



**UFOP**

Universidade Federal  
de Ouro Preto

**Universidade Federal de Ouro Preto  
Instituto de Ciências Exatas e Aplicadas  
Departamento de Computação e Sistemas**

**Modelando e analisando padrões de  
mobilidade por meio de  
representações de *node embedding***

**Matheus Fellipe do Carmo Barros**

**TRABALHO DE  
CONCLUSÃO DE CURSO**

**ORIENTAÇÃO:**

Carlos Henrique Gomes Ferreira

**COORIENTAÇÃO:**

Bruno Pereira dos Santos

**Setembro, 2021  
João Monlevade–MG**

**Matheus Fellipe do Carmo Barros**

**Modelando e analisando padrões de mobilidade  
por meio de representações de *node embedding***

Orientador: Carlos Henrique Gomes Ferreira

Coorientador: Bruno Pereira dos Santos

Monografia apresentada ao curso de Engenharia de computação do Instituto de Ciências Exatas e Aplicadas, da Universidade Federal de Ouro Preto, como requisito parcial para aprovação na Disciplina “Trabalho de Conclusão de Curso II”.

**Universidade Federal de Ouro Preto**

**João Monlevade**

**Setembro de 2021**

## SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

B277m Barros, Matheus Fellipe do Carmo.

Modelando e analisando padrões de mobilidade por meio de representações de node embedding. [manuscrito] / Matheus Fellipe do Carmo Barros. - 2021.

50 f.: il.: color., gráf., tab..

Orientador: Prof. Me. Carlos Henrique Gomes Ferreira.

Coorientador: Prof. Dr. Bruno Pereira dos Santos.

Monografia (Bacharelado). Universidade Federal de Ouro Preto. Instituto de Ciências Exatas e Aplicadas. Graduação em Engenharia de Computação .

1. Telefonia - Rede móvel. 2. Redes complexas. 3. Teoria dos grafos. I. Ferreira, Carlos Henrique Gomes. II. Santos, Bruno Pereira dos. III. Universidade Federal de Ouro Preto. IV. Título.

CDU 004.738.5:621.39

Bibliotecário(a) Responsável: Sione Galvão Rodrigues - CRB6 / 2526



MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DE OURO PRETO  
REITORIA  
INSTITUTO DE CIÊNCIAS EXATAS E APLICADAS  
DEPARTAMENTO DE COMPUTAÇÃO E SISTEMAS



**FOLHA DE APROVAÇÃO**

**Matheus Fellipe do Carmo Barros**

**Modelando e analisando padrões de mobilidade por meio de representações de *node embedding***

Monografia apresentada ao Curso de Engenharia da Computação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Engenharia de Computação

Aprovada em 02 de Setembro de 2021

**Membros da banca**

Mestre - Carlos Henrique Gomes Ferreira - Orientador Universidade Federal de Ouro Preto  
Doutor - Bruno Pereira dos Santos - Co-orientador Universidade Federal de Ouro Preto  
Doutor - Luiz Carlos Bambirra Torres - Universidade Federal de Ouro Preto  
Doutor - Lourenço Alves Pereira Júnior - Instituto Tecnológico de Aeronáutica

Prof. Carlos Henrique Gomes Ferreira, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 28/09/2021.



Documento assinado eletronicamente por **Carlos Henrique Gomes Ferreira, PROFESSOR DE MAGISTERIO SUPERIOR**, em 28/09/2021, às 17:51, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [http://sei.ufop.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0226299** e o código CRC **EE48D487**.

*Dedico esse trabalho primeiramente a Deus, aos meus avós, família, amigos e professores  
que muito me ajudaram ao longo dessa jornada*

# Agradecimentos

Agradeço primeiramente a Deus por ter me dado força e sabedoria.

Agradeço aos meus avós José Marcelino (*in Memoriam*) e Zélia do Carmo (*in Memoriam*) por terem me dado toda a base para ser quem sou, por todo o amor dedicado em minha criação e por terem me incentivado a nunca desistir dos estudos.

Agradeço a minha mãe Rosângela e meus irmãos, Gabriel e Gustavo, pelo suporte e por todo amor e carinho.

Agradeço a minha esposa Lorena e aos meus filhos, Pietro e Heitor, por todo amor, carinho, companheirismo e por serem o meu refúgio nos momentos difíceis

Agradeço aos meus colegas de curso, em especial a Vinícius Gomes, por toda ajuda e incentivo ao longo do curso.

Agradeço a todos os professores que tive no curso.

Agradeço especialmente o meu orientador Carlos Henrique Gomes Ferreira e o meu coorientador Bruno Pereira dos Santos por todo tempo dispensado em me orientar, ajudar e corrigir os meus erros durante a elaboração desse trabalho.

A todos o meu mais sincero Muito Obrigado !!!

*“Importante não é ver o que ninguém nunca viu, mas sim pensar o que ninguém nunca pensou sobre algo que todo mundo vê.”*

— Arthur Schopenhauer (1788 – 1860),  
*in: Parerga und Paralipomena, 1851.*

# Resumo

Motivados pelo crescente número de dispositivos móveis capazes de se conectar e trocar mensagens, nós propomos uma metodologia para modelagem e análise da mobilidade dos nós em redes móveis. Especificamente, nós notamos que muitas soluções existentes na literatura, comumente, usam medidas topológicas diretamente computadas sob o grafo para capturar a noção de importância do nó em termos de conectividade. No entanto, cada medida possui sua especificidade, portanto, elas falham em capturar aspectos gerais e, muitas vezes, temporais da rede analisada. Diferentemente dessas medidas, nossa metodologia é baseada em uma estratégia de *embeddings*, que é uma forma de mapear variáveis categóricas/discretas em um vetor (de baixa dimensão) de números contínuos. Ao usar *embeddings*, nós somos capazes de representar uma sequência de redes de contato preservando a dimensão espaço-temporal dos nós. Em um estudo de caso realizado, os resultados mostram que nossa abordagem fornece uma representação rica para extração de diferentes padrões de mobilidades, que podem ser úteis para diversas aplicações atreladas a redes, cuja entidades, são móveis. Finalmente, nós comparamos os padrões extraídos pela nossa metodologia com várias medidas topológicas, comumente usadas em diversas soluções deste contexto. Os resultados mostram uma correção moderada com subconjunto delas, o que revela que os padrões extraídos pela nossa abordagem consegue agregar diferentes noções de importância de um nó na rede simultaneamente.

**Palavras-chaves:** Redes Móveis; Redes Complexas; *Node Embedding*; Mobilidade;



# Abstract

Motivated by the growing number of mobile devices capable of connecting and exchanging messages, we propose a methodology aiming at modeling and analyzing the nodes' mobility in mobile networks. Specifically, we note that many existing solutions in the literature commonly use topological measures directly computed over the graph to capture the notion of node importance in terms of connectivity. However, each measure has its specificity and, therefore, fails to capture general aspects. Unlike these metrics, our methodology is based on a *embeddings*, which is a way to map categorical/discrete variable into a vector of low-dimension of continuous numbers. By using embeddings, we are able to represent and extract patterns from a sequence of contact networks preserving the node's spatio-temporal dimension. In a case study carried out, our results show that our approach provides a rich representation for extracting different mobility patterns, which can be helpful for various applications linked to networks whose entities are mobile. Finally, we compare the patterns extracted by our methodology with several topological measures commonly used in several solutions in this context. The results show a moderate correction with a subset of them. It means that the patterns extracted by our approach can simultaneously aggregate different notions of a node's importance in the network.

**Key-words:** Mobile Networks; Complex Networks; Node Embedding; Mobility.

# Lista de ilustrações

Figura 1 – Visão geral da metodologia adotada neste trabalho. . . . .	20
Figura 2 – Exemplo de representações obtidas pelo <i>node2vec</i> em janelas de tempo distintas. . . . .	27
Figura 3 – Exemplo de representações obtidas pelo <i>DynamicNode2vec</i> em janelas de tempo distintas. . . . .	28
Figura 4 – Exemplo de mobilidade de um nó no grafo sendo mapeada na representação latente. . . . .	28
Figura 5 – Exemplo da diversidade estrutural dos grafos de mobilidade obtidos. . .	31
Figura 6 – Distribuição das medidas topológicas gerais das redes modeladas. . . .	33
Figura 7 – Distribuição das medidas topológicas relacionadas ao agrupamento durante o período observado. . . . .	34
Figura 8 – Representação 2-D do espaço vetorial obtido. . . . .	35
Figura 9 – Representação 2-D do espaço vetorial obtido. Apenas os nós com a mobilidade máxima e mínima no período observado são representados. . . . .	36
Figura 10 – CDF da média e do coeficiente de variação da distância de cosseno. . .	37
Figura 11 – Mapas de calor da distância de cosseno de um nó. . . . .	38
Figura 12 – CDF da média e do coeficiente de variação da norma dos nós. . . . .	39
Figura 13 – Mapa de calor da norma dos nós ao longo do período observado. . . . .	40

# Lista de tabelas

Tabela 1 – Parâmetros de entrada do GRM com 100 nós. Adaptado de (NUNES et al., 2017).	22
Tabela 2 – Correlação entre a norma e o C.V..	41
Tabela 3 – Correlações entre as medidas topológicas e do <i>embedding</i> .	42

# Sumário

1	INTRODUÇÃO . . . . .	12
1.1	Definição do problema . . . . .	13
1.2	Objetivos . . . . .	14
1.3	Estrutura do documento . . . . .	15
2	TRABALHOS RELACIONADOS . . . . .	16
3	METODOLOGIA . . . . .	20
3.1	O modelo <i>Group Regularity Mobility – (GRM)</i> . . . . .	21
3.2	Modelagem da rede Temporal . . . . .	23
3.3	DynamicNode2Vec . . . . .	25
4	RESULTADOS . . . . .	33
4.1	Caracterização das redes de contatos . . . . .	33
4.2	Representação do modelo espaço-temporal . . . . .	35
4.3	Compreensão da dinâmica dos usuários . . . . .	37
5	CONCLUSÃO . . . . .	43
	REFERÊNCIAS . . . . .	45

# 1 Introdução

Os dispositivos móveis têm mudado a vida humana, permitindo a conectividade e comunicação ubíqua entre usuários fisicamente móveis. O desenvolvimento e melhoria de soluções neste contexto, tais como, Internet das coisas (IoT), computação vestível, redes oportunísticas, cidades inteligentes, entre outros, exploram a mobilidade no sentido de que conexões de rede são estabelecidas quando os dispositivos móveis ficam no alcance de transmissão um do outro. Consequentemente, a compreensão da mobilidade dos usuários na rede é algo fundamental, pois se relaciona diretamente com o projeto, gerenciamento e desempenho das aplicações e serviços oferecidos por uma rede móvel (HESS et al., 2016). Entre as várias possibilidades, é possível citar aplicações que demandam um roteamento eficaz em redes oportunistas no contexto de cidades inteligentes, e que possuem como foco a disseminação/coleta de informações, localização e comunicação veicular (CAI et al., 2021; SYED et al., 2021).

Com a finalidade de compreender como as entidades se movem, dados históricos (também conhecidos como rastros) são extraídos de infra-estruturas de rede ou dos próprios dispositivos. Entre as várias possibilidades de obtenção desses dados, é possível citar as redes de celulares, mídias sociais, sistemas de transportes, e histórico de posicionamento dos próprios dispositivos (HESS et al., 2016). Ainda, a obtenção desses dados pode ser realizada de duas maneiras. Ativa, quando o usuário compartilha a informação de forma intencional, ou passiva, quando a infraestrutura da rede é explorada e a informação é coletada sem que o usuário saiba ou queira (CELES et al., 2017). Uma vez extraídos, esses dados se tornam importantes facilitadores de estudos nessa direção, pois permitem reconstruir o comportamento das entidades móveis, extrair padrões de mobilidade e gerar dados sintéticos em um ambiente simulado que pode variar em diferentes quesitos tais como espaço, tempo, e quantidade de entidades. Dessa forma é possível, por exemplo, avaliar o desempenho de soluções de redes em diferentes condições bem como propor novas abordagens sem a necessidade da coleta de dados de um cenário real (NUNES et al., 2017).

Tipicamente, estudos de mobilidade sobre as diferentes categorias de redes móveis são realizados por meio da modelagem de grafos temporais. Neste tipo de abordagem, os nós representam as entidades comunicantes e as arestas as interações entre elas ao longo do tempo. A principal vantagem desses modelos, é que medidas e modelos amplamente difundidos no contexto da ciência das redes podem ser diretamente aplicados as redes modeladas. Isso inclui métricas, locais e globais, de centralidade, análise de vizinhança, formação e evolução de comunidades, e modelos de propagação (VASTARDIS; YANG, 2013; ROSSETTI; CAZABET, 2018; BARABÁSI et al., 2016).

Especificamente no contexto de roteamento e disseminação de mensagens na rede, vários esforços foram propostos usando como base diferentes medidas de centralidade, estrutura de comunidades, e várias outras métricas sociais diretamente extraídas dos grafos (SOBIN et al., 2016; DALY; HAAHR, 2007; WAN et al., 2017; ZHANG et al., 2017; NUNES et al., 2018; SANTOS; VIEIRA; LOUREIRO, 2020; GUO; DONG; WEN, 2021). No entanto, essas medidas topológicas capturam aspectos particulares dos nós na rede e são incapazes de generalizar um conceito de importância (VALENTE et al., 2008). Por exemplo, a *centralidade de grau* (do inglês, *Degree Centrality*, considera que os nós importantes são aqueles que possuem o maior número de arestas incidentes a eles, portanto, os mais conectados. Já a centralidade de intermediação (do inglês, *Betweenness*) captura o conceito de que nós importantes são aqueles que funcionam como *pontes* entre uma região e outra na rede (BARABÁSI et al., 2016). Essas especificidades estendem-se para várias outras medidas, discutidas posteriormente neste trabalho, usadas em diversas categorias de soluções para redes móveis. Além disso, capturar essa noção de importância de um nó na rede em termos de conectividade considerando a dimensão temporal torna-se ainda mais desafiador. Numa rede móvel, alguns nós tendem a apresentar padrões de mobilidade regulares e consistentes, já outros esporádicos (NUNES et al., 2017). Portanto, sendo uma alternativa conhecer os padrões de mobilidades destes nós observando-os por um período, e usar este conhecimento para o desenvolvimento de novas soluções como é feito em (OLIVEIRA; VAZAO, 2016; SANTOS; VIEIRA; LOUREIRO, 2020).

