# Digg it Up! Analyzing Popularity Evolution in a Web 2.0 Setting

**Symeon Papadopoulos**[1 2] and **Athena Vakali**[1] and **Ioannis Kompatsiaris**[2]

**Abstract.** The recent advent and wide adoption of Social Bookmarking Systems (SBS) has disrupted the traditional model of online content publishing. Until recently, the majority of content consumed by people was published as a result of a centralized selection process. Nowadays, large-scale adoption of the Web 2.0 paradigm has diffused the content selection to the masses. As a result, the evolution of online content popularity nowadays constitutes an overly complex phenomenon involving both semantic and social aspects.

Work in the era before the wide adoption of Web 2.0 mostly focused on estimating web resource popularity under a web graph model. Recently, several aspects of the popularity dynamics emerging in the context of Web 2.0 applications have attracted significant research interest. Here, we introduce a formalism that enables the study of the dynamics underlying the popularity evolution of online content within an SBS. Based on this formalism, we carry out a study of the digg™ popularity dynamics. Our study confirms the power law nature of content popularity in SBS, and presents new insights into the temporal aspects of popularity under the influence of the social factor.

## 1 INTRODUCTION

The startling success of Web 2.0 applications during the last years has reshaped our views on how information is generated and distributed to the masses. While in the traditional publishing model, the content selection has been carried out by a single editor or a small editing committee, the advent of Web 2.0 applications, such as the Social Bookmarking Systems (SBS), has allowed mass participation in the content selection process.

In an SBS, users upload and save links to web pages (bookmarks) that they deem interesting. These bookmarks are usually public: Once they are submitted by a user, the rest of the community is able to view them and if they consider them interesting, they may give a 'thumbs up'. Thus, these applications, originally meant to provide a convenient tool to organize one's bookmarks, are nowadays acting as a content popularity (or 'interesting-ness') ranking mechanism. Examples of such systems are del.icio.us, StumbleUpon and digg™.

Traditionally, the popularity of web pages has been inferred by means of its connectivity properties within snapshots of the web graph. The main idea behind page ranking schemes such as PageRank [15] and HITS [12] lies in that an important page is one that is linked-to by many other important pages. This is an implicit way of defining popularity and it has so far served well the search needs of web users. However, the emergence of Web 2.0 applications has led to a much richer user experience which in turn has stimulated users to provide explicit ratings/opinions on digital resources (e.g. news articles, blog posts, video clips, products, etc.). Thus, new content rating and ranking frameworks are nowadays possible that can exploit the explicit opinion of the masses as expressed within *Social Web* and *Social Media* applications.

The work presented in this paper has been mainly motivated by the fact that the digital resource rating processes, such as the ones taking place in a range of Web 2.0 applications, have not been sufficiently studied so far from the temporal and social point of view. For this reason, we established a convenient framework for the study of such aspects in an SBS and applied the proposed framework for analyzing the content rating dynamics emerging in the digg™ application[3]. More specifically, the work described in this paper makes the following contributions:

- Inspired by the *Folksonomy* framework of [13] and [8], the *Diggsonomy* framework is introduced to enable the study of web resource popularity in SBS;
- The power law nature of popularity in an SBS is confirmed through the collection and analysis of an extensive data set from digg™;
- Insights to the temporal evolution of popularity are provided and preliminary observations on the influence of the social factor on popularity are presented.

We consider the work presented here of significance for the following groups of users:

- *Administrators of SBS.*
  Since the operation of such a service involves a large investment in server and network equipment, better understanding of page popularity could contribute to smoother service provision and/or reduced infrastructure costs by means of efficient data partitioning and page caching schemes.
- *Developers of recommender systems.*
  Insight to the temporal dynamics of page popularity can be exploited to improve the precision and relevance of recommendations for systems where the temporal aspect is of significance (e.g. news articles, ads).

The rest of this paper is structured as follows. Section 2 provides a short overview of existing work in the fields of mining web 2.0 data and modelling related temporal phenomena. Further, in Section 3 the proposed analysis and modelling framework is introduced. The data collection and the associated experimentation as well as a discussion of the findings are provided in Section 4. Finally, Section 5 concludes the paper.

