

Um Método de Detecção de Regiões Funcionais Utilizando Dados de Redes Sociais

Alice A. F. Menezes¹, Josiel W. V. Santos¹, Bruno Á. Souza¹, Thais G. Almeida¹,
Fabiola G. Nakamura¹, Eduardo F. Nakamura¹, Carlos M. S. Figueiredo²

¹Universidade Federal do Amazonas (UFAM) – Manaus, AM – Brasil

²Universidade do Estado do Amazonas (UEA) – Manaus, AM – Brasil

{alice.menezes, jwvs, bruno.abia, tga, fabiola, nakamura}@icompufam.edu.br,
cfigueiredo@uea.edu.br

Abstract. *Social Sensing is an emergent research area due to the amount of useful information that are provided by users of social networks. Several works in literature present studies with applications in natural disaster detection, traffic monitoring or analyzing dynamic of cities. In this way, this work presents a method to detect functional regions using data from location-based social networks. A dataset of 157,054 check-ins from Foursquare was used to characterize subregions in an urban area, determining functions for them. Comparing the results with government data, we verified that an accuracy of 86% was obtained for the proposed method. In addition, we also present two study cases for it: one related to traffic occurrences, and other related to urban area dynamics.*

Resumo. *O Sensoriamento Social é uma área de pesquisa emergente devido à quantidade de informações úteis que são fornecidas pelos usuários de redes sociais. Vários trabalhos na literatura apresentam estudos com aplicações em detecção de desastres naturais, monitoramento de tráfego ou análise da dinâmica de cidades. Nesta direção, este trabalho apresenta um método para a detecção de regiões funcionais utilizando dados de redes sociais baseadas em localização. Foi utilizado um conjunto de dados de 157.054 check-ins do Foursquare, com o objetivo de caracterizar subregiões em uma área urbana, determinando funções para as mesmas. Comparando os resultados com dados governamentais, verificamos que foi obtida uma acurácia de 86% para o método proposto. Além disso, também apresentamos dois estudos de caso para o mesmo: um relacionado à análise de ocorrências de trânsito, e outro relacionado à dinâmica das áreas urbanas.*

1. Introdução

Celulares e redes sociais, originalmente concebidos como ferramentas de comunicação, estão sendo cada vez mais utilizados como ferramentas inovadoras de coleta de dados para aplicações de monitoramento social e urbano [Steenbruggen et al. 2015]. Celulares, além de possuírem sensores como GPS, acelerômetro, microfone, câmera e giroscópio, estão cada vez mais conectados e seus usuários compartilhando dados voluntariamente em redes sociais. Tal fenômeno possibilitou o surgimento de métodos que permitem a observação do comportamento de cidades e seus ambientes de forma dinâmica e automatizada, o que vem sendo chamado de Sensoriamento Social [Aggarwal and Abdelzaher 2013].

Aliado a técnicas de Mineração de Dados, o Sensoriamento Social pode ser utilizado como uma fonte de informação útil sobre diversos aspectos do nosso cotidiano. Desta forma, aplicações interessantes vêm surgindo, como estudos de comportamentos epidemiológicos [Madan et al. 2010], de monitoramento socioeconômico [Liu et al. 2015], de comportamento social [Moturu et al. 2011] e de análise da dinâmica das cidades [Silva et al. 2012]. Tais estudos podem contribuir significativamente para auxiliar a tomada de decisões sobre planejamento urbano, tornando as cidades mais inteligentes e, conseqüentemente, melhorando a qualidade de vida de seus habitantes.

Dentre os trabalhos da literatura, há vários que aplicam sensoriamento social a partir do Twitter¹ e do Foursquare² devido a sua popularidade [Sakaki et al. 2010, Aiello et al. 2013, Silva et al. 2013]. Em particular, o Foursquare é uma rede social baseada em localização [Traynor and Curran 2012] que permite que seus usuários realizem *check-ins* e expressem suas opiniões, fazendo um breve resumo das experiências vividas no local por meio de comentários [Almeida et al. 2016, Menezes et al. 2016]. Isto possibilita estudos como a análise de mobilidade e detecção de regiões funcionais, alvo do estudo deste artigo.

Regiões funcionais são áreas, em um perímetro urbano, nas quais alguma função ou característica é associada [Cranshaw et al. 2012]. Esta função pode estar relacionada a aspectos econômicos, sociais, de mobilidade e etc. Como exemplos de função, podemos citar se uma área é residencial, relacionada a comércio ou diversão noturna. A verificação da distribuição de diferentes regiões funcionais é um importante fator em estudos urbanos, como para o planejamento do crescimento urbano, mas que antes da ascensão das redes sociais consistia na realização de custosos censos, que poderiam levar alguns anos para serem atualizados [Zhi et al. 2016]. A detecção de regiões funcionais utilizando dados sociais permite um monitoramento dinâmico e de baixo custo, o que facilita a tomada de decisões e execução de ações preventivas de acordo com o objeto de estudo.

Neste contexto, este trabalho apresenta um método de detecção de regiões funcionais que utiliza dados do Foursquare. O método associa dados sociais com regiões em uma determinada área urbana. Assim, o cotidiano das pessoas que habitam a área irá definir as funções das regiões. Como resultado, obtivemos uma acurácia de 86% usando como alvo a cidade de Nova Iorque, cidade que possui dados oficiais de planejamento urbano disponíveis. Além disso, este trabalho também apresenta como contribuição duas aplicações de estudo de caso: uma para auxiliar na análise de acidentes de trânsito utilizando regiões funcionais, tornando possível o melhor direcionamento de ações para a prevenção destes, e outra para o monitoramento da dinâmica urbana, que permite observar como a cidade evolui e, conseqüentemente, como planejar seu futuro.

