Análise de Redes Sociais e Teoria de Grafos como Suporte para *Consumer Insights*

RESUMO
O processo de *insight* pode sofrer viés cognitivo pela ação dos participantes. Crenças, valores e experiências pregressas podem influenciar o processo de análise dos dados para se conseguir um conhecimento que não é simplesmente uma correlação. Uma alternativa para minimizar este viés cognitivo é a utilização de ferramentas de *analytics*. Essas ferramentas permitem trabalhar com um elevado número de variáveis e dados estruturados e não-estruturados a partir de um suporte computacional não permitido em um passado recente. Essa facilidade permite que voz, texto e imagem sejam utilizados como dados para análises, permitindo avançar nos tipos e complexidades de análises. Atualmente mais de 70% dos dados disponíveis para análises são não-estruturados (IDC, 2010; GARTNER, 2014). Este estudo apresenta uma proposta para mitigar o problema do viés cognitivo, a partir de técnicas de *analytics*, em busca de padrões que possam identificar as oportunidades para suportar a elaboração de *insights*. Foram utilizados dados não-estruturados de 26.514 conversas telefônicas realizadas em Dezembro de 2016, provenientes do *contact center* de uma empresa do mercado financeiro. A metodologia analítica consistiu na transcrição de voz para texto e no uso associado de *Social Network Analysis* e da Teoria de Grafos, sem a identificação prévia do cliente ou incorporação de dados de sistemas CRM, propositalmente. Os resultados obtidos permitiram identificar *insights*, a partir dos elementos de maior centralidade identificados nas redes de associação de palavras. Implica-se que o uso de técnicas analíticas em dados não estruturados pode permitir a obtenção de achados com menor influência do viés cognitivo.

Palavras-chave: Análise de Redes Sociais. *Consumer Insights*. Viés cognitivo.

1. INTRODUÇÃO

Não é possível entender o ser humano como uma máquina que maximizará suas decisões (KAHNEMAN, 2012; MURAMATSU, 2008; RITA, 2005), em especial em processos que envolvam análises de dados para se alcançar *insights* (CNUDDE, 2015). Novas profissões estão surgindo, e serão as profissões do século XXI que utilizarão significativamente dados e tecnologia, como os Cientistas de Dados (DAVENPORT, 2012). Este trabalho se propôs conciliar a análise de dados com o objetivo de encontrar padrões nesses dados que suportem *insights* de negócio (BREADING, 2011; ISHIKIRIYAMA, 2015).

O artigo de Guszcza, datado de 2015, é uma importante inspiração para esta pesquisa. Segundo o autor, as técnicas de *Analytics*, sob a perspectiva da Ciência do Comportamento, podem auxiliar a melhorar as tomadas de decisão com o objetivo de influenciar o comportamento restringindo a ocorrência de vieses cognitivos.

Os dados existem, e não apenas os dados estruturados, com características bem definidas, organizadas e disponibilizadas em tabelas e bancos de dados (BAARS e KEMPER, 2008). Existe uma grande variedade de dados não estruturados, tais como os arquivos em formato PDF, as imagens, os arquivos de voz e de textos em *blogs* (BREADING, 2011). De acordo com Francisco (2014), também é possível utilizar mapas, dados em formato HTML (Hyper Text Markup Language) (a linguagem utilizada na construção de páginas na Web) ou XML (eXtensible Markup Language), esta utilizada para compartilhar informações na Web (HOLZSCHLAG, 2001), como insumo na análise de dados. Em 2020, existirão no mundo 50 vezes mais dados do que existiam em 2010, e algo entre 70% e 80% desses dados serão dados não estruturados. Esses serão os dados com que as empresas trabalharão no futuro (HOLZINGER et. al, 2013). Devido ao volume e à variedade de dados não estruturados, é pertinente desenvolver pesquisas que permitam avançar com o conhecimento nessa área.

Isso não elimina o uso dos dados estruturados nos dias atuais. Tais dados podem suportar análises de negócio de forma inovadora, tal qual utilizar o consumo de energia elétrica como indicador socioeconômico (FRANCISCO, 2010) em um exemplo de uma análise geoespacial e estatística espacial.

O desafio é utilizar as técnicas de *Analytics* para sustentar um modelo de *Consumer Insight* que limite o viés cognitivo, o que poderia comprometer o projeto de *insight* direcionando os profissionais para uma conclusão errada (GUSZCZA, 2015).

No processo de tomada de decisão é possível que a análise de um grande número de opções seja inadequada, sendo em si um limitador da racionalidade (SIMON, 1955). Por isso, quando somos expostos a esse cenário podemos tomar decisões erráticas, não seguindo um modelo de decisão único. O sentido da palavra "decisão", formada por de, que em latim significa "parar", e caedere, também do latim, significando "cortar", compõem a acepção "parar de cortar" (GOMES, 2002). O decisor pode apresentar uma dificuldade de "parar de cortar" e deixar fluir o processo decisório quando está exposto a um cenário complexo e de múltiplas variáveis.

Segundo a Teoria do Prospecto (KHANEMAN, 2012) ou no Limite da Racionalidade (SIMON, 1955 e 1956), aumentar o número de opções pode não ser a estratégia mais adequada para sustentar uma decisão melhor. O cérebro otimiza as decisões através de heurísticas, que são atalhos do cérebro para tomar decisões a partir de um repertório de experiências e informações acumuladas ao longo da vida. Porém, nem sempre essas experiências ou informações conduzirão à resposta correta, ou de melhor retorno (KAHNEMAN, 2012).

Em um ambiente corporativo, uma decisão tomada sob a influência de um viés cognitivo pode gerar perda para a empresa, gerando consequências para adversas para empresa. Tal possibilidade cresce quando o processo de decisão ocorrerá em um cenário que envolve uma grande quantidade de dados, variáveis e informações imperfeitas. Um exemplo é a técnica de *Data Mining*, em que se busca encontrar padrões a partir de uma base de dados (CNUDDE, 2015). O processo de *Data Mining* contém em si algum nível de subjetividade, pois a partir de uma base de dados ele buscará padrões, e a escolha dos dados que serão utilizados pode já estar contaminada por um viés cognitivo. Ao encontrar o padrão nos dados, será possível realizar uma análise com o objetivo de alcançar um *insight*, o que implica uma associação de ideias que darão suporte a um novo produto ou serviço que não é diretamente percebido (LAUGHLIN, 2015). No exemplo descrito, ainda que exista uma técnica de análise de dados para suportar o processo, existem várias fases em que os vieses cognitivos podem influenciar o processo de tomada de decisão, seja na escolha da base de dados com que se vai trabalhar ou na associação de ideias, a partir do resultado alcançado com a técnica de *Data Mining*. Esta pesquisa se propõe a apresentar uma alternativa para minimizar o viés cognitivo no processo de encontrar padrões em dados que deem sustentação a um processo de *Insight*, especificamente na área de relacionamento com o consumidor denominado *Consumer Insight*.

