

Aprendizagem Não-Supervisionada para Avaliação de Métricas Topológicas de Redes Ópticas

Danilo R. B. Araújo*, Joaquim F. Martins-Filho[‡], Carmelo J. A. Bastos-Filho[†]

*Universidade Federal Rural de Pernambuco, Recife 52171-900 Pernambuco, Brasil
Tel: +55 81 21193842, e-mail: danilo@deinfo.ufrpe.br

[‡]Universidade Federal de Pernambuco, Recife 50740-550 Pernambuco, Brasil
Tel: +55 81 21267784, e-mail: jfmf@ufpe.br

[†]Universidade de Pernambuco, Recife 50720-001 Pernambuco, Brasil
Tel: +55 81 31847542, e-mail: carmelofilho@ieee.org

Resumo—Diversos problemas podem ser modelados por meio de grafos, tais como redes sociais, redes de telecomunicações e estruturas biológicas. Com o objetivo de apoiar a análise de redes complexas, diversas métricas topológicas e modelos de geração de grafos têm sido propostos no últimos anos. Este artigo propõe um método sistemático para análise de diferentes redes ópticas de transporte, baseado no algoritmo *k-means* e explora as potencialidades de uma métrica topológica baseada na entropia da transformada de Fourier dos autovalores do Laplaciano, $I(\hat{\mathcal{F}})$. A métrica $I(\hat{\mathcal{F}})$ e outras três métricas disponíveis na literatura são aplicadas para identificação do modelo canônico que foi usado para gerar redes de transporte. De acordo com os resultados obtidos, o agrupamento de redes realizado pelo *k-means* facilita a análise das características topológicas das redes e percebe-se que a métrica $I(\hat{\mathcal{F}})$ possui maior poder explicativo sobre a topologia das redes analisadas.

I. INTRODUÇÃO

Ciência das Redes é uma área de investigação interdisciplinar que estuda redes complexas presentes em diferentes domínios de aplicação, tais como as redes de telecomunicações, redes biológicas, redes neurais, redes sociais, dentre outras. O *National Research Council* define ciência das redes como o estudo das representações de rede de fenômenos físicos, biológicos e sociais, levando a modelos preditivos desses fenômenos [1]. Avanços importantes na Ciência das Redes estão relacionados com a proposição de modelos de geração de redes para criar grafos com propriedades topológicas semelhantes às propriedades encontradas em redes do mundo real. Em 1940, Erdos e Renyi apresentaram estudos importantes relacionados a redes aleatórias [2]. Watts e Strogatz propuseram, em 1988, o primeiro procedimento para geração de redes que apresentam o efeito de mundo pequeno (*small-world* - SW) [3]. Em 1999, Barabási e Albert apresentaram um modelo baseado em conexão preferencial (*preferential attachment*) para gerar redes livres de escala (*scale-free* - SF) [4]. Diversas variantes destes modelos já foram propostas nos últimos anos [5]. Na maior parte dos casos, as redes reais não apresentam uma topologia com características aleatórias, como as redes geradas pelo modelo Erdos-Renyi (ER). Em geral, as redes reais apresentam características mais próximas de redes regulares, redes com efeito de mundo pequeno, redes livre de escala ou uma composição destas características. Portanto, as características topológicas podem ser usadas para classificar as redes em famílias de grafos.

Além disso, estudos realizados para redes de telecomunicação americanas e européias indicam que existe uma forte relação entre aspectos econômicos e demográficos, e a malha de fibra óptica que é implantada para atender as demandas por tráfego [6], [7], [8]. Moss e Townsend [6] estudaram o desenvolvimento da Internet nos Estados Unidos da América no período de 1977 até 1999 e estabeleceram que há uma forte correlação entre a topologia física e lógica das redes de transporte com a presença de empresas que possuem a informação como principal ativo. Recentemente, Tranos e Gillespie [7] estudaram os fatores que determinam a distribuição espacial das redes de transporte na Europa e concluíram que diversas variáveis afetam a topologia física, mas, em geral, os nós apresentam maior conectividade em grandes regiões metropolitanas. Tranos [8] comparou a infraestrutura de redes de aviação com a da Internet na Europa e concluiu que a infraestrutura aérea possui uma distribuição em lei de potência, mas as redes de transporte ópticas são mais estruturadas. Cardenas *et. al* [9] apresentaram um estudo em que mostram que a distribuição de grau dos nós de redes SDH operadas pela Telefônica-Espanha também é caracterizada por lei de potência. Um estudo sobre a emergência de modelos canônicos de redes em topologias de redes da Internet de diversos países foi realizado por Knight *et. al* [10], se mostrando não conclusivo sobre quais modelos canônicos são mais adequados para representar uma rede de transporte, pois algumas redes apresentam distribuição em lei de potência e outras são mais regulares.

Os estudos já realizados em análise topológica de redes de transporte geralmente usam diversas métricas com o objetivo de identificar qual o melhor modelo canônico que explica uma rede existente. Contudo, há uma lacuna sobre um método sistemático que possa agrupar redes de transporte com características semelhantes e analisar as propriedades topológicas no contexto de cada agrupamento considerado. Este trabalho propõe o uso da técnica de aprendizagem não-supervisionada *k-means* [11] para agrupar redes com características semelhantes e explora as potencialidades de uma métrica topológica recentemente proposta, baseada na entropia da transformada de Fourier dos autovalores da matriz Laplaciana [12].

O restante deste artigo está organizado da seguinte forma: a Seção II fornece uma breve revisão da literatura sobre métricas topológicas para caracterização de redes; a Seção III explica em detalhes a métrica $I(\hat{\mathcal{F}})$; a Seção IV detalha a metodologia

proposta para análise dos dados. A Seção V apresenta os resultados obtidos. Finalmente, a Seção VI apresenta as conclusões e sugestões para trabalhos futuros.

II. CARACTERÍSTICAS TOPOLÓGICAS DE REDES

Uma rede pode ser modelada como um grafo $G = (N, L)$, em que N é um conjunto rotulado de nós ou vértices e L é um conjunto de enlaces ou arestas que representam a ligação entre os nós da rede. Para quantificar o número de nós e o número de enlaces da rede usa-se $n = |N|$ e $l = |L|$, respectivamente. O relacionamento entre os nós da rede pode ser estudado por meio de métricas que são calculadas em função dos elementos do conjunto L . Um formato típico de representação de G é por meio de matrizes de adjacências. Uma matriz de adjacências A indica a ligação entre dois nós i e j de G se um elemento $a_{i,j} = 1$, e indica ausência de conectividade se $a_{i,j} = 0$. Se a rede representada por G possui apenas conexões bidirecionais, a matriz A é simétrica.

Um grafo é dito conectado quando existe um caminho entre todos os possíveis pares de nós origem-destino. Quando não há caminho para conectar pelo menos um par de nós da rede, o grafo é dito desconectado. A conectividade por enlace (k_l) é o número mínimo de enlaces que precisam ser removidos de modo a transformar um grafo conectado em um grafo desconectado. A conectividade por nós é definida de forma análoga [13]. A **densidade** de uma rede (q) é definida como a taxa entre o número de enlaces que existem na rede e o número máximo de enlaces que poderiam existir. O **grau de um nó** i (d_i) representa o número de enlaces que ligam um nó i aos nós vizinhos. O número médio de enlaces que estão conectados a um nó é denominado **grau médio** e é definido pela média aritmética de d_i para todo i . A **distribuição de grau dos nós** de G (g') define a probabilidade $Pr(d)$ de que um nó selecionado ao acaso, possua um certo grau d .

O **coeficiente de assortatividade** de uma rede ($-1 \leq r \leq 1$) avalia se há uma tendência de que os nós se conectem com nós que possuem grau semelhante ($r \geq 0$). Se $r \leq 0$, há uma tendência de que os nós com grau elevado se conectem com nós que possuem grau baixo.

O **caminho mais curto** (SP) descreve o número mínimo de saltos necessário para caminhar entre dois pares de nós origem-destino. O maior valor de caminho mais curto na rede, quando são considerados todos os pares origem-destino, é conhecido como **diâmetro da rede** ($diam(G) = SP_{max}$). O **comprimento médio dos caminhos** mínimos (\bar{c}) é a média de todos os SP entre todos os possíveis pares origem-destino.

O coeficiente de agrupamento de um nó i (cc_i) é a taxa entre o número de triangulações que contém o nó i e o número de triangulações que poderiam existir se todos os nós vizinhos de i estivessem conectados. Considerando um nó i , existem d_i vizinhos e potencialmente $d_i(d_i - 1)/2$ enlaces conectando todos os vizinhos. Supondo que estes vizinhos compartilham c enlaces, pode-se calcular cc_i por meio de

$$cc_i = \frac{2c}{d_i(d_i - 1)}. \quad (1)$$

O **coeficiente de agrupamento** da rede (CC) é a média de todos os possíveis c_i da rede. Uma rede com elevado

coeficiente de agrupamento indica que há alguns grupos de nós na rede que são altamente conectados entre si, mas a ligação entre estes grupos é feita por meio de poucos enlaces.

O grau de aleatoriedade de uma rede pode ser quantificado por meio da **entropia** da rede ($I(G)$). Se a distribuição de grau dos nós de uma rede é $g' = \{h_1, h_2, \dots, h_m\}$, $I(G)$ é calculada por meio de

$$I(G) = - \sum_{i=1}^m h_i [\log_2(h_i)]. \quad (2)$$

Outro grupo de métricas usado para análise de redes é derivado dos autovalores de matrizes que representam a rede. Além da matriz de adjacências (A), são usadas na análise espectral a matriz de grau (D) e a matriz Laplaciana da rede (\mathcal{L}). D é uma matriz diagonal $n \times n$ que contém o grau dos nós na diagonal principal da matriz. A matriz \mathcal{L} é definida como $\mathcal{L} = D - A$. O conjunto ordenado de n autovalores de A ou de \mathcal{L} é conhecido como espectro de G . Se existem dois grafos com conjuntos similares de autovalores, é provável que estes dois grafos apresentem uma estrutura semelhante ou um isomorfismo [14]. Araújo *et. al* [12] propuseram uma nova métrica $I(\hat{\mathcal{F}})$, baseada na entropia da transformada de Fourier dos autovalores da matriz Laplaciana. $I(\hat{\mathcal{F}})$ considera simultaneamente a aleatoriedade da distribuição de grau dos nós e quais as ligações estão presentes ou ausentes e fornece uma medida mais elucidativa quando comparada com a entropia tradicional. A Seção III apresenta mais detalhes sobre esta métrica. As métricas apresentadas nesta seção são muito usadas em diversos estudos para indicar propriedades de desempenho esperadas em redes reais e qual modelo canônico melhor representa estas redes.

III. A MÉTRICA TOPOLÓGICA $I(\hat{\mathcal{F}})$

Trabalhos recentes demonstram que a aplicação da Transformada Discreta de Fourier (TDF) sobre os autovalores da matriz Laplaciana podem ajudar na análise de redes e identificar características não percebidas quando o conjunto original de autovalores é usado. Araújo *et. al* [15] propuseram o uso de pontos característicos da curva da TDF para classificar redes de acordo com modelos canônicos. Contudo, as métricas propostas em [15] podem ser aplicadas apenas para classificar redes esparsas. A métrica $I(\hat{\mathcal{F}})$, proposta por Araújo *et. al* [12], é baseada na entropia dos coeficientes da Transformada de Fourier do espectro do Laplaciano e pode ser aplicada tanto para redes esparsas quanto redes densas. \mathcal{L} fornece um resumo da topologia da rede, pois contém informações sobre os graus dos nós e sobre os enlaces presentes na rede. Portanto, uma métrica derivada do espectro de \mathcal{L} pode resumir adequadamente a topologia da rede. Este aspecto já foi explorado para proposição de diversas outras métricas [5], mas muitas destas não são válidas para determinados graus de conectividade, como a densidade de enlaces.

Portanto, é possível usar a entropia dos coeficientes da TDF dos autovalores do Laplaciano para classificar grafos de acordo com a topologia e o nível de aleatoriedade das redes. A métrica é calculada de acordo com o Algoritmo 1. A Equação (3) resume o cálculo numérico da métrica.

Algorithm 1 O algoritmo usado para calcular $I(\hat{\mathcal{F}})$.

- 1: **procedimento** CALCULARIF
 - 2: Seja \mathcal{A} a matriz de adjacências do grafo G ;
 - 3: Calcule a matriz de grau \mathcal{D} ;
 - 4: Calcule a matriz Laplaciana $\mathcal{L} = \mathcal{D} - \mathcal{A}$;
 - 5: Calcule os autovalores reais de \mathcal{L} e armazene em E ;
 - 6: Calcule a Transformada Discreta de Fourier (*TDF*) de E e armazene os valores em \mathcal{F} ;
 - 7: Normalize o conjunto \mathcal{F} de modo a obter valores no intervalo $[0,1]$ e armazene estes valores em $\hat{\mathcal{F}}$;
 - 8: Calcule a entropia de $\hat{\mathcal{F}}$ de acordo com a Equação (3);
 - 9: **fim procedimento**
-

$$I(\hat{\mathcal{F}}) = - \sum_{i=1}^{|\hat{\mathcal{F}}|} (\hat{\mathcal{F}}_i \cdot \log_2 \hat{\mathcal{F}}_i). \quad (3)$$

A Equação (3) possui uma estrutura semelhante à encontrada para o cálculo tradicional de entropia em teoria da informação e para a entropia dos graus dos nós, já conhecida em Ciência das Redes e apresentada na Equação (2). A métrica para cálculo de entropia avalia a quantidade de informação, grau de incerteza ou previsibilidade sobre o valor numérico de uma determinada grandeza física. Contudo, a adaptação do conceito de entropia para o valor normalizado do conjunto \mathcal{F} remete simultaneamente à medida de incerteza sobre o grau dos nós e sobre a assortatividade de uma rede, visto que incorpora de forma simultânea: (i) o mecanismo de ligação entre os nós por meio da matriz Laplaciana (subsídio para avaliação de assortatividade); e (ii) a avaliação de aleatoriedade sobre o grau dos nós por considerar que a entropia também atua sobre o grau dos nós que está presente na diagonal principal da matriz Laplaciana. Além disso, a mesma motivação apresentada aqui poderia ser aplicada diretamente ao conjunto dos autovalores, sem uso de uma transformação adicional (*TDF*). Contudo, a *TDF* proporciona uma análise coerente da métrica sobre uma faixa larga de densidades e a aplicação da entropia diretamente sobre o conjunto normalizado dos autovalores não possui uma mesma interpretação quando redes com densidades diferentes são consideradas. A Fig. 1 ilustra o uso de três diferentes medidas de entropia *versus* densidade de enlaces, para redes com 100 nós geradas por diferentes modelos canônicos. A Fig. 1 reforça estes argumentos apresentados para justificativa do uso de \mathcal{F} . Por exemplo, para a entropia do grau dos nós, ilustrada na Fig. 1b, há diferentes regras de classificação de redes para $q < 0,10$; $0,10 < q < 0,96$; e $q > 0,96$. Além disso, a entropia dos autovalores de \mathcal{L} , ilustrada na Fig. 1c, também não possibilita uma regra de classificação única para todas as faixas de densidade de enlaces.

A métrica $I(\hat{\mathcal{F}})$ foi aplicada em [16] e em [17] para auxiliar na estimativa de probabilidade de bloqueio de redes ópticas. Outro estudo fez uso das métricas $I(\hat{\mathcal{F}})$, CC , r e \bar{c} para análise de cinco topologias de redes brasileiras [18]. Contudo, todos estes estudos aplicaram a métrica $I(\hat{\mathcal{F}})$ na análise de um conjunto muito pequeno de cenários. Este estudo propõe uma metodologia sistemática para avaliação da capacidade explicativa de métricas topológicas e aplica a metodologia proposta sobre um conjunto de mais de 100 redes ópticas

implantadas em diversos países.

IV. METODOLOGIA PROPOSTA

O problema estudado neste artigo pode ser descrito da seguinte forma: dado um conjunto de topologias de redes (grafos) com diferentes características, o objetivo é agrupar as topologias fornecidas de acordo com o modelo canônico que mais aproxima cada rede analisada. No estudo são considerados dois modelos canônicos, o modelo de Barabási-Albert (BA) e o modelo de Watts-Strogatz (WS). As redes geradas pelo modelo BA representam redes *scale-free* típicas e as redes geradas pelo modelo WS possuem o efeito *small-world* como característica predominante [5]. São criadas diversas redes usando os dois modelos e em seguida estas redes são agrupadas usando o algoritmo *k-means* [11]. O algoritmo *k-means* permite o agrupamento de padrões em k grupos de acordo com características previamente escolhidas. Então, são usadas diversas métricas topológicas de redes com o objetivo de avaliar se há alguma métrica que possui maior capacidade explicativa sobre o modelo canônico que originou a rede analisada. Na implementação adotada do *k-means*, os centróides são inicializados com os k primeiros padrões da base de dados e o algoritmo é finalizado quando os centróides não são modificados em duas iterações sucessivas.

Nos experimentos considerados, as redes criadas pelos modelos canônicos possuem diferentes graus médios (\bar{d}). Considerando a necessidade de agrupamento de redes em esparsas ou densas, e geradas a partir dos modelos BA ou WA, o estudo adota por convenção $k = 4$ grupos. As métricas consideradas no processo de aprendizagem não-supervisionada são: grau médio \bar{d} , entropia do grau dos nós $I(G)$, coeficiente de assortatividade r , grau do nó mais conectado $d(h)$ e entropia da *TDF* dos autovalores do Laplaciano $I(\hat{\mathcal{F}})$. Para identificação da capacidade de um conjunto de métricas de agrupar corretamente as redes de acordo com o modelo canônico, são adotadas algumas definições e procedimentos. Define-se como **modelo canônico preferencial** o modelo relacionado ao procedimento que gerou a rede mais próxima do centróide de cada grupo identificado pelo algoritmo *k-means*. Define-se como **taxa de sucesso de grupo** o percentual de redes de um grupo que foram geradas pelo modelo canônico preferencial do grupo considerado. Define-se como **capacidade explicativa** das métricas, a média das taxas de sucesso de grupo considerando todos os k grupos. O objetivo deste estudo é, portanto, identificar o conjunto de métricas topológicas que possui maior capacidade explicativa para identificação do procedimento gerador de uma topologia de rede.

O estudo é dividido em duas etapas. Na primeira etapa são geradas diversas topologias de redes usando os modelos canônicos. É considerada a geração de topologias para uma rede de 32 nós, que é o número médio de nós das redes encontradas na base de dados *Internet Topology Zoo* [10]. No arranjo experimental é considerada a geração de 500 redes de cada modelo com variação no grau médio de 2,0 até 5,0. O procedimento de agrupamento considera, portanto, um total de 1000 redes. O modelo WS é usado com uma probabilidade de religação $rp = 0,10$ e a rede regular usada no início do algoritmo é escolhida de acordo com a densidade especificada no experimento. O modelo BA também considera um número adicional de arestas Δm proporcional à

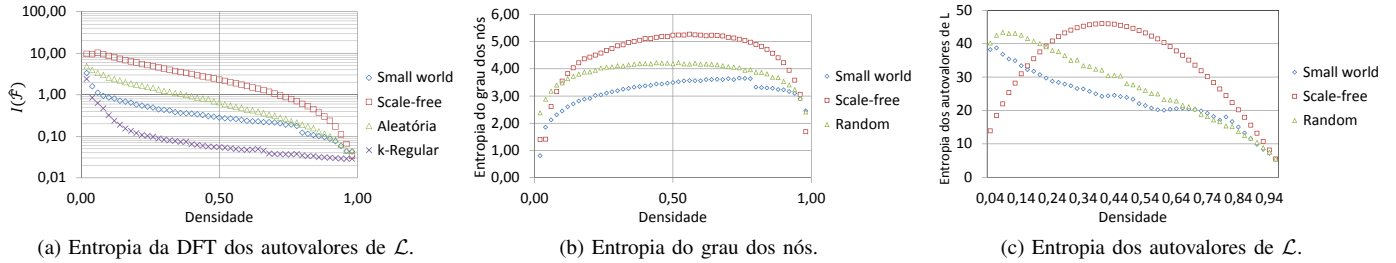


Figura 1: Diferentes medidas de entropia para redes *scale-free*, redes aleatórias e redes *small-world*, de 100 nós.

densidade indicada na geração de cada rede. A segunda etapa do estudo consiste na aplicação do conjunto das métricas com maior capacidade explicativa (de acordo com a etapa 1) para agrupar 107 redes ópticas em operação em diversos países e disponíveis publicamente na base de dados *Internet Topology Zoo* [10]. Foram utilizadas apenas as redes desta base de dados que correspondem a grafos conectados. O objetivo da primeira etapa é realizar uma análise sobre redes nas quais o modelo canônico é conhecido e permitir extrair a capacidade explicativa das métricas. O objetivo da segunda etapa do estudo é conhecer as propriedades topológicas de uma base de dados ampla de redes ópticas e avaliar se, a metodologia baseada em *k-means* e nas métricas definidas na primeira etapa, permite a separação das redes segundo o modelo canônico.

V. RESULTADOS

A Tabela I apresenta um resumo do processo de agrupamento usando o *k-means* para topologias de redes geradas pelos modelos canônicos BA e WS para 1000 diferentes topologias de uma rede com 32 nós. Todas as linhas da Tabela I usam a métrica de grau médio da rede (\bar{d}) para permitir a separação entre redes esparsas e densas. As demais métricas são usadas para separação das topologias de acordo com o modelo canônico. Com base nos resultados obtidos, é possível perceber que o agrupamento por \bar{d} e $I(\hat{\mathcal{F}})$ é o que possui a maior capacidade explicativa quando comparada com as demais opções de duas variáveis. A separação mais precisa ocorre quando são usadas 4 métricas (\bar{d} , $I(\hat{\mathcal{F}})$, $I(G)$, $d(h)$) e este melhor conjunto de 4 métricas oferece uma capacidade explicativa de 0,9993. A inclusão de uma quinta métrica não melhorou a separação dos grupos por modelo canônico. A melhoria da capacidade explicativa quando as diversas linhas da Tabela I são comparadas indica que as métricas mais expressivas para representar os modelos canônicos analisados são, nesta ordem: $I(\hat{\mathcal{F}})$, $I(G)$, $d(h)$ e r . Considerando que o grau médio das redes geradas variam de 2 até 5, o ponto médio de grupos relacionados com redes esparsas e densas deveriam fornecer centróides com graus médios de 2,75 e 4,25, respectivamente. Portanto, é possível perceber na maioria das linhas da Tabela I quatro grupos bem definidos relacionados com redes BA esparsas, redes BA densas, redes WS esparsas e redes WS densas. Por exemplo, a primeira linha da Tabela I separou redes WS esparsas no grupo 1, redes BA esparsas no grupo 2, redes BA densas no grupo 3 e redes WS densas no grupo 4. Com base nos resultados da Tabela I, pode-se concluir que a melhor combinação de variáveis para separação de redes que apresentam diferentes densidades, e com base nos modelos

canônicos BA ou WS é: \bar{d} e $I(\hat{\mathcal{F}})$, quando 2 métricas forem usadas; \bar{d} , $I(\hat{\mathcal{F}})$ e $I(G)$, quando 3 métricas forem usadas; e \bar{d} , $I(\hat{\mathcal{F}})$, $I(G)$ e $d(h)$, quando 4 métricas forem usadas. Além disso, a melhoria em termos de capacidade explicativa quando a separação usa 4 métricas ao invés de duas é de apenas 2%. Portanto, o conjunto \bar{d} e $I(\hat{\mathcal{F}})$ pode ser uma boa opção para indicação do modelo canônico de uma rede quando for desejável o uso de um mecanismo simples de comparação.

A Tabela II apresenta um resumo do procedimento de agrupamento sobre uma base de 107 redes de fibra óptica [10], usando $k = 8$ e as variáveis n , q e $I(\hat{\mathcal{F}})$. Nesta etapa n também foi usado, pois as redes da base de dados apresentam diferentes números de nós. O perfil das redes foi construído comparando as métricas com o valor médio esperado para a base de dados, ou seja, é considerado que uma rede possui poucos nós se a quantidade obtida foi abaixo da média para a base inteira (que é de 32 nós). Considerando que $I(\hat{\mathcal{F}})$ varia com o número de nós e a densidade da rede, para determinar se o valor obtido é baixo ou alto, o valor numérico foi comparado com o obtido para uma rede ER de mesmo tamanho e densidade, conforme sugerido em [12]. Com base nos resultados apresentados na Tabela II é possível perceber que existem redes de fibra óptica implantadas atualmente que possuem características tanto de redes BA quanto de redes WS. Contudo, a quantidade de redes separadas em cada grupo da Tabela II não foi uniforme. Por exemplo, o grupo 1 possui 27 redes e o grupo 7 possui apenas 2 redes. Os grupos que possuem valor elevado de $I(\hat{\mathcal{F}})$ totalizam 29 redes. Ou seja, com base nos resultados obtidos, é possível dizer que 27% das redes da base de dados analisada possuem característica *scale-free* predominante e presença de concentradores. Por outro lado, 73% das redes da base de dados são mais regulares e sem presença de concentradores. A Fig. 2 apresenta um mapa geolocalizado das redes correspondentes aos centróides da Tabela II. É possível identificar a presença de concentradores nos mapas das Figs. 2c, 2d, e 2g. Nas demais topologias, a distribuição de grau dos nós é mais uniforme. A rede americana ATT (Fig. 2c) possui um concentrador com grau 9 localizado em Dallas e outros dois concentradores importantes em Chicago e São Francisco, ambos com grau 8. A rede brasileira RNP (Fig. 2d) possui um concentrador com grau 5 em Brasília, que possui uma quantidade de enlaces muito superior à média da rede, que é de aproximadamente 2 enlaces por nó. A rede grega FORTHNET (Fig. 2g) possui diversos concentradores, com destaque para o nó localizado em Atenas, do qual partem 17 enlaces. Uma análise interessante resulta da comparação entre as redes brasileiras RNP (Fig. 2d) e a rede da

Tabela I: Resultado do processo de agrupamento por k -means para diferentes conjuntos de métricas topológicas.

Métricas consideradas	Capac. explic.	Grupo 1			Grupo 2			Grupo 3			Grupo 4		
		BA	WS	Centróide	BA	WS	Centróide	BA	WS	Centróide	BA	WS	Centróide
\bar{d}, r	0,9390	3	171	2,00; 0,20	248	14	2,56;-0,30	288	59	4,13;-0,23	1	296	3,88;0,13
$\bar{d}, d(h)$	0,9396	35	185	2,00;4,00	252	0	2,75;12,00	221	0	4,31;16,00	32	355	3,88;6,00
$\bar{d}, I(G)$	0,9654	222	0	2,56;2,20	25	186	2,00;0,88	293	6	4,13;2,77	0	348	3,88;1,75
$\bar{d}, I(\mathcal{F})$	0,9776	7	185	2,00;1,76	259	0	2,56;2,80	254	0	4,13;2,30	20	355	3,88;1,28
$\bar{d}, I(\mathcal{F}), r$	0,9887	2	182	2,00;0,18;1,78	268	3	2,75; -0,33;2,76	263	1	4,13;-0,29;2,28	7	354	3,94;0,08;1,25
$\bar{d}, I(\mathcal{F}), d(h)$	0,9925	3	185	2,00;3,00; 1,76	256	0	2,56;11,00;2,66	276	0	4,13;16,00;2,27	5	355	3,88;6,00;1,27
$\bar{d}, I(\mathcal{F}), I(G)$	0,9939	0	185	2,00;0,88;1,76	237	0	2,56;2,16;2,80	299	4	4,13;2,77;2,33	4	351	3,94;1,77;1,24
$\bar{d}, I(\mathcal{F}), I(G), r$	0,9967	0	184	2,00;0,84; 0,14;1,76	233	1	2,44;2,04; -0,32;2,87	305	1	4,00;2,79; -0,30;2,29	2	354	4,00;1,79; 0,10;1,28
$\bar{d}, I(\mathcal{F}), I(G), d(h)$	0,9993	0	185	2,00;0,88; 3,00;1,76	240	0	2,44;2,17; 11,00;2,84	299	0	4,00;2,81; 15,00;2,26	1	355	3,94;1,77; 6,00;1,24
$\bar{d}, I(\mathcal{F}), I(G), d(hub), r$	0,9993	0	185	2,00;0,88;0,13; 3,00;1,77	240	0	2,44;2,17;-0,29; 11,00;2,84	299	0	4,13;2,83;-0,25; 15,00;2,46	1	355	4,00;1,79;0,10; 6,00;1,28

EMBRATEL (Fig. 2h). Embora ambas possuam abrangência nacional, a da EMBRATEL possui mais nós e maior densidade de enlaces. Contudo, embora a métrica $I(\mathcal{F})$ possua valor superior para a rede da EMBRATEL em comparação com a rede RNP, é possível perceber que isto se deve ao número de nós adicional que a EMBRATEL possui. Por isso, é importante usar redes ER como *benchmarking* quando redes com diferentes tamanhos são comparadas. Uma rede ER equivalente à RNP e à EMBRATEL possui $I(\mathcal{F})$ igual a 1,70 e 4,02, respectivamente. Como a RNP apresenta $I(\mathcal{F})$ acima do valor de referência, e a EMBRATEL apresenta $I(\mathcal{F})$ abaixo do valor de referência, conclui-se que estas redes são melhor aproximadas pelos modelos BA e WS, respectivamente. Na Fig. 2h é possível perceber maior regularidade nas conexões da rede EMBRATEL, pois há anéis interligados nas regiões NE e S, além da região SE.

Tabela II: Resultado do algoritmo k -means usando n, \bar{d} e $I(\mathcal{F})$, para uma base contendo 107 redes de fibra óptica [10].

ID	Perfil das redes do grupo	Centróide			
		Rede	n	\bar{d}	$I(\mathcal{F})$
1	Poucos nós, esparsas, entr. baixa	BELNET	19	2,10	1,31
2	Poucos nós, densas, entr. baixa	HIBERNIA	20	2,70	1,36
3	Poucos nós, densas, entr. alta	ATT	25	4,48	1,65
4	Poucos nós, esparsa, entr. alta	RNP	28	2,21	1,77
5	Muitos nós, densas, entr. baixa	BICS	33	2,91	2,18
6	Muitos nós, esparsa, entr. baixa	IRIS	51	2,51	2,80
7	Muitos nós, esparsa, entr. alta	FORTHNET	60	1,97	5,80
8	Muitos nós, densa, entr. baixa	EMBRATEL	76	2,63	3,46

É interessante enfatizar que as redes destacadas na Fig. 2 são estruturalmente distintas, pois possuem diferenças em termos de quantidade de nós, densidades de enlaces ou na forma como os nós estão conectados. Portanto, este conjunto de redes pode ser aplicado em estudos nos quais deseja-se avaliar o impacto de novos algoritmos sobre topologias de redes diferentes, assim como na proposição de novos algoritmos de roteamento.

VI. CONCLUSÕES

Neste artigo foi analisada a capacidade explicativa de diversas métricas topológicas com relação ao modelo canônico que mais aproxima uma dada rede. Para esta avaliação foi usado o algoritmo de aprendizagem não-supervisionada k -means. De acordo com os resultados obtidos para diversas

redes de fibra óptica, o procedimento de agrupamento proporcionado pelo k -means facilitou a análise das características topológicas das redes e a métrica $I(\mathcal{F})$ é a métrica com mais maior capacidade explicativa individual quando comparada com as demais. Novos estudos podem ser realizados para avaliação de outras características de redes de transporte, como a abrangência geográfica e a capacidade de distribuição de carga de tráfego. Além disso, outros modelos de geração de grafos com características topológicas diferentes podem ser aplicados para avaliação da metodologia proposta neste artigo.

AGRADECIMENTOS

Os autores agradecem o suporte financeiro do CNPq, CAPES, UFPE, UFRPE e UPE.

REFERÊNCIAS

- [1] National Research Council Committee on Network Science for Future Army Applications. *Network Science*. The National Academies Press, 2005.
- [2] P. Erdos and A. Rényi. On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, pages 17–61, 1960.
- [3] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440442, 1998.
- [4] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
- [5] T. G. Lewis. *Network Science - Theory and Applications*. John Wiley & Sons, 2009.
- [6] M. L. Moss and A. M. Townsend. The Internet backbone and the American metropolis. *The Information Society Journal*, 16(1):35–47, 2000.
- [7] E. Tranos and A. Gillespie. The spatial distribution of Internet backbone networks in Europe: A metropolitan knowledge economy perspective. *European Urban and Regional Studies*, 16(4):423–437, 2009.
- [8] E. Tranos. The topology and the emerging urban geographies of the Internet backbone and aviation networks in Europe: a comparative study. *Environment and Planning*, 43(2):378–392, 2011.
- [9] J. P. Cardenas, R. M. Benito, M. L. Mouronte, and V. Feliu. The effect of the complex topology on the robustness of Spanish SDH network. In *Networking and Services, 2009. ICNS '09. Fifth International Conference on*, pages 86–90, 2009.
- [10] S. Knight, H. X. Nguyen, N. Falkner, R. Bowden, and M. Roughan. The Internet topology zoo. *Selected Areas in Communications, IEEE Journal on*, 29(9):1765–1775, 2011.
- [11] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297. University of California Press, 1967.

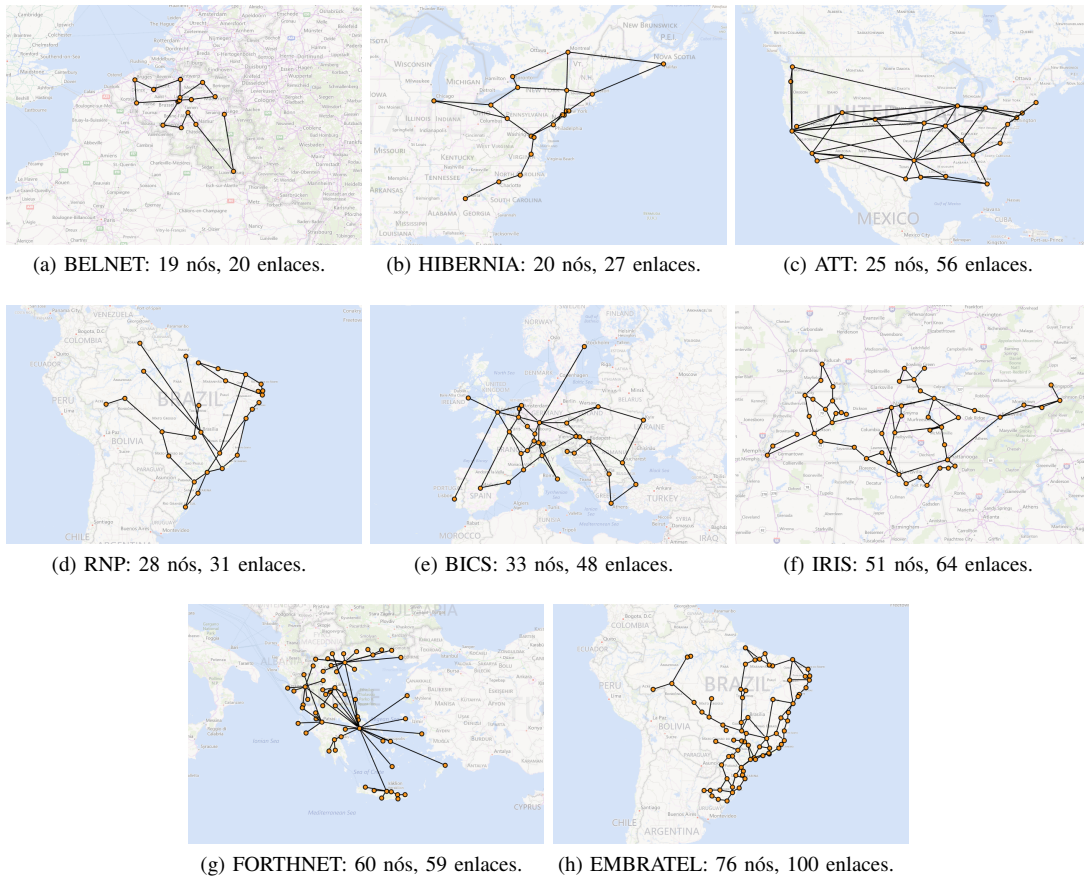


Figura 2: Redes centrais obtidas pelo k -means sobre um conjunto de 107 redes ópticas usando as métricas n , \bar{d} e $I(\hat{\mathcal{F}})$.

- [12] D. R. B. Araújo, C. J. A. Bastos-Filho, and J. F. Martins-Filho. *Studies in Computational Intelligence. Complex Networks V*, volume 549, chapter Using the Entropy of DFT of the Laplacian Eigenvalues to Assess Networks, pages 209–216. Springer, 2014.
- [13] B. Bollobás. *Modern graph theory*. Springer-Verlag, 1998.
- [14] E. R. van Dam and W. H. Haemers. Which graphs are determined by their spectrum? *Linear Algebra and its Applications*, 373:241–272, 2003.
- [15] D. R. B. Araújo, C. J. A. Bastos-Filho, and J. F. Martins-Filho. Towards Using DFT to Characterize Complex Networks. In *XXXI Simpósio Brasileiro de Telecomunicações (SBrT2013)*, pages 1–5, 2013.
- [16] D. R. B. Araújo, C. J. A. Bastos-Filho, and J. F. Martins-Filho. Methodology to Obtain a Fast and Accurate Estimator for Blocking Probability of Optical Networks. *Journal of Optical Communications and Networking*, 7(5):380–391, May 2015.
- [17] D. R. B. Araújo, C. J. A. Bastos-Filho, and J. F. Martins-Filho. Artificial neural networks to estimate blocking probability of transparent optical networks: A robustness study for different networks. In *Proceedings of the 17th International Conference on Transparent Optical Networks ICTON 2015*, July 2015.
- [18] D. R. B. Araújo, C. J. A. Bastos-Filho, and J. F. Martins-Filho. Métricas de redes complexas para análise de redes Ópticas. *Revista de Tecnologia da Informacao e Comunicacao*, 4(2):1–8, 2014.