

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
CAMPUS DE SÃO JOÃO DA BOA VISTA

JULIANA SEMÍRAMIS MENZINGER

**Otimização de redes de acesso ópticas em regiões
rurais utilizando agrupamento multinível**

São João da Boa Vista

2022

Juliana Semíramis Menzinger

**Otimização de redes de acesso ópticas em regiões
rurais utilizando agrupamento multinível**

Trabalho de Graduação apresentado ao Conselho de Curso de Graduação em Engenharia Eletrônica e de Telecomunicações do Campus de São João da Boa Vista, Universidade Estadual Paulista, como parte dos requisitos para obtenção do diploma de Graduação em Engenharia Eletrônica e de Telecomunicações .

Orientador: Profº Dr. Ivan Aritz Aldaya Garde

Coorientador: Profº Dr. Rafael Abrantes Penchel

São João da Boa Vista

2022

M551o	Menzinger, Juliana Semíramis Otimização de redes de acesso ópticas em regiões rurais utilizando agrupamento multinível / Juliana Semíramis Menzinger. -- São João da Boa Vista, 2022 32 p. : il., tabs. Trabalho de conclusão de curso (Bacharelado - Engenharia de Telecomunicações) - Universidade Estadual Paulista (Unesp), Faculdade de Engenharia, São João da Boa Vista Orientador: Ivan Aritz Aldaya Garde Coorientador: Rafael Abrantes Penchel 1. Telecomunicações Tráfego. 2. Telecomunicações. 3. Comunicação. 4. Planejamento da comunicação. I. Título.
-------	--

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Engenharia, São João da Boa Vista. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE ENGENHARIA
CAMPUS DE SÃO JOÃO DA BOA VISTA
GRADUAÇÃO EM ENGENHARIA ELETRÔNICA E DE TELECOMUNICAÇÕES

TRABALHO DE CONCLUSÃO DE CURSO

OTIMIZAÇÃO DE REDES DE ACESSO ÓPTICAS EM REGIÕES RURAIS
UTILIZANDO AGRUPAMENTO MULTINÍVEL

Aluna: Juliana Semiramis Menzinger
Orientador: Prof. Dr. Ivan Aritz Aldaya Garde

Banca Examinadora:

1. Ivan Aritz Aldaya Garde (Orientador)
2. José Augusto de Oliveira (Examinador)
3. Marcelo Luis Francisco Abbade (Examinador)

A ata da defesa com as respectivas assinaturas dos membros encontra-se no prontuário do aluno (Expediente nº 008/2020)

São João da Boa Vista, 25 de fevereiro de 2022

*"O aprendizado é contínuo e o próximo nível nunca será o último."
(Autor desconhecido)*

AGRADECIMENTOS

Primeiramente ao meu pai, minha mãe e minhas irmãs que tanto me ajudaram, dando-me as melhores oportunidades de estudo mesmo quando passávamos por os problemas financeiros, cada suporte foi essencial para que eu chegasse até aqui, sem eles nada disso seria possível. Eles são o motivo pelo qual estou aqui e pelo o qual eu irei mais longe.

A minha vó Eulalia, por todo cuidado e orações que fez por mim para que eu me mantivesse firme perante as dificuldades, a minha tia Henriete e meus primos que sempre torceram muito por mim e me deram animo quando precisei.

Aos professores e a todos os funcionários da UNESP que fizeram parte da minha educação durante a graduação, principalmente ao meu orientador Prof. Dr. Ivan Aritz Aldaya Garde e meu coorientador Prof. Dr. Rafael Abrantes Penchel, que me orientaram, incentivaram e me aconselharam da melhor forma, tanto neste trabalho de conclusão de curso como na minha vida profissional.

Ao Processo nº 2017/05988-5, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Auxílio à Pesquisa - Modelo físico de linha de transmissão com componentes de alto fator de qualidade.

Ao CNPq pelo apoio financeiro para que fosse possível realizar a pesquisa de iniciação científica que deu origem a este trabalho de conclusão de curso.

A Siemens e a todos meus colegas de trabalho, em especial ao Alexandre Rodrigues da Silva e ao Paulo Eduardo Galvão que além de todo o suporte durante o meu estagio me deram todo suporte necessário para que eu concluísse este trabalho de conclusão de curso. E agradeço também a todos aqueles que fizeram parte da minha graduação e que de certa forma me ajudaram quando precisei.

RESUMO

Neste trabalho estudou-se a otimização de uma rede de acesso óptica em regiões rurais. Utilizou-se o agrupamento heurístico k-means com a topologia estrela multinível neste estudo, por acreditar-se que este tipo de agrupamento é ideal em um projeto como este que visa levar fibra óptica em regiões mais afastadas com maior qualidade possível e com um custo mais baixo. Para isso simulou-se através do Software Matlab um cenário ideal e um cenário real, no cenário ideal gerou-se através do software coordenadas aleatórias afim de obter dados suficientes para o estudo do caso, em um segundo momento no cenário ideal as coordenadas deixam de ser geradas aleatoriamente e recebem dados reais coletados da região de Nazaré Paulista, compreendia entre (23° 9' 20.74"S e 46° 16' 15.85"O). Em ambos os casos estudou-se a clusterização formada através dos pontos (usuários), a topologia criada (nodo divisor) e o quanto de potência é perdida na fibra, nas divisões e a total.

PALAVRAS CHAVE: COMUNICAÇÕES ÓPTICAS; K-MEANS; CLUSTERIZAÇÃO; FIBRA ÓPTICA; MULTINÍVEL; REDES ÓPTICAS PASSIVAS.

ABSTRACT

In this work, the optimization of an optical access network in rural regions was studied. Use the k-means heuristic clustering with the multilevel star topology in this study, because it is believed that this type of clustering is ideal in a project like this one that aims to take optical fiber to more distant regions with the highest possible quality and at a lower cost . For this, an ideal scenario and a real scenario were simulated through the Software Matlab, in the ideal scenario random coordinates were generated through the software in order to obtain enough data for the case study, in a second moment in the ideal scenario as coordinates no longer be randomly generated and provide real data collected from the region of Nazaré Paulista, included between (23° 9'20.74"S e 46°16'15.85"O). In both cases, the clustering formed through the points (users), a created topology (divisor node) and how much power is lost in the fiber, in the divisions and the total was studied.

KEYWORDS: OPTICAL COMMUNICATIONS; K-MEANS; CLUSTERING; OPTICAL FIBER; MULTILEVEL; PASSIVE OPTICAL NETWORKS.

LISTA DE ILUSTRAÇÕES

Figura 1	Topologia típica de uma rede PON incluindo a OLT, o nó divisor e as ONUs. . .	13
Figura 2	Topologia em estrela de um único nível e de múltiplos níveis	14
Figura 3	Diagrama de fluxo do algoritmo de Lloyd.	18
Figura 4	Resultado após cada iteração do algoritmo de Lloyd.	18
Figura 5	Representação gráfica do processo de clusterização multi-nível. (a) Posição dos elementos a clusterizar. (b) Primeiro nível de clusterização. (c) Segundo e último nível de clusterização.	19
Figura 6	Posição das ONUs e histogramas bidimensionais da densidade de ONUs para (a-b) 100 usuários, (b-c) 200 usuários, (d-e) 400 usuários.	21
Figura 7	Posição das ONUs consideradas para a avaliação do algoritmo num cenário realista (Fonte: Google Maps).	23
Figura 8	Posição das ONUs antes de aplicar o algoritmo	24
Figura 9	Rede gerada após aplicação do algoritmo, indicando as ONUs, os nó divisor e os enlaces que os unem.	25
Figura 10	Perdas da conexão entre o nó central e cada ONU.	25
Figura 11	<i>Box plots</i> do comprimento total e das perdas médias para 100 cenários aleatoriamente gerados para (a-b) 100 ONUs, (c-d) 200 ONUs e (e-f) 400 ONUs.	27
Figura 12	Perdas médias vs. comprimento total de fibra instalada para 100 cenários arbitrariamente gerados considerando (a) 100, (b) 200, e (c) 400 ONUs.	28
Figura 13	Posição das ONUs reais antes de aplicar o algoritmo	29
Figura 14	Rede gerada após aplicação do algoritmo real	29
Figura 15	Perdas de cada ONU para o cenário realista.	30
Figura 16	a)Box plot do comprimento de fibra instalado, b)Box plot da perda média . . .	30
Figura 17	Perdas médias vs comprimento para um cenário real	31

LISTA DE ABREVIATURAS E SIGLAS

PON - *Passive Optical Network*

OLT - *Optical Line Terminal*

ONU - *Optical Network Unit*

EDFAs - *Amplificadores ópticos a fibra dopada com érbio*

WDM - *Multiplexação por comprimento de onda*

PCM - *Pulse Code Modulation*

GaAs - *Arseneto de Gálio*

InP - *Fosfeto de índio*

InGaAsP - *Fosfeto de arsênio de gálio indium*

LISTA DE SÍMBOLOS

B - Taxa de bits

L - Espaçamento entre repetidores

λ - Comprimento de onda

P_{in} - Potência de entrada

d_i - Distância inicial

S - Sudeste

O - Oeste

Lat - Latitude

Lon - Longitude

R - Raio da Terra

M - Matriz Central

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Evolução histórica	12
1.2	Redes Ópticas Passivas	13
1.3	Justificativa do trabalho	14
1.4	Contribuição	14
1.5	Organização do documento	14
2	OTIMIZAÇÃO DE REDES PON MULTINÍVEL BASEADA EM K-MEANS	16
2.1	Algoritmo de otimização para redes PON multi-nível	16
2.1.1	Aprendizado de máquinas para otimização de topologia de redes	16
2.1.2	Clusterização mediante k-means	17
2.1.3	Aplicação de k-means para otimização de redes PON multi-nível	18
2.2	Geração de cenários	19
2.2.1	Cenários aleatórios	19
2.2.2	Cenário realista	20
2.3	Métricas de avaliação	20
2.3.1	Comprimento total de fibra instalada	20
2.3.2	Perda média	23
3	RESULTADOS DE SIMULAÇÃO	24
3.1	Resultados com cenários gerados de forma aleatória	24
3.1.1	Análise de um caso particular	24
3.1.2	Análise estatístico do conjunto de cenários	26
3.2	Resultados com cenário realista	26
3.2.1	Análise de uma solução particular do cenário realista	26
3.2.2	Análise estatística das soluções para o cenário realista	26
4	CONCLUSÕES E TRABALHO FUTURO	32
	REFERÊNCIAS	33