Não obstante, não será objeto de análise desta pesquisa elencar de forma exaustiva as possíveis técnicas existentes para realizar associações de ideias em busca de um *insight*. Busca-se, de maneira complementar, indicar um caminho prescritivo para tomadores de decisão em contextos organizacionais por meio do uso de técnicas recentes de análise de dados, eminentemente não estruturados, minimizando o viés cognitivo.

1.1 Problema de Pesquisa

Inicia-se este tópico com a definição da palavra "*insight*" oferecida pelo Dicionário Aulete Digital (2017): "1. Psic. Compreensão ou resolução de um problema, ger. intuitiva, decorrente da repentina percepção mental dos elementos e das relações apropriados à sua solução. 2. P.ext. Revelação súbita que vem à mente de um indivíduo; ESTALO; ILUMINAÇÃO". Quando compomos a ideia de *Insight*, esperamos ter um conhecimento intuitivo para a solução de um problema, que será, por exemplo, criar ou melhorar um produto ou serviço. Neste caso, dados

são importantes, mas não suficientes para alcançarmos um *insight* (FISHER, 2010). Nesse terreno fértil, o viés cognitivo pode influenciar e direcionar o processo de *insight*.
É possível entender que no cotidiano das empresas constantemente tomam-se decisões cujo objetivo é, teoricamente, maximizar o lucro da empresa. Nesse processo, cada profissional é submetido a vieses cognitivos. Essa situação é potencializada em circunstâncias em que o objetivo é criar algo novo, para as quais se busca um *insight*.
As empresas trabalham com os dados disponíveis para suportar suas decisões de negócio. De acordo com Helbing *et al.* (2017), a quantidade de dados produzida anualmente dobra: "em outras palavras em 2016 nós produzimos mais dados do que em toda a história da humanidade até 2015" (HELBING *et al.*, 2017). Esses dados podem representar os dados estruturados que as empresas utilizam, tais como o cadastro de clientes, as transações comerciais, as informações bancárias (FRANCISCO, 2014), ou os dados não estruturados: textos livres, imagens, vídeos e mapas (BREADING, 2011). A grande variedade de dados, a velocidade com que crescem, corresponde ao insumo de uma área denominada *Analytics* (FRANCISCO, 2014), que tem a missão de tornar diversidade nos tipos de dados em algo que possa ser utilizado pelas empresas.
A partir do cenário descrito, o problema de pesquisa consiste em: Como técnicas de *Analytics* podem reduzir a ocorrência de vieses cognitivos nos processos de tomada de decisão para *insights* de negócio, em especial considerando dados não estruturados?

1.2 Objetivo

O objetivo geral desta pesquisa é identificar a possibilidade de usar técnicas de *Analytics* para restringir as oportunidades de ocorrer um viés cognitivo ao longo do processo de tomada de decisão. Os objetivos específicos são:

1. Analisar as técnicas da Análise de Redes Sociais (ou *Social Network Analysis*, SNA) e Teoria de Grafos para dados não estruturados de modo a restringir a ocorrência de viés humano;
2. Buscar demonstrar a possibilidade do uso das técnicas de SNA e Teoria de Grafos que permita encontrar padrões nos dados não estruturados, sem a necessidade do profissional realizar associações de ideias e, por conseguinte, enviesar cognitivamente o processo de *Insight*.

1.3 Justificativa

Os seres humanos não conseguem tomar decisões baseadas em um livre-arbítrio sem influências externas. Camargo (2010) defende quatro pontos sobre o processo decisório humano:

a. A biologia é inerente ao ser humano e, portanto, produz comportamento;
b. O livre-arbítrio não existe da maneira como gostaríamos, no sentido de sermos livres para escolher sem nenhuma influência de processos internos anteriores ao comportamento aparente;
c. O ser humano não é totalmente diferente e superior às criaturas vivas do planeta em relação ao seu comportamento e, muitas vezes, incorre em erros por esquecer-se de que houve uma evolução, ou seja, de onde veio e aspectos e características físicas e comportamentais que carrega;
d. A cultura e a sociedade não são os únicos fatores influenciadores do nosso comportamento pois não somos uma "tábula rasa", temos uma história

Não é possível entender o ser humano como uma máquina que maximizará suas decisões (KAHNEMAN, 2012; MURAMATSU, 2008; RITA, 2005), em especial em processos que envolvam análises de dados para se alcançar *insights* (CNUDDE, 2015). A oportunidade que está sendo apresentada é a de conciliar a análise de dados com o objetivo de encontrar padrões nesses dados que suportem *insights* de negócio (BREADING, 2011; ISHIKIRIYAMA,2015).
O artigo de Guszcza, datado de 2015, é uma importante inspiração para esta pesquisa. Segundo o autor, as técnicas de *Analytics*, sob a perspectiva da Ciência do Comportamento, podem

auxiliar a melhorar as tomadas de decisão com o objetivo de influenciar o comportamento restringindo a ocorrência de vieses cognitivos.
Os dados existem, e não apenas os dados estruturados, com características bem definidas, organizadas e disponibilizadas em tabelas e bancos de dados (BAARS e KEMPER, 2008). Existe uma grande variedade de dados não estruturados, tais como os arquivos em formato PDF, as imagens, os arquivos de voz e de textos em *blogs* (BREADING, 2011). De acordo com Francisco (2014), também é possível utilizar mapas, dados em formato HTML (Hyper Text Markup Language, a linguagem utilizada na construção de páginas na Web) ou XML (eXtensible Markup Language), esta utilizada para compartilhar informações na Web (HOLZSCHLAG, 2001), como insumo na análise de dados. Devido ao volume e à variedade de dados não estruturados, é pertinente desenvolver pesquisas que permitam avançar com o conhecimento nessa área.
O desafio é utilizar as técnicas de *Analytics* para sustentar um modelo de tomada de decisão que limite o viés cognitivo, o que poderia comprometer o projeto de *insight* direcionando os profissionais para uma conclusão errada (GUSZCZA, 2015).

