



UNIVERSIDADE FEDERAL DO MARANHÃO
Programa de Pós-Graduação em Ciência da Computação

Stefano Walker Pereira Pontes

Modelagem Baseada em Grafo para Detecção de Regiões de Anomalias em Termografia de Mama

São Luís - MA
2019

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

STEFANO WALKER PEREIRA PONTES

**MODELAGEM BASEADA EM GRAFO PARA DETECÇÃO DE
REGIÕES DE ANOMALIAS EM TERMOGRAFIA DE MAMA**

SÃO LUÍS – MA

2019

STEFANO WALKER PEREIRA PONTES

**MODELAGEM BASEADA EM GRAFO PARA DETECÇÃO DE
REGIÕES DE ANOMALIAS EM TERMOGRAFIA DE MAMA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFMA como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Tiago Bonini Borchardt

SÃO LUÍS – MA

2019

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Núcleo Integrado de Bibliotecas/UFMA

Pontes, Stefano Walker Pereira.

Modelagem Baseada em Grafo para Detecção de Regiões de Anomalia em Termografia de Mama / Stefano Walker Pereira
Pontes. - 2019.

102 f.

Orientador(a): Tiago Bonini Borchartt.

Dissertação (Mestrado) - Programa de Pós-graduação em Ciência da Computação/ccet, Universidade Federal do Maranhão, São Luís, 2019.

1. Câncer de mama. 2. Grafos. 3. Redes complexas. 4. Superpixel. 5. Termografia infravermelha. I. Borchartt, Tiago Bonini. II. Título.

STEFANO WALKER PEREIRA PONTES

**MODELAGEM BASEADA EM GRAFO PARA DETECÇÃO DE
REGIÕES DE ANOMALIAS EM TERMOGRAFIA DE MAMA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFMA como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Dissertação aprovada em ____ de _____ de _____.

Prof. Dr. Tiago Bonini Borchardt
Orientador

Prof. Dr. Geraldo Braz Junior
Universidade Federal do Maranhão

Prof. Dr. Marcus Costa de Araújo
Universidade Federal de Pernambuco

SÃO LUÍS – MA
2019

“Nada temos a temer quanto ao futuro, a menos que nos esqueçamos como Deus tem nos conduzido no passado.”

(Ellen G. White)

AGRADECIMENTOS

Agradeço à minha mãe e aos meus irmãos por me apoiarem nessa caminhada. Família é nosso bem mais importante.

Agradeço à minha esposa, por ser compreensiva e companheira, por ajudar a controlar minhas horas de sono, e por segurar as pontas nos momentos mais difíceis.

Agradeço ao professor Tiago Bonini, por ser, mais uma vez, um orientador atencioso, conselheiro e revisor deste trabalho. Agradeço ao professor Geraldo Braz pela sua disponibilidade, me ajudando em algumas tentativas desse trabalho. Ao professor e amigo Raimundo Campos, do DEDIR, pelos bons papos acerca da elaboração de trabalhos acadêmicos. E aos demais professores e amigos do PPGCC/UFMA que fizeram dessa jornada uma experiência única.

Agradeço, também, à gentil pesquisadora Claire Donnat, da Universidade de Stanford, pelos esclarecimentos acerca da biblioteca SNAP e seus projetos, e pelos preciosos conselhos para este trabalho.

A Deus pelo sustento, paz e saúde que concedeu a mim em todo esse tempo.

RESUMO

O tipo de câncer mais comum entre as mulheres ocorre nas mamas. Como a prevenção ainda é difícil, por vários fatores, o diagnóstico precoce é primordial para aumentar a taxa de sobrevivência dos acometidos. Por isso são dedicados muitos esforços para aprimorar os métodos de diagnósticos, melhorando a qualidade, precisão e custo dos exames de diagnóstico por imagem. Nesse cenário surge termografia como um exame promissor, com uma série de vantagens, mas que ainda está em evolução. Esse trabalho tem como objetivo oferecer uma metodologia de detecção de região de anomalias em termografia de mama. Nesta metodologia são combinadas segmentação de superpixel, construção do grafo e análise de medidas de redes complexas. Na construção do grafo, cada superpixel se torna um nó do grafo e as relações de adjacência entre as regiões se tornam as arestas, cujos pesos são determinados através de funções. Aos grafos gerados são aplicadas variações de limiarização para realçar as características de redes de pequeno mundo. Após vários experimentos envolvendo medidas de grafo e análise de comunidades, as regiões detectadas são avaliadas conforme método avaliativo proposto. A avaliação da detecção realizada nos experimentos da metodologia forneceu resultados que ainda podem ser aprimorados, em relação à detecção da exata região das anomalias, com percentual de 59,1%. Contudo a mesma propiciou resultados expressivos, no sentido de indicar em qual das mamas há anomalia, com percentual de acerto em 86,4%.

Palavras-chave: Termografia infravermelha, câncer de mama, grafo, redes complexas, superpixel, SLIC, centralidade de autovetor.

ABSTRACT

The most common type of cancer, among women, occurs in the breasts. As the prevention is still difficult, due to several factors, early diagnosis is primordial in order to increase the survival rate of those affected. Therefore, much effort is dedicated to improving diagnostic methods by increasing of accuracy, and reduction of the imaging tests cost. In this context, thermography appears as a promising examination, with a series of advantages, but still in evolution. This work aims to provide a methodology for detection of abnormalities regions in breast thermography. In this methodology are combined superpixel segmentation, graph construction and analysis of complex network measurements. In the construction of the graph, each superpixel becomes a node of the graph and the adjacency relations between the regions becomes the edges, whose weights are determined by functions. The generated graphs are applied threshold variations to highlight the characteristics of small world networks. After several experiments involving graph measurements and community analysis, the detected regions are evaluated according to the proposed evaluation method. The evaluation of the detection performed in the methodology experiments provided results that can still be improved, in relation to the detection of the exact region of the anomalies, with a percentage of 59.1 %. However, it produced significant results, in order to indicate in which breast there is an anomaly, with a success rate of 86.4%.

Keywords: Infrared thermography, breast cancer, graph, complex networks, superpixel, SLIC, eigenvector centrality.

LISTA DE ILUSTRAÇÕES

Figura 1 – Ilustração da estrutura interna da mama feminina.	26
Figura 2 – Divisão da mama em quadrantes	28
Figura 3 – Exemplos de grafo não-direcionado (a) e grafo direcionado (b).	35
Figura 4 – Exemplo de um grafo desconexo com 3 componentes conectados.	36
Figura 5 – Exemplo de transformação dos tipos de redes quanto à natureza de suas arestas.	36
Figura 6 – (a) Um grafo aleatório, modelo de Erdős e Rényi (1959). (b) A distribuição de graus para uma rede com 10.000 nós, usando uma probabilidade $p=0,2$. Cada ponto no gráfico corresponde a média de 10 redes.	38
Figura 7 – A construção de uma rede de pequeno mundo de acordo com Watts and Strogatz (1998). Para cada aresta é aplicado uma religação aleatória com probabilidade p	38
Figura 8 – O modelo de rede <i>small world</i> de Watts e Strogatz: (a) um exemplo de rede com $N=64$ nós, $k=2$, $p=0,1$. (b) Uma distribuição da graus para uma redes com 10.000 nós, usando uma probabilidade $p=0,2$. Cada ponto no gráfico corresponde a média de 10 redes.	39
Figura 9 – (a) Exemplo de uma rede livre de escala (modelo de Barabási e Albert).(b) Distribuição dos graus de conexão para uma rede livre de escala. A distribuição segue uma lei de potência. Eixos em escala logarítmica. Gráfico referente a média de 10 redes.	40
Figura 10 – Exemplos de tipos de centralidade para um mesmo grafo: (a) Centralidade de Grau, (b) Centralidade de Proximidade, (c) Centralidade de Intermediação, (d) Centralidade de autovetor. Os pontos vermelhos possuem maior pontuação de centralidade.	50
Figura 11 – Exemplo de uma rede com comunidades destacadas.	51
Figura 12 – (a) Uma imagem real com a tesselação <i>quadtrees</i> . (b) O grafo de adjacência da região associado à partição <i>quadtrees</i>	52
Figura 13 – Duas diferentes tesselações irregulares de uma imagem usando métodos de segmentação de super-pixel: (a) Watershed (Vincent; Soille, 1991) e (b) SLIC (ACHANTA et al., 2011).	53

Figura 14 – Comparação de técnicas de geração de superpixels. Turbopixels (LEVINSHTEIN et al., 2009), Quick-shift (VEDALDI; SOATTO, 2008) e SLIC (ACHANTA et al., 2011).	53
Figura 15 – Etapas da Metodologia proposta	58
Figura 16 – Resultado da segmentação da termografia da mama: (a) o termograma original obtido pela câmera térmica; (b) a representação da imagem em escala de cinza obtida na etapa de Aquisição; (c) a máscara da ROI; (d) a aplicação da máscara da ROI sobre a imagem em escala de cinza.	59
Figura 17 – Divisão da ROI em regiões de superpixels: (a) a imagem original; (b) a ROI segmentada e com ajuste de contraste e filtro da média; (c) aplicação de SLIC com aproximadamente 30 superpixels e (d) com aproximadamente 250 superpixels.	60
Figura 18 – Exemplos de aplicação de SLIC em 4 termogramas diferentes, com 200 sementes, e compactação $m = 10$	60
Figura 19 – Aplicação de SLIC com variação na quantidade de sementes. (a) termograma inicial com CLAHE e filtro da média; (b) aplicação de SLIC com 100 sementes, (c) 200, (d) 400, (e) 600 e (f) 800 sementes.	60
Figura 20 – Processo de construção de grafo a partir das regiões de superpixels. (a) uma amostra de regiões; (b) representação de um grafo considerando somente as arestas do nó central; (c) o subgrafo final com os pesos das arestas calculados e normalizados.	62
Figura 21 – Processo de construção do modelo de grafo, seguindo a modelagem direta. (a) termografia inicial; (b) após aplicar SLIC com 100 sementes, gerando 67 superpixels; (c) o grafo construído plotado em cima da imagem inicial, seguindo o esquema de Modelagem 1; (d) o grafo destacado.	63
Figura 22 – Resultados gerados a partir dos dois tipos de modelagens.	63
Figura 23 – Processo de limiarização de grafo. (a) amostra do grafo original com pesos normalizados; (b) o resultado de limiarização com $L=0,056$ correspondendo ao percentil de 20%, eliminando as arestas com o peso menor que 0,056; (c) o resultado para $L=0,238$, que corresponde ao percentil de 60%.	65
Figura 24 – Diferentes graus de limiarização com a modelagem direta.	65

Figura 25 – Diferentes graus de limiarização com a modelagem inversa.	66
Figura 26 – Resultado do efeito de adição de critério de limiarização na modelagem inversa. Ampliação mostrando parte da mama esquerda, que contém anomalia nos quadrantes superiores. Ambas imagens com 400 sementes, limiarização com $L=60\%$. No grafo à esquerda (a), existe um espaço sem conexões. No grafo à direita (b), a mesma região não sofre limiarização por causa da adição do critério de não eliminação das arestas entre os nós mais quentes.	66
Figura 27 – Diagrama de validação de detecção para os todos os quadrantes da mama esquerda.	69
Figura 28 – Diagrama de validação de detecção para casos em que anomalia estiver quadrante central (a), ou em dois quadrantes inferiores (b).	69
Figura 29 – Comparação de segmentação de comunidades para cada modelagem.	72
Figura 30 – Segmentação de comunidades a partir dos grafos. SLIC com 600 sementes com (a) 0%, (b) 20%, (c) 40%, (d) 60% e (e) 80% de limiarização.	72
Figura 31 – Experimento I-a sobre o exame n° 202: (a) termograma com demarcação da anomalia no QSE-MD; (b) grafo correspondente; (c) divisão em comunidades; (e - h) em cor branca as comunidades que se destacaram em relação a algumas medidas.	74
Figura 32 – Experimento I-a sobre o exame n° 286: (a) termograma com demarcação da anomalia no QSE-ME; (b) grafo correspondente; (c) divisão em comunidades; (e - h) em cor branca as comunidades que se destacaram em relação a algumas medidas.	75
Figura 33 – Experimento I-a sobre o exame n° 292: (a) termograma com demarcação da anomalia no QSE-MD e QSD-MD; (b) grafo correspondente; (c) divisão em comunidades; (e - h) em cor branca as comunidades que se destacaram em relação a algumas medidas.	75
Figura 34 – Experimento II-a sobre o exame n° 286: (a) termograma com demarcação da anomalia no QSE-ME; (b) grafo; (c) divisão em comunidades; (e - h) em cor branca as comunidades que se destacaram em relação a algumas medidas.	77

Figura 35 – Experimento III sobre o exame nº 291 com $S = 600$. O maior ponto azul plotado possui a maior centralidade e aponta para a região com suspeita de anomalia. O retângulo indica o quadrante correto que contém a anomalia. Níveis de avaliação: (a) D; (b) D; (c) D; e (d) A.	79
Figura 36 – Experimento III sobre o exame nº 281 com $S = 600$. Níveis de avaliação: (a) D; (b) D; (c) A; e (d) A.	79
Figura 37 – Experimento III sobre o exame nº 210 com $S = 600$. Níveis de avaliação: (a) D; (b) D; (c) D; e (d) A.	79
Figura 38 – Experimento IV sobre o exame nº 281 com $S = 600$. Níveis de avaliação: (a) D; (b) A; (c) A; e (d) A.	81
Figura 39 – Experimento IV sobre o exame nº 210 com $S = 600$. Níveis de avaliação: A para todos os limiares.	81
Figura 40 – Experimento IV sobre o exame nº 291 com $S = 600$. Níveis de avaliação: B para todos os limiares.	81
Figura 41 – Experimento IV sobre o exame nº 291 com $S = 200$. Níveis de avaliação: A para todos os limiares.	82
Figura 42 – Experimento V sobre o exame nº 291 com $S = 200$. Níveis de avaliação: (a) D; (b) D; (c) D; e (d) A.	83
Figura 43 – Experimento V sobre o exame nº 281 com $S = 200$. Níveis de avaliação: (a) D; (b) D; (c) D; e (d) A.	84
Figura 44 – Experimento V sobre o exame nº 210 com $S = 100$. Níveis de avaliação: (a) D; (b) D; (c) D; e (d) A.	84
Figura 45 – Experimento VI sobre o exame nº 281 com $S = 600$. Níveis de avaliação: (a) B; (b) A; (c) A; e (d) A.	85
Figura 46 – Experimento VI sobre o exame nº 210 com $S = 600$. Níveis de avaliação: A para todos os limiares.	85
Figura 47 – Experimento VI sobre o exame nº 291 com $S = 600$. Níveis de avaliação: (a) A; (b) B; (c) B; e (d) B.	86
Figura 48 – Experimento VI sobre o exame nº 291 com $S = 800$. Níveis de avaliação: A para todos os limiares.	86
Figura 49 – Experimento VI sobre vários exames com $S = 600$. Níveis de avaliação: A = verde, B = amarelo, D = vermelho.	88

LISTA DE TABELAS

Tabela 1 – Resultados de Classificação nas Metodologias dos Trabalhos Relacionados	23
Tabela 2 – Comparação dos tipos de exames de mama.	32
Tabela 3 – Comparação dos esquemas de modelagem de grafo.	63
Tabela 4 – Lista de experimentos de detecção.	67
Tabela 5 – Níveis de avaliação da detecção para os casos genéricos.	68
Tabela 6 – Modularidade da rede para cada quantidade de sementes de SLIC, considerando todos os exames selecionados.	71
Tabela 7 – Resultados de avaliação do Experimento I-a. Parâmetros: modelagem direta; S=600; L=40%.	73
Tabela 8 – Resultados de avaliação do Experimento I-b. Parâmetros: modelagem direta; S=600; L=60%.	73
Tabela 9 – Resultados de avaliação do Experimento II-a. Parâmetros: modelagem inversa; S=600; L=40%.	76
Tabela 10 – Resultados de avaliação do Experimento II-b. Parâmetros: modelagem inversa; S=600; L=60%.	76
Tabela 11 – Resultados de avaliação do Experimento III. Parâmetros: modelagem inversa; S=600; L de 0 a 80%.	77
Tabela 12 – Resultados de avaliação do Experimento IV. Centralidade de grau. .	80
Tabela 13 – Resultados de avaliação do Experimento V. Centralidade de proximidade.	83
Tabela 14 – Resultados de avaliação do Experimento VI. Centralidade de autovetor.	87
Tabela 15 – Percentual de distribuição de anomalia entre os quadrantes das termografias utilizadas.	88
Tabela 16 – Resumo dos melhores resultados de avaliação de todos os experimentos.	89
Tabela 17 – Trabalhos que contribuem com detecção da região da anomalia (em algum dos níveis de aproximação) em comparação com o presente trabalho.	89

LISTA DE ABREVIATURAS E SIGLAS

DMR	<i>Digital Mastological Research</i> (base de termografias)
INCA	Instituto Nacional do Câncer
QC	Quadrante Central
QI	Quadrantes Inferiores
QIE	Quadrante Inferior Externo
QII	Quadrante Inferior Interno
QS	Quadrantes Superiores
QSE	Quadrante Superior Externo
QSI	Quadrante Superior Interno
TE	Termografia Estática
TD	Termografia Dinâmica
ROI	<i>Region of Interest</i>
SLIC	<i>Simple Linear Iterative Clustering</i>

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Motivação	16
1.2	Objetivos	18
1.2.1	Objetivos Específicos	18
1.3	Contribuições do trabalho	19
1.3.1	Organização do Trabalho	19
2	TRABALHOS RELACIONADOS	20
2.1	Trabalhos relacionados a termografia de mama	20
2.1.1	Resultados e considerações sobre os trabalhos relacionados a termografia de mama	23
2.2	Trabalhos relacionados a grafos	24
3	FUNDAMENTAÇÃO TEÓRICA	26
3.1	Mama	26
3.1.1	Estrutura Interna da Mama	26
3.1.2	Quadrantes da mama	27
3.1.3	Tumores de mama	28
3.1.3.1	Anomalias Malignas	29
3.1.3.2	Anomalias Benignas	30
3.1.4	Exames para detecção de câncer de mama	31
3.1.4.1	Termografia de mama	33
3.2	Teoria dos grafos	34
3.2.1	Definições básicas	34
3.2.2	Modelo de redes complexas	36
3.2.3	Medidas de Redes Complexas	40
3.2.3.1	Grau	41
3.2.3.2	Ordem, tamanho e densidade de grafo	42
3.2.3.3	Coefficiente de agrupamento	42
3.2.3.4	Entropia e energia	44
3.2.3.5	Coefficiente de Pearson	44
3.2.3.6	Centralidade	46

3.2.3.7	Cliques	49
3.2.3.8	Comunidades	50
3.2.4	Representação de imagem por modelagem baseada em grafo .	51
4	MATERIAIS E MÉTODOS	56
4.1	Base de termografias	56
4.1.1	Seleção das termografias	56
4.2	Metodologia proposta	57
4.2.1	Pré-processamento	58
4.2.1.1	Conversão das matrizes de temperatura em imagens	58
4.2.1.2	Extração da Região de Interesse	59
4.2.2	Segmentação de regiões de superpixels	59
4.2.3	Modelagem de grafo a partir de termograma	61
4.2.4	Limiarização de grafo	64
4.2.5	Detecção da região da mama com anomalia	66
4.2.6	Avaliação da detecção da região da mama suspeita de anomalia	68
5	Resultados e Discussão	70
5.1	Experimento I	71
5.2	Experimento II	75
5.3	Experimento III	77
5.4	Experimento IV	80
5.5	Experimento V	82
5.6	Experimento VI	84
5.7	Discussões gerais acerca dos experimentos	86
6	CONCLUSÕES	90
	Referências	92

1 INTRODUÇÃO

O câncer de mama é o segundo tipo de câncer que mais faz vítimas em todo o mundo (BRAY et al., 2018) e é o tipo mais comum entre as mulheres (World Health Organization, 2017). De acordo com o Instituto Nacional do Câncer, no Brasil, corresponde a cerca de 25% dos casos novos a cada ano. E estima-se que 29,5% dos casos de câncer entre as mulheres aparecerá nas mamas (INCA, 2019). Para 2018 eram esperados 59.700 casos novos de câncer de mama no Brasil. Excluído o câncer de pele não melanoma, é o mais frequente nas mulheres das regiões Sul, Sudeste, Centro-Oeste e Nordeste. O câncer de mama também acomete homens, porém é pouco frequente, representando apenas 1% do total de casos da doença.

A prevenção desta doença é ainda difícil por causa da ampla variedade de fatores de risco e características genéticas que a causam. Na maioria dos casos o tratamento desta doença é eficaz quando é detectado em estágios iniciais, aumentando a possibilidade de tratamentos menos agressivos e ampliando as taxas de sucesso. O diagnóstico precoce do câncer de mama ocorre quando o paciente ou o profissional de saúde reconhecem os sintomas iniciais do câncer antes que ocorra o avanço da doença. Isso permite maior eficácia e menor custo do tratamento.

Há também o rastreamento por exames de rotina, que é feito por meio de exames de triagem em pessoas que não apresentam sintomas, pretendendo identificar pessoas com anomalias nas mamas, mesmo antes que se manifeste qualquer sinal de câncer. Após análise mais detalhada por meio de diversos exames, as mamas com anomalias têm diagnóstico da doença confirmado ou rejeitado (PENG et al., 2016).

Dentre a variedade de exames, a mamografia é o principal meio usado para detectar o câncer precocemente. Contudo, a mamografia apresenta algumas limitações para a detecção de tumores em pacientes jovens por causa da densidade da mama, que interfere na identificação de massas e microcalcificações por raio-X (WISHART et al., 2010; USUKI et al., 1998; YASMIN; SHARIF; MOHSIN, 2013).

Seguindo a orientação da Organização Mundial da Saúde e de países que adotam o rastreamento mamográfico, o Instituto Nacional do Câncer sugere que a realização de exame de mamografia seja feito a cada dois anos para mulheres entre 50 e 69 anos (INCA, 2019). Segundo Suhrke et al. (2011), pacientes jovens devem evitar a repetição

de exames de mamografia em intervalo de tempo curto. Segundo os autores, a exposição frequente à radiações ionizantes como o raio-X pode elevar em até cinco vezes as chances de desenvolverem câncer.

1.1 Motivação

Recentemente a termografia infravermelha de mama (também chamada de imagem térmica de mama) tem tomado certa atenção nesse campo de estudo, por oferecer a capacidade de ser ferramenta alternativa ou complementar à mamografia em determinados casos (SATTAROV; MCINTYRE; MOTOWIDLO, 2015; WANG et al., 2014; LUNA; DELGADO, 2010). Contudo, esses dois tipos de exames têm características distintas. Enquanto a mamografia é um exame estrutural, a termografia é um exame fisiológico e funcional (BEZERRA et al., 2013).

A imagem térmica permite mapear um corpo ou uma região com a intenção de distinguir as diferentes temperaturas. A análise da termografia se baseia em um ensaio não destrutivo, pois utiliza câmeras e sensores infravermelhos para detectar a radiação infravermelha emitida pelas superfícies dos corpos e, assim, calcular a temperatura real dos objetos (KENNEDY; LEE; SEELY, 2009). A câmera de termovisão transforma uma radiação infravermelha invisível ao olho humano em uma imagem visível. Detecta a energia emitida por um objeto, modifica a frequência da energia recebida e produz uma imagem correspondente, contudo na faixa visível do espectro eletromagnético (FLIR, 2017).

Para as mamas, a termografia se tornou um método promissor de auxílio à detecção do câncer da mama (GERASIMOVA et al., 2013), pois alterações fisiológicas, tais como um aumento anormal da atividade metabólica e de circulação vascular causam anomalias na anatomia dos tecidos da mama. Essas anomalias se refletem em alterações de temperatura. Um termograma anormal é considerado um relevante indicador de risco para a existência ou desenvolvimento de patologias mamárias (WIECEK et al., 1999).

Termografia é indicado para pacientes de qualquer idade, não possuindo efeitos colaterais, podendo ser realizado quantas vezes for necessário. Este exame não tem sua performance afetada por causa da densidade das mamas de mulheres mais jovens (NG, 2009). A termografia também pode ser um exame complementar para detectar alterações em mamas submetidas a quimioterapia, quando os seios estão bastante debilitados para

serem submetidos a mamografia. Quando associado à mamografia, a termografia aumenta os índices de detecção de câncer em até 95% (KEYSERLINGK et al., 2000). Nesta última década, os trabalhos para a detecção de tumores com termografia têm alcançado índices de acurácia acima de 90% (NG, 2009) (SILVA et al., 2016).

Sabe-se que a imagem infravermelha aplicada à medicina se tornou objeto de estudo em meados de 1950. Um dos primeiros relatos em 1956 cita a imagem infravermelha envolvida em trabalhos visando a detecção precoce do câncer de mama (WEIDNER et al., 1991). Anos mais tarde, estendeu-se a outras áreas médicas, como doenças dos sistemas esquelético e neuromuscular, reumatologia, trombose venosa, manejo e controle da dor, temperatura da pele durante o exercício, cirurgias de revascularização miocárdica, cirurgia ortopédica, etc (HILDEBRANDT; RASCHNER; AMMER, 2010) (SENGUPTA; WEBB, 2001).

Contudo, em 1979, um projeto de detecção de câncer de mama relatou resultados negativos (GUTIERREZ-DELGADO et al., 2011). Por essa causa, o interesse por essa modalidade de imagem diminuiu, apontando o custo elevado como principal fator de desvantagem. Então os esforços foram concentrados em exames que apresentavam resultados mais precisos, como a mamografia. Após isso, os avanços na termografia e a redução nos custos dos equipamentos, renovaram o interesse.

Ainda sobre a época dos estudos pioneiros, entre 1956 e 1963 foi demonstrado que o câncer de mama aumenta a temperatura local da pele e pode ser medido com alta precisão por imagens infravermelhas (WEIDNER et al., 1991) (LAWSON; CHUGHTAI, 1963). Também foi sugerido que as mudanças de temperatura captadas poderiam ser explicadas tanto pelo aumento do fluxo vascular, quanto pelo aumento do metabolismo. Verificou-se que a elevação na temperatura poderia ocorrer devido à hipervascularidade, à angiogênese tumoral, ou devido à vasodilatação resultante da liberação de óxido nítrico pelas células do tumor de mama. Tal teoria foi confirmada anos mais tarde por outros trabalhos (WEIDNER et al., 1991) (FOLKMAN, 1971).

Processamento de imagens e detecção assistida por computador são ferramentas viáveis que desempenham um importante papel na análise de termografias. Uso de técnicas de processamento de imagens atreladas a métodos de aprendizado de máquina têm ajudado a reduzir a taxa de diagnósticos falso-positivos e evitar o viés associado à análise do médico das imagens em infravermelho. Análises estatísticas, por assimetria, por classificação de redes neurais artificiais e reconhecimento automático de tumor são tecnologias ainda em

estudo e desenvolvimento (GUTIERREZ-DELGADO et al., 2011).

Nas regiões isoladas dos grandes centros urbanos, em que as pessoas têm pouco ou nenhum acesso aos serviços de exames de rastreamento (sobretudo áreas subdesenvolvidas, áreas rurais, ou áreas de acesso precário a saúde), a termografia pode se tornar o exame mais viável para os serviços de rastreamento, pois é um exame não-invasivo, indolor e não possui restrição de repetição. Além de possuir um custo muito menor (de investimento e portabilidade de instalação), quando comparado com métodos tradicionais como a mamografia, ultrassonografia, ou ressonância magnética (GUTIERREZ-DELGADO et al., 2011).

1.2 Objetivos

Tomando em consideração as motivações apresentadas, essa dissertação tem como objetivo geral a proposição de uma metodologia de auxílio a detecção de áreas suspeitas de anomalias em termografias de mama, a partir da modelagem das regiões de imagem em forma de grafo e posterior análise de medidas de redes complexas. O presente trabalho possibilita um complemento a esta área de pesquisa, pois pode fornecer às metodologias de classificação existentes um aporte para indicação das regiões com suspeita de anomalia, visto que essas tais metodologias não oferecem a detecção da região tumoral.

1.2.1 Objetivos Específicos

Na trajetória em busca do objetivo geral, foram listados como objetivos específicos os seguintes:

- Desenvolver modelagem de grafo a partir de imagens utilizando: técnicas de pré-processamento de imagens; técnicas de segmentação de regiões com agrupamento de pixels; e construção de representação de imagem por meio de grafo.
- Investigar técnicas de análise de grafo que mais se adaptam ao problema.
- Experimentar técnicas de análise de grafos para rastrear as regiões da mama com suspeita de anomalias.

- Validar a performance dessa metodologia por meio de verificação dos diagnósticos presentes na base de termografias.

1.3 Contribuições do trabalho

Há na literatura trabalhos que tratam de temas diversos relacionados a termografia, dentre eles: criação de base de imagens térmicas, métodos de segmentação de mamas e métodos de classificação. Contudo, são poucos os trabalhos que buscam a detecção do local da anomalia, considerando o tipo de exame em questão. Na área de grafos, existem trabalhos promissores que usam da ideia de representar imagem por meio de grafo, afim de extrair características de redes complexas. Desse modo, é possível fazer análises de pontos da rede que se destacam para essas características. Por esses motivos, como acréscimo ao conhecimento acadêmico, este trabalho traz as seguintes contribuições:

- A proposição de uma metodologia que usa grafos como estruturação do modelo, para posterior análise de regiões de imagens.
- Uma metodologia de rastreamento de anomalias em imagens térmicas de mama, que se utiliza da aplicação de teoria de redes complexas no objetivo de obter regiões de mama suspeitas de possuir anomalia.

1.3.1 Organização do Trabalho

Esta dissertação contém ainda mais 5 capítulos, totalizando 6. No capítulo 2 são apresentados os trabalhos relacionados com esta pesquisa, tanto para área da termografia como para área de grafos. O capítulo 3 detalha todos os conceitos sobre anatomia da mama, tumores de mama e termografia. Também é apresentada a fundamentação matemática e computacional necessária para compreensão das etapas da metodologia. No capítulo 4 são descritos os materiais e métodos empregados neste trabalho. No capítulo 5 são expostos os resultados obtidos juntamente com as discussões acerca dos experimentos realizados. Ao final, no capítulo 6, são apresentadas as conclusões deste trabalho.

2 TRABALHOS RELACIONADOS

A fim de considerar todos os trabalhos relacionados, foi realizada uma revisão sistemática de literatura baseada nas diretrizes definidas nos trabalhos Biolchini et al. (2005) e Kitchenham et al. (2004). Observou-se que na literatura existem diversos trabalhos que apresentam metodologias para diagnóstico utilizando classificação entre mamas com anomalia versus mamas normais. A maioria desses trabalhos utiliza a base de imagens mastológicas da Universidade Federal Fluminense, *Database for Mastology Research with Infrared Image* (DMR), que possui termografia com dois tipos de protocolo de aquisição: o protocolo estático (doravante abreviado por DMR-TE) e o protocolo dinâmico (abreviado por DMR-TD) (SILVA et al., 2014b).

Como trabalhos relacionados, neste capítulo serão apresentadas as pesquisas mais relevantes, divididas nas áreas de classificação de termografia de mama (Seção 2.1) e modelagem de grafo (Seção 2.2).

2.1 Trabalhos relacionados a termografia de mama

Em Serrano (2010) e Resmini (2011), usando a base DMR-TE, são propostas metodologias para o diagnóstico precoce de câncer utilizando várias medidas, como coeficiente de Hurst, fractais, lacunaridade, etc. Na tese de Borchardt (2013), é descrita uma metodologia para auxílio à detecção de anomalias nas mamas, que, segundo o autor, pode ser aplicada a imagens termográficas provenientes de vários tipos de protocolos. Com a base DMR-TE, foi efetuada uma etapa de extração de características com medidas estatísticas, histograma, dimensão fractal de Higuchi, Índice de Moran, Coeficiente de Geary e função K de Ripley. Posteriormente foi feita a classificação com Máquina de Vetores de Suporte (SVM), que teve seus resultados de classificação otimizados por uma técnica de algoritmo genético. A classificação foi validada com medidas de acurácia, especificidade e sensibilidade.

No trabalho de Gerasimova et al. (2012) é mencionada a utilização de 46 imagens com tumor entre malignos e benignos. Na etapa de extração de características, o trabalho usa métodos com análise de Fourier, transformação de Wavelet e diagramas de fase. As conclusões argumentaram que o uso de termografia era viável e serviram de inspiração

para os trabalhos de outros autores em anos seguintes.

Em outro trabalho, Gerasimova et al. (2013) apresentam detalhes de sua aquisição, onde utiliza 9 termografias dinâmicas, sendo 6 doentes e 3 saudáveis. Foi utilizado uma análise multifractal sobretudo com o método *Wavelet Transform Modulus Maxima* (WTTM), para extração de características, sem passar por uma etapa de segmentação. Foram realizados vários testes que buscam os valores dos melhores coeficientes de WTTM que melhor classificam os exames.

