



PUC Minas

IEC – Instituto de Educação Continuada
Pós-Graduação em Ciência de Dados e Big Data

Recuperação da Informação na Web e em Redes Sociais

Análise de Tweets durante o carnaval de BH

Aluno: Matheus D’Almeida Vieira (Matrícula: 96463)

Professor: Zilton Cordeiro Jr.

**Março
2019**



PUC Minas

**IEC – Instituto de Educação Continuada
Pós-Graduação em Ciência de Dados e Big Data**

Projeto Final

Análise de Tweets durante o carnaval de BH

Trabalho apresentado ao Instituto de Educação Continuada (IEC) da Pós-Graduação em Ciência de dados e Big Data da PUC Minas, como requisito parcial para obtenção de créditos na disciplina de Recuperação da Informação na Web e Redes Sociais.

Aluno: Matheus D’Almeida Vieira (Matrícula: 96463)

Professor: Zilton Cordeiro Jr.

**Março
2019**

Sumário

1. Resumo	1
2. Introdução	2
3. Descrição Das Atividades	2
3.1. Ferramentas utilizadas.....	2
3.2. Pipeline e Metodologia.....	3
4. Análise dos Resultados.....	10
4.1. Conclusão sobre o processo	10
4.2. Conclusão sobre os dados.....	11
5. Trabalhos Futuros	19
6. Bibliografia.....	20

1. Resumo

Este trabalho tem como objetivo aplicar todas as técnicas de recuperação, análise de dados não estruturados, mineração de texto e visualização de dados aprendidas no curso de Recuperação de informações na Web e Redes Sociais.

Para o desenvolvimento deste trabalho foi realizado o monitoramento do Twitter dos dias 01/03/2019 a 04/03/2019 coletando todos os tweets relacionados às hash tags “#Carnaval” e “#Carnaval BH” e aos textos “CarnavalBH” e “Carnaval BH”. Uma vez que os tweets foram coletados durante o período do carnaval, realizamos uma análise de sentenças, Ngrams e frequência de termos para tentar perceber os acontecimentos do carnaval que foram referenciados no tweet, após essa análise criamos um grafo de tweets para verificar os principais influenciadores do Twitter durante o carnaval de Belo Horizonte.

2. Introdução

No primeiro trimestre de 2018, segundo a IT Midia e o Twitter, a plataforma de microblog Twitter atingiu 336 milhões de usuários ativos reportando um crescimento de 21% na receita quando comparado ao mesmo período de 2017. Este crescimento de receita e aumento do número de usuários é justificado pelo fato do mesmo ter se tornado um canal digital de comunicação direta, rápida e fácil entre entidades (empresas, marcas, órgãos governamentais) e pessoas de fácil acesso a todos, de forma pública e em tempo real.

A liberdade conferida aos usuários da plataforma que permite emitir opiniões ou escrever simples textos de forma pública faz com que a plataforma seja uma ótima fonte de dados para monitoramento de marcas, identificação de perfis de potenciais compradores, monitoramento de experiência dos clientes, identificar formadores de opinião e várias outras formas de monitoramento atividades de rede sociais. Tudo isso é possível não somente pela simplicidade da plataforma, mas principalmente devido à vontade das pessoas em emitir opiniões sobre temas, marcas, produtos, eventos e experiências de forma aberta e direta.

Deste modo, o fluxo de texto em tempo real do Twitter é muito mais significativo e muito mais rico que a grande maioria do público usuário da plataforma percebe. A partir deste fluxo de informações e aplicando as técnicas corretas de mineração de texto é possível filtrar os ruídos (neste caso mensagens sem informação relevante) de uma determinada discussão ou tópico e gerar informações relevantes que indicam por exemplo: qual o tema foi mais discutido, quais são as palavras mais usadas pelos usuários do serviço ou produto, se as palavras são positivas ou negativas, a repercussão de uma ação perante a sociedade dentre outras possibilidades.

Sendo assim, neste trabalho final, foi desenvolvido uma sequência de coleta de dados dos tweets relacionados ao Carnaval de Belo Horizonte durante o período de 01/03/2019 a 04/03/2019 com o objetivo de identificar fatos relevantes que aconteceram durante o evento, identificar as impressões o público presente e as definir entre impressões positivas e negativas, entender se houveram influenciadores digitais atuando na rede e identificar os principais tópicos que foram discutidos ou que fizeram parte do carnaval.

Para realizar a identificação dos itens anteriormente citados, foi criado um pipeline de coleta, tratamento, mineração de texto que foram sendo adequados e customizados conforme a necessidade de análise gerando principalmente visualizações em formato de Tag Clouds para realização das análises pertinentes e aplicando deste modo todas a técnicas aprendidas ao longo do curso.

3. Descrição Das Atividades

3.1. Ferramentas utilizadas

Para o desenvolvimento deste trabalho foram utilizados os seguintes softwares para construção do pipeline, mineração de texto e visualização de dados:

- **Knime versão 3.7.1** → Software usado para configuração de todo pipeline de processamento dos dados do Twitter com os plugins de conexão do Twitter e Text Processing.

- **Sqlite** → Fornecido pelo Knime e utilizado como repositório dos Tweets recuperados pelo pipeline durante os 4 dias de coletas consecutivas.
- **Notepad++** → Utilizado como editor dos arquivos CSV (MPQA Lexicom) para incrementar a base traduzida e fazer modificações nas traduções erradas.
- **Fluxos aprendidos em aula** → Reaproveitamento e adaptação de todos os fluxos apresentados em aula para obter as informações tal como suas bases de dados (Ex.: Base de palavras usado no fluxo do boticário).
- **Twitter (Conta e API)** → Foi criada uma conta de estudo e um aplicativo para acesso a API de coleta de dados do Twitter (conta específica do aluno).

3.2. Pipeline e Metodologia

Para realização da análise foi criado um pipeline que contempla a coleta de dados, processamento e tratamento de textos, tratamento de exceções com acertos pontuais nos termos e palavras-chaves extraídas e uma etapa específica para cada visualização de dados.

Pipeline de coleta de dados

O pipeline de coleta foi configurado para coletar os tweets da Hash Tag “#CarnavalBH” e do termo “Carnaval BH” separando em uma base única os tipos “Mixed” e “Recent” a cada hora, iniciando as 09:00 da manhã até as 21:00 horas da noite durante todo o período do carnaval. Para os tweets do tipo “Popular” foi realizada uma coleta a cada 3 horas de 09:00 da manhã até as 21:00 horas da noite.

