Identificação e Análise do Comportamento de Perfis de Usuários em um Serviço de Mensagens Curtas

Adriane Gomes[1], Lesandro Ponciano[1], Wladmir Brandão[1]

[1]Pontifícia Universidade Católica de Minas Gerais (PUC Contagem)
Rua Rio Comprido, 4.580, CEP 32010-025 - Contagem - MG - Brasil
Bacharelado em Sistemas de Informação

ajmgomes@sga.pucminas.br, {lesandrop, wladmir}@pucminas.br

Resumo. *Serviços de Mensagens Curtas são amplamente utilizados e geram um volume expressivo de dados. A interatividade fornecida por tais serviços, como salas de bate-papo, proporciona o uso intensivo entre os jovens. A garantia da segurança nestes ambientes apresenta-se como um desafio, sendo que parte deste processo consiste na identificação de perfis mal-intencionados. A identificação de tais perfis tem o intuito de contribuir para a detecção de conteúdo mal-intencionado, além de promover maior retenção do cliente ao serviço da empresa prestadora mediante a melhoria da segurança ofertada. O objetivo deste artigo é identificar e analisar o comportamento dos usuários de um serviço de bate-papo SMS. Aplicando o algoritmo de mineração de dados obteve-se cinco perfis distintos quanto à utilização do serviço que foram definidos como: Infrequentes, Regulares, Moderados, Frequentes e Hiperativos. Observou-se ainda que os usuários denominados Frequentes e Hiperativos são os mais propensos a enviarem mensagens com conteúdo Mal-intencionado - respectivamente, 30,23% e 25% - e que, portanto, poderiam ser submetidos a um acompanhamento específico ao utilizarem as salas de bate-papo SMS com a finalidade de tornar o ambiente mais seguro.*

1. Introdução

Serviços de Mensagens Instantâneas (MI), também conhecidos por *Instant Messaging*, permitem o envio e o recebimento de mensagens de texto em tempo real com um baixo custo. No Brasil, a troca de MI concentra o principal uso de celulares e *smartphones*. Dados do grupo de consultoria de informações Kantar Worldpanel indicaram que 83,3% dos lares brasileiros utilizaram aplicativos de MI em 2016 [Rossi 2017]. A interatividade fornecida por este serviço, como as salas de bate-papo (do inglês *chat rooms*), ocasiona o seu uso intensivo entre crianças e adolescentes (aqui referenciados como jovens).

Semelhantes ao MI, os Serviços de Mensagens Curtas (SMS, do inglês *Short Message Service*) são amplamente usados e geram um volume expressivo de dados. De acordo com uma pesquisa realizada em 2018 pela Mobile Time e Opinion Box, 51% dos entrevistados brasileiros recebem SMS todos os dias e 24% das pessoas questionadas disseram enviar SMS diariamente. Aproximadamente 781 bilhões de mensagens de texto foram enviadas no mês de junho de 2017 nos Estados Unidos

[Statistic Brain 2017]. No Brasil, ainda que o SMS não seja o principal meio de comunicação de usuários domésticos, de acordo com uma pesquisa do Instituto Brasileiro de Opinião e Estatística (Ibope) cerca de 10 milhões de brasileiros o utilizam [Costa 2017].

As salas de bate-papo SMS podem ser acessadas por qualquer pessoa, incluindo jovens. Segundo a associação civil sem fins lucrativos SaferNet, a Central Nacional de Denúncias de Crimes Cibernéticos recebe em média 2.500 denúncias por dia contendo evidências de crimes de Pornografia Infantil ou Pedofilia[1] dentre outros [SaferNet 2008]. Desta maneira, a garantia da segurança apresenta-se como um desafio a ser atingido. O **problema tratado neste trabalho** é a escassez de estratégias para identificar potenciais perfis mal-intencionados dentro do processo de garantia da segurança dos usuários em salas de bate-papo SMS. No contexto do SMS, acompanhar como se dá a utilização deste serviço é de interesse das operadoras de telefonia móvel. O estudo do comportamento de seus usuários pode ser utilizado para evoluções no serviço ofertado, como na melhoria dos mecanismos de controle e no nível de segurança.

O **objetivo deste artigo** é analisar o comportamento dos usuários e propor uma estratégia para identificação de perfis mal-intencionados em salas de bate-papo SM**S**. O trabalho será realizado a partir de um serviço de mensagens curtas de uma operadora de telefonia móvel que atua no Brasil. Este estudo é baseado na análise das interações e agrupamento das informações obtidas, o que permitirá a identificação de perfis distintos de usuários. Propõe-se também um método para classificação manual das mensagens de texto, visando à identificação de mensagens com conteúdo mal-intencionado. Por meio desta classificação, analisa-se a relação entre as mensagens de texto rotuladas como mal-intencionadas e o perfil do usuário responsável pelo seu envio. O estudo desta relação permite a sinalização de um potencial risco à maliciosidade a partir de um dado perfil de usuário no serviço de bate-papo SMS.

A identificação de perfis mal-intencionados tem o intuito de contribuir para a detecção de conteúdo impróprio. E, desta maneira, colaborar na previsão da vulnerabilidade ao assédio no serviço de mensagens curtas antes que o mesmo ocorra, promovendo maior retenção do cliente ao serviço da empresa prestadora mediante a melhoria da segurança ofertada. Cabe ressaltar que há um interesse crescente na detecção de conteúdo impróprio [Schmidt 2017] e que são poucos os estudos e ferramentas que se dedicam à análise de conversas em serviços de bate-papo SMS [Yu et al. 2015] de modo a auxiliar à descoberta de distintos perfis e na identificação da conversação que estes usuários estabelecem entre si.

A análise descritiva do serviço de bate-papo SMS evidenciou que o maior volume de mensagens é enviado no domingo e na terça-feira e que há altos índices de utilização durante a noite. Além disso, as salas de bate-papo mais acessadas são as que o tema envolve relacionamentos e os usuários geralmente se comunicam por meio de Mensagens de Sala - mensagens direcionadas a um destinatário específico. Os resultados do algoritmo de agrupamento K-means apontaram a existência de cinco

[1] Transtorno em que um adulto possui fantasias com foco em jovens como parceiros sexuais [Lanning et al. 2010].

agrupamentos distintos de usuários quanto à utilização do serviço, denominados Infrequente, Regular, Moderado, Frequente e Hiperativo. Por meio do método de caracterização manual do conteúdo discutido nas salas de bate-papo, os Usuários Frequentes e Hiperativos foram identificados como os perfis que poderiam possuir um acompanhamento específico ao utilizarem as salas de bate-papo SMS, uma vez que são os mais propensos a enviarem mensagens com conteúdo Mal-intencionado (respectivamente, 30,23% e 25%).

O presente trabalho está estruturado em 6 seções, organizado como descrito a seguir: a seção 2 apresenta o referencial teórico, onde destacam-se os conceitos relevantes para o estudo. A seção 3 descreve os principais trabalhos relacionados à caracterização de perfis de usuários. Na seção 4 discutem-se os materiais e métodos empregados na solução do problema proposto. A seção 5 apresenta uma análise dos resultados alcançados. E, por fim, na seção 6 é apresentada a conclusão geral e as contribuições desta pesquisa, além de possíveis trabalhos futuros.

2. Referencial Teórico

Nesta seção são descritos os principais conceitos e técnicas no contexto de comportamento mal-intencionado, identificação e caracterização de perfis de usuários. O propósito é discutir sobre tais conceitos destacando-se aspectos que são relevantes para compreensão da pesquisa conduzida neste trabalho.

2.1. Comportamento Mal-intencionado

De acordo com Saquib e Ali (2015), grande parte dos usuários conectados em redes sociais e em serviços de comunicação executam atividades normais, entretanto, determinados usuários praticam atos incomuns e suspeitos, considerados comportamentos *mal-intencionados* ou *maliciosos*. Os mais vulneráveis a tais comportamentos são os menores de idade. Geralmente, o contato é realizado por mensagens em salas de bate-papo, nas quais os jovens são vítimas de pedófilos que mantêm conversas de cunho sexual [Bogdanova et al. 2012].

