Sporepedia: A Case Study in the Popularity Dynamics of Social Media

Michael Twardos

October 20, 2009

The increasing amount of user-generated content uploaded to online content services poses a significant problem for content discovery. The ability of online services to display content that maximizes their community's attention ultimately determines their engagement and retention. Harnessing the information collected from viewers through mechanisms such as ratings to compute a "most popular" list is a common solution in determining what content to display. Using the data provided by Sporepedia, the online segment of the PC video game Spore, we study the content that is displayed and the rate at which that content changes. We capture the detailed motion of individual assets in order to assess the freshness or stagnation of the content displayed. This study can be used to determine the effectiveness of Spore's ranking algorithm and as a future point of comparison for other online media and content services.

1 Introduction

The amount of user-generated content available online is unprecedented. Hours of youtube videos are uploaded every minute and millions of twitter status updates are posted everyday. These new forms of media are produced and consumed in drastically different ways than preceding media formats like television and print. The consumption of online content has shifted from a scarcity of information to a scarcity of user attention [2]. On the production side, creators often create large quantities of content with the hope that at least one of their creations will "stick" or have large viewership instead of focusing on a single masterpiece [1]. In many cases, the massive amount of online content is not optimally organized and displayed. Online traffic is often skewed to favor a small number of items, leaving the majority unnoticed [3]. Depending on the way content is displayed strongly affects the consumption behavior of individuals [4, 5] and determines the community's engagement. Therefore, the ability to identify the content most successful in capturing the attention of the audience takes on great importance and ultimately determines the success of the content distribution service.

One way to organize the display of content is through a list sorted by a quality metric, i.e. a "popularity ranking", that describes the relative significance of a given item among its competitors. In the entertainment industries, consumers use ranked lists to determine the best content while market analysts use them to discover the characteristics for successful products. The popularity of an item is determined by competing forces such as its past popularity and its current novelty. In a true "winner take all society" [6], new items are not able to gain popularity as the most popular items only grow more popular over time. In other cases, items that naturally receive the most attention are the newest items, i.e. attention is a function of novelty. Understanding the relationship between the rate at which popularity and novelty evolve can allow content services to decide what to present to users to maximize their attention [7]. Determining the movement of items in rank can shed light on the effects from these competing forces.

We investigate the motion of ranked items by looking at the most popular user generated content in the pc video game Spore. This content consists of three dimensional objects that users can build using the "creation editors" available in the game. Users are able to make creatures as shown in Figure 1, as well as buildings and vehicles of high levels of complexity and personalization . After creation, the objects are procedurally animated and texturized by the Spore game engine and can be used in gameplay.

Figure 1. A few examples of creatures made with the Spore Creature Creator. The high level of creativity and diversity of the Spore community is easy to notice.

Creations are shared with the Spore community via upload to Spore servers and displayed on the Spore website in a segment called “Sporepedia”. Over the last year, approximately 200,000 creations are uploaded on a daily basis, about 70% of which are creatures, 12% vehicles, 12% buildings and 6% ufos. This has amounted to over 100 million pieces of content that were uploaded in the last year. Users can explore this content on the website and sort the it via filters such as “most popular”, “most popular new”, “newest” and “by creator”. Users can interact with this data by commenting, tagging and rating (thumbs up or thumbs down) and “following” their creators. The ratings of an asset provided by users as well as the age of the asset are used to algorithmically rank the assets.

Figure 2. Spore's online content viewer, Sporepedia. Items are sortable by popularity, novelty and per creator as well as filterable by content type (creature, building, etc.) 20 Items are displayed per page and the user is able to "page back" to look at more content.

In addition to Sporepedia, Spore also provides data about user created content via a public api (www.sporeapi.com). In this study, we use the Spore API to study the dynamics of the most popular Spore content over time as generated by the Spore popularity algorithm. The data used in this study consists of the top 500 ranked assets of all items from the time range of June 1, 2009 to October 1, 2009. We characterize the changes in ranking of the most popular items and quantify the motion of rank reordering over time. How easy are new items able to "penetrate" the list and how "viscous" is the motion of unpopular content back down the ranks? To what extent is this movement optimal in sharing new content and displaying items of highest quality? These are the questions we intend to address in this study.

2 Rank Dynamics

Figure 3 shows the “rank trajectories” of a few assets by plotting their rank as a function of time from mid July to the end of September. The rise in rank is often characterized by a large jump (from above 500) as opposed to a slow steady rise. This increase in popularity is often followed by a time period of steady or slowly decreasing rank values and then often a large jump back above 500 occurs. The items shown in this plot persist in the top 500 for a time scale of months. Surely, other content may stay in top ranks for much longer or shorter time, but a general trend of the rising then falling of a content’s rank is observable in the time scale of weeks.

