
O fim do virtual: os métodos digitais ¹

Richard Rogers ²

Uma distinção ontológica entre o que é digital e aquilo que é digitalizado precisa ser feita. De modo mais claro, é preciso diferenciar objetos de estudo, conteúdos, dispositivos³ e ambientes que “nascem” no novo meio e aqueles que “migraram” para ele. Como veremos, essa distinção suscita a questão do método para pesquisas relacionadas à internet. Assim, os métodos atuais de estudo deveriam mudar, sutil ou indiscriminadamente, tendo em vista apreender tanto os objetos quanto os conteúdos que são “do meio”? Essa questão envolve os “métodos virtuais”, os quais que se valem de métodos provenientes das ciências humanas e sociais para utilizá-los no próprio meio. Isso quer dizer que a distinção entre o que é digital e aquilo que é digitalizado também poderia ser aplicada aos métodos de pesquisa vigentes. Desse modo, podemos perguntar: que tipo de pesquisa pode ser realizada com métodos que foram digitalizados (como diretórios e entrevistas on-line) face àqueles que são digitais (como sistemas de recomendação e folksonomia)?

A fim de responder à questão indicada, eu gostaria de propor que a pesquisa web tenha novos usos e que os objetos digitais sejam enfatizados, assim como os métodos que habitualmente fazem uso deles. O meu esforço é, então, deslocar a atenção das novas mídias e da cultura digital de modo geral, e na pesquisa web, de modo mais específico, para longe das oportunidades oferecidas pela digitalização e,

¹ Publicado originalmente como capítulo 1 do livro *Digital Methods* (MIT Press, 2013). Tradução de Carlos d’Andréa (docente pelo PPGCOM/UFMG) e Tiago Barcelos P. Salgado (doutorando pelo PPGCOM/UFMG e bolsista CAPES/PROEX UFMG). Ambos são pesquisadores pelo NucCon, grupo de pesquisa vinculado ao CCNM/UFMG. E-mail: carlosfbd@gmail.com; tigubarcelos@gmail.com.

² Professor do Departamento de *Media Studies* da Universidade de Amsterdã (UVA), na Holanda, e coordenador do projeto *Digital Methods Initiative* (DMI). E-mail: R.A.Rogers@uva.nl.

³ Nota dos tradutores (NT): o termo dispositivo, tradução do original *device*, refere-se aos aparatos tecnológicos e não a outras concepções de dispositivo, como as elaboradas por Michel Foucault, Gilles Deleuze e Giorgio Agamben.

assim, investigar como a pesquisa em internet pode ir além do mero estudo da cultura on-line. Tendo isso em vista, como coletar e analisar links, *tags*, resultados de mecanismos de busca, sites arquivados, perfis de sites de redes sociais, edições na Wikipédia e outros objetos digitais? Como alguém pode aprender a respeito do uso que os dispositivos on-line (ex.: sistemas de recomendação) fazem de objetos digitais e, fundamentalmente, como esse uso pode ser readaptado para a pesquisa social e cultural? Finalmente, proponho uma prática de pesquisa que aprenda com os métodos dos dispositivos on-line, readapte-os e busque fundamentar afirmações sobre a mudança cultural em dados web, introduzindo o termo “fundamentação em dados on-line” (*online groundedness*). O objetivo geral é retrabalhar os métodos relacionados à pesquisa em internet desenvolvendo uma nova vertente de estudos, os métodos digitais.

Até hoje, os métodos utilizados em pesquisas na internet têm servido para criticar a constante ideia de que a internet é uma realidade virtual à parte. Esse pensamento surgiu a partir do discurso em torno da realidade virtual, ao final dos anos 1980 e início dos anos 1990, quando a internet correspondia a um domínio virtual, com oportunidades para a redefinição da consciência, da identidade, da corporalidade, da comunidade, da cidadania e da política (BARLOW, 1996; BENEDIKT, 1991; DIBBELL, 1998; RHEINGOLD, 1991; RHEINGOLD, 1993; SHAVIRO, 2008; STONE, 1995; TURKLE, 1995). De fato, em 1999, em um dos primeiros esforços para sintetizar a pesquisa em internet, Steve Jones, pesquisador em Comunicação, convidou pesquisadores a irem além da perspectiva que considerava a internet como um domínio à parte e iniciou a discussão a respeito do método (JONES, 1999). Como os cientistas sociais poderiam estudar a internet se eles não confiavam nas abordagens elaboradas até então, como a interação homem-computador, a psicologia social e os estudos em cibercultura (HINE, 2000)?

No trabalho inovador sobre o uso da internet em Trindade e Tobago, os etnógrafos Daniel Miller e Don Slater confrontaram a ideia do ciberespaço – uma realidade à parte em que todos aqueles que nela “habitam” experimentam suas *affordances* – com o potencial de transformação de identidades, não importando o lugar onde aqueles que navegavam pelo ciberespaço estavam (MILLER; SLATER, 2000). Slater e Miller (2000) fundamentaram suas pesquisas na internet,

argumentando que os habitantes de Trindade e Tobago se apropriaram desse meio, adaptando-o às suas práticas culturais. Como um estudo de caso, o ponto principal dessa pesquisa foi seu potencial para generalizações. Se os moradores de Trindade e Tobago usavam a internet para encenar a cultura do país, a expectativa era que outras culturas também faziam o mesmo.

O importante programa *Sociedade Virtual?* (1997-2002) marcou um outro momento decisivo na pesquisa em internet ao desmistificar as capacidades transformadoras do ciberespaço por meio de múltiplos estudos empíricos sobre quem utilizava a internet. No fim das contas, esse programa elaborou cinco “regras da virtualidade” (WOOLGAR, 2002). Naquilo que tem sido atualmente nomeado como a clássica crítica da divisão digital, pesquisadores argumentam que o uso das novas mídias se baseia em situações individuais (questões de acesso) e que receios e riscos são divididos de maneira desigual (questões de competência). No que diz respeito à relação entre real e virtual, as interações virtuais são complementares ao “real” e não as suas substitutas. Igualmente, as interações virtuais suscitam mais interações reais e não um isolamento.

Como conclusão, a pesquisa descobriu que as identidades se fundamentam tanto no on-line quanto no off-line. De maneira significativa, o programa passou a vislumbrar as abordagens que têm sido caracterizadas como métodos virtuais, com um instrumental para estudar quem utiliza a internet. *Surveys*, entrevistas, observação participante e não participante, tudo isso se tornou o conjunto de métodos preferenciais para as investigações. Nas humanidades, os subsequentes estudos focados em usuários – concentrados em amadores, fãs e produzíveis – também passaram a lidar com a divisão real/virtual, buscando demonstrar e criticar a reputação do *status* da cultura on-line (JENKINS, 2006; KEEN, 2007; BRUNS, 2008). O argumento que defendemos neste trabalho é que os métodos virtuais e os estudos de usuários nas ciências humanas e sociais têm deslocado a atenção para longe dos dados presentes nos meios e das oportunidades de pesquisa para além da cultura on-line.

Como os estudos de usuários com base em dados coletados rotineiramente por *softwares* podem ser repensados? Até o presente momento, esses estudos têm se

baseado em descrições que privilegiam a observação, as entrevistas e os questionários, diferenciando-se da computação social e das humanidades computacionais, por um lado, e das grandes instituições comerciais, com suas notáveis coleções de dados arquivados, por outro. De certo modo, empresas como a *Google*, a *Amazon* e outras que dominam o mercado de dispositivos web atualmente conduzem estudos sobre quem usa a internet. No entanto, o termo “estudo de usuários” é pouco utilizado nesse contexto.

As entradas de dados (*inputs*) pelos usuários, como preferências, histórico de busca, histórico de compra e localização são coletadas e analisadas de modo a ajustar os resultados. Tomando como exemplo esse trabalho de adequação, há várias reivindicações constantes por uma virada metodológica na pesquisa em internet, pelo menos no que tange à coleta de dados. Com as “análises culturais”, nomeadas após o lançamento do *Google Analytics*, a proposta é elaborar serviços massivos para a coleta, arquivamento e análise de dados para as humanidades digitais (MANOVICH, 2007). Uma maneira de descrever a virada metodológica é considerar o seu início a partir da necessidade (negociada) de acesso a conjuntos de dados comerciais, como o conjunto de dados de atividades de milhares de usuários do *Second Life* ou do jogo *Everquest*, da *Sony* (CONTRACTOR, 2009). As análises culturais gostariam de desempenhar o papel de *big science* (LAZER et al., 2009). Conforme Manovich (2008), “as visualizações deveriam ser elaboradas para aproveitarem ao máximo a grande resolução (*gigapixels*) dos monitores disponíveis hoje”.

De certo modo, os programas de pesquisa são uma resposta à questão: “O que a *Google* faria?”. Esses programas de pesquisa poderiam ser situados no amplo contexto de extensão e de efeitos da “googlelização” (VAIDHYANATHAN, 2011). Até a presente data, a crítica à “googlelização”, originária da reação à entrada das empresas que desenvolvem mecanismos de busca on-line no mercado de arquivamento de livros (o projeto *Google Livros*), tem examinado a crescente “deformação” da *Google*, do modelo de negócios dessa empresa, bem como de sua estética por meio das indústrias de informação e conhecimento (JEANNENEY, 2007; VAIDHYANATHAN, 2007). Os estudiosos em biblioteconomia, em especial, preocupam-se com a mudança dos locais de acesso à informação e ao conhecimento (de arquivos públicos extensos para servidores comerciais). Os “efeitos do buscador

Google” são efeitos das mídias. Esses efeitos podem ser expressos em termos da suplantação da navegação por mecanismos de busca. Igualmente, eles podem ser estudados em termos da “morte do editor-perito” e da ascensão do algoritmo de *back-end*, temas que serão retomados ao longo deste texto. Neste tópico, no entanto, trata-se de considerar que esses efeitos também podem ser estudados em termos de modelos para a pesquisa – modelos que buscam replicar a escala da coleta de dados, bem como a análise desses dados.

A proposta apresentada neste trabalho é mais modesta, mas está de acordo com os estudos de usuários que se voltam para o registro do uso que eles fazem da internet. Os dispositivos e *softwares* instalados em computadores (os navegadores, por exemplo) capturam o uso cotidiano de pessoas que os utilizam por meio do que é designado como “interatividade de registro” (JENSEN, 1998). Nesse sentido, o histórico de navegadores se tornaria um meio para o estudo do uso que as pessoas fazem desses dispositivos e *softwares*. A discórdia mais evidente na virada metodológica descrita acima se refere ao fato de que a coleta de dados poderia se beneficiar da reflexão sobre como a computação pode ter técnicas que podem ser reutilizadas para pesquisa. Assim, a proposta é considerar, em primeiro lugar e acima de tudo, a disponibilidade de “técnicas” de computação.

Desse modo, eu gostaria de ajudar a definir uma nova era na pesquisa em internet, a qual não mais se preocupe com a divisão entre real e virtual. Trata-se de uma pesquisa que passa a se preocupar com mudanças nos tipos de questões colocadas ao estudo da internet. A internet é usada como um espaço de pesquisa, para além da cultura on-line. A questão não é mais o quanto a sociedade e a cultura estão on-line, mas sim como diagnosticar mudanças culturais e condições sociais por meio da internet.

O ponto de partida conceitual é o reconhecimento de que a internet não é simplesmente um objeto de estudo, mas também uma fonte para os estudos. O conhecimento pode ser construído, então, com base em dados coletados e analisados por dispositivos, como os mecanismos de busca. Um dos exemplos mais notáveis é o *Google Flu Trends*, um projeto não comercial da *Google.org*. Lançado em 2008, esse projeto antecipa possíveis locais de surto de *influenza*, mensurando o número de

buscas realizadas para os termos “gripe” e “sintomas de gripe”, entre outros relacionados a esses, bem como localiza geograficamente os lugares em que essas buscas foram realizadas (TABELA 1). Isso complementa e desafia os métodos de coleta de dados existentes (número de hospitalizações, de pacientes e de relatórios laboratoriais). Do mesmo modo, reabre a discussão da web como meio de antecipação (ROGERS, 2003; *Centers for Disease Control and Prevention*, 2011).

Onde se originaram a “web fundamentada em dados” (*grounded web*) e as práticas de pesquisa de localização geográfica associadas a ela? O “fim do ciberespaço” como um espaço não espacial (nos termos de Manuel Castells) pode ser relacionado aos desfechos técnicos da famosa ação judicial movida por duas organizações não governamentais em 2000 na França (CASTELLS, 1996; GOLDSMITH; WU, 2006). À época, usuários franceses estavam aptos a acessar páginas sobre acontecimentos memoráveis do Nazismo no *site Yahoo.com* nos Estados Unidos. No mesmo período, duas organizações não governamentais francesas quiseram tornar essas páginas inacessíveis quando o acesso fosse realizado na França. Seguindo a ação judicial, a tecnologia *IP-to-geo* (localização de endereço) foi promovida para canalizar o conteúdo em termos nacionais, ou seja, quando alguém que está na França digita o endereço *google.com* em algum navegador, a busca agora retorna, como padrão, o endereço *google.fr*. Essa fundamentação da web em dados (*grounding*) foi implementada pelos maiores projetos de organização de conteúdo, como o *YouTube*. Igualmente, a televisão on-line também é acessada geograficamente.

Table 1.1
U.S.-Based Flu-Related Queries, the Geolocation Results of Which Correlate with Surveillance
Data from the Centers for Disease Control, 2011

treatment for flu
how to treat the flu
treat flu
cure flu
cold or flu
treat the flu
how to treat flu
dangerous fever
remedies for flu
influenza type a
human temperature
flu medicine
symptoms of flu
is flu contagious
flu and fever
length of flu
flu or cold
flu duration
cure the flu
duration of flu
when is the flu contagious
how long is flu contagious
medication for flu
flu contagious
thermoscan
flu vs cold
remedies for the flu
fever and flu
is it the flu
flu fever

Source: Google Correlate results for Influenza-like Illness (CDC), <http://correlate.googlelabs.com/>,
May 26, 2011.

Tabela 1: Buscas sobre gripe em bases de dados americanas.

Fonte: CDC – <http://correlate.googlelabs.com>

Baseado em coleta de dados referentes a práticas on-line, o trabalho de diagnóstico por meio do qual se reivindica condições sociais conduz a novas formulações teóricas. Para o terceiro período da pesquisa em internet, o programa dos métodos digitais introduz o termo “fundamentação em dados on-line” para conceituar a pesquisa que segue as mídias, apreende suas dinâmicas e faz reivindicações fundamentadas sobre mudanças culturais e sociais. De fato, o objetivo teórico maior dos métodos digitais é pensar uma nova relação entre a web e o que a fundamenta. Como os etnógrafos que vieram antes dos métodos digitais para o

programa britânico *Virtual Society?*, alguém precisou visitar os fundamentos dos dados (*the grounded*) para estudar a web.

Nesse quesito, o programa de pesquisa complica a lógica na qual as descobertas de perspectivas mais tradicionais estão fundamentadas (MARRES; ROGERS, 2008). Por exemplo, agora que a internet se tornou uma significativa “metafonte”, o jornalismo passa a ter necessidades metodológicas que deem conta da veracidade das fontes. A passagem de informações e conteúdos de fonte para fonte era antes uma abordagem de rede social para a checagem de informações, para falar em termos metodológicos.

“Com quem mais devo conversar?”. Esta é a questão que conclui uma entrevista que discutia se uma confiança entre fonte e jornalista tinha sido ou não estabelecida. A relação entre “Com quem eu devo conversar?” e “Com quem mais você está *linkado*?” é assimétrica para o jornalismo, mas esta última é a questão formulada pelos mecanismos de busca quando recomendam informações. Como pensar as diferenças entre as recomendações de fontes entre os links verbais e os on-line? A busca é o começo da consulta por informações, consulta esta que termina com algumas entrevistas fundamentadas realizadas fora da internet, mantendo assim as divisões entre real e virtual? Ou isto é muito simplista? Nossa divisão ideal do conjunto de fontes (real e virtual, fundamentada ou *googlada*) levanta a questão do que vem a seguir. O que nós procuramos nas conclusões da entrevista para checar a realidade? Para alguns, a internet pode não estar mudando a hierarquia das fontes (por exemplo, as restrições para citar a Wikipédia em alguns modelos educacionais), mas ela pode estar muito bem mudando a ordem da checagem e a relação da web com sua fundamentação em dados.

Eu desenvolvi a noção de “fundamentação em dados on-line” após a leitura de um estudo feito pelo jornal holandês NRC *Handelsblad*. A investigação dos grupos de direita e de ódio na Holanda indagou se a linguagem usada por eles foi se tornando mais extremista ao longo do tempo, talvez indicando um endurecimento das culturas de direita e de ódio. Significativamente, os pesquisadores optaram pelo uso do *Internet Archive* ao invés de um pesquisador especialista, *surveys* entre especialistas ou os panfletos, *flyers* e outras efemérides disponíveis no Instituto de História Social

(DOHMEN, 2007). Eles localizaram e analisaram as mudanças no tom da linguagem ao longo do tempo, descobrindo que os *sites* de direita estavam empregando de modo crescente uma linguagem mais extremista. Assim, as descobertas feitas sobre cultura são fundamentadas na análise de websites. De maneira mais expressiva, o on-line se tornou a referência com a qual se poderia julgar a extensão de uma condição social percebida.

Siga o meio: a abordagem de pesquisa dos métodos digitais

Por que seguir o meio? Um ponto inicial para se responder a essa questão é reconhecer que pesquisar a internet é se deparar constantemente com objetos de estudo instáveis. Essa instabilidade é discutida de maneira frequente em termos de efemeridade dos websites e de outras mídias digitais, bem como das complexidades associadas à sua “fixação”, termo tomado de empréstimo do campo da fotografia.

Como tornar as mídias digitais permanentes, a fim de serem cuidadosamente estudadas? Em um tipo de abordagem, o arquivamento de dados web é periodicamente confrontado com o dilema da captura de websites, por um lado; por outro, esse arquivamento tem de lidar com a manutenção de sua vivacidade, uma vez que os dados se alteram a todo instante. Em outro tipo de abordagem, *hardwares* e *softwares* antigos são mantidos para impedir que as mídias “morram”. Em outra perspectiva, praticada também por estudos de jogos eletrônicos (*game studies*), a questão da efemeridade é tratada com base na simulação e na emulação, as quais mantêm programas nostálgicos como jogos de *Atari* rodando nos *hardwares* atuais. Essa questão, entretanto, é muito mais ampla do que as questões de conservação. Quem pesquisa a *internet* e com ela é repetidamente surpreendido pelo que ocorre no meio, como as atualizações de *softwares* que abruptamente interrompem e, algumas vezes, “esvaziam” a pesquisa de outrem.

Como prática de pesquisa, a ação de seguir o meio, oposta àquela de esforçar-se para fixá-lo, ou seja, torná-lo estável, também deve ser discutida pelo uso de um termo emprestado do jornalismo e da sociologia da ciência: “dar um furo”. Ser o primeiro a publicar algo é “conseguir o furo”. “Furar alguém” se refere a alguma pessoa que publica achados antes de outrem. O sociólogo da ciência Michael Lynch aplicou esse termo na situação em que os temas de pesquisa de alguém chegam às

mesmas ou a conclusões similares que outros pesquisadores que as registraram antes. O resultado é que “[os temas de pesquisa] reconfiguram o campo no qual nós pensávamos que nosso estudo teria sido situado” (LYNCH, 1997).

Na pesquisa em internet, “ser furado” é comum. Analistas da indústria, cães de guarda (*watchdogs*) e blogueiros cunham novos termos todo o dia (ex.: “googlização”) e chegam a conclusões que moldam os trabalhos acadêmicos em andamento. No entanto, eu argumento que a realização de furos (no sentido jornalístico) também é feita por objetos, os quais são continuamente reconfigurados. Por exemplo, o *Facebook* tem sido considerado como um exemplo de “jardim murado”, ou seja, um sistema relativamente fechado, em que, por padrão, apenas amigos podem visualizar informações e atividades de outras pessoas amigas. Trata-se de uma série de círculos concêntricos: um usuário deve ter uma conta para ter acesso a esse site. Igualmente, deve se tornar amigo de outras pessoas para visualizar os perfis delas e precisa mudar as suas configurações padrões de privacidade para permitir que amigos vejam o seu próprio perfil. A exposição máxima diz respeito à abertura de perfis a amigos de amigos.

Em março de 2009, no entanto, o *Facebook* alterou uma de suas configurações. A partir de então, os usuários poderiam abrir seus perfis a todos os outros usuários com contas nessa plataforma, em oposição aos perfis abertos apenas para amigos ou amigos de amigos, como nas configurações anteriores (SLEE, 2009). Que tipo de pesquisa seriam “furadas” pela mudança brusca efetuada pelo *Facebook*? Quem seria beneficiado? O *Facebook* exemplifica a notável reconfiguração repentina de um objeto de pesquisa, aspecto que é comum ao meio investigado.

De modo mais teórico, seguir o meio é uma forma particular de pesquisar as especificidades dele. A especificidade do meio não se refere apenas a subdivisão das disciplinas nos estudos de mídia, de acordo com os objetos de estudo: vídeo, rádio, televisão etc. Refere-se também às distinções ontológicas das mídias, no sentido em que as ontologias são construídas de modo distinto. Para o erudito literário e pesquisador das mídias Marshall McLuhan, as mídias são específicas no modo como elas engajam os sentidos (MCLUHAN, 1964). Profundidade, resolução e outras propriedades estéticas têm efeitos no modo como, de forma ativa ou passiva, alguém

processa as mídias. Dito de outra maneira, ou alguém é preenchido pelas mídias ou as preenche. Para o teórico cultural Raymond Williams, a especificidade do meio reside em outro lugar. As mídias se especificam a partir das formas que assumem – formas estas que são moldadas pelos atores dominantes para servir seus interesses (WILLIAMS, 1974). Por exemplo, a criação do “fluxo”, o termo pelo qual a televisão sequencia a programação de modo a manter os espectadores assistindo, serve para medições de audiência e publicidade. Assim, para Williams (1974), as mídias não são distintas umas das outras *a priori*, mas podem ser levadas a tal.

Para Katherine Hayles, as mídias têm características em suas materialidades: um livro específica, enquanto um texto, não (HAYLES, 2004). A proposta dessa autora para uma “análise da especificidade das mídias” é um programa de estudos comparativos entre elas, que considera as características materiais delas (por exemplo, o hipertexto nas mídias digitais) e investiga sua presença (simulada) em outras mídias (por exemplo, no impresso). Seria possível pegar outros traços das mídias e estudá-los de forma comparativa. Por exemplo, como argumentou Alexander Galloway, o fluxo está presente não apenas no rádio e na televisão, mas também na web, onde links quebrados interrompem a navegação (GALLOWAY, 2004).

O ponto de partida de Hayles (2004) pode ser visto no estudo de Mathew Fuller sobre o *Microsoft Word* e o *Adobe Photoshop* sobre como um *software* específico restringe ou ativa um texto (FULLER, 2003). Para Fuller (2003), um documento do *Word* ou uma imagem do *Photoshop* são resultados específicos dos *softwares* e, portanto, distintos de um outro documento ou imagem. Um programa de pesquisa contínuo estudaria os efeitos dos recursos dos *softwares*, como também Lev Manovich aponta em seu trabalho sobre a especificidade das mídias computacionais. A ontologia de Manovich (2008) se move para além dos resultados das mídias (impressão hipertextual de Hayles, documento *Word* e imagem do *Photoshop* segundo Fuller) e propõe o termo “metamídia”. As mídias computacionais são metamídias na medida em que incorporam formatos anteriores, em consonância com a tese da remediação proposta por Bolter e Grusin (1999). Para Manovich (2008), entretanto, as mídias computacionais não apenas remoldam os resultados de outras mídias; elas também incorporam suas formas de produção.

A especificidade do meio proposta aqui não se apoia tanto no engajamento dos sentidos de McLuhan (1964), nas formas socialmente moldadas discutidas por Williams (1974), na materialidade de Hayles (2004) ou nas propriedades e características de outros teóricos, tanto no que diz respeito aos resultados (formas culturais) quanto às formas de produção. A especificidade do meio que proponho refere-se ao método, tanto no estudo dos modos preferidos para estudar mídias específicas (registros em diários na pesquisa de audiência nos estudos de TV, por exemplo) quanto no que tange aos métodos do meio.

Previamente, eu descrevi esse trabalho como “epistemologia da web” (ROGERS, 2004). Na web, informação, conhecimento e sociabilidade são organizados por sistemas de recomendação – algoritmos e *scripts* preparam e oferecem relações de URLs, arquivos de mídia, amigos etc. De certo modo, Manovich (2008) direcionou a discussão para este lado, tanto com seu foco nas formas de produção (método em um sentido artesanal) quanto com a virada metodológica associada com a iniciativa de análise cultural. Esse autor concentrou-se, no entanto, em conteúdos digitalizados e não em conteúdos digitais. Eu gostaria de levar essa virada mais adiante e propor que os pouco questionados métodos da web são dignos de estudo, tanto em si quanto nos efeitos que espalham por outras mídias (por exemplo, programas de TV recomendados para usuários do *TiVo* com base em seus perfis).

O trabalho inicial na área da epistemologia da web estava situado no contexto das políticas dos mecanismos de busca (INTRONA; NISSENBAUM, 2000) e buscou entender os significados pelos quais as fontes são julgadas pelas ferramentas. Por que, em março de 2003, a Casa Branca, a CIA, o FBI, a Fundação *Heritage* (uma organização de centro-direita) e organizações jornalísticas como a CNN estavam no topo das buscas feitas com o termo “terrorismo”? De certa maneira a resposta está em como os hiperlinks são manipulados. Os hiperlinks, no entanto, são apenas um objeto digital para o qual podem ser acrescentados o tópico, *tag*, *PageRank*, edição da Wikipédia, postagem, comentário, *trackback*, *pingback*, endereço de IP, URL, *whois*, *timestamp*, link permanente, *social bookmark* e perfil. Sem uma ordem específica, a lista continua. O propósito é estudar como esses objetos são manipulados especificamente no meio e aprender com o método do meio.

Em seguida, eu gostaria de introduzir uma série de objetos, formatos, dispositivos, assim como plataformas, primeiro abordando rapidamente como eles são frequentemente estudados com os métodos digitalizados, assim como os pontos de partida conceituais que estão fora do meio. Depois, eu gostaria de discutir a diferença que faria para a pesquisa se seguissemos o meio – aprendendo como os objetos digitais (como os hiperlinks) são tratados pelos dispositivos, como websites são arquivados, como mecanismos de busca ordenam a informação e como a tecnologia de localização geo-IP oferece conteúdos de acordo com a nacionalidade ou o idioma. Que tipos de pesquisa podem ser feitas por meio da análise de hiperlinks, repondo *insights* a partir dos algoritmos dominantes? Como trabalhar com o *Internet Archive* para pesquisa social? Por que capturar históricos de websites? Como os resultados de mecanismos de busca podem ser estudados de modo a mostrar mudanças nas hierarquias de credibilidade e as diferenças na dependência de fontes em relação à esfera da web, da esfera jornalística e da blogosfera? A tecnologia de localização geo-IP pode ser retrabalhada de modo a perfilar países e culturas? Como o estudo de sites de rede social podem revelar gostos e preferências culturais? Como robôs estão mudando o modo como a qualidade da Wikipédia é mantida? O que um robô de pesquisa faria?

Desse modo, do micro ao macro, eu trato o hiperlink, o website, o mecanismo de busca, as esferas e a web (ou webs, incluindo as nacionais). Finalmente, eu me volto para as plataformas – sites de mídia social e a Wikipédia – e busco aprender a partir da criação de perfis e das culturas dos robôs, respectivamente, repensando como implantá-las de forma analítica. O propósito geral de seguir o meio é reorientar a pesquisa em internet para considerá-la como uma fonte de dados, um método e uma técnica.

O link

Existem pelo menos duas abordagens predominantes para se estudar os hiperlinks: a teoria literária do hipertexto e a teoria das redes sociais, incluindo as “redes de pequeno mundo” e a Teoria do Labirinto (*Path Theory*) (LANDOW, 1994; WATTS, 1999; PARK; THELWALL, 2003). Para os teóricos literários do hipertexto, conjuntos de hiperlinks formam uma multidão de caminhos distintos pelo texto.

Navegar ou clicar na navegação do texto pode desencadear uma história pela escolha de rotas (por meio de múltiplos cliques) pelo texto (ELMER, 2001). Desse modo, são de interesse os novos significados da autoria, assim como a história contada por meio da navegação entre links. Para os teóricos das redes de pequeno mundo, os links que formam caminhos mostram a distância entre os atores. Os analistas de redes sociais recorrem ao pensamento baseado em caminhos e se concentram em como os laços, unidirecionais ou bidirecionais, posicionam os atores (KREBS, 2002). Um vocabulário especial foi desenvolvido para caracterizar a posição de um ator em uma rede, especialmente sua centralidade. Por exemplo, um ator está “extremamente entre” (*highly between*) se há uma alta probabilidade de que outros atores precisem passar por ele para se alcançarem.

Como os mecanismos de busca tratam os links? Indiscutivelmente por uma abordagem da cienciometria (e da sociologia associativa). Assim como na análise de redes sociais, o interesse é na posição de um ator, mas não necessariamente em termos de distância de um do outro ou os meios pelos quais um ator pode ser alcançado pelo estabelecimento de conexões em rede. Antes, os laços são indicadores de reputação e podem ser tomados para definir a posição do ator. Adicionalmente, essa abordagem não assume que os laços entre os atores são amigáveis ou, pelo contrário, são utilitários no sentido de prover caminhos ou pistas poderosas para o estabelecimento de ligações “bem-sucedidas”.

Aqui eu gostaria de seguir o modo como os buscadores tratam links como marcadores de impacto e reputação. Como a reputação de um ator poderia ser caracterizada pelos tipos de hiperlinks estabelecidos e recebidos? No algoritmo básico de busca, atores podem ser perfilados não apenas pela quantidade de links recebidos diretamente e da quantidade recebida de outros que também receberam muitos links. Atores podem ser perfilados também pela análise de quais links eles estabelecem e recebem de maneira específica (BEAULIEU, 2005).

Em uma pesquisa anterior, colegas e eu encontramos tendências de estabelecimento de links entre tipos de domínios, isto é, governos tendem a *linkar* apenas outros sites governamentais. Os sites não governamentais tendem, por sua vez, a *linkar* uma variedade de sites, ocasionalmente até críticos a eles. Websites

corporativos tendem, por seu turno, a não *linkar* outros sites, com a exceção de seus coletivos, como no caso dos sites de comércio industrial. Sites acadêmicos e educacionais tipicamente *linkam* para parceiros e iniciativas criadas por eles. Consideradas juntas, essas tendências de estabelecimento de links por tipos de organizações revelam uma “política de associação” do cotidiano (MARRES; ROGERS, 2000; ROGERS, 2002). Ao caracterizar um ator de acordo com os links de entrada e de saída, nota-se que há alguma divergência quanto aos padrões e, mais genericamente, se links particulares que são recebidos falam pela reputação de um ator. Uma organização não governamental recebendo um link de um site governamental poderia ser interpretado como um impulsionador da reputação.⁴

Além de capturar a micropolítica dos hiperlinks, as análises de links também podem ser usadas em trabalhos mais sofisticados de amostragem. Neste trabalho, a distinção entre métodos digitais e métodos digitalizados se destaca de maneira mais acentuada. O *Open Net Initiative* (ONI) da Universidade de Toronto conduz uma pesquisa sobre a censura na internet construindo listas de websites (contidos em diretórios on-line como o *Open Directory Project* e o *Yahoo*) e depois checando se os sites estão bloqueados em vários países. Esse é um importante trabalho que ilumina o escopo e a infraestrutura técnica das práticas de censura estatal da internet mundo afora (DEIBERT et al., 2007). Nas análises, os sites são agrupados por categorias: blogueiros famosos, sites governamentais, sites de direitos humanos, direitos das mulheres etc. São aproximadamente quarenta categorias. Assim, os padrões de censura podem ser pesquisados por tipo de site entre os países.

A lista completa de websites por país (cerca de 3.000) é uma amostra que cobre, é claro, apenas uma pequena fração de todos os websites de uma categoria específica. Como seria uma amostra de websites em um método que segue o meio,

⁴ O *software Issue Crawler* foi desenvolvido especificamente para permitir essa análise de hiperlinks. Por meio dele, websites são capturados e links são reunidos e armazenados. Os módulos de análise dos rastreamentos são adaptações da cienciometria (análise de correlação de links) e da análise de redes sociais. Após uma rede ser localizada pelo *Issue Crawler*, atores individuais podem ser perfilados usando a ferramenta “*actor profiler*”. Essa ferramenta mostra, em um gráfico, os links de entrada e de saída dos dez principais atores da rede. Outras técnicas para perfilar atores são um extrator que captura todos os links externos de um site e um extrator de um mecanismo de busca, o *Yahoo! inlink ripper*, que provê uma lista das listas feitas para um website.

Figura 1: Técnica de descoberta de censura na internet

Mapa das redes sociais, políticas e religiosas de sites iranianos, mostrando sites acessíveis (azul), bloqueados (vermelho) e novos sites bloqueados (vermelho com alfinetes amarelos). O resultado de amostra de dinâmica de URLs, um método de análise de hiperlinks com auxílio do *Issue Crawler* e descoberta de censura na internet com auxílio da ferramenta *Censorship Explorer*. Análise e gráfico pelo *Govcom.org*, 2006.

Fonte: © Govcom.org Foundation, Amsterdam, 2006.

O website

Até agora, as pesquisas de websites têm sido dominadas por estudos de usuários ou do tipo “rastreamento óptico”, nas quais uma poética da navegação é combinada com ideias sóbrias como “não me faça pensar” (KRUG, 2000; DUNNE, 2005). Muitos dos métodos para estudar websites estão localizados acima dos ombros:⁵ observa-se a navegação ou o uso de um mecanismo de busca e depois se realiza uma entrevista com as pessoas investigadas.

Naquilo que pode ser chamado de abordagem clássica de registro, uma técnica popular é a do rastreamento óptico. Enquanto os sites carregam suas páginas, os olhos dos usuários se movem para o alto e para a esquerda da tela, movimento conhecido como o “triângulo de ouro” da busca. Os resultados são mapas de calor que provem pistas para o redesenho do site e uma referência de valor, para fins comerciais, das diferentes seções de uma página. Outra vertente dominante dos estudos de websites reside em uma análise das características: sites são comparados e constatados com base nos níveis de interatividade, capacidade para *feedback* do usuário etc. (FOOT; SCHNEIDER, 2006). Essas pesquisas indagam se um conjunto particular de recursos resultaria em mais usuários e mais atenção. Nessa tradição, websites são frequentemente arquivados para estudos posteriores. Assim, grande parte do trabalho reside no arquivamento de sites antes da análise.

Uma das tarefas cruciais a seguir é uma reflexão adicional sobre os modos com os quais os sites são capturados e armazenados, bem como o modo de disponibilizar os dados nos quais as descobertas se baseiam. Assim, o programa de pesquisa dos métodos digitais se engaja especificamente com o site como um objeto arquivado, feito acessível, mais legível, por meio do *Wayback Machine* do *Internet Archive*.

⁵ NT: O autor se utiliza da expressão “*over the shoulders*”, indicando um tipo de plano cinematográfico em que algo ou alguém é filmado pela perspectiva ou ângulo da câmera colocada sobre o ombro de outra pessoa.

Que tipos de abordagem de pesquisa são favorecidas pela atual organização de websites pelo *Internet Archive*? Com o *Wayback Machine*, alguém pode estudar a evolução de uma página avulsa (ou de múltiplas páginas) ao longo do tempo, por exemplo lendo ou coletando telas a partir das datas em que a página foi indexada. Como esse arranjo de sites históricos pode ser colocado em uso? Previamente eu mencionei o trabalho de reportagem investigativa feita pelo NRC *Handelsblad* em suas análises sobre o surgimento da linguagem extremista na Holanda. Os jornalistas leram algumas centenas de websites disponíveis no *Internet Archive*, alguns de uma década atrás. Esse é um trabalho que deveria ser melhor baseado em termos metodológicos e técnicos. Alguém poderia capturar, do *Internet Archive*, as páginas dos sites de direita e dos extremistas, colocar o texto (e imagens) em um banco de dados e, sistematicamente, consultá-lo quanto à presença de palavras-chave particulares ao longo do tempo. Como a NRC *Handelsblad* fez, alguém poderia determinar mudanças nas condições sociais por meio de conjuntos particulares de sites arquivados.

Como realizar mais pesquisas com o *Internet Archive*? O programa dos métodos digitais desenvolveu meios de capturar a história de sites capturando telas e as combinando em um filme, como uma fotografia em *time-lapse*. Para demonstrar como usar o *Internet Archive* para capturar essas histórias de evolução, colegas e eu capturamos telas das páginas principais do buscador *Google* de 1998 até o final de 2007. A análise dizia respeito às sutis alterações introduzidas na interface, em particular as abas. Nós descobrimos que o projeto de diretório, a organização da web em tópicos por editores humanos, está em declínio.

Estariam as histórias dos mecanismos de busca, capturadas pelas evoluções das suas interfaces, indicando mudanças em como, de maneira mais geral, a informação e o conhecimento são ordenados? Uma abordagem de estudos comparativos de mídia seria útil, tendo os jornais on-line como um dos casos mais pungentes. Na versão on-line do jornal *The New York Times*, por exemplo, artigos ainda estão posicionados na página principal e nas seções, mas também são listados como “mais enviados” ou “mais compartilhados”, provendo um sistema de recomendação específico do meio para navegar pelas notícias. O impacto dos sistemas de recomendação – o meio dominante pelo qual a informação e o

conhecimento são ordenados – pode também ser estudado por meio das expectativas dos usuários. Estariam os usuários esperando ordenações típicas da web em arquivos, livrarias, centros de informação turística, e outros espaços de consulta à informação e ao conhecimento?

Os mecanismos de busca e as esferas

O estudo de mecanismos de busca foi abalado pela agora famigerada divulgação de dados do mecanismo de busca AOL, em 2006, quando buscas de 650.000 usuários ao longo de três meses foram disponibilizadas on-line, com assustadoras e muitas vezes salientes repercussões na imprensa sobre o nível de detalhe íntimo revelado sobre os usuários da busca, mesmo que seus históricos tenham sido anonimizados (sem nomes) e desvinculados da geografia (sem endereço de IP). Pode-se interpretar as descobertas do caso AOL como uma mudança em como se considerar a presença on-line, se é que esse termo permanece apropriado. Se uma pessoa é *googlada*, os registros criados por ela frequentemente aparecem no topo dos resultados, ou próximo a ele, enquanto o que outros escreveram sobre a pessoa aparece mais abaixo na classificação. No entanto, com as consultas de busca armazenadas, um terceiro conjunto de rastros podem vir à tona para definir um indivíduo: seu histórico de buscas. Isso levanta questões políticas: por quanto tempo uma empresa de busca deveria guardar os históricos de busca? Assim, mecanismos de busca estão sendo estudados em termos legais, especialmente no que tange ao modo como leis de retenção de dados podem ser aplicadas à busca.

Outra vertente nos estudos de mecanismo de busca, resumida no termo “googlização”, é uma crítica político-econômica que considera como o modelo de serviço gratuito da *Google* para o usuário final pode se espalhar pelas indústrias e culturas do *software*. A seguir, eu discuto a crítica e proponho uma pesquisa que considera tanto a googlização do *front-end* quanto do *back-end*. A googlização do *front-end* incluiria o estudo das políticas de informação da interface. Googlização *back-end* diz respeito à ascensão do algoritmo que recomenda as fontes hierarquicamente, tanto com base em tudo que foi clicado ou lido (“mais enviados”) quanto baseado no histórico individual de leitura (“recomendado para você”, como a versão on-line do jornal *The New York Times* faz).

O significado de estudar as novas hierarquias informacionais dos mecanismos de busca deveria ainda ser vista à luz dos estudos de usuário. Um pequeno percentual de usuários configura suas preferências para exibir mais do que dez resultados por página. Em geral, os usuários não olham além da primeira página de resultados e, cada vez mais, eles clicam nos resultados que aparecem no topo (SPINK; JANSEN, 2004). Desse modo, o poder de um mecanismo de busca reside na combinação de suas práticas de classificação (inclusão de uma fonte nos resultados do topo) junto com o aparente “respeito” dos usuários com essa ordenação (não olhando adiante). O modelo do buscador *Google* apoia-se ainda no registro da interatividade, na qual a preferência de um usuário assim como seu histórico são registrados, armazenados e empregados, crescentemente, para fornecer resultados personalizados. Se em algum momento as consultas poderiam retornar as mesmas informações para todos os usuários, agora os resultados são gerados dinamicamente com base no registro das preferências, histórico e localização destes.

As diferentes ordenações de fontes pelos mecanismos são pouco estudadas, em grande parte porque não são armazenadas e disponibilizadas para pesquisa, com exceção da liberação de dados pela AOL ou outros acordos negociados com empresas de busca. Uma vez, a *Google* disponibilizou uma API (*Application Programming Interface*) que permitiu a coleta de dados. Com isso, um número limitado de consultas poderia ser feito por dia e os resultados reaproveitados. Pesquisadores que confiaram na API foram “furados” pela *Google* quando esta descontinuou ou “depreciou” esse serviço ao final de 2006. Na reintrodução desse serviço, ofertado de maneira diferente em 2009, a *Google* enfatizou, no entanto, que buscas automatizadas e o armazenamento permanente de resultados eram contra os termos de serviço. Como estudar um mecanismo de busca nessas condições? Colegas e eu mineramos os dados da *Google* e colocamos um aviso apreciando a tolerância do buscador *Google*.⁶

O que poderia ser encontrado nos resultados da busca do *Google*? Como observei, os mecanismos de busca, um ponto de entrada crucial na web, são máquinas epistemológicas na medida em que rastreiam, indexam, armazenam em

⁶ O aviso aparece na página de créditos da ferramenta *Issue Dramaturg*, disponível em: <<https://dramaturg.issuecrawler.net/>>.

cache e, em última instância, ordenam o conteúdo. Anteriormente, eu descrevi a web, particularmente a web baseada nos mecanismos de busca, como um espaço de potencial colisão para relatos alternativos da realidade (ROGERS, 2004). O fraseado foi construído sobre o trabalho do sociólogo C. Wright Mills, que caracterizou o propósito da pesquisa social como “nada menos do que apresentar definições conflitantes da própria realidade” (MILLS, 1971, p. 212). Os mecanismos de busca colocam relatos alternativos da realidade lado a lado ou os resultados se alinham com o oficial e o *mainstream*? Armazenar e analisar os resultados dos mecanismos de busca pode responder a essas questões. É importante apontar que as posições superiores das buscas são altamente procuradas; organizações fazem uso de técnicas de otimização de busca para aumentar a visibilidade de seus sites. Existem técnicas “mocinhas” e “vilãs” com essa finalidade, ou seja, aquelas aceitas pelos mecanismos e outras que induzem os motores para retirar os websites dos resultados até que eles cumpram novamente com as regras do mecanismo.

No projeto *Issue Dramaturg* (discutido no capítulo 4 do livro *Digital Methods*), colegas e eu armazenamos os resultados da busca do *Google* para a consulta do termo [9/11], assim como outras palavras-chave, indagando sobre as hierarquias da fonte. Que fontes são privilegiadas? Quais estão “vencendo” a competição para ser as fontes situadas no topo de algumas consultas? Outro propósito foi mapear fontes particulares, na abordagem de estudos de mecanismos de busca que nomeei como “distância da fonte”. Para a consulta do termo [9/11], quão distante do topo dos resultados de busca estão atores significativos nos relatos do 11 de setembro, como a prefeitura da cidade de Nova York e o jornal *The New York Times*? Essas fontes são proeminentes ou elas aparecem lado a lado com fontes que desafiam as visões mais oficiais e mais familiares? Assim, além da prefeitura de Nova York e o do *The New York Times*, outro ator que monitoramos foi o “9/11 truth movement” (911truth.org).

Entre março e setembro de 2007, o site do movimento apareceu entre os cinco resultados mais populares para a consulta do termo [9/11], enquanto a prefeitura e o jornal estavam bem abaixo do quinto lugar. Meados de setembro de 2007, período próximo ao aniversário desse acontecimento, foi dramático. O site 911truth.org caiu rapidamente para a 200ª posição e, subsequentemente, saiu dos 1000 primeiros

lugares, que é número máximo de resultados fornecidos pelo *Google*. Eu acredito que esse é um dos primeiros casos totalmente documentados de aparente remoção de um website pelo Google – de um lugar entre os cinco primeiros durante seis meses para fora do *ranking* dos 1.000 primeiros resultados. O caso levanta questões sobre a estabilidade e volatilidade dos resultados dos mecanismos de busca e apresenta uma área de estudo.

Apesar de ser dominante, há outros mecanismos de busca além do *Google*. O que é menos apreciado, talvez, é que existem outras ferramentas para temas ou esferas particulares da web. Para a blogosfera, existe o *Technorati* e o *Google Blog Search*; para o âmbito noticioso, o *Google Notícias* e, para os favoritos, on-line, o *Delicious*. De fato, pensar a web em termos de esferas se refere inicialmente ao nome de uma das mais conhecidas, a blogosfera, assim como à pesquisa que procura definir outra, a esfera web (FOOT; SCHNEIDER, 2002). A “esfera” na “blogosfera” refere-se, em espírito, à esfera pública; ela também sugere a forma geométrica, na qual todos os pontos da superfície estão à mesma distância do centro ou do núcleo. Pode-se pensar nessa equidistância como um ideal igualitário, no qual todo *blog*, ou mesmo toda fonte de informação, pode vir a estar conectado ao núcleo e vice-versa.

Foi descoberto, no entanto, que certas fontes são centrais na web. Eles recebem uma vasta maioria de links e de visitas. Seguindo o princípio de que o rico fica mais rico (também conhecido como Efeito de Matthew e distribuição *power law*), os sites que já recebem atenção tendem a conquistar mais. A distância entre o centro e outros nós pode apenas crescer, com a esfera idealizada se tornando uma ficção, porém em algo mais útil. Eu gostaria de levar adiante a abordagem que muda a questão da distância do centro para a periferia e a operacionaliza como uma mensuração das diferenças entre *rankings* de fontes por esfera. Análise esférica é um método digital para mensurar e aprender a partir da distância entre fontes em diferentes esferas na web.

Conceitualmente, uma esfera pode ser considerada um conjunto de recursos relacionados ao dispositivo, isto é, o puro *PageRank* de todas as fontes na web (sites mais influentes pelos links recebidos) ou na verdade *PageRanks* análogos de todas as fontes calculadas pelos mecanismos dominantes por esfera (*Technorati*, *Google New*

e *Delicious*). Assim, para estudar uma esfera, eu proponho primeiro permitir que os mecanismos possam demarcá-lo. Na análise de esferas, considera-se que fontes são mais influentes, não apenas em termos gerais, mas por consulta. A análise entre esferas compara as fontes recuperadas em cada esfera em uma mesma consulta, o que pode depois ser visto como uma pesquisa de *rankings* comparativos. Mais importante: com a análise entre esferas, pode-se pensar nas consequências do modo como cada mecanismo trata os links, recarregamento automático de páginas, *tags* etc. Fontes particulares tendem a estar no núcleo de uma esfera e não em outras? Que comparações entre fontes, e distâncias entre fontes pelas esferas nos dizem sobre a qualidade da nova mídia? O que elas nos dizem sobre atuais os compromissos informacionais em culturas particulares?

Em uma análise preliminar, colegas e eu estudamos quais animais são mais associados às mudanças climáticas na web de língua inglesa, nas notícias e na blogosfera. Nós descobrimos que a esfera web tinha o conjunto mais diverso de animais associados ao tema. A esfera de notícias favorecia o urso polar e a blogosfera ampliou, ou tornou mais proeminente, a seleção da esfera noticiosa. Concluímos aqui, de maneira cautelosa, que a esfera web pode ser menos propensa à criação de ícones midiáticos do que a esfera de notícia, o que tem implicações para os estudos de mídia que tomam como ponto de partida uma cultura da publicidade. A blogosfera pareceu, além disso, parasitar mais o noticiário do que oferecer uma alternativa a ele.

As webs

Como mencionado anteriormente, a pesquisa em internet tem sido assombrada pela divisão entre virtual e real. Uma das razões para essa divisão diz respeito às disposições técnicas da internet e como elas foram associadas como um domínio virtual, o ciberespaço. Assim, fazia sentido fazer alguma distinção, em termos tecnológicos, sobre o ciberespaço (CHUN, 2006). Os princípios e os protocolos, particularmente a comutação de pacotes e o princípio das pontas (*end-to-end principle*), inicialmente levaram à compreensão da noção de ciberespaço como um domínio livre de restrições físicas. A indiferença técnica da internet para a localização geográfica de seus usuários gerou ideias de desespacialização; em sua arquitetura, ela teria sido supostamente feita para uma libertação dos espaços dos

estados-nação e suas formas divergentes de tratar os fluxos de informação (MARVIN, 2004).

Podemos lembrar o famoso comentário atribuído a John Gilmore, cofundador da *Electronic Frontier Foundation* juntamente com John Perry Barlow: “a internet trata a censura como um defeito e se desvia dela” (BOYLE, 1997). A geografia, no entanto, foi incorporada ao ciberespaço desde o início, se considerarmos as localizações dos treze servidores de raiz originais, a distribuição desigual do tráfego por país, assim como a atribuição de endereços de IP em intervalos, que depois permitiram a aplicação da tecnologia de localização de endereços geo-IP para fins de propagandas e direitos autorais. A tecnologia geo-IP e outras tecnologias locativas também podem ser usadas para a pesquisa que toma a internet como um espaço de estudos e que questiona o que pode ser aprendido sobre as condições sociais entre países. No programa de pesquisa em métodos digitais, colegas e eu nomeamos esse trabalho de “estudos da web nacional” (*national web studies*).

Discuti acima a pesquisa feita por etnógrafos britânicos que fundamentaram (*grounded*) o ciberespaço por meio do trabalho empírico sobre como os usuários de internet no Caribe se apropriaram do meio para adequá-lo às suas práticas culturais. Este é, sem dúvida, um exemplo de estudos da web nacional, embora usando métodos de observação externos ao meio. Para estudar a web em termos nacionais, pode-se ainda investigar os dados que são coletados rotineiramente, por exemplo por grandes empresas como a classificação dos sites mais acessados disponibilizada pelo site *Alexa*, de acordo com o tráfego por país. Que sites são mais visitados por país e o que a visita de sites diz sobre a cultura informacional de um país?

O site *Alexa* foi pioneiro na coleta de dados de registro com a barra de ferramentas instaladas pelos usuários em seus navegadores. A barra fornecia estatísticas sobre o site acessado pelo navegador, assim como sobre seu recarregamento automático. Todos os websites que o usuário carregou ou nos quais navegou também seriam registrados e, então, comparados com as URLs que já estavam na base de dados do site *Alexa*. As URLs que não estavam na base de dados seriam rastreadas e buscadas. Assim, nasceu em 1996, o *Internet Archive*.

Esse projeto foi desenvolvido durante o período da história da internet que se pode chamar de ciberespaço. Para ilustrar a diferença no *design* e na concepção entre o *Internet Archive* e os arquivos da web nacional que estão brotando em muitos países, pode-se apontar que o primeiro foi construído para a navegação em sites – um tipo de uso da internet que abriu caminho para a busca (SHIRKY, 2005). No *Wayback Machine* do *Internet Archive* pode-se digitar uma simples URL, ver páginas disponibilizadas e navegar por elas. Se alguém acessa um link externo, o *Internet Archive* identifica a página mais próxima da data do site do qual se está saindo e a recarrega. Se nenhuma versão do site existe no *Internet Archive*, ele se conecta com a versão atual do website. É a continuidade do fluxo, de website para website, que é preservada (GALLOWAY, 2004). Arquivos nacionais da web, por outro lado, deixaram de pensar na web em termos de ciberespaço. Seus respectivos propósitos são preservar as webs nacionais. Com o propósito de contribuir metodologicamente com a pesquisa em internet, a questão inicial é: como seria demarcada uma web nacional?

Na Biblioteca Nacional da Holanda, por exemplo, a abordagem é similar a dos pesquisadores de censura na internet discutida acima. É um método digitalizado, isto é, um modelo de diretório no qual um especialista escolhe sites significativos com base em critérios editoriais. Esses sites são continuamente arquivados com tecnologia originalmente desenvolvida no projeto *Internet Archive*. No momento de escrita deste trabalho, aproximadamente 998 websites nacionais eram arquivados na Holanda – muito longe do que é salvo no *Internet Archive*. Ao explicar a diferença nas abordagens e resultados dos dois projetos, eu gostaria de observar que o fim do virtual e o fim do ciberespaço não foram gentis com o arquivamento da web. O retorno do estado-nação e a aplicação de certos regimes políticos (especialmente direitos autorais) têm reduzido drasticamente os esforços de arquivamento. Os métodos digitais poderiam ajudar a corrigir essa situação? Eu gostaria de encorajar arquivistas de webs nacionais a considerar uma abordagem baseada nos registros, isto é, o modelo *Alexa* adaptado para o contexto nacional. Os resultados podem ser salutares.

Sites de mídia social e pós-demografia

Sites de rede social como o *Dutch Hyves*⁷ e seu precursor norte-americano, o *Facebook*, são plataformas populares pelas oportunidades que oferecerem para a pesquisa social e cultural. Até agora, as principais pesquisas focaram em como usuários se apresentam e gerenciam suas identidades e privacidade, bem como a amizade on-line se relaciona com ser amigo no sentido tradicional (BOYD; ELLISON, 2007). Trabalhos de larga escala, com conjuntos de *big data*, comparam redes on-line com redes off-line ou outras formas pré-existentes de redes sociais. Outra abordagem, que tem origem na ciência da informação e da computação, busca usar as enormes quantidades de dados que as pessoas disponibilizam em seus perfis. As informações pessoais de cada perfil contêm dados demográficos tradicionais como gênero, idade e localização, mas também informações pós-demográficas como interesses, gosto musical, livros favoritos e programas de TV. Assim, sites de rede social oferecem novas oportunidades para pesquisa e, talvez, especialmente para pesquisa sobre públicos. Enquanto a pesquisa de opinião pública foi associada com *surveys* e as avaliações e *shares* de audiência televisiva, poderia a informação contida nos perfis de redes sociais on-line oferecer diferentes tipos de *insights* para a composição e caracterização dos públicos? A questão diz respeito a que formas de análise podem ser realizadas, inclusive de modo ético.⁸

Para este trabalho, eu escolhi uma abordagem experimental, situada mais na tradição das artes, de modo a criar perfis agregados e compará-los, por exemplo, quanto aos interesses e aos gostos midiáticos dos amigos do presidente Barack Obama e do senador John McCain. Será que os “amigos” de Obama e McCain assistem aos mesmos programas de TV, leem os mesmos livros e têm outros interesses similares? Assim como sexo, idade, localização e outra informação demográfica, pode-se adicionar o consumo de mídia e outros interesses, descrevendo públicos de modo a mostrar similaridades quando a expectativa é de divisão, por exemplo.

⁷ Site de rede social sediado na Holanda (<https://en.wikipedia.org/wiki/Hyves>), descontinuado em 2013.

⁸ Tema discutido no capítulo de conclusão do livro *Digital Methods*.

Eu exploro ainda os perfis dos nomes de usuários, isto é, os apelidos que as pessoas escolhem junto aos seus nomes reais ao fazer contas on-line e assinaturas de serviços. Tipicamente uma pessoa tem dois nomes na web, seu próprio nome e um apelido. Que combinações de serviços um nome de usuário ou agregados de nomes de usuários usam? Aqui está o domínio de pesquisa em temas relacionados, os quais os métodos da web frequentemente usam para recomendar produtos, serviços, informação e amigos, mas que podem ser repropostos.

Colegas e eu desenvolvemos ferramentas que se situam além de sites de redes sociais (como o *MySpace*) e ferramentas relacionadas às mídias sociais (como *User Name Check*) para mostrar como os resultados das consultas podem ser repropostos. O site *Elfriendo.com* é recuperado pelas ferramentas antes do *MySpace* e cria perfis agregados com base em um conjunto de amigos neste site. O site *Leakygarden.net* cria uma lista de serviços web para o qual um nome de usuário ou apelido se cadastrou, criando assim um metaperfil de um usuário. As duas ferramentas se esforçam para mostrar composições de públicos de acordo com suas preferências, em oposição à demografia, abrindo assim uma linha de investigação chamada pós-demografia.

Wikipédia e o conteúdo em rede

Até o momento, as abordagens para o estudo da Wikipédia seguiram certas características da enciclopédia on-line, as quais parecem contraintuitivas à primeira vista. Uma dessas características é que a Wikipédia é editada pelos chamados amadores e, ainda assim, é surpreendentemente enciclopédica, não apenas em sua forma, mas também em sua precisão (GILES, 2005). O maior debate sobre a qualidade da Wikipédia frente à *Encyclopaedia Britannica* trouxe à tona questões relevantes para os métodos digitais, na medida em que o modelo coletivo de edição permitido pela web desafiou o trabalho de um conjunto de especialistas. No entanto, pesquisas identificaram que há apenas uma pequena proporção de editores dentre os usuários das plataformas da web 2.0, incluindo a Wikipédia, o que ilustra o que ficou conhecido com o mito do conteúdo gerado por usuário (SWARTZ, 2006).

O cofundador da Wikipédia, Jimmy Wales, tem observado com frequência que a comunidade dedicada ao projeto é de fato relativamente pequena, com pouco mais

de 500 membros. Assim, o pequeno grupo de editores da Wikipédia poderia ser considerado uma nova elite. Um exercício de pesquisa consiste, portanto, em relativizar as alegadas diferenças entre amadores e especialistas por exemplo por meio da realização de um estudo de demografia de wikipedistas (VAN DIJCK, 2009). Outro aspecto contra-intuitivo da Wikipédia é o fato de editores não remunerados ainda serem comprometidos e altamente vigilantes. A vigilância da multidão, como é denominada, é parte de um aspecto mítico de uma rede de produção de qualidade até que alguém considere como essa vigilância é efetivada.

Quem está editando? Uma abordagem para essa questão está no projeto *Wikiscanner*, desenvolvido a partir de 2007 no *California Institute of Technology* por Virgil Griffith. Esse projeto identifica editores anônimos identificando seu endereço IP e o cruzando com um banco de dados com as localizações dos endereços de IP (tecnologia IP-to-Geo). Para Griffith, a qualidade da Wikipédia é assegurada escandalizando editores que fazem mudanças de interesse próprio, como um membro da família real holandesa que “embelezou” um artigo e acabou na primeira página do jornal depois que um jornalista usou a ferramenta *Wikiscanner*.

A fim de manter aberto o debate sobre a qualidade, de que outro modo os vândalos são mantidos à distância na Wikipédia, incluindo os interessados em experimentações e os pesquisadores que fazem alterações erradas em um artigo ou criam um artigo ficcional (CHESNEY, 2006; READ, 2006; MAGNUS, 2008)? Colegas e eu contribuimos para a pesquisa sobre a qualidade da Wikipédia ao introduzirmos termos como “conteúdo em rede” e a “tecnicidade do conteúdo”. O primeiro tema se refere ao conteúdo trabalhado conjuntamente por autores humanos e vigias não humanos, incluindo *bots* e alertas que revertem edições ou notificam wikipedistas sobre alterações realizadas nesse site. O segundo termo se concentra nos *bots*, deixados de lado pela maioria das análises a respeito da Wikipédia, como os estudos que se concentram na vigilância da multidão. Ao olhar para as estatísticas sobre o número de edições por usuário, é notório que os *bots* são, de longe, os maiores editores. A implicação, que tem sido pesquisada no programa dos métodos digitais, é que os *bots* e os *softwares* de alerta são agentes significativos de vigilância (NIEDERER; VAN DIJCK, 2010).

Como o projeto *Wikiscanner* e as estatísticas dos *bots* nos lembram, a Wikipédia é uma síntese de atividades e eventos em rede registrados e disponibilizados como grandes conjuntos de dados. A Wikipédia tem também um processo interno reflexivo, como mostra o processo pelo qual a criação de um novo artigo é aprovada, algo que se perdeu nas enciclopédias e outros projetos já finalizados. Pode-se estudar o processo de maturação de um artigo, pois as informações estão no histórico de edições, mas também na página de discussão do artigo e, talvez, em seu histórico de disputas, suas proteções e reaberturas. Outra abordagem para os dados da Wikipédia se basearia nos registros de edição de um ou mais artigos e em outra proposição de *insights* técnicos do *Wikiscanner* olhando para os trechos em que as edições foram feitas. Os “locais de edições mostram que temas suscitam preocupações e demandam *expertise* por organização e por país.

O fim do virtual: fundamentar as reivindicações on-line

Meu objetivo é colocar em andamento uma transformação em como e porque são feitas pesquisas com a internet. O primeiro passo é mover a discussão para fora das limitações do virtual (o quanto a cultura e a sociedade estão on-line) para as limitações do método atual (como estudar cultura e sociedade e fundamentar as descobertas com a internet).

Gostaria de concluir com uma breve discussão sobre essas limitações da pesquisa em internet assim como com uma proposta de renovação. Aponto primeiro que o fim do ciberespaço e sua desespacialização e o fim do virtual como um domínio à parte é lamentável – para algumas abordagens e projetos de pesquisa em particular. De certo modo, a divisão entre real e virtual serviu a práticas específicas de pesquisa.⁹ Anteriormente, eu mencionei que o arquivamento da internet floresceu no ciberespaço e, mais recentemente, sofreu sem ele. Se o ciberespaço tornou uma vez possível a ideia de um arquivamento massivo de websites, a web fundamentada nos dados e as webs nacionais estão hoje diminuindo as coleções.

⁹ O fim do ciberespaço também não foi bom para os projetos que se apoiam na característica clássica do usuário anônimo. Por exemplo, organizações e governos proibiram funcionários de editar a Wikipédia no trabalho, uma vez que as localizações das edições podem ser rastreadas e torná-los sujeitos de escândalos.

Na verdade, eu argumentaria que é possível se aprender a partir dos métodos utilizados nos meios, movendo a discussão da teoria sobre a especificidades dos meios de uma ontologia (propriedades e características) para uma epistemologia (método). A internet e a web, mais especificamente, têm seus objetos ontológicos, como os links e as *tags*. Uma epistemologia da web é, entre outras coisas, o estudo de como esses objetos digitais são manejados pelos dispositivos. Os *insights* de tal tipo de estudo conduzem para importantes distinções metodológicas, assim como para *insights* sobre o propósito da pesquisa em internet.

Quando a diferenciação metodológica é considerada, poder-se-ia ver os métodos atuais da internet tanto como aqueles que seguem os meios (e as técnicas dominantes empregadas em produzir ou organizar informação, conhecimento ou sociabilidade) quanto como aqueles que remédiam ou digitalizam métodos já existentes. A diferença no método pode ter resultados significativos. Uma razão para a diminuição nos esforços de arquivamento da web pode estar na escolha de um método digitalizado (seleção editorial) em detrimento de um digital (coleção de dados de registro), como aqueles empregados no projeto original do *Internet Archive*, em que os sites pelos quais usuários navegavam eram gravados. Na verdade, eu empreguei o termo “métodos digitais” para que pesquisadores considerem o valor e os resultados de uma abordagem sobre a outra. Como um dos casos em questão, a escolha de uma amostra de uma URL dinâmica ao invés de um modelo editorial poderia beneficiar a pesquisa sobre censura na internet.

Em terceiro lugar, finalmente, eu gostaria de argumentar que a internet é um espaço de pesquisa que vai além da cultura on-line e de seus usuários. Com o fim da divisão entre real e virtual, a internet pode ser repensada como uma fonte de dados sobre a sociedade e a cultura. Coletar e analisar esses dados para a pesquisa social e cultural requer não apenas uma nova perspectiva sobre a internet, mas também novos métodos para fundamentar as descobertas. Fundamentar as reivindicações nas especificidades do on-line é uma mudança maior no propósito da pesquisa em internet, na medida em que não se trata tanto de pesquisar a internet e seus usuários, mas sim estudar a cultura e a sociedade com a internet.

Referências

- BARLOW, John P. **A Declaration of the Independence of Cyberspace**. Davos, Switzerland, 1996. Disponível em: <<https://projects.eff.org/~barlow/Declaration-Final.html>>. Último acesso em: 19 out. 2012.
- BEAULIEU, Anne. Sociable Hyperlinks: An Ethnographic Approach to Connectivity. In: HINE, Christine (Ed.). **Virtual Methods: Issues in Social Research on the Internet**. Oxford: Berg, 2005. p. 183 a 197.
- BENEDIKT, Michael. Cyberspace: Some Proposals. In: BENEDIKT, Michael (Ed.). **Cyberspace**. Cambridge, MA: MIT Press, 1991. p. 119-224.
- BOLTER, Jay D., GRUSIN, Richard. **Remediation: Understanding New Media**. Cambridge, MA: MIT Press, 1999.
- BOYD, Danah; ELLISON, Nicole. Social Network Sites: Definition, History, and Scholarship. **Journal of Computer-Mediated Communication**, 2007, v. 13 - n. 1.
- BOYLE, James. Foucault in Cyberspace. **University of Cincinnati Law Review**, 1997, v. 66 - p. 177 a 205.
- BRUNS, Axel. **Blogs, Wikipedia, Second Life, and Beyond: From Production to Produsage**. New York: Peter Lang, 2008.
- CASTELLS, Manuel. **The Information Age: Economy, Society and Culture – The Rise of the Network Society**. Malden, MA: Blackwell, 1996.
- Centers for Disease Control and Prevention. **Overview of Influenza Surveillance in the United State**. Atlanta, GA, 2011. Disponível em: <<http://www.cdc.gov/flu/weekly/overview.htm>>. Último acesso em: 18 out. 2012.
- CONTRACTOR, Noshir. Digital Traces: An Exploratorium for Understanding and Enabling Social Networks. Presentation at the annual meeting of the **American Association for the Advancement of Science**. 2009.
- CHESNEY, Thomas. An Empirical Examination of Wikipedia's Credibility. **First Monday**, 2006, v. 11 - n. 11.
- CHUN, Wendy. **Control and Freedom: Power and Paranoia in the Age of Fiber**. Cambridge, MA: MIT Press, 2006.
- DEIBERT, Ronald J. et al. (Eds.). **Access Denied: The Practice and Policy of Global Internet Filtering**. Cambridge, MA: MIT Press, 2007.
- DIBBELL, Julian. **My Tiny Life: Crime and Passion in a Virtual World**. New York: Henry Holt, 1998.
- DOHMEN, Joep. Opkomst en ondergang van extreemrechtse sites. **NRC Handelsblad**, 25 ago. 2007.
- DUNNE, Anthony. **Hertzian Tales: Electronic Products, Aesthetic Experience, and Critical Design**. Cambridge, MA: MIT Press, 2005.
- ELMER, Greg. **Hypertext on the Web: The Beginnings and Ends of Web Pathology**. Space and Culture, v. 10, 2001, p. 1 a 14.

-
- FOOT, Kirsten; SCHNEIDER, Steven. Online Action in Campaign 2000: An Exploratory Analysis of the U.S. Political web Sphere. **Journal of Broadcasting and Electronic Media**, 2002, v. 46 - n. 2 - p. 222 a 244.
- FOOT, Kirsten; SCHNEIDER, Steven. **Web Campaigning**. Cambridge, MA: MIT Press, 2006.
- FULLER, Matthew. **Behind the Blip: Essays on the Culture of Software**. Brooklyn: Autonomedia, 2003.
- GALLOWAY, Alexander. **Protocol: How Control Exists after Decentralization**. Cambridge, MA: MIT Press, 2004.
- GILES, Jim. Internet Encyclopedias Go Head to Head. **Nature**, 2005, v. 438 - n. 7070 - p. 900 a 1000.
- GOLDSMITH, Jack, WU, Tim. **Who Controls the Internet? Illusions of a Borderless World**. New York: Oxford, 2006.
- HAYLES, N. K. Print Is Flat, Code Is Deep: The Importance of Media-Specific Analysis. **Poetics Today**, 2004, v. 25 - n.1 - p. 67 a 90.
- HINE, Christine. **Virtual Ethnography**. London: Sage, 2000.
- INTRONA, Lucas; NISSENBAUM, Helen. Shaping the Web: Why the Politics of Search Engines Matters. **Information Society**, 2000, v. 16 - n. 3- p. 1 a 17.
- JEANNENEY, Jean-Noel. **Google and the Myth of Universal Knowledge: A View from Europe**. Chicago: University of Chicago Press, 2007.
- JENSEN, Jens F. "Interactivity": Tracking a New Concept in Media and Communication Studies. **Nordicom Review**, 1998, v. 19 - n.1 - p. 185 a 204.
- JENKINS, Henry. **Convergence Culture: Where Old and New Media Collide**. New York: NYU Press, 2006.
- JONES, Steve. Studying the Net: Intricacies and Issues. In: JONES, Steve (Ed.). **Doing Internet Research: Critical Issues and Methods for Examining the Net**. London: Sage, 1999. p. 1 a 28.
- KEEN, Andrew. **The Cult of the Amateur: How Today's Internet Is Killing Our Culture**. London: Nicholas Brealey, 2007.
- KREBS, Valdis. Mapping Networks of Terrorist Cells. **Connections**, 2002, v. 24 - n. 3 - p. 43 a 52.
- KRUG, Steve. **Don't Make Me Think! A Common Sense Approach to Web Usability**. Indianapolis: New Riders, 2000.
- LANDOW, George. **Hyper/Text/Theory**. Baltimore: Johns Hopkins University Press, 1994.
- LAZER, David et al. Computational Social Science. **Science**, 2009, v. 323 - n. 5915 - p. 721 a 723.
- LYNCH, Michael. A Sociology of Knowledge Machine. **Ethnographic Studies**, 1997, n. 2, p. 16 a 38.

-
- MCLUHAN, Marshall. **Understanding Media: The Extensions of Man**. New York: McGraw Hill, 1964.
- MAGNUS, P. D. Early Response to False Claims in Wikipedia. **First Monday**, 2008, v. 13 - n. 9.
- MANOVICH, Lev. **Cultural Analytics**. Unpublished ms., 2007. Disponível em: <http://www.manovich.net/cultural_analytics.pdf>. Último acesso em: 28 jan. 2009.
- MANOVICH, Lev. **Software Takes Command**. Unpublished ms., 2008. Disponível em: <<http://www.manovich.net/>>. Último acesso em: 10 abr. 2009.
- MARRES, Noortje; ROGERS, Richard. Depluralising the Web, Repluralising Public Debate: The GM Food Debate on the Web. In: ROGERS, Richard (Ed.). **Preferred Placement**. Maastricht: Jan van Eyck Editions, 2000. p. 113 a 135.
- MARRES, Noortje; ROGERS, Richard. Subsuming the Ground: How Local Realities of the Ferghana Valley, Narmada Dams and BTC Pipeline Are Put to Use on the Web. **Economy and Society**, 2008, v. 37 - n. 2 - p. 251 a 281.
- MARVIN, Carolyn. Peaceable Kingdoms and New Information Technologies: Prospects for the Nation-State. In: STURKEN, Marita; THOMAS, Douglas; BALL-ROKEACH, Sandra J. (Eds.). **Technological Visions: The Hopes and Fears That Shape New Technologies**. Philadelphia: Temple University Press, 2004. p. 240 a 254.
- MILLER, Daniel; SLATER, Don. **The Internet: An Ethnographic Approach**. Oxford: Berg, 2000.
- MILLS, C. W. **The Sociological Imagination**. Harmondsworth: Penguin, 1971.
- NIEDERER, Sabine; VAN DIJCK, José. Wisdom of the Crowd or Technicity of Content? Wikipedia as Socio-technical System. **New Media and Society**, 2010, v. 12 - n. 8 - p. 1368 a 1387.
- PARK, Han W.; THELWALL, Mike. Hyperlink Analyses of the World Wide Web: A Review. **Journal of Computer-Mediated Communication**, jul. 2003, v. 8 - n. 4.
- READ, Brock. Can Wikipedia Ever Make the Grade? **Chronicle of Higher Education**, 2006, v. 53 - n. 10.
- RHEINGOLD, Howard. **The Virtual Community: Homesteading on the Electronic Frontier**. Reading, MA: Addison-Wesley, 1993.
- RHEINGOLD, Howard. **Virtual Reality: Exploring the Brave New Technologies**. New York: Summit, 1991.
- ROGERS, Richard. **Information Politics on the Web**. Cambridge, MA: MIT Press, 2004.
- ROGERS, Richard. Operating Issue Networks on the Web. **Science as Culture**, 2002, v. 11 - n. 2 - p. 191 a 214.
-

-
- ROGERS, Richard. The Viagra Files: The Web as Anticipatory Medium. **Prometheus**, 2003, v. 21 - n. 2 - p. 195 a 212.
- SHAVIRO, Steven. Money for Nothing: Virtual Worlds and Virtual Economies. In: IPE, Mary (Ed.). **Virtual Worlds**. Hyderabad: Icfai University Press, 2008. p. 53 a 67.
- SHIRKY, Clay. Ontology Is Overrated: Categories, Links, and Tags. **The Writings of Clay Shirky**, 2005. Disponível em: <http://www.shirky.com/writings/ontology_overrated.html>. Último acesso em: 28 jan. 2009.
- SLEE, Mark. Opening More Control for Everyone. **The Facebook Blog**, 16 mar. 2009. Disponível em: <<https://blog.facebook.com/blog.php?post=60186587130>>. Último acesso em: 21 fev. 2012.
- SPINK, Amanda; JANSEN; Bernard J. **Web Search: Public Searching on the Web**. Dordrecht: Kluwer, 2004.
- STONE, Allucquère R. **The War of Desire and Technology at the Close of the Mechanical Age**. Cambridge, MA: MIT Press, 1995.
- SWARTZ, Aaron. Who Writes Wikipedia? Blog post, **Raw Thoughts**, 4 set. 2006. Disponível: <<http://www.aaronsw.com/weblog/whowriteswikipedia>>. Último acesso em: 18 out. 2012.
- TURKLE, Sherry. **Life on the Screen: Identity in the Age of the Internet**. New York: Simon and Schuster, 1995.
- VAIDHYANATHAN, Siva. Where is This Book Going? **The Googlization of Everything Blog**, 25 set. 2007. Disponível em <http://www.googlizationofeverything.com/2007/09/where_is_this_book_going.php>. Último acesso em: 22 dez. 2008.
- VAIDHYANATHAN, Siva. **The Googlization of Everything**. Berkeley: University of California Press, 2011.
- VAN DIJCK, José. Users Like You? Theorizing Agency in User-Generated Content. **Media Culture and Society**, 2009, v. 31 - n. 1 - p. 41 a 58.
- WATTS, Duncan. **Small Worlds**. Princeton: Princeton University Press, 1999.
- WILLIAMS, Raymond. **Television: Technology and Cultural Form**. London: Fontana, 1974.
- WOOLGAR, Steve. Five Rules of Virtuality. In: WOOLGAR, Steve (Ed.). Virtual Society? **Technology, Cyberbole, Reality**. Oxford: Oxford University Press, 2002. p. 1 a 22.