O comportamento mal-intencionado no ambiente *online*, especificamente em serviços de conversa por SMS, também inclui ações antiéticas executadas para manipular o processo de raciocínio de outros usuários, satisfazendo assim interesses pessoais de um determinado indivíduo. Tal comportamento inclui o envio de mensagens insultuosas de cunho violento e difamador, ameaças, perseguições e uso de linguagem ofensiva e imprópria no conteúdo veiculado [Saquib e Ali 2015; Oliveira et al. 2015]. O estudo do comportamento mal-intencionado de um usuário pode ser realizado nas seguintes categorias: análise de conteúdo e detecção de usuário mal-intencionado [Saquib e Ali 2015]. Esta pesquisa aborda a rotularização do conteúdo das salas de bate-papo e a detecção de usuários mal-intencionados que realizam as ações relatadas anteriormente.

2.2. Perfis de Usuários

Perfis podem ser utilizados para identificar pessoas com comportamentos em comum [Benevenuto et al. 2011]. E, para parte dos usuários, existe uma relação entre a

identidade do indivíduo real e seu perfil no meio digital [Boyd 2007]. Neste contexto, há distintos perfis e, para a garantia da segurança dos usuários, é fundamental obter uma compreensão do padrão de acesso dos perfis mal-intencionados [Yu et al. 2015].

A identificação de padrões de comportamento dos usuários é uma análise obtida por meio da extração e tratamento das informações [Benevenuto et al. 2011], dentre as técnicas comumente utilizadas destaca-se a técnica de mineração de dados denominada análise de agrupamento ou clusterização (do inglês *clustering*). Organizar dados em grupos é um dos principais modos de compreensão e aprendizagem [Jain e Dubes 1988], cujo objetivo é encontrar uma organização conveniente e válida dos dados. Esta técnica consiste em separar objetos em grupos, de maneira que em um determinado grupo os objetos serão similares, de acordo com alguns atributos pré-determinados [Linden 2009]. Os mecanismos de identificação examinam as múltiplas dimensões do comportamento do usuário e realizam a composição de tais características com indicadores inter-relacionados [Ferrara et al. 2016; Boshmaf et al. 2012].

2.3. Identificação de Perfis de Usuários por Agrupamento

A adoção do agrupamento para caracterização do perfil de usuários por meio de grupos significativos, consiste no processo de agrupamento de dados tal que os objetos dentro de um grupo possuem um alto grau de semelhança entre eles e uma grande diferença em relação aos objetos de outros grupos. O grau de similaridade entre os perfis de usuários de um grupo é medido e quanto mais perto os indivíduos são mais semelhantes, no entanto quanto maior a distância menor é a similaridade dos indivíduos.

Existem diferentes algoritmos de agrupamento disponíveis na literatura, dentre eles, pode-se destacar o algoritmo *K-means*. O algoritmo de clusterização *K-means* utiliza o conceito de centroide, que é um ponto imaginário no espaço n-dimensional que reúne as propriedades médias de um determinado grupo [Jain et al. 1999]. O algoritmo *K-means* possui como entrada o número de grupos desejados (k) e particiona o conjunto de n objetos em k grupos, de forma que a similaridade intergrupo seja baixa e a similaridade intragrupo seja alta, sendo que esta é avaliada considerando o valor médio dos objetos do grupo (centroide). Assim, cada objeto pertence ao grupo do centroide mais próximo a ele.

O funcionamento do algoritmo se dá pela seguinte execução iterativa [Castro e Ferrari 2016]: os k centroides iniciais dos grupos são determinados aleatoriamente, em seguida, calcula-se a distância entre os objetos da base e cada um dos centroides e atribui-se cada objeto ao centroide mais próximo. Como dito anteriormente, os novos centroides são calculados tomando-se a média dos objetos pertencentes a cada centroide promovendo um reposicionamento dos centroides e uma nova alocação de objetos a grupos. O algoritmo converge quando não há mais alterações nos centroides e mudanças nas alocações de objetos aos grupos, assim o algoritmo *K-means* terá minimizado o erro quadrático calculado entre as instâncias e os centróides dos grupos.

3. Trabalhos Relacionados

Nesta seção são apresentados trabalhos relacionados a caracterização de usuários por meio de técnicas de agrupamento e a identificação de conversas com indícios de

comportamento mal-intencionado em aparelhos móveis. As pesquisas a seguir apresentam diversas informações que nortearam a abordagem proposta neste trabalho.

Oliveira et al. (2015) propõem e avaliam uma estratégia para determinar se os usuários apresentam comportamento malicioso (envio de mensagens com conteúdo impróprio para menores, uso de linguagem ofensiva, mensagens de cunho violento ou difamador e spammers) a partir de perfis de navegação e utilização de um serviço de MI. Esse serviço baseado em SMS foi caracterizado e para a descoberta dos perfis distintos de usuários os autores utilizaram o algoritmo de agrupamento *X-means*. Analisando os padrões de comportamento dos usuários foi possível identificar a existência de quatro grupos distintos de usuários. E, de maneira particular, aprofundou-se o estudo sobre o comportamento dos usuários de um determinado grupo. Os resultados do estudo apontaram que os usuários deste grupo apresentam um perfil de utilização do serviço extremo, enviando um alto volume de mensagens e também com um elevado número de acessos. Também foi possível identificar que os usuários deste grupo possuem um foco claro de utilização do serviço: relacionamento. Para a caracterização de suas mensagens foi realizada a submissão dos textos a um classificador manual para formação da base de treino e posteriormente ao classificador *Support vector machine*. Com o método proposto os autores atingiram uma precisão de 91% na classificação das mensagens.

Benevenuto et al. (2012) apresentam os resultados de uma análise que aborda a caracterização do comportamento de usuários em redes sociais. De maneira distinta a estudos que reconstruíram as ações do usuário a partir de artefatos "visíveis" como comentários e depoimentos, o estudo foi baseado em dados do fluxo de cliques de 37.024 usuários que acessaram ao longo de 12 dias as redes sociais Orkut, MySpace, Hi5 e LinkedIn. O objetivo do estudo foi aumentar a eficiência das redes sociais estudadas e melhorar a experiência dos usuários. A análise dos autores revelaram que os principais recursos utilizados nas redes sociais são a frequência e tempo de conexão dos usuários às redes sociais, além dos tipos e sequências de atividades que os usuários realizam. As interações sociais foram classificadas em atividades publicamente visíveis e atividades silenciosas. Os resultados do estudo destacaram a presença de ações "silenciosas" do usuário, como navegação em páginas de perfis ou visualização de fotos de um amigo.

Bogdanova et al. (2012) abordam o problema da detecção automática de pedófilos *online* com técnicas de processamento de linguagem natural. A proposta é investigar se recursos baseados em emoções auxiliam na detecção. Considerando que a expressão de certas emoções no texto podem ser úteis para identificação de pedófilos nas mídias sociais, os autores sugerem uma lista de recursos de alto nível baseada em sentimentos e emoções. Os resultados experimentais mostram que a tarefa de categorização de texto binário baseada em tais características discrimina os pedófilos de não pedófilos com alta precisão. A classificação de Naive Bayes (NB) com base nas características propostas pelos autores atingiu precisões de até 94%, enquanto que os recursos de baixo nível atingem apenas 72% de precisão nos mesmos dados.

Rodrigues et al. (2009) abordaram o problema de detecção de usuários mal-intencionados que postam vídeos com conteúdo impróprio como divulgação de propagandas e distribuição de pornografia. A partir dos atributos dos usuários e dos

vídeos do YouTube, os usuários foram classificados manualmente como legítimos, spammers e promotores de vídeos. Além disso, uma caracterização de aspectos de diferenciação entre os usuários desses grupos foi realizada e tais aspectos foram aplicados em um algoritmo de classificação. O mecanismo de detecção de usuários maliciosos foi capaz de identificar corretamente 97% dos promotores, 54% de spammers, errando apenas 5,4% de usuários legítimos. Por meio da simulação observou-se que uma pequena porcentagem de promotores de vídeos é capaz de aumentar a quantidade de conteúdo impróprio.

