Volume, velocidade, variedade, veracidade e valor:

como os 5 Vs do big data

estão impactando as organizações e a sociedade

STANLEY LOH

1ª edição

Porto Alegre
Edição do Autor
2019

Prefixo Editorial: 924726
Número de ISBN: 978-85-924726-2-7

Copyright © by Stanley Loh
Todos os direitos reservados.

Formas de citação:

LOH, Stanley. Volume, Velocidade, Variedade, Veracidade e Valor: Como os 5 Vs do Big Data estão impactando as Organizações e a Sociedade. Porto Alegre, 2019.

Loh, S. (2019). **Volume, Velocidade, Variedade, Veracidade e Valor**: Como os 5 Vs do Big Data estão impactando as Organizações e a Sociedade. Porto Alegre, 166p.

RESUMO

A sociedade, as organizações e as pessoas estão cada vez mais gerando, armazenando e usando mais dados. O grande volume de dados é a primeira característica do fenômeno pós-moderno conhecido como Big Data. A quantidade de dados melhora a tomada de decisão mas exige cuidados como análises corretas e foco para evitar prejuízos com a chamada sobrecarga. Este livro discute o impacto do crescimento exponencial, as diferentes formas e tecnologias para coleta de dados, as dificuldades para lidar com tamanho volume e ferramentas que podem nos ajudar a encontrar utilidade nos dados.

A segunda característica do Big Data é a grande velocidade com que dados são trocados entre pessoas, organizações e países, através de tecnologias da informação e comunicação. A velocidade ajuda a tornar as decisões mais rápidas, mas também faz com que eventos de um lado do mundo impactem o outro lado em minutos. Além de discutir o impacto desta velocidade na sociedade, este livro discute também as formas de disseminação da informação e os fatores que influenciam a velocidade de transmissão.

A terceira característica original do Big Data é a variedade dos dados. Hoje podemos coletar e armazenar dados em formatos estruturados ou não estruturados, como textos, imagens e sons. O livro apresenta os diferentes formatos de dados (incluindo grafos, mapas, redes, etc.) e ferramentas para lidar com eles. O livro também discute os benefícios de se ter dados tão variados e as dificuldades que advém desta diversidade.

A quarta característica do Big Data a ser analisada é a veracidade dos dados. Quando podemos confiar numa informação ? O que é uma fonte confiável ? Como confirmar uma informação, como saber se algo é verdade ? São questões discutidas neste livro, além dos males que fake News e informações não confiáveis estão causando na sociedade.

A quinta característica do Big Data é o valor dos dados. De nada adianta ter dados se eles não puderem ser transformados primeiro em informação, depois em conhecimento e por fim servirem para resolver problemas (inteligência e sabedoria). Este livro discute o que é uma informação de valor e como medir tal valor. Também são discutidos impactos que dados valorizados podem causar em relações entre empresas, pessoas e sociedades.

Por fim, o livro aborda a questão da complexidade de dados e como isto está gerando sistemas e organizações mais complexas, mais difíceis de serem entendidas e gerenciadas.

Na introdução, apresentamos os benefícios do Big Data e discutimos o uso de dados com estatísticas em contraposição a intuições.

Sumário

RESUMO 3
INTRODUÇÃO – OS 5Vs 8
Mundo VUCA 9
Dados X Informação X Conhecimento X Inteligência/Sabedoria 10
Dados x Intuições 11
Benefícios do Big Data 13
Aplicações de Big Data 15
Este livro 17
PARTE I – VOLUME DE DADOS 19
Raridade x popularidade 19
Crescimento Exponencial de dados – mundo exponencial 20
Difusão e Propagação de Tecnologias 24
Retornos Acelerados e Vantagens Cumulativas 26
O ponto da virada: de linear para exponencial 28
Só volume não basta - tem que estar perto e associado 29
Sabedoria das massas ou multidões 32
Crowdsourcing 33
O Prosumidor 34
Formas de Coleta de informações: explícita X implícita X por inferência 35
Novas tecnologias para coletar e monitorar dados 36
Enriquecimento de dados – cruzamento de bases 38
Visão 360º - saber tudo sobre as pessoas 40
Identificação de gestos e ações 41
Limitações da memória humana 42
Storytelling – como associar dados para uma visão holística 46
Sobrecarga de informações 47
A Cauda Longa 49
A Atenção 50
Os Superficiais de Nicholas Carr 52
Como encontrar dados em grandes volumes 54
Sistemas de recomendação 56
Volume de dados precisa análise 57
Volume de dados precisa sínteses e resumos 59

Seleção de Amostras 60
PARTE II – VELOCIDADE DOS DADOS 63
Velocidade da Globalização 63
Velocidade das Inovações 64
Atualidade da informação x informação obsoleta 65
Velocidade para Encontrar ou Buscar Informações 66
Velocidade dos Boatos 67
Memes 67
Tecnologias para Velocidade na Comunicação – o Imediatismo 68
Computação ubíqua e pervasiva – Internet das Coisas (IoT) 69
Mais canais de comunicação e interação 73
Velocidade de Cooperação 73
Tipos de redes 74
A força das conexões na rede 75
Imitação e Moda – O Poder do Contexto 76
Os eleitos e influenciadores – o ponto da virada 77
Velocidade na Tomada de Decisões 80
Velocidade de análise de dados 82
Simulações rápidas e simulações sociais 82
Prazer e Recompensas Imediatas 83
Planejamento e desenvolvimento ágil 84
PARTE III – VARIEDADE NOS DADOS 87
O Colapso do Consenso 89
Dados quantitativos 90
Dados qualitativos 92
Análise Qualitativa X Quantitativa 95
Informações Estruturadas x Não Estruturadas 96
Sinais Fracos 98
Bancos de Dados e Planilhas 99
Dados Textuais 100
Dados em Imagens 101
Dados em Sons 104
Imagens X textos X sons 106
Mapas e Informações Geográficas 107

Linguagens gráficas e diagramas 110
Grafos, Redes, Mapas Mentais e Mapas Conceituais 110
Variedade de fontes de dados 114
Integração entre dados variados 115
PARTE IV – VERACIDADE DOS DADOS 116
Análise da informação e suas partes 116
Modelos e padrões 117
Informações imprecisas ou incompletas 118
Informações duplicadas e inconsistentes 119
Provas e confirmação de teorias 119
Amostragem e erro estatístico 121
Credibilidade da fonte 122
Sabedoria das Massas 124
Hipótese nula 124
Quebra de paradigmas e verdades 125
O 4º Paradigma 126
Fatos X Opiniões 127
Viés da conformidade 128
Dificuldade para corrigir dados errados 128
Integridade de Dados 129
PARTE V – VALOR DOS DADOS 131
Quem paga por informação? 132
Custo da Informação 133
Quantidade de informação 134
Utilidade esperada 134
Qualidades da Informação 135
Medidas de Interestingness 135
Capital Intelectual 136
Informação é poder - o capital atual 138
Por que compartilhamos informações de graça ? 139
Análise do Fluxo de Informação nas Organizações 140
Individualistas X Coletivistas 142
Os donos dos dados 142
Privacidade de dados 144

Transparência nos dados 145
PARTE VI – COMPLEXIDADE E BUSCA DE ENTENDIMENTO 147
Complexidade e Cooperação 147
Ordem x caos em sistemas complexos 148
Dificuldades para identificar padrões - pessoas e sistemas complexos 149
Antifrágil para sobreviver no caos e complexidade 152
BIBLIOGRAFIA 153

INTRODUÇÃO – OS 5Vs

Originalmente, o fenômeno Big Data (Tole, 2013) possui 3 características principais conhecidas como 3Vs das informações. São elas:

- volume: cada vez mais a Humanidade gera e armazena mais informações. As tecnologias como rádio, TV e telefone geraram mudanças na comunicação (velocidade e volume de informações). Mas a Internet aumentou tudo isto e ainda permite que pessoas gerem informações que ficam disponíveis publicamente. É a explosão de informações;

- velocidade: hoje em dia é possível escrever um post num blog e divulgá-lo instantaneamente para milhares de pessoas em diferentes partes do mundo;

- variedade: a variedade de informações acontece porque nosso mundo está se tornando mais complexo. Geramos e armazenamos informações em formatos diferentes como textos, diagramas, figuras, sons, planilhas, tabelas, gráficos, etc.

Além dos 3Vs originais, alguns autores ainda incluíram mais 2 Vs, que são Veracidade e Valor. Estes 2 também serão tratados neste livro, porém não deveriam ser considerados como características do Big Data, porque toda informação precisa ter valor e ser verdadeira.

O **volume de dados** aumenta, tanto pela geração quanto pela facilidade em armazenamento e disseminação. Empresas sabem tudo sobre seus clientes. A privacidade acabou. Cada vez teremos mais sensores espalhados para monitorar nosso mundo. São sensores de luminosidade para acender ou pagar lâmpadas nas ruas automaticamente, sensores de temperatura para acionar condicionadores de ar, sensores de chuva e intensidade solar para melhor adaptar nossa agricultura. E tudo isto feito com diferentes máquinas interligadas. É a Internet das Coisas (Internet of Things – IoT).

Para poder entender tantos dados sendo gerados e armazenados vamos precisar de técnicas de análise de dados mais modernas e eficientes, como Data Mining e BI. Além disto, a complexidade do ser humano foi passada para a Tecnologia da Informação. Hoje podemos armazenar dados não estruturados, ou seja, imagens, vídeos, sons e textos.

E some-se a isto tudo a possibilidade de análises mais complexas com o desenvolvimento de softwares com funções de Inteligência Artificial. Se antes, os gestores apenas queriam encontrar endereço de clientes num banco de dados, hoje querem saber qual a faixa de idade que mais compra os produtos de uma certa faixa de preço e isto tudo apresentado por loja, cidade e país. As necessidades das pessoas estão mais complexas.

A **velocidade** de coleta e transmissão de dados poderá gerar decisões em tempo real, antes que eventos negativos aconteçam.

A diversidade é boa. Ela aumenta a criatividade, amplia oportunidades de negócios. A padronização não funciona mais. O que se quer hoje é personalização. Mas a padronização é boa porque facilita nossa vida. Imagina atravessar ruas numa cidade onde em cada

esquina há um tipo diferente de sinaleira ? A diversidade é boa porque mantém individualidades e regionalidades, mas dificulta nosso viver. Se não houver padronização, não conseguimos atravessar uma rua; precisaríamos perder tempo para entender o contexto para tomar decisões simples em cada situação do dia a dia. Ou seja, a Humanidade procura padronização. Foi por isto que inventamos e nos adaptamos bem à agricultura. Para entender a natureza e poder obter alimento de uma forma mais tranquila e regular. Por outro lado, pessoas buscam atendimento personalizado, grupos étnicos buscam espaço e preservação de culturas, minorias querem ser ouvidas e entendidas. Por isso, a **variedade de dados** permitirá representar e entender situações mais complexas.

Entretanto, de nada adianta ter grandes quantidades de dados e não ter qualidade. Por isto, será necessário garantir a **veracidade dos dados** antes de qualquer análise. Novas técnicas para verificação de veracidade poderão dar maior confiabilidade a análises e decisões.

Assim também, dados sem valor apenas ocupam espaço e geram custos. Por isso, o quinto V trata do **valor dos dados**, ou seja, não só de como extrair valor mas também como dar valor quantitativo aos dados.

Mundo VUCA

Alguns autores citam a sigla VUCA para descrever o que está acontecendo no mundo hoje. A sigla se refere a:

- volatilidade (volatility): conceitos, padrões, valores, etc. estão em constante mudança; nada é definitivo; isto acontece na Humanidade há muito tempo, pois passamos de uma terra plana para redonda depois elíptica, de geocentrismo para heliocentrismo, de um universo para multiversos, de leis de Newton para a relatividade de Einstein e agora para a Física Quântica, de criacionismo para evolucionismo, etc.; o problema é que as mudanças estão ocorrendo muito mais rapidamente que tempos atrás; o filósofo Zygmunt Bauman também escreveu sobre isto em livros como Modernidade Líquida, Vida a Crédito e outros tantos. Para ele, os conceitos estão difusos, há mudanças acontecendo nas instituições e tradições.

- incerteza (uncertainty): a consequência da volatilidade é a incerteza; a verdade é verdade apenas em alguns contextos e talvez por algum tempo; Thomas Kuhn ("A Estrutura das Revoluções Científicas") descreve a mudança de paradigmas (conjuntos de verdades) nas Ciências;

- complexidade (complexity): a quantidade de informações, conceitos, elementos, etc. aumenta exponencialmente, bem como as interações entre tudo isto. A consequência é um mundo difícil de entender; até mesmo o clima está mudando e já não conseguimos entendê-lo simplesmente olhando para o céu como faziam nossos ancestrais;

- ambiguidade (ambiguity): a incerteza e a volatilidade geram definições parciais, conceitos difusos, éticas locais, apesar de vivermos num mudo globalizado; ao mesmo tempo que culturas se mesclam, há muitas diferenças que devem ser mantidas e já desigualdades para serem reparadas.

O mundo VUCA tem muita ver com Big Data, podendo ser causa e consequência ao mesmo tempo. Tudo isto será discutido neste livro.

Dados X Informação X Conhecimento X Inteligência/Sabedoria

É importante distinguir dados, informação, conhecimento e acrescentar o conceito de inteligência. A Figura 1 apresenta uma tabela. O valor 35 na 2a linha com a 2a coluna é um dado. Dados são representações de informações. Sozinhos não dizem nada. Quando entendemos que o 35 significa a idade do cliente José, em anos, estamos transformando o dado em informação.

As tecnologias armazenam e lidam com dados. Mas as pessoas trabalham com informações. As pessoas buscam significados naquilo que recebem através dos sentidos. Dados ou sinais que não podem ser entendidos não nos servem. Mas as tecnologias lidam com representações das informações, o que então chamamos dados.

Por outro lado, informações sem utilidade também não servem para nada. Queremos extrair conhecimento das informações. Um exemplo de conhecimento na Figura 1 é: "Todos os clientes da cidade de SP têm saldo médio maior que 9 mil reais". Notem, isto não é uma informação explícita na tabela. Só conseguimos chegar a este conhecimento se cruzarmos informações diferentes. Conhecimento, portanto, vem das informações, mas está acima. As pessoas recebem muitas informações no seu dia a dia, mas nem tudo fica retido, nem tudo é útil, nem tudo será utilizado mais adiante. O que resta, o que é útil, o que é utilizado forma o conhecimento desta pessoa.

Cliente	Idade	Saldo Médio	Cidade
José	35	9000	SP
João	30	4000	Santos
Ana	25	8600	Rio
Maria	23	3000	Ribeirão Preto
Carlos	34	9700	SP

Figura 1: Dados X Informação X Conhecimento/Sabedoria

Já o conceito de Inteligência (alguns chamam Sabedoria) está acima de conhecimento. Imagine um grupo de pessoas numa sala fechada (nada entra ou sai) recebendo uma tarefa: quebrar a cadeira onde estão sentados. Admitamos que todos possuem a mesma força física e foram criados e educados em famílias e escolas semelhantes. Ou seja, possuem o mesmo nível de conhecimento, obtido por estudos nas escolas, leituras em casa, viagens, experiências, etc. Algumas destas pessoas conseguirão resolver o problema e outras não. Mas por que, se todas possuem a mesma força física e os mesmos conhecimentos ? A diferença está na forma como cada um utiliza o conhecimento que tem e as conexões que faz em seu cérebro. Isto é inteligência, ou seja, saber resolver problemas utilizando o conhecimento que possui. E isto se aplica também a poder resolver problemas novos, usando adaptações, analogias, etc.

O grande objetivo então dos dados é ajudar pessoas e organizações a resolverem seus problemas e alcançarem seus objetivos, acabando ou diminuindo o "achismo" ou "empirismo". Houve um caso em que os ouvidores de uma concessionária de rodovias achavam que o trecho mais problemático era um. Quando foram feitas análises estatísticas sobre as ocorrências registradas, descobriu-se que o trecho com mais problemas era outro. Dados e análise estatística refutam opiniões. Dados são representações dos fatos da vida real.

O conhecimento nos faz mais inteligentes; pessoas e empresas que aprendem. Como Kuhn relata, até a metade do século 19 não se usava conhecimento na indústria, somente nas Ciências. Assim como o conhecimento científico mudou o paradigma da Ciência na idade média, o uso intensivo de conhecimento acelerou inovações e permitiu à indústria aproveitar os que as metodologias científicas ensinavam nas ciências, fazendo a prática da indústria menos empírica. E isto tudo fez surgir a Era do Conhecimento, apoiada pelas chamadas tecnologias da informação.

Dados x Intuições

Se não tivermos dados, toda decisão será um jogo de sorte ou azar. Por isto, dados e estatísticas são muito importantes. Por exemplo, no Brasil, o técnico de vôlei Bernardinho e sua equipe têm conseguido grandes resultados para o time nacional de vôlei usando estatísticas. Eles monitoram tudo o que é feito por cada jogador do time do Brasil e também dos adversários. Registram todos os tipos de jogadas, se resultaram em fracasso ou sucesso, como estava a posição dos jogadores, e com isto extraem relatórios de que jogadores estão melhor e quais estão com pior desempenho. Então, quando um brasileiro for "sacar", eles analisam em tempo real as estatísticas e verificam para que adversário deve ser direcionado o saque e de que forma (tipo de saque). E isto é feito para outras estratégias além do saque.

Michael Lewis (2004), no livro Moneyball (que virou filme com Brad Pitt), faz uma grande discussão sobre esta dicotomia entre usar ou não estatísticas. Ele discorre sobre o caso real do Oakland Athletics, time de baseball americano, para expor seus argumentos. A questão toda se desenrola na diferença entre olheiros humanos e sistemas estatísticos para fazer

previsões sobre jovens jogadores. Cada time escolhe os jogadores mais promissores no início da temporada. A grande maioria dos clubes utiliza, até hoje, os olheiros (scouts).

Dados são melhores para apoiar decisões porque evitam o “achômetro” ou “achismo”. Carl Sagan (“O mundo assombrado por demônios”) e Shermer (2011) criticam o crescente uso de pseudociências no lugar da Ciência e do método científico.
(Ver a reportagem “Einstein e Newton estavam errados: estimulada por políticos nacionalistas, 'pseudociência' avança na Índia” https://www.bbc.com/portuguese/geral-46780542)

A falta de dados e de métodos científicos pode levar a grandes erros (ver reportagem sobre jornalista contrária a vacinas que morreu de H1N1).
https://exame.abril.com.br/mundo/apresentadora-antivacinas-morre-de-gripe-suina-aos-26-anos-nos-eua/?fbclid=IwAR2DVfAzAiKeWNa-TMHWIMr25o5H2a_j20UKAu22r5y2AGpMB8H7BSkBmok

Isto não significa que intuições e sentimentos não possam se usados no processo de decisão. Segundo Sinclair e Ashkanasy (2005), intuição é um modo não sequencial de processamento de informações que combina elementos cognitivos e afetivos e resulta em conhecimento direto sem uso de raciocínio consciente. Intuição é diferente de heurísticas, que são estratégias racionais de pouco esforço (segundo Tversky e Kahnemann).

Intuição é um palpite, mas não uma adivinhação. Ela é usada numa decisão sem muita explicação de onde veio, se ela está certa ou não ou por que devemos utilizá-la. É como saber algo sem saber explicar como. O ser humano possui uma certa capacidade para tomar decisões rápidas com pouca informação. A intuição não deve ser confundida com caminho mais fácil (preguiça). Ela deve ser usada quando a racionalidade está limitada. Ela deve ser precedida por dados.

Gunther (2013) acredita que usamos dados do inconsciente, que foram colhidos e armazenados antes, mas que não temos consciência de quando os estamos usando. É como reconhecer um amigo na rua ou a voz de alguém no telefone. Não tem explicação, mas a gente faz e na maioria das vezes não erra. Gunther cita Alfred P. Sloan, ex-executivo da GM: "o ato final da decisão é intuitivo". Isto porque é uma escolha entre alternativas. Ninguém sabe qual a melhor alternativa ou se uma delas vai dar certo ou não. Se soubéssemos, não seria decisão e sim "bola de cristal".

Entretanto, o próprio Gunther (2013) recomenda não confiar na primeira impressão, e sugere que coletemos muitos dados. Kahneman também concorda: é um grande risco tomar decisões usando a área preguiçosa e irracional do cérebro.

Por exemplo, grandes negócios são fechados somente após o encontro presencial entre as partes. Os homens de negócios dizem que é importante "olhar nos olhos". Isto também serve para contratações para empregos. Gladwell (2005), no livro “Blink – a decisão num piscar de olhos”, conta diversos casos onde especialistas tomam decisões baseados em intuições e não conseguem explicar como tomaram as decisões (corretas).

A intuição inclusive já evitou uma guerra nuclear. O militar russo Petrov avaliou intuitivamente os dados que tinha e combinou com avaliações anteriores sobre falhas no sistema, concluindo que não devia apertar o botão para enviar mísseis. (https://pt.wikipedia.org/wiki/Stanislav_Petrov).

Ayres (2008) compara decisões tomadas com estatística x intuitivas. A conclusão de seus estudos é que números são melhores como base para análise, mas a intuição é importante para levantar possibilidades. O ideal é tomar decisão sobre fatos e dados confirmados.

Mas a intuição ajuda a identificar hipóteses e criar alternativas, além de ser útil para avaliar quais variáveis influenciam o processo e também para entender e interpretar os resultados estatísticos. Segundo Silver (2013), meteorologistas melhoram em 25% as previsões de precipitações feitas por computador e em 10% as da temperatura. Neste caso, as informações visuais são melhores interpretadas por seres humanos do que pelo computador. É por isto que

Por outro lado, avaliadores humanos muitas vezes erram porque se preocupam mais com aparências. Então os sistemas baseados em estatísticas podem ser melhores pois não são influenciados por ruídos e variáveis que não implicam em resultados e conseguem se adaptar melhor a pequenas variações nos parâmetros. Por outro lado, humanos vão melhor em alguns casos porque usam uma abordagem híbrida, com uma quantidade maior de informações do que a oferecida apenas pelas estatísticas. E ainda acumulam informações com o passar do tempo (não são sistemas estáticos). Um bom especialista humano também consegue informações privilegiadas, que a maioria não pode obter (por exemplo, no baseball, dados sobre a situação social e familiar do jogador). Se o investidor de bolsa de valores utilizar somente as informações públicas, a que todos têm acesso, não terá nenhuma vantagem. Os investidores precisam encontrar detalhes de informações que os outros não possuem.

Uma boa ideia então é combinar dados estatísticos com intuição, e não somente usar um ou outro. Onde a intuição não é detalhista, os dados podem nos ajudar a lembrar detalhes. Onde a estatística não é completa, a observação humana pode completar uma análise. Não há nada que garanta o resultado, seja utilizando dados estatísticos ou intuições. Mas é melhor para uma decisão ter mais dados (sejam confirmados ou não) e um bom especialista humano para analisar os dados.

Benefícios do Big Data

Como visto antes, a primeira grande vantagem de se ter Big Data é poder coletar e armazenar grandes volumes de dados. Isto permite uma visão mais ampla e completa de uma situação, como por exemplo comportamento de clientes ou movimentação de mercados.

Além disso, é possível coletar dados em tempo real e fazê-los chegar ao destino na velocidade em que eles surgem. Sensores espalhados geograficamente permitem controlar à distância situações, sem deslocamentos físicos. E ações podem ser disparadas também à distância.

Outra vantagem é ter dados mais precisos. Por exemplo, variáveis numéricas podem ser armazenadas e analisadas com mais casas decimais. Além disso, mais variáveis podem ser consideradas em fórmulas e o processamento matemático continua rápido. Dados mais acurados e confiáveis permitem análises melhores, como por exemplo controle e entendimento de comportamentos de mercados.

O Big Data também proporciona dados diversificados. A diversidade é importante para incorporar mais variáveis ao um processo de análise e também para poder analisar um número maior de alternativas ou possibilidades. Isto é importante para quem faz planejamentos ou simulações, para melhor definir onde se quer chegar e como.

Uma consequência disso tudo dito acima é que as análises são feitas com maiores quantidades de dados e com técnicas melhores. Modelos descritivos, como as técnicas de Data Mining e Business Intelligence, permitem entender o passado e encontrar padrões no histórico de dados. Uma vantagem do volume é guardar históricos mais antigos para análise (ex. mudanças climáticas). A análise permite identificar padrões mais exatos e ordem nos dados com amostras maiores. O ser humano é bom em identificar padrões, mas quando há muitos dados, isto gera confusão e conclusões erradas.

Também é possível medir indicadores ou métricas com rapidez em tempo real, permitindo alertas instantâneos. Avaliar métricas permite um melhor controle (se estamos no caminho certo).

O Big Data também facilita o uso de modelos preditivos para antecipar eventos ou calcular valores futuros para variáveis (predição ou previsão). E ainda, Big Data permite elaborar e usar modelos de decisão e apoio à decisão para avaliar e comparar um número maior de alternativas para escolher a melhor. Simulações com mais dados e variáveis permitem avaliar o que pode acontecer num futuro. Funciona como um labirinto de caminhos possíveis, que podem ser testados em segundos. É assim que funcionam alguns algoritmos automatizados para jogar xadrez.

As análises com Big Data facilitam encontrar correlações e causas entre variáveis. Por exemplo, uma empresa descobriu que 5 pontos de aumento na atitude comportamental dos empregados implicava em 1,3 ponto de incremento na satisfação dos clientes, e isto aumenta em 0,5% o faturamento da empresa.

O 2º V do Big Data (velocidade) permite transmitir dados mais rapidamente para que cheguem na hora e no lugar certo, para que pessoas possam tomar decisões melhores e mais rapidamente. E também para que mais pessoas recebam a informação ao mesmo tempo.

Mais dados e dados com melhor qualidade (veracidade e valor) resultam em melhores decisões. A verdade vem dos dados e não de intuições ou suposições. As teorias de Newton, Copérnico e Galileu só puderam ser provadas quando mais dados foram coletados e analisados, após observações e experimentos. Mesmo as hipóteses iniciais das novas teorias só puderam ser elaboradas porque havia dados discordantes com a teoria atual. Esta é a variedade, o 3º V do Big Data.

Outros benefícios do Big Data serão discutidos na próxima seção, sobre aplicações.

Para terminar, vale citar autores que defendem benefícios do Big Data.

Yuri van Geest chama de benefícios 5P do big data: produtividade, prevenção, participação, personalização e previsão (Ismail, Malone e Geest, 2015):

- Produtividade: fazer mais com menos recursos; mais dados para focar nos recursos mais importantes; menos dados pois podemos filtrar as características que realmente impactam uma situação;
- Prevenção: antecipação de riscos e problemas, antes que aconteçam; melhor planejamento de caminhos para alcançar objetivos;
- Participação: mais pessoas ou entidades colaborando, fornecendo dados ou dando sugestões; trocas mais rápidas de dados;
- Personalização: atendimentos diferentes para cada pessoa, de acordo com necessidades específicas e com técnicas mais adequadas;
- Previsão: análises e simulações, antecipando eventos ou calculando com mais precisão valores para variáveis.

Reed (2018) cita os seguintes benefícios para os negócios:

- Transparência:
- Acurácia de dados:
- Precisão em produtos e serviços;
- Decisões melhor orientadas;
- Inovação.

Aplicações de Big Data

Os 5Vs do Big Data estão impactando empresas, famílias, a sociedade, os países, o mundo. A tomada de decisão está mais complexa, usando mais dados e necessitando mais ferramentas e apoios para entendimento disto tudo. É disto que trata este livro.

A primeira grande aplicação de Big Data é na tomada de decisões que envolvam muitos dados (grande quantidade e variedade). Por exemplo, departamentos financeiros e empresas de seguro precisam coletar, armazenar, tratar e analisar dados de clientes e parceiros, além de detalhes de custos e orçamentos e descrições de eventos (ex. inadimplentes e sinistros). Também precisam fazer simulações ou gerar cálculos complexos que acabam gerando mais

dados (volume) e de tipos mais diversos (ex. juros, correções monetárias, atualizações, projeções, etc.).

Devido ao grande volume de dados, outra grande aplicação de Big Data é para previsões climáticas. Sensores espalhados por diversas regiões geográficas poderão coletar dados em tempo real e transmiti-los para centrais, onde estes dados serão analisados também em tempo real. Modelos matemáticos complexos poderão atuar sobre estes dados com maior velocidade, inclusive permitindo simulações. Isto poderá ajudar a antecipar eventos extremos e catástrofes como tsunamis, terremotos e erupções de vulcões.

Campanhas de marketing precisarão entender os diferentes segmentos de clientes. Haverá mais microssegmentação, sendo que produtos e serviços poderão ser gerados para grupos pequenos de clientes (nichos de mercado). A inovação de produtos e serviços ficará cada vez mais especializada, para finalidades e públicos específicos.

Além disso, A variedade de dados permitirá entender por completo as necessidades de clientes, por exemplo a chamada "visão 360 graus" do cliente. Com isso, o atendimento a clientes passará a ser mais personalizado. Jeff Bezos da Amazon disse que precisa ter 3 milhões de lojas diferentes se tiver 3 milhões de clientes.

Também será possível analisar reclamações de clientes feitas, por exemplo em redes sociais ou na Internet, através de textos, emojis ou emoticons. Gravações em grupos de Whatsapp também poderão ser analisadas e cruzadas com anotações textuais e até mesmo com vídeos de clientes nas lojas.

Sistemas de recomendação poderão entender melhor as necessidades e interesses das pessoas e oferecer produtos ou serviços mais adequados, evitando spam ou mesmo evitando sugerir alternativas erradas.

Também deverá ser personalizada a educação e o treinamento de pessoas. O Big Data permitirá levar em conta as diferenças entre alunos e contextos sociogeográficos diferentes. Cada aluno poderá receber conteúdo diferente e de uma forma diferente, de acordo com seu perfil ou modo de estudo.

Nas empresas, departamentos de controle de estoque poderão prever exatamente quando e onde cada produto irá ser vendido. Departamentos de logística poderão fazer simulações melhores para rotas de entrega e até mesmo para escolha de parceiros ou recursos próprios. Em tempo real, clientes poderão saber onde estão os produtos e quanto tempo irá demorar para chegar a entrega. E as empresas que vendem poderão dar uma melhor previsão para a chegada de produtos e calcular fretes mais justos.

Em indústrias, sensores espalhados pelo chão de fábrica ou conectados a máquinas poderão dar o status atual da linha de produção, inclusive indicando onde há possíveis problemas. Com softwares inteligentes, ações automáticas podem ser acionadas ou até mesmo robôs de manutenção.

Em produtos eletrônicos e veículos, sensores poderão coletar dados em tempo real, para auxiliar na detecção de problemas ou causas de mau funcionamento, e até mesmo disparar dispositivos corretores ou de adaptação, como as suspensões dinâmicas em automóveis. Produtos eletrônicos poderão transmitir, em tempo real, dados para centrais que analisam problemas genéricos ou de dispositivos específicos. Já funcionam em computadores, quando ocorre problema. Agora imagine qualquer produto eletrônico coletando dados de funcionamento e acionando centrais inteligentes para melhorar o produto ou corrigir erros *in loco*.

Especialistas humanos em auditoria e segurança poderão analisar melhor as operações e até mesmo evitar fraudes antes que aconteçam. Desta forma também, os departamentos de Inteligência de governos poderão prever atentados terroristas com boa antecedência.

A área de medicina poderá entender melhor como tratamentos e medicamentos atuam sobre pacientes, podendo prever quando pessoas ficam doentes e quando epidemias poderão ocorrer (ou estiverem ocorrendo). Hospitais usarão sistemas com Inteligência Artificial para analisar grandes quantidades e variedades de dados sobre pessoas e situações, além de usar robôs que atuem imitando operações de sucesso que foram usadas no passado por especialistas humanos.

Cidades poderão ser melhor administradas, com coleta seletiva e rápida de lixo e destinação adequada e rápida também. Tráfegos poderão ser melhorados, e os recursos econômicos poderão ser melhor investidos onde realmente façam a diferença.

Governos se valerão de dados para garantir segurança pública, transparência e democracia, com participação ativa de todos os segmentos, inclusão social, melhor distribuição de recursos, menos corrupão e sem fraudes em eleições.

Este livro

Este livro procura investigar os impactos causados pelo volume, pela velocidade e pela variedade de dados que estamos produzindo e recebendo nos dias atuais. Vivemos afogados em dados, incluindo números, estatísticas, opiniões, notícias, verdades e mentiras. As informações correm muito rápidas e nos chegam em formatos diferentes (gráficos, textos, imagens, vídeos, mensagens curtas).

Em parte, os 3 Vs são resultado da nossa sociedade atual. Em parte, são as tecnologias que impulsionam os 3 Vs. Em parte, o mundo e a sociedade líquida, como bem dito por Zygmunt Bauman, estão concretizando os 3 Vs.

Um dos objetivos deste livro é preparar empresas, profissionais e pessoas em geral para viver numa sociedade regida pelos 3 Vs.

O leitor não verá soluções prontas neste livro. Mas uma discussão sólida sobre o que estamos vivendo.

A Parte I fala do volume de dados, que é a primeira característica do fenômeno pós-moderno conhecido como Big Data. A quantidade de dados melhora a tomada de decisão mas exige cuidados como análises corretas e foco para evitar prejuízos com a chamada sobrecarga. Este livro discute o impacto do crescimento exponencial, as diferentes formas e tecnologias para coleta de dados, as dificuldades para lidar com tamanho volume e ferramentas que podem nos ajudar a encontrar utilidade nos dados.

A Parte II trata da velocidade com que dados são trocados entre pessoas, organizações e países, através de tecnologias da informação e comunicação. A velocidade ajuda a tornar as decisões mais rápidas, mas também faz com que eventos de um lado do mundo impactem o outro lado em minutos. Além de discutir o impacto desta velocidade na sociedade, este livro discute também as formas de disseminação da informação e os fatores que influenciam a velocidade de transmissão.

A Parte III apresenta a variedade nos dados. Hoje podemos coletar e armazenar dados em formatos estruturados ou não estruturados, como textos, imagens e sons. O livro apresenta os diferentes formatos de dados (incluindo grafos, mapas, redes, etc.) e ferramentas para lidar com eles. O livro também discute os benefícios de se ter dados tão variados e as dificuldades que advém desta diversidade.

A Parte IV discute a veracidade dos dados. Quando podemos confiar numa informação ? O que é uma fonte confiável ? Como confirmar uma informação, como saber se algo é verdade ? São questões discutidas neste livro, além dos males que fake News e informações não confiáveis estão causando na sociedade.

A Parte V é sobre o valor dos dados. De nada adianta ter dados se eles não puderem ser transformados primeiro em informação, depois em conhecimento e por fim servirem para resolver problemas (inteligência e sabedoria). Este livro discute o que é uma informação de valor e como medir tal valor. Também são discutidos impactos que dados valorizados podem causar em relações entre empresas, pessoas e sociedades.

A última parte, Parte VI, aborda a questão da complexidade de dados e como isto está gerando sistemas e organizações mais complexas, mais difíceis de serem entendidas e gerenciadas.

PARTE I – VOLUME DE DADOS

Estamos vivendo numa era de grandes volumes de informações. O volume de informações já está sendo medido em exabytes. A escala é assim: bit, byte, kylobyte, megaybte, gigabyte, terabyte, petabyte, exabyte, zettabyte, yottabyte.

Chamam isto de Big Data (Tole, 2013), mas anos atrás Korth e Silberschatz (1997) já falavam sobre isto e chamavam esta nova revolução de "explosão de informações". Sim eles comparavam estes novos acontecimentos a revoluções como a invenção da imprensa por Gutenberg (distribuição de informações a todo canto do mundo) e invenção do telefone por Graham Bell (informação distribuída imediatamente, em tempo real).

Há inúmeras estatísticas explicando quanto de informações estamos produzindo e armazenando. Para ter um exemplo, 385 terabytes guardam todo o catálogo da Biblioteca do Congresso americano, a maior do mundo, enquanto que 1,8 zettabyte armazena todos os dados acumulados pela civilização em um ano. Comparando com os 3 exabytes que a humanidade conseguia guardar em 1986, hoje produzimos quase o dobro disto em 2 dias (Revista Veja, ed.2321, n.20, ano 46, maio de 2013). Estamos vivendo em tempos exponenciais como será discutido neste capítulo.

O volume aumenta a cada ano pelas seguintes razões. Primeiro, o armazenamento de dados hoje é barato. Basta ver o preço de discos rígidos e DVDs para backup. Há serviços grátis de hospedagem de dados nas nuvens, como Dropbox, Google Drive, etc, que permitem armazenar documentos e fotos. Segundo, as pessoas estão mais familiarizadas com a tecnologia e consequentemente geram e armazenam mais informações. Crianças de 2 anos já sabem usar celulares e computadores e a 3a idade está menos tecnofóbica. Terceiro, há uma tendência atual de "não jogar nada fora", que começou com o Gmail dizendo que ninguém precisava "deletar' seus e-mails. E por último, há mais possibilidades de serviços para publicar e difundir informações, tais como blogs, twitter, e-mail, redes globais, conexões sem fio, etc. Isto permite que pessoas criem, armazenem e divulguem, sem custos para si, informações. Basta ver as postagens em serviços como Facebook, Twitter, Instagram, etc.

É por isto que alguns países recentemente estabeleceram leis mundiais (como a LGPD) para que o usuário possa solicitar que informações sejam apagadas. É o “Direito ao esquecimento”.

Raridade x popularidade

Antigamente, o valor de um bem era proporcional à sua raridade, por exemplo, o ouro e a prata. Ou seja, se fosse difícil encontrar o bem, seu preço subia. É também o caso de uma

pintura de um artista já falecido. Por outro lado, Kevin Kelly (1998) defende que estamos vivendo uma nova economia, a economia de rede. Nela, quanto mais pessoas usando (mais oferta), mais valor as coisas terão. O teclado Qwerty e outros formatos (VHS sobre Betamax; PC aberto sobre computadores proprietários) vingaram porque as pessoas já estavam acostumadas e não queriam trocar. E as empresas aceitaram um paradigma de menor qualidade porque era mais requisitado.

Em alguns casos, a proporção é mais radical ainda. Um bem só teria valor se um mínimo de pessoas estivesse usando. Por exemplo, o valor do aparelho de fax e do telefone só foi percebido depois que um bom número de pessoas estava usando. Algumas redes sociais só ficam famosas quando já há um grande número de usuários, ou seja, alguém entra só se seus amigos já estão. Mas quem são os primeiros a entrar ? Everett Rogers propôs um modelo para explicar a difusão de inovações, em formato de curva de sino.

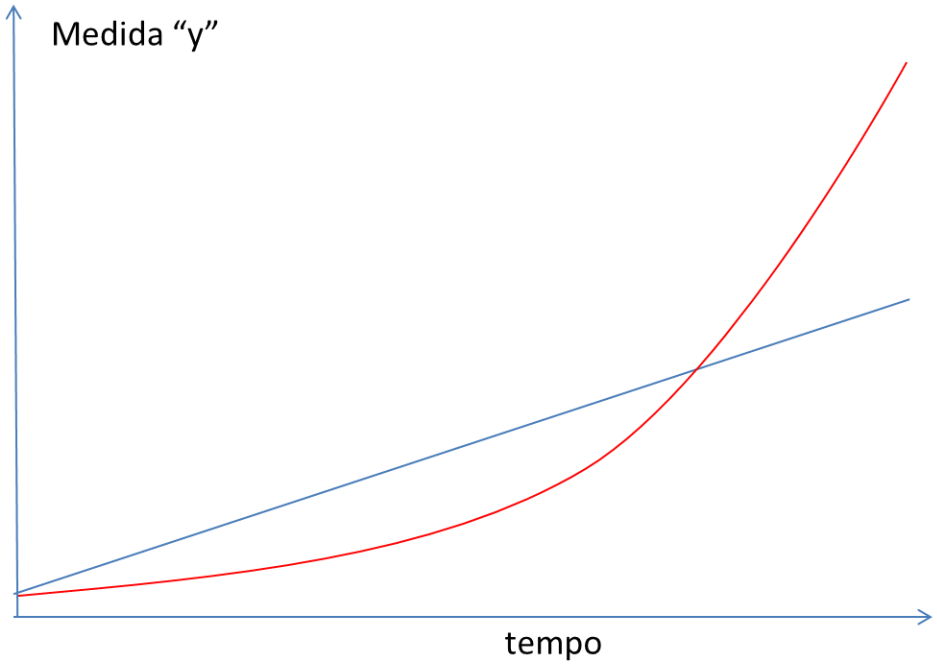
Aplicativos como Waze e redes sociais só têm valor se muitas pessoas aceitarem e usarem. O Youtube só foi vendido por 1,6 bilhão de dólares porque tinha muitos usuários e fiéis. O Google tinha o Google Vídeos com tecnologia melhor, mas bem menos usuários. Um cliente só adere a uma companhia de celular se esta tiver muitas antenas espalhadas. E a companhia só implanta novas antenas se tem perspectiva de haver mais clientes.

Todos querem estar na moda, ninguém que ser muito diferente. Mas também ninguém aceita que todos sejam iguais. Acontece também que a maior participação das massas faz as elites desertarem. Esse é o esgotamento ou imunidade, como definido por Malcolm Gladwell (2013), no livro "O ponto da virada": as pessoas perdem interesse em algumas tecnologias quando há muitos usuários. E aí tendem a trocar por algo mais novo ou inovador. Por exemplo, poucas pessoas atendem ao telefone fixo hoje em sua residência ou mesmo preferem não ter telefone fixo em casa, porque as empresas de telemarketing estão esgotando a paciência de todos.

Robin Dunbar (1992) identificou a regra dos 150: quando a rede ou grupo chega ao número de 150 membros ou elementos, fica difícil coordenar ou participar ou entender as relações, e os relacionamentos enfraquecem. E com isto, o interesse diminui.

Crescimento Exponencial de dados – mundo exponencial

A figura abaixo apresenta uma função linear (em azul) e uma função exponencial (em vermelho). O desempenho das funções muda ao longo do tempo (que aumenta para a direita). Pode-se ver que a função linear (azul) cresce em linha reta com uma certa inclinação fixa. Já a função exponencial cresce mas com inclinações diferentes ao longo do tempo. A função exponencial possui uma inclinação pequena na primeira parte (até menor que a inclinação da função linear). Entretanto, a partir de um certo ponto (quase metade da linha de tempo), a inclinação da função exponencial muda, aumentando consideravelmente. Por esta razão, podemos ver que os valores da medida “y” são maiores na função linear no início da linha de tempo, mas depois os valores da função exponencial ficam bem acima.



A História da Humanidade é contada em linhas exponenciais.

Veja gráficos sobre a evolução da população mundial ao longo do tempo:
http://miningtext.blogspot.com/2014/07/a-evolucao-da-populacao-mundial-desde.html

Veja gráficos sobre ganhos e benefícios para o bem estar humano a partir da Revolução Industrial:
https://www.vox.com/future-perfect/2018/11/8/18052076/human-history-in-one-chart-industrial-revolution

De 2015 a 2025, teremos um salto de 8 bilhões de dispositivos conectados à Internet para 50 bilhões; e em 2035 estima-se que será 1 trilhão (Ismail, Malone e Geest, 2015).

Vejam exemplos comprovados de que vivemos num mundo exponencial:

- população
- expectativa de vida, renda
- tamanho, preço, velocidade e capacidade de processamento dos computadores
- capacidade e custo de armazenamento (storage)
- no Inglês, existem aprox. 540 mil palavras, 5x mais que na época de Shakespeare
- uma semana do jornal The NY Times contém mais informações do que uma pessoa no século 18, durante toda a sua vida
- 4 exabytes de informações novas serão postadas este ano; isto é mais que nos últimos 5 mil anos.

Ridley (2014) apresenta outros números: “Desde 1900, o mundo aumentou sua população em 400%; a área agrícola, em 30%; a produtividade média, em 400%; e a safra total de grãos, em 600%. Então, a produção de comida per capita subiu 50%... hoje as pessoas cultivam (isto é, aram, colhem ou pastoreiam) 38% das terras do planeta, donde se conclui que, com as produtividades de 1961, elas teriam de cultivar 82% para alimentar a população atual. A intensificação salvou 44% do planeta para as florestas.”

Peter Diamandis e Steven Kotler (2012) veem os crescimentos exponenciais como benefícios para a Humanidade.

Watts (2015) nos explica que a sinergia foi um dos princípios que deram sucesso à evolução biológica: "Cada gene interage com outro em uma cadeia complexa de ativação e supressão para expressar traços fenotípicos que não são reduzíveis a propriedades de nenhum gene em particular. Cada planta e cada animal interage um com o outro de formas complexas, por via de relações entre caça e caçador, simbiose, competição e cooperação, para produzir propriedades no nível de ecossistema que não podem ser entendidas em termos de uma espécie em particular. "

Diamandis e Kotler (2012) lembram que a Universidade da singularidade escolheu 8 áreas com crescimento exponencial par serem o núcleo do currículo da universidade: biotecnologia e bioinformática, sistemas computacionais, redes e sensores, inteligência artificial, robótica, fabricação digital, medicina, e nanomateriais e nanotecnologia. “Cada um deles tem o potencial de afetar bilhões de pessoas, resolver grandes desafios e reinventar indústrias. Esses oito campos são tão importantes para nosso potencial de abundância”. O objetivo é reunir diferentes áreas e ao contrário de tratar cada área de forma independente e especializada.

Pentland (2014) lembra que o famoso experimento da DARPA para encontrar 10 balões meteorológicos vermelhos (Galen Pickard, Wei Pan, Iyad Rahwan, Manuel Cebrian, Riley Crane, Anmol Madan, Alex Pentland. Time-Critical Social Mobilization. Science, v.334, n.509, 2011) foi vencido por uma estratégia baseada em pirâmide, onde os incentivos eram distribuídos de forma recursiva, à medida que cada membro do time vencedor ia angariando mais componentes para ajudar. Esta estratégia foi melhor até o altruísmo (o grupo que iria repassar o prêmio para a Cruz Vermelha ficou em 2º lugar). A equipe vencedora encontrou os 10 balões em 8 horas, 52 minutos e 41 segundos.

Matemáticos até já identificaram certos padrões em situações exponenciais e deram nomes. Lei de Potência e Lei de Zipf são exemplos. E estes padrões já puderam ser identificados em diversas situações. Por exemplo, 20% das pessoas disseminam 80% das informações em algumas redes sociais.

Outra famosa séria é a de Fibonacci (o próximo número é a soma dos 2 anteriores): 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144.

Note que em 5 passos é possível multiplicar por 10. Imagine 1 pessoa que avisa 2 outras, e cada uma destas avisa 2 ou 3 e assim por diante. É por isso que as pilhas de arroz

desmoronam quando se acrescenta um grão somente. E por isto também uma gota faz um copo de água transbordar. Talvez uma borboleta no Brasil mude o clima no Japão. Um % duplicado sete vezes é 100%.

A boa notícia é que aumentos exponenciais de bens ou serviços nem sempre precisam de aumento de custos na mesma proporção. Este é o cálculo do custo marginal. Por exemplo, uma sala de espera que suportava 10 pessoas pode suportar 20 pessoas sem que precisemos dobrar o tamanho dela e os móveis.

Veja a Figura 2. Suponha que a linha preta corresponda ao crescimento de clientes de uma empresa ao longo do tempo. Para atender os novos clientes no mesmo padrão dos clientes anteriores, a empresa terá que aumentar seus recursos. Se for um restaurante, este precisará ter mais mesas e cadeiras, mais pratos, mais garçons e comprar mais comida. Mas talvez não precise aumentar o número de cozinheiros. Então, as linhas vermelha, amarela e verde representam as situações possíveis para aumento dos recursos (e custos). O pior caso (vermelho) é quando o aumento de custos supera o aumento de clientes. Neste caso, provavelmente não haverá aumento de lucro (pode até haver prejuízo). O caso amarelo é quando o aumento é proporcional. Isto pode ser preocupante: pouco aumento de lucro mas aumento de riscos (pois há mais recursos envolvidos). O caso ideal (verde) é quando o aumento de clientes pode ser atendido sem grande aumento de custos. O exemplo pode ser adaptado para produção de bens industriais, para serviços, para órgãos públicos, etc.

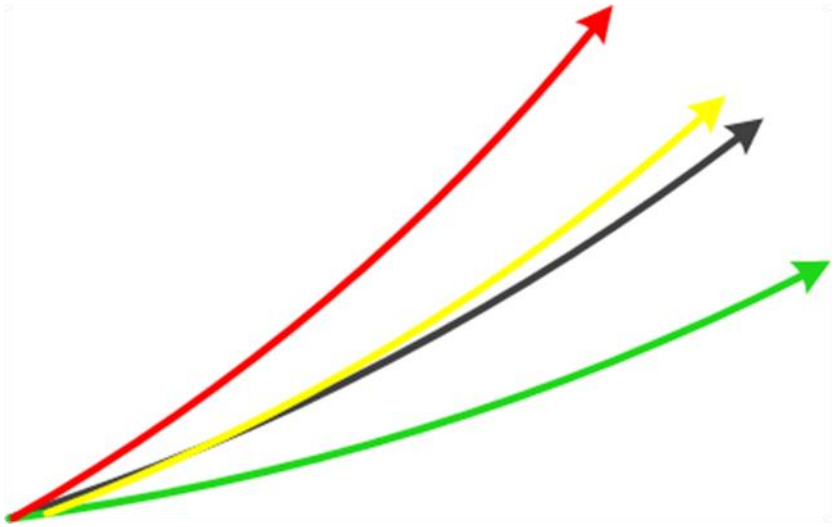


Figura 2: Custo Marginal x escalabilidade

Jeremy Rifkin (2015) escreveu sobre o custo marginal zero e a Economia colaborativa. O objetivo é aumentar o volume de recursos disponíveis sem aumento de custo para a empresa. Isto já foi alcançado por empresas como Uber (sem carros próprios), Facebook (sem produção de conteúdo), AirBnB (sem hospedagens próprias) e Netflix (sem produção de vídeos). Há também empresas de cursos online que podem ter mais alunos sem precisar aumentar na mesma medida seus custos. E tirar 1 foto ou 1 mil não aumenta seu custo.

Os custos estão caindo exponencialmente também. Segundo Ray Kurzweil, num "ambiente baseado em informação, o ritmo de desenvolvimento entra em uma trajetória de crescimento exponencial e a relação preço/ desempenho dobra a cada um ou dois anos." Por exemplo:
Impressão 3D: de US $ 40 mil (2007) para US $ 100 (2014);
Robôs industriais de US $ 500 (2008) para US $ 22 mil (2013).

Difusão e Propagação de Tecnologias

Além de os custos da tecnologia estarem caindo exponencialmente, outro fenômeno que acelera o Big Data é a propagação rápida de tecnologias.

