

## **Análise de Usuários que Conversam sobre Cerveja no Twitter<sup>1</sup>**

### *Statistical Analysis of Users who Chatting about Beer on Twitter*

Submissão: 28/mar./2014 - Aprovação: 14/abr./2014

#### **Rodrigo Otávio de Araújo Ribeiro**

Doutor e Mestre em Engenharia de Produção pela Universidade Federal Fluminense - UFF. Bacharel em Estatística pela Escola Nacional de Ciências Estatísticas - ENCE/IBGE. Especialista na aplicação de modelos estatísticos em grandes bases de dados. Diretor de Inteligência de Marketing no IBOPE DTM.

**E-mail:** rodrigo.ribeiro@ibopedtm.com

**Endereço profissional:** IBOPE DTM - Rua Voluntários da Pátria - nº 89 - sala 803 - 22270-000 - Botafogo - Rio de Janeiro/RJ – Brasil.

#### **Tarsila Gomes Bello Tavares**

Bacharel em Estatística pela Escola Nacional de Ciências Estatísticas ENCE/IBGE. Especialista na aplicação de modelos estatísticos em grandes bases de dados. Coordenadora de Inteligência de Marketing no IBOPE DTM.

**E-mail:** tarsila.tavares@ibopedtm.com

#### **Daniel de Oliveira Cohen**

Bacharel em Estatística pela Universidade Estadual de Campinas – UNICAMP. Realiza análises estatísticas como regressão, segmentação e análise de redes sociais em dados coletados por meio de pesquisas quantitativas. Estatístico no IBOPE Inteligência.

**E-mail:** daniel.cohen@iboointeligencia.com

---

<sup>1</sup> Este foi um dos trabalhos apresentados no 6º Congresso Brasileiro de Pesquisa - Mercado, Opinião e Mídia da ABEP (realizado em 24 e 25 de março de 2014), vencedor do Prêmio “Alfredo Carmo”, transformado em artigo por seu(s) autor(es), submetido à PMKT e aprovado para publicação.

## **RESUMO**

A identificação de usuários influentes nas mídias sociais é um assunto que tem gerado grande interesse por parte das empresas nos últimos anos. Este artigo visa avaliar esta influência por meio da utilização de grafos para entendimento da estrutura relacional existente entre os usuários, estabelecida por suas conversas no Twitter. A análise exploratória de dados e as técnicas de Mineração de Textos foram utilizadas para conclusões complementares acerca do assunto. O ambiente de conversas escolhido para avaliação foi o das cervejas brasileiras, sendo as buscas realizadas por palavras relacionadas às principais marcas atuantes no mercado nacional. A avaliação foi realizada, considerando uma amostra de 25 dias entre os meses de dezembro de 2013 e janeiro de 2014.

## **PALAVRAS-CHAVE:**

Cerveja, Twitter, Análise de Redes Sociais.

## **ABSTRACT**

*The identification of influential users in social media is a subject that has generated great interest by companies in recent years. This work aims to evaluate this influence through the use of graphs for understanding the existing relational structure between users, established through their conversations on Twitter. Exploratory data analysis and text mining techniques have been used to further conclusions about the subject. The "conversation environment" was chosen is Brazilian beer, and the search related words were the major active brands in domestic market. The evaluation was performed considering a sample of 25 days between the months of December 2013 and January 2014.*

## **KEYWORDS:**

*Beer, Twitter, Social network analysis.*

## 1. INTRODUÇÃO

Este artigo tem como objetivo principal a identificação dos usuários mais influentes no Twitter quando o assunto é cerveja. As informações utilizadas foram relativas à coleta de uma amostra de 25 dias entre os meses de dezembro de 2013 e janeiro de 2014, considerando apenas postagens realizadas em português, no Brasil.

O conteúdo das conversas também foi avaliado por meio de aplicação de algoritmos de *Text Mining*. Foi realizada uma análise descritiva acerca do comportamento geral dos usuários do Twitter que conversam sobre o assunto, visando o entendimento de aspectos relacionados à utilização, à repercussão das diferentes marcas e ao perfil dos usuários.

A maior quantidade de postagens sobre o assunto ocorreu na parte da tarde e da noite, existindo uma forte assimetria em relação à distribuição da quantidade de mensagens postadas pelos usuários; a grande maioria realizou apenas uma única postagem durante o período. Pela observação dos picos na série histórica da quantidade total de mensagens postadas, foi possível avaliar o efeito das datas comemorativas, sendo o efeito do Ano Novo, muito próximo ao que foi observado no Natal.

Quanto à avaliação semântica das postagens sobre cerveja, foram identificados diversos subtópicos (temas) dentro do assunto principal. Este tipo de informação pode auxiliar empresas no direcionamento de suas estratégias e acompanhamento contínuo do comportamento do consumidor. Percebeu-se que, muitos usuários quando postam mensagem sobre cerveja, também atrelam a informação de onde, com quem ou até mesmo quando irão consumi-la, além de citarem as marcas de sua preferência.

A análise de influência de usuários nas redes sociais possibilita a criação de diversas estratégias de marketing. Usuários mais influentes sobre determinado assunto podem ser contatados por empresas para divulgação de suas marcas, sendo utilizados como elos entre as empresas e os demais consumidores finais.

A medição da influência realizada neste trabalho foi feita com base na quantidade de conexões que os usuários apresentaram durante o período do estudo. No Twitter, usuários podem direcionar suas mensagens uns para os outros e repassar informações divulgadas por alguma de suas conexões (*retweets*).

Uma das formas para se avaliar o grau de influência de usuários consiste na verificação da quantidade de conexões que repassam suas mensagens ou pela quantidade de conexões com as quais eles direcionam suas mensagens. Neste artigo buscou-se avaliar esses dois pontos de vista.

O presente estudo foi estruturado da seguinte forma: após a introdução, na segunda etapa, o referencial teórico apontou as principais características das diferentes técnicas utilizadas na análise. Na terceira, realizou-se uma pequena explanação acerca da mídia social Twitter, em forte crescimento no Brasil. Na quarta, foi feita a contextualização do mercado de cerveja nacional, sua evolução, tendências e principais marcas atuantes. Na quinta, detalhou-se a metodologia analítica aplicada, esclarecendo as questões respondidas pelo estudo. Na sexta, realizou-se a explicação sobre a coleta de informações. Na sétima, foram mostrados os resultados da análise de dados. Na oitava, apresentaram-se as principais conclusões e, por último, a limitação e sugestão de novas pesquisas.

## 2. REFERENCIAL TEÓRICO

### 2.1 SOCIAL NETWORK ANALYSIS

Uma rede social é determinada por um conjunto de atores (ou nós) e relações preestabelecidas e definidas entre eles (WASSERMAN; FAUST, 1994). Os atores podem assumir diversas formas e representar indivíduos de diferentes grupos como pessoas, empresas e entidades. Justamente devido à grande flexibilidade, uma análise de redes sociais (comumente também chamada de *Social Network Analysis* – SNA), pode ser aplicada em, praticamente, qualquer contexto.

Geralmente, quando estudadas as redes sociais a partir de técnicas de SNA, são representadas visualmente por “grafos”. Nos grafos, os atores ou nódulos são representados por pontos e o relacionamento entre um par de nós é definido por arestas ou conexões. As conexões podem ser diretas quando for importante destacar qual foi o ator de origem desse relacionamento (WASSERMAN; FAUST, 1994). Segundo os mesmos autores, além da forma visual de ser apresentada, uma rede social pode ser descrita por uma matriz  $n \times n$ , em que  $n$  é o número total de atores naquela rede.

A existência de relacionamento entre o par de nós  $A$  e  $B$  seria dada pelo valor 1, na célula correspondente da matriz. A leitura pode ser feita da seguinte forma: as linhas representam nós de onde o relacionamento sai (atores de origem) e, as colunas, nós onde o relacionamento chega (atores de destino). Dessa forma, uma rede social indireta irá apresentar sempre uma matriz simétrica.

