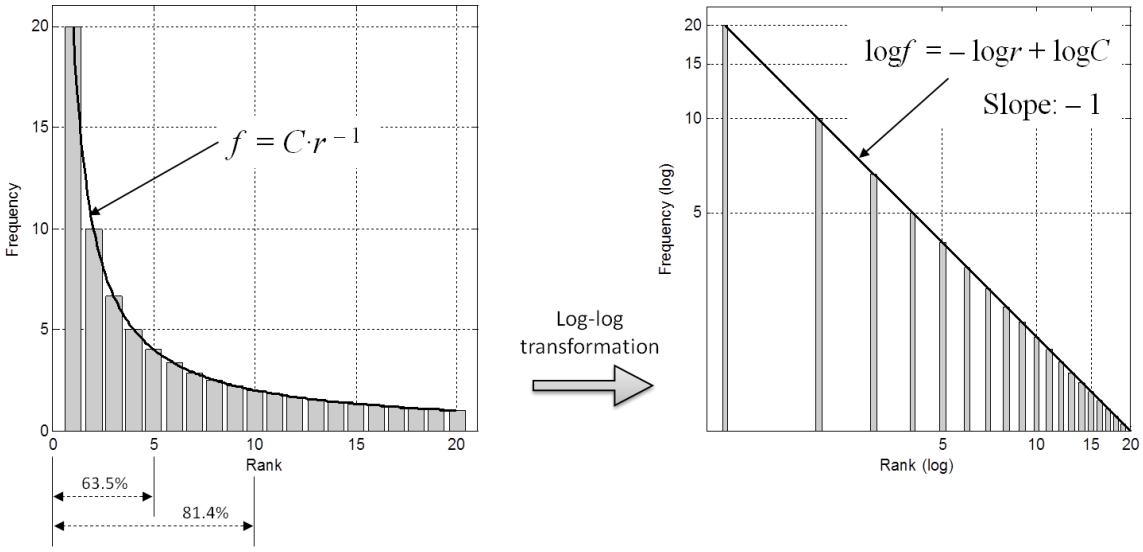


Figure 3. An example of Zipf's (1949) "ideal" distribution.

However, the Zipf-Mandelbrot model has several limitations when applied to social tagging systems. First, the principle of least effort views human communication processes as sequences of perfectly rational decisions. It assumes that every individual knows in advance which words to use exactly how many times to minimize his or her efforts. Second, the Zipf-Mandelbrot model treats a given text as a static entity, excluding any possibility that its volume grows over time. However, it is more reasonable to view a social tagging system as a growing text rather than one with a fixed volume. Third, the Zipf-Mandelbrot model focuses only on the text corpora created by single authors. However, a recent study (Zanette & Montemurro, 2005) found serious deviations from the ideal distribution when several novels, each of which was written by a different author, were aggregated into a single corpus, although the word-frequency distribution of each novel follows the ideal distribution. When a social tagging system is considered as an aggregate of texts written by different authors, it is expected that the aggregated tag-frequency distribution will not follow the Zipf-Mandelbrot model.

The Zipf-Mandelbrot model is essentially a normative model, where every individual is assumed to be perfectly rational, and, thus, the model does not fully reflect reality. However, the Zipf-Mandelbrot model, or any normative model in general, is a useful reference point for the analysis of how irrational the final collective outcomes are. For instance, if the observed slope of the tag-frequency distribution of a social tagging system deviates from the ideal slope, this indicates that the system is not optimally functional.

Yule-Simon Model: Preferential Attachment

Simon (1955) modified Yule's (1925) model of biological speciation to suggest an ingenious explanation for the occurrence of power-law word-frequency distributions, rejecting Mandelbrot's (1953) optimization approach.[3] The Yule-Simon model was later adopted to explain the mechanism by which power-law degree distributions arise in large-scale networks, which is now known as "preferential attachment" (Barabási & Albert, 1999). Unlike the Zipf-Mandelbrot model, the Yule-Simon model assumes that the volume of a text increases over time: (1) a corpus grows by one word at a time, as a word is added to the existing text; (2) the added word is either a new one that has never been used before with a probability of a or an existing one with the complementary probability of $(1 - a)$; and (3) the probability that an existing word will be used again at a given point in time is proportional to its frequency of occurrences in the existing text. The Yule-Simon model shows that the limiting slope of a word-frequency distribution is solely determined by a:

$$\text{Limiting slope} = -1/(1 - a).$$

As a increases, the limiting slope becomes steeper, which means that the word-frequency distribution further deviates from the ideal distribution (Figure 4b) and that the costs for encoding and decoding become imbalanced (Figure 4d). When $a = 0$, the slope monotonically decreases, and the functionality monotonically increases (Figures 4a and 4c). In contrast, when $a > 0$, both the slope and functionality show more complicated behavior (Figures 4b and 4d).

[3] The interested reader who wishes to pursue this matter is referred to the debate between Simon and Mandelbrot in *Journal of Information and Control,* volumes 4 (1960) and 5 (1961).

