



Sensoriamento Participativo na Composição de Pontos de Interesse

Marina Iolanda Oliveira¹

Felipe Domingos da Cunha²

Resumo: O crescente uso de aparelhos celulares smartphones, tem contribuído para um aumento na adesão às Redes sociais baseadas em localização. Este tipo de rede possibilita o sensoriamento participativo, o que torna possível entender o comportamento dos eventos nos grandes centros urbanos. Para realização do sensoriamento é necessária a participação ativa dos usuários compartilhando dados, ou seja, os dados são enviados em diferentes momentos e armazenados durante o período desejado, possibilitando assim a análise e estudo das informações obtidas. A proposta deste trabalho é aplicar redes de sensoriamento participativo para usufruir das informações geradas voluntariamente por usuários de mídias sociais. Coletou-se o total de 3.379.842 *tweets* realizados em 10 dos principais pontos turísticos do mundo, em um período de 20 dias. Após o tratamento dos dados coletados e a partir da aplicação de testes estatísticos foi possível caracterizar o padrão de visitação nestes lugares, apresentando os horários com maior e menor fluxo de movimentação baseados na média de *tweets* realizados em cada hora do dia.

Palavras-chave: Coleta de Dados, Sensoriamento Participativo, *Tweets*, Análise de Informações.

¹ Ciência da Computação, Universidade de Itaúna, marinaioliveira9@gmail.com.

² Doutor em Ciência da Computação, Universidade de Itaúna, felipe@uit.br.

1. Introdução

Com o crescimento abrangente da tecnologia, a computação ubíqua está cada vez mais presente e pode ser definida como a tecnologia tão necessária no dia-a-dia do ser humano que se torna indistinguível, sendo cada vez mais importante e necessária para as realizações de tarefas diárias (Weisner, M.,1996). Com a disponibilidade ubíqua de tecnologia computacional, como smartphones, tablets e outros dispositivos portáteis, bem como a popularização de redes sociais, é possível estar conectado e compartilhar dados constantemente.

Diariamente usuários de mídias sociais geram e compartilham um grande volume de dados (Alecrim, Emerson, 2015; França et al., 2014). Por se tratar de uma grande quantidade de dados, utiliza-se rede de sensoriamento participativo, o qual possibilita a coleta dos dados a fim de obter o contexto de entidades (Silva et al, 2013, Silva et al., 2014). Em Abwod et al., 1999, os autores definem contexto como qualquer informação que caracteriza uma entidade. Uma entidade é uma pessoa, lugar ou objeto que é considerada relevante para a interação entre usuário e aplicação.

O sensoriamento participativo é definido como um processo distribuído de coleta de dados. Para realizar o sensoriamento é necessário a participação ativa das pessoas,



compartilhando voluntariamente informações de contexto (Burke et al., 2006). As informações são enviadas em diferentes momentos, os dados são armazenados e após um determinado período podem ser analisados.

A popularização de dispositivos móveis tem aumentado o uso de redes sociais baseadas em localização, como *Foursquare*, *Instagram*, *Facebook*, *Twitter* e *Waze*. Redes sociais baseadas em localização são exemplos de Redes de Sensores Participativos, onde nós sensores são formados por usuários que enviam dados sobre seu contexto utilizando seus dispositivos (Silva et al., 2014). Os usuários enviam informações similar a sensores em uma rede de sensores tradicionais (Silva et al., 2014).

A FIGURA 1 ilustra a rede de sensoriamento participativo, onde cada usuário com seus dispositivos móveis envia informações e suas localizações. As informações são enviadas em diferentes momentos, os dados são armazenados e após um determinado período e podem ser analisados. O usuário não necessariamente participa em todos os momentos, mas gera dados voluntariamente de acordo com seu cotidiano. Depois de um dado tempo, é possível analisar estes dados coletados.