As próximas seções deste trabalho estão divididas da seguinte forma: na seção 2 apresentamos os trabalhos relacionados, na seção 3 apresentamos o método proposto. Na seção 4, apresentamos a base de dados utilizada, os experimentos realizados e o resultados obtidos. Na seção 5, apresentamos duas aplicações que utilizam regiões funcionais. Por fim, na seção 6, apresentamos as conclusões e as direções futuras para este trabalho.

¹<http://twitter.com/>

²<https://pt.foursquare.com/>

2. Trabalhos Relacionados

Os dados de redes sociais tornaram possível a análise da dinâmica das cidades, para a caracterização de usuários e regiões. Neste sentido, no trabalho de Noulas et al. [Noulas et al. 2011a], é apresentado o primeiro estudo sobre comportamento humano com dados provenientes do Foursquare. Os autores coletaram aproximadamente 12.000.000 *check-ins* de usuários da rede social. É apresentada uma análise da dinâmica geo-temporal de atividades coletivas de usuários, a fim de demonstrar como *check-ins* provêm meios para descoberta de padrões de comportamento humano diário e semanal, propriedades urbanas de vizinhanças e transições recorrentes entre diferentes atividades.

Em outro trabalho, Noulas et al. [Noulas et al. 2011b] demonstra como informações semânticas sobre localidades e atividades sociais podem ser exploradas no contexto de aplicações móveis e pesquisa científica. Os autores propõem o uso da categoria de lugares para criação de "impressões digitais" de usuários e localidades. Desta forma, pessoas podem ser caracterizadas a partir dos tipos de lugares que elas visitam, enquanto localidades podem ser modeladas de acordo com seus locais. Para tanto, os autores utilizam um algoritmo de agrupamento espectral [Shi and Malik 1997] para agrupar localidades e usuários.

No trabalho de Cranshaw et al. [Cranshaw et al. 2012], os autores apresentam uma nova metodologia para o estudo da dinâmica, estrutura e características de uma cidade em larga escala. Para isso, eles utilizaram uma abordagem na qual, a partir de uma quantidade massiva de dados geolocalizados provenientes do Foursquare, foi desenvolvido um algoritmo que mapeia áreas geográficas distintas de uma cidade fornecendo um visão em tempo real de qualquer mundaça no padrão de atividade da população. Tal algoritmo utiliza técnicas de agrupamento espectral e leva em consideração tanto a proximidade espacial dos usuários, quanto a proximidade social.

No trabalho de Silva et al. [Silva et al. 2012], os autores introduzem um nova técnica de visualização de dados, chamada *City Image*, cujo objetivo consiste em auxiliar o entendimento sobre a dinâmica de cidades. Para validar sua técnica, os autores utilizaram como base de dados 4,7 milhões de *check-ins* do Foursquare de usuários de oito cidades diferentes. O resultado de tal técnica é uma matriz quadrada que sumariza a dinâmica de cidades. Tal matriz é gerada com base em grafos de transição, modelados a partir da mobilidade de usuários de acordo com as categorias de localidades da rede social Foursquare.

Em Zang et al. [Zhang et al. 2013], os autores argumentam que a informação temporal referente à mobilidade humana é de extrema importância para análises urbanas, e portanto, deve ser considerada em conjunto com informações espaciais. Os autores utilizaram dados do Foursquare, provenientes das cidades de Nova Iorque e São Francisco, para analisar a influência da dinâmica temporal de atividades no estudo de mobilidade humana.

Em comparação com este trabalho, os trabalhos citados anteriormente não apresentam um estudo que relaciona as regiões funcionais e as atividades de seus habitantes, apesar de também utilizarem dados sociais.

No artigo de Zhi et al. [Zhi et al. 2016], os autores introduziram um novo modelo para detectar regiões funcionais baseado no modelo LRA (*Low Rank Approximation*). Para

tanto, utilizaram uma base de dados de aproximadamente 15 milhões de *check-ins* provenientes da cidade de Shanghai. Foram identificadas estruturas espaço-temporais latentes, que quando analisadas revelaram uma série de associações entre as atividades espaciais e temporais de cidadãos. Por meio do algoritmo K-means, os autores obtiveram 5 tipos de clusters que estavam diretamente relacionados com a combinação temporal de atividades dos cidadãos. Desta forma, os autores destacaram a correlação direta entre grupos de regiões e diferentes atividades durante períodos variados no decorrer no dia.

O trabalho citado determina funções para as regiões diante 5 aspectos, enquanto que este trabalho baseia-se nas categorias do Foursquare. Além disso, nosso método permite que as funções das regiões possam ser analisadas tanto em conjunto quanto de forma individual. Por fim, ressaltamos que este trabalho apresenta possíveis aplicações para a utilização das regiões funcionais.

Existem ainda trabalhos que utilizam dados de dispositivos móveis para determinar a característica das regiões. Este processo é mais custoso e a mobilidade dos usuários não está explicitamente ligada a uma atividade. No trabalho de Xiang et al. [Xiang et al. 2013], os autores realizam partições de regiões por meio da análise da trajetória de cidadãos utilizando um modelo de tópicos latentes (LDA), no qual são consideradas distribuições de probabilidades ao particionar regiões. A base de dados utiliza contém 500 milhões de chamadas de celulares provenientes de cidadãos da China. Ao invés de utilizar algoritmos de agrupamento baseados em similaridade espacial, os autores consideram informações temporais a fim de determinar regiões funcionais.

