



RAC - Revista de Administração
Contemporânea

ISSN: 1415-6555

rac@anpad.org.br

Associação Nacional de Pós-Graduação e
Pesquisa em Administração
Brasil

da Silva Añaña, Edar; Vieira, Leandro Maurício; de La Martinière Petroll, Martin; Petersen-Wagner,
Renan; Simm Costa, Ricardo

As Comunidades Virtuais e a Segmentação de Mercado: uma Abordagem Exploratória, Utilizando
Redes Neurais e Dados da Comunidade Virtual Orkut

RAC - Revista de Administração Contemporânea, vol. 12, núm. Esp, 2008, pp. 41-63
Associação Nacional de Pós-Graduação e Pesquisa em Administração
Rio de Janeiro, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=84009403>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica
Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

As Comunidades Virtuais e a Segmentação de Mercado: uma Abordagem Exploratória, Utilizando Redes Neurais e Dados da Comunidade Virtual *Orkut*

Virtual Communities and Market Segmentation: an Exploratory Approach, Using Neural Networks and Virtual Community's Data

Edar da Silva Añaña *

Doutorando pelo PPGA/EA/UFRGS.
Professor do UFPEL, Pelotas/RS, Brasil.

Leandro Maurício Medeiros Vieira

Mestre pelo PPGA/EA/UFRGS.
Editor do Portal Administradores.com.br, João Pessoa/PB, Brasil.

Martin de La Martinière Petroll

Mestre pelo PPGA/EA/UFRGS, Porto Alegre/RS, Brasil.

Renan Petersen-Wagner

Mestre pelo PPGA/EA/UFRGS, Porto Alegre/RS, Brasil.

Ricardo Simm Costa

Doutorando do PPGA/EA/UFRGS, Porto Alegre/RS, Brasil.

*Endereço: Rua Gonçalves Chaves, 3063, apto. 201-B, Centro, Pelotas/RS, 96010-570.
E-mail: edar@ufpel.edu.br

RESUMO

A existência de uma infinidade de dados, nos últimos anos, tem disponibilizado para a área de marketing novos horizontes para a definição e segmentação de mercados, auxiliada por bancos de dados, tanto privados e internos da organização, quanto públicos e externos. Um exemplo do segundo tipo são as comunidades virtuais, sendo o *Orkut* um dos seus expoentes. Este artigo, de caráter exploratório, tem o intuito de apresentar duas formas alternativas de segmentação de mercado (regressão logística e redes neurais), por meio de dados secundários coletados no *Orkut*, pressupondo que é possível prever determinadas atitudes de consumo explicitadas por membros das comunidades virtuais. Para tanto foram escolhidas duas comunidades autodenominadas **eu amo cerveja gelada** e **eu odeio cerveja**, das quais foram coletadas amostras probabilísticas aleatórias de 400 membros cada. Com o processamento e a análise dos dados desenvolvidos em três fases - limpeza de dados e seleção de variáveis de interesse, análise discriminante e análise através de redes neurais - confirmou-se a possibilidade de se determinar atitudes de consumo e, com isso, utilizar comunidades virtuais como bancos de dados para segmentação. As demais contribuições, limitações e implicações constam no estudo.

Palavras-chave: segmentação; marketing; comunidade virtual; redes neurais; regressão logística.

ABSTRACT

During the last few years, endless data have supplied the marketing area with new perspectives for the definition and segmentation of markets, in addition to the company's private and internal data banks as well as of public and external ones. An example of the latter are the virtual communities, and Orkut is one of its exponents. This article, offers an exploratory approach to market segmentation using two alternative methods (logistic regression and neural networks) through secondary data collected from Orkut, assuming the possibility to predict certain consumption attitudes described by members of virtual communities. This way, two communities self denominated **I love cold beer** and **I hate beer** were picked up, and each one supplied random probabilistic samples among 400 members. Processing and analyzing the data collected in three phases - data cleaning and selection of variables of interest, discriminating analysis, and analysis through neural networks - confirmed the possibility of determining consuming attitudes, and consequently of using the Orkut virtual community as a data bank for segmentation. Further contributions, limitations and implications are part of this study.

Key words: segmentation; marketing; virtual community; neural networks; logistic regression.

INTRODUÇÃO

Os profissionais de marketing nunca tiveram acesso a tantos dados para tomar decisões como atualmente. Com o avanço da microinformática, da Internet e o desenvolvimento de sistemas específicos para gestão de banco de dados, a área de marketing passou a contar com uma estrutura bastante robusta para definir e segmentar mercados. De maneira mais específica, o surgimento das comunidades virtuais mostra-se uma fonte rica e ainda pouco explorada para definir diferentes segmentos de mercado e desenvolver estratégias para abordá-los de forma mais efetiva. Os membros deste tipo de comunidade explicitam características pessoais, sociais e psicológicas que, se devidamente avaliadas pelos profissionais de marketing, podem constituir interessante fonte de informação para a segmentação de mercado, mormente porque a participação em comunidades virtuais torna possível a caracterização dos consumidores com base em seus estilos de vida e interesses.

Para que isto seja possível, no entanto, é necessário possuir métodos, técnicas e ferramentas que permitam o adequado tratamento e uso dessas bases de dados. Assim, é objetivo deste trabalho apresentar um comparativo entre duas técnicas de tratamento de dados: a regressão logística e as redes neurais; tem-se em vista verificar as forças e limitações de cada uma delas na identificação de segmentos de consumidores.

O estudo é apresentado da seguinte maneira: na seção 2, é realizada uma revisão da literatura com relação ao uso e aplicações da segmentação de mercado, apresentando as principais bases de segmentação. Na seção 3, discutem-se as comunidades virtuais e o comportamento de seus membros. Em seguida, são apresentados alguns dados da comunidade virtual, na qual se enfoca este trabalho, e as variáveis utilizadas para a segmentação. Na seção 4, é realizado um comparativo entre os métodos de segmentação utilizados, enfatizando os procedimentos realizados; e, na seção seguinte, apresentam-se os resultados obtidos. Finalmente, na seção 6 são realizadas as conclusões e encaminhamentos cabíveis.

SEGMENTAÇÃO

A segmentação de mercado, segundo Kamineni (2005), foi e é fundamental para a estratégia de marketing de determinada organização, sendo o seu sucesso

vinculado à complexa tarefa de compreender os consumidores que possuem características diferentes entre si e que, portanto, podem reagir de maneiras diferentes a um determinado apelo (Schiffman & Kanuk, 2000). Logicamente, se todos os indivíduos fossem iguais, tivessem as mesmas necessidades e reagissem de maneiras semelhantes a um mesmo apelo, de nada adiantaria uma organização realizar a segmentação de mercado de seus produtos.

