

Análise semântica a partir de Sistemas Complexos na base de dados do Disque Denúncia

Walkir Alexandre Toscano de Brito

Doutorando em Informática do Programa de Pós-Graduação em Informática – PPGI, na Universidade Federal do Rio de Janeiro – UFRJ, na área de Sistemas Complexos Adaptativos. Mestre em Informática pela UFRJ e Estatístico formado pela Escola Nacional de Ciências Estatísticas – ENCE.

Claudia L. R. Motta

Professora do Programa de Pós-Graduação em Informática – PPGI, no Instituto Tércio Pacitti de Aplicações e Pesquisa Computacionais – NCE/UFRJ na Universidade Federal do Rio de Janeiro.

Zeca Borges

Coordenador e responsável pelo Disque Denúncia desde 1995, Engenheiro Civil formado pela Universidade Federal do Rio de Janeiro.

Resumo

Esta proposta parte da tentativa de enfrentamento à criminalidade no estado do Rio de Janeiro e do papel desempenhado pelo sistema de Disque Denúncia (DD), cujo trabalho com suas bases de dados de denúncias está aliado a novas tecnologias que propiciam melhorias para a área de inteligência no combate à criminalidade e ao crime organizado. Acreditamos na tendência mundial de utilização de novas tecnologias e serviços de segurança no combate e prevenção da criminalidade, buscando novas soluções para o melhor tratamento e entendimento dos dados coletados pela base de dados do DD. Propomos a utilização de Sistemas Complexos para mineração e análise da complexidade dos dados obtidos nas denúncias e subdenúncias de crimes ou delitos, a fim de identificar palavras, expressões e conexões entre os diferentes elementos relacionados a elas. Assim, trataremos neste trabalho apenas das denúncias e dinâmicas referentes a tráfico de drogas, drogas ilícitas, *modus operandi* e dinâmicas do tráfico e/ou atividades criminosas e suas ramificações e facções. Esperamos com os resultados contribuir para as tomadas de decisão nas esferas de segurança pública, com vistas ao enfrentamento da violência e criminalidade no Rio de Janeiro.

Palavras-Chave

Sistemas Complexos, semântica, inteligência, segurança pública, Disque Denúncia.

Introdução

Um setor que mundialmente vem buscando tecnologias mais eficientes para dar suporte às suas áreas de operações é o de segurança pública. Praticamente todas as instituições e/ou agências que atuam nessa área possuem, em sua estrutura, uma ou mais unidades de inteligência, dedicadas ao trato de grandes quantidades de dados. Atualmente, as atividades de inteligência, investigação, denúncias, delitos e ocorrências envolvem receber, armazenar e processar dados das mais diversas fontes (estruturadas e não estruturadas), nos seus dois ramos, Inteligência e Contrainteligência.

Nesse sentido, Paula *et al.* (2012) descrevem que a atividade de inteligência deve ser compreendida como um sistema complexo adaptativo, em que os processos de construção, produção e gestão da informação e do conhecimento possam otimizar sua organização e utilização na segurança pública em defesa do Estado, da sociedade e do cidadão. Para tanto, utiliza-se um conjunto de distintas fontes de entrada de dados, abertas ou fechadas, tais como: perfis em redes sociais, dados cadastrais de pessoas físicas e jurídicas, informações sobre pessoas, fotos, veículos, áudios obtidos mediante processos de interceptação legal, informações via *web*, aplicativos móveis, dentre outras tantas fontes disponíveis. São basicamente orientadas para a produção de conhecimentos necessários para subsidiar os tomadores de decisão na construção e execução de políticas de segurança pública e de suas ações de prever, prevenir, neutralizar e reprimir, utilizando e fomentando a transformação e a produção do conhecimento em inteligência desde que seja preciso, útil, oportuno e significativo.

No levantamento da literatura de trabalhos relacionados à temática aqui abordada, percebe-se o uso crescente de técnicas de *machine learning* para que, sem intervenção humana, seja possível classificar qualquer denúncia anônima utilizando apenas sua natureza textual (PINHO, 2017). Quando os dados são textuais e, portanto, dados não estruturados, pode-se dispor de métodos e modelos de Mineração de Textos (MT), que têm sido bastante explorados em muitos problemas em diversas áreas (classificação de documentos, extração de informação, sumarização, análise de sentimentos, dentre outros).

Motivação

Segurança pública é inegavelmente um tema que nos últimos anos tem adquirido cada vez mais importância em diferentes contextos, e falar dela, quando se trata do estado do Rio de Janeiro e de sua capital, torna-se ainda mais interessante e desafiador ao considerar os aspectos socioeconômicos e geográficos do problema da violência. Esses aspectos permitem aproximação e interação entre favela e asfalto, e o convívio cotidiano entre diferentes classes sociais.

Historicamente, desde a década de 1990, observa-se uma ascensão da criminalidade no Rio de Janeiro e uma crescente crise das instituições de segurança pública, o que vêm se somar à ligação entre tráfico de drogas e armas, contexto este que caracterizou uma profunda mudança na dinâmica da criminalidade e da violência na cidade. Tal transformação vem se agravando devido à crise econômica que acomete o estado nos últimos anos e à mudança de posicionamento na área de segurança pública para se adequar à falta de verbas, com consequentes mudanças de políticas públicas referentes à segurança no estado.