1 INTRODUÇÃO

1.1 EVOLUÇÃO HISTÓRICA

O início da pesquisa em comunicação por fibra óptica teve início na década dos anos 1970s (AGRAWAL, 2014). O progresso das comunicações ópticas pode ser dividido em cinco gerações. Na primeira geração ocorre um aumento do produto taxa de bits-distância, $B \times L$, em que B é a taxa de transmissão de bits e L o espaçamento entre repetidores, que começa a saturar à medida que a tecnologia avança. A primeira geração dos sistemas de comunicações ópticas usava *lasers* semicondutor arseneto de gálio (GaAs) e operava na faixa de 0,8 μm . Esses sistemas atingiam uma taxa de transmissão de 45 Mb/s e permitiam espaçamento entre repetidores de até 10 km (AGRAWAL, 2014). Embora para os padrões atuais, o produto $B \times L$ resultante pareça muito baixo, antes da primeira geração o maior espaçamento entre repetidores nos sistemas baseados em cabo coaxial era de 1 km. Portanto, a introdução desta primeira geração representou um avanço importante para a tecnologia. Com a finalidade de aumentar o produto capacidade-comprimento a segunda geração dos sistemas de comunicações ópticas usava *lasers* de Fosfeto de índio (InP) capazes de emitir a um comprimento de onda de 1,3 μm , região em que o coeficiente de atenuação da fibra óptica é menor do que 1 dB/km. Os sistemas de segunda geração conseguiam taxas de transmissão de até 1,7 Gb/s e possuíam uma separação entre repetidores de cerca de 50 km (AGRAWAL, 2014). Com a finalidade de reduzir a perda de transmissão na fibra óptica, a terceira geração de sistemas de ondas luminosas, os sistemas da terceira geração migraram para o comprimento de onda de 1.55 μm , atingindo taxas de transmissão de 2,5 Gb/s. O melhor desempenho era alcançado pelo emprego de fibras de dispersão deslocadas em combinação com *lasers* de InGa-AsP. Porém, mesmo operando no comprimento de onda em que a fibra apresenta menor atenuação, os sistemas de terceira geração ainda precisavam de repetidores optoeletrônicos espaçados, tipicamente, por 60 km ou 70 km (AGRAWAL, 2014). A quarta geração de sistemas de comunicação utilizavam amplificadores ópticos a fibra dopada com érbio (EDFAs, do inglês *erbium-doped fiber amplifiers*), para amplificar simultaneamente múltiplos canais na janela de 1.55 μm multiplexados em comprimento de onda (WDM, do inglês *wavelength division multiplexing*), aumentando assim a taxa de transmissão total por fibra (AGRAWAL, 2014). A quinta geração, atualmente utilizada em sistemas de longa distância e metropolitanos, faz uso de receptores coerentes digitais para aumentar a eficiência espectral. O foco dos atuais sistemas da quinta geração é o aumento desta eficiência espectral de sistemas WDM. Mesmo apresentando melhoraria na sensibilidade do receptor, os sistemas de quinta geração continuam requerendo EDFAs, ao menos nos sistemas de longa distância e metropolitanos. (AGRAWAL, 2014).

Observa-se que cada nova geração trouxe modificações fundamentais que ajudaram a melhorar o desempenho dos sistemas das comunicações ópticas, reduzindo desta forma os custos de instalação e manutenção associada as fibras e a cada repetidor. Levando em consideração o avanço da comunicação óptica onde as redes de longa distância se estenderam para as redes metropolitanas e posteriormente para as redes de acesso, tem-se que as redes de acesso possuem restrições em relação ao custo. Com isto grande parte das redes de acesso implantadas continuam se baseando em detecção direta. Com a finalidade de minimizar o custo de manutenção da rede de distribuição em sistemas de acesso, grande parte destes não incluem elementos ativos, implementando as denominadas redes ópticas passivas (PONs, do inglês *optical passive networks*). Desta forma, as PONs tornaram-se viáveis economicamente para atender clientes que se encontram distante das centrais de atendimento, apresentando desta forma demandas para altas taxas de transmissão de dados e diferentes tipos de serviços. Levando em consideração o conceito de redes econômicas, é de suma importância que uma boa parte do investimento em infraestrutura seja focada na parte da rede óptica passiva que chega até o usuário final. Embora as PONs sejam comumente utilizadas em regiões urbanas, com a utilização da clusterização multinível K-means é possível utilizar as redes ópticas passivas também em regiões rurais, apesar dos entornos rurais apresentarem o problema de uma maior distância e uma distribuição

irregular dos usuários. (AGRAWAL, 2014).

1.2 REDES ÓPTICAS PASSIVAS

Como mencionado anteriormente, as redes ópticas podem ser classificadas em três categorias principais: redes baseadas em enlaces de longa distância, redes de área metropolitana, e PONs. O conceito de PON foi introduzido em meados dos anos 1980s na Inglaterra (BARBOSA; METZ; FERNANDES, 2020). O termo "rede óptica passiva" deve-se, principalmente, ao fato das PONs não contarem com elementos ativos que precisam de energia elétrica no seu funcionamento, utilizam-se assim fibras ópticas, sem amplificadores ópticos. Geralmente as redes PONs também não contem dispositivos de chaveamento e, por conseguinte, não dispõem de proteção como outras redes. A maioria das redes PONs possui uma topologia estrela, isto é uma configuração ponto-multiponto, em que um nó central denominado terminal de enlace óptico (OLT, do inglês *optical link terminal*) distribui informação para múltiplos usuários denominados unidades de rede ópticas (ONUs, do inglês *optical network units*). Na configuração básica, mostrada na Figura 1, a OLT transmite a informação comum para todos os usuários utilizando uma única fibra a um nó divisor, o qual tem enlaces dedicados para cada usuário. Desta forma, numa parte da rede, o sinal óptico é compartilhado entre todos os usuários e interligados por essa rede, reduzindo, assim, a quantidade de equipamentos e o comprimento de fibra instalado, além de possuir um número reduzido de eventuais intervenções de manutenção, sendo assim vantajoso por ser capaz de ser menos dispendioso (BARBOSA; METZ; FERNANDES, 2020).

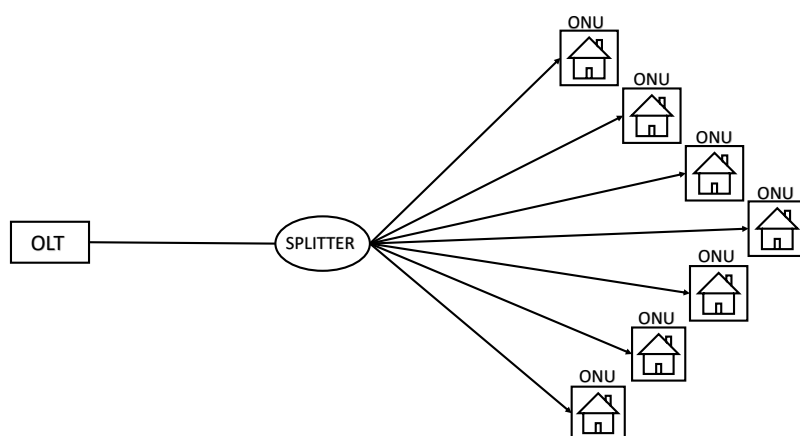


Figura 1 – Topologia típica de uma rede PON incluindo a OLT, o nó divisor e as ONUs.

O comprimento da fibra instalada é algo que deve ser levado muito em consideração em relação ao conceito das PONs e, portanto, ser algo que deve ser otimizado, pois representa um dos fatores de maior impacto no custo total de implantação. A topologia em estrela é geralmente usada em redes de acesso óptico, nas quais um único nó distribui sinais para diferentes nós. Porém, quando o número de ONUs é grande, uma topologia em estrela simples pode não atender aos requisitos do custo baixo desejado. Com isso a evolução da topologia em estrela simples para a topologia em estrela de vários estágios é natural. Na Figura 2 apresentamos um exemplo de topologia em estrela de um único nível e de múltiplos níveis. É importante observar que a flexibilidade fornecida pelo uso de particionamento de vários níveis torna o projeto de rede mais desafiador, porque a localização ideal do nó divisor não é trivial.

Nas PONs, as derivações são feitas por meio do uso de divisores ópticos passivos, comumente chamado de *splitter*, para conectá-la às ONUs. O *splitter* é um elemento passivo que realiza a divisão do sinal óptico proveniente de uma fibra para diferentes valores de divisão desta fibras, em razões usuais de 1:2, 1:4, 1:8, 1:16, 1:32 e 1:64 (TAKEUTI, 2005).

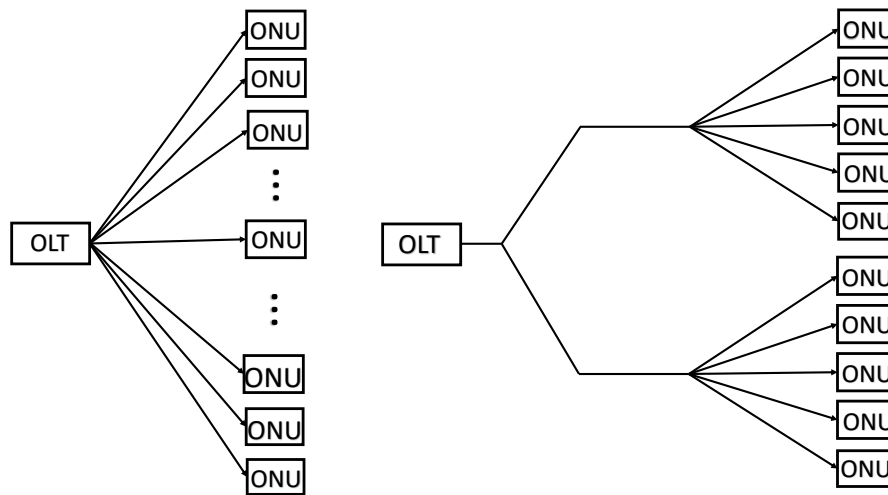


Figura 2 – Topologia em estrela de um único nível e de múltiplos níveis

1.3 JUSTIFICATIVA DO TRABALHO

A principal motivação para a realização deste trabalho é poder ser capaz de ajudar a sociedade através dos conhecimentos adquiridos durante a graduação. Atualmente no Brasil as regiões mais afastadas não possuem a infraestrutura necessária para que uma parcela da sociedade tenha acesso a fibra óptica, portanto, pensou-se em um trabalho capaz de levar fibra óptica até os moradores dessas regiões mais afastadas.

Para isto, é necessário realizar um trabalho que seja capaz de levar fibra óptica de forma menos dispendiosa possível, visto que em regiões rurais existe uma baixa densidade de usuários. Já existem alguns algoritmos capazes de projetar redes regulares, como é o caso por exemplo do *grid* Manhattan regular, que ajudam a projetar de forma eficiente redes passivas em entornos urbanos e suburbanos. O projeto de redes em entornos urbanos de alta densidade difere significativamente do projeto de redes em entornos rurais. Por um lado, em regiões urbanas, os usuários tendem a estar mais uniformemente distribuídos. Por outro lado, em regiões rurais, esses usuários não são uniformemente distribuídos. Devido a isto a métrica de distância utilizada em entornos urbanos é geralmente a distância Manhattan, e em entornos rurais a mais adequada é a distância euclidiana.

1.4 CONTRIBUIÇÃO

Neste trabalho desenvolveu-se um algoritmo que permite achar a posição dos nós divisores, para um cenário aleatório e para um cenário real. Para isso, baseou-se na clusterização k-means, com agrupamento multi-nível. Este algoritmo é capaz também de calcular o uso total da fibra e potência perdida ao longo da fibra e em cada nó divisor. Além disso, por meio de um estudo estatístico analisaremos o efeito da razão de divisão nos nós divisores nas métricas de desempenho da rede, possibilitando otimizar este parâmetro.

1.5 ORGANIZAÇÃO DO DOCUMENTO

Este trabalho de conclusão de curso está estruturada em quatro capítulos; No Capítulo 1 encontram-se a justificativa e objetivos do trabalho, a evolução histórica dos sistemas de comunicações baseados em fibras ópticas e o início das PONs, bem como a contribuição e a estrutura de apresentação deste documento. No Capítulo 2, explica-se o algoritmo para a arquitetura da rede com topologia em estrela multi-nível, assim como as métricas a serem utilizadas para avaliar a performance da arquitetura da rede. O Capítulo 3 apresenta-se os resultados numéricos obtidos usando cenários sintetizados artificialmente

e um cenário realista. Finalmente, no Capítulo 4 apresentamos as conclusões e indicamos alguns possíveis trabalhos futuros.

