



Programa Interdisciplinar de Pós-Graduação em

# **Computação Aplicada**

**Mestrado Acadêmico**

Richard Mateus Altmayer

**GEOSOCIAL: Um modelo de análise e agrupamento de população de pessoas baseado em hábitos de frequência e semântica de locais**

São Leopoldo, 2018



Richard Mateus Altmayer

**GEOSOCIAL:  
Um modelo de análise e agrupamento de população de pessoas baseado em hábitos de  
frequência e semântica de locais**

Proposta de Dissertação apresentada como  
requisito parcial para a obtenção do título de  
Mestre, pelo Programa Interdisciplinar de  
Pós-Graduação em Computação Aplicada da  
Universidade do Vale do Rio dos Sinos —  
UNISINOS

Orientador:  
Prof<sup>ª</sup>. Dr<sup>ª</sup>. Marta Becker Villamil

São Leopoldo  
2018

A468g

Altmayer, Richard Mateus.

GeoSocial : um modelo de análise e agrupamento de população de pessoas baseado em hábitos de frequência e semântica de locais / Richard Mateus Altmayer. – 2018.

116 p. : il. ; 30 cm.

Dissertação (mestrado) – Universidade do Vale do Rio dos Sinos, Programa Interdisciplinar de Pós-Graduação em Computação Aplicada, 2018.

“Orientador: Profa. Dra. Marta Becker Villamil.”

1. Análise visual. 2. Coordenadas paralelas. 3. Clusterização.  
4. Rede social. I. Título.

CDU 004

Dados Internacionais de Catalogação na Publicação (CIP)  
(Bibliotecário: Flávio Nunes – CRB 10/1298)

Richard Mateus Altmayer

GEOSOCIAL:

Um modelo de análise e agrupamento de população de pessoas baseado em hábitos de frequência e semântica de locais

Dissertação apresentada à Universidade do Vale do Rio dos Sinos – Unisinos, como requisito parcial para obtenção do título de Mestre em Computação Aplicada.

Aprovado em 12/04/2018

BANCA EXAMINADORA

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Marta Becker Villamil – UNISINOS

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Isabel Manssour – PUCRS

---

Prof. Dr. Jorge Luis Victória Barbosa – UNISINOS

Prof<sup>a</sup>. Dr<sup>a</sup>. Marta Becker Villamil (Orientadora)

Visto e permitida a impressão  
São Leopoldo,

Prof. Dr. Rodrigo da Rosa Righi  
Coordenador PPG em Computação Aplicada



## **AGRADECIMENTOS**

Gostaria de agradecer a todos que, de alguma maneira, contribuíram para a realização desta pesquisa, em especial:

À família, principalmente à minha esposa Taís, pelo incondicional apoio prestado durante o desenvolvimento deste trabalho.

À minha orientadora, professora Dr<sup>a</sup>. Marta Becker Villamil, por todo suporte prestado e confiança demonstrada durante estes dois anos de realização desta pesquisa.

À CAPES, pelo auxílio à esta pesquisa.



## RESUMO

A utilização de informações sobre comportamento de navegação de usuários na web tem sido amplamente utilizada para traçar perfis comportamentais de usuários com o intuito de oferecer anúncios publicitários por segmentos ou categorias. Nesta mesma linha, hábitos de comportamento baseado em locais que um indivíduo frequenta no seu cotidiano também podem ser analisados. Este trabalho propõe um modelo de agrupamento de indivíduos de uma população para posterior análise de seus hábitos de frequência a locais (GeoSocial). Os padrões de frequência dos grupos formados representam características de comportamento da população e podem ajudar a identificar oportunidades mercadológicas ou auxiliar aos tomadores de decisão ligados ao governo proporem determinadas melhorias/mudanças na infra-estrutura de uma determinada cidade. As informações dos locais de interesse frequentados pelos usuários são capturadas por coordenadas GPS via aplicativo móvel desenvolvido. O aplicativo rastreia e armazena as localidades que o indivíduo frequenta, permite visualizar o seu tempo e locais de permanência e pode conectá-lo à uma rede social formada a partir das similaridades entre seus hábitos e de outros indivíduos. O modelo proposto engloba: *i.* um módulo de clusterização de usuários que utiliza a técnica *Affinity Propagation*; *ii.* um módulo de visualização interativa para análise dos grupos por meio da técnica de Coordenadas Paralelas. O GeoSocial é avaliado mediante a utilização de diferentes cenários, fazendo uso de dados artificiais gerados. A avaliação evidencia o potencial de adaptação do modelo à diferentes objetivos de análise.

**Palavras-chave:** Análise visual. Coordenadas Paralelas. Clusterização. Rede Social.



## ABSTRACT

Information about user navigation behavior on the web has been widely used to draw user behavioral profiles in order to offer advertisements segmented by categories. In this same line, behavior habits based on places that an individual attends in their daily life can also be analyzed. This paper proposes a clustering model of individuals for further analysis of their habits of frequency in places (GeoSocial). Patterns of the formed groups represent characteristics of population's behavior and can help to identify market opportunities or to help decision makers linked to government to propose improvements/changes in the infrastructure of a city. Users information about their frequented interest places are captured by GPS coordinates by a mobile app developed. App tracks and storages places that are frequent individuals. It allows visualize their time permanency on places and connect they to a social network formed from the similarities between their habits and the others. The proposed model includes: *i.* a user clustering module based on *Affinity Propagation* technique; *ii.* an interactive visualization module to analyze individual data correlation of groups based on Parallel Coordinates technique. GeoSocial is evaluated by different scenarios, making use of artificial data generated. Evaluation indicates the possibility of the model to a multitude of objectives.

**Keywords:** Visual analytics. Parallel Coordinates. Clustering. Social Network.



## LISTA DE FIGURAS

|    |  |    |
|----|--|----|
| 1  | Evolução dos algoritmos de clusterização . . . . .   | 25 |
| 2  | Etapas do processo de clusterização . . . . .  | 25 |
| 3  | <i>Affinity Propagation</i> - Iterações até a convergência para formalização dos <i>clusters</i> . . . . .   | 27 |
| 4  | Processo de troca de mensagens entre os pontos de dados . . . . .  | 28 |
| 5  | Escopo do processo de análise visual . . . . .   | 30 |
| 6  | Processo de análise visual . . . . .   | 31 |
| 7  | Geometria das Coordenadas Paralelas . . . . .  | 31 |
| 8  | Múltiplas dimensões representadas por Coordenadas Paralelas . . . . .  | 32 |
| 9  | Limites de valores definidos para os eixos das Coordenadas Paralelas . . . . .   | 33 |
| 10 | Técnica de <i>brushing</i> aplicada sobre os eixos das Coordenadas Paralelas . . . . .   | 34 |
| 11 | Translação dos eixos das Coordenadas Paralelas . . . . .   | 34 |
| 12 | <i>Flipping</i> sobre o eixo potência das Coordenadas Paralelas . . . . .  | 35 |
| 13 | Representação de rede social . . . . .   | 36 |
| 14 | Representação de detecção de comunidades em rede social . . . . .  | 36 |
| 15 | Rede heterogênea apresentando informações sociais e espaciais . . . . .  | 40 |
| 16 | Filtragem semântica utilizando ontologia em grafo . . . . .  | 41 |
| 17 | Gráfico de tempo mostrando atividade de determinado usuário . . . . .  | 42 |
| 18 | Anel de comportamento . . . . .  | 42 |
| 19 | Projeção do comportamento de indivíduo utilizando diferentes número de <i>eigenbehaviors</i> . . . . .   | 44 |
| 20 | Rede social da população do projeto <i>Reality Mining</i> . . . . .  | 44 |
| 21 | Trajetórias GPS e interesses dos usuários . . . . .  | 45 |
| 22 | Definição de ponto de permanência a partir de trajetória GPS . . . . .   | 46 |
| 23 | Apuração da máxima combinação de viagens . . . . .   | 47 |
| 24 | Escopo, dividido em módulos, do método proposto . . . . .  | 51 |
| 25 | Processo de equiparação dos perfis dos usuários . . . . .  | 54 |
| 26 | Processo de transformação das matrizes em vetores . . . . .  | 55 |
| 27 | Vetor resultante do processo de clusterização <i>Affinity Propagation</i> . . . . .  | 57 |
| 28 | Aplicação de filtros na ferramenta de análise visual . . . . .   | 58 |
| 29 | Seleção de registro(s) da tabela . . . . .   | 59 |
| 30 | Técnica de <i>brushing</i> aplicado ao sistema de Coordenadas Paralelas . . . . .  | 59 |
| 31 | Quantidade de usuários por <i>cluster</i> originado no cenário de avaliação 1 . . . . .  | 71 |
| 32 | Comportamento de usuários em locais categorizados como “academia” . . . . .  | 72 |
| 33 | Análise do sistema de coordenadas paralelas sobre o contexto de períodos do dia do cenário de avaliação 1 . . . . .  | 72 |
| 34 | Indícios de correlação negativa entre os eixos do sistema de coordenadas paralelas sobre o contexto de períodos do dia . . . . .   | 73 |
| 35 | Quantidade de usuários por <i>cluster</i> mediante adoção da técnica <i>brushing</i> sobre eixo do sistema de coordenadas paralelas no cenário de avaliação 1 . . . . .      | 73 |
| 36 | Sexo dos usuários por <i>cluster</i> mediante adoção da técnica <i>brushing</i> sobre eixo do sistema de coordenadas paralelas no cenário de avaliação 1 . . . . .           | 74 |
| 37 | Faixas etárias dos usuários por <i>cluster</i> mediante adoção da técnica <i>brushing</i> sobre eixo do sistema de coordenadas paralelas no cenário de avaliação 1 . . . . . | 74 |
| 38 | Quantidade de usuários por <i>cluster</i> originado no cenário de avaliação 2 . . . . .  | 75 |
| 39 | Atividades dos usuários considerados no cenário de avaliação 2 . . . . .   | 76 |

|    |   |     |
|----|---|-----|
| 40 | Técnica de <i>brushing</i> aplicada no sistema de coordenadas paralelas sob o contexto referente aos meses do ano em bares . . . . .                                    | 77  |
| 41 | Quantidade de usuários por <i>cluster</i> existentes no cenário de avaliação, considerando <i>clusters</i> com padrões expressivos . . . . .                            | 77  |
| 42 | Quantidade de usuários por formação acadêmica considerados no cenário de avaliação 2 . . . . .  | 78  |
| 43 | Técnica de <i>brushing</i> nos sistemas de coordenadas paralelas construídas no cenário de avaliação 2 . . . . .  | 79  |
| 44 | Quantidade de usuários por formação acadêmica considerada no cenário de avaliação 2 . . . . .   | 79  |
| 45 | Quantidade de usuários por faixa etária considerada no cenário de avaliação 2   | 80  |
| 46 | Quantidade de usuários por <i>cluster</i> originado no cenário de avaliação 3 . . . .   | 80  |
| 47 | Técnica de <i>brushing</i> sobre os eixos da análise realizada no cenário de avaliação 3 . . . . .  | 82  |
| 48 | Quantidade de usuários por faixas de renda consideradas no cenário de avaliação 3 . . . . .   | 82  |
| 49 | Seleção de registros pertencentes à um <i>cluster</i> com padrão expressivo considerados no cenário de avaliação 3 . . . . .  | 83  |
| 50 | Sistemas de coordenadas paralelas resultantes do processo de retroalimentação, evidenciando o subconjunto de indivíduos considerado no cenário de avaliação 3 . . . . . | 84  |
| 51 | Aplicativo móvel - Diferentes apresentações do menu ao usuário . . . . .  | 90  |
| 52 | Aplicativo móvel - Credenciais do usuário . . . . .   | 90  |
| 53 | Aplicativo móvel - Definição de informações básicas do usuário . . . . .  | 91  |
| 54 | Aplicativo móvel - Definição de endereços do usuário . . . . .  | 92  |
| 55 | Aplicativo móvel - Escolha da foto do usuário . . . . .   | 93  |
| 56 | Aplicativo móvel - Processo de <i>login</i> . . . . .   | 93  |
| 57 | Aplicativo móvel - Redefinição da senha do usuário . . . . .  | 94  |
| 58 | Região de ponto de permanência . . . . .  | 95  |
| 59 | Aplicativo móvel - Atividade apresentada na linha do tempo do usuário . . . .   | 95  |
| 60 | Aplicativo móvel - Processo de alteração de local . . . . .   | 96  |
| 61 | Aplicativo móvel - Representação através de gráfico pizza . . . . .   | 97  |
| 62 | Aplicativo móvel - Representação através de gráfico de barras . . . . .   | 98  |
| 63 | Aplicativo móvel - Rede de vínculos formalizados com os demais usuários . .   | 98  |
| 64 | Aplicativo móvel - Relação de locais frequentados pelo usuário . . . . .  | 99  |
| 65 | Aplicativo móvel - Processo de convite de novos usuários . . . . .  | 100 |
| 66 | Aplicativo móvel - Processo de alteração de senha do usuário . . . . .  | 100 |
| 67 | Aplicativo móvel - Processo de <i>logout</i> . . . . .  | 101 |

## LISTA DE TABELAS

|    |  |     |
|----|--|-----|
| 1  | Comparativo entre recursos utilizados na concepção dos trabalhos relacionados                                  | 48  |
| 2  | Processo de decomposição da localização em intervalos de horários . . . . .                                    | 53  |
| 3  | Processo de cálculo da similaridade entre dois usuários . . . . .  | 56  |
| 4  | Atribuição de valores na matriz de similaridades $S$ . . . . .   | 56  |
| 5  | Atribuição das preferências da matriz de similaridades $S$ . . . . .   | 56  |
| 6  | Taxa de utilização, por parte de usuários reais, do aplicativo móvel desenvolvido . . . . .                    | 61  |
| 7  | Faixas de renda consideradas na geração de dados artificiais . . . . .   | 64  |
| 8  | Formações acadêmicas consideradas na geração de dados artificiais . . . . .                                    | 64  |
| 9  | Dias da semana de estudo considerados na geração de dados artificiais . . . . .                                | 65  |
| 10 | Turnos de estudo considerados na geração de dados artificiais . . . . .  | 65  |
| 11 | Cursos considerados na geração de dados artificiais . . . . .  | 66  |
| 12 | Dias de trabalho considerados na geração de dados artificiais . . . . .  | 67  |
| 13 | Horário de início das jornadas de trabalho consideradas na geração de dados artificiais . . . . .              | 67  |
| 14 | Profissões consideradas na geração de dados artificiais . . . . .  | 68  |
| 15 | Locais de lazer, e suas distribuições de probabilidade, considerados na geração de dados artificiais . . . . . | 68  |
| 16 | <i>Clusters</i> mais expressivos obtidos no cenário de avaliação 1 . . . . .                                   | 71  |
| 17 | <i>Clusters</i> mais expressivos obtidos no cenário de avaliação 2 . . . . .                                   | 75  |
| 18 | <i>Clusters</i> mais expressivos obtidos no cenário de avaliação 3 . . . . .                                   | 81  |
| 19 | Semânticas de locais fornecidos pela API Google Maps . . . . .   | 104 |
| 20 | Relação de atributos que constituem o perfil dos usuários . . . . .  | 106 |
| 21 | Cenários previamente avaliados . . . . .   | 108 |



## **LISTA DE SIGLAS**

- API Application Programming Interface
- CSV Comma-Separated Values
- GPS Global Position System
- PCA Principal Component Analysis



## SUMÁRIO

|          |  |            |
|----------|--|------------|
| <b>1</b> | <b>INTRODUÇÃO</b>  | <b>19</b>  |
| 1.1      | Contribuição   | 21         |
| 1.2      | Objetivos  | 21         |
| 1.3      | Estrutura  | 22         |
| <b>2</b> | <b>FUNDAMENTAÇÃO TEÓRICA</b>   | <b>23</b>  |
| 2.1      | Clusterização  | 23         |
| 2.2      | Visualização   | 28         |
| 2.3      | Redes Sociais  | 35         |
| <b>3</b> | <b>TRABALHOS RELACIONADOS</b>  | <b>39</b>  |
| 3.1      | <i>MobiVis: A Visualization System for Exploration Mobile Data</i>           | 39         |
| 3.2      | <i>Eigenbehaviors: identifying structure in routine</i>                      | 43         |
| 3.3      | <i>Findind Similar Users Using Category-Based Location History</i>           | 45         |
| 3.4      | Comparativo entre trabalhos relacionados                                     | 48         |
| <b>4</b> | <b>MÉTODO</b>  | <b>51</b>  |
| 4.1      | Clusterização  | 51         |
| 4.2      | Visualização   | 57         |
| 4.3      | Captura das localizações dos usuários  | 60         |
| <b>5</b> | <b>ANÁLISE DOS RESULTADOS</b>  | <b>63</b>  |
| 5.1      | Geração de dados artificiais   | 63         |
| 5.2      | Cenários de avaliação  | 70         |
| <b>6</b> | <b>CONSIDERAÇÕES FINAIS</b>  | <b>85</b>  |
|          | <b>APÊNDICE A – APLICATIVO MÓVEL</b>   | <b>89</b>  |
|          | <b>APÊNDICE B – SEMÂNTICAS DE LOCAIS FORNECIDOS PELA API GOOGLE MAPS</b>     | <b>103</b> |
|          | <b>APÊNDICE C – ATRIBUTOS UTILIZADOS NA MODELAGEM DE PERFIL DOS USUÁRIOS</b> | <b>105</b> |
|          | <b>APÊNDICE D – CENÁRIOS PREVIAMENTE AVALIADOS</b>                           | <b>107</b> |
|          | <b>REFERÊNCIAS</b>   | <b>109</b> |



## 1 INTRODUÇÃO

*Digital shadow* refere-se ao conjunto de atividades digitais rastreáveis, ações, contribuições e comunicações que são manifestados na internet ou em dispositivos digitais (GANTZ; REINSEL, 2011) (SARMA; GIRÃO, 2009). Este rastreamento, aliado a uma integração e análise dos dados, tem sido utilizado por editores e anunciantes de *sites online* com o objetivo de aumentar a eficácia de seus anúncios, utilizando informações sobre o comportamento de navegação dos usuários na *web*.

Quando consumidores visitam um *site*, as páginas que visitam, a quantidade de tempo que visualizam cada página, os *links* que acessam, as pesquisas que realizam, e os recursos com os quais interagem, permitem aos *sites* coletar dados, além de outro fatores, para criar um “perfil” (SARMA; GIRÃO, 2009). Como resultado, editores de *site* podem usar estes dados para agrupar pessoas baseado em visitantes que apresentem perfis similares.

Esta quantidade de informação é enorme, porém insuficiente para analisar todas as facetas da personalidade de um indivíduo. Os históricos de contextos dos locais visitados pelos indivíduos pode ser considerado uma derivação de *digital shadow*, uma vez que permitem a extração de informação referente às características de tais locais durante um determinado período de tempo (BARBOSA et al., 2016) (PAPADOGKONAS; ROUSSOS; LEVENE, 2008). É extremamente comum, nos dias atuais, indivíduos utilizarem *smartphones* para diversos propósitos, criando assim uma variedade de dados, incluindo suas localizações. A utilização de dispositivos móveis, atuando como ferramenta de captura de dados de entrada, torna-se uma ótima alternativa, já que o indivíduo se encontra em um ambiente ubíquo (ALJAWARNEH; DHOMEJA; MALKANI, 2016) (JUNIOR; BARBOSA, 2015).

O conceito fundamental da computação ubíqua está na compreensão das características do ambiente em que o usuário esteja (BARBOSA, 2015), retratando assim, as reais situações pelas quais o mesmo esteja inserido. A ideia por trás deste conceito é realizar o processamento com a mínima exigência da interação com o indivíduo (WEISER, 2002). Em diversos casos, o objetivo de sistemas que utilizam, em sua concepção, recursos de ubiquidade é identificar e analisar o comportamento social humano ou compreender a interação sobre o ambiente monitorado (PAPADOGKONAS; ROUSSOS; LEVENE, 2008). A localização dos indivíduos pode ajudar o *marketing* comportamental a alcançar pessoas baseado no contexto em que ele se encontra naquele exato momento.

O trabalho aqui proposto, por sua vez, talvez possa ser considerado a união entre os conceitos de *digital shadow* e análise dos históricos de contexto de indivíduos.

Quando analisamos uma população com o objetivo de identificar o seu padrão de comportamento, os contextos por ela frequentados, bem como o tempo permanecido neles permitem o agrupamento dos indivíduos por afinidades relacionadas ao seu estilo de vida.

Por exemplo, podemos identificar com que frequência a população de uma cidade costuma ir à restaurantes. Além disto, pode-se identificar se estas pessoas são saudáveis através da

frequência com que vão à academias ou hospitais. Outra possibilidade é a análise de suas jornadas de trabalho (se são jornadas matutinas, vespertinas, ou noturnas). Através destas análises, um investidor que chega em uma nova cidade pode decidir que tipo de negócio ele pode iniciar e em que local. Ele pode observar, por exemplo, que as pessoas que moram em um bairro estão comprando em outro lugar porque há uma falta de lojas no bairro onde habitam.

O governo das cidades também pode criar rotas (ou alterá-las em determinado período do dia) que permitam um fluxo de tráfego eficiente através da observação da integração do tempo que toda a população gasta em engarrafamentos ou deslocando-se em determinados horários. Além disso, empresas, escolas e escritórios públicos podem alternar seus horários para colaborar para um melhor tráfego.

Para o agrupamento de indivíduos em diferentes conjuntos, mantendo em um mesmo grupo os indivíduos que apresentem semelhanças de comportamento entre si, foi usada a técnica de clusterização *Affinity Propagation*, que atua baseada em exemplares (SHANG et al., 2012). A técnica não faz avaliação da estrutura dos objetos analisados, apenas recebe, como dados de entrada, uma matriz contendo os indicadores de similaridade entre objetos envolvidos na análise (FREY; DUECK, 2007) (XIA et al., 2008) (SHEA; HASSANABADI; VALAEE, 2009). A utilização desta técnica não requer o número específico de *clusters* a serem originados como resultado da análise realizada (FREY; DUECK, 2007) (XIA et al., 2008). Contudo, para a formalização dos *clusters*, o procedimento adotado na concepção desta técnica consiste na troca de mensagens entre os pontos de dados analisados, transmitindo informações, através de fórmulas específicas, do quão bem um ponto de dado serve para ser o exemplar do outro, assim como quão bom seria para o um ponto escolher outro como seu exemplar (FREY; DUECK, 2007).

Com o objetivo de melhorar a compreensão sobre os dados capturados e sua clusterização, este trabalho utiliza a técnica de análise visual de informações denominada Coordenadas Paralelas, popularizada a partir de trabalhos desenvolvidos por Inselberg (1985) e Wegman (1990). A finalidade deste procedimento é prover uma visualização de dados multivariados e altamente dimensionados em um espaço bidimensional (HEINRICH; WEISKOPF, 2013) (ZHOU et al., 2008). Contudo, a visualização através do sistema de Coordenadas Paralelas, assim como outra ferramenta de análise visual, está intrinsecamente relacionada à capacidade cognitiva do analista. Tal representação permite que a interação humana ocorra sobre as amostras visualizadas e sobre os eixos do sistema, dependendo da análise a ser realizada (HEINRICH; WEISKOPF, 2013).

Com o advento da internet e com o avanço das tecnologias computacionais, as redes sociais têm recebido, nas últimas décadas, uma maior notoriedade, possibilitando o surgimento de uma grande gama de aplicações que atuam sobre as redes sociais (VADISALA; VATSAVAYI, 2017) (HAJIBAGHERI et al., 2012). A estrutura formalizada por uma rede social é composta por atores e as relações que os conectam, as quais representam interesses em comum entre os atores (VADISALA; VATSAVAYI, 2017) (ELBARAWY; MOHAMED; GHALI, 2014).

Através da estrutura formada no âmbito das redes sociais, torna-se possível a realização

de análises dos componentes que a compõem, proporcionando a exploração das características que permeiam a estrutura, como padrões de relações sociais existentes (CURRAN; CURRAN, 2014). Dentre os processos de análise de redes sociais, a detecção de comunidades, presentes na estrutura da rede, objetiva a identificação de grupos de indivíduos que compartilhem interesses entre si (BOUHALI; ELLOUZE, 2015) (ASMI; LOTFI; EL MARRAKI, 2016). Este trabalho tem por objetivo agrupar pessoas de maneira espontânea baseada em seus hábitos e comportamentos.

## 1.1 Contribuição

A contribuição deste trabalho se situa na criação de um modelo para análise comportamental e agrupamento de uma população de pessoas, baseado em hábitos cotidianos de frequência e permanência em locais de seu interesse, analisados de forma semântica.

Não há, em trabalhos relacionados à ideia proposta, trabalho cujo objetivo seja idêntico à proposta deste trabalho. Existem trabalhos que atuam na visualização e análise dos hábitos de uma população, porém limitando-se à um grupo restrito de pessoas relacionadas a um contexto, não podendo analisar, desta forma, todo o comportamento da população. Por outro lado, existem trabalhos que identificam a similaridade entre os indivíduos que compõem uma população baseado na semântica dos locais em que estes usuários estiveram presentes, mas não disponibilizam uma forma de visualização das informações, não permitindo a realização de uma análise da correlação de seus dados.

Para que uma análise de comportamento de população direcionada à um objetivo específico não se torne intratável devido ao tamanho dos bancos de dados, sua complexidade e sua especificidade relacionada à cultura local, é necessário que a mesma seja realizada de forma casada entre homem e máquina. Nenhum trabalho, até o momento, supriu esta necessidade utilizando-se de técnicas de análise visual como a proposta neste trabalho.

## 1.2 Objetivos

O objetivo geral deste trabalho é propor um modelo de análise de comportamento e agrupamento de uma população de pessoas, considerando os hábitos cotidianos de permanência em locais pré-classificados de seus indivíduos.

Destacam-se como objetivos específicos deste trabalho:

- Revisar trabalhos relacionados a análise comportamental de população.
- Analisar e avaliar os conceitos básicos de redes sociais, clusterização e análise visual para escolha das técnicas de desenvolvimento do trabalho.
- Agrupar indivíduos conforme suas similaridades de hábitos cotidianos de acordo com a permanência em locais de interesse.

- Construir ferramenta de análise visual dos dados.
- Capturar dados referentes ao histórico de contexto de indivíduos. Fabricar e analisar dados artificiais que representem usuários e informações a respeito de sua permanência em diferentes localizações.
- Avaliar o modelo com dados artificiais, os quais representem os hábitos de uma população.

O escopo deste trabalho não considera aspectos de privacidade nas capturas de localidades pelas quais o usuário encontra-se inserido, ou seja, são coletadas todas as localidades que o mesmo estiver presente. Contudo, este trabalho não tem o intuito de criar estereótipos para as pessoas com base nos dados coletados a partir do dispositivo móvel do usuário. Com isso, suas funcionalidades estão restritas a apresentar ao usuário, através do aplicativo móvel desenvolvido, o seu histórico de localizações frequentadas, bem como utilizar os dados coletados como entrada da ferramenta de análise proposta neste trabalho.

### **1.3 Estrutura**

Este trabalho está organizado em mais cinco capítulos. No capítulo seguinte é abordada a fundamentação teórica realizada na concepção deste trabalho, evidenciando as características das técnicas e conceitos adotados neste trabalho, sendo clusterização, visualização e redes sociais. No terceiro capítulo são evidenciados os trabalhos relacionados à ideia proposta neste trabalho. No quarto capítulo é apresentado o modelo proposto, evidenciando os métodos utilizados no desenvolvimento das ferramentas desenvolvidas na concepção deste trabalho. O quinto capítulo apresenta a análise dos resultados, oriundos da geração de dados artificiais, buscando a avaliação do modelo proposto no trabalho. Por fim, o último capítulo apresenta as considerações finais do trabalho desenvolvido.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo é destinado à abordagem dos principais conceitos presentes no desenvolvimento deste trabalho, destacando suas fundamentais características referentes ao seu contexto. O tema referente ao processo de clusterização visa fornecer subsídios para a completa compreensão do tema presente no trabalho proposto. O processo de análise visual, tema que compreende o principal objetivo deste trabalho, é abordado no intuito de explicar suas funcionalidades e como é aplicada ao seu contexto. Por fim, o tema de redes sociais é explanado em virtude deste trabalho possibilitar a geração de uma rede social espontânea, através do processo de clusterização.

### 2.1 Clusterização

Nesta seção são apresentados os principais conceitos relativos ao processo de clusterização, ressaltando a sua importância na área de mineração de dados. Em seguida, é apresentado o processo de clusterização *Affinity Propagation*, método adotado na concepção deste trabalho. Nesta seção, são retratadas as principais características deste método, evidenciando a forma como é realizado o funcionamento da técnica. Dentre as peculiaridades deste método, o fato de não necessitar a definição do número de *clusters* a serem gerados no processo foi determinante para sua adoção no desenvolvimento deste trabalho.

A técnica de clusterização *Affinity Propagation* está presente em diversas áreas do conhecimento. Dentre os cenários de aplicação, a técnica foi utilizada na classificação de imagens, em trabalhos propostos por Zhang, Tang e Corpetti (2016), Jia et al. (2008) e Dueck e Frey (2007). Na segmentação de imagens, o método foi aplicado por Xiao et al. (2007). No âmbito de classificação de estruturas genéticas, o algoritmo foi empregado em trabalhos propostos por Meng et al. (2017) e Hong et al. (2017). Tal método também foi utilizado na clusterização de textos por Guan et al. (2011). Além destas aplicações, a técnica também foi empregada na clusterização de informações sobre mobilidade veicular, em trabalhos propostos por Shahwani et al. (2017) e Shea, Hassanabadi e Valaee (2009). Alguns trabalhos desenvolvem uma extensão baseada na técnica *Affinity Propagation* para uma melhor adequação às necessidades do trabalho proposto.

#### 2.1.1 Mineração de dados

Diante do atual cenário em que estamos inseridos, é notório o grande aumento no volume de dados gerados nos últimos anos<sup>1</sup> (ARORA; CHANA, 2014) (PALMA et al., 2008). Devido à este grande volume de dados, o termo “Big Data” está presente em várias áreas do conheci-

---

<sup>1</sup>Diariamente, 2.5 quintilhões de *bytes* de dados são produzidos e 90% do volume de dados mundial foi produzido dentro dos últimos dois anos (IBM, 2017).

mento, sendo caracterizado por três aspectos: velocidade, variedade e volume (SOWMYA; SUNEETHA, 2017). Diante do exposto, ferramentas capazes de encontrar, analisar e proporcionar a visualização das informações obtidas com o processamento dos dados se fazem desejáveis (SOWMYA; SUNEETHA, 2017) (PALMA et al., 2008).

A mineração de dados consiste no processo de extração de informação útil ou relações escondidas entre as informações contidas no conjunto de dados analisado, com o intuito de proporcionar o conhecimento necessário para a tomada de decisão em ações futuras (ARORA; CHANA, 2014). Clusterização é uma das técnicas presentes no processo de mineração de dados (ARORA; CHANA, 2014) (XIA et al., 2008).

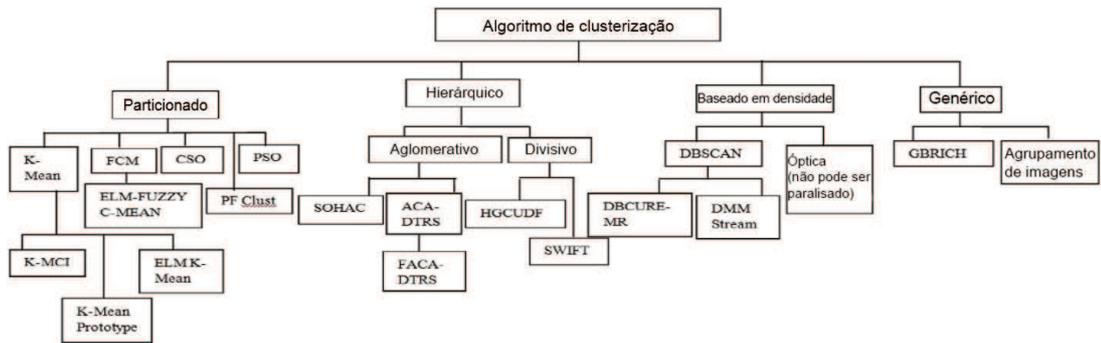
A técnica de clusterização objetiva o agrupamento de dados em diferentes conjuntos de acordo com suas características. Através desta técnica, os objetos que apresentam similaridades entre si são posicionados em um mesmo grupo, enquanto os objetos que apresentam dissimilaridades permanecem em grupos distintos (ARORA; CHANA, 2014) (BERKHIN, 2006). Assim, clusterização é denominada como um processo não supervisionado de aprendizado de conceitos escondidos presentes nos dados. Ressalta-se que a representação de dados através de um número reduzido de *clusters* implica, necessariamente, na perda de detalhes na informação (BERKHIN, 2006).

Os algoritmos de clusterização são desenvolvidos, em sua maioria, seguindo três modelos: hierárquico, particionado e baseado em densidade (ARORA; CHANA, 2014), conforme evidenciado na Figura 1, que demonstra a evolução dos algoritmos de clusterização. O método hierárquico consiste em formar, na construção do *cluster*, um dendograma. Diante do fato de o *cluster* ser construído em forma hierárquica, este modelo permite analisar os dados em diferentes níveis (BERKHIN, 2006) (XU; WUNSCH, 2005). O método hierárquico pode ser classificado em aglomerativo ou divisivo (ARORA; CHANA, 2014) (BERKHIN, 2006) (XU; WUNSCH, 2005). O método particionado busca a divisão dos objetos dentre os  $k$  grupos (BERKHIN, 2006). Contudo, este método exige a definição do número de *clusters* a serem obtidos (JAIN; MURTY; FLYNN, 1999). No método baseado em densidade, um *cluster* é formado de acordo com o nível de concentração de observações. Este método possui a vantagem de ser capaz de descobrir *clusters* em uma forma irregular de distribuição (BERKHIN, 2006).

