UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ADMINISTRAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO

METODOLOGIA DE SEGMENTAÇÃO DE MÍDIA SOCIAL

Luiz Wanderley Tavares

Orientadora: Prof^a. Dr^a. Maria Aparecida Gouvêa.

SÃO PAULO

2017

Prof. Dr. Marco Antonio Zago
Reitor da Universidade de São Paulo

Prof. Dr. Adalberto Américo Fischmann
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. Roberto Sbragia
Chefe do Departamento de Administração

Prof. Dr. Moacir Miranda de Oliveira Junior
Coordenador do Programa de Pós-Graduação em Administração

LUIZ WANDERLEY TAVARES

METODOLOGIA DE SEGMENTAÇÃO DE MÍDIA SOCIAL

Dissertação apresentada à Faculdade de Economia, Administração e Contabilidade da USP, como parte das exigências para a obtenção do título de Mestre em Ciências.

Área de Concentração: Métodos Quantitativos e Informática.

Orientadora: Prof^a. Dr^a. Maria Aparecida Gouvêa.

Versão Original

SÃO PAULO

2017

Tavares, Luiz Wanderley
Metodologia de segmentação de mídia social / Luiz Wanderley Tavares. -- São Paulo, 2017.
84 p.

Dissertação (Mestrado) – Universidade de São Paulo, 2017.
Orientador: Maria Aparecida Gouvêa.

1. Mídias sociais 2. Netnografia 3. Twitter 4. Tribos 5. Filtragem colaborativa I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

CDD – 303.4833

“I believe in intuition and inspiration.... At times, I feel certain I am right while not knowing the reason. When the eclipse of 1919 confirmed my intuition, was not in the least surprised. In fact, I would have been astonished had it turned out otherwise. **Imagination is more important than knowledge.** For knowledge is limited, whereas imagination embraces the entire world, stimulating progress, giving birth to evolution. It is, strictly speaking, a real factor in scientific research” (Einstein, 1931, p. 96-97).

AGRADECIMENTOS

Em 54 A.C., Cícero (1951) realizou a defesa de seu amigo Plancius, onde citou a gratidão não apenas sendo a maior de todas as virtudes; mas, como a mãe de todas elas.

A gratidão pode ser descrita como o reconhecimento de uma ação ou benefício recebido por alguém; porém, agradecer não é uma atividade simples e é fácil pecar com o esquecimento.

Não se pode esquecer as pessoas responsáveis em fazer você chegar neste momento da vida, ajudando na criação e educação básica. Nem se pode falhar ao não citar amigos e professores forjadores da sua capacidade de pensar e agir ainda antes de iniciar a vida adulta.

O fim do primeiro grande ciclo de formação inicia com novas interações e responsabilidades e a vida ganha complexidade, novas pessoas participam de sua existência e a necessidade de adaptação é constante, se faz necessário agradecer a todos os provocadores de mudanças, sem eles não existiria evolução.

A sequência da vida parece mostrar uma rotina previsível e monótona; porém, novos ciclos quebram a rotina quando a paternidade lhe é oferecida por uma luz em forma de gente. Como não ter gratidão por uma pessoa capaz de gerar vida, ainda mais quando esta nova vida é uma continuação da sua?

As novas vidas criam o início de novos ciclos, onde os verbos criar e educar ganham novos significados. Sim, sempre se deve agradecer aos filhos, sem eles a rotina previsível e monótona teria destruído a imaginação. São eles, nascidos em uma nova realidade, capazes de questionar o passado de seus ancestrais e provocar mudanças no futuro ao seu redor.

O futuro não existe e pode ser moldado a todo momento, sendo apenas necessário ter coragem para enfrentar novas aventuras. Em todos os momentos existem pessoas

apresentando novos rumos; aceitos ou não. É obrigatório agradecer os responsáveis em fazer você embarcar nos maiores desafios, neste momento são considerados amigos, durante são “traidores” e depois os criadores de uma nova perspectiva de vida.

Para terminar, durante todos os processos de transformação existem novos orientadores cruzando o seu caminho, pessoas dispostas em doar atenção e conhecimento para o seu crescimento pessoal. Como esquecer de agradecer pessoas capazes de semear e cultivar conhecimento na imaginação?

Todos são importantes e estão lembrados.

Obrigado.

RESUMO

As primeiras mídias sociais da internet surgiram há pouco mais de duas décadas, segunda metade dos anos 90. Em comparação com a evolução humana, isso seria algo como um milésimo de segundo de sua existência. Neste período, vários estudos procuram entender o comportamento e o agrupamento dos seres humanos nesta nova forma de comunicação. Teorias sobre formas de analisar as pessoas neste meio e como elas se agrupam e criam novos modos de comunicação e propagação de suas ideias florescem e iluminam este desconhecido caminho a ser criado e percorrido. Os métodos de identificação do comportamento humano criados antes das mídias sociais ganham uma nova forma de serem utilizados. Estudos sobre o “eu” (Belk, 1988), tribalismo (Cova, B., 1997), etnografia (Danzig, 1985), netnografia (Kozinets, 1998) e filtragem colaborativa (Golberg, Nichols, Oki e Terry, 1992) entram em cena para colocar uma luz no estudo das relações humanas no mundo digital. A internet revolucionou o modo de as pessoas interagirem e a evolução constante da tecnologia vem incessantemente gerando profundas implicações para o marketing. A rede mundial passou a ser um canal global pelo qual as empresas podem divulgar e vender seus produtos. No entanto, mesmo oferecendo um enorme potencial para as empresas, a internet aumentou a complexidade de identificar os clientes. Os usuários presentes nas mídias sociais estão menos interessados nos produtos e valorizam mais as identidades e os laços sociais gerados em torno de seus assuntos de interesse. Estas tribos eletrônicas ultrapassam as fronteiras geográficas e independem de raça, sexo e aspectos culturais de seus integrantes. Este trabalho apresenta um método para identificar tribos nas mídias sociais. O método foi aplicado na identificação da tribo de *MMA* (*Mixed Martial Arts*, em tradução livre, Artes Marciais Mistas) no *Twitter*. A validação foi realizada usando a plataforma de anúncios do *Twitter*, enviando durante 72 horas uma publicidade para mais de 600 mil usuários, divididos em grupo de controle e segmentações do *Twitter* e do método proposto *DNA*. O estudo comparou os resultados obtidos pelo método proposto *DNA* com os resultados do grupo de controle e da segmentação realizada pelo *Twitter*. Os resultados obtidos apontaram o aumento de interações dos usuários identificados como pertencentes a tribo de *MMA*, validando o método.

Palavras-chave: Mídia Social. Netnografia. *Twitter*. Tribos. Filtragem Colaborativa.

ABSTRACT

The first Internet social media emerged just over two decades ago, at the second half of 90's. Compared to human evolution, this would be something like a millisecond of its existence. In this period, several studies try to understand the behavior and grouping of human beings in this new form of communication. Theories about ways of analyzing people in this environment and how they group themselves and create new ways of communication and propagation their ideas flourish and illuminate this unknown pathway to be created and traveled. Methods of identifying human behavior created before social media receive a new way of being used. Studies on the "self" (Belk, 1988), tribalism (Cova, B., 1997), ethnography (Danzig, 1985), netnography (Kozinets, 1998) and collaborative filtering (Golberg, Nichols, Oki and Terry, 1992) come on the scene to shed light on the study of human relations in the digital world. The Internet has revolutionized people's way of interacting and the constant evolution of technology generates profound implications for the marketing. The worldwide network has become a global channel through which companies can disclose and sell their products. However, while offering tremendous potential to businesses, the Internet has increased the complexity of identifying customers. Users present in social media are less interested in products and value more the identities and social ties generated around their subjects of interest. These electronic tribes transcend the geographical borders and are independent of race, sex and cultural aspects of its members. This paper presents a method to identify tribes in social media. The method was applied in the identification of the MMA (Mixed Martial Arts) tribe on Twitter. The validation was done using the Twitter ads platform, sending 72 hours of advertisement for more than 600 thousand users, divided in control group and segmentations of Twitter and the proposed method. The study compared the results obtained by the proposed method with that of the control group and the segmentation created by Twitter. The obtained results pointed out the increase of interactions of the users identified as belonging to the MMA tribe validating the method.

Keywords: Social Media. Netnography. *Twitter*. Tribes. Collaborative Filtering.

SUMÁRIO

1.	INTRODUÇÃO	13
1.1.	Problema da pesquisa	18
1.2.	Objetivo da pesquisa	20
2.	FUNDAMENTAÇÃO TEÓRICA	23
2.1.	A Internet e as mídias sociais	23
2.2.	Tribalismo	29
2.3.	Netnografia	32
2.4.	Filtragem Colaborativa	34
3.	ASPECTOS METODOLÓGICOS	39
3.1.	Tipologia da Pesquisa	39
3.2.	Procedimento de segmentação	40
3.2.1.	Definição das sementes	43
3.2.2.	Definição dos geradores	49
3.2.3.	Definição do <i>DNA</i> da tribo	50
3.3.	Amostragem	53
3.4.	Experimento	54
3.4.1.	Primeira campanha publicitária: Controle	55
3.4.2.	Segunda campanha publicitária: <i>Twitter</i> segmentando o público de <i>MMA</i> .	56
3.4.3.	Terceira campanha publicitária: 40.000 usuários selecionados pelo método proposto <i>DNA</i>	57
4.	ANÁLISE DOS RESULTADOS	59
4.1.	Comparação entre horários com e sem evento ao vivo	60
4.1.1.	Resultado utilizando o método <i>DNA</i> com e sem evento ao vivo	61
4.1.2.	Resultado utilizando o método <i>Twitter</i> com e sem evento ao vivo	62

4.2. Resultado comparativo do período total pesquisado	64
4.2.1. <i>Twitter</i> x Grupo de Controle	65
4.2.2. Método <i>DNA</i> x Grupo de Controle	65
4.2.3. Método <i>DNA</i> x <i>Twitter</i>	66
4.3. Interações a cada 6 horas	67
4.3.1. <i>Twitter</i> x Grupo de Controle	68
4.3.2. Método <i>DNA</i> x Grupo de Controle	69
4.3.3. Método <i>DNA</i> x <i>Twitter</i>	70
CONCLUSÕES	73
REFERÊNCIAS BIBLIOGRÁFICAS	75
Apêndice I – Dados coletados	81
Apêndice II – Glossário	82
Apêndice III – Exemplo de memória de cálculo	84
FIGURAS, QUADROS E TABELAS	
Figura 1: Matriz de avaliações de itens por usuários.	35
Figura 2: Esquema do método	41
Figura 3: Matriz colaborativa do método proposto	42
Figura 4: Conjuntos de definição do peso $w_{t,i}$	42
Figura 5: Campanha publicitária	54
Figura 6: Percentual de interações a cada seis horas	67
Quadro 1: Consequências da Internet (Katz e Rice, 2002)	24
Quadro 2: Classificação das mídias sociais por presença social / riqueza da mídia e auto apresentação / auto divulgação (Kaplan e Haenlein, 2010)	28
Quadro 3: Programação do canal de TV nos períodos estudados	61

Tabela 1: Sementes selecionadas pela busca na mídia social	44
Tabela 2: Usuários mais comentados pelas primeiras 10 sementes selecionadas	45
Tabela 3: Lista final das sementes selecionadas	46
Tabela 4: Sementes com as informações normalizadas	48
Tabela 5: Sementes classificadas em ordem de relevância	49
Tabela 6: Quantidade de geradores por quantidade de sementes em conjunto	49
Tabela 7: Matriz de correlação dos dados do <i>DNA</i>	52
Tabela 8: Tabela com os dados da 1ª campanha contabilizados em períodos de 6 horas	56
Tabela 9: Tabela com os dados da 2ª campanha contabilizados em períodos de 6 horas	57
Tabela 10: Tabela com os dados da 3ª campanha contabilizados em períodos de 6 horas	58
Tabela 11: Teste de proporção do método <i>DNA</i> entre a faixa do evento ao vivo e as outras faixas	61
Tabela 12: Teste de proporção do método <i>Twitter</i> entre a faixa do evento ao vivo e as outras faixas	63
Tabela 13: Teste de proporção do grupo de Controle entre a faixa ao vivo e as outras faixas	64
Tabela 14: Teste de proporção do período total pesquisado	64
Tabela 15: Teste de proporção entre a segmentação do <i>Twitter</i> e o grupo de Controle	68
Tabela 16: Teste de proporção entre o método <i>DNA</i> e o grupo de Controle	69
Tabela 17: Teste de proporção entre o método <i>DNA</i> e a segmentação do <i>Twitter</i> ..	70

1. INTRODUÇÃO

A necessidade de pertencer ou a necessidade de formar e manter uma quantidade mínima de relacionamentos interpessoais é inata nos seres humanos. A formação dos laços sociais é um dos tópicos discutidos no estudo sobre a “necessidade de pertencer” de Baumeister e Leary (1995). O estudo cita a facilidade de como os laços são formados e como as pessoas pertencem a pequenos grupos em sua sociedade.

Os grupos passaram a ter a internet como uma nova forma de interação com os seus assuntos de interesse. Esta nova possibilidade de interação aumentou a quantidade de grupos acessados pelas pessoas. Katz e Rice (2002) apontaram as controvérsias existentes com o efeito da internet no envolvimento social e sua contribuição para a cooperação entre duas partes, o Capital Social. Este não necessariamente positivo, pois os contatos entre as pessoas podem gerar situações negativas. Os autores separaram as duas perspectivas opostas, distópica e utópica, em três importantes casos: acesso, envolvimento cívico e comunitário e interação social e expressão. Maiores detalhes na fundamentação teórica, seção 2.1.

Os efeitos gerados por estes casos advém do surgimento das mídias sociais e a nova forma de os consumidores interagirem com os conteúdos e pessoas de seus interesses. Em virtude da difusão do seu uso, a internet passou a ser uma poderosa ferramenta para as pessoas se unirem e atuarem juntas em tribos sem restrição de tempo e espaço, e este fenômeno traz à tona a necessidade de as empresas desenvolverem uma abordagem social conectada a esta nova ressocialização dos indivíduos (Cova, B. e Cova, V., 2002).

Estas novas uniões passaram a estimular as empresas a procurarem novas formas de identificar os consumidores. Estes relacionamentos nas mídias sociais foram chamados de *tribos eletrônicas* (Adams e Smith, 2008).

Para Wenger (2000, p. 227) a atuação nas tribos pode ter várias formas e vários níveis entre uma interação local e uma participação global, distinguindo três modos de envolvimento: engajamento, imaginação e alinhamento.

- Engajamento: os participantes das tribos fazem coisas juntos, falam e ajudam-se mutuamente.
- Imageminação: eles constroem suas próprias imagens, da tribo e do mundo, para orientar e refletir sobre a situação e explorar as possibilidades.
- Alinhamento: eles têm certeza do suficiente alinhamento dos processos com os locais de atividades para o engajamento ser efetivo.

A visão de Wenger (2000) é similar a de outros autores em estudos mais recentes (Kamel Boulos e Wheeler, 2007 e Pahl-Wostl, 2015).

As tribos estão espalhadas em várias mídias sociais, e muitas foram criadas com um propósito específico, como a mídia social específica para apreciadores de vinho chamada de Vivino. Apesar de ser especializado, o Vivino também publica comentários de seus usuários em outras mídias sociais e é comum estes usuários também trocarem informações sobre vinhos em outras plataformas, como no *Twitter*, Instagram, Facebook, entre outras.

Como os usuários podem estar em mais de uma tribo (Cova, B. e Cova, V., 2002), pode-se pensar também em verificar as referências entre elas, e desta forma melhorar as ofertas de produtos e serviços.

Nas mídias sociais, os usuários fazem conexões com outros usuários, além de realizar comentários, publicar imagens, mostrar interesse em comentários de outros usuários ou mesmo republicá-los. Apesar destas manifestações, pouco se sabe sobre os usuários. Os aplicativos possuem poucas informações, como nome (nem sempre verdadeiro) e outros dados não obrigatórios como sexo, data de nascimento e localização

geográfica; segmentação consiste no procedimento de coleta de informações normalmente utilizadas para classificar os usuários em pequenos grupos ou segmentos de usuários com características iguais. Com este panorama, segmentar os usuários na mídia social passou a ser um desafio para as empresas e seus desenvolvedores. Os métodos e algoritmos utilizados pelos desenvolvedores das mídias sociais para realizar a segmentação não são divulgados.

Uma das maneiras de segmentar os usuários na mídia social é estudando as suas conexões e os assuntos que eles comentam.

