

WILLIAM DE OLIVEIRA SOUZA

MONITORAMENTO E ANÁLISE DO MECANISMO DE INCENTIVO  
PRESIDÊNCIA EM ESTABELECIMENTOS DO APLICATIVO SWARM

*(Pre-examination version, compiled at 31 de Outubro de 2018)*

Proposta para o desenvolvimento do Trabalho de Conclusão do Curso de Bacharelado em Sistemas de Informação, apresentado à UTFPR como requisito parcial para obtenção do título de bacharel em Sistemas de Informação.

Field: *Sistemas de Informação*.

Advisor: Thiago Henrique Silva.

CURITIBA PR  
2018

# Resumo

Este trabalho descreve um estudo sobre redes sociais e os mecanismos de incentivo presentes para incentivar os usuários. O foco será dado para a rede social Foursquare, que possui uma extensão para *smartphones* especialmente para realizar *check-ins*, chamada *Swarm*. O *Swarm* também possui seus mecanismos de incentivo e entre esses mecanismos temos prefeitura e medalhas. Prefeitura é o termo utilizado para o usuário que mais realizou *check-ins* nos últimos 30 dias em determinado local. Através do aplicativo é possível detectar como ocorrem as alternâncias de prefeitos de locais e as possíveis causas. Isso é possível por meio de programas computacionais criados especificamente para este fim, denominados *web crawlers*. Estão sendo coletados diariamente dados de mais de 5.000 locais diferentes e de categorias distintas, para proporcionar uma análise rica do objeto de estudo. Com esses dados, esperamos entender melhor o mecanismo de incentivo de prefeitura do *Swarm*. Isso é importante por várias razões, por exemplo, outros sistemas interessados nesse mecanismo podem se beneficiar desse conhecimento na sua implementação. O diferencial para os demais trabalhos relacionados é o estudo aprofundado do mecanismo de incentivo chamado de prefeitura, e também pela quantidade superior de dados que estão sendo coletados, sendo aproximadamente 1.000 locais para cada uma das 5 cidades, entre Brasil, Estados Unidos e Japão.

**Palavras-chave:** Foursquare, Swarm, Prefeitura, Check-in, Mecanismo de incentivo.

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>7</b>
1.1	Objetivos Gerais . . . . .	8
1.1.1	Objetivos Específicos . . . . .	8
1.2	Organização do Documento . . . . .	8
<b>2</b>	<b>Referencial Teórico</b>	<b>10</b>
2.1	Redes Sociais Baseadas em Localização . . . . .	10
2.2	Mecanismos de Incentivo . . . . .	11
2.2.1	Medalhas . . . . .	11
2.2.2	Prefeitura . . . . .	12
2.3	Coleta de Dados Automática da Web . . . . .	12
2.4	Banco de Dados NoSQL . . . . .	13
<b>3</b>	<b>Estado da Arte</b>	<b>14</b>
<b>4</b>	<b>Metodologia</b>	<b>18</b>
4.1	Dados a Serem Coletados . . . . .	18
4.2	Coletor de Dados . . . . .	19
4.2.1	Obtenção dos Dados . . . . .	19
4.3	Infraestrutura . . . . .	19
4.3.1	Agendador de Tarefas . . . . .	20
4.3.2	Armazenamento de Dados . . . . .	20
4.4	Análise dos Dados . . . . .	20
<b>5</b>	<b>Análises</b>	<b>22</b>
<b>6</b>	<b>Recursos de Hardware e Software</b>	<b>24</b>
6.1	Recursos de Hardware . . . . .	24
6.2	Recursos de Software . . . . .	24
<b>7</b>	<b>Viabilidade e Cronograma Preliminar</b>	<b>25</b>

<b>8 Conclusão</b>	<b>27</b>
<b>Bibliografia</b>	<b>28</b>

# Lista de Figuras

2.1	Check-ins de um dia na mesma cidade de um determinado usuário . . . . .	10
2.2	Check-ins em cidades diferentes . . . . .	11
2.3	Exemplo de arquivo formato JSON . . . . .	13
4.1	Etapas para obtenção dos dados . . . . .	18
5.1	Disputa no Parque Barigui . . . . .	22
5.2	Disputa na Universidade Tecnológica Federal do Paraná . . . . .	22
5.3	<i>Check-ins</i> por categoria de local . . . . .	23
7.1	Cronograma de Atividades . . . . .	26

# Lista de Acrônimos

API	<i>Application Programming Interface</i>
HTML	<i>HyperText Markup Language</i>
IP	<i>Internet Protocol</i>
JSON	<i>JavaScript Object Notation</i>
LBSN	<i>Location-Based Social Networks</i>
NoSQL	<i>Not Only SQL</i>
SGBD	Sistema Gerenciador de Banco de Dados SQL
<i>Structured Query Language</i>	
PSN	<i>Participatory Sensor Network</i>

# Capítulo 1

## Introdução

Com a crescente utilização das redes sociais também aumentou o número de serviços baseados em redes sociais. Entre elas existe um tipo particular de rede social chamada rede social baseada em localização (LBSN, do inglês *Location-Based Social Network*) que é baseada em localização, assim como o Foursquare. Entre as LSBNs que mais cresceram nos últimos anos pode-se citar o Foursquare <sup>1</sup> que passou de 5 milhões de usuários em 2010 (Lindqvist et al., 2011) para 10 milhões de usuários em 2011 (Pontes et al., 2012), e tem como foco mapear estabelecimentos em geral para fornecer melhores informações para seus usuários, mas também permite a interação entre os usuários, o que acaba tornando essa rede social ainda mais utilizada não somente no Brasil como em muitos países como mostram Pontes et al. (2012).

Um dos diferenciais do Foursquare é a gamificação, que é a presença de aspectos frequentes em jogos, e que tem por objetivo gerar disputa entre os usuários, o que pode estimular o uso do sistema. O Foursquare possui uma plataforma para *smartphones* chamada *Swarm*, através da qual é possível acompanhar as atividades dos amigos, realizar *check-ins*, visualizar histórico de *check-ins* próprios e dos amigos, entre outras coisas. O aspecto da motivação se dá nas bonificações com que o aplicativo recompensa seus usuários. Dentre essas recompensas temos: moedas, medalhas e prefeituras.

Moedas são adquiridas de várias formas dentro do *Swarm*, realizar *check-in* em locais é apenas uma forma, mas a quantidade de moedas que se ganha depende bastante do local e da frequência com que são feitos esses *check-ins*. Semanalmente é gerado um *ranking* de moedas adquiridas somente entre os amigos para comparar quem obteve mais moedas na última semana, e de acordo com a classificação no *ranking*, os usuários recebem mais ou menos moedas.

