Digging up the Dirt on User Generated Content Consumption

Jake T. Lussier, Troy Raeder, Nitesh V. Chawla

Interdisciplinary Center for Network Science and Applications
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, IN 46556 USA
{jlussier, traeder, nchawla}@nd.edu

Abstract

Knowledge-sharing online social networks are becoming increasingly pervasive and popular. However, consumption of user generated content in these networks has not been studied extensively despite its significant influence on many social network behaviors. In this work, we study data gathered from `digg.com` and `ireport.com` for studying the creation and consumption behavior. We build on our prior work to explore how and why individuals consume content. We also extract creation and consumption features and construct a simple classification model for predicting whether a submission will be consumed in the future.

1 Introduction

The recent emergence of online social networks (OSNs) has changed the manner in which web content is both created and used. In some cases, web sites have created entirely new classes of virtual content. For example, social networks on Facebook, dynamic career profiles on LinkedIn, and even separate virtual worlds in Second Life have all affected, and in some cases dominated, human-to-human interactions. Regardless of the specific application, nearly all OSNs allow users the opportunity to create and consume content. Given the incredibly diverse range of existing OSNs, this content, commonly referred to as User Generated Content (UGC), may be the only common thread in these networks.

Accordingly, it is useful to characterize online social networks by the role that the UGC plays. One useful distinction, as described by Guo et al. in [8], is whether an online social network is oriented toward networking or knowledge-sharing. Networking oriented OSNs are those in which the formation and sustenance of social links are the primary purpose, and the sharing of UGC is only a consequence of this. Some examples of these networks are Facebook, Twitter, MySpace, and LinkedIn. In knowledge-sharing oriented networks, on the other hand, the creation and consumption of UGC is most important, and people form social ties in order to facilitate these processes. Some examples of these networks in popular culture are iReport, Digg, and Youtube. It should be noted that there is no hard line between the two types of OSNs. For example, a Facebook user may occasionally friend another user simply to share information, or a blogger might friend another blogger because of a "real-life" friendship with no intent of information sharing in mind. However, it is still the primary purpose and role of an OSN that defines it.

International Conference on Management of Data
COMAD 2010, Nagpur, India, December 8–10, 2010

We consider two knowledge-sharing oriented OSNs – Digg and CNN's iReport. While we are the first study to include the iReport data (to the best of our knowledge), there have been a few papers analyzing Digg content consumption [17, 18]. These works deal primarily with the characterization of future consumption behavior based on past behavior, touching only tangentially on network aspects. Our work is broader in scope, seeking to understand and characterize consumption of UGC as a dynamic process with both network and external components.

To this end, we first consider the network statistics and explain how and why consumption patterns and social ties are intimately related. We point the reader to our prior work for the patterns that demonstrate *homophily*, the *effect of network* in popularization of the content, and the notion of *imbalanced reciprocity* in the relationships on Digg [15]. In our prior work, we demonstrated that individuals' consumption habits influence their friend networks, consistent with the concept of homophily. We also showed that one's social network can also influence the consumption of a submission through the activation of an extended friend network. Finally, we investigated the level of reciprocity, or balance, in the network and uncover relationships that are significantly less balanced than expected. In this paper, we present activity patterns which support the intuition that consumers typically intake content, process it, and then sometimes react to it (see Section 4). We also construct a simple classification model to demonstrate the predictability of popularity range for a story based on various characteristics.

2 Website and Dataset Specifics

Launched in 2004, `digg.com` was intended to democratize digital media. Digg allows users to discover and share content from anywhere on the web by posting a URL, indicating whether it is a story, video, or image, and then providing a short description. Other users then comment on the content, or simply "digg" (like) or "bury" (dislike) it. Once a submission has earned enough diggs, it becomes "popular" and jumps to the homepage in its category. Stories that are not yet popular are listed in the "upcoming" section. Finally, Digg allows users to add others to their social networks. If user A adds user B to his or her social network, A becomes a *fan* of B. This unidirectional link allows the initiator to monitor the other's activity. Specifically, once A nominates B, A can see any stories that B submits or diggs through a special "friend" interface. If B reciprocates and returns A's friendship, then A and B are called *friends*. Since its launch, Digg has grown to over two million users and has prompted the creation and growth of other social networking sites centered on story creation and dispersion.