2. REFERENCIAL TEÓRICO

2.1 *Consumer Insights*

O dicionário Merriam-Webster (on-line) define como *buzzword* uma palavra ou frase que se torna popular em determinada época. Em especial, para esta pesquisa, existe uma expressão popular nos dias de hoje denominada *Consumer Insights*. É uma nova *buzzword* que define uma estratégia de pesquisa do consumidor. A literatura no Brasil sobre esse assunto ainda é incipiente.
Laughlin (2014, p. 76) define *Consumer Insight* como "a compreensão não-óbvia (sic) de seus clientes que, se posta em prática, tem o potencial de mudar o seu comportamento para benefício mútuo". Essa definição é poderosa por chamar a atenção para três pontos importantes no entendimento de qualquer estratégia de *Consumer Insights*:

i. Compreensão não óbvia: essa definição é importante porque descarta os modelos em que a busca se dava por compreender algum comportamento a partir de uma análise de dados disponíveis ou mesmo através de pesquisas com objetivo de validar um conceito ou proposição. A definição remete à ideia de buscar algo que não é uma óbvia análise dos dados disponíveis. O próprio autor cita que existe uma confusão do *Consumer Insights* com pesquisas, análises, dados ou *Database Marketing*, o que é, em certa medida, uma redução do termo;
ii. Potencial de mudar o comportamento: essa parte da definição indica que o resultado tem de ser prático. O *Consumer Insight* deve levar a um resultado prático para alterar o comportamento do consumidor por exemplo, ao entender a demanda do consumidor por um novo produto em especial. É pertinente lembrar que o *Consumer Insight* é um processo de mercado; por isso a preocupação conexa com a praticidade do *Insight*. As empresas não podem, a rigor, investir em pesquisas pelo simples desenvolvimento do conhecimento, sendo direcionadas para o lucro;
iii. Para benefício mútuo: Ainda que pareça um pouco utópica a ideia que se busca transmitir é a de existir uma premissa de que o resultado de todo trabalho deve suportar uma mudança vantajosa para as duas partes. Seguindo essa linha de raciocínio, o consumidor tem de ter um ganho com o *Insight*. Não caberia nesse processo, por exemplo, um projeto com um resultado limitado e restrito à redução de custo para empresa se este não for de alguma forma revertido em ganho para o consumidor. Entende-se que reduzir o custo de produção ou de um serviço é relevante para as empresas; porém, isso não é *Consumer Insight*.

Para que o processo de *Consumer Insight* funcione adequadamente, Laughlin defende a necessidade de existirem quatro áreas técnicas de suporte que consistem na Qualidade dos dados, os profissionais envolvidos no processo, Pesquisa do Consumidor e o *Database Marketing*. Todo o processo tem início com os dados que serão utilizados no processo, se os dados forem de baixa qualidade a análise deverá seguir a baixa qualidade dos dados. Os profissionais devem possuir conhecimento na área técnica para suportar as análises e também de negócio, pois além dos dados básicos de pesquisa, deverá saber buscar *insights*. Não basta, neste caso, ter o domínio em estatística ou computação para ter sucesso no processo. Existe um viés de excesso de confiança nos dados. Os profissionais que trabalham com dados em muitos casos têm um excesso de confiança de que os dados podem ter a resposta para as perguntas que são feitas. Entretanto, isso em alguns casos tem se mostrado falho (GUSZCZA, 2015). Para o caso de *Consumer Insights*, são fundamentais as pesquisas que suportem o entendimento do comportamento do consumidor. Laughlin defende que os consumidores são humanos, e nem sempre suas decisões serão racionais e apresentarão um padrão possível de ser identificado a partir de uma base de dados de negócio. A estratégia de aproximar-se do consumidor para entendê-lo enriquece sobremaneira a base de dados a ser utilizada. Nesse ponto, é importante voltar a definição para recordar que o *Consumer Insight* não é uma análise óbvia, não é a simples correlação de dados para encontrar padrões comportamentais. O *Database Marketing* ainda tem muito a contribuir. As empresas continuam utilizando internamente ou contratando serviços de empresas especialistas no serviço de *Database Marketing* e, ao longo desse processo, geram uma base de dados e um histórico de relacionamento importante que podem sustentar análises e correlações não diretamente observáveis.

Sustentado em Laughlin (2015), o processo de *Consumer Insight* deve gerar como resultado um conhecimento que não corresponde a mera correlação de dados. Devido a isso, armadilhas cognitivas podem enviesar a análise. Existem diversos vieses cognitivos estudados em profundidade por psicólogos (Dan Ariely, Daniel Kahneman) e economistas (Richard Thaler). Um que pode ocorrer no processo de *insight* é o viés da ancoragem. A exposição prévia a uma informação leva a considerá-la fortemente na tomada de decisão ou na formulação de estimativas, independentemente de sua relevância para o que é decidido ou estimado (KAHNEMAN, 2012; RITA, 2005 e ARIELY, 2008). Conforme exposto o processo de *Consumer Insight* não é uma simples e direta associação de dados (LAUGHLIN, 2015), por isso, um viés de ancoragem pode acontecer no momento em que os profissionais estão associando ideias em busca de um *insight*.

Para evitar o viés da ancoragem, é possível utilizar alguns procedimentos, como (KAHNEMAN, 2012):

a. Avaliar os dados considerados como referência, verificando se têm fundamento sólido ou se são valores arbitrários, utilizados simplesmente como âncoras;
b. Questionar as premissas, certificando-se de que são relevantes para a tomada de decisão e que não são utilizadas apenas para corrigir uma possível lacuna de informação;
c. Evitar tomar decisões ou chegar a conclusões por impulso e sem informações suficientes, uma vez que, em um cenário de informação imperfeita, seu cérebro utilizará o que estiver mais facilmente à disposição, representando o Sistema 1 explicado por Kahneman (2012).

2.2 Teoria de Redes e Teoria de Grafos

Existe uma confusão taxonômica ao se referir a Teoria de Redes (RECUERO, 2015). A Teoria de Rede é uma técnica que une as Ciências Sociais, Antropologia e a Psicologia Social sob um suporte de uma área da matemática denominada Teoria de Grafos (RECUERO, 2015), que teve o seu início em 1736 na cidade de Königsberg na antiga Prússia com o matemático Leonhard

Euler. Nessa cidade haviam sete pontes separando uma ilha do continente. A questão levantada era se seria possível atravessar as sete pontes, passando apenas uma vez por cada ponte. A solução de Euler para o problema permitiu a criação da Teoria de Grafos, conforme ilustrado na Figura 1.

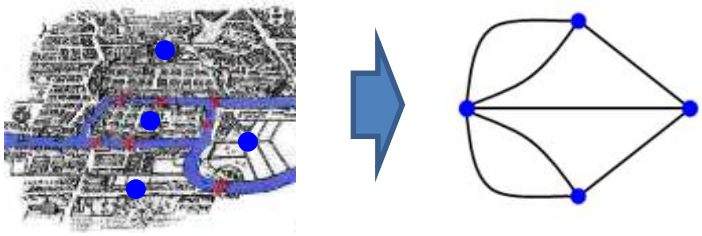


Figura 1. A cidade de Königsberg, e sua representação na forma de um Grafo
Fonte: Adaptado de Universo Racionalista (2016)
Obs: Os nós são as regiões, e as arestas são as pontes.

Na terminologia da Teoria de Grafos e da Teoria de Redes, as origens são os nós, ou mais recentemente denominados também atores, na Análise de Redes (RECUERO, 2015; BARABÁSI, 2009). Os caminhos que ligam os nós são as arestas, que no problema da cidade de Königsberg eram as pontes. Euler demonstrou que para se passar em cada ponte uma única vez seria preciso haver exatamente zero ou dois nós dos quais saíssem um número ímpar de arestas para que isso acontecesse.

