

UNIVERSIDADE FEDERAL DO ABC (UFABC)

CONECTIVIDADE E INCLUSÃO DIGITAL PARA SÃO PAULO

RELATÓRIO R3: A FORMAÇÃO DA OPINIÃO PÚBLICA EM
RELAÇÃO ÀS PRAÇAS DIGITAIS

Resultado 3 (R3): Relatório sobre o acompanhamento da opinião dos cidadãos em relação à política de conectividade e inclusão digital.

Prof. Sérgio Amadeu da Silveira, UFABC (coordenador)

Santo André, Fevereiro de 2015.

EQUIPE DE PROJETO

Coordenador

Prof. Dr. Sérgio Amadeu da Silveira – CECS/UFABC

Pesquisadores Principais

Prof. Dr. Claudio Luis de Camargo Penteadó – CECS/UFABC

Prof. Dr. Carlos Alberto Kamienski – CMCC/UFABC

Colaboradores

Waleska Barbosa da Silva

Juliano Ratusznei

Geovani Anacleto da Silva

Nilton Queiroz Pinheiro

Raul Iago Ataíde de Souza Melo

Paulo Roberto Elias de Souza

Renata Faleiros Camargo Moreno

Ariane de Andrade Quinalha

Vanessa Cristina do Nascimento

Gustavo Frazato Mobrício

Henrique Vander Galdino dos Santos

Isadora Castanhedi

Jonatas Silveira de Souza

Julia Moreno Rosin

Ligia Machiavelli de Lima

Luana Hanaê Gabriel Homma

Lucca Amaral Tori

Rafael Akio de Miranda Pinto

Victor Machado de Franca

Wesley de Oliveira

SUMÁRIO

EQUIPE DE PROJETO	2
1. INTRODUÇÃO	4
1.1. A PRESENÇA DO TWITTER NA INTERNET	6
1.2. A OPINIÃO E O COMENTÁRIO NO TWITTER.....	6
2. A EXTRAÇÃO DE DADOS DAS REDES SOCIAIS.....	9
3. METODOLOGIA	13
3.1. EXTRAÇÃO DE DADOS DO TWITTER.....	13
3.2. ANÁLISE DOS DADOS EXTRAÍDOS	16
4. ANÁLISE DOS DADOS	21
4.1. ANÁLISE DA EXTRAÇÃO POR TERMOS	21
4.1.1. ANÁLISE DE USO DE PALAVRAS-CHAVE E FREQUÊNCIA	21
4.1.2. CLASSIFICAÇÃO DE SENTIMENTO SEGUNDO ANÁLISE AMOSTRAL	23
4.1.3. ANÁLISE DE TERMOS RELACIONADOS AO PROGRAMA WIFI LIVRE.....	25
4.1.4. ANÁLISE DE TERMOS CHAVE PARA A PREFEITURA DE SÃO PAULO	26
4.1.5. RELAÇÕES ENTRE PALAVRAS-CHAVE ATRAVÉS DO GRAFO	32
4.2. ANÁLISE DA EXTRAÇÃO POR PRAÇAS	33
4.2.1. ANÁLISE DA FREQUÊNCIA DE UTILIZAÇÃO DAS PRAÇAS	33
4.2.2. RELAÇÃO DA UTILIZAÇÃO DAS PRAÇAS ATRAVÉS DO GRAFO	36
5. REFERÊNCIAS.....	39
ANEXOS.....	41
Anexo 1	41

1. INTRODUÇÃO

Este é o terceiro relatório do projeto “Conectividade e Inclusão Digital para São Paulo” é uma parceria da Universidade Federal do ABC (UFABC) com a Prefeitura de São Paulo que visa estudar os efeitos da Internet, em suas múltiplas dimensões, na vida dos cidadãos e das comunidades atendidas pela política de abertura de sinal de rede sem fio do programa Wi-Fi Livre SP. Esses estudos têm a finalidade de gerar dados para fomentar a implementação e avaliação de uma política pública de acesso gratuito à Internet por meio de rede sem fio. Os estudos e dados obtidos são importantes ferramentas para a gestão da política de conectividade e inclusão digital, que além de medir os efeitos da política, também permite fazer os ajustes e correções necessárias, condições necessárias para uma maior efetividade e eficácia da ação, como também auxiliar no desenvolvimento de novos projetos voltados para o aprofundamento da inclusão digital e promoção da cidadania.

Neste relatório pretende-se realizar o acompanhamento da opinião dos cidadãos em relação à política de conectividade e inclusão digital. Para tal, identificou-se uma das principais práticas de convivência na Internet e de formação de opinião pública no Município de São Paulo, a saber, o uso das redes de sociais online. Neste sentido, optamos por organizar um processo de coleta, acompanhamento e análise das opiniões sobre as Praças Digitais do programa Wi-Fi Livre SP em uma das principais redes sociais utilizadas pelos paulistanos: o Twitter. Desse modo, podemos observar o que é dito, com que intensidade, com que impacto ou repercussão entre os demais usuários da rede, em geral, formadores de opinião do Município.

O Twitter é uma das principais redes de relacionamento social no país. Ele possui características distintas de outras redes sociais, tais como o Facebook ou LinkedIn. Sua principal distinção está na inserção da informação que deve respeitar o limite de 140 caracteres. Por isso, as postagens são curtas e visam alertar e informar algo com grande velocidade. Em geral, serve para criar ondas de atenção a partir da publicação de links que nos remetem para blogs e outros sites. Devido a essa estrutura e dinâmica, o Twitter é classificado como nanoblog ou micro-blogging.

Durante as eleições norte-americanas de 2008, o Twitter ganhou destaque mundial devido ao intenso uso que obteve pelo então candidato Barack Obama para informar sua agenda e realizar declarações

fundamentais sobre suas decisões políticas. Nos protestos ocorridos no Egito no início de 2011, o Twitter teve um papel importante de informação da opinião pública internacional e de articulação dos ativistas egípcios. Antes de ser deposto, o regime de Mubarak mandou bloquear o Twitter no dia 26 de janeiro de 2011 devido a sua grande importância na comunicação realizada entre os revoltosos.

Devido a adesão de grandes celebridades do mundo midiático, o Twitter passou a ser muito utilizado para jovens acompanharem seus ídolos, fãs saberem dos passos de seus artistas e esportistas preferidos. Assim, diversos atores de telenovelas e programas de entretenimento brasileiros conquistaram milhões de seguidores no nanoblog. Uma das práticas mais comuns no Twitter é o comentário crítico de eventos transmitidos em tempo real, por streaming ou pela TV aberta, principalmente jogos de futebol, novelas, séries de TV, programas de entrevistas. O sistema de divulgação dos assuntos mais comentados no Twitter, em cada momento, reforça essa tendência de comentar acontecimentos transmitidos em outras mídias.

O sistema de ranqueamento, denominado pelo Twitter de Trend Topics, pode ser local, nacional ou internacional. As disputas pelos assuntos de maior relevância levaram a diversos embates entre correligionários das principais candidaturas a cargos majoritários nas eleições brasileiras de 2012 e 2014. Como um comentário pode ser replicado pelas pessoas que o leram, o Twitter trabalha com o chamado efeito viral nas redes. Um leitor de uma mensagem contagia um outro ao republicá-la ou retuitá-la, conforme o jargão da rede. Isso tem levado à constituição de importantes ondas de opinião na plataforma do nanoblog.

Mais recentemente, o Twitter tem incorporado elementos de outras redes sociais, tais como a possibilidade de inserção de memes, fotos e vídeos. Apesar disso, a composição dos usuários da rede tende a reunir mais os chamados formadores de opinião. Trata-se de uma rede de replicação veloz de informações e de orientação das atenções. Um dos principais usos do Twitter é o de repercutir os conteúdos de outras mídias, bem como, de relatar as experiências vividas em tempo real, seja de pessoas comuns ou de personalidades. Sua presença na cidade de São Paulo é relevante e indica as tendências dos líderes de opinião.

1.1. A PRESENÇA DO TWITTER NA INTERNET

A pesquisa do Comitê Gestor da Internet, realizada em 2013, indica que 18% dos brasileiros que acessam à Internet utilizam o Twitter. Entre os brasileiros conectados com curso superior, 27% possuem perfis no nanoblog. Este número cai para 16% entre os conectados que possuem apenas o ensino médio. Há uma forte correlação entre maior escolaridade e uso do Twitter. O grupo etário que possui maior adesão ao Twitter está na faixa de 16 a 24 anos, com 24% de presença entre os da mesma idade que acessam a Internet. Encontramos 19% dos brasileiros conectados com idade entre 25 e 34 anos, 17% entre 10 e 15 anos, 14% entre 35 e 44 anos, 11% entre 45 e 59 anos e 7% com 60 anos ou mais.

O uso do Twitter é mais presente entre as pessoas com maior renda. Dos brasileiros com acesso à Internet que ganham até 2 salários mínimos, 14% são usuários do Twitter. Há um crescimento para 17% entre os internautas que ganham de 2 a 3 salários que sobe para 19% na faixa de 3 até 5 salários mínimos. Nos conectados que recebem entre 5 e 10 salários, o Twitter está presente em 22%. O maior conjunto de usuários de Twitter está entre os internautas que recebem mais de 10 salários mínimos, ou seja, 26% destes. Maior escolaridade e maior renda, conformam uma rede social que concentra mais os formadores de opinião que possuem uma maior propensão para obter informações em tempo real de fontes que são protagonistas dos eventos (artistas, políticos, esportistas, ativistas) e que querem ou necessitam informar suas opiniões e atividades.

1.2. A OPINIÃO E O COMENTÁRIO NO TWITTER

O processo pelo qual se forma a opinião pública e como ela influencia as decisões coletivas é demasiadamente complexo e envolve a denominada esfera pública. Se alargarmos o modelo habermasiano de esfera pública e incluirmos nele as conversações mais generalizadas e não apenas ao uso racional de argumentos voltados ao entendimento (8), poderemos observar as redes sociais como arenas discursivas fundamentais para a constituição da vontade coletiva em nossa cidade. Principalmente, devido à grande presença de paulistanos nas redes sociais, 86,5%, conforme constatou-se no primeiro relatório apresentado desta pesquisa. Entre os conectados no Município de São Paulo, 23,8% afirmaram utilizar o nanoblog Twitter em seu cotidiano.

