



Anderson dos Santos Ferreira

Análise de dados de redes Wi-Fi por meio de redes de correlação

Recife

Dezembro de 2021

Anderson dos Santos Ferreira

Análise de dados de redes Wi-Fi por meio de redes de correlação

Artigo apresentado ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Universidade Federal Rural de Pernambuco – UFRPE

Departamento de Estatística e Informática

Curso de Bacharelado em Sistemas de Informação

Orientador: Glauco E. Gonçalves
Coorientador: Victor W. C. de Medeiros

Recife

Dezembro de 2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

- F383a dos Santos Ferreira, Anderson
Análise de dados de redes Wi-Fi por meio de redes de correlação / Anderson dos Santos Ferreira. - 2021.
19 f. : il.
- Orientador: Glauco Estacio Goncalves.
Coorientador: Victor Wanderley Costa de Medeiros.
Inclui referências.
- Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Rural de Pernambuco, Bacharelado em
Sistemas da Informação, Recife, 2021.
1. Rede Wi-Fi. 2. Redes de correlação. 3. Redes complexas. 4. Matriz de correlação. 5. Série temporal. I. Goncalves,
Glauco Estacio, orient. II. Medeiros, Victor Wanderley Costa de, coorient. III. Título

Anderson dos Santos Ferreira

Análise de dados de redes Wi-Fi por meio de redes de correlação

Artigo apresentado ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Aprovado em: 17 de Novembro de 2020.

BANCA EXAMINADORA

Glauco Estácio Gonçalves (Orientador)
Instituto de Tecnologia
Universidade Federal do Pará

Victor Wanderley Costa de Medeiros (Coorientador)
Departamento de Estatística e Informática
Universidade Federal Rural de Pernambuco

Silvana Bocanegra
Departamento de Estatística e Informática
Universidade Federal Rural de Pernambuco

Análise de dados de redes Wi-Fi por meio de redes de correlação

Anderson dos Santos Ferreira¹, Glauco E. Gonçalves², Victor W. C. de Medeiros¹

¹Departamento de Estatística e Informática – Universidade Federal Rural de Pernambuco
Rua Dom Manuel de Medeiros, s/n, - CEP: 52171-900 – Recife – PE – Brasil

²Instituto de Tecnologia – Universidade Federal do Pará
Rua Augusto Corrêa, 01, - CEP: 66.075-110 – Belém – PA – Brasil

andersoon_12@hotmail.com, glaucogoncalves@ufpa.br, victor.wanderley@ufrpe.br

Abstract. *The growing use of Wi-Fi networks has generated a large volume of data that allow studies to analyze human behavior. One of the ways to study these data is through a complex network created from correlation coefficients. Data from the UFRPE Wi-Fi network were collected during 6 days, the creation of time series allowed the analysis of the occupation graphs of the headquarters buildings, thus enabling the creation of complex networks based on correlation coefficients. The occupancy graphs showed the times and buildings with the highest occupancy for each day, while the complex network metrics showed a strong correlation of buildings with different degree values.*

Resumo. *A crescente utilização de redes Wi-Fi tem gerado um grande volume de dados que possibilitam estudos para analisar o comportamento humano. Um dos meios de estudar estes dados é através de uma rede complexa criada a partir de coeficientes de correlação. Dados da rede Wi-Fi da UFRPE foram coletados durante 6 dias, a criação de séries temporais permitiu a análise dos gráficos de ocupações dos prédios do campus Sede, assim possibilitando a criação de redes complexas com base em coeficientes de correlação. Os gráficos de ocupação mostraram os horários e prédios com as maiores ocupações para cada dia, enquanto que as métricas das redes complexas mostrou uma forte correlação de prédios com valores de grau distintos.*

1. Introdução

A existência de redes Wi-Fi nos mais diversos lugares, como aeroportos, hotéis, restaurantes e lojas, aliado ao grande número de dispositivos móveis; em Junho de 2021, o Brasil registrou 346 milhões de dispositivos móveis, só de *smartphone* são 242 milhões de unidades, o que representa mais de um aparelho por habitante [Fernando 2021]; contribui para o crescimento do uso de redes Wi-Fi, assim gerando um grande volume de dados. Os dados originados da utilização de redes Wi-Fi apresentam características espaço-temporais, pois registram atividades de pessoas em determinados locais por um determinado tempo e que mudam de maneira dinâmica com o passar do tempo.

Compreender a conduta humana e a dinâmica populacional por meio de dados de dispositivos Wi-Fi, tornou-se uma relevante linha de pesquisa na área da computação [Kontokosta and Johnson 2017]. A utilização de dados de rede Wi-Fi para monitoramento

e análise do comportamento populacional vem sendo abordado, pelo menos, desde a metade da primeira década do século 21. Um mapeamento de dados da rede Wi-Fi do campus do MIT foi utilizado com objetivo de buscar uma melhor compreensão dos padrões de atividades da comunidade acadêmica, com base na captura realizada por 3 mil pontos de acesso [Sevtsuk et al. 2008].

Estudar e compreender padrões de mobilidades em ambientes urbanos com alta densidade é um desafio para os municípios que buscam aperfeiçoar o planejamento e a gestão urbana. A coleta em 54 pontos de acesso de rede Wi-Fi, durante uma semana, em uma parte de Lower Manhattan em Nova York, resultou em mais de 30 milhões de capturas de mais de 800.000 dispositivos móveis únicos. A análise permitiu identificar níveis de intensidade de uso das ruas, padrões de mobilidade e comportamento humano em ambientes urbanos heterogêneos [Traunmueller et al. 2018].

Outra forma de analisar dados de rede Wi-Fi é através da modelagem destes por meio de redes complexas. Redes complexas são representadas por grafos que contêm nós e enlaces, que indicam uma relação entre os nós. É possível modelar coisas que possam representar os nós e que possuam relações entre si, assim podendo ser modelado dados de redes sociais, transporte público, distribuição de energia ou água, rede de aeroportos e até proteínas.

As redes Wi-Fi são formadas por pontos de acesso, que podem representar um nó, enquanto que as movimentações dos usuários da rede entre os pontos de acessos, podem representar os enlaces entre os nós. Calcular a correlação entre duas ou mais variáveis permite identificar o grau de interferência de uma variável em outra. O coeficiente de correlação de Pearson mede o grau de correlação de variáveis, através de valores positivos, negativos e/ou nulos. Por meio de métodos de redes complexas é possível estudar os dados de uma rede Wi-Fi, considerando os coeficientes de correlação, sob uma nova perspectiva [Wang 2021].

O objetivo deste trabalho é analisar dados espaço-temporais de uma rede Wi-Fi sob a perspectiva de uma rede complexa de correlação de seus atributos. Uma vez que o grande volume de dados gerados por uma rede Wi-Fi, exigem técnicas avançadas de mineração, visualização e análise, que permitem examinar condutas humanas através de métodos de redes complexas.

Este artigo é composto por mais 5 seções. A seção 2 contém detalhes do funcionamento de uma rede Wi-Fi de grande escala e conceitos importantes do coeficiente de correlação e de redes complexas. A seção 3 detalha a obtenção dos dados e a modelagem aplicada para obtenção da quantidade de pessoas nos prédios, coeficiente de correlação e métricas das redes complexas. A seção 4 é designada para a apresentação dos resultados obtidos. A seção 5 discute os resultados obtidos. Por fim, a seção 6 diz respeito as conclusões obtidas após a análise dos resultados, além de apresentar sugestões para trabalhos futuros.

2. Referencial Teórico

Esta seção aborda o funcionamento de redes Wi-Fi e os principais conceitos envolvidos no uso de redes complexas para representação de dados.

2.1. Rede Wi-Fi

Diferentemente do que acontece em uma rede Wi-Fi residencial, onde a associação, autenticação e conexão à rede é realizada por um modem; uma rede Wi-Fi de grande escala utiliza pontos de acesso e controladora. Os pontos de acesso são dispositivos de rede de comunicação que permite aos dispositivos móveis conectar-se à rede [Cisco 2018]. A controladora é um dispositivo de rede de comunicação responsável pela configurações, associações, autenticações e gerenciamento dos pontos de acesso [Cisco 2009]. Em resumo, os usuários se conectam aos pontos de acesso, que trocam informações com a controladora para permitir ou não a conexão do usuário.

