



**UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE
UNIVERSIDADE FEDERAL RURAL DO SEMIÁRIDO
MESTRADO EM CIÊNCIA DA COMPUTAÇÃO**



ADELINE MARINHO MACIEL

**MINERAÇÃO DE GRAFOS EM DADOS ESPACIAIS DE
DESMATAMENTO**

MOSSORÓ – RN

2012

ADELINE MARINHO MACIEL

**MINERAÇÃO DE GRAFOS EM DADOS ESPACIAIS DE
DESMATAMENTO**

Dissertação apresentada ao Mestrado de
Ciência da Computação – associação ampla
entre a Universidade do Estado do Rio Grande
do Norte e a Universidade Federal Rural do
Semiárido, para a obtenção do título de Mestre
em Ciência da Computação.

Orientador: Prof. Dr. Marcelino Pereira dos
Santos Silva - UERN.

MOSSORÓ – RN

2012

Catálogo da Publicação na Fonte.

Maciel, Adeline Marinho.

Mineração de grafos em dados espaciais de desmatamento. / Adeline Marinho Maciel. – Mossoró, RN, 2012.

94 f

Orientador(a): Prof. Dr. Marcelino Pereira dos Santos Silva.

Dissertação (Mestrado em Ciência da Computação). Universidade do Estado do Rio Grande do Norte. Departamento de Ciência da Computação.

1. Grafos - Dissertação. 2. Sensoriamento remoto - Dissertação. 3. Mineração de dados - Dissertação. I. Silva, Marcelino Pereira dos Santos. II. Universidade do Estado do Rio Grande do Norte. III. Título.

UERN/BC

CDD 005.1

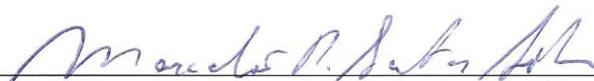
ADELINE MARINHO MACIEL

**MINERAÇÃO DE GRAFOS EM DADOS ESPACIAIS DE
DESMATAMENTO**

Dissertação apresentada ao Mestrado em
Ciência da Computação para a obtenção do
título de Mestre em Ciência da Computação.

APROVADA EM: 18/04/2012.

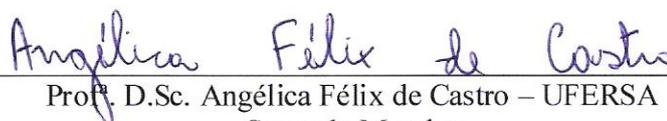
BANCA EXAMINADORA



Prof. D.Sc. Marcelino Pereira dos Santos Silva – UERN
Presidente



Prof.^a. D.Sc. Maria Isabel Sobral Escada – INPE
Primeiro Membro



Prof.^a. D.Sc. Angélica Félix de Castro – UFRSA
Segundo Membro

*A meus pais,
FRANCISCO SIMÕES MACIEL e
SILVANA MARIA MARINHO MACIEL*

AGRADECIMENTOS

A realização deste trabalho só foi possível graças ao Pai Celestial por estar sempre presente em minha vida.

Agradeço aos meus pais, Simões e Silvana, pela confiança depositada em mim, e aos meus irmãos, Aveline e Lincoln, pelas horas de distração proporcionadas durante os momentos de preocupação.

Agradeço a turma pela diversão, convivência, aprendizado, confiança e pela amizade.

Agradeço a Universidade do Estado do Rio Grande do Norte e a Universidade Federal Rural do Semiárido, pelo apoio Institucional, por oferecer uma estrutura e os meios necessários para que fosse possível desempenhar todas as atividades requeridas pelo Programa de Pós-Graduação em Ciência da Computação (MCC). A secretária Rosita por sempre atender minhas solicitações. A Capes, pela bolsa concedida.

Agradeço a todos os professores do MCC, por proporcionarem a abertura de novos horizontes por meio do conhecimento transmitido nas diferentes disciplinas.

Agradeço à Dr. Sc. Isabel Escada, pesquisadora do INPE, que mesmo atarefada com suas pesquisas, disponibilizou um tempo para transmitir conhecimentos que foram de fundamental importância para o sucesso deste trabalho.

Agradeço ao Dr. Sc. Marcelino Pereira, pela orientação, apoio, confiança, paciência, por estar sempre presente, apesar de possuir uma agenda cheia. Pelo conhecimento transmitido, pelo incentivo constante, tempo e disponibilidade prestados a mim.

Agradeço a todos pela compreensão e que fizeram parte da minha vida durante o tempo do mestrado.

“E, se algum de vós tem falta de sabedoria, peça-a a Deus, que a todos dá liberalmente, e o não lança em rosto, e ser-lhe-á dada.

Peça-a, porém, com fé, não duvidando; pois o que duvida é semelhante à onda do mar, que é levada pelo vento de uma para outra parte”.

Tiago 1:5-6.

RESUMO

O combate ao desmatamento na Amazônia Legal é uma prioridade para o governo, para a sociedade e para as organizações ambientais. Para isto é preciso a criação de metodologias que possibilitem realizar o planejamento regional e diagnóstico de desmatamento em dados de sensoriamento remoto. No Brasil, um dos grandes desafios da pesquisa em computação é a modelagem computacional de sistemas complexos artificiais, naturais e socioculturais e da interação homem-natureza. Perante o crescente aumento de volume e complexidade das bases de dados, a pesquisa de novas técnicas de mineração de dados tem sido enfatizada. Muitas destas bases possuem características estruturais, nas quais seus dados são compostos por segmentos e relacionamentos. Assim, a mineração de grafos tem fornecido novos recursos para a busca de informações estratégicas em bases de dados com características estruturais, permitindo a descoberta de padrões. Diante deste contexto, este trabalho visa utilizar grafos para representar relações temporais entre objetos espaciais de desmatamento e, por conseguinte, extrair subestruturas frequentes nestes dados, utilizando a mineração de grafos. O estudo propõe uma metodologia para descoberta de relacionamentos entre os padrões de desmatamento, utilizando um protótipo computacional. Para isto, utiliza dados espaciais de desmatamento, os quais retratam a ocupação do município de Vale do Anari desde o início da implantação dos projetos de assentamento planejado pelo INCRA nesta área, abordando dados catalogados trienalmente desde 1985 até 2000. A mineração de grafos revelou diversas subestruturas frequentes, apresentando padrões recorrentes nas regiões selecionadas, principalmente características sobre evolução temporal dos padrões. A análise das relações entre os padrões de desmatamento e a distância de ocorrência de cada objeto contribuiu para agregar conhecimento aos dados de sensoriamento remoto, provendo novo método de descoberta de relacionamentos relevantes que envolvam a modelagem ambiental, além de subsidiar o desenvolvimento de novas técnicas na busca por padrões de desmatamento.

Palavras-Chave: Mineração de dados, grafos, sensoriamento remoto, padrões de desmatamento, subestruturas frequentes.

ABSTRACT

The fight against deforestation in the Legal Amazon is a priority for the environmental organizations, society and government. It demands the creation of methodologies that enable them to perform the regional planning and diagnosis of deforestation in remote sensing data. In Brazil, one of the major challenges of research in Computer Science is the computational modeling of complex systems: artificial, natural, socio-cultural and human-nature interactions. Given the increasing volume and complexity of databases, the search for new techniques of data mining has been emphasized. Many of these databases have structural features which data are composed of segments and relationships. Thus, graph mining has provided new resources to search for strategic information in databases with structural characteristics, allowing the discovery of patterns. Given this context, this work aims to use graphs to represent temporal relationships among deforestation spatial objects and, therefore, extract frequent substructures from these data using graph mining. The study proposes a methodology for finding relationships among the deforestation patterns, using a software prototype. For that, spatial data from deforestation are used, which depict the occupation of the municipality of Vale do Anari since the beginning of the planned rural settlements by INCRA in this area, addressing data cataloged every three years from 1985 to 2000. The graph mining revealed several frequent substructures, with recurring patterns in the regions selected, mainly features of temporal evolution of such patterns. The relations among the deforestation patterns and distance of occurrence of each object contributed to aggregate knowledge to remote sensing data, providing a new method for finding relevant relationships involving environmental modeling, supporting also the development of new techniques for the search of deforestation patterns.

Keywords: Data mining, graph, remote sensing, deforestation patterns, frequent substructures.

LISTA DE TABELAS

Tabela 1 - Conjuntos de métricas da paisagem.	30
Tabela 2 - Rótulo canônico.....	54
Tabela 3 - Rótulo canônico (final).....	54
Tabela 4 - Distância de 250 m para detecção de padrões Irregulares ligados a Irregulares	76
Tabela 5 - Distância de 500 m para detecção de padrões Irregulares ligados a Irregulares	76
Tabela 6 - Distância de 500 m para detecção de padrões Lineares ligados a Irregulares	78
Tabela 7 - Distância de 250 m para detecção de padrões Irregulares ligados a Irregulares	80
Tabela 8 - Distância de 500 m para detecção de padrões Irregulares ligados a Irregulares	81
Tabela 9 - Distância de 1 km para detecção de padrões Irregulares ligados a Irregulares	81
Tabela 10 - Distância de 500 m para detecção de padrões Lineares ligados a Irregulares	82
Tabela 11 - Distância de 4 km para detecção de padrões Geométricos ligados a Irregulares..	82

LISTA DE FIGURAS

Figura 1 - Elementos de Sensoriamento Remoto	19
Figura 2 - Tipos de resultados gerados a partir de bases de dados coletadas pelos sistemas sensores, respectivamente temos: imagem, gráfico e tabela.	20
Figura 3 - Níveis de coleta de dados.....	21
Figura 4 - Elementos da representação vetorial.....	22
Figura 5 - Exemplo de uma imagem da região de Manaus com composição colorida TM-LANDSAT	23
Figura 6 - Tipologia de padrões espaciais de desmatamento (começando pela esquerda): corredor, difuso, espinha de peixe, geométrico.	28
Figura 7 - As Pontes de Königsberg.....	32
Figura 8 - Grafo $G=(V, E)$	33
Figura 9 - Definições básicas de um grafo.	34
Figura 10 - Lista de Adjacência para grafo não orientado.	36
Figura 11 - Lista de Adjacência para grafo orientado.	36
Figura 12 - Matriz de adjacência para um grafo não orientado.....	37
Figura 13 - Matriz de adjacência para grafo direcionado.....	37
Figura 14 - Matriz de incidência para grafo não orientado.	38
Figura 15 - Matriz de incidência para grafos orientados.	38
Figura 16 - Grafo.	39
Figura 17 - Isomorfismo.	40
Figura 18 - Exemplo de uma árvore implementada com a biblioteca JUNG.....	41
Figura 19 - Etapas do processo de KDD.	44
Figura 20 - Conjunto de dados agrupados em três grupos.	47
Figura 21 - Extração de subestruturas frequentes.....	50
Figura 22 - Geração de candidatas.	51
Figura 23 - Subgrafo Frequente.....	53
Figura 24 - Taxa de desmatamento anual na Amazônia Legal (em km ²).....	58
Figura 25 - Brasil, estado de Rondônia e o município de Vale do Anari.....	59
Figura 26 - Fluxograma de atividades	60
Figura 27 - Procedimentos realizados.	61
Figura 28 - Fluxo de dados entre as ferramentas utilizadas	62

Figura 29 - Tipologia de padrões espaciais de desmatamento: irregular, linear e geométrico.	63
Figura 30 - Exemplo de fluxo ligando alguns dos municípios que compõem o estado de Rondônia.....	65
Figura 31 - Interface desenvolvida para converter os arquivos para o tipo padrão de entrada do FSG.....	66
Figura 32 - Exemplo de um arquivo padrão do FSG.....	67
Figura 33 - Exibição do padrão de grafo com 3 arestas.	68
Figura 34 - Visualização de um dos padrões frequentes encontrados.	68
Figura 35 - Arquivo formato Shape/ArcView do Vale do Anari.	69
Figura 36 - Taxa da média anual de desmatamento no município de Vale do Anari (1982-2000).....	70
Figura 37 - Gráfico de padrões espaciais no município de Vale do Anari no período de 1985-2000.	71
Figura 38 - Matriz de confusão	72
Figura 39 - Resultado da classificação usando o GeoDMA.....	72
Figura 40 - Árvore de decisão do município de Vale do Anari.....	73
Figura 41 - Seleção de polígonos de desmatamento, após a classificação, delimitados em células e criação do fluxo entre todos os objetos de desmatamento de uma região.	74
Figura 42 - Subestrutura frequente de ligação entre padrões irregulares	77
Figura 43 - Subestrutura frequente relacionando a ligação entre um padrão linear a um irregular	78
Figura 44 - Gráfico de padrões espaciais no PA Machadinho no período de 1985-2000	79
Figura 45 - Gráfico de frequência de subestruturas frequentes para distância de 500 m, utilizando suporte mínimo de 10%.....	79

LISTA DE SIGLAS E ABREVIACOES

AGM - *Apriori based Graph Mining*

DCBD - *Descoberta de Conhecimento em Bancos de Dados*

DNA - *deoxyribonucleic acid*

FFSM - *Fast Frequent Subgraph Mining*

FSG - *Frequent SubGraphs*

Gaston - *GrAph, Sequences and Tree extractiON algorithm*

GeoDMA - *Geographical Data Mining Analyst*

GPS - *Global Positioning System*

gSpan - *graph-based Substructure pattern*

INPE - *Instituto Nacional de Pesquisas Espaciais*

ISR - *Imagem de Sensoriamento Remoto*

KDD - *Knowledge Discovery in Databases*

MDL - *Minimum Description Length*

MIGDAC - *Mining Graph DAta for Classification*

PA - *Projeto de Assentamento*

PRODES - *Monitoramento da Floresta Amaznica Brasileira por Satlite*

SGBD - *Sistema Gerenciador de Bancos de Dados*

SQL - *Structured Query Language*

Subdue - *SUBstructure Discovery Using Examples*

SUMÁRIO

1	INTRODUÇÃO	15
1.1	OBJETIVOS	16
1.1.1	Objetivos Específicos	16
1.2	ESTRUTURA DO DOCUMENTO.....	17
2	IMAGENS DE SENSORIAMENTO REMOTO.....	18
2.1	SENSORIAMENTO REMOTO.....	18
2.1.1	Sistema Sensor.....	19
2.1.2	Aquisição de Dados	20
2.2	REPRESENTAÇÃO COMPUTACIONAL DE DADOS GEOGRÁFICOS	21
2.2.1	Representação Vetorial.....	21
2.2.2	Representação Matricial - Imagens	22
2.3	APLICAÇÕES DE IMAGENS DE SENSORIAMENTO REMOTO.....	23
2.4	GEOPROCESSAMENTO	24
2.4.1	Sistemas de Informações Geográficas.....	24
2.5	MINERAÇÃO DE DADOS GEOGRÁFICOS	25
2.5.1	GeoDMA	26
2.6	PRODES -- DADOS DE DESMATAMENTO.....	26
2.7	PADRÕES DE DESMATAMENTO	27
2.7.1	Métricas da Paisagem	29
2.8	TRABALHOS RELACIONADOS	31
3	GRAFOS	32
3.1	HISTÓRICO	32
3.2	DEFINIÇÕES BÁSICAS	33
3.3	TIPOS DE GRAFOS	34
3.4	REPRESENTAÇÕES DOS GRAFOS	35
3.4.1	Lista de Adjacência ou Dicionário	35
3.4.2	Matriz de Adjacência.....	36
3.4.3	Matriz de Incidência	37
3.5	BASES TEÓRICAS PARA MINERAÇÃO DE GRAFOS	38
3.5.1	Subgrafos.....	39

3.5.2	Isomorfismo.....	40
3.6	ÁREAS DE APLICAÇÃO	40
3.7	FERRAMENTAS PARA MANIPULAÇÃO DE GRAFOS.....	41
4	MINERAÇÃO DE DADOS	42
4.1	DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS (DCBD)	43
4.1.1	Etapas do Processo de KDD	44
4.2	TÉCNICAS DE MINERAÇÃO DE DADOS	45
4.2.1	Regras de Associação	45
4.2.2	Classificação	46
4.2.3	Padrões em Dados de Séries Temporais.....	46
4.2.4	<i>Clustering</i> (Agrupamento).....	46
4.3	MINERAÇÃO DE GRAFOS	47
4.3.1	Mineração de Subestruturas Frequentes	48
4.3.2	<i>Frequent SubGraphs</i> (FSG)	51
4.3.3	Pesquisas Relacionadas	54
5	MINERAÇÃO DE GRAFOS EM DADOS ESPACIAIS DE DESMATAMENTO .55	
5.1	DEFINIÇÃO DA PROPOSTA.....	55
5.2	AMAZÔNIA LEGAL.....	57
5.2.1	Área de Estudo Analisada.....	58
6	METODOLOGIA.....	60
6.1	PROTÓTIPO DESENVOLVIDO – FindFRUG	61
6.2	PROCEDIMENTOS REALIZADOS.....	62
6.2.1	Aquisição dos Dados de Desmatamento	62
6.2.2	Seleção dos Dados	63
6.2.3	Definição dos Padrões de Desmatamento	63
6.2.4	Criação das Ligações Entre os Objetos de Desmatamento.....	64
6.2.5	Consultas aos Dados	65
6.2.6	Criação de Arquivos para o FSG.....	66
6.2.7	Mineração de Grafos e Visualização	67
6.3	ESTUDO DE CASO.....	69
6.3.1	Área de Estudo	69
6.3.2	Evolução do Desmatamento	69
6.3.3	Projetos de Assentamento.....	70

6.3.4	Classificação Realizada no Município	71
7	RESULTADOS	75
7.1	SUBESTRUTURAS FREQUENTES – PA MACHADINHO	75
7.1.1	Padrão Irregular	75
7.1.2	Padrão Linear.....	77
7.1.3	Análise dos Resultados PA Machadinho.....	78
7.2	SUBESTRUTURAS FREQUENTES – PA VALE DO ANARI.....	80
7.2.1	Padrão Irregular	80
7.2.2	Padrão Linear.....	81
7.2.3	Padrão Geométrico	82
7.3	DISCUSSÃO SOBRE OS RESULTADOS OBTIDOS.....	83
8	CONCLUSÕES E TRABALHOS FUTUROS.....	85
	REFERÊNCIAS BIBLIOGRÁFICAS	86

1 INTRODUÇÃO

Atualmente, existe uma grande quantidade e diversidade de dados provenientes dos satélites que, periodicamente, capturam imagens da superfície terrestre. Esses dados são armazenados nos institutos acadêmicos de pesquisa, órgãos não governamentais, órgãos públicos ou mesmo privados. As organizações utilizam informações específicas dos dados processados como fonte para diferentes pesquisas, por exemplo, diagnóstico de desmatamento ou queimadas em áreas de proteção ambiental.