Entre as técnicas possíveis para segmentar os usuários, a análise de conglomerados (*cluster analysis*) é utilizada para agrupar indivíduos, pessoas, empresas ou objetos em grupos homogêneos em função da similaridade de suas variáveis (Fávero, Belfiore, Silva e Chan, 2009). No caso das mídias sociais, os indivíduos em um mesmo *cluster* terão conexões em comum e comentarão sobre os mesmos assuntos.

A análise de conglomerados começou a ser amplamente utilizada na segunda metade dos anos 60 e durante os anos 70 e surgiram autores questionando o seu uso como Cormack (1971), Everitt (1979) e Ng, Jordan e Weiss (2001). Cormack (1971) fez uma revisão ampla das técnicas criadas desde suas medidas e índices de similaridade como suas estratégias de classificação, dissertando sobre suas distorções e a possibilidade de as técnicas criarem conglomerados inexistentes. Em 1979, Everitt dissertou sobre os problemas não resolvidos na análise de conglomerados, fazendo um estudo comparativo da análise de conglomerados com as outras técnicas de análises multivariadas para apontar suas diferenças, fraquezas e virtudes. Entre os problemas apontados por Everitt (1997) está a definição do número de conglomerados. Ng et al. (2001) falam sobre a dificuldade de encontrar bons conglomerados e apontam a necessidade de fazer suposições simplificadoras e de ter múltiplos reinícios por causa da possibilidade de

existirem muitos mínimos locais na probabilidade de logs, e citam que algoritmos como *k-means* têm problemas semelhantes.

No caso das mídias sociais, elas não possuem dados pessoais e sociais suficientes para separar os usuários. As conexões e interações feitas nos aplicativos são as informações disponíveis para aplicar as regras de *cluster*, surgindo o problema da capacidade analítica, pois analisar mais de 1 bilhão de usuários, com mais de 600 bilhões de conexões, comentando os mais variados assuntos é algo realmente complexo e desafiador.

O *Twitter* é um serviço por meio do qual amigos, familiares e colegas de trabalho podem se comunicar e se manter conectados, trocando mensagens rápidas e frequentes. As pessoas publicam *Tweets*, que podem conter fotos, vídeos, links e até 140 caracteres de texto. Essas mensagens são publicadas em seu perfil e enviadas aos seus seguidores. Os usuários utilizam com frequência as seguintes áreas: notícias, política, esportes, cultura pop, influenciadores e serviços públicos (página da empresa na Internet, 2017).

Estudo realizado em 2012 pela Universidade de Massachusetts Dartmouth apontou o uso do *Twitter* por 73% das empresas da *Fortune 500* e as 10 maiores empresas consistentemente publicavam em suas contas no *Twitter*. Além de publicar informações sobre seus produtos e serviços, as empresas também começaram a publicar sobre os planos de carreiras, ofertas de empregos e ações de responsabilidades sociais e comunitárias (Barnes, Lescault e Andonian, 2012).

O ecossistema da mídia social centraliza nas experiências dos usuários e as empresas passam a oferecer conteúdos, produtos e serviços voltados a aumentar as relações de afinidade entre eles (Hanna, Rohm e Crittenden, 2011). O conceito de um ecossistema de mídia social permite aos profissionais de marketing pensarem primeiro em termos de estratégia geral, e não sobre as táticas. Trabalhar dentro do ecossistema permite aos gerentes de marketing fazerem perguntas críticas como:

- Quem é / são o (s) alvo (s)?
- Em quais plataformas de mídia tradicional e social os alvos estão?
- Que conteúdo de marketing (estória) a empresa quer contar?
- Como os profissionais de marketing podem propagar ou alimentar esse conteúdo em todo o ecossistema?

Somente após 2008 começou a existir uma substancial quantidade de estudos indicando as mídias sociais como veículos de comunicação. Estes estudos são restritos sempre a um único país. O ritmo lento e gradual nos gastos com publicidade nesses meios e a falta de atenção das pesquisas não surpreendem, devido às dificuldades de fazer estudos internacionais onde a complexidade da coleta de dados e a quantidade de informações acabam sobrecarregando a capacidade analítica (Okazaki e Taylor, 2013).

Esta dissertação foi realizada com a análise de pouco mais de 2,5 milhões de usuários do *Twitter* (quase 8,0% do total de usuários ativos conforme a empresa). Cada um destes usuários possui em média pouco mais de 630 conexões, indicando a existência de mais de 200 bilhões de conexões totais no *Twitter* mostrando as afinidades dos usuários com outros usuários, empresas e serviços. Porém, a despeito da quantidade de conexões dos usuários do *Twitter*, as suas interações seguem um padrão onde existem usuários publicando mensagens constantemente e outros simplesmente observando (Gruzd et al., 2011). A identificação das afinidades dos usuários está relacionada com suas conexões e assuntos de interesse, e estes podem revelar as tribos às quais eles pertencem.

A identificação de uma tribo nas mídias sociais começa com a procura de seus influenciadores (usuários, produtos ou serviços). Vários fatores podem explicar as razões de uma entidade ter a habilidade de influenciar outras pessoas, entre eles, expertise, laços

afetivos, habilidades de liderança, similaridade demográfica, grupo de referência e afinidade (Langner, Hennigs e Wiedmann, 2013).

Conhecer o perfil dos participantes de uma tribo pode ajudar as empresas na exposição de suas marcas e serviços. Estudos empíricos mostram a mídia social influenciando fortemente na imagem da marca (Bruhn, Schoenmueller e Schäfer, 2012), reforçando a necessidade das empresas em buscar melhores formas de identificar seu público-alvo nas mídias sociais.

1.1. Problema da pesquisa

Uma forma de verificar se o procedimento adotado para segmentar um público alvo atinge o objetivo é medindo o retorno das interações dos usuários com as mensagens realizadas nas mídias sociais. Os anúncios apresentados nas mídias sociais são frequentemente ignorados pelos usuários. Hadija, Barnes e Hair (2012) realizaram uma pesquisa com universitários e obtiveram um resultado expressivo sobre o assunto: somente 20% dos respondentes notam os anúncios nas mídias sociais, enquanto 40% comentam não notar ou notar ocasionalmente, 60% dos respondentes nunca visitaram o *web site* do produto após serem expostos à propaganda na mídia social, enquanto 70% deles nunca interagiu em uma propaganda na mídia social e somente 10% dos respondentes acreditam nas redes sociais como a melhor forma em promover as marcas. Considera-se uma interação quando o usuário acessa um anúncio apresentado no aplicativo da mídia social. De acordo com o site Adweek (<http://www.adweek.com/>), as interações dos usuários com as publicidades realizadas no *Twitter* oscilam entre 1% e 3%.

Sendo assim, uma nova forma de segmentação pode identificar melhor o público-alvo, comparando com os métodos utilizados pelos aplicativos de mídias sociais? Poderia

a identificação de uma tribo fornecer uma seleção de usuários com maior afinidade com uma determinada mensagem de produto ou serviço?

Neste contexto, a questão que irá nortear este trabalho pode ser enunciada da seguinte forma:

“Como identificar os participantes de uma tribo nas mídias sociais com o objetivo de aumentar as interações com as mensagens no *Twitter*? “

O aumento de publicidade nas mídias sociais vem crescendo a cada ano. No estudo de Zhang e Mao (2016) sobre a motivação em interagir com uma publicidade na internet cita gastos de US\$ 121 bilhões em publicidade na internet em 2014. Além de pesquisar como as motivações na internet podem fazer os clientes acessarem os anúncios, interagir com o público-alvo correto na internet passou a ser uma das necessidades básicas dos anunciantes e atingir somente o público compatível com os seus interesses reduz custos e tempo das ações publicitárias.

Estas considerações impulsionaram o interesse pelo desenvolvimento deste estudo, abrindo caminho para criar uma alternativa viável e prática na identificação das tribos e seus influenciadores. Um método computacional automático em larga escala pode dar suporte aos vários setores da indústria, mas principalmente para o marketing passar a ter uma metodologia para ser aplicada nas mídias sociais de forma sistêmica, aumentando a precisão do perfil do público-alvo para a divulgação de novos produtos e serviços.

1.2. Objetivo da Pesquisa

Este trabalho tem o objetivo de desenvolver um método para segmentar os usuários da mídia social do *Twitter*, descrevendo os procedimentos de classificação do interesse destes usuários em cada uma das tribos das quais eles participam.

O estudo irá utilizar a filtragem colaborativa como instrumento para identificar o perfil dos usuários de uma determinada tribo. Com base neste perfil da tribo, busca-se identificar os outros usuários na mídia social com maior afinidade com a tribo analisada.

Para verificar se esta forma de segmentação está identificando o público desejado corretamente, será apresentada uma mensagem publicitária aos usuários escolhidos pelo método proposto, pela plataforma de segmentação do próprio *Twitter* e para um grupo de controle na plataforma do próprio *Twitter*.

Os usuários serão convidados a interagir com a mensagem e será comparada a quantidade de pessoas engajadas pelo método proposto *DNA*, pela segmentação realizada pelo *Twitter* e no grupo de controle.

É razoável supor uma maior interação dos usuários com as mensagens, caso o grupo de usuários escolhidos tenha maior afinidade com o assunto de interesse.

A partir da reflexão sobre a comparação das quantidades de interações entre o método proposto e o grupo de controle emerge a hipótese HC:

HC₀ – Os usuários escolhidos pelo método proposto realizam a mesma quantidade de interações em comparação com os usuários do grupo de controle.

HC₁ - Os usuários escolhidos pelo método proposto realizam mais interações em comparação com os usuários do grupo de controle.

Quanto à comparação das quantidades de interações entre o método proposto e a segmentação do *Twitter* emerge a hipótese HT:

HT₀ - Os usuários escolhidos pelo método proposto realizam a mesma quantidade de interações em comparação com os usuários escolhidos pelo *Twitter*.

HT₁ - Os usuários escolhidos pelo método proposto não realizam a mesma quantidade de interações em comparação com os usuários escolhidos pelo *Twitter*.

O desenvolvimento deste estudo permitirá a verificação destas hipóteses e a ampliação da área de conhecimento referente às mídias sociais.

2. FUNDAMENTAÇÃO TEÓRICA

Teorias sobre as novas formas de analisar as pessoas nas mídias sociais e como elas se agrupam e criam novos meios de comunicação e propagação de suas ideias florescem e iluminam um caminho desconhecido a ser percorrido. As atividades sociais, o conceito de posse, a formação de grupos e a forma de compartilhar as memórias vêm recebendo novos contornos no mundo digital completamente diferentes quando comparados com o mundo físico. Os métodos de identificação do comportamento humano criados antes das mídias sociais ganham uma nova forma de serem aplicados. Estas linhas de raciocínio são descritas a seguir e criam a base teórica deste estudo.

2.1. A Internet e as mídias sociais

O uso da internet modificou a forma de os seres humanos se comunicarem e terem acesso às informações. A criação de aplicativos de mídias sociais eletrônicas promoveu uma maior interação entre as pessoas, onde novas formas de interação para troca de informações e experiências foram criadas. As consequências deste novo caminho para gerar relacionamentos vêm sendo estudadas. Katz e Rice (2002) apontam três questões centrais sobre o uso da internet: acesso, envolvimento cívico e comunitário, interação social e expressão.

Acesso é definido pela ação do usuário em conectar na rede de computadores para encontrar material de seu interesse ou comunicar-se com outros usuários. O envolvimento implica a participação de atividades sociais, interagindo com outros usuários, gerando benefícios individuais e coletivos, enquanto a interação social é mais focada no relacionamento entre os indivíduos, onde a expressão refere-se ao material criado pelos usuários ou grupos para refletir seus interesses ou talentos.

Estas três questões são analisadas em duas perspectivas, distópica e utópica, cujas consequências são listadas no Quadro 1.

Quadro 1 – Consequências da Internet (Katz e Rice, 2002).

	Distopia	Utopia
Acesso	<ul style="list-style-type: none"> ▪ Diminui a gama de participantes ▪ Fomenta divisão racial ▪ Limita o discurso ▪ Limita as oportunidades econômicas ▪ Reduz a autoestima ▪ Debilita a voz política ▪ Corrói tradições culturais ▪ Cria balcanização¹ cibernética ▪ Limita benefícios ▪ É inconveniente ▪ Corrói a legitimidade política devido à falta de representatividade ▪ Reduz a privacidade individual e global. 	<ul style="list-style-type: none"> ▪ Supera diferenças tais como: geografia, classe social, raça e etnia, idade, sexo, fuso, cronologia (materiais mais antigos estão disponíveis, e as contribuições são menos efêmeras do que com outros meios de comunicação) e Ideologia. ▪ Oferece mais oportunidades para a participação ▪ Identifica novos talentos, o que enriquece a cultura ▪ Incentiva maior tolerância
Envolvimento cívico e comunitário	<ul style="list-style-type: none"> ▪ Destrói culturas locais e indígenas ▪ Explora pessoas ▪ Reduz a qualidade de vida ▪ Reduz o envolvimento da comunidade ▪ Fragmenta a comunidade, levando à busca de interesses estreitos e balcanização¹ cibernética ▪ Limita as conexões sociais (isolamento e anomia²) ▪ Estimula conexões sociais excessivas (vício) ▪ Leva à perseguição cibernética e até mesmo assassinatos ▪ Provoca conflitos étnicos e raciais ▪ Incita ódio dos outros 	<ul style="list-style-type: none"> ▪ Supera restrições físicas e temporais nos limites da comunidade ▪ Revive as comunidades ▪ Aumenta o voluntariado ▪ Cria informações compartilhadas e opiniões comunitárias ▪ Estimula ativismo além de uma região local ▪ Enriquece a produção cultural ▪ Complementa as relações <i>off-line</i>
Interação social e expressão	<ul style="list-style-type: none"> ▪ Dominado por empresas multinacionais ▪ Incentiva pornografia infantil ▪ Incentiva predação sobre as crianças ▪ Convida às fraudes emocionais ▪ Desenvolve vários “eu” que levam à confusão ▪ Mata a criatividade ▪ Leva à aprendizagem mecânica ▪ Reduz a qualidade dos produtos intelectuais ▪ Permite o plágio ▪ Fraca integridade artística ▪ Propaga vícios (ao sexo, jogos de azar, interação, jogos violentos, os jogos não-violentos, fantasias) 	<ul style="list-style-type: none"> ▪ É social, diversificada e frequente ▪ Complementa e fortalece as interações <i>off-line</i> ▪ Permite formar amizades ▪ Mantém ou restaura as relações com a família e amigos ▪ Ajuda o usuário jovem a desenvolver sua identidade e tornar-se socializado como adulto ▪ Monitora bebês, babás e salas de aula ▪ Regenera a escrita e a correspondência ▪ Permite com que novas formas criativas de arte sejam expressadas

¹ balcanização: separação ou fragmentação de um estado em vários.² anomia: ausência de leis ou de organização = anarquia.

As consequências listadas no Quadro 1 podem ser vistas frequentemente nos noticiários, tanto as utópicas como as distópicas. Elas convivem junto na internet e nas mídias sociais, do mesmo modo incentivam a tolerância, incitam o ódio, permitem novas formas de arte e também o plágio. Os conflitos existentes geram dúvidas quanto à veracidade das atitudes das pessoas nas mídias sociais, dificultando a real identificação dos usuários na rede.

Em 1988, Belk apresentou seu estudo sobre a extensão do “eu” (*self*), que descreveu como as pessoas consideram objetos como parte de sua identidade, normalmente por trazer reconhecimento ou mostrar posição social. Com a internet muitos deles foram digitalizados (músicas, livros, fotografias, etc.) e as mudanças observadas fizeram Belk (2013) rever o estudo sobre o “eu” estendido para o mundo digital, que incluiu 5 (cinco) dimensões digitais que elucidam muitas das atitudes dos usuários nas mídias sociais, as quais apresentam-se a seguir.

Dematerialization: Livros, músicas, fotos, foram transformados em arquivos de dados e são armazenados em repositórios na rede (como o *YouTube*, *Dropbox*, *Flickr*, *Instagram*, etc.), os quais o usuário não sabe onde fisicamente estão guardados. Os usuários podem enviar as suas músicas e fotos para seus contatos, sem que fiquem impossibilitados de usá-los. O ato de emprestar um álbum de música e ficar sem ele até que fosse devolvido, deixou de existir, e a posse destes objetos perdeu sua força no “eu”. Mesmo assim, a posse de objetos virtuais pode também gerar a sensação de *status*, porém não da mesma maneira que a posse física de um objeto.

Reembodiment: As pessoas passaram a ter mais de uma identificação na internet, e neste quesito, as mídias sociais são as maiores incentivadoras. O usuário pode ser quem ele quiser, construir um indivíduo completamente diferente dele próprio, construindo uma nova visão do que é o “eu”. Mudanças como nome, imagem (*Avatars*, como são

chamadas no mundo digital), são manipuladas para apresentar seus usuários de uma forma igual ou completamente diferente de sua imagem no mundo real.