"... em sociedades complexas, a esfera pública forma uma estrutura intermediária que faz a mediação entre o sistema político, por um lado, e os setores privados e sistemas de ação especializados em termos de funções, por outro lado" (HABERMAS, p.41)

"Nas sociedades de larga escala, os cidadãos não podem se juntar num único fórum deliberativo (como a imagem da ágora sugere) em nenhum sentido realista. O processo de debate é inevitavelmente disperso através de uma diversidade de fóruns." (MAIA, p.72)

Devido às suas características, o Twitter e as redes sociais, em geral, não podem ser facilmente enquadradas na tipologia que Habermas criou no livro "Direito e Democracia: entre facticidade e validade". Para Habermas, a esfera pública episódica é composta dos bares, cafés, encontros de rua. A esfera pública de presença organizada envolve o encontro de pais, um público que frequenta o teatro, concertos de rock e reuniões de partido, entre outros exemplos. Já a esfera pública abstrata é aquela produzida pela mídia, sendo composta dos leitores, ouvintes ou espectadores dispersos (Habermas, p.107).

Manuel Castells trabalha o conceito de autocomunicação de massas que pode ser muito útil para compreendermos as transformações ocorridas a partir da emergência das redes sociais online. (Castells, p.58) No twitter, em determinadas condições, as pessoas podem ter suas opiniões pessoais transformadas em mensagens para uma ampla massa. Isso tem ocorrido com muita frequência. Uma constatação, um protesto, uma frase de indignação ou uma piada sobre uma personalidade pode ganhar milhões de leitores em poucos minutos, a depender do seu apelo intersubjetivo. Dado ao processo de interação no Twitter e devido as possibilidades de replicação veloz de mensagens, podemos observar a opinião sendo formada em uma esfera pública interconectada.

Neste sentido, esse relatório busca lançar um processo de acompanhamento da avaliação pública sobre as praças digitais em São Paulo. Para tal, acompanha aqui as opiniões postadas por pessoas que integram os formadores de opinião do município, bem como, seus debates. Os conteúdos coletados no nanoblog, são expostos a seguir e permitem observar o atual estágio de opinião a respeito da política pública de conectividade e inclusão digital da Prefeitura do Município de São Paulo. A coleta de postagens sobre

outros temas relativos à Prefeitura foi realizada para permitir a comparação e aferição dos diferentes níveis de relevância dos assuntos debatidos pelos formadores de opinião, usuários do Twitter residentes em São Paulo, no início de janeiro de 2015.

2. A EXTRAÇÃO DE DADOS DAS REDES SOCIAIS

Todas as mensagens publicadas no Twitter são chamadas de *Tweets*, e possuem limitação de 140 caracteres por mensagem. Os *Tweets* são comunicações rápidas e concisas entre os usuários, e podem ser monitorados em tempo real com grande velocidade de comunicação através da plataforma do Twitter. Tal monitoramento permite a criação de diversas aplicações, desde interação em tempo real de programas de TV com sua massiva audiência, até análises de comportamento e de intenção.

Para a extração dos dados do Twitter, é necessário a utilização de uma ferramenta disponibilizada pela empresa que permite realizar a extração de maneira segura, mantendo a integridade dos dados, e que esteja de acordo com os termos de utilização do serviço. Esta ferramenta permite ao desenvolvedor realizar a manipulação dos dados de um serviço online se chama API¹.

Utilizando a API do Twitter é possível, com a utilização de comandos específicos, desenvolver algoritmos para realizar requisições de informações e receber as respostas para estas requisições em formato padronizado.

Um exemplo é dado abaixo.

Exemplo de algoritmo de uso de uma API

1º passo: estabelecer conexão com o serviço

2º passo: realizar autenticação no serviço através de uma credencial (token²)

3º passo: requisitar informações

¹ **API: Application Programming Interface** (ou **Interface de Programação de Aplicativos**), é um conjunto de rotinas e padrões estabelecidos por um software para a utilização das suas funcionalidades por aplicativos que não pretendem envolver-se em detalhes da implementação do software, mas apenas usar seus serviços. Em um contexto de desenvolvimento voltado para a Internet, a API é um conjunto definido de funções de requisição e de resposta em um servidor.

² **Token:** no caso descrito, o Token significa uma chave única de acesso para o serviço.

4º passo: manipular as informações recebidas

5º passo: gravar as informações

As informações que são recebidas da API são descritas por meio de identificadores. O identificador utilizado para a manipulação de informações sobre o usuário é o *user*, que terá seu conteúdo gravado em uma variável³.

Na variável *user* serão recebidas as informações solicitadas, que neste caso são as informações do perfil do usuário. Na maioria dos casos, para facilitar a leitura e manipulação dos dados por uma ampla gama de tecnologias compatíveis, essa informação de resposta vem em um formato de linguagem de marcação padronizado como o XML⁴ ou o JSON⁵.

No Twitter existem dois tipos de API's especializadas para diferentes funções. São elas: **REST API** e **Streaming API**.

- **REST API:**

A REST API permite acesso de leitura e escrita no Twitter. Com ela podemos realizar a leitura dos Tweets já postados por determinado usuário, fazer buscas dentro do Twitter e até mesmo enviar mensagens pelo Twitter sem estar diretamente conectado na página.

Com esta API é possível realizar buscas nos Tweets que já foram publicados na plataforma. Contudo, possui uma restrição de data para essas buscas, e os retornos são filtrados para as mensagens de maior relevância, ou seja, não é possível buscar os Tweets em sua totalidade por meio desta interface.

³ **Variável** é um objeto utilizado para armazenar um valor ou expressão dentro de um algoritmo.

⁴ **XML: eXtensible Markup Language** é uma linguagem de marcação para necessidades especiais. De forma simplificada, um XML é um arquivo que não tem um formato definido, você pode criar campos e atributos conforme sua necessidade. Vários arquivos que usamos no dia-a-dia são baseados no XML, como o docx ou xlsx das versões mais recentes do Microsoft Word e do Microsoft Excel, respectivamente.

⁵ **JSON: JavaScript Object Notation** é um formato leve para intercâmbio de dados. Funciona como uma alternativa mais leve e mais fácil de se escrever do que o XML, além de ser um formato mais fácil de se trabalhar com linguagens de programação.

- **Streaming API:**

A Streaming API permite o acesso ao fluxo de dados que o Twitter de forma praticamente instantânea, permitindo que as informações publicadas no serviço possam ser obtidas em tempo real. Sua grande vantagem é que todos os Tweets são capturados, o que permite acesso a um maior volume de dados.

Essa API é dividida em três partes, que monitoram funções diferentes dentro do serviço:

- Public Streams;
- User Streams;
- Site Streams.

A **Public Streams** permite fazer o monitoramento de assuntos específicos, através de palavras-chave, utilizando o banco de dados global de mensagens públicas. É possível facilmente obter todos os Tweets de usuários que utilizaram uma palavra chave específica em qualquer região do planeta.

A **User Streams** e a **Site Streams** são bastante parecidas, a **User Streams** permite o monitoramento da atividade de um usuário específico, retornando todas as suas atualizações. Por outro lado, a **Site Streams** permite o monitoramento em massa, de vários usuários simultâneos. Para a realização da extração dos dados foi necessária a utilização da **Streaming API** com a leitura da **Public Streams**.

Em conjunto com a **Coordenadoria de Conectividade e Convergência Digital da Secretaria Municipal de Serviços de São Paulo**, foi definida a lista de **palavras-chave** que seriam acompanhadas nas mensagens do Twitter. A listagem de palavras-chave também denominada de **termos de busca** está disponível no **ANEXO I**.

Com a definição das palavras a serem buscadas e a utilização da **Public Streams** foi possível localizar e extrair todos os Tweets relacionados com nossa pesquisa em tempo real nos limites geográficos da cidade. Após o processamento, os dados são gravados em um banco de dados para uma análise posterior.

A análise dos dados capturados é realizada regularmente, de forma a criar uma série histórica consistente com possibilidade de realizar estudos comparativos de maneira horizontal. Os dados coletados são agrupados pelos temas utilizados, o que permite mensurar a frequência com que os temas são mencionados. Outra análise relevante, baseada em uma amostra aleatória dos Tweets do período, é uma *análise de intenção*, onde as mensagens são classificadas de acordo com seu teor positivo, negativo ou neutro.

Todo o procedimento de extração através da **Streaming API** com a **Public Streams** será melhor detalhado na seção da *Metodologia*.

3. METODOLOGIA

3.1. EXTRAÇÃO DE DADOS DO TWITTER

Para determinar progressivamente as mudanças temporais das mensagens coletadas, por exemplo, uma variação no volume de comentários por um determinado assunto ou variação de opiniões destas mensagens, é necessário a extração e análise dos dados em períodos regulares, mensais e quinzenais. A cada período, uma análise deve ser realizada, permitindo a posterior comparação entre as séries históricas dos dados compilados.

Para acessar o conteúdo público do Twitter foi desenvolvido um robô que verificou automaticamente os Tweets à medida em que eles eram publicados de acordo com determinados termos de busca (palavras-chave). Segundo a documentação do Twitter (disponível em <http://dev.twitter.com>), existem duas funcionalidades para leitura das mensagens:

O “search/tweets”, da **REST API**, cuja busca por tweets é baseada em uma solicitação padronizada de acesso aos dados. Entretanto, a documentação ressalta que esta funcionalidade não é adequada para coleta de dados em tempo real. Além disso, a fonte de dados deste recurso é composta por mensagens indexadas⁶ de acordo com os parâmetros de popularidade e relevância do Twitter, o que restringe a ampla busca das mensagens.