Clark (2007) estudou o tempo (em anos) para algumas tecnologias saírem da Inglaterra (onde foram criadas) e chegarem até algumas regiões do planeta. Veja os dados:

A = Fiação de algodão (1771)
B = Máquina a vapor (1775)
C = Locomotiva a vapor (1825)

	A	B	C
Áustria	30	42	13
Bélgica	28	16	10
Brasil	75	35	29
Canadá	--	36	11
Dinamarca	--	29	19
França	7	3	7
Alemanha	13	8	12
Hungria	--	28	21
Índia	46	30	28
Irlanda	19	15	9
Itália	--	12	14
México	64	43	48
Holanda	24	10	14
Portugal	--	28	31
Rússia	22	23	11
Espanha	--	7	23
Suíça	23	49	22
Suécia	--	23	30
EUA	20	28	5

De maneira semelhante, um artigo da Newsweek, de abril de 1998 ("Reeling in the Years") apresenta o número de anos para que uma tecnologia conseguisse conquistar ¼ da população dos EUA:

Data	Invenção	Número de anos

1873	Eletricidade	46
1876	Telefone	35
1886	Automóvel a gasolina	55
1906	Rádio	22
1926	Televisão	26
1953	Forno de microondas	30
1975	Computado Pessoal	16
1983	Telefone móvel	13
1991	Web	7

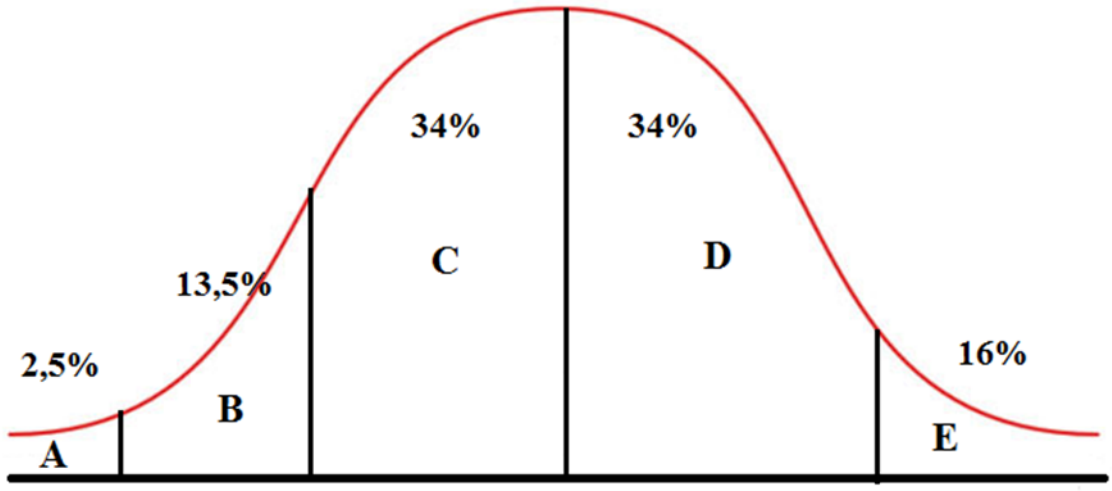
(Fonte: https://www.highbeam.com/doc/1G1-20473137.html)

Vejam quanto anos algumas tecnologias demoraram para alcançar 50 milhões de usuários:

- rádio 38 anos
- TV 13 anos
- Internet 4 anos
- IPOD 3 anos
- Facebook: 2 anos; 100 milhões de usuários em menos de 9 meses

Rogers (2003) propôs uma curva em formato de "sino" para mostrar quantas pessoas utilizam uma inovação ao longo do tempo (ver Figura 3). O grupo identificado como "A" é dos inovadores (innovators). São os primeiros a utilizar uma inovação. Seu perfil é de pessoas com mais conhecimentos (formação), mais capital e são orientados a risco. Pode-se notar que os inovadores utilizam a inovação somente por pouco tempo, pois logo depois vão em busca de outras novidades. Eles funcionam como "testadores" da inovação. O grupo "B" são os chamados "primeiros a adotar" a inovação (early adopters). Eles começam a utilizar a inovação logo após os inovadores a deixarem. Geralmente são jovens, tendem a ser líderes e possuem muito conhecimento (formação. O grupo "C" formam a "maioria inicial" (early majority). São mais conservadores e por isto só começam a utilizar a inovação depois que um bom número de pessoas já a utiliza. São também abertos a novas ideias, ativos na comunidade e influenciam pessoas próximas. O grupo "D" é a "maioria tardia" (late majority). São pessoas de mais idade, menos formação, muito conservadores e menos ativos socialmente. O último grupo, o "E", é dos chamados "retardatários" (laggards). São os últimos que começam a usar uma tecnologia. São extremamente conservadores, possuem pouco capital e bem mais idade, além de menor formação.

Figura 3: Curva de adoção e difusão de inovações – percentual de adotantes em cada divisão



A questão é que este ciclo está acontecendo de forma muito rápida, ou seja, uma nova tecnologia surge para substituir outra quando a primeira está ainda na etapa B. É o fenômeno da "obsolescência" (seja planejada ou não). Schumpeter ficou famoso por cunhar o termo "destruição criativa" (creative destruction), referindo-se às inovações que acabam com as anteriores (as tornam obsoletas). E este seria o motor do capitalismo, segundo o próprio Schumpeter.

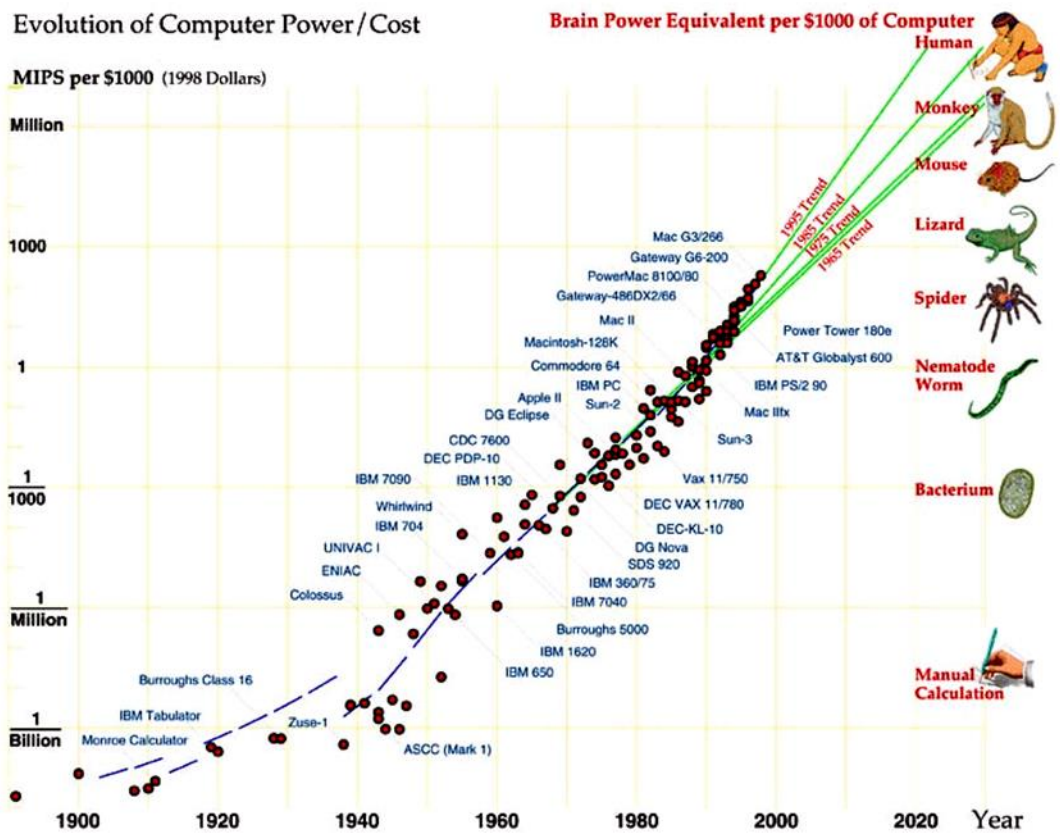
Uma pesquisa da Microsoft constatou que 50% da sua receita vinham de produtos lançados nos últimos 12 meses, que 50 novos produtos eram lançados a cada ano, que atualizações de produtos eram feitas a cada 18 meses e que 2/3 dos componentes de um produto atualizado eram novos.

Segundo Annie Leonard, 99% dos produtos viram lixo após 6 meses. https://www.youtube.com/watch?v=zriUGfQ2V5w
https://www.youtube.com/watch?v=PeGxBh3r6Ek

Retornos Acelerados e Vantagens Cumulativas

Raymond Kurzweil desenvolveu uma teoria que diz que o poder de criação aumenta exponencialmente (não é linear). Isto acontece porque o conhecimento que retorna (feedback) ajuda a selecionar melhores caminhos (se há um grande número de opções). A Figura 4 exemplifica a teoria de Kurzweil no caso da evolução do computador.

Figura 4: A Lei do Retorno Acelerado de Ray Kurzweil



Fonte: site de Ray Kurzweil

Outro exemplo da teoria é no projeto Genoma, para sequenciar o genoma humano. Nas palavras de Kurzweil: “No começo do projeto genoma previ que levaríamos 15 anos para sequenciar todo o genoma humano. Os céticos argumentavam que levaria centenas, milhares de anos. Uns sete anos depois do início do projeto, só 1% estava concluído. Vieram atrás de mim argumentando que, se levou sete anos para fazer 1%, vai levar 700 para completar 100%. Respondi que não. Afinal, quando se chega a 1% da execução de um projeto, cuja velocidade aumenta exponencialmente, você só precisa de sete duplicações na velocidade para chegar aos 100%. Foi o que aconteceu.”

Segundo Diamandis e Kotler (2012), economistas da London School of Business and Finance calcularam que acrescentar 10 telefones por 100 pessoas aumenta em 0,6% o PIB de um país em desenvolvimento. A abundância de informação/comunicação, além de promover a especialização (expandindo as oportunidades educacionais), permite aos especialistas o intercâmbio de especialidades, criando assim o que o economista Friedrich Hayek denominou catalaxia: a possibilidade de expansão ilimitada gerada pela divisão de trabalho.

Outro fator que está aumentando exponencialmente é a produtividade do trabalho. Brynjolfsson e McAfee (2011) calcularam que a produtividade estava aumentando 1% ao

ano, desde 1800. Assim, o índice levava 70 anos para dobrar. A partir de 2010, o índice começou a crescer 4% ao ano.

O ponto da virada: de linear para exponencial

Diversos autores comentam sobre o fenômeno conhecido como "os ricos ficam mais ricos" (rich get richer), entre eles Barabasi (2003) e Easley e Kleinberg (2010). Não estamos falando só de dinheiro.

Segundo Ridley (2014), a disseminação de conhecimento faz com que esse conhecimento útil gere mais conhecimento útil, gerando rendimentos crescentes. Watts (2015) lembra que o sociólogo Robert Merton chamou isto de Efeito Mateus, baseado numa frase do Evangelho de Mateus, da Bíblia: "A todos aqueles que têm mais lhes será dado, e eles terão em abundância; mas para aqueles que nada têm, até mesmo isso lhes será tomado."

Numa rede social, aqueles com mais conexões acabam aumentando suas conexões mais que os que possuem menos conexões. Gladwell (2011), no livro "Outliers", também defende a importância do networking para o sucesso, assim como Max Gunther no livro "O Fator Sorte".

Gladwell (2013), no livro "O Ponto da Virada", apresenta a história do paciente zero de AIDS, o comissário de bordo franco-canadense Gaetan Dugas, que disse ter tido 2.500 parceiras em toda a América do Norte e que estava associado a pelo menos 40 dos primeiros casos dessa síndrome na Califórnia e em Nova York. Esse é o tipo de pessoa que impulsiona as epidemias de doenças.

O ponto da virada inclui pessoas que atraem facilmente outros, seja por carisma, poder, posição social ou outro fator. Gladwell (2013) os chama de eleitos: pessoas que fazem a diferença na disseminação ou propagação de informações, de doenças e outros tipos de influências em rede. São também chamados de Formadores de opinião (opinion leaders) ou Influenciadores. Gladwell descreve 3 tipos de eleitos:
a) os hubs: são os que possuem muitas conexões e ajudam a disseminar rapidamente uma ideia;
b) os experts: são vistos como autoridades num assunto e assim suas ideias penetram melhor;
c) os vendedores: aqueles que sabem persuadir e por isto possuem poder de convencer outros.

Max Gunther (2013) diz que os eleitos aceitam estar em qualquer grupo, iniciam conversas e não recusam convites para encontros e reuniões. Um resumo do livro de Gunther está em http://miningtext.blogspot.com.br/2013/07/5-fatores-que-distinguem-sortudos-de.html

Segundo Pentland (2014), pessoas carismáticas possuem uma habilidade fora do normal para convencer outras a tentarem um comportamento novo. São pessoas que ouvem bastante, possuem retórica fácil e circulam em vários grupos sociais. Contagiam pessoas

com bom humor e atitudes positivas, aumentando a confiança dos outros e encorajando a participação social.

Thompson (2018) lembra o poder da exposição repetida, como um fator para o sucesso. Um dos exemplos que ele usa são as obras de arte. Por que a Monalisa é a obra mais vista? Porque uma tem que ser. Aquela mais vista, continuará sendo a mais vista (rich get richer).

Entretanto, a transposição de linear para exponencial pode ser negativa, ou seja, uma queda brusca. Este é o chamado Efeito Sêneca: quando há um crescimento lento e depois uma queda brusca. Isto pode acontecer com os chamados eventos extremos no clima. E é também o que acontece com o Peru no Natal: durante o ano, o peru tem uma vida tranquila, é bem alimentado e cuidado pela família. Mas isto só dura até o Natal (ou Dia de Ação de Graças).

Taleb (2014), no livro "Antifrágil", classifica as situações ou eventos em 2 tipos:

- Mediocristão: um processo dominado pela mediocridade, com poucos sucessos ou fracassos extremos (digamos, a renda de um dentista). Nenhuma observação particular pode afetar significativamente o conjunto. Também chamado de "cauda fina", ou integrante da família gaussiana de distribuições.
- Extremistão: um processo em que o todo pode ser compreensivelmente impactado por uma única observação (por exemplo, a renda de um escritor). Também chamado de "cauda grossa". Inclui a família de distribuições fractais, ou da lei de potência.

Só volume não basta - tem que estar perto e associado

Uma das razões para o aumento no volume de dados são as grandes concentrações de pessoas em cidades, porque aumentam a interação entre pessoas fomentando a criatividade, a troca de conhecimento e consequentemente gerando mais dados e criações. A criatividade também é acelerada pelo volume de dados em grandes cidades (Diamond, 1997; Weiner, 2016; Johnson, 2011).

Geoffrey West defende que a urbanização traz melhores retornos em economias de escala (infraestrutura). West explica por que as cidades aumentam cada vez mais: quando uma cidade dobra de tamanho, ela só precisa aumentar 15% de sua infraestrutura, e também ocorre um aumento de 15% em sua renda, riqueza e inovação (Diamandis e Kotler, 2012). E segundo as estatísticas, isto acontece em qualquer cidade do mundo, não importando o país, renda, posição geográfica, história, etc. Nos seus estudos, West descobriu diversas características que seguem uma proporção linear ao número de habitantes de uma cidade: número de postos de gasolina, índices de crimes, PIB, número de patentes, etc. Dobrar o tamanho de uma cidade sistematicamente aumenta renda, riqueza, número de patentes, número de colégios, número de pessoas criativas, policiais, taxa de crimes, casos de AIDS e gripe, quantidade de lixo.

Geoffrey West também acredita que grandes concentrações de pessoas aumentam a criatividade humana. Simonton (2002) concorda, ele acha que uma população maior

favorece o nascimento de gênios. Isto acontece, segundo Simonton (2002) por 3 razões: por haver mais papeis diferentes (especialização), por pressão malthusiana sobre a tecnologia (e a necessidade emergente de inovações para não faltar recursos) e por maior variação nos atributos da população (entre eles, o de inteligência).

Conforme Diamandis e Kotler (2012), "dois terços de todo crescimento ocorre nas cidades porque, com sua densidade populacional elevada, os nossos espaços urbanos são perfeitos laboratórios da inovação." Segundo Weiner (2016), isto acontece devido à diversidade, citando o psicólogo Keith Sawyer: "gênios em grupo só acontecem se os cérebros do time não possuem os mesmos ingredientes".

Steven Johnson escreveu um artigo (Anatomy of an idea) demonstrando a importância do contato e cooperação entre pessoas, e não somente por meios digitais mas também pessoalmente. Marissa Meyer, diretora doYahoo e ex-Google, é contra o teletrabalho porque ela acredita que isto diminui a interação e a cooperação entre os funcionários, e portanto diminui a criatividade.
(http://exame2.com.br/mobile/negocios/noticias/marissa-mayer-elimina-o-home-office-no-yahoo;
http://www.computerworld.com.pt/2013/03/04/ceo-do-yahoo-verificou-registos-da-vpn-antes-de-proibir-teletrabalho)

Thompson (2018), em "Hit makers", fala do compositor Savan Kotecha indo pra Suécia porque lá havia mais gente e organizações envolvidas com o ramo da música. Weiner (2016), ao pesquisar a origem geográfica dos gênios, lembra que muitas coisas boas aconteceram em Viena no final do século XIX a partir das cafeterias: "algumas das melhores ideias da cidade ... surgiram inicialmente na atmosfera enfumaçada das cafeterias... Era para lá que as pessoas iam para saber o que estava acontecendo na esquina ou no outro lado do mundo... Além de informações, opiniões também circulavam nas cafeterias.. A genialidade da Viena fin de siècle não estava em nenhuma disciplina específica, mas em uma energia intelectual e artística que se embrenhava por todos os cantos da cidade... Viena é a prova concreta de que a criatividade é contagiosa, que genialidade gera mais genialidade."

Scaruffi lembra que o sucesso do Vale do Silício se deu (e se dá até hoje) porque ali há um ecossistema que reúne:
- empresas: as empresas "conversam" entre si, trocando conhecimento mas também utilizando produtos umas das outras. Há uma simbiose, que ora se manifesta como uma cooperação ora como uma competição, onde uma tenta copiar e fazer melhor algo que outra inventou. As empresas guardam seus segredos industriais, mas olham para o que outras estavam fazendo, e não têm medo de abrir seus projetos para trabalhar em conjunto com outras empresas. A concentração física das empresas também funciona como vitrine para olhares de investidores e facilita atrair mão de obra altamente qualificada;
- Os Hobbistas: são pessoas que tem por hobby trabalhar com eletrônica e tecnologias, usando kits amadores e construindo novos produtos. O mais conhecido deles talvez tenha sido Steve Wozniak que criou junto com Steve Jobs o primeiro computador pessoal. Eles

trocavam conhecimentos em Comunidades e davam suporte. Produzem e consomem computadores, modems, software, etc., além de dar suporte aos novatos;
- Pequenas empresas e spin-offs: várias empresas pequenas com menos de 10 funcionários criaram inovações. Algumas saíram de garagens; outras foram criadas a partir de projetos em Universidades, as chamadas spin-offs. Muitas delas não sobreviveram mas deram origem a novas empresas, ou então, seus engenheiros e empreendedores acabaram em outras empresas do Vale;
- pessoal técnico e capacitado: as empresas do Vale ficaram conhecidas (e são até hoje) por contratarem os melhores dos melhores. Apesar da alta rotatividade de pessoal técnico, a mudança de técnicos de uma empresa para outra gerou troca e aperfeiçoamento de conhecimentos;
- Universidades: elas têm um papel fundamental no Vale. Dali partiram as ideias iniciais das grandes inovações. Muito investimento recebido foi utilizado para criar laboratórios de inovações nas Universidades. Professores são contratados e projetos financiados. Alunos se engajam curiosos no início, mas depois tornam-se engenheiros de computação, eletrônica ou de software. Nas Universidades do Vale, o espírito é bem diferente do quase feudalismo nas universidades da Europa, onde o Professor só fala com os alunos com reuniões marcadas. Durante a Grande Depressão que iniciou em 1929, muitos professores estimularam seus alunos a criarem seus próprios negócios. E este sentimento continua até hoje;
- Capital de risco: muito do dinheiro que chegou no Vale foi pelas mãos de investidores anjos. Ou de empresas de Venture Capital criadas por milionários. Muitos destes se fizeram no Vale mesmo. E daí assumiram a condição de serem responsáveis por retribuir investindo em novos empreendedores, como os mecenas nas artes. Uma espécie de gratidão pelo Vale ter lhes feito milionários. A concentração de empresas e criatividade facilita a atração de investidores;
- Governo: ao contrário do que defendia Adam Smith e corroborando a teoria de Keynes, o Governo foi muito importante para o desenvolvimento do Vale do Silício. Primeiro porque o Governo era o principal cliente durante as duas guerras mundiais e foi também cliente durante a guerra fria. Os departamentos de defesa, espacial, de inteligência, etc, sabiam que só poderiam estar à frente dos inimigos ou competidores se tivessem novas e melhores tecnologias. A guerra tornou grandes as pequenas empresas. Segundo o Governo interviu através de políticas, tais como dando ou evitando monopólios. A AT&T ganhou monopólio e com isto conseguiu desenvolver tecnologias de comunicação que foram importantes para a guerra e acabaram chegando ao mercado. Por outro lado, o PC só se popularizou porque a IBM foi obrigada pelo Governo a abrir a arquitetura do PC, o que permitiu que várias empresas (inclusive pequenas) criassem peças e software para microcomputadores. O Governo também estimulou a cooperação entre universidades e empresas, desempenhando um papel importante ao juntar ambos lados em projetos estratégicos. As inovações geradas para a área militar acabavam voltando para as empresas, gerando novos produtos para o mercado consumidor (por exemplo, o GPS). O Governo também incentivou investimentos baixando o imposto sobre o capital de risco, o que fez com que muitas pessoas e empresas investissem nas startups do Vale. E até mesmo chegou a arcar com uma parte dos investimentos. Durante a Guerra Fria, criou uma lei em 1958 pela qual o governo colocava 3 dólares para cada dólar investido numa startup por uma instituição financeira.

Sabedoria das massas ou multidões

Mais dados podem gerar melhores resultados (mais completos, precisos ou confiáveis). A ideia de que algo feito por várias pessoas tende a ser melhor do que se fosse feito por apenas uma não é nova e virou dito popular. Em “Politics”, Aristóteles defendeu uma forma de democracia referenciada como “sabedoria das massas”: “Cada indivíduo será pior juiz que os especialistas, mas quando todos trabalham juntos, eles são melhores ou no mínimo iguais (com desempenho semelhante)”.

Surowiecki (2004) comenta diversos casos (não comprovados cientificamente) onde a coletividade suplanta os resultados de indivíduos. Há o caso de um experimento simples durante uma palestra, em que a plateia consegue acertar as medidas de uma pessoa. Cada pessoa sugere valores com base em sua intuição ou experiência, e a média das suposições gera erro zero. Surowiecki avaliou diversas situações num jogo de auditório tipo Quiz, nos EUA, semelhante ao Show do Milhão aqui no Brasil. Os especialistas (“universitários”) estavam corretos em 65% dos casos enquanto que a plateia (pela escolha da maioria) estava correta em 91% dos casos, sem comunicação entre eles.

A partir disto, Surowiecki utilizou o termo Sabedoria das Massas ou Multidões (Wisdom of Crowds), para explicar este fenômeno. Segundo este conceito, a opinião da maioria das pessoas leva ao que é correto ou ao melhor caminho. Surowiecki acredita que a maioria poderia acertar o número de feijões numa caixa, sem contagem prévia; bastaria utilizar a média de todos os valores fornecidos individualmente por cada elemento do grupo. Sob certas circunstâncias, grupos são mais inteligentes do que os indivíduos mais inteligentes agindo sozinhos. Quando nossos julgamentos imperfeitos são agregados do modo certo, nossa inteligência coletiva é frequentemente excelente. Os erros acontecem em todas as direções e um neutraliza o outro.

A Sabedoria das Massas também está presente na análise de consultas feitas em mecanismos de busca (*query mining*). Ela ajuda a identificar as palavras mais usadas em conjunto, para sugerir expansões de consulta (reformular consultas acrescentando outras palavras, para dar maior precisão). Spink et al. (2001) analisaram estatisticamente diversas consultas feitas em mecanismos de busca e chegaram às seguintes conclusões:

- o número médio de consultas por sessão é de 4,86;
- usuários utilizam de 2 a 3 termos na consulta (média de termos por consulta = 2,7);
- 52% das consultas em mecanismos de busca são reformuladas;
- 32,5% das consultas modificadas sofreram alterações nos termos submetidos, mas não no número total de termos;
- 41,6% das consultas modificadas incluíram termos novos;
- 25,9% das consultas modificadas excluíram termos.

Outro termo utilizado com significado semelhante é "Inteligência Coletiva", proposto por Pierre Lévy. Ele tem sido usado em diferentes aplicações, mas sempre tendo em comum a ideia de que “o todo é mais que a mera soma das partes”, ou seja, várias partes integradas podem realizar um objetivo que nenhuma parte conseguiria em separado (Lévy, 1998). Este também é o fundamento do conceito de Sinergia, da Teoria Geral de Sistemas. Por

exemplo, a molécula de água (H2O) é formada por dois átomos de Hidrogênio e um átomo de Oxigênio. Hidrogênio e Oxigênio são gasosos em seu estado natural; se olharmos somente as partes, tenderemos a acreditar erroneamente que o estado natural da água é gasoso. A ideia de Sinergia é que elementos em conjunto podem fazer mais que cada um em separado se estiverem atuando de forma integrada (exemplo: mover uma pedra).

A Inteligência Coletiva está fortemente calcada no fundamento de que cada indivíduo só tem parte do conhecimento e a inteligência está distribuída entre os elementos do grupo. Este é também o princípio dos sistemas multiagente inteligentes da Inteligência Artificial. Imagine alguém tentando controlar o trânsito em uma grande cidade. Não há como olhar todas os cruzamentos, mesmo numa sala central com diversas câmeras. Pior ainda tenta sincronizá-las. Mas se cada sinaleira for inteligente o suficiente para saber se há carros passando em cada direção e qual a velocidade (através de sensores), ela poderá comunicar-se com outras sinaleiras e trocar informações. Assim, poderá manter uma direção aberta por mais tempo, se há carros vindo em sua direção e se não há engarrafamento mais adiante. Ou poderá segurar o tráfego em caso contrário. Poderá até mesmo indicar desvios por ruas menos movimentadas, caso receba tal informação de outras sinaleiras.

Premissas para Sabedoria ou Inteligência das Multidões

Surowiecki (2004) propõe 4 premissas para a Sabedoria das Massas:
1. Diversidade de opiniões: cada um sabe uma parte do todo e um complementa o outro;
2. Independência: opiniões individuais não devem ser influenciadas por outras opiniões;
3. Descentralização: experiências locais (individuais) ajudam no contexto global;
4. Agregação: deve haver um mecanismo para juntar as decisões individuais numa decisão coletiva.

Tumer e Wolpert (2004) veem 4 características principais num sistema de inteligência coletiva:
1. Há vários agentes (ou processadores) atuando de forma concorrente, realizando ações que afetam os demais;
2. Há pouca ou nenhuma comunicação centralizada, sendo que cada agente se comunica com alguns poucos agentes, de formas diferentes;
3. Há pouco ou nenhum controle centralizado, sendo que cada agente controle ou influencia outros agentes de formas diferentes, mas não há um agente central controlador de todo o sistema;
4. Existe um objetivo bem especificado, que é o foco do comportamento de todo o sistema.

Crowdsourcing

Na mesma linha de pensamento da Sabedoria das Massas, o uso de muitas fontes de dados podem gerar soluções melhores. Essas fontes podem ser pessoas gerando ideias. O termo lançado por Jeff Howe e Mark Robinson (Wired Magazine, Junho de 2006) se refere a uma chamada aberta para realizar uma função que deveria ser feita internamente. A ideia é que o

público possa participar com opiniões, sugestões e mesmo com a escolha de melhores alternativas. Por exemplo, os responsáveis pelo salgadinho Doritos fizeram um concurso para receber sugestões de propagandas. Reuniram as melhores ideias e fizeram a peça publicitária que foi ao aro no Super Bowl americano. A empresa GoldCorp Challenge distribuiu para geólogos dados geológicos para que examinassem e dessem pareceres. A reunião das informações ajudou a empresa a resolver um problema específico.

As etapas do Crowdsourcing são:
1. Postar um problema no meio público;
2. Receber sugestões de soluções vindas de diversos indivíduos;
3. Recompensar algumas ideias;
4. Utilizar as ideias para benefício da empresa.

É diferente de Open Innovation, que é feito de forma mais organizada e entre empresas.

O Prosumidor

Outra grande causa por que temos muitos dados é que as pessoas querem produzir além de consumir. Alvin Toffler já previa no seu livro 3a Onda, escrito na década de 80, que haveria um novo tipo de cliente, o chamado prosumidor. Este reúne características de consumidor e de produtor. Ou seja, são pessoas que não se contentam em consumir (seja produtos, serviços ou informações), mas também querem oferecer isto a outras pessoas ou produzir elas mesmas seus produtos ou serviços. Isto inclui a imensa cadeia de "selfservices", que estão presentes em restaurantes (você mesmo faz seu prato em buffets), em postos de gasolina (você enche o tanque do seu carro), em lojas de roupas e conveniências (você pega um cestinho, vai até as prateleiras, pega os produtos e coloca no cestinho). O prosumidor também se faz presente em produtos tipo "faça você mesmo" ou "monte em casa". O ato de cortar seu próprio cabelo com uma maquininha também caracteriza um prosumidor. E mais recentemente, com a chegada de impressoras 3D, o próprio cliente vai poder produzir em casa seus produtos.

O prosumidor também está presente em redes sociais e blogs, onde cada usuário pode gerar conteúdo e divulgar para outras pessoas lerem. Podemos dizer que Andy Warhol estava certo quando fez a previsão de que "no futuro todo mundo teria seus 15 minutos de fama".
As empresas estão-se valendo deste tipo de comportamento para dar vazão aos desejos de pessoas que querem produzir algo para outras pessoas e com isto também faturar. Já há sites que permitem ao cliente desenhar uma camiseta para si mas que também ficará à disposição de outros clientes.

Steve Jobs também se valeu do conceito de prosumidor quando lançou o iPad. Ele precisava que seus aparelhos tivessem aplicativos disponíveis para baixar, senão o iPad seria um dispositivo sem utilidade. Mas a Apple não podia criar todos os aplicativos necessários. Em parceria com empresas, Jobs criou a Apple Store. Empresas podem desenvolver aplicativos para iOS (o sistema operacional dos dispositivos móveis da Apple)

e estes aplicativos ficam disponíveis na AppStore para serem baixados, alguns grátis, outros pagos. Assim, a empresa que produz os aplicativos também pode faturar. E a Apple fatura duas vezes (com a venda de aparelhos e com participação nas vendas de aplicativos).

Formas de Coleta de informações: explícita X implícita X por inferência

Há várias formas de coletar dados. A seguir, discutimos as principais.

A coleta de dados explícita acontece quando perguntamos algo a alguém (num entrevista ou questionário) e a pessoa nos dá os dados em forma de resposta. Ou então quando alguém preenche um formulário na Web ou nos diz algo, mesmo sem a gente pedir.

Já a coleta implícita é aquela que utiliza a observação. Não conheço estabelecimento que faça isto, mas é um futuro provável: quando você paga em dinheiro num supermercado, este só registra o que você comprou e como; não ficam registrados dados como seu sexo, idade, etc. Mas imagine que o operador do caixa (check-out) possa observar o cliente e utilizar códigos para dar entrada no sistema de dados que ele está vendo (sexo, faixa etária, estilo de se vestir, se está acompanhado ou não).

Num futuro um pouco mais distante isto já poderá ser feito através da análise de imagens gravadas com câmeras. Já foi feito um experimento que, pelo contorno da pessoa diante de um banner, era possível identificar o sexo e a faixa etária. Paco Underhill e parceiros fazem consultoria para empresas de varejo analisando estatisticamente o comportamento de clientes em lojas. As informações são coletadas por observação direta no ambiente ou em gravações de imagens.

Com esta onda de Big Data por aí, está todo mundo coletando dados sobre todos. A operadora de celular sabe por onde a gente anda e quando. Qual o caminho que costumamos fazer, por onde costumamos andar em cada dia da semana e horário. E se instalarmos aplicativos tipo o Waze no nosso celular, a Google (que comprou o Waze) vai saber até a que velocidade estamos andando. E daí inferir se estamos a pé ou de carro, ou num engarrafamento. Aí alguém inventou a tecnologia de RFID, e ela está em cartões com chips, carros, produtos novos e vai estar em sacolas, carrinhos de supermercados, etc. Então não é só por celular. Os aplicativos e softwares que usamos em celulares, tablets, notebooks e etc também estão avisando onde estamos, se estivermos conectados via Wifi, 3G ou 4G.

Este tipo de coleta também é considerada implícita, apesar de não usar a observação humana. Neste caso, a observação é feita sobre dados eletrônicos, capturados por dispositivos eletrônicos.

Inferir é gerar uma informação a partir de outra. Se você compra muito produto congelado no supermercado, a análise destes dados pode ajudar a inferir que:
a) você tem um bom freezer em casa;
b) você não sabe cozinhar ou não gosta;

c) você é uma pessoa muita atarefada e não tem tempo nem para cozinhar.

A coleta por inferência então é quando o sistema gera informações novas a partir de outras. O nível de inferência é subjetivo de cada organização e certamente aumenta a incerteza sobre a veracidade da informação. Mas muitas empresas assumem o risco desta incerteza, porque mais incerto ainda é não saber nada sobre o cliente.

Utilizando a sabedoria das massas, se numa loja de supermercado a venda de água mineral foi muito acima do normal, é porque faltou água neste bairro. E se na mesma cidade, várias farmácias estão vendendo antigripal, é porque há um surto de gripe. E provavelmente a temperatura também esfriou ou a umidade aumentou.

E isto já chegou à Internet. O Facebook já consegue inferir nossa orientação sexual e tendência política só analisando nossas "curtidas" (ler a reportagem "Estudo mostra que botão 'Curtir' do Facebook revela muito mais do que se imagina sobre o usuário http://oglobo.globo.com/tecnologia/estudo-mostra-que-botao-curtir-do-facebook-revela-muito-mais-do-que-se-imagina-sobre-usuario-7812419).
Um exemplo caso aconteceu em algumas sinaleiras de grandes cidades. Uma pessoa passava pelos carros perguntando ao motorista se queria ganhar um brinde. A grande maioria das pessoas dizia que sim, mesmo que desconfiadas. Então o "entrevistador de sinaleiras" pedia o nome e o telefone do motorista, alegando que depois entraria em contato.

A princípio, parece que só foi utilizada a coleta explícita (perguntas e respostas). Mas se pararmos para pensar, a pessoa só se dirigia a certos tipos de carros. Além disto, anotava mais que o nome e o telefone. Ela anotava o tipo de carro e outros dados que conseguisse coletar (adesivos informando que há bebês no carro, sobre estacionamentos hospitalares, associações e clubes, etc). Então este é um tipo de coleta implícita, por observação.

Além disto, os dados iam para centrais onde eram então analisados. A partir dos dados coletados explícita ou implicitamente, alguém iria fazer uma inferência. Por exemplo, a partir do selo de estacionamento de um hospital colado no vidro de um carro, pode-se inferir que o dono do carro é um médico; daí tem-se o perfil de pessoas com boa renda e alto senso crítico. Se o carro tinha cadeira de bebês, infere-se que há uma família por trás.

Novas tecnologias para coletar e monitorar dados

Novas tecnologias estão surgindo para coletar dados. Chips e antenas de RFID permitem rastrear produtos e até mesmo pessoas (bem como GPS e celulares). A análise de vídeos (imagens) permite capturar movimentos e gestos. Capturas de sons permitem a posterior análise e o reconhecimento de fala. Já há diversos dispositivos para identificação de pessoas por biometria (até mesmo tatuagens já servem para isto).

Diversos sensores estão sendo fabricados e utilizados nas mais diversas situações. Sensores de movimento alertam para intrusos. Sensores de umidade e luminosidade são utilizados na agricultura de precisão. Sensores de rotação são comuns em jogos em aparelhos móveis, mas também servem para estabilizar veículos. Computadores de bordo também usam sensores de proximidade para estacionar de forma autônoma um carro. A medicina no futuro irá utilizar sensores para medir sinais de saúde nas pessoas.

O professor Pentland faz pesquisas com sensores para coletar expressões faciais e utilizar isto para melhorar a comunicação. O futurista Michio Kaku fala que haverá em breve diagnóstico médico por imagens capturadas pelo espelho do banheiro ou pela câmera do celular.

Na China, a análise automatizada de imagens permite reconhecer expressões faciais de alunos em sala de aula para entender seu comportamento, a fim de que professores possam melhorar seus métodos.
https://www.techjuice.pk/this-school-scans-classrooms-every-30-seconds-through-facial-recognition-technology/

O pesquisador Kevin Warwick implantou sensores em seu corpo (ver Figura 5). Isto permite a ele controlar dispositivos à distância, tal como uma mão robótica imitando seus gestos.

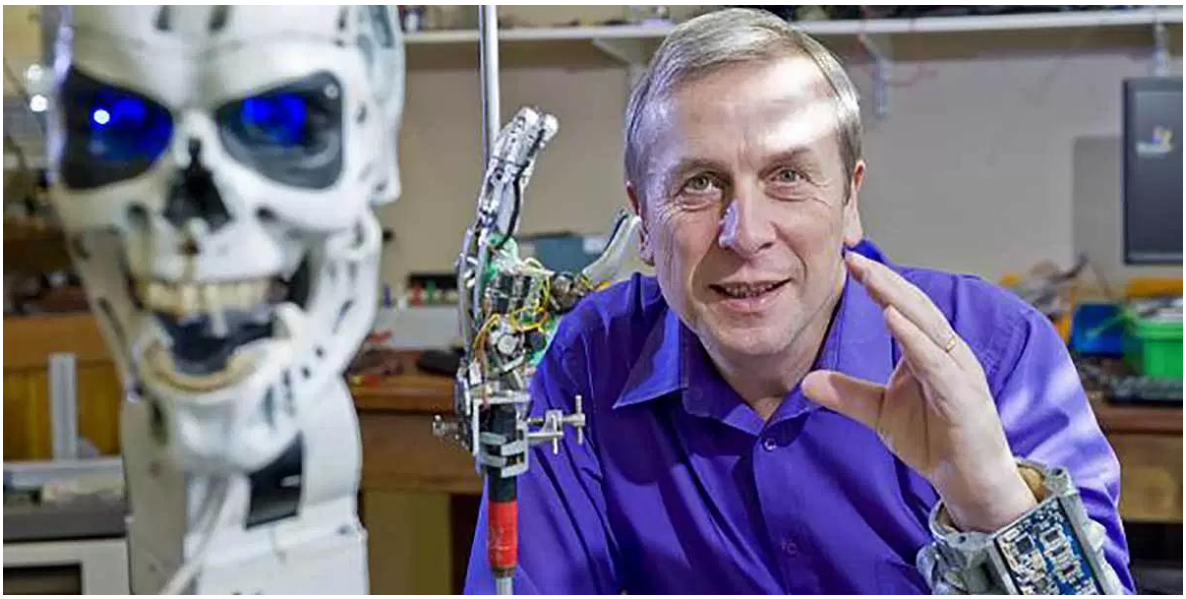


Figura 5: Prof. Kevin Warwick e seus implantes cibernéticos

O pesquisador A. J. Jacobs usa software para coletar dados sobre seu comportamento e com isto então descobrir o que é bom para ter corpo saudável (ver Figura 6). Ele mesmo considera isto útil mas concorda que gera muito stress mental.

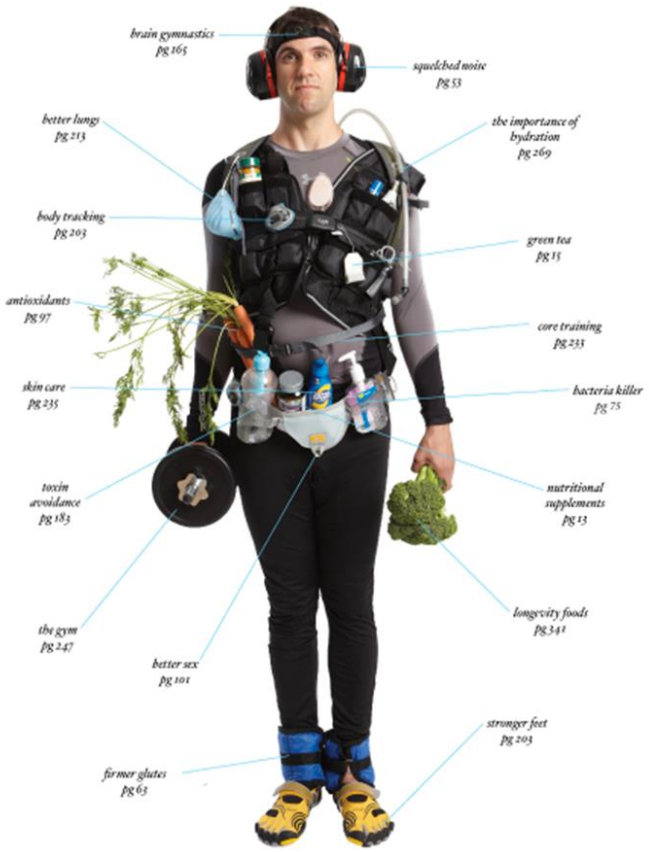


Figura 6: O homem Big Data – Jacobs e seus sensores

Enriquecimento de dados – cruzamento de bases

O enriquecimento de dados compreende acrescentar dados a uma base existente. Por exemplo, se tivermos dados cadastrais de clientes, podemos incorporar dados externos da empresa, por exemplo vindos de outras empresas parceiras ou mesmo de comportamentos capturados fora da empresa.

A vantagem do enriquecimento é ter mais dados para análise estatística, aumentando as chances de encontramos padrões. Por exemplo: uma base de vendas contém dados como data da venda, número de nota fiscal, os itens adquiridos, valor total pago. Se incorporarmos dados dos clientes (cidade, idade, sexo) e dados dos produtos (preço, categoria, tamanho), há mais chances de haver repetições. Além disto, pela técnica de associação, poderemos cruzar dados de produto com dados de clientes e, por exemplo, encontrar padrões entre faixa etária e faixa de preço (ex.: jovens tendem a adquirir produtos de menor valor).

Outro exemplo de enriquecimento: cada click de uma pessoa num site é monitorado. Aí estes dados são cruzados com o que a pessoa comprou pela internet. E depois estes dados são cruzados com dados dos cadastros de lojas físicas. E então a estes dados são somados dados sobre as compras que esta pessoa fez na loja física, fora da Internet. E tudo isto é complementado com dados vindos dos perfis da pessoa nas redes sociais e com o que a pessoa diz em fóruns e blogs.

Isto então passou a ser chamado de CRM Social. A partir de uma identificação do cliente, as empresas podem visitar o perfil do cliente em redes sociais e complementar os dados demográficos. Nem sempre sistemas automáticos tipo robôs e crawlers funcionam, porque muitos destes sites só permitem acessar perfis por senha e por pessoas relacionadas ao cliente. Existe muita informação estruturada, como nome, sexo, estado civil, cidade, idade, etc., mas também há dados não estruturados como fotos, figuras, vídeos e mensagens de texto. Também há comunidades e até mesmo as páginas "curtidas" podem ser utilizadas para traçar o perfil da pessoa. Além disto, as empresas estão também preocupadas em conhecer os amigos dos clientes, quem são e do que gostam. Dizem que amigos possuem gostos semelhantes. Assim, se alguém comprou um produto X, é muito possível que ele divulgue isto para seus amigos e uma boa parcela destes também acabe comprando o mesmo produto. E também, muitas pessoas divulgam suas reclamações e insatisfações por meio de redes sociais ou blogs. A sugestão é que as empresas utilizem técnicas de análise de sentimentos em textos publicados na Internet.

Além disto, é possível pegar dados públicos, disponíveis livremente na Internet. Estes dados não identificam pessoas individualmente, mas dão estatísticas sobre grupos de pessoas. Uma empresa pode comprar uma lista telefônica com nome, endereço e telefone de clientes. Mas não sabe classificar os clientes por dados sócio-demográficos. Então, a empresa pode consultar bases públicas sobre setores censitários. Um setor censitário é diferente de um bairro ou quadra; é uma região, geralmente menor que um bairro mas podendo abranger partes de 2 bairros, que foi pesquisada pelo censo do IBGE. Então, há informações estatísticas sobre cada setor específico. Imagine que a empresa então possui os seguintes dados sobre uma pessoa: o nome é José da Silva e mora na Rua X, n.41. Bom, usando um sistema de GIS simples, pode-se saber o setor censitário onde ela mora. Depois, procuram-se dados estatísticos sobre este setor e, digamos, temos que neste setor:

- 100% das residências possuem 3 TVs;
- 98% possuem 2 banheiros;
- 90% possuem aparelhos de DVD;
- 90% possuem TVs LCD;
- etc.

Agora, de posse destas informações estatísticas, podemos estimar alguns dados sobre José da Silva. Que ele tem 3 TVs, com 100% de chances, que há 98% de chances de ele ter 2 banheiros em casa, e assim por diante.

Então, desta forma, uma empresa combina a lista telefônica com dados censitários e poderá obter um banco de dados de clientes potenciais.

Empresas parceiras também costumam compartilhar dados sobre clientes, por exemplo, administradoras de cartões de crédito, instituições financeiras, redes de varejo, escolas, postos de gasolina, editoras, etc. E há empresas que vendem este tipo de informação (cadastros).

E a cada pesquisa que participamos, com o objetivo de concorrer a prêmios, estamos fornecendo mais dados sobre nós.

Mas não precisa ser só enriquecimento de dados sobre pessoas. Se tivermos o campo cidade em alguma base de dados, podemos incorporar dados sobre as cidades, tais como número de habitantes, geografia, economia principal, nível de escolaridade, renda per capita, índices sócio-culturais como IDH e outros.

Neste caso, pode-se cruzar a cidade do cliente com dados dos produtos adquiridos. Isto nos permitirá, por exemplo, descobrir que tipo de cidade compra mais um certo tipo de produto. Num caso real, uma empresa de comércio eletrônico descobriu que somente clientes de cidades pequenas (com menos de 50 mil habitantes) compravam produtos eletrônicos mais caros (depois descobriu-se que a razão era porque naquelas cidades não havia lojas físicas vendendo tais produtos; enquanto que em cidades maiores, o preço do produto estava muito alto em relação a um concorrente com loja física).

Visão 360º - saber tudo sobre as pessoas

O Big Data também se manifesta no armazenamento de dados sobre clientes. Hoje em dia, as empresas utilizam sistemas de CRM para saber tudo sobre o cliente, e assim atendê-lo melhor e poder satisfazer melhor suas necessidades. Isto se chama a visão 360º do cliente. Ele funciona como o quitandeiro (dono do armazém) da esquina que sabe nosso nome, onde moramos, o que gostamos, como costumamos pagar e até mesmo pergunta por parentes que estejam doentes.

O objetivo não é invadir privacidade, mas poder entender melhor o cliente e evitar produtos ou serviços que não sejam adequados ou mesmo evitar propaganda mal direcionada. Para isto, além de coletar dados demográficos sobre clientes, é preciso saber seus gostos, preferências e hábitos. Tudo isto se consegue mantendo um histórico dos relacionamentos. Não basta só registrar o que o cliente comprou, o que ele quis comprar e não comprou (porque não tinha em estoque ou porque mudou de ideia), também são informações importantes (análise de descartes).

É claro que esta coleta de informações não pode ser feita de uma vez só, para não aborrecer o cliente. Por isto, costuma-se usar o que se chama de "irrigação gota-a-gota", que é coletar poucas informações a cada contato. Também podem ser utilizada as chamadas "golden questions", que são perguntas que acabam trazendo informações diversas. Por exemplo, a profissão da pessoa pode dizer algo sobre seu status social, seu modo de vida e trabalho ou seu nível de exigência e tolerância. Pode-se também usar técnicas de coleta implícita ou por inferência, como já discutido anteriormente.

Além disto, há tecnologias modernas para identificação de pessoas por biometria: identificação por digital, palma da mão, íris, formato do rosto, modo de caminhar ou escrever, voz.

O uso de tecnologias associadas à Internet, tais como blogs, twitter, redes sociais (ex.: Facebook, Linkedin, ResearchGate) e mídias sociais (ex.: Youtube, Flickr, Picasa) permitem conhecer melhor cada cliente. A ideia inicial era utilizar estes recursos apenas para interagir com clientes, por exemplo, gerando promoções ou divulgando a marca. Entretanto, com o avanço das tecnologias, a ideia evolui para complementar o cadastro dos clientes e ajudar a entender melhor seus gostos e comportamentos. Isto então passou a ser chamado de CRM Social.

A partir de uma identificação do cliente, as empresas podem visitar o perfil do cliente em redes sociais e complementar os dados demográficos. Nem sempre sistemas automáticos tipo robôs e crawlers funcionam, porque muitos destes sites só permitem acessar perfis por senha e por pessoas relacionadas ao cliente.

Existe muita informação estruturada, como nome, sexo, estado civil, cidade, idade, etc., mas também há dados não estruturados como fotos, figuras, vídeos e mensagens de texto. Também há comunidades e até mesmo as páginas "curtidas" podem ser utilizadas para traçar o perfil da pessoa.

Além disto, as empresas estão também preocupadas em conhecer os amigos dos clientes, quem são e do que gostam. Dizem que amigos possuem gostos semelhantes. Assim, se alguém comprou um produto X, é muito possível que ele divulgue isto para seus amigos e uma boa parcela destes também acabe comprando o mesmo produto.

Identificação de gestos e ações

Imagine uma loja poder saber quantos clientes tocaram num determinado produto, quantos tiraram o produto da estante, quantos colocaram o produto no cestinho ou carrinho, quantos devolveram. Imagine poder saber se os clientes estão mais interessados nos produtos das prateleiras mais altas, mais baixas ou do meio. Quanto tempo o cliente ficou com o produto na mão, talvez lendo as informações do rótulo. Quantos clientes passaram na frente de uma gôndola. Quantos pararam.

Paco Underhill (1999) e colegas de empresa analisam este tipo de dado mas usando câmeras comuns (ver livro Why we buy). Ele filmam clientes em lojas (com consentimento ou aviso), depois observam os vídeos anotando os tipos de ações e atributos dos clientes. Não há identificação do cliente. Apenas generalizam padrões de comportamento. Eles usam perto de 200 variáveis para caracterizar comportamentos de clientes em lojas. Incluindo como o cliente estava vestido, que caminho fez dentro da loja, e outros já citados acima.

Pois bem, este também é o objetivo a que se propõe o trabalho de conclusão de Alessandro Tyska Menezes. Usando um aparelho Kinect da Microsoft (dos que vem com o jogo Xbox mesmo), posicionado acima de uma estante ou gôndola, e um software criado por ele mesmo, é possível identificar movimentos de pessoas em frente ao aparelho. Depois, o software faz a análise estatística dos movimentos identificados, indicando quais movimentos foram feitos com mais frequência.

Mas o mais legal é que o software também minera padrões sequenciais, utilizando o algoritmo "a priori" de Agrawal e Srikant (1995). Este algoritmo identifica sequências de ações mais frequentes. Ou seja, quantas pessoas colocaram o produto no cesto após retirá-lo da estante. E isto permite estimar a probabilidade de uma pessoa colocar um produto no cesto, após retirá-lo da estante. Isto é possível analisando a sequência de ações. Se uma pessoa retira um produto da estante, o que ela faz depois ? Que ações são posteriores a uma determinada ação ?

As ações referem-se a movimentos que podem ser identificados pelos sensores 3D do Kinect (Figura 7), tais como levantar o braço, esticar, coçar a cabeça, abaixar o braço, além de saber distinguir braço direito do esquerdo e também poder observar quantas pessoas estão em frente ao aparelho.

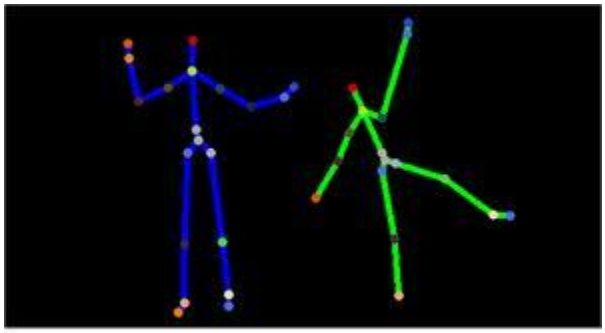


Figura 7: Movimentos identificados pelo Kinect

No caso do trabalho do Alessandro, o objetivo é poder auxiliar especialistas em marketing e promotores a definir estratégias para colocação de produtos ou para promoção e visualização em lojas de varejo. Link para o TCC do Alessandro Tyska Menezes.
http://www.intext.com.br/tcc-tyska.pdf

Limitações da memória humana

A nossa memória nos engana. Só guardamos partes das histórias e fatos que ouvimos, lemos ou vemos. As memórias são muito mais sensíveis à incorporação de informação adicional nos primeiros minutos ou horas após a aquisição. Um comentário pejorativo, não-factual, pode interferir sensivelmente com a memória de fatos concretos aprendidos poucos minutos ou horas antes.

A recuperação de eventos diretamente experimentados pela pessoa é chamada memória episódica, enquanto que conhecimentos gerais formam a memória semântica. O primeiro tipo é mais fácil de ser recuperar (Swap et al., 2001). A memória implícita (ou não-declarativa), refere-se ao que se sabe, embora talvez de forma não consciente. Este tipo de memória é importante para processos cognitivos tais como habilidades procedurais, hábitos, aprendizados emocionais, habilidades motoras e sensoriais (Izquierdo, 2002).

Também temos uma memória de trabalho, que serve para manter durante alguns segundos ou minutos, a informação que está sendo processada no momento (Izquierdo, 2002). Este tipo de memória é útil para conservar na consciência, por exemplo, a terceira palavra da frase anterior. Uma de suas limitações é que ela não pode conter demasiada informação de uma vez só (Smith, 1997). De acordo com Kleiman (1993), a memória de trabalho pode trabalhar com aproximadamente 7 unidades ao mesmo tempo.

