Social Attention in the Age of the Web

Bernardo A. Huberman

The Internet is slowly but irreversibly changing ideas we've had for centuries about libraries as unique repositories of knowledge. What started as a digital medium for transmitting data and computer programs soon morphed into the World Wide Web, which in about a decade transformed forever the way people think of information and the ways in which they access it. We no longer need to physically enter a library to obtain the latest news or to read a scholarly journal. A simple search through any computer or mobile device having a browser and a keyboard is enough to put at our disposal not only what we search for but also a trove of related findings that increase our curiosity and expand our horizons. Add to that the ubiquity of e-mail and instant messaging, and we find ourselves in a world of instant connectivity and potentially productive connections with social networks across the globe.

What are we to make of such a world? To start with, instant and free access to information across geographic and institutional boundaries has made its value plummet in an economic sense. We value what is scarce, not what is plentiful, and the precious entity is now *attention*, which is always finite and claimed by many sources at the same time. The Web has made possible the creation and display of content that, it is hoped, multitudes will attend to. Thus the keen competition for people's attention, manifested through e-mails, blogs, and manuscripts that keep appearing on our screens.

The kind of attention that I have in mind is social in nature, and while recognizing that the perceptual component of individual attention is central to the whole process, I will focus on the intensity (i.e., the number of visits, links, and citations) of signals pointing to a given idea, result, paper, Web site, etc. This in turn brings into focus the role that social networks of all kinds play in the amount of attention allocated to topics of interest to a discussion group.

Attention is so important in the world of academia[1] that I'd venture to state that it is often its main currency: we publish to get the attention of others, we cite so that other researchers' work gets attention, and we cherish the prominence of great work if only because of the attention it gathers. This phenomenon has been taking place since the establishment of learned societies and academic disciplines, but it has not been analyzed systematically until recently. Recent work (Goldhaber 1998; Franck 1999; Klamer and Van Dalen 2002; Falkinger 2007; Huberman and Wu 2008; Wu and Huberman 2007) is starting to frame this problem in the context of the new digital medium while providing insights on the role that attention plays both in the Web and in electronic publishing. Richard Lanham (2006) has eloquently described the significant role that the arts and letters play in this attention economy by creating attention structures that often trump style over content.

A recent study we performed at HP Labs provides a stark example of how attention drives content creation outside the academy or enterprise (Huberman, Romero, and Wu 2008). Analysis of a massive YouTube data set revealed that the productivity of those uploading videos strongly depends on attention, as measured by the number of downloads. Conversely, a lack of attention leads to a decrease in the number of videos uploaded and a consequent drop in productivity, which in many cases asymptotes to no uploads whatsoever. Most of the YouTube content shares this fate, as the consumption of uploaded content is highly skewed. Whereas most videos are never downloaded, a few get a disproportionate share of the audience's attention.

This result offers insight into how the tragedy of the digital commons is partially averted. Basically, contributors treat the digital commons as a private good, in which payment for their efforts is in the form of the attention that their content gathers. The result is a massive contribution to the public good.

The relevant question is how attention allocation, and its impact on research, is transformed by the advent of digital media and the consequent flood of information that streams into our senses, as well as by the new modalities exemplified by Wikipedia and Web-based social networks that span the globe.

For academics, the panacea offered by the Web is tempered by the fact that even the best Web sites are at times insufficient to provide the information they seek while filtering out unnecessary content. While some sites decide what to display on the basis of an objective criterion (e.g., novelty of a paper or image, page rank in search, popularity of a topic, or the salience of news), they do not necessarily maximize the user's value. For example, an algorithm such as Google's page rank inserts the most linked-to pages in the first page of a query result, but other links in other pages often con-

---

[1] That attention is a valued resource in general, and that people are willing to forsake financial gain to obtain it has been empirically demonstrated by B. A. Huberman, C. Loch, and A. Onculer. 2004. Status as a Valued Resource. *Social Psychology Quarterly* 67(1): 103–114.

tain incipiently valuable information that is not available to the user just because they are buried further down the list.

Another problem stems from the finite number of items that a user can attend to in a given time interval. This psychological constraint is compounded by a strong empirical regularity observed in Web browsing that goes under the name of "the law of surfing" (Huberman et al. 1998; Huberman 2002). This law states that the probability of a user accessing a number of Web pages in a single session markedly decays with the number of pages, thereby constraining the amount of information that ever gets explored in a single surfing session. A typical user seldom visits pages beyond the first one displaying search results; consequently, a page ranked near the bottom by a search engine is unlikely to be viewed by many users. This behavior tends to reinforce the leading position of those top items and to further increase their popularity, which in turn penalizes content that is not yet well known. Thus, it is easy for an item to get locked in a top ranking and hard for other bottom items to surface, even though the latter can often be more valuable.