Devido ao fato dessas imagens possuírem características únicas, cada vez mais ganham importância nas atividades ligadas à pesquisa e inovação. Com isso, cresce o volume de dados dos grandes repositórios de imagens, como exemplo, o do Instituto Nacional de Pesquisas Espaciais (INPE), que possui mais de 1.849.971 imagens distribuídas e utilizadas para os mais diferentes fins (INPE, 2011). Esta situação faz com que seja necessário o desenvolvimento de técnicas, ferramentas e metodologias para o tratamento das imagens, de média/alta resolução, para que possam ser extraídas informações úteis.

O combate ao desmatamento no Brasil é uma prioridade para o governo, para a sociedade e para as organizações ambientais. Para isso são necessárias monitoração e fiscalização efetivas, possibilitando a aplicação de sanções que possam reprimir e diminuir este problema. Deste modo, é preciso criar metodologias e técnicas que possibilitem diagnosticar e fiscalizar, de forma eficiente e a custos razoáveis, áreas de desmatamento que podem comprometer importantes biomas como, por exemplo, o bioma amazônico.

No Brasil, um dos grandes desafios da pesquisa em computação é a modelagem computacional de sistemas complexos artificiais, naturais e socioculturais e da interação homem-natureza (MEDEIROS, 2008). A complexidade deste tipo de pesquisa aumenta à medida que crescem os volumes dos repositórios de imagens, pois com o crescimento das bases de dados torna-se necessário o desenvolvimento de técnicas que possibilitem a extração de informações úteis dos grandes repositórios (HAN e KAMBER, 2006). Assim, a modelagem computacional de dados de sensoriamento remoto envolve um grande conjunto de algoritmos e técnicas de manipulação de dados.

Um grafo é uma abstração matemática apropriada para resolver diversos tipos de problemas. Por sua vez a mineração de dados é o processo não trivial de identificar, em dados, padrões que sejam válidos, novos (previamente desconhecidos), potencialmente úteis e compreensivos, visando melhorar o entendimento de um problema ou um procedimento de tomada de decisão (SILVA, 2004a). Diante do crescente aumento das bases de dados e da sua

complexidade, novas técnicas de mineração de dados são desenvolvidas, pois grande parte dessas bases possuem características estruturais, cujos dados são compostos por segmentos e relacionamentos entre estes. Dessa forma, a mineração de grafos envolve diversas tarefas sobre bases de dados estruturais. Dentre estas abordagens podemos citar a mineração de subestruturas frequentes, a qual visa identificar subgrafos semelhantes presentes em um conjunto de grafos, permitindo abstrair estruturas detalhadas e representar conceitos estruturais nos dados (SILVA, 2004b). Logo, elaborar a representação de problemas reais por intermédio de grafos constitui-se em uma estratégia atraente, uma vez que é possível minerar estes grafos.

Este estudo propõe uma metodologia para descoberta de padrões de relacionamentos em dados de desmatamento. Para isto utiliza dados espaciais de desmatamento, técnicas de mineração de dados e de mineração de grafos, aplicadas a um estudo de caso. A metodologia requer interação com um especialista do domínio utilizado neste trabalho, uma vez que o especialista subsidiará a análise dos relacionamentos descobertos no estudo de caso.

1.1 OBJETIVOS

Diante deste contexto, esse trabalho tem por objetivo o desenvolvimento de um método para extrair padrões a partir de dados de sensoriamento remoto, utilizando um formalismo matemático computacional (grafos) para representar dados de desmatamento, com o intuito de realizar a busca inteligente de padrões (mineração de grafos). Esta abordagem possibilita a identificação e análise da relação entre padrões de desmatamento a partir de técnicas de mineração de dados, métricas da paisagem e técnicas de mineração de grafos, levando em consideração fatores como estrutura e atores do processo, além da distribuição espacial dos polígonos de desmatamento e o período de ocorrência de tais eventos.

A aplicação deste método contribui para agregar conhecimento aos dados de desmatamento tendo em vista o planejamento regional e o diagnóstico de relações temporais entre áreas de desmatamento.

1.1.1 Objetivos Específicos

- Analisar diferentes ferramentas para manipulação e mineração de dados geográficos, utilizando ferramentas livres e que sejam amplamente conhecidas e empregadas em estudos que tratam dados georeferenciados;

- Desenvolver um método que possibilite modelar dados espaciais de desmatamento como grafos para identificar a existência de subestruturas frequentes entre os objetos de desmatamento;
- Avaliar resultados da aplicação de algoritmos de mineração de grafos, verificando seu comportamento nas bases de dados, e selecionar o que melhor se ajuste às características dos dados obtidos;
- Identificar diferentes padrões de desmatamento, com base em pesquisas que desenvolvem tipologias de padrões de desmatamento bem fundamentadas para a área de estudo;
- Validar resultados obtidos neste trabalho através da análise de áreas de desmatamento abordadas em outras pesquisas, as quais possuam resultados válidos e relevantes, permitindo avaliar a efetividade da proposta.

1.2 ESTRUTURA DO DOCUMENTO

O trabalho está estruturado em 8 capítulos. No capítulo 1 é apresentada uma introdução definindo pontos principais sobre o assunto abordado nesta dissertação.

No capítulo 2 são apresentados conceitos com relação a dados de sensoriamento remoto, características, ferramentas, revisão bibliográfica sobre conceitos de uso de terra, métricas da paisagem, padrões de desmatamento e trabalhos correlatos. O capítulo 3 apresenta conceitos, aplicações e representações de grafos. No capítulo 4 são apresentados conceitos sobre mineração de dados, conceitos sobre mineração de grafos e descoberta de padrões frequentes.

O capítulo 5 discorre sobre a metodologia proposta e descreve a área de estudo escolhida. O capítulo 6 apresenta a metodologia desenvolvida para esta pesquisa, relatando as etapas realizadas para a extração de padrões de desmatamento, criação de fluxos entre objetos de desmatamento, mineração de grafos e descoberta de subestruturas frequentes.

No capítulo 7 são apresentados os resultados obtidos para estudo de caso, exibindo uma análise dos relacionamentos frequentes, descobertos após a mineração de grafos; e uma descrição dos pontos relevantes.

Por fim, no capítulo 8, é apresentada a conclusão do trabalho e sugestão de trabalhos futuros.

2 IMAGENS DE SENSORIAMENTO REMOTO

As Imagens de Sensoriamento Remoto (ISR) são amplamente utilizadas como fonte de dados para diferentes pesquisas. Essas pesquisas podem envolver diversos tipos de domínios, como: ambientais, agrícolas, cartográficas, urbanos, entre outros. Representando assim, uma forma viável de monitoramento, seja na observação ambiental ou urbano, permitindo encontrar características ou mudanças que ocorram no local. Antes de seguir com este tema é necessário contextualizar o que seria Sensoriamento Remoto, uma vez que as imagens e dados são obtidos por meio dessa atividade.

2.1 SENSORIAMENTO REMOTO

Sensoriamento Remoto pode ser entendido como um conjunto de atividades, as quais permitem a obtenção de informações dos objetos que compõem a superfície terrestre sem que haja a necessidade de contato físico com esses objetos (SCHOWENGERDT, 2007). Dentre essas atividades estão envolvidas a detecção, aquisição e análise (extração de informação e interpretação), da energia eletromagnética (ou radiação eletromagnética) que é emitida ou refletida pelos objetos terrestres e que são armazenadas por sensores remotos (MORAES, 2002).

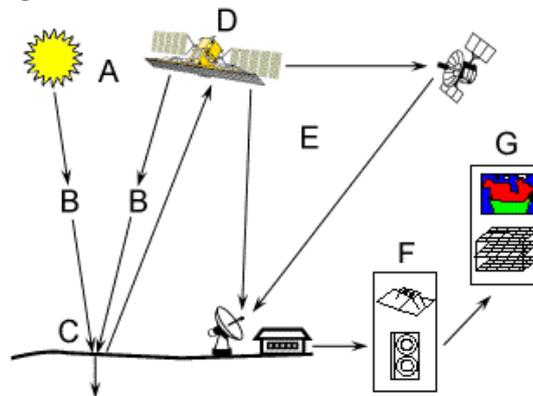
Essa energia eletromagnética que é refletida ou emitida pelos objetos, que compõem a superfície terrestre, é quantificada e registrada na imagem como nível de cinza e utilizada para identificar e classificar diferentes alvos. Deste modo os sensores remotos são ferramentas imprescindíveis para a realização do mapeamento e o monitoramento de recursos naturais.

Em sensoriamento remoto alguns elementos devem ser destacados, como sintetizado na Figura 1 (CCRS, 2007). A seguir serão listadas as principais etapas para aquisição e uso de imagens:

- **(A) Fonte de energia ou iluminação** – uns dos primeiros requisitos para obtenção de dados em sensoriamento remoto é a disponibilidade de uma fonte de energia que ilumina o objeto a ser observado.
- **(B) Radiação e atmosfera** – a energia se desloca da sua fonte para os objetos monitorados e sofrerá interações com a atmosfera pela qual passa. Ao atingir a superfície a energia será refletida pelos objetos e retornará para o sensor após interagir novamente com a atmosfera.

- **(C) Interação com o objeto monitorado** – após a passagem pela atmosfera, a energia interage com o objeto. As interações dependem das características do objeto e da radiação.
- **(D) Registro da energia pelo sensor** – após a reflexão da energia pelo objeto ou após a emissão pelo objeto, é utilizado um sensor para coletar e mensurar o fluxo da radiação eletromagnética.
- **(E) Transmissão, recepção, e processamento** – a energia registrada pelo sensor deve ser transmitida, geralmente de forma eletrônica, para uma estação de recepção e processamento. Nessa estação ocorrerá o processamento dos dados e, posteriormente, a geração de uma imagem.
- **(F) Interpretação e Análise** – a imagem processada é interpretada, visualmente e/ou digitalmente, para extração de informações sobre os objetos de interesse.
- **(G) Aplicação** – as informações geradas a partir das imagens processadas são utilizadas para diferentes tipos de estudos e aplicações.

Figura 1 - Elementos de Sensoriamento Remoto

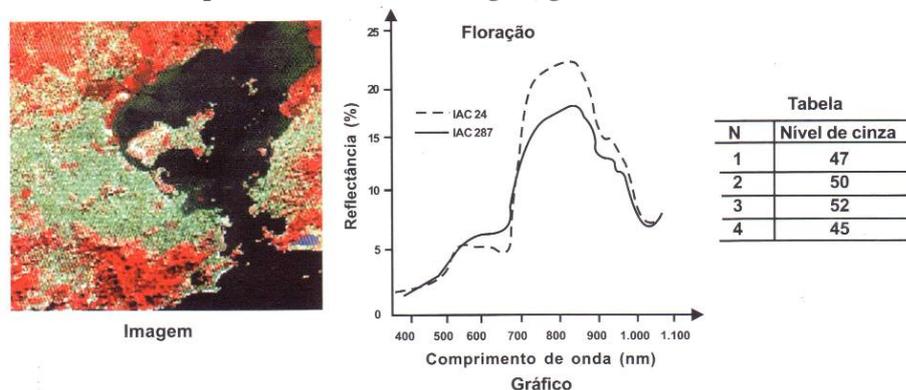


Fonte: (CCRS, 2007)

2.1.1 Sistema Sensor

Os dispositivos capazes de detectar a energia eletromagnética provenientes de um objeto são definidos como sensores remotos. Estes dispositivos são capazes de detectar e registrar a energia eletromagnética, em determinada faixa do espectro eletromagnético gerando, assim, informações que possam ser transformadas em um produto passível de interpretação, seja na forma gráfica, na forma de imagem ou de outro produto (MOREIRA, 2007; MORAES, 2002). A Figura 2 exhibe os diferentes tipos de resultados obtidos pelos sistemas sensores.

Figura 2 - Tipos de resultados gerados a partir de bases de dados coletadas pelos sistemas sensores, respectivamente temos: imagem, gráfico e tabela.



Fonte: (MOREIRA, 2007)

A diferença entre a energia eletromagnética da superfície observada pode ser coletada por sistemas sensores imageadores ou sistemas sensores não-imageadores. Os sistemas imageadores proveem como resultado uma imagem da superfície analisada, por exemplo, as câmeras fotográficas. Já nos sistemas não-imageadores, os resultados são apresentados no formato de dígitos ou gráficos (MOREIRA, 2007).

Os sensores podem ser classificados em duas classes: ativos e passivos. Os sensores ativos possuem (microondas) uma fonte de energia eletromagnética própria que emite essa energia para a superfície que será imageada, detectando assim partes da energia refletida por estes na direção dos sensores. Os sensores passivos não possuem fonte de energia eletromagnética própria (MORAES, 2002).

2.1.2 Aquisição de Dados

Os sensores podem ser mantidos no nível orbital (satélites) ou suborbital, em aeronaves ou mantidos ao nível do solo (MORAES, 2002).

Ao nível do solo, a aquisição é realizada em campo ou em laboratório cujas medidas são obtidas utilizando-se equipamentos radiômetros ou espectroradiômetros. Ao nível de aeronaves, os dados são obtidos por meio de sistemas sensores de varredura óptico-eletrônico, sistemas fotográficos ou radar. A resolução espacial dos dados dependerá da altura do voo do meio utilizado. Ao nível orbital, os sistemas sensores estão localizados em satélites, e este tipo de sensoriamento permite a repetição das informações, assim como uma melhor monitoração de grandes áreas de superfície terrestre. Na Figura 3 são apresentados os níveis orbital e suborbital de coleta de dados de sensoriamento remoto.

Figura 3 - Níveis de coleta de dados.



Fonte: (MOREIRA, 2007)

2.2 REPRESENTAÇÃO COMPUTACIONAL DE DADOS GEOGRÁFICOS

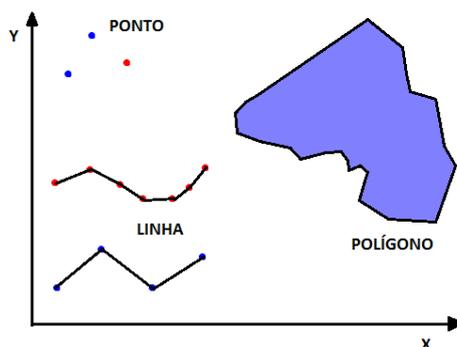
Ao se utilizar computadores como instrumentos na representação de dados georeferenciados, torna-se necessário o estudo de diferentes formas de representação computacional do espaço geográfico. Assim, os dados de sensoriamento remoto após o processamento, seja pela segmentação ou classificação, podem ser convertidos e representados, geometricamente, no formato vetorial e matricial.

2.2.1 Representação Vetorial

O conceito de representação vetorial consiste na representação de um elemento ou objeto, reduzido a três formas básicas: pontos, linha e áreas ou polígonos, como pode ser visto na Figura 4. A representação vetorial de um elemento ou objeto consiste na tentativa de reproduzi-lo o mais próximo possível da realidade (CÂMARA e MONTEIRO, 2001).

- *Ponto* - será um par ordenado (x,y) de coordenadas espaciais. Outros dados não espaciais, denominados atributos, também podem ser arquivados para indicar o tipo de ponto que está sendo tratado;
- *Linha* - também conhecida como arcos ou elementos lineares, consiste em um conjunto de pontos conectados. Nas linhas são armazenadas, além das coordenadas dos pontos que as compõem, os seus atributos;
- *Áreas ou Polígonos* - correspondem à região do plano que é limitada por uma ou mais linhas poligonais conectadas de maneira que o último ponto de uma linha seja o primeiro ponto da próxima linha.