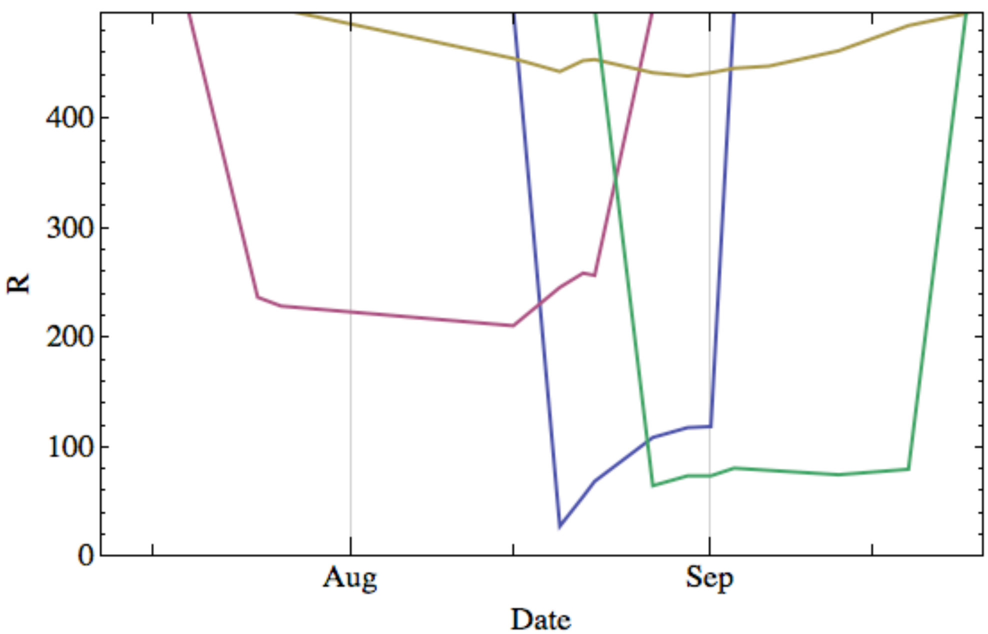

Figure 3. “Rank trajectories” of Spore content. The blue line is the "Abyssal Leviathan" (id 500427484569) created by “TheGuy185” on 8/16/2009. Red is "Acanthopsis Choirorhynchevolu" (id 500213971320) created by Spooradit on 12/19/2008. Gold is "A gift for PandaKing" (id 500278704455) made by BellaBlim on 2/17/2009. Green is "Ancient Char" (id 500434170404) made by Kemeiki on 8/24/2009.

Next, we ask the question: For a given ranking, R, what will its most likely ranking be tomorrow? Given that we can track individual assets through the ranked list, we are able to determine the distribution of the rate of change of rank, Delta R (in the units of change of rank per day), given that an asset is at a level R. A “phase space plot” of change in rank, Delta R, vs rank, R, of all the motions recorded for all assets is shown in Figure 4. This data can be interpreted as a transition probability to another state for each R. Note that these measurements of Delta R only use data from the highest 500 levels of ranking. Any movement of content that is on or off the list of the top 500 (suddenly appears or disappears from the list) is not considered in this measurement because we do not know where they came from or where they went. Surely there are large swings of popular items that move from very low (unknown) popular rankings to the top 500 in a short amount of time, but those motions are beyond the course of the study.

In any case, data points above Delta R = 0 indicate that an asset's next move was up in rank, while points below imply a decrease in ranking. The asymmetry in the graph can clearly be seen: assets are more likely to move in larger jumps down the ranks and make small incremental steps up the list. For low levels of R (high rank) motion appears smaller than for more moderate levels of R. Visualizing the relationship between R and Delta R in Spore content paints an overall picture of the "fluidity" of the ranking system.