A Teoria de Redes, também recentemente denominada *Social Network Analysis*, ou Análise de Redes Sociais, teve a sua fundamentação matemática na Teoria dos Grafos, tendo as duas representações gráficas extremamente semelhantes, incluindo os mesmos elementos, com alterações ocasionais de nomes, como nó e ator. Porém, a Teoria de Redes começou a ser utilizada no início do século XX para explicar a relação entre grupos sociais (RECUERO, 2015). Essa área desenvolveu-se com o objetivo de buscar explicações para relações complexas, tais como tentar explicar quantos graus, nós, separam qualquer pessoa no planeta. Esse experimento foi conduzido pelo psicólogo Stanley Milgram da Universidade de Yale e denominado "Problema do pequeno mundo" (tradução livre do autor para *The small world problem*), que consistiu em enviar cartas para pessoas aleatoriamente com o objetivo de enviarem as cartas a um determinado destinatário. Se elas não conhecessem o destinatário deveriam enviar as cartas a alguém que elas acreditavam ter a maior probabilidade de conhecer o destinatário, com o objetivo de garantir que o este recebesse a carta. Ao final da pesquisa, observou-se que entre primeira pessoa e os destinatários existiam em média seis graus ou nós entre eles. Com o advento das mídias sociais tornou-se mais comum a interação entre indivíduos ou grupos, distantes, que não necessariamente se conhecem diretamente. São atores na rede interconectados virtualmente. Com as mídias sociais, o "Problema do pequeno mundo" de Milgram foi estudado novamente, e agora, admite-se, que existem três graus, em média, que separam qualquer pessoa no mundo (RECUERO, 2015).

A Teoria de Redes continuou a avançar, e segundo Barabási (2009), estudos foram mostrando a significância das relações sociais. Por isso, atores com uma rede mais ampla têm maior facilidade de ampliar a rede do que atores com uma rede com menos nós. Barabási sustentou o pressuposto de que os "ricos tendem a ficar mais ricos, enquanto os pobres tendem a permanecer pobres" (BARABÁSI, 2009). Este conceito é importante na perspectiva de *embeddedness* de Granovetter (2003), no qual é apresentado o impacto de um elemento da rede (nó) na rede na qual faz parte, e vice versa. O efeito e a causa não são mais possíveis de serem desassociados.

A Teoria de Redes é um ótimo modelo para representar sistemas complexos e o fluxo de informações em cada nó (KOLACZYK; CSÁRDI, 2009). Uma forma de representar a relação

entre os nós de uma rede é através de uma rede de adjacência. Essa rede pode ser representada por uma matriz *n* x *n* com 1 e 0, no qual um 1 representa a conexão entre os nós e 0 por sua vez representa a inexistência dessa conexão.

termos	Parceiro A	parcelada	compra	site	comprar	limite	compras	cartao
Parceiro A	0	1	1	1	1	1	1	1
parcelada	1	0	1	0	1	1	1	0
compra	1	1	0	1	1	1	1	1
site	1	0	1	0	1	1	0	1
comprar	1	1	1	1	0	1	1	1
limite	1	1	1	1	1	0	1	0
compras	1	1	1	0	1	1	0	0
cartao	1	0	1	1	1	0	0	0

Figura 2. Exemplo de Matriz *n* x *n* com a relação entre sete palavras da pesquisa
Fonte: Elaborada pelo autores

Na Figura 2, é possível observar a relação entre as palavras. Cada palavra será considerada um nó da rede. Como exemplo, é possível observar que as palavras "Parceiro A" tem relação com a palavra "cartao". Entretanto, a palavra "cartao" não tem relação com a palavra "parcelada".

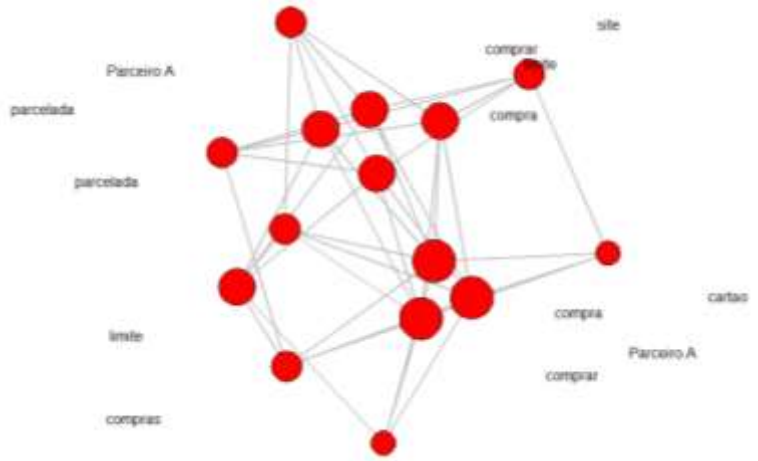


Figura 3: Representação gráfica da rede da matriz da Figura 2
Fonte: Elaboração dos autores

A Figura 3 mostra o exemplo de uma rede com base na Matriz de Adjacência representada na Figura 2. O tamanho dos nós representa a centralidade de grau do nó (ator). A Teoria dos Grafos contribui para algumas medidas que representam a relevância de cada nó (ator) na rede.

2.3 Métricas da Teoria de Grafos

São inúmeras as métricas disponíveis na Teoria de Grafos. Algumas serão relacionadas conforme a conexão com a Teoria de Rede descrita anteriormente.

a. Grau do nó (*degree*)

Descreve o número de conexões de arestas que o nó (ator) possui, podendo ser descrito apenas como grau. As arestas que ligam os nós (atores da rede) podem ser não direcionadas, onde não possuem direção ou direcionada. A direção das arestas informa a direção do fluxo (RECUERO *et al.*, 2015).



Figura 4: Representação dos tipos de arestas
Fonte: Elaboração dos autores

O número de conexões de arestas no nó (ator) permite calcular o grau do nó, que independe do tipo de aresta, se direcionada ou não direcionada.

b. *Closeness*

É a medida que descreve o quão próximo o nó está dos demais nós da rede. É uma medida importante, pois nós com menor *closeness* (proximidade) descrevem pontos onde ocorrerá a difusão com maior velocidade. Desta forma ao avaliar o poder de difusão de informação, calcular quais os nós da rede com menor *closeness* permitirá identificar os nós que merecem maior atenção para acelerar o processo (RECUERO *et al.*, 2015; BARABÁSI, 2009).

$$C = \frac{n-1}{\sum_{j=1}^{n} d(i,j)} \qquad \text{(I)}$$

Na equação (I) temos C como o valor de *closeness*, n como o número de nós da rede, $\sum_{j=1}^{n} d(i,j)$ representa a soma de todas a distâncias para todos os nós e *d(i,j)* é o número de rota mais curtas que liga os nós *i* e *j* (RECUERO *et al*., 2015).

c. *Betweenness*

É o grau de intermediação de um nó na rede. É uma medida do número de rotas mínimas entre nós que passam por um determinado nó da rede. Quanto mais central for o nó, mais rotas mínimas passam por ele. Esta é uma medida importante para demonstrar os nós que possuem a conexão que mantém a integridade de uma parte da rede. É comum que nós de maior *betweenness* se eliminados, sejam responsáveis por isolar outros nós da rede, com o maior impacto, do que os nós com menor *betweenness*. Desta forma quanto maior o grau de *betweenness*, maior o estrago na coesão da rede.

$$b = \sum \theta\, ij\, (v) \div \theta\, ij \qquad \text{(II)}$$

Na equação (II) que calcula o grau de *betweenness b* é o valor do grau, $\theta ij(v)$ é o número de rotas mínimas entre os nós *i* e *j* que passam pelo nó *v* e θij é o número de todas as rotas mínimas entre *i* e *j* (RECUERO *et al.*, 2015).