4. Metodologia

Nesta seção apresentam-se os materiais e métodos utilizados para atingir a solução do problema proposto. Como este estudo faz uso da quantificação no tratamento das informações o mesmo pode ser definido como quantitativo, no qual utiliza-se técnicas estatísticas objetivando resultados que evitem distorções de análise e interpretação [Dalfovo et al. 2008].

De um modo geral, a abordagem metodológica possui como suporte a análise de um serviço de bate-papo SMS. A partir dessa análise, é proposto um estudo das interações e agrupamento das informações obtidas para compreensão do comportamento dos usuários, além da caracterização manual do conteúdo discutido nas salas de bate-papo e, consequente, dos usuários que nelas interagem. Ao longo do estudo foram realizadas as etapas descritas nos próximos parágrafos.

4.1. Base de Dados

O conjunto de dados utilizado nesta pesquisa foi fornecido por uma operadora de telefonia móvel do Brasil e contém registros de mensagens de texto enviadas e recebidas por usuários de um serviço de bate-papo SMS, entre os dias 1º e 30 de setembro de 2017. A princípio, obteve-se o conhecimento necessário sobre o conjunto de dados disponibilizado.

```
Sessão:                  {9BCDEFAD-A036-4390-8662-0DAF13178707}
Apelido do remetente:    douglas01
Categoria:               55
Nome da categoria:       Alagoas
Mensagem de texto:       vamos tc no resevado
Tipo de evento:          3
Nome do evento:          SendPublicMessage
Data do envio:           2017-09-24 18:27:42
Tipo da sala:            1
Sala:                    24058
Nome da sala:            Alagoas 1
Apelido do destinatário: dhiana
```

Figura 1. Exemplo de tipos de dados e dados associados a um registro de mensagem de texto SMS

Na Figura 1 é exemplificada a maneira como uma mensagem de texto é registrada na base de dados do serviço. A seguir, são apresentadas as informações disponibilizadas para cada registro e os domínios de cada atributo do conjunto de dados:

- ***Sessão*:** identificador único da sessão (uma sessão é criada quando um usuário inicia a navegação no serviço de bate-papo SMS);
- ***Apelido do remetente*:** identificador do usuário responsável pelo envio da mensagem de texto;
- ***Categoria***: identificador da categoria (assunto) da sala de bate-papo;
- ***Nome da categoria***: identificador textual da categoria;
- ***Mensagem de texto***: mensagem de texto enviada pelo usuário (remetente);
- ***Tipo de evento***: classificação do tipo da mensagem de texto em *Mensagens de Sala* e *Mensagens Públicas*. Mensagens de Sala são direcionadas a um determinado usuário mas visualizadas por todos os participantes da sala de bate-papo SMS. Em contrapartida, Mensagens Públicas não são direcionadas a usuários específicos;
- ***Nome do evento***: identificador textual do tipo de evento;
- ***Data do envio***: data e hora em que a mensagem de texto foi enviada;
- ***Sala***: identificador único da sala de bate-papo SMS. As salas de bate-papo são classificadas por seus respectivos assuntos e organizadas em 63 categorias. Essas categorias foram organizadas para auxílio na análise:

  (*i*) *Localidade*: as mensagens de texto enviadas em salas desta categoria são relacionadas aos estados do país.

  (*ii*) *Pessoal*: as mensagens de texto são enviadas e recebidas em salas de bate-papo SMS criadas pelos próprios usuários do serviço.

  (*iii*) *Relacionamento*: as mensagens são trocadas em salas de bate-papo SMS nas quais o tema é paquera, balada, sexo, amizade, entre outros.

  (*iv*) *Assuntos Gerais*: as mensagens de texto enviadas nesse tipo de sala de bate-papo SMS são relacionadas a religião, esporte, música ou qualquer assunto que não se enquadre nas categorias acima.
- ***Nome da sala***: identificador textual da sala de bate-papo SMS;
- ***Apelido do destinatário***: identificador do usuário que recebe a mensagem de texto.

4.2. Pré-processamento dos Dados

O objetivo das técnicas de pré-processamento é o preparo dos dados brutos para serem analisados. Desta maneira, esse processo foi necessário para correção de problemas presentes no conjunto de dados e que necessitavam de tratamento específico [Castro e Ferrari 2016]. Neste estudo foram realizadas a manipulação e a transformação dos dados de maneira que o conhecimento passasse a ser obtido de maneira ágil e correta.

Dentre os métodos do processo de preparação da base de dados, foram executadas a limpeza e a transformação dos dados brutos. A **limpeza dos dados** foi realizada para correção de registros duplicados, extração de caracteres especiais e registros com incompletude. Inicialmente, o conjunto de dados aglomerava 7.640.079

registros e, após a realização da limpeza dos dados, 207.618 registros foram removidos. Restando 7.432.461 registros no conjunto de dados.

Em seguida, os dados foram submetidos ao processo de **normalização** para serem transformados com o objetivo de torná-los apropriados à aplicação do algoritmo de agrupamento *K-means* (conforme especificado na seção 2.2). A técnica de normalização utilizada foi a *Max-Min* no intervalo de [0, 1], responsável por realizar uma transformação linear nos dados brutos a partir da Equação 1 [Castro e Ferrari 2016]. De maneira que max_a e min_a são, respectivamente, os valores máximo e mínimo de determinado atributo; e um valor a em um valor a' é mapeado no domínio.

$$a' = \frac{a - min_a}{max_a - min_a} \quad (1)$$

4.3. Análise Descritiva dos Dados e Técnica de Agrupamento para Identificação dos Perfis de Usuários

A análise descritiva dos dados permite organizar, resumir e descrever os aspectos importantes de um conjunto de dados por meio de gráficos e tabelas [Castro e Ferrari 2016]. A descrição dos dados também tem como objetivo identificar anomalias e dados dispersos que não seguem a tendência geral do restante do conjunto de dados. Esse processo foi realizado utilizando-se medidas de resumo (tais como média, mediana e desvio padrão) e visualização dos dados com o intuito de simplificar, sumarizar e descrever as principais características da base de dados fornecida pela operadora de telefonia móvel. Formando, assim, uma base de análise quantitativa dos dados.

Para auxiliar na resolução do problema proposto, fez-se a caracterização do comportamento dos usuários e a identificação de distintos perfis de usuários com o emprego de técnicas de agrupamento por meio do algoritmo *K-means*. Uma das deficiências do algoritmo de agrupamento *K-means* é o fato que o número de grupos a serem identificados deve ser fornecido previamente. Diante desta situação, fez-se necessária a aplicação do método Elbow com o intuito de identificar o número ideal de grupos a ser fornecido. O método Elbow funciona de maneira a aumentar gradativamente o número de grupos e analisar o resultado de cada incremento. No momento em que o benefício do incremento estagnar ou deixar de ser relevante, consequentemente, a diferença da distância Euclidiana será insignificante. Neste ponto a quantidade de grupos é a ideal para segmentar os dados, ou seja, a adição de outro grupo não irá gerar um agrupamento melhor para os dados.

4.4. Classificação Manual das Mensagens

O desenvolvimento do software para a classificação manual das mensagens de texto teve como objetivo a realização de uma análise quantitativa da relação entre as mensagens de texto rotuladas com conteúdo mal-intencionado e o perfil do usuário responsável pelo seu envio. A análise desta relação consiste em uma estratégia que permite a sinalização de um risco em potencial a partir do perfil de um usuário mal-intencionado.