O trabalho aqui apresentado diferencia-se dos anteriores por levar em consideração o cotidiano dos habitantes da região estudada, mostrando que as regiões funcionais podem alterar sua função ao longo do tempo, permitindo que os aspectos da dinâmica das cidades seja analisado de forma espaço-temporal.

3. Detecção de Regiões Funcionais

Neste artigo, propomos um método para a detecção de regiões funcionais em uma área urbana da seguinte forma (Figura 1):

1. Definição de dados sociais geolocalizados, com funções relacionadas a eles;
2. Definição de uma grade com um limite definido sobre uma determinada área urbana;
3. Associação de dados sociais com as células da grade, de forma que cada célula possua diversas funções;
4. Ranqueamento das funções em cada célula, verificando a categoria do Foursquare com a quantidade predominante de *check-ins* e fazendo com que a mesma seja determinada a função principal da região;
5. Verificação e agrupamento de regiões adjacentes que possuam a mesma função principal.

Como resultado, obtemos regiões compostas por polígonos, formados por coordenadas, e funções extraídas de dados sociais compartilhados por usuários da área urbana estudada. As funções das regiões definidas podem variar em dias e meses por exemplo, devido às mudanças dos dados das redes sociais ao longo do tempo. Desta forma, podemos dizer que as regiões funcionais são estabelecidas através de atividades humanas.

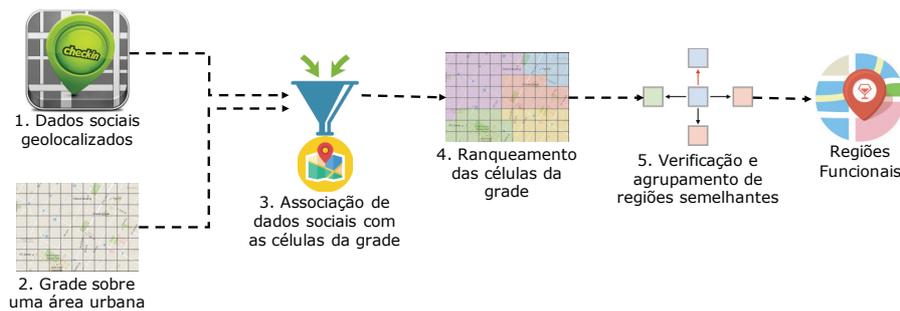


Figura 1. Método proposto.

De acordo com o método proposto, definimos que as regiões funcionais possuem as seguintes variáveis:

- Áreas formadas por uma lista de polígonos P , de forma que $P = \{p_1, p_2, p_3, \dots, p_i\}$ e i é o total de células da grade. O valor de i é definido no passo 2, pois ao estabelecermos a área em que a grade será formada, informamos quatro pontos, NLa , SLa , LLo e OLo referentes as latitudes norte e sul e as longitudes leste e oeste, respectivamente. Em seguida, esta área é dividida em subregiões pertencentes a lista P .
- Funções predominantes formadas por dados sociais geolocalizados F , em que $F = \{f_1, f_2, f_3, \dots, f_j\}$ e j é a quantidade de funções referentes a cada polígono pertencente a lista P . No passo 3, diversos dados sociais são associados a cada célula da grade, fazendo com que as mesmas possuam diversas funções. Desta forma, as funções predominantes são determinadas após o ranqueamento das funções das células da grade, no passo 4, em que apenas a função de maior ocorrência permanece.
- Intervalos de tempo T , de forma que $T = \{t_1, t_2, t_3, \dots, t_k\}$ e k é a quantidade de intervalos de tempo. Estes intervalos de tempo podem ser estabelecidos em horas, dias, semanas, meses e etc.

Assim, antes do passo 5, possuímos uma matriz que associa P a F para cada intervalo de tempo em T , de forma que $M = \{PF_{t_1}, PF_{t_2}, PF_{t_3}, \dots, PF_{t_k}\}$.

Por fim, para os intervalos de tempo em T , verificamos para cada região em P se as regiões adjacentes possuem a mesma função contida em F , de forma que as regiões semelhantes sejam interpoladas formando as matrizes contidas em M' , de forma que $M' = \{P'F'_{t_1}, P'F'_{t_2}, P'F'_{t_3}, \dots, P'F'_{t_k}\}$, P' possui tamanho $i' < i$ e F' possui tamanho $j' < j$.

O método proposto também permite que possamos analisar as regiões funcionais tanto em conjunto (todos os valores em F) quanto de forma individual. Assim, também é possível formar os polígonos em P , e posteriormente em P' , baseando-se apenas em uma função de estudo.

4. Experimentos e Resultados

Neste trabalho, definimos que as funções contidas em F são referentes às categorias do Foursquare (Tabela 1). Escolhemos esta rede social pois a mesma é baseada em

localização. Além disso, suas categorias podem estabelecer a função de uma região de acordo com as atividades dos habitantes da mesma.

Desta forma, as regiões funcionais geradas pelo método proposto estarão sempre de acordo com a atualidade, facilitando estudos relacionados, que em sua maioria são custosos se realizados por meios tradicionais.

Tabela 1. Exemplos de locais existentes nas categorias do Foursquare.

Categoria do Foursquare	Exemplos de locais
<i>Arts and Entertainment</i>	Cinemas, museus e casinos.
<i>College and University</i>	Escolas, laboratórios universitários e centros de estudo.
<i>Food</i>	Restaurantes, cafeterias e padarias.
<i>Nightlife Spot</i>	Bares, clubes de <i>rock</i> e boates.
<i>Outdoors and Recreation</i>	Parques, academias e praias.
<i>Professional and Other Places</i>	Fábricas, auditórios e centros médicos.
<i>Shop and Service</i>	Lojas, salões de beleza e supermercados.
<i>Travel and Transport</i>	Aeroportos, estradas e hotéis.
<i>Residence</i>	Propriedades privadas, prédios e reboques residenciais.