O “statuses/filter”, da **Streaming API**, cuja busca por mensagens é baseada em uma lista de termos. Nesta opção, é possível ter acesso em tempo real a todas as mensagens públicas sem restrições envolvendo indexação. No entanto, para a devida utilização deste recurso, é necessário que o robô automatizado possua tecnologia para lidar com sobrecargas de memória e processamento, além de manter a leitura ininterrupta dos dados, garantindo segurança e integridade das informações.

⁶ **Indexar** significa criar índices para algo, a fim de aumentar a velocidade de leitura de uma informação posteriormente.

Portanto, o endpoint⁷ escolhido para o desenvolvimento do robô foi o “statuses/filter”, usado com a **Streaming API**. O acesso a esta plataforma pode ser realizado por diversas linguagens de programação, entre elas o PHP⁸, a linguagem elegida para os estudos deste projeto.

Uma observação importante sobre a implementação da linguagem PHP é a utilização do CLI⁹, que possui vantagens em relação aos servidores HTTP convencionais. Executar o script¹⁰ em uma requisição HTTP traz algumas complicações como tempo de conexão entre cliente (browser, curl) e servidor (Apache, IIS), monitoramento da conexão para realizar uma nova requisição no caso de haver falhas, entre outras. Como é necessário que a conexão com o Twitter, na Streaming API, seja contínua e por tempo indeterminado, usar um processo em linha de comando como o CLI é traz maior confiança e disponibilidade de recursos para o monitoramento.

Para realizar a conexão das funcionalidades do PHP com o Twitter, pode ser utilizada uma biblioteca de acesso a API do site que já possui recursos de gestão de conexões, recuperação de falhas e outros recursos importantes para garantir o monitoramento contínuo. Uma das ferramentas mais utilizadas para este fim é o **Phirehose**.

O **Phirehose** traz em sua documentação algumas recomendações para o melhor aproveitamento do recurso. Entre elas, o enfileiramento dos JSONs recebidos, que são as mensagens padronizadas recebidas

⁷ **Endpoint** é uma operação específica dentro de uma API.

⁸ **PHP (Hypertext Preprocessor)** é uma linguagem de programação voltada para o desenvolvimento de conteúdo dinâmico para páginas e sistemas da Internet.

⁹ **CLI (Command Line Interface)** uma interface para execução de algum script via linha de comando. Pode ser através de um terminal, em sistemas Unix, ou do prompt de comando, em sistemas Windows, ou qualquer outra plataforma que permita executar o script diretamente, passando parâmetros para ele.

¹⁰ **Script**, arquivo que contém o algoritmo na linguagem de programação utilizada, no caso, os scripts são os algoritmos de extração em PHP.

pela API do Twitter. Com o enfileiramento e a possibilidade de processamento contomitante a coleta, foi necessário também o uso de threads¹¹ para não gerar erros no processamento para a leitura

Para gravar os JSONs coletados foi utilizado PDO¹² com uma conexão num banco de dados com tecnologia MySQL¹³. Todos os dados coletados foram gravados integralmente no banco de dados para permitir análises mais detalhadas posteriormente.

Para a implementação do **Phirehouse**, foi criada uma classe¹⁴ chamada *TwitStream* com os recurso disponíveis no **Phirehouse** em *OauthPhirehose*. A herança¹⁵ dessa classe implica na implementação do método *enqueueStatus*, responsável pelo enfileiramento dos JSONs recebidos da API. Para inicialização do método de leitura de dados, é necessário informar credenciais de autenticação, definir a lista dos termos de busca com o método *setTrack* e iniciar efetivamente com o método *consume*.

Para usar as credenciais de autenticação, as seguintes etapas foram realizadas:

- Criação de conta no Twitter.
- Registro de uma aplicação pelo endereço <https://apps.twitter.com/>.
- Dentro da página da nova aplicação criada, em API Keys, foi gerado uma chave de acesso.
- No script, às constantes "TWITTER_CONSUMER_KEY", "TWITTER_CONSUMER_SECRET", "OAUTH_TOKEN" e "OAUTH_SECRET" foram atribuídos, respectivamente, "API key", "API secret", "Access token" e "Access token secret", que são mostrados na página da

¹¹ **Thread**: Uma operação dentro de um programa, que é executada a parte, sem que a execução principal dele pare para esperar o processamento dela acabar. Vários threads podem ser executados ao mesmo tempo, sendo que uma não interfere na outra e nem na principal.

¹² **PDO: PHP Data Objects** é uma extensão do PHP que define uma interface de acesso a um banco de dados.

¹³ **MySQL** é um SGBD (Sistema de Gerenciamento de Banco de Dados) que utiliza a linguagem SQL (Linguagem de Consulta Estruturada) como interface.

¹⁴ **Classe** em programação orientada a objetos é uma estrutura que contém um conjunto de objetos com característica similares.

¹⁵ **Herança** é um princípio da programação orientada a objetos que permite que classes compartilhem atributos e métodos,

aplicação. A escolha do nome dessas constantes foi feita com base nos exemplos do Phirehose.

Na função *enqueueStatus*, foi feito o uso de um Worker, que é um objeto responsável por empilhar tarefas (Stackable), para serem executadas à medida que forem sendo adicionadas à pilha. Os Stackables foram implementados com a classe *TwitProcessor*, onde ocorre o processamento dos JSONs recebidos da API. Cada *TwitProcessor* tem um indicador para verificar se o processamento foi concluído, assim é possível limpar a pilha do Worker, evitando uma sobrecarga na memória.

No método “run” do *TwitProcessor*, o JSON recebido é inserido no banco de dados, sem modificações, através de uma conexão MySQL com PDO. Após a inserção no banco, a tarefa é concluída e marcada com a flag “done”. Com futuras análises das informações que vêm em cada JSON, será possível adicionar mais procedimentos ao *TwitProcessor* para refinar a estruturação dos dados.

Existem dois processos concomitantes para a leitura dos dados do Twitter, um deles responsável por filtrar os Tweets em relação às coordenadas geográficas da cidade de São Paulo e outro para captura de todas as mensagens enviadas de acordo com a às coordenadas geográficas de todas as praças do programa WiFi Livre SP.

As duas extrações simultâneas serão analisadas separadamente após o término de sua coleta.

3.2. ANÁLISE DOS DADOS EXTRAÍDOS

Para a análise dos dados, são utilizadas as mensagens capturadas pelo robô em intervalos regulares de tempo. As abordagens de análise podem ser descritas desta maneira:

- 1. Das mensagens contendo as palavras-chave dentro do perímetro da cidade de São Paulo:**
 - 1.1.** Análise de uso de palavras-chave e frequência;
 - 1.2.** Classificação de sentimento segundo análise amostral;
 - 1.3.** Análise de termos relacionados ao programa WiFi Livre;
 - 1.4.** Análise de termos chave para a prefeitura de São Paulo;

1.5. Relações entre as palavras-chave utilizadas através de grafo.

2. Das mensagens originadas nas localidades das praças digitais:

2.1. Análise de frequência de utilização das praças;

2.2. Relação entre a utilização das praças através de grafo.

Para a análise das mensagens contendo as palavras-chave, descrevemos os métodos listados da seguinte forma:

➤ **Análise de uso de palavras-chave e frequência:**

Uma lista de ocorrências das palavras-chave encontradas nas mensagens e o cálculo de sua porcentagem em relação ao total de mensagens obtidas pela extração.

➤ **Classificação de sentimento segundo análise amostral:**

É retirada uma amostra baseada no número total de mensagens obtidas e as classificamos de acordo com o sentimento que a mensagem passa em relação ao assunto caracterizado.

Para a amostra, são capturados de maneira aleatória um número de Tweets de acordo com a equação:

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

Onde:

n é a amostra calculada;

N é a população, em nosso caso, a quantidade de Tweets extraídos no período;

Z é a variável normal padronizada associada ao nível de confiança, tem valores tabelados exibidos na tabela abaixo;

p é a probabilidade do evento ser verdadeiro, ou seja, a quantidade de acertos esperada;

e é o nível de precisão ou erro amostral.

Nível de Confiança	Valor de Z
99%	2,57
95%	1,96
90%	1,64
80%	1,28

Tabela 1: Relação entre nível de confiança e valores da variável normal padronizada.

Podemos definir todos os valores previamente, com exceção do tamanho da população, portanto, os valores utilizados nessa equação são:

- Nível de Confiança de 99%, logo, pela Tabela 1, $Z = 2,57$;
- Probabilidade do evento ser verdadeiro, ou seja, do comentário ser ou não útil para a análise, $p = 50$;
- Erro amostral comumente é utilizado entre 3% e 10%, normalmente se usa 5%, logo, $e = 5$.

Após o cálculo da amostra ter sido realizado, retiraremos um número n de Tweets do banco de dados utilizando o PHP e o MySQL.

Desse modo, os Tweets são extraídos e submetidos a uma análise manual. Nela é realizada a categorização dos assuntos contidos na lista de palavras-chave e a **análise de sentimento** em relação aos assuntos que pode ser positiva, negativa, neutra ou não relacionada:

- **Positivo** significa que o comentário representa um elogio ou uma aprovação ao assunto em questão;
- **Negativo** significa que o comentário representa uma crítica ou uma reprovação ao assunto em questão;
- **Neutro** significa que não é possível interpretar o sentimento do comentário em relação ao assunto;

- **Não Relacionado à Prefeitura** significa que o comentário tem o mesmo significante do termo buscado, mas refere-se a um outro contexto, sem nenhuma relação com a Prefeitura.

Desta forma, é possível analisar a intenção dos usuários ao criar mensagens relacionadas aos diversos assuntos listados na lista de palavras-chave.

➤ **Análise de termos relacionados ao programa WiFi Livre:**

É buscado em todas as mensagens palavras-chave relacionadas ao programa WiFi Livre. Os resultados são analisados e através dessa análise é possível identificar qual o impacto do programa nas redes sociais.