Já a memória de curta duração, diferentemente da memória de trabalho, estende-se desde os primeiros segundos ou minutos seguintes ao aprendizado e pode durar até 3-6 horas (Izquierdo, 2002). Todo conhecimento de uma pessoa passa primeiro pela memória de curta duração e depois se consolida na memória de longo prazo. Ao contrário da informação na memória de curto prazo, a informação na memória de longo prazo exige uma ação positiva para recuperá-la (Smith, 1997). Quando acrescentamos algo aos nossos conhecimentos, modificamos a informação já existente. Para poder recuperar uma informação, é preciso relacioná-la com o que já temos armazenado. Uma informação nova, se não puder ser relacionada a algo pré-existente, será mais difícil de recuperar.

Mlodinow (2013) argumenta que "nosso inconsciente capta os dados incompletos fornecidos por nossos sentidos, completa o que está faltando e passa essa percepção à nossa mente consciente. Quando olhamos uma cena, achamos que vemos uma imagem nítida e bem-definida, como uma fotografia, mas na verdade vemos apenas uma pequena porção da imagem com clareza, e nosso cérebro subliminar pinta o resto... Quando tentamos dar uma explicação para nossos sentimentos e comportamentos, o cérebro realiza uma ação que sem dúvida o surpreenderia: faz uma busca no seu banco de dados mental de normas culturais e escolhe algo plausível... nosso inconsciente pode escolher entre um número enorme de interpretações para alimentar nossa mente consciente. No fim, sempre vamos achar que estamos lidando com fatos, embora na verdade estejamos lidando com a conclusão preferida."

A repetição também é um fator importante para fixação de ideias (Gladwell, 2005; Thompson, 2018). Entretanto, a ideia é não repetir da mesma forma, mas utilizar meios diferentes, para não "enjoar" nem gerar o efeito "burnout" (as pessoas já sabem do que se trata e desviam a atenção; como as propagandas antes dos vídeos no Youtube; a maioria das pessoas fica esperando aparecer o botão "pular").

Os tipos de memórias segundo Ivan Izquierdo são:

- Memória sensorial: informação de entrada, do meio-ambiente, recebida pelos sentidos;

- Memória Rápida ou de Curto Termo: duração de 10 a 30 segundo podendo ser retida até por 6 horas;
- Memória de Trabalho ou Temporária: é o "meio de campo" entre memória rápida e permanente; organiza os dados recebidos, filtra e manda para a memória permanente; recupera da permanente para a rápida; faz associações;
- Memória Permanente ou de Longo Termo.

Conforme Mlodinow (2013), "as pessoas têm uma boa lembrança dos aspectos principais dos eventos, mas uma má lembrança dos detalhes; segundo, quando pressionadas pelos detalhes não lembrados, mesmo pessoas bem-intencionadas, fazendo sinceros esforços para ser precisas, e sem querer, preenchem os detalhes inventando coisas; terceiro, as pessoas acreditam nas lembranças que inventam... nossos erros de memória têm uma origem comum – todos são artefatos das técnicas que nossa mente usa para preencher as inevitáveis lacunas. Essas técnicas incluem confiar em nossas expectativas, nos nossos sistemas de valores, de forma geral, e em nossos conhecimentos prévios. Como resultado, quando nossas expectativas, nossos valores e conhecimento prévio estão às turras com os acontecimentos reais, nosso cérebro pode ser enganado."

As memórias adquiridas em estado de alerta e com certa carga emocional ou afetiva são melhor lembradas que as memórias de fatos inexpressivos ou adquiridas em estado de sonolência. Situações onde o corpo libera adrenalina e beta-endorfina são mais eficazes para preservar a memória do evento (Izquierdo, 1989).

Conforme Gais et al. (2006), a recuperação da memória é desencadeada por algum estímulo semelhante àqueles inerentes a cada experiência. Algumas memórias requerem muitos estímulos para serem lembradas: é o caso das memórias de atos-motores elaborados ou complexos, ou de procedimentos com muitas regras. Por isto, a mensagem deve ser sempre simples e procurar remeter ao contexto do usuário. Além disto, Gais e parceiros também descobriram que cochilos ajudam a memorizar.

Uma mensagem bem estruturada será melhor lembrada. Isto porque organizamos a atividade mental em hierarquias de eventos, colocando fronteiras para dividir e entender partes. Isto ajuda a memorizar e aprender. Para melhorar a memória, precisamos entender o contexto e a sequência de eventos e minimizar as interrupções (Zacks, 2010).

Segundo Kandel (2009), a memória implícita, que é responsável "não apenas por habilidades perceptuais e motoras simples, mas, em princípio, também pelas piruetas de Margot Fonteyn, pela técnica de Wynton Marsalis ao trompete, pelos precisos golpes de fundo de quadra de Andre Agassi e pelos movimentos das pernas de um adolescente andando de bicicleta", é uma forma de memória que não é recuperada conscientemente. "A memória implícita nos guia naquelas rotinas bem estabelecidas que não são controladas conscientemente.". Já a memória explícita "requer atenção seletiva para a codificação e para a recuperação de informações... A repetição constante pode transformar a memória explícita em memória implícita. Isto é o hábito."

Conforme Kandel (2009), “embora a memória de longo prazo geralmente exija o treinamento repetido, espaçado, com intervalos de descanso, ela às vezes ocorre em seguida a uma única exposição que não se apresenta emocionalmente carregada.”

Segundo o neurocientista Antônio Damasio (2005), a recuperação de memórias é facilitada por marcadores somáticos. “Os marcadores somáticos são um caso especial do uso de sentimentos gerados a partir de emoções secundárias. Essas emoções e sentimentos foram ligados, pela aprendizagem, a resultados futuros previstos de determinados cenários. Quando um marcador somático negativo é justaposto a um determinado resultado futuro, a combinação funciona como uma campainha de alarme. Quando, ao contrário, é justaposto um marcador somático positivo, o resultado é um incentivo... Os marcadores-somáticos são, portanto, adquiridos por meio da experiência, sob o controle de um sistema interno de preferências e sob a influência de um conjunto externo de circunstâncias que incluem não só entidades e fenômenos com os quais o organismo tem de interagir, mas também convenções sociais e regras éticas.”

O cérebro funciona através de associações, ou sinapses entre neurônios. As informações que estão armazenadas ou sendo processadas também são influenciadas por associações. Primação é um efeito experimental que acontece quando um evento antecedente (prime) influencia o desempenho de um evento posterior (alvo). Um exemplo prático seria o caso de uma palavra ser acessada mais rapidamente se for precedida da palavras relacionadas, seja por semântica (carro-estrada), fonologia (cadeia-cadeira) ou morfologia (cantar-canto). O efeito de primação pode ser obtido por meio de estímulos auditivos ou visuais, linguísticos ou por figuras.

Outra forma de primação é o uso de mensagens subliminares. Nosso consciente não percebe informações que tenham menos de 50 milissegundos de exposição. Mas o inconsciente percebe e tende a recuperar esta memória mais tarde. Os casos mais famosos incluem inserir imagens no meio de filmes. Um filme em geral é apresentado através da sequência de 30 imagens (quadros por segundo). Se trocarmos uma destas 30 imagens, o consciente não perceberá, mas a imagem será observada por nosso sentido de visão.

Primação semântica é quando a apresentação de uma informação associada a um conceito torna mais acessível palavras semanticamente relacionadas com esse conceito (Garcia-Marques, 2005). Já a primação afetiva é o impacto gerado nos pensamentos a partir de um estímulo afetivo como por exemplo um sorriso, um olhar de tristeza (Garcia-Marques, 2005).

A primação acontece no nosso dia a dia porque o cérebro procura atalhos para poupar energia. E isto influencia o modo como vemos o mundo. Pode ajudar a recuperar uma memória, como pode também desfocar nossa atenção. Uma boa maneira de memorizar (decorar) conteúdos é fazendo associações.

Um caso negativo de associação é o efeito Stroop. É uma interferência no tempo de reação de uma tarefa. Por exemplo, quando uma palavra como azul é impressa numa cor diferente

da expressa pela palavra, ocorre um atraso no processamento mental, gerando um tempo de reação mais lento e uma possibilidade maior de erros.

Por estas razões, os seres humanos precisam de técnicas especiais e ferramentas artificiais para conseguir recuperar ou gerenciar grandes volumes de dados armazenados na memória natural (como discutiremos mais tarde).

Carey (2015) discute diversas técnicas para melhorar a memorização e a recuperação de informações.

No livro abaixo são discutidas algumas ferramentas para recuperação de informações explícitas:
LOH, Stanley. **Memória organizacional:** como coletar, armazenar e recuperar conhecimentos explícitos e tácitos. Porto Alegre, 2018.

Storytelling – como associar dados para uma visão holística

Storytelling é o termo utilizado hoje em dia para representar a técnica de transmissão de conhecimento através de histórias. Se você quer que as pessoas lembrem de algo e acredite na mensagem, é melhor dar a informação no formato de uma história. Quanto mais detalhes (quanto mais rica a narrativa), mais facilmente a história será tomada como verdadeira.

Histórias ajudam a dar coesão aos dados. Por isto os mitos funcionam bem, porque criam e contam histórias que permitem ao ser humano entender melhor coisas fora do alcance da lógica e da Física (Campbell, 1995). Segundo Campbell (2002), a mitologia facilita a "organização de imagens e narrativas simbólicas", gerando "metáforas das possibilidades da experiência humana e a realização de uma dada cultura num determinado tempo".

Segundo Shermer (2011), "os humanos são animais buscadores de padrões, contadores de histórias, que estão à procura do significado profundo atrás de eventos na aparência aleatórios do dia a dia."

Como Duhigg (2016) diz: os seres humanos "estão sempre tentando descobrir como as informações se combinam." E para isto, as histórias funcionam bem como modelos mentais, que combinam dados permitindo ao ser humano entender melhor o conjunto de dados e dar explicações lógicas.

Segundo Swap et al. (2001), histórias são parcialmente episódicas e por isto são melhores de serem lembradas. Histórias são utilizadas para descrever assuntos complicados, expor situações e/ou comunicar lições aprendidas ou, ainda, para interpretar mudanças culturais. Jesus transmitiu conhecimentos através de parábolas, que eram histórias com um ensinamento.

E por que as pessoas gostam de histórias ? Segundo Bloom (2012), a nossa atividade mais prazerosa é a imaginação, que a gente desenvolve através de histórias que nos levam a outros mundos.

Além disto, as histórias também facilitam a memorização. Os Incas não usavam registros escritos ou desenhos para transmitir suas memórias. Mas nem por isto suas histórias se perderam, porque usavam muito bem a transmissão oral. Para fortalecer a memória de quem ouvia, o conhecimento era transmitido através de histórias. E não foram só os Incas que fizeram isto. Muitas das histórias religiosas que conhecemos hoje em diferentes religiões (mitos ou não) foram transmitidas por via oral, até que o pergaminho, o papiro e depois o papel e o livro fossem inventados.

Conforme Swap et al. (2001), a recuperação de eventos que foram diretamente experimentados se chama "memória episódica". Já o conhecimento geral sobre o mundo é a "memória semântica". Segundo os mesmos autores, a memória que vem com a experiência pessoal é mais facilmente recuperada, devido a algum tipo de associação (Damásio chama de marcadores somáticos).

Para Zacks e Swallow (2007, 2010), o ser humano organiza a atividade mental em hierarquias de eventos, colocando fronteiras para dividir e entender as partes. Isto ajuda a memorizar e aprender, a entender o contexto e as sequências e lidar com interrupções. Por isto, o registo de memórias através de histórias facilita a recuperação.

Além disto, por razões que a neurociência explica, ao ouvir ou ler histórias, estimulamos as mesmas áreas cerebrais como se estivéssemos tendo a experiência e assim assimilamos melhor as memórias recuperadas (Speer et al., 2009).

As histórias podem ser repassadas de forma oral, mas a modernidade permitiu fazer isto de formas mais divertidas, tais como histórias em quadrinhos e vídeos. Segundo McLuhan (2014), as histórias em quadrinhos "possuem uma forma de expressão altamente participante, perfeitamente adaptada à forma em mosaico do jornal. Dão também um sentido de continuidade de um dia para o outro. Também as notícias sobre pessoas são de baixo teor informacional e por isso convidam a que o leitor as preencha, exatamente como acontece com a imagem da televisão e a radiofoto. Eis por que a TV constitui um golpe fatal nas estórias em quadrinhos, das quais foi antes um rival do que um complemento."

Sobrecarga de informações

Alguém vai dizer que o volume de informações é bom, porque as pessoas e organizações possuem mais informação para tomar decisões. Entretanto, muitas informações atrapalham as tarefas, porque o ser humano tem limitações para lidar com grandes volumes de informações, como discutido antes. Acabamos por nos "afogar" com tantos dados. Herbert Simon (citado em Goleman, 2013) alertou para o fato de que o que a informação consome é "a atenção de quem a recebe. Eis por que a riqueza de informações cria a pobreza de atenção".

Isto gera stress mental por termos mais opções para escolher, mais informações para ler, mais conhecimento para aprender e por não conseguirmos lidar com tanta informação disponível e nem mesmo conseguir encontrar as informações que precisamos (*information overload*). É como uma mesa cheia de papéis e a gente sabendo que a informação que a gente procura está em algum destes papéis nesta mesa. Este é o efeito da sobrecarga de informações. Temos dificuldades para distinguir o que é mais importante. Talvez, nestas situações o hábito, as intuições e instintos funcionem melhor.

A sobrecarga de informações (information overload) deixa as pessoas perdidas com tanta informação, pois não conseguem encontrar o que querem. A sobrecarga faz muita coisa relevante parecer irrelevante e o contrário também. Como há falta de tempo e recursos para analisar todos os dados e alternativas, as pessoas filtram informações e alternativas. Mas muitas vezes, as heurísticas utilizadas não são as melhores, como advertem Tversky e Kahnemann nos seus artigos.

O psicólogo Richard Nisbett (citado por Tetlock e Gardner, 2016) alerta que informações irrelevantes deturpam os julgamentos e decisões. Participantes de um estudo que recebiam informação irrelevante acabavam perdendo a confiança.

Outra consequência negativa da sobrecarga é o stress de querer estar sempre atualizado e recolhendo dados, também conhecido pela sigla FOMO (Fear Of Missing Out).

De acordo com Mlodinow (2013), "o sistema sensorial do homem envia ao cérebro cerca de 11 milhões de bits de informação por segundo. Porém, qualquer pessoa que um dia tenha tomado conta de algumas crianças que falam ao mesmo tempo pode testemunhar como nossa mente consciente não consegue processar algo próximo desse número. A verdadeira quantidade de informação com que podemos lidar foi estimada em algo entre dezesseis e cinquenta bits por segundo. Portanto, se nossa mente consciente tentasse processar toda essa informação enviada pelo sistema sensorial, nosso cérebro travaria, como um computador sobrecarregado."

Segundo Alves (2014), "cada vez mais pessoas se queixam de fadiga mental, ou seja, queixam-se dos turbilhões que chegam à mente sem que possam se sentir aliviadas em momento nenhum. Elas fazem pequenas escolhas e tomadas de decisão o dia todo e, então, quando chegam ao final do dia, estão esgotadas, com as energias sugadas e, evidentemente, sem pique para brincar com os filhos, fazer uma boa refeição, estudar, ler um bom livro ou até mesmo se divertir (inclua nessa categoria fazer sexo)".

Sobrecarregar o consumidor com informações não ajuda a vender. Pode paralisar o cliente. Bons argumentos e informações complementares devem ser a exposição. Além disto, o cliente quer ter uma boa experiência de consumo e não somente adquirir algo que lhe satisfaça uma necessidade. Os meios são tão importantes quanto os fins.

Gladwell (2005), no livro Blink, afirma que, quando estamos em situações de estresse, a nossa mente foca e reduz informações. Ele conta vários casos de policiais relatando

situações de estresse, atirando contra ou abordando criminosos; o excitamento deixava as pessoas cegas. Veem somente um cenário principal e descartam o que está em volta, inclusive sons.

Eppler e Mengis (2004) estudaram de forma extensa o problema de sobrecarga de informações. Eles elencaram diversas causas incluindo:

- fatores pessoais (limitações para processamento, atitudes, motivação, experiência)
- características da informação (complexidade, ambiguidade, qualidade, incerteza)
- parâmetros de tarefas e processos (complexidade, rotina, pressões, interrupções, padrões)
- características organizacionais (trabalho colaborativo, centralização, acúmulo para demonstra poder)
- tecnologias da informação (e-mails, sistemas de pushing, canais, baixo custo de duplicação, velocidade de acesso)

Eppler e Mengis (2004) também listam sintomas e efeitos e dão dicas de maneiras para combater a sobrecarga, como por exemplo: treinamento, prioridades, técnicas para gestão do tempo, melhorar qualidade da informação, foco, regras de uso, categorização, técnica para visualização, interfaces inteligentes, padronização de procedimentos.

Karr-Wisniewski e Lu (2010) estudaram a perda de produtividade devido à sobrecarga de informações. Concluíram que as causas principais estão relacionadas às pessoas que usam tecnologias e não à tecnologia em si. Eles dão dicas de soluções para o problema, incluindo: mecanismos de busca, resumos, personalização e recomendação, portais e ferramentas para visualização (dashboards), etc.

A Cauda Longa

Hoje em dia, há muitas opções de filmes, músicas, produtos e atrações para escolha das pessoas. Em parte, isto acontece porque as empresas estão oferecendo produtos para microssegmentos de clientes. Este fenômeno pode ser visto em livrarias e revistarias (há livros e revistas sobre assuntos muito específicos, como revistas de náutica para quem gosta de lanchas rápidas) e também em sites de produtos esportivos (há chuteiras de todas as cores e modelos).

Em geral, as empresas preferem comercializar somente os produtos que saem mais. Por que arriscar com produtos que não vendem muito, podendo a empresa ficar com o produto encalhado. Só que, com a Internet e a logística atual, é possível vender também os produtos menos vendidos e entregá-los em qualquer parte do mundo. Este fenômeno foi chamado de "a cauda longa" ("the long tail") por Chirs Anderson (2006).

Se fizermos um gráfico de Paretto para os produtos vendidos por uma empresa, relacionando produtos (eixo X) e quantidade vendida (eixo Y), os produtos mais vendidos ficarão mais à esquerda e os menos vendidos mais à direita, formando um ranking, como mostra a Figura 8. Desta forma, pode-se notar que a linha do gráfico ou as barras começam

no alto (com os produtos mais vendidos) e vão diminuindo de tamanho, formando uma linha longa de produtos que são pouco vendidos (a tal da “cauda longa”).

Empresas como NetShoes e Amazon se utilizam da cauda longa. A Amazon é considerada a livraria sem livros, pois seu catálogo é imenso. Quando o cliente clica para comprar um produto, aí é que a Amazon sai em busca do produto. Ela não precisa ter o produto em estoque. Basta ter fornecedores (saber o que ela pode vender) e um bom esquema logístico, utilizando parceiros de transporte, para poder entregar o produto em qualquer parte do mundo. A NetShoes também possui um catálogo com diferentes marcas, modelos, cores, tamanhos e até mesmo permite a personalização dos calçados que comercializa.

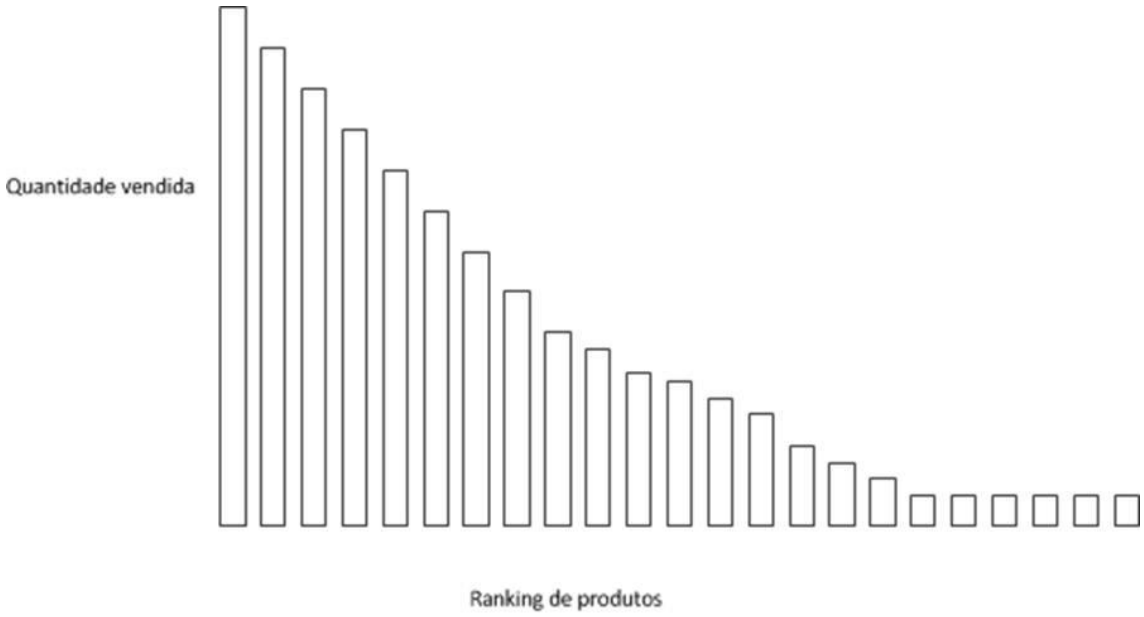


Figura 8: A Cauda Longa para produtos

A Atenção

Em meio a tantas mensagens, tantas informações, num mundo tão cheio de opções e numa vida tão corrida, é muito difícil conseguir a atenção das pessoas.

Nirenberg (1981) comenta que há 3 níveis de atenção:
a) a não-atenção: quando a pessoa não recebe a mensagem;
b) o apenas receber: a pessoa pode até repetir ou lembrar a mensagem mas não entende as ideias;
c) o pensar: quando a pessoa raciocina sobre o que está recebendo.

Segundo Carr (2010), quando alcançamos o limite de nossa memória de trabalho, fica mais difícil distinguir o que é informação relevante das irrelevantes (sinal do ruído).

Segundo Simons e Ambinder (2005), costumamos colocar foco em objetos centrais. Entretanto, isto tira a atenção dos detalhes. Na maioria das situações, as pessoas procuram categorizar situações e não ver detalhes. Uma mensagem muito complexa não será entendida por completo.

Conforme Goleman (2013), "pensar profundamente exige manter a mente focada. Quanto mais distraídos estamos, mais superficiais são as nossas reflexões. Da mesma forma, quanto mais curtas as nossas reflexões, mais triviais elas tendem a ser."

Segundo Gladwell (2005), no livro "Blink – decisões num piscar de olhos", o stress deixa a pessoa focada; ela não vê nada ao redor. Isto acontece muito em casos com policiais, mas quem vive em cidades grandes, com tantos compromissos, vive focado e muito possivelmente não conseguirá ver muitas das mensagens que recebe no seu dia a dia.

Levitin (2015) afirma que "o cérebro humano evoluiu para esconder de nós as coisas em que não estamos prestando atenção. Em outras palavras, muitas vezes possuímos um ponto cego cognitivo: não sabemos o que estamos perdendo porque nosso cérebro consegue ignorar completamente aquilo que não representa uma prioridade para ele no momento — mesmo que esteja bem diante de nossos olhos... graças ao filtro de atenção, acabamos vivenciando boa parte do mundo no piloto automático, sem registrar a complexidade, as nuances e muitas vezes a beleza do que está diante de nossos olhos." Por isto, ocorre um grande número de falhas de atenção. Esta chamada "cegueira por desatenção" é famosa no vídeo de basquete onde se pede para as pessoas contarem os passes dados pelos jogadores de camiseta branca e acabam não notando uma pessoa fantasiada de gorila passando no meio da ação.

Steven Johnson fala que estamos mais distraídos. Isto é devido a mais informações que estamos recebendo. E também pela vontade de fazer mais coisas ao mesmo tempo, como por exemplo, olhar TV, acessar Facebook, ouvir música, digitar no Whatsapp.

Segundo Goleman (2013), trabalhar em multitarefa é ruim. A troca rápida de atenção enfraquece o envolvimento completo e concentrado. Rosen (2008) complementa que isto prejudica o aprendizado, a retenção, a memória e o QI.

Por outro lado, conforme Goleman (2013), "a atenção firmemente focada se cansa — muito parecido com o que ocorre com um músculo que trabalha demais — quando a forçamos ao ponto da exaustão cognitiva. Os sinais de fadiga mental, como uma queda na efetividade e um aumento da distração e da irritabilidade, significam que o esforço necessário para manter o foco esgotou a glicose que alimenta a energia neural. O antídoto para a fadiga da atenção é o mesmo para a fadiga física: descansar."

Falando em descansar, Johnson (2012) afirma que jogos ajudam a manter foco e atenção. Goleman (2013) concorda: "de um modo mais geral, foi identificado que vários games melhoram a acuidade visual e a percepção espacial, a mudança de atenção, a tomada de decisão e a capacidade de acompanhar objetos (embora muitos desses estudos não nos permitam saber se as pessoas atraídas pelos games já eram um pouco melhores do que a

média nessas habilidades mentais ou se foram os games que as melhoraram). Games que oferecem desafios cognitivos cada vez mais difíceis — a necessidade de tomar decisões mais precisas e desafiadoras, e de ter reações em velocidades mais rápidas, com a atenção completamente focada, aumentando a envergadura da memória de trabalho — promovem mudanças cerebrais positivas. Vídeo games focam a atenção e nos fazem repetir movimentos sem parar, de modo que são tutoriais poderosos. Isso apresenta uma oportunidade para o treinamento do cérebro."

Penteado (1976) dá algumas dicas de estímulos para atenção:
- a intensidade: seja de luz, som, etc.;
- a repetição;
- a modificação: repetir de formas diferentes para evitar a monotonia;
- o contraste: o que é diferente se sobressai; deve-se procurar isolar um objeto para conseguir atenção sobre ele.

Heat e Heath (2007) sugerem 6 características para uma ideia chamar atenção: simplicidade, curiosidade (inesperado), ideia concreta (sem abstração), credibilidade (estatísticas), emoções e histórias (para ativar o cérebro).

Tecnologias também podem ajudar a chamar a atenção de pessoas. Por exemplo, há carros equipados com sensores que detectam se os olhos do motorista estão abertos ou fechados, e avisam o motorista para evitar desastres. Também há sensores para chamar atenção de motoristas com altos teores de álcool.

Apesar de as tecnologias ajudarem a chamar ou manter atenção de pessoas, Duhigg (2016) lembra casos em que a automação deixou as pessoas mais desatentas devido à restrição cognitiva ("um curto-circuito mental que às vezes acontece quando o cérebro é obrigado a passar abruptamente de um estado relaxado de automação para um ansioso de atenção"). Segundo Duhigg, "a restrição cognitiva pode fazer as pessoas se concentrarem demais naquilo que estiver bem diante de seus olhos ou se distraírem com tarefas imediatas." E isto pode gera acidentes, como os relatados por Duhigg (2016): "estudos das universidades Yale, Harvard e Berkeley, da UCLA, da Nasa, do National Institutes of Health e de outras organizações constataram uma probabilidade particularmente maior de erro quando as pessoas são obrigadas a se alternar entre a automaticidade e o foco, e esses erros representam um perigo especial quando sistemas automáticos se infiltram em aviões, carros e outros ambientes em que equívocos podem ser trágicos."

Os Superficiais de Nicholas Carr

Com o advento da Internet, dos mecanismos de busca e aplicativos de mensagens e postagens rápidas, como Twitter, Instagram, Whatsapp e Facebook, estamos sempre fazendo leituras rápidas ou lendo somente manchetes. É uma nova forma de leitura dinâmica, com novas conexões, juntando compreensão de texto, imagens e padrões visuais ao mesmo tempo, de forma não sequencial.

Carr (2010) cita estudos de O´Shea sobre jovens: estão lendo em diagonal, pulando coisas e procurando por informações importantes. Isto não é de todo ruim. Gary Small (citado por Carr, 2010) afirma que usuários experientes do Google demonstram mais atividades cerebrais e em novas áreas; é um novo cérebro. Quando pessoas fazem buscas na Internet o cérebro exibe um padrão diferente da atividade de leitura de um texto ou livro. A atividade de leitura está associada com regiões cerebrais relativas a linguagem, memória e processamento visual, mas não aciona regiões pré-frontais associadas com tomada de decisão e resolução de problemas, como fazem as atividades de busca na Internet.

Nicholas Carr (2010, 2008) lembra que os Fenícios inventaram o alfabeto que deu origem aos alfabetos ocidentais justamente porque precisavam acelerar a escrita e leitura. E Gutenberg inventou a prensa móvel para imprimir mais rapidamente livros. O vocabulário de todas as línguas também evoluiu, gerando novas palavras, para sintetizar pensamentos que antes precisavam ser expressos com muitas palavras. É como McLuhan já dizia: o meio ou as mídias influenciam o modo como pensamos. Estamos evoluindo para ser caçadores e coletores de informação na selva eletrônica.

McLuhan (1962) comenta o caso de povos africanos que não conseguem entender filmes (imagens em movimentos). O mesmo autor atesta que palavras escritas perdem as emoções da palavra falada. Um sentido pode anestesiar outro. O visual faz perder outros sentidos. Por isto, Cezanne pintava diferente, valorizando características dos objetos (volume, peso) que não apareciam nas imagens.

A evolução das mídias e dos meios eletrônicos. Segundo McLuhan (2014), as histórias em quadrinhos "possuem uma forma de expressão altamente participante, perfeitamente adaptada à forma em mosaico do jornal. Dão também um sentido de continuidade de um dia para o outro." A nova onda de prosumidores, permitindo que as pessoas comentem tudo, resulta em informações que são verdadeiros quebra-cabeças, criados a partir de pequenas partes, contribuições de diversas pessoas através de meios colaborativos.

Apesar da quantidade e da diversidade de dados em imagens, sons, textos, etc. (como discutiremos adiante), precisamos de ferramentas para síntese de informações na linguagem e na comunicação (como será discutido adiante). Por isto, surgiram emoticons e emojis, softwares para gerar resumos, software para gerar gráficos e dashboards, algoritmos para encontrar padrões (data mining), etc.

O livro "Abundância" é otimista, mas Diamandis e Kotler (2012) alertam para problemas que podem ocorrer em consequência deste novo modo de obter informações. Eles contam de uma pesquisa feita pela National Governors Association, que entrevistou 300 professores universitários sobre suas aulas para turmas de calouros. "Os resultados: 70% disseram que os estudantes não conseguiam entender materiais de leitura complexos, 66% disseram que os estudantes não conseguiam pensar analiticamente, 62% disseram que os estudantes escreviam mal, 59% disseram que os estudantes não sabem fazer pesquisas e 55% disseram que os estudantes não conseguiam aplicar seus conhecimentos."

Com as tecnologias modernas, não precisamos mais guardar tantas informações na memória; basta encontrar pelo Google ou guardar em aplicativos no smartphone. Não precisamos mais quebrar a cabeça com questões de raciocínio; basta procurar no Google a resposta. Nicholas Carr discute se o Google está nos deixando burros: https://www.theatlantic.com/magazine/archive/2008/07/is-google-making-us-stupid/306868/

A questão não é deixar de usar as tecnologias, mas saber utilizá-las para melhorar nossa vida, incluindo memorização, raciocínio e atenção. Johnson (2012), em "Tudo o que é ruim é bom pra você", discute como jogos, TV, Internet, etc. podem nos ajudar: "Novas tecnologias que desafiem a mente sem exigir demais dela, lançadas em ciclos cada vez mais curtos, farão a linha que aponta nossa capacidade de sondar e dominar sistemas complexos subir continuamente".

"Páginas em branco, fotos coloridas" é como Humberto Gessinger resumiu bem o momento atual: as pessoas não querem mais ler; só ver figuras ou filmes. Séries atraem mais que livros. Nosso cérebro pode estar mudando. Talvez a gente regrida 50 mil anos, para só se comunicar por imagens, como no tempo das cavernas. Ou como disse José Saramago, "De degrau em degrau, vamos descendo até o grunhido".

Como encontrar dados em grandes volumes

Informações podem ser armazenadas de duas formas: estruturadas ou não estruturadas. As informações estruturadas são chamadas assim porque são facilmente interpretadas ou entendidas. Dados estruturados normalmente são armazenados em tabelas, planilhas ou arquivos XML (arquivos textos com marcas indicando o significado de cada parte do conteúdo). Por exemplo, quando temos uma tabela com dados de amigos, sabemos que, em cada linha, teremos dados de um único amigo. E também sabemos que nas colunas teremos cada tipo ou parte de informação. Para encontrar informações estruturadas dispomos de linguagens de consulta tipo SQL (Structured Query Language), QBE (Query By Example) e linguagens específicas para documentos tipo XML (XPath e XQuery).

Para encontrar documentos por características do documento, existem os sistemas de GED (Gestão Eletrônica de Documentos). Eles servem para armazenar e recuperar documentos, utilizando bancos de dados. A recuperação é feita por atributos do documento. Exemplo: encontre o contrato feito com a empresa XYZ no ano de 2002. A catalogação consiste em incluir cada documento eletrônico associado a seus atributos. Isto significa que antes de colocar o GED em funcionamento é necessário definir que atributos serão armazenados. Isto pode incluir data do documento, número de processo, número de cliente ou dono do documento, valor, tipo do documento, origem, autor, setor, etc.

Em se tratando de informações textuais, existe a área de Recuperação de Informações (RI - Information Retrieval). O objetivo desta recuperação é encontrar documentos textuais com conteúdo de informação que seja relevante a uma necessidade de informação de algum usuário. Um sistema de RI é responsável pelo processo completo de analisar documentos,

armazenar suas representações e permitir a recuperação de documentos ou partes que satisfaçam a consulta (que representa uma necessidade de informação).

As ferramentas de RI geralmente trabalham com técnicas de indexação, capazes de indicar e acessar mais rapidamente documentos textuais. Existem três tipos de indexação, conforme Yates (1996):

* indexação tradicional: é aquela onde uma pessoa determina os termos descritivos ou caracterizadores dos documentos, os quais farão parte do índice de busca;
* indexação full-text (ou indexação do texto todo): onde todos os termos que compõem o documento fazem parte do índice; e
* indexação por tags (por partes do texto): onde apenas algumas partes do texto são escolhidas, automaticamente, para gerar as entradas no índice (somente aquelas consideradas mais importantes ou mais caracterizadoras).

Índices do tipo thesaurus são tipos específicos de dicionários e seguem a técnica do vocabulário controlado. Segundo Fosket (1997), thesaurus é um dispositivo de controle de termos usado na representação de documentos. Um thesaurus contém termos relacionados semântica e genericamente, baseados na Linguagem Natural, mas num conjunto restrito. As relações podem definir hierarquias entre conceitos (divisão de conceitos em menores). A vantagem do uso deste tipo de indexação é a consistência através de vocabulário padrão, o controle de sinônimos e a facilidade de localização de novos conceitos.

Uma das etapas do processo de RI é a busca por documentos relevantes à consulta, através da comparação entre a representação da consulta e as representações dos documentos da coleção. Os documentos resultantes são apresentações segundo ordenações ou rankings, que permitem distinguir graus diferentes de relevância, mesmo para documentos da mesma classe.

O Google é o serviço de Internet mais utilizado hoje em dia justamente porque as pessoas precisam encontrar informações na Internet. Mas esta é caótica, sem padrões e possui um volume muito grande de dados. Google utiliza a indexação full text e ainda usa critérios para montar o ranking final (técnicas de SEO descrevem estes critérios).

Em se tratando de imagens e sons, as técnicas de recuperação são um pouco diferentes. Existem algoritmos que tentam identificar o conteúdo de uma imagem de forma automática. Como já existem tecnologias para identificar rostos (Google e Facebook permitem marcar rostos de pessoas nas fotos), o desafio atual dos algoritmos de recuperação de imagens é identificar partes do corpo humano, animais, gestos, sorrisos, etc. A recuperação de imagens estáticas baseada em conteúdo prevê a indexação automática da imagem por suas propriedades (regiões, cores, formas, brilho, arestas, texturas, objetos) e a recuperação por estas propriedades ou através de relacionamentos entre os objetos (estrutura da imagem).

Já a recuperação de áudio pode ser feita de modo digital (analisando padrões nas representações por bits), por conteúdo (padrões semânticos nos sons) ou por análise textual (transformando as falas em textos escritos e aplicando técnicas de RI). Para armazenamento

de áudio é necessária a transformação de um sinal analógico em dados digitais, seguindo duas etapas:

a) amostragem: representação de sinais temporais contínuos em números reais, sendo a taxa de captura dada pelo número de bits por amostra;

b) quantização: representação dos números reais em um conjunto finito, resultando em uma distorção controlada.

Em geral, utiliza-se um espectrograma para representar sons, especialmente de voz. O espectrograma é um gráfico que mostra a intensidade por meio do escurecimento ou coloração do traçado, as faixas de frequência no eixo vertical e o tempo no eixo horizontal, demonstrando visualmente as características acústicas da emissão. Tecnologias para reconhecimento da fala (speech recognition) fazem a recuperação dos áudios por fonemas ou palavras identificados nos sons. O reconhecimento da fala é feito pela análise do espectro de áudio, dependente do tempo, fazendo o mapeamento para possíveis unidades lógicas (fonemas, palavras, sub-palavras). Para tanto, utiliza-se uma máquina de reconhecimento de padrões (pattern matching), treinada para cada instância (geralmente, fonemas).

Algumas destas técnicas são apresentadas especificamente na parte II, quando discutiremos as variedades de tipos de dados (ex. busca por imagens, sons, mapas, etc.).

Para maiores detalhes sobre armazenamento e recuperação de diferentes tipos de dados, ver Loh (2018).

Sistemas de recomendação

Quando há muitas opções, fica impossível analisar todas para tomar uma boa decisão. Por isto, existem sistemas de recomendação para auxiliar as pessoas na escolha entre muitas alternativas. O objetivo é indicar automaticamente o que mais convém a uma pessoa, seja baseado na média geral de outras pessoas (sabedoria das massas), seja baseado no perfil do indivíduo (seus interesses explícitos ou inferidos pelo seu histórico de ações).

Os itens possíveis de indicação incluem livros, produtos, discos, restaurantes, lugares turísticos, meios de transporte, páginas web, serviços de empresas ou simplesmente informações.

As técnicas baseadas na individualidade montam um perfil do usuário a partir da coleta explícita ou implícita de informações sobre os clientes e também pode fazer uso de inferências. O perfil deve ser montado a partir do histórico de relacionamento do cliente com a empresa. Algumas técnicas ainda fazem uso da sabedoria das massas ou de informações vindas de pessoas similares. Assim, a origem da recomendação pode ser um indivíduo, como por exemplo, especialistas, profissionais certificados, autoridades, usuários que utilizam bastante o serviço (heavy users), pessoas próximas ou semelhantes (com mesma função, de mesma área científica ou área demográfica, mesma experiência, características ou comportamentos semelhantes). Mas as recomendações também podem vir

de grupos de pessoas, tais como comunidades ou simplesmente por maioria (ex: livros mais vendidos).

Este post traz um resumo da evolução dos sistemas de recomendação.
http://miningtext.blogspot.com/2018/12/evolucao-dos-sistemas-de-recomendacao.html

Volume de dados precisa análise

O Big Data por si só não será útil se não conseguirmos extrair conhecimentos deste conjunto. É por isto que a análise de dados é muito importante. Através de análises, podemos entender o passado, identificar e construir modelos, fazer classificações, reconhecer padrões e por fim prever o futuro.

Com dados e análise, podemos identificar quais são os melhores métodos de gestão organizacional, quais os melhores tratamentos para curar doenças, como tratar melhores clientes, como fazer melhores produtos, etc.

Isto tudo não é novidade. A identificação de padrões é parte da nossa vida desde milhares de anos atrás. Nossos antepassados conseguiam prever as variações do tempo, as estações, os ciclos das plantações, as fases lunares e os eclipses, e até mesmo o surgimento de reis.

Segundo Shermer (2011), "evoluímos para nos tornarmos habilidosos, buscadores de padrões, criaturas que descobrem causas. Aqueles que eram melhores na descoberta de padrões (ficar a favor do vento é ruim para caçar animais, o esterco de vaca faz bem aos cultivos) produziam a maior descendência. Somos os seus descendentes."

E hoje em dia não é diferente. Quem não dá palpites sobre como será o tempo, se vai chover, fazer sol, calor, observando as nuvens ? Ou se o próximo inverno será mais frio ou menos frio do que o ano anterior, pelo que viu no outono ? Se um local público vai lotar ou não para um evento, observando o movimento das pessoas chegando ? Ou quantas pessoas há num concerto ao ar livre num parque público, lembrando o último evento que ocorreu ali ? Mesmo algumas superstições são exemplos de padrões, que acreditamos que irão se repetir. Numa entrevista de negócios, usar a mesma roupa de um acontecimento bom. Sentar no mesmo lugar do último título para torcer por seu time. Não quebrar espelho, pois quando isto ocorreu, um evento de má sorte também ocorreu junto.

A classificação é um instinto do ser humano. Tentamos colocar tudo em grupos (pessoas, produtos, eventos, animais, plantas, etc.). Mesmo num texto como este, as informações estão agrupadas. Acreditamos que podemos reduzir tudo a um modelo único ou a poucas regras. Esta é a busca eterna dos físicos, para entender a Natureza e o Universo. Einstein acreditava que há uma ordem na desordem, mas que os padrões ainda devem ser descobertos.

A classificação facilita nosso entendimento do mundo e agiliza nossa tomada de decisão. Os padrões servem para minimizar a incerteza. Se encontramos uma situação nova e verificamos que ela se encaixa num padrão já entendido, já sabemos que atitudes tomar naquela situação. Este é um dos conceitos de inteligência: saber adaptar-se a novas situações e conseguir resolver problemas novos. Isto não significa que vamos usar exatamente as mesmas ações. A inteligência humana pressupõe a adaptação dos padrões para novas realidades.

A melhor forma de entender um conjunto de dados é estabelecer um modelo para ele. O modelo explicaria as características comuns aos dados, as relações entre os dados, as relações de causalidade e influência ao longo do tempo. O ser humano busca padrões no seu contexto porque se sentirá mais parte do contexto e menos um alienígena. É como uma necessidade humana, para não ficarmos loucos. O que não se encaixa nos nossos padrões, como por exemplo eventos paranormais, acabamos considerando como bruxarias.

A finalidade dos modelos é permitir o entendimento de um conjunto de eventos, poder comunicar a outros, poder reproduzir este comportamento. Os modelos são construídos a partir de experiências passadas, de registros de casos que já aconteceram, com suas características descritas (o que, quando, onde, por que, com quem e como aconteceram - os 5W e 1H). Sem registros históricos não há como identificar padrões e daí montar modelos.

Mas os modelos estão também associados ao futuro. Eles nos servem para direcionar nossas decisões e ações. Por isto, usamos muitas vezes o termo "modelo de predição", porque usando modelos podemos "prever" o futuro (ou tentar, pelo menos). Modelos são utilizados para previsão do tempo, para previsão de colheitas, de níveis de vendas, de quebra de máquinas, possibilidade de voto numa eleição, de um cliente pagar ou não um empréstimo, e para outros tantos fins como veremos neste livro.

Softwares de data mining e Business Intelligence ajudam pessoas a encontrarem padrões nos dados e a visualizarem informações de forma a entender melhor históricos e tendências.

Há software também para previsões, para ajudar a predizer valores para clima, vendas e eventos. Conforme Ismail, Malone e Geest (2015), "graças ao aumento de dados e ao poder de computação, as previsões dos meteorologistas ficaram 50% mais precisas..."

O meu livro "BI na era do big data para cientistas de dados - indo além de cubos e dashboards na busca pelos porquês, explicações e padrões" (Loh, 2014) apresenta e discute, em detalhes, técnicas e ferramentas para análise de dados, incluindo Data Mining e Business Intelligence.

Volume de dados precisa sínteses e resumos

Após os esforços para encontrar informações em grandes quantidades, uma outra forma de análise é a geração de sínteses ou resumos dos dados.

Para dados estruturados, podemos gerar gráficos. Hoje temos uma variedade de recursos gráficos para apresentar informações, tais como gráficos em barras, linhas, pizza, coordenadas paralelas, redes e hierarquias, 3D, mapas, etc. Estas representações ou resumos gráficos compõem o que hoje se chama de "dashboard" (painel de informações).

A área que pesquisa como apresentar dados em imagens é a "visualização de informações". Quem tiver maior interesse por como construir gráficos pode ler a seguinte referência:
BERTIN, Jacques. Semiology of Graphics: Diagrams, Networks, Maps. University of Wisconsin Press, 1983.

Entretanto, muitas informações estão em formato não estruturado, incluindo textos, imagens, vídeos e sons (como será discutido no capítulo sobre diversidade de dados).

Para textos, podemos gerar resumos ou sumários. Sumário ou resumo é uma síntese do conteúdo. Podem ser criados manualmente por especialistas ou pelo próprio autor ou fonte das informações. O interessante seria extrair o resumo automaticamente dos dados.

Um resumo pode conter os pontos ou tópicos principais ou então apresentar os pontos periféricos. Por exemplo, um resumo com tópicos principais sobre uma doença pode falar de sintomas, sinais, causas e tratamentos. Um resumo com pontos periféricos sobre o mesmo assunto falaria de reações adversas, doenças associadas e interações com outros tipos de tratamentos. Saggion e Carvalho (1995) distinguem sumários informativos de sumários indicativos. Os do primeiro tipo representam o conteúdo do documento de forma sucinta, enquanto que os do último tipo apresentam somente os tópicos abordados.

Em se tratando de textos, um resumo pode conter os títulos e subtítulos, as frases que mais aparecem, as linhas iniciais de parágrafos e seções.

Outra forma de apresentar um resumo de um texto é através de uma nuvem de palavras (tag cloud). Este tipo de representação mostra as palavras de um texto, sendo que as mais frequentes aparecem em tamanho maior (como pode ser visto na Figura 9). Uma ferramenta par gerar tag clouds a partir de textos pode ser encontrada em http://www.wordle.net/.

Figura 9: Exemplo de Tag Cloud (nuvem de palavras chave)

Atualmente, sentimentos estão sendo apresentados através de emoticons e emojis ao invés de palavras. Se continuar assim, como disse José Saramago, “de degrau em degrau, vamos descendo até o grunhido”.

Seleção de Amostras

É muito difícil analisar todos os dados coletados e armazenados em um Big Data. Em geral, é preciso fazer uma seleção inicial. Isto porque alguns dados, mesmo presentes por direito na base, talvez não sirvam os propósitos. Por exemplo, produtos que não são mais vendidos e não interferem mais no processo, e por consequente não são interessantes para serem analisados, devem ser excluídos.

O primeiro passo então é selecionar um conjunto de dados (uma amostra) sobre os quais serão aplicadas as técnicas de análise ou mineração.

Existem 4 tipos de técnicas de seleção de amostras. Discutiremos elas através de um exemplo: uma loja querendo analisar a satisfação de seus clientes. Também discutiremos duas situações possíveis: a loja já ter um cadastro de clientes e o caso de a loja não

conhecer seus clientes (porque entram e saem da loja sem mesmo a loja saber se são homens ou mulheres).

- Amostras aleatórias

Neste caso, são selecionados aleatoriamente elementos do universo (conjunto todo). Por exemplo, a loja determina o tamanho da amostra (valor N) e a seleção é feita pegando-se os N primeiros clientes da base de dados (do cadastro) ou são selecionados N elementos dentro do cadastro, pulando de forma aleatória. Se a loja não tiver um cadastro, ela irá selecionar clientes que saiam da loja com sacolas (produtos comprados), "atacando" N clientes pulando alguns (a critério da pessoa que fará a abordagem).
Este tipo de amostra pode trazer problemas, pois imagine que os N selecionados são todos do mesmo tipo (homens X mulheres, classe A ou classe C, etc). E pior ainda se forem selecionadas justamente as exceções.
Alguns pesquisadores julgam a técnica eficiente pois acreditam na distribuição aleatória (aquela velha história da moedinha, se jogarmos uma moeda 1000 vezes e só analisarmos os 100 primeiros resultados, a distribuição será a mesma).
Entretanto, para que a técnica seja utilizada adequadamente, a aleatoriedade deve ser total. No caso de clientes saindo da loja, não se pode selecionar clientes apenas num dia. Deve-se levar em conta as variedades (dia da semana, dia do mês, mês, turno, etc).
Esta técnica só deve ser usada quando não se pode utilizar uma técnica melhor.

- Amostras por conveniência

Neste caso, a seleção é feita pelo que for mais fácil. Por exemplo, a loja seleciona os N primeiros clientes que saírem da loja num determinado dia ou liga para N clientes cadastrados que tiverem telefone e só utiliza dados dos N primeiros que atenderem o telefone.
É a pior técnica pois não há critério algum, nem mesmo a aleatoriedade, o que pode levar a tendências (selecionar somente elementos de um tipo).
Esta técnica só deve ser usada quando não se pode utilizar uma técnica melhor.

- Amostras por julgamento

As amostras por julgamento são formadas por elementos que satisfaçam regras previamente determinadas. Por exemplo, analisar somente a satisfação de clientes mulheres que compraram mais de um produto até uma semana após o Dia das Mães.
Neste caso, o critério de seleção está bem definido e é justificado (por exemplo, só querer analisar certos tipos de elementos do conjunto todo). E portanto os resultados da análise serão condizentes somente com as regras definidas (não valem para o universo todo).
Podem ser utilizada regras de seleção ou de exclusão. O segundo caso pode ser melhor para se ter uma visão melhor do todo. Por exemplo, a loja pode querer analisar todos os tipos de clientes, mas vai excluir quem só veio uma vez por ano ou quem comprou num valor muito abaixo da média de gasto.

- Amostras estratificadas

Esta é a forma correta de gerar amostras. Para tanto, precisa-se identificar que variáveis podem interferir na análise. Por exemplo, no caso da loja, atributos como sexo, idade, classe sócio-econômica, bairro e cidade, valor gasto e forma de pagamento podem fazer

diferença para entender os tipos de clientes. E talvez altura, peso e escolaridade não sejam diferenciais para campanhas de marketing ou para entender comportamentos de compra.
Depois de identificadas as variáveis, precisa-se saber a proporção de elementos no universo todo para cada variável. Por exemplo, digamos que há 60% de mulheres e 40% de homens entre todos os clientes da loja, e que 25% são da classe A, 50% da classe B e 25% da classe C, e assim por diante nas demais variáveis.
Então, a amostra será definida com a mesma proporção que a divisão no universo. Ou seja, a amostra deve conter 60% de mulheres, 40% de homens, 25% de pessoas da classe A, 50% de pessoas da classe B, 25% da classe C e assim por diante.

Como determinar o número ideal de elementos numa amostra ? Se olharmos para as pesquisas para presidente do Brasil, a amostra normalmente é composta por aproximadamente 2 mil pessoas. Isto quer dizer que cada pessoa representa em torno de 50 mil outras.
O cálculo estatístico do tamanho da amostra depende do erro amostral (a diferença entre o valor estimado pela pesquisa e o verdadeiro valor e isto pode ser um valor estabelecido como meta); do nível de confiança (a probabilidade de que o erro amostral efetivo seja menor do que o erro amostral admitido pela pesquisa); da população (número de elementos existentes no universo da pesquisa, valor que pode não ser conhecido); entre outros (percentuais máximo e mínimo). Há uma calculadora online para fazer tais cálculos: http://www.calculoamostral.vai.la/

Tversky e Kahneman (1971) discutem os problemas com amostras muito pequenas. Por exemplo, se você jogar uma moeda não viciada três vezes e der duas vezes cara e uma vez coroa, você estará inclinado a acreditar que a probabilidade é 66,66% contra 33,33%. Mas se jogar mil vezes a mesma moeda, certamente haverá uma proporção próxima de 50/50. Pior seria se nas três primeiras jogadas, desse somente um lado. Como sabemos que, no caso da moeda, a probabilidade é 50/50, isto pode gerar a chamada "falácia do jogador": acreditar que o jogo vai mudar para reverter uma tendência e voltar ao padrão estatístico. Por exemplo, jogando 5 vezes a mesma moeda e dando sempre o mesmo lado (digamos, cara), vamos acreditar que na 6a vez irá dar o outro lado (coroa). E na 7a também vamos estar inclinados que dará coroa para equilibrar o jogo e voltar à proporção 50/50. Entretanto, a proporção só acontece com amostrar maiores. Então, as próximas jogadas só minimizam os desvios e não os corrigem logo em seguida.