Com o objetivo de auxiliar no entendimento das relações entre os atores, existem algumas métricas que podem ser da rede, como um todo ou de cada nó, em específico. Entre elas se destacam:

- Grau (*degree*): quantidade de arestas ligadas a cada nó.
- *PageRank*: medida espectral de popularidade definida em grafos diretos com pesos de ligações não negativos (PAGE et al., 1998), e pode ser dado por:

$$G = (1 - \alpha)\bar{D}^{-1}A + \left(\frac{\alpha}{n}\right)J_{n \times n}$$

Em que:

$n$  = total de nós na rede.

$A$  = matriz  $n \times n$  com valores iguais a 1, quando a ligação existe entre os nós respectivos de um par linha/coluna.

$\bar{D}$  = matriz absoluta diagonal definida por  $\bar{D}_{ii} = \sum_j |A_{ij}|$ .

$J_{n \times n}$  = matriz composta apenas por valores iguais a 1,  $0 < \alpha < 1$  um parâmetro aleatório.

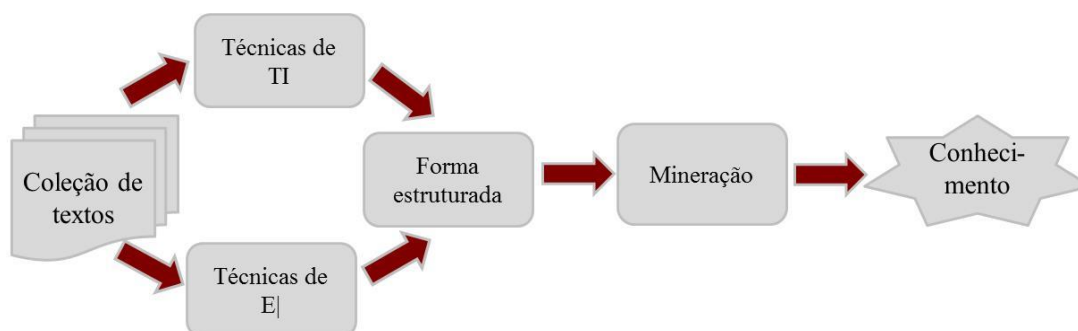
A matriz  $G$  é tal que, cada uma de suas linhas, soma 1 (KUNEGIS; LOMMATZSCH; BAUCKHAGE, 2009).

O *software* utilizado para desenvolver os grafos deste estudo foi o Gephi, um *freeware* que permite diversas formas de edição e personalização dos resultados finais. Ele pode ser utilizado na criação dos grafos e no cálculo das métricas da análise.

## 2.2 TEXT MINING

A Mineração de Textos, também conhecida como *Text Mining*, é o processo de extração de informação útil ou conhecimento de documentos de textos estruturados ou não (BARION; LAGO, 2008). No contexto desse estudo, essa técnica será aplicada para identificar padrões de comentários e opiniões emitidas por usuários do Twitter sobre o mercado nacional de cervejas.

Técnicas de Recuperação de Informação são aplicadas sobre um conjunto de textos, com o objetivo de torná-lo estruturado. A partir desses dados estruturados, são aplicadas técnicas de Mineração de Dados para se obter informações relevantes, conforme mostra a Figura 1.



Fonte: BARION, E. C. N.; LAGO, D. Mineração de textos. *Revista de Ciências Exatas e Tecnologia*, 2008.

### FIGURA 1

Processo de *Text Mining* ou Mineração de Textos.

O primeiro passo da Mineração é a indexação, processo que armazena uma estrutura de índices, a partir das palavras dos textos, e viabiliza a pesquisa de documentos por meio de todos os termos contidos ali (SALTON; MCGILL, 1983). Algumas etapas para uma análise de *Text Mining* (BARION; LAGO, 2008):

- **Análise Léxica:** converte uma sequência de caracteres numa sequência de palavras que serão candidatas a termos do índice. Nesta fase, é separado o alfabeto de entrada em caracteres e separadores de palavras.
- **Remoção de *Stop-words*:** remove um conjunto de palavras que aparecem com frequência em textos, mas não possuem valor semântico, tais como: preposições, artigos e conjunções. Essa fase é de extrema importância, pois diminui a base a ser indexada e facilita a mineração.
- ***Stemming*:** remove todas as variações de palavras, deixando apenas a raiz de cada uma, por exemplo, a palavra “amamos” passa a se identificar como a raiz “ama”.
- **Seleção dos termos-índice:** determina quais palavras ou radicais serão utilizados como elementos de indexação. Estas palavras são selecionadas de acordo com o peso atribuído a elas.
- ***Bag of Words* - BOW:** consiste em uma matriz na qual cada termo diferente presente na coleção de documentos é indexado. A partir desta indexação, cada documento pode ser representado por um vetor  $1 \times n$ , em que  $n$  é o número total de termos; cada entrada desse vetor será o número de vezes que os termos aparecem nesse documento (SIVIC, 2009).
- **Determinação dos pesos:** o preenchimento da matriz BOW é feito com base em métricas que ponderam a frequência de ocorrência dos termos nos documentos e na coleção total (conjunto de todos os documentos). A métrica mais comumente utilizada para esta finalidade é denominada *tf-idf* (*term frequency inverse document frequency*).

- Correlação (similaridade) entre termos: baseando-se na matriz BOW, pode-se calcular a correlação de Pearson entre diferentes termos, com a finalidade de medir quão relacionado eles estão pela fórmula (HUANG, 2008):

$$SIM_p \left( \vec{t}_a, \vec{t}_b \right) = \frac{m \sum_{t=1}^m w_{t,a} \times w_{t,b} - \sum_{t=1}^m w_{t,a} \times \sum_{t=1}^m w_{t,b}}{\sqrt{[m \sum_{t=1}^m w_{t,a}^2 - \sum_{t=1}^m w_{t,a}^2][m \sum_{t=1}^m w_{t,b}^2 - \sum_{t=1}^m w_{t,b}^2]}}$$

Em que:

$\vec{t}_x$  = vetor criado por meio do BOW.

$m$  = total de termos distintos na coleção toda de documentos.

$w_{t,a}$  = peso (*tf-idf*) do termo  $t$  no documento  $a$ .

Essa estatística pode variar de -1 a 1 e, quanto maior, mais correlacionados são os vetores; se essa correlação assumir valor igual a 1, os vetores são idênticos.

### 3. O TWITTER

O Twitter foi criado em 2006 pelos sócios Jack Dorsey, Evan Williams, Biz Stone e Noah Glass, em San Francisco, nos Estados Unidos. O serviço é uma rede social que permite aos usuários postarem e lerem *tweets*, que nada mais são do que mensagem de até 140 caracteres. Seu acesso pode ser feito diretamente em algum *browser* de internet, por aplicativos no celular e, em alguns países, as postagens podem ser feitas por SMS. A ideia rapidamente se espalhou e ganhou popularidade no mundo todo: em 2012, eram mais de 500 milhões de usuários registrados que postavam 340 milhões de *tweets* por dia (LUNDEN, 2012). De acordo com o *site* de informações de acessos em páginas da *Web* (<www.alexa.com>), o Twitter foi uma das dez páginas mais acessadas do mundo nesse ano.

Uma vez cadastrado, o usuário define um endereço no *site* que ainda não esteja sendo utilizado. A partir de então, ele será sempre conhecido por esse endereço precedido do símbolo @ pelos outros usuários.

Definido esse endereço e cadastrada a conta, o usuário poderá “seguir” ou “ser seguido” por outras contas. Isso significa que, cada vez que usuários seguidos postam algo, a mensagem aparece diretamente na sua página (também chamada de *timeline*). Por *default*, *tweets* são visíveis publicamente. No entanto, é possível restringir a visualização das mensagens para apenas seus seguidores. Outra possibilidade de mensagem é repostar o que já foi postado por alguém, prática também conhecida como *retweet*, e que é caracterizada pela sigla RT. O objetivo, nesse caso, é o usuário repassar esse determinado texto para todos que o seguem (STRACHAN, 2009).