Assim, a proposta da segmentação de mercado, nome este surgido em 1956 por Smith (como citado em Lin, 2002, p. 249), é identificar a taxonomia dos padrões de consumo, mediante a divisão do mercado em alguns submercados homogêneos, onde se inserem os consumidores que possuem necessidades, características e comportamentos semelhantes. Kotler (2000) e Schiffman e Kanuk (2000) ressaltaram alguns critérios que precisam ser levados em consideração, no momento em que se decide partir para a segmentação de mercado. É necessário, na visão dos autores, que os segmentos, por meio de suas características, sejam: (1) identificáveis mediante algum método; (2) de tamanho suficiente, ou seja, que possuam um número adequado de indivíduos, justificando, assim, economicamente uma ação de marketing específica; (3) estáveis no seu comportamento por determinado período de tempo, de modo que a ação de marketing empregada tenha o efeito planejado; e (4) acessíveis à organização, em termos de comunicação e distribuição, a um custo viável.

Kamineni (2005) afirma que o segmento de mercado deve ser também heterogêneo em relação aos demais segmentos, ou seja, que as características, necessidades e comportamentos dos consumidores de um segmento sejam, quanto possível, diferentes daqueles dos consumidores de outros segmentos do mercado. A diferenciação entre conjuntos de indivíduos advém da seleção de bases de segmentação. A seguir, apresentam-se as principais bases de segmentação.

Bases de Segmentação

Schiffman e Kanuk (2000) identificaram sete bases para a segmentação do mercado. São elas: (1) geográfica; (2) demográfica; (3) psicográfica; (4) sociocultural; (5) relacionada com o uso; (6) por uso-situação; e (7) por benefício. Já Kotler (2000) e Kamineni (2005), apesar de abordarem intrinsecamente as sete bases de segmentação acima, conferiram a elas uma divisão mais concisa, com mesmo conteúdo e, pela visão dos autores deste trabalho, mais interessante: (1) segmentação geográfica; (2) demográfica; (3) psicográfica; e (4) comportamental.

A **segmentação geográfica** leva em consideração a localização do indivíduo e parte da premissa de que indivíduos residentes em regiões semelhantes

tendem a possuir características similares, que diferem dos residentes de outras regiões.

Por sua vez, a **segmentação demográfica**, uma das bases mais utilizadas (Kotler, 2000), em função de sua facilidade na quantificação e classificação do consumidor, identifica características como idade, gênero, estado civil, número de filhos e escolaridade. Todavia, muitas vezes, a segmentação demográfica é pouco sensível no que se refere ao entendimento de todas as variáveis que interferem no processo de tomada de decisão do consumidor, não sendo, portanto, capaz de retratá-lo por completo (Oates, Shufeldt, & Vaught, 1996). Isso acontece, segundo Lin (2002), porque os consumidores pertencentes ao mesmo grupo demográfico podem apresentar diferentes perfis psicográficos.

Dessa forma, os dados demográficos são, muitas vezes, utilizados na **segmentação psicográfica**, que não avalia somente quem está no segmento, mas principalmente o porquê do consumidor estar em tal segmento (Thompson & Kaminski, 1993). Assim, as variáveis psicográficas levam em consideração atividades, interesses, opiniões (*Activities, Interests and Opinions* [AIO]), necessidades, valores, atitudes e traços de personalidade do consumidor (Kamineni, 2005). Inseridos, portanto, na segmentação psicográfica, estão os padrões de estilo de vida do consumidor, que provêm compreensão mais ampla do comportamento dos indivíduos, o que enriquece o processo de definição do mercado-alvo pelo profissional de marketing (Shufeldt, Oates, & Vaught, 1998).

Os padrões de estilo de vida do consumidor são definidos por meio da inclusão do já referido AIO onde: (1) as atividades são o modo como o consumidor despende seu tempo e dinheiro; (2) os interesses são tudo aquilo que está à sua volta e que considera mais ou menos importante; e (3) opiniões são como os consumidores se vêem e como eles vêem o mundo à sua volta (Gonzalez & Bello, 2002). Dessa forma, o uso do estilo de vida apresenta uma visão geral do mercado sob forma multidimensional, tornando-se importante fonte de informação para o profissional de marketing (Shufeldt, Oates, & Vaught, 1998).

Além dos estilos de vida, a segmentação psicográfica também possui outra faceta: a da personalidade (Lin, 2002). Nesse sentido, mesclam-se as duas, resultando na chamada *Values and Lifestyle Segmentation* [VALS]. Segundo Kahle e Kennedy (1989), a idéia principal empunhada pela VALS era de que os consumidores comprem produtos em parte para refletir seus valores e estilos de vida. Assim, está intrínseco o princípio da abstração, ou seja, associando um conceito abstrato (por exemplo, um valor) a algo específico (por exemplo, um produto) faz-se com que o algo específico do produto **receba** afeições positivas, associadas ao conceito abstrato do valor. A VALS, desenvolvida por Mitchell em 1983, baseou-se na hierarquia das necessidades de Maslow e no conceito da

característica social (Kahle, Beatty, & Homer, 1986). Todavia, após críticas de muitos estudiosos, foi desenvolvida, em 1989, a VALS2. Esta, além de definir melhor os segmentos de mercado, leva em consideração mudanças sociais e econômicas. A diferença, portanto, entre a VALS e a VALS2 é que a primeira se preocupa com o que o consumidor valoriza, enquanto a segunda explica por que e como o consumidor decide comprar certo produto (Gates, 1989).

Já a *List of Values* [LOV], ou seja, a lista de valores, foi desenvolvida por Kahle em 1983 e serviu como instrumento de medida do valor no estudo das tendências das similaridades e diferenças do consumidor, ou seja, na natureza do consumidor (Kahle & Kennedy, 1989). A vantagem de se utilizar LOV ao invés de VALS é que esta última estuda demasiadamente aspectos demográficos, enquanto LOV estuda, com mais afinco, o comportamento do consumidor. Além disso, LOV é mais interessante que *Rokeach Value System* [RVS], pois ela, além de ser baseada na RVS, relaciona-se mais proximamente com a vida e com as situações diárias das pessoas, além de não ter certos problemas metodológicos que a RVS apresenta (ver Kahle & Kennedy, 1989).

Um dos pontos, contudo, que dificulta a utilização das bases psicográficas nas organizações é a dificuldade de identificação e mensuração das variáveis que as compõem (Schiffman & Kanuk, 2000). No meio acadêmico, o desenvolvimento de escalas específicas e a proliferação de técnicas de validação de escalas multi-ítems, tais como o alfa de Cronbach e a análise fatorial, tornaram bastante atrativas as variáveis psicográficas, que na sua maioria não são mensuráveis diretamente por meio de uma questão (Wind, 1978).

Com a divulgação da microinformática e o advento de técnicas estatísticas avançadas, parece não haver mais razão para utilizar uma única base de segmentação, já que são muitos os critérios que determinam a resposta do consumidor ao estímulo. Entretanto, conforme o que foi escrito até aqui, parece que a maioria dos pesquisadores ainda utiliza uma única base de segmentação.