Usuários adquirem medalhas de acordo com o *check-in*, em casos de eventos especiais, marcação de amigos nos *check-ins* e *check-ins* em diversas categorias de estabelecimentos. De acordo com as medalhas ganhas, é possível notar se um usuário pode ser referência para dicas em determinados locais. Por exemplo, um usuário que realiza *check-ins* em diversas categorias tende a ser uma fonte mais confiável para dar indicações de tipos de locais em comparação

---

<sup>1</sup><https://foursquare.com>

a uma pessoa que realiza *check-ins* somente em uma determinada categoria, isso também foi abordado no artigo de Wang et al. (2017) que realizou uma pesquisa profunda sobre mecanismos de incentivo.

São denominados *prefeitos* de estabelecimentos os usuários que mais realizaram *check-ins* nos últimos 30 dias no local em questão, as prefeituras podem ser disputadas tanto entre amigos quanto entre pessoas desconhecidas, pelo fato do local ser acessível para qualquer pessoa, ser prefeito sugere que o usuário é denominado especialista do local, como explicita o artigo de Wang et al. (2017) com o termo *Market Expert*.

Juntos, esses três tipos de recompensa, compõem os mecanismos de incentivo para a utilização do aplicativo. Neste trabalho será estudado o mecanismo de disputa de prefeituras. Especificamente, será analisada a alternância de prefeitos para os locais, podendo ser locais com alta alternância de prefeitos ou onde a disputa não é tão acirrada. Com isso será possível detectar quais as características de locais onde ocorrem disputas acirradas e se isso tem a ver com a categoria ou localização, por exemplo.

## 1.1 Objetivos Gerais

Estudar a alternância de prefeitos no aplicativo *Swarm*, além disso comparar esses resultados entre as cidades que serão estudadas, sendo elas Curitiba (Brasil), São Paulo (Brasil), Chicago (Estados Unidos), Nova Iork (Estados Unidos) e Tokyo (Japão).

### 1.1.1 Objetivos Específicos

- Detectar os possíveis motivos e a frequência para a alternância de prefeitos.
- Evidenciar quais as categorias com maior número de *check-ins* e relacionar com o número de prefeitos das mesmas categorias.
- Estabelecer parâmetros de comparação e diferenciação entre as cidades estudadas, de acordo com a análise de *check-ins*.
- Discutir possíveis implicações para sistemas que desejam implantar o mecanismo de prefeitura.

## 1.2 Organização do Documento

Na sequência deste trabalho estão presentes o Capítulo 2, que descreve as bases teóricas que fundamentam a proposta apresentada e também a obtenção dos dados. No Capítulo 3 está apresentado o Estado da Arte, com as obras investigadas para a pesquisa. A metodologia se encontra no Capítulo 4, apresentando passos essenciais para a realização do estudo. O Capítulo 5



apresenta análises iniciais dos dados. Os recursos de *hardware* e *software* serão abordados no Capítulo 6 juntamente com informações do computador utilizado. Na sequência será apresentado o Capítulo 7 com a descrição do cronograma de atividades e a viabilidade do estudo. O Capítulo 8 terá a finalização da proposta, com foco nos resultados obtidos.

## Capítulo 2

# Referencial Teórico

### 2.1 Redes Sociais Baseadas em Localização

Entre as definições de rede social, uma delas se refere ao local onde pessoas podem socializar, interagir e se ajudar, não necessariamente usando internet, podendo ser um grupo de pessoas que se reúnem com objetivos em comum. Entre os tipos de redes sociais, uma delas é baseada em localização e se chama Redes Sociais Baseadas em Localização (LBSN). O Foursquare é um exemplo, a diferença de LBSN e rede social é que rede social é a definição ampla e engloba, por exemplo, redes sociais de entretenimento <sup>1</sup>, profissionais <sup>2</sup> e compartilhamento de imagens <sup>3</sup>, ou seja, redes sociais cujo o foco não é no aspecto geográfico, como a localização, onde os usuários realizam as atividades.

Dados de LBSNs são um recurso muito utilizado para determinar padrões de comportamento de usuários, pois ao compartilhar dados contendo a sua localização pode ser possível determinar por exemplo a rota feita no dia ou em outro período de tempo como nas figuras 2.1 e 2.2 , e também é possível saber por quais tipos de locais o usuário tem preferência, por exemplo: bares; parques; restaurantes e casas noturnas.

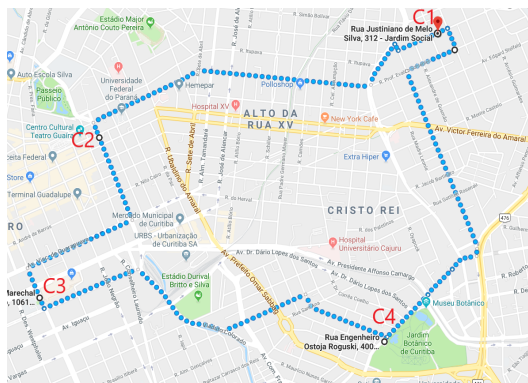


Figura 2.1: Check-ins de um dia na mesma cidade de um determinado usuário

<sup>1</sup>www.facebook.com

<sup>2</sup>www.linkedin.com

<sup>3</sup>www.instagram.com



Figura 2.2: Check-ins em cidades diferentes

No Swarm esse mecanismo é utilizado para encontrar estabelecimentos próximos por meio da localização atual. O Swarm registra todos os *check-ins* realizados pelo usuário na própria conta do usuário, e assim pode ser estabelecido um padrão de locais mais frequentados e quais os tipos desses locais. Com essas informações pode ser possível recomendar locais semelhantes a fim de oferecer ao usuário mais opções.

Explorar esses dados pode habilitar novos serviços. Por exemplo detectar áreas que possuem estabelecimentos com maior número de *check-ins*, um serviço pode sugerir qual a melhor área para se iniciar um novo negócio, ou estudar em quais áreas faltam determinados tipo de negócio. Por exemplo: Restaurante próximo a centros empresariais; farmácia próxima de hospitais e lanchonete próxima à casas noturnas ou locais de muita movimentação.

## 2.2 Mecanismos de Incentivo

Mecanismos de incentivo são elementos que fazem com que usuários continuem a realizar determinada tarefa por gostar ou receber algo em troca, como reconhecimento, moedas virtuais, *likes* e popularidade. No Swarm os tipos de mecanismos de incentivo são: Prefeituras; medalhas e moedas virtuais.

Para as redes sociais adquirirem cada vez mais usuários ativos é de extrema importância ter uma boa motivação para o usuário utilizar a aplicação, esse incentivo pode vir de várias formas, como mencionado acima. Entretanto, isso tudo só é possível, pois os usuários interagem entre si e trocam informações, compartilham conhecimentos e se atualizam com notícias.