Diante desse contexto conturbado, o Rio de Janeiro conta com o Disque Denúncia (DD), uma central de atendimentos criada inicialmente para ajudar as polícias Federal, Civil e Militar no esclarecimento de crimes e delitos, mediante o recebimento de denúncias feitas por meio de ligações anônimas da população. Este instrumento tem sido caracterizado, ao longo dos seus 23 anos de funcionamento, por receber as mais variadas demandas que, apesar de não serem apenas criminais, guardam relação direta com os sentimentos de temor, insegurança e medo da população carioca ao longo dos anos.

Criado como uma alternativa para contribuir para a diminuição da violência na cidade, o Disque Denúncia começou a funcionar em 1995, com o objetivo de receber da população informações sobre atividades criminosas e casos de violação aos direitos humanos, para que fossem transmitidas aos órgãos de segurança pública. O serviço, que tem como princípio mais importante a garantia do anonimato de quem fornece as informações, atende desde reclamações sobre a conduta de um vizinho até dados completos sobre quadrilhas e crimes ligados ao tráfico de drogas e de armas, sobre atividades criminosas, sobre o modo de agir de determinadas facções, dentre outras.

Assim, no estado do Rio de Janeiro, o Disque Denúncia atua como um canal entre a população e os agentes de segurança pública no combate à criminalidade e na solução de problemas que atigem comunidades, bairros e cidades. Ao receber denúncias anônimas e coletar informações de importância fundamental na apuração de ações criminosas e delitos, contribui consequentemente para a redução da criminalidade¹. Recebendo mais de 100 mil relatos anônimos de crimes e/ou delitos por ano, este é um programa bastante semelhante ao *American Crime Stoppers*², por exemplo. A informação desses relatórios de crimes é, pois, extremamente importante para reduzir os níveis de criminalidade, ajudando a polícia e a inteligência, assim como outras autoridades, a realizar operações mais eficientes contra criminosos.

O problema

Diante do panorama atual de setores de segurança pública dos governos ao redor do mundo buscarem tecnologias mais eficientes para dar suporte

1 - Disponível em: <<http://disquedenuncia.org.br/o-disque-denuncia>>. Acessado em 28/05/2018.

2 - Disponível em: <http://www.crimestoppersusa.org/>>. Acessado em 29/05/2018.

às suas operações na área de investigação e inteligência, destaca-se o caso especial do estado do Rio de Janeiro, cujas especificidades de geografia física, urbana e humana representam grandes desafios para os agentes envolvidos no contexto da segurança pública do estado. Neste sentido, levando em consideração que pretendemos trabalhar com armazenamento, tratamento e associação em grandes bases de dados, devemos considerar que esses dados precisam ser compreendidos a partir da natureza de suas fontes e suas implicações, tanto sociais como aquelas relacionadas à perspectiva do processo de análise desses dados. Para tanto, precisamos recorrer aos Sistemas Complexos, a fim de abarcar a complexidade tanto das informações e dados quanto da interação entre os diferentes órgãos responsáveis pela inteligência e pela segurança pública (FURTADO, SAKOWSKI e TÓVOLLI, 2015).

Assim, apesar da contribuição para a resolução de crimes que o Disque Denúncia vem trazendo, o serviço ainda encontra alguns obstáculos que limitam o alcance pleno de seu objetivo principal, que é ser um agente facilitador para a resolução rápida e eficiente de crimes e violações aos direitos do cidadão. O primeiro obstáculo diz respeito ao volume de informações recebidas pela central de atendimento diariamente. Como ocorre uma maciça divulgação do serviço pela mídia, além do aumento da criminalidade na cidade, os atendentes recebem aproximadamente 350 denúncias por dia. Este grande número de casos recebidos faz com que a atividade de inteligência tenha um gasto excessivo de tempo para organizar e analisar as informações recebidas. E há um trabalho ainda maior em associá-las aos respectivos sistemas classificatórios para que elas sejam separadas e enviadas às autoridades policiais responsáveis por cada denúncia. Cabe lembrar que existem policiais, delegacias ou batalhões especializados para cada tipo de crime e essas informações geradas pela área de inteligência, que são utilizadas na concepção das cenas ou criação dos cenários da ocorrência e compõem respectivamente os mapas de inteligência, devem ser geradas da maneira mais organizada e eficiente possível, e levantadas todas as informações necessárias à investigação.