2 OTIMIZAÇÃO DE REDES PON MULTINÍVEL BASEADA EM K-MEANS

No presente capítulo apresenta-se e descreve-se as ferramentas necessárias para a otimização de redes PON multi-níveis. Na Secção 2.1 introduz-se os conceitos básicos de aprendizado de máquinas, particularizando-se para o caso do método de clusterização *k-means*. Este método de clusterização é utilizado como base para implementar um algoritmo que calcula as localizações dos nós distribuidores de uma rede PON multi-nível. A geração dos cenários para a avaliação do algoritmo proposto é descrita na Secção 2.2. Finalmente, na Secção 2.3, apresenta-se as métricas utilizadas para avaliar o desempenho do algoritmo.

2.1 ALGORITMO DE OTIMIZAÇÃO PARA REDES PON MULTI-NÍVEL

2.1.1 Aprendizado de máquinas para otimização de topologia de redes

Primeiramente, é importante introduzir alguns conceitos básicos de aprendizado de máquina, assim como algumas definições importantes. O aprendizado de máquina compreende um conjunto de algoritmos baseados em estatísticas que podem identificar padrões de distribuição de dados (ALPAYDIN, 2010). Apesar de atualmente este tema ser uma tendência global, ele tem sido estudado desde 1959. Em um artigo publicado nesse ano de 1959, foi proposto um método para que os computadores aprendam a jogar xadrez (SAMUEL, 1959). A popularidade atual se deve aos avanços tecnológicos nas áreas de eletrônica e telecomunicações e, mais especificamente, computação distribuída. O aprendizado ocorre em um processo denominado treinamento, em que estes algoritmos geralmente iteram para achar os parâmetros do modelo que melhor fitam o sistema a emular. O objetivo dessa otimização é modelar os padrões existentes no conjunto de dados, chamados de exemplos de treinamento.

Existem muitos exemplos de uso de algoritmos de aprendizado de máquina na vida diária, como mecanismos de pesquisa para encontrar os melhores resultados, bem como anúncios para usuários individuais, filtros usados em e-mails, reconhecimento facial durante a filmagem e reconhecimento de voz em telefones celulares (SHALEV-SHWARTZ; BEN-DAVID, 2014). Nestas aplicações, é vantajoso usar tais algoritmos, pois com tais algoritmos é possível achar soluções aceitáveis a problemas de alta complexidade de forma mais eficiente que outros métodos clássicos de otimização (SHALEV-SHWARTZ; BEN-DAVID, 2014). Em particular, o aprendizado de máquinas permite resolver problemas em que a dinâmica subjacente é desconhecida ou, mesmo sendo conhecida, resulta muito complexa de realizar um algoritmo de simulação.

Os métodos de aprendizado de máquinas podem ser classificados dependendo do domínio da variável ou variáveis de saída. Portanto, se a saída tem um domínio contínuo, diz-se que o método é de regressão. Se o domínio da variável ou variáveis de saída é discreto, então o método é denominado de classificação. A filtragem de spam é um exemplo de classificação, em que as entradas são as mensagens de *emails* (ou outros) e as classes são *spam* ou *não spam*. Enquanto à regressão, alguns exemplos são as previsões do preço de um imóvel, da demanda de venda, a probabilidade de superlotação de um estoque, o tempo que levará para uma máquina apresentar defeito, uma nota de um cliente para aumentar seu limite de crédito, ou qualquer outro valor quantitativo contínuo. Outro critério de classificação dos algoritmos de aprendizado de máquinas é o tipo de aprendizado. Em sistemas relativamente simples, existem dois tipos de aprendizado: supervisionada e não supervisionada. Na aprendizagem supervisionada, os dados introduzidos pelo algoritmo são rotulados, ou seja, são conhecidos os valores desejados dos exemplos de treinamento. No caso de aprendizagem não supervisionada, a saída dos dados de treinamento não é conhecida *a priori*. Este último tipo de algoritmo é utilizado, não somente quando não se tem informação mas também em casos onde não existe uma estrutura subjacente dos dados, portanto os resultados serão agrupados de acordo com a similaridade dos dados (SHALEV-SHWARTZ; BEN-DAVID, 2014).

Um tipo particularmente importante para o desenvolvimento do presente trabalho são os métodos de classificação não supervisionada, também denominados de algoritmos de clusterização. Estes algoritmos classificam os dados em grupos chamados *clusters* com base na similaridade das características de dados. Os algoritmos de clusterização, por sua vez, podem ser divididos em algoritmos baseados em distância, como k-means, ou baseados em densidade como clusterização espacial baseada em densidade de aplicações com ruído (DBSCAN, do inglês *Density-Based Spatial Clustering of Applications with Noise*) ou clusterização baseada em histograma (HBC, do inglês *Histogram-based clustering*). Os algoritmos mencionados são exemplos de clusterização não paramétrica, pois não requerem achar os parâmetros das distribuições subjacentes. Por outro lado, maximização de expectativa (EM, do inglês *expectation maximization*) é um método de clusterização em que os parâmetros de um conjunto de distribuições gaussianas são otimizados de forma iterativa.

Dos algoritmos mencionados, k-means apresenta certas vantagens que são desejáveis para a implementação de uma rede de múltiplos níveis. Por um lado, k-means minimiza a distância a certos pontos, que é interessante pois esses pontos podem ser considerados como concentradores/divisores de cada *cluster*. Por outro lado, em k-means o número de agrupamentos é estabelecido, permitindo assim controlar a razão de divisão. Porém, o algoritmo k-means considera o agrupamento em um único nível, o qual faz necessário a modificação deste para considerar múltiplos níveis.

2.1.2 Clusterização mediante k-means

Antes de descrever o algoritmo utilizado para encontrar a arquitetura multi-nível, descreveremos o algoritmo k-means original. K-means é um algoritmo de agrupamento de dados iterativo para particionar um conjunto de dados. Ele foi proposto num trabalho pioneiro de S. Lloyd em 1959 como uma técnica para modulação de código de pulso (PCM, do inglês *Pulse Code Modulation*), porém só foi publicado no ano de 1982. Esse algoritmo busca minimizar a distância dos elementos de um conjunto de dados a k centroides de forma iterativa (LLOYD, 1982).

Existem diferentes implementações de K-means, sendo a mais popular, o algoritmo de Lloyd. Este algoritmo é composto por uma primeira etapa em que os centroides, pontos de inicialização onde os demais dados serão agrupados, são inicializados. Existem vários tipos de inicialização, sendo a aleatória a mais comumente utilizada. É importante notar que a inicialização aleatória confere a k-means uma natureza heurística, isto é, para o mesmo conjunto de dados, sucessivas aplicações do algoritmo podem resultar em soluções diferentes. Uma vez os centroides tenham sido inicializados, o algoritmo itera entre duas etapas: a etapa de atribuição, na qual o conjunto de dados é particionado associando os dados aos centroides mais próximos considerando uma determinada norma, geralmente a distância euclidiana. A etapa de atualização é responsável por atualizar os centroides calculando o centro de gravidade de cada um dos centroides. O algoritmo converge quando ao aplicar uma atualização dos centroides não tem valores diferentes dos já obtidos para a posição destes. Desta forma, o algoritmo k-means realiza uma partição dos N dados em k agrupamentos. Cada agrupamento está representado por o seu centro de massa, denominado centroide, de forma que todos os elementos estão associados ao centroide mais próximo conforme a uma norma determinada (a norma utilizada neste trabalho é a a norma euclidiana). Em k-means, as posições dos centroides são otimizadas para minimizar a distância total entre os centroides e os elementos das agrupações associados a eles (PEREIRA; JAMHOUR, 2021). A Figura 3 representa o diagrama de fluxo do algoritmo de Lloyd.

Com a finalidade de entender melhor o funcionamento do algoritmo de Lloyd, na Figura 4 apresenta-se os resultados do algoritmo após cada iteração até convergir. Com intuito de facilitar a compreensão, considera-se uma clusterização com apenas duas classes. O resultado do algoritmo é indicando cada classe numa cor diferente, enquanto que as posições dos centroides são identificadas com marcadores x. Como pode se observar, a posição dos centroides é atualizada até alcançar as posições que minimizam a distância total até o conjunto de dados associado a cada *cluster*.

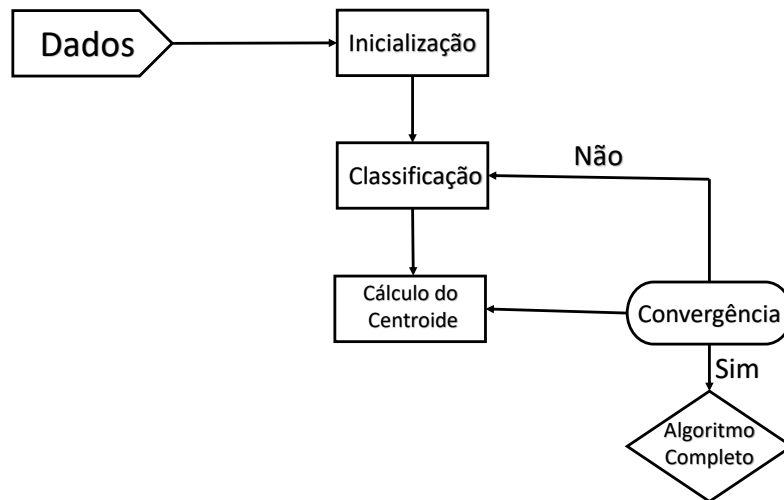


Figura 3 – Diagrama de fluxo do algoritmo de Lloyd.

2.1.3 Aplicação de k-means para otimização de redes PON multi-nível

Como temos descrito, a clusterização k-means divide o conjunto de dados em k -subconjuntos, cada um deles associado a um centroide. Este método pode ser utilizado para achar, por exemplo o nó distribuidor numa rede em estrela de um único nível, mas pode ser que não seja possível indicar os nós distribuidores em redes em estrela multi-níveis. Para projetar redes multi-nível, precisamos desenvolver um algoritmo que considere não somente um nível mas que se adapte a múltiplos níveis. No nosso caso, basearemos o algoritmo multi-nível na clusterização k-means pelas propriedades previamente indicadas.

O algoritmo desenvolvido aplica k-means sucessivamente, clusterizando cada *cluster* em menores *clusters* até que cada *cluster* resultante tenha menos de k elementos. Assim, considerando um conjunto de dados $x = \{p_1, p_2, \dots, p_n\}$, estes são divididos inicialmente em k *clusters* formados por n_i elementos, sendo $1 \leq i \leq k$. Cada um destes *clusters* é subdividido em outros k clusters. Desta forma, cria-se uma estrutura composta pelos centroides dos sucessivos processos de clusterização que correspondem aos nós divisores dos diferentes níveis. O processo de clusterização hierárquico ou multi-nível é representado graficamente na Figura 5. Na Figura 5(a) representam-se os elementos (ONUs) a ser conectadas. Após a primeira clusterização, estes elementos são divididos em quatro grupos, os quais são mostrados na Figura 5(b). Posteriormente, cada grupo é clusterizando chegando ao resultado final representado na Figura 5(c).

Um dos parâmetros mais importantes de uma estrutura multi-nível é a denominada razão de divisão,

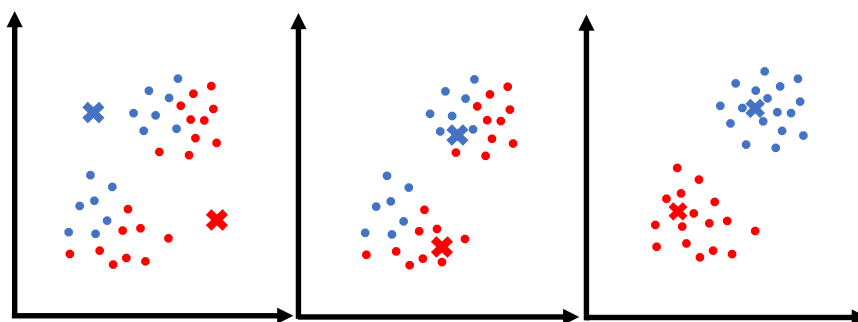


Figura 4 – Resultado após cada iteração do algoritmo de Lloyd.