[1] Aristotle University, Thessaloniki, Greece, avakali@csd.auth.gr
[2] Informatics and Telematics Institute, Thessaloniki, Greece, {papadop, ikom}@iti.gr

[3] *http://www.digg.com*

## 2 RELATED WORK

Early work in the area of web content popularity has focused on the analysis of page popularity under a graph model [1], where the web constitutes the graph. Popular web page ranking algorithms, such as PageRank [15] and HITS [12] have served the search engine developers extremely well in their effort to infer the popularity of a web page based on its connectivity properties on a snapshot of the web graph. Such work is not directly related to the one presented here, but it is mentioned in order to provide additional historical background to the discussion.

The recent emergence of Web 2.0 applications has spawned a new wave of research, focused on the analysis of users' behavior within a web application and the semantics emerging from that behavior. More specifically, the advent of Social Tagging Systems (STS) motivated researchers to formalize the notion of a Folksonomy [13]. Within the Folksonomy framework, novel information retrieval paradigms were proposed [8] and established data mining methods were applied to gain better insights to the so-called *Social Web* data [4]. Such works have mostly focused on the analysis of the semantic aspects of Social Web data, with a primary focus on the tagging patterns appearing in STS; however, they do not address the evolution of such patterns in time.

The temporal aspect of Social Web data has only lately attracted significant interest from the research community. The studies in [6], [9], [7] and [5] deal with the temporal aspects of tags and user communities, but they disregard the popularity evolution of web resources, which constitutes a different phenomenon. Some large scale experimental work has been recently presented in [3], where the phenomenon of popularity evolution in *User Generated Content* systems, such as YouTube and Daum, is partly studied. A comprehensive study of the temporal aspects of web resource popularity is published in [10] and [11] where the evolution of comment activity in Slashdot is analyzed. The aforementioned studies, however, disregard the influence of the social factor on the evolution of popularity. Finally, a smaller-scale study of 'collective attention' in digg™ was presented in [17].

To our knowledge, this is the first study of popularity evolution within an SBS, comprising both temporal and social aspects. Our contributions, i.e. the extension of the well-established Folksonomy framework with a temporal component, the confirmation of the power law nature of popularity, the temporal analysis and the quantification of the social influence on popularity, hopefully provide new insights to the dynamic phenomena that appear in the context of such systems.

## 3 POPULARITY ANALYSIS FRAMEWORK

This section briefly describes the major concepts of digg™ and establishes a framework to facilitate the analysis of popularity in such a system. Note that the same framework can be applied in a multitude of similar SBS, e.g. del.icio.us, StumbleUpon or Flickr.

The motivation behind digg™ is the empowerment of simple users with the means to collectively decide upon the significance (or interesting-ness) of web items (mostly news items, images and videos; for convenience, these will be referred to as 'stories' in the rest of the paper); digg™ attempts to promote the vision of Social Media. This is achieved by providing registered users with two basic rights: (a) submitting stories that they deem interesting, (b) voting on previously submitted links. Votes can be either positive (i.e. the user 'diggs the story') or negative (i.e. the user 'buries the story'), but only the positive ones, namely the 'Diggs', are available for public view. A user may vote in favor or against a particular story only once.

### 3.1 Diggsonomy, a time-aware Folksonomy

Registered users of digg™, i.e. *diggers*, similarly to users of other Web 2.0 applications, form social networks with the intention of keeping track of other diggers, whose submissions they find interesting. Since digg™ offers two kinds of relations between users, namely 'has-friend' and 'is-fan-of', the resulting network has the form of a directed graph.

One would expect that the *digging* behavior of users is partly influenced by the digging patterns of their social network. In order to study the social influence on digging in a formal way, we first define the Diggsonomy formalism in the following. We consider an SBS and the finite sets $U, R, T, S, D$ which describe the set of users, resources, timestamps, social relations and user Diggs on resources respectively. Note that $T$ is an ordered set.