PARTE II – VELOCIDADE DOS DADOS

Desde que a humanidade descobriu que há padrões temporais na agricultura (tempo de cultivo e colheita) e na pecuária (tempo de acasalamento e nascimento de bebês), estamos estudando o tempo. Até descobrir que o tempo é relativo, segundo Einstein. Mas o grande objetivo foi e sempre será controlar o tempo.

Há milhares de anos atrás, nossos ancestrais desenvolveram cerimônias e rituais para tentar vencer a morte. Foi por isto que as pirâmides do Egito foram construídas. E até hoje há certos tipos de sepultamentos e técnicas de clonagem e criogenia para que um dia possamos ressuscitar e viver eternamente.

Na Grécia antiga, as pessoas queriam aproveitar da melhor maneira possível o tempo. O trabalho era perda de tempo. O tempo era usado para pensar e criar coisas novas. Hoje o trabalho acontece o tempo todo. E mesmo quem não precisa trabalhar para ganhar dinheiro e viver não está parado, ou seja, está usando o tempo para fazer algo. Domenico de Masi falou do "ócio criativo". Temos que aproveitar o tempo para trabalhar, aprender e nos divertir. Tudo isto ao mesmo tempo.

Os sumérios começaram a medir o tempo nos dados, por isto temos um sistema baseado no número 60 (12 x 5). Foram eles também que criaram a rotina diária, para aproveitar melhor a luz solar. Havia tempo para comer, trabalhar, descansar, etc.

Hoje, o tempo virou o novo capital do pós modernismo. O tempo é um novo tipo de moeda. "Tempo é dinheiro". Crianças correm para estudar e gerar riqueza para seu país. Empreendedores correm para desenvolver inovações antes que os concorrentes. Empresas correm para lançar novos produtos e serviços. Marketeiros correm para alcançar clientes e vender. Pessoas vendem seu tempo por salários em empregos ou em serviços de consultoria.

Mas tempo também é saúde e sabedoria. A meditação está crescendo porque ajuda pessoas a controlarem ansiedade, traumas, emoções indesejadas, etc. O tempo nos dá sabedoria e tempo para tomar melhores decisões. Mas o tempo também "escorre pelas mãos". O que foi não volta.

Velocidade da Globalização

Conforme Stiglitz (2002), a Globalização é definida pela remoção de barreiras ao livre comércio e pela integração das economias nacionais. Seus benefícios incluem:
- crescimento das exportações (gera riqueza);
- mais empregos em regiões pobres (fábricas de países ricos em países pobres);

- compartilhamento de conhecimento;
- projetos em saúde, educação, irrigação, etc.;
- investimentos estrangeiros;
- redução de custos de transporte e comunicação;
- instituições transnacionais (ONU, OMS).

Frieden (2008) explica a Globalização através do exemplo da boneca Barbie: "sua produção começou com os moldes que a Mattel, dos Estados Unidos, colocou à disposição de fábricas no sudeste da Ásia. Taiwan e Japão forneciam o plástico e o cabelo. Empresas chinesas supriam o tecido de algodão para os vestidos. As bonecas eram montadas na Indonésia, Malásia e China e depois despachadas para Hong Kong, e de lá para os consumidores da Mattel nos EUA e em vários outros países."

Atualmente vivemos tão dependentes uns dos outros que produzir sapatos com baixo custo na China pode eliminar empregos em Sapiranga no Rio Grande do Sul e afetar um pequeno comerciante nesta cidade. Uma crise de petróleo no Oriente Médio pode afetar a produção de comida na América do Sul, já que muitos fertilizantes, que melhoram em muito a produtividade na produção de alimentos, são feitos de petróleo. As crises antes nacionais agora são mundiais. Podem virar marolas ou tsunamis no outro lado do mundo.

Com as modernas tecnologias da informação e comunicação, os laços estão mais curtos. Ou seja, os dados trafegam mais rapidamente, o que faz com que as influências planetárias sejam quase instantâneas. Por isto, as bolsas reagem instantaneamente de um mercado para outro, de um lado do globo terrestre a outro. E por isto também informações influenciam pessoas de um país a outro, como ocorreu com a Primavera Árabe, quando manifestações no Egito se propagaram por outros países árabes.

Alguém lembra que 40 milhões de livros foram impressos 50 anos após a invenção da prensa por Gutenberg. E isto acelerou a velocidade para disseminar argumentos e convencer pessoas, tanto que temos como uma grande consequência disto a Reforma Luterana (Burke, 2002).

Velocidade das Inovações

A Lei de Moore trata da evolução exponencial da capacidade de processamento dos computadores. Mas as inovações surgem rapidamente em todas as áreas. A Figura 10 mostra (na linha vermelha) o momento em que um produto novo é lançado para substituir um anterior. Pode-se notar que é o momento em que o produto anterior está no auge. Isto permite à empresa continuar com receitas crescentes. Esta velocidade de inovações pode ser vista quando vamos ao supermercado. Sempre há produtos novos. E é difícil que um produto permaneça por muito tempo; no mínimo, haverá mudança de embalagem ou da "fórmula".

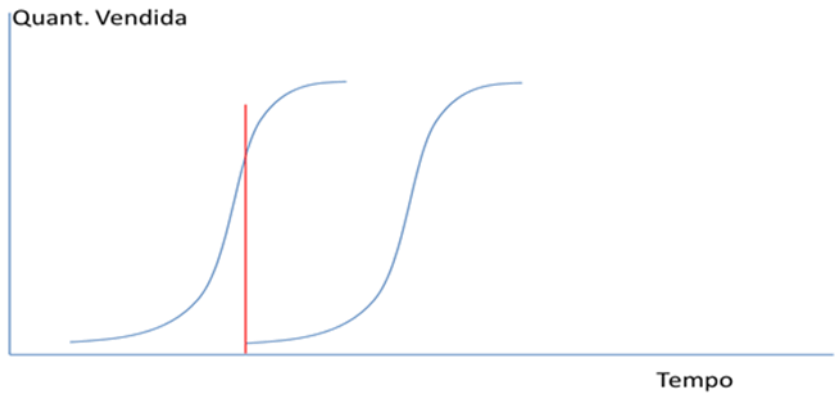


Figura 10: Momento para lançamento de um produto substituto

Conforme Ismail, Malone e Geest (2015) nos lembram, "muitos produtos são lançados mais cedo – inacabados e em beta perpétuo – com a única finalidade de coletar dados dos usuários o quanto antes para estabelecer como 'terminar' o produto. Os dados coletados a partir desses primeiros usuários são rapidamente analisados para obter ideias sobre bugs que precisam ser resolvidos e sobre os recursos mais desejados pelos usuários. Uma vez que as mudanças são implementadas, o produto é relançado e analisado... e o processo continua."

Atualidade da informação x informação obsoleta

Uma das qualidades da informação é a atualidade. Mas a velocidade com que a informação deve ser atualizada depende da necessidade.

Se você demorar um mês para fazer uma pesquisa de preço, no final do processo, os preços coletados no início já não serão mais os mesmos. Isto não significa que a informação precisa ser atualizada de imediato; a regularidade de atualização ou o quanto uma informação deve estar atualizada depende do processo e da necessidade para a decisão. Por exemplo, a temperatura de uma caldeira deve ser controlada em intervalos de minutos, porque, se ela subir muito, a caldeira poderá explodir. Então, é preciso saber logo se a temperatura está subindo muito e em que velocidade, para tomar decisões o quanto antes.

Por outro lado, empresas multinacionais reúnem dados de vendas ao final do mês, especialmente se querem avaliar tendências de vendas (subida ou descida) de seus produtos. Então, não é necessário atualizar a cada dia ou semana o banco de dados centralizado.

Outra dificuldade é manter histórico das mudanças ou evoluções dos dados. Por exemplo, em sistemas médicos, é importante saber os indicadores de saúde de uma pessoa ao longo de sua vida. Isto permite analisar a evolução e também saber qual a atualidade do dado (uma espécie de selo de tempo – timestamp – para cada valor). Além disto, é importante

para processos de rastreabilidade, para saber como as alterações foram feitas (e por quem), o que permite ainda "voltar atrás" em um valor. Por isso, tecnologias como blockchain estão sendo muito utilizadas pois são propícias para isto (rastreabilidade e segurança das alterações).

Talvez a grande dificuldade seja saber quando uma informação está obsoleta. Primeiro, é preciso saber se a informação mudou ou não. Isso significa monitorar regularmente a informação, ou seja, coletar novamente o dado de tempos em tempos. Segundo, se estamos falando de documentação numa empresa que segue normas de qualidade, é preciso avisar as pessoas envolvidas que houve mudança na informação e não permitir que diferentes pessoas ou setores utilizem versões diferentes do mesmo dado. E isto se aplica também a novas versões de métodos ou conjuntos de dados. Agora, imagine a dificuldade de monitorar se surgiu algum método novo que substitua o método anterior, agora ultrapassado. Thomas Kuhn (2011) discute as mudanças de paradigmas em teorias científicas.

Paradigmas de qualidade exigem o controle de versões de documentos, evitando que documentos obsoletos fiquem disponíveis e que os documentos com a versão atual fiquem indisponíveis. Também é necessário criar uma Tabela de Temporalidade de Documentos: é o instrumento aprovado por autoridade competente que determina prazos de guarda e destinação final dos documentos produzidos (recolhimento ou eliminação).

Outro problema é utilizar a informação na velocidade certa. Jack Welch, antigo CEO da GE, dizia que vantagem competitiva se consegue com:
a) aprender mais rápido que a concorrência;
b) transformar mais rápido este conhecimento em ações.
Por exemplo, uma empresa que descubra que um cliente está querendo comprar um determinado produto, deve usar esta informação o mais rápido possível para tentar concretizar esta venda. Se demorar, alguma outra empresa pode fazê-lo antes. E aí, aquela informação não será mais útil.

Velocidade para Encontrar ou Buscar Informações

Com o aumento do volume de dados, faz-se necessário ter mecanismos para encontrar informações de maneira mais rápida. Para isto podem ser utilizados índices (para dados estruturados, dados textuais ou multimídia) e também taxonomias e tags (marcadores ou palavras chave) para organizar informações.

Em Loh (2018), são discutidos diferentes modos de recuperação de informações.

O tempo de busca por uma informação pode ser a diferença entre vida e morte. Há sistemas críticos que precisam de informações atualizadas em milissegundos. Um astronauta em uma espaçonave que saia da rota pré-determinada não pode esperar que cálculos para uma nova rota demorem para ser feitos. Um policial precisa também de dados imediatos. Operadores

em bolsas de valores também precisam encontrar informações de maneira rápida. Como já discutido antes, resumos e índices podem ser muito úteis nestes momentos, bem como as tecnologias de informação.

Velocidade dos Boatos

Allport e Postman concluíram que a intensidade (ou força) de um boato é diretamente proporcional à importância do tema e à ambiguidade de evidência ou certeza. Então, um boato "corre" mais rápido se não há informações para negá-lo e se é um tema importante ou preocupante para as pessoas envolvidas.

O rumor se espalha porque serve para (a) explicar eventos mal entendidos e (b) liberar tensões emocionais sentidas pelos indivíduos.

Além disto, há dispositivos fortalecedores dos boatos tais como a crença no boato por parte de quem repassa e detalhes que podem indicar sua veracidade (dados, estatísticas, relatos, imagens).

Ellwardt (2011) e Foster (2004) lembram que boatos e fofocas são instrumentos para integrar pessoas em um grupo ou para fortalecimento de laços internos (aumenta a coesão).

Memes

O termo "meme" foi cunhado por Richard Dawkins no seu bestseller "O Gene Egoísta" (2007). O termo é hoje utilizado amplamente nas redes sociais. Ele se refere à distribuição e replicação de informações pelas redes e é considerado um elemento chave no marketing viral.

Para Dawkins, o meme é para a memória o análogo do gene na genética. É uma unidade de informação que se multiplica de cérebro em cérebro, contagiando outras pessoas. O meme pode ser uma ideia ou parte de uma ideia, línguas, sons, desenhos, capacidades, valores estéticos e morais, ou qualquer outra coisa que possa ser aprendida facilmente e transmitida enquanto unidade autônoma. Até mesmo a roupa que vestimos, nossos hábitos, o que comemos, podem ser considerados memes, conforme Susan Blackmore (*apud* MORTON).

O interessante nos memes, é que há um ganho de tempo já que não se precisa muito esforço para aprender, é algo mais natural. Os memes devem ser considerados como estruturas vivas; quando um meme fértil infecta uma mente, parasita o cérebro, transformando-o num veículo para a propagação do meme, exatamente como um vírus pode parasitar o mecanismo genético de uma célula hospedeira.

A principal característica do meme é que ele é uma informação copiada como num processo evolucionário, com variação e seleção. E o meme não reside somente na mente humana. Por exemplo, a dança das abelhas para indicar a localização de comida é replicada com alguma modificação (não é cópia fiel). O mesmo acontece com outros animais para avisar de predadores (Blackmore, 2001).

A replicação normalmente acontece por imitação. Entretanto, nem sempre é uma cópia fiel que se propaga, pois pode haver mutações (como na brincadeira do telefone sem fio). Alguns memes são mais bem sucedidos na replicação do que outros, como na seleção natural de Darwin.

Tecnologias para Velocidade na Comunicação – o Imediatismo

Antigamente, havia corredores físicos para transporte de materiais e informações. Rios, estradas e comerciantes foram o meio de comunicação. Com o modernismo, surgiram espaços estáticos em cidades, como os salões do iluminismo e os cafés parisienses (Johnson, 2011). Estes espaços reuniam intelectuais e assim facilitaram a troca de informações e o aumento da criatividade, como bem contado no filme de Woody Allen "Meia noite em Paris". Por isto, empresas hoje em dia valorizam a happy hour e as salas de convivência. E por isto também, Marissa Meyer, quando era CEO do Yahoo, proibiu o tele-trabalho, já que os funcionários não se encontravam mais presencialmente e isto diminuiu a geração de ideias.

Atualmente, as tecnologias da informação são o meio para transferir dados mais rapidamente e entre mais pessoas, através de cabos, wifi, 3G/4G, satélite, rádio, etc. A consequência é que os dados podem ir direto da fonte de coleta para o consumidor (seja uma pessoa, organização ou sistema). E isto está alterando como organizações funcionam e como pessoas trabalham.

Antigamente, um cliente enviava uma carta com suas reclamações para a empresa. Depois veio o telefone, através de SAC e setores 0800. Com a Internet, os sites das empresas e os e-mails passaram a ser o canal. Mas ainda assim, a resposta não era imediata. Hoje, com Whatsapp e aplicativos tipo Messenger, a pessoa que envia uma mensagem espera a resposta em poucos minutos.

A consequência é uma enxurrada de mensagens, que pessoas e organizações não conseguem responder. Por isto, veja este artigo "Não fale conosco pelo fale conosco".
https://webinsider.com.br/nao-fale-conosco-pelo-fale-conosco-entende/

Dentro das empresas também houve mudança de cultura. Antes, alguém preenchia um formulário em papel para pedir algo a outro setor. Depois, vieram os formulários eletrônicos e e-mails, bem como software do tipo Workflow. Então a pessoa manda um pedido e logo em seguida liga para avisar e exigir a resposta.

Computação ubíqua e pervasiva – Internet das Coisas (IoT)

A **pervasividade** é a característica de a tecnologia estar presente em todo lugar, sem ser notada. Ela é invisível no sentido de que o usuário não precisa se dar conta que a tecnologia existe ou que está usando tecnologias. Ela pode estar embutida nos mais diversos dispositivos incluindo o celular, a própria roupa, qualquer acessório como o relógio ou óculos e até mesmo o nosso corpo. Já a **ubiquidade** (computação ubíqua) tem a ver com a possibilidade de termos acesso às informações, independente da nossa localização (*anytime anywhere*).

É possível experimentar tais fenômenos ouvindo músicas, vendo filmes ou jogando jogos. Antes, era necessário comprar o conteúdo, recebê-lo num meio físico, instalar ou copiar e então acessá-lo num dispositivo próprio (DVD, CD player, TV, vídeo game, etc). Hoje estas aplicações estão disponíveis em qualquer dispositivo, e podemos fazer tais ações onde estivermos (na rua, no ônibus, em casa, etc.). Não há mais distinção entre dia e noite, casa x trabalho x lazer, etc.

Além disto, tais conteúdos podem ser adquiridos por sistema "on demand", ou seja, pagos por uso. Podemos ter na hora e local que quisermos, sem precisar esperar a entrega. É tudo instantâneo. Até mesmo a produção de um produto pode acontecer em nossa casa com impressoras 3D. Por isto, a nova onda é "DIY" (Do It Yourself – faça você mesmo).

E também há mais dispositivos em rede disponíveis para uso próprio ou coletivo. Conforme Ismail, Malone e Geest (2015), "dez anos atrás, havia 500 milhões de dispositivos conectados à internet. Hoje, existem cerca de oito bilhões. Em 2020, haverá 50 bilhões e, uma década depois, teremos um trilhão de dispositivos conectados à internet à medida que, literalmente, todos os aspectos do mundo estiverem habilitados para informação com a internet das coisas (Internet of Things)... nós estaremos saltando de oito bilhões de dispositivos conectados à internet hoje para 50 bilhões até 2025, e para um trilhão apenas uma década mais tarde."

A Figura 11 apresenta um exemplo de pervasividade e ubiquidade na área médica ou hospitalar. Na figura podemos ver uma pessoa (pode ser um médico) num estádio recebendo no seu celular informações coletadas por sensores localizados junto a pacientes num hospital. Ou então, um médico num trem, usando seu tablet para acessar imagens de exames de pacientes. Estes dados podem estar armazenados nas nuvens e não no próprio prédio do hospital. Do outro lado, uma TV recebe vídeos através de sinais wireless. O celular do primeiro médico se comunica com o tablet do segundo e trocam informações via bluetooth. Abaixo, uma cena já mais antiga, um administrador do hospital acessando relatórios a partir de um computador na sua casa ou escritório.

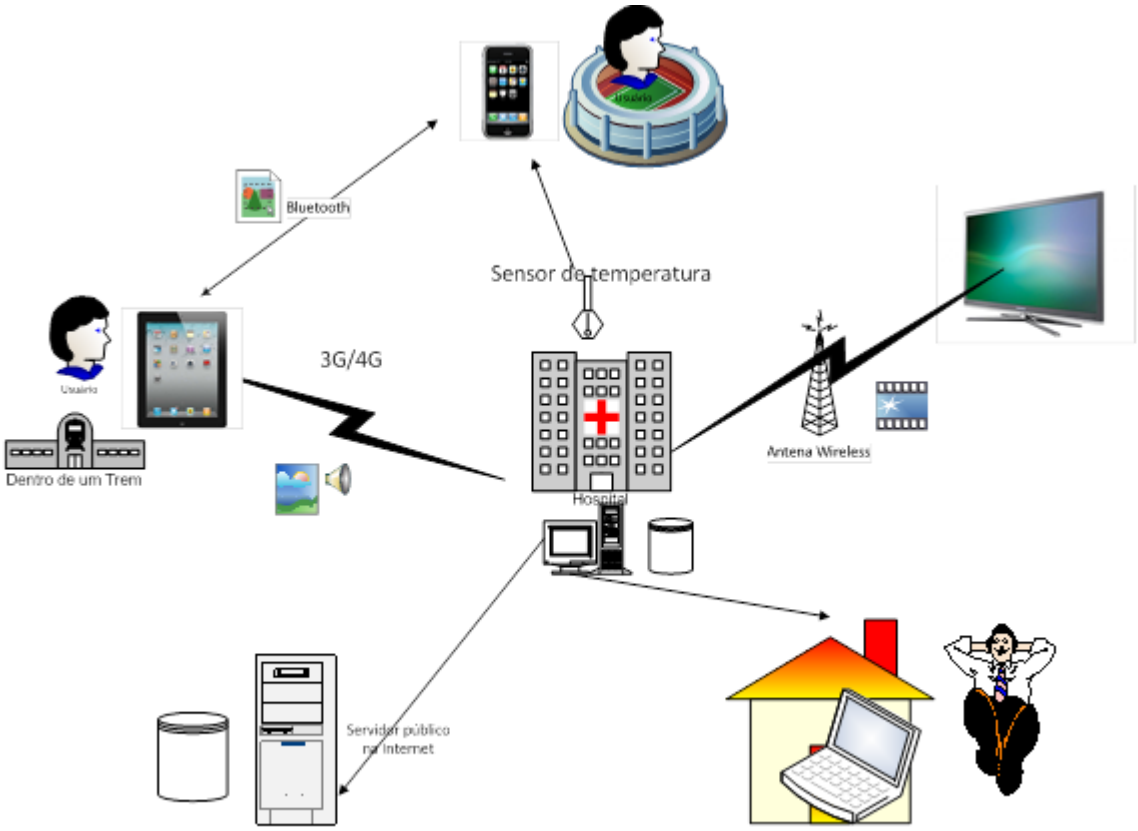


Figura 11: Esquema exemplo de sistemas pervasivos e ubíquos

A seguir, são discutidas algumas características dos sistemas pervasivos e ubíquos.

- heterogeneidade

Este tipo de sistema deve integrar dispositivos de todas as naturezas, sistemas operacionais dos mais variados tipos e tecnologias antigas e também as mais modernas.

- mobilidade

Esta é a característica principal. Hoje em dia as pessoas usam dispositivos móveis para tudo, em todo lugar e a qualquer hora. Com os diferentes tipos de conexões wireless (wifi, 3G/4G, bluetooth), tais dispositivos estão conectados 24/7 (24 horas por dia, 7 dias da semana).

- redes P2P (pier-to-pier)

Em contraposição às redes cliente-servidor, as redes P2P permitem que um mesmo dispositivo possa pedir uma informação (ser cliente) e em outro momento fornecer informações (ser servidor). No exemplo da figura, o tablet do médico no trem é cliente ao solicitar exames armazenados no hospital, mas também pode repassar tais dados ao celular do médico no estádio. Este última não precisa acessar o servidor do hospital ou talvez até não tenho conexão para isto; então por que não se pode pegar a informação do local mais fácil ?

- Near Field Communication

Tecnologias deste tipo (como bluetooth, por exemplo), permitem a troca de dados entre dispositivos que estejam próximos. Basta ligar a função, e um dispositivo pode encontrar outros. Após o pareamento entre eles, com uso de senhas para permissões, a troca de dados pode ser feito em ambos os sentidos.

- disponibilidade de informações

Os Administradores de Bancos de Dados (DBAs) devem preocupar-se com a tolerância a falhas. Os dados devem estar disponíveis 24/7 para qualquer tipo de dispositivo ou sistema operacional.

- troca de informações não padronizadas

Neste sentido, os dados podem estar armazenados em formatos e modelos diferentes, devido à heterogeneidade dos hardwares e softwares sendo utilizados. A troca de informações pode ser feita por arquivos XML, que são arquivos texto mas com marcas de significado. Estas marcas são evoluções das marcas HTML, que eram somente para definir a forma de apresentação das informações. As marcas XML definem do que se tratam os dados, para que possam ser entendidos pelos sistemas automáticos. Assim, se quisermos passar o nome de um paciente usaremos uma marca tipo <nome_do_paciente> José da Silva </nome_do_paciente>. Os dados também podem ser cambiados via Web Services. A troca aqui neste caso é mais complexa. Um Web Service, como o nome diz, permite que um sistema acesse serviços de outro sistema. Por exemplo, as APIs do Google Maps, para que outros sistemas podem recuperar mapas do Google, são exemplos de Web Services. Neste caso, há parâmetros de entrada e informações na forma de resultados. É portanto necessário conhecer tais parâmetros, além do endereço do Web Service.

- consciência de contexto

Este tipo de característica permite que os sistemas possam entender o contexto do usuário. Por contexto, entende-se: o tipo de dispositivo (hardware e software) do usuário, a configuração da conexão (tamanho da banda, tipo), o local (geográfico e semântico) em que se encontra o usuário, o horário local, as condições climáticas, etc. Por exemplo, o sistema descrito na Figura 11 pode reconhecer o tipo de celular utilizado pelo primeiro médico (incluindo hardware e software), identificar que ele está num estádio de futebol e que naquele momento já um jogo em andamento, reconhecer que há outros dispositivos próximos e os tipos de conexões disponíveis.

- adaptação

A característica de adaptação complementa a característica anterior. É importante reconhecer o contexto do usuário para fornecer a informação na medida certa. Assim, as imagens devem ser apresentadas no dispositivo destino conforme o tamanho de tela, resolução e número de cores. Se o ambiente local é barulhento (identificado por inferência ou por sensores), as mensagens importantes e os alertas para o usuário devem ser comunicadas de uma forma que ele ouça. Por outro lado, se ele estiver numa Igreja, talvez a forma de comunicação deva ser outra. Se um usuário pede para ser informado a cada início de dia, das principais notícias locais, o sistema deve reconhecer a posição geográfica e o

fuso horário onde a pessoa está, e também coletar as notícias dos sites de empresas jornalísticas deste local. E ainda deve saber se o usuário sabe ler na língua local, caso contrário deve traduzir as notícias. A adaptação ainda pode reconhecer o estado físico e emocional do usuário, incluindo temperatura, batimentos cardíacos, etc.

- sensores para coleta automática de dados

Hoje em dia a tecnologia evoluiu para que possamos ter diferentes tipos de sensores para coleta de dados de forma automática (sem intervenção humana). Há sensores de luminosidade, umidade, movimento e presença (se houve movimento num local ou se alguém está ali presente), GPS, sensor de nível, bússola, etc.

- agentes inteligentes

Quando um usuário solicita uma informação, ele não deve ter que dizer onde está a informação. É responsabilidade do sistema pervasivo e ubíquo localizar a informação e fazê-la chegar onde está o solicitante e no seu dispositivo. Para isto, deve haver softwares inteligentes que tenham capacidade para entender o que a pessoa quer e depois poder localizar tal informação na rede de dispositivos conectados.

- comunidades de agentes (sistemas multiagente)

Estes agentes inteligentes podem trocar informações sem que as pessoas precisem intervir ou mesmo ficar sabendo. Eles trocam informações para localizar respostas a questionamentos dos usuários que eles representam. Podem até mesmo repassar as solicitações para outros agentes, formando uma rede inteligente. Em casos mais evoluídos, podem até mesmo negociar entre si, dados parâmetros fornecidos pelos usuários. Deste forma, num futuro breve, quando uma pessoa quiser comprar algo, simplesmente deverá informar seu agente inteligente. Este irá comunicar-se com outros agentes à procura de alguém que tenha informado seu agente que deseja vender algo. Com algumas regras básicas, os agentes podem negociar preços e fechar a venda, sem intervenção humana. Para isto, serão necessários protocolos de cooperação e troca, e recursos inteligentes para negociação.

- conteúdos nas nuvens (cloud computing)

O armazenamento de dados nas nuvens libera empresas de terem que ter uma grande infraestrutura para armazenamento; basta computadores ou dispositivos móveis conectados à Internet. A questão de desempenho e segurança fica a cargo do fornecedor do serviço. E poderemos acessar os dados de qualquer lugar no planeta.

- Internet das coisas (Internet of Things)

Fabricantes já estão incorporando tecnologias para que cada máquina tenha um número IP que a identifique e recursos para conexões na Internet. Desta forma, um eletrodoméstico qualquer poderá ser acessado via Internet e reprogramado. Itens como TVs, geladeiras, ar-condicionados, rádios de automóveis, relógios, câmeras serão todos dotados destas características. E estarão também se conectando de forma inteligente, através de agentes inteligentes. Esta é a chamada Internet das Coisas.

Mais canais de comunicação e interação

O volume de dados também aumentou porque é possível se comunicar com mais pessoas em menos tempo. As tecnologias estão aproximando pessoas e diminuindo as distâncias. Veja o número de "amigos" ou "seguidores" que um usuário comum possui no Facebook, Twitter ou Instagram. E há tecnologias para teleconferência em tempo real (ex. Google talk, Skype).

Uma das causas deste fenômeno é que o número de canais de comunicação aumenta conforme a fórmula: n (n-1) / 2, onde n = número de pessoas. Podemos ver isto na Figura 12. Três elementos permitem 3 canais de comunicação. Quando se acrescenta 1 elemento a mais (total de 4 elementos), o número de canais já dobra (para 6 canais).

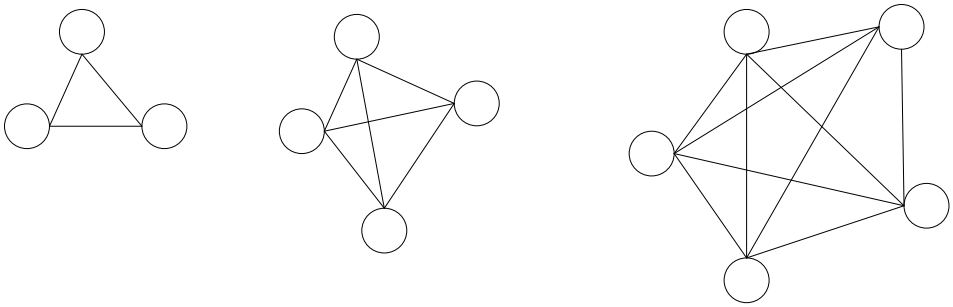


Figura 12: Número de canais de comunicação x número de elementos

Quando há muitos canais para fluxo da informação, ela não chega aonde devia chegar. Por isto, organizações com hierarquias muito altas (espaço vertical) e muitos departamentos e setores (espaço horizontal) acabam tendo problemas que são vistos em programas como "Chefe Espião" (em inglês, Undecover Boss). Não só problemas não ficam conhecidos por quem decide, mas também as ideias e inovações acabam se perdendo.

Por isto, Ismail, Malone e Geest (2015) falam das vantagens das startups com estruturas menos rígidas e funções mais próximas.

Velocidade de Cooperação

Mais canais de comunicação, maior a cooperação e a geração de ideia, soluções e inovações. Segundo Koestler (1964) e Johnson (2011), 2 elementos principais são necessários para a geração de ideias: maturação de ideias e associação de planos ou contextos diferentes (procure no Youtube o vídeo "De onde vem as boas ideias" com animação e legendas, que é um resumo do livro de Johnson). E isto normalmente se consegue com grupos de pessoas ou pesquisadores.

Estudos apresentados em Sawyer (2006) concluíram que os grandes resultados advêm de pequenas contribuições, como na construção de um muro onde cada indivíduo coloca um tijolo. Então, os agraciados com o prêmio Nobel seriam os que colocaram o último tijolo. Isto não significa que todas as contribuições são consideradas.

Para Simonton (2002), não existe gênio individual na criatividade científica. Newton já dizia que tinha tido sucesso por estar sobre ombros de gigantes. A Comunidade de pesquisa (empresas e pessoas) faz a seleção e variação de ideias, podem gerar, combinar, recombinar, criar, integrar, eliminar, refazer, etc. diferentes ideias.

Diamandis e Kotler (2012) lembram que a vida multicelular aconteceu porque "as células começaram a se especializar, e essas células especializadas aprenderam a cooperar de um modo extraordinário... Um tipo de célula lidava com a locomoção, enquanto outro desenvolvia a capacidade de sentir gradientes químicos." Exemplos recentes e bem sucedidos de cooperação incluem o sistema operacional Linux e a Wikipédia. "A Wikipedia levou 100 milhões de horas de tempo de voluntários para ser criada", conforme Shirky (citado por Diamandis e Kotler, 2012). "Se abríssemos mão de nosso vício da TV por apenas um ano, o mundo teria mais de um trilhão de horas de excedente cognitivo para dedicar aos projetos compartilhados."

Alex Pentland (2014), no livro "Social Physics", apresenta resultados de experimentos reais provando que uma boa rede de pessoas e interação pode aumentar inovação, empreendedorismo e retorno no investimento. Quanto maior o número de conexões e mais ativas elas forem, as ideias fluem melhor e vão se refinando. Pentland fala em resultados positivos com a tunagem (tuning) da rede, como por exemplo:
a) acrescentando incentivos para pessoas participarem mais;
b) controlando o fluxo de ideias, para evitar que a mesma ideia (sem melhoramentos) continue circulando de forma repetida;
c) controlando o fluxo de ideias para que ideias que pouco circularam possam ganhar mais chances de avaliação.

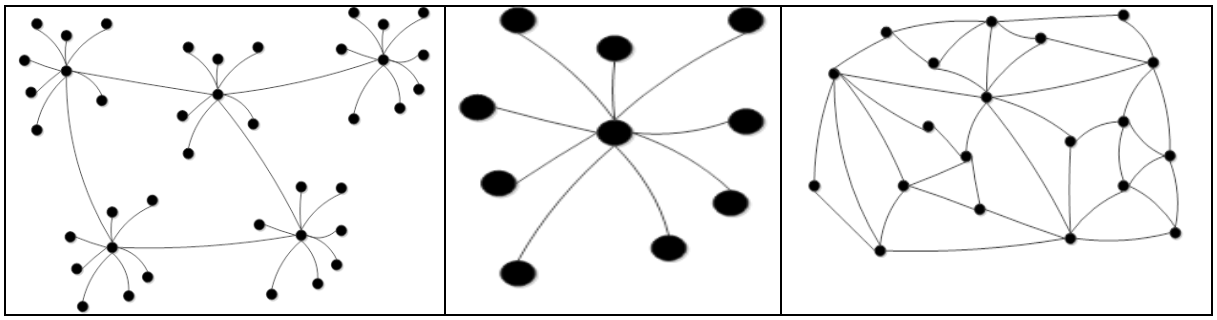
Alstott e parceiros (2013) publicaram um artigo onde investigam a velocidade das mobilizações feitas por redes sociais, chegando a conclusões sobre o que influencia (país de origem, sexo de quem posta e de quem recebe, idade, etc.).

Tipos de redes

A velocidade com que dados se disseminam depende do tipo de rede por onde os dados irão trafegar e dos elementos ou nodos que formam esta rede. Os elementos podem ser pessoas, tecnologias, empresas ou veículos de mídia.

Existem diversos tipos de redes (ver Baran, 1962). Uma rede centralizada possui um nodo central ligado a todos os outros e cada nodo que não é central está ligado ao nodo central. Já uma rede descentralizada não possui um nodo central. Uma rede distribuída é

caracterizada por caminhos alternativos para chegar de um nodo a outro. Já uma rede hierárquica possui nodos de ponta ligados a um único nodo que por sua vez se liga a somente um nodo e assim por diante, formando uma hierarquia.



Redes igualitárias, segundo Barabasi (2003) são aquelas cujos nodos (todos os nodos) possuem o mesmo número de conexões. Redes aleatórias são aquelas cujos nodos possuem uma média de conexões sem muito desvio. Já as redes Livres de Escala possuem alguns nodos com alto número de conexões, e outros com poucas. Os nodos com mais conexões podem chegar a ter milhões de conexões, ou seja, sem escala (Barabasi e Bonabeau, 2003). As redes Livres de Escala são importantes para o marketing digital, porque permitem disseminar rapidamente uma informação, bastando alcançar os nodos com maior número de conexões. Como Gladwell (2013) comenta em seu livro, estes nodos centrais podem servir para o mal (por exemplo, ser responsáveis pela disseminação de doenças) ou para o bem (por exemplo, para divulgar informações ou convencer pessoas).

Há também as redes chamadas de Pequeno Mundo, onde um nodo pode alcançar qualquer outro nodo por poucos laços. Milgram (1967) foi responsável por um experimento que demonstrou que qualquer pessoa no mundo estaria separada de qualquer outra por até 6 laços. Esta é a teoria dos 6 graus de separação. Barabasi comenta que experimentos já foram feitos com o Facebook, provando que esta separação é em média de 5,73 laços, e a distância máxima foi de 12 laços. Experimentos semelhantes foram feitos com o Twitter (5,2 bilhões de relações), chegando-se a uma distância média de 4,67 laços.

Numa rede, pode haver grupos isolados e estes são chamados Demes (como o conceito da Biologia, para populações isoladas que só cruzam entre si).

Segundo Barabasi, a Regra de Pareto (dos 20/80) também aparece nas redes de informações e conexões. Por exemplo, dizem que 80% dos links da web apontam para apenas 15% de páginas, que 80% das citações vão para 38% dos cientistas e que 80% dos links em Hollywood estão conectados a 30% dos atores.

A força das conexões na rede

As redes são formadas por elementos ou nodos que estão ligados entre si. Estas ligações ou laços possuem um grau ou força. Isto pode representar o número de interações ou a

confiança de um nodo em outro. Quanto mais duas pessoas interagem entre si, mais forte o laço entre elas se torna. De acordo com Granovetter (1973), a força de um laço é uma combinação de quantidade de tempo, intensidade emocional, intimidade (confidências mútuas) e serviços recíprocos. Laços assimétricos acontecem quando a força num sentido é diferente da força no outro sentido (Recuero, 2009).

Laços fortes são importantes para integração em comunidades. Isto dá coesão ao grupo. O networking é um dos 5 fatores de sorte ou sucesso conforme Max Gunther (2013).

Entretanto, laços fracos não devem ser menosprezados, pois servem para iniciar uma integração enquanto que os laços fortes a reforçam (Goldenberg et al., 2001). Até mesmo a fofoca pode ajudar a integrar uma pessoa numa rede. Foster (2004) diz que a fofoca ajuda a integrar pessoas num grupo, reforça laços entre pessoas, aumenta a coesão do grupo e não fofocar pode excluir alguém.

Por outro lado, conforme Goldenberg et al. (2001), quando o número de laços fracos aumenta, a força dos laços fortes diminui. Este fenômeno acontece no Facebook. Quando uma pessoa aumenta suas conexões mas tem muitas amizades fracas, as interações com amigos mais próximos acaba diminuindo também, e com isto o interesse pelo que é postado. E isto pode estar causando o declínio do Facebook.

Os tipos de laços podem caracterizar similaridades entre pessoas (comportamento, atributos demográficos), proximidade ou distância geográfica (onde moram ou trabalham), relações de amizade ou parentesco e também tipos de atividade ou gosto (operações de compra e venda, comunidades de interesse).

Pesquisas de Marder e outros (2012) concluíram que quanto maior o número de amigos no Facebook, maior o nível de stress, porque há mais chances de ofensas. Esta pesquisa, com mais de 300 pessoas cadastradas no Facebook, na maioria estudantes com idade média de 21 anos, descobriu a média de pessoas em diferentes grupos de contatos, a saber:

- conhecidos offline: 97%
- família: 81%
- irmãos: 80%
- amigos de amigos: 69%
- colegas: 65%

Outro problema que pode acontecer é quando “rich get richer”, ou seja, quem tem muitas conexões ganha mais novas conexões que os que têm menos. O resultado é uma rede altamente centralizada, e isto pode gerar dependência a estes nodos centrais, como já comentado antes.

Imitação e Moda – O Poder do Contexto

Muitas coisas se espalham por imitação. Pessoas especiais inspiram outros para imitação. A quantidade de pessoas utilizando algo a seu redor, acaba por influenciar seu uso. Veja o caso de viciados que não conseguem largar o vício porque continuam no mesmo grupo social de viciados. Assim também é a moda. E também assim funciona a obsolescência percebida. Quando alguém vê outro usando algo novo, fica com vontade de experimentar. E se o que você está usando não é mais utilizado pelo grupo social, você mesmo se sente envergonhado de usar aquilo.

Christakis e Fowler (2009) concluem que a probabilidade de uma pessoa fumar em contato com outro fumante é de 61% maior que a aleatoriedade. A probabilidade é 29% maior que a média aleatória se algum amigo do amigo fuma e cai para 11% num relacionamento indireto de grau 3. Com 4 graus (laços) de separação, não há incremento no risco, acima da aleatoriedade.

É assim também que as redes sociais crescem rapidamente. Você só entra numa rede ou por querer ser pioneiro ou porque seus amigos já estão nela. E não utilizar a rede significa ser excluído pelo grupo, não receber os mesmos convites, enfraquecer os laços. Uma rede social com muitos usuários incentiva novatos (o velho ditado de “rich get richer”).

O psicólogo social Janis (1972, 1982) apresenta a teoria conhecida como "groupthinking", segundo a qual a pressão do grupo sobre indivíduos pressiona por comportamentos padrão em conformidade com o grupo. Muitas vezes, o “groupthinking” gera decisões equivocadas. Indivíduos agem ou tomam decisões influenciados pelo grupo, mesmo contra seus interesses ou crenças, por necessidade de aceitação no grupo. O comportamento das massas pode ser algo bom para as empresas, mas também pode tornar-se uma histeria coletiva ou "tolice das massas". Gladwell fala que jovens fumam não porque estão vendo adultos fumarem, mas sim outros jovens.

Os eleitos e influenciadores – o ponto da virada

Segundo Malcolm Gladwell (2013), a velocidade de disseminação de algo passa de linear para exponencial num determinado ponto, que ele chamou de “ponto da virada”. Gladwell acredita que certos tipos de pessoas desencadeiam tal comportamento. Estes são chamados por ele de “os eleitos”.

Quando se trata de disseminação de informações, a velocidade de transmissão através de elementos de uma rede aumenta em muito quando a informação chega até um tipo de “eleito”. Desta feita, se alguém quer aumentar a velocidade da informação, deve fazê-la chegar aos eleitos. Mas antes é preciso identificá-los.

Os tipos de eleitos segundo Gladwell são: os comunicadores (hubs), os especialistas (autoridades) e os vendedores (usam de persuasão).

- Comunicadores (hubs)

São pessoas que possuem muitos contatos, que se relacionam com outras pela amizade ou contexto (morar perto, trabalhar na mesma área, fazer atividades comuns, frequentar mesmos lugares) e não por atributos comuns. Se estão em vários contextos diferentes, fazem conexões entre grupos diferentes, multiplicando suas conexões e fazendo a informação se espalhar entre diversos pequenos grupos (expansores de fronteiras -boundary spanners).

- Experts (autoridades)

São formadores de opinião e definidores de tendências. São aquelas pessoas que todo mundo respeita porque sabem muito sobre um determinado assunto ou porque possuem muita experiência.

- Vendedores

São aqueles que sabem utilizar de persuasão para convencer outros. São bons argumentadores, usam bem a retórica e sabem se expressar.

Em redes sociais, os chamados “influenciadores digitais” ampliam a disseminação de informações e memes. Segundo Godin (2000), há algumas maneiras para atrair e manter um influenciador:

- fazer grandes promessas;
- mostrar a eles como fazer em larga escala;
- descrever um caminho viável;
- propagandear alguém que conseguiu êxito.

Segundo a Roper Consulting, os influenciadores são pessoas que:

- Utilizam múltiplas fontes de informação;
- Colocam pessoas em primeiro lugar;
- Acreditam que o boca-a-boca é mais importante que mídia tradicional;
- Leem muito.
- Acreditam em compartilhar o que sabem;
- Acham que a propaganda não os influencia;
- Querem melhores produtos e serviços;
- Confiam nos seus instintos;
- Usam de ativismo; fazem mais do que anos atrás e estão envolvidos em trabalho voluntário e comunidades, discutindo problemas e soluções;
- Pensam fora da "caixinha";
- Dão ênfase em serem efetivos e relevantes;
- Gostam de preservar tradições e costumes;
- Veem a falta de moral como um grande problema social;
- Acreditam que pais e líderes de comunidade têm responsabilidade por guiar as crianças;
- Não acreditam que devam saber tudo mas buscam conhecimento para preencher suas lacunas;
- Procuram auto-melhoria e aprendizagem.

Existem influenciadores em cada área de interesse.

- Influenciadores sociais e influenciadores de categoria lidam com questões do dia a dia;
- Influenciadores de negócios lidam com questões profissionais: ;
- Influenciadores digitais lidam com questões digitais:

São características dos influenciadores sociais:
- São alguém que as pessoas confiam (confiável);
- Possuem muitas conexões e são bem-informados;
- São líderes do boca-a-boca;
- Indicam tendências e dão opiniões sobre produto, serviços e ideias;
- Possuem 2 vezes mais chances de serem procurados para dar opinião;
- Possuem 2 vezes mais chances de fazer recomendações;
- Participam de almoços e jantares de negócios;
- Fazem palestras;
- Escrevem para um jornal, revista ou participar de programas de TV;
- Fazem grandes doações para organizações;
- Fazem reclamações;
- Participam de reuniões públicas e eventos políticos;
- Participam de comitês;
- Fazem perguntas em encontros públicos;
- Organizam eventos sociais;
- São membros ativos de algum grupo lobista.

A Roper Consulting sugere algumas ações para alcançar os influenciadores:
- Mostrar para eles que o interesse não é no lucro, mas contribuir com causas e comunidades;
- Oferecer ações que deem suporte a grupos, eventos, jogos, fundações educacionais.

Aral e Walker (2012) fizeram uma pesquisa com usuários do Facebook para saber quem era mais suscetível a ser influenciado. As conclusões foram de que usuários jovens são mais suscetíveis a influências do que os mais velhos, homens são mais suscetíveis que mulheres, mulheres influenciam mais os homens do que outras mulheres, casados são os menos suscetíveis a adotar produtos, quem influencia não é muito influenciado e vice-versa.

Outro tipo de ator importante para disseminação de informações em redes sociais são os formadores de opinião. Não necessariamente são líderes formais ou por autoridade. Geralmente exercem tal papel por (Kang e Kim, 2009):
- características pessoais: são confiáveis, possuem conhecimentos superiores, novas práticas, forte personalidade, são menos dogmáticos, mais inovadores e mais aventureiros;
- teoria do capital social: possuem uma localização estrutural nas redes que faz a diferença; por exemplo, possuem alto grau de centralidade nas redes (número de conexões).

Uma das características dos formadores de opinião é que publicam regularmente sobre um assunto. Além disto, eles externalizam opiniões não tradicionais e utilizam inovações na forma de comunicação. Outro diferencial é apresentar dados, estatísticas e gráficos para suportar suas opiniões.

Velocidade na Tomada de Decisões

Uma das consequências negativas da velocidade da informação é a falta de tempo para pensar, avaliar e tomar decisões. Isto nos impele a seguir com as massas. Somos homens e mulheres massa como dizia Ortega y Gasset. Este sentimento de pensar como o grupo, e não querer discordar, ajuda a viver, pois não temos conflitos nem gastamos energia pensando. Aí entra a civilização do espetáculo que Vargas Llosa bem descreveu. Somos "maria vai com as outras" (ou "josé vai com os outros").

Esta falta de tempo nos faz acelerar a tomada de decisão. O ideal seria tomar decisões de forma racional, coletando e analisando dados, avaliando e comparando diferentes alternativas, usando a lógica, etc. Mas nem sempre isto é possível. O que acontece na prática é que as pessoas limitam os passos ou o conjunto de informações e possibilidades, ou porque não possuam tempo e recursos ou por que não vale a pena coletar todos os dados necessários e verificá-los. Esta é a teoria da racionalidade limitada nas decisões (Simon, 1972).

Tversky e Kahneman (este último, ganhado de prêmio Nobel) apresentam 3 princípios que limitam as decisões. O primeiro, o da disponibilidade, pressupõe uma certa capacidade para recuperarmos informações (discutido no próximo item). O segundo é o princípio da representatividade, pelo qual as pessoas fazem classificações e utilizam avaliações de similaridade, por exemplo para comparar situações atuais com anteriores, para reusar informações ou mesmo decisões já tomadas antes. O problema neste caso é que nem sempre as categorizações são feitas de forma correta ou completa. Por exemplo, pode-se contratar uma pessoa achando que ela será um bom funcionário porque seu currículo e atitudes são muito parecidos com os de um bom funcionário que já trabalhou na empresa. Entretanto, como dito antes, apenas alguns aspectos (por mais numerosos que sejam) foram utilizados neste processo. E além disto, há muita informação imprecisa e subjetiva.

O terceiro princípio é o da ancoragem e ajustamento, segundo o qual as pessoas utilizam pontos de referência e ajustam os caminhos a partir destes pontos. Isto pode envolver análises de dados históricos e modelos onde estes dados se encaixam. O problema é que o mundo, a vida, as pessoas, o mercado são muito complexos para seguirem tão perfeitamente modelos criados por humanos. Os pontos de referência podem ter sido mal definidos. Por exemplo, tomar como base as vendas do último ano, sem se dar conta que este ano de referência foi atípico. E há ainda erros na ajustagem. Se todo ano há um aumento de 10% nas vendas de um determinado produto, é o que se espera para o próximo ano. Mas as condições podem ser alteradas e o resultado pode vir a ser outro.

O livro "Rápido e Devagar" de Kahneman (2012) discute nossos 2 modos de tomar decisões: o sistema rápido e o sistema devagar (racional). O sistema rápido é utilizado por exemplo para reconhecer rostos. Até bebês o usam. E a gente não precisa raciocinar, é automático, sem esforço. Utiliza associações e reconhecimento de padrões, sendo difícil de

controlar ou modificar. Já o sistema devagar é usado para, por exemplo calcular quantas horas tem em 4 dias. Ele é serial, controlável, flexível, governado por regras e exige muito esforço.

O sistema rápido também utiliza intuições, emoções e instintos. O raciocínio lógico e o método científico sucumbem ante argumentos como “eu acho”, “eu sinto que”, “me parece que”. Malcolm Gladwell, no livro Blink (2005), fala de experimentos de psicólogos analisando vídeos de casais conversando e tentando prever se o casal iria continuar junto ou não depois de 15 anos. Ao analisar 1 hora de vídeo, eles conseguiram uma acurácia de 95%, enquanto que analisando apenas 15 minutos de vídeos, atingiram 90% de precisão nas predições. Ou seja, não são necessários muitos dados nesta situação. Padrões podem ser identificados em resumos. Gladwell também comenta sobre técnicas utilizadas por americanos para reconhecer operadores alemães de código Morse. Como saber distinguir operados numa tarefa tão rápida como transmitir código Morse ? É algo que não pode ser explicado conscientemente. Uchida, Kepecs e Mainen (2006) concluem que as pessoas vão acumulando dados, a partir de experiências e sentidos, os quais vão sendo agregados até o momento em que uma decisão é tomada. Mas isto acontece em frações de segundo.

Wilson (2004) discute o inconsciente adaptativo, um sistema de percepção não consciente, que utiliza funções de menor ordem (percepção, compreensão da linguagem), ao contrário de funções de alta ordem, envolvendo raciocínio. Segundo Wilson, nossos sentidos recebem 11 milhões de pedaços de informação num dado momento, nossos olhos recebem e enviam para cérebro 10 milhões de sinais a cada segundo, mas só conseguimos processar 40 partes de informação por segundo, de forma consciente. Por exemplo, se você pedir para um pianista explicar que sequência de teclas ele usa numa música que saiba tocar sem partitura (de memória ou de cabeça), dificilmente ele conseguirá explicar, ou pelo menos, levará um bom tempo tentando relembrar. Mas no momento de tocar a música, a sequência vem sem ele precisar pensar sobre isto.

Além disso, emoções, segundo Damasio (2005), também influenciam nossas decisões. Emoções são reações que acontecem no nosso corpo em razão da influência do meio em que estamos, do que estamos recebendo por nossos 5 sentidos. Já os sentimentos são o que o nosso cérebro pensa a respeito destas percepções do meio. Emoções são marcadores. Ajudam a lembrar coisas boas e ruins. Servem como alertas ou incentivos, através de dor ou prazer. Por exemplo, o coração bater mais forte, a boca ficar seca, a pele pálida, e os músculos se contraírem num momento de pânico.

E também tomamos decisões por instintos. Conforme Dawkins (2007) e Winston (2006) muito do nosso comportamento é moldado por instintos. Os instintos são reflexos de nossos genes e alguns estão enraizados em nós desde o tempo das savanas (antes das cavernas), sendo transmitidos de geração em geração por milhares de anos.

A falta de tempo ou urgência para tomada de decisão favorece o uso do sistema rápido em detrimento do raciocínio lógico para a tomada de decisão. Ayres (2008) compara decisões tomadas com estatística x intuitivas. A conclusão de seus estudos é que números são melhores como base para análise, mas a intuição é importante para levantar possibilidades.

Velocidade de análise de dados

Se tivermos pouco tempo para avaliar informações para a tomada de decisão, o ideal seria ter ferramentas que nos ajudassem nisso. Os softwares de Business Intelligence e de visualização de dados em gráficos melhoram a análise de dados.

Mas esta análise precisa ser feita em tempo real. Quem investe em bolsas ou criptomoedas quer ser avisado o quanto antes quando seus ativos perdem força. Consumidores querem saber logo das promoções para não perder a compra. Gerentes precisam saber dos problemas logo que acontecem. Médicos, enfermeiras e os próprios pacientes querem saber logo como está a saúde do paciente.