Algumas ideias de automatização desse processo de classificação vêm sendo testadas (KU e LEROY, 2014), no intuito de reduzir esse tempo e permitir que a atividade de inteligência amplie os conteúdos informacionais que são facilitadores das atividades de prevenção e da resolução dos problemas vinculados à violência e à criminalidade. No levantamento da literatura de trabalhos relacionados, percebemos que em alguns deles técnicas de *machine learning* foram utilizadas para que, sem intervenção humana, seja possível classificar qualquer denúncia anônima utilizando apenas sua natureza textual (PINHO, 2017). Utilizou-se redes neurais sem peso para treinar um sistema de classificação automatizado, obtendo resultados experimentais com acurácia elevada. Entretanto, existem poucos estudos sobre linguagem natural para a língua portuguesa ao compararmos com a língua inglesa. No trabalho mencionado foram abordadas as mesmas técnicas de processamento de linguagem natural em língua inglesa. Todo o trabalho prévio de tratamento das denúncias é

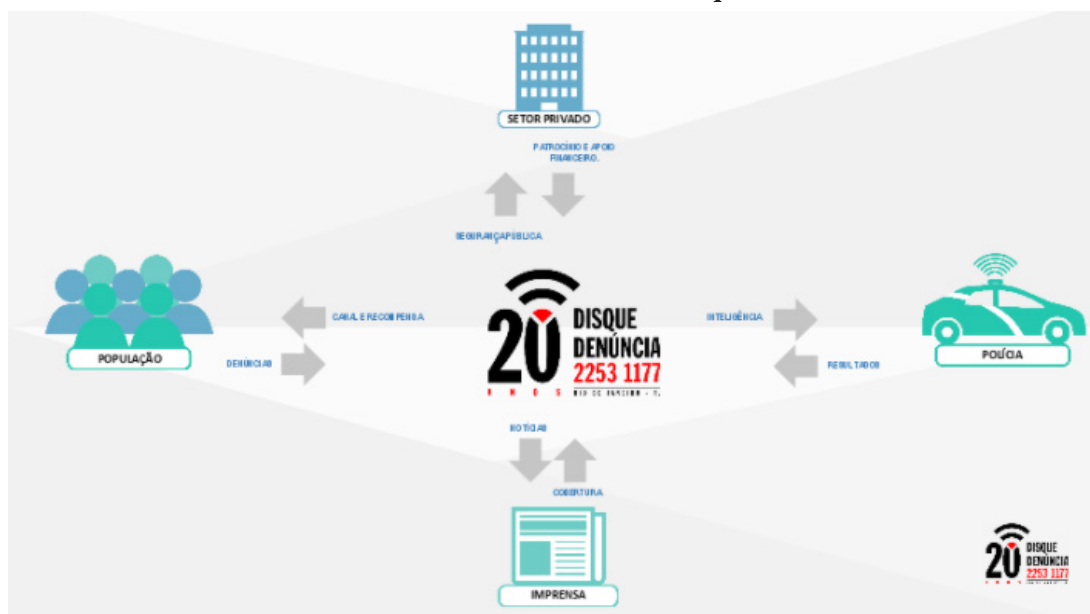
muito similar ao realizado pelo DD e os desafios de trabalhar em língua portuguesa foram maiores devido às características específicas da própria linguagem.

Entretanto, quando tratamos de análises mais profundas, armazenamento e tratamento de dados, surgem dificuldades de natureza operacional derivadas da limitação tecnológica: desde a coleta de informações e dados até a sua análise, passando pelo tratamento, armazenamento e seus filtros, assim como a manutenção de suas estruturas e equipamentos existentes. Por conta do agravamento da crise do estado, novos equipamentos não podem ser comprados nem ter seus licenciamentos de suprimentos. Suas licenças não são atualizadas, e não há suporte e manutenção de todo o equipamento necessário. Desse modo, faz-se necessário o desenvolvimento de um sistema de apoio à decisão (DSS), combinando técnicas de processamento de linguagem natural (PNL), medidas de similaridade, mineração de texto, mapas de inteligência e tratamento de dados que acompanhem o estudo e análise de tais informações, permitindo uma melhor disponibilidade das mesmas nas áreas de inteligência e difusão.

A denúncia produzida pela central configura-se como uma ferramenta dinâmica, que não se esgota no texto produzido pelo atendente, nem é responsável direta por seus resultados: depende de outros processos para consolidar sua legitimidade como instrumento de produção de informação, sobretudo quando caracterizada como delito ou ato criminal.

Este trabalho, assim, apresenta parte da descrição do histórico de denúncias e de seus contextos sociais e políticos, as quais tratam da segurança pública no estado do Rio de Janeiro e de seus atores em um grande sistema complexo. Assim, este trabalho analisa as implicações de tal sistema classificatório como um instrumento de colaboração para a segurança pública do estado, contribuindo para a inteligência e sendo vital às necessidades da população. As denúncias recebidas todos os dias pelo serviço são avaliadas por analistas e encaminhadas às autoridades e aos sistemas de inteligência da Polícia Militar ou da Polícia Civil para serem investigadas (Figura 1).

Figura 1
Dinâmica de relacionamento do Disque Denúncia



Fonte: Disque Denúncia.