As palavras-chave utilizadas são “WiFi Livre”, “WiFi SP”, “Praças Digitais” e “Praças Digitais SP”.

➤ **Análise de termos chave para a Prefeitura de São Paulo:**

São analisados todos os Tweets relacionados à alguns termos específicos que tenham relação com a aprovação do trabalho da prefeitura na cidade.

Esses termos são “Ciclovias”, “Haddad”, “Praças”, “Prefeitura” e “Subprefeitura”.

➤ **Relações entre palavras-chave utilizadas através de grafo:**

Utilizando um grafo, as palavras chave dos Tweets serão relacionadas entre si, permitindo realizar uma análise entre palavras-chave que ocorrem em um mesmo Tweet ou que comumente aparecem juntas.

Para a análise das mensagens dentro das localizações das praças, descrevemos os métodos listados da seguinte forma:

➤ **Análise da frequência de utilização das praças:**

Uma lista com o número de ocorrências de utilização das praças e sua porcentagem em relação ao número total de mensagens de todas as praças.

➤ **Relação entre a utilização das praças através de grafo:**

Utilizando um grafo, os usuários e as praças serão relacionados entre si, permitindo verificar a quantidade de usuários que utilizaram as praças e a conexão entre as praças, ligadas por usuários que utilizaram mais de uma praça.

4. ANÁLISE DOS DADOS

Após o término do período de extração de dados, que foi realizada entre os dias 16 e 31 de Janeiro, foram obtidos 22.490 Tweets nas buscas por termos e 16.458 Tweets na busca pelas praças digitais.

4.1. ANÁLISE DA EXTRAÇÃO POR TERMOS

4.1.1. ANÁLISE DE USO DE PALAVRAS-CHAVE E FREQUÊNCIA

A distribuição percentual dos termos captados é:

Posição	Palavra-Chave	Ocorrências	Porcentagem
1	Luz	6066	26,97%
2	Lixo	5199	23,12%
3	Wifi	2460	10,94%
4	Escuro	1241	5,52%
5	Fio	761	3,38%
6	Bike	701	3,12%
7	Bicicleta	543	2,41%
8	Poste	453	2,01%
9	Roubo	427	1,90%
10	Prefeitura	413	1,84%
11	Haddad	353	1,57%
12	Mendigo	323	1,44%
13	Masp	286	1,27%
14	Assalto	257	1,14%
15	Biblioteca	213	0,95%
16	Mercado Municipal	192	0,85%
17	Sujeira	183	0,81%
18	Largo Da Batata	170	0,76%
19	Centro Cultural	125	0,56%
20	Ciclovía	117	0,52%
21	Furto	43	0,19%
22	Morador De Rua	31	0,14%
23	Subprefeitura	30	0,13%
24	Reciclagem	24	0,11%
25	Parque Da Luz	21	0,09%

26	Poda	21	0,09%
27	Escuridao	19	0,08%
28	Parque Dom Pedro li	14	0,06%
29	Zelador	13	0,06%
30	Largo Do Arouche	13	0,06%
31	Largo Do Paissandu	10	0,04%
32	Imundice	10	0,04%
33	Seguranca	9	0,04%
34	Praca Ramos De Azevedo	7	0,03%
35	Jardinagem	7	0,03%
36	Largo Da Matriz	6	0,03%
37	Cidade De Sao Paulo	6	0,03%
38	Prefeito De Sp	5	0,02%
39	Pracas	4	0,02%
40	Parque Raul Seixas	4	0,02%
41	Wifi Livre	4	0,02%
42	Parque Da Aclimacao	4	0,02%
43	Parque Santo Dias	3	0,01%
44	Parque Chico Mendes	2	0,01%
45	Largo Dona Ana Rosa	2	0,01%
46	Praca Da Liberdade	2	0,01%
47	Praca Da Sé	2	0,01%
48	Prefeito Sp	2	0,01%
49	Telecentro	2	0,01%
50	Internet Sem Fio	2	0,01%
51	Parque Linear Tiquatira	2	0,01%
52	Pateo Do Colegio	1	0,00%
53	Praca Do Samba	1	0,00%
54	Prefeitura Sp	1	0,00%
55	Praia Do Sol	1	0,00%
56	Largo Do Cambuci	1	0,00%
57	Prefeitura De S. Paulo	1	0,00%

Tabela 2: Ocorrências e Porcentagens das palavras-chave utilizadas.

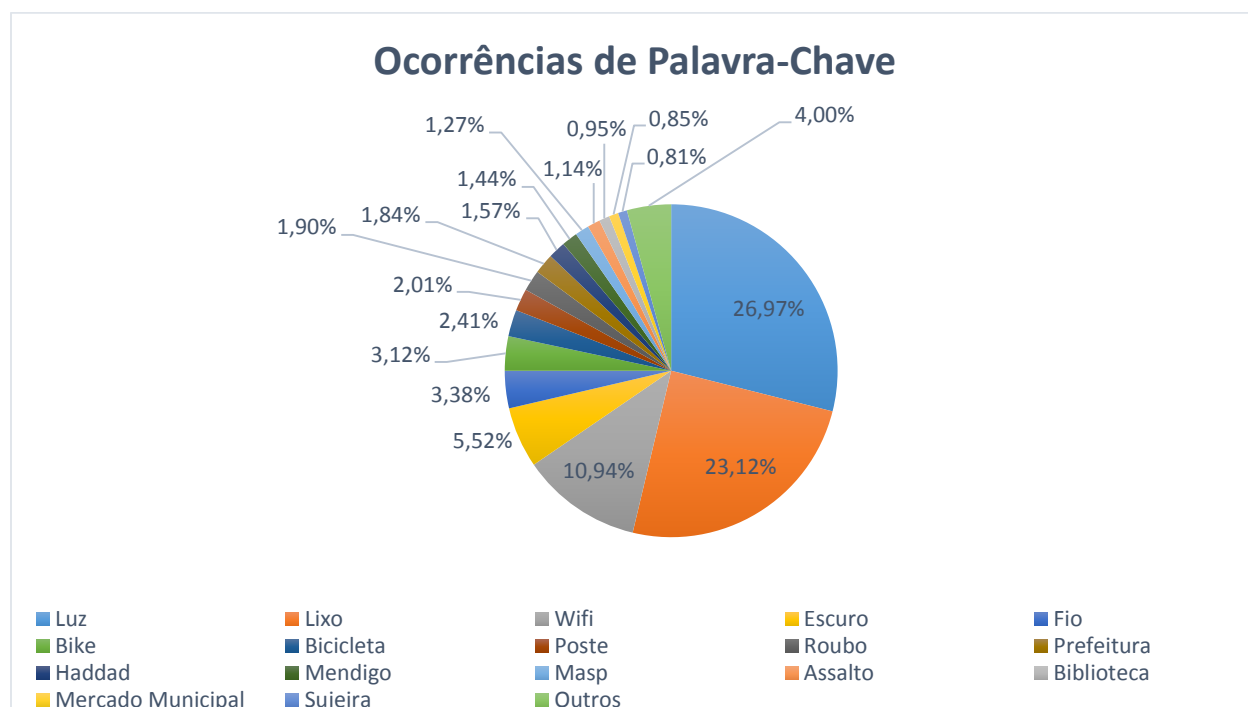


Figura 1: Gráfico de porcentagens de ocorrências das palavras-chave.

Olhando os valores percentuais e também a distribuição gráfica disso, podemos observar que dois termos são muito maiores que os outros e que somados formam uma parcela superior à 50% dos Tweets, que são **Luz** e **Lixo**.

Por serem muito genéricos e portarem uma grande diversidade semântica, os termos de busca **lixo** e **luz** tiveram um número de captações desproporcional se comparado com as outras palavras-chave utilizadas.

4.1.2. CLASSIFICAÇÃO DE SENTIMENTO SEGUNDO ANÁLISE AMOSTRAL

Foi calculada uma **amostra de 642 desses Tweets** e a partir da análise da mesma podemos ter uma base do que vem sendo falado sobre esses termos na última quinzena.

Abaixo, temos uma tabela com uma síntese das análises feitas com a amostra dos dados:

Assunto	Ocorrências	Positivos	Negativos	Neutros	Não Relacionados
Ciclovias	3	0	0	3	0
Cultura	1	1	0	0	0
Infraestrutura	6	1	3	0	2
Lugares	5	4	1	0	0
Luz	31	0	20	9	2
Obras	1	0	1	0	0
Política	11	0	7	1	3
Saúde	1	0	0	1	0
Segurança	9	1	5	3	0
Trânsito	1	0	0	1	0
Transporte	6	0	2	2	2
WiFi*	27	-	-	-	-
Ruído	540	-	-	-	-

Tabela 3: Análise de sentimento da amostra calculada.

* Durante a análise da amostra, notou-se um tipo de Tweet que foi repetido 27 vezes numa amostra de 642 (4,2%), usuários comentando sobre a qualidade de sua rede WiFi privada.

Por exemplo, o usuário @_carolinasouto disse: “q raiva desse wifi”, enquanto o usuário @leleochaves disse: “Não tô nem acreditando que tô com wifi agora”.

Dentro dessa amostra é possível observar o tamanho do ruído (cerca de 84% do total), que apresenta esse valor muito elevado por dois motivos:

- A porcentagem muito elevada de termos suscetíveis à captura de ruído na pesquisa, como os termos Lixo e Luz;
- A aleatoriedade dos Tweets capturados para a amostra.

Se utilizássemos outra amostra aleatória, teríamos números compatíveis, mas talvez uma certa redução da porcentagem de ruído e aumento de outros assuntos.

Sobre os assuntos que entraram na amostra, podemos concluir que:

- Uma grande parcela (cerca de 5%) realizaram comentários negativos ou críticos ao sistema de energia elétrica e luz;
- Outra grande parte das postagens eram de opiniões políticas, a maioria eram negativas à prefeitura de São Paulo de forma geral.