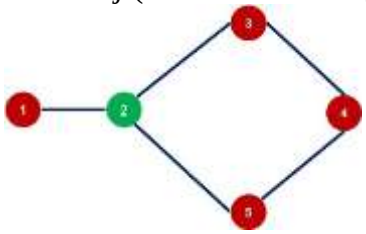


Figura 5: Representação de uma rede
Fonte: Elaboração dos autores

Na Figura 5 é possível verificar o nó com o número 2 se for retirado da rede irá retirar o nó 1 da rede. O nó número 2 é um bom exemplo do impacto da remoção de um nó com elevado *betweenness*.
A descrição das medidas de graus existentes não foi exaustiva. Somente três métricas foram apresentadas com o objetivo de realizar um paralelo com conceitos da Teoria de Redes e propor uma possibilidade de medir o impacto de cada nó numericamente nos conceitos apresentados.
Outra perspectiva para o grau de *betweenness* pode ser na avaliação conceito de *structural holes*. Os nós da rede que podem ser considerados como *structural holes* precisam necessariamente ter um elevado grau de *betweenness*, pois estes nós são como "porteiros" que recebem informações de agrupamentos (*clusters*) que estão conectados através dele (AHUJA, 2000). O conceito de *structural holes* carrega alguma semelhança com o conceito de *Weak Ties* de Granovetter (1973).

d. *Weak Ties* e *Structural Holes*

Granovetter (1973) apresentou o conceito que demonstrou ser importante as relações com os nós mais distantes da rede. Os "*strong ties*" são os mais próximos a nó origem e possuem acesso aos mesmos dados e informações. Granovetter (1973) descreve como os "*strong ties*" como sendo, por exemplo, a família e amigos. Enquanto os "*weak ties*" são os colegas de trabalho e conhecidos. No artigo Granovetter (1973) apresenta que os "*weak ties*" são mais importantes, no artigo, para quem procura emprego, do que os "*strong ties*" por terem acesso a oportunidades de emprego, que os mais próximos na rede não possuem. No caso estudado por Granovetter (1973) as pessoas que procuravam emprego tinham maior probabilidade de sucesso se procurassem colegas distantes, pois estes teriam acesso a oportunidades de emprego diferentes das que o indivíduo que estava procurando emprego tinha. Enquanto que os amigos mais próximos, como os familiares, por estarem dentro de um meio social muito similar, tinham acesso as mesmas oportunidades de emprego que o indivíduo que procurava emprego.
Em uma perspectiva é possível encontrar similaridade com o conceito de *Weak Ties* e *Structural Holes*. Conforme descrito por AHUJA (2000) existe um papel importante para os nós (atores) que conectam grupos (*clusters*) distintos sem outra conexão. Grupos mais distantes em uma estrutura de rede possuem um resultado semelhante o conceito de *Weak Ties* de Granovetter. Possuem acesso a dados e informações diferentes. A perspectiva diferente para os *structural holes* é a importância dos nós (atores) que realizam a conexão entre os grupos mais distantes.

3. METODOLOGIA, PREPARAÇÃO DOS DADOS E RESULTADOS

Optou-se por descrever de forma combinada a metodologia e os resultados subsequentes de forma a tornar este trabalho prescritível em contextos similares de replicação.
Os dados utilizados foram as transcrições das chamadas do grupo de atendimento ao cliente de uma empresa do Mercado Financeiro, doravante referenciada como EMPRESA.
Os dados consistem de uma amostra probabilística de 26.514 ligações realizadas em um período compreendido entre 01 e 31 de dezembro de 2016, obtidas a partir de uma população de 120.410 ligações. O número de chamadas foi restrito à capacidade de transcrição para a entrega do projeto. Não existe limitação tecnológica que impeça trabalhar com 100% das chamadas. De fato, trabalhar com um número maior de chamadas e com um período de tempo maior do que um mês é aconselhável para obter maior base de dados e ter a possibilidade de capturar possíveis eventos sazonais. A opção pelo recorte de dados de apenas um mês deveu-se ao limite imposto por dois fatores: o espaço disponível para gravação e o tempo necessário e esforço computacional para se realizar as transcrições de voz para texto.
As 26.514 chamadas transcritas produziram um texto com 6.856.100 palavras. Este texto contém 39.393 palavras únicas. Cada chamada transcrita tinha em média 259 palavras. Cada chamada foi transcrita em um arquivo separado.

O Corpus composto para essa análise consistiu de 26.514 arquivos. Após o tratamento dos dados, foi composta por 26.514 linhas, representando cada arquivo de transcrição das chamadas, e 943 colunas, que correspondem às palavras resultantes do tratamento dos dados. Através do processo de retirar as palavras definidas dentro das técnicas de *Text Mining* como *stopwords*, que consiste basicamente de pronomes, preposições, artigos definidos e indefinidos e alguns verbos resultou na redução de 39.393 palavras únicas para as 943 que compuseram a matriz de palavras.
A ferramenta utilizada para fazer as análises foi a linguagem de programação R (versão 3.3.2). Os procedimentos utilizados são descritos a seguir, a partir da instalação da ferramenta e da instalação e disponibilização de *packages* para análise de rede (sna) e *text mining* (tm).

```
> install.packages("sna")
> install.packages("tm")
> library(sna)
> library(tm)
```

A fase de tratamento dos dados no *text mining* tem como objetivo reduzir o número de palavras para a quantidade necessária ao processo de análise. Podem existir no texto a ser analisado, além das palavras, números, acentos, símbolos e palavras sem significado. O tratamento dos dados consiste em deixar somente o que é objeto de análise. Este processo é lento e muitas vezes precisa ser reiniciado quando uma ação ocorre, pois, o tratamento realizado anteriormente pode ter perdido a efetividade.

```
text_corpus <- tm_map(text_corpus,rm_accent)
```