Numa tentativa de ampliar o espectro da segmentação, composição híbrida, Wind (1978) apresentou dois métodos capazes de oferecer contribuições importantes: (1) segmentação *a priori*, em que o pesquisador determina previamente as variáveis de interesse, levando em consideração diretamente as bases de segmentação - para, após, classificar os diferentes segmentos; e (2) segmentação *a posteriori*, em que o pesquisador seleciona um conjunto de variáveis inter-relacionadas, que podem pertencer a diferentes bases de segmentação, e busca definir segmentos por meio de análises estatísticas ou de agrupamentos, a partir das similaridades. Na segmentação *a posteriori*, portanto, a definição de bases de segmentação é relegada a um segundo plano, e a análise estatística passa a ter papel mais importante.

A partir da segmentação *a posteriori*, apresentam-se os procedimentos metodológicos para a segmentação de mercado: (1) decidir as variáveis-bases (dependentes) e as que as descrevem (independentes) para realizar a segmentação; (2) decidir a metodologia de análise de dados; (3) aplicar a metodologia para identificar vários segmentos; (4) descrever todos os segmentos, usando as variáveis dependentes e as variáveis independentes; (5) selecionar os melhores segmentos para direcionar atividades de marketing; e (6) desenvolver o marketing *mix* para cada segmento-alvo (Porto & Torres, 2005).

Nesta linha, Ramaswamy, Chatterjee e Cohen (1996) apresentam uma metodologia inovadora para a identificação de segmentos de mercado, que vai além da segmentação descritiva. A partir de um conjunto de variáveis inter-relacionadas, os autores procuraram, além de agrupar indivíduos com características semelhantes, definir equações capazes de prever o comportamento de cada segmento definido, por meio de **modelos de segmentação latente**. Tal metodologia, apesar de ainda estar em fase de consolidação, possui potencial de aplicação e vem recebendo a atenção de estudos na área. Um desses estudos, que converge com o foco do presente trabalho, é o de Fonsêca e Omaki (2004), um dos poucos trabalhos publicados no Brasil sobre a identificação e previsão de segmentos de mercado mediante o uso das Redes Neurais.

COMUNIDADES VIRTUAIS

Segundo Tajra (2002), o termo **comunidades virtuais** foi criado em 1993, com o seguinte significado: “(...) agregações sociais que surgem da Internet, quando pessoas suficientes mantêm suficientes debates públicos, com suficiente sentimento humano, para formar teias de relacionamento no **ciberespaço**” (p. 37). Nota-se, primeiramente, que mesmo nas relações decorrentes de ambientes ditos virtuais, haveria criação de sentimentos entre os integrantes das comunidades. Em segundo lugar, no significado está presente a conotação de público, o inverso de privado, fazendo com que tanto os dados quanto as informações e os debates etc. sejam de livre acesso.

A origem das comunidades virtuais, conforme Castells (2003), está muito associada aos movimentos de contracultura e às maneiras alternativas de vida que emergiram no final dos anos 60. Além disso, com o advento da Internet, elas se tornaram instrumento essencial para arquivar, informar e trocar conhecimento no mundo corporativo, reproduzindo comunidades reais típicas, como vizinhanças, clubes ou associações (Turban & King, 2004).

Assim como as comunidades reais, as virtuais se agrupam segundo afinidades de interesses, de conhecimentos, sobre projetos mútuos, em um processo de cooperação ou de troca, não havendo, por fim, barreiras geográficas (Levy, 1999). Por serem comunidades por escolha, diferentemente das comunidades **compulsórias** reais, os membros das comunidades virtuais são capazes de abandoná-las sem aviso e com pouco custo pessoal. Porém eles se filiam a elas porque se identificam com os seus propósitos e valores, permitindo que as comunidades virtuais exerçam considerável influência na definição de quem uma pessoa é como ser humano. É o caso, por exemplo, do *Orkut*, comunidade virtual com enorme penetração no Brasil.

O Exemplo do *Orkut*

O *Orkut* é um dos mais recentes fenômenos de agremiação em escala mundial. Trata-se de um *software* norte-americano, criado por Orkut Buyukokkten e lançado pelo *Google* em janeiro de 2004. Recuero (2006) o descreve como uma espécie de conjunto de perfis de pessoas e suas comunidades virtuais. O *Orkut* foi desenvolvido com base na idéia de “*software* social”, sistema que visa proporcionar conexões entre as pessoas, incentivando a formação de grupos e comunidades, tal qual uma organização social.

Para fazer parte do *Orkut*, é necessário ser convidado por um membro efetivo. O primeiro integrante foi o seu criador, que convidou seus conhecidos que, por sua vez, convidaram seus conhecidos e assim sucessivamente, até envolver praticamente todos os países do mundo com acesso à Internet.

O *software* permite ao usuário criar o seu perfil, incluindo fotos, descrição, atributos físicos, detalhes de sua personalidade, preferências, entre outros aspectos, além de possibilitar a formação de uma rede de **amigos** e a criação e participação de **comunidades**, com propósitos variados o mais possível. Esse perfil é visível a todos os demais membros, que até podem visualizar as conexões diretas (amigos) e indiretas (amigos dos amigos), e convidar os amigos dessa pessoa a fazerem parte de sua rede pessoal. Além disso, há diversas formas de interação, por meio de sistemas de fóruns, mensagens privadas e públicas etc.

Apesar de ter surgido nos Estados Unidos, os brasileiros são maioria no *Orkut*. Segundo dados divulgados no próprio *Orkut*, 74% dos usuários são brasileiros, enquanto os norte-americanos ocupam a segunda posição na escala de nacionalidades, com apenas 9% de participação.

Sendo a imensa maioria dos membros de nacionalidade brasileira, o *Orkut* pode ser considerado um *software* valioso de marketing no país, possibilitando nova forma de conversação e sendo capaz de gerar reações sociais em cadeia

em torno de um produto, de uma idéia, de uma marca, além de fornecer dados potencialmente preciosos no que tange à segmentação de mercado.

Comunidades Virtuais x Marketing

A análise das comunidades virtuais sob a ótica do marketing é fenômeno novo (Solomon, 2002). Trabalhos nacionais estudaram outras questões como a relação de *websites* com *brand equity* (Cunha & Cunha, 1999), com valor para o cliente (Koetz, 2004), como forma de vantagem competitiva (Yamashita & Gouvea, 2004). Além destes, outros estudos referiram-se às ferramentas de marketing e sua potencial utilização em ambientes virtuais (Gonçalves, Gonçalves, Soares, & Santo, 1998; Ramos-e-Silva & Oliveira, 1997), às mudanças estruturais nas organizações com o ambiente virtual, como forma de marketing (Cunha, 2000), além de técnicas de pesquisa que poderiam ser utilizadas em ambiente virtual, como o *laddering* (Leão & Mello, 2001, 2002) e a *netnografia* (Rocha, Barros, & Pereira, 2005).