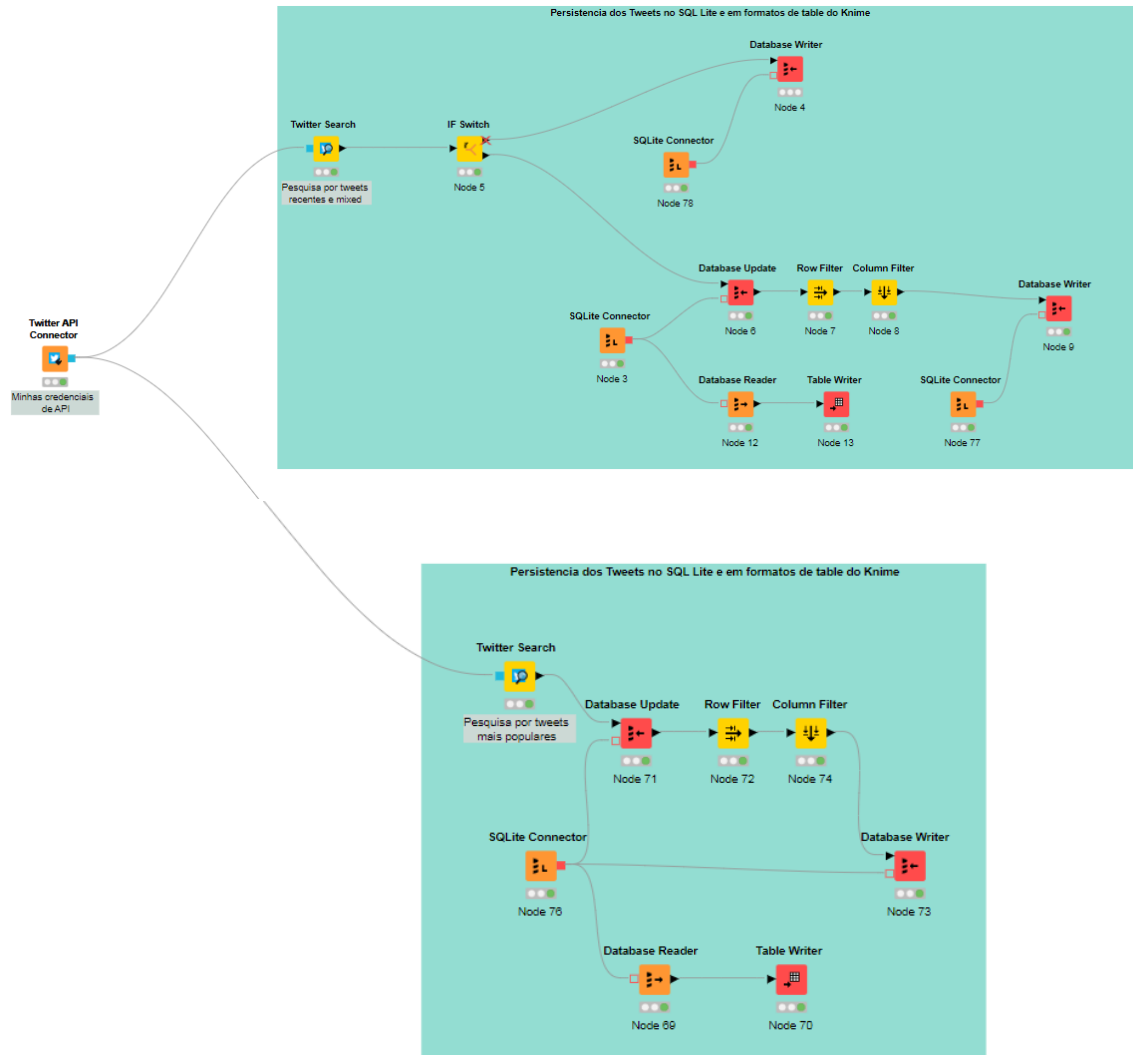
Os fluxos foram totalmente separados (conforme imagens a seguir) de modo a criar o banco de dados no SQLite e manter a base sempre atualizada com a coleta sem haver duplicidade de postagens do Twitter para que isso não afete a relevância dos termos durante a análise. Ao final de todo processamento de coleta, são gerados 2 arquivos tipo “.table” para ser processado no pipeline de tratamento e mineração de texto e analisados em várias etapas seguintes.

Devido ao período de coleta ser mais de um dia e não pontual, houve a necessidade de armazenar os tweets em bases intermediárias conforme citado anteriormente para permitir um processamento posterior da mesma.

Foi priorizada a coleta nos dias do evento para evitar a coleta de tweets enviados após o final do carnaval visando minimizar o esforço de tratamento e filtragem de dados após a coleta ter sido realizada diminuindo a chance de erro e aumentando a precisão da etapa de análise.

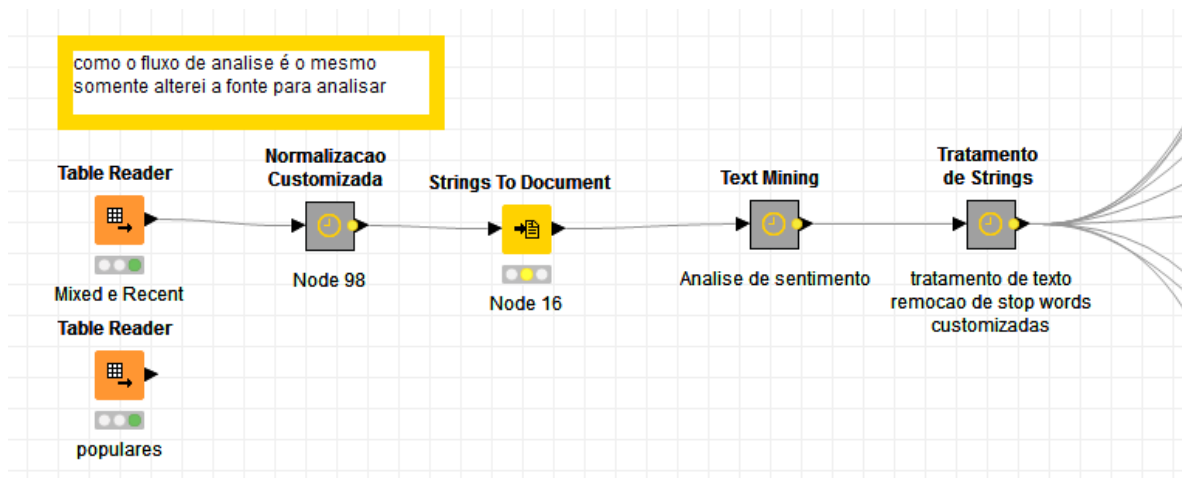
Durante a etapa de coleta não houve nenhuma modificação do dado coletado ou aplicação de filtro pois assim conseguimos maximizar a abrangência da coleta com o objetivo de ter a maior assertividade possível na análise dos textos. Para a fase de análise foram considerados aproximadamente 1.350 tweets coletados das Hash Tags citadas para “Mixed” e “Recent” e 325 para “Most Popular”.

Todas as bases geradas estão armazenadas na pasta “MyDataBase” localizada no diretório raiz do workflow deste projeto do Knime.



Pipeline de Processamento - Etapas Comuns de tratamento

Para cada uma das bases geradas pelas etapas de coleta, dividimos e agrupamos os tratamentos em metanodos específicos no Knime para agrupar e facilitar o entendimento e visualização do fluxo de tratamento. É importante centralizar esse tratamento em um único ponto para que o resultado da visualização seja o mais correto possível entre as técnicas de análise e para garantir a mesma base de dados para todas as técnicas de análise.



As etapas em comum consistem na leitura da base gerada pela coleta e o repasse para os seguintes nodes:

- **Normalização customizada** → Nesta etapa realizamos alguns tratamentos diretamente no texto dos tweets removendo caracteres não contemplados no stop words, caracteres especiais e corrigindo palavras em que a capitalização falhou.
- **Strings To Document** → Após os tratamentos de dados brutos, criamos a estrutura de documento que será utilizado para os demais nodos e processos exigidos pelas estruturas no Knime.
- **Text Mining** → São adicionadas todas as tags que serão utilizadas para as análises nas etapas posteriores.