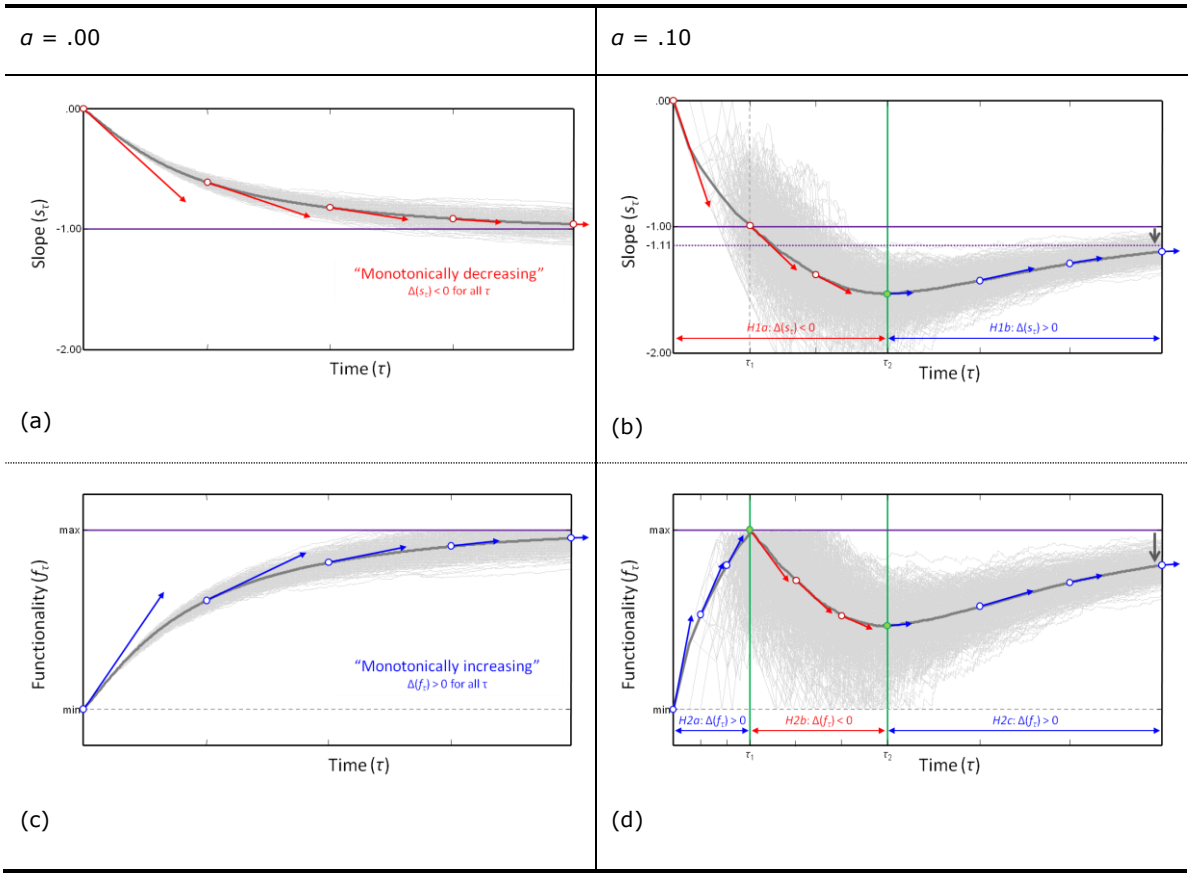


Figure 4. The outcomes of the Monte Carlo simulations of the Yule-Simon process following Simon and van Wormer (2007): (a) and (b) are the changes in the slope of tag-frequency distribution when a = .00 and a = .10 (H1); (c) and (d) are the changes in the closeness to the ideal distribution when a = .00 and a = .10 (H2).

The Yule-Simon model shows that the optimal vocabulary balance can be achieved simply by *reproducing* the existing vocabulary structure. To reach the ideal distribution, individuals simply use frequent words more often than infrequent ones, a case of preferential attachment in which "the rich get richer." In addition, the Yule-Simon model attributes the deviation from the ideal distribution to the introduction of new words, which deepens the established power-law structure (i.e., the slope becomes less than -1). Because, by definition, new words have never been used before, it is unlikely that they will be used again due to the tendency of preferential attachment (i.e., "the poor get poorer").

The mechanism of text growth makes the Yule-Simon model useful and applicable in exploring the structural changes of social tagging systems. First, when social tagging systems are treated as corpora that consist of tags, it is reasonable to expect that their volume will increase over time (Chi & Mytkowicz,

2008). The other assumption of the Yule-Simon model—preferential attachment in word selection—also reflects the tag-selection behavior of users in social tagging systems. Cattuto et al. (2007), for example, reported that when individual users select tags, they tend to choose those that have been frequently used by others.

The tag-selection bias reported in previous studies can be explained by the literature on social influence. When people need to make judgments, they tend to imitate what the majority have done, considering the majority behavior as social proof (Cialdini & Trost, 1998). Such a tendency becomes more evident when the given situation is novel, ambiguous, or uncertain (Tanford & Penrod, 1984). Recently, Gigerenzer and Gaissmaier (2011) have conceptualized the notion of ecological rationality as the idea that others' behavior has informational values for judgment making, and, thereby, the imitation of others' behavior enhances the efficiency of cognitive processes. Further, Boyd and Richerson (2005) contended that the tendency to imitate the majority in peer groups is "programmed" in human instincts as an efficient survival strategy through the evolutionary process of human species. Interestingly, they point out that these imitation heuristics are the most effective when the environment is relatively stable, which is consistent with the Yule-Simon model. If no new words are introduced into the system (i.e., $a = 0$: a stable environment), the simple preferential attachment mechanism is sufficient to achieve the optimal vocabulary balance. When a number of new words are continually introduced into the system (i.e., $a > 0$), the system becomes more unstable, which causes the deviation from the ideal distribution.

Hypotheses

The preceding analysis provides the basis for the formulation of two hypotheses regarding the change in functionality of social tagging systems over time (Figure 4). At the beginning, it is presumed that only a few tags are used with about equal frequency, and, therefore, the tag frequency follows a uniform distribution and its slope is close to 0. As the social tagging system grows, tags will be used at different rates. Thus, the slope of the tag-frequency distribution will begin to decrease. Further, the preferential attachment mechanism makes it likely that frequent tags are used more often and the infrequent tags are used less often, making the slope even steeper. At the same time, the introduction of new tags leads to the further decrease in the slope of the tag-frequency distribution, even after the slope passes by -1. However, the decrease rate of the slope will slow down, and the slope will eventually converge to $-1/(1 - a)$, where a is the introduction rate of the new tags. Hypothesis 1 articulates this prediction:

Hypothesis 1: *The slope of the tag-frequency distribution of social tagging systems will initially decrease substantially but increase again after passing its minimum and eventually stabilize over time (Figure 4b).*

This hypothesis can be decomposed into three subhypotheses and formulated in mathematical terms. First, the slope of the tag-frequency distribution will decrease until the trajectory hits its bottom at a point in time, τ_2. Therefore,

Hypothesis 1a: *The first-order difference of the slope function with respect to time will be negative until τ_2, that is, $s_{\tau+1} - s_\tau = \Delta(s_\tau) < 0$, where $\tau \in [\tau_0, \tau_2]$.*

Second, the slope will increase again after τ_2. Therefore,

Hypothesis 1b: *The first-order difference of the slope function with respect to time will be positive after τ_2, that is, $s_{\tau+1} - s_\tau = \Delta(s_\tau) > 0$, where $\tau \in [\tau_2, \infty)$.*

Third, although the slope of the tag-frequency distribution will increase after τ_2, the increase rate will slow down, implying that the trajectory will stabilize over time. In other words, the magnitude of the first-order difference of the slope function will diminish over time after τ_2, that is, $\lim_{\tau\to\infty}|\Delta(s_\tau)| = 0$. Therefore,

Hypothesis 1c: *The magnitude of the first-order difference of the slope function, $|\Delta(s_\tau)|$, will be negatively correlated to time after τ_2, where $\tau \in [\tau_2, \infty)$.*

The closeness of the observed slope to the ideal slope is considered as the proxy for the functionality of social tagging systems. At the beginning, the functionality will be minimal because there will be maximal discrepancy. As the slope of the tag-frequency distribution decreases toward -1, the functionality will accordingly increase. However, after the slope becomes -1 at a certain point in time, τ_1, the functionality will keep decreasing but increase again after τ_2 and converge to $-1/(1 - a)$. Thus, the second hypothesis is established as:

Hypothesis 2: *Social tagging systems will increase in functionality over time up to a maximum defined by the slope of the "ideal" distribution followed by an immediate decrease and convergence to the point lower than the maximum functionality (Figure 4d).*

Hypothesis 2 can be decomposed into four subhypotheses in a similar way to Hypothesis 1. First, the functionality will increase until the slope passes by -1 at τ_1. Thus,

Hypothesis 2a: *The first-order difference of the functionality function with respect to time will be positive before τ_1, that is, $f_{\tau+1} - f_\tau = \Delta(f_\tau) > 0$, where $\tau \in [\tau_0, \tau_1]$.*

Second, the functionality will decrease between τ_1 and τ_2. Thus,

Hypothesis 2b: *The first-order difference of the functionality function with respect to time will be negative between τ_1 and τ_2, that is, $f_{\tau+1} - f_\tau = \Delta(f_\tau) < 0$, where $\tau \in [\tau_1, \tau_2]$.*

Third, the functionality will increase again after τ_2. Thus,

Hypothesis 2c: *The first-order difference of the functionality function with respect to time will be positive after τ_2, that is, $f_{\tau+1} - f_\tau = \Delta(f_\tau) > 0$, where $\tau \in [\tau_2, \infty)$.*

However, the functionality will eventually stabilize, converging to a point lower than the maximum functionality. In other words, the magnitude of the first-order difference of the functionality function will diminish over time after τ_2, that is, $\lim_{\tau\to\infty}|\Delta(f_\tau)| = 0$. Therefore,

Hypothesis 2d: *The magnitude of the first-order difference of the functionality function, $|\Delta(f_\tau)|$, will be negatively correlated to time after τ_2, where $\tau \in [\tau_2, \infty)$.*

Method

Data

The hypotheses were tested by analyzing the event-log file of an online academic community, *nanoHUB* (http://www.nanohub.org). This community was established in 2001 and now has more than 122,000 users (about 10,000 registered users). The users share academic documents, simulation tools, and other resources about nanotechnologies. The data consisted of tags applied to these resources by the community members. Data were collected over 34 months from September 18, 2006, to June 15, 2009. Since the social tagging service was launched on September 18, 2006, the number of tagged resources has increased from 0 to 1,937, and the number of tags has also increased over time and reached 896 (Figure 6a). When a tag is placed on a resource, it is defined as a tagging event. A total of 13,972 tagging events were observed over the 34 months.

The current data set is relatively small compared to those analyzed in previous studies. Nonetheless, the current data set offers two important advantages. First, despite the larger data set size, previous studies analyzed only subsets of the entire tag-resource structures. Moreover, those subsets were not representative of the entire structures. See, for example, the top 100 most frequent tags in Razikin et al. (2011) and the top 750 tags in Halpin, Robu, and Shepherd (2007). Second, because most previous studies analyzed cross-sectional data, they failed to capture the dynamic processes by which social tagging systems evolve over time. In contrast, the current data set allowed complete identification of every change in the social tagging system over the whole 34-month period.

At every tagging event τ, an $n \times m$ binary matrix $\{a_{ij}\}_\tau$ was constructed, where each row and each column corresponded to a tag and an information resource, respectively (Figure 5). If the ith tag was attached to the jth resource at time τ, then the entry $a_{ij\tau}$ was denoted by 1; otherwise, by 0. The number of rows, n, equals the number of distinctive tags, whereas that of columns, m, equals the number of tagged resources. As the social tagging system grew, the matrix was also extended by adding another row when a new tag was introduced or by adding another column when a resource was newly tagged. During the period of the study, 896 new tags occurred, and old tags were used 13,076 times. Thus, the complete sequence of matrices representing the entire data set contained 13,972 matrices and approximated the introduction rate of new tags, $\alpha = .064 = 896/13972$.

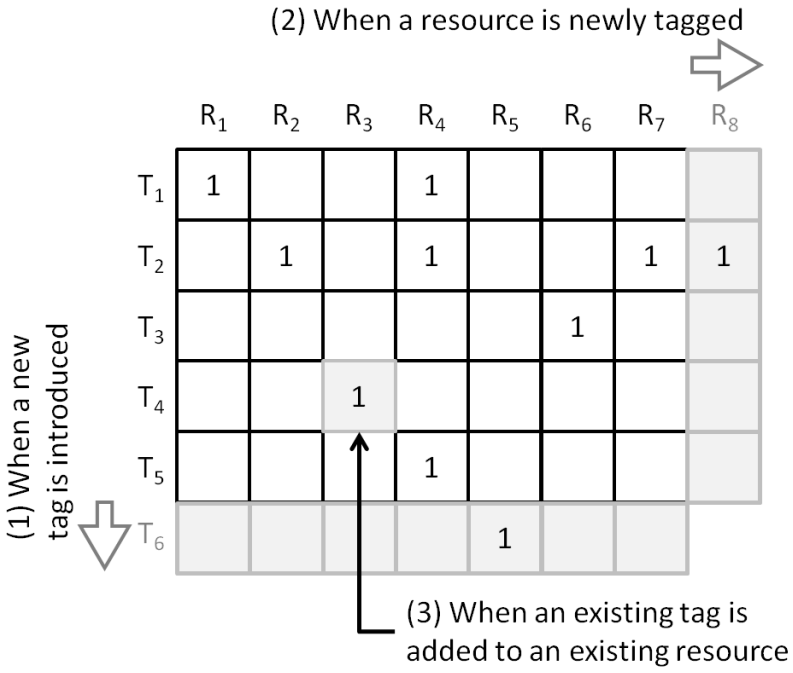


Figure 5. Construction and expansion of the tag-resource association matrix.

Measures

Slope of the tag-frequency distribution. From the row sum of a matrix, $\sum_j a_{ij}$, the frequency of occurrence of each word (*f*) and its rank (*r*) were obtained and transformed into natural logarithms. Then the slope of the linear regression of logarithms of *f* on *r* was obtained by the ordinary least square (OLS) method[4] for each of the 13,972 matrices.

[4] Clauset, Shalizi, and Newman (2009) warned that the OLS method could produce substantially inaccurate estimates of parameters for power-law distributions, suggesting the maximum likelihood (ML) method as a better alternative, in part because the residual is not normally distributed, which violates the assumption of the OLS method. However, when the number of observations is relatively small (i.e., less than 1,000), the ML method does not necessarily outperform the OLS method. Also, the ML estimates are less robust than those of the OLS method. Because the data used in this study contained a relatively small

Functionality of a social tagging system. The closeness of the slope of observed tag-frequency distributions to the ideal slope can be considered as a proxy for the functionality of social tagging systems. However, because the closeness is not a standardized metric, it does not tell us precisely how much less functional observed social tagging systems are than the ideal one. Thus, it was appropriate to use more direct measures of the functionality.