Sharing: A facilidade em compartilhar informações reforça a mudança de comportamento das pessoas com as mídias sociais. Fotos, vídeos, músicas, dados pessoais, etc. são todos compartilhados em tempo real. Isto não garante a veracidade de todas as informações compartilhadas. Confissões, histórias e estórias são contadas conforme as imagens criadas do “eu”, e os usuários podem perder os seus controles.

Co-Construction of self: A aceitação e o envolvimento de outros usuários aos dados compartilhados (fotos, comentários, etc.) fazem parte da construção do “eu” nas mídias sociais. No processo de criação de um *Avatar*, o usuário vai compondo o seu “eu” conforme a aceitação e rejeição de seus amigos virtuais. Este procedimento é muito realizado nos jogos *online*, onde o seu personagem precisará ter uma personalidade conforme seus objetivos no mundo virtual criado pelo jogo.

Distributed memory: Neste novo mundo virtual, as memórias das pessoas passaram a ser fragmentos na Internet. Suas posses digitais, as mensagens trocadas, fotos, vídeos, etc. estão todos espalhados em vários repositórios. E a recuperação destas memórias pode ser totalmente seletiva, pois as pessoas possuem vários usuários nas diversas mídias sociais e podem ou não incluir todos estes usuários nas suas lembranças digitais, perdendo no passado as memórias indesejadas.

A expansão do mundo digital trouxe uma massiva utilização das mídias sociais e as empresas passaram a estudar formas para estarem próximas aos seus clientes nesta nova plataforma de comunicação. Kaplan e Haenlein (2010) realizaram um estudo sobre as oportunidades e os desafios da mídia social. Eles definem mídias sociais como locais com Conteúdo Gerado pelo Usuário (CGU).

Em sua definição, o CGU precisa cumprir três requisitos básicos:

- Precisa ser publicado em um site acessível ao público ou em um site de rede social acessível a um grupo selecionado de pessoas;
- Precisa mostrar uma certa quantidade de esforço criativo;
- Precisa ter sido criado fora das rotinas e práticas profissionais.

A primeira condição exclui o conteúdo trocado em e-mails ou mensagens instantâneas. A segunda, exclui a mera replicação de conteúdo já existente; e o terceiro, todo o conteúdo criado com um contexto comercial em mente.

Além disso, eles classificaram 6 formas de mídias sociais por: “presença social / riqueza da mídia” e “auto-apresentação / auto-divulgação” (Quadro 2). As 6 formas de mídias analisadas foram: plataforma colaborativa, *blogs*, comunidades de conteúdo, sites de redes sociais, mundos de jogos virtuais e mundos sociais virtuais.

As plataformas colaborativas habilitam a criação simultânea de conteúdo por vários usuários, e vêm sendo utilizadas para a propagação de conhecimento entre as empresas, seus colaboradores e clientes.

Os *blogs* representam o início das mídias sociais e podem ser comparados com as atuais páginas dos usuários das redes sociais. Normalmente são gerenciados por um único usuário e têm um conteúdo específico. Várias empresas utilizam *blogs* para divulgar informações para seus acionistas e colaboradores. Assim como as plataformas colaborativas, eles também possuem riscos de os usuários manifestarem descontentamento ou protestos em suas páginas.

As comunidades de conteúdo compartilham todo o tipo de mídia nas plataformas oferecidas. Assim como citado no estudo de Belk (1988), vídeos, fotos, livros, apresentações, etc. são compartilhados abertamente ou em comunidades fechadas nas mídias sociais específicas do tipo de conteúdo ou não. As empresas passaram a usar as mais variadas plataformas para promover e divulgar seus produtos e serviços, assim

como utilizam plataformas de vídeos para divulgar suas marcas e explicar o funcionamento e utilização de seus produtos e serviços.

As redes sociais são ambientes onde os usuários criam perfis com informações pessoais, convidam amigos e colegas para acessarem seus perfis e enviarem mensagens entre eles. Como já comentado, estas informações podem ou não representar a verdade sobre seus usuários. Assim como nas outras formas de mídias sociais, as empresas buscam este ambiente para divulgar suas marcas, produtos e serviços.

As outras duas formas consideradas de alta “presença social / riqueza de mídia” são ligadas em plataformas comerciais específicas e são exploradas pelas empresas proprietárias dos sistemas. Os seus usuários também criam seus *Avatares* e podem conquistar, adquirir ou comprar capacidades para interagir nestes mundos.

		Presença social / Riqueza da mídia		
		Baixo	Médio	Alto
Auto-apresentação / Auto-divulgação	Alto	<i>Blogs</i>	Sites de redes sociais (Ex. <i>Facebook</i>)	Mundos sociais virtuais (Ex. <i>Twinity</i>)
	Baixo	Plataforma colaborativa (Ex. <i>Wikipedia</i>)	Comunidades de Conteúdo (Ex. <i>YouTube</i>)	Mundos de jogos virtuais (Ex. <i>Warcraft</i>)

Quadro 2 – Classificação das mídias sociais por presença social / riqueza da mídia e auto-apresentação / auto-divulgação (Kaplan e Haenlein, 2010).

As empresas buscam a participação nas mídias sociais pelo fato conhecido de as marcas formarem uma extensão do conceito do “eu” do consumidor (Belk, 1988), oferecendo um método de auto-expressão e auto-definição. As marcas permitem que os consumidores expressem a adesão a um grupo. Estes grupos ou tribos existem quando os membros se identificam, e as marcas facilitam esta ação social coletiva (Goulding, Shankar e Canniford, 2013).

Como visto anteriormente, com a crescente utilização das variadas mídias sociais, os seus usuários passaram a formar grupos sobre os mais variáveis assuntos. As participações nos grupos identificam suas preferências e afinidades. Esta movimentação social passou a ser estudada e foi nominada como tribalismo (Cova, B., 1997).

2.2. Tribalismo

No estudo sobre comunidade e consumo de 1997, Cova, B. utiliza o termo tribalismo para contrapor o individualismo de adquirir produtos e serviços de suas próprias casas sem a necessidade de um contato social. Este processo de egocentrismo, incentivado pelo desenvolvimento e o largo uso de computadores, gerou uma grave dissolução social e um severo individualismo. Porém, o individualismo gerou um movimento contrário, onde as pessoas passaram a buscar contatos sociais. Este movimento passou a ser chamado de tribalismo, ou tribos pós-modernas (Cova, B. e Cova, V., 2002). As dinâmicas sociais pós-modernas podem metaforicamente ser definidas como "tribos" porque, como as tribos das sociedades arcaicas:

- Eles não podem confiar no poder central para manter a ordem social ou coagir seu círculo eleitoral em submissão a regras coletivas (raramente eles têm regras claramente codificadas para as quais a submissão poderia ser exigida);

- Constituem um ato coletivo que representa um contrapoder para o poder institucional;

- Eles reúnem as pessoas em torno de algo racional e moderno - um projeto, uma ocupação profissional, a noção de progresso, mas em torno de elementos não racionais e arcaicos - localidade, parentesco, emoção e paixão; e

- Estão próximos de clãs e outros agrupamentos étnicos.

Cova, B. e Cova, V. (2002) também chamam atenção para as diferenças entre pessoas de uma tribo e as escolhidas por segmentação;

- Um segmento é definido como um grupo de pessoas homogêneas (compartilham as mesmas características) não ligadas umas às outras; um segmento não é capaz de ações coletivas, os seus membros são simples consumidores;

- Uma tribo é definida como uma rede de pessoas heterogêneas em termos de idade, sexo, renda, etc., porém, ligadas por uma paixão ou emoção compartilhada; uma tribo é capaz de ação coletiva, seus membros não são simples consumidores, eles também defendem a tribo.

As tribos pós-modernas possuem algumas diferenças claras com as tribos arcaicas:

- São grupos efêmeros e não totalizadores. As tribos arcaicas eram permanentes e totalizavam;
- Uma pessoa pode pertencer a várias tribos pós-modernas. Em uma tribo arcaica, uma pessoa só poderia pertencer a uma tribo;
- Os limites de uma tribo pós-moderna são conceituais. Eles eram físicos nas tribos arcaicas;
- Os membros de uma tribo pós-moderna estão relacionados por sentimentos compartilhados e sinais apropriados. Membros de tribos arcaicas foram relacionados por parentesco e dialeto.

No livro “*Linked*”, Barabási (2014) faz uma compilação de vários estudos sobre assuntos já comentados nesta seção. Em um destes estudos, Cass Sustein (2001), professor de direito da Universidade de Chicago fez um questionário com 60 sites sobre política. Nas respostas, 60% das conexões destes sites foram criadas por terem afinidade sobre o mesmo assunto, enquanto somente 15% conectaram-se com outros de opinião contrária, levantando a suspeita sobre a internet encorajar a segregação e a fragmentação, como já foi comentado anteriormente. Porém, esta fragmentação reforça a definição das tribos, onde um usuário se conecta com assuntos de seu interesse, indicando que para definir as tribos basta verificar a lista de conexões realizadas por seus participantes nas mídias sociais.

Em outra passagem, Barabási (2014) comenta o problema de definir as comunidades baseadas na *WEB*, onde “nós todos as conhecemos ao vê-las, mas todo mundo tem critérios ligeiramente diferentes para defini-las”. Esta razão reforça a situação de não existirem fronteiras nítidas entre as várias comunidades, podendo existir sobreposição entre elas, dificultando a utilização das técnicas de agrupamento para separar os usuários por áreas de interesse.

Booth e Matic (2011) sugerem a utilização de influenciadores para identificar e atuar em um grupo específico, promovendo as marcas, produtos e serviços. Com estratégias de marketing bem definidas nas mídias sociais e utilizando influenciadores como parte da comunicação pode-se minimizar os impactos gerados caso algum usuário faça um comentário negativo sobre alguma parte da ação de marketing planejada.

Escalas e Bettman (2015) retornam ao conceito trabalhado por Belk (1988) onde as pessoas se envolvem em comportamentos de consumo em parte para construir seus próprios conceitos e criar identidades pessoais; porém, estudaram o endosso das celebridades na construção destas identidades, pois as celebridades personificam várias características e propriedades simbólicas sendo úteis aos consumidores na construção dos “eu”. Marcas com consumidores e admiradores com forte envolvimento emocional formam tribos com fortes laços e recebem atributos com um significado intrínseco muito maior quando comparado com o valor de suas características ou benefícios instrumentais. Porém, o uso de celebridades para aumentar o engajamento da tribo com a marca só é percebido como positivo se o uso da marca ou produto pela celebridade faz sentido. Sendo assim, a autenticidade da celebridade ou influenciador é um fator crítico para criar ligações emocionais positivas com os usuários de uma tribo (Kowalczyk, et al., 2016).

Considerando as conexões criadas entre os usuários e suas interações nas mídias sociais, a filtragem colaborativa surge como uma alternativa para identificar os usuários

pertencentes a uma mesma tribo, onde suas conexões em comum apresentam uma tendência do grupo. Este assunto está mais detalhado na seção 2.4.

2.3. Netnografia

O termo netnografia começou a ser utilizado no final do século 20 e foi descrito por Kozinets (1998 e 2002) como um novo método qualitativo adaptado da etnografia e concebido especificamente para investigar o comportamento do consumidor nas culturas e comunidades presentes na Internet.

Para se conceituar netnografia, é importante entender o significado de etnografia como a análise do comportamento social de um público em seu ambiente natural, em que o pesquisador deve se envolver com o grupo analisado, aprender sua linguagem em uso (dialetos, jargões, especial uso das palavras e neologismos), imergir por grande período de tempo para que aumente a probabilidade de encontrar momentos importantes nos eventos comuns do grupo analisado (Elliott e Jankel-Elliott, 2003).

Willis e Trondman (2000) descrevem a etnografia como uma família de métodos que envolvem contato social direto e sustentado com os agentes, e de ricamente redigir o encontro, respeitando, registrando, representando pelo menos parcialmente em seus próprios termos, a irredutibilidade da experiência humana. A etnografia é a disciplinada e deliberada testemunha registradora de eventos humanos.

Os estudos etnográficos buscam sempre observar os indivíduos em seus ambientes naturais para evitar as limitações verificadas nos “questionários de pesquisa” e os casos onde “as pessoas nem sempre fazem o que dizem” (Elliott e Jankel-Elliott, 2003).

Elliott e Jankel-Elliott (2003) falam sobre os princípios da etnografia, como o estudo do comportamento em ambientes naturais. A necessidade de compreender o mundo simbólico dos sujeitos estudados, vendo o mundo pelos seus olhos e aprendendo

o idioma em uso (dialeto, jargão, usos especiais de palavras e neologismos). Presença prolongada no campo, aumentando a probabilidade de encontrar espontaneamente momentos importantes nos eventos comuns diários. Por fim, a participação na vida cultural para desenvolver uma compreensão dos significados culturais e simbólicos das regras locais.

A internet fornece várias formas de analisar um público em seu “*habitat natural*”. Os usuários começaram a se agrupar desde a criação dos servidores de listas de emails, salas de *chat* e plataformas colaborativas. Kozinets (1998) considera que estes grupos sociais têm uma real existência para seus participantes, e portanto têm efeitos em muitos aspectos dos seus comportamentos.

Os métodos netnográficos de análise dos grupos nas mídias sociais não necessitam de muitos investimentos, sendo mais rápidos, discretos e fornecem muito mais informações quanto comparados com os métodos etnográficos tradicionais como *focus groups* ou entrevistas (Costello, McDermott e Wallace, 2017)

Alguns dos benefícios da utilização da netnografia incluem anonimato, cocriação, abundantes formas de comunicação e variedade de dados. O pesquisador pode “vestir” um *Avatar* apropriado à sua pesquisa e interagir com uma comunidade ou simplesmente permanecer anônimo observando as atividades dos participantes sem ser notado.

Costello et al. (2017) citam a crescente utilização do uso da cocriação com comunidades de marcas em tempo real onde a netnografia mostra ser o método mais adequado para examinar esta atividade. Este estudo de 2017 também verificou os recentes estudos netnográficos publicados e aponta o crescimento da preferência em utilizar a netnografia “observacional” e “não-participativa” (passiva), e aproveitar as oportunidades para usar as participativas em tempo real (ativas) em benefício das comunidades das mídias sociais, como no caso da cocriação.

A busca em identificar as comunidades virtuais passa pela ação de identificar suas celebridades e influenciadores, assim como, encontrar os usuários mais mencionados e assuntos mais comentados. A coleta das informações é realizada pelas API fornecidas pelas mídias sociais e por meio dos dados lidos pode-se compreender o comportamento dos seus usuários, sendo um típico caso da aplicação da netnografia.

2.4. Filtragem colaborativa

A filtragem colaborativa é um método normalmente utilizado para recomendar itens aos consumidores em sites de compras e serviços, realizando a análise dos hábitos de compras e as classificações dadas aos produtos pelos consumidores; assim, quando o consumidor realiza a compra de um novo produto, o sistema verifica quais produtos se adequam ao seu perfil, oferecendo produtos e serviços adquiridos por pessoas de perfil similar (Anderson, Ball, et al., 2003).

O sistema de recomendação Tapestry foi o primeiro a utilizar o termo “filtragem colaborativa” (Golberg, Nichols, Oki e Terry, 1992). O sistema foi projetado para recomendar documentos criados a partir de um *streaming* de notícias. O conceito foi criado para alavancar a colaboração social, prevenindo os usuários de ficarem inundados pelo grande volume do *streaming* de dados.

O conceito utilizado pelo método consiste em coletar as opiniões dos usuários para os produtos de um grupo selecionado, explorando as semelhanças no comportamento de classificação entre vários usuários na análise do produto (Sammut e Webb, 2010).

Uma representação da utilização do método pode ser observada na Figura 1. A matriz de n usuários por m itens representa as preferências, onde cada célula $v_{u,i}$ contém as avaliações dadas para o item i pelo usuário u . Na maioria das vezes esta matriz é bem espaçada, uma vez que a maioria dos usuários não avaliam todos os itens. Sendo assim,

recomendar é a tarefa de prever qual seria a avaliação que o usuário u daria para um item não avaliado.

		<i>Itens</i>						
		1	2	3	$..$	i	$..$	m
<i>Usuários</i>	1	4	5	2		3	4	
	2	3		4	3			2
	3		3			2	3	
	$:$				2			
	u	2		1			1	
	$:$		1		2			
	a	4	4			?	2	1
	$:$		3	2		2		3
	n	3				4	4	1

Figura 1: Matriz de avaliações de itens por usuários.

Um método muito utilizado é baseado na vizinhança (*neighborhood-based method*), também chamado de abordagem baseada em memória (*memory-based approach*) (Breese, Heckerman, & Kadie, 1998).