Quando uma postagem é feita em cima de um tópico específico, o usuário pode fazer uso de uma técnica chamada *hashtag* – frases ou palavras que começam com o símbolo # (STRACHAN, 2009). Da mesma forma, se o interesse for visualizar apenas mensagens daquele tópico, uma busca pode ser feita utilizando o mesmo termo em *hashtag*.

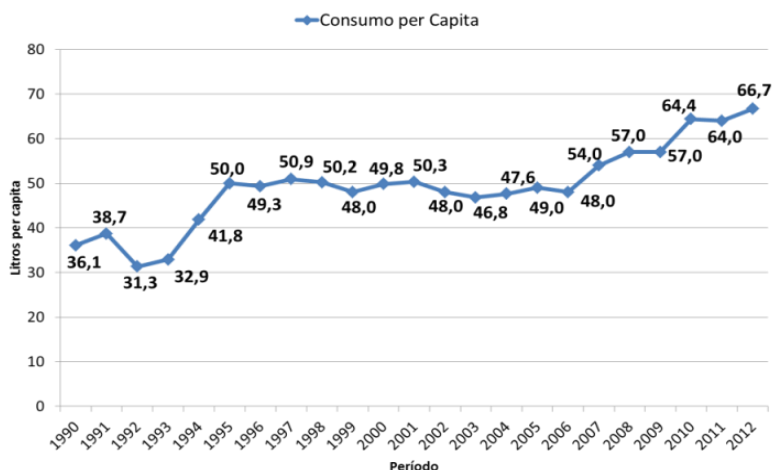
Uma palavra, frase ou tópico (*hashtag*), quando muito mencionada ao mesmo tempo por um número grande de usuários diferentes, pode ser considerado um *trending topic* (CHOWDHURY, 2009). *Trending topics* acontecem, geralmente, quando esforços de um grupo de usuários com

interesse em comum, são reunidos em prol de algum objetivo ou quando grandes e populares eventos estão acontecendo.

#### 4. MERCADO DE CERVEJAS

Atualmente, o Brasil possui um mercado de cerveja extremamente competitivo, no qual se destacam empresas como Ambev, Brasil Kirin e Grupo Petrópolis. Com um faturamento de R\$ 63 bilhões em 2012, o país é o terceiro maior fabricante mundial de cerveja e o 26º no *ranking* internacional de consumo (VALOR ECONÔMICO, 2013).

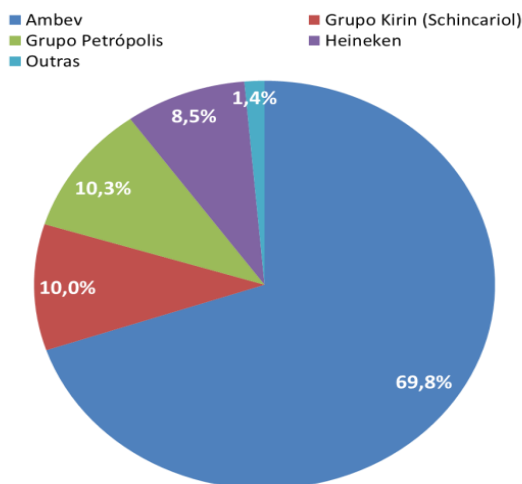
O *market share* deste mercado no Brasil está concentrado nas cervejarias AmBev, Grupo Kirin e Grupo Petrópolis que, juntas têm 90% do mercado. Outra informação importante é o consumo *per capita*, medido em litros e por ano. Em 2012, o consumo chegou a 66,7 litros *per capita* (Gráfico 1).



#### GRÁFICO 1

Consumo nacional de cerveja (litros *per capita*).

Desde 2008 o consumo de cervejas no Brasil tem apresentado acentuado crescimento (Gráfico 2).



#### GRÁFICO 2

Market share do mercado brasileiro de cerveja.



Devido à relevância do mercado de cerveja na economia brasileira e seu contínuo crescimento decidiu-se elaborar esse estudo, no qual foi realizado o monitoramento das seguintes marcas: Antarctica, Baden Baden, Bohemia, Brahma, Budweiser, Eisenbahn, Itaipava, Nova Schin, Serramalte, Skol e Stella Artois, além da palavra cerveja e duas de suas variações regionais: cerva e breja.

## 5. METODOLOGIA ANALÍTICA

A metodologia analítica aplicada consiste na execução de três passos: o primeiro refere-se à análise do comportamento geral e do perfil dos usuários quanto à utilização do Twitter para realizar *posts* sobre cerveja; o segundo, à análise semântica que se baseia na utilização de técnicas de *Text Mining* e estatística multivariada para a identificação dos temas mais pertinentes dentro do ambiente de conversa de cerveja e, por último, à avaliação sobre a influência dos usuários.

### 5.1 COMPORTAMENTO GERAL E PERFIL DOS USUÁRIOS QUANDO O ASSUNTO É CERVEJA

Na primeira etapa analítica, buscou-se avaliar as principais métricas agregadas presentes no trabalho agrupado no tempo. As mais importantes foram as seguintes:

- Quantidade de *posts*: avalia o número total de postagens realizadas por intervalo de tempo.
- Quantidade de usuários distintos: avalia o número total de usuários distintos que realizaram postagens por intervalo de tempo.
- Média de *posts* por usuário: calculada por meio da divisão da quantidade de *posts* pela quantidade de usuários distintos.
- Percentual de *posts*: proporção de postagens classificadas em cada uma das categorias existentes.

A análise da quantidade total de *posts* torna possível a avaliação da intensidade total dos impactos ocorridos durante o período observado. Por meio da média de *posts* por usuário é possível verificar, de forma genérica, o grau da intensidade de divulgação do assunto dentre os usuários considerados, de forma que, quanto mais próxima de 1 estiver a média, menor será essa intensidade. Já o percentual de *posts*, avalia o peso de cada categoria existente dentro de determinada variável categórica sob o total de postagens considerado.

A avaliação dessas métricas serve para o entendimento das características acerca do comportamento geral dos usuários sobre o assunto. A identificação dos picos foi feita pela visualização da série temporal da quantidade de *posts*. O mesmo procedimento deve ser realizado para a avaliação da curva horária.

Alternando-se a quantidade de *posts* para a média de *post* por usuário foi possível avaliar mudanças quanto ao comportamento individual dos usuários. Muitas vezes, ocorrem grandes variações dessa métrica em determinados intervalos de tempo, devido a eventos específicos em que usuários tendem a realizar mais postagens acerca de um mesmo assunto.

O Twitter possibilita a utilização de métricas específicas que denotam os diferentes tipos de comportamento de seus usuários, dentre elas é possível destacar a penetração (proporção de *posts* com determinada característica). As características avaliadas foram as seguintes:



- RTs: *twetts* que repassaram uma mensagem que já havia sido postada anteriormente por outro usuário.
- @: direcionamento de mensagens para outra pessoa.
- Http: *twetts* que possuem informações contidas em *sites* da internet.
- Hashtag (#): grupo de discussão sobre algum assunto específico.
- Outros: *twetts* que não possuem nenhuma das características citadas.

## 5.2 ANÁLISE DE INFLUÊNCIA

A análise de influência é feita a partir de uma rede de conversas na qual foram observados dois níveis distintos de influência, o primeiro avalia quanto determinado *tweet* é passado à frente, sendo retuitado. O outro nível de influência engloba os *twetts* enviados diretamente a outros usuários.

No primeiro nível, constatou-se quão influente um usuário é verificando quantos outros usuários fizeram o retuite de sua postagem. No segundo, verificou-se a influência pela quantidade de conversas direcionadas entre os usuários. Neste artigo, foram considerados os dois casos e todos os tipos de conexões existentes entre usuários. Contudo, em termos práticos, o efeito dos retuites é sempre mais impactante.