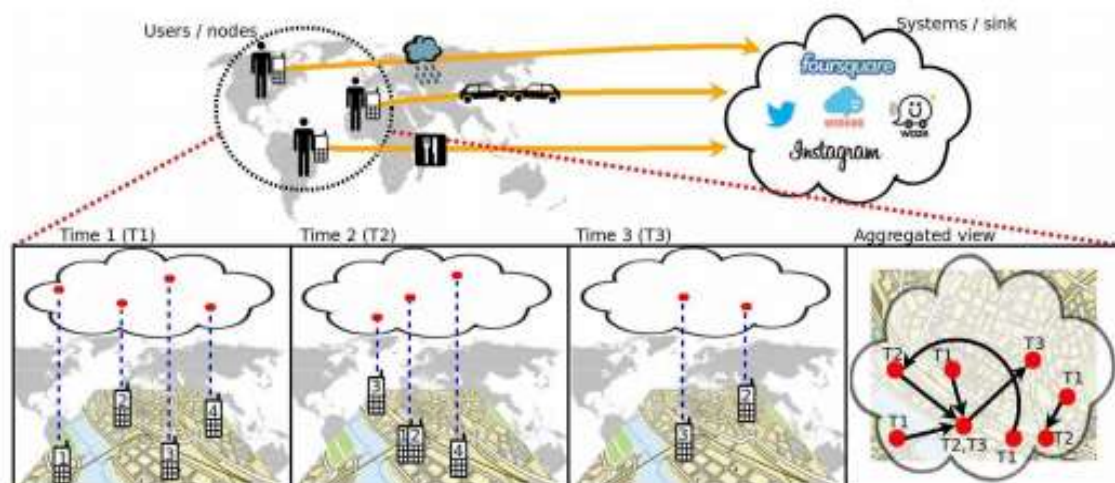


FIGURA 1 - Redes de Sensoriamento Participativo (Silva et al., 2014).

Neste contexto, a proposta deste trabalho é aplicar redes de sensoriamento participativo para usufruir das informações geradas voluntariamente por usuários de mídias sociais, quantificando *check-ins* realizados em determinados ambientes. A mineração dos dados coletados, proporciona resultados aproximados do fluxo de movimentação do ambiente no momento e horários com maiores movimentações. Desta forma, os dados foram coletados em diferentes pontos de interesse no mundo, durante 20 dias e a partir da aplicação de testes estatísticos estudou-se o padrão de visitação em alguns dos principais pontos turísticos do mundo. Sendo eles: *Big Ben*, *Burj Khalifa*, Cataratas do Iguazu, *Central Park*, *CN Tower*, Cristo Redentor, *Musée Du Louvre*, *Oxford Street*, *Statue Of Liberty* e *One World Trade Center*.

Pessoas geram dados a todo instante. Por meio de tais informações é possível analisar o comportamento humano a fim de entendê-lo e criar soluções que os atendam. O trabalho permite que de qualquer lugar obtenha-se informações mundiais por meio das redes sociais.

Sendo possível planejar a visitação de pontos turísticos com base na média de visitação dos mesmos.

A quantidade de pessoas em um local auxilia na obtenção de informações para consumidores e fornecedores sobre o nível de interesse no ambiente, uma vez que, se o ambiente possui um fluxo alto de movimentação, o mesmo apresenta um diferencial entre os demais. A movimentação influencia no critério de escolha quanto ao lugar que deseja ir. Com os resultados obtidos, é possível planejar o horário de visitação, baseado em seu horário de maior ou menor movimentação.

2. Metodologia

Nesta seção serão descritas as etapas realizadas no desenvolvimento do trabalho. Será apresentado a configuração do ambiente e o armazenamento, coleta e tratamentos dos dados.

De acordo com a FIGURA 2, configurou-se um ambiente em nuvem para execução do *script* de coleta e o armazenamento de todos os dados coletados. Coletou-se *tweets* recentes com geolocalização por meio da *API* disponibilizada pelo *Twitter*. Aplicou-se o processo de Mineração de Dados (*Data Mining*) para o tratamento dos dados coletados. Após toda análise, a aplicação retornou a quantidade de pessoas nos ambientes estudados baseados na média dos *tweets* coletados. Os resultados obtidos foram representados graficamente para melhor entendimento e visualização do fluxo de movimentação do ambiente.



FIGURA 2 - Desenvolvimento.

2.1 Configurações do Ambiente

Para realização do trabalho, seria necessário um ambiente o qual permanece operante durante todo o período da coleta, por se tratar de uma coleta diária. Coletou-se assim, um grande volume de *tweets*. Por esse motivo, optou-se em realizar a coleta em um ambiente em nuvem. Por meio do ambiente configurado realizou-se a coleta dos dados durante os 20 dias pré-estabelecidos, conectado durante todos os dias sem sobrecarregar uma máquina física.

2.2 Armazenamento dos Dados

Atualmente, para o estudo de Big Data não é usual utilizar bancos de dados relacionais, visto que o grande volume de dados e a variedade dos mesmos tornam a usabilidade de tais bancos difíceis e não tão eficientes (Vieira et al. 2012). Para o armazenamento e processamento de tantas informações propôs-se o banco de dados *NoSQL* (*Not only SQL*) (Vieira et al. 2012).



Para o armazenamento dos dados da pesquisa, utilizou-se o banco de dados não relacional, *MongoDB*. Juntamente com o *Robomongo*, *software* utilizado para processamento e gerenciamento dos dados armazenados no *MongoDB*.

2.3 Coleta do Dados

Para a implementação do *script* de coleta, utilizou-se a linguagem *Python*, uma das mais populares linguagens de programação ao se trabalhar com Big Data e mineração de dados. Por meio da API disponibilizada pelo *twitter*, coletou-se *tweets* realizados recentemente no raio de 50 metros em relação a latitude e longitude dos mesmos. Realizou-se requisições ao *Twitter* com tais informações, assim que encontrado cada *tweet* compatível com os parâmetros definidos, os mesmos foram armazenados em sua respectiva coleção, de acordo com sua localização.