Este código elimina as letras com acento pela mesma letra sem acento. Isto é importante porque o pacote tm é sensível a acentos. O objetivo desse comando é eliminar palavras que possam ter sido acentuadas incorretamente na transcrição de voz para texto. Esse processo foi utilizado nesta pesquisa devido ao fato de os dados nela utilizados serem uma transcrição de voz para texto. Para pesquisas em que os dados sejam primariamente texto, ou seja, não tenham sido transcritos a partir de voz, por exemplo, esse comando deve ser utilizado com cuidado no idioma português pois temos palavras em que o acento muda o sentido da palavra.

```
text_corpus <- tm_map(text_corpus,stripWhitespace)
```

Remove os espaços entre as palavras que excedem o normal de um espaço

```
text_corpus <- tm_map(text_corpus,tolower)
```

Transforma todas as letras maiúsculas em minúsculas.

```
text_corpus <- tm_map(text_corpus,removeNumbers)
```

Remove os números em forma de algarismos e deixa apenas números em forma de palavra

```
text_corpus <- tm_map(text_corpus,removePunctuation)
```

Remove todos os tipos de pontuação

```
text_corpus <- tm_map(text_corpus,removeWords,stopwords("portuguese"))
```

Remove as palavras comuns e muito frequentes que não têm significado que possa gerar uma tendência, *insight* ou classificação que diferencie um grupo do outro. Por exemplo, preposições, artigos, contrações e conjunções, entre outros. O pacote tm tem uma lista própria de *stop words*. Ao longo do processo, pode ser necessário o pesquisador ler as palavras que resultaram do trabalho e preparar uma lista específica daquelas que devem ser retiradas, especialmente para a pesquisa em andamento. Esse procedimento foi adotado pelo pesquisador que leu todas a palavras restantes e retirou manualmente as que não tinham significado, devido ao erro na transcrição de voz para texto, ou que seguiam o modelo das *stop words*, sendo comuns e com pouco significado.
Estes são os códigos de tratamento dos dados mais comuns (MEYER *et al.*, 2008) utilizados na pesquisa.
A EMPRESA tem empresas parceiras para as quais fornece o serviço financeiro, oferecendo produtos financeiros. As empresas parceiras e os produtos financeiros oferecidos pelas EMPRESA aos clientes das empresas parceiras serão o objeto de análise.

De um total de 29 empresas parceiras, apenas 6 foram identificadas na base de ligações transcritas.
Para esta pesquisa, doravante as 6 empresas parceiras identificadas serão denominadas simplesmente Parceiro A, Parceiro B, Parceiro C, Parceiro D, Parceiro E e Parceiro F.
A lista de produtos disponibilizada pela EMPRESA é: Empréstimo pessoal, Parcelamento de Fatura, Compra mais segura, Compra mais segura premiada, Renda extra hospitalar, Proteção global, Proteção *Service*, Seguro odontológico, Cartão protegido, Cartão protegido premiado, Suporte sorte, Perda e Roubo, Viva mais segura, Futuro mais seguro, Estudo mais seguro, Ativação da entrega da fatura por e-mail, Ativação da entrega da fatura por Correio, Ativação para avaliação emergencial de crédito, Ativação para comunicação por SMS, Ativação para comunicação por *Telemarketing* e Ativação para comunicação por mala direta.
Nas Tabelas 1 e 2 é possível verificar a lista de empresas parceiras e dos produtos disponibilizados com as suas frequências relativas e absolutas, que foram elaboradas a partir do discurso dos clientes. É possível observar que do total de 26.514 chamadas, em 4.930 os clientes citam ao menos o nome de uma empresa parceira e em 3.935 citam ao menos um produto que a EMPRESA disponibiliza.

Tabela 1. Parceiros e as frequências absolutas e relativas capturadas no processo de análise de Text Mining

Empresas Parceiras	Frequência absoluta	Frequência relativa
Parceiro A	4.178	84, 75 %
Parceiro B	640	12,98 %
Parceiro C	84	1,70 %
Parceiro D	19	0,39 %
Parceiro E	8	0,16 %
Parceiro F	1	0,02 %
Total	**4.930**	

Fonte: Os autores.

Tabela 2. Produtos oferecidos pela EMPRESA aos clientes das empresas parceiras e as frequências absolutas e relativas capturadas no processo de análise de Text Mining

Produtos	Frequência absoluta	Frequência relativa
Parcelamento de Fatura	996	25,31 %
Empréstimo Pessoal	904	22,97 %
Compra mais segura	660	16,77 %
Compra mais segura premiada	431	10,95 %
Cartão protegido	416	6,89 %
Cartão protegido premiado	271	2,36 %
Super sorte	93	2,01 %
Perda e Roubo	79	2,01 %
Proteção global	59	1,50 %
Seguro Odontológico	15	0,38 %
Proteção service	11	0,28 %
Total	**3.935**	

Fonte: Os autores.

Construiu-se a rede de relação entre as palavras referentes ao Parceiro A, responsável pela maior frequência de citações nas chamadas transcritas. Essa rede é apresentada na Figura 7, abaixo.

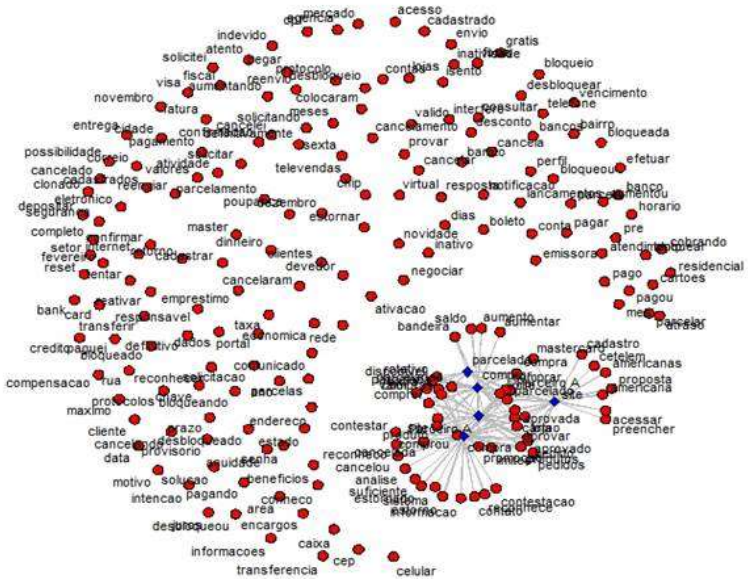


Figura 6. Relação das palavras com relação ao Parceiro A
Fonte: Elaboração dos autores
Observação: Os nós representados pelos losangos azuis são os termos de maior frequência (notadamente, limite, comprar, Parceiro A, compras e pedido)

Destacam-se as palavras com maior centralidade de *closeness*, na rede representada na Figura 6: pedido, limite e compras – com significativa importância no fluxo de informação (RECUERO, 2015; BARABÁSI, 2009).