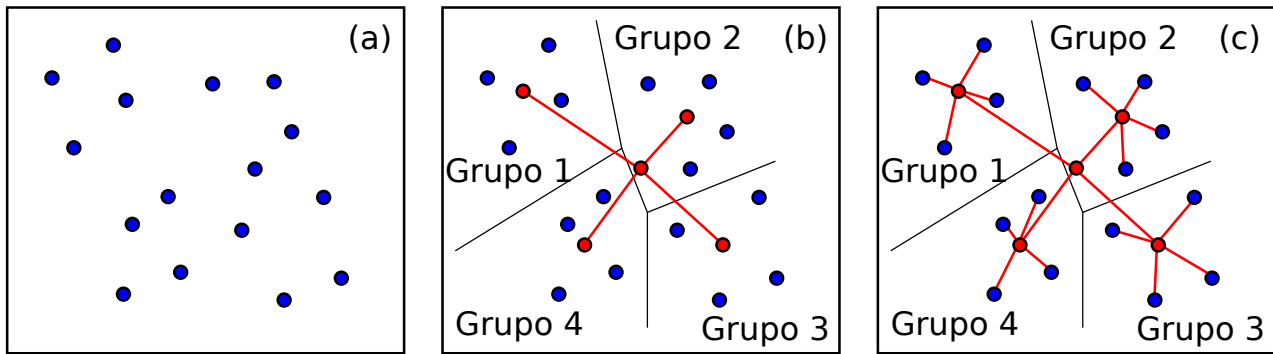


Figura 5 – Representação gráfica do processo de clusterização multi-nível. (a) Posição dos elementos a clusterizar. (b) Primeiro nível de clusterização. (c) Segundo e último nível de clusterização.

isto é, a quantos do nível inferior é conectado um centroide de um determinado nível. Este parâmetro pode ser controlado por meio do valor de k utilizado nas clusterizações k -means. Para aplicações práticas, o valor de k assume um valor de potência de 2 pois os divisores de feixe ópticos que são mais comumente utilizados no mercado, tem uma entrada e uma potência de 2 de saídas. Cabe também mencionar que poderiam-se utilizar diferentes razões de divisão para os diferentes níveis e, inclusive, dentro de cada nível.

2.2 GERAÇÃO DE CENÁRIOS

Para avaliar o desempenho do algoritmo, utilizou-se o software MATLAB da empresa Mathworks para a criação de cenários aleatórios e a implementação do cenário real. Este é um software voltado para cálculo numérico e otimizado para operar com matrizes.

2.2.1 Cenários aleatórios

Para a criação de um cenário aleatório, determinou-se o número de usuários, o número de áreas de alta densidade, a razão entre o número de usuários e o número de áreas de alta densidade, assim como o controle de espalhamento dos centros e dos usuários entorno dos centros. Para isso, selecionam-se os centros de alta densidade de forma aleatória. Uma vez criados os centros das regiões de alta densidade, geraram-se os usuário associados a cada uma dessas regiões considerando uma distribuição gaussiana bidimensional com parâmetros de dispersão previamente configurados. Para este trabalho consideramos 4 áreas de alta densidade com os parâmetros listados na Tabela 1. Após a determinação destas variáveis fez-se a normalização do cenário encontrado e depois realizou-se o plot do cenário normalizado e seu histograma. Para o escalonamento da área e a normalização, encontrou-se os valores máximos e mínimos das coordenadas de X e Y, estipulou-se uma área total na qual deseja-se ter como parâmetro para escalonar e normalizar ,depois disto, utilizando a formula com valores máximos e mínimos escalonou-se e normalizou-se.

Região de alta densidade	Número de usuários	Dispersao (σ)
I	25	2
II	25	2
III	25	2
IV	25	2

Tabela 1 – Parâmetros utilizados para a geração dos cenários aleatórios.

A fim de poder extrair uma informação geral sobre o desempenho do algoritmo selecionado, gerou-se um conjunto de cenários e aplicou-se o algoritmo a cada um deles. Nesta fase do trabalho o foco é na criação de cenários aleatórios, para diferentes números de usuários e diferentes parâmetros criou-se cenários aleatórios diferentes, com isso manteve-se os parâmetros gerais mas utilizou-se diferentes número de usuários, com o intuito de visualizar a rede formada e analisar o melhor nó divisor para diferentes situações, portanto com isso obteve-se os cenários gerados para 100, 200 e 400 usuários, repetindo o código apenas uma vez e representando o resultado deste cenário criado e o seu histograma gerado apenas uma vez também, com intuito de neste primeiro momento apenas visualizar os cenários criados. Abaixo encontram-se os cenários aleatórios criados para 100, 200 e 400 usuários e seus respectivos histogramas. A Figura 6 mostra exemplos de cenários aleatórios junto com os histogramas bidimensionais representando a densidade de usuários para (a-b) 100 usuários, (c-d) 200 usuários e (e-f) 400 usuários.

2.2.2 Cenário realista

Além dos cenários aleatórios utilizados para extrair informação geral sobre o algoritmo desenvolvido, aplicou-se o algoritmo a um cenário realista. A partir da determinação de um ponto inicial, compreendido na região de Nazaré Paulista/SP com Latitude: 23°9'20.74"S e Longitude: 46°16'15.85"O, realizou-se mais 99 pontos ao entorno, totalizando 100 pontos que seriam nossas ONUs (usuários). Com todos esses pontos determinados foi coletado e armazenado a longitude e latitude de todos estes pontos. As posições das ONUs são mostradas na Figura 7, indicando cada ONU sobreposta no mapa satelital.

O próximo passo é transformar esses valores de latitude e longitude para coordenadas do plano cartesiano, de forma que as posições sejam compatíveis com a notação matricial do algoritmo desenvolvido. Para realizar esta transformação é necessário saber que o raio da terra (R) é de aproximadamente 6371km, e que as coordenadas cartesianas x e y estão relacionadas com a latitude (Lat) e longitude (Lon) mediante:

$$x = R \cdot \cos(Lat) \cdot \cos(Lon), \quad (2.1)$$

e

$$y = R \cdot \cos(Lat) \cdot \sin(Lon). \quad (2.2)$$

É importante notar que antes de aplicar as funções trigonométricas é necessário transformar graus em radianos. Depois de ter todos estes valores calculados foi possível criar uma matriz com a localização das ONUs reais, para que seja possível projetar-se a rede multi-nível. Também cabe mencionar que para facilitar tanto a representação como os cálculos, as coordenadas x e y foram centralizadas subtraindo o valor médio de cada uma delas. Os valores em graus e em coordenadas cartesianas das 100 ONUs são listados na tabela seguinte.

2.3 MÉTRICAS DE AVALIAÇÃO

Neste trabalho consideraremos duas métricas para avaliar o desempenho do algoritmo. Por um lado, computará-se o comprimento total de fibra instalada e, por outro lado, as perdas médias sofridas por todas as ONUs.

2.3.1 Comprimento total de fibra instalada

O comprimento total de fibra instalada considera a soma dos comprimentos de fibras necessários para comunicar os nós distribuidores entre os diferentes níveis, assim como os nós distribuidores do último nível e as ONUs. É importante destacar que este comprimento não corresponde a soma dos comprimentos de fibra até cada ONU pois parte dos enlaces são compartilhados nos níveis superiores.

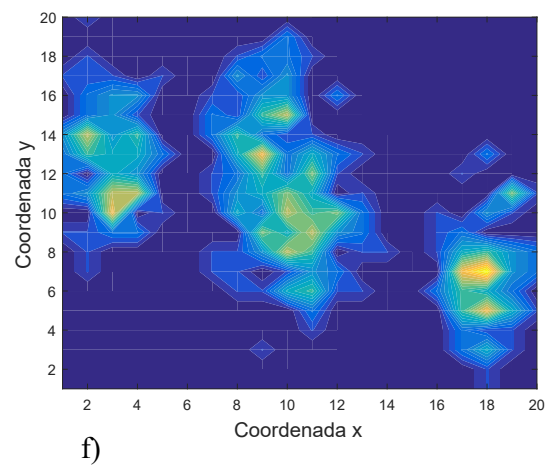
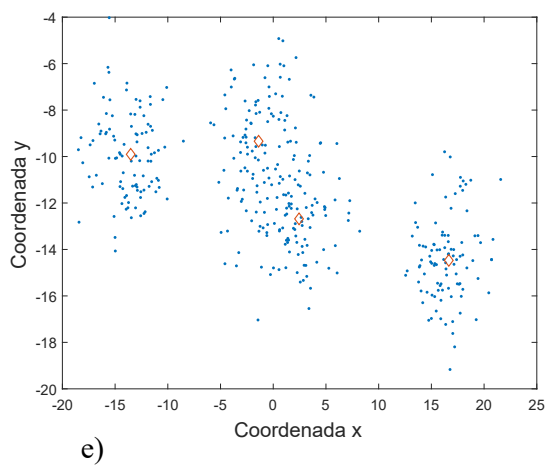
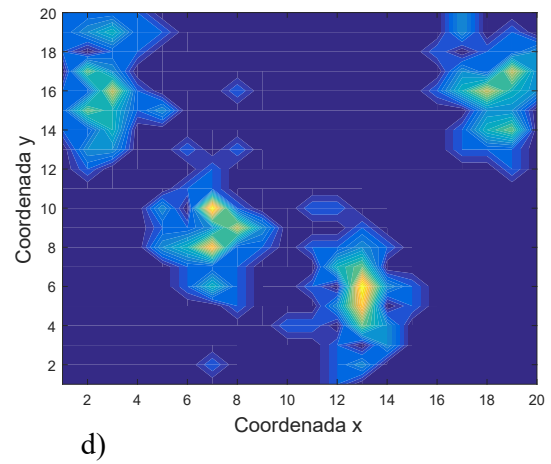
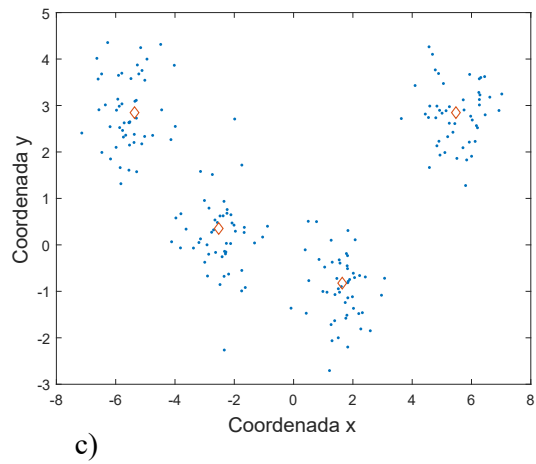
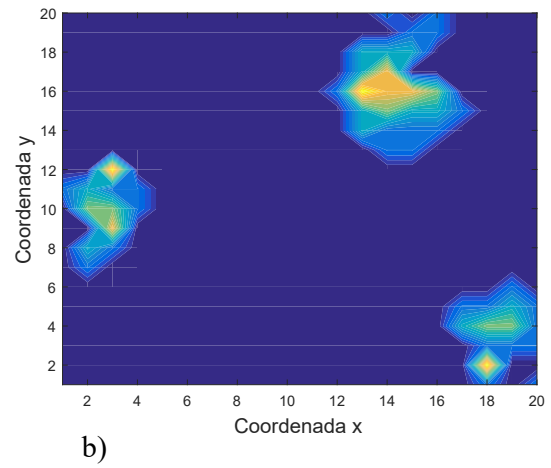
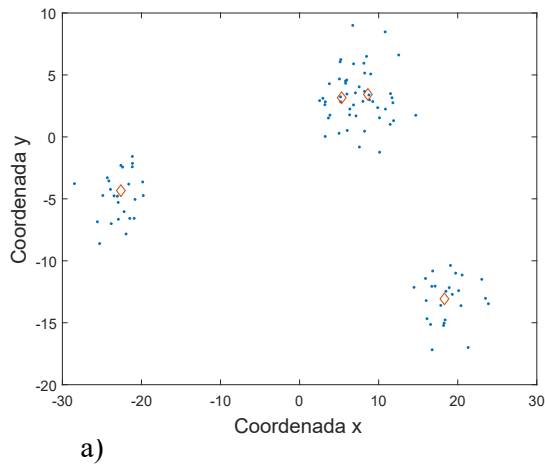


Figura 6 – Posição das ONUs e histogramas bidimensionais da densidade de ONUs para (a-b) 100 usuários, (b-c) 200 usuários, (d-e) 400 usuários.