Figura 4 - Elementos da representação vetorial



Fonte: Adaptado de (CÂMARA e MONTEIRO, 2001)

2.2.2 Representação Matricial - Imagens

As imagens digitais da superfície terrestre que são gravadas por sensores orbitais, como o sensor TM do LANDSAT e o HRV do Spot, são transmitidas para as estações localizadas na terra que, por sua vez, transformam essas imagens em dois tipos de produtos: analógico, pouco usado atualmente, e digital (MOREIRA, 2007).

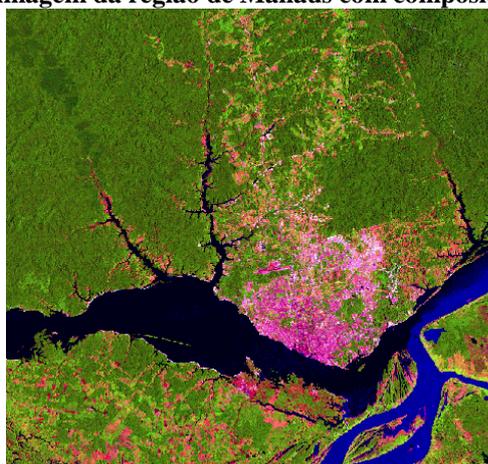
Os produtos analógicos são tratados pelo processo de interpretação visual, que consiste na extração de informações alvo da superfície terrestre, tomando como base a suas respostas espectrais, quando observados nas imagens (MOREIRA, 2007). Os produtos digitais, são tratados por meio de métodos específicos de análise de dados, implementados em computadores, ou seja, essa imagem pode ter um melhoramento de alguns aspectos de suas feições, para posteriormente ser realizada uma interpretação visual ou classificação digital.

As imagens representam formas de captura indireta de informação espacial e além da obtenção por satélites, podem ser adquiridas por fotografias aéreas ou *scannes* aerotransportadores (CÂMARA e MONTEIRO, 2001). Cada elemento de uma imagem (*pixel*) possui um valor proporcional à energia eletromagnética refletida ou emitida pela área da superfície terrestre que possui correlação. Esses elementos são armazenados como matrizes, como pode ser visto na Figura 5, que apresenta uma composição colorida das bandas 3 (Azul), 4 (Verde) e 5 (Vermelho) do satélite TM- LANDSAT, para uma região de Manaus no estado do Amazonas (CÂMARA e MONTEIRO, 2001). Dentre algumas características importantes das imagens de satélites temos:

- *Resolução Espectral* – corresponde ao número e a largura de bandas do espectro eletromagnético imageadas;
- *Resolução Espacial* – corresponde a menor área da superfície terrestre observada instantaneamente por cada sensor;

- *Resolução Radiométrica* – tem relação com o nível de quantização registrado pelo sistema sensor;
- *Resolução Temporal* – é o intervalo de tempo entre duas passagens do satélite pelo mesmo ponto.

Figura 5 - Exemplo de uma imagem da região de Manaus com composição colorida TM-LANDSAT



Fonte: (CÂMARA e MONTEIRO, 2001)

Alguns fatores são importantes para o sucesso na análise das imagens de sensoriamento remoto, dentre eles podemos citar a época da aquisição dessas imagens. Pois, por exemplo, dependendo da cultura estudada podem ocorrer fenômenos naturais como a presença de nuvens e a fenologia das plantas, bem como fatores humanos, como período de plantio e colheita. Para que não ocorram problemas na análise dessas imagens, indica-se o uso, por exemplo, de calendários agrícolas, no caso dos ambientes que sejam ocupados por plantio de soja, cana-de-açúcar, café, entre outros (MOREIRA, 2007).

2.3 APLICAÇÕES DE IMAGENS DE SENSORIAMENTO REMOTO

A interpretação é vital para que profissionais de diferentes áreas possam extrair informações presentes nas imagens. A utilização e relevância das imagens obtidas por sensoriamento remoto são importantes para os diferentes campos de pesquisa (SAUSEN, 2005). Dentre estes campos temos:

- Meteorologia – que envolve a previsão do tempo, acompanhamento de mudanças atmosféricas, controle de poluentes, medição do efeito estufa e do buraco na camada de ozônio;

- Defesa Civil – que tem grande relevância na previsão de catástrofes naturais, permitindo, assim, a tomada de medidas preventivas para minimizar os impactos;
- Planejamento e acompanhamento de culturas agrícolas;
- Planejamento e monitoramento do crescimento urbano;
- Monitoramento de áreas florestais para detecção de queimadas e outras formas de desmatamento;
- Segurança Nacional (militares) – espionagem, acompanhamento de movimentação de inimigos e planejamento estratégico do posicionamento de tropas.

2.4 GEOPROCESSAMENTO

O surgimento dos computadores teve impacto nas mais diferentes áreas da ciência. A representação por meio de *bits* tornou possível reproduzir dados de diversas naturezas: textos de filosofia, fórmulas matemáticas, áudios de língua estrangeira; possibilitando o uso do computador como uma ferramenta de apoio. Além disso, este fato permitiu o surgimento de novos campos de pesquisa. Dentre estes campos encontra-se o Geoprocessamento, que une o computador com técnicas de outras disciplinas do conhecimento para estudar fenômenos relacionados ao espaço físico (CÂMARA e DAVIS, 2001).

As ferramentas de geoprocessamento possibilitam representar os dados do mundo real em sistemas computacionais, de forma a estabelecer estratégias para o planejamento da superfície terrestre por meio de dados georeferenciados. O conceito de Geoprocessamento pode ser visto em (CÂMARA e MONTEIRO, 2001), na qual é definida pelo autor como uma tecnologia interdisciplinar para os estudos de fenômenos ambientais e urbanos.

O Geoprocessamento faz uso de tecnologias para coletar, armazenar, processar e disponibilizar as informações espaciais. Algumas dessas tecnologias são: Sistemas de Posicionamento Global (GPS, do inglês *Global Positioning System*), Bancos de Dados Geográficos e Sistemas de Informações Geográficas (SIG).

2.4.1 Sistemas de Informações Geográficas

A compreensão e representação de dados que abordam eventos ocorridos no espaço geográfico é uma necessidade que denota cada vez mais relevância no mundo atual. Esses eventos abrangem questões, das quais necessitam de ferramentas específicas para a

representação de informações envolvendo dados referenciados por coordenadas espaciais ou geográficas (CÂMARA e QUEIROZ, 2001).

Diante desta necessidade, foram desenvolvidos os SIG, que são capazes de trabalhar com dados georeferenciados, buscando simular a realidade do espaço geográfico. Exemplos desses dados são: dados cartográficos, dados de censo e cadastro urbano, imagens de satélite e modelos de terreno, dentre outros. O termo Sistemas de Informação Geográfica aplica-se aos sistemas que realizam o tratamento computacional de dados geográficos, além de realizar o armazenamento da geometria e dos atributos dos dados que são manipulados. Os dados em um SIG estão localizados na superfície terrestre e representados em uma projeção cartográfica (CÂMARA, MONTEIRO, *et al.*, 2004).

Os SIG podem ser utilizados de três maneiras: como ferramenta para produção de mapas; como uma ferramenta de suporte para análise espacial e fenômenos; e como um banco de dados geográficos, o qual possui funções de armazenamento e recuperação de informação espacial (QUEIROZ e CÂMARA, 2001).

Dentre alguns SIG existentes podemos citar: TerraView (TERRAVIEW, 2011), Spring (SPRING, 2010), MapServer (MAPSERVER, 2010) e GeoServer (GEOSERVER, 2010).

O *software* TerraView foi selecionado como o SIG padrão deste trabalho por melhor atender as necessidades, e tarefas específicas, que foram exigidas no decorrer desta pesquisa. O TerraView é um aplicativo desenvolvido de acordo com a biblioteca de geoprocessamento TerraLib. Essa ferramenta possui como principais atributos uma fácil visualização e manipulação de dados geográficos e contempla também recursos de consulta e análise de dados. Além de manipular dados vetoriais (pontos, linhas e polígonos) e matriciais (imagens e grades), armazenando-os nos SGBR relacionais mais difundidos do mercado (TERRAVIEW, 2011).

2.5 MINERAÇÃO DE DADOS GEOGRÁFICOS

O volume de dados cartográficos teve um aumento considerável nos últimos anos. A ocorrência de tal aumento se deu devido a fatores como a constante captura de imagens da superfície terrestre, muitas vezes diárias; aos novos desenvolvimentos tecnológicos e computacionais, dentre outros fatores. Diante destes acontecimentos, houve um aumento considerável do volume de dados espaciais e, conseqüentemente, da necessidade de aumentar

a capacidade de armazenamento dos repositórios para onde estes dados são destinados (MILLER e HAN, 2009).

Os dados de sensoriamento remoto possuem características e atributos importantes, das quais, quando analisadas adequadamente, revelam informações, antes implícitas, que podem ser importantes. Diante do volume geográfico dos dados, da heterogeneidade, da diferença entre os tipos de dados e da complexidade dos dados é necessário desenvolver ferramentas que sejam capazes de identificar e mapear padrões (MILLER e HAN, 2009). Para isso, existem diferentes ferramentas que realizam a busca por padrões, tanto no domínio espaço-temporal, como em imagens de sensoriamento remoto. Dentre estas ferramentas podemos citar: *A Visualization System for Space-Time and Multivariate Patterns* (VIS-STAMP) (GUO, CHEN, *et al.*, 2006), *Multivariate Mapping and Visualization* (SOMVIS) (GUO, GAHEGAN, *et al.*, 2005), *A System Prototype for Spatial Data Mining* (GeoMiner) (HAN, KOPERSKI e STEFANOVIC, 1997), *Regionalization with Constrained Clustering and Partitioning* (RedCap) (GUO, 2008) e *Geographical Data Mining Analyst* (GeoDMA) (KORTING, FONSECA, *et al.*, 2008), esse último foi utilizado e descrito.

2.5.1 GeoDMA

A partir das ideias propostas por (SILVA, CÂMARA, *et al.*, 2008), foi desenvolvida por (KORTING, FONSECA, *et al.*, 2008), uma ferramenta denominada *Geographical Data Mining Analyst* (GeoDMA), a qual é um *plugin* do SIG TerraView. Esse sistema realiza análise em imagens de sensoriamento remoto baseando-se em técnicas de mineração de dados. Para isso utiliza todas as fases de processamento necessárias para manipulação de dados de sensoriamento remoto, incluindo processos de segmentação, treinamento, classificação, extração de atributos e análise exploratória dos dados, sendo utilizado para as mais diversificadas aplicações.

2.6 PRODES -- DADOS DE DESMATAMENTO

O Sistema de Monitoramento da Floresta Amazônica Brasileira por Satélite, denominado PRODES, é um sistema desenvolvido pelo Instituto Nacional de Pesquisas Espaciais, que realiza o monitoramento do desmatamento das formações florestais na Amazônia Legal, e desde sua criação, em 1998, realiza o levantamento de dados, utilizando imagens de sensoriamento remoto e técnicas de processamento digital de imagens. Por meio

desse sistema são fornecidas estimativas anuais de taxas de desmatamento, detectando exclusivamente desmatamentos tipo “corte raso¹” e que sejam superiores a 6.25 hectares. Para isso, o sistema utiliza imagens dos satélites LANDSAT e CBERS, com resolução espacial de 30 e 20 metros, respectivamente. Todos os dados são disponibilizados para consulta e por meio da Internet é possível obter também o banco de dados digital e multitemporal contendo dados de: imagens, tabelas e mapas vetoriais (INPE, 2011b).

Dados provenientes do PRODES foram utilizados para identificar padrões de desmatamento associados aos diferentes tipos e trajetórias de padrões de ocupação em uma região da Amazônia Legal no trabalho elaborado por (SAITO, ESCADA, *et al.*, 2011), utilizando mineração de dados e métricas de ecologia da paisagem.

Em (VASCONCELOS e NOVO, 2003) os pesquisadores tinham por objetivo detectar a incidência da malária na região da barragem de Tucuruvi, na Amazônia brasileira, utilizando dados como o desmatamento, precipitação e oscilação do nível de água do reservatório. O estudo utilizou dados de desmatamento obtidos pelo PRODES. Como resultado dessa pesquisa, os dados mostraram uma correlação positiva entre o aumento do desmatamento e a incidência de malária, já que as áreas de contato entre o homem e o mosquito são propensas a sua propagação.

2.7 PADRÕES DE DESMATAMENTO

Como as imagens de sensoriamento remoto possuem propriedades espectrais e espaciais que possibilitam compreender os fatores que determinam a mudança de uso de solo em florestas tropicais, estas são utilizadas para possibilitar um acompanhamento da evolução do uso da terra² daquele território, bem como identificar diferentes processos de transformação de uso da terra (ESCADA, 2003). Uma determinada localidade pode sofrer mudança de uso de terra quando ocorre uma troca (conversão) de um tipo de uso para outro tipo ou então quando ocorre uma maior utilização (alteração) do uso da terra corrente.

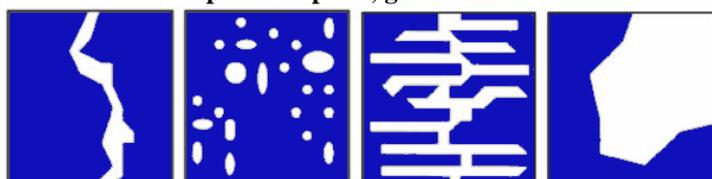
Diversas pesquisas indicam que atores sociais distintos envolvidos em algum tipo de mudança de uso da terra como, por exemplo, pequenos agricultores, grandes fazendeiros, criadores de gado, podem ser identificados por meio de diferentes padrões de uso de terra

¹ Corte raso - é a eliminação de toda e qualquer vegetação existente sobre uma área, ou seja, desmatamento total de um ambiente florestal (IAP, 2010).

² Uso da terra – como a superfície está sendo utilizada pelo homem. Uso da terra inclui: cultivo agrícola, pastagem, recreação, entre outros. (ESCADA, 2003).

(LAMBIN, GEIST e LEPERS, 2003). Os autores observaram padrões de cobertura de terra³ que são mais comuns em áreas com processos de desmatamento nas florestas tropicais e identificaram configurações espaciais que foram associados a processos de uso de terra (ESCADA, 2003). Por meio de modelos visuais esta tipologia de padrões de desmatamento pode ser visualizada, como mostra a Figura 6.

Figura 6 - Tipologia de padrões espaciais de desmatamento (começando pela esquerda): corredor, difuso, espinha de peixe, geométrico.



Fonte: (LAMBIN, GEIST e LEPERS, 2003)

Cada padrão de uso de terra está associado a um ator e processo específicos:

- *Corredor* – geralmente associado à colonização espontânea de migrantes, ao longo de estradas e rios;
- *Difuso* – usualmente relacionado à agricultura de subsistência;
- *Espinha de Peixe* – corresponde aos esquemas de assentamento planejado;
- *Geométrico* – clareiras grandes de desmatamento para atividades econômicas em larga-escala.

Na tipologia apresentada na Figura 6, a escala é um fator relevante, pois, por exemplo, entre os padrões do tipo *geométrico* e *corredor*, o elemento que irá definir o ator associado a uma determinada configuração espacial será o tamanho e forma das partes desmatadas. Analisando-se a forma e tamanho das partes desmatadas, pode-se concluir que o padrão *geométrico* está associado a grandes proprietários enquanto o padrão *corredor* está relacionado aos pequenos proprietários. A diferença entre os padrões *corredor* e *espinha de peixe* está na configuração espacial. O padrão *corredor* está associado a um tipo de ocupação espontânea e o padrão *espinha de peixe*, está associado a uma ocupação do tipo planejada, porém, ambas as tipologias estão associadas aos pequenos proprietários (ESCADA, 2003).

Como a tipologia de padrões espaciais de desmatamento em florestas tropicais proposta por (LAMBIN, GEIST e LEPERS, 2003) captura o processo de desmatamento no âmbito mundial, ela não é suficiente para ser aplicada na apresentação de todos os processos

³ Cobertura de terra – como a superfície é revestida, por exemplo, tipo de vegetação, rochas e águas (ESCADA, 2003).

de desmatamento que ocorrem na Amazônia Legal. Por esta área ser muito heterogênea e extensa, seria preciso levar em consideração fatores regionais do território, como histórico de ocupação e atividades econômicas presentes, para que fosse realizada a associação de padrões espaciais a processos de mudança de uso da terra.

