

Dissertação de Mestrado

**Um Padrão de Agrupamento de Cidades a Partir da
Dinâmica Social Urbana Extraída de Dados
Provenientes de Redes de Sensores Participativos**

Vilker Tenório Cabral Lobo
vilker.tenorio@gmail.com

Orientador:
Heitor Soares Ramos Filho

Maceió, Janeiro de 2017

Vilker Tenório Cabral Lobo

**Um Padrão de Agrupamento de Cidades a Partir da
Dinâmica Social Urbana Extraída de Dados
Provenientes de Redes de Sensores Participativos**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Modelagem Computacional de Conhecimento do Instituto de Computação da Universidade Federal de Alagoas.

Orientador: Heitor Soares Ramos Filho

Maceió, Janeiro de 2017

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central

Bibliotecária Responsável: Janaina Xisto de Barros Lima

L779p Lobo, Vilker Tenório Cabral.

Um padrão de agrupamento de cidades a partir da dinâmica social urbana extraída de dados provenientes de redes de sensores participativos/ Vilker Tenório Cabral Lobo. – 2018.

81 f. : il.

Orientador: Heitor Soares Ramos Filho.

Dissertação (Mestrado em Modelagem Computacional de Conhecimento) – Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2017.

Bibliografia: f. 77–81.

1. Computação urbana. 2. Redes de sensores participativos 3. Cadeias de Markov. 4. Dinâmica das cidades. I. Título.

CDU: 004.911



UNIVERSIDADE FEDERAL DE ALAGOAS/UFAL
Programa de Pós-Graduação em Modelagem Computacional de Conhecimento
Avenida Lourival Melo Mota, Km 14, Bloco 12, Cidade Universitária
CEP 57.072-900 – Maceió – AL – Brasil
Telefone: (082) 3214-1364/1825



Membros da Comissão Julgadora da Dissertação de Mestrado de Vilker Tenório Cabral Lobo, intitulada: “Um padrão de agrupamento de cidades a partir da dinâmica social urbana extraída de dados provenientes de redes de sensores participativos”, apresentada ao Programa de Pós-Graduação em Modelagem Computacional de Conhecimento da Universidade Federal de Alagoas, em 29 de setembro de 2017, às 10h00min, na sala 15 do Centro de Pesquisas em Matemática Computacional da Ufal.

COMISSÃO JULGADORA

Prof. Dr. Heitor Soares Ramos Filho

Ufal – Instituto de Computação

Orientador

Prof. Dr. André Luiz Lins de Aquino

Ufal – Instituto de Computação

Examinador

Prof. Dr. Ivan César Martins

Ufal – Pesquisador Pós-Doutorado

Examinador

Prof. Dr. Thiago Henrique Silva

UTFPR – Departamento Acadêmico de Informática

Examinador

Maceió, setembro de 2017.

RESUMO

A dinâmica das cidades vem sendo estudada ao longo dos anos, tendo diversas aplicações, como planejamento urbano, estudo de propagação de doenças, sistemas de previsão de tráfego, de recomendações para locais e estudos do comportamento social humano. No entanto, com a evolução tecnológica, em especial, a popularização do uso dos *smartphones* e a massificação do acesso à Internet, é apresentada uma nova oportunidade para se realizar tal tipo de estudo, o uso de dados de mídia social para o estudo populacional. Nesse contexto, este trabalho se propõe a apresentar uma nova maneira de se comparar cidades, utilizando como medida de similaridade o padrão de mobilidade social dos seus habitantes. Como modo de validar o estudo, foram utilizados 1.601.323 *check-ins* do *Foursquare* distribuídos por 10 cidades em um período de 33 dias, a fim de - utilizando as métricas propostas - apresentar diferentes padrões de agrupamento das cidades.

Palavras-chave: Computação Urbana. Redes de Sensores Participativos. Cadeias de Markov. Dinâmica das Cidades.

ABSTRACT

The dynamics of cities have been studied over the years, with various applications such as urban planning, disease propagation studies, traffic forecasting systems, local recommendations and studies of human social behavior. However, with the technological evolution, especially the popularization of the use of smartphones and the massification of Internet access, a new opportunity is presented to carry out such a study, the use of social media data for the population study. In this context, this paper aims to present a new way to compare cities, using as similarity measure the pattern of social mobility of its inhabitants. As a way to validate the study, 1.601.323 *Foursquare check-ins* were used spread over 10 cities in a period of 33 days in order to - using the proposed metrics - present different patterns of grouping the cities.

Keywords: Urban Computing. Participatory Sensing. Markov Chain. Dynamics of Cities.

AGRADECIMENTOS

Agradeço a todos os que me apoiaram durante essa longa jornada, em especial aos meus professores, que foram fonte inestimável de conhecimento, ao meu orientador, que foi fundamental durante esse processo e aos amigos e familiares pelo apoio a mim dispensado.

LISTA DE FIGURAS

1.1	Fluxo de desenvolvimento do trabalho.	11
3.1	Estatísticas do uso de internet pelo mundo.	24
3.2	Árvore de Categorias - <i>Shop & Service</i>	26
4.1	Modelo de uso <i>TWITTER STREAMING API</i>	28
4.2	Histograma da frequência diária de <i>check-ins</i> , de 25-04-2014 à 18-06-2014	29
4.3	Exemplo de trecho de <i>check-in</i> , salientando o link para página no <i>Swarm</i>	30
4.4	Exemplo de código fonte da página do <i>Swarm</i> onde se encontram as categorias.	31
4.5	Grafo de transições entre categorias de um usuário no dia 18/05/2014 (domingo).	34
5.1	Histogramas da cidade de São Paulo apresentando o comportamento ao longo do tempo de 2 categorias em 3 diferentes semanas.	37
5.2	Agrupamento Hierárquico (método Centróide), exemplo de divisão em 3 grupos, usando <i>p-valor</i> do teste Qui-quadrado como parâmetro de agrupamento.	57
5.3	Agrupamento hierárquico (método Centróide), exemplo de divisão em 3 grupos, usando o <i>p-valor</i> do teste para Cadeias de Markov como parâmetro de agrupamento.	58

LISTA DE TABELAS

4.1	Exemplo de dados extraídos de <i>check-ins</i>	32
4.2	Exemplo do padrão de mobilidade social de um usuário no dia 18/05/2014 (domingo).	33
4.3	Matrizes de Frequência da cidade de São Paulo	35
5.1	Representação de valores números por meio de escala de cores de vermelho e azul	38
5.2	Matrizes de escala de vermelho e azul que representam os resultados da comparação entre as Matrizes de Frequência de dias de uma mesma semana de New York	39
5.3	Matrizes de escala de vermelho e azul que representam os resultados da comparação entre as Matrizes de Frequência de dias de uma mesma semana de São Paulo	40
5.4	Matrizes de escala de vermelho e azul que representam os resultados da comparação entre as Matrizes de Frequência de dias de uma mesma semana de outras cidades	41
5.5	Matrizes de escala de vermelho e azul que representam os resultados da comparação entre as Matrizes de Frequência do mesmo dia de diferentes semanas de Los Angeles	42
5.6	Matrizes de escala de vermelho e azul que representam os resultados da comparação entre as Matrizes de Frequência do mesmo dia de diferentes semanas de New York	42
5.7	Matrizes de escala de vermelho e azul que representam os resultados da comparação entre as Matrizes de Frequência do mesmo dia de diferentes semanas de São Paulo	42
5.8	Representação de valores números por meio de escala de cores de cinza	44
5.9	Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de segunda-feira	45
5.10	Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de terça-feira	45
5.11	Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de quarta-feira	45
5.12	Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de quinta-feira	46
5.13	Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de sexta-feira	46
5.14	Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de sábado	46
5.15	Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de domingo	47
5.16	Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com New York em um dia de segunda-feira	48

5.17	Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com New York em um dia de terça-feira	48
5.18	Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com New York em um dia de quarta-feira	48
5.19	Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com New York em um dia de quinta-feira	49
5.20	Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com New York em um dia de sexta-feira	49
5.21	Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com New York em um dia de sábado	49
5.22	Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com New York em um dia de domingo	50
5.23	Matriz de Transição de Probabilidade de New York em um dia de segunda-feira . .	51
5.24	Matriz de escala de cinza que representa os resultados da comparação entre Cadeias de Markov de diferentes cidades com New York	52
5.25	Matriz de escala de cinza que representa os resultados da comparação entre Cadeias de Markov de diferentes cidades com São Paulo	53
5.26	New York - Mais Similares	53
5.27	New York - Menos Similares	53
5.28	São Paulo - Mais Similares	53
5.29	São Paulo - Menos Similares	53
5.30	Matriz de escala de vermelho e azul que representa os resultados da comparação entre Cadeias de Markov de diferentes turnos da da cidade de New York	54
5.31	Matriz de escala de vermelho e azul que representa os resultados da comparação entre Cadeias de Markov de diferentes turnos da da cidade de São Paulo	54
5.32	Matriz de escala de cinza que representa os resultados da comparação entre Cadeias de Markov de diferentes cidades com New York no turno matutino	55
5.33	Matriz de escala de cinza que representa os resultados da comparação entre Cadeias de Markov de diferentes cidades com New York no turno vespertino	55
5.34	Matriz de escala de cinza que representa os resultados da comparação entre Cadeias de Markov de diferentes cidades com New York no turno noturno	55
A.1	Matrizes com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência diferentes dias de uma mesma semana da cidade de New York	61
A.2	Matrizes com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência diferentes dias de uma mesma semana da cidade de São Paulo	62
A.3	Matrizes com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência diferentes dias de uma mesma semana de outras cidades	63
A.4	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência diferentes dias um mesmo dia de diferentes semanas da cidade de Los Angeles	63

A.5	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência diferentes dias um mesmo dia de diferentes semanas da cidade de New York	64
A.6	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência diferentes dias um mesmo dia de diferentes semanas da cidade de São Paulo	64
A.7	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de segunda-feira	64
A.8	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de terça-feira	65
A.9	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de quarta-feira	65
A.10	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de quinta-feira	65
A.11	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de sexta-feira	66
A.12	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de sábado	66
A.13	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de domingo	66
A.14	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com New York em um dia de segunda-feira	67
A.15	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com New York em um dia de terça-feira	67
A.16	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com New York em um dia de quarta-feira	67
A.17	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com New York em um dia de quinta-feira	68
A.18	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com New York em um dia de sexta-feira	68
A.19	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com New York em um dia de sábado	68
A.20	Matriz com os resultados da comparação, <i>p-valores</i> de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com New York em um dia de domingo	69
A.21	Matriz com os resultados da comparação, <i>p-valores</i> de testes de Kullback et al. (1962), entre as Cadeias de Markov de diferentes cidades com New York	69

A.22	Matriz com os resultados da comparação, <i>p-valores</i> de testes de Kullback et al. (1962), entre as Cadeias de Markov de diferentes cidades com São Paulo	69
A.23	Matriz com os resultados da comparação, <i>p-valores</i> de testes de Kullback et al. (1962), entre as Cadeias de Markov de diferentes períodos do dia cidade de New York	70
A.24	Matriz com os resultados da comparação, <i>p-valores</i> de testes de Kullback et al. (1962), entre as Cadeias de Markov de diferentes períodos do dia cidade de São Paulo	70
A.25	Matriz com os resultados da comparação, <i>p-valores</i> de testes de Kullback et al. (1962), entre as Cadeias de Markov de diferentes cidades com New York no turno matutino	70
A.26	Matriz com os resultados da comparação, <i>p-valores</i> de testes de Kullback et al. (1962), entre as Cadeias de Markov de diferentes cidades com New York no turno vespertino	71
A.27	Matriz com os resultados da comparação, <i>p-valores</i> de testes de Kullback et al. (1962), entre as Cadeias de Markov de diferentes cidades com New York no turno noturno	71
A.28	Matrizes de Transição de Probabilidade de New York	72

SUMÁRIO

1	INTRODUÇÃO	9
1.1	Motivação	9
1.2	Objetivo	10
1.3	Metodologia	10
1.4	Contribuições	11
1.5	Organização do Trabalho	13
2	REVISÃO DE LITERATURA	14
2.1	Dinâmica Populacional Humana	14
2.1.1	Ciência Social Computacional	15
2.2	Computação Urbana	15
2.2.1	Redes de Sensores Tradicionais	16
2.2.2	Infraestrutura das Cidades	17
2.3	Redes Sensoriamento Participativos	18
2.3.1	A Dinâmica da Propagação de Doenças	18
2.3.2	Classificação de Regiões Por Pontos de Interesse	18
2.3.3	Formação de Grupos Sociais	19
2.3.4	A Dinâmica Social Urbana	19
3	SENSORIAMENTO PARTICIPATIVO	21
3.1	Características de Sensoriamento Participativo	21
3.2	Redes Sociais Baseadas em Localização	23
3.2.1	Foursquare	25
4	CARACTERIZAÇÃO DOS DADOS	27
4.1	Processamento dos Dados	27
4.1.1	Extração dos Dados	27
4.1.2	Tratamento dos Dados	30
4.1.3	Agrupamento de Dados	32
5	RESULTADOS E ANÁLISES	36
5.1	Comparações entre Matrizes de Frequência de uma Mesma Cidade	36
5.2	Comparações entre Matrizes de Frequência de Diferentes Cidades	43
5.3	Comparações entre Matrizes de Markov de Diferentes Cidades	50
5.3.1	Comparação das Cidades Por Turno	54
5.4	Exemplo de Aplicação	56
6	CONCLUSÕES E TRABALHOS FUTUROS	59
A	Resultados Numéricos dos Testes de Hipótese	61
A.1	Representação numérica do grau de similaridade entre diferentes amostras de uma mesma cidade	61
A.2	Representação numérica do grau de similaridade de amostras de diferentes cidades	64
A.2.1	Representação por meio dos <i>p-valores</i> de testes Qui-quadrado	64
A.2.2	Representação por meio dos <i>p-valores</i> de testes para Cadeias de Markov	69

A.3 Matrizes de Transição de Probabilidade Estimadas	72
Referências Bibliográficas	73

1

INTRODUÇÃO

1.1 Motivação

COM o crescente aumento da concentração da população em ambientes urbanos, as cidades se tornaram ambientes complexos e dinâmicos. Contudo, apesar de oferecer melhorias e facilidades para a condição de vida, as cidades passaram a apresentar muitos problemas, como excesso de tráfego, poluição e violência. Para sanar tais problemas, oferecendo uma melhor alocação de recursos, os administradores passaram a ter a necessidade de entender profundamente a dinâmica das cidades.

A partir daí, o estudo populacional passou a ser uma área de vital importância para a sociedade, no entanto, apesar de ter havido uma enorme evolução nas técnicas estatísticas aplicadas à análise dos dados utilizados nesses tipos de estudos, na grande maioria dos casos, tais dados ainda são captados por meio de entrevistas e formulários respondidos pela população, o que, conforme [Kostakos et al. \(2009\)](#), implica grande consumo de recursos e tempo, dificulta estudos longitudinais e possibilita que os dados sejam influenciados por erro de autorrelato.

Sendo assim, conseguir dados que retratem de maneira fiel a população estudada é, normalmente, o maior desafio desse tipo de pesquisa. Nesse contexto, alavancado pela massificação do uso dos *smartphones* e pelo aumento da cobertura do sinal de internet móvel, uma promissora fonte de dados tem surgido como alternativa para realizar tais estudos: as *Redes de Sensores Participativos (RSP)*.

Sensoriamento Participativo é um tipo de sensoriamento onde pessoas agem como sensores, transmitindo, de forma voluntária, suas sensações sobre o ambiente em que se encontram. Quando um grupo de indivíduos trabalha em conjunto forma-se uma *RSP*, onde cada um passa a atuar como um nó móvel da rede, gerando informações que, uma vez reunidas, formam uma base de conhecimento sobre o fenômeno que está sendo observado.

Analisando a definição acima, é possível enxergar as *Redes Sociais Baseadas em Localização* (RSBL) (*Instagram, Foursquare e Waze*) como exemplos de RSPs, pois funcionam por meio de informações fornecidas voluntariamente pelos usuários. De fato, segundo (Burke et al., 2006), as RSBL são os exemplos mais populares de RSPs.

Com o uso cada vez maior de mídias sociais, as RSPs têm se tornado uma alternativa bastante viável ao estudo populacional e apresenta grandes vantagens – que serão discutidas com mais detalhes no capítulo 3.1 – em relação às fontes tradicionais de dados.

Em especial, por trazer uma grande variedade, periodicidade e escalabilidade, o uso de dados de RSPs, revolucionou os estudos do comportamento populacional, pois proporcionou uma nova maneira de analisar a população, por meio da dinâmica social urbana.

Enquanto muitos trabalhos tradicionais – realizados por meio de formulários de pesquisa aplicados à população – normalmente realizam estudos a partir do agrupamento da população utilizando como parâmetros o comportamento socioeconômico (Tammaru et al., 2016), religioso (Kosmin & Lachman, 2011) e étnico (Waters, 1994), essa nova vertente de estudo faz uso dos dados sensoriais cedidos continuamente pela população, mediante SSPs, para explorar a dinâmica social urbana ao longo do tempo, tentando, assim, entender/explicar a evolução do padrão de comportamento coletivo da população nas cidades.

1.2 Objetivo

O presente trabalho se propõe a, por meio do uso de dados provenientes de uma RSP, mais especificamente o *Foursquare*, estudar a dinâmica social urbana para oferecer uma nova metodologia para agrupamento das cidades, que leve em consideração o grau de similaridade das atividades exercidas cotidianamente pela população. Para atingir tal objetivo, o trabalho responderá os seguintes questionamentos:

1. Como agrupar os dados populacionais por padrões de atividades sociais?
2. O comportamento da população estudada segue rotinas diárias/semanais?
3. Como comparar cidades por características sociais específicas (como atividades ligadas à educação ou ao trabalho)?
4. Como utilizar as principais atividades sociais da população para criar um modelo capaz de comparar o padrão de mobilidade social das cidades?

1.3 Metodologia

A metodologia adotada para o desenvolvimento do presente trabalho pode ser expressa pela imagem 1.1.

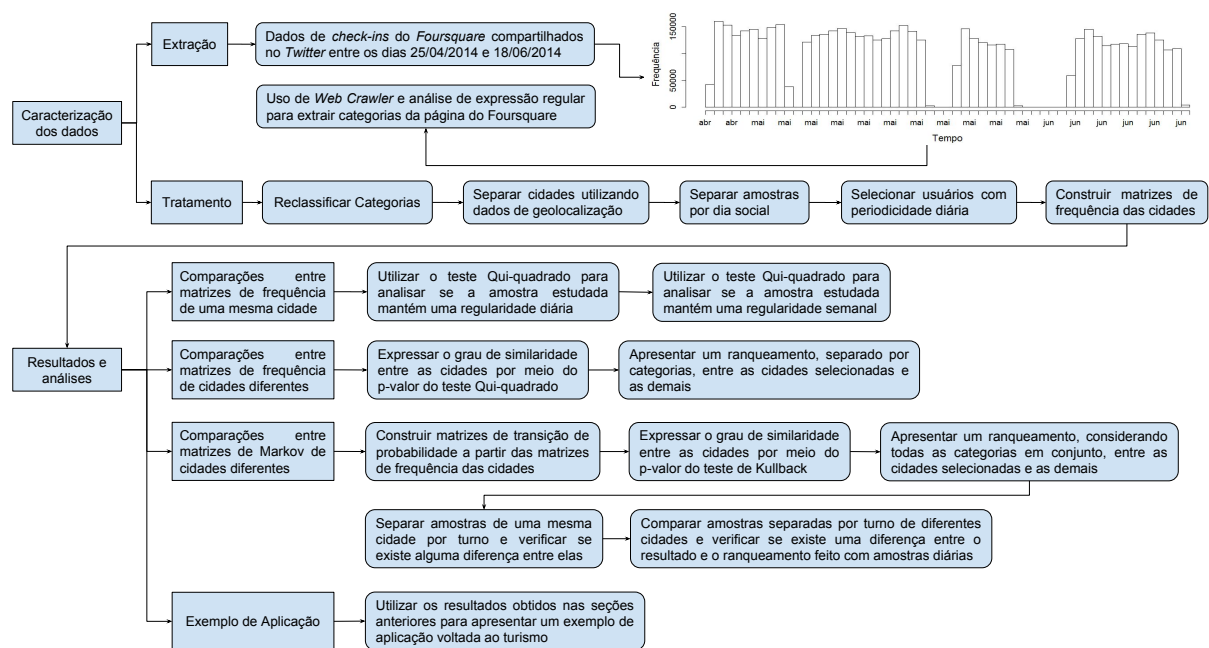


Figura 1.1: Fluxo de desenvolvimento do trabalho.

Como é possível ver na imagem acima, a primeira etapa do trabalho foi caracterizar os dados que foram utilizados na pesquisa. Para isso, foram extraídos dados de *check-ins* de usuários de uma *RSP* em diversas cidades e, em seguida, foram aplicadas todas as transformações necessárias para que os dados ficassem aptos a serem utilizados. Com os dados no formato adequado para uso, foram realizadas várias comparações. Primeiramente investigou-se a possível associação dos dados de uma mesma cidade em dias diferentes; na sequência, comparou-se dados de diferentes cidades para descobrir a verossimilhança entre elas, separando-as por características específicas e, depois, levando em consideração todas as características em conjunto. Todo esse processo foi efetuado tanto nas amostras diárias quanto nas amostras separadas por turno. Por fim, foi apresentado um exemplo de aplicação que faz uso das técnicas de comparação apresentadas para auxiliar turistas na escolha de cidades para visitar, levando em consideração experiências anteriores em outras cidades.

É importante salientar que os resultados apresentados se referem à população estudada, ou seja, os usuários do *Foursquare*. Qualquer generalização feita em relação à real população das cidades deve ser encarada apenas como mera suposição.

1.4 Contribuições

As principais contribuições ligadas aos resultados obtidos por este trabalho podem ser resumidas conforme explicitado a seguir.

Foi mostrado que, apesar de um moderado grau de similaridade, estatisticamente não é possível afirmar que a população de uma dada cidade em dois dias de uma mesma semana

seguem rotinas sociais semelhantes. Ou seja, o padrão de mobilidade urbano varia de acordo com o dia da semana analisado. Tal resultado fica bastante claro ao se analisar os exemplos apresentados das cidades de São Paulo e de New York.

Além disso, foi demonstrado que, apesar de não ser possível agrupar os dados por rotinas diárias, existe uma certa semelhança no comportamento da população em algumas classes de atividades sociais, sendo possível separá-las em três diferentes grupos de acordo com o padrão de oscilação dos dados.

Ainda por meio de análise estatística, foi apresentado que as cidades estudadas possuem um grau relativamente elevado de similaridade entre as amostras semanais, ou seja, existe semelhança no comportamento da social da população no mesmo dia da semana de diferentes semanas, por exemplo, segunda-feira de uma semana com segunda-feira de outra semana.

Foi proposta também uma metodologia de comparação entre cidades que tem como métrica as características sociais específicas de cada local e, utilizando os resultados obtidos, criado um agrupamento de cidades que leva em consideração tais características. Tal agrupamento apresentou resultados interessantes em relação às cidades estudadas, por exemplo, para atividades ligadas à educação, as que tiveram maior grau de similaridade com São Paulo foram Moscow, Rio de Janeiro e Recife.

Por fim, foi gerada uma representação das cidades por meio do uso de *Cadeias de Markov*, capaz de retratar as principais características sociais em conjunto - e não mais separadamente, como realizado na abordagem anterior - a fim de que, utilizando novamente uma abordagem estatística, fosse criada uma comparação e, posteriormente, um agrupamento de cidades a partir de suas dinâmicas sociais.

Além das contribuições obtidas por meio dos resultados, o presente trabalho tem como frutos indiretos e diretos, respectivamente, as publicações:

- TAVARES, M.; CARDOSO, I.; LOBO, V. T. C.; FERREIRA, T.; RAMOS, H. S.. “Avaliação e Paralelização de Algoritmos de Mapeamento para Grafos de Visibilidade”. Em: *Escola Regional de Alto Desempenho (ERAD) - Região Nordeste*. Maceió, Brasil. 2015.
- LOBO, V. T. C.; RAMOS, H. S.; SILVA, T. H.. “Um Padrão de Agrupamento de Cidades a Partir da Dinâmica Social Urbana Extraída de Dados Provenientes de Redes de Sensores Participativos”. Em: *XXXV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos - I Workshop de Computação Urbana (SBRC - CoUrb)*. Belém, Brasil. 2017.

Essa última, por se destacar entre as melhores publicações apresentadas no *workshop*, foi convidada à submissão de uma versão estendida na revista *IEEE Latin America Transactions*, a qual, até o presente momento, está em fase de análise.

1.5 Organização do Trabalho

Na ordem do estudo, são apresentadas, na seção *Revisão de Literatura* (Capítulo 2), publicações relevantes nas áreas nas quais o presente trabalho se insere; na seção *Sensoriamento Participativo* (Capítulo 3), a lista de conceitos necessários para entender esse tipo de estudo; na seção *Caracterização dos Dados* (Capítulo 4), a explicação de todas as etapas necessárias à adequação dos dados utilizados e uma posterior análise dos mesmos; na seção *Análises e Resultados* (Capítulo 5), os diferentes métodos de análise de dados aplicados e os resultados obtidos, e, por fim, na seção *Conclusões e Trabalhos Futuros* (Capítulo 6), uma breve explanação final sobre todo o trabalho e a apresentação dos próximos passos desta pesquisa.

2

REVISÃO DE LITERATURA

ESTE capítulo é organizado da seguinte maneira: na seção 2.1 é feita uma breve iniciação ao estudo da dinâmica populacional urbana, salientando a importância dos dados nesse tipo de estudo e apresentando uma promissora nova área de pesquisa. Já na seção 2.2, são explanados conceitos ligados à computação urbana e o que essa área tem a oferecer em relação ao estudo da dinâmica populacional, apresentando diferentes fontes de dados utilizadas em estudos dessa natureza e sempre citando trabalhos impactantes em cada uma delas. Por fim, na seção 2.3, são discutidos trabalhos promissores que fazem uso de dados de *RSP*, destacando o estudo da dinâmica social urbana, área na qual esse trabalho se insere.

2.1 Dinâmica Populacional Humana

O estudo da dinâmica populacional humana é uma subárea das ciências sociais que tem como objetivo entender/explicar como a população se movimenta ao longo do tempo. Esse tipo de pesquisa tem diversas aplicações, como planejamento urbano, estudo de propagação de doenças, sistemas de previsão de tráfego, recomendações para locais e estudos do comportamento social humano.