Os mecanismos de incentivo que o Swarm utiliza envolve *check-ins* pelos usuários, que ao compartilhar sua localização recebem moedas virtuais em troca, ao realizar *check-ins* em locais diferenciados, em eventos ou marcando amigos se ganha medalhas personalizadas (Medalhas), e quando o usuário é a pessoa que mais realizou *check-ins* em determinado local, ela é reconhecida como perita do local (Prefeitura).

### 2.2.1 Medalhas

Para incentivar pessoas a usar o Swarm, a rede social possui mecanismos de incentivo, um deles é recompensar os usuários com medalhas de acordo com *check-ins*, para ganhar medalhas o

usuário deve realizar *check-ins* em locais novos, categorias novas de locais (restaurantes, parques, aeroportos, mercados e agências bancárias), eventos e/ou marcar determinado número de pessoas em um *check-in*. É possível utilizar as medalhas já adquiridas para diferenciar *check-ins*.

### 2.2.2 Prefeitura

Outro mecanismo de incentivo é a prefeitura, muito cobiçado pelos usuários, que consiste na pessoa que realizou mais *check-ins* em um local específico nos últimos 30 dias, a pessoa denominada prefeito do local é vista como uma autoridade do local e pode até ganhar desconto ao comprar algo no local onde é prefeito pelo aplicativo. Um usuário pode ser coroado como prefeito em vários locais ao mesmo tempo, inclusive da própria casa, basta cadastrar o local no aplicativo.

A alternância de prefeitos se dá no exato momento em que um usuário realiza *check-in*, e na verificação do *Swarm* ele agora possui mais de um *check-in* de diferença que o atual prefeito pois pode ser que o atual prefeito realizará seu *check-in* mais tarde no mesmo dia e com isso ficarão com o mesmo número de *check-ins* para o dado local.

## 2.3 Coleta de Dados Automática da Web

A coleta de dados é uma etapa de extrema importância, e deve ser realizada de forma confiável, a forma mais fácil de obter os dados é através de APIs, que são estruturas onde os dados já são padronizados, formatados e preparados para serem coletados e estudados por pesquisadores. Também se pode utilizar *web crawlers* que são programas computacionais que buscam os dados através de padrões definidos por quem o cria. A diferença entre API e *web crawler* é que com a utilização da API os dados já estão prontos para serem coletados e com o *web crawler* o trabalho é maior, pois quem cria o padrão é quem desenvolve o programa, não tendo nada já formatado nem padronizado.

Para sites que não possuem API, a opção para coletar os dados é utilizar *web crawlers*, que são programas desenvolvidos e configurados previamente para coletar dados desejados pelo usuário, a coleta é feita através de requisições ao site, após receber a confirmação da requisição o *web crawler* obtém todas as informações da página e a partir disso se inicia o tratamento dos dados.

Para realizar requisições tanto via API quanto utilizando o *web crawler*, é necessário respeitar a quantidade das requisições, definidas ou um limite razoável que não irá atrapalhar o serviço. Isso ajuda evitar que o *web crawler* seja bloqueado por exemplo. Nesta proposta as requisições foram feitas via API disponibilizada pelo Foursquare juntamente com um *web crawler* para tratar os dados, que foi desenvolvido de forma eficiente e contorna alguns tipos de erros como queda de conexão com a rede e dados inconsistentes.

## 2.4 Banco de Dados NoSQL

Entre os tipos de banco de dados existem os dois mais comuns são: Dados estruturados ou dados não estruturados, o modelo estruturado como o próprio nome já diz, possui uma estrutura, geralmente de linhas, colunas, possuem chave primária para buscas mais rápidas e chaves estrangeiras, que são chaves primárias em outras tabelas para usar como referência, para isso, todos os dados obtidos devem ter um mesmo padrão, não podendo ter campos a mais, e se for a menos o campo ficará vazio e ocupando memória em vão. Para realizar consultas em banco de dados estruturados é utilizada uma linguagem de consulta chamada SQL, a partir disso é possível unir conteúdo de várias tabelas e trabalhar com esses dados.

No caso de banco de dados não estruturado (NoSQL), se não utiliza SQL como linguagem de consulta, os dados são na maioria das vezes em formato JSON (exemplo na figura 2.3) sendo que cada JSON é um item. Os dados de cada item podem variar, sendo que uns podem possuir mais dados e outros menos, por exemplo o número de dicas para cada local do Foursquare. O banco de dados utilizado no estudo é o MongoDB, que é do tipo não relacional, com a estrutura flexível o grande ganho do banco de dados não relacional é em velocidade, armazenamento e escalabilidade, pois não vai ocupar memória desnecessária. As consultas são feitas através de chave: valor, onde é buscada determinada chave é o retorno é o valor correspondente. Uma consulta no MongoDB para capturar os dados do local "Universidade Tecnológica Federal do Paraná" na cidade de Curitiba ficaria:

```
db.Curitiba.find({name:"UTFPR"})
```

onde Curitiba é o nome da coleção, onde está sendo registrado os dados de todos os estabelecimentos da cidade e UTFPR é um local que se encontra em Curitiba.

Uma das vantagens do banco de dados relacional é a padronização e normalização, pois suportam as propriedades ACID (Atomicidade, Consistência, Isolamento e Durabilidade).

```
{
  "name" : William Souza,
  "age" : "23",
  "e-mail" :william@alunos.utfpr.edu.br,
  "info" : {
    "hometown" : "Curitiba",
    "job" : "Student"
  }
}
```

Figura 2.3: Exemplo de arquivo formato JSON

## Capítulo 3

### Estado da Arte

Neste capítulo serão abordados trabalhos relevantes para a elaboração deste artigo, trabalhos que possuem pontos em comum e que ajudaram na elaboração do estudo.

Dentre os temas de análise de dados de redes sociais, *check-ins* e mecanismos de incentivo por exemplo, inúmeros artigos foram encontrados, porém ajustando os termos de busca para o propósito do trabalho, foi levado em consideração outros temas, por exemplo, prefeitura do aplicativo Swarm, com isso foram encontrados vários artigos que tinham, de algum modo, relação com o tema.

Entre os trabalhos similares um deles é o desenvolvido por Neves et al. (2016), que teve o objetivo de analisar a relação entre os *check-ins* do Foursquare e series temporais de popularidade de locais fornecidos pelo Google <sup>1</sup>. Para obter esses dados a estratégia é digitar o nome de um estabelecimento na busca do Google e se a informação existir ela será exibida. Para conseguir comparar as duas ferramentas foi desenvolvido um *web crawler* para obter *check-ins* do Foursquare e depois comparar as duas fontes. Os autores descobriram que a similaridade entre a série popular do Google e as series temporais de popularidade criadas com *check-ins*, geraram gráficos bem similares tanto para a cidade de Curitiba quanto para Chicago (Neves et al., 2016).