Ponto de atenção: As listas de palavras positivas e negativas tiveram que ser revisadas (algumas adicionadas e outras removidas) para adequação ao contexto da análise.

Como por exemplo a tradução do termo “Stood” que estava como “Se” no arquivo que deve ser excluída.

```

2307 "stirring", "agitação"
2308 "stirringly", "convincendo"
2309 "stood", "se"
2310 "straight", "reta"
  
```

A palavra “chama” teve que ser desconsiderada devido ao contexto existir um bloco chamado “Chama o síndico”

```

979 "flame", "chama"
980 "flame", "chama"
981 "flatter", "alisador"
  
```

Foi necessário incrementar o registro de palavras negativas para adequar ao contexto

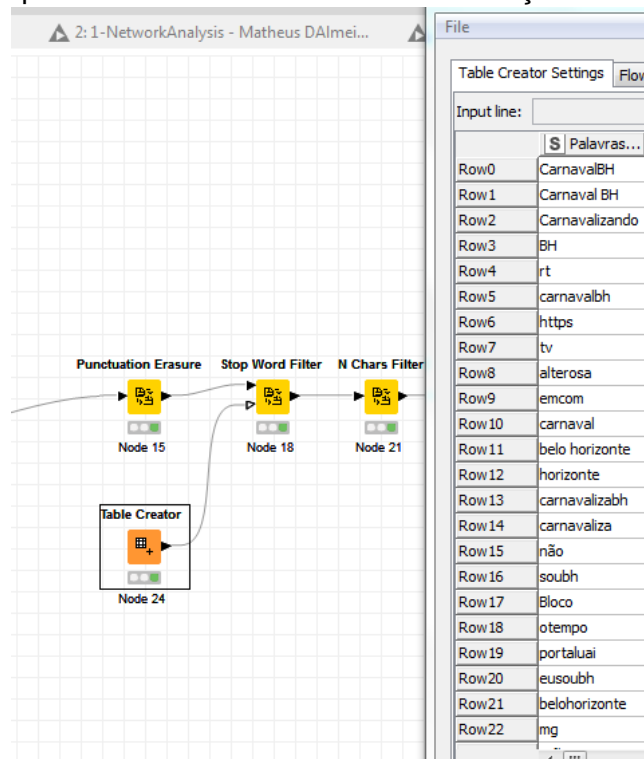

```

4901 "zealously","zelosamente"
4902 "milicia","milicia"
4903 "miliciano","miliciano"
4904

```

- **Tratamento de strings**→ Contempla todo tratamento necessário para os processamentos posteriores.

Ponto de atenção: Foi necessário informar palavras extras para serem removidas como stop words. Isso foi necessário para tirar o viés de análise colocado pelos termos pesquisados. Ou seja, uma vez que procuramos por “#CarnavalBH” e “Carnaval BH” com certeza esses termos terão a maior frequência por serem objetos de pesquisa contidos em todos os textos do tweet. Além disso, essa etapa foi usada para remover palavras que não representavam nenhuma semântica para análise (ex: https e rt) e removemos algumas palavras que são nomes de canais de comunicação.

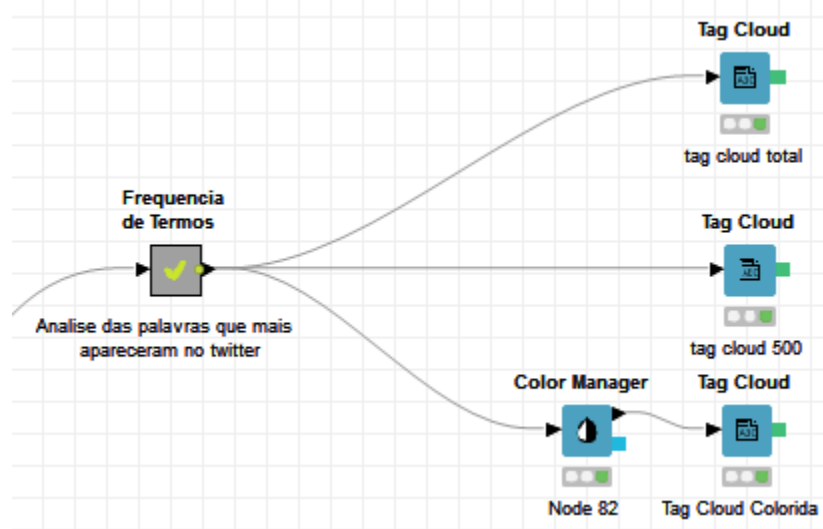


Pipeline e etapas de Análise

Uma vez que todos os tratamentos de palavras foram realizados e padronizados para todos os fluxos, foi criada uma fase de análise com um pipeline específico para cada técnica utilizada.

Para iniciar as análises a primeira inspeção foi calcular a frequência de termos sem nenhuma categorização ou Tag de identificação dos mesmos. Ou seja, o objetivo era ver as palavras que mais se destacaram obtendo uma visão geral de todos os Tweets coletados

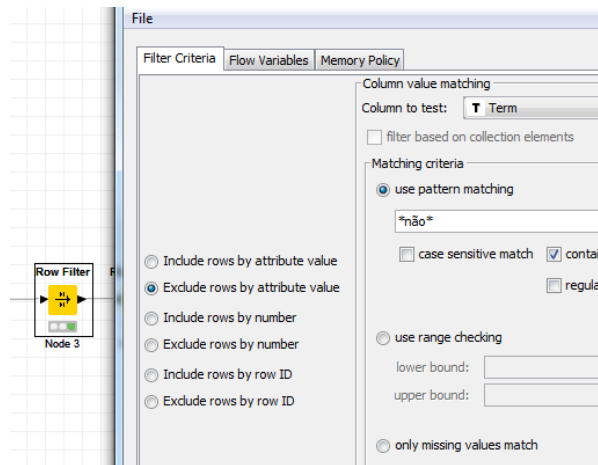
e principalmente verificar se tínhamos alguma palavra indevida que estivesse comprometendo o cálculo da frequência de termos.



Nesta etapa foi encontrado problemas e realizados ajustes nas etapas anteriores:

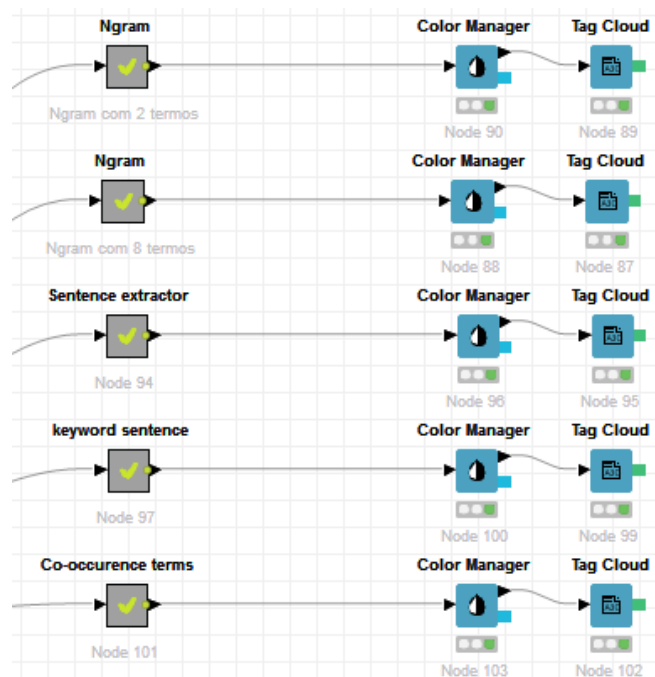
- Problemas em palavras que não estavam ficando totalmente minúsculas (ex.: Bloco e bloco) afetando o cálculo da frequência.
- Stop words não removidas “Se” e “SE” sendo necessário incluir na lista customizada.
- A palavra “Não” com uma frequência muito alta mas com baixo valor semântico para análise sendo necessário excluir essa palavra
- Palavras como “RT” e “BH” e algumas outras que não tinham valor semântico mas estavam com uma frequência alta foram retiradas como stop words customizadas.