O software (disponibilizado em: http://classificadormanual-com.umbler.net/) apresenta as mensagens de texto a serem rotuladas por voluntários, no entanto, os demais dados citados na seção 4.1 foram anonimizados. Três voluntários receberam

orientações sobre a utilização do software e, além disso, houve o nivelamento sobre o conceito de Maliciosidade adotado na metodologia desta pesquisa (conforme discutido na seção 2.1). Cabe ressaltar que cada voluntário realizou a rotularização das mensagens de texto individualmente. Além disso, cada mensagem foi classificada três vezes por voluntários distintos e, como demonstrado na Figura 2, a partir das seguintes premissas:

- Cada voluntário possui duas possibilidades de rotularização da mensagem de texto: "*Malicioso*" ou "*Não Malicioso*";
- Se três (3) voluntários rotularam uma determinada mensagem de texto com *Maliciosa*, a mesma será considerada **Maliciosa**;
- Se dois (2) indivíduos rotularam uma determinada mensagem de texto com *Maliciosa*, a mesma será considerada **Duvidosa**;
- Se um (1) indivíduo rotular uma determinada mensagem de texto com *Maliciosa*, a mesma será considerada **Não Maliciosa;**
- E, por fim, se nenhum indivíduo rotular uma determinada mensagem de texto com *Maliciosa*, a mesma também será considerada **Não Maliciosa**.

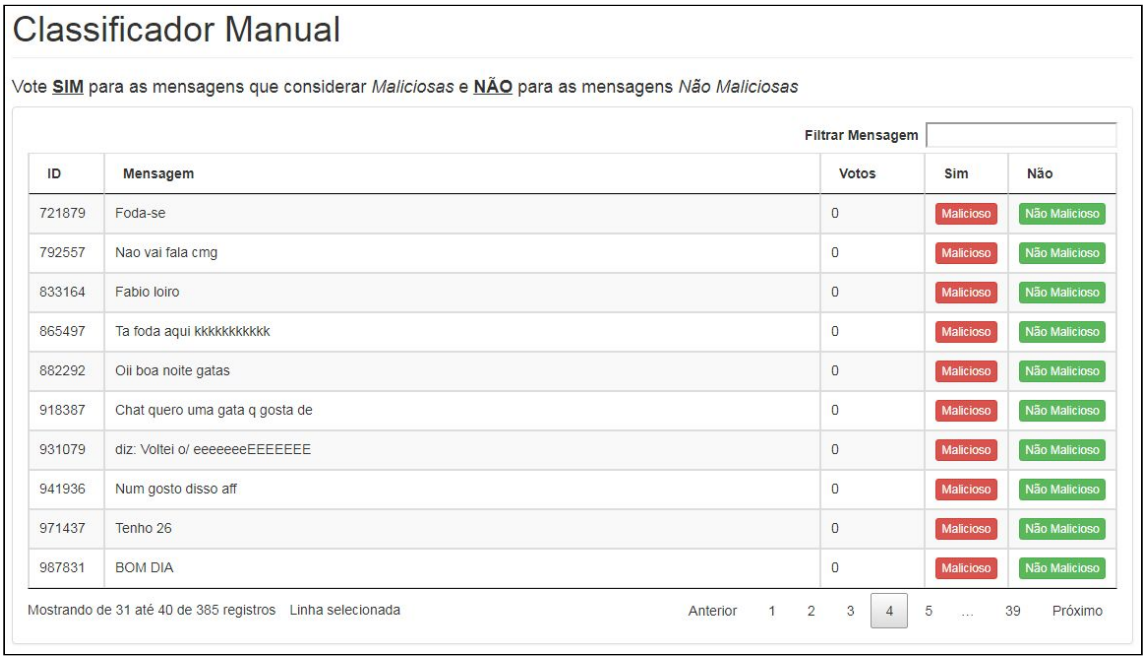

Figura 2. Software Utilizado na Classificação Manual das Mensagens

Para a identificação de potenciais perfis mal-intencionados foi necessária a rotularização das mensagens de texto e, respectivamente, a análise do perfil do usuário responsável pelo seu envio. Como o conjunto de dados utilizado nesta pesquisa possui um alto volume de registros (cerca de 7.432.461 mensagens de texto), a técnica de amostragem foi utilizada para se determinar a quantidade mínima de registros capaz de representar o volume total de mensagens. Assim, aplicando-se à Equação 2 [Castro e Ferrari 2016], definiu-se que uma amostra de no mínimo 664 mensagens de texto seria necessária para representar o conjunto de dados, considerando um valor de confiança de 99% e erro amostral de 5%. Desta maneira, foram selecionadas aleatoriamente 866 mensagens de texto para serem rotuladas de acordo com o seu conteúdo.

$$n = \frac{N . Z^2 . p . (1-p)}{Z^2 . p . (1-p) + e^2 . (N-1)} \quad (2)$$

Onde:

- ***n***: quantidade mínima de elementos da amostra.
- ***N***: população total.
- ***Z***: variável normal padronizada associada ao nível de confiança. Para esta pesquisa considerou-se 99% (Z = 2,575).
- ***p***: população proporcional de elementos que pertence à categoria que se deseja estudar (assumiu-se o valor de 0,5).
- **(*1 - p*)**: população proporcional de elementos que não pertence à categoria que se deseja estudar (assumiu-se o valor de 0,5).
- ***e***: margem de erro amostral (considerada 5%).

5. Resultados

Nesta seção são apresentados os resultados obtidos com os experimentos realizados. A princípio é apresentada a análise da utilização do serviço de bate-papo SMS. Em seguida, são descritos os resultados da identificação e da análise dos perfis de usuários do serviço estudado. E, por fim, é apresentada a estratégia para a identificação de potenciais perfis mal-intencionados do serviço de bate-papo SMS.

5.1. Análise da Utilização do Serviço de Bate-papo SMS

As informações geradas a partir da análise descritiva da utilização do serviço de bate-papo SMS serão apresentadas nesta seção por intermédio de gráficos e diagramas. A apresentação dos dados usando técnicas de visualização tem o objetivo de se entender a natureza da distribuição dos dados, desta maneira, o conhecimento passa a ser extraído mais facilmente.

A base de dados utilizada nesta pesquisa reúne 7.432.461 mensagens de texto enviadas e recebidas por 38.967 usuários, em 322.747 sessões distintas criadas no serviço. Observou-se que em média 247.749 mensagens foram enviadas por dia no mês de setembro de 2017. A Figura 3 apresenta uma visualização da troca de mensagens em uma perspectiva semanal, como pode ser observado os dias com maior volume de mensagens são domingo e terça-feira com 256.220 e 252.720 mensagens de texto enviadas, respectivamente.

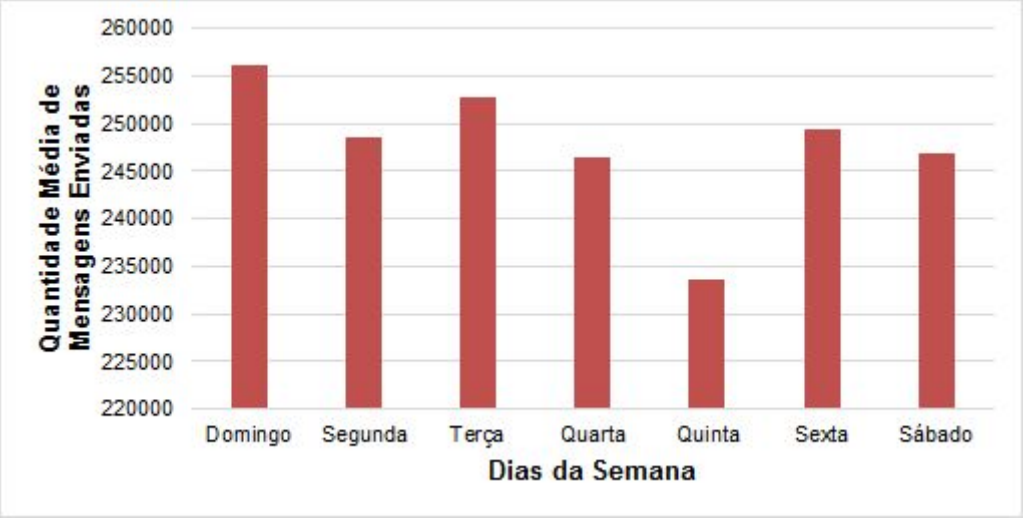

Figura 3. Quantidade Média de Mensagens Enviadas por Dia da Semana

A partir da Figura 4 pode-se observar que o volume de mensagens nas salas de bate-papo cuja categoria é Relacionamento representa 46% de todas as mensagens do período observado. E apenas 16% das mensagens de texto foram enviadas em salas de bate-papo SMS cuja a categoria é Pessoal.

