



Padrões de conectividade e links na web: uma proposta teórico-metodológica para avançar a Webometria e a Análise de Hiperlinks¹

Suely Fragoso²
Unisinos

Resumo

Este trabalho propõe uma abordagem comunicacional para a análise da conectividade de websites. A primeira seção consiste em uma breve revisão das principais correntes de estudos de hiperlinks na web, cujas limitações são apontadas e discutidas. A segunda seção contempla os princípios fundamentais da proposta de que a adoção de uma abordagem comunicacional seria capaz de revitalizar os estudos de Webometria e Análise de Hiperlinks. A terceira seção exemplifica a proposta a partir de um estudo em andamento, no qual se aborda a conectividade internacional de sites com domínio brasileiro (.br) procedendo análises quali-quantitativas em diferentes escalas: das ccTLDs, dos websites, das páginas e dos hiperlinks.

Palavras-chave: hiperlink; web; conectividade; webometria; análise de hiperlinks

Concebida como uma estrutura hipertextual global para compartilhamento de informação (Berners-Lee, 1990, 1996), a web cresceu para tornar-se o maior e mais conhecido sistema hipertextual e a parcela mais proeminente da internet. Nos dias de hoje, a web não é apenas mais ampla que sua proposta inicial, mas também mais diversificada. Algumas de suas possibilidades originais desapareceram, outras foram modificadas, mas o papel central dos hiperlinks não se alterou.

ESTUDOS DE HIPERLINKS

Hiperlinks são conexões automatizadas que, quando acionadas, dão acesso a outro módulo de informação, não necessariamente em ordem linear (Weinreich, Obendorf e Lamersdorf, 2001, p. 19; Primo, Recuero e Araújo, 2004). Definidos assim, os links parecem ser muito simples – e, de certo modo, são mesmo. Essa simplicidade é o que garante sua enorme flexibilidade e permite que eles sejam utilizados de muitas formas, com diferentes objetivos e distintas motivações. Não surpreende, portanto, que seja mais fácil entender o que os links *são* do que o que eles *fazem* ou o que *significam* (Halavais, 2008).

Não é incomum tentar entender novidades tecnológicas lançando mão de comparações com elementos já conhecidos. Muitas áreas de conhecimento abordaram os hiperlinks desse modo, cada uma definindo os termos das analogias de acordo com suas próprias

¹ Trabalho apresentado no GP Cibercultura, IX Encontro dos Grupos/Núcleos de Pesquisas em Comunicação, evento componente do XXXII Congresso Brasileiro de Ciências da Comunicação.

² Professora Titular do Programa de Pós-Graduação em Ciências da Comunicação e do Curso de Graduação em Comunicação Digital da Unisinos, RS.



premissas, métodos e objetivos. Para Letras e Linguística, por exemplo, links são semelhantes a formas de pontuação (Bolter, 1997); para a Biblioteconomia, a citações (Rousseau, 1997; Thelwall, 2003; Smith, 2004). As ciências sociais os consideram análogos a canais de comunicação (Park, 2003), fluxos de informação (Barnett e Park, 2003 e 2004; Barnett e Sung, 2006), etiquetas identitárias (Tomasini, 2006; Adamic, 2008), laços sociais (Adamic e Adar, 2001; Recuero, 2009), etc.

A analogia entre hiperlinks e citações foi, até aqui, a mais produtiva. A idéia já estava implícita na descrição do Memex por Vannevar Bush (1945) e deu origem a um novo campo de estudos, denominado Webometria³. A comparação com citações coloca ênfase em um sub-conjunto das possíveis motivações e usos dos hyperlinks, mas resguarda alguma flexibilidade, já que as citações, como os links, não são todas iguais: referências ao trabalho de outros podem endossar, reconhecer como fonte de conceitos ou dados, criticar, analisar, persuadir, etc (Brooks *apud* Halavais, 2008, p. 41). O que permanece é a idéia de que o trabalho referido é suficientemente importante para ser mencionado. Combinado com técnicas quantitativas de bibliometria e transposto para os estudos de hiperlinks, este princípio se traduz no pressuposto de que o número de links dirigidos para um site ou página é um indicador de sua relevância. Esta idéia já estava presente no algoritmo de hierarquização do buscador *Lycos* (Mauldin, 1997) e alcançou enorme popularidade com o algoritmo *PageRank*, do *Google* (Brin e Page, 1998), desencadeando um grande interesse nos estudos quantitativos de hiperlinks.

Os estudos quantitativos da web também foram favorecidos pela existência de um arcabouço teórico-metodológico adequado, a Teoria das Redes. Todas as redes são compostas dos mesmos elementos básicos: nós e conexões, e compartilham muitas propriedades estruturais. Análises baseadas na Teoria das Redes revelaram, por exemplo, que, como outras redes dinâmicas, a web é multicentrada e distribuída em ‘cauda longa’⁴ (Barabási, 2002, p. 65-92). Sem negar que análises realizadas nesse grau de abstração revelam características importantes, é preciso reconhecer que elas são limitadas, já que deixam de levar em conta justamente o que diferencia a web das demais redes.

Resultados mais específicos vêm sendo obtidos com um subconjunto da Teoria das Redes conhecido como Teoria Social das Redes, que se ocupa dos padrões estruturais das relações sociais. A Análise de Redes Sociais combina o poder das abordagens

3 Webometria é “o estudo dos aspectos quantitativos da construção e uso de recursos informacionais, estruturas e tecnologias na Web, baseado em bibliometria e infometria” (Björneborn, 2004, p. 12)

4 Mais estritamente, é uma rede sem escala distribuída em curva zipf.



estruturais com o reconhecimento da natureza social da web, assumindo que “um mapa da rede de comunicação é mais ou menos isomórfico à estrutura dos relacionamentos entre os usuários” (Garrido e Halavais, 2003, p 172). Esta comparação está na origem da analogia entre websites e atores sociais (Park, 2003) bem como entre links e laços sociais, relações ou interações (Adamic e Adar, 2001; Adamic, 2008; Prior, 2008; Recuero, 2009). Tomadas com excessiva literalidade, essas correspondências podem levar a interpretações duvidosas dos padrões estruturais revelados com técnicas de análise de redes sociais.

A disjunção macro/micro

Boa parte dos estudos de hiperlinks realizados na área de ciências sociais é baseada em Análise de Redes Sociais ou em Bibliometria. Em ambos os casos, operações quantitativas próprias de aparatos metodológicos pré-estabelecidos são apropriadas com base em analogias pouco rigorosas entre hiperlinks e citações (na tradição da Bibliometria) ou interações e laços sociais (nos estudos inspirados na Análise de Redes Sociais). Essas correspondências imprecisas não são evidentes nos resultados estruturais, mas se revelam na hora de sua interpretação.

O preço que a análise de hiperlinks paga por sua frágil compreensão dos nós e conexões que articulam a web é a dificuldade de progredir das descrições dos padrões estruturais para sua interpretação e discussão. Thelwall (2006) acredita que a solução resida no desenvolvimento de uma “Teoria da Linkagem” capaz de “prover uma explicação definitiva que possa ser usada para a interpretação dos achados de pesquisas de análise de links”; teoria esta que, no entanto, ele considera impossível. É duvidoso que a existência de um conjunto claramente definido de teorias e métodos a serem repetidamente aplicados a uma sequência de casos semelhantes entre si dotaria os estudos de hiperlink de maior cientificidade, ou conduziria a melhores interpretações dos resultados. O próprio Thelwall denomina sua abordagem ‘altamente positivista’ e, nesse paradigma, atribui a ineficiência da análise de links ao caráter dinâmico e à heterogeneidade da web – duas características que não são estranhas às ciências sociais, já que a própria sociedade existe em permanente modificação e é altamente heterogênea.

As colocações acima têm relação direta com a histórica tensão entre as análises sociais em escala micro e macro, frequentemente traduzidas na preferência por métodos quantitativos e qualitativos. No caso, a fragilidade dos estudos de links aparece na



passagem das abordagens quantitativas para as ações interpretativas, produzindo uma falsa impressão de que o problema se restringe à escala micro, ou às operações qualitativas. Porém, essa dicotomia macro-micro é uma fratura imposta por diferentes posturas epistemológicas (Alexander e Giesen, 1987), ou seja, é o nosso modo analítico de compreender o mundo que impõe à realidade essa idéia de camadas em que os fenômenos ocorrem, separados em escalas distintas. Além disso, macro e micro são então divisões amplas,

entre as quais alguém poderia introduzir uma imagem de níveis interdependentes muito mais precisa. (...) Os termos ‘micro’ e ‘macro’ são completamente relativos. O que é macro em um nível será micro em outro. Diferentes propriedades são associadas a diferentes níveis, e problemas específicos podem demandar a conversão de parâmetros em variáveis conforme o movimento para maiores ou menores unidades de referência. Cada nível, entretanto, é homólogo a cada um dos outros, e não há processo da vida empírica que possa subsistir em um único nível. Paradoxalmente, cada fenômeno empírico pode ser compreendido em qualquer nível de análise; apenas é compreendido de um modo diferente. Essas distintas compreensões revelam propriedades importantes do processo (Alexander, 1987, p. 291)

A reintegração das dicotomias macro-micro e quantitativo-qualitativo requer a articulação de construções metodológicas capazes de operar em escalas diversas, conforme as intenções e necessidades da pesquisa. Abordagens multi-escala, por sua vez, implicam que o equilíbrio de causalidades varia ao longo do tempo e conforme se alterem as condições (Gerstein, 1987, p. 89-90), exigindo o apoio de um arcabouço teórico suficientemente sólido e flexível. A próxima seção deste texto propõe que uma abordagem comunicacional da web poderia desempenhar esse papel.

ABORDAGEM COMUNICACIONAL DA WEB

A web é chamada de mídia nos meios acadêmicos e pelo público em geral. À primeira vista, pode parecer que se trata de mais uma analogia para explicar uma novidade em termos de algo previamente conhecido. A relação entre ‘web’ e ‘mídia’, entretanto, não é comparativa (como seria, por exemplo, entre a web e a televisão), ela indica que a web é um entre vários elementos que pertencem a uma categoria. Se a web é mídia, então as teorias da comunicação, mais precisamente as teorias da mídia, devem ser um arcabouço teórico adequado para estudá-la.

Mesmo que outros, ou muitos, já o tenham feito, certamente não basta apenas afirmar que a web é mídia. É preciso explicar *como* e *porque* a web é mídia – um desafio e tanto, tendo em vista que a própria definição de mídia é motivo de intensas disputas. Uma análise detalhada das propriedades midiáticas da web seria proveitosa para



entender a rede e também a mídia, mas a empreitada está fora do escopo deste trabalho. Para o momento, será suficiente traçar as linhas gerais dessa idéia. O que é essencial para a proposta apresentada é que as muitas definições de mídia existentes refletem uma diversidade de posturas epistemológicas e de escalas de análise.

A mídia web e a Análise de Hiperlinks

Uma abordagem comunicacional da web reposiciona as dificuldades enfrentadas pela Análise de Hiperlinks em termos dos estudos midiáticos. Havíamos mencionado anteriormente que a equiparação entre websites e atores sociais constituía um ponto particularmente frágil da Análise de Hiperlinks. A teoria da comunicação ajuda a perceber que os websites não são atores sociais, são *produtos midiáticos*, explícita e deliberadamente *criados por atores sociais* lançando mão dos dispositivos tecnológicos e das técnicas próprias do meio. Assim como outros produtos midiáticos, os sites são portadores não apenas dos conjuntos de significações conscientemente desejados por seus criadores, mas também de um segundo nível de conteúdo, relativo a (e indicativo de) suas condições de produção e existência. Por isso, uma abordagem midiática requer o reconhecimento das diversas mediações (da linguagem e também técnicas, sociais, culturais, políticas e econômicas) (Martín-Barbero, 2003) que afetam e configuram a web.

Como muitos outros mídias, a web foi criada com certas características e finalidades que foram posteriormente alteradas pela apropriação social; no caso, foi concebida como um repositório acadêmico (Berners-Lee, 1990) e posteriormente tornou-se mídia. No processo, boa parte do perfil colaborativo do projeto original da web perdeu força diante de seu potencial econômico. O número de websites comerciais superou o de universidades e institutos de pesquisa e também as *homepages* pessoais, direcionando a web para um modelo de publicidade e serviços em que os sites funcionam como vitrines dos produtos e serviços das empresas e estão frequentemente subordinados aos departamentos de marketing (Thelwall, 2009, p. 45). Poucos anos mais tarde, o estouro da ‘bolha pontocom’ e um conjunto de avanços técnicos impulsionaram a retomada da vocação colaborativa da web. A popularização de sistemas baseados em conteúdos criados pelos usuários teve um impacto sobre o modelo distributivo da web, bem como sobre o conteúdo e a forma dos produtos midiáticos nela disponíveis. Websites institucionais são bastante mais semelhantes ao que se podia encontrar em mídias anteriores, enquanto a inovação é mais evidente nos sistemas colaborativos. Esta



diferença ajuda a perceber que, como a maioria das mídias, a web apresenta diversos tipos de sites, que são produzidos por diferentes atores sociais (empresas, governos, indivíduos, etc.) com finalidades diversas, que se refletem em diferenças de conteúdo, estrutura, retórica, etc. Nesse sentido, a palavra ‘website’ denomina uma categoria que abrange uma variedade de tipos e sub-tipos de artefatos – corresponderia, em um paralelo com outra mídia, a ‘programa de TV’ por exemplo.

‘Hiperlink’ é uma outra categoria que abrange uma variedade de manifestações de um certo tipo de elemento, no caso, interno aos websites. Assim como os sites, os links são *criados* deliberadamente⁵ e apresentam muitos significados, funções, modos de apropriação das linguagens, etc. A constante que caracteriza o hiperlink, e o diferencia dos demais elementos dos sites, é sua capacidade de dar acesso a outras partes ou a outros produtos midiáticos abrigados na mesma mídia (no caso, a web). Esse acesso é uma materialização técnica das referências a outros produtos midiáticos comuns no universo da mídia. Tratam-se de referências de nível simbólico, por exemplo entre programas de televisão (novelas que mencionam fatos abordados nos telejornais, propagandas que referem situações de programas de auditório, etc.). Assim como os links, essas referências são elementos constituintes dos programas de televisão em que se inserem, e variam em intensidade, intenção, forma, etc. O universo de práticas é amplo, mas alguns padrões são evidentes: os programas mais populares são os mais referidos, programas de conteúdo semelhante fazem referências mútuas, etc.

As considerações acima pretendem ser balizamentos iniciais para uma Análise de Hiperlinks norteada pela abordagem midiática. A próxima seção consiste de um exemplo de tentativa de atravessar escalas de observação dos hiperlinks na web sem perda de coerência, tendo como base o arcabouço teórico proposto.

EXEMPLO DE CONSTRUÇÃO MULTIMÉTODO MULTI-ESCALA BASEADA NA ABORDAGEM MIDIÁTICA DA WEB

Esta seção apresenta um exemplo de construção teórico-metodológica a partir de uma investigação em curso, norteada pela abordagem midiática da web. O ponto de partida são resultados obtidos com Análises de Hiperlinks realizadas por outros autores.

Origem e Construção do Problema de Pesquisa

⁵ Existem links que são produzidos automaticamente, mas não espontaneamente, pois obedecem a especificações de programação e decorrem de ações do usuário ou do próprio sistema.



Halavais (1998), Barnett e Park (2003) e Barnett e Jun (2004) realizaram contagens de hiperlinks entre websites com ccTLDs⁶ de diferentes países com a intenção de visualizar a conectibilidade internacional da web. Entre seus resultados, foi possível perceber que os sites .br apresentavam alta conectividade com sites de outros países, inclusive recebendo mais *inlinks*⁷ internacionais que enviando *outlinks* para sites de outros países (Fragoso, 2005; 2006; 2007).

Para Halavais (1998), o número de links entre ccTLDs reflete a existência de canais de comunicação entre atores sociais dos Estados-Nação correspondentes. Barnett, Park e Jun (2003 e 2004) consideram que o grau de interconectibilidade dos ccTLDs é indicativo de centralidade no sistema-mundo (Wallerstein, 1979). Em ambos os casos, a interpretação dos resultados se baseia em uma analogia entre links e fluxos de informação. Compreender a web como mídia não contradiz essa analogia, mas a reposiciona. Em termos comunicacionais, os resultados obtidos por aqueles autores significam que existe um número maior de referências a produtos midiáticos brasileiros em produtos de outros países do que o inverso. Traduzindo em termos de outra mídia, corresponderia a dizer que os jornais de outros países fazem mais menções aos jornais brasileiros que vice-versa, por exemplo.

Esse quadro suscita um conjunto de perguntas: que referências internacionais são essas? Em que tipos de produtos midiáticos se encontram e a que tipo se dirigem? Por que foram criadas? O que significam? Que implicações têm? Essas questões correspondem justamente à passagem do nível macro para o nível micro que, como mencionamos anteriormente, constitui um dos pontos fracos da Análise de Hyperlinks.

O primeiro passo para abordar essa elevada conectibilidade internacional dos sites brasileiros consistiu na construção de uma amostra composta por websites com ccTLD .br e sites com outras ccTLDs a eles conectados. Por se tratar da primeira Análise de Hiperlinks com abordagem comunicacional que realizamos, optamos por uma amostra de sites mais semelhantes a produtos midiáticos pré-web, ou seja, websites institucionais. Isso implicou a exclusão de sites colaborativos (blogs, sites de rede social, wikis, etc.)⁸.

⁶ ccTLD é a sigla para *country code Top Level Domain*. Nomes de Domínio denominam computadores específicos na internet, diferenciando cada um dos demais. A última parte de um Nome de Domínio é um *Top Level Domain* (TLD). Existem dois tipos de TLDs, os genéricos e os de código de país. TLDs genéricos são gTLDs (como .com, .org, .net) e são usados pelo público em geral. Os ccTLDs são associados a países ou áreas geográficas (.br, .de, .to).

⁷ Em relação a um site A, *inlink* é um link em outro site que aponta para A e *outlink* um link em A que aponta para outro site.

⁸ Esta opção não implica que as estratégias metodológicas adotadas e propostas não seriam válidas para sistemas colaborativos, nem que a abordagem comunicacional da Web não abrange os mesmos.



Construção da Amostra

Uma amostra representativa do ponto de vista comunicacional deveria ser composta por websites .br com boa visibilidade, o que sugeriu sua localização através de ferramentas de busca. A construção de amostras com buscadores é mais comum nas chamadas Avaliações de Impacto na Web (Thelwall, 2009, p.9-26) e pode ser problemática na Análise de Hiperlinks, já que eles indexam apenas uma fração da web (Gulli e Signorini, 2005, p. 2), são linguística e nacionalmente tendenciosos (Vaughan e Thelwall, 2004, p. 13) e seus critérios de inclusão, exclusão e hierarquização dos resultados não são conhecidos (Fragoso, 2007). Entretanto, a quase totalidade dos usuários de internet utiliza ferramentas de busca para a localização de páginas, mesmo quando acessam sites anteriormente conhecidos, sendo que, “para a maioria dos usuários, tudo se passa como se a web se restringisse ao conteúdo dos bancos de dados dos grandes buscadores” (Fragoso, 2007). Uma vez que a visibilidade dos sites é um critério central para uma análise comunicacional, optou-se pela construção da amostra com buscadores. Foram utilizadas as duas ferramentas de busca mais populares, *Yahoo!* e *Google*, produzindo um total de 25,200 URLs .br cuja comparação revelou um subconjunto de 102 sites com visibilidade particularmente elevada. Uma amostra dessas dimensões certamente não possui representatividade estatística – o que, de todo modo, não é realmente viável em um meio com a amplitude e o dinamismo da web⁹. Desde o princípio, a intenção foi construir uma amostra *qualitativamente* significativa, para viabilizar o trabalho com os sites, páginas e links. Todos os sites selecionados foram checados individualmente para conferir sua concordância com os critérios de amostragem, com 100% de confirmação.

A distribuição dos sub-domínios pelos 102 sites foi equilibrada, com um número maior de sites .br (domínio utilizado pela maioria das grandes instituições de ensino e pesquisa) e .gov.br (refletindo o alcance e a penetração do projeto brasileiro para implantação do governo eletrônico (Pinto e Fernandes, 2005). Os demais sub-domínios apareceram em menor proporção. Embora o domínio de segundo nível mais frequente no ccTLD .br seja o .com.br (91,94% em junho de 2009, Registro.br, 2009), era esperado que a baixa conectibilidade dos sites comerciais (Thelwall, 2009, p. 45) prejudicasse sua colocação em buscas por ccTLD. Todos os sites selecionados tinham conteúdo em português e compatível com seus nomes de domínio.

9 O universo de sites .br era de 757.925 em junho de 2005 (época da primeira amostragem) e 1.715.798 em junho de 2009 (Registro.br, 2005; 2009).



Conectibilidade Internacional

A conectibilidade internacional de uma amostra qualitativa, comunicacionalmente relevante e focada em sites .br não corresponderia necessariamente aos resultados encontrados por Halavais (1998) e Barnett e Park e Jun (2003; 2004). Nos trabalhos daqueles autores, os sites .br apareceram colateralmente e em função justamente de receberem *inlinks* de sites de outros países – isso não implicava que fossem justamente os sites .br mais visíveis¹⁰. Assim, nossa amostra poderia apresentar um padrão de conectividade internacional que não corresponderia necessariamente à estrutura geral da web, mas que, entretanto, dados os dados os critérios de amostragem, seriam mais relevantes para a abordagem comunicacional dos sites com domínio .br.

Inlinks vindos de sites internacionais foram localizados com buscadores e *outlinks* com crawlers¹¹. Os resultados das buscas e mapeamentos foram analisados por amostragem para exclusão dos sites com gTLDs ou com ccTLD não-br com nomes de domínio em língua portuguesa (como seria o caso, por exemplo, de <http://www.midiasdigitais.org> ou <http://www.alexprimo.com>), de modo a evitar a possível distorção dos resultados por um alto número de conexões de e para sites brasileiros registrados com domínios diferentes do .br¹².

Foram encontrados 78.989 *inlinks* e *outlinks* internacionais, confirmando a elevada conectibilidade internacional dos sites .br também em amostras qualitativas de sites com alta visibilidade (Fragoso, 2007). A quantidade elevada de links internacionais encontrados apontava problemas operacionais para a próxima etapa, que consistiria na identificação mais detalhada dos *inlinks* e *outlinks*, passando das operações de contagem no nível macro (dos TLDs) para análises nos níveis intermediário e micro (dos sites, páginas e links). Por esse motivo, amostra inicial de sites foi recortada com os seguintes critérios: eliminação dos sub-domínios que apresentaram conectibilidade mais baixa que os demais¹³ e verificação de permanência entre os melhor colocados nas ferramentas de busca. Novas buscas com os mesmos critérios utilizados anteriormente permitiram identificar que 44 dos 77 com maior número de conexões internacionais na amostra original permaneciam entre os 50 primeiros resultados nos dois buscadores. Uma nova

¹⁰ Embora desde o sucesso do *PageRank*, os buscadores levem em conta o número de *inlinks* para hierarquizar os resultados das buscas, esse critério não é utilizado de forma exclusiva e não se restringe aos *inlinks* internacionais.

¹¹ Detalhes dos procedimentos e critérios foram detalhados em Fragoso, 2005; 2006 e 2007.

¹² As TLDs genéricas .com, .org, .net são de uso mundial (embora, por razões históricas, estejam mais concentradas nos EUA). No Brasil, a incidência de sites abrigados em domínios genéricos é particularmente alta devido à dificuldade para registro de domínios .br.

¹³ A amostra inicial abrangeu o domínio .br e seis sub-domínios: .com.br, .org.br, .gov.br, .edu.br, .ind.br e .inf.br. Três deles (.edu.br, .ind.br e .inf.br) tinham conectibilidade bem menor que os demais e foram eliminados da segunda amostra.



amostra foi constituída com esses 44 sites, cujos *inlinks* e *outlinks* internacionais (41.731) foram também atualizados com buscadores e crawlers.

A categoria ‘hiperlink’ abriga uma variedade de tipos de hiperlinks, o que dificulta a interpretação e discussão da alta conectibilidade dos sites .br. Para compreender os links é preciso conhecer o contexto e as condições em que eles ocorrem, ou seja, identificar os sites e as páginas onde se encontram.

Análise nos níveis intermediário e micro (sites, páginas e links)

A análise para a caracterização dos hiperlinks da pesquisa utilizada para exemplificar a proposta de abordagem comunicacional encontra-se em fase inicial e terá lugar em três escalas distintas: sites, páginas de ocorrência e links propriamente ditos. Esses níveis de análise serão concentrados sobre as ccTLDs que apresentam maior conectividade com a .br e serão realizados por ccTLD. A primeira ccTLD, para a qual os procedimentos já foram iniciados, é a da Alemanha (.de), que apresentou alta troca de links com sites .br. A caracterização dos websites por ccTLD ainda implica o manuseio de um número considerável de dados¹⁴, sugerindo a sistematização por Análise de Conteúdo, uma abordagem metodológica bastante utilizada nos estudos de mídia (Riffe, Stephen e Fico, 2005) e que já foi anteriormente testada com Análise de Hiperlinks (Adamic e Adar, 2001; Bar-Ilan, 2004; Thelwall, 2009). Essa técnica permite combinar a avaliação qualitativa individual dos elementos da amostra com uma sistematização quantitativa, por categorização, compatível com o volume de dados envolvidos. A Análise de Conteúdo dos sites será realizada a partir das homepages¹⁵ e conforme os seguintes critérios: (a) língua predominante no conteúdo do site; (b) tipo de ator social responsável pelo site; (c) tipo de site (conforme Bar-Ilan, 2004). Sites em língua diferente do país da ccTLD não serão considerados nas próximas etapas da observação, mas sua presença será levada em conta por ocasião da interpretação e discussão dos resultados.

A segunda aproximação de escala terá como objeto as páginas em que os links se encontram ou para as quais se dirigem. No caso de pares de links indo e vindo de/para várias páginas de um mesmo site, será dada preferência à referência de menor nível de profundidade, separadamente na origem e destino¹⁶. O grande número de páginas a

¹⁴ Em relação a .de, por exemplo, são 587 links (426 *outlinks* e 161 *inlinks*).

¹⁵ Em casos duvidosos ou fronteiriços, páginas adicionais serão avaliadas.

¹⁶ Por exemplo, em 3 pares (hipotéticos de links (1) de <http://www.cnpq.br/editais/bolsas> para <http://www.daad.de>; (2) de <http://www.cnpq.br> para <http://www.daad.de/deustchland> e (3) de <http://www.cnpq.br/editais/bolsas/alemanha>



serem observadas recomenda novamente a categorização por Análise de Conteúdo, que levará em conta os seguintes critérios: (a) língua predominante na página; (b) estrutura da página (Haas e Grams, 2000); (d) tipo de conteúdo da página (Bar-Ilan, 2004); (e) relação entre origem e destino (Bar-Ilan, 2004); (f) unidade entre o conteúdo da página e o tipo do site.

O terceiro nível de análise corresponde à observação dos próprios links, a ‘escala micro’ da Análise de Hiperlinks tradicional. Aqui, serão combinadas taxonomias específicas para a categorização de links e Análise de Conteúdo. Os parâmetros de categorização serão (a) a condição de produção do link (relativa ao tipo de sistema hipermídia, Bodner e Chignell, 1999; De Bra, Brusilovsky e Houben, 1999; Not e Zancanaro, 2000); (b) a linguagem da âncora de saída (Fragoso, 2001); (c) a explicitude da âncora de saída (adaptado de De Rose, 1989); (d) a função do link (Fragoso, 2001; Harrison, 2002; Bar-Ilan 2004, Obendorf e Weinreich, 2003).

CONSIDERAÇÕES FINAIS

A compreensão midiática da web orienta todas as etapas do exemplo apresentado, permitindo e articulando a operação em múltiplas escalas. O cruzamento final dos vários níveis de análise também será viabilizado pela compreensão integrada de todos os elementos envolvidos: os links são componentes das páginas que são componentes dos sites que foram deliberadamente criado por atores sociais geograficamente situados em relação aos Estados-Nação representados pelas ccTLDs. Como os resultados da pesquisa utilizada como exemplo ainda não são conhecidos, não é possível exemplificar a partir deles como o processo interpretativo será perpassado pelo reconhecimento das especificidades do próprio meio (McLuhan, 1998; 2002) e das mediações (Martín-Barbero, 2003) que o envolvem e atravessam.

É preciso estar atento para as limitações das técnicas adotadas, especialmente a Contagem de Links e a Análise de Conteúdo. Ambas são reducionistas, no sentido de que equiparam os fenômenos comunicacionais aos dados trabalháveis (Riffe, Lacy e Fico, 2005, p 23), com alto risco de perder de vista a complexidade da realidade empírica. Além disso, a operacionalização em categorias facilita a tradução das fronteiras estabelecidas entre os tipos e sub-tipos de dados para o fenômeno em si, impondo uma ordem que, ao final, pode esconder mais do que revela.

A arquitetura desse exemplo reflete nossa convicção de que a boa ciência requer um

para <http://www.daad.de/brazil>; serão analisadas as páginas correspondentes aos links 1 (destino mais superficial) e 2 (origem mais superficial), mas não as do par 3.



sólido arcabouço teórico-metodológico mas é um empreendimento criativo. Nessa concepção, o problema e os objetivos da pesquisa, as estratégias metodológicas e a interpretação dos resultados são sempre interdependentes. Por isso, os procedimentos adotados neste caso são um exemplo, mas não pretendem servir de modelos prontos para a replicação em futuras análises de hiperlinks norteadas pela visão comunicacional da web. Cada novo problema de pesquisa postula novos objetivos, que vão requerer estratégias de amostragem e análise específicas e produzirão resultados de natureza distinta, cujas interpretações precisam ser consistentes com as escolhas anteriormente realizadas. O que pretende ser perene é a proposta de utilizar os estudos de mídia como um arcabouço teórico para potencializar as Análises de Hiperlinks.

REFERÊNCIAS BIBLIOGRÁFICAS

ADAMIC, Lada. The Social Hyperlink. In. TUROW, Joseph; TSUI, Lokman. **The Hyperlinked Society: questioning connections in the digital age**. University of Michigan Press, 2008, pp. 227-249.

ADAMIC, Lada; ADAR, Eytan. You are what you link. **10th annual International World Wide Web Conference**, Hong Kong, 2001. Disponível em <<http://www10.org/program/society/yaww1/YouAreWhatYouLink.htm>>. Acesso em: 11 jan. 2009.

ALEXANDER, Jeffrey C. Action and Its Environments. In. ALEXANDER, Jeffrey C. et al. (Eds). **The Micro-Macro Link**. University of California Press, 1987, p. 289-318.

ALEXANDER, Jeffrey C.; GIESEN, Bernhard. From Reduction to Linkage: the long view of the micro-macro debate. In. ALEXANDER, Jeffrey C. et al. (Eds). **The Micro-Macro Link**. University of California Press, 1987, p. 1-44.

BARABASI, Albert-László. **Linked**. New York, Plume, 2002.

BAR-ILAN, Judit. The Use of Web Search Engines in Information Science Research. **Annual Review of Information Science and Technology (ARIST)**, v38, 2004, p. 231-88.

BARNETT, G. A., and SUNG, E. (2005). Culture and the structure of the international hyperlink network. **Journal of Computer-Mediated Communication**, vol. 11, n. 1, artigo 11. Disponível em: <<http://jcmc.indiana.edu/vol11/issue1/barnett.html>> Acesso em: 21 jun. 2009.

BARNETT, George e PARK, Han Woo. The Structure of International Internet Flows, **Proceedings of IR 4.0: broadening the band, 2003**. Association of Internet Researchers. Disponível em <<http://www.aoir.org/>> Acesso em: 21 mar. 2004. [acesso restrito].

BARNETT, George e SEUNG Joon Jun. An Examination of the Determinants of International Internet Structure. **Proceedings of IR 5.0: ubiquity?, 2004**. Association of Internet Researchers. Disponível em <<http://www.aoir.org/>> Acesso em: 13 set. 2005. [acesso restrito].

BERNERS-LEE, Tim. **Information Management: A Proposal**. CERN, Março de 1989. Disponível em: <<http://www.w3.org/History/1989/proposal.html>> Acesso em: 26 jun. 2009.

BERNERS-LEE, Tim. **The web: Past, Present and Future**, 1996. Disponível em <<http://www.w3.org/People/Berners-Lee/1996/ppf.html>> Acesso em: 26 jun. 2009.

BERNERS-LEE, Tim; Cailliau, Robert. **WorldWideWeb: Proposal for a HyperText Project**. 1990. Disponível <<http://www.w3.org/Proposal.html>> Acesso em: 23 jun. 2009.

BJÖRNEBORN, Lennart. **Small-World Link Structures across an Academic Web Space: A Library and Information Science Approach**. 2004. Tese de Ph.D. apresentada à Royal School



of Library and Information Science, Dinamarca. Disponível em <<http://www.db.dk/lb/phd>> Acesso em: 8 jun. 2009.

BODNER, Richard; CHIGNELL, Mark. Dynamic Hypertext: querying and linking. **ACM Computing Surveys** v. 31 n. 4, 1999. Disponível em <http://www.cs.brown.edu/memex/ACM_HypertextTestbed/papers/39.html> Acesso em: 26 jun. 2009.

BOLTER, Jay David. **Writing Space: the computer, hypertext and the history of writing**. Lawrence Erlbaum Associates, 1991.

BRIN, S. e L. PAGE, The Anatomy of a Large-Scale Hypertextual Web Search Engine, **Seventh International Conference on World Wide Web**, Brisbane, Australia, 1998. Disponível online em <<http://infolab.stanford.edu/~backrub/google.html>> Acesso em: 26 jun. 2009.

BUSH, Vannevar. As we may think. **The Atlantic Monthly**, Julho de 1945. Disponível em <<http://www.theatlantic.com/doc/194507/bush?ca=VPDFbQ%2BoWxp52HBQmdMjcm0HoF5tJ4jht1rUKgmY6S4%3D>> Acesso em: 25 fev. 2008.

DE BRA, Paul; BRUSILOVSKY, Peter, HOUBEN, Geert-Jan. Adaptive Hypermedia: From Systems to Framework. **ACM Computing Surveys** v. 31 n. 4, 1999. Disponível em <http://www.cs.brown.edu/memex/ACM_HypertextTestbed/papers/25.html> Acesso em: 26 jun. 2009.

DE ROSE, Steven. Expanding the notion of links. **Proceedings of the second annual ACM conference on Hypertext**. Pennsylvania, 1989, p. 249 – 257. Disponível em <<http://portal.acm.org/citation.cfm?doi=74224.74245>> Acesso em: 26 jun. 2009.

FRAGOSO, Suely . Sampling the Web: discussing strategies for the selection of Brazilian websites for quanti-qualitative analysis. In: Consalvo, Mia; Haythornthwaite, Caroline. (Eds.). **AoIR Internet Research Annual** v. 4. New York: Peter Lang, 2006, p. 195-208.

FRAGOSO, Suely . What we give and what you get. **IR 8.0 Let's Play**, 8th International Conference of the Association of Internet Researchers, AoIR. Vancouver, 2007. Disponível em < <http://docs.oir.org/members/ir8-fragoso.pdf>> Acesso em 27 jun. 2009. [acesso restrito]

FRAGOSO, Suely. **Hipersemioses: proposta para a incorporação de elementos tridimensionais a sistemas hipermídia**. Relatório Final de Pesquisa apresentado ao Centro de Ciências da Comunicação da Unisinos, 2001. Inédito.

FRAGOSO, Suely. Mapping Brazil's connectivity - do we really get more than we give?. **IR 6.0 - Internet Generations**, 5th International Conference of the Association of Internet Researchers, AoIR. Chicago, 2005. Disponível em < <http://docs.oir.org/members/ir6-fragoso.pdf>> Acesso em 27 jun. 2009. [acesso restrito]

GERSTEIN, Dean. To Unpack micro and macro: link small with large and part with whole. In. Alexander, Jeffrey C. et al. (Eds). **The Micro-Macro Link**. University of California Press, 1987, p. 86-111.

GULLI, Antonio; SIGNORINI, Alessio. The Indexable Web is more than 11.5 billion pages. **International Conference on the WWW 2005**, 10 a 14 de Maio, 2005, Chiba, Japão. Disponível em <<http://www.cs.uiowa.edu/~assignori/web-size/size-indexable-web.pdf>> Acesso em: 25 dez. 2006.

HAAS, Stephanie W.; GRAMS, Erika S. Page and Link Classifications: connecting diverse resources. **Proceedings of the third ACM conference on Digital libraries**, Pennsylvania, 1998. Disponível em: <<http://portal.acm.org/citation.cfm?id=276675.276686>> Acesso em: 26 jun. 2009.

HAAS, Stephanie W.; GRAMS, Erika S. Readers, authors, and page structure: A discussion of four questions arising from content analysis of Web pages. **Journal of the American Society for Information Science** v. 51, n. 2, 2000, p. 181-192.



- HALAVAIS, Alexander e GARRIDO, Maria. Mapping Networks of Support for the Zapatista Movement: Applying Social Network Analysis to Study Contemporary Social Movements. In McCaughey, Martha e Ayers, Michael D. (eds). **Cyberactivism: Online Activism in Theory and Practice**. New York: Routledge, 2003, pp. 165-184.
- HALAVAIS, Alexander. The Hyperlink as Organizing Principle. In. Turow, Joseph; Tsui, Lokman. **The Hyperlinked Society: questioning connections in the digital age**. University of Michigan Press, 2008, pp. 39-55.
- HARRISON, Claire. Hypertext Links: Whither Thou Goest, and Why. **First Monday**, v. 7, n. 10, Outubro 2002. Disponível em http://firstmonday.org/issues/issue7_10/harrison/index.html Acesso em: 11 jan. 2009.
- HENZINGER, Monika. Link Analysis in Web Information Retrieval. **Proceedings of the sixteenth ACM conference on Hypertext and hypermedia**. Austria, 2005. Disponível em <http://portal.acm.org/citation.cfm?id=1083357> Acesso em: 9 jun. 2009. [acesso restrito]
- KLEINBERG, Jon M. Authoritative Sources in a Hyperlinked Environment. **Journal of the ACM (JACM)** v. 46 n. 5, 1999, p. 604-632.
- KUMAR, Ravi *et al.* The Web as a graph. PODS '00: **Proceedings of the nineteenth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems**. Maio de 2001. Disponível em <http://www.cs.brown.edu/research/webagent/pods-2000.pdf> Acesso em: 11 jan. 2009.
- MARTÍN-BARBERO, Jesus. **Dos Meios às Mediações: comunicação, cultura e hegemonia**. Rio de Janeiro: Editora UFRJ, 2003.
- MAULDIN, Michael L., Lycos: Design choices in an Internet search service. **IEEE Expert**, Jan-Fev 1997, p. 8-11. Disponível em <http://www.fuzine.com/liti/pub/ieee97.html> Acesso em 10 jan. 2007.
- MCLUHAN, Marshall. **The Gutenberg Galaxy: the making of typographic man**. University of Toronto Press, 2002.
- MCLUHAN, Marshall. **Understanding Media: the extensions of man**. MIT Press, 1998.
- MODESTO, Marco *et al.* Um Novo Retrato da Web Brasileira. **Anais do Seminário Integrado de Software e Hardware SEMISH XXXII** São Leopoldo, 2005. Disponível em <http://200.18.98.97/intranet/documentos/papers/csbc2005/pdf/arq0163.pdf> Acesso em: 3 set. 2005.
- NOT, Elena; ZANCANARO, Massimo. The MacroNode Approach: mediating between adaptive and dynamic hypermedia. **Lecture Notes In Computer Science; v. 1892 Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems**, Springer-Verlag, 2000, p. 167-178, p. 167 – 178. Disponível em <http://portal.acm.org/citation.cfm?id=727902> Acesso em: 26 jun. 2009. [acesso restrito]
- OBENDORF, Hartmut; WEINREICH, Harald. Comparing Link Marker Visualization Techniques – Changes in Reading Behavior. **Proceedings of the 12th International World Wide Web Conference - WWW 2003**, Budapeste, 2003, p. 736-745. Disponível em <http://vsis-www.informatik.uni-hamburg.de/getDoc.php/publications/119/www2003.pdf> Acesso em: 9 jun. 2009.
- PARK, Han Woo. Hyperlink Network Analysis: A New Method for the Study of Social Structure on the Web. **Connections** v. 25, n.1, p. 49-61, 2003.
- PINTO, Solon L; FERNANDES, Ciro C.C. Institucionalização do governo eletrônico: o caso do Brasil. **X Congresso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública**, Santiago, Out. 2005. Disponível em <http://www.clad.org.ve/fulltext/0052752.pdf> Acesso em: 11 jan. 2009.
- PRIMO, Alex; RECUERO, Raquel; ARAÚJO, Ricardo Matsumura de . Co-Links: Proposta de uma nova tecnologia para a escrita coletiva de links multidirecionais. **Revista Fronteira**, São Leopoldo, v. VI, p. 91-113, 2004.



PRIOR, Markus Are Hyperlinks ‘Weak Ties’? In. Turow, Joseph; Tsui, Lokman. **The Hyperlinked Society: questioning connections in the digital age**. University of Michigan Press, 2008, pp. 227-249.

RECUERO, Raquel. **Redes Sociais na Internet**. Porto Alegre: Sulina, 2009.

REGISTRO.BR. **Registro.br – Estatísticas**. Domínios Registrados por DPN - 07/05/2005 23:05:56. 2005. Disponível em <<http://registro.br/estatisticas.html>> Acesso em: 8 mai. 2005.

REGISTRO.BR. **Registro.br – Estatísticas**. Domínios Registrados por DPN - 25/06/2009 00:00:00. 2009. Disponível em <<http://registro.br/estatisticas.html>> Acesso em: 25 jun. 2009.

RIFFE, Daniel; LACY, Stephen; FICO, Frederick G. **Analyzing media messages: using quantitative content analysis in research**. Lawrence Erlbaum Associates, 2005.

ROUSSEAU, Ronald. Situations: an exploratory study. **Cybermetrics, International Journal of Scientometrics, Informetrics and Bibliometrics**. V. 1, n.1. 1997. Disponível em <<http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html>> Acesso em: 17 jan. 2009.

SMITH, A.G. Web links as analogues of citations. **Information Research**, v. 9, n. 4, paper 188, 2004. Disponível em <<http://InformationR.net/ir/9-4/paper188.html>> Acesso em: 17 jan. 2009.

THELWALL, Mike. (2003). What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation. **Information Research**, v. 8, n.3, paper no. 151, 2003. Disponível em <<http://informationr.net/ir/8-3/paper151.html>> Acesso em: 11 jan. 2009.

THELWALL, Mike. Interpreting social science link analysis research: A theoretical framework. **Journal of the American Society for Information Science and Technology**, vol 57 n. 1, 2006, p. 60-68.

THELWALL, Mike. **Introduction to Webometrics: Quantitative Web Research for the Social Sciences**. Morgan & Claypool Publishers, 2009.

TOMASINI, Ana Cláudia Valente. **Globalização e Nacionalismo: um estudo de caso da comunidade virtual "Brasil" sobre a identidade brasileira no Orkut**. 2006. Dissertação de Mestrado apresentada ao Programa de Pós Graduação em Ciências da Comunicação da Unisinos, São Leopoldo, RS.

VAUGHAN, Liwen; THELWALL, Mike. Search engine coverage bias: evidence and possible causes. **Information Processing & Management** v. 40, n. 4, 2004, p. 693-707.

VELOSO, Eveline et al. Um retrato da web brasileira, **Anais do XXI Seminário Integrado de Hardware e Software SEMISH 00**, Curitiba, 2000. Disponível em <<http://stat.akwan.com.br/~golgher/semish00.ps.gz>> Acesso em: 15 jul. 2005.

W3C, **W3C Glossary and Dictionary**. Disponível em <<http://www.w3.org/2003/glossary/>> Acesso em 26 jun. 2009.

WALLERSTEIN, Immanuel. **El Moderno Sistema Mundial**. Madrid, Siglo Veintiuno Editores, 1979.

WEINREICH, Harald; OBENDORF, Hartmut; LAMERSDORF, Winfried. The Look of the Link - Concepts for the User Interface of Extended Hyperlinks **Proceedings of the 12th ACM conference on Hypertext and Hypermedia**. Dinamarca, 2001. Disponível em: <<http://www.obendorf.de/studium/projekte/lookoflink-ht01.pdf>> Acesso em: 9 jun. 2009.