A área de estudo foi definida na cidade de Nova Iorque e regiões adjacentes, gerando uma lista P com 2.208 polígonos com lados de 1 quilômetro de comprimento. Em relação aos períodos de tempo em T , foram definidos 6, sendo eles: 00:00 a 03:59, 04:00 a 07:59, 08:00 a 11:59, 12:00 a 15:59, 16:00 a 19:59 e 20:00 a 23:59.

Escolhemos polígonos de 1 quilômetro de comprimento para uma análise detalhada das regiões. Já as seis faixas de horários foram escolhidas para que fosse possível verificar como as regiões mudam sua função ao longo do dia de acordo com a rotina dos habitantes. Ressaltamos que tanto o tamanho das células quanto os horários são configuráveis.

4.1. Descrição dos Dados

Para a realização dos experimentos, foram coletados 157.054 *check-ins* do Foursquare entre abril de 2014 e agosto de 2014. Estes dados foram coletados no respectivo ano para posterior validação do método, que foi comparado com dados governamentais do mesmo período. Na Figura 2, apresentamos a quantidade de *check-ins* por categoria do Foursquare com os mesmos períodos de tempo contidos em T .

Podemos observar que a categoria com o maior número de *check-ins* é *Food*, enquanto que a menor é *Residence*. Isto ocorre porque muitos usuários evitam registrar suas propriedades privadas em redes sociais por motivos de segurança. Apesar disso, podemos observar nos resultados que algumas regiões funcionais são referentes à áreas residenciais.

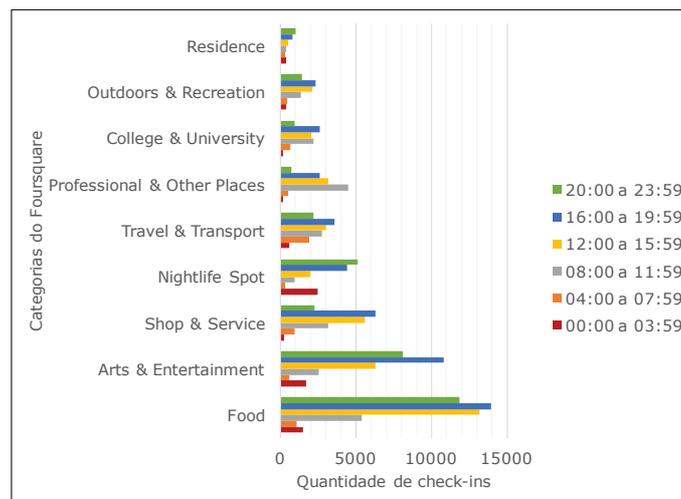


Figura 2. Quantidade de *check-ins* por categoria do Foursquare ao longo do dia.

4.2. Comparação com Dados Governamentais

Após a utilização dos dados sociais, o método proposto gerou regiões funcionais que mudam suas funções ao longo do tempo, de acordo com as atividades dos habitantes das mesmas. Como exemplo, podemos verificar na Figura 3, as regiões funcionais estabelecidas nos períodos 00:00 a 03:59 e 12:00 a 15:59.

No primeiro período (Figura 3(a)), as categorias com mais *check-ins* foram *Nightlife Spot*, *Arts & Entertainment* e *Food*, nesta ordem. Isto é refletido nas regiões funcionais estabelecidas. Da mesma forma, no segundo período (Figura 3(b)), as categorias com mais *check-ins* foram *Food* e *Shop & Service*, nesta ordem.

Além disso, podemos verificar que algumas áreas estão vazias. Isto ocorre porque, nestes períodos, não houveram dados sociais que refletissem a função da região. Isto pode ser resolvido aumentando-se a quantidade de dados sociais para estudo. Também é possível verificar o resultado da interpolação das regiões, que apresenta os polígonos de P' em diferentes formas.

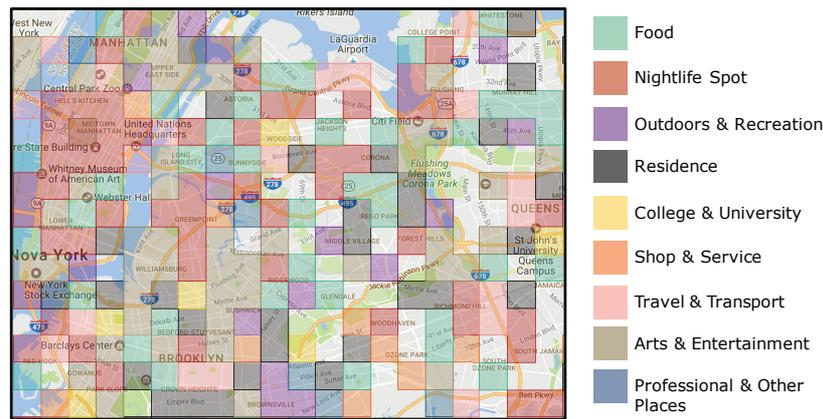
Através destes resultados, podemos verificar que o método proposto permite que possamos realizar diferentes estudos de uma região, pois a função da mesma se altera de acordo com a época do ano e o período do dia, de forma que reflete a rotina dos usuários que nela residem.

Para a validação do método proposto, foram utilizados dados do Departamento de Planejamento Urbano da cidade de Nova Iorque, denominados *City Owned and Leased Properties*³. Estes dados podem ser visualizados na aplicação *Zoning and Land Use Application*⁴.