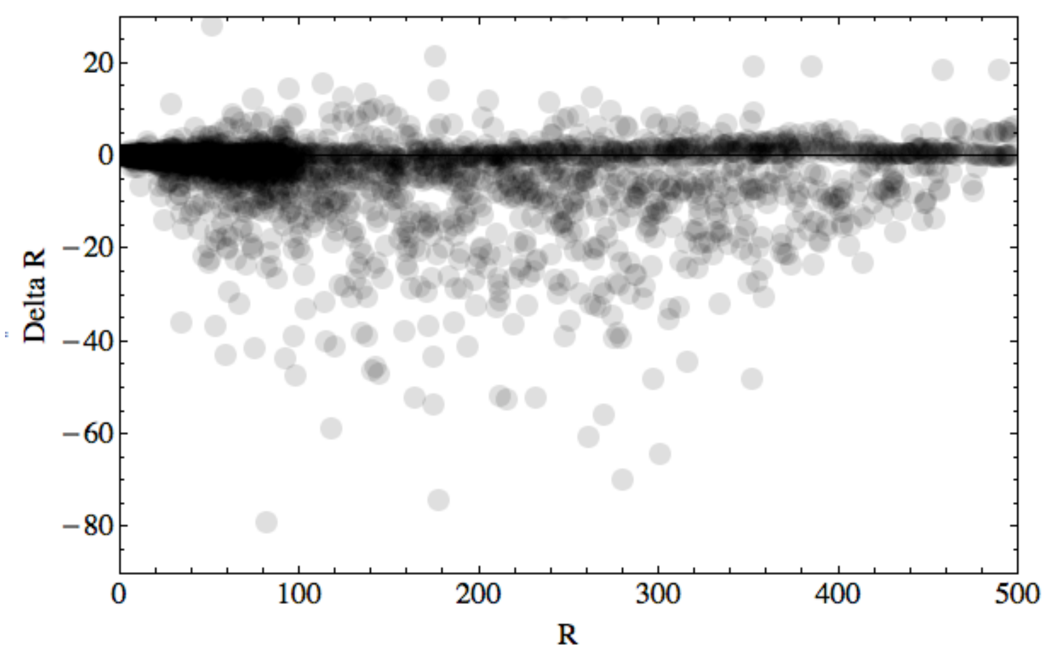

Figure 4. A "phase space plot" of change in rank, Delta R, vs rank, R, of all the motions recorded for all assets over the time period studied. A significant portion of the points are below Delta R = 0 indicating large drops in ranking of items.

The following plot shows the average change in rank, <Delta R>, as a function of R. For most of the plot the average change in rank is negative. Starting at R=1, we see the average change in rank is very small and approaches 0. The steady ranking of Charles Darwin at #1 since the launch of Spore in summer of 2008 attests to this measurement. This implies that the algorithm used to rank Spore assets has a difficult time making the most popular content change. This is most likely a consequence of the top user contributed ratings of the asset providing too much influence to their perceived popularity. <Delta R> reaches a minimum at about 200 and then decreases, crossing 0 and then rising to a positive value before cutting off at 500. The width and height of this inverted hump can serve as a metric to characterize the overall "slipperiness" of the ranking algorithm that "pushes" assets down the list.

