Revista da UIIPS – Unidade de Investigação do Instituto Politécnico de Santarém,
Vol. 5, N. º 2, 2017, pp. 117-124
ISBN: 2182-9608
http://ojs.ipsantarem.pt/index.php/REVUIIPS

ON THE EVOLUTION OF HYPERLINKING

Artur Marques

Escola Superior de Gestão e Tecnologia, Instituto Politécnico de Santarém, Portugal

artur.marques@esg.ipsantarem.pt

ABSTRACT

Across time, the hyperlink object has supported different applications and studies. This is one perspective on the evolution of the hyperlinking concept, its context and related behaviors. Through a spectrum of hyperlinking applications and practices, the article contrasts the status quo with its related, broader, conceptual roots; it also bridges to some theorized and prototyped hyperlink variations, namely "stigmergic hyperlinks", to make the case that the ubiquitousness of some objects and certain usage patterns can obfuscate opportunities to (re)think them. In trying to contribute an answer to "what has the common hyperlink (such an apparently simple object) done to society, and what has society done to it?", the article identifies situations that have become so embedded in the daily routine, that it is now hard to think of hyperlinking alternatives.

Keywords: hyperlink, hyperlinking, search, stigmergic hyperlink, www

1. INTRODUCTION

Vannevar Bush was an electrical engineer, appointed MIT's Vice President in 1932, and then head of the OSRD (U.S. Office of Scientific Research and Development) during World War II. OSRD was the office which supported scientific and military projects such as the atomic bomb / "Manhattan Project" and the radar. In 1945 Bush was already asking what significant problems would researchers face after the war, and answering the question with the challenge of how to promote cross-awareness of knowledge and intellectual production in different scientific areas, because the then existing "methods of transmitting and reviewing the results of research" were, in his perspective, "generations old" and "totally inadequate" (Bush, 1945, 1996). Bush considered that conventional indexes/citations embody such a distance to the originals/cited works that they are creativity-unfriendly. Since the mind "operates by association", a natural solution would be to create links between recorded ideas in context.

Bush's plan of a "memex" machine and its ability to link any points between microfilmed documents is the hyperlink precursor which inspired Theodore Nelson, Douglas Engelbart, and others, even if using a much different memory support and an association representation that most current users couldn't begin to imagine because so many of us take the contemporary hyperlink for granted, as a dead simple interface to something that sometimes is even no longer perceived as a problem. Still the original association challenge persists and the current WWW hyperlink usage is happening on a massive and industrialized scale from which emerge hyperlinking motivations and hyperlinking use cases that strongly deviate from plain citation and navigation, let alone knowledge association.

The common hyperlink is not a one-size-fits-all device, has limitations, can disappoint even its creators in certain ways and, in certain scenarios, can fall victim to its own success.

One challenge in researching, extending, changing or just thinking about the hyperlink is how deceptively simple "it" appears to be. However, when looking past the surface layer, one should gradually be able to sense its depth and great importance in today's world, as envisioned by scientists such as Vannevar Bush from the mid-1940s and now voiced in publications as generalist as the New York Times Magazine, all suggesting the concept and the object as one of all times' greatest inventions.

Besides its everyday star role on the Web, its place in computer science in general and hypertext in particular, the hyperlink as an entity and as an interface mechanism doesn't rest confined to such fields. Hyperlinks have become a pervasive tool that "changed the fundamental dynamics of human communication" (Hespos, 2008) be it in information sciences, political sciences, sociology or media studies, etc., usually exposing the same familiar attributes and functionality, which makes them easy to use but also somewhat conditioning because of their impregnated usage pattern that makes it hard to think of them differently, "out of the box", in particular when required to do so without disruptive consequences for the users. "Hyperlinks are more than technical artefacts" (Maeyer, 2013) supporting a gamut of situations in a way which could eventually be different if other attributes and behaviors were also available.

Some hyperlinking behaviors may now cause frustration - e.g. broken links and invasive ads publishing - or even social damage, as in phishing attacks. Solutions such as "Stigmergic hyperlinks" (Marques & Figueiredo, 2010), or "stighs", can be of immediate help in providing relief to some of the "bad" outcomes of the common hyperlink: in particular, they constitute a fully automatic response to broken links in pages. While many solutions exist for identifying broken hyperlinks and other structural issues (Haslhofer & Popitsch, 2009; Ingham, Caughey, & Little, 1997; Kovilakath & Kumar, 2012; Morishima, Nakamizo, Iida, Sugimoto, & Kitagawa, 2009), very few exist for helping in fixing the problem. The relatively scarce responses to identify and solve the broken links usually follow a recommender system approach, which depends on users to rank alternative substitute resources, before committing to a fix. Such approach is applicable to all situations where indeed there is a replacement to be provided and human labor available to be invested. The fully autonomous solutions, which identify and solve broken references without human assistance, are much rarer and limited to specific domains.