Como toda pesquisa, os estudos populacionais também seguem o método científico, tendo como premissa inicial imaginar/construir uma hipótese sobre algum aspecto que deseja ser observado na população; relacioná-la com teorias e modelos para tornar tal hipótese operacional; coletar os dados necessários para validar a hipótese proposta. Contudo, dentro do contexto de estudos demográficos, conseguir dados que retratem de maneira fiel a dinâmica da população é, normalmente, o maior desafio dessa área de pesquisa.

Com a evolução da tecnologia, no entanto, a população passou a viver conectada em rede. Atividades corriqueiras - como fazer uma ligação, realizar compras com cartão de crédito, interagir com amigos por meio de redes sociais e até mesmo marcar uma consulta médica - deixam traços digitais que, segundo (Lazer et al., 2009), podem ser compilados em imagens

abrangentes de comportamento individual e em grupo da população, tendo o potencial de transformar a compreensão da vida humana, suas organizações e sociedades.

2.1.1 Ciência Social Computacional

Nesse contexto, uma nova área de pesquisa tem ganhado mais importância no cenário mundial, a Ciência Social Computacional (CSC). Essa, segundo [Amaral \(2017\)](#), é um campo científico interdisciplinar e emergente que usa métodos computacionais para analisar e modelar fenômenos sociais, estruturas sociais e comportamento coletivo.

Enquanto estudos sociais tradicionais se baseiam principalmente em dados de auto-relatos sobre comportamento social ([Lazer et al., 2009](#)). A grande vantagem da CSC é a capacidade de reunir e analisar enormes quantidades de dados de diferentes fontes, como, diferentes mídias sociais.

Esses dados, como dito anteriormente, são gerados sempre que usuários interagem com a rede, e, por meio dessa interação, passam a registrar seus traços na mesma, como uma marca digital que é deixada, e, quando estudada ao longo do tempo, pode ser utilizada para analisar diversos aspectos sociais do comportamento dos usuários, por exemplo, como se movem ([Zhao et al., 2015](#); [Gonzalez et al., 2008](#)) e com que finalidade ([Cho et al., 2011](#)); o que comem ([Colombo et al., 2012](#); [Silva et al., 2014a](#)) e como interagem em grupo ([Eagle et al., 2009](#); [Nunes et al., 2016](#)). Quando a análise é feita em larga escala, é possível investigar as atividades predominantes em uma dada região ([Frias-Martinez et al., 2012a](#); [Cranshaw et al., 2012](#)) e, até mesmo, entender como funciona a dinâmica populacional de uma cidade ([Yuan et al., 2012](#); [Cranshaw et al., 2012](#); [Silva et al., 2014b](#)).

2.2 Computação Urbana

Uma característica em comum na maioria dos estudos de CSC é que são realizados em ambientes urbanos, pois, devido à natureza dos dados utilizados, é necessária uma infra-estrutura mínima para sua existência, por exemplo, torres de distribuição de sinais de redes móveis, câmeras de monitoramento de tráfego ou sensores que meçam a emissão de poluentes. Por conta disso, mesmo com o avanço da propagação do sinal de acesso à internet em ambientes rurais, as pesquisas ainda se concentram principalmente em ambientes urbanos, já que há uma maior abrangência de fontes de dados e assim garantir uma maior confiabilidade nos resultados.

Sendo assim, é possível verificar que grande parte dos trabalhos de CSC também estão inseridos na área de Computação Urbana, a qual, segundo ([Zheng et al., 2014](#)), pode ser definida como um processo de aquisição, integração e análise de um grande volume de dados

heterogêneos gerados por diversas fontes em espaços urbanos, tais como sensores, veículos e seres humanos, para ajudar na solução de diversos problemas que as cidades enfrentam, tais como congestionamento de trânsito, poluição do ar, falta de água e aumento do consumo de energia.

Mais especificamente, tratando-se de dados em ambientes urbanos, podemos afirmar, segundo [Silva & Loureiro \(2015\)](#), que existem quatro principais fontes de dados urbanos: (1) Dados estatísticos oficiais; (2) Redes de sensores tradicionais; (3) Infraestrutura das cidades e (4) Redes de sensoriamento participativos.

No entanto, no que se refere à mobilidade urbana, dados estatísticos oficiais, em geral, não conseguem caracterizar de forma adequada a dinâmica inerente da população estudada, já que normalmente não são fruto de estudos contínuos da população ao longo do tempo.

As próximas seções deste capítulo têm como objetivo apresentar trabalhos que obtiveram resultados relevantes em diferentes aplicações dentro da área do estudo da dinâmica urbana, discutindo quais fontes de dados foram utilizadas para a realização dos experimentos e salientando os resultados obtidos em relação à quantidade de dados usados, a escalabilidade do experimento, o tempo para ser realizado e, quando possível, o custo de implementação.

2.2.1 Redes de Sensores Tradicionais

O uso de Redes de Sensores tradicionais para o estudo da mobilidade urbana tem como principal vantagem a exatidão dos resultados obtidos, pois, em geral, os sensores são desenvolvidos com uma finalidade específica.

No entanto, dentre as principais desvantagens desse tipo de abordagem se destacam o custo significativo de construção das redes de sensoriamento, principalmente quando feito em larga escala, e a dificuldade de acesso aos dados, já que normalmente a implantação de sensores em ambientes urbanos, por conta de questões de segurança e/ou burocráticas, são permitidas apenas a órgãos governamentais, tornando muito difícil o acesso aos mesmos.

Dentre as principais aplicações na área de mobilidade urbana que fazem uso desse tipo de fonte de dados, sem dúvida, se destaca a análise da dinâmica do tráfego nas cidades, trabalhos como [\(Coifman, 2006\)](#) e [\(Ma et al., 2015\)](#), por exemplo, estudam previsão do trânsito por meio de dados de sensores *Microwave*.

Em especial, destacam-se trabalhos de autores como [Xie & Wang \(2014\)](#), que utilizam dados de *RSP* juntamente com dados de sensores físicos para obter uma maior precisão nos resultados, associando, para isso, 125 milhões de dados veiculares capturados por sensores de tráfego SURTRAC - é uma abordagem inovadora do controle de sinal de trânsito em tempo real, que combina pesquisas em inteligência artificial e teoria do trânsito - com 3.4 milhões de postagens, provenientes do *Twitter*, *Foursquare*, *Flickr*, *Picasa* e *Panoramio*, com intuito de estudar possíveis soluções para a melhoria da mobilidade urbana.

2.2.2 Infraestrutura das Cidades

A exploração de dados da infraestrutura das cidades é uma alternativa inteligente ao uso dos recursos já existentes nas cidades, uma vez que, diferentemente do exemplo anterior, se utiliza de uma infraestrutura já existente, mas que foi criada com outros propósitos, como redes de telefonia móvel e *GPS*.

Talvez uma das maiores contribuições dos últimos tempos no estudo da dinâmica populacional tenha sido alcançada com o uso dessa fonte de dados para criação de modelos de representação da dinâmica humana ao longo do tempo. Anteriormente realizados por meio de simulações computacionais do comportamento humano, esses modelos não levavam em consideração dados reais para serem construídos. O uso de dados da infraestrutura da cidade, em especial, de dados de registro da rede de telefonia móvel, ajudaram a representar de maneira mais fidedigna tal comportamento.

É possível ver exemplos disso em trabalhos como (Song et al., 2010), (Becker et al., 2013) e, especialmente, em Gonzalez et al. (2008), que estuda o padrão de mobilidade dos indivíduos a partir de duas bases de dados, a primeira, contém seis meses de registros telefônicos (ligações e mensagens de texto) de 100.000 indivíduos selecionados aleatoriamente de uma amostra de mais de 6 milhões de usuários anônimos, a segunda captura a localização de 206 aparelhos a cada duas horas em um período de uma semana. Analisando a dinâmica dos registros, os autores afirmam que a mobilidade dos usuários pode ser caracterizada como uma Caminhada de Lévy e aproximada por uma Lei de Potência Truncada. Utilizando essa técnica de análise, apresentam resultados bastante relevantes, como a existência de uma forte tendência das pessoas retornarem a locais visitados anteriormente - resultado coerente com o de Song et al. (2010) que encontrou um alto grau de previsibilidade em sua base de usuários - mostrando, assim, a recorrência e a periodicidade temporal inerentes à mobilidade humana.

Apesar dos bons resultados apresentados, a modelagem por meio de dados de registros telefônicos apresenta um problema, por calcular as localizações por meio das torres que proveem o sinal ao aparelho, não conseguindo capturar a movimentação do usuário entre locais próximos, logo, movimentações de curto espaço não podem ser consideradas nesse tipo de abordagem. Tal problema pode ser verificado no trabalho de Gonzalez et al. (2008), por exemplo, onde a cobertura máxima das torres varia entre 1km^2 e 3km^2 .

Esse problema pode ser contornado com o uso de dados de *GPS*, que oferece, normalmente, uma maior precisão em relação à localização. No entanto, a maioria dos trabalhos que utilizam tal fonte de dados, o fazem por meio de dados de mídia social, ou seja, dados de *RSP's*, abordagem que será discutida em detalhes na próxima seção.

No entanto, é possível citar alguns trabalhos, como (Liu et al., 2012; Castro et al., 2013), que fazem uso de dados de *GPS* veiculares. Em especial, tais trabalhos se destacam por utilizar dados captados a partir de *GPS* de táxis, os quais, por se manterem em constante movimento ao longo do dia, com diferentes passageiros, que têm diferentes pontos de partida e chegada,

representam uma significativa fonte de dados para o estudo do padrão de mobilidade urbana.

2.3 Redes Sensoriamento Participativos

O uso de dados de *RSP*, por ter várias vantagens em relação às demais fontes de dados, é, sem dúvida, uma das promissoras alternativas para o estudo da dinâmica populacional. Dentre essas vantagens, é possível destacar: um menor custo em relação à 2.2.1; maior facilidade de acesso aos dados do que em relação à apresentada na seção 2.2.2; e maior abrangência em relação a apresentada por ambas.

As peculiaridades do uso de dados de *RSP* serão discutidas com mais detalhes no próximo capítulo, sendo o objetivo dessa seção apenas apresentar os principais trabalhos que fazem uso dessa fonte de dados para o estudo da dinâmica populacional e suas diferentes aplicações.

2.3.1 A Dinâmica da Propagação de Doenças

Inúmeros trabalhos visam estudar diferentes aspectos da dinâmica humana. Autores como (Achrekar et al., 2011; Sadilek et al., 2012) se valem do fato de os mecanismos de controle epidemiológico estarem diretamente ligados a como as pessoas se movimentam ao longo do tempo para analisar a propagação de doenças por meio de dados de *RSP*.

Gomide et al. (2011), por exemplo, propõem uma ferramenta de monitoramento da dengue usando como base dados postagens do *Twitter* realizadas no Brasil e distribuídos entre 2009 e 2011. Já Lampos & Cristianini (2010), estudam a propagação de gripe também com dados de *Twitter*, mas no Reino Unido, em um período de 24 semanas, durante um surto de *H1N1*, atingindo resultados bastante expressivos com uma alta correlação linear entre os seus dados e os da Agência de Proteção de Saúde.

2.3.2 Classificação de Regiões Por Pontos de Interesse

Alguns trabalhos, como (Noulas et al., 2011b; Cranshaw et al., 2012; Yuan et al., 2012; Hu et al., 2015), visam pesquisar como utilizar dados de *RSP* para inferir as atividades existentes em determinadas regiões das cidades. Em especial, Cranshaw et al. (2012) estuda as características das cidades as dividindo em *Livehoods* – uma nova forma de dividir a cidade pelas atividades predominantes em cada região – para isso, o autor analisa 18 milhões de postagens feitas no *Foursquare* e valida seus resultados comparando os grupos formados na cidade de Pittsburgh com a opinião de 27 moradores da cidade.

Já Frias-Martinez et al. (2012b), tenta caracterizar ambientes urbanos por meio de dados extraídos do *Twitter* em um período de 49 dias. Para separar a ilha de Manhattan em áreas de acordo com suas atividades, os autores utilizam técnicas como *Self-Organizing Map* e *Diagrama de Voronoi*, construindo, assim, uma representação da distribuição dos dados como

um mapa. Por fim, criam vetores que representam o comportamento médio dos *tweets* de cada segmento de terra e aplicam *K-means* para agrupar as áreas que têm vetores parecidos. É feita análise de cada cluster formado para, então, ser comparado com resultados do *NYC Department of City Planning* e do *NYC Department of Parks & Recreation*.

2.3.3 Formação de Grupos Sociais

Outros autores, como [Cho et al. \(2011\)](#); [Brown et al. \(2013\)](#); [Gao et al. \(2013\)](#); [Pham et al. \(2016\)](#); [Nunes et al. \(2016\)](#), propõem-se a investigar a formação de grupos sociais a partir do estudo da dinâmica populacional. Dentre esses, destacam-se [Nunes et al. \(2016\)](#), que utilizam dados de *MIT Reality Mining* ([Eagle & Pentland, 2006](#)) para estudar o padrão de mobilidade social da população em questão, tendo como foco do trabalho, mais especificamente, entender o impacto de grupos sociais nesse padrão de mobilidade. Para isso, os autores estudam a dinâmica de evolução dos grupos ao longo do tempo, levando em consideração os seguintes aspectos: (i) tamanho; (ii) taxas de mudança de estrutura de crescimento, redução, nascimento e morte; (iii) periodicidade da reunião de grupo e (iv) duração das reuniões do grupo e sua correlação com a força dos vínculos do grupo e a estabilidade da composição. Apesar de utilizar uma base de dados com poucos usuários (80 usuários que residem em dois edifícios universitários diferentes), obtém resultados interessantes, como o fato de que os contatos do grupo apresentam não só periodicidade diária, mas também uma fração de periodicidade semanal.

2.3.4 A Dinâmica Social Urbana

Essa classe de trabalho se diferencia dos estudos tradicionais da dinâmica urbana por não ter como objetivo principal o estudo da dinâmica espacial da população, sendo assim, os dados populacionais não são analisados em relação à distância dos deslocamentos e sim em relação aos tipos de locais frequentados.

Nesse contexto, trabalhos como os de ([Noulas et al., 2011a](#); [Silva et al., 2012, 2014a,b](#); [Manca et al., 2017](#)) visam estudar o comportamento social das cidades por meio de dados de *RPS*. Dentre esses, é possível destacar ([Silva et al., 2014a](#)), que usa observações capturadas do *Twitter* com a *tag* do *Foursquare*, contendo as categorias *Drink*, *Fast-Food* e *Slow-Food* para estudar a preferência dos usuários em relação a esses três aspectos em diferentes locais do mundo. Para validar os resultados obtidos, os autores comparam seus resultados com um *survey* ([Inglehart & Welzel, 2010](#)) que estudou aspectos religiosos, políticos, econômicos e de estilo de vida da população de alguns países no período de 2005 à 2008.

O presente trabalho também objetiva estudar a dinâmica social das cidades por meio de dados de *RSP*. Difere, contudo, de outros como ([Silva et al., 2012](#)), ([Silva et al., 2014a](#)) e ([Silva et al., 2014b](#)), por oferecer uma investigação mais aprofundada das particularidades dos dados.

Especificamente, apresenta evidências estatísticas de que o comportamento social dos dias da semana e, até mesmo, dos turnos do dia, diferem entre si, enquanto os trabalhos citados não fazem tais distinções. Além disso, realiza um estudo abrangente, investigando mais vertentes da dinâmica urbana (estudando cada característica do comportamento social da população separadamente) e propõe um modelo matemático mais consistente para representação das características da população em conjunto, retratando os dados mediante o uso de *Cadeias de Markov* e utilizando testes estatísticos específicos para esse tipo de modelagem, a fim de comparar diferentes cidades.

Os trabalhos citados, calculam a semelhança entre as cidades por meio de cálculo de distância Euclidiana (Silva et al., 2014b) ou por meio de correlações de Spearman (Silva et al., 2014a), o que implica alguns problemas, uma vez que ambas as abordagens calculam a distância entre duas cidades linearizando dados que são originalmente multidimensionais, perdendo, assim, informações passíveis de exploração e gerando eventuais falsas associações. Além disso, as métricas utilizadas para cálculo de distância não são específicas para o tipo de dados utilizado, sendo a primeira um simples cálculo de distância, não conseguindo identificar outros tipos de associações entre os dados que não seja linear. Já a segunda, apesar de ser melhor que a primeira, avalia apenas associações entre os dados que tenham relações monotônicas.

A abordagem aqui empregada, por sua vez, modela os dados por meio de Cadeias de Markov, representando as probabilidades de deslocamento dos habitantes de uma cidade entre todos os tipos de locais (estados da cadeia) como uma Matriz de Transição de Probabilidades, conseguindo representar a interação de todos os estados de maneira conjunta. Ademais, para comparar as cidades é utilizado um teste específico para Cadeias de Markov, o qual permite calcular a verossimilhança entre elas como um todo e a expressar o resultado por meio de uma única variável, o *p-valor*.

SENSORIAMENTO PARTICIPATIVO

NESTA seção são discutidos conceitos necessários para se entender o que é *SP*, peculiaridades como abrangência de dados e confiabilidade, além das vantagens e desvantagens de sua utilização. É apresentada, ainda, uma classificação de sistemas que usam dados de *SP*, dando ênfase aos que oferecem serviços baseados em localização, em especial o *Foursquare*.

3.1 Características de Sensoriamento Participativo

O *Sensoriamento Participativo* pode ser definido, segundo [Burke et al. \(2006\)](#), como um processo distribuído de coleta de dados pessoais e sobre diversos aspectos das cidades. Tal processo requer a participação ativa das pessoas para compartilhar voluntariamente informação contextual e/ou tornar seus dados sensoriados disponíveis.

Apesar de esse ser um conceito antigo e bastante amplo, com o surgimento das mídias sociais passou a ter bem mais importância, pois essas se utilizam de dados sensoriais de seus usuários para seu funcionamento e podem ser conceituadas como *RSPs*.

As *RSPs*, impulsionados pela massificação do uso dos smartphones e de outros dispositivos conectados em rede, se tornaram muito populares nesta década, notadamente por mesclarem aspectos de redes sociais e mecânicas de jogo em seu funcionamento.

Tais sistemas geralmente se baseiam e refletem as relações sociais da vida real por meio de plataformas online, como um site, nas quais os usuários podem compartilhar ideias, atividades, eventos e interesses por meio da internet ([Zheng et al., 2011](#)).

Um reflexo direto das pessoas passarem mais tempo conectadas à Internet é o aumento do uso de *RSPs*, a exemplo do *Instagram*, que, conforme estatística fornecida pelo próprio aplicativo, teve um crescimento no número de usuários, chegando a marca de 500 milhões, que compartilham mais de 95 milhões de fotos por dia.

Segundo matéria do *New York Times* (Stewart, 2016), os usuários passam, em média, 50 minutos por dia conectados aos *RSPs*, isso considerando apenas as plataformas *Facebook*, *Messenger* e *Instagram*, sem falar nas ferramentas, como *Youtube*, *Qzone*, *WhatsApp*, *FourSquare* e *Google+*. Esse valor pode até não parecer elevado, mas supera quase todas as outras atividades de lazer, se aproximando, até mesmo, do tempo gasto em necessidades básicas como comer e beber, que somam 1,07 horas diárias.

Um outro fator importante é a variedade e a qualidade das informações compartilhadas pelos usuários. Atualmente, os dispositivos móveis acoplam diversos tipos de sensores e de dispositivos úteis em tarefas como *GPS*, acelerômetro, giroscópio, câmera e microfone. Ademais, algumas *RSPs* também utilizam sinais da rede *WI-FI* e de redes móveis para melhorarem a precisão da estimativa de localização do dispositivo.

Os dados de sensoriais, por sua vez, são obtidos de duas formas: (i) automática (Sensoriamento Oportunista), quando o sistema os recolhem sem a necessidade de auxílio do usuário, como, por exemplo, dados de *GPS* compartilhados pelo aplicativo *Waze*, que os utiliza para analisar o tráfego nas regiões próximas ao usuário; (ii) e com auxílio humano (Sensoriamento Participativo), quando o usuário utiliza o *RSP* para compartilhar informações do seu interesse, como ideias ou informações sensoriais do ambiente em que se encontra.

Existem vantagens e desvantagens na comparação de as ambas abordagens. No caso do Sensoriamento Oportunista, a qualidade das informações extraídas automaticamente pelas plataformas, via de regra, são de um grau elevado de exatidão e possuem maior periodicidade, pois precisam apenas que o dispositivo esteja ligado e conectado à rede para sua captação, não tendo a necessidade da interação direta com o usuário.

Enquanto isso, as informações sensoriais compartilhadas por intermédio dos usuários, ou seja, Sensoriamento Participativo, apesar de serem mais passíveis ao erro – principalmente por conta da má utilização da ferramenta, como quando um usuário partilha uma foto qualquer afirmando ser do local em que se encontra – oferecem uma visão bem mais abrangente do ambiente, uma vez que utilizam sensações captadas pelos usuários por meio dos seus sentidos para caracterizar o local em que se encontram, como o que estão vendo, ouvindo, tocando e até mesmo o gosto que estão sentindo. Tais sensações, mesmo com a evolução de técnicas de Inteligência Artificial, dificilmente podem ser obtidas com a mesma riqueza de detalhes sem a interação humana e, por conta disso, se tornam informações inestimáveis para o estudo do comportamento humano.

Além disso, um outro fator a favor do uso de dados de *SP* é que os eventuais erros de compartilhamento de informações têm diminuído proporcionalmente, pois, visando contornar tais problemas, os *RSPs* passaram a implementar várias medidas de proteção, como algoritmos de verificações de conteúdo (como análise de imagens e textual), além de ranqueamento de usuários e políticas de uso restritivas, que provocam, em alguns casos, exclusão de conteúdo, e, em casos extremos, banimento de contas que não se adequam à política de uso da ferramenta.

No entanto, mesmo com as melhorias citadas, por conta do fator humano envolvido, o trabalho com dados de *RSP* ainda é uma tarefa desafiadora, pois, ao contrário de outras fontes de dados, além da possibilidade de erro na informação disponibilizada, também não há garantia da periodicidade da mesma, principalmente quando necessita da intervenção direta do usuário para ser compartilhada, já que não há obrigatoriedade por parte do usuário de fornecer a informação, nem mesmo de manter o dispositivo operante. Tais características devem ser sempre levadas em consideração na construção dos experimentos que utilizem esse tipo de fonte de dados.

Contudo, apesar dos problemas apresentados, o uso de dados de *RSP's* ainda possui grandes vantagens como o custo e a abrangência da informação, principalmente se comparadas com as Redes de Sensores Tradicionais, onde existe um alto valor envolvido na implantação da rede para o sensoriamento em larga escala. Tal valor está ligado não somente à compra de sensores, mas também à própria implementação e manutenção da rede. Já ao se utilizar dados de *RSP's* é possível o acesso a um grande volume de dados tendo como único custo a captação dos mesmos, uma vez que os dispositivos da rede funcionam de maneira autônoma e não apresentam problemas comuns em redes sem fio tradicionais, como vida útil do dispositivo e gasto de energia.

Tratando-se de abrangência da informação, é facilmente perceptível a tendência de crescimento desse tipo de fonte de dados. Atualmente, o uso de *smartphones* se estende a todas as classes sociais e as operadoras oferecem uma cobertura de rede cada vez maior e de melhor qualidade para os dispositivos. Segundo a [ICT Data and Statistics Division \(2016\)](#), sete bilhões de pessoas (95% da população mundial) vivem em áreas cobertas por uma rede móvel-celular. Quando se trata especificamente de redes móveis de banda larga (3G ou superior) os números chegam a uma cobertura de 84% da população mundial, com 67% de cobertura em áreas rurais.

Infelizmente, apesar de uma grande cobertura de rede, ainda segundo a [ICT Data and Statistics Division \(2016\)](#), 3.9 bilhões de pessoas ainda não possuem acesso à internet (53% da população mundial). Além disso, como se pode ver na Imagem 3.1, os números globais escondem uma grande disparidade, pois, enquanto no continente africano 75% da população não faz uso da internet, no europeu esse número é de apenas 21%. Como veremos posteriormente, tais números podem dificultar e, em alguns casos, até mesmo impossibilitar o uso de dados de *RSP* de algumas localidades em pesquisas que tenham como objetivo realizar estudos de larga escala populacional.

3.2 Redes Sociais Baseadas em Localização

Após discutir detalhes e peculiaridades de se trabalhar com dados de *SP*, apresentam-se agora os diferentes tipos de *RSPs*, separando-os pelas atividades que oferecem em suas

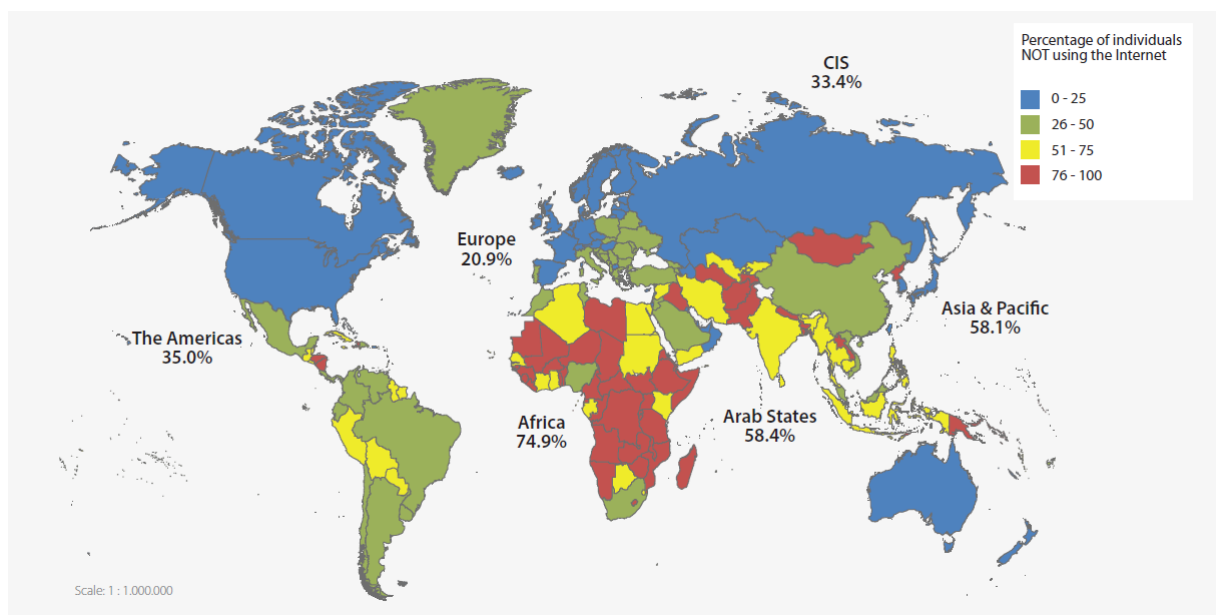


Figura 3.1: Estatísticas do uso de internet pelo mundo.

plataformas. Vale salientar que nessa separação muitos *RSPs* apareceram em mais de um grupo, pois, devido ao grande interesse dos usuários e com a evolução dos dispositivos móveis, que acomplam um número cada vez maior de tecnologias, passou a existir uma maior diversificação dos serviços por eles oferecidos. Atualmente, podemos agrupar as principais *RSPs* pelos seguintes serviços:

- Serviço de Compartilhamento de Localização;
 - Exemplo: Foursquare e Facebook (Places).
- Serviço de Compartilhamento de Fotos;
 - Exemplo: Instagram e Flickr.
- Serviço de Compartilhamento de Vídeos;
 - Exemplo: Instagram e Facebook.
- Serviço de Alerta de Tráfego;
 - Exemplo: Waze.