De acordo com Xu e Wunsch (2005), o processo de clusterização passa por quatro etapas básicas, conforme ilustrado na Figura 2:

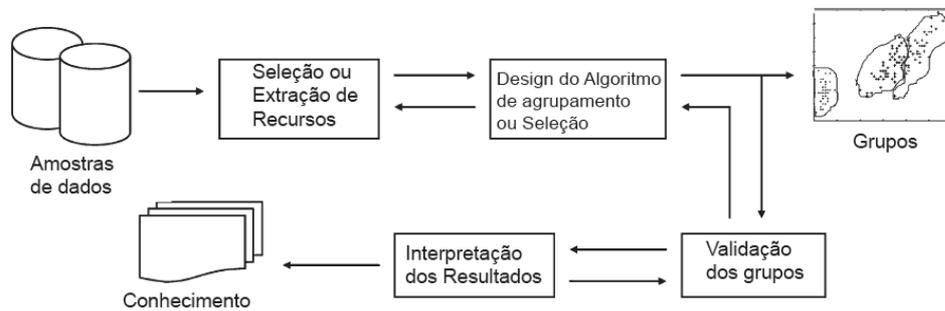
1. Seleção ou extração de características: a seleção de características visa a identificação de características distintas a partir do conjunto de observações, enquanto a extração faz uso de transformações para gerar novas características a partir das originais. Ambas abordagens são extremamente importantes para o processo, a definição correta das características implica em um melhor desempenho durante o processo de clusterização.
2. Desenvolvimento ou seleção do algoritmo de clusterização: a partir da análise das características do problema a ser estudado, faz-se necessária a avaliação do algoritmo de

Figura 1: Evolução dos algoritmos de clusterização



Fonte: Adaptado de Arora e Chana (2014)

Figura 2: Etapas do processo de clusterização



Fonte: Adaptado de Xu e Wunsch (2005)

clusterização a ser utilizado. Esta etapa é comumente executada em conjunto com a definição da medida de proximidade entre as observações contidas no conjunto de dados.

3. Validação do *cluster*: a partir do conjunto de dados, cada algoritmo de clusterização pode gerar uma separação na estrutura. Além disto, diferentes abordagens podem implicar em diferentes resultados, ou seja, gerando *clusters* distintos. Com isto, a existência de formas de avaliação eficazes são extremamente importantes para informar aos usuários um grau de confiança sobre os *clusters* resultantes do processo executado a partir do algoritmo utilizado. Tal avaliação deve ser objetiva e não tendenciosa.
4. Interpretação dos resultados: o objetivo principal do processo de clusterização é fornecer aos usuários conhecimento significativo sobre os dados estudados, provendo subsídios para que os indivíduos possam aplicar este conhecimento na resolução do problema a ser avaliado.

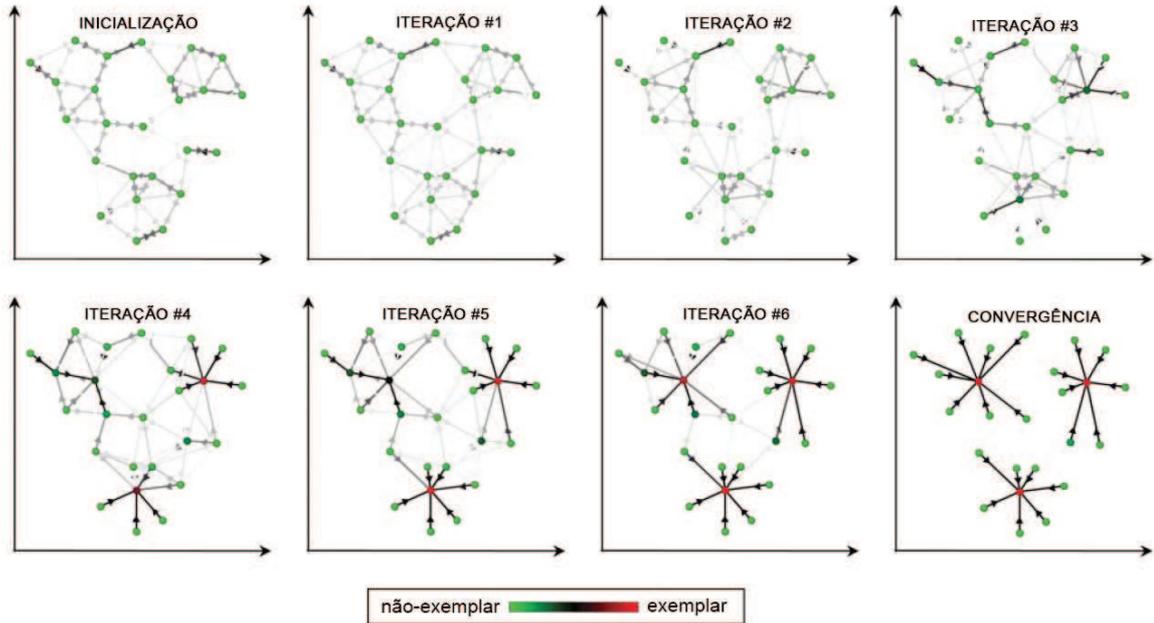
### 2.1.2 *Affinity Propagation*

Proposto por Frey e Dueck (2007), *Affinity Propagation* é um algoritmo de clusterização baseado em exemplares (SHANG et al., 2012). Como dados de entrada, o método recebe uma matriz de valores reais  $s$ , que mapeia a similaridade entre pontos de dados. No algoritmo,  $s(i, k), i \neq k$  indica o valor da similaridade computada entre os pontos  $i$  e  $k$  (FREY; DUECK, 2007) (XIA et al., 2008) (SHEA; HASSANABADI; VALAEE, 2009).

Esta técnica é caracterizada por não requisitar o número de *clusters* a serem formalizados. Em virtude disto, o método recebe um conjunto de valores reais  $s(k, k)$  para cada ponto  $k$  da análise. Tais valores são denominados “preferências” (FREY; DUECK, 2007) (XIA et al., 2008). O valor estipulado para as preferências influencia no número de *clusters* a serem originados (FREY; DUECK, 2007) (SHEA; HASSANABADI; VALAEE, 2009) (ZHANG; TANG; CORPETTI, 2016). Em casos onde o indicado seja o algoritmo considerar todos os pontos como possíveis exemplares, o valor de preferência estipulado para cada ponto deve ser o mesmo. Em relação à definição da preferência, quanto menor o valor, menor será o número de *clusters* formados e, quanto maior o valor, conseqüentemente, o número de *clusters* formalizados será maior. A partir de tal concepção, a utilização do valor da mediana dos valores da matriz  $s$  resulta em um número moderado de *clusters* (FREY; DUECK, 2007). Logo, definindo o valor de preferência como zero, ao final do processo cada ponto será o seu próprio exemplar.

O processo de clusterização desta técnica consiste na troca de mensagens, de forma iterativa, entre os pontos de dados durante um número de iterações estabelecido ou até o momento em que a definição de exemplares mantenha-se constante ao longo de um número determinado de iterações (FREY; DUECK, 2007), conforme demonstrado na Figura 3, que demonstra o processo de formalização dos *clusters* através das iterações realizadas. Existem dois tipos de mensagens transmitidas entre os pontos de dados durante o processo. A primeira, denominada

Figura 3: *Affinity Propagation* - Iterações até a convergência para formalização dos *clusters*



Fonte: Adaptado de Frey e Dueck (2007)

“responsabilidade”  $r(i, k)$ , é uma mensagem enviada do ponto  $i$  para o ponto  $k$ , candidato a ser o ponto exemplar, indicando o quão adequado o ponto  $k$  é para tornar-se o ponto exemplar de  $i$ . A segunda, denominada “disponibilidade”  $a(i, k)$ , é uma mensagem enviada do ponto  $k$ , candidato a ser o ponto exemplar, para o ponto  $i$ , denotando o quão apropriado seria para o ponto  $i$  definir o ponto  $k$  como exemplar (FREY; DUECK, 2007) (XIA et al., 2008).

De acordo com Frey e Dueck (2007), o processo de clusterização previsto no algoritmo *Affinity Propagation*, ilustrado pela Figura 4a e Figura 4b, que demonstram o processo de troca de mensagens entre os pontos de dados analisados, utiliza diferentes fórmulas para definição do valor a ser passado de um ponto para outro nos diferentes tipos de mensagens. Inicialmente, a matriz de disponibilidades é inicializada com o valor zero:  $a(i, k) = 0$ . A partir das iterações, a matriz de responsabilidades é apurada de acordo com a fórmula:

$$r(i, k) \leftarrow s(i, k) - \max_{k' \text{ s.t. } k' \neq k} \{a(i, k') + s(i, k')\}$$

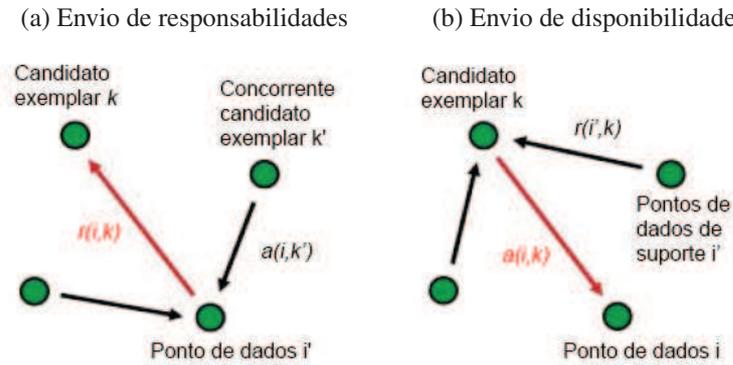
Por outro lado, a matriz de disponibilidades é computada a partir das fórmulas:

$$a(i, k) \leftarrow \min\{0, r(k, k) + \sum_{i' \text{ s.t. } i' \notin \{i, k\}} \max\{0, r(i', k)\}\}$$

$$a(k, k) \leftarrow \sum_{i' \text{ s.t. } i' \neq k} \max\{0, r(i', k)\}$$

O resultado das apurações de responsabilidade e disponibilidade combinadas fornece subsídio para que a técnica consiga determinar o exemplar de cada ponto contido na análise. Para

Figura 4: Processo de troca de mensagens entre os pontos de dados



Fonte: Adaptado de Frey e Dueck (2007)

um ponto  $i$ , o valor de  $k$  que maximiza o resultado de  $a(i, k) + r(i, k)$  determina o ponto  $i$  como sendo exemplar se  $k = i$ , ou identifica o exemplar para o ponto  $i$  (FREY; DUECK, 2007) (XIA et al., 2008).

## 2.2 Visualização

Esta seção contextualiza o processo de análise visual, demonstrando sua motivação, suas características e seus principais objetivos. Além disto, a metodologia utilizada no desenvolvimento deste processo é evidenciada. Posteriormente, é destacada a técnica de análise visual denominada por Coordenadas Paralelas, técnica utilizada na concepção deste trabalho, abordando suas principais características e o princípio do funcionamento da técnica.

Diversos trabalhos, em diferentes cenários, têm sido desenvolvidos voltados para a área de análise visual. No cenário de visualização de dados geográficos, técnicas de visualização têm sido empregadas em trabalhos propostos por Alonso, Espinoza-Molina e Datcu (2017), Yu et al. (2015), Kang et al. (2010), Andrienko, Andrienko e Wrobel (2007), Compieta et al. (2007), Andrienko e Andrienko (2005) e Kapler e Wright (2005). Trabalhos no âmbito de análise do comportamento humano com base em dados telefônicos têm sido propostos por Zheng et al. (2016) e Lu et al. (2016). Já na análise de estruturas de rede sociais, métodos de visualização são utilizados em trabalhos desenvolvidos por Heer e Perer (2014) e Bista, Nepal e Paris (2014). Por fim, trabalho proposto por Kumar e Garland (2006) atua na análise de dados variáveis com relação ao tempo.

Especificamente, a adoção da técnica de visualização de Coordenadas Paralelas tem sido amplamente utilizada em diversos cenários de atuação, diante do exposto em trabalhos propostos por Zhou et al. (2016), Zhang et al. (2014), Barlowe et al. (2008) e Rubel et al. (2008). No âmbito de análise de dados geográficos, a técnica foi utilizada em proposta desenvolvida por Jern e Franzen (2006). Já no contexto de descoberta de *clusters* presentes em dados analisados, a técnica foi empregada no trabalho proposto por Artero, Oliveira e Levkowitz (2004). Baciú,

Wang e Li (2016) propuseram o emprego de Coordenadas Paralelas na avaliação da utilização de recursos em plataformas de computação na nuvem. Wang et al. (2017) e Steed et al. (2013) desenvolveram propostas de utilização desta técnica em simulações climáticas, com o intuito de realizar a sua previsão. A técnica de Coordenadas Paralelas também tem sido utilizada na análise de dados forenses, em trabalho proposto por Wang et al. (2014). Além disto, esta técnica é utilizada na realização de análise de dados de clientes, como demonstrado em trabalho proposto por Zhu e Xu (2016).

### 2.2.1 Análise Visual

Atualmente, estamos lidando com altos volumes de dados gerados diariamente (CHENG et al., 2016) (KEIM et al., 2008a) (ANDRIENKO; ANDRIENKO; WROBEL, 2007). Geralmente, estes dados são armazenados sem a aplicação de um critério de filtragem das informações (KEIM et al., 2008a). Esta situação acarreta em uma realidade onde nossa capacidade de coletar e armazenar dados aumenta mais rápido do que a nossa habilidade de analisá-los (KEIM et al., 2006). Entretanto, o valor do dado em si não é tão valioso quanto a informação que pode ser extraída a partir deste dado através da análise realizada (KEIM et al., 2010).

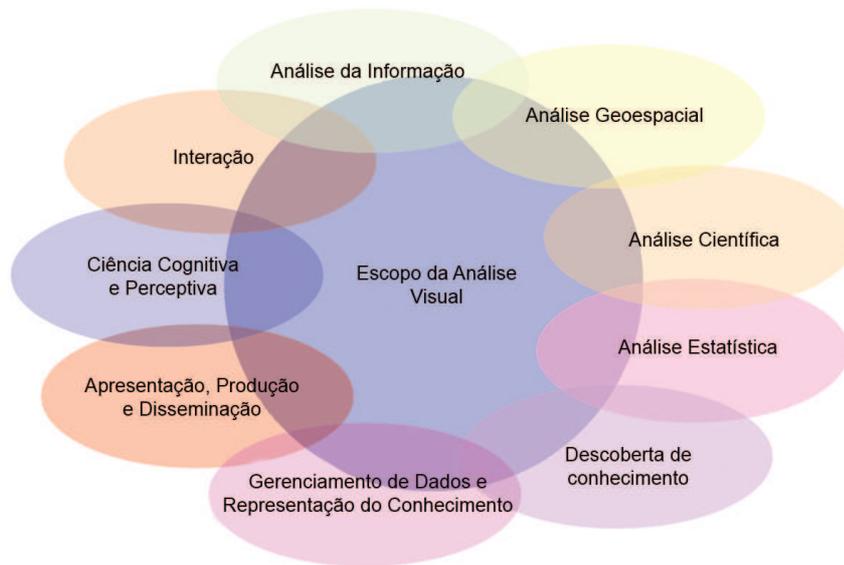
Diante disto, o processo de análise visual consiste em representar visualmente as informações, permitindo a direta interação humana com a representação, o que objetiva um melhor discernimento sobre as informações retratadas pelos dados, auxiliando na tomada das melhores decisões (KEIM et al., 2006) (KEIM et al., 2008a) (WU et al., 2016a). O termo “análise visual” começou a ser profundamente difundido a partir do ano de 2005 (COOK; THOMAS, 2005).

De modo geral, a análise visual objetiva dois propósitos: compreensão e previsão (ANDRIENKO; ANDRIENKO; WROBEL, 2007). Dessa forma, a abordagem utilizada na análise pode indicar dois cenários diferentes: novos problemas e novas oportunidades. O primeiro indica situações em que o problema não esteja suficientemente compreendido e/ou mal definido. O segundo, por sua vez, indica situações em que novos dados ou novas tecnologias são utilizados para resolver problemas existentes de uma melhor maneira (ANDRIENKO et al., 2017).

A análise visual é uma atividade interdisciplinar, sendo caracterizada por estar presente em, principalmente, três grupos de diferentes áreas do conhecimento, conforme demonstrado pela Figura 5 (JERN; FRANZEN, 2006). O primeiro grupo refere-se à visualização, englobando metodologias de análise da informação, geoespacial e científica (KEIM et al., 2008b). A visualização é o instrumento pelo qual a interação humana e os recursos computacionais atuam em conjunto para alcançar resultados efetivos (ANDRIENKO et al., 2010). O segundo grupo corresponde à fatores humanos, compreendendo áreas como interação, cognição, percepção, colaboração, apresentação e disseminação. Por fim, o último grupo é determinado pela análise de dados, através das áreas de gestão de dados e representação do conhecimento, descoberta de conhecimento e análise estatística (KEIM et al., 2008b).

Para a obtenção da descoberta de conhecimento, o processo de análise visual segue uma me-

Figura 5: Escopo do processo de análise visual



Fonte: Adaptado de Keim et al. (2008b)

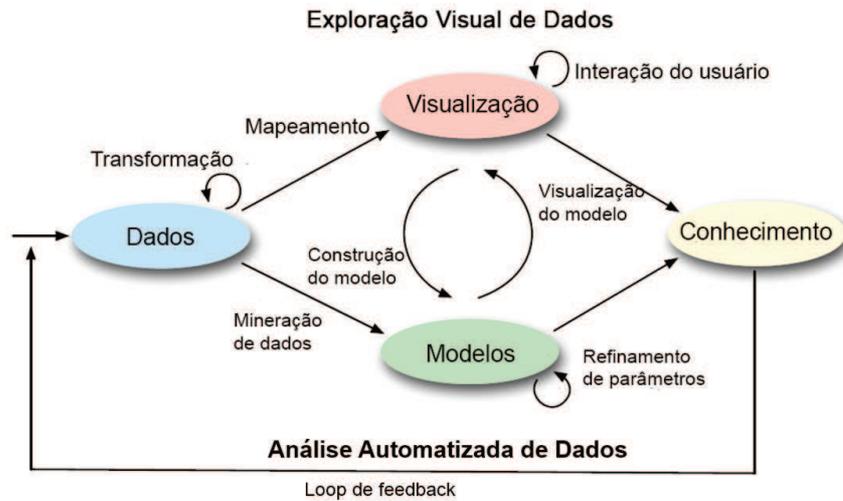
Metodologia conforme apresentado na Figura 6, que apresenta os diferentes estágios do processo e as transições que ocorrem entre eles. Este processo é determinado pelas interações realizadas entre os dados, as visualizações, os modelos constituídos a partir dos dados e os usuários (KEIM et al., 2010).

Diante do fato de os dados a serem analisados poderem ser oriundos de fontes heterogêneas, o primeiro estágio do processo consiste em realizar um pré-processamento dos dados e/ou transformá-los em outra estrutura melhor formulada para uma análise futura. A partir de tal operação, o processo pode seguir dois caminhos diferentes. O primeiro utiliza técnicas de mineração de dados para constituir modelos sobre os dados originais. Através da interação com os dados, este modelo pode ser avaliado e refinado. Por outro lado, a visualização permite que o usuário interaja através da modificação de parâmetros pertinentes à análise. Assim, a visualização de modelos pode ser utilizada para avaliar os modelos gerados. A realização alternada destes dois métodos é desejável para a obtenção de um contínuo aprimoramento dos resultados preliminares. Como resultado destes procedimentos, o processo de análise visual objetiva o ganho de conhecimento extraído dos dados analisados (KEIM et al., 2010).

Shneiderman (1996) defende que o processo de análise visual proceda através de uma inicial visão geral sobre os dados, uma posterior aplicação de filtro/zoom e, finalmente, a obtenção de detalhes a partir de um item ou subconjunto dos dados. Contudo, o processo de visualização de um grande volume de dados pode resultar em um não reconhecimento de interessantes padrões sobre os dados a serem analisados (KEIM et al., 2010).

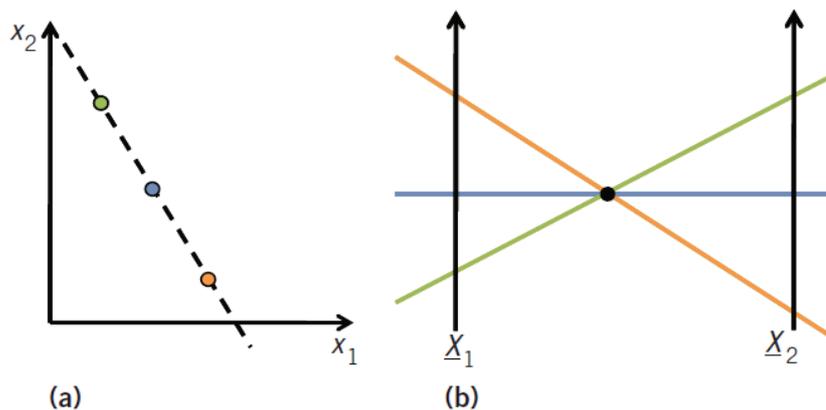
Existem vários desafios a serem alcançados na área de análise visual, tornando desejável o desenvolvimento e melhorias de técnicas e aplicações (KEIM et al., 2008b). Keim et al. (2008b) identifica diversas áreas em que o processo de análise visual pode ser melhor explorado, como por exemplo: física e astronomia, finanças, segurança, monitoramento de ambientes, gestão de

Figura 6: Processo de análise visual



Fonte: Adaptado de Keim et al. (2010)

Figura 7: Geometria das Coordenadas Paralelas



Fonte: Heinrich e Weiskopf (2015)

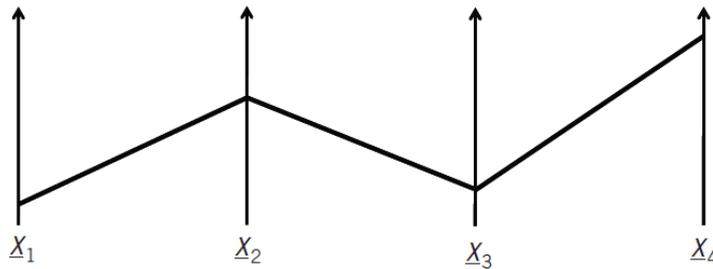
desastres e emergências, *softwares* de análise e, por fim, biologia, medicina e saúde.

### 2.2.2 Coordenadas Paralelas

A técnica de visualização de dados, denominada Coordenadas Paralelas, teve seu conceito originado por d'Ocagne (1885). No entanto, tal conceito somente foi popularizado a partir de trabalhos desenvolvidos por Inselberg (1985) e Wegman (1990). Este método atua na visualização de dados multivariados e altamente dimensionados, representando dados com  $n$  dimensões em um espaço bidimensional (HEINRICH; WEISKOPF, 2013) (ZHOU et al., 2008). O processo de transformação geométrica de pontos de dados para os segmentos de reta no contexto das Coordenadas Paralelas é ilustrado pela Figura 7.

Diante do fato do processo analítico de uma representação de Coordenadas Paralelas envolver fatores humanos, o sucesso da análise está intrinsecamente relacionado à capacidade

Figura 8: Múltiplas dimensões representadas por Coordenadas Paralelas



Fonte: Heinrich e Weiskopf (2015)

cognitiva do indivíduo que esteja realizando a análise. Dessa forma, a possibilidade de extração de informação útil a partir dos dados torna-se mais propensa através da análise de usuários experientes (HEINRICH; WEISKOPF, 2013).

#### 2.2.2.1 Construção

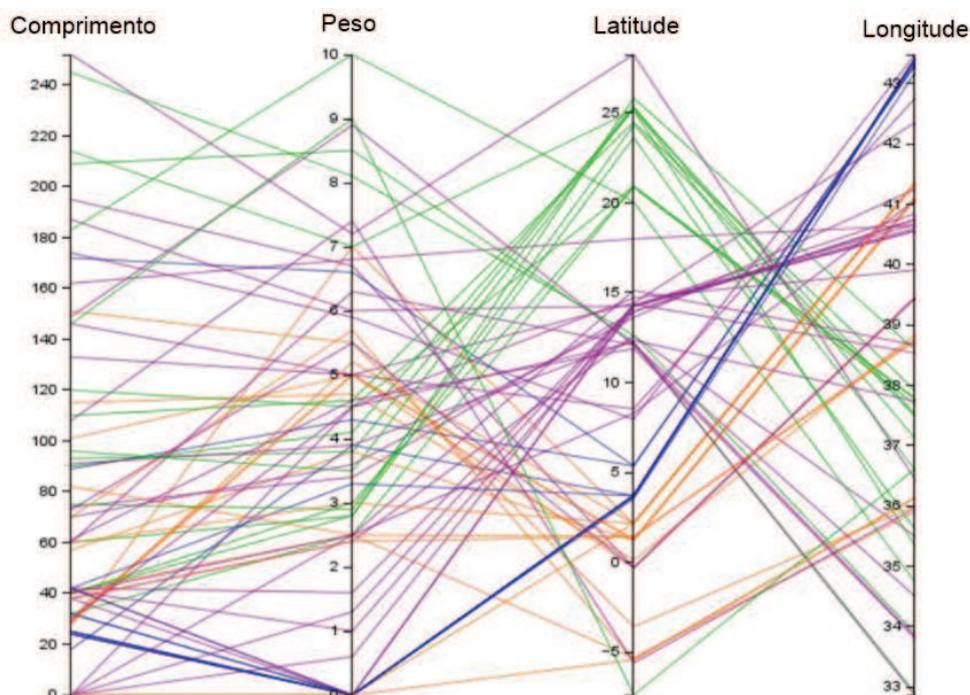
O objetivo desta técnica é transformar as  $n$  dimensões pertencentes ao dado em um conjunto de segmentos de reta transpassantes sobre os  $n$  eixos paralelos que, por sua vez, denotam cada uma das dimensões dos dados (HEINRICH; WEISKOPF, 2013) (SIIRTOLA, 2003). Conforme exemplificado na Figura 8, cada segmento de reta intersecta cada eixo na altura apropriada que represente o valor do dado na respectiva dimensão (BERTHOLD; HALL, 2003). Em virtude de tais características, situações que necessitem a inclusão de novas dimensões são facilmente resolvidas com a adição de novos eixos paralelos (HEINRICH; WEISKOPF, 2013) (SIIRTOLA, 2003).

A orientação dos eixos presentes no sistema de Coordenadas Paralelas pode ser apresentada em *layout* horizontal, sendo esta a forma comumente utilizada, ou vertical. A definição do *layout* depende das características do cenário de visualização, ou seja, o número de eixos, intervalo dos dados, a dimensão do dispositivo utilizado e preferência do usuário (SIIRTOLA, 2003). Os possíveis valores, compreendidos ao longo de cada eixo, são delimitados pelos valores mínimo e máximo da respectiva dimensão dentre os pontos de dados presentes no conjunto analisado (ANDRIENKO; ANDRIENKO, 2001), logo, o menor valor intersecta o eixo na parte inferior e o maior na parte superior (HEINRICH; WEISKOPF, 2013), conforme demonstrado pela Figura 9. Todos os eixos existentes da visualização devem estar uniformemente espaçados entre si (CUZZOCREA; ZALL, 2013) (HEINRICH; WEISKOPF, 2013).

#### 2.2.2.2 Interação

A geometria desenvolvida na representação de Coordenadas Paralelas permite que o usuário interaja de diferentes maneiras. A primeira consiste na interação sobre as amostras presentes

Figura 9: Limites de valores definidos para os eixos das Coordenadas Paralelas



Fonte: Adaptado de Vatin e Napoli (2013)

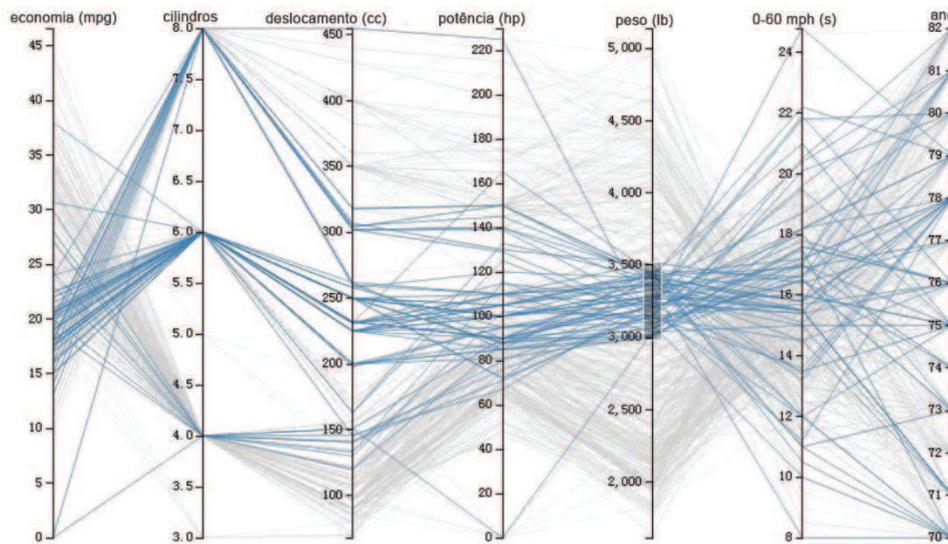
no conjunto de dados, através da técnica denominada *brushing*. Por outro lado, a representação também permite a interação sobre seus eixos que, por sua vez, pode ocorrer através das ações de translação e alteração na escala de valores (HEINRICH; WEISKOPF, 2013).

A técnica de *brushing* é forma de interação mais comum utilizada pelos usuários (HEINRICH; WEISKOPF, 2013) (SIIRTOLA, 2000). O método consiste na seleção de um intervalo de valores contidos nos eixos do sistema, evidenciando o subconjunto de dados que satisfazem os critérios do intervalo selecionado, indicado pelo retângulo gerado sobre o eixo, conforme demonstrado na Figura 10 (HUANG; WEI; DU, 2016) (HEINRICH; WEISKOPF, 2013).

O procedimento de translação consiste na mudança da ordem de apresentação dos eixos na representação das Coordenadas Paralelas. Tal operação torna-se desejável na investigação de padrões específicos presentes nos dados ou pela simples reordenação dos eixos (HEINRICH; WEISKOPF, 2013), conforme demonstrado na Figura 11, que exhibe o mesmo conjunto de dados, o qual é representado através de diferentes ordenações dos seus eixos.

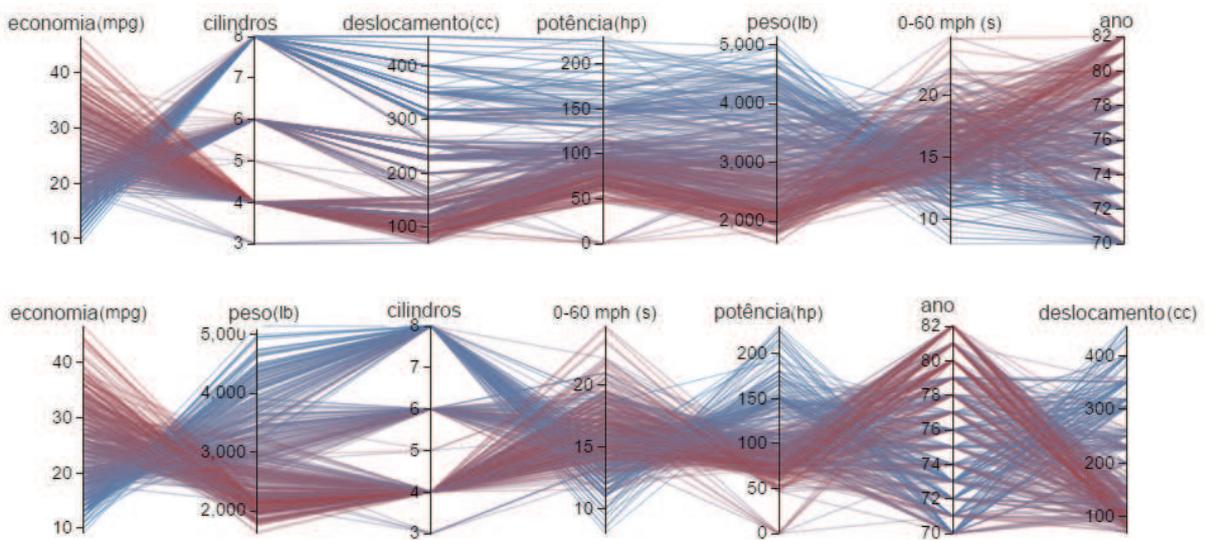
A mudança na escala objetiva a alteração da forma como os valores dos eixos do sistema de Coordenadas Paralelas são apresentados, buscando identificar padrões emergentes dos dados. Tal operação pode ser equiparada com a aplicação de uma função executada em todos os valores pertinentes ao eixo. Dentre as possíveis operações sobre os valores, a operação *flipping* realiza a negação da ordem com que todos os valores são visualizados, invertendo a relação dos valores maiores na parte superior e menores na parte inferior (HEINRICH; WEISKOPF, 2013), conforme demonstrado pela Figura 12, que aplica o método *flipping* no eixo referente à dimensão potência.

Figura 10: Técnica de *brushing* aplicada sobre os eixos das Coordenadas Paralelas



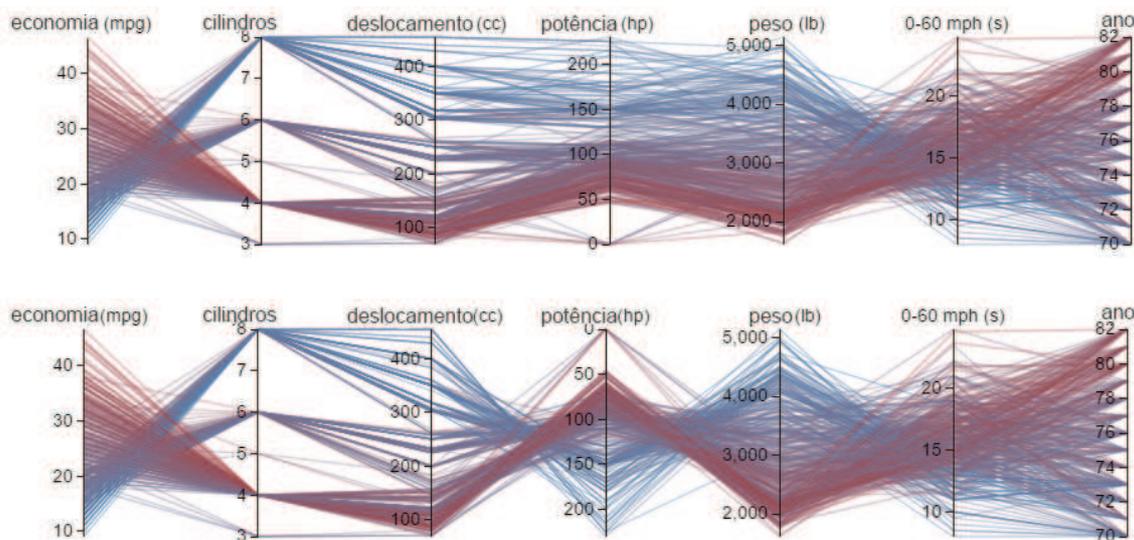
Fonte: Adaptado de Huang, Wei e Du (2016)

Figura 11: Translação dos eixos das Coordenadas Paralelas



Fonte: Elaborada pelo autor.

Figura 12: *Flipping* sobre o eixo potência das Coordenadas Paralelas



Fonte: Elaborada pelo autor.