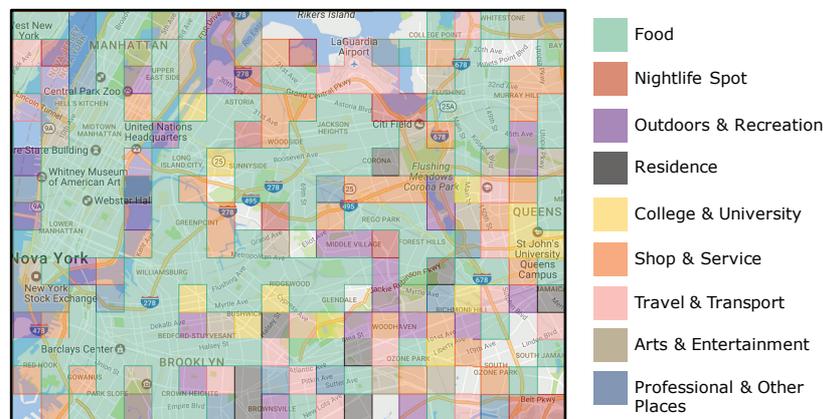
Os dados foram fornecidos em forma de uma planilha e estavam divididos por categorias que podem ser mapeadas para as categorias do Foursquare. Porém, os dados não eram geolocalizados, tornando necessária a utilização de uma API de mapa para transformar os endereços dos locais em coordenadas.

³<http://www1.nyc.gov/site/planning/about/publications/colp.page>

⁴<http://maps.nyc.gov/doitt/nycitymap/template?applicationName=ZOLA>



(a) 00:00 a 03:59



(b) 12:00 a 15:59

Figura 3. Resultado do método proposto para diferentes períodos de tempo, apresentando a mudança nas funções das regiões.

Em seguida, estes dados governamentais foram inseridos nas regiões funcionais cujas coordenadas geográficas coincidem. Por fim, mapeamos as categorias de locais nos dados governamentais para as categorias macro do Foursquare, de acordo com a hierarquia de categorias desta rede social⁵. A partir disso, criamos e comparamos as regiões funcionais de ambas as fontes de dados. Regiões idênticas eram contadas como acerto, enquanto regiões diferentes eram contadas como erro. Desta forma, apuramos que o nosso método possui uma acurácia de 86% para a caracterização de uma área urbana.

Com a similaridade encontrada, temos a contribuição de que o método proposto pode ser útil para entender o comportamento de cidades mesmo quando não se tem recursos necessários para realização de censos. Além disso, verificamos que os dados sociais podem ser utilizados como meio alternativo para o auxílio no planejamento urbano de uma cidade.

5. Utilização das Regiões Funcionais

Nesta seção, apresentamos duas possíveis aplicações para o método proposto, a análise de acidentes de trânsito e o monitoramento dinâmico das cidades. Estas aplicações mos-

⁵<https://developer.foursquare.com/categorytree>

tram como a utilização de regiões funcionais pode facilitar a tomada de decisões e ações preventivas em uma determinada região, de forma dinâmica e com baixo custo. Apesar do foco nestas duas aplicações, o método proposto também pode ser empregado para a avaliação econômica, social e de mobilidade da área urbana estudada.

5.1. Aplicação I: Análise de Acidentes de Trânsito

Com o objetivo de criar soluções eficientes para a prevenção de acidentes de trânsito, órgãos governamentais buscam analisar os fatores relacionados aos mesmos em uma determinada região. Aspectos como a localização e o horário em que os acidentes ocorreram são levados em consideração, assim como as condições das vias, condutores e veículos.

Desta forma, uma possível aplicação para as regiões funcionais, além do auxílio ao planejamento urbano, é a verificação dos fatores relacionados aos acidentes de trânsito em cada região detectada que subdivide uma área urbana. Para isto, definimos as regiões funcionais através dos dados do Foursquare (conforme apresentado na seção 4). Em seguida, utilizamos dados de acidentes de trânsito fornecidos pelo departamento de polícia da cidade de Nova Iorque⁶, com o objetivo de verificarmos a relação entre as regiões funcionais e os aspectos dos acidentes de trânsito.

Nestes dados do departamento de polícia, obtivemos informações sobre a data e a hora em que o acidente ocorreu, além da latitude, longitude, envolvidos no acidente e fatores que contribuíram para o mesmo. Após uma análise, dividimos os fatores em 4 categorias:

- **Direção imprópria:** ações como dirigir enquanto utiliza algum dispositivo móvel, fazer manobras proibidas na via e alta velocidade estão inclusas nesta categoria;
- **Motorista não condicionado:** condutores que estavam sob o efeito de bebidas alcoólicas ou drogas, estavam cansados ou perderam a consciência no momento do acidente estão inseridos nesta categoria;
- **Externo:** vias em condições ruins, animais e distrações provocadas por passageiros estão nesta categoria;
- **Veículo não condicionado:** falhas no acelerador, freios ou outras partes do veículo estão inseridas nesta categoria.

Além dos fatores descritos acima, também verificamos se o acidente ocasionou apenas danos materiais ou envolveu a morte do condutor ou de terceiros, como pedestres e ciclistas. No primeiro caso, o acidente foi classificado com **leve**. No segundo, como **grave**.

Como as regiões funcionais são representadas por polígonos com dados sociais relacionados a eles, aplicamos um algoritmo para verificar em qual região funcional os dados geolocalizados dos acidentes de trânsito (latitude e longitude) estavam inseridos. Assim, tornou-se possível associar os dados do departamento de polícia com os dados do Foursquare.