## 5.3 ANÁLISE DE SEMÂNTICA

A análise da correlação entre tópicos foi feita seguindo o seguinte processo: primeiramente foi realizada a análise léxica. Em um segundo momento, foi feita a limpeza de *stopwords* (palavras sem valor semântico), para posterior execução do algoritmo de *stemming* (extração de radicais).

Após esses passos, a matriz BOW foi calculada. Nesta matriz, cada termo considerado corresponde a uma coluna e cada linha a um documento (*tweet*).

A medida utilizada para avaliação foi o *tf-idf* (*term frequency inverse document frequency*). Com base na informação da matriz foi possível a obtenção das medidas mais associadas à determinada palavra, esta similaridade foi avaliada pela correlação de Pearson.

Já a classificação dos *posts* por tema foi gerada por meio da elaboração de uma heurística de identificação baseada na seleção de palavras-chave definidas por especialistas.

O processo de avaliação das palavras a serem consideradas ocorre da seguinte maneira:

- Passo 1: definição de palavras-chave que caracterizem determinado tema.
- Passo 2: elaboração de algoritmo para contagem das palavras-chave definidas no passo 1.
- Passo 3: repetir os passos 1 e 2 até que a proporção de *posts* classificados em algum tema seja considerada satisfatória.

Geralmente, a proporção mínima de *posts* classificados em temas para obtenção de resultados consistentes é de 50%.

## 6. COLETA DE DADOS

A extração das informações foi feita por meio de um programa desenvolvido pelo IBOPE DTM que se conecta diretamente à API do Twitter.

Com base na distribuição do *market share* de cerveja no mercado brasileiro, foi decidido estudar apenas as marcas das empresas de maior relevância no segmento: Ambev, Grupo Kirin e Grupo Petrópolis. Logo, realizou-se o monitoramento das seguintes marcas: Antártica, Baden Baden, Bohemia, Brahma, Budweiser, Eisenbahn, Itaipava, Nova Schin, Serramalte, Skol e Stella Artois, além da palavra cerveja e duas de suas variações regionais: cerva e breja.

Os dados referem-se a todas as postagens realizadas durante o período avaliado que continham as palavras especificadas.

Após 25 dias de monitoramento, foram obtidos 438.507 *tweets* (*posts*) relacionados à cerveja. Entretanto, como o estudo foi feito com foco na divulgação no Brasil, só foram considerados os *posts* na língua portuguesa, logo, passou-se a trabalhar com 291.043 *posts* (66,4%).

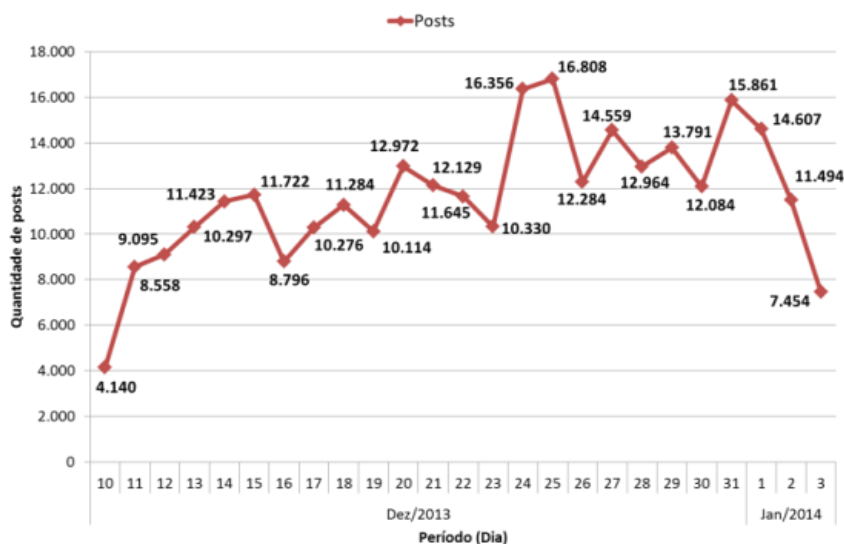
O período de monitoramento, de 10/12/2013 a 03/01/2014, foi escolhido com base na hipótese de que as datas comemorativas de final do ano: Natal e Ano Novo influenciavam no número de postagens sobre cerveja no Twitter.

## 7. ANÁLISE DE DADOS

A análise realizada seguiu a mesma estrutura da metodologia apresentada. Primeiramente foi avaliada a distribuição geral das postagens.

### 7.1 COMPORTAMENTO GERAL E PERFIL DOS USUÁRIOS QUANDO O ASSUNTO É CERVEJA

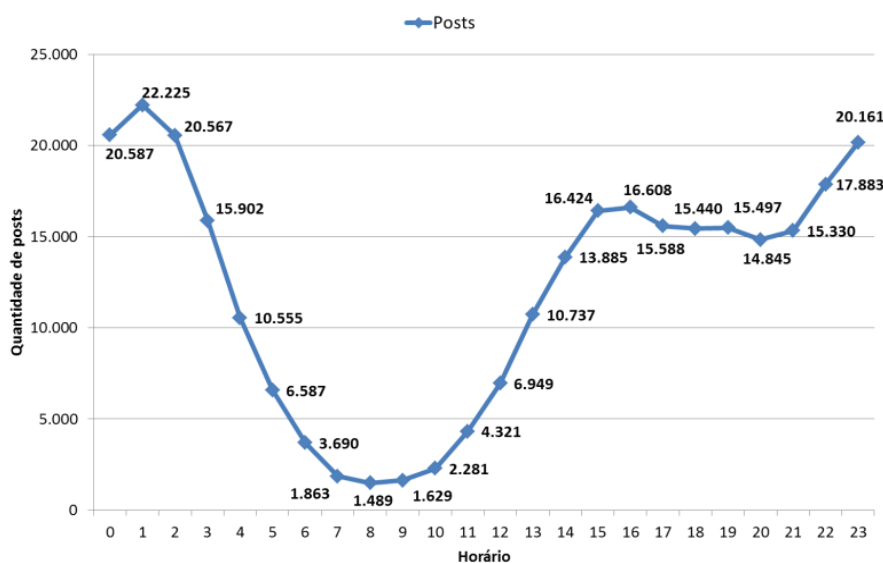
Os impactos causados pelas datas comemorativas exerceram grande variação na quantidade diária de *tweets* postados. No Gráfico 3, é possível verificar que os dias que tiveram incidência de picos de postagens foram 24, 25 e 31 de dezembro, ou seja, véspera de Natal, Natal e véspera de Ano Novo, em que observou-se um aumento de mais de 4 mil *posts* em relação à média total do período.



#### GRÁFICO 3

Distribuição de *posts* sobre cerveja no Twitter.

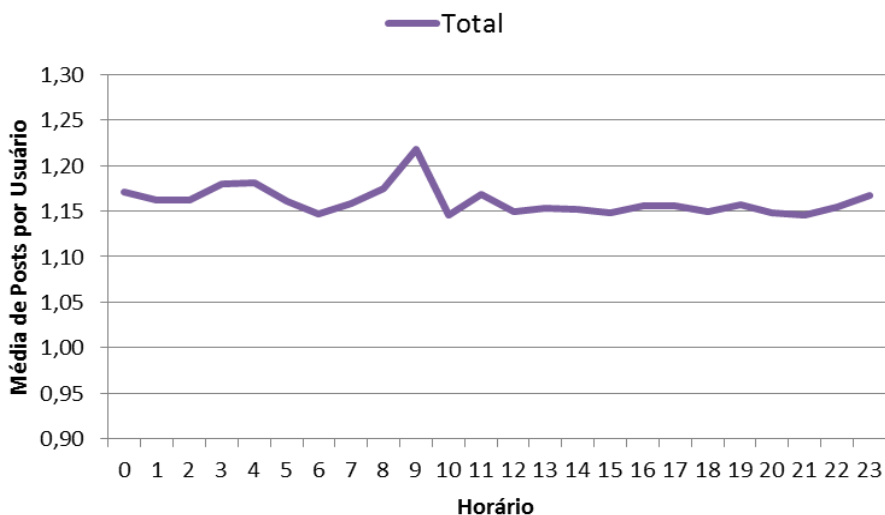
Quanto ao horário das postagens (Gráfico 4), foi observado um crescimento acentuado a partir das 10 horas da manhã, o qual se estabilizou entre às 15 e 21 horas.