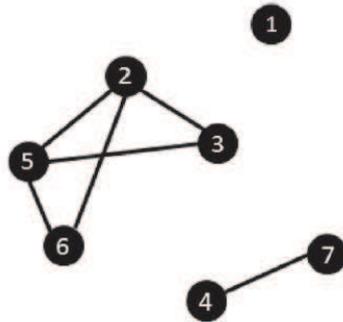
### 2.3 Redes Sociais

A utilização da expressão “rede” para representar um complexo sistema de relações sociais existentes entre indivíduos tem sido utilizada há muito tempo (MITCHELL, 1969). A partir do advento da internet, juntamente com a evolução das tecnologias computacionais, as redes sociais tornaram-se cada vez mais presentes, através de uma enorme variedade de aplicações (VADISALA; VATSAVAYI, 2017) (HAJIBAGHERI et al., 2012) (AGGARWAL, 2011) (BORGATTI et al., 2009). Redes sociais exercem uma importante função social, pois através da internet, facilitam a interação entre pessoas geograficamente distantes (GUESMI; TRABELSI; LATIRI, 2014).

A estrutura de uma rede social é representada, basicamente, por atores e relações, as quais conectam os atores com base nos interesses que compartilham, conforme ilustrado pela Figura 13 (VADISALA; VATSAVAYI, 2017) (ELBARAWY; MOHAMED; GHALI, 2014) (AGGARWAL, 2011) (HAJIBAGHERI et al., 2012) (NEWMAN, 2006) (JAMALI; ABOLHASANI, 2006). Atores podem ser representações de pessoas de um grupo, departamentos de uma empresa ou nações do mundo (WASSERMAN; FAUST, 1994). Geralmente, a estrutura de uma rede social apresenta-se de maneira dinâmica, ou seja, indivíduos juntam-se ou afastam-se da rede a partir de mudanças em seus interesses. Dessa forma, o aspecto temporal deve ser considerado na interpretação da estrutura da rede (TANTIPATHANANANDH; BERGER-WOLF, 2011). Contudo, o termo “rede social” não está restrito somente ao estudo de comunidades virtuais, este conceito está presente no campo da sociologia, analisando estruturas genéricas que envolvem interações entre atores (AGGARWAL, 2011).

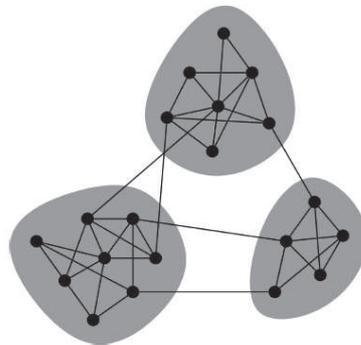
O cenário de uma rede social fornece subsídios para a aplicação de análise sobre a estrutura da rede. A análise de redes sociais pode ser utilizada para estudar as características presentes na

Figura 13: Representação de rede social



Fonte: Hao et al. (2015)

Figura 14: Representação de detecção de comunidades em rede social



Fonte: Newman (2006)

estrutura, como padrões de relações sociais, familiares e mobilidade social (CURRAN; CURRAN, 2014). Através de um amplo conhecimento da dinâmica social presente na rede, torna-se possível a utilização destas informações em aplicações práticas (LV; ZHANG; ZHANG, 2016). A detecção de comunidade é um grande desafio presente na área de análise de redes sociais (LIU et al., 2016) (HAO et al., 2015) (BOUJLALEB et al., 2015) (HAJIBAGHERI et al., 2012).

O processo de detecção de comunidade consiste em identificar, dentro da estrutura da rede social, grupos de pessoas que possuam um alto índice de relação, o que pode ser considerado um processo de clusterização dos indivíduos da rede em comunidades, de acordo com o demonstrado na Figura 14 (BOUHALI; ELLOUZE, 2015) (ASMI; LOTFI; EL MARRAKI, 2016) (GUESMI; TRABELSI; LATIRI, 2014). O processo de detecção de comunidades considera que uma rede social é, presumivelmente, subdividida em grupos. O número e tamanho das comunidades originadas depende exclusivamente das características da rede (NEWMAN, 2006).

Diversas abordagens têm sido utilizadas na concepção de algoritmos para detectar comunidades dentro de uma rede (ASMI; LOTFI; EL MARRAKI, 2016) (LIU et al., 2016) (LIU; LIU; JIANG, 2014) (GUESMI; TRABELSI; LATIRI, 2014). No entanto, a maioria das técnicas de detecção de comunidades em redes sociais utilizam, principalmente, informações sobre a estrutura topológica. Em casos onde o processo de detecção de comunidades esteja sendo

aplicado à redes sociais baseadas em localização, as informações sobre a posição geográfica dos indivíduos, juntamente com as informações referentes à topologia da estrutura, tornam-se imprescindíveis para o correto processo de descoberta de comunidades (LIU et al., 2016).

Dentre diferentes domínios de aplicação, processos de detecção de comunidades a partir de redes sociais estão altamente difundidos dentro da comunidade acadêmica, sendo utilizados na concepção de trabalhos propostos por Asmi, Lotfi e El Marraki (2016), Liu et al. (2016), Krishnan, Karthika e Bose (2016), Lv, Zhang e Zhang (2016), Li et al. (2016), Boujlaleb et al. (2015), Hao et al. (2015), Liu, Liu e Jiang (2014), ElBarawy, Mohamed e Ghali (2014), Guesmi, Trabelsi e Latiri (2014), Huang et al. (2013), Hajibagheri et al. (2012) e Chang e Lin (2010).



### 3 TRABALHOS RELACIONADOS

Na comunidade acadêmica, diversos trabalhos que utilizam recursos de geolocalização têm sido desenvolvidos buscando a compreensão de padrões sociais através das atividades diárias de usuários, como por exemplo, os trabalhos propostos por Eagle e Pentland (2006), Gonzalez, Hidalgo e Barabasi (2008) e Cho, Myers e Leskovec (2011). A utilização de dados coletados a partir de dispositivos móveis, proporcionando uma análise comportamental humana sobre tais dados, foi evidenciada nas propostas de Liu e Magno (2015) e Laasonen (2005). A inferência de relações entre diferentes indivíduos, através da utilização de serviços baseados em localização, está amplamente presente, conforme evidenciado pelos trabalhos de Navarro et al. (2015), Zheng et al. (2011), Li et al. (2008), Boujlaleb et al. (2015) e Wu et al. (2016b). Recursos de visualização de dados têm sido elaborados, com o intuito de gerar conhecimento, na concepção de trabalhos propostos por Kang et al. (2010), Kumar e Garland (2006) e Kapler e Wright (2005).

Neste capítulo são descritos três trabalhos relacionados à ideia proposta neste trabalho, os quais fazem uso de dados georreferenciados para disponibilizar uma forma de agrupamento e interpretação do comportamento dos indivíduos de uma população. Os trabalhos relacionados detalhados foram selecionados mediante a característica fundamental de considerar, em sua concepção, dados geográficos do histórico de contexto dos indivíduos que, traduzidos, refletem os seus interesses de acordo com a perspectiva semântica dos locais frequentados. A partir de tal característica, considerou-se trabalhos que apresentassem primordialmente, em seu desenvolvimento, diretivas que apontassem para um processo de agrupamento de indivíduos considerando os interesses destas pessoas, representados pelas características dos locais frequentados. Adicionalmente, embora não obrigatoriamente, foram considerados trabalhos que apresentassem um processo de análise visual do comportamento dos indivíduos seguindo as mesmas características.

O primeiro trabalho apresentado, proposto por Shen e Ma (2008), apresenta o projeto *MobiVis*, que analisa padrões de comportamento de um seletivo grupo de indivíduos, sendo estes estudantes do MIT<sup>1</sup>. O segundo trabalho, proposto por Eagle e Pentland (2009), refere-se ao projeto *Eigenbehaviors*, o qual utiliza a técnica de análise de componentes principais para interpretar os hábitos de interesse de uma população, também representada por um grupo de alunos do MIT. O terceiro trabalho relacionado, idealizado por Xiao et al. (2010), aborda um projeto que realiza uma análise, de forma semântica, dos locais de interesse dos indivíduos.

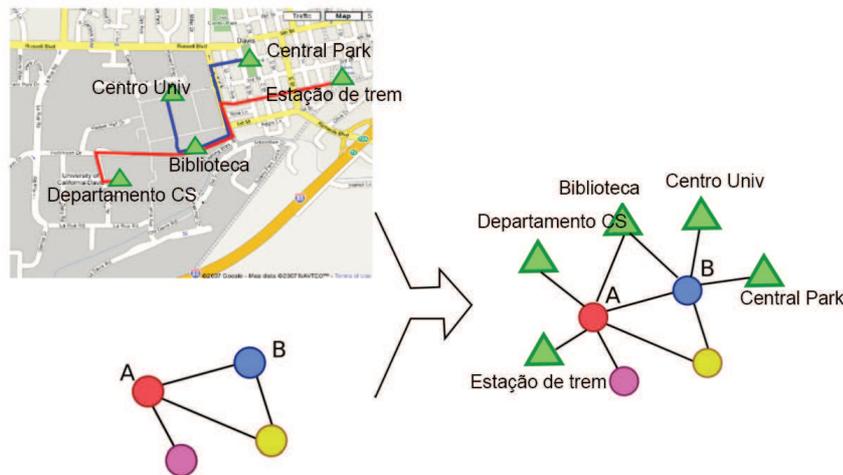
#### 3.1 *MobiVis: A Visualization System for Exploration Mobile Data*

Diante da alta disseminação do uso de dispositivos móveis por parte da população e, das oportunidades que surgem através da sua utilização, Shen e Ma (2008) apresentam neste traba-

---

<sup>1</sup>Instituto de Tecnologia de Massachussets

Figura 15: Rede heterogênea apresentando informações sociais e espaciais



Fonte: Adaptado de Shen e Ma (2008)

lho o MobiVis, um sistema de análise visual desenvolvido para explorar e compreender a relação existente entre indivíduos no âmbito sócio-espacio-temporal. Com a coleta e análise de dados obtidos através do dispositivo móvel dos usuários, torna-se possível a descoberta de comunidades, o entendimento de comportamentos sociais e a inferência de relações entre determinados eventos.

Dados coletados de dispositivos móveis retratam informações espaciais e sociais. Enquanto dados de GPS registram as localizações geográficas dos indivíduos, a proximidade *bluetooth* e as chamadas telefônicas realizadas podem representar relações sociais entre determinados indivíduos. A proposta consiste em visualizar as informações em um grafo não direcionado. Para as informações sociais, os dados de proximidade *bluetooth* e chamadas telefônicas são representadas pelas arestas, sendo variáveis em relação ao tempo, enquanto os vértices remetem aos indivíduos. MobiVis incorpora os grafos social e espacial em uma única visualização em uma estrutura de rede, conforme demonstrado na Figura 15. A visualização de uma rede heterogênea variável no tempo é o objetivo principal do projeto MobiVis.

O conjunto de dados utilizados no experimento foi obtido através da utilização de um *software* de registro de informações pré-instalado em dispositivos móveis, os quais foram distribuídos para alunos do MIT durante os anos letivos de 2004 e 2005, superando 500.000 horas de dados coletados sobre o comportamento humano.

### 3.1.1 Técnicas de Filtragem

A captura contínua de dados sobre o comportamento humano através de dispositivos móveis tende a gerar um grande volume de dados. Uma ferramenta de análise visual deve permitir aos usuários filtrar e isolar dados de acordo com seu interesse (SHEN; MA, 2008). Neste projeto, são apresentadas duas técnicas de filtragem: semântica e temporal.

Figura 16: Filtragem semântica utilizando ontologia em grafo



Fonte: Adaptado de Shen e Ma (2008)

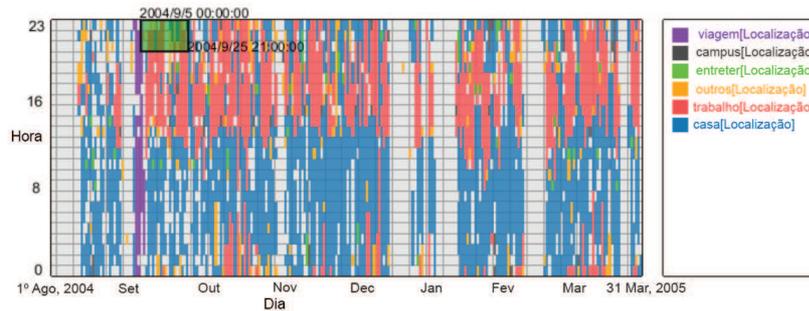
A filtragem semântica visa solucionar problemas encontrados em cenários onde a coleta de dados, oriundos de dispositivos móveis, pode acarretar em um grande volume de informação, tornando difícil sua visualização em uma tela com espaço limitado. Nesta abordagem, utilizando ontologia em grafo, a informação está contida nos vértices e arestas, possibilitando a filtragem dos dados para encontrar subgrupos de interesse, readequando a estrutura à seleção realizada, conforme demonstrado na Figura 16 (SHEN; MA, 2008).

A filtragem temporal, por sua vez, permite a apresentação e exploração da informação sob a perspectiva temporal. Considerando que o comportamento humano costuma apresentar repetições nos padrões apresentados, é extremamente desejável que sistemas de visualização tornem salientes tais padrões e permitam aos usuários selecionar intervalos de tempo que representem seu interesse. Tal abordagem utiliza um gráfico 2D, onde os dois eixos retratam diferentes escalas de tempo e os blocos internos representam os valores, conforme demonstrado na Figura 17. Esta visualização também permite que o usuário selecione uma área do gráfico, delimitando um intervalo de tempo específico (SHEN; MA, 2008).

### 3.1.2 Anéis de Comportamento

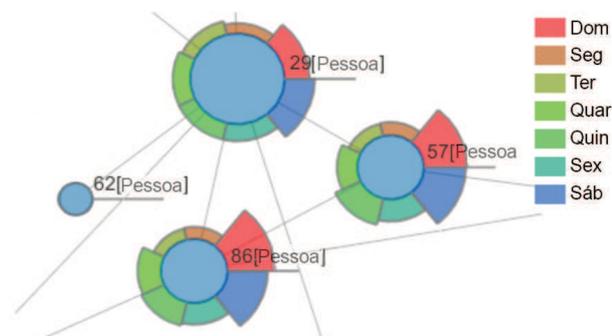
A visualização através de gráfico de tempo se mostrou uma importante ferramenta para análise do comportamento. No entanto, esta forma de visualização não é a mais eficaz em cenários onde deseja-se comparar o comportamento de diferentes sujeitos ou grupos. Com isso,

Figura 17: Gráfico de tempo mostrando atividade de determinado usuário



Fonte: Adaptado de Shen e Ma (2008)

Figura 18: Anel de comportamento



Fonte: Adaptado de Shen e Ma (2008)

uma nova forma de visualização, denominada anéis de comportamento, foi desenvolvida com o intuito de facilitar a comparação entre as informações. Os anéis de comportamento possuem um *layout* radial em forma de pizza, onde cada uma de suas cunhas representa a informação variável com relação ao tempo. O tamanho da cunha indica o valor do item analisado, conforme demonstrado na Figura 18 (SHEN; MA, 2008).

Além da representação individual, em uma rede de anéis de comportamento, o usuário pode, a partir da interatividade presente na visualização, associar diferentes nodos para enxergar o padrão de comportamento presente em um grupo de indivíduos. Assim, é formado um macro anel de comportamento que engloba os sujeitos selecionados pelo usuário, que apresenta a atividade analisada de forma agregada.

### 3.1.3 Resultados

Um dos principais objetivos da análise de rede social é a percepção de amizade entre os usuários da rede. Através da ferramenta de visualização projetada, é possível chegar à determinada conclusão a partir da análise do comportamento dos indivíduos observados.

Outro interesse do projeto é possibilitar a comparação dos padrões de comportamento diário de diferentes grupos de indivíduos. Para tal, são disponibilizados anéis de comportamento

considerando as relações de proximidade existentes entre os sujeitos pertencentes aos grupos.

### 3.2 *Eigenbehaviors: identifying structure in routine*

Dados geográficos possuem informações significativas sobre o cenário em que o indivíduo esteja inserido. De forma geral, a vida das pessoas é repleta de atividades rotineiras e de fácil identificação. Este preceito torna-se mais evidente quando o comportamento está em um contexto temporal, espacial e social. O principal objetivo do trabalho é identificar a estrutura presente em padrões de comportamento de indivíduos ou comunidades dentro de uma população (EAGLE; PENTLAND, 2009).

A partir do conjunto de dados que remetem ao comportamento dos usuários, a estrutura comportamental dos indivíduos é constituída através das suas componentes principais, onde *eigenbehaviors* é o termo utilizado para representar o vetor de características. *Eigenbehaviors* são os autovetores da matriz de covariância dos dados. A utilização de *eigenbehaviors* pode não somente dar indícios sobre o comportamento individual, mas também sobre o comportamento de comunidades dentro de uma rede social (EAGLE; PENTLAND, 2009).

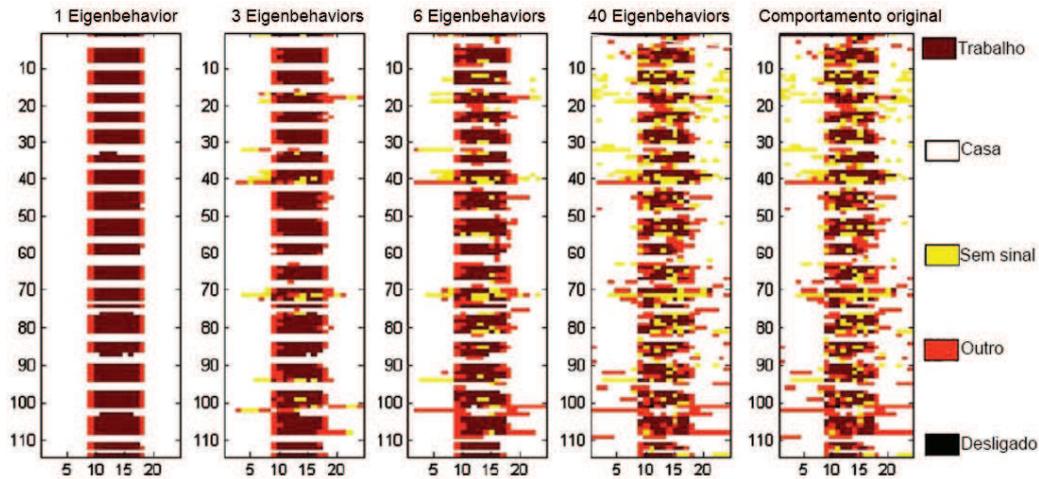
Para o experimento, foi utilizado um grande conjunto de dados obtido através do projeto *Reality Mining*, o qual representa o comportamento de 100 estudantes do MIT durante o período acadêmico de 2004 a 2005. Os dados foram obtidos através de um *software* de registro de informações instalado em dispositivos móveis, contendo dados de localizações do indivíduo, chamadas telefônicas e relações de proximidade *bluetooth*. O objetivo deste trabalho não é indicar que os indivíduos contidos no conjunto de dados do projeto *Reality Mining* são o reflexo comportamental da sociedade. Contudo, para muitas pessoas existe uma certa regularidade no comportamento diário (EAGLE; PENTLAND, 2009).

#### 3.2.1 *Eigenbehaviors para indivíduos*

Cada indivíduo do projeto *Reality Mining* é representado por um conjunto de comportamentos diários, onde a informação sobre o comportamento diário do indivíduo é retratado por um vetor. Através da realização da técnica de análise de componentes principais (PCA) sobre os vetores de comportamento, são gerados novos vetores que melhor descrevem o comportamento do indivíduo. Estes vetores correspondem aos autovetores da matriz de covariância utilizada na técnica PCA (EAGLE; PENTLAND, 2009).

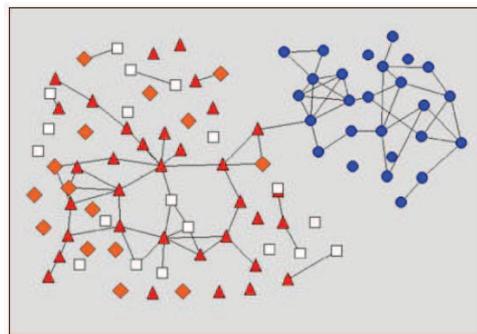
Os *eigenbehaviors* de um indivíduo caracterizam um espaço em que todos os dias da apuração podem ser projetados. A Figura 19 retrata a projeção de cada dia em espaços criados a partir da utilização de diferentes números de *eigenbehaviors* para um indivíduo. Através da visualização, pode-se perceber que com a utilização de 40 *eigenbehaviors*, a representação aproximou-se perfeitamente da reprodução original dos dados (EAGLE; PENTLAND, 2009).

Figura 19: Projeção do comportamento de indivíduo utilizando diferentes número de *eigenbehaviors*



Fonte: Adaptado de Eagle e Pentland (2009)

Figura 20: Rede social da população do projeto *Reality Mining*



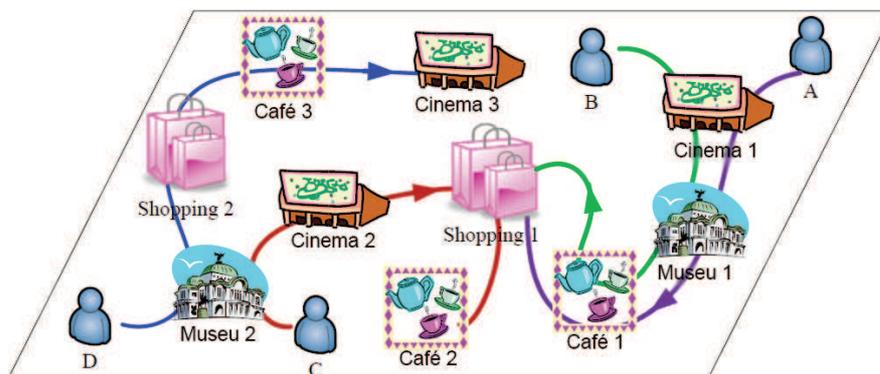
Fonte: Eagle e Pentland (2009)

### 3.2.2 Eigenbehaviors para redes sociais

No âmbito de redes sociais, a proposta do trabalho implica em possibilitar a inferência de relações em uma rede social baseada na comparação dos *eigenbehaviors* estruturados a partir das informações dos indivíduos. É correto considerar que cada um dos diferentes grupos de indivíduos presentes na rede social, formada através do projeto *Reality Mining*, conforme demonstrado na Figura 20, possuem características comportamentais semelhantes entre si. Assim, é possível detectar os *eigenbehaviors* destas comunidades e vislumbrar indivíduos que possuam hábitos compatíveis com os presentes em tais comunidades. A indicação de que o comportamento da comunidade explica o comportamento do indivíduo é medida pela distância euclidiana entre as componentes principais do comportamento do indivíduo e da comunidade (EAGLE; PENTLAND, 2009).

O comportamento existente em uma comunidade é proveniente das características comportamentais dos indivíduos que a formam. Diante disto, torna-se possível mensurar a semelhança

Figura 21: Trajetórias GPS e interesses dos usuários



Fonte: Adaptado de Xiao et al. (2010)

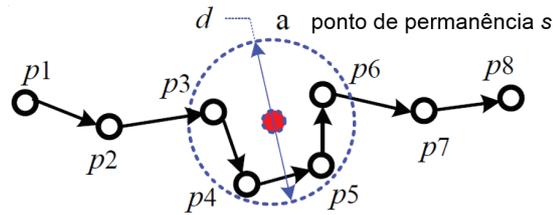
entre os membros da comunidade através do cálculo de similaridade entre os *eigenbehaviors* dos sujeitos comparados com os *eigenbehaviors* da comunidade. Para estabelecer a similaridade entre dois sujeitos em uma comunidade é utilizada a técnica da distância euclidiana (EAGLE; PENTLAND, 2009).

### 3.3 Findind Similar Users Using Category-Based Location History

É notório que, recentemente, houve um grande aumento do número de pessoas que utilizam dispositivos com recursos de geolocalização, os quais são capazes de registrar o histórico de localizações dos indivíduos. Além dos dados geográficos oriundos deste monitoramento, tais informações podem ser utilizadas para retratar os interesses e preferências das pessoas, conforme demonstrado pela Figura 21 (XIAO et al., 2010) que, diante do cenário apresentado, três diferentes situações são evidenciadas:

1. Sobreposições geográficas: pessoas que frequentam os mesmos locais, em um mesmo espaço geográfico, tendem a apresentar interesses em comum. Esta abordagem considera que pessoas compartilham dos mesmos interesses se ambas frequentarem o mesmo cinema, o mesmo museu ou mesma cafeteria, por exemplo.
2. Sobreposições semânticas: pessoas que apresentam padrões de mobilidade semelhantes e frequentam locais de mesma categoria apresentam interesses em comum. Neste contexto, as pessoas compartilham dos mesmos interesses, caso costumem frequentar, mesmo que em espaços geográficos diferentes, cinemas, museus e cafeterias, por exemplo.
3. Sequência de localizações: A ordem com que os indivíduos frequentam determinados locais pode inferir a existência, ou não, de similaridade entre os mesmos. Assim, mesmo que diferentes indivíduos tenham frequentado, ao longo do período, as mesmas categorias de locais, a diferença na sequência das localizações pode refletir uma baixa significância na relação entre estes indivíduos.

Figura 22: Definição de ponto de permanência a partir de trajetória GPS



Fonte: Adaptado de Xiao et al. (2010)

Segundo Xiao et al. (2010), o principal propósito deste projeto é determinar a similaridade entre os indivíduos de acordo com as localizações históricas sob a perspectiva semântica. Para os autores, o processo de definição da similaridade entre os usuários passa por duas etapas fundamentais: (i) modelagem dos interesses dos usuários com base em dados coletados do dispositivo móvel dos indivíduos e; (ii) mensurar a similaridade entre os usuários baseado em seus históricos de localizações.

Base para a concepção deste trabalho, trajetória GPS,  $Tra = p_0 \rightarrow p_1 \rightarrow \dots \rightarrow p_k$ , é uma sequência de pontos geográficos, ordenada cronologicamente, contendo a informação temporal em que o usuário esteve presente na respectiva posição geográfica, onde  $p_i = (x, y, t)$ , representa latitude, longitude e o instante de tempo, respectivamente. Um ponto de permanência  $s$  é uma região geográfica onde o usuário tenha permanecido durante um determinado limite de tempo  $\theta_t$  e de um limite de distância  $\theta_d$ , conforme demonstrado pela Figura 22.

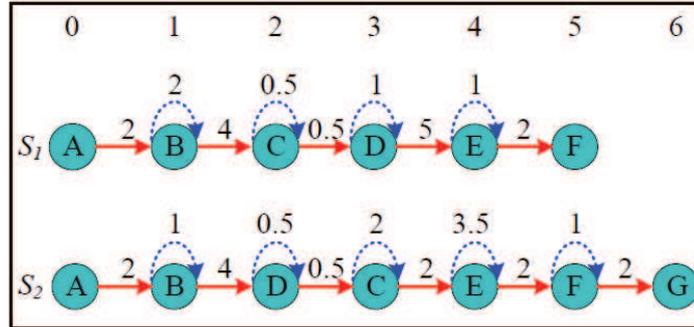
### 3.3.1 Modelagem do histórico de localizações

O processo de modelagem do histórico de localizações dos indivíduos passa por diferentes estágios. Na etapa de representação dos pontos de permanência (*Stay Points Representation*), conforme método demonstrado na Figura 22, são extraídos os pontos de permanência dos usuários com base nos dados de trajetórias do dispositivo GPS. O trabalho propõe a construção de vetores de características de acordo com os pontos de interesse presentes na região de permanência. O vetor de características de uma região de permanência refere-se ao peso de uma categoria de um ponto de interesse dentro de determinada região.

Durante o processo de geração de histórico de localizações dos usuários, as regiões de permanência são agrupadas de acordo com seus vetores de característica, onde as regiões pertencentes ao mesmo *cluster* possuem características semânticas semelhantes, definindo assim, uma localização semântica. Para Xiao et al. (2010), os seres humanos conseguem distinguir, com mais facilidade, usuários semelhantes mediante categorias com uma granularidade mais fina, ou seja, a categoria *restaurante* indica pessoas que fazem suas refeições fora do ambiente doméstico, enquanto a categoria *restaurante japonês* especifica pessoas interessadas em um cardápio característico.

O processo de construção do histórico de localizações é realizado para cada indivíduo, com

Figura 23: Apuração da máxima combinação de viagens



Fonte: Xiao et al. (2010)

base em sua hierarquia de localização semântica e seus pontos de permanência. A proposta consiste em substituir o ponto de permanência pela localização semântica a qual o vetor de característica do respectivo ponto de permanência está contido. A partir de tal abordagem, o histórico de localizações de diferentes usuários tornam-se comparáveis (XIAO et al., 2010). Assim, um histórico de localização semântica é uma sequência de localizações semânticas. Dessa forma, o histórico de localizações semântica do usuário  $m$  pode ser definido por  $H = \{S^2, S^3\}$ , onde

$$S^2 = (c_{20} \xrightarrow{\Delta t_1} c_{20} \xrightarrow{\Delta t_2} c_{21} \xrightarrow{\Delta t_3} c_{21} \xrightarrow{\Delta t_4} c_{21} \xrightarrow{\Delta t_5} c_{21} \xrightarrow{\Delta t_6} c_{20}),$$

$$S^3 = (c_{30} \xrightarrow{\Delta t_1} c_{30} \xrightarrow{\Delta t_2} c_{32} \xrightarrow{\Delta t_3} c_{32} \xrightarrow{\Delta t_4} c_{33} \xrightarrow{\Delta t_5} c_{33} \xrightarrow{\Delta t_6} c_{30})$$

### 3.3.2 Casamento entre histórico de localizações

Para Xiao et al. (2010), o casamento entre os históricos de localizações passa pelo processo de encontrar a máxima combinação de viagens comum entre indivíduos. A combinação de viagens considera a ordem de visitaç o dos locais e tamb m o tempo decorrido de um ponto ao outro. Assim, dadas duas sequ ncias de localiza o  $S_1 = [a_1, a_2, \dots, a_k]$  e  $S_2 = [b_1, b_2, \dots, b_k]$ , uma correspond ncia m xima de viagem ocorre quando a ordem das categorias de locais visitados e o tempo de deslocamento de um local ao outro, mediante toler ncia  $p$ , s o semelhantes, conforme demonstrado na Figura 23.

A similaridade entre os usu rios, calculada a partir dos dados hist ricos de localiza es, considera a similaridade sem ntica em cada uma das diferentes camadas da estrutura. Assim, a similaridade entre duas sequ ncias de localiza es sem nticas   dada pelo somat rio da similaridade apurada em cada uma das camadas (XIAO et al., 2010).

Tabela 1: Comparativo entre recursos utilizados na concepção dos trabalhos relacionados

| Recurso/Aspecto                        | SHEN; MA<br>2008 | EAGLE; PENTLAND<br>2009 | XIAO et al.<br>2010        | GeoSocial                   |
|--|------------------|-------------------------|----------------------------|-----------------------------|
| Formalização de rede social            | Visualização     | Distância euclidiana    | Não                        | <i>Affinity Propagation</i> |
| Cálculo da similaridade entre usuários | Não              | Distância euclidiana    | Sequências de localizações | Distância euclidiana        |
| Visualização de dados                  | Grafo/Gráfico 2D | <i>Eigenbehaviors</i>   | Não                        | Coordenadas Paralelas       |
| Ferramenta de análise visual           | Grafo/Gráfico 2D | Não                     | Não                        | Coordenadas Paralelas       |
| Abrangência espacial                   | Restrita         | Restrita                | Global                     | Global                      |
| População                              | Restrita         | Restrita                | Global                     | Global                      |

Fonte: Elaborada pelo autor.

### 3.4 Comparativo entre trabalhos relacionados

Para finalizar a apresentação dos trabalhos correlatos ao GeoSocial, é apresentado, na Tabela 1, um comparativo dos recursos existentes neste trabalho, comparados com os recursos implementados nos trabalhos relacionados descritos, considerando os seguintes aspectos:

1. Formalização de rede social: determina o método de formalização de rede social existente no trabalho, quando houver.
2. Cálculo da similaridade entre usuários: estabelece a forma de cálculo da similaridade entre indivíduos adotado no trabalho, caso houver.
3. Visualização de dados: indica a forma de visualização de dados proposta pelo trabalho, caso seja implementada.
4. Ferramenta de análise visual: estabelece a forma de interação da ferramenta de análise visual proposta no trabalho, caso houver.
5. Abrangência espacial: indica a abrangência espacial adotada na concepção do trabalho (restrita ou global).
6. População: indica a população de indivíduos considerada na concepção do trabalho relacionado (restrita ou global).

Através da apresentação do projeto MobiVis, proposto por Shen e Ma (2008), é possível perceber características semelhantes entre o respectivo projeto e a proposta deste trabalho. Ambas as abordagens fazem uso de dados coletados através de dispositivos móveis, atuando no contexto social, espacial e temporal. Além disto, as duas propostas utilizam-se da classificação semântica dos locais frequentados pelos indivíduos possibilitando a compreensão do comportamento dos mesmos. Mesmo que através de técnicas distintas, os trabalhos apresentam formas de visualização dos dados que permitem a análise e interpretação das informações capturadas.

Contudo, diferente da ideia apresentada no projeto MobiVis, que analisa seus grupos de usuários determinados pelas diferentes turmas de estudantes do MIT, este trabalho busca a clusterização dos indivíduos em grupos de acordo com seus hábitos, ou seja, informações geográficas que representem os interesses gerais dos usuários, não estando delimitados em uma área espacial determinada.

Baseado nas informações à respeito do projeto defendido por Eagle e Pentland (2009), é possível identificar importantes similaridades entre os projetos. Ambas as propostas utilizam-se de dados coletados a partir do dispositivo móvel dos indivíduos. Embora que realizado de formas diferentes, as duas propostas consistem em modelar os interesses dos usuários e prover subsídios que permitam, através da visualização dos dados, a análise do comportamento dos indivíduos de uma população. Além disto, ambas as abordagens conseguem mensurar a similaridade entre os indivíduos através de seus interesses. Entretanto, diferente da proposta apresentada no projeto *Eigenbehaviors*, que possui seus membros concentrados em estudantes no MIT, este trabalho visa a identificação de grupos de indivíduos da população em geral, baseado em informações geográficas que representem os seus interesses, considerando uma população e abrangência espacial global. Além disto, o trabalho propõe-se ao desenvolvimento de uma ferramenta de análise visual, visando a exploração dos dados na busca pela identificação de oportunidades ou lacunas com base em tais dados.

Conforme características apresentadas sobre o projeto proposto por Xiao et al. (2010), pode-se perceber pontos em comum com o apresentado na proposta deste trabalho. As propostas assemelham-se no fato de utilizarem dados coletados a partir do dispositivo móvel dos usuários para retratar as localizações que representam os interesses dos indivíduos, retratados sob a perspectiva semântica. Além disto, as propostas apresentam uma importante semelhança no quesito de modelagem do perfil do usuário com base nas informações geográficas coletadas, especialmente no processo de definição de um ponto de permanência do indivíduo. Adicionalmente, a funcionalidade de estipular o nível de similaridade entre os usuários está presente em ambas as propostas. Contudo, diferente da proposta relacionada, este trabalho propõem-se à disponibilizar uma forma de visualização e análise dos dados, permitindo a análise das características dos indivíduos pertencentes ao projeto.