Desta forma, analisamos 73.182 acidentes de trânsito entre abril de 2014 e agosto de 2014, o mesmo período dos dados do Foursquare apresentados na seção 4. O resultado da análise foi dividido em 6 períodos de tempo para a verificação da mudança nas associações dos dados ao longo do dia (Figura 4).

⁶<https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95>

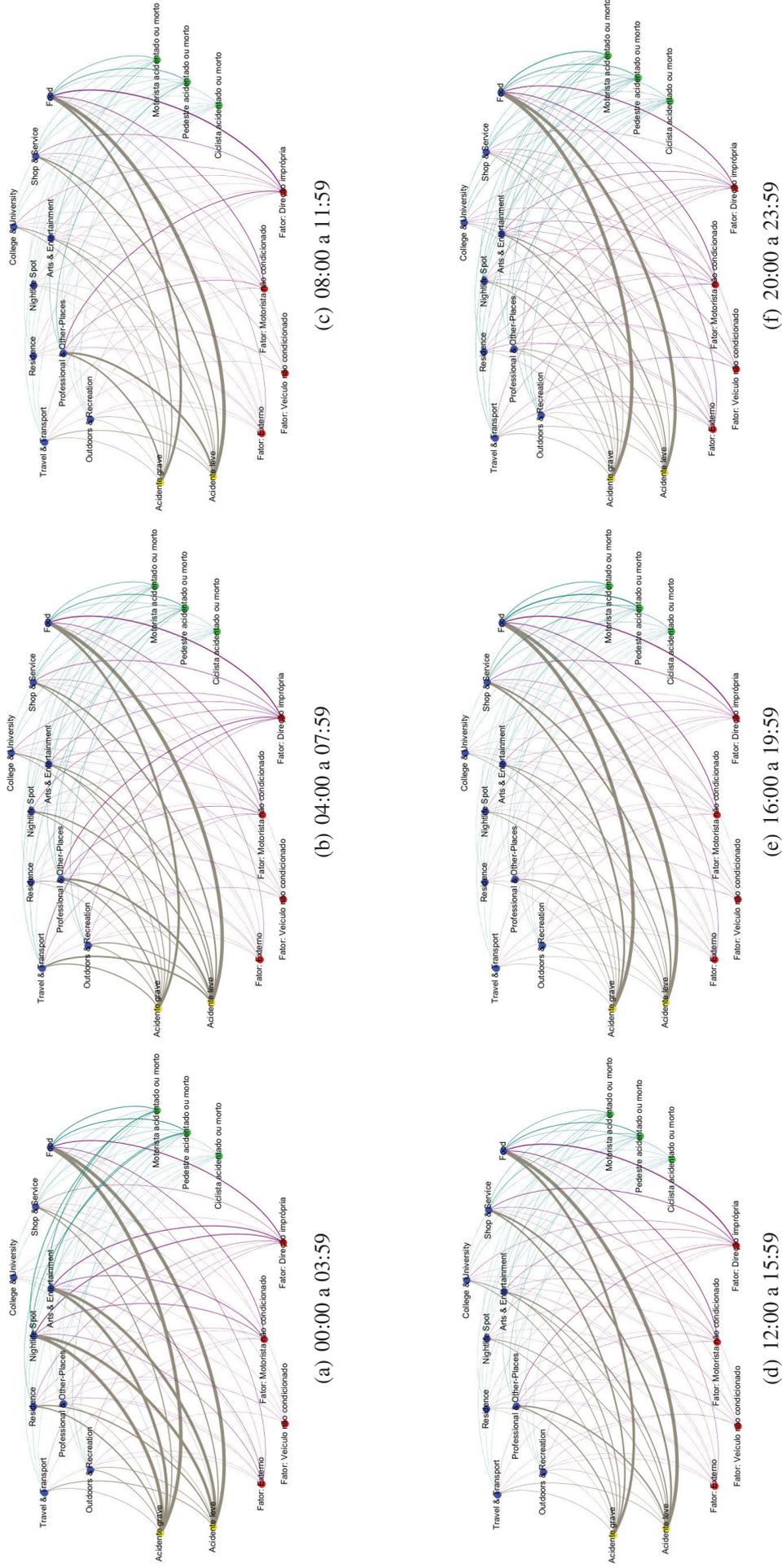


Figura 4. Representação da associação entre as regiões funcionais e os aspectos relacionados a acidentes de trânsito ao longo do dia.

Através das associações representadas na Figura 4 (sendo que quanto maior a incidência da associação, maior é a espessura da aresta), fizemos as seguintes observações relacionadas aos acidentes de trânsito:

- No período de 00:00 a 03:59 (Figura 4(a)), percebemos que a maioria dos acidentes, tanto graves quanto leves, estão associados às categorias *Nightlife Spot*, *Arts & Entertainment* e *Food*, assim como os fatores *Motorista não condicionado* e *Direção imprópria*. Isto sugere que os acidentes nestas regiões estão relacionados ao uso de substâncias não permitidas, como o álcool, por exemplo. Além disso, verificamos que estes mesmos acidentes ocasionaram a morte de motoristas e pedestres.
- No período de 04:00 a 07:59 (Figura 4(b)), percebemos que os fatores relacionados aos acidentes começam a ser direcionados para outras categorias do Foursquare, como *Professional & Other Places*, *College & University* e *Shop & Service*. Isto está diretamente relacionado à rotina das pessoas na região estudada, que passam a realizar suas atividades cotidianas neste período de tempo.
- No período de 08:00 a 11:59 (Figura 4(c)), percebemos que os acidentes passam a se concentrar nas categorias *Professional & Other Places* e *Food*, devido ao horário do *rush*, e que a maioria dos acidentes está relacionado a direção imprópria. Neste período, o índice de acidentes leves é maior do que o de acidentes graves, o que indica que, na maioria dos acidentes, ocorrem apenas danos materiais. Um padrão semelhante é apresentado na Figura 4(d).
- No período de 16:00 a 19:59 (Figura 4(e)), percebemos que a maioria dos acidentes estão concentrados nas categorias *Food* e *Shop & Service*, e estão relacionados aos fatores *Motorista não condicionado* e *Direção imprópria*. Além disso, o número de mortes de motoristas e pedestres é maior em relação aos períodos anteriores. Um padrão semelhante é apresentado na Figura 4(f), porém apenas com a categoria *Food*.