However, it is important to understand that such solutions are not intended as replacements of the common hyperlink! They are only adequate for authors with particular needs, e.g. wanting to have "publish and forget" autonomous systems, wanting to have web metrics at the hyperlink-level or wanting to provide different contextual "searches". The hyperlinking behavior here discussed is mostly based on the common hyperlink.

The hyperlink's ubiquitousness would only come with HTML and the WWW. Such is the penetration of the hyperlink as structured in HTML that it became hard to conceive it in a different dress, with other attributes and other interactions. Despite everyday problems - e.g. how easy it is to "phish", the dependency on browsers for travelling back to sources, the unverified destination pointer and the dead link disruption (Leuf, 2006) - the dominant hyperlink solution remains mostly unchanged and unchallenged, for reasons that include its simplicity.

At a lower level, the everyday hyperlink does serve for citation and navigation, but on a higher level those purposes are being leveraged for serving other intentions not immediately inferable from the root design, namely "gatekeeping" sites, "computing" contents and "measuring" or "tracking" users. These functions are byproducts of the underlying business, technology and marketing forces. The hyperlink is literally the entity, by which some of the world's biggest organizations live or die, from search-engine based businesses to the contents industry, to the entertainment world in general: failing to link or to be linked, regardless of other core merits, can push one into dire straits.

This paper first delves into some motivations for hyperlinking, across time; then it approaches the subject of "Web Metrics" for the understanding of the hyperlinking behavior. It concludes with one perspective on the consequences of the current hyperlinking practices.

2. MOTIVATIONS FOR HYPERLINKING

Why do authors create hyperlinks? What are their possible motivations?

This is an active research question, with partial answers that vary across the timeline: the main reasons for hyperlinking in the mid-1990s may be different from the ones during the blogs' boom of the mid-2000s, and are probably distinct from what drives authors and publishers today.

In Ciszek and Fu (2005) the authors share the perspective that hyperlinks "have evolved beyond being citation knowledge tools to become social linking mechanisms". The statement captures that in the early Web many hyperlinks were plain citations-as-references, from a document to other documents or to parts of itself. If hyperlinks are authored with a citation functionality in mind, then some identified motivations (Kim, 2000) are: to guide the reader to background information, to provide an example, to give supportive evidence, to provide a definition, for historical background, or to compare with an alternative. Also implicit in the statement is that other factors, beyond those identified in academic referencing, would determine hyperlink authoring.

Thelwall (2003) argues that "academic hyperlink creation" and "citations in journals" can however be "very different phenomena" from a scholar's professional point-of-view, in the sense that the hyperlinks present in university Web pages can exist for purposes other than assisting the reader, namely connecting to other institutions, employers, partners and personal bookmarks. The author presents a classification which distinguishes "ownership links" (for acknowledging authorship or co-authorship), "social links" (those with a social reinforcement role), "general navigation", and "gratuitous" links (those serving no communication function at all). This taxonomy is limited to links in academic Web pages.

In Miles-Board, Carr and Hall (2002) the authors identify the most frequent "linking practices" in "genuine associative-linked pages", after narrowing an initial sample of 770,992 random Web pages to 576 meeting certain criteria of link density and link distribution, in order to avoid dubious or non-associative content. They conclude that "reference links", which tie nouns such as names of people, institutions, products, places, etc., to corresponding resources are the most frequent, occurring in 84% of the filtered "well-linked pages"; followed by "deep links" which associate keywords or concepts to deeper explanations (37%); followed by "glossary links" which connect alphabetical arranged terms (35%); followed by "citation links" as in endnotes (32%); followed by "structural links" which relate parts of content, for example performing navigation between sections, or pointing to figures (17%). It should be noted that these conclusions are specific to what the authors call the "well-linked pages" and not to all pages in general; if no filters were in place, then "structural links" would probably be the most observed, because while publishers can make the editorial control option to rarely have explicit external links, as some newspapers did (Barnhurst, 2002; Dimitrova, Connolly-Ahern, Williams, Kaid, & Reid, 2003; Tremayne, 2005), there is no escape from navigational structures, if the site is to be usable.