Mas o mais importante: as pessoas não querem receber quais são as alternativas mas sim qual é a melhor alternativa. Ou seja, o software do futuro deverá utilizar técnicas de Inteligência Artificial para coletar dados, analisar possibilidades e indicar a melhor opção para cada pessoa em cada situação.

Simulações rápidas e simulações sociais

Entender o comportamento de multidões é um desafio. Por esta razão, há especialista planejando o design de locais públicos não pelo estabelecimento de regras ou considerando uma coordenação centralizada, mas simplesmente observando o comportamento das multidões quando atitudes aleatórias e descoordenadas acontecem. Numa situação de incêndio, informações vindas de uma coordenação local não serão ouvidas e talvez nem mesmo os avisos luminosos sejam percebidos. A entrada e saída de multidões em eventos públicos deve ser pensada como um evento complexo onde pessoas com racionalidade limitada e atitudes diversas agem (como as renas fugindo dos lobos). Dizem que o Coliseu de Roma original, com capacidade para 50 mil pessoas, podia ser evacuado em 3 minutos devido a sua estrutura bem planejada. E naquela época não havia microfones, alto-falantes e avisos luminosos.

Além disto, a ação de uma pessoa acaba por influenciar a decisão dos que estão próximos. Isto pode modificar o comportamento dos outros, que podem imitar ou fazer algo bem diferente. Por vezes, algumas decisões de pessoas pensando no benefício próprio e único podem prejudicar ainda mais o sistema. Tomar decisões de forma independente, talvez não seja a melhor alternativa, conforme a teoria do Equilíbrio de John Nash. Talvez a melhor alternativa para todos seja cada um “perder” um pouco de algo para todos “ganharem”. As técnicas relativas à Teoria dos Jogos ajudam a entender os resultados nestes tipos de sistemas complexos.

A importância da simulação é que as pessoas podem avaliar resultados futuros sem ainda tomar as decisões no presente. A simulação é uma forma de experimentação prática. Por exemplo, são muito comuns simulações de resgate em caso de acidentes e simulações de evacuação de prédios em caso de incêndio. Isto permite avaliar como os elementos (por exemplo, pessoas) se comportam ou o que acontecerá quando os elementos forem

colocados juntos. As simulações são muito úteis para sistemas complexos, onde mesmo sendo conhecidos os elementos não se entende bem as relações entre os elementos.

Simulações sociais podem prever resultados a partir da aglomeração de seres vivos, bastando definir comportamentos simples a cada indivíduo e vendo o que acontece com o todo. Isto tudo sem precisar reunir pessoas ou animais em ambientes reais.

A simulação pode ser feita em ambiente real ou preparado, com elementos reais ou preparados (por exemplo, protótipos). Por exemplo, protótipos de carros podem ser testados em túneis de vento, para avaliação de sua aerodinâmica. Uma pequena inovação (mudança) no design pode ser testada para se saber se melhora ou piora o desempenho. Um software que permite fazer simulações sociais é o NetLogo (https://ccl.northwestern.edu/netlogo).

Simulações feitas em computadores permitem avaliar situações em menos tempo que o caso real. Por exemplo, para testar a evacuação de um prédio em chamas, não é preciso colocar pessoas num ambiente real. Podemos montar um modelo computacional e embutir regras de comportamento. Parâmetros podem ser alterados e a simulação pode ser refeita em segundos.

McNeely (2013) explica que o laboratório recria um ambiente real num espaço controlado, permitindo refazer experiências, observar melhor detalhes e até mesmo experimentar com modificações do meio ou das variáveis para testar comportamentos futuros e prever consequências ainda não observadas. Uma das vantagens é uma aceitação melhor do conhecimento descoberto, uma vez que pode ser observado por vários pesquisadores.

Entretanto, um cuidado que se deve ter é que um laboratório não consegue reproduzir fiel ou completamente o mundo real. Mesmo um ecossistema hermeticamente fechado não contém todos os elementos reais, seja em quantidade ou variedade. Portanto, os resultados devem ser interpretados sob as circunstâncias e limitações onde e quando os experimentos foram feitos.

O bom disto tudo é que num futuro próximo as simulações poderão ser feitas em tempo real no momento que forem necessárias. Então, um software poderá indicar qual a melhor saída num incêndio, sem que tenham sido feitos testes ou simulações anteriores.

Prazer e Recompensas Imediatas

Uma das consequências nefastas do aumento da velocidade no nosso mundo é a necessidade de recompensas imediatas, prazer momentâneo e resultados de curto prazo. Ninguém mais tem paciência para esperar.

A espera é um suplício. Antes era um prazer. Esperar o Papai Noel era um momento de ansiedade e ao mesmo tempo alegria. A dopamina a mil. E depois o prazer de recebê-lo e receber os presentes. Antes a gente comia o feijão com arroz de olho na sobremesa. Agora queremos a sobremesa logo.

Nossos jovens e crianças (a geração Y) buscam a sexualização e o amadurecimento precoce, funcionam em multitarefa, não querem mais estudar e sim ganhar dinheiro (vide aumento de investidores na bolsa e em bitcoins e o sonho de criar um novo Facebook).

Por isto o aumento de antidepressivos e outras drogas usadas para felicidade ao modo fácil.

O filósofo Zygmunt Bauman (2010) lembra no livro "Vida a crédito": "Com um cartão de crédito, é possível inverter a ordem dos fatores: desfrute agora e pague depois! Com o cartão de crédito você está livre para administrar sua satisfação, para obter as coisas quando desejar, não quando ganhar o suficiente para obtê-las. Vivemos a crédito: nenhuma geração passada foi tão endividada quanto a nossa – individual e coletivamente (a tarefa dos orçamentos públicos era o equilíbrio entre receita e despesa; hoje em dia, os 'bons orçamentos' são os que mantêm o excesso de despesas em relação a receitas no nível do ano anterior). Viver a crédito tem seus prazeres utilitários: por que retardar a satisfação? Por que esperar se você pode saborear as alegrias futuras aqui e agora? Reconhecidamente, o futuro está fora do nosso controle... E se o futuro se destina a ser tão detestável quanto se supõe, pode-se consumi-lo agora, ainda fresco e intacto, antes que chegue o desastre e que o futuro tenha a chance de mostrar como esse desastre pode ser detestável... As cadernetas de poupança se desenvolvem e se alimentam de um futuro em que se pode confiar – um futuro cuja chegada é certa e que, tendo chegado, não será muito diferente do presente. Um futuro do qual se espera que valha o que nós valemos – e assim respeite as poupanças do passado e recompense seus portadores. As cadernetas de poupança também prosperam na esperança/expectativa/confiança de que – graças à continuidade entre o agora e o 'depois' – o que está sendo feito neste momento, no presente, irá se apropriar do 'depois', amarrando o futuro antes que ele chegue. O que fazemos agora vai 'fazer a diferença', determinar a forma do futuro."

Mas há movimentos contrários como slow food e espaços gourmet, para cozinhar com amigos e com prazer, técnicas e espaços para meditação, templos religiosos para rezar, livros para ler e filmes para simplesmente chorar.

E Paul Zak, o Dr. Love, recomenda que alimentemos a oxitocina (hormônio que reforça o comportamento social, a empatia e a confiança) no nosso corpo de maneira natural com abraços, mensagens e filmes.

Planejamento e desenvolvimento ágil

Apesar da importância do planejamento, pelo que foi exposto acima, há uma necessidade grande de gerar resultados o quanto antes. Por isto, as metodologias chamadas "ágeis" estão em moda hoje em dia. A ideia é oferecer entregas rápidas de um produto com características mínimas (MVP ou produto mínimo viável). Para tanto, as reuniões são rápidas e os ciclos de desenvolvimento de um projeto também são curtos (chamados Sprints na metodologia Scrum).

O Manifesto Ágil propõe: (http://www.manifestoagil.com.br/)

- Indivíduos e interações mais que processos e ferramentas;
- Produto em funcionamento mais que documentação abrangente;
- Colaboração com o cliente mais que negociação de contratos;
- Responder a mudanças mais que seguir um plano.

Abrahamson e Freedman (2008), no livro "Uma bagunça perfeita" lembram que, às vezes, não vale pena organizar e planejar tudo (a relação custo X benefício talvez não seja satisfatória e o esforço não compense).

Para agilizar a comunicação de informações, muitos autores sugerem equipes ágeis nas empresas.

O Projeto Aristóteles do Google (baseado em centenas de entrevistas com funcionários e a análise de dados de mais de 100 equipes de trabalho na empresa) conclui que há 5 principais dinâmicas que definem o sucesso de uma boa equipe:

- Segurança psicológica: podemos assumir riscos sem se sentir inseguro ou envergonhado?
- Confiabilidade: podemos contar com os outros para fazer um trabalho de alta qualidade em tempo?
- Estrutura e clareza: são claros os objetivos, funções e planos?
- Significado do trabalho: será que estamos trabalhando em algo que é pessoalmente importante para cada um de nós?
- Impacto do trabalho: será que acreditamos no trabalho que estamos fazendo?

Duhigg (2016) explica que segurança psicológica é uma "crença compartilhada, comum aos integrantes de uma equipe, de que o grupo é um lugar seguro para correr riscos". "As equipes têm êxito quando todos sentem que podem se expressar e quando cada integrante demonstra sensibilidade em relação aos sentimentos dos demais." O objetivo é dar aos indivíduos uma maior sensação de controle, permitindo que todos tenham voz e estimulando sensibilidade social entre os companheiros de equipe ("os integrantes não precisam ser amigos... precisam ter sensibilidade social e permitir que todo mundo se sinta ouvido.

Segundo Sutherland (2016), no artigo "The New New Product Development Game" [O novo jogo de desenvolvimento de novos produtos], Takeuchi e Nonaka listaram as características das equipes que encontraram nas melhores empresas do mundo:

1. Transcendentes: elas têm uma noção de propósito que vai além do comum. Esse objetivo autoconcebido lhes permite ultrapassar o trivial e alcançar o extraordinário.
2. Autônomas: As equipes são auto-organizadas e se autogerenciam. Podem decidir como executar o trabalho e têm o poder de fazer com que suas decisões sejam cumpridas.
3. Multifuncionais: As equipes têm todas as habilidades necessárias para completar o projeto. Planejamento, design, produção, vendas, distribuição. E essas habilidades alimentam e reforçam umas às outras.

Aí aqui vão algumas dicas para que as ideias possam fluir melhor dentro de empresas ou entre empresas:

1. Aumentar a diversidade (entropia) do grupo, reunindo pessoas com diferentes perfis, background, formações, opiniões, papeis, funções, origens.
2. Forçar a participação de todos, evitando que pessoas ou pequenos grupos dominem discussões.
3. Utilizar brainstorming para levantar o maior número de ideais diferentes, sem avaliá-las.
4. Utilizar a técnica dos 6 chapéus para que pessoas com opiniões divergentes se encontrem e se manifestem, forçando a colisão de ideias.
5. Incentivar a circulação de ideias, evitando concentração em pessoas tipo "hub". Se possível, fazer com que cada elemento da rede tenha aproximadamente o mesmo número de conexões. As conexões podem ser medidas por trocas diretas entre pessoas ou por número de participações (por exemplo, número de mensagens enviadas numa lista de discussão ou fórum).
6. Categorizar ideias para evitar repetições e permitir que ideias semelhantes possam ser combinadas. Isto também permitirá avaliar a entropia das ideias (com o objetivo de aumentar a diversidade das ideias).
7. Fazer um banco de ideias, categorizadas, explicadas e descritas, para manter o histórico e permitir que ideias antigas possam ser reutilizadas para refinar ideias mais recentes.
8. Utilizar mapas mentais ou grafos para conectar ideias. Nas ligações, podem ser utilizados labels ou mnemônicos que justifiquem as ligações. Veja o filme de Sherlock Holmes e observe como ele conectava informações através de colagens e fios num quarto de sua casa.
9. Faça perguntas ao grupo, seja um "advogado do Diabo" na discussão (5W + 2H).
10. Leve as pessoas para campo, para que observem situações cotidianas. Faça-as descrever detalhes, principalmente aqueles mais óbvios ou menos relatados. Muitas vezes a inovação consiste em ver o que todos viram e ninguém notou.

PARTE III – VARIEDADE NOS DADOS

Conforme Levitin (2015), em 1550, "eram conhecidas quinhentas espécies de plantas no mundo. Em 1623, este número aumentara para 6 mil. Hoje, conhecemos 9 mil espécies só de grama, 2700 tipos de palmeiras, 500 mil diferentes espécies de plantas. E a quantidade não para de crescer."

Para poder viver num mundo tão diverso, a Humanidade começou a classificar as coisas e identificar padrões.

A classificação é um instinto do ser humano. Tentamos colocar tudo em grupos (pessoas, produtos, eventos, animais, plantas, etc.). Mesmo num texto como este, as informações estão agrupadas. Acreditamos que podemos reduzir tudo a um modelo único ou a poucas regras. Esta é a busca eterna dos físicos, para entender a Natureza e o Universo. Einstein acreditava que há uma ordem na desordem, mas que os padrões ainda devem ser descobertos.

A classificação e a padronização facilitam nosso entendimento do mundo e agilizam nossa tomada de decisão, e servem para minimizar a incerteza. Se encontramos uma situação nova e verificamos que ela se encaixa num padrão ou modelo já entendido, já sabemos que atitudes tomar naquela situação.

Entretanto, o 3º V do Big Data não está aí por acaso. A variedade de informações é boa para os seres humanos e para a Sociedade.

Segundo Koestler (1964) e Johnson (2011), a variedade de informações é um dos motores da criatividade, permitindo gerar mais ideias, soluções e inovações. Arhtur Koestler fala em bissociação de matrizes (bisociation of matrices); Johnson, em colisão de ideias (collision of hunches). Koestler descreve como passar repentinamente de um plano (assunto) para outro (como Arquimedes), conectando as partes e gerando uma solução nova. Johnson diz que é preciso completar nossas teorias com as ideias de outros. Eric Weiner (2016) e Dean Simonton (2002) confirmam. Devemos lembrar que Steve Jobs estudou caligrafia e artes; Galilleu, artes visuais; Da Vinci, anatomia; Einstein, música; Freud, arqueologia (era colecionador de antiguidades); Tom Jobim, os pássaros.

Segundo estes autores, o processo criativo inclui a busca por informações, associações e combinações de informações. Podemos usar analogias e mesmo pegar ideias emprestadas de outras áreas (por exemplo, adaptando-as para novas situações). Simonton (2002) lembra que muitas invenções surgiram quando o inventor teve visões (insights) fazendo outras coisas (Gutenberg estava amassando uvas para fazer vinho, e Arquimedes estava tomando banho nas banheiras públicas). Para Koestler (1964), o ato de criação não é criar do zero, do nada; é combinar, descobrir, selecionar, organizar o que já existia (fatos, ideias, habilidades).

Apesar dos benefícios da padronização, o que se quer hoje é personalização. Os marketeiros falam em microssegmentação de clientes. A Acxiom trabalha com 70 perfis de clientes.
https://c.ymcdn.com/sites/dema.site-ym.com/resource/resmgr/Member_Resources/Lifestage_Clustering.pdf?fbclid=IwAR2WOVVzKCINrjnh39qHlsI8I0CU9C2Ga_8BAPTuU3FFHrynrVJYtjgh8A4

Mas, pelo ponto de vista social, também precisamos da diversidade de produtos e serviços para atender pessoas e necessidades especiais. A inovação de produtos e serviços ficará cada vez mais especializada, para finalidades e públicos específicos.

Além disto, a variedade de informações nos tira das "bolhas". Algoritmos de aplicações de redes sociais filtram conteúdos de acordo com interesses dos usuários. Isto significa que cada usuário acaba recebendo somente coisas que "cabem" nos seus interesses ou ideologias. Mas, como já dizia Nelson Rodrigues, o consenso é burro.
https://canaltech.com.br/redes-sociais/algoritmos-de-redes-sociais-formam-bolha-politica-em-torno-dos-usuarios-60755/

Mas a variedade também existe para manter vivas algumas tradições. Conforme Pagel e Mace (2004), 700 das mil línguas diferentes faladas na Terra (aproximadamente 15%) são faladas na Nova Guiné. Para comparação, há apenas 90 línguas diferentes na China apesar de sua população e tamanho. Uma das possíveis razões para aquilo acontecer na Nova Guiné é que as tribos procuram se diferenciar umas das outras e também dificultar que forasteiros entrem no grupo.

É por razões assim que surgem diferentes culturas. Tooby e Cosmides (1992) definem cultura como um conjunto de "práticas de comportamento amplamente distribuídas ou quase universais de grupo, incluindo crenças, sistemas de ideação, sistemas de símbolos significantes ou substâncias informacionais de algum tipo." A cultura é mantida e transmitida pelo grupo, explicando assim similaridades intra-grupo e diferenças entre grupos. A cultura é replicada de geração para geração, através de aprendizado, podendo ser chamado de socialização. O indivíduo é mais ou menos um receptor passivo da cultura e é produto da cultura.

Para Hofstede (2001), cultura é a programação coletiva da mente que distingue um grupo ou categoria de outro. É um fenômeno coletivo, porque é compartilhado por pessoas que vivem ou viveram no mesmo ambiente social, no qual a cultura foi aprendida. A cultura é aprendida e não herdada; deriva do ambiente social e não dos genes.

Para Hofstede (2001) existem diferentes camadas culturais, sendo que cada pessoa pertence a diferentes grupos ou categorias ao mesmo tempo. As camadas culturais incluem a família, o país, a região onde a pessoa nasceu ou mora, a geração (faixa etária), o gênero (se masculino ou feminino), o nível ou classe social e a cultura organizacional (interna à empresa).

O Colapso do Consenso

Alvin Toffler, na década de 80 no seu livro "A 3a Onda", escreveu uma seção com este título: o colapso do consenso. Ele já previa o surgimento cada vez maior de muitas minorias diferentes e a dificuldade de integração e convívio. É a superespecialização. Podemos ver isto na moda, nas tribos urbanas, nas religiões, partidos e ideologias políticas, nas diferentes filosofias, teorias científicas e educacionais que aparecem nos livros, ou seja, até mesmo nos produtos que consumimos (existem revistas de Náutica especificamente sobre lanchas, comida indiana sem glúten, roupa masculina para ginástica em academias fechadas). E isto tudo também se reflete nas diferentes formas de agir do ser humano.

Segundo Nelson Rodrigues, "o consenso é burro". A democracia melhorou as decisões nas antigas civilizações, evitando que tiranos centralizadores tomassem decisões sozinhos, sem olhar para o bem de todos. A liberdade de expressão é defendida até hoje como algo imprescindível ao bom convívio e necessária para que o bem sobreviva e os erros não sejam escondidos. A diversidade de opiniões é saudável e, assim como o casamento entre raças diferentes propicia uma melhor seleção de genes, pode gerar ideias e decisões melhores. O problema é que a diversidade está crescendo de forma exponencial.

Depois, esta superdiversidade se encontra nas escolas, nos shoppings e até nas manifestações de rua. Numa escola, professores devem aprender a lidar com crianças com etnias e credos muito diferentes. Para o marketing, está cada vez mais difícil fazer campanhas. Hoje se fala em microssegmentação do mercado, em atendimento e produtos personalizados.

Vivemos numa época realmente difícil de definir e pré-determinar tudo. As definições hoje são nebulosas e não únicas ou imutáveis. O que é uma família, uma raça, um casamento, um casal, um filho, uma escola, uma empresa ? Empresas e sistemas devem aprender a trabalhar com conceitos cada mais subjetivos. Por exemplo, alguns hospitais trabalham com até 12 tipos de sexo (gênero) diferentes. As empresas e órgãos públicos estão agora obrigados a atender cidadãos pelo seu nome social. Registros de filiação não podem mais se limitar a nome de pai e mãe, pois pode haver 2 pais homens ou duas mães mulheres (e o que dizer da possibilidade de não ser uma dupla mas um trio ?).

A variedade de palavras aumentou com o tempo porque precisamos inventar palavras novas para designar o que só poderia ser explicado com muitas palavras.

Para encerrar, um conselho dado por Bertrand Russell, numa entrevista (Face a Face com Bertrand Russell, BBC), em 1959:
"*Neste mundo que está ficando cada vez mais interconectado, nós temos que aprender a tolerar uns aos outros. Nós temos que aceitar o fato de que algumas pessoas dizem coisas que não gostamos. Nós só podemos viver juntos desta forma, se nós vivermos juntos e não morrermos juntos. Nós precisamos aprender a bondade da caridade e da tolerância. O que é absolutamente vital para a continuação da vida humana neste planeta*."

Dados quantitativos

A maioria das técnicas de análise de dados é baseada em técnicas estatísticas. E estas por sua vez, devem ser aplicadas sobre dados quantitativos ou estruturados. Dados quantitativos incluem variáveis que podem ser expressas com valores numéricos (ex.: idade, quantidade de produtos em estoque, quantidade vendida, tempo de permanência de pacientes em hospitais), valores temporais (data e hora, por exemplo) ou valores conhecidos com nominais, categóricos ou qualitativos (ex.: bairro, cidade, sexo, classe social). Estes últimos são valores selecionados de um conjunto limitado, e não incluem atributos textuais que podem ser preenchidos com texto livre (ex.: descrição de um problema ou uma solução). Os dados nominais são semelhantes a dados numéricos porque poderiam ter um correspondente numérico. Por exemplo, sexo pode ser armazenado com um número representado os diferentes tipos (e há bases de dados que trabalham com mais de 2 sexos). Satisfação de clientes poderia ser representada por escalas numéricas; bairros, cidades e países podem ser representados por códigos numéricos.

"Todas as coisas são números", já dizia Pitágoras há mais de 2 mil anos atrás. Se as coisas não nascem números, nós as transformamos em números. A representação por números começou para facilitar as comparações e depois o comércio. E isto permitiu identificar padrões, para entendermos o passado e podermos nos preparar para o futuro. Assim foi contando dias, estações e anos. Só assim entendemos os ciclos da agricultura e de morte-vida de animais, inclusive nós mesmos. E assim também a Física conseguiu explicar eletromagnetismo, atração entre corpos, movimentos e tudo o que é matéria. "A Matemática é o alfabeto com o qual Deus escreveu o Universo", disse Galileu Galilei.

Mais tarde os números nos permitiram dar valor às coisas, sejam materiais (quanto custa, por quanto posso trocar) e mesmo às nossas vidas (por quanto tempo vamos viver, quanto tempo ainda temos, quanto vale nosso dia, nossas horas e nosso trabalho). É mais fácil para entender a natureza, os sistemas, os comportamentos e relações, e até mesmo a complexidade.

A Teoria do Caos diz que há padrão em tudo, até mesmo no nosso livre arbítrio. E tudo se reduz a funções matemáticas. A dificuldade não é nem encontrar a função que rege cada sistema, mas sim saber quais variáveis influenciam cada resultado, e depois conseguir coletar em tempo hábil cada medida. O Prêmio Nobel de Economia geralmente é dado a matemáticos, porque estes descobrem funções matemáticas para explicar comportamentos econômicos. Um destes casos é o de John Nash. No filme "Uma Mente Brilhante", a vida deste gênio é bem retratada. Em algumas passagens podemos ver como sua mente funcionava, tentando encontrar padrões matemáticos em tudo, por exemplo, pássaros voando, pessoas caminhando no campus da universidade.

Imagens podem ser representadas por funções matemáticas (modelo vetorial). Algumas são representadas por conjuntos de funções ou fórmulas, como por exemplo rostos. Isto permite comparar imagens de rostos de uma maneira muito rápida e precisa. Documentos e imagens também podem ser representados por uma sequência de números. Este é o método da assinatura. Uma função matemática pega o documento original, aplica operações e gera um

resultado numérico (sequência de números ou bits). Este resultado é usado para representar unicamente cada documento ou imagem. Assim, é possível comparar documentos para saber se são iguais ou mesmo comparar versões do mesmo documento, para saber se houve alterações. É assim que os mecanismos de busca (search engines) sabem se um documento na Web foi alterado e se é necessário reindexá-lo. Imagens complexas de fractais podem ser geradas por uma única função matemática.

Números também podem ser utilizados para representar relações entre elementos. O mais famoso é o Pi, que denota a razão entre a circunferência de qualquer círculo e seu diâmetro. Uma série famosa é a que virou a lei de Titius-Bode. Ela foi formulada inicialmente por Johann Titius em 1776 e depois formulada como uma expressão matemática por J. E. Bode em 1778. Esta série descreveria a distância dos planetas ao sol (com algum ajuste). Com esta série William Herschel descobriu um novo planeta além de Saturno: Urano (Losee, 2001). Também procuraram o planeta que faltava entre Marte e Júpiter e descobriram os asteroides Ceres e Pallas.

A famosa série de Fibonacci foi encontrada em vários casos na natureza (sementes de girassóis, caracóis, alinhamento de planetas). A regra desta série é de que o próximo número é soma dos 2 anteriores (0, 1, 1, 2, 3, 5, 8, 13, 21, ...). Ela era utilizada na arquitetura antiga como uma forma de estética e beleza. Outra série famosa é a que virou a lei de Titius-Bode. Ela foi formulada inicialmente por Johann Titius em 1776 e depois formulada como uma expressão matemática por J. E. Bode em 1778. Esta série descreveria a distância dos planetas ao sol (com algum ajuste). Com esta série William Herschel descobriu um novo planeta além de Saturno: Urano (Losee, 2001). Também procuraram o planeta que faltava entre Marte e Júpiter e descobriram os asteroides Ceres e Pallas. É claro que os céticos falam em coincidência.

O desenho do Homem Vitruviano de Leonardo Da Vinci (ver Figura 13) apresenta as proporções do corpo humano. E Da Vinci se inspirou em Vitrivius, que acreditava que a arquitetura deveria imitar as proporções da natureza. (Christianson, 2012)

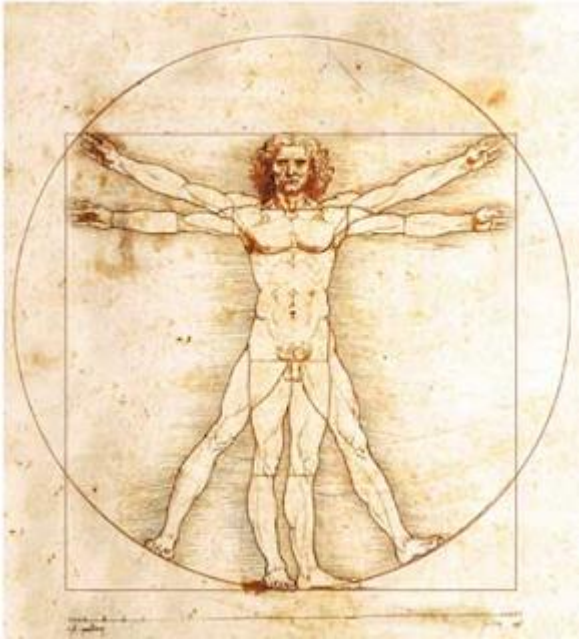


Figura 13: Homem Vitruviano de Leonardo Da Vinci

Também não se pode deixar de falar do famoso teorema de Pitágoras com diversas aplicações como por exemplo cálculo de distâncias, cálculo de menor caminho, etc. Já as leis gerais da geometria do grego Euclides são usadas até hoje para posicionamento em mapas, inclusive em espaços com mais de 3 dimensões. Até em funções computacionais de avaliação de similaridade e raciocínio baseado em casos (RBC), a distância euclidiana é utilizada.

Dados qualitativos

Por outro lado, há também análises qualitativas. Estas têm por objetivo encontrar as variáveis envolvidas, para depois então serem aplicadas técnicas quantitativas. Por exemplo, uma pesquisa sobre refrigerantes preferidos por uma população pode começar por uma pesquisa qualitativa, para que fossem identificados as diferentes preferências. Também pode-se fazer uma análise qualitativa para identificar possíveis motivos para cada preferência. Após então, pode-se conduzir pesquisas quantitativas para determinar quantidades (quantas pessoas preferem cada tipo e quantas vezes cada motivo foi citado). A análise qualitativa pode ser feita de forma manual ou intelectual por humanos, mas já há ferramentas de software que auxiliam tal processo. Neste caso, normalmente a análise qualitativa é feita sobre dados chamados não-estruturados, os quais incluem textos, sons e imagens (figuras, desenhos, diagramas, fotos, vídeos, etc.).

Um processo de análise de dados normalmente aplica técnicas quantitativas sobre variáveis. Mas não deve excluir análises qualitativas. Isto envolve, por exemplo, a descoberta de quais variáveis devem ser incluídas em um modelo para análise, que eventos podem interferir nos

resultados e como representar quantitativamente cada atributo. Por exemplo, a idade de pessoas pode ser representada utilizando um valor numérico absoluto (ex.: 35 anos), um valor relativo (ex.: mais jovens, mais velhos), uma faixa ou intervalo de valor (ex.: pessoas entre 15 e 20 anos) ou uma categoria ou valor nominal (ex.: crianças, jovens, adultos, terceira idade).

Tornar o subjetivo em objetivo, o abstrato em mensurável, o incompreendido e intocável em algo simples: este é o desafio de transformar o qualitativo em quantitativo..

Os programas que "escutam" um trecho de música e nos dizem que música é, transformam música (sons) em números para poder fazer a comparação rápida. Os softwares de biometria (identificação por características físicas da pessoa) também transformam um ser humano em números. Nossas características (traços do rosto ou das impressões digitais, atributos de nossa voz ou pupilas) são transformadas para equações matemáticas para uma rápida análise.

No nosso dia a dia, também usamos simplificações deste tipo. Por exemplo, para representar a qualidade de um filme ou uma música, usamos estrelas. Quanto mais estrelas, melhor a qualidade. Mas como traduzir um conjunto de diretores, atores, temas, cenários, etc. a um único número. Já há estudos para análise automatizada de textos e imagens. Isto certamente passa por números e fórmulas matemáticas. Só assim poderemos num futuro breve pesquisar no Youtube por um vídeo onde apareça um casal numa praia com coqueiros, ao entardecer.

O humor da humanidade já pode ser representado por uma série temporal, a partir da análise de postagens no Twitter. E já se pode entender o aumento de vendas pela análise dos gráficos do Google Trends. E isto inclui prever a bilheteria de um filme analisando blogs. Os artigos de Bollen et al. (2011), Choi e Varian (2012) e Mishne (2006) explicam o que eu estou dizendo.

Até a vida das pessoas está sendo representada em números. O método do biorritmo pretende mostrar as características físicas, emocionais, intelectuais e intuitivas em momentos no tempo. Isto permitiria considerar o momento de cada dimensão para tomar decisões. O pressuposto é que a vida de uma pessoa, representada por estas 4 dimensões, segue ciclos de períodos regulares (contados em dias). A Figura 14 abaixo mostra o meu Biorritmo no dia 18 de agosto de 2013, calculado pelo site http://www.profcardy.com/numerologia/biorritmo.php

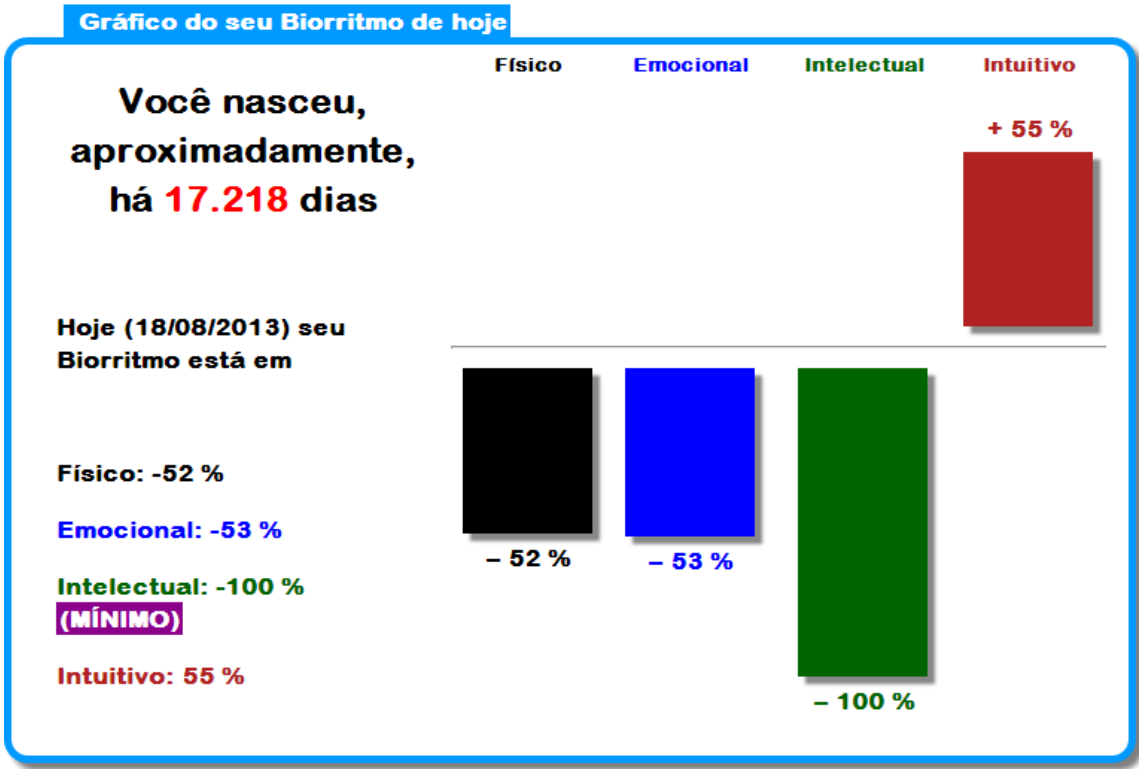


Figura 14: biorritmo num determinado dia

Já a Figura 15 mostra a previsão para os próximos 2 meses. Quando as 4 linhas estiverem lá embaixo, não vou marcar nenhum compromisso.

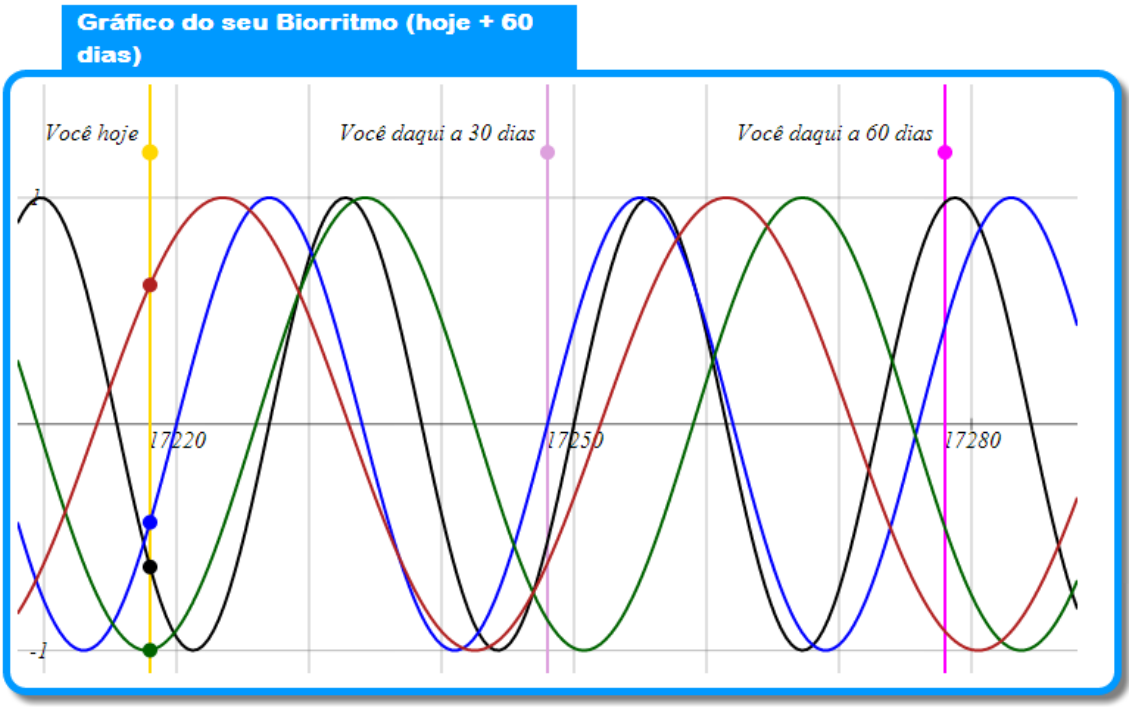


Figura 15: biorritmo para vários dias

Nem sempre os números são melhores que nossas intuições e sentimentos. Nate Silver relata que olheiros humanos tiveram melhores desempenhos que as estatísticas do sistema Pecota em vários casos no baseball (Silver, 2013). Apesar de Lewis (2004) relatar alguns casos contrários, (em Moneyball, as estatísticas foram melhores que olheiros), a conclusão é que ambos devem se ajudar. A prova disto é que a equipe de Obama mesclou dados e sentimentos das pessoas para fazer uma campanha vitoriosa (Moraes, 2012).

Análise Qualitativa X Quantitativa

A primeira etapa num processo de análise de dados é descobrir quais são as variáveis envolvidas: isto é uma pesquisa qualitativa. Por exemplo, a Figura 16 apresenta o resultado de uma pesquisa de campo onde foi solicitado ao entrevistado que dissesse qual o refrigerante que prefere beber. A pesquisa não era induzida, ou seja, não eram apresentadas opções; o entrevistado deveria escolher a resposta por contra própria. Na figura, podemos ver quais os produtos mais citados. Nesta pesquisa, a primeira variável definida foi o "produto" a ser pesquisa (no caso, o refrigerante que as pessoas bebem). Isto é uma análise qualitativa feita previamente. A segunda análise feita é a identificação dos itens mencionados, que também passam a ser variáveis. Nesta mesma figura, também temos um resultado quantitativo, pois a lista dos itens mencionados segue um ranking de frequência, ou seja, quantas vezes cada produto foi mencionado. Os produtos no topo da lista foram os mais lembrados.

1997	1998	1999	2000
Lemmon	Lemmon	Lemmon	Lemmon
Orange	Orange	Other-Cola	Other-Cola
Dig-Cola	Other-Cola	Dig-Cola	Dig-Cola
Other-Cola	Dig-Cola	CitriCola	CitriCola

Figura 16: Resultado de pesquisa não induzida

Já a Figura 17 apresenta o resultado de uma segunda pesquisa qualitativa no mesmo contexto. O objetivo era saber por que as pessoas preferem um determinado refrigerante. Note que, de novo, não foram indicados motivos, ou seja, a pesquisa foi do tipo não

induzida; o entrevistado tinha que lembrar ou identificar o motivo. Estes motivos então também formam uma nova variável. E a figura também mostra os resultados quantitativos, pois os motivos estão ordenados de cima para baixo, do mais citado para o mesmo citado.

Dig-Cola	Lemmon	Other-Cola	CitriCola
Embalagem bonita	Muito gás	Gosto	Preço
Preço	Gosto	Preço	Gosto
Gosto			Sempre encontro

Figura 17: Resultado de pesquisa qualitativa

Em resumo, as pesquisas e análises de dados em geral utilizam etapas qualitativas combinadas com etapas quantitativas. Se quisermos avaliar a satisfação de clientes com um determinado produto, vamos perguntar primeiro se ele gosta ou não do produto. Aí já temos a definição da primeira variável: gosto ou satisfação. E ela é binária, pois só aceita um de dois valores: sim ou não. Mas a pesquisa é quantitativa pois queremos sabre a quantidade de respostas "sim" e de "não".

Depois, será necessário fazer uma pesquisa qualitativa para saber o motivo de gostar ou não. E também será feita uma contagem (análise quantitativa) de tipos mencionados.

Informações Estruturadas x Não Estruturadas

Informações podem ser armazenadas de duas formas: estruturadas ou não estruturadas.

Dados estruturados normalmente são armazenados em tabelas, planilhas ou arquivos XML (arquivos textos com marcas indicando o significado de cada parte do conteúdo).

As informações estruturadas são chamadas assim porque são facilmente interpretadas ou entendidas. Por exemplo, quando temos uma tabela com dados de amigos, sabemos que, em cada linha, teremos dados de um único amigo. E também sabemos que nas colunas teremos cada tipo ou parte de informação (como na tabela a seguir).

NOME	**CIDADE**	**IDADE**	**SEXO**
José	São Paulo	35	M
Maria	Rio de Janeiro	23	F

João	Curitiba	43	M

Então, normalmente as informações estruturadas aparecem na forma de tabelas ou planilhas.

Um cuidado é que as pessoas que farão a interpretação dos dados nas tabelas ou planilhas entendam o que significa cada coluna, pois somente o nome pode ser ambíguo. Por exemplo, na tabela anterior, a coluna "cidade" significa "a cidade onde o amigo mora" e não onde ele nasceu.

Também podemos ter dados estruturados no formato de textos. Para isto, é preciso incluir marcas que representem o significado de cada dado ou parte da informação. Um exemplo disto são os arquivos XML muito utilizados na Internet. Eis um exemplo de parte de um arquivo XML com informações sobre livros:

```
<livros>
<título> O Ateneu</título>
<autor>Raul Pompeia</autor>
<ano>1888</ano>
<título> O Cortiço</título>
<autor>Aluísio Azevedo</autor>
<ano>1890</ano>
</livros>
```

A marca <livros> indica que vem uma lista de livros. A marca </livros> indica o final da lista. A marca <título> informa que a seguir vem o título de um livro (e que começa uma nova lista de informações sobre um livro). E assim por diante.

As informações estruturadas são mais fáceis de serem tratados por meios computacionais, porque existem linguagens formais (como SQL e QBE) que permitem sua manipulação e consulta de forma mais concisa e precisa. Entretanto, há o trabalho inicial de se definir a estrutura destes dados e depois coletá-los e incluí-los nos Bancos de Dados de maneira manual (através de trabalho humano).

A maioria das informações disponíveis na Internet está em formato não estruturado, tais como textos, imagens, gráficos, vídeos ou sons. Há quem diga que 80% das informações não estão estruturadas. Isto dificulta o tratamento por meios computacionais, apesar de o ser humano saber lidar e interpretar tais informações, relativamente bem.

Dados estruturados normalmente são armazenados em textos livros, imagens (figuras, desenhos, fotos, filmes) e sons.

Os dados não estruturados necessitam de mecanismos computacionais diferentes dos tradicionalmente usados para dados estruturados, para que possam ser coletados, armazenados, manipulados e consultados.

Sinais Fracos

Existem informações que sozinhas não significam muito, mas quanto integradas podem ajudar a predizer eventos. Estes são os chamados sinais fracos (weak signals) segundo Ansoff (1980). Sinais fracos são aqueles pedaços de informação, ambíguos, vagos, incompletos, imprecisos e controversos. Não são claros; são quase mudos. Estão normalmente escondidos no ruído e não recebem muita atenção no processo de decisão. Não são certezas, mas pistas. Podem surgir na forma de frases, fotos, cheiros, imagens, desenhos, pedaços de artigos ou qualquer observação pronunciada por alguém. Nesses fragmentos esparsos pode residir um potencial informativo importante para a investigação. Eles gradualmente se integram para formar um padrão de inteligência, que dão alertas de necessidades de mudanças. Tornam-se fortes quando combinados com outros sinais. "Uma andorinha sozinha não faz verão, mas vários bandos voando na mesma direção em dias diferentes e por um certo tempo devem significar algo importante. É por isto que as postagens em twitter, blogs, Facebook e outras redes sociais estão sendo monitorados por departamentos de inteligência, seja em empresas, governo, partidos políticos e até pais e familiares (a importância da análise de textos na Web será discutida mais adiante).

Sinais fracos podem gerar grandes influências nos resultados. A Teoria do Caos (Gleick, 1989) explica que pequenas alterações em algumas variáveis podem modificar completamente o resultado final. Daí é que surge o tal efeito borboleta (uma borboleta voando no Brasil pode gerar uma tempestade no Texas). Gladwell, no livro Ponto da Virada (2013) também comenta sobre pequenos eventos que desencadeiam grandes revoluções, como aconteceu com a Primavera Árabe e a 1ª Guerra Mundial. Nate Silver fala de um sinal que foi desconsiderado para um terremoto na Itália: sapos deixaram de desovar 5 dias antes. Outro caso interessante é com relação à análise de atrasos em voos. Muitas companhias descobriram que a etapa de limpeza da aeronave nas escalas era um determinante para os atrasos. Antes relegada a um fator de pouca importância no tempo da viagem, a etapa de limpeza recebeu foco de equipes de planejamento. A TAM então passou a usar um tapete vermelho para clientes limpares os pés na entrada. A GOL projetou um esquema em que os clientes ajudam na limpeza interna. E tudo isto deu certo.

Já Pentland estuda outros tipos de sinais fracos, o que ele chama de “sinais honestos”. São sinais que aparecem nos rostos das pessoas, impercebíveis no cotidiano pelo olhar humano, muito porque acontecem num tempo menor que um piscar de olhos. Quando assistimos vídeos em câmera lenta, tais sinais aparecem claramente. Pentland e sua equipe utilizam tecnologias para detectar estes sinais honestos. Os sinais podem ser demonstrações de empatia para facilitar e encorajar comunicação (ex.: acenos com cabeça) ou podem indicar estresse. E não aparecem somente no rosto, mas são demonstrados por todo o corpo humano. A linguagem dissimula emoções, mas o corpo não as consegue esconder. Já há até taxonomias para análise de expressões faciais (Ekman e parceiros, 1997; Kring e Sloan, 2007). O ser humano intuitivamente consegue identificar tais sinais fracos. É o que muitos dizem de uma conversa olho no olho para conhecer melhor uma pessoa. As decisões

referentes a escolha ou avaliação de pessoas são feitas assim. Mas também servem para avaliar veracidade de argumentos e informações que os outros nos passam. Saber reconhecer tais sinais pode melhorar nossa tomada de decisão. Os estudos de Pentland concluíram que empregados que se valem de interações cara a cara acabam sendo 30% mais produtivos.

Bancos de Dados e Planilhas

Dados estruturados são formatados em tabelas ou planilhas, pois se sabe exatamente o que cada coluna significa ou representa.

Dados armazenados em Bancos de Dados ficam dispostos em tabelas (este é o modelo relacional). Tabelas são compostas de linhas, representando entidades ou instâncias. E cada linha é dividida em colunas, que são os campos do Banco de Dados (partes de informações ou unidades de dados).

Abaixo temos um exemplo de tabela para armazenar dados sobre pessoas.

NOME	**CIDADE**	**SEXO**	**IDADE**	**PROFISSÃO**
José	São Paulo	M	23	Médico
Maria	Curitiba	F	45	Médica
João	Santos	M	67	Enfermeiro

A vantagem dos Bancos de Dados é poder armazenar dados de uma maneira simples e estruturada, permitindo a fácil recuperação posterior, inclusive por sistemas automatizados.

A recuperação de dados armazenados em Bancos de Dados é feita através de uma linguagem de programação específica para este fim, conhecida como SQL. Um comando de recuperação em SQL precisa informar basicamente que tabelas precisam ser consultadas, quais as condições de recuperação (só serão recuperadas as linhas que satisfazem as condições booleanas) e que colunas (atributos ou campos) devem ser recuperadas.

Outra forma de recuperar dados estruturados em tabelas é usando exemplos. Para isto há a linguagem conhecida como QBE (Query By Example). A consulta precisa montar um exemplo de linha a ser recuperada. Abaixo há um exemplo de consulta em QBE que permite recuperar o nome das pessoas que são do sexo masculino, que possuem mais de 20 anos e que tem como profissão “médico”.

NOME	**CIDADE**	**SEXO**	**IDADE**	**PROFISSÃO**
?	-	M	> 20	Médico

Para recuperar dados armazenados em estruturas tipo XML, há as linguagens XQuery e XPath, as quais precisam especificar as tags a serem recuperadas, além de informar condições em algumas tags e também poder navegar pela estrutura hierárquica de tags (se houver).

Dados Textuais

A área de Extração de Informação (EI) procura solucionar o problema de achar informações específicas dentro dos textos (Spark-Jones e Willet, 1997), como por exemplo pessoas ou entidades citadas e características de pessoas ou entidades ou objetos (ex. idade, endereço, ano de fabricação). O objetivo é converter dados não estruturados descritos com palavras em dados estruturados (como em planilhas ou bancos de dados).

Por exemplo, veja o texto abaixo:
"José da Silva é funcionário da Empresa XYZ, mora na Rua X número 31, tem 35 anos ..."

Este texto contém dados estruturados como o nome da pessoa, a empresa onde trabalha, seu endereço e sua idade. Na Web, temos exemplos de sistemas que fazem extração de preços de produtos em sites de comércio eletrônico.

Scarinci (1997) afirma que os processos de extração podem ser:
- estatísticos: com base em freqüências de ocorrência de padrões;
- léxicos: com base na ocorrência de termos simples ou compostos ou por formatos de termos;
- sintáticos: com base nas relações entre termos; ou
- baseados em conhecimento: com uso de regras pré-definidas.

Gramáticas regulares, *parsers* (analisadores de estruturas gramaticais) e autômatos finitos podem ser utilizados para o reconhecimento das informações através de análise léxico-sintática (estudo dos termos e suas relações nas frases). Entretanto, algumas técnicas podem fazer uso de técnicas mais complexas como análise semântica (estudo do significado de termos, frases e textos).

Os sistemas para extração de informações estruturadas a partir de textos podem ser:
- sistemas de reconhecimento de padrões (*pattern matching*): baseados na presença de termos, em combinações booleanas de termos e em combinações com peso
- sistemas dirigidos à sintaxe: avaliando as representações sintáticas das sentenças com o uso de *parsers* (parciais ou completos);
- sistemas dirigidos à semântica: utilizando técnicas como as anteriores, mais regras semânticas sobre um modelo do domínio.

É bom fazer distinção com a área de Recuperação de Informações (*Information Retrieval*), que tem por objetivo encontrar textos com informações relevantes. As ferramentas de Recuperação de Informações, geralmente, trabalham com técnicas de indexação. Existem

três tipos de indexação:

* indexação tradicional: é aquela onde uma pessoa determina os termos descritivos ou caracterizadores dos documentos, os quais farão parte do índice de busca; também podem ser utilizadas classificações ou taxonomias (como thesaurus);
* indexação *full-text* (ou indexação do texto todo): onde todos os termos que compõem o documento fazem parte do índice; e
* indexação por *tags* (por partes do texto): onde apenas algumas partes do texto são escolhidas, automaticamente, para gerar as entradas no índice (somente aquelas consideradas mais importantes ou mais caracterizadoras).

Índices do tipo *thesaurus* talvez sejam os mais conhecidos e usados para a indexação tradicional. São tipos específicos de dicionários e seguem a técnica do vocabulário controlado. Segundo Fosket (1997), *thesaurus* é um dispositivo de controle de termos usado na representação de documentos. Um *thesaurus* contém termos relacionados semântica e genericamente, baseados na Linguagem Natural, mas num conjunto restrito. As relações podem definir hierarquias entre conceitos (divisão de conceitos em menores). Ainda conforme o mesmo autor, os *thesauri* proveem mapas de conhecimento, apresentando conceitos ou ideias e indicando relações entre eles. A vantagem do uso deste tipo de indexação é a consistência através de vocabulário padrão, o controle de sinônimos e a facilidade de localização de novos conceitos. Um thesaurus também define os termos usados para descrever um conceito.

Já os sistemas de recuperação do tipo full text buscam por palavras presentes nos textos. Estas ferramentas indexam automaticamente todas as palavras do documento, gerando índices sobre palavras. O índice de palavras nada mais é que um banco de dados contendo as palavras encontradas nos documentos e apontadores para onde elas aparecem.

O terceiro tipo de indexação (por *tags*) procura indexar somente as partes relevantes do documento. Para tanto, ferramentas automatizadas podem analisar cada documento, procurando por marcas (*tags*) que identificam estas partes mais importantes. Ou então, humanos podem selecionar as palavras que melhor descrevem o conteúdo do texto (não necessariamente palavras que aparecem no título ou dentro do texto).

Para análise estatística de grandes conjuntos de textos, existem as técnicas e ferramentas para Text Mining. O meu livro "BI na era do big data para cientistas de dados - indo além de cubos e dashboards na busca pelos porquês, explicações e padrões" (Loh, 2014) apresenta e discute, em detalhes, técnicas e ferramentas para text mining (ferramentas também podem ser encontradas para baixar e testar em www.intext.com.br).