Em especial, para este trabalho, os olhares serão voltados para os *RSPs* que oferecem serviços de compartilhamento de localização, em especial o *Foursquare*. Esses sistemas, por misturarem aspectos de redes sociais em seu funcionamento, são também conhecidos como *Redes Sociais Baseadas em Localização (RSBL)*.

Segundo Noulas et al. (2011b), *RSBL* são serviços construídos em cima da noção de aproximar locais que visitamos com os amigos aos quais estamos conectados. Devido a sua popularidade crescente, eles oferecem uma base de dados de atividade humana promissora.

Todavia, é importante perceber que dados de uma *RSBL* não representam apenas o compartilhamento de informações de vários usuários associado às suas respectivas localizações, mas também uma nova forma de enxergar o comportamento social, pois os usuários passam a deixar seu rastro de atividades diárias nos sistemas, permitindo a representação do seu comportamento social por meio do padrão de compartilhamento de informação.

À medida que os *RSBLs* passam a acumular um grande volume de dados, capturados a partir do compartilhamento de atividades diárias de uma vasta base de usuários ativos, é possível, a partir de deles, derivar uma *RSP* e, por meio dessa rede, inferir não apenas o comportamento social dos indivíduos, isoladamente considerados, mas também a dinâmica social de toda uma região, cidade ou até mesmo um país.

3.2.1 Foursquare

O *Foursquare* é uma *RSBL* lançada em 11 de março de 2009 e sua plataforma é dividida em duas aplicações, o *Foursquare* e o *Swarm*, sendo acessível via navegador e possuindo versões de aplicativo para as principais plataformas móveis (*iOS*, *Android*, *Windows Phone*, *Blackberry* e *Symbian*). A divisão da plataforma se dá pelo foco de cada aplicação, enquanto o *Swarm* é voltado para *check-ins* e interação entre usuários, o *Foursquare* se dedica a descobertas e recomendações de locais.

Ao passo que outros *RSPs* têm interesse apenas por dados dos usuários, o *Foursquare* tem como diferencial a sua vasta base de dados de locais catalogados. Conforme dados divulgados por *Foursquare* (2016), essa *RSBL* possui:

- Comunidade: mais de 50 milhões de pessoas usam o *Foursquare* e o *Swarm* a cada mês, em *desktops*, *mobile web* e *mobile apps*. Pessoas fizeram *check-in* mais de 8 bilhões de vezes mundialmente.
- Plataforma: mais de 65 milhões de locais formam o mapa de negócios no mundo.

Outro ponto importante é que a classificação dos locais é feita de maneira hierárquica e leva em consideração a atividade que é predominantemente exercida no local em questão. Ao todo, existem 715 categorias, agrupadas em 10 diferentes estruturas de árvores, onde as raízes são consideradas as categorias primárias das estruturas e, em geral, apresentam uma classificação mais ampla da atividade do local, enquanto as folhas são classificações bastante específicas das atividades.

O *Musée du Louvre*, por exemplo, pertence à categoria *Art Museum*, que é um nó folha e uma categoria filha da subárvore *Museum*, a qual, por sua vez, é uma categoria filha da categoria primária *Arts & Entertainment* que é a raiz de uma das 10 árvores.

4

CARACTERIZAÇÃO DOS DADOS

NESTE capítulo serão discutidos todos os aspectos referentes aos dados utilizados no experimento. Em especial, serão apresentados todos os processos que foram necessários para a utilização de dados de *RSP*, como extração, tratamento e agrupamento de dados.

4.1 Processamento dos Dados

Nesta seção, serão expostos os processos de extração dos dados, mostrando como foi realizada a captura dos dados utilizados no trabalho; o tratamento dos dados; as modificações necessárias que foram realizadas, como remoção imperfeições e formatação; e, por fim, o agrupamento de dados, ou seja, como foram associados por meio de padrões.

4.1.1 Extração dos Dados

A extração de dados de *SSPs* não é um processo simples, pois demanda tempo, alto custo computacional e, algumas vezes, custo financeiro com a compra de dados. Existem diversas formas de coleta de dados de *SSP*, as mais comuns são por meio de *APIs*, *Web Crawler* e aplicações.

Dentre as *APIs*, existem dois tipos mais usados para captura de dados, são elas *REST* e *STREAMING*. Em ambos os casos, existem limites severos de quantidade de dados que podem ser extraídos. Podem-se listar diversos motivos para a criação dessas limitações, tais como a economia de banda de Internet por parte das plataformas e, principalmente, a venda de dados para terceiros que desejam minerá-los, uma vez que é uma das fontes de renda dessas ferramentas.

No caso específico do *Foursquare*, sua *API* fornece uma quantidade muito limitada de dados gratuitos, como se pode ver em [Foursquare \(2017\)](#). Contudo, é possível utilizar outro

canal para adquirir tais dados, que é por intermédio de *check-ins* do *Foursquare* que também foram compartilhados no *Twitter*. Baseado no trabalho de Noulas et al. (2011a), é esperado que amostras coletadas utilizando tal mecanismo correspondam de 20% a 30% da quantidade total de dados compartilhados pela plataforma.

Utilizando *API* de *Streaming* do *Twitter* é possível capturar mensagens praticamente em tempo real. Mais detalhadamente, segundo (Twitter, 2017), o processo de *streaming* recebe os *tweets* de entrada e não executa qualquer análise, filtragem e/ou agregação antes de armazenar o resultado em um banco de dados. O processo de tratamento *HTTP* consulta o banco de dados para obter resultados em resposta a solicitações de usuários. Enquanto esse modelo é mais complexo do que a *API REST*, os benefícios de se ter um fluxo de dados em tempo real fazem a integração valer a pena para muitos tipos de aplicações. Um exemplo do processo de consulta a *API STREAMING* pode ser visto na imagem 4.1.

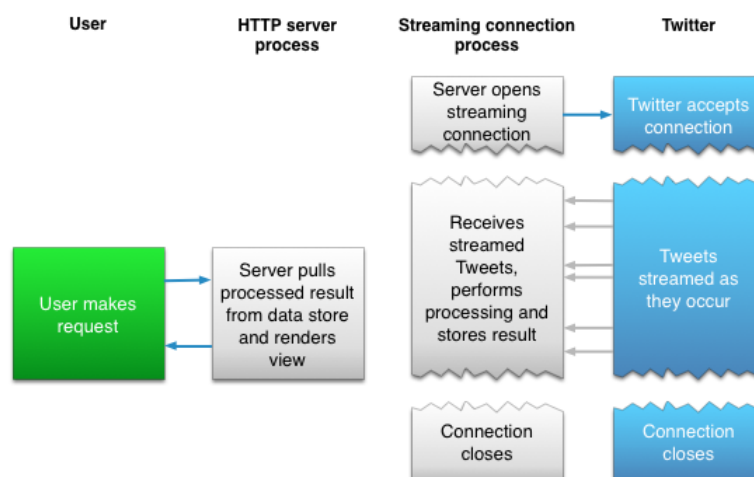


Figura 4.1: Modelo de uso *TWITTER STREAMING API*

Para este trabalho, foram armazenados 33 dias do fluxo de *check-ins* do *Foursquare* compartilhados no *Twitter*, no período entre os dias 25-04-2014 e 18-06-2014. Dentre esses dias, ocorreram interrupções na captura de dados - entre os dias 04-05-2014 e 06-05-2014; 20-05-2014 e 22-05-2014; e 30-05-2014 e 04-06-2014 - sendo possível ver essas interrupções, bem como a frequência diária de *check-ins* em todo o período de extração de dados na Imagem 4.2.

O *Foursquare* armazena uma grande quantidade de informações dentro de cada *check-in*. Quando um usuário faz uma postagem – dentre outros parâmetros, não explorados nesse trabalho – são automaticamente armazenadas informações sobre o usuário, a data e a hora em que o *check-in* foi feito e dados de geolocalização do usuário no exato momento da publicação. Por conta desse último parâmetro, o *Foursquare* é também capaz de adicionar um link para uma página dentro do seu domínio, contendo diversas informações sobre o

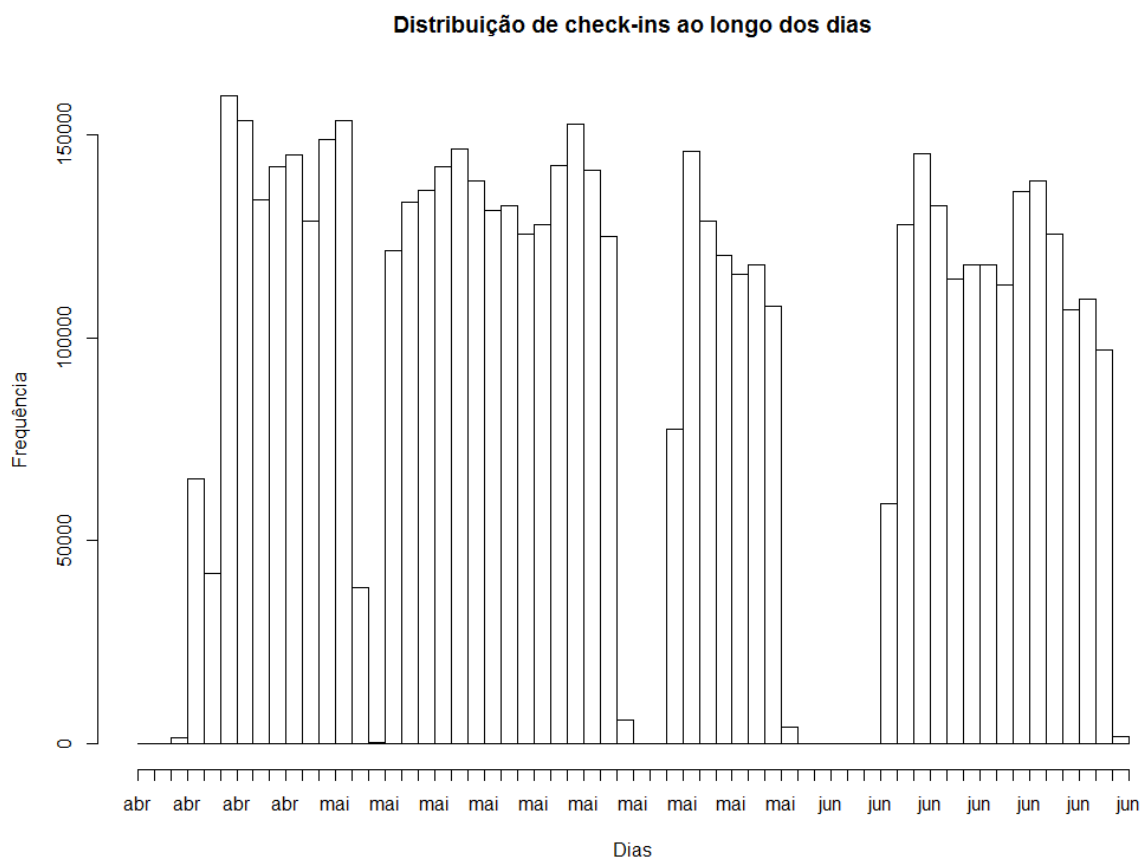


Figura 4.2: Histograma da frequência diária de *check-ins*, de 25-04-2014 à 18-06-2014

local onde o *check-in* foi realizado, como é possível ver na imagem 4.3. Mais especificamente, nessas páginas são exibidos mapas, fotos e avaliações do local feitas pelos usuários durante cada *check-in*. Além disso, também são exibidas as categorias, que são explicadas em detalhes na seção 3.2.1. Seguindo o princípio utilizado em trabalhos como o de Noulas et al. (2011b), Phithakitnukoon & Olivier (2011) e Silva et al. (2012), tais categorias foram utilizadas como forma de agrupamento de dados.

No entanto, para ter acesso às categorias, foi feito um novo processo de extração de dados, pois, como dito anteriormente, o *check-in* que é capturado pela API apenas traz o link para a página que contém tais dados, imagem 4.4. Nesse processo, foram utilizadas técnicas de *Web Crawler* juntamente com análise de expressão regular, para extrair a informação de dentro da página. Esse procedimento foi executado para todos os *check-ins* armazenados e a informação anexada as demais, capturadas na primeira etapa de extração.

Após a fase de coleta de dados no fluxo de postagens no *Twitter* e a posterior busca das categorias no *Foursquare*, foram escolhidas dez cidades para que, com o auxílio dos dados de geolocalização, latitude e longitude, contidos em cada *check-in*, fosse possível criar um subconjunto de dados com mensagens postadas apenas nessas cidades.

```
retweeted : false
▶ coordinates {2}
timestamp_ms : 1435280034196
▼ entities {5}
▶ user_mentions [0]
▶ symbols [0]
▶ trends [0]
▶ hashtags [0]
▼ urls [1]
▼ 0 {4}
url : ██████████
▼ indices [2]
0 : 60
1 : 82
expanded_url : http://4sq.com/██████████
display_url : 4sq.com/1Gvutc5
in_reply_to_screen_name : null
id_str : 614235099006128128
retweet_count : 0
in_reply_to_user_id : null
favorited : false
```

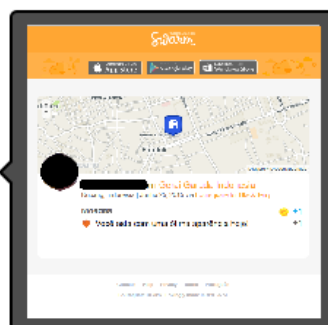


Figura 4.3: Exemplo de trecho de *check-in*, salientando o link para página no *Swarm*.

Para isso, foram cruzadas as localizações de cada mensagem compartilhada com os *Bounding Box* das cidades - caixas imaginárias que definem os limites geográficos em termos de latitude e longitude - resultando em um novo conjunto de dados formado apenas por *check-ins* realizados em uma dada cidade.

4.1.2 Tratamento dos Dados

Após o processo de extração de dados, deu-se início à fase de tratamento dos mesmos, que foi dividida em três partes: *i.* seleção, onde foram escolhidos os dados que fizeram parte do experimento; *ii.* ajustes, nessa etapa os dados foram modificados para se adequarem às condições necessárias para a realização do trabalho; *iii.* exclusão, estágio onde foram identificadas e removidas possíveis inconsistências.

Seleção de Dados

A primeira etapa em um processo de tratamento de dados é a seleção. Para esse trabalho, foram escolhidos apenas dados - contidos no *check-in* - que possam ajudar a representar a dinâmica social urbana, são eles: identificação do usuário (IDUSER), data e hora (TIMESTAMP), cidade (CITY), classificação do local (CATEGORY) e fuso horário (TIMEZONE). É possível ver

```

    "photos": {
      "count": 0,
      "items": []
    },
    "venue": {
      "name": "Gerai Garuda Indonesia",
      "beenHere": {},
      "stats": {
        "checkinsCount": 235,
        "usersCount": 2,
        "tipCount": 0
      },
      "location": {
        "city": "Kupang",
        "lng": 123.61058519070481,
        "contextLine": "Kupang, Indonésia",
        "state": "Nusa Tenggara Timur",
        "country": "Indonésia",
        "address": "Jl R W Monginsidi Ruko Lontar Permai",
        "cc": "ID",
        "lat": -10.160275272388564
      },
      "id": "551213b4498ef8fef55c2b3a",
      "canonicalPath": "\v\gerai-garuda-indonesia\551213b4498ef8fef55c2b3a",
      "canonicalUrl": "https://foursquare.com\v\gerai-garuda-indonesia\551213b4498ef8fef55c2b3a",
      "categories": [{
        "pluralName": "Escritórios",
        "name": "Escritório",
        "icon": {
          "prefix": "https://ss3.4sqi.net/img/categories_v2/building/default_",
          "mapPrefix": "https://ss3.4sqi.net/img/categories_map/building/default",
          "suffix": ".png"
        },
        "id": "4bf58dd8d48988d124941735",
        "shortName": "Escritório",
        "primary": true
      }
    ]},
    "verified": false,
    "photos": {
      "count": 48,
      "groups": [{
        "type": "venue",
        "name": "Fotos do lugar",
        "count": 48,
        "items": [{
          "source": {

```

Figura 4.4: Exemplo de código fonte da página do *Swarm* onde se encontram as categorias.

uma pequena amostra desses dados na tabela 4.1.

Dados referentes à localização, como latitude e longitude, foram excluídos, pois já haviam sido utilizados para capturar a cidade em que o *check-in* foi realizado. Também foram excluídos todos os campos referentes a dados pessoais dos usuários, restando apenas o *ID*'s que o *Foursquare* utiliza para identificá-los, os quais serão usados em um próximo passo de tratamento de dados.

Ajuste na Classificação de Locais

O *Foursquare*, como mostrado anteriormente na seção 3.2.1, tem um vasto banco de dados de locais cadastrados, organizados de maneira hierárquica por meio de categorias. Neste trabalho, iremos utilizar apenas nove das dez categorias primárias do *Foursquare*, quais sejam: *Arts & Entertainment (A&E)*, *College & University (C&U)*, *Food (FD)*, *Nightlife Spot*

Tabela 4.1: Exemplo de dados extraídos de *check-ins*

IDUSER	TIMESTAMP	CITY	CATEGORY	TIMEZONE
XXXXX	2014-04-27 00:00:05	Bangkok	Residential Building	420
XXXXX	2014-04-27 00:00:08	Los Angeles	General College & University	-420
XXXXX	2014-04-27 00:00:08	Tokyo	Karaoke Bar	540
XXXXX	2014-04-27 00:00:20	Bangkok	Gay Bar	420
XXXXX	2014-04-27 00:00:42	New York	American Restaurant	-240

(NL), *Outdoors & Recreation (O&R)*, *Professional & Other Places (P&O)*, *Residence (R)*, *Shop & Service(S&S)* e *Travel & Transport (T&T)*. A exclusão da categoria *Event* se deu por conta da não periodicidade da ocorrência de *check-ins* na mesma, já que caracteriza atividades esporádicas e que, em geral, não se repetem ou se repetem com um período maior do que a amostra que será utilizada, como é o caso de festivais de música, conferências e convenções.

Em seguida, uma transformação importante foi realizada, a reclassificação dos locais de *check-ins* onde as categorias mais específicas foram substituídas pelas categorias mais generalistas, ou seja, todos os *check-ins* das bases de dados passaram a pertencer a uma das nove categorias listadas no parágrafo anterior. Essa transformação foi possível porque, conforme foi visto na seção 3.2.1, categorias que pertencem a uma mesma árvore têm uma mesma característica principal, e essa é representada por sua categoria primária. Sendo assim, *check-ins* feitos no *Museu do Louvre* deixaram de ter como categoria *Art Museum* e foram reclassificadas como *Arts & Entertainment*.

A princípio, tal transformação pode parecer uma perda de informação, mas, muitas vezes, para realizar análises em grande escala, é necessário sumarizar os dados para, só então, ser possível enxergar padrões.

Exclusão de Dados

Como o foco desse trabalho é estudar a dinâmica das cidades, foram considerados apenas *check-ins* de usuários com uma certa periodicidade. Mais especificamente, foram consideradas postagens feitas por um mesmo usuário em um mesmo dia social e com até no máximo 6 horas de diferença e, ainda, com diferença mínima de 5 minutos caso as categorias fossem diferentes e de 30 minutos se fossem iguais. Tal escolha foi feita por assim existir uma maior probabilidade dessas postagens representarem o real padrão de mobilidade do usuário.

4.1.3 Agrupamento de Dados

Ao término do processo de tratamento dos dados, foi realizado o agrupamento dos mesmos, dividido em três partes. Primeiro, os dados foram agrupados por cidade e dia da semana; em seguida, foram criadas sequências de transição entre categorias; e, por fim, utilizando as sequências criadas foram construídas *Matrizes de Frequência(MFs)* que representam o padrão

de mobilidade das cidades.

Agrupamento por Cidade e Dia da Semana

Inicialmente, os dados foram agrupados por cidade e, então, separados em diferentes amostras. Assim, o conjunto de dados foi distribuído em dez partes, referentes a cada uma das cidades selecionadas: São Paulo, Rio de Janeiro, Recife, New York, Los Angeles, Mexico City, Bangkok, Manila, Moscow e St. Petersburg.

É importante salientar que, a princípio, o volume total de dados era 1.676.054 *check-ins* dividido da seguinte forma: São Paulo(179.071), Rio de Janeiro (114.674), Recife (83.994), New York (165.301), Los Angeles (64.227), Mexico City (393.235), Bangkok (317.032), Manila (159.400), Moscow (136.835) e St. Petersburg (62.285). Após toda a fase tratamento, o conjunto da dados passou a ter 1.654.044 *check-ins*, distribuído da seguinte forma: São Paulo (173.710), Rio de Janeiro (113.695), Recife (82.477), New York (163.513), Los Angeles (63.601), Mexico City (392.139), Bangkok (305.813), Manila (168.625), Moscow (131.755) e St. Petersburg (58.716).

Com os dados separados por cidade, então, foi aplicado o fuso horário e, em seguida, separado as amostras por dia, no entanto, não foi escolhido o formato tradicional de representação do dia, que começa a meia noite e vai até a meia noite do outro dia, foi escolhido trabalhar com o formato de dia social, que vai das 04h às 04h da manhã. Esta escolha foi feita principalmente por conta da categoria *Nightlife Spot*, que representa locais de vida noturna, comumente frequentados até muito após a meia noite. Vale salientar que, como muitos desses locais também funcionam pelo dia, como alguns bares e casas de shows, por exemplo, todas as publicações feitas nesses locais fora do horário noturno, ou seja, das 18h às 04h, foram desconsideradas.

Transição Entre Categorias

Já que o conjunto de dados conta apenas com *check-ins* feitos por usuários com uma frequência diária de postagens e os locais onde tais *check-ins* foram realizados estão todos classificados por categorias do *Foursquare*, foi possível, agrupando as amostras por usuário e data, construir uma sequência de transições de tipos de locais por onde os usuários estiveram ao longo de um determinado dia. É possível ver um exemplo dessa sequência na Tabela 4.2.

Tabela 4.2: Exemplo do padrão de mobilidade social de um usuário no dia 18/05/2014 (domingo).

IDUSER	TIMESTAMP-BEGIN	TIMESTAMP-END	CATEGORY-BEGIN	CATEGORY-END
XXXXXX	2014-05-18 09:28:33	2014-05-18 10:36:23	Food	Residence
	2014-05-18 10:36:23	2014-05-18 12:34:25	Residence	Arts and Entertainment
	2014-05-18 12:34:25	2014-05-18 13:20:23	Arts and Entertainment	Residence
	2014-05-18 13:20:23	2014-05-18 14:29:47	Residence	Food
	2014-05-18 14:29:47	2014-05-18 15:47:08	Food	Travel and Transport
	2014-05-18 15:47:08	2014-05-18 16:32:40	Travel and Transport	Residence

Para melhor visualização, os dados da Tabela 4.2 foram representados por meio de um grafo (Imagem 4.5), onde os vértices são as categorias e o peso das arestas as horas entre cada *check-in*.

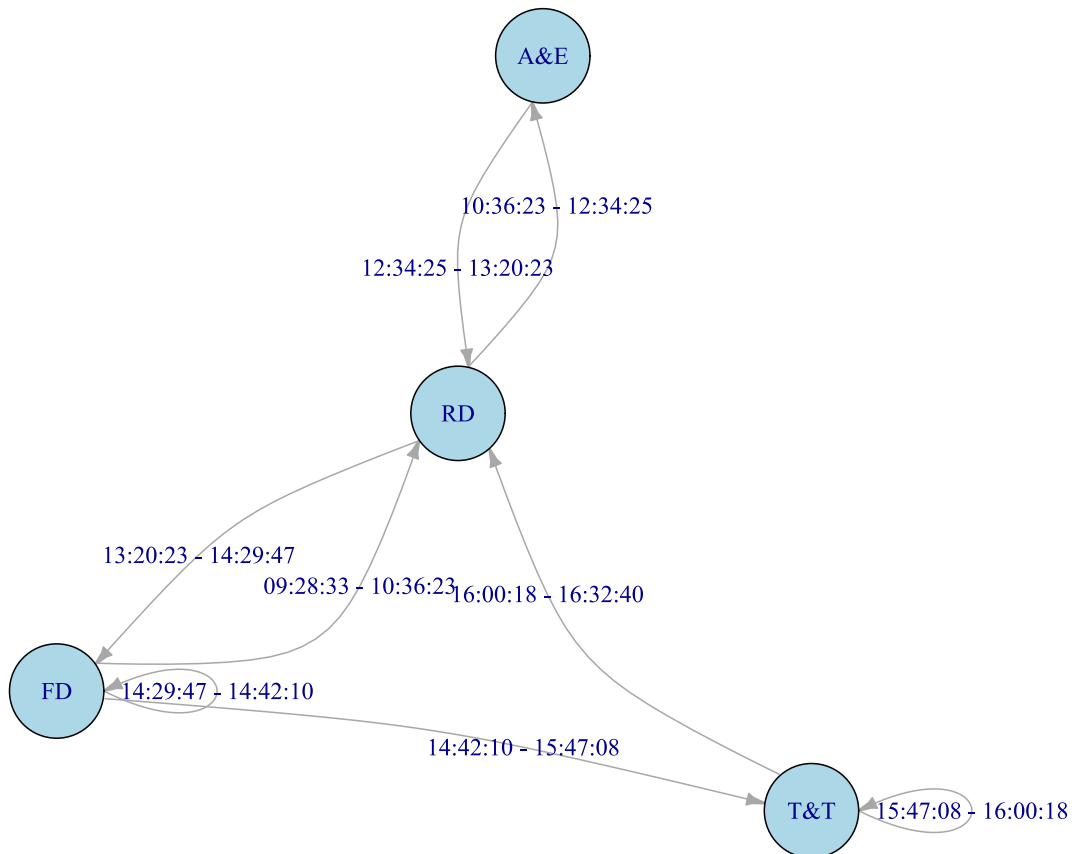


Figura 4.5: Grafo de transições entre categorias de um usuário no dia 18/05/2014 (domingo).