Expectations of reciprocity, as in B expecting a link from A because B links to him, is one emergent social behavior around the hyperlink (L. Adamic, 2008; L. A. Adamic, 2009). "Emergent" because nothing in hyperlink authoring implies such consequence; still there was an époque of "guestbook" proliferation, pages of "favorites ", and "circles" of sites on a common theme, and there are modern equivalents in social networks.

Besides citation, hyperlinks always supported navigation. Navigation became less innocent as the Web grew in its number of consumers and producers. At first, navigation enabled nothing but a guiding structure to "consume" a site's available contents and "teleport" the audience to other cited content providers. Then, the growing number of Web content publishers increased the awareness that visitors might leave and never return, and that one could be left "talking" alone. For authors who were just enjoying the technology and whose drive didn't depend on an audience nothing changed, but for all the others who wanted readers and/or viewers, it became clear that it would be necessary to manage processes of acquisition and retention of visitors.

Hyperlink authoring would become less spontaneous; authors started having second thoughts before publishing external links: what if the destination was badly perceived? Would the negative perception harm the publisher somehow? What if the destination was favorably perceived? Could it mean traffic lost for a rival? What if the external resource is moved or deleted? These questions point that when audience matters, there is value in traffic volume and traffic quality.

Value means money: faster than many established business models could adapt, the Web was hosting old and new businesses. With money at play - a powerful motivator - accounting leapfrogs in relevance. Accounting for the Web means "Web metrics" which can be considered as the measurement of Web properties.

In leveraged applications, the hyperlink can translate to a sale in commerce, to a subscription, to a donation, to an advertisement or to valuable information about visitors. These activities correspond to all the possible revenue streams in any type of business (Krishnamurthy, 2003), thus they constitute hyperlinking with a monetization motivation. Some revenue streams which exist since the dawn of money but could not scale up to support big operations, namely donation-based and freemium models, can now deliver in the WWW-based hyperlinking society.

The monetization possibilities of the hyperlink are the cause of its industrialization, happening in a "marketplace of attention" (Webster, 2008). This is a dichotomized marketplace: as a whole, in providing ever more contents, Web publishers are a force of attention fragmentation; however people have a tendency to socialize and consume with and what they are already familiar with, which is a factor of attention concentration/polarization. Combining the bias for attention polarization with the winner-takes-all form that the public WWW is assuming in a number of situations (in e-commerce, in Web search services, in blogging, etc.) contributes for information "enclaves", pushing in the opposite direction of fragmentation.

The marketplace of attention happens in a context of "convergence", "abundance" and "scarcity". Convergence because all content, new and old, is being made available in the same global network. "Abundance" because the volumes of original content and new content being imported from other media grows by numbers which can only be abstracted. "Scarcity" because the number of hours in a day hasn't changed and, in potential, "a wealth of information creates a poverty of attention" (Simon, 1971). These long established ideas of the challenges posed by an abundance of information have never been more pertinent, because "the real scarce resource will always be human attention and attracting that attention will be the necessary precondition of social change" (Cowen, 2007).

This is why any filters between content producers and content consumers have become so important: the market is so competitive and the human attention so relatively inelastic that, from a producer's perspective, missing a chance for visibility by being filtered-out, or getting a chance for visibility by being filtered-in, can be decisive. On the same token but from the consumption side, failing to filter-in preferred information or to filter-out less desirable alternatives, can translate to a bad attention investment - the scarcest of the investable resources in the new economy. Therefore, the power of search engines such as Google, who perform as filtering devices in the marketplace of attention, is just immense and the public space can benefit from other devices and other search possibilities.

In automatically eliminating ill-favored links, systems of stigmergic hyperlinks function as filtering devices in the market of attention. They are valuable in assisting communities of users who consistently express preferences in the same polarizing direction. On the other hand, if the consumers/users don’t really constitute a community, by having fragmented interests; or the producers/publishers aren't being able to attract a stable audience, then fragmentation takes place and the effects of running this and other equivalents of recommender systems can be erratic.

Web metrics

In (Dhyani, Ng, & Bhowmick, 2002) the authors classify the metrics for the Web in six categories: "Web graph properties", for quantifying structural properties of networks; "Web page significance", for formalizing the concepts of "quality" and "relevance"; "usage characterization" for measuring user behavior; "Web page similarity" for quantifying the relatedness of pages; "Web page search and

retrieval", for comparing different search and retrieval services; and "information theoretic", for capturing properties of information needs, production and consumption.