Sobre este último, Rocha *et al.* (2005) consideraram a **netnografia** como a parte escrita de uma **cibercultura** compreendida pela antropologia cultural. E, tal qual outras técnicas qualitativas, a **netnografia** passa pelos mesmos problemas e preocupações referentes à validade dos seus achados (Kozinets, 1997). Apesar disso, ele descobriu em seu estudo, uma subcultura de admiradores do seriado *X-Files*, unidos por sentimentos, compartilhando gosto pelo mistério, pelo estético, e um desejo de consumir bens relacionados ao seriado. Nesse último, percebe-se a troca de informações sobre produtos, e até a troca de produtos entre os membros das comunidades, aspectos esses interessantes para o profissional de marketing.

Já Ward (1999) apresentou a **cyber-etnografia**, termo por ele cunhado para diferenciá-la da etnografia tradicional. Segundo a autora, a diferença está na ausência de delimitação do espaço físico, já que a **cyber-etnografia** estuda comunidades que estão baseadas em um ciberespaço. Ward (1999) ainda criticou o uso indiscriminado de técnicas construídas para serem utilizadas em ambientes reais em ambientes virtuais. Por fim, a autora define **cyber-etnografia** como o estudo de interações *on-line*, permitindo que os sujeitos estudados interajam durante o processo de pesquisa, parte esta essencial daquele método.

Por fim, existem outros estudos inseridos no marketing e, mais especificamente, no comportamento do consumidor. Um exemplo é o estudo de Langerack, Verhoef, Verlegh e Valck (2004), que buscou entender as razões para a participação dos usuários em comunidades virtuais, que seriam a satisfação do contato com outros membros, entre o organizador da comunidade e os membros, e com o *site* que hospeda a comunidade virtual. Outro exemplo é o estudo de Mattar (2003), o qual defende que as subculturas são entendidas, nas comunidades virtuais, como

grupos de pessoas que compartilham crenças e valores confinados a um espaço virtual compartilhado por pessoas de todo o mundo.

Nota-se que os estudos que envolvem comunidades virtuais e marketing estão muito ligados a métodos interpretativos (ex.: **netnografia** e **cyber-etnografia**), enquanto estudos que utilizam os dados postados pelos membros dessas comunidades (por exemplo o perfil e os interesses) como fonte secundária para trabalhos de natureza quantitativa, ainda não foram desenvolvidos em sua plenitude. Desta maneira, o presente trabalho busca suprir, mediante a tabulação de padrões de consumo e do adequado processamento dos seus dados secundários, esta importante lacuna de conhecimento.

METODOLOGIA

O presente trabalho é pesquisa exploratória que utiliza dados secundários existentes na comunidade de relacionamento *Orkut*, para determinar possíveis atitudes em relação a um produto de consumo (cerveja), dados estes referentes ao perfil, preferências e hábitos, que são voluntariamente veiculados pelos próprios interessados.

A pesquisa parte do pressuposto de que é possível prever determinadas atitudes de consumo, como gostar ou não de cerveja, a partir do cruzamento e interpretação de outras informações explicitadas pelos membros das comunidades virtuais. Para isso busca identificar, no emaranhado de dados existentes no *Orkut*, algumas variáveis que, tratadas estatisticamente e utilizando técnicas de mineração de dados, permitam prever aquelas atitudes de consumo, com vistas à segmentação de mercado.

Seleção e Coleta das Amostras

O trabalho utiliza duas amostras probabilísticas de 400 elementos cada, colhidas das comunidades autodenominadas **eu amo cerveja gelada** (52.000 membros) e **eu odeio cerveja** (32.000 membros). A amostragem probabilística aleatória simples, conforme Malhotra (2001), é aquela em que “cada elemento da população tem uma probabilidade conhecida e igual de escolha e na qual a amostra é extraída de um arcabouço amostral por processo aleatório” (p. 309). Para a pesquisa, foram colhidos aleatoriamente os *logins* de 400 usuários de cada uma das duas comunidades incluídas no estudo, através do sorteio de 400 telas, cada tela com cerca de 15 usuários, das quais foram capturados os *links* dos usuários localizados no canto superior esquerdo. Após

a captura dos *links*, realizada numa única passagem para garantir a aleatoriedade, foi feita a coleta dos dados correspondentes a 787 usuários, uma vez que dos 800 links sorteados, 13 não apresentavam dados ou não estavam mais disponíveis por ocasião da coleta, correspondendo a uma mortalidade de 1,6%.

Processamento e Análise dos Dados

A mineração de dados tem orientação tipicamente exploratória de busca por conhecimento obscurecido pelos complexos padrões de associação e grandes quantidades de dados. Segundo Hair, Anderson, Tatham e Black (2005), “umas poucas ‘pépitais’ podem ser encontradas somente depois de se processar vastas quantidades de informação” (p. 538) e, para isto, diversas técnicas analíticas podem ser empregadas, desde as mais básicas (abordagens descritivas ou gráficas) até análises multivariadas sofisticadas: análise de agrupamentos, de regressão ou logística, além de redes neurais e algoritmos genéticos.

O processamento e a análise dos dados utilizados no presente trabalho foram desenvolvidos em três fases: 1) limpeza de dados, dicotomização e identificação das variáveis de interesse; 2) análise discriminante; e 3) análise com redes neurais. Na primeira fase foram utilizados os testes *t* e Qui-Quadrado para limpar os dados e descartar as variáveis não significativas que, por apresentarem características comuns a ambos os grupos, não contribuiriam para a explicação do fenômeno em estudo. No segundo momento foi realizada a análise discriminante com o pacote estatístico SPSS, e finalmente foi utilizado um algoritmo de redes neurais como técnica alternativa, cujo poder de explicação e previsão é comparado à análise discriminante convencional.

O trabalho utiliza um conjunto de dados, com variáveis contínuas e categóricas, tratadas como variáveis independentes; e uma variável dicotômica que representa uma atitude em relação ao consumo de cerveja, odiar ou gostar de, assumida como variável dependente. A fim de facilitar a inclusão nos modelos multivariados, as variáveis independentes categóricas, que apresentaram significância estatística entre os grupos, foram convertidas em variáveis binárias do tipo ausência-presença de determinada característica (*dummies*). A transformação de variáveis categóricas em variáveis do tipo *dummy* e sua inclusão em modelos de regressão logística foi preconizada por Hosmer e Lemeshow (2000), os quais defenderam que “se algumas variáveis independentes forem discretas, variáveis nominais como sexo, grupo racial e outras, não são apropriadas para incluir no modelo como se fossem variáveis de intervalo” (p. 32), uma vez que os números utilizados para representar as categorias são meros identificadores, não possuindo nenhum

significado numérico. Neste caso, dizem aqueles autores, a solução é utilizar uma série de *dummies* ou variáveis de desenho, obtidas a partir da transformação das variáveis categóricas em binárias (0 e 1). No caso presente, das 63 variáveis categóricas, 61 foram transformadas em *dummies*. Assim, por exemplo, a variável **arte no corpo**, que permitia o preenchimento de seis opções (*piercing* na orelha, *piercing* na língua etc.), foi convertida em **uso de *piercing* na orelha** (S/N), **uso de *piercing* na língua** (S/N) etc.