Conceitos básicos

Inteligência e segurança pública

Cabe ressaltar que no Brasil utilizamos a definição de inteligência do Subsistema de Inteligência de Segurança Pública (SISP), criado pelo Decreto nº 3.695, em 21 de dezembro de 2000³, e que consiste em: “identificar, acompanhar e avaliar ameaças reais ou potenciais de segurança pública e produzir conhecimentos e informações que subsidiem ações para neutralizar, coibir e reprimir atos criminosos de qualquer natureza”. Nesse contexto, e para que se compreenda a dimensão da atividade de inteligência, ocorre-se ao conceito disposto no artigo 1º, § 2º, do Decreto nº 4.376 de 2002, que regulamentou a Lei nº 9.883 de 1999, onde inteligência é a atividade de obtenção e análise de dados e informações e de produção e difusão de conhecimentos, dentro e fora do território nacional, relativos a fatos e situações de imediata ou potencial influência sobre o processo decisório, ação governamental, a salvaguarda e a segurança da sociedade e do Estado⁴.

O estado do Rio de Janeiro foi o primeiro a criar, em 01 de janeiro de 1995, com base na doutrina da Inteligência Militar, uma agência de inteligência diretamente ligada ao secretário de segurança, chamado Centro de Inteligência de Segurança Pública (CISP), atualmente denominado Subsecretaria de Inteligência (SSINTE). Tem por objetivo planejar, dirigir e executar a atividade de inteligência em segurança pública,

3 - Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto/d3695.htm>. Acessado em 28/05/2018.

4 - Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto/2002/D4376compilado.htm>. Acessado em 28/05/2018.

assim como regula, coordena e supervisiona o Sistema de Inteligência de Segurança Pública do Estado do Rio de Janeiro⁵.

A concepção da atividade de inteligência ao longo do tempo também sofreu um processo de transformação, sendo que na atualidade aponta-se para uma perspectiva na qual a inteligência não tem poder de polícia, usa-se o cérebro para avaliar as informações. Estas podem ser classificadas de diversas maneiras, tais como: informação militar, tática, geral, diplomática, política, econômica, social, biográfica, científica, tecnológica e informação sobre comunicações e transportes. Seu processo envolve as seguintes fases: necessidade de conhecimento; coleta de dados na imprensa ou outros similares, incluindo coleta de dados não disponíveis; processamento dos dados; e disseminação do conhecimento ao usuário, para a tomada de decisão. A atividade deve ser centralizada e seu quadro de profissionais deve ser preenchido por pessoas íntegras e com bons propósitos⁶.

Historicamente, o sistema classificatório do Disque Denúncia sofreu uma grande reformulação a partir de maio de 2002. Se em sua criação existiam apenas 30 tipos diferentes de assuntos (antes de 2005), quando da realização deste estudo o sistema apresenta 19 classes e 187 assuntos de distribuição, definidos conforme a Figura 2.

Podemos observar que o sistema classificatório do Disque Denúncia apresenta características e categorias de outros sistemas classificatórios, não necessariamente relacionados à questão criminal. Abrange a soma dos códigos referentes ao atual Código Penal Brasileiro, aliando parte das categorias usadas pelas polícias, parte do Estatuto da Criança e do Adolescente, o Código do Consumidor, a Lei de Contravenções Penais, o Código de Posturas, além do que os integrantes da central considerem relevante incluir, a partir das diferentes e crescentes demandas que recebem. Da mesma forma, novas leis ou estatutos precisam ser incorporadas ao sistema classificatório para ampliar sua representação de categorias existentes. Com o aumento das categorias utilizadas no sistema, o operador pode escolher mais de um assunto, indicando o mais importante, dependendo do relato realizado na denúncia, permitindo a inclusão de outras denúncias associadas ao primeiro cadastro. Gera-se, assim, o que chamamos de subdenúncias ou denúncias associadas.

Podemos comprovar, observando o sistema de classificação atual contendo 187 diferentes assuntos, que eventos de desordem, contravenções e litígios em geral, não necessariamente criminosos, mas que guardam em si manifestações de insegurança e violência, passaram a integrar o conjunto de categorias do sistema.

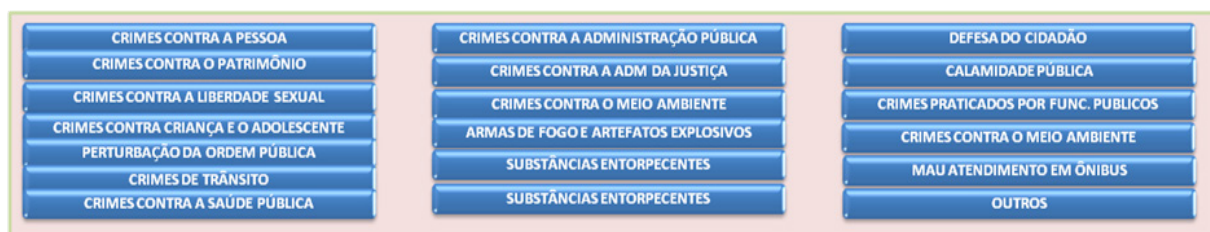
As classes e os tipos de assuntos, a sequência em que são listados e como são apresentados também exprimem o grau de importância e o lugar que os eventos ocupam na hierarquia do sistema. Em primeiro lugar, estão os crimes contra a pessoa, seguidos pelos crimes contra o patrimônio, em terceiro os crimes contra a liberdade sexual e em quarto os crimes contra a criança e o adolescente, realizando uma estrutura indicativa ao Código Penal Brasileiro. Em quinto lugar está a perturbação da ordem

5 - Doutrina de Inteligência em Segurança Pública – DISPERJ. Disponível em: <<http://www.rj.gov.br/web/seseg/exibeConteudo?article-id=1516603>>. Acessado em 04/06/2018.