One such site is `ireport.com`, CNN's public journalism initiative that allows users to post news-related stories or videos that other users can then comment on. Launched in 2006, iReport utilizes the reporting power of the masses, and thus enables CNN to obtain unique first-hand accounts of breaking news reports that can then be displayed on CNN's own site and broadcast on its news channel. And indeed, this tool has proved an excellent resource. In July 2009 alone, iReport attracted over 320,000 submissions worldwide, and CNN had featured 699 of these stories.

Before examining topics directly related to consumption, we will first describe the social network structures to illustrate how they compare to previously studied OSNs. Accordingly, the first dataset we will consider is from a single crawl of iReport.com, which returned 21,436 stories, 77,943 comments, and the activity of 26,150 unique users overall. More specifically, we will examine several networks constructed from this data, all of which consist of nodes which represent users and directed links which represent comments from one user on another user's post. One of these networks will represent the entire iReport community, and the others will only include users who post or comment on stories with specific tags (Obama and Weather in this paper).

We will then examine the data of a single crawl of digg.com, which returned 6,073,456 friend relationships and 564,193 users. Here, too, we will represent users as nodes and friendships as directed links. More specifically, we consider a directed edge from A to B if A has added B to his network of friends.

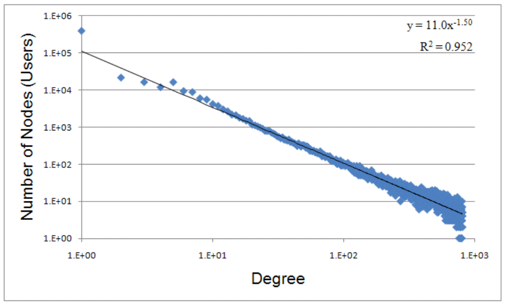
After this social network analysis, we will move on to study Digg consumption patterns. Because submission-date and comment-date information on iReport is unreliable (dates are given relative to the crawl time at very weak granularity such as "6 months ago") we are unable to study consumption on the iReport dataset. As a result, *only results from the Digg data are presented.*

3 Social Network Analysis

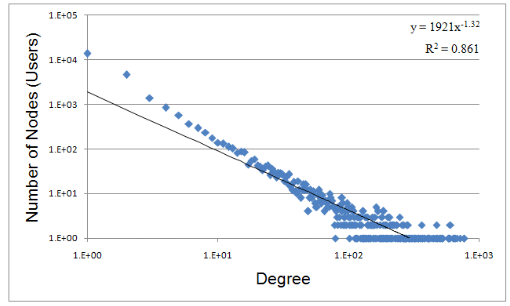
In this section, we situate this work relative to related works by analyzing the social networks. First, we consider the networks derived from the iReport data. All networks are constructed as mentioned earlier, with A linked to B by a directed edge if A commented on B's story. However, each network has different restrictions regarding whether or not to include a particular node in the network. One network includes all nodes and thus offers a complete representation of the iReport community found in the crawl. Three other networks were constructed from users that submitted posts or comments on stories with a particular tag. The tags used to construct these networks were "Obama" and "Weather." Some simple statistics for these networks are displayed in Table 1 where "Complete" refers to the statistics for the entire network and all other column labels refer to the particular tag used to construct that network. For a given network s, $N(s)$ is the number of nodes, $E(s)$ is the number of edges, $C(s)$ is the clustering coefficient, and $D(s)$ is the average node degree. We also calculated node-degree distributions for each network.

Our crawl of Digg.com also allowed us to construct a network to represent all the users and interactions captured in the crawl, with the structure as described previously. The statistics for this network are shown in Table 1. Moreover, the node-degree distribution for the entire Digg network is shown in Figure 1.