A análise discriminante e a regressão logística são as técnicas estatísticas apropriadas, quando a variável dependente é categórica, nominal ou não métrica, e as variáveis independentes são métricas (Hair *et al.*, 2005). A análise discriminante envolve determinar uma variável estatística, a combinação linear de duas ou mais variáveis independentes, que discriminarão melhor entre os grupos definidos *a priori*. A discriminação é conseguida estabelecendo-se os pesos da variável estatística para cada variável, para maximizar a variância entre os grupos relativa à variância dentro dos grupos. A função discriminante normalmente assume a forma:

$$Z_{jk} = a + W_1 X_{1k} + W_2 X_{2k} + \dots + W_n X_{nk}$$

onde

Z_{jk} = escore Z discriminante da função discriminante j para o objeto k .

a = intercepto vertical.

W_i = peso discriminante para a variável i .

X_{ik} = variável independente i para o objeto k .

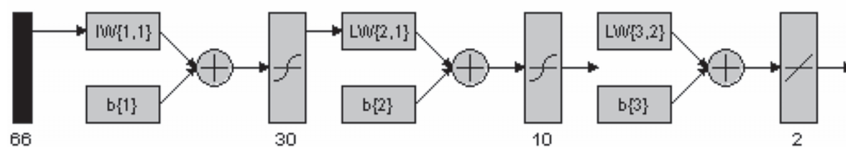
Redes neurais são uma das ferramentas mais adequadas para serem associadas com a mineração de dados (Hair *et al.*, 2005). O modelo é composto de nós, que atuam como entradas, saídas ou processadores intermediários. Cada nó conecta-se ao próximo conjunto de nós por uma série de caminhos ponderados, semelhantes aos pesos em um modelo de regressão (Haykin, 2001). São chamadas neurais ou neuronais, por serem inspiradas na funcionalidade das células nervosas no cérebro. Semelhantes aos humanos, as redes neurais podem aprender a reconhecer padrões pela exposição repetida de exemplos diferentes. Elas podem ser usadas para reconhecer padrões ou salientar características, quer sejam caracteres manuscritos, fisionomias ou outras características pessoais. Assim como os seres humanos têm a capacidade de reconhecer um caráter manuscrito ou a voz de uma pessoa que lhe é familiar, também as redes neurais podem reconhecer padrões em dados que são inexatos e incompletos, como o caso presente.

Similar às redes biológicas, uma RNA pode ser organizada de vários modos diferentes (topologias); isto é, os neurônios podem ser conectados de vários modos. Portanto as RNAs aparecem em muitas configurações. No processamento da informação, muitos dos elementos deste executam seus cálculos ao mesmo tempo. Este processamento paralelo assemelha-se ao modo como o cérebro trabalha, e difere do processamento serial dos cálculos tradicionais. Um algoritmo de aprendizado muito usado é o Algoritmo *Back-propagation* (Dandolini, 1997).

No algoritmo *Back-Propagation*, um padrão de entrada é apresentado na camada de entrada e é propagado através das camadas intermediárias, passando por todos os processos, para produzir uma saída. Esta saída é então comparada com a saída objetivo, e o erro é propagado para trás (da última camada para a camada de entrada) através das camadas da rede. O erro propagado é usado para ajustar os pesos das conexões. Este processo de treinamento é então repetido com novo par de entrada, e novo erro é propagado para trás. Este processo é repetido muitas vezes com muitos pares de amostras de padrões, até que o erro seja minimizado. A possibilidade que tem a rede de lidar com dados ruidosos e resolver problemas não linearmente separáveis deriva da capacidade que têm os neurônios das camadas ocultas de ativarem ou não determinadas áreas e assim organizarem-se conforme os padrões oriundos da camada de entrada (Freeman & Skapura, 1991).

Neste estudo, foi utilizada uma rede multicamadas (*Multi-Layer Perceptron* [MLP]), com algoritmo *back-propagation*, cuja estrutura se encontra representada na Figura 1. Ao longo do trabalho foram testadas diversas arquiteturas e configurações; o melhor rendimento, em termos de ajuste e de capacidade de generalização, foi obtido com duas camadas ocultas, com 30 e 10 neurônios, respectivamente, ambas utilizando a tangente hiperbólica como função de transferência, e uma camada de saída com dois neurônios e função linear, já que os dados deviam ser classificados em dois grupos.

Figura 1: Rede Neural MLP *Back-propagation Feedforward*, de 3 Camadas



RESULTADOS

O levantamento inicial continha variáveis numéricas e categóricas, das quais foram aproveitadas 7 variáveis originalmente dicotômicas e geradas outras 59 *dummies*; ao todo foram incluídas 66 variáveis na análise. Entre as variáveis numéricas encontra-se o número de comunidades a que o indivíduo pertence, o número de amigos, o número de recados existentes na sua página, o número de fãs, o número de testemunhos, o seu carisma⁽¹⁾, a exposição que faz da sua intimidade⁽²⁾ e o número de fotos disponíveis. As variáveis categóricas compreendem as características das fotos divulgadas, o tipo de relacionamento atual (casado, solteiro, namorando etc.); se comunica ou não a idade, se divulga ou não endereço de contato, se tem filhos, se fuma, se bebe, se gosta ou não de animais de estimação, com quem mora, os interesses na comunidade virtual, o humor, a etnia, a religião, a orientação sexual, a visão política, o estilo, a aparência física etc.

A análise discriminante foi feita em duas etapas; na primeira delas foram incluídos todos os 787 casos: 394 membros da comunidade “1”, que se dizem apreciadores de cerveja, e 393 da comunidade “2”, que dizem odiar aquela bebida; na segunda fase, a amostra foi “particionada” em duas, para testar o poder preditivo da função discriminante em dados independentes e, dessa forma, possibilitar a comparação com a rede neural.

A primeira fase da análise discriminante, com todos os dados da amostra, não estava interessada no cálculo da variável estatística em si, mas na verificação dos pressupostos de normalidade e de igualdade das matrizes de variância e covariância entre os grupos. O teste de Kolmogorov-Smirnov demonstrou que todas as 66 variáveis apresentam assimetria significativa entre as duas caudas, em níveis inferiores a 0,001, sendo, portanto, normais. A igualdade de matrizes de variância e covariância foi comprovada por meio do teste M de Box, cujo resultado demonstrou a existência de significância estatística entre os grupos, o que neste caso não era desejável. Tal significância, contudo, segundo Hair *et al.* (2005), não impede a continuidade da análise, contanto que se busquem estimativas de variância separadas por variáveis, na fase de estimação da função discriminante. A violação do pressuposto de igualdade entre as matrizes de variância e covariância foi superada por meio de uma segunda análise, desta vez utilizando o método Stepwise pela distância de Mahalanobis, que confirmou existência de 12 variáveis significantes.