O dispositivo móvel de um usuário de uma rede Wi-Fi antes de ter acesso à internet passa por um dos seguintes estados:

- A. Não autenticado ou não associado
- B. Autenticado, contudo ainda não associado
- C. Autenticado e associado.

Caso seja o primeiro contato do dispositivo na área da rede Wi-Fi, ele estará no estado A e enviará um quadro do tipo *Probe Request* (requisição de sondagem), item 1 da Figura 1, que é a requisição que um dispositivo móvel realiza de maneira periódica, para que pontos de acesso, em sua proximidade, enviem informações sobre redes sem fio disponíveis [Oliveira et al. 2019].

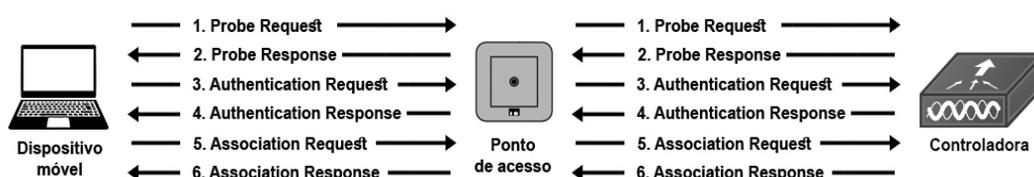


Figura 1. Representação dos elementos da arquitetura e quadros de gerência de redes Wi-Fi.

A Figura 1 mostra um caso em que todos os quadros são repassados do ponto de acesso para a controladora, contudo não é uma regra e que pode ser diferente a depender do fabricante e das configurações dos equipamentos [L. Yang 2005].

O *Probe Request* contém: endereço MAC, IP, taxas de dados e as configurações de segurança aceitas pela placa de rede do dispositivo móvel. Um ponto de acesso na proximidade de um dispositivo móvel recebe o quadro *Probe Request* e o envia para controladora verificar se o dispositivo móvel possui pelo menos uma taxa de dados suportada pela rede Wi-Fi. Um *Probe Response* (resposta de sondagem), item 2 da Figura 1, será retornado para o ponto de acesso, que o repassa para o dispositivo móvel caso as taxa de dados sejam compatíveis. Neste quadro serão enviadas informações de taxas de dados e tipos de criptografia compatíveis, bem como outros recursos da rede Wi-Fi.

A escolha da rede móvel a qual o dispositivo móvel se conectará é tomada com base no *Probe Response* que é recebido. Quando uma rede compatível é descoberta, o dispositivo móvel envia um quadro *Authentication Request* (requisição de autenticação) para um ponto de acesso, item 3 da Figura 1, que repassa o quadro para controladora verificar os dados e definir se retornará um *Authentication Response* (resposta de autenticação)

positivo ou negativo para o ponto de acesso repassar para o dispositivo móvel, item 4 da Figura 1. Quando um ponto de acesso recebe um *Authentication Request* ou um *Probe Request* de um dispositivo não autenticado, ele responde com um quadro de desautenticação, colocando a dispositivo móvel no estado A.

Quando um dispositivo móvel define a qual ponto de acesso gostaria de se associar, ele envia um *Association Request* (requisição de associação), item 5 da Figura 1, ao ponto de acesso. O ponto de acesso repassa o quadro para controladora verificar os dados contidos no *Association Request* e em caso de compatibilidade, cria um ID de associação ao dispositivo móvel. A controladora envia um quadro *Association Response* positivo para o ponto de acesso, que repassa a informação para o dispositivo móvel, item 6 da Figura 1, concedendo acesso à rede, assim permitindo a transferência de dados com o dispositivo móvel [Meraki 2020].

Além dos pacotes citados, existe o *Reassociation Request* que é um pacote enviado por um dispositivo móvel quando ele sai do alcance de um ponto de acesso e se conecta a outro ponto de acesso [Cisco 2021]. Os quadros de *Probe Request*, *Probe Response*, *Authentication Request*, *Authentication Response*, *Association Request* e *Reassociation Request* contêm as informações necessárias para cada um dos estados: horário de envio e/ou recebimento do quadro, MAC do dispositivo móvel e MAC do ponto de acesso ao qual o dispositivo está se comunicando. Informações que são essenciais para o desenvolvimento do trabalho.

2.2. Redes Complexas

Por lidar com conceitos de sistemas dinâmicos, estatística e teoria dos grafos, a ciência das redes despertam o interesse de diversas áreas do conhecimento, como física, matemática, biologia, sociologia e computação. A ciência das redes pode ser enquadrada dentro da chamada análise de sistemas complexos, que tem como objetivo estudar sistemas com múltiplos componentes, o que exige habilidades diferentes em três grandes áreas: matemática, estatística e computação [Lewis 2009].

Os métodos de redes complexas são úteis para o entendimento profundo da interação de um sistema com diversos componentes. Interações que acontecem entre elementos de sistemas reais distintos a princípio podem parecer que não possuam nenhuma correlação entre eles, contudo um estudo com uso de modelos em redes, nos mostra que muitas propriedades são compartilhadas [Sayama 2015]. As redes complexas tem como base a teoria de grafos [Barabási 2021], um ramo da matemática e da computação, logo assimilar suas definições e conceitos é importante para entender os termos que serão utilizados ao logo do trabalho.

Um grafo é formado por dois conjuntos finitos N e E , onde N é o conjunto de nós e E é o conjunto de enlaces. A conexão entre dois nós é feita por um enlace, logo todo enlace também é uma dupla $\{i, j\}$ onde $i \in N$ e $j \in N$. Um exemplo de um grafo seria definido por $G = (N, E)$, onde $N = \{1, 2, 3, 4, 5, 6\}$ e $E = \{\{1,2\}, \{1,5\}, \{2,3\}, \{2,5\}, \{3,4\}, \{4,5\}, \{4,6\}\}$, representado na Figura 2.

Grafos podem ser dirigidos ou não dirigidos, a diferença fica pela presença ou ausência de sentido nos enlaces do grafo. Nos grafos dirigidos, os enlaces (i, j) são representados por pares ordenados, enquanto que em grafos não-dirigidos são representados por conjuntos $\{i, j\}$. Um exemplo de um grafo dirigido seria definido por $D = (N, E)$, onde

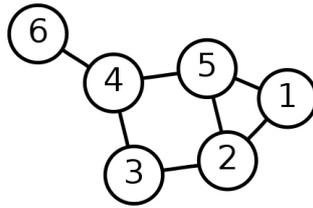


Figura 2. Representação do grafo não-dirigido G.

$N = \{ 0, 1, 2, 3, 4, 5, 6, 7 \}$ e $E = \{(0,2), (0,4), (1,3), (2,7), (3,6), (4,5), (4,7), (5,1), (5,7), (6,0), (6,2), (6,4), (7,3), (7,5)\}$, representado na Figura 3

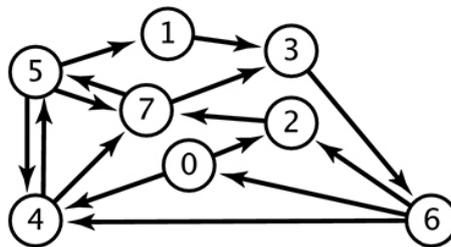


Figura 3. Representação de grafo dirigido D.

Redes complexas podem apresentar um nó ou um grupo de nós que não se conectam com os demais, assim formando grafos desconexos. Cada grupo de nós conectados é chamado de componente conectada do grafo desconexo. A componente conectada que reunir a maior quantidade de nós é chamada de maior componente conectada.

A quantidade de enlaces que um nó possui permite medir a importância ou a popularidade de um nó na rede; quanto mais enlaces estiverem conectados a um nó, maior será a sua importância na rede; sendo denominada de grau do nó. Outras medidas que podem ser obtidas a partir do grau de um nó é a do grau médio, que é média de graus dos nós em um grafo, enquanto que grau máximo indica a número máximo de enlaces que um nó ou mais nós possuem.