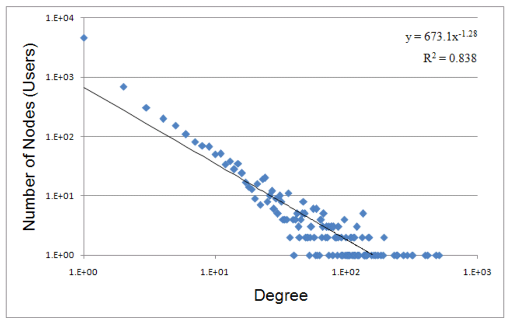
An interesting finding shown in Table 1 is that all of these social networks have small clustering coefficients. This indicates that inidividuals in these networks do not tend to cluster together. However, in examining the node-degree distributions, it is also obvious that all of the networks have recognizable "fat tails," indicating that although most users have relatively few connections, there are a substantial number of users who are extremely well connected. One can interpret this phenomenon as meaning that news flows throughout these networks in such a way that many individuals learn of a story but very few learn of it first-hand. Moreover, the distribution suggests that a loss of a



(a)



(b)



(c)

Figure 1: Degree distributions for (a) Digg users, (b) iReport users, and (c) iReport users who post/comment on Obama stories.

Table 1: Statistics for Social Networks. N(s): number of nodes, E(s): number of edges, C(s): Average clustering coefficient, D(s): Average degree.

	Digg	iReport	Obama	Weather
N(s)	564,193	26,150	7,239	2,309
E(s)	6,073,456	77,942	19,091	3,065
C(s)	0.075	0.106	0.048	0.019
D(s)	16.146	5.961	5.274	2.655

small group of iReporters or Diggers on average will not drastically affect the network. The only significant damage could be done if one of the "power users" in the fat tail were to be removed from the network. This effect would certainly be damaging for these networks since the observed clustering coefficients are unusually low, meaning that the neighbors of most hubs are generally unconnected. In this way, the loss of a single hub would mean the disconnection of entire groups of hub neighbors.

These network phenomena can be seen in our visualization of the Weather network shown in Figure 2. First, the majority of the nodes in this visualization are connected to a small set of neighbors (2.665 on average) and do not cluster particularly closely, consistent with the small clustering coefficient. Second, some of the nodes along the outside of the visualization are extremely well-connected hubs that demonstrate the presence of the "power users" indicated by the "fat tail" distributions.

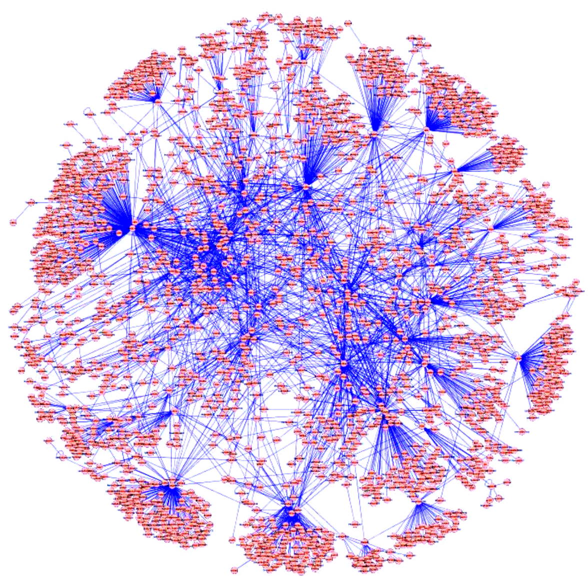


Figure 2: Visualization of Weather network.

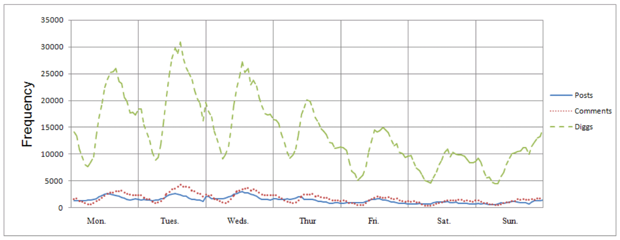


Figure 3: Average weekly activity on Digg.