Mais recentemente, o uso de abordagens baseadas em *embeddings* tem sido adotadas como uma alternativa viável para diversos estudos de mobilidade incluindo, predição de pontos de interesses (POI) (YANG et al., 2019; YANG et al., 2020), recomendação de POIs (RAHMANI et al., 2019; AGRAWAL; ROY; MITRA, 2021) e planejamento da mobilidade urbana (ZHANG; DUAN; LI, 2021; WANG et al., 2019; XU et al., 2020). Os *embeddings* compreendem representações vetoriais de baixa dimensão, as vezes referidas como *espaço latente*. Tais representações oferecem uma forma eficiente de codificar a estrutura e as propriedades do grafo com intuito de revelar padrões difíceis de serem capturados somente na topologia (LU et al., 2019). Entre as várias vantagens desse tipo de abordagem, estão a redução do alto custo computacional de análises feitas diretamente nos grafos e a melhor eficiência em tarefas como classificação, mineração, e agrupamento de nós (GOYAL; FERRARA, 2018). Apesar disso, poucos trabalhos têm explorado tal modelagem no contexto de mobilidade, principalmente, no contexto de disseminação/roteamento de mensagens.

## 1.1 Definição do problema

Neste trabalho, nós estamos interessados em modelar e analisar padrões de mobilidade em redes móveis capturados por técnicas de *node embedding*. Como mencionado,

soluções focando em propriedades topológicas específicas são incapazes de capturar uma noção de *importância* coletiva e que favoreça a disseminação e roteamento de mensagens na rede. Ademais, a dinâmica das entidades móveis faz com que essas propriedades topológicas mudem constantemente, sendo difícil rastreá-las (FERREIRA et al., 2019). Portanto, é necessária uma abordagem que consiga modelar os padrões de mobilidades dessas entidades considerando a dimensão espaço-temporal. Assim, é possível identificar potenciais nós que favoreçam a disseminação e roteamento de informações na rede de forma eficiente. Neste contexto, a questão norteadora deste trabalho é definida da seguinte forma:

*Dado um rastro descrevendo uma sequência de contatos realizados por entidades móveis durante um período de interesse, como identificar potenciais entidades que facilitem a disseminação/roteamento de mensagens considerando a dimensão espaço-temporal?*

O problema em questão nos leva ao desafio de obter uma representação completa da rede em termos de conectividade e evolução temporal. Para endereçar tal desafio, nós estabelecemos alguns objetivos a seguir.

## 1.2 Objetivos

O objetivo deste trabalho consiste em **modelar interações dos usuários através de técnicas de *node embedding* respeitando a dimensão espaço-temporal com que elas ocorrem.**

Ainda, este objetivo geral pode ser decomposto nos seguintes objetivos específicos:

- **Obter uma representação da mobilidade dos usuários baseada em *node embedding* que considere a estrutura temporal da rede:** Nós pretendemos adotar uma abordagem baseada em *node embedding* para representar os contatos e possíveis trocas de informações dos usuários preservando as propriedades topológicas da rede, especificamente, a distância e importância da conectividade entre os nós;
- **Extrair e avaliar padrões de mobilidade dos usuários na representação obtida:** Em seguida, nós propomos extrair diferentes propriedades de mobilidade dos usuários a partir da representação obtida. Particularmente, propriedades relacionadas a dinamicidade e importância das conexões estabelecidas por eles;
- **Contrastar os padrões extraídos na representação obtida com medidas topológicas tipicamente usadas em soluções de disseminação/roteamento:** Por fim, nós almejamos comparar os padrões extraídos pela nossa abordagem com

medidas topológicas que, tipicamente, são usadas em diversas soluções para redes móveis. Dessa forma, nós pretendemos verificar os benefícios da nossa proposta.

A compreensão dos aspectos sociais em torno da mobilidade têm sido de suma importância para aplicações de redes móveis. Quando isso acontece, surge a oportunidade de novas abordagens capazes de prover melhorias nestes ambientes, tais como, protocolos de comunicação, projetos arquiteturais, alocação e compartilhamento de recursos na rede e aplicações de diferentes naturezas (VASTARDIS; YANG, 2013; XIA et al., 2016). Além disso, desafios clássicos em redes dessa natureza e que afetam a qualidade do serviço, podem ser mais facilmente endereçados. Portanto, a proposta deste trabalho oferece uma gama de possibilidade de estudos posteriores como, por exemplo, identificação de propriedades sociais das entidades móveis, problemas de congestionamento, egoísmo na transmissão de informações por parte dos indivíduos, escalabilidade e tolerância a falhas (XIA et al., 2013; SEGUNDO et al., 2015).

### 1.3 Estrutura do documento

O restante dessa monografia está organizada da seguinte forma. O capítulo 2 apresenta uma breve revisão da literatura no contexto de soluções de *embeddings* aplicadas a diferentes estudos de mobilidade. Em seguida, o capítulo 3 descreve a nossa metodologia dividida em três partes: Obtenção de rastros de mobilidade; Modelagem das redes móveis; e obtenção dos *embeddings*. Os resultados obtidos são apresentados no capítulo 4. Por fim, nós concluímos o nosso trabalho no capítulo 5.



## 2 Trabalhos Relacionados

Relembre que o objetivo deste trabalho é modelar padrões de interações entre usuários móveis considerando a dimensão espaço-tempo através de técnicas de representação de nós em espaços vetoriais de baixa dimensão, conhecido como *embeddings*. A modelagem da mobilidade e interações entre entidades apresenta uma aplicabilidade rica. Neste sentido, soluções cientes da mobilidade foram feitas, por exemplo, métricas de roteamento (BATABYAL; BHAUMIK, 2015; REIS; SANTOS; VIEIRA, 2021; SANTOS et al., 2018), protocolos e ações para coordenação (ALAJEELY; DOSS; AHMAD, 2018), bem como na modelagem e análise do comportamento dos indivíduos em uma rede móvel (HESS et al., 2015). Apesar desta ampla aplicabilidade, soluções apoiadas por técnicas de *embeddings* tem sido pouco exploradas. No entanto, *embeddings* se mostra como uma solução viável para diferentes estudos de mobilidade como discutiremos a seguir. Visto que muitos estudos de mobilidade fazem o uso de modelos de redes para representar os contatos entre entidades (usuários, objetos, etc.) de um dado ambiente ou sistema sendo analisado, torna-se viável o uso de técnicas dessa natureza. Nesta seção, nós descrevemos alguns dos principais esforços da literatura que adotaram estratégias de *embeddings* para modelar a mobilidade dos usuários sob diferentes perspectivas e motivados por diferentes tarefas. Por fim, nós colocamos a nossa contribuição em perspectiva comparando-a a esses esforços pré-existentes.

Yu et al. (YU; WANG; LI, 2018) argumentaram que pode não ser realístico usar a frequência de contatos entre entidades (ex., usuários) como métrica para inferir relações sociais. Isso porque usuários podem frequentemente visitar um mesmo lugar, sem ter nenhum tipo de relação, e por alguma coincidência, entrarem em contato. Por outro lado, outras duas entidades podem raramente visitar um dado local, mas esse local pode ser mais informativo sobre as preferências e particularidades da mobilidade delas. Neste contexto, os autores propuseram inferir a força da relação entre os usuários (expressas pelos contatos que eles tiveram) bem como a motivação por trás delas. Para tanto, é usado uma abordagem baseada em *embedding* que considera uma rede modelando a frequência de contato entre duas entidades cujo peso da aresta é ajustado com base nas informações contextuais da localização do contato.

Yang et al. (YANG et al., 2019) focaram nas tarefas de previsão de novas amizades e de previsão de localização dos usuários. Para isso os autores propuseram LBSN2Vec, uma abordagem de *embedding* que considera uma rede de contatos modelada como um hipergrafo. LBSN2Vec baseia-se em uma estratégia de passeio aleatório sobre o hipergrafo para amostrar o grafo e obter uma representação vetorial (isto é, um *embedding*) que preserve as propriedades dos nós na rede (por exemplo, a proximidade entre eles) no espaço vetorial

obtido. Mais recentemente (YANG et al., 2020), os autores incorporaram informação contextual na abordagem possibilitada pela criação de hiperarestas considerando não só contatos dos usuários, mas também pontos de interesses e tempo. A abordagem, denominada LSBN2Vec++, mostrou ser capaz de fornecer uma representação mais rica e eficiente do que a versão anterior para ambas as tarefas em questão.

Técnicas de *embeddings* no contexto de mobilidade também tem sido utilizadas em tarefas associadas a recuperação de informação. Especificamente no contexto de recomendação de POIs, Rahmani et al. (RAHMANI et al., 2019) propuseram CATAPE, um método que incorpora simultaneamente sequências de localizações dos usuários e categorias dos POIs visitados por eles para recomendar potenciais novos pontos de interesses. A ideia é que as sequências de localizações capturem a influência geográfica enquanto as categorias dos POIs de interesse forneçam características desses pontos. Outros trabalhos como o de Agrawal et al. (AGRAWAL; ROY; MITRA, 2021) são ainda mais restritivos e totalmente baseados em informações contextuais já que não usam nenhuma informação sobre contatos. Neste caso, os autores usaram a nota dada pelo usuário após visitar um POI para fornecer novas recomendações (AGRAWAL; ROY; MITRA, 2021). De forma similar, Xie et al. exploraram o potencial de recomendação do próximo POI a um usuário dado o seu histórico de mobilidade (XIE et al., 2016).

Ainda com foco na recomendação de pontos de interesses, Yuan et al. propuseram utilizar informações de consultas anteriores, consultas de outros usuários no sistema e histórico de localização como informações para obtenção de uma representação vetorial (YUAN et al., 2020). O foco está em recomendar a um usuário um POI através de palavras chaves em uma consulta incompleta. Já Wang et al. focaram simultaneamente na recomendação e predição de pontos de interesses dos usuários (WANG et al., 2019). A abordagem baseada em *embeddings* proposta por eles considera um processo de amostragem em um grafo heterogêneo capaz de representar informações do tempo, usuários, POIs e suas categorias.

Outros trabalhos produziram soluções baseadas em *embeddings* motivadas por problemas de mobilidade urbana. Por exemplo, Zhang et al. usaram uma estratégia de *embedding* para extrair agrupamentos de linhas de transportes conforme os dados de mobilidade dos usuários no sistema de transporte (ZHANG et al., 2019). Em (ZHANG; DUAN; LI, 2021), os autores consideraram também dados de POIs para a construção de redes de mobilidade urbana dos usuários. O objetivo era obter uma representação *embedding* capaz de codificar as similaridades entre os padrões de mobilidades dos usuários considerando ambas as fontes de dados. Em um estudo de caso real e com dados de usuários de transporte público urbano, foi possível encontrar comunidades formadas por usuários que possuem os mesmos padrões de mobilidade. De maneira similar, Wang et al. propuseram um arcabouço baseado em *embedding* cujo foco é analisar a dinâmica de comunidades formadas pela mobilidade entre POIs em diferentes períodos de interesse

(por exemplo, dias úteis e finais de semana) (WANG et al., 2018).

Já o trabalho de Xu *et al.* (XU et al., 2020) apresenta o *SUME*, que é uma estratégia baseada em *embeddings* para inferência demográfica de usuários a partir POIs. O *SUME* recebe como entrada um modelo de rede heterogênea em que diversas categorias de arestas entre usuários e locais são criadas. Uma parte delas codifica co-visitações e a outra parte informações semânticas dos POIs. Em seguida, essas redes são representadas no espaço vetorial cujo objetivo é preservar a semelhança dos usuários nos padrões de mobilidade urbana.

É possível observar que o uso de estratégias baseadas em *embeddings* tem sido aplicadas de forma diversificada em problemas relacionados a mobilidade de entidades. A seguir, colamos o trabalho aqui proposto em perspectiva destacando as principais diferenças em relação aos trabalhos acima discutidos. Várias abordagens recorreram a informações contextuais além da rede de contato (YU; WANG; LI, 2018; YANG et al., 2019; YANG et al., 2020). De fato, nós percebemos que tais informações oferecem maiores possibilidades de análises e extração de padrões. Por outro lado, isso impõe desafios extras desde a coleta e armazenamento, visto que tais dados nem sempre estão disponíveis e abertos, até questões relacionadas a privacidade e segurança. Como discutiremos no próximo capítulo, nossa abordagem independe de informações contextuais e baseia-se exclusivamente em uma sequência de grafos de contatos que representa a organização topológica da rede. Isso é motivado pelo problema aqui tratado, que consiste em identificar potenciais nós para auxiliar no roteamento de mensagens. Neste esforço inicial, nós usamos apenas das informações topológicas da rede, embora nossa abordagem possa ser estendida para considerar informações contextuais.

Nós também observamos que algumas abordagens adotam estratégias de amostragem através de passeios aleatórios, similar a nossa proposta, apesar de considerarem grafos heterogêneos (múltiplas categorias de nós e/ou de arestas) (WANG et al., 2018; WANG et al., 2019; YANG et al., 2020). De fato, passeios aleatórios tem possibilitado identificar importantes propriedades acerca dos nós no grafo. Por exemplo, conectividade entre os nós (proximidade) e nós que desempenham papéis similares de conectividade (*Hubs* e periféricos) (GROVER; LESKOVEC, 2016). Na abordagem aqui proposta, um grafo tradicional e homogêneo foi usado no processo de amostragem, isto é, nós representam usuários e arestas representam o contato entre estes usuários. Além de não utilizarmos informações contextuais, a dimensão temporal é considerada através de uma sequência de grafos modelando janelas de tempo (*snapshots*) distintas, que nos permite avaliar a dimensão espaço-tempo.

Em termos de aplicações, os trabalhos discutidos aqui compreendem em esforços ortogonais aos nossos. No que diz respeito a trabalhos focados em predição (YANG et al., 2019; YANG et al., 2020), embora essa tarefa esteja fora do escopo deste trabalho,

vale ressaltar que nossa abordagem é baseada na premissa de que encontros ocorrem regularmente, inerente a modelos de mobilidade tradicionais (NUNES et al., 2017). Portanto, nosso trabalho pode ser estendido para endereçar tal tarefa. Observamos trabalhos com foco em recomendação de POIs (RAHMANI et al., 2019; AGRAWAL; ROY; MITRA, 2021; YANG et al., 2020). Isso é feito usando dados contextuais, que também está fora do nosso escopo. Por fim, vimos esforços relacionados a melhoria da mobilidade urbana (ZHANG; DUAN; LI, 2021; WANG et al., 2019; XU et al., 2020). Em relação a esses esforços, nossa proposta pode ser vista como uma tentativa inicial em direção ao planejamento e projeto de aplicações inerentes a um roteamento efetivo em redes oportunísticas, que oferece uma gama de aplicações em cidades inteligentes tais como disseminação/coleta de informações, localização, comunicação veicular (CAI et al., 2021; SYED et al., 2021).