Dentre os trabalhos que utilizaram a base DMR-TD, Silva et al. (2014a) apresenta a maneira como constrói as séries temporais. Depois extrai características das séries temporais, que são: complexidade e mobilidade do sinal, entropia e dimensão fractal. Para cada uma dessas características aplicou as seguintes medidas estatísticas para a avaliação e classificação: média, desvio padrão, amplitude total, assimetria e curtose.

Utilizando estratégia um pouco semelhante às anteriores (em 2012 e 2013), Gerasimova et al. (2014) usa uma base composta por 33 termografias com tumor, com formato 256 x 320 pixels. Os resultados reportados correspondem ao uso de espectro de potência médio, funções de partição, espectros de singularidade e *Wavelet Probability Distribution Function* (WPDF), sobre 64 termografias, com janela de tamanho 8 x 8.

Os autores Saniei et al. (2015) relatam a aquisição de 50 termogramas dinâmicos (sendo 25 com tumor). Nesse trabalho, primeiro a região do corpo é separada do fundo por meio de componentes conectados. Depois, a região das mamas é separada do restante do corpo utilizando técnicas de aproximação de curvas parabólicas e áreas circulares. Em seguida é feita outra etapa de segmentação visando esqueletizar a rede de vasos sanguíneos. Por fim, os autores extraem características (*thermal minutia points - TMP*) dos pontos de ramificação das redes vasculares das mamas. Posteriormente usa técnica *TMP matching*, que é comumente usada para reconhecimento de impressões digitais.

Em Alves (2015) foram extraídos grupos de características, como estatística baseada em agrupamento, comparação de histogramas, geometria fractal, índices de diversidade e estatística espacial. Dentre essas características cita-se lacunaridade, temperatura média, energia horizontal e valor do maior centroide. A partir destas medidas, foram implementados métodos para extrair 6 outras características das respectivas séries temporais.

Em Silva (2016) (que é extensão da tese de Silva (2015)), usando também a base DMR-TD com 70 imagens, os autores propõem um novo método de análise, para

indicar o risco de câncer, usando técnicas de aprendizado de máquina supervisionado e não-supervisionado, chamando, assim, a metodologia como híbrida. Ao montar as séries de temperaturas, são utilizados quadrados de 11 x 11 pixels, tomando a temperatura mais alta de cada uma dessas janelas. Depois é feita a clusterização das séries de temperaturas e avaliação dos clusteres. Após isso, faz seleção de característica e construção do modelo de classificação.

O trabalho Andrade et al. (2017) também seguiu a mesma linha de segmentação e extração de característica, diferenciando pelo fato de usar a análise das mamas esquerda e direita de forma individual. Após alguns testes primários, utilizou-se a estratégia de reamostragem dos exames dividindo a série em janelas de 2 tempos. As características usadas foram comprimento da linha da série, área sob a curva, energia média, amplitude média dos vales, amplitude média dos picos, média, desvio padrão, assimetria e curtose.

Por fim, como último trabalho usando a base DMR-TD, relata-se o trabalho de Graduño-Ramón (2017). Nessa pesquisa a segmentação acontece com a separação do fundo e posteriormente com a segmentação da ROI, separando a mama em esquerda e direita. Esta técnica de segmentação automática consiste em detectar as bordas que definem a prega mamária, abaixo dos seios. Com essas curvas, usa-se equações polinomiais de segundo grau, estimadas através de regressão linear. Também é utilizado a técnica de *watershed* para segmentar as regiões mais quentes, para futuras análises acerca do local do câncer. Para classificação dos exames, a mama esquerda é comparada com a direita observando as medidas estatísticas: média, temperatura máxima, temperatura mínima e desvio padrão de cada região da mama segmentada. Se a diferença entre uma dessas medidas for menor que 1 °C, as mamas são classificadas como saudável.

Seguindo a linha de análise de variáveis anatômicas, o trabalho de Wilkinson, Amri e Pulko (2016) tem como objetivo o estudo sobre a influência das dimensões do tumor (calor metabólico, profundidade do tumor e diâmetro de um termograma) através de termografias, além de verificar o efeito da densidade da mama no contraste de termografias. De forma geral verificou-se que a espessura da camada de gordura atrapalha o aferimento do tumor. Houve uma certa dificuldade para calcular o diâmetro do tumor, ou a distribuição isotérmica de um ponto quente quando esse possui padrão vascular superlativo. E também percebeu-se a dificuldade de associar o diâmetro do tumor com a sua localização em profundidade.

2.1.1 Resultados e considerações sobre os trabalhos relacionados a termografia de mama

Para facilitar a comparação entre os resultados de classificação de todos os trabalhos apresentados, a Tabela 1 apresenta as seguintes métricas de avaliação: Acurácia¹, Sensibilidade², Especificidade³ e Precisão⁴. Os trabalhos que não possuem resultados com essas métricas se limitaram a realizar pesquisas cujos resultados foram apresentados como análise interpretativa, para fundamentar possíveis correlações sobre os dados obtidos.

Tabela 1 – Resultados de Classificação nas Metodologias dos Trabalhos Relacionados

TRABALHO	Dataset	Acurácia	Sensibilidade	Especificidade	Precisão
Serrano (2010)	DMR-TE	-	-	-	-
Resmini (2011)	DMR-TE	82,14%	91,7%	25%	85,8%
Borchardt (2013)	DMR-TE	85,7%	95,8%	25%	-
Gerasimova et al. (2012) *	indefinida	-	-	-	-
Gerasimova et al. (2013) *	indefinida	-	-	-	-
Gerasimova et al. (2014a) *	DMR-TE	-	-	-	-
Gerasimova et al. (2014)	indefinida	-	76%	86%	-
Saniei et al. (2015)	indefinida	-	86%	38%	-
Alves (2015)	DMR-TD	100%	100%	100%	100%
Silva (2015)	DMR-TD	100%	100%	100%	100%
Amri e Pulko (2016) #	-	-	-	-	-
Andrade et al. (2017)	DMR-TD	95,8%	93,6%	95,9%	-
Garduño-Ramón et al. (2017)	DMR-TE	-	78%	81%	-

* trabalhos cujos resultados são apresentados sob forma de análise e discussões

trabalhos que não tinham objetivo de classificação

DMR-TE: *Digital Mastological Research* - Termografia Estática.

DMR-TD: *Digital Mastological Research* - Termografia Dinâmica.

Como justificativa para a elaboração do trabalho descrito por esta dissertação, a base de imagens adotada foi a DMR-TD, por possuir mais imagens, por ser a base com a aquisição de melhor resolução, por ter um protocolo mais rígido e padronizado, e por ter sido usado nos trabalhos que apresentaram melhores resultados, conforme mostra a Tabela 1.

Ressalta-se que os trabalhos de Alves (2015) e Andrade et al. (2017) foram os que mais contribuíram em relação à indicação de qual das duas mamas contém anomalia, em cada exame, pois esses trabalhos classificam cada mama separadamente. E como o

¹ Acurácia é o percentual de amostras classificadas corretamente considerando todas as classes.

² Sensibilidade mede o percentual de casos positivos corretamente classificados como positivos, ou seja, a presença de anomalias nas mamas.

³ Especificidade avalia a proporção de casos negativos classificados corretamente como casos negativos.

⁴ Precisão ou Valor Preditivo Positivo (VPP) indica a proporção de resultados positivos obtidos que realmente são positivos, e também indica a probabilidade de repetir os mesmos resultados de acurácia.

objetivo de classificação já foi alcançado satisfatoriamente em vários trabalhos, sendo que em dois desses foram completamente alcançados, torna-se admissível a finalidade deste trabalho, que é propor uma metodologia de detecção da região do tumor. Dessa forma, este trabalho contribuirá de forma complementar aos trabalhos de melhores resultados, visto que nenhum dos trabalhos dessa linha se ateu à detecção da localização.

2.2 Trabalhos relacionados a grafos

Foram elencados, a seguir, alguns dos trabalhos mais relevantes que fazem uso de redes complexas e que inspiraram o uso de técnicas de grafo na metodologia apresentada na Seção 4.2. Ressalta-se que os conceitos e a fundamentação teórica sobre grafos mencionados nesta subseção está disposta mais adiante, na Seção 3.2.

Como primeira pesquisa a ser apresentada, de maneira sucinta, o trabalho de Zarbakhsh et al. (2018) faz análise de imagens de mamografia, onde as regiões de cada imagem são mapeadas em um grafo, que posteriormente passa por uma seleção de características, para depois serem classificadas em uma rede neural artificial.

A fim de obter uma abordagem baseada em redes complexas para análise de formas de contornos, os autores Backes et al. (2009) utilizam medições de grau em redes para compor um conjunto de descritores de forma. Em outra pesquisa, Harchaoui e Bach (2007) usaram as representações de imagem em forma de grafo para tarefas de aprendizado, visando a segmentação de regiões alvo, utilizando um método de relacionamento entre núcleos de imagens.

No trabalho de Stam et al. (2006) foi possível mostrar que um cérebro acometido pela Doença de Alzheimer é caracterizado pela perda de características de redes *small world*. Foi comprovado através de análise do grafo modelado (em forma de matrizes de conectividade funcional de canais de eletroencefalografia) que as medidas de comprimento e de agrupamento foram as que mostraram mais evidências da presença da doença. Em um trabalho semelhante, Mahmoudi et al. (2012) investigam o papel funcional de áreas e redes cerebrais.

Um dos trabalhos mais relevantes, que inspiraram a metodologia desta dissertação, foi o trabalho de Velame (2016), que desenvolveu um método de segmentação de imagens com extração de superpixels e detecção de comunidades de grafos. Nesse trabalho usou-se

superpixels reduzindo o número de nós do grafo e, posteriormente, usou algoritmos de detecção de comunidades. Ao final, a metodologia foi avaliada por meio de análise do percentual de acerto sobre o *ground-truth* das imagens.

Conforme apresentado acima, pela versatilidade, grafos podem ser usados para modelar os mais diversos tipos de problemas, provando ser um artifício muito útil para área de visão computacional e o processamento de imagens.

3 FUNDAMENTAÇÃO TEÓRICA

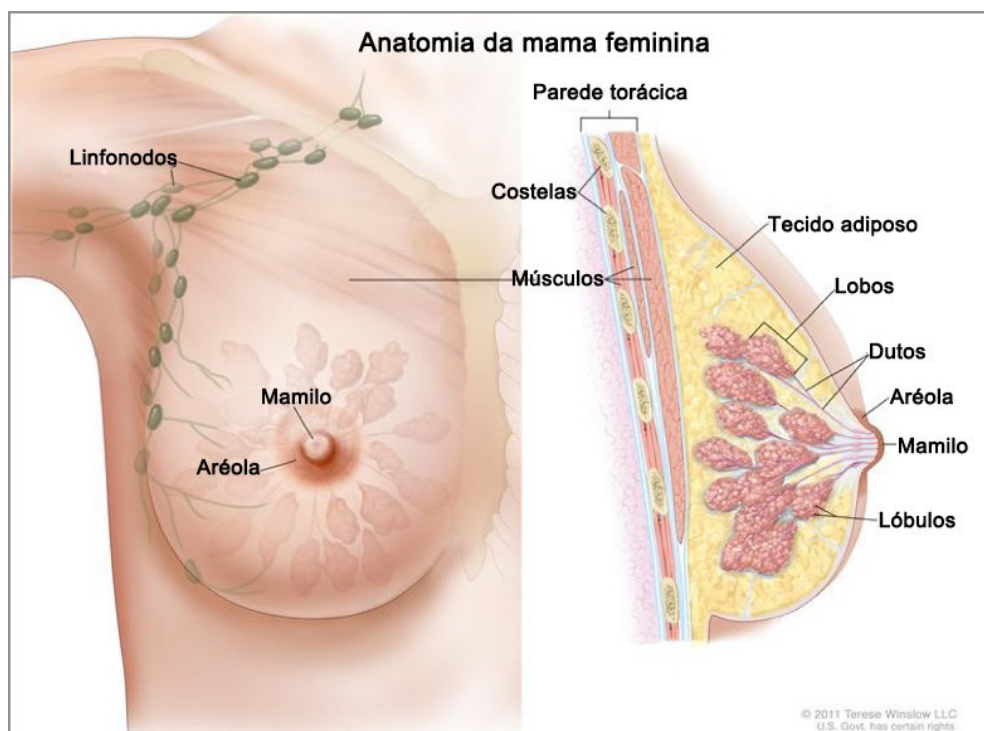
Este capítulo expõe conceitos importantes sobre os temas usados como fundamentação teórica. Tais tópicos se fazem necessários para compreensão das técnicas utilizadas na elaboração da metodologia proposta. Abordam-se os conceitos relacionados a anatomia e tumores de mama, termografia, teoria de grafos e medidas de redes complexas.

3.1 Mama

3.1.1 Estrutura Interna da Mama

As mamas femininas são estruturas glandulares pares situadas na parede anterior e superior do tórax, que derivam de glândulas sudoríparas modificadas (BERNARDES, 2011). As glândulas mamárias têm como principal função a secreção do leite materno, contendo todos os nutrientes e minerais necessários ao recém-nascido. A Figura 1 ilustra as partes mais relevantes da estrutura interna da mama feminina. Algumas dessas são detalhadas a seguir:

Figura 1 – Ilustração da estrutura interna da mama feminina.



Fonte: Adaptado de WOOD (1994)

- Lóbulo mamário (alvéolo): conjunto de ácinos, que são responsáveis pela produção do leite durante a lactação;
- Lobo mamário: conjunto de lóbulos mamários que se liga à papila (mamilo) através de um ducto. Cada lobo é uma glândula mamária, pois cada porção possui sua parte secretora e seu ducto próprio;
- Dutos mamários (Canais lactíferos): conduzem o leite até à papila, através de 15 a 20 canais;
- Aréola: estrutura central da mama onde se projeta a papila, que é a protuberância elástica onde desembocam os ductos mamários;
- Vasos linfáticos: conduzem a linfa da glândula mamária até os gânglios linfáticos (linfonodos), que estão situados perto das axilas, próximo à clavícula.

Em resumo, a mama é composta internamente por tecido glandular, tecido fibroso e tecido adiposo. O tecido glandular é composto por cerca de 20 lobos, cujo topo encontra-se voltado para a superfície, e em conjunto, essa estrutura forma a mama. No tecido fibroso existem os ligamentos de Cooper, que são responsáveis pela retração cutânea nos casos de câncer de mama e são expansões fibrosas que se projetam na glândula mamária. Já o tecido adiposo se refere a todo o resto da mama e serve para preencher o espaço entre os elementos funcionais e estruturais. A quantidade deste tipo de tecido depende da idade da mulher, estado nutricional e de outras características físicas (WOOD, 1994).

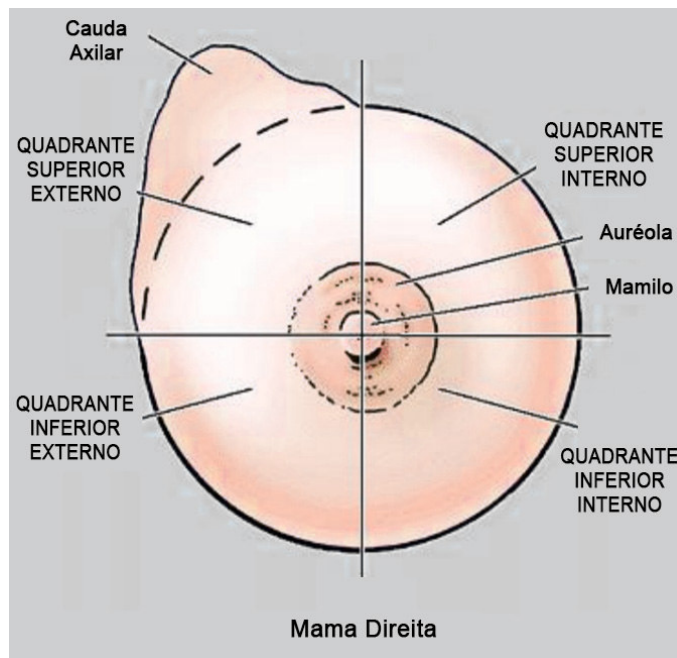
As mulheres mais jovens possuem mamas mais firmes e densas porque apresentam maior quantidade de tecido glandular. Quanto mais a idade avança em direção a menopausa, o tecido mamário vai se atrofiando e sendo trocado progressivamente por tecido gorduroso. Na pós-menopausa as mamas são constituídas quase totalmente por tecido adiposo (TORTORA; PETTI, 2002).

3.1.2 Quadrantes da mama

Para a localização anatômica das alterações observadas na mama e para uma padronização de descrição de tumores, o esquema habitualmente utilizado para a divisão da superfície da mama é em quatro quadrantes: superior externo (QSE), superior interno

(QSI), inferior externo (QIE) e inferior interno (QII) (SILVA, 2013). A Figura 2 mostra uma representação visual dessa divisão.

Figura 2 – Divisão da mama em quadrantes



Fonte: Adaptado de MOORE; DALLEY; AGUR (2006)

A maior quantidade de dutos está localizada no quadrante superior externo da mama (próximo à axila). Como a maioria dos tumores cancerosos são provenientes dos dutos, é nesse quadrante que ocorre a maioria dos tumores malignos da mama, correspondendo a aproximadamente 50%. O restante do percentual de incidência está distribuído quase que igualmente entre a região dos outros quadrantes e do mamilo (MONTEIRO et al., 2003).

3.1.3 Tumores de mama

A multiplicação desordenada de células caracteriza o câncer. Esse processo pode ser causado por mutações genéticas em que células sofrem alterações no seu código genético e perdem sua função essencial passando a se reproduzir exageradamente, formando um tumor, que vai desestruturando e destruindo os tecidos (ZELMANOWICZ, 2002).

O excesso de reprodução celular no local exige a maior quantidade de nutrientes para a região, assim novos vasos sanguíneos vão sendo criados ao redor das células para alimentá-las. Este processo chamado de angiogênese é natural em situações como

cicatrizações de cortes, onde a criação dos novos vasos sanguíneos faz com que a região receba um fluxo maior de sangue. Um suprimento maior de sangue e vasos capilares é solicitado ao organismo e quanto mais nutriente o organismo prover para essas células, mais agressivo será o crescimento do tumor (BUITRAGO; UEMURA; SENA, 2002). Há vários tipos de câncer de mama. Por isso, a doença pode evoluir de diferentes formas. Alguns tipos têm crescimento acelerado, enquanto outros tem desenvolvimento mais lento. Tais comportamentos estão relacionados às características de cada tumor. E o surgimento é bastante influenciado pelo fatores¹ de risco que os provocam.

Além da classificação quanto ao tipo de câncer, existe a categorização por estágios, que permite aos médicos definirem o grau de avanço do câncer e assim indicarem o tratamento mais apropriado (INCA, 2019). No estágio 0 o tumor é do tipo carcinoma *in situ*; No estágio I o tumor tem diâmetro inferior a 2 cm e não se propagou para além da mama; No estágio II o tumor possui diâmetro entre 2 cm e 5 cm ou se espalhou pelo menos a um linfonodo axilar; No estágio III o tumor tem diâmetro superior a 5 cm ou está propagado aos linfonodos; No estágio IV, independentemente do tamanho, sofreu metástase.

Para melhor entendimento e para ajudar na investigação acerca da localização da região da mama onde as anomalias ocorrem, é premente a necessidade de definição dos diversos tipos de anomalias de mamas e de suas características. Alguns tipos mais importantes são apresentados nas subseções seguintes.

3.1.3.1 Anomalias Malignas

O carcinoma é o nome dado ao tipo de câncer mais comum nos humanos. Surge no tecido epitelial, que está distribuído difusamente pelo nosso organismo, pois ele é o responsável pelo revestimento da nossa pele e órgãos (ZELMANOWICZ, 2002).

Os carcinomas de mama são classificados de acordo com a estrutura interna da mama onde eles ocorrem. Eles aparecem comumente nos dutos e nos lóbulos. O tipo mais comum de câncer de mama é o **carcinoma ductal** (CD), que começa nas células dos dutos. Há também o chamado **carcinoma lobular** (CL), que começa nos lobos ou lóbulos

¹ Fatores que favorecem o surgimento de câncer: 1ª menstruação (menarca) muito cedo (antes dos 12 anos); Nunca ter engravidado; Ter engravidado com mais de 30 anos; Última menstruação (menopausa) acima de 55 anos; Doenças benignas na mama (cistos); Terapia de reposição hormonal; Histórico familiar, para ocorrências antes da menopausa; Uso abusivo de álcool; Vida sedentária/obesidade; Nutrição inadequada (MONTEIRO et al., 2003).

e é encontrado na maioria das vezes em ambas as mamas, ao contrário dos outros tipos de câncer. Existe também o **carcinoma lobular *in situ*** (CLIS), que ocorre somente nos lóbulos. A existência de um CLIS em uma mama aumenta o risco do desenvolvimento de câncer na outra mama. Os CLIS podem se tornar **carcinoma lobular invasivo** (CLI) quando as células cancerígenas se espalham a partir dos lóbulos para tecidos adjacentes (BUITRAGO; UEMURA; SENA, 2002). Existem outras nomenclaturas para caracterizar tumores específicos, alguns deles estão detalhados a seguir:

Existe o **carcinoma inflamatório** (CI), que raramente apresenta sinais hormonais, classificado como triplo negativo. Essa é uma das formas mais agressiva de câncer de mama, mas é a forma mais rara. Se apresenta como uma grande extensão de inflamação na mama. Se inicia nas glândulas que produzem leite. Possui uma alta chance de produzir metástase. Nesse tipo de câncer, em estágio avançado, a mama apresenta-se vermelha, inchada e bastante quente, sendo por isso identificado facilmente por meio da termografia. Nesse caso, a pele da mama pode também apresentar uma textura semelhante à casca de laranja (ZELMANOWICZ, 2002).

O **carcinoma mucinoso** (CM) ou coloide da mama é um subtipo de carcinoma ductal invasivo (CDI). Ocorre em larga faixa etária, possuindo sua incidência maior em mulheres de maior idade, correspondendo a cerca de 1% a 7% do total de neoplasias da mama. Do ponto de vista histológico, esta variante é caracterizada por associar-se, na grande maioria dos casos, ao carcinoma ductal *in situ* (CDIS). É detectado pela quantidade da substância mucina, produzida pelo tumor (BIAZÚS; MELO; ZUCATTO, 2009).

Outro tipo de tumor maligno chamado **tumor floide** é definido como um nódulo mamário palpável de grandes dimensões (maior que 3 cm) e de rápido crescimento. Possui aparência similar ao fibroadenoma: lesão oval, redonda ou lobulada. O prognóstico é bom, a não ser que haja metástases (PEREIRA et al., 2012).

3.1.3.2 Anomalias Benignas

Os tumores benignos também podem apresentar crescimento anormal, mas não apresentam um quadro muito grave, visto que o crescimento do tumor se dá por expansão, mas com um ritmo mais lento do que os malignos. A maioria dessas anomalias benignas é limitada a um invólucro fibroso, que não permite a propagação das células defeituosas. As protuberâncias formadas não possuem muita diferença em relação ao tecido sadio.

Operações cirúrgicas para remover o tumor resolvem a maioria dos casos (ROCHA, 2013). Alguns dos tumores benignos mais comuns são destacadas a seguir.

Fibroadenoma são os tumores de mama benignos mais comuns que ocorre com mais frequência em mulheres jovens, principalmente em mulheres com idade entre 15 e 35 anos. Um fibroadenoma se assemelha a um nódulo mole ou firme, com contornos bem definidos. É indolor e apresenta mobilidade quando tocado. Os fibroadenomas em geral não passam de 2 cm. A localização mais recorrente é nas regiões superiores externas da mama. Em 20% dos casos há lesões múltiplas uni ou bilateramente (BERNARDES, 2011).

A **Doença Fibrocística da Mama** apresenta como sintomas nódulos e dores incomuns. Surgem em alguns casos quando as mulheres, antes da menopausa, desenvolvem cistos preenchidos por líquidos ou regiões fibrosas que podem ficar dolorosas imediatamente antes da menstruação. Embora a maioria relate pequeno incômodo mamário durante a vida, algumas mulheres têm, mensalmente, uma dor intensa. Entretanto não é considerado uma doença e não aumenta o risco de câncer mamário, por isso são denominados genericamente por nódulos benignos (BERNARDES, 2011).

Hiperplasia denota um grande e repentino crescimento das células em determinada parte do corpo. Quando acomete as mamas pode acontecer tanto nos ductos como nos lóbulos. É uma anomalia benigna, não produz sintomas nem causa dores. Em casos de incômodo, às vezes é necessário intervenção cirúrgica para remoção da área afetada (BERNARDES, 2011).

A **esteatonecrose da mama** (necrose gordurosa) é um processo patológico que ocorre quando há saponificação localizada, formando áreas esbranquiçadas no tecido adiposo. É um processo inflamatório benigno que está se tornando cada vez mais comum devido à cirurgia conservadora da mama e aos procedimentos de mamoplastia (HORTOLAM; JALES, 2016).

3.1.4 Exames para detecção de câncer de mama

Existe o **autoexame**, que é recomendado para todas as mulheres partir de 20 anos de idade e deve ser feito pelo menos uma vez por mês. Neste autoexame a mulher, usando a polpa dos dedos, apalpa a mama, a aréola e as regiões axilares, pressionando levemente e fazendo movimentos circulares. Caso a mulher encontre caroços, secreções,

proeminências ou retrações na pele, deve procurar um mastologista, que irá realizar exame clínico, podendo indicar a realização de outros exames para diagnóstico por imagem (IBCC, 2019).

Quanto ao tipo de imagem, os exames para o auxílio ao diagnóstico de patologias da mama são classificados como estruturais ou funcionais. Nos exames estruturais são evidenciadas as estruturas internas da mama. Já nos exames funcionais são percebidos o funcionamento dos órgãos e movimentação de fluídos. Como exemplo de exames estruturais temos: mamografia, ultrassonografia e ressonância magnética. Já os funcionais são: ultrassonografia (se encaixa nas duas categorias), ressonância magnética (se realizada com uso de contraste) e termografia. A fim de comparação, a seguir serão detalhados alguns dos exames mais usados para diagnóstico de doenças de mama. As principais características são sumarizados na Tabela 2.

Tabela 2 – Comparação dos tipos de exames de mama.

Exame	Ressonância magnética	Ultrassonografia	Mamografia	Termografia
Tipo	Estrutural e Funcional	Estrutural e Funcional	Estrutural	Funcional
Invasivo	Sim, se fizer uso de contraste	Não	Sim, devido a radiação emitida	Não
Funcionamento	Gera um campo magnético	Captura ondas de ultrassom	Emite radiação ionizante	Captura a radiação infravermelha
Desconforto	Aplicação de contraste intravenoso	Aplicação de gel	Dor pela compressão da mama	Não
Contra-indicação	Alergia ao contraste	Depende da destreza do operador	Não recomendado para mama densa	Não existe
Tipo de imagem	Destaca estrutura de fluxos de sangue	Destaca estrutura de fluxos líquidos	Estruturas e calcificações internas	Mapa de calor da região analisada
Custo	800x	100x	100x	1x

Fonte: Adaptado de Silva (2015)

A **mamografia** é feita por meio da compressão da mama da paciente pelo mamógrafo e pela radiação ionizante incidida sobre a mama. Os feixes de Raios-X lançados sobre a mama são retidos de forma proporcional à densidade dos tecidos mamários. Como pacientes com idade superior a 40 anos possuem mamas menos densas, a mamografia é o método mais recomendado, pois permite identificar com precisão as estruturas internas da mama (SERPEJANTE, 2014).

A **ressonância magnética** (RM) também é usada para diagnóstico de doenças da mama. Possui alta resolução de imagens, em relação aos outros exames, alcançando

ótima sensibilidade na detecção de câncer de mama e lesões que poderiam passar despercebidas em exames tradicionais como a mamografia (BRUENING et al., 2012). A RM produz imagens do interior do corpo humano observando diferentes densidades em diferentes tipos de tecido, mapeando a posição de moléculas de água desses tecidos. No organismo os núcleos dos átomos de hidrogênio se alinham e formam pequenos ímãs, por causa do campo magnético criado pelo aparelho. Então, são aplicadas ondas de rádio através da parte do corpo examinada, produzindo vibrações que são captadas. O computador analisa os sinais captados e os transforma em imagem, mostrando lesões em tecidos sem expor o paciente à radiação. Às vezes é necessária a aplicação de contraste intravenoso. A principal desvantagem da RM é o alto custo do exame, resultante do custo muito elevado do equipamento e sua manutenção.

Ultrassonografia é um exame de diagnóstico por imagem que utiliza emissores sofisticados de feixes sonoros de alta frequência que se propagam através do corpo humano, sendo totalmente inofensivo para os órgãos e além de ser de baixo custo. O exame de ultrassom é feito pelo próprio médico, que manuseia a sonda/sensor sobre a pele utilizando-se gel à base de água, a fim de se eliminar o efeito da impedância sonora do ar no espaço entre o sensor e a região a ser analisada. As imagens são formadas simultaneamente na tela do aparelho. Os ecos de reflexão das ondas sonoras pelo corpo são captados pelo sensor e são transformados em imagens. Em tons de cinza, podem aparecer massas, lesões ou presença de líquidos, assim como alterações na estrutura, altura e composição de órgãos e tecidos moles (IBCC, 2019).

3.1.4.1 Termografia de mama

A termografia é uma técnica que permite mapear a superfície da região de um corpo com o intento de distinguir áreas de diferentes temperaturas, sendo portanto uma técnica que permite a visualização artificial da luz dentro do espectro infravermelho. É através da aplicação da termografia que se obtêm os termogramas (ou imagens térmicas), que são as imagens que foram convertidas pelos sensores termográficos. Estas traduzem em sinais elétricos as radiações infravermelhas emitidas pelas diferentes partes de um corpo. Os termogramas são a visualização artificial da luz dentro do espectro infravermelho, portanto representam o mapa de temperatura em pseudocores, conforme a variação das zonas de temperatura capturada (FLIR, 2017; FLIR, 2015).

As termografias podem ser classificadas em duas formas quanto a sua aquisição. No modo de aquisição estática é capturada apenas uma imagem estando a paciente em equilíbrio térmico com o ambiente. Na aquisição dinâmica, são registradas várias imagens, obedecendo a um intervalo de tempo definido entre cada aquisição. Durante esta aquisição é observada a recuperação da temperatura da pele após um estresse térmico (resfriamento ou aquecimento) (HERMAN, 2013).

A aquisição de várias imagens térmicas no decorrer do tempo, a uma dada taxa de quadros, faz que a variação local das temperaturas de um tecido anormal mostre comportamento diferente, quando comparado a tecido normal. Essas anomalias se justificam pelo fato de que os tumores apresentam grande vascularização anormal, quando comparado a um tecido saudável (ANBAR et al., 2001; SILVA, 2015). Dessa forma, para um dado exame dinâmico, é possível fazer tanto a análise de uma imagem individual, como comparar todas as imagens entre si.

Com a evolução dos equipamentos de captura de termografia e com o aumento na precisão de captação de temperatura (cerca de 0,08 °C de precisão), a termografia se tornou uma tecnologia bastante atrativa e acessível (GERASIMOVA et al., 2013).

3.2 Teoria dos grafos

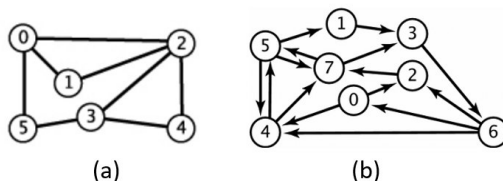
A teoria dos grafos é um ramo da matemática que estuda a relação entre os objetos de um determinado conjunto. Ela pode ser aplicada em várias áreas desde, o estudo de algoritmos, até aplicações na sociologia, biologia e física, trazendo uma verdadeira multidisciplinaridade a essa área (COSTA et al., 2007). Para a área de processamento de imagens, os grafos são geralmente usados em técnicas de segmentação ou representação de modelos.

3.2.1 Definições básicas

Um grafo G consiste em um conjunto de nós ou vértices V e um conjunto de conexões, ligações, arcos ou arestas E entre esses nós: $G(V, E)$. As arestas podem ser denominadas arestas direcionadas ou setas, quando representam uma direção de ligação entre vértices. Nesse caso, o grafo é chamado de grafo direcionado ou digrafo (*directed*

graph). O grafo pode também ser ponderado, quando pesos (w) são atribuído a cada aresta. Tal valor pode representar a força da conexão ou alguma distância física entre os dois vértices conectados (WEST et al., 2001).

Figura 3 – Exemplos de grafo não-direcionado (a) e grafo direcionado (b).