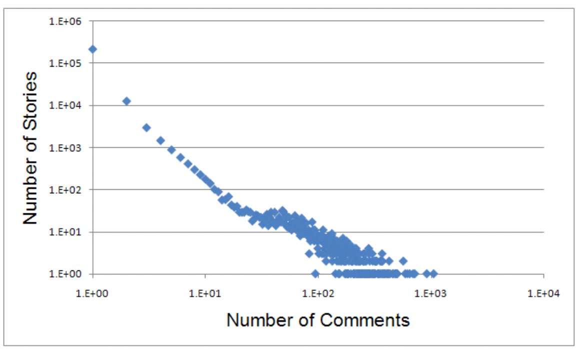


Figure 4: Distribution of number of comments on stories.

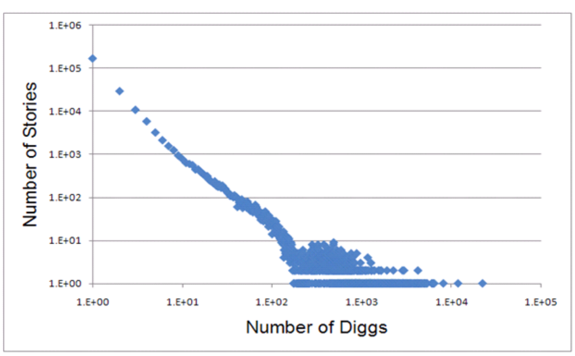


Figure 5: Distribution of number of diggs on stories.

4 Creation and Consumption of UGC

With the social networks clearly explained, we now study how and why individuals consume content. To do so, we provide findings regarding consumption patterns, story-digg/comment distributions, and story-lifetime distributions.

4.1 Creation and Consumption Patterns

Guo et al. [8] illustrate that UGC creation in blog, social bookmarking, and Q&A sites follows strong daily and weekly patterns. In other words, the frequency of content creation for a given application has typical times of maximum and minimum activity over the course of a day and over the course of the week. Moreover, different patterns emerge on weekdays than weekends.

We further investigate this phenomenon by studying activity not only for creation (submissions), but also for consumption (diggs and comments). We do this by separately binning submissions, comments, and diggs into the 24 hours of the seven days of the week over the entire period crawled. For example, we bin all posts between 12:01 AM and 1:00 AM on any Monday together, and would do similarly for any other day of the week for submissions, diggs, or comments. The results can be seen in Figure 3.

As can be seen, the posts do indeed appear to follow a weekly pattern, with peaks occurring every day and weekends exhibiting less activity than weekdays. While these results simply confirm the observations reported in [8], the consumption behaviors offer new insights into how users react to creation online. First, the similar shapes of the curves indicate that higher/lower creation activity generally corresponds to higher/lower consumption activity. Second, peak submission activity is generally followed by peak digg activity which is then followed by peak comment activity. More specifically, if we take each individual day and consider the hour with the highest number of submissions as that day's peak submission hour, and do likewise for comments and diggs, we find that the peak digg and comment hours typically follow the peak submission hours by about 1.5 and 2.5 hours, respectively. This suggests that most submissions are not immediately consumed, but rather require a certain amount of reaction time during which consumers first read or view the content, then digg it, and finally comment on it. Although "digging" is unique to digg.com, we can generalize this chain of events for any knowledge-sharing OSN by breaking it down into intake, processing, and response.

A third insight from these results is the seemingly intimate relationship between creators and consumers in knowledge-sharing oriented OSNs. To see this, first note that Figure 3 indicates a fairly short time lag between creation and consumption in these networks

OSNs. If creation and consumption were largely decoupled as (imagine a newspaper), we might expect to see large gaps between the time the information is created and when it is first consumed. Thus, our result suggests that in the world of UGC, there is no significant distinction between creators and consumers. Rather, the creators are consumers and the consumers are creators.