No banco de dados *MongoDB* criou-se uma coleção para cada ponto turístico a fim de obter melhor organização das informações e para facilitar na manipulação dos dados posteriormente. Após a coleta exportou-se cada coleção para um arquivo “.csv” para o processo de *Data Mining*.

Coletou-se um total de 3.379.842 *tweets* de alguns dos principais pontos turísticos do mundo. Sendo eles: *Big Ben* (510.534\$), *Burj Khalifa* (468.273), *Cataratas do Iguaçu* (402.278), *Central Park* (52.103\$), *CN Tower* (447.502), *Cristo Redentor* (534.904), *Musée Du Louvre* (26.481), *Oxford Street* (549.459), *Statue Of Liberty* (184.546) e *One World Trade Center* (203.762). Conforme apresentado na FIGURA 3 a qual exibe a quantidade total de *tweets* coletados dos pontos turísticos durante os 20 dias. A FIGURA 4, exibe a média diária de *tweets* dos pontos turísticos sendo *Big Ben* (25.527), *Burj Khalifa* (23.414), *Cataratas do Iguaçu* (20.114), *Central Park* (2.605), *CN Tower* (22.375), *Cristo Redentor* (26.745), *Musée Du Louvre* (1.324), *Oxford Street* (27.473), *Statue Of Liberty* (9.227) e *One World Trade Center* (10.188).

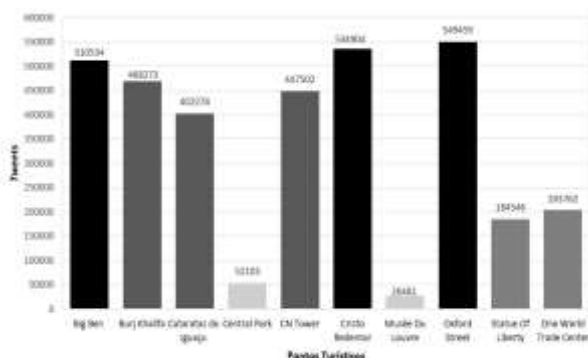


FIGURA 3 -Total de *tweets* coleta por pontos turísticos.

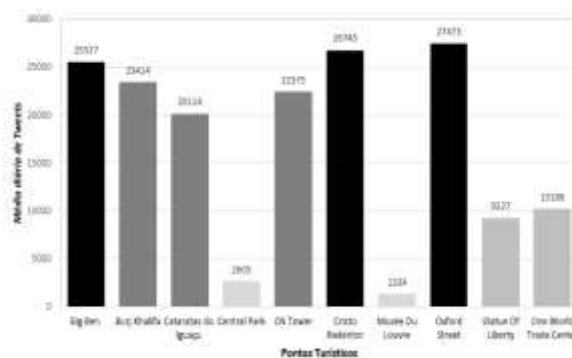


FIGURA 4 - Média diária de *tweet* coletados.

2.4 Pre-processamento

Devido à grande variedade dos dados armazenados, não seria confiável realizar o estudo das informações adquiridas sem realizar o tratamento dos dados primeiramente. Aplicou-se assim, as etapas do processo de mineração de dados, a fim de obter-se informações confiáveis para os resultados e validações da pesquisa.



A etapa de seleção dos dados, realizou-se durante a execução de um *script*, onde os dados relevantes para o processo de mineração de dados foram selecionados e coletados. Os arquivos foram exportados, cada um segundo sua coleção, para seu respectivo arquivo *CSV* (*Comma-separated values*) para realizar o tratamento dos dados. No Pré-processamento e formatação, executou-se um *script* para formatação dos dados, uma vez que os mesmos estavam com caracteres especiais dificultando assim a leitura da data de criação do *tweet*. Utilizou-se também um *script* para realizar a quebra da variável data de criação do *tweet*, permitindo utilizar apenas o dia e hora de cada *tweet* coletado.

Na etapa de mineração de dados, os dados foram organizados e formatados, excluindo-se então *tweets* duplicados, ou seja, *tweets* de um mesmo usuário e com o mesmo texto com a mesma data de criação. Para realizar tal formatação, utilizou-se o *software Weka* (*Waikato Environment for Knowledge Analysis*), que permite a análise e estatística de dados. Na etapa de Interpretação/Avaliação, observou-se que após todas as etapas do processo de mineração de dados, os dados obtidos eram aceitáveis para realização da proposta do trabalho.

2.5 Extração do Conhecimento

Sendo assim, após todo o processo de coleta e tratamento dos dados, os mesmos foram avaliados a fim de verificar o resultado obtido. Conforme a FIGURA 5 os resultados foram representados graficamente para uma melhor visualização. O eixo x do gráfico representa o período do dia em horas (0 a 23). O eixo y representa a média de *tweets* realizados na hora do dia.