Na pesquisa realizada por (LORENA e LAMBIN, 2007; LORENA, 2008) foi proposta uma metodologia baseada em dados socioeconômicos de pesquisa domiciliar combinados com imagens de sensoriamento remoto. O objetivo da pesquisa era investigar as ligações entre padrões espaciais de desmatamento (como visualizado em imagens de satélites) e os diferentes agentes e suas atividades econômicas específicas. Para tal fim, diferentes padrões espaciais de desmatamento foram identificados por meio de uma tipologia específica, depois alguns padrões foram analisados e separados de acordo com aspectos específicos. Por meio de mapas temáticos, foi verificada a evolução do uso de terra por intermédio das características dos moradores da região. Posteriormente, foi realizada uma análise de *cluster* usando os dados socioeconômicos (pesquisa domiciliar) e dados do uso de terra (imagens de satélite) na busca por grupos homogêneos em relação ao padrão espacial. Por fim, a análise das diferenças entre os padrões espaciais foi ratificada por meio de análise de variância. O estudo mostrou que os diferentes padrões espaciais podem ser relacionados com as diferentes atividades econômicas, porém, a configuração espacial, não é uma consequência apenas das principais atividades econômicas, mas também do projeto de assentamento e dos atores presentes.

Tipologias regionais foram propostas de acordo com os atores envolvidos no processo de desmatamento, tamanho das propriedades, uso da terra e padrões de desmatamento para as duas regiões abordadas: a região denominada “Terra do Meio”, no estado do Pará e a região que contempla o assentamento planejado pelo Instituto Nacional de Colonização e Reforma Agrária (INCRA) de Vale do Anari no estado de Rondônia (ESCADA, 2003). No estado de Rondônia o INCRA fez a implantação de diversas configurações espaciais de lotes de colonos e projetos de assentamento.

2.7.1 Métricas da Paisagem

Em algumas etapas deste trabalho foi necessário utilizar métricas da paisagem para possibilitar a realização da identificação e análise dos padrões de desmatamento, os quais estavam associados aos diferentes tipos de uso de terra na Amazônia Legal.

Métricas da paisagem são utilizadas para calcular atributos que distinguirão os diferentes tipos de padrões de cobertura de terra (SILVA, 2006). A ecologia da paisagem tem

como fundamento a noção de que os processos ecológicos são influenciados fortemente pelos padrões ambientais, ou seja, envolve a análise da interação entre o processo ecológico e o padrão espacial da paisagem, que são as causas e consequências da heterogeneidade espacial em uma dada escala (TURNER, GARDNER e O'NEILL, 2001). Assim, a paisagem pode ser entendida como um mosaico heterogêneo que contém partes homogêneas de acordo com algumas características específicas (FORMAN, 1995).

A identificação de padrões da paisagem é realizada por meio da extração de informações dos padrões utilizando-se métricas da paisagem, que caracterizam propriedades geométricas e espaciais e permitem compreender a distribuição espacial da estrutura da paisagem. Além disso, as métricas da paisagem podem ser usadas para descrever mudanças na estrutura da paisagem e como ferramentas para diagnóstico ambiental (SAITO, 2011; SILVA, 2006; FERRAZ, VETTORAZZI, *et al.*, 2005). As métricas espaciais calculam os atributos da paisagem para regiões, as quais possuem propriedades espaciais similares para categorizá-los (BEKALO, 2009). O trabalho realizado por (SILVA, CÂMARA, *et al.*, 2008) fez uma descrição concisa de diversas métricas que foram utilizadas neste trabalho, dentre elas podemos citar: área, perímetro, compacidade, contiguidade, circularidade (BEKALO, 2009).

Em (MCGARIGAL e MARKS, 1995) foi representada, em termos matemáticos, uma série de métricas para identificar padrões da paisagem. Algumas dessas métricas foram utilizadas por (SILVA, CÂMARA, *et al.*, 2005) e, posteriormente, foram implementadas no *software* GeoDMA desenvolvido por (KORTING, FONSECA, *et al.*, 2008), viabilizando assim a mineração de dados em imagens segmentadas. A Tabela 1 mostra algumas dessas métricas.

Tabela 1 - Conjuntos de métricas da paisagem.

MÉTRICA	VARIAÇÃO	SIGNIFICADO
PERIM - (<i>Perimeter</i>)	$0 < \text{PERIM} < \infty$	Perímetro do objeto da paisagem incluindo toda sua região interna.
AREA - (<i>Area</i>)	$0 < \text{AREA} < \infty$	Área interna do objeto da paisagem.
PARA - (<i>Perimeter-Area Ratio</i>)	$0 < \text{PARA} < \infty$	Medida da complexidade de forma da região.
SHAPE - (<i>Shape Index</i>)	$1 \leq \text{SHAPE} < \infty$	Índice de formato da região, igual a 1 quando a região é mais compacta, e cresce sem limite de acordo com a irregularidade da forma.
CIRCLE - (<i>Related Circumscribing Circle</i>)	$0 \leq \text{CIRCLE} \leq 1$	Quando igual a 0 para regiões circulares e próximos de 1 ocorrem para regiões lineares.
ANGLE - (<i>Angle</i>)	$0 < \text{ANGLE} < 360$	Medida angular do objeto da paisagem.

FRAC - (<i>Fractal Dimension</i>)	$1 \leq \text{FRAC} \leq 2$	Índice que mede a complexidade de forma do polígono. Valores próximos a 1 indicam formas simples enquanto os próximos a 2 para formas mais complexas.
COMPACITY - (<i>Compacity</i>)	$0 < \text{COMPACITY} < \infty$	Mede a compactação do objeto.
RECTANGULARITY - (<i>Rectangularity</i>)	$0 \leq \text{RECTANGULARITY} \leq 1$	Mede a retangularidade da forma do objeto.

Fonte: (MCGARIGAL e MARKS, 1995; SAITO, KORTING, *et al.*, 2010; SILVA, CÂMARA, *et al.*, 2005)

2.8 TRABALHOS RELACIONADOS

Em (SILVA, CÂMARA, *et al.*, 2008; SILVA, 2006) foi desenvolvido uma metodologia, que tinha como objetivo compreender padrões de mudança de cobertura de terra utilizando imagens de sensoriamento remoto, mineração de dados, conceitos de ecologia da paisagem e de processamento digital de imagens. Para tal propósito, foi desenvolvido na pesquisa, um protótipo computacional que utilizou algumas métricas da paisagem a partir de dados de desmatamento, possibilitando a análise e identificação de processos de desmatamento em áreas amazônicas.

3 GRAFOS

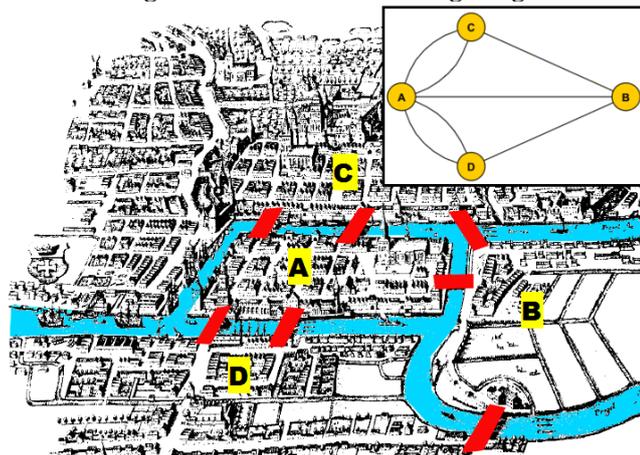
Diferentes estruturas que possuem relacionamentos por meio de ligações podem ser visualizadas facilmente no cotidiano, por exemplo, linhas telefônicas, redes de computadores, transporte coletivo de metrô, circuitos elétricos das residências, além de estruturas menos implícitas para o conhecimento populacional, como os relacionamentos entre os usuários de redes sociais e rotas aéreas. No entanto o que não é pensado é que as estruturas fazem parte de algo do qual os matemáticos denominam como grafos.

3.1 HISTÓRICO

Os primeiros trabalhos de teoria dos grafos tiveram origem no século XVII, no qual vários autores publicaram artigos, com destaque para o problema descrito por Euler, conhecido como o problema das Pontes de Königsberg em 1736 (CARVALHO, 2005).

Numa cidade denominada Königsberg, existiam sete pontes que estabeleciam ligações entre as margens do Rio Pregel e duas de suas ilhas. O problema consistia em, a partir de um determinado ponto, passar por todas as pontes somente uma vez e retornar ao ponto inicial. Esse problema ficou conhecido como Problema do caminho Euleriano.

Figura 7 - As Pontes de Königsberg.

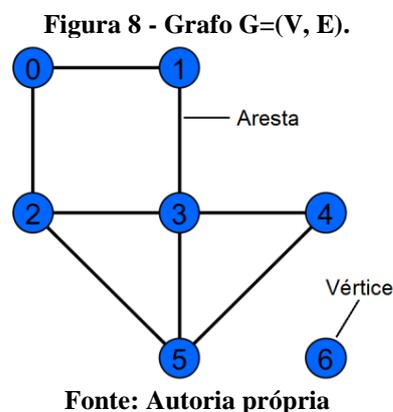


Fonte: Adaptado de (DIESTEL, 2005)

Euler resolveu esse problema criando um grafo em que a terra firme é o vértice e as pontes são as arestas. Então quando se caminha por um vértice, é preciso entrar e sair deles (ou vice-versa, no caso do ponto inicial), o que significa que usamos um número par de arestas cada vez que passamos por um vértice. O grafo da Figura 7 possui vértices com

número ímpar de arestas, a resposta para o problema de passar por todas as pontes somente uma vez e retornar ao ponto inicial não teve solução. Este foi considerado um dos primeiros resultados que abordou teoria dos grafos.

Um grafo é uma estrutura $G=(V, E)$, em que V é o conjunto finito de vértices (nós ou pontos) de um grafo G , os elementos de E são definidos como arestas (arcos ou linhas) válidos para G , estes elementos são definidos por meio de uma função que deriva de V (DIESTEL, 2005). O conjunto de arestas E vem do inglês, *Edges*. A Figura 8 ilustra um grafo que possui vértices e arestas.

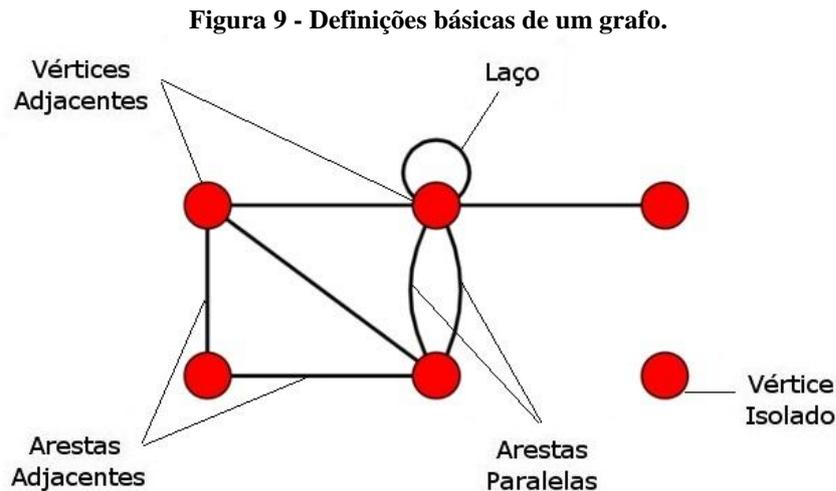


3.2 DEFINIÇÕES BÁSICAS

Alguns conceitos básicos com relação a grafos podem ser definidos e são ilustrados na Figura 9 (MARTINS, 2004):

- Laço - uma aresta que incide com um único vértice;
- Orientação – é a direção para a qual uma seta aponta. Se a aresta não tiver seta, diz-se que ela não tem orientação;
- Incidente – diz-se que uma aresta é incidente com os vértices que ela liga (não importa a orientação);
- Vértices Adjacentes – dois vértices são adjacentes se estão ligados por uma aresta;
- Arestas Adjacentes - duas arestas são adjacentes se elas têm ao menos um vértice em comum;
- Vértice isolado – Um vértice é dito isolado se não existe aresta incidente sobre ele;
- Arestas paralelas – Duas arestas incidentes nos mesmos vértices (não importa a orientação). Ou seja, se $a_1 = (v,w)$ e $a_2 = (v,w)$, então as arestas a_1 e a_2 são paralelas;
- Grau do vértice - É o número de arestas que incidem em cada vértice.

- Um laço, em um grafo não orientado, o grau do vértice será igual ao número de vértices adjacentes. Um laço, em um grafo orientado, o grau do vértice será igual a 2, pois um laço soma um ao grau de entrada e um ao grau de saída, ou seja, o grau do vértice é o número de arestas que saem mais o número de arestas que chegam ao vértice.



Fonte: Autoria própria

3.3 TIPOS DE GRAFOS

O grafo pode ser dirigido e não dirigido. O grafo orientado ou dígrafo é aquele em que é necessário estabelecer um sentido (orientação) para as arestas. O sentido da aresta é indicado por meio do uso de setas. Um grafo é não-dirigido, ou não orientado, quando não é necessário ser estabelecido sentido para as arestas, indicando conexão dupla (LOPES, 1999).

Um grafo bipartido é aquele cujo conjunto de vértices V pode ser particionado em dois conjuntos V_1 e V_2 , de modo que toda aresta de G conecta os vértices de V_1 aos vértices de V_2 .

Grafo simples é o grafo que não contém nenhum par de arestas paralelas ou laços.

Grafo completo é um grafo simples que possui uma aresta entre cada par de seus vértices. Um grafo completo de n vértices é denominado clique.

Um grafo G' é dito complementar de G se possui a mesma ordem de G e se uma aresta $(v,w) \in G$, então $(v,w) \notin G'$ e vice-versa. No caso de grafos dirigidos, o grafo complementar G' será aquele que possui os mesmos vértices de G , possui todas as arestas não presentes em G e possui o reverso das arestas de G (CARVALHO, 2005). O grafo complementar G' será composto exatamente pelas arestas que faltam em G para formar um grafo completo.

Um grafo é considerado rotulado, quando seus vértices e/ou arestas são distinguidos uns dos outros por rótulos, ou seja, um nome, um número ou conjunto de letras e números.

Um grafo $G=(V,E)$ é dito ser valorado quando existe uma ou mais funções relacionando V e/ou E com um conjunto de números. Por exemplo, dado um $G=(V,E)$, em que $V=\{v|v \text{ é uma cidade com aeroporto}\}$ e $E=\{(v,w,t)| \text{ há uma linha aérea ligando } v \text{ a } w, \text{ em que } t \text{ será igual ao tempo de espera do voo}\}$.

Grafo regular é aquele em que todos os vértices possuem grau r , ou seja, todos os vértices possuem o mesmo grau.

Seja $G=(V,E)$ um grafo, um subgrafo G' de G é um grafo $G'=(V',E')$, tal que $V' \subseteq V$ e $E' \subseteq E$.

3.4 REPRESENTAÇÕES DOS GRAFOS

Existem basicamente três formas para representação de grafos: Lista de Adjacência ou dicionário, Matriz de Adjacência e Lista de Incidência.

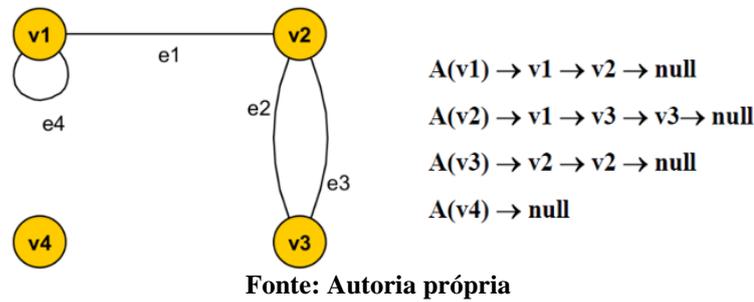
3.4.1 Lista de Adjacência ou Dicionário

“É a forma mais conveniente para a entrada dos dados, pela economia e simplicidade da sua representação” (NETTO, 2003); é construída como um conjunto de listas de vértices, cada lista sendo formada por um vértice e pelo conjunto de vértices que recebem dele um arco ou que com ele partilham uma aresta. Quando o grafo é um dígrafo, existe a opção de se formar a lista com os vértices que enviam um arco ao vértice inicial; um grafo orientado possui, portanto duas listas de adjacência equivalentes.

Em geral a representação de lista de adjacência é preferida por fornecer um modo compacto de representação de grafos esparsos, ou seja, grafos para os quais a quantidade de arestas é bastante inferior ao número de vértices (CORMEN, LEISERSON, *et al.*, 2002).

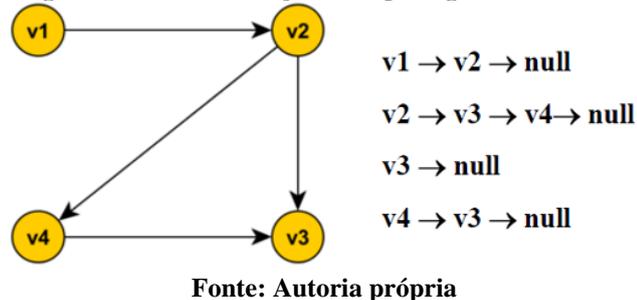
A Figura 10 ilustra a representação da lista de adjacência para um grafo não-orientado. Podemos visualizar que o $v1$ é adjacente ao vértice $v1$, por causa do laço que existe no vértice $v1$, e é adjacente ao vértice $v2$, pois os vértices estão ligados pela aresta $e1$. O vértice $v2$ é adjacente ao vértice $v1$, devido a aresta $e1$ e é adjacente ao vértice $v3$ por duas vezes, porque existem arestas paralelas entre os vértices $v2$ e $v3$. O mesmo ocorre ao vértice $v3$ que é adjacente ao vértice $v2$ duas vezes porque entre os vértices $v2$ e $v3$ existem arestas paralelas. O vértice $v4$ possui o valor nulo na lista de adjacência por ele ser um vértice isolado.