Dados em Imagens

Imagens estáticas (figuras, fotos, gráficos, desenhos, etc.) podem ser armazenadas de dois modos: Matricial (Raster) ou Vetorial. No modo Matricial, os objetos que formam a imagem são representados por um ponto ou por conjunto de pontos. Cada ponto possui

coordenadas (exemplo: linha-coluna, latitude-longitude) e atributos ou características (ex: cor, nome). Por exemplo, um prédio seria representado por pontos entre as linhas 5 e 8 e as colunas de 10 a 14. Já na representação Vetorial, os objetos podem ser representados por pontos, retas ou polígonos. Pode-se ter a formação de objetos complexos em níveis (por exemplo, uma entidade formada por um polígono, que é formado por linhas, que por sua vez são formadas por pontos). Também são permitidas representações em níveis ou escala: uma cidade pode ser representada por um ponto numa escala e por um polígono noutra. Basicamente, a representação matricial precisa guardar informações sobre todos os pontos que compõem uma imagem, enquanto que no modo vetorial é armazenada uma representação da imagem (por exemplo, uma função matemática).

Há algumas vantagens da representação vetorial sobre a matricial:

a) o armazenamento de dados na representação vetorial é mais conciso, pois pode ser feito por equações matemáticas; por exemplo, uma estrada pode ser representada por diversos segmentos de retas e cada um destes é representado por uma equação de reta; se fôssemos usar a representação matricial, precisaríamos armazenar inúmeras coordenadas de pontos;
b) em caso de serem utilizadas escalas (para visualização com *zoom in/ou*), quando há o armazenamento da representação de uma entidade apenas numa escala, para esta entidade ser visualizada em outras escalas é necessário realizar certas operações matemáticas; por exemplo, um estádio de futebol, perfeitamente circular, é representado no modo matricial como um conjunto de pontos: se precisarmos aumentar o tamanho do estádio (como na Figura 18), será que o sistema saberá que deverá manter a figura de um círculo ? no caso da representação vetorial, o estádio é representado pela equação de um círculo, então o dito problema não ocorre em operações de zoom (in ou out);
c) já que as imagens são representadas geralmente por equações ou fórmulas, fica mais fácil realizar operações matemáticas como cálculos de área, distância, comprimento, etc.

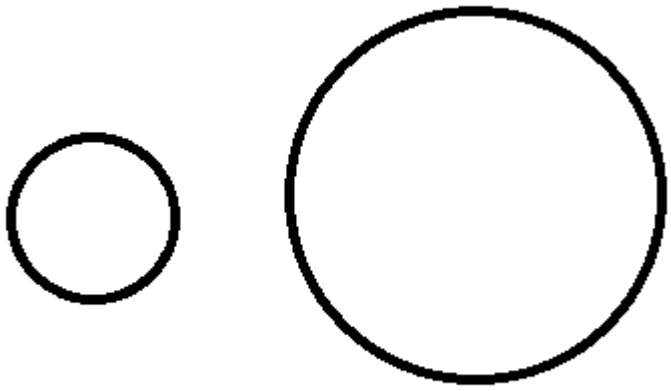


Figura 18: Um círculo na escala normal e depois de uma operação de zoom

No modelo matricial, as imagens são formadas por linhas e colunas de pixels, sendo um pixel uma unidade da imagem (picture element). Geralmente, os pixels são capturados e representados através de 3 sensores, um para cada cor primária (vermelho, verde, azul - formando a sigla RGB = red/green/blue). Assim, cada imagem é uma combinação de 3 imagens monocromáticas (uma para cada cor primária). Segundo Soares (1992), alternativamente, pode-se utilizar outros meios de caracterização dos pixels (ou píxeis) de uma imagem, sendo o padrão luminância-crominância considerado o melhor. Isto porque o olho humano porque olho humano é mais sensível a variações de luminância e porque numa imagem há pouca diferença em cor, além deste padrão poder recuperar valores de RGB. Cada pixel então é caracterizado por um conjunto de bits, informando sua cor, brilho, etc. No padrão luminância-crominância, utiliza-se maior número de bits para a cor verde, porque contribui mais para a luminância da imagem. O número de bits utilizados define o número possível de cores (por exemplo: 4 bits permitem 16 cores, 8 bits = 256 cores, etc.). A informação de cor pode ser definida com o uso de uma tabela ou palheta de cores (padrão usado em monitores e televisores). Cada pixel no monitor é dividido em 3 partes independentes (uma para cada cor de RGB), sendo sua cor dada pela variação da intensidade destas partes (representadas por um valor numérico).

Imagens em movimento são armazenadas através de sequências de imagens estáticas ou fixas. O normal é utilizar para filmes 30 imagens (quadros) para serem apresentados num segundo de filme.

Uma forma de diminuir o tamanho de arquivos de vídeo é armazenar somente as modificações de um quadro para o anterior. Assim, se um quadro possui um ponto no céu e no quadro seguinte este ponto está um pouco mais à direita, basta armazenar a imagem completa original e depois ir armazenando somente os pixels diferentes da imagem seguinte em relação à anterior (e assim por diante). Outra possibilidade é diminuir a resolução dos quadros (imagens estáticas) ou o número de quadros por segundo.

Imagens podem ser armazenadas e recuperadas com sistemas tipo GED. Entretanto, a recuperação não é feita pelo conteúdo do documento mas por atributos associados ao documento. Uma forma de utilizar GED para recuperar imagens pelo conteúdo é associar assuntos (tags) como atributos do documento. Mas a tarefa de definir os assuntos associados ao conteúdo e escolher as palavras a serem cadastradas ainda deve ser feita por humanos.

O Google Imagens também se vale de palavras associadas às imagens para fazer sua busca (nome do arquivo, palavras próximas à imagem ou que possuem link para uma imagem).

Já com o Google Lens, o usuário tira uma foto de um objeto e o aplicativo procura imagens semelhantes e tenta ainda descrever o conteúdo que aparece (por exemplo, se for um produto, ele dirá que produto é e dará algumas características como marca).

Logo algoritmos baseados em Inteligência Artificial conseguirão recuperar imagens através de seu conteúdo. Como já existem tecnologias para identificar rostos (Orkut e Facebook permitem marcar rostos de pessoas nas fotos), identificar partes do corpo humano, animais,

gestos, sorrisos, etc., no futuro será possível procurar imagens por exemplo por objetos que estão presentes ("encontre uma imagem de uma praia com o sol ao fundo, um barco passando e um coqueiro na praia").

Dados em Sons

A recuperação de áudio pode ser feita de modo digital (analisando padrões nas representações por *bits*), por conteúdo (padrões semânticos nos sons) ou por análise textual (transformando as falas em textos escritos e aplicando técnicas de RI). Maybury (1997) propõe a representação de áudio por seu conteúdo através da análise das propriedades acústicas do áudio (duração, *loudness*, brilho, tamanho da banda).

Segundo Soares (1992), para armazenamento de áudio é necessária a transformação de um sinal analógico em dados digitais, seguindo duas etapas:

a) amostragem: representação de sinais temporais contínuos em números reais, sendo a taxa de captura dada pelo número de bits por amostra;

b) quantização: representação dos números reais em um conjunto finito, resultando em uma distorção controlada.

Em geral, utiliza-se um espectrograma para representar sons, especialmente de voz. O espectrograma é um gráfico que mostra a intensidade por meio do escurecimento ou coloração do traçado, as faixas de frequência no eixo vertical e o tempo no eixo horizontal, demonstrando visualmente as características acústicas da emissão (Valentim et al, 2010). Um exemplo pode ser visto na Figura 19.

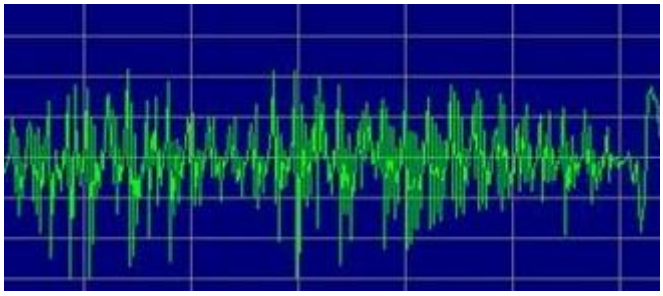


Figura 19: Sinal de voz armazenado

Tecnologias para armazenamento e recuperação de sons também estão avançando mas, como no caso de imagens, ainda geram muitas confusões. Alguns autores sugerem o reconhecimento da fala (*speech recognition*). O índice gerado automaticamente para a posterior recuperação dos áudios utiliza os fonemas ou palavras identificados nos sons como indexadores ou descritores dos documentos (áudio). As consultas então devem ser transcritas para fonemas (por um dicionário de pronúncia) ou palavras. Em geral, a primeira

alternativa (indexação por fonemas) é escolhida, com as seguintes vantagens em relação a transformar o áudio para textos (reconhecimento) e depois aplicar técnicas tradicionais de RI:

a) há poucos fonemas para treino;
b) permite a recuperação por sub-palavras (por exemplo, radicais).

O reconhecimento da fala é feito pela análise do espectro de áudio, dependente do tempo, fazendo o mapeamento para possíveis unidades lógicas (fonemas, palavras, sub-palavras). Para tanto, utiliza-se uma máquina de reconhecimento de padrões (*pattern matching*), treinada para cada instância (geralmente, fonemas).

Alguns cuidados devem ser tomados, tais como erros no reconhecimento dos fonemas (usar técnicas contextuais para diminuir ambiguidades ou reconhecer padrões difíceis) e o uso de nomes próprios.

O modelo mais utilizado nestas situações é o Hidden Markov Model (HMM), um modelo em escala, definido por autômatos finitos, formando um algoritmo adaptável com sub-modelos para formar palavras. Redes de Markov, que são modelos estatísticos que avaliam qual a probabilidade de um evento (neste caso, um fonema) aparecer após outro. Por exemplo, é pouco provável que um som de "z" apareça após um som de "r" na língua portuguesa. Isto elimina ambiguidades no reconhecimento, gerando maior precisão.

Sparck-Jones e outros (1997) complementam argumentando que a identificação de cada unidade da fala é computada usando uma sequência de probabilidades de transição com base nos predecessores. O problema deste tipo de modelo é que as pessoas costumam utilizar pronúncias diferentes.

O Youtube já possui uma tecnologia para reconhecimento de fala (entender o texto falado numa gravação sonora). Podemos ver tais tipos de sistemas também em celulares. No início, era preciso gravar um nome para poder fazer a recuperação de contatos por voz. Os sons são armazenados digitalmente pelas suas características (intensidade, frequência, etc.), como no exemplo da Figura 19. Assim, quando um nome era falado, o sistema do celular transformava o sinal analógico para sinais digitais e comparava com o que estava gravado na base de dados. O sinal mais similar seria recuperado.

Atualmente, a tecnologia está evoluindo. Não é mais necessário gravar nomes ou comandos. Há uma base de "falas" já previamente gravadas nos aparelhos (vêm de fábrica). O problema é que variações de timbre, língua, sotaque e barulhos no ambiente podem reduzir a eficiência.

O software Via Voice da IBM permite reconhecer falas com bastante precisão. Entretanto, como ele utiliza redes neurais, deve haver um processo de treinamento da rede anterior a seu uso. E depois de treinado, o software só reconhece frases da pessoa que o treinou.

Alguns Centros de Atendimento por Telefone (Call Center) permitem que o cliente possa falar o que deseja. Neste caso, não é necessário treinamento do cliente. Porém, a limitação é

que apenas algumas palavras podem ser identificadas. Como o contexto geralmente é limitado, o reconhecimento funciona com uma boa eficiência e baixa margem de erro.

O problema de identificar qualquer palavra falada a partir de qualquer emissor humano está a caminho de ser solucionado. Para que tais soluções funcionem, é preciso ter uma base de fonemas (sons básicos) para cada língua. Depois, estes fonemas são representados digitalmente, como na Figura 19. Ao se receber uma frase falada, esta é representada também digitalmente. Após, ela deve ser quebrada em partes e cada parte deve ser comparada com os fonemas básicos.

Sistemas que possam armazenar e recuperar gravações sonoras poderão ser úteis para repórteres, emissoras de TV e rádio, mas também nas empresas poderão ser utilizados. Para documentar reuniões, apresentações e treinamentos, muitas vezes é mais fácil gravar do que anotar o que é dito. Mas na hora de recuperar, as pessoas precisam procurar por gravações que contenham certas frases ou palavras.

O interessante dos sistemas de recuperação é que podem combinar textos e sons. Por exemplo, um texto pode ser recuperado de forma audível, através de tecnologias de síntese de voz. Por exemplo, já há muitos anos, existem sistemas de resposta audível para consultar saldo bancário por telefone. Uma voz sintetizada lê um texto. O resultado é ainda um pouco robótico, porque o sistema precisaria imitar a forma de seres humanos falarem, principalmente incluindo entonação e junção de palavras.

Diversos aplicativos já fazem síntese de voz. A Odd Cast possui uma tecnologia muito interessante que permite a escolha da língua e do sotaque, incluindo o sexo da foz sintetizada. O resultado é muito bom, mesmo quando se pede para um francês falar um texto em português.

Imagens X textos X sons

Na evolução do ser humano, a imagem veio antes da escrita; mas sons vieram ainda antes. McLuhan (1962) comenta o caso de povos africanos que não conseguem entender filmes (imagens em movimentos). O mesmo autor atesta que palavras escritas perdem as emoções da palavra falada. Um sentido pode anestesiar outro. O visual faz perder outros sentidos. Por isto, Cezanne pintava diferente, valorizando características dos objetos (volume, peso) que não apareciam nas imagens.

Pieters e Wedel (2004) afirmam que imagens são melhores para capturar atenção, independente do tamanho. Já os elementos de texto capturam melhor atenção na proporção direta do seu tamanho. Uma marca (brand) transfere atenção para outros elementos. Para estes autores, existem dois tipos de atenção: a básica e a incremental. Isto se compara a chamar atenção X manter a atenção. Enquanto as imagens ajudam a chamar a atenção (básica), os textos ajudam na atenção incremental (manter a atenção).

Carr (2010) acredita que o uso da Internet estimula a inteligência visual e espacial, em detrimento da capacidade de análise, reflexão e pensamento crítico.

Mapas e Informações Geográficas

Uma forma muito comum de armazenar informações é através de mapas. Podemos armazenar rotas, posições de entidades, distâncias e proximidades, regiões, etc. O diferencial dos mapas é trazer informações que não aparecem em dados estruturados tradicionais, tais como distância, proximidade, tipo de terreno, etc.

A principal característica de um sistema geográfico (GIS) é o armazenamento de entidades geográficas (qualquer coisa que possa ser localizada num mapa ou imagem) e a integração de informações descritivas (atributos das entidades geográficas) com informações gráficas ou espaciais (mapas e outras imagens). A representação de entidades geográficas pode ser feita de modo matricial ou vetorial (já vimos antes a diferença de armazenamento de imagens nestes dois modos). Além disto, é possível realizar o chamado geoprocessamento, que consiste em análises sobre as informações geográficas, tais como cálculos de área, distâncias, proximidades, cruzamentos, etc.

Uma maneira simples de localizar uma entidade num mapa ou imagem é possuir um banco de dados com as entidades geográficas e sua localização fixa na imagem. Neste caso, podemos utilizar as representações matriciais ou vetoriais discutidas acima.

Entretanto, o interessante dos sistemas geográficos é poder encontrar dinamicamente e em tempo real a localização de uma entidade no globo terrestre. Para isto, precisamos de um sistema de referência espacial e de pontos de referência num mapa. A chamada geolocalização pode ser feita de diversas formas. As pessoas leigas normalmente acreditam que somente a tecnologia de GPS pode ser utilizada, mas isto está errado.

Os aparelhos de GPS se comunicam com satélites (normalmente 3 no mínimo) para gerar coordenadas de latitude e longitude. Com estas coordenadas, é possível ir para mapas e encontrar o ponto que representa tais coordenadas. O Google Maps, por exemplo, permite que o usuário forneça estas coordenadas para apontar um resultado no mapa. Os mapas devem possuir pontos de referência.

Outra forma de fazer geolocalização é através de antenas de celulares. O celular capta o sinal da antena com uma determinada intensidade. Isto permite saber em que raio de distância encontra-se o aparelho em relação à antena (ver Figura 20). Feito isto, por um esquema de triangulação de antenas e sinais, é possível determinar onde encontra-se o aparelho com margens de erro de poucos metros. De novo, será necessário ter pontos de referências em mapas (podem ser as próprias antenas), para apontar a localização num mapa.

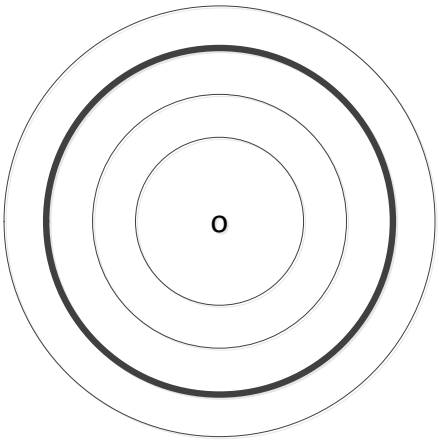


Figura 20: Círculos representando a intensidade de sinal de um celular em relação a uma antena

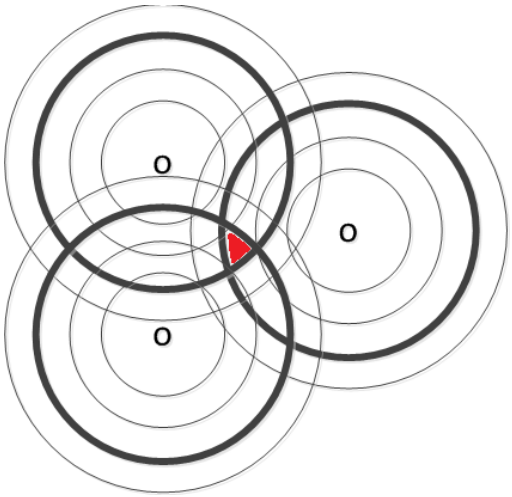


Figura 21: Triangulação de antenas

Outra alternativa para geolocalização mas também semelhante ao sistema de antenas de celulares é utilizar a comunicação de dispositivos móveis com antenas Wireless (por exemplo, pela tecnologia de Wifi). Se não houver antenas próximas, pelo menos pode-se saber em que região (área de um círculo) encontra-se o aparelho.

E a mais recente tecnologias de RFID também está sendo utilizada para geolocalização. Da mesma forma que as antenas Wifi, os chips de RFID podem ser geolocalizados. Isto permite encontrar produtos em armazéns e bois numa fazenda.

Note que o alcance das antenas diminui do primeiro caso para este último (nesta escala: GPS, celular, wifi e RFID).

A geolocalização de entidades também pode ser feita tendo como entrada um endereço (logradouro + localização vertical), por exemplo: Rua X, n. 41. É este o modo de operação

mais comum no Google Maps.

Para poder realizar tal operação, é preciso ter (ou comprar) mapas de cidades que associam posições de latitude e longitude a números em ruas (localização horizontal), como vemos na Figura 22). Neste exemplo, se quisermos localizar o número 200 da rua ABC, podemos utilizar simplesmente uma regra de 3; então o número procurado encontra-se no meio da quadra. Notem que é importante que haja estes pontos de referência. Para isto, existem empresas que fazem tais mapas. Entretanto, elas não possuem tantos pontos assim registrados, ainda mais em cidades pequenas. É esta a razão de muitas vezes o Google Maps apontar um endereço uma ou mais quadras distante da localização real.

Outro cuidado é que nem todas as cidades respeitam a regra de separar números pares e ímpares do mesmo lado da rua (esquerda ou direita). E isto também gera confusões como as que podemos notar no Google Maps.

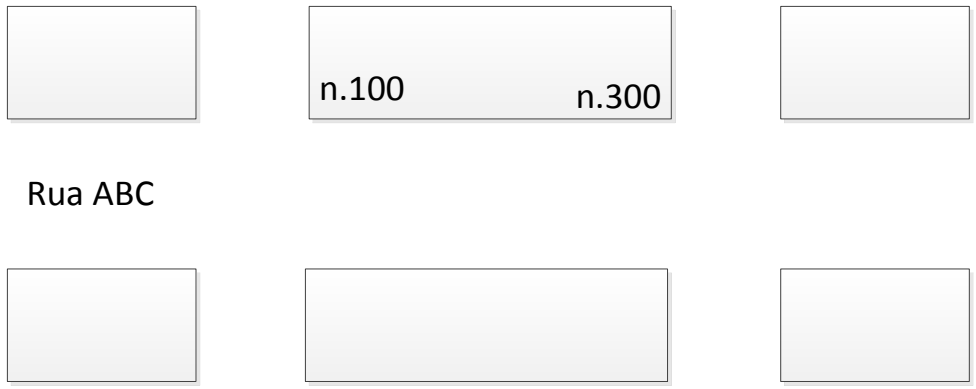


Figura 22: Geolocalização a partir de um endereço

Outro cuidado é que a regra de 3 funciona bem para ruas retas, sem muitas curvas. Por esta razão, já foram localizadas pessoas morando no mar na enseada de Botafogo (ver Figura 23

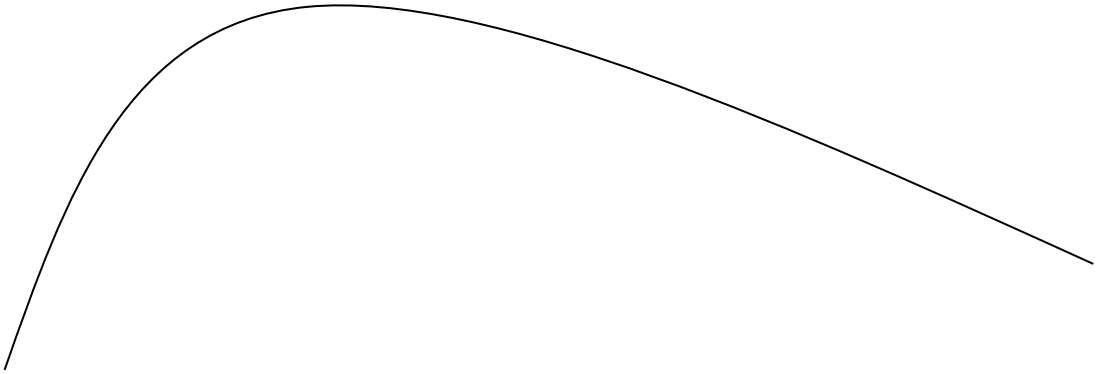


Figura 23: Causa comum para erros de localização de endereços em mapas (uma rua curva)

Outra possibilidade dos mapas é permitir a usuários selecionarem um ponto ou polígono no mapa e o sistema informar do que se trata (uma empresa, uma escola, etc), O Google Maps possui um recurso que indica o endereço de um ponto no mapa (rua e número). Entretanto, nem sempre ele consegue indicar qual a entidade geográfica correspondente. Para isto, é necessário ter um banco de dados sobre entidades geográficas, além do mecanismo de reconhecer coordenadas no mapa. O Google está tentando criar um banco de dados assim através de colaboração (próprias empresas e pessoas indicando) ou descobrindo por endereços associados a entidades (empresas, associações, etc).

Linguagens gráficas e diagramas

Uma forma de armazenamento de conhecimento é através de símbolos gráficos baseados em uma gramática, como o caso de diagramas.

Existem diagramas para representar esquemas elétricos, plantas baixas, fluxogramas, etc. Diagramas de Venn representam conjuntos e elementos. A tabela periódica representa elementos químicos e relações entre eles. A notação BPMN permite representar processos através de sequência de tarefas.

Árvores são um caso especial de diagramas. São usadas para representar genealogias e esquemas evolucionários de seres vivos (segundo a teoria de Charles Darwin). Mas há também casos interessantes como as partituras para representar sons e músicas, e a álgebra, que é uma linguagem matemática mas que usa a disposição dos elementos com base em regras gramaticais.

Grafos, Redes, Mapas Mentais e Mapas Conceituais

Outra forma de representar conhecimento não estruturado é utilizando imagens. Leia o livro de Christianson (2012) para entender como os diagramas mudaram o mundo.

Entretanto, as imagens devem ter algum tipo de organização para facilitar o entendimento do conhecimento explicitado.

Desta forma, surgiram os mapas mentais ou mapas conceituais. São estruturas tipo grafo, onde nodos estão conectados por arestas. Os nodos representam conceitos ou ideais e representam unidades de conhecimento. As arestas (ligações) representam associações entre os conceitos e podem ser de qualquer tipo (causa, efeito, consequência, explicação, similaridade, contrariedade, etc.). As arestas podem ser direcionadas ou não e podem ter significado ou algum tipo de associação ou uma classificação, daí portanto as palavras que aparecem junto às arestas.

Os mapas também representar fluxo de informações. Por exemplo, imagine que os nodos representem pessoas numa organização e as arestas direcionadas (setas) indicam de onde a informação vem (de que pessoa) e para onde vai (para que pessoa). Os nodos também poderiam representar grupos de pessoas, tais como setores ou departamentos (ao invés de indivíduos).

Os nodos podem possuir características ou atributos e isto pode ser representado por cores ou pelo tamanho do nodo. Então podemos representar competências das pessoas com cores diferentes, ou o setor de cada indivíduo. O tamanho do nodo pode indicar o tempo da pessoa na empresa (mais antigos = círculos maiores) ou o tamanho de sua competência. Ou então, ao invés de usar círculos, podem ser usados outros tipos de ícones, podendo representar cargos ou funções ou máquinas.

A Teoria dos Grafos surgiu com o artigo de Leonhard Euler, publicado em 1736, sobre o problema das sete pontes de Königsberg. O original em latim está no link abaixo:
http://www.math.dartmouth.edu/~euler/docs/originals/E053.pdf

A Figura 24 apresenta o mapa mental com conhecimentos sobre o Mal de Parkinson. Pode-se entender rapidamente deste gráfico que os distúrbio emocionais afetam esta doença, e que esta doença afeta o cérebro.

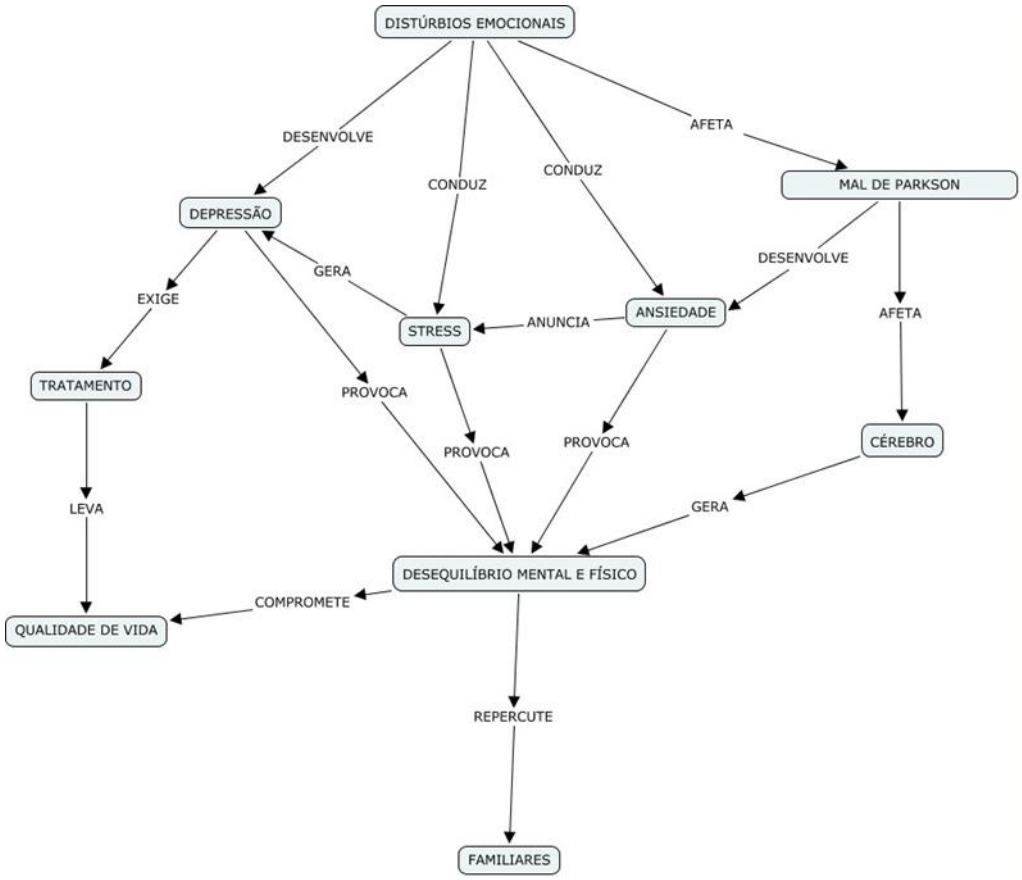


Figura 24: Exemplo de mapa mental para o Mal de Parkinson

O mais interessante no fato de que mapas mentais ou conceituais podem representar conhecimento é que podemos fazer análises e inferências em cima do mapa. E isto inclusive pode nos levar a novos conhecimentos ou pelo menos hipóteses. Seria como não precisar ler um livro para poder levantar hipóteses novas sobre o conhecimento que ele representa (se todo o conteúdo do livro pudesse ser representado por um mapa mental).

A Figura 25 apresenta o mapa mental com conhecimentos sobre impressoras. Imagine que você tenha um problema com sua impressora e precisa solucioná-lo. Verifique se o mapa lhe ajuda a encontrar uma solução.

Exemplos de falhas que podem ocorrer: (verifique o que fazer, recuperando conhecimento a partir do mapa mental):

- A máquina parou de funcionar.
- Mando imprimir mas não sai nada.
- Mando imprimir e sai folhas em branco.
- Às vezes imprime direito, às vezes imprime com falhas, depende do dia.
- Papel está saindo mascado.

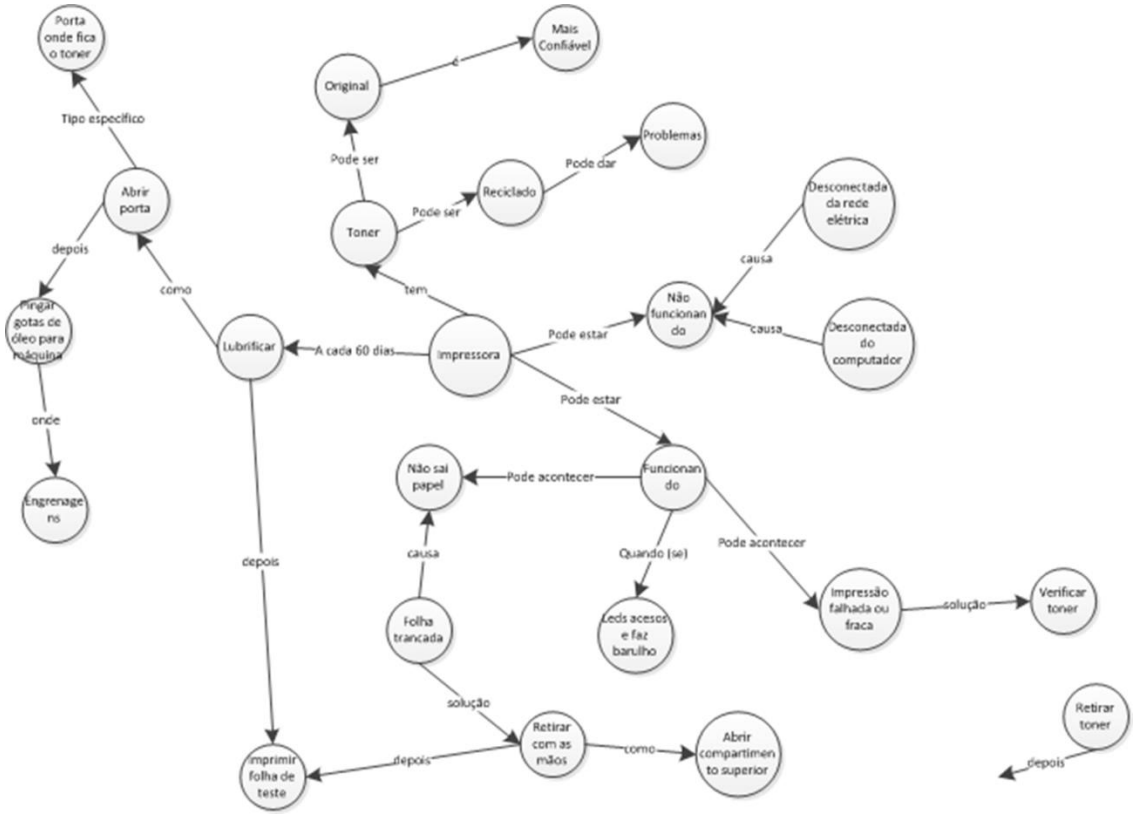


Figura 25: Exemplo de conhecimento para resolver problemas com impressoras

Tempos atrás o Google estava desenvolvendo um projeto para construir um Grafo do Conhecimento ("Knowledge Graph"), um modo mais rápido e fácil de obter respostas e explorar informações dentro do seu buscador. A ideia era conectar todo tipo de conhecimento recuperado em uma busca. Isto significa associar imagens, sons, vídeos, textos, etc. E mais, apresentar estes diferentes tipos de conhecimento através de um mapa ou grafo.

Figura 26: Mapa de Conhecimento

Assim, se você pesquisar sobre Machado de Assis, poderá encontrar um grupo de conhecimentos contendo seus livros, outro grupo contendo seus poemas, mas também suas biografias. A partir das biografias, poderia ver outros conhecimentos associados, como a cidade onde nasceu (textos e imagens), a época em que viveu (notícias, fotos e mapas da época) e assim por diante.

Mais detalhes sobre este projeto podem ser encontrados em
https://www.google.com/intl/pt-Br/insidesearch/features/search/knowledge.html

A recuperação de conhecimento estruturado como redes ou grafos é feita através das associações (links) entre os nodos (unidades de conhecimento).

Variedade de fontes de dados

A variedade também diz respeito a ter fontes de dados diferentes. Tal estratégia permite obter informações mais completas e também mais consistentes (discutiremos a seguir, o caso de informações conflitantes em fontes diferentes).

A coleta de dados em fontes diferentes permite, por exemplo, saber que produtos uma pessoa pesquisou na Internet e comparar com os produtos que ela realmente comprou na loja. Permite saber em que hotéis uma família ficou e quais lugares ela visitou durante esta estadia.

A diferenciação em fontes de informação também dá velocidade à transmissão de dados. Sistemas de CRM (Customer Relationship Management) integram diferentes canais de

comunicação com clientes para justamente poder atender melhor os clientes. Por exemplo, um cliente de banco pode pedir uma informação via site do Banco e obter a resposta quando estiver no caixa. Pode receber oferta de uma aplicação via SMS e aceitá-la pelo aplicativo móvel.

Thompson (2018) lembra que a emissora de esporte ESPN manda avisos para smartphones de clientes e isto ajuda a manter o interesse do cliente, além de atrair para o canal na TV. As pessoas escolhem canais e programas para saber mais sobre algo que já conhecem.

Integração entre dados variados

O uso de dados de diferentes tipos enriquece o conhecimento que pode ser extraído. Por exemplo, uma pesquisa de satisfação de clientes que colete dados estruturados (por exemplo, dados demográficos dos clientes) e dados não estruturas (por exemplo, reclamações registradas textualmente), permitiria associar os tipos de reclamações mais comuns aos tipos (perfis) de clientes. Uma pesquisa semelhante com clientes (e muito comum) é coletar satisfação de clientes através de notas (valores numéricos) para produtos e serviços e as reclamações textuais.

Outra possibilidade seria integrar dados de imagens. Por exemplo, rostos de pessoas podem identificar clientes que visitaram a loja e assim poderiam ser associados a dados de compras recentes destes mesmos clientes. E ainda: podemos identificar o sentimento do cliente ao analisar a expressão facial.

Outros exemplos de integração de diferentes tipos de dados já foram explicados na seção “Enriquecimento de dados – cruzamento de bases”.

PARTE IV – VERACIDADE DOS DADOS

A verdade é uma só ou não ? Pode haver duas ou mais verdades ? Por exemplo, se eu largar uma maçã que esteja segurando com a mão, o que vai acontecer com ela? A regra newtoniana nos diz que ela vai cair reta. Mas e se houver um vento lateral muito forte? Ou um sugador muito forte acima dela? Então o certo seria dizer: "ela vai cair reta se tais condições forem mantidas". Ou seja, a verdade é uma só sob certas condições, num dado contexto, momento e espaço.

A sociedade vai evoluindo e seus conceitos também. Antes havia dois tipos de sexo (gênero) para informar em cadastros: masculino ou feminino. Antes havia só dois campos para informar filiação: pai e mãe. Hoje temos nomes sociais, casais de mesmo sexo, famílias com dois pais ou duas mães, talvez até mais que isto.

A sociedade com a variedade discutida no capítulo anterior admite várias verdades.

Entretanto, para algumas situações não podemos ser tão flexíveis e admitir que cada um tenha sua verdade. Exemplos:
- um medicamento cura ou não uma doença?;
- a questão de um concurso tem qual resposta correta?;
- o réu é culpado ou inocente?;
- se enviarmos uma pessoa num foguete assim e com tais suprimentos, ela chegará em Marte e conseguirá sobreviver X dias?

Fake News e mensagens difamatórias estão arruinando vidas, influenciando eleições e separando grupos. Decisões, como discutido antes, podem ser tomadas com sinais fracos e tendências. Mas o método científico ainda é necessário para tomarmos melhores decisões. Uma sociedade mais justa não se consegue com subjetividade. Até porque seria difícil medir justiça e desigualdade sem indicadores bem definidos e sem instrumentos precisos de medição. Inovações e soluções só resolverão problemas se forem baseadas em conhecimento verdadeiro e não em suposições.

Análise da informação e suas partes

Platão, no livro Teeto (o conhecimento), ao registrar a discussão de Sócrates com Teodoro e Teeteto, lembra que conhecimento é a opinião verdadeira que possa ser explicada com a razão. Nestes diálogos também, eles concordam que a explicação racional pode ser feita pela decomposição de algo mais complexo em conhecimentos menores ou mais simples, até chegar a unidades que não precisariam ser definidas ou provadas. É assim com as letras para formar palavras; não precisamos definir as letras mas entendemos as palavras pela combinação das letras.

Então, uma das maneiras de confirmar uma informação é verificando a veracidade de suas unidades (o método conhecido como analítico, também proposto por Descartes e outros). Por exemplo, se alguém diz "que Fulano construiu um hospital enquanto era prefeito da cidade X", temos que verificar:

a) Fulano foi mesmo prefeito na cidade X ?
b) Foi ele que construiu o hospital ?
c) Foi no período em que ele era prefeito ?

Algumas verdades básicas podem ser assumidas como premissas para não se precisar provar tudo. Se formos céticos com tudo, seremos chatos e ineficientes. Se vamos calcular a área de um quadrado, sabemos que a fórmula é o tamanho do lado multiplicado por ele mesmo. Não é necessário verificar esta fórmula, pois já foi provada anteriormente.

É preciso também lembrar que quanto mais detalhes forem dados na informação, menor será a probabilidade de ser verdade. Por exemplo, se alguém disser que "haverá um terremoto na cidade X no dia 23 de novembro de 2020 às 9 horas", os seguintes fatos devem acontecer:

a) Terremoto na cidade X
b) Terremoto no referido dia
c) Terremoto às 9 horas

E tudo isto deve ocorrer junto (é uma conjunção).

A validação de uma informação pode utilizar estatísticas anteriores para dar maior credibilidade. No exemplo acima, se a cidade X fica no Japão, a probabilidade de um terremoto é maior do que se a cidade ficasse no Brasil.

Modelos e padrões

A finalidade dos modelos é permitir o entendimento de um conjunto de eventos, poder comunicar a outros, poder reproduzir este comportamento. Os modelos são construídos a partir de experiências passadas, de registros de casos que já aconteceram, com suas características descritas (o que, quando, onde, por que, com quem e como aconteceram - os 5W e 1H). Sem registros históricos não há como identificar padrões e daí montar modelos.

Os modelos então servem para entendimento de alguns aspectos da realidade (talvez a maioria, mas não todos), para predizer algumas situações (não todas, nem algumas poucas com total acurácia), para que possamos diferenciar contextos (gerais ou específicos), para que nossa vida não seja uma total escuridão e nossos caminhos possam ser trilhados com um mínimo de planejamento.

Modelo não inclui tudo, é uma representação da realidade, de parte dela, para um fim especifico. Todo modelo é uma especialização ou generalização da realidade, e toda especialização é uma abstração, ou seja, só absorve parte da realidade. Um protótipo de automóvel de tamanho reduzido terá apenas o design exterior do produto final, mas servirá

para avaliar a aerodinâmica do projeto. Já um protótipo em tamanho real deste carro servirá para crash testes mas não terá os acessórios interiores, nem a pintura final. Por outro lado, uma classificação étnica é uma generalização, pois tenta encaixar todas as pessoas em algum grupo existente ou pré-definido. Portanto, o modelo deve ser estudado dentro do seu contexto específico.

Entretanto, entender os limites (escopo) do contexto não é uma tarefa fácil. Se temos um modelo que prediz o quanto um cliente com perfil Y irá gastar no Dia das Mães, temos que entender que este modelo de predição só serve para os parâmetros definidos no modelo (sexo, renda, idade, estado civil, etc. do cliente). Se algum outro atributo fora deste contexto (por exemplo, altura do cliente) puder influenciar os resultados, a predição dada pelo modelo conterá uma margem de erro. E se houver vários destes atributos, a margem de erro aumenta.

Os modelos podem não ser precisos. É preciso avaliar quando, onde, como e por que os modelos acertam ou erram. Para tanto, devem ser feitos experimentos controlados. Fazendo novas observações, poderemos verificar se elas se encaixam no modelo. Se sim, confirmam o modelo. Se não, exigem algum refinamento do modelo ou mesmo a desistência dele. Entretanto, é impossível fazer todos os testes necessários ou coletar ou observar todos os eventos necessários. Hans Reichenbach comenta o caso de avaliar remédios utilizando placebo; isto pode durar muito tempo ou não ser possível de ser realizado pela dificuldade em encontrar cobaias.

Um cuidado: reduzir sistemas complexos a sistemas simples ou a um conjunto de regras simples pode ser perigoso. Podemos estar fazendo suposições erradas, julgando com critérios errados em cima de fatos não observados ou mal interpretados. E o resultado pode ser desastroso, como na caça às bruxas e na ascensão de ditadores. Daqui a pouco vamos estar categorizando e estereotipando todas as pessoas, como fazemos com grupos musicais e criações artísticas. O perigo do rótulo é não conseguir sair dele. Esta é uma tendência perigosa do ser humano: criar um modelo ou teoria e sair procurando casos que confirmem a sua validade, tentando encaixar tudo no modelo. E se algo não se encaixar ? Forçaremos o encaixe ou mudaremos nosso modelo ?

Informações imprecisas ou incompletas

Informações podem se tornar não confiáveis não por serem inverdades mas por estarem imprecisas ou incompletas (alguns chamam de “meias verdades”). Por exemplo, o endereço de uma pessoa necessita das seguintes partes para ser considerada precisa: um país, uma cidade, um logradouro, uma localização horizontal e uma localização vertical. Se uma destas partes faltar, não podemos dizer que a informação está errada, mas incompleta ou imprecisa. E isto pode prejudicar ações baseadas nestas informações. A imprecisão também pode acontecer com valores matemáticos (casas após a vírgula, temperatura, medidas de comprimento, etc.).

Uma informação desatualizada também pode não ser mais confiável. Assim como também a falta de informações por indisponibilidade. Não adianta o dado existir se não pudermos coletá-lo.

Informações duplicadas e inconsistentes

Imagine que recebamos dois endereços diferentes relativos à mesma pessoa. Qual é o verdadeiro? Podemos considerar a data de coleta ou atualização, a confiabilidade da fonte de coleta ou até mesmo algum documento comprobatório.

Em alguns casos, talvez possamos considerar como correto o conjunto de informações mais preciso ou completo (ex. endereço com localização horizontal e vertical se sobreporia a ter somente localização horizontal).

Provas e confirmação de teorias

Uma maneira eficiente de confirmar uma informação é fazendo uma prova ou refazendo a situação. É como funciona o método científico. E isto já era usado pelos primeiros cientistas da Idade Média. Eles observavam um fenômeno, construíam uma teoria e depois faziam novos experimentos ou observações para tentar validar a teoria.

Segundo Hempel (*apud* Thomas Kuhn) há 3 fases na avaliação de hipóteses científicas:
1. acumular observações, descrevendo status das observações ou experimentos;
2. verificar se estas observações confirmam, contradizem ou são neutras em relação às hipóteses;
3. decidir aceitar, rejeitar ou suspender julgamento das hipóteses.

Se alguém quer saber se uma teoria (por exemplo, "todo pássaro voa") é verdadeira, deverá procurar por casos que a confirmem e por casos que a contrariem ou neguem. Se não for encontrado nenhum caso deste último tipo, a tendência é confirmar a teoria.

Entretanto, Poincaré (citado em Stewart, 1991) lembra: "Quando queremos testar uma hipótese, que devemos fazer? Não podemos verificar todas as suas consequências, uma vez que elas seriam em número infinito; contentamo-nos em verificar algumas delas e, se somos bem-sucedidos, declaramos a hipótese confirmada."

Então uma maneira de validar teorias é o método "ten fold". Divide-se o conjunto de casos em 10 partes. Nove partes são usadas para gera o modelo ou teoria, e uma delas será usada para testar a teoria ou modelo. Este método é útil quando não se pode avaliar casos futuros. Então o teste é feito com casos existentes. Seria mais ou menos assim em diagnóstico médico: o modelo de diagnóstico é usado em uma décima parte de pacientes, assumindo-se que não tenham sido avaliados ainda. O resultado do modelo com estes pacientes é comparado ao diagnóstico real dado aos mesmos pacientes.

Desta forma, um modelo ou teoria talvez só funcione em algumas situações ou condições. E pode haver uma temporalidade. Até quando o modelo vai durar? O peru acredita que a família gosta dele pois lhe dá alimento e cuida de sua saúde. Mas isto só dura até o Dia de Ação de Graças.

Portanto, como diz Ian Stewart (1991), “nunca se pode ter certeza de que uma teoria é absolutamente correta, ainda que ela resista a um milhão de testes experimentais; pois – quem sabe? – poderá fracassar no milionésimo primeiro.”

Assim, talvez não possamos dizer que uma informação é verdadeira. O que devemos dizer é: há uma probabilidade X de que esta informação seja verdadeira. Conforme Descartes nos lembra: “quando não está em nosso poder o distinguir as opiniões mais verdadeiras, devemos seguir as mais prováveis”.

A prova matemática ou lógica também pode ajudar a validar um conhecimento. As transformações ajudam a chegar a novas conclusões ou a regredir a causas e premissas. A técnica de Redução ao Absurdo consiste em extrair conclusões a partir de premissas e então avaliar estas novas conclusões. Por exemplo, se alguém acredita que a Terra é plana, deverá pensar nas consequências disto, como por exemplo “a água cai para os lados e irá falta água no planeta”. Ao chegar a esta conclusão, a avaliação é que esta nova conclusão é um absurdo, que não pode ocorrer. Então a premissa ou teoria é falsa.

E há também a análise contrária. “Depois de você excluir o impossível, o que sobra, por mais improvável que seja, deve ser a verdade.” (Ellenberg, 2015).

A Falseabilidade ou Falsificação de Karl Popper consistem em fazer uma observação ou fazer uma experiência física que tente mostrar que a teoria o modelo é falso. Por exemplo, a asserção "todos os corvos são pretos" poderia ser falseada pela observação de um corvo vermelho. Como provar que horóscopo não funciona ? Se não existir um procedimento que permita provar isto, não é uma "boa" teoria. Se tiver como provar, e não se conseguir provar, a teoria está correta. Falsas teorias podem explicar praticamente tudo em seus respectivos campos.

Se há uma teoria de que todos os cisnes são brancos, para refutá-la precisaríamos encontrar um cisne que não fosse branco. Após exaustiva busca, se não acharmos damos como verdadeira aquela tese. Entretanto, talvez não tenhamos procurado direito ou só olhamos para uma certa região ou confundimos os cisnes negros com patos pretos.

Taleb (2008) explica tal situação no livro “A lógica do cisne negro”. O melhor é exemplo é o que acontece com o peru; ele pensa assim: “eu fui muito bem tratado durante o ano; a família desta casa me deu comida e cuidou da minha segurança e da minha saúde”. E acha que isto vai continuar para sempre. Mas só até o Dia de Ação de Graças.

Portanto, nem sempre é possível provar uma teoria ou que um conhecimento é verdadeiro. Talvez o seja só por um período de tempo ou numa região.

Amostragem e erro estatístico

Outro cuidado que devemos ter é com conhecimentos extraídos de amostras. Temos que lembrar que o resultado de uma análise é condizente com a amostra escolhida. Se ela foi mal preparada ou selecionada, o conhecimento resultante pode não ser confiável. Imagine que uma empresa faça uma pesquisa de satisfação com seus clientes. Se a empresa só convidar um certo grupo de pessoas, o resultado pode ter um viés que não reflete a realidade. Este geralmente é o problema com as pesquisas eleitorais. A intenção é boa, mas a amostra é ruim.

Também devemos atentar para o tamanho da amostra, como já discutido antes. Isto determinará a margem de erro dos resultados estatísticos.

Segundo a Wikipédia, a significância estatística é um "*procedimento para verificar a discrepância de uma hipótese estatística em relação aos dados observados, utilizando uma medida de evidência (p-valor). Este procedimento foi definido por Ronald Fisher no início do século XX como teste de significância... O nível de significância é a probabilidade de rejeitar a hipótese nula quando ela é verdadeira (conhecido como erro do tipo I). Em testes de hipóteses estatísticos, diz-se que há significância estatística ou que o resultado é estatisticamente significante quando o p-valor observado é menor que o nível de significância definido para o estudo. O nível de significância é geralmente determinado pelo pesquisador antes da coleta dos dados e tradicionalmente fixado em 0,05 ou menos, dependendo da área de estudo. Em muitas áreas de estudo, resultados com nível de significância de 0,05 (probabilidade de erro de 5%) são considerados estatisticamente relevantes. O p-valor (nível descritivo ou probabilidade de significância) é a probabilidade de se obter uma estatística de teste igual ou mais extrema que a estatística observada a partir de uma amostra de uma população quando a hipótese nula é verdadeira. Em outras palavras, o p-valor é o menor nível de significância para o qual se rejeita a hipótese nula. Por exemplo, a hipótese nula é rejeitada a 5% quando o p-valor é menor que 5%*."

Um bom livro sobre distorções causadas por análises estatísticas falhas é “Como mentir com estatística” de Darrell Huff (2016).

Além disto, temos que lembrar que não estamos trabalhando com um mundo fechado. Enquanto as análises estão sendo feitas sobre os dados coletados, outros dados estão sendo acrescentados ou aqueles coletados já se modificaram ou mesmo não existem mais. Por exemplo, após fazer uma pesquisa sobre a satisfação de clientes com o atendimento de um loja, ações continuam ocorrendo que podem interferir no resultado, tanto para melhorar quanto para piorar. Da mesma forma, se coletarmos dados com frequentados de um parque, após um mês talvez os frequentadores já sejam outros e tenham perfil diferente.

Isto acontece porque as análises são feitas sobre os dados armazenados na tecnologia (por exemplo, nos bancos de dados das empresas) e não sobre os dados da realidade. Todo processo de observação e coleta de dados acaba deixando dados de fora. Não há como coletar todos os dados; por isto, nosso mundo não é fechado. Até o planeta Terra troca energia e matéria com resto do Universo. E o acelerador de partículas do CERN na Suíça

precisa de alguém para ligar (e há sensores também). Não há como coletar toda a realidade. Não há como montar um ambiente ou laboratório para testes ou simulações que seja idêntico ao mundo real.

Funciona como esquematizado na Figura 27. Coletamos dados do mundo real através de diferentes formas e os armazenamos em bancos de dados (tecnologia). As análises são feitas sobre estes dados armazenados. Os padrões encontrados nos dados são interpretados gerando conhecimento novo. E aí acreditamos que este conhecimento explica o mundo real. O problema é que o conhecimento só explica os dados armazenados. O ideal seria coletar todos os dados relevantes para entendimento dos resultados das análises.

Por isto, temos que ter ciência de que os resultados das análises dizem respeito tão somente aos dados analisados, isto é, às amostras analisadas. O conhecimento descoberto é então uma hipótese ou tendência, que deverá ser confirmada analisando o mundo real ou através de tentativa e erro.