### 3 Metodologia

Neste capítulo, nós detalhamos a metodologia seguida neste trabalho, composta pelas três primeiras etapas apresentadas na Figura 1. A primeira delas, consiste na obtenção dos dados de mobilidade. Para isso, recorremos a um modelo de mobilidade disponível na literatura. Esse modelo possibilita a geração de rastros sintéticos de mobilidade que mimetizam características de mobilidade humana que são reais. Este e outros aspectos referentes a essa etapa são discutidos na Seção 3.1. Uma vez obtidos os rastros de mobilidade, nós então modelamos uma sequência de redes de contato entre os nós, dividindo o período observado em janelas de tempos de tamanho fixo e sem sobreposição como descrito na Seção 3.2. Em seguida, nós propomos na Seção 3.3 o uso de *DynamicNode2Vec*, uma técnica capaz de gerar representações vetoriais de baixa dimensão, para os nós dessas redes (também conhecidas como *embeddings*) preservando a dimensão espaço-temporal. Finalmente, no Capítulo 4, investigamos a efetividade da nossa proposta para extrair padrões de mobilidade. Para comparação, nós também computamos algumas medidas estruturais da rede com o objetivo de entender as diferenças entre os padrões capturados no espaço latente proposto e da estrutura do grafo.

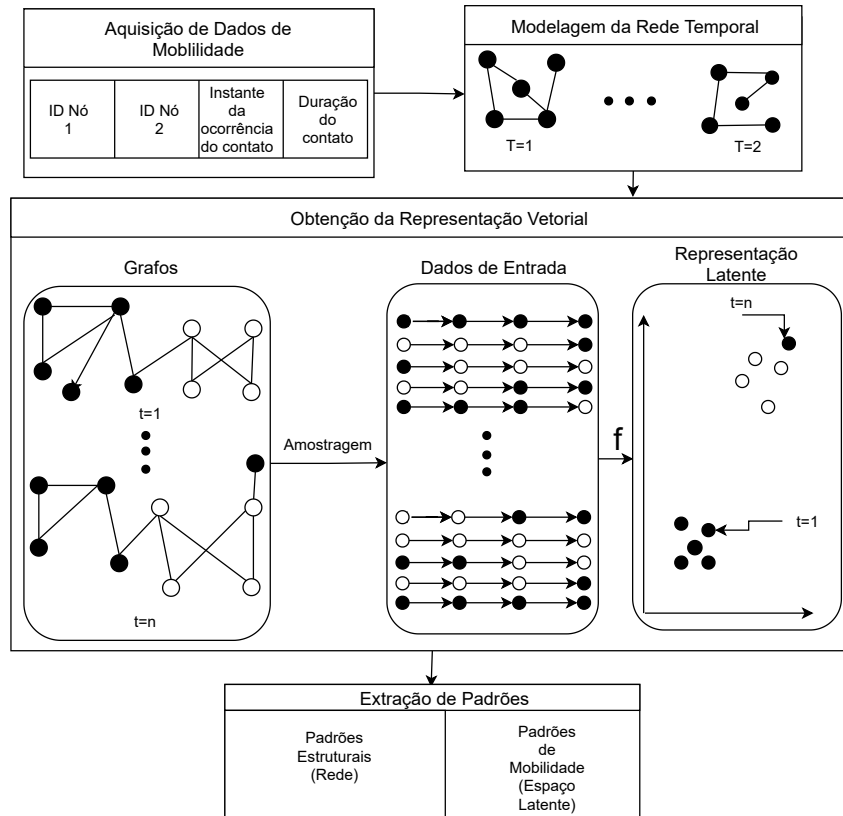


Figura 1 – Visão geral da metodologia adotada neste trabalho.

### 3.1 O modelo *Group Regularity Mobility* – (GRM)

O estudo de mobilidade é importante em diversos campos do conhecimento. Entretanto, obter registros de mobilidade de entidades, quer sejam elas humanas ou não, não é uma tarefa trivial. Entre os principais motivos, podemos citar questões de privacidade, segurança, armazenamento e processamento desses registros. Oportunamente, pesquisadores tem desenvolvido modelos de mobilidade para contornar esses problemas. Modelos de mobilidade mimetizam o comportamento de entidades móveis (humanas ou não) e nos permite gerar rastros de mobilidade<sup>1</sup> em diversas dimensões: espacial, temporal e tamanho. As utilidades desses modelos são diversas, variam do campo de estudo de redes oportunísticas, ao campo da Internet das Coisas (IoT) e outros (MOTA et al., 2014; SANTOS; VIEIRA; LOUREIRO, 2020; SANTOS et al., 2021). Neste trabalho, nós utilizamos um modelo que possibilita estudar mobilidade em diferentes escalas sendo esse modelo apresentado a seguir.

Neste trabalho, nós usamos o GRM (do inglês, *Group Regularity Mobility Model*) proposto em (NUNES et al., 2017). GRM é um modelo de mobilidade humana que busca apresentar a dinâmica e a regularidade de encontros de grupos de usuários de uma determinada rede. Esse modelo simula o padrão de mobilidade humana considerando não só o comportamento cíclico, comumente apresentado por humanos, como também o aspecto de encontros esporádicos em reuniões, na academia, no trabalho e outros. Essas características diferem o GRM da maioria dos demais modelos disponíveis na literatura. Dessa forma, o uso deste modelo nos permite simular, em larga escala e com baixo custo computacional, entidades que representam humanos, experimentar e avaliar propriedades relacionadas a mobilidade das entidades na perspectiva espaço temporal.

Para gerar um rastro, GRM precisa de três informações. Primeiro, um modelo de rede social de contatos fornecido pelo usuário ou pré-definido na literatura. Além do modelo de rede social, GRM precisa de um conjunto de configurações de simulação que inclui o tamanho da área simulada (*Grade, TamanhodaCélula*), a duração da simulação (*Dur.daSimulação*), o número de nós (*#Nós*) e o número de grupos (*#Grupos*).

Por fim, um conjunto de parâmetros estatísticos relacionados as distribuições estatísticas que descrevem o rastro de mobilidade usado (NUNES et al., 2017). São eles:

- **Horários dos Encontros dos Grupos (GMT):** Propriedade que trata do intervalo de tempo entre os encontros dos usuários. Cada grupo recebe um tempo médio de inter-contato. O valor deste tempo é gerado a partir de uma distribuição lei de potência cujo expoente ( $\alpha_{gmt}$ ) e o valor de corte (ou truncamento) ( $\beta_{gmt}$ ) são especificados no modelo. Além disso, o GRM considera que cada grupo possui um fator de regularidade definido por  $K$ . Por exemplo, grupos com um fator  $k = 24h$

<sup>1</sup> Registros de posicionamento, relações e/ou interações entre entidades.

se reúnem a cada, 24, 48 ou 72h. Este parâmetro funciona como um multiplicador que irá gerar o comportamento periódico dos traços reais. Sendo assim, é necessário especificar a distribuição dos valores de  $K$  para os grupos;

- **Duração dos Encontros dos Grupos (CD):** Propriedade que trata a duração de cada encontro realizado por grupo. Importante em redes móveis para saber a quantidade de dados que podem ser transferidos durante um determinado contato e identificar a confiança de que uma mensagem chegará em um dado nó. Com base em estudos anteriores, essa propriedade também é definida com base em distribuições de leis de potência. Portanto, similar ao GMT, dois parâmetros são especificados: o expoente ( $\alpha_{dur}$ ) e o valor de corte ( $\beta_{dur}$ );
- **Estrutura dos Grupos e Contexto Social (GESC):** Propriedade responsável por definir o tamanho dos grupos. Também com base em estudos anteriores, GRM usa uma distribuição de lei de potência truncada para gerar os tamanhos dos grupos. Similar aos parâmetros anteriores, essa propriedade exige o uso de dois parâmetros: o expoente ( $\alpha_{size}$ ) e o valor de corte ( $\beta_{size}$ ).

No contexto deste trabalho, nós utilizamos um rastro sintético contendo 100 entidades móveis cobrindo um período de 90 dias, disponibilizado pelos autores (NUNES et al., 2017). Este rastro é baseado em um modelo já conhecido na literatura, o modelo agrupamento Gaussiano (BRANDES; GAERTLER; WAGNER, 2003). É importante destacar que a escolha deste rastro, em específico, não impede que nossa abordagem seja usada em outros rastros. Por fim, a Tabela 1 descreve os valores dos parâmetros usados pelo GRM para gerar o rastro sintético:

Tabela 1 – Parâmetros de entrada do GRM com 100 nós. Adaptado de (NUNES et al., 2017).

Parâmetros	Valores
# de Nós	100
# de Grupos	500
Dur. da Simulação	90 dias
K	70% -24h; 15% -7 dias; 15% - 6h
Grade	30 x 30
Tamanho da Célula (m)	50
$\alpha_{gmt}$	3
$\beta_{gmt}$	30 dias
$\alpha_{dur}$	3
$\beta_{dur}$	30 dias
$\alpha_{size}$	2.24
$\beta_{size}$	30
Rede Social	Gaussian Random Partition



Uma vez definido o rastro de mobilidade a ser estudado, nós movemos para a próxima seção onde explicamos a modelagem da rede temporal a partir do traço.

## 3.2 Modelagem da rede Temporal

Antes de apresentar o modelo de rede estudado nesse trabalho, nós brevemente revisitamos o conceito de rede. De forma simples, uma rede é representada por um grafo  $G = (V, A)$  definido por um par de conjuntos  $V$  e  $A$ , onde  $V$  é o conjunto de nós (ou vértices) e  $A$  um conjunto de arestas. Uma aresta é dita não direcionada se ela representar uma ligação, em ambos sentidos, entre dois nós. Em uma rede de contatos, por exemplo, cada usuário pode ser tratado como um nó e os contatos entre os usuários são representados por arestas não direcionadas.

No entanto, como explica Rossetti *et al.* (ROSSETTI; CAZABET, 2018), uma rede naturalmente muda temporalmente. Consequentemente, nós são adicionados ou retirados da rede e as ligações entre eles mudam. Para capturar essa dinâmica natural na rede, existe uma variedade de modelos de redes temporais, cada um com sua vantagem e desvantagem. Aqui nós escolhemos o modelo baseado em janelas de tempos discretas fixas e sem sobreposição, também conhecido como um modelo baseado em (snapshots) que explicaremos agora no contexto do nosso trabalho.

Para discretizarmos o período de interesse, que compreende em 87 dias<sup>2</sup>, em janelas de tempo não sobrepostas e de duração de um dia representadas pelo conjunto  $T = \{1, 2, 3, \dots, 87\}$ . Para cada janela de tempo  $t$  nós modelamos um grafo  $G_t$  não direcionado  $G_t = (V_t, A_t)$ , tal que  $V_t = \{v_1, v_2, v_3, \dots, v_i\}$  é um conjunto de nós representando indivíduos no instante de tempo  $t$  e  $A_t = \{a_1, a_2, a_3, \dots, a_n\}$  um conjunto de arestas ligando dois quaisquer nós  $v_i$  e  $v_j$  para  $v_i \neq v_j$  no instante de tempo  $t$  se eles estiveram próximos em um raio de 100 metros. Na prática, isso representa uma troca de mensagem ou uma potencial troca de mensagens entre dois nós quaisquer para cenários e aplicações específicas como, por exemplo, conexões Bluetooth 5.0 ou Wi-Fi 6, (YAAKOP *et al.*, 2017; YIN *et al.*, 2019). Considerando a área de 2,25km<sup>2</sup> definida pela base de dados utilizada, tal cenário representa aplicações práticas em áreas como *universidades, indústrias, shoppings*, etc.

Com a modelagem da rede, uma possibilidade a ser explorada é a extração de propriedades puramente estruturais que podem ser úteis para o problema aqui tratado. Entre as várias possibilidades, estão as medidas de centralidade e agrupamento, extraídas diretamente da estrutura dos grafos. Vários trabalhos da literatura já destacaram a importância de várias dessas medidas para o processo de roteamento/disseminação de informação e propuseram diversas soluções baseadas nelas (DALY; HAAHR, 2007; WAN *et*

<sup>2</sup> Embora o rastro original forneça um período de observação de 90 dias, nós notamos que os últimos 3 dias apresentam um pequeno número de nós com contato (<8). Por isso optamos por não analisar estes últimos 3 dias.



al., 2017; ZHANG et al., 2017; NUNES et al., 2018; GUO; DONG; WEN, 2021). Portanto, nós utilizamos algumas delas para comparar o quanto a nossa abordagem diverge dessas medidas quando o objetivo é considerar potenciais nós para disseminação da informação bem como a importância deles. Especificamente, nós utilizamos as seguintes medidas (NEWMAN, 2005; GOLBECK, 2013; BARABÁSI et al., 2016; HANSEN et al., 2019):

- **Centralidade de Grau** (do inglês, *Degree Centrality*): Representa o número de arestas incidentes ao nó. Para o problema tratado aqui, um nó com grau alto, que significa ter mais conexões, possui maior capacidade de troca de mensagens por ter vários vizinhos imediatos na rede. Note que esta é uma medida totalmente local, por não considerar as outras conexões dos demais nós na rede. Dessa forma, o resultado é um valor absoluto que é o número de arestas incidentes ao nó;
- **Coeficiente de Agrupamento** (do inglês, *Clustering Coefficient*): Mede o grau em que os nós em um grafo tendem a se agrupar. Para cada nó, o coeficiente de agrupamento é computado pela razão entre o número de triângulos (conjuntos de três nós conectados) ligados ao nó e o número total de triângulos possíveis (abertos e fechados) considerando o grau do nó. Pode ser usado também na versão *média*, onde a média do coeficiente de agrupamento de todos os nós é computada. Um coeficiente de agrupamento médio alto (próximo de 1) indica robustez na rede, ou seja, os nós estão bem conectados com seus vizinhos, portanto, exibe muitas possibilidades de troca de mensagens. Já um valor próximo de 0, indica nós pouco conectados e com baixa tendência a se conectarem;
- **Centralidade de Intermediação** (do inglês, *Betweenness Centrality*): É uma medida da centralidade calculada por meio da fração dos caminhos mais curtos entre os pares de nós que passam pelo nó de interesse. Essa métrica pode ser utilizada para identificar qual o grau de influência um determinado nó tem na distribuição de informações na rede. O valor varia entre 0 e 1, onde 1 indica que um nó possui bastante importância para o fluxo de informação na rede, enquanto um valor próximo de 0 indica o contrário;
- **Centralidade da Proximidade** (do inglês, *Closeness Centrality*): Indica a proximidade de um nó para todos os outros nós da rede. É obtida calculando o comprimento do caminho mais curto de um dado nó a todos os outros nós da rede. Na versão normalizada, seu valor varia entre 0 e 1, onde um valor próximo de 0 indica que um nó é totalmente periférico na rede e não é central em termos de distância dos outros nós. Já valor próximo de 1 indica que o nó facilmente acessa todos os outros nós da rede por ter baixos caminhos até eles;
- **Centralidade do Autovetor** (do inglês, *Eigenvector Centrality*): Essa métrica considera não somente informações dos vizinhos imediatos, mas também a importân-

cia dos nós aos quais os vizinhos imediatos estão conectados, e assim sucessivamente. Captura situações em que uma conexão com um indivíduo popular é mais importante do que uma conexão com um indivíduo solitário. Basicamente analisa não a quantidade de nós que um determinado nó se relaciona, mas a “qualidade” desses nós no sentido de disseminação de informações na rede. Valor próximo de 1 significa que um nó está conectado a muitos nós que também possuem uma alta importância para conectividade da rede. Já um valor próximo de 0, indica o contrário;

Outra propriedade bastante interessante em redes é a formação de comunidades. Entre as várias definições de comunidade na literatura, nós consideramos a de que uma comunidade é formada por nós massivamente conectados entre si e pouco conectado com nós de outras comunidades (FORTUNATO; HRIC, 2016). Com base nisso, gostaríamos de entender como as comunidades estão definidas no rastro utilizado, quando os contatos são observados em janelas de tempo diárias. Para isso, propomos extrair comunidades em cada grafo  $G_t$  usando o algoritmo de Louvain (BLONDEL et al., 2008), uma solução bastante conhecida e utilizada na literatura.