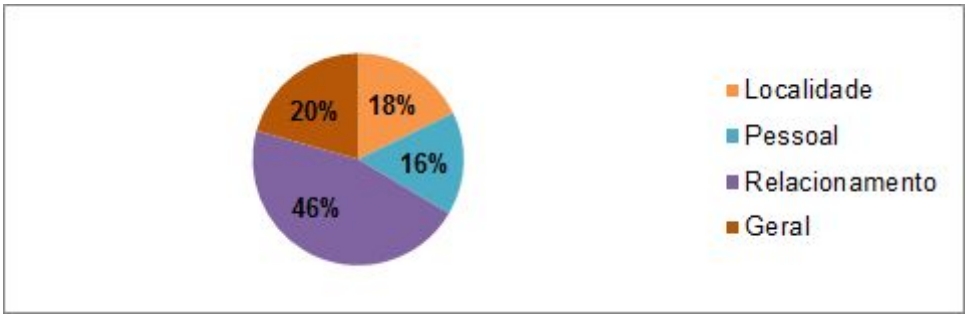

Figura 4. Mensagens Enviadas por Categoria

A Figura 5 permite visualizar que altos índices de utilização são comuns durante a noite, entre 18:00hrs e 23:59hrs. Neste intervalo de tempo, acontecem cerca de 35% de todas as trocas de mensagens no serviço. Durante o período da tarde, o volume de troca de mensagens também é significante, correspondendo a 30%.

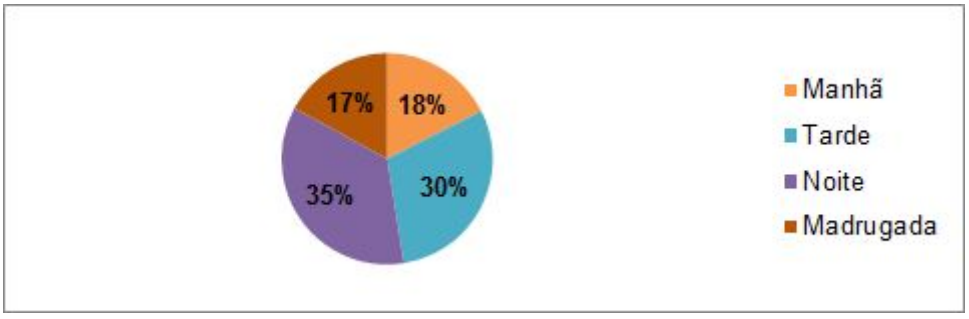

Figura 5. Mensagens Enviadas ao Longo do Dia

Analisando a Figura 6, nota-se que em mais de 53,32% das sessões criadas pelos usuários as mensagens trocadas são exclusivamente Mensagens de Sala. Além disso, em

42,33% das sessões as mensagens de texto trocadas são do tipo Mensagens de Sala e Mensagens Públicas. Tais comportamentos sugerem que os usuários geralmente se comunicam com mensagens direcionadas a um destinatário específico.

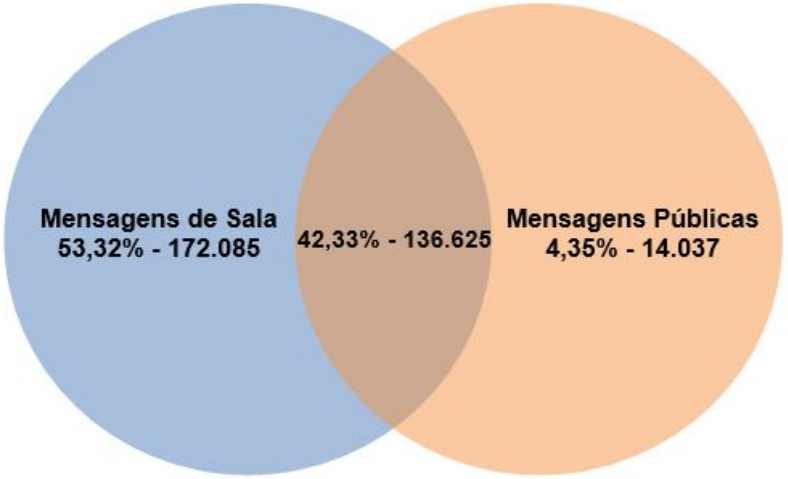

Figura 6. Número de Sessões Criadas pelos Usuários por Tipo de Mensagem

5.2. Análise dos Perfis de Usuários do Serviço de Bate-papo SMS

Com o intuito de identificar perfis de usuários, foram definidas e levantadas três métricas na interação entre os participantes do serviço para serem utilizadas como entrada do algoritmo de agrupamento *K-means*. Sendo elas:

I. ***Mensagens/sessão***: quantidade de mensagens de texto enviadas pelo usuário por sessão.
II. ***Sessões/semana***: número de sessões criadas pelo usuário por semana.
III. ***Mensagens/minuto***: frequência de envio das mensagens de texto por minuto.

A Tabela 1 apresenta a média, mediana e desvio padrão para cada métrica utilizada no processo de agrupamento dos dados. Tais valores correspondem à medidas de tendência central e de dispersão, e consistem em valores típicos que foram considerados na análise dos grupos de usuários. Além disso, como representada na Figura 7, a matriz de gráficos de dispersão permite avaliar a relação entre os pares de métricas em um mesmo instante.

Tabela 1. Medidas de Tendência Central e Dispersão das Métricas Utilizadas no Agrupamento

Métricas	Dados Não Normalizados			Dados Normalizados		
	Média	Mediana	Desvio Padrão	Média	Mediana	Desvio Padrão
Mensagens/sessão	15.850	9,000	21.621	0.016	0,008	0.023
Sessões/semana	2.095	0,500	4.757	0.021	0,002	0.053
Mensagens/minuto	0.709	0,559	0.618	0.017	0,013	0.015

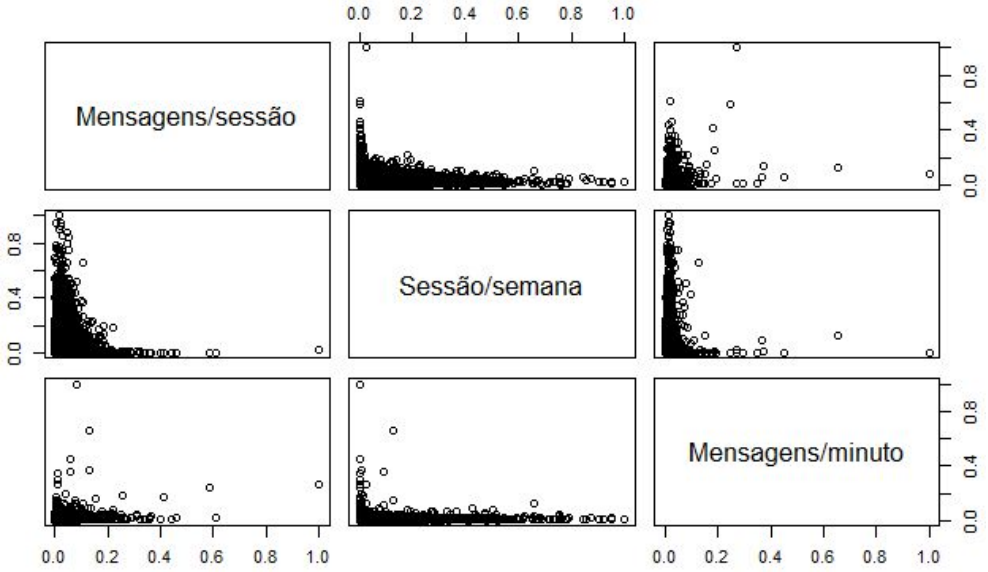

Figura 7. Matriz de Gráfico de Dispersão para as Métricas Utilizadas no Agrupamento

A partir da Figura 8, pode-se observar que quando o número de grupos está no intervalo entre 1 e 4, por exemplo, há uma grande variação entre a soma dos quadrados. No entanto, quando o número de grupos é equivalente a 5 observa-se que o ganho marginal é estagnado e a soma dos quadrados cai abruptamente para 35,48. Desta maneira, evidencia-se que 5 é o número de grupos apropriado para ser fornecido ao algoritmo de agrupamento *K-means*.

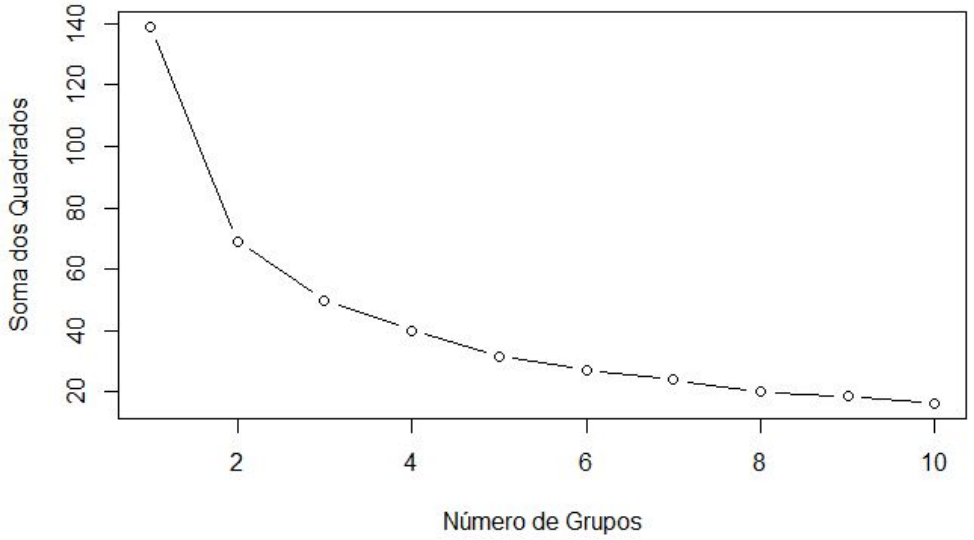

Figura 8. Soma dos Quadrados x Número de Grupos

Os resultados do algoritmo de agrupamento *K-means* apontaram a existência de agrupamentos distintos de usuários quanto à utilização do serviço de bate-papo SMS, denominados "**Infrequente**", "**Regular**", "**Moderado**", "**Frequente**" e "**Hiperativo**". A Tabela 2 mostra os cinco grupos e seus respectivos centros, além do percentual de usuários em cada grupo e a Figura 9 apresenta os valores das métricas normalizadas para cada um dos perfis de usuários identificados.

Tabela 2. Detalhes dos Grupos de Usuários Identificados

Grupos		Usuários (%)	Dados Não Normalizados			Dados Normalizados		
			Mensagens /sessão	Sessões/ semana	Mensagens /minuto	Mensagens /sessão	Sessões/ semana	Mensagens /minuto
1	Infrequente	8,00	4,55	0,34	2,11	0,004	0,001	0,052
2	Moderado	75,00	10,69	1,06	0,57	0,065	0,018	0,017
3	Regular	9,00	62,96	1,81	0,71	0,010	0,009	0,014
4	Frequente	7,00	23,91	11,60	0,54	0,024	0,127	0,013
5	Hiperativo	1,00	27,15	35,81	0,65	0,028	0,398	0,016

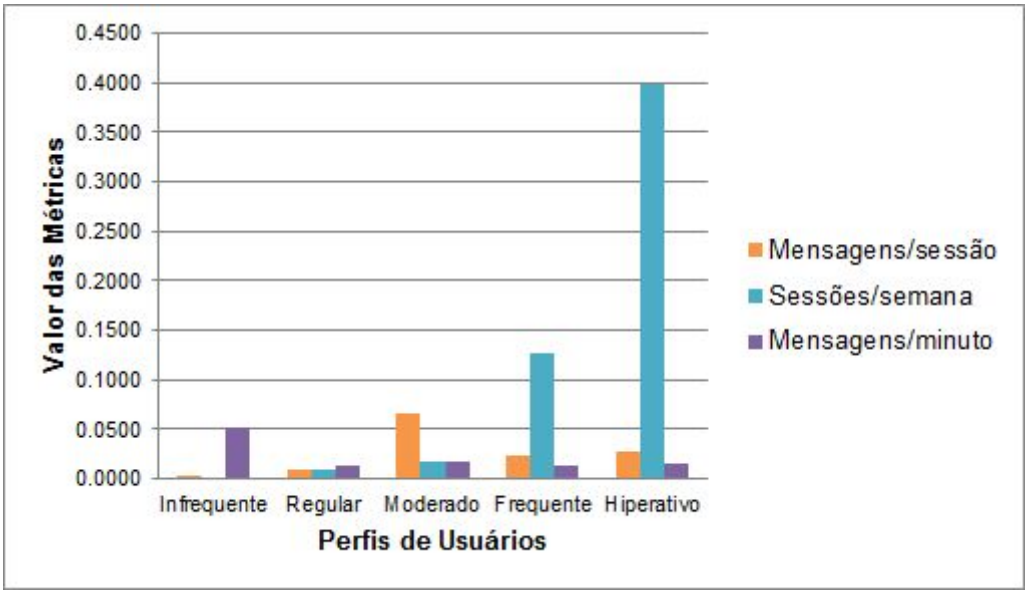

Figura 9. Visão Geral dos Grupos de Usuários

Cerca de 3.194 usuários do serviço (8%) encontram-se no grupo 1 da Tabela 2. Além disso, indivíduos deste grupo enviam em média 4,55 mensagens por sessão, aproximadamente três vezes menos que a média da métrica. Esses usuários também possuem a menor média de utilização do serviço, acessando-o 0,34 vezes por semana. No entanto, a frequência no envio de mensagens neste grupo é a maior do serviço. Cerca de 2 mensagens por minuto são enviadas, frequência três vezes maior que a média da métrica. No mês de setembro de 2017 tais usuários foram responsáveis pela menor quantidade de mensagens enviadas no serviço, apenas 1%. O envio de mensagens por este grupo de usuário, segundo a Figura 10, predomina em categorias de

Relacionamento e Assuntos Gerais, respectivamente. Frente a tais características, este grupo de usuário é identificado como "**Infrequente**".

O grupo 2 da Tabela 2 é o mais representativo, uma vez que é composto por 29.169 usuários (75%). Devido a sua numerosidade, tais indivíduos foram responsáveis pelo envio de 22% das mensagens setembro de 2017. No entanto, os mesmos não utilizam o serviço mais de 1 vez por semana. Além disso, este grupo envia uma pequena quantidade de mensagens, em média 10,69 por sessão e apenas 0,57 mensagens por minuto. O envio de mensagens por integrantes deste grupo é voltado às categorias de Relacionamento e Assuntos Gerais. No entanto, de maneira distinta ao Usuário Infrequente, o menor número de mensagens é enviado na categoria Pessoal - como pode ser observado na Figura 10. As métricas utilizadas no processo de clusterização referentes a este grupo, não apresentam valores destoantes entre si, mas valores intermediários se comparados às médias das métricas. Dessa forma, os integrantes desta classe de usuários são classificados como "**Moderados**".