Esta aplicação é capaz de identificar os aspectos relacionados aos acidentes de trânsito e o que eles ocasionam, e pode ser utilizada para auxiliar na definição de soluções que minimizem os mesmos.

5.2. Aplicação II: Monitoramento dinâmico da cidade

Com a detecção de regiões funcionais por meio de redes sociais, tornou-se possível que o levantamento de informações necessárias para o planejamento urbano fosse realizado dinamicamente, conforme a evolução das cidades. Além disso, este método facilita a execução desta tarefa, uma vez que não há dados e censos bem estabelecidos e com baixo custo.

Outro fator importante é que os levantamentos necessários são realizados entre longos intervalos de tempo, de 2 a 4 anos, por exemplo. Com os dados provenientes de redes sociais, as regiões funcionais podem ser definidas e alteradas conforme os usuários que habitam a região compartilham suas informações.

Para mostrar como este monitoramento dinâmico é viável, apresentamos a Figura 5, na qual é feita uma comparação da formação das regiões funcionais entre os anos de 2014 e 2016. Para isto, foram coletados 70.339 *check-ins* do Foursquare entre julho de 2016 e agosto de 2016 na mesma área urbana.

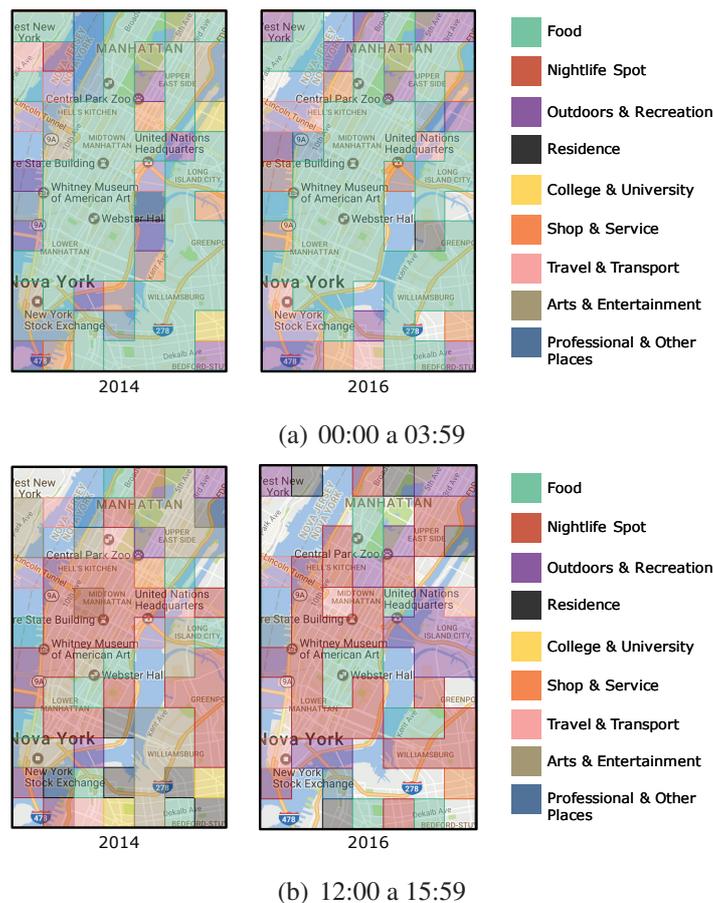


Figura 5. Comparação entre as regiões funcionais dos anos de 2014 e 2016 em diferentes períodos de tempo.

Podemos perceber, na Figura 5(a), que a categoria *Food* prevalece sobre as demais tanto no ano de 2014, quanto no ano de 2016. Isto é ocasionado devido ao horário em que as regiões funcionais foram formadas, de 12:00 a 15:59. Além disso, a presença de regiões funcionais com a categoria *Outdoors & Recreation* aumentou no ano de 2016 em relação ao ano de 2014. Também verificamos que, nas duas regiões funcionais em que anteriormente prevalecia a categoria *College & University*, agora prevalecem as categorias *Shop & Service* e *Outdoors & Recreation*. Estas mudanças podem ocorrer devido a expansão da cidade, construção de novos conjuntos residenciais e a criação de novos empreendimentos, por exemplo.

Na Figura 5(b), a comparação é realizada em um horário noturno, de 00:00 a 03:59. Neste horário, também verificamos o aumento na aparição de regiões funcionais com a categoria *Outdoors & Recreation*, apesar da categoria *Nightlife Spot* prevalecer sobre as demais. A categoria *College & University* também foi eliminada.

Desta forma, com esta aplicação, verificamos uma forma alternativa, de baixo custo, para obter as informações necessárias para o auxílio ao planejamento de uma área urbana. Isto torna possível não só que este planejamento seja realizado de forma dinâmica, como viabiliza que áreas urbanas com poucos recursos também consigam estas informações.