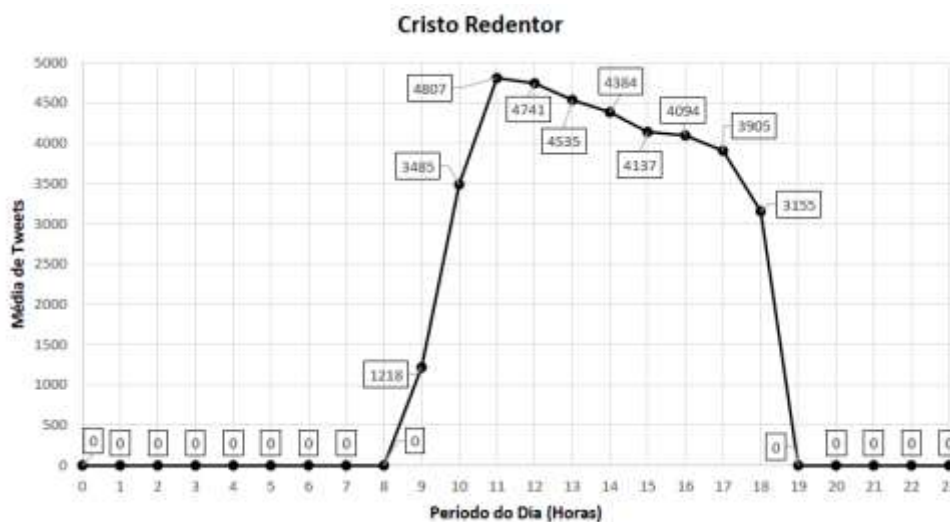


FIGURA 5 - Predição de Visitação do Cristo Redentor.

A Figura 5 representa a predição de visitação do ponto turístico Cristo Redentor. É possível observar a partir dos dados coletados, que o mesmo não possui movimentação das 00:00 às 08:00 e das 19:00 às 23:00. A explicação para a falta de informações (*tweets*) nesse horário, está representada na FIGURA 6, que representa o horário de funcionamento do Cristo Redentor, o está aberto das 08:00 às 19:00 horas se segunda-feira a domingo, confirmando o motivo de ausência de informações coletadas.

Horário: domingo	08:00–19:00
segunda-feira	08:00–19:00
terça-feira	08:00–19:00
quarta-feira	08:00–19:00
quinta-feira	08:00–19:00
sexta-feira	08:00–19:00
sábado	08:00–19:00



FIGURA 6 - Horário de funcionamento do Cristo Redentor. FIGURA 7 – Horário de Pico do Cristo Redentor.

Por meio da FIGURA 5 é possível observar também que o horário com a média de maior movimentação, seria no horário de 11:00 horas, com uma média de 4807 *tweets*. Após esse horário de maior movimentação, é possível observar uma queda no fluxo, porém, das 13:00 as 19:00, horário de encerramento, é possível observar um fluxo contínuo de visitação.

Para validação do fluxo de movimentação, foi verificada as informações obtidas e representadas na FIGURA 6 é comparada com o horário de pico do Cristo Redentor conforme FIGURA 7. Tais informações são disponibilizadas pelo *Google Maps*, serviço de geolocalização da *Google*, o qual disponibiliza mapas, rotas, horário de funcionamento, horário de pico entre outras informações do local desejado.

3. Resultados

Para obtenção dos resultados, executou-se um *script* contabilizador, o qual soma a quantidade de *tweets* realizados por dia em cada hora do dia. Utilizou-se o cálculo da média aritmética simples para obter o resultado, ou seja, a média de *tweets*. Sendo assim a média de *tweets* é igual à quantidade total de *tweets* por hora dividido pela soma dos dias que houveram *tweets* neste horário.

Os *tweets* coletados possuem a data de criação formatados conforme a ISO 8601, a qual padroniza os horários conforme o Meridiano de *Greenwich*. Sendo assim, fez-se a modificação das datas de criação mediante ao fuso horário de cada país: *Big Ben* -1:00, *Burj Khalifa* -4:00, Cataratas do Iguazu +3:00, *Central Park* -4:00, *CN Tower* -4:00, Cristo Redentor +3:00, *Musée Du Louvre* -2:00, *Oxford Street* -1:00, *Statue Of Liberty* -4:00 e *One World Trade Center* -4:00.

Durante os dias 12 a 27 de março, foram coletados 510.534 *tweets* realizados no ponto turístico *Big Ben*. A FIGURA 8 aponta que das 12 horas as 23 horas ocorre uma grande movimentação de visitantes no local. Atingindo uma grande concentração às 21:00 e 22:00 horas.