Outra medida importante para o trabalho é a densidade do grafo, que é a relação da quantidade de enlaces presentes na rede e da quantidade de enlaces possíveis na rede. Redes complexas reais costumam ter baixa densidade, em outras palavras, possuem muito menos enlaces do que poderia ter, considerando-se o número de nós.

Saber o grau de um nó é importante para descobrir a importância dele dentro do grafo, a maneira de medir a importância de um nó é através do coeficiente de aglomeração, que mede o quanto a vizinhança de um nó é coesa. Quanto mais vizinhos de um nó estiverem conectados entre si, maior será seu coeficiente de aglomeração, que geralmente apresenta um alto valor em redes complexas.

Estudar os caminhos entre os nós de uma rede é importante para saber o quais os menores caminhos entre dois nós, uma vez que permite calcular o tamanho médio dos caminhos (TMC), que é obtido através da média dos menores caminhos entre todo par de nós presentes em um grafo. Redes complexas costumam apresentar um baixo valor para de TMC.

Uma outra maneira de avaliar a importância de um nó é através do *Betweenness* ou intermediação, que soma a fração dos menores caminhos que passam por um nó f levando em consideração os menores caminhos entre todos os pares de nós. Se entre dois nós g e h existem dois menores caminhos e apenas um passa por f , logo será contabilizado $1/2$ a soma de todas as frações que determina a intermediação.

2.3. Redes de Correlação

É comum que estudos que reúnam duas ou mais variáveis desperte o interesse em conhecer o relacionamento entre elas. O coeficiente de correlação é uma medida que revela o grau de relacionamento entre duas variáveis, exibindo como elas variam sem a necessidade de explicar qual variável é dependente e qual é independente [Lira 2004].

Uma maneira de calcular a correlação entre duas variáveis é através do coeficiente de correlação de Pearson (ρ), que varia entre -1 e 1, com o sinal indicando a direção do relacionamento como negativa ou positiva e o valor indicando a força da relação entre as variáveis. Correlação com valor zero mostra que não existe correlação entre as variáveis. Correlação com valor negativo indica que quando o valor de uma variável aumenta, o valor da outra variável diminui. Correlação com valor positivo indica que quando o valor de uma variável aumenta, o valor da outra variável também aumenta [Paranhos et al. 2014].

Elaborar uma rede a partir da correlação de séries temporais é comum em diversas áreas como na ciência da saúde, agronomia, mercado financeiro entre outros. [Souza et al. 2021] criou redes de correlação para detectar alertas precoces do surgimento de uma pandemia com base em dados da COVID-19. [Silva 2017] buscou verificar a viabilidade do melhoramento de soja com uso de rede de correlações.

3. Materiais e Métodos

Esta seção descreve os materiais utilizados e os métodos aplicados para analisar os dados da rede Wi-Fi da UFRPE.

3.1. Base de dados

Sabendo-se que um dispositivo móvel envia e recebe pacotes para ter acesso à internet disponibilizada por uma rede sem fio, possibilita que a captura destes pacotes permita descobrir: data, horário e localização do usuário. Para que a coleta dos dados fosse realizada, foi feito um contato com a coordenação de redes e comunicação da UFRPE para a captura de dados da rede Wi-Fi presente no campus Sede.

As capturas dos quadros foram realizadas entre os dias 07/06/2019 (Sexta-feira) às 17h05min01s e 12/06/2019 (Quarta-feira) às 04h32mi55s, o horário de encerramento das capturas se deu pela limitação de espaço em disco da máquina local. Arquivos de captura de pacotes de redes corporativas do porte da rede de um campus universitário tendem a crescer rapidamente devido ao grande volume de pacotes trafegados na rede, por isso o espaço em disco disponível no momento da captura se apresentou como uma limitação neste experimento. No entanto, deve-se observar que a captura realizada oferece uma quantidade de dados significativa para este estudo, já que permite observar o comportamento dos usuários em um dia típico.

O processo de capturas foi realizado pelos coordenadores de redes e comunicação da UFRPE, que espelharam uma máquina local na porta da VLAN (rede local virtual)

a qual se encontrava conectada a controladora, assim todos os dados que chegavam ou saíam da controladora, poderiam ser capturados pela máquina local.

Para o trabalho foram utilizados dados dos quadros de *Association Request* e *Re-association Request*, que são de usuários com permissão de acesso à rede Wi-Fi. Pelo fato de apenas docente, discente e técnico administrativo da UFRPE possuírem acesso à rede Wi-Fi disponibilizada pela UFRPE, a utilização dos quadros possibilitou o cumprimento da finalidade de representar ocupantes dos prédios do campus Sede da UFRPE através dos dados de usuários da rede Wi-Fi.

3.2. Métodos

De posse dos dados, o próximo passo consistiu na filtragem dos quadros, uma vez que a controladora captura todos os quadros que chegam até ela. Utilizando o Wireshark, programa que analisa protocolos de redes, foi possível separar todos os quadros de *Association Request* e *Reassociation Request*, assim diminuindo a quantidade de quadros de pouco mais de 232 milhões para 1 milhão 288 mil e 403 quadros.

A filtragem realizada no Wireshark gerou arquivos no formato *.csv* que agrupavam os quadros por dia de captura, contudo quadros com horário próximo ao final de um dia eram agrupados nos arquivos correspondente ao dia seguinte. Como cada pacote possui a informação do horário e do dia da captura, através de um *script* escrito na linguagem Python foi possível agrupar corretamente cada pacote ao seu respectivo dia de captura.

Os coordenadores de redes e comunicação da UFRPE forneceram informações sobre localizações de cada um dos 162 *Access Points* presentes no campus Sede. Cada *Access Point* está associado a um único prédio, que pode conter um ou mais *Access Points*. Para análises de dados mais precisas, alguns prédios foram unificados, levando em consideração a distância entre os prédios e a quantidade de *Access Points* em cada um deles. Depois de unidos, os prédios receberam uma numeração para uma melhor visualização e identificação de cada prédio, como mostrado na Figura 4. O número do prédio e seu respectivo nome estão na Tabela 1.

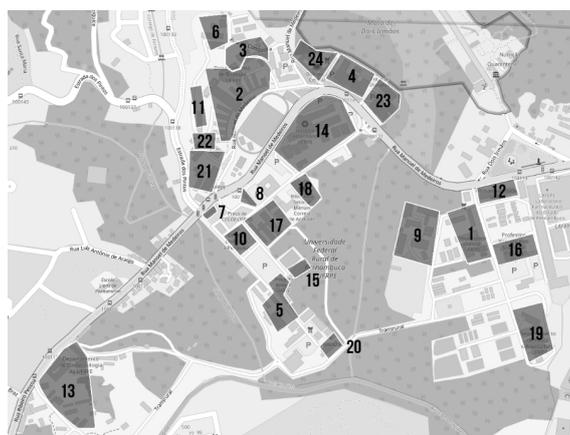


Figura 4. Numeração de cada prédio no campus Sede da UFRPE.

Cada pacote traz diversas informações, contudo as mais importantes para o trabalho são: horário de captura do pacote, endereço MAC do dispositivo móvel e endereço MAC do *Access Point* ao qual o dispositivo móvel está tentando se conectar. Com as

Tabela 1. Numeração e respectivo nome de cada prédio do campus Sede da UFRPE.

1 - DCFL-DEPA	7 - CASA 4 FEMININA	13 - DZ	19 - DEPT PESCA
2 - PC	8 - DCD	14 - DMV	20 - PREFEITURA
3 - DB	9 - CCEAA-DTR	15 - DEFIS	21 - CEGEN-DEINFO
4 - PRO REITORIA	10 - CENAPESQ	16 - CEAGRI	22 - DQ-AGRICOLA-EAD
5 - DED-PADR-DADM	11 - DCM-DCE	17 - CEGOE	23 - PRAE-DQV
6 - DB-NB	12 - FITOSSANIDADE	18 - BIBLIOTECA SETORIAL	24 - BIBLIOTECA CENTRAL

informações citadas, criou-se séries temporais com duração de 5 minutos que contêm as quantidades de pessoas em cada *Access Point* do campus Sede da UFRPE.