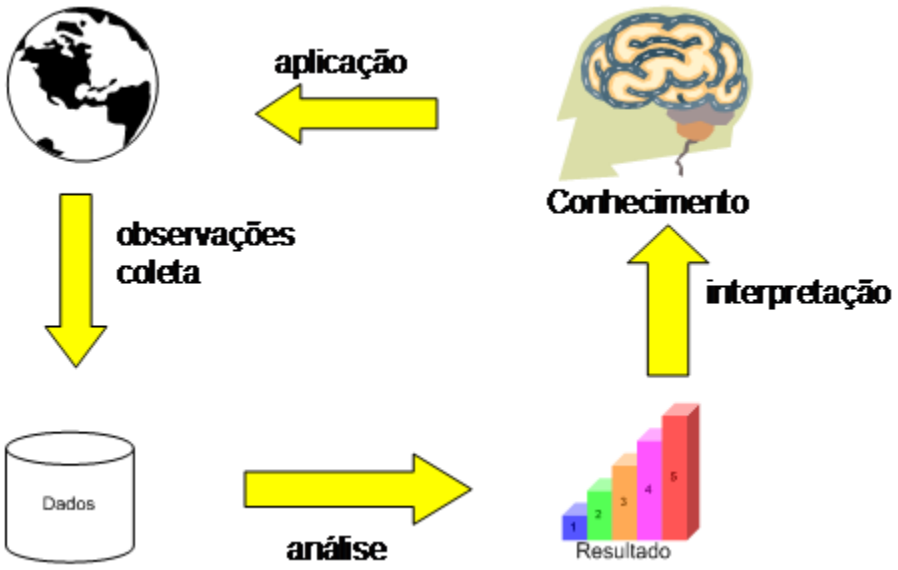


Figura 27: Teoria do Mundo Fechado

Credibilidade da fonte

A confiabilidade de um conhecimento está muito relacionada à confiabilidade da fonte que o transmite ou gera. É por isto que o mecanismo de busca do Google procura colocar no topo dos resultados as páginas ou documentos mais apontados. A ideia é identificar as chamadas autoridades, ou seja, quem é mais citado ou apontado por seus pares (páginas do mesmo assunto). Isso se assemelha a identificar quem está falando de quem. Em termos técnicos, basta saber que páginas apontam para que páginas, ou seja, que tem algum link

para outra página dentro do código html. As mais apontadas ou que recebem mais links, ou seja, as mais votadas, são autoridades.

No meio acadêmico, para avaliar a qualidade das publicações utilizados índices ou fatores de impacto, que contabilizam o número de citações de cada artigo. Na Figura 28, podemos ver um exemplo de contabilização de citações no Google Acadêmico (scholar.google.com.br).

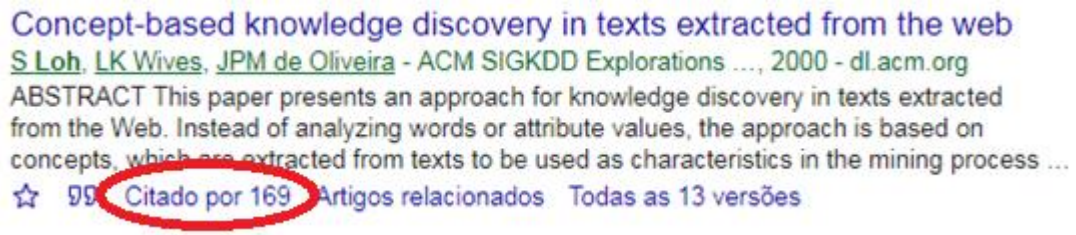


Figura 28: Resultado do Google Acadêmico com citações de um artigo

Conforme Wittgenstein (*apud* Epstein, 1986), "*o significado de uma palavra é o seu uso na linguagem; depende de quem a usa, quando, onde, com que objetivos e em que circunstâncias*". E McLuhan já dizia que o meio é a mensagem. Isto significa que a fonte e o meio de comunicação são importantes para gerar credibilidade sobre a informação transmitida. O meu livro sobre a comunicação entre usuários e analistas de sistemas (2014) discute com mais detalhes os problemas da comunicação.

Por isto, uma maneira de validar conhecimentos é analisando a fonte que o gerou ou transmitiu. Alguns artigos já atestaram a confiabilidade da Wikipédia (https://blog.wikimedia.org/2012/08/02/seven-years-after-nature-pilot-study-compares-wikipedia-favorably-to-other-encyclopedias-in-three-languages/), tendo um deles sido publicado na conceituada revista Nature. Uma das razões de sua credibilidade é que ela é pública e transparente. Apesar de qualquer usuário poder editar os textos, há revisores especialistas que validam as edições. Além disto, qualquer um pode apontar erros ou sugerir complementos ou correções. Este é justamente um dos fatores que fazem do sistema operacional Linux tão confiável.

Outra forma de avaliar o autor é pelo seu currículo. Michael Shermer ("Por que as pessoas acreditam em coisas estranhas: pseudociência, superstição e outras confusões dos nossos tempos") traz uma definição para pessoas inteligentes. São aquelas que possuem graus acadêmicos, cargos universitários (especialmente em instituições reconhecidas e de prestígio), publicações que são revistas por pares. Por outro lado, o mesmo autor adverte: embora "a expertise seja útil para separar o joio do trigo, é perigosa, pois podemos: 1) aceitar uma ideia errada só porque está sendo apoiada por alguém que respeitamos (falso positivo) ou 2) rejeitar uma ideia correta só porque é apoiada por alguém que não respeitamos (falso negativo)."

O índice h é um indicador numérico para reputação de cientistas.
https://pt.wikipedia.org/wiki/%C3%8Dndice_h

Uma discussão mais detalhada sobre isto está em:
http://miningtext.blogspot.com/2018/11/o-que-e-uma-informacao-confiavel.html

Sabedoria das Massas

Atualmente, muitos usam a chamada sabedoria das massas/multidões (wisdom of crowds) para chegar a soluções ou respostas a questões. A argumentação é que a média dos saberes estaria mais correta que o saber de uma pessoa (mesmo sendo um especialista). Pois quando um erra para mais, outro erra para menos e no total ou na média, a informação estaria mais correta. Isto funciona bem em alguns casos como descrito por Surowiecki no livro homônimo. Por outro lado, há também a burrice ou tolice das massas. Ver http://miningtext.blogspot.com/2013/07/sabedoria-das-massas-e-inteligencia.html

John Dominic Crossan utilizou, no livro "Jesus Histórico", uma metodologia para saber que fatos eram verdadeiros sobre Jesus. A teoria é que, se uma informação aparece em duas ou mais fontes independentes, então a informação seria verdadeira, porque seria muita coincidência que o mesmo fato fosse "criado" por duas fontes. Então ele analisou primeiro que fontes eram dependentes de que fontes (quem tinha lido ou se baseado em quem), e depois analisou os eventos descritos em fontes independentes.

Podemos usar a mesma metodologia para confirmação de informações. Se a informação aparece em mais de uma fonte e estas fontes não estão conectadas (uma não se baseia na outra), então podemos confirmar a informação. De novo, temos aí a sabedoria das massas.

O Google utiliza um método baseado na sabedoria das massas para montar o ranking de páginas nos resultados de cada busca. Filtros antispam também usam a indicação de várias pessoas para saber o que é spam. E tem funcionado bem.

Hipótese nula

Uma das formas de avaliar padrões estatísticos é definir uma hipótese que não deve ser verdade (a hipótese nula) e avaliar sua probabilidade.

Por exemplo, há uma teoria que podem ser encontrados nomes de pastores atuais na Bíblia. Os métodos usados comprovam que isto é verdade. Entretanto, a hipótese nula é que tantos outros nomes também podem ser encontrados com os mesmos métodos. Se isto for verdade, a teoria é nula (Ellenberg, 2015).

Outro exemplo. Suponhamos que se descubra que "100% das mulheres loiras compram o produto X". A hipótese nula é avaliar o que acontece com mulheres que não são loiras. Se for tudo 100%, não há novidade no 1o padrão.

Então uma boa maneira de confirmar um modelo é compará-lo a padrões estatísticos aleatórios. Por exemplo, se criássemos um método para selecionar e recomendar produtos a clientes por análise de perfil de cada cliente, iríamos registrar o percentual de aceitação das recomendações com este método e comparar com os resultados de um método aleatório, que seleciona randomicamente qualquer produto para indicar.

Quebra de paradigmas e verdades

Hoje em dia, vemos muita discussão sobre qual seria a verdade verdadeira em muitas áreas. Começa com a discussão entre Criacionismo vs. Big Bang e Evolução. Segue com a disputa entre religiões e provas da existência de Deus. Até chegar às teorias de Física, Química, Biologia, etc. Pois bem, o mais correto seria dizer "no momento, a teoria mais aceita é ...". Isto porque não há provas de uma verdade ou outra. Há indícios e há opiniões (maioria x minoria).

Já vimos várias teorias sendo quebradas: Terra plana, Sol x Terra, átomo indivisível, universo x multiverso, matéria x antimatéria, leis de Newton X Einstein X teoria quântica, etc. Até Stephen Hawking está revendo seus conceitos sobre buracos negros. Recentemente, o Museu de História Natural de Londres retirou a estátua de Lamarck e colocou a de Darwin. Mas já existem teorias surgindo dizendo que talvez algumas ideias de Lamarck estejam certas (características adquiridas passando para a prole). Por exemplo, talvez informações passem pelo DNA e aí explicariam a sensação de "déjà-vu".

Thomas Kuhn, no livro "A estrutura das revoluções científicas", estuda o surgimento de novos paradigmas. Quando observações não se encaixam na teoria antiga ou vigente, é necessário adaptar a teoria ou criar uma nova. É como funciona a tese x antítese, dando origem a uma nova teoria para acomodar ambos os lados (a dialética também faz isto).

Conforme Shermer (2011), "todos os fatos na ciência são temporários e sujeitos a contestação e portanto o ceticismo é um método que leva a conclusões provisórias... A ciência é progressiva porque os seus paradigmas dependem de conhecimento cumulativo, obtido por meio de experimentação, corroboração e refutação... o progresso científico [é] definido como o crescimento cumulativo de um sistema de conhecimento ao longo do tempo, no qual os aspectos úteis são mantidos e os não úteis são abandonados, com base na rejeição ou confirmação de conhecimento testável."

A quebra de paradigma não é fácil, porque as pessoas confiam em suas crenças e ideologias. Shermer (2011) afirma: "desenvolvemos uma 'imunidade' a novas ideias que não corroborem as anteriores. Historiadores da ciência chamam isso de Problema de Planck, referência ao físico Max Planck, autor desta observação sobre o que precisa ocorrer para haver inovação na ciência: 'Uma inovação científica importante raramente avança persuadindo e convertendo gradualmente os seus opositores: é raro que Saulo se transforme

em Paulo. O que ocorre é que seus oponentes aos poucos vão morrendo e a geração seguinte passa a se familiarizar com a ideia desde o início'."

Arbesman (2013) acredita que as teorias deixam de ser verdade com alguns padrões temporais. Ele usa o termo "meia-vida" dos fatos como uma analogia ao decaimento radioativo. Sua teoria é que fatos e teorias se tornam obsoletos. Por exemplo, seus estudos concluíram que metade das asserções médicas (ex. sobre cirrose e hepatite) duram em torno de 45 anos. A meia-vida de um artigo de Física é de 13,07 anos, de um em Matemática é de 9,17 anos e em Psicologia cai para 7,15 anos.

O 4º Paradigma

O 4o Paradigma de Jim Gray, chamado de eScience, vê a Ciência como uma exploração de dados, unificando teorias, experimentos e simulações. As ferramentas permitem analisar grandes volumes de dados, a fim de unificar teorias, explicar resultados de experimentos e simulações e ajudar na construção de novos modelos.

Milhares de anos atrás, a ciência era empírica, descrevendo apenas fenômenos naturais. Este foi o 1º Paradigma. Por exemplo, o método da sangria era experimentação pura, uma espécie de tentativa e erro (ou, faça e veja o que acontece). E isto durou até a Renascença e o Iluminismo. Hoje temos tecnologias para coleta de dados que incluem sensores, câmeras e Internet das Coisas (IoT). Podemos armazenar dados não estruturados como imagens, sons e textos.

Há poucas centenas de anos, ramos teóricos surgiram usando modelos e generalizações. Este foi o 2º Paradigma. Com ele, surgiram as leis da Física e outras equações e teorias que explicam os fenômenos observados e podem prever eventos futuros. Hoje temos tecnologias como BI e Data Mining que podem ajudar a encontrar padrões estatísticos e modelos matemáticas de forma quase automática. Softwares de predição podem ajudar a predizer eventos ou valores numéricos usando modelos e equações matemáticas.

Com o avanço dos computadores e do software (por exemplo, processamento paralelo e inteligência artificial), foi possível elaborar teorias complexas e testá-las com simulações computacionais. Este é o 3º Paradigma. Imagine poder testar a divulgação e a venda de novos produtos da seguinte forma: se fizermos um post no Facebook para 1000 pessoas, se uns 20% compartilharem, se 2% dos que receberem a postagem comprarem o produto, se 10% dos que comprarem divulgarem, e se etc etc. qual será o resultado final ?

O 4º Paradigma, a eScience, é um processo de exploração massiva de dados, combinando a identificação e uso de teorias, observações e experimentos e a simulação computacional.

Para testar ou avaliar modelos e teorias, os cientistas realizam análises de Big Data, armazenados em bancos de dados não estruturados, capturados por instrumentos e sensores de última geração, usando computadores de alto desempenho para simulações, técnicas de

gestão de informações e estatísticas, armazenados nas nuvens e construídos de forma colaborativa.

A capacidade de processamento paralelo, em clusters e grids de computadores (e ainda processos de comunicação machine to machine) somada à inteligência artificial tem proporcionado análises mais complexas, levantamento pela força bruta de relações escondidas, validação de fatos e modelos, e a captura de dados reais em ambientes reais.

Fatos X Opiniões

Em toda comunicação as pessoas tendem a misturar emoções, opiniões, interpretações pessoais e sua própria experiência. Na verdade, quando falamos sobre algo, estamos dando o nosso ponto de vista, o qual está carregado de sentimos e experiências. Quando ouvimos, temos a tendência de procurar contextualizar as informações no nosso mundo, conforme o que já sabemos.

Por exemplo, imagine que você está numa sala de aula e uma pedra atravessa o vidro e cai perto do quadro. Este é o fato, mas cada pessoa presente na sala irá relatar este fato de uma maneira diferente. Alguns vão dizer: "sabe ... hoje lá na aula alguém jogo uma pedra pelo vidro". Outro dirão: "hoje quebraram o vidro da sala de aula com uma pedra". E alguns podem dizer assim: "olha... quase morri hoje com uma pedrada". Isto acontece porque as pessoas não transmitem fatos e sim as suas impressões em relação aos fatos.

O Dicionário Oxford elegeu como a palavra do ano em 2016: pós-verdade. Pós-verdade é um neologismo que descreve a situação na qual, na hora de criar e modelar a opinião pública, os fatos objetivos têm menos influência que os apelos às emoções e às crenças pessoais (Wikipédia). Isto pode ser notado nas discussões em redes sociais. As pessoas compartilham opiniões e não fatos.

Segundo experimentos de Skinner em laboratório (citado por Shermer, 2011), "a mente humana procura relações entre eventos e com frequência as encontra, mesmo quando elas não estão presentes... Pessoas inteligentes acreditam em coisas estranhas porque são competentes para defender crenças que elas acabaram adotando por razões não inteligentes."

Shu e outros (2017) dizem que as pessoas tendem a acreditar em conhecimentos que condizem com suas crenças e conhecimentos prévios. As pessoas formam grupos ideológicos onde opiniões similares ecoam. E todos nós temos a propensão a acreditar em conhecimentos que vêm de pessoas de nossas relações e que são tidas como confiáveis por nós. Além disto, quando ouvimos repetidamente uma informação, temos a tendência a acreditar nela.

Ecker e outros (2011) concluem que repetir a informação falsa aumenta a familiaridade, reforçando-a (mesmo sendo para corrigi-la). Há uma frase atribuída a Joseph Goebbels, ministro da propaganda de Hitler: "uma mentira contada mil vezes se torna uma verdade".

Portanto, quando recebemos informações de pessoas devemos coletar evidências ou sinais que corroborem o que foi dito, seja através de observação do ambiente ou por outras fontes.

Viés da conformidade

O viés da conformidade ou da confirmação acontece porque costumamos dar maior ênfase aos argumentos que reforçam nossas crenças. Argumentos contrários são muito mais difíceis de serem aceitos. Conforme Mlodinow (2009): "Se os detalhes que recebemos se adequarem à imagem mental que temos de alguma coisa, então, quanto maior o número de detalhes numa situação, mais real ela parecerá."

Para evitar este tipo de problema, é recomendado que um grupo de pessoas tenha pontos de vista diversos para melhorar a discussão e troca de ideias. A técnica dos 6 chapéus é útil para isto. Segundo ela, o grupo deve ter pessoas que desempenhem diferentes papeis, por exemplo:
- o chapéu AMARELO representa o pensamento lógico, o estudo de viabilidade e a visão positivista através da argumentação com benefícios;
- o chapéu PRETO representa a visão pessimista e cautelosa, a análise crítica, a suposição de problemas, a identificação de riscos e obstáculos;
- o chapéu VERDE representa a criatividade, as novas ideias, a sugestão de diferentes alternativas;
- o chapéu BRANCO representa o uso de dados, estatísticas, a análise de fatos;
- o chapéu VERMELHO representa o uso de emoções, intuições, sentimentos e palpites;
- o chapéu AZUL representa a organização, o controle, a coordenação do processo e a visão holística.

Pode-se notar que o chapéu branco se contrapõe ao vermelho quanto ao embasamento para uma análise ou decisão. Ambos são salutares e devem equilibrar forças, para que o resultado não penda para um lado somente. O mesmo acontece entre o chapéu amarelo (otimista) e o chapéu preto (pessimista). O amarelo também conflita com o verde e o vermelho por ser mais "pé no chão". E o chapéu azul deve integrar os demais para que haja equilíbrio e um resultado melhor seja alcançado.

Dificuldade para corrigir dados errados

Segundo Shu e outros (2017), estudos psicológicos concluem que a correção de informações falsas não ajuda muito e pode até mesmo aumentar os enganos. Segundo Ecker e outros (2011), mitos simples são mais atrativos (cognitivamente) do que refutações complicadas. A dica é corrigir com histórias curtas. Ecker e outros (2011) e Lewandowsky e outros (2012) dizem que o conhecimento errado, mesmo após sua correção, continua a influenciar a memória e o raciocínio. É necessária uma forte retratação, usando explicações

diferentes, usando poucos argumentos e simples e colocando ênfase nos fatos, evitando assim repetir ou reforçar o mito.

Ecker, Lewandowky e parceiros (2011, 2012) alertam também que, mesmo a tentativa de corrigir uma informação acaba reforçando a informação original, mesmo ela sendo errada, simplesmente porque ela está sendo repetida, aumentando a familiaridade e reforçando-a.

Integridade de Dados

Organizações que trabalham com bancos de dados costumam utilizar restrições de integridade para evitar que sejam armazenados dados sem qualidade. Por exemplo, num hospital, se um procedimento de cesariana foi feito, então o sexo do paciente tem que ser feminino. Este tipo de restrição pode ser aplicada no momento em que uma operação é feita sobre os dados (por exemplo, inclusão de um registro de cesariana) ou pode ser verificada num certo momento sobre o estado atual dos dados (por exemplo, mesmo depois que a operação no banco de dados já foi feita). Quando a verificação for feita na execução de uma operação sobre dados, a restrição pode ser verificada antes ou depois da alteração.

As restrições de integridade podem ser definidas com recursos como triggers e stored procedures em Sistemas de Gerência de Bancos de Dados (SGBDs) e controlam automaticamente sem necessidade de intervenção humana as operações que estão sendo feitas sobre os dados. Podem ser criadas restrições tais como:

- sobre estados de dados: ex. salário não pode ser negativo;
- sobre tipos de dados: ex. mês na data deve ser armazenado como número;
- sobre formatos de dados; ex. ano de nascimento deve ser armazenado com 4 dígitos;
- sobre transições de dados ou estados: ex. salário não pode diminuir;
- sobre associações entre dados: ex. se a cidade é Caxias do Sul, então a UF tem que ser RS;
- sobre operações: ex. um usuário de Biblioteca não pode retirar mais de 3 livros ao mesmo tempo;
- sobre temporalidade dos dados: ex. contas inativas por mais de 5 anos devem ser excluídas;
- sobre permissões de acesso a dados: ex. cadastros de clientes só podem se excluídos por gerentes de lojas;

O maior desafio é pensar e definir as diversas restrições que devem ser controladas sobre os dados. Isto exige conhecimento sobre o domínio dos dados e experiência sobre o histórico dos dados.

Uma forma de fazer avaliação automática da integridade dos dados é extrair padrões estatísticos dos dados através de técnicas de Data Mining, como análise de distribuição e associações. Um exemplo do primeiro tipo é “30% dos clientes são mulheres” e um exemplo do segundo é “90% dos clientes homens compram no mínimo 2 produtos”.

Após esta mineração, entra o trabalho humano de verificar as regras e padrões descobertos. Padrões com probabilidade muito alta podem indicar que há exceções que podem ser erros. Por exemplo, se encontrarmos um padrão em que "99,9% das vendas que continham o produto X foram feitas na loja Y", seria interessante verificar o que está acontecendo com os 0,1%. Pode ser que este produto só seja vendido na loja Y e que os 0,1% de exceção sejam erros de registro nos dados de vendas.

Na verdade, foi desta forma que se iniciou o Data Mining. As empresas queriam avaliar a qualidade dos dados mas não tinham condições de definir todas as regras de integridade. Então desenvolveram algoritmos para analisar estatisticamente os dados e apontar as probabilidades dos padrões encontrados. Regras banais como "se a cidade é Rio de Janeiro, a UF é RJ" eram encontradas. Se a probabilidade fosse 100% estava tudo certo. Se não fosse, havia problemas de integridade. O "pulo do gato" foi descobrir regras interessantes que não eram conhecidas. Esta é a origem dos algoritmos de Data Mining.

PARTE V – VALOR DOS DADOS

O valor da informação se dá de várias formas. Primeiro, a informação é importante para si (pessoa física ou empresa). É por meio dela que empresas melhoram sua gestão, criam inovações, entendem e atendem melhor seus clientes e alcançam qualidade através de métodos de melhoria contínua. Já as pessoas aumentam suas competências através de treinamento e outras formas de aquisição de conhecimento.

Mas informações também têm seu valor ao permitir que pessoas ou entidades manipulem ou influenciem outros (pessoas ou entidades). Começa pela manipulação da informação, inclusive com distorções da História (como previu George Orwell em "1984"). E termina pela manipulação das pessoas alvo, por meio de informações e métodos selecionados para isto.

A habilidade para exercer controle sobre outros depende de: (a) habilidade para constituir redes e programar a rede em termos de objetivos estabelecidos para a rede, e (b) habilidade para conectar e assegurar cooperação de diferentes sub-redes através do compartilhamento de objetivos comuns e da combinação de recursos, enquanto diminuindo a competição através de alianças. Para programar redes, é preciso implementar comunicação e persuasão alinhada aos projetos dos possíveis participantes. Outra fonte de poder vem do controle de pontos de conexão entre sub-redes, formados pelos expansores de fronteiras (Castells, 2011).

Castells (2011) resume em quatro tipos o poder das redes sociais:
1. Poder em rede (Networking Power): poder dos atores e organizações que participam da rede sobre aqueles que não estão na rede;
2. Poder da rede (Network Power): o poder imposto pelas redes sobre participantes, sobrepondo regras de participação;
3. Poder enredado (Networked Power): o poder de atores sociais sobre outros que participam da mesma rede;
4. Poder de criar redes (Network-making Power): poder de governar redes específicas de acordo com interesses e valores de quem governa.

É tanto verdade que as redes sociais podem alterar o comportamento e humor das pessoas que o Facebook fez um experimento em 2012, manipulando o feed de notícias de alguns usuários, aumentando ou diminuindo postagens positivas e negativas. A conclusão foi que realmente as pessoas se sentiram mais felizes ou tristes (detectado pelas postagens originadas pelo próprio usuário). Ver reportagem na Forbes:
http://www.forbes.com/sites/kashmirhill/2014/06/30/facebook-only-got-permission-to-do-research-on-users-after-emotion-manipulation-study/

Castells (2003) acredita que a Internet passou de uma ferramenta para produtividade organizacional e de um meio de comunicação para uma alavanca de transformação social.

Podemos ver isto na prática pelos eventos tipos Flash mob que são organizados por redes sociais, e mais recentemente, as manifestações ocorridas em junho de 2013 no Brasil, as quais foram organizadas pelas redes sociais. A campanha de Obama em 2012 utilizou diferentes meios de contato: cartas, site na Internet, twitter, telefone e também as velhas visitas pessoais com direito a chá com bolacha.

Conforme Castells, “o que muda atualmente é que os cidadãos têm um instrumento próprio de informação, auto-organização e automobilização que não existia. Antes, se estavam descontentes, a única coisa que podiam fazer era ir diretamente para uma manifestação de massa organizada por partidos e sindicatos, que logo negociavam em nome das pessoas. Mas, agora, a capacidade de auto-organização é espontânea. Isso é novo e isso são as redes sociais. E o virtual sempre acaba no espaço público. Essa é a novidade. Sem depender das organizações, a sociedade tem a capacidade de se organizar, debater e intervir no espaço público."

Quem paga por informação?

A informação é diferente de todos os outros ativos organizacionais. Materiais, energia, recursos humanos e materiais são perdidos à medida que os usamos ou os repassamos a outros. Mas com a informação é diferente: podemos fazer cópias. E quanto mais usamos informações, mais ganhamos informação. A análise de informações permite descobrir padrões, regras, exceções, e isto tudo pode levar a aumento de vendas ou a encontrar um nicho de mercado. E como um bem corporativo, a informação pode ser comprada e vendida.

Nós, clientes, pagamos por informação. Pagamos para obter conteúdo de canais de TV, para obter música, notícias, filmes, etc. As empresas pagam para obter cadastros de clientes potenciais, para obter orientações de consultorias, para obter dados de mercados e concorrentes (inteligência competitiva), para ter fotos para suas publicidades, etc.

Se estamos recebendo algo de graça, o produto somos nós e nossos dados. Harari (2018) lembra que “as pessoas ficam contentes de ceder seu ativo mais valioso — seus dados pessoais — em troca de serviços de e-mail e vídeos de gatinhos fofos gratuitos.” Como já discutimos acima, o capital de relacionamento, ou seja, as informações dos clientes são uma moeda de troca muito valiosa. Pode determinar que uma empresa adquira outra, pode fazer com que empresas pequenas ganhem dinheiro vendendo informações. E pode ser útil para pessoas físicas ganharem “brindes” em troca de informações.

Yuval Noah Harari (2018) lembra que “vender anúncios pode ser necessário para sustentar os gigantes [empresas como Google e Facebook] no curto prazo, mas eles frequentemente avaliam aplicativos, produtos e companhias em função dos dados que colhem deles, e não do dinheiro que eles geram. Um aplicativo popular pode não ter um bom modelo de negócios e até mesmo perder dinheiro no curto prazo, mas na medida em que absorver dados pode valer bilhões.”

Outra forma de transformar informações em dinheiro é através de patentes. Segundo o INPI (Instituto Nacional de Propriedade Industrial), patente é "um título de propriedade temporária sobre uma invenção ou modelo de utilidade, outorgados pelo Estado aos inventores ou autores ou outras pessoas físicas ou jurídicas detentoras de direitos sobre a criação. Em contrapartida, o inventor se obriga a revelar detalhadamente todo o conteúdo técnico da matéria protegida pela patente".

O sistema de patentes iniciou no século XV em Veneza, segundo Chang (2004), quando artesãos estrangeiros com conhecimento especializado foram atraídos para o local. Para que pudessem vir, seu conhecimento precisava ser protegido. Antes mesmo da Revolução Industrial, o mundo já sentia a necessidade de impor direitos de propriedade, especialmente para proteger os resultados dos grandes inovadores.

O objetivo das patentes é proteger invenções, principalmente permitindo ao inventor obter retorno pelo custo e esforço dispendidos no desenvolvimento da invenção. Isto tende a incentiva a inovação, uma vez que custos e esforços investidos na invenção não se percam com cópias desautorizadas. Por outro lado, a patente permite a difusão da invenção, para que outros utilizem pagando royalties.

Desta forma, se temos uma patente, podemos licenciá-la. Isto é como vender informação. Só que ainda continuamos com ela e com os direitos dela.

Custo da Informação

A informação não é obtida sem custo. Há custo para coletar, guardar ou analisar dados. Há custo para gerar inovações. Há custo para gerenciar informações e pessoas.

O custo de pensar talvez seja o mais caro. Por isto Duhigg (2012) disserta sobre o poder do hábito. Os hábitos nos aliviam de ter que raciocinar a todo momento. Em especial, também ajudam quando se precisa de velocidade para tomar decisões. Mas mesmo um processo de treinamento de hábitos possui custo, pela repetição necessária (curva de aprendizado). Mas o resultado pode ser compensador: hábitos podem aumentar a produtividade e diminuir erros (aumentar a qualidade).

Por outro lado, Abrahamson e Freedman (2008) comentam no livro "Uma Bagunça Perfeita", que nem sempre vale a pena gastar tempo organizando coisas. O certo seria fazer uma avaliação comparando o custo da organização com a redução de tempo e outras coisas como resultado da organização.

Anderson (2011) (no livro "Grátis") lembra a história do almoço grátis: "os subsídios cruzados constituem a essência da frase 'não existe almoço grátis'. Isso significa que, de uma forma ou de outra, a comida precisa ser paga, se não diretamente por você, então por alguém cujo interesse é dar-lhe comida de graça." Um exemplo de "subsídios cruzados" é quando uma empresa dá informações de graça. Na verdade, ela está ganhando em

publicidade ou venda de informações sobre os clientes. Antes eram rádio e TV, agora são os blogueiros. “A atenção que você dá a um blogueiro, seja numa visita ou com um link, aumenta a reputação dele... Ele pode usar essa reputação para conseguir um emprego melhor, ampliar sua rede de contatos ou encontrar mais clientes. Algumas vezes, esses créditos de reputação podem transformar-se em dinheiro, mas raramente é possível prever a trajetória exata – é diferente a cada vez.” (Anderson, 2011).

Quantidade de informação

Conforme a Teoria da Informação de Shannon, a quantidade de informação está associada ao inesperado. Quanto mais improvável a informação recebida, mais quantidade ela terá. Isto quer dizer que uma afirmação como “2 + 2 = 4” tem menos informação que “amanhã haverá um terremoto no lugar X”. Shannon foi além disto definindo uma fórmula matemática para medir a quantidade de informação.

Além disto, ter ou não ter uma informação pode resultar em ganhos ou perdas. Veja o caso de informações privilegiadas na Bolsa de Valores. Isto é proibido porque pessoas que possuem estas informações (tidas como sigilosas) possuem vantagens sobre as demais.

Akerlof (1970) explicou bem este tipo de problema falando da assimetria de informações. Ele usou como exemplo o processo de compra e venda de carros usados. Se uma pessoa conhece melhor do que outra o carro sendo negociado, então esta primeira tem vantagem na negociação.

O valor da informação que não temos também fica claro no exemplo dado por Ellenberg (2015) quando o exército americano estava analisando furos de balas nos aviões que retornavam. Havia muitos tiros em várias partes menos no motor. Então a blindagem foi feita na região do motor, ou seja, nas partes onde não havia furos, assumindo-se que as balas alemãs atingiam mais o motor e isto fazia com que os aviões não retornassem.

Utilidade esperada

Ellenberg (2015) explica como devemos analisar dados probabilísticos para avaliar uma aposta ou risco. A utilidade esperada de uma alternativa é calculada como:

(valor positivo a ganhar * probabilidade de acontecer) +
(valor negativo a perder * probabilidade de acontecer)

Por exemplo, se há uma possibilidade de ganhar 200 mil gastando 100 mil em apostas, vamos assumir que os 200 mil valem 400 UTIs (unidades úteis). Se o resultado de perder 100 mil gera -1.000 UTIs (pois resultaria em muitas dívidas e na impossibilidade de conseguir pagar), temos que
Utilidade esperada = (400 * 50%) + (-1000 * 50%) = 200 – 500 = - 300

E a conclusão é que a aposta não vale a pena pois deu resultado negativo.

Qualidades da Informação

Só quantidade de informações não basta. É preciso que as informações tenham qualidades. Eis algumas:

- Objetividade e simplicidade: já discutimos que a sobrecarga de informações não ajuda mas atrapalha. Deve haver foco para informar;
- Granularidade adequada: na mesma linha acima, relatórios podem ser abrangentes ou resumidos. Por exemplo, podemos ver as vendas de uma empresa totalizadas por país, região, cidade ou por filial. O nível de detalhe (granularidade) depende da necessidade;
- Precisão: um número pode ser apresentado com várias casas decimais; qual é o nível mais adequado?
- Completeza: a falta de informações deturpa decisões como já discutido antes;
- Atualidade: informação obsoleta não tem valor, como já discutido antes;
- Disponibilidade: a informação deve estar pronta para uso, sem sujeiras, inconsistências e no formato adequado.

Medidas de Interestingness

Como medir se uma informação é interessante ou não ? Que medidas podem ser usadas ?

Podemos começar usando a novidade (por exemplo, a quantidade de informação sugerida por Shannon, como discutido antes). Outra medida é a probabilidade de ser verdade ou de acontecer.

Padrões (ou seja, o que acontece com a maioria) são informações interessantes. Mas as exceções podem ser úteis em alguns casos. Em geral, o interessante é algo inesperado, que contradiz as expectativas. Pode ser um padrão (ordem) para a maioria dos casos ou simplesmente algo que sai do padrão, como uma exceção. Descobrir que a maioria dos clientes de um supermercado compra em média 2 kg de feijão é interessante. Mas também é interessante observar quem está comprando abaixo ou acima disto. O que sai da média, o que está fora do previsto, também pode ser interessante.

Para tanto, precisamos de um sistema de crenças, com conceitos básicos ou primitivos ou atômicos que formem um senso comum (ou conhecimento comum ou ordinário). Alguma coisa que, quando solta no ar, sobe contradiz nossos conhecimentos sobre gravidade. Isto é algo interessante que merece ser investigado.

Os povos são cheios de crenças populares e superstições (sabedoria popular). Isto poderia ser incorporado num sistema de crenças, para ajudar a descobrir contradições ou exceções.

Ou então a empresa poderia gerar um conjunto de regras de negócio e comparar com padrões encontrados em seus dados. Por exemplo, houve o caso de uma empresa de BPM (Business Process Management) que aplicou Data Mining em nas medições de processos. Ela descobriu uma sequência muito frequente de tarefas que ia contra suas regras de negócio. Ela admitia exceções em alguns processos, mas a exceção ser mais frequente que a regra, isto sim era interessante.

Geng e Hamilton (2006) propõem 9 critérios para determinar se um padrão é interessante ou não. Aí vão eles:

- concisão: um padrão que trata de poucos atributos é mais interessante porque é mais fácil de ser entendido; por exemplo, o que se entende uma regra que diz que "89% dos clientes que compram refrigerante, carne, salada e leite num supermercado, também compram queijo" ? Agora, se a regra for "89% dos clientes que compram presunto também compram queijo", aí fica mais fácil de se entender o padrão e tomar algumas atitudes;
- cobertura ou generalidade: um padrão é geral se cobre um conjunto grande de dados; é o caso contrário da exceção como já discutido antes;
- confiabilidade: um padrão é confiável se tem suporte maior, ou seja, se ocorre com alta frequência ou percentual (em grande parte dos casos);
- raridade: um padrão é interessante se se distancia muito dos demais padrões (é caso das exceções);
- diversidade: um padrão que trata de atributos bem diferentes dos que são tratados em outros padrões é considerado diverso e por isto tem um certo grau de importância;
- novidade: se um padrão não puder ser inferido de outros padrões, então ele é interessante;
- surpresa: é o caso já comentado de contradizer as crenças ou expectativas;
- utilidade: é útil se contribui para alcançar um objetivo;
- aplicabilidade: se ajuda em alguma tomada de decisão ou em ações futuras.

Capital Intelectual

Quanto vale uma empresa ? Como determinar o valor de uma empresa ? Atualmente, o valor de uma empresa não se dá somente pela soma de seus bens materiais e ativos financeiros. Existem 3 tipos de Capital: bens (móveis e imóveis), recursos financeiros e conhecimento (o chamado capital intelectual).

Segundo Thomas Stewart (1998), Capital Intelectual é o conhecimento útil que gera valor para a empresa. Portanto, não é qualquer conhecimento. Thomas Stewart cita alguns autores que acreditam que os ativos intelectuais valham 3 ou 4 vezes mais que o valor contábil da empresa.

No Brasil, os ativos intangíveis já podem até mesmo estar presentes na contabilidade oficial de uma empresa. A Figura 29 apresenta tipos de ativos intangíveis (Kayo et al., 2006).

Figura 29: Tipos de Ativos Intangíveis

Família	Alguns ativos intangíveis
Ativos Humanos	• conhecimento, talento, capacidade, habilidade e experiência dos empregados; • administração superior ou empregados-chave; • treinamento e desenvolvimento; • entre outros.
Ativos de Inovação	• pesquisa e desenvolvimento; • patentes; • fórmulas secretas; • *know-how* tecnológico; • entre outros.
Ativos Estruturais	• processos; • *softwares* proprietários; • bancos de dados; • sistemas de informação; • sistemas administrativos; • inteligência de mercado; • canais de mercado; • entre outros.
Ativos de Relacionamento (com públicos estratégicos)	• marcas; • logos; • *trademarks*; • direitos autorais (de obras literárias, de *softwares*, etc); • contratos com clientes, fornecedores, etc.; • contratos de licenciamento, franquias, etc.; • direitos de exploração mineral, de água, etc.; • entre outros.

Fonte: Kayo (2002, p. 19).

Os tipos de capital intelectual são:

- Capital Humano:

inclui conhecimento e competências individuais dos funcionários (refere-se ao valor que a empresa perde quando funcionários vão embora); inclui também ATITUDES (saber/poder/quere fazer)
Ex.: um time de futebol formado por jogadores com baixos salários pode ganhar de um time só de "estrelas"; uma startup com poucos funcionários pode valer mais que uma empresa com centenas ou milhares de funcionários.

- Capital Estrutural ou Organizacional:

inclui conhecimento ou competência coletiva, como processos, know-how, marcas e patentes, documentos (é o valor que fica quando funcionários vão embora);
Ex: organização e processos da Fedex ou Amazon para entregas certas e rápidas; a forma de colaboração entre funcionários de uma empresa ágil e produtiva.

- Capital do Cliente ou de Relacionamentos:

inclui conhecimento e vantagens advindas dos clientes (é o valor que se ganha com relacionamentos com clientes).

Ex: o valor do Youtube antes da compra pelo Google se deu pelo número de usuários.

Os dados são o novo petróleo ("data is the new oil"). Seja para organizar uma equipe, seja para desempenhar um processo de maneira mais rápida e barata, seja para gerar melhorias, o conhecimento é o tipo de bem ou capital que faz a diferença hoje em dia.

Informação é poder - o capital atual

"Capital" é tudo que dá vantagem de uns sobre outros. Quando os humanos eram nômades, o capital era a força, o fogo e as armas. Com o advento da agricultura, o capital passou a ser a terra. Que tinha terra, podia produzir, ter comida, excedentes para troca. E isto durou toda a idade média, inclusive com guerras por terras.

Depois, o capital passou a ser quem tinha dinheiro para fazer comércio ou para emprestar (os bancos). Com a revolução industrial, o capital passou a ser a posse de máquinas (bens de produção). Atualmente, o capital é a informação e o conhecimento. Tudo isto é bem explicado no livro "A 3ª Onda" de Alvin Toffler (da década de 80).

"Informação é dinheiro". A informação e o conhecimento são importantes para: gerar inovações, reduzir custos, aumentar receitas, conquistar clientes, obter vantagens competitivas no mercado. Isto é bom para empresas.

Mas engana-se quem acha que a informação é o novo capital só por isto. Informação é poder. Mas como meio de manipulação. O assunto da moda é Fake News.

Ortega y Gasset ("A rebelião das massas") escreveu sobre o homem-massa, o "maria vai com as outras". E como este homem-massa é utilizado como massa de manobra para rebeliões.

George Orwell ("1984") descreveu a importância de manipular informações, mas não só sobre o presente. A grande artimanha era recontar a história de maneira a obter vantagens.

Mario Vargas Llosa ("A Civilização do espetáculo") descreve a decadência da cultura pelo mau gosto da maioria. E como pessoas comuns acabam seguindo a massa sem se dar conta.

Surowiecki escreveu sobre "A sabedoria das massas/multidões". Mas há também a burrice ou tolice das massas. Nem sempre, a voz do povo é a voz de Deus.

Malcolm Gladwell ("O ponto da virada") fala da importância dos eleitos para difundir informações. Entre eles incluem-se: os que possuem muitas conexões sociais (entre estes, os influenciadores digitais), os especialistas em assuntos e aqueles que conseguem usar de persuasão sobre outros.

Somos prosumidores de informação. Consumimos e produzimos informações. E também difundimos o que os outros produzem. O meio é a mensagem (McLuhan). Dominar os

meios de comunicação é tão importante hoje em dia quanto ter dinheiro. Se bem que, um puxa o outro.

Por que compartilhamos informações de graça ?

Se informação é poder e dinheiro, porque as pessoas dão ou compartilham informações de graça? Onde está o valor para esta pessoa ?

Se a teoria de Richard Dawkins, em "O Gene Egoísta" (2007), estiver certa, somos todos descendentes e portadores de genes egoístas. A teoria consiste em que os seres egoístas têm mais chances de sobreviver e assim passar seus genes para seus descendentes. Dawkins argumenta e defende sua teoria até mesmo com exemplos aparentes de altruísmo (quando alguém dá sua vida ou a põe em risco por causa de outros). Uma gazela que salta na frente de um predador não está querendo avisar seus semelhantes muito menos provocando o predador. Sua intenção é mostrar ao predador que ela sabe saltar alto e pode fugir facilmente, sugerindo que o predador procure outra vítima.

Há interesses egoístas que não afetam o grupo todo. Por exemplo, pessoas escrevem textos ou fazem vídeos por puro prazer ou para descarga de energia e tensões. Os estudos neurológicos de Luke Chang, Alec Smith, Martin Dufwenberg e Alan Sanfey (2011) concluem que as pessoas colaboram para não se sentirem mal (aversão à culpa). Algo como "estar de bem consigo mesmo". Eles acreditam que a cooperação humana seja motivada por sentimentos morais. Mas a moral vem do grupo e aí retornamos ao altruísmo recíproco. "A sociedade civilizada é baseada na cooperação e na confiança", dizem.

Mas há razões individuais. A pessoa que compartilha informações pode estar buscando valorização individual dentro do grupo (motivação interesseira). Então uma das razões e motivação para compartilhar informações pode estar no interesse egoísta de "querer aparecer" ou na busca pela fama. Isto pode ocorrer porque o indivíduo quer ser notado ou se diferenciar dos demais do grupo. É exatamente o comportamento inverso da gazela. A pessoa compartilha para aparentar ser melhor que os colegas, e então "cair nas graças" do chefe, para manter sua empregabilidade.

Segundo o biólogo Richard Alexander, os seres humanos cooperam para competir. Ele e o psicólogo Nicholas Humphrey compartilham a teoria de que o cérebro humano se desenvolveu rapidamente pela necessidade de nos relacionar, para sermos mais fortes. Entender as relações internas no grupo foram vitais para melhor comunicar ideias, compreender os outros e organizar o grupo para traçar estratégias contra rivais.

Um comportamento semelhante (e egoísta) é motivado para receber reconhecimento entre os colegas (melhorar sua reputação e garantir benefícios futuros). Aí os olhos que se buscam não são o do chefe, mas dos pares. Isto pode ajudar no momento de promoções por votação, ou mesmo quando avaliações são feitas pelos colegas.

Outro exemplo de motivação egoísta é compartilhar informações para receber informações de volta. Isto nos faz retomar o caso do altruísmo recíproco: o indivíduo faz porque acredita que os demais colegas também irão retribuir. Este é o princípio que alimenta o chamado Crowdsourcing. Em muitos casos, um indivíduo sozinho não consegue resolver problemas. Então o grupo, pela sabedoria das massas ou inteligência coletiva, pode gerar melhores soluções. E seria um interesse egoísta se a solução resolve um problema de quem iniciou o processo.

Há também outro comportamento egoísta que envolve o grupo. Acontece quando alguém compartilha para testar ideias ou sondar o que os outros estão pensando (coletar feedback). Isto permitira descobrir quem é quem no grupo, quais os papéis, quem são os influenciadores, quem está em cima do muro, etc.

Análise do Fluxo de Informação nas Organizações

A análise do fluxo de informações dentro de uma organização pode dar uma ideia de poder.

Minu Ipe (2003) acredita que indivíduos com baixo status e poder direcionam informação para cima, enquanto que indivíduos com mais status e poder direcionam para seus pares (informação não desce).

Ardichvili e colegas (2006) concluíram que poder e status determinam motivação para compartilhar e também a direção do fluxo. O fluxo vertical acontece mais entre pessoas que aceitam diferenças de status e hierarquias. Em geral, pessoa de níveis mais acima não participam online e barram os de nível mais baixo. Status e atribuições são determinados por idade, sexo, riqueza e experiência. Já o fluxo horizontal acontece quando o status é determinado por conquistas passadas (meritocracia).

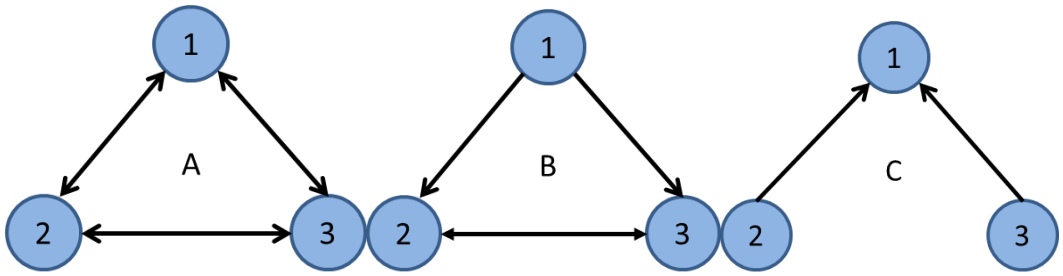
A Figura 30 representa 3 esquemas de troca de informações (A, B e C) entre pessoas (nodos ou círculos) de uma organização. As setas indicam de onde e para onde está indo a informação (seguindo sugestão de Formanski et al., 2012).

No esquema A, pode-se ver que todo mundo ajuda todo mundo fornecendo informações. Há troca entre todas as pessoas (reciprocidade total), em ambos os sentidos.

No esquema B, a pessoa 1 supre as outras duas e estas (2 e 3) também trocam informações entre si. Entretanto, 2 e 3 não fornecem informações para 1. Isto pode caracterizar que 1 está passando orientações.

No esquema C, as pessoas 2 e 3 não se comunicam. Porém, ambas suprem a pessoa 1 com informações (que podem ser relatórios). E a pessoa 1 não devolve nenhuma informação para 2 ou 3.

Figura 30: Análise do fluxo de informações



O fornecimento de informação (através de uma seta) pode indicar que uma pessoa (origem) está subordinada a outra (destino) ou que uma pessoa (origem) está passando orientações para um subordinado (destino).

A Figura 31 mostra outro exemplo. Nodos representam pessoas e arestas (setas) representam o fluxo de informação de uma pessoa para outra. As cores indicam o departamento ou setor de cada pessoa. A largura da seta representa o quanto de informação passou naquele canal.

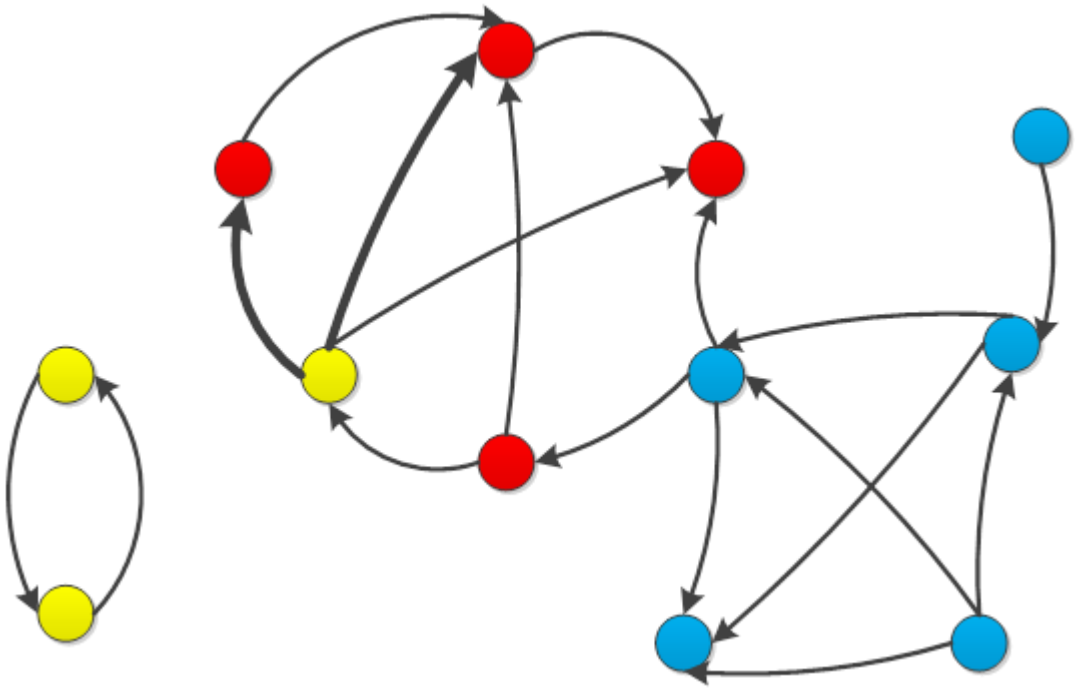


Figura 31: Grafo de comunicação entre membros de equipes

Várias análises podem ser feitas a partir deste mapa de fluxo:

a) podemos notar uma sub-rede isolada à esquerda com duas pessoas do departamento amarelo. Elas não trocam informações com pessoas de outro departamento. Além disto, há

uma pessoa do departamento "amarelo" que não interage com estas duas para trocar informações, mas que está bem "enturmada" com pessoas de outro departamento (vermelho). Isto pode indicar um problema se os 2 primeiros amarelos formam “ilhas” de informação e não compartilham com os demais (caso deveriam).

b) podemos notar que há uma pessoa (azul em baixo, bem à esquerda) que só recebe informações. Pode ser um novato, ainda aprendendo. E há alguém (azul em baixo, bem à direita) que só fornece; pode ser alguém experiente, mas será que ele ou ela não deve receber algum tipo de informação de alguma outra pessoa ?

c) uma pessoa do grupo azul está interligando duas sub-redes, a azul e a vermelha, provavelmente um elo de ligação importante para juntar duas áreas de conhecimento.

d) uma pessoa do grupo azul (em cima, bem à direita) está isolada, tendo somente contato com uma pessoa. Pode ser que seja um aprendiz.

Individualistas X Coletivistas

Ardichvili e colegas (2006) pesquisaram comportamentos de pessoas individualistas e dos coletivistas nas organizações. Os primeiros consideram-se independentes do grupo, preferem comunicação formal (escrita), buscam prestígio e reconhecimento, fazem perguntas no grupo e demonstram sentimentos negativos com os de fora do grupo. Já aquelas pessoas que possuem mais espírito coletivista colocam prioridade no objetivo do grupo, preferem comunicação não verbal e informal (face a face), têm medo da má reputação, fazem perguntas em reservado e demonstram mais modéstia (não querem parecer superiores).

Segundo Kollock e Smith (1996), os individualistas são os que fazem perguntas mas não respondem, reúnem informações mas não distribuem, participam das discussões online mas não contribuem ali com informações.

Adamic e colegas (2008) encontraram um padrão comum nos fóruns do Yahoo: nas discussões técnicas, os novatos são ajudados por outros mais experientes.

Ardichvili e colegas (2003) concluíram que empregados compartilham informações porque chegam num momento na carreira (quando atingem certo status) em que acreditam que devem retribuir, ajudando novos empregados.