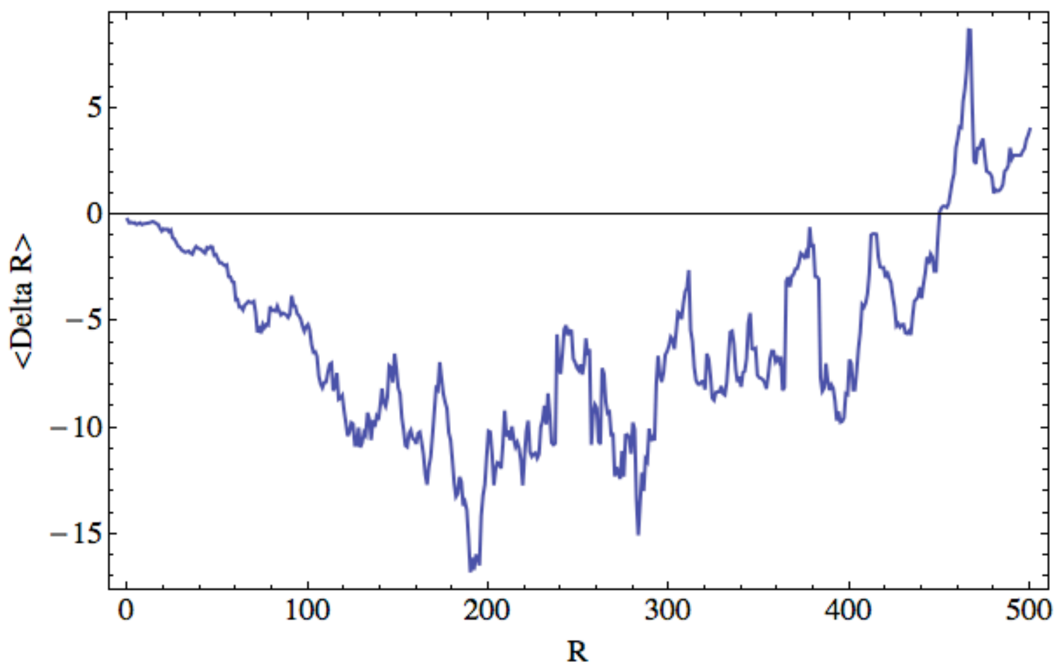

Figure 5. The average change in rank, <Delta R>, as a function of R. Assets are more likely to move down in rank and the effect is maximal at about R = 200. For values above R=450, <Delta R> shifts to positive values, indicating the general flow of content.

Next, we compare the average movement up the ranking system given that an item moves up, Delta R+, to the average movement down given that it moves down, Delta R-. The results are shown in Figure 6. For assets moving up in rank, they average about Delta R+ = 3 for the entire span of R. Delta R+ drops to zero as R=0 is approached which clarifies the concept of a of a "sticky" or viscous top rank area. On the other hand, Delta R- is much more pronounced than Delta R+: near R=0, Delta R- dips below -10 and then nears -5 for larger values of R. This dip of Delta R- shows another perspective of the slipperiness away from the direction of top ranking. Overall, we can conclude from this chart that there appears to be different mechanisms for assets growing and decaying in popularity. While increases in popularity are slow and more regular across the range, decreases in popularity are larger and vary more as a function of R.

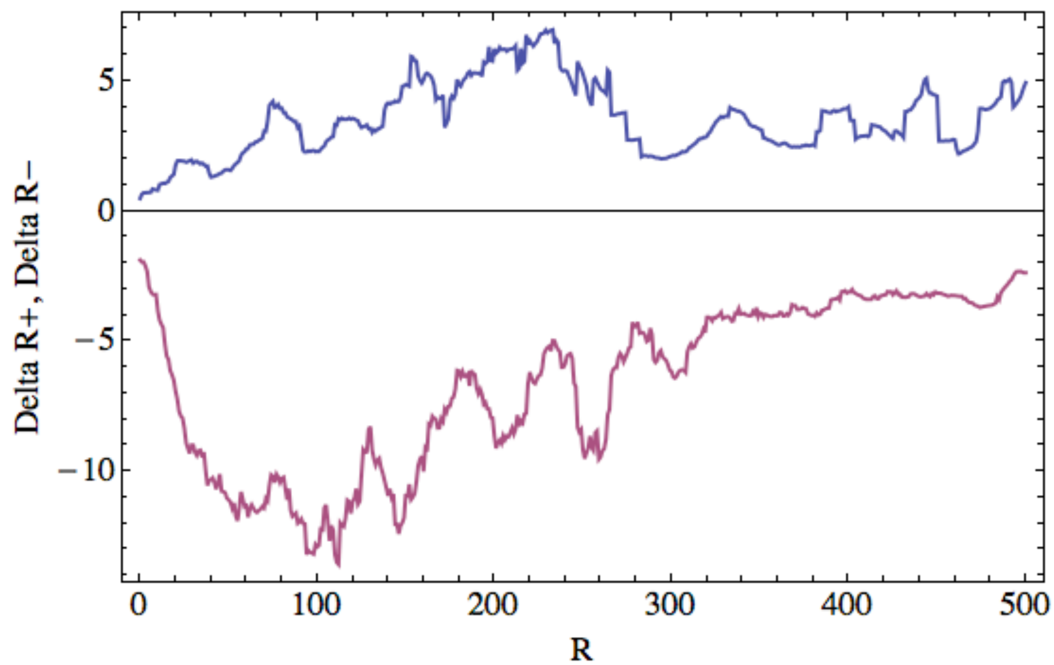

Figure 6. The average change in rank, <Delta R+>, and <Delta R-> as a function of R.

Another way to visualize the motion of ranked content is to look at the fraction of items that moved up (compared to down) as a function of the rank. This is shown in the following figure. Since values under 0.5 imply more movement down than up, the top ranked assets are most likely to move down to lower ranking value while lower ranked assets are likely to move up.. Surely the dynamics are more complex, but this implies a general trading of higher ranked assets for lower ones. For rank values between 100 and 300, the ratio hovers close to 0.5 implying equal probability of upward and downward movement. The slope of this line implies the degree to "mixing" occurs in this top ranked group.