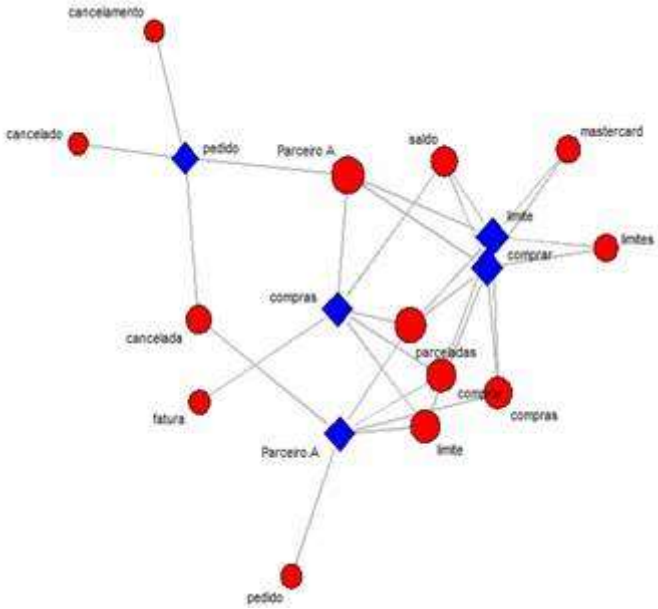


Figura 7. Relação das palavras com relação ao Parceiro A com menor medida de *closeness*
Fonte: Elaboração própria

O código de obtenção da rede é:

```
parceiroa <- read.table("c:/SNA/snaa.csv",header=TRUE,sep=";", dec=",")
gparceiroa <- parceiroa[,2:6]
```

```
rownames(gparceiroa) <- parceiroa[,1]
gplot(gparceiroa)
gplot(gparceiroa,gmode="twomode",displaylabels = TRUE)
gplot(gparceiroa,gmode="twomode",displaylabels = TRUE,
      edge.col="gray",label.cex = 0.7,usearrows=FALSE)
degree(gparceiroa,gmode="twomode",cmode="indegree")
closeness(gparceiroa,gmode="twomode")
betweenness(gparceiroa,gmode="twomode")
```

4. OBTENÇÃO DOS *INSIGHTS* E PRINCIPAIS CONCLUSÕES

A técnica de SNA identificou alguns padrões sem apresentar vieses cognitivos, conforme foi proposto como problema a ser resolvido neste trabalho. Identificaram-se três padrões comportamentais dos clientes, unicamente a partir da análise dos dados. Com o conhecimento desses padrões, foi possível evoluir para *insights* de negócio. Os padrões serão apresentados à empresa como oportunidades para aumentar a receita, reduzir o custo e preservar a marca.

Na oportunidade para aumentar a receita, foi possível identificar a possibilidade de concentrar os esforços iniciais de vendas de serviço em um único parceiro e com um único produto. As chamadas transcritas permitiram concluir que apenas seis empresas parceiras foram citadas nas transcrições. Sem cruzar os dados cadastrais dos clientes não é possível afirmar que somente entraram em contato clientes dessas seis empresas parceiras. Entretanto, uma empresa parceira, Parceira A, respondeu por 85% das citações, conforme Tabela 1. Na rede é possível observar que somente o Parceiro A é um dos termos com menor centralidade *closeness* e aparece conectado com palavras relacionadas a compras ou limite. Foi possível observar que os demais Parceiros não possuem baixa centralidade de *closeness*. Corroborando a importância no relacionamento com o Parceiro A. As palavras associadas ao Parceiro A possuem conexão com geração de receita: compras, limite e pedido.

Quanto à oportunidade de reduzir custos, foi possível observar que o termo "fatura" está conectado a "valor" e "limite", que, por sua vez, estão conectados a "compras". É possível identificar que existe uma dúvida dos clientes sobre o valor da fatura e o limite disponível para compras. Ações para evitar o contato dos clientes, como disponibilizar os valores automaticamente quando o cliente ligar para o *call center* ou enviar uma mensagem de texto para o celular cadastrado poderão reduzir os custos envolvidos no Relacionamento com o Cliente. O contato por telefone (voz) é um dos contatos mais custosos para a empresa (MADRUGA, 2008).

Com relação à preservação da marca, foi importante observar que um número significativo de chamadas se referia a cancelamento de pedidos, com os termos "pedido" e "cancelamento", "cancelado" ou "cancelada". Pode ser uma oportunidade avaliar o porquê de aparecer baixa centralidade de *closeness*, que pode indicar um elevado fluxo de ligações com estes termos.

Foi possível transformar em texto as chamadas de voz feitas ao *call center*. Com o auxílio de técnicas de *Analytics*, explorou-se os dados sem uma pergunta predefinida, tendo como fundamento a técnica de *Data Mining* (TAN, 2009) para explorar dados de modo a extrair um conhecimento e transformá-lo em vantagens de negócio (EREVELLES *et al.*, 2016). A técnica possibilita que essas vantagens não sejam de "compreensão óbvia" (LAUGHLIN, 2014), o que suporta serem *Consumer Insights* fundamentados em *Analytics*, com significativa minimização do viés cognitivo dos profissionais envolvidos.

5. CONTRIBUIÇÃO GERENCIAL

Em certa medida, os modelos quantitativos geram ansiedade nos profissionais. Ao citar SNA espera-se que os profissionais que não sejam especificamente de áreas analíticas tenham uma certa aversão ou sejam até mesmo incrédulos. Esses termos dão a impressão de ficarem

circunscritos à Academia, ou de terem um caráter tão técnico que se transformam em algo esotérico, hermético e destinado a poucos.
A contribuição desta pesquisa é "levantar esse véu" e mostrar que é possível realizar análises robustas com técnicas avançadas, sem a necessidade de uma formação em ciências exatas. Este é um viés cognitivo importante no mercado de trabalho brasileiro. As análises muitas vezes são realizadas sob uma perspectiva simples, sem buscar um aprofundamento.
Esta pesquisa forneceu algumas contribuições importantes:
• *Contact Center* como fonte de dados: muitas empresas têm como canal de relacionamento com o cliente os *Contact Centers*. Como eles recebem uma grande quantidade de chamadas, podem ser uma fonte rica para análises, conforme demonstrou-se neste projeto;
• Desmistificação dos modelos quantitativos: ao final desta pesquisa, o produto final foi de uso prático para as empresas. Pode ocorrer de as empresas terem interesse em usar as ligações dos seus *Contact Centers* para *insights*, ou qualquer empresa que tenha acesso a *blogs*, a *sites* da internet ou aos perfis em redes sociais e não tenha conhecimento de como transformar esses textos livres em informação. Deseja-se que esta pesquisa seja considerada um guia prático, não teórico, de como transformar palavras em informação de negócio. O cerne desta pesquisa foi a técnica de *Consumer Insight*, porém todas as técnicas que foram utilizadas também podem servir para projetos de melhoria de processo, redução de custo, pesquisa de mercado entre outras;
A aplicabilidade desta proposta é primordial para o sucesso da pesquisa aqui desenvolvida: um projeto aplicável e fundamentado em um referencial teórico reconhecido. A ele associaram-se dados não estruturados, representando o maior volume de dados disponíveis hoje nas empresas (GANDOMI, 2014). Aliadas ao projeto e aos dados, estão as técnicas de *Analytics*, há muito estudadas mas que apenas agora, com desenvolvimento tecnológico, estão disponíveis para as empresas (TAN, 2009).