6 - Idem.

pública (análogo à Lei de Contravenções Penais) e em sexto os crimes de trânsito. Em sétimo, oitavo e nono lugares estão, respectivamente, os crimes contra a saúde pública, os crimes contra a administração pública e os crimes contra a administração da justiça. Em seguida, em décimo lugar, estão os crimes contra o meio ambiente (referindo-se à Lei nº 9.605 de 1998), em décimo primeiro as armas de fogo e artefatos explosivos e somente em décimo segundo a classe sobre substâncias entorpecentes⁷. Em décimo terceiro estão as substâncias tóxicas/explosivas, em décimo quarto a defesa do cidadão, em décimo quinto calamidade pública e em décimo sexto crimes praticados por funcionários públicos.

Figura 2
Sistema de classificação do Disque Denúncia



Fonte: Disque Denúncia.

Processamento de linguagem natural

Uma etapa muito importante para o estudo como um todo é o pré-processamento dos dados antes de qualquer análise. O objetivo desta etapa é tratar os dados em linguagem natural para que possam ser inseridos nos algoritmos computacionais necessários para o estudo. É utilizado o modelo *bag-of-words*, segundo o qual se utiliza uma matriz com referência a todas as palavras das denúncias. Como passo inicial, as denúncias são separadas umas das outras, e cada uma é repartida em unidades de palavras. Logo em seguida, cada unidade é normalizada por meio da conversão de todos os textos para letras minúsculas e são removidos acentos, pontos, números e caracteres especiais. São removidas também as *stop words*, ou seja, palavras que não agregam valor para o documento, como, por exemplo, pronomes e preposições. Na última fase do tratamento dos textos, com o objetivo de igualar palavras com o mesmo significado, é aplicada a técnica de *stemming*, segundo a qual cada palavra é convertida para sua raiz gramatical. Por exemplo, as palavras ‘assassinato’, ‘assassino’ e ‘assassínio’ são todas convertidas para ‘assassi’. De posse do *bag-of-words* e com todas as remoções necessárias realizadas, é feita a conversão do conjunto de palavras (*tokens*) em um grande vetor de palavras. Nesta etapa são gerados

7 – Neste caso, o fato do código ser o número 12 pode estar relacionado com o artigo 12 da lei de entorpecentes, Lei nº 6.368 de 1976.

vários vetores esparsos, onde cada linha da matriz faz referência a um documento e cada elemento faz referência a uma palavra do *bag-of-words*, se ela existe ou não e sua importância. Cabe ressaltar que neste estudo levou-se em consideração que a utilização de ponderação⁸ para cada termo seria de grande valia, pois poderiam ser selecionados somente as palavras ou termos mais importantes para a posterior avaliação.

Metodologia

A metodologia proposta neste trabalho envolve o ciclo de tratamento dos históricos de denúncias recebidas pelo serviço do DD, de uma forma ampla, com toda a classificação de denúncias para os anos de 2015, 2016 e 2017. Desde a extração de radicais temáticos, a construção de dicionários especializados para apoio à extração de entidades contidas nas denúncias e o desenvolvimento de possíveis cenários que compõem os crimes narrados em formato de uma matriz de relacionamentos, com as principais palavras utilizadas em cada denúncia e subdenúncia. Assim, gera-se uma nuvem de palavras por classificações e temas para a criação de cenários. Os cenários são, então, convertidos em mapas de inteligência destinados à análise de vínculos das denúncias e subdenúncias.

O método proposto nesta pesquisa permite a visualização de indícios ocultos pela complexidade das informações textuais que envolvem a denúncia e propicia a descoberta de conhecimento entre associações criminais ao aplicar, para esta finalidade, um modelo de mineração de texto composto de algoritmos híbridos. Vale ressaltar que o anonimato dos denunciadores é mantido, pois este se constitui como o principal pilar estratégico para a existência do DD.

No cenário atual, as informações passadas à central DD por meio de ligações anônimas são filtradas e classificadas em 187 assuntos, divididos em 19 classes (ver Figura 1). Realizamos este estudo considerando os dados desde 2015 até abril 2018, baseado no sistema de classificação de denúncias desenvolvido e utilizado pelos gestores do DD. Visando a uma melhor compreensão da complexidade inerente à segurança pública e às diferentes naturezas dos delitos denunciados, decidimos apresentar essa visão geral do estudo apenas para os anos iniciais (Figuras 3 e 4).

Assim, abordaremos com maior intensidade temas ligados ao combate à criminalidade e ao crime organizado, cujos principais ilegítimos são o tráfico de drogas, o contrabando de armas, a corrupção institucional e a lavagem de dinheiro. Empregamos para isto consulta interna à base de classificação do Disque Denúncia e relacionamos os principais temas de denúncia análogas a este estudo, e decidimos abordar os seguintes temas: apologia ao tráfico, bomba/granada/morteiro, consumo de drogas, obstrução de vias públicas, posse ilícita de armas de fogo, tráfico de drogas e uso ilícito de armas, desde janeiro de 2015 até abril de 2018.

8 - A forma de ponderação utilizada foi a *tf-idf* (abreviação do termo em inglês term frequency-inverse document frequency), uma medida estatística que indica a importância de uma palavra de um documento em relação a uma coleção de documentos ou em um corpus linguístico. Ela é frequentemente utilizada como fator de ponderação na recuperação de informações na mineração de dados. O valor *tf-idf* de uma palavra aumenta proporcionalmente à medida que o número de ocorrências dela cresce em um documento. No entanto, este valor é equilibrado pela frequência da palavra no corpus. Isto auxilia a distinguir o fato da ocorrência de algumas palavras serem geralmente mais comuns de que outras. É dada pela seguinte fórmula:

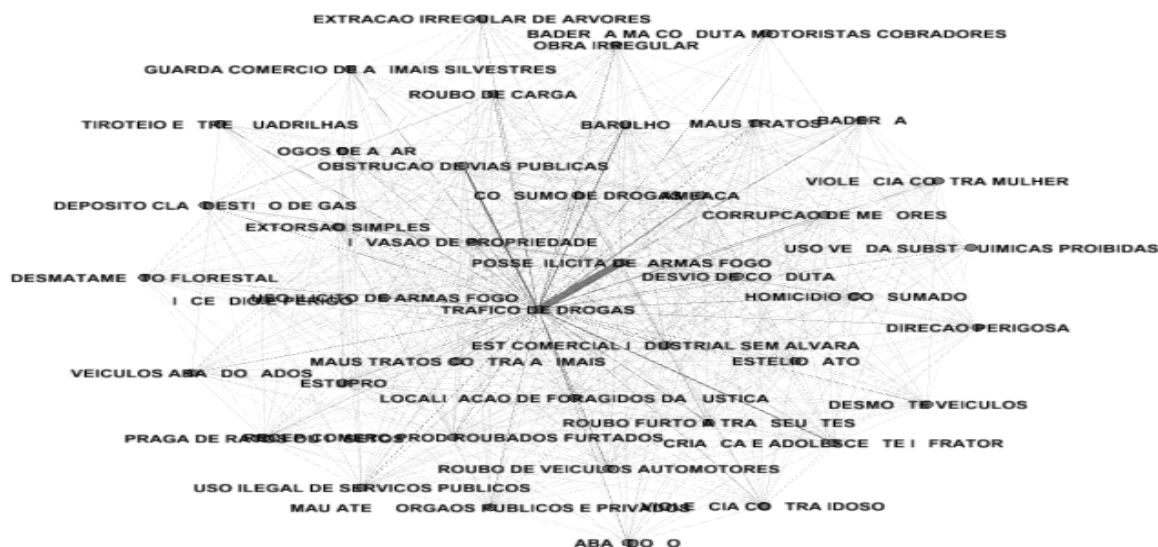
$$w_{i,d} = tf \times idf = tf_{i,d} \times \log(n / df_i)$$

Especificamente, no caso das principais denúncias recebidas, é possível perceber que há relações e padrões que podem ser observados e analisados que relacionam mutuamente denúncias de crime organizado às denúncias escolhidas para este estudo. Ainda, é possível perceber como tais denúncias refletem as dinâmicas sociais associadas ao diferentes crimes, bem como seus impactos na sociedade. Neste primeiro momento, não foi analisada a complexidade da rede gerada pelas conexões entre as categorias de denúncias, mas apenas a semântica formada pelas palavras mais frequentes desde o início da análise para a geração da nuvem. Isto permitiu montar uma análise das palavras ao longo do tempo, como uma *timeline* das principais palavras, semânticas e expressões empregadas na classificação das denúncias. Assim, pode-se observar suas possíveis alterações no decorrer do tempo e as principais implicações dessas mudanças (Figuras 6, 7 e 8).

Cabe ressaltar que foi realizada a análise para todas as classes de denúncias existentes, mas optamos por mostrar apenas as que possuem aderência ao tema do estudo presente. Utilizaremos para este fim, portanto, as nuvens de palavras para as seguintes categorias: substâncias entorpecentes, substâncias tóxicas/explosivas, e armas de fogo e artefatos explosivos.

Um caso observado durante o estudo realizado, que podemos citar como exemplo da observação dessas palavras ao longo do tempo, diz respeito a palavras como escola, colégio e igreja. Começaram a aparecer com uma maior frequência em denúncias relacionadas a armas de fogo e artefatos explosivos a partir de janeiro de 2017. Indicando, assim, alguma modificação no comportamento de algumas atividades criminosas, que utilizavam escolas, igrejas e/ou colégios para esconder e/ou comercializar drogas, sempre com a utilização de armas de fogo ou artefatos explosivos em determinadas regiões.

Figura 3
Maiores relações de classificação de denúncias de 2015



Fonte: Disque Denúncia.