No estudo de Wang et al. (2017) que teve como objetivo analisar a efetividade de contribuidores "*word-of-mouth*" (boca-a-boca) juntamente com elementos de gamificação, como prefeituras e medalhas, os autores utilizaram a os aspectos presentes em jogos para incrementar o efeito "boca-a-boca". Isso envolve os dois tipos de influências citados no artigo, um deles é o *market maven*, que é a pessoa que conhece vários locais diferentes, possui uma variedade de experiências e a segunda é o *market expert* que é conhecedor de determinados locais ou regiões específicas. *Market maven* pode ser comparado com as medalhas adquiridas pelo usuário ao realizar *check-ins* em diferentes locais e *market expert* se compara ao usuário que é prefeito de locais de categorias distintas.

O método de avaliação contou com a ajuda de mais de duzentos voluntários divididos em 2 grupos. Para cada voluntário era solicitado a análise de páginas específicas de estabelecimentos,

---

<sup>1</sup> [www.google.com](http://www.google.com)

comentários e prefeituras desses locais e em seguida eram questionados se o comentário ou algo assim podia ter influenciado outras pessoas. Como resultado desse estudo foi possível perceber que os usuários preferem seguir dicas de quem possui medalhas ao invés de seguir a dica do prefeito do local, pois quem é um usuário específico não tem muitos meios de comparação do que com pessoas que já visitaram vários outros locais.

Outros termos diferentes além de prefeitura e medalhas são abordados no artigo de Cramer et al. (2011), por exemplo, mecanismos de incentivo que também será estudado adiante. O foco deste estudo foi analisar *check-ins* de usuários nas redes sociais e os motivos para os mesmos. Foi usado a rede social Facebook e também o Twitter <sup>2</sup> já que ambas possuíam integração com o Foursquare. Foram escolhidas pessoas aleatórias que usavam as redes sociais e essas pessoas foram submetidas a uma entrevista com perguntas e análises.

Como resultado foi detectado que pessoas podem visualizar *check-ins* de outras, e através disso mandam solicitação de amizade mesmo sem conhecer quem realizou o *check-in*. O estudo comprovou que a maioria das pessoas não aceita convite de estranhos. Outro ponto é a motivação para realizar *check-in*, onde determinados locais tendem a ter mais *check-ins* do que outros, dependendo se o usuário quer ou não compartilhar seu local, pessoas podem realizar *check-ins* apenas para ajudar outras, como se fosse o *like* do Facebook.

A pesquisa de Pontes et al. (2012) analisa as opções disponíveis a se fazer na rede social Foursquare como prefeituras, dicas de locais e invasão de privacidade, por exemplo. Com uma busca abrangente foi possível notar milhões de usuários distribuídos no mundo inteiro, a dispersão de prefeituras e dicas de locais por cidades também.

Os autores identificaram que usuários com prefeitura moram próximos do local onde são prefeitos. Além disso, foi mostrado que através dos *check-ins* pode-se descobrir a cidade onde o usuário reside. Foram coletadas informações referentes a cidade de cada usuário e capturado todas as atividades do mesmo. Os grupos de estudo eram três: Classe 0: usuários com atividades singulares como prefeituras ou dicas; Classe 1: usuário com múltiplas atividades e com uma localização predominante no meio das atividades; Classe 2: são usuários com múltiplas atividades mas sem nenhuma localização predominante. Os resultados foram bastante expressivos para as classes 0 e 1, pois nessas classes estão englobados os prefeitos da maioria dos locais e pessoas que realizam *check-ins* com mais frequência.

No artigo de Lindqvist et al. (2011), os autores tinham como objetivo descobrir o motivo de pessoas compartilharem localizações e porque as definiam como públicas ou privadas. O método aplicado foi um questionário de 55 perguntas, que foram envolvidos 219 usuários, as perguntas abordavam a rede social Foursquare e suas funcionalidades. Como resultado desse estudo foi possível detectar o motivo pelo qual pessoas não fazem *check-ins* em determinados locais, por exemplo pessoas relataram não realizar *check-in* em redes de *fast food* pois para elas não é algo bom outros saberem que elas passaram por lá. Os locais com mais *check-ins* são bares e restaurantes e os locais com menor número de *check-ins* são residências privadas, consultórios e

---

<sup>2</sup>[www.twitter.com](http://www.twitter.com)

locais de trabalho. Dentre as razões para pessoas usarem o Foursquare estão: Elementos de jogo (gamificação); acompanhar aonde os amigos estão ou para ver locais pelos quais já passaram.

Em Santos et al. (2017) os autores se propuseram a estudar os mecanismos de incentivo do Foursquare, sendo eles prefeitura e medalhas. Para se tornar prefeito de um determinado local o usuário tem que ser a pessoa que mais fez *check-ins* no local durante os últimos 30 dias, as medalhas são adquiridas de acordo com o uso do aplicativo em geral, realizando *check-ins* em locais de categorias diferentes, marcando determinado número de pessoas e realizando *check-ins* em eventos ou datas comemorativas. O estudo foi feito baseado em novos usuários que foram monitorados durante 13 semanas, deles foram coletados: número de *check-ins*; locais de prefeitura e medalhas para mensurar a interação deles com o aplicativo.

Para analisar os dados foi necessário dividir os usuários em 3 grupos sendo o Grupo 1 composto pelos usuários que realizaram até 250 *check-ins*, Grupo 2 entre 250 e 500 *check-ins* e Grupo 3 de 500 *check-ins* em diante. Através dessa divisão percebeu-se que usuários do Grupo 1 poderiam não estar muito motivados e consequentemente realizando poucos *check-ins* possuíam poucas medalhas e prefeituras, já o Grupo 2 apresentou um grupo mais ativo e que até certo ponto os usuários estavam motivados, possuíam mais medalhas e prefeituras em relação ao Grupo 1, por fim, o Grupo 3 foi eleito como grupo mais ativo e foi detectado maior número de prefeituras e medalhas em disparado, englobando assim os usuários super motivados.

Outro artigo que diz respeito à redes sociais baseadas em localização e análise de padrões de atividades de usuários é descrito em Lin et al. (2017), onde o aplicativo *Swarm* do Foursquare é utilizado e os *check-ins* por categoria de estabelecimentos são analisados em tempos diferentes do dia para saber quais categorias de locais são mais frequentados em quais horários do dia. Por exemplo, casas noturnas e baladas têm grande número de *check-ins* durante o período da noite e locais como restaurantes tem picos de *check-ins* durante o dia e noite. Fora as categorias de restaurantes e casas noturnas, outras categorias foram utilizadas totalizando 10. Intervalos de tempo para a coleta de dados foram 6, divididos a cada 4 horas com início das 6 horas até 10 horas como período da manhã.