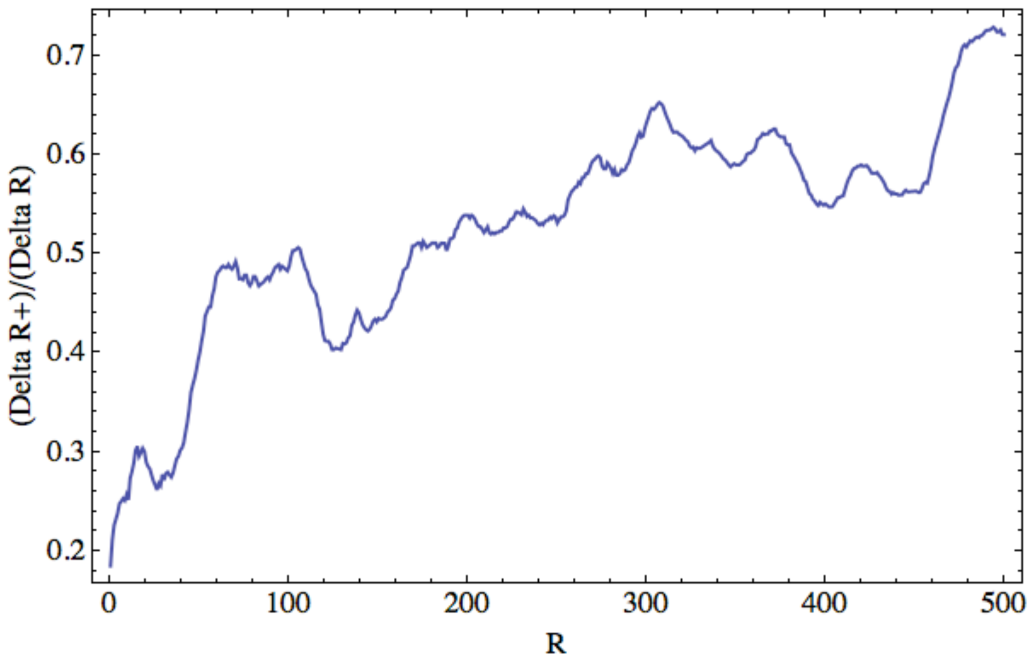

Figure 7. The probability an item moves up in rank as a function of its rank. For the highest ranks, the probability is below 0.5, meaning it is more likely to move to lower ranking. At about R= 250, the probability passes 0.5 making movement to higher levels of rank more probable.

Common throughout this discussion is the concept of how difficult is it for new items to reach high levels of ranking, i.e. penetration. In the following figure, we show the probability that an asset reaches a given maximal rank, R_max. We can conclude from this plot that for all the items that make it into the top 500, only half of them make it into the top 400 and half of them make it to the top 300, etc. What is most striking about this plot is the downward concavity of this plot's slope. A trend that is linear as a function of R would imply direct proportionality of the probability of a top rank value. It will be interesting to compare this trend with the ability for other types of media to penetrate their ranked list