Diante do exposto no detalhamento dos trabalhos relacionados citados, percebe-se a utilização das características fundamentais, utilizadas na concepção do trabalho proposto, presentes na elaboração dos trabalhos relacionados. Contudo, pode-se identificar importantes lacunas a serem preenchidas, como por exemplo, a utilização conjunta de um processo automatizado de identificação de comunidades existentes dentro de uma população, formalizada a partir de informações que remetem aos hábitos dos indivíduos, e de um processo de análise visual que, por sua vez, permita a visualização e a identificação de dados, referentes aos hábitos dos indivíduos, que forneçam subsídios para a extração de informação útil a partir da interação com a interface visual projetada. Assim, a proposta deste trabalho objetiva a criação de um modelo de análise e agrupamento de uma população de pessoas, através das características dos locais frequentados

pelos indivíduos, os quais representam seus interesses, analisados sob a perspectiva semântica. O objetivo do trabalho é compreender os hábitos da população, alcançando uma abrangência espacial global. Além disto, as atividades de toda e qualquer pessoa pode estar sendo analisada através do modelo proposto.

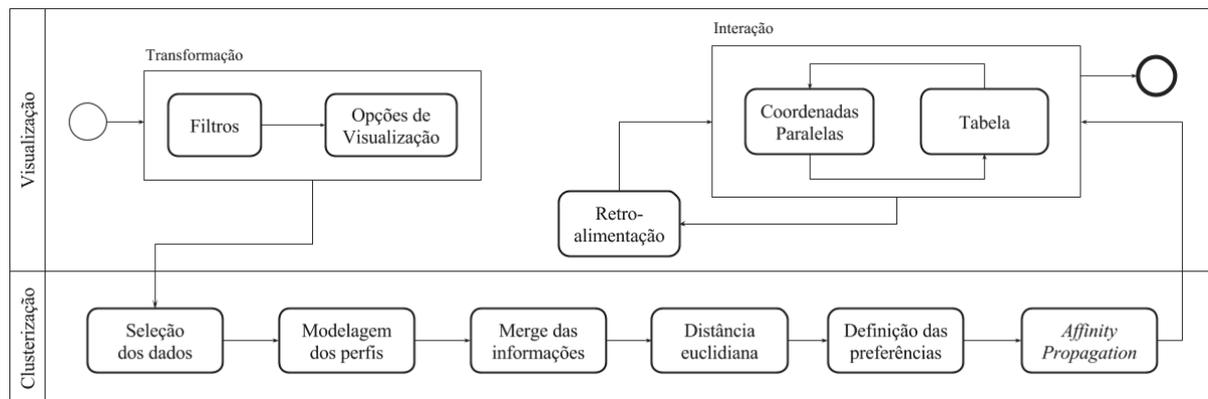
Para tal, o módulo de clusterização, considerando as diferentes etapas contempladas no processo, fazendo uso da técnica de clusterização *Affinity Propagation*, aliado ao módulo de visualização proposto que, por sua vez, estabelece o processo de análise visual, o qual utiliza, em sua concepção, a técnica de coordenadas paralelas, reflete a principal contribuição científica do trabalho proposto. Assim, o agrupamento de indivíduos resultante permite a constituição de uma rede social espontânea, ou seja, a formalização de uma rede social cujos critérios de agrupamento são baseados única e exclusivamente pelos hábitos dos indivíduos que a compõe, onde tais hábitos são interpretados pelo histórico de contexto de cada indivíduo pertencente à rede, analisados sob as informações a respeito dos locais de interesse frequentados.

Em contrapartida, a ferramenta de análise visual proposta neste trabalho possibilita, através da manipulação e interação com os dados obtidos, a identificação de novas oportunidades, bem como a compreensão de carências existentes nas áreas analisadas. Aliada à análise visual, a análise da estrutura formada pela rede social permite a caracterização de nichos específicos para os diferentes segmentos considerados na visualização dos dados, possibilitando a identificação de grupos de indivíduos que melhor correspondem aos critérios do analista, evidenciando o enorme potencial existente no modelo proposto neste trabalho.

## 4 MÉTODO

Este capítulo descreve os métodos e técnicas utilizadas na elaboração deste trabalho, os quais são divididos em módulos, de acordo com o ilustrado pela Figura 24. Conforme demonstrado, a proposta consiste no desenvolvimento de um módulo de clusterização, que visa o agrupamento dos dados mediante suas características, bem como um módulo de visualização que, por sua vez, permite a realização de uma análise visual por parte de analistas.

Figura 24: Escopo, dividido em módulos, do método proposto



Fonte: Elaborada pelo autor.

### 4.1 Clusterização

O processo de clusterização dos indivíduos é dividido em diferentes etapas, conforme demonstrado na Figura 24. Primeiramente, o processo inicia com a seleção das informações dos usuários pertencentes à análise. Neste processo são consultadas as localizações em que o usuário esteve presente durante o período de utilização do aplicativo.

A partir da seleção dos dados dos usuários, a etapa seguinte refere-se à modelagem do perfil de cada usuário presente na análise. Este procedimento classifica a atividade dos usuários conforme a semântica dos locais disponibilizados pela base de dados da Google, apresentado no Apêndice B, juntamente com as categorias pré-determinadas na concepção deste trabalho, sendo elas, residência, trabalho, ensino e deslocamento.

Na constituição do perfil do usuário somente são consideradas as classificações semânticas dos locais em que o indivíduo tenha realizado alguma atividade, ou seja, onde exista registro de localização armazenado para a respectiva representação semântica de local. Diante disto, a quantidade de informação contida na representação de um indivíduo torna-se variável. A composição das informações referente à cada semântica de localização, presente na modelagem do perfil dos usuários, dá-se de acordo com uma relação de atributos definidos na concepção deste trabalho, indicados no Apêndice C, os quais representam a quantidade total de tempo, em

minutos, de permanência do usuário na respectiva categoria de local conforme o contexto de cada atributo.

A definição de valores dos diferentes atributos que compõem o perfil dos usuários ocorre através de diferentes estágios, de acordo com o apresentado na Tabela 2.

O primeiro consiste na decomposição do período de permanência em cada local, considerando cada intervalo de uma hora do dia, conforme horários informados nas colunas “Horário inicial” e “Horário final”. Em seguida, são contabilizados os tempos, em minutos, em que o horário da localização esteve compreendido dentro de cada intervalo, obtido através do processo de decomposição do período de permanência nos locais frequentados. Com isto, tem-se a definição do valor ilustrado pela coluna “Minutos”, indicado na Tabela 2.

A partir da decomposição do período de permanência no local frequentado pelo usuário, o tempo, em minutos, apurado em cada intervalo estabelecido, é atribuído aos atributos que compõem o perfil dos indivíduos, de acordo com o seu contexto, onde:

- Dia/Mês: refere-se ao atributo, sob o contexto de dias do mês, que receberá o valor, em minutos, computado no respectivo intervalo. Assim, conforme destacado na Tabela 2, dois intervalos, totalizando 90 minutos, indicam que seus tempos apurados serão atribuídos ao atributo que representa o 28º dia do mês, enquanto três intervalos, totalizando 165 minutos, estabelecem que seus valores serão atribuídos ao atributo que representa o 1º dia do mês.
- Dia/Semana: indica o atributo, sob o contexto de dias da semana, que irá receber o valor, em minutos, estipulado no intervalo. Conforme indicado no exemplo retratado pela Tabela 2, dois intervalos, totalizando 90 minutos, indicam que seus valores serão atribuídos ao atributo do perfil do usuário que representa o 4º dia da semana (quarta-feira), enquanto os demais intervalos, totalizando 165 minutos, indicam que seus valores serão atribuídos ao atributo que representa o 5º dia da semana (quinta-feira).
- Hora/Dia: estabelece o atributo, sob o contexto de horas do dia, que receberá o valor, em minutos, estipulado no respectivo intervalo. De acordo com o demonstrado na Tabela 2, cada um dos cinco intervalos originados, indica, respectivamente, que serão atribuídos 30 minutos ao atributo que refere-se à 23ª hora do dia (entre 22:00 e 23:00), 60 minutos ao atributo que refere-se à 24ª hora do dia (entre 23:00 e 00:00), 60 minutos ao atributo que refere-se à 1ª hora do dia (entre 00:00 e 01:00), 60 minutos ao atributo que refere-se à 2ª hora do dia (entre 01:00 e 02:00) e 45 minutos ao atributo que refere-se à 3ª hora do dia (entre 02:00 e 03:00).
- Período/Dia: refere-se ao atributo, sob o contexto de períodos do dia, que receberá o valor, em minutos, estipulado em cada um dos intervalos computados. Diante disto, conforme exemplificado na Tabela 2, dois intervalos, totalizando 90 minutos, indicam que seus tempos apurados serão atribuídos ao atributo do perfil do usuário que representa o

Tabela 2: Processo de decomposição da localização em intervalos de horários

| Intervalos      |                |         | Atributos do perfil do usuário |            |             |             |     |
|-----------------|----------------|---------|--------------------------------|------------|-------------|-------------|-----|
| Horário inicial | Horário final  | Minutos | Dia/Mês                        | Dia/Semana | Hora/Dia    | Período/Dia | Mês |
| 28/02/18 22:30  | 28/02/18 23:00 | 30      | 28                             | 4          | 22:00-23:00 | 18:00-00:00 | 2   |
| 28/02/18 23:00  | 01/03/18 00:00 | 60      | 28                             | 4          | 23:00-00:00 | 18:00-00:00 | 2   |
| 01/03/18 00:00  | 01/03/18 01:00 | 60      | 1                              | 5          | 00:00-01:00 | 00:00-06:00 | 3   |
| 01/03/18 01:00  | 01/03/18 02:00 | 60      | 1                              | 5          | 01:00-02:00 | 00:00-06:00 | 3   |
| 01/03/18 02:00  | 01/03/18 02:45 | 45      | 1                              | 5          | 02:00-03:00 | 00:00-06:00 | 3   |

A tabela exemplifica o processo de decomposição de uma localização em intervalos, considerando o período de 28/02/2018 22:30 à 01/03/2018 02:45. Cada valor, originado a partir da decomposição, é atribuído aos atributos do perfil do usuário, de acordo com o seu contexto. Este processo é cumulativo, logo, para novas localizações do mesmo categoria, os valores são adicionados aos atributos do perfil para o respectivo tipo de local.

Fonte: Elaborada pelo autor.

período do dia entre os horários de 18:00 e 00:00 (noite), enquanto os demais intervalos apurados, totalizando 165 minutos, indicam que seus valores serão atribuídos ao atributo que representa o período do dia entre os horários de 00:00 e 06:00 (madrugada).

- **Mês:** indica o atributo, sob o contexto de meses do ano, que irá receber o valor, em minutos, computado no respectivo intervalo. Assim, de acordo com o demonstrado na Tabela 2, dois intervalos, totalizando 90 minutos, indicam que seus tempos estimados serão atribuídos ao atributo do perfil do usuário que remete ao 2º mês do ano (fevereiro), por outro lado, os três atributos restantes, totalizando 165 minutos, indicam que seus valores serão atribuídos ao atributo do perfil do usuário que representa o 3º mês do ano (março).

Importante ressaltar que o perfil dos usuários possui um atributo que indica o tempo total apurado, através da mensuração do tempo de permanência nos locais frequentados pelo indivíduo. Assim, de forma obrigatória, este atributo é incrementado a partir de todos os intervalos apurados, considerando todas as localizações existentes.

Ao final do procedimento de modelagem do perfil do usuário, o mesmo é retratado por uma matriz $_{n \times m}$ , onde  $n$  indica o número de diferentes semânticas de locais em que o usuário tenha apresentado alguma atividade e  $m$ , por sua vez, indica o número dos diferentes atributos (81) que compõem o perfil do usuário.

A terceira etapa do processo de clusterização consiste na formulação da matriz de similaridades  $S$  entre os usuários. Tal informação é utilizada como dado de entrada do algoritmo de clusterização *Affinity Propagation*, método utilizado na concepção deste trabalho. A matriz resultante deste procedimento é uma matriz quadrada de ordem  $n$ , onde  $n$  refere-se ao número de usuários existentes na análise realizada. A composição dos valores da matriz de similaridades

Figura 25: Processo de equiparação dos perfis dos usuários

Perfis dos usuários originais

| Usuário 1     |           |    |    |
|---------------|-----------|----|----|
| Tipo de local | Atributos |    |    |
| 1             | 100       | 20 | 50 |
| 2             | 80        | 15 | 35 |
| 3             | 60        | 40 | 10 |

| Usuário 2     |           |    |    |
|---------------|-----------|----|----|
| Tipo de local | Atributos |    |    |
| 1             | 75        | 35 | 5  |
| 3             | 65        | 30 | 25 |
| 5             | 90        | 55 | 45 |

Perfis após o processo de equiparação

| Usuário 1     |           |    |    |
|---------------|-----------|----|----|
| Tipo de local | Atributos |    |    |
| 1             | 100       | 20 | 50 |
| 2             | 80        | 15 | 35 |
| 3             | 60        | 40 | 10 |
| 5             | 0         | 0  | 0  |

| Usuário 2     |           |    |    |
|---------------|-----------|----|----|
| Tipo de local | Atributos |    |    |
| 1             | 75        | 35 | 5  |
| 2             | 0         | 0  | 0  |
| 3             | 65        | 30 | 25 |
| 5             | 90        | 55 | 45 |

Fonte: Elaborada pelo autor.

$S$  é dada por

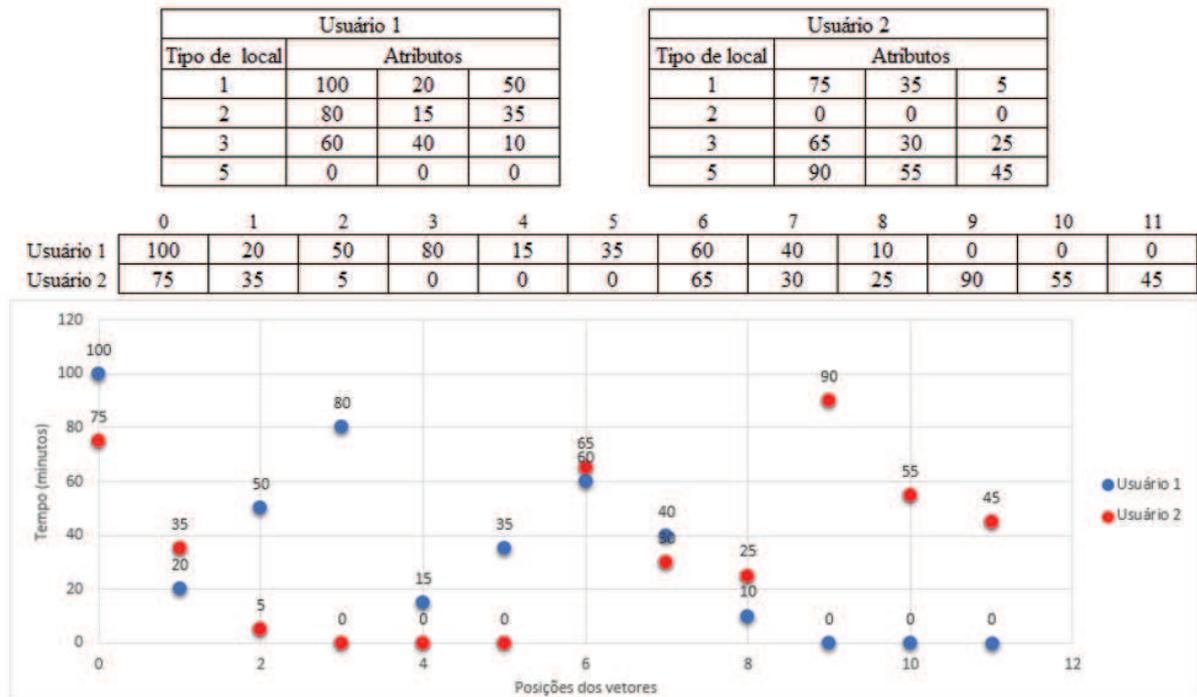
$$S_{n \times m} = \begin{cases} \text{valor da similaridade entre os usuários, } n \neq m, \\ \text{valor da preferência, } n = m \end{cases}$$

O cálculo do valor da similaridade é realizado para todos os diferentes pares de usuários existentes no processo de clusterização. O cálculo consiste em definir o quão semelhantes são os dois indivíduos de acordo com o perfil de cada usuário, constituído no estágio anterior. Contudo, em virtude de tais perfis, possivelmente, apresentarem estruturas diferentes com relação às semânticas de locais frequentados, o processo equipara suas estruturas, complementando as informações das semânticas de locais que estejam presentes em um perfil e que não estejam em outro, e vice-versa, conforme demonstrado na Figura 25, que exemplifica, de forma resumida, o processo de equiparação das informações contidas nos perfis dos usuários envolvidos no cálculo de similaridade.

Após a preparação dos dados ser realizada, através da equiparação dos perfis, os perfis de ambos os usuários são representados por matrizes de mesmas dimensões. Tais matrizes  $u_{n \times m}$  são convertidas em vetores  $v$  de  $n \times m$  posições, conforme exemplificado pela Figura 26, que demonstra, juntamente com o processo de conversão de matriz para vetor, um gráfico exemplificando a dispersão dos valores mediante os atributos que compõem o perfil de ambos os usuários através da representação vetorial.

Conforme apresentado no gráfico da Figura 26, o eixo das abscissas representa as posições dos vetores. Diante do fato da distância ser medida entre os dois pontos localizados sobre a mesma coordenada  $x$ , a distância euclidiana obtida com o cálculo é unidimensional, ou seja, não há distinção entre as coordenadas  $x$  dos pontos envolvidos no cálculo, logo, a distância

Figura 26: Processo de transformação das matrizes em vetores



Fonte: Elaborada pelo autor.

euclidiana unidimensional entre dois pontos é dada por  $d[i] = |v_1[i] - v_2[i]|$ . Pelo fato de as diferenças serem tratadas como uma penalidade, as similaridades entre os usuários devem ter seus valores atribuídos de forma negativa. Diante disto, o cálculo do somatório dos valores referentes à distância euclidiana unidimensional, considerando todas as posições pertencentes aos vetores, resulta no valor da similaridade entre os usuários, demonstrado pela Tabela 3, é dado por

$$sim = - \sum_{i=0}^{(nxm)-1} |v_1[i] - v_2[i]|$$

Em virtude de a matriz de similaridades  $S$  apresentar-se de forma simétrica, o valor da similaridade apurado entre os usuários  $n$  e  $m$  é o mesmo valor apurado entre os usuários  $m$  e  $n$ , logo, a mesma apuração define o valor de duas células ( $S[nxm]$  e  $S[mxn]$ ) da matriz de similaridades  $S$ , evitando reprocessamento. Dessa forma, o valor resultante é atribuído na matriz de similaridades  $S$  conforme a identificação de ambos os usuários, de acordo com o demonstrado na Tabela 4, que apresenta a atribuição da similaridade entre dois usuários.

A partir do momento em que a similaridade entre todos os usuários esteja calculada, o processo entra na próxima etapa, definindo o valor de preferência a ser utilizado no processo de clusterização. O processo considera como preferência o valor da mediana, obtido com base nos valores presentes na matriz de similaridades  $S$ , sendo este valor atribuído à diagonal principal da matriz, conforme demonstrado na Tabela 5.

Após a completa formalização da matriz de similaridades, é iniciado o estágio final do

Tabela 3: Processo de cálculo da similaridade entre dois usuários

| Posição | Usuário (vetor) 1 | Usuário (vetor) 2 | Distância euclidiana unidimensional<br>$ v_1[i] - v_2[i] $ |
|---------|-------------------|-------------------|--|
| 0       | 100               | 75                | 25   |
| 1       | 20                | 35                | 15   |
| 2       | 50                | 5                 | 45   |
| 3       | 80                | 0                 | 80   |
| 4       | 15                | 0                 | 15   |
| 5       | 35                | 0                 | 35   |
| 6       | 60                | 65                | 5  |
| 7       | 40                | 30                | 10   |
| 8       | 10                | 25                | 15   |
| 9       | 0                 | 90                | 90   |
| 10      | 0                 | 55                | 55   |
| 11      | 0                 | 45                | 45   |
| Soma    |                   |                   | -435   |

Fonte: Elaborada pelo autor.

Tabela 4: Atribuição de valores na matriz de similaridades  $S$ 

|           | Usuário 1 | Usuário 2 | Usuário 3 |
|-----------|-----------|-----------|-----------|
| Usuário 1 |           | -435      |           |
| Usuário 2 | -435      |           |           |
| Usuário 3 |           |           |           |

Fonte: Elaborada pelo autor.

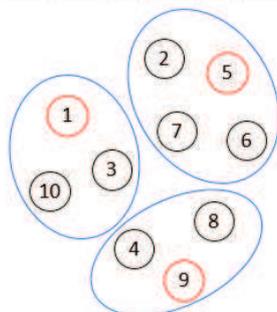
Tabela 5: Atribuição das preferências da matriz de similaridades  $S$ 

|           | Usuário 1 | Usuário 2 | Usuário 3 |
|-----------|-----------|-----------|-----------|
| Usuário 1 | -310      | -435      | -123      |
| Usuário 2 | -435      | -310      | -310      |
| Usuário 3 | -123      | -310      | -310      |

Fonte: Elaborada pelo autor.

Figura 27: Vetor resultante do processo de clusterização *Affinity Propagation*

|            |   |   |   |   |   |   |   |   |   |    |
|------------|---|---|---|---|---|---|---|---|---|----|
| Usuários   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Índices    | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9  |
| Exemplares | 4 | 0 | 4 | 8 | 0 | 0 | 0 | 8 | 8 | 4  |



Fonte: Elaborada pelo autor.

processo de clusterização. Nesta etapa, a matriz de similaridades  $S$  é utilizada como dado de entrada do algoritmo *Affinity Propagation*, visando a realização da efetiva definição dos *clusters* existentes dentre os usuários considerados na análise. Como resultado deste procedimento, é obtido um vetor de  $n$  posições, onde  $n$  indica o número usuários considerados na apuração dos *clusters*.

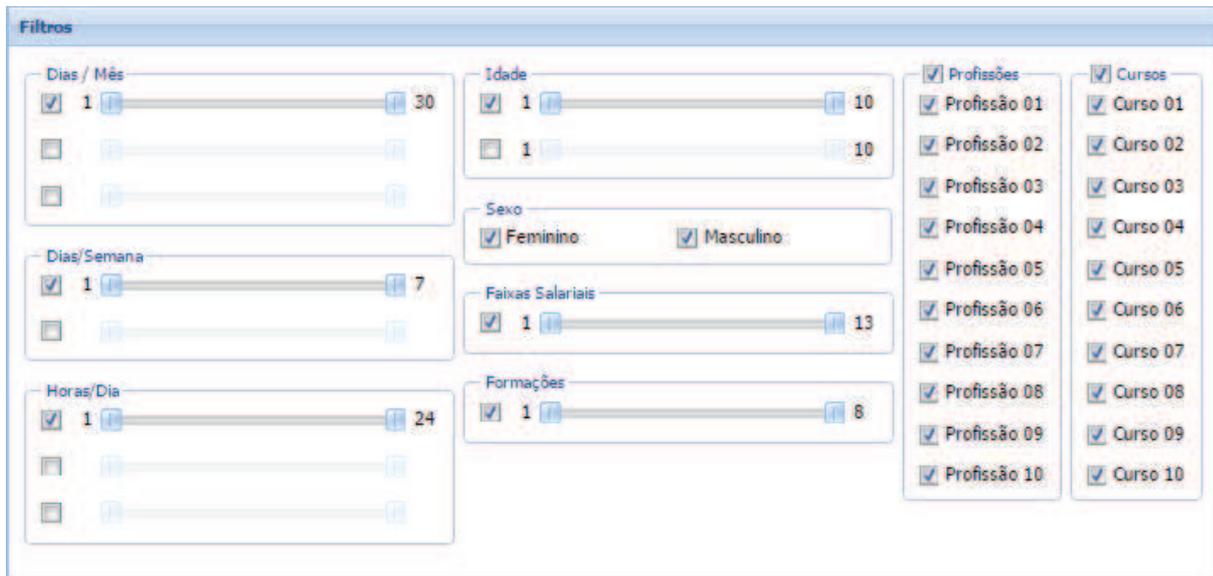
O vetor oriundo do processamento do algoritmo indica quais usuários referem-se aos exemplares dos demais usuários. Com isso, a ordem com que estes usuários são dispostos na matriz de similaridades  $S$  indica qual o seu índice dentro do vetor resultante. Dessa forma, o valor contido em cada posição do vetor representa a posição do usuário definido como exemplar para o usuário da respectiva posição, conforme ilustrado na Figura 27. Assim, o usuário 5, representado pelo índice 4 do vetor, é considerado exemplar do usuário 1, representado pelo índice 0, por exemplo.

## 4.2 Visualização

O processo de visualização é dividido em diferentes estágios, conforme ilustrado pela Figura 24. Inicialmente, o processo apresenta ao usuário ferramentas de transformação dos dados a serem analisados, compreendendo a aplicação de filtros e definição das informações a serem visualizadas.

A aplicação de filtros objetiva a restrição dos registros de pessoas a serem considerados na apuração dos *clusters* através da técnica *Affinity Propagation* e, conseqüentemente, a visualização no sistema de Coordenadas Paralelas. Os filtros são aplicados sob diferentes tipos de dados relacionados aos usuários, no que diz respeito aos seus dados pessoais, e locais, através de dados que remetem ao seu contexto, restringindo assim o retorno dos dados de acordo com a preferência/necessidade do usuário, conforme exemplificado na Figura 28. O segundo procedimento que visa a transformação dos dados é a definição das opções de visualização das

Figura 28: Aplicação de filtros na ferramenta de análise visual



Fonte: Elaborada pelo autor.

informações, onde o usuário pode estabelecer que informações são pertinentes à sua análise.

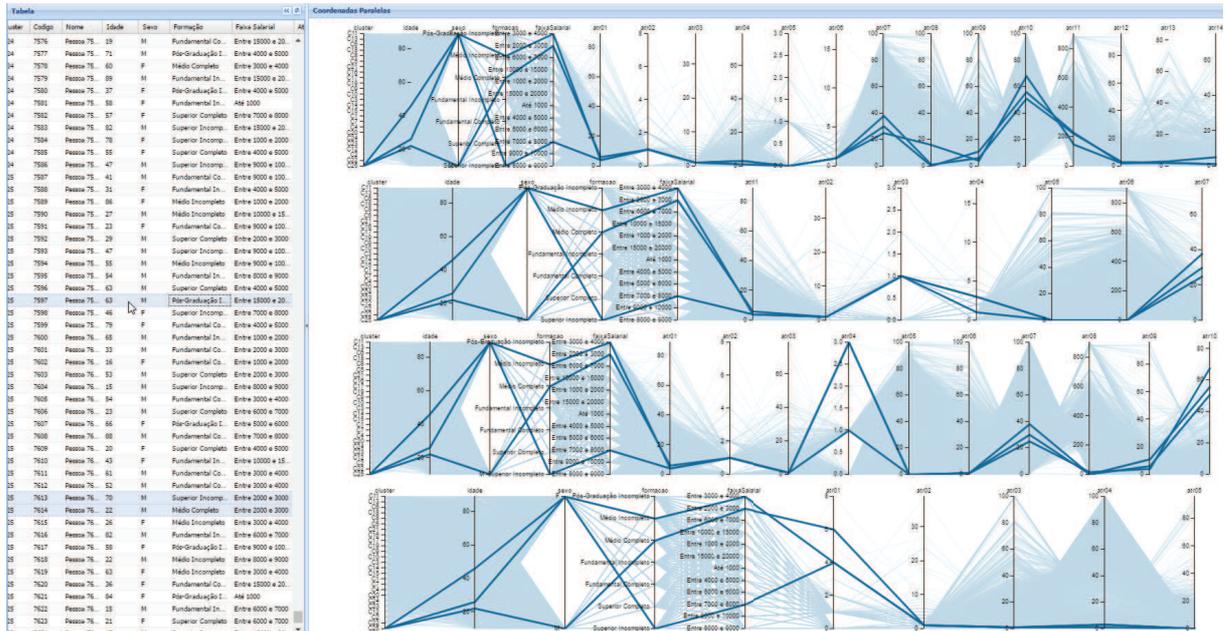
Com a definição dos filtros e das opções de visualização, o módulo de clusterização proposto é realizado. Neste caso, diferente da operação realizada para formalizar a rede no aplicativo, que utiliza todas as informações disponíveis para identificar os *clusters* de usuários existentes, no processo de análise visual a apuração dos *clusters* é realizada considerando os critérios definidos no estágio de transformação dos dados. A adoção de filtros implica diretamente no processo de modelagem do perfil dos usuários. Dessa forma, usuários pertencentes à determinados *clusters* constituídos no aplicativo podem passar a integrar diferentes *clusters*. Após a geração dos dados, os registros das pessoas são populados na tabela e as representações, através de Coordenadas Paralelas, são construídas. O número de representações de Coordenadas Paralelas formadas na análise é dinâmico, definido com base nos dados informados no processo de transformação dos dados.

Conforme evidenciado na Figura 24, o processo de análise visual dá-se através da análise da estrutura formalizada nas diferentes Coordenadas Paralelas em conjunto com a tabela populada com os dados dos usuários. Diversas formas de interação são disponibilizadas na ferramenta com o intuito de extrair informações úteis a partir dos dados analisados.

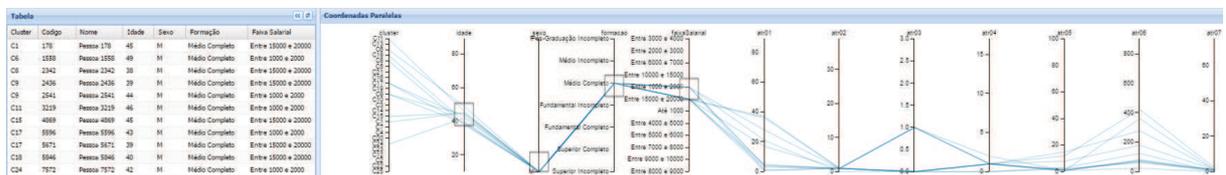
Dentre as interações a serem realizadas sobre a tabela, o usuário pode realizar a seleção de registros, de forma única ou múltipla, destacando as respectivas linhas, referentes aos registros selecionados, de todas as Coordenadas Paralelas existentes, conforme demonstrado na Figura 29. Por outro lado, não havendo seleção de registros na tabela, a ferramenta apresenta detalhes do registro, de forma individual, em que o cursor do mouse estiver posicionado sobre.

Sobre o sistema de Coordenadas Paralelas, a interação pode ocorrer de diferentes maneiras. O usuário pode aplicar a técnica de *brushing* sobre os eixos do sistema, aplicando filtro sobre

Figura 29: Seleção de registro(s) da tabela



Fonte: Elaborada pelo autor.

Figura 30: Técnica de *brushing* aplicado ao sistema de Coordenadas Paralelas

Fonte: Elaborada pelo autor.

os dados na respectiva dimensão. A partir de tal operação, os dados obtidos com a aplicação do critério estabelecido com a utilização da técnica *brushing* são atualizados na tabela, tornando equivalentes os dados contidos na tabela e os dados evidenciados no sistema de Coordenadas Paralelas, conforme demonstrado pela Figura 30. As demais formas de interação pertinentes ao sistema de Coordenadas Paralelas são a translação dos eixos do sistema e a aplicação da técnica de *flipping*, que inverte a ordem dos valores dos eixos do sistema de Coordenadas Paralelas.

Por fim, a ferramenta de análise desenvolvida permite ao usuário realizar a retroalimentação dos dados, constituídos a partir da interação com o sistema de Coordenadas Paralelas. Através deste procedimento, as diferentes Coordenadas Paralelas são reconstruídas com base nos dados presentes na tabela da ferramenta. Esta reconstrução torna possível o reinício da análise dos dados para extração de conhecimento.

### 4.3 Captura das localizações dos usuários

A concepção do modelo proposto prevê a utilização de dados que remetem ao histórico do contexto dos usuários, no que refere-se aos locais frequentados pelos mesmos, indicando os locais de interesse dos indivíduos. Assim, este trabalho propõe o desenvolvimento de um aplicativo móvel, descrito no Apêndice A, com a finalidade de identificar e armazenar as informações referentes às localizações que tenham sido frequentadas pelos usuários. O aplicativo móvel foi disponibilizado para usuários de dispositivos Android na plataforma Play Store da Google.

Em conjunto com o aplicativo móvel, foi desenvolvida uma aplicação de camada servidor, prevendo a centralização do processamento das informações, bem como a conexão com a base de dados, fazendo a leitura e escrita dos dados dos usuários. O ambiente da camada servidor, contemplando servidor de aplicação e base de dados, foi hospedado através do serviço de computação na nuvem da Amazon (Amazon Web Services). Nesta aplicação, foi implementado o módulo referente ao processo de clusterização dos usuários através da técnica *Affinity Propagation*, realizado com base nas informações dos usuários contidas na base de dados da aplicação.

O processo de experimentação utilizando aplicativo móvel desenvolvido a partir da proposta deste trabalho consiste na sua utilização por parte de usuários reais. O experimento contou com a colaboração de 17 usuários, onde foi possível obter informações à respeito dos seus históricos de contexto, relacionados aos locais de interesse frequentados. O período de utilização do aplicativo móvel desenvolvido esteve compreendido entre os dias 01/08/2017 e 30/11/2017, contemplando quatro meses de utilização.

Contudo, grande parte dos usuários apresentou uma ínfima taxa de utilização do aplicativo, conforme demonstrado na Tabela 6, tornando imprecisa a análise de suas atividades. A taxa de utilização foi estipulada com base no tempo total, em minutos, compreendido entre o período de utilização do aplicativo móvel desenvolvido, comparado ao tempo total, em minutos, das informações referentes aos históricos de contextos de cada indivíduo, retratados pelo tempo de permanência em seus locais de interesse. Assim, optou-se pela não utilização de tais dados na avaliação do modelo proposto neste trabalho. Como alternativa, fez-se uso de geração de dados artificiais para atender à tal propósito.

Tabela 6: Taxa de utilização, por parte de usuários reais, do aplicativo móvel desenvolvido

| Usuário | Taxa de utilização (%) |
|---------|------------------------|
| 1       | 80,956                 |
| 2       | 59,212                 |
| 3       | 0,027                  |
| 4       | 12,499                 |
| 5       | 0,168                  |
| 6       | 0,040                  |
| 7       | 0,047                  |
| 8       | 0,104                  |
| 9       | 0,195                  |
| 10      | 0,089                  |
| 11      | 1,340                  |
| 12      | 1,070                  |
| 13      | 34,577                 |
| 14      | 0,364                  |
| 15      | 0,004                  |
| 16      | 10,334                 |
| 17      | 0,220                  |

Fonte: Elaborada pelo autor.