O algoritmo de Louvain funciona maximizando uma medida definida como *modularidade*. Tal medida quantifica o quanto os nós de uma comunidade estão densamente conectados, em comparação a como seria a situação deles em uma rede aleatória. Ela é definida na faixa de -0.5 a 1, embora valores entre 0.3 e 0.7 representem comunidades bem formadas. Em resumo, o algoritmo Louvain opera primeiro encontrando pequenas comunidades otimizando a modularidade localmente em todos os nós. Em seguida, cada pequena comunidade é fundida em um meta-nó e a primeira etapa é repetida. O número final de comunidades é o resultado de um procedimento de otimização. Nós recomendamos a leitura de (BLONDEL et al., 2008) para maiores detalhes.

### 3.3 DynamicNode2Vec

Com intuito de extrair padrões de mobilidade dos usuários na sequência de redes definidas na seção anterior, nós usamos *DynamicNode2Vec* (FERREIRA et al., 2019). Essa técnica permite representar os nós de uma dada sequência de grafos em um espaço vetorial de baixa dimensão (embeddings) preservando as propriedades existentes no grafo. *DynamicNode2Vec* é baseada em duas técnicas. A primeira delas compreende o estado da arte na obtenção de *embeddings* de nós no contexto estático, denominada *Node2Vec* (GROVER; LESKOVEC, 2016). A outra, denominada *DynamicWord2Vec*, permite obter uma representação temporal de *embeddings* no contexto de palavras (do inglês, *word embeddings*) (YAO et al., 2018). Dessa forma, para compreender como *DynamicNode2Vec* funciona, nós brevemente o descrevemos considerando ambas as técnicas.

*Node2vec* é uma das mais populares técnicas de *node embedding* da literatura. A ideia principal dessa técnica consiste em usar um algoritmo de passeio aleatório (do inglês, *random walk*) enviesado, que perfaz caminhos no grafo retornando uma sequência de nós visitados e repetido várias vezes. Esse processo de amostragem é executado a partir de todos os nós. Dessa forma, os dois primeiros parâmetros a serem definidos compreendem no número de *passeios* por nó ( $nw$ ) e o número de passos que cada um dá ( $wl$ ).

É possível configurar ainda o tipo de amostragem a ser realizada, que pode seguir uma busca em largura (do inglês, *breadth-first sampling* (BFS)) ou uma busca em profundidade (do inglês, *depth-first sampling* (DFS)). Se o processo de amostragem for baseado em BFS, a vizinhança a ser visitada a partir de um determinado nó de origem é restrita aos nós mais próximos, enquanto o DFS consiste em nós amostrados sequencialmente em distâncias crescentes. Aqui, nós estamos interessados na amostragem baseada em BFS, que visa capturar a proximidade e a importância das conexões entre os nós a ser representada no espaço vetorial. Para controlar esse comportamento, outros dois parâmetros são usados,  $p$  e  $q$ . O primeiro determina a probabilidade de voltar imediatamente a um nó já visitado enquanto o segundo controla se o passeio fica perto do nó de origem, explorando a mesma vizinhança (ou seja, correspondendo a BFS), ou se deve andar mais longe, explorando outros nós (ou seja, correspondendo a DFS). Seguindo as recomendações dos autores para o nosso objetivo (GROVER; LESKOVEC, 2016), nós definimos  $p = 1$  e  $q = 0.5$ .

Após a definição de tais parâmetros, os caminhos são gerados de forma que nós com maior proximidade e vizinhança em comum, possuem maior probabilidade de aparecerem próximos nos múltiplos caminhos amostrados no grafo. O fluxo natural por trás do *node2vec* seria obter uma representação de  $d$  dimensões ( $d < |V|$ ) de forma a maximizar a representação de propriedades contidas no grafo no espaço vetorial, ou seja, a distância de um dado nó para os demais nós. Preservar a proximidade dos nós na rede no espaço vetorial é apenas uma das possibilidades de representação dessa técnica. Propriedades como, por exemplo, homofilia também podem ser exploradas. Para essa e outras aplicações, nós recomendamos a leitura do artigo original (GROVER; LESKOVEC, 2016).

No entanto, *Node2Vec* funciona apenas para o contexto estático e não pode ser aplicado para o contexto temporal. Este é um problema conhecido na literatura como sendo *problema de alinhamento* (YAO et al., 2018). Na prática, isso significa que quando aplicado a uma sequência de grafos, as representações são obtidas de forma independente e não são mapeadas no mesmo espaço vetorial. Como consequência, a representação de um nó obtida em uma dada janela de tempo não pode ser comparada com outras representações geradas em outras janelas de tempo. Assim, não é possível rastrear um elemento (ou seja, um nó) em janelas de tempo distintas. As Figuras 2a e 2b exemplificam este problema para dois dados instantes de tempo  $t$  e  $t+1$ .

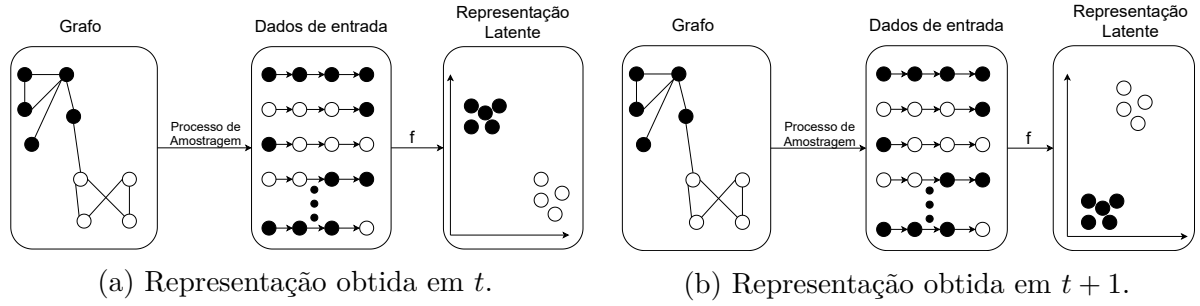


Figura 2 – Exemplo de representações obtidas pelo *node2vec* em janelas de tempo distintas.

Na Figura 2a nós temos um grafo de mobilidade no instante de tempo  $t$ , seguido do processo de amostragem do *Node2Vec*. Os caminhos amostrados pelos passeios aleatórios são tomados como entrada para uma função de otimização que irá produzir o espaço vetorial. Para efeitos de visualização, nós representamos este espaço  $n$ -dimensional em duas dimensões. Isso pode ser alcançado com técnicas de redução de dimensionalidade, por exemplo, t-SNE ou PCA (MAATEN; HINTON, 2008; JOLLIFFE; CADIMA, 2016). Dessa forma, nós podemos ver como as vizinhanças são preservadas no espaço latente. A Figura 2b representa o mesmo processo para a janela de tempo  $t+1$ . Observe ainda que o grafo de contatos não muda e, embora o processo de amostragem resulte em entradas similares, o que dificilmente acontece na prática por se tratar de um processo estocástico, a representação dos nós no espaço latente difere-se da anterior. Para uma análise temporal, os nós deveriam ter representações equivalentes, consistente com a ideia de mobilidade que queremos capturar. Este é o principal desafio do *problema de alinhamento*, preservar semelhanças e revelar diferenças da vizinhança temporalmente no mesmo espaço latente.

Para endereçar tal desafio, *DynamicNode2Vec* incorpora outra técnica denominada *Dynamic Word2Vec* (YAO et al., 2018), responsável por observar a associação dos nós ao longo do tempo conforme os caminhos amostrados, mapeando-os em um espaço latente equivalente. Resumidamente, isso é alcançado da seguinte forma. Para cada janela de tempo observada, uma matriz de co-ocorrência dos nós é computada através de uma janela deslizante de tamanho 5 que percorre os caminhos amostrados pelo *Node2Vec* e verifica quantas vezes dois dados nós são visitados dentro dessa janela. Dessa forma, quanto maior a co-ocorrência entre estes dois nós, mais próximo eles estão na rede e mais importante é a conexão entre eles, o que torna essa matriz bastante informativa. Após computar uma sequência de matrizes de co-ocorrências correspondentes a sequência de redes observadas, uma representação compacta de  $d$  dimensão é obtida para cada matriz através de decomposição dessa matriz (do inglês, *low-rank factorization* (MF)). A decomposição de matriz é uma ferramenta eficaz para analisar dados de modo a descobrir as interações entre duas entradas. Por fim, o alinhamento temporal é obtido por meio da solução de um problema de otimização resolvido através da técnica *stochastic gradient descent*. Dois principais parâmetros de ajuste são adicionados a uma função objetivo para

tratar, respectivamente, de problemas de *sobreaajuste* ( $\lambda$ ) e *alinhamento* ( $\tau$ ).

As Figuras 3a e 3b exemplificam como *DynamicNode2Vec* funciona. De forma similar ao exemplo anterior, nós temos dois grafos de contatos observados em duas janelas de tempo distintas. Cada um deles passa por um processo de amostragem independente, porém a função de mapeamento dos nós para o espaço vetorial considera ambas as janelas ao obter uma representação para os nós. Como resultado, a mesma representação é obtida nas duas janelas de tempo observadas. Isso nos permite observar que nada mudou em termos de mobilidade já que os nós na rede possuem as mesmas conexão de antes.

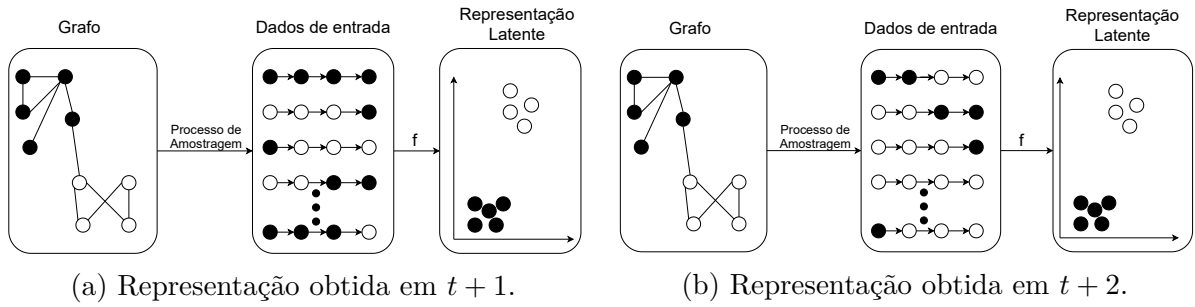


Figura 3 – Exemplo de representações obtidas pelo *DynamicNode2Vec* em janelas de tempo distintas.

Por fim, nós apresentamos outro exemplo nas Figuras 4a e 4b para mostrarmos como o fenômeno de interesse (a mobilidade dos nós) é capturado. As setas vermelhas em ambas as figuras destacam um nó móvel na rede de contatos entre uma janela e outra. Olhando na rede modelada representada pelo grafo, podemos observar como o nó mudou radicalmente seu único contato. É possível observar também que a distância para o conjunto de nós de cor preta aumentou significativamente. De forma similar, nós podemos observar a representação latente obtida em que o mesmo comportamento é mapeado em relação a este nó. Isso nos permite quantificar e mapear mudanças dos nós no espaço vetorial, como sendo um reflexo da mobilidade dos nós mapeadas pelas redes de contatos.

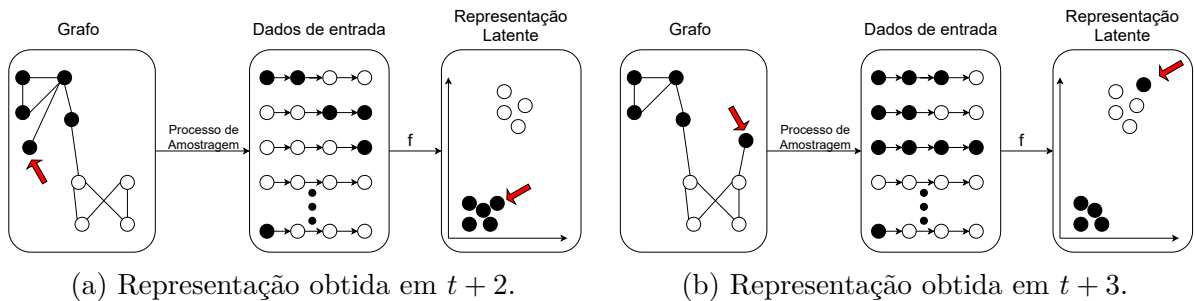


Figura 4 – Exemplo de mobilidade de um nó no grafo sendo mapeada na representação latente.

Para quantificar o nível de mobilidade na rede, nós usamos duas métricas (FERREIRA et al., 2019; YAO et al., 2018):

- **Distância de Cosseno:** A distância de cosseno quantifica a similaridade entre dois vetores em um mesmo espaço vetorial. Dessa forma, dados dois vetores de um nó (usuário)  $v_i$  representados por duas janelas de tempo quaisquer  $t_1$  e  $t_2$ , é possível calcular a sua mobilidade usando a distância de cosseno:

$$\cos(v_{i(t_1)}, v_{i(t_2)}) = 1 - \frac{\mathbf{u}_{v_i}(t_1) \cdot \mathbf{u}_{v_i}(t_2)}{\|\mathbf{u}_{v_i}(t_1)\| \|\mathbf{u}_{v_i}(t_2)\|}. \quad (3.1)$$

A distância de cosseno varia de 0 a 1. Valores próximos a 0 indicam que o nó não apresentou nenhuma mudança de contato entre as duas janelas de tempo comparadas. Por outro lado, valores próximos a 1 indicam que o nó mudou drasticamente sua rede de contatos no período.