Fonte: Sedgewick e Wayne (2011).

A **distância** entre um par de vértices em um grafo (ou distância geodésica) é o número de arestas em um caminho mínimo conectando o par. Em outras palavras, a distância $d(v, u)$ de um grafo é o comprimento do menor caminho entre v e u (SZWARCFITER, 1988). Se não existe caminho algum conectando dois vértices, ou seja, se eles pertencem a diferentes componentes conectados, então convencionalmente a distância entre eles é definida como infinita. Em termos de distância, podem ser definidas a excentricidade, diâmetro e raio.

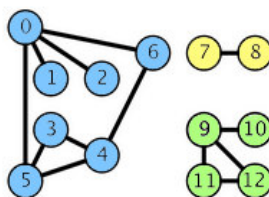
A **excentricidade** ε de um vértice v é definida como a maior distância geodésica entre v e cada um dos outros vértices. Isto é, a distância de um nó ao nó mais distante dele no grafo.

Dito isso, a excentricidade máxima de qualquer vértice do grafo define o **diâmetro de um grafo**, ou seja, a maior distância entre os pares de vértices. Já o **raio de um grafo** é a excentricidade mínima de qualquer vértice do grafo. Se $\varepsilon(v) \approx \text{diam}(G)$, então v é vértice periférico. Se $\varepsilon(v) \approx \text{raio}(G)$, então v é um vértice central.

Para um grafo não-direcionado G , dois vértices quaisquer u e v são classificados como conectados se G contém um caminho de u para v . Caso contrário, eles são ditos desconectados. Seguindo esse raciocínio, um grafo é chamado conexo se todos os vértices estão ligados uns aos outros por um caminho (SEDGWICK; WAYNE, 2011). Um **componente conectado** (ou componente conexo) é um maximal subgrafo conexo de G . Por essa definição, cada vértice e cada aresta pertence exatamente a um componente conectado. Caso o grafo inteiro não seja conexo, ele poderá ter vários componentes conectados. A Figura 4 mostra um exemplo de um grafo e os seus componentes conectados.

A partir do modelo de grafos que possui arestas direcionadas e pesos, é possível

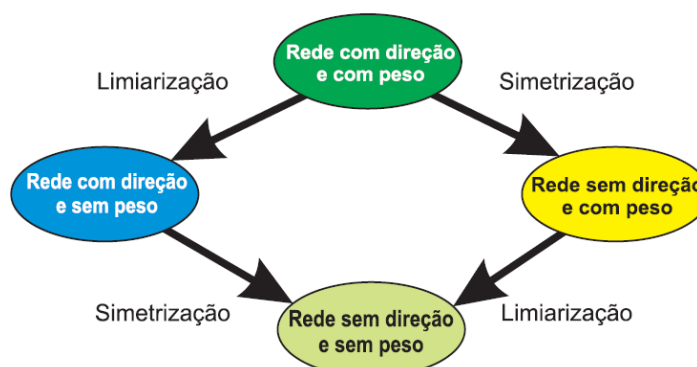
Figura 4 – Exemplo de um grafo desconexo com 3 componentes conectados.



Fonte: Sedgewick e Wayne (2011).

obter mais configurações com o uso das operações de limiarização e simetrização. Segundo Costa et al. (2007), na limiarização é feita a remoção de arestas, cujos valores de peso são menores a um determinado limiar (também pode ser aplicado a lógica inversa, considerando eliminação de valores maiores ao limiar). Depois, atribui-se um valor unitário às arestas que restaram. Por outro lado, a simetrização ocorre para transformação de arestas direcionadas em arestas não-direcionadas. A Figura 5 apresenta um esquema que relaciona essas operações.

Figura 5 – Exemplo de transformação dos tipos de redes quanto à natureza de suas arestas.



Fonte: Costa et al. (2007) apud Rodrigues (2007).

3.2.2 Modelo de redes complexas

Da interseção entre a teoria dos grafos e a mecânica estatística surgem as redes complexas, que podem modelar bem alguns dos problemas de processamento de imagem, devido a grande capacidade que elas têm de generalização e adaptabilidade (COSTA; ROCHA, 2006).

O termo rede complexa refere-se a um grafo que apresenta uma topografia não trivial. Algumas propriedades que têm recebido atenção na literatura são: coeficiente de

aglomeração; distribuição de graus de vértice; resistência da rede quanto às remoções de alguns vértices, sem que haja perda de sua funcionalidade; variedade e mistura de padrões (METZ et al., 2007).

Para estudar as propriedades das redes reais em termos de topologia, vários modelos de redes foram definidos. Percebeu-se que as redes complexas poderiam ser classificadas quanto às propriedades estatísticas em relação ao grau de conexão² ou ao coeficiente de agrupamento³ (COSTA et al., 2007). Os modelos mais conhecidos e estudados são: rede regular, rede aleatória, rede de pequeno mundo e rede livre de escala. Além desses, existem outros modelos que se adaptam bem a problemas específicos, como redes com estrutura de comunidade. A seguir, serão apresentados alguns modelos, dos quais foram utilizadas características que mais se adaptaram ao propósito desta dissertação e outros modelos que serão apresentados apenas para auxílio à compreensão de conceitos importantes acerca de medidas de redes.

Em contraste aos modelos de redes complexas, há o modelo de **redes regulares**, que possui todos os nós com o mesmo grau de conectividade. Tal modelo pode ser aplicado ao estudo de modelos atômicos para área da física.

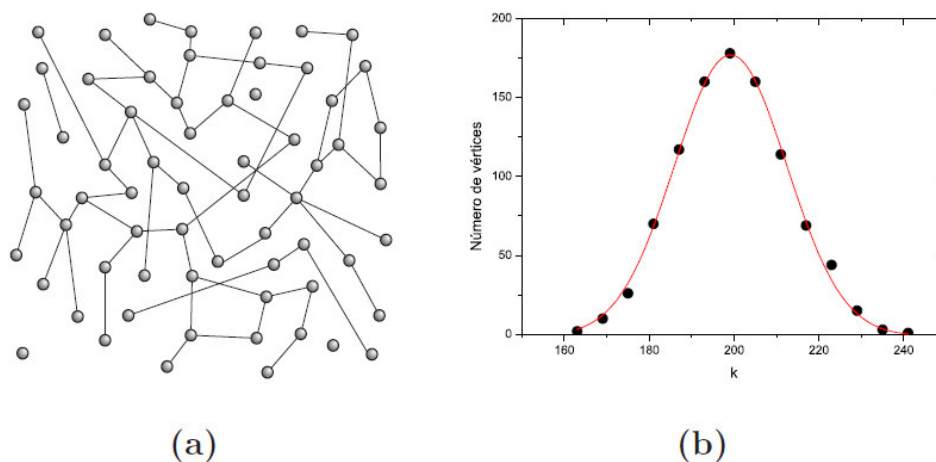
Como mais simples modelo de rede complexa, existe o **grafo aleatório** definido por Erdős e Rényi (1959). Esse modelo consiste em N vértices e E arestas: começando com todos os N nós desconectados. A rede é construída pelo estabelecimento de E arestas ao acaso, (desconsiderando múltiplas arestas para o mesmo par de nós, ou auto-ligações). A Figura 6 (a) mostra um exemplo desse tipo de rede, sendo que para este modelo, $P(k)$ (grau de distribuição) obedece a uma distribuição de Poisson.

Outro modelo de rede bastante conhecido é o de ***small world network*** ou **rede de pequeno mundo**. Proposto por Watts e Strogatz (1998) é um tipo de modelo matemático de grafo, em que a maior parte das conexões são estabelecidas entre os vértices mais próximos, semelhante às redes regulares, contudo possui um grau de aleatoriedade, por possuir algumas ligações aleatórias com pares de nós distantes. Dessa forma, o grafo se comporta como um "mundo-pequeno", pois diversas situações do mundo real podem ser mapeadas por esse modelo (redes de estradas, redes de infraestrutura elétrica, redes de telecomunicação, redes de neurônios e sinapses, rede social, etc.).

² O *grau de conexão* de um nó se refere à quantidade de conexões que ele possui. Tal medida será detalhada na Subseção 3.2.3.1.

³ O *coeficiente de agrupamento* mede o agrupamento de um subgrafo formado por um nó e a sua vizinhança. Esta medida será melhor definida na Subseção 3.2.3.3.

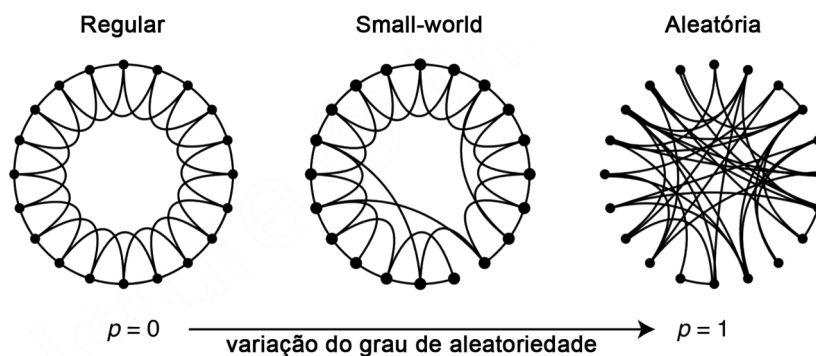
Figura 6 – (a) Um grafo aleatório, modelo de Erdős e Rényi (1959). (b) A distribuição de graus para uma rede com 10.000 nós, usando uma probabilidade $p=0,2$. Cada ponto no gráfico corresponde a média de 10 redes.



Fonte: Adaptado de Rodrigues et al. (2007)

Para ser obtida uma rede aos moldes do modelo *small world*, toma-se como exemplo uma rede regular, de topologia circular, contendo N nós conectados a k vizinhos mais próximos para as duas direções, obtendo-se $2k$ arestas iniciais (considerando que $N \gg k \gg \log(N) \gg 1$). A partir daí, cada aresta é reconectada de forma aleatória com uma probabilidade fixa p . Tudo isso para acrescentar um aspecto aleatório à rede. Se $p=0$ a rede se mantém regular, apresentando alta quantidade de ciclos e diâmetro alto, e se $p=1$, a rede é aleatória, apresentando baixa quantidade de triângulos (ciclos de ordem três), mas contendo diâmetro de grafo. Com p entre 0 e 1, a rede *small world* está no meio termo entre a ser rede regular e ser aleatória (COSTA et al., 2007). As Figuras 7 e 8 ajudam na compreensão desse exemplo.

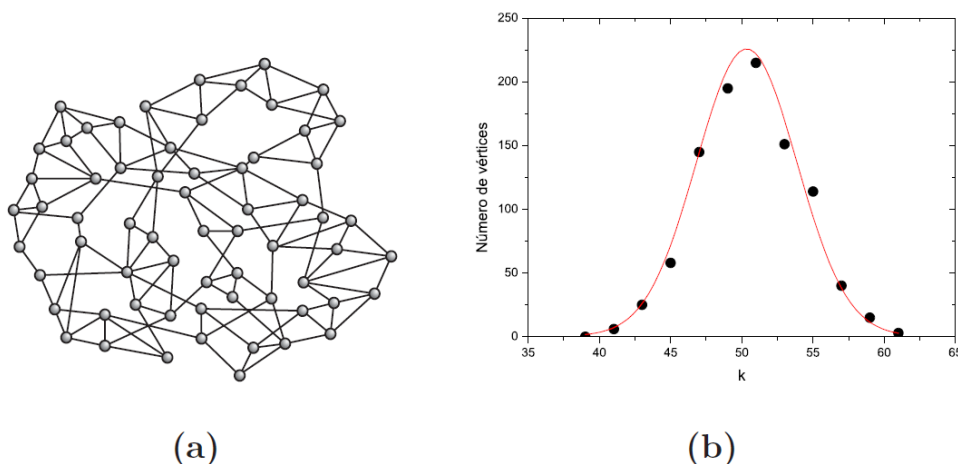
Figura 7 – A construção de uma rede de pequeno mundo de acordo com Watts and Strogatz (1998). Para cada aresta é aplicada uma religação aleatória com probabilidade p .



Fonte: adaptado de Watts e Strogatz (1998).

Uma rede *small world* é caracterizada por possuir a distância média entre dois

Figura 8 – O modelo de rede *small world* de Watts e Strogatz: (a) um exemplo de rede com $N=64$ nós, $k=2$, $p=0,1$. (b) Uma distribuição da graus para uma redes com 10.000 nós, usando uma probabilidade $p=0,2$. Cada ponto no gráfico corresponde a média de 10 redes.



Fonte: Adaptado de (RODRIGUES, 2007)

vértices quaisquer não superior a um número pequeno de vértices. Esse tipo de rede é construído com base no modelo clássico de Erdős-Rényi, apresentando um diâmetro de rede pequeno e um agrupamentos local. Ou seja, há claramente conjunto de nós com laços mais fortes, que constituem pequenas comunidades na rede. Tal fato é validado pelo coeficiente de agrupamento.

Em (BACKES; CASANOVA; BRUNO, 2009) é aplicada a técnica de limiarização ou Evolução Dinâmica de Rede, em que a partir da variação de um Limiar L , é produzido um novo conjunto de arestas para a rede. Tal estratégia visa ajustar a rede regular em um modelo de rede de mundo-pequeno (WATTS; STROGATZ, 1998), a fim de analisar o coeficiente de aglomeração. Já na abordagem deste presente trabalho deseja-se utilizar a Evolução Dinâmica da Rede para criar variações de um mesmo grafo através da eliminação de arestas de alto peso, delimitadas por cada valor de limiar L .

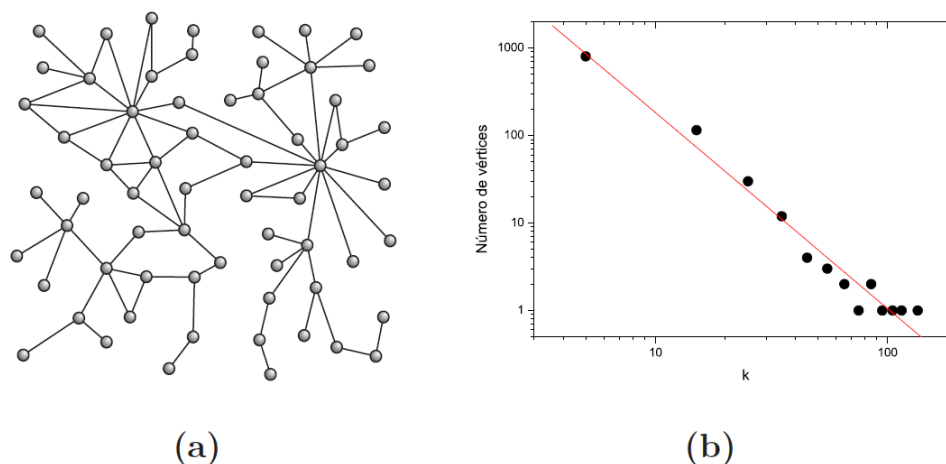
O último tipo de modelo de rede aqui analisado são as **redes livre de escala** ou redes sem escala⁴, que são redes complexas que seguem a *lei de potência*⁵ para o grau de distribuição de conexão. A maior parte dos nós possui poucas arestas, e existem outros poucos nós que têm um número alto de ligações. Isto é, existe uma alta probabilidade de

⁴ *Redes livre de escala* recebem esse nome devido a característica de invariância de escala, onde objetos ou leis não mudam quando suas variáveis são multiplicadas por um fator comum, representando assim uma universalidade (GUERRIERO, 2012).

⁵ Para a física, ocorre *lei de potência* quando a relação entre dois escalares x e y pode ser descrita na forma $y = ax^k$, onde a e k são constantes. Esta lei pode ser expressa por equação linear (linha reta) em um gráfico com eixos log-log, pois a equação anterior pode ser escrita como $\log(y) = k \log(x) + \log(a)$ (GUERRIERO, 2012).

um nó se ligar a outro de mesmo grau. Por isso, essas redes possuem nós denominados de *hubs*, que são os de alta conectividade. De forma geral, essas redes têm mais resistência a falhas por remoção de ligações ou nós, mas são suscetíveis a ataques em conjunto.

Figura 9 – (a) Exemplo de uma rede livre de escala (modelo de Barabási e Albert).(b) Distribuição dos graus de conexão para uma rede livre de escala. A distribuição segue uma lei de potência. Eixos em escala logarítmica. Gráfico referente a média de 10 redes.



Fonte: Adaptado de Rodrigues et al. (2007)

Um ponto importante no estudo de Barabási (1999), além de as redes complexas possuírem uma certa dinâmica de estruturação, possuíam também um grau de aleatoriedade influenciado pelo grau de influência ou popularidade que os nós possuíam. Essa regra de estruturação ficou marcada pela expressão do autor: "o rico fica cada vez mais rico".

3.2.3 Medidas de Redes Complexas

Em diversos trabalhos⁶ encontram-se dezenas de medidas capazes de expressar as características de topologia de uma rede complexa. A partir desse ponto, escolher as medidas para análise é primordial, pois alguns conjuntos de medidas podem ser insuficientes, ou mesmo se apresentarem como redundantes, levando a classificações ou detecção errôneas. Além do mais, as medidas devem permitir a representação topológica e funcional mais completa possível.

Percebe-se que há um consenso entre os principais trabalhos no sentido de analisar, não somente, uma característica isolada em cada problema, mas um conjunto de medidas,

⁶ Alguns dos mais relevantes trabalhos sobre redes complexas: (BACKES; CASANOVA; BRUNO, 2009; CHALUMEAU et al., 2008; JORGE-HERNANDEZ et al., 2014; STAM et al., 2006; COSTA; ROCHA, 2006; COSTA et al., 2007; RODRIGUES, 2007)

para assim, analisar e classificar redes. Por isso, em alguns desses trabalhos, são abordados os tipos de redes que podem servir como melhores modelos para cada categoria de problema, e de forma mais específica são listadas medidas que melhor se adaptam a cada classe de problema.

Sobretudo, para esta dissertação, buscou-se medidas que levassem em conta características relacionadas a conectividade, correlação, ciclos e centralidade. Nas subseções seguintes estão detalhadas as medidas de grafo que se mostraram mais relevantes para o presente trabalho. Para facilitar as demonstrações, as fórmulas estão definidas considerando a estruturação de grafo em termos de matriz de adjacência⁷.

3.2.3.1 Grau

O **Grau** k_i de um nó i ou **conectividade** é o número de arestas conectadas a este nó e que é definida em termos de matriz de adjacência A , como:

$$k_i = \sum_{j=1}^N a_{ij}, \quad (3.1)$$

onde N é a quantidade de nós no grafo.

Grau é uma das características mais importantes de um nó. Medidas de todo o grafo podem ser extraídas por meio dos graus de todos os nós. Provenientes do grau de um nó, duas simples medidas são o **Grau Máximo** (Eq. 3.2) e o **Grau Médio** (Eq. 3.3), apresentadas a seguir:

$$k_{max} = \max(k_i) \quad (3.2)$$

$$k_{\mu} = \frac{1}{N} \sum_{i=1}^N k_i \quad (3.3)$$

É possível caracterizar diversos tipos de redes através da medida de conectividade e da sua distribuição. Além de determinar se a rede possui conexões que se configuram de forma aleatória ou por meio de alguma lei de formação. Tais medidas, por muito tempo foram usadas como as medidas mais importantes para estudo de redes complexas (COSTA et al., 2007).

⁷ A *matriz de adjacência* registra as ligações da rede. Os elementos a_{ij} na matriz serão igual a 1 se existir uma conexão entre i e j , caso contrário serão 0.

3.2.3.2 Ordem, tamanho e densidade de grafo

Ordem de um grafo, também chamada de cardinalidade do conjunto V , representa a sua quantidade de vértices, denotada por $|V|$ ou pela letra N . O **tamanho de um grafo**, também chamada de cardinalidade do conjunto E , representa a sua quantidade de arestas, denotado por $|E|$ ou pela letra M . A **densidade de um grafo** é definida por meio da razão entre quantidade de arestas e vértices, ou seja, $Dens = |E|/|V|$. Então, um grafo é dito denso quando possui grande número de arestas para pouca quantidade de vértices. Ao contrário, o grafo é chamado de esparso (SEDGWICK; WAYNE, 2011).

3.2.3.3 Coeficiente de agrupamento

O coeficiente de agrupamento (ou coeficiente de clusterização) é considerado uma medida da conectividade local. É uma medida do subgrafo formado por um nó e a sua vizinhança. Essa medida para um nó da rede é dado por:

$$C_i = \frac{N_{\Delta}(i)}{N_3(i)}, \quad (3.4)$$

onde $N_{\Delta}(i)$ é o número de triângulos envolvendo o nó i , e $N_3(i)$ é o número de trios conectados tendo i como nó central:

Considerando que a_{ij} indica, através de uma matriz de adjacência, que há ligação entre nó i e seu nó vizinho j , então o valor de a_{ij} será 1 se houver aresta entre os nós i e j . $N_{\Delta}(i)$ é um triângulo se houver ligação entre cada par formado por i , j e m , ou seja:

$$N_{\Delta}(i) = \sum_{m \neq j \neq i} a_{ij} a_{im} a_{jm} \quad (3.5)$$

Já o número de trios conectados $N_3(i)$ para um grafo é definido por:

$$N_3(i) = \sum_{m \neq j \neq i} a_{ij} a_{im} \quad (3.6)$$

Considerando que k_i é o grau de conectividade do nó i (o número de vizinhos), o número

de trios N_3 também pode ser definido como: $N_3(i) = k_i(k_i - 1)/2$. Nesse caso, o coeficiente de agrupamento tem sua fórmula adaptada para:

$$C_i = \frac{l_i}{\frac{k_i(k_i-1)}{2}} \quad (3.7)$$

, onde $N_\Delta(i)$ conta o número de arestas entre os vizinhos de i , simbolizado por l_i . Assim, o **Coefficiente de Agrupamento de Rede** é definido como:

$$C_c = \frac{1}{N} \sum C_i. \quad (3.8)$$

É importante ressaltar que um alto coeficiente de agrupamento está associado à robustez de uma rede, que é a resiliência contra danos aleatórios na rede representada pelo grafo. Aplicado a processamento de imagem, representa o quão uniforme são as regiões da imagem. Um alto valor de agrupamento indica que as conexões próximas entre as regiões possuem aspectos uniformes, mas também pode indicar um problema de conectividade com regiões distantes. Se um determinado grafo possui um índice de agrupamento alto em determinado ponto da rede, significa que aquela região possui uma propriedade de ligação valorizada em detrimento das outras regiões.

O coeficiente de agrupamento varia significativamente em redes do tipo comunidade, sendo maior nos vértices que integram as comunidades e menor nos vértices que interligam comunidades (*hubs*). Isto significa que apresentam subgrafos (comunidades) com grau alto e, ao mesmo tempo, poucas arestas entre o subgrafo e o restante da rede.

Para redes ponderadas, as definições de grau dadas acima podem ser usadas considerando o peso. Nesse caso o grau de um nó i é chamado de **força de conexão de nó** (*strength*) s_i , definida como a soma dos pesos das arestas de i :

$$s_i = \sum_j w_{ij}. \quad (3.9)$$

Caso seja considerado o peso no cálculo do coeficiente de agrupamento, deve considerar a seguinte formulação:

$$C_i^w = \frac{1}{s_i(k_i - 1)} \sum_{m \neq j} \frac{w_{ij} + w_{im}}{2} a_{ij} a_{im} a_{jm}, \quad (3.10)$$

onde o fator de normalização da força do vértice $s_i(k_i - 1)$ assegura que $0 < C_i^w < 1$ (COSTA et al., 2007). A partir dessa equação, uma das definições de **coeficiente de agrupamento ponderado** é:

$$C^w = \frac{1}{N} \sum C_i^w. \quad (3.11)$$

3.2.3.4 Entropia e energia

Nas propriedades estruturais das redes, o tipo de correlação mais usada é por meio de dois nós conectados por uma aresta. Esta correlação pode ser expressa pela probabilidade de um nó i ter um vizinho com o mesmo grau, ou seja, $P(k_i, k')$, sendo $k_i = k'$. Para analisar as probabilidades de distribuição de graus foram escolhidas Entropia e Energia.

Entropia é frequentemente associada como grau de desordem. Para análise de redes, em termos práticos, é montado um vetor de frequência de graus de nós. Esse vetor é normalizado para se obter a probabilidade P de ligação para cada grau. Então é aplicada a Equação 3.12.

$$H = - \sum_{i=1}^N P(k_i, k')_i \log_2(P(k_i, k')) \quad (3.12)$$

Para o cálculo da **Energia** de uma rede, é aplicada uma pequena variação da equação anterior para obtenção da Equação 3.13. Assim, a energia é definida a partir da soma a probabilidade P de ligação para cada grau, elevado ao quadrado (GUTMAN; ZHOU, 2006).

$$E = \sum_{i=1}^N (P(k_i, k')_i)^2 \quad (3.13)$$

3.2.3.5 Coeficiente de Pearson

Uma forma de descrever a probabilidade de um vértice qualquer ter determinado grau de conectividade k é através da distribuição dos graus $P(k)$, expresso pela fração de vértices em uma rede com grau k . A Figura 6 (b) apresenta a distribuição de probabilidades para redes de modelo aleatório. Ou seja, se a rede possui uma distribuição de Poisson,

essa pode ser associada ao modelo Erdős-Rényi (SEDGWICK; WAYNE, 2011).

O coeficiente de Pearson pode ser aplicado para se obter um valor que representa a correlação entre a probabilidade $P(k)$ e a conectividade k na escala logarítmica, escala que descreve a retidão r da distribuição. A correlação de Pearson, de forma genérica, é obtida por meio da fórmula:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \langle x \rangle)(y_i - \langle y \rangle)}{\sqrt{\sum_{i=1}^n (x_i - \langle x \rangle)^2} \sqrt{\sum_{i=1}^n (y_i - \langle y \rangle)^2}}, \quad (3.14)$$

onde os valores do i -ésimo elemento dos vetores x e y são representados por x_i e y_i , de tamanho n , e $\langle x \rangle$ e $\langle y \rangle$ representa as médias de x e y . Para se obter a retidão de distribuição para redes, basta substituir x pelo logaritmo de k (conectividade) e substituir y pelo logaritmo $P(k)$ (probabilidade de k) (RODRIGUES, 2007).

O valor r_{xy} varia de -1 a 1. Se $r = -1$, é indicação de que a rede possui distribuição de graus como uma rede livre de escala (o gráfico de distribuição é uma reta decrescente, conforme Figura 9 (b)), pois quando duas distribuições são completamente anti-correlacionadas, o Coeficiente de Pearson tem valor mínimo. Esse estudo sobre as conexões pode determinar as correlações entre os nós, pois quando a probabilidade de dois nós se ligarem for independente do grau de cada nó, apresenta-se um caso de não-correlação (NEWMAN, 2003).

Relacionado com a noção de grau e conectividade de rede há o conceito de mistura ou **assortatividade** (*assortivity*). Essa medida verifica o quanto os vértices com um alto grau tendem a ser conectados a outros vértices com um alto grau, e também o quanto vértices com um grau baixo tendem a se conectar a outros vértices de baixo grau (BONDY; MURTY et al., 1976). A assortatividade é determinada pelo coeficiente de Pearson com modificação da Equação 3.14, obtendo-se a seguinte equação:

$$r = \frac{\frac{1}{M} \sum_{j>i} k_i k_j a_{ij} - [\frac{1}{M} \sum_{j>i} \frac{1}{2}(k_i + k_j) a_{ij}]^2}{\frac{1}{M} \sum_{j>i} \frac{1}{2}(k_i^2 + k_j^2) a_{ij} - [\frac{1}{M} \sum_{j>i} \frac{1}{2}(k_i + k_j) a_{ij}]^2} \quad (3.15)$$

Nesta fórmula, M representa o número de arestas, i e j os nós participantes interligados pelas arestas, k_i é o grau do nó i , a_{ij} é o peso da aresta. Caso a assortatividade seja positiva, os nós com graus de conexão similares tendem a se ligar e a rede é chamada

assortativa. Se a assortividade é negativa, nós com grau de conexão maior tendem a se ligar com os nós de grau menor, então a rede é denominada dissassortativa. E se essa medida for 0, a rede é dita não-relacionada, então não há qualquer correlação entre as conectividades (RODRIGUES, 2007).

3.2.3.6 Centralidade

As medidas de centralidade determinam a relevância relativa, que aplicado à contextos reais podem, por exemplo, definir qual é a pessoa mais influente dentro de uma rede social, ou o quanto é importante uma rota dentro de uma rede de caminhos. Diversos conceitos de centralidade foram desenvolvidos visando a análise de redes sociais e relações de ordem sociológica (NEWMAN, 2003). Das principais medidas para centralidade as mais utilizadas são: centralidade de grau, centralidade de proximidade, centralidade de intermediação e centralidade de vetor próprio. Essas medidas são detalhadas a seguir.

A **centralidade de grau** é definida como o número de ligações incidentes de um vértice. O grau pode ser interpretado como a probabilidade que o vértice tem de receber alguma informação da rede, no caso de uma rede orientada (onde os laços tem direção) (OPSAHL; AGNEESSENS; SKVORETZ, 2010).

A **centralidade de proximidade** é baseada, também, em caminhos mais curtos, ou seja, *um nó importante está próximo dos outros*. O afastamento de um nó é calculado pela soma de suas distâncias para todos os nós da rede. E a proximidade é definida como o inverso do afastamento. Dessa forma, quanto menor é a distância total para todos os outros nós, mais central é o nó. Com isso, proximidade pode ser considerada uma medida de rapidez, para determinar a velocidade que ela necessitará para difundir informações de um nó a todos os outros nós sequencialmente, por isso, é também conhecida por centralidade de informação (NEWMAN, 2005).

Como outra medida de centralidade, a **centralidade de intermediação** (*betweenness centrality*) é baseada em caminhos mais curtos e é usada para medir o fluxo que passa por uma dada aresta ou nó (FREEMAN, 1977). Se a rede for analisada levando em conta as relações de transporte ou comunicação, algumas arestas ou nós podem receber um tráfego mais intenso do que outros. Este fato significa que tais elementos são os mais importantes e estão situados na maioria dos menores caminhos, ou seja, *um nó importante faz parte de muitos caminhos* (BRANDES, 2001). Essa medida é calculada da seguinte

maneira:

$$B_u = \sum_{ij} \frac{\sigma(i, u, j)}{\sigma(i, j)}, \quad (3.16)$$

onde $\sigma(i, u, j)$ representa a quantidade de menores caminhos entre os nós i e j que passam pelo nó u , e $\sigma(i, j)$ é o total de menores caminhos entre i e j . A soma é feita sobre todos os pares distintos i e j de nós. A média do grau de intermediação pode ser utilizada como uma medida de caracterização global da rede:

$$\langle B \rangle = \frac{1}{N} \sum_i^N B_i. \quad (3.17)$$

Com o grau de intermediação, é possível obter a dominância do ponto central, que funciona como uma medida global, que é definida pela equação:

$$c_D = \frac{1}{N-1} \sum_i^N (B_{max} - B_i), \quad (3.18)$$

onde B_{max} se refere ao maior valor da centralidade da rede. Quando a rede está completamente conectada a dominância do ponto central será 0. Quando a rede está com um formato semelhante a uma topologia de estrela (todos os nós só se conectam ao nó central), a dominância do ponto central assume o valor 1 (RODRIGUES, 2007).

Centralidade de autovetor (*eigenvector centrality*), mede a importância de um nó numa rede em função da influência de seus vizinhos, se baseando na ideia de que *um nó importante tem vizinhos importantes*. Com essa medida, atribui-se pontuações a todos os nós da rede, considerando que um determinado nó de alta pontuação contribui muito para aumentar a pontuação de seus nós vizinhos, e um nó de baixa pontuação contribui pouco para valorizar na pontuação de seus vizinhos. Como se baseia em autovalores e autovetores de matrizes simétricas, isso implica em dizer que, apesar de que um nó tenha baixa centralidade de grau, (ou seja, está conectado a apenas alguns poucos nós da rede), se os nós vizinhos forem importantes, por conseguinte, o nó em questão será importante, possuindo uma alta centralidade de autovetor (SABIDUSSI, 1966).

À luz da álgebra linear, temos a definição de autovalores (ou valor característico) e autovetores (ou vetor característico). Um escalar λ é chamado de autovalor de A , se

houver um vetor X diferente de zero, tal que:

$$AX = \lambda X \quad (3.19)$$

Nesse caso, o vetor X é chamado de autovetor de A correspondente a λ (KOSCHÜTZKI et al., 2005).