4.2 Story-Digg/Comment Distributions

Considering creation and content alongside one another leads to the question, "How is content typically consumed?" One way to answer this question in terms of the Digg data is to look at how many diggs / comments are typically made on a submission. Accordingly, Figures 4 and 5 show the Story-Comment and Story-Digg distributions, respectively, where the x-axis denotes the number of comments / diggs on a story, and the y-axis shows the frequency of that number of comments / diggs. As can be seen in these figures, each distribution has a "fat tail", indicating that a small number of stories that attract a disproportionately large number of posts. This can signify either that there is a core set of users to sustain consumption activity throughout small fluctuations, or that a core set of stories dominate activity.

There is unusual behavior in the middle of these distributions. In particular, the Story-Comment distribution levels off around the 25-50 comment range. The Story-Digg distribution, on the other hand, illustrates that there are a higher number of stories with a little under 100 diggs than one might expect. Additionally, there is a steep drop off after 100 diggs. In looking at these phenomena, we conjecture that Digg bots may post content and then recruit other bots to comment on the story and digg the story to approximately 100 diggs. This explanation seems likely, especially since Digg has publicly acknowledged issues with bots. Indeed, digg oftentimes bans users who exhibit any form of suspicious activity in order to prevent scripts from being used to automatically digg stories[1]. In light of this, it would be interesting to construct a network and look at users whose stories are always dugg by a core group of users, or look at users whose stories often receive around 100 diggs, to determine if this could offer any insights into bot detection.

4.3 Story-Lifetime Distribution

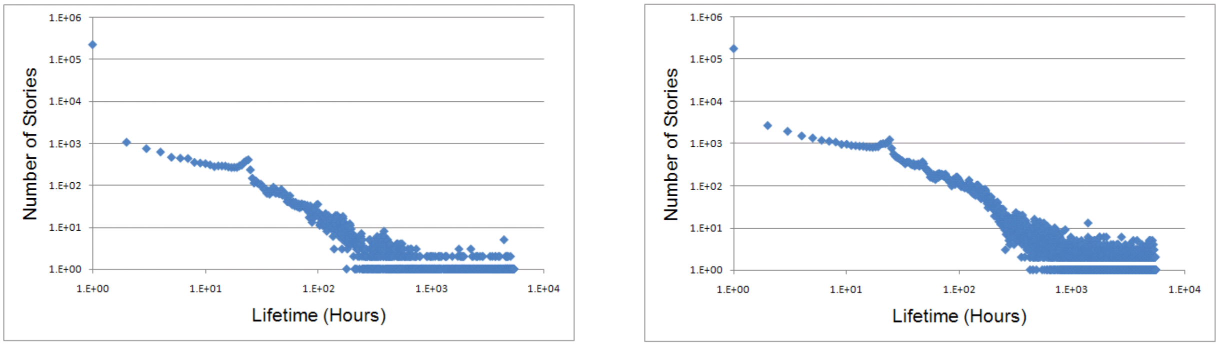
Another way to answer the "How is content typically consumed?" question in terms of the Digg data is to consider the duration over which a story is dugg / commented on. For this purpose, we define the lifetime of a story as the duration between the time of the story submission and the last comment / digg. With this, Figure 6 shows two Story-Lifetime distributions that define lifetime in terms of comment and digg duration, respectively. As can be seen, each distribution starts at a maximum, meaning that most stories live very short lives. From there, there is a steep decrease followed by an increase to a local maximum at about 24 hours. Around this point, the distribution decreases steeply again for a short while, only to level off to a power-law-like distribution with an extremely fat tail.