- **Norma do vetor no espaço vetorial:** Por definição, quanto mais um nó aparece nos caminhos amostrados no grafo, maior é a *norma* do seu vetor. Dessa forma, é possível verificar os nós que possuem conexões importantes com outros nós, visto que eles estão frequentemente inseridos nos caminhos amostrados. A norma de um vetor correspondente a um nó  $v_i$  no instante de tempo  $t$  é definida por  $\|\mathbf{u}_{v_i}(t)\|$ .

A principal vantagem da utilização da norma em relação à frequência com que os nós aparecem nos caminhos amostrados como uma medida de importância de conectividade, é que ela é mais robusta a perturbações esporádicas. Por exemplo, se um determinado nó aparece exorbitantemente nos caminhos amostrados em uma única janela de tempo, esse valor geraria um ruído ao observarmos um período por se tratar de um valor absoluto. Consequentemente, essa pequena perturbação o indicaria, por exemplo, como um bom candidato para o roteamento de mensagens considerando o período observado, o que pode não ser verdade. Por outro lado, a norma suaviza essa perturbação e evita este tipo de problema.

Note que essas duas métricas são complementares e capturam padrões distintos. A distância de cosseno revela a similaridade das conexões de um nó em janelas de tempo distintas, permitindo revelar o seu grau de mobilidade. Já a *norma* captura o quão importantes essas conexões são. Um nó pode mudar muito as suas conexões e, simultaneamente, não ser importante para o roteamento a depender da aplicação.

Por fim, discutimos a parametrização de *DynamicNode2Vec*. Os primeiros parâmetros a serem definidos são  $nw$  e  $wl$  que correspondem, respectivamente, ao número de passeios aleatórios por nó e o número de passos que cada um dá. Em nossos experimentos, observamos que 4 passeios por nó e 8 passos por passeio, são suficientes para realizar o processo de amostragem em nosso estudo de caso. O aumento do número de passeio e a quantidade de caminhos que cada um faz causou um aumento proporcional na co-ocorrência de nós nos caminhos amostrados, sem trazer mais informações.

Parâmetros como o número de dimensões dos *embeddings* ( $d$ ), *sobreaajuste* ( $\lambda$ ) e *alinhamento* ( $\tau$ ) da função de otimização são obtidos através de uma busca exaustiva realizada da seguinte maneira. Primeiro, uma faixa de valores foram definidos para cada um deles. Para  $\lambda$  e  $\tau$ , foram testados valores entre 0 e 100, variando a cada 5. Já para  $d$ , nós testamos valores entre 10 e 50, variando a cada 10. Embora isso resulte em um excesso de experimentos, os resultados tendem a convergir de forma rápida como mostrado em (YAO et al., 2018; FERREIRA et al., 2019).

Em seguida, nós identificamos comunidades no grafo e no espaço vetorial para cada janela de tempo analisada. No grafo, o algoritmo de Louvain explicado na Seção 3.2 é usado. Já para o espaço vetorial, *DynamicNode2Vec* aplica o algoritmo *K-means spherical* que considera a distância de cosseno para encontrar as comunidades. O parâmetro  $K$  do *K-means* é configurado conforme o número de comunidades encontradas pelo algoritmo de Louvain no grafo da respectiva janela de tempo. Na prática, nós observamos a estrutura de comunidade existente na rede e tentamos recuperá-la no espaço vetorial. Aquela parametrização que melhor recupera as comunidades é selecionada. Essa estimativa de qualidade do modelo é feita usando duas métricas:

- *Informação mútua normalizada (NMI)* (do inglês, Normalized Mutual Information): Medida baseada na entropia de Shannon (SHANNON, 2001) que permite avaliar o quanto dois dados conjuntos de nós e suas partições (comunidades) se parecem uma com a outra. O NMI varia entre 0 e 1, onde 0 indica que não há nenhuma semelhança entre as duas partições obtidas e 1 indica que elas são perfeitamente iguais. Dessa forma, nós computamos o NMI entre as comunidades identificados pelo algoritmo de Louvain na rede e as comunidades encontradas pelo *k-means* nos embeddings produzidos por nosso modelo para cada parametrização testada. O espaço vetorial mais representativo é aquele que melhor recupera os grupos originalmente definidos na rede, produzindo assim um resultado de NMI médio mais alto.

Porém, uma desvantagem dessa métrica é que ela não considera o tamanho das comunidades. Portanto, se os nós de uma comunidade muito pequena em uma dada partição são atribuídos a outras comunidades na outra partição em questão, a métrica tende a ser bastante penalizada mesmo que essa comunidade represente uma fração mínima de nós na rede. Outro problema do NMI é que se o grafo não possui estrutura de comunidade forte o suficiente, dificilmente elas serão recuperadas no espaço vetorial. Por isso, é necessária uma medida adicional descrita a seguir.

- *Correlação entre (NMI) e a Modularidade*: Visto que o NMI é bastante sensível a estrutura de comunidades na rede, nós então o correlacionamos com a modularidade do grafo ao longo do período analisado. Na prática, nós queremos entender, para além da média, como a presença de comunidades fortemente estruturadas em janelas



de tempo específicas resultam em um espaço vetorial mais representativo e, portanto, em um NMI alto. A correlação de Pearson foi usada (BENESTY et al., 2009). O coeficiente de Pearson varia entre -1 a 1, onde -1 indica uma correlação negativa perfeita entre as duas variáveis, 0 indica que as duas variáveis não são linearmente dependentes e, 1 indica que há uma correlação positiva perfeita entre duas variáveis. Dessa forma, mesmo que o NMI médio não seja alto pelo fato da rede apresentar em algumas janelas de tempos específicas comunidades fracamente estruturadas, nós podemos observar o quanto isso interfere no NMI. Se a correlação entre as duas métricas for alta, isso indica que quando a estrutura de comunidade é forte (modularidade alta), o NMI também é alto no espaço vetorial.

Por meio desse processo, nós encontramos a seguinte parametrização:  $\lambda = 50$ ,  $\tau = 15$  e  $d = 50$ . Isso produziu um NMI médio de 0.46 com desvio padrão de 0.17. Já a correlação entre o NMI e a Modularidade para as janelas de tempo observadas resultou em um coeficiente de Pearson de 0.91 significativo para um  $p - value = 0.001$ . Isso mostra que para janelas de tempos onde a estrutura de comunidade é forte, é possível recuperar as comunidades de forma satisfatória. Já em janelas de tempo sem uma estrutura de comunidade bem definida na rede, naturalmente também não é possível encontrá-las no espaço vetorial, por questão de sobreposição ou aleatoriedade das conexões dos nós. Isso é exemplificado pelas Figuras 5a e 5b.

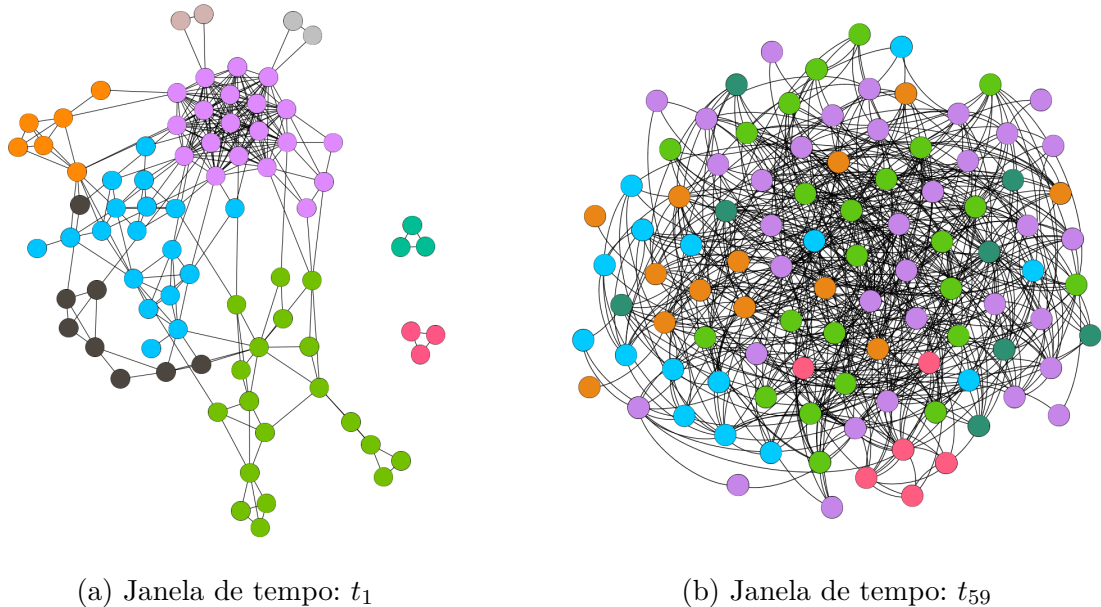


Figura 5 – Exemplo da diversidade estrutural dos grafos de mobilidade obtidos.

Note que o grafo da janela de tempo  $t_1$  revela uma estrutura de comunidades bem definida segundo o algoritmo de Louvain, cuja modularidade encontrada é de 0.51. Para essa janela de tempo específica, ao comparar essas comunidades com aquelas recuperadas



no espaço vetorial equivalente pelo *K-means*, obtemos como retorno um NMI 0.76. Dada que a medida não é ajustada para o tamanho das comunidades, este número é considerado de moderado a alto. Por outro lado, a janela de tempo  $t_{59}$  exemplifica uma rede com uma estrutura de comunidade mais fraca, cuja modularidade retornada pelo algoritmo de Louvain é de 0.23. Isso produz um NMI de 0.29. Dessa forma, podemos observar como a média do NMI é afetada pela estrutura das redes observadas e a importância de correlacionar ambas as medidas para garantir que quando possível, o modelo consiga representar os nós preservando as propriedades originais do grafo.

Uma vez definida a metodologia deste trabalho, nos movemos para o próximo capítulo para investigarmos possíveis padrões de mobilidades no cenário definido.

## 4 Resultados

Neste capítulo, nós apresentamos os principais resultados do nosso trabalho. Primeiro fornecemos uma caracterização das redes de contato modeladas na seção 4.1. Em seguida ilustramos na seção 4.2 como a nossa abordagem consegue incorporar tais redes considerando a mobilidade dos usuários sob a perspectiva espaço-tempo. Por fim, nós analisamos, na seção 4.3, a dinâmica e a importância dos usuários em termos de mobilidade.

### 4.1 Caracterização das redes de contatos

Com objetivo de entender como as redes de contato estão estruturadas, nós começamos analisando algumas propriedades topológicas. Primeiro, nós olhamos para as medidas topológicas mais gerais, cuja as distribuições para todas as janelas de tempo observadas são apresentadas na Figura 6. Para facilitar a visualização usamos *boxplots*. O traço preto central representa a mediana e as caixas vão do 1º ao 3º quartil. Os *whiskers* representam o 5<sup>th</sup> e o 95<sup>th</sup> percentis, respectivamente.

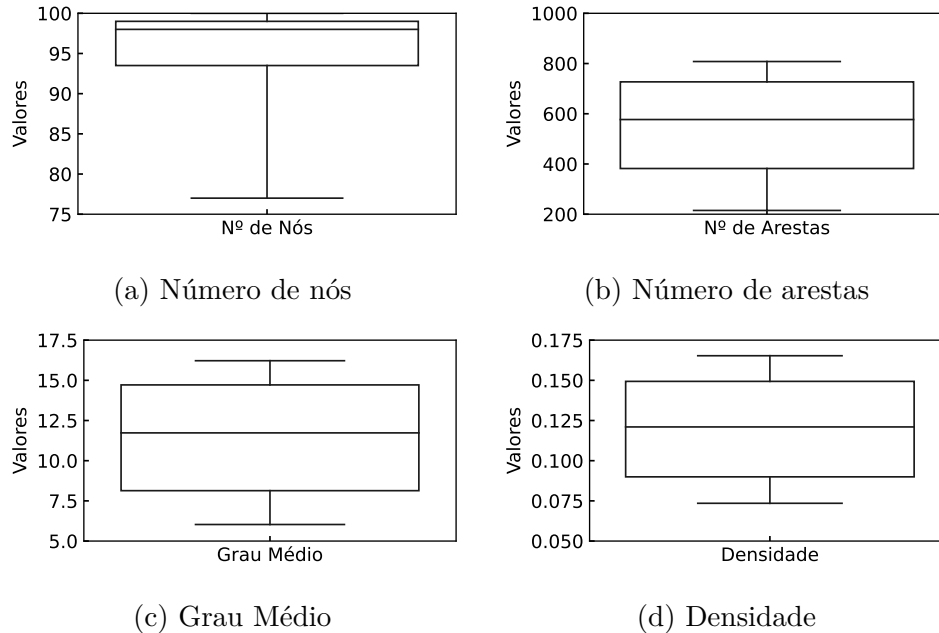


Figura 6 – Distribuição das medidas topológicas gerais das redes modeladas.

A Figura 6a mostra que apesar de considerarmos um cenário contendo 100 usuários, nem sempre estes usuários, representados pelos nós na rede, estabeleceram conexões quando observados em janelas de tempo diárias. Em 25% do período observado, no máximo 93 dos nós (ou usuários) realizaram ou tiveram a oportunidade de realizar algum contato. Note que este percentual exclui outliers. No entanto, quando observada a distribuição de

forma geral, nós podemos notar que em 50% do período analisado, pelo menos 98 dos nós, apresentaram potencial para troca de mensagens. Isso indica que grande parte deles interagiram ou tiveram a oportunidade de interagir na maior parte do período observado. Já a forma com que essas potenciais interações ocorreram é bastante dinâmica. O número de arestas criadas estabelecidas varia (Figura 6b), consequentemente, o grau médio (Figura 6c) e a densidade (6d) da rede também, revelando assim uma alta heterogeneidade em termos de conectividade.

Movendo para as medidas que explicam como os nós estão agrupados e como estes agrupamentos se dividem na rede, a Figura 7a mostra a distribuição do coeficiente de agrupamento médio e a *Modularidade* do período observado. De forma geral, podemos notar uma assimetria maior na distribuição do coeficiente de agrupamento médio. A mediana de 0,41 mostra que em 50% do período considerado, esses nós na média não estão fortemente conectados, o que reduz as opções de trocas de mensagens. Isso reflete também na distribuição da *Modularidade*. Relembre que esta medida é obtida por meio do Algoritmo de Louvain (explicado na seção 3.2) e está relacionada a qualidade das comunidades formadas por essas conexões. Nós podemos notar que em 50% dos casos, a estrutura de comunidades é fraca, revelando que a fraca conectividade entre os nós na rede não permite dividi-los em comunidades fortemente estruturadas, o que também prejudica aplicações voltadas para roteamento e disseminação de mensagens de forma rápida e eficiente na rede. Por fim, nós analisamos a distribuição do número de comunidades encontradas na Figura 7b, que, em geral, é pequeno e varia pouco.