Como exemplo, temos que a matriz $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ é autovetor de $A = \begin{bmatrix} 3 & 2 \\ 3 & -2 \end{bmatrix}$ para $\lambda = 4$, pois:

$$\begin{bmatrix} 3 & 2 \\ 3 & -2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 4 \begin{bmatrix} 2 \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 3 \cdot 2 + 2 \cdot 1 \\ 3 \cdot 2 + (-2) \cdot 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix} \Rightarrow \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix} \quad (3.20)$$

Sendo um grafo $G = (V, E)$ com $|V|$ vértices, e considerando que $A = (a_{v,t})$ seja a matriz de adjacência (ou seja, $a_{v,t} = 1$ se vértice v é ligado ao vértice t , e $a_{v,t} = 0$ se não houver ligação entre v e t), a pontuação de centralidade relativa de um vértice v pode ser definida como um rearranjo da Equação 3.19, nos seguintes termos:

$$X_v = \frac{1}{\lambda} \sum_{t \in M(v)} X_t \equiv \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t, \quad (3.21)$$

onde $M(v)$ é o conjunto de vizinhos de v e λ é uma constante calculada através do módulo do vetor X , ou seja $\|X\|$.

Por métodos tradicionais, o valor de λ seria calculado de maneira muito onerosa, envolvendo adaptações com matriz identidade e cálculo de determinantes, o que iria resultar em equações de grau altíssimo, além de precisar da resolução de sistemas de equação⁸. Há muitos autovalores diferentes λ para os quais existe uma solução de autovetor diferente de zero. No entanto, o requisito adicional de que todas as entradas no autovetor sejam não-negativas implica (pelo teorema de Perron-Frobenius) que apenas o maior autovalor resulta na medida de centralidade desejada, então, nesse caso, aplica-se o algoritmo da iteração de energia de autovalor para encontrar esse autovetor dominante (MATCOVSCHI; PASTRAVANU, 2007).

⁸ Um grafo de n nós produziria para a fórmula 3.19, um sistema de equações com n variáveis.

O autovetor é definido apenas até um fator comum, portanto, apenas as proporções das centralidades dos vértices são bem definidas. Para definir uma pontuação absoluta, deve-se normalizar o autovetor. A iteração de energia é um dos muitos algoritmos de autovalor que podem ser usados para encontrar esse autovetor dominante. Tal artifício é usado da seguinte maneira: tornando $\lambda = \|AX'\|$ e substituindo na Equação 3.19, de forma adaptada temos: $\frac{AX'}{\|AX'\|} = X$, logo o algoritmo se executa de acordo com os seguintes passos:

1. X' recebe inicialmente o valor de um vetor coluna com $|V|$ elementos com valor 1.
2. Enquanto não extrapolar o número máximo de iterações OU até X convergir:
 - a) X recebe valor de $\frac{AX'}{\|AX'\|}$.
 - b) X' recebe valor de X .
 - c) (refaça o passo 2).

Ao final se tem um vetor X que possui um valor de centralidade para cada respectivo nó, o que representa o quão influente é cada nó em relação à medida de centralidade de autovetor. Outras medidas que se assemelham à medida de centralidade de autovetor são a centralidade de Katz e algoritmo de PageRank (no sistema de busca do Google) (AUSTIN, 2006).

A Figura 10 fornece uma noção visual das diversas variações de medidas de centralidade.

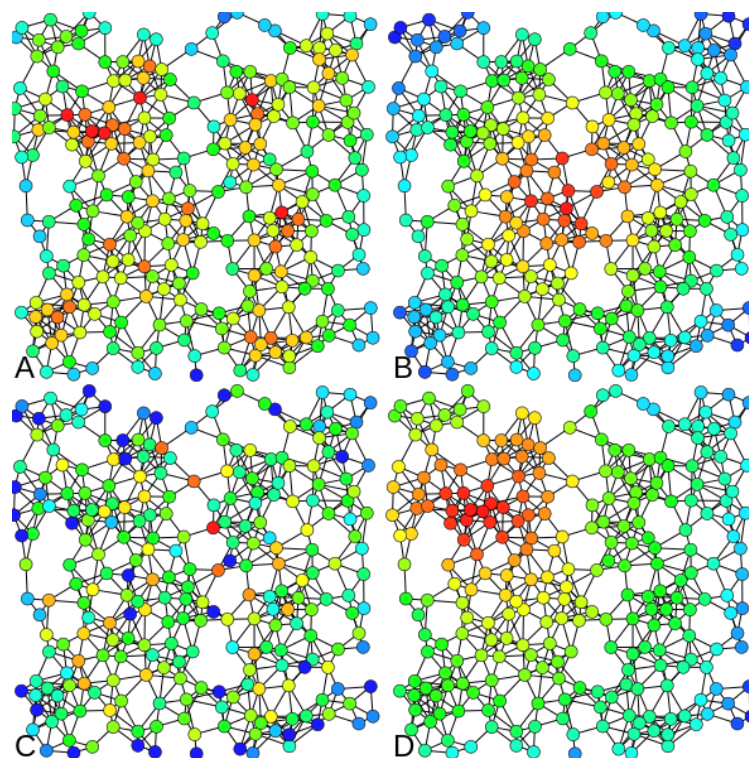
3.2.3.7 Cliques

Um clique é um subconjunto de nós totalmente conectados, dado um conjunto de nós. Em outras palavras, um conjunto de vértices é um clique se: para todo par v, u de vértices distintos neste conjunto, existe uma aresta entre v e u . Também são considerados os cliques de 1-vértice, que são compostos pelos seus próprios vértices individualmente.

Sobre cliques advém dois conceitos importantes: clique maximal e clique máximo. Um clique é dito maximal se não pode ser estendido, adicionando um novo nó adjacente. E um clique máximo é o maior clique possível dentro de um grafo (problema NP-completo).

Como medidas de clique utilizadas nesta dissertação, foram escolhidas o **número de cliques maximais** e o **tamanho do clique máximo**. Para diminuir a complexidade

Figura 10 – Exemplos de tipos de centralidade para um mesmo grafo: (a) Centralidade de Grau, (b) Centralidade de Proximidade, (c) Centralidade de Intermediação, (d) Centralidade de autovetor. Os pontos vermelhos possuem maior pontuação de centralidade.



Fonte: adaptado de (ROCCHINI, 2017)

de cálculo, ambas medidas foram aplicadas a subgrafos provenientes de comunidades de um grafo geral .

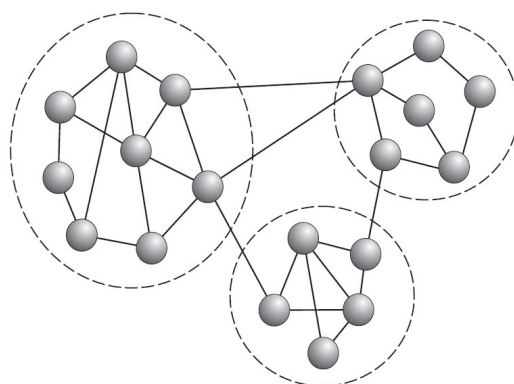
3.2.3.8 Comunidades

Grande parte das redes complexas possui estrutura modular, onde, as conexões são densamente distribuídas entre nós que pertençam a um mesmo grupo e esparsamente entre os nós de grupos distintos como exemplificado na Figura 11. Os módulos são formados por nós que possuem alguma relação de similaridade. Esses módulos foram conceituados por DANON et al. (2005) como comunidades, que são definidas por nós altamente conectados entre si e pouco conectados com o restante da rede. Contudo, o processo de reconhecimento das comunidades é um problema NP-completo, de difícil solução, o que gera uma desvantagem, pois, em geral não há uma estimativa sobre a quantidade de comunidades que constituem a rede.

O método mais popular usado no reconhecimento de comunidades é baseada no grau de intermediação (NEWMAN, 2003). Neste método, são removidas as arestas de

maior grau de intermediação, ou seja, aqueles que estão entre as comunidades. A cada iteração, o processo é refeito até que a rede se divida em nós isolados. No momento em que houver um maior valor de modularidade, então é registrada a divisão que possui as maiores comunidades. Esse método possui tempo de computação de ordem $O(MN^2)$, já que o cálculo do caminho mínimo entre um par qualquer de nós (que geralmente é feito com busca em largura) tem complexidade de ordem $O(M)$. Assim, como existem $O(N^2)$ pares de nós, a complexidade total, para o cálculo da centralidade de intermediação, está na ordem de $O(MN^2)$.

Figura 11 – Exemplo de uma rede com comunidades destacadas.



Fonte: Rodrigues et al. (2007)

3.2.4 Representação de imagem por modelagem baseada em grafo

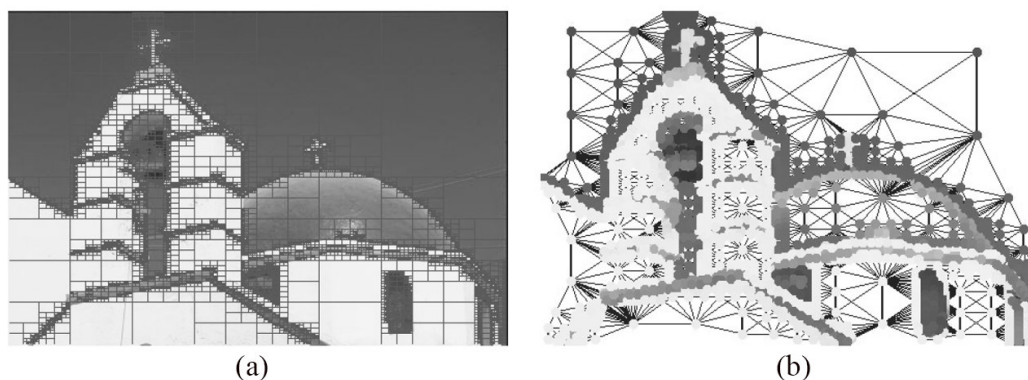
Tesselação⁹ é o nome dado às técnicas usadas a fim de modelar imagens em grafos (LÉZORAY; GRADY et al., 2012). Existe a **tesselação regular** (*regular image tessellation*), que divide a imagem em áreas de tamanhos iguais (ou em pixel) e transforma cada região como um nó de grafo. Em oposição a essa, há a **tesselação irregular**, que é o nome dado à técnica de mesclar regiões de pixels adjacentes usando um algoritmo de segmentação e etiquetagem de imagem. Contudo, o grafo final gerado não tem mais a garantia de ser regular devido à variação do espaço da imagem e do padrão de mistura dos pixels subjacentes.

O trabalho de Montanvert, Meer e Rosenfield (1991) foi um dos primeiros a propor uma nova técnica de análise de imagens baseada em hierarquias de tesselações

⁹ Tesselação (do inglês *tesselation*, pelo latim *tessella*) é o cobrimento de uma superfície bidimensional, usando polígonos, que indica a pavimentação de uma região através de peças de mosaico. Foi utilizada a palavra tesselação pois, em português, tecelagem (de tecer – entrelaçar fios) não teria o mesmo sentido.

irregulares. Com essa técnica foi possível analisar componentes conectados e obter o grafo de adjacência das classes segmentadas. Em outro trabalho, Samet (1984) utilizou a tesselação de *quadtree*¹⁰, cujo resultado e modelagem em grafo podem ser visualizados na Figura 12.

Figura 12 – (a) Uma imagem real com a tesselação *quadtree*. (b) O grafo de adjacência da região associado à partição *quadtree*.



Fonte: Adaptado de Lézoray et al. (2012).

Uma tesselação de forma genérica resulta em decomposições de componentes que podem ser chamados de superpixels. Dado um grafo, onde cada nó é identificado com um pixel, uma partição em regiões conectadas pode ser identificada como um novo grafo, onde cada partição de V é identificada com um nó. Cada V_i contíguo é chamado de superpixel. Superpixels são uma tendência cada vez mais popular em visão computacional e processamento de imagens.

Os algoritmos baseados em grafos podem ser usados para operar eficientemente nessas estruturas irregulares, em oposição aos algoritmos tradicionais de processamento de imagens que exigiam que cada elemento fosse organizado em uma grade. De fato, modelos baseados em grafos podem fornecer radicais aumentos de velocidade ao tornar grafos baseados em pixels para gráficos baseados em superpixel, por causa da drástica redução no número de nós. Os algoritmos mais comuns para geração de superpixels são: *Normalized Cuts* (Jianbo Shi; Malik, 2000), *Watershed* (Vincent; Soille, 1991), *Turbo Pixels* (Levinshtein et al., 2009) e SLIC (ACHANTA et al., 2011). Um exemplo de aplicação do *Watershed* e do SLIC apresentados na Figura 13 (b).

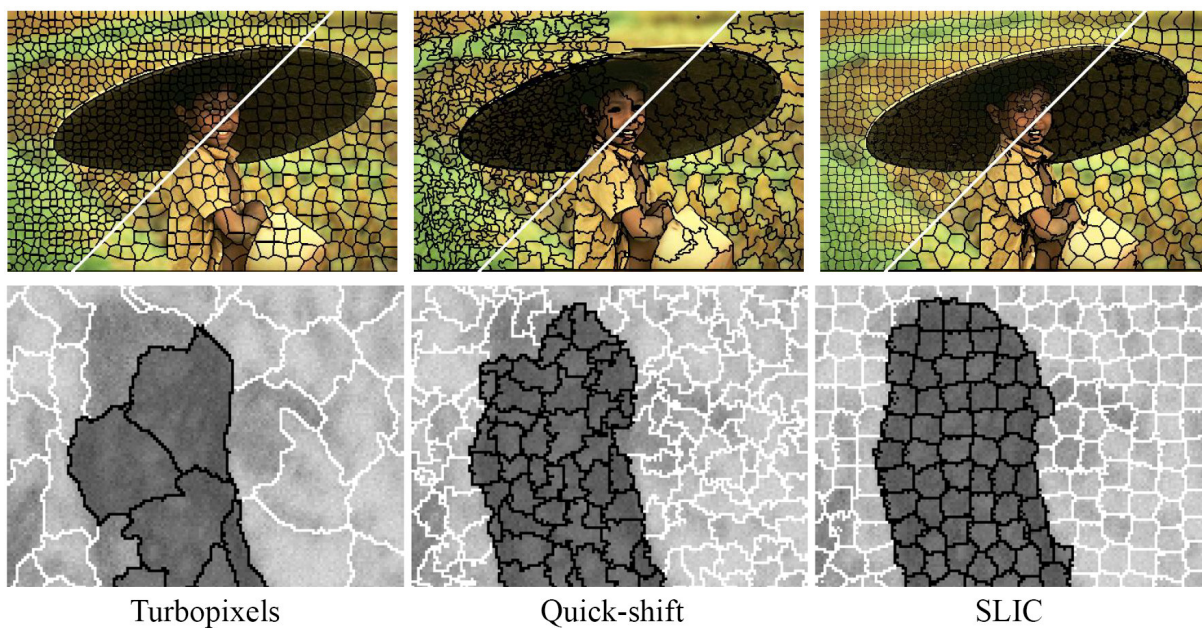
¹⁰ Uma região *quadtree* é uma representação de imagem hierárquica derivada da subdivisão recursiva do espaço 2D em quatro quadrantes de tamanho igual até que cada quadrado contenha uma região homogênea (com base nas propriedades de aparência de cada região da grade retangular subjacente) ou contenha no máximo um pixel. Uma região de tesselação *quadtree* pode ser facilmente representada por uma árvore onde cada nó da árvore corresponde a um quadrado.

Figura 13 – Duas diferentes tesselações irregulares de uma imagem usando métodos de segmentação de super-pixel: (a) Watershed (Vincent; Soille, 1991) e (b) SLIC (ACHANTA et al., 2011).



Fonte: Adaptado de Lezoray et al. (2012).

Figura 14 – Comparação de técnicas de geração de superpixels. Turbopixels (LEVINSHTEIN et al., 2009), Quick-shift (VEDALDI; SOATTO, 2008) e SLIC (ACHANTA et al., 2011).



Fonte: Adaptado de Achanta (2011).

Como parte da metodologia, apresentada na seção subsequente, foi escolhido o algoritmo *Simple Linear Iterative Clustering* (SLIC) (ACHANTA et al., 2011), como etapa de segmentação, visando a tesselação. Este algoritmo para segmentar superpixels se mostrou mais vantajoso para ser integrado à metodologia deste trabalho, tanto por oferecer melhor performance computacional, como por possuir parâmetros de ajuste (em relação a quantidade e formas dos superpixel), possibilitando flexíveis adaptações.

A Figura 14 mostra a comparação entre SLIC com dois dos melhores métodos que competem para o mesmo objetivo, Turbopixels (LEVINSHTEIN et al., 2009) e Quick-shift

(VEDALDI; SOATTO, 2008). O SLIC se mostrou superior a todos os métodos¹¹ existentes, de acordo com os critérios: controle de parâmetros, acurácia de segmentação, complexidade computacional e rapidez de execução.

O SLIC executa um agrupamento local de pixels no espaço 5-D definido pelos valores L, a, b do espaço de cores CIELAB¹² e coordenadas x, y dos pixels. O SLIC recebe como parâmetro o número de sementes K e o grau de compactação m . O número de sementes corresponde a mesma quantidade de superpixels gerados ao final. Considerando que uma imagem tem N pixels, então cada superpixels terá aproximadamente N/K pixels. O tamanho do diâmetro aproximado de cada uma desses superpixels seria aproximadamente $S = \sqrt{N/K}$. Em relação à dimensão, os clusters são definidos por $C_k = [l_k, a_k, b_k, x_k, y_k]$, com $k = [1, K]$, sendo l_k, a_k, b_k, x_k e y_k as representações dos componentes respectivos l, a, b, x e y , contidos no espaço do cluster C_k . A extensão espacial de qualquer cluster é aproximadamente S^2 , inicialmente.

A medida de distância D_S é definida como segue:

$$\begin{aligned} d_{lab} &= \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2} \\ d_{xy} &= \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \\ D_S &= d_{lab} + \frac{m}{S} d_{xy}, \end{aligned} \tag{3.22}$$

onde D_S é a soma da distância de lab (d_{lab}) e a distância do plano xy (d_{xy}) multiplicada por m/S . A variável m permite controlar a compactação do superpixel. Quanto maior o valor de m , a proximidade espacial é valorizada, sendo mais compacto o cluster. O valor de m pode assumir o intervalo $[1, 20]$ e por padrão é 10. A seguir o algoritmo é esquematizado em passos simplificados:

1. Primeiramente os centros de clusters C_k são inicializados.

¹¹ Quick-shift não possui parâmetro de controle de compactação do superpixel e nem número de superpixels, enquanto Turbopixels e SLIC possuem. Apesar de Turbopixels e SLIC terem a mesma complexidade ($O(N)$) o SLIC executa 10 vezes mais rápido do que Turbopixels. Quick-shift possui complexidade N^2 . No requisito de reconhecimento de classe de objeto (índice de erro de segmentação), o SLIC alcançou acurácia de segmentação 76.9%, enquanto que os índices de Turbopixels e Quick-shift foram respectivamente de 75,1% e 62,0% (ACHANTA et al., 2011).

¹² A CIE (Commission Internationale de l'Eclairage), uma organização sem fins lucrativos, definiu os espaços de cores CIE $L^*a^*b^*$, para expressar a cor dos objetos. Neste espaço de cor, L indica luminosidade, a^* é coordenada vermelho/verde e b^* é a coordenada amarelo/azul, dentro da escala do espectrofotômetro circular (MINOLTA, 2006).

2. Repita até a estabilidade ou até um dado limite de iterações:

a) Para cada cluster C_k :

i. Encontre similares pixels numa vizinhança $2S \times 2S$ ao redor do centro do superpixel no plano xy , de acordo com a Equação 3.22.

b) Computa novos centros e o erro residual entre os novos centros e os anteriores.

Após essa explanação acerca da fundamentação teórica, no próximo capítulo, serão abordados os materiais e os métodos utilizados para o desenvolvimento deste trabalho.

4 MATERIAIS E MÉTODOS

Este capítulo contempla os procedimentos que compõe a metodologia desenvolvida para detecção das regiões de anomalia em termografia de mama. Inicialmente é apresentada a bases de termografias utilizada nos experimentos. Posteriormente, são descritas as etapas da metodologia contendo os métodos e seus detalhamentos.

4.1 Base de termografias

As termografias utilizadas nesses trabalho pertencem a DMR-TD, que possui 70 exames de termografia dinâmica, sendo 35 com presença confirmada de alguma anomalia e 35 saudáveis (SILVA et al., 2014b). Essa base é a que contém maior quantidade de termografias, possui a aquisição de boa resolução, tem um protocolo mais rígido e padronizado e é usada nos trabalhos que apresentaram melhores resultados na literatura. Por tais motivos foi eleita para ser usada neste trabalho.

As termografias dessa base foram adquiridas por um protocolo definido, com controle de temperatura do ambiente e monitoramento da temperatura do paciente. Foi usada a câmera FLIR, de modelo SC620, que possui sensibilidade de 0,04 °C e produz imagens com dimensões de 640 x 480 pixels. As termografias são salvas diretamente pela câmera em arquivo como matriz de temperatura, com precisão de três casas decimais. Esse armazenamento foi possível através do *software* TermoCad, desenvolvido com base no kit de desenvolvimento de *software* (SDK), fornecido pelo fabricante da câmera (BORCHARTT et al., 2013).

4.1.1 Seleção das termografias

A metodologia, aqui apresentada, depende diretamente da localização do tumor para que possa ser feita a avaliação da detecção. Contudo, a base não possui a localização do tumor definida para todos os exames. Por isso as termografias tiveram que passar por uma triagem, a fim de selecionar somente os exames que possuem a descrição da localização do tumor. Dos 35 exames diagnosticados como doentes, apenas 22 exames possuem alguma descrição da localização do tumor. Em relação à qualidade da descrição

da localização, os exames selecionados podem ser divididos da seguinte maneira:

- 16 exames possuem descrição completa da localização do tumor;
- 2 exames possuem descrição pouco precisa;
- 4 exames possuem descrição imprecisa.

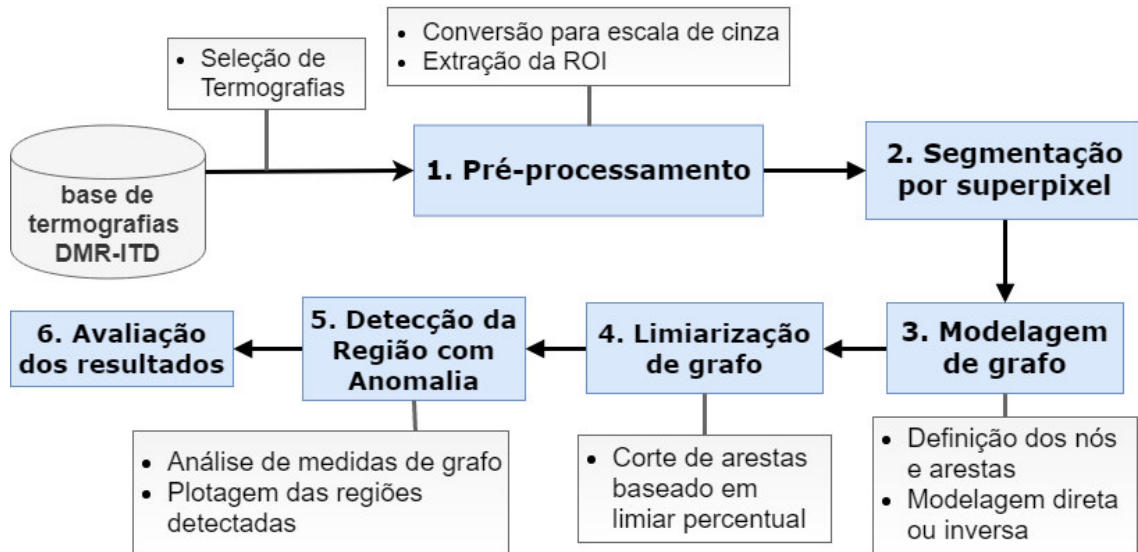
Nesta metodologia foi considerada uma descrição de boa qualidade quando há indicação do quadrante da região da anomalia (superior externo (QSE), superior interno (QSI), inferior externo (QIE) e inferior interno (QII)). No caso de descrição pouco precisa ou totalmente imprecisa, usou-se informações textuais extras que puderam ajudar na definição da localização. As características peculiares de cada tipo tumor (apresentadas na Subseção 3.1.3) também auxiliaram a suprir deficiência de descrição.

O Apêndice A contém uma tabela com descrição de localização da anomalia para cada exame selecionado. Como exemplo de descrições imprecisas, tem-se o Exame nº 250, que não possui o quadrante definido explicitamente, mas possui a descrição do resultado da biópsia: "Carcinoma lobular in situ. Próximo do mamilo D". Então, considerando essa descrição, deduz-se que o quadrante com anomalia é o Quadrante Central da Mama Direita. Como outro exemplo no Apêndice A, o exame nº 204 também não define o quadrante declaradamente, contudo apresenta a descrição: "Descarga papilar de líquido seroso em mama direita. Resultado da peça cirúrgica após retirada da mama". Nesse caso, conclui-se que a anomalia está localizada em toda a região que fazia parte da mama direita.

4.2 Metodologia proposta

A metodologia proposta consiste em 6 etapas consecutivas: (1) pré-processamento, (2) segmentação por superpixel, (3) modelagem de grafo, (4) limiarização de grafo, (5) detecção da região com anomalia e (6) avaliação do resultados. A sequência da realização dessas etapas é exposta na Figura 15. Cada etapa é descrita, respectivamente, nas subseções seguintes, juntamente com as técnicas utilizadas.

Figura 15 – Etapas da Metodologia proposta



4.2.1 Pré-processamento

Na etapa de pré-processamento, fez-se necessária a preparação das imagens. Para isso, foi aplicado um método de conversão das matrizes de temperatura em imagens e um método para extração da região de interesse, subetapas que são detalhadas a seguir.

4.2.1.1 Conversão das matrizes de temperatura em imagens

Para os resultados desta metodologia não sofrerem com perdas de dados causados pela conversão de temperatura para valores em níveis de cinza, a posterior análise de medidas de grafo é feita com o uso das temperaturas. Por outro lado, a fim de realizar a etapa de modelagem de grafo, é necessário usar imagens em escala de cinza, resultantes da conversão das matrizes. As imagens foram geradas pela adaptação da conversão de escalas termométrica, conforme apresentado na equação 4.1.

$$p_{ij} = 255 \cdot \frac{t_{ij} - t_{\min}}{t_{\max} - t_{\min}} \quad (4.1)$$

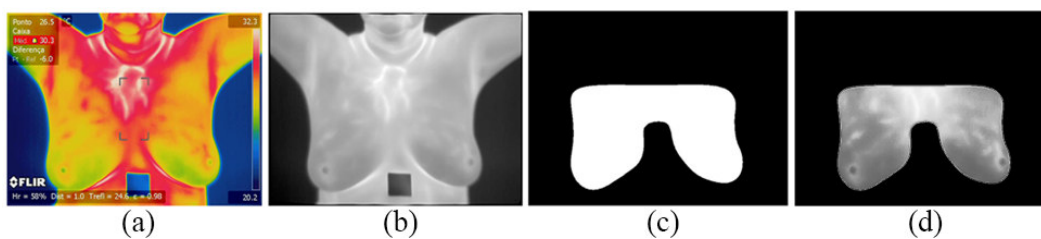
, onde p_{ij} representam os valores dos pixels da imagem a serem obtidos na conversão dos valores de temperatura t_{ij} da matriz. Já t_{\max} e t_{\min} são respectivamente a maior e menor temperaturas encontradas entre todos os exames e 255 é o maior valor de cinza possível.

O valor de t_{max} encontrado foi de 35,92 °C. Já o valor de t_{min} foi de 22,56 °C, mas foi ajustado para 22,40 °C, para evitar que, na conversão, a menor temperatura resulte em um pixel com valor 0, se confundindo com o fundo da imagem.

4.2.1.2 Extração da Região de Interesse

A *Region of Interest* (ROI) de cada termografia foi extraída por meio da subtração entre o termograma em escala de cinza e a máscara da ROI, que é fornecida pela base. Este resultado é demonstrado na Figura 16 (d).

Figura 16 – Resultado da segmentação da termografia da mama: (a) o termograma original obtido pela câmera térmica; (b) a representação da imagem em escala de cinza obtida na etapa de Aquisição; (c) a máscara da ROI; (d) a aplicação da máscara da ROI sobre a imagem em escala de cinza.



Antes de ser extraída a ROI, as imagens em escala de cinza passam por um melhoramento de contraste com o método *Contrast-Limited Adaptive Histogram Equalization* (CLAHE) e, logo em seguida, por um filtro da mediana com *kernel* igual a 5 para suavizar os ruídos provocados pela aplicação do CLAHE (PIZER et al., 1990) (GONZALEZ; WOODS, 2012).

4.2.2 Segmentação de regiões de superpixels

Nesta etapa, o termograma em escala de cinza passa pelo processo de tesselação, sendo dividido em regiões de superpixels, como preparação para etapa posterior, conforme técnicas referenciadas na Seção 3.2.4.

Optou-se por realizar uma tesselação irregular, por superpixels, uma vez que prioriza o relacionamento geográfico das regiões dos termogramas e valoriza as diversas formas das subregiões das imagens. Então foi aplicado o algoritmo SLIC, proposto por Achanta et al. (2011) e detalhado na Subseção 3.2.4. A Figura 17 mostra os resultados de aplicação de SLIC para diversas quantidades de superpixels.

Figura 17 – Divisão da ROI em regiões de superpixels: (a) a imagem original; (b) a ROI segmentada e com ajuste de contraste e filtro da média; (c) aplicação de SLIC com aproximadamente 30 superpixels e (d) com aproximadamente 250 superpixels.

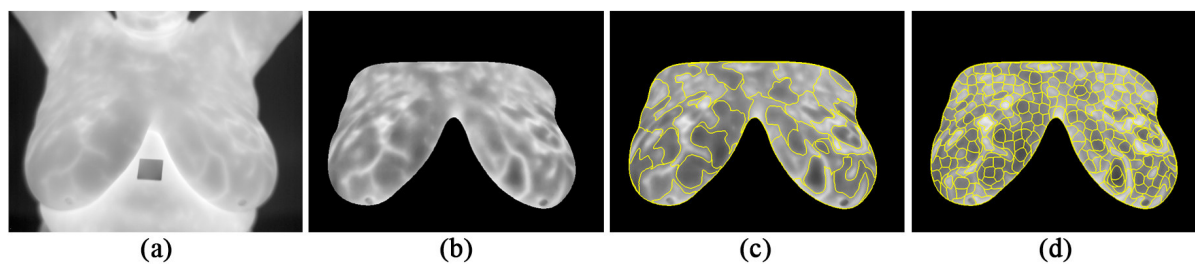


Figura 18 – Exemplos de aplicação de SLIC em 4 termogramas diferentes, com 200 sementes, e compactação $m = 10$.

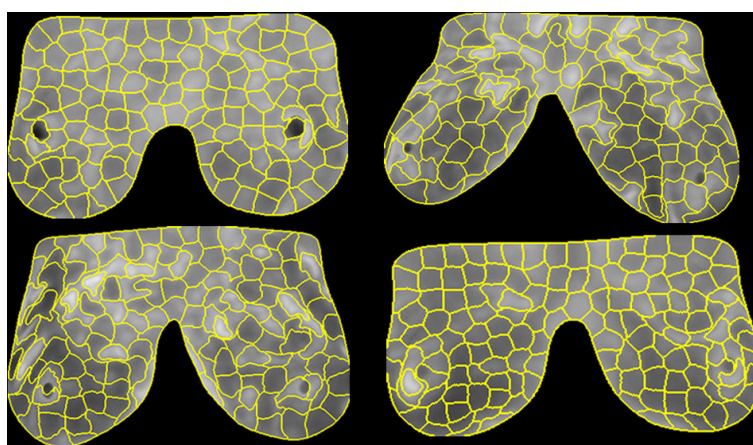
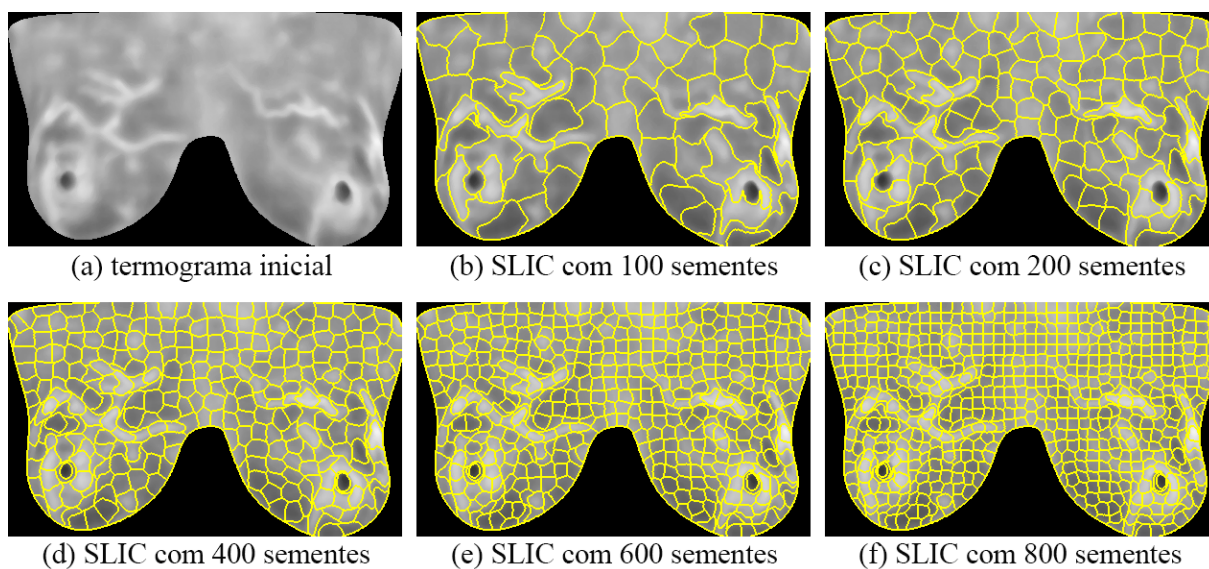


Figura 19 – Aplicação de SLIC com variação na quantidade de sementes. (a) termograma inicial com CLAHE e filtro da média; (b) aplicação de SLIC com 100 sementes, (c) 200, (d) 400, (e) 600 e (f) 800 sementes.