No grupo 3 da Tabela 2 tem-se 3.443 dos usuários (9%). Cabe ressaltar que tais usuários são responsáveis pelo volume mais significativo de mensagens por sessão, em média 62,96 envios (valor cerca de 4 vezes maior que a média da métrica). Além disso, acessam o serviço apenas 1,81 vezes por semana e possuem a frequência de envio de mensagens acima da média, equivalente a 0,71. Os integrantes deste grupo foram responsáveis pelo envio de 19% das mensagens em setembro de 2017. Além disso, similar ao comportamento do Usuário Regular, o envio de mensagens predomina nas categorias de Relacionamento e Assuntos Gerais - como pode ser observado na Figura 10. Por este comportamento, o grupo de usuário foi nomeado como "**Regular**".

O grupo 4 da Tabela 2 aglomera 2.486 usuários da base (7%) e apresenta a menor frequência de envio de mensagens do serviço, cerca de 0,54 mensagens por minuto. Com relação às métricas mensagens enviadas por sessão e sessões criadas por semana, em ambas o grupo se apresenta acima da média. Tais usuários enviam 23,91 mensagens por sessão e, geralmente, utilizam o serviço 12 vezes por semana. Cabe ressaltar que os indivíduos deste grupo foram responsáveis pelo maior envio de mensagens no serviço em setembro de 2017, cerca de 38% de todas as mensagens trocadas. Além disso, como é observado na Figura 10, o envio de mensagens por este grupo predomina em salas de bate-papo cuja categoria é Relacionamento e, de maneira distinta a todos os outros grupos identificados, também em salas Pessoais. Por tais características esse grupo de usuário é nomeado de "**Frequente**".

Por fim, no grupo 5 da Tabela 2 tem-se 389 dos usuários (1%). Tais usuários enviam cerca de 27 mensagens por sessão, valor acima da média da métrica. A frequência média de envio de mensagens para este grupo é baixa (0,65 mensagens por minuto). Contudo, esses usuários são os que utilizam o serviço de maneira mais intensa e apresentam o maior número de sessões criadas por semana (por volta de 35,81). Apesar de ser o menor grupo, em setembro de 2017 foi responsável pelo envio de 20% das mensagens do serviço. Assim como os usuários moderados, o maior envio de mensagens se concentra nas categorias de Relacionamento e Assuntos Gerais, e apresenta o menor número de mensagens na categoria Pessoal - como pode ser observado na Figura 10. A atividade semanal desse grupo é extremamente alta se

comparada aos demais e essa característica justifica o nome que a classe de usuários recebeu de "**Hiperativo**".

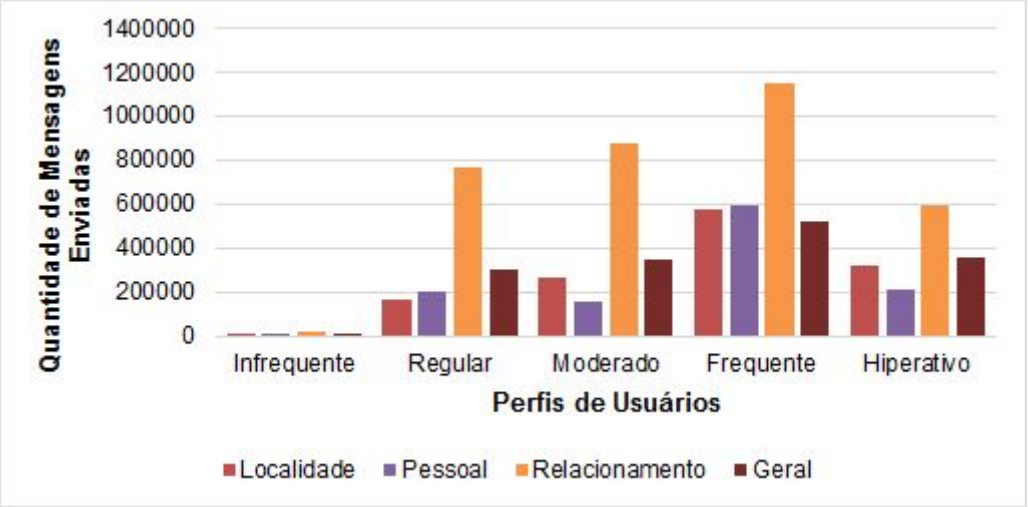

Figura 10. Quantidade de Mensagens Enviadas por Categoria em Cada Grupo

5.3. Potenciais Perfis Mal-intencionados do Serviço de Bate-papo SMS

Todos os perfis de usuários identificados na seção 5.2 possuíram representatividade na amostra extraída, de maneira a corresponder a seguinte distribuição: Usuário Infrequente (6%), Usuário Regular (12%), Usuário Moderado (15%), Usuário Frequente (45%) e Usuário Hiperativo (22%). De acordo com os dados da classificação manual realizada por três voluntários, grande parte das mensagens foram rotuladas como Não Mal-intencionadas (por volta de 72% da amostra). Além disso, 20% das mensagens de texto foram consideradas Mal-intencionadas e 8% das mensagens de texto foram rotuladas como Duvidosas.

A partir da Figura 11 pode-se observar a probabilidade de um dado perfil de usuário enviar uma mensagem com conteúdo Mal-intencionado, Duvidoso ou Não Mal-intencionado, calculada a partir da quantidade total de mensagens de texto enviadas por cada grupo de usuário. O perfil de usuário que possui a maior probabilidade (96%) de enviar mensagens Não Mal-intencionadas é o Usuário Infrequente. Pode-se observar também que o perfil com a menor probabilidade de enviar mensagens com conteúdo Mal-intencionado é o Usuário Moderado. Em contrapartida, o Usuário Frequente e Hiperativo são os mais propensos a enviarem mensagens com conteúdo Mal-intencionado - respectivamente, 30,23% e 25%. Além disso, o Usuário Hiperativo possui ainda a maior probabilidade de enviar mensagens de texto rotuladas como Duvidosas. Consequentemente, os Usuários Frequentes e Hiperativos poderiam ser submetidos a um acompanhamento específico ao utilizarem as salas de bate-papo SMS com a finalidade de tornar o ambiente mais seguro.

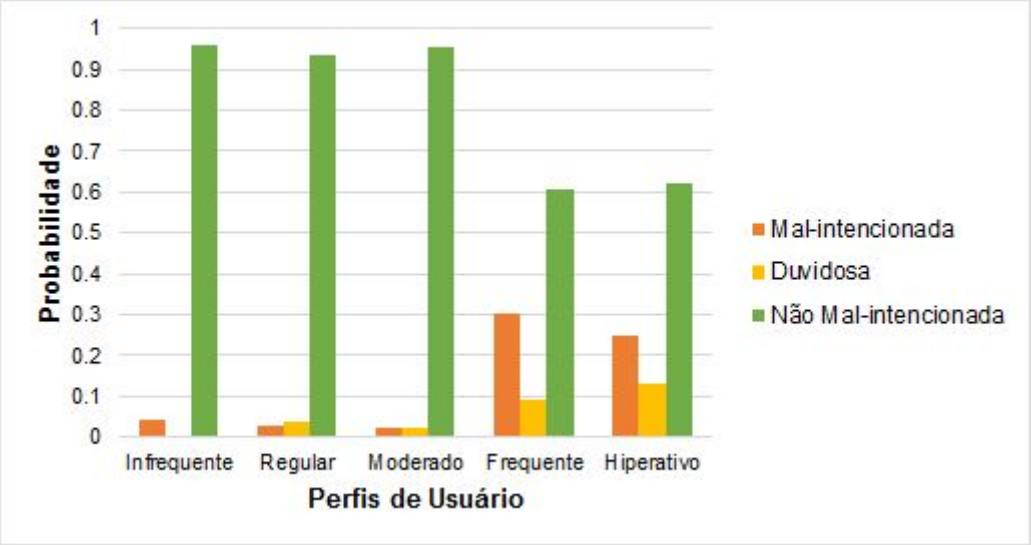

Figura 11. Probabilidade no Envio de Mensagens de Texto com Conteúdo Mal-intencionado, Duvidoso e Não Mal-intencionado

6. Conclusão

Este trabalho apresentou um estudo do comportamento dos usuários de um serviço de bate-papo SMS, no qual analisou-se um conjunto de dados fornecido por uma operadora de telefonia móvel do Brasil com aproximadamente 7 milhões de mensagens enviadas e recebidas por cerca de 39 mil usuários durante o mês de setembro de 2017. A partir da análise descritiva do serviço observou-se que o maior volume de mensagens é enviado no domingo e na terça-feira e que há altos índices de utilização durante a noite e a tarde. Além disso, observou-se que em salas de bate-papo cujo tema envolve relacionamentos são as mais acessadas e que os usuários geralmente se comunicam por meio de Mensagens de Sala - mensagens direcionadas a um destinatário específico.