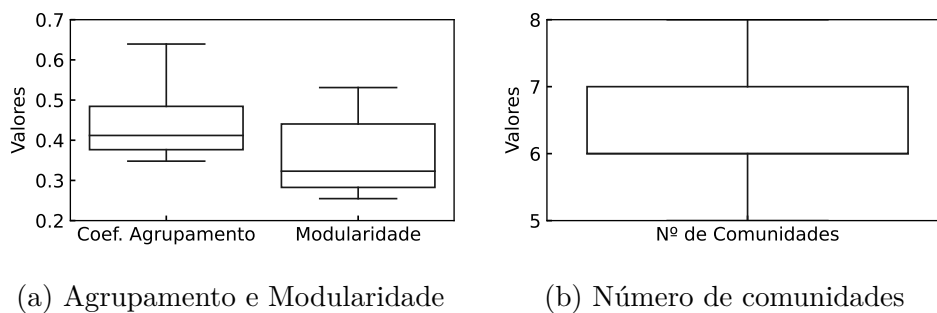


Figura 7 – Distribuição das medidas topológicas relacionadas ao agrupamento durante o período observado.

Em suma, nós podemos observar que as redes de mobilidade modeladas aqui são bastante heterogêneas em termos de estrutura e organização. Em alguns dos períodos observados, os nós estão mais conectados indicados por um grau médio mais alto, coeficiente de agrupamento alto, e comunidades bem definidas. No entanto, na outra parte do período essa estrutura mais consistente não é observada, o que isso impõe desafios no contexto de mobilidade devida a tentativa de encontrar nós estratégicos que facilitem a disseminação das mensagens por toda a rede.

## 4.2 Representação do modelo espaço-temporal

Como explicado na seção 3.3, nós geramos um espaço vetorial de baixa dimensão preservando a dimensão espaço-temporal através de sequência de redes modeladas usando o *DynamicNode2Vec*. Para ilustrar como nossa proposta funciona, nós usamos às seis primeiras janelas de tempo como exemplo antes de investigar os aspectos de mobilidade de forma geral.

Relembre ainda, que a parametrização adotada pelo *DynamicNode2Vec* produziu uma representação dos nós em um espaço de 50 dimensões. Para observar a disposição dos nós em um espaço vetorial de duas dimensões e exemplificar a sua mobilidade, nós usamos o t-SNE, uma técnica de redução de dimensionalidade para exploração e visualização de dados (MAATEN; HINTON, 2008). t-SNE consegue preservar a estrutura local e global de pontos em um dado espaço vetorial qualquer de n-dimensões em poucas dimensões. Dessa forma, pontos que estão próximos uns dos outros em um espaço de alta dimensão, tendem a estar próximos uns dos outros no espaço de mais baixa dimensão representado.

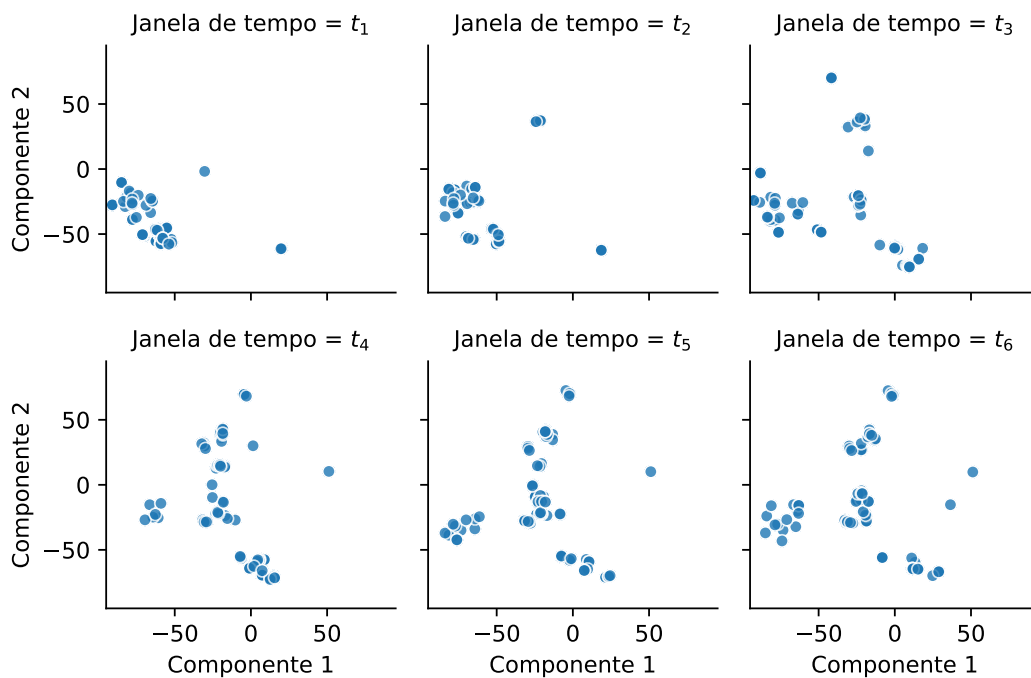


Figura 8 – Representação 2-D do espaço vetorial obtido.

A Figura 8 mostra a representação em duas dimensões para todos os nós nas seis primeiras janelas de tempo analisadas. Cada ponto representa um nó e sua localização no espaço vetorial conforme o dia. Observando na primeira janela de tempo  $t_1$ , note que muitos nós estão sobrepostos e próximos no espaço representado. Essa disposição indica que eles também estão próximos na rede — em termos de conectividade — e revela a importância da conectividade de uns com os outros. Ao passar do tempo, alguns se distanciam mais e

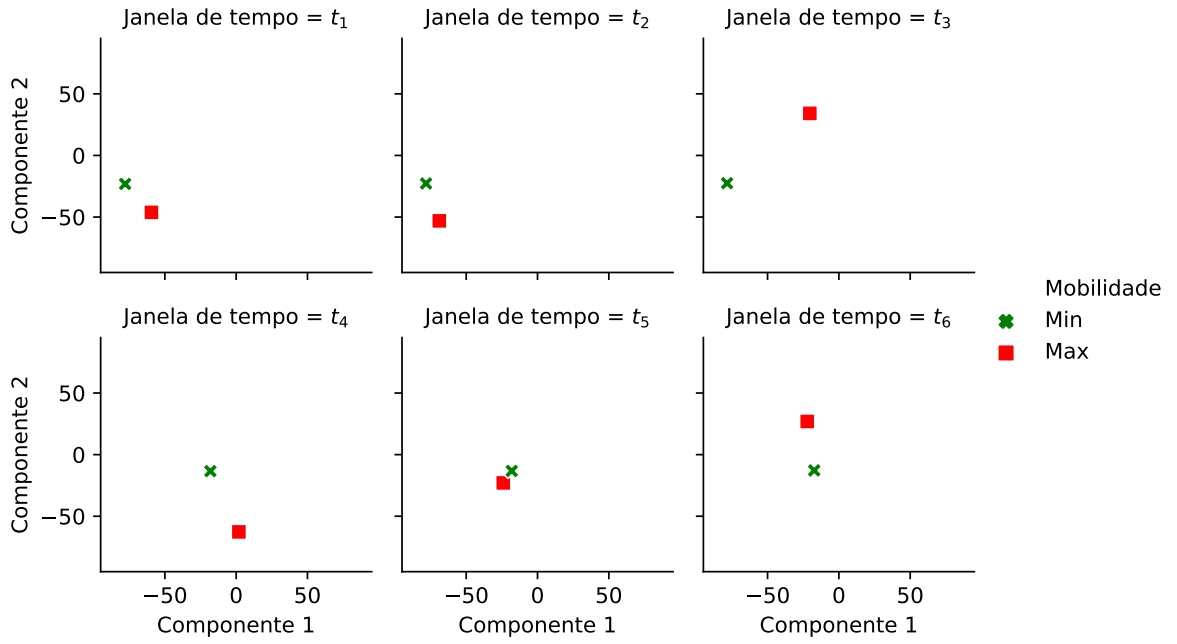


Figura 9 – Representação 2-D do espaço vetorial obtido. Apenas os nós com a mobilidade máxima e mínima no período observado são representados.

gradualmente vão preenchendo outras regiões da representação obtida. Este processo está diretamente relacionado a dinâmica das conexões estabelecidas nas redes modeladas.

Para exemplificar como a dinâmica dos nós pode ser capturada neste espaço estabelecido, nós selecionamos dois nós que, durante às seis janelas de tempo mostradas, apresentaram a maior e a menor mobilidade média dia após dia. Relembre que podemos capturar essa dinâmica por meio da distância de cosseno, como explicado na seção 3.3. Portanto, nós computamos a distância de cosseno dia após dia para todos os nós e, em seguida, computamos a média dessas distâncias para cada um deles. O nó com menor mobilidade média neste período apresentou uma distância de cosseno média de 0.10. Já o de maior mobilidade, apresentou uma distância de cosseno média de 0.43.

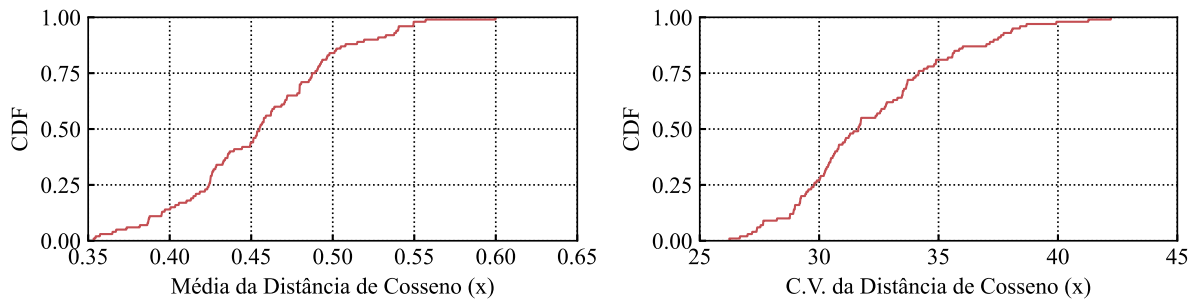
Usando a mesma representação t-SNE obtida anteriormente para efeito de consistência mostramos na Figura 9 como estes nós se deslocam no período exemplificado. Focando no nó que apresentou a menor média (min) podemos observar nas três primeiras janelas de tempo uma certa estabilidade visto que ele se encontra praticamente na mesma posição se observarmos os eixos x e y das três figuras. Entre as janelas de tempo  $t_3$  e  $t_4$  um deslocamento é observado indicando uma mudança de conectividade deste nó. Por fim, uma estabilidade até a janela  $t_6$  é novamente vista. Por outro lado, ao analisarmos o nó com a maior mobilidade média (max) nessas seis primeiras janelas de tempo, é possível notar uma grande variabilidade dia após dia. Dessa forma podemos distinguir os nós com diferentes níveis de mobilidade e, caso necessário, separá-los em períodos específicos conforme a necessidade das aplicações.

Após ilustrarmos o modelo espaço-temporal proposto, nós aprofundamos a nossa análise para todo o período aqui estudado buscando uma compreensão geral da dinâmica dos usuários na próxima seção.

### 4.3 Compreensão da dinâmica dos usuários

Com intuito de compreender os padrões de mobilidade no cenário estudado, nós iniciamos investigando o nível de mobilidade dos usuários em todo o período observado. Relembre que possíveis aplicações da nossa abordagem são decorrentes das redes oportunísticas ou tolerantes a atrasos. Portanto, nós analisamos o nível de mudança dos usuários nas redes de contato sob a perspectiva de que as novas conexões - representadas por mudanças topológicas na rede - podem ocorrer com um certo atraso. Especificamente, nós calculamos a distância de cosseno de uma janela de tempo  $t_i$  em relação a todas as janelas de tempo subsequentes  $t_j$ , tal que,  $i < j$ . Assim, nós estamos avaliando o grau de mobilidade e o potencial que um nó tem de estabelecer novas conexões não somente de um dia para o outro, mas também permitindo que estes aspectos sejam capturados considerando o atraso, ou seja, em janelas de tempo distantes.

A Figura 10a mostra a função de distribuição cumulativa (CDF) da distância de cosseno média dos nós na rede. Se olharmos para os 10% dos nós mais estáveis em termos de conectividade na rede, temos uma distância de cosseno média de no máximo 0.38. Por outro lado, se olharmos os 10% dos nós que mais mudam, temos uma distância de cosseno média superior a 0.54, o que representa um nível de mobilidade 42% maior.

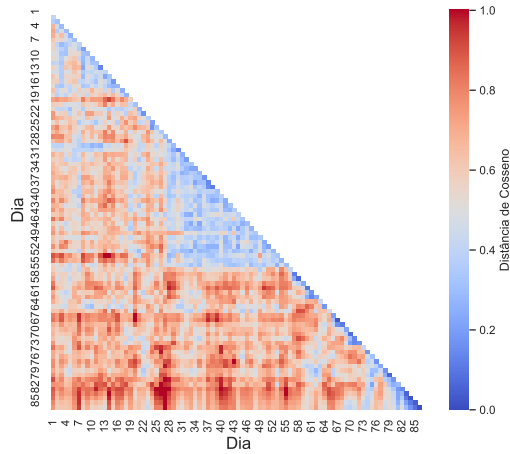


(a) Distribuição da média da distância de cosseno (b) Distribuição do C.V. da distância de cosseno

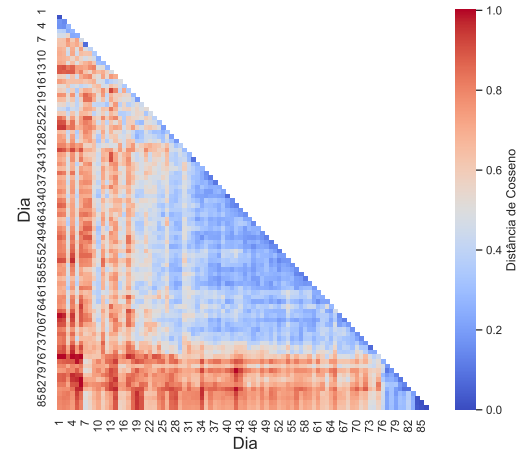
Figura 10 – CDF da média e do coeficiente de variação da distância de cosseno.