Ponto de atenção: Mesmo utilizando Regex, substituição de string para algumas palavras o knime não conseguiu realizar a correção, portanto foi necessário remover os termos na etapa de análise para não comprometer o resultado. **Por este motivo somente utilizamos o TF com frequência absoluta e nunca relativa na análise.**



Para avaliar as frequências das palavras foram criadas 2 tag clouds, uma considerando todas as palavras e uma estipulando um range de 500 para que seja possível visualizar as variações de frequência e corrigir exceções como citadas anteriormente. Deste modo foi possível, em âmbito geral, identificar todos os temas e palavras com significância nos tweets, sendo possível evoluir para a segunda etapa de análise que se refere ao entendimento do contexto das palavras.

Para análise de contexto, foi utilizado uma análise por N Gram e Sentence Extraction com o objetivo de comparar as palavras com maior frequência e se elas aparecem nas sentenças mais relevantes do carnaval e quais são os tópicos e contextos mais discutidos durante o evento.



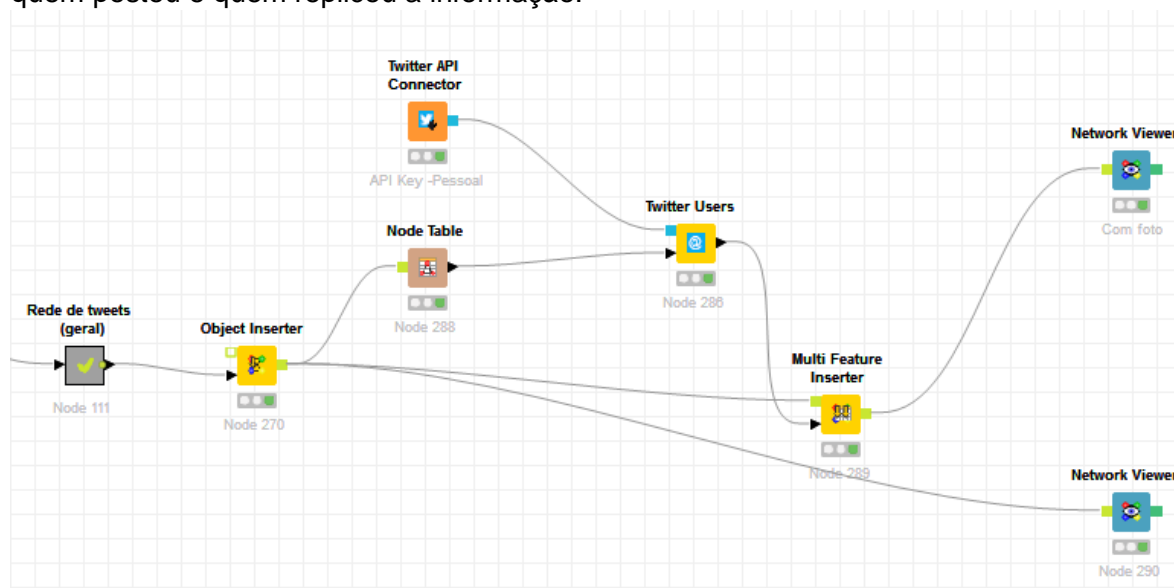
Nesta etapa foi iniciada com a análise de NGram de 2 termos, para obter os termos em pares mais utilizados e verificar se era possível identificar tanto a semântica quanto o contexto dos termos. Para confirmar o contexto estendemos o NGram para 8 termos para

captar uma frase pequena com alta frequência mesmo que a ordem das palavras não sejam as mesmas.

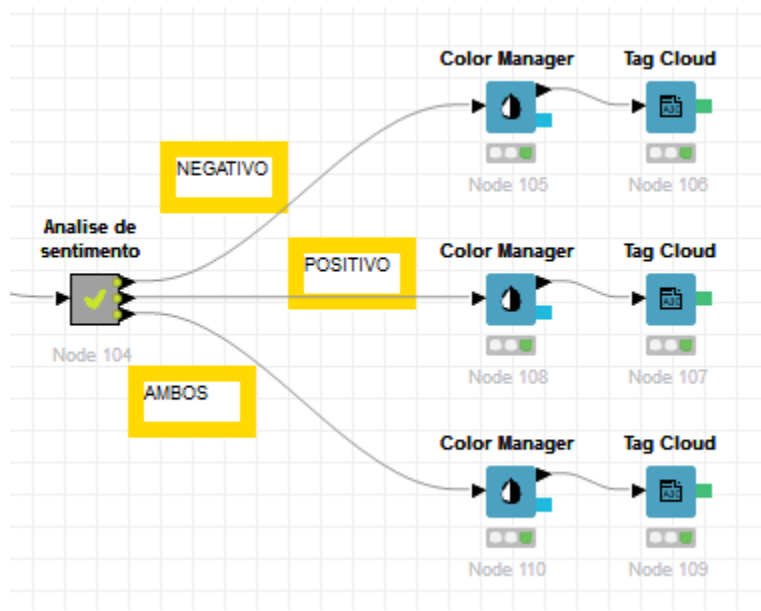
Para confirmar o NGram, usamos um Sentence Extractor e um co-occurrence terms, assim podemos ter certeza de todo o contexto e significado das expressões identificadas com alta frequência.

Por fim, uma vez que temos as confirmações das sentenças passamos para etapa de identificação de formadores de opinião, produtores de conteúdo e pessoas mais referenciadas durante o carnaval.

Esta etapa foi executada, através da criação da rede de retweets e relacionando quem postou e quem replicou a informação.



Deste modo também separamos os tweets positivos e negativos para identificar os perfis do twitter a intensidade que informações positivas e negativas se propagam e qual intenção se destacou mais. Neste caso sempre excluindo tweets neutros ou sem classificação.



4. Análise dos Resultados

4.1. Conclusão sobre o processo

Tendo-se executado todas as etapas descritas acima, podemos tirar as seguintes conclusões.

A primeira conclusão que se chega sobre o trabalho, diz respeito ao processo de análise em si, pode-se afirmar que o trabalho de mineração de texto é um processo totalmente interativo e incremental pois em sua essência temos um formato livre e não estruturado de dados que por si próprio adicionam uma dificuldade considerável intensificado pelo idioma português.

A cada etapa de análise, até a última, foi necessário retornar as listas de palavras positivas/negativas, listas customizadas de StopWords, analisar o range de palavras da tag cloud várias vezes até que o dado final estivesse totalmente tratado e os resultados fossem confiáveis.