Para compreensão do comportamento dos usuários, aprofundou-se o estudo das interações e o agrupamento dessas informações com o algoritmo *K-means*, a partir das seguintes métricas: quantidade de mensagens de texto enviadas por sessão (mensagens/sessão), número de sessões criadas por semana (sessões/semana) e frequência de envio das mensagens de texto por minuto (mensagens/minuto). Os resultados do algoritmo mostraram a existência de cinco perfis de usuários distintos quanto à utilização do serviço de bate-papo SMS, denominados: Infrequente (8%), Regular (75%), Moderado (9%), Frequente (7%) e Hiperativo (1%).

Por meio do método de caracterização manual do conteúdo discutido nas salas de bate-papo foi possível a sinalização de um risco em potencial a partir do perfil de um usuário mal-intencionado. Para a identificação desses perfis foi necessária a rotularização das mensagens de texto e, respectivamente, a análise do perfil do usuário responsável pelo seu envio. Desta maneira, os Usuários Frequentes e Hiperativos foram identificados como perfis que poderiam possuir um acompanhamento específico ao utilizarem as salas de bate-papo SMS do serviço, uma vez que são os mais propensos a enviarem mensagens com conteúdo Mal-intencionado (respectivamente, 30,23% e 25%).

Os resultados obtidos foram satisfatórios e a principal contribuição desta pesquisa é a identificação de distintos perfis de usuários e a detecção de perfis potencialmente mal-intencionados. Além disso, o conhecimento extraído por esta pesquisa pode ser facilmente utilizado pela operadora de telefonia móvel para evoluções no serviço ofertado e promoção de maior retenção do cliente ao seu serviço mediante a melhoria da qualidade e do processo de garantia da segurança.

Como possíveis evoluções da pesquisa podem-se destacar o aprimoramento do processo de agrupamento com a utilização de outros algoritmos de clusterização e a adoção de novas métricas, além da análise do comportamento dos perfis de usuários em perspectiva diária. Como trabalhos futuros pretende-se propor um método de classificação automática para a detecção de usuários que enviam mensagens com conteúdo mal-intencionado. Além disso, pretende-se realizar o desenvolvimento de uma ferramenta capaz de sinalizar um risco em potencial a partir do perfil de um determinado usuário. Os dados e códigos utilizados nesta pesquisa foram disponibilizados em <https://is.gd/y6Pg6u> para, assim, serem utilizados em eventuais evoluções deste estudo.

Referências Bibliográficas

Benevenuto, F., Almeida, J., Silva, A. (2011). Coleta e análise de grandes bases de dados de redes sociais online. *Jornadas de Atualização em Informática* (JAI), p. 11-57.

Benevenuto, F., Pereira, A., Rodrigues, T., Almeida, V., Almeida, J., Gonçalves, M. (2010). Avaliação do perfil de acesso e navegação de usuários em ambientes web de compartilhamento de vídeos. *Brazilian Symposium on Multimedia Systems and Web. (WebMedia)*, p. 149–156.

Benevenuto, F., Rodrigues, T., Cha, C. Almeida, V. (2012). Characterizing user navigation and interactions in online social networks. *Information Sciences*, v. 195, p. 1–24.

Bogdanova, D., Rosso, P., e Solorio, T. (2012). On the impact of sentiment and emotion based features in detecting online sexual predators. *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, p. 110 – 118, Jeju, Korea. Association for Computational Linguistics.

Boshmaf, Y., Muslukhov, I., Beznosov, K., Ripeanu, M. (2012). Design and analysis of a social botnet. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, v. 57, February, p. 556-578.

Boyd, D. (2007). Why youth (heart) social network sites: The role of networked publics in teenage social life. *MacArthur foundation series on digital learning–Youth, identity, and digital media volume*, p. 119-142.

Castro, L. N., Ferrari, D. G. (2016). Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações. Editora Saraiva, São Paulo, Brasil.

Costa, G. (2017). Administradores - O Portal da Administração. Em tempos de WhatsApp, 10 milhões de internautas no Brasil ainda preferem SMS. Disponível em:

http://www.administradores.com.br/artigos/cotidiano/whatsapp-e-o-app-preferido-de-92-milhoes-de-brasileiros-e-cerca-de-10-milhoes-nao-ouviram-o-gemidao-do-zap/106537/. Acessado em 05 de março de 2018.

Dalfovo, M., Lana, R., Silveira, A. (2008). Métodos quantitativos e qualitativos: um resgate teórico. *Revista Interdisciplinar Científica Aplicada*, Blumenau, v. 2, n. 4, p. 01-13. ISSN 1980-7031.

Ferrara, E., Varol, O., Davis, C., Menczer, F., Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, v. 59, n. 2.

Jain, A., Dubes, R. (1988). Algorithms for Clustering Data. Prentice-Hall. Englewood Cliffs.

Jain, A., Murty, M., Flynn, P. (1999). Data Clustering: A Review. *ACM Computing Surveys (CSUR)*.

Lanning, K. (2010). Child molesters: A behavioral analysis for professionals investigating the sexual exploitation of children. *National Center for Missing & Exploited Children with Office of Juvenile Justice and Delinquency Prevention*, Virginia, USA.

Linden, R. (2009). Técnicas de Agrupamento. *Revista de Sistemas de Informação da FSMA*, n. 4, p. 18-36.

Oliveira, R., Brandão, W., Marques-Neto, H. (2015). Characterizing User Behavior on a Mobile SMS-Based Chat Service. XXXIII *Brazilian Symposium on Computer Networks and Distributed Systems*, Vitoria, 2015, p. 130-139.doi: 10.1109/SBRC.2015.25.

Rodrigues, T., Benevenuto, F., Almeida, V., Almeida, J., Gonçalves, M. (2009). Detecting spammers and content promoters in online video social networks. *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, p. 620-627.

Rossi, D. (2017). KANTAR Brasil Insights. Troca de mensagens instantâneas é um dos principais usos do celular entre brasileiros. Disponível em: https://br.kantar.com/tecnologia/móvel/2017/troca-de-mensagens-instantâneas-é-um-dos-principais-usos-do-celular-entre-brasileiros-comtech/. Acessado em 28 de março de 2018.

SaferNet. (2008). Central Nacional de Denúncias de Crimes Cibernéticos. Disponível em: http://www.safernet.org.br/site/institucional/projetos/cnd. Acessado em 28 de fevereiro de 2018.

Saquib, S., Ali, R. (2015). Malicious Behavior in Online Social Network. *Proc. of the IEEE Workshop on Computational Intelligence: Theories Applications and Future Directions* (IEEE WCI 2015), December.

Schmidt, A. Wiegand, M. (2017). A survey on hate speech detection using natural language processing. *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*. Association for Computational Linguistics, Valencia, Spain, p. 1–10.

Statistic Brain. (2017). Text Message Statistics – United States. Disponível em: https://www.statisticbrain.com/text-message-statistics/. Acessado em 28 de fevereiro de 2018.

Yu, S., Wang, G., Zhou, W. (2015). Modeling Malicious Activities in Cyber Space. *IEEE Netw.*, v. 29, no. 6, p. 83-87.