5.1 Recomendação para pesquisas futuras

Seria oportuno estender o uso da técnica de Teoria de Redes além do pacote `tm` do R. Dentro da própria linguagem R, ou na linguagem Python, que tem apresentado um desenvolvimento interessante para algumas técnicas, em especial a Teoria de Redes, o Reconhecimento de Imagens e as Redes Neurais Artificiais, entre outras. O pacote `tm` em si pode ser mais explorado para identificar padrões comportamentais.
Uma pesquisa futura poderia avaliar em especial as técnicas de Análise de Agrupamentos para Teoria de Redes. Como o número de observações é muito grande, com os termos de um texto transcrito de chamadas em um *contact center*, ainda que tenhamos conseguido reduzir de 39.393 palavras únicas para uma matriz de 943 palavras, ainda é um número muito grande para que a técnica de Análise de Agrupamentos seja eficiente.
Replicações desta pesquisa, no contexto de outros negócios, são altamente recomendáveis para que se tenha maior evidência da identificação de *insights* de negócios com minimização do viés cognitivo. Replicações que envolvam relacionamento por voz transcrito para texto, ou mesmo nativamente oriundo de textos, como *blogs*, relatórios setoriais, etc. são altamente desejáveis.

REFERÊNCIAS BIBLIOGRÁFICAS

ARIELY, D. Previsivelmente irracional. [s.l.]: Elsevier, 2008.

AULETE DIGITAL. Disponível em: <http://www.aulete.com.br/>. Acesso em: 10. dez. 2016.

BAARS, H.; KEMPER, H.-G. Management support with structured and unstructured data: an integrated business intelligence framework. Information System Management. Issue 25. p. 132-148. Primavera. 2008.

BREADING, MARK. The Unstructured World. What do insurers do with all those images, email, and tweets? Techdecisions. PropertyCasualty360.com. Outubro, 2011.

CAMARGO,PEDRO. Comportamento do consumidor: a biologia, anatomia e fisiologia do consumo. São Paulo: Novo Conceito, 2010.

CNUDDE, SOFIE, MARTENS, DAVID. Loyal to your city? A data mining analysis of a public service loyalty program. Decision Support Systems, n 73. P. 74–84. 2015.

EREVELLE,S.; FUKAWA, N.; SWAYNE, L. Big Data consumer analytics and the transformation of marketing. Journal of Business Research 69 (2016) 897–904.

FISHER, D.; SMITH, S.. How do people think marketing works? teaching students how to uuild theory, develop imagination, and discover deeper consumer insights. Marketing Education Review, vol. 20, n. 1, pp. 65–70, 2010.

FRANCISCO, E. de R. Indicadores de renda baseados em consumo de energia elétrica: abordagens domiciliar e regional na perspectiva da estatística espacial. São Paulo: FGV, 2010. Tese – Pós Graduação em Administração, Fundação Getulio Vargas, São Paulo, 2010.

_____. Big Data ou Big Problems? Revista ESPM. Ano 20, Edição 95, n. 5, setembro/outubro 2014.

GANDOMI, A., HAIDER,M. Big data concepts, methods and analytics. International Journal of Information Management. 35. Pag 137 – 144. 2014.

GARTNER. <http://www.gartner.com/it-glossary/analytics>. Acesso em: 08 jul. 2015.

GARTNER. Market Guide for Contact Center Speech Analytics. Gartner database, ID G00261569. 07 mar. 2014.

GOMES,HELDER J. C. Text Mining: análise de sentimentos na classificação de notícias. Mestrado em Estatística e Gestão de Informação. Universidade Nova de Lisboa. Novembro de 2012.

GOMES, L. F. A. M.; GOMES, C. F. S.; ALMEIDA, A. T. Tomada de decisão gerencial. São Paulo: Atlas, 2002.

GUSZCZA, JAMES. The last-mile problem. How data science and behavioral science can work together. Deloitte Review. Issue 16. 2015.

HELBING, D.; FREY, B. S.; GIGERENZER,G.; HAFEN, E., HAGNER, M.,

HOLZINGER, A.; STOCKER, C.; OFNER, B.; PROHASKA, G.; BRABENETZ, A.;

HOLZSCHLAG, M. XML, HTML, XHTML magic. [s.l.]: New Riders, 2001

ISHIKIRIYAMA, C. S.; MIROB, D., GOMES, C. F.. Text Mining Business Intelligence: a small sample of what words can say. Information Technology and Quantitative Management. Procedia Computer Science 55. p. 261-267. 2015.

KAHNEMAN, D. Rápido e devagar: duas formas de pensar. Rio de Janeiro: Objetiva, 2012.

KAHNEMAN, D.; TVERSKY, A.. Prospect Theory: An analysis of decision under risk. Econometrica. Volume 47. Issue 2. p. 263-292. Março 1979.

LAUGHLIN, P. Holistic Customer Insight as an engine of growth. Institute of Direct & Digital Marketing (IDM) Journal of Direct, Data & Digital Marketing Practice Vol. 16 Issue 2. 2015

MADRUGA, R. Gestão moderna de call center e telemarketing. 2. ed. São Paulo: Atlas, 2008.

MEYER, D.; HORNIK, K.; FEINERER, I. (2008) Text Mining Infrastructure in R. Journal of Statistical Software, 25 (5). pp. 1-54. ISSN 1548-7660

DICIONÁRIO CALDAS AULETE. Disponível em: <www.aulete.com.br>.

MURAMATSU, ROBERTA, FONSECA, PATRICIA. Um enigma do comportamento do consumidor no brasil: Análise comportamental do consumo procrastinado em um cenário inflacionário. XXXII Encontro da ANAPAD. Rio de Janeiro, setembro 2008.
RITA, VERA. Psicologia Econômica: origens, modelos e propostas. Tese de Doutorado. PUC-SP. 2005
SIMON, H. A., A Behavioral model of rational choice. The Quarterly Journal of Economics. Vol LXIX, p. 99-118. Feb. 1955.
_____. Rational choice and the structure of the environment. Psychological Review. Vol. 63. n. 2. p. 129-138. 1956.
SUTHERLAND, R.. Rory Sutherland knows how to save marketing. Wired. Junho de 2013.
TAN, P.-N.; STEINBACH,M.; KUMAR,V. Introdução ao Data Mining: mineração de dados. Rio de Janeiro: Ciência Moderna,, 2009.
WIRED. Rory Sutherland knows how to save marketing. Disponível em: <http://www.wired.co.uk/magazine/archive/2013/06/feature-rory-sutherland/rory-sutherland-knows-how-to-save-marketing>. Acesso em: 10 jul. 2015.