Figura 10 - Lista de Adjacência para grafo não orientado.



A Figura 11 ilustra a representação da lista de adjacência para um grafo orientado. É possível visualizar que o $v1$ é adjacente ao vértice $v2$, por $v2$ ser o vértice de destino do vértice $v1$. Assim ocorrendo sucessivamente com as outras listas de adjacências. O vértice $v2$ será adjacente a $v3$ e $v4$, pois eles serão os vértices de $v2$. O vértice $v3$ não possui lista de adjacência, pois ele não possui outros vértices como destino.

Figura 11 - Lista de Adjacência para grafo orientado.



3.4.2 Matriz de Adjacência

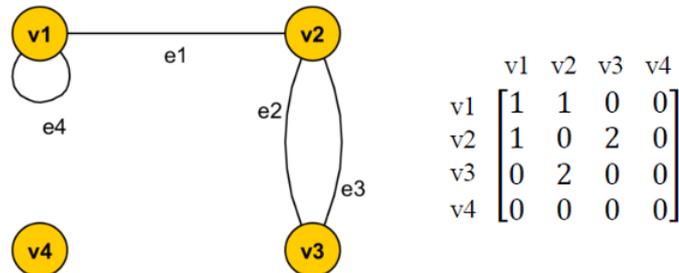
A Matriz de Adjacência é uma matriz de ordem n (a mesma do grafo) na qual se associa cada linha e cada coluna a um vértice (NETTO, 2003). Os dados estruturais correspondem a valores nulos associados à ausência de ligações e a valores não nulos (iguais a 1 quando o grafo não for valorado) nas posições (i,j) associadas à presença de arcos, ou seja:

$$A=[a_{ij}] \quad \begin{aligned} a_{ij} &= 1 \Leftrightarrow \exists (i,j) \in U \\ a_{ij} &= 0 \Leftrightarrow \neg \exists (i,j) \in U \end{aligned}$$

A Figura 12 ilustra a representação da matriz de adjacência para um grafo simples. Podemos visualizar que a matriz de adjacência, formada pelo o número de arestas que unem v_i e v_j , que é composta pelos vértices $|v1| \times |v1|$ possui apenas 1 aresta ligando ambos os vértices, já os vértices $|v3| \times |v2|$ possui 2 arestas ligando ambos, assim sucedendo com as

demais relações $|v| \times |v|$. A relação $|v4| \times |v3|$ possui valor nulo, pois não existem arestas ligando os vértices.

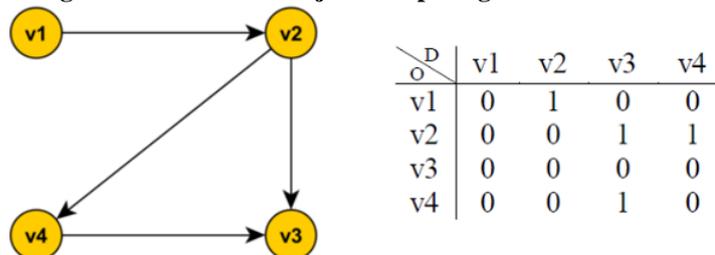
Figura 12 - Matriz de adjacência para um grafo não orientado.



Fonte: Autoria própria

Para os grafos orientados, é preciso observar o sentido do caminho entre os nós e adotar um padrão para o sinal dos pesos (CARVALHO, 2005). Em grafos orientados a matriz não é mais simétrica. A Figura 13 ilustra a representação de uma matriz de adjacência para um grafo orientado. Podemos observar a relação de origem e destino dos vértices, por exemplo, a relação entre os vértices $|v1| \times |v2|$ possui apenas o valor 1, pois v2 é o vértice de destino do vértice v1. Nos vértices $|v2| \times |v3|$ possui o valor 1, pois v3 é o vértice de destino do vértice de origem v2. As relações do vértice v3 com os outros vértices são nulas, pois v3 não tem vértices de destino, do qual ele seja a origem.

Figura 13 - Matriz de adjacência para grafo direcionado.



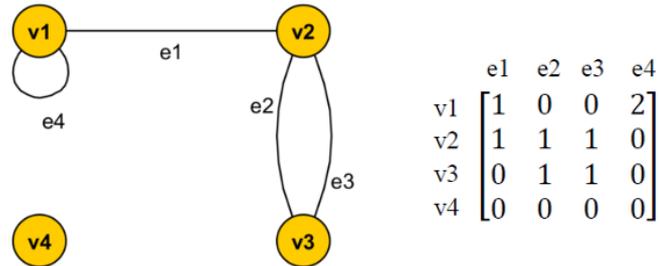
Fonte: Autoria própria

3.4.3 Matriz de Incidência

Matriz de incidência é uma matriz de dimensões $n \times m$, na qual cada linha corresponde a um vértice e cada coluna a uma ligação, ou seja, dado um grafo $G = (V, E)$ a matriz de incidência $MI(G)=[a_{ij}]$ é uma matriz de $|v| \times |e|$ em que a_{ij} é o número de vezes $\{ 0, 1, 2, .. n \}$ em que a aresta e_j é incidente ao vértice v_i (NETTO, 2003).

A Figura 14 ilustra a representação da matriz de incidência para um grafo não orientado. Podemos observar que na matriz de incidência do grafo será representado pelo número de vezes que a aresta incide no vértice, ou seja, a relação $|v1| \times |e1|$, possui o valor 1, pois $e1$ é a aresta que incide em $v1$, já a relação $|v1| \times |e4|$ possui o valor 2, por $e4$ ser um laço. Assim ocorrendo sucessivamente com os outros elementos da matriz.

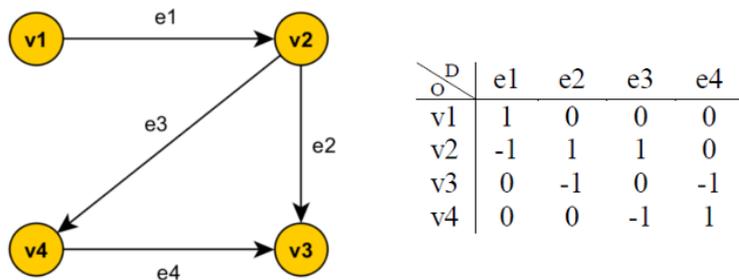
Figura 14 - Matriz de incidência para grafo não orientado.



Fonte: Autoria própria

Para os grafos orientados se v_i é origem e v_j é destino da aresta e_k , então $m_{ik}=1$ e $m_{jk}=-1$. A Figura 15 ilustra a representação da matriz de incidência para grafos orientados. Pode ser observada que a relação $|v1| \times |e1|$, possui valor 1 por o vértice $v1$ ser a origem da aresta $e1$ e a relação $|v2| \times |e1|$ tem como valor -1, por o vértice $v2$ ser o vértice de destino da aresta $e1$. Ocorrendo o mesmo tipo de representação para as outras relações.

Figura 15 - Matriz de incidência para grafos orientados.



Fonte: Autoria própria

3.5 BASES TEÓRICAS PARA MINERAÇÃO DE GRAFOS

Diante do aumento de dados gerados, buscaram-se formas de extrair informações destes dados, inicialmente com mineração e depois com mineração de dados baseada em grafos (WASHIO e MOTODA, 2003). A primeira tem por objetivo extrair regras de conhecimento, enquanto a segunda tem por objetivo obter a topologia dos dados.

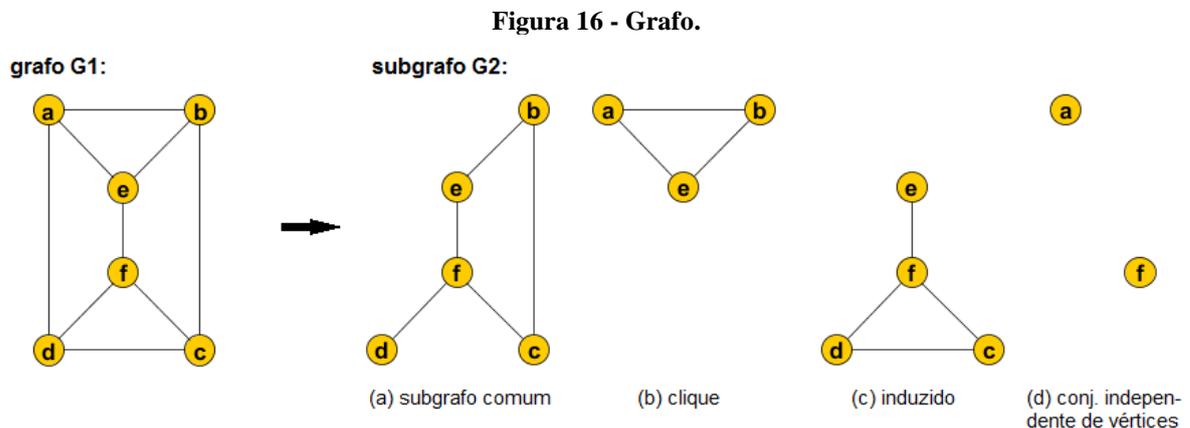
Para entender a mineração de grafos, que será abordada em um capítulo mais adiante, é necessário o conhecimento de algumas bases teóricas relacionadas à teoria dos grafos como: Subgrafos e Isomorfismo, que serão descritas a seguir.

3.5.1 Subgrafos

Por definição, um grafo $G_2=(V_2,E_2)$ é um subgrafo de um grafo $G_1=(V_1,E_1)$ se $V_2 \subseteq V_1$ e $E_2 \subseteq E_1$ (CARVALHO, 2005). Assim é possível observar que:

- Todo grafo é subgrafo dele mesmo.
- O subgrafo de um subgrafo de G é um subgrafo de G .
- Um vértice de G é um subgrafo de G .
- Uma aresta de G com os dois vértices que ele liga é um subgrafo de G .
- Existem alguns subgrafos interessantes para serem observados como:
 - Clique: Um clique é um subgrafo que é completo.
 - Subgrafo induzido: Seja $H = (W, F)$ um subgrafo de $G = (V, E)$. Uma aresta entre dois vértices de W existe se e somente se essa aresta existe em V , digamos que H é um subgrafo induzido por W .
 - Conjunto independente de vértices: Um subgrafo induzido de G que não contém nenhuma aresta.

A Figura 16 ilustra um grafo G_1 com seus respectivos subgrafos G_2 . Podemos observar em $G_2(a)$ temos um subgrafo comum, em $G_2(b)$ um clique, em $G_2(c)$ um grafo induzido pelos vértices c, d, e e f e em $G_2(d)$ um conjunto independentes de vértices.



Fonte: Autoria própria

3.5.2 Isomorfismo

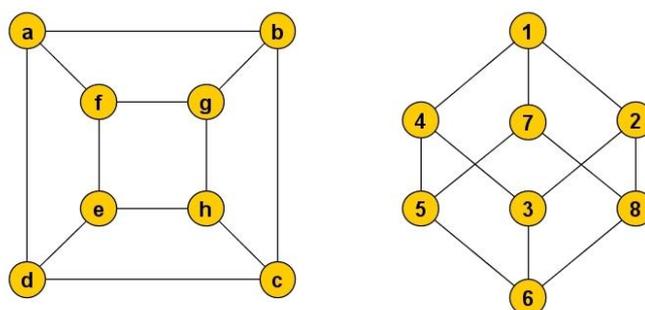
Dois grafos $G_1=(V_1,E_1)$ e $G_2=(V_2,E_2)$ são ditos isomorfos entre si se existe uma correspondência entre os seus vértices e arestas de tal maneira que a relação de incidência seja preservada (CARVALHO, 2005).

Temos $|V_1|=|V_2|$ e existe uma função unívoca $f: V_1 \rightarrow V_2$, tal que (i,j) é elemento de E_1 se e somente se $(f(i),f(j))$ é elemento de E_2 .

Na representação de grafos por matrizes de adjacência, pode-se concluir que dois grafos G_1 e G_2 são isomórficos se, para alguma ordem de seus vértices, suas matrizes forem iguais (CARVALHO, 2005). Uma maneira simples, mas eficiente, de testar o isomorfismo é provar exatamente o contrário por meio do conceito de Invariante. Invariante é uma propriedade que é preservada pelo isomorfismo como, por exemplo, número de nós, grau e determinante da matriz de adjacência.

Para auxiliar na visualização do isomorfismo da Figura 17, pode-se utilizar a função: $f(a) = 1, f(b) = 2, f(c) = 3, f(d) = 4, f(e) = 5, f(f) = 7, f(g) = 8, f(h) = 6$

Figura 17 - Isomorfismo.



Fonte: Autoria própria

3.6 ÁREAS DE APLICAÇÃO

Grafos são bons modelos para diferentes problemas, dentre eles podemos destacar (FEOFILOFF, KOHAYAKAWA e WAKABAYASHI, 2011):

- Informática e computação, para a representação de redes de computadores, redes sociais, também no desenvolvimento de circuitos impressos por intermédio dos grafos planares. Além de ser aplicado no planejamento de projetos por meio das redes PERT⁴ que utilizam grafos para planejar e visualizar a coordenação das atividades do projeto.

⁴ Acrônimo para *Program Evaluation and Review Technique* (FORTIN, ZIELÍNSKI, et al., 2010).

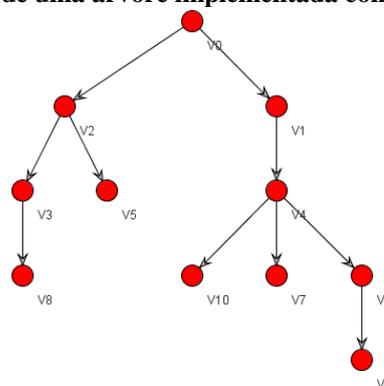
- A bioquímica molecular utiliza os algoritmos de grafos para estudos como os da estrutura do DNA e da composição das proteínas (NETTO, 2003).

3.7 FERRAMENTAS PARA MANIPULAÇÃO DE GRAFOS

Existe uma quantidade significativa de bibliotecas de software e ferramentas para manipulação, criação e visualização de grafos. Essas ferramentas auxiliam de maneira efetiva na automatização tanto do desenho do grafo quanto na execução de algoritmos, por exemplo, problema do caixeiro viajante. Dentre algumas dessas ferramentas podemos citar: “Graphviz” (GRAPHVIZ, 2011), software para visualização de grafos, “Grafos” (VILLALOBOS, 2010), software direcionado para o setor educacional.

A ferramenta de manipulação de grafos escolhida para este trabalho, JUNG (*Java Universal Network/Graph Framework*) trata-se de uma biblioteca de software de código aberto que provê uma série de recursos para modelagem, análise e visualização de estruturas de dados que possam ser representados como grafos ou redes. Essa biblioteca é escrita em linguagem de programação Java, possibilitando assim que a API⁵ do Java também seja utilizada por aplicações desenvolvidas em JUNG. Possui diversos algoritmos de teoria dos grafos, mineração de dados e análise de redes sociais, incluindo rotinas de *clustering*, otimização, geração de grafos aleatórios, análise estatística, e cálculo de distância de redes, fluxos e medidas de importância (centralidade, PageRank, HITS, etc.) (BERNSTEIN, 2011; JUNG, 2010). A Figura 18, ilustra um exemplo de uma árvore que foi implementada utilizando recursos da biblioteca JUNG.

Figura 18 - Exemplo de uma árvore implementada com a biblioteca JUNG.



Fonte: (JUNG, 2010)

⁵ *Application Programming Interface* – é um conjunto de instruções de programação e de padrões estabelecidos por um software para a utilização de suas funções por outros aplicativos, a empresa do software fornece a API para que programadores possam desenvolver suas aplicações (FOLDOC, 2010).

4 MINERAÇÃO DE DADOS

Nos últimos anos o volume de dados tem aumentado consideravelmente, tornando cada vez mais complexa a transformação dessa informação em um conhecimento que tenha utilidade para as organizações que gerenciam essa quantidade de dados. Outro fator importante que causou um crescimento da quantidade de dados foi ascensão do desenvolvimento da área tecnológica, incluindo os meios de armazenamento de informações, que a cada ano aumentam sua capacidade de armazenamento e reduzem os custos (WITTEN e FRANK, 2005). Além disso, a revolução da Internet no final da década de 1990 aumentou ainda mais o acesso de usuários a banco de dados, proporcionando também o aumento dos dados existentes nessas plataformas (SILBERCHATZ, KORTH e SUDARSHAN, 2006).