Observando a Imagem 4.5 é possível saber quais os tipos de locais que foram visitados pelo usuário acompanhando a ordem dos *check-ins* feitos por ele, por exemplo, é possível saber que seu primeiro *check-in* é feito pela manhã (09:28:33) na categoria *Food*, em seguida, seu próximo *check-in* é na categoria *Residence* (10:36:23), levando a crer que esse foi seu primeiro deslocamento no dia, para acompanhar o restante de seu deslocamento basta seguir a ordem de tempo representadas nas arestas.

É importante perceber que nem sempre é possível capturar todo o rastro de transições que os usuários deixaram ao longo de suas atividades diárias, mas, levando em consideração todo o processo de tratamento de dados apresentado até aqui, é factível supor que as amostras ainda presentes nas bases de dados representam potencialmente transições feitas em sequência. Para esse trabalho, essa suposição é tratada como verdade.

Construindo Matrizes de Frequência

Para finalizar o agrupamento dos dados, foram criadas *MFs* de transições entre categorias, as quais guardam informações acerca do deslocamento diário de um usuário entre elas. Assim, é possível ter ideia do padrão de mobilidade de cada usuário dentro de sua cidade.

Tabela 4.3: Matrizes de Frequência da cidade de São Paulo

		Segunda-feira											Sábado								
		A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T			A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
A&E		13	3	6	1	1	6	4	7	1		A&E	8	2	26	14	12	4	2	3	5
C&U		1	28	18	0	6	13	29	10	7		C&U	3	2	15	0	4	5	3	9	3
FD		10	23	51	4	22	40	14	34	13		FD	28	4	65	18	19	17	29	50	9
NL		0	2	2	1	1	1	2	1	0		NL	6	0	7	12	3	0	2	1	1
O&R		4	6	20	0	19	24	19	14	18		O&R	13	5	24	7	27	7	16	23	11
P&O		7	21	67	2	17	53	29	26	18		P&O	5	2	29	4	7	19	17	20	8
RD		1	8	3	0	10	14	11	6	1		RD	6	0	16	6	12	12	28	10	6
S&S		9	10	54	2	18	19	15	64	10		S&S	22	0	101	10	20	14	26	116	12
T&T		1	7	20	3	11	23	18	23	64		T&T	9	7	15	4	20	8	7	21	36

Contudo, como esse trabalho visa realizar um estudo em larga escala, foram sumarizadas as informações diárias de todos os usuários de uma cidade em uma só matriz, que passa a representar a dinâmica social dessa cidade. Essa abordagem é semelhante a de [Silva et al. \(2012\)](#), que as utiliza para construir o que chama de *Image of City*.

As matrizes da Tabela 4.3 representam um exemplo das *MFs* geradas a partir da aplicação do processo explicado à cidade de São Paulo nos dias 12/05/2014 e 17/05/2014.

O uso de tais matrizes para criar diferentes modelos de agrupamento de cidades, que leva em consideração a dinâmica social urbana, é o foco deste trabalho e será apresentado em detalhes no próximo capítulo.

5

RESULTADOS E ANÁLISES

ESTE capítulo é dividido em quatro partes principais: A Seção 5.1 compara características de uma mesma cidade em diferentes dias da semana com o objetivo de descobrir qual a maneira mais adequada de se agrupar os dados. Na Seção 5.2 é, primeiramente, apresentada uma maneira de se comparar cidades por meio de características específicas e, em seguida, discutidos os resultados da comparação entre New York e São Paulo com as demais cidades selecionadas. Já seção 5.3 é sugerida uma nova modelagem para representação da dinâmica das cidades que viabiliza a comparação simultânea de todas características sociais das cidades, apresentando o resultado de tal técnica tanto na comparação de cidades por dia e por turno. Por fim, na seção 5.4, é sugerido um exemplo de aplicação que faz uso dessas técnicas de comparação entre cidades.

5.1 Comparações entre Matrizes de Frequência de uma Mesma Cidade

Como mostrado anteriormente, o processo de tratamento teve como resultado matrizes que representam o padrão de mobilidade das cidades. Essas matrizes são separadas por cidade e dia. Nesta seção, elas foram investigadas com o objetivo de responder a duas perguntas: (i) o comportamento social das pessoas independe do dia da semana? Ou seja, é possível agrupar as amostras diárias sem haver distorção da informação? É possível separar por dias de semana e finais de semana? (ii) A população segue uma rotina semanal? Ou melhor, é possível associar os dados de um mesmo dia da semana, mas de semanas diferentes?

Antes de responder a tais perguntas, é preciso ter em mente que as cidades se comportam de maneiras diferentes em relação a cada categoria, ou seja, o comportamento social ao longo do tempo das pessoas quanto à alimentação (categoria *FD*), por exemplo, é logicamente diferente se comparado ao comportamento em relação ao trabalho (categoria *P&O*), o que

pode ser facilmente percebido analisando verticalmente os histogramas da Imagem 5.1. Então, só é aconselhável considerar dois dias semelhantes se todas as categorias, de ambos os dias, tiverem comportamentos similares.

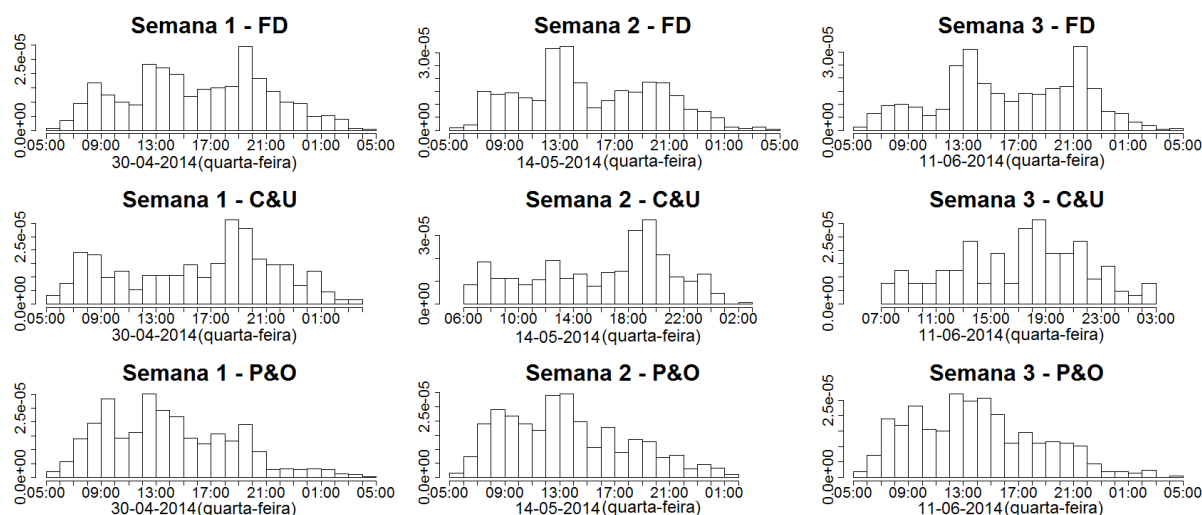


Figura 5.1: Histogramas da cidade de São Paulo apresentando o comportamento ao longo do tempo de 2 categorias em 3 diferentes semanas.

Dito isso, uma dedução lógica – e que pode ser facilmente confirmada verificando a diferença nos valores das matrizes da Tabela 4.3 – é que as chances de transições entre as categorias são diferentes, isto é, a probabilidade dos usuários saírem de casa (categoria *RD*) para o trabalho (categoria *C&U*), por exemplo, é diferente da deles irem ao cinema (categoria *A&E*). No entanto, não é possível afirmar facilmente que tais probabilidades se mantêm independentemente do dia da semana.

Então, para responder a primeira pergunta, foi inicialmente selecionada uma cidade e, em seguida, comparadas, categoria a categoria e dia a dia, todas as matrizes de frequência em um período de uma semana. Para isso, foi utilizado um teste Qui-quadrado com o intervalo de confiança de 99%, onde a hipótese nula pode ser enunciada como: as frequências de uma determinada linha l (categoria) da matriz M_{seg} são iguais às frequências da linha l da matriz M_{ter} . Esse teste foi aplicado para todos os dias da semana e para todas as cidades. A fim de analisar seu resultado, foram criadas tabelas para cada categoria, com o resultado do nível descritivo (*p-valor*) da comparação de todos os dias da semana.

Para melhor ilustrar a técnica, considere-se o exemplo de New York. Foram selecionadas todas as matrizes de frequência diárias da cidade, e, após, comparada a linha correspondente a categoria *A&E* de segunda-feira com a linha correspondente à mesma categoria na matriz de terça-feira. Esse processo foi repetido para todos os dias e categorias e resultou uma matriz de *p-valor* que é simétrica, pois não há diferença na comparação da segunda com a terça-feira e de terça com a segunda-feira da mesma semana.

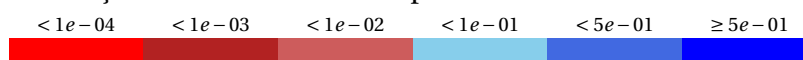
Como as tabelas construídas são de difícil visualização, devido a grande quantidade de

números com várias casas decimais, foram criadas matrizes de escala de cores para melhor visualização. Em tais matrizes, as variações da cor azul representam os valores que passaram no teste, já as da cor vermelha as que não passaram. A Tabela 5.1 apresenta em mais detalhes a representação numérica de cada tonalidade de cor. Essa escala será adotada para todos resultados de comparação entre amostras de uma mesma cidade contidos neste trabalho.

As Tabelas A.1 e A.2, que representam os *p-valores* do teste para todas as categorias das cidades de New York (NY) e São Paulo (SP), respectivamente, foram adicionadas ao apêndice juntamente com a Tabela A.3, que representa o *p-valores* dos testes que obtiveram resultados interessantes e/ou diferentes do comum em algumas categorias das demais cidades selecionadas.

Já as matrizes de cores das cidades de NY e SP são representadas pelas Tabelas 5.2 e 5.3, respectivamente, com as matrizes de cores das categorias destacadas das demais cidades sendo representadas pela Tabela de 5.4.

Tabela 5.1: Representação de valores números por meio de escala de cores de vermelho e azul



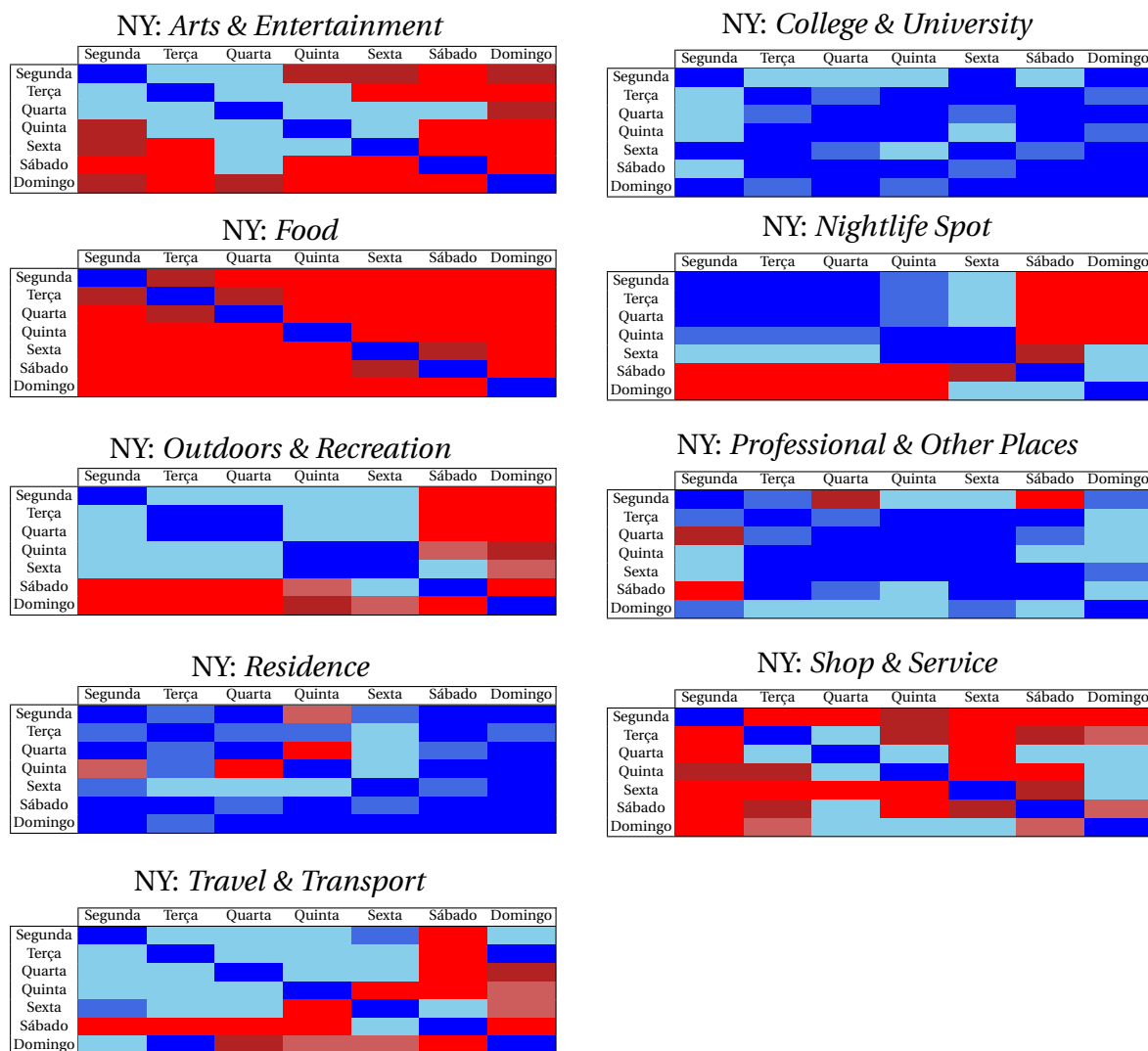
Verificando alguns resultados, foi possível perceber que, por exemplo, o comportamento da categoria *C&U* tende sempre a se manter bastante regular em todos os dias da semana, inclusive nos finais de semana, como visto em New York e Moscow (Tabelas 5.2 e 5.4), onde a população deve frequentar bibliotecas e outros locais de estudo. No entanto, esse comportamento não se repete em algumas cidades, como Manila (Tabela 5.4), onde há uma clara diferença entre dias de semana e finais de semana.

O mesmo comportamento pode ser observado na categoria *RD*, como nos resultados do teste em cidades como New York (Tabela 5.2) e Bangkok (Tabela 5.4). Contudo, ao contrário da maioria das cidades, o padrão de mobilidade de São Paulo (Tabela 5.3) em relação a essa categoria é bastante instável, mostrando, assim, o porquê de ser conhecida como “A cidade que nunca dorme”. Das cidades analisadas, outra que obteve o comportamento semelhante em relação a tal categoria foi a Cidade do México (Tabela 5.4).

Já a categoria *NL*, na maioria dos casos analisados, segue um padrão de comportamento bastante homogêneo durante os dias da semana. Por sua vez, nos finais de semana, há um grande aumento nos valores do teste, como podemos ver nas cidades de New York (Tabela 5.2) e Rio de Janeiro (Tabela 5.4). Esse resultado é bastante plausível, eis que, durante a semana, em razão do trabalho, a população tende a não frequentar a vida noturna, atividade essa que aumenta bastante aos finais de semana, em especial aos sábados.

Uma diferença percebida foi em relação à cidade de São Paulo (Tabela 5.3), na qual, apesar

Tabela 5.2: Matrizes de escala de vermelho e azul que representam os resultados da comparação entre as Matrizes de Frequência de dias de uma mesma semana de New York

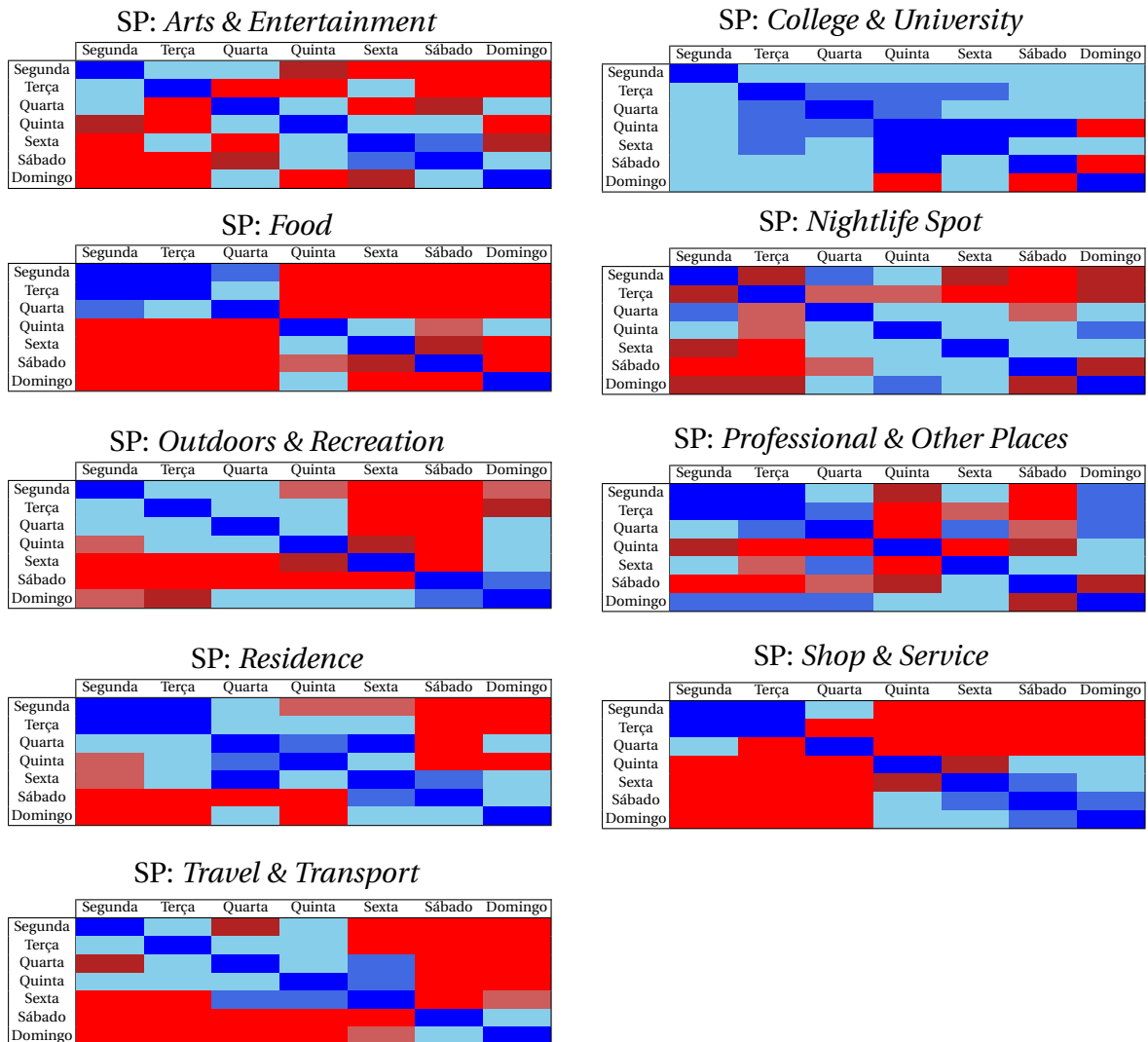


de ser notada uma clara mudança do comportamento a medida que se aproxima o final de semana, as comparações com dia de terça-feira apresentam uma diferença de valores bem elevada em relação aos outros dias, levando a crer que possivelmente nesse dia específico existiu um evento relacionado a essa categoria.

Curiosamente, outra peculiaridade notada é que em grande parte das cidades estudadas, como em New York (Tabela 5.2) e São Paulo (Tabela 5.3), o comportamento da população no domingo se assemelha mais ao de sexta. Resultado esse que provavelmente deve se dar em razão de que, tanto o domingo como a sexta são dias que nem todos escolhem sair à noite, já que muitos trabalham no sábado e querem descansar no domingo. No entanto, no Rio de Janeiro (Tabela 5.4) esse comportamento não é observado, sendo sábado, notadamente, o dia com o comportamento mais próximo ao de domingo.

Como na categoria *NL*, a categoria *T&T* também apresenta uma diferença no comporta-

Tabela 5.3: Matrizes de escala de vermelho e azul que representam os resultados da comparação entre as Matrizes de Frequência de dias de uma mesma semana de São Paulo

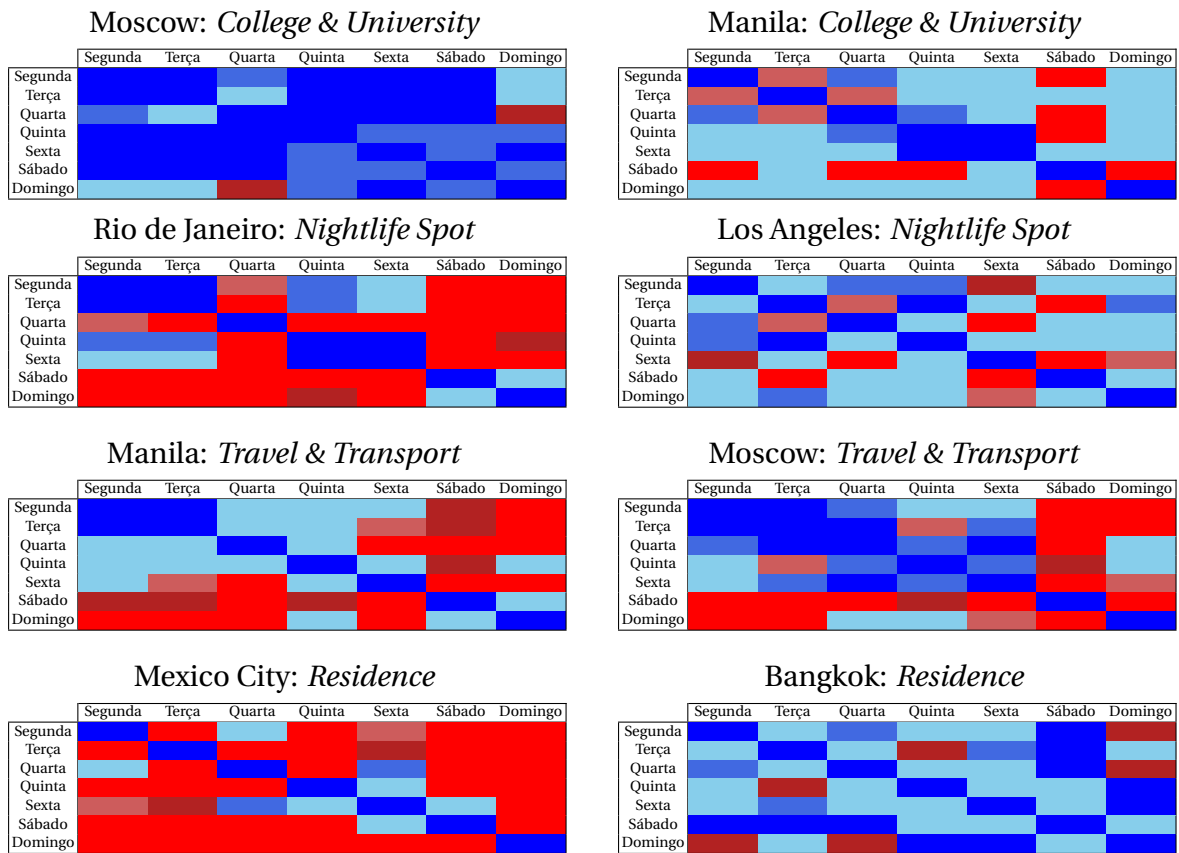


mento da população entre os dias da semana e finais de semana. Em cidades como New York (Tabela 5.2), São Paulo (Tabela 5.3), Manila (Tabela 5.4) e Moscow (Tabela 5.4) é possível ver grau de similariedade considerável entre os dias da semana, mas uma desordem elevada em finais de semana, principalmente em São Paulo, Manila e Moscow.

Na categoria S&S (Tabelas 5.2 e 5.3), é notado um alto grau de irregularidade entre os resultados. Tal comportamento, apesar de inconveniente, já era esperado, pois se trata de uma categoria relacionada a compras, onde não existe uma periodicidade de comportamento, principalmente em se tratando de grandes metrópoles como as estudadas.

De fato, analisando todas as tabelas é possível perceber que o comportamento varia muito de acordo com a categoria estudada, sendo possível, normalmente, dividi-las em três grupos: (1) categorias C&U, P&O e RD têm um comportamento mais linear, pois representam a rotina diária da população; (2) S&S e FD têm uma maior variação, uma vez que são atividades que

Tabela 5.4: Matrizes de escala de vermelho e azul que representam os resultados da comparação entre as Matrizes de Frequência de dias de uma mesma semana de outras cidades

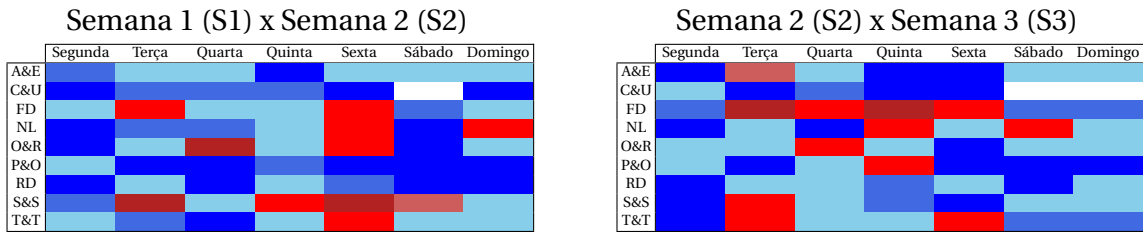


não fazem parte do dia-a-dia; (3) *A&E, NL, O&R e T&T* são atividades que seguem padrões bem diferentes em dias da semana e em finais de semana, já que representam, em geral, atividades de lazer.