#### GRÁFICO 4

Quantidade de *posts* por hora.

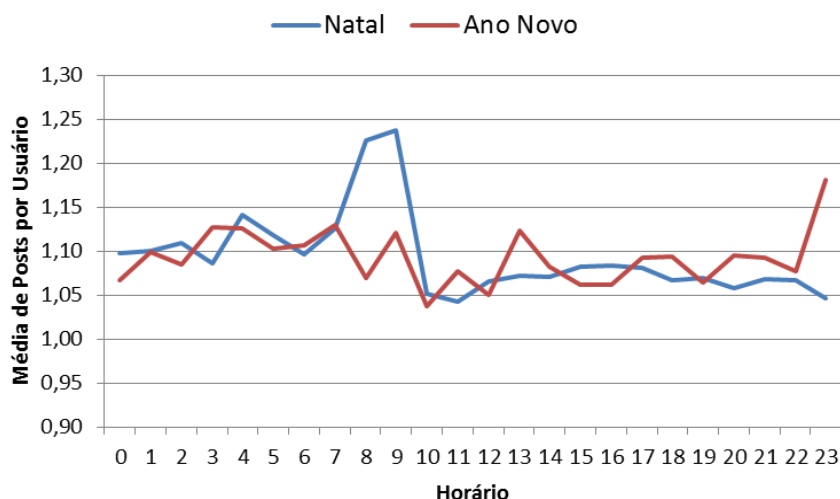
Ao analisar a postagem média, ou seja, quantidade média de *post* por usuário, pode-se ver que houve um pico na média às 9 horas da manhã (Gráfico 5).



#### GRÁFICO 5

Quantidade de *posts* médio por Total, Natal e Ano Novo.

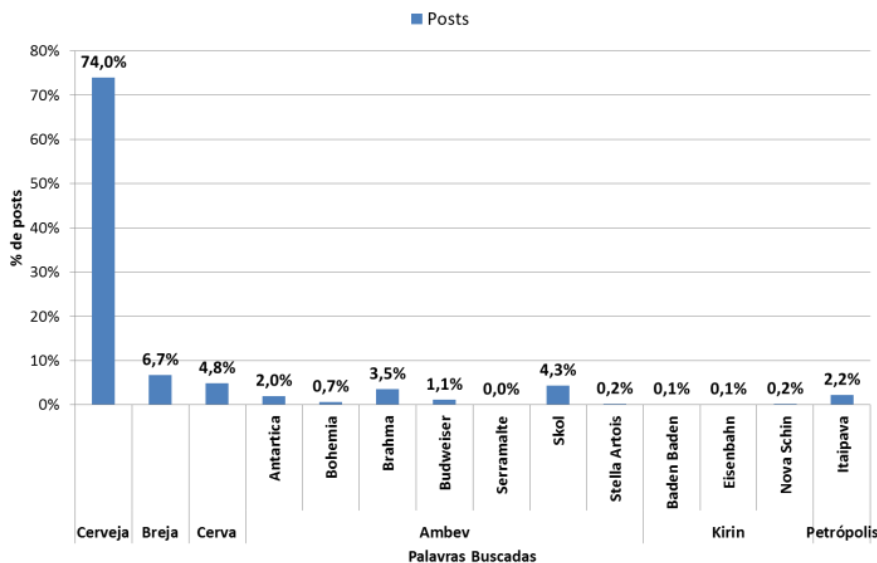
Entretanto, quando foram detalhadas as datas comemorativas do Natal e do Ano Novo foi visto que, no Natal, a maior incidência de *posts* médios ocorreu entre as 8 e 9 horas, enquanto no Ano Novo essa maior incidência de *posts* médios ocorreu no período das 23 horas, conforme mostra o Gráfico 6.



### GRÁFICO 6

Quantidade de *posts* médio por Total, Natal e Ano Novo.

No Gráfico 7, é possível ver que, 85,5% dos *posts* referentes à cerveja não citam uma marca específica. Entretanto, dos 14,5% dos *posts* com citação de algumas marcas buscadas, a Skol é a que possui maior expressão no Twitter com 4,3%, seguida pela Brahma com 3,5% e Itaipava com 2,2% dos *posts*.



### GRÁFICO 7

Percentual de *posts* por palavras buscadas.

Quanto às métricas individuais (Tabela 1), foi visto que a única marca que teve destaque em *posts* com *hashtag* (#) foi a Eisenbahn com 17,9% dos *posts*, ou seja, é a única marca que gera alguma discussão sobre algum outro assunto específico.

As marcas divulgadoras de informações contidas em *site* (*http*) foram as marcas: Baden Baden com 41,4%, Eisenbahn com 36,5%, Antartica com 32,4% e Stella Artois com 29,4% dos *posts*. No

âmbito mensagens direcionadas @, a marca Serramalte se destacou com 32,5%, seguida da Nova Schin com 23,4% dos *posts*.

Por último, o repasse de mensagens já anteriormente postadas (RT) foram maiores nas marcas Budweiser em 26,9% e a Antartica em 21,1% dos *posts*.

**TABELA 1**

Quantidade de *posts* por palavra buscada e suas métricas individuais.

PALAVRA BUSCADA	POSTS	%	% PENETRAÇÃO TIPO DE POST					
			RT	@	H1P	HASTAG	OUTROS	
CERVEJA	215.229	74,0%	21,1%	13,2%	8,0%	3,8%	56,3%	
BREJA	19.537	6,7%	11,0%	16,9%	6,1%	3,5%	64,4%	
CERVA	14.112	4,8%	11,2%	15,8%	5,7%	3,8%	65,7%	
AMBEV	ANTARCTICA	5.781	2,0%	21,1%	11,3%	32,4%	4,5%	33,7%
	BOHEMIA	1.943	0,7%	6,2%	13,2%	17,4%	9,0%	59,1%
	BRAHAMA	10.234	3,5%	15,7%	13,9%	16,1%	6,7%	52,0%
	BUDWEISER	3.232	1,1%	26,9%	8,7%	12,1%	9,5%	50,5%
	SERRAMALTE	114	0,0%	4,4%	32,5%	13,2%	12,3%	45,6%
	SKOL	12.632	4,3%	15,4%	13,5%	12,9%	9,0%	55,7%
	STELLA ARTOIS	574	0,2%	9,2%	7,0%	29,4%	6,3%	52,4%
KIRIN	BADEN BADEN	331	0,1%	3,0%	16,6%	41,4%	6,9%	38,1%
	EISENBAHN	263	0,1%	6,1%	11,8%	36,5%	17,9%	44,9%
	NOVA SCHIN	538	0,2%	15,2%	23,4%	9,3%	3,2%	50,2%
PETRÓPOLIS	ITAIPA VA	6.523	2,2%	10,2%	14,2%	19,5%	3,7%	54,9%
<b>TOTAL</b>	<b>291.043</b>	<b>100,0%</b>	<b>19,1%</b>	<b>13,6%</b>	<b>9,2%</b>	<b>4,2%</b>	<b>56,5%</b>	

Com foco nos usuários do Twitter que fizeram algum comentário sobre cerveja, foi possível ver que apenas uma pessoa foi responsável por 43,3% dos *posts* sobre cerveja (Tabela 2), porém ele possui apenas 153 seguidores, ou seja, apenas seus 153 seguidores viram diretamente a informação divulgada pelo *post*.

**TABELA 2**

Ranking dos dez usuários com maior quantidade de postagem.