A principal razão pela qual a Mineração de Dados (em inglês, *Data Mining*) tem atraído muita atenção na indústria da informação nos últimos anos é devido à grande disponibilidade de enormes quantidades de dados e a necessidade iminente para transformar esses dados em informações e conhecimentos úteis (HAN e KAMBER, 2006).

O termo Mineração de Dados refere-se, em geral, ao processo de analisar grandes bancos de dados de forma semi-automática para encontrar padrões úteis. Assim como a descoberta de conhecimento na inteligência artificial (também chamada aprendizado de máquina) ou na análise estatística, a mineração de dados tenta descobrir regras e padrões a partir dos dados. Porém, a mineração de dados difere do aprendizado de máquina e da estatística porque lida com grandes volumes de dados, armazenados principalmente no disco. Ou seja, a mineração de dados lida com “descoberta de conhecimento nos bancos de dados” (SILBERCHATZ, KORTH e SUDARSHAN, 2006).

Nos diversos segmentos da sociedade, as instituições buscam na tecnologia recursos que acrescentem em valor aos seus negócios, seja agilizando operações, suportando ambientes ou viabilizando inovações. No entanto, instituições científicas, indústrias, corporações e governos cada vez mais acumulam grandes volumes de dados, impulsionados também pela versatilidade e alcance proporcionados pela Internet (SILVA, 2004a). Muitos dados podem significar menos informações, devido ao fato dessas informações, estarem escondidas. Diante do aumento das bases de dados torna-se relevante o desenvolvimento de técnicas que possam extrair as informações úteis desses repositórios de dados.

A abundância de dados, juntamente com a necessidade de poderosas ferramentas de análise de dados, tem sido descrita como uma situação de dados ricos, mas pobres em informação. O rápido crescimento, a grande quantidade de dados coletados e armazenados em repositórios de dados, ultrapassou em muito a capacidade humana de compreensão. Como

resultado, os dados recolhidos em grandes repositórios de dados tornam-se “dados túmulos”, ou seja, arquivos de dados que raramente são visitados. Por conseguinte, as decisões importantes são muitas vezes feitas com base nas informações de dados não-ricos armazenados em repositórios de dados, por não possuírem ferramentas para extrair o valioso conhecimento incorporado nos extensos volumes de dados (HAN e KAMBER, 2006).

A análise de grandes quantidades de dados pelo homem é inviável sem o auxílio de ferramentas computacionais apropriadas. Portanto, torna-se imprescindível o desenvolvimento de ferramentas que auxiliem o homem, de forma automática e inteligente, na tarefa de analisar, interpretar e relacionar esses dados para que se possa desenvolver e selecionar estratégias de ação em cada contexto de aplicação (GOLDSCHMIDT e PASSOS, 2005).

Ferramentas de mineração de dados realizam análise de dados e podem descobrir padrões de dados importantes, contribuindo grandemente para estratégias de negócios, bases de conhecimento, investigação científica, médica, controle de produção e análise de mercado ao projeto de engenharia e exploração científica.

4.1 DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS (DCBD)

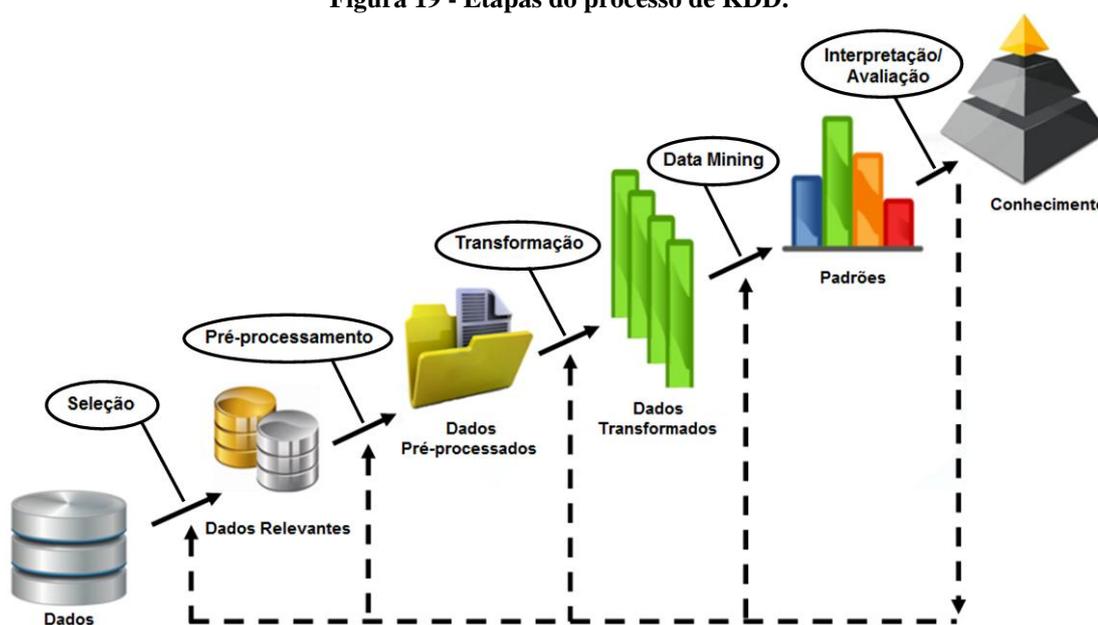
Com a finalidade de subsidiar na interpretação, análise e transformação dos dados em conhecimento útil, houve o surgimento de uma área de estudo denominada Descoberta de Conhecimento em Bancos de Dados - DCBD (do inglês *Knowledge Discovery in Databases - KDD*) (GOLDSCHMIDT e PASSOS, 2005). A DCBD pode ser definida como o processo não trivial que tem por intuito identificar padrões em dados, os quais sejam válidos, novos (desconhecidos), e que sejam potencialmente úteis e compreensíveis, auxiliando dessa forma no entendimento de certo problema ou em algum procedimento de tomada de decisão (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996; SILVA, 2004a).

O termo KDD foi formalizado em 1989 em referência ao amplo conceito de procurar conhecimento por meio de base de dados (PARSAYE, 1989). O KDD é interativo por mostrar a necessidade da inclusão humana controlando o processo e iterativo por indicar a possibilidade de repetições totais ou parciais do processo visando a busca por um conhecimento que seja satisfatório a partir de refinamentos sucessivos (GOLDSCHMIDT e PASSOS, 2005).

4.1.1 Etapas do Processo de KDD

O processo de KDD é iterativo, cognitivo, interativo e exploratório, pois envolve vários passos com diversas decisões sendo realizadas pelo analista, em que esse analista é um especialista do domínio dos dados ou um especialista de análise de dados. Esse processo é composto por um conjunto de etapas (Figura 19), conforme descrito:

Figura 19 - Etapas do processo de KDD.



Fonte: Adaptado de (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996)

- **Seleção de Dados** - Consiste em selecionar um conjunto de dados, ou focar num subconjunto de variáveis ou amostras de dados, que a descoberta deve ser realizada (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996).
- **Pré-Processamento** - Relacionada a operações tais como remoção de ruídos, ou seja, limpeza dos dados, com o objetivo de adequá-los para a etapa de mineração de dados. Nesta etapa devem-se decidir quais as estratégias serão tomadas para tratar os dados ausentes, formatação dos dados, de modo a permitir o ajuste dos dados à etapa de mineração (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996). A qualidade dos dados influencia na qualidade dos modelos de conhecimento a serem absorvidos a partir destes dados. Se os dados informados ao processo de DCBD não possuem uma boa qualidade, os modelos de conhecimentos gerados também não terão uma boa qualidade (GOLDSCHMIDT e PASSOS, 2005).

- **Transformação** - Será feita a localização de características úteis para representar os dados dependendo do objetivo da tarefa, visando à redução do número de variáveis e/ou instâncias a serem consideradas para o conjunto de dados, bem como o enriquecimento semântico das informações (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996). Os dados já formatados serão transformados ou consolidados de forma a serem organizados para uma ferramenta e/ou técnica para a realização da mineração nos dados.
- **Mineração de Dados** – Essa etapa tem por objetivo selecionar os métodos a serem utilizados para localizar padrões nos dados, seguida da busca por padrões de interesse numa forma particular de representação ou conjunto de representações; essa etapa também busca pelo melhor ajuste dos parâmetros do algoritmo para a tarefa em questão (SILVA, 2004a). Métodos específicos serão aplicados para extração dos padrões dos dados, tais como: classificação, clusterização, associação dentre outros (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996).
- **Interpretação dos padrões minerados** - Nesta etapa é realizada a identificação se os padrões realmente são interessantes, ou seja, se de fato constituem uma nova descoberta para representação do conhecimento, possibilitando reiniciar as etapas anteriores para outras interações caso seja necessário (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996).
- **Apresentação da Informação Descoberta** - Nessa etapa é realizada a incorporação do conhecimento obtido à atuação do sistema, ou documentá-lo e reportá-lo às partes interessadas (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996). São usadas técnicas de visualização para que o usuário visualize os padrões. Se necessário, pode-se reiniciar algumas das etapas anteriores para mais iterações.

4.2 TÉCNICAS DE MINERAÇÃO DE DADOS

4.2.1 Regras de Associação

Análise de regras de associação é uma técnica que tem como objetivo a identificação de regras que descrevam os relacionamentos entre componentes de uma base de dados que ocorram com uma mesma frequência. Dentre as diferentes aplicações podemos citar:

transações de compra, supermercados, marketing e controle de estoque (APPEL, 2010; AGRAWAL e SRIKANT, 1994).

4.2.2 Classificação

A classificação pode ser definida como um processo, que tem por objetivo encontrar modelos ou funções que descrevem e distinguem conceitos ou classes. O modelo criado deve ser capaz de prever classes. Essas classes são representadas por um atributo discreto dos elementos que possuem este atributo desconhecido. Assim, os modelos são derivados de um conjunto de dados anteriormente classificado (APPEL, 2010). Dessa forma é possível determinar o valor de um atributo por intermédio dos valores de um subconjunto dos demais atributos da base de dados.

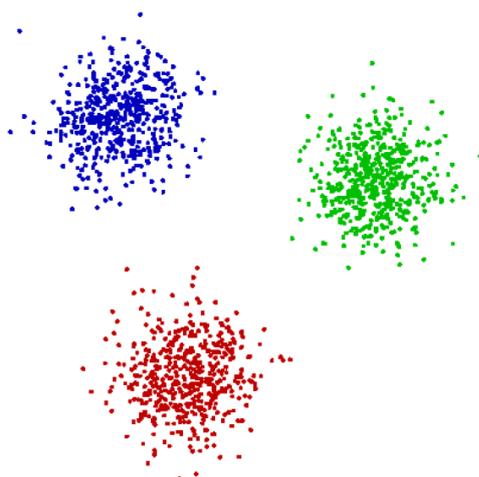
4.2.3 Padrões em Dados de Séries Temporais

Um banco de dados de série temporal consiste em uma sequência de valores ou eventos obtidos no decorrer do tempo, cujos valores são medidos em intervalos de tempo iguais. A análise de séries temporais tem por objetivo encontrar modelos, semelhanças nessa sequência de período de tempo, descobrindo assim o elemento gerador daquele fenômeno (HAN e KAMBER, 2006).

4.2.4 *Clustering* (Agrupamento)

Processo de organização de itens de um conjunto em grupos, no qual os elementos possuem similaridade em algumas características. Os objetos são agrupados com base no princípio de maximizar a similaridade intraclasse e minimizar a similaridade interclasse. Os aglomerados são formados de modo que os elementos dentro do cluster possuam alta similaridade de uns em relação aos outros, mas que sejam diferentes de elementos pertencentes a outros grupos (APPEL, 2010; HAN e KAMBER, 2006). A Figura 20 exibe um exemplo de formação de alguns dados de agrupamento com três grupos, para isso foi utilizado características de proximidade.

Figura 20 - Conjunto de dados agrupados em três grupos.



Fonte: (APPEL, 2010)

4.3 MINERAÇÃO DE GRAFOS

Diferentes dados na *Web* podem ser representados como grafos, especialmente, hiperdocumentos e outros tipos de ligações (por exemplo, redes sociais, mensagens eletrônicas com a relação entre destinatários e remetentes, coautores e colaboradores em pesquisas científicas). Devido ao crescente aumento e a complexidade das bases de dados, isso tem motivado a pesquisa de novas técnicas de mineração de dados. Muitas destas bases possuem características estruturais, dos quais dados são compostos por segmentos e relacionamentos entre estes. Deste modo, a mineração de grafos (do inglês, *graph mining*) tem abordado diferentes tarefas sobre bases de dados estruturais (SILVA, 2004b).

Como uma estrutura de dados em geral, os grafos tornaram-se cada vez mais importantes na modelagem de estruturas sofisticadas e suas interações, pois diversos problemas práticos de interesse podem ser modelados segundo sua teoria e sua estrutura pode assumir numerosos significados, apresentando assim um amplo poder de representação devido à flexibilidade proporcionada por este tipo de estrutura (LAST, KANDEL e BUNKE, 2004). Alguns exemplos de aplicações que utilizam modelagem por intermédio de grafos e suas propriedades são: análise de redes (sociais, internet, transporte), representação de estrutura molecular em física e química, projetos de circuitos eletrônicos, processamento de imagens, visão computacional, fluxos de trabalho e documentos XML (HAN e KAMBER, 2006; GRACIANO, 2007; COOK e HOLDER, 2007).

Assim a mineração de grafos tem por objetivo minerar dados que são compostos por elementos (nós ou vértices) e ligações entre esses elementos (arestas), devido a estas

características os algoritmos tradicionais para mineração de dados não podem ser aplicados para esse tipo de estrutura (APPEL, 2010).

O contexto das redes sociais tem sido extensivamente estudado, sua análise tem sido frequentemente referida como análise de redes sociais. Além disso, em um banco de dados relacional, os objetos são semanticamente ligados utilizando múltiplas relações. Mineração em um banco de dados relacional exige muitas vezes que a mineração seja feita por meio de múltiplas relações entre si, que são semelhantes à mineração em grafos ou redes conectadas. Tal tipo de mineração de dados por intermédio de relações é considerada de mineração de dados multirrelacional (HAN e KAMBER, 2006).

A mineração de grafos pode ser dividida em dois grupos. No primeiro, a mineração é realizada em uma base de dados composta por vários grafos, tendo como objetivo a mineração de subestruturas frequentes. Nesta etapa, um grafo será submetido como consulta a uma base que possui vários grafos. O resultado é composto por todos os grafos que possuem o grafo que foi consultado como subgrafo (KURAMOCHI e KARYPIS, 2004; APPEL, 2010; INOKUCHI, WASHIO e MOTODA, 2000). O segundo grupo consiste na descoberta de padrões, das propriedades estatísticas e leis de potência, em um ou mais grafos, usualmente com milhares de nós (APPEL, 2010).

Diversos algoritmos de busca em grafo têm sido desenvolvidos na informática, química, visão computacional, na indexação de vídeo e recuperação de texto (COOK e HOLDER, 2007). Com a crescente demanda na análise de grandes quantidades de dados estruturados, mineração de grafos tornou-se um tema importante e ativo na mineração de dados. Algoritmos de mineração de dados tradicional também são aplicados na mineração de grafos, porém devido algumas características estruturais desta representação, técnicas vêm sendo pesquisadas e aplicadas à mineração de grafos, dentre elas temos: classificação, *clustering*, mineração de padrões frequentes (subestruturas frequentes) (AGGARWAL e WANG, 2010). Neste trabalho abordaremos a mineração de padrões frequentes, ou seja, a busca por subestruturas frequentes em grafos.

4.3.1 Mineração de Subestruturas Frequentes

Entre os vários tipos de padrões em grafos, subestruturas frequentes são padrões muito básicos que podem ser descobertos em um conjunto de grafos. Eles são úteis para a caracterização de conjuntos de grafos, diferenciando os diversos grupos, classificando e agrupando grafos, e na construção de índices (COOK e HOLDER, 2007).

Há estudos sobre a utilização de subestruturas frequentes como características para classificar os compostos químicos, sobre a técnica de mineração de padrões frequentes de grafos para estudar famílias de proteínas estruturais e também para sua utilização na indexação e busca de similaridade em bases de dados de grafo (HAN e KAMBER, 2006).

Uma abordagem elaborada por (HOLDER, COOK, *et al.*, 2002) visa à descoberta de padrões em dados relacionais representados por grafos, com base no princípio *Minimum Description Length* (MDL) que mede o quão bem os padrões reduzem uma base de dados original. Segundo a heurística MDL, a melhor subestrutura é aquela que minimiza o tamanho da descrição do grafo quando comprimido por meio dessa subestrutura. Esta abordagem está implementada no Sistema Subdue (COOK e HOLDER, 2000), que além de minerar subestruturas e padrões, realiza agrupamento, compressão, aprendizagem relacional e isomorfismo inexato de grafos.