## 5 ANÁLISE DOS RESULTADOS

Este capítulo é destinado à análise dos resultados, oriundos da adoção do modelo de análise e agrupamento de uma população de indivíduos proposto neste trabalho. Para tal, o processo inicia-se com a geração de dados artificiais que representem indivíduos e suas atividades, buscando retratar a realidade da melhor forma. A partir da geração dos dados, estes são submetidos aos processos de clusterização e análise visual propostos na concepção deste trabalho.

### 5.1 Geração de dados artificiais

Com o intuito de realizar a avaliação do GeoSocial considerando um amplo conjunto de indivíduos, realizou-se o procedimento de análise do modelo proposto através da geração de dados artificiais, os quais representam as atividades de um conjunto de indivíduos, objetivando comprovar o funcionamento do modelo.

Para a geração dos dados artificiais considerou-se diferentes populações, contendo 2000 indivíduos cada, o que pode ser considerado como uma amostragem de populações de duas diferentes cidades, por exemplo. Tal cenário permite a avaliação dos hábitos dos indivíduos destas duas populações. As informações geradas contemplam um ano de atividade. Ressalta-se que o experimento não considera período de férias dos indivíduos, ou seja, todos os dias do ano são elegíveis para frequentarem o local de trabalho ou a instituição de ensino, conforme regras a serem definidas.

A geração das informações que compõem o perfil dos indivíduos contempla uma relação de atributos, os quais seus valores são determinados mediante distribuições de probabilidades, definidas e parametrizadas de acordo com cada atributo que compõe o perfil dos indivíduos, conforme:

- Idade: o atributo referente à idade dos indivíduos é definido através da distribuição de probabilidade uniforme, tendo seus valores gerados entre os limites 15 e 70.
- Sexo: o atributo que define o sexo dos indivíduos, o qual é gerado conforme distribuição de probabilidade uniforme, limitado aos valores 0, que indica o sexo feminino, e 1 que, por sua vez, remete ao sexo masculino.
- Faixa de renda: a faixa de renda dos indivíduos é estipulada mediante distribuição de probabilidade normal, considerando média igual à R\$ 4.000,00 e desvio padrão igual à R\$ 2.500,00. O valor gerado resultante é classificado por meio das opções apresentadas na Tabela 7, que relaciona as faixas de renda consideradas no experimento.
- Formação acadêmica: a formação acadêmica dos indivíduos é definida através de distribuição de probabilidade uniforme, considerando 8 valores possíveis. Assim, o valor

Tabela 7: Faixas de renda consideradas na geração de dados artificiais

| Valor | Faixa de Renda (R\$)        |
|-------|-----------------------------|
| 1     | Até 1.000,00                |
| 2     | Entre 1.000,00 e 2.000,00   |
| 3     | Entre 2.000,00 e 3.000,00   |
| 4     | Entre 3.000,00 e 4.000,00   |
| 5     | Entre 4.000,00 e 5.000,00   |
| 6     | Entre 5.000,00 e 6.000,00   |
| 7     | Entre 6.000,00 e 7.000,00   |
| 8     | Entre 7.000,00 e 8.000,00   |
| 9     | Entre 8.000,00 e 9.000,00   |
| 10    | Entre 9.000,00 e 10.000,00  |
| 11    | Entre 10.000,00 e 15.000,00 |
| 12    | Entre 15.000,00 e 20.000,00 |
| 13    | Acima de 20.000,00          |

Fonte: Elaborada pelo autor.

Tabela 8: Formações acadêmicas consideradas na geração de dados artificiais

| Valor | Formação Acadêmica       |
|-------|--------------------------|
| 1     | Fundamental Completo     |
| 2     | Fundamental Incompleto   |
| 3     | Médio Completo           |
| 4     | Médio Incompleto         |
| 5     | Superior Completo        |
| 6     | Superior Incompleto      |
| 7     | Pós-graduação Completo   |
| 8     | Pós-graduação Incompleto |

Fonte: Elaborada pelo autor.

gerado, limitado entre os valores 1 e 8, é classificado conforme opções apresentadas na Tabela 8, que relaciona as formações acadêmicas consideradas no experimento.

- **Discente:** o próximo passo é definir se o indivíduo refere-se à um discente. Tal característica é definida através de uma distribuição de probabilidade uniforme, onde 40% dos valores possíveis indicam que o indivíduo é discente de alguma instituição de ensino, enquanto 60% dos valores possíveis refletem que o indivíduo não exerce tal atividade. Para os indivíduos discentes, são geradas informações adicionais que indicam a periodicidade com que os indivíduos frequentam e o tempo que permanecem na instituição de ensino, retratando assim, os hábitos dos indivíduos referentes à sua permanência em locais de ensino.
- **Número de dias de estudo:** a próxima etapa consiste em definir o número de dias da semana que cada indivíduo frequenta a instituição de ensino. Tal atributo é classificado através da distribuição de probabilidade normal, considerando média igual à 3 e desvio

Tabela 9: Dias da semana de estudo considerados na geração de dados artificiais

| Dia da semana | Probabilidade (%) |
|---------------|-------------------|
| Segunda-feira | 20                |
| Terça-feira   | 20                |
| Quarta-feira  | 20                |
| Quinta-feira  | 20                |
| Sexta-feira   | 15                |
| Sábado        | 5                 |

Fonte: Elaborada pelo autor.

Tabela 10: Turnos de estudo considerados na geração de dados artificiais

| Turno (horário de início) | Probabilidade (%) |
|---------------------------|-------------------|
| Manhã (08:00)             | 15                |
| Tarde (13:00)             | 15                |
| Noite (19:30)             | 70                |

Fonte: Elaborada pelo autor.

padrão igual à 2. Ressalta-se que o valor gerado deve estar compreendido entre os valores 1 e 6, em virtude de o experimento considerar pelo menos 1 dia e no máximo 6 dias de estudo.

- Dias de estudo: com o número de dias de estudo definido, o processo passa a definir em quais dias da semana o indivíduo habitualmente frequenta a instituição de ensino. Para tal, os dias da semana são estipulados mediante distribuição de probabilidade uniforme, conforme distribuição apresentada na Tabela 9, onde cada dia da semana possui uma probabilidade, dentre os valores possíveis, de ser gerado. Ressalta-se que este é um processo iterativo, ou seja, este procedimento será repetido de acordo com o número de dias da semana em que o indivíduo frequenta a instituição de ensino. No entanto, a relação de dias da semana gerada somente considera valores únicos, não sendo permitida a duplicidade de dias da semana em que o indivíduo frequenta a instituição de ensino.
- Turno de estudo: a seguinte característica a ser definida é o turno de estudo dos indivíduos, ou seja, o horário pré estabelecido no qual os indivíduos passam a frequentar a instituição de ensino. Para a definição de tal atributo, utilizou-se a distribuição de probabilidade uniforme, conforme distribuição apresentada na Tabela 10, onde cada turno previsto no experimento possui uma probabilidade, dentre os valores possíveis, de ser gerado.
- Horário de chegada à instituição de ensino: com o intuito de melhor adequar os dados à realidade, durante o processo de geração dos dados referentes à cada localização em que os indivíduos permanecem na instituição de ensino, foi considerada a adição de um valor, em minutos, que expressa a diferença em relação ao horário de início do turno de estudo. Tal valor é estipulado através da distribuição de probabilidade normal, considerando o

Tabela 11: Cursos considerados na geração de dados artificiais

| Valor | Curso                 |
|-------|-----------------------|
| 1     | Administração         |
| 2     | Biomedicina           |
| 3     | Ciência da Computação |
| 4     | Comércio Exterior     |
| 5     | Direito               |
| 6     | Enfermagem            |
| 7     | Engenharia Civil      |
| 8     | Jornalismo            |
| 9     | Psicologia            |
| 10    | Recursos Humanos      |

Fonte: Elaborada pelo autor.

valor de média igual à 0 e desvio padrão igual à 20. Logo, caso o valor gerado seja negativo, indica que o indivíduo ingressou na instituição de ensino antes do horário de início do turno e, em caso de positivo, indica que ingressou após o horário de início do turno.

- Tempo de permanência na instituição de ensino: para cada localização referente à permanência dos indivíduos na instituição de ensino, o tempo, em minutos, de permanência em tal localização é estipulado com base em distribuição de probabilidade normal, utilizando o valor de média igual à 150 e de desvio padrão igual à 30. O valor gerado é adicionado ao horário de ingresso na localização, estipulado anteriormente.
- Curso: o curso realizado pelos indivíduos discentes do experimento é definido com base em distribuição de probabilidade uniforme, estando os possíveis valores limitados entre os valores 1 e 10. Cada valor gerado é classificado conforme cursos apresentados na Tabela 11, que relaciona o conjunto de 10 diferentes cursos considerados no experimento.
- Atividade profissional: o próximo passo é definir se o indivíduo exerce atividade profissional. Esta característica é obtida através da adoção da distribuição de probabilidade uniforme, onde 75% dos valores possíveis indicam que o indivíduo exerce atividade profissional, enquanto 25% dos valores possíveis apontam que o indivíduo não exerce tal atividade. Para os indivíduos que exercem atividade profissional, são geradas informações adicionais que indicam a periodicidade com que os indivíduos frequentam e o tempo que permanecem no trabalho, retratando assim, os hábitos dos indivíduos referentes à sua permanência em tais locais.
- Número de dias de trabalho: a próxima etapa consiste em definir o número de dias da semana que cada indivíduo frequenta o local de trabalho. Tal atributo é classificado através da distribuição de probabilidade uniforme, conforme distribuição apresentada na Tabela

Tabela 12: Dias de trabalho considerados na geração de dados artificiais

| Dias de trabalho        | Probabilidade (%) |
|-------------------------|-------------------|
| Seg, Ter, Qua, Qui, Sex | 80                |
| Seg, Qua, Sex, Dom      | 10                |
| Ter, Qui, Sab           | 10                |

Fonte: Elaborada pelo autor.

Tabela 13: Horário de início das jornadas de trabalho consideradas na geração de dados artificiais

| Horário início jornada | Probabilidade (%) |
|------------------------|-------------------|
| 07:00                  | 40                |
| 08:00                  | 40                |
| 14:00                  | 10                |
| 20:00                  | 10                |

Fonte: Elaborada pelo autor.

12, onde cada conjunto de dias da semana possui uma probabilidade, dentre os valores possíveis, de ser gerado.

- Horário de início da jornada de trabalho: o horário de início da jornada dos indivíduos que exercem atividade profissional é definido com base em distribuição de probabilidade uniforme, de acordo com o demonstrado na Tabela 13, onde cada um dos horários previstos no experimento possui uma probabilidade, dentre os valores possíveis, de ser gerado.
- Horário de chegada ao local de trabalho: buscando tornar os dados artificiais gerados condizentes com a realidade, durante o processo de geração dos dados artificiais que referem-se às localizações em que os indivíduos permanecem nos locais de trabalho, considerou-se a inclusão de um valor, definido em minutos, que representa a diferença em relação ao horário de início da jornada de trabalho do indivíduo. Este valor é calculado mediante a adoção de distribuição de probabilidade normal, considerando o valor de média igual à 0 e desvio padrão igual à 10. Desta forma, sendo definido um valor negativo, este valor indica que o indivíduo ingressou no local de trabalho antes do horário de início da jornada de trabalho e, em caso de positivo, o valor indica que ingressou após o horário de início da jornada de trabalho.
- Tempo de permanência no local de trabalho: para cada localização que se refere à permanência dos indivíduos em seus locais de trabalho, o tempo, em minutos, de permanência é definido através da utilização de distribuição de probabilidade normal, considerando o valor de média igual à 480 e desvio padrão igual à 60. O valor obtido é adicionado ao horário de ingresso nesta localização.
- Profissão: a profissão dos indivíduos que exercem atividade profissional é estabelecida através da adoção de distribuição de probabilidade uniforme, tendo os possíveis valores

Tabela 14: Profissões consideradas na geração de dados artificiais

| Valor | Profissão            |
|-------|----------------------|
| 1     | Advogado             |
| 2     | Analista de Sistemas |
| 3     | Babá                 |
| 4     | Dentista             |
| 5     | Enfermeira           |
| 6     | Engenheiro           |
| 7     | Mecânico             |
| 8     | Médico               |
| 9     | Professor            |
| 10    | Programador          |

Fonte: Elaborada pelo autor.

Tabela 15: Locais de lazer, e suas distribuições de probabilidade, considerados na geração de dados artificiais

| Local |             | População 1 |               | População 2 |               |
|-------|-------------|-------------|---------------|-------------|---------------|
|       |             | Média       | Desvio Padrão | Média       | Desvio Padrão |
| 1     | Academia    | 300         | 30            | 90          | 50            |
| 2     | Bar         | 150         | 15            | 280         | 80            |
| 3     | Restaurante | 100         | 10            | 150         | 70            |
| 2     | Cinema      | 150         | 15            | 50          | 25            |
| 3     | Shopping    | 90          | 9             | 240         | 75            |

Fonte: Elaborada pelo autor.

delimitados pelos valores 1 e 10. O valor estipulado é categorizado conforme profissões apresentadas na Tabela 14, que relaciona o conjunto de 10 diferentes profissões consideradas no experimento.

Após definidos os localizações referentes à permanência dos indivíduos nos seus respectivos locais de trabalho e instituições de ensino, o processo de geração de dados artificiais passa a produzir dados que remetem à permanência dos indivíduos em locais de seu interesse. Para tal, o experimento considera um conjunto de 5 diferentes tipos de locais, conforme apresentado na Tabela 15. Por considerar 2 populações de pessoas, o tempo de permanência em cada categoria de local é estipulado de forma distinta para cada população. Assim, a Tabela 15 também apresenta os valores de média e desvio padrão utilizados na adoção de distribuição de probabilidade normal para definição do tempo, em minutos, que cada indivíduo permaneceu em cada localidade.

- Número de locais de lazer frequentados: inicialmente, o processo define o número de locais de lazer frequentados por cada indivíduo. Para esta finalidade, o processo utiliza a distribuição de probabilidade normal, considerando o valor de média igual à 200 e desvio padrão igual à 50.

- **Localização de lazer:** a próxima etapa consiste em, através de um processo iterativo com base no número de locais estipulado, gerar informações sobre os locais de lazer frequentados. Primeiro, o processo define a categoria do local de lazer, através de distribuição de probabilidade uniforme, delimitando os valores gerados entre os limites 1 e 5, conforme dados relatados na Tabela 15. Em seguida, é definida a data em que o indivíduo tenha frequentado o local, sendo gerado o dia através de distribuição de probabilidade uniforme, estando limitado entre os valores 1 e 365. Com a data definida, o processo define o horário de ingresso na localização de lazer. Para tal, utiliza a distribuição de probabilidade uniforme para definir a hora, delimitado entre os valores 5 e 23, e o minuto do horário, delimitado entre os valores 0 e 59. Com a data inicial definida, o processo apura o tempo de permanência, que por consequência define a data final da permanência em tal localização, sendo definido através de distribuição de probabilidade normal, considerando os valores de média e desvio padrão estipulados para a população a qual pertence o indivíduo. Com a localização definida, o processo verifica se o período de permanência na localização conflita com o período das localizações geradas anteriormente. Estando em conflito, a localização é desconsiderada e o processo de geração de dados artificiais passa a gerar novamente a localização que represente uma localidade de lazer do indivíduo.

Por fim, o processo inicia a geração de dados referentes às localizações em que o indivíduo tenha permanecido em sua residência. Para cada dia do ano, do primeiro ao último, o experimento apura o provável tempo em que o indivíduo tenha permanecido em sua residência. Para tal, são verificados os locais anteriormente gerados em que o indivíduo esteve presente na respectiva data.

Não havendo locais frequentados, o processo estima um horário inicial e o tempo de permanência do usuário em sua residência, sendo definidos em duas etapas: *i.* definição do tempo, em minutos, de permanência através de distribuição de probabilidade normal, considerando o valor de média igual à 650 e desvio padrão igual à 250, estando limitado ao valor de 1440, o que representa o valor máximo de minutos em um dia. *ii.* estima a quantidade de minutos do horário inicial, através da distribuição de probabilidade uniforme, estando o valor delimitado entre os valores 0 e a diferença entre o valor 1440 e o tempo de permanência, estipulado na etapa anterior. Com posse de tais dados, o horário inicial da localização é definido, tomando como base a respectiva data, sendo o horário de início definido a partir da conversão da quantidade de minutos definida em horas (ex: 120 → 02:00). A partir da definição do horário inicial, o horário final é definido pela adição do tempo de permanência gerado, em minutos.

Por outro lado, havendo locais gerados em que o indivíduo esteve presente na respectiva data, o experimento considera a lista de locais frequentados, ordenados pelo horário que ocorreram. De forma iterativa, o processo considera o intervalo entre as localizações para definir o tempo de permanência que o indivíduo esteve em sua residência. Assim, o horário inicial é definido com base no horário final da localização anterior, ou às 00:00 em se tratando da primeira apuração do dia. O tempo de permanência dá-se através de distribuição de probabilidade

normal, contudo, os valores de média e desvio padrão considerados são variáveis para cada intervalo, em minutos, apurado. Dessa forma, o valor da média é definido como  $1/2$  do intervalo, enquanto o desvio padrão é definido como  $1/3$  do mesmo intervalo, delimitando o valor gerado no próprio valor total do intervalo. A partir da definição dos dados, o horário final é estipulado com base no horário inicial obtido, sendo adicionado o tempo de permanência estipulado, em minutos.

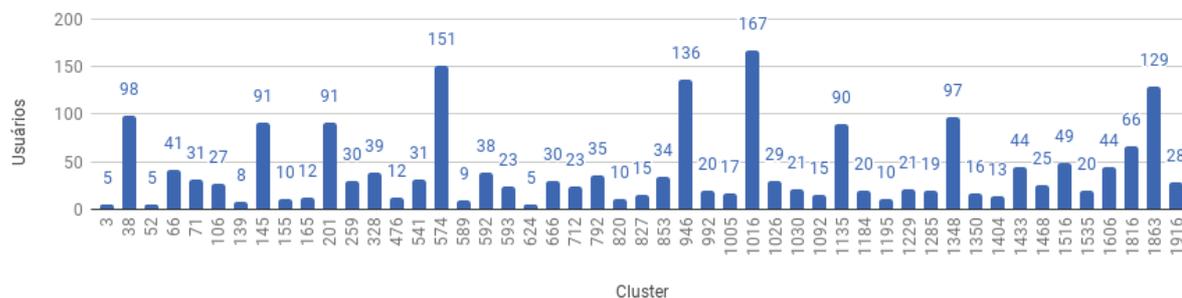
## 5.2 Cenários de avaliação

Com o aporte do fato de a comunidade acadêmica ter-se utilizado de cenários para avaliar aplicações baseadas em contexto e computação ubíqua (SATYANARAYANAN, 2001)(FRANCO et al., 2011)(OLIVEIRA et al., 2013), esta seção é destinada à apresentação e avaliação através da utilização de cenários simulados, onde fez-se uso de geração de dados artificiais com o intuito de avaliar o modelo proposto neste trabalho. Para tal, foram estabelecidos três diferentes cenários, sendo cada um composto por uma população de pessoas com diferentes características, as quais representam os hábitos dos indivíduos da população no que se refere aos locais de interesse frequentados, conforme destacado na descrição de cada cenário de avaliação, evidenciando o funcionamento dos processo de clusterização e análise visual propostos neste trabalho.

Conforme característica existente na técnica de clusterização *Affinity Propagation*, cada *cluster* originado possui um elemento que remete ao seu exemplar que, por sua vez, reúne as características que melhor definem as características do respectivo *cluster*. Assim, a forma de identificação de cada *cluster* adotada nos cenários de avaliação considera a identificação de seu exemplar.

Considerando outra característica fundamental da técnica de clusterização *Affinity Propagation*, a definição do número de iterações utilizadas no processamento dos *clusters* a serem originados é algo crucial em virtude da qualidade na formalização dos *clusters* a serem gerados. Assim, utilizando um número de iterações menor que o ideal pode resultar em um cenário final onde o processo de clusterização não agrupe corretamente os indivíduos, de acordo com suas características. Em virtude disto, foram realizados testes para verificar a eficácia do processo de clusterização, onde chegou-se ao número de 30 iterações, originando o cenário ideal de *clusters* formalizados.

Importante salientar que, além dos cenários abordados neste trabalho, diversos outros podem ser explorados com a utilização da ferramenta de análise visual proposta neste trabalho. Dessa forma, evidencia-se possibilidade de descoberta de oportunidades/lacunas diante do cenário analisado, sendo esta a característica fundamental do processo de análise visual.

Figura 31: Quantidade de usuários por *cluster* originado no cenário de avaliação 1

Fonte: Elaborada pelo autor.

Tabela 16: *Clusters* mais expressivos obtidos no cenário de avaliação 1

| <i>Cluster</i> | Usuários | Percentual |
|----------------|----------|------------|
| 574            | 151      | 7,55       |
| 946            | 136      | 6,80       |
| 1016           | 167      | 8,35       |
| 1863           | 129      | 6,45       |
| Total          | 583      | 29,15      |

Fonte: Elaborada pelo autor.

### 5.2.1 Cenário 1 - Todas as atividades dos usuários

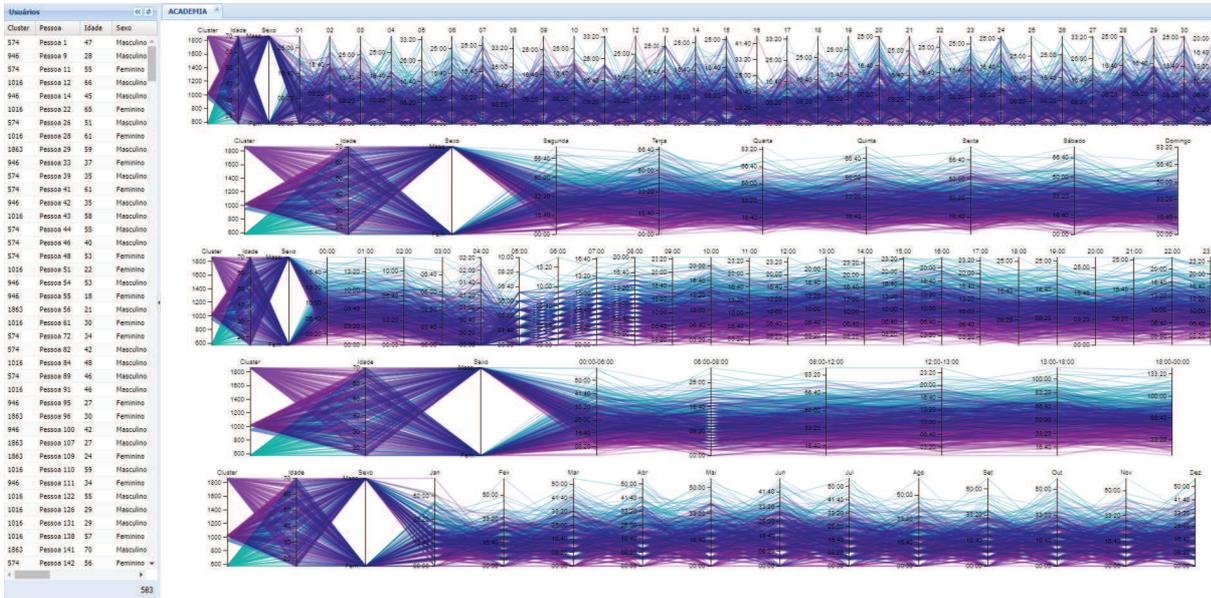
O primeiro cenário de avaliação consiste em considerar as atividades geradas para a população 1, cuja atividades foram geradas de acordo com o descrito na Seção 5.1. Neste cenário foram consideradas todas atividades dos usuários, sem aplicar restrição ao conjunto de dados artificiais gerado.

A avaliação deste cenário inicia com a demonstração dos *clusters* originados a partir do processo de clusterização dos usuários, mediante suas atividades. Através do processo de clusterização obteve-se o total de 49 *clusters* originados, conforme destacado na Figura 31.

Do total de *clusters* gerados, 4 *clusters* (574, 946, 1016, 1863) apresentaram uma maior concentração de usuários, ou seja, apresentam um padrão expressivo perante os demais *clusters*, considerando uma concentração superior ou igual à 5% do número de pessoas da população, representando 29,15% do número total de usuários considerados no cenário de avaliação, conforme demonstrado na Tabela 16, que indica o número de usuários presentes em cada um dos *clusters* mais populosos, bem como o respectivo percentual representado pelo número de usuários presentes em cada *cluster*.

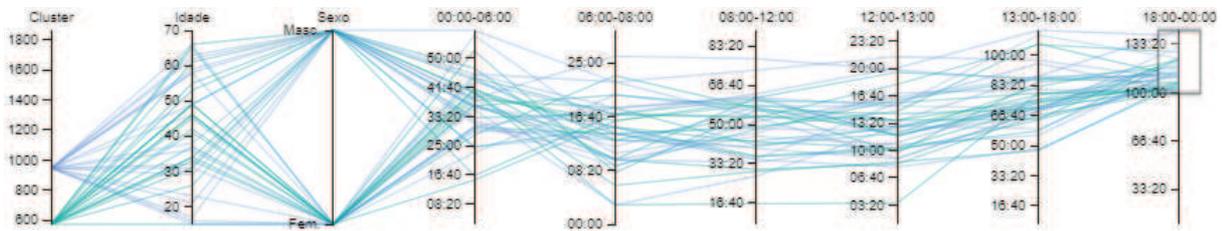
Com o intuito de compreender o comportamento dos indivíduos pertencentes aos *clusters* que apresentaram um padrão expressivo perante os demais *clusters*, estes foram explorados através da ferramenta de análise visual proposta no trabalho, conforme demonstrado na Figura 32, onde é apresentado o comportamento dos indivíduos pertencentes aos *clusters*, considerando suas atividades em locais categorizados como “academia”. Na Figura 32, é possível constata-

Figura 32: Comportamento de usuários em locais categorizados como “academia”



Fonte: Elaborada pelo autor.

Figura 33: Análise do sistema de coordenadas paralelas sobre o contexto de períodos do dia do cenário de avaliação 1



Fonte: Elaborada pelo autor.

tar, através da diferenciação de cores, as diferenças nos padrões de hábitos dos integrantes da população.

Através da análise do sistema de coordenada paralelas sobre o contexto de períodos do dia, é possível perceber que as pessoas que costumam frequentar academias com uma maior frequência entre os horários de 18:00 e 00:00, também apresentam uma alta frequência em academias entre os horários de 13:00 e 18:00, conforme apresentado na Figura 33. Além disso, é possível constatar que todas as pessoas que possuem uma alta frequência em academias entre os horários 18:00 e 00:00 pertencem somente à 2 clusters (574, 946).

Outra possibilidade de análise a ser realizada diante do cenário apresentado, é a indicação de correlação entre os eixos do sistema de coordenadas paralelas apresentadas no cenário de avaliação. Assim, é possível constatar a existência de indícios que revelam correlação negativa entre os eixos que remetem aos períodos do dia 00:00 à 06:00 e 06:00 à 08:00, conforme demonstrado na Figura 34.

Com a interação sobre o conjunto de sistemas de coordenadas paralelas previsto na interface

Figura 34: Índícios de correlação negativa entre os eixos do sistema de coordenadas paralelas sobre o contexto de períodos do dia

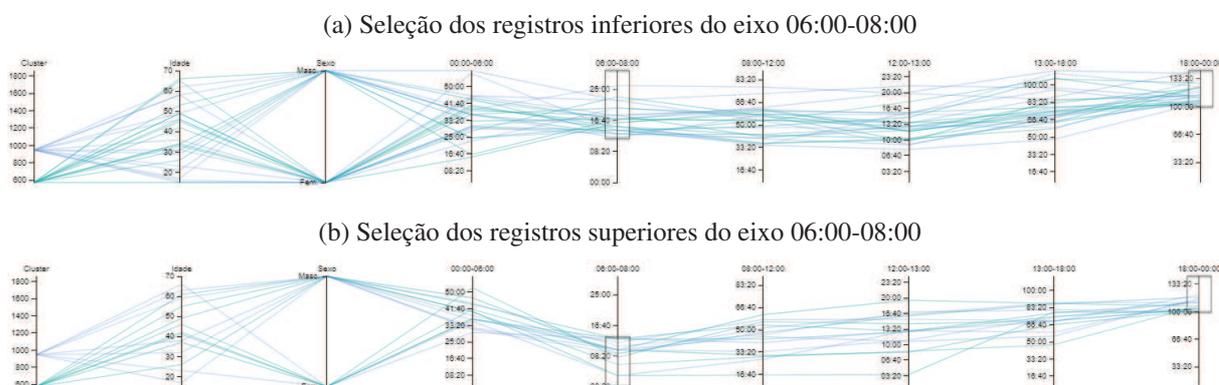
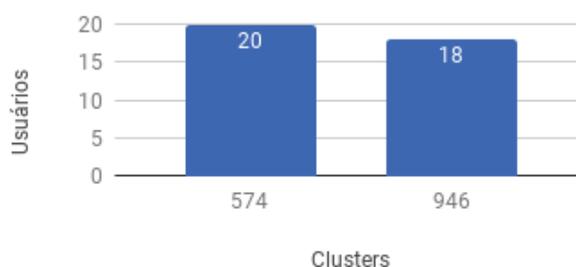


Figura 35: Quantidade de usuários por *cluster* mediante adoção da técnica *brushing* sobre eixo do sistema de coordenadas paralelas no cenário de avaliação 1



Fonte: Elaborada pelo autor.

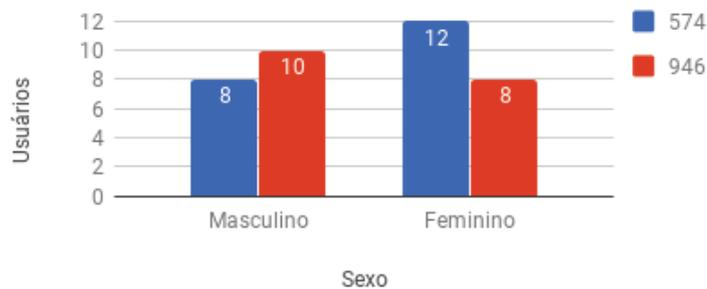
da ferramenta de análise visual proposta neste trabalho, é possível acompanhar, mediante a tabela da interface, os registros que são filtrados através da adoção da técnica de *brushing* sobre os sistemas de coordenadas paralelas. Com isto, torna-se possível também a extração de métricas sobre os dados que não refletem diretamente ao tempo de permanência nos locais que representam os interesses dos indivíduos.

A Figura 35, por sua vez, exibe o número de usuários por *cluster* resultantes da adoção da técnica de *brushing* sobre o eixo do sistema de coordenadas paralelas apresentadas neste cenário de avaliação.

Diante do exposto no cenário de avaliação, considerando as pessoas que apresentam uma alta frequência de permanência academias durante o período do dia entre os horários de 18:00 às 00:00, é possível analisar o número de pessoas por *cluster* formalizado através do sexo dos seus integrantes. Dessa forma, conforme demonstrado na Figura 36, há uma maior concentração de pessoas do sexo masculino no *cluster* 946, enquanto há uma maior concentração de pessoas do sexo feminino no *cluster* 574.

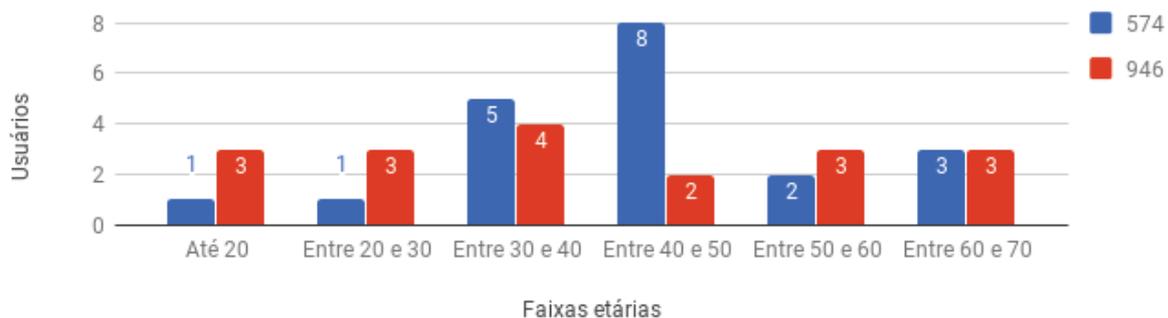
Através do cenário apresentado, torna-se possível a análise das faixas etárias das pessoas que possuem uma alta frequência em academias durante o período do dia que compreende os horários das 18:00 às 00:00, resultantes do cenário explorado na ferramenta de análise visual.

Figura 36: Sexo dos usuários por *cluster* mediante adoção da técnica *brushing* sobre eixo do sistema de coordenadas paralelas no cenário de avaliação 1



Fonte: Elaborada pelo autor.

Figura 37: Faixas etárias dos usuários por *cluster* mediante adoção da técnica *brushing* sobre eixo do sistema de coordenadas paralelas no cenário de avaliação 1



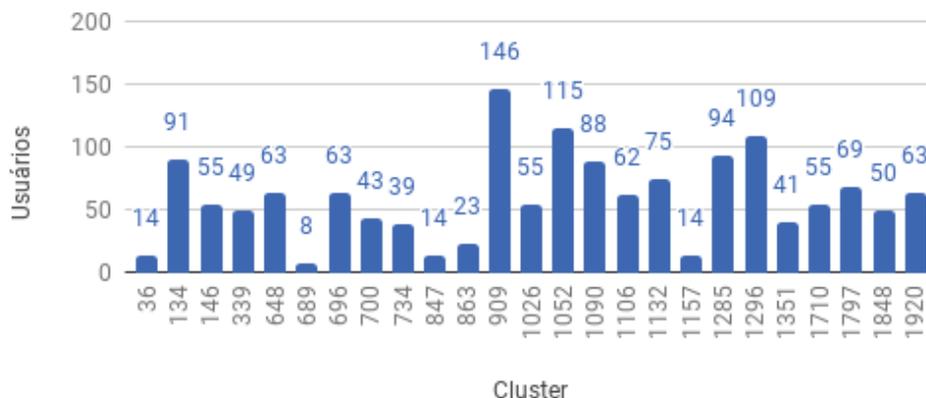
Fonte: Elaborada pelo autor.

Assim, de acordo com a Figura 37, é possível constatar que a grande maioria das pessoas pertencentes ao *cluster* 574 estão compreendidos entre as faixas etárias de 30 a 40 anos e de 40 a 50 anos. Por outro lado, as pessoas pertencentes ao *cluster* 946 estão divididos de forma praticamente igual entre as diferentes faixas etárias consideradas no cenário de avaliação.