RANK	USUÁRIOS (NOME TWITTER )	POSTS	SEGUIDORES NO TWITTER
1	BEEININDEX	1461	153
2	SKOL_	443	107
3	DJ_RICARDOO	348	512
4	CERVEJA_DUFF	208	155
5	RENATORDM	188	514
6	ITAIPA VA_	185	415
7	PREDRERO	162	28.107
8	MARCIO_SKOL	157	171
9	SERRALHERO	107	2.181
10	GORONAH	105	769
<b>TOTAL</b>		<b>3364</b>	

Seguindo essa linha de raciocínio, a cantora Claudia Leitte fez um *post* sobre cerveja, mas essa informação foi vista por seus 7.869.106 seguidores (Tabela 3).

**TABELA 3**

Ranking dos dez usuários com maior quantidade de seguidores no Twitter.

RANK	USUÁRIOS (NOME TWITTER )	POSTS	SEGUIDORES NO TWITTER
1	CLAUDIALEITTE	1	7.869.106
2	DANILOGENTILI	1	5.324.329
3	SPIDERANDERSON	1	4.226.383
4	CLARORONALDO	1	3.625.623
5	PRETAGIL	2	3.450.693
6	PORTALR7	5	2.835.528
7	VEJA	2	2.825.215
8	BGAGLIASSO	1	2.735.376
9	GI	11	2.220.615
10	SIGNOSFODAS	1	1.432.674
<b>TOTAL</b>		<b>26</b>	

Para analisar a influência dos usuários, foi feito um *ranking* dos 20 usuários de maior PageRank. A usuária “frasesdebebada” possui um PageRank de 0,007 e 365 ligações (Tabela 4), ou seja, ela teve a maior influência na rede. Também é possível ver na Tabela 4, a presença de dois usuários que falaram sobre cerveja no Twitter, que se encontram no *ranking* dos 10 usuários com maior quantidade de seguidores (Tabela 3).

**TABELA 4**

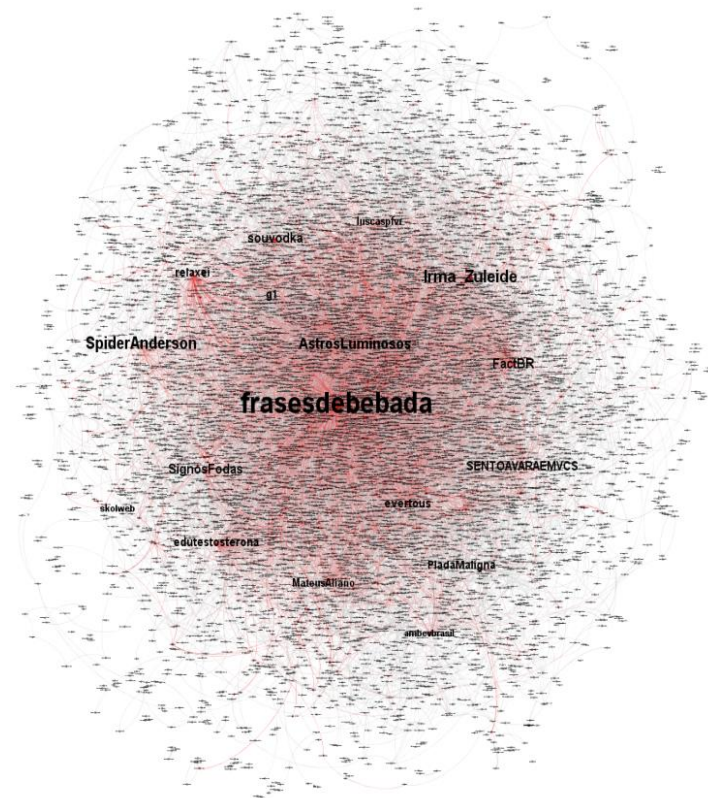
Ranking dos 20 usuários com maior PageRank.

RANK	USUÁRIOS	TIPO DE PESSOA	DEGREE	PAGERANK
1	FRASESDEBEBADA	FÍSICA	365	0,0070
2	IRMA_ZULEIDE	FÍSICA	51	0,0033
3	SPIDERANDERSON	FÍSICA	40	0,0029
4	ASTROSLUMINOSOS	JURÍDICA	73	0,0024
5	SIGNOSFODAS	JURÍDICA	48	0,0021
6	FACTBR	JURÍDICA	160	0,0020
7	SOUVODKA	FÍSICA	60	0,0018
8	SENTOAVARAEMVCS	FÍSICA	32	0,0017
9	EDUTESTOSTERONA	FÍSICA	98	0,0016
10	EVERTOUS	FÍSICA	108	0,0016
11	PIADAMALIGNA	FÍSICA	19	0,0015
12	GI	JURÍDICA	89	0,0014
13	RELAXEI	FÍSICA	96	0,0013
14	MATEUSALIANO	FÍSICA	93	0,0012
15	LUCASPFVR	FÍSICA	49	0,0011
16	FELIXPASSIVA	FÍSICA	22	0,0010
17	BITCH_MALVADA	FÍSICA	15	0,0010
18	EUZOERO	FÍSICA	24	0,0010
19	PREDRERO	JURÍDICA	25	0,0009
20	UMVINGADOR	FÍSICA	12	0,0009

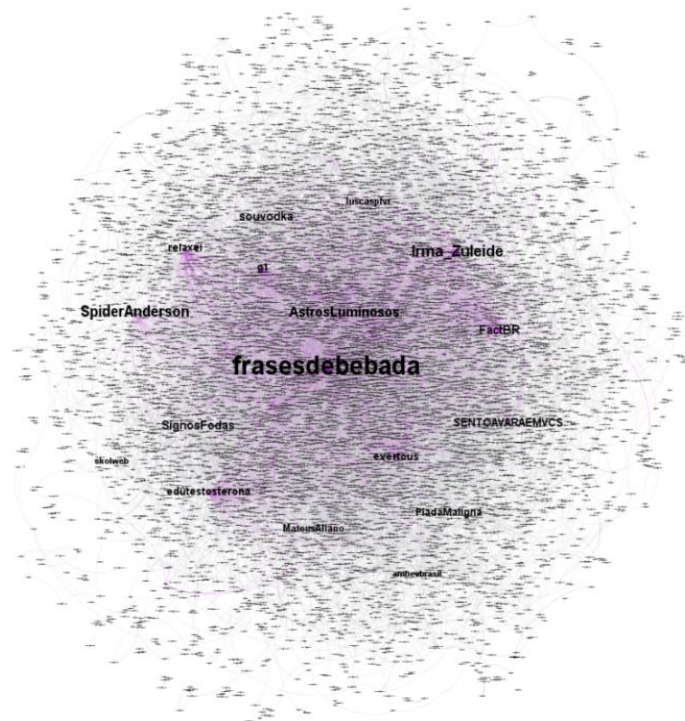




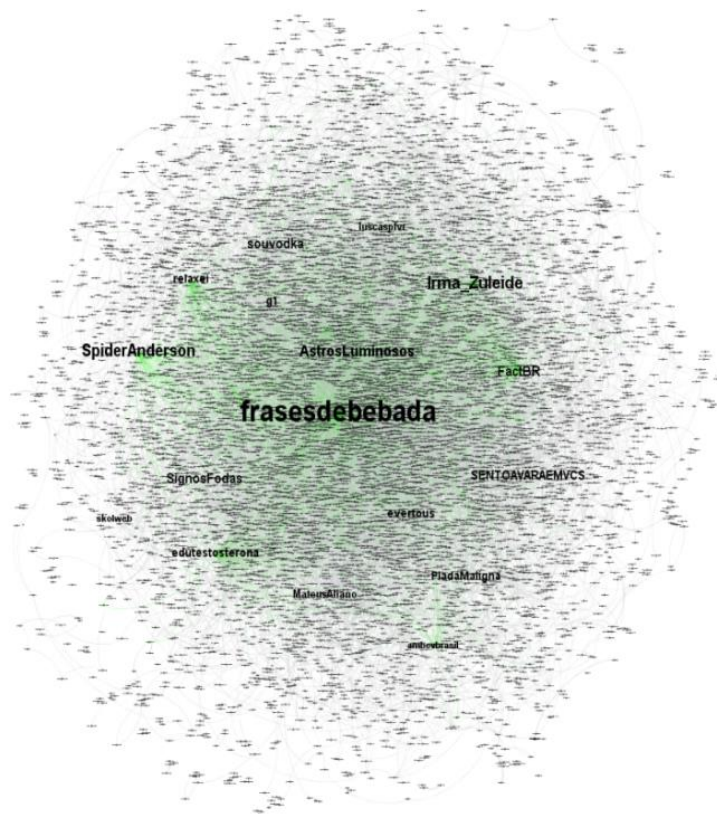




**FIGURA 3**  
Rede da usuária “frasesdebebada”.



**FIGURA 4**  
Rede da usuária “Irma\_Zuleide”.



**FIGURA 5**  
Rede do usuário “SpiderAnderson”.

Na análise semântica, é possível verificar que não existe uma palavra que se associe a mais de uma marca.

