

# Sistemas de recomendação de notícias nas mídias sociais buscam substituir o *gatekeeping* dos meios de comunicação de massa

*Recommender systems of news in social media intend to substitute the gatekeeping of mass media communication*

**Walter Teixeira Lima Junior**

Pós-doutor em Comunicação e Tecnologia, pela Universidade Metodista de São Paulo - Umesp; doutor em Jornalismo Digital, pela Escola de Comunicações e Artes da Universidade de São Paulo - ECA/USP; docente do Programa de Pós-Graduação da Faculdade Cásper Líbero.

E-mail: [digital@walterlima.jor.br](mailto:digital@walterlima.jor.br)

**Cláudia Ferraz Castelo Branco**

Jornalista; mestre em Jornalismo, pela Faculdade Cásper Líbero.

**Pablo Barbosa**

Graduado em Comunicação Social/Jornalismo, pela Faculdade de Comunicação da Universidade Federal da Bahia - Facom/UFBA.

Artigo recebido em: 10/10/2009

Artigo aceito em: 03/11/2009

## Resumo

O presente trabalho realiza uma análise sobre as possibilidades de auto-organização da informação, por meio do sistema de recomendação, como utilizado pelo *Digg*. O estudo compara teoricamente os processos de decisão, que compõem a cadeia de produção de material jornalístico nos meios de comunicação de massa (*gatekeeping*), com os sistemas mediados por computador (mídias sociais) que utilizam filtros colaborativos, identificando suas potencialidades e inconsistências.

**Palavras-chave:** jornalismo, mídia social, sistema de recomendação, colaboração, *gatekeeping*.

## Abstract

The paper makes an analysis about the possibilities of self-organization of information, through the Recommender System, as used by *Digg*!. It compares theoretically the decision-making up the production chain of the articles in the mass media (*gatekeeping*), with computer-mediated systems (social media) that use collaborative filtering, identifying their strengths and inconsistencies.

**Keywords:** journalism, social media, recommender system, collaboration, *gatekeeping*.

### Introdução

Os meios de comunicação de massa (MCM), que têm no jornal impresso o seu modelo mais antigo, com mais de 400 anos, desenvolveram e refinaram os seus processos de produção de notícias ao longo de décadas. Através do método empírico, as teorias sobre os processos de comunicação de massa têm se consolidado por intermédio da observação de como esses processos se comportaram e se adaptaram às mudanças culturais e às novas inserções tecnológicas. Esse modelo midiático, que pode ser denominado como analógico, estruturou uma forma de relacionamento com a sua audiência no que concerne ao problema da interação. A questão da transmissão da informação de um emissor para muitos – ou seja, o modelo clássico emissor à meio à receptor – sedimentou uma forma de relacionamento entre a mídia e sua audiência, pois não havia a necessidade de considerar o retorno da audiência para a remodelação do conteúdo. O retorno é só considerado por meio de sistemas que baseiam os modelos de negócio em função do aporte publicitário, como a medição de índices de audiência (*people meter*), o índice de verificação de circulação (IVC) e as pesquisas de audiência e de níveis de leitura (entrevistas).

Sem a possibilidade do impacto do retorno da audiência, até porque os canais de retorno dos veículos, ditos tradicionais, são bem “estreitos” para que o *feedback* aconteça, como a sessão de cartas ao leitor (jornal e revista) ou via telefone (rádio e televisão), os processos internos de produção do conteúdo informativo de relevância social (notícias) se moldaram dentro de um sistema de edição de conteúdo de forma hierárquica.

Entretanto, uma nova variável tecnológica coloca todo esse espectro comunicacional em outra esfera. Com o advento das redes computacionais, a comunicação mediada por computador (CMC) revela-se como uma nova dimensão social, afetando as estruturas tradicionais de comunicação e sociabilidade como um todo. A Internet, ao permitir a digitalização de toda produção simbólica, possibilita práticas até então secundarizadas pelo padrão de produção do modelo de comunicação de massa (MCM).

As redes, por intermédio de computadores, criam um novo ambiente informacional, permitindo o surgimento de novas plataformas midiáticas, de modo a inserir velocidade no consumo, na produção e na distribuição de conteúdos informativos. Algumas dessas novas experimentações são delineadas como revoluções,

seja no campo do processo de produção do Jornalismo, seja na recepção de notícias pelo usuário.

### Redes sociais

Um desses campos é denominado rede social, ambiente *on-line* de interação social (*social on-line environments, social networking websites*). Apesar de o conceito não ser novo, a história dessas ferramentas sociais revela o constante desenvolvimento de *softwares* programados para permitir que o usuário insira um conteúdo e o compartilhe com os demais usuários da rede. A evolução dessas ferramentas acontece desde os anos 1970, tendo como um dos seus precursores o *Electronic Information Exchange System* (Eies)<sup>1</sup>. Segundo a *Techencyclopedia*<sup>2</sup>, o *software* colaborativo é o sistema que permite que pessoas trabalhem juntas nos mesmos documentos e projetos por meio de redes locais ou remotas. *Lotus Notes* é considerado o pai do *Groupware* (*software* para grupos de trabalho, categoria de aplicativos destinados a auxiliar grupos de usuários que trabalham juntos em rede ou de forma corporativa), que foi o primeiro termo cunhado para descrever um *software* colaborativo. Também chamado de *social software*, o *software* colaborativo adota os sistemas de comunicação, incluindo *e-mail*, videoconferência, mensagens instantâneas e *chat*.