Foram coletados dados de mais de 19 mil usuários do *Swarm* das cidades de Nova Iork (EUA), São Francisco (EUA) e Hong Kong (China), todos com consentimento do estudo, sabendo que seus dados iriam ser usados pelos solicitante.

Foram utilizadas fórmulas separadas para detectar a diferença entre dias da semana e finais de semana, com o objetivo de distinguir categorias de locais. Por exemplo, *check-ins* em baladas ou casas noturnas são muito maiores no período da noite e no final de semana do que em dias de semana, por outro lado, *check-ins* em locais como trabalho e colégio/faculdade são muito maiores em dias de semana em relação à finais de semana. Na categoria de alimentação a diferença é pouca entre dias de semana e finais de semana, mas em comparação com outras categorias é a que mais possui *check-ins* do total.

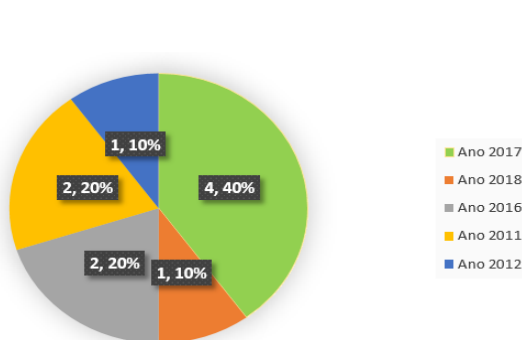
A diferença entre os padrões de *check-ins* entre cidades mostrou algumas diferenças como na cidade de Hong Kong, os valores de compras no período da tarde são muito maiores



do que nas outras duas cidades e o motivo pode ter sido causado por um estabelecimento em específico chamado Shopping Heaven que atrai muitos turistas. Viagens pela manhã também foram mais registradas em Hong Kong em comparação com Nova Iork e São Francisco. Grandes diferenças não foram encontradas entre Nova Iork e São Francisco pois se encontram no mesmo país.

Alguns autores encaram dados provenientes de LBSNs como uma fonte de sensoriamento chamada de *Participatory Sensor Network* (PSN), que permite o estudo de sociedades urbanas quase que em tempo real e em larga escala. Silva et al. (2016) abordou conceitos de PSNs e realizou uma análise sobre comportamento social urbano usando o Instagram, Foursquare e Waze <sup>3</sup>. Os dados foram coletados das redes sociais através de *web crawlers* construídos para esse propósito. A dinâmica das cidades também foi um tema bastante estudado, pois os dados considerados podem permitir detectar diferentes regiões da cidade com características diferentes ou semelhantes baseado na movimentação das pessoas. Como conclusão do estudo, pode-se citar que LBSNs podem promover oportunidades para acessar dados em escala global e mecanismos de incentivo são fundamentais para tornar essa fonte de dados sustentável (Silva et al., 2016).

Pode-se notar que temos uma diversidade de temas, áreas abordadas e publicação de artigos de diversos anos. Os trabalhos utilizados foram encontrados através de busca em bases públicas de artigos com as palavras chave: Foursquare; *incentive mechainism*; *rewards*; *Swarm*; *Mayorship* e *Check-in*. Após a seleção dos artigos percebeu-se que a maioria deles foi publicado em 2017, mas por outro lado também foram encontrados artigos com assuntos em comum publicados em 2012. Mais detalhes sobre o ano de publicação dos artigos na figura 3.1(a).



(a) Ano de publicação dos artigos



(b) Palavras-chave dos artigos

Com as palavras mais citadas dentre todos os artigos analisados foi possível perceber uma predominância das palavras *Mayorships* (que significa prefeitura e é um dos temas essenciais para o trabalho atual) e *Badges* (medalhas), ambos são mecanismos de incentivo para os usuários, LBSN, *Check-ins* e *Gamification*, entretanto outras palavras importantes foram detectadas conforme figura 3.1(b).

<sup>3</sup>www.waze.com

# Capítulo 4

## Metodologia

Os passos para a realização do estudo são: Para iniciar é apresentado os dados do Swarm a serem coletados, que estão com mais detalhes no capítulo 4.1; logo após são descritos alguns detalhes de como foram elaborados os coletores de dados no capítulo 4.2; a infraestrutura dedicada que armazena os dados e é responsável pela execução diária do coletor será descrita no capítulo 4.3; o agendador de tarefas utilizado para executar a rotina diária de coleta de dados está presente no capítulo 4.3.1; para armazenamento, o capítulo 4.1 irá conter qual a estrutura escolhida para armazenamento dos dados e por último a descrição da análise dos dados, no capítulo 4.4.

A figura 4.1 deixa mais claro os passos para a realização do trabalho.

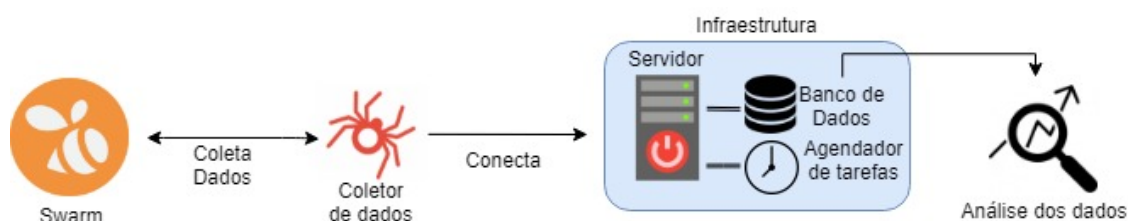


Figura 4.1: Etapas para obtenção dos dados

### 4.1 Dados a Serem Coletados

Dados são de suma importância para o andamento do trabalho e se não estiverem corretos podem levar a conclusões equivocadas, para isso foram realizados testes antes de colocar os *web crawlers* em produção. A estrutura utilizada no trabalho em maioria foi o formato JSON (do inglês *JavaScript Object Notation*). O *web crawler* faz a requisição para um página feita em HTML (*HyperText Markup Language*) mas obtém os dados no formato JSON através de uma biblioteca disponível para a linguagem de programação utilizada.

Os principais dados de interesse são: Nome do local; endereço; avaliações; comentários; *check-ins*; fotos; categoria do local; preço e, principalmente, o nome do prefeito se houver. Além disso é importante acrescentar mais um campo com a data atual para controle e também para

ajudar nas análises. Na sequência os dados são inseridos no banco de dados automaticamente pelo próprio código.