4.1.3. ANÁLISE DE TERMOS RELACIONADOS AO PROGRAMA WIFI LIVRE

Foi realizada a leitura de todos os Tweets que continham referência direta ao programa WiFi Livre. Utilizando os termos “WiFi Livre”, “WiFi SP”, “Praças Digitais” e “Praças Digitais SP”, obtivemos apenas um resultado que tem relação com o programa WiFi Livre, o Tweet é o seguinte: “*WiFi livre de São Paulo salvando vidas*”, enviado pelo usuário @barbosiix.

Com base na coleta de mensagens realizada na rede social de formadores de opinião, o Twitter, durante a segunda quinzena de janeiro, podemos concluir que o Programa WiFi Livre não é um tema relevante para a opinião pública paulistana. Com a continuidade mensal de dados captados utilizando o mesmo método de extração e aprimorando a eliminação de ruídos será possível observar a evolução da percepção dos formadores de opinião sobre o programa WiFi Livre.

Também é possível concluir que o Programa WiFi Livre que está sendo implantado é pouco conhecido da opinião pública paulistana. Um programa novo e de conectividade possui um grande potencial de impacto mais positivo entre os moradores do Município.

4.1.4. ANÁLISE DE TERMOS CHAVE PARA A PREFEITURA DE SÃO PAULO

Escolhemos alguns termos utilizados na pesquisa que poderiam dar informações sobre a popularidade e aprovação da prefeitura, relacionada com alguns assuntos específicos. Esses termos foram: **Ciclovía, Haddad, Praças, Prefeitura e Subprefeitura.**

Ciclovía

Foram obtidos 117 Tweets sobre o assunto Ciclovía, a distribuição de opinião sobre esse assunto se dá da seguinte forma:

Avaliação	Ocorrências	Porcentagem
Positiva	20	17%
Negativa	26	22%
Neutra	52	44%
Não Relacionada	19	16%
Total	117	

Tabela 4: Análise de sentimento do termo Ciclovía.

O que mostra que as opiniões são divididas, mas a maioria dos usuários não criticam, nem elogiam, apenas notificam o uso das ciclovias ou comentam a respeito. Um gráfico que mostra de forma melhor a divisão dessa opinião é mostrado abaixo.

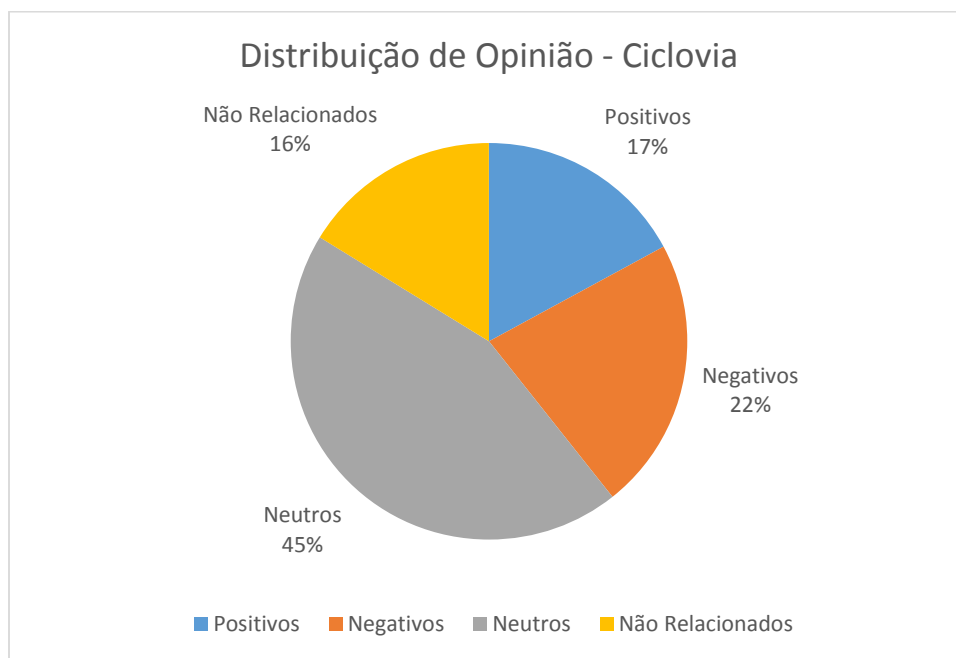


Figura 2: Distribuição de opinião para o termo Ciclovía.

Haddad

Para este termo, foram encontrados 353 Tweets. Em sua maioria esses Tweets se referiam de forma negativa à ciclovias e a nomeação do ex-senador Eduardo Suplicy e do candidato a governador do estado de São Paulo, Alexandre Padilha à secretaria da prefeitura de São Paulo.

A distribuição de opinião desses Tweets é dada na tabela abaixo:

Avaliação	Ocorrências	Porcentagem
Positiva	32	9%
Negativa	171	48%
Neutra	93	26%
Não Relacionada	57	16%
Total	353	

Tabela 4: Análise de sentimento do termo Haddad.

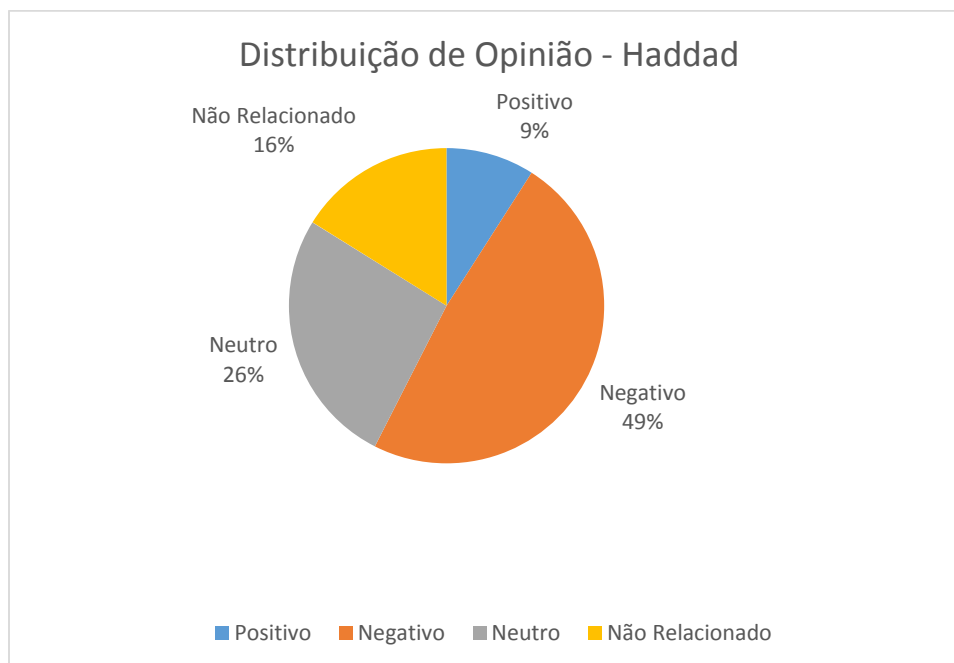


Figura 3: Distribuição de opinião para o termo Haddad.

Praças

Para o termo Praças, tivemos 401 ocorrências. Como alguns dos outros termos apresentados, os resultados dessa não tratam nem de forma positiva, nem de modo negativo o estado atual das praças. A maior parte dos Tweets relacionados ao termo são marcações de localização através da ferramenta de geolocalização do Twitter.

A distribuição se dá na tabela a seguir:

Avaliação	Ocorrências	Porcentagem
Positiva	3	1%
Negativa	2	0,5%
Neutra	331	82,5%
Não Relacionada	65	16%
Total	401	

Tabela 5: Análise de sentimento do termo Praças.

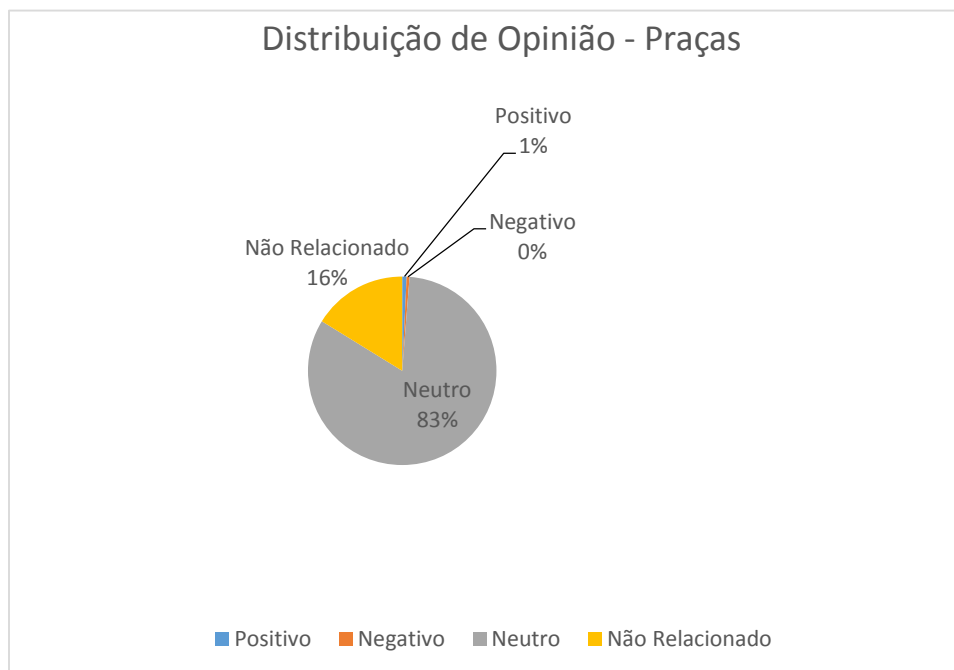


Figura 4: Distribuição de opinião para o termo Praças.