Além da média também observamos o coeficiente de variação (C.V.) como medida de dispersão, que corresponde ao desvio padrão expresso como uma porcentagem média<sup>1</sup>. Assim podemos observar se o nível de mobilidade dos nós é constante ou decorre de rajadas em períodos específicos. Os valores do C.V. são apresentados na Figura 10b. Valores

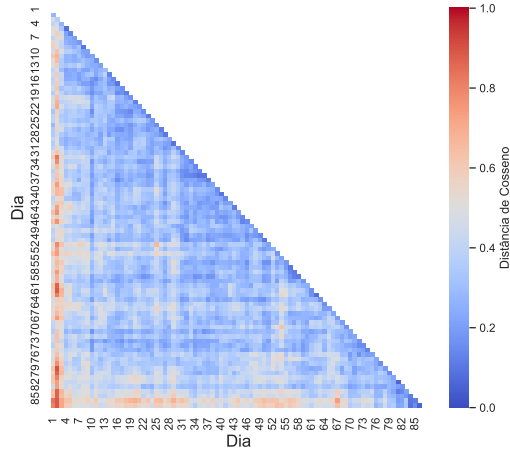
<sup>1</sup> O C.V. é computado pela razão entre o desvio padrão e a média, por fim multiplicado por cem.



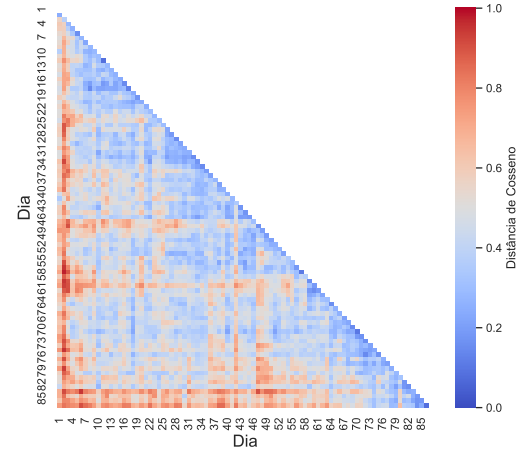
(a) Nó com maior Média



(b) Nó com maior C.V.



(c) Nó com menor média



(d) Nó com menor C.V.

Figura 11 – Mapas de calor da distância de cosseno de um nó.

maiores do que 30, na prática, são considerados altos coeficientes de variação. Portanto, é possível observar que mais de 75% dos nós na rede apresentam as maiores mudanças em períodos específicos.

Para investigar como tais mudanças ocorrem selecionamos os nós com maior e menor valor para ambas as métricas (Média e C.V.) em relação à distância de cosseno. Em seguida, nós construímos mapas de calor para mostrar como essas mudanças estão distribuídas ao longo do período observado de acordo com esses critérios. As Figuras 11a e 11b mostram os mapas de calor para os nós com maior média e maior C.V., respectivamente. Os eixos X e Y indicam os dias em que a distância de cosseno foi computada. Já a cor indica o nível de mudança, revelando o grau de dispersão delas no período observado. Quanto mais vermelho e intenso o ponto, maior é a diferença das conexões do nó em relação aos dois dias. Por outro lado, quanto mais azul e intenso, menor é essa diferença.

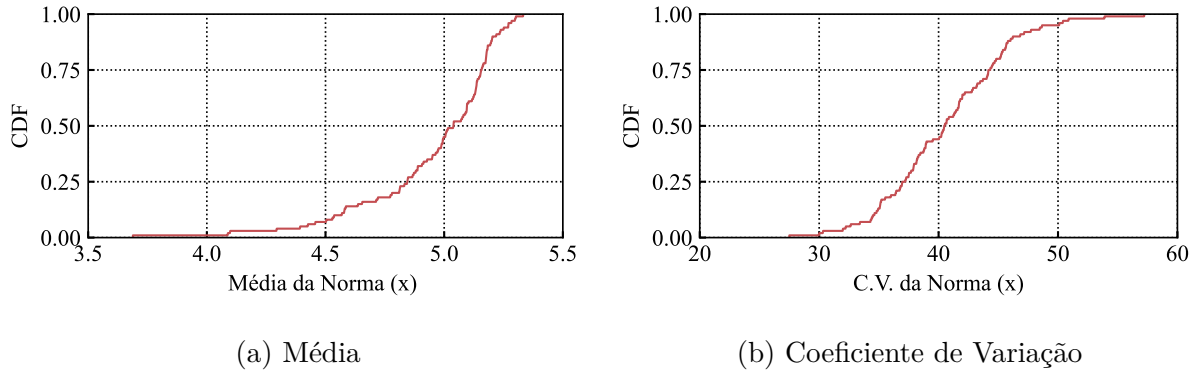


Figura 12 – CDF da média e do coeficiente de variação da norma dos nós.

Em geral, é possível notar que o nó com maior distância de cosseno média, Figura 11a, apresenta um nível de mobilidade alto para a maioria dos pares de dias observados. Altos valores da distância de cosseno são observados de forma dispersa, indicando que este nó é bastante dinâmico e está sempre renovando suas conexões. Por outro lado, o nó com maior C.V. (Figura 11b) concentra suas mudanças no início e no fim do período analisado, como mostra a região avermelhada no mapa de calor. Isso explica o alto coeficiente de variação. Já os nós mais estáveis, com menor média e menor C.V. da distância de cosseno, apresentam de forma geral poucas mudanças como mostram as Figuras 11c e 11d.

A distância de cosseno fornece um indicativo da mobilidade dos nós em termos de mudança de conectividade. No entanto, não podemos afirmar que um nó que muda muito as suas conexões necessariamente forneça as conexões mais importantes para que uma mensagem alcance toda a rede. No contexto de redes oportunísticas, isso envolve propriedades sociais. Por exemplo, algumas pessoas se relacionam com várias outras pessoas, mas essas várias pessoas também podem se relacionar com outras, e assim por diante. Neste caso, tais grupos podem apresentar uma dinamicidade muito alta, possibilitando o envio e recebimento de mensagens através de diferentes conexões. Note que os nós da rede neste caso, tendem a ter uma distância de cosseno alta. Em outro cenário, um grupo de pessoas pode interagir muito entre si e pouco com grupos externos. Ainda, essa pouca interação pode ser estável (sempre as mesmas pessoas) e ser realizada por membros específicos. Isso faz com que a outra parte deste grupo precise exclusivamente dos membros com conexões externas para receber ou enviar mensagens. Em cenário como este, essas *pontes* não apresentam um nível de mobilidade alto, já que as conexões externas são sempre as mesmas. Portanto, eles tendem a ter uma distância de cosseno baixa, mas apresentam conexões fundamentais para que uma mensagem chegue a uma determinada região. Por este motivo avaliamos a norma dos vetores obtidos para os nós durante o período analisado, que como explicado na seção 3.3, captura a importância de um nó em termos de conectividade.

A Figura 12a mostra a distribuição da média da norma dos vetores dos nós. Os 10% com menor valor apresentam uma norma média menor ou igual a 4.45. Já os 10% com



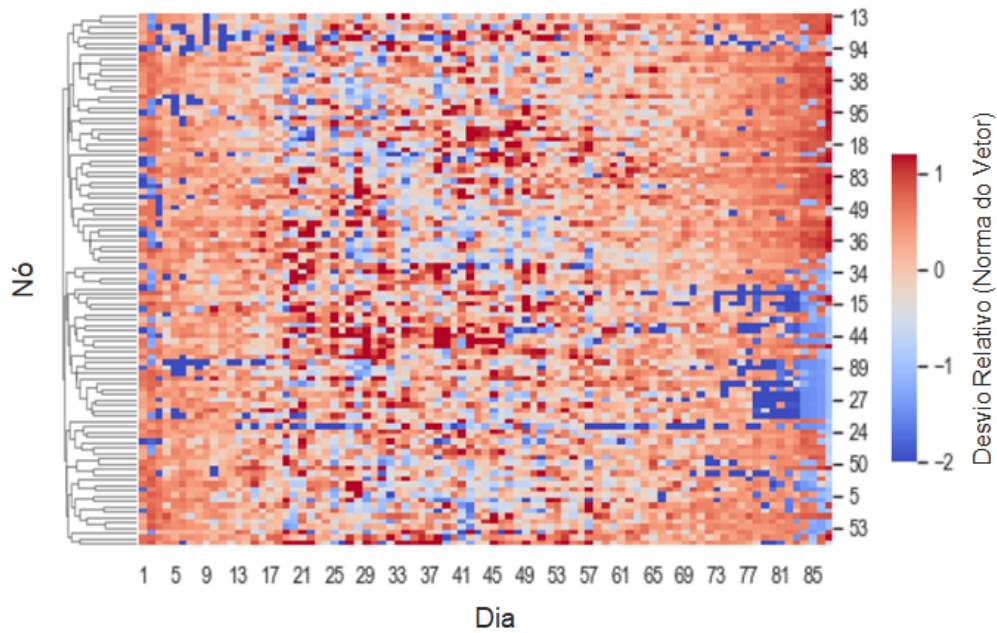


Figura 13 – Mapa de calor da norma dos nós ao longo do período observado.

maior valor médio apresentam uma norma média maior ou igual a 5.3, resultando em uma diferença de quase 20%. Em termos absolutos essa diferença não é muita, mas, na prática pode corresponder ao potencial de uma mensagem chegar ou não em uma determinada região da rede. Nós com maior norma são mais importantes e centrais em termos de conectividade. Novamente, como medida de dispersão em relação à média, avaliamos o coeficiente de variação das normas obtidas por nó e apresentamos a distribuição na Figura 12b. Diferente do cosseno, 98% dos nós apresentam um C.V. alto (maior do que 30). Isso nos fornece indícios de que alguns nós são bastante importantes para a troca de informação na rede e essa importância tende a ocorrer em momentos específicos.

Com o propósito de entender a importância da conectividade dos nós considerando a dimensão temporal, criamos um mapa de calor com a norma de cada nó para cada dia do período analisado, apresentado na Figura 13. As linhas representam os nós e as colunas os dias. Cada coluna (dia) foi normalizada utilizando o *escore z* (do inglês, *z-score*). Assim, cada valor é subtraído pela média da coluna e dividido pelo desvio padrão da coluna. Dessa forma nós podemos observar o quanto a importância do nó desvia relativamente da importância média de todos os nós na rede para cada dia. A cor de cada célula reflete o valor da norma normalizado. Por fim, para medir a semelhança dos nós em termos de importância, nós computamos um dendrograma sob as linhas, agrupando os nós que possuem importância similar em dias específicos. Como métrica de similaridade para o dendrograma, foi usada a correlação de Pearson entre as normas dos nós.

De maneira geral podemos observar que alguns nós apresentam uma importância maior do que os demais (indicada por pontos vermelhos) em períodos específicos, confirmando a nossa suspeita a cerca do alto C.V. da norma. Ademais, o dendrograma

computado revela a existência de um subconjunto de nós com importância similar em dias específicos.

Até o momento observamos que a distância de cosseno e a norma fornecem análises complementares. Nós então investigamos como essas medidas se relacionam. Para isso mensuramos a correlação de Pearson de forma pareada entre a média e o C.V. de ambas as medidas. Empiricamente, nós buscamos entender o quanto uma dessas medidas permite inferir a outra numa tentativa de sumarizar a interpretação de ambas.

A Tabela 2 sumariza a correlação entre elas. A correlação mais forte é negativa e ocorre entre a média da norma e o C.V. da distância de cosseno. Isso sugere que nós com conexões mais importantes apresentam um padrão de mobilidade estável. Em seguida, existe uma correlação moderada/alta entre a média da distância de cosseno e o C.V. da norma. Logo, nós que consistentemente mudam muito as suas conexões, tendem a manter conexões às vezes importante e às vezes não. As demais correlações não são significativas para  $p = 0.01$ .

Tabela 2 – Correlação entre a norma e o C.V..

Métricas		Correlação
Média da Dist. de Cosseno	C.V. da Dist. de Cosseno	-0.04
Média da Dist. de Cosseno	C.V. da Norma	0.70
Média da Norma	C.V. da Dist. de Cosseno	-0.95
Média da Norma	C.V. da Norma	-0.10

Finalmente, nós encerramos a nossa análise buscando responder se medidas topológicas do grafo, tipicamente usadas em estratégias para redes móveis (DALY; HAAHR, 2007; WAN et al., 2017; ZHANG et al., 2017; NUNES et al., 2018; GUO; DONG; WEN, 2021), conseguem capturar os padrões que revelamos aqui por meio do espaço latente construído. Para isso, nós consideramos as medidas topológicas discutidas na seção 3.2, que capturam diferentes noções de centralidade e de conectividade dos nós na rede, portanto, a noção de importância. Em seguida computamos a correlação pareada entre essas medidas, apresentada na Tabela 3.

Focando nas medidas de centralidade (Grau, Intermediação, Proximidade e Auto-vetor), é possível observar uma correlação negativa moderada com a média da distância de cosseno. Isso sugere que quanto mais dinâmico um nó é, menor é a sua centralidade de acordo com diferentes visões. Já em relação à norma, as correlações encontradas com as medidas topológicas são positivas, porém baixa. Isso acontece porque medidas de centralidade, tipicamente, capturam noções de centralidade e importância exclusivas (VALENTE et al., 2008). Portanto, elas não possuem a capacidade de generalizar um conceito de centralidade que seja universalmente importante no grafo. Por fim, nós consideramos o coeficiente de agrupamento, que também apresenta uma correlação baixa/moderada com

Tabela 3 – Correlações entre as medidas topológicas e do *embedding*.

Métricas no <i>Embedding</i>	Métricas Topológicas	Correlação
Média da D. de Cosseno	Média do Grau	-0.64
Média da D. de Cosseno	Média do Coef. de Intermediação	-0.69
Média da D. de Cosseno	Média da Proximidade	-0.64
Média da D. de Cosseno	Média do Autovetor	-0.65
Média da D. de Cosseno	Média do Coef. de Agrupamento	0.51
Média da Norma	Média do Grau	0.33
Média da Norma	Média do Coef. de Intermediação	0.50
Média da Norma	Média da Proximidade	0.28
Média da Norma	Média do Autovetor	0.32
Média da Norma	Média do Coef. de Agrupamento	-0.55

ambas as métricas dos nós no espaço latente. Em relação à distância de cosseno, essa correlação sugere que nós que apresentam conectividade com os vizinhos, (alto coeficiente) tendem a mudar as conexões. Por outro lado, nós menos conectados em relação aos seus vizinhos, exibem conexões muito importantes na rede.

## 5 Conclusão

Neste trabalho, nós propomos modelar e analisar a mobilidade de entidades móveis usando técnicas de *node embedding*. Especificamente, nós propomos a utilização de uma técnica já existente na literatura (*DynamicNode2Vec*) com objetivo de revelar padrões de mobilidade. Como estudo de caso usamos um rastro de mobilidade oriundo de um modelo de geração de rastros sintéticos com propriedades de mobilidade presentes em cenários reais. Ao caracterizarmos a estrutura topológica das redes de contato modeladas, nós observamos propriedades bastante dinâmicas e heterogêneas. Embora isso torne o treinamento dos *embeddings* um desafio, nossa proposta conseguiu representar tais redes de maneira satisfatória, capturando a dinâmica e a importância dos nós em termos de mobilidade e conectividade na rede.