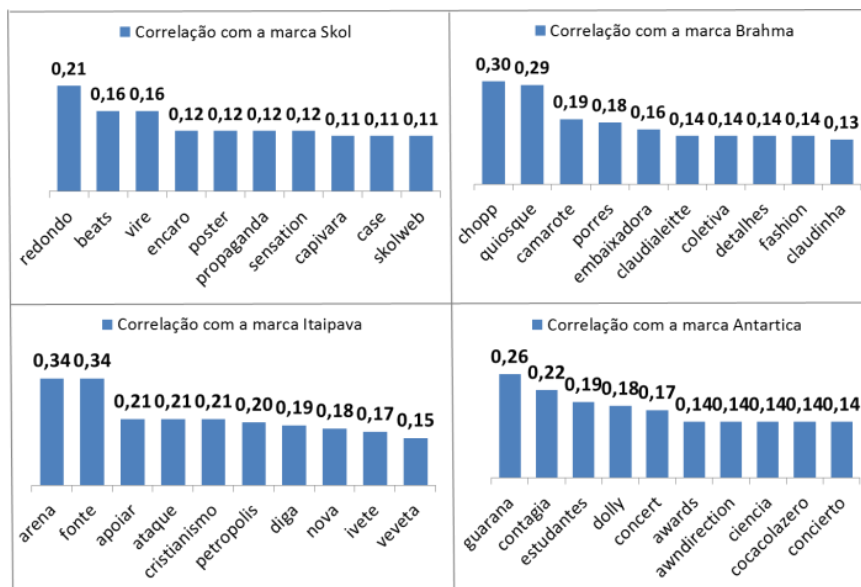
Logo, para facilitar a visualização do estudo, foram selecionadas as dez palavras de maior associação com as marcas estudadas. As marcas foram escolhidas de acordo com seu volume de *posts*.

Foi constatado que a Skol foi responsável por 4,3% dos *posts* relacionados a cerveja, a Brahma com 3,5%, a Itaipava com 2,2% e a Antarctica com 2,0%.

A palavra mais associada à Skol foi redondo, com uma correlação de Pearson igual a 0,21, seguido das palavras *beats* e *vire* com correlação de 0,16 (Gráfico 8).

Um diferencial que a marca Brahma teve sobre as demais marcas foi que a garota propaganda da marca, Claudia Leitte, apareceu na 6ª posição das palavras mais associadas com 0,14 de correlação (Gráfico 8).

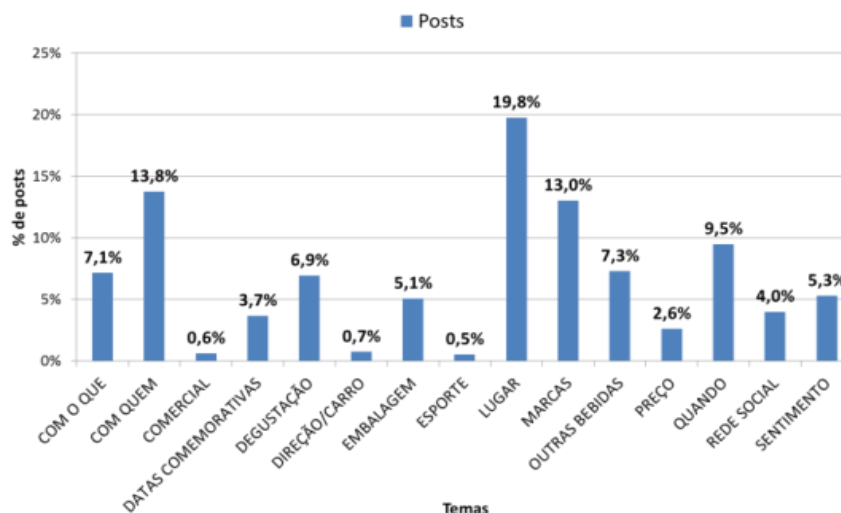
No caso da marca Antarctica, da Ambev, tem-se uma correlação maior sobre o refrigerante e seus concorrentes do que com a cerveja especificamente (Gráfico 8).



### GRÁFICO 8

Palavras (*top 10*) com maior correlação com as marcas.

Um grupo de especialistas em semântica foi responsável pela seleção de palavras-chave agrupadas em alguns temas considerados como principais quando o assunto é cerveja. Foi obtido um total de 39,2% de *posts* sem nenhuma classificação. Esses *posts*, em geral, possuem informação sobre cerveja, mas sem conteúdo relevante. Entretanto, é possível observar no Gráfico 9, a distribuição dos 60% dos *posts* com classificação. A partir daí, foi vista uma concentração de *posts* nos assuntos referentes ao lugar onde foi consumida a bebida (19,8%), com quem beberam (13,8%) e sobre as marcas especificamente (13,0%).

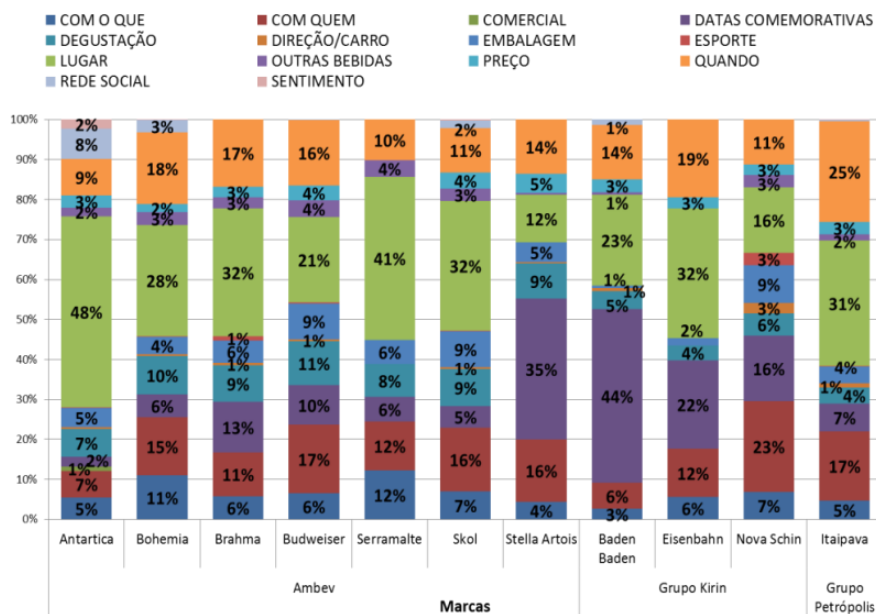


### GRÁFICO 9

Percentual de *posts* por tema.