Os donos dos dados

Em geral, a informação como bem (conhecimento explícito) é visto como sendo propriedade da organização. Já a informação como expertise (conhecimento tácito) é visto

como propriedade da pessoa. O mesmo conhecimento (por exemplo, um padrão estatístico) pode estar nas duas formas. Como um relatório, é visto como conhecimento da organização. Na cabeça da pessoa, é visto como conhecimento pessoal. A informação como produto é tangível e pode ser vendida. A informação como expertise é intangível e muitas vezes nem pode ser transferida de um meio ou pessoa para outro(a).

Quando algo é público ou emprestado, as pessoas não dão valor. Esta é a "tragédia dos comuns", que acontece quando um recurso é compartilhado por várias pessoas (Ostrom, 1990). O mesmo tipo de problema pode acontecer com o conhecimento público, que não é sinônimo de acesso aberto (open access), conforme Hess e Ostrom (2007).

Por isso, se a informação é propriedade de alguém, este alguém irá cuidar melhor dela (Knight et al., 1995). Por isto, umas das formas de resolver o problema da "tragédia dos comuns" é definir bem as regras de propriedade, incluindo uso e co-participação e também definindo punições.

Uma das características da informação pública é que não se pode excluir indivíduos para não vê-la ou se beneficiar dela (Kollock, 1999).

E a informação postada na Internet tem dono ? Todas as obras intelectuais, sejam livros, vídeos, filmes, fotos, obras de artes plásticas, música, etc., podem ser protegidas pelo direito autoral e não podem ser utilizadas sem prévia autorização. Além disto, qualquer texto, página na web ou site completo, quando apresentar criatividade e forma original, são protegidos. Também estão protegidos elementos individuais como imagens e sons (Martins Filho, 1998).

Os direitos autorais têm por objetivo proteger as criações originais de obras artísticas ou intelectuais, expressas por qualquer meio, sejam em livros, jornais, revistas ou meio digital. A legislação brasileira trata da proteção dos direitos autorais na lei n. 9.610/98 e no Código Civil de 2002. Além disto, os direitos autorais na Constituição Federal brasileira de 1988. O direito autoral se caracteriza por dois aspectos (Martins Filho, 1998):
1. o moral: que garante ao criador o direito de ter seu nome impresso na divulgação de sua obra e o respeito à integridade desta, além de lhe garantir os direitos de modificá-la, ou mesmo impedir sua circulação; e
2. o patrimonial: que regula as relações jurídicas da utilização econômica das obras intelectuais.

A informação em si não é protegida por direitos autorais e há o direito de citação para fins de estudo, crítica ou polêmica. Entretanto, é proibido reproduzir na íntegra obra que não pertença ao domínio público, com o pretexto de anotá-la, comentá-la ou melhorá-la. Deve-se publicar os comentários em separado. É permitido o uso de obra protegida desde que restrito a pequenos trechos e desde que a reprodução em si não seja o objetivo principal da obra nova e que não prejudique o uso comercial da obra reproduzida (Branco Jr., 2007). Em relação ao tamanho do trecho que pode ser reproduzido, não há uma quantificação pré-determinada.

Não caracteriza ofensa aos direitos, a reprodução de notícia, artigo informativo, discursos pronunciados em reuniões públicas publicadas em jornais ou revistas, desde que se mencione o nome do autor, se assinados, ou da publicação de onde foram transcritos. Também é lícito citar trechos de qualquer obra, para fins de estudo, crítica ou polêmica, desde que indicando o nome do autor e as fontes bibliográficas da obra (Martins Filho, 1998).

Muitos criticam a concentração de dados pelas empresas gigantes da Internet, como Google, Facebook, etc. Yuval Noah Harari (2018) lembra que "os gigantes dos dados poderão penetrar nos mais profundos segredos da vida, e depois usar esse conhecimento não só para fazer escolhas por nós ou nos manipular mas também na reengenharia da vida orgânica e na criação de formas de vida inorgânicas."

Segundo Yuval Noah Harari (2018), "se quisermos evitar a concentração de toda a riqueza e de todo o poder nas mãos de uma pequena elite, a chave é regulamentar a propriedade dos dados. Antigamente a terra era o ativo mais importante no mundo, a política era o esforço por controlar a terra, e se muitas terras acabassem se concentrando em poucas mãos — a sociedade se dividia em aristocratas e pessoas comuns. Na era moderna, máquinas e fábricas tornaram-se mais importantes que a terra, e os esforços políticos focam-se no controle desses meios de produção. Se um número excessivo de fábricas se concentrasse em poucas mãos — a sociedade se dividiria entre capitalistas e proletários. Contudo, no século XXI, os dados vão suplantar tanto a terra quanto a maquinaria como o ativo mais importante, e a política será o esforço por controlar o fluxo de dados. Se os dados se concentrarem em muito poucas mãos — o gênero humano se dividirá em espécies diferentes."

Por esta razão, Thompson (2018) defende que mais canais de comunicação e mais informação irão gerar mais democracia: "...marcas corporativas são mais poderosas em mercados em que os consumidores têm poucas informações ... pode ser devido ao fato de que o produto é, de alguma forma, técnico ... ou porque o produto é refinado. Assim como a TV a cabo e a Internet eliminaram o poder de autoridades políticas, desafiando a teoria de que o partido decide, o fluxo de informações da Internet também está diluindo o poder de marco de muitos produtos de consumo... Quando os consumidores não conhecem o verdadeiro valor dos produtos que estão procurando, eles confiam na iconografia corporativa para guiá-los".

Privacidade de dados

Com o desenvolvimento das tecnologias digitais, a privacidade acabou. Primeiro, há tecnologias para coleta e armazenamento de dados. Como já discutido antes, está todo mundo coletando dados sobre todos. A operadora de celular sabe por onde a gente anda e quando, qual o caminho que costumamos fazer, por onde costumamos andar em cada dia da semana e horário. E se instalarmos aplicativos tipo o Waze no nosso celular, a Google (que comprou o Waze) vai saber até a que velocidade estamos andando. E daí inferir se estamos

a pé ou de carro, ou num engarrafamento. Aí alguém inventou a tecnologia de RFID, e ela está em cartões com chips, carros, produtos novos e vai estar em sacolas, carrinhos de supermercados, etc. Então não é só por celular. Os aplicativos e softwares que usamos em celulares, tablets, notebooks e etc também estão avisando onde estamos, se estivermos conectados via Wifi, 3G ou 4G.

Em segundo lugar, há o enriquecimento de dados. Cada click na internet é monitorado. Aí eles cruzam estes dados com o que a gente comprou pela internet. Aí eles cruzam estes dados com nossos cadastros em lojas físicas. Aí eles cruzam com o que a gente comprou na loja física, fora da Internet. Aí eles complementam estes dados sobre nós com nossos perfis nas redes sociais e com o que estamos dizendo em fóruns e blogs (é o Social CRM). E é possível pegar dados públicos, disponíveis livremente na Internet. Estes dados não identificam pessoas individualmente, mas dão estatísticas sobre grupos de pessoas. Empresas parceiras também costumam compartilhar dados sobre clientes, por exemplo, administradoras de cartões de crédito, instituições financeiras, redes de varejo, escolas, postos de gasolina, editoras, etc.

Em terceiro, os dados servem para análises e inferências. Inferir é gerar uma informação a partir de outra. O nível de inferência é subjetivo de cada organização e certamente aumenta a incerteza sobre a veracidade da informação. Mas muitas empresas assumem o risco desta incerteza, porque mais incerto ainda é não saber nada sobre o cliente.

A privacidade então acabou . O que existe é política de privacidade. Onde a empresa diz que dados está coletando das pessoas, como e para quê. E a pessoa assina ou clica aceitando a política de privacidade. Dizem que apenas 0,4% dos visitantes de um site leem a política de privacidade do site. E quem lê as várias páginas com linhas pequenas nos contratos de serviços ?

Para proteger pessoas e organizar o trabalho das empresas no uso de tecnologia e coleta de dados sobre pessoas, diversos países (inclusive o Brasil) estão definindo legislação para proteção de dados privados, a chamada Lei Geral de Proteção de Dados.

Transparência nos dados

Segundo Wallin (2014), o que faz da Suécia um dos países menos corruptos do mundo é a lei de acesso público aos documentos oficiais. “Essa lei, criada na Suécia há mais de 200 anos, evita os abusos do poder. Se os cidadãos ou a mídia quiserem, podem verificar ... salário, ... gastos e as despesas de ... viagens a trabalho [de políticos].” Há uma Agência Nacional Anticorrupção cujo foco principal é o suborno. “Pode-se dizer que o suborno, tanto na esfera pública como no setor privado, é um câncer para qualquer sistema. Mesmo quando o valor do suborno é muito baixo, ele pode influenciar uma licitação no valor de um bilhão de coroas suecas. No setor público, é importante que as compras de bens e serviços sejam realizadas de modo correto. A construção de um novo hospital, por exemplo, pode custar cerca de 1,7 bilhão de coroas suecas (cerca de 260 milhões de dólares). Quando uma

agência do setor público lida com um contrato desse porte, é importante que haja uma distância entre a empresa que vai construir o hospital e os funcionários públicos que vão aprovar tal contrato."

No Brasil, segundo Wallin (2014), sob alegação de segurança nacional, "95% dos gastos presidenciais são mantidos em sigilo. Por lei, essas despesas deveriam perder o caráter confidencial ao fim do mandato, mas isso não acontece."

Por outro lado, merecem destaques algumas iniciativas brasileiras como a Lei de Acesso a Informação Pública (Lei 12.527/2011, complementada pela LC 131/2009, pelo decreto 7.724/2-12 e pela portaria institucional 233/2012), o portal da transparência (http://www.portaltransparencia.gov.br/) e o Portal Brasileiro de Dados Abertos (dados.gov.br), uma espécie de banco de dados do governo federal com mais de mil conjuntos de dados disponíveis publicamente e de maneira fácil.

Estes portais com dados públicos permitem não só aos cidadãos conhecerem o que está sendo feito e como, bem como permitem que os dados sejam analisados numa espécie de crowdsourcing (esforço coletivo e público para realizar uma tarefa), para que novos conhecimentos sejam descobertos para bem o comunitário.

PARTE VI – COMPLEXIDADE E BUSCA DE ENTENDIMENTO

O fenômeno Big Data é causado por um mundo mais complexo e também ajuda o mundo a ser mais complexo. A complexidade aumenta na proporção direta a:

- número de elementos ou itens envolvidos;
- diferentes tipos dos elementos;
- número de relações entre os elementos;
- diferentes tipos de relações entre os elementos.

Uma equipe de trabalho com homens e mulheres de diferentes profissões é mais complexo que um time de futebol masculino. Uma sala de aula com 20 alunos é menos complexa que uma sala com 40 alunos. Uma rede social que permite ligar amigos e inimigos é mais complexa que uma rede onde somente amigos estão relacionados.

O mundo está mais complexo porque:

- há mais pessoas no mundo;
- nossos grupos sociais estão aumentando, passando o limite de 150 para nossa capacidade intelectual (conforme Robin Dunbar);
- as interações entre as pessoas estão mais diversificadas usando diferentes mídias, tecnologia e redes sociais;
- nossos conceitos estão evoluindo para estados menos estáveis ou concretos. Por exemplo: como definir hoje em dia uma família ? E um casamento ? Como armazenar a filiação de uma pessoa ? Seriam 2 ou 3 valores possíveis ? Pode haver uma pessoa só com um pai ou uma mãe ? E sexo então. Há hospitais que trabalham com 12 diferentes tipos de sexo.

Por consequência disso, também temos conjuntos de dados mais complexos, com tipos mais variados, mais relações entre os dados.

Complexidade e Cooperação

O aumento de complexidade é algo natural segundo a teoria da evolução da vida. António Damásio (2018) lembra que "células nucleadas cooperam para constituir tecidos que, posteriormente, cooperam para formar órgãos e sistemas." Foi assim que os humanos evoluíram a partir de seres monocelulares, passando por multicelulares, peixes, anfíbios, macacos, etc.

Nowak (2006) identificou 5 regras da evolução da cooperação entre seres vivos:

1. Altruísmo entre parentes
2. Reciprocidade direta ("eu te coço, tu me coça")
3. Reciprocidade indireta ("alguém vai me coçar")

4. Reciprocidade de rede ou espacial (ajudar aos mais próximos)
5. Seleção de grupo (em grupo, somos mais fortes)

Wilson (2013) lembra que a seleção multinível (seleção de grupo e seleção individual combinadas) é a estratégia vencedora da evolução. “Os leões rugem, os lobos uivam e os pássaros cantam para anunciar que estão em seus territórios e desejam que os membros concorrentes da mesma espécie guardem distância... Bandos e comunidades de bandos com melhores combinações de inovações culturais tornaram-se mais produtivos e mais bem equipados para a competição e para a guerra. Seus rivais os copiavam ou então eram desalojados, tendo seus territórios tomados. Desse modo, a seleção de grupo impeliu a evolução da cultura.”

A vida em grupo exige certas abnegações. Segundo Damásio (2018), “o princípio é sempre o mesmo: organismos abrem mão de alguma coisa em troca de algo que outros podem lhes oferecer... abrem mão é da sua independência; o que recebem em troca é acesso ao ‘rossio’, aos produtos que são gerados por um arranjo cooperativo, compostos de nutrientes indispensáveis ou condições gerais favoráveis, por exemplo, acesso a oxigênio ou vantagens climáticas.”

O perigo de um cérebro centralizador, que tudo sabe e tudo controla, é acontecer de termos um Big Brother, como profetizado por George Orwell no livro “1984”.

Apesar de a evolução da vida ter gerado um cérebro centralizador para controlar os diversos elementos, há casos de sistemas complexos que funcionam bem sem um controle central. Por exemplo, as sociedades de formigas e abelhas, apesar de terem uma rainha, não precisam que esta diga o que cada um tem que fazer.

Duncan Watts (2015) utiliza o conceito de sinergia para explicar o que é um sistema complexo: "Cada gene interage com outro em uma cadeia complexa de ativação e supressão para expressar traços fenotípicos que não são reduzíveis a propriedades de nenhum gene em particular. Cada planta e cada animal interage um com o outro de formas complexas, por via de relações entre caça e caçador, simbiose, competição e cooperação, para produzir propriedades no nível de ecossistema que não podem ser entendidas em termos de uma espécie em particular."

Ordem x caos em sistemas complexos

Segundo Stewart (1991), caos é um comportamento estocástico (aleatório) que ocorre num sistema determinístico. O determinismo clássico é caracterizado, segundo Stewart (1991), por equações que estabelecem a evolução dos sistemas de uma maneira única, sem nenhuma influência externa aleatória. Desta forma, é possível entender o comportamento do sistema e até mesmo prever estados futuros. O problema é que um sistema determinístico pode se comportar de maneira aparentemente randômica.

Em alguns casos, mesmo que tenhamos a ordem de um sistema (por exemplo, as equações que descrevem seu comportamento), ainda assim será difícil prever todas as situações futuras. Em parte, isto pode acontecer devido ao "efeito borboleta" descoberto por Lorenz: pequenas diferenças levavam a alterações grandes no resultado. Por exemplo, ínfimas alterações atmosféricas, como as causadas pelo bater das asas de uma borboleta, poderiam ter um grande efeito nos subsequentes padrões atmosféricos globais (Mlodinow, 2009).

Conforme Mlodinow (2009), "o determinismo é um modelo fraco para descrever a experiência humana." Para tanto, usamos processos probabilísticos, cujo futuro é difícil de prever, mas cujo passado é fácil de entender.

Segundo Stewart (1991), podemos descobrir a ordem de um sistema complexo determinístico simples, com poucos graus de liberdade. Já para sistemas complexos, com muitos graus de liberdade, é necessário utilizar estatística: "Não podemos determinar o comportamento de cada item ou indivíduo mas podemos entender a média.". Por exemplo, o comportamento de um sistema gasoso é complexo; podemos prever o movimento de algumas moléculas mas não de todas juntas.

Waldrop (1993) escreve sobre a previsão climática: o clima nunca se repete de maneira exata; nós podemos identificar características importantes e entender sua dinâmica, mas não podemos fazer previsões completas. Mas mesmo assim estes estudos são importantes porque predição não é a essência da Ciência mas sim compreensão e explanação.

Dificuldades para identificar padrões - pessoas e sistemas complexos

O problema de prever eventos futuros é que o futuro é feito *COM* pessoas. A maioria dos modelos incluem pessoas. Se precisamos saber a causa por que máquinas quebram, temos que lembrar que elas são operadas por pessoas, pessoas fazem sua manutenção, pessoas as programam. Se quisermos prever índices de vendas, temos que lembrar que são as pessoas que compram produtos e serviços, e há também vendedores, promotores, publicitários, especialistas em moda e por fim administradores determinando preços.

Apesar de todos os autômatos embutidos em sistemas computacionais, das regras e procedimentos de qualidade, das interfaces planejadas para guiar o usuário, ainda sim as pessoas agem de forma não planejada, não prevista. E o "*ser humano é um ser racional e irracional, capaz de medida e desmedida; sujeito de afetividade intensa e instável*" (Morin, 2000, p.60)

Não temos como prever o que as pessoas farão em qualquer situação. Os genes condicionam vários comportamentos dos seres humanos (Winston, 2006; Dawkins, 2007) e muitas vezes agimos por instintos bastante primitivos, enraizados em nós nos tempos das savanas (Winston, 2006). Mas as pessoas também são condicionadas ou influenciadas pelo meio que as cerca, podem receber treinamento para fazerem algo dentro de certos procedimentos e há ainda as várias possibilidades do erro humano. E por fim, ainda há o

livre arbítrio: os genes nos moldam como roteiristas de filmes mas o resultado final é nós que decidimos, porque os genes nos dão modelos de decisão e não a decisão final (Winston, 2006; Dawkins, 2007).

Nate Silver (2013) conta o caso dos modelos utilizados pelos cientistas políticos prevendo a vitória esmagadora de Al Gore na eleição presidencial de 2000 nos Estados Unidos. Mas quem ganhou as eleições foi George W. Bush, e um dos motivos foi a cédula de votação, com marcadores mal associados aos nomes, confundindo eleitores que iriam votar em Al Gore.

A raiz do problema está em que estamos tratando com sistemas complexos. Sistemas complicados são aqueles compostos por muitas partes, mas para os quais podemos prever o resultado final, se cada parte funcionar de forma planejada. Se conhecermos os dados de entrada, as condições ambientes e o sistema funcionar segundo o padrão conhecido, ou seja, um contexto estável, é certo que saberemos o resultado final (Sargu and McGrath, 2011). Um exemplo de sistema complicado é um carro: um mecânico conhece as partes, suas interações e consegue prever o funcionamento. Se algum problema ocorrer, ele poderá determinar a causa usando seus conhecimentos e coletando alguns dados diagnósticos.

Por outro lado, sistemas complexos podem até ter poucas partes mas as interações entre as partes podem causar funções ou resultados imprevisíveis. As partes interagem de forma inesperada e por isto seu comportamento passado não pode ser usado para antecipar eventos futuros com acurácia (Sargu and McGrath, 2011). Sistemas complexos contêm interações dinâmicas e portanto as mesmas condições de entrada podem levar a resultados diferentes em tempos diferentes. Há 3 características que determinam um sistema complexo: multiplicidade (relativa ao número de elementos ou partes do sistema), interdependência (o nível de conexões entre as partes) e a diversidade (heterogeneidade dos elementos). Conforme Sagu e McGrath, quanto maior o nível de cada característica, mais complexo será o sistema. Um exemplo de sistema complexo foi a campanha (ou guerra) contra pardais na China em 1958. Os pardais estavam atacando as plantações de arroz e então o Governo chinês fez uma campanha para dizimar os pardais. O problema é que, após a eliminação dos pardais, os gafanhotos é que começaram a comer grãos de arroz, porque os pardais eram predadores naturais dos gafanhotos.

As loucuras que acontecem nos mercados econômicos e nas bolsas de valores também são resultados dos comportamentos complexos das multidões. Muitas vezes não há uma explicação lógica para a correria de venda ou compra nos mercados. Simples boatos podem se difundir rapidamente e levantar medo na população, gerando comportamentos ilógicos de indivíduos e levando as massas para direções inesperadas.

Entender o comportamento de multidões é um desafio. Conforme a teoria de Herbert Simon (1972), o ser humano toma decisões sob uma Racionalidade Limitada à informação disponível, à capacidade cognitiva das mentes e ao tempo disponível. Na maioria das vezes não vale a pena (pelo custo ou tempo) coletar todas as informações necessárias para tomar uma decisão. Por exemplo, se uma pessoa quiser comprar um sapato, pensará em verificar na cidade qual a loja com o preço mais barato. Entretanto, se for avaliar o preço de cada

loja, ao terminar o processo, terá levado tanto tempo que os primeiros preços consultados já poderão ter sido alterados e o custo total de deslocamentos e perda de tempo não valerá o desconto que conseguir. É impossível que o indivíduo conheça todas as alternativas para uma decisão e que possa avaliar todas as suas consequências. A tendência do ser humano é simplificar as escolhas. Isto quer dizer que não temos como saber se a decisão tomada foi a mais acertada antes de tomá-la; somente após saberemos se deu certo ou não. E mesmo tendo alcançado êxito, talvez não tenhamos certeza se foi a melhor alternativa.

Em geral então, as pessoas procuram diminuir a incerteza das decisões mas assumem certos riscos pela racionalidade limitada. Por exemplo, se alguém quiser traçar uma rota de fuga em caso de incêndio num prédio, talvez não consiga avaliar todas as alternativas possíveis (local de início do fogo, quantidade de pessoas, etc.). E no momento da situação de incêndio, o ser humano tem que simplificar ao máximo seu processo de decisão para acelerar as ações. Isto quer dizer que os planos iniciais podem ter sido esquecidos ou terão que ser simplificados. E assim, as atitudes planejadas mudam pela racionalidade limitada. E o ser humano se torna imprevisível. Tversky e Kahneman (1974, 1983) discutem o problema de avaliações probabilísticas erradas em decisões humanas. Eles apresentam diversos experimentos que comprovam que o ser humano avalia de forma errada muitas situações, usando modelos probabilísticos errados ou incompletos.

Além disto, a ação de uma pessoa acaba por influenciar a decisão dos que estão próximos. Isto pode modificar o comportamento dos outros, que podem imitar ou fazer algo bem diferente. Por vezes, algumas decisões de pessoas pensando no benefício próprio e único podem prejudicar ainda mais o sistema. Há o famoso caso do paradoxo de Braess, que diz que criar atalhos em redes rodoviárias pode não diminuir o tempo médio, porque a maioria das pessoas irá escolher o atalho, gerando novos engarrafamentos. Tomar decisões de forma independente, talvez não seja a melhor alternativa, conforme a teoria do Equilíbrio de John Nash. Talvez a melhor alternativa para todos seja cada um “perder” um pouco de algo para todos “ganharem”.

As técnicas relativas à Teoria dos Jogos ajudam a entender os resultados nestes tipos de sistemas complexos. A Teoria dos Jogos é uma tentativa de tentar prever resultados em sistemas complexos. Através da análise da combinação de diferentes estratégias dos jogadores (componentes do sistema que possuem poder de decisão), pode-se prever os resultados possíveis. A dificuldade está em prever as decisões que serão tomadas.

Apesar das dificuldades, das incertezas, mesmo assim precisamos procurar padrões para entender a complexidade dos sistemas. Isto nos ajudará em situações futuras, mesmo que não nos permitindo chegar a previsões exatas. Ghani e Simmons (2004), por exemplo, conseguiram prever com 96% de acerto, o preço final em leilões no eBay, um tipo de situação bastante complexa, envolvendo diversas variáveis e além disto intuições, sentimentos, percepções e escolhas humanas.

Antifrágil para sobreviver no caos e complexidade

Se não é possível prever ou controlar o futuro ou um sistema complexo, então que estejamos preparados para o que vier.

Esta é a lição de Taleb (2014) no livro “Antifrágil”. Sistemas frágeis quebram quando sofrem pressão. Sistemas resilientes não se modificam sob pressão. O antifrágil se modifica para se adaptar às pressões, tirar proveito e se tornar melhor.

Esta é também a ideia do pensamento não linear: a direção em que você deve ir depende de onde você já está (Ellenberg, 2015).

BIBLIOGRAFIA

ABRAHAMSON, Eric; FREEDMAN, David H. Uma Bagunça Perfeita. Rocco, 2008.

ADAMIC, Lada A. et alli. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. WWW Conference, Beijing, China, April 2008.

AGRAWAL, R.; SRIKANT, R. Mining Sequential Patterns. In: Proceedings of the International Conference on Data Engineering, 11º, 1995, Taipei, Taiwan, p. 3-14.

AKERLOF, George A. The Market for "Lemons: Quality Uncertainty and the Market Mechanism. The Quarterly Journal of Economics, v.84, n.3, Agosto de 1970, p.488-500.

ALVES, Renato. O cérebro como foco e disciplina. São Paulo: Editora Gente, 2014.

ALLPORT, Gordon W.; POSTMAN, Leo J. The basic psychology of rumor. Transactions of the New York Academy of Sciences, series II, 1945, p.61-81.

ALLPORT, Gordon W.; POSTMAN, Leo. The Psychology of Rumor. Henry Holt and Company, 1947.

ALSTOTT, Jeff; MADNICK, Stuart; VELU, Chander. Predictors of Social Mobilization Speed. http://arxiv.org/ftp/arxiv/papers/1303/1303.3805.pdf

ANDERSON, Chris. A Cauda Longa - do mercado de massa para o mercado de nicho. Rio de Janeiro: Campus, 2006.

ANDERSON, Chris. Free - grátis: o futuro dos preços. Rio de Janeiro: Elsevier, 2011.

ANSOFF, H. Igor. Strategic issue management. Strategic Management Journal, v.1, n.2, April/June 1980, p.131–148.

ARAL, Sinan; WALKER, Dylan. Identifying Influential and Susceptible Members of Social Networks. Science, v.337, n.6092, 20 de Julho de 2012, p.337-341.

ARBESMAN, Samuel. The Half-Life of Facts: Why Everything We Know Has an Expiration Date. Penguin, 2013.

ARDICHVILI, Alexander; PAGE, Vaughn; WENTLING, Tim.Motivation and barriers to participation in virtual knowledge-sharing communities of practice. Journal of Knowledge Management, v.7, n.1, 2003, p.64.

ARDICHVILI, Alexandre et alli. Cultural influences on knowledge sharing through online communities of practice. Journal of Knowledge Management, v.10, n.1, 2006, p. 94-107.

ASUR, Sitaram; HUBERMAN, Bernardo A. Predicting the Future with Social Media. Proceedings WI-IAT '10 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - v.1, 2010, p. 492-499.

AYRES, Ian. Super Crunchers: por que pensar com números é a nova maneira de ser inteligente. Ediouro, 2008.

BARABASI, Albert-Laszlo; BONABEAU, Eric. Scale-free networks. Scientific American, n.288, Maio de 2003, p.50-59.

BARABASI, Albert-László. Linked: How Everything is Connected to Everything Else and What It Means for Business, Science, and Everyday Life." New York: Plume, 2003.

BARAN, Paul. On distributed communications networks. First Congress of the Information Systems Sciences. Novembro 1962.

BAUMAN, Zygmunt. Vida a crédito: Conversas com Citlali Rovirosa-Madrazo. Zahar, 2010.

BERTIN, Jacques. Semiology of Graphics: Diagrams, Networks, Maps. University of Wisconsin Press, 1983.

BLACKMORE, Susan. Evolution and memes: the human brain as a selective imitation device. Cybernetics and Systems, v.32, 2001, p.225-255.

BLOOM, Paul. Who Cares about the Evolution of Stories? Critical Inquiry, v.38, n.2, Winter 2012), p. 388-393.

BOLLEN, Johan; MAO, Huina; ZENG, Xiao-Jun. Twitter mood predicts the stock market. Journal of Computational Science, 2(1), March 2011, 1-8.

BOWLES, Samuel. Individual interactions, group conflicts and the evolution of preferences. Em: Steven Durlauf and Peyton Young, Social Dynamics. Cambridge, MIT Press, 2000.

BRANCO Jr., Sérgio Vieira. Direitos autorais na Internet e o uso de obras alheias. Rio de Janeiro, Lumen Juris, 2007.

BRYNJOLFSSON, Erik; McAFEE, Andrew. Race Against The Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy. Digital Frontier Press, 2011.

BURKE, Peter. Problemas causados por Gutenberg: a explosão da informação nos primórdios da Europa moderna. Estudos Avançados, v.16, n.44, Jan./Apr. de 2002.

CAMPBELL, Joseph. O herói de mil faces. 11ª ed. Pensamento, 1995. (original de 1949)

CAMPBELL, Joseph. Isto és tu - Redimensionando a metáfora religiosa. Landy Editora, 2002.

CANTOR, Mauricio; WHITEHEAD, Hal. The interplay between social networks and culture: theoretically and among whales and dolphins. Philosophical Transactions of the Royal Society, n.368, Abril de 2013.

CAREY, Benedict. Como aprendemos: a surpreendente verdade sobre quando, como e por que o aprendizado acontece. Elsevier, 2015.

CARR, Nicholas. Is Google Making Us Stupid? What the Internet is doing to our brains. The Atlantic, Julho/Agosto de 2008.

CARR, Nicholas. The Shallows: What the Internet Is Doing to Our Brains. W. W. Norton & Co., 2010.

CASTELLS, Manuel. A Galáxia Internet: Reflexões sobre a Internet, os Negócios e a Sociedade. Rio de Janeiro: Jorge Zahar Ed., 2003.

CASTELLS, Manuel. A Network Theory of Power. International Journal of Communication, v.5, 2011, p.773–787.

CHANG, Ha-Joon. Chutando a Escada: a estratégia do desenvolvimento em perspectiva histórica. São Paulo: Editora UNESP, 2004.

CHANG, Luke J.; SMITH, Alec; DUFWENBERG, Martin; SANFEY, Alan G. Triangulating the Neural, Psychological, and Economic Bases of Guilt Aversion. Neuron, v.70, n.3, Maio de 2011, p.560-572.

CHOI, H.; VARIAN, H. Predicting the Present with Google Trends. Economic Record, special issue selected Papers from the 40th Australian Conference of Economists, v. 88, n.1, p.2–9, June 2012.

CHRISTAKIS, N. A.; FOWLER, J. H. Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives. Nova Iorque: Little, Brown and Company, 2009.

CHRISTIANSON, Scott. 100 diagrams that changed the world: from the earliest cave paintings to the innovation of the ipod. New York: Penguin books, 2012.

CLARK, Gregory. A farewell to alms: a brief economic history of the world. Princeton University Press, 2007.

CROSSAN, John Dominic. O Jesus histórico: a vida de um camponês judeu no mediterrâneo. Rio: Imago, 1994 (traduzido do original inglês The Historical Jesus: the life of a Mediterranean Jew peasant, 1991).

DAMASIO, Antonio. Descartes' Error: Emotion, Reason, and the Human Brain Paperback. Penguin Books, 2005.

DAMASIO, António. A estranha ordem das coisas: As origens biológicas dos sentimentos e da cultura. Cia das Letras, 2018.

DAWKINS, Richard. O Gene Egoísta. Companhia das Letras, 2007.

DIAMANDIS, Peter H.; KOTLER, Steven. Abundância: o futuro é melhor do que você imagina. HSM Editora, 2012.

DUGAS, A. F. et al. Influenza Forecasting with Google Flu Trends. PLoS One, 8(2), 2013. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3572967/

DUHIGG, Charles. O Poder do Hábito - Por que fazemos o que fazemos na vida e nos Negócios. Objetiva, 2012.

DUHIGG, Charles. Mais rápido e melhor: os segredos da produtividade na vida e nos negócios. Objetiva, 2016.

DUNBAR, Robin I. M. Neocortex size as a constraint on group size in primates. Journal of Human Evolution, v.20, n.6, 1992, p.469-493.

DUNBAR, Robin I. M. The social brain hypothesis. Evolutionary Anthropology, Issues, News, and Reviews, v.6, n.5, 1998, p.178-90.

ECKER, Ullrich; LEWANDOWSKY, Stephan; SWIRE, Briony; CHANG, Darren. Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction. Psychonomic Bulletin & Review, v.18, n.3, Fevereiro de 2011, p.570-8.

EKMAN, Paul; ROSENBERG, Erika L. (ed.) What the Face reveals - basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). New York: Oxford University Press Inc., 1997.

ELLENBERG, Jordan. O poder do pensamento matemático: a ciência de como não estar errado. Zahar, 2015.

ELLWARDT, Lea. Gossip in Organizations: A Social Network Study. Ridderprint, 2011.

EPPLER, Martin J.; MENGIS, Jeanne. The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. The Information Society, v.20, n.5, 2004, p.325-344.

FORMANSKI, José Gilberto; FORMANSKI, Filipi Naspolini; RODRIGUEZ y RODRIGUEZ, Martius Vicente. A contribuição da análise de redes sociais na identificação dos conhecimentos críticos em uma organização: um estudo de caso. Anais do KM Brasil 2012. São Paulo: SBGC, agosto 2012.

FOSKET, D. J. Theory of clumps. In: SPARCK-JONES, Karen; WILLET, Peter (eds). Readings in Information Retrieval. San Francisco: Morgan Kaufmann, 1997.

FOSTER, Eric K. Research on Gossip: Taxonomy, Methods, and Future Directions. Review of General Psychology, v.8, n.2, 2004, p.78–99.

FRIEDEN, Jeffry A. Capitalismo Global – história econômica e política do século XX. Zahar, 2008.

GENG, Liqiang; HAMILTON, Howard J. Interestingness Measures for Data Mining: A Survey. ACM Computing Surveys, v.38, n.3, 2006.

GHANI, Rayid; SIMMONS, Hillery. Predicting the End-Price of Online Auctions. International Workshop on Data Mining and Adaptive Modelling Methods for Economics and Management held in conjunction with the 15th European Conference on Machine Learning (ECML/PKDDD), Pisa, Itália, 2004.

GIGERENZER, Gerd; GAISSMAIER, Wolfgang. Heuristic Decision Making. Annual Review of Psychology, v.62, 2011, p.451–482.

GLADWELL, Malcolm. Blink - a decisão num piscar de olhos. Rocco, 2005.

GLADWELL, Malcolm. Outliers - the story of success. Back Bay Books, 2011.

GLADWELL, Malcolm. O ponto da virada - como pequenas coisas podem fazer uma grande diferença (original: the tipping point). Rio de Janeiro: Sextante, 2013.

GLEICK, James. Caos - a criação de uma nova ciência. Rio de Janeiro: Campus, 1989.

GODIN, Seth. Unleashing the Ideavirus. Do You Zoom Inc., 2000.

GOLDENBERG, Jacob; LIBAI, Barak; MULLER, Eitan. Talk of the network: a complex systems look at the underlying process of word-of-mouth. Marketing Letters, v.12, n.3, 2001, p.211-223.

GOLEMAN, Daniel. Foco: a atenção e seu papel fundamental para o sucesso. Objetiva, 2013.

GORR, Wilpen L. et al. Forecasting Crime. 1999.

GRANOVETTER, Mark S. The strength of weak ties. American Journal of Sociology, v.78, n.6, Maio de 1973, p.1360-1380.

GUNTHER, Max. O Fator Sorte. Rio de Janeiro: Best Business, 2013 (original: The luck factor, 1977).

HARARI, Yuval Noah. 21 lições para o século 21. Companhia das Letras, 21018.

HESS, Charlotte; OSTROM, Elinor. Introduction: An Overview of the Knowledge Commons. Em: Understanding Knowledge as a Commons - From Theory to Practice. Editado por Charlotte Hess e Elinor Ostrom. The MIT Press, 2007, p.3-26.

HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin. The Fourth Paradigm: data-intensive scientific discovery. Redmond: Microsoft Research, 2009.

HILL, R.A.; DUNBAR, Robin I. M. Social network size in humans. Human Nature, v.14, n.1, 2003, p.53-72.

HOFSTEDE, Geert H. Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations. SAGE, 2001.

HOFSTEDE, Geert H. Dimensionalizing Cultures: The Hofstede Model in Context. Online Readings in Psychology and Culture, v.2, n.1, 2011.

HUFF, Darrell. Como Mentir Com Estatística. Intrínseca, 2016.

IPE, Minu. Knowledge Sharing in Organizations: A Conceptual Framework. Human Resource Development Review, v.2, n.4, Dezembro de 2003, p.337-359.

ISMAIL, Salim; MALONE, Michael S.; GEEST, Yuri Van. Organizações Exponenciais. HSM Editora, 2015.

JOHNSON, Steven. De onde vêm as boas ideias. Zahar, 2011.

JOHNSON, Steven. Anatomy of an idea. 2011.
http://www.stevenberlinjohnson.com/2011/12/anatomy-of-an-idea.html

JOHNSON, Steven. Tudo que é ruim é bom pra você. Zahar, 2012.

KAHNEMAN, Daniel. Rápido e Devagar - Duas Formas de Pensar. Objetiva, 2012.

KANDEL, Eric R. Em busca da memória: o nascimento de uma nova ciência da mente. Companhia das Letras, 2009.

KANG, Sung Choon; KIM, Tai Gyu. Opinion leaders, social capital and innovations in teams. Seoul Journal of Business, v.15, n.2, Dezembro de 2009.

KARR-WISNIWESKI, P.; LU, Y. When more is too much: Operationalizing technology overload and exploring its impact on knowledge worker productivity. Computers in Human Behavior, v.26, 2010, p.1061-1072.

KAYO, Eduardo Kazuo; KIMURA, Herbert; MARTIN, Diógenes Manoel Leiva; NAKAMURA, Wilson Toshiro. Ativos Intangíveis, Ciclo de Vida e Criação de Valor. Revista de Administração Contemporânea, v. 10, n. 3, Jul./Set. 2006, p.73-90.

KELLY, Kevin. New Rules for the New Economy – 10 radical strategies for a connected world. New York: Penguin Group, 1998.

KNIGHT, Chris; POWER, Camilla; WATSS, Ian. The Human Symbolic Revolution: A Darwinian Account. Cambridge Archaeological Journal v.5, n.1, 1995, p.75-114.

KELLY, Kevin. Bootstrapping Complexity. Amazon (e-book Kindle), 2011.

KOESTLER, Arthur. The Act of Creation - a study of the conscious and unconscious processes in humor, scientific discovery and art. New York: Arkana (The Penguin Group), 1964.

KOLLOCK, Peter; SMITH, Marc. Managing the Virtual Commons: Cooperation and Conflict in Computer Communities. Em: Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives, edited by Susan Herring. Amsterdam: John Benjamins. 1996. p. 109-128.

KOLLOCK, Peter. The Economies of Online Cooperation: Gifts and Public Goods in Cyberspace. Em: Marc Smith and Peter Kollock (editores). Communities in Cyberspace. London: Routledge, 1999.

KORTH, Henry F.; SILBERSCHATZ, Abraham. Database research faces the information explosion. Communications of the ACM, v.40, n.2, Fevereiro de 1997, p.139-142.

KRING, Ann M.; SLOAN, Denise M. The Facial Expression Coding System (FACES): development, validation, and utility. Psychological Assessment, v.19, n.2, Junho de 2007, p.210-24.

KUHN, Thomas S. A Estrutura das Revoluções Científicas. 10.ed. São Paulo: Perspectiva, 2011 (original: 1962).

LENAT, Douglas B. The nature of Heuristics. Artificial Intelligence, v.19, n.2, Outubro de 1982, p.189-249.

LEVITIN, Daniel J. A mente organizada: como pensar com clareza na era da sobrecarga de Informação. Rio de Janeiro: Objetiva, 2015.

LEWANDOWSKY, Stephan; ECKER Ullrich K. H.; SEIFERT, Colleen M.; SCHWARZ, Norbert; COOK, John. Misinformation and Its Correction: Continued Influence and Successful Debiasing. Psychological Science in the Public Interest, v.13, n.3, 2012, p.106-131.

LEWIS, Michael. Moneyball: The Art of Winning an Unfair Game. W. W. Norton & Company, 2004.

LOH, Stanley. BI na era do big data para cientistas de dados - indo além de cubos e dashboards na busca pelos porquês, explicações e padrões. Porto Alegre, 2014.

LOH, Stanley. Memória organizacional: como coletar, armazenar e recuperar conhecimentos explícitos e tácitos. Porto Alegre, 2018.

LORENZI, Fabiana, LOH, S. Best Practices for Social CRM In: Organizations and Social Networking: Utilizing Social Media to Engage Consumers. Hershey : IGI Global, 2013, p. 273-284.

LORENZI, Fabiana; SILVEIRA, Sidnei Renato. Desenvolvimento de sistemas de informação inteligentes. Porto Alegre: Ed. UniRitter, 2011.

LOSEE, John. A Historical Introduction to the Philosophy of Science. 4a.ed. New York: Oxford University Press, 2001. (original 1972)

MARDER, Ben; JOINSON, Adam; SHNAKAR, Avi. More Facebook friends means more stress, says report. The University of Edinburgh, News Release, n.26, novembro de 2012.

MARTINS Filho, Plínio. Direitos autorais na Internet. Ciência da Informação, v.27, n.2, 1998, p.183-188.

McLUHAN, Marshall. The Gutenberg Galaxy: the making of typographic man. Toronto: University of Toronto Press, 1962.

McLUHAN, Marshall. Os Meios de Comunicação como Extensões do Homem. 18.ed. Cultrix, 2014. (original de 1969)

MALTZ, Michael D.; KLOSAK-MULLANY, Jacqueline. Visualizing Lives: New Pathways for Analyzing Life Course Trajectories. Journal of Quantitative Criminology, v.16, n.2, June 2000, p.255-281.

MARDER, Ben; JOINSON, Adam; SHNAKAR, Avi. More Facebook friends means more stress, says report. The University of Edinburgh, News Release, n.26, novembro de 2012.

MAYBURY, Mark T. Intelligent multimedia information retrieval. Menlo Park: AAAI Press, 1997.

McLUHAN, Marshall. The Gutenberg Galaxy: the making of typographic man. Toronto: University of Toronto Press, 1962.

McLUHAN, Marshall. Os Meios de Comunicação como Extensões do Homem. 18.ed. Cultrix, 2014. (original de 1969)

McPHERSON, Miller; SMITH-LOVIN, Lynn; COOK, James M. Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology, v.27, 2001, p.415-444.

McNEELY, Ian F. A reinvenção do conhecimento: de Alexandria à Internet. Record, 2013.

MILGRAM, Stanley. The Small World Problem. Psychology Today, v.2, 1967, p.60–67.

MISHNE, Gilad. Predicting movie sales from blogger sentiment. In AAAI Spring Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW) 2006.

MORAES, Maurício. Big Brother Obama. Revista Info, Editora Abril, n.324, dezembro de 2012.

MLODINOW, Leonard. O andar do bêbado: como o acaso determina nossas vidas. Zahar, 2009.

MLODINOW, Leonard. Subliminar - Como o inconsciente influencia nossas vidas. Zahar, 2013.

MORAES, Maurício. Big Brother Obama. Revista Info, Editora Abril, n.324, dezembro de 2012.

MORIN, Edgar. Os Sete Saberes Necessários à Educação do Futuro. 2.ed. São Paulo: Cortez; Brasília: UNESCO, 2000.

MORTON, Tom. The Descent of Man, Episode 3: Is Anyone in There? Broadcast on The Science Show on ABC Radio National.

NIRENBERG, Jesse S. A psicologia da comunicação. IBRASA, 1981.

NOWAK, Martin A. Evolutionary dynamics - exploring the equations of life. Harvard University Press, 2006.

OSTROM, Elinor. Governing the Commons - The Evolution of Institutions for Collective Action. Cambridge University Press, 1990.

PAGEL, Mark; MACE, Ruth. The cultural wealth of nations. Nature, v.428, n.18, Março de 2004.

PENTEADO, José R. W. A técnica da comunicação humana. 5.ed. Biblioteca Pioneira de Administração e Negócios, 1976.

PENTLAND, Alex. Honest Signals: How They Shape Our World. MIT, 2008.

PENTLAND, Alex. To Signal Is Human. American Scientist, v.98, n.3, Maio-Junho de 2010, p. 203-211.

PENTLAND, Alex. Social Physics: how good ideas spread - the lessons from a new Science. Penguin, 2014.

PIETES, Rik; WEDEL, Michel. Attention capture and transfer in advertising: brand, pictorial, and text-size effects. Journal of Marketing, v.68, Abril de 2004, p.36-50.

POPPER, Karl. The logic of scientific discovery. Londres: Hutchinson & Co., 1959.

POPPER, Karl. Conjecturas e Refutações. Brasília: Editora da UnB. 1980.

RADINSKY, Kira; HORVITZ, Eric. Mining the web to predict future events. Proceedings WSDM '13 Proceedings of the sixth ACM international conference on Web search and data mining, 2013, p. 255-264.

RECUERO, Raquel. Redes sociais na Internet. Porto Alegre: Sulina, 2009 (Coleção Cibercultura)

REED, Jeff. Data Analytics: Applicable Data Analysis to Advance Any Business Using the Power of Data Driven Analytics. Amazon, 2018.

RIDLEY, Matt. O otimista racional – por que o mundo melhora. Record, 2014.

RIFKIN, Jeremy. The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism. Palgrave MacMillan, 2015.

ROGERS, Everett M. Diffusion of innovations. 5.ed. New York: Free Press, 2003.

ROSEN, Christine. The Myth of Multitasking. The New Atlantis, n.20, Spring 2008, p. 105-110.

SAGGION, Horacio; CARVALHO, Ariadne. Análise textual visando a tradução automática. In: CASTILHO, José M. V. (ed). XXI Conferência Latino-Americana de Informática - PANEL'95. Anais... Canela, UFRGS, 1995.

SARGUT, Gökçe; McGRATH, Rita Gunther. Learning to Live with Complexity. Harvard Business Review, special issue on Complexity, September 2011.

SAWYER, Robert Keith. Explaining creativity: the science of human innovation. Oxford University Press, 2006.

SCARINCI, Rui G. **SES - Sistema de Extração Semântica de Informações**. Porto Alegre: CPGCC - Universidade Federal do Rio Grande do Sul. Março de 1997. (dissertação de mestrado)

SCARUFFI, Piero. A History of Silicon Valley - Almost a 3rd Edition (ebook).

SHERMER, Michael. Por que as pessoas acreditam em coisas estranhas: pseudociência, superstição e outras confusões dos nossos tempos. São Paulo: JSN Editora, 2011.

SHU, Kai; LIU, Huan. A Quick Guide to Fake News Detection on Social Media. Outubro de 2017. https://www.kdnuggets.com/2017/10/guide-fake-news-detection-social-media.html

SHU, K.; SILVA, A.; WANG, S.; TANG, J.; LIU, H. Fake News Detection on Social Media: A Data Mining Perspective. ACM SIGKDD Explorations Newsletter, v.19, n.1, 2017, p.22-36.

SILVER, Nate. O sinal e o ruído: por que tantas previsões falham e outras não. Rio de Janeiro: Intrínseca, 2013.

SIMON, Herbert A. Theories of Bounded Rationality. In C. B. McGuire & R. Radner (Ed.), Decision and Organization. Amsterdam: North-Holland Publishing Company. 1972.

SIMONTON, Dean Keith. A origem do gênio: perspectivas darwinianas sobre a criatividade. Record, 2002.

SINCLAIR, Marta; ASHKANASY, Neal M. Intuition: Myth or a Decision-making Tool? Management Learning, v.36, n.3, Setembro de 2005, p.353-370.

SOARES, Luiz F. G. et alli. Fundamentos de sistemas multimídia. VIII Escola de Computação, Gramado: II/UFRGS, 1992.

SPARCK-JONES, Karen; WILLET, Peter (eds). **Readings in Information Retrieval**. San Francisco: Morgan Kaufmann, 1997.

SPEER, Nicole K.; REYNOLDS, Jeremy R.; SWALLOW, Khena M.; ZACKS, Jeffrey M. Reading Stories Activates Neural Representations of Visual and Motor Experiences. Psychological Science, v.20, n.8, 2009, p.989–999.

SPINK, Amanda; WOLFRAM, Dietmar; JANSEN, Major B. J.; SARACEVIC, Tefko. Searching the web: The public and their queries. Journal of the American Society for Information Science and Technology, v. 52, n.3, 2001, p. 226–234.

STEWART, Ian. Será que Deus joga dados? A nova matemática do caos. Zahar, 1991.

STEWART, Thomas A. Capital Intelectual - a nova vantagem competitiva das empresas. Campus, 1998.

STEWART, Thomas R. Uncertainty, judgment and error in prediction. In: SAREWITZ, D.; PIELKE, R. A.; BYERLEY, R. Prediction: Science, Decision Making and the Future of Nature. Washington: Island Press, 2000, p. 41-57.

STIGLITZ, Joseph E. A globalização e seus malefícios - a promessa não cumprida de benefícios globais. Futura, 2002.

SUROWIECKI, James. The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations. Little, Brown. 2004.

SUTHERLAND, Jeff. Scrum: a arte de fazer o dobro do trabalho na metade do tempo. 2.ed. Leya, 2016.

SWAP, Walter; LEONARD, Dorothy; SHIELDS, Mimi; ABRAMS, Lisa. Using mentoring and storytelling to transfer knowledege in the workplace. Journal of Management Information Systems, v.18, n.1, Summer 2001, p.95-114.

TALEB, Nassim Nicholas. A Lógica do Cisne Negro. Best Seller, 2008.

TALEB, Nassim Nicholas. Antifrágil - Coisas que se Beneficiam com o Caos. Best Business, 2014.

TETLOCK, Philip; GARDNER, Dan. Superprevisões: a arte e a ciência de antecipar o futuro. Objetiva, 2016.

TOFFLER, Alvin. A Terceira Onda. 30.ed. Record, 2007.

TOLE, A. A. Big Data Challenges. Database Systems Journal, v. IV, n. 3, 2013, p.31-40.

THOMPSON, Derek. Hit Makers: como nascem as tendências. HarperCollins, 2018.

TOOBY, John; COSMIDES, Leda. The psychological foundations of culture. p.19-136. In: BARKOW, Jerome H.; COSMIDES, Leda; TOOBY, John (eds.). The Adapted Mind: Evolutionary Psychology and the Generation of Culture. New York: Oxford University Press, 1992.

TVERSKY, Amos; KAHNEMAN, Daniel. Judgment under uncertainty: heuristics and biases. Science, n.185, 1974, p.1124-1131.

TVERSKY, Amos; KAHNEMAN, Daniel. Extensional versus intuitive reasoning: the conjunction fallacy in probabilistic reasoning. Psychological Review, n.90, 1983, p.293-315.

UCHIDA, Naoshige; KEPECS, Adam; MAINEN, Zachary F. Seeing at a glance, smelling in a whiff: rapid forms of perceptual decision making. Neuroscience, v.7, Junho de 2006, p.485-491.

UNDERHILL, Paco. Why we buy: the science of shopping. Simon & Schuster, 1999.

VALENTIM, Amanda Freitas; CÔRTES, Marcela Guimarães; GAMA, Ana Cristina Côrtes. Análise espectrográfica da voz: efeito do treinamento visual na confiabilidade da avaliação. Revista da Sociedade Brasileira de Fonoaudiologia, v.15, n.3, 2010, p.335-42.

WALDROP, M. Mitchell. Complexity: The Emerging Science at the Edge of Order and Chaos. Simon and Schuster, 1993.

WALLIN, Claudia. Um país sem excelências e mordomias. Geração, 2014.

WATTS, Duncan J. Tudo é óbvio: quando você sabe a resposta. Rio de Janeiro: Paz e Terra, 2015.

WEINER, Eric. The Geography of Genius: A Search for the World's Most Creative Places from Ancient Athens to Silicon Valley. Simon & Schuster, 2016.

WILSON, Timothy D. Strangers to Ourselves: Discovering the Adaptive Unconscious. Belknap Press of Harvard University Press, Maio de 2004.

WILSON, Edward O. A conquista social da terra. Companhia das Letras, 2013.

WINSTON, Robert. Instinto humano. São Paulo: Globo, 2006.

WOLF, Gary. The Data-Driven Life - What happens when technology can analyze every quotidian thing that happened to you today ? The New York Times Magazine Maio de 2010.

YATES, Ricardo Baeza. An extended model for full text databases. Journal of the Brazilian Computer Society, v.2, n.3, Abr 1996.

ZACKS, Jeffrey M. How We Organize Our Experience into Events. Psychological Science, April 2010.

ZACKS, Jeffrey M.; SWALLOW, Khena M. Event Segmentation. Current Directions in Psychological Science, v.16, n.2, Abril de 2007, p.80–84.