**Definition 1** (DIGGSONOMY OF AN SBS) *Given a Social Bookmarking System (SBS), its derived Diggsonomy $\boldsymbol{B}$ is defined as the tuple $\boldsymbol{B} = (U, R, T, S, D)$, where $S \subseteq U \times U$ is the social network of the SBS users, and $D \subseteq U \times R \times T$ is the users' voting set* [4] *$D$, modeled as a triadic relation between the other sets.*

**Definition 2** (PERSONOMY) *The personomy $\boldsymbol{P}_u$ of a given user $u \in U$ is the restriction of $\boldsymbol{B}$ to $u$, i.e. $\boldsymbol{P}_u = (R_u, S_u, D_u)$ with $D_u = \{(r,t) \in R \times T | (u,r,t) \in D\}$, $S_u = \pi_U(S)$ and $R_u = \pi_R(D_u)$.*

**Definition 3** (DIGG HISTORY) *The Digg history for a particular resource (story) $r$, denoted as $H_r$ is equal to the projection of $D$ on $U \times T$ restricted on $r$, i.e. $H_r = \pi_{U\times T}(D|r) \subseteq U \times T$. The user $u_0$ for whom the statements $(u_0, t_0) \in H_r$ and $\forall t \in \pi_T(H_r), t_0 < t$ hold is called the submitter of the story.*

This formalism is similar to the Folksonomy definitions appearing in [13] and [8]. The major difference is that the Diggsonomy formalism focuses mostly on the temporal aspect of popularity, therefore we employ the ordered set $T$ of timestamps in the definitions. Also, the set of tags which was part of the Folksonomy formalism does not appear here.

### 3.2 The power law nature of popularity

A widely researched and empirically supported model for popularity is the power law distribution. A comprehensive review of the properties observed in such distributions is provided in [14]. According to this model, the probability density function of the number of Diggs that a story will receive should follow the following law:

$$p(x) = Cx^{-\alpha}, \tag{1}$$

where $\alpha$ is called the exponent of the power law (the constant C is part of the model in order to satisfy the requirement that the distribution $p(x)$ sums to 1). Note that the random variable $X$ we are trying to model by the power law distribution is essentially the cardinality of the Digg history set of Definition 3.

A straightforward way to empirically identify a power-law in a measured quantity is to plot its histogram. However, this might be tricky in practice since the tail of the distribution would appear very

[4] Only positive votes are considered.

noisy (due to the regular histogram binning which is not appropriate for functions following the power-law). A potential solution to this problem would be to employ logarithmic binning; however, a more elegant way to deal with the problem is to calculate and plot the cumulative distribution function of $p(x)$, $P(x) = \int_0^\infty p(x')dx'$, where the random variable $X$ in our study is the number of Diggs that a story receives. Assuming that $p(x)$ follows a power law as described by Equation 1, we get:

$$P(x) = C \int_0^\infty x'^{-\alpha} dx' = \frac{C}{a-1} x^{-(\alpha-1)}.$$

Thus, the cumulative distribution function $P(x)$ also follows a power law, but with a different exponent $\alpha - 1$.

The value for the exponent $\alpha$ of the power law can be approximated by use of the formula proposed in [14]:

$$\alpha = 1 + n \left[ \sum_{i=1}^{n} ln \frac{x_i}{x_{min}} \right]^{-1}. \qquad (2)$$

Here, the quantities $x_i, i = 1...n$ are the measured values of x and $x_{min}$ is the minimum value of $x$ for which the power-law holds. The result of fitting the measured Diggs distribution to this model appear in the next section of the paper.

## 3.3 Temporal evolution of popularity

Most research related to time series analysis has focused on the assumption of stationarity for the time series under study [2]. Furthermore, these series span usually long intervals, so it is possible to derive stable statistics about their temporal behavior. Contrary to these assumptions, digital content popularity is highly non-stationary and has a transient temporal nature. This is recognized by [11], where the user reaction time to a discussion thread is modelled by means of a log-normal distribution.