## 4.2 Coletor de Dados

Os *web crawlers* foram desenvolvidos em linguagem Python. Para se obter os dados descritos no item anterior, foi necessário criar uma conta para desenvolvedores junto ao Foursquare e usar um *login* e senha fornecidos para realizar todas as requisições, servindo como um meio de segurança para controlarem o número de requisições diárias para cada usuário.

Após a autenticação, para cada requisição todo o conteúdo disponibilizado na página do local é coletado, contendo todos os detalhes do local. Com todos os locais estipulados já percorridos e inseridos no banco de dados, é disparado um e-mail com a confirmação da execução com o *web crawler*. No capítulo 4.2.1 são apresentados mais detalhes sobre como foi feita a obtenção dos dados.

### 4.2.1 Obtenção dos Dados

Para coletar as alternâncias de prefeituras foi desenvolvido um *web crawler*, para cada cidade a ser analisada. Neste trabalho os dados são coletados diariamente em horários espaçados para não sobrecarregar a API oferecida pelo Foursquare. Através dessa API e com uma extensão é possível coletar dados do *Swarm* (que contém a informação do prefeito do local). Os dados que estão sendo coletados contém toda a página do local. As cidades monitoradas são: Curitiba (Brasil); São Paulo (Brasil); Chicago (Estados Unidos); Nova Iork (Estados Unidos) e Tokyo (Japão).

Como os *web crawlers* e o banco de dados estão armazenados em um servidor, para executar os programas foi utilizado um agendador de tarefas que sempre executa os *web crawlers* no mesmo horário todos os dias e dentro de cada *web crawler* foi configurado para assim que terminar a execução disparar um e-mail de confirmação, sendo assim possível o monitoramento das coletas de dados.

Para cada cidade foram pesquisadas as mesmas categorias para o resultado ser eventualmente analisado de acordo com a categoria, são cerca de 1.000 locais para cada cidade distribuídos em aproximadamente 30 categorias, ou seja, juntando as 5 cidades são aproximadamente 5.000 locais distintos que são coletados diariamente. Todos esses dados são armazenados em um banco de dados não relacional que possui armazenamento de dados mais compacto em relação ao modelo relacional, mas o grande ganho está na escalabilidade do modelo não relacional.

## 4.3 Infraestrutura

Para ser possível a realização do trabalho, além de ter uma segurança maior em relação aos dados, o *web crawler* e o banco de dados foram hospedados em um servidor dedicado com sistema

operacional Linux, que sempre está em funcionamento, com acesso à Internet e possui grande capacidade de armazenamento, para evitar ao máximo de ocorrer problemas com as coletas de dados. Esse servidor está disponível 24 horas por dia nos 7 dias da semana (24X7). No servidor todos os dados conforme já informado estão sendo coletados diariamente, e estão sendo armazenados em um banco de dados, para a execução dos coletores de dados, são utilizados agendadores de tarefas, a qual permite programar tarefas para executar de acordo com determinada programação, podendo ser executar uma mesma tarefa a cada 5 minutos ou uma vez por semana em um horário específico.

A seguir serão apresentadas as seções 4.3.1 que se refere a rotina de execução do *web crawler* e na seção 4.3.2 é apresentado o MongoDB e mais detalhes sobre onde os dados estão sendo armazenados.

### 4.3.1 Agendador de Tarefas

Para realizar o agendamento da execução de cada um dos *web crawlers* foi utilizada a *cron* do servidor, que é um serviço de agendamento de tarefas, a *cron* possui horários específicos para cada operação, antes de iniciar a execução do *web crawler*, é aberta a conexão com o banco de dados, e após os cinco *web crawlers* terem terminado a execução, a conexão com o banco de dados é fechada. Um agendamento do web crawler para a cidade de Curitiba, pode ser analisado a seguir:

```
00 04 * * * python /home/sda3/william/curitiba.py
```

onde 00 04 se refere ao horário de execução, sendo minutos e horas correspondentemente, os asteriscos são para outro tipo de agendamento por dias da semana ou por intervalos de horário específicos, na sequência, Python é o comando que será executado e por último o local do arquivo que será executado.

### 4.3.2 Armazenamento de Dados

O modelo de banco de dados utilizado foi o não relacional, ou não estruturado, de forma que os dados coletados não precisam estar sempre no mesmo formato e no mesmo padrão, o SGBD escolhido foi o MongoDB que trabalha muito bem com o formato JSON e possui integração de forma simples com a linguagem de programação utilizada. A organização no banco de dados foi feito por cidade, cada uma das cinco cidades possui uma base diferente de dados para facilitar na análise e organização e também para dividir a sobrecarga de dados.

## 4.4 Análise dos Dados

As fórmulas utilizadas para embasar os gráficos avaliam a alternância de prefeitos em dois casos, uma das fórmulas avalia apenas o número de alterações de prefeitos, sem avaliar quem são

(fórmula 4.1), e a segunda avalia o número de reaparições de prefeitos (fórmula 4.2), para casos em que a disputa é entre dois usuários ou mais que se revezam no topo.

As variáveis analisadas são denominadas P, R e N:

- P = Número de mudança de prefeitura;
- R = Número de reaparições;
- N = Número de *check-ins*.

Será sempre analisado uma janela deslizante de 30 dias, pois é assim que o aplicativo *Swarm* avalia os *check-ins* realizados pelos usuários em busca da prefeitura, ou seja, a variável N será referente ao total de *check-ins* desses 30 dias, o que pode dar maior pontuação para locais que estejam com a média de *check-ins* entre 25 a 30 do que locais que possuem média menor.

A variável P analisa a disputa do usuário que se torna prefeito, sem levar em conta quem, apenas o número bruto de trocas, sendo o valor mínimo entre zero, no caso de não haver troca de prefeito no local, e no máximo 29, (pois a janela analisada é de 30 dias). Outro tipo de análise é feita ao trocar a variável P pela variável R, pois R verifica o número de reaparições de prefeitos, ou seja, para ter alguma reaparição terá que ter no mínimo duas trocas de prefeitura, pois um usuário necessariamente tem que desaparecer e reaparecer em seguida para contabilizar como uma reaparição.

As duas fórmulas apesar de serem similares, podem ser usadas para complementar os resultados, pois ambas vão levar em consideração o número de *check-ins* do local e vai variar o tipo de troca de pessoa que houve no local.

- Fórmula para analisar a troca de prefeitos.

$$P * \log(N + 1) \quad (4.1)$$

- Fórmula para detectar o número de reaparições de usuários.

$$R * \log(N + 1) \quad (4.2)$$

## Capítulo 5

### Análises

Os dados estão sendo coletados para as cinco cidades, conforme descrito no Capítulo 1, desde o dia 7 de maio de 2018, como exemplo para os gráficos, foi utilizada a cidade de Curitiba, e já se pode notar certa disputa de prefeituras como apresentam as figuras 5.1 e 5.2.