Com as 12 variáveis identificadas foi possível calcular a variável estatística, a qual conseguiu classificar corretamente 74,5% com todos os dados incluídos e

72,6% na validação cruzada, quando o dado testado é deixado de parte. Tal resultado, contudo, tem caráter meramente informativo, uma vez que o propósito maior deste trabalho era verificar o poder preditivo do modelo com dados independentes e não na amostra inteira. O processamento relatado até aqui tem apenas o propósito de verificar a qualidade das variáveis, quanto ao atendimento dos pressupostos teóricos da Análise Discriminante.

Após a confirmação da adequação dos dados, a amostra foi “particionada” aleatoriamente em duas: uma para estimação da variável estatística da análise discriminante e treinamento da rede neural (669 casos), e outra (118 casos) para validação do poder preditivo de ambos. O novo arranjo amostral, após a partição, encontra-se na Tabela 1.

Tabela 1: Composição das Amostras após a Partição

Comunidades	Amostra de estimação/treino		Amostra de validação	
	Frequência	Percentual	Frequência	Percentual
Eu amo cerveja	332	49,6	62	52,5
Eu odeio cerveja	337	50,4	56	47,5
	669	100,0	118	100,0

A fim de possibilitar a comparação da efetiva capacidade preditiva do modelo com a rede neural, que será apresentada logo a seguir, tanto a estimação da variável estatística, quanto os testes da análise discriminante foram refeitos, novamente utilizando o método Stepwise pela distância de Mahalanobis, desta vez utilizando apenas os 669 dados da amostra de estimação. Com a nova amostra, 13 variáveis foram incluídas no modelo e conseguiram classificar corretamente 75,3% e 74,9%, dos casos, respectivamente com e sem a inclusão do dado na estimativa; o Lambda de Wilks foi de 0,697 e a significância entre os grupos $< 0,000$. Os centróides dos grupos foram estimados em -663 e 553, respectivamente para os grupos 1 (apreciadores de cerveja) e 2 (avessos à cerveja). A variável estatística padronizada corresponde a: $(fuma * -0,369) + (gênero * 0,602) + (Foto_c_outros * -0,174) + (Extrovert_estravag * -0,195) + (Est_natural * -0,198) + (Mora_c_AnimEstim * 0,197) + (Mora_c_pais * 0,413) + (Amig_visit_freq * -,169) + (Relat_cinema * 0,358) + (tatoo_visivel * -0,188) + (piercing_lingua * -0,169) + (Alto_Carisma * 0,289) + (Bxssma_Publicida * -0,161)$.

Assim, considerando o intervalo médio entre os centróides $(-0,10)$ e a equação estimadora da função de separação, obtidos da amostra de estimação, foi feita a classificação dos dados da amostra de validação. A função estimada conseguiu classificar corretamente $(54+25)/118 = 67\%$ do dados totais da amostra

independente, porém com graus muito diferentes de exatidão entre os dois grupos (40,3% no Grupo 1 e 96,4% no Grupo 2), o que limita bastante a sua aplicação. O resultado encontra-se na Tabela 2.

Tabela 2: Resultado da Classificação Feita através da Função Discriminante Estimada

			Gp_previsto		Total
			Eu amo cerveja	Eu odeio cerveja	
Gp_Real	Eu amo cerveja	Count	25	37	62
			40,3%	59,7%	100,0%
	Eu odeio cerveja	Count	2	54	56
			3,6%	96,4%	100,0%
Total		Count	27	91	118

A última etapa do trabalho consistiu na classificação dos dados por meio de um algoritmo de redes neurais, com vistas a comparar o resultado de ambos os modelos. Para tanto foram apanhadas as duas subamostras anteriormente preparadas, sendo a primeira delas (669 respostas) destinada ao treinamento, e a segunda (118 respostas) ao teste da rede.

A modelagem de redes neurais foi feita por meio de uma rede MLP de três camadas, com algoritmo *back-propagation*, rodando sobre o pacote MatLab. Foram utilizados 30 e 10 neurônios, respectivamente, nas duas camadas ocultas, e dois neurônios na camada de saída; as funções de transferência utilizadas foram a tangente hiperbólica, nas duas camadas ocultas, e a linear na camada de saída; o treinamento foi feito em 5.000 épocas.

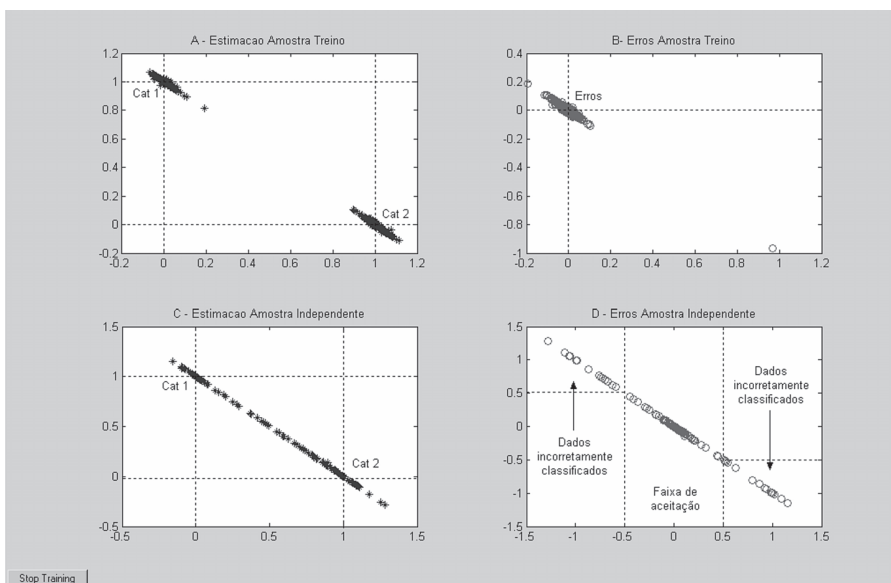
O resultado do processamento feito com a rede neural demonstrou que a ferramenta é capaz de lidar com problemas complexos, como é o caso da produção de informações a partir de dados fragmentados, colhidos de fontes secundárias. Conforme se vê na Tabela 3, a rede neural superou a capacidade preditiva da análise discriminante em dois pontos importantes: (a) ao prever corretamente 71,19% dos dados da amostra independente, contra 67% da análise discriminante; e (b) ao prever corretamente mais de 70% dos dados de ambos os grupos e não apenas de um deles.