### Mídias sociais

Para definir mídias sociais, é necessário comparar as novas possibilidades oferecidas pela mídia conectada via redes e a mídia tradicional analógica ou eletrônica. Nas mídias tradicionais, o conteúdo não pode ser alterado, sendo os níveis de interatividade bem menores que os proporcionados pelas mídias sociais conectadas, cujo ambiente de integração surge na forma de agregação de conteúdo e conhecimento, convergência de mídias, divergência de meios (distribuição), participação, colaboração e recuperação da informação.

Com parca conceituação teórica, sua definição pode ser encontrada na *Wikipedia*<sup>3</sup> como um “termo guarda-

<sup>1</sup> Disponível em: <[http://www.lifewithalacrity.com/2004/10/tracing\\_the\\_evo.html](http://www.lifewithalacrity.com/2004/10/tracing_the_evo.html)>. Acesso em: 8 de outubro de 2009.

<sup>2</sup> Disponível em: <<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=collaborativesoftware>>. Acesso em: 8 de outubro de 2009.

<sup>3</sup> Disponível em: <[http://pt.wikipedia.org/wiki/M%C3%ADdias\\_sociais](http://pt.wikipedia.org/wiki/M%C3%ADdias_sociais)>. Acesso em: 8 de outubro de 2009.

chuva que define várias atividades que integram tecnologia, interação social e construção de textos, imagens, vídeos e áudio”. Entretanto, Lima Junior (2009) apresentou uma definição.

Na tentativa de delimitar um recorte adequado para o uso das ferramentas que suportam as redes sociais com o objetivo de produção de mídia de relevância social, podem-se elencar alguns atributos necessários para que esse objetivo seja alcançado: a mídia social é um formato de comunicação mediada por computador (CMC) que permite a criação, compartilhamento, comentário, avaliação, classificação, recomendação e disseminação de conteúdos digitais de relevância social de forma descentralizada, colaborativa e autônoma tecnologicamente. Tem como principal característica a participação ativa (síncrona e/ou assíncrona) da comunidade de usuários na integração de informações, visando à formação de uma esfera pública interconectada (LIMA JUNIOR, 2009: 8).

### Filtros colaborativos

O crescimento da utilização das redes sociais como mídia tem um ingrediente tecnológico fundamental. A possibilidade de identificar a informação que é de interesse daquela coletividade, que colabora com um determinado *website* informativo, por exemplo, é uma das características mais importantes da comunicação mediada por computador. Esses sistemas, denominados como “*social filtering*” e “*social information filtering*”, garantem uma nova forma de produção e compartilhamento da informação em um suporte midiático.

O princípio básico do filtro está em desenvolver um sistema de avaliação para adequar o material inserido, com base em diferentes tipos de algoritmos. “Colaborativo” significa que um grupo de pessoas interessadas em um tema define as suas preferências e “levanta” o sistema. Os filtros colaborativos são usados especialmente como “sistemas de recomendação” que podem, por exemplo, melhorar a experiência de um *website* pela sugestão de música ou filmes de que o usuário poderia gostar<sup>4</sup>. Trata-se de uma solução atualmente adotada por serviços na *Web*, que dependem de potencializar relações de confiança entre os usuários para que mais

peças os utilizem. Esses sistemas lançam mão de mecanismos de coleta de informações sobre o indivíduo de acordo com seu histórico de comportamento, suas contribuições e referências de outras pessoas.

### Sistemas de recomendação

Entre os mecanismos de avaliação, no campo do uso das redes sociais como mídia, surgem os sistemas de recomendação (SR). Certos *websites* têm permitido, por intermédio desses sistemas, que a seleção de serviços, conteúdos e informações (inclusive jornalísticas) se apoie em mais filtros com base no perfil de interesse, no comportamento e em outras informações disponíveis dos usuários. Um SR é um sistema computacional que tenta predizer itens nos quais o usuário possa estar interessado.

Há poucos anos, em 1996, o interesse em filtros colaborativos levou a um *workshop* sobre o tópico na University of California, Berkeley. Os resultados desse *workshop* foram publicados em março de 1997, na edição especial da revista da *Association for Computing Machinery* (ACM). Nela foi cunhada a expressão sistemas de recomendação, por Resnick & Varian, no artigo introdutório da publicação. Eles escolheram a frase sistemas de recomendação ao invés de filtros colaborativos, porque os “recomendadores” não precisam explicitamente colaborar com os aqueles que recebem a recomendação. Além disso, as recomendações se referem a sugestões de artefatos interessantes, além de filtrar objetos indesejados. Resnick & Varian definem um recomendar como um sistema que aceita modelos de usuários como contribuição, agregando-lhes e devolvendo a recomendação aos usuários (PERUGINI, GONÇALVES & FOX, 2004: 6).

Para Perugini, Gonçalves & Fox, sistemas de recomendação também são programados “para conectar grupos de indivíduos com interesses similares e para influenciar a experiência coletiva ao invés de somente focar na busca da meta-informação de um indivíduo específico” (2004: 7).

Uma função cada vez mais importante desses sistemas é a de auxiliar no controle do que interessa e de quem possa interessar, como um *gatekeeper*<sup>5</sup>. Neste

<sup>4</sup> Disponível em: <<http://www.techweb.com/encyclopedia/defineterm.jhtml?term=collaborativefiltering>>. Acesso em: 8 de outubro de 2009.

<sup>5</sup> Sua função está intimamente ligada à ação pessoal deste profissional, que é o responsável pela seleção das notícias que vão ser publicadas.

sentido, ao recomendar outros *blogs* em suas páginas, os blogueiros expandem a visibilidade e o raio de ação dos *links*. Isso funciona como uma eleição informal, já que “linkar” é oferecer, sugerir, assinalar positivamente.

O *Digg*, um dos *sites* informativos mais visitados dos EUA<sup>6</sup>, é hoje um dos exemplos mais conhecidos da utilização de sistemas de recomendação. De modo integrado, o público acrescenta conteúdo, opina e elege o que é mais relevante. Isso gera um espaço de troca, explorando mais efetivamente a liberdade que caracteriza a Internet. Antes de prosseguir, é oportuno apresentar como funciona o *Digg*, suas formas de publicação, compartilhamento e organização.

### Como funciona o Digg

O funcionamento do *Digg* é “amigável”. Seus usuários podem participar a partir de envio de *links* para postagens de *blogs*, notícias, *sites*, vídeos e demais tipos de conteúdo que serão avaliados também pelos usuários. As notícias mais “aclamadas” pela comunidade de usuários chegam a alcançar mais de mil votos (*diggs*), e suas notícias estão distribuídas em basicamente seis categorias: *Technology*, *Science*, *World & Business*, *Sports*, *Entertainment* e *Gaming*.

O usuário que postou a notícia não recebe nada em troca (brindes ou dinheiro) e quem votou para que fosse enviada a capa, também não. É preciso informar que o *Digg* não se baseia totalmente em conteúdo: o aspecto social do *site* é muito forte. Existe uma disputa entre os usuários para saber quem insere notícias mais relevantes, além de um sistema de comentários bem movimentado com perfis de usuários.

Diante dessa autonomia menos suscetível à manipulação por uma classe “legalmente definida”, é necessário verificar os parâmetros que governam a gestão da produção comum nesse espaço. Pela descrição do sistema de recomendação do *Digg*, dois critérios são determinantes: o direito de colaborar mediante um registro simples e a prevalência de uma avaliação coletiva do valor dessas colaborações para sua organização segundo um índice de relevância. Por que foram feitas estas escolhas, e não outras? O que está por trás deste

modelo? Os sistemas de recomendação suscitam outra questão importante sobre a mídia tradicional. Até que ponto as empresas tradicionais de mídia “manipulam” os destaques noticiosos de suas capas?

### Sistemas de recomendação versus gatekeeping

Uma questão importante, em qualquer trabalho acadêmico sobre os processos internos orientados para a produção de conteúdo jornalístico, é a contextualização sobre as pressões editoriais (mercadológicas e/ou ideológicas) sofridas pelos profissionais que atuam diretamente na engrenagem de elaboração da notícia.

Estudos sobre a prática jornalística, do ponto de vista de produção de conteúdo informativo nas redações, detectaram problemas com a questão da liberdade de atuação dos profissionais nos diversos patamares hierárquicos de uma organização midiática. Fatores que determinam os tipos de cerceamento da liberdade de atuação profissional vão desde o posicionamento mercadológico de uma empresa de comunicação de massa, levando-se em consideração o seu modelo de negócio, sua área de atuação e seu impacto comunicacional (concentração), até a adoção de uma linha editorial e um alinhamento ideológico. No livro *Conceitos de jornalismo: norte e sul*, Michael Kunczik (1997) discorreu, no terceiro item do capítulo 7 – “O controle social entre as equipes jornalísticas” –, sobre a “orientação intencional ou involuntária do indivíduo pelo grupo”. O referido autor observou que:

Embora a liberdade de imprensa signifique a liberdade para intervenção das autoridades e a coação política, assim como a ausência de pressão econômica (nesse aspecto, a questão da concentração da imprensa representa uma papel decisivo), o conceito de liberdade interna de imprensa se relaciona com a liberdade e os deveres internos dos meios de comunicação de massa (KUNCZIK, 1997: 168).

Outra questão abordada por Kunczik refere-se à hierarquia. Os organogramas das empresas de comunicação de massa têm um formato peculiar, imposto pelo modelo de negócio, porém muito similares, apesar dos diferentes tipos de veículos. Mudam-se as nomenclaturas, mas os processos de decisão sempre levam ao que é conceituado de *gatekeeping*.

Não se pode esquecer que os jornalistas são empregados que têm que se submeter à hierarquia da sala de redação e aceitar seus níveis de autoridade. As relações entre a equipe jornalística podem ser amistosas, mas no caso de um conflito a hierarquia predomina: o

<sup>6</sup> Relatório divulgado pela Universidade de Harvard. Disponível em: <<http://noticias.uol.com.br/ultnot/efe/2007/08/16/ult1766u23230.jht>>. Acesso em: 8 de outubro de 2009.

editor, o chefe de redação, o chefe de editoria e o escritor com o poder de orientação (KUNCZIK, 1997: 174).

Nessa hierarquia, há vários pontos de checagem do material que está sendo produzido e por onde ele passará para aprovação. Essas “passagens” foram traduzidas na teoria denominada de “*gatekeeping*”. Kurt Lewin (1947: 145) foi, aparentemente, o primeiro a sugerir o termo “*gatekeeping*”, que usou para descrever uma mulher ou mãe como uma pessoa que decide qual comida irá colocar na mesa do jantar

Em suma, o *gatekeeping* é a pessoa que decide o que irá passar através de cada sessão de passagem (decisão), sendo que como em qualquer processo, há muitas. Embora o conceito fosse aplicado originalmente para uma cadeia de alimentação, ele acrescentou que o processo pode incluir uma notícia que entra no circuito de canais de comunicação de um grupo<sup>7</sup>.

O conceito também leva em consideração a questão da audiência, pois ela determina que tipo de conteúdo será feito na redação e a angulação que deverá ter. Contudo, um ponto de crítica, principalmente pelos simpatizantes dos sistemas de recomendação, versa sobre os critérios não jornalísticos que influenciam na decisão de como deve ser pautada uma matéria, produzida, editada e divulgada.

As críticas, nesse sentido, apontam que as escolhas do *gatekeeper* constituem uma complexa teia de influências, preferências, motivos e valores em comum. “O *gatekeeping* é inevitável e, em algumas circunstâncias, pode ser útil. O sistema de *gatekeeping* também pode ser perigoso, desde que ele possa levar ao abuso de poder por decidir que tipo de informação será descartada e o que passará. Todavia, o sistema é, muitas vezes, uma rotina, guiada por estabelecimento de perguntas padrão”<sup>8</sup>.

Quando uma nova questão é percebida por um grupo pequeno da população, a mesma somente será colocada em uma discussão mais ampla caso se consiga disseminar essa ideia ou caso o pequeno grupo que controla os meios da mídia seja, por algum motivo, sensível a

essa questão. O primeiro processo requer alguma organização para ocorrer. O segundo processo também apresenta problemas, como uma possível discriminação que surge, na medida em que problemas ou opiniões ligadas a assuntos pertinentes ao grupo que controla os meios da mídia acabam por receber destaque mais rapidamente do que assuntos pertinentes a outros grupos.

Tanto a filtragem por relevância quanto a por credibilidade das informações e opiniões a serem apresentadas na mídia sofrem também com essa concentração do poder de decisão sobre esses aspectos. A relevância e a credibilidade passam a ser medidas pela ótica de um grupo restrito em oposição a serem medidas pela população em geral. A possibilidade de interação da audiência com o conteúdo a ser disponibilizado é denominado, por alguns autores, como mídia *bottom-up*. Os MCM utilizariam o processo *top-down*, quando o conteúdo vem de cima, é reinterpretado<sup>9</sup> e redistribuído para baixo, entendeu Jenkins (2006: 7).

### *Ponderações sobre a Organização dos SR*

Uma das questões levantadas pelos críticos dos sistemas de recomendação é que eles favorecem uma tendência. Os usuários tendem a gostar de notícias submetidas por amigos, e os usuários tendem a gostar de notícias que os amigos leem ou de que gostam. Isso é um filtro social, pois, sendo alguns usuários mais ativos do que outros, a implantação direta dos filtros sociais pode levar à “tirania da minoria” em que as notícias de capa são oriundas de usuários com uma ativa rede social (PERUGINI, GONÇALVES & FOX, 2004: 2). Aliás, a própria sequência dos resultados de uma pesquisa no Google já se vale de critérios de eleição e recomendação que, se bem manipulados, podem jogar para o topo da página endereços da Internet com muito acesso e pouca relevância. Para Perugini, Gonçalves & Fox, “os sistemas de recomendação tentam imitar, naturalmente, a relação social e automatizar esse processo” (2004: 2), sendo que os SR envolvem o uso de modelagem, que inclui a representação de preferências de usuário e interesses.

<sup>7</sup> Teoria disponível em: <<http://www.tcw.utwente.nl/theorieenoverzicht/Theory%20clusters/Media,%20Culture%20and%20Society/gatekeeping.doc/>>. Acesso em: 8 de outubro de 2009.

<sup>8</sup> Disponível em: <<http://www.tcw.utwente.nl/theorieenoverzicht/Theory%20clusters/Media,%20Culture%20and%20Society/gatekeeping.doc/>>. Acesso em: 8 de outubro de 2009.

<sup>9</sup> Em *Convergence culture*, Henry Jenkins desenvolveu considerações importantes, ao relacionar exemplos da interação entre a mídia *top-down* (de cima para baixo) e a *bottom-up*, e ao vincular o fenômeno a um processo cultural desencadeado pela tecnologia. A ocorrência da tomada do *Digg* pelos usuários no caso HD/DV define a noção de cultura de convergência.

A semelhança entre a dinâmica estabelecida por agentes múltiplos que compõem redes sociais e a complexidade de organismos vivos – por isso, a menção a termos como *bottom-up* e emergência – é uma tendência na associação de sistemas informatizados, que ajudam a formação de redes sociais, com os sistemas que estruturam os organismos vivos. Essas “pontes” são importantes, pois muitos sistemas informatizados conectados em redes são baseados no comportamento humano. Entretanto, a observação do comportamento humano e a sua tradução em um modelo de sistema informatizado são bem menos complexos no entendimento do que a utilização de termos da biologia em sistemas computacionais conectados em redes. Essa associação é decorrente da teoria da biologia chamada organicismo.

Os organicistas afirmam que as características únicas dos organismos vivos não se devem à sua composição, e sim à sua organização. A partir dessa nova fronteira nas ciências biológicas, William E. Ritter (*apud* MAYR, 1997: 38) cunhou, em 1919, o termo organicismo: “os todos são tão relacionados com as suas partes que não só a existência do todo depende da ordenada cooperação e interdependência de suas partes, mas também o todo exerce uma medida de controle determinantes sobre suas partes”. Porém, essa associação é inconsistente, pois o organicismo inclui o reconhecimento de que a existência de um “programa genético é uma característica importante para o novo paradigma” (MAYR, 1997: 39).

Os programas que estruturam redes sociais não conseguem realizar tal processo, mesmo de modo reduzido e metafórico. A base do organicismo, pontuou Mayr, “é o fato de que os seres vivos possuem uma organização. Eles não são apenas montes de características ou de moléculas, uma vez que sua função depende inteiramente de sua organização, de suas inter-relações, interações e interdependências” (MAYR, 1997: 41).

Também transportados da área da biologia, os termos “emergência” e “*bottom-up*” são derivados da teoria do organicismo. Utilizado com frequência, o conceito de emergência foi bastante difundido por Steve Johnson (2003), que:

(...)relata exemplos de sistemas auto-organizados que privilegiam as sequências, em detrimento da lógica, e nos quais se dispensa a presença de um controle centralizado para haver ação. Surgem de um nível de elementos relativamente simples

em direção a formas de comportamento mais sofisticados e por isso são chamados sistemas emergentes (JOHNSON, 2003)<sup>10</sup>.

Metaforicamente, há certas semelhanças entre o termo “emergência”, que é utilizado para sistemas informacionais auto-organizados, e a teoria sobre emergência, oriunda das ciências biológicas. As semelhanças estão na ausência de um controle centralizado que atue no processo de decisão para a “subida” da informação. Mas as semelhanças param nesse campo. Na biologia, Jacob (*apud* MAYR, 1997) descreveu a emergência da seguinte forma:

Em cada nível, unidades de tamanho relativamente bem definido e de estrutura quase idêntica se associam para formar uma unidade do nível acima. Cada uma dessas unidades formadas pela integração de subunidades pode receber a designação geral de “integron”. Um “integron” é formado pela organização de “integrans” dos níveis abaixo dele, ele participa da construção do “integron” no nível acima dele. Cada “integron” tem novas características e capacidades, que não estavam presentes em nenhum nível mais baixo de integração; dessas características e capacidades diz-se que ele as emergiram (*sic*) (*apud* MAYR, 1997: 42).

O termo *bottom-up* também tem a mesma característica metafórica. Steve Johnson descreveu o conceito da seguinte forma: “*bottom-up* são modelos onde não há líderes, e a dinâmica se estabelece a partir da interação entre agentes múltiplos, sem perceber instruções de nível mais alto (JOHNSON, 2003: 15). Nos casos observados por Johnson, isso resulta algum tipo de macrocomportamento observável, o que é identificado pelo autor como uma propriedade emergente. Entretanto, ao analisar-se como as ciências biológicas conceituam o termo, tornam-se nítidas as diferenças entre as duas utilizações. Zhang & Sun, em “*Top-down versus bottom-up learning in cognitive skill acquisition*”, afirmaram que o conhecimento implícito já adquirido no nível *bottom* pode ser utilizado na aprendizagem do conhecimento explícito no nível *top* por intermédio da aprendizagem *bottom-up* (SUN *et al.*, 1998, 2001). Da mesma forma, “o conhecimento explícito já adquirido no nível *topo* poder ser utilizado na aprendizagem do

<sup>10</sup> Disponível em <[http://www.relativa.com.br/livros\\_template.asp?Codigo\\_Produto=14279#200](http://www.relativa.com.br/livros_template.asp?Codigo_Produto=14279#200)>. Acesso em: 8 de outubro de 2009.

conhecimento implícito no nível *bottom* através do aprendizado *top-down*” (ZHANG & SUN, 2004: 67).

### *O problema da objeção de Babel*

A perspectiva de Primo (2007) é importante para entender que a abertura para o trabalho colaborativo não oferece uma auto-organização mágica que faria sempre emergir a verdade a partir de vozes espontâneas, interessadas no bem comum. Não como ferramentas democratizantes, mas como um método funcional para ambientes descentralizados; uma solução para o problema identificado como objeção de Babel. De acordo com a objeção de Babel, quando todos podem falar, ninguém consegue ouvir, e regressa-se a uma cacofonia ou ao dinheiro como o fator distintivo entre declarações que são ouvidas e aquelas que caem na obscuridade. Yochai Benkler (2006) argumentou que as pessoas estão começando a ver o “surgimento de fontes alternativas e fora do sistema de mercado, produzida entre pares para a filtragem e o credenciamento ao invés de alternativas de mercado” (BENKLER, 2006: 7).

### *Identificando parâmetros e obscuridades: limites do SR do Digg*

Qual a relação entre o *site* de buscas *Google* e o candidato à presidência dos Estados Unidos, Barack Obama? *A priori* incoerente, o questionamento ilustra uma falha na mais recente investida do *Digg*. Denominada *Recommendation Engine*, a ferramenta em fase de testes amplia o propósito de existência do *site*, cujo princípio inicial era “ajudar as pessoas a descobrir novas informações, permitindo-lhes a partilhar aquilo que elas acham interessante com o outro”<sup>11</sup> e passa a recomendar, automaticamente, notícias aos usuários. Percebe-se, portanto, que as duas técnicas de sistemas de recomendação são utilizadas no *Digg*: a filtragem colaborativa e a filtragem baseada em conteúdo.

Tecnicamente, a ferramenta traça um perfil de leitura e “interesse” a partir da análise de três variantes: navegação, preferências pessoais explícitas no cadastro e recomendações de amigos. Esta combinação tornaria

possível apontar as notícias que um determinado usuário gostaria de ler. Entretanto, o algoritmo que pretende revelar nexos entre fatos para indicar tendências e predisposições aparentemente comete falhas que superam o relacionamento da informação por palavras-chave ou *tags*.

Antes de tentar responder ao questionamento inicial, são utilizados os valores-notícia (atributos do acontecimento) elencados para hierarquizar, relacionar e recomendar conteúdo para o usuário. É notório, entre os pesquisadores de comunicação, que não há consenso sobre as características que levam um fato a se tornar notícia e, principalmente, a hierarquia entre os vários fatos.

Estudiosos como Gislene da Silva (2005) tentaram compilar as contribuições de diversos autores sobre o tema, tendo como resultado a percepção de que os critérios de noticiabilidade agem de modo combinado e sofrem interferência do contexto social e editorial. Para complexificar ainda mais a discussão, Gislene da Silva citou Vicent Campbell (2004), quando discutiu *News values and news selection*. Um deles é o de que os valores notícia determinam a seleção dos acontecimentos e, ao mesmo tempo, a seleção de fatos noticiosos também determina os valores-notícia.

Não é uma tarefa simples valorar e sistematizar os atributos de uma notícia de modo a eliminar os traços subjetivos e de contexto. Apesar de o objetivo do *paper* não ter como foco os gargalos da construção de uma base de conhecimento<sup>12</sup> (BC), é válido ressaltá-las e, simultaneamente, amenizar as críticas aos programadores do *Digg*, pois encontrar a sinergia entre engenheiros de conhecimento e especialistas de um domínio exige dedicação, planejamento e uma dose de boa vontade.

Enquanto os engenheiros de conhecimento são responsáveis por transformar as ideias, os conceitos e os modos de racionalizar o mundo em um modelo processável computacionalmente, o especialista em um domínio precisa se esforçar para traduzir seu conhecimento em uma linguagem clara e objetiva, além de avaliar e apontar os deslizos cometidos pelo sistema com base nas respostas obtidas. Isso significa dizer que é um ciclo contínuo de aperfeiçoamento (BARBOSA & LIMA JUNIOR, 2007: 4).

<sup>11</sup> Tradução livre: *Digg is founded on the principles of helping people to discover new information by enabling them to share what they find interesting with one another*. Disponível em: <<http://www.digg.com/privacy>>.

<sup>12</sup> É uma base de dados ou conhecimento acumulado sobre um determinado assunto. Essas informações podem ser utilizadas na solução dos problemas apresentados pelos clientes e/ou usuários, por meio de ferramentas de inteligência artificial (IA) ou sistemas especialistas.

O *Digg* afirma que não existe edição de material, seja no envio, seja na promoção ou na exclusão de um artigo. Não existe censura dos envios, a não ser um filtro que administra as exclusões da mesma maneira que administra qualquer outra coisa, com um algoritmo proprietário. Como todo sistema proprietário, o departamento técnico não fala muito sobre a sua tecnologia, mas é possível reunir algumas informações.

Em um *podcast* do *DiggNation*, gravado em 14 de junho de 2006, Kevin Rose disse que o número total de servidores era de 75. Em uma entrevista feita por e-mail para o site *HowStuffWorks*, o CEO<sup>13</sup> do *Digg*, Jay Adelson disse que seu sistema não precisa armazenar muita informação, já que a maior parte do conteúdo é texto. As configurações, entretanto, podem aumentar indefinidamente.

Outro ponto importante está na série de verificações de legitimidade, que o *Digg* chama de “*karma checks*”, embutidos em diversos pontos do sistema. Quando um usuário tenta enviar um artigo, por exemplo, o sistema verifica se o envio é legítimo, se não é *spam* e se o usuário é válido (não foi banido ou tem ações limitadas). De acordo com o *Digg*, o sistema sabe a diferença entre os usuários que foram criados só para promover uma notícia e os usuários que interagem com o *site*. O sistema também impediria o “*autodigging*”, uma maneira computadorizada e fraudulenta de promover uma notícia. Tudo isso soa muito democrático, mas a questão do controle editorial é a base de muitas reclamações apontadas por usuários.

Nesse sentido, os pesquisadores desse *paper* criaram uma conta no *Digg* para testar a ferramenta de recomendação no dia 14 de julho de 2008. Foi realizada uma pesquisa com o termo *Google* e clicou-se em três notícias. Ei-las: (1) *Web 3.0 – Could it unseat Internet giants like Google?*; (2) *Google gives parents the shaft, 75% price hike in daycare*; e (3) *Google changes home page, adding link to Privacy Policy*.

Optou-se por não incluir informações adicionais ao cadastro, tais como filtrar as preferências de visualização (*Filter your topics – news/videos/images*) ou, ainda, criar vínculos (cadastrar amigos) o que, efetivamente, poderia influenciar as indicações. Como resultado, o *Digg* recomendou as notícias a seguir: (1) *Hypnotized by Apple*; (2) *iPhone 3G: the object of MY desire*;

e (3) *CNN exclusive: Obama on Foreign Policy*, sendo, respectivamente, categorizadas em *most recent*, *most matches* e *most diggs*.

No dia 15 de julho de 2008, foi realizado o mesmo procedimento e clicou-se nas seguintes notícias: (1) *No Google source code for Viacom, just 12TB of YouTube data*; (2) *Google now offers Google Talk on the iPhone*; (3) *Google, Yahoo spiders crawl through flash sites*. Não obstante, as matérias indicadas foram as que seguem: (1) *Apple sues clone maker psystar*; (2) *Is the phone killing terrestrial radio?*; (3) *New Yorker editor defends Obama cover*. No esforço de evitar especulações desnecessárias, mas sim propor reflexões, é possível delinear três explicações para as sugestões da ferramenta *Recommendation Engine*. Em primeiro lugar, apesar da proposta de automoderação do *site*, não se pode descartar a hipótese de má-fé ou interesses pessoais envolvidos por parte dos administradores do *Digg*, já que os mesmos podem interferir tanto no processo de recomendação quanto na gestão de conteúdo, utilizando-se do “direito de impor limites sobre determinadas características e serviços ou restringir o acesso sem aviso ou responsabilidade”<sup>14</sup>.

Quanto à segunda possibilidade, os aperfeiçoamentos necessários de uma ferramenta em fase de testes podem comprometer a eficácia e a eficiência, seja por meio da análise semântica seja por intermédio da estatística. Em outras palavras, a primeira diz respeito a uma avaliação da sequência dos termos no contexto da frase (no caso específico do *Digg*, no título), enquanto a segunda dedica-se a contabilizar o número de vezes que um termo aparece no texto (BARBOSA & LIMA JUNIOR, 2007).

Por fim, é possível que haja, efetivamente, algum padrão oculto (descoberto através de mineração na *Web* ou outra técnica que utilize recursos de estatística e/ou inteligência artificial) entre os termos *Google* e Barack Obama. Neste ponto, faz-se necessária uma breve explicação sobre o que seria mineração.

Como conceito que envolve estatística, inteligência artificial e *machine learning* (aprendizado de máquina), a mineração na *Web* garimpa informações de valor estratégico que estão “invisíveis” nos registros, permitindo a identificação de tendências para uma visão

<sup>13</sup> Chief executive officer.

<sup>14</sup> *Digg may also impose limits on certain features and services or restrict your access to parts or all of the services without notice or liability.* Disponível em: <<http://www.digg.com/tos>>.



antecipada de cenários futuros e a descoberta de novos padrões entre dados, nem sempre perceptíveis ao analista humano. Em outras palavras, a mineração na *Web* pode ser conceituada como a descoberta e a análise inteligente de informações úteis no *World Wide Web* (COOLEY, MOBASHER & SRIVASTAVA, 1997).

Apesar do enorme potencial de organizar e extrair informação relevante do caos refletido no WWW, a utilização dessa tecnologia pelo *Digg* parece improvável. Os diversos mecanismos de salvaguarda explícitos nos termos de uso<sup>15</sup> evitam que processos judiciais responsabilizem o *Digg* pelo conteúdo contido em seu interior ou acessado por meio do *site* ou de serviços. Isto reforça a ideia da inoperabilidade do sistema de relacionamento e descoberta de informação. Como exemplo, o *Digg* se esquia de qualquer responsabilidade sobre os *links* disponíveis, sejam fornecidos por terceiros, sejam pelo próprio *site*.

No parágrafo 11 dos termos de uso, o *Digg* afirma que “não é responsável pela disponibilidade desses *sites* ou recursos externos e não endossa nem é responsável por qualquer conteúdo, publicidade, produtos ou demais materiais disponíveis ou originados em tais *sites* ou recursos”. E continua: “você também está ciente e concorda que o *Digg* não será responsável, direta ou indiretamente, por qualquer perda ou dano causado ou alegado ser causado por ou em conexão com o uso da confiança em tal conteúdo, bens ou serviços disponíveis em tais *sites* ou recursos”. O fato é que as inúmeras cláusulas e a linguagem jurídica dos termos de uso e da política de privacidade<sup>16</sup> prejudicam o bom entendimento dos contratos. Há, também, um movimento automático, por parte dos usuários na rede, que os aceitam sem lê-los na íntegra. Como consequência destes dois fatos, o *Digg* se aproveita para impor regras desfavoráveis aos leitores-colaboradores.

### Considerações finais

Os sistemas de recomendação são mecanismos computacionais que mostram ter um grande potencial, pois fazem o ranqueamento dos conteúdos e permitem a inserção deles pelos usuários dos portais, utilizando a filtragem colaborativa. Esses sistemas estão se contrapondo aos processos tradicionais de edição, como o *gatekeeping*, utilizados pelos meios de comunicação de massa. Portanto, o procedimento usado pelos MCM estaria fadado ao insucesso mercadológico, pois a “transparência” contida na possibilidade de não intermediário, para a formação do discurso, colocaria em xeque o modelo tradicional.

Verificou-se também, no caso do *Digg*, que seu valor informativo é fortemente mobilizador no sentido de refletir-se em outros *sites*. É crescente a adoção de recursos que facilitam o compartilhamento de conteúdo para destinos diversos outros canais, como atalhos para integração com o *Facebook* e o *Twitter*. Do ponto de vista dos processos de inovação, sempre procurados pelo homem, seja por quem cria, seja por quem consome, os SR estão na vanguarda de sistemas que favorecem a auto-organização da informação jornalística. Apesar do claro avanço que os sistemas de recomendação impõem sobre os modelos adotados pelos MCM, ainda há muito que avançar tecnologicamente para afirmar com precisão que se trata de um sistema “transparente” de produção de jornalismo ou que é um processo que causa “emergência” ou demonstra “inteligência coletiva”. Aliás, termos importados das ciências biológicas e utilizados de forma indiscriminada fornecem a falsa sensação de que os filtros colaborativos funcionam e tomam decisões como se fossem seres vivos.

Apesar das inconsistências ainda apresentadas por tais sistemas, eles se revelam como uma nova dimensão social, afetando as estruturas tradicionais de comunicação e sociabilidade como um todo, de maneira a possibilitar práticas até então secundarizadas pelo padrão de produção do modelo impetrado e consolidado pelos meios de comunicação de massa.

<sup>15</sup> Disponível em: <<http://www.digg.com/tos>>.

<sup>16</sup> Disponível em: <<http://www.digg.com/privacy>>. Acesso em: 8 de outubro de 2009.

## Referências

- BARBOSA, Pablo & LIMA JUNIOR, Walter Teixeira. Mineração de dados e textos e suas possibilidades aplicadas ao processo de produção da notícia. In: 5º ENCONTRO NACIONAL DE PESQUISADORES EM JORNALISMO – SBPJOR. *Anais eletrônicos...* Aracaju: SBPJor/UFS, 2007. Disponível em: <<http://www.lopezfreire.files.wordpress.com/2007/11/ind-pablo-barbosa.pdf>>. Acesso em: 08 de junho de 2008.
- BENKLER, Yochai. *The wealth of networks: how social production transforms markets and freedom*. New Haven and London: Yale University Press, 2006.
- CAMPBELL, Vicent. *News values and news selection in: information age journalism*. London: Arnold, 2004.
- CASTELLS, Manuel. *A sociedade em rede*. Tradução de Roneide Venâncio. São Paulo: Paz e Terra, 2003.
- COOLEY, Robert W.; MOBASHER, Bamshad & SRIVASTAVA, Jaideep. Web mining: information and pattern Discovery on the World Wide Web. In: 9th INTERNATIONAL CONFERENCE ON TOOLS WITH ARTIFICIAL INTELLIGENCE – ICTAI. *Proceedings...* Newport Beach, CA: IEEE Computer Society, 1997. p. 558-567.
- JENKINS, Henry. *Convergence culture: where old and new media collide*. New York: New York University Press, 2006.
- JOHNSON, Steven. *Emergência: a dinâmica de rede em formigas, cérebros, cidades e softwares*. Tradução de Maria Carmelita Pádua Dias. Rio de Janeiro: Jorge Zahar, 2003.
- KUNCZIK, Michael. *Conceitos de jornalismo: norte e sul*. Tradução de Rafael Varela Júnior. São Paulo: Edusp, 1997.
- LEWIN, Kurt Z. Frontiers in group dynamics: concept, method and reality in social science; social equilibria and social change. *Human Relations*, v.1, n. 1, p. 5-41, June, 1947.
- LIMA JUNIOR, Walter Teixeira. Produção de informação de relevância social no ambiente da mídia social conectada. In: XXXII CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO – INTERCOM. Curitiba: Intercom/Universidade Positivo, 2009. Disponível em <<http://www.intercom.org.br/papers/nacionais/2009/resumos/R4-1859-1.pdf>>. Acesso em: 8 de outubro de 2009.
- MARINHO, Leandro Balby & GIRARDI, María del Rosario. *Mineração na Web*. Reic – Revista *Eletrônica de Iniciação Científica*, v. III, n. II, p. 1, Porto Alegre, junho, 2003. Disponível em: <<http://www.sbc.org.br/index.php?language=1&subject=101&content=magazine&option=content&id=8>>. Acesso em: 27 de julho de 2008.
- MAYR, Ernst. *Isto é biologia: a ciência do mundo vivo*. Tradução de Claudio Ângelo. São Paulo: Schwarcz, 1997.
- MIELNICZUK, Luciana P. & SILVEIRA, Stefanie C. da. Interação mediada por computador e jornalismo participativo nas redes digitais. In: PRIMO, Alex; OLIVEIRA, Ana Claudia de; NASCIMENTO, Geraldo C. do & RONSINI, Veneza Mayora (Orgs.). *Comunicação e interações*. Livro da Compós 2008. 1. ed. Porto Alegre: Sulina, 2008.
- PALACIOS, Marcos. Jornalismo *on-line*, informação e memória: apontamentos para debate. Revista PJ:Br – *Jornalismo Brasileiro*, ed. 4, segundo semestre de 2004. Disponível em: <[http://www.eca.usp.br/pjbr/arquivos/artigos4\\_f.htm](http://www.eca.usp.br/pjbr/arquivos/artigos4_f.htm)>. Acesso em: 12 de julho de 2008.
- PERUGINI, Saverio; GONÇALVES, Marcos André & FOX, Edward A. Recommender systems research: a connection-centric survey. *Journal of Intelligent Information Systems*, v. 23, n. 2, p. 107-143, 2004.
- PRIMO, Alex. O aspecto relacional das interações na Web 2.0. *E-Compós*, v. 9, p. 1-21, Brasília, 2007. Disponível em: <<http://www6.ufrgs.br/limc/PDFs/web2.pdf>>. Acesso em: 17 de julho de 2008.
- SILVA FILHO, Wagner D. da & CAZELLA, Sílvio César. Um *framework* para recomendação de artigos científicos baseado na relevância da opinião dos usuários em filtragem colaborativa. In: V ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL – ENIA. São Leopoldo: Unisinos, 2005. Disponível em: <[http://www.inf.unisinos.br/~cazella/papers/ENIA\\_Danda\\_Cazella2005.pdf](http://www.inf.unisinos.br/~cazella/papers/ENIA_Danda_Cazella2005.pdf)>. Acesso em: 21 de junho de 2008.
- SILVA, Gislene da. Valores-notícia: atributos do acontecimento (Para pensar critérios de noticiabilidade I). In: IV ENCONTRO DOS NÚCLEOS DE PESQUISA DA INTERCOM. Trabalho apresentado ao NP02 – Jornalismo. Porto Alegre: Intercom/PUC-RS, 2004. Disponível em: <<http://www.intercom.org.br/papers/nacionais/2005/resumos/R0797-1.pdf>>.
- ZHANG, Xi & SUN, Ron. *Top-down versus bottom-up learning in cognitive skill acquisition*. *Cognitive Systems Research*, v. 5, n. 1, p. 63-89, Elsevier, 2004.