Um *script* na linguagem Python foi criado para realizar a contagem, contendo uma condição que permite que apenas quadros com endereço MAC de *Access Point* do campus Sede sejam analisados para contagem. Um usuário é contado em um prédio caso atinja uma das seguintes condições: 1) No último prédio onde foi identificado durante uma série temporal. Caso o usuário transite por mais de um prédio durante uma série temporal, será contabilizado apenas no último prédio no qual foi identificado durante a série temporal referente ao horário do momento da identificação. 2) Ser identificado em um prédio e após 5 minutos não ser visto em nenhum outro prédio.

Foi assumido que um usuário seria contado na série atual e na série seguinte, caso seja identificado em um prédio na série atual e não seja identificado na série seguinte, por exemplo: um usuário foi visto no prédio CEAGRI às 07h13min32s, referente ao intervalo atual que é entre 07h10min00s até 07h14min59s, do dia 10/06/2019 (Segunda-feira), e não foi identificado no intervalo seguinte que é entre 07h15min00s e 07h19min59s, logo será contado no CEAGRI no dia 10/06/2019 no intervalo atual que é entre 07h10min00s até 07h14min59s e no intervalo seguinte, entre 07h15min00s e 07h19min59s.

Dentro de 24 horas existem 288 intervalos de 5 minutos, logo foi preciso atribuir as respectivas quantidades de pessoas aos 288 intervalos de cada um dos prédios. Utilizando a linguagem Python, para cada um dos dias uma lista foi criada para cada um dos prédios, que recebeu a quantidade de pessoas referente aos seus 288 intervalos, assim permitindo gerar gráficos com curvas de ocupação e o cálculo do coeficiente de correlação. Através da função *corrcoef* da biblioteca *numpy* da linguagem Python, foi passado como parâmetro da função uma lista com todas as listas com as quantidades de pessoas de cada prédio. O retorno da função foi uma matriz com os coeficientes de correlação dos prédios.

Dada as séries temporais de ocupação de cada prédio foi possível computar uma matriz de correlação destas séries. Uma rede não dirigida e sem pesos foi elaborada. Criando um nó para cada prédio e criando um enlace entre dois nós a depender do valor do coeficiente de correlação entre as séries temporais de ocupação dos respectivos prédios. Um limiar foi usado para determinar se o enlace será criado ou não. Aqueles prédios que não possuem enlaces (grau zero) são então removidos da rede. Neste experimento o limiar foi variando de 0 a 0.9 de modo a observar o impacto do limiar em cada uma das medidas das redes formadas, as interpretações dos valores dos coeficientes de correlação mostrada na Tabela 2. As redes foram criadas através da biblioteca *networkx* da linguagem Python, assim possibilitando obter métricas como densidade, grau, coeficiente de

aglomeração e Betweenness, que permitem medir e analisar a importância dos nós para rede.

Tabela 2. Interpretações dos valores dos coeficientes de correlação.

Valor de ρ	Interpretação	Valor de ρ	Interpretação
-0.8 a -1	correlação negativa muito forte	0.1 a 0.19	correlação positiva muito fraca
-0.6 a -0.79	correlação negativa forte	0.2 a 0.39	correlação positiva fraca
-0.4 a -0.59	correlação negativa moderada	0.4 a 0.59	correlação positiva moderada
-0.2 a -0.39	correlação negativa fraca	0.6 a 0.79	correlação positiva forte
-0.1 a -0.19	correlação negativa muito fraca	0.8 a 1	correlação positiva muito forte

4. Resultados

Os mais de 1 milhão de quadros capturados durante 6 dias, são oriundos de 14.927 usuários únicos, levando em consideração que uma pessoa possui apenas 1 dispositivo móvel e que o MAC de um dispositivo móvel representa 1 pessoa. A divisão das capturas é mostrada na Tabela 3, a quantidade de dados obtidos durante a sexta-feira é interessante, uma vez que os dados só começaram a ser capturados às 17h05min01s e ainda foi possível capturar uma quantidade consideravelmente alta de dados, quando comparado ao Sábado, Domingo e Quarta-feira.

Tabela 3. Quantidade de usuários únicos e quadros de *Association Request* e *Reassociation Request* obtidos de acordo com o dia.

Dias	Quantidade de capturas	Quantidade de usuários únicos
Sexta	135.235	5.598
Sábado	24.374	1.198
Domingo	8.951	502
Segunda	571.123	10.825
Terça	547.657	11.186
Quarta	1.063	147
	Total: 1.288.403	

A quantidade de dados capturados no Sábado, Domingo e Quarta estão dentro do esperado. O baixo valor do Sábado e Domingo é explicado pelo fato de não haver aula, enquanto que a quantidade da Quarta é justificada pela captura ter sido encerrada às 04h32min55s, o período da madrugada tem a presença de poucas pessoas no campus Sede.

4.1. Gráficos de ocupação por horário

Os gráficos com as taxas de ocupação trazem resultados interessantes e que merecem um pouco mais de análise. Os dias com mais dados (Sexta-feira, Segunda-feira e Terça-feira) apresentam resultados similares, apesar da Sexta-feira não possuir dados de uma

dia inteiro. O destaque fica por conta dos prédios PC, CEGOE e CEGEN-DEINFO, os únicos a possuírem picos de ocupação durante a análise, com o PC ficando por maior tempo alternando entre subidas e descidas, entre 17h05min e 19h20min, como mostrado na Figura 5.

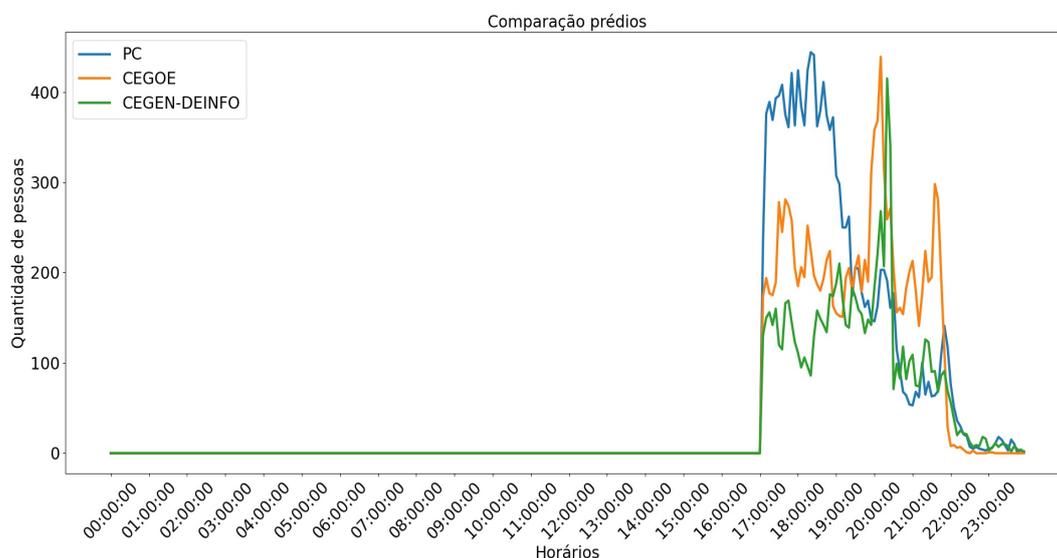


Figura 5. Curvas de ocupação dos prédios na Sexta-feira.

O gráfico do Sábado tem dois prédios com as maiores ocupações, apesar de ser um dia de final de semana, onde a quantidade de pessoas é menor no campus. O CEGEN teve uma maior ocupação entre às 07h30min e 12h05min, enquanto que a ocupação do CEGOE alternou entre subidas e descidas entre às 07h30min e 16h35min, porém com uma quantidade acima de 10 pessoas, com exceções às 08h30min e às 12h40min, além de ter tido o maior pico às 13h55min, quando foi contabilizado 65 pessoas, como mostrado na Figura 6.

A ocupação do Domingo é ainda menor que a do Sábado, o maior pico é atingido pelo prédio do DMV com 13 pessoas às 13h35min. Dos 24 prédios, apenas três (CEGEN-DEINFO, PC e DMV) tiveram picos com pelo menos 10 pessoas, com as maiores ocupações acontecendo entre às 08h55min e 17h15min, como mostrado na Figura 7.