In spite of all these obstacles, we somehow manage to remain up-to-date with our work; once in a while, we even discover interesting facts and ideas that are relevant to our intellectual endeavors. We often accomplish this through a social network of like-minded academics, colleagues, and friends who quickly propagate novel ideas and their opinions about them. These networks, sometimes called "informal colleges" or "communities of practice," were identified a long time ago as important channels for the dissemination and validation of new results in a given discipline (Crane 1972; Crozier 1964). The advent of the Web has increased the scope and swiftness of these networks by several orders of magnitude.

Social networks are not restricted to academia. Any infrastructure that provides opportunities for communication among its members is eventually threaded by communities of people who have similar goals and a shared understanding of their activities. These informal networks coexist with the formal structure of any organization and serve many purposes, such as deciding on the relative worth of given results (and at times the reputations of the authors of these results), solving problems more efficiently (Feld 1981), and furthering the interests of their members. Despite their lack of official recognition, informal networks can provide effective ways of learning and actually enhance the productivity of the formal organization.

In the digital domain, we are witnessing a proliferation of social networks, such as Facebook, Myspace, Linkedin, and Hnet, that connect very large and geographically extended social groups while providing them with a sense of immediacy that fosters the exchange of information on myriad topics and types of media.

This new social trend has sparked a keen interest in identifying online communities. Some of this work finds that online relationships do indeed reflect actual social relationships, thus adding to the "social capital" of a community. Mailing lists and personal Web pages also serve as proxies for social relationships, and the commu-

nities identified from these online proxies resemble the actual social communities of the represented individuals.

Research on the role of social networks in the dissemination of ideas, purchases, and reputations is also accelerating because of the ease with which data can be gathered and analyzed on a scale that was impossible using traditional methods (Wasserman and Faust 1994). As an example, the figure below displays the results of an analysis of a network of recommendations responsible for the purchase of books. The study focused on data from Amazon, containing 15 million recommendations of books recommended to more than 5 million people who purchased them (Leskovek, Adamic, and Huberman 2007). By studying the networks that grew up around each book—who bought and recommended it, and who responded to the recommendation—we learned that social networks take on different characteristics depending on the type of books that were recommended. In the figure, red dots and lines indicate people who purchased a product while blue dots and lines represent people who received a recommendation.[2] The network around a medical book (small graph in the upper left-hand corner) shows a scattered network where recommendations, on average, don't travel very far. On the other hand, the network surrounding a Japanese graphic novel, which occupies the central part of the picture, shows a thick flow of information among densely connected groups of people.

---
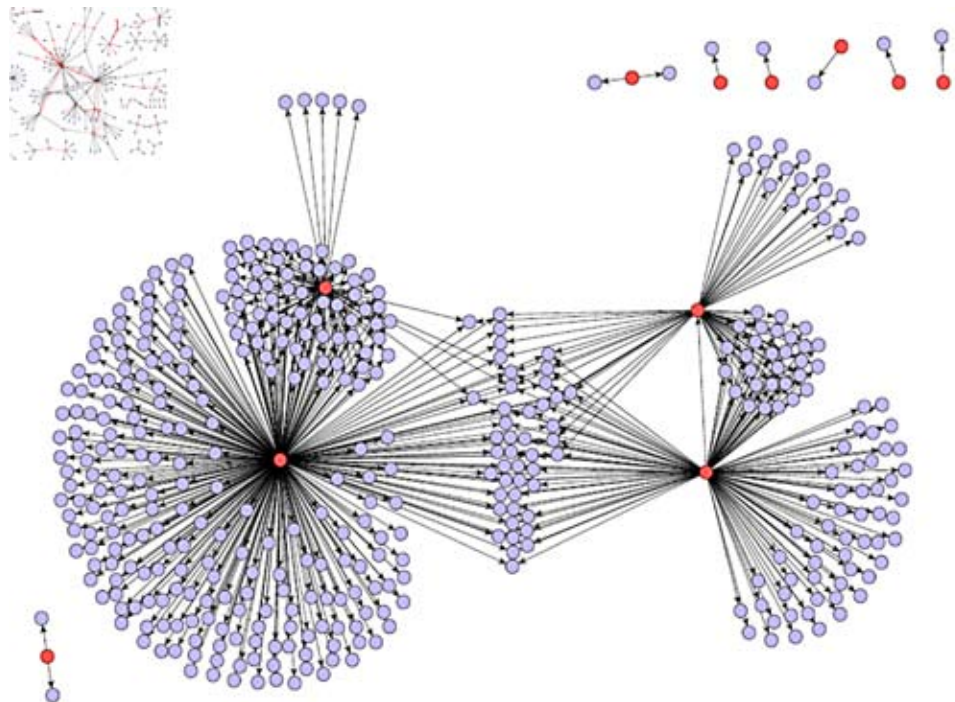
[2] Please see the online version of this publication for a color rendition of this figure above, available at http://www.clir.org/pubs/abstract/pub145abst.html.