A técnica baseada na vizinhança começa com a escolha de um grupo de usuários similares com o usuário em consideração para recomendação (usuário ativo), e uma combinação ponderada de suas avaliações é utilizada para prever a avaliação dos itens não avaliados. Os procedimentos comuns da técnica são: atribuir um peso aos usuários em relação à proximidade do usuário ativo, selecionar K usuários mais similares, calcular a predição com a combinação ponderada das avaliações dos usuários selecionados.

O banco de dados consiste em um conjunto de avaliações $v_{u,i}$ correspondendo às avaliações realizadas pelo usuário u para o item i . Se I_u é um conjunto de itens avaliados pelo usuário u , então a média das avaliações do usuário u pode ser escrita como:

$$\bar{v}_u = \frac{1}{|I_u|} \sum_{i \in I_u} v_{u,i}$$

Sendo o usuário ativo definido como a , e I o grupo de itens avaliados por ambos os usuários (a e u), tem-se $p_{a,i}$ como a predição da avaliação de a para o item i e $w_{a,u}$ a avaliação do item i dada pelo usuário u . Utilizando o coeficiente da correlação de Pearson entre os dois usuários (Resnick et al., 1994), pode-se definir:

$$w_{a,u} = \frac{\sum_{i \in I} (v_{a,i} - \bar{v}_a)(v_{u,i} - \bar{v}_u)}{\sqrt{\sum_{i \in I} (v_{a,i} - \bar{v}_a)^2 \sum_{i \in I} (v_{u,i} - \bar{v}_u)^2}}$$

$$p_{a,i} = \bar{v}_a + \frac{\sum_{u \in K} (v_{u,i} - \bar{v}_u) \times w_{a,u}}{\sum_{u \in K} w_{a,u}}$$

A similaridade baseada na correlação de Pearson mede a extensão da associação linear entre duas variáveis. Como alternativa, pode-se tratar as avaliações de dois usuários como um vetor em um espaço m -dimensional e calcular a similaridade baseada no cosseno do ângulo entre eles, dada por:

$$w_{a,u} = \cos(\vec{v}_a, \vec{v}_u) = \frac{\vec{v}_a \times \vec{v}_u}{\|\vec{v}_a\|^2 \times \|\vec{v}_u\|^2}$$

$$= \frac{\sum_{i=1}^m v_{a,i} \times v_{u,i}}{\sqrt{\sum_{i=1}^m v_{a,i}^2} \times \sqrt{\sum_{i=1}^m v_{u,i}^2}}$$

Neste método não podem existir avaliações negativas, e os itens não avaliados são tratados como avaliados com zero.

Os algoritmos baseados na vizinhança não escalam muito bem se aplicados em milhões de usuários e itens, devido à complexidade computacional da pesquisa da similaridade de usuários. Em 2003, Linden, Smith e York apresentaram uma proposta de fazer a filtragem colaborativa item por item, onde ao invés de procurar os usuários similares apropriados, eles combinaram os itens avaliados pelos usuários com itens similares, deixando os sistemas *online* mais rápidos e muitas vezes com melhores recomendações.

As similaridades entre pares de itens i e j são processadas usando correlação de Pearson, como segue:

$$w_{i,j} = \frac{\sum_{u \in U} (v_{u,i} - \bar{v}_i)(v_{u,j} - \bar{v}_j)}{\sqrt{\sum_{u \in U} (v_{u,i} - \bar{v}_i)^2} \sqrt{\sum_{u \in U} (v_{u,j} - \bar{v}_j)^2}}$$

onde U é o conjunto de todos os usuários avaliadores de ambos os itens i e j , $v_{u,i}$ é a avaliação dada pelo usuário u ao item i , e $\bar{v}_i$ é a média da avaliação do i -ésimo item entre os usuários.

Considerando K o conjunto dos k itens avaliados por a e mais similares a i , a predição pode usar uma simples média ponderada como:

$$p_{a,i} = \frac{\sum_{j \in K} (v_{a,j} \times w_{i,j})}{\sum_{j \in K} |w_{i,j}|}$$

A filtragem colaborativa vem sendo utilizada como método para identificar itens similares de um grupo de usuários. A ideia da similaridade pode ser utilizada para identificar vários tipos de produtos (Anderson, Ball, et al., 2003).

Esta característica do método pode ser utilizada para identificar usuários com similaridades de conexões e assuntos nas mídias sociais, como se as conexões e os assuntos fossem produtos e as quantidade de usuários conectados a eles fossem suas pontuações.

3. ASPECTOS METODOLÓGICOS

A seguir serão descritos os procedimentos utilizados para a criação do método de segmentação proposto, assim como a pesquisa para a sua verificação.

O método leva em consideração o comportamento dos usuários nas redes sociais, onde os estudos descritos na seção 2 são verificados e aplicados na identificação das tribos formadas entre os usuários com o intuito de criar uma forma para identificar outros usuários dessas tribos.

3.1. Tipologia da Pesquisa

A primeira parte deste estudo consiste em criar a lista dos usuários mais influentes para um determinado assunto ou tribo. A lista será gerada a partir dos dados secundários obtidos na rede social utilizada, o *Twitter*, e servirá como base para identificar os usuários simpatizantes com tribo analisada.

A abordagem utilizada é qualitativa e consiste em analisar as atitudes de uma tribo na mídia social. Esta abordagem foi explicada na seção 2 e é denominada netnografia. Este estudo irá verificar as conexões existentes mais evidentes na tribo, para analisar a influência dos usuários selecionados dentro do grupo analisado, definindo os mais engajados como os que têm mais conexões importantes dentro da tribo.

Portanto, com a definição da lista de usuários mais influentes de uma tribo, pode-se comparar qualquer usuário da mídia social em relação a esta lista visando detectar os de maior afinidade com ela. Esta afinidade será calculada conforme pesos atribuídos para cada um dos usuários da lista segundo sua influência, e assim cada usuário analisado terá um valor que determinará a afinidade que ele tem com a tribo comparada.

O conceito utilizado é o mesmo abordado pela filtragem colaborativa. Nela, as análises de um grupo de usuários servem para prever as preferências dos outros usuários similares. Em toda a compra realizada por um usuário, o método compara o perfil dele

com os perfis de todos os outros usuários da base de dados, e assim, o método sugere quais produtos ele pode se interessar em adquirir, prevendo seu interesse com base nos itens já adquiridos por usuários similares a ele.

O método proposto pode ser validado com a realização de um experimento para medir o engajamento de usuários selecionados como tendo afinidade com a tribo escolhida, *versus* usuários selecionados pelo método de segmentação do *Twitter*.

O experimento consiste em enviar uma publicidade para os usuários selecionados e esperar o engajamento deles com o anúncio apresentado. Engajamento pode ser definido como qualquer forma de interação com o anúncio, como gostar do anúncio (*like*), republicar o anúncio (*retweet*), clicar para ver o anúncio em destaque, clicar no endereço eletrônico proposto, seguir o usuário anunciante, clicar nos termos destacados (*# - hashtags*) e/ou clicar nos usuários mencionados (*@ - arrobas*) (Evans, 2010).

3.2. Procedimento de segmentação

A ideia central da segmentação utilizando a filtragem colaborativa passa pela necessidade de identificar um grupo de usuários influenciados pelo tema escolhido. Assim como é necessário ter um grupo de produtos analisados pelos usuários para identificar as pessoas similares, o método precisa identificar quais serão os “produtos” obrigatórios que definem a tribo, neste caso os usuários mais influentes identificados como pertencentes à tribo.

Após a identificação dos usuários definidores da tribo, cria-se a lista de usuários “analisadores”. Na filtragem colaborativa eles serão os usuários similares dos produtos já analisados e em conjunto serão a forma de identificar os outros usuários similares a eles.

Portanto, o método tem três etapas: a seleção dos usuários influenciadores, chamados de “sementes”; a seleção do grupo de usuários “avaliadores” da tribo, aqueles

que seguem todas as sementes, e aqui chamados de “geradores”; por fim, a criação da lista de usuários mais seguidos pelos geradores e a identificação da influência de cada um deles no grupo, chamada de “*DNA da tribo*” ou somente “*DNA*”.

A Figura 2 mostra um esquema representativo do método, exemplificando a relação entre as sementes, os geradores e o *DNA da tribo*.

Figura 2: Esquema do método.

Fazendo uma analogia com a Figura 1, os *DNAs* seriam os “usuários”, e os “itens” todos os usuários da mídia social a serem classificados por cada *DNA*, definindo, assim, sua afinidade com eles. Porém, os *DNAs* são criados pelas preferências dos geradores, e estes passam a ser os “usuários” de uma outra estrutura de filtragem colaborativa, onde os “itens” são os participantes do *DNA*. Por fim, a predição sobre um usuário qualquer da mídia social pertencer à tribo é feita pela soma dos pesos dados aos participantes do *DNA* seguidos por ele.

Portanto, o método consiste em uma filtragem colaborativa tridimensional, onde existe o eixo das tribos (nomes atribuídos a elas), o do *DNA* e o dos usuários da mídia social.

A Figura 3 exemplifica a matriz das tribos e usuários, e na Figura 4 está representada a definição do peso do usuário a como a intersecção dos conjuntos dos seus “seguindo” (usuários seguidos pelo usuário a) com o DNA .

		<i>Usuários</i>						
		1	2	3	$..$	a	$..$	m
<i>Tribos</i>	1		80	20		20	90	
	2		10		47		75	98
	3	25				39		
	$:$		36		97			89
	t		49	47		?	78	
	$:$	68		75	63			76
	n	89		77			39	

Figura 3: Matriz colaborativa do método proposto

Figura 4: Conjuntos de definição do peso $w_{t,i}$

Assim, a afinidade de cada usuário da mídia social com uma tribo pode ser descrita como a predição desta matriz tridimensional, como segue:

$$p_{t,a} = \sum_{i \in D \cap A} w_{t,i}$$

onde t é a tribo, a o usuário da mídia social a ser avaliado, A o conjunto dos seguidos por a , D o conjunto de usuários do DNA e i os seus participantes. O peso $w_{t,i}$

será definido pela influência dos participantes do *DNA* perante a tribo e será descrito a seguir.

3.2.1. Definição das sementes

O conceito de sementes poderia ser confundido como a definição dos influenciadores do assunto proposto. Em parte, esta analogia não está de todo errada; porém, muitos dos influenciadores entram em outra categoria identificada como celebridades. Uma celebridade pode até influenciar uma tribo; porém, ela possui muitos seguidores não pertencentes à tribo. Os atletas celebridades são um bom exemplo; eles podem até influenciar tribos do esporte que praticam, mas muitos de seus seguidores podem não o praticar ou até chegar ao extremo de odiá-lo.

Existem alguns algoritmos criados para identificar influenciadores, e um deles foi descrito no estudo realizado por Booth e Matic (2011), onde é criada uma lista de atributos a serem considerados para a escolha de um bom influenciador. As informações quantitativas de cada atributo são pesquisadas na internet e nas mídias sociais, sendo atribuído um percentual subjetivo para a participação de cada um destes atributos na definição do peso do influenciador.

Uma particularidade do método é a necessidade de todas as sementes serem seguidas pelos usuários identificados como geradores. Sendo assim, a seleção das sementes passa a ser a busca por usuários engajados e influentes com a tribo selecionada, que não podem, porém, ser celebridades.

Não é difícil identificar usuários envolvidos com um assunto específico, mas estes também devem ter um grau de influência na tribo. A influência pode ser medida analisando as publicações do usuário na mídia social, verificando se ele escreve sobre assuntos de interesse do grupo e se seus comentários possuem engajamentos. As

sementes devem estar ativas, sendo usuários que publicam e recebem engajamentos constantemente.

Com as recomendações listadas nesta seção, as sementes iniciais podem ser escolhidas conforme a lista de atributos a seguir:

- Sejam identificadas pelos mecanismos de busca da mídia social com os termos associados à tribo,
- Tenham mais de 500 seguidores,
- Tenham publicado nos últimos 15 dias.

Estas recomendações não garantem que as sementes serão as melhores para definir a tribo; porém, as selecionadas inicialmente podem ajudar a escolher as melhores.

O método consiste em selecionar 10 (dez) semente com uma busca na mídia social, utilizando os termos mais usados pela tribo. Exemplo, “MMA Brasil” para identificar pessoas com afinidade com as artes marciais e lutas de *MMA (Mixed Martial Arts)*, em tradução livre, Artes Marciais Mistas).

Na Tabela 1 estão as 10 sementes selecionadas pelo exemplo apresentado nesta seção.

Nome de Tela	Nome	Quantidade de Seguidores
mmabrazil	MMA Brasil	30.561
MMAPARA	MMA PARÁ	2.981
pezao011	Marcos Pezão	2.063
ricardodemente	Ricardo Demente	3.095
ShootoMMABrasil	Shooto Brasil MMA	4.001
SuperesportesMG	Superesportes MG	101.079
TeamNogueirasc	Team Nogueira Jlle	994
TMarretaMMA	Thiago Marreta	4.379
ufc_brasil	UFC BRASIL	1.591.000
Valmir_MAISMMA	MAISMMA	2.822

Tabela 1: Sementes selecionadas pela busca na mídia social.

Uma vez identificadas as 10 (dez) primeiras sementes utilizando a interface de busca disponibilizada pela mídia social e seguindo os critérios descritos nesta seção, um programa lê todas as mensagens publicadas pelos usuários escolhidos nos últimos 15 dias.

As mensagens publicadas por estas 10 sementes podem ser utilizadas para identificar a existência de influenciadores entre elas. Os usuários mencionados por mais de uma semente são candidatos a usuários conhecidos na tribo escolhida, e podem ser incluídos na lista de sementes, caso ainda não estejam presentes na lista das 10 sementes escolhidas inicialmente.

Sendo assim, o mesmo programa pode selecionar todas as menções aos outros usuários (@), e caso algum outro usuário não pertencente à lista inicial tenha sido mencionado por mais de 1 (um) dos 10 (dez) usuários selecionados inicialmente, ele será incluído na lista e suas últimas mensagens serão coletadas. Com isso, verifica-se novamente a existência de novas sementes. Este processo será repetido até não existir mais nenhum usuário novo a ser incluído.

Neste exemplo, os usuários mencionados por mais de 1 (um) usuário da lista de 10 (dez) usuários iniciais são apresentados na Tabela 2:

Nome de Tela	Quantidade de comentários	Quantidade de usuários que comentaram
canalCombate	44	4
demianmaia	10	4
FabricioWerdum	26	3
bethecorreia	14	3
JessicaMMApro	22	2

Tabela 2: Usuários mais comentados pelas primeiras 10 sementes selecionadas.

Após a inclusão destas novas sementes, passamos a ter uma lista com 15 sementes e o sistema pesquisa novamente os usuários mais seguidos por mais de uma semente nesta nova lista. Neste caso, todos os usuários seguidos por mais de um usuário já estavam na lista das 15 sementes. A lista final das 15 sementes selecionadas está na Tabela 3.

Nome de Tela	Nome	Quantidade de Seguidores	Quantidade de publicações nos últimos 15 dias
bethecorreia	Bethe	26.827	38
canalCombate	Combate	616.688	387
demianmaia	Demian Maia	476.274	96
FabricioWerdum	Fabricio Werdum	297.811	16
JessicaMMApro	Jessica Andrade	26.752	21
mmabrasil	MMA Brasil	30.561	242
MMA PARA	MMA PARÁ	2.981	134
pezao011	Marcos Pezão	2.063	30
ricardodemente	Ricardo Demente	3.095	10
ShootoMMABrasil	Shooto Brasil MMA	4.001	7
SuperesportesMG	Superesportes MG	101.079	934
TeamNogueirasc	Team Nogueira Jlle	994	13
TMarretaMMA	Thiago Marreta	4.379	56
ufc_brasil	UFC BRASIL	1.591.000	397
Valmir_MAISMMA	MAISMMA	2.822	61

Tabela 3: Lista final das sementes selecionadas.

Os geradores são usuários seguidores das sementes; porém, ninguém foi identificado seguindo todas as sementes da lista de 15 selecionadas. Devido a esse cenário, passa a ser necessário escolher quais das 15 sementes e quantas delas serão necessárias para fazer a lista de geradores.

As sementes mais importantes devem ter seguidores mais engajados com elas e publicar sobre assuntos e pessoas mais relevantes na tribo, além de ter uma quantidade de seguidores e publicações coerentes com os outros usuários da tribo (Booth e Matic, 2011).