When using a hyperlink, users manifest behaviors which can be measured with tools that fit the "usage characterization" category; on the long run, it should be possible to better understand their information preferences with some "information theoretic" reasoning. For most hyperlinking use-cases, these are the most relevant categories, but it is almost consequential to put other types of measurements to work: for example, when a system of stigmergic hyperlinks is to replace one link whose vitality nulled, it computes its replacement from the network of "survivors". There is a chance for increased "Web page similarity" across the linked destinations, strictly based on the nodes' proximity, yet there was no content analysis and the network graph emerges from basic reward/decay rules, in bio-inspired stigmergic fashion. Identifying the survivors and their hyperlinks is a graph based measurement.

Behind the scenes, most clicks now spark computations based on data ranging from the basic variables provided by every client-server environment (local time, IP address, referrer, etc.) to the complex, specific to each application's purpose.

Measuring hyperlinking in the "Deep Web" is much harder (Bergman, 2001; He, Patel, Zhang, & Chang, 2007). The retrieval of deep webs where different contents can have the same URL is a challenge that cannot be approached with horizontal solutions dependent on a 1:1 correspondence between address and content, requiring very specific approaches or a radically new general search engine which can index content not only by node address, but also by trails of user interactions which open the content: I call it the trails-based search engine ("TB search engine"). There is no such solution yet and the consciousness of its need won't reach critical mass while the information needs of most people are being satisfied with the current services, who deliver in quantity and with a quality that can't be contrasted with alternatives that don't exist. However, some technical client-side factors for hidden data, such as content induced Javascript, have begun to be handled - see the "CrawlJax" tool by (Mesbah, Deursen, & Lenselink, 2012). Others are working on the front-end of search, trying to think tomorrow's Web with more engaging "interactive visualizations" (Dork, Williamson, & Carpendale, 2012).

3. SOME CONSEQUENCES OF THE STATUS QUO

Such is today's hyperlinked society that an entire new industry was created to support the contemporary hyperlink needs and wants, and research new ones. At the core of this industry, Google is probably the strongest example of a hyperlink-based business. Most of Google's revenue comes from ad selling, with ads being delivered via sponsored search results presented as hyperlinks, or via the AdWords + AdSense programs. AdWords enables do-it-yourself Web campaigns, while AdSense enables participants to be part of an ad space network. One fundamental difference between selling ads online and selling ads on more mature media like newspapers, TV and radio, is that the Web is currently the only medium with a prevalent two-stage business model for advertisements, charging the advertiser not for the mere advertisement display, but only for subsequent interactions with it, in the form of clicks. In other media, the business model charges per audience quantity and not per interaction. Not only there is this significant difference, but the available Internet technologies provide real time, finer granularity, measuring tools that have the potential to support more detailed, better informed, more transparent campaigns.

Google, online ad networks in general, and sponsored web search results are examples of relatively new businesses. Orbiting the search engine players yet on the same marketplace for attention, a myriad of other organizations sell services which were unimaginable a few years ago: there are markets for domain names, infomediaries in buying/selling, appraisal services for all types of content for reasons which include placing higher in search engines' rankings, security services, and many other neo offerings, including new educational curricula tailored for the hyperlink-based economy. In all the diversity there is one common goal: a "better" hyperlink, with "better" having many subjective meanings, e.g. "shorter, easier to remember", "better placed, better redacted, worthy of higher rank", "immune to link rot", or "less prone to phishing".

The point is that while the hyperlink is always at the center, there are many different, sometimes contradictory, reasons for its publication. The industry is gaming the hyperlink: for example, when a Google search returns some results, it is doing its algorithmic best to deliver value to the user; but, at the same time, some firm in the Search Engine Optimization (SEO) business, will try to push its customer's site higher and higher, for certain search expressions, regardless of alternatives, playing every artificial hyperlink trick in the book, if necessary. There is a huge disproportion between the scale and objectivity of these tasks: Google's is massive and spread, trying to serve millions, while the SEO player's is singular and focused, with a single client to please. The resources available to each are even more disproportionate; in the end this is a cat and mouse situation with two very different motivations for hyperlink publishing. Every time such extreme disproportions appear, as in this particular hyperlink use case, one has to ask: is there room in the middle?