The functionality of social tagging systems was defined as the degree to which they facilitate the principle of least effort in both the encoding and decoding processes. Cancho and Solé's (2003) entropy measures were well suited to this application with some modification. *Encoding effort* was defined in terms of the uniformity of tags, here measured by means of tag entropy:

$$H_E(T) = -\sum_i p(t_i)\log_2 p(t_i), \quad (4)$$

where $p(t_i) = \sum_i a_{ij}/\sum_j\sum_i a_{ij}$. If a single tag were used for all the information resources (i.e., the completely unified tag set), then the encoding effort would be minimal, $H_E(T) = \log_2 1 = 0$. On the other hand, when every tag is used at the same frequency, the encoding effort would be maximal, and $H_E(T) = \log_2 n$.

Decoding effort was defined in terms of the diversity of tags, specifically, the entropy of resources conditional on the tag set, *T*. First, when a correct tag is chosen, the conditional entropy of resources is

$$H_D(R|t_i) = -\sum_j p(r_j|t_i)\log_2 p(r_j|t_i), \quad (5)$$

where $p(r_j|t_i)$ is the conditional probability of discovering r_j when t_i is used. The conditional probability is defined as:

$$p(r_j|t_i) = a_{ij}/(\sum_i a_{ij}\cdot\sum_j a_{ij}), \quad (6)$$

where $\sum_i a_{ij}$ is the number of tags that are commonly attached to r_j (i.e., synonyms). Next, the entropy of resources conditional on the tag set, *T*, is

$$H_D(R|T) = -\sum_j\{p(r_j)\cdot H_D(R|t_i)\}. \quad (7)$$

By Equations (6) and (7), when every resource has a unique tag (i.e., $\sum_i a_{ij} = 1$ for any *j*, and $\sum_j a_{ij} = 1$ for any *i*; maximally diversified), the decoding effort is minimal, $H_D(R|T) = \log_2 1 = 0$. On the other hand, when a single tag is used for all the information resources, the decoding effort is maximal, $H_D(R|T) = \log_2 m$.

Finally, *functionality* was measured as the proportion of the encoding entropy relative to the total entropy:

---

number of observations (i.e., 896 tags at most), and we attempted to trace the gradual change in the slope, we decided to use the OLS method rather than the ML method.

$$\text{Functionality} = H_D(R|T)/\{H_D(R|T)+H_E(T)\}. \quad (8)$$

The functionality measure ranges from 0 to 1. When the encoding effort, $H_E(T)$, and the decoding effort, $H_D(R|T)$, are equal to each other, the functionality is .5, meaning that the optimal vocabulary balance is achieved.

First-order differences of the slope and the functionality functions. For each tagging event, both slope and functionality were computed according to the formulae in the preceding subsections, and the ordered sequences of the slopes $\{s_T\}$ and the functionalities $\{f_T\}$ were constructed. The first-order differences of slope and functionality were defined as $\Delta(s_T) = s_{T+1} - s_T$ and $\Delta(f_T) = f_{T+1} - f_T$, respectively. For *H1c* and *H2d*, the magnitudes of the first-order difference were measured as their absolute values (i.e., $|\Delta(s_T)|$ and $|\Delta(f_T)|$).

Hypotheses were tested with single-sample *t* tests and correlation analyses. Significance levels were set at .05 for one-tailed tests. All analyses were conducted by using *MATLAB R2009b*. The program code is available upon request.

Results

The Change in the Slope of the Tag-Frequency Distribution

Hypothesis 1 predicted that the slope of the tag-frequency distribution of social tagging systems would initially decrease but increase again and eventually stabilize at a point below -1. Figure 6b presents the changes in the slope of the observed tag-frequency distribution between September 18, 2006, and June 15, 2009. The observed slope decreased from the beginning, passed by -1 at the 842nd tagging event, and continued to decrease until the 1,550th tagging event, reaching its minimum. As time elapsed, the slope gradually increased over time and stabilized around -1.37.

H1a predicted that the first-order difference of the slope function would be negative until it reached an inflection point, the 1,550th tagging event. The average $\Delta(s_T)$ until the inflection point at the 1,550th tagging event was significantly less than 0 [M = -0.00112; sd = 0.0104; $t(1,548)$ = -4.26; $p < .001$], supporting *H1a*. Although the average change in the slope was very small, its effect was accumulated over 1,550 tagging events and had a meaningful effect on the entire system.

Second, *H1b* predicted that the first-order difference of the slope function would be positive after the inflection point. The average $\Delta(s_T)$ after the 1,550th tagging event was significantly greater than 0 [M = 0.000289; sd = 0.000238; $t(12,421)$ = 13.34; $p < .001$], supporting *H1b*. Finally, *H1c* predicted that the slope function would stabilize after the inflection point in the tagging events, and thus, that the magnitude of the first-order difference would diminish over time. The correlation between the magnitude of first-order difference and time was r = -.43 (df = 12,420, $p < .001$), supporting *H1c* (Figure 6c).