Ao se analisar os temas mais comentados nas marcas estudadas (Gráfico 10), foi visto que, das cervejas produzidas pela Ambev, a marca Stella Artois, possui 35% dos *posts* no tema Datas Comemorativas (Gráfico 10), ao contrário das demais marcas dessa empresa com *posts* no tema

Lugar. Já as cervejas do Grupo Kirin ficaram divididas em três temas: Baden Baden com 44% dos *posts* em Datas Comemorativas, a Eisenbahn com 32% no tema Lugar e a Nova Schin com 23% no tema Com Quem. No Grupo Petrópolis a cerveja Itaipava obteve 31% dos *posts* em Lugar contra 25% no tema Quando.



**GRÁFICO 10**  
 Percentual de *posts* por tema aberto pelas marcas de cerveja.

## 8. CONCLUSÕES

Foi observado neste artigo que datas comemorativas exercem grande influência na quantidade de postagens relacionadas à cerveja, chegando a aumentos superiores a 35% na quantidade total de postagens diárias.

Durante o dia, de maneira geral, existe um aumento na quantidade de postagens nos períodos da tarde e da noite. Os horários com maior intensidade de postagens foram entre as 23 e 2 horas da manhã.

A análise da rede social identificou de maneira eficiente os usuários influentes por meio da quantidade e qualidade de conexões durante o período. Vários influenciadores foram identificados, dentre eles se destacam Anderson Silva, que enviou um *tweet* agradecendo seus patrocinadores antes da luta, e o G1, empresa de comunicação.

A análise semântica dos *posts* para identificação de assuntos relacionados à cerveja, demonstrou que existe uma concentração de *posts* relacionados ao lugar de consumo da bebida, com quem consumiram e quais foram as marcas consumidas.

Do grupo Kirin, cada uma teve uma incidência maior em temas diferentes: a Baden Baden teve maior quantidade de *posts* associados a Datas Comemorativas, a Eisenbahn com *posts* associados a Lugar e a Nova Schin com *posts* associados ao tema Com Quem. Já a Itaipava, do Grupo Petrópolis, teve uma incidência maior em *posts* com o tema Lugar.



## 9. LIMITAÇÃO E SUGESTÃO DE NOVAS PESQUISAS

Como não houve rupturas bruscas na série histórica da quantidade de postagens, entende-se que não houve problema de queda de conexão com a API do Twitter, portanto pode-se confiar na consistência e qualidade das informações utilizadas neste estudo.

Sugere-se que, em estudo futuro, seja interessante a realização da análise com um histórico maior de informações, a fim de entender se existe alguma sazonalidade quanto ao comportamento dos usuários acerca do tema.

Outra hipótese a ser estudada consiste na avaliação da diferença entre os horários de consumo e a postagem.

## 10. REFERÊNCIAS BIBLIOGRÁFICAS

ALEXA. Disponível em: <[www.alexa.com](http://www.alexa.com)>. Acessado em: 6 jan. 2014.

BAEZA-YATES, R.; RIBEIRO NETO, B. *Modern information retrieval*. Addison-Wesley, 1999.

BARION, E. C. N.; LAGO, D. Mineração de textos. *Revista de Ciências Exatas e Tecnologia*, 2008.

BAVELAS, Alex. *A mathematical model for group structure*. Applied Anthropology 7, 1948.

CERVBRASIL. *A Cerveja – Contribuição econômica*, s. d. Disponível em: <<http://www.cervbrasil.org.br/a-cerveja/contribuicao-economica/>>. Acessado em: 6 jan. 2014.

CERVEJAS DO MUNDO. *História da cerveja*, 2009. Disponível em: <<http://www.cervejasdomundo.com/Brasil.htm>>. Acessado em: 6 jan. 2014.

CHOWDHURY, A. *Top Twitter Trends of 2009*. Twitter Blog, 15 dez. 2009. Disponível em: <<https://blog.twitter.com/2009/top-twitter-trends-of-2009>>. Acessado em: 3 fev. 2014.

CORRÊA, A. C. G. *Recuperação de documentos baseada em Informação Semântica no Ambiente AMMO*. UFSCAR, 2003.

COUTINHO, C. A. T.; QUINTELLA, C. A. S.; PANZANI, M. M. *História da Cerveja no Brasil*. Portal São Francisco, s. d. Disponível em: <<http://www.portalsaofrancisco.com.br/alfa/historia-da-cerveja/historia-da-cerveja-no-brasil.php>>. Acessado em: 6 jan. 2014.

HUANG, A. *Similarity Measures for Text Document Clustering*. Department of Computer Science, The University of Waikato, 2008.

KUNEGIS, J.; LOMMATZSCH, A.; BAUCKHAGE, C. *The Slashdot zoo: mining a social network with negative edges*. Track: Social Networks and Web 2.0 / Session: Interactions in Social Communities, 2009.

LIU, Bing. *Web Data Mining: exploring hyperlinks, contents, and usage data*. Springer, 2011.

LUNDEN, I. *Analyst: Twitter Passed 500M Users In June 2012, 140M Of Them In US; Jakarta 'Biggest Tweeting' City*. TechCrunch, 30 jul. 2012. Disponível em: <<http://techcrunch.com/2012/07/30/analyst-twitter-passed-500m-users-in-june-2012-140m-of-them-in-us-jakarta-biggest-tweeting-city/>>. Acessado em: 3 fev. 2014.

MANNING, C. D.; RAGHAVAN, P.; SCHUTZE, H. *Scoring, term weighting, and the vector space model: introduction to information retrieval*. Stanford, 2008.

MELO, I. D. et al., *Análise de Redes Sociais*. Universidade Federal da Paraíba, 2013.

MOURA, M. F. *Proposta de utilização de mineração de textos para seleção, classificação e qualificação de documentos*. Campinas: Embrapa Informática Agropecuária, 2004.

NÚCLEO EDUCACIONAL DE BROGLIE. *Produção e consumo de cerveja no Brasil e no mundo*, 2013. Disponível em: <<http://www.nucleodebroglie.com/2013/03/producao-e-consumo-de-cerveja-no-brasil.html>>. Acessado em: 6 jan. 2014.

PAGE, L. et al. *The PageRank citation ranking: bringing order to the web*. Technical report, Stanford Digital Library Technologies Project, 1998.

QUEIROZ, D. F. *Análise estrutural do setor cervejeiro*. FAEC – Departamento de Economia, 2010. Disponível em: <<http://pt.slideshare.net/diegofelinto/monografia-2010-anlise-estrutural-do-setor-cervejeiro-no-brasil-diego-queiroz>>. Acessado em: 6 jan. 2014.

SALTON, G.; MCGILL, M. J. *Introduction to modern information retrieval*. Computer Science Series, USA: McGraw-Hill, 1983.

SILVA, Anderson. (SpiderAnderson) tweets. Disponível em: <live <http://t.co/2aBqwULK>>. Acessado em: 15 abr. 2014.

SANTOS, M. A. M. R. *Extraindo regras de associação a partir de textos*. PUC, 2002.

SINDICATO NACIONAL DA INDÚSTRIA DA CERVEJA – SINDICERV. *Mercado*, s. d. Disponível em: <<http://www.sindicerv.com.br/mercado.php>>. Acessado em: 6 jan. 2014.

SIVIC, J. *Efficient visual search of videos cast as text retrieval*. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, v. 31, n. 4, IEEE, 2009.

STRACHAN, D. *Twitter: how to set up your account*. Telegraph, 19 fev. 2009. Disponível em: <<http://www.telegraph.co.uk/travel/4698589/Twitter-how-to-set-up-your-account.html>>. Acessado em: 3 fev. 2014.

TWITTER, *Finding your Twitter short or long code*. Disponível em: <<http://help.twitter.com/entries/14226-how-to-find-your-twitter-short-long-code>>. Acessado em: 3 fev. 2014.

VALOR ECONÔMICO. *Ritmo de produção de cerveja cai em 2013*. 2013. Disponível em: <<http://www.valor.com.br/empresas/3221828/ritmo-de-producao-de-cerveja-cai-em-2013>>. Acessado em: 6 jan. 2014.

WASSERMAN, Stanley; FAUST, Katherine. *Social network analysis: methods and applications*. Cambridge: Cambridge University Press, 1994.