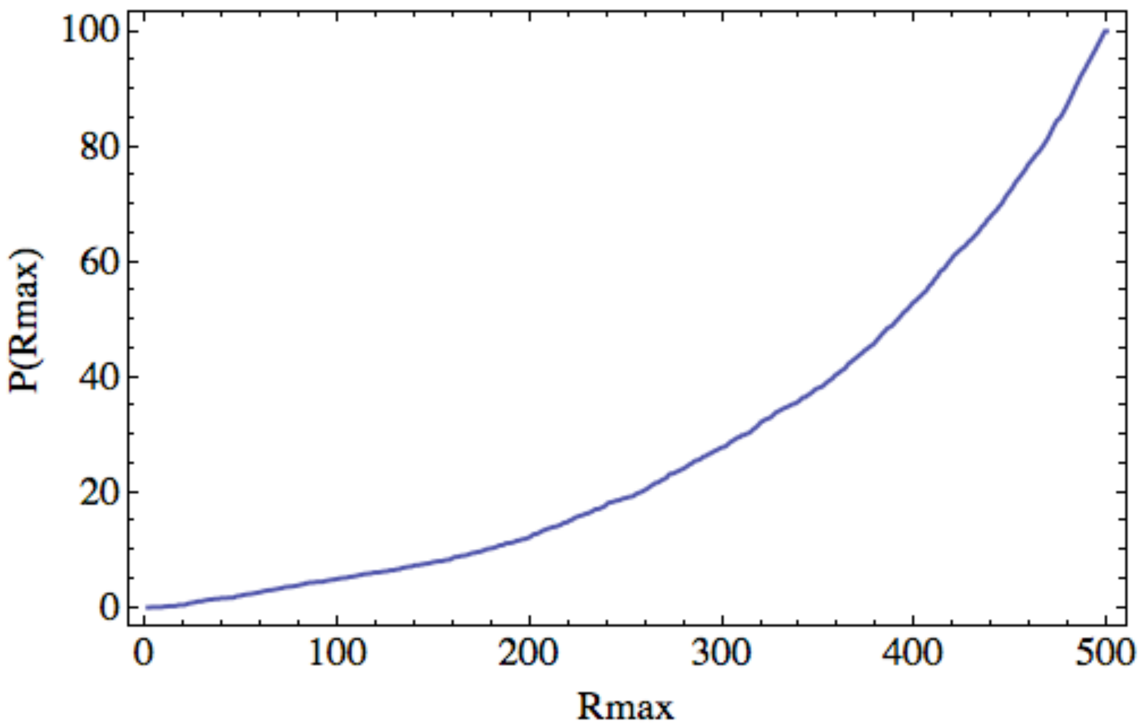

Figure 8. The probability an item maximum rank is P(R_max) given that the item reaches the top 500. There is about a 10% chance that the item reaches the top 200 items and about a 5% chance it reaches the top 100.

3 Conclusion

In this study, we presented key metrics to describe the change in rank over time of user generated content in the video game Spore. We have identified several observables of the change in rank over time including the expected change in rank, the expected change in rank up Delta R+ and change in rank down Delta R- and the probability that an item moves up or down the list. We used these variables to make temporal visualizations of the rank motion of spore assets.

Several interesting features in Sporepedia's asset ranking have been observed. First, the "trajectories" of assets were characterized by an increase and then decrease in ranking observable over a time frame of a couple of months for most content. Second, we observed the asymmetry in rank motion as assets are more likely to make big movements down and small movements up the ranked list indicated in the phase plot of Delta R vs. R. Third, the diminished motion of items at high levels of ranking (close to R = 0) implies a “viscosity” as a function of rank. Fourth, we observed the “inverted hump” of ranking, identified first in the Delta R vs. R plot that showed a maximal negative average change in ranking that increased to 0 for both higher and lower ranking levels. This hump was seen clearly in the plot filtering for only downward movements Delta R- vs. R. Finally, we observed the concave probability distribution of P(R_max) as function R indicating an disproportionate difficulty in reaching a higher rank.

For any content discovery system, the rankings intend to properly define the balance between quality and new content to produce the optimal experience for its viewing audience. Engineering the placement of items in a ranked list attempts to maximize a user’s attention by displaying the most attention grabbing content. Studies like this can be used to draw conclusions about the effectiveness of content display algorithms in reaching this goal. In addition, this study poses many new questions. How do other entertainment media that use consumers to filter their content compare? What are the optimal rank dynamics for a satisfied user base and an engaged community? How does the filtering, organizational and navigation tools create the best experience for content discovery? How do the sizes of both the producers and consumers effect the efficiency of the ranking process? What effect does a large disproportion between producers and consumers have? Given the motion of ranked assets described here, what advertising campaigns are most effective at popularizing new content? What effect does the past success of a user’s creations have on their future popularity? Future studies are encouraged to address these questions.

References

[1] F. Wu, B, Huberman, *Persistence and Success in the Attention Economy* (2009)

[2] B. Huberman, F. Wu, *The Economics of Attention: Maximizing User Value in Information Rich Environments*, Advances in Complex Systems, Vol 11, No. 4 (2008) 487-496

[3] C.Y. Lee, S. Kim, *Stochastic Dynamics of the Popularity of Growing Websites*, Journal of the Korean Physical Society, Vol. 44, No. 2, February 2004, pp. 442

[4] Symeon Papadopoulos, Athena Vakali, and Ioannis Kompatsiaris, *The Dynamics of Content Popularity in Social Media*, International Journal of Data Warehousing and Mining (accepted for publication, December 2008).

[5] Bernardo A. Huberman, Peter L. T. Pirolli, James E. Pitkow, and Ra jan M. Lukose. *Strong regularities in World Wide Web surfing*. Science, vol. 280, no. 5360, pp. 95–97, 1998.

[6] R. Frank and P. Cook, *The Winner-take-all Society* (Free Press, New York, 1995)

[7] F. Wu , B. Huberman, *Popularity, Novelty and Attention* (2008)