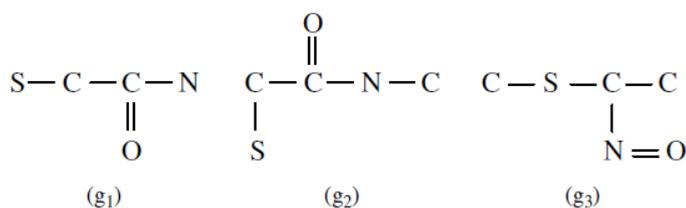
A descoberta dessas subestruturas consiste em duas etapas. Na primeira etapa, os candidatos a subestruturas frequentes são gerados, enquanto a frequência de cada candidato é checada em uma segunda etapa. Muitas pesquisas que visam à descoberta de subestruturas frequentes costumam focar na primeira etapa, pois a segunda etapa envolve isomorfismo de subgrafos que é de complexidade NP-completo⁶ (COOK e HOLDER, 2007; INOKUCHI, WASHIO e MOTODA, 2000).

A mineração de subestruturas frequentes busca subgrafos semelhantes, ou seja, que é frequente em um conjunto de grafos. Um subgrafo frequente pode ser definido como: dado um conjunto de vértices de um grafo g por $V(g)$ e o conjunto de arestas por $E(g)$, um grafo g será subgrafo de outro grafo g' se existir isomorfismo de g em g' . Dado um grafo rotulado, $D = G_1, G_2, G_3, \dots, G_n$, o suporte (g) (ou *frequência*(g)) pode ser definido como a porcentagem ou quantidade de grafos em D cujo g é subgrafo (WASHIO e MOTODA, 2003). Logo, um subgrafo frequente pode ser definido como um grafo no qual o seu suporte não pode ser menor que o suporte mínimo estabelecido (HAN e KAMBER, 2006).

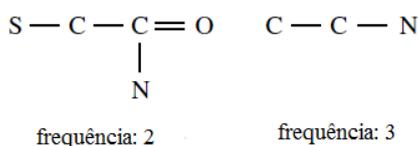
Dessa forma, as subestruturas descobertas são utilizadas para comprimir os dados originais, permitindo abstrair estruturas detalhadas e representar conceitos estruturais nesses dados. A Figura 21 representa a definição de subestrutura frequente, pode-se observar um conjunto de dados de uma estrutura química e a extração de dois subgrafos frequentes, dado o suporte mínimo igual a 66,6%.

⁶ Problemas NP pertencem ao conjunto de todos os problemas de decisão, os quais existem verificadores polinomiais. "NP" é abreviatura de "*nondeterministic polynomial*". Assim, um problema X é NP-completo (ou completo em NP) se X está em NP e todos os outros problemas em NP pode ser polinomialmente reduzido a X , ou seja, não são mais difíceis que X (FEOFILOFF, 2011).

Figura 21 - Extração de subestruturas frequentes.



Amostra de um conjunto grafos



Grafos frequentes

Fonte: Adaptado de (HAN e KAMBER, 2006)

Por meio do agrupamento de subestruturas frequentes descobertas é possível obter uma melhor compreensão dos dados, destacando topologias hierárquicas e representando conhecimento que seja relevante ao usuário.

Na literatura existem duas abordagens básicas que tratam dos problemas de mineração de subestruturas frequentes: primeira, a abordagem *a priori* que realiza a busca por subgrafos frequentes com grafos que possuem um tamanho pequeno, procedendo de uma forma *bottom-up*, que a cada iteração o tamanho dos novos subgrafos descobertos é incrementado em um. Esses novos subgrafos são gerados pela junção de dois subgrafos frequentes descobertos anteriormente, e logo após isso, a frequência deles é checada e comparada ao suporte mínimo. Segunda, a abordagem *pattern-growth*, que realiza a extensão de um subgrafo frequente, adicionando uma nova aresta em todas as posições possíveis, e a cada novo subgrafo frequente descoberto nessa extensão o processo é executado novamente de forma recursiva (HAN e KAMBER, 2006).

Existem diferentes algoritmos para extração de subestruturas frequentes citados nas literaturas (ZOU e HOLDER, 2010; SHRIVASTAVA, KULSHRESTHA, *et al.*, 2010), dentre eles podemos citar: o *Apriori based Graph Mining* (AGM) (INOKUCHI, WASHIO e MOTODA, 2000), o *Frequent SubGraphs* (FSG) (KURAMOCHI e KARYPIS, 2001), o *graph-based Substructure pattern* (gSpan) (YAN e HAN, 2002), *Fast Frequent Subgraph Mining* (FFSM) (HUAN, WANG e PRINS, 2003), *GrAph, Sequences and Tree extractiON algorithm* (Gaston) (NIJSSEN e KOK, 2004) e *SUBstructure Discovery Using Examples*

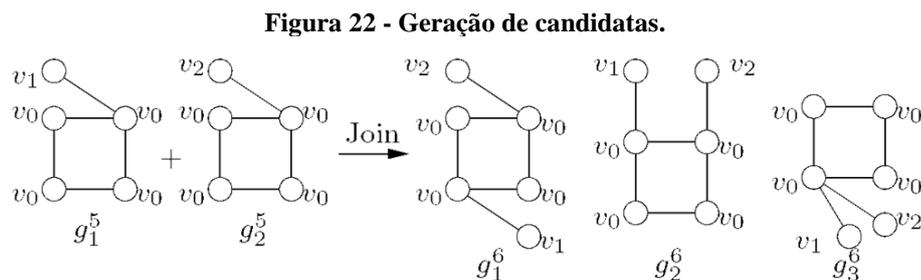
(Subdue) (COOK e HOLDER, 2000). O algoritmo utilizado nessa pesquisa é o *Frequent SubGraphs* (FSG), que faz uso da abordagem *a priori*.

4.3.2 *Frequent SubGraphs* (FSG)

Esse algoritmo utiliza um método de geração de candidatos baseado em aresta, que aumenta o tamanho da subestrutura em uma aresta a cada iteração do algoritmo (KURAMOCHI e KARYPIS, 2001).

Dois padrões de tamanho k são combinados apenas se eles compartilharem o mesmo subgrafo contendo $k-1$ arestas, que recebe o nome de núcleo. No caso do FSG, o tamanho do grafo faz referência ao número de arestas nele contido. O novo candidato a subgrafo frequente gerado, inclui o núcleo e mais duas arestas dos padrões combinados.

Por exemplo, a Figura 22 apresenta esse processo de geração de candidatos à subgrafo frequente em que podemos observar que um núcleo sozinho pode ter múltiplos automorfismo⁷ e cada automorfismo pode conduzir para diferentes $(k+1)$ candidatos. O núcleo (um quadrado de 4 vértices rotulado com v_0) tem mais que um automorfismo que resulta em 3 diferentes candidatos de tamanho 6. Depois da geração dos candidatos, o suporte deles é conferido e o processo se repete até que não haja mais subgrafos frequentes encontrados (HAN e KAMBER, 2006).



Fonte: (KURAMOCHI e KARYPIS, 2001)

Em (PRADHAN, 2006; KURAMOCHI e KARYPIS, 2001) é descrito como o FSG encontra todos os subgrafos que aparecem frequentemente em uma base de dados. Neste caso um subgrafo é considerado como um subgrafo frequente quando ele possui um percentual de frequência (em todo o conjunto do grafo) superior ou igual ao percentual especificado pelo

⁷ Um automorfismo de um grafo G é um isomorfismo de G em G , ou seja, ele é um isomorfismo de grafos de G para ele mesmo (FEOFILOFF, KOHAYAKAWA e WAKABAYASHI, 2011).

usuário. Os objetos no FSG são grafos rotulados não direcionados, cada vértice representa um elemento e cada aresta representa a relação entre estes dois elementos.

O FSG é equivalente à mineração de regras de associação de conjunto de itens frequentes. Se um conjunto de transações é modelado como um grafo, então o FSG pode descobrir subgrafos frequentes em que cada vértice corresponderá a um elemento e cada aresta representará uma coexistência entre dois elementos em uma transação. Basicamente o FSG pode ser declarado da seguinte maneira, dado um conjunto de dados de grafo $G=\{G1,G2,G3,\dots\}$ o problema consiste em descobrir subgrafos que tenham frequência (dentro de todo conjunto de dados) maior que o valor do suporte especificado pelo usuário (PRADHAN, 2006).

Dois etapas são necessárias para a descoberta dos padrões frequentes:

- *Geração das candidatas* - Dois subgrafos frequentes de tamanho k que tem o mesmo núcleo de tamanho $k-1$ são combinados para gerar as possíveis candidatas de tamanho $k+1$; uma rotulagem canônica dos candidatos é atribuída para evitar duplicatas e é imposta uma restrição de propriedade para podar falsas candidatas.
- *Contagem de frequência e poda do subgrafo* - Rótulos canônicos são utilizados para verificar o isomorfismo do subgrafo; subgrafos tendo frequência maior ou igual ao valor do suporte especificado pelo usuário são mantidos, desta maneira somente estes subgrafos participam de uma nova expansão de subgrafos.

Cada uma das transações, candidatos e subestruturas descobertas são armazenadas usando a representação de listas de adjacência. Depois de determinar o rótulo canônico para um subgrafo, a matriz de adjacência canônica é convertida em lista de adjacência, economizando memória, quando as transações dos grafos de entrada são esparsas, aumentando o desempenho computacional (KURAMOCHI e KARYPIS, 2001). O ponto forte do FSG é a rotulagem canônica de subgrafos, a qual é utilizada para verificar as candidatas duplicadas, bem como verificar o isomorfismo no grafo para a contagem de frequência (PRADHAN, 2006; KURAMOCHI e KARYPIS, 2001).

Realização da Rotulagem Canônica

A rotulagem canônica é um código único, dado ao grafo para comprimir a matriz de adjacência. Uma vez que, o mesmo grafo pode estar representado em sequências diferentes de vértices na matriz de adjacência, mais de um rótulo pode ser gerado para a representação da

matriz de adjacência do grafo. Assim, para calcular o rótulo canônico de um grafo, todas as permutações dos vértices do grafo são experimentadas, visando obter a melhor ordem entre os vértices que retorne como resultado o rótulo mínimo lexicograficamente. Para diminuir o espaço de busca, os vértices são divididos de acordo com seus graus e rótulos usando uma técnica chamada vértice invariantes. Então todas as permutações possíveis de vértices dentro de cada partição são geradas. Isso permite que dois grafos isomorfos tenham o mesmo rótulo canônico (PRADHAN, 2006; KURAMOCHI e KARYPIS, 2001).

Por exemplo, para obtenção do rótulo canônico do grafo da Figura 23, são necessárias duas etapas:

- *Divisão por grau do vértice* – primeiro a matriz de adjacência do grafo da Figura 23 é dividida em dois grupos de acordo com o grau de cada vértice, o primeiro grupo contendo grau um e outro com grau dois. Observando a Tabela 2, os dois grupos são: partição 0 e 1, com grau 1 e partição 2 com grau 2;
- *Divisão por rótulo do vértice* – em seguida, a primeira partição é dividida em duas partições novamente, por comparação de strings foi verificado que o rótulo $v_0 < v_1$. Assim observando a Tabela 2, as duas novas partições são: partição 0, possuindo rótulo v_0 e partição 1 com rótulo igual a v_1 ;

Como não existem mais divisões possíveis utilizando a técnica de vértice invariante, e após todos os testes de permutações possíveis, foram obtidas duas permutações diferentes dos vértices, como apresentado na Tabela 2 e Tabela 3. A matriz da Tabela 2 possui rótulo de “000 $e_1e_0e_0$ ”, enquanto a matriz da Tabela 3 possui rótulo igual a “000 $e_0e_1e_0$ ”. Por comparação de strings, o rótulo da Tabela 3, torna-se o rótulo canônico, devido ao fato de $e_0 < e_1$ (KURAMOCHI e KARYPIS, 2001).

Figura 23 - Subgrafo Frequente.

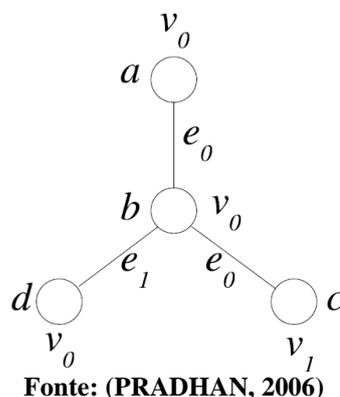


Tabela 2 - Rótulo canônico.

id	d	a	c	b
rótulo	v_0	v_0	v_1	v_0
partição	0		1	2
d	0	0	0	e_1
a	0	0	0	e_0
c	0	0	0	e_0
b	e_0	e_1	e_0	0

Fonte: Adaptado de (PRADHAN, 2006)

Tabela 3 - Rótulo canônico (final).

id	a	d	c	b
rótulo	v_0	v_0	v_1	v_0
partição	0		1	2
a	0	0	0	e_0
d	0	0	0	e_1
c	0	0	0	e_0
b	e_0	e_1	e_0	0

Fonte: Adaptado de (PRADHAN, 2006)

4.3.3 Pesquisas Relacionadas

Algoritmos de mineração de dados são crescentemente aplicados em banco de dados de grafos biológicos. Entretanto, enquanto algoritmos de mineração de grafos podem identificar ocorrência de subgrafos frequentes, estes não necessariamente representam padrões úteis. Nessa abordagem os autores propuseram um novo algoritmo de mineração de grafos, denominado MIGDAC (*Mining Graph DATA for Classification*), que aplica técnicas de teoria dos grafos que são interessantes para a descoberta de subgrafos. Aplicando o algoritmo para descoberta de padrões em dados de compostos químicos, primeiramente os compostos químicos são interpretados como grafos e em seguida transformados em grafos hierárquicos, os quais podem representar mais informações, além de simplificar as estruturas dos grafos complexos. Dessa forma o algoritmo MIGDAC pode ser aplicado para extrair o conjunto de padrões específicos (LAM e CHAN, 2008b).

5 MINERAÇÃO DE GRAFOS EM DADOS ESPACIAIS DE DESMATAMENTO

Pesquisas desenvolvidas no âmbito da mineração de dados geográficos abordam a análise de informações, que são obtidas por meio da extração de padrões em imagens de sensoriamento remoto. Para esta tarefa são utilizadas métricas da paisagem, as quais são usadas para a classificação dos objetos da imagem. O padrão de cobertura de terra é derivado da tipologia de padrões de desmatamento para florestas tropicais, o qual pode ser aplicado à região de estudo. Dentre algumas pesquisas que abordaram esse tipo de análise, podemos citar os trabalhos elaborados por (SAITO, 2011; SAITO, KORTING, *et al.*, 2010; SILVA, CÂMARA, *et al.*, 2008).

No campo de mineração de grafos, em geral, as aplicações são voltadas para pesquisas em problemas que envolvam as ciências químicas e biológicas, por exemplo, na montagem de dados do genoma (MANOCH, TONGSIRI, *et al.*, 2011). Como estes ramos de pesquisa tornaram-se áreas ativas na mineração de dados, então existem inúmeras abordagens que discorrem sobre estes temas (KARWATH e RAEDT, 2004; WASHIO e MOTODA, 2003; LAM e CHAN, 2008b).

Dentre outras pesquisas que abordam a mineração de grafos, pode-se citar: redes sociais, cujos vértices são os indivíduos e as arestas são as relações sociais, por exemplo, redes de coautoria, do qual o vértice em potencial contém filiação e temas de interesse (SILVA, JR. e ZAKI, 2010; LAM e CHAN, 2008a); detecção de operações suspeitas que possam estar relacionadas a atividades criminosas (MICHALAK e KORCZAK, 2011); no controle de candidatas, detecção do desperdício de memória e redes biológicas, para a identificação de mudança de biosistemas (JIANG, COENEN e ZITO, 2010; MAXWELL, BACK e RAMAKRISHNAN, 2010; YOU, HOLDER e COOK, 2008).

Como observado, existe um grande interesse na busca por padrões, uma vez que a interpretação destes representa um avanço significativo na resolução e análise de problemas relevantes.

5.1 DEFINIÇÃO DA PROPOSTA

A pesquisa por padrões de desmatamento torna-se, cada vez mais, importante para o planejamento e monitoramento do meio ambiente. Mediante o estudo de padrões é possível utilizar os meios apropriados para interceder na recuperação do ambiente, além de descobrir os atores que ocasionam o desmatamento.

Diante deste contexto, representar problemas reais por intermédio de grafos constitui-se em estratégia atraente, uma vez que estes grafos possam ser minerados, e assim revelar padrões antes ocultos. A modelagem de dados de sensoriamento remoto por meio de grafos pode revelar ligações relevantes entre os objetos, pois a mineração de grafos pode subsidiar na elucidação, permitindo uma interpretação dos relacionamentos entre os objetos presentes no ambiente ou a confirmação da trajetória de ocupação realizada no decorrer dos anos.

Geralmente um grafo é modelado em formato específico e a sua representação utiliza, na maioria das vezes, formalismos matemáticos de representação como: matriz de adjacência, listas de adjacência, matriz de incidência ou outros recursos que possam viabilizar a representação desta estrutura. Análogo à representação de grafos, a exposição de padrões minerados também ocorre seguindo esse formalismo matemático, o que pode acarretar dificuldades à compreensão dos padrões encontrados, pois estas representações não são facilmente interpretadas pelos usuários.