As elaborated in (Marques, 2014), there might indeed be room in the middle. One idea is to reinforce the original association purpose of the hyperlink by enabling custom in-context search solutions. However, in the current "search environment", heavily absorbed by a handful of large scale solution providers, it is very hard for these topical solutions to find a place. The need for vertical solutions is sometimes not perceived at all because of the amount of already available indexed content - we can only pay attention to so much, and there is a reasonable chance that in the already charted WWW there is more than enough to distract us from unfulfillment. The industry knows our human limits and comfort zones: for example, the default 10 search results per page that most search engines will display is an option which finds support in TV research suggesting that, no matter how many content channels might be available, the majority of viewers will only manage and consume 10 to 15 of them (Jeffres, 2001), at most.

And who may be those authoring custom search/hyperlinking solutions? Following Benkler's (Benkler, 2006) view of "ideal-type information production strategies", they'll probably approach, but not necessarily match, the "Joe Einsteins" or the "Romantic Maximizers", depending if their main driver is "fame" or "fortune", respectively. Benkler thinks of "Joe Einsteins" as those willing to "give away information for free in return for status, benefits to reputation, and value of the innovation to themselves". The "romantic maximizers" are authors who sell their creations, for example to publishers. In both cases issues of intellectual property can make the situation less clear because Joe's "free" solution might infringe unanticipated publisher rights which can force him to have to charge to be able to pay, or desist; and the romantic's product might violate copyleft licenses which force him to open his code and to have to adapt to a different revenue model, for example support-based.

It is worth noting that "open source software" and "free software" are legal concepts, corresponding to software with a license in which the copyright holder provides at least the rights to use, modify, copy and distribute the software, to anyone, for any purpose. Some licenses "attach" to derivative works (copyleft licenses) so that future versions and/or branches must reciprocate the conditions the author first received. The information economy is still relatively young and many questions of intellectual property haven't been ruled in court at all; some decisions will influence information products and services (Lindberg, 2008).

If the most fundamental object in hypertextuality could exhibit a customizable behavior, then hyperlink authors and publishers would have reasons to think beyond the citation & navigation seminal purposes, beyond the current tracking and ad selling functions, and could even rethink what "search" is: search does not have to be literally about finding something at a destination resource; it can be also about identifying alternatives to the resource, or providing richer visualizations. For example, imagine yourself reading some article that, at a point, links to somewhere with an anchor like "<movie name here> is frightening"; in today's WWW the usage pattern creates the expectation that clicking the anchor will bring you to some site about the film and that "searching" it is to look for some expression there, but the search behavior could handle the string differently, e.g. as a request to suggest/compare with other frightening movies using some fright-scale, or to produce some representation of classifiers associated with the movie, highlighting the searched one; or to find community discussions in that context. So, there are alternatives to centralized horizontal solutions, there are alternatives to literal textual searches (Dork et al., 2012), and there are behavioral possibilities besides the vanilla hyperlink's features.

Hyperlinking is just beginning, but has already become so embedded in our lives that having an awareness of how we got here and being able to escape established patterns should be relevant.

4. REFERENCES

Adamic, L. (2008). The social hyperlink *The Hyperlinked Society*.

Adamic, L. A. (2009). *The social hyperlink*. Paper presented at the Proceedings of the 20th ACM conference on Hypertext and hypermedia, Torino, Italy.

Barnhurst, K. G. (2002). News Geography & Monopoly: the form of reports on US newspaper internet sites. *Journalism Studies, 3*(4), 477-489. doi: 10.1080/1461670022000019218

Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*: Yale University Press.

Bergman, M. K. (2001). The deep web: Surfacing hidden value. *Journal of Electronic Publishing, 7*(1), 07-01.

Bush, V. (1945). As We May Think. *The Atlantic*.

Bush, V. (1996). As we may think. *interactions, 3*(2), 35-46. doi: 10.1145/227181.227186

Ciszek, T., & Fu, X. (2005). An annotation paradigm: The social hyperlink. *Proceedings of the American Society for Information Science and Technology, 42*(1), n/a-n/a. doi: 10.1002/meet.14504201288

Cowen, T. (2007). Richard A. Lanham, The Economics of Attention: Style and Substance in the Age of Information. *Journal of Cultural Economics, 31*(2), 167-169. doi: 10.1007/s10824-007-9036-9

Dhyani, D., Ng, W. K., & Bhowmick, S. S. (2002). A survey of Web metrics. *ACM Comput. Surv., 34*(4), 469-503. doi: 10.1145/592642.592645