6. Conclusão e Trabalhos Futuros

Neste trabalho, apresentamos um método para a detecção de regiões funcionais através de redes sociais baseadas em localização. O método consiste na utilização de polígonos para a formação das regiões, além dos dados sociais e de intervalos de tempo. Desta forma, as regiões funcionais estão diretamente relacionadas a rotina dos habitantes da área urbana estudada, capturando inclusive sua dinâmica.

Como experimento, o método foi aplicado à cidade de Nova Iorque, e os resultados obtidos foram comparados com dados governamentais oficiais disponíveis. Com isso, apuramos que o método proposto obteve uma acurácia de 86% na definição das regiões funcionais, o que o torna uma alternativa viável para o auxílio no planejamento urbano em outras cidades onde não há tal monitoramento de forma sistemática. Suas principais vantagens são o pequeno custo de monitoramento e a possibilidade de observar o comportamento dinâmico das cidades e seus habitantes em intervalos de tempos muito menores que os métodos tradicionais. Não apresentamos experimentos com outras localidades, como as cidades brasileiras, porque apesar de ser possível obter dados do Foursquare para as regiões funcionais, não haveria como validar os resultados obtidos, pois os dados governamentais disponíveis são poucos, além de serem dispersos e incompletos.

Por fim, apresentamos duas aplicações para a qual as regiões funcionais podem ser utilizadas, além do planejamento urbano. Em uma das aplicações, é apresentado um estudo sobre os aspectos de acidentes de trânsito, relacionando-os com as regiões funcionais encontradas. Na outra, é apresentado como o monitoramento dinâmico de uma área urbana pode ser realizado através do método proposto.

Como trabalhos futuros, pretendemos analisar outras formas de definir a função predominante de uma região e verificar outras redes sociais baseadas em localização, que podem complementar os resultados encontrados, tendo em vista que diferentes países e cidades podem utilizar mais uma rede social do que outra. Além disso, iremos demonstrar através de outras aplicações a extensão deste estudo para avaliações econômicas, sociais e de mobilidade urbana.

7. Referências

- Aggarwal, C. and Abdelzaher, T. (2013). Social sensing. *Managing and Mining Sensor Data*, pages 237–297.
- Aiello, L., Petkos, G., Martin, C., Corney, D., Papadopoulos, S., Skraba, R., Göker, A., Kompatsiaris, I., and Jaimes, A. (2013). Sensing trending topics in twitter. *IEEE Transactions on Multimedia*, 15:1268–1282.
- Almeida, T., Souza, B., Menezes, A., Figueiredo, C., and Nakamura, E. (2016). Sentiment analysis of portuguese comments from foursquare. In *Proceedings of the 22nd Brazilian Symposium on Multimedia and the Web*.
- Cranshaw, J., Schwartz, R., Hong, J., and Sadeh, N. (2012). The livelihoods project: Utilizing social media to understand the dynamics of a city. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.
- Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., and Shi, L. (2015). Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, 105:512–530.

- Madan, A., Cebrian, M., Lazer, D., and Pentland, A. (2010). Social sensing for epidemiological behavior change. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*.
- Menezes, A., Almeida, T., Gatto, B., Santos, E., Figueiredo, C., and Nakamura, E. (2016). Poi: uma aplicação de detecção de pontos de interesse. In *Proceedings of the 13th Simpósio Brasileiro de Sistemas Colaborativos (SBSC)*.
- Moturu, S., Khayal, I., Aharony, N., Pan, W., and Pentland, A. (2011). Using social sensing to understand the links between sleep, mood, and sociability. In *Proceedings of 3th IEE International Conference on Social Computing Privacy, Security, Risk and Trust (PASSAT)*.
- Noulas, A., Scellato, S., Mascolo, C., and Pontil, M. (2011a). An empirical study of geographic user activity patterns in foursquare. In *Proceedings of the 15th International AAAI Conference on Weblogs and Social Media*.
- Noulas, A., Scellato, S., Mascolo, C., and Pontil, M. (2011b). Exploiting semantic annotations for clustering geographic areas and users in location-based social networks. In *Proceedings of the 15th International AAAI Conference on Weblogs and Social Media*.
- Sakaki, T., Okazaki, M., and Matsuo, Y. (2010). Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*.
- Shi, J. and Malik, J. (1997). Normalized cuts and image segmentation. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*.
- Silva, T., Vaz de Melo, P., Almeida, J., Salles, J., and Loureiro, A. (2013). A comparison of foursquare and instagram to the study of city dynamics and urban social behavior. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*.
- Silva, T. H., de Melo, P. O. V., Almeida, J. M., Salles, J., and Loureiro, A. A. (2012). Visualizing the invisible image of cities. In *Green Computing and Communications (GreenCom), 2012 IEEE International Conference on*, pages 382–389. IEEE.
- Steenbruggen, J., Tranos, E., and Nijkamp, P. (2015). Data from mobile phone operators: A tool for smarter cities? *Telecommunications Policy*, 39:335–346.
- Traynor, D. and Curran, K. (2012). Location-based social networks. *From Government to E-Governance: Public Administration in the Digital Age*, page 243.
- Xiang, F., Tu, L., Huang, B., and Yin, X. (2013). Region partition using user mobility patterns based on topic model. In *Proceedings of the 16th IEEE International Conference on Computational Science and Engineering (CSE)*.
- Zhang, K., Jin, Q., Pelechris, K., and Lappas, T. (2013). On the importance of temporal dynamics in modeling urban activity. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*.
- Zhi, Y., Li, H., Wang, D., Deng, M., Wang, S., Gao, J., Duan, Z., and Liu, Y. (2016). Latent spatio-temporal activity structures: a new approach to inferring intra-urban functional regions via social media check-in data. *Geo-spatial Information Science*, pages 1–12.