Como foi possível ver, apenas algumas das categorias mantiveram um padrão ao longo dos dias, as demais têm uma dinâmica bastante diferente. Para o intervalo de confiança definido (99%), no grupo 1, não existem evidências conclusivas, na grande maioria dos resultados, para rejeitar o teste. No entanto, no grupo 2, existem claras evidências para se rejeitar o teste em praticamente todos os resultados. Já no grupo 3, as categorias aparentam poder ser associadas nos dias da semana, contudo, em finais de semana seguem um padrão de mobilidade bastante diferente. Em suma, conforme dito anteriormente, não é conveniente agregar dois dias, a menos que todas as categorias desses sejam semelhantes, então nesse trabalho foram analisados todos os dias da semana separadamente.

Em relação ao questionamento quanto a população seguir ou não uma rotina semanal, foram separadas três semanas (*S1, S2 e S3*) com amostras diárias de todas as categorias e, de maneira análoga à anterior, executou-se um teste Qui-quadrado para comparar novamente, categoria a categoria, o mesmo dia da semana das três diferentes semanas selecionadas.

Tabela 5.5: Matrizes de escala de vermelho e azul que representam os resultados da comparação entre as Matrizes de Frequência do mesmo dia de diferentes semanas de Los Angeles



Analisando os resultados do exemplo da cidade de Los Angeles (Tabela A.4), foi possível perceber que na comparação da semana *S1* com a semana *S2*, para o intervalo de confiança de 99%, não existem evidências para se refutar o teste em aproximadamente 81% dos casos. Já na comparação das semanas *S2* e *S3* o valor para os mesmos 99% de nível de significância é de 76%. Vale salientar que os valores em branco nas tabelas não atingiram as condições mínimas para uma execução fidedigna do teste.

Tabela 5.6: Matrizes de escala de vermelho e azul que representam os resultados da comparação entre as Matrizes de Frequência do mesmo dia de diferentes semanas de New York

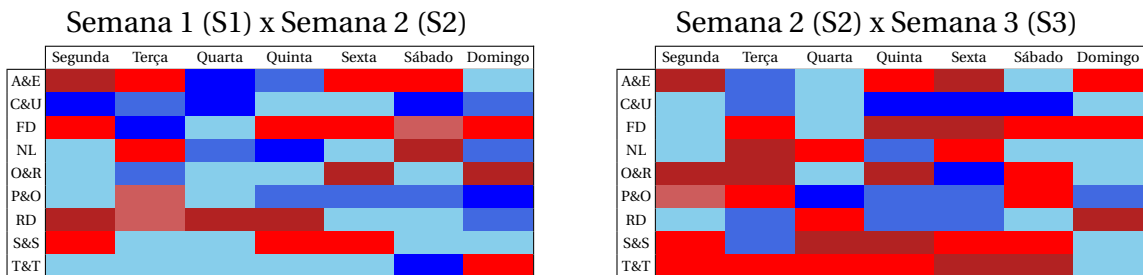
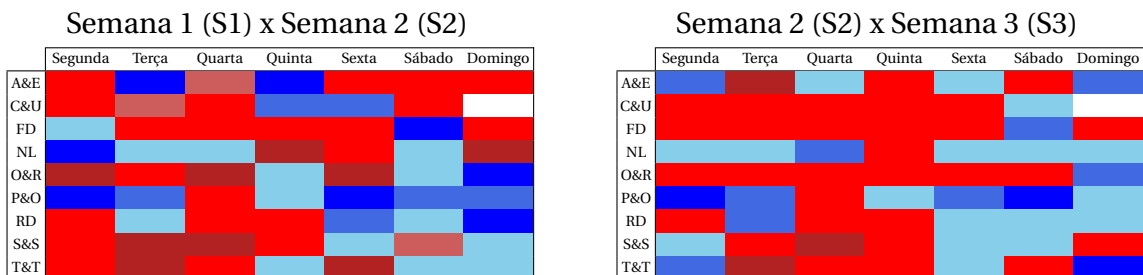


Tabela 5.7: Matrizes de escala de vermelho e azul que representam os resultados da comparação entre as Matrizes de Frequência do mesmo dia de diferentes semanas de São Paulo



Já avaliando os resultados para as cidades de New York (Tabela A.5) e São Paulo (Tabela A.6), novamente com um intervalo de confiança de 99%, os números são mais baixos. Em New York, na comparação entre a semana *S1* e *S2* não há evidências conclusivas para se refutar o teste em 65% dos resultados. Já na cidade de São Paulo, os números são de, 44%. Por sua vez,

na comparação entre as semanas S2 e S3 os resultados são de 49% para New York e 46% para São Paulo.

Verificando os números acima, é possível perceber que existe uma semelhança entre as semanas. É importante salientar que já era esperada uma relativa variação entre as amostras, em razão da variabilidade intrínseca à população estudada e os ruídos nas amostras coletadas – a exemplo dos eventos ocorridos entre os dias amostrados. Neste trabalho foi realizada a junção das semanas, mas, em trabalhos futuros, pretende-se estudar a evolução dos dados ao longo do tempo, o que exigirá, contudo, uma amostra com uma fração de tempo maior que a ora utilizada.

5.2 Comparações entre Matrizes de Frequência de Diferentes Cidades

Extrapolando a análise da seção anterior, foram comparadas, por categorias e dias da semana, diferentes cidades com o objetivo de se descobrir o grau de semelhança entre elas quanto a características específicas.

Para ilustrar o método, foram novamente selecionadas as cidades de São Paulo e New York. As Tabelas de A.7 à A.13 e de A.14 à A.20 apresentam, respectivamente, os resultados (*p-valores*) do uso do teste Qui-quadrado para a comparação dessas duas cidades com as demais. Vale salientar que o teste é aplicado para cada categoria separadamente, ou seja, a categoria *i* da cidade *j* é comparada com a categoria *i* das ($n - j$) cidades estudadas. Como é possível perceber pelo número de tabelas de cada exemplo, esse processo é repetido para todos os sete dias da semana.

Cabe observar que não se trata de um verdadeiro teste de hipótese para comparar se duas cidades são iguais, o *p-valor*, nesse caso, apenas é utilizado como um parâmetro de representação de verossimilhança entre elas.

A título de exemplo, tem-se a comparação entre a cidade de São Paulo e do Rio de Janeiro, em um dia de segunda-feira. Os seguintes procedimentos foram adotados para a construção da comparação: escolha da categoria a ser comparada, no caso, *A&E*; seleção da Matriz de Frequência de segunda-feira de São Paulo; separação da linha referente à categoria a ser comparada, no caso a linha 1 da matriz; seleção da mesma categoria na Matriz de Frequência de segunda-feira de Rio de Janeiro; por fim, comparação, elemento a elemento, entre as linhas das matrizes. Ao final do processo, o resultado foi o elemento $M_{1,7}$ da Tabela A.7, sendo esse o grau de similaridade (*p-valor*) entre as duas cidades em relação à categoria *A&E*. Para a construção do restante da Tabela, tal processo foi repetido para as demais cidades e categorias.

Como na seção anterior, para facilitar a visualização dos resultados, foram criadas matrizes

de cores (Tabelas de 5.9 a 5.15 e 5.16 a 5.22) para representar as Tabelas de *p*-valor, as quais, do mesmo modo, foram adicionadas como apêndice no trabalho.

No entanto, como o teste de hipótese é utilizado, nesse caso, como artifício para comparação entre cidades que são logicamente diferentes, não seria razoável separar os resultados em aprovados e reprovados. Sendo assim, foi criada apenas uma escala de cinza (Tabela 5.8) para representar o grau de semelhança entre as cidades – essa escala será adotada para todos os resultados de comparação entre cidades diferentes constantes neste trabalho.

Os *p*-valores, então, foram utilizados para ranquear os resultados. Aplicou-se a análise a todos os dias da semana e, posteriormente, foi feito um ranking das cidades que mais se assemelham à São Paulo e New York em cada categoria, ou seja, que mais vezes aparecem como semelhantes durante todos dias da semana ou dos fins de semana.

Em relação à São Paulo, alguns resultados dessa análise merecem destaque:

- Levando em consideração todos os dias úteis, as cidades mais próximas em relação à categoria *P&O* - que representa todas as categorias ligadas a atividades de trabalho - são: Moscow, Los Angeles e Mexico City.
- Ainda tendo em conta apenas dias úteis, nas categorias ligadas à educação (*C&U*) as cidades mais parecidas são: Moscow, Rio de Janeiro e Recife.
- Já em relação aos finais de semana, na categoria que classifica locais de vida noturna, como bares e casas de shows (categoria *NL*), as mais próximas são: Bangkok, Rio de Janeiro e St. Petersburg.
- Também em relação aos finais de semana, só que na categoria que representa os hábitos alimentares da população (Categoria *FD*), as cidades mais parecidas são: Moscow, Recife e Rio de Janeiro.
- E, por fim, nos finais de semana, as cidades mais próximas na categoria que representa locais que oferecem conteúdo de arte e entretenimento (Categoria *A&E*), como museus, cinemas e teatros, são: Los Angeles, Moscow e Mexico City.

Tabela 5.8: Representação de valores números por meio de escala de cores de cinza



Tabela 5.9: Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de segunda-feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	Light		Light	Light		Light			
Los Angeles	Light	Light		Light			Light		
Manila	Light			Light					
Mexico City	Light		Light	Light	Light		Light	Light	
Moscow	Light	Light		Light	Light			Light	
New York	Light		Light	Light	Light				
Rio de Janeiro	Light	Light		Light			Light		
Recife	Light	Light		Light	Light		Light		
St. Petersburg	Light			Light	Light				

Tabela 5.10: Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de terça-feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	Light			Light					
Los Angeles	Light			Light			Light		
Manila				Light					
Mexico City	Light			Light	Light	Light		Light	
Moscow	Light	Light		Light					Light
New York	Light			Light			Light		
Rio de Janeiro	Light			Light	Light		Light		
Recife	Light	Light		Light	Light		Light		
St. Petersburg	Light			Light	Light				

Tabela 5.11: Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de quarta-feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	Light			Light	Light		Light		
Los Angeles	Light			Light			Light		
Manila	Light			Light					
Mexico City				Light	Light		Light		Light
Moscow	Light	Light		Light	Light		Light	Light	
New York	Light			Light	Light				
Rio de Janeiro	Light			Light	Light		Light		
Recife	Light			Light	Light		Light		
St. Petersburg	Light	Light		Light	Light		Light		

Tabela 5.15: Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de domingo

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok									
Los Angeles									
Manila									
Mexico City									
Moscow									
New York									
Rio de Janeiro									
Recife									
St. Petersburg									

Quanto à aplicação da análise a New York, os resultados destacados são os seguintes:

- Na categoria *A&E*, apesar de haver uma diferença elevada nos resultados se comparados aos obtidos pela cidade de São Paulo - apresentando, assim, um padrão de comportamento diferenciado para essa categoria em relação às demais cidades - Los Angeles sempre figura entre as mais próximas, não importando o dia.
- Analisando a categoria *NL*, os resultados dos testes mostram Mexico City e, novamente, Los Angeles entre as mais parecidas, sendo Mexico City mais semelhante nos dias úteis e Los Angeles aos finais de semana.
- Já na categoria *O&R*, destaca-se a aparição de São Paulo entre as três cidades mais próximas, acompanhada, mais uma vez, de Los Angeles e Mexico City.
- As categorias *S&S* e *T&T* seguem a mesma tendência das demais, com Los Angeles e Mexico City sendo as mais semelhantes a New York, apenas permutando as posições, com Los Angeles aparecendo à frente nos dias de semana e Mexico City aos finais de semana.
- Verificando os resultados da categoria *FD*, é notada uma grande diferença entre o comportamento das cidades, resultado semelhante ao apresentado em São Paulo. No entanto, mais uma vez, dentre as cidades com maior similaridade, estão Los Angeles e Mexico City.
- Talvez como maior surpresa, Moscow aparece entre as cidades entre as cidade mais próxima de New York em relação à categoria *C&U*.

Tabela 5.22: Matriz de escala de cinza que representa os resultados da comparação entre as Matrizes de Frequência de diferentes cidades com New York em um dia de domingo

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok									
Los Angeles									
Manila									
Mexico City									
Moscow									
São Paulo									
Rio de Janeiro									
Recife									
St. Petersburg									

Analisando os resultados é possível perceber, tanto nas comparações com São Paulo - onde aparecem, com frequência, as cidades brasileiras do Rio de Janeiro e Recife entre as mais similares - como em New York - onde aparecem Los Angeles e Mexico City entre as mais semelhantes em quase todas as categorias - que fatores geográficos demonstram uma forte relação com o comportamento social da população.

A princípio, alguns resultados podem causar estranheza, porém, muitas dessas cidades são grandes centros históricos, culturais e econômicos, só que nem sempre conhecidos. Vale salientar, contudo, que esse trabalho não tem como objetivo descobrir as causas das semelhanças entre as cidades, tendo como único objetivo analisar os resultados extraídos dos dados que foram compartilhados pelos usuários.

5.3 Comparações entre Matrizes de Markov de Diferentes Cidades

Ao estudar o comportamento das cidades separando-as em categorias foi possível entender a permutação entre uma categoria específica e as demais, todavia, para poder realmente entender a dinâmica de uma cidade, é necessário estudar o comportamento da população em relação a todas as categorias em conjunto.

Para retratar tal comportamento, foram utilizadas *Cadeias de Markov de Tempo Discreto (CMTD)*, onde cada categoria passa a representar um estado das *Matrizes de Transição de Probabilidade (MTP)*. Como logicamente não é possível contar com toda a população necessária para construir tais matrizes, essas foram estimadas a partir dos dados amostrais coletados de cada cidade.

A escolha por *CMTD* foi feita, em especial, por três características encontradas nesses tipos de cadeias, (i) é um processo estocástico bastante utilizado para representar estados discretos; (ii) têm capacidade de representar de forma semelhante dados com frequências de amostragem diferentes; e (iii) em *processos markovianos* a probabilidade condicional

de qualquer evento futuro depende somente do estado presente, sendo mais fácil de se representar dados temporais quando se tem apenas amostras de tempo.

Existem diferentes métodos para se estimar as *Matrizes de Transição* de *CMTDs*, neste trabalho utilizamos a abordagem descrita em Pardo (2005), como podemos ver na equação 5.1, a qual apresenta um estimador de máxima verossimilhança para as probabilidades de transição de estado da matriz de transição.

$$\hat{p}(i, j) = \frac{v_{ij}}{v_{i*}}, \text{ onde:} \tag{5.1}$$

$$v_{i,j} = \sum_{k=2}^n I_{\{(i,j)\}}(Y_{k-1}, Y_k) \text{ e } v_{i*} = \sum_{j=1}^M v_{ij}$$

Sendo $I_{\{(i,j)\}}(Y_{k-1}, Y_k)$ uma função definida por: $I_{\{(i,j)\}}(Y_{k-1}, Y_k) =$

$$I_{\{(i,j)\}}(Y_{k-1}, Y_k) = \begin{cases} 1 & \text{se } Y_{k-1} = i \text{ e } Y_k = j \\ 0 & \text{caso contrário} \end{cases}$$

É importante ressaltar que o estimador descrito em Pardo (2005) é um estimador de *Máxima Verossimilhança*, sendo consistente, porém tendencioso, com o viés tendendo a zero à medida que o tamanho da amostra aumenta. Assim, é possível estimar uma *Matriz de Transição* de maneira consistente caso se tenha uma amostra grande o suficiente.

A Tabela 5.23 apresenta um exemplo de *MTP* estimada a partir de dados de um dia de semana da cidade de New York. Como é possível perceber, existem probabilidades associadas à transição entre todos os estados da *Cadeia de Markov*, então é possível estimar, por exemplo, a probabilidade dos habitantes de New York irem a um bar (Categoria *NL*) após o trabalho (Categoria *P&O*) em um dia de segunda-feira, apenas verificando a intercessão entre os dois estados na matriz.

As *MTPs* estimadas dos demais dias da semana, que estão representadas pela Tabela A.28, foram adicionadas ao apêndice do trabalho.

Tabela 5.23: Matriz de Transição de Probabilidade de New York em um dia de segunda-feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
A&E	0.071	0.018	0.429	0.071	0.071	0.071	0.036	0.161	0.071
C&U	0.125	0.156	0.250	0.000	0.031	0.188	0.031	0.125	0.094
FD	0.055	0.043	0.250	0.070	0.117	0.148	0.070	0.148	0.098
NL	0.051	0.000	0.231	0.462	0.051	0.026	0.103	0.026	0.051
O&R	0.000	0.022	0.283	0.076	0.120	0.141	0.054	0.130	0.174
P&O	0.065	0.039	0.275	0.033	0.144	0.209	0.039	0.111	0.085
RD	0.056	0.000	0.278	0.000	0.111	0.167	0.056	0.222	0.111
S&S	0.031	0.041	0.227	0.026	0.206	0.113	0.046	0.247	0.062
T&T	0.024	0.006	0.213	0.024	0.116	0.116	0.024	0.104	0.372

Após todas as Matrizes de probabilidade terem sido estimadas, foi utilizado o teste de hipótese para *Cadeias de Markov* de [Kullback et al. \(1962\)](#) com intuito de calcular o grau de similaridade entre uma dada cidade e as demais. Onde, na equação 5.2, $i = 1, 2, \dots, s$ é o número de exemplos de CMTD que deseja comparar (número de cidades) e r é a cardinalidade do estado de estados (número de categorias). Esse teste, que, como mostrado na equação, é baseado em um teste Qui-quadrado, serve para verificar se as cadeias estimadas s foram geradas a partir de uma mesma CMTD.

$$2 * \sum_{i=1}^s \sum_{j=1}^r \sum_{k=1}^r f_{ijk} \ln \frac{n * f_{ijk}}{f_{i..} * f_{.jk}} \sim \chi^2 (r * (r - 1)) \tag{5.2}$$

Vale salientar que o teste foi aplicado separadamente para cada dia da semana, ou seja, a similaridade entre as cidades depende do dia da semana considerado.

Ademais, conforme mencionado na seção anterior, novamente não será aplicado verdadeiramente um teste de hipótese, sendo o *p-valor* utilizado como medida de verossimilhança entre as cidades.

Nas Tabelas 5.24 e 5.25 é possível ver os resultados do teste aplicado para comparar todas as cidades selecionadas à New York e São Paulo, respectivamente, por meio de matrizes de escala de cinza. Como o agrupamento é feito por dia da semana, existe variação na ordem de proximidade das cidades de acordo com o dia analisado. As matrizes de *p-valor* utilizadas para a construção das referidas tabelas se encontram no apêndice.

Tabela 5.24: Matriz de escala de cinza que representa os resultados da comparação entre Cadeias de Markov de diferentes cidades com New York

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Bangkok	Dark	Light	Light	Light	Light	Light	Light
Los Angeles	Dark	Light	Light	Light	Dark	Dark	Dark
Manila	Light	Light	Dark	Light	Light	Light	Light
Mexico City	Light	Light	Light	Light	Dark	Dark	Dark
Moscow	Light	Light	Light	Light	Light	Light	Light
Rio de Janeiro	Light	Light	Light	Light	Light	Light	Light
Recife	Light	Light	Light	Light	Light	Light	Light
St. Petersburg	Light	Light	Light	Light	Light	Light	Light
São Paulo	Light	Light	Light	Light	Light	Light	Light

Tal variação, por exemplo, pode ser percebida verificando os dias de segunda-feira - onde as três cidades mais próximas de New York são Los Angeles, Manila e Mexico City - e de quinta-feira, onde, apesar da primeira se manter a mesma, a segunda passa a ser São Paulo, com Mexico City como terceira e Manila apenas como a sexta mais próxima.

Essa constatação pode ser verificada de maneira ainda mais evidente caso se analise os resultados da cidade de São Paulo, onde Recife, apesar de ser claramente uma das cidades mais similares a São Paulo - figurando como a mais próxima em três ocasiões (segunda, terça

Tabela 5.25: Matriz de escala de cinza que representa os resultados da comparação entre Cadeias de Markov de diferentes cidades com São Paulo

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Bangkok							
Los Angeles							
Manila							
Mexico City							
Moscow							
New York							
Rio de Janeiro							
Recife							
St. Petersburg							

e domingo) e como a segunda mais próxima em outra (quarta) - aparece em sétimo lugar em duas ocasiões (quinta e sexta).

As tabelas 5.26 e 5.28 apresentam uma ordenação das três cidades mais parecidas a New York e São Paulo, respectivamente, em todos os dias da semana estudada. De maneira análoga, as tabelas 5.27 e 5.29 representam um ranqueamento das três cidades que mais diferem de New York e São Paulo.

Tabela 5.26: New York - Mais Similares

	1ª	2ª	3ª
Segunda	Los Angeles	Manila	Mexico City
Terça	Los Angeles	Mexico City	Manila
Quarta	Manila	Mexico City	Los Angeles
Quinta	Los Angeles	São Paulo	Mexico City
Sexta	Los Angeles	Mexico City	São Paulo
Sábado	Los Angeles	Mexico City	São Paulo
Domingo	Los Angeles	Mexico City	São Paulo

Tabela 5.27: New York - Menos Similares

	7ª	8ª	9ª
Segunda	Rio de Janeiro	Recife	Moscow
Terça	Rio de Janeiro	St. Petersburg	Recife
Quarta	Recife	Rio de Janeiro	St. Petersburg
Quinta	Bangkok	Recife	Rio de Janeiro
Sexta	Recife	Rio de Janeiro	Moscow
Sábado	Recife	Rio de Janeiro	Bangkok
Domingo	Rio de Janeiro	Recife	St. Petersburg

Tabela 5.28: São Paulo - Mais Similares

	1ª	2ª	3ª
Segunda	Recife	Mexico City	Manila
Terça	Recife	Moscow	Mexico City
Quarta	Moscow	Recife	Bangkok
Quinta	Rio de Janeiro	Manila	St. Petersburg
Sexta	Manila	Bangkok	Rio de Janeiro
Sábado	Mexico City	Moscow	St. Petersburg
Domingo	Recife	Manila	St. Petersburg

Tabela 5.29: São Paulo - Menos Similares

	7ª	8ª	9ª
Segunda	Los Angeles	New York	Bangkok
Terça	Los Angeles	New York	Bangkok
Quarta	Los Angeles	New York	Mexico City
Quinta	Bangkok	Recife	Moscow
Sexta	Recife	New York	St. Petersburg
Sábado	Bangkok	Recife	Manila
Domingo	Bangkok	New York	Los Angeles

Estudando os resultados, fica evidente que são compatíveis com os apresentados na seção anterior (5.2), no entanto, enquanto esse resultado é bastante perceptível quando verificado em New York (tabelas 5.26 e 5.27) - tendo Los Angeles e Mexico City aparecendo entre as cidades mais similares em quase todos os dias, tanto na análise separada por categorias, quanto na em conjunto - na cidade de São Paulo (tabelas 5.28 e 5.29) não seria tão fácil sugerir quais cidades obtiveram valores mais similares apenas verificando os testes separados por categorias. Inclusive, seria bem provável a sugestão de resultados incorretos, como a

indicação do Los Angeles como uma das cidades próximas a São Paulo, já que aparece em duas categorias (*P&O* e *A&E*) dentre as três cidades mais similares, mas, na verdade, ao analisar todas as categorias em conjunto é possível notar que a cidade tem um padrão de mobilidade social bastante diferente do de São Paulo.

5.3.1 Comparação das Cidades Por Turno

Até o momento foi realizada comparação da dinâmica social urbana diária, no entanto, é plausível supor que pessoas se comportem de maneira diferente de acordo com o horário do dia. Tal suposição se confirma se verificarmos, por exemplo, os histogramas apresentados na Imagem 5.1, onde, conforme a hora do dia, há uma forte oscilação na frequência de *check-ins*.

Para atestar essa diferença de maneira mais clara, primeiramente foram separados os *check-ins* por turno, manhã (04:00-12:00), tarde (12:00-18:00) e noite (18:00-04:00), para, então, como no exemplo anterior, estimar suas *CMTD*.

Executando o teste para *CMTD* (Equação 5.2), a fim de comparar os diferentes períodos do dia, é possível perceber, verificando os resultados nas Tabelas 5.30 e 5.31, que há uma clara diferença entre eles, principalmente em relação aos turnos matutino e noturno, como se pode ser notado pela incidência de tons da cor vermelha.

Já o turno vespertino, talvez por estar inserido entre os outros dois, obteve resultados um pouco melhores na comparação com os demais, principalmente com o turno matutino o que se pode perceber pela maior incidência de tons da cor azul.

Tabela 5.30: Matriz de escala de vermelho e azul que representa os resultados da comparação entre Cadeias de Markov de diferentes turnos da da cidade de New York

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Manhã x Tarde							
Manhã x Noite							
Tarde x Noite							

Tabela 5.31: Matriz de escala de vermelho e azul que representa os resultados da comparação entre Cadeias de Markov de diferentes turnos da da cidade de São Paulo

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Manhã x Tarde							
Manhã x Noite							
Tarde x Noite							

Após mostrar que os períodos do dia apresentam comportamentos distintos, com intuito de investigar se as cidades mais similares a *NY* se manteriam, não importando o turno, foi novamente utilizado o teste de *CMTD* para comparar turno a turno a cidade com as demais.

Os resultados de tal comparação são apresentados nas Tabelas 5.32, 5.33 e 5.34, os resultados de *p-valor*, por sua vez, estão nas Tabelas A.25, A.26 e A.27 localizadas no apêndice.