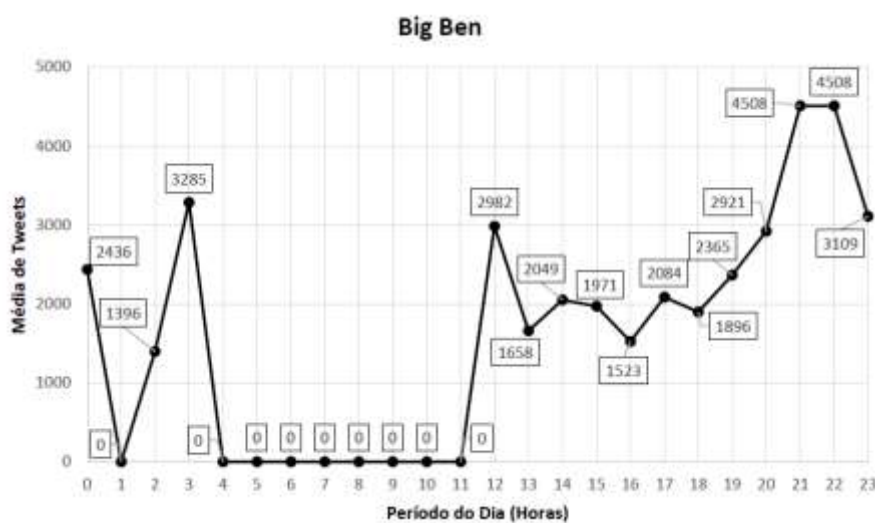


FIGURA 8 - Visitação Big Ben.

O *Burj Khalifa* abre as 4:30 da manhã para visitantes que desejem observar o nascer do sol e encerra seu funcionamento às 23:00 horas. É possível observar conforme na FIGURA 9, que há um grande fluxo de movimentação no maior arranha-céus do mundo, ocorre durante as 17 e 18 horas, horário para admirar o pôr-do-sol. Caindo o fluxo a partir das 19 horas às 23 horas, horário de encerramento das visitas.

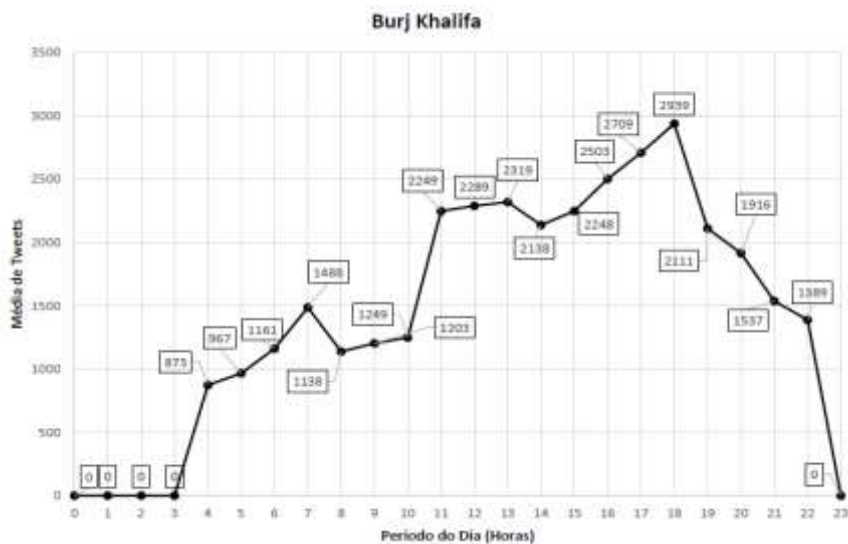


FIGURA 9 - Visitação Burj Khalifa.

Com o horário de funcionamento entre as 9:00 horas e as 17:00 horas, é possível observar na FIGURA 10, que das 9:00 as 13:00 horas a um crescimento na visitação das Cataratas do Iguaçu, atingindo a maior média de visitação às 13:00 horas. A partir desse horário há uma queda de visitação até o horário de encerramento.



FIGURA 10 - Visitação Cataratas do Iguaçu.

Central Park, que possui horário de funcionamento de 06:00 horas as 01:00 hora. Segundo as informações disponibilizadas pelo *Google Maps*, o mesmo possui horário de pico das 12:00 horas às 16:00 horas. É possível analisar essa maior movimentação, pelo gráfico indicado na FIGURA 11.