Em conformidade com isso, são observados no conteúdo da Tabela 3 usuários com muitos seguidores e outros nem tanto. A mesma análise pode ser feita com a quantidade de publicações no período definido. A média de seguidores é pouco mais de 212 mil e a de publicações 163. Utilizando a técnica de identificação de *outlier* (Fávero, Belfiore, Silva e Chan, 2009), são observados 3 usuários nesta condição. Os usuários *ufc_brasil* e *canalCombate* são *outliers* pela quantidade de seguidores (mais de 494.055,5) e o usuário *SuperesportesMG* pela quantidade de publicações (mais de 442,25).

Para classificar as outras 12 sementes é necessário verificar suas publicações e o seu poder de engajar os usuários. Seguindo o conceito criado por Booth e Matic (2011), foram identificadas 4 informações, extraídas de um conjunto de 15 variáveis.

As variáveis usadas, coletadas nos cadastros dos usuários e nas suas publicações são:

- Quantidade total de seguidores (1)
- Quantidade de publicações (2)
- Quantidade de pessoas que gostaram das publicações (*like*) (3)
- Quantidade de republicações realizadas por outros usuários (*retweet*) (4)
- Identificados os 5 termos destacados (#) mais usados pelas sementes, verifica-se quais deles a semente utilizou, atribuindo 1 para a *hashtag* utilizada e 0 para a não utilizada (5-9)
- Identificados os 5 usuários mais mencionados (@) pelas sementes, verifica-se quais deles a semente mencionou, atribuindo 1 para a *arroba* utilizada e 0 para a não utilizada (10-14)
- Quantidade das outras sementes que são seguidoras da semente analisada (15)

As 4 informações extraídas são:

- Quantidade de seguidores (1)
- Relevância da publicação: soma da quantidade de *likes* e *retweets* dividida pela quantidade de publicações ($(3+4) / 2$)
- Quantidade dos assuntos mais comentados: soma da quantidade de # e @ utilizadas nas publicações (soma de 5 a 14)
- Quantidade das outras sementes seguidoras da semente analisada (15)

Após a coleta dos dados, as 4 informações são normalizadas e analisadas pela soma das informações normalizadas por cada semente e a redução fatorial das 4 informações para um único fator. Aplicando este procedimento ao exemplo, tem-se a Tabela 4:

Nome de Tela	Nome	Seguidores	Relevância Publicação	Quantidade # e @	Seguidores Sementes	Total	Fator
demianmaia	Demian Maia	2,65434	2,52589	-0,14952	1,54220	6,57290	2,54821
FabricioWerdum	Fabricio Werdum	1,47908	0,92444	-0,14952	1,15665	3,41065	1,33470
mmabrazil	MMA Brasil	-0,28088	-0,61733	-0,14952	1,54220	0,49446	0,16583
Valmir_MAISMMA	MAISMMA	-0,46356	-0,71598	1,64477	0,00000	0,46524	-0,46970
MMA PARA	MMA PARÁ	-0,46251	-0,72489	1,64477	0,00000	0,45737	-0,47269
TMarretaMMA	Thiago Marreta	-0,45331	0,23573	0,74762	-0,38555	0,14450	-0,22500
bethecorreia	Bethe	-0,30547	0,52138	-0,14952	-0,38555	-0,31917	-0,05114
pezao011	Marcos Pezão	-0,46856	-0,65890	0,74762	0,00000	-0,37984	-0,44471
JessicaMMApro	Jessica Andrade	-0,30597	0,48694	-1,04667	-0,38555	-1,25125	-0,05932
ricardodemente	Ricardo Demente	-0,46176	-0,48749	-1,04667	-0,38555	-2,38148	-0,49457
ShootoMMABrasil	Shooto Brasil MMA	-0,45579	-0,74588	-1,04667	-0,77110	-3,01944	-0,71946
TeamNogueirasc	Team Nogueira Jlle	-0,47560	-0,74393	-1,04667	-1,92775	-4,19395	-1,11217

Tabela 4: Sementes com as informações normalizadas.

Analisando os dados da Tabela 4, existem 6 sementes com a coluna Total positiva e 3 sementes com a coluna Fator positiva. Sendo assim, consideraram-se 6 o número máximo de sementes relevantes e 3 o mínimo. A ordem de utilização das sementes é obtida primeiramente classificando as sementes positivas nas duas colunas descendentemente pelo Fator. Em seguida, são classificadas descendentemente pela coluna Fator as sementes com Fator positivo e Total negativo e por último classificadas pelo campo Total as sementes com Total positivo e Fator negativo.

Aplicando na Tabela 4 as classificações nas sementes com Total ou Fator positivos, obtém-se a Tabela 5:

Nome de Tela	Nome	Total	Fator
demianmaia	Demian Maia	6,572903	2,548210
FabricioWerdum	Fabricio Werdum	3,410645	1,334700
mmabrazil	MMA Brasil	0,494457	0,165830
Valmir_MAISMMA	MAISMMA	0,465240	-0,469700
MMA PARA	MMA PARÁ	0,457371	-0,472690
TMarretaMMA	Thiago Marreta	0,144504	-0,225000

Tabela 5: Sementes classificadas em ordem de relevância.

3.2.2. Definição dos geradores

Escolhidas as sementes, o próximo passo é a seleção dos geradores. Para isso, utilizam-se programas de coleta de dados para ler todos os usuários seguidores de cada uma das sementes selecionadas. Com todos os seguidores identificados, o processo consiste em verificar quantos seguidores estão em todas as sementes e nas n primeiras sementes juntas até o número mínimo de sementes determinado. No exemplo da Tabela 5, verificam-se as 6, 5, 4 e 3 primeiras sementes em conjunto.

Para não existirem geradores dissonantes perante os demais, são excluídos os considerados *outliers* pelas quantidades de seguidores e seguidos.

A Tabela 6 indica a quantidade de seguidores mútuos das sementes escolhidas.

Quantidade de Sementes	Quantidade de Geradores	Quantidade de Geradores sem Outliers
6	0	0
5	54	38
4	951	860
3	3.835	3.678

Tabela 6: Quantidade de geradores por quantidade de sementes em conjunto.

Para a geração do DNA pode-se escolher os seguidores mútuos de todas as sementes ou somente de 3 sementes. O primeiro caso é descartado por não existir seguidores mútuos das 6 sementes.

Conforme a filtragem colaborativa, quanto mais informação existente sobre os “produtos”, melhor é a identificação dos produtos similares. Neste caso, quanto mais sementes utilizadas, mais os geradores têm afinidade com o assunto escolhido.

Ainda conforme a filtragem colaborativa, uma maior quantidade de geradores irá criar um *DNA* mais afinado com as preferências da tribo de *MMA*.

Em razão destes fatos, a escolha de 4 sementes indica ser mais apropriada e utilizam-se 860 geradores para a criação do *DNA* de *MMA*.

3.2.3. Definição do *DNA* da tribo

Com os geradores identificados, os programas de leitura começam o trabalho de ler todas as informações dos usuários seguidos pelos geradores, assim como todas as suas publicações dos últimos 15 dias para garantir que os usuários mais atuantes na mídia social sejam os mais importantes no *DNA*.

Após a leitura de todos os dados, é feita uma tabela com todos os usuários lidos contendo as seguintes informações:

- Identificação do usuário (1)
- Quantidade de geradores seguidores (2)
- Quantidade total de seguidores (3)
- Quantidade de publicações realizadas (4)
- Quantidade de pessoas que gostaram das publicações (*like*) (5)
- Quantidade de republicações realizadas por outros usuários (*retweet*) (6)

- Identificados os 5 termos destacados (#) mais usados por todos os usuários lidos, verifica-se quais o usuário utilizou, atribuindo 1 para a *hashtag* utilizada e 0 para a não utilizada (7-11)
- Identificados os 5 usuários mais mencionados (@) por todos os usuários lidos, verifica-se quais a semente mencionou, atribuindo 1 para a *arroba* utilizada e 0 para a não utilizada (12-16)
- Quantidade dos outros usuários lidos seguindo o usuário analisado (17)

E da mesma forma feita com as sementes, são extraídas as seguintes 4 informações:

- Relevância de seguidores: Quantidade de geradores seguidores dividida pela quantidade total de seguidores (2 / 3)
- Relevância da publicação: soma da quantidade de *likes* e *retweets* dividida pela quantidade de publicações ((5+6) / 4)
- Quantidade dos assuntos mais falados: soma da quantidade de # e @ utilizadas nas publicações (soma de 7-16)
- Quantidade dos outros usuários lidos seguindo o usuário analisado (17)

Assim como foi feito, tanto na escolha das sementes, como na escolha de geradores, verifica-se a existência de *outliers* pela quantidade total de seguidores na lista de usuários do *DNA*, garantindo a coerência dos usuários remanescentes.

Outra necessidade é a definição da quantidade de usuários participantes do *DNA*. Esta quantidade pode ser definida escolhendo um número fixo de usuários ou a quantidade mínima de geradores seguindo o usuário. Como a lista de geradores é variável de acordo com a quantidade de seguidores das sementes, o *DNA* também não pode ser fixo e deve variar conforme a quantidade de geradores seguidores.

Na filtragem colaborativa as notas atribuídas aos produtos pelos usuários determinam suas preferências. Considerando a quantidade de geradores seguidores de um usuário do DNA como as notas recebidas, o corte da quantidade mínima de geradores seguidores determina as suas preferências por estes usuários do DNA.

Para este estudo, foram considerados os usuários do DNA seguidos por pelo menos 30% dos geradores como uma quantidade mínima para considerá-los como uma preferência válida do grupo.

No caso do *MMA*, foram selecionados todos os usuários seguidos por no mínimo 258 geradores.

Após a exclusão dos *outliers* e dos usuários seguidos por menos de 258 seguidores, o *DNA* da tribo *MMA* ficou com 263 usuários. Mas, ainda é necessário definir o peso de cada usuário no *DNA*, e a forma escolhida foi a aplicação da redução das 4 informações normalizadas em um único fator.

Executando a análise fatorial nos dados dos 263 usuários, tem-se a seguinte matriz de correlação na Tabela 7:

Matriz de Correlação					
		Rel. Seguidores	Rel. Publicações	Rel. # e @	Rel. Seg. DNA
Correlação	Rel. Seguidores	1,00000	-0,21000	0,00800	-0,34500
	Rel. Publicações	-0,21000	1,00000	0,00500	0,17900
	Rel. # e @	0,00800	0,00500	1,00000	0,23900
	Rel. Seg. DNA	-0,34500	0,17900	0,23900	1,00000

Tabela 7: Matriz de correlação dos dados do *DNA*

Para usar o valor do fator como peso, algumas transformações matemáticas se fazem necessárias, principalmente pela diferença do tamanho entre os *DNAs*, e pelo gradiente entre o maior e o menor fator poder variar conforme a tribo. Esta variação impossibilita a comparação entre os pesos atribuídos às tribos dos usuários da mídia

social. Para ser possível a comparação entre os pesos das tribos, é melhor a utilização de valores ponderados. Sendo assim, o peso $w_{t,i}$ pode ser escrito como:

$$w_{t,i} = \frac{f_{t,i} + 1 - \min(f_t)}{\sum_{d \in D} (f_{t,d} + 1 - \min(f_t))}$$

onde f é o valor atribuído pelo fator ao usuário do *DNA* e $\min(f)$ é o menor valor atribuído pelo fator a um usuário do *DNA*. A fórmula nunca irá assumir o valor zero e o maior valor atribuído a um usuário da mídia social para uma tribo será 1 (um), caso ele siga todos os usuários da tribo. Sendo assim, quanto maior a pontuação, maior é a afinidade do usuário com a tribo e a predição do usuário da mídia social $p_{a,i}$ passa a ser:

$$p_{t,a} = \sum_{i \in D \cap A} \left(\frac{f_{t,i} + 1 - \min(f_t)}{\sum_{d \in D} (f_{t,d} + 1 - \min(f_t))} \right)$$

Outra grande vantagem do método de atribuição do peso é garantir maiores pesos para os usuários mais influentes de forma matemática, sem subjetividade.

Portanto, qualquer usuário da mídia social pode ter sua afinidade com as tribos calculada pela comparação dos seus seguidos com a lista de usuários contida no *DNA* da tribo escolhida.

3.3. Amostragem

Criado o *DNA* da tribo de *MMA*, o experimento começa com a escolha de um grupo de usuários com afinidade com este esporte.

A escolha foi realizada de um universo de pouco mais de 2,5 milhões de usuários brasileiros do *Twitter*, selecionados de forma aleatória. Estes usuários possuem mais de 1,5 bilhões de conexões, as quais foram utilizadas para calcular a afinidade destes usuários em comparação com o *DNA*.

Para a realização da pesquisa foram selecionados os 40.000 usuários identificados com peso calculado pelo método proposto *DNA* maior do que 0,10 (item 3.2) e tenham interagido no *Twitter* nos últimos 30 dias, para garantir a utilização apenas de usuários ativos e utilizando a plataforma recentemente.

3.4. Experimento

Para verificar a afinidade da lista de 40.000 usuários com a tribo de *MMA*, foi criado 1 (um) experimento com 3 (três) campanhas publicitárias. O experimento utilizou a plataforma publicitária do *Twitter*.

As campanhas publicitárias continham as mesmas mensagens e imagem e somente eram diferenciadas pelo endereço eletrônico divulgado na página publicada na internet (figura 5). Esta diferença permitiu verificar as interações dos usuários com cada uma das três campanhas, pois cada endereço possuía uma codificação diferente para permitir a identificação da origem do acesso realizado à página da empresa.

Figura 5: Campanha publicitária.

O experimento foi iniciado às 07:00:00 do dia 18 de novembro de 2016 e encerrado às 06:59:59 do dia 21 de novembro de 2016. Este intervalo foi escolhido em virtude da realização de uma competição de *MMA* na cidade brasileira de São Paulo entre

os dias 19 e 20 de novembro de 2016, gerando uma grande movimentação nas mídias sociais sobre o assunto.

As três campanhas publicitárias iniciaram e terminaram no mesmo momento.

3.4.1. Primeira campanha publicitária: Controle.

A campanha de controle consiste em enviar a publicidade para usuários brasileiros da mídia social sem identificação de segmentação. Ela foi utilizada para verificar se o envio da mensagem para o grupo segmentado descrito na seção 3.3 fornece mais interações com a peça publicitária.

Método

Público e Procedimento: a plataforma publicitária do *Twitter* foi programada para enviar a campanha para todos os brasileiros de ambos os sexos durante o período do experimento. A escolha dos usuários foi realizada somente pelo *Twitter*.

Conforme os usuários utilizavam a mídia social, eles recebiam o anúncio da Figura 5 junto com outras mensagens de seus interesses. Cada interação realizada pelo usuário com o anúncio foi contabilizada pela plataforma publicitária.

O mesmo usuário pode ter recebido a mensagem mais de uma vez.

Resultados

A mídia social enviou um total 456.418 mensagens durante o período proposto e elas tiveram 5.962 interações dos usuários, sendo, portanto, uma taxa final de interações de 1,306%.

Período		CONTROLE		
Data	Período	Impressões	Interações	Taxa CONTROLE
18.11.2016	07:00 às 12:59	80.376	1.117	1,390%
18.11.2016	13:00 às 18:59	41.080	376	0,915%
18.11.2016	19:00 às 00:59	38.174	442	1,158%
19.11.2016	01:00 às 06:59	24.997	417	1,668%
19.11.2016	07:00 às 12:59	57.421	784	1,365%
19.11.2016	13:00 às 18:59	38.078	378	0,993%
19.11.2016	19:00 às 00:59	40.658	434	1,067%
20.11.2016	01:00 às 06:59	22.085	429	1,942%
20.11.2016	07:00 às 12:59	26.556	581	2,188%
20.11.2016	13:00 às 18:59	23.433	255	1,088%
20.11.2016	19:00 às 00:59	39.112	390	0,997%
21.11.2016	01:00 às 06:59	24.448	359	1,468%
Total Geral		456.418	5.962	1,306%

Tabela 8: Tabela com os dados da 1ª campanha contabilizados em períodos de 6 horas

Na Tabela 8 estão descritas as quantidades de impressões, interações e a taxa obtida em faixas de 6 horas. Ela será utilizada nas seções 4.3.1. e 4.3.2.

3.4.2. Segunda campanha publicitária: *Twitter* segmentando o público de *MMA*.

A plataforma publicitária do *Twitter* permite várias formas de segmentar um público. Esta campanha foi realizada para identificar se a segmentação pode aumentar a quantidade de interações dos usuários com a publicidade, considerando que os usuários escolhidos têm afinidade com o assunto.

Método

Público e Procedimento: a plataforma publicitária do *Twitter* foi programada para enviar a campanha para todos os brasileiros de ambos os sexos que têm afinidade com o esporte MMA, e/ou utilizaram o termo “MMA Brasil” em suas mensagens, durante o período do experimento. A escolha dos usuários foi realizada somente pelo *Twitter*.