Os principais motivos que tornaram o processo iterativo e responsável pelo maior esforço no tratamento de dados no trabalho foram:

- Necessidade de adequação e criação de Tags em português, tornando esse processo manual.
- Revisão de modo iterativo de lista customizada de stopwords, pois termos como “se”, “RT”, “https”, “carnaval”, etc neste contexto não poderiam ser consideradas na frequência e algumas realmente não são consideradas stopwords, mas no contexto tratado e para o objetivo do trabalho deveriam ser retiradas.
- Problemas na padronização de alguns termos, em parte resolvido com o Stemmer e outros através do filtro do termo após a criação da bag of words.
- Necessidade de revisão e modificação da lista de palavras positivas e negativas para o contexto estudado.

Uma vez que os itens acima foram devidamente tratados a análise dos dados foi realizada com segurança e de modo assertivo.

4.2. Conclusão sobre os dados

As conclusões a seguir foram tiradas analisando a fonte de dados “Mixed e Recent” e também “Most Popular”, entretanto o resultado das duas foi muito semelhante visto que os tweets mais populares têm um impacto significativo na frequência de termos. Deste modo se decidiu usar a base com maior abrangência e maior diversidade para uma análise mais ampla sobre o carnaval de BH.

Frequência de termos de todos os tweets

Nesta primeira análise, o objetivo é analisar em todos os tweets coletados quais os termos que mais apareceram sendo obtida a seguinte tag cloud.



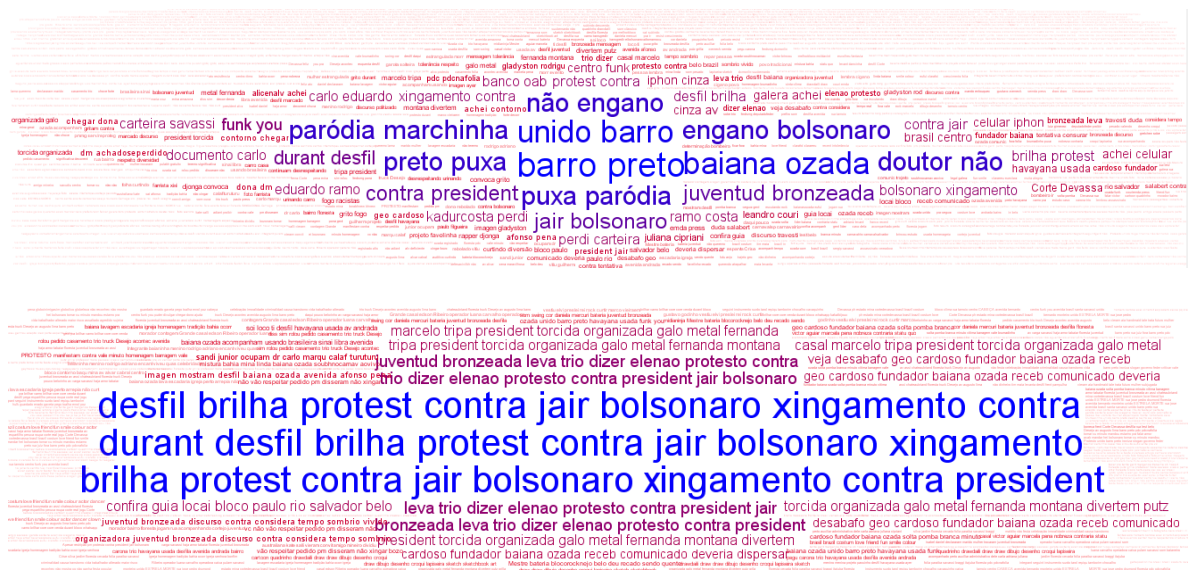
Observamos nessa tag cloud podemos perceber claramente as palavras: Bolsonaro, Contra, paródia, presidente, desfile, marchinha, baiana, Ozada que inicialmente fazem todo o sentido no contexto do evento. Podemos relacionar os destaques da Tag Cloud do seguinte modo (hipóteses):

- Baiana Ozadas é o maior bloco de belo horizonte que neste carnaval ocupou praticamente toda a avenida Afonso Pena e Savassi com maior duração do desfile.
- Foi noticiado pela mídia a onda de protestos ao Bolsonaro e seu governo em praticamente todos os blocos de Belo Horizonte através de paródias.
- O bloco Brilha foi um dos 3 blocos maiores de BH que ativamente criticou o governo de Bolsonaro durante o desfile.

Essas são as primeiras suposições obtidas pela análise dos tweets, entretanto vemos algumas palavras que a priori não têm conexão com esses fatos, inclusive noticiados na imprensa, que são: Carteira, Eduardo, Kadurcosta e OAB.

brunoni
canta grafit brubaspino
carlinhos grafito silveira que lindezaeap
cartosanbarraigugrupo corpea querido andraverick penuca garimpar brilho
camavaleria guia real aovivo avenida alto pega garcego briga
camarau homenagemrocanegai bairro bolsonaro you visto zuera pdcontorno galo brasil
castelo in reciclado blogdoarcano briha barro imagen xingamento vamo yurimouraa pdcnataa funcionamentonoferiadoozonario
celebr jufantini ricardopintoli brazil convoca cobertura afonso perdi xuxa whoismaarco vale war pib toha boca
censurar jentibhago rodrigo calcinha durant djonga banco ramo pena som vapor usada verancones pausa floresta bilquinho
cinza ju roupa encher estado carlo mina unido transgredir vandacruzelo paulo feriatodos bilique
cliente kakal saibam tco tempo usando passai eustaquio barbaenabim
corde salvador deputadofeder ferdsmgpreto carteira contra desfil oab leandro rua rod Unidosdosambaquelethom berem
Deossa continentalabianca roa fuma caducostea caducostea caducostea
PROTESTO kikloz star eibolsonarovatomarcou fogo rapper confira kadurcosta couri racistas receb forma nelmocaci espanhilo batena
Parabéns continua kekik supriatadomim
abrimo corte limatibara subemadmo havayana funk relato costa savassi eduardo centro protest pulah travesti mtoquerido equip barragem
aghaasag demonin local tcojmetip homenageia gladyston soubhnocarnav documento belo proibir olodum travel mistura ei barbaenai
and denisejutalini tcoatunim
art desha mahinfuenciadojodymyle linda grito acompanh president mostram tomar minuto eh barbaenab
arthurhebraek discurso maior tconduyent lucimar jair miguelarcanyoj the melhor dua barbaenai
averiguar diversão manifestamtoconkissuwo mend tentativa mauoludom domingo baiansocada
ayer danieli marcou tcoyriqulwa matheusantão dmselabaha

Visando entender o significado destas palavras, utilizamos em seguida o Ngram com 2 e 8 termos para confirmar e verificar se o que identificamos na primeira análise estava correto.

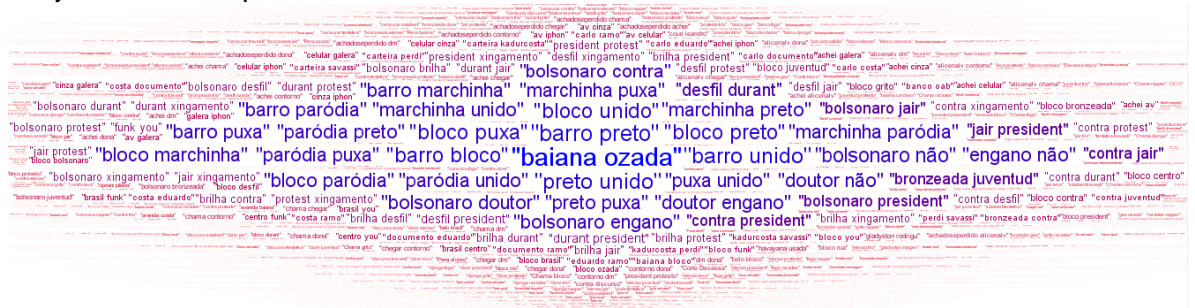


12

bereni kakalgoncalv arthurhebreak
 hifm kakalv andrewerick justapeta amandufcal
 brancaomg letacania boradoido chain blocobianero barro castelo vumoura iorithiago agathaasg
 brubsprimo lucianwall deniseeu cidad ju belo ricardopintoth garimpar beijar bocaParabéns tuma limpeza wmicent jfbantini adsump
 brunemarqu mahinfluncia domingo tcofrydyndjw curta melhor jeito belo brazil in belo brazil confira bloco rua baianaizada bhsurpreendit toctuhqibghomenageia olodum Como abrindo porta vssali garcezoq Du
 1 avariqual pñ bem elevado fantasia multa alegria equip marcou bloco baiana ozada centrobaiana ozada vapor afonso Pena desfil soubhnocarnav aovivo bak aplicativo slu mobil tempo real festa não bora pra
 ferdsmg dcomonit subvernario não adiant proibitconfira guia local bloco paulo no salvador belo confira acompanh tempo real cobertura PROTESTO manifestam contra vale minuto homenagem barra
 ei bozonaro tomar desfil pega espartilho peruca roupa corte real jogu Corte Devassaa baiana ozada encher centro acompanh tco vamo transgredir eibolsonarovaitomocuu and the tra
 oco consehojeidi star war blogdoarcanjo**baiana ozada som xuxa rapper djonga convoca bloco Chama grito fogo racistas**imagen mostram desfil baiana
banco oab kadurcosta perdi carteira savassi documento carlo eduardo ramo costa visto alto afonso Pena receb baiana
 unidosdosambaqueixinho celebr grupocorpo c
durant desfil brilha protest contra jair bolsonaro xingamento contra
 atheuscarlo **confira relato estado mina cobertura baiana ozada unido barro preto havayana usada funk you ferdsmg deputadofeder**
 não preocup cinza abrirrem
 ia curtindo diversão leandro couri **bloco reciclado rua tamoiu centro funk you avenida brasil santa savassi unido barro preto rua juiz fora barro preto pdc pdcnafolia**
 imagen gladyston rodrigu desfil desfil bloco havayana usada avenida andrada santa **baiana ozada vapor afonso Pena desfil soubhnocarnav** passei carnavaur briga zuera dmsss kkkkk melhor
 calcinha calcinha pulabh gratuito tcofdaurbnifoto bitar paiva pulabh gratuito tcofwzmggygbmistura bahia mina linda baiana ozada soubhnocarnav aovivo discurso travesti duda salabert contra tentativa censu
 linlul mauuloculim gaio continentalaleitaco lucimar mend tcofjelmelp client linda usando kakal bijoux bloco quando cidad desfila rua bairro floresta kakalbijoux carnavaiderua bloquinho brincodepaet brihobolocob
 carlosanibanga mtoquendo grafi art pausa recanegar baiaia feriadobh barbeianibh eustaquioobarbearia funcionamentonoferiado barbeianibhivismoarco lucimar mend tcofpepziak doctormat
 carnavalesmh nondorenaul helima sensacional rock sinton acess limababara stephanekhorni didonal
 camavauuuuu netomaozair imagen ayzer salesbedgar dehimanun
 celorochemello rubinha continua
 imesadonari denegcontredobro documento
 dcomonit subvernario não adiant proibitconfira guia local bloco paulo no salvador belo confira acompanh tempo real cobertura PROTESTO manifestam contra vale minuto homenagem barra
 confira bloco rua baianaizada bhsurpreendit toctuhqibghomenageia olodum Como abrindo porta vssali garcezoq Du
 banco oab **kadurcosta perdi carteira savassi documento carlo eduardo ramo costa visto alto afonso pena receb baiana c**
 PROTESTO manifestam contra vale minuto honer

durant desfil brilha protest contra jair bolsonaro xingamento cc
 imagen gladyston rodrigu **confira relato estado mina cobertura baiana ozada unido barro preto havayana usada funk you baiana ozada vapor**
 discurso travesti duda sa
 quetacorn kadupitobro garimpar beijar bocaParabéns tuma limpeza wmicent jfbantini adsump
 quetacorn kadupitobro garimpar beijar bocaParabéns tuma limpeza wmicent jfbantini adsump
 oco consehojeidi star war blogdoarcanjo**baiana ozada som xuxa rapper djonga convoca grito fogo racistas**imagen mostram desfil baiana ozada avenida afons
 banco oab **kadurcosta perdi carteira savassi documento carlo eduardo ramo costa visto alto afonso pena receb baiana c**
 PROTESTO manifestam contra vale minuto honer

Geramos também a Co-occurrence terms para ver qual termos aparecem em conjunto mais frequentemente.



Com base nas quatro tag clouds geradas conseguimos entender o contexto de basicamente todos os 1300 tweets realizados sobre o carnaval de BH na hash tag #CarnavalBH sendo eles:

- O Bloco Unidos do Bairro Preto e o bloco Brilha foram os blocos que mais tiveram um ativismo contra o governo e Jair Bolsonaro usando paródias e marchinhas.

Evidência e comprova a análise:

<https://www.uai.com.br/app/noticia/carnaval/2019/03/04/noticias-carnaval,242495/folioes-chamam-bolsonaro-de-miliciano-no-bloco-unidos-do-barro-preto.shtml>

Folhões chamam Bolsonaro de miliciano no bloco Unidos do Barro Preto

Bolsonaro foi alvo de muitos protestos e xingamentos, em meio ao repertório com sucessos de Chico Science, Nação Zumbi, maracatu

por Flávia Ayer
 04/03/2019 17:20

Bloco Unidos do Barro Preto puxa paródia com a marchinha do ...

<https://www.youtube.com/watch?v=NoHMzWyuGao> ✓



5 de mar de 2019 - Vídeo enviado por Arthur Mayer

Bloco Unidos do Barro Preto puxa paródia com a marchinha do carnaval.

Arthur Mayer. Loading ...



Portal Uai ✓

4 de março às 09:31 · 🌐

#CarnavalBH Bloco Unidos do Barro Preto puxa paródia com a marchinha do carnaval. "Doutor, eu não me engano, o Bolsonaro é miliciano", cantam. 📷 Flávia Ayer/E.M/D.A Press

- Bloco Baiana Ozadas foi realmente o maior bloco de BH e o mais comentado.