Tabela 5.32: Matriz de escala de cinza que representa os resultados da comparação entre Cadeias de Markov de diferentes cidades com New York no turno matutino

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Bangkok							
Los Angeles							
Manila							
Mexico City							
Moscow							
Recife							
Rio de Janeiro							
St. Petersburg							
São Paulo							

Tabela 5.33: Matriz de escala de cinza que representa os resultados da comparação entre Cadeias de Markov de diferentes cidades com New York no turno vespertino

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Bangkok							
Los Angeles							
Manila							
Mexico City							
Moscow							
Recife							
Rio de Janeiro							
St. Petersburg							
São Paulo							

Tabela 5.34: Matriz de escala de cinza que representa os resultados da comparação entre Cadeias de Markov de diferentes cidades com New York no turno noturno

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Bangkok							
Los Angeles							
Manila							
Mexico City							
Moscow							
Recife							
Rio de Janeiro							
St. Petersburg							
São Paulo							

Da análise das Tabelas 5.32, 5.33, 5.34, extrai-se alguns resultados interessantes:

1. Manhã: Apesar de Los Angeles e Mexico City serem claramente as cidades mais próximas a New York, quando analisada amostras diárias (Tabela XIX), levando em consideração só o turno matutino, São Paulo se destaca, aparecendo entre as três mais similares em 6 dos 7 dias analisados.
2. Tarde: No entanto, em relação ao período vespertino, Los Angeles e Mexico City voltam a dominar a lista das três cidades mais similares, com 6 aparições da primeira e 5 da segunda. São Paulo não aparece nenhuma vez na lista, ficando, inclusive, entre as três menos semelhantes em dois dias da semana, terça (9^a) e quinta (7^a).
3. Noite: Já no período noturno, fica claro o porquê de Los Angeles e Mexico City serem tão semelhantes a New York, já que mais uma vez são as cidades que mais aparecem na lista das três mais similares. No entanto, diferente do período da tarde, há um destaque bem maior para Mexico City, estando entre as cidades mais similares todos os dias da semana, sendo, inclusive, a mais equivalente em três deles, quais sejam, terça, quinta e sexta.

A partir dos resultados dessa comparação foi possível perceber que existe uma grande equivalência com os resultados da análise diária. No entanto, alguns resultados trazem informações que passariam despercebidos se não fosse a separação por turnos, como o caso de São Paulo, que figura na lista das cidades mais semelhantes em análises diárias, mas, quando comparada por turno, só apresenta resultados muito similares no período da manhã.

Tendo em vista que ao se comparar cidades foi criado um grau de proximidade entre elas, que é expresso pelo *p-valor*, um caminho evidente a seguir é agrupar tais cidades por este valor, sendo um exemplo de aplicação para os resultados desse processo apresentado na seção 5.4. No entanto, é importante salientar que tais resultados foram obtidos analisando uma amostra de dados associada a um período de tempo e esses podem, e devem, sofrer alterações caso sejam utilizados dados de outra faixa temporal. Como explicado anteriormente, tais alterações podem ser fruto de inúmeras causas, então deve-se entender esta análise como uma representação das cidades associadas a um período de tempo, como um *snapshot* das cidades.

5.4 Exemplo de Aplicação

Inúmeras aplicações podem se beneficiar do agrupamento de cidades. Dependendo da aplicação, pode ser interessante investigar o grau de semelhança entre elas. Por exemplo, a utilização da técnica de análise proposta pode ser utilizada como um engenho para sistemas de recomendação voltados ao turismo. Nesse caso, o estudo pode ser realizado de duas

formas diferentes: agrupando as cidades por algumas categorias específicas ou por todas as categorias em conjunto.

Por exemplo, dado que um usuário já tenha viajado para New York e gostado da cidade em razão dos seus museus, praças históricas e locais relacionados à cultura, seria possível gerar um agrupamento de cidades semelhantes à New York nesses aspectos. Para isso, far-se-ia necessário, primeiro, verificar quais os dias da semana que o usuário pretendia passar no local, e, então, realizar um processo similar ao feito na seção 5.2, utilizando os *p-valores* resultantes do teste Qui-quadrado para as categorias *Arts & Entertainment* e *Outdoors & Recreation* naqueles determinados dias, e, a partir daí, agrupar o resultado utilizando como parâmetro os valores obtidos, conforme realizado no dendrograma da Imagem 5.2.

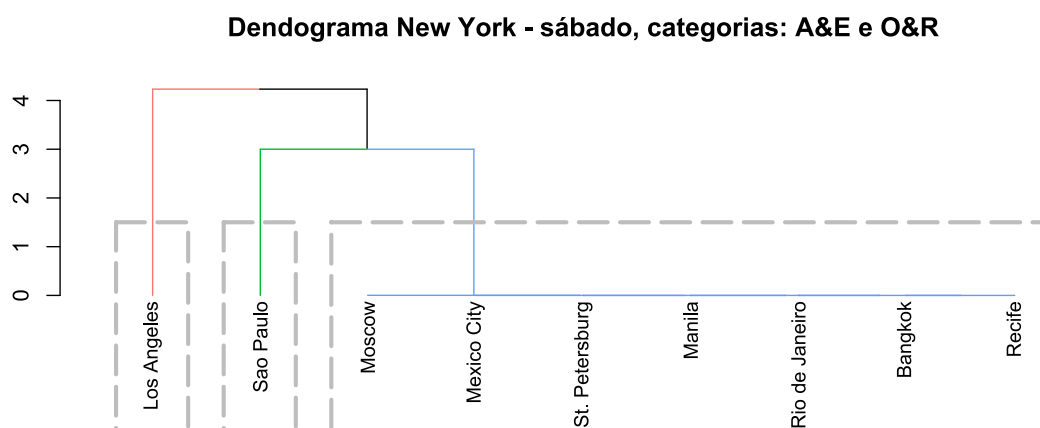


Figura 5.2: Agrupamento Hierárquico (método Centróide), exemplo de divisão em 3 grupos, usando *p-valor* do teste Qui-quadrado como parâmetro de agrupamento.

Nesse exemplo, caso o usuário pretendesse viajar em um dia de sábado, seria possível sugerir a visita à cidade de Los Angeles, pois, de acordo com os resultados visíveis na Imagem 5.2, é a que mais se aproxima de New York em relação às categorias ligadas às atividades preferidas pelo usuário, sendo a cidade de São Paulo também uma boa opção, ocupando o segundo lugar no ranking.

De outro modo, se o usuário desejasse descobrir, dentre um rol de cidades, aquela que possui, em conjunto, todas as categorias semelhantes as da cidade de New York, bastaria, após selecionar os dias da viagem, fazer uma comparação entre as *Cadeias de Markov* das cidades naqueles dias da semana específicos, como as realizadas na seção 5.3.

É possível ver essa segunda abordagem aplicada à cidade de New York, nos dias de sexta e sábado, na Imagem 5.3, da qual se depreende que, mais uma vez, a cidade de Los Angeles seria a mais indicada para o usuário, seguida da cidade de Mexico City.

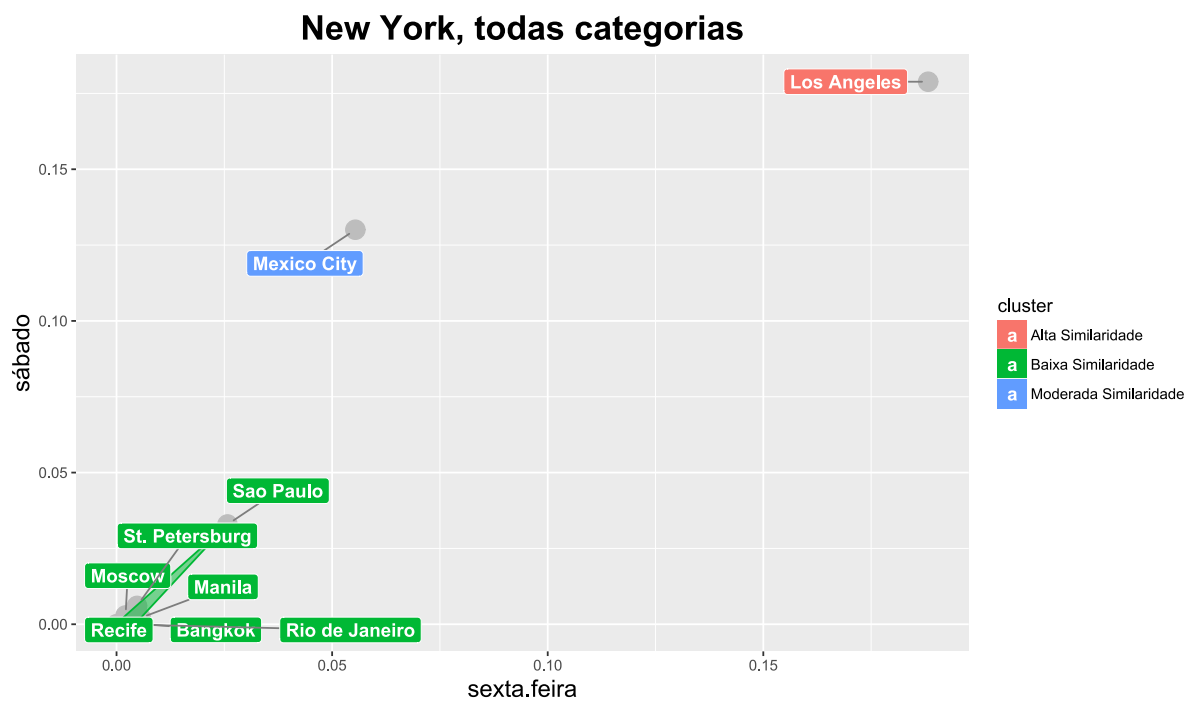


Figura 5.3: Agrupamento hierárquico (método Centróide), exemplo de divisão em 3 grupos, usando o p -valor do teste para Cadeias de Markov como parâmetro de agrupamento.

6

CONCLUSÕES E TRABALHOS FUTUROS

COM a popularização do uso de dispositivos móveis e o crescimento da cobertura do sinal de Internet, as *RSPs* têm se tornado mais uma fonte de dados inestimável para o estudo populacional.

Análises que fazem uso dessa fonte de dados, como as apresentadas neste trabalho, vem se mostrando opções bastante promissoras de pesquisa, por atingir resultados significantes com um baixo custo de desenvolvimento, curto tempo para realização e, principalmente, pela possibilidade de serem efetuadas em larga escala.

Em especial, este trabalho se aproveita da dinamicidade dos dados de *RSPs* para o estudo da dinâmica social urbana. Mais especificamente, após responder algumas questões sobre a natureza dos dados utilizados, compara diversas cidades - a partir de métricas extraídas do comportamento social da população - e, então, as agrupa por fatores de semelhança.

Por meio da análise dos resultados obtidos, não só foi possível saber como funciona a dinâmica da população de uma cidade em relação a um tipo de ambiente social específico (como locais de vida noturna), mas também descobrir quais cidades têm o comportamento mais semelhante ao dela em relação a esse ambiente. Além disso, mediante uma modelagem que faz uso de *Cadeias de Markov* para representação dos dados, foi possível, também, descobrir quais cidades possuem o comportamento mais semelhante considerando todos os tipos de ambientes de forma associada.

No entanto, muito além dos resultados obtidos, é apresentada uma nova maneira de enxergar a dinâmica social das cidades, que tem potencial de ser explorada em diversas áreas, como turismo, investimentos, governança pública e até controle epidemiológico.

Como continuação deste trabalho, é possível listar a criação de uma ferramenta de análise contínua, que realize a captação dos dados de forma constante, a fim de oferecer resultados mais confiáveis e em tempo real. Por meio de tal ferramenta, será possível retratar a evolução da dinâmica social das cidades ao longo do tempo, e, até mesmo, prever tendências futuras de comportamento das cidades.

Outro ponto que deve ser estudado é a possibilidade de classificar dados de outros *SSP's* por categorias do *Foursquare*, tendo, assim, um conjunto de dados mais abrangente e confiável.

Com um volume maior de dados, será possível, ainda, estudar de maneira mais específica a dinâmica social da população. Para isso, serão exploradas as transições entre as subcategorias do *Foursquare*, para se ter resultados mais fidedignos nas comparações entre as cidades.

Por fim, será averiguada a viabilidade do desenvolvimento da aplicação proposta na Seção 5.4 como uma plataforma web de recomendação de locais com base nos lugares anteriormente visitados. Tal aplicação teria como objetivo não só oferecer uma comparação dos estados atuais das cidades, como também possibilitar o uso do histórico de resultados, tornando factível a comparação entre cidades em intervalos de tempo específicos (como, por exemplo, o período de visita do usuário ao local).

Apêndice A

Resultados Numéricos dos Testes de Hipótese

A.1 Representação numérica do grau de similaridade entre diferentes amostras de uma mesma cidade

Tabela A.1: Matrizes com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência diferentes dias de uma mesma semana da cidade de New York

NY: Arts & Entertainment

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Segunda	1	4.174e-02	2.007e-02	1.864e-03	8.785e-04	3.671e-08	1.166e-03
Terça	4.174e-02	1	1.650e-02	4.499e-02	3.622e-04	9.509e-07	3.808e-07
Quarta	2.007e-02	1.650e-02	1	4.351e-02	2.404e-02	1.640e-02	1.745e-03
Quinta	1.864e-03	4.499e-02	4.351e-02	1	7.654e-02	5.010e-11	3.468e-05
Sexta	8.785e-04	3.622e-04	2.404e-02	7.654e-02	1	2.735e-05	1.838e-06
Sábado	3.671e-08	9.509e-07	1.640e-02	5.010e-11	2.735e-05	1	1.408e-04
Domingo	1.166e-03	3.808e-07	1.745e-03	3.468e-05	1.838e-06	1.408e-04	1

NY: College & University

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Segunda	1	1.325e-01	1.592e-01	2.518e-02	8.763e-01	1.777e-01	8.973e-01
Terça	1.325e-01	1	3.968e-01	6.254e-01	8.302e-01	5.578e-01	2.636e-01
Quarta	1.592e-01	3.968e-01	1	5.749e-01	3.743e-01	7.847e-01	8.973e-01
Quinta	2.518e-02	6.254e-01	5.749e-01	1	4.470e-02	7.952e-01	2.433e-01
Sexta	8.763e-01	8.302e-01	3.743e-01	4.470e-02	1	4.232e-01	7.518e-01
Sábado	1.777e-01	5.578e-01	7.847e-01	7.952e-01	4.232e-01	1	7.518e-01
Domingo	8.973e-01	2.636e-01	8.973e-01	2.433e-01	7.518e-01	7.518e-01	1

NY: Food

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Segunda	1	7.972e-04	6.602e-13	6.601e-20	1.523e-18	5.714e-47	5.848e-11
Terça	7.972e-04	1	9.052e-04	3.381e-11	3.481e-07	2.415e-21	5.306e-05
Quarta	6.602e-13	9.052e-04	1	1.256e-13	2.842e-12	4.191e-35	3.331e-13
Quinta	6.601e-20	3.381e-11	1.256e-13	1	2.266e-07	1.820e-31	8.589e-11
Sexta	1.523e-18	3.481e-07	2.842e-12	2.266e-07	1	5.043e-04	1.669e-05
Sábado	5.714e-47	2.415e-21	4.191e-35	1.820e-31	5.043e-04	1	1.252e-08
Domingo	5.848e-11	5.306e-05	3.331e-13	8.589e-11	1.669e-05	1.252e-08	1

NY: Nightlife Spot

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Segunda	1	6.557e-01	7.122e-01	3.132e-01	1.623e-02	6.435e-08	2.659e-06
Terça	6.557e-01	1	7.180e-01	4.861e-01	3.620e-02	1.351e-05	5.078e-07
Quarta	7.122e-01	7.180e-01	1	2.616e-01	3.299e-02	3.254e-05	2.327e-05
Quinta	3.132e-01	4.861e-01	2.616e-01	1	5.802e-01	1.061e-07	4.072e-04
Sexta	1.623e-02	3.620e-02	3.299e-02	5.802e-01	1	7.615e-04	1.077e-01
Sábado	6.435e-08	1.351e-05	3.254e-05	1.061e-07	7.615e-04	1	2.155e-02
Domingo	2.659e-06	5.078e-07	2.327e-05	4.072e-04	1.077e-01	2.155e-02	1

NY: Outdoors & Recreation

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Segunda	1	6.735e-02	7.499e-02	1.706e-02	1.471e-01	3.667e-06	1.766e-05
Terça	6.735e-02	1	6.769e-01	8.912e-02	5.507e-02	2.270e-05	7.171e-05
Quarta	7.499e-02	6.769e-01	1	2.887e-02	2.375e-02	9.105e-07	2.044e-07
Quinta	1.706e-02	8.912e-02	2.887e-02	1	5.958e-01	8.188e-03	1.452e-03
Sexta	1.471e-01	5.507e-02	2.375e-02	5.958e-01	1	4.180e-02	6.020e-03
Sábado	3.667e-06	2.270e-05	9.105e-07	8.188e-03	4.180e-02	1	2.487e-05
Domingo	1.766e-05	7.171e-05	2.044e-07	1.452e-03	6.020e-03	2.487e-05	1

NY: Professional & Other Places

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Segunda	1	3.162e-01	2.814e-03	7.869e-02	4.240e-02	2.288e-05	2.269e-01
Terça	3.162e-01	1	3.001e-01	8.480e-01	9.817e-01	9.402e-01	1.528e-01
Quarta	2.814e-03	3.001e-01	1	7.929e-01	6.713e-01	3.349e-01	1.134e-01
Quinta	7.869e-02	8.480e-01	7.929e-01	1	5.464e-01	1.918e-01	7.875e-02
Sexta	4.240e-02	9.817e-01	6.713e-01	5.464e-01	1	6.327e-01	2.445e-01
Sábado	2.288e-05	9.402e-01	3.349e-01	1.918e-01	6.327e-01	1	1.733e-02
Domingo	2.269e-01	1.528e-01	1.134e-01	7.875e-02	2.445e-01	1.733e-02	1

Tabela A.3: Matrizes com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência diferentes dias de uma mesma semana de outras cidades

Moscow: *College & University*

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Segunda	1	7.923e-01	2.845e-01	8.136e-01	9.662e-01	6.485e-01	1.808e-02
Terça	7.923e-01	1	1.624e-01	7.435e-01	9.192e-01	9.532e-01	2.704e-02
Quarta	2.845e-01	1.624e-01	1	9.082e-01	5.395e-01	5.754e-01	4.531e-03
Quinta	8.136e-01	7.435e-01	9.082e-01	1	4.779e-01	2.725e-01	3.209e-01
Sexta	9.662e-01	9.192e-01	5.395e-01	4.779e-01	1	4.795e-01	7.255e-01
Sábado	6.485e-01	9.532e-01	5.754e-01	2.725e-01	4.795e-01	1	3.103e-01
Domingo	1.808e-02	2.704e-02	4.531e-03	3.209e-01	7.255e-01	3.103e-01	1

Manila: *College & University*

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Segunda	1	7.047e-03	2.217e-01	1.726e-01	2.147e-02	1.309e-04	2.317e-02
Terça	7.047e-03	1	8.843e-03	5.455e-02	1.043e-01	2.207e-02	1.090e-02
Quarta	2.217e-01	8.843e-03	1	3.709e-01	2.940e-02	4.936e-08	1.847e-01
Quinta	1.726e-01	5.455e-02	3.709e-01	1	5.823e-01	3.214e-04	1.503e-02
Sexta	2.147e-02	1.043e-01	2.940e-02	5.823e-01	1	1.954e-02	1.847e-01
Sábado	1.309e-04	2.207e-02	4.936e-08	3.214e-04	1.954e-02	1	1.550e-04
Domingo	2.317e-02	1.090e-02	1.847e-01	1.503e-02	1.847e-01	1.550e-04	1

Rio de Janeiro: *Nightlife Spot*

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Segunda	1	8.416e-01	6.858e-03	3.899e-01	1.902e-02	1.518e-09	2.545e-06
Terça	8.416e-01	1	1.234e-04	3.933e-01	8.231e-02	3.892e-11	1.405e-07
Quarta	6.858e-03	1.234e-04	1	1.979e-05	2.235e-06	3.595e-18	1.871e-07
Quinta	3.899e-01	3.933e-01	1.979e-05	1	6.037e-01	3.349e-08	9.742e-04
Sexta	1.902e-02	8.231e-02	2.235e-06	6.037e-01	1	9.576e-07	2.455e-04
Sábado	1.518e-09	3.892e-11	3.595e-18	3.349e-08	9.576e-07	1	1.253e-01
Domingo	2.545e-06	1.405e-07	1.871e-07	9.742e-04	2.455e-04	1.253e-01	1

Los Angeles: *Nightlife Spot*

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Segunda	1	9.822e-02	3.992e-01	4.746e-01	8.388e-04	7.398e-02	6.592e-02
Terça	9.822e-02	1	9.886e-03	6.068e-01	4.233e-02	5.551e-05	2.283e-01
Quarta	3.992e-01	9.886e-03	1	1.351e-02	4.835e-07	1.737e-02	5.180e-02
Quinta	4.746e-01	6.068e-01	1.351e-02	1	2.837e-02	1.284e-01	7.659e-02
Sexta	8.388e-04	4.233e-02	4.835e-07	2.837e-02	1	6.125e-05	5.135e-03
Sábado	7.398e-02	5.551e-05	1.737e-02	1.284e-01	6.125e-05	1	2.621e-02
Domingo	6.592e-02	2.283e-01	5.180e-02	7.659e-02	5.135e-03	2.621e-02	1

Manila: *Travel & Transport*

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Segunda	1	5.666e-01	1.918e-02	1.438e-01	1.597e-02	9.827e-04	3.015e-07
Terça	5.666e-01	1	1.573e-02	2.051e-02	9.620e-03	7.809e-04	1.182e-09
Quarta	1.918e-02	1.573e-02	1	3.204e-02	2.345e-07	1.326e-09	1.349e-11
Quinta	1.438e-01	2.051e-02	3.204e-02	1	1.126e-02	1.905e-03	5.524e-02
Sexta	1.597e-02	9.620e-03	2.345e-07	1.126e-02	1	7.410e-05	4.052e-06
Sábado	9.827e-04	7.809e-04	1.326e-09	1.905e-03	7.410e-05	1	1.448e-01
Domingo	3.015e-07	1.182e-09	1.349e-11	5.524e-02	4.052e-06	1.448e-01	1

Moscow: *Travel & Transport*

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Segunda	1	8.501e-01	4.721e-01	2.464e-02	1.712e-01	5.314e-24	1.386e-07
Terça	8.501e-01	1	5.857e-01	7.840e-03	4.132e-01	1.552e-22	5.017e-08
Quarta	4.721e-01	5.857e-01	1	2.862e-01	9.573e-01	4.508e-11	1.190e-02
Quinta	2.464e-02	7.840e-03	2.862e-01	1	2.714e-01	1.351e-03	1.056e-01
Sexta	1.712e-01	4.132e-01	9.573e-01	2.714e-01	1	9.063e-13	7.900e-03
Sábado	5.314e-24	1.552e-22	4.508e-11	1.351e-03	9.063e-13	1	3.415e-04
Domingo	1.386e-07	5.017e-08	1.190e-02	1.056e-01	7.900e-03	3.415e-04	1

Mexico City: *Residence*

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Segunda	1	3.556e-04	2.150e-02	2.813e-07	7.058e-03	2.005e-04	1.122e-09
Terça	3.556e-04	1	4.149e-05	4.168e-04	9.901e-04	4.473e-14	9.383e-18
Quarta	2.150e-02	4.149e-05	1	4.672e-04	4.387e-01	1.216e-05	3.297e-04
Quinta	2.813e-07	4.168e-04	4.672e-04	1	1.168e-02	1.830e-04	6.862e-06
Sexta	7.058e-03	9.901e-04	4.387e-01	1.168e-02	1	2.221e-02	9.577e-05
Sábado	2.005e-04	4.473e-14	1.216e-05	1.830e-04	2.221e-02	1	1.440e-04
Domingo	1.122e-09	9.383e-18	3.297e-04	6.862e-06	9.577e-05	1.440e-04	1

Bangkok: *Residence*

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Segunda	1	1.016e-01	3.521e-01	3.086e-02	9.824e-02	5.261e-01	4.694e-03
Terça	1.016e-01	1	1.256e-02	1.402e-03	3.224e-01	5.765e-01	1.725e-02
Quarta	3.521e-01	1.256e-02	1	5.120e-02	1.893e-02	5.388e-01	4.352e-03
Quinta	3.086e-02	1.402e-03	5.120e-02	1	1.938e-01	6.708e-02	5.928e-01
Sexta	9.824e-02	3.224e-01	1.893e-02	1.938e-01	1	1.707e-01	8.291e-01
Sábado	5.261e-01	5.765e-01	5.388e-01	6.708e-02	1.707e-01	1	6.474e-02
Domingo	4.694e-03	1.725e-02	4.352e-03	5.928e-01	8.291e-01	6.474e-02	1

Tabela A.4: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência diferentes dias um mesmo dia de diferentes semanas da cidade de Los Angeles

Semana 1 (S1) x Semana 2 (S2)

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
A&E	2.746e-01	2.767e-02	6.290e-02	8.488e-01	1.686e-02	6.284e-02	7.303e-02
C&U	9.460e-01	3.307e-01	2.564e-01	4.895e-01	6.149e-01		5.637e-01
FD	1.337e-02	7.026e-11	1.164e-02	5.326e-02	2.092e-04	2.243e-01	1.987e-01
NL	8.052e-01	3.509e-01	3.813e-01	6.873e-02	1.340e-05	5.106e-01	1.178e-04
O&R	9.263e-01	6.604e-02	1.112e-03	9.430e-02	2.668e-05	5.531e-01	2.172e-02
P&O	3.705e-02	5.248e-01	9.649e-01	4.444e-01	7.102e-01	5.886e-01	6.547e-01
RD	7.893e-01	6.948e-02	6.881e-01	7.745e-02	2.865e-01	8.614e-01	6.939e-01
S&S	4.326e-01	7.665e-04	7.653e-02	8.367e-07	1.723e-03	5.736e-03	7.513e-02
T&T	5.439e-02	4.869e-01	6.497e-01	2.383e-02	3.102e-04	2.033e-02	1.609e-01

Semana 2 (S2) x Semana 3 (S3)

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
A&E	9.665e-01	8.898e-03	6.426e-02	9.067e-01	5.600e-01	1.486e-02	3.486e-02
C&U	3.719e-02	8.825e-01	3.173e-01	5.153e-01	8.254e-01		
FD	2.716e-01	8.048e-04	5.207e-05	3.147e-03	2.157e-05	3.167e-01	2.675e-01
NL	7.220e-01	7.746e-02	7.766e-01	1.798e-05	1.166e-01	7.420e-05	1.311e-01
O&R	1.083e-01	1.110e-01	6.540e-05	1.668e-01	7.582e-01	1.144e-02	3.320e-02
P&O	5.878e-02	6.473e-01	1.857e-02	9.796e-06	7.003e-01	5.866e-01	7.237e-01
RD	7.912e-01	8.816e-02	8.326e-02	3.679e-01	8.486e-02	7.578e-01	7.710e-02
S&S	9.013e-01	2.032e-10	4.455e-02	3.967e-01	5.210e-01	2.006e-02	7.809e-02
T&T	7.304e-01	1.958e-04	1.869e-02	3.355e-02	3.522e-04	3.886e-01	4.099e-01

Tabela A.5: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência diferentes dias um mesmo dia de diferentes semanas da cidade de New York

Semana 1 (S1) x Semana 2 (S2)								Semana 2 (S2) x Semana 3 (S3)							
	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo		Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
A&E	1.874e-03	9.940e-06	5.635e-01	3.084e-01	7.826e-13	6.232e-11	1.512e-01	A&E	2.372e-03	3.159e-01	8.625e-02	1.468e-07	3.204e-03	2.555e-02	4.463e-06
C&U	5.702e-01	4.739e-01	9.188e-01	1.330e-01	3.987e-02	5.469e-01	2.636e-01	C&U	1.713e-02	2.771e-01	1.043e-01	9.360e-01	9.454e-01	1.000e+00	1.797e-01
FD	2.751e-07	5.406e-01	1.373e-02	8.171e-09	1.873e-05	5.759e-03	3.358e-04	FD	2.287e-02	3.460e-06	8.562e-02	1.246e-03	6.703e-04	1.193e-04	2.068e-04
NL	3.642e-02	7.878e-08	2.038e-01	9.678e-01	1.044e-02	1.401e-03	2.102e-01	NL	1.193e-01	2.018e-03	7.371e-06	3.646e-01	1.522e-04	1.226e-02	1.323e-01
O&R	6.312e-02	2.645e-01	5.538e-02	2.326e-02	4.076e-03	3.241e-02	2.595e-03	O&R	1.172e-03	3.587e-03	8.664e-02	5.138e-04	8.479e-01	9.791e-16	4.776e-02
P&O	1.037e-01	8.261e-03	9.525e-02	2.538e-01	2.920e-01	4.621e-01	7.891e-01	P&O	8.005e-03	1.933e-04	8.844e-01	2.799e-01	3.229e-01	3.796e-05	4.116e-01
RD	6.715e-04	9.317e-03	7.333e-04	2.494e-03	1.677e-01	4.298e-02	3.138e-01	RD	5.434e-02	4.844e-01	7.426e-06	2.123e-01	2.889e-01	1.479e-01	8.049e-04
S&S	3.715e-06	2.092e-02	6.907e-02	1.135e-04	5.912e-07	9.913e-02	1.461e-01	S&S	6.007e-09	3.309e-01	4.594e-03	3.730e-03	2.900e-05	1.399e-12	1.878e-01
T&T	1.199e-02	4.188e-02	8.950e-02	2.460e-02	1.845e-01	7.304e-01	6.635e-19	T&T	4.312e-10	9.284e-07	1.532e-06	7.725e-07	4.001e-03	2.461e-03	1.452e-02

Tabela A.6: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência diferentes dias um mesmo dia de diferentes semanas da cidade de São Paulo

Semana 1 (S1) x Semana 2 (S2)								Semana 2 (S2) x Semana 3 (S3)							
	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo		Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
A&E	1.764e-09	5.940e-01	5.766e-03	5.164e-01	1.619e-05	3.040e-06	7.223e-07	A&E	2.476e-01	9.588e-04	3.044e-02	8.527e-08	2.132e-02	3.723e-04	2.929e-01
C&U	1.197e-04	6.109e-03	1.985e-08	4.185e-01	3.605e-01	3.339e-11		C&U	9.401e-08	7.354e-07	5.489e-13	4.828e-05	5.885e-09	2.024e-02	
FD	9.889e-02	1.436e-13	4.924e-08	2.235e-10	2.323e-04	5.571e-01	8.463e-14	FD	2.587e-16	1.102e-06	1.681e-35	2.414e-43	3.091e-04	2.120e-01	7.023e-06
NL	6.706e-01	1.247e-02	1.700e-02	1.331e-03	4.661e-08	4.895e-02	2.471e-03	NL	1.956e-02	6.007e-02	3.239e-01	3.667e-10	3.110e-02	1.599e-01	1.022e-02
O&R	1.712e-03	2.371e-04	7.562e-04	6.208e-02	8.859e-04	1.794e-02	8.077e-01	O&R	2.441e-04	2.079e-04	2.768e-08	1.089e-41	2.483e-08	4.855e-04	2.233e-01
P&O	9.811e-01	4.808e-01	3.428e-10	5.866e-02	6.301e-01	2.332e-01	3.232e-01	P&O	5.637e-01	2.640e-01	7.333e-15	1.554e-01	4.859e-01	5.653e-01	7.424e-02
RD	1.023e-11	1.316e-02	7.299e-10	8.825e-16	3.427e-01	5.303e-02	5.111e-01	RD	2.929e-09	3.568e-01	7.850e-05	1.257e-43	2.295e-02	1.969e-02	3.409e-02
S&S	1.706e-04	1.747e-03	2.870e-03	3.905e-08	2.457e-02	8.040e-03	1.175e-01	S&S	4.069e-02	3.646e-04	8.073e-04	4.228e-05	7.758e-02	7.327e-02	1.470e-07
T&T	1.952e-04	7.755e-04	1.700e-16	1.150e-01	1.901e-03	1.139e-02	1.830e-01	T&T	3.549e-01	1.869e-03	1.042e-20	1.092e-09	2.025e-02	9.008e-24	5.648e-01

A.2 Representação numérica do grau de similaridade de amostras de diferentes cidades

A.2.1 Representação por meio dos *p-valores* de testes Qui-quadrado

Tabela A.7: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de segunda-feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	4.128e-03	1.182e-20	1.276e-04	8.595e-04	2.505e-14	1.869e-04	1.234e-05	1.805e-16	2.051e-12
Los Angeles	8.016e-04	1.347e-04	5.219e-17	3.624e-02	2.909e-06	5.544e-06	9.377e-03	2.370e-15	1.572e-29
Manila	9.874e-02	6.360e-15	1.898e-05	8.316e-01	7.100e-08	2.981e-12	8.317e-08	1.358e-37	3.051e-11
Mexico City	3.056e-02	6.841e-11	2.084e-02	5.587e-01	4.583e-02	6.971e-10	7.105e-04	1.948e-01	4.341e-12
Moscow	3.655e-01	2.586e-01	1.400e-08	9.811e-01	1.518e-02	6.168e-06	7.067e-08	1.034e-04	3.874e-05
New York	5.860e-03	4.498e-23	1.215e-03	5.015e-02	1.353e-03	4.746e-08	2.465e-12	2.638e-05	1.628e-52
Recife	3.284e-01	2.577e-03	6.002e-10	4.724e-01	3.425e-13	6.615e-07	1.313e-01	6.673e-10	3.296e-07
Rio de Janeiro	3.181e-01	4.477e-02	2.172e-14	5.637e-01	7.860e-02	1.944e-09	1.120e-02	2.165e-07	6.766e-28
St. Petersburg	1.557e-04	1.218e-07	2.412e-15	3.458e-01	8.977e-03	6.903e-07	2.613e-05	6.459e-10	4.222e-07

Tabela A.8: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de terça-feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	2.191e-01	4.834e-37	5.592e-08	4.302e-03	2.263e-38	5.133e-10	1.833e-07	3.138e-20	9.464e-07
Los Angeles	4.189e-01	1.012e-16	1.748e-20	6.110e-02	2.770e-06	9.098e-09	3.543e-02	5.448e-09	3.156e-06
Manila	8.465e-08	8.935e-11	5.619e-13	3.189e-01	1.067e-18	2.424e-11	1.639e-15	2.789e-41	1.420e-10
Mexico City	5.141e-02	4.047e-19	3.871e-05	4.079e-02	1.280e-03	3.514e-04	1.279e-09	1.753e-01	8.206e-08
Moscow	1.822e-01	6.274e-04	8.940e-05	3.511e-01	9.666e-07	2.302e-08	3.987e-05	1.869e-10	1.172e-01
New York	2.134e-02	7.837e-28	1.489e-12	1.718e-07	5.693e-07	1.479e-10	1.040e-02	2.031e-09	1.684e-05
Recife	2.445e-02	3.530e-06	2.919e-06	1.708e-02	4.364e-04	2.059e-14	4.260e-01	9.406e-10	9.217e-08
Rio de Janeiro	6.154e-02	6.341e-02	1.034e-07	3.547e-01	1.231e-04	1.173e-10	1.737e-02	6.235e-07	2.159e-17
St. Petersburg	7.087e-03	7.590e-09	1.247e-09	1.426e-01	2.496e-03	3.182e-11	9.318e-05	2.111e-08	1.467e-08

Tabela A.9: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de quarta-feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	4.778e-03	6.021e-44	4.594e-08	9.588e-02	4.176e-04	3.385e-12	2.448e-03	3.000e-13	6.716e-14
Los Angeles	9.333e-01	1.678e-15	1.076e-22	4.309e-01	6.610e-05	6.035e-10	2.321e-03	1.384e-13	1.509e-14
Manila	4.116e-02	3.092e-16	6.570e-09	5.113e-03	1.744e-07	3.454e-15	3.468e-09	5.203e-13	1.030e-18
Mexico City	1.049e-05	1.182e-15	4.978e-15	1.296e-04	7.456e-04	4.199e-26	4.608e-02	2.346e-10	3.729e-03
Moscow	2.726e-01	6.343e-04	1.485e-06	6.588e-01	3.254e-01	8.255e-15	1.288e-03	6.291e-04	1.405e-10
New York	4.737e-02	8.473e-05	1.628e-25	1.496e-10	4.422e-02	7.539e-09	9.090e-23	2.766e-10	9.092e-15
Recife	3.039e-01	3.821e-11	1.736e-05	1.293e-01	1.526e-03	1.208e-22	3.003e-02	1.112e-05	1.317e-10
Rio de Janeiro	2.788e-02	7.394e-10	5.785e-15	1.519e-01	8.604e-02	6.763e-12	3.585e-01	7.224e-08	3.728e-17
St. Petersburg	7.516e-01	4.013e-04	3.837e-23	1.258e-01	9.106e-02	6.815e-13	1.911e-04	5.452e-07	5.164e-12

Tabela A.10: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de quinta-feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	8.324e-31	7.974e-02	2.301e-20	3.220e-05	1.148e-27	2.045e-03	2.275e-07	2.137e-11	2.118e-09
Los Angeles	2.698e-08	6.171e-01	3.505e-08	2.154e-01	2.398e-03	2.106e-03	7.031e-09	4.261e-08	4.272e-07
Manila	7.918e-10	1.000e+00	7.174e-09	9.904e-07	1.314e-06	3.343e-05	2.427e-08	5.681e-06	5.267e-03
Mexico City	6.214e-03	2.636e-01	7.486e-20	1.894e-01	8.834e-09	2.463e-02	4.102e-05	4.919e-06	1.204e-02
Moscow	2.255e-06	9.592e-01	1.473e-12	1.499e-03	3.608e-03	8.093e-02	3.379e-08	2.793e-18	7.049e-04
New York	3.075e-09	8.825e-01	6.388e-55	9.018e-03	3.320e-02	2.486e-09	5.634e-01	1.237e-13	3.046e-08
Recife	2.801e-05	6.793e-01	7.237e-43	2.851e-01	1.779e-08	2.858e-15	4.504e-03	9.012e-06	5.512e-02
Rio de Janeiro	7.543e-08	7.251e-01	3.935e-10	6.615e-02	9.177e-03	3.250e-04	4.486e-03	1.766e-01	2.293e-26
St. Petersburg	1.986e-25	7.788e-01	6.496e-13	1.044e-02	2.618e-02	2.471e-05	3.957e-01	1.481e-20	4.270e-09

Tabela A.11: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de sexta-feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	1.907e-02	4.010e-03	1.955e-05	3.280e-01	6.690e-20	1.955e-10	1.659e-04	1.967e-05	4.986e-05
Los Angeles	6.262e-04	5.809e-01	2.884e-13	1.629e-01	2.080e-04	3.946e-11	1.163e-01	2.512e-05	2.782e-22
Manila	3.453e-03	2.642e-01	1.575e-07	7.969e-07	2.451e-25	1.314e-13	1.505e-05	9.103e-14	9.457e-05
Mexico City	5.053e-01	1.764e-03	1.702e-10	5.025e-04	1.308e-07	3.414e-04	1.999e-01	3.788e-06	1.212e-02
Moscow	8.380e-05	8.214e-01	2.265e-13	1.848e-01	2.781e-10	1.451e-10	2.849e-03	8.601e-13	1.585e-07
New York	1.249e-05	4.729e-01	1.681e-26	1.885e-07	1.394e-06	7.239e-05	3.900e-05	4.525e-17	1.227e-03
Recife	1.487e-10	1.045e-01	1.496e-20	5.672e-04	8.877e-07	3.467e-13	2.946e-03	3.677e-09	1.669e-13
Rio de Janeiro	3.322e-07	8.188e-01	5.875e-09	8.066e-01	8.855e-06	1.425e-03	6.080e-01	2.264e-07	1.200e-10
St. Petersburg	2.271e-07	1.588e-01	2.977e-32	6.198e-01	1.483e-09	1.451e-18	1.424e-02	2.601e-07	1.745e-18

Tabela A.12: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de sábado

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	4.948e-17	1.154e-03	9.178e-19	3.494e-05	1.553e-11	2.415e-14	1.071e-11	1.624e-05	1.025e-12
Los Angeles	2.217e-01	1.997e-01	8.593e-15	3.093e-01	3.022e-07	4.347e-11	1.313e-02	4.772e-09	9.023e-17
Manila	9.045e-16	3.227e-02	4.817e-16	3.813e-05	1.358e-02	1.532e-03	3.362e-08	1.061e-04	2.025e-11
Mexico City	5.592e-02	1.161e-02	3.832e-08	2.741e-03	1.024e-09	6.988e-03	2.534e-04	1.584e-06	4.250e-03
Moscow	4.230e-01	2.925e-01	7.228e-03	1.194e-03	8.766e-02	4.858e-05	1.123e-12	4.929e-18	3.004e-02
New York	2.360e-13	4.753e-01	6.360e-12	2.578e-02	5.736e-08	4.114e-12	4.718e-04	2.915e-09	2.983e-18
Recife	3.554e-05	3.170e-01	2.066e-14	3.707e-03	1.518e-13	1.813e-06	6.907e-06	8.497e-11	9.464e-07
Rio de Janeiro	2.320e-04	5.744e-17	2.810e-11	1.040e-05	5.433e-07	3.264e-04	2.924e-01	2.671e-07	1.167e-03
St. Petersburg	9.900e-02	9.189e-01	6.516e-13	1.856e-01	2.692e-06	1.147e-07	1.916e-02	5.796e-13	6.531e-04

Tabela A.13: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com São Paulo em um dia de domingo

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	6.988e-11	4.446e-01	1.052e-18	1.552e-01	1.788e-24	1.855e-08	1.278e-07	1.632e-05	2.083e-13
Los Angeles	4.582e-02	2.850e-01	2.259e-19	9.026e-02	1.582e-09	4.149e-01	1.425e-07	7.138e-06	4.106e-09
Manila	9.975e-07	9.358e-01	1.325e-19	2.533e-02	1.177e-06	1.994e-03	2.163e-01	2.177e-04	2.395e-11
Mexico City	6.480e-11	4.966e-01	1.510e-15	1.074e-02	7.446e-19	1.946e-02	4.760e-04	1.596e-01	2.079e-02
Moscow	5.945e-03	5.125e-01	6.700e-05	1.379e-01	4.581e-04	3.591e-02	4.200e-01	1.756e-21	1.714e-09
New York	1.676e-20	6.831e-01	4.121e-15	7.341e-02	5.325e-09	7.017e-18	5.098e-02	1.328e-03	5.118e-06
Recife	2.029e-09	8.467e-03	2.100e-02	4.214e-02	2.390e-14	7.462e-01	9.439e-01	2.329e-07	3.979e-08
Rio de Janeiro	3.399e-03	3.283e-03	1.646e-07	1.419e-02	8.993e-02	6.148e-06	2.725e-02	2.425e-15	5.543e-10
St. Petersburg	3.009e-02	1.649e-01	1.142e-07	1.084e-01	1.861e-02	1.714e-01	8.949e-01	2.026e-31	9.474e-06

Tabela A.14: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com New York em um dia de segunda-feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	4.084e-02	1.068e-03	1.163e-20	4.894e-09	1.603e-07	1.370e-15	4.203e-01	3.166e-38	6.840e-13
Los Angeles	1.391e-04	1.877e-01	1.423e-08	4.525e-04	1.728e-04	6.120e-03	9.306e-02	7.049e-12	5.196e-05
Manila	4.419e-01	1.274e-09	1.553e-17	1.980e-03	1.784e-09	1.958e-16	3.497e-01	7.643e-30	4.872e-12
Mexico City	1.335e-04	9.431e-05	4.380e-07	2.766e-01	1.458e-12	3.668e-04	3.703e-01	4.933e-04	1.461e-07
Moscow	3.800e-12	3.143e-04	9.736e-18	1.887e-06	8.218e-15	1.273e-10	1.845e-05	5.421e-12	5.733e-16
Recife	4.299e-09	3.978e-06	7.504e-18	9.537e-06	3.788e-156	9.816e-13	3.308e-01	2.640e-09	1.823e-23
Rio de Janeiro	5.009e-09	9.201e-04	1.026e-17	1.280e-06	1.996e-09	2.159e-08	1.841e-05	9.343e-23	9.798e-20
São Paulo	2.627e-06	5.782e-03	3.044e-20	1.447e-10	3.943e-07	4.559e-13	1.173e-08	2.068e-05	1.318e-15
St. Petersburg	1.715e-04	1.557e-04	6.619e-12	3.145e-02	1.874e-17	4.024e-07	5.931e-01	1.715e-20	1.854e-10

Tabela A.15: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com New York em um dia de terça-feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	6.639e-06	1.532e-02	6.192e-29	3.486e-01	4.327e-06	5.562e-09	9.366e-04	4.623e-15	5.288e-09
Los Angeles	5.079e-02	3.293e-03	2.377e-02	3.146e-01	4.710e-08	1.871e-05	1.116e-01	2.476e-07	5.242e-01
Manila	8.446e-06	9.301e-04	3.965e-50	1.337e-02	3.589e-03	1.179e-13	3.650e-03	5.538e-20	3.041e-03
Mexico City	5.868e-04	2.824e-02	9.890e-06	4.126e-01	2.073e-04	8.012e-02	1.206e-02	3.613e-17	1.352e-01
Moscow	6.801e-03	1.720e-01	1.611e-26	3.173e-02	3.353e-11	9.578e-13	4.779e-01	5.405e-16	5.786e-04
Recife	7.235e-07	2.032e-07	4.678e-10	3.795e-01	2.140e-14	1.272e-19	4.727e-01	3.354e-25	1.366e-10
Rio de Janeiro	1.945e-08	1.293e-02	8.094e-32	7.218e-03	1.622e-14	1.261e-08	2.285e-01	1.530e-17	3.285e-09
São Paulo	4.480e-02	2.275e-02	5.878e-33	3.419e-08	8.650e-04	9.591e-08	1.178e-01	1.862e-10	1.767e-03
St. Petersburg	8.795e-07	1.823e-04	3.531e-11	1.309e-20	1.618e-06	2.851e-07	1.552e-01	1.955e-18	4.892e-07

Tabela A.16: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com New York em um dia de quarta-feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	7.306e-19	8.399e-06	6.634e-17	2.447e-04	4.487e-02	4.875e-07	1.300e-02	8.977e-06	3.838e-07
Los Angeles	2.171e-04	1.229e-01	2.041e-10	3.887e-01	5.009e-01	2.903e-01	1.971e-01	2.688e-01	6.056e-12
Manila	1.168e-03	4.730e-01	2.380e-05	6.133e-03	2.578e-02	5.784e-03	3.972e-01	3.025e-03	4.664e-02
Mexico City	6.977e-04	4.513e-01	5.949e-05	5.061e-05	1.770e-01	1.554e-03	3.155e-01	2.454e-05	7.535e-02
Moscow	1.665e-20	9.074e-01	5.843e-15	8.355e-09	5.618e-01	7.326e-16	1.465e-01	8.262e-04	4.736e-04
Recife	5.016e-11	4.651e-05	1.617e-29	4.457e-05	1.324e-02	1.465e-09	1.774e-01	3.525e-10	1.421e-10
Rio de Janeiro	1.013e-18	1.161e-01	1.967e-16	8.532e-14	2.534e-01	4.999e-08	4.787e-03	6.860e-09	8.148e-04
São Paulo	4.359e-03	2.191e-01	1.318e-18	3.552e-03	6.859e-01	5.133e-02	1.742e-02	2.664e-03	1.106e-02
St. Petersburg	4.821e-06	5.626e-01	2.112e-21	2.340e-07	4.445e-02	2.420e-10	6.948e-02	1.933e-11	1.395e-06

Tabela A.17: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com New York em um dia de quinta-feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	2.038e-24	3.957e-02	1.991e-84	3.570e-05	5.121e-21	9.062e-21	3.192e-06	4.785e-30	9.295e-27
Los Angeles	2.256e-08	1.594e-02	5.951e-29	1.748e-02	5.743e-18	1.150e-06	2.331e-03	1.291e-04	1.129e-08
Manila	1.123e-20	7.344e-01	1.241e-43	7.675e-10	5.181e-18	7.702e-12	7.606e-02	1.815e-18	4.751e-17
Mexico City	3.023e-28	2.945e-02	8.561e-32	1.820e-01	4.762e-07	5.122e-06	2.007e-02	6.215e-11	6.117e-09
Moscow	1.105e-37	1.209e-02	1.135e-63	1.801e-05	1.469e-14	4.222e-10	1.910e-01	1.412e-02	7.314e-03
Recife	1.777e-03	5.198e-01	2.655e-36	3.225e-03	4.964e-10	1.419e-27	3.832e-01	1.151e-17	2.634e-14
Rio de Janeiro	1.910e-14	5.637e-01	7.737e-61	2.396e-01	1.333e-10	2.483e-24	6.221e-01	1.096e-14	1.424e-35
São Paulo	1.053e-07	8.465e-01	5.009e-36	2.706e-02	1.966e-04	2.037e-10	9.163e-01	5.524e-10	6.208e-14
St. Petersburg	5.412e-14	6.989e-01	3.959e-39	9.010e-04	4.153e-02	5.249e-10	9.127e-01	5.366e-18	1.624e-16

Tabela A.18: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com New York em um dia de sexta-feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	1.289e-11	3.589e-02	1.678e-50	2.036e-12	6.782e-70	1.725e-18	5.415e-01	2.178e-09	7.648e-10
Los Angeles	2.302e-02	1.414e-01	1.678e-09	1.798e-02	2.320e-04	1.793e-09	2.203e-01	8.797e-03	6.291e-03
Manila	2.506e-06	4.277e-01	2.371e-70	2.150e-01	2.917e-64	1.559e-26	3.957e-03	5.317e-17	3.650e-03
Mexico City	6.257e-08	1.068e-01	2.123e-10	2.196e-07	3.500e-08	1.273e-03	3.865e-01	7.377e-06	1.101e-01
Moscow	1.074e-37	1.669e-01	4.651e-29	3.944e-15	4.575e-16	1.727e-03	2.633e-04	1.926e-14	3.058e-15
Recife	9.477e-12	4.898e-02	1.201e-152	5.547e-08	6.746e-23	5.374e-09	2.220e-04	1.802e-17	2.080e-24
Rio de Janeiro	3.176e-17	1.576e-01	2.003e-34	1.260e-08	6.459e-17	3.195e-18	2.806e-02	3.939e-14	2.877e-13
São Paulo	6.359e-10	3.339e-01	9.910e-26	3.411e-10	3.832e-12	2.930e-12	2.230e-01	7.226e-07	3.118e-02
St. Petersburg	2.909e-70	1.261e-01	5.095e-16	6.359e-12	8.916e-04	4.337e-06	9.263e-04	4.390e-12	6.557e-19

Tabela A.19: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com New York em um dia de sábado

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	4.852e-38	2.309e-02	2.852e-118	3.001e-17	1.179e-30	4.269e-31	5.452e-02	5.508e-25	2.735e-25
Los Angeles	7.302e-10	6.789e-02	1.335e-06	9.562e-03	3.188e-02	4.772e-01	7.322e-01	3.516e-10	5.386e-05
Manila	1.171e-15	3.437e-01	3.832e-151	1.279e-10	1.196e-09	5.339e-14	5.623e-01	1.905e-46	2.766e-26
Mexico City	8.490e-08	3.322e-01	3.423e-26	1.652e-05	1.241e-06	1.611e-02	3.993e-01	5.065e-04	7.388e-03
Moscow	4.266e-08	3.813e-01	5.879e-37	5.335e-17	1.819e-05	1.632e-04	3.838e-01	5.169e-86	8.245e-20
Recife	1.377e-15	3.679e-01	7.045e-87	3.989e-09	6.676e-31	3.899e-30	2.936e-01	7.898e-17	3.593e-27
Rio de Janeiro	4.998e-14	1.685e-03	1.177e-25	1.509e-38	6.864e-13	5.208e-13	2.444e-01	2.285e-15	6.026e-29
São Paulo	1.417e-03	4.555e-01	7.162e-16	9.920e-04	1.677e-04	4.408e-06	4.910e-02	7.130e-10	1.409e-15
St. Petersburg	1.243e-08	8.071e-01	1.153e-39	1.839e-18	2.532e-07	1.729e-02	8.376e-01	2.126e-17	6.660e-05

Tabela A.20: Matriz com os resultados da comparação, *p-valores* de testes Qui-quadrado, entre as Matrizes de Frequência de diferentes cidades com New York em um dia de domingo