Conforme os usuários escolhidos utilizavam a mídia social, eles recebiam o anúncio da Figura 5 junto com outras mensagens de seus interesses. Cada interação realizada pelo usuário com o anúncio foi contabilizada pela plataforma publicitária.

O mesmo usuário pode ter recebido a mensagem mais de uma vez.

Resultados

A mídia social enviou 199.620 mensagens durante o período proposto e elas tiveram 3.731 interações dos usuários, acarretando uma taxa final de interações de 1,869%.

Período		Twitter		
Data	Período	Impressões	Interações	Taxa TWT
18.11.2016	07:00 às 12:59	32.533	657	2,019%
18.11.2016	13:00 às 18:59	28.187	518	1,838%
18.11.2016	19:00 às 00:59	26.372	428	1,623%
19.11.2016	01:00 às 06:59	9.057	203	2,241%
19.11.2016	07:00 às 12:59	16.960	306	1,804%
19.11.2016	13:00 às 18:59	10.810	162	1,499%
19.11.2016	19:00 às 00:59	7.933	170	2,143%
20.11.2016	01:00 às 06:59	7.149	208	2,909%
20.11.2016	07:00 às 12:59	15.190	381	2,508%
20.11.2016	13:00 às 18:59	15.847	233	1,470%
20.11.2016	19:00 às 00:59	21.919	311	1,419%
21.11.2016	01:00 às 06:59	7.663	154	2,010%
Total Geral		199.620	3.731	1,869%

Tabela 9: Tabela com os dados da 2ª campanha contabilizados em períodos de 6 horas

Na Tabela 9 estão descritas as quantidades de impressões, interações e a taxa obtida em faixas de 6 horas. Ela será utilizada nas seções 4.3.1. e 4.3.3.

3.4.3. Terceira campanha publicitária: 40.000 usuários selecionados pelo método proposto *DNA*.

Esta campanha foi criada para enviar a publicidade para os 40.000 usuários selecionados pelo método proposto *DNA* neste estudo.

Assim como na segunda campanha, a expectativa é o aumento de interações com a publicidade pelo fato de os 40.000 usuários terem alta afinidade com o assunto MMA.

Método

Público e Procedimento: a plataforma publicitária do *Twitter* foi programada para enviar a campanha para os 40.000 usuários durante o período do experimento.

Conforme os usuários escolhidos utilizavam a mídia social, eles recebiam o anúncio da Figura 5 junto com outras mensagens de seus interesses. Cada interação realizada pelo usuário com o anúncio foi contabilizada pela plataforma publicitária.

O mesmo usuário pode ter recebido a mensagem mais de uma vez.

Resultados

A mídia social enviou 18.701 mensagens durante o período proposto e elas tiveram 461 interações dos usuários, gerando uma taxa final de interações de 2,465%.

Período		Método DNA		
Data	Período	Impressões	Interações	Taxa DNA
18.11.2016	07:00 às 12:59	1.602	28	1,748%
18.11.2016	13:00 às 18:59	2.276	47	2,065%
18.11.2016	19:00 às 00:59	1.952	55	2,818%
19.11.2016	01:00 às 06:59	349	11	3,152%
19.11.2016	07:00 às 12:59	1.278	19	1,487%
19.11.2016	13:00 às 18:59	2.115	54	2,553%
19.11.2016	19:00 às 00:59	2.288	63	2,753%
20.11.2016	01:00 às 06:59	900	38	4,222%
20.11.2016	07:00 às 12:59	809	31	3,832%
20.11.2016	13:00 às 18:59	2.428	52	2,142%
20.11.2016	19:00 às 00:59	2.285	50	2,188%
21.11.2016	01:00 às 06:59	419	13	3,103%
Total Geral		18.701	461	2,465%

Tabela 10: Tabela com os dados da 3ª campanha contabilizados em períodos de 6 horas

Na Tabela 10 estão descritas as quantidades de impressões, interações e a taxa obtida em faixas de 6 horas. Ela será utilizada nas seções 4.3.2. e 4.3.3.

4. ANÁLISE DOS RESULTADOS

Os resultados dos experimentos estudados foram analisados utilizando o teste de proporção entre duas amostras onde:

$$Z_{\text{observado}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_c \cdot \hat{q}_c \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

onde $\hat{p}_1$ e $\hat{p}_2$: proporções nas duas amostras,

n_1 e n_2 : tamanho das duas amostras e

$$\hat{p}_c = \frac{X_1 + X_2}{n_1 + n_2}, \quad X_1 = \hat{p}_1 \cdot n_1, \quad X_2 = \hat{p}_2 \cdot n_2$$

$$\hat{q}_c = 1 - \hat{p}_c$$

As hipóteses do teste de proporção para realizar as comparações entre os métodos de segmentação com o grupo de controle e entre as faixas de horários no mesmo método são:

$$H_0: p_1 = p_2 \quad H_1: p_1 > p_2$$

Na comparação entre os métodos com o grupo de controle:

p_1 = amostra do método proposto *DNA* ou do *Twitter*.

p_2 = amostra do grupo de controle.

Na comparação entre as faixas de horário:

p_1 = amostra do período NOITE de 19 de novembro.

p_2 = amostra dos outros períodos.

Para a hipótese H_0 ser verdadeira com grau de significância de 0,05, tem-se:

$$Z_{\text{observado}} < +1,64$$

No caso de comparar o método de segmentação proposto com a segmentação realizada pelo *Twitter* são:

$$H_0: p_1 = p_2 \quad H_1: p_1 \neq p_2$$

onde:

p_1 = amostra do método proposto *DNA*.

p_2 = amostra do *Twitter*.

Para a hipótese H_0 ser verdadeira com grau de significância de 0,05, tem-se:

$$-1,96 < Z_{\text{observado}} < +1,96$$

Os dados das Tabelas 8, 9 e 10 (seções 3.4.1., 3.4.2. e 3.4.3, respectivamente) estão agrupados em períodos de 6 horas para a realização dos estudos comparativos entre os métodos.

Porém, os dados foram coletados a cada hora e podem ser reagrupados para realizar outros estudos. Um dos estudos pertinentes é a comparação entre as faixas de horários e dias em um mesmo método, lembrando o fato de ter acontecido um evento ao vivo de *MMA* na cidade de São Paulo, já citado na seção 3.4.

4.1. Comparação entre horários com e sem evento ao vivo

Foram criados dois grupos para o estudo entre as faixas de horários com e sem evento. O primeiro na faixa entre as 09:00 e 14:59, chamado de DIA; e o segundo entre as 21:00 e 02:59, chamado de NOITE.

O evento ocorreu na noite de 19 de novembro de 2016. Porém, em todos os períodos noturnos existem programas e reprises de lutas passando em um canal de televisão a cabo (Quadro 3).

Programação do canal TV especializado em lutas		
Data	DIA	NOITE
18.11.2016	Programa Jornalístico	Programa de Jogos de Lutas
19.11.2016	Programa Jornalístico	MMA Ibirapuera SP
20.11.2016	Programa Jornalístico	Reprise MMA Ibirapuera SP

Quadro 3: Programação do canal de TV nos períodos estudados.

O *Twitter* se auto intitula uma “segunda tela” em sua página institucional na internet, considerando “segunda tela” quando os usuários utilizam esta plataforma para trocarem mensagens durante os eventos acontecendo “ao vivo” em outras mídias como: TV, rádio, etc.

A expectativa nesta análise é verificar se durante o evento existiu maior número de interações dos usuários nos métodos segmentados, indicando a possibilidade de o evento ter atraído um maior número de pessoas engajadas com *MMA* para a mídia social *Twitter*, podendo assim confirmar a informação fornecida pela empresa.

4.1.1. Resultado utilizando o método *DNA* com e sem evento ao vivo

Neste primeiro estudo sobre a possibilidade de existir um maior número de usuários interagindo na mídia social durante o evento ao vivo, observa-se a comparação da faixa NOITE de 19 de novembro (momento do evento ao vivo) com as outras faixas de horários do período de 72 horas pesquisado.

Data	Período	DNA			19/11/2016 - NOITE		
		Impressões	Interações	P _{DNA}	Impressões	Interações	P _{DNA}
					2.107	69	3,275%
18/11/16	DIA	2.247	41	1,825%	$Z_{\text{observado}} = 3,05$		rejeita H_0
	NOITE	1.518	46	3,030%	$Z_{\text{observado}} = 0,41$		H_0 verdadeira
19/11/16	DIA	1.821	30	1,647%	$Z_{\text{observado}} = 3,24$		rejeita H_0
	NOITE	2.107	69	3,275%			
20/11/16	DIA	1.428	33	2,311%	$Z_{\text{observado}} = 1,68$		rejeita H_0
	NOITE	1.712	45	2,629%	$Z_{\text{observado}} = 1,17$		H_0 verdadeira

Tabela 11: Teste de proporção do método *DNA* entre a faixa do evento ao vivo e as outras faixas.

Na Tabela 11 observa-se o teste de proporção entre as faixas sem evento ao vivo com a faixa do momento do evento no Ibirapuera em São Paulo.

Considera-se a seguinte hipótese para investigar a maior interação dos usuários durante o evento:

$$H_0: p_{\text{evento}} = p_2 \quad H_1: p_{\text{evento}} > p_2$$

onde,

p_{evento} = amostra da faixa NOITE de 19 de novembro de 2016.

p_2 = amostra das outras faixas.

Para a hipótese ser verdadeira com grau de significância de 0,05, temos:

$$Z_{\text{observado}} < +1,64$$

A rejeição da hipótese ocorre claramente na comparação das faixas DIA dos dias 18 e 19 de novembro e é muito fraca na faixa DIA do dia 20.

A faixa NOITE do dia do evento não rejeita a hipótese quando comparada com as outras faixas NOITE. Novos estudos podem ser realizados para verificar o quanto a programação noturna do canal especializado pode estar gerando fluxo de usuários durante a noite, causando a confirmação da hipótese.

Da mesma forma, o período DIA de 20 de novembro pode ter recebido um maior número de usuários como reflexo do dia anterior, onde usuários poderiam estar ainda buscando informações sobre o evento nas mídias sociais e especializada, carecendo de mais estudos sobre este caso.

4.1.2. Resultado utilizando o método *Twitter* com e sem evento ao vivo

Considerando a análise realizada na seção 4.1.1, pode-se esperar um resultado parecido, onde as faixas NOITE tenham proporções consideradas iguais e as outras faixas não.

					19/11/2016 - NOITE		
Data	Período	Twitter			Impressões	Interações	PTWT
		Impressões	Interações	PTWT			
					6.151	151	2,455%
18/11/16	DIA	30.389	624	2,053%	$Z_{\text{observado}} = 1,99$	rejeita H_0	
	NOITE	18.586	343	1,845%	$Z_{\text{observado}} = 2,96$	rejeita H_0	
19/11/16	DIA	15.604	233	1,493%	$Z_{\text{observado}} = 4,85$	rejeita H_0	
	NOITE	6.151	151	2,455%			
20/11/16	DIA	19.627	442	2,252%	$Z_{\text{observado}} = 0,93$	H_0 verdadeira	
	NOITE	22.137	356	1,608%	$Z_{\text{observado}} = 4,43$	rejeita H_0	

Tabela 12: Teste de proporção do método *Twitter* entre a faixa do evento ao vivo e as outras faixas.

Da mesma forma, considera-se a seguinte hipótese para confirmar a maior interação dos usuários durante o evento:

$$H_0: p_{\text{evento}} = p_2 \quad H_1: p_{\text{evento}} > p_2$$

onde,

p_{evento} = amostra da faixa NOITE de 19 de novembro de 2016.

p_2 = amostra das outras faixas.

Para a hipótese ser verdadeira com grau de significância de 0,05, tem-se:

$$Z_{\text{observado}} < +1,64$$

Diferentemente da análise da seção 4.1.1., a Tabela 12 mostra uma realidade diferente, onde o evento ao vivo teve uma maior interação em todos os períodos, com exceção da faixa DIA de 20 de novembro, onde a hipótese de a faixa ser igual à faixa do evento ao vivo se confirmou.

Como já explicado na seção 4.1.1., esta confirmação de H_0 da faixa DIA pode ser reflexo da luta do dia anterior, mas o fato de ter rejeitada a hipótese nas outras faixas mostra uma dependência do evento ao vivo maior pelo método do *Twitter* em comparação com o método *DNA*, pois ao confrontar os resultados encontrados nos dois métodos, as interações nas faixas NOITE não têm as hipóteses rejeitadas no método *DNA*, apontando maior homogeneidade entre o público escolhido pelo método *DNA*. Todavia, no método *Twitter* as faixas NOITE foram rejeitadas.

					19/11/2016 - NOITE		
Data	Período	CONTROLE			Impressões	Interações	P _{controle}
		Impressões	Interações	P _{controle}	36.369	450	1,237%
18/11/16	DIA	37.854	392	1,036%	Z _{observado} = 2,59		rejeita H ₀
	NOITE	37.208	469	1,260%	Z _{observado} = -0,28		H ₀ verdadeira
19/11/16	DIA	57.973	774	1,335%	Z _{observado} = -1,29		H ₀ verdadeira
	NOITE	36.369	450	1,237%			
20/11/16	DIA	29.823	623	2,089%	Z _{observado} = -8,63		H ₀ verdadeira
	NOITE	29.526	292	0,989%	Z _{observado} = 3,00		rejeita H ₀

Tabela 13: Teste de proporção do grupo de Controle entre a faixa do evento ao vivo e as outras faixas.

Como forma de mostrar o comportamento diferenciado do público selecionado pelos métodos de segmentação, o mesmo estudo foi realizado no grupo de controle e seus resultados estão na Tabela 13. Nitidamente os comportamentos das interações não podem ser explicados com base no evento ao vivo. A faixa DIA do dia 18 de novembro seguiu o mesmo padrão dos outros 2 testes; porém, as outras duas tiveram a hipótese confirmada e ainda com $Z_{\text{observado}}$ negativo, indicando uma reação completamente contrária à esperada. Reação repetida na faixa NOITE do dia 18 de novembro. Diferentemente, na faixa NOITE de 20 de novembro a hipótese foi rejeitada, reforçando a falta de padrão quando comparado com os resultados obtidos nos métodos segmentados.

4.2. Resultado comparativo do período total pesquisado

Os resultados totais mostram o aumento percentual das interações dos usuários com as publicidades quando estes são escolhidos por algum método de segmentação.

Na Tabela 14 seguem os testes de proporções entre o grupo de controle e os métodos de segmentação para certificar se existem variações significativas entre eles.

Teste de Proporção Duas Amostras Nível de significância de 0,05			DNA			Twitter		
			Impressões	Interações	Taxa	Impressões	Interações	Taxa
			18.701	461	2,465%	199.620	3.731	1,869%
Twitter			$Z_{\text{observado}} \text{ DNA x Twitter}$ 5,679656 $Z_{\text{observado}} > 1,96$. Rejeita HT_0			Teste de duas amostras para proporção: $HC_0: P_1 = P_2$ $HT_0: P_1 = P_2$ $HC_1: P_1 > P_2$ $HT_1: P_1 \neq P_2$		
Impressões	Interações	Taxa	$Z_{\text{observado}} \text{ DNA x CONTROLE}$ 13,450167 $Z_{\text{observado}} > 1,64$. Rejeita HC_0			$Z_{\text{observado}} \text{ Twitter x CONTROLE}$ 17,383406 $Z_{\text{observado}} > 1,64$. Rejeita HC_0		
199.620	3.731	1,869%						
CONTROLE								
Impressões	Interações	Taxa						
456.418	5.962	1,306%						

Tabela 14: Teste de proporção do período total pesquisado.

4.2.1. *Twitter* x Grupo de Controle

Para considerar a segmentação do *Twitter* melhor em comparação ao grupo de controle, utiliza-se o teste de proporção com a seguinte hipótese:

$$H_0: p_{twitter} = p_{controle} \quad H_1: p_{twitter} > p_{controle}$$

Resultado:

$$Z_{observado} = 17,3834$$

A hipótese é rejeitada porque $Z_{observado} > 1,64$ (significância de 0,05).

Este resultado confirma a existência da diferença entre as interações com a publicidade no grupo segmentado pelo *Twitter* e o grupo de controle. Sendo assim, a segmentação realizada pelo *Twitter* consegue selecionar um grupo de usuários mais engajados com o assunto, quando comparada aos usuários do grupo de controle.

4.2.2. Método *DNA* x Grupo de Controle

Para considerar a segmentação do método *DNA* melhor em comparação ao grupo de controle, utiliza-se o teste de proporção com a seguinte hipótese:

$$H_0: p_{DNA} = p_{controle} \quad H_1: p_{DNA} > p_{controle}$$

Resultado:

$$Z_{observado} = 13,4502$$

A hipótese é rejeitada porque $Z_{observado} > 1,64$ (significância de 0,05).