Prefeitura

Para o termo “Prefeitura”, obtivemos 413 ocorrências. Os resultados foram variados, desde comentários sobre o trabalho em relação às ciclovias, o passe livre para estudantes, até o debate sobre as manifestações contra o aumento da tarifa do transporte público. Muitos comentários tiveram apenas caráter informativo, sem expressar opinião e muitos sem conexão com a prefeitura de São Paulo.

A distribuição da opinião sobre esse termo é mostrada na tabela abaixo:

Avaliação	Ocorrências	Porcentagem
Positiva	16	3,9%
Negativa	89	21,5%
Neutra	144	34,9%
Não Relacionada	164	39,7%
Total	413	

Tabela 6: Análise de sentimento do termo Prefeitura.

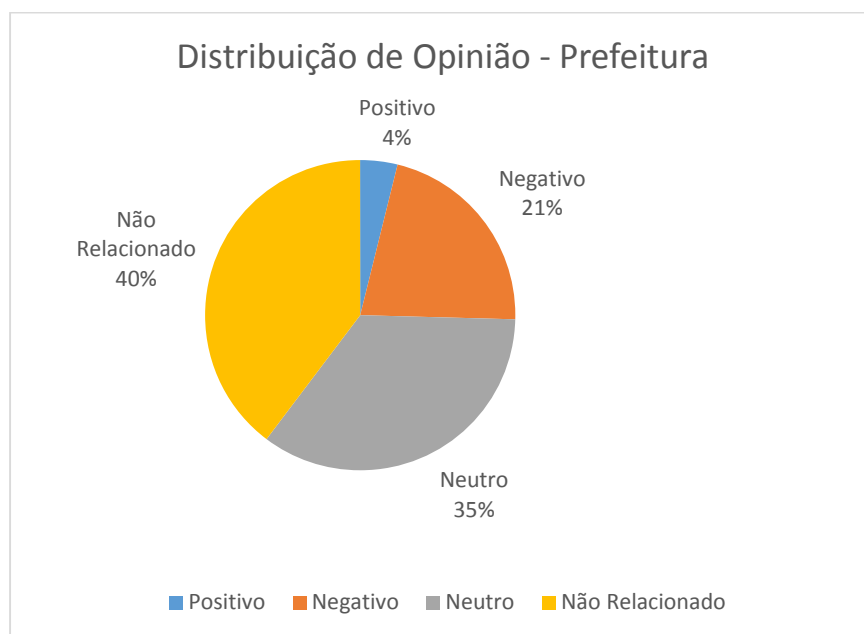


Figura 5: Distribuição de opinião para o termo Prefeitura.

Subprefeitura

Sobre o termo “Subprefeitura” foram encontrados 30 tweets. A maior parte desses tweets foi apenas marcações de localização, sem apresentar muito conteúdo para avaliar uma opinião geral sobre este termo.

A distribuição de opiniões se dá a seguir:

Avaliação	Ocorrências	Porcentagem
Positiva	0	0%
Negativa	2	7%
Neutra	18	60%
Não Relacionada	10	33%
Total	30	

Tabela 6: Análise de sentimento do termo Subprefeitura.

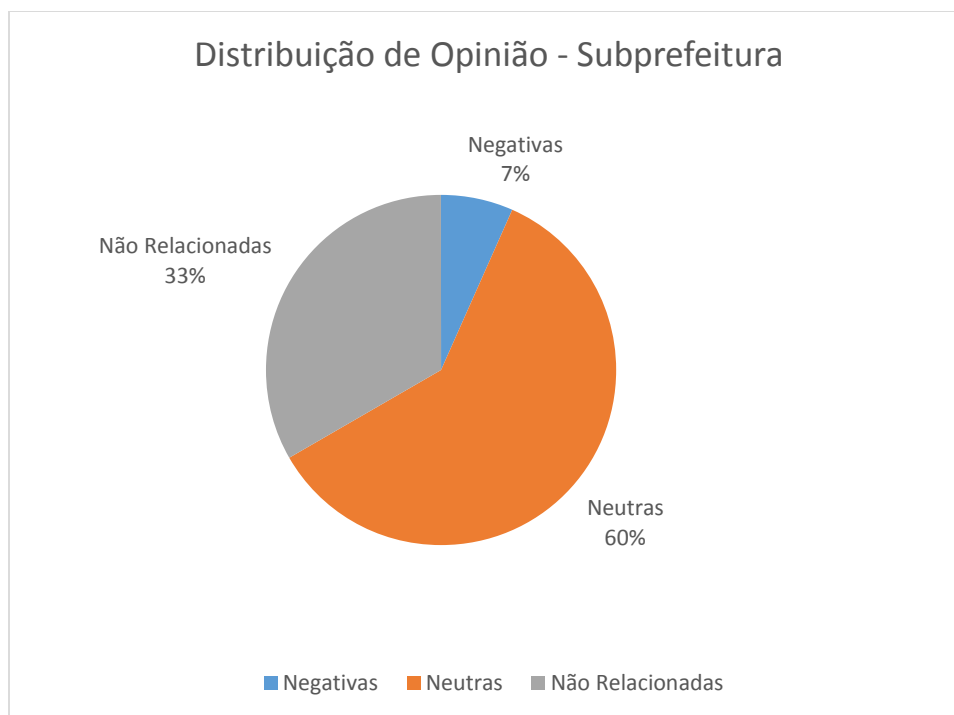


Figura 6: Distribuição de opinião para o termo Subprefeitura.

4.1.5. RELAÇÕES ENTRE PALAVRAS-CHAVE ATRAVÉS DO GRAFO

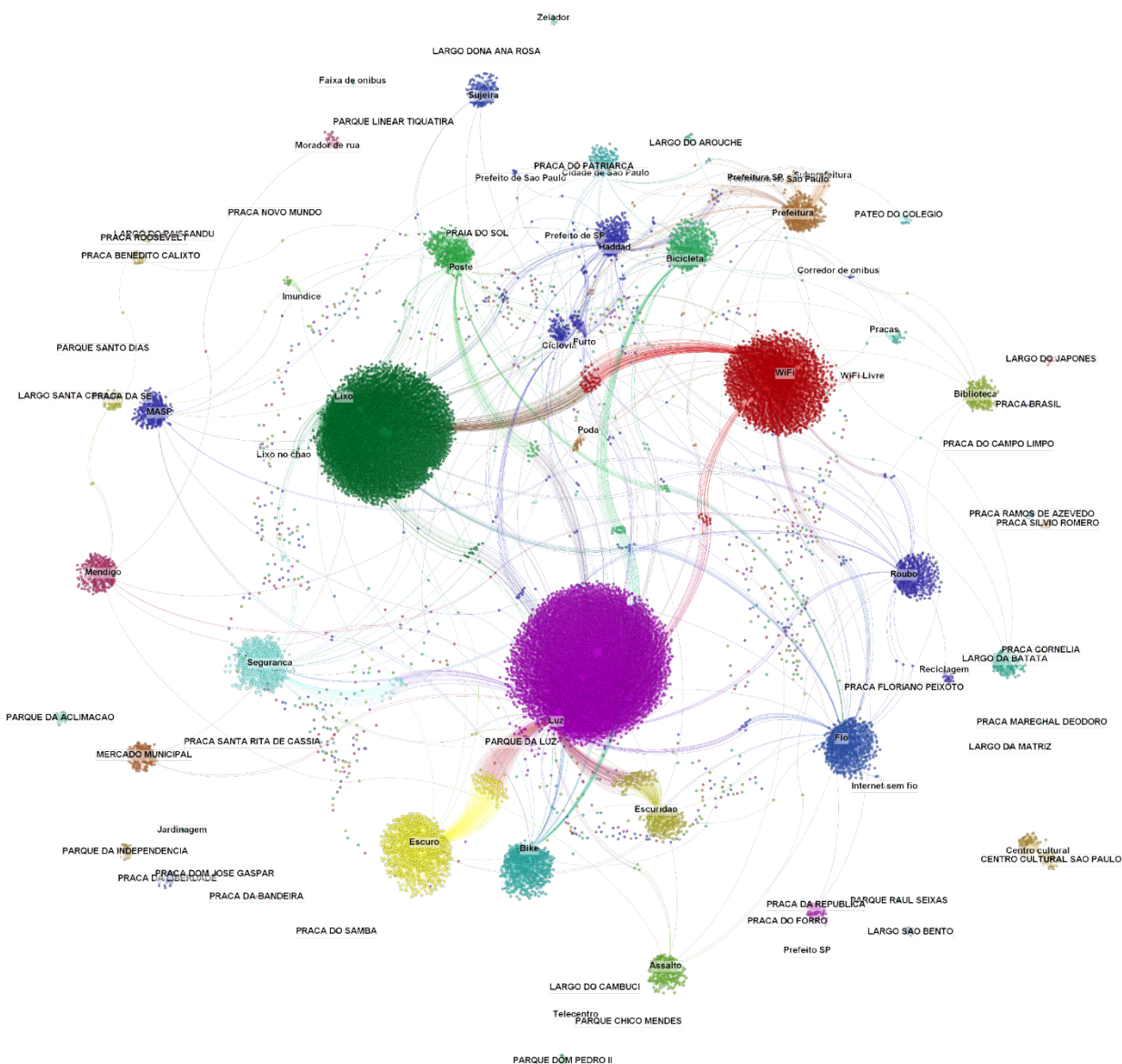


Figura 7: Grafo de relacionamento de Palavras-Chave.

Neste grafo de relacionamento entre palavras-chave utilizadas nas mensagens originadas dentro do perímetro da cidade de São Paulo, pode ser facilmente verificado a formação de clusters, ou seja, aglomerações bem definidas de mensagens sobre determinado assunto. Pela análise do grafo, os assuntos

com maior frequência de palavras-chave foram "Luz", "Lixo" e "Wifi", dado condizente com a análise de frequência constante na página 20, 21 e 22;

Algumas mensagens, no entanto, abordam mais de um assunto, e por isso possuem interconexões. A disposição espacial do relacionamento dos pontos do grafo facilita a compreensão do relacionamento entre eles. Por exemplo, verifica-se que o termo "Bicicleta" é um dos termos de maior interconexão com o termo "Haddad", dentre os termos selecionados no escopo desta pesquisa. Outro termo, "MASP", possui relações com os termos "Imundice", "Lixo" e "Bike". A utilização do termo "WiFi Livre" foi inexpressiva e não relacionada aos outros termos.