Dimitrova, D. V., Connolly-Ahern, C., Williams, A. P., Kaid, L. L., & Reid, A. (2003). Hyperlinking as Gatekeeping: online newspaper coverage of the execution of an American terrorist. *Journalism Studies, 4*(3), 401-414. doi: 10.1080/14616700306488

Dork, M., Williamson, C., & Carpendale, S. (2012). Navigating tomorrow's web: From searching and browsing to visual exploration. *ACM Trans. Web, 6*(3), 1-28. doi: 10.1145/2344416.2344420

Haslhofer, B., & Popitsch, N. (2009). *DSNotify - Detecting and Fixing Broken Links in Linked Data Sets*. Paper presented at the Proceedings of the 2009 20th International Workshop on Database and Expert Systems Application.

He, B., Patel, M., Zhang, Z., & Chang, K. C.-C. (2007). Accessing the deep web: a survey. *Communications of the ACM, 50*(5), 94-101.

Hespos, T. (2008). How Hyperlinks Ought to Change the Advertising Business. In J. T. L. Tsui (Ed.), *The Hyperlinked Society: Questioning Connections in the Digital Age* (pp. 137-144): The University of Michigan Press.

Ingham, D. B., Caughey, S. J., & Little, M. C. (1997). Fixing the" broken-link" problem: The W3Objects approach. *TECHNICAL REPORT SERIES-UNIVERSITY OF NEWCASTLE UPON TYNE COMPUTING SCIENCE*.

Jeffres, K. A. N. D. J. A. L. W. (2001). Reconceptualizing Channel Repertoire in the Urban Cable Environment. *Journal of Broadcasting & Electronic Media, 45*(3), 464-482.

Kim, H. J. (2000). Motivations for hyperlinking in scholarly electronic articles: a qualitative study. *Journal of the American Society for Information Science, 51*(10), 887-899.

Kovilakath, V. P., & Kumar, S. D. M. (2012). *Semantic broken link detection using structured tagging scheme*. Paper presented at the Proceedings of the International Conference on Advances in Computing, Communications and Informatics, Chennai, India.

Krishnamurthy, S. (2003). E-Commerce Management : Text and Cases / Slatanic - the disastrous move from free to paid content: Thomson / South-Western College Pub.

Leuf, B. (2006). Enhancing the Web *The Semantic Web - Crafting Infrastructure for Agency* (pp. 3-30): Wiley.

Lindberg, V. (2008). *Intellectual Property and Open Source*: O'Reilly Media.

Maeyer, J. D. (2013). Towards a hyperlinked society: A critical review of link studies. *New Media & Society, 15*(5), 737-751. doi: 10.1177/1461444812462851

Marques, A. (2014). *Stigmergic Hyperlinks' contributes to Web search*. Paper presented at the CAPSI, Santarém, Portugal.

Marques, A., & Figueiredo, J. d. (2010). Stigmergic Hyperlink: a new social web object. *IJISSC - International Journal of Information Systems and Social Change, 2*(4), 31-43. doi: 10.4018/jissc.2011100103

Mesbah, A., Deursen, A. v., & Lenselink, S. (2012). Crawling Ajax-Based Web Applications through Dynamic Analysis of User Interface State Changes. *ACM Trans. Web, 6*(1), 1-30. doi: 10.1145/2109205.2109208

Miles-Board, T., Carr, L., & Hall, W. (2002). *Looking for linking: associative links on the Web*. Paper presented at the Proceedings of the thirteenth ACM conference on Hypertext and hypermedia, College Park, Maryland, USA.

Morishima, A., Nakamizo, A., Iida, T., Sugimoto, S., & Kitagawa, H. (2009). *Bringing your dead links back to life: a comprehensive approach and lessons learned*. Paper presented at the Proceedings of the 20th ACM conference on Hypertext and hypermedia, Torino, Italy.

Simon, H. A. (1971). Designing Organizations for an Information-Rich World *Computers, communications, and the public interest*: Johns Hopkins University Press.

Thelwall, M. (2003). What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation. *Information Research, 8*(3).

Tremayne, M. (2005). News Websites as Gated Cybercommunities. *Convergence: The International Journal of Research into New Media Technologies, 11*(3), 28-39. doi: 10.1177/135485650501100303

Webster, J. G. (2008). Structuring a Marketplace of Attention. In J. T. L. Tsui (Ed.), *The Hyperlinked Society* (pp. 23-38): University of Michigan Press.