A Segunda-feira oferece um gráfico interessante para análise, uma vez que mostra dados de um dia normal no campus. A ocupação dos prédios ganha intensidade às 06h55min, depois mantendo variações que duram até às 12h00min, quando é notado uma queda na ocupação dos prédios, com exceção do PC, que cresce.

Uma possível justificativa para isso está no fato do prédio ser caminho para o Restaurante Universitário. A queda na ocupação dos demais prédios segue até às 12h35min, quando volta a crescer de forma gradual. A linha que mais chama atenção é a do prédio DED-PADR-DADM, pois apresenta 13 picos com pelo menos 800 pessoas, como mostrado na Figura 8.

Durante o horário noturno, 19h05min e 22h05min, o prédio CEGOE é o que mais registra ocupação, o detalhe fica por conta do DED-PADR-DADM, que para o intervalo apresenta uma ocupação sempre inferior a 200 pessoas e com linhas de ocupação seme-

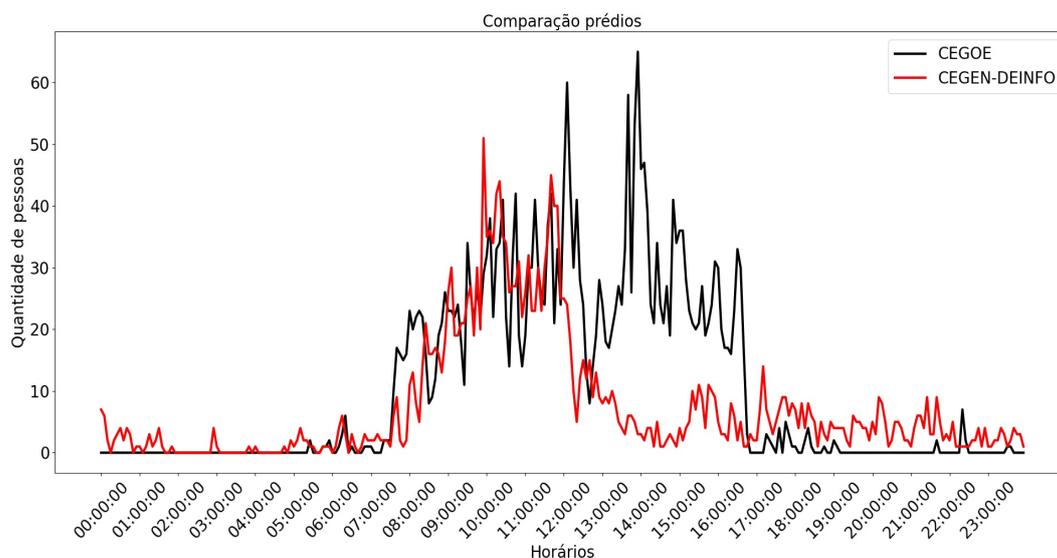


Figura 6. Curvas de ocupação dos prédios no Sábado.

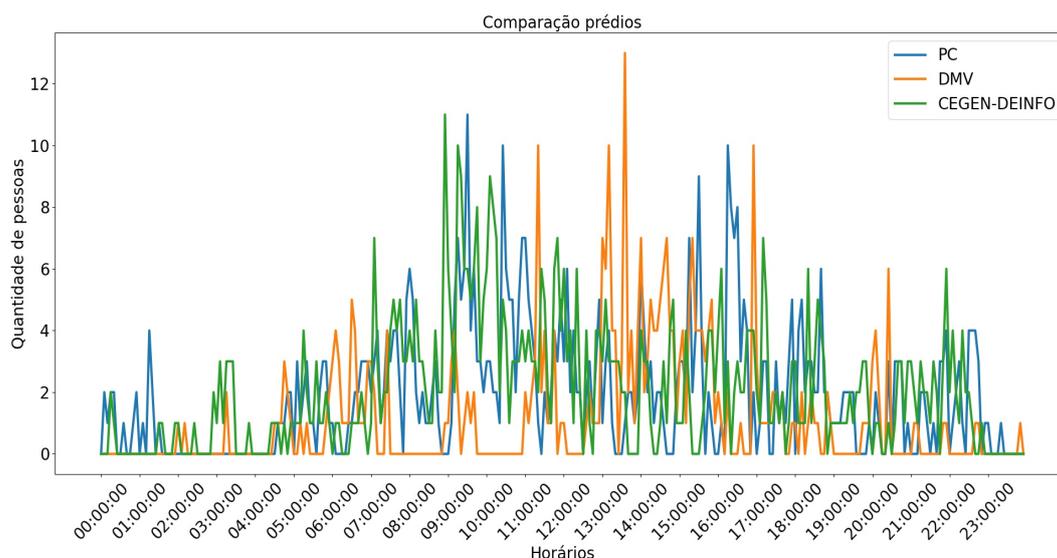


Figura 7. Curvas de ocupação dos prédios no Domingo.

lhantes aos outros prédios analisados.

O gráfico da Terça-feira, mostrado na Figura 9, mostra que os prédios começam a registrar mais pessoas às 05h45min, com o CEGOE liderando. Entre às 07h e 17h55min, o PC, CEAGRI, CEGOE e DMV são os prédios que apresentam as maiores taxas de ocupação. Os maiores picos ficam por conta do DED-PADR-DADM, o primeiro tem início às 09h, chega ao máximo às 09h35min e depois cai de forma acentuada até às 09h55min.

A ocupação dos prédios, de maneira semelhante a Segunda-feira, apresenta uma queda a partir das 12h, assim como na Segunda-feira, porém a volta do crescimento de ocupação inicia às 12h45min. O DED-PADR-DADM volta a alternar entre picos e quedas

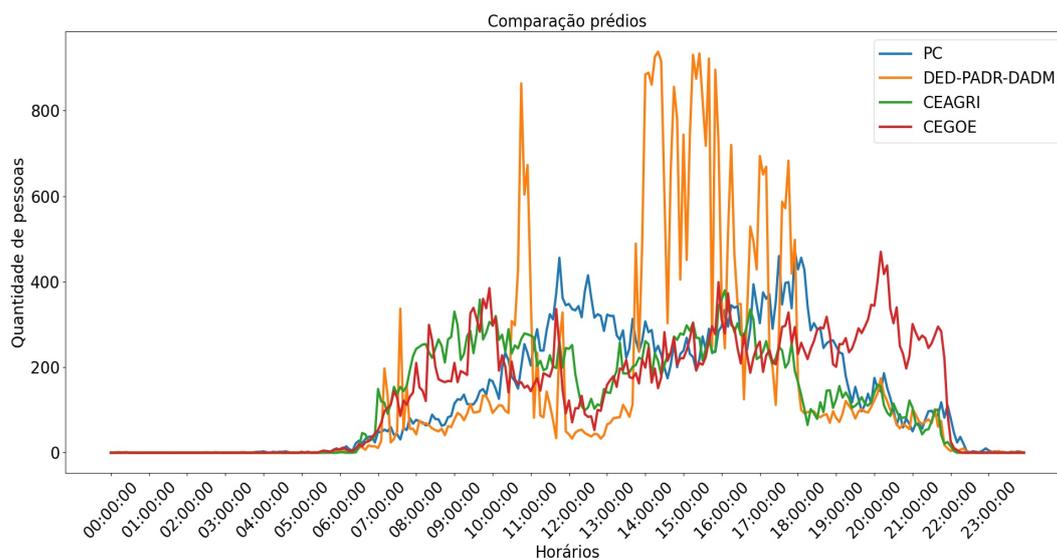


Figura 8. Curvas de ocupação dos prédios na Segunda-feira.