Este resultado confirma a existência da diferença entre as interações com a publicidade no grupo segmentado pelo método *DNA* e o grupo de controle. Sendo assim, a segmentação realizada pelo método *DNA* consegue selecionar um grupo de usuários mais engajados com o assunto, quando comparada aos usuários do grupo de controle.

4.2.3. Método *DNA* x *Twitter*

Para verificar se a segmentação do método *DNA* é diferente em comparação com a segmentação realizada pelo *Twitter*, utiliza-se o teste de proporção com a seguinte hipótese:

$$H_0: p_{DNA} = p_{twitter} \quad H_1: p_{DNA} \neq p_{twitter}$$

Resultado:

$$Z_{\text{observado}} = 5,6796$$

A hipótese é rejeitada porque $Z_{\text{observado}} > 1,96$ (significância de 0,05). Este resultado confirma a existência de diferença entre as interações com a publicidade no grupo segmentado pelo método *DNA* e o grupo segmentado pelo método do *Twitter*.

Este estudo realizado utilizando as 72 horas de pesquisa pode confirmar ou não as hipóteses propostas; porém, outras análises podem ser realizadas sobre os mesmos dados procurando confirmar os resultados em faixas de horários menores.

4.3. Interações a cada 6 horas

Os resultados analisados na seção anterior foram obtidos pelo total de interações realizadas em cada um dos métodos no período total da pesquisa (72 horas). Pode-se realizar a mesma análise em intervalos menores para verificar se existe uma variação de interações significativa dependendo do horário. Na Figura 6 observa-se a taxa de interações a cada seis horas nos três grupos estudados. Com exceção do período 5, os comportamentos das linhas são similares. Esta similaridade entre as amostras utilizadas mostra uma coerência no comportamento das pessoas atingidas pelas propagandas e ilustra o aumento de interações proporcionado pela escolha dos usuários engajados com o assunto.

Figura 6: Percentual de interações a cada seis horas.

As verificações em períodos menores podem confirmar ou rejeitar as hipóteses da seção 4.2., onde todas as hipóteses foram rejeitadas.

As tabelas 8 (seção 3.4.1.), 9 (seção 3.4.2.) e 10 (seção 3.4.3.) contêm os dados da pesquisa acumulados a cada seis horas e serão usadas nas próximas seções.

4.3.1. *Twitter* x Grupo de Controle

As hipóteses enunciadas na seção 4.2.1. para analisar se a segmentação do *Twitter* é melhor em comparação ao grupo de controle, foram as seguintes:

$$H_0: p_{twitter} = p_{controle} \quad H_1: p_{twitter} > p_{controle}$$

Aplicando o teste de proporção entre a Tabela 8 da seção 3.4.1. (grupo de controle) e a Tabela 9 da seção 3.4.2. (*Twitter*), obtêm-se:

Período		<i>Twitter</i> x CONTROLE	
Data	Período	$Z_{\text{observado}}$	$H_0: p_{twitter} = p_{controle}$
18/11/16	07:00 às 12:59	7,706709	rejeita H_0
18/11/16	13:00 às 18:59	10,566462	rejeita H_0
18/11/16	19:00 às 00:59	5,036950	rejeita H_0
19/11/16	01:00 às 06:59	3,495474	rejeita H_0
19/11/16	07:00 às 12:59	4,179242	rejeita H_0
19/11/16	13:00 às 18:59	4,441615	rejeita H_0
19/11/16	19:00 às 00:59	7,908652	rejeita H_0
20/11/16	01:00 às 06:59	4,867582	rejeita H_0
20/11/16	07:00 às 12:59	2,099071	rejeita H_0
20/11/16	13:00 às 18:59	3,354056	rejeita H_0
20/11/16	19:00 às 00:59	4,690760	rejeita H_0
21/11/16	01:00 às 06:59	3,297195	rejeita H_0

Tabela 15: Teste de proporção entre a segmentação do *Twitter* e o grupo de Controle.

As estatísticas dos testes realizados e descritos na Tabela 15 confirmam a diferença das interações com a publicidade entre o grupo segmentado pelo *Twitter* e o grupo de controle. Comparando os períodos de seis em seis horas, pode-se observar a rejeição da hipótese em todas as faixas, confirmando maior interação no grupo segmentado.

Este resultado corrobora a conclusão da seção 4.2.1.

4.3.2. Método *DNA* x Grupo de Controle

As hipóteses enunciadas na seção 4.2.2. para analisar se a segmentação do *DNA* é melhor em comparação ao grupo de controle, foram as seguintes:

$$H_0: p_{DNA} = p_{controle}$$

$$H_1: p_{DNA} > p_{controle}$$

Fazendo a mesma análise entre a Tabela 8 da seção 3.4.1. (grupo de controle) e a Tabela 10 da seção 3.4.3. (Método *DNA*), obtêm-se:

Período		<i>DNA</i> x CONTROLE	
Data	Período	$Z_{\text{observado}}$	$H_0: p_{DNA} = p_{\text{controle}}$
18/11/16	07:00 às 12:59	1,209335	H_0 verdadeira
18/11/16	13:00 às 18:59	5,431996	rejeita H_0
18/11/16	19:00 às 00:59	6,466943	rejeita H_0
19/11/16	01:00 às 06:59	2,136328	rejeita H_0
19/11/16	07:00 às 12:59	0,369362	H_0 verdadeira
19/11/16	13:00 às 18:59	6,774209	rejeita H_0
19/11/16	19:00 às 00:59	7,337066	rejeita H_0
20/11/16	01:00 às 06:59	4,751730	rejeita H_0
20/11/16	07:00 às 12:59	3,115369	rejeita H_0
20/11/16	13:00 às 18:59	4,562313	rejeita H_0
20/11/16	19:00 às 00:59	5,396612	rejeita H_0
21/11/16	01:00 às 06:59	2,732348	rejeita H_0

Tabela 16: Teste de proporção entre o método *DNA* e o grupo de Controle.

Na maioria dos períodos o grupo segmentado pelo método *DNA* mostrou ter maior interação com a publicidade em comparação com o grupo de controle. Porém, a hipótese foi confirmada em dois períodos dos doze estudados (Tabela 16). Devido ao fato de o *Twitter* não informar quais usuários receberam a publicidade da lista de 40.000 usuários selecionada pelo método *DNA*, não é possível chegar a uma conclusão plausível para estes dois eventos. Mas, também se pode observar maior quantidade de interações no grupo segmentado pelo método *DNA* em comparação com o grupo de controle.

E assim, este resultado corrobora a conclusão da seção 4.2.2.

4.3.3. Método *DNA* x *Twitter*

Finalizando esta análise, as hipóteses enunciadas na seção 4.2.3. para analisar se a segmentação do *DNA* é diferente em comparação com a segmentação do *Twitter*, foram as seguintes:

$$H_0: p_{DNA} = p_{twitter}$$

$$H_1: p_{DNA} \neq p_{twitter}$$

Segue o teste entre a Tabela 10 da seção 3.4.3. (Método *DNA*) e a Tabela 9 da seção 3.4.2. (*Twitter*), obtêm-se:

Período		<i>DNA x Twitter</i>	
Data	Período	$Z_{\text{observado}}$	$H_0: p_{DNA} = p_{twitter}$
18/11/16	07:00 às 12:59	-0,757000	H_0 verdadeira
18/11/16	13:00 às 18:59	0,773126	H_0 verdadeira
18/11/16	19:00 às 00:59	3,933941	rejeita H_0
19/11/16	01:00 às 06:59	1,119377	H_0 verdadeira
19/11/16	07:00 às 12:59	-0,827465	H_0 verdadeira
19/11/16	13:00 às 18:59	3,460030	rejeita H_0
19/11/16	19:00 às 00:59	1,723840	H_0 verdadeira
20/11/16	01:00 às 06:59	2,156205	rejeita H_0
20/11/16	07:00 às 12:59	2,316044	rejeita H_0
20/11/16	13:00 às 18:59	2,486285	rejeita H_0
20/11/16	19:00 às 00:59	2,887167	rejeita H_0
21/11/16	01:00 às 06:59	1,531404	H_0 verdadeira

Tabela 17: Teste de proporção entre o método *DNA* e a segmentação do *Twitter*.

Nesta comparação pode-se observar a proximidade entre os métodos (Tabela 17). Nos testes, a hipótese foi confirmada em seis dos doze períodos analisados, ou seja, em 50% dos casos. Como não se tem a informação sobre quais usuários receberam a publicidade e como existe a possibilidade de um mesmo usuário ter recebido mais de uma vez o mesmo anúncio, não se pode afirmar se um método é melhor em comparação com o outro.

As análises realizadas mostram o aumento das interações dos usuários com as publicidades apresentadas nos dois métodos de segmentação, método *DNA* e *Twitter*, em comparação com o grupo de controle. Porém, a comparação direta entre os dois métodos de segmentação não se mostra conclusiva. Ao fazer o teste de proporção utilizando os totais, pode-se verificar a existência da diferença entre os métodos; porém, ao analisar períodos menores, tem-se a confirmação das diferenças em 50% dos períodos analisados.

Neste caso, o estudo dos intervalos a cada seis horas não confirma totalmente o resultado obtido no total da pesquisa.

Sendo assim, pode-se rejeitar a hipótese HC_0 definida na seção 1.2., onde os usuários escolhidos pelo método proposto *DNA* geraram um maior número de interações quando comparados com os usuários do grupo de controle.

Porém, não se pode rejeitar a hipótese HT_0 definida na seção 1.2. Sendo assim, ambos os métodos podem ser considerados similares.

De acordo com os resultados obtidos, pode-se confirmar o objetivo principal de desenvolver um método para segmentar os usuários da mídia social do *Twitter*, confirmando com a tribo eletrônica escolhida (*MMA*) o aumento de interações dos usuários segmentados em comparação os usuários do grupo de controle.

Este método abre novas possibilidades de estudos nas mídias sociais, pois identificar as tribos de preferências dos usuários pode auxiliar na identificação de sua personalidade.

CONCLUSÕES

Este estudo tem respaldo em teorias consagradas da academia e apresentou uma forma de analisar as pessoas no mundo digital. O comportamento dos consumidores está sendo amplificado no mundo virtual e rompendo todas as fronteiras tradicionais. As tribos eletrônicas formadas nas mídias sociais não possuem bordas definidas por países, nem são limitadas por idioma, religião ou raça. Os participantes das mídias sociais unem-se em volta de vários assuntos de interesse, podendo ser estes assuntos as únicas afinidades entre eles. Os usuários vestem vários trajes, mostram suas crenças e encontram seus pares em movimentos caóticos e ao mesmo tempo homogêneos relativamente aos seus interesses.

A identificação de padrões comportamentais utilizando métodos como a filtragem colaborativa ajuda a classificar usuários desconhecidos com pouco uso de processamento de dados. A proximidade entre os usuários auxilia a entender a tribo e identificar assuntos e influenciadores em comum. Os laços criados formam grupos capazes de impulsionar ou prejudicar marcas e produtos.

Este estudo criou uma metodologia para identificar as tribos e testou uma tribo isoladamente (usuários simpatizantes com *MMA*); porém, novas pesquisas podem testar o comportamento de usuários com afinidades múltiplas (pertencentes a mais de uma tribo) e suas implicações na definição da personalidade do usuário.

Para as empresas o trabalho realizado mostra uma forma prática para a identificação do público-alvo com o objetivo de aumentar o engajamento entre as pessoas e os anúncios realizados nas mídias sociais. Os estudos acadêmicos mostram a baixa aderência do público com os anúncios nas mídias sociais; entretanto, este trabalho confirmou o aumento de engajamento quando o público escolhido tem mais envolvimento com o principal assunto vinculado.

Outros estudos podem ampliar o conhecimento sobre a capacidade de aumentar o engajamento, como envolver influenciadores e celebridades na divulgação e verificar se a participação destes atores pode aumentar o engajamento. Kowalczyk et al. (2016) apontam como a não aderência de uma celebridade pode não ajudar uma marca ou produto; porém, mesmo uma celebridade engajada pode não aumentar as interações dos usuários.

Outro ponto observado é a necessidade de entender melhor se a vinculação de anúncios em horários diferentes pode realmente mudar a quantidade de interações. Este fato foi observado com aumento de interações durante o evento ao vivo e em momentos de cruzamento de horários com programas em outros meios de comunicação.

Os conceitos apresentados e a pesquisa realizada fornecem uma diretriz sobre como definir as tribos e abre caminho para encontrar formas de identificar as múltiplas preferências dos usuários, servindo de base para estudos mais profundos sobre a composição e formação das personalidades de seus integrantes no mundo virtual, buscando formas de engajar e criar laços entre os usuários e a mensagem propagada.

REFERÊNCIAS BIBLIOGRÁFICAS

- Adams, T.L. and Smith, S.A. (2008). *Electronic tribes: the virtual worlds of geeks, gamers, shamans, and scammers*. Austin, TX: University of Texas Press.
- Anderson, M., Ball, M., Boley, H., Greene, S., Howse, N., McGrath, S., & Lemire, D. (2003). Racofi: A rule-applying collaborative filtering system.
- Barabási, Albert-László and Frangos, Jennifer. (2014). *Linked*. New York, NY: Basic Books.
- Barnes, N. G., Lescault, A. M., & Andonian, J. (2012). Social media surge by the 2012 Fortune 500: Increase use of blogs, Facebook, Twitter and more. *Charlton College of Business Center for Marketing Research*.
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3), 497-529
- Belk, R. W. (1988). Possessions and the extended self. *Journal of consumer research*, 15(2), 139-168.
- Belk, R. W. (2013). Extended self in a digital world. *Journal of Consumer Research*, 40(3), 477-500.
- Booth, N., & Matic, J. A. (2011). Mapping and leveraging influencers in social media to shape corporate brand perceptions. *Corporate Communications: An International Journal*, 16(3), 184-191.
- Breese, J. S., Heckerman, D., & Kadie, C. (1998, July). Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence* (pp. 43-52). Morgan Kaufmann Publishers Inc..

- Bruhn, M., Schoenmueller, V., & Schäfer, D. B. (2012). Are Social Media replacing traditional media in terms of brand equity creation? *Management Research Review*, 35(9), 770-790.
- Cicero, M. T. (1951). *Pro Plancio*. University Tutorial Press.
- Cormack, R. M. (1971). A review of classification. *Journal of the Royal Statistical Society. Series A (General)*, 321-367.
- Costello, L., McDermott, M. L., & Wallace, R. (2017). Netnography: Range of Practices, Misperceptions, and Missed Opportunities. *International Journal of Qualitative Methods*, 16(1), 1609406917700647.
- Cova, B. (1997). Community and consumption: Towards a definition of the “linking value” of product or services. *European journal of marketing*, 31(3/4), 297-316.
- Cova, B., & Cova, V. (2002). Tribal marketing: The tribalisation of society and its impact on the conduct of marketing. *European journal of marketing*, 36(5/6), 595-620.
- Damien Wilson, Sarah Quinton, (2012),"Let's talk about wine: does *Twitter* have value?", *International Journal of Wine Business Research*, Vol. 24 Iss 4 pp.271-286
- Daniel Lemire Harold Boley Sean McGrath Marcel Ball, (2005),"Collaborative filtering and inference rules for context-aware learning object recommendation", *Interactive Technology and Smart Education*, Vol. 2 Iss 3 pp. 179-188
- Danzig, A. B. (1985). Ethnography and Qualitative Design in Educational Research: Judith P. Goetz and Margaret D. LeCompte. *Anthropology & Education Quarterly*, 16(1), 76-77.
- Ducheneaut, N., Wen, M. H., Yee, N., & Wadley, G. (2009, April). Body and mind: a study of avatar personalization in three virtual worlds. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1151-1160). ACM.