The same methodology used to discover the social network underlying the propagation of recommendations may be used for any other kind of information linking people. For example, several years ago we developed a fully automated method for identifying communities of practice within an organization by studying the patterns of e-mail exchanges among its members (Tyler, Wilkinson, and Huberman 2003; Huberman and Adamic 2004). The method uses e-mail data to construct a network of correspondences, and then discovers the communities by partitioning this network in a particular way, as described below. The only pieces of information used from each e-mail were the names of the sender and receiver (i.e., the "To:" and "From:" fields), enabling the processing of a large number of e-mails while minimizing privacy concerns.

Using this method and a standard desktop PC, we were able to identify small communities within a globally distributed organization in a matter of hours. Interviews validated the results obtained by our automated process and provided interesting perspectives on the communities identified. Other approaches have used coauthorship of papers to identify social networks (Kempe, Kleinberg, and Tardos 2003), which can also be useful if one is interested in tracking the evolution of cooperation within disciplines. And since social structure affects the flow of information, knowledge of the communities that exist within a network can also be used for navigating the networks when searching for individuals or resources (Huberman and Adamic 2004; Kempe, Kleinberg, and Tardos 2003).

An important aspect of social networks is how they direct attention to given topics or results while ignoring others. Collective attention is at the heart of the spread of ideas and the reputations of people, and it has been studied at the individual and small-group levels by a number of psychologists, economists, and researchers in the area of marketing and advertising. Attention also affects the propagation of information in social networks, determining the effectiveness of advertising and viral marketing. While progress on this problem has been made in small laboratory studies and in the theoretical literature of attention economics, it is only recently that we have obtained empirical results from very large groups in a natural, nonlaboratory setting (Wu and Huberman 2007).

To understand how social networks mediate the allocation of attention, consider how a news story spreads. When it first comes out, the story catches the attention of a few, who may pass it on to others in their social network if they find it interesting enough. If a lot of people start to pay attention to this story, its exposure in the media will continue to increase. In other words, a positive reinforcement effect sets in: the more popular the story becomes, the faster it spreads. This growth is counterbalanced by the fact that the novelty of a story tends to fade with time and that people therefore pay less attention to it.

Thus, with respect to the dynamics of collective attention, two competing effects are present: (1) the growth in the number of people that attend to a given story; and (2) the habituation that makes the

same story less likely to be attractive as time goes on. This process becomes more complex when multiple items or stories appear at the same time and people must decide which stories to attend to. However simplistic this description might be, it allows for the construction of a mathematical model that predicts how attention is allocated among many items, links, and other factors, and how those items become less novel over time (Wu and Huberman 2007).

The predictions of this theory, which were empirically tested with a million users of a popular news site (digg.com) are as follows: (1) the distribution of attention among a set of items is log-normally distributed, i.e., it is highly skewed in such a way that most stories get a typical small number of "diggs" (as a measure of the attention they receive), whereas a few receive a lot of attention (a winner-take-all scenario); and (2) collective attention decays slowly, specifically in the form of a stretched exponential function of time.
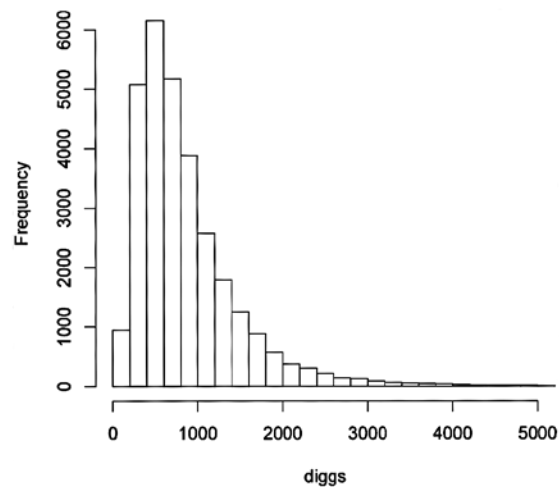
The figure above, which shows the distribution of attention over all stories in digg.com, clearly displays the skewed behavior just described. This distribution, with its long tail, provides another plausible explanation to the question of why the large majority of articles in the sciences receive so little attention whereas a small percentage (i.e., those in the tail of the distribution) make the grade in terms of a large number of citations.