Tabela 3: Resultado da Classificação Feita através da Rede MLP

		Grupo previsto		Total
		Eu amo cerveja	Eu odeio cerveja	
Grupo Real	Eu amo cerveja	Count 44 71,0%	18 29,0%	62 100,0%
	Eu odeio cerveja	Count 16 28,6%	40 71,4%	56 100,0%
Total		Count 60	58	118

A análise das saídas da rede neural, por outro lado, demonstra a grande dificuldade de generalizar o modelo para dados externos, uma característica absolutamente previsível, tendo em vista tratar-se de dados colhidos anonimamente, sem qualquer contato com os **respondentes**, o que impede qualquer tipo de controle sobre a qualidade do dado. O resultado, conforme se vê na Figura 2, revela excelente concentração, tanto dos dados (concentrados nos pontos 1;0 e 0;1) como dos erros (concentrados no centro 0;0), na amostra de treino (A e B), o que já não ocorre na amostra de teste (C e D).

Figura 2: Saídas da Rede Neural: os Painéis A e B Correspondem à Simulação dos Dados da Própria Amostra de Treino, e os Painéis C e D às Saídas da Amostra de Validação



Mesmo revelando alguma dificuldade de generalização, ainda assim o resultado é encorajador, na medida em que a previsão realizada com a RNA foi capaz de prever um razoável volume de dados externos, ao longo de todo o espectro de respostas. Conforme pode ser visto nos painéis C e D da Figura 2, os dados estimados concentram-se mais em torno dos pontos (1;0) e (0;1) do que longe deles; e o contrário ocorre com os erros, que se concentram na faixa intermediária (onde não comprometem a previsão) e se tornam rarefeitos nas proximidades dos pontos.

Comparando-se os dois métodos de estimação, verifica-se que, com os dados utilizados, a rede neural apresenta poder discriminante e preditivo superior ao da análise discriminante. Conforme a Tabela 4, mesmo com dados pobres e fragmentados, capturados de fonte secundária de domínio público, ambas as técnicas apresentaram razoável capacidade de previsão, o que corrobora um dos pressupostos do presente trabalho: de que é possível prever atitudes e comportamentos latentes de consumo, a partir da análise e do processamento de outras informações disponíveis, mesmo algumas aparentemente inocentes, como é o caso das informações publicadas voluntariamente nas comunidades virtuais.

Tabela 4: Comparação entre o Desempenho da Análise Discriminante e da Rede Neural

Técnica	Análise discriminante	Rede neural
Número de variáveis incluídas no modelo final	13	66
Número de neurônios e de camadas	-	Entr(66); C1(30); C2(10); C3(2)
Melhor índice de acerto com dados de treino/ajuste	74,9%	99,85%
Melhor índice de acerto com dados de teste/validação (independentes)	67%	71,19%

CONCLUSÕES, LIMITAÇÕES E SUGESTÕES PARA PESQUISAS FUTURAS

Por se tratar de um estudo meramente exploratório, este não tinha o objetivo de modelar equação ou de descobrir fórmula que fosse capaz de prever, com total segurança, qualquer comportamento ou atitude. Dessa forma, os achados, por mais instigantes que sejam, encontram-se limitados pela impossibilidade de se compreender, em toda a sua dimensão, as razões que levam alguém a beber (pouco, moderadamente ou em demasia), ou a rejeitar, de forma peremptória, o consumo de algum tipo de bebida. Tais limitações, ao mesmo tempo, recomendam cautela como os resultados do presente trabalho, encorajam o aprofundamento da investigação e o desenvolvimento de futuras pesquisas, tanto de cunho quantitativo quanto qualitativo.

Ainda que limitações existam, os resultados foram bastante encorajadores, na medida em que, mesmo com variáveis escolhidas praticamente ao acaso, foi possível modelar relações que, no seu conjunto, permitiram prever a existência de uma atitude latente de consumo de boa parte dos membros de uma comunidade virtual. A possibilidade de prever atitudes e comportamentos de consumo, a partir da busca exaustiva e do processamento adequado de informações de domínio público, confere aos profissionais de marketing a possibilidade de prever necessidades e oportunidades de mercado antes da concorrência e assim atendê-las mais prontamente e melhor.

O trabalho realizado demonstra claramente a melhor capacidade de lidar com dados fragmentados e ruidosos e a superioridade do poder preditivo das redes neurais em relação à análise discriminante, conforme suposição inicial. Pelo resultado já exposto, a rede neural apresentou grau de ajustamento superior, tanto na amostra de treino, quanto na amostra independente (dados de teste), o que encoraja a utilização desta última pelos profissionais de marketing, mormente por aqueles envolvidos na pesquisa e prospecção de novas oportunidades de negócio.

Futuros estudos que envolvam outras arquiteturas de rede neural poderão responder se a configuração usada no presente trabalho é mesmo a melhor opção, ou se outras arquiteturas ou mesmo a modificação de alguns parâmetros não são capazes de oferecer ajustes ainda melhores. Da mesma forma, outros dados, como variáveis obtidas a partir da conversão de campos textuais, ou mesmo de outros cadastros (outras comunidades, listas de discussão, *blogs*, *flogs* etc.) poderão ser incluídos no modelo, com vistas a aumentar o ajuste e melhorar o poder explicativo das variáveis independentes.

Ainda que a coleta tenha adotado critérios rigorosos de aleatoriedade, não há nenhuma garantia de que os dados utilizados espelhem a real atitude dos usuários em relação ao consumo de cerveja, tendo em vista que a coleta foi feita diretamente da Internet, sem qualquer contato com os respondentes. Por se tratar de uma divulgação voluntária, é de se supor que as pessoas tornem públicas apenas as características pessoais com que realmente se identificam, não havendo, assim, por que dizer-se adepto do consumo de cerveja ou avesso a ele, caso isso não seja realmente importante para cada um. Entretanto há que considerar o risco de algum usuário aderir a uma determinada comunidade por modismo, comodidade ou influência de outros membros da mesma faixa etária ou círculo de relações, independentemente da sua real atitude em relação ao produto em si. No caso específico de comunidades ditas avessas ao consumo de álcool, existe ainda o risco de alguém aderir por motivo de crença religiosa, por pressão do grupo de referência e porque a bebida lhe faria mal, a ponto de ser odiada ou simplesmente para fazer campanha contra o consumo por considerá-lo prejudicial à saúde.

Tais limitações, contudo, não diminuem a importância do trabalho; pelo contrário, encorajam a realização de trabalhos futuros, com amostras controladas e outras técnicas de processamento, como forma de corroborar ou negar aquilo que acaba de ser relatado.

Artigo recebido em 29.06.2006. Aprovado em 14.02.2007.

NOTAS

¹ Variável numérica resultante da divisão do número de fãs pelo número de amigos.

² Variável numérica resultante da divisão do número de recados existentes na página pelo número de amigos.

REFERÊNCIAS BIBLIOGRÁFICAS

Castells, M. (2003).

A galáxia da internet: reflexões sobre a internet, os negócios e a sociedade. Rio de Janeiro: Jorge Zahar.

Cunha, M. V. M., Jr. (2000, setembro).

Is the internet really changing the markets and the marketing management? A structural change approach. *Anais do Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Administração*, Florianópolis, SC, Brasil, 24.

Cunha, A. L., &

Cunha, M. V. M., Jr. (1999, setembro).