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
Bangkok	4.311e-45	9.081e-01	1.771e-35	7.556e-02	2.392e-86	1.666e-14	5.409e-03	1.585e-14	9.798e-11
Los Angeles	2.846e-04		1.993e-02	1.474e-04	3.397e-07	1.157e-03	4.488e-01	4.336e-08	3.783e-02
Manila	4.516e-11		1.288e-47	2.889e-09	5.408e-28	2.016e-06	7.724e-02	4.290e-14	6.275e-11
Mexico City	5.759e-08	2.636e-01	4.755e-22	6.809e-06	2.991e-12	1.885e-06	1.645e-04	2.510e-09	1.153e-04
Moscow	7.349e-03	4.386e-01	5.959e-28	5.411e-07	8.452e-10	2.086e-12	2.025e-01	1.435e-20	5.269e-10
Recife	2.644e-10	7.518e-01	7.083e-40	2.607e-15	1.249e-50	1.932e-08	3.204e-02	8.940e-15	6.751e-13
Rio de Janeiro	9.800e-07		2.675e-33	7.325e-18	1.878e-09	1.866e-18	8.222e-01	4.570e-23	3.088e-10
São Paulo	1.282e-07	7.518e-01	5.222e-15	2.307e-01	8.352e-06	1.386e-08	1.436e-02	2.750e-15	1.544e-03
St. Petersburg	2.744e-06		5.234e-34	7.041e-07	3.247e-10	9.660e-12	4.792e-02	4.176e-22	1.848e-09

A.2.2 Representação por meio dos *p-valores* de testes para Cadeias de Markov

Tabela A.21: Matriz com os resultados da comparação, *p-valores* de testes de [Kullback et al. \(1962\)](#), entre as Cadeias de Markov de diferentes cidades com New York

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Bangkok	1.297e-02	4.145e-03	2.489e-03	1.631e-04	3.160e-03	6.374e-04	6.279e-03
Los Angeles	4.397e-02	4.935e-02	1.634e-02	5.033e-02	1.774e-01	1.910e-01	2.510e-01
Manila	1.950e-02	5.194e-03	1.092e-01	3.426e-03	5.789e-03	1.815e-03	4.369e-03
Mexico City	1.541e-02	1.809e-02	3.319e-02	2.585e-02	6.215e-02	1.229e-01	5.402e-02
Moscow	4.757e-04	1.409e-03	9.154e-04	4.381e-03	2.551e-03	2.349e-03	2.497e-03
Rio de Janeiro	3.511e-05	5.384e-05	6.927e-05	1.700e-03	8.049e-04	7.548e-05	1.562e-05
Recife	2.511e-04	1.075e-03	1.731e-05	1.336e-03	6.288e-06	5.980e-06	1.356e-04
St. Petersburg	2.143e-03	2.071e-04	2.313e-04	9.101e-03	4.189e-03	6.990e-03	1.014e-03
São Paulo	5.696e-03	1.325e-03	2.953e-03	3.353e-02	2.270e-02	3.119e-02	1.819e-02

Tabela A.22: Matriz com os resultados da comparação, *p-valores* de testes de [Kullback et al. \(1962\)](#), entre as Cadeias de Markov de diferentes cidades com São Paulo

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Bangkok	1.638e-02	3.548e-03	2.849e-02	1.158e-02	8.785e-02	8.235e-03	1.253e-02
Los Angeles	1.966e-04	2.068e-04	1.652e-03	2.713e-02	4.248e-02	4.757e-02	2.670e-02
Manila	4.127e-02	1.375e-02	1.825e-02	8.670e-02	9.115e-02	3.466e-02	6.521e-02
Mexico City	6.146e-02	3.204e-02	2.717e-03	6.015e-02	3.247e-02	1.228e-01	4.796e-02
Moscow	4.536e-02	5.471e-02	5.798e-02	3.012e-02	3.330e-02	8.923e-02	4.909e-02
New York	4.770e-03	1.191e-03	2.272e-03	2.907e-02	2.412e-02	3.326e-02	1.142e-02
Rio de Janeiro	3.376e-02	2.170e-02	1.969e-02	8.679e-02	5.037e-02	4.007e-02	5.009e-02
Recife	9.485e-02	5.640e-02	3.151e-02	2.190e-02	1.030e-02	5.126e-03	9.174e-02
St. Petersburg	1.698e-02	2.629e-02	1.686e-02	7.528e-02	2.931e-02	7.288e-02	5.901e-02

Tabela A.23: Matriz com os resultados da comparação, *p-valores* de testes de [Kullback et al. \(1962\)](#), entre as Cadeias de Markov de diferentes períodos do dia cidade de New York

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Manhã x Tarde	1.756e-02	2.435e-02	8.378e-03	1.907e-02	9.746e-02	8.614e-02	3.005e-01
Manhã x Noite	1.708e-03	5.822e-06	5.042e-06	7.175e-04	1.102e-04	1.254e-03	5.517e-03
Tarde x Noite	6.419e-02	1.742e-03	1.015e-02	9.396e-03	8.002e-03	1.891e-01	2.889e-02

Tabela A.24: Matriz com os resultados da comparação, *p-valores* de testes de [Kullback et al. \(1962\)](#), entre as Cadeias de Markov de diferentes períodos do dia cidade de São Paulo

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Manhã x Tarde	4.271e-03	1.032e-03	7.615e-03	0	6.585e-02	1.021e-01	2.073e-02
Manhã x Noite	4.173e-09	3.656e-06	3.061e-05	2.772e-07	1.058e-05	1.034e-03	3.571e-04
Tarde x Noite	2.300e-04	1.287e-03	5.220e-04	0	4.201e-03	7.243e-03	2.912e-02

Tabela A.25: Matriz com os resultados da comparação, *p-valores* de testes de [Kullback et al. \(1962\)](#), entre as Cadeias de Markov de diferentes cidades com New York no turno matutino

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Bangkok	7.475e-03	2.546e-03	1.037e-03	1.218e-05	1.063e-03	1.310e-02	2.792e-04
Los Angeles	2.074e-03	4.036e-03	2.408e-03	1.376e-03	3.229e-04	1.849e-01	3.890e-02
Manila	4.713e-03	1.511e-02	2.814e-02	2.636e-03	4.081e-04	4.801e-03	1.299e-02
Mexico City	2.009e-02	4.479e-03	2.469e-02	1.801e-03	4.418e-04	3.061e-02	9.993e-02
Moscow	2.044e-04	1.326e-03	4.056e-04	2.138e-05	3.380e-04	7.358e-05	1.006e-05
Recife	5.982e-04	1.824e-04	2.179e-04	1.954e-06	2.170e-05	2.278e-09	1.131e-09
Rio de Janeiro	3.114e-05	7.795e-05	2.640e-03	0.000e+00	5.710e-04	1.443e-06	2.421e-06
St. Petersburg	5.964e-05	9.832e-05	4.486e-06	5.298e-10	4.586e-06	4.663e-05	1.148e-08
São Paulo	3.315e-02	4.674e-02	7.019e-03	2.383e-03	5.474e-03	1.566e-02	1.422e-03

Tabela A.26: Matriz com os resultados da comparação, p -valores de testes de [Kullback et al. \(1962\)](#), entre as Cadeias de Markov de diferentes cidades com New York no turno vespertino

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Bangkok	5.307e-03	7.032e-03	9.488e-04	1.192e-04	3.764e-02	9.275e-03	5.469e-03
Los Angeles	1.181e-02	5.042e-05	7.127e-03	2.363e-02	4.406e-02	5.301e-02	1.608e-01
Manila	4.177e-03	4.669e-03	8.859e-03	4.667e-04	1.120e-02	7.261e-02	5.482e-03
Mexico City	3.177e-03	6.270e-03	6.161e-03	4.130e-03	4.738e-02	1.067e-01	1.275e-01
Moscow	1.638e-03	9.611e-04	1.563e-05	1.560e-02	1.404e-02	1.974e-02	1.666e-03
Recife	3.901e-06	1.324e-06	1.208e-07	5.794e-04	3.538e-12	8.664e-04	6.093e-05
Rio de Janeiro	6.420e-07	2.097e-07	1.481e-06	6.773e-04	3.759e-05	1.130e-03	9.934e-06
St. Petersburg	3.397e-04	4.846e-05	1.322e-05	4.869e-03	3.801e-03	5.960e-03	1.997e-02
São Paulo	6.316e-04	1.017e-05	2.101e-05	0.000e+00	2.628e-02	2.938e-02	1.640e-02

Tabela A.27: Matriz com os resultados da comparação, p -valores de testes de [Kullback et al. \(1962\)](#), entre as Cadeias de Markov de diferentes cidades com New York no turno noturno

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
Bangkok	4.657e-04	3.188e-04	3.202e-06	1.911e-05	1.334e-04	1.715e-07	4.048e-07
Los Angeles	2.218e-07	2.293e-04	9.956e-05	1.290e-04	5.189e-02	3.128e-02	0.000e+00
Manila	1.324e-04	8.849e-08	1.120e-02	2.819e-04	2.702e-04	2.065e-06	9.972e-08
Mexico City	4.768e-05	7.616e-04	1.451e-03	1.034e-02	1.181e-01	2.243e-02	7.901e-06
Moscow	1.165e-10	1.027e-04	3.901e-05	1.884e-04	1.076e-04	8.368e-07	2.291e-05
Recife	2.769e-10	2.602e-07	6.796e-11	2.774e-09	8.105e-09	0.000e+00	6.956e-10
Rio de Janeiro	8.453e-11	1.157e-07	8.272e-10	0.000e+00	4.187e-04	2.242e-08	0.000e+00
St. Petersburg	5.299e-07	1.537e-07	3.696e-06	4.587e-03	2.796e-04	0.000e+00	1.893e-08
São Paulo	4.924e-08	1.548e-10	2.263e-07	2.636e-03	2.889e-03	1.208e-03	1.395e-05

A.3 Matrizes de Transição de Probabilidade Estimadas

Tabela A.28: Matrizes de Transição de Probabilidade de New York

Segunda-Feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
A&E	0.071	0.018	0.429	0.071	0.071	0.071	0.036	0.161	0.071
C&U	0.125	0.156	0.250	0.000	0.031	0.188	0.031	0.125	0.094
FD	0.055	0.043	0.250	0.070	0.117	0.148	0.070	0.148	0.098
NL	0.051	0.000	0.231	0.462	0.051	0.026	0.103	0.026	0.051
O&R	0.000	0.022	0.283	0.076	0.120	0.141	0.054	0.130	0.174
P&O	0.065	0.039	0.275	0.033	0.144	0.209	0.039	0.111	0.085
RD	0.056	0.000	0.278	0.000	0.111	0.167	0.056	0.222	0.111
S&S	0.031	0.041	0.227	0.026	0.206	0.113	0.046	0.247	0.062
T&T	0.024	0.006	0.213	0.024	0.116	0.116	0.024	0.104	0.372

Terça-Feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
A&E	0.136	0.000	0.322	0.119	0.034	0.153	0.034	0.085	0.119
C&U	0.034	0.172	0.276	0.034	0.103	0.069	0.034	0.207	0.069
FD	0.121	0.042	0.229	0.107	0.107	0.117	0.033	0.145	0.098
NL	0.111	0.000	0.267	0.511	0.022	0.000	0.022	0.022	0.044
O&R	0.042	0.014	0.250	0.069	0.139	0.097	0.097	0.153	0.139
P&O	0.031	0.000	0.351	0.053	0.076	0.252	0.038	0.137	0.061
RD	0.000	0.029	0.176	0.000	0.176	0.147	0.235	0.147	0.088
S&S	0.064	0.023	0.291	0.064	0.093	0.087	0.052	0.267	0.058
T&T	0.066	0.041	0.174	0.041	0.091	0.107	0.050	0.091	0.339

Quarta-Feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
A&E	0.263	0.000	0.434	0.118	0.026	0.026	0.000	0.053	0.079
C&U	0.048	0.190	0.190	0.143	0.143	0.143	0.000	0.143	0.000
FD	0.176	0.011	0.262	0.096	0.032	0.176	0.032	0.128	0.086
NL	0.200	0.000	0.371	0.371	0.000	0.000	0.029	0.029	0.000
O&R	0.024	0.024	0.238	0.024	0.167	0.167	0.095	0.095	0.167
P&O	0.016	0.016	0.352	0.041	0.074	0.238	0.049	0.115	0.098
RD	0.038	0.038	0.115	0.077	0.154	0.269	0.038	0.154	0.115
S&S	0.033	0.013	0.382	0.026	0.112	0.066	0.059	0.250	0.059
T&T	0.034	0.008	0.203	0.017	0.102	0.161	0.068	0.068	0.339

Quinta-Feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
A&E	0.153	0.000	0.318	0.165	0.118	0.071	0.024	0.035	0.118
C&U	0.000	0.250	0.312	0.000	0.188	0.062	0.000	0.062	0.125
FD	0.069	0.012	0.458	0.090	0.106	0.090	0.019	0.097	0.058
NL	0.141	0.000	0.295	0.423	0.013	0.026	0.038	0.038	0.026
O&R	0.110	0.012	0.262	0.085	0.159	0.104	0.067	0.110	0.091
P&O	0.060	0.006	0.317	0.066	0.108	0.168	0.024	0.156	0.096
RD	0.026	0.000	0.184	0.053	0.132	0.026	0.316	0.184	0.079
S&S	0.046	0.005	0.291	0.036	0.158	0.082	0.082	0.276	0.026
T&T	0.076	0.006	0.171	0.029	0.147	0.112	0.029	0.053	0.376

Sexta-Feira

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
A&E	0.183	0.000	0.275	0.202	0.073	0.028	0.055	0.073	0.110
C&U	0.000	0.105	0.316	0.053	0.053	0.158	0.053	0.211	0.053
FD	0.106	0.014	0.353	0.163	0.090	0.059	0.029	0.124	0.061
NL	0.136	0.006	0.299	0.448	0.019	0.013	0.026	0.013	0.039
O&R	0.105	0.015	0.310	0.075	0.185	0.090	0.040	0.110	0.070
P&O	0.058	0.022	0.326	0.072	0.130	0.152	0.036	0.138	0.065
RD	0.042	0.000	0.250	0.125	0.042	0.083	0.125	0.333	0.000
S&S	0.030	0.013	0.312	0.056	0.108	0.039	0.048	0.377	0.017
T&T	0.033	0.007	0.197	0.072	0.086	0.086	0.046	0.105	0.368

Sábado

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
A&E	0.155	0.000	0.426	0.169	0.047	0.007	0.007	0.128	0.061
C&U	0.000	0.000	0.400	0.100	0.200	0.200	0.000	0.100	0.000
FD	0.110	0.002	0.307	0.207	0.082	0.027	0.042	0.154	0.068
NL	0.130	0.000	0.280	0.522	0.000	0.012	0.012	0.019	0.025
O&R	0.076	0.013	0.299	0.040	0.228	0.058	0.022	0.210	0.054
P&O	0.124	0.021	0.278	0.134	0.082	0.134	0.021	0.155	0.052
RD	0.000	0.000	0.278	0.111	0.167	0.056	0.000	0.222	0.167
S&S	0.046	0.004	0.350	0.100	0.104	0.042	0.042	0.277	0.035
T&T	0.039	0.008	0.289	0.102	0.117	0.070	0.016	0.117	0.242

Domingo

	A&E	C&U	FD	NL	O&R	P&O	RD	S&S	T&T
A&E	0.158	0.009	0.377	0.079	0.132	0.035	0.009	0.158	0.044
C&U	0.060	0.000	0.400	0.040	0.000	0.300	0.200	0.000	0.000
FD	0.109	0.000	0.336	0.088	0.111	0.059	0.049	0.191	0.057
NL	0.038	0.019	0.404	0.385	0.038	0.000	0.038	0.019	0.058
O&R	0.117	0.018	0.301	0.055	0.276	0.055	0.018	0.092	0.067
P&O	0.074	0.000	0.495	0.042	0.095	0.032	0.042	0.158	0.063
RD	0.042	0.000	0.250	0.125	0.125	0.000	0.000	0.250	0.208
S&S	0.060	0.008	0.321	0.034	0.109	0.045	0.068	0.321	0.034
T&T	0.105	0.000	0.189	0.028	0.077	0.084	0.042	0.119	0.357

REFERÊNCIAS BIBLIOGRÁFICAS

- Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.-H. & Liu, B. (2011), Predicting flu trends using twitter data, *in* 'Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on', IEEE, pp. 702–707.
- Amaral, I. (2017), 'Computational social sciences', *Encyclopedia of Big Data*.
- Becker, R., Cáceres, R., Hanson, K., Isaacman, S., Loh, J. M., Martonosi, M., Rowland, J., Urbanek, S., Varshavsky, A. & Volinsky, C. (2013), 'Human mobility characterization from cellular network data', *Communications of the ACM* **56**(1), 74–82.
- Brown, C., Noulas, A., Mascolo, C. & Blondel, V. (2013), A place-focused model for social networks in cities, *in* 'Social Computing (SocialCom), 2013 International Conference on', IEEE, pp. 75–80.
- Burke, J. A., Estrin, D., Hansen, M., Parker, A., Ramanathan, N. & Reddy, S. (2006), 'Participatory sensing', *Workshop on World-Sensor-Web (WSW'06) at ACM Sensys (SenSys'06)*.
- Castro, P. S., Zhang, D., Chen, C., Li, S. & Pan, G. (2013), 'From taxi gps traces to social and community dynamics: A survey', *ACM Computing Surveys (CSUR)* **46**(2), 17.
- Cho, E., Myers, S. A. & Leskovec, J. (2011), Friendship and mobility: user movement in location-based social networks, *in* 'Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, pp. 1082–1090.
- Coifman, B. (2006), 'Vehicle level evaluation of loop detectors and the remote traffic microwave sensor', *Journal of transportation engineering* **132**(3), 213–226.
- Colombo, G. B., Chorley, M. J., Williams, M. J., Allen, S. M. & Whitaker, R. M. (2012), You are where you eat: Foursquare checkins as indicators of human mobility and behaviour, *in* 'Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on', IEEE, pp. 217–222.
- Cranshaw, J., Schwartz, R., Hong, J. I. & Sadeh, N. (2012), 'The livelihoods project: Utilizing social media to understand the dynamics of a city', *International AAAI Conference on Weblogs and Social Media* p. 58. URL <http://ssrn.com/abstract=2168428>.
- Eagle, N. & Pentland, A. S. (2006), 'Reality mining: sensing complex social systems', *Personal and ubiquitous computing* **10**(4), 255–268.

- Eagle, N., Pentland, A. S. & Lazer, D. (2009), 'Inferring friendship network structure by using mobile phone data', *Proceedings of the national academy of sciences* **106**(36), 15274–15278.
- Foursquare (2016), 'Foursquare website'. URL <https://pt.foursquare.com/about>, Acessado em: 18-12-2016.
- Foursquare (2017), 'Foursquare para desenvolvedores: Rate limits'. URL <https://developer.foursquare.com/overview/ratelimits>, Acessado em: 20-02-2017.
- Frias-Martinez, V., Soto, V., Hohwald, H. & Frias-Martinez, E. (2012*a*), Characterizing urban landscapes using geolocated tweets, in 'Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)', IEEE, pp. 239–248.
- Frias-Martinez, V., Soto, V., Hohwald, H. & Frias-Martinez, E. (2012*b*), Characterizing urban landscapes using geolocated tweets, in 'Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)', pp. 239–248.
- Gao, S., Liu, Y., Wang, Y. & Ma, X. (2013), 'Discovering spatial interaction communities from mobile phone data', *Transactions in GIS* **17**(3), 463–481.
- Gomide, J., Veloso, A., Meira Jr, W., Almeida, V., Benevenuto, F., Ferraz, F. & Teixeira, M. (2011), Dengue surveillance based on a computational model of spatio-temporal locality of twitter, in 'Proceedings of the 3rd international web science conference', ACM, p. 3.
- Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A.-L. (2008), 'Understanding individual human mobility patterns', *arXiv preprint arXiv:0806.1256*.
- Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W. & Prasad, S. (2015), 'Extracting and understanding urban areas of interest using geotagged photos', *Computers, Environment and Urban Systems* **54**, 240–254.
- ICT Data and Statistics Division (2016), 'ICT Facts & Figures', *14th World Telecommunication/ICT Indicators Symposium (WTIS)*.
- Inglehart, R. & Welzel, C. (2010), 'Changing mass priorities: The link between modernization and democracy', *Perspectives on Politics* **8**(02), 551–567.
- Kosmin, B. A. & Lachman, S. P. (2011), *One Nation Under God: Religion in Contemporary American Society*, Crown Publisher.

- Kostakos, V., Nicolai, T., Yoneki, E., O'Neill, E., Kenn, H. & Crowcroft, J. (2009), 'Understanding and measuring the urban pervasive infrastructure', *Personal and Ubiquitous Computing* **13**(5), 355–364. URL <http://dx.doi.org/10.1007/s00779-008-0196-1>.
- Kullback, S., Kupperman, M. & Ku, H. H. (1962), 'Tests for contingency tables and markov chains', *Technometrics* **4**(4), 573–608. URL <http://www.jstor.org/stable/1266291>.
- Lamos, V. & Cristianini, N. (2010), Tracking the flu pandemic by monitoring the social web, *in* 'Cognitive Information Processing (CIP), 2010 2nd International Workshop on', IEEE, pp. 411–416.
- Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M. et al. (2009), 'Life in the network: the coming age of computational social science', *Science (New York, NY)* **323**(5915), 721.
- Liu, Y., Kang, C., Gao, S., Xiao, Y. & Tian, Y. (2012), 'Understanding intra-urban trip patterns from taxi trajectory data', *Journal of geographical systems* **14**(4), 463–483.
- Ma, X., Tao, Z., Wang, Y., Yu, H. & Wang, Y. (2015), 'Long short-term memory neural network for traffic speed prediction using remote microwave sensor data', *Transportation Research Part C: Emerging Technologies* **54**, 187–197.
- Manca, M., Boratto, L., Roman, V. M., i Gallissà, O. M. & Kaltenbrunner, A. (2017), 'Using social media to characterize urban mobility patterns: State-of-the-art survey and case-study', *Online Social Networks and Media* **1**, 56–69.
- Noulas, A., Scellato, S., Mascolo, C. & Pontil, M. (2011a), 'An empirical study of geographic user activity patterns in foursquare.', *ICWSM* **11**, 70–573.
- Noulas, A., Scellato, S., Mascolo, C. & Pontil, M. (2011b), 'Exploiting semantic annotations for clustering geographic areas and users in location-based social networks.', *The social mobile web*.
- Nunes, I. O., de Melo, P. O. V. & Loureiro, A. A. (2016), Group mobility: Detection, tracking and characterization, *in* 'Communications (ICC), 2016 IEEE International Conference on', IEEE, pp. 1–6.
- Pardo, L. (2005), *Statistical inference based on divergence measures*, CRC Press.
- Pham, H., Shahabi, C. & Liu, Y. (2016), 'Inferring social strength from spatiotemporal data', *ACM Transactions on Database Systems (TODS)* **41**(1), 7.
- Phithakkitnukoon, S. & Olivier, P. (2011), Sensing urban social geography using online social networking data, *in* 'The Social Mobile Web', pp. 36–39.

- Sadilek, A., Kautz, H. A. & Silenzio, V. (2012), Predicting disease transmission from geo-tagged micro-blog data., *in* 'AAAI'.
- Silva, T. H. & Loureiro, A. (2015), 'Computação urbana: Técnicas para o estudo de sociedades com redes de sensoriamento participativo', *Anais da XXXIV JAI* **8329**, 68–122.
- Silva, T. H., d. Melo, P. O. S. V., Almeida, J. M., Salles, J. & Loureiro, A. A. F. (2012), Visualizing the invisible image of cities, *in* 'Green Computing and Communications (GreenCom), 2012 IEEE International Conference on', pp. 382–389.
- Silva, T. H., de Melo, P. O., Almeida, J., Musolesi, M. & Loureiro, A. (2014a), 'You are what you eat (and drink): Identifying cultural boundaries by analyzing food & drink habits in foursquare', *arXiv preprint arXiv:1404.1009*.
- Silva, T. H., Vaz de Melo, P. O., Almeida, J. M., Salles, J. & Loureiro, A. A. (2014b), 'Revealing the city that we cannot see', *ACM Transactions on Internet Technology (TOIT)* **14**(4), 26.
- Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. (2010), 'Limits of predictability in human mobility', *Science* **327**(5968), 1018–1021.
- Stewart, J. B. (2016), 'Facebook has 50 minutes of your time each day. it wants more.'. URL <https://www.nytimes.com/2016/05/06/business/facebook-bends-the-rules-of-audience-engagement-to-its-advantage.html>, New York Times Online, acessado em: 04-03-2017.
- Tammaru, T., Ham, M. V., Marcinczak, S. & Musterd, S. (2016), *Socio-Economic Segregation in European Capital Cities: East Meets West*, Routledge.
- Twitter (2017), 'Twitter developer documentation - streaming APIs'. URL <https://dev.twitter.com/streaming/overview>, Acessado em: 20-02-2017.
- Waters, M. C. (1994), 'Ethnic and racial identities of second-generation black immigrants in new york city', *The International Migration Review* **28**(4), 795–820. URL <http://www.jstor.org/stable/2547158>.
- Xie, X.-F. & Wang, Z.-J. (2014), 'Combining physical and participatory sensing in urban mobility networks', *Workshop on Big Data and Urban Informatics (BDUIC)* pp. 859–876.
- Yuan, J., Zheng, Y. & Xie, X. (2012), Discovering regions of different functions in a city using human mobility and pois, *in* 'Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, pp. 186–194.
- Zhao, K., Musolesi, M., Hui, P., Rao, W. & Tarkoma, S. (2015), 'Explaining the power-law distribution of human mobility through transportation modality decomposition', *Scientific reports*.

- Zheng, Y., Capra, L., Wolfson, O. & Yang, H. (2014), 'Urban computing: concepts, methodologies, and applications', *ACM Transactions on Intelligent Systems and Technology (TIST)* **5**(3), 38.
- Zheng, Y., Liu, Y., Yuan, J. & Xie, X. (2011), Urban computing with taxicabs, *in* 'Proceedings of the 13th international conference on Ubiquitous computing', ACM, pp. 89–98.

Este trabalho foi redigido em \LaTeX utilizando uma modificação do estilo IC-UFAL. As referências bibliográficas foram preparadas no JabRef e administradas pelo \BIBTeX com o estilo LaCCAN. O texto utiliza fonte Fourier-GUTenberg e os elementos matemáticos a família tipográfica Euler Virtual Math, ambas em corpo de 12 pontos.