ONU	Lat. [°]	Lon. [°]	x [km]	y [km]
1	-23,156	-46,271	0,125	0,390
2	-23,155	-46,271	0,146	0,422
3	-23,155	-46,271	0,106	0,454
4	-23,156	-46,271	0,154	0,405
5	-23,155	-46,271	0,109	0,482
6	-23,155	-46,271	0,114	0,511
7	-23,155	-46,272	0,061	0,479
8	-23,156	-46,271	0,078	0,356
9	-23,156	-46,271	0,103	0,350
10	-23,156	-46,271	0,136	0,362
11	-23,156	-46,270	0,201	0,392
12	-23,156	-46,271	0,183	0,361
13	-23,156	-46,270	0,212	0,412
14	-23,156	-46,270	0,197	0,375
15	-23,156	-46,270	0,197	0,342
16	-23,156	-46,270	0,210	0,324
17	-23,156	-46,270	0,288	0,353
18	-23,156	-46,270	0,275	0,331
19	-23,156	-46,270	0,287	0,339
20	-23,156	-46,269	0,322	0,352
21	-23,156	-46,269	0,311	0,392
22	-23,156	-46,269	0,316	0,414
23	-23,156	-46,269	0,372	0,370
24	-23,155	-46,270	0,299	0,439
25	-23,155	-46,270	0,283	0,440
26	-23,155	-46,270	0,245	0,439
27	-23,155	-46,270	0,243	0,459
28	-23,155	-46,270	0,248	0,465
29	-23,155	-46,270	0,244	0,478
30	-23,155	-46,270	0,256	0,491
31	-23,155	-46,270	0,227	0,513
32	-23,155	-46,268	0,438	0,451
33	-23,155	-46,268	0,464	0,468
34	-23,155	-46,269	0,404	0,453
35	-23,156	-46,269	0,385	0,355
36	-23,157	-46,270	0,264	0,261
37	-23,157	-46,270	0,252	0,291
38	-23,157	-46,270	0,256	0,245
39	-23,157	-46,270	0,258	0,231
40	-23,157	-46,270	0,230	0,212
41	-23,158	-46,270	0,201	0,194
42	-23,157	-46,271	0,157	0,210
43	-23,157	-46,271	0,108	0,199
44	-23,157	-46,271	0,127	0,261
45	-23,158	-46,270	0,198	0,147
46	-23,158	-46,270	0,215	0,131
47	-23,158	-46,270	0,247	0,164
48	-23,158	-46,270	0,273	0,162
49	-23,158	-46,270	0,255	0,135
50	-23,160	-46,272	0,044	-0,059

ONU	Lat. [°]	Lon. [°]	x [km]	y [km]
51	-23,160	-46,272	0,01	-0,079
52	-23,160	-46,272	-0,02	-0,053
53	-23,160	-46,273	-0,04	-0,061
54	-23,160	-46,273	-0,06	-0,049
55	-23,160	-46,273	-0,04	-0,087
56	-23,160	-46,272	-0,02	-0,101
57	-23,160	-46,272	0,01	-0,123
58	-23,161	-46,272	0,05	-0,153
59	-23,161	-46,272	0,03	-0,191
60	-23,161	-46,271	0,09	-0,175
61	-23,161	-46,271	0,10	-0,202
62	-23,163	-46,272	0,00	-0,468
63	-23,164	-46,272	-0,01	-0,520
64	-23,163	-46,272	0,07	-0,379
65	-23,163	-46,271	0,09	-0,375
66	-23,165	-46,274	-0,20	-0,671
67	-23,166	-46,275	-0,29	-0,759
68	-23,166	-46,275	-0,26	-0,803
69	-23,166	-46,273	-0,14	-0,770
70	-23,167	-46,274	-0,21	-0,890
71	-23,167	-46,275	-0,28	-0,852
72	-23,167	-46,275	-0,33	-0,852
73	-23,168	-46,274	-0,20	-0,983
74	-23,169	-46,276	-0,39	-1,057
75	-23,169	-46,275	-0,36	-1,11
76	-23,169	-46,275	-0,34	-1,127
77	-23,170	-46,275	-0,34	-1,143
78	-23,170	-46,275	-0,35	-1,155
79	-23,169	-46,276	-0,42	-1,103
80	-23,169	-46,276	-0,44	-1,087
81	-23,169	-46,276	-0,46	-1,102
82	-23,169	-46,276	-0,46	-1,120
83	-23,171	-46,276	-0,37	-1,260
84	-23,169	-46,283	-1,169	-1,103
85	-23,169	-46,282	-1,133	-1,078
86	-23,169	-46,282	-1,136	-1,055
87	-23,169	-46,282	-1,113	-1,043
88	-23,168	-46,283	-1,217	-1,020
89	-23,168	-46,283	-1,155	-1,007
90	-23,169	-46,283	-1,211	-1,037
91	-23,152	-46,266	0,71	0,782
92	-23,153	-46,266	0,74	0,739
93	-23,151	-46,270	0,21	0,938
94	-23,150	-46,271	0,15	0,988
95	-23,150	-46,271	0,14	1,010
96	-23,151	-46,271	0,09	0,975
97	-23,148	-46,271	0,13	1,303
98	-23,147	-46,272	0,03	1,410
99	-23,146	-46,273	-0,06	1,422
100	-23,146	-46,270	0,22	1,431

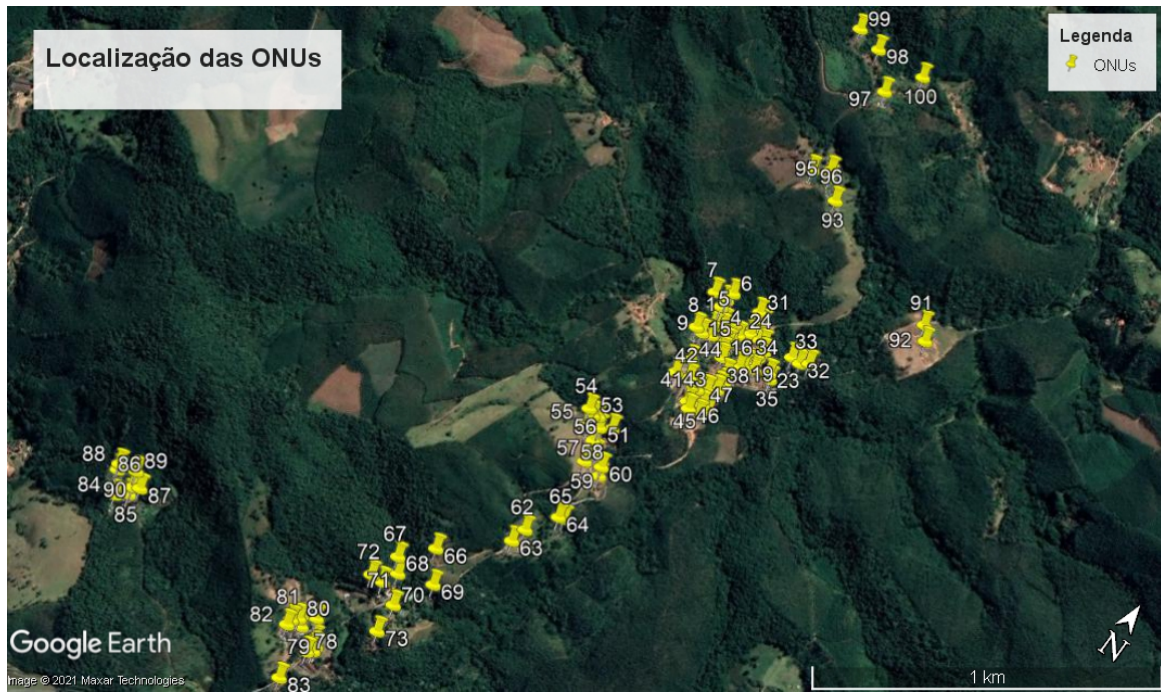


Figura 7 – Posição das ONUs consideradas para a avaliação do algoritmo num cenário realista (Fonte: Google Maps).

2.3.2 Perda média

Para computar as perdas média de uma determinada arquitetura, deve ser computado a perda de cada ONU, a qual, por sua vez, tem duas componentes: as perdas de propagação e as perdas por divisão. Desta forma, em escala logarítmica temos que a perda para uma determinada ONU é:

$$L_{Total}|_{dB} = L_{transmissão}|_{dB} + L_{divisão}|_{dB}, \quad (2.3)$$

sendo $L_{Total}|_{dB}$, $L_{transmissão}|_{dB}$ e $L_{divisão}|_{dB}$ as perdas totais, de transmissão e de divisão. As perdas por transmissão para uma determinada ONU, podem ser calculadas multiplicando o comprimento de fibra até essa ONU e o coeficiente de atenuação da fibra:

$$L_{transmissão}|_{dB} = \alpha|_{dB/km} \cdot L_{fibra}|_{km}, \quad (2.4)$$

em que $\alpha|_{dB/km}$ é o coeficiente de atenuação da fibra, considerado de 0,2 dB/km e $L_{fibra}|_{km}$ é o comprimento de fibra calculado a partir da estrutura achada mediante o algoritmo de clusterização multi-nível.

A perda por divisão nos nós divisores, para uma razão de divisão k , pode ser calculada da seguinte forma:

$$L_{divisão}|_{dB} = 10 \cdot \log_{10}(k). \quad (2.5)$$

De forma aproximada, temos que para $k=2$, $L_{divisão}|_{dB} \approx 3$ dB, para $k=4$, $L_{divisão}|_{dB} \approx 6$ dB, para $k=8$, $L_{divisão}|_{dB} \approx 9$ dB e para $k=16$, $L_{divisão}|_{dB} \approx 12$ dB.

3 RESULTADOS DE SIMULAÇÃO

Nesse capítulo aplicamos o algoritmo considerado para otimizar a topologia em estrela multi-nível de uma rede PON considerando diferentes tipos de cenários. Em particular na Seção 3.1 utilizam-se múltiplos cenários arbitrariamente gerados para obter métricas de ensemble que permitam uma generalização da dependência entre a razão de divisão e as métricas de desempenho, isto é, perdas médias e comprimento total de fibra instalada. Já na Seção 3.2, aplica-se o algoritmo ao cenário realista descrito no capítulo anterior.

3.1 RESULTADOS COM CENÁRIOS GERADOS DE FORMA ALEATÓRIA

3.1.1 Análise de um caso particular

A fim de observar o funcionamento do algoritmo, como exemplo aplicamos este a um cenário arbóreo com 100 ONUs, mostrado na Figura 8. Considerando um valor de $k=4$, o resultado do algoritmo é representado na Figura 9, em que as posições das ONUs são indicadas com marcadores azuis, os nós divisores são indicados em preto e os enlaces de fibra em vermelho. Observa-se na imagem plotada que é respeitada a razão de nó divisor que foi estipulada inicialmente, pois que o ponto central de cor preta (primeiro nó divisor) é dividido em 4 enlaces os enlaces, depois esse enlace é dividido em mais 4 enlaces, formando assim a rede completa para atender todas as ONUs.