Brand equity and the internet: the impact of the internet presence sites in the development and strenght of brand associations. *Anais do Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Administração*, Foz do Iguaçu, PR, Brasil, 23.

Dandolini, G. A. (1997).

Um procedimento para avaliação da saúde financeira de pequenas empresas: estudo de um caso usando redes neuronais artificiais. Dissertação de mestrado. Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil.

Fonsêca, F. R. B., &

Omaki, E. T. (2004).

Redes neurais artificiais e segmentação psicográfica em marketing: um ensaio sobre a aplicação de RNAs para segmentar os clientes do mercado industrial baseado no risco percebido da compra. *Anais do Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Administração*, Curitiba, PR, Brasil, 28.

- Freeman, J. A., & Skapura, D. M. (1991). *Neural networks: algorithms, applications and programming techniques*. USA: Addison-Wesley.
- Gates, M. (1989). Vals changes with the times. *Incentive*, 163(6), 27-31.
- Gonçalves, F. C., Gonçalves, C. A., Soares, F. C., & Santo, N. C. E. (1998, setembro). Comércio eletrônico na internet: uma pesquisa exploratória no mercado consumidor. *Anais do Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Administração*, Foz do Iguaçu, PR, Brasil, 22.
- Gonzalez, A. M., & Bello, L. (2002). The construct "lifestyle" in market segmentation: the behaviour of tourist consumers. *European Journal of Marketing*, 36(1-2), 51-85.
- Hair, J., Anderson, R., Tatham, R., & Black, W. (2005). *Análise multivariada de dados*. São Paulo: Bookman.
- Haykin, S. (2001). *Redes neurais*. Porto Alegre: Bookman.
- Hosmer, D. W., & Lameshow, S. (2000). *Applied logistic regression*. New York: John Wiley & Sons.
- Kahle, L., Beatty, S. E., & Homer, P. (1986). Research in brief alternative measurement approaches to consumer values: the list of values (LOV) and values and life style (VALS). *Journal of Consumer Research*, 13(3), 405-409.
- Kahle, L., & Kennedy, P. (1989). Using the list of values (LOV) to understand consumers. *The Journal of Consumer Marketing*, 6(3), 5-12.
- Kamineni, R. (2005). The next stage of psychographic segmentation: usage of enneagram. *Journal of American Academy of Business*, 6(1), 315-320.
- Koetz, C. I. (2004, novembro). O valor para o consumidor virtual – análise da utilização da internet por meio de websites. *Anais do Encontro de Marketing da Associação Nacional de Pós-Graduação e Pesquisa em Administração*, Porto Alegre, RS, Brasil, 1.
- Kotler, P. (2000). *Administração de marketing* (10a ed.). São Paulo: Prentice Hall.
- Kozinets, R. V. (1997). I want to believe: a netnography of the x-phil's' subculture of consumption. *Advances in Consumer Research*, 24, 470-475.

- Langerak, F., Verhoef, P. C., Verlegh, P. W. J., & Valck, K. (2004). Satisfaction and participation in virtual communities. *Advances in Consumer Research*, 31, 56-57.
- Leão, A. L. M. S., & Mello, S. C. B. de (2001, setembro). Mensurando customer value através do método laddering: uma proposta da aplicação da técnica para o ambiente virtual. *Anais do Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Administração*, Campinas, SP, Brasil, 25.
- Leão, A. L. M. S., & Mello, S. C. B. de (2002, setembro). Conhecendo o valor do cliente virtual: uma análise utilizando a teoria de cadeias meios-fim. *Anais do Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Administração*, Salvador, BA, Brasil, 26.
- Levy, P. (1999). *Cibercultura*. São Paulo: Editora 34.
- Lin, C. F. (2002). Segmenting customer brand preference: demographic or psychographic. *The Journal of Product and Brand Management*, 11(4-5), 249-269.
- Malhotra, N. K. (2001). *Pesquisa em marketing: uma orientação aplicada*. Porto Alegre: Bookman.
- Mattar, Y. (2003). Virtual communities and hip-hop music consumers in Singapore: interplaying global, local and subcultural identities. *Leisure Studies*, 22(4), 283-300.
- Oates, B., Shufeldt, L., & Vaught, B. (1996). A psychographic study of the elderly and retail store attributes. *The Journal of Consumer Marketing*, 13(6), 14.
- Porto, R. B., & Torres, C. V. (2005, setembro). Valores humanos, atributos do produto e variáveis sociodemográficas: Predições e comparações entre preferência e posse de carro. *Anais do Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Administração*, Brasília, DF, Brasil, 29.
- Ramaswamy, V., Chatterjee, R., & Cohen, S. H. (1996). Joint segmentation on distinct interdependent bases with categorical data. *Journal of Marketing Research*, 33(3), 337-350.
- Ramos-e-Silva, J. A., & Oliveira, L. C. V. (1997, setembro). O marketing na internet.BR. Uma avaliação da presença empresarial na world wide web. *Anais do Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Administração*, Rio das Pedras, RJ, Brasil, 21.

- Recuero, R. C. (2006).
Redes sociais na internet. Biblioteca On-Line de Ciências da Comunicação. Recuperado em 07 fevereiro, 2006, de <http://bocc.ubi.pt/pag/recuero-raquel-redes-sociais-na-internet.pdf>
- Rocha, E. P. Q.,
Barros, C., &
Pereira, C. (2005, setembro).
Perspectivas do método etnográfico em marketing: consumo, comunicação e netnografia. *Anais do Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Administração*, Brasília, DF, Brasil, 29.
- Schiffman, L. G., &
Kamuk, L. L. (2000).
Comportamento do consumidor (6a ed.). Rio de Janeiro: LTC.
- Shufeldt, L.,
Oates, B., &
Vaught, B. (1998).
Is lifestyle an important factor in the purchase of OTC drugs by the elderly? *The Journal of Consumer Marketing*, 15(2), 111.
- Solomon, M. R. (2002).
O comportamento do consumidor: comprando, possuindo e sendo. Porto Alegre: Bookman.
- Tajra, S. F. (2002).
Comunidades virtuais: um fenômeno na sociedade do conhecimento. São Paulo: Érica.
- Thompson, A. M., &
Kaminski, P. F. (1993).
Psychographic and lifestyle antecedents of service quality expectations: a segmentation approach. *The Journal of Services Marketing*, 7(4), 53-61.
- Turban, E., &
King, D. (2004).
Comércio eletrônico: estratégia e gestão. São Paulo: Prentice Hall.
- Ward, K. J. (1999).
Cyber-ethnography and the emergence of the virtually new community. *Journal of Information Technology*, 14(1), 95-105.
- Wind, Y. (1978).
Issues and advances in segmentation research. *Journal of Marketing Research*, 15(3), 317-337.
- Yamashita, S. S., &
Gouvêa, M. A. (2004, setembro).
Impactos e implicações da Internet sobre o marketing de relacionamento. *Anais do Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Administração*, Curitiba, PR, Brasil, 28.