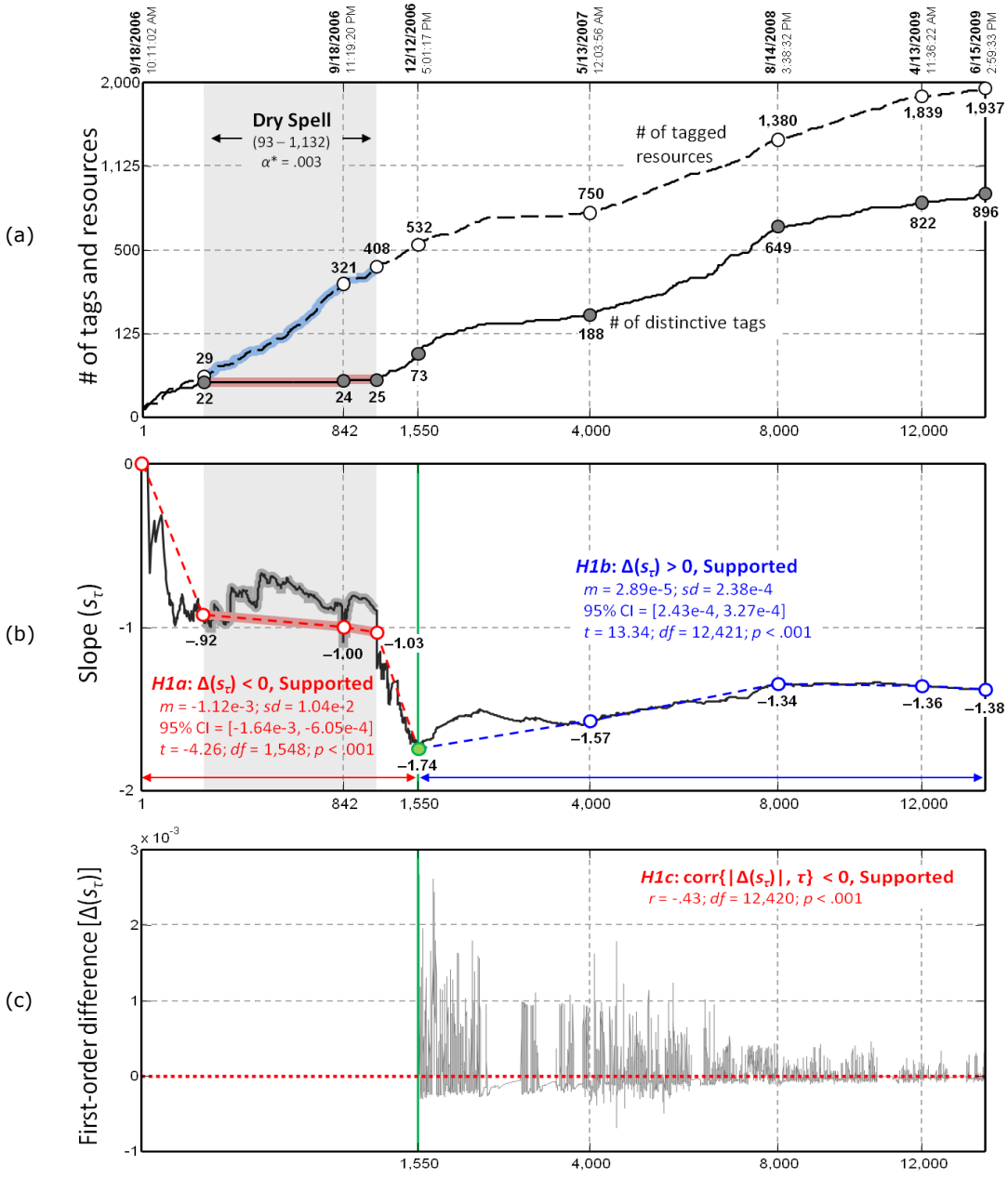


Figure 6. Testing Hypothesis 1: (a) The increase in the numbers of distinctive tags and tagged resources; (b) the change in the slope of the tag-frequency distribution; (c) the first-order difference of the slope function after the 1,550th tagging event.

The Change in the Functionality of the Social Tagging System

Hypothesis 2 predicted that social tagging systems would increase in functionality up to a maximum, followed by an immediate decrease and convergence to less than the maximum functionality. First, *H2a* predicted that the first-order difference of the functionality function would be positive before the time when slope reaches its "ideal," which in this case was the 842nd event. The average $\Delta(f_T)$ until the 842nd event was significantly greater than 0 [M = 0.000247; sd = 0.00249; $t(841) = 2.85$; $p < .001$], supporting *H2a*. Second, *H2b* predicted that the first-order difference would be negative between the 842nd and the 1,550th tagging events. The average $\Delta(f_T)$ between the 842nd and the 1,550th tagging events was significantly less than 0 [M = -0.0000606; sd = 0.000180; $t(706) = -8.96$; $p < .001$], supporting *H2b*. Third, *H2c* predicted that the first-order difference would be positive after the 1,550th tagging event. However, the average $\Delta(f_T)$ after the 1,550th tagging event was less than 0 [M = -0.0000136; sd = 0.0000362; $t(12,421) = -41.89$; $p < .001$] showing the opposite direction to the prediction. The functionality of the social tagging system decreased even further after the slope of tag-frequency distribution reached its bottom (Figure 7a). Finally, *H2d* predicted the functionality would stabilize over time, and thus, the magnitude of the first-order difference would diminish over time. As shown in Figure 7b, the magnitude decreased over time. The correlation between the magnitude of first-order difference and time was r = -.53 (df = 12,420, $p < .001$), supporting *H2d*.