In digg™, the popularity of a submitted resource may evolve in two ways: (a) the resource reaches a plateau of popularity while in the 'Upcoming' section of the site and remains there until it is completely removed in case it does not receive any Diggs for a long time, (b) the resource attains the 'Popular' status after some time and is moved to the 'Popular' section, where it undergoes a second-stage popularity growth of a much higher magnitude. Figure 1 depicts the two possible ways in which the popularity of a story may evolve. For convenience, we denote the set of stories of the first type as $R_U$ and the set of popular stories as $R_P$. Note that it is not the number of Diggs collected by a story that determines whether it will jump to the 'Popular' section (although it is certainly taken into account). The digg™ administrators use a set of proprietary criteria and heuristics to make that decision. Obviously, sharing such knowledge would render the system prone to malicious attacks (artificial population boosting or burying).

In order to study the temporal evolution of story popularity at a mass level, we consider the projection on $T$ of the Digg history set $H_r$ of Definition 3 for each story, denoted by $T_r = \pi_T(H_r)$. This projection can be considered as a random variable with a discrete probability density function (pdf) $p(t = T)$. For each story, we perform the following transformation of $T_r$:

$$T'_r = \begin{cases} \frac{T_r - min(T_r)}{max(T_r) - min(T_r)} & if \quad r \in R_U, \\ \frac{T_r - min(T_r)}{tp(T_r) - min(T_r)} R[T_r, min(T_r), tp(T_r)] & if \quad r \in R_P. \end{cases} \qquad (3)$$

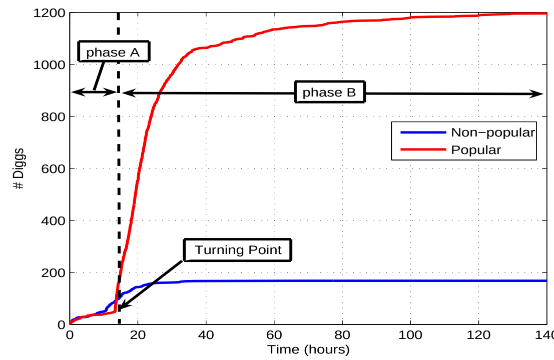


**Figure 1.** Two possible popularity evolution patterns. Note the two different phases in the popularity evolution for popular stories.

where $min(T)$, $max(T)$ return the minimum and maximum timestamp values of the input variable, $tp(T)$ is the function returning the timestamp corresponding to the turning point of popularity (i.e. the moment when the story is moved from the 'Upcoming' to the 'Popular' section), and $R[t, t_0, t_1]$ is the rectangle function, with non-zero values only for $t \in [t_0, t_1)$.

Effectively, our transformation maps the description of the full lifetime of the non-popular stories and the first growth-stage of the popular ones (phase A) from the timestamp domain to the $[0, 1]$ interval. That permits a direct timestamp pdf comparison between the non-popular and popular stories as will be presented in the next section. At a second level, and in order to study the evolution of story popularity for those story that attain the status of 'Popular', we apply the following transformation:

$$T'_r = \frac{T_r - tp(T_r)}{max(T_r) - tp(T_r)} H[T_r - tp(T_r)] \quad if \quad r \in R_P, \qquad (4)$$

where H[t] is the Heaviside step function. This transformation maps the second popularity growth phase (phase B) of popular stories also to the $[0, 1]$ interval, thus permitting comparison with the other popularity phases.

## 3.4 Social network influence on digging

Based on the Diggsonomy formalism, we also define two interesting measures of the social influence on digging behavior, namely the *user social susceptibility* $I_u$, and the *story social influence gain* $I_r$.

**Definition 4** (USER SOCIAL SUSCEPTIBILITY) *The social susceptibility of a given user u ($USS_u$) quantifies the extent to which his/her voting behavior follows (in a temporal sense) the behavior of his/her friends' voting behavior.*

$$I_u = \frac{|D'_u|}{|D_u|}, \qquad where \qquad (5)$$
$$D'_u = \{(r_u, t_u) \in D_u | \exists f \in S_U, (r_u, t_f) \in D_f, t_f < t_u\}$$

**Definition 5** (STORY SOCIAL INFLUENCE GAIN) *The social influence gain for a given story r ($SSIG_r$) is a measure of the extent to which r has benefited from the social network of the story submitter.*

$$I_r = \frac{|H'_r|}{|H_r|}, \qquad where \qquad (6)$$
$$H'_r = \{(u, t_k) \in H_r | \exists (u_0, t_0) \in H_r, u_0 \in S_u, t_0 < t_k\},$$

*and $u_0$ is the submitter of the story as defined in Definition 3.*

| General Topics | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Statistics** | Technology | World & Business | Science | Gaming | Lifestyle | Entertainment | Sports | Offbeat | TOTAL |
| # Stories | 16,257 | 25,894 | 4,504 | 6,182 | 18,070 | 15,088 | 5,604 | 13,509 | 105,108 |
| # Domains | 6,999 | 8,048 | 2,094 | 2,071 | 7,811 | 4,775 | 2,052 | 5,260 | 30,944 |
| # Users | 7,604 | 9,786 | 2,573 | 2,712 | 7,966 | 6,073 | 2,382 | 7,251 | 34,593 |
| Avg # Diggs | 17.31 | 14.19 | 28.69 | 18.43 | 8.47 | 9.28 | 9.11 | 22.12 | 14.60 |
| Max # Diggs | 6,886 | 4,964 | 2,991 | 5,442 | 4,419 | 7,842 | 3,236 | 8,517 | 8,517 |
| Avg St. Life (hours) | 29.93 | 29.05 | 37.96 | 26.99 | 22.79 | 21.31 | 26.29 | 27.53 | 26.92 |
| Max St. Life (hours) | 397.84 | 406.60 | 390.35 | 395.93 | 406.73 | 395.91 | 394.01 | 398.90 | 358.33 |

**Table 1.** Data set statistics

$USS$ may be attributed to a combination of the following: (a) an inherent tendency of friends to have similar interests and (b) the possibility for users to see through the digg™ interface which stories their friends have already dugg. Similar observations hold for $SSIG_r$.

# 4 EXPERIMENTS

This section describes the data collection and the experimentation phase carried out in the context of this study based on the introduced framework. Due to paper size limitations, parts of the results and the associated discussion had to be omitted.

## 4.1 Data set collection

The collection of data from digg™ was carried out by means of the public API which has been made available from the service. During the week between 24 and 30 April 2008, the stories appearing in the 'Upcoming' section of the site were downloaded, locally stored and monitored for the two following weeks. A total of 109,360 stories appeared in the site during this period, which indicates an average story submission rate of over 15,000 stories per day (or 650 stories per hour). Out of this initial story set, 105,108 stories were retained, since 3.88% of the initially collected story set vanished from digg™ very soon. Table 1 summarizes the statistics of this data set.

The story popularity tracking was carried out by querying the API at predetermined intervals for potential new Diggs for each of the initially stored stories. As the stories grew older these intervals were increased since, as it can be deduced from Figure 3, the majority of Diggs for a given story are received on average during the first 10% of the story lifetime. In this way, redundant calls to the service were kept to a minimum.

## 4.2 Results

Analysis of the collected data revealed that the distribution of Diggs in stories follows a power law form as described by Equation 1. Figure 2 depicts in logarithmic scales the cumulative distribution of the number of Diggs that stories receive. In the same figure, the power-laws for $\alpha = 0.84$ and $\alpha = 1.0$ are plotted. The value $\alpha = 0.84$ was estimated through Equation 2 with $x_{min} = 100$. Since this is the exponent of the cumulative distribution of Diggs, the actual distribution has an exponent of $\alpha' = \alpha + 1 = 1.84$ as discussed in the previous section.

The temporal study of story popularity curves was subsequently carried out according to the methodology presented in the previous

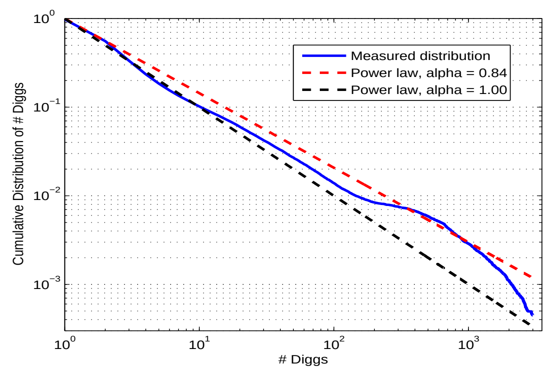


**Figure 2.** The Power Law nature of Digg popularity

section. The first objective of the temporal study was to compare the Digg arrival time distribution between the popular and non-popular stories. The histograms of 5,469 non-popular and 839 popular stories were aggregated [5]. After applying the transformation of Equation 3 and aggregating the resulting histograms over the two phases of popular stories and the full lifetimes of non-popular stories, we ended up with the distributions of Figure 3. The figure clearly illustrates the fact that while the non-popular stories gather the majority of their Diggs during the very first moments of their lifetime (first two bins in the histogram), the popular ones manage to sustain a relatively high Digg influx for the whole duration of their first phase of popularity growth.

Moreover, Figure 3 reveals an additional interesting aspect of the story popularity evolution, when the full lifetime of unpopular stories is compared to the high-growth phase (phase B) of popular stories (phase B corresponds to the pdf of the variable in Equation 4).

It appears that once a story is moved to the 'Popular' section of digg™, its popularity growth takes place extremely fast. After such a growth period, only minor popularity gains are possible. This is not surprising, since stories remain in the first places of the 'Popular' section of the site only for a few minutes, during which they are exposed to the largest possible user base (since the first page of the 'Popular' section is the default home page of digg™). Similar patterns were reported in the case of sponsored links appearing next to search engine results, where the vast majority of click-throughs came from just the top two or three ads on the page [16].

The last part of the experimental phase of our work involved the

[5] Only stories with $|H_r| \geq 20$ were studied to prevent noisy histograms from distorting the resulting aggregate histogram of the non-popular stories.

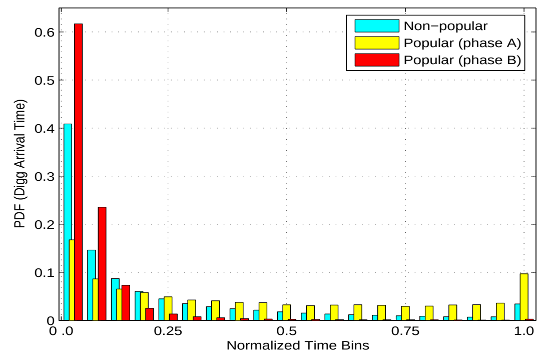


**Figure 3.** Digg arrival time distributions of popular vs. non-popular stories

estimation of the SSIG values, as introduced in Equation 5, for the available stories [6]. The histogram of Figure 4 clearly depicts the difference in story social influence gain between the popular and the non-popular stories. The histogram indicates that the event of a story with $I_r > 0.4$ becoming popular is highly unlikely. Instead of concluding that high social influence gain for a story implies low probability of becoming popular, we would rather speculate that digg™ employs filters based on measures similar to $I_r$ to prevent groups of 'friends' from gaining control over which stories appear in the 'Popular' section [7].

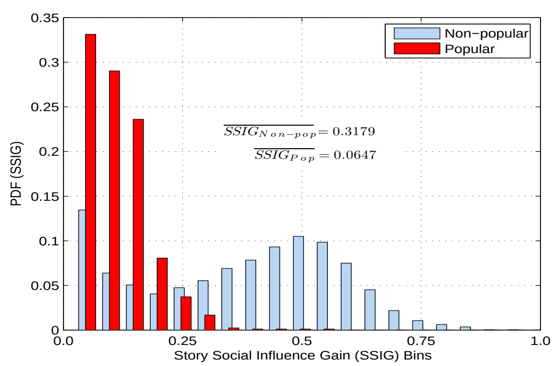


**Figure 4.** Story Social Influence Gain (SSIG)

## 5 CONCLUSION

We presented a study of web resource popularity within the context of an SBS, namely digg™. The study was based on the Diggsonomy formalism which endows the well-established Folksonomy with a temporal aspect. The analysis of the collected data confirmed the power law nature of popularity and provided new insights to the temporal aspect of the phenomenon. Finally, preliminary results were reported on the significance of two measures for the evolution of resource popularity in an SBS: the social susceptibility of a user and the social information gain of a story.

In the future, we plan to extend the depth of this study and to investigate the feasibility of predicting web content popularity (within an SBS context) by taking into account different features of web resources, namely temporal, semantic, and social. Furthermore, we intend to expand the scope of our studies to data sets of larger amounts and of different origin (e.g. del.icio.us, flickr).

[6] The filtering rule of $|H_r| \geq 20$ (same as above) was applied here too.
[7] The post in *http://blog.digg.com/?p=106* reinforces this speculation.

## ACKNOWLEDGEMENTS

This work was supported by the BOEMIE, MESH and WeKnowIt projects, partially funded by the European Commission, under contract numbers FP6-027538, FP6-027685 and FP7-215453 respectively. Furthermore, improvements to this paper were possible thanks to the valuable comments of the reviewers.

## REFERENCES

[1] A. L. Barabasi and R. Albert, 'Emergence of scaling in random networks', *Science*, **286**(5439), 509–512, (October 1999).

[2] George Edward Pelham Box and Gwilym M. Jenkins, *Time Series Analysis: Forecasting and Control*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 1994.

[3] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon, 'I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System', in *ACM Internet Measurement Conference*, (October 2007).

[4] A. Vakali E. Giannakidou, V. Koutsonikola and I. Kompatsiaris, 'Co-clustering tags and social data sources', in *9th International Conference On Web-Age Information Management (WAIM 2008)*.

[5] Tanja Falkowski and Myra Spiliopoulou, 'Users in volatile communities: Studying active participation and community evolution', *User Modeling 2007*, 47–56, (2007).

[6] Scott A. Golder and Bernardo A. Huberman, 'Usage patterns of collaborative tagging systems', *J. Inf. Sci.*, **32**(2), 198–208, (2006).

[7] Harry Halpin, Valentin Robu, and Hana Shepherd, 'The complex dynamics of collaborative tagging', in *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pp. 211–220, New York, NY, USA, (2007). ACM Press.

[8] Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme, 'Information retrieval in folksonomies: Search and ranking', in *Proceedings of the 3rd European Semantic Web Conference*, volume 4011 of *LNCS*, pp. 411–426, Budva, Montenegro, (June 2006). Springer.

[9] Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme, 'Trend detection in folksonomies', in *Proceedings of the First International Conference on Semantics and Digital Media Technologies (SAMT 2006)*, pp. 56–70. Springer, (December 2006).

[10] Andreas Kaltenbrunner, Vicenç Gómez, and Vicente López, 'Description and prediction of slashdot activity', in *Proceedings of the 5th Latin American Web Congress (LA-WEB 2007)*, Santiago de Chile, (2007). IEEE Computer Society.

[11] Andreas Kaltenbrunner, Vicenç Gómez, Ayman Moghnieh, Rodrigo Meza, Josep Blat, and Vicente López, 'Homogeneous temporal activity patterns in a large online communication space', in *Proceedings of the BIS 2007 Workshop on Social Aspects of the Web (SAW 2007)*, Poznan, Poland, (2007). CEUR-WS.

[12] Jon M. Kleinberg, 'Authoritative sources in a hyperlinked environment', *Journal of the ACM*, **46**(5), 604–632, (1999).

[13] Peter Mika, 'Ontologies are us: A unified model of social networks and semantics', in *International Semantic Web Conference*, pp. 522–536, (2005).

[14] M. Newman, 'Power laws, Pareto distributions and Zipf's law', *Contemporary Physics*, **46**, 323–351, (September 2005).

[15] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd, 'The pagerank citation ranking: Bringing order to the web', Technical report, Stanford Digital Library Technologies Project, (1998).

[16] Matthew Richardson, Ewa Dominowska, and Robert Ragno, 'Predicting clicks: estimating the click-through rate for new ads', in *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pp. 521–530, New York, NY, USA, (2007). ACM.

[17] Fang Wu and Bernardo A. Huberman, 'Novelty and collective attention', *Proceedings of the National Academy of Sciences*, **104**(45), 17599–17601, (November 2007).