Evidência e comprova a análise:

<https://www.hojeemdia.com.br/horizontes/baianas-ozadas-agita-500-mil-pessoas-em-desfile-pelo-centro-de-belo-horizonte-1.698399>

Baianas Ozadas agita 500 mil pessoas em desfile pelo Centro de Belo Horizonte

José Vitor Camilo

jcamilo@hojeemdia.com.br

04/03/2019 - 16h57 - Atualizado 17h18

- Bloco Chama o Síndico teve movimentos contra o racismo

Evidência e comprova a análise:

<https://noticias.uol.com.br/carnaval/2019/noticias/redacao/2019/03/04/angola-janga-pede-fogo-nos-racistas-e-grita-por-marielle-em-desfile-em-bh.htm>

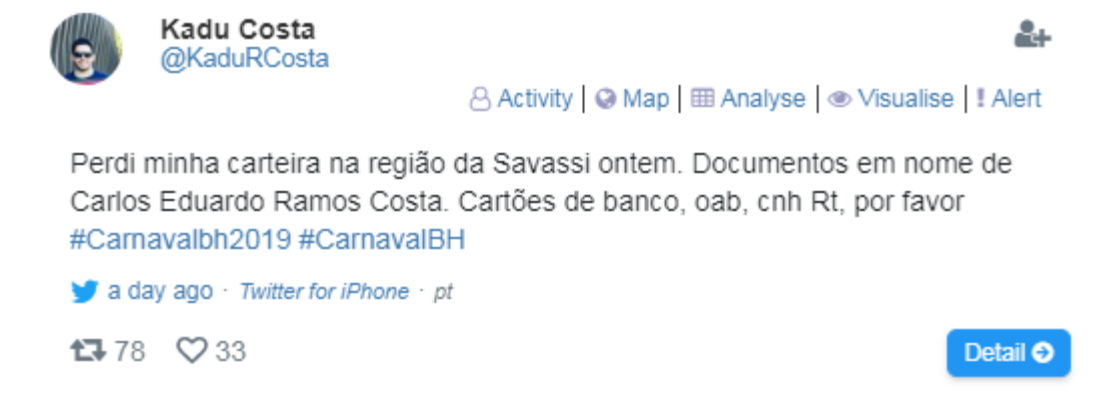
Angola Janga pede fogo nos racistas e grita por Marielle em desfile ...

<https://noticias.uol.com.br/.../angola-janga-pede-fogo-nos-racistas-e-grita-por-marielle...> ✓ ▼

4 de mar de 2019 - Angola Janga pede fogo nos racistas e grita por Marielle em desfile em BH ... O verso "Fogo nos Racistas", do hit "Olho de Tigre" do mineiro Djonga, foi o que mais levantou o Jorge Ben Jor chama o síndico no Carnaval do Mirante BH ... PM proíbe grito político no bloco Tchazinho Zona Norte em BH.

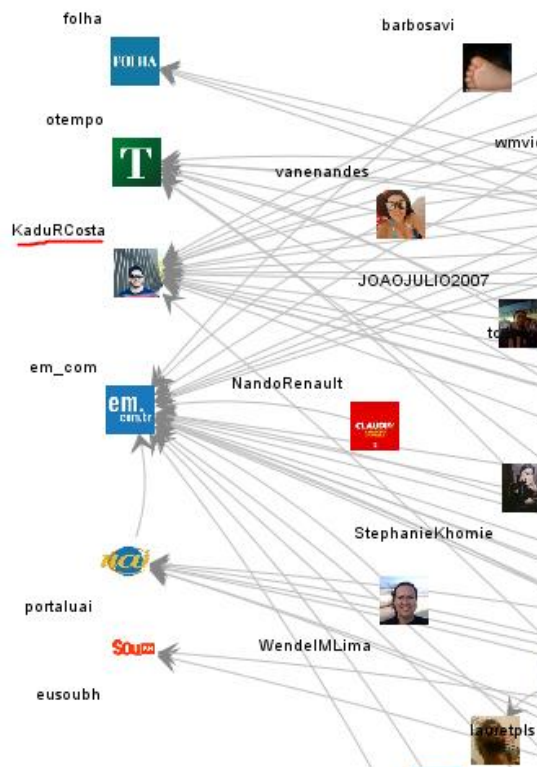
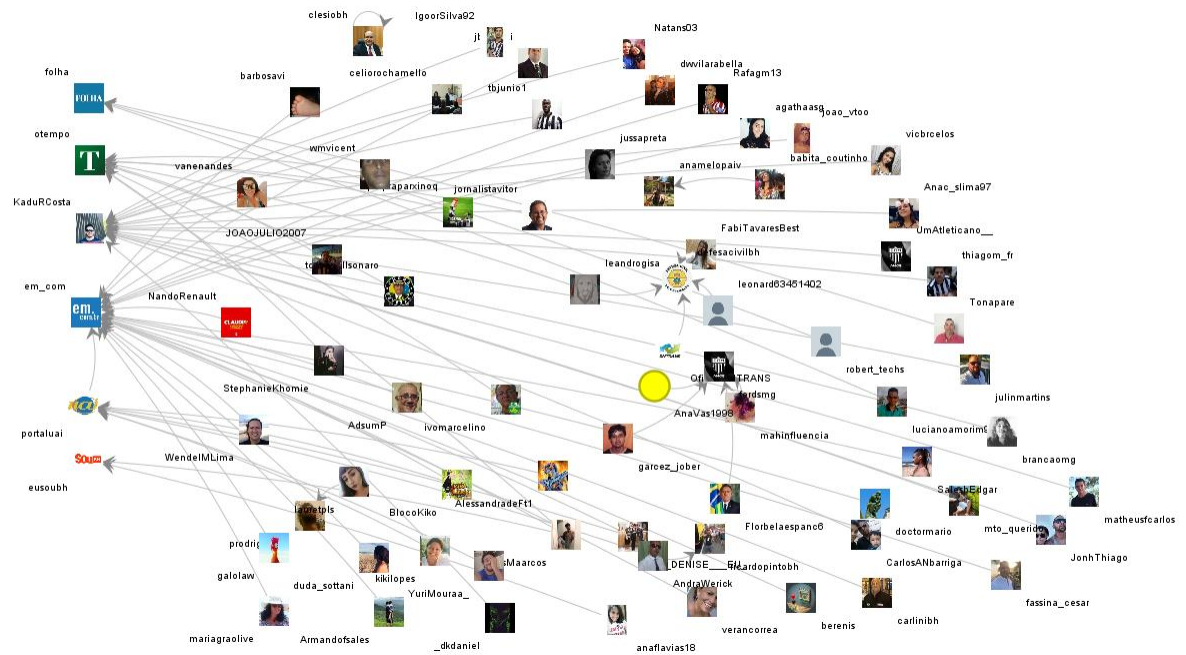
- Conseguimos entender a relação entre as palavras que estavam soltas no contexto. Se tratam do Carlos Eduardo Ramos que perdeu sua carteira contendo todos os documentos inclusive cartões de banco e carteira da OAB na região na Savassi e pediu ajuda através do twitter para encontrar a mesma.

Evidência e comprova a análise:



Redes de conexão entre perfis

Uma vez que comprovamos todos os acontecimentos relevantes e através dos tweets coletados, vamos analisar a rede de relacionamento entre as pessoas no Twitter. O caso que mais chama atenção e que queremos comprovar é o KaduRCosta, por não se tratar de um cenário político, mas uma solicitação de ajuda via Twitter.