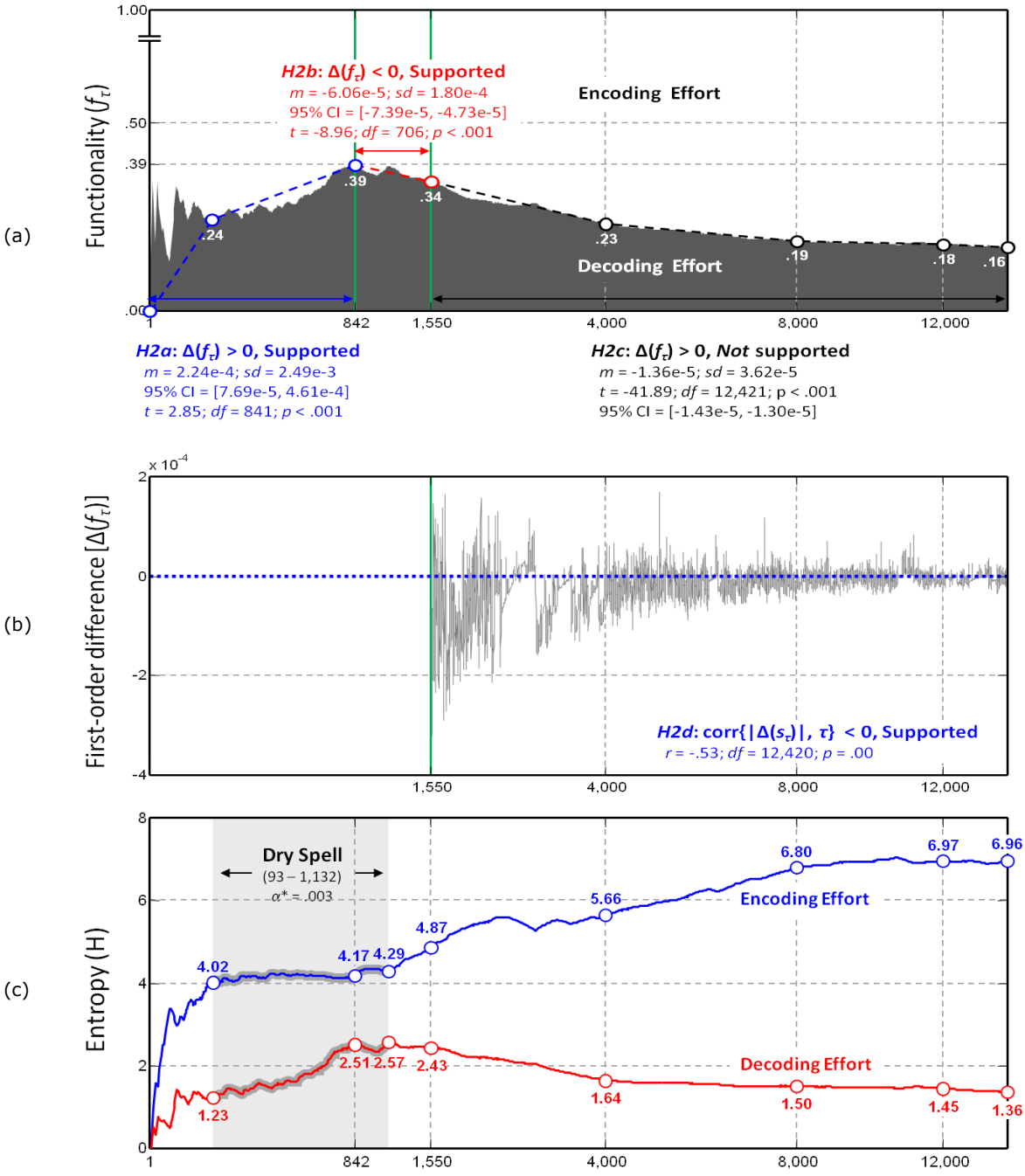


Figure 7. Testing Hypothesis 2: (a) the change in functionality; (b) the first-order difference of the functionality function after the 1,550th tagging event; (c) the changes in encoding and decoding effts over time.

Post Hoc Analysis

Impacts of the "dry spell." Over the 34-month period, the social tagging system of *nanoHUB.org* showed gradual increases in both the numbers of distinctive tags and tagged resources (Figure 6a). However, the introduction rate of new tags a was not constant over time. Specifically, between the 93rd and the 1,132nd tagging events, only 3 new tags were introduced, while old tags were reused 1,037 times ("Dry Spell" in Figure 6a). During this period, a was approximately .003 = 3/(3+1037), which is effectively equivalent to the case of a = 0 (Figure 4a). During the dry spell, the observed slope hardly changed, remaining close to the ideal slope (Figure 6b). From this observation, it can be conjectured that controlling the number of tags would be desirable to maintain the maximal functionality of the system.

This conjecture became more evident when the changes in encoding and decoding efforts were examined individually. As shown in Figure 7c, the encoding effort increased over time as the number of tags increased. This occurred because the more tags were used and, thereby, the more diverse the tag set became, the more effort individual users needed to spend in finding proper tags. On the other hand, the decoding effort increased at the early stage but decreased as the number of tags increased. This suggests that the more diverse the tag set became, the less effort individual users would spend on finding target information resources under the condition that they choose correct tags. The controlled number of tags during the dry spell had significant impacts on the changes in encoding and decoding efforts, revealing two interesting facts. During this period, the encoding effort did not significantly increase (4.02 to 4.29), even though the number of information resources organized by the tagging system substantially increased (29 to 408: about 14 times). On the other hand, the decoding efforts significantly increased (1.23 to 2.57) as the number of resources in the tagging system increased. Therefore, the tag set appeared to temporarily function as "controlled vocabulary," balancing the encoding and decoding efforts.

External constraints by the tripartite network structure. Although the Yule-Simon model predicts the limiting slope of the tag-frequency distribution will eventually converge to the point of -1.07 = -1/(1-.064) after hitting its bottom, the observed slope was trapped around -1.37 and did not increase any further after the 8,000th tagging event. This result failed to support *H2c* that the functionality would increase again after the slope of the tag-frequency distribution reached its minimum.

To fully understand this discrepancy, the external structural constraints in which social tagging systems are embedded were taken into account. Although the Yule-Simon model does not include the possibility of the influences of external forces, social tagging systems actually evolve under the constraints imposed by a tripartite network structure in which the system is embedded: the sets of users, tags, and resources (Lambiotte & Ausloos, 2006; Lu, Hu, & Park, 2011). In that case, the change of the tag-frequency distribution is highly constrained by the structures of the two other sets (Figure 8). As presented at the bottom of Figure 8, there were already extreme power-law structures developed in both the user-login-duration distribution (slope = -3.92) and the resource-hit distribution (slope = -2.04). That is, the tagging system was dominated by a few very active users who primarily paid attention to only a few popular resources. Such highly biased structures in the user and resource sets could be a possible explanation as to why the slope of the tag-frequency distribution failed to further increase, and, thereby, the anticipated increase in functionality did not occur.