4.2. ANÁLISE DA EXTRAÇÃO POR PRAÇAS

4.2.1. ANÁLISE DA FREQUÊNCIA DE UTILIZAÇÃO DAS PRAÇAS

A distribuição percentual da utilização das praças é:

Praça	Ocorrências	Porcentagem
PRAÇA DOM JOSÉ GASPAR	2145	13,03%
PARQUE LINEAR TIQUATIRA	1705	10,36%
PARQUE RAPOSO TAVARES	1648	10,01%
PRAÇA DA SÉ	1126	6,84%
PRAÇA DA BANDEIRA	1067	6,48%
PARQUE ORLANDO VILLAS BÔAS	975	5,92%
PARQUE DA LUZ	870	5,29%
PRAÇA DA REPÚBLICA	838	5,09%
MASP	673	4,09%
PRAÇA ROOSEVELT	423	2,57%
LARGO DO AROUCHE	361	2,19%
PRAÇA RAMOS DE AZEVEDO	350	2,13%
MERCADO MUNICIPAL	311	1,89%
CENTRO CULTURAL SÃO PAULO	264	1,60%
PARQUE DA INDEPENDÊNCIA	261	1,59%
PARQUE DOM PEDRO II	180	1,09%
PRAÇA CRAVEIRO DO CAMPO	178	1,08%
PRAÇA VEREADOR ANTONIO SAMPAIO / AV ULTRAMARINO	149	0,91%
PRAÇA BENEDITO CALIXTO	147	0,89%

PARQUE DO NABUCO	140	0,85%
PRAÇA DA LIBERDADE	126	0,77%
PRAÇA DO PÔR-DO-SOL	124	0,75%
PRAÇA NOSSA SENHORA DOS PRAZERES	107	0,65%
PRAÇA FORTUNATO DA SILVEIRA / UNICSUL	107	0,65%
LARGO DO PAISSANDÚ	106	0,64%
PRAÇA SAMPAIO VIDAL	102	0,62%
PARQUE DA ACLIMAÇÃO	101	0,61%
PARQUE SANTO DIAS	99	0,60%
PRAÇA DON ORIONE	97	0,59%
PRAÇA NA RUA DAS IMBIRAIARAS	90	0,55%
LARGO SANTA CECILIA	86	0,52%
PÁTIO DO COLÉGIO	83	0,50%
PARQUE DOMINGOS LUÍS	82	0,50%
LARGO SÃO BENTO	79	0,48%
PARQUE ECOLÓGICO DA VILA PRUDENTE	73	0,44%
PRAÇA SILVIO ROMERO	69	0,42%
LARGO DA BATATA	60	0,36%
PRAÇA ALFREDO EGYDIO DE SOUZA ARANHA (PARQUE LINA E PAULO RAIA)	56	0,34%
LARGO DONA ANA ROSA	56	0,34%
CALÇADÃO CULTURAL DO GRAJAÚ	55	0,33%
LARGO DA CONCÓRDIA	53	0,32%
PRAÇA ESCOLAR	47	0,29%
PARQUE SANTA AMÉLIA	46	0,28%
PRAÇA DO 65 / POMBAS URBANAS	37	0,22%
PRAÇA TUNEY ARANTES	35	0,21%
PARQUE CHICO MENDES	33	0,20%
PRAÇA DO PATRIARCA	33	0,20%
PRAÇA DOUTOR JOÃO BATISTA VASQUES	33	0,20%
PRAÇA LUISA P DOS SANTOS (PRAÇA ARARIBA)	32	0,19%
PRAÇA PADRE ALEIXO (DO FORRÓ)	30	0,18%
PRAÇA NA RUA GREGÓRIO RAMALHO	29	0,18%
LARGO DE MOEMA	28	0,17%
LARGO SÃO FRANCISCO	26	0,16%
PRAÇA ENG. NORIWUKI YAMAMOTO	26	0,16%
LARGO DO CAMBUCI	26	0,16%
PARQUE RAUL SEIXAS	25	0,15%
PRAÇA ZILDA NATEL	19	0,12%

PRAÇA CONDE FRANCISCO MATARAZZO	19	0,12%
PRAÇA NOVO MUNDO	18	0,11%
POLO CULTURAL DO JARDIM NGELA	16	0,10%
PRAÇA ELIS REGINA	15	0,09%
PRAÇA ROTARY	15	0,09%
PRAÇA FLORIANO PEIXOTO	14	0,09%
PRAÇA CIRO PONTES / AO LADO DO SENAI	14	0,09%
PRAÇA BACHAREL FERNADO BRAGA PEREIRA DA ROCHA	13	0,08%
PARQUE ALFREDO VOLPI	13	0,08%
PRAÇA DA ESTAÇÃO VILA MARA	13	0,08%
PRAÇA SANTA RITA DE CÁSSIA	12	0,07%
PRAÇA DA CONQUISTA (PRAÇA VEREADOR JOÃO APARECIDO DE PAULA)	12	0,07%
PRAÇA MARIQUINHA SCIASCIA	10	0,06%
PRAÇA DO LARGO DE PIRAPORINHA	10	0,06%
PRAÇA PROFESSORAS (NA AV. DAS ALAMANDAS)	10	0,06%
PRAÇA VICENTE FALCETTA	9	0,05%
PRAÇA KANTUTA	8	0,05%
PRAÇA FELISBERTO FERNANDES DA SILVA / LARGO SÃO MATEUS	8	0,05%
PRAÇA DILVA GOMES MARTINS (COHAB 1)	8	0,05%
PRAÇA DIRCEU DE CASTRO FONTOURA	8	0,05%
PRAÇA LOURENÇO DE BELLIS	7	0,04%
TERMINAL CASA VERDE	6	0,04%
PRAÇA DO TERMINAL METALÚRGICOS	6	0,04%
PRAÇA OSVALDO LUÍS DA SILVEIRA	6	0,04%
LARGO DO JAPONES	6	0,04%
PRAÇA GEN. PORTO CARREIRO	5	0,03%
PRAÇA OSCAR DA SILVA	4	0,02%
PRAÇA ARLINDO ROSSI	4	0,02%
PRAÇA JOÃO BOLDO	4	0,02%
PRAÇA MARCO ANTONIO PRIMON MAESTRE	4	0,02%
PRAÇA ALEXANDRE FRANCISCO REBOUÇAS	3	0,02%
CDC CAMPO BELO	3	0,02%
PRAÇA JESUS TEIXEIRA DA COSTA	3	0,02%
PRAÇA FUHAD H. SMAIRE	3	0,02%
PRAÇA GIOVANI FANI	3	0,02%
PRAÇA JOÃO BEIÇOLA	2	0,01%
PRAÇA WILSON MOREIRA DA COSTA	2	0,01%
PRAÇA CORNÉLIA	2	0,01%

PRAÇA GENERAL GUIMARÃES	2	0,01%
PRAÇA JÚLIO CÉSAR DE CAMPOS	1	0,01%
PRAÇA SÃO LUIS DO CURU	1	0,01%
CENTRO DE CONVIVÊNCIA EDUCATIVO E CULTURAL DE HELIÓPOLIS	1	0,01%
PRAÇA AUGUSTO DOMINGUES ALVES MAIA	1	0,01%
PRAIA DO SOL	1	0,01%
LARGO DA MATRIZ	1	0,01%
PRAÇA PADRE DAMIÃO	1	0,01%
PRAÇA JOÃO TADEU PRIOLLI (PRAÇA DO CAMPO LIMPO)	1	0,01%
LARGO SANTA ANGELA	1	0,01%
PRAÇA JAGUAMITANGA	1	0,01%
PRAÇA GENERAL HUMBERTO DE SOUSA MELLO	1	0,01%

Tabela 7: Porcentagens de ocorrências das praças.

