



Cibergeografia Midiática: proposta de confluência de quatro abordagens quantitativas com vistas à construção de uma metodologia quanti-qualitativa de análise da World Wide Web^{*} ^{}**

Suely Fragoso^{***}

Unisinos

Resumo

O presente texto revisa quatro abordagens quantitativas do ciberespaço (cibergeografia, teoria matemática das redes, análise de hiperlinks e webometria) e propõe que a confluência de pressupostos e técnicas das mesmas, quando orientada pelo reconhecimento do caráter midiático da Web, pode resultar numa metodologia extremamente potente para apoiar empiricamente investigações do ciberespaço.

Palavras-chave

World Wide Web; Cibergeografia; Análise de Hiperlinks, Teoria Matemática das Redes, Webometria

O caráter dinâmico, a extensão e a heterogeneidade das redes digitais de comunicação dificultam bastante as investigações científicas da Internet. Sobretudo quando se trata de trabalhos empíricos, que demandam a construção de amostragens representativas e a seleção ou o desenvolvimento de metodologias de pesquisa suficientemente robustas, o ciberespaço se oferece como um investigável particularmente arisco. Diante de um conjunto de impasses que em muitos casos parecem ser intransponíveis, as soluções têm oscilado entre o desenvolvimento de pesquisas estritamente quantitativas e a escolha de observáveis notadamente anedóticos. Prestando tributo à importante gama de investigações que têm conseguido evitar as armadilhas situadas em cada um desses extremos e procurando contribuir para a longevidade das mesmas, este texto se propõe a explicitar e discutir quatro tipos de abordagem quantitativa do ciberespaço considerados particularmente produtivos: a Cibergeografia (que alguns textos equiparam à Geografia Virtual), a Teoria Matemática das Redes, a Análise de Hiperlinks e a Webometria. O grande motivador desta empreitada é a antevisão de que, se orientada pelo

^{*} Trabalho apresentado ao NP 08 - Tecnologias da Informação e da Comunicação, do V Encontro dos Núcleos de Pesquisa da Intercom.

^{**} Este trabalho é resultado parcial de uma pesquisa apoiada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq.

^{***} Professora Titular e Pesquisadora do Programa de Pós-Graduação em Ciências da Comunicação desde 1999, atualmente também Coordenadora Executiva do Programa. Pesquisadora CNPq. Ph.D. em Comunicação (*Institute of Communications Studies, The University of Leeds*, Inglaterra) e Mestre em Comunicação e Semiótica (PUC/SP). <suely@unisinos.br>



reconhecimento do caráter midiático da *World Wide Web*, a confluência de pressupostos e técnicas daquelas linhas de investigação pode resultar num arcabouço teórico-metodológico extremamente potente para os estudos do ciberespaço.

Introdução: Cibergeografia

Até recentemente, as análises e discussões sobre o impacto das tecnologias digitais da comunicação (TICs) na organização socioeconômica e cultural mundial caracterizaram-se pela polarização. Ainda na década de 1990, as óbvias diferenças entre entusiastas e apocalípticos chegavam a obscurecer as fragilidades da visão que os irmanava: para ambos, o futuro acenava com um planeta homogeneamente enredado pelos fios e cabos das redes digitais e, conseqüentemente, com o espraiamento dos padrões e valores por elas incorporados, até a completa erosão das singularidades e diferenças regionais. O final dos anos 90 viu ruir o determinismo de tais abordagens, que tendiam a compreender as intrincadas relações e interdependências entre o chamado ciberespaço¹ e o mundo 'fora' das redes digitais em termos do impacto das forças daquele ('virtual') sobre a fragilidade deste ('real').

A partir do reconhecimento das disparidades na disponibilidade de acesso e das diferentes escalas e padrões de utilização das TICs ao redor do mundo, constata-se uma distribuição francamente heterogênea da infra-estrutura de suporte à comunicação digital, inclusive dos terminais e aplicativos essenciais para o acesso. Já era sabido que a grandeza dos números absolutos de usuários e os indicadores que apontam o crescimento acelerado das redes digitais escondiam o fato de que as TICs constituem privilégio de uma pequeníssima proporção de pessoas ao redor do globo². Não surpreende, também, que os padrões de acesso às TICs reflitam as disparidades econômicas entre os continentes (com óbvia concentração do acesso às redes digitais de comunicação nos países do hemisfério norte) e, em escalas mais próximas, também reiterem as diferenças em desempenho econômico dos diferentes países de cada continente e das diferentes regiões de cada país³ (Fragoso, 2004, p. 110-111). De fato, a

1 A palavra ciberespaço é utilizada aqui para denominar o conjunto de informações codificadas binariamente que transita em circuitos digitais e redes de transmissão, inclusive sem fio. A efetiva espacialidade desse construto, que o presente texto toma por pressuposta, foi discutida em trabalhos anteriores, particularmente Fragoso, 2001 e Fragoso, 2003.

2 Em números absolutos, estima-se que a quantidade de pessoas com acesso às redes tenha saltado de cerca de 21 milhões em 1995 (NUA, 2003, s.p.) para espantosos 934 milhões em 2004 (*ClicZ Network*, 2005, s.p.). Proporcionalmente, no entanto, trata-se ainda de menos de 15% da população total do planeta, estimada em 6.4 bilhões de pessoas em meados de 2005 (IBGE, 2005, s.p.).

3 No Brasil, por exemplo, dados de 2003 (*Mapa da Exclusão Digital*) apontam para 13,5% e 10,41% dos domicílios com computadores e acesso à Internet nas regiões Sul e Sudeste (respectivamente) contra apenas 3,68% na região



proporcionalidade entre a abundância quantitativa e qualitativa de infra-estrutura para acesso às redes digitais de comunicação e o poder econômico dos diferentes continentes, nações e regiões é inegável. Ocorre que, apesar da intangibilidade do espaço informacional perpetrado pela circulação de dados nas redes telemáticas, as redes digitais de comunicação permanecem ancoradas a diversas instâncias do mundo físico, sendo imprescindível atentar para a interferência do espaço geográfico 'tradicional' sobre os fluxos informacionais – e vice-versa.

Dado que “os principais centros metropolitanos em todo o mundo continuam a acumular fatores indutores de inovação e a gerar sinergia na indústria e serviços avançados” (Castells, 1999, p. 416), cai por terra a suposição de que, com a compressão espaço-temporal, na era das comunicações digitais 'a geografia não mais importa'⁴. Afinal, “existe um mundo entre as TICs e o ciberespaço na forma de outras infra-estruturas, redes sociais face-a-face, trabalhadores capacitados, acesso a materiais e mercados globais e locais” (Dodge e Kitchin, 2001, p. 14) que impede o mero desaparecimento da 'antiga geopolítica'. Embora possa tender a tornar-se menos polarizada entre os hemisférios Norte e o Sul do planeta, a existência de desigualdades locais e regionais tende a persistir graças à acentuação do desenvolvimento dos “segmentos e territórios dinâmicos das sociedades em todos os lugares” *versus* “aqueles que correm o risco de tornar-se não pertinentes sob a perspectiva da lógica do sistema” (Castells, 1999, p. 22).

Voltada para a problemática das injunções do espaço geográfico tradicional sobre o ciberespaço e vice-versa, a cibergeografia tem se concentrado sobre a distribuição espacial da infra-estrutura que garante a existência do ciberespaço e sobre o tráfego de informação entre servidores. Os resultados dessa linha de investigações vêm demonstrando, de forma bastante consistente, que as redes digitais de comunicação operam como variáveis geográficas complementares, capazes de afetar o processo de configuração das “novas centralidades e novas periferias” (Cunha Filho, 2003) no mundo contemporâneo conforme reiterem ou contradigam as configurações anteriormente em vigor (independente de as mesmas terem sido resultantes de ações políticas ou de acidentes geológicos).

A World Wide Web e a Teoria Matemática das Redes

Norte. A média nacional estimada por aquele estudo foi de 8% das casas com computadores e acesso à Internet (GAID-FGV, 2003, pp. 43-45.)

⁴ Afirmações que tais marcaram um período das análises do ciberespaço e podem ser encontradas em diversos autores que discutem o impacto sociocultural das tecnologias digitais de informação e comunicação, sobretudo na década de 1990 (por exemplo Lévy, 1998 e 1999; Trivinho, 1998 e Virilio, 1999).



Em paralelo à intersecção entre o ciberespaço e o mundo físico, que constitui o objeto mais freqüentemente abordado pela cibergeografia, uma outra configuração espacial aguarda para ser confrontada com a geografia tradicional. Trata-se da espacialidade instituída pelas relações entre os vários elementos que constituem o ciberespaço, e que pode ser traduzida graficamente em mapas dos espaços informacionais propriamente ditos. Embora a complexidade espacial das relações estruturantes do ciberespaço convide à elaboração de gráficos com um número cada vez maior de variáveis e dimensões (mapas dinâmicos, tridimensionais, hiperbólicos, etc), a espacialidade de frações específicas dos diferentes tipos de espaço informacional envolvidos na configuração do ciberespaço pode, muitas vezes, ser apreendida com mapeamentos relativamente simples, inclusive os bidimensionais e estáticos. Muitos e distintos tipos de elementos (caracterizados por diferentes modos de articular as unidades informacionais) compõem tipos diversos de espaços informacionais (WWW, WAIS, telnet, email, etc) os quais, em conjunto, instituem o ciberespaço. Dentre eles, a *World Wide Web* se oferece, com máxima proeminência, ao mapeamento.

A criação da WWW foi a principal alavanca do crescimento exponencial do ciberespaço (e do acesso a ele) durante a década de 90. Incorporando o conceito de hipertexto a um sistema de domínio público, o projeto da Web garantiu não apenas a viabilização da transferência de arquivos entre computadores com sistemas operacionais incompatíveis (característica fundamental do projeto original, cf. Berners-Lee, 1996, s.p) mas também a integração de texto, imagem e som a documentos inter-relacionados por hiperconexões possíveis de ‘percorrer’ com o apoio de navegadores⁵. Impulsionada pela convergência de vários fatores de indução (tecnológicos, políticos, econômicos e culturais), a utilização da Web aumentou vertiginosamente em poucos anos, introduzindo para milhões de pessoas ao redor do mundo as práticas da comunicação mediada por computador e a hipermídia. Atualmente, a grande maioria dos usuários da Internet acessa apenas a Web e a identifica genericamente como sendo ‘o’ ciberespaço.

A espacialidade da WWW é instituída pelas hiperconexões (Fragoso, 2001) entre elementos com dimensionalidades diversas: ‘páginas’ estritamente bidimensionais se conectam a outras nas quais textos convivem com animações e seqüências de vídeo e com representações em perspectiva de modelos tridimensionais e assim por diante. Do

⁵ O primeiro navegador, o *Mosaic*, foi desenvolvido pela equipe do *National Center for Supercomputing Applications* da *University of Illinois* e disponibilizado pela universidade em 1993. Sua interface gráfica altamente intuitiva foi fundamental para a popularização da Web Web.

lado de cá da tela, o usuário percorre com os olhos a superfície na qual as diferentes páginas são enunciadas separadamente (uma por janela do *browser*) e seleciona links que deslocam aquela parcela do ciberespaço relativamente a seu observador. Uma vez que os caminhos percorridos na WWW fazem parte de uma estrutura espacial particularmente complexa, ao contrário do que ocorre com as travessias às cegas no mundo físico (por exemplo em elevadores ou veículos sem janelas), é bastante difícil para o usuário prever os destinos de links semelhantes, ou os vários pontos de chegada disponíveis a partir de cada página. O desnorteamento daí decorrente é uma das forças impulsionadoras do desenvolvimento de diversas estratégias para o mapeamento de web sites, que variam desde singelas realizações artesanais, planas e bidimensionais até as mais elaboradas ferramentas para geração de mapas de sites tridimensionais e dinâmicos⁶.

Mesmo porque em sua maioria concebidos como ferramentas para auxiliar a navegação, grande parte dos mapas de websites baseiam-se em analogias com espacialidades conhecidas, por exemplo o espaço tridimensional euclidiano, a crosta terrestre ou o espaço sideral (Figura 1). Entre as exceções, destacam-se alguns dos mapeadores concebidos para ajudar os *webmasters* na manutenção dos sites. Ao concentrar-se sobre as relações entre os diferentes documentos de um dado site e não sobre o conteúdo dos mesmos, muitos desses aplicativos colocam a ênfase nos elementos em função e a partir dos quais realmente se institui a espacialidade da Web: os links. Os mapas resultantes não raro revelam as irregularidades da essência relacional do ciberespaço na mesma medida em que se afastam de noções anteriores de espaço: caem por terra, no processo, tanto concepções mais conservadoras, que antevêm na Web um grande espaço tridimensional de cunho euclidiano, quanto as atraentes proposições de espaços rizomáticos onde ‘tudo se conecta com tudo’⁷ de forma essencialmente aleatória.

Ainda em 1997, quando era comum encontrar autores que consideravam a Internet uma rede descentralizada e igualitária, na qual nenhum nó poderia ser considerado central ou principal, um trabalho seminal de Jakob Nielsen alertava para o fato de que o acesso à

6 Um conjunto de exemplos variados e que ainda não perderam sua atualidade é descrito em Dodge e Kitchin, 2001, pp. 113-124. Parte deste material encontra-se também no website *An Atlas of Cyberspace*, disponível em <http://www.cybergeography.org/atlas/atlas.htm> [jun 2005]

7 A metáfora do rizoma teria sido proposta por Deleuze e Guatarri como um exemplo de estrutura natural em que, ao contrário do que ocorre com as raízes [e caules, troncos, ramos], pontos quaisquer conectam-se a outros, independente inclusive de serem de naturezas distintas. Aplicada ao hipertexto digital por Lévy, a imagem passou a ser evocada principalmente em discursos laudatórios do potencial libertador e anárquico das tecnologias, em especial a Internet. (Primo e Recuero, 2004, s.p.).



Web se dá conforme uma distribuição do tipo Zipf⁸. O autor intuía, à época, que também seguiriam uma distribuição Zipf as participações em grupos da Usenet (poucas pessoas postam a maioria das mensagens e muitas raramente se manifestam) e os links entre websites (“tanto no sentido de que existem alguns poucos sites para os quais todo mundo envia links e muitos sites com os quais quase ninguém estabelece links” quanto no sentido de que “a maior parte das visitas a qualquer dado site chegam a partir de links enviados por alguns poucos outros sites, enquanto pequenas quantidades de tráfego vêm de uma grande variedade de sites” cf. Nielsen, 1997). A confirmação empírica (e em várias frentes) desta hipótese veio confirmar a suspeita de que as redes digitais de comunicação não estão constituídas de forma aleatória. Afinal, se as conexões entre as diferentes páginas da Web (os nós da rede Web) fossem estabelecidas aleatoriamente, a projeção em gráfico do número de links de um conjunto de Websites suficientemente grande e diversificado para ser representativo deveria resultar em uma curva em forma de sino (a maioria dos sites teria um número médio de links, enquanto alguns poucos teriam muitos links e outro pequeno grupo teria poucos links). Uma tal distribuição implicaria, por outro lado, que a maior parte dos sites tivesse mais ou menos o mesmo número de links. Há indicadores de que, de fato, a distribuição de links na Web segue uma distribuição do tipo Zipf, o que se traduz em um pequeno número de páginas com um número muito elevado de links, um número médio de páginas com um número médio de links e um grande número de páginas com pouquíssimos links. (Barabási, 2002, p 70).

A distribuição exponencial da curva que representa a distribuição dos links por website evidencia a não aplicabilidade de alguns pressupostos do modelo aleatório das redes à Web. Em primeiro lugar, a Web é uma rede em desenvolvimento, isto é, novos nós continuam a ser adicionados a todo instante. Isoladamente este fator implicaria, no entanto, que os sites mais antigos teriam sempre maior número de links que os sites mais novos, pois numa rede em desenvolvimento verdadeiramente aleatória, todos os nós teriam igual oportunidade de receber um link a cada momento, o que implicaria em vantagem para os nós mais antigos (cuja longevidade garantiria a presença em um maior número de eventos de distribuição de links). Não é difícil perceber, no entanto, que este não é o caso na WWW: basta pensar no grande número de links que apontam para sites

⁸ Aplicável a uma série de fenômenos, a distribuição é denominada Zipf em homenagem a George Kingsley Zipf (1902-1950). Trata-se de uma distribuição em que a frequência de ocorrência de um evento (P), como função do escalão (i), quando (i) é determinado pela mesma frequência de ocorrência, é uma função exponencial $P_i \sim 1/i^a$ em que o expoente a se aproxima da unidade.

relativamente jovens, como o *Google*, em oposição a uma quantidade certamente mais modesta de conexões para o mais antigo nó da Web, no CERN.

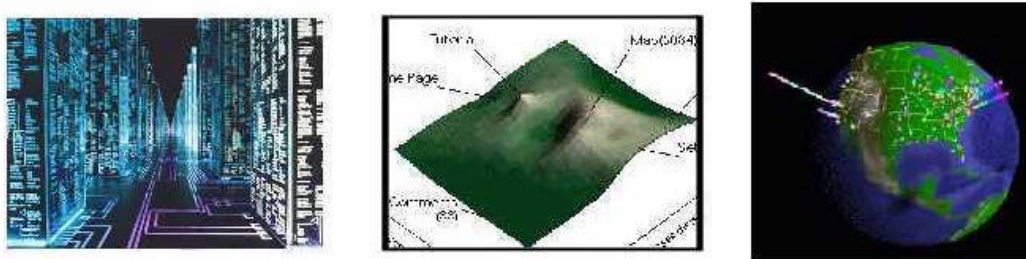


Figura 1: Exemplos de mapas do ciberespaço que lançam mão de analogias com noções conhecidas de espaço: à esquerda, ‘Cidade de Texto’, uma concepção ficcional (do filme *Hackers*, dirigido por Iain Softley, 1995). Ao centro, projeção topográfica dos padrões de acesso a um website (*Alexandria Digital*), por Sara I. Fabrikant, 2001. À direita, representação gráfica do tráfego da WWW em tempo real, por Stephen E. Lamm e Daniel A. Reed, 1996. Os mapas da esquerda e da direita foram reproduzidos desde o site *An Atlas of Cyberspace* (<http://www.cybergeography.org/atlas/atlas.htm>). O mapa central foi reproduzido de *Spatialization...* (<http://www.geog.ucsb.edu/%7Esara/html/research/diss/spatialization.html>)

Ocorre que a Web não é uma estrutura igualitária: seus diversos nós (páginas) recebem valorações distintas: “quando escolhem entre duas páginas, uma com duas vezes mais links que a outra, quase o dobro das pessoas linkam para a página mais conectada” (Barabási, 2002, p. 85). Estabelece-se assim um padrão preferencial de conectibilidade, que ajuda a compreender porque “a arquitetura da Web é dominada por alguns poucos nós altamente conectados”. Esses sites que abrigam uma grande quantidade de conexões, chamados conectores ou *hubs*, são extremamente visíveis. A esse respeito, Barabási afirma: “[p]ara todos os propósitos práticos, páginas linkadas por um ou dois documentos não existem. É quase impossível encontrá-las” (Barabási, 2002, p. 58). É preciso ressaltar que, evidentemente, tal colocação desconsidera o fato que constitui o ponto de partida do presente texto, a saber: que as redes digitais de comunicação (incluída aí a *World Wide Web*) não existem isoladas do mundo dito ‘real’. Assim como as múltiplas interferências dos espaços informacionais sobre a vida social e cultural não podem ser desprezadas, também não se pode deixar de levar em conta as ingerências da vida ‘fora da rede’ sobre o ciberespaço. Ao voltar os olhos apenas para a World Wide Web propriamente dita, isolando-a de seu contexto de existência e apreensão, Barabási deixa escapar algumas evidências um tanto óbvias, por exemplo a de que os websites não recebem indicações apenas de outros sites, a partir de links, mas também de outras fontes, midiáticas ou não, muitas delas externas à WWW. Enquanto permanecer garantida aos usuários a possibilidade de digitar endereços (URLs) e acessar



diretamente qualquer documento na Web, será sempre elevado o nível de perturbação que o mundo ‘do lado de cá da tela’ haverá de impor à ordem um tanto cristalina que as teorias matemáticas procuram encontrar na rede.

Não deixa de ser verdadeiro, entretanto, que especificidades relativas ao tipo de hiperconexão possível de estabelecer entre webpages contribuem para o estabelecimento de hierarquias na WWW. Segundo Barabási, o fato de que os links possíveis de estabelecer na Web como a conhecemos hoje são unidirecionais (ou seja, estabelecem uma única mão de direção na qual é possível prosseguir) implica em sua necessária fragmentação: “[r]edes direcionadas como a *World Wide Web* se repartem naturalmente em alguns continentes facilmente identificáveis” O primeiro desses continentes, chamado Núcleo Central, abrigaria todos os grandes sites (*Yahoo!*, *CNN.com*, *Amazon.com* etc). A “característica mais marcante” do Núcleo Central

é que ele é facilmente navegável, pois há um caminho entre quaisquer dois documentos que pertençam a ele. Isto não significa que haja um link direto entre quaisquer dois nós do núcleo central, mas que há [pelo menos] um caminho através de nós pertencentes ao núcleo que permite *surf*ar entre quaisquer dois nós (Barabási, 2002, p. 167).

Outros dois principais continentes, chamados de *In* e *Out*, caracterizam-se pela existência de 'mãos de direção':

[a] partir das páginas do continente *In* é possível chegar ao núcleo central, mas uma vez no núcleo não há caminhos que levem de volta ao continente *In*. Em contraste, os nós que pertencem ao continente *Out* podem ser facilmente atingidos a partir do núcleo central, mas uma vez fora do núcleo não há caminhos de retorno (Barabási, 2002, p. 167).

O quarto continente é um arquipélago de pequenas ilhas desconectadas, grupos isolados de páginas interconectadas que são impossíveis de alcançar desde o Núcleo Central e não possuem links que levem a ele. (Barabási, 2002, p. 166)

Decorre que a acessibilidade de uma *webpage* depende em grande parte de sua localização 'continental' na Web: uma página localizada no Núcleo Central tem muito mais chance de ser alcançada *a partir de hiperconexões* do que uma outra, situada no continente *Out*. Páginas do continente *In* e dos 'arquipélagos', por sua vez, só serão acessíveis para quem souber seu endereço: apenas digitando a URL é possível atingi-las. Em outras palavras, embora seja verdadeiro que qualquer um (que disponha de aparatos e alfabetidade tecnológicos suficientes) pode publicar na WWW, é preciso lembrar que não apenas a profusão de informações disponíveis nas redes digitais garante a manutenção da diferença entre publicar e ser visto ou ouvido (Vaz, 2000, s.p.; Serra, s.d., s.p.), mas também a topologia e estrutura das redes influenciam a acessibilidade de um documento na Web.

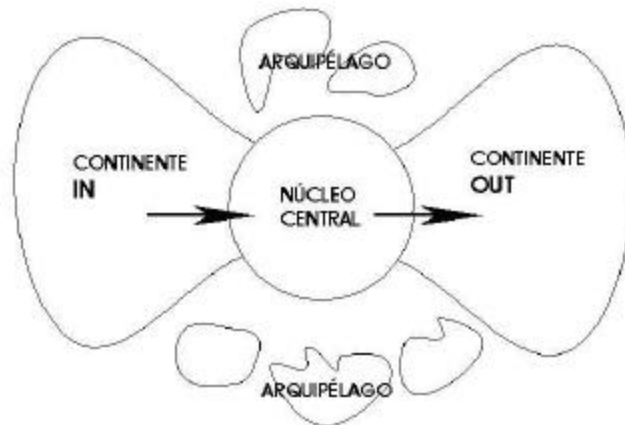


Figura 2: Representação esquemática dos principais 'continentes' em que se dividem as redes direcionais (conforme Barabási, 2002, p. 166).

Com isso, o estabelecimento de uma hiperconexão implica como que um 'endosso' da página de destino pelo site que abriga o link de saída (*outlink*): quando coloco um *outlink* para a Intercom em minha página pessoal *crio valor* para o site da Intercom pois o demarco como um destino que considero pertinente para o público de minha página Web. É evidente que o valor que uma dada página é capaz de agregar a outra pelo estabelecimento de um link será proporcional ao valor da própria página que abriga o *outlink*: uma conexão na primeira página do *Yahoo!* (que recebe uma grande quantidade de visitantes por dia) agrega mais valor ao site da Intercom do que uma conexão em minha página pessoal (que passa dias sem receber uma só visita). É também evidente que a maior visibilidade do *Yahoo!* faz com que ele seja alvo de muito mais links de chegada (*inlinks*), ofertados por outros sites, do que minha página pessoal: assim, a visibilidade do *Yahoo!* aumenta muito mais rapidamente que a de minha página pessoal e a diferença entre a visibilidade daquele portal e a minha página pessoal tende sempre a aumentar. Desenvolve-se assim, na Web, uma economia peculiar, cuja unidade simbólica de valor é o hiperlink. “Com esse valor [quase] econômico, eles [os links] têm também poder, afetando a acessibilidade e o conhecimento na Web” (Walker, 2002). Caracteriza-se assim uma política de linkagem em que “a criação de um link para além de um site seria uma concessão intencional de poder. Ou seja, não se trata apenas do exercício de uma possibilidade tecnológica, mas também de um ato de solidariedade com o criador da página-destino...” (Primo e Recuero, 2004, s.p.) Nos termos de Walker, quando eu estabeleço um link da minha página A para uma outra página B eu



estou *dando um link* para B. Esse link se traduz em um valor preciso em diversos sistemas de indexação, e portanto o “link tem um valor mais claro para B do que o conteúdo da página B tem para mim ou para meus leitores. Eu pago B pelo conteúdo de B com o meu link” (Walker, 2002, s.p.)

Ferramentas de Busca, Análise de Hiperlinks e Webometria

Herdeiros dos antigos indexadores *Archie*, *Gopher* e *Veronica*, os sistemas de busca baseiam-se largamente em *bots*⁹ que percorrem a Web saltando de um site para o outro e coletando dados sobre as páginas que encontram. A primeira ferramenta possível de ser assim definida teria sido o *World Wide Web Wanderer*, que entrou em operação em junho de 1993. Concebido para medir o crescimento da Web, o *Wanderer* não permitia que os usuários acrescentassem páginas a seu banco de dados. Esta lacuna logo seria suprida pelo *ALIWEB*, do mesmo ano, que permitiu também que os usuários acrescentassem uma breve descrição aos registros das páginas. Antes do final de 1993, novas ferramentas refinaram a busca na Web lançando mão de análises estatísticas de ocorrências de palavras e organizando os resultados encontrados conforme critérios de relevância. A maior parte delas armazenava apenas o título, o cabeçalho e/ou a URL das páginas que encontravam, mas as primeiras ferramentas que armazenavam o conteúdo completo das páginas não demorariam a surgir¹⁰.

Concebido no final da década de 90, o *Google* introduziu uma profunda alteração no paradigma das buscas na web. Pioneiro na incorporação de técnicas de análise de links para realizar não só as buscas, mas principalmente a hierarquização de resultados, o *Google* calcula uma classificação para cada página que encontra baseando-se não apenas no conteúdo dos documentos, mas sobretudo em sua posição na estrutura de links da Web. O procedimento é realizado conforme um algoritmo sugestivamente denominado *PageRank*:

O *PageRank* pode ser pensado como um modelo de comportamento de usuário. Nós assumimos que existe um ‘internauta randômico’ que recebe uma página aleatória e continua clicando em links, nunca selecionando ‘voltar’, mas que eventualmente se aborrece e recomeça em outra página aleatória. A probabilidade que o ‘internauta randômico’ visite uma dada página é o *PageRank* dela.

...

9 As versões para ferramentas de busca são também conhecidas como *spiders* ou *crawlers*.

10 Datam de 1993 o *Jump Station*, que coletava informações do título e do cabeçalho das páginas e o *WWW Worm*, que indexava os títulos e as URLs. Ambos listavam os resultados conforme a ordem em que encontravam os sites, mas no mesmo ano surgiria o *RSBE*, com um sistema classificatório. O WebCrawler viria inaugurar a coleta de dados em toda a página toda. No ano seguinte (1994) surgiria o *Yahoo!*, talvez o mais longo sistema de buscas que permanece popular ainda hoje.

Outra justificativa intuitiva é que uma página pode ter um *PageRank* alto se existirem muitas [outras] páginas apontando para ela ou se algumas páginas que apontam para ela tiverem um *PageRank* alto. Intuitivamente, páginas que são bem citadas por muitas outras na Web valem a pena. Também as páginas que têm talvez uma única citação vinda de um lugar como a página inicial do *Yahoo!* geralmente valem a pena. Se uma página não fosse de grande qualidade, ou fosse um link quebrado, o mais provável é que a homepage do *Yahoo!* não se linkaria com ela. O *PageRank* lida com ambos esses casos e toda a gama entre um e o outro recursivamente propagando pesos através da estrutura da web (Brin e Page, 1998, s.p.)

Num cenário em que a quase totalidade (98,8%) dos usuários da Internet afirmam utilizar ferramentas de busca e a grande maioria (81,7%) dos usuários só dá atenção no máximo às três primeiras páginas de resultados (iProspect, 2004, p. 6 e p. 4 respectivamente), o sucesso do *Google* deu novo fôlego a toda uma série de investigações sobre o ciberespaço que tinham como foco os hiperlinks. Aspectos da funcionalidade do *PageRank* foram apropriados por diversas pesquisas, especialmente as filiadas à *Hyperlink Analysis* (Análise de Hiperlinks ou AH) e à Webometria.

A Análise de Hiperlinks considera os links entre websites (ou webpages) como indicativos de vínculos sociais e procura estudar as relações aí representadas aplicando técnicas-padrão da Análise de Redes Sociais (*Social Networks Analysis*) a esta nova fonte de dados que é a WWW. A Webometria, por outro lado, tem primado pela combinação de técnicas bastante mais simples e concentrado esforços na investigação da confiabilidade e validade dos dados e das hipóteses aventadas sobre as interpretações dos mesmos (Park e Thelwall, 2003, s.p.).

Tanto a Análise de Hiperlinks quanto a Webometria têm como premissa fundamental que as configurações de hiperlinks representam conexões entre pessoas, empresas, organizações, cidades ou nações, e como tal podem fornecer indicadores sobre as relações sociais estabelecidas por atores nesses diversos níveis. Um subconjunto particularmente interessante da AH propõe o cruzamento dos dados de linkagem com os mapas geopolíticos ‘tradicionais’ enfatizando as trocas entre nações. Exemplos desse tipo de aplicação são os trabalhos de Halavais (2001) e Barnett e Park (2003). Os autores que trabalham nesta linha não consideram os links como um sistema de referência (valoração por citação) mas como indicadores da existência de relações extra-ciberespaço entre os países que sediam os websites interconectados. Assim, os padrões de linkagem internacional constituem verdadeiras fronteiras geográficas no ciberespaço. Embora não sem problemas metodológicos, sobretudo no que se refere à coleta de dados sobre os padrões de linkagem, as projeções dos resultados de investigações desse tipo sobre os mapas geopolíticos tem fornecido indicadores intrigantes. Tipicamente filiadas

à *World System Theory* (Wallerstein, 1979), tais pesquisas têm interpretado os resultados conforme uma noção de globalização que se caracteriza por trocas desiguais entre os países informacionalmente ricos e os informacionalmente pobres, deixando de lado outras formulações.

Por uma Cibergeografia Midiática

Em consonância com o diagnóstico traçado por Park e Thelwall (2003) em sua análise comparativa da Análise de Hiperlinks e Webometria, é possível afirmar que a primeira (AH) parte de um arcabouço teórico-metodológico extremamente poderoso (*a Análise de Redes Sociais*), mas tem suas conclusões fragilizadas pelo uso de técnicas um tanto precárias para a coleta de dados e pela tendência a assumir que o significado dos links é evidente ou, por vezes, até mesmo independente de seu contexto de ocorrência. A Webometria, por outro lado, tem se esmerado no desenvolvimento de métodos para processamento e validação de dados e tem procurado relacionar cuidadosamente os significados dos links a suas condições de criação e existência. No processo tem também, no entanto, restringido seu campo de aplicação às Ciências da Informação (sua filiação original), com o que toma um tanto literalmente (e, muitas vezes, sem maiores questionamentos) a analogia entre hiperlinks e citações, entendendo-os como resultados de atos de valoração de um documento por outro¹¹. Descarta, assim, toda uma outra série de possibilidades de interpretação dos padrões de linkagem, derivada da compreensão dos links como índices de outras ordens de relação social estabelecidas dentro ou fora do ciberespaço.

De qualquer modo, a crença de que as configurações de hiperlinks entre sites que representam pessoas, empresas, organizações, cidades ou nações podem fornecer indicadores sobre as relações sociais por eles estabelecidas vale tanto para as abordagens mais amplas praticadas pela Análise de Hiperlinks quanto para as mais restritas, da Webometria. Porque não dão a devida atenção a quanto a noção de ‘representação’ é fundamental para a compreensão e operacionalidade de tal premissa, as pesquisas de ambas as filiações constantemente desconsideram o caráter comunicacional de problemáticas que se propõem a abordar.

Reconhecer o caráter midiático da *World Wide Web* significa admitir que os websites representam não apenas elementos estritamente vinculados aos conteúdos que seus

¹¹ Uma medida do quanto essa analogia é tomada de forma literal advém da constatação de que a maioria absoluta dos trabalhos de Webometria tem se detido justamente sobre os padrões de linkagem verificados em websites acadêmicos (Park e Thelwall, 2003, s.p.).



enunciadores efetivamente desejaram expressar, mas que incorporam também todo um outro conjunto de marcos e indicativos de suas condições de produção e existência. Em outras palavras, implica reconhecer a efetiva existência e atuação, em todas as fases do processo comunicacional estabelecido em torno dos produtos midiáticos disponíveis na Web, dos agentes (aquelas mesmas pessoas, empresas, organizações, cidades ou nações) representados pelos websites. Em paralelo ao instrumental quantitativo que hoje lhe é ‘tradicional’, uma Análise de Hiperlinks orientada por tal pressuposto logo se veria conclamada a convocar também o apoio de métodos qualitativos, por exemplo para iluminar a opção (sempre necessária e difícil) entre montar o *corpus* utilizando ferramentas de busca públicas (*Google, Yahoo, AltaVista*) ou *crawlers* ou para discernir com maior nível de detalhe o que caracteriza os diferentes agrupamentos de páginas cujos padrões de linkagem se procura averiguar.

Orientada pela mesma premissa, a confluência de pressuposições e técnicas de pesquisa pode ir, no entanto, mais além: entende-se que o reconhecimento do caráter midiático dos websites permitiria mobilizar de forma um tanto mais produtiva a Teoria Matemática das Redes, pois funcionaria como um antídoto para os desvios tipicamente decorrentes do tratamento da WWW como uma mera abstração. Finalmente, a proposta de uma Análise de Hiperlinks com os olhos voltados para a interface entre o mundo ‘real’ e o ciberespaço convida, também, à incorporação do conjunto de conhecimento hoje abrigado pela Cibergeografia. Mais uma vez, a abordagem dos websites como produtos midiáticos ajudaria a compreender porque a projeção dos mapeamentos de hiperlinks sobre dados da geografia física e política tem se mostrado capaz de fornecer indicadores relevantes sobre a atuação de condicionantes culturais, econômicos e políticos na interface entre o ciberespaço e o mundo ‘real’, como é o caso por exemplo das pesquisas já mencionadas de Halavais (2001) e Barnett e Park (2003).

Uma vez que o reconhecimento do caráter midiático da WWW é proposto como espinha dorsal da confluência entre os instrumentais teórico-metodológicos da Cibergeografia, da Teoria Matemática das Redes, da Análise de Hiperlinks e da Webometria, sugere-se denominar a abordagem daí originada ‘Cibergeografia Midiática’. Resta enfrentar a tarefa, nada singela, de dar corpo ao conjunto de pressupostos e técnicas quanti-qualitativas que se anuncia a partir de um tal movimento de convergência.

Referências bibliográficas

BARABASI, A.-L., *Linked*. New York, Plume, 2002.



BARNETT, G.A. e PARK, H.W., “The Structure of Internet International Flows”, *IR 4.0, Broadening the Band, IV International Conference of the Association of Internet Researchers, AoIR*. Toronto, Canadá, 2003. Disponível online a partir de <http://www.aoir.org> [acesso restrito] [nov 2003].

BATTY, M., Virtual Geography, *Futures*, vol 29, n° 4/5, Pergamon Press, 1997, pp. 337-352.

BRIN, S. e PAGE, L., The Anatomy of a Large-Scale Hypertextual Web Search Engine *in WWW7 / Computer Networks* 30, 1998, p. 107-117. Disponível online em <http://dbpubs.stanford.edu:8090/pub/1998-8> [jun 2005]

CASTELLS, M., *A Sociedade em Rede. Economia, Sociedade e Cultura na Era da Informação*, vol. 1. São Paulo, Paz e Terra, 1999.

CLICZNETWORK, Trends & Statistics: The Web's Richest Source, *ClicZStats*, s.d. Disponível online em http://www.clickz.com/stats/web_worldwide/ [jun 2005]

CUNHA, P., “Espaço tecnológico e espaço comunicacional: ciberespaço, novas centralidades, novas periferias” in A. LEMOS. e P. CUNHA (orgs), *Olhares sobre a cibercultura*. Porto Alegre, Sulina, 2003, p. 197-211.

DODGE, M. e R. KITCHIN, *Mapping Cyberspace*. London, Routledge, 2001.

FRAGOSO, S., Espaço, Ciberespaço, Hiperespaço, *revista Textos em Comunicação e Cultura*, n° 42, 2000, p. 105-113.

FRAGOSO, S., “Um e Muitos Ciberespaços” in A. LEMOS. e P. CUNHA (orgs), *Olhares sobre a cibercultura*. Porto Alegre, Sulina, 2003, p. 212-231.

FRAGOSO, S., As Múltiplas Faces da Exclusão Digital, *Ícone* vol 6, n° 7, Julho de 2004. Recife, Universidade Federal de Pernambuco, 2004, p. 110-122.

GAID-FGV, *Mapa da exclusão digital*. Coordenação M. C. Neri. Rio de Janeiro, FGV/IBRE, CPS, 2003. Relatório disponível online em <http://www.fgv.br> [mai 03] e arquivos complementares em CD-ROM.

HALAVAIS, A. M. C., *Measuring National Borders on the World Wide Web*. Dissertação de mestrado apresentada à *School of Communications* da *University of Washington*, 1998. Disponível online em <http://alex.halavais.net/research/thesis.pdf> [set 2003]

IBGE, *PopClock*, Disponível online em <http://www.ibge.gov.br/> [jun 2005]

iProspect, *iProspect's Search Engine User Attitudes Survey Results*, Abril-Maio de 2004. Disponível online a partir de <http://www.iProspect.com> [jun 2005]

LÉVY, P., *Cibercultura*. São Paulo, Editora 34, 1999.

NIELSEN, J., Do Websites Have Increasing Returns? *Alertbox*, 15 de Abril de 1997. Disponível online em <http://www.useit.com/alertbox/9704b.html> [jun 2005]

NIELSEN, J., *Zipf Curve and Website Popularity*, 1997. Disponível online em <http://www.useit.com/alertbox/zipf.html> [dez 2002]



NUA, *Internet Surveys: How Many Online?*, 2003. Disponível online em http://www.nua.com/surveys/how_many_online/world.html [jun 03]

PARK, H.W. e THELWALL, M., Hyperlink Analyses of the World Wide Web: a Review, *Journal of Computer Mediated Communication* vol. 8, n°4, julho de 2003. Disponível online em <http://www.ascusc.org/jcmc/vol8/issue4/park.html> [mai 2005]

PRIMO, A. e RECUERO, R., Co-Links: Proposta de uma nova tecnologia para a escrita coletiva de links multidirecionais, *Anais da XIII Compós – Encontro Anual da Associação Nacional de Programas de Pós-Graduação em Comunicação*, Grupo de Trabalho Tecnologias Informacionais de Comunicação e Sociedade. São Bernardo do Campo, UMESP, 2004. CD-ROM.

SERRA, J. P., “A Internet e o Mito da Visibilidade Universal”, *Agora.net#1*, s.d. Disponível online em <http://www.labcom.ubi.pt/agoranet> [dez 2003]

TRIVINHO, E., *Redes: obliterações no fim de século*. São Paulo, Annablume/FAPESP, 1998.

VIRILIO, P., *O Espaço Crítico*. São Paulo, Editora 34, 1999.

WALKER, J., Links and Power: The Political Economy of Linking on the Web, *Hypertext'02. Encontro anual do ACM SIGWeb*, Baltimore, USA. 2002. Disponível na biblioteca online da Association for Computing Machinery, a partir de <http://www.acm.org> [acesso restrito] [mar 2004]

WALLERSTEIN, I. *El moderno sistema mundial: la agricultura capitalista y los orígenes de la economía-mundo europea en siglo XVI*. México, Siglo Veintiuno, 1979.

WENTIAN L., *References on Zipf's Law*, 1999. Disponível online em <http://linkage.rockefeller.edu/wli/zipf/> [mar 2003]