Dentre os trabalhos que utilizam imagens de sensoriamento remoto para análise, por exemplo, de mudanças que ocorrem no meio ambiente, existem alguns que contemplam uma abordagem embasada em formalismos matemáticos como grafos (ERSAHIN, CUMMING e WARD, 2010; YEH e WU, 2010; FELZENSZWALB e HUTTENLOCHER, 2004; MICHEL e INGLADA, 2008). Os trabalhos que abordam o uso de subestruturas frequentes, podemos citar o trabalho desenvolvido por (TILTON, COOK e KETKAR, 2008) que propôs uma abordagem para representar segmentação de imagens hierárquicas produzidas pelo RHSEG por meio de grafos, compreensível para o sistema Subdue, o qual realiza a busca da melhor subestrutura frequente. No trabalho apresentado por (ZAMALIEVA, AKSOY e TILTON, 2009), é proposto um método genérico não supervisionado para a descoberta de objetos compostos. O método traduz imagens segmentadas para um grafo relacional e aplica o algoritmo de descoberta de conhecimento baseado em grafo Subdue para encontrar subestruturas interessantes e frequentes, que possam corresponder aos objetos compostos. Na estrutura relacional do grafo os nós correspondem às regiões e as arestas representam os relacionamentos entre estas regiões. Os objetos das regiões que aparecem juntos mais frequentemente podem ser considerados fortemente relacionados. Outra abordagem discutida por (DOGRUSOZ e AKSOY, 2007) é apresentado um novo método para modelagem de arranjos de objetos geoespaciais. Os autores propõem o uso de objetos como primitivas texturais e exploram seus padrões espaciais. As primitivas são detectadas usando processamento espectral e morfológico, as quais formam os nós do grafo. Este grafo é

clusterizado e os clusters resultantes são classificados como regulares ou irregulares de acordo com a distribuição dos ângulos entre os nós vizinhos.

Entretanto, mesmo com estas aplicações, ainda não é realizada uma análise utilizando relações entre dados de desmatamento na observação de processos e padrões de ocupação em uma área de assentamento planejado. Este trabalho visa utilizar grafos para representar relações entre padrões de desmatamento e, conseqüentemente, extrair padrões destes relacionamentos, utilizando para isto a mineração de grafos. Os principais relacionamentos que desejamos encontrar é a *relação temporal dos diferentes padrões de desmatamento associados a padrões de ocupação* considerando toda área de estudo. Para a realização deste trabalho, foi escolhida uma área estudada por (ESCADA, 2003; MOTA, CÂMARA, *et al.*, 2009; SILVA, CÂMARA, *et al.*, 2008), na qual foi empregada a metodologia desenvolvida, computacionalmente implementada e aplicada na área de estudo.

5.2 AMAZÔNIA LEGAL

A Floresta Amazônica é a maior floresta tropical do planeta, com uma extensão territorial de 5,5 milhões de quilômetros quadrados na América do Sul, no qual 60% estão localizadas no Brasil e o restante está distribuído entre os países do Peru, Colômbia, Bolívia, Venezuela, Guiana, Suriname, Equador e Guiana Francesa. No Brasil a Amazônia Legal abrange os estados do Acre, Amapá, Amazonas, Mato Grosso, Pará, Rondônia, Roraima, Tocantins e grande parte do Maranhão. Ela corresponde a cerca de 61% do território brasileiro (5.217.423 km²) (SUDAM, 2011).

O combate ao desmatamento no Brasil é uma prioridade para o governo, a sociedade e para as organizações ambientais. E para isto são necessárias monitoração e fiscalização efetivas, com a aplicação de sanções que possam reprimir e diminuir este problema. O Projeto de Monitoramento da Floresta Amazônica Brasileira por Satélite (PRODES) realizou uma estimativa entre agosto de 2010 e julho de 2011 e foi verificado que a área desmatada atingiu uma área de 6.238 quilômetros quadrados (Figura 24). Esta é considerada a menor área desmatada desde que o sistema PRODES iniciou o monitoramento do desmatamento na região da Amazônia Legal, em 1988 (INPE, 2011b; MENDES, 2011).

Para este trabalho foi escolhida uma área de estudo específica, a qual contém uma formação de ocupação estruturada devido ao tipo de assentamento que foi implantado no local. Por meio de revisão bibliográfica, pode-se constatar que outras abordagens haviam realizado uma extensa investigação, inclusive com pesquisas de campo. Estas pesquisas

resultaram na análise de temas relevantes como: descoberta e caracterização dos agentes que originam o desmatamento, influência destes agentes nas áreas de pesquisa e estudos sobre o crescimento do desmatamento nestas regiões, corroborando assim com os resultados deste trabalho (ESCADA, 2003; MOTA, CÂMARA, *et al.*, 2009; PEREIRA, ESCADA e RENNÓ, 2007; SILVA, CÂMARA, *et al.*, 2008).



Fonte: (INPE, 2011b; MENDES, 2011)

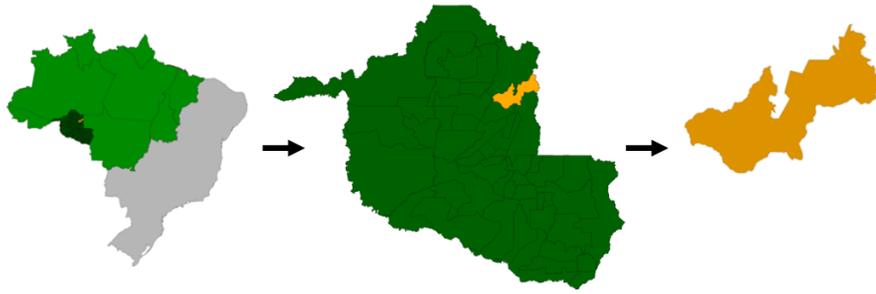
Outro fator que contribuiu para a escolha da área de pesquisa foi a disponibilidade relacionada à aquisição dos dados de desmatamento, o qual foi gentilmente fornecido por (ESCADA, 2003) do Instituto Nacional de Pesquisas Espaciais (INPE). Os dados obtidos possuem como principal característica a apresentação dos objetos de desmatamento desde o início da implantação do projeto de assentamento elaborado pelo Instituto Nacional de Colonização e Reforma Agrária (INCRA) em 1982 no município de Vale do Anari.

Deste modo, esse trabalho visa agregar conhecimento ao estudo já realizado, pois possibilita uma nova fonte de aquisição de informações relacionadas ao desmatamento, permitindo a descoberta de relacionamento entre os padrões de desmatamento utilizando para isto mineração de grafos.

5.2.1 Área de Estudo Analisada

A área de estudo escolhida é o município de Vale do Anari, localizado no estado de Rondônia. Possui uma área de 3.135 km² e uma população superior a 9 mil habitantes, de acordo com o último Censo brasileiro realizado em 2010 (IBGE, 2011a; IBGE, 2011b). A Figura 25 apresenta o Estado de Rondônia e o município de Vale do Anari em destaque.

Figura 25 - Brasil, estado de Rondônia e o município de Vale do Anari.



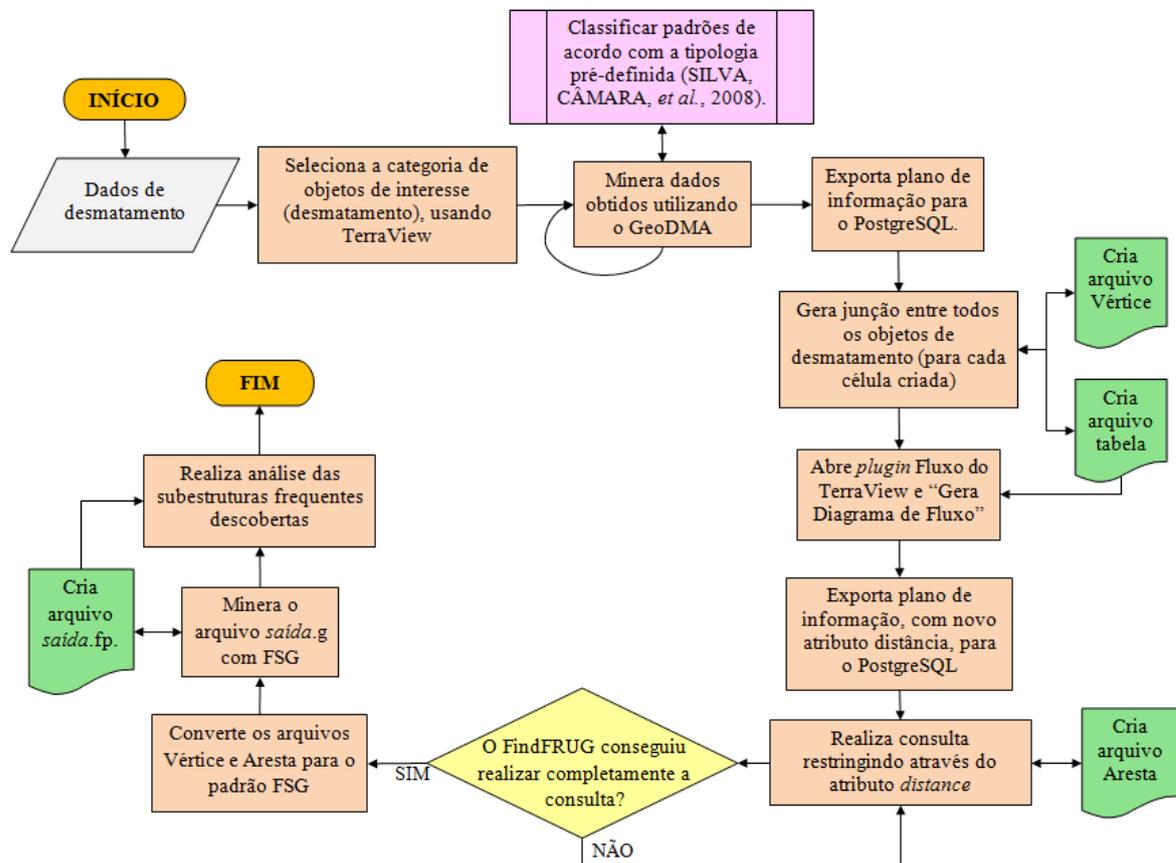
Fonte: (IBGE, 2011a)

O Projeto de Assentamento (PA) predominante deste município é originário de um projeto de assentamento planejado pelo Instituto Nacional de Colonização e Reforma Agrária (INCRA). A partir de 1982 este projeto de assentamento foi estabelecido na região, com lotes de tamanho médio de 50 ha. O processo de ocupação do PA Vale do Anari pode ser descrito como uma ocupação sucessiva de lotes de terra mediante a acessibilidade e melhoria da infraestrutura, por exemplo, melhoria de estradas ou do acesso aos núcleos urbanos, em que existe uma maior concentração de objetos de desmatamento (ESCADA, 2003). Diferentes estudos sobre o processo de ocupação foram realizados para esta área (ESCADA, 2003; SILVA, CÂMARA, *et al.*, 2005).

6 METODOLOGIA

Os procedimentos que foram realizados neste trabalho estão esquematizados na Figura 26. Dentre as atividades desempenhadas, existem etapas iterativas, que preveem a repetição de algum passo, caso necessário; e algumas atividades que terão seguimento após uma tomada de decisão. Além disso, o fluxograma apresenta um esboço do uso de cada ferramenta adotada neste trabalho expondo, por exemplo, em quais procedimentos cada uma das ferramentas foi aplicada visando à descoberta das subestruturas frequentes e inclui também as ferramentas implementadas, por exemplo, o conversor de arquivos.

Figura 26 - Fluxograma de atividades

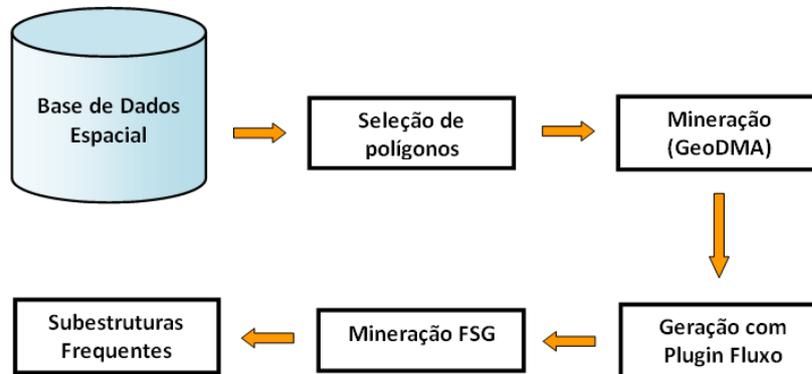


Fonte: Autoria própria

De maneira menos minuciosa a Figura 27 sintetiza em poucas etapas os procedimentos quem foram realizados neste trabalho. Em síntese as etapas foram: aquisição dos dados espaciais de desmatamento, seleção dos dados de desmatamento, criação de mosaico com as cenas adquiridas, extração de padrões de ocupação, consultas SQL no FindFRUG, geração de fluxo com o *Plugin Fluxo*, e por fim a mineração de grafos, com a descoberta das

subestruturas frequentes. Na seção 6.2, cada etapa será detalhada, apresentando os passos e as ferramentas utilizadas no decorrer da pesquisa.

Figura 27 - Procedimentos realizados.



Fonte: Autoria própria

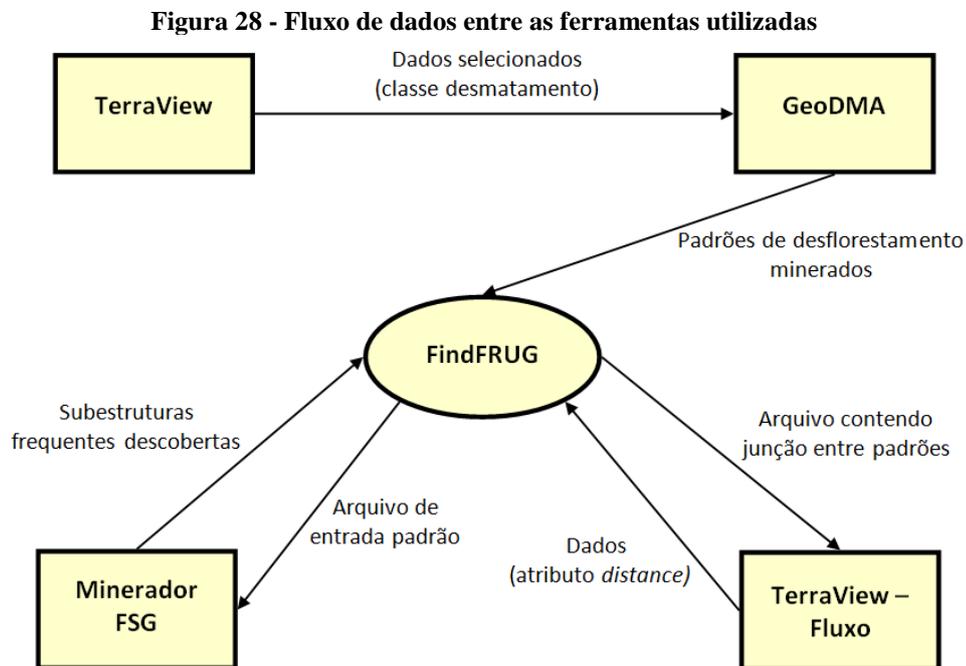
6.1 PROTÓTIPO DESENVOLVIDO – FindFRUG

Devido ao número de recursos utilizados no desenvolvimento, na fase de teste e validação da metodologia apresentada, foi implementado um protótipo denominado FindFRUG (do inglês *Finder of Frequent Relationships Using Graphs* - Buscador de Relações Frequentes Utilizando Grafos). Neste protótipo foram abordados componentes de software que possuíam funções requeridas pela metodologia: conversor de arquivos, que retorna o formato padrão de entrada requerido pela implementação do FSG; leitor e manipulador de banco de dados, o qual permitiu acesso e consulta aos dados exportados para o SGBD PostgreSQL e um visualizador específico que realiza a leitura e modelagem dos arquivos de saída do FSG para um formato gráfico, este último desenvolvido e fornecido por (FRANÇA, 2011).

O protótipo, implementado em Java, interage com os seguintes softwares: PostgreSQL, GeoDMA, TerraView e a implementação do FSG. De acordo com as necessidades particulares de cada etapa realizada neste trabalho, cada ferramenta possuiu uma utilização específica. Para a análise e o armazenamento dos dados espaciais de desmatamento, foi utilizado o Sistema de Informação Geográfica (SIG) TerraView versão 4.1.0 (TERRAVIEW, 2011). Para a mineração de dados (classificação dos objetos) foi utilizado o *Geographical Data Mining Analyst* (GeoDMA) (KORTING, FONSECA, *et al.*, 2008; KORTING, 2007), um *plugin* do TerraView. Para a criação dos fluxos entre os objetos de desmatamento (grafo), foi utilizado o *Plