Garimpo de dados na rede: o uso da Internet como instrumento predictivo

Luciane Lucas*

RESUMO

O texto se propõe a discutir a relação entre a Internet e os bancos de dados corporativos. Ao colocar em discussão questões como privacidade e nomadismo na Web, busca-se analisar a discursificação por trás desta tecnologia, a partir da comparação entre Internet e rizoma. Dando aos softwares de inteligência artificial o status de “fio de Ariadne” – em função de sua capacidade de rastreamento de dados e conexões significativas –, o *data mining* (garimpo de dados) surge como mola propulsora dos bancos de dados e da própria Internet, revalorizando a já antiga discussão sobre as possibilidades de algoritmização do conhecimento humano.

Palavras-chave: banco de dados; *data mining* Internet.

SUMMARY

This paper aims at discussing the relation between the Internet and corporate databank. Raising questions as Web's privacy and nomad features, it discusses the concept underlying such technology, out of comparison between the Internet and the rhizome. Providing the artificial intelligence software with a dueing characteristic – due to its traceability capacity for significant data and connections – the data mining arises as a driving force for databank and the Internet, revalorizing the old discussion on the possibilities of algorithm-making in relation to human knowledge.

Keywords: databank, data mining, Internet.

RESUMEN

El texto se propone a discutir la relación entre la Internet y los bancos de datos corporativos. Al poner en discusión cuestiones como privacidad y nomadismo en la Web, se busca analizar la discursificación detrás de esta tecnología, partiéndose de la comparación entre Internet y rizoma. Atribuyendo a los soportes lógicos de inteligencia artificial el status de “hilo de Ariadna” – en función de su capacidad de rastreo de datos y conexiones significativas –, el data mining (mineraje de datos) surge como mola propulsora de los bancos de datos y de la misma Internet, revalidando la ya antigua discusión sobre las posibilidades de algoritmización del conocimiento humano.

Palabras-clave: banco de datos; *data mining* Internet.

“**M**antém estes pensamentos noite e dia à disposição (procheiron); põe-nos por escrito, faz-lhes a leitura; que eles sejam o objeto das conversas contigo mesmo, com um outro [...] se te suceder um daqueles episódios que chamamos indesejáveis, logo encontrarás alívio no pensamento de que não era inesperado.”
Epicteto. Diálogos III, 24 (103).

Na contemporaneidade, os bancos de dados corporativos têm se firmado, progressivamente, como um novo padrão comunicacional. Fundamentando estratégias de personalização e fidelização, os assim chamados *databases* corporativos constituem um dos principais ícones da sociedade de consumo, na medida em que estabelecem formas cada vez mais eficazes de conhecer em detalhes o nível de consumo, o estilo de vida e os hábitos de compra dos indivíduos.

Questões múltiplas surgem neste contexto, uma vez que tal modelo aponta para uma mudança de natureza paradigmática. Se a comunicação até então era dirigida para uma massa sem rosto, dado o interesse de atingir o maior número possível de pessoas, agora o mais importante passa a ser identificar os rostos por trás da massa amorfa, buscando constituir uma relação mais estreita com cada um deles. Acredita-se, por conta disso, que, na próxima década, os bancos de dados já tenham se constituído como uma tendência consolidada, estabelecendo um novo paradigma comunicacional, com implicações substanciais no consumo, enquanto *ethos* da sociedade contemporânea.

Ao articularem os conceitos de previsibilidade e planejamento prévio, os bancos de dados possibilitaram a simulação e a correção antecipada de rumos, legitimando os mecanismos de uma emergente sociedade de controle. Definindo perfis, cruzando informações e agrupando consumidores em células segundo características comuns – a partir de dados pessoais, como grau de fidelidade, gostos e preferências, frequência de compra, recentidade e valor monetário –, torna-se possível promover um melhor monitoramento dos níveis e das oscilações de consumo, dentro das prerrogativas de uma sociedade de controle. Deleuze descreve bem este momento de transformação social, enunciando as bases do novo modelo: “Estamos entrando nas sociedades de controle, que funcionam não mais por confinamento, mas por controle contínuo e comunicação instantânea” (1992: 216).

Se, por um lado, é possível, diante desta personalização crescente, atingir níveis mais precisos de satisfação do consumidor, por outro, sabe-se que o

sistema busca, a todo momento, estriar os processos de subjetivação do indivíduo, a fim de arrebatá-lo para a moral hedonista que caracteriza a contemporaneidade. Esta espécie de “controle em meio aberto” (Deleuze, 1992: 216) acaba por trazer à tona assuntos correlatos, colocando em cena questões importantes, como privacidade, articulação do par desejo/risco, cuidados com os mecanismos de “estriamento da subjetividade”, seletividade diante do desenvolvimento de uma economia da atenção (Lévy), entre outras.

Estas novas preocupações se dão na medida em que a tecnologia dos bancos de dados inaugura um modo novo de interface com o consumidor. Porém, deve-se considerar que mais importantes do que os dados privativos do indivíduo são as operações possíveis de simulação e cruzamento que, ao estriarem trajetórias rizomáticas de sentido, encontram nexos infundáveis entre dados aparentemente sem ligação. Ou seja, o poder dos bancos de dados não está necessariamente no volume de informações que conseguem reunir sobre cada consumidor, mas antes, nas inúmeras correlações que podem ser estabelecidas entre os dados, a fim de gerar padrões inicialmente desconhecidos. São estas informações – obtidas a partir de *softwares* inteligentes e tarefas que aplicam modelos matemáticos e estatísticos –, então, que vão fornecer o embasamento para a montagem de estratégias de estímulo e personalização no consumo. Mais do que os dados em si, são os mecanismos de leitura e cruzamento de dados que conferem aos *databases* o poder predictivo de que se revestem. Não é o dado frio que vale, por mais que ele revele algo sobre o comportamento de consumo do indivíduo: são as conexões sempre renováveis e os trajetos que ele percorre, segundo a lógica do rizoma, que importam. É este trabalho de ler criativamente entre as dobras¹, de descobrir os pontos de contato invisíveis entre as informações armazenadas, que torna o banco de dados um agenciamento. Os dados só adquirem importância, portanto, na medida em que se podem identificar cadeias de significação entre eles, ou seja, laços invisíveis que se formam, costurando nexos.

O *data mining* – técnica que permite buscar, em uma grande base de dados, informações que aparentemente estão camufladas ou escondidas...² – realiza nos *databases* uma quantidade infinita de operações, apurando dados cuja natureza a inteligência humana nem sempre consegue explicar (o caso das redes neurais, por exemplo). Deste modo, ele não só possibilita que as engrenagens de consumo calculem, por meio de modelos matemáticos e estatísticos, a lealdade do consumidor, como identifica também padrões ainda desconhecidos e tendências no comportamento de consumo, além de prever riscos.

Revalorizando conhecimentos como o Cálculo, a Probabilística, a Estatística e a Lógica, o *data mining* – enquanto método de leitura inteligente dos dados – prenuncia e fortalece um movimento já iniciado com a montagem dos próprios *databases* corporativos: a valorização de um comportamento

predictivo no cenário contemporâneo. Buscando, por meio de algoritmos e estatísticas, prever tendências e antecipar-se a elas – em um contexto no qual o risco deve ser frequentemente administrado pelo próprio indivíduo –, as ferramentas *data mining* se articulam para descobrir padrões de informação até então ocultos, prever probabilidades, analisar riscos e identificar comportamentos de consumo. Utilizando redes neurais, algoritmos genéticos, árvores de decisão e instrumentos de inteligência artificial diversos, tais técnicas, na verdade, apontam para um novo paradigma temporal, próprio da contemporaneidade: aquele em que o futuro, de onde está, influencia e reorganiza as ações no presente.

As novas tecnologias da informação, principalmente os bancos de dados, possibilitam e, ao mesmo tempo, dão consistência a este hábito, agora incorporado ao *ethos* contemporâneo, de buscar detectar tendências e predisposições, a fim de minimizar os riscos. Este aparente jogo de espelhos – fundamentado em sucessivas simulações – tem a ver com uma cultura que se alicerça em um progressivo esvaziamento, uma vez que funciona na ordem do sinal, e não na do sentido. O que as tecnologias fazem hoje é tentar dar a este sinal visibilidade, procurando cadeias de sentido que o justifiquem. A proliferação dos bancos de dados como ferramenta tecnológica capaz de prever tendências, não se dá por acaso. A tentativa de rastreamento “telúrico” do sinal, buscando-se descobrir seu significado no meio de uma cadeia rizomática de sentido, tem por fim assegurar um *modus operandi* que perpetue e estimule a sociedade de consumo. E, neste contexto, os bancos de dados adequam-se perfeitamente a este trabalho de prospecção.

A simulação desempenha, portanto, um papel importante neste cenário. É por meio dela que estes sinais ganham corpo dentro de uma perspectiva temporal. Pode-se dizer que a simulação, então, faz problema porque é a sua operação que elucida a trajetória de um dado fenômeno na cadeia de tempo. Ou seja, é ela quem coloca, lado a lado, o presente como abertura e o futuro como possível e transformável.

Sabe-se hoje – e sobretudo diante dos recursos de simulação do aparato tecnológico disponível – que o futuro pode, antes mesmo de concretizar-se como tal, transformar o presente. De onde o futuro constitui mera probabilidade, ele aciona e determina a ação. Curiosamente, invertem-se as posições. São as perspectivas de futuro que, dispostas matematicamente como possibilidades, determinam o curso dos acontecimentos, alterando não só o presente (de modo que presente e futuro trocam de lugar), como também o grau de importância atribuído ao passado. Antes, o presente encontrava suas causas no passado. Agora, o futuro passa a funcionar como uma instância que promove, de dentro do presente, uma mudança e uma rearticulação de suas ações.

Para entender melhor como a questão do tempo se modifica diante das inovações tecnológicas, é preciso pôr em questão a causalidade. Segundo os gregos – e a partir,

sobretudo, do pensamento aristotélico –, o tempo é que estava atrelado a uma dimensão causal. Diferente do que fazemos hoje, ao entendermos os fenômenos dentro de uma temporalidade que asfixia, os gregos não viam o tempo como uma “dimensão essencial própria”. Há que se pensar, portanto, no que acontece com este conceito de tempo no contexto da contemporaneidade, ou como quer o nome, quando todos os tempos aparecem juntos. Há que se pensar se, fora do conceito moderno de Tempo, não seríamos obrigados a recriar uma nova acepção, mais adequada à dimensão temporal que se desenvolve na chamada atualidade. Diante de fenômenos como a Internet, por exemplo, novas noções de espacialidade e temporalidade se fazem necessárias, a fim de exprimir com clareza a forma como experimentamos os novos modos de convivência e relação. Lévy, falando sobre esta necessidade de um novo olhar diante da virtualização, coloca: “[...] a virtualização submete a narrativa clássica a uma prova rude: unidade de tempo sem unidade de lugar (graças às interações em tempo real por redes eletrônicas, as transmissões ao vivo, aos sistemas de telepresença), continuidade de ação, apesar de uma duração descontínua (como na comunicação por secretária eletrônica ou por correio eletrônico). A sincronização substitui a unidade de lugar, e a interconexão, a unidade de tempo” (Lévy, 1996: 21).

Estas noções, que vão se construindo à medida que novas experiências de tempo e espaço se articulam na contemporaneidade, demandam, por sua vez, novas formas de lidar com o novo, a incerteza e a obsolescência crescentes. Diante dos desafios que o cenário da atualidade nos impõe – em que a aceleração e o ritmo de mudança exigem um grau avançado de adaptação –, as novas tecnologias incorporam cada vez mais um componente predictivo, na expectativa de se anteciparem ao futuro, construindo com menos riscos o presente. Os bancos de dados, neste ponto, apresentam-se como um dos ícones principais desta nova configuração tecnológica.

Internet: banco de dados labiríntico

De todas as ferramentas tecnológicas hoje disponíveis, a Internet parece constituir uma espécie de Babel eletrônica, capaz de reunir as características principais de uma infinidade delas. Na verdade, ela funciona como um grande e labiríntico banco de dados, no qual as informações são dispostas de modo a facilitar uma seleção de seus conteúdos de acordo com os gostos individuais de quem a consulta. É interessante observar que a mecânica de rede não é prerrogativa da Internet, podendo-se verificar sua estrutura em outras ferramentas tecnológicas ou mesmo em configurações sociais. Este entendimento mais amplo da rede pode ser identificado nas palavras de Rosenstiehl:

“A nossa época será marcada pelo ‘fenômeno rede’. Como todos os fenômenos morfológicos profundos, de caráter universal, o fenômeno rede per-

tence não só à ciência, mas também à vida social. Cada um de nós se situa em redes, correspondendo cada rede a um tipo de comunicação, de frequência, de associação simbólica” (1988: 228-229).

Assim, o *ethos* rizomático da Internet, modular como é e caracterizando o objeto rede, serve de parâmetro para a análise de várias tendências na contemporaneidade. Inclusive para entender o estado da arte na aplicação das ferramentas *data mining* a partir das quais se buscam informações mais precisas e personalizadas sobre os hábitos de consumo e o estilo de vida dos indivíduos. Isto porque a Internet, mais do que qualquer outro banco de dados, consegue prospectar, voluntária e detalhadamente, grandes volumes de potenciais consumidores. Considerando-se ainda o tamanho gigantesco deste armazém de dados, ela disponibiliza um território fértil e de proporções adequadas para a prática da “mineração” – termo pelo qual se conhece o ato de extrair o ouro da canga, ou se preferirmos, as informações predictivas em meio à profusão de dados disponíveis em rede.

Levando-se em conta as possibilidades de controle por ela enunciadas – segundo Deleuze, são as máquinas cibernéticas e os computadores que caracterizam a emergente sociedade de controle –, a *Web* pode ser lida também como uma espécie de máquina de guerra. Sendo autônoma na sua força e no seu modo próprio de manifestação, esta máquina – lembremos – pode ser estriada por uma subjetividade capitalística³, o que não significa que ela não tenha suas próprias táticas *desviacionistas* (Certeau), de modo a garantir o exercício de produção da subjetividade. Assim, rizomática por natureza, a máquina de guerra constitui uma boa metáfora para entendermos como funciona a Internet. Isto porque tal qual a máquina de guerra deleuziana, “os elementos de sua obra são o segredo, a velocidade e o afecto” (Deleuze, 1997: 17). É interessante observar que o sucesso da Internet hoje, sobretudo no que se refere ao seu uso corporativo, deve-se exatamente a estes três ingredientes. Ao personalizar as estratégias de sedução segundo as expectativas e desejos dos consumidores, as empresas podem promover uma comunicação mais intimista, não de todo visível, mas perfeitamente articulada. É a arte do segredo. A comunicação de massa deixa de ser eficaz por se fundamentar, basicamente, na disseminação de uma mesma e única mensagem para todos, colocando-se visível em toda a sua estratégia para seus oponentes. A comunicação de massa, portanto, não constitui uma máquina de guerra autêntica, ao contrário dos bancos de dados e da Internet (que constitui uma espécie de hibridismo comunicacional).

Do mesmo modo, encontra-se aí o elemento velocidade. Por se caracterizar como rede – acentrada e labiríntica (portanto, com várias alternativas de trajeto) –, a Internet se adequa às demandas da contemporaneidade: permite uma troca acelerada de dados, ao mesmo tempo em que possibilita articulações outras entre tempo e espaço, o que amplia o alcance e o poder das comunicações.

Por fim, mas não menos importante, aparece o elemento “afeto”, este talvez o maior responsável pelo sucesso da *Web* como instrumento de controle e levantamento de dados. Buscando-se o que identifica a Internet como instrumento de comunicação, tem-se que é a interatividade que lhe confere um caráter singular. Nenhuma outra ferramenta de uso corporativo parece envolver tamanha carga de ludicidade, exigindo do internauta que com ela mantenha efetivo contato e grau de interação. É no afeto, portanto, que está o poder de estriamento da Internet, quando ela toma a forma de uma máquina de guerra, capturada pela mecânica da sociedade de consumo. Não é à toa que os *sites* mais visitados são justamente aqueles que incorporam as suas páginas toda uma série de serviços e elementos lúdicos, como *chats* e fóruns em tempo real, promoções *on-line*, formação de comunidades virtuais, entre outros recursos que conferem certa dinâmica às *homepages* corporativas. Dentro deste quadro, o *site* da Brahma, por exemplo, ainda em 1999, oferecia a possibilidade de um “churrasco virtual”: “[...] o churrasqueiro Marcos Bassi, dono de uma das mais famosas casas de carnes do país, que leva seu sobrenome, estará dando aulas multimídia pelo *site* [da Brahma]. Ele também estará comandando o Clube do Churrasco, cujos membros plenos, que quiserem pagar uma taxa de filiação, ganham direito a um *kit* e a encontros mensais em um dos endereços da Bassi; e os virtuais participam de *chats* com os outros membros e podem tirar suas dúvidas com o especialista pela *Web*, em datas predeterminadas”⁴.

Mas a Internet não se fundamenta apenas no lúdico. Ela se alicerça, igualmente, em duas outras bases, constituindo assim um tripé: o conceito de utilidade pública, que lhe atribui credibilidade, e o levantamento identitário dos usuários, baseado principalmente no cadastramento *on-line*. Revestidos do caráter lúdico e cumprindo com estes dois pontos, os *sites* costumam a possibilidade de personalização com a percepção de utilidade, despertada no consumidor. Este, por sua vez, passa a usar a *homepage* corporativa como um grande portal. E não há nada de mais significativo para uma empresa *pointcom* do que ter um usuário que coloque seu *site* na caixa de favoritos. Isto só acontece, naturalmente, quando o elemento “afeto” é urdido de forma contínua, até que uma relação personalizada se tenha construído. Há que se avaliar, naturalmente, os impactos e as conseqüências destas novas técnicas de sedução que se delineiam no cenário contemporâneo. Sobretudo na Internet, onde o nomadismo – estado natural do cibernauta – é posto permanentemente em xeque, em função de uma tentativa contínua de estriamento, ou seja, de mapeamento das trajetórias rizomáticas por ele cartografadas.

Estes três elementos – ludicidade, cadastramento eletrônico e utilidade pública – são os que conferem substância aos *sites* corporativos, calcando-se esta tríade em dois pressupostos básicos que merecem destaque na contemporaneidade. O primeiro é da ordem da economia do dom, e o segundo se fundamenta na economia da atenção (enunciada por Pierre Lévy).

Pode-se dizer que, no contexto atual, as estratégias de sedução da sociedade de consumo tornam-se bem-sucedidas na medida em que investem numa política do dom. Isto significa que o consumidor é seduzido pela sensação de que recebe algo a mais, como na lógica do presente. É este apelo que, tornando um *site* diferente daquele que com ele concorre, faz com que o internauta opte por manter o vínculo. A percepção de ganho é, portanto, fundamental para o estabelecimento dos laços. Antropologicamente, a economia do dom⁵ já constituía uma premissa verificada em algumas sociedades. Um exemplo clássico é o da festa *kula*, na qual braceletes e colares eram trocados pelos trobriandenses, adquirindo tanto mais valor quanto mais circulassem entre os integrantes do grupo.

No caso específico da Internet, não é raro encontramos exemplares que tomem por base uma orientação calcada na política do dom. Tudo o que de alguma forma possa ser transformado em utilidade para o internauta é imediatamente assimilado dentro desta premissa. Por utilidade devemos entender aqui tudo o que possa ser percebido como adicional ao conteúdo básico esperado pelo visitante. Estes elementos extras conferem credibilidade às empresas e, por vezes, servem mesmo para transformá-las em uma espécie de referência na área em que atuam.

Neste contexto, encontram-se também os *links* disponíveis nas páginas do *site* – quanto mais traduzirem a amplitude das necessidades do usuário, dentro de um mesmo quadro temático (ou etimológico, digamos assim), mais úteis eles serão. Deste modo, uma vinícola pode disponibilizar, por exemplo, *links* que levem o internauta a conhecer não só dicas sobre o universo dos *sommeliers*, como também informações que lhe permitam fazer um *fondue* ou preparar receitas especiais envolvendo a gastronomia francesa. Esta proposta de personalização, que caminha para um quadro de individualização – no qual boletins eletrônicos sob medida são enviados para o e-mail do usuário –, pode ser identificada facilmente em algumas páginas da *Web*.

“Fundos de previdência privada, como o Flexprev, do Itaú, e o Prever, do Unibanco, oferecem a possibilidade de simular um plano para garantir a renda desejada em 30 anos. No *site* da Amil, médicos poderão selecionar artigos de publicações especializadas, nacionais e internacionais” (Como caçar clientes no ciberespaço. *Revista Exame*, 21/05/1997).

Do mesmo modo, todo o clima de espetáculo que sustenta alguns *sites* tem por pressuposto uma máxima bastante comum na atualidade, que é a da economia da atenção. Sabe-se que, diante da proliferação de estímulos visuais e mensagens a que é submetido diariamente o homem contemporâneo, a capacidade de assimilação dos códigos vai se tornando cada vez menor, o que implica a necessidade de selecionar aquilo a que se dará atenção. Não tendo mais condições de reter informações no ritmo acelerado com que as recebe, o indivíduo põe-se a selecionar, segundo seus padrões, aquilo que julga de maior utilidade para si.

Em função desta tendência que Lévy convencionou chamar economia da atenção, os *sites* procuram, a todo momento, preencher suas *homepages* com estímulos que garantam maior tempo de permanência do visitante e maior número de consultas aos *sites*. Vale observar, inclusive, que muitos já começam a investir na formação de pequenas tribos virtuais, com a intenção de combater o nomadismo na *net* e criar laços com os consumidores atuais e potenciais. Esta é, sem dúvida, uma forma de estriar o espaço liso da *Web*, na medida em que pequenos centros de aglutinação são formados para garantir um maior controle sobre o trajeto e o tempo de permanência do usuário na Internet.

Esta tentativa de mapear e circunscrever, de certo modo, o traçado sempre novo que o internauta constrói ao percorrer o “labirinto da *Web*”, é sintoma bastante claro do uso da Internet como ferramenta da sociedade de controle. Busca-se, a todo o momento, criar mecanismos que desestimulem o internauta de “flanar” livremente pela *Web*. Há sempre o interesse de mantê-lo o maior tempo possível à vista ou de construir para ele um “fio de Ariadne” que garanta seu retorno (através de *banners*, *interpages*, *mini-sites*), o que constitui, no fim das contas, uma tentativa de estriamento do seu processo de subjetivação. Do mesmo modo, promoções *on-line*, *chats* com especialistas e personalidades públicas, boletins eletrônicos personalizados, *e-mails* promocionais com *link* para a página, todas estas ferramentas servem para manter o consumidor em “solo corporativo”. Se, por um lado, isto demonstra que as empresas estão mais aptas a oferecer serviços de melhor qualidade a seus clientes, por outro, é uma demonstração do quão fortes têm se tornado as estratégias de sedução da sociedade de consumo. Acerca do modo de vida nômade, que bem caracteriza a natureza do papel do internauta, fala Deleuze:

“O nômade tem um território, segue trajetos costumeiros, vai de um ponto ao outro, não ignora os pontos [...]. Mas a questão é diferenciar o que é princípio do que é somente consequência da vida nômade [...]. O ponto de água só existe para ser abandonado, e todo ponto é uma alternância e só existe como alternância. Um trajeto está sempre entre dois pontos, mas o entredois tomou toda a consistência, e goza de uma autonomia, bem como de uma direção própria. A vida do nômade é *intermezzo* [...]” (Deleuze, 1997: 50-51).

A rigor, é importante que se diga, o papel de nômade do internauta nunca poderá ser totalmente estriado, dada a configuração singular e rizomática da Internet. Entretanto, inúmeras técnicas são desenvolvidas com o objetivo de reconhecer o internauta, para garantir seu retorno e permanência no *site*. Para criar táticas desviacionistas (Certeau) que permitam ao consumidor manter-se ao largo das tentativas de monitoramento, é preciso que ele busque assegurar sua posição de *flanêur* no espaço virtual, alisando com seu passo descompromissado o território previamente estriado. É preciso que ele seja um anti-

Ulisses⁶, não se pondo a conquistar o território em que navega, mas, antes, adotando um passo leve, ao sabor do rizoma.

Sabe-se, entretanto, que as estruturas de consumo buscam formas de identificar permanentemente o indivíduo. Na *Web*, isto se dá por meio do cadastramento eletrônico e pela identificação dos *cookies*, espécie de pegada/passos que todo internauta deixa de si ao entrar ou sair das páginas que visita. É pelo cadastramento eletrônico ou pela identificação dos *cookies* que o *site* personaliza suas ofertas e sua comunicação.

Considerando as vantagens dos produtos personalizados, *on-line* ou não, ninguém pensa ser possível ou mesmo viável frear os avanços da Internet nas técnicas de reconhecimento dos seus usuários. Por outro lado, questões pertinentes à privacidade aparecem imediatamente junto com tal procedimento, uma vez que a legislação no Brasil não é clara quanto aos limites desta possível “invasão de privacidade”. Os *cookies*, por exemplo, permitem que as empresas ofereçam produtos e serviços sob medida, segundo a identificação das necessidades e expectativas dos consumidores, conferindo-lhes um caráter benéfico. Entretanto, algumas questões envolvendo os *cookies* merecem atenção. A primeira refere-se ao modo como as empresas utilizam tais pistas que estão nos computadores dos usuários. Um exemplo disso é o cruzamento das informações que constam nos cadastros *on-line* e as que ficam armazenadas nos *cookies* – prática considerada incorreta, pois implica ligar informações anônimas a pessoas reais. A fim de ilustrar como tal procedimento pode de fato ser levado adiante, observe-se o que foi publicado em respeitada revista de negócios, em matéria de capa, sobre a questão da privacidade:

“Alguns *sites* têm armazenado nos *cookies*, sem autorização, informações pessoais dadas pelos usuários nos cadastros *on-line*. Se alguém usar uma única vez o serviço de *e-mail* do *site* brasileiro TurismoNet [...] é exatamente isto que acontece. Sem pedir licença ao internauta, o TurismoNet passa informações do cadastro pessoal ao *cookie*, e reconhece o internauta pelo nome assim que ele retorna ao *site* numa nova visita” (A morte da privacidade? *Revista Exame*, junho de 2000: 34).

A questão não se resume apenas ao cruzamento destas informações, mas envolve, inclusive, a mecânica de rastreamento e incorporação dos dados que os consumidores dão sobre si nos cadastros *on-line*. Por vezes, o internauta desiste de se cadastrar em determinada *homepage* ou, ainda, opta por não mais prosseguir numa compra *on-line*, mas não consegue efetivamente passar anônimo pelo *site*. Este é um caso em que definitivamente não se respeita a privacidade do usuário, quando ele está se cadastrando, independente de ele continuar ou não seu pedido, o que só deveria ser feito após o envio do formulário:

“Quem nunca desistiu de preencher um cadastro no meio do caminho? No mundo do papel, basta rasgar a folha. Mas na Internet não. Na corretora InvestShop, as informações dos internautas

são armazenadas no banco de dados antes mesmo de se pressionar o botão Enviar. Propositadamente, o *e-mail* é o primeiro item solicitado. No que se configura uma invasão óbvia de privacidade, quando alguém desiste do cadastro no meio do caminho, o InvestShop entra em contato por e-mail, convidando a pessoa a finalizar a sua ficha” (p. 35).

Há que se perguntar, no entanto, o que leva uma pessoa, com não pouca frequência, a se cadastrar eletronicamente nos *sites* que visita. Há que se entender, ainda, qual a mecânica que faz com que os consumidores disponibilizem seus dados na Internet. As estatísticas comprovam que, na maior parte das vezes, os internautas não se opõem a falar de si, estando inclusive mais aptos a responder a pesquisas e cadastros *on-line* do que por mala direta ou telefone. Sabemos que, em parte, isto se deve a uma política do dom, na qual os cadastros são preenchidos em troca de brindes, participação em sorteios e promoções, boletins eletrônicos com dicas, entre outras vantagens.

Curiosamente, o anonimato total – comportamento que seria próprio de um nômade como o internauta – não é bem-visto ou aceito na *net*. Mesmo nas salas de bate-papo é comum vermos pessoas usando o mesmo *nick*, na tentativa de criarem um padrão, assumindo uma *persona*. De um modo geral, é até considerado comum deixar pistas, construir um trajeto, valer-se do fio – haja vista o número de integrantes do ICQ (espécie de instrumento de localização instantânea de pessoas que estão *on-line*). Algo da ordem do anonimato parece enunciar a impossibilidade de o indivíduo constituir-se como presença no mundo virtual.

Pode-se argumentar que esta tentativa de barrar o anonimato seria uma forma das empresas preservarem a ordem e manterem seguros os internautas que estão *on-line*. Entretanto, há que se considerar também as aplicações até então desconhecidas destas informações extraídas dos bancos de dados. Cruzamentos de *databases* diversos, bem como envio indiscriminado de correspondência eletrônica, são práticas consideradas indevidas, mas os pequenos detalhes envolvendo estas transações, nós não os conhecemos. Isto significa que o limite, não sendo definido coletivamente, não pode ser monitorado. Obviamente, o anonimato na *Web* é de toda forma desestimulado. O internauta que se aventura a deletar os *cookies*, capazes de rastrear suas preferências, chama para si uma espécie de castigo:

“*Cookies* em si não atrapalham ninguém, se apropriadamente usados. Mas tente desabilitá-los [...] e prepare-se para pagar caro pelo anonimato. Mergulha-se num mundo de senhas, preferências que precisam ser renovadas a cada visita, páginas que não abrem. Os portais latino-americanos StarMedia e O Site simplesmente barram a entrada de qualquer pessoa sem *cookies*. Participar de um bate-papo do UOL, o maior provedor de acesso e conteúdo brasileiro, com cerca de 700.000 assinantes? Impossível. Ali os sem-*cookies* não entram [...]” (*ibidem*, 33).

Existem formas na *Web* de se alisar este espaço profundamente estriado. Programas com agentes inteligentes do tipo *Webwasher* simplesmente apagam os vestígios deixados na *net*, inclusive os *cookies*. O internauta volta então à condição de anonimato, uma espécie de *flanêur* passeando pelos caminhos rizomáticos da *Web*. Este é um procedimento que, sem dúvida, constrói linhas de fuga, capazes de quebrar as linhas de poder que se arranjam em torno da sociedade de consumo. Cabe avaliar, entretanto, até que ponto o sujeito está preparado para viver o anonimato. Ou melhor, cabe investigar até que ponto o indivíduo está pronto a abdicar de sua privacidade em função das vantagens que um relacionamento com as estruturas de consumo pode lhe proporcionar.

Bancos de dados, Internet e aplicações *data mining*: ponto de encontro

Partindo do fato de que a Internet consiste, na verdade, em um grande banco de dados, no qual quase todo o tipo de informação sobre consumidores e *prospects* pode ser extraída, resta falar sobre os mecanismos de leitura dos dados que nela se encontram.

A Internet, em função do volume de dados que reúne, é o modelo ideal para prospecção de informações via *data mining*. Sabe-se que a aplicação destas ferramentas e técnicas com base em modelos matemáticos e estatísticos é tanto mais eficaz quanto maior for a quantidade de dados reunidos. Vale lembrar que, embora se utilize de algoritmos, redes neurais, modelagem predictiva e outros procedimentos estatísticos, o *data mining* é mais do que um conjunto de ferramentas ou técnicas. Na verdade, ele é um processo que, na medida em que se torne uma filosofia corporativa, tenderá a reformular os padrões de comunicação entre as estruturas de consumo e os *prospects*, uma vez que efetivamente propiciará um modo novo de comunicação. Com ele, já é possível ler as informações invisíveis entre dados não-correlacionados, prever tendências de consumo, analisar riscos e formar células específicas de consumidores. Jesus Mena, explicando as premissas do *data mining* na Internet, enuncia o conceito que está por trás da prática. O *data mining* é muito mais do que uma ferramenta de *software*; é um processo que envolve métodos e processos para extrair e preparar dados. O *data mining* pode envolver a captura estratégica de informações do visitante por meio de formulários de registro e do cruzamento desses dados *on-line* com estatísticas adicionais, elaborando uma análise por meio de poderosos *softwares* de reconhecimento de padrão para a criação de modelos predictivos (1999:6).

Numerosas são as aplicações do *data mining* no campo do consumo. Se os bancos de dados inauguraram um modo novo de conhecer os clientes e estruturar a comunicação segundo as prerrogativas individuais de cada um deles, o uso do *data mining* permitiu ir além, prospectando

em meio a estes dados, aquilo que realmente constitui o nexo entre as informações. Um exemplo clássico de como rastrear dados invisíveis pode ser identificado no caso da Wall Mart, maior cadeia de varejo do mundo:

“Com base em técnicas de inteligência artificial [...] [estes softwares] ficam vasculhando os dados em busca de informações que possam ser de interesse, de acordo com critérios pré-determinados. São usados por grandes cadeias varejistas para descobrir, por exemplo, que quem compra fraldas descartáveis é um consumidor potencial de cerveja ou que quem leva para casa sandálias havaianas pode estar interessado em CDs do Gilberto Gil” (O que cerveja tem a ver com fraldas. *Revista Exame*, s.d.).

O que confere ao *data mining* singularidade é justamente esta sua capacidade de ler entre as dobras, de descobrir nexos entre informações aparentemente sem ligação, permitindo detectar ou criar parâmetros que auxiliem na análise do comportamento do consumidor. Por esta razão, pode-se com ele prever tendências neste comportamento, identificar pesos de determinadas variáveis na definição de perfis de consumo, bem como analisar possibilidades de fraude e riscos. Do mesmo modo, o *data mining* permite fazer associações, descobrindo pontos de ligação entre o consumo de dois ou mais produtos.

De modo geral, portanto, a aplicação de suas ferramentas e técnicas tem por finalidade: 1) descobrir padrões escondidos aleatoriamente (“descobrimento é o processo de *examinação* em um banco de dados para encontrar padrões escondidos sem uma idéia ou hipótese pré-determinada sobre o que são esses padrões”);⁷ 2) modelagem de prognóstico; 3) análise prévia (“é o processo de aplicação dos padrões para encontrar anomalias ou elementos de dados raros”)⁸.

Aplicando modelos matemáticos e estatísticos que utilizam correlações, análise de variância, regressão linear, métodos de predição (*forecast method*) e *data visualization*, entre outros, o *data mining* enuncia uma possibilidade de monitoramento dos níveis e das oscilações de consumo que os *databases* corporativos simples não conseguem realizar. Trata-se, portanto, de um mecanismo de estriamento muito mais poderoso porque identifica, na base, as cadeias rizomáticas de sentido, mesmo quando apontam para correlações aparentemente desconexas para o entendimento humano.

Eis por que o uso destas ferramentas traz para a agenda de discussões da contemporaneidade a questão da algoritmização do conhecimento. Inúmeros pontos acerca da capacidade de simulação da mente humana (ou mesmo de, aprendendo como ela, superá-la no descobrimento das cadeias de sentido) demandam maior estudo. Considerando-se os avanços da Probabilística e da Lógica, bem como da Inteligência Artificial, uma multiplicidade de novas questões surge diante das possibilidades bem mais refinadas de estriamento dos processos de subjetivação. Uma vez que cadeias de sentido podem se fazer revelar mesmo na impossibilidade de compreensão do homem – modelo do qual inicialmente se partiu para fundamentar o aprendizado das redes neurais –, há que se tentar descobrir que padrão é este diante do qual mesmo o pensamento do homem parece estar disposto a sucumbir. Sob pena de tudo não passar de mero jogo de simulação.

Notas

¹ Vale aqui uma referência à leitura deleuziana sobre a obra de Leibniz, que utiliza permanentemente o conceito de dobra. Se redimensionarmos o termo hoje, veremos que o conceito de dobra pode adaptar-se perfeitamente ao contexto contemporâneo, tanto como figura mental que traduz os infundáveis caminhos rizomáticos da mecânica de rede (e assim funcionam os bancos de dados), como também enquanto “dobra que o pensamento faz no processo de subjetivação” (expressão usada por Roberto de Maggiori). Diz Deleuze sobre o

conceito de dobra em Leibniz: “Desde o pré-formismo do século XVII até a genética de hoje, a dobra mudou de natureza, de função, de sentido. Mas, vejamos: Leibniz mesmo não inventou a noção e a operação da dobra [...]. No entanto, foi o primeiro pensador a ‘liberar’ a dobra, levando-a ao infinito [...]” (Deleuze, 1992:197).

² Para obter dados adicionais sobre o *data mining*, ver monografia publicada na Internet. O endereço é: <http://w.w.w.infolink.com.br/~mpolito/mining/mining.htm>

³ Para maiores informações, consultar “Tratado de Nomadologia”. In: *Mil Platôs*, vol. 5.

⁴ Ver http://www.about-net.com.br/001/2_nmkt/2_07_15.htm

⁵ Termo adotado por Henrique Antoun.

⁶ Termo adotado por Henrique Antoun.

⁷ Ver <http://www.infolink.com.br/~mpolito/mining/mining.htm>

⁸ *Ibidem*

Bibliografia

- AMARAL, M.T. *Contemporaneidade e novas tecnologias*. Rio de Janeiro: Sette Letras, 1996.
- ATLAN, Henri. *Entre le cristal et la fumée*. Paris: Seuil, 1979.
- COLLI, Giorgio. *O nascimento da Filosofia*. Lisboa: Ed. 70, 1998.
- DELEUZE, Gilles. “Controle e devir”. *Conversações*. Rio de Janeiro: Ed. 34, 1992.
- _____. “*Post-scriptum* sobre as sociedades de controle”. *Conversações*. Rio de Janeiro: Ed. 34, 1992.
- _____. “O que é um dispositivo?” *O mistério de Ariana*. Lisboa: Vega, 1997.
- DELEUZE, Gilles & GUATTARI, Félix. *Mil Platôs – Capitalismo e esquizofrenia*. Vol. 1 e 5. Rio de Janeiro: Ed. 34, 1995/1997.
- FOUCAULT, Michel. *O que é um autor?* Lisboa: Vega, 1992.
- LÉVY, Pierre. *O que é o virtual?* São Paulo: Ed. 34, 1996.
- _____. *A máquina universo – Criação, cognição e cultura informática*. Porto Alegre: ArtMed, 1998.
- _____. *As tecnologias da inteligência – O futuro do pensamento na era da informática*. Rio de Janeiro: Ed. 34, 1993.
- MENA, Jesus. *Data mining your website*. Digital Press, 1999.
- ROSENSTIEHL, Pierre. “Rede”. *Enciclopédia Einaudi*. Porto: Impr. Nacional, 1988.

*Luciane Lucas é Mestre pela ECO/UFRJ e doutoranda em Comunicação e Cultura pela mesma instituição.