Para o parâmetro do SLIC que controla a quantidade de sementes, foram escolhidos 5 valores: 100, 200, 400, 600 e 800. A comparação dos resultados podem ser

vistos na Figura 19. Destaca-se que o método SLIC é aplicado sobre o *bounding box* da ROI. Esta, por sua vez, possui espaços vazios, representados por fundo de cor preta. Os superpixels que recaem sobre essas regiões vazias são descartados. Por este motivo, a quantidade de superpixels produzidos é menor do que a quantidade de sementes utilizadas no SLIC. Como no exemplo observado na Figura 19 (b), para 100 superpixels, ao descartar os superpixels fora da ROI, sobram apenas 73. E na Figura 19 (c), para 200 superpixels, sobram 154 superpixels dentro da ROI.

O parâmetro do SLIC que controla a compactação dos superpixels foi escolhido como 10 (de uma faixa entre 1 a 20), valor padrão sugerido pelos autores do método. Valores baixos ponderam mais a proximidade de cor, fornecendo superpixels com forma mais diversificadas. Enquanto que, valores altos resultam em superpixels de forma mais quadrada. Percebe-se, pela Figura 18, que o mesmo parâmetro de compactação promoveu uma boa adequação às diferentes formas das regiões dos termogramas.

Nos diversos experimentos, todas as etapas da metodologia são repetidas para cada um desses valores de sementes de SLIC, com o objetivo de verificar qual valor desse parâmetro permite obter os melhores resultados de detecção.

4.2.3 Modelagem de grafo a partir de termograma

Para os exames serem analisados conforme as medidas de redes complexas apresentadas na Seção 3.2, é fundamental a geração do grafo como forma de representação do termograma. Nesta etapa, para cada exame é construído um grafo ponderado e não-direcionado, cujos nós representam os superpixels e as arestas representam as relações de adjacência entre os superpixels. Os pesos das arestas são calculados por função de similaridade de temperatura. A seguir são apresentadas as duas variações de modelagens utilizadas nesta etapa.

Na primeira modelagem desenvolvida, nomeada **modelagem direta**, cada superpixel é representado por um nó. A relação de vizinhança se dá quando um superpixel é fronteiro a outro, representada por uma aresta não-direcionada. O peso w_{uv} da aresta entre um nó u e um nó v é definida como o valor absoluto da diferença entre as maiores

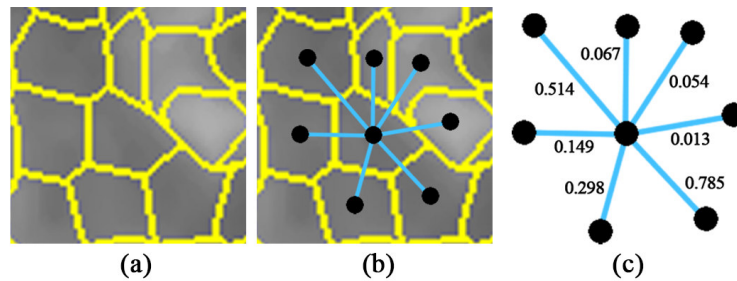
temperaturas de cada superpixel, conforme definida pela Equação 4.2:

$$w_{uv} = \gamma - |t_{max(u)} - t_{max(v)}| \quad (4.2)$$

, onde γ é uma constante que assume o valor igual 10, que corresponde um valor seguro, superior à diferença entre a maior e a menor temperaturas máximas de todos os superpixels. Quando u e v possuem temperaturas máximas aproximadas, o peso será alto. E quando há uma diferença muito grande entre as temperaturas máximas de u e v , o peso w_{uv} terá valor baixo. Ou seja, essa fórmula provê que sejam valorizadas as conexões entre regiões que mais se assemelham. Após a geração do grafo, os pesos de todas as arestas são normalizados para valores entre 0 e 1.

A Figura 20 ilustra de forma simplificada o processo da modelagem em grafo, na perspectiva de um nó. Na Figura 20 (c) é apresentado, o resultado do grafo gerado através da modelagem direta. Para o grafo dessa figura, os pontos são plotados exatamente no centroide dos superpixels.

Figura 20 – Processo de construção de grafo a partir das regiões de superpixels. (a) uma amostra de regiões; (b) representação de um grafo considerando somente as arestas do nó central; (c) o subgrafo final com os pesos das arestas calculados e normalizados.



Na segunda modelagem, chamada de **modelagem inversa**, os passos da primeira modelagem são repetidos, diferenciando apenas ao inverter a maneira de calcular o peso das arestas:

$$w_{uv} = \beta + |t_{max(u)} - t_{max(v)}| \quad (4.3)$$

, onde β , é uma constante que assume o valor de 0,001, para evitar que a diferença seja 0, quando as áreas adjacentes u e v tenham os seus maiores valores de temperatura iguais. Quando há uma diferença muito grande entre as temperaturas máximas de u e v , o peso w_{uv} terá valor alto. Nessa segunda modelagem, as conexões entre regiões diferentes têm peso maior do que as conexões entre regiões semelhantes. Na Figura 21 é apresentado o processo de construção do modelo de grafo. Os resultados dos dois esquemas

Figura 21 – Processo de construção do modelo de grafo, seguindo a modelagem direta. (a) termografia inicial; (b) após aplicar SLIC com 100 sementes, gerando 67 superpixels; (c) o grafo construído plotado em cima da imagem inicial, seguindo o esquema de Modelagem 1; (d) o grafo destacado.

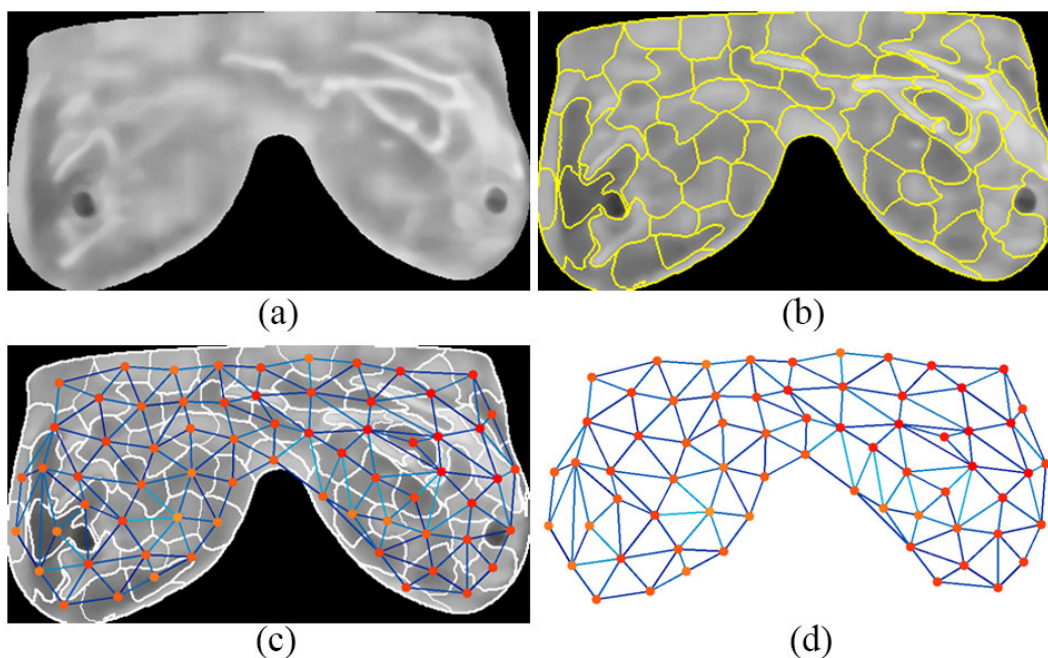
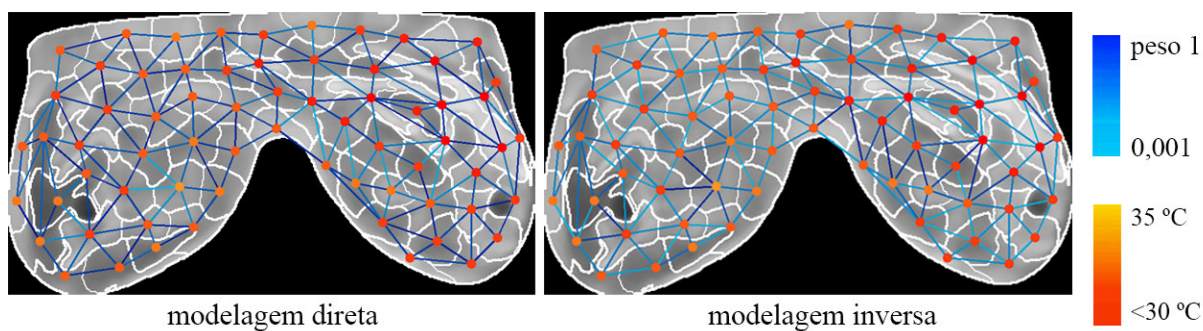


Figura 22 – Resultados gerados a partir dos dois tipos de modelagens.



de modelagens podem ser vistos na Figura 22. E as principais diferenças são sumarizadas na Tabela 3.

Tabela 3 – Comparação dos esquemas de modelagem de grafo.

relação entre vértices u e v	Modelagem direta	Modelagem inversa
	peso de aresta	
$t_{max(u)} \approx t_{max(v)}$	peso alto	peso baixo
$t_{max(u)} \not\approx t_{max(v)}$	peso baixo	peso alto
Cálculo do peso da aresta	Equação 4.2	Equação 4.3

4.2.4 Limiarização de grafo

Inicialmente, o grafo obtido contém um conjunto de arestas que conectam todos os nós entre si, demonstrando que a rede tem certa regularidade, pois a maioria dos nós possuem quase a mesma quantidade de ligações. Se o grafo continuar dessa maneira, apresentará poucas propriedades relevantes que permitem observar padrões distintivos entre as regiões. Por isso, é preciso transformar o grafo para que este ganhe mais aspectos de redes de pequeno mundo.

A fim de obter esse resultado, é aplicada a técnica de limiarização de grafo (detalhada na Subseção 3.2) para criar variações de um mesmo grafo, usando eliminação de arestas de determinada faixa de peso.

A limiarização, aqui definida, elimina as arestas menores que o limiar L . Em termos de matriz de adjacência A , a equação que condiciona a limiarização é definida da seguinte maneira:

$$a_{uv} = \begin{cases} 0, & \text{se } w_{uv} \leq L \\ a_{uv}, & \text{caso contrário} \end{cases} \quad (4.4)$$

Para a modelagem direta, as arestas com peso baixo (que representam ligações entre regiões muito distintas) são eliminadas, restando somente as ligações entre as regiões mais homogêneas, que são as arestas com peso alto. Então, essa modelagem, atrelada à limiarização, é ideal para fazer comparações entre componentes conectados e/ou comunidades, usando como critério as medidas de conectividade local, estrutura de rede e correlação.

Na modelagem inversa, as arestas com peso baixo (que representam ligações entre regiões muito semelhantes) são eliminadas, restando somente as ligações entre as regiões mais distintas, que são as arestas com peso alto. Essa modelagem, junto com a limiarização, vai valorizar a ligação entre nós de características distintas. Para ambos tipos de modelagem, as arestas eliminadas na limiarização serão as arestas com menor peso, provocando um cenário ideal para análise de medidas de agrupamento.

Por meio de testes preliminares, foram escolhidos 5 valores distintos para o limiar L . Como os pesos das arestas são normalizados, L assumirá valores relativos aos percentis 0%, 20%, 40%, 60% e 80%. Exemplificando, o limiar de valor relacionado com o percentil

20% irá eliminar 20% das arestas, mantendo apenas 80% das ligações de maior peso (que dependendo do esquema de modelagem, podem representar as ligações entre regiões mais semelhantes ou mais distintas). A Figura 23 ilustra o processo de limiarização para um dado subgrafo. As Figuras 24 e 25 mostram o resultado de diferentes graus de limiarização aplicados a cada um dos esquemas de modelagens sobre uma mesma termografia.

Figura 23 – Processo de limiarização de grafo. (a) amostra do grafo original com pesos normalizados; (b) o resultado de limiarização com $L=0,056$ correspondendo ao percentil de 20%, eliminando as arestas com o peso menor que 0,056; (c) o resultado para $L=0,238$, que corresponde ao percentil de 60%.

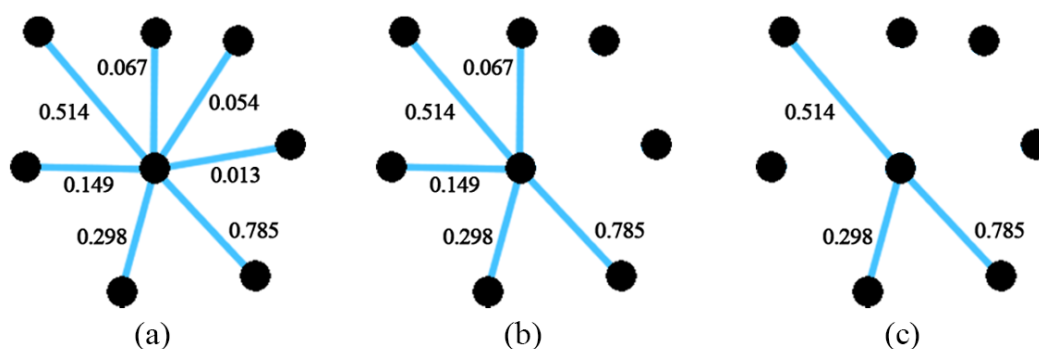
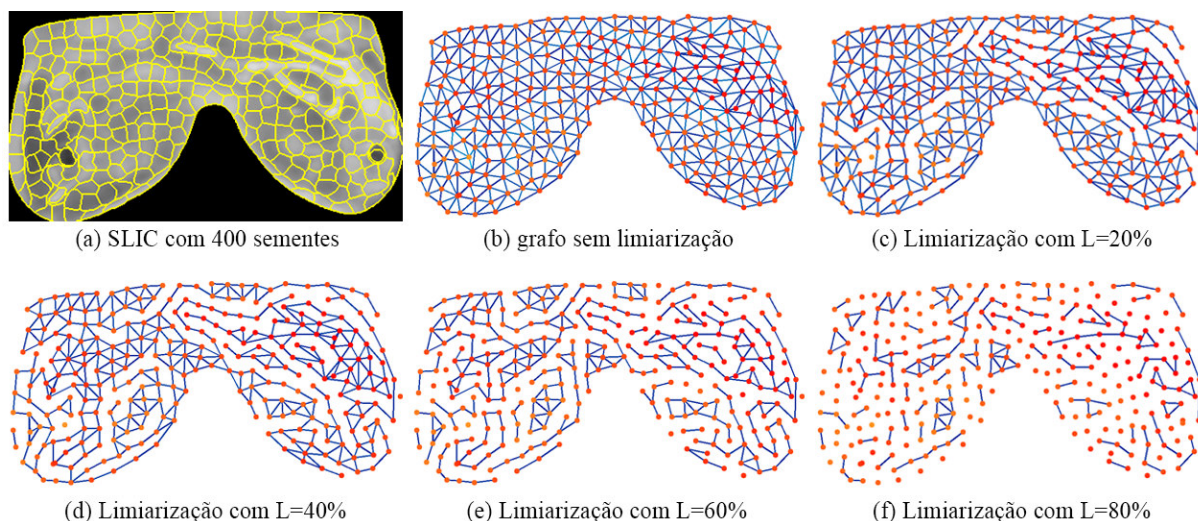


Figura 24 – Diferentes graus de limiarização com a modelagem direta.



Em meio aos testes iniciais com limiarização, percebeu-se a necessidade de realizar um ajuste na modelagem inversa, acrescentando um critério: se um nó faz parte do percentil dos 5% dos nós com maior temperatura máxima, as arestas que interligam esses nós não sofrem limiarização. Um acontecimento influenciou para o estabelecimento desse critério: durante a limiarização, em alguns grafos foram criados alguns “buracos” no grafo, nas regiões internas de agrupamentos muito grandes, demonstrado na Figura 26

Figura 25 – Diferentes graus de limiarização com a modelagem inversa.

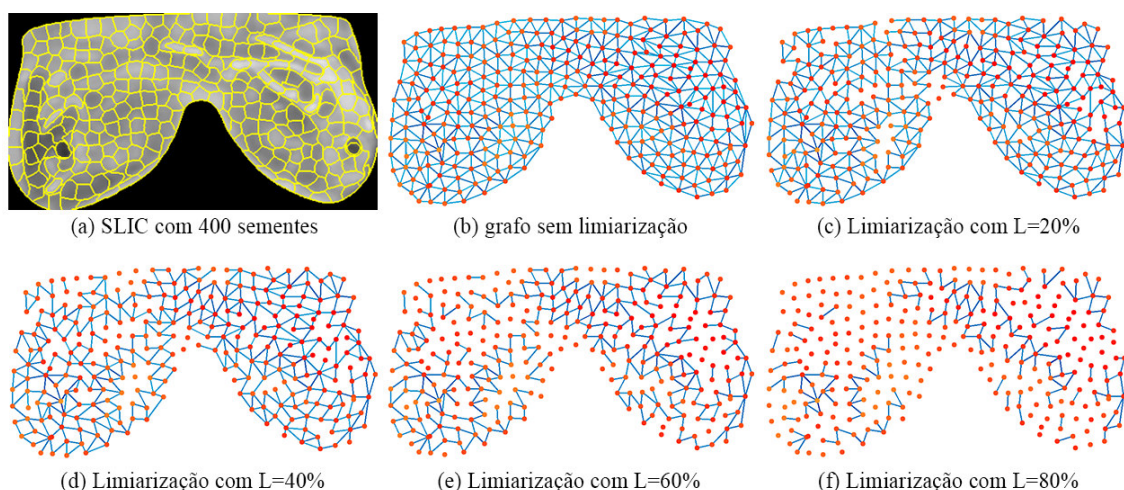
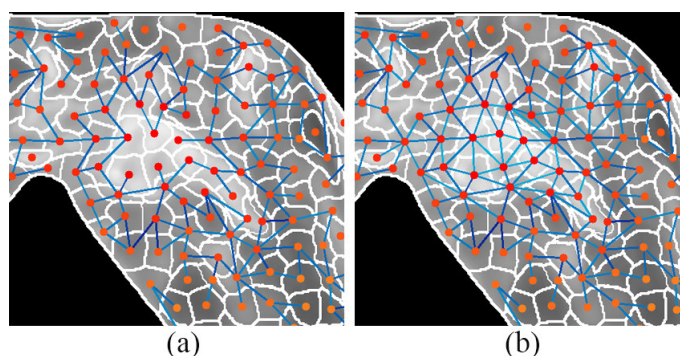


Figura 26 – Resultado do efeito de adição de critério de limiarização na modelagem inversa. Ampliação mostrando parte da mama esquerda, que contém anomalia nos quadrantes superiores. Ambas imagens com 400 sementes, limiarização com $L=60\%$. No grafo à esquerda (a), existe um espaço sem conexões. No grafo à direita (b), a mesma região não sofre limiarização por causa da adição do critério de não eliminação das arestas entre os nós mais quentes.



(a). Tais espaços vazios, sem vértices interligados, em muitos casos, coincidiam com parte da região da localização da anomalia, o que prejudicava na detecção desta. Considerando que os tumores, provavelmente (mas nem sempre), estão presente em regiões de maior temperatura essa foi uma estratégia que resolveu parte do problema. Após alguns testes, o valor 5% foi escolhido por se adaptar melhor a este cenário. Na Seção de resultados será relatada a influência que a adição desse critério causou à modelagem inversa.

4.2.5 Detecção da região da mama com anomalia

Esta etapa da metodologia consiste em verificar qual das medidas de grafo apontam para a região suspeita de anomalia. Para a realização desta etapa é necessário realizar vários experimentos para cada uma das variações de modelagens de grafo e depois

confrontar as áreas indicadas com as áreas na etapa seguinte, na validação dos resultados.

Com base nas metodologias apresentadas em alguns trabalhos relacionados, foram elaboradas algumas estratégias que se desdobram em experimentos, que são elencados na Tabela 4. Nesses experimentos foram usadas as medidas que, teoricamente, se mostram mais relevantes para o problema. E à medida em que os experimentos foram sendo realizados, foram percebidas características interessantes, as quais influenciaram a exploração de novas estratégias e adaptação para os experimentos, buscando otimizar os resultados.

Nos **experimentos I e II**, a forma de análise é baseada na comparação entre comunidade e componentes conectados. Primeiramente os grafos são divididos em comunidades e/ou componentes conectados, formando subgrafos. Cada um dos subgrafos é submetido a análise de **14 medidas**, definidas na Subseção 3.2.3 e listadas a seguir: grau máximo, coeficiente de clusterização, energia, entropia, grau médio da vizinhança de cada nó, coeficiente de Pearson (assortividade), número de vértices (ordem), densidade, centralidades (de grau, de intermediação, de proximidade, de autovetor), tamanho do clique máximo e número de cliques maximais. Nesses experimentos, para um determinado grafo, uma medida é aplicada a cada um de seus subgrafos, afim de saber qual subgrafo se destaca para tal medida, e posteriormente é verificado se esse subgrafo coincide com a região da anomalia.

Tabela 4 – Lista de experimentos de detecção.

Experimento	Forma de análise	Medidas	Modelagem
I	comparação de comunidades	14 medidas	Direta
II		14 medidas	Inversa
III	centralidade	cent. de intermediação	Inversa
IV		cent. de grau	
V		cent. de proximidade	
VI		cent. de autovetor	

Nos **experimentos III, IV, V e VI**, a forma de análise é feita por meio de medidas de centralidade aplicadas ao grafo inteiro, sem dividir em subgrafos. Para esses experimentos, são apresentados somente resultados com uso de grafo provenientes da modelagem inversa, pois as propriedades das medidas de centralidade exigem, preferencialmente, que muitos nós do grafo estejam conectados. A modelagem inversa atende melhor a essa exigência, mesmo quando submetida a graus de limiarização altos. Esse fato pode ser comprovado comparando grafos de diferentes modelagens para um mesmo limiar: a Figura 24 (e) possui maior quantidade de componentes desconectados e pequenos, enquanto a

Figura 25 (e) mostra menos componentes desconectados e mostra componentes conectados maiores.

4.2.6 Avaliação da detecção da região da mama suspeita de anomalia

A base DMR-TD utiliza o esquema de divisão de região das mamas em quadrantes para indicar a região com anomalia, não sendo possível saber as coordenadas XY da localização da anomalia e nem a sua extensão ou forma. Assim, não é possível avaliar com precisão, em termos numéricos, os índices de performance da metodologia, usando métricas de segmentação conhecidas. A fim de resolver o problema de avaliação da detecção, foi preciso definir um método de avaliação que pudesse ser aplicado a todos os exames, neste trabalho.

Os acertos de detecção são avaliados conforme o grau de proximidade entre o quadrante sinalizado e o quadrante correto, denotados pelos conceitos A, B, C e D, detalhados da Tabela 5 e ilustrado na Figura 27. Com essa avaliação, é possível mensurar o índice de exatidão para cada detecção, com base no nível de aproximação. Além disso, é possível quantificar a avaliação de todos os experimentos, verificando quais combinações de parâmetros (tipo de modelagem e grau de limiarização) obtiveram melhores resultados, considerando cada conceito separadamente.

Tabela 5 – Níveis de avaliação da detecção para os casos genéricos.

Nível	Significado do nível de aproximação de acerto
A	A região sinalizada pela detecção é a mesma região descrita no diagnóstico.
B	A região sinalizada pela detecção é adjacente ao quadrante correto, em relação às posições horizontais e verticais. Indica que houve um acerto de aproximação, com pouca exatidão.
C	A região sinalizada pela detecção é um quadrante oposto ao quadrante correto, em relação à diagonal, indicando que um houve acerto apenas da mama com anomalia, mas sem exatidão da localização, quanto aos quadrantes.
D	Sinaliza que houve erro de detecção da mama com anomalia.

No caso em que a anomalia estiver localizada nos dois quadrantes superiores (ao mesmo tempo), esses são sinalizados com nível A, em caso de coincidência do acerto de detecção. Então os quadrantes inferiores são avaliados com nível B. E a mama oposta que não tem anomalia é avaliada com nível D. Ao inverso, caso a anomalia estiver nos quadrantes inferiores, esses são avaliados por A, e os superiores por B. Sob essas condições inexiste o nível C. A Figura 28 (b) ilustra essa situação.

Figura 27 – Diagrama de validação de detecção para os todos os quadrantes da mama esquerda.

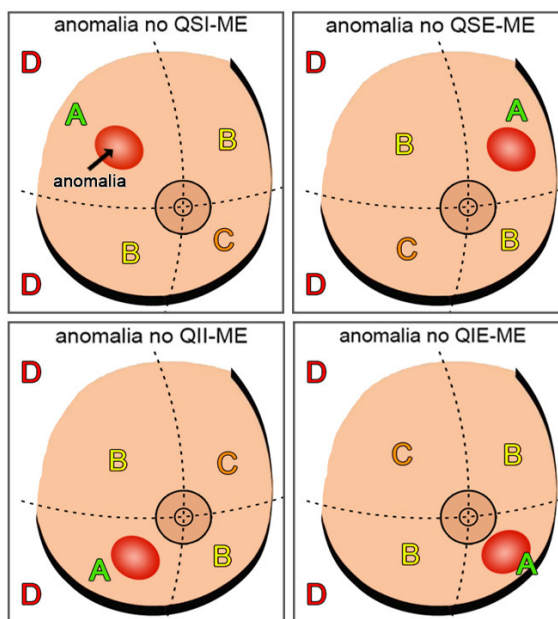
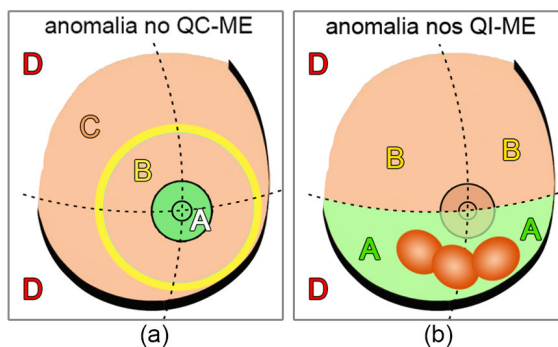


Figura 28 – Diagrama de validação de detecção para casos em que anomalia estiver quadrante central (a), ou em dois quadrantes inferiores (b).



Em alguns exames, a localização da anomalia é dada no quadrante central, que é a região da auréola mamilar. Para esses casos, a detecção é considerada como: nível A, se a região indicada faz parte da auréola mamilar; nível B, se estiver em um raio r , sendo r a distância entre o mamilo e a borda (da mama) mais próxima; nível C, se acertou a mama afetada; e nível D se errou a localização da mama. A Figura 28 (a) ilustra esse caso.

Existem também casos com anomalias de grande tamanho, que acometem toda a região de uma mama. Nesses casos há somente dois níveis: A para acerto e D para erro.

5 RESULTADOS E DISCUSSÃO

Neste capítulo são listados os resultados obtidos e feitas discussões acerca de cada experimento aplicado à metodologia proposta para detecção de regiões de anomalia em termografia de mama.

Seguindo os passos da metodologia, primeiramente, a base de termografias passou por uma seleção de exames diagnosticados como doentes e que continham descrição do tumor, totalizando 22 termografias. Depois as termografias foram submetidas à primeira etapa da metodologia, o pré-processamento (Seção 4.2.1), onde foram feitos os procedimentos: conversão de matriz de temperatura para imagem em escala de cinza; melhoramento de contraste e filtro de redução de ruídos; e extração da ROI.

Na etapa subsequente, é aplicado o algoritmo SLIC para geração dos superpixels (Seção 4.2.2). Esta etapa foi repetida para 5 variações de quantidade de sementes S (100, 200, 400, 600 e 800). Usou-se $m = 10$ como parâmetro de compactação da forma do superpixel. Foi escolhido o número de iterações do algoritmo igual a 12, pois, a partir dessa quantidade, o SLIC não apresentava tanta diferença quando se comparava regiões de superpixels produzidas por cada iteração.

Com os superpixels gerados, foi possível implementar a etapa de modelagem de grafo (Seção 4.2.3), com testes usando as duas variações de modelagem. Logo após, foi realizada a etapa de limiarização de grafo para cada um dos 5 limiares L (Seção 4.2.4). Então, na etapa de detecção, foram realizados os experimentos listados, com análises de medidas de grafo (Seção 4.2.5). E, na etapa final da metodologia, ocorreu a avaliação de cada experimento, e foram aferidos os níveis de aproximação de acerto, com base nos diagramas de avaliação elaborados (Seção 4.2.6).

Ressalta-se que, em alguns experimentos, a metodologia foi aplicada para cada uma das variações de parâmetros dos principais métodos. Isto é, foram gerados 2 repositórios: um para a modelagem direta e outro para a modelagem inversa. Para cada um dos tipos de modelagem foram criados 25 subrepositórios, que correspondem a multiplicação dos 5 valores possíveis de sementes S usados na geração dos superpixels e os 5 valores para limiarização L , totalizando, assim, 50 repositórios de grafos, provenientes do mesmo grupo de termografias selecionadas.

Nas seções seguintes, são expostos os resultados dos experimentos, seguindo a

lista de configuração estabelecida na Tabela 4. Juntamente, serão feitas discussões sobre esses experimentos, demonstrando os exemplos mais relevantes para cada caso.

5.1 Experimento I

Realizando o primeiro experimento, na etapa de detecção da região com anomalia, a forma de análise ocorreu por meio de comparação de comunidades. Cada grafo foi dividido em comunidades (conforme método da Seção 3.2.3.8), e em cada uma das comunidades foram aplicadas 14 medidas de grafo. A comunidade que teve maior índice para determinada medida foi escolhida como região detectada para aquela medida. Após a etapa de detecção, aconteceu a etapa de avaliação, onde foram contados os acertos de detecção para cada uma das medidas.

Para se saber qual era o repositório ideal para a análise nesse experimento, buscou-se quantificar a modularidade dos grafos de cada repositório, pois os repositórios de grafos, cujas quantidades de sementes resultem em modularidade alta, têm suas comunidades mais bem definidas e segregáveis. A Tabela 6 mostra que para $S = 800$ foi produzida uma maior modularidade de comunidade. Observa-se também que os grafos com $S = 600$ apresentaram valor de modularidade muito próximo aos grafos com $S = 800$, sendo que esses dois grupos produziram quantidade de comunidades semelhante. Além disso, 600 sementes resultam em um grafo com menos nós, diminuindo assim, o custo computacional na geração das comunidades e aferição das medidas. Por estes motivos, 600 foi o valor escolhido como número de sementes para os experimentos I e II.

Tabela 6 – Modularidade da rede para cada quantidade de sementes de SLIC, considerando todos os exames selecionados.

Qtd. sementes	Modularidade da rede	Média de comunidades
100	0.599	4.2
200	0.658	4.8
400	0.705	5.1
600	0.722	5.2
800	0.725	5.2

Já os percentuais de limiarização escolhidos foram 40% e 60%, pelo fato de produzirem uma boa quantidade de componentes, sem produzir muitos nós desconectados. A Figura 29 apresenta a comparação entre as comunidades geradas com as duas modelagens e a Figura 30 demonstra a comparação entre as comunidades geradas com grafos de cada

percentual de limiarização. Nesta última figura, percebe-se que os limiares 0 e 20% influenciaram para que as comunidades se apresentem muito grandes. Nesse caso, uma comunidade abrange mais de uma região de quadrante, o que traria dificuldade na etapa de avaliação de detecção, por isso os limiares 0 e 20% foram rejeitados nos experimentos I e II. Por outro lado, o limiar 80% (Figura 30 (f)) foi rejeitado também, porque influenciou na obtenção de muitos nós isolados e produziu comunidades de tamanho muito pequeno.

Figura 29 – Comparação de segmentação de comunidades para cada modelagem.

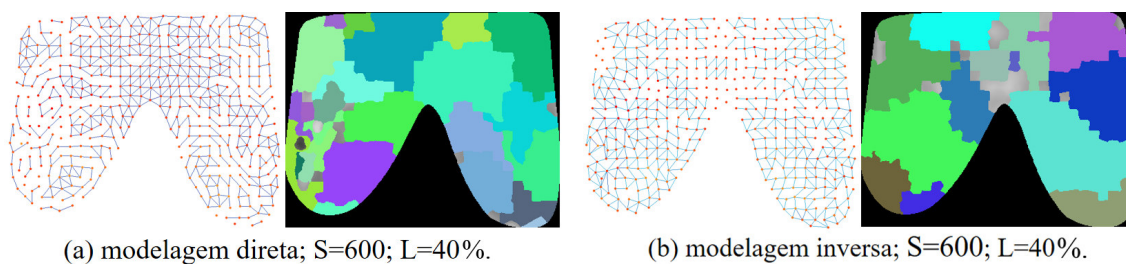
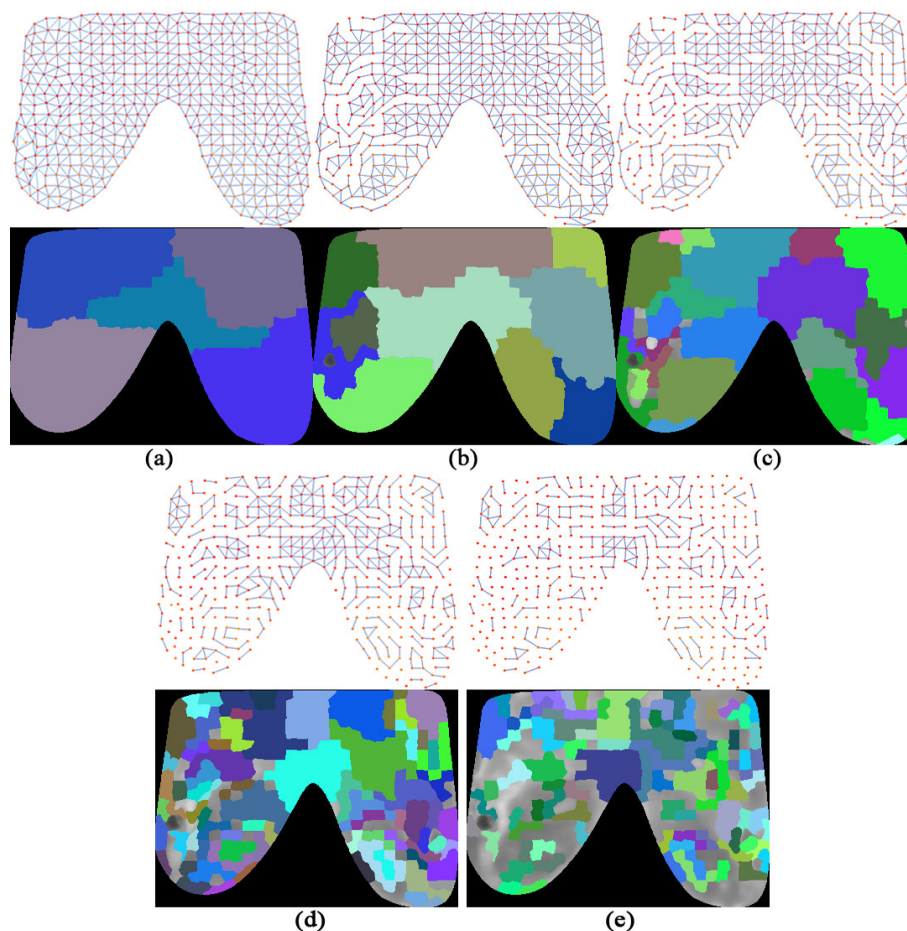


Figura 30 – Segmentação de comunidades a partir dos grafos. SLIC com 600 sementes com (a) 0%, (b) 20%, (c) 40%, (d) 60% e (e) 80% de limiarização.



Após a segmentação de comunidades de grafo e a comparação entre cada comunidade em relação às medidas, os índices de avaliação de acerto de detecção são apresentados

nas Tabelas 7 e 8.

No experimento I-a, foram usados os grafos modelados com os parâmetros: $S = 600$ e $L = 40\%$. Observa-se pela Tabela 7 que as medidas que alcançaram neste experimento maior índice de acerto foram grau máximo e grau médio de vizinhança. Ambas empatadas, alcançaram percentual de acerto de 36,4% para nível A. Os níveis agrupados AUB e AUBUC, obtiveram percentuais de nível de proximidade a 63,6% e 68,2%, respectivamente, demonstrando que há uma razoável tendência para que as comunidades identificadas apontem para a proximidade da região de anomalia.

Tabela 7 – Resultados de avaliação do Experimento I-a. Parâmetros: modelagem direta; $S=600$; $L=40\%$.

medidas	avaliações em níveis				níveis agrupados (%)		
	A	B	C	D	A	AUB	AUBUC
Grau máximo	8	6	1	7	36,4	63,6	68,2
Coef. de agrupamento	7	6	1	8	31,8	59,1	63,6
Energia	7	4	1	10	31,8	50,0	54,5
Entropia	7	4	1	10	31,8	50,0	54,5
Grau médio de vizinhança	8	6	1	7	36,4	63,6	68,2
Coefficiente de Pearson	2	2	2	16	9,1	18,2	27,3
Número de vértices	5	6	1	10	22,7	50,0	54,5
Densidade	1	2	2	17	4,5	13,6	22,7
Centralidade de grau	5	4	0	13	22,7	40,9	40,9
Central. de intermediação	6	5	0	11	27,3	50,0	50,0
Central. de proximidade	6	4	0	12	27,3	45,5	45,5
Central. de autovetor	6	1	2	13	27,3	31,8	40,9
Tamanho do clique máximo	6	3	1	12	27,3	40,1	45,5
Número de cliques maximais	6	5	1	10	27,3	50,0	54,5

Tabela 8 – Resultados de avaliação do Experimento I-b. Parâmetros: modelagem direta; $S=600$; $L=60\%$.

medidas	avaliações em níveis				níveis agrupados (%)		
	A	B	C	D	A	AUB	AUBUC
Grau máximo	7	3	2	10	31,8	45,5	54,5
Coef. de agrupamento	6	6	1	9	27,3	54,5	59,1
Energia	7	3	1	11	31,8	45,5	50,0
Entropia	7	3	1	11	31,8	45,5	50,0
Grau médio de vizinhança	7	5	1	9	31,8	54,5	59,1
Coefficiente de Pearson	1	4	2	15	4,5	22,7	31,8
Número de vértices	7	4	1	10	31,8	50,0	54,5
Densidade	5	4	1	12	22,7	40,9	45,5
Centralidade de grau	5	4	0	13	22,7	40,9	40,9
Central. de intermediação	6	8	1	7	27,3	63,6	68,2
Central. de proximidade	8	2	0	12	36,4	45,5	45,5
Central. de autovetor	7	5	0	10	31,8	54,5	54,5
Tamanho do clique máximo	8	3	0	11	36,4	50,0	50,0
Número de cliques maximais	5	6	1	10	22,7	50,0	54,5

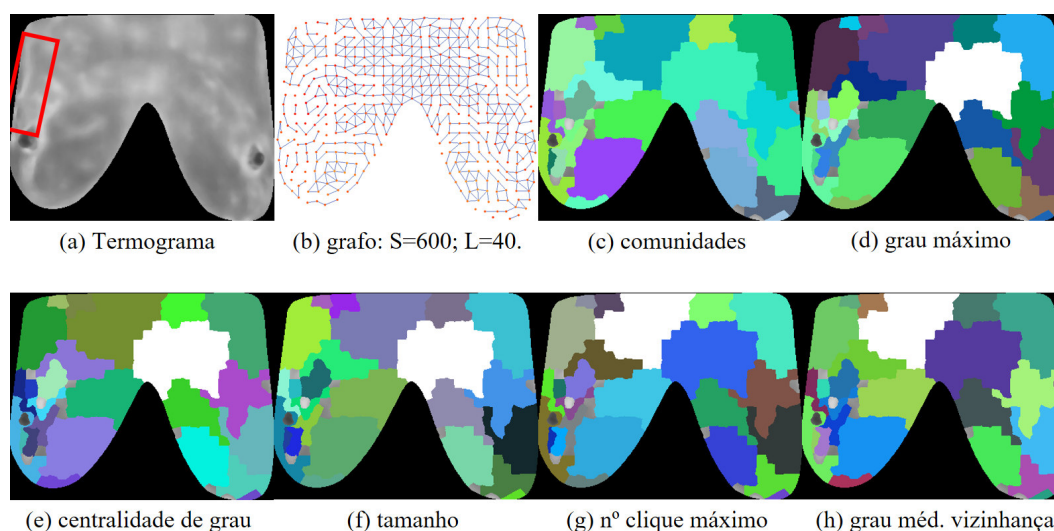
No experimento I-b, cujos resultados são mostrados na Tabela 8, o parâmetro S

foi mantido, e alterado L para 60%. As medidas que mais se destacaram foram centralidade de proximidade e tamanho do clique máximo, ambas para percentual de proximidade de nível A, com 36,4%. Já a centralidade de intermediação forneceu maiores acertos para os níveis de proximidade AUB e AUBUC, com 63,6% e 68,2%, nessa ordem.

Nota-se que as variações (a e b) do experimento I resultaram em percentuais de acertos iguais para cada nível de proximidade, entretanto as medidas de maior desempenho foram diversas.

Como exemplo, mostra-se a Figura 31 com o exame nº 202. Para esse, no experimento I não houve acerto de nível A, sendo todos classificados com nível D (erro), pois as comunidades que se destacaram para cada medida não indicaram corretamente a região de anomalia. Observa-se que, para este caso, a região de anomalia corresponde à parte do grafo que contém poucas conexões, assim todas as medidas certamente tendem a não se mostrar com valores elevados nessa região. Isso mostra que a modelagem direta contribuiu negativamente. O mesmo ocorre com o exemplo da Figura 32.

Figura 31 – Experimento I-a sobre o exame nº 202: (a) termograma com demarcação da anomalia no QSE-MD; (b) grafo correspondente; (c) divisão em comunidades; (e - h) em cor branca as comunidades que se destacaram em relação a algumas medidas.



Como exemplo de resultados positivos, tem-se a Figura 33 com o exame nº 292. Nesse, com experimento I, houve alguns acertos de nível A (Figura 33 (d) e (g)), sendo a maioria classificada com nível D (erro). Nota-se que a região de anomalia teve suas arestas eliminadas durante a limiarização, assim nem todas as medidas possuem valores altos nessa região. Mais uma vez mostrando que a modelagem direta não contribuiu neste cenário.

Figura 32 – Experimento I-a sobre o exame n° 286: (a) termograma com demarcação da anomalia no QSE-ME; (b) grafo correspondente; (c) divisão em comunidades; (e - h) em cor branca as comunidades que se destacaram em relação a algumas medidas.

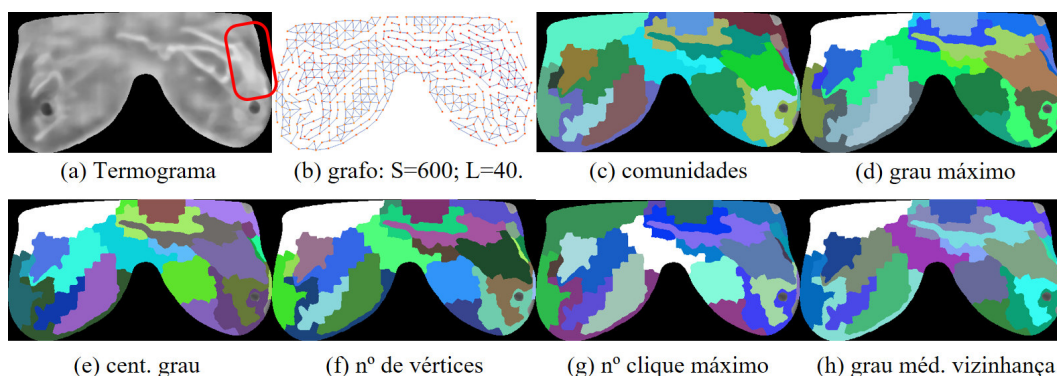
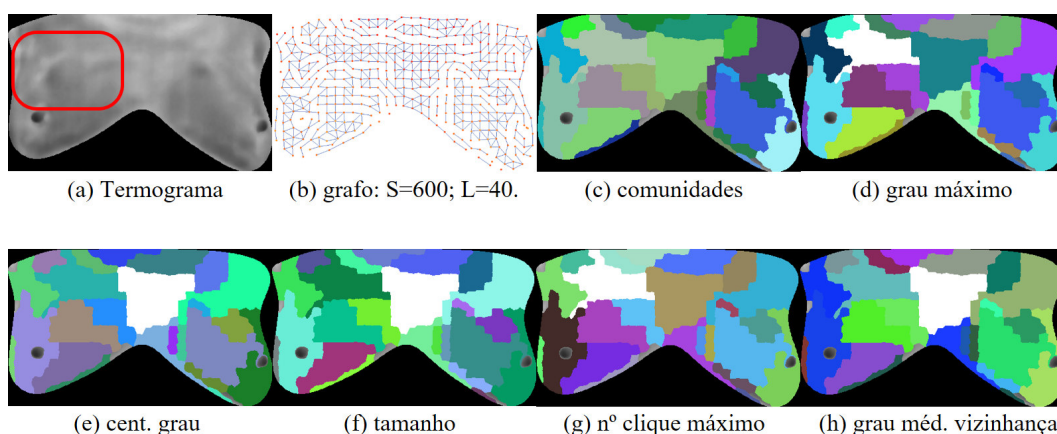


Figura 33 – Experimento I-a sobre o exame n° 292: (a) termograma com demarcação da anomalia no QSE-MD e QSD-MD; (b) grafo correspondente; (c) divisão em comunidades; (e - h) em cor branca as comunidades que se destacaram em relação a algumas medidas.



5.2 Experimento II

Neste segundo experimento, o procedimento é idêntico ao experimento I, alterando apenas o tipo de modelagem para a modelagem inversa. Os resultados de avaliação são mostrados na Tabela 9 (experimento II-a, com parâmetro $L = 40$) e na Tabela 10 (experimento II-b, com parâmetros $L = 60$).

Para os experimentos II-a e II-b, a característica que apresentou maior índice de acertos foi o número de vértices, com avaliação de nível A em 40,9% e 45,5%, respectivamente. Como resultados para os níveis de proximidade AUB e AUBUC, com o coeficiente de agrupamento e o grau médio de vizinhança, os valores se fixaram em 68,2% em II-a e 63,6% em II-b.

Tabela 9 – Resultados de avaliação do Experimento II-a. Parâmetros: modelagem inversa; S=600; L=40%.

medidas	avaliações em níveis				níveis agrupados (%)		
	A	B	C	D	A	AUB	AUBUC
Grau máximo	5	5	1	11	27,3	45,5	50,0
Coef. de agrupamento	7	8	0	7	31,8	68,2	68,2
Energia	7	3	2	10	31,8	45,5	54,5
Entropia	7	3	2	10	31,8	45,5	54,5
Grau médio de vizinhança	8	7	0	7	36,4	68,2	68,2
Coeficiente de Pearson	1	3	4	14	4,5	18,2	36,4
Número de vértices	9	3	1	9	40,9	54,5	59,1
Densidade	4	4	1	13	18,2	36,4	40,9
Centralidade de grau	7	3	1	11	31,8	45,5	50,5
Central. de intermediação	3	2	3	14	13,6	22,7	36,4
Central. de proximidade	6	2	1	12	27,3	36,4	45,5
Central. de autovetor	5	2	3	12	22,7	31,8	45,5
Tamanho do clique máximo	6	3	2	11	27,3	40,9	50,0
Número de cliques maximais	5	6	2	9	22,7	50,0	59,1

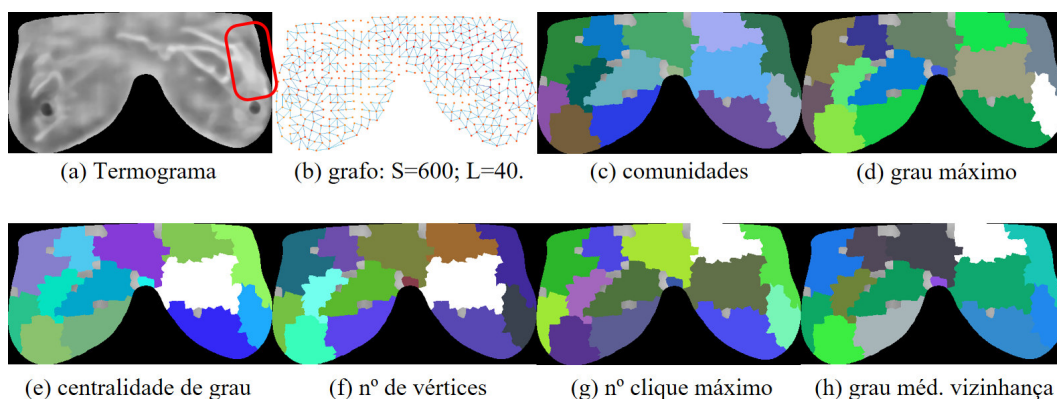
Tabela 10 – Resultados de avaliação do Experimento II-b. Parâmetros: modelagem inversa; S=600; L=60%.

medidas	avaliações em níveis				níveis agrupados (%)		
	A	B	C	D	A	AUB	AUBUC
Grau máximo	6	5	1	10	27,3	50,0	54,5
Coef. de agrupamento	8	6	0	8	36,4	63,6	63,6
Energia	8	3	1	10	36,4	50,0	54,5
Entropia	8	3	1	10	36,4	50,0	54,5
Grau médio de vizinhança	8	5	1	8	36,4	59,1	63,6
Coef. de Pearson	1	2	4	15	4,5	13,6	31,8
Número de vértices	10	2	0	10	45,5	54,5	54,5
Densidade	5	5	0	12	22,7	45,5	45,5
Centralidade de grau	7	3	0	12	31,8	45,5	45,5
Central. de intermediação	4	2	3	13	18,2	27,3	40,9
Central. de proximidade	7	3	1	11	31,8	45,5	50,0
Central. de autovetor	6	3	2	11	27,3	40,9	50,0
Tamanho do clique máximo	8	3	0	11	36,4	50,0	50,0
Número de cliques maximais	7	3	1	11	31,8	45,5	50,0

Observa-se que o experimento II apresentou resultados superiores em relação ao experimento I. Como a única diferença entre os dois experimentos é a forma de modelagem, entende-se que a modelagem inversa provou ser mais adequada para a finalidade de analisar medidas de grafo.

A Figura 34 mostra o exame de nº 286, com análise no experimento II. Este obteve um acerto de nível A (de forma parcial), com a medida grau máximo. E a maior parte das outras classificações das avaliações foi nível B, pois as comunidades que se destacaram para cada medida indicaram uma região de anomalia aproximada da região correta. Nesse sentido, o experimento II se saiu superior em relação ao experimento I.

Figura 34 – Experimento II-a sobre o exame n° 286: (a) termograma com demarcação da anomalia no QSE-ME; (b) grafo; (c) divisão em comunidades; (e - h) em cor branca as comunidades que se destacaram em relação a algumas medidas.



5.3 Experimento III

Nos experimentos de III a VI foram usados somente grafos gerados pela modelagem inversa, por razões apresentadas na Seção 4.2.5, além do fato de a modelagem inversa apresentar melhores resultados no experimento II. Em cada um dos experimentos de III a VI foi aplicada uma medida diferente de centralidade, visando detectar a região suspeita como a região de maior medida de centralidade. Uma vez que os experimentos com centralidade possuem apenas uma medida por experimento, então, para algumas centralidades foi possível explorar todas as 25 combinações de parâmetro de SLIC e limiarização, aplicados à modelagem inversa.

No experimento III, foi analisada a medida de centralidade de intermediação. Seus valores de avaliação são exibidos na Tabela 11, onde se mostra que, para o nível A, o limiar de 80% se mostrou sobressalente em relação aos demais limiares, chegando a 31,8%. No nível agrupado AUBUC atingiu 59,1%. Por outro lado, em AUB alcançou 54,5% com os limiares 20% e 60%.

Tabela 11 – Resultados de avaliação do Experimento III. Parâmetros: modelagem inversa; S=600; L de 0 a 80%.

S=600 Limiar	avaliações em níveis				níveis agrupados (%)		
	A	B	C	D	A	AUB	AUBUC
0%	6	4	2	10	27,3	45,5	54,5
20%	4	8	0	10	18,2	54,5	54,5
40%	2	6	0	14	9,1	36,4	36,4
60%	6	3	1	12	18,2	54,5	54,5
80%	7	4	2	9	31,8	50,0	59,1
acum. total (%)	22,7	22,7	4,5	50,0	22,7	45,5	50,0

Esses resultados comprovam que a centralidade de intermediação é útil para analisar os nós que mais aparecem na maioria dos menores caminhos, ou seja, os nós do tipo *hub* de rede. Pela estrutura da modelagem inversa, é esperado que os nós que façam parte da região do tumor não sejam nós do tipo *hub*. Esse argumento é comprovado pelo fato de que os poucos acertos de nível A se deram para o limiar 80%, que é, justamente, a ocasião em que houve muita desconexão de nós, aumentando a probabilidade de um acerto eventual, por mera sorte.

As Figuras 35, 36 e 37, comprovam esse fato. Essas três figuras contêm três exames diferentes, todas com os limiares 20, 40, 60 e 80%, aplicados ao experimento III. Esses exames contêm suas anomalias localizadas respectivamente em QII, QSE e QS das mamas esquerdas. Nessas figuras foram plotados pontos que representam os (5 a 10% dos) nós de maior centralidade de intermediação, sinalizados pelos pontos azuis. Os pontos com cores mais claras são os nós de maior centralidade, liderados pelo nó de maior tamanho. O maior ponto azul-claro plotado possui a maior centralidade e aponta para a região com suspeita de anomalia. O retângulo indica o quadrante correto que contém a anomalia, sendo esse de cor vermelha quando há uma avaliação com nível D (erro de detecção), cor amarela se há uma avaliação de nível B (uma detecção aproximada), retângulo azul para nível C, e cor verde indicando um acerto de detecção, quando o ponto de maior centralidade coincide com a região, conquistando avaliação de nível A. Outro aspecto interessante é o fato de os limiares 20 e 40% apontarem o nó de maior centralidade na região central da imagem, entre as duas mamas. Isso mais uma vez comprova a argumentação sobre as características de *hub* que a centralidade de intermediação mede. Então, por esses resultados e análises feitas, constata-se que essa característica de centralidade não é uma boa medida para ajudar a alcançar o objetivo deste trabalho.

Figura 35 – Experimento III sobre o exame n° 291 com $S = 600$. O maior ponto azul plotado possui a maior centralidade e aponta para a região com suspeita de anomalia. O retângulo indica o quadrante correto que contém a anomalia. Níveis de avaliação: (a) D; (b) D; (c) D; e (d) A.

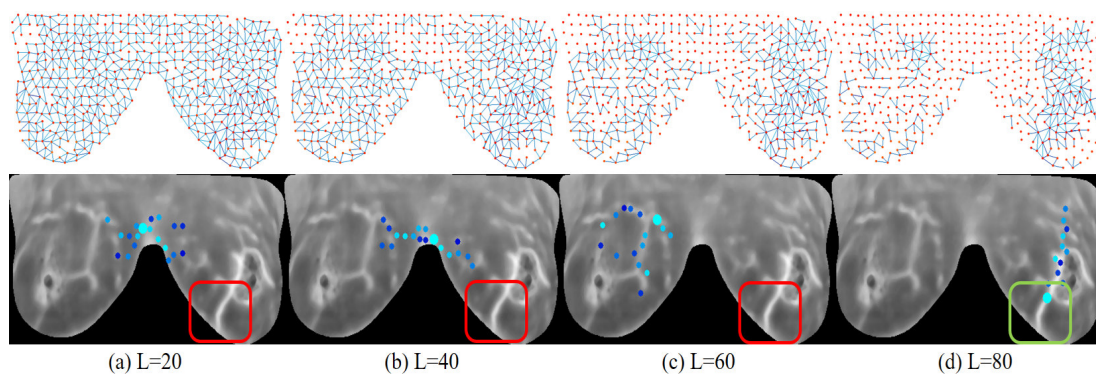


Figura 36 – Experimento III sobre o exame n° 281 com $S = 600$. Níveis de avaliação: (a) D; (b) D; (c) A; e (d) A.

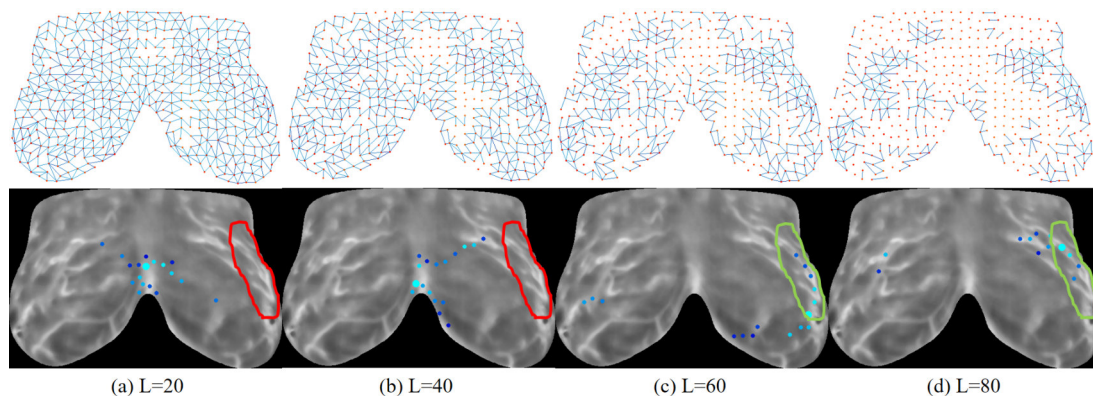
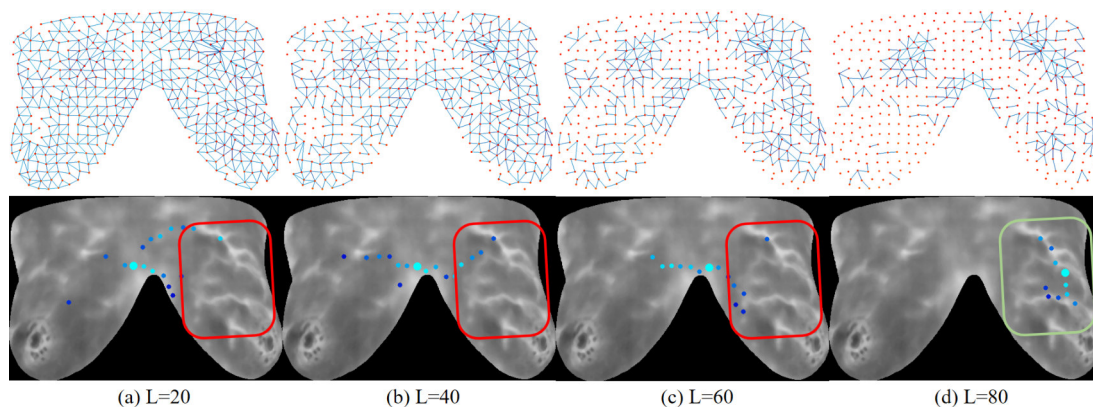


Figura 37 – Experimento III sobre o exame n° 210 com $S = 600$. Níveis de avaliação: (a) D; (b) D; (c) D; e (d) A.



5.4 Experimento IV

No quarto experimento, com uso da centralidade de grau, foram testadas todas as 25 combinações entre S e L , visto que apresentou resultados promissores em alguns testes iniciais. Pelos resultados listados na Tabela 12, nota-se que o experimento IV apresentou pequena melhoria em relação ao experimento II, ultrapassando os 50% para o nível A com $S = 200$ e $L = 40\%$. Contudo os valores de S para 400, 600 e 800 permitiram os melhores índices de acerto no nível de proximidade AUB, com 72,7%.

Tabela 12 – Resultados de avaliação do Experimento IV. Centralidade de grau.

modelagem inversa		avaliações em níveis				níveis agrupados (%)		
SLIC	Limiar L	A	B	C	D	A	AUB	AUBUC
S=100	0%	6	6	2	8	27,3	54,5	63,6
	20%	6	7	2	7	27,3	59,1	68,2
	40%	7	6	2	7	31,8	59,1	68,2
	60%	8	7	0	7	36,4	68,2	68,2
	80%	9	6	0	7	40,9	68,2	68,2
	acumulado (%)	32,7	29,1	5,5	32,7	32,7	61,8	67,3
S=200	0%	8	4	2	8	36,4	54,5	63,6
	20%	9	5	3	5	40,9	63,6	77,3
	40%	11	4	2	5	50,0	68,2	77,3
	60%	10	3	2	7	45,5	59,1	68,2
	80%	10	3	1	8	45,5	59,1	63,6
	acumulado (%)	43,6	17,3	9,1	30,0	43,6	60,9	70,0
S=400	0%	8	5	3	6	36,4	59,1	72,7
	20%	6	8	2	6	27,3	63,6	72,7
	40%	9	7	2	4	40,9	72,7	81,8
	60%	8	7	2	5	36,4	68,2	77,3
	80%	7	8	2	5	31,8	68,2	77,3
	acumulado (%)	34,5	31,8	1,0	23,6	34,5	66,4	76,4
S=600	0%	9	6	0	7	40,9	68,2	68,2
	20%	9	5	2	6	40,9	63,6	72,7
	40%	10	6	1	5	45,5	72,7	77,3
	60%	9	6	2	5	40,9	68,2	77,3
	80%	9	6	3	4	40,9	68,2	81,8
	acumulado (%)	41,8	26,4	7,3	24,5	41,8	68,2	75,5
S=800	0%	9	6	2	5	40,9	68,2	77,3
	20%	7	6	1	8	31,8	59,1	63,6
	40%	9	5	0	8	40,9	63,6	63,6
	60%	10	6	0	6	45,5	72,7	72,7
	80%	10	6	0	6	45,5	72,7	72,7
	acumulado (%)	40,9	26,4	2,7	30,0	40,9	67,3	70,0

Por meio das Figuras 38, 39 e 40 é possível identificar que há uma melhoria significativa na quantidade de acertos, em relação às mesmas imagens do experimento III. Nota-se também que essa característica elege, como o nó de maior centralidade, um nó

que se mantém o mesmo ao longo da variação do limiar. Na Figura 40 percebe-se que não há um acerto de nível A, mas há acertos de proximidade B, sinalizados pelo retângulo amarelo. Contudo, alterando o parâmetro S para 200, o mesmo exame n° 291 passa a ter todas as detecções avaliadas como corretas, conforme mostra a Figura 41.

Figura 38 – Experimento IV sobre o exame n° 281 com $S = 600$. Níveis de avaliação: (a) D; (b) A; (c) A; e (d) A.

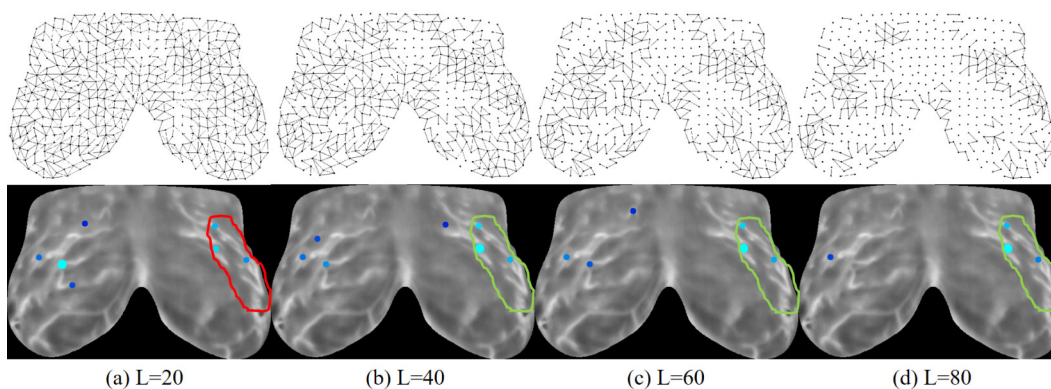


Figura 39 – Experimento IV sobre o exame n° 210 com $S = 600$. Níveis de avaliação: A para todos os limiares.

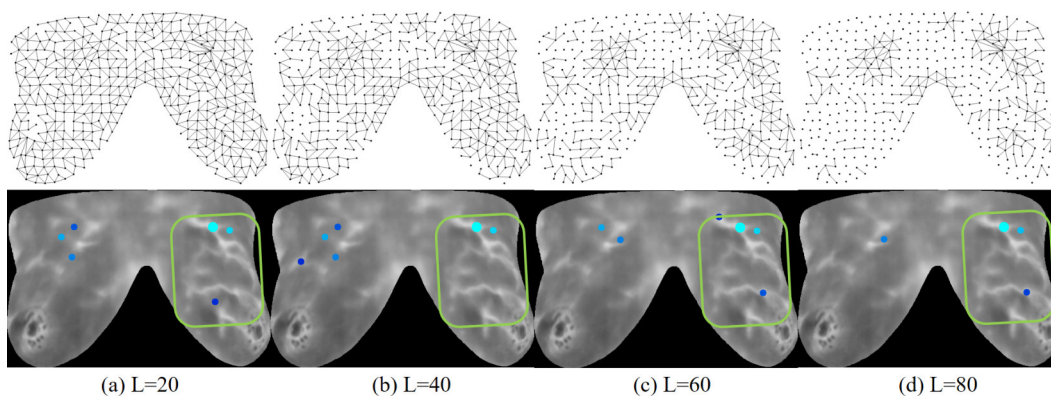


Figura 40 – Experimento IV sobre o exame n° 291 com $S = 600$. Níveis de avaliação: B para todos os limiares.

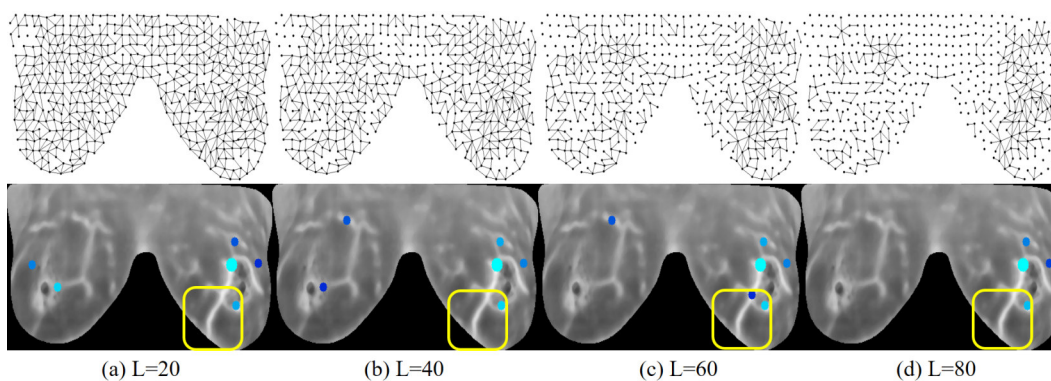
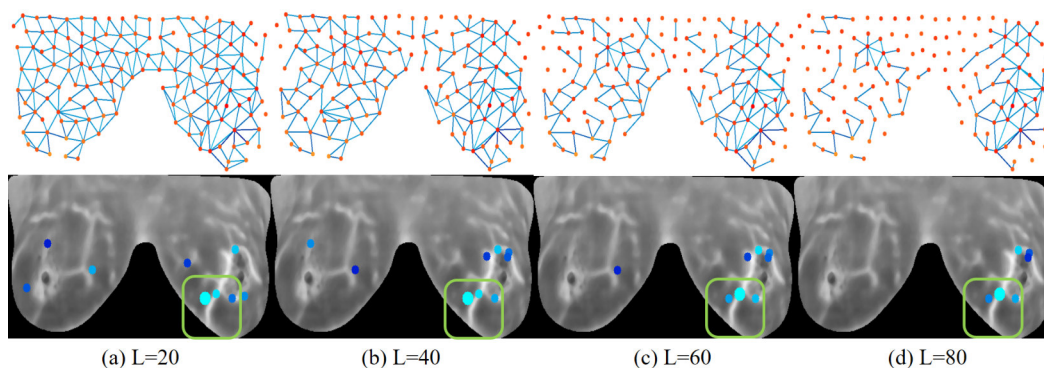


Figura 41 – Experimento IV sobre o exame n° 291 com $S = 200$. Níveis de avaliação: A para todos os limiares.



5.5 Experimento V

No quinto experimento foi aplicada a centralidade de proximidade. Também foram testadas todas as 25 combinações entre S e L , conforme mostra Tabela 13. Os resultados desse experimento superaram os do experimento anterior, chegando a 54,5% de avaliação de nível A, com $S = 100$ e $L = 80\%$. Com esses mesmos parâmetros, os outros níveis de proximidade também alcançaram os valores mais altos até então. Com 77,3% para os níveis AUB e AUBUC, nota-se que, de forma unânime, o limiar 80% apresentou os mais altos resultados para todas as variações de sementes do SLIC, nesse experimento. E também percebe-se que os dois menores limiares resultaram em valores de avaliação parecidos, com 0% para o nível A.

As Figuras 42, 43 e 44 ilustram bem essas considerações. Nelas é possível notar que houve acertos para o limiar 80%, ratificando o resultado da Tabela 13, que apresenta os maiores percentuais de acerto para o limiar 80%.

O experimento V, de forma geral, constata que a centralidade de proximidade é ideal para detectar os nós mais centrais em aspectos geográficos, ou seja, detectar os nós que estão no centro geográfico da rede. Assim, essa medida não se adequa bem ao propósito de detecção de tumores que podem estar em qualquer local das mamas.

Tabela 13 – Resultados de avaliação do Experimento V. Centralidade de proximidade.

modelagem inversa		avaliações em níveis				níveis agrupados (%)		
SLIC	Limiar	A	B	C	D	A	AUB	AUBUC
S=100	0%	0	1	4	17	0	4,5	22,7
	20%	0	2	3	17	0	9,1	22,7
	40%	7	6	2	7	31,8	59,1	68,2
	60%	8	5	1	8	36,4	59,1	63,6
	80%	12	5	0	5	54,5	77,3	77,3
	acumulado (%)	20,0	12,7	10,9	56,4	20,0	32,7	43,6
S=200	0%	0	1	2	19	0	4,5	13,6
	20%	0	1	2	19	0	4,5	13,6
	40%	4	2	0	16	18,2	27,3	27,3
	60%	3	4	4	11	13,6	31,8	50,0
	80%	10	5	1	6	45,5	68,2	72,7
	acumulado (%)	15,5	11,8	8,2	64,5	15,5	27,3	35,5
S=400	0%	0	1	2	19	0	4,5	13,6
	20%	0	1	2	19	0	4,5	13,6
	40%	1	3	2	16	4,5	18,2	27,3
	60%	6	7	0	9	27,3	59,1	59,1
	80%	9	4	1	8	40,9	59,1	63,6
	acumulado (%)	14,5	14,5	6,4	64,5	14,5	29,1	35,5
S=600	0%	0	2	1	19	0	9,1	13,6
	20%	0	2	1	19	0	9,1	13,6
	40%	1	2	1	18	4,5	13,6	18,2
	60%	6	6	2	8	27,3	54,5	63,6
	80%	9	7	1	5	40,9	72,7	77,3
	acumulado (%)	14,5	17,3	5,5	62,7	14,5	31,8	37,3
S=600	0%	0	2	1	19	0	9,1	13,6
	20%	0	2	1	19	0	9,1	13,6
	40%	1	3	2	16	4,5	18,2	27,3
	60%	7	6	1	8	31,8	59,1	63,6
	80%	9	6	1	6	40,9	68,2	72,7
	acumulado (%)	15,5	15,5	7,3	61,8	15,5	30,9	38,2

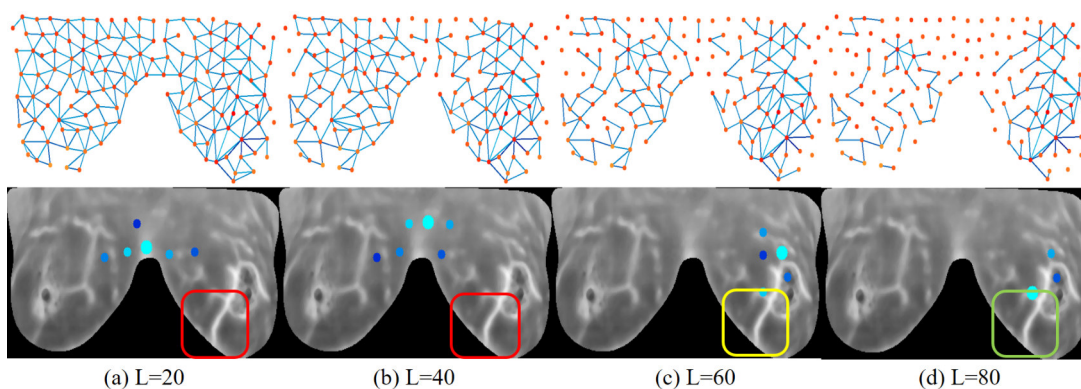
Figura 42 – Experimento V sobre o exame n° 291 com $S = 200$. Níveis de avaliação: (a) D; (b) D; (c) D; e (d) A.

Figura 43 – Experimento V sobre o exame n° 281 com $S = 200$. Níveis de avaliação: (a) D; (b) D; (c) D; e (d) A.

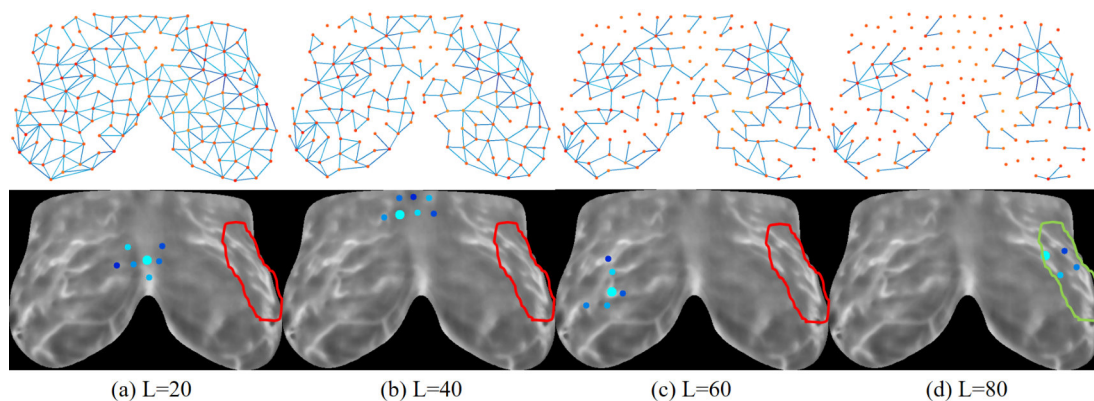
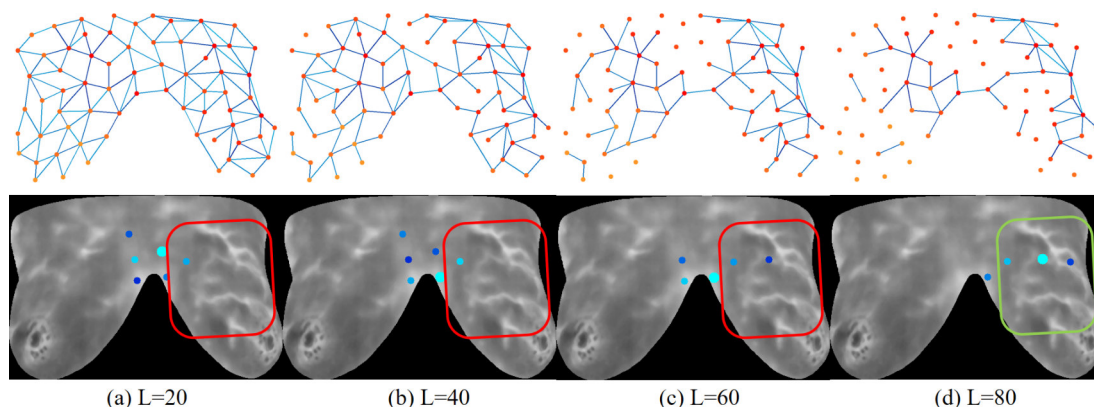


Figura 44 – Experimento V sobre o exame n° 210 com $S = 100$. Níveis de avaliação: (a) D; (b) D; (c) D; e (d) A.



5.6 Experimento VI

O sexto experimento faz o uso de centralidade de autovetor, na tentativa de medir a importância de um nó no grafo, levando em conta a influência de seus vizinhos. Na Tabela 14 são exibidos os resultados, onde o maior valor obtido foi 59,1% para o nível A, e 86,4% para os níveis AUB e AUBUC. Esses resultados foram para os parâmetros $S = 600$ e $L = 20$. Ainda com $S = 600$, percebe-se que com valores de L a 20 e 80% foi possível alcançar percentuais iguais para nível A. Contudo os resultados com 20% foram adotados como melhores pelo fato de fornecer um maior percentual de acertos dos níveis de avaliação agrupados AUB e AUBUC.

Considerando todos os melhores resultados para cada valor de S , nota-se que o mais alto resultado do experimento VI foi superior aos melhores resultados de todos os experimentos anteriores. Assim como foi mostrado nos experimentos anteriores, os

mesmos exames são usados como exemplo, por meio das Figuras 45, 46 e 47. Sendo que na Figura 47 percebe-se que nem todas as avaliações de detecção são de nível A. Contudo, com o parâmetro $S = 800$, o mesmo exame passa a ter todas as detecções avaliadas como corretas, conforme mostra a Figura 48.

No experimento VI, foi identificado que, para cada valor de S separadamente, os valores de L a 20% e a 40% se apresentaram como os maiores resultados para este experimento. Esse acontecimento anda em concordância com a medida de centralidade de autovetor, pois essa medida é muito afetada quando o corte de arestas é alto. Ressalta-se, também, que o acumulado percentual (que mede o percentual geral de acertos para cada nível de avaliação, considerando todos limiares para um mesmo valor S) é maior no experimento VI, em detrimento dos outros experimentos, mostrando que a centralidade de autovetor forneceu resultados menos variantes, considerando cada valor de limiar.

Figura 45 – Experimento VI sobre o exame n° 281 com $S = 600$. Níveis de avaliação: (a) B; (b) A; (c) A; e (d) A.

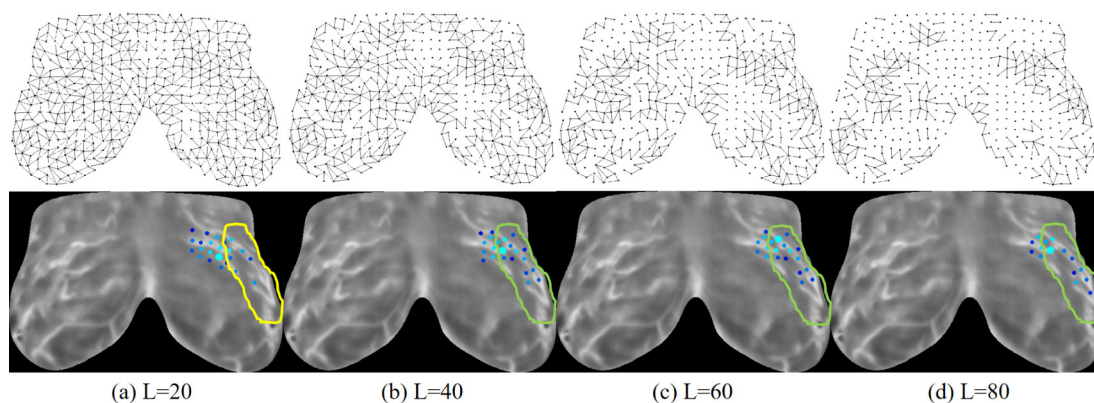


Figura 46 – Experimento VI sobre o exame n° 210 com $S = 600$. Níveis de avaliação: A para todos os limiares.

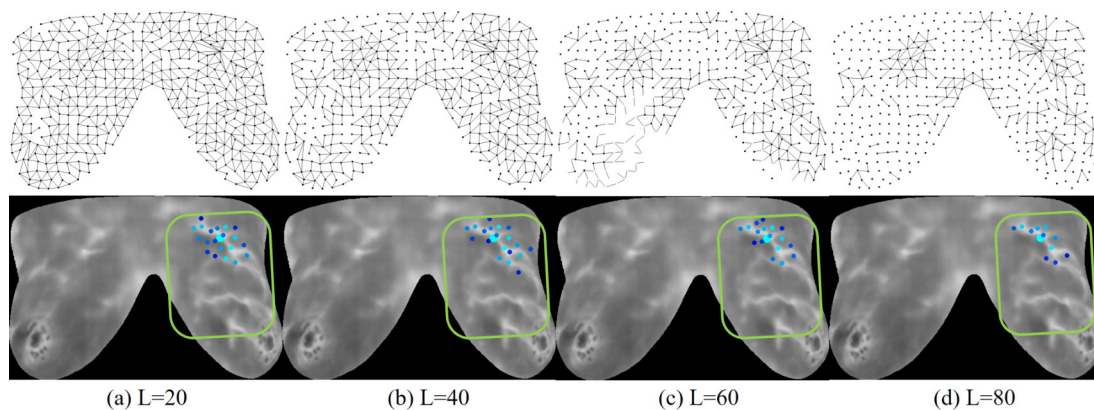


Figura 47 – Experimento VI sobre o exame n° 291 com $S = 600$. Níveis de avaliação: (a) A; (b) B; (c) B; e (d) B.

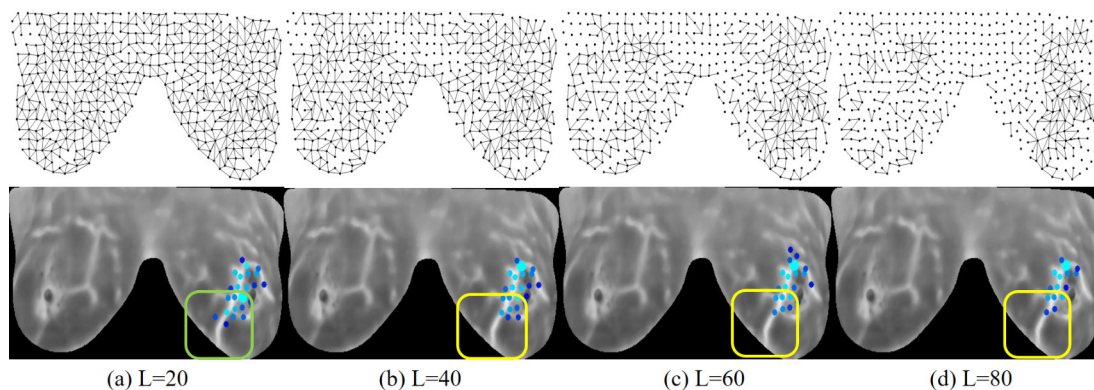
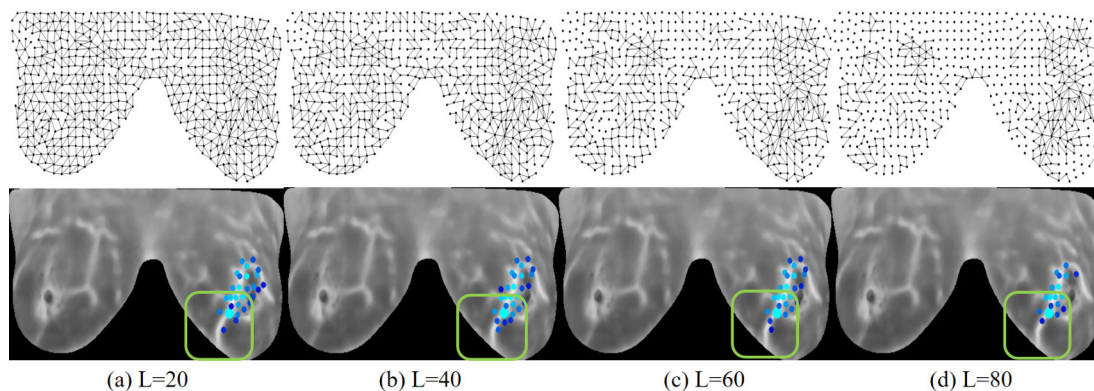


Figura 48 – Experimento VI sobre o exame n° 291 com $S = 800$. Níveis de avaliação: A para todos os limiares.



5.7 Discussões gerais acerca dos experimentos

Como a base não tem demarcação da localização exata, foi necessário desenvolver o método de avaliação utilizado. Essa avaliação é totalmente manual e depende da observação minuciosa por parte do avaliador, tornando a avaliação subjetiva, tanto na questão de divisão dos quadrantes, quando no julgamento dos níveis de acerto. Todos esses fatores representam uma desvantagem para o tipo de avaliação utilizada, contudo o avaliador procurou ser o mais justo possível, durante a aplicação do esquema de avaliação.

Um fator que dificultou a aplicação dos critérios de avaliação é a posição do mamilo. A maioria das mamas é de aspecto flácido e, como as temografias são capturadas com a paciente em pé, então os mamilos geralmente ficam localizados muito abaixo do ponto central da mama. Isso contribuiu para que os quadrantes inferiores sejam bem

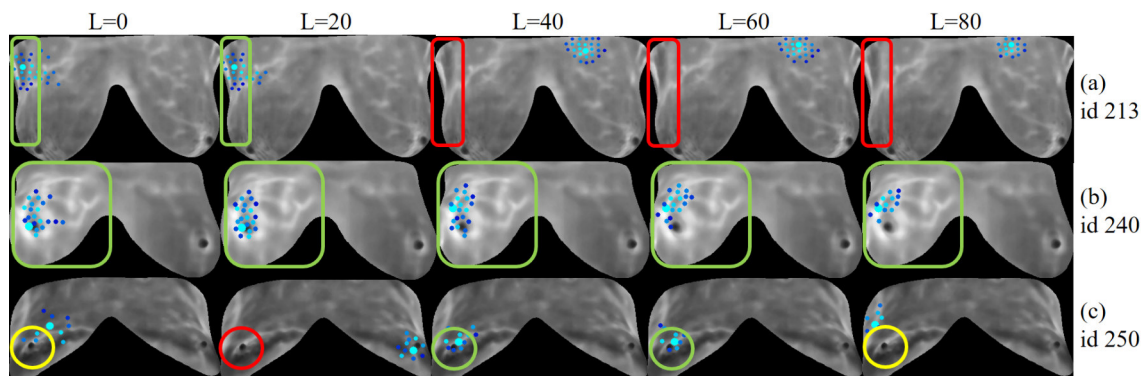
Tabela 14 – Resultados de avaliação do Experimento VI. Centralidade de autovetor.

modelagem inversa		avaliações em níveis				níveis agrupados (%)		
SLIC	Limiar	A	B	C	D	A	AUB	AUBUC
S=100	0%	8	4	2	8	36,4	54,5	63,6
	20%	12	3	2	5	54,5	68,2	77,3
	40%	13	2	2	5	59,1	68,2	77,3
	60%	12	5	1	4	54,5	77,3	81,8
	80%	11	4	2	5	50,0	68,2	77,3
	acumulado (%)	50,9	16,4	8,2	24,5	50,9	67,3	77,5
S=200	0%	9	3	1	9	40,9	54,5	59,1
	20%	10	5	1	6	45,5	68,2	72,7
	40%	12	3	2	5	54,5	68,2	77,3
	60%	11	5	1	5	50,0	72,7	77,3
	80%	11	4	1	6	50,0	68,2	72,7
	acumulado (%)	48,2	18,2	5,5	28,2	48,2	66,4	71,8
S=400	0%	11	5	1	5	50,0	72,7	77,3
	20%	11	5	1	5	50,0	72,7	77,3
	40%	12	4	1	5	54,5	72,7	77,3
	60%	11	5	1	5	50,0	72,7	77,3
	80%	11	5	1	5	50,0	72,7	77,3
	acumulado (%)	50,9	21,8	4,5	22,7	50,9	72,7	77,3
S=600	0%	12	5	0	5	54,5	77,3	77,3
	20%	13	6	0	3	59,1	86,4	86,4
	40%	12	5	0	5	54,5	77,3	77,3
	60%	12	5	0	5	54,5	77,3	77,3
	80%	13	5	0	4	59,1	81,8	81,8
	acumulado (%)	56,4	23,6	0	20,0	56,4	80,0	80,0
S=800	0%	12	4	0	6	54,5	72,7	72,7
	20%	13	4	0	5	59,1	77,3	77,3
	40%	12	3	0	7	54,5	68,2	68,2
	60%	12	4	0	6	54,5	72,7	72,7
	80%	12	4	0	6	54,5	72,7	72,7
	acumulado (%)	55,5	17,3	0	27,3	55,5	72,7	72,7

menores do que os superiores e, em alguns casos, alguns quadrantes inferiores têm tamanho desprezível. A Figura 49 (a) ilustra um caso desse tipo, com o exame de nº 213.

A Figura 49 (b) mostra o resultado de detecção do exame nº 240, que é diagnosticado com carcinoma ductal infiltrante, apresentando localização em toda mama esquerda. Já a Figura 49 (c) apresenta o caso nº 250, em que a anomalia está localizada no quadrante central da mama, o que exige um esquema de classificação diferenciado, relatado na Seção 4.2.6.

Figura 49 – Experimento VI sobre vários exames com $S = 600$. Níveis de avaliação: A = verde, B = amarelo, D = vermelho.



Analisando os resultados obtidos em cada um dos experimentos é possível observar que o melhor resultado para os níveis secundários de acerto (AUB ou AUBUC) foi expressivo, significando que a metodologia conseguiu acertar a mama com anormalia, a uma taxa satisfatória de 86,4%, apesar de não ter exatidão de detecção (em nível A). A Tabela 15 mostra que se, no melhor dos casos, fosse feita aposta de que todas as anomalias estivessem presentes na mama direita, o percentual de acerto seria de 60,4%. Entretanto a metodologia proporcionou uma taxa de acerto de 86,4% em relação ao nível de AUB, conforme mostra a Tabela 16.

Tabela 15 – Percentual de distribuição de anomalia entre os quadrantes das termografias utilizadas.

	quadrante	(%)	total (%)
Mama Direita	QSE	18,9	60,4
	QSI	18,9	
	QIE	11,3	
	QII	11,3	
Mama Esquerda	QSE	15,1	39,6
	QSI	11,3	
	QIE	5,7	
	QII	7,5	

** os QCs foram contados como QSE, QSI, QIE e QII, ao mesmo tempo.*

A probabilidade de acertar um dos quadrantes de forma aleatória seria $1/8$ (12,5%). Mas, segundo relatado na literatura médica, como a chance de o tumor aparecer no quadrante superior externo é de 50%, então caso seja escolhido apenas um QSE entre as duas mamas, a chance de acerto é de 25%. Levando em conta as estatísticas de localização do tumor, mostrada na Tabela 15, se for feita aposta de que todos os exames possuem anomalia no QSE-MD, o percentual de acerto será de apenas 18%, um valor ainda bem

abaixo de 36,4%, apresentado no experimento I, e mais longe ainda dos experimentos IV a VI, que passaram de 50%.

Tabela 16 – Resumo dos melhores resultados de avaliação de todos os experimentos.

Experim.	Medidas	S	$L(\%)$	níveis agrupados (%)		
				A	AUB	AUBUC
I	grau máx.	600	40	36,4	63,6	68,2
II	núm. de nós	600	60	45,5	54,5	54,5
III	cent. de grau	600	80	31,8	50,0	59,1
IV	cent. de intermed.	200	40	50,0	68,2	77,3
V	cent. de proximid.	100	80	54,5	77,3	77,3
VI	cent. de autovetor	600	20	59,1	86,4	86,4

Ressalta-se que, durante a seleção de imagens da base, muitas termografias que visualmente continham uma clara localização tiveram que ser descartadas, porque não tinham boa descrição da sua localização definida por diagnóstico. Tais imagens poderiam aumentar a taxa de acertos dessa metodologia.

Dentre os trabalhos elencados na Seção 2, os trabalhos de Alves (2015) e Andrade et al. (2017) foram os que tiveram alguma contribuição em relação à indicação de qual mama continha anomalia. Esses trabalhos possuem o objetivo de detecção da presença de anomalia, e não a detecção da região, onde são analisadas as mamas esquerda e direita como se fossem exames diferentes. Então, o grau de aproximação de indicação da área suspeita desses trabalhos equivale ao nível C de aproximação de detecção do presente trabalho. Por outro lado, a metodologia aqui apresentada ainda oferece mais dois níveis de detecção (A e B) de maior aproximação. A Tabela 17 esquematiza essa comparação, onde o nível agrupado AUBUC foi comparado ao índice de Sensibilidade dos outros trabalhos que tinham objetivo de classificação.

Tabela 17 – Trabalhos que contribuem com detecção da região da anomalia (em algum dos níveis de aproximação) em comparação com o presente trabalho.

Trabalho	Objetivo	níveis de aproximação de detecção (%)		
		A	AUB	AUBUC
Alves (2015)	Classificação	-	-	100
Andrade et al. (2017)	Classificação	-	-	93,6
Presente Trabalho	Detecção	59,1	86,4	86,4

6 CONCLUSÕES

Esta dissertação apresentou uma metodologia de detecção de regiões suspeitas de anomalias, em termografias de mama, a partir da modelagem das regiões em forma de grafo e posterior análise de medidas de redes complexas. O presente trabalho também possibilitou um complemento a esta área de pesquisa, pois apresentou a capacidade de fornecer, às metodologias de classificação existentes, uma indicação sugestiva da regiões com suspeita de anomalia.

O fato de o câncer de mama ser um problema de saúde pública justifica os esforços ao redor do mundo para a obtenção de métodos de detecção mais robustos. Além disso, a mama tem como função principal a produção do leite para a amamentação, sem falar da importância psicológica, representando papel essencial na formação da autoestima da mulher.

Com a ascensão dos métodos de diagnóstico por termografia de mama, espera-se que se torne um exame mais difundido. Pois, além de suas vantagens agregadas (ser mais barato, não invasivo, ter repetição ilimitada, etc), a termografia pode se tornar uma ferramenta no auxílio da detecção de anomalias em mama, de forma complementar aos outros exames consagrados. Tudo isso para aumentar a taxa de detecção do câncer em estágios iniciais, aumentando a chance de sobrevivência dos pacientes.

Como parte dos objetivos, foi desenvolvida uma modelagem de grafo a partir de imagens utilizando: técnicas de pré-processamento de imagens; técnicas de segmentação de regiões com agrupamento de pixels; e construção de representação de imagem por meio de grafo. Através de investigação de trabalhos relacionados na literatura, pôde-se aplicar técnicas de análise de grafo pertinentes ao objetivo do trabalho. E, por meio de exaustivos experimentos, foi possível explorar a influência de diversas medidas na realização de uma análise que trouxesse bons resultados de detecção.

Foi desenvolvido um método de avaliação de detecção baseado no níveis de proximidade em termos de quadrante. Com o método de avaliação o avaliador foi capaz de confrontar a região de detecção com a região de diagnóstico fornecido pela base de termografias, sendo possível validar e medir a performance desta metodologia.

Por se tratar de uma abordagem que usa modelagem de grafo, a representação de regiões superpixels destacou as formas e valorizou as correlações geográficas das regiões

da termografia, além de reduzir a complexidade computacional, caso cada pixel fosse representado por um nó de grafo.

Baseado nos pontos discutidos e nos resultados apresentados, pode-se concluir que a metodologia forneceu resultados que ainda precisam ser melhorados para detecção da região exata das anomalias, com percentual de 59,1%. No entanto a mesma propiciou resultados expressivos e promissores, no sentido de indicar em qual das mamas há anomalia, ou a localização aproximada da anomalia, com percentual de acerto em 86,4%.

Muito ainda pode ser explorado com relação à linha de trabalhos que usam modelagem em grafo para processamento de imagem combinado com detecção ou segmentação de regiões. Quanto à detecção de termografia de mama, há a necessidade de elaboração de métodos que aumentem a eficiência da detecção. Estão listadas, a seguir, algumas sugestões de trabalhos futuros.

- Elaborar outras estratégias de modelagem de grafo, como, por exemplo, uma modelagem que permita ligação de nós não adjacentes, com a intenção de estabelecer relacionamentos entre regiões mais distantes umas das outras;
- Realizar testes e ajustes na metodologia para incorporar outros tipos técnicas de análises de grafo;
- Testar a metodologia proposta com bases de demarcação precisa, com formas e limites bem definidos, com intuito de permitir uma avaliação com alta precisão numérica e que possa ser aferida pelas métricas convencionalmente utilizadas;
- Testar a metodologia proposta com outras bases de termografia de mama;
- Testar outras técnicas de detecção não necessariamente relacionadas com grafo como, por exemplo, reconhecimento de padrões;
- Aplicar essa metodologia a problemas de outras áreas relacionadas a exames médicos.

REFERÊNCIAS

- ACHANTA, R.; SHAJI, A.; SMITH, K.; LUCCHI, A.; FUA, P.; SÜSSTRUNK, S. Finding objects of interest in images using saliency and superpixels. **EPFL**, Lausanne, p. 183, 2011. Disponível em: <<http://infoscience.epfl.ch/record/153491>>.
- ALVES, T. E. S. **Uma Metodologia de Auxílio ao Diagnóstico de Doenças de Mama a partir de Termografias Dinâmicas**. Tese (Doutorado) — PhD thesis, Universidade Federal Fluminense, Instituto de Computação, Niterói, RJ, Brasil, 2015.
- AMRI, A.; PULKO, S. H.; WILKINSON, A. J. Potentialities of steady-state and transient thermography in breast tumour depth detection: A numerical study. **Computer methods and programs in biomedicine**, Elsevier, v. 123, p. 68–80, 2016.
- ANBAR, M.; MILESCU, L.; NAUMOV, A.; BROWN, C.; BUTTON, T.; CARLY, C.; ALDULAIMI, K. Detection of cancerous breasts by dynamic area telethermometry. **IEEE Engineering in Medicine and Biology Magazine**, IEEE, v. 20, n. 5, p. 80–91, 2001.
- ANDRADE, F.; PAIVA, A.; SILVA, A. C. Análise de imagens de termografia dinâmica para classificação de alterações na mama usando séries temporais. In: **Workshop of Undergraduate Works (WUW) in the 30th Conference on Graphics, Patterns and Images (SIBGRAPI'17)**. Niterói, RJ, Brazil: SIBGRAPI, 2017. Disponível em: <<http://sibgrapi2017.ic.uff.br/>>.
- AUSTIN, D. How google finds your needle in the web's haystack. **American Mathematical Society Feature Column**, v. 10, n. 12, 2006. Disponível em: <<http://www.ams.org/publicoutreach/feature-column/fcarc-pagerank>>.
- BACKES, A. R.; CASANOVA, D.; BRUNO, O. M. A complex network-based approach for boundary shape analysis. **Pattern Recognition**, v. 42, n. 1, p. 54 – 67, 2009. ISSN 0031-3203. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0031320308002793>>.
- BARABÁSI, A.; ALBERT, R. Emergence of scaling in random networks. **Science**, American Association for the Advancement of Science, v. 286, n. 5439, p. 509–512, 1999.
- BERNARDES, A. Anatomia da mama feminina. **Manual de Ginecologia**, Permanyer Portugal, Lisboa, Portugal, v. 2, n. 12, p. 12–24, 2011.
- BEZERRA, L.; OLIVEIRA, M.; ARAÚJO, M.; VIANA, M. J.; SANTOS, L. C.; SANTOS, F.; ROLIM, T.; LYRA, P.; LIMA, R. C.; BORCHARTT, T. B.; RESMINI, R.; CONCI, A. Infrared imaging for breast cancer detection with proper selection of properties: From acquisition protocol to numerical simulation. In: **Multimodality Breast Imaging: Diagnosis and Treatment**. Washington, USA: Ng, E.Y.K; Acharya, U.R.; Rangayyan, R.M.; Suri, J.S., 2013. PM227, p. 285–322. ISBN 978-0-8194-9294-4.
- BIAZÚS, J. V.; MELO, M. P. de; ZUCATTO, Â. E. **Cirurgia da mama**. Porto Alegre, Rio Grande do Sul, Brasil: Artmed Editora, 2009.
- BIOLCHINI, J.; MIAN, P. G.; NATALI, A. C. C.; TRAVASSOS, G. H. Systematic review in software engineering. **System Engineering and Computer Science Department COPPE/UFRJ, Technical Report ES**, v. 679, n. 05, p. 45, 2005. Disponível em: <<http://www.cos.ufrj.br/uploadfile/es67905.pdf>>.

- BONDY, J. A.; MURTY, U. S. R. et al. **Graph theory with applications**. Citeseer, 1976. v. 290. Disponível em: <<http://www.freetechbooks.com/graph-theory-with-applications-t559.html>>.
- BORCHARTT, T. B.; CONCI, A.; LIMA, R. C.; RESMINI, R.; SANCHEZ, A. Breast thermography from an image processing viewpoint: A survey. **Signal Processing**, v. 93, n. 10, p. 2785–2803, 2013. ISSN 0165-1684. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0165168412002794>>.
- BRANDES, U. A faster algorithm for betweenness centrality. **The Journal of Mathematical Sociology**, Routledge, v. 25, n. 2, p. 163–177, 2001. Disponível em: <<https://doi.org/10.1080/0022250X.2001.9990249>>.
- BRAY, F.; FERLAY, J.; SOERJOMATARAM, I.; SIEGEL, R. L.; TORRE, L. A.; JEMAL, A. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: a cancer journal for clinicians**, Wiley Online Library, Hoboken, Nova Jersey, EUA, v. 68, n. 6, p. 394–424, 2018. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/30207593>>.
- BRUENING, W.; UHL, S.; FONTANAROSA, J.; RESTON, J.; TREADWELL, J.; SCHOELLES, K. Noninvasive diagnostic tests for breast abnormalities: update of a 2006 review. Agency for Healthcare Research and Quality (US), Rockville (MD), 2012. Disponível em: <www.ncbi.nlm.nih.gov/books/NBK84530/>.
- BUITRAGO, F.; UEMURA, G.; SENA, M. C. F. **Fatores prognósticos no câncer da mama feminina**. 2002. Disponível em: <http://bvsmms.saude.gov.br/bvs/artigos/fatores_prognosticos.pdf>.
- CHALUMEAU, T.; COSTA, L. d. F.; LALIGANT, O.; MERIAUDEAU, F. Texture discrimination using hierarchical complex networks. In: **Signal Processing for Image Enhancement and Multimedia Processing**. New York City, EUA: Springer, 2008. p. 95–102.
- CHERNOV, G.; CHERNOV, V.; FLORES, M. B. 3d dynamic thermography system for biomedical applications. In: **Application of Infrared to Biomedical Sciences**. [S.l.]: Springer, 2017. p. 517–545.
- COSTA, L. d. F.; ROCHA, L. E. C. da. A generalized approach to complex networks. **The European Physical Journal B-Condensed Matter and Complex Systems**, Springer, v. 50, n. 1-2, p. 237–242, 2006.
- COSTA, L. d. F.; RODRIGUES, F. A.; TRAVIESO, G.; BOAS, P. R. V. Characterization of complex networks: A survey of measurements. **Advances in physics**, Taylor & Francis, v. 56, n. 1, p. 167–242, 2007.
- DANON, L.; DIAZ-GUILERA, A.; DUCH, J.; ARENAS, A. Comparing community structure identification. **Journal of Statistical Mechanics: Theory and Experiment**, IOP Publishing, v. 2005, n. 09, p. P09008, 2005. Disponível em: <<https://iopscience.iop.org/article/10.1088/1742-5468/2005/09/P09008/meta>>.
- ERDON, P.; RENYI, A. On random graphs. **Publ. Math. Debrecen. Publicationes Mathematicae**, v. 6, p. 290–297, 1959.

FLIR, S. **Sc620 infrared camera system**. Flir - Technical Specifications, 2015. Disponível em: <<http://www.ir.com/cs/apac/en/view/?id=41413>>.

_____. **3 principais aplicações da análise termográfica**. Flir - Technical Specifications, 2017. Disponível em: <<http://www.flir.com.br/home/news/details/?ID=80767>>.

FOLKMAN, J. Tumor angiogenesis: therapeutic implications. **New england journal of medicine**, Mass Medical Soc, v. 285, n. 21, p. 1182–1186, 1971.

FREEMAN, L. A set of measures of centrality based on betweenness. **Sociometry**, v. 40, p. 35–41, 03 1977.

GARDUÑO-RAMÓN, M. A.; VEGA-MANCILLA, S. G.; MORALES-HENÁNDEZ, L. A.; OSORNIO-RIOS, R. A. Supportive noninvasive tool for the diagnosis of breast cancer using a thermographic camera as sensor. **Sensors**, v. 17, n. 3, 2017. ISSN 1424-8220. Disponível em: <<http://www.mdpi.com/1424-8220/17/3/497>>.

GERASIMOVA, E.; AUDIT, B.; ROUX, S.; KHALIL, A.; ARGOUL, F.; NAIMARK, O.; ARNEODO, A. Multifractal analysis of dynamic infrared imaging of breast cancer. **EPL (Europhysics Letters)**, IOP Publishing, v. 104, n. 6, p. 68001, 2013.

GERASIMOVA, E.; AUDIT, B.; ROUX, S. G.; KHALIL, A.; GILEVA, O.; ARGOUL, F.; NAIMARK, O.; ARNEODO, A. Wavelet-based multifractal analysis of dynamic infrared thermograms to assist in early breast cancer diagnosis. **Frontiers in physiology**, Frontiers Media SA, v. 5, 2014.

GERASIMOVA, E.; PLEKHOV, O.; BAYANDIN, Y.; NAIMARK, O.; FREYND, G. Identification of breast cancer using analysis of thermal signals by nonlinear dynamics methods. **International Conference on Quantitative InfraRed Thermography**, 2012.

GONZALEZ, R. C.; WOODS, R. E. **Digital image processing**. London, England: Pearson, 2012.

GUERRIERO, V. Power law distribution: Method of multi-scale inferential statistics. **Journal of Modern Mathematics Frontier**, Science and Engineering Publishing Company, v. 1, n. 1, p. 21–28, 2012.

GUTIERREZ-DELGADO, F. et al. Feasibility of new-generation infrared imaging screening for breast cancer in rural communities. **International Health**, 2011.

GUTMAN, I.; ZHOU, B. Laplacian energy of a graph. **Linear Algebra and its applications**, Elsevier, v. 414, n. 1, p. 29–37, 2006. Disponível em: <<https://aimath.org/WWN/matrixspectrum/gutman2006.pdf>>.

HARCHAOUI, Z.; BACH, F. Image classification with segmentation graph kernels. In: IEEE. **Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on**. 2007. p. 1–8. ISSN 1063-6919. Disponível em: <<https://ieeexplore.ieee.org/document/4270074>>.

HERMAN, C. The role of dynamic infrared imaging in melanoma diagnosis. **Expert review of dermatology**, Taylor & Francis, v. 8, n. 2, p. 177–184, 2013.

HILDEBRANDT, C.; RASCHNER, C.; AMMER, K. An overview of recent application of medical infrared thermography in sports medicine in austria. **Sensors**, Molecular Diversity Preservation International, v. 10, n. 5, p. 4700–4715, 2010.

HORTOLAM, J.; JALES, R. M. **Esteatonecrose em exames por imagem da mama**. 2016. Disponível em: <<https://www.fcm.unicamp.br/drpixel/conteudo/esteatonecrose-em-exames-por-imagem-da-mama>>.

IBCC. **O câncer de mama no alvo da moda - Exames**. Instituto Brasileiro de Controle do Câncer, 2019. Disponível em: <<http://www.ibcc.org.br/exames/>>.

INCA. **Estimativa 2018: Incidência de Câncer no Brasil**. Rio de Janeiro, Brasil: Instituto Nacional do Câncer, 2019. Disponível em: <<https://www.inca.gov.br/publicacoes/livros/estimativa-2018-incidencia-de-cancer-no-brasil>>.

INCA. **Tipos de Câncer**. Rio de Janeiro, Brasil: Instituto Nacional do Câncer, 2019. Disponível em: <<https://www.inca.gov.br/tipos-de-cancer/cancer-de-mama>>.

Jianbo Shi; Malik, J. Normalized cuts and image segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22, n. 8, p. 888–905, Aug 2000. ISSN 0162-8828. Disponível em: <<https://ieeexplore.ieee.org/document/868688>>.

JORGE-HERNANDEZ, F.; CHIMENO, Y. G.; GARCIA-ZAPIRAIN, B.; ZUBIZARRETA, A. C.; BELDARRAIN, M. A. G.; FERNANDEZ-RUANOVA, B. Graph theory for feature extraction and classification: A migraine pathology case study. **Bio-medical materials and engineering**, IOS Press, v. 24, n. 6, p. 2979–2986, 2014.

KENNEDY, D. A.; LEE, T.; SEELY, D. A comparative review of thermography as a breast cancer screening technique. **Integrative cancer therapies**, Sage Publications Sage CA: Los Angeles, CA, v. 8, n. 1, p. 9–16, 2009.

KEYSERLINGK, J.; AHLGREN, P.; YU, E.; BELLIVEAU, N.; YASSA, M. Functional infrared imaging of the breast. **IEEE Engineering in Medicine and Biology Magazine**, IEEE, v. 19, n. 3, p. 30–41, 2000.

KITCHENHAM, B. Procedures for performing systematic reviews. **Keele, UK, Keele University**, v. 33, n. 2004, p. 1–26, 2004. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.3308&rep=rep1&type=pdf>>.

KOSCHÜTZKI, D.; LEHMANN, K. A.; PEETERS, L.; RICHTER, S.; TENFELDE-PODEHL, D.; ZLOTOWSKI, O. Centrality indices. In: **Network analysis**. Springer, 2005. p. 16–61. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-540-31955-9_3>.

LAWSON, R. N.; CHUGHTAI, M. Breast cancer and body temperature. **Canadian Medical Association Journal**, Canadian Medical Association, v. 88, n. 2, p. 68, 1963.

LEVINSHTEIN, A.; STERE, A.; KUTULAKOS, K. N.; FLEET, D. J.; DICKINSON, S. J.; SIDDIQI, K. Turbopixels: Fast superpixels using geometric flows. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 31, n. 12, p. 2290–2297, 2009. Disponível em: <[Turbopixels:Fastsuperpixelsusinggeometricflows](#)>.

Levinshtein, A.; Stere, A.; Kutulakos, K. N.; Fleet, D. J.; Dickinson, S. J.; Siddiqi, K. Turbopixels: Fast superpixels using geometric flows. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 31, n. 12, p. 2290–2297, Dec 2009. ISSN 0162-8828.

LÉZORAY, O.; GRADY, L. et al. Graph theory concepts and definitions used in image processing and analysis. **Image processing and analysing with graphs: theory and practice**. CRC Press, Boca Raton, FL, p. 1–24, 2012. Disponível em: <Graphtheoryconceptsanddefinitionsusedinimageprocessingandanalysis>.

LUNA, J.; DELGADO, F. Feasibility of new-generation infrared imaging screening for breast cancer in rural communities. **US Obstetrics & Gynecology**, v. 1, p. 52–55, 2010.

MAHMOUDI, A.; TAKERKART, S.; REGRAGUI, F.; BOUSSAOUD, D.; BROVELLI, A. Multivoxel pattern analysis for fmri data: a review. **Computational and mathematical methods in medicine**, Hindawi, v. 2012, 2012. Disponível em: <<https://www.hindawi.com/journals/cmmm/2012/961257/>>.

MATCOVSCHI, M.; PASTRAVANU, O. Perron-frobenius theorem and invariant sets in linear systems dynamics. In: **2007 Mediterranean Conference on Control Automation**. Atenas, Grécia: Mediterranean Conference on Control and Automation, 2007. p. 1–6.

METZ, J.; CALVO, R.; SENO, E. R.; ROMERO, R. A.; LIANG, Z. Redes complexas: conceitos e aplicações. **Relatórios Técnicos do ICMC-USP São Carlos**, 2007.

MINOLTA, K. **Compreendendo o Espaço de Cor CIE L C h**. Konica Minolta Sensing Americas, 2006. Disponível em: <<http://sensing.konicaminolta.com.br/2015/08/compreendendo-o-espaco-de-cor-cie-lch/>>.

Montanvert, A.; Meer, P.; Rosenfield, A. Hierarchical image analysis using irregular tessellations. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 13, n. 4, p. 307–316, April 1991. ISSN 0162-8828. Disponível em: <<https://ieeexplore.ieee.org/document/88566>>.

MONTEIRO, A. P. d. S.; ARRAES, E. P. P.; PONTES, L. d. B.; CAMPOS, M. d. S. d. S.; RIBEIRO, R. T.; GONÇALVES, R. E. B. Auto-exame das mamas: frequência do conhecimento, prática e fatores associados. **Rev Bras Ginecol Obstet**, v. 25, n. 3, p. 201–205, 2003. Disponível em: <http://www.jornaldosamigos.com.br/cancer_de_mama2.htm>.

MOORE, K. L.; DALLEY, A. F.; AGUR, A. M. **Anatomia orientada para a clínica**. Rio de Janeiro, Brasil: Guanabara koogan, 2006.

NEWMAN, M. E. The structure and function of complex networks. **SIAM review**, SIAM, v. 45, n. 2, p. 167–256, 2003. Disponível em: <<https://epubs.siam.org/doi/10.1137/S003614450342480>>.

_____. A measure of betweenness centrality based on random walks. **Social networks**, Elsevier, v. 27, n. 1, p. 39–54, 2005. Disponível em: <<https://arxiv.org/abs/cond-mat/0309045>>.

NG, E.-K. A review of thermography as promising non-invasive detection modality for breast tumor. **International Journal of Thermal Sciences**, Elsevier, Amsterdã, Países Baixos, v. 48, n. 5, p. 849–859, 2009.

OPSAHL, T.; AGNEESSENS, F.; SKVORETZ, J. Node centrality in weighted networks: Generalizing degree and shortest paths. **Social networks**, Elsevier, v. 32, n. 3, p. 245–251, 2010. Disponível em: <<https://toreopsahl.com/2010/04/21/article-node-centrality-in-weighted-networks-generalizing-degree-and-shortest-paths/>>.

PENG, L.; CHEN, W.; ZHOU, W.; LI, F.; YANG, J.; ZHANG, J. An immune-inspired semi-supervised algorithm for breast cancer diagnosis. **Computer methods and programs in biomedicine**, Elsevier, Amsterdã, Países Baixos, v. 134, p. 259–265, 2016.

PEREIRA, K. F.; SILVA, C. D. V.; CASTILHO, T. J. de; SILVA, L. V. da; RIBEIRO, R.; LINHARES, J. C.; HATSCHBACH, S. B. B. Tumor fíloides de mama: série de casos e revisão de literatura. **Rev Bras Mastologia**, v. 22, n. 2, p. 42–45, 2012. Disponível em: <http://www.mastology.org/wp-content/uploads/2015/06/MAS_v22n2_42-45.pdf>.

PIZER, S. M.; JOHNSTON, R. E.; ERICKSEN, J. P.; YANKASKAS, B. C.; MULLER, K. E. Contrast-limited adaptive histogram equalization: speed and effectiveness. In: IEEE. **Visualization in Biomedical Computing, 1990., Proceedings of the First Conference on**. Nova Jersey, EUA, 1990. p. 337–345.

RESMINI, R. **Análise de imagens térmicas da mama usando descritores de textura**. Tese (Doutorado), 2011.

RIYAH, S.; AGOSTINI, V.; MOLINARI, F.; KNAFLITZ, M. et al. Comparison of time-series registration methods in breast dynamic infrared imaging. **Opto-Electronics Review**, De Gruyter Open, v. 23, n. 1, p. 68–77, 2015.

ROCCHINI, C. **Centralidade**. Wikipedia, 2017. Disponível em: <<https://pt.wikipedia.org/wiki/Centralidade>>.

ROCHA, R. M. **Quais as diferenças entre os tumores benigno e maligno?** Saúde Abril, 2013. Disponível em: <<https://saude.abril.com.br/bem-estar/quais-as-diferencas-entre-os-tumores-benigno-e-maligno/>>.

RODRIGUES, F. A. **Caracterização, Classificação e Análise de Redes Complexas**. Universidade de São Paulo. Instituto de Física de São Carlos. Tese (Doutorado) — USP, 2007. Disponível em: <<http://conteudo.icmc.usp.br/pessoas/francisco/works/tese.pdf>>.

SABIDUSSI, G. The centrality index of a graph. **Psychometrika**, v. 31, n. 4, p. 581–603, 1966. Disponível em: <[https://link.springer.com/article/10.1007%2F02289527?LI=true](https://link.springer.com/article/10.1007%2F2F02289527?LI=true)>.

SAMET, H. The quadtree and related hierarchical data structures. **ACM Comput. Surv.**, ACM, New York, NY, USA, v. 16, n. 2, p. 187–260, jun. 1984. ISSN 0360-0300. Disponível em: <<http://doi.acm.org/10.1145/356924.356930>>.

SANIEI, E.; SETAYESHI, S.; AKBARI, M. E.; NAVID, M. A vascular network matching in dynamic thermography for breast cancer detection. **Quantitative InfraRed Thermography Journal**, Taylor & Francis, v. 12, n. 1, p. 24–36, 2015.

SATTAROV, A.; MCINTYRE, P.; MOTOWIDLO, L. High-field open MRI for breast cancer screening. **IEEE Transactions on Applied Superconductivity**, v. 25, n. 3, p. 1–5, June 2015. ISSN 1051-8223.

SEDGEWICK, R.; WAYNE, K. **Algorithms**. 4th. ed. Londres, Inglaterra: Addison-Wesley Professional, 2011. ISBN 032157351X, 9780321573513.

SENGUPTA, S.; WEBB, D. R. Use of a computer-controlled bipolar diathermy system in radical prostatectomies and other open urological surgery. **ANZ journal of surgery**, Wiley Online Library, Hoboken, Nova Jersey, EUA, v. 71, n. 9, p. 538–540, 2001.

SERPEJANTE, C. **Mamografia: exame detecta o câncer de mama**. Minha Vida, 2014. Disponível em: <<https://www.minhavidade.com.br/saude/tudo-sobre/16864-mamografia>>.

SERRANO, R. **Análise da viabilidade do uso do Coeficiente de Hurst e da Lacunaridade no auxílio ao diagnóstico precoce de patologias da mama**. Tese (Doutorado) — Dissertação de Mestrado, Instituto de Computação, Universidade Federal, 2010.

SILVA, L. F. **Uma Análise Híbrida para Detecção ao de Anomalias da Mama usando Séries Temporais de Temperatura**. Tese (Doutorado) — PhD thesis, Universidade Federal Fluminense, Instituto de Computação, Niterói, RJ, Brasil, 2015.

SILVA, L. F.; OLIVERA, G. O.; GALVAO, J. B. S.; SANTOS, A. A. S.; MUCHALUAT-SAADE, D. C.; CONCI, A. Análise de séries temporais de sinais térmicos da mama para detecção de anomalias. 2014. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/wim/2014/015.pdf>>.

SILVA, L. F.; SAADE, D. C. M.; SEQUEIROS, G. O.; SILVA, A. C.; PAIVA, A. C.; BRAVO, R. S.; CONCI, A. A new database for breast research with infrared image. **Journal of Medical Imaging and Health Informatics**, American Scientific Publishers, v. 4, n. 1, p. 92–100, 2014.

SILVA, L. F.; SANTOS, A. A. S.; BRAVO, R. S.; SILVA, A. C.; MUCHALUAT-SAADE, D. C.; CONCI, A. Hybrid analysis for indicating patients with breast cancer using temperature time series. **Computer methods and programs in biomedicine**, Elsevier, v. 130, p. 142–153, 2016.

SILVA, M. D. S. d. **Análise Computacional de Imagens de Mamografia para Apoio ao Diagnóstico: Detecção e Caracterização de Calcificações**. Tese (Doutorado) — Dissertação de Mestrado, Faculdade de Engenharia da Universidade do Porto, 2013. Disponível em: <https://web.fe.up.pt/~tavares/downloads/publications/relatorios/Monografia_Marlene_Silva.pdf>.

STAM, C. J.; JONES, B.; NOLTE, G.; BREAKSPEAR, M.; SCHELTENS, P. Small-world networks and functional connectivity in alzheimer’s disease. **Cerebral cortex**, Oxford University Press, v. 17, n. 1, p. 92–99, 2006. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/16452642>>.

SUHRKE, P.; MÆHLEN, J.; SCHLICHTING, E.; JØRGENSEN, K. J.; GÖTZSCHE, P. C.; ZAHL, P.-H. Effect of mammography screening on surgical treatment for breast

- cancer in norway: comparative analysis of cancer registry data. **Bmj**, British Medical Journal Publishing Group, v. 343, p. d4692, 2011.
- SZWARCFITER, J. L. **Grafos e algoritmos computacionais**. 2nd. ed. Rio de Janeiro, Brasil: Editora Campus, 1988. ISBN 8570013418.
- TORTORA, G. J.; PETTI, K. **Principles of human anatomy**. Hoboken, Nova Jersey, EUA: John Wiley & Sons, 2002.
- USUKI, H.; IKEDA, T.; IGARASHI, Y.; TAKAHASHI, I.; FUKAMI, A.; YOKOE, T.; SONOO, H.; ASAISHI, K. What kinds of non-palpable breast cancer can be detected by thermography? **Biomed Thermology**, v. 18, n. 4, p. 8–12, 1998. ISSN 0916-6238. Disponível em: <<http://sciencelinks.jp/j-east/article/200013/000020001300A0252772.php>>.
- VEDALDI, A.; SOATTO, S. Quick shift and kernel methods for mode seeking. In: SPRINGER. **European Conference on Computer Vision**. 2008. p. 705–718. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-540-88693-8_52>.
- VELAME, V. M. G. **Segmentação de imagens utilizando superpixels e redes complexas**. São Paulo, Brasil.: Biblioteca Digital de Trabalhos Acadêmicos da USP., 2016.
- VIGNESHWARAN, B.; SUNDARAM, M.; MENAKA, M.; VENKATRAMAN, B.; SANGEETA, D.; MALATHY, K. Classification of normal and abnormal patterns in medical thermography for the screening of breast cancer. **Quantitative InfraRed Thermography**, 2019. ISSN 0166-218X.
- Vincent, L.; Soille, P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 13, n. 6, p. 583–598, June 1991. ISSN 0162-8828.
- WANG, L.; BOHLER, T.; ZOHRER, F.; GEORGII, J.; RAUH, C.; FASCHING, P.; BREHM, B.; SCHULZ-WENDTLAND, R.; BECKMANN, M.; UDER, M.; HAHN, H. A hybrid method towards automated nipple detection in 3D breast ultrasound images. In: **2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. Chicago, Illinois, EUA: IEEE EMBC., 2014. p. 2869–2872. ISSN 1557-170X.
- WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. **nature**, Nature Publishing Group, v. 393, n. 6684, p. 440, 1998.
- WEIDNER, N.; SEMPLE, J. P.; WELCH, W. R.; FOLKMAN, J. Tumor angiogenesis and metastasis—correlation in invasive breast carcinoma. **New England Journal of Medicine**, Mass Medical Soc, v. 324, n. 1, p. 1–8, 1991.
- WEST, D. B. et al. **Introduction to graph theory**. 2nd. ed. London, England: Pearson, 2001.
- WIECEK, B.; ZWOLENIK, S.; JUNG, A.; ZUBER, J. Advanced thermal, visual and radiological image processing for clinical diagnostics. In: IEEE. [**Engineering in Medicine and Biology, 1999. 21st Annual Conference and the 1999 Annual Fall Meeting of the Biomedical Engineering Society**] BMES/EMBS

Conference, 1999. Proceedings of the First Joint. Atlanta, Georgia, EUA, 1999. v. 2, p. 1108–vol.

WISHART, G.; CAMPISI, M.; BOSWELL, M.; CHAPMAN, D.; SHACKLETON, V.; IDDLES, S.; HALLETT, A.; BRITTON, P. The accuracy of digital infrared imaging for breast cancer detection in women undergoing breast biopsy. **European Journal of Surgical Oncology**, v. 36, n. 6, p. 535–540, jun. 2010. ISSN 1532-2157.

WOOD, P. **Scientific illustration: A guide to biological, zoological, and medical rendering techniques, design, printing, and display.** Hoboken, Nova Jersey, EUA: John Wiley & Sons, 1994.

World Health Organization. **Global Health Observatory Database: Cause-specific mortality 2017.** 2017. Disponível em: <http://www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html>.

YASMIN, M.; SHARIF, M.; MOHSIN, S. Survey paper on diagnosis of breast cancer using image processing techniques. **Research Journal of Recent Sciences**, International Science Congress Association, v. 2, n. 10, p. 88–98, 2013.

ZARBAKSH, P.; ADDEH, A. et al. Breast cancer tumor type recognition using graph feature selection technique and radial basis function neural network with optimal structure. **Journal of cancer research and therapeutics**, Medknow Publications, v. 14, n. 3, p. 625, 2018.

ZELMANOWICZ, A. de M. **Câncer de mama.** 2002. Disponível em: <<https://www.abcdasaude.com.br/cancerologia/cancer-de-mama>>.

**APÊNDICE A – Tabela de descrição de localização da anomalia.
Base DMR-TD (doentes).**

Id	Classif.	Localiz.	Descrição do resultado da biópsia	Tumor / caract. predispostas
138	Benigno	QIE-MD	Fibroadenoma em QIE na mama D, com 0,5x 0,6cm, Birads 4.	Fibroadenoma
179	Benigno	QSI-MD	Possui nódulo de 2,1-5,5 cm em QSI de mama direita.	nódulo indefinido
180	Maligno	QS-ME	Carcinoma ductal infiltrante com 7cm em união dos quadrantes superiores de mama E.	CDI - bem realçado e grande
181	Maligno	QS-MD	Mama direita: Carcinoma Ductal Infiltrante grau 3 com 2cm em União dos quadrantes superiores.	CDI
192	Maligno	ME	Carcinoma ductal infiltrante localmente avançado com cerca de 5-10cm em mama E, inicialmente em união de quadrantes inferiores, porém atualmente lesão ulcerada ao redor do complexo areolo-mamilar(CAM).	CDI
198	Maligno	QS-ME QS-MD *	Carcinoma residual, correspondendo a 60% de celularidade em leito tumoral que 2,5x2,0 cm, sendo 2% representado por carcinoma intra-ductal angio-linfática.	CDS pequeno, acima dos mamilos.
202	Maligno	QSE-MD	Carcinoma ductal in situ, localmente avançado no quadrante superior externo de mama direita.	CDS
204	Benigno	MD *	Descarga papilar de líquido seroso em mama direita. Resultado da peça cirúrgica após retirada da mama: Carcinoma ductal in situ de baixo grau.	CDS a secreção apenas escurece a região central do mamilo.
209	Maligno	QI-ME	Carcinoma ductal infiltrante grau 2, nódulo com cerca de 4cm em união dos quadrantes internos de mama E.	CDI
210	Maligno	QS-ME	Carcinoma ductal in situ de alto grau, união dos quadrantes superiores com 1,1 - 2 cm, mama E.	CDS avançado
213	Maligno	QSE - MD	Carcinoma papilífero em mama D, QSE, nódulo com 1,5cm. Axila esquerda e direita com linfonodos suspeitos.	CDI
240	Maligno	MD	Carcinoma ductal infiltrante infiltrando a derme, com graduação histológica Nottingham II, (grau arquitetural 2, nuclear 2 índice mitótico 3). MDT.	CDI
249	Benigno	QC-MD #	Tumor filóide de mama, de baixo potencial de malignidade. Próximo do mamilo D.	Tumor Filóide
250	Maligno	QC-MD #	Carcinoma lobular in situ. Próximo do mamilo D.	CLS
263	Benigno	MDT #	Esteatonecrose com área de fibrose hialina e calcificação.	Esteatonecrose
280	Maligno	QS-ME	Carcinoma Ductal Infiltrante no QS-ME.	CDI
281	Benigno	QSE-ME	Tumor filóide no QSL-ME.	Tumor Filóide
282	Benigno	QII - ME	Citologia - lesão líquida (Benigno) no QII-ME.	Lesão Líquida
284	Benigno	QS-MD #	Hiperplasia ductal florida/alteração de célula colunar .	Hiperplasia
286	Maligno	QSE-ME	Carcinoma Ductal Infiltrante no QSL-ME.	CDI
291	Benigno	QII - ME	Fibroadenoma no QII-ME.	Fibroadenoma
292	Maligno	QS- MD	Carcinoma Ductal Infiltrante no QS-MD.	CDI

* Exames com descrição pouca precisa.

Exames com descrição imprecisa.

Para ambos casos, a região do quadrante não é explícita. Mas as informações extras, como descrições, informações visuais e tipo de tumor, ajudaram a deduzir a localização do quadrante.

Siglas da tabela:

QSE - quadrante superior externo

QSI - quadrante superior interno

QIE - quadrante inferior externo

QII - quadrante inferior interno

QS - quadrantes superiores

QC - quadrante central

ME - mama esquerda

MD - mama direita

CDI - carcinoma ductal infiltrante

CDS - carcinoma ductal in situ

CLS - carcinoma lobular in situ