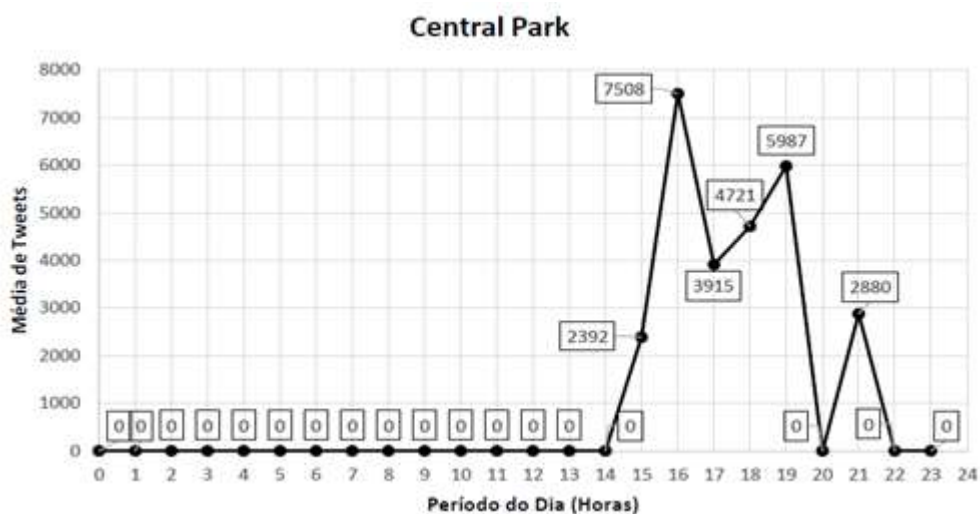


FIGURA 11 - Visitação *Central Park*.

Com os resultados apresentados na FIGURA 12 referente ao *CN Tower*, é possível perceber que a movimentação do mesmo ocorre em grande quantidade a partir das 15 horas e mantém recebendo um grande volume de visitantes durante todo o período da tarde até 21 horário de encerramento de visitação do parque.

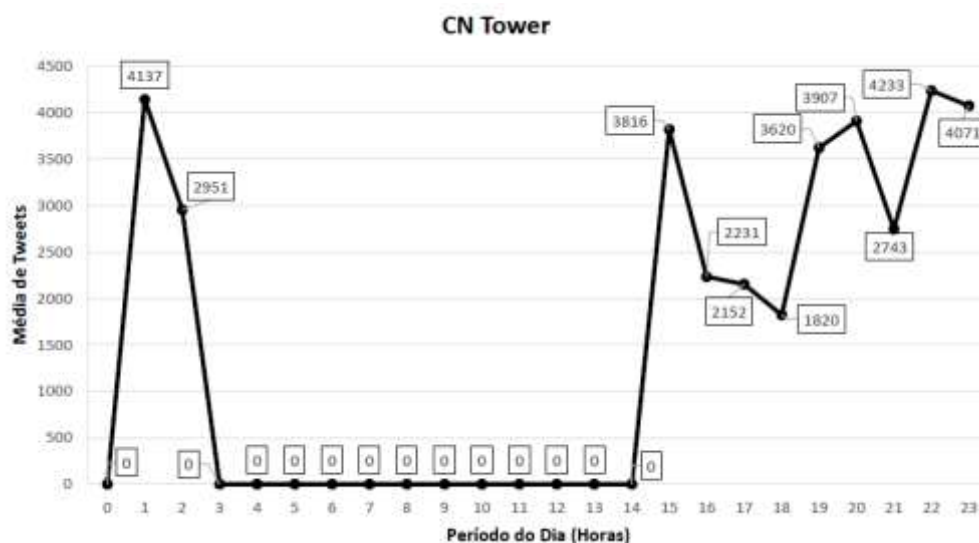


FIGURA 12 - Visitação *CN Tower*.

A FIGURA 13 representa a predição de visitação do Cristo Redentor. Pode-se observar que durante seu horário de funcionamento das 13:00 horas às 19:00 horas uma concentração contínua de visitas. Durante todo o dia, o ponto turístico possui uma grande movimentação, atingindo um maior fluxo as 11:00 horas.



FIGURA 13 - Visitação Cristo Redentor.

Ao se coletar dados de pontos turísticos de partes diferentes do mundo, é possível observar que os visitantes do *Museu do Louvre* não realizam *tweets* durante as visitas, dificultando assim a comparação e análise da movimentação dos visitantes. A FIGURA 14 exibe uma média de visitação às 12 horas e as 18 horas, próximo ao horário de encerramento.