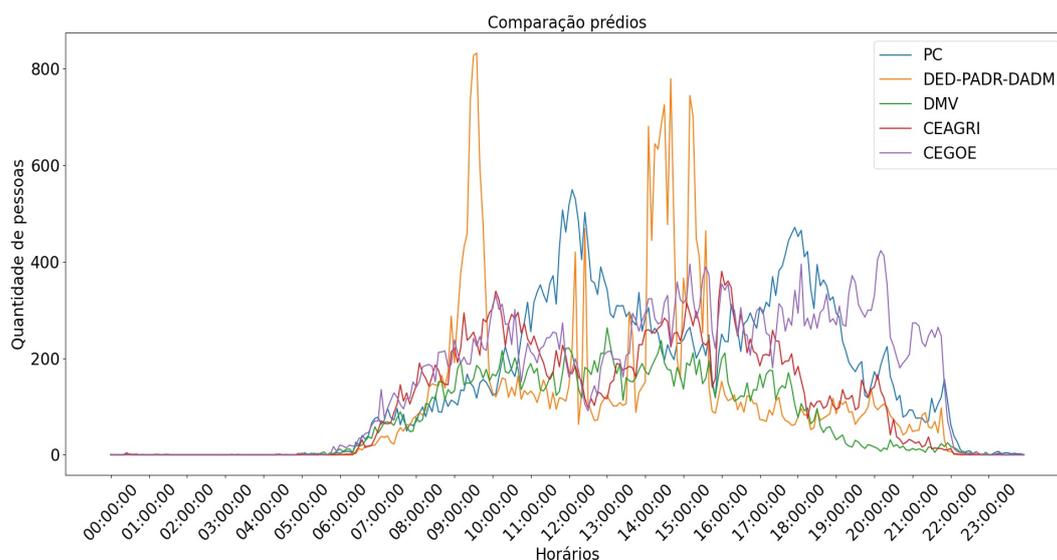


Figura 9. Curvas de ocupação dos prédios na Terça-feira.

bruscas das 14h05min até às 15h35min, quando passa a apresentar variações similares aos prédios com ocupações entre 100 e 200 pessoas.

As linhas do gráfico da Quarta-feira, mostrado na Figura 10, apresentam prédios com pouca ou nenhuma ocupação. A maior ocupação é da CASA-4-FEMININA, com 6 pessoas às 00h20min.

4.2. Redes complexas com base em rede de correlação

Os gráficos gerados com as métricas das redes complexas apresentam linhas semelhantes, seja para os dias com mais registros (Sexta-feira, Segunda-feira e Terça-feira) ou para os dias com menos registros (Sábado, Domingo e Quarta-feira), com exceção do Domingo que apresenta linhas distintas para *Betweenness* médio, quantidade de componentes e

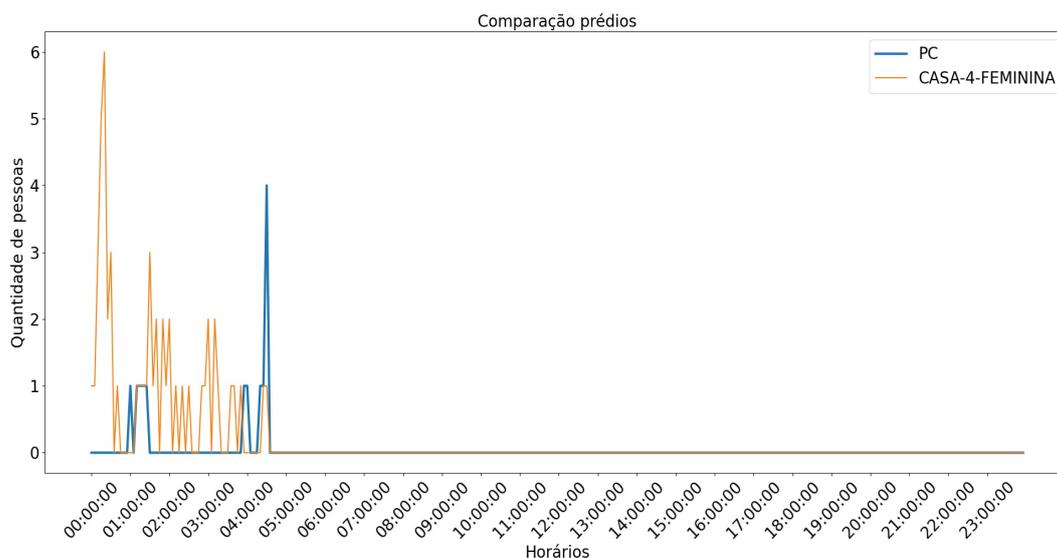


Figura 10. Curvas de ocupação dos prédios na Quarta-feira.

TMC.

Os gráficos a seguir trazem dados de todos os dias, contudo a rede gerada e suas análises são baseadas nos dados da Terça-feira, o segundo dia com mais registros, e no valor de 0.6 para o limiar do coeficiente de correlação, que representa correlação forte e muito forte, a matriz de correlação obtida na Terça-feira é mostrada na Figura 11.

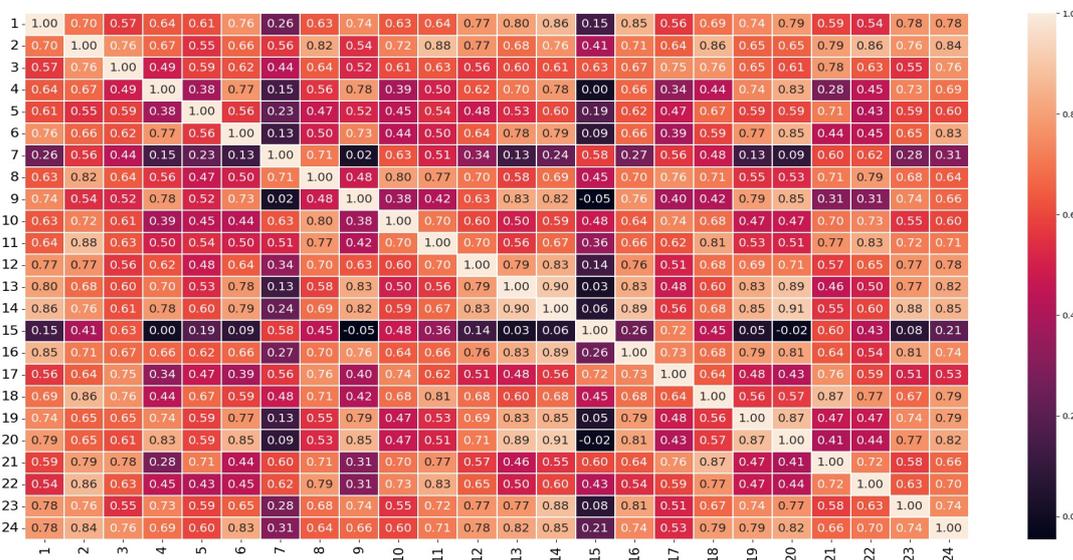


Figura 11. Matriz com os coeficientes de correlação dos prédios obtida com as quantidades de pessoas da Terça-Feira.

A rede é composta por 24 nós, gráfico (a) da Figura 12; mostrando que a rede reúne nós que representam todos os prédios do campus Sede. A rede possui 158 enlacs, gráfico (b) da Figura 12, que representam as conexões dos prédios de acordo com os valores de coeficientes de correlação, para o caso da rede complexa com valor de limiar

de 0.6, por exemplo, um enlace mostra que um par de prédios possui o coeficiente de correlação com valor de pelo menos 0.6.

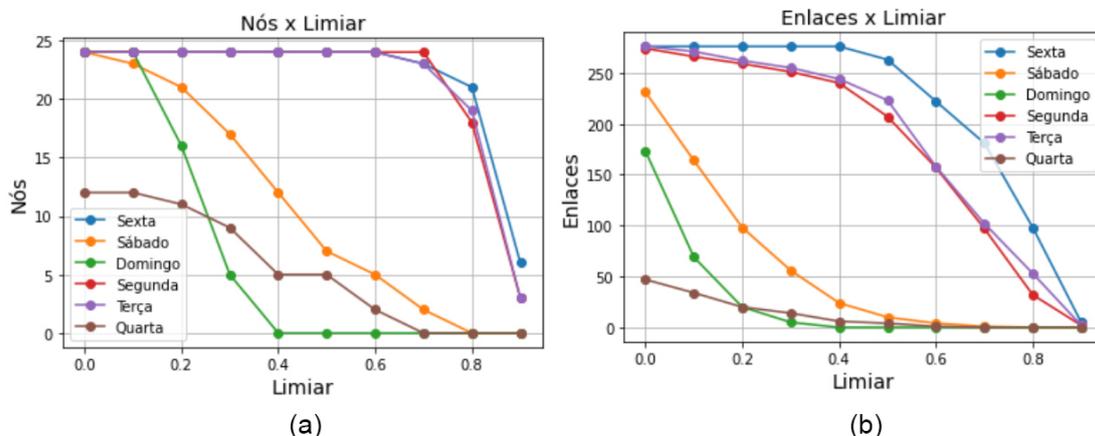


Figura 12. Gráficos da quantidade de nós (a) e enlaces (b) nas redes complexas de acordo o limiar para cada dia de captura.