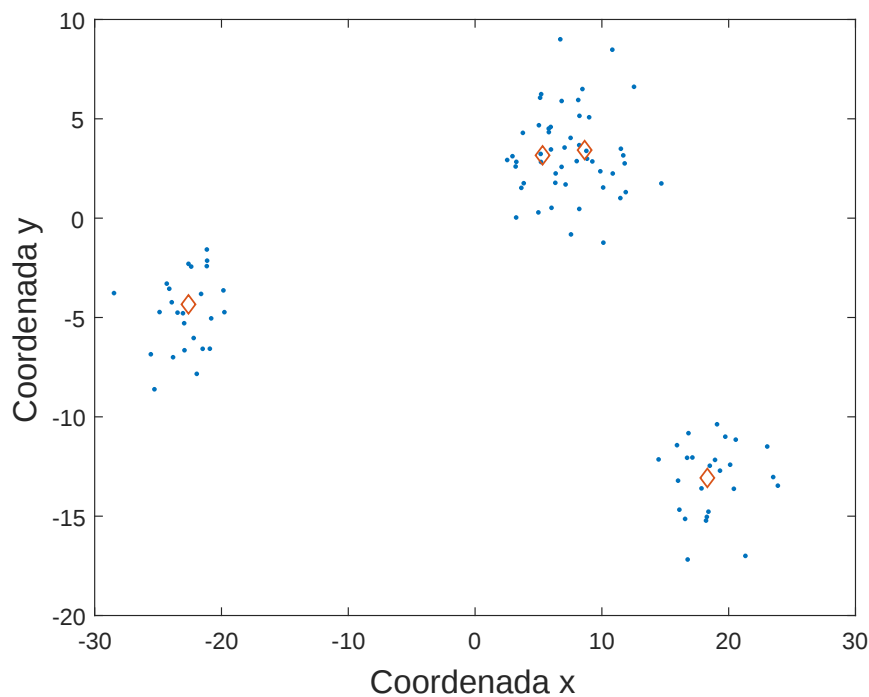


Figura 8 – Posição das ONUs antes de aplicar o algoritmo

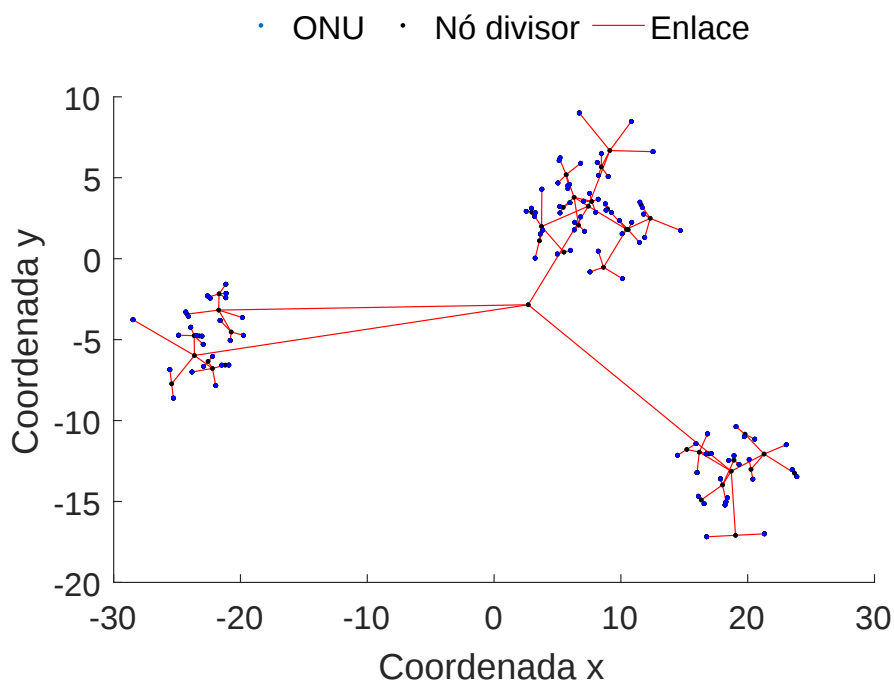


Figura 9 – Rede gerada após aplicação do algoritmo, indicando as ONUs, os nodo divisor e os enlaces que os unem.

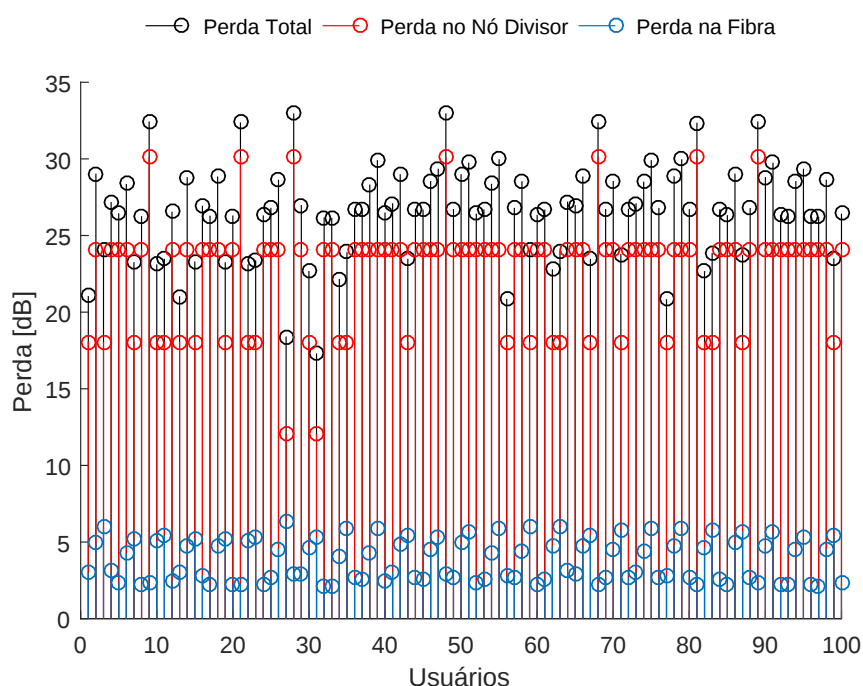


Figura 10 – Perdas da conexão entre o nó central e cada ONU.

Realizou-se também o cálculo do comprimento da fibra para cada ONU, assim como o número de divisores e comprimento total de fibra instalada. Os resultados são apresentados na figura Figura 10, em que o eixo horizontal representa o identificador da ONU enquanto que o eixo vertical representa as diferentes perdas em dB para cada ONU: em azul representa-se a perda de transmissão por fibra; em vermelho tem-se a perda nos nodos divisores, e em preto tem-se a soma dessas duas perdas, por comprimento da fibra e por nodo divisor. Observando as diferentes perdas, podemos comprovar que tanto as perdas por transmissão como as de divisão variam dependendo da ONU. A variação

da primeira é lógica devido aos diferentes comprimentos de fibra desde o nó central até cada ONU. Enquanto às perdas por divisão, pode-se perceber que a variação aparenta estar quantificada. Este comportamento pode ser explicado notando que nem todas as rotas até as ONUs atravessam a mesma quantidade de nós divisores e que cada nó divisor representa uma perda extra de 6 dB (pois $k=4$).

3.1.2 Análise estatístico do conjunto de cenários

Para extrair informação geral do desempenho do algoritmo de clusterização multi-nível e do efeito da razão de divisão, assim como a dependência com relação ao número do ONUs a agrupar, se criaram 100 cenários aleatórios para 100, 200 e 400 ONUs. Cada um desses cenários foi clusterizando utilizando valores de razão de divisão de 2, 4, 8, 16 e 32. Para uma melhor análise dos dados, utilizamos em primeiro lugar gráficos *box plot* tanto para o comprimento total de fibra instalada como para as perdas médias de cada configuração, a qual permite uma análise gráfica da distribuição dos valores. Assim, na Figura 11 mostramos os diferentes cenários, junto com os *outliers* para o comprimento total de fibra instalada e a perda média em termos da razão de extinção para (a-b) 100 ONUs, (c-d) 200 ONUs e (e-f) 400 ONUs.

Comparando os resultados de comprimento instalado e perdas médias, podemos observar que a primeira métrica tem uma dependência clara da razão de divisão, enquanto que as perdas médias não apresentam uma dependência tao evidente. Assim, focando-se no comprimento total de fibra instalada, pode-se concluir que tanto uma razão de divisão de 2 como de 4 apresentam resultados similares em termos de valor médio e dispersão, sendo o valor de 4 ligeiramente melhor. Este comportamento é comum aos cenários com 100, 200 e 400 ONUs.

Mesmo que os diagramas de *box plot* representem uma ferramenta válida para visualização de dados, não conseguem considerar a relação entre diferentes variáveis. Desta forma, para poder observar a relação entre comprimento total de fibra instalada e perdas médias, na Figura 12 representamos a primeira em termos da segunda métrica. Uma das primeiras características que pode-se observar é o aumento da correlação a medida que aumenta a razão de divisão. Também é interessante ver que quando o número de ONUs aumenta as perdas para razões de divisão altas diminui. Mesmo assim, para todos os números de ONUs considerados, a razão de divisão que resulta em um desempenho ótimo em função do comprimento total de fibra instalada e de perdas médias é $k=4$.

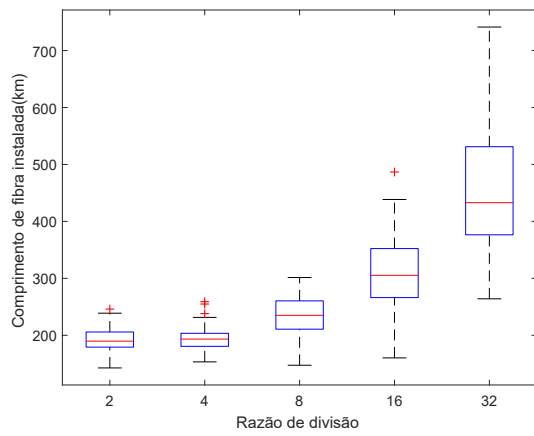
3.2 RESULTADOS COM CENÁRIO REALISTA

3.2.1 Análise de uma solução particular do cenário realista

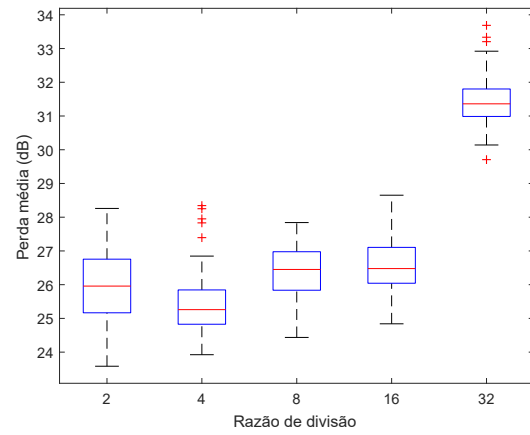
Na Figura 13 mostra-se o mapa das 100 ONUs consideradas no cenário realista, em que cada marcador indica a posição de uma ONU. É importante mencionar que, diferente da análise com cenários aleatórios, neste caso o cenário é único. Na Figura 14 representamos a rede gerada pelo algoritmo de clusterização multi-nível. Assim como no caso do cenário aleatórios, analisamos as perdas de cada ONU, as quais são mostradas na Figura 15. Em comparação com as perdas do cenário aleatoriamente gerado, a contribuição das perdas de transmissão em relação as perdas por divisão é menor devido ao menor diâmetro da rede.

3.2.2 Análise estatística das soluções para o cenário realista

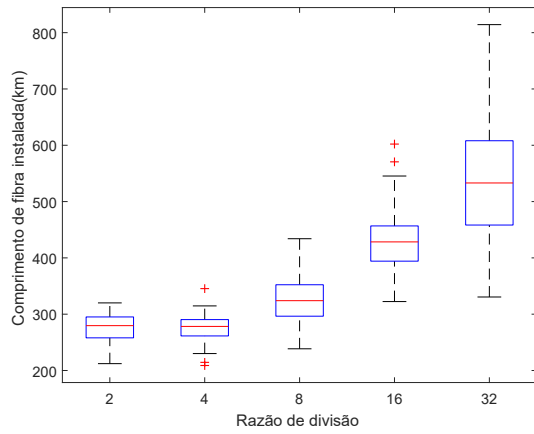
É importante ressaltar que mesmo sendo um cenário único, a natureza estocástica do método faz com que diferentes execuções do algoritmo resultem em diferentes arquiteturas propostas. Para evitar que a inicialização dos centroides afete significativamente os resultados, repetimos a clusterização 100 vezes, obtendo os resultados mostrados na Figura 16. Uma diferença significativa em relação as Figuras obtidas para cenários arbitrários é que neste caso, tanto as perdas médias como o comprimento total de fibra instalada apresentam uma forte dependência da razão de divisão. Porém, assim como no caso



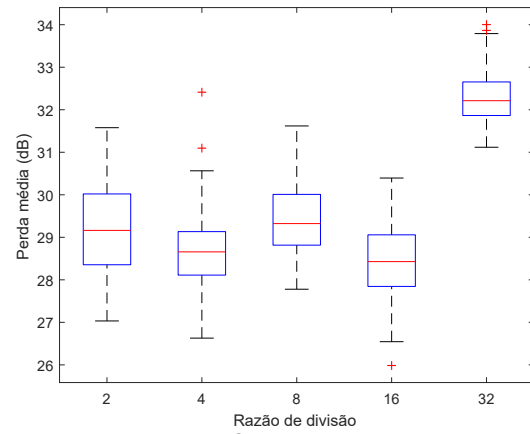
a)



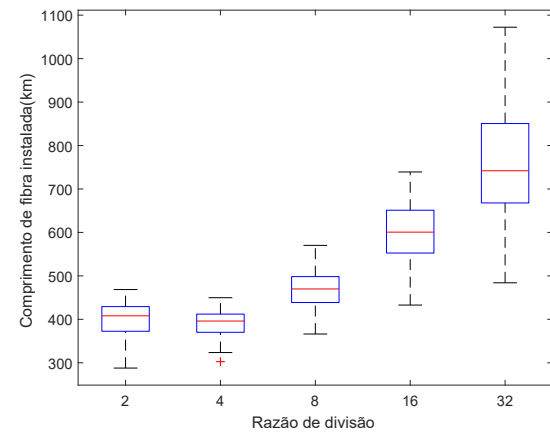
b)



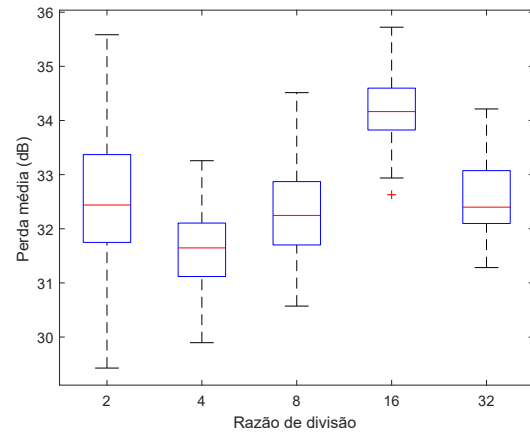
c)



d)

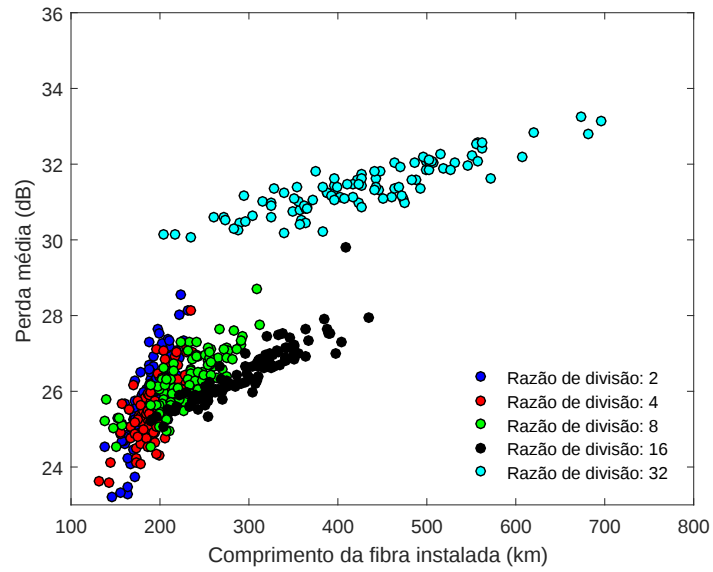


e)

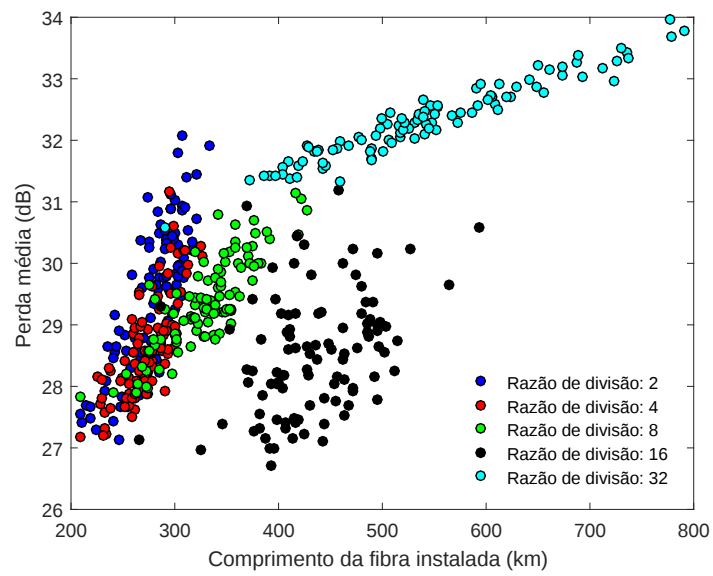


f)

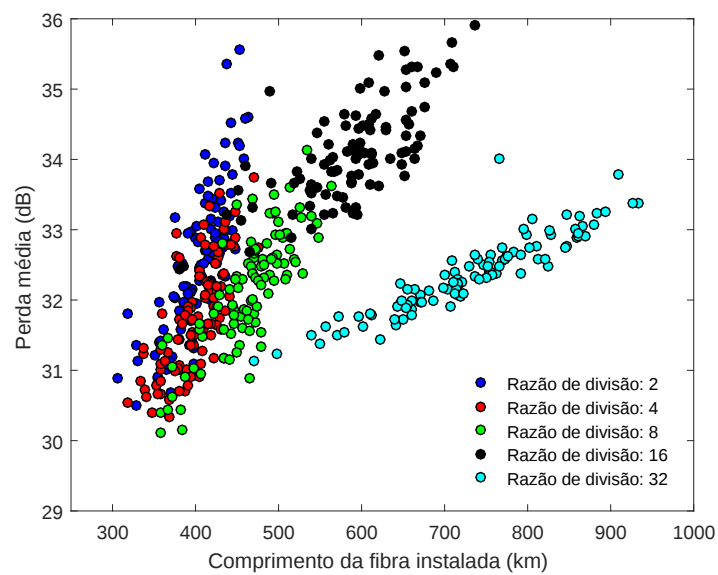
Figura 11 – *Box plots* do comprimento total e das perdas médias para 100 cenários aleatoriamente gerados para (a-b) 100 ONUs, (c-d) 200 ONUs e (e-f) 400 ONUs.



(a)



(b)



(c)

Figura 12 – Perdas médias vs. comprimento total de fibra instalada para 100 cenários arbitrariamente gerados considerando (a) 100, (b) 200, e (c) 400 ONUs.

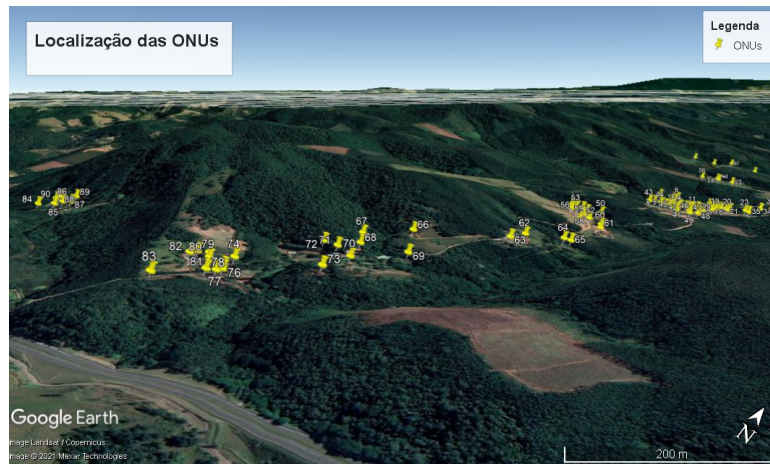


Figura 13 – Posição das ONUs reais antes de aplicar o algoritmo

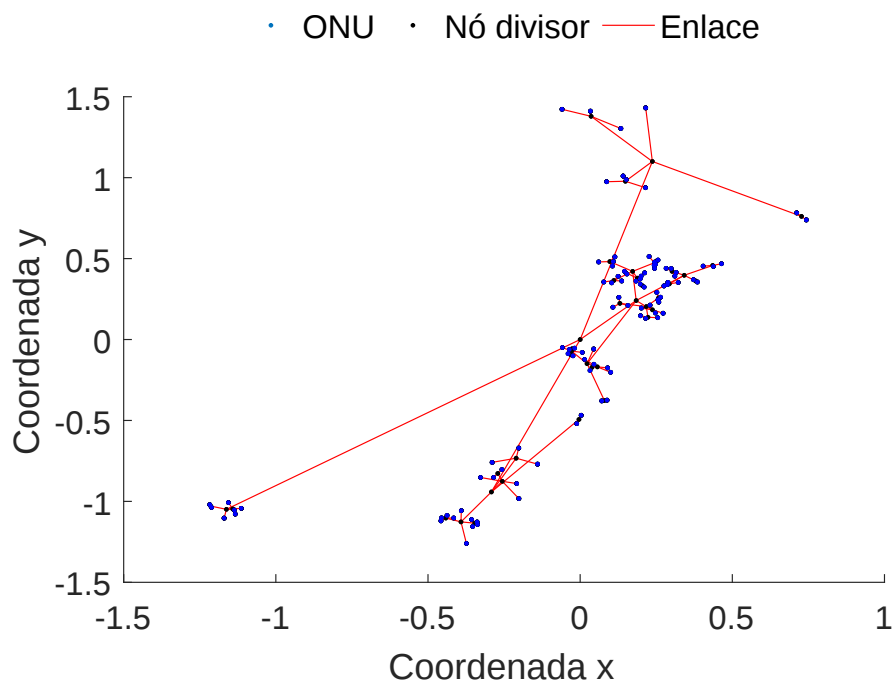


Figura 14 – Rede gerada apos aplicação do algoritmo real

anterior, os melhores resultados correspondem a razões de divisão de 2 e 4. Quando representamos as perdas médias em termos do comprimento total de fibra instalada, vemos que, como esperado, os resultados correspondentes a $k = 2$ e 4 apresentam bons resultados enquanto que maiores razões de divisão resultam em arquiteturas de redes menos eficientes. Neste ponto, é importante notar que a menor dispersão dos pontos pode ser explicada pois todos correspondem ao mesmo cenário.

Pode-se observar o mesmo comportamento para todos os casos apresentados, tanto em um cenário aleatório como em um cenário real, as diferenças apresentadas para um mesmo nó divisor relacionam-se com a quantidade de usuários, fazendo com que aumente ou diminua o comprimento da fibra instalada, pois quando aumenta-se o número de usuários aumenta-se também o comprimento de fibra instalado e as perdas médias.

Observando a Figura 16 e analisando a posição nota-se que o pior caso tanto para um comprimento de fibra instalada quando para a perda média em dB, é o caso para o nó divisor 1:32, os demais seguem mais concentrados no canto inferior esquerdo, nota-se também que a melhor relação está no nó divisor 1:4 que está mais concentrado no canto inferior esquerdo, mostrando assim que possui um baixo

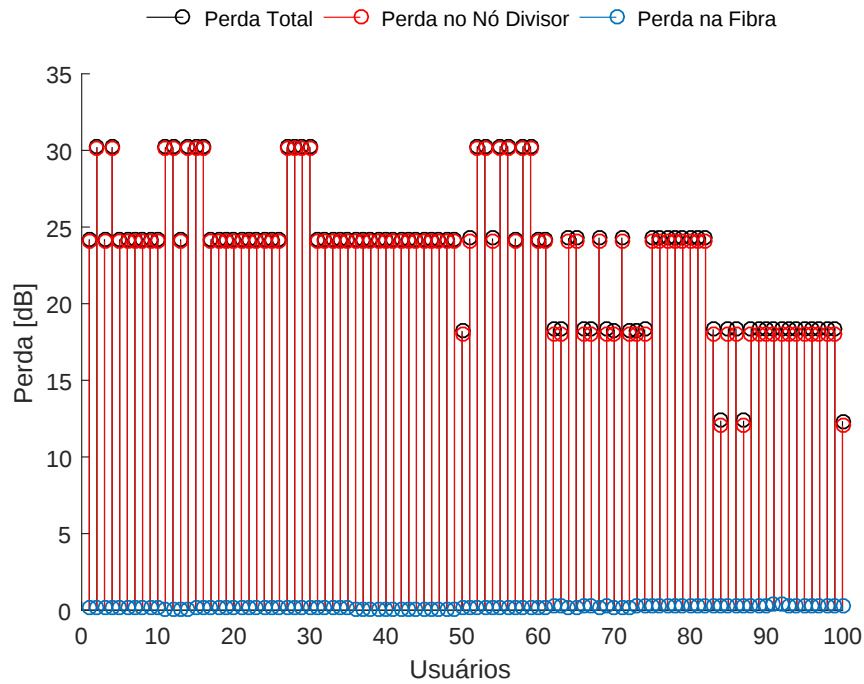


Figura 15 – Perdas de cada ONU para o cenário realista.

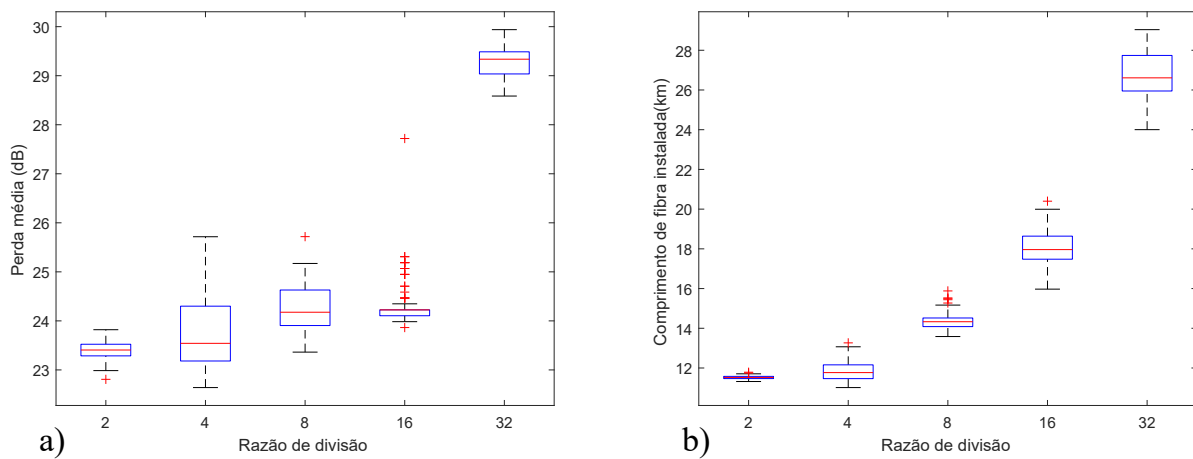


Figura 16 – a)Box plot do comprimento de fibra instalado, b)Box plot da perda média

comprimento de fibra instalado para uma baixa perda média. Observa-se que em todos os casos de 100, 200 e 400 usuários o melhor caso para um menor comprimento de fibra instalada e menor perda média é o do nó divisor 1:4.

Analisando a Figura 17 nota-se que o nó de razão de divisão 1:32 é o mais disperso entre todos os outros nós, sendo assim é o que possui um valor de perda média e de comprimento de fibra instalado, sendo assim também o pior caso como um projeto de rede PON. Os demais seguem mais concentrados no canto inferior esquerdo, como é o caso do nó de razão de divisão 1:4, que está mais concentrado no canto esquerdo, possuindo assim um baixo comprimento de fibra instalado para uma baixa perda média.

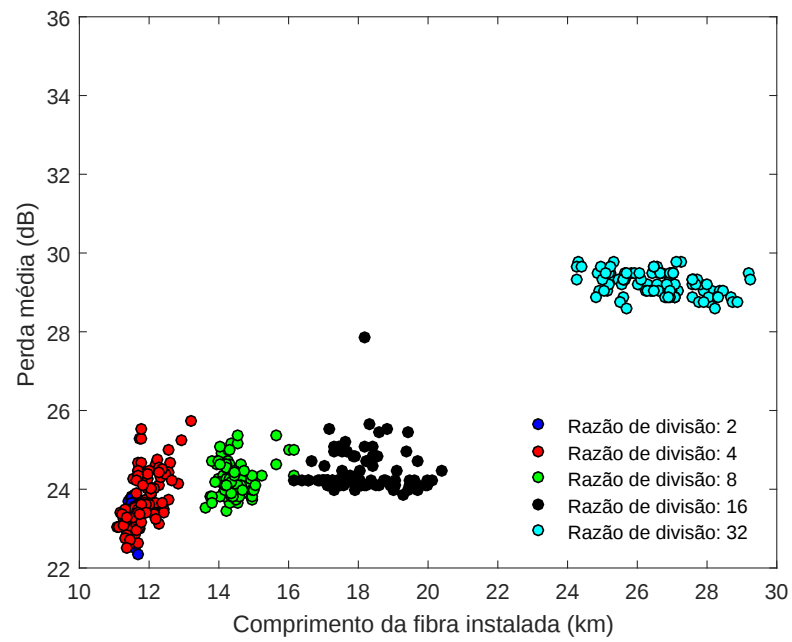


Figura 17 – Perdas médias vs comprimento para um cenário real

4 CONCLUSÕES E TRABALHO FUTURO

Neste trabalho foi desenvolvido um algoritmo baseado em clusterização k-means capaz de projetar uma rede PON utilizando agrupamento multinível, permitindo estudar e analisar o desempenho da rede em termos do comprimento total de fibra instalada e as perdas médias, e sua dependência da razão de divisão. Para avaliar o algoritmo desenvolvido, utilizamos tanto cenários gerados de forma aleatória como um cenário realista.

As simulações realizadas no *software* Matlab com 100 cenários aleatórios e com 100 repetições do cenário realista mostraram que em todos os casos uma razão de divisão de 2 e 4 apresentam um ótimo desempenho enquanto a comprimento total de fibra instalada e perdas médias.

Sendo o foco do estudo encontrar o resultado em que a rede PON projetada tivesse a menor perda utilizando a menor quantidade de fibra necessária. A criação de um cenário aleatório e a implementação e um cenário real foi feita, e a hierarquização da rede foi feita a partir do agrupamento hierárquico. A criação do cenário e estrutura da rede PON foi feita por meio da implementação do algoritmo k-means pelo *software* Matlab.

Introduziu-se com conceitos de aprendizado de máquinas em específico os algoritmos baseados aprendizagem não supervisiona com a técnica de clusterização. Por meio do código criado foi possível obter resultados satisfatórios com o uso do algoritmo k-means mesmo para redes de grade irregular, como é o caso de zonas rurais. Não se encontraram estudos que fazem utilização desse algoritmo para utilização de projetos de redes de grade irregular, como é o caso do código implementado, entretanto tal algoritmo se mostrou inovador na utilização de projetos de redes para grades irregulares, apresentando um desempenho satisfatório e uma baixa complexidade.

Ao utilizar a clusterização k-means, foi possível alcançar uma razão de divisão ótima em termos de comprimento e de perda média, simulou-se inúmeras vezes, alternando-se diversos parâmetros importantes, e em todos os casos encontrou-se o menor comprimento de fibra instalada e menor perda média na razão de divisão quatro (1:4).

O trabalho trata-se de um código implementado pelo *software* Matlab, capaz de projetar a melhor rede de acesso óptico utilizando agrupamento multinível em regiões rurais. Trabalhos futuros deverão levar em consideração o relevo da área onde será projetada a rede PON.

Como já explicado, no presente trabalho utilizaram-se a perda média e o comprimento total de fibra instalada como métricas de avaliação das arquiteturas geradas, desconsiderando assim qualquer métrica associada à confiabilidade. Propomos como trabalho futuro incluir o estudo de confiabilidade da rede e estudar o efeito da razão de divisão nesta métrica.

REFERÊNCIAS

- AGRAWAL, G. **Sistemas de comunicação por fibra óptica**. [S.l.]: Elsevier Brasil, 2014. v. 4.
- ALPAYDIN, E. **Introduction to Machine Learning**. 2nd. ed. [S.l.]: The MIT Press, 2010. ISBN 026201243X, 9780262012430.
- BARBOSA, V. d. A.; METZ, M. M.; FERNANDES, G. T. d. S. Redes óptica passivas (pon). o futuro das redes e suas tendências mercadológicas. **Revista Científica Multidisciplinar Núcleo do Conhecimento**. Ano, v. 3, p. 158–185, 2020.
- LLOYD, S. Least squares quantization in pcm. **IEEE transactions on information theory**, IEEE, v. 28, n. 2, p. 129–137, 1982.
- PEREIRA, F. C. D. L.; JAMHOUR, E. Proposta de algoritmo de posicionamento de divisor de potência ótica para redes gpon. **Brazilian Journal of Development**, v. 7, n. 4, p. 43454–43464, 2021.
- SAMUEL, A. L. Some studies in machine learning using the game of checkers. **IBM Journal on Research and Development**, v. 3, p. 210–229, 1959.
- SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding machine learning : from theory to algorithms**. Cambridge University Press, 2014. ISBN 9781107057135 1107057132. Disponível em: <http://www.worldcat.org/search?qt=worldcat_org_all&q=9781107057135>.
- TAKEUTI, P. Projeto e dimensionamento de redes ópticas passivas (pons). **São Carlos, SP**, 2005.