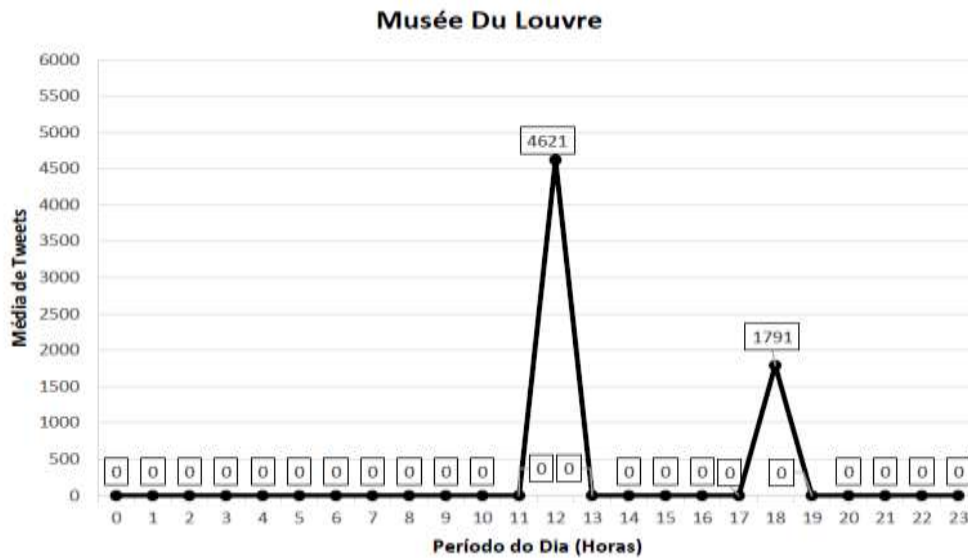


FIGURA 14 - Visitação *Musée Du Louvre*.

Ao analisar uma das avenidas mais movimentada do mundo e a mais movimentada da Europa é possível observar que a movimentação na *Oxford Street* inicia-se a partir de 1:00 hora da manhã tem uma queda de movimentação e consegue manter um grande fluxo de movimentação durante todo o dia das 10 horas às 22 horas, conforme é possível observar na FIGURA 15.

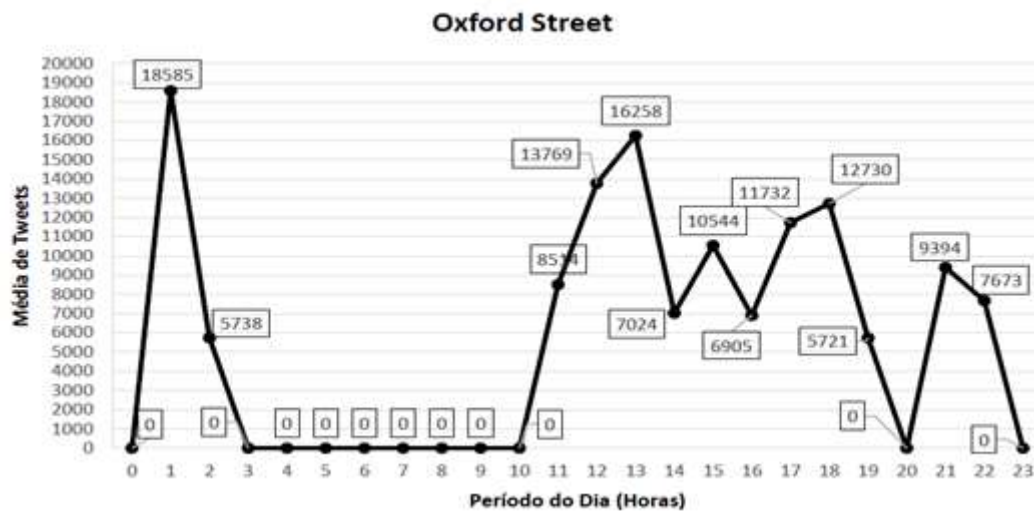


FIGURA 15 - Visitação *Oxford Street*.

Foram coletados 184546 da Estátua da Liberdade no decorrer dos dia 2 ao dia 23 de março. Por meio da FIGURA 16 observa-se que das 8:00 horas às 13:00 horas o aumento da movimentação gradativamente, tendo uma média maior movimentação das 14:00 horas às 15:00 horas.

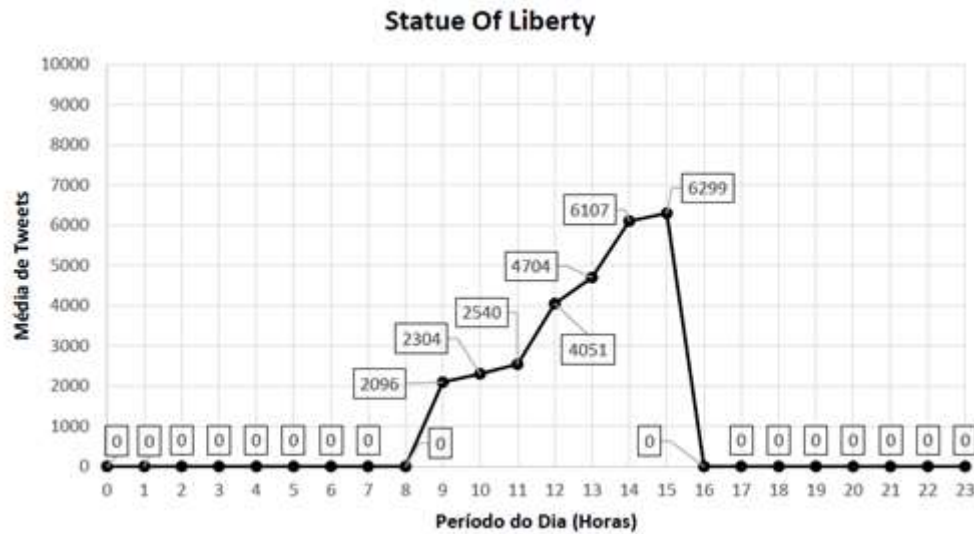


FIGURA 16 - Visitação *Statue Of Liberty*.