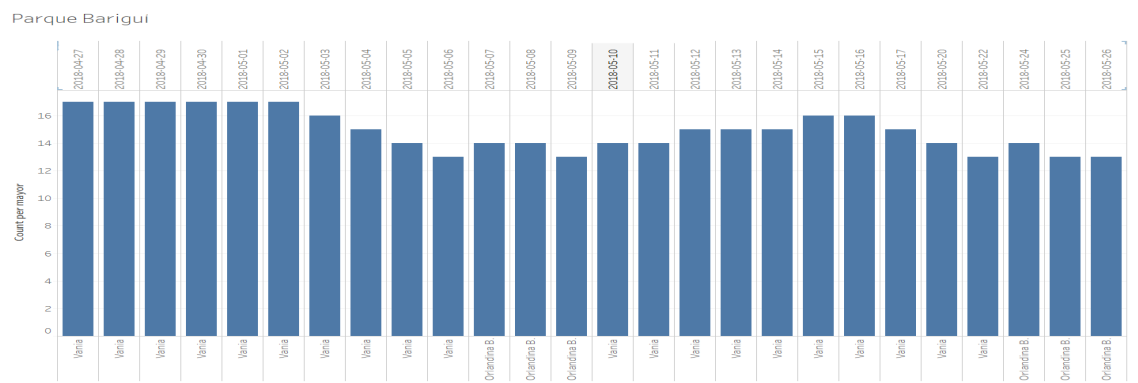


Figura 5.1: Disputa no Parque Barigui

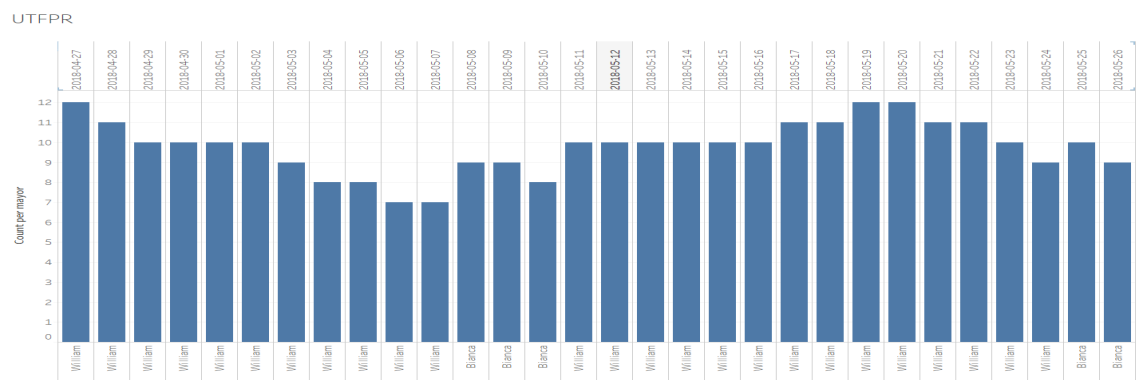


Figura 5.2: Disputa na Universidade Tecnológica Federal do Paraná

Essas disputas podem ser devidas à esquecimento por parte do atual prefeito, pois percebe-se que a troca é logo após uma queda no número de *check-ins*.

Apenas olhando para as duas figuras acima pode-se perceber que a figura 5.1 possui mais *check-ins* do que a figura 5.2, mas ambas possuem o mesmo número de troca de prefeitos

(variável P) e de reaparição de prefeitos (variável R), com isso, ao aplicar as fórmulas 4.1 e 4.2 temos os resultados:

- Parque Barigui (figura 5.1):

Número de alterações de prefeitos: 7,76

Número de reaparições de prefeitos: 5,17

- Universidade Tecnológica Federal do Paraná (figura 5.2):

Número de mudança de prefeitos: 7,4

Número de reaparições de prefeitos: 4,93

Com os resultados acima podemos perceber que a prefeitura no Parque Barigui é mais concorrida em comparação à Universidade Tecnológica Federal do Paraná, apesar de terem o mesmo valor de P(3) e R(2), mas o que acabou diferenciando foi o total de *check-ins* entre os dois locais.

Também é possível notar a preferência de *check-ins* de acordo com as categorias na figura 5.3, onde o número de *check-ins* realizados na categoria de parques é muito maior do que demais.

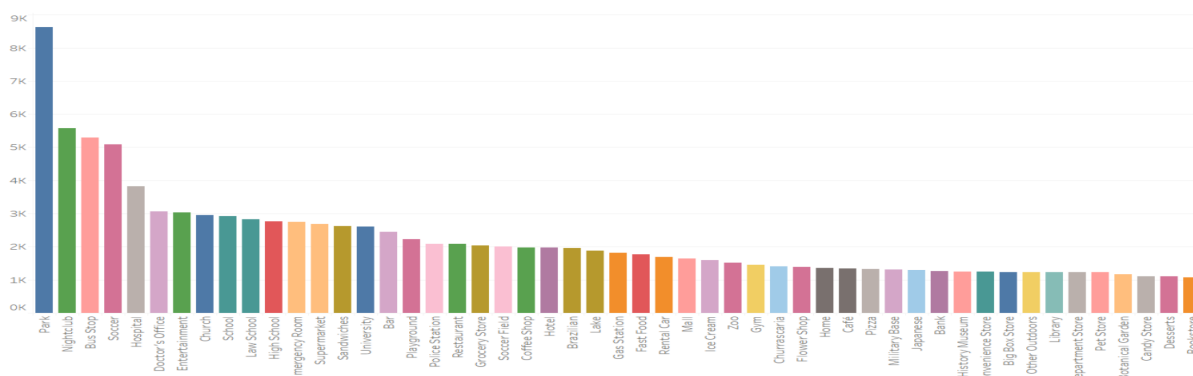


Figura 5.3: *Check-ins* por categoria de local

## Capítulo 6

# Recursos de Hardware e Software

### 6.1 Recursos de Hardware

Os requisitos para rodar as aplicações que estão sendo utilizadas no trabalho não são muito exigentes, como *hardware* o mais importante é armazenamento e segurança para conseguir guardar todos os dados corretamente para análises posteriores e ter certeza que o programa será executado sem quedas de energia e Internet. O tratamento dado para queda de energia ou Internet não envolve nenhum equipamento adicional, e sim temporizadores no código, que fica em *loop* até conseguir a requisição ou depois de algumas tentativas é passado para a próxima requisição, para não causar maiores problemas com temporizadores ou problemas com o travamento de uma requisição em específico. O computador que está sendo utilizado para este trabalho possui um processador da marca Intel I5 com armazenamento interno de 128GB SSD, 8GB de memória RAM. As configurações mínimas de hardware para simular o ambiente desta proposta são:

- Processador marca Intel I3 ou similar;
- 128 GB de armazenamento;
- 4 GB de memória RAM.

### 6.2 Recursos de Software

Na parte de software o computador possui sistema operacional Windows 10 64 bits e os programas utilizados são: Banco de dados MongoDB; Tableau Desktop para criação de gráficos mais elaborados e Python como linguagem de programação, além disso foram utilizadas bibliotecas como *requests*, para fazer as requisições das páginas, *json* para realizar o tratamento dos dados da requisição e *matplotlib* para elaborar os gráficos via código.



## Capítulo 7

### Viabilidade e Cronograma Preliminar

Para a realização do projeto com êxito, não será necessário nenhum investimento de *hardware* adicional, visto que o pesquisador já se tenha um computador conforme configuração mínima, como descrito no Capítulo 6. Para executar os programas utilizados não foi necessário um tipo específico de sistema operacional, pois todos os programas utilizados possuem disponibilidade e adaptabilidade para ambientes Windows, Linux e Mac.

As situações que poderiam comprometer a coleta dos dados estão relacionados à robustez do *web crawler* e a disponibilidade do servidor. O *web crawler* foi desenvolvido de forma segura visando contornar eventuais quedas de conexão com a rede de Internet, bem como realizar o envio de mensagens quando a execução estiver finalizada, no servidor o serviço deve estar ligado 24 horas por dia pois os 5 *web crawlers* são executados diariamente em horários distintos, iniciando 6 horas da manhã e finalizando 10 horas da noite.

O número de requisições poderia interferir na coleta dos dados, pois são quase 5.000 requisições diárias, o problema é que o Foursquare poderia bloquear as requisições ao detectar que partem sempre de um mesmo computador, contudo o *web crawler* possui pausas aleatórias de tempo para cada requisição, a fim de não demonstrar que é algo programado e contínuo, e sim pessoas comuns visitando páginas.

Para dar mais embasamento ao trabalho foi utilizada uma fórmula, a qual é possível avaliar se os locais possuem alta ou baixa troca de proprietários, analisando algumas variáveis em comum para todos os locais.

Para controlar a viabilidade foi desenvolvido um cronograma de atividades, para ajudar no controle das tarefas as decorrer do trabalho, como mostra a figura 7.1.

Atividade	Abril	Maio	Junho	Julho	Agosto	Setembro	Outubro	Novembro
Criação dos crawlers	X							
Seleção das cidades e categorias	X							
Criação do ambiente	X	X						
Verificar andamento do crawler	X	X	X	X	X	X	X	X
Escrita do artigo	X	X	X	X	X	X	X	
Análises iniciais		X	X					
Correção de textos		X	X		X	X	X	X
Análise com gráficos		X	X			X	X	
Defesa inicial			X					
Elaboração da fórmula			X	X	X			
Análise com fórmulas					X	X	X	
Análise com fórmulas e gráficos						X	X	
Entrega do artigo final								X
Defesa final								X

Figura 7.1: Cronograma de Atividades

## Capítulo 8

### Conclusão

Com relação aos dados iniciais percebeu-se uma disputa pela prefeitura em determinados locais, isso demonstra que pessoas diferentes estão interessadas em um mesmo local, a partir daí se caracteriza a disputa, nas figuras 5.1 e 5.2 é possível perceber a alternância de prefeitos devido a queda no número de *check-ins* do atual prefeito, o que pode ter sido esquecimento. Na figura 5.3 fica claro que a categoria Parques possui número de *check-ins* maior até o momento, e isso evidencia o motivo pelo qual a média de *check-ins* da figura 5.1 é maior que 5.2.

O estudo da alternância prefeituras é relevante, pois é a causa de milhares de *check-ins* diários, isso evidencia a importância desta proposta juntamente com mecanismos de incentivo e recompensas, pois os usuários se sentem mais à vontade ao ser eleito prefeito de determinado local.

O objetivo deste trabalho foi evidenciar a alternância de prefeituras do aplicativo *Swarm* e detectar possíveis motivos para a alternância de acordo com número de *check-ins* e categoria do local. Isso é considerado um problema pois não se tinha certeza de que categorias que possuem maior número de *check-ins* teriam mais prefeitos e disputas mais acirradas, e o meio para solucionar isto foi coletando dados e os estudando através de gráficos e fórmulas matemáticas.

## Bibliografia

- Cramer, H., Rost, M., and Holmquist, L. E. (2011). Performing a check-in: Emerging practices, norms and 'conflicts' in location-sharing using foursquare. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI '11*, pages 57–66, New York, NY, USA. ACM.
- Lin, S., Xie, R., Xie, Q., Zhao, H., and Chen, Y. (2017). Understanding user activity patterns of the swarm app: A data-driven study. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, UbiComp '17*, pages 125–128, New York, NY, USA. ACM.
- Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J., and Zimmerman, J. (2011). I'm the mayor of my house: Examining why people use foursquare - a social-driven location sharing application. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, pages 2409–2418, New York, NY, USA. ACM.
- Neves, Y. C. B., Sindeaux, M. P., Souza, W., Kozevitch, N. P., Loureiro, A. A., and Silva, T. H. (2016). Study of google popularity times series for commercial establishments of curitiba and chicago. In *Proceedings of the 22Nd Brazilian Symposium on Multimedia and the Web, Webmedia '16*, pages 303–310, New York, NY, USA. ACM.
- Pontes, T., Vasconcelos, M., Almeida, J., Kumaraguru, P., and Almeida, V. (2012). We know where you live: Privacy characterization of foursquare behavior. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing, UbiComp '12*, pages 898–905, New York, NY, USA. ACM.
- Santos, F. A., Silva, T. H., Braun, T., Loureiro, A. A. F., and Villas, L. A. (2017). Towards a sustainable people-centric sensing. In *2017 IEEE International Conference on Communications (ICC)*, pages 1–6.
- Silva, T. H., Celes, C., and Vincius Mota, J., Cunha, F., Ferreira, A., Ribeiro, A., de Melo, P., Almeida, J., and Loureiro, A. (2016). Users in the urban sensing process: challenges and research opportunities. *Pervasive Computing: Next Generation Platforms for Intelligent Data Collection, Academic Press*, pages 45–95.

Wang, L., Gunasti, K., Gopal, R., Shankar, R., and Pancras, J. (2017). The impact of gamification on word-of-mouth effectiveness: Evidence from foursquare. In *Proceedings of the 50th Hawaii International Conference on System Sciences*.