Nossas principais análises no estudo de caso considerado revelaram diferentes padrões de mobilidade dos nós. Particularmente, ao obter representações da mobilidade dos usuários observamos que, alguns preservam suas conexões ao longo do tempo enquanto outros as mudam. Ainda, ao analisarmos a variação com que essas mudanças ocorrem, nós observamos que elas podem ser constantes ou esporádicas. Em seguida, investigamos a importância das conexões estabelecidas pelos nós para a conectividade da rede, que pode ser fundamental para uma tarefa de disseminação/roteamento de mensagens. Similar ao que ocorre com a mobilidade, observamos grupos de nós com diferentes níveis de importância em termos de conectividade durante o período observado. No entanto, nós notamos uma variação ainda maior nessa medida, indicando que estes grupos tendem a ser importantes para a conectividade em determinados momentos.

Além disso, nós também analisamos como o nível de mudança das conexões dos nós se relaciona com a importância dessas conexões para a conectividade da rede. Para o cenário estudado, os resultados sugerem que nós mais estáveis tendem a apresentar conexões mais importantes na maior parte do tempo, enquanto nós mais dinâmicos apresentam uma maior variabilidade na importância dessas conexões. Finalmente, contrastamos esses padrões obtidos com típicas medidas topológicas da rede. De maneira geral, observamos que essas medidas não conseguem generalizar a noção de importância de conectividade capturada por nossa metodologia, evidenciando as contribuições obtidas.

Como trabalhos futuros, nossa metodologia abre uma série de caminhos para estudos adicionais. Uma possibilidade imediata é investigar a sensibilidade de alguns parâmetros como, por exemplo, o tamanho da janela de tempo considerada na modelagem da rede de contato. Na mesma direção, analisar outros rastros com diferentes propriedades. Além disso, existe a possibilidade de nossa metodologia ser incorporada em aplicações

que demandam uma disseminação e roteamento efetivo de mensagens bem como em soluções para cidades inteligentes. Por fim, outra possibilidade interessante consiste em estender a modelagem da rede para incluir informações contextuais dos locais de visita dos nós. Isso também viabiliza que novas tarefas sejam realizadas, incluindo predição e recomendação de pontos de interesses aos usuários.

# Referências

- AGRAWAL, S.; ROY, D.; MITRA, M. Tag embedding based personalized point of interest recommendation system. *Information Processing & Management*, Elsevier, v. 58, n. 6, p. 102690, 2021. Citado 3 vezes nas páginas 13, 17 e 19.
- ALAJEELY, M.; DOSS, R.; AHMAD, A. Routing protocols in opportunistic networks—a survey. *IETE Technical Review*, Taylor & Francis, v. 35, n. 4, p. 369–387, 2018. Citado na página 16.
- BARABÁSI, A.-L. et al. *Network Science*. [S.l.]: Cambridge University Press, 2016. Citado 3 vezes nas páginas 12, 13 e 24.
- BATABYAL, S.; BHAUMIK, P. Mobility models, traces and impact of mobility on opportunistic routing algorithms: A survey. *IEEE Communications Surveys & Tutorials*, IEEE, v. 17, n. 3, p. 1679–1707, 2015. Citado na página 16.
- BENESTY, J. et al. Pearson correlation coefficient. In: *Noise reduction in speech processing*. [S.l.]: Springer, 2009. p. 1–4. Citado na página 31.
- BLONDEL, V. D. et al. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, IOP Publishing, v. 2008, n. 10, p. P10008, oct 2008. Citado na página 25.
- BRANDES, U.; GAERTLER, M.; WAGNER, D. Experiments on graph clustering algorithms. In: SPRINGER. *European Symposium on Algorithms*. [S.l.], 2003. p. 568–579. Citado na página 22.
- CAI, Y. et al. A survey on routing algorithms for opportunistic mobile social networks. *China Communications*, IEEE, v. 18, n. 2, p. 86–109, 2021. Citado 2 vezes nas páginas 12 e 19.
- CELES, C. et al. Big data analytics no projeto de redes móveis: Modelos. *Protocolos e Aplicações*, p. 1–58, 2017. Citado na página 12.
- DALY, E. M.; HAAHR, M. Social network analysis for routing in disconnected delay-tolerant manets. In: *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*. [S.l.: s.n.], 2007. p. 32–40. Citado 4 vezes nas páginas 13, 23, 24 e 41.
- FERREIRA, C. H. G. et al. Modeling dynamic ideological behavior in political networks. *The Journal of Web Science*, v. 7, 2019. Citado 4 vezes nas páginas 14, 25, 28 e 30.
- FORTUNATO, S.; HRIC, D. Community detection in networks: A user guide. *Physics reports*, Elsevier, v. 659, p. 1–44, 2016. Citado na página 25.
- GOLBECK, J. *Analyzing the social web*. [S.l.]: Newnes, 2013. Citado na página 24.
- GOYAL, P.; FERRARA, E. Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems*, Elsevier, v. 151, p. 78–94, 2018. Citado na página 13.

- GROVER, A.; LESKOVEC, J. node2vec: Scalable feature learning for networks. In: *ACM. Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. [S.l.], 2016. p. 855–864. Citado 3 vezes nas páginas 18, 25 e 26.
- GUO, X.; DONG, C.; WEN, W. Dynamic computation offloading strategy with dnn partitioning in d2d multi-hop networks. In: *2021 9th International Conference on Communications and Broadband Networking*. [S.l.: s.n.], 2021. p. 172–178. Citado 4 vezes nas páginas 13, 23, 24 e 41.
- HANSEN, D. et al. *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. [S.l.]: Morgan Kaufmann, 2019. Citado na página 24.
- HESS, A. et al. Data-driven human mobility modeling: a survey and engineering guidance for mobile networking. *ACM Computing Surveys (CSUR)*, ACM New York, NY, USA, v. 48, n. 3, p. 1–39, 2015. Citado na página 16.
- HESS, A. et al. Data-driven human mobility modeling: a survey and engineering guidance for mobile networking. *ACM Computing Surveys (CSUR)*, ACM, v. 48, n. 3, p. 38, 2016. Citado na página 12.
- JOLLIFFE, I. T.; CADIMA, J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, The Royal Society Publishing, v. 374, n. 2065, p. 20150202, 2016. Citado na página 27.
- LU, Y. et al. Temporal network embedding with micro-and macro-dynamics. In: *International Conference on Information and Knowledge Management*. [S.l.: s.n.], 2019. Citado na página 13.
- MAATEN, L. Van der; HINTON, G. Visualizing data using t-sne. *Journal of machine learning research*, v. 9, n. 11, 2008. Citado 2 vezes nas páginas 27 e 35.
- MOTA, V. F. et al. Protocols, mobility models and tools in opportunistic networks: A survey. *Computer Communications*, Elsevier, v. 48, p. 5–19, 2014. Citado na página 21.
- NEWMAN, M. E. A measure of betweenness centrality based on random walks. *Social networks*, Elsevier, v. 27, n. 1, p. 39–54, 2005. Citado na página 24.
- NUNES, I. O. et al. Combining spatial and social awareness in d2d opportunistic routing. *IEEE Communications Magazine*, IEEE, v. 56, n. 1, p. 128–135, 2018. Citado 4 vezes nas páginas 13, 23, 24 e 41.
- NUNES, I. O. et al. Grm: Group regularity mobility model. In: *ACM. Proceedings of the 20th ACM International Conference on Modelling, Analysis and Simulation of Wireless and Mobile Systems*. [S.l.], 2017. p. 85–89. Citado 6 vezes nas páginas 10, 12, 13, 19, 21 e 22.
- OLIVEIRA, A.; VAZAO, T. Low-power and lossy networks under mobility: A survey. *Computer networks*, Elsevier, v. 107, p. 339–352, 2016. Citado na página 13.
- RAHMANI, H. A. et al. Category-aware location embedding for point-of-interest recommendation. In: *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*. [S.l.: s.n.], 2019. p. 173–176. Citado 3 vezes nas páginas 13, 17 e 19.

- REIS, G.; SANTOS, B.; VIEIRA, L. MobETX: uma métrica de roteamento ciente da mobilidade para Internet das Coisas. 2021. Citado na página 16.
- ROSSETTI, G.; CAZABET, R. Community discovery in dynamic networks: a survey. *ACM Computing Surveys (CSUR)*, ACM, v. 51, n. 2, p. 35, 2018. Citado 2 vezes nas páginas 12 e 23.
- SANTOS, B. P. et al. Mobile matrix: routing under mobility in iot, iomt, and social iot. *Ad Hoc Networks*, Elsevier, v. 78, p. 84–98, 2018. Citado na página 16.
- SANTOS, B. P. et al. A internet das coisas está preparada para mobilidade? In: . [s.n.], 2021. Disponível em: <<https://www.sbrc2021.facom.ufu.br/pt-br/programacao/workshops/v-workshop-de-computacao-urbana-courb-2021>>. Citado na página 21.
- SANTOS, B. P.; VIEIRA, L. F. M.; LOUREIRO, A. A. F. Routing and Mobility Management in the Internet of Things. In: *XXXVIII SBRC*. [S.l.: s.n.], 2020. Citado 2 vezes nas páginas 13 e 21.
- SEGUNDO, F. R. et al. Roteamento em redes tolerantes a atrasos e interrupções: uma abordagem baseada em redes neurais. 2015. Citado na página 15.
- SHANNON, C. E. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, ACM New York, NY, USA, v. 5, n. 1, p. 3–55, 2001. Citado na página 30.
- SOBIN, C. et al. A survey of routing and data dissemination in delay tolerant networks. *Journal of Network and Computer Applications*, Elsevier, v. 67, p. 128–146, 2016. Citado na página 13.
- SYED, A. S. et al. Iot in smart cities: A survey of technologies, practices and challenges. *Smart Cities*, Multidisciplinary Digital Publishing Institute, v. 4, n. 2, p. 429–475, 2021. Citado 2 vezes nas páginas 12 e 19.
- VALENTE, T. W. et al. How correlated are network centrality measures? *Connections (Toronto, Ont.)*, NIH Public Access, v. 28, n. 1, p. 16, 2008. Citado 2 vezes nas páginas 13 e 41.
- VASTARDIS, N.; YANG, K. Mobile social networks: Architectures, social properties, and key research challenges. *IEEE Communications Surveys & Tutorials*, IEEE, v. 15, n. 3, 2013. Citado 2 vezes nas páginas 12 e 15.
- WAN, L. et al. A doa estimation approach for transmission performance guarantee in d2d communication. *Mobile Networks and Applications*, Springer, v. 22, n. 6, p. 998–1009, 2017. Citado 4 vezes nas páginas 13, 23, 24 e 41.
- WANG, P. et al. Learning urban community structures: A collective embedding perspective with periodic spatial-temporal mobility graphs. *ACM Transactions on Intelligent Systems and Technology (TIST)*, ACM New York, NY, USA, v. 9, n. 6, p. 1–28, 2018. Citado na página 18.
- WANG, Y. et al. A heterogeneous graph embedding framework for location-based social network analysis in smart cities. *IEEE Transactions on Industrial Informatics*, IEEE, v. 16, n. 4, p. 2747–2755, 2019. Citado 4 vezes nas páginas 13, 17, 18 e 19.



- XIA, F. et al. Pis: A multi-dimensional routing protocol for socially-aware networking. *IEEE Transactions on Mobile Computing*, IEEE, v. 15, n. 11, p. 2825–2836, 2016. Citado na página 15.
- XIA, F. et al. Socially aware networking: A survey. *IEEE Systems Journal*, IEEE, v. 9, n. 3, p. 904–921, 2013. Citado na página 15.
- XIE, M. et al. Graph-based metric embedding for next poi recommendation. In: SPRINGER. *International Conference on Web Information Systems Engineering*. [S.l.], 2016. p. 207–222. Citado na página 17.
- XU, F. et al. Sume: Semantic-enhanced urban mobility network embedding for user demographic inference. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, ACM New York, NY, USA, v. 4, n. 3, p. 1–25, 2020. Citado 3 vezes nas páginas 13, 18 e 19.
- YAAKOP, M. B. et al. Bluetooth 5.0 throughput comparison for internet of thing usability a survey. In: IEEE. *2017 International Conference on Engineering Technology and Technopreneurship (ICE2T)*. [S.l.], 2017. p. 1–6. Citado na página 23.
- YANG, D. et al. Revisiting user mobility and social relationships in lbsns: a hypergraph embedding approach. In: *The world wide web conference*. [S.l.: s.n.], 2019. p. 2147–2157. Citado 3 vezes nas páginas 13, 16 e 18.
- YANG, D. et al. Lbsn2vec++: Heterogeneous hypergraph embedding for location-based social networks. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, 2020. Citado 4 vezes nas páginas 13, 17, 18 e 19.
- YAO, Z. et al. Dynamic word embeddings for evolving semantic discovery. In: *Proceedings of the eleventh acm international conference on web search and data mining*. [S.l.: s.n.], 2018. p. 673–681. Citado 5 vezes nas páginas 25, 26, 27, 28 e 30.
- YIN, J. et al. A survey on bluetooth 5.0 and mesh: New milestones of iot. *ACM Transactions on Sensor Networks (TOSN)*, ACM New York, NY, USA, v. 15, n. 3, p. 1–29, 2019. Citado na página 23.
- YU, Y.; WANG, H.; LI, Z. Inferring mobility relationship via graph embedding. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, ACM New York, NY, USA, v. 2, n. 3, p. 1–21, 2018. Citado 2 vezes nas páginas 16 e 18.
- YUAN, Z. et al. Spatio-temporal dual graph attention network for query-poi matching. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. [S.l.: s.n.], 2020. p. 629–638. Citado na página 17.
- ZHANG, T.; DUAN, X.; LI, Y. Unveiling transit mobility structure towards sustainable cities: An integrated graph embedding approach. *Sustainable Cities and Society*, Elsevier, v. 72, p. 103027, 2021. Citado 3 vezes nas páginas 13, 17 e 19.
- ZHANG, T. et al. Integrating geovisual analytics with machine learning for human mobility pattern discovery. *ISPRS International Journal of Geo-Information*, Multidisciplinary Digital Publishing Institute, v. 8, n. 10, p. 434, 2019. Citado na página 17.

---

ZHANG, Z. et al. Srsn-based adaptive relay selection for d2d communications. *IEEE Internet of Things Journal*, IEEE, v. 5, n. 4, p. 2323–2332, 2017. Citado 4 vezes nas páginas 13, 23, 24 e 41.