### 5.2.2 Cenário 2 - Atividades de pessoas que exercem atividade profissional e frequentam bares

O segundo cenário de avaliação analisa a população 2, considerando somente registros de atividades de usuários que exercem atividade profissional e que tenham frequentado locais de lazer categorizados como “bares”, de acordo com os dados artificiais retratados na Seção 5.1.

Diante do previsto no modelo proposto, definiu-se, na etapa de transformação, a parametrização de acordo com o cenário previsto na análise que, por sua vez, realiza o procedimento de clusterização considerando as restrições estipuladas na ferramenta. Assim, somente serão consideradas as atividades que remetem à permanência de pessoas no local de trabalho e em bares, para aquelas que exercem atividade profissional. Como resultado do pré processamento adotado na etapa de transformação dos dados, obteve-se atividades de 1498 usuários, contemplando

Figura 38: Quantidade de usuários por *cluster* originado no cenário de avaliação 2

Fonte: Elaborada pelo autor.

Tabela 17: *Clusters* mais expressivos obtidos no cenário de avaliação 2

| <i>Cluster</i> | Usuários | Percentual |
|----------------|----------|------------|
| 909            | 146      | 9,75       |
| 1052           | 115      | 7,68       |
| 1296           | 109      | 7,28       |
| Total          | 370      | 24,70      |

Fonte: Elaborada pelo autor.

74,90% do número total de pessoas da população.

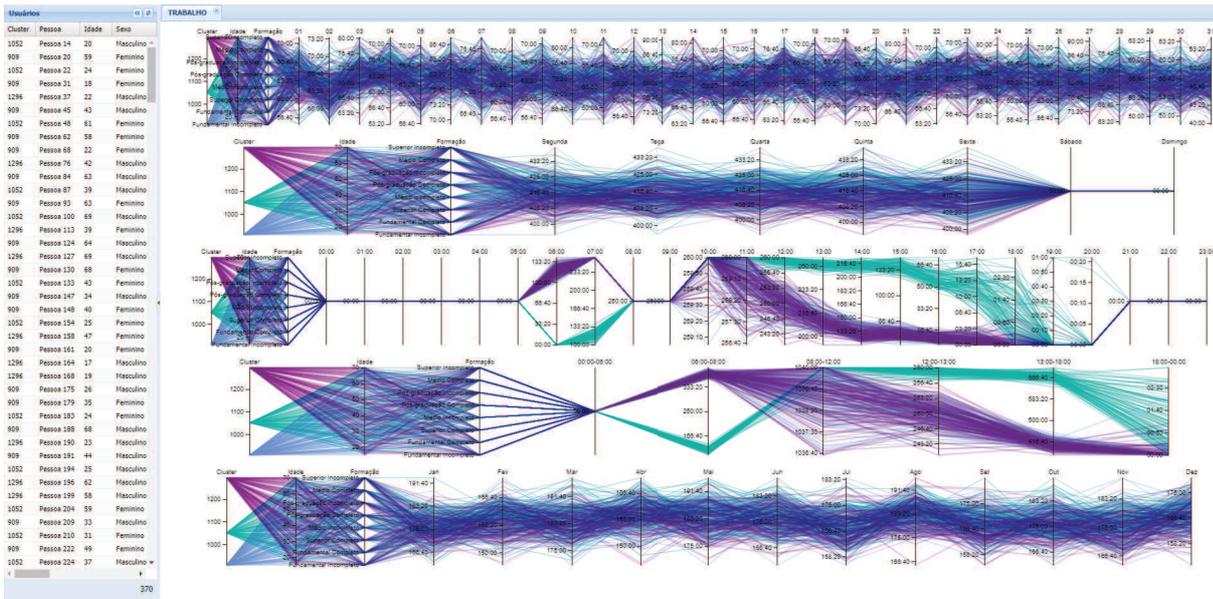
A partir do contexto estipulado, a análise do cenário de avaliação inicia-se com a constatação do número de *clusters* originados no processo de clusterização, computado através das atividades dos usuários. Com a adoção do processo de clusterização, obteve-se o total de 25 *clusters* originados, conforme demonstrado na Figura 38.

Diante do número total de *clusters* originados, 3 *clusters* (909, 1052, 1296) destacam-se em virtude de apresentarem uma quantidade de usuários superior à 5% do número total de pessoas da população, o que representa aproximadamente 6,7% da população com a adoção das restrições previstas no cenário de avaliação. Tais *clusters* contemplam 370 pessoas do total de 1498 indivíduos considerados no cenário de avaliação, o que representa 24,70% da população do cenário, conforme demonstrado na Tabela 17, que indica o número de usuários presentes em cada um dos *clusters* mais populosos, bem como o respectivo percentual representado pelo número de usuários presentes em cada *cluster*.

Buscando compreender as atividades dos usuários pertencentes aos *clusters* que apresentam, dentre os usuários que o compõem, um expressivo padrão em suas atividades, estas são submetidas à ferramenta de análise visual proposta neste trabalho, conforme demonstrado na Figura 39, onde é possível notar a clara distinção entre as atividades dos usuários de diferentes *clusters*, identificados por cores, o que sugere o correto funcionamento do processo de clusterização adotado na concepção deste trabalho.

Figura 39: Atividades dos usuários considerados no cenário de avaliação 2

(a) Trabalho



(b) Bar

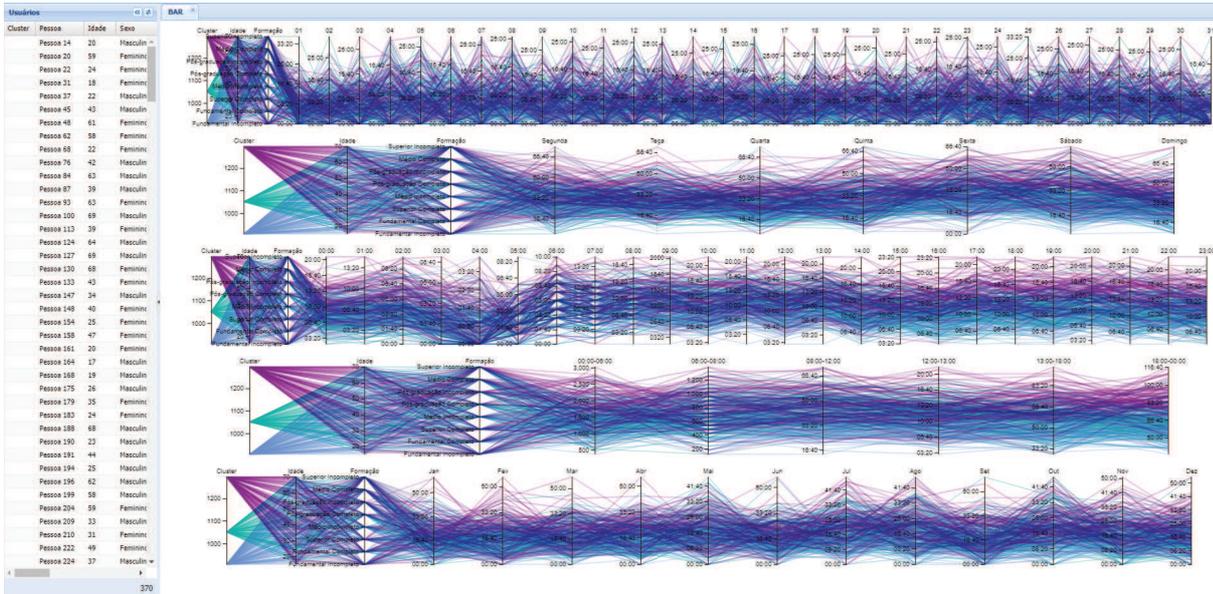
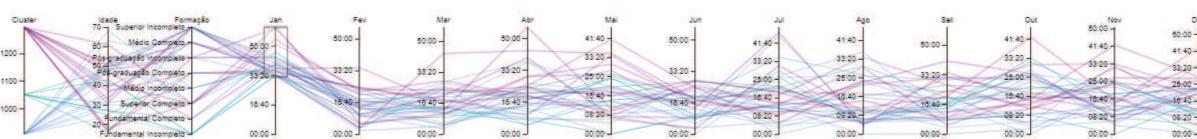
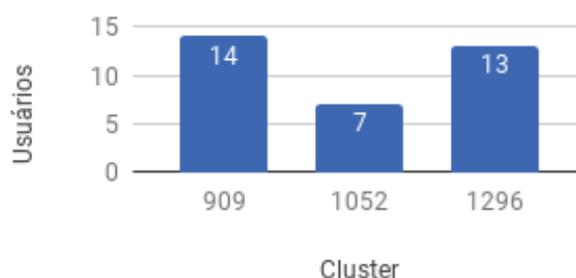


Figura 40: Técnica de *brushing* aplicada no sistema de coordenadas paralelas sob o contexto referente aos meses do ano em bares



Fonte: Elaborada pelo autor.

Figura 41: Quantidade de usuários por *cluster* existentes no cenário de avaliação, considerando *clusters* com padrões expressivos



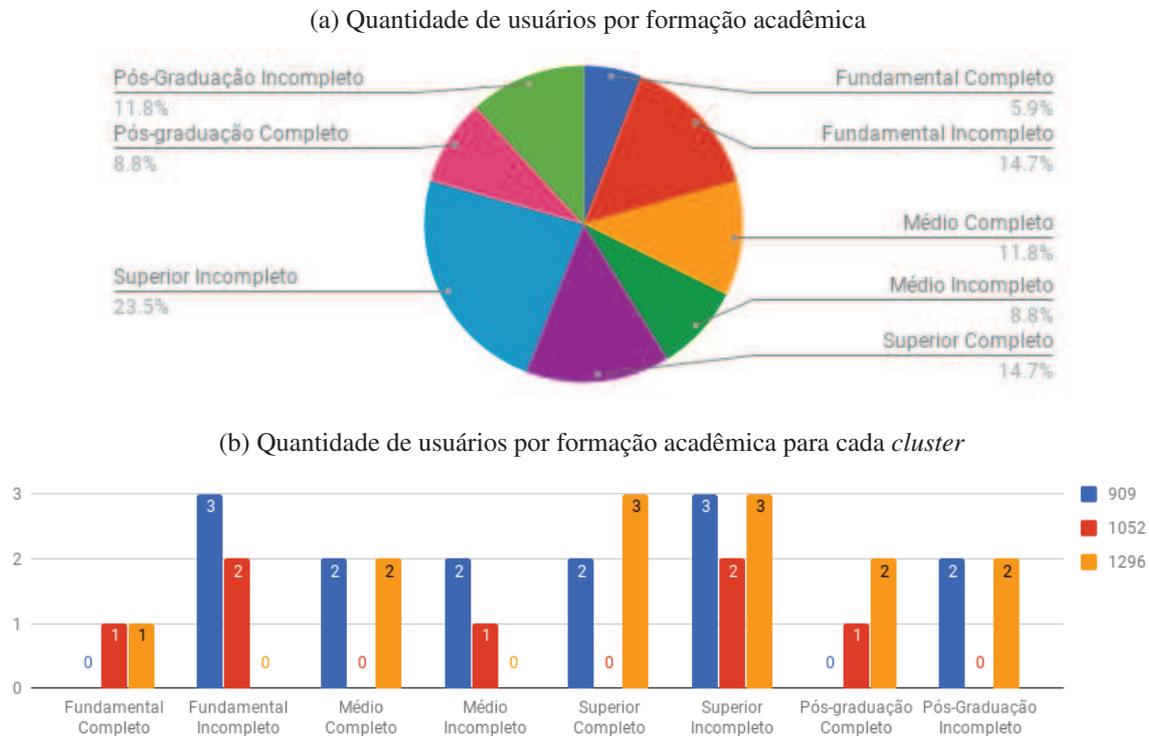
Fonte: Elaborada pelo autor.

Ao analisar as atividades das pessoas em locais de lazer, os quais referem-se à bares, sob a perspectiva que analisa o tempo de permanência considerando os meses do ano, é possível notar que, através da adoção da técnica de *brushing*, conforme indicado na Figura 40, pessoas que apresentam uma maior permanência em bares no mês de janeiro, tendem a demonstrar uma baixa permanência na mesma categoria de local nos dois meses subsequentes (fevereiro e março). O subconjunto de indivíduos que apresentam um alto nível de permanência em bares durante o mês de janeiro contempla 34 indivíduos, o que representa, de forma aproximada, somente 2,3% do total de indivíduos considerados no cenário de avaliação. Assim, a Figura 41 exibe, de forma gráfica, o número de indivíduos pertencentes à cada cluster, respeitando as condições exploradas no cenário.

Diante do cenário de avaliação apresentado, considerando indivíduos que apresentam uma alta permanência em locais de lazer categorizados como bares, torna-se possível a realização de uma análise sobre a formação acadêmica dos indivíduos pertencentes ao cenário de avaliação. Desta forma, constata-se, através da Figura 42a, que as formações acadêmicas que abrangem um maior número de indivíduos são superior incompleto (23,5%), superior completo (14,7%) e fundamental incompleto (14,7%), totalizando 52,9% do total de indivíduos considerados no cenário de avaliação. Por outro lado, a Figura 42b detalha o número de pessoas, pertencentes à cada *cluster*, classificados pela formação acadêmica que possuem.

A ferramenta de análise proposta neste trabalho permite a utilização conjunta dos sistemas de coordenadas paralelas, analisando cenários que englobam diferentes categorias de locais. Assim, uma segunda possibilidade de análise a ser realizada sobre o cenário de avaliação ex-

Figura 42: Quantidade de usuários por formação acadêmica considerados no cenário de avaliação 2



posto, é a análise dos hábitos dos usuários, diante do contexto de locais de lazer categorizados como bares, no entanto, considerando aqueles indivíduos que possuem uma alta presença no local de trabalho, identificados pela análise do período do dia que compreende os horários das 13:00 às 18:00.

Dessa forma, conforme demonstrado pela Figura 43a, que ilustra a adoção da técnica de *brushing* sobre o eixo das coordenadas paralelas referente ao período do dia analisado, no âmbito das atividades referentes ao local de trabalho, é possível identificar os indivíduos que enquadram-se na restrição imposta no cenário de avaliação. Com isto, é possível estabelecer que tais indivíduos pertencem ao *cluster* 1052 que, por sua vez, contém 115 indivíduos, o que corresponde à 7,68% da população total considerada no cenário de avaliação.

Considerando somente os indivíduos revelados através dos sistema de coordenadas paralelas que referem-se às atividades dos usuários no âmbito do local de trabalho, é realizada uma análise considerando suas atividades em locais de lazer categorizados como bares. Assim, conforme demonstrado na Figura 43b, que restringe o conjunto de usuários, considerando somente aqueles que possuem formação acadêmica de pós graduação, completa ou incompleta, resultando no total de 34 indivíduos, o que representa cerca de 2,3% da população total considerada no cenário de avaliação.

Diante do exposto na ferramenta de análise visual, é possível perceber que, de forma geral, os indivíduos evidenciados na visualização tendem a apresentar uma frequência média/baixa em

Figura 43: Técnica de *brushing* nos sistemas de coordenadas paralelas construídas no cenário de avaliação 2

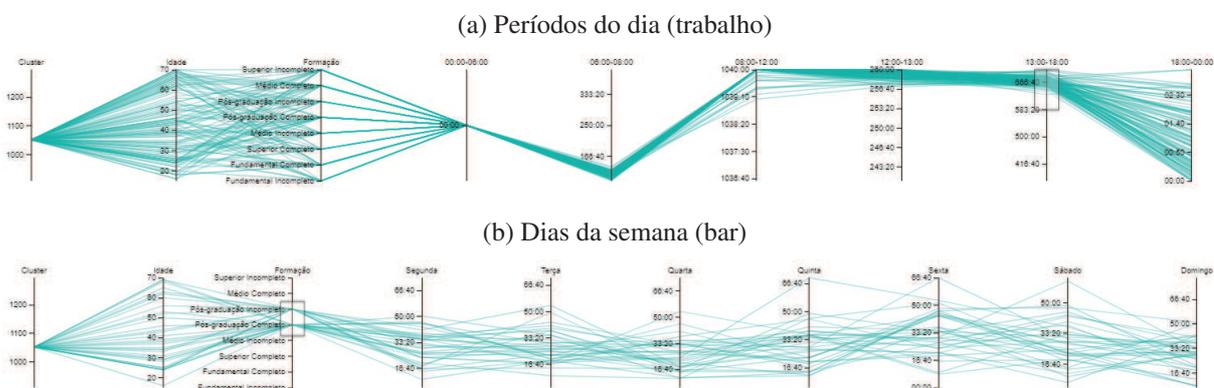
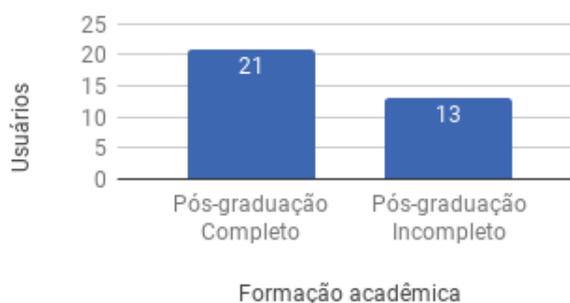


Figura 44: Quantidade de usuários por formação acadêmica considerada no cenário de avaliação 2



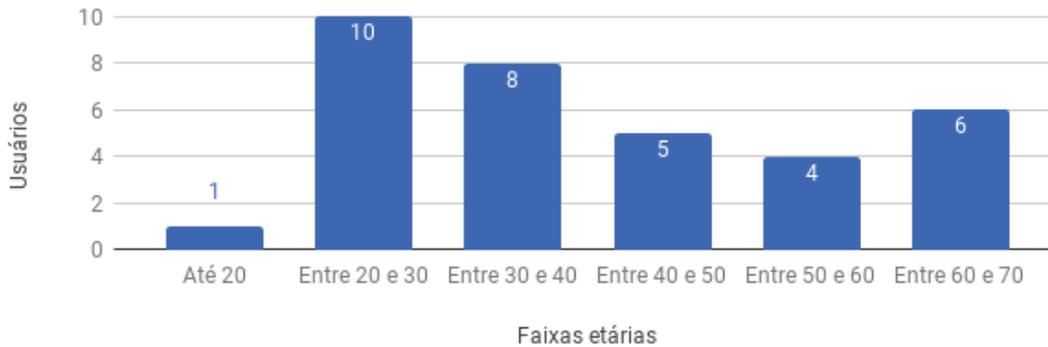
Fonte: Elaborada pelo autor.

relação à sua permanência em bares durante os dias da semana que correspondem às segundas-feiras, terças-feiras e quartas-feiras. No entanto, apresenta um gradativo aumento na permanência nesta mesma categoria de local durante os dias da semana que correspondem às quintas-feiras, sextas-feiras e sábados, tornando a apresentar uma frequência média/baixa aos domingos, conforme demonstrado na Figura 43b.

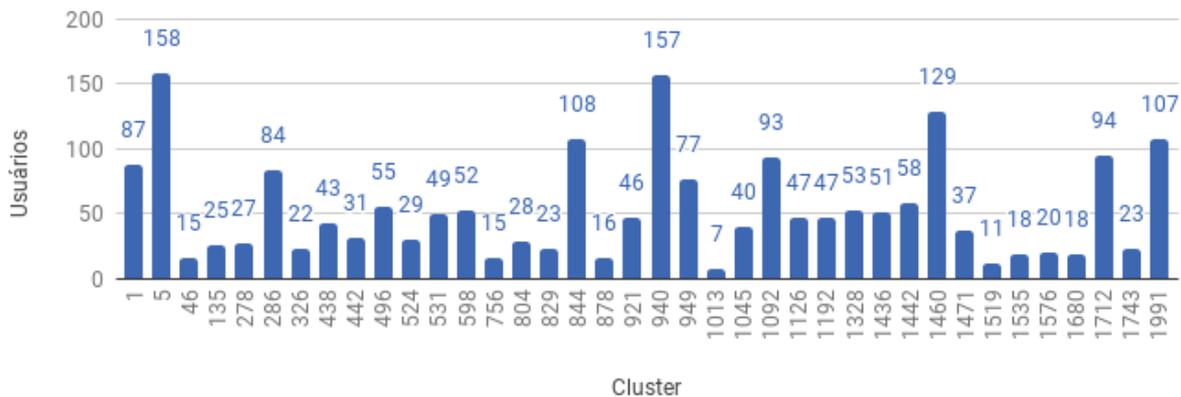
Diante das características do cenário de avaliação analisado, é possível mensurar as características dos usuários no que refere-se à formação acadêmica dos usuários existentes no cenário de avaliação. Assim, conforme demonstrado na Figura 44, percebe-se que, do total de 34 pessoas incluídas na análise, 21 (61,76%) possuem pós-graduação completa, enquanto 13 (38,24%) possuem pós-graduação incompleta.

Em contrapartida, também é possível realizar a análise sob a perspectiva da idade dos usuários pertencentes ao cenário de avaliação. Assim, de acordo com o ilustrado pela Figura 45, fica evidenciado que 10 usuários (29,41%) que enquadram-se no cenário analisado possuem sua idade compreendida na faixa etária que compreende as idades entre de 20 e 30 anos, enquanto apenas 1 usuário (2,94%) possui idade inferior ou igual à 20 anos.

Figura 45: Quantidade de usuários por faixa etária considerada no cenário de avaliação 2



Fonte: Elaborada pelo autor.

Figura 46: Quantidade de usuários por *cluster* originado no cenário de avaliação 3

Fonte: Elaborada pelo autor.

### 5.2.3 Cenário 3 - Todas as atividades dos usuários, clusterizadas considerando o valor do 1º quartil como valor de preferência

O terceiro cenário de avaliação contempla uma nova população, onde as atividades dos locais de lazer são geradas com base nas distribuições de probabilidade normal, as quais são descritas na Tabela 15, no entanto, para a concepção desta população utilizou-se o valor de média da população 1 e de desvio padrão da população 2. Neste cenário foram consideradas todas atividades dos usuários, sem aplicar restrição ao conjunto de dados artificiais gerado.

Neste cenário de avaliação, diferente do proposto nos cenários de avaliação anteriores, onde foi utilizado o valor de preferência como sendo o valor da mediana dos valores de similaridade estipulado entre os usuários, este cenário de avaliação considera o valor referente ao primeiro quartil para ser definido como o valor de preferência utilizado no processo de clusterização.

A avaliação deste cenário inicia com a demonstração dos *clusters* originados a partir do processo de clusterização dos usuários, mediante suas atividades. Através do processo de clusterização obteve-se o total de 38 *clusters* originados, conforme destacado na Figura 46.

Tabela 18: *Clusters* mais expressivos obtidos no cenário de avaliação 3

| <i>Cluster</i> | Usuários | Percentual |
|----------------|----------|------------|
| 5              | 158      | 7,90       |
| 844            | 108      | 5,40       |
| 940            | 157      | 7,85       |
| 1460           | 129      | 6,45       |
| 1991           | 107      | 5,35       |
| Total          | 659      | 32,95      |

Fonte: Elaborada pelo autor.

Do total de *clusters* gerados, 5 *clusters* (5, 844, 940, 1460, 1991) apresentaram uma maior concentração de usuários, ou seja, apresentam um padrão expressivo perante os demais *clusters*, considerando uma concentração superior ou igual à 5% do número de pessoas da população, representando 32,95% do número total de usuários considerados no cenário de avaliação, conforme demonstrado na Tabela 18, que indica o número de usuários presentes em cada um dos *clusters* mais populosos, bem como o respectivo percentual representado pelo número de usuários presentes em cada *cluster*.

Analisando os hábitos dos indivíduos pertencentes aos *clusters* que apresentaram um padrão expressivo perante os demais *clusters*, estes foram explorados através da ferramenta de análise visual proposta no trabalho, considerando suas atividades em locais categorizados como “restaurantes”.

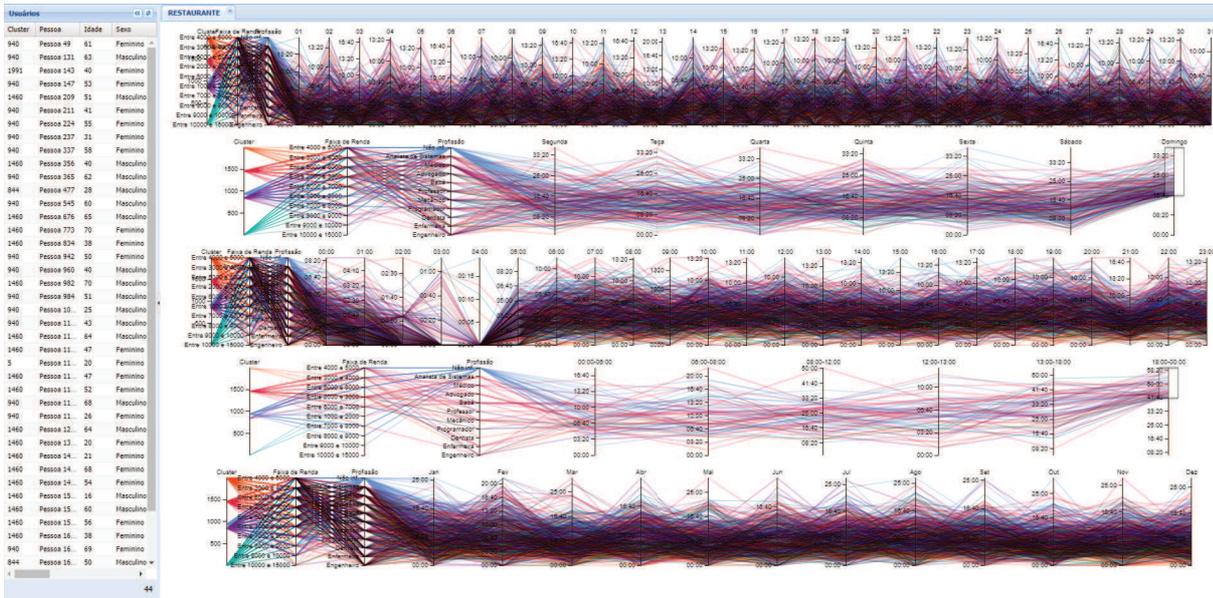
Desta forma, ao analisar o comportamento dos indivíduos que apresentam uma alta permanência em restaurantes aos domingos durante o período do dia que compreende os horários entre 18:00 e 00:00, evidenciou-se os indivíduos que correspondem ao critérios da análise, conforme demonstrado na Figura 47, onde é demonstrada a adoção da técnica *brushing* sobre os eixos que correspondem aos contextos de dias da semana e períodos do dia do sistema de coordenadas paralelas, os quais atuam no âmbito das atividades dos indivíduos em locais categorizados como restaurantes. Tal análise resulta em um total de 44 indivíduos, o que corresponde à 2,2% do número total de indivíduos existentes neste cenário de avaliação.

Diante do exposto pelo cenário de avaliação, é possível analisar os hábitos dos indivíduos e classificá-los de acordo com a sua faixa de renda, de acordo com o considerado no cenário. A Figura 48 retrata a quantidade de usuários existentes no cenário de avaliação por faixa de renda considerada, onde constata-se que a maioria das faixas de renda dos indivíduos está compreendida entre os valores de R\$ 2.000,00 e R\$ 6.000,00.

Adicionalmente, como funcionalidade presente no processo de análise visual proposto na ferramenta, é possível realizar a seleção de  $n$  registros com o objetivo de visualizar as atividades de um conjunto de diferentes indivíduos, expostos no sistema de coordenadas paralelas analisado, conforme demonstrado pela Figura 49, que ilustra a seleção dos registros que referem-se às atividades dos dois indivíduos pertencentes ao *cluster* 844.

Conforme previsto no modelo proposto neste trabalho, a utilização do procedimento de

Figura 47: Técnica de *brushing* sobre os eixos da análise realizada no cenário de avaliação 3



Fonte: Elaborada pelo autor.

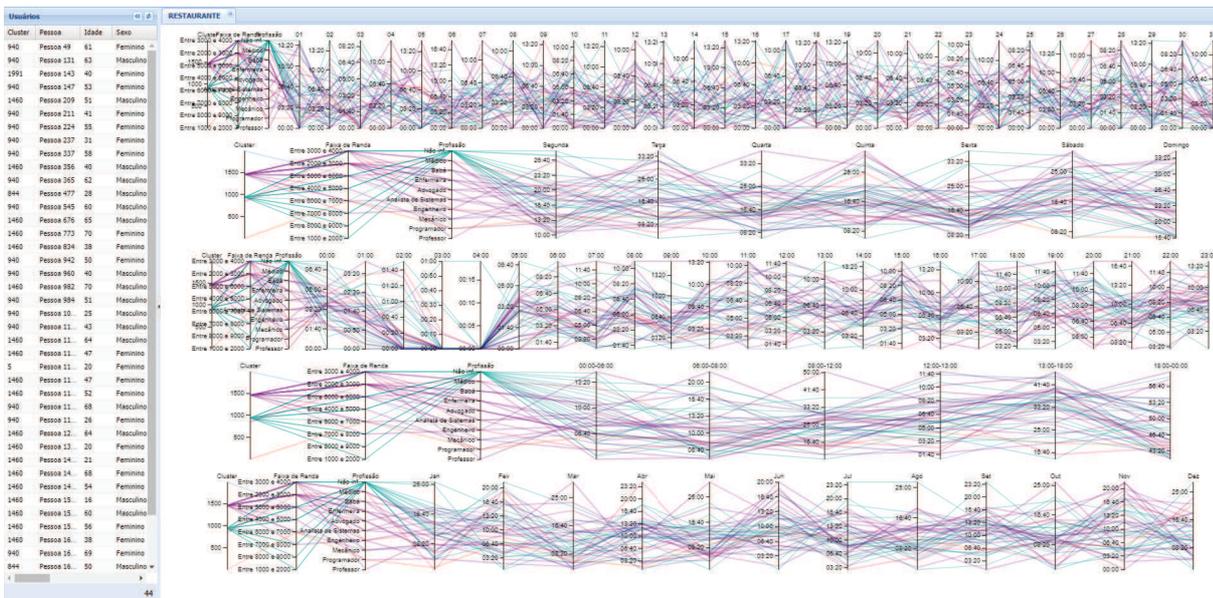
Figura 48: Quantidade de usuários por faixas de renda consideradas no cenário de avaliação 3



Fonte: Elaborada pelo autor.



Figura 50: Sistemas de coordenadas paralelas resultantes do processo de retroalimentação, evidenciando o subconjunto de indivíduos considerado no cenário de avaliação 3



Fonte: Elaborada pelo autor.

## 6 CONSIDERAÇÕES FINAIS

O modelo de análise e agrupamento de populações de pessoas proposto neste trabalho se utilizou de técnicas de clusterização e análise visual. A população foi analisada sob o ângulo de seus hábitos de frequência à locais de interesse. O Geosocial é constituído por dois módulos principais: o primeiro é o módulo de clusterização que utiliza a técnica de clusterização denominada *Affinity Propagation*. O segundo é o módulo de visualização que utiliza a técnica de análise visual, denominada Coordenadas Paralelas.

Através do procedimento de clusterização proposto, torna-se possível a identificação de subconjuntos de indivíduos de uma população, os quais apresentam um determinado nível de similaridade entre si no que se refere aos seus hábitos, computados a partir da frequência com que visitam e permanecem em seus locais de interesse, analisados sob a perspectiva semântica.

O mapeamento das atividades dos usuários compõe o perfil destes indivíduos. Este são agrupados de forma que a análise visual extraia *insights* que permitam a identificação de padrões de comportamento do grupo de indivíduos que compõem a análise. Diante do cenário resultante, o usuário analista pode identificar um nicho de mercado existente ou detectar certas carências de um grupo de pessoas de determinada região, por exemplo. As carências identificadas podem ser comerciais, um novo supermercado na região, por exemplo; ou de estrutura, podendo ser novas vias públicas, creches, escolas entre outras.

Por meio da técnica de clusterização utilizada houve a possibilidade da formação de uma rede social espontânea, considerando os hábitos e comportamentos dos usuários, ou seja, a formalização da rede dá-se de forma autônoma, de acordo com as informações à respeito do cotidiano das pessoas, considerando as características dos seus locais de interesse.

O Geosocial foi submetido à avaliação através da utilização de dados artificiais gerados com o intuito de representar as atividades de indivíduos que compõem populações de pessoas, sendo esta uma alternativa à utilização do aplicativo desenvolvido durante o desenvolvimento deste trabalho.

A avaliação do modelo constitui-se na geração de atividades de indivíduos compondo três diferentes populações, considerando um ano de atividade. Através da utilização da ferramenta de visualização de dados projetada, foi possível comprovar seu objetivo onde, através da técnica de análise visual utilizada na concepção deste trabalho, pode-se explorar os dados em busca de identificação de padrões existentes nos dados que, por sua vez, podem indicar lacunas a serem tratadas e oportunidades a serem exploradas. Outro objetivo do trabalho que pode ser observado pela ferramenta proposta foi o resultado do processo de clusterização proposto, identificando subconjuntos de indivíduos, dentro das respectivas populações, os quais apresentaram níveis de similaridades adequados.

Importante ressaltar que, além dos cenários abordados neste trabalho, diversos outros podem ser explorados com a utilização da ferramenta de análise visual proposta neste trabalho. Dessa forma, evidencia-se a possibilidade de descoberta de oportunidades/lacunas diante do cenário

analisado, sendo esta a característica fundamental do processo de análise visual. Assim, o modelo elaborado neste trabalho pode servir de auxílio à tomada de decisões para empreendedores e gestores públicos no que diz respeito à uma possível nova oportunidade de negócios ou na identificação de uma possibilidade de tornar um serviço mais eficaz à determinada população. Com isto, as avaliações dos cenários apresentados neste trabalho, bem como a possibilidade de análise de diversos outros cenários comprovam o potencial existente no modelo proposto neste trabalho.

O trabalho propôs o desenvolvimento de um aplicativo móvel que atuou na captura das informações referentes ao histórico de contexto dos indivíduos, contemplando os dados sobre as localizações frequentadas. Além disso, buscando tornar o aplicativo atrativo, incorporou-se uma série de funcionalidades no aplicativo com o intuito de disseminar seu uso em uma população com usuários reais. Contudo, não houve um grande número de usuários que aderiram ao uso do aplicativo desenvolvido, resultando em uma cenário onde o volume de dados coletados dos usuários não foi suficiente para a realização da análise da população, objetivo fundamental proposto neste trabalho. Assim, optou-se pela avaliação do modelo desenvolvido através da definição de cenários, utilizando dados gerados de forma artificial.

Ao analisar possibilidades de trabalhos futuros que podem ser propostos a partir da ideia defendida neste trabalho, é possível vislumbrar um cenário onde a captura das localizações que os usuários tenham frequentado possam ser obtidas de forma diferente da maneira proposta no trabalho. Uma possibilidade iminente, diante do contínuo avanço da área, é a utilização de reconhecimento facial utilizado para identificar indivíduos inseridos no contexto dos estabelecimentos frequentados.