Palavras negativas

caiu
canibal irreverência assassina
cansado irritar Cuidado imoralidade ameaçar
caos juro assassinato feio CONFUSÃO gritar ameaça
chato local baixo protestos resistência explosão Apesar fome ai
Acidente chocalho mal bomba roubado vão grito atrapalhar Dúvidas suspeito fascista acidente
Ai maldito ditador cheio cobertura Corte paródia racistas PROTESTO ódio surdo falta
Falta mar ferver confusão perto Devassa contra engano proibir Crise tristeza sujeira faca
Imóveis corte morte intolerância Abuso pouco protesto desrespeitando Carnage tirano seca estragar
Jam criticar pausa morrendo deficiente transgredir perde ruim saudade errado Selvagem
Lixo denunciar pressa perda luta pesado risco entrar Secreto
MORTE destruir proibido perder quebrar dores Protesto
Morte destruição propaganda dor Problemas
Perdido dificuldades Perto

Palavras positivas

Civil animado CABEÇA
Dança anjo luz amo Boa
Defesa bastante magia trabalhador livre amizade Amizade
Diversão branco maravilhosa verdade tradicional sustentável limpeza Super Alegria
ESTRELA brincadeira moda verdadeira Gosto lindo sensacional super incrível Simples
Emocionante cabeça objetivo vontade Liberdade amigo maior grande restrição suficiente harmonia Sim
celebração paciência Integrante Pronto apenas quente ficar forte feliz puro simples gostar Parabéns
Maduro Ribeiro mesmo Mestre melhor homenagem dança clássico especial melhoria fácil
responsável fama Misericórdia
Segurança acima respeitar bem tolerância som diversão determinação amigos
Estrela compacta pagar beijo ajuda significativa sim SENSACIONAL Melhor maravilhoso reconhecimento
conferir parabéns bom crescimento tradição fantasia respeito alegria Grande Felicidade lugar querer dourado
FANTASIA coragem paz calor delírio vivo líder Desejo segurança integrante pronto direito Maravilhosa
Feliz céu perdão celebridade legal Bom princípio flor por favor desejando Justiça
Gratidão dedicado pergunto coração madrugada destaque poder democrática Inspiração

Palavras positivas x Palavras negativas

Parabéns
Perdido céu Morte
Perto dedicado magia criticar Misericórdia
Problemas democrática mal suspeito luz corte Mesmo
Protesto denunciar maldito sustentável perder surdo local coragem Maravilhosa
Secreto desejando mar trabalhador pesado Melhor perda super livre conferir MORTE
Selvagem destruir maravilhosa verdade puro amigos PROTESTO Felicidade morrendo sujeira limpeza compacta Lixo
Acidente Sim destruição moda verdadeira restrição cheio SENSACIONAL determinação Grande segurança melhoria suficiente juro chocalho Justiça
Ai Simples dificuldades morte vontade ruim confusão atrapalhar dança Corte alegria Desejo roubado maravilhoso simples irritar chato Jam
Alegria objetivo Apesar sensacional especial clássico ficar Devassa grito líder Crise protestos lugar seca irreverência celebração Inspiração
Super direito paciência CONFUSÃO tirano feliz forte desrespeitando paródia transgredir Carnage princípio intolerância saudade incrível caos Imóveis
pagar Cuidado tradicional grande maior melhor protesto homenagem fantasia Bom madrugada
acidente dor parabéns Integrante tristeza lindo quente proibir contra engano sim vivo legal integrante risco imoralidade
Amizade ai dores pausa Maduro ódio Abuso resistência diversão pouco tradição feio flor responsável harmonia cansado
Boa ameaça dourado paz Segurança Dúvidas amigo
CABEÇA ameaçar elevado perdão assassinato Gosto apenas vão tolerância som racistas perto significativa explosão ferver recuperar gritar canibal Gratidão
Civil amizade entrar pergunto baixo Liberdade deficiente Mestre respeito cobertura respeitar delírio ditador reconhecimento gostar caiu Feliz
Dança amo errado poder beijo Pronto luta bem perde crescimento destaque querer fácil cabeça Falta
Defesa animado estragar por favor bom Ribeiro mesmo ajuda coração quebrar fome brincadeira FANTASIA
Diversão anjo faca pressa bomba acima celebridade propaganda fascista branco Estrela
ESTRELA assassina falta proibido calor pronto fama bastante Emocionante

Com base na visualização das tags clouds acima, podemos concluir que a parte negativa do carnaval se destacou no Twitter em sua grande maioria pelos protestos de diversas formas ocorridos nos blocos, sendo o de maior destaque as paródias contra o governo e os discursos ocorridos em alguns blocos sobre problemas sociais como o racismo e discriminação.

Em compensação, fica claro analisando as palavras positivas que as homenagens ocorridas no evento (ex.: as vítimas de brumadinho), alegria, respeito, dança, diversão e música foram pontos positivos no carnaval. Além disso, a alta frequência de palavras como Tradição, Melhor e sensacional comprovam o sucesso do evento na visão popular.

Ao compararmos os pontos positivos e negativos em uma única cloud tag, vemos que por distribuição, aparentemente pontos positivos e negativos são equilibrados, mas fica evidente a forte presença de movimentos e protestos como característica marcante do evento ocorrido em Belo Horizonte devido ao destaque das palavras negativas demonstrado na tag cloud.

5. Trabalhos Futuros

Como continuação deste trabalho podemos realizar as seguintes análises para melhorar o estudo. Em relação a análise de sentimentos dos textos foi realizado uma análise em termos individuais e não contemplando todo o significado do tweet como um todo.

Neste caso uma sugestão seria para cada tweet coletado e com termos identificados e associado a uma tag, atribuir um peso considerando a intensidade positiva ou negativa da palavra e através da soma dos valores obtidos podemos classificar a frase como um todo sendo positiva ou negativa e ter uma análise mais exata da impressão do público envolvido no evento. Para que a essa análise seja exata além da necessidade de atribuir a intensidade a adjetivos e palavras positivas e negativas temos que implementar no Knime um POS em português com uma abrangência significativa para o contexto analisado de modo a tratar os tweets com exatidão.

Assim seria possível usar com maior confiabilidade a última rede criada de retweets separando tweets negativos de positivos para entender como cada pessoa influenciou as opiniões sobre o carnaval de BH. Este item não entrou na conclusão pois foram identificados tweets com palavras negativas que não tinham o conceito totalmente negativo tornando a análise um pouco imprecisa.

Uma sugestão para acelerar esse trabalho seria criar um node do Knime usando Polyglot (análise de texto em python) que já oferece um grande suporte ao idioma português (<https://polyglot.readthedocs.io/en/latest/modules.html>).

6. Bibliografia

DARZI, Rodrigo. **A importância do monitoramento de tweets**. Disponível em: <<https://agenciaimma.com.br/monitoramento-de-tweets/>> Acesso em: 10 de fevereiro de 2019

ITMIDIA. **Twitter cresce base de usuários e reporta lucro pela segunda vez consecutiva**. Disponível em: <<https://itmidia.com/twitter-cresce-base-de-usuarios-e-reporta-lucro-pela-segunda-vez-consecutiva/>> Acesso em: 10 de fevereiro de 2019

PEARL, Cathy. **Using KNIME to Find Out What your users are thinking**. Disponível em: <<https://www.knime.com/blog/using-knime-to-find-out-what-your-users-are-thinking>> Acesso em 22 de fevereiro e 2019.