But this is still not the complete story. Other drivers can be as effective as novelty in eliciting social attention. One is popularity, which often leads us to read and examine ideas if only because others do. Another is style, as is the case when visual elements make an idea or presentation initially compelling because of its elegance or esthetic value. Much research is needed to elucidate all these aspects, and we are currently examining some others as well, such as the role of attention in opinion formation on the Web and its role in the productivity of individuals.

In conclusion, I hope to have shown that in the age of the Web, social attention and its swift allocation through vast social networks

plays a central role in the dissemination and validation of ideas and results within the academic community. Two very successful examples bracket my statement. Wikipedia has already shown the power of an interactive medium in creating a vast landscape of knowledge, even when the threshold for contributions is negligible and authorship remains anonymous. At the other extreme, many practitioners of a highly technical branch of the hard sciences, superstring theory in particle physics, have opted out of the traditional publication venues and chosen to exchange their manuscripts through an electronic preprint repository (arxiv.org) without going through standard refereeing procedures. In both cases, the intense chatter of these worldwide communities brings attention to relevant results and serves as a good quality filter.

And given that this essay is about attention and that I'm keenly aware of its fleeting nature, I think that it would be unwise to continue writing beyond this point.

References

Crane, D. 1972. *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*. Chicago: University of Chicago Press.

Crozier, M. 1964. *The Bureaucratic Phenomenon*. Chicago: University of Chicago Press.

Falkinger, J. 2007. Attention Economies. *Journal of Economic Theory* 133: 266–294.

Feld, S. L. 1981. The Focused Organization of Social Ties. *American Journal of Sociology* 86: 1015-1035.

Franck G. 1999. Science Communication, a Vanity Fair. *Science* 286: 53–55.

Goldhaber, M. H. 1998. Attention Economics and the Net. *First Monday* 2.

Huberman, B. A. and F. Wu. 2008. The Economics of Attention: Maximizing User Value in Information-Rich Environments. *Advances in Complex Systems* 11(4): 487-496.

Huberman, B. A., D. M. Romero, and F. Wu. 2008. Crowdsourcing, Attention and Productivity, *arXiv:0809.3030v1* (2008).

Huberman, B.A., and L. Adamic. 2004. Information dynamics in the networked world. In *Complex Networks*, edited by E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai, eds., 371-395. Springer.

Huberman, B. A. 2002. *The Laws of the Web: Patterns in the Ecology of Information*. Cambridge, Mass.: MIT Press.

Huberman, B. A., P. L. T. Pirolli, J. E. Pitkow, and R. M. Lukose. 1998. Strong Regularities in World Wide Web Surfing. *Science* 280 (5360): 95–97.

Kempe, D., J. Kleinberg, and E. Tardos. 2003. Maximizing the Spread of Influence through a Social Network. *Proceedings of the Ninth ACM International Conference on Knowledge Discovery and Data Mining*, Washington, D. C., August 24-27, 2003. Klamer, A., and H. P. Van Dalen. 2002. Attention and the Art of Scientific Publishing. *Journal of Economic Methodology* 9: 289–315.

Lanham, Richard. 2006. *The Economics of Attention: Style and Substance in the Age of Information.* Chicago: University of Chicago Press.

Leskovek, J., L. Adamic, and B. A. Huberman. 2007. The Dynamics of Viral Marketing. *ACM Transactions on the Web* 1(1): 5.

Tyler, J. R., D. M. Wilkinson, and B. A. Huberman. 2003. Email as Spectroscopy: Automated Discovery of Community Structure within Organizations. In *Proceedings of the International Conference on Communities and Technologies*, edited by M. H. Huysman, E. Wenger, and V. Wulf, 81-96. Springer.

Wasserman. S., and K. Faust. 1994. *Social Network Analysis*. Cambridge and New York: Cambridge University Press.

Wu, Fang, and Bernardo A. Huberman. 2007. Novelty and Collective Attention. *Proceedings of National Academy of Sciences* (USA) 104: 17599–17601.