One might assume that such behavior is the result of something like the "Top in 24 Hr" tab on Digg. After all, it seems reasonable that after stories fall off the "Most Recent" page, they are rarely commented on, unless they are seen again on the "Top in 24 Hr" page. Such a phenomenon would explain why many stories have short lives and a good number live a little under 24 hours. However, only stories that obtain "popular" status reach this page, and even plots of story-lifetime distribution for stories that are never popular have this same behavior (not shown in interest of space). Accordingly, a different explanation is needed, and for this, we turn again to the relationship between creation and consumption patterns. As described in Section 4.2, Guo et al. showed that creation patterns follow daily and weekly trends. Moreover, we showed the consumption patterns are directly related to these creation patterns, oftentimes exhibiting similar behavior with a slight lag due to necessary reaction time. Therefore, it follows that consumption patterns must also show daily and weekly patterns. In particular, consumption activity would typically peak once a day. In this way, consumption activity can be seen as forming a sort of standing wave with a period of about 24 hours. We can thus hypothesize that the unusual spike in lifetimes around 24 hours is due to stories that miss the wave maximum upon submission, but are then consumed the next time the wave rises. This behavior then stops after 24 hours because UGC turnover is so fast that anything over a day old that has not received any attention is highly unlikely to receive any after this.

5 Lifetime Prediction

As mentioned in the introduction, one use for understanding UGC consumption patterns is to be able to discriminate between promising and dying content. To this end, we use initial generation and consumption patterns to determine whether or not the content will persist. Given the temporal aging of the content, we consider this as a classification problem in which we attempt to use the first 12 hours of a submission's life to predict whether or not it is going to receive any more comments. Thus, our framing of the problem results in a binary class – life beyond 12 hours (class 1) or not (class 0). We consider a number of features that are indicative of the activity on the content posted. Note that this binary class construction led to an imbal-

[1] `http://promote-my-site.com/index.php/258-Digg-is-on-a-Banning-Rampage.html`



(a) (b)

Figure 6: Lifetime distributions with lifetime defined by (a) comment duration and (b) digg duration.

anced dataset – 232,500 stories had no comments and 10,083 stories had at least one more comment after the first 12 hours.

5.1 Feature Extraction

If the first 12 hours after a given submission are to tell us anything about the remainder of the UGC's life, one could assume such an insight would be because of the existing generation and/or consumption information. Accordingly, we extract features of both types and assess their effectiveness. Note that other aspects of these social networks could also be incorporated into our method, but our goal is to assess the effectiveness of creation and consumption features alone.

We consider the following creation features: the container label used for the story; the *digg age* of the author of the story, where the *digg age* is defined as the time difference between the author's first post and the author's last post; and the submission date. Likewise, we consider the following consumption features: the number of diggs up to the current time, the number of comments up to the current time, the time of the most recent digg, and the time of the most recent comment. As can be seen in Table 2, the consumption features offer greater information gain in all cases. Moreover, the digg features perform significantly better than the comment features, presumably because the ease of "digging" results in a much wider range of digg behaviors and thus provides a better discriminating factor. These features clearly indicate that the initial consumption activity plays a fundamental role in the long-term popularity of the story.

The prominence of the consumption factors makes intuitive sense in light of Digg's public statements about their algorithm. The Digg FAQ[2] states that their promotion algorithm "takes several factors into consideration, including (but not limited to) the number and diversity of diggs, buries, the time the story was submitted and the topic." However, we present these findings simply to demonstrate that content consumption, rather than creation, is a critical driver in the operation of UGC sites.

[2] `http://digg.com/faq`

Table 2: Creation/Consumption Features InfoGain

Feature	InfoGain
Number of Diggs	0.2959
Most Recent Digg	0.2889
Most Recent Comment	0.1412
Number of Comments	0.1326
Submission Time	0.053
Container	0.024
Author Digg Age	0.014

Table 3: Prediction Results

Precision	0.952
Recall	0.865
F-Measure	0.899
ROC Area	0.839

5.2 Prediction Method

With the features selected, we then use them in a C4.5 decision tree [16], largely due to a decision tree's simplicity and comprehensibility. Given the high class imbalance, we undersampled the majority (no comments) class. The results are shown in Table 3. These results are definitely encouraging. We believe by appropriately treating the data for class imbalance using advanced sampling methods and/or appropriate learning algorithms, we can further increase the performance. And while a rigorous mathematical model might also perform quite well on the lifetime prediction problem, this method effectively demonstrates the utility of the process.