- Einstein, Albert. (1931). *Cosmic religion: With other opinions and aphorisms*. New York: Covici-Friede.
- Elliott, R., & Jankel-Elliott, N. (2003). Using ethnography in strategic consumer research. *Qualitative market research: An international journal*, 6(4), 215-223.
- Escalas, J. E., & Bettman, J. R. (2015). Managing Brand Meaning through Celebrity Endorsement. In *Brand Meaning Management* (pp. 29-52). Emerald Group Publishing Limited.
- Evans, D. (2010). *Social media marketing: the next generation of business engagement*. John Wiley & Sons.
- Everitt, B. S. (1979). Unresolved problems in cluster analysis. *Biometrics*, 169-181.
- Fávero, L. P., Belfiore, P., Silva, F. D., & Chan, B. L. (2009). Análise de dados. *Modelagem multivariada para tomada de decisões*. Rio de Janeiro: Campus.
- Goldberg, D., Nichols, D., Oki, B., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the Association of Computing Machinery*, 35(12), 61-70
- Goulding, C., Shankar, A. and Canniford, R. (2013), “Learning to be tribal: facilitating the formation of consumer tribes”, *European Journal of Marketing*, Vol. 47 Nos 5/6, pp. 813-832.
- Gruzd, A., Wellman, B. and Takhteyev, Y. (2011), “Imagining *Twitter* as an imagined community”, *American Behavioural Scientist*, Vol. 55 No. 10, pp. 1294-318.
- Gupta, P., Goel, A., Lin, J., Sharma, A., Wang, D., & Zadeh, R. (2013, May). Wtf: The who to follow service at *Twitter*. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 505-514). ACM.
- Hanna, R., Rohm, A., & Crittenden, V. L. (2011). We’re all connected: The power of the social media ecosystem. *Business horizons*, 54(3), 265-273.

- Harari, Y. N. (2014). *Sapiens: A Brief History of Humankind*. Random House.
- Herlocker, J. L., Konstan, J. A., Borchers, A., & Riedl, J. (1999, August). An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 230-237). ACM.
- Kamel Boulos, Maged N., and Steve Wheeler. "The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education1." *Health Information & Libraries Journal* 24.1 (2007): 2-23.
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, 53(1), 59-68.
- Katz, J. E., & Rice, R. E. (2002). *Social consequences of Internet use: Access, involvement, and interaction* (pp. 1-463). Cambridge, MA: MIT press.
- Kowalczyk, C. M., Kowalczyk, C. M., Pounders, K. R., & Pounders, K. R. (2016). Transforming celebrities through social media: the role of authenticity and emotional attachment. *Journal of Product & Brand Management*, 25(4), 345-356.
- Kozinets, R. V. (1999). E-tribalized marketing?: The strategic implications of virtual communities of consumption. *European Management Journal*, 17(3), 252-264.
- Kozinets, R. V. (1998). On netnography: Initial reflections on consumer research investigations of cyberculture. *NA-Advances in Consumer Research Volume 25*.
- Kozinets, R. V. (2002). The field behind the screen: using netnography for marketing research in online communities. *Journal of marketing research*, 39(1), 61-72.
- Langner, S., Hennigs, N., & Wiedmann, K. P. (2013). Social persuasion: targeting social identities through social influencers. *Journal of Consumer Marketing*, 30(1), 31-49.

- Lemire, D., Boley, H., McGrath, S., & Ball, M. (2005). Collaborative filtering and inference rules for context-aware learning object recommendation. *Interactive Technology and Smart Education*, 2(3), 179-188.
- McNamee, Peter e Celona, John. (2007). *Decision analysis for the professional*. 4ª edição, SmartOrg.
- Ng, A. Y., Jordan, M. I., & Weiss, Y. (2001, December). On spectral clustering: Analysis and an algorithm. In NIPS (Vol. 14, No. 2, pp. 849-856).
- Norm O'Reilly Ida E. Berger Tony Hernandez Milena M. Parent Benoit Segui. (2012). Understanding adolescent sport participation through online social media. *Sport, Business and Management: An International Journal*, Vol. 2 Iss 1 pp. 69-81.
- Pahl-Wostl, C. (2015). The Role of Institutions, Actors and Social Networks in Societal Change. In *Water Governance in the Face of Global Change* (pp. 51-83). Springer International Publishing.
- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994, October). GroupLens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work* (pp. 175-186). ACM.
- Sammut, C., Webb, G. (2010). *Encyclopedia of Machine Learning*, Springer-Verlag Berlin Heidelberg.
- Shintaro Okazaki, Charles R. Taylor, (2013), "Social media and international advertising: theoretical challenges and future directions", *International Marketing Review*, Vol. 30 Iss 1 pp. 56 – 71.
- Sustein, Cass. *Republic.com*. Princeton: Princeton University Press, 2001.
- Wenger, E. (2000). Communities of practice and social learning systems. *Organization*, 7(2), 225-246.
- Willis, P., & Trondman, M. (2000). Manifesto for ethnography. *Ethnography*, 1(1), 5-16.

Yazdanfar, N., & Thomo, A. (2013). Link recommender: Collaborative-Filtering for recommending URLs to *Twitter* users. *Procedia Computer Science*, 19, 412-419.

Zhang, J., & Mao, E. (2016). "From Online Motivations to Ad Clicks and to Behavioral Intentions: An Empirical Study of Consumer Response to Social Media Advertising". *Psychology & Marketing*, 33(3), 155-164.

Zeljka Hadija, Susan B. Barnes, Neil Hair, (2012), "Why we ignore social networking advertising", *Qualitative Market Research: An International Journal*, Vol. 15 Iss 1 pp. 19 – 32

Internet links

<http://www.adweek.com/>

<https://about.twitter.com/pt/company>

<https://support.twitter.com/articles/280973#>

“<http://www.archive.org/details/ciceroproplancioOOcicerich>”

Apêndice I – Dados coletados

Período		DNA			TWITTER			CONTROLE		
Dia	Hora	Impressões	Interações	DNA Taxa	Impressões	Interações	TWT Taxa	Impressões	Interações	CNT Taxa
18.11.16	7	51	1	1,961%	4.028	83	2,061%	27.681	459	1,658%
18.11.16	8	177	3	1,695%	7.950	172	2,164%	27.451	424	1,545%
18.11.16	9	295	6	2,034%	8.375	193	2,304%	14.327	143	0,998%
18.11.16	10	339	4	1,180%	8.651	156	1,803%	6.073	45	0,741%
18.11.16	11	346	3	0,867%	1.909	36	1,886%	2.344	24	1,024%
18.11.16	12	394	11	2,792%	1.620	17	1,049%	2.500	22	0,880%
18.11.16	13	417	10	2,398%	1.404	28	1,994%	2.124	25	1,177%
18.11.16	14	456	7	1,535%	8.430	194	2,301%	10.486	133	1,268%
18.11.16	15	335	7	2,090%	10.324	180	1,744%	1.864	15	0,805%
18.11.16	16	353	7	1,983%	3.459	49	1,417%	4.377	23	0,525%
18.11.16	17	361	9	2,493%	1.782	37	2,076%	9.217	69	0,749%
18.11.16	18	354	7	1,977%	2.788	30	1,076%	13.012	111	0,853%
18.11.16	19	380	12	3,158%	4.631	50	1,080%	9.819	81	0,825%
18.11.16	20	295	4	1,356%	7.046	123	1,746%	3.051	52	1,704%
18.11.16	21	378	10	2,646%	3.608	61	1,691%	5.029	40	0,795%
18.11.16	22	396	11	2,778%	3.723	50	1,343%	9.074	112	1,234%
18.11.16	23	326	10	3,067%	1.244	14	1,125%	2.366	34	1,437%
19.11.16	0	177	8	4,520%	6.120	130	2,124%	8.835	123	1,392%
19.11.16	1	168	4	2,381%	3.783	87	2,300%	8.318	114	1,371%
19.11.16	2	73	3	4,110%	108	1	0,926%	3.586	46	1,283%
19.11.16	3	35	2	5,714%	1.749	46	2,630%	2.709	51	1,883%
19.11.16	4	25	1	4,000%	1.298	27	2,080%	4.679	92	1,966%
19.11.16	5	29	1	3,448%	374	6	1,604%	3.513	81	2,306%
19.11.16	6	19	-	0,000%	1.745	36	2,063%	2.192	33	1,505%
19.11.16	7	47	1	2,128%	3.529	85	2,409%	3.541	51	1,440%
19.11.16	8	103	-	0,000%	830	19	2,289%	4.727	56	1,185%
19.11.16	9	200	6	3,000%	2.316	36	1,554%	10.018	119	1,188%
19.11.16	10	268	6	2,239%	4.840	83	1,715%	23.620	361	1,528%
19.11.16	11	311	3	0,965%	4.047	74	1,829%	11.102	142	1,279%
19.11.16	12	349	3	0,860%	1.398	9	0,644%	4.413	55	1,246%
19.11.16	13	369	4	1,084%	1.592	14	0,879%	3.743	48	1,282%
19.11.16	14	324	8	2,469%	1.411	17	1,205%	5.077	49	0,965%
19.11.16	15	316	19	6,013%	1.850	20	1,081%	8.171	85	1,040%
19.11.16	16	376	9	2,394%	2.682	50	1,864%	7.147	57	0,798%
19.11.16	17	341	12	3,519%	2.128	38	1,786%	4.963	39	0,786%
19.11.16	18	389	2	0,514%	1.147	23	2,005%	8.977	100	1,114%
19.11.16	19	373	6	1,609%	1.680	29	1,726%	8.834	92	1,041%
19.11.16	20	382	13	3,403%	3.388	69	2,037%	5.401	40	0,741%
19.11.16	21	437	11	2,517%	1.059	23	2,172%	5.725	71	1,240%
19.11.16	22	416	12	2,885%	38	2	5,263%	5.265	44	0,836%
19.11.16	23	420	14	3,333%	-	-	0,000%	6.217	66	1,062%
20.11.16	0	260	7	2,692%	1.768	47	2,658%	9.216	121	1,313%
20.11.16	1	295	8	2,712%	1.513	33	2,181%	4.922	67	1,361%
20.11.16	2	279	17	6,093%	1.773	46	2,594%	5.024	81	1,612%
20.11.16	3	191	5	2,618%	964	28	2,905%	3.876	97	2,503%
20.11.16	4	68	5	7,353%	947	26	2,746%	4.854	102	2,101%
20.11.16	5	39	3	7,692%	1.209	51	4,218%	1.867	44	2,357%
20.11.16	6	28	-	0,000%	743	24	3,230%	1.542	38	2,464%
20.11.16	7	56	1	1,786%	52	3	5,769%	340	9	2,647%
20.11.16	8	75	6	8,000%	64	1	1,563%	564	4	0,709%
20.11.16	9	126	7	5,556%	224	6	2,679%	1.405	18	1,281%
20.11.16	10	151	8	5,298%	561	10	1,783%	1.286	17	1,322%
20.11.16	11	-	-	0,000%	-	-	0,000%	-	-	0,000%
20.11.16	12	401	9	2,244%	14.289	361	2,526%	22.961	533	2,321%
20.11.16	13	381	7	1,837%	2.609	48	1,840%	1.631	29	1,778%
20.11.16	14	369	2	0,542%	1.944	17	0,874%	2.540	26	1,024%
20.11.16	15	381	9	2,362%	2.565	30	1,170%	5.583	53	0,949%
20.11.16	16	370	9	2,432%	3.685	41	1,113%	7.159	78	1,090%
20.11.16	17	457	13	2,845%	3.514	79	2,248%	2.726	25	0,917%
20.11.16	18	470	12	2,553%	1.530	18	1,176%	3.794	44	1,160%
20.11.16	19	391	9	2,302%	2.480	29	1,169%	9.697	102	1,052%
20.11.16	20	417	5	1,199%	2.005	26	1,297%	9.412	107	1,137%
20.11.16	21	471	4	0,849%	4.073	64	1,571%	3.588	28	0,780%
20.11.16	22	392	10	2,551%	5.204	70	1,345%	7.337	52	0,709%
20.11.16	23	395	11	2,785%	2.784	39	1,401%	3.114	36	1,156%
21.11.16	0	219	11	5,023%	5.373	83	1,545%	5.964	65	1,090%
21.11.16	1	150	7	4,667%	2.835	66	2,328%	7.252	83	1,145%
21.11.16	2	85	2	2,353%	1.868	34	1,820%	2.271	28	1,233%
21.11.16	3	47	3	6,383%	1.236	19	1,537%	3.290	68	2,067%
21.11.16	4	44	1	2,273%	582	16	2,749%	4.977	96	1,929%
21.11.16	5	45	-	0,000%	645	6	0,930%	3.576	43	1,202%
21.11.16	6	48	-	0,000%	497	13	2,616%	3.082	41	1,330%
Total	DNA			TWITTER			CONTROLE			
	Impressões	Interações	Taxa	Impressões	Interações	Taxa	Impressões	Interações	Taxa	
	18.701	461	2,465%	199.620	3.731	1,869%	456.418	5.962	1,306%	

Apêndice II - Glossário

Símbolos

@ - O sinal @ (arroba) é usado para destacar nomes de usuários. As pessoas usam o @nomedeusuario para mencionar, enviar mensagens ou criar um *link* para o perfil de outros usuários.

@nomedeusuario – É a identificação dos usuários nas mídias sociais. Sempre está imediatamente precedido pelo símbolo @.

A

API – É um conjunto de rotinas e padrões de programação para acesso a um aplicativo de software ou plataforma baseado na *Web*. A sigla *API* refere-se ao termo em inglês "*Application Programming Interface*" ou em português "Interface de Programação de Aplicativos".

Avatar - significa manifestação corporal de um ser superpoderoso e imortal, na religião hindu. No mundo digital, *Avatar* é o corpo virtual criado pelo usuário para projetar sua identidade e ações no mundo virtual (Ducheneaut, Wen, Yee e Wadley, 2009).

E

Engajamento - Todos os cliques nas mensagens, incluindo *Retweets*, respostas, seguidores e cliques em *hashtags* e imagens.

H

Hashtag ou # - É qualquer palavra ou frase imediatamente precedida pelo símbolo #. Quando o usuário clica ou toca em uma *hashtag*, vê outras mensagens contendo a mesma palavra-chave ou tópico.

M

Menção - A inclusão do sinal @ seguido diretamente pelo nome de usuário em uma mensagem é chamada de "menção".

N

Números de seguindo e seguidores - Esses números indicam quantas pessoas o usuário segue e quantas pessoas seguem o usuário.

P

Perfil - O perfil exibe informações escolhidas pelo usuário para compartilhar publicamente.

R

Retweet (s.) - Um *Tweet* existente encaminhado por outro usuário para seus seguidores é conhecido como *Retweet*. Geralmente usados para transmitir notícias ou outras descobertas no *Twitter*.

S

Sala de Chat - *Chat* é um termo da língua inglesa traduzido como “bate-papo” (conversa). Na internet é uma ferramenta (ou fórum) que permite a comunicação por escrito e em tempo real de dois ou mais usuários.

Seguidor (es) - Um seguidor surge quando alguém segue um usuário para receber suas mensagens na *timeline* da Página Inicial dele.

T

Timeline - A *timeline* é um fluxo de mensagens transmitidas em tempo real. A *timeline* é onde o usuário vê todas as mensagens compartilhadas pelas pessoas seguidas.

Tweet (s.) - Um *Tweet* pode conter fotos, vídeos, links e texto com até 140 caracteres.

Tweets Promovidos - Os *Tweets* Promovidos são *Tweets* pagos. Eles aparecem na *timeline* dos usuários e são visivelmente marcados como texto "Promovido".

U

URL, URLs - Um URL (*Uniform Resource Locator*) é um endereço da *Web* que aponta para uma página exclusiva na Internet.

Apêndice III – Exemplo de memória de cálculo

Exemplo de memória de cálculo do teste de proporção de duas amostras do período total pesquisado.

Teste de Proporção				DNA			Twitter		
Duas Amostras				Impressões (n1)	Interações	Taxa (p1)	Impressões (n1)	Interações	Taxa (p1)
Nível de significância de 0,05				18.701	461	2,465%	199.620	3.731	1,869%
Twitter				Teste DNA - Twitter			Fórmulas das células		
Impressões (n2)	Interações	Taxa (p2)		4192,000000	0,000058	0,005961	$Xn = p1 \cdot n1 + p2 \cdot n2$	$R = 1/n1 + 1/n2$	$P = p1 \cdot p2$
199.620	3.731	1,869%		0,019201	0,980799	5,679656	$pc = Xn / (n1+n2)$	$qc = 1 - pc$	$Z = P / \text{raiz}(pc \cdot qc \cdot R)$
CONTROLE				Teste DNA x CONTROLE			Teste Twitter x CONTROLE		
Impressões (n2)	Interações	Taxa (p2)		6423,000000	0,000056	0,011589	9693,000000	0,000007	0,005628
456.418	5.962	1,306%		0,013519	0,986481	13,450167	0,014775	0,985225	17,383406

$$Z_{\text{observado}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_c \cdot \hat{q}_c \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

onde $\hat{p}_1$ e $\hat{p}_2$: proporções nas duas amostras,

n_1 e n_2 : tamanho das duas amostras e

$$\hat{p}_c = \frac{X_1 + X_2}{n_1 + n_2}, \quad X_1 = \hat{p}_1 \cdot n_1, \quad X_2 = \hat{p}_2 \cdot n_2$$

$$\hat{q}_c = 1 - \hat{p}_c$$