Apesar da possibilidade de seu uso implicar em problemas que afetam a privacidade das pessoas, o uso da tecnologia de reconhecimento facial, realizado de forma consciente, pode resultar em um cenário eficaz na descoberta dos hábitos das pessoas no que refere-se aos seus locais de interesse frequentados. A viabilidade da utilização de técnicas de reconhecimento facial está amplamente fundamentada, onde constata-se que, mediante características da imagem, é possível realizar o reconhecimento de uma pessoa através de uma imagem com precisão de até 99% (KLONTZ; JAIN, 2013).

Assim, através de imagens obtidas por câmeras existentes no estabelecimentos frequentados pelos indivíduos, e posterior processamento destas por meio de algoritmos capazes de realizar a identificação das pessoas, é possível conceber uma nova forma de identificar os locais de interesse frequentados pelas pessoas, contabilizando o tempo em que permanecem nestes locais.

Desta forma, é extremamente importante salientar que o mapeamento das atividades de uma população de pessoas, através da utilização de técnicas de reconhecimento facial, as quais identificam os indivíduos em seus locais de interesse, não implicam em alterações a serem realizadas no modelo de análise e agrupamento de uma população de pessoas proposto neste trabalho, evidenciando o potencial do modelo proposto na realização de análises de uma população de pessoas, baseadas em informações que remetem aos hábitos desta população.

Outra possibilidade de trabalhos futuros pertinentes a partir da proposta deste trabalho, é a inclusão de indicadores estatísticos a partir da interação do usuário analista com a interface da ferramenta de análise visual proposta. Assim, o usuário poderia, com maior precisão, identificar indícios da população que o auxilie no processo de tomada de decisão, através da realização de inferências estatísticas, mediante os dados da população analisada. A adição de tal recurso incorporaria à ferramenta a capacidade de transcrição do cenário apresentado pelos conjuntos de sistemas de coordenadas paralelas em números, os quais refletiriam os dados estatísticos do cenário avaliado de forma visual.



## APÊNDICE A – APLICATIVO MÓVEL

Este apêndice é dedicado à apresentação do aplicativo móvel desenvolvido, evidenciando suas funcionalidades. Além das funcionalidades cruciais para a constituição deste trabalho, o desenvolvimento deste aplicativo compreende funcionalidades que forneçam aos usuários um a visualização das informações referentes aos seus históricos de localizações.

- Menu: o menu do aplicativo móvel é apresentado ao usuário de duas formas. A primeira ocorre caso o usuário não esteja registrado no aplicativo<sup>1</sup>, conforme demonstrado na Figura 51a. Neste momento, o usuário pode optar por duas ações: realizar o *login*, caso o mesmo já tenha realizado seu cadastro previamente e; realizar seu cadastro. Por outro lado, caso o usuário esteja registrado no aplicativo, o usuário pode optar por visualizar/editar seus dados cadastrais, alterar sua senha, convidar pessoas conhecidas a participar do projeto, visualizar os demais usuários pertencentes à sua rede, visualizar os locais frequentados e realizar *logout* do aplicativo móvel, conforme ilustrado na Figura 51b.
- Cadastro/Meus Dados: para estar apto a utilizar o aplicativo, o usuário deve possuir uma conta registrada na aplicação. Diante disto, o fluxo de cadastro está disponível para que o indivíduo possa realizar seu registro ou alterar seus dados cadastrais. A diferenciação ocorre mediante registro do usuário no aplicativo móvel. Não estando registrado no aplicativo, o usuário inicia o processo de cadastro através da opção “Cadastro” contida no menu da aplicação. Caso contrário, o usuário pode alterar seus dados cadastrais através da opção “Meus Dados” do menu. Independentemente do cenário a ser utilizado, o processo de registro de informações do usuário é dividido em diferentes etapas.

Em se tratando do processo de cadastro de um novo usuário, o indivíduo define suas credenciais através de seu e-mail e senha a serem utilizados para acesso ao aplicativo, conforme demonstrado na Figura 52. Esta etapa não é disponibilizada ao usuário em casos de alteração dos dados cadastrais.

Em seguida, o usuário informa seus dados pessoais básicos, sendo estes nome, sobrenome, data de nascimento, sexo, sua formação e sua faixa salarial, de acordo com o ilustrado na Figura 53a. Nesta etapa, as informações a respeito da formação acadêmica e da faixa salarial do usuário são opcionais. No próximo estágio, como evidenciado na Figura 53b, o usuário seleciona a sua profissão dentre as opções disponíveis. Em seguida, conforme apresentado na Figura 53c, o usuário seleciona, dentre as opções, o curso que esteja realizando ou já tenha realizado. Esta etapa somente será apresentada ao usuário caso sua formação acadêmica seja referente à Graduação ou Pós-Graduação.

---

<sup>1</sup>O usuário encontra-se registrado no aplicativo móvel através da realização de *login* ou da finalização do seu cadastro. Nestes cenários, um registro com as informações do usuário é inserido na base de dados local do dispositivo móvel.

Figura 51: Aplicativo móvel - Diferentes apresentações do menu ao usuário



Fonte: Elaborada pelo autor.

Figura 52: Aplicativo móvel - Credenciais do usuário



Fonte: Elaborada pelo autor.

Figura 53: Aplicativo móvel - Definição de informações básicas do usuário

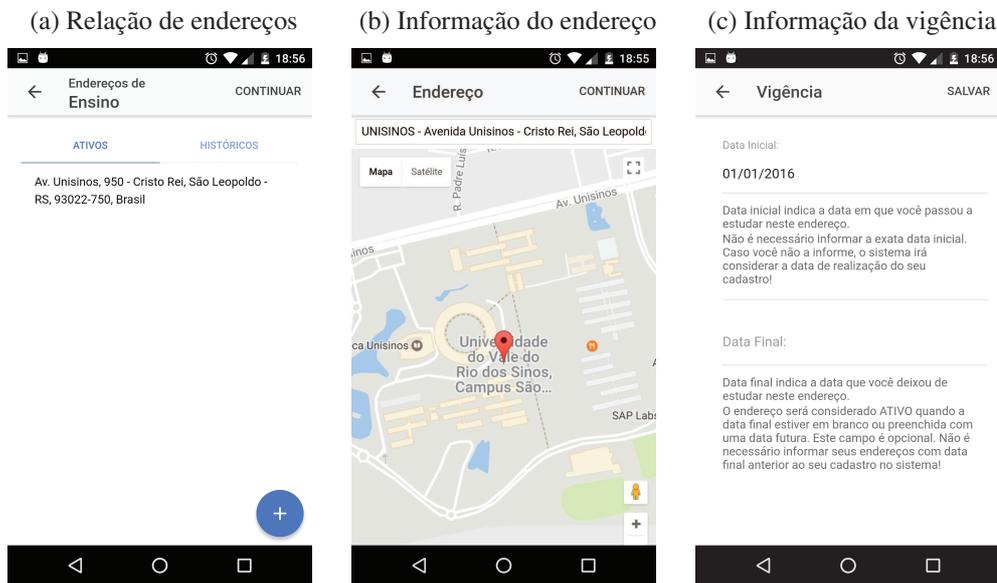


Fonte: Elaborada pelo autor.

Dando continuidade ao seu cadastro, o usuário passa a informar os endereços que fazem parte do seu cotidiano, divididos em três categorias: residenciais, de trabalho e de ensino. Estas informações são extremamente importantes pois indicam os locais onde, costumeiramente, os indivíduos passam a grande parte do seu tempo. A proposta implementada no aplicativo permite aos usuários cadastrarem mais de um endereço para cada uma das respectivas categorias. Isso deve-se ao fato de que uma mesma pessoa possa residir, trabalhar ou estudar em diferentes endereços durante determinados períodos de sua vida, garantindo assim, a consistência da análise dos dados a ser realizada posteriormente.

O processo de cadastro de um endereço é dividido em três etapas. A primeira etapa, conforme ilustrado na Figura 54a, consiste na visualização dos endereços já cadastrados, os quais são separados em endereços ativos e históricos. Esta separação mostrou-se necessária para manter a consistência dos dados durante o período de utilização do aplicativo, uma vez que o usuário pode alterar seu endereço durante este período. A partir da ação, presente na primeira etapa, de adicionar um novo endereço, o usuário inicia o segundo estágio do processo de definição de um endereço. Dessa forma, o usuário é redirecionado para uma segunda página, conforme ilustrado pela Figura 54b, onde o mesmo pesquisa pelo endereço/estabelecimento desejado. Após a seleção e representação do endereço através de um marcador no mapa, o usuário prossegue com o processo, concluindo a segunda etapa. Por fim, de acordo com o apresentado na Figura 54c, o usuário informa a vigência do endereço cadastrado. O intuito desta definição é estabelecer diferenciação entre os endereços ativos e históricos. Um endereço é considerado histórico quando sua data final estiver informada e a mesma for inferior à data atual. Este fluxo é seguido para definição das três categorias de endereços previstas na implementação do aplicativo

Figura 54: Aplicativo móvel - Definição de endereços do usuário



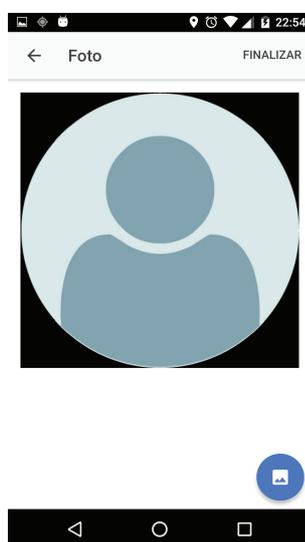
Fonte: Elaborada pelo autor.

móvel.

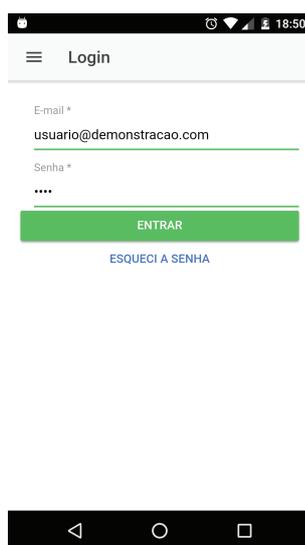
Para finalizar o seu cadastro, o usuário informa, opcionalmente, uma foto a ser associada ao seu perfil. Para tal, o usuário pode selecionar uma foto existente na galeria do seu dispositivo ou capturar uma imagem a partir da câmera, conforme demonstrado na Figura 55. Após a conclusão do cadastro do usuário, o usuário é redirecionado para a página da sua linha do tempo e o aplicativo inicia o processo de captura das informações geográficas referentes à sua localização.

- *Login*: mediante prévio cadastro realizado, o usuário pode efetuar o processo de *login* no aplicativo através da página evidenciada pela Figura 56. Após informar suas credenciais o usuário será redirecionado para a página referente à sua linha do tempo. Neste momento, o processo de monitoramento das atividades do usuário implementado no aplicativo móvel será iniciado.
- *Esqueci a senha*: caso o usuário, no decorrer da sua utilização do aplicativo, esqueça a senha definida por si no momento de seu cadastro e não esteja registrado no aplicativo, o mesmo pode utilizar a opção “Esqueci a senha”, disponível na página de “*Login*”, conforme demonstrado na Figura 56. Para receber uma nova senha, o usuário deve informar o e-mail registrado em suas credencias para qual será enviada uma nova senha, de acordo com o apresentado na Figura 57. Este processo, altera a senha do usuário para uma senha aleatória e a envia para o e-mail do usuário.
- *Página inicial*: na página inicial do aplicativo móvel são apresentados ao usuário, os resultados das apurações realizadas com base no monitoramento de suas atividade, permitindo

Figura 55: Aplicativo móvel - Escolha da foto do usuário



Fonte: Elaborada pelo autor.

Figura 56: Aplicativo móvel - Processo de *login*

Fonte: Elaborada pelo autor.

Figura 57: Aplicativo móvel - Redefinição da senha do usuário



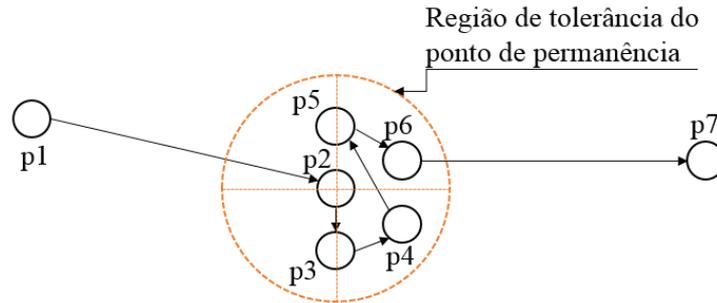
Fonte: Elaborada pelo autor.

que o indivíduo visualize e manipule tais informações. O histórico das atividades dos usuários são apresentadas de duas diferentes maneiras, através de sua linha do tempo e de representações gráficas.

1. Linha do tempo: a linha do tempo do usuário apresenta, de forma cronológica decrescente, os registros de sua atividade. Nela são apresentados os detalhes da apuração dos locais frequentados pelos indivíduos e também dos períodos em que o usuário esteve deslocando-se de um local à outro.

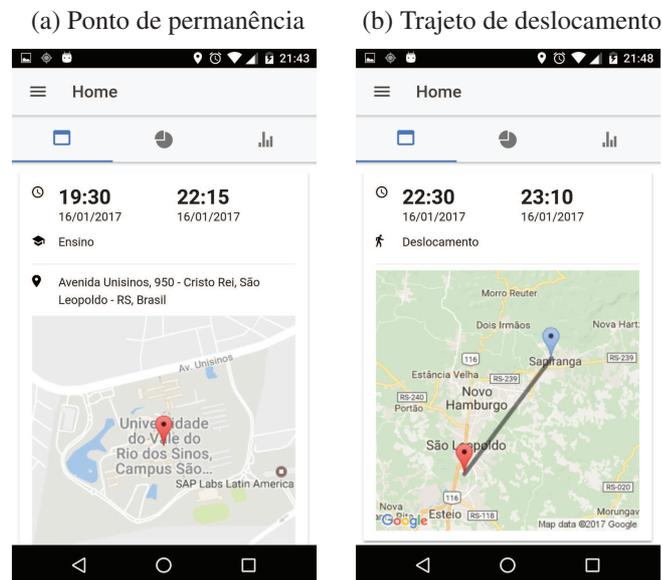
A linha do tempo categoriza os locais frequentados a partir dos registros de informações geográficas apurados através do dispositivo móvel do usuário. Tal categorização pode indicar diferentes tipos de locais. Através dos endereços cadastrados pelo usuário, o aplicativo pode identificar que o local trata-se da residência, do trabalho ou da instituição de ensino do usuário. Caso a localização apurada não se refira à um endereço cadastrado pelo usuário, o aplicativo utiliza a API Google Maps para consultar, dentro de determinada região, locais cadastrados na base de dados do Google para identificar possíveis locais frequentados pelo indivíduo. A partir de tal consulta, a aplicação define o local com a menor distância, calculada em linha reta, até o coordenada geográfica da localização apurada. Assim, o local apurado é categorizado de acordo com o tipo de local cadastrado na base de dados do Google. Além da apuração do tempo de permanência em determinados locais, o aplicativo móvel também estabelece os períodos em que o usuário esteve deslocando-se de uma localização à outra. Esta definição ocorre em virtude de o processo de apuração das localizações constatar que a distância, medida em linha reta, entre a coordenada geográfica da localização e o ponto apurado em dado instante ultrapassam a distância

Figura 58: Região de ponto de permanência



Fonte: Elaborada pelo autor.

Figura 59: Aplicativo móvel - Atividade apresentada na linha do tempo do usuário



Fonte: Elaborada pelo autor.

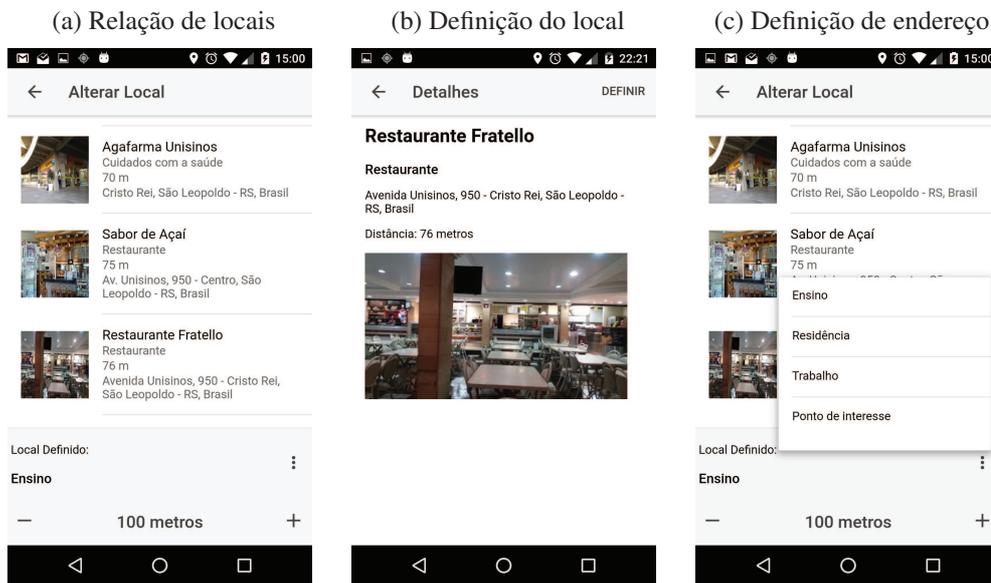
mínima configurada, conforme demonstrado na Figura 58.

Diante disto, as representações de atividade em diferentes categorias de locais são exibidas de maneiras distintas, de acordo com o observado na Figura 59a, onde é apresentada a representação de um ponto de permanência do usuário, e na Figura 59b, que ilustra o trajeto percorrido pelo usuário do ponto de partida ao ponto de chegada.

A partir da sua linha do tempo, o usuário pode alterar os dados da apuração de localizações frequentadas. Tal funcionalidade é desejável em virtude de o processo de captura das informações geográficas em que o usuário está localizado não ser um processo exato.

A alteração do local apurado é permitido apenas em apurações que identifiquem a permanência do usuário em determinados locais. Este processo permite a reclassificação desta localização, mediante dois diferentes fluxos. O primeiro fluxo consiste

Figura 60: Aplicativo móvel - Processo de alteração de local



Fonte: Elaborada pelo autor.

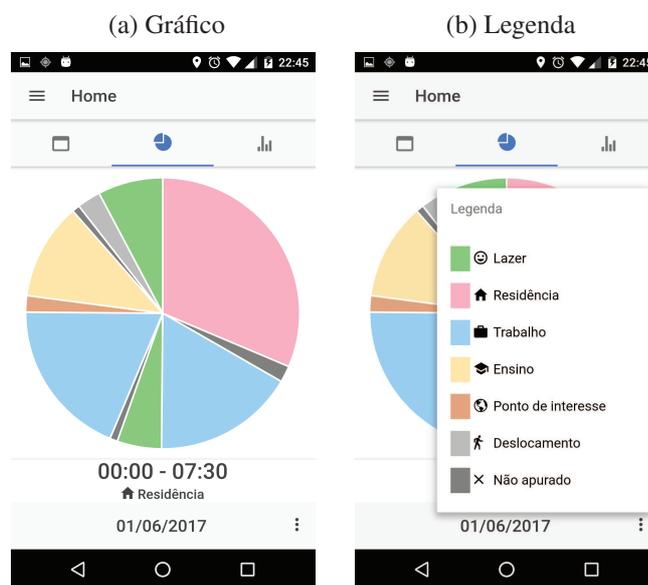
na definição de um local disponibilizado pela base de dados do Google, sendo este processo dividido em duas etapas. A primeira etapa consiste na escolha do local correto a ser definido, sendo permitida a manipulação da área de pesquisa de locais próximos, conforme demonstrado pela Figura 60a. Na segunda etapa, ao escolher um local dentre as opções, o usuário visualiza os detalhes do local escolhido, conforme apresentado na Figura 60b. Por outro lado, o segundo fluxo fundamenta-se na definição de um dos endereços do usuário cadastrados, conforme demonstrado pela Figura 60c. Em se tratando de um local não disponibilizado pela base de dados do Google e não sendo um endereço do usuário, a localização pode ser classificada como “Ponto de interesse”.

Com a redefinição do local, o tipo de localização é reclassificado conforme semântica do novo local selecionado. Assim, uma localização anteriormente classificada como ponto de interesse pode passar a representar uma localização de lazer, por exemplo. Esta redefinição é considerada no processo de apuração das localizações futuras com o intuito de estipular com maior precisão o local frequentado pelo usuário.

2. Gráficos: as representações gráficas permitem aos usuários visualizarem suas atividades durante determinados períodos de tempo, viabilizando uma melhor análise do tempo empregado nas diferentes categorias de locais determinadas pelo aplicativo. A representação visual está disponível através de dois diferentes tipos de gráficos: pizza e barras. Ambas representações gráficas permitem que o usuário navegue, através de eventos de toque, por diferentes períodos de apuração.

A visualização dos dados através do gráfico de pizza exhibe informações, de forma

Figura 61: Aplicativo móvel - Representação através de gráfico pizza



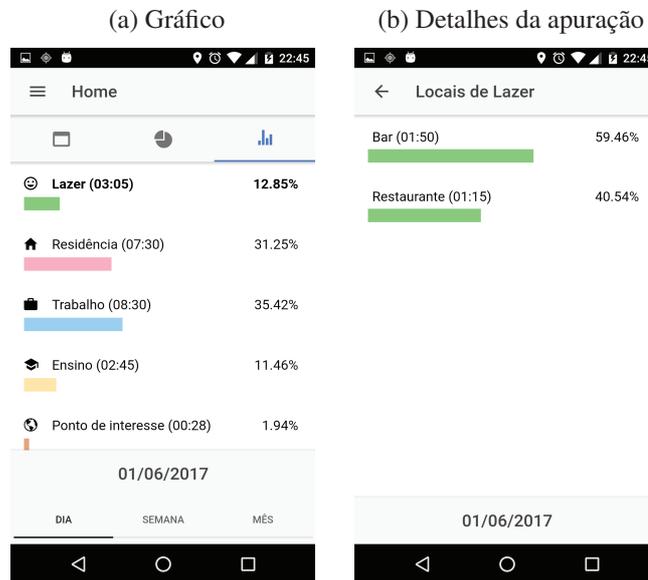
Fonte: Elaborada pelo autor.

diária, sobre a apuração das categorias de locais frequentados pelo usuário. O gráfico permite interação com usuário que, ao selecionar as diferentes fatias do gráfico, visualiza informações referentes ao intervalo de tempo selecionado, conforme demonstrado na Figura 61a. As diferentes categorias de locais que podem ser apresentadas neste gráfico são acessíveis através da funcionalidade “Legenda”, de acordo com o ilustrado pela Figura 61b.

A segunda forma de representação visual presente no aplicativo móvel refere-se à visualização através de gráficos de barras. Nesta abordagem, o aplicativo permite a visualização resumida das atividades do usuário em três diferentes períodos: diário, semanal e mensal, conforme evidenciado pela Figura 62a. Esta funcionalidade totaliza o tempo computado nas diferentes categorias de locais contidos na implementação do aplicativo móvel, onde o comprimento das barras retrata o percentual de cada uma das categorias perante o tempo total do período selecionado. A categoria “Lazer” permite que o usuário visualize detalhes da apuração realizada nesta categoria, exibindo, de forma semântica, os diferentes tipos de locais frequentados, de acordo com o demonstrado na Figura 62b.

- **Rede:** a rede do usuário consiste em apresentar, em ordem alfabética, os vínculos formados pelo usuário registrado no aplicativo móvel com os demais usuários do projeto, conforme demonstrado na Figura 63. Os vínculos entre os usuários são dinâmicos e emergem automaticamente, não necessitando de solicitação e aprovação para serem formalizados.
- **Locais:** visando disponibilizar outra forma de visualizar o histórico de locais frequentados pelo usuário, foi implementada uma funcionalidade que apresenta a listagem dos

Figura 62: Aplicativo móvel - Representação através de gráfico de barras



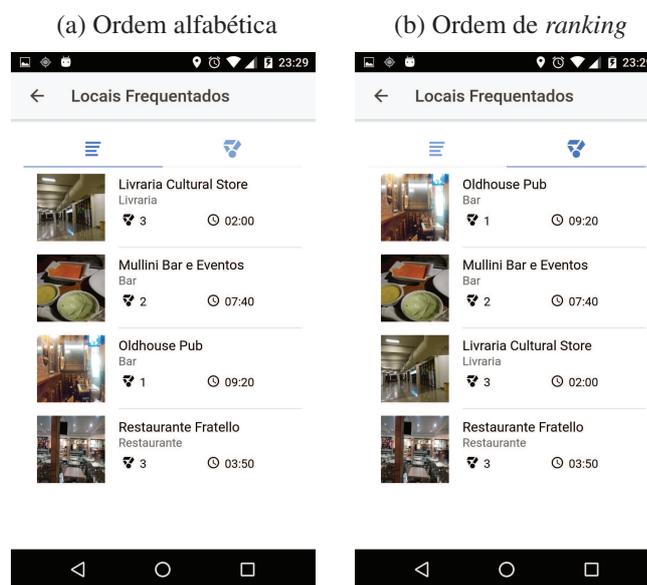
Fonte: Elaborada pelo autor.

Figura 63: Aplicativo móvel - Rede de vínculos formalizados com os demais usuários



Fonte: Elaborada pelo autor.

Figura 64: Aplicativo móvel - Relação de locais frequentados pelo usuário

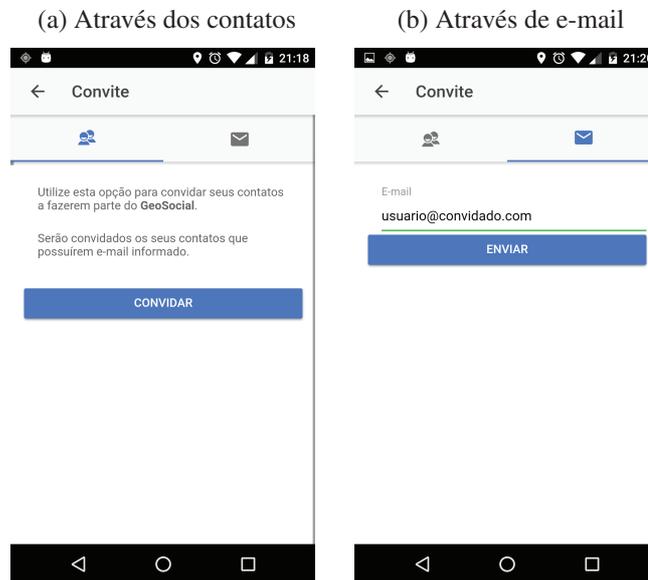


Fonte: Elaborada pelo autor.

diferentes locais que o usuário tenha frequentado. Esta visualização permite que os locais sejam visualizados de duas maneiras diferentes. A primeira abordagem consiste na visualização dos locais, classificados em ordem alfabética pelo seu nome, conforme demonstrado na Figura 64a. Por outro lado, a segunda maneira exibe os locais ordenados pelo *ranking* do usuário, ou seja, classifica a lista de locais com base na posição do usuário perante os demais, com relação ao tempo total frequentado em determinado local, conforme demonstrado na Figura 64b.

- **Convite:** com o intuito de disseminar a utilização do aplicativo móvel entre a população, foi disponibilizada a funcionalidade para que os usuários registrados no aplicativo possam convidar seus conhecidos a participar do projeto. Esta funcionalidade permite o envio de convite de duas diferentes maneiras. A primeira corresponde ao envio de forma individual, conforme demonstrado na Figura 65a, onde o usuário informa o e-mail do destinatário que irá receber o convite para participar do projeto. A segunda forma, de acordo com o demonstrado na Figura 65b, visa o envio do convite para todos os contatos telefônicos existentes no dispositivo móvel do usuário. Tal opção considera somente os contatos telefônicos que possuem e-mail cadastrado.
- **Alterar a senha:** prevendo a possibilidade de o usuário alterar sua senha, foi realizada a implementação da funcionalidade de alteração da senha do usuário no aplicativo móvel. Dessa forma, conforme demonstrado na Figura 66, é necessário que o indivíduo informe sua senha atual e também a nova senha desejada, juntamente com confirmação da nova senha para realizar a alteração.
- **Sair:** estando registrado no aplicativo, o usuário pode optar por deixar de ter sua atividade

Figura 65: Aplicativo móvel - Processo de convite de novos usuários

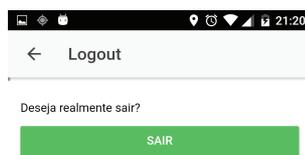


Fonte: Elaborada pelo autor.

Figura 66: Aplicativo móvel - Processo de alteração de senha do usuário



Fonte: Elaborada pelo autor.

Figura 67: Aplicativo móvel - Processo de *logout*

Fonte: Elaborada pelo autor.

monitorada pelo aplicativo, restringindo assim, sua participação no projeto. Através da opção “Sair” do aplicativo móvel, o usuário confirma sua saída, removendo seu registro do aplicativo, conforme ilustrado na Figura 67. O usuário pode, em qualquer momento, retomar o monitoramento de sua atividade no aplicativo realizando o processo de *login* com suas credenciais, conforme mencionado anteriormente.



**APÊNDICE B – SEMÂNTICAS DE LOCAIS FORNECIDOS PELA API GOOGLE MAPS**

Na Tabela 19 são apresentadas as semânticas de locais disponibilizados pela biblioteca Google Maps. Na relação, são listadas as chaves, fornecidas pela biblioteca, e as respectivas descrições das semânticas de locais disponibilizadas.

Tabela 19: Semânticas de locais fornecidos pela API Google Maps

| Chave                   | Descrição                   | Chave                  | Descrição                  |
|-------------------------|-----------------------------|------------------------|----------------------------|
| gym                     | Academia                    | lawyer                 | Advocacia                  |
| airport                 | Aeroporto                   | insurance_agency       | Agência de seguros         |
| travel_agency           | Agência de viagens          | post_office            | Agência dos correios       |
| real_estate_agency      | Alimentação                 | food                   | Agência imobiliária        |
| lodging                 | Alojamento                  | car_rental             | Aluguel de carros          |
| movie_rental            | Aluguel de filmes           | aquarium               | Aquário                    |
| campground              | Área de camping             | storage                | Armazenamento              |
| bank                    | Banco                       | bar                    | Bar                        |
| fire_station            | Bombeiros                   | hair_care              | Cabelereiro                |
| cafe                    | Café                        | atm                    | Caixa eletrônico           |
| funeral_home            | Casa funerária              | night_club             | Casa noturna               |
| casino                  | Casino                      | cemetery               | Cemitério                  |
| locksmith               | Chaveiro                    | movie_theater          | Cinema                     |
| accounting              | Contabilidade               | roofing_contractor     | Contratante de telhados    |
| health                  | Cuidados com a saúde        | dentist                | Dentista                   |
| doctor                  | Doutor                      | electrician            | Eletricista                |
| embassy                 | Embaixada                   | general_contractor     | Empreiteiro geral          |
| moving_company          | Empresa de mudanças         | plumber                | Encanamento                |
| meal_delivery           | Entrega de refeição         | school                 | Escola                     |
| local_government_office | Escritório do governo local | establishment          | Estabelecimento            |
| subway_station          | Estação de metrô            | bus_station            | Estação de ônibus          |
| train_station           | Estação de trem             | parking                | Estacionamento             |
| stadium                 | Estádio                     | pharmacy               | Farmácia                   |
| finance                 | Financeira                  | physiotherapist        | Fisioterapeuta             |
| florist                 | Florista                    | art_gallery            | Galeria de arte            |
| hospital                | Hospital                    | church                 | Igreja                     |
| car_wash                | Lavagem de carros           | laundry                | Lavanderia                 |
| library                 | Livraria                    | book_store             | Loja de livros             |
| place_of_worship        | Local de culto              | store                  | Loja                       |
| pet_store               | Loja de animais             | home_goods_store       | Loja de artigos para o lar |
| liquor_store            | Loja de bebidas alcoólicas  | bicycle_store          | Loja de bicicletas         |
| convenience_store       | Loja de conveniência        | department_store       | Loja de departamento       |
| electronics_store       | Loja de eletrônicos         | hardware_store         | Loja de ferragens          |
| jewelry_store           | Loja de jóias               | furniture_store        | Loja de móveis             |
| clothing_store          | Loja de roupas              | shoe_store             | Loja de sapatos            |
| mosque                  | Mesquita                    | museum                 | Museu                      |
| bakery                  | Padaria                     | park                   | Parque                     |
| amusement_park          | Parque de diversões         | rv_park                | Parque de estacionamento   |
| painter                 | Pintura                     | bowling_alley          | Pista de boliche           |
| police                  | Polícia                     | taxi_stand             | Ponto de taxi              |
| gas_station             | Posto de gasolina           | city_hall              | Prefeitura                 |
| meal_takeaway           | Refeição para viagem        | car_repair             | Reparo de carros           |
| restaurant              | Restaurante                 | beauty_salon           | Salão de beleza            |
| shopping_mall           | Shopping                    | synagogue              | Sinagoga                   |
| spa                     | Spa                         | grocery_or_supermarket | Supermercado               |
| hindu_temple            | Templo hindu                | courthouse             | Tribunal                   |
| university              | Universidade                | car_dealer             | Venda de carros            |
| veterinary_care         | Veterinária                 | zoo                    | Zoológico                  |

Fonte: Elaborada pelo autor.

**APÊNDICE C – ATRIBUTOS UTILIZADOS NA MODELAGEM DE PERFIL DOS USUÁRIOS**

A Tabela 20 apresenta a relação dos 81 (oitenta e um) atributos que constituem o perfil dos usuários. O tempo descrito de cada atributo é computado em minutos de acordo com o seu contexto.