O edifício *One World Trade Center*, edifício localizado no mesmo local que o World Trade Center, possui horário de funcionamento das 8:00 horas às 22:00. Durante o período de coleta foi observado que os horários de maior visitação são entre as 11:00 horas às 13:00 horas, conforme exibido na FIGURA 17.

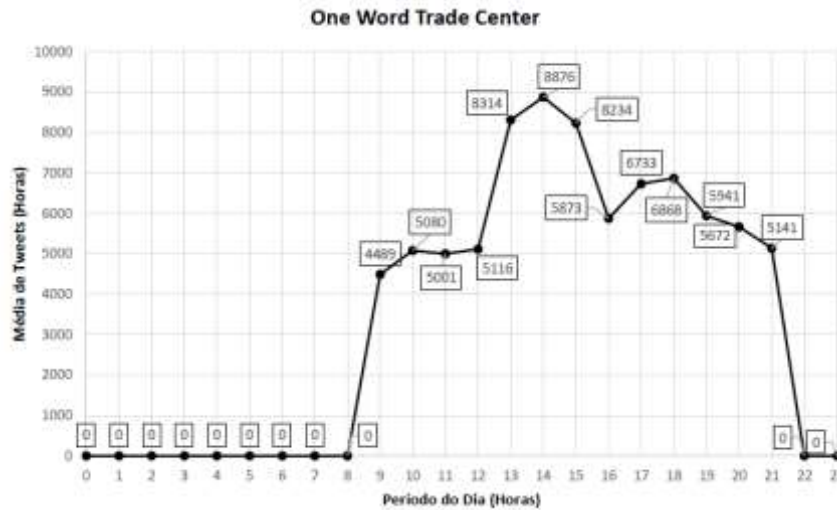


FIGURA 17 - Visitação *World Trade Center*.

4. Conclusão

Por meio do trabalho realizado e dos resultados obtidos, conclui-se que atualmente dados de larga escala são possíveis de ser analisados de forma rápida e eficaz por meio do sensoriamento participativo. Sendo possível de apenas um único lugar seja possível coletar, analisar e criar soluções de vários outros lugares.

Com os resultados obtidos é possível tanto para visitantes quanto para fornecedores analisar o fluxo de movimentação dos pontos turísticos e como agir de acordo com maior ou menor movimentação do mesmo. Tais informações são importantes aos turistas, que poderiam



SICT
Semana de Iniciação
Científica e Tecnológica

25 a 29 de setembro de 2017
Engenharias e Computação

 Universidade de Itaúna

evitar tumultos e ter um melhor proveito do local visitado. Ou caso prefiram, o grande fluxo de movimentação também pode gerar um grande fluxo de culturas diferentes, enriquecendo ainda mais a visita. Para os fornecedores, esses dados seriam valiosos para verificar expediente de trabalho, quantidade de funcionários e estoque.

5. Referências

ABOWD, G. D., DEY, A. K., BROWN, P. J., DAVIES, N., SMITH, M., and STEGGLES, P. (1999). Towards a better understanding of context and context-awareness. In International Symposium on Handheld and Ubiquitous Computing, pages 304–307. Springer.

ALECRIM, E. (2015). *O que é Big Data?*

BURKE, J. A.; ESTRIN, D.; HANSEN, M.; PARKER, A.; RAMANATHAN, N.; REDDY, S. and SRIVASTAVA, M. B. (2006). *Participatory Sensing*. Center for Embedded Network Sensing.

FRANÇA, T. C.; DE FARIA, F. F.; RANGEL, F. M.; DE FARIAS, C. M. and OLIVEIRA, J. (2014). *Big Social Data: Princípios sobre Coleta, Tratamento e Análise de Dados Sociais*. XXIX Simposio Brasileiro de Banco de Dados–SBBD, 14.

SILVA, T. H.; DE MELO, P. O. V.; ALMEIDA, J. M.; VIANNA, A. C.; SALLES, J. and LOUREIRO, A. A. (2014b). *Definição, Modelagem e Aplicações de Camadas de Sensoriamento Participativo*.

VIEIRA, M. R.; FIGUEIREDO, J.; LIBERATTI, G. and VIEBRANTZ, A. F. M. (2012). *Bancos de dados nosql: conceitos, ferramentas, linguagens e estudos de casos no contexto de big data*. Simpósio Brasileiro de Bancos de Dados.

WEISNER, M. (1996). *Ubiquitous Computing*. Mark Weisner Personal Homepage.