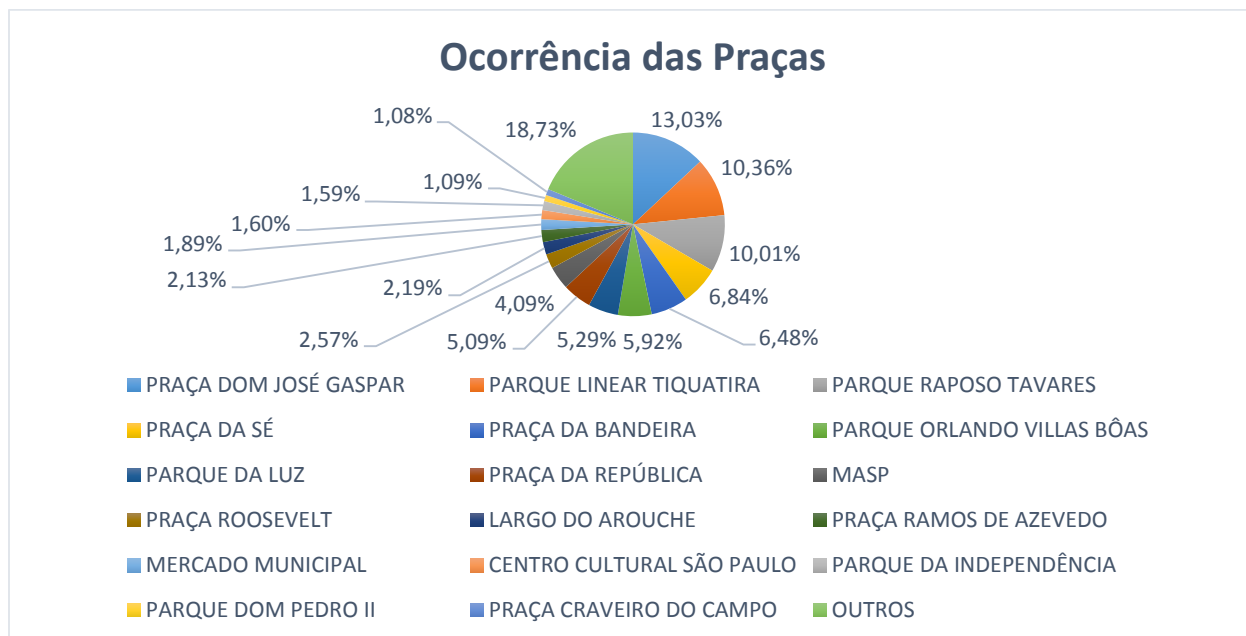
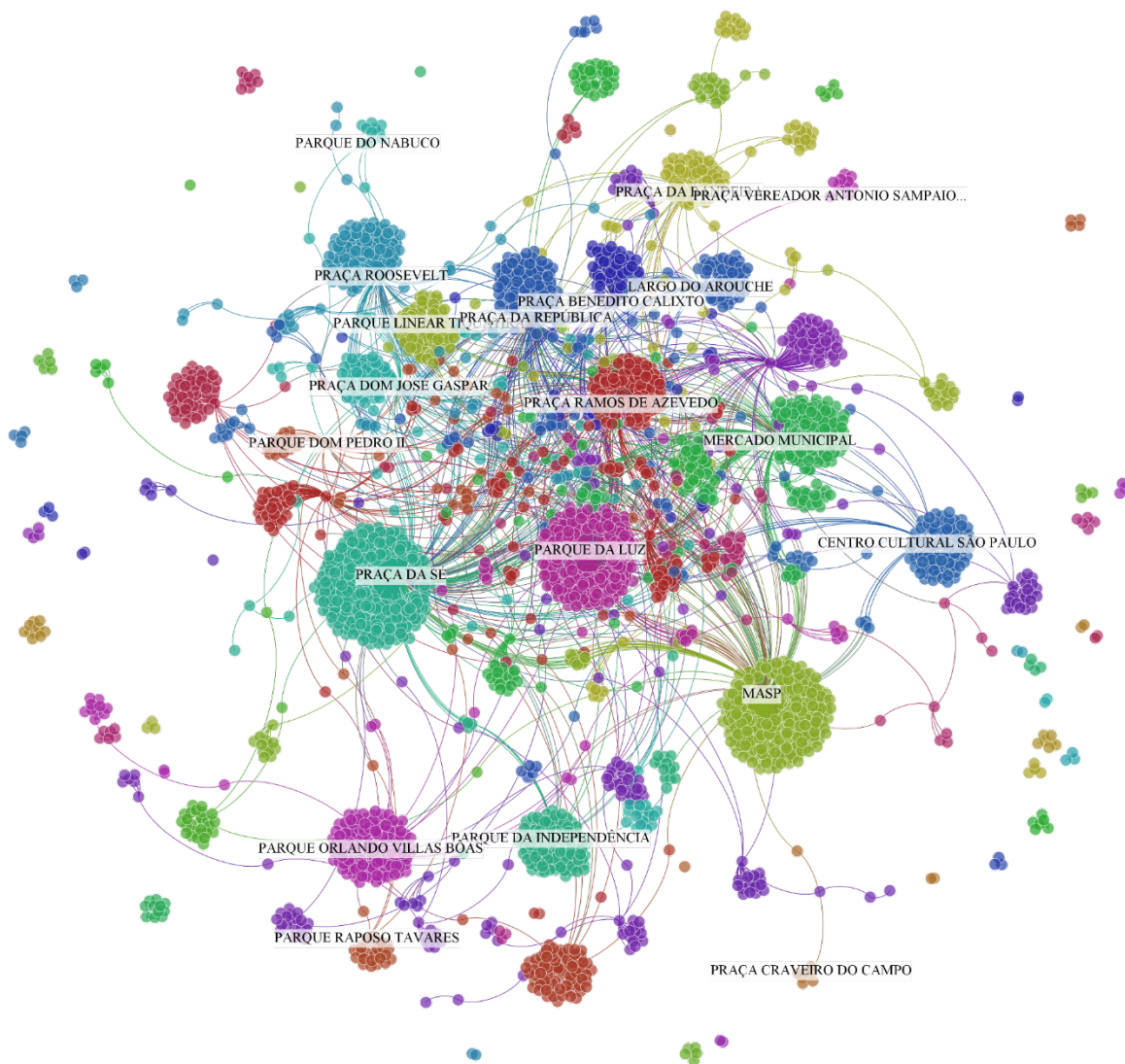


Figura 8: Distribuição de ocorrência de praças.

4.2.2. RELAÇÃO DA UTILIZAÇÃO DAS PRAÇAS ATRAVÉS DO GRAFO



Neste grafo são mostradas as relações entre usuários que utilizam o Twitter com o recurso de geolocalização habilitado e em qual praça ocorreu tal utilização. É possível verificar a formação de agrupamentos em relação às localidades onde já existe ou que será implantado o WiFi Livre.

Os maiores agrupamentos correspondem aos locais com maior utilização da rede social. Se destacam no grafo, a "Praça da Sé", "Parque da Luz", "MASP", "Parque Orlando Villas Boas", "Mercado Municipal" entre outras. É possível também verificar a movimentação dos usuários entre as praças. Por exemplo, pela relação existente entre "Praça da Sé" e "Parque da Luz" e "Masp" é possível verificar que dentro do período de análise, o mesmo usuário esteve em ambas localidades utilizando o Twitter.

5. REFERÊNCIAS

- Bailey, Fenn. 2011. Introduction - fennb/phirehose Wiki. *GitHub*. [Online] 2 de Agosto de 2011. [Citado em: 21 de Julho de 2014.] <https://github.com/fennb/phirehose/wiki/Introduction>.
- _____. 2013. phirehose/example at master - fennb/phirehose. *GitHub*. [Online] 27 de Dezembro de 2013. [Citado em: 21 de Julho de 2014.] <https://github.com/fennb/phirehose/tree/master/example>.
- BENKLER, Yochai. 2006. *The wealth of networks: how social production transforms markets and freedom*. New Haven: Yale University.
- CASTELLS, Manuel. 2009. *Communication Power*. New York: Oxford University Press Inc.
- Facebook. Facebook Platform Upgrade Guide. *Facebook Developers*. [Online] Facebook. [Citado em: 21 de 07 de 2014.] https://developers.facebook.com/docs/apps/upgrading#upgrading_v2_0_graph_api.
- _____. Using the Graph API. *Facebook Developers*. [Online] Facebook. [Citado em: 21 de 07 de 2014.] <https://developers.facebook.com/docs/graph-api/using-graph-api/v2.1#search>.
- HABERMAS, Jurgen. 1997. *Direito e democracia: entre facticidade e validade*. Rio de Janeiro: Tempo Brasileiro.
- Hitwise. 2014. Últimas Notícias. *Serasa Experian*. [Online] Serasa Experian, 25 de 06 de 2014. [Citado em: 02 de 09 de 2014.] <http://noticias.serasaexperian.com.br/facebook-e-lider-entre-redes-sociais-em-maio-no-brasil-de-acordo-com-hitwise/>.
- MAIA, Rousiley C.M. 2008. *Política deliberativa e reconsiderações acerca do conceito de esfera pública*. In: *Mídia e Deliberação / Rousiley C.M. Maia (coordenadora)*. Rio de Janeiro: Editora FGV.
- Pew Research Center. 2014. Emerging Nations Embrace Internet, Mobile Technology. *Pew Global*. [Online] Pew Research Center, 13 de 02 de 2014. [Citado em: 25 de 01 de 2015.] <http://www.pewglobal.org/2014/02/13/emerging-nations-embrace-internet-mobile-technology/>.
- Twitter. 2014. GET search/tweets. *Twitter Developers*. [Online] Twitter, 12 de Junho de 2014. [Citado em: 20 de Julho de 2014.] <https://dev.twitter.com/docs/api/1.1/get/search/tweets>.
- _____. 2013. Public streams. *Twitter Developers*. [Online] Twitter, 13 de 03 de 2013. [Citado em: 20 de Julho de 2014.] <https://dev.twitter.com/docs/streaming-apis/streams/public>.
- _____. 2012. The Streaming APIs. *Twitter Developers*. [Online] Twitter, 24 de Setembro de 2012. [Citado em: 20 de Julho de 2014.] <https://dev.twitter.com/docs/api/streaming>.

Watkins, Joe. 2014. pthreads/Stacking.php at master - krakjoe. *GitHub*. [Online] 12 de Março de 2014. [Citado em: 21 de Julho de 2014.]
<https://github.com/krakjoe/pthreads/blob/master/examples/Stacking.php>.

ANEXOS

Anexo 1 – LISTA DE PALAVRAS-CHAVE

Assalto;	Praças mais cuidadas;
Biblioteca;	Prefeitura;
Centro cultural;	Reciclagem;
Ciclovia;	Roubo;
Corredor de ônibus;	Secretaria de Serviços;
Escuridão;	Segurança;
Escuro;	Subprefeitura;
Faixa de ônibus;	Sujeira;
Faixa exclusiva de ônibus;	Telecentro;
Fio;	WiFi;
Furto;	WiFi Livre;
Haddad;	WiFi Livre SP;
Iluminação pública;	Zelador;
Imundice;	Zeladoria;
Internet sem fio;	Bike;
Jardinagem;	Bicicleta;
Lixo;	Prefeitura de São Paulo;
Lixo no chão;	Prefeitura de S. Paulo;
Luz;	Prefeitura SP;
Mendigo;	Prefeito de São Paulo;
Morador de rua;	Prefeito de S. Paulo;
Poda;	Prefeito SP;
Poste;	Prefeito de SP;
Praças;	Cidade de São Paulo;
Praças digitais;	

Praça Digital SP;

Inclusão Digital São Paulo;

Inclusão Digital SP.