O resultado de 0.57 para densidade, como mostrando no gráfico (c) da Figura 13, é considerado alto quando comparado com outras redes reais que possuem centenas ou milhares de nós. O coeficiente de aglomeração com valor de 0.77, gráfico (d) da Figura 13, mostrando que na rede analisada existe uma boa possibilidade de que os nós vizinhos de um determinado nó sejam vizinhos entre si, o que leva a perceber que a maneira como os prédios vizinhos são ocupados são semelhantes.

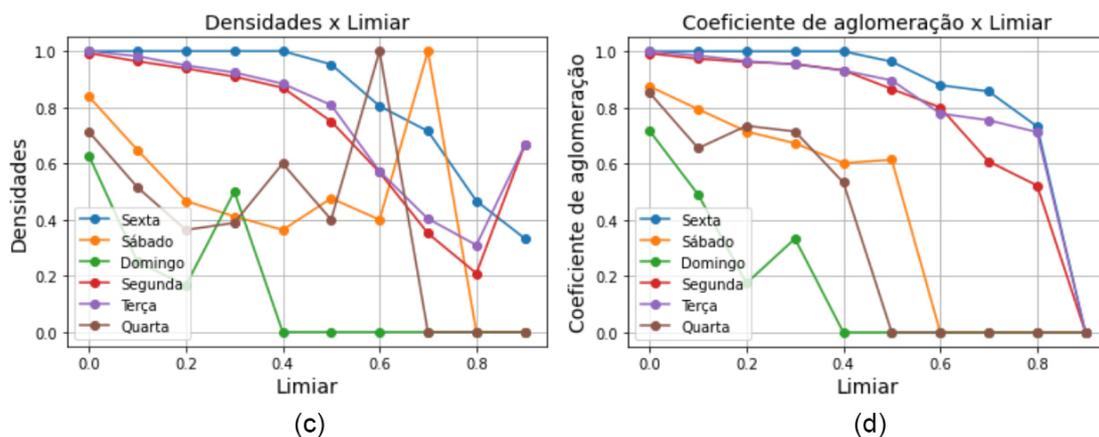


Figura 13. Gráficos com valores de densidades (c) e Coeficiente de aglomeração (d) nas redes complexas de acordo o limiar para cada dia de captura.

O grau médio da rede é 13, como mostra o gráfico (e) da Figura 14, o que representa que um nó se correlaciona com pouco mais de 50% dos demais nós da rede. O gráfico (f) da Figura 14 indica que o maior grau da rede tem o valor 19 e pertence a dois nós, um nó que representa o prédio do DMV e o nó que representa o prédio da BIBLIOTECA CENTRAL.

Além dos prédios com maior grau, foi identificado os nós com os menores graus: CASA-4-FEMININA e DED-PADR-DADM com grau 5 e DEFIS com grau 4. O detalhe

fica pela proximidade geográfica entre os prédios com maior grau, assim como acontece com os prédios com menores graus, como mostra na Figura 15. A observação reforça a interpretação do coeficiente de aglomeração, que mostra que existe uma boa possibilidade do comportamento de ocupação dos prédios vizinhos serem iguais.

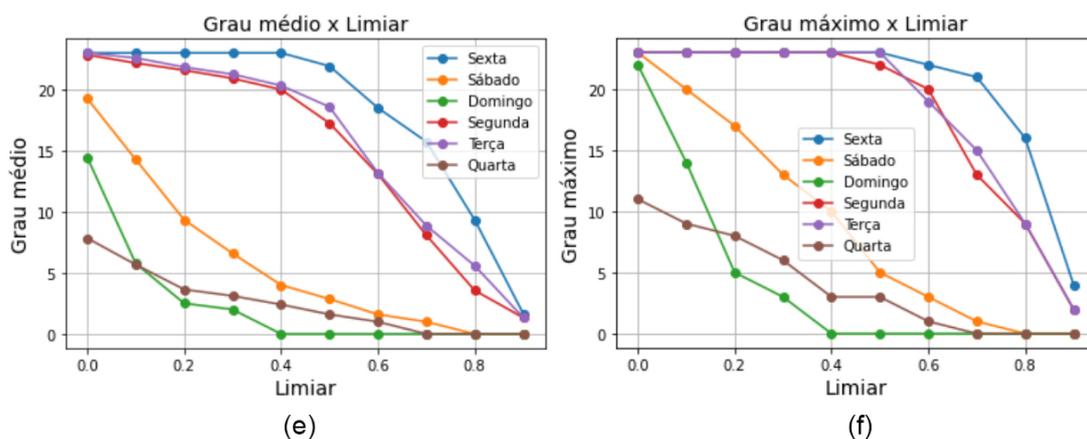


Figura 14. Gráficos com valores de grau médio (e) e grau máximo (f) nas redes complexas de acordo o limiar para cada dia de captura.

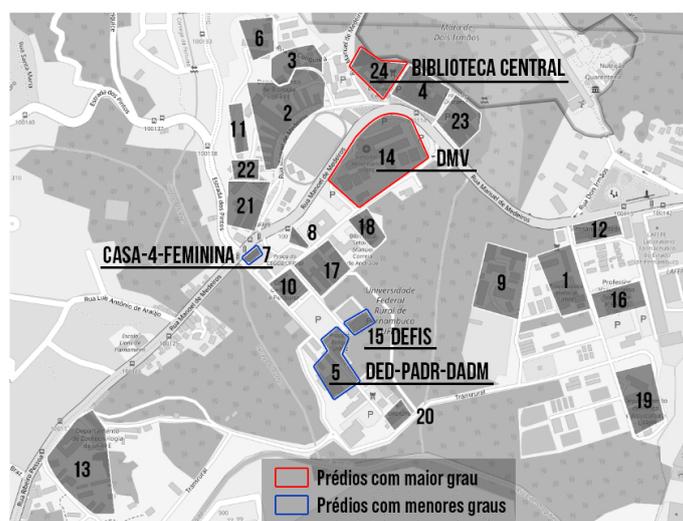


Figura 15. Localização dos prédios com maior grau e menores graus.

Os *Betweenness* médios da rede, gráfico (g) da Figura 16 mostram resultados com valores bem pequenos, o valor encontrado para a rede com limiar 0.6 da Terça-feira foi de 0.02. A quantidade de componentes, gráfico (h) da Figura 16, só apresenta alterações, para os dias com mais movimentos, quando o valor do limiar é de 0.9.

O tamanho médio dos caminhos, que calcula a média dos menores caminhos, é de 1.46, como mostra a Figura 16, o que significa que para se deslocar de um nó para qualquer outro é preciso se mover, em média, apenas uma vez.

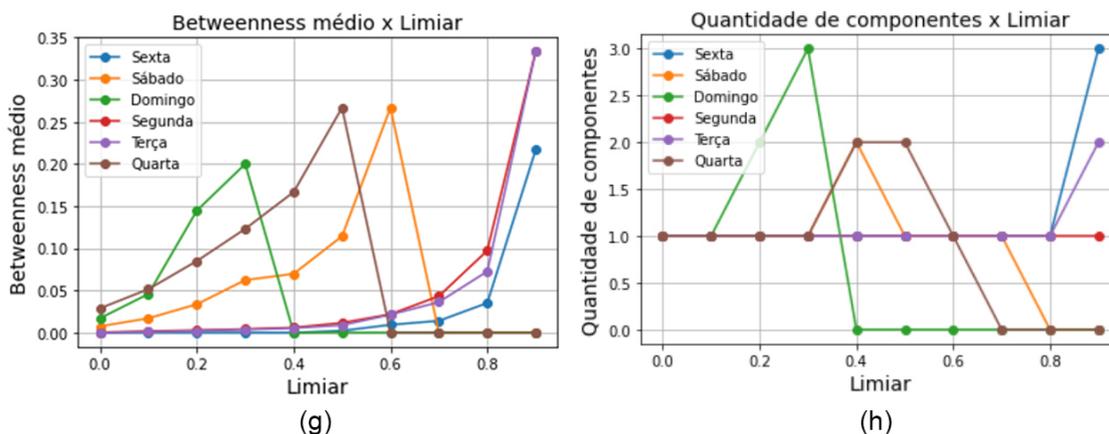


Figura 16. Gráficos com valores dos *Betweenness* médio (g) e da Quantidade de componentes (h) nas redes complexas de acordo o limiar para cada dia de captura.