These results demonstrate that creation and consumption patterns can be used to construct powerful predictive methods that could lead to practical applications. Commercially, advertisements could be strategically placed based on how long a given submission is predicted to attract attention. Likewise, the sites themselves could use this knowledge to determine where and when submissions ought to be placed on the site.

6 Related Work

Much of the work done in knowledge-sharing oriented OSNs has focused on the formation [5], diffusion [2, 6, 10, 13], and growth [9, 12, 14] of social networks, while neglecting issues pertaining to creation and consumption. One such study by Leskovec et al. [12] focused on local behavior in four online social networks, three of which are knowledge-sharing oriented, to develop a model of network evolution in which nodes select their lifetimes and then proceed to attempt to "close triangles" in the network. Likewise, Adamic and Glance [1] study the extent to which politically conservative and liberal communities in the blogosphere interact with one another, and find that the communities are generally separated with few links between them. This work, and many others like it, study knowledge-sharing oriented social networks, but focus mainly on social networking phenomena that are consequences of creation and consumption.

Moreover, most of the work that is directly related to UGC either focuses on creation patterns alone, or deals only superficially with consumption patterns. Cheng et al. [4] study Youtube and conclude that "related" videos have strong correlations with each other. Leskovec et al. [11] study the diffusion of news across web sites and discover that blogs generally lag mainstream news sites by only a few hours. Guo et al. [8] study UGC creation patterns and find regular temporal patterns and stretched-exponential posting behavior, suggesting that a small set of power users in knowledge-sharing oriented OSNs cannot dominate as they can in a network fitting a power-law. Agarwal et al [3] propose a method for identifying influential contributors to blogs. Certain aspects of the model (number of *in-links* a blog post receives and the number of comments it generates) are directly related to consumption. However, the authors do not study these consumption patterns directly; they merely use them as part of a larger model. Lastly, Guo et al. [7] touch on ideas related to consumption when they examine media access patterns. However, media access is simply the viewing of any form of media available on the web whether it is user generated content or not.

The studies that have examined the consumption of user-generated content focus primarily on characterizing the future consumption patterns of stories based on past consumption. Wu and Huberman [18] model the popularity of stories on `digg.com` and find that the number of diggs N_t that a story receives after time t is modeled by a simple multiplicative process. In particular, N_t increases with the *popularity* of the story at time $t-1$ and the *novelty* of the story (i.e. activity saturates with time). They further extend this work [19] to develop algorithms for maximizing attention to user-generated content.

Hogg and Lerman [17] develop a stochastic modeling framework for user-generated content and use `digg.com` as an example. The Digg model considers two factors: *visibility* and *popularity*. Visibility depends on both position within the list of submitted stories (which decays with time) and the network connections of people who have dugg the story (which increases). The authors show that the model fits the consumption pattern of a few select stories.

We reported results pertaining to homophily and network activation, as well as insights into non-reciprocal relationships, in a previous paper [15]. More specifically, our previous study used Digg.com to focus entirely on the social networking aspects of UGC consumptuon. In this paper, we also use iReport and explain novel findings by offering network statistics, consumption patterns, and a classification framework that uses creation and consumption statistics.

7 Conclusions & Future Work

We studied the *consumption* of user-generated content in online social networks, primarily in the context of the social bookmarking website `digg.com`.

First, we studied the Digg and iReport social networks in order to situate this work relative to the related works. We reported that these networks have small clustering coefficients and node-degree distributions with recognizable "fat tails." We hypothesized that these properties suggest that news flows throughout these networks in such a way that many individuals learn of a story but very few learn of it first-hand.

Second, we studied the creation and consumption patterns and hypothesized that event-flow in knowledge-sharing OSNs can be broken down into intake, processing, and response. We also found unusual behavior in digg and comment distributions, suggesting the activity of Digg bots. Furthermore, the local maximum at about 24 hours in the lifetime distribution suggests a "wave-like" pattern for consumption in which most stories are consumed during peak hours of activity.

Finally, we extracted consumption and creation features from the Digg data and found that consumption features offer the greatest information gain. We then used these features to build a decision tree which demonstrates powerful predictive capabilities.

While the range of potential research topics in UGC consumption is vast, there are several topics which our findings suggest could be of particular interest. For one, since this work showed that consumers generally react to creation, it would be interesting to understand exactly how this occurs. In particular, one may be able to investigate whether the consumer population reacts to increased activity by consuming a large number of the creations or by recognizing and imitating what others are consuming. Another topic that ought to be studied is the anomalous consumption behavior reported above which seems to suggest bot activity. A deeper understanding of this issue could lend to refined methods for bot detection and exclusion. In general, any of the topics examined in this paper could potentially be studied in greater depth and on a larger scale.

Acknowledgments

This work was supported in part by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053, the NSF Grant BCS-0826958, the Arthur J. Schmitt Foundation and the Ateyeh Foundation. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

[1] L. Adamic and N. Glance. The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43. ACM New York, NY, USA, 2005.

[2] E. Adar and L. Adamic. Tracking information epidemics in blogspace. In *Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on*, pages 207–214, 2005.

[3] N. Agarwal, H. Liu, L. Tang, and P. Yu. Identifying the influential bloggers in a community. In *Proceedings of the international conference on Web search and web data mining*, pages 207–218. ACM, 2008.

[4] X. Cheng, C. Dale, and J. Liu. Statistics and social network of youtube videos. In *Quality of Service, 2008. IWQoS 2008. 16th International Workshop on*, pages 229–238, 2008.

[5] R. Gross, A. Acquisti, and H. J. Heinz, III. Information revelation and privacy in online social networks. In *ACM workshop on Privacy in the Electronic Society*, pages 71–80, New York, NY, USA, 2005. ACM.

[6] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *Proceedings of the 13th international conference on World Wide Web*, pages 491–501. ACM New York, NY, USA, 2004.

[7] L. Guo, E. Tan, S. Chen, Z. Xiao, and X. Zhang. The stretched exponential distribution of internet media access patterns. In *Proceedings of the twenty-seventh ACM symposium on Principles of distributed computing*, pages 283–294. ACM New York, NY, USA, 2008.

[8] L. Guo, E. Tan, S. Chen, X. Zhang, and Y. Zhao. Analyzing patterns of user content generation in online social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 369–378. ACM New York, NY, USA, 2009.

[9] R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 611–617. ACM New York, NY, USA, 2006.

[10] J. Leskovec, L. Adamic, and B. Huberman. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1(1):5, 2007.

[11] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the Dynamics of the News Cycle. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 497–506. ACM New York, NY, USA, 2009.

[12] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins. Microscopic evolution of social networks. In *Proceedings of KDD '08*, pages 462–470, New York, NY, USA, 2008. ACM.

[13] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst. Cascading behavior in large blog graphs. In *SIAM International Conference on Data Mining (SDM 2007)*, 2007.

[14] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. Geographic routing in social networks. *Proceedings of the National Academy of Sciences*, 102(33):11623–11628, 2005.

[15] J. Lussier, T. Raeder, and N. Chawla. User Generated Content Consumption and Social Networking in Knowledge-Sharing OSNs. *Advances in Social Computing*, pages 228–237, 2010.

[16] J. Quinlan. *C4. 5: programs for machine learning*. Morgan Kaufmann, 2003.

[17] Tad Hogg and Kristina Lerman. Stochastic Models of User-Contributory Web Sites. In *AAAI Conference on Weblogs and Social Media*, 2007.

[18] F. Wu and B. Huberman. Novelty and collective attention. *Proceedings of the National Academy of Sciences*, 104(45):17599, 2007.

[19] F. Wu and B. Huberman. Popularity, novelty and attention. In *ACM Conference on Electronic Commerce*, 2008.