Tabela 20: Relação de atributos que constituem o perfil dos usuários

| Atributo   | Descrição           | Atributo | Descrição           |
|------------|---------------------|----------|---------------------|
| total      | Total               | hr02     | Entre 02:00 e 03:00 |
| dia01      | 1º dia do mês       | hr03     | Entre 03:00 e 04:00 |
| dia02      | 2º dia do mês       | hr04     | Entre 04:00 e 05:00 |
| dia03      | 3º dia do mês       | hr05     | Entre 05:00 e 06:00 |
| dia04      | 4º dia do mês       | hr06     | Entre 06:00 e 07:00 |
| dia05      | 5º dia do mês       | hr07     | Entre 07:00 e 08:00 |
| dia06      | 6º dia do mês       | hr08     | Entre 08:00 e 09:00 |
| dia07      | 7º dia do mês       | hr09     | Entre 09:00 e 10:00 |
| dia08      | 8º dia do mês       | hr10     | Entre 10:00 e 11:00 |
| dia09      | 9º dia do mês       | hr11     | Entre 11:00 e 12:00 |
| dia10      | 10º dia do mês      | hr12     | Entre 12:00 e 13:00 |
| dia11      | 11º dia do mês      | hr13     | Entre 13:00 e 14:00 |
| dia12      | 12º dia do mês      | hr14     | Entre 14:00 e 15:00 |
| dia13      | 13º dia do mês      | hr15     | Entre 15:00 e 16:00 |
| dia14      | 14º dia do mês      | hr16     | Entre 16:00 e 17:00 |
| dia15      | 15º dia do mês      | hr17     | Entre 17:00 e 18:00 |
| dia16      | 16º dia do mês      | hr18     | Entre 18:00 e 19:00 |
| dia17      | 17º dia do mês      | hr19     | Entre 19:00 e 20:00 |
| dia18      | 18º dia do mês      | hr20     | Entre 20:00 e 21:00 |
| dia19      | 19º dia do mês      | hr21     | Entre 21:00 e 22:00 |
| dia20      | 20º dia do mês      | hr22     | Entre 22:00 e 23:00 |
| dia21      | 21º dia do mês      | hr23     | Entre 23:00 e 00:00 |
| dia22      | 22º dia do mês      | periodo1 | Entre 00:00 e 06:00 |
| dia23      | 23º dia do mês      | periodo2 | Entre 06:00 e 08:00 |
| dia24      | 24º dia do mês      | periodo3 | Entre 08:00 e 12:00 |
| dia25      | 25º dia do mês      | periodo4 | Entre 12:00 e 13:00 |
| dia26      | 26º dia do mês      | periodo5 | Entre 13:00 e 18:00 |
| dia27      | 27º dia do mês      | periodo6 | Entre 18:00 e 00:00 |
| dia28      | 28º dia do mês      | mes01    | Janeiro             |
| dia29      | 29º dia do mês      | mes02    | Fevereiro           |
| dia30      | 30º dia do mês      | mes03    | Março               |
| dia31      | 31º dia do mês      | mes04    | Abril               |
| diaSemana1 | Segunda-feira       | mes05    | Maio                |
| diaSemana2 | Terça-feira         | mes06    | Junho               |
| diaSemana3 | Quarta-feira        | mes07    | Julho               |
| diaSemana4 | Quinta-feira        | mes08    | Agosto              |
| diaSemana5 | Sexta-feira         | mes09    | Setembro            |
| diaSemana6 | Sábado              | mes10    | Outubro             |
| diaSemana7 | Domingo             | mes11    | Novembro            |
| hr00       | Entre 00:00 e 01:00 | mes12    | Dezembro            |
| hr01       | Entre 01:00 e 02:00 |          |                     |

Fonte: Elaborada pelo autor.

## APÊNDICE D – CENÁRIOS PREVIAMENTE AVALIADOS

O principal objetivo dos cenários previamente analisados foi o balizamento do procedimento de clusterização proposto neste trabalho, mediante avaliação de diferentes cenários. Dessa forma, utilizando os dados artificiais gerados, são apresentados na Tabela 21, os resultados obtidos através do processo de clusterização, onde foram considerados os seguintes aspectos:

1. População: indicação da população utilizada no cenário de avaliação.
2. Usuários: quantidade de usuários gerados para a realização da análise do processo de clusterização.
3. Preferência: métrica definida como valor de preferência, utilizada para compor a matriz de similaridades do processo de clusterização. Tal métrica é característica fundamental do procedimento de clusterização através da técnica *Affinity Propagation*.
4. *Clusters*: quantidade total de *clusters* gerados a partir do processo de clusterização.
5. *Clusters* com 1 usuário: quantidade de *clusters* gerados a partir do processo de clusterização, onde, ao final do processo, somente um usuário esteve alocado no respectivo *cluster*.
6. Característica: características das atividades dos usuários gerados para a realização do procedimento de clusterização avaliado. Além das características dos locais frequentados, também pode-se avaliar o procedimento de clusterização considerando um subconjunto de atributos do perfil dos usuários para compor o procedimento de clusterização.

Tabela 21: Cenários previamente avaliados

| População | Usuários | Preferência | Clusters | Clusters com 1 usuário | Características   |
|-----------|----------|-------------|----------|------------------------|---|
| 1         | 500      | Mediana     | 19       | 0                      | Todas as atividades   |
| 1         | 500      | Mediana     | 46       | 12                     | Atividades de lazer   |
| 1         | 500      | Mediana     | 19       | 0                      | Atividades de trabalho, residência e ensino                 |
| 1         | 500      | Mediana     | 32       | 0                      | Atividades de lazer   |
| 1         | 500      | Mediana     | 68       | 15                     | Atividades de ensino e bar                                  |
| 1         | 500      | Mediana     | 60       | 26                     | Atividades em restaurantes aos finais de semana             |
| 1         | 500      | Mediana     | 44       | 4                      | Atividades de lazer (clusterização por dias da semana)      |
| 1         | 500      | 3º Quartil  | 86       | 46                     | Atividades de lazer   |
| 1         | 500      | 1º Quartil  | 26       | 2                      | Atividades de lazer   |
| 1         | 500      | Mediana     | 48       | 0                      | Atividades em academias (clusterização por períodos do dia) |
| 1         | 500      | 3º Quartil  | 45       | 7                      | Todas as atividades   |
| 1         | 500      | 1º Quartil  | 18       | 0                      | Todas as atividades   |
| 2         | 500      | Mediana     | 25       | 0                      | Todas as atividades   |
| 2         | 500      | Mediana     | 38       | 7                      | Atividades de lazer   |
| 2         | 500      | Mediana     | 23       | 0                      | Atividades de trabalho, residência e ensino                 |
| 2         | 500      | Mediana     | 36       | 0                      | Atividades de lazer   |
| 2         | 500      | Mediana     | 67       | 18                     | Atividades de ensino e bar                                  |
| 2         | 500      | Mediana     | 58       | 22                     | Atividades em restaurantes aos finais de semana             |
| 2         | 500      | Mediana     | 40       | 0                      | Atividades de lazer (clusterização por dias da semana)      |
| 2         | 500      | 3º Quartil  | 88       | 45                     | Atividades de lazer   |
| 2         | 500      | 1º Quartil  | 25       | 3                      | Atividades de lazer   |
| 2         | 500      | Mediana     | 46       | 0                      | Atividades em academias (clusterização por períodos do dia) |
| 2         | 500      | 3º Quartil  | 39       | 4                      | Todas as atividades   |
| 2         | 500      | 1º Quartil  | 17       | 0                      | Todas as atividades   |
| 3         | 500      | Mediana     | 23       | 0                      | Todas as atividades   |
| 3         | 500      | Mediana     | 43       | 10                     | Atividades de lazer   |
| 3         | 500      | Mediana     | 22       | 0                      | Atividades de trabalho, residência e ensino                 |
| 3         | 500      | Mediana     | 28       | 0                      | Atividades de lazer   |
| 3         | 500      | Mediana     | 64       | 19                     | Atividades de ensino e bar                                  |
| 3         | 500      | Mediana     | 64       | 30                     | Atividades em restaurantes aos finais de semana             |
| 3         | 500      | Mediana     | 41       | 1                      | Atividades de lazer (clusterização por dias da semana)      |
| 3         | 500      | 3º Quartil  | 95       | 56                     | Atividades de lazer   |
| 3         | 500      | 1º Quartil  | 23       | 0                      | Atividades de lazer   |
| 3         | 500      | Mediana     | 45       | 0                      | Atividades em academias (clusterização por períodos do dia) |
| 3         | 500      | 3º Quartil  | 42       | 5                      | Todas as atividades   |
| 3         | 500      | 1º Quartil  | 15       | 0                      | Todas as atividades   |

## REFERÊNCIAS

- AGGARWAL, C. C. An introduction to social network data analytics. **Social network data analytics**, [S.l.], p. 1–15, 2011.
- ALJAWARNEH, M.; DHOMEJA, L. D.; MALKANI, Y. A. Context-aware service composition of heterogeneous services in pervasive computing environments: a review. In: MULTI-TOPIC CONFERENCE (INMIC), 2016 19TH INTERNATIONAL, 2016. **Anais...** [S.l.: s.n.], 2016. p. 1–6.
- ALONSO, K.; ESPINOZA-MOLINA, D.; DATCU, M. Multilayer architecture for heterogeneous geospatial data analytics: querying and understanding eo archives. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, [S.l.], v. 10, n. 3, p. 791–801, 2017.
- ANDRIENKO, G.; ANDRIENKO, N. Constructing parallel coordinates plot for problem solving. In: INTERNATIONAL SYMPOSIUM ON SMART GRAPHICS, 1., 2001. **Anais...** [S.l.: s.n.], 2001. p. 9–14.
- ANDRIENKO, G.; ANDRIENKO, N. Visual exploration of the spatial distribution of temporal behaviors. In: INFORMATION VISUALISATION, 2005. PROCEEDINGS. NINTH INTERNATIONAL CONFERENCE ON, 2005. **Anais...** [S.l.: s.n.], 2005. p. 799–806.
- ANDRIENKO, G.; ANDRIENKO, N.; WROBEL, S. Visual analytics tools for analysis of movement data. **ACM SIGKDD Explorations Newsletter**, [S.l.], v. 9, n. 2, p. 38–46, 2007.
- ANDRIENKO, G. et al. Space, time and visual analytics. **International Journal of Geographical Information Science**, [S.l.], v. 24, n. 10, p. 1577–1600, 2010.
- ANDRIENKO, G. et al. Visual analytics of mobility and transportation: state of the art and further research directions. **IEEE Transactions on Intelligent Transportation Systems**, [S.l.], 2017.
- ARORA, S.; CHANA, I. A survey of clustering techniques for big data analysis. In: CONFLUENCE THE NEXT GENERATION INFORMATION TECHNOLOGY SUMMIT (CONFLUENCE), 2014 5TH INTERNATIONAL CONFERENCE-, 2014. **Anais...** [S.l.: s.n.], 2014. p. 59–65.
- ARTERO, A. O.; OLIVEIRA, M. C. F. de; LEVKOWITZ, H. Uncovering clusters in crowded parallel coordinates visualizations. In: INFORMATION VISUALIZATION, 2004. INFOVIS 2004. IEEE SYMPOSIUM ON, 2004. **Anais...** [S.l.: s.n.], 2004. p. 81–88.
- ASMI, K.; LOTFI, D.; EL MARRAKI, M. An approach based on the clustering coefficient for the community detection in social networks. In: SECURITY OF SMART CITIES, INDUSTRIAL CONTROL SYSTEM AND COMMUNICATIONS (SSIC), 2016 INTERNATIONAL CONFERENCE ON, 2016. **Anais...** [S.l.: s.n.], 2016. p. 1–8.
- BACIU, G.; WANG, Y.; LI, C. Cognitive visual analytics of multi-dimensional cloud system monitoring data. In: COGNITIVE INFORMATICS & COGNITIVE COMPUTING (ICCI\* CC), 2016 IEEE 15TH INTERNATIONAL CONFERENCE ON, 2016. **Anais...** [S.l.: s.n.], 2016. p. 310–315.

- BARBOSA, J. L. V. Ubiquitous computing: applications and research opportunities. In: COMPUTATIONAL INTELLIGENCE AND COMPUTING RESEARCH (ICCIC), 2015 IEEE INTERNATIONAL CONFERENCE ON, 2015. **Anais...** [S.l.: s.n.], 2015. p. 1–8.
- BARBOSA, J. L. V. et al. Trailtrade: a model for trail-aware commerce support. **Computers in Industry**, [S.l.], v. 80, p. 43–53, 2016.
- BARLOWE, S. et al. Multivariate visual explanation for high dimensional datasets. In: VISUAL ANALYTICS SCIENCE AND TECHNOLOGY, 2008. VAST'08. IEEE SYMPOSIUM ON, 2008. **Anais...** [S.l.: s.n.], 2008. p. 147–154.
- BERKHIN, P. A survey of clustering data mining techniques. In: **Grouping multidimensional data**. [S.l.]: Springer, 2006. p. 25–71.
- BERTHOLD, M. R.; HALL, L. O. Visualizing fuzzy points in parallel coordinates. **IEEE Transactions on Fuzzy Systems**, [S.l.], v. 11, n. 3, p. 369–374, 2003.
- BISTA, S. K.; NEPAL, S.; PARIS, C. Multifaceted visualisation of annotated social media data. In: BIG DATA (BIGDATA CONGRESS), 2014 IEEE INTERNATIONAL CONGRESS ON, 2014. **Anais...** [S.l.: s.n.], 2014. p. 699–706.
- BORGATTI, S. P. et al. Network analysis in the social sciences. **science**, [S.l.], v. 323, n. 5916, p. 892–895, 2009.
- BOUHALI, S.; ELLOUZE, M. Community detection in social network: literature review and research perspectives. In: SERVICE OPERATIONS AND LOGISTICS, AND INFORMATICS (SOLI), 2015 IEEE INTERNATIONAL CONFERENCE ON, 2015. **Anais...** [S.l.: s.n.], 2015. p. 139–144.
- BOUJLALEB, L. et al. User-centric approach of detecting temporary community. In: COMPLEX SYSTEMS (WCCS), 2015 THIRD WORLD CONFERENCE ON, 2015. **Anais...** [S.l.: s.n.], 2015. p. 1–6.
- CHANG, W.-L.; LIN, T.-H. A cluster-based approach for automatic social network construction. In: SOCIAL COMPUTING (SOCIALCOM), 2010 IEEE SECOND INTERNATIONAL CONFERENCE ON, 2010. **Anais...** [S.l.: s.n.], 2010. p. 601–606.
- CHENG, S. et al. Model-driven visual analytics for big data. In: SCIENTIFIC DATA SUMMIT (NYSDES), 2016 NEW YORK, 2016. **Anais...** [S.l.: s.n.], 2016. p. 1–2.
- CHO, E.; MYERS, S. A.; LESKOVEC, J. Friendship and mobility: user movement in location-based social networks. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 17., 2011, New York, NY, USA. **Proceedings...** ACM, 2011. p. 1082–1090. (KDD '11).
- COMPIETA, P. et al. Exploratory spatio-temporal data mining and visualization. **Journal of Visual Languages & Computing**, [S.l.], v. 18, n. 3, p. 255–279, 2007.
- COOK, K. A.; THOMAS, J. J. **Illuminating the path**: the research and development agenda for visual analytics. [S.l.]: Pacific Northwest National Laboratory (PNNL), Richland, WA (US), 2005.
- CURRAN, K.; CURRAN, N. Social networking analysis. In: **Big data and internet of things**: a roadmap for smart environments. [S.l.]: Springer, 2014. p. 367–378.

- CUZZOCREA, A.; ZALL, D. Parallel coordinates technique in visual data mining: advantages, disadvantages and combinations. In: INFORMATION VISUALISATION (IV), 2013 17TH INTERNATIONAL CONFERENCE, 2013. **Anais...** [S.l.: s.n.], 2013. p. 278–284.
- D'OCAGNE, M. **Coordonnées parallèles & axiales**: méthode de transformation géométrique et procédé nouveau de calcul graphique déduits de la considération des coordonnées parallèles. [S.l.]: Gauthier-Villars, 1885.
- DUECK, D.; FREY, B. J. Non-metric affinity propagation for unsupervised image categorization. In: COMPUTER VISION, 2007. ICCV 2007. IEEE 11TH INTERNATIONAL CONFERENCE ON, 2007. **Anais...** [S.l.: s.n.], 2007. p. 1–8.
- EAGLE, N.; PENTLAND, A. S. Reality mining: sensing complex social systems. **Personal and ubiquitous computing**, [S.l.], v. 10, n. 4, p. 255–268, 2006.
- EAGLE, N.; PENTLAND, A. S. Eigenbehaviors: identifying structure in routine. **Behavioral Ecology and Sociobiology**, [S.l.], v. 63, n. 7, p. 1057–1066, 2009.
- ELBARAWY, Y. M.; MOHAMED, R. F.; GHALI, N. I. Improving social network community detection using dbscan algorithm. In: COMPUTER APPLICATIONS & RESEARCH (WSCAR), 2014 WORLD SYMPOSIUM ON, 2014. **Anais...** [S.l.: s.n.], 2014. p. 1–6.
- FRANCO, L. K. et al. Mucs: a model for ubiquitous commerce support. **Electronic Commerce Research and Applications**, [S.l.], v. 10, n. 2, p. 237 – 246, 2011. Special Issue on Electronic Auctions: Strategies and Methods.
- FREY, B. J.; DUECK, D. Clustering by passing messages between data points. **science**, [S.l.], v. 315, n. 5814, p. 972–976, 2007.
- GANTZ, J.; REINSEL, D. Extracting value from chaos. **IDC iview**, [S.l.], v. 1142, n. 2011, p. 1–12, 2011.
- GONZALEZ, M. C.; HIDALGO, C. A.; BARABASI, A.-L. Understanding individual human mobility patterns. **Nature**, [S.l.], v. 453, n. 7196, p. 779–782, 2008.
- GUAN, R. et al. Text clustering with seeds affinity propagation. **IEEE Transactions on Knowledge and Data Engineering**, [S.l.], v. 23, n. 4, p. 627–637, 2011.
- GUESMI, S.; TRABELSI, C.; LATIRI, C. Fca for common interest communities discovering. In: DATA SCIENCE AND ADVANCED ANALYTICS (DSAA), 2014 INTERNATIONAL CONFERENCE ON, 2014. **Anais...** [S.l.: s.n.], 2014. p. 449–455.
- HAJIBAGHERI, A. et al. Social networks community detection using the shapley value. In: ARTIFICIAL INTELLIGENCE AND SIGNAL PROCESSING (AISP), 2012 16TH CSI INTERNATIONAL SYMPOSIUM ON, 2012. **Anais...** [S.l.: s.n.], 2012. p. 222–227.
- HAO, F. et al.  $k$ -clique community detection in social networks based on formal concept analysis. **IEEE Systems Journal**, [S.l.], 2015.
- HEER, J.; PERER, A. Orion: a system for modeling, transformation and visualization of multidimensional heterogeneous networks. **Information Visualization**, [S.l.], v. 13, n. 2, p. 111–133, 2014.

HEINRICH, J.; WEISKOPF, D. State of the art of parallel coordinates. In: EUROGRAPHICS (STARS), 2013. **Anais...** [S.l.: s.n.], 2013. p. 95–116.

HEINRICH, J.; WEISKOPF, D. Parallel coordinates for multidimensional data visualization: basic concepts. **Computing in Science & Engineering**, [S.l.], v. 17, n. 3, p. 70–76, 2015.

HONG, H. et al. Predicting potential gene ontology from cellular response data. In: INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND COMPUTATIONAL BIOLOGY, 5., 2017. **Proceedings...** [S.l.: s.n.], 2017. p. 5–10.

HUANG, J.; WEI, Y.; DU, X. Pcnnet: exploratory visual analysis of large-scale network. In: BIG DATA ANALYSIS (ICBDA), 2016 IEEE INTERNATIONAL CONFERENCE ON, 2016. **Anais...** [S.l.: s.n.], 2016. p. 1–6.

HUANG, Y. et al. An effective community detection algorithm of the social networks. In: INFORMATION SCIENCE AND TECHNOLOGY (ICIST), 2013 INTERNATIONAL CONFERENCE ON, 2013. **Anais...** [S.l.: s.n.], 2013. p. 824–827.

IBM. **Bringing big data to the enterprise: what is big data?** 2017.

INSELBERG, A. The plane with parallel coordinates. **The visual computer**, [S.l.], v. 1, n. 2, p. 69–91, 1985.

JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. **ACM computing surveys (CSUR)**, [S.l.], v. 31, n. 3, p. 264–323, 1999.

JAMALI, M.; ABOLHASSANI, H. Different aspects of social network analysis. In: WEB INTELLIGENCE, 2006. WI 2006. IEEE/WIC/ACM INTERNATIONAL CONFERENCE ON, 2006. **Anais...** [S.l.: s.n.], 2006. p. 66–72.

JERN, M.; FRANZEN, J. "geoanalytics-exploring spatio-temporal and multivariate data. In: INFORMATION VISUALIZATION, 2006. IV 2006. TENTH INTERNATIONAL CONFERENCE ON, 2006. **Anais...** [S.l.: s.n.], 2006. p. 25–31.

JIA, Y. et al. Finding image exemplars using fast sparse affinity propagation. In: ACM INTERNATIONAL CONFERENCE ON MULTIMEDIA, 16., 2008. **Proceedings...** [S.l.: s.n.], 2008. p. 639–642.

JUNIOR, F. G.; BARBOSA, J. L. V. U-deal: a decentralized trail-aware model for business opportunities identification. **IEEE Latin America Transactions**, [S.l.], v. 13, n. 5, p. 1640–1648, 2015.

KANG, C. et al. Analyzing and geo-visualizing individual human mobility patterns using mobile call records. In: GEOINFORMATICS, 2010 18TH INTERNATIONAL CONFERENCE ON, 2010. **Anais...** [S.l.: s.n.], 2010. p. 1–7.

KAPLER, T.; WRIGHT, W. Geotime information visualization. **Information Visualization**, [S.l.], v. 4, n. 2, p. 136–146, 2005.

KEIM, D. A. et al. Challenges in visual data analysis. In: INFORMATION VISUALIZATION, 2006. IV 2006. TENTH INTERNATIONAL CONFERENCE ON, 2006. **Anais...** [S.l.: s.n.], 2006. p. 9–16.

- KEIM, D. et al. Visual analytics: definition, process, and challenges. In: **Information visualization**. [S.l.]: Springer, 2008. p. 154–175.
- KEIM, D. et al. Visual analytics: scope and challenges. **Visual data mining**, [S.l.], p. 76–90, 2008.
- KEIM, D. et al. **Mastering the information age solving problems with visual analytics**. [S.l.]: Eurographics Association, 2010.
- KLONTZ, J. C.; JAIN, A. K. A case study of automated face recognition: the boston marathon bombings suspects. **Computer**, [S.l.], v. 46, n. 11, p. 91–94, Nov 2013.
- KRISHNAN, S. G.; KARTHIKA, S.; BOSE, S. Detection of communities in dynamic social networks. In: RECENT TRENDS IN INFORMATION TECHNOLOGY (ICRTIT), 2016 INTERNATIONAL CONFERENCE ON, 2016. **Anais...** [S.l.: s.n.], 2016. p. 1–6.
- KUMAR, G.; GARLAND, M. Visual exploration of complex time-varying graphs. **IEEE Transactions on Visualization and Computer Graphics**, [S.l.], v. 12, n. 5, p. 805–812, 2006.
- LAASONEN, K. Clustering and prediction of mobile user routes from cellular data. In: EUROPEAN CONFERENCE ON PRINCIPLES OF DATA MINING AND KNOWLEDGE DISCOVERY, 2005. **Anais...** [S.l.: s.n.], 2005. p. 569–576.
- LI, Q. et al. Mining user similarity based on location history. In: ACM SIGSPATIAL INTERNATIONAL CONFERENCE ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 16., 2008. **Proceedings...** [S.l.: s.n.], 2008. p. 34.
- LI, X. et al. Community detection in large social networks based on relationship density. In: COMPUTER SOFTWARE AND APPLICATIONS CONFERENCE (COMPSAC), 2016 IEEE 40TH ANNUAL, 2016. **Anais...** [S.l.: s.n.], 2016. v. 1, p. 524–533.
- LIU, C.; LIU, J.; JIANG, Z. A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks. **IEEE transactions on cybernetics**, [S.l.], v. 44, n. 12, p. 2274–2287, 2014.
- LIU, J. et al. Community detection in location-based social networks: an entropy-based approach. In: COMPUTER AND INFORMATION TECHNOLOGY (CIT), 2016 IEEE INTERNATIONAL CONFERENCE ON, 2016. **Anais...** [S.l.: s.n.], 2016. p. 452–459.
- LIU, Y.; MAGNO, C. A case study to apply mobile technology into individual's local community. In: BIG DATA (BIG DATA), 2015 IEEE INTERNATIONAL CONFERENCE ON, 2015. **Anais...** [S.l.: s.n.], 2015. p. 2910–2912.
- LU, M. et al. A visual analytics approach for exploring individual behaviors in smartphone usage data. In: PACIFIC VISUALIZATION SYMPOSIUM (PACIFICVIS), 2016 IEEE, 2016. **Anais...** [S.l.: s.n.], 2016. p. 224–228.
- LV, P.; ZHANG, J.; ZHANG, H. A social network graphics segmentation algorithm based on community-detection. In: COMPUTER AND COMMUNICATIONS (ICCC), 2016 2ND IEEE INTERNATIONAL CONFERENCE ON, 2016. **Anais...** [S.l.: s.n.], 2016. p. 619–623.
- MENG, J. et al. Parallel gene selection and dynamic ensemble pruning based on affinity propagation. **Computers in Biology and Medicine**, [S.l.], 2017.

MITCHELL, J. C. **Social networks in urban situations**: analyses of personal relationships in central african towns. [S.l.]: Manchester University Press, 1969.

NAVARRO, N. d. A. B. et al. A context-aware spontaneous mobile social network. In: UBIQUITOUS INTELLIGENCE AND COMPUTING AND 2015 IEEE 12TH INTL CONF ON AUTONOMIC AND TRUSTED COMPUTING AND 2015 IEEE 15TH INTL CONF ON SCALABLE COMPUTING AND COMMUNICATIONS AND ITS ASSOCIATED WORKSHOPS (UIC-ATC-SCALCOM), 2015 IEEE 12TH INTL CONF ON, 2015. **Anais...** [S.l.: s.n.], 2015. p. 85–92.

NEWMAN, M. E. Modularity and community structure in networks. **Proceedings of the national academy of sciences**, [S.l.], v. 103, n. 23, p. 8577–8582, 2006.

OLIVEIRA, R. R. et al. Swtrack: an intelligent model for cargo tracking based on off-the-shelf mobile devices. **Expert Systems with Applications**, [S.l.], v. 40, n. 6, p. 2023 – 2031, 2013.

PALMA, A. T. et al. A clustering-based approach for discovering interesting places in trajectories. In: ACM SYMPOSIUM ON APPLIED COMPUTING, 2008., 2008. **Proceedings...** [S.l.: s.n.], 2008. p. 863–868.

PAPADOBKONAS, D.; ROUSSOS, G.; LEVENE, M. Analysis, ranking and prediction in pervasive computing trails. In: IET 4TH INTERNATIONAL CONFERENCE ON INTELLIGENT ENVIRONMENTS, 2008., 2008. **Anais...** [S.l.: s.n.], 2008. p. 1–8.

RUBEL, O. et al. High performance multivariate visual data exploration for extremely large data. In: HIGH PERFORMANCE COMPUTING, NETWORKING, STORAGE AND ANALYSIS, 2008. SC 2008. INTERNATIONAL CONFERENCE FOR, 2008. **Anais...** [S.l.: s.n.], 2008. p. 1–12.

SARMA, A. C.; GIRÃO, J. Identities in the future internet of things. **Wireless personal communications**, [S.l.], v. 49, n. 3, p. 353–363, 2009.

SATYANARAYANAN, M. Pervasive computing: vision and challenges. **IEEE Personal Communications**, [S.l.], v. 8, n. 4, p. 10–17, Aug 2001.

SHAHWANI, H. et al. A stable clustering algorithm based on affinity propagation for vanets. In: ADVANCED COMMUNICATION TECHNOLOGY (ICACT), 2017 19TH INTERNATIONAL CONFERENCE ON, 2017. **Anais...** [S.l.: s.n.], 2017. p. 501–504.

SHANG, F. et al. Fast affinity propagation clustering: a multilevel approach. **Pattern recognition**, [S.l.], v. 45, n. 1, p. 474–486, 2012.

SHEA, C.; HASSANABADI, B.; VALAEE, S. Mobility-based clustering in vanets using affinity propagation. In: GLOBAL TELECOMMUNICATIONS CONFERENCE, 2009. GLOBECOM 2009. IEEE, 2009. **Anais...** [S.l.: s.n.], 2009. p. 1–6.

SHEN, Z.; MA, K. L. Mobivis: a visualization system for exploring mobile data. In: IEEE PACIFIC VISUALIZATION SYMPOSIUM, 2008., 2008. **Anais...** [S.l.: s.n.], 2008. p. 175–182.

SHNEIDERMAN, B. The eyes have it: a task by data type taxonomy for information visualizations. In: VISUAL LANGUAGES, 1996. PROCEEDINGS., IEEE SYMPOSIUM ON, 1996. **Anais...** [S.l.: s.n.], 1996. p. 336–343.

- SIIRTOLA, H. Direct manipulation of parallel coordinates. In: INFORMATION VISUALIZATION, 2000. PROCEEDINGS. IEEE INTERNATIONAL CONFERENCE ON, 2000. **Anais...** [S.l.: s.n.], 2000. p. 373–378.
- SIIRTOLA, H. Combining parallel coordinates with the reorderable matrix. In: COORDINATED AND MULTIPLE VIEWS IN EXPLORATORY VISUALIZATION, 2003. PROCEEDINGS. INTERNATIONAL CONFERENCE ON, 2003. **Anais...** [S.l.: s.n.], 2003. p. 63–74.
- SOWMYA, R.; SUNEETHA, K. Data mining with big data. In: INTELLIGENT SYSTEMS AND CONTROL (ISCO), 2017 11TH INTERNATIONAL CONFERENCE ON, 2017. **Anais...** [S.l.: s.n.], 2017. p. 246–250.
- STEED, C. A. et al. Big data visual analytics for exploratory earth system simulation analysis. **Computers & Geosciences**, [S.l.], v. 61, p. 71–82, 2013.
- TANTIPATHANANANDH, C.; BERGER-WOLF, T. Y. Finding communities in dynamic social networks. In: DATA MINING (ICDM), 2011 IEEE 11TH INTERNATIONAL CONFERENCE ON, 2011. **Anais...** [S.l.: s.n.], 2011. p. 1236–1241.
- VADISALA, J.; VATSAVAYI, V. K. Challenges in social network data privacy. **International Journal of Computational Intelligence Research**, [S.l.], v. 13, n. 5, p. 965–979, 2017.
- VATIN, G.; NAPOLI, A. Guiding the controller in geovisual analytics to improve maritime surveillance. In: GEOPROCESSING 2013: THE FIFTH INTERNATIONAL CONFERENCE ON ADVANCED GEOGRAPHIC INFORMATION SYSTEMS, APPLICATIONS, AND SERVICES, 2013. **Anais...** [S.l.: s.n.], 2013. p. 26–31.
- WANG, J. et al. Multi-resolution climate ensemble parameter analysis with nested parallel coordinates plots. **IEEE Transactions on Visualization and Computer Graphics**, [S.l.], v. 23, n. 1, p. 81–90, 2017.
- WANG, W. B. et al. Improving performance of forensics investigation with parallel coordinates visual analytics. In: COMPUTATIONAL SCIENCE AND ENGINEERING (CSE), 2014 IEEE 17TH INTERNATIONAL CONFERENCE ON, 2014. **Anais...** [S.l.: s.n.], 2014. p. 1838–1843.
- WASSERMAN, S.; FAUST, K. **Social network analysis: methods and applications**. [S.l.]: Cambridge university press, 1994. v. 8.
- WEGMAN, E. J. Hyperdimensional data analysis using parallel coordinates. **Journal of the American Statistical Association**, [S.l.], v. 85, n. 411, p. 664–675, 1990.
- WEISER, M. The computer for the 21st century. **IEEE pervasive computing**, [S.l.], v. 1, n. 1, p. 19–25, 2002.
- WU, M. et al. Friend recommendation algorithm for online social networks based on location preference. In: INFORMATION SCIENCE AND CONTROL ENGINEERING (ICISCE), 2016 3RD INTERNATIONAL CONFERENCE ON, 2016. **Anais...** [S.l.: s.n.], 2016. p. 379–385.
- WU, Y. et al. A survey on visual analytics of social media data. **IEEE Transactions on Multimedia**, [S.l.], v. 18, n. 11, p. 2135–2148, 2016.

XIA, D.-y. et al. Local and global approaches of affinity propagation clustering for large scale data. **Journal of Zhejiang University-Science A**, [S.l.], v. 9, n. 10, p. 1373–1381, 2008.

XIAO, J. et al. Joint affinity propagation for multiple view segmentation. In: **COMPUTER VISION, 2007. ICCV 2007. IEEE 11TH INTERNATIONAL CONFERENCE ON**, 2007. **Anais...** [S.l.: s.n.], 2007. p. 1–7.

XIAO, X. et al. Finding similar users using category-based location history. In: **SIGSPATIAL INTERNATIONAL CONFERENCE ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS**, 18., 2010, New York, NY, USA. **Proceedings...** ACM, 2010. p. 442–445. (GIS '10).

XU, R.; WUNSCH, D. Survey of clustering algorithms. **IEEE Transactions on neural networks**, [S.l.], v. 16, n. 3, p. 645–678, 2005.

YU, L. et al. iviztrans: interactive visual learning for home and work place detection from massive public transportation data. In: **VISUAL ANALYTICS SCIENCE AND TECHNOLOGY (VAST), 2015 IEEE CONFERENCE ON**, 2015. **Anais...** [S.l.: s.n.], 2015. p. 49–56.

ZHANG, J. et al. Big data density analytics using parallel coordinate visualization. In: **COMPUTATIONAL SCIENCE AND ENGINEERING (CSE), 2014 IEEE 17TH INTERNATIONAL CONFERENCE ON**, 2014. **Anais...** [S.l.: s.n.], 2014. p. 1115–1120.

ZHANG, Z.; TANG, P.; CORPETTI, T. Satellite image time series clustering via affinity propagation. In: **GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS), 2016 IEEE INTERNATIONAL**, 2016. **Anais...** [S.l.: s.n.], 2016. p. 2419–2422.

ZHENG, Y. et al. Recommending friends and locations based on individual location history. **ACM Trans. Web**, New York, NY, USA, v. 5, n. 1, p. 5:1–5:44, 2011.

ZHENG, Y. et al. Telcoflow: visual exploration of collective behaviors based on telco data. In: **BIG DATA (BIG DATA), 2016 IEEE INTERNATIONAL CONFERENCE ON**, 2016. **Anais...** [S.l.: s.n.], 2016. p. 843–852.

ZHOU, H. et al. Visual clustering in parallel coordinates. In: **COMPUTER GRAPHICS FORUM**, 2008. **Anais...** [S.l.: s.n.], 2008. v. 27, n. 3, p. 1047–1054.

ZHOU, J. et al. Visual analytics of relations of multi-attributes in big infrastructure data. In: **BIG DATA VISUAL ANALYTICS (BDVA), 2016, 2016**. **Anais...** [S.l.: s.n.], 2016. p. 1–2.

ZHU, X.; XU, G. Applying visual analytics on traditional data mining process: quick prototype, simple expertise transformation, and better interpretation. In: **ENTERPRISE SYSTEMS (ES), 2016 4TH INTERNATIONAL CONFERENCE ON**, 2016. **Anais...** [S.l.: s.n.], 2016. p. 208–213.