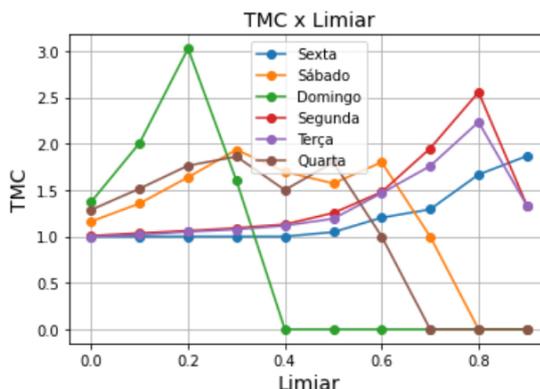


Figura 17. Gráficos da valores de TMC nas redes complexas de acordo o limiar para cada dia de captura.

5. Discussão

Os resultados das curvas de ocupação permitem entender melhor o funcionamento dos prédios, assim possibilitando um estudo sobre como melhorar a distribuição de cursos e/ou alunos pelo campus, além de auxiliar no planejamento da segurança do campus, uma vez que é possível conhecer os caminhos com maiores e menores tráfegos de pessoas.

Os formatos que as curvas de ocupação adotam mostram o que acontece nos prédios ao longo de um dia e indicam como estes estão sendo usados pela comunidade acadêmica. O quantitativo de pessoas em si, além dos momentos em que há aumento ou decréscimo súbito destas quantidades (tal como ocorreu no prédio DED-PADR-DADM) demandam um aprofundamento da investigação na busca de um melhor entendimento do comportamento do prédio naquele dia.

As redes complexas trouxeram resultados interessantes como os graus de prédios, contudo algumas métricas não apresentaram maiores informações sobre as ocupações dos prédios, como o Coeficiente de aglomeração, Tamanho médio dos caminhos e *Betweenness*. Normalmente estas medidas oferecem melhores *insights* acerca da rede quando

utilizadas em redes complexas cujo número de nós é muito maior do que o encontrado neste trabalho.

6. Conclusões e trabalhos futuros

O objetivo do trabalho relatado neste artigo foi de analisar dados de redes Wi-Fi por meio de redes de correlação. O método utilizado permitiu conferir as curvas de ocupação dos prédios ao longo de cada um dos dias de análises. Os resultados permitiriam um melhor entendimento da ocupação dos prédios de acordo com os horários dos dias, além de descobrir os horários dos prédios com os maiores registros de ocupação.

Concluimos também que analisar as redes complexas geradas a partir do valor da correlação entre os prédios nos permitiu identificar prédios com forte correlação necessariamente não possuem grau com alto valor, como conferido na Terça-feira, onde a CASA-4-FEMININA (grau 4) e o PC (grau 17), possuem forte correlação. Entretanto outras métricas não trouxeram resultados satisfatórios para uma análise mais profunda.

Para trabalhos futuros a ideia é trabalhar com pacotes de *Probe Request*, uma vez que possibilita analisar capturar dados de requisições de dispositivos móveis com acesso a rede sem fio em lugares com disponibilidade de rede Wi-Fi e não apenas para dispositivos já associados a rede. Outro ponto é realizar o trabalho em um rede Wi-Fi com muitos mais lugares que possam ser representados por um nó, assim métricas como *Betweenness* e Coeficiente de aglomeração vão possibilitar uma melhor análises da importância dos nós da rede complexa.

Referências

- Barabási, A.-L. (2021). The bridges of Königsberg. Disponível em: <http://networksciencebook.com/chapter/2#bridges>. Acesso em: 01.12.2021.
- Cisco (2009). *Perguntas frequentes sobre o Wireless LAN Controller (WLC)*. Disponível em: https://www.cisco.com/c/pt_br/support/docs/wireless/4400-series-wireless-lan-controllers/69561-wlc-faq.html. Acesso em: 19.11.2021.
- Cisco (2018). Estabelecer uma rede wireless usando um ponto de acesso wireless (o wap). Disponível em: http://www.cisco.com/c/pt_br/support/docs/smb/wireless/cisco-small-business-100-series-wireless-access-points/smb5530-set-up-a-wireless-network-using-a-wireless-access-point-wap.html. Acesso em: 19.11.2021.
- Cisco (2021). *Chapter: 802.11 Network Security Fundamentals*. Disponível em: http://www.cisco.com/c/en/us/td/docs/wireless/wlan_adapter/secure_client/5-1/administration/guide/SSC_Admin_Guide_5_1/C1_Network_Security.html. Acesso em: 14.11.2021.
- Fernando, d. S. M. (2021). *Pesquisa Anual do Uso de TI*. Disponível em: <https://eaesp.fgv.br/producao-intelectual/pesquisa-anual-uso-ti>. Acesso em: 08.12.2021.
- Kontokosta, C. E. and Johnson, N. (2017). Urban phenology: Toward a real-time census of the city using wi-fi data. *Computers, Environment and Urban Systems*, 64:144–153.

- L. Yang, P Zerfos, E. S. (2005). *Architecture Taxonomy for Control and Provisioning of Wireless Access Points (CAPWAP)*. Disponível em: <https://datatracker.ietf.org/doc/html/rfc4118>. Acesso em: 08.12.2021.
- Lewis, T. (2009). *Network Science: Theory and Applications*. Wiley.
- Lira, S. A. (2004). Análise de correlação: abordagem teórica e de construção dos coeficientes com aplicações. *Universidade Federal do Paraná*.
- Meraki, C. (2020). 802.11 association process explained. Disponível em: http://documentation.meraki.com/MR/WiFi_Basics_and_Best_Practices/802.11_Association_Process_Explained. Acesso em: 19.11.2021.
- Oliveira, L., Schneider, D., De Souza, J., and Shen, W. (2019). Mobile device detection through wifi probe request analysis. *IEEE Access*, 7:98579–98588.
- Paranhos, R., Figueiredo Filho, D. B., Rocha, E. C. d., Silva Júnior, J. A. d., Neves, J. A. B., and Santos, M. L. W. D. (2014). Desvendando os mistérios do coeficiente de correlação de pearson: o retorno. *Leviathan (São Paulo)*, (8):66–95.
- Sayama, H. (2015). *Introduction to the Modeling and Analysis of Complex Systems*. Open Textbook Library. Open SUNY Textbooks, Milne Library, State University of New York at Geneseo.
- Sevtsuk, A., Huang, S., Calabrese, F., and Ratti, C. (2008). Mapping the mit campus in real time using wifi.
- Silva, F. C. d. S. (2017). Rede de correlações, eficiência da seleção visual e indireta via capacidade de ramificação e índices de seleção em soja.
- Souza, D. B., da Cunha, J. T. S., dos Santos, E. F., Correia, J. B., da Silva, H. P., de Lima Filho, J. L., Albuquerque, J., and Santos, F. A. N. (2021). Using discrete ricci curvatures to infer COVID-19 epidemic network fragility and systemic risk. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(5):053501.
- Traunmueller, M. W., Johnson, N., Malik, A., and Kontokosta, C. E. (2018). Digital footprints: Using wifi probe and locational data to analyze human mobility trajectories in cities. *Computers, Environment and Urban Systems*, 72:4–12.
- Wang, Li-Na, T. G.-M. Z. C.-R. (2021). A network method to identify the dynamic changes of the data flow with spatio-temporal feature. *Applied Intelligence*.