Figura 4
Maiores relações de classificação de denúncias de 2016



Fonte: Disque Denúncia.

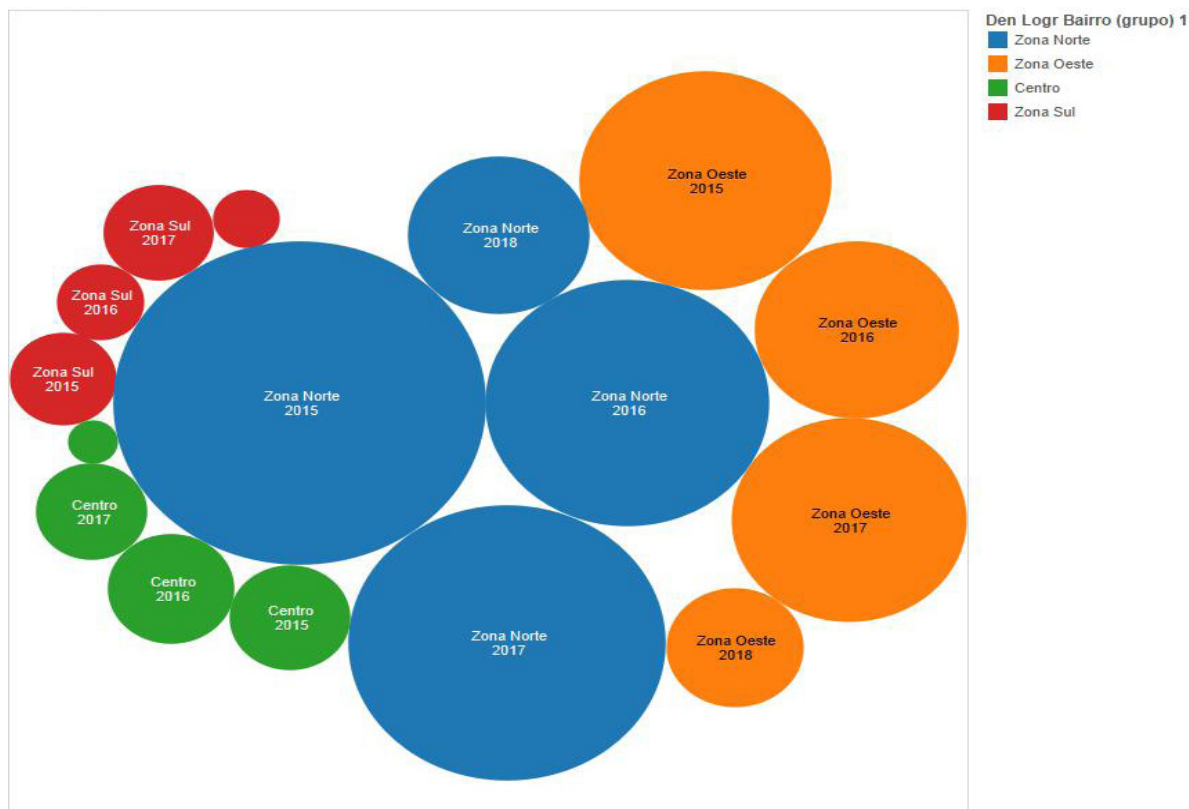
Resultados

Conforme vimos na metodologia, a proposta abarca todo o ciclo de tratamento dos históricos de denúncias, desde a extração de radicais temáticos formados pelas principais palavras relacionadas às diferentes categorias de denúncia, passando pela construção de dicionários especializados para apoio à extração de entidades. Os cenários foram convertidos em mapas de inteligência destinados à análise de vínculos oriundos das denúncias e à descoberta de conhecimento para investigação e possível elucidação dos delitos. As nuvens de palavras extraídas a partir das relações das denúncias e subdenúncias foram representadas por redes de vínculos, posteriormente tratados como um grafo capacitado para os anos de 2015 e 2016 (Figuras 3 e 4 acima).

Os conceitos produzidos pelas denúncias têm similaridade com a teoria dos grafos, segundo a qual os objetos e eventos representam os nós do grafo e os relacionamentos representam os seus arcos (Figuras 3 e 4). Se obtivermos uma organização topológica formada pelos elementos que integram a denúncia, por meio de uma representação gráfica ordenada em um mapa, teremos um grafo, constituído de todas as suas propriedades inerentes, como conectividade, acessibilidade, representação matricial, temporalidade e arborescência, etc. A conversão dos elementos textuais descritivos da denúncia para elementos sintetizados pela simbologia representada pelo diagrama relacional é denominada de estruturação do

conhecimento, e este resultado é formado por uma nuvem composta das principais ocorrências de palavras (*stems*). A seguir, temos o resultado do estudo pelas regiões de ocorrência desses temas ao longo dos anos e as nuvens de palavras associadas (Figuras 5 a 8).

Figura 5
Distribuição por área geográfica — 2015 a 2018



Fonte: Disque Denúncia.

[illegible][illegible]

Cadernos de Segurança Pública | Ano 10 • Número 10 • agosto de 2018 | www.isp.rj.gov.br

Figura 8



Figura 9



Considerações finais

O tema da segurança compreende a participação de diversos atores e espaços de análise, de reflexão e de socialização, não pertencendo apenas a um órgão específico do Estado. Assim, incorporam-se novas e diferentes formas de enfrentamento ao fenômeno da insegurança, da violência e da criminalidade por meio da aplicação e do uso de tecnologia nas atividades de inteligência.

O problema da (in)segurança vem causando preocupação recorrente nos governos ao redor do mundo, ensejando que a atividade de inteligência não esteja restrita a questões que dizem respeito apenas à defesa do Estado, mas também da sociedade. E isto inclui a busca de um conjunto de diagnósticos e projeções no sentido de imaginar cenários de risco e mitigar situações de conflitos, oferecendo, por meio da tecnologia, atividades de inteligência no Brasil, e podendo, assim, agir em prol da defesa do Estado, da sociedade e do cidadão.

As contribuições teóricas pretendidas com esta pesquisa dizem respeito às possibilidades que as atividades de inteligência e do uso de tecnologia podem trazer para o campo da segurança pública e da defesa do cidadão. Indo desde o aspecto da prevenção à repressão criminal qualificada, visando a apontar, sobretudo, medidas que, pela via da democratização de suas atividades em prol dos direitos de cidadania e emancipação humana, possam prevenir e minimizar os efeitos dos conflitos. Isso possibilita aos analistas de inteligência que possuem acesso a essas informações, e que antes agiam por *feeling* em relação às palavras e expressões de referências a crimes e delitos, agora possam, por meio da informação gerada pela tecnologia sobre cada expressão, frase, contexto e circunstâncias das denúncias, traçar seus mapas de inteligência, montar suas linhas de análise, investigar a diferença entre as novas e as antigas expressões empregadas em cada delito e traçar formas e o *modus operandi* referente a cada grupo criminal investigado.

Consideramos que este estudo, além de ter se mostrado útil para a compreensão da complexidade envolvida na problemática da segurança pública e de suas operações, pode gerar trabalhos futuros. Temas como a verificação da semântica das palavras das nuvens, com foco em temas específicos, utilizando expressões, semânticas e palavras requerem do trabalho uma contínua atualização dessas expressões, frases e análises. Assim, também propomos a análise das novas expressões das entidades de relacionamento contidas nas denúncias de acordo com seus radicais temáticos e por meio de outras fontes de dados como aplicativos, redes sociais e dados abertos. Este trabalho, pois, está no escopo de uma pesquisa que aborda o uso da tecnologia cognitiva na gestão do conhecimento e integração entre as instituições da segurança pública, no contexto de suas bases de dados, aplicada à inteligência.

Referências Bibliográficas

FURTADO, B. A., SAKOWSKI, P. A. M. & TÓVOLLI, M. H. **Modelagem de Sistemas Complexos para Políticas Públicas**. IPEA, 2015. Disponível em: <http://www.ipea.gov.br/portal/index.php?option=com_content&id=25860&Itemid=383>. Acessado em 29/05/2018.

PÁSSARO, S., KLEIN, E. & LOPER, E. **Natural Language Processing with Python: analyzing text with the natural language toolkit**, 2009. Disponível em: https://books.google.com.br/books?hl=en&lr=&id=KGIfiiP1i4C&oi=fnd&pg=PR5&dq=%20Natural+language+processing+with+Python%3A+analyzing+text+with+the+natural+language+toolkit&ots=Y3Alv8MGH3&sig=qVmMsCwsIso9UO5XM22uttVYz_s#v=onepage&q=Natural%20language%20processing%20with%20Python%3A%20analyzing%20text%20with%20the%20natural%20language%20toolkit&f=false>. Acessado em 29/05/2018.

PAULA, G. de, DANDOLINI, G. A. & SOUZA, J. A. Tecnologia da Informação e Comunicação e as atividades de Inteligência. **Revista Ordem Pública**, Vol. 5, N.1, Semestre I, 2012. Disponível em: <<https://rop.emnuvens.com.br/rop/article/view/48>>. Acessado em 29/05/2018.

PINHO, R. A., BRITO, W. A. T., MOTTA, C. L. R. & LIMA, P. V. Automatic Crime Report Classification through a weightless Neural Networks. ESANN 2017 proceedings, **European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning**. Bruges (Belgium), 26-28 April 2017. Disponível em: <<https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2017-42.pdf>>. Acessado em 29/05/2018.

KU, C. H. & LEROY, G. A. **Decision support system: automated crime report analysis and classification for e-government**. Elsevier, 2014. Disponível em: <<https://pdfs.semanticscholar.org/c186/7cace312aa54c5b089570289d4edc4c2cbd2.pdf>>. Acessado em 29/05/2018.

WU, H. C., LUK, R. W. P., WONG, K. F. & KWOK, K. L. Interpreting tf-idf term weights as making relevance decisions. **ACM Transactions on Information Systems (TOIS)**, 26(3), 13, 2008. Disponível em: <<https://dl.acm.org/citation.cfm?id=1361686>>. Acessado em 29/05/2018.