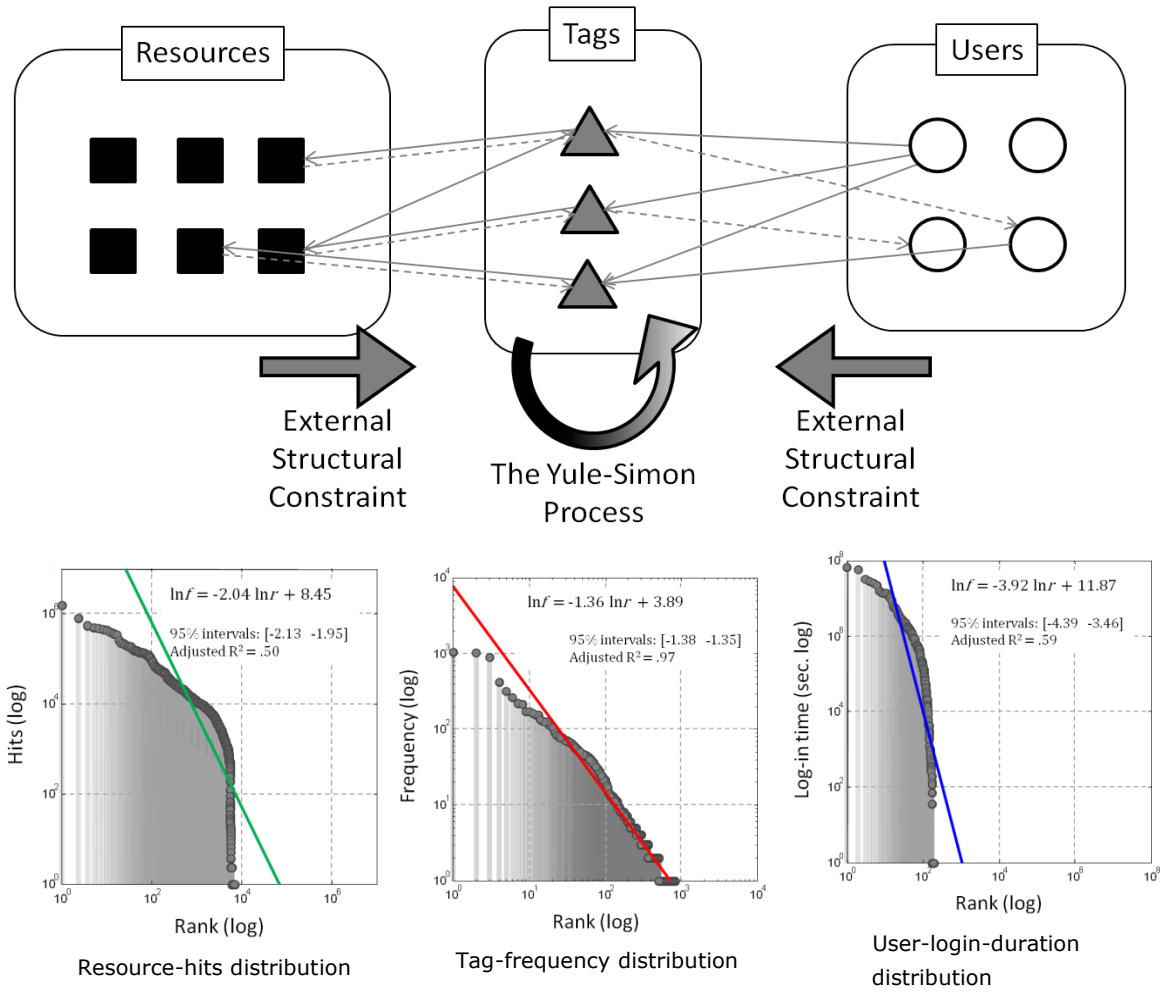


Figure 8. Structural constraints imposed on the Yule-Simon process.

Discussion and Conclusions

From three theoretical models, Shannon's (1948) general communication model, the Zipf-Mandelbrot model (Mandelbrot, 1953; Zipf, 1949), and the Yule-Simon model (Simon, 1955), two hypotheses were developed on the changes in the functionality of social tagging system. The first hypothesis predicted that the slope would initially decrease but increase again and stabilize at a point below the ideal slope. The second hypothesis predicted that the system functionality would initially increase but decrease, increase again, and stabilize at a point less than its maximum. The two hypotheses were reformulated into three and four subhypotheses, respectively. All the subhypotheses were

empirically supported except *H2c*. Although *H2c* predicted that functionality would increase after the slope reaches its bottom (at the 1,550th tagging event), the data showed that the system functionality decreased after the 1,550th tagging event. The post hoc analysis suggested that the failure to support *H2c* was due to the constraints imposed by the tripartite network structure.

One major finding of this study is that the functionality of social tagging systems is neither constant nor continuously increasing or decreasing. Instead, the functionality increased at the early stage, but declined later, and eventually stabilized. In short, the functionality is time-varying. This provides a potential explanation for the inconsistency among previous studies. At the early stage (before the 842nd event), when the slope of the tag-frequency distribution was greater than -1, individual users chose tags according to personal preferences without looking at others' choices. Therefore, a number of the tags that appeared during this period were highly subjective (e.g., *introductory*, *ideas*, *learning*) or general (e.g., *physics*, *nano*) so that they were of little use for others (Begelman et al., 2006; Macgregor & McCulloch, 2006). Later, the functionality gradually increased by the preferential attachment mechanism and showed its maximal functionality at the 842nd tagging event, when the encoding and decoding efforts were optimally balanced. This suggests the possibility that poorly informed individuals can sufficiently produce highly functional systems without centralized control, realizing the "wisdom of crowds" (Zauder et al., 2007). However, the achieved functionality was not sustained for a long time. After the 842nd tagging event, the functionality gradually decreased due to the failure of controlling undesirable noise, such as misspelling (e.g., *histroy* [history], *protiens* [proteins]), unnecessary plural forms (e.g., *crystals* [crystal], *reflections* [reflection]), and technical synonyms (e.g., *finite-difference method* and *finite-difference scheme*), which is consistent with the findings of Chi and Mytkowicz (2008). To summarize, the current study suggests that the inconsistent findings of previous studies appear to be compatible with each other when the possibility that functionality of social tagging systems changes over time is considered.

Second, the present study showed that the self-correcting capability of the social tagging system diminished over time, and the system was eventually trapped in a suboptimal state, despite its structural flexibility and decentralization. The observed slope of the tag-frequency distribution stabilized at a point much lower than the ideal slope (i.e., -1) or even one that the Yule-Simon model predicted (i.e., -1.07). This observation can be explained by the notion of self-organization of complex systems. According to Kauffman's (1993) NK model, a high level of interdependence among components of a system (i.e., the number of "epistatic links" $K > 3$) is a *sufficient* condition for complex systems being trapped in suboptima, which is called "complexity catastrophes" (p. 36). When the NK model is applied to social tagging systems, the structural flexibility and decentralization of social tagging systems facilitate interactions among users. This, in turn, increases the interdependence among the users. Accordingly, the increased interdependence enhances the likelihood of the entire system being trapped in a suboptimum. In short, the structural flexibility and decentralization, paradoxically, suppresses the self-correcting capability of social tagging systems.[5] As illustrated in Figure 8, the interdependences *within* the user set

[5] The failure of a complex system to reach its optimal state can be explained by the notions of path-dependency (for a recent review, see Schreyogg & Sydow, 2011) and Pareto-inferior Nash equilibrium (Holt & Roth, 2004). However, both of them, and many other complexity theories, commonly point out the interdependency among elements of a system as the major cause of the failure.

and the resource set and *among* the three sets (i.e., the tripartite network of users, resources, and tags) inhibited the further increase in the functionality during the late stage, which the Yule-Simon model fails to capture.[6]

One might interpret the current findings as indicating that the social tagging system became more functional over time as a resource discovery tool keeping the decoding effort at a low level, while the encoding effort increased (Figure 7a). However, this interpretation should be made with great caution, because the decoding effort was measured by the *conditional* entropy (Equation 7). That is, the measure of decoding effort assumes the condition under which users selected a *correct* tag to discover their target information resources. Therefore, if individual users do not always select correct tags, the actual decoding effort should be much higher than the conditional entropy measure. This is one of the significant limitations of the entropy measures used in this study. Hence, future research should take into account the actual patterns of tag usage to estimate both encoding and decoding costs.

Social tagging systems have made possible what was considered to be impossible even a decade ago: the creation of highly functional socially generated information organization/retrieval systems where collective contributions provide equally good, if not better, information than authority-based systems. The time has come to acknowledge and rigorously explore the potential of social tagging systems and to identify how best they can be used in conjunction with or as alternatives to existing systems. For this, the present study contributes to a better understanding of social tagging systems. Social tagging systems were examined as a whole rather than only either a resource organization or as a resource discovery tool by conceptualizing a social tagging system as a special form of communication system among users. In addition, the entire history and change in functionality of a social tagging system was examined. This facilitated integration of previously inconsistent findings into a single theoretical framework using the Yule-Simon model. Nonetheless, further research on users' actual usage patterns is necessary to attain a complete understanding of the functionality of social tagging systems.

References

[6] A proposal for the extension of the original Yule-Simon model applicable to tripartite structures is available upon request.

Ashby, R. (1957). *An introduction to cybernetics*. London, UK: Chapman & Hall.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*, 509–512. doi:10.1126/science.286.5439.509

Begelman, G., Keller, P., & Smadja, F. (2006, May). Automated tag clustering: Improving search and exploration in the tag space. Presented at the Collaborative Web Tagging Workshop, Edinburgh, Scotland.

Berkvens-Stevelinck, C. (2004). *Magna commoditas: A history of Leiden University library 1575–2000*. Leiden, The Netherlands: Primavera Pers.

Boulos, K. M. N., & Wheeler, S. (2007). The emerging Web 2.0 social software: An enabling suite of sociable technologies in health and health care education. *Health Information and Libraries Journal*, *24*, 2–23. doi:10.1111/j.1471-1842.2007.00701.x

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*. New York, NY: Oxford University Press.

Cancho, R. F. i, & Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, *100*, 788–791. doi:10.1073/pnas.0335980100

Cattuto, C., Loreto, V., & Pietronero, L. (2007). Semiotic dynamics and collaborative tagging. *Proceedings of the National Academy of Sciences*, *104*(5), 1461–1464. doi:10.1073/pnas.0610487104

Chi, E. H., & Mytkowicz, T. (2008). Understanding the efficiency of social tagging systems using information theory. In P. Brusilovsky & H. C. Davis (Eds.), *Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia* (pp. 81–88). New York, NY: ACM Press.

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (4th ed., Vol. 2, pp. 151–192). Boston, MA: McGraw-Hill.

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review, 51,* 661–703. doi:10.1137/070710111

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482. doi:10.1146/annurev-psych-120709-145346

Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, *32*, 198–208. doi:10.1177/0165551506062337

Halpin, H., Robu, V., & Shepherd, H. (2007). The complex dynamics of collaborative tagging. In C. Williamson & M. E. Zurko (Eds.), *Proceedings of the 16th International Conference on World Wide Web* (pp. 211–220). New York, NY: ACM Press. doi:10.1145/1242572.1242602

Hendry, D., Jenkins, J., & McCarthy, J. (2006). Collaborative bibliography. *Information Processing and Management*, *42*(3), 805–825.

Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*, *332*, 60–65. doi:10.1126/science.1200970

Holt, C. A., & Roth, A. E. (2004). The Nash equilibrium: A perspective. *Proceedings of the National Academy of Sciences*, *101*(12), 3999–4002.

Jacob, E. (2004). Classification and categorization: A difference that makes a difference. *Library Trends*, *52*, 515–540.

Kammerer, Y., Nairn, R., Pirolli, P., & Chi, E. H. (2009). Signpost from the masses: Learning effects in an exploratory social tag search browser. In D. R. Olsen & R. B. Arthur (Eds.), *Proceedings of the 27th International Conference on Human Factors in Computing Systems* (pp. 625–634). New York, NY: ACM Press. doi:10.1145/1518701.1518797

Kauffman, S. A. (1993). *The origins of order: Self-organization and selection in evolution*. New York, NY: Oxford University Press.

Lambiotte, R., & Ausloos, M. (2006, May). *Collaborative tagging as a tripartite network.* Presented at the International Conference on Computational Science, 2006, Reading, UK.

Lipczak, M., & Milios, E. (2011). Efficient tag recommendation for real-life data. *ACM Transactions on Intelligent Systems and Technology*, *3*, 1–21.

Lu, C., Hu, X., & Park, J. (2011). Exploiting the social tagging network for Web clustering. *IEEE Transactions on Systems Man and Cybernetics—Part A: Systems and Humans*, *41*(5), 840–852.

Macgregor, G., & McCulloch, E. (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review*, *55*(5), 291–300. doi:10.1108/00242530610667558

MacLeod, R. M. (2004). *The Library of Alexandria centre of learning in the ancient world*. London, UK: I. B. Tauris.

Mandelbrot, B. (1953). An information theory of the statistical structure of language. In W. Jackson (Ed.), *Communication theory* (pp. 486–502). New York, NY: Academic Press.

Michener, W. (2006). Meta-information concepts for ecological data management. *Ecological Informatics*, *1*, 3–7. doi:10.1016/j.ecoinf.2005.08.004

Pierce, J. (1980). *An introduction to information theory: Symbols, signals, and noise* (2nd ed.). New York, NY: Dover Publications.

Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information*. New York, NY: Oxford University Press.

Razikin, K., Goh, D. H., Chua, A. Y. K., & Lee, C. S. (2011). Social tags for resource discovery: A comparison between machine learning and user-centric approaches. *Journal of Information Science*, *37*, 391–404. doi:10.1177/0165551511408847

Schreyogg, G., & Sydow, J. (2011). Organizational path dependence: A process view. *Organization Studies*, *32*, 321–335. doi:10.1177/0170840610397481

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, *42*, 425–440.

Simon, H. A., & Van Wormer, T. A. (2007). Some Monte Carlo estimates of the Yule distribution. *Behavioral Science*, *8*, 203–210. doi:10.1002/bs.3830080305

Sinclair, J., & Cardew-Hall, M. (2007). The folksonomy tag cloud: When is it useful? *Journal of Information Science*, *34*, 15–29. doi:10.1177/0165551506078083

Spink, A. (2010). *Information behavior: An evolutionary instinct*. New York, NY: Springer.

Sumpter, D. J. T. (2010). *Collective animal behavior*. Princeton, NJ: Princeton University Press.

Tanford, S., & Penrod, S. (1984). Social influence model: A formal integration of research on majority and minority influence processes. *Psychological Bulletin*, *95*, 189–225.

Wolf, T., & Holvoet, T. (2005). Emergence versus self-organisation: Different concepts but promising when combined. In S. A. Brueckner, G. Marzo Serugendo, A. Karageorgos, & R. Nagpal (Eds.), *Engineering self-organising systems* (Vol. 3464, pp. 1–15). Berlin, Germany: Springer.

Yule, G. U. (1925). A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, *213*, 21–87.

Zanette, D., & Montemurro, M. (2005). Dynamics of text generation with realistic Zipf's distribution. *Journal of Quantitative Linguistics*, *12*, 29–40. doi:10.1080/09296170500055293

Zauder, K., Lazic, J. L., & Zorica, M. B. (2007). Collaborative tagging supported knowledge discovery. In V. Luzˇar-Stiffler & V. H. Dobrić (Eds.), *Proceedings of the 29th International Conference on Information Technology Interfaces* (pp. 437–442). Zagreb, Croatia: University of Zagreb. doi:10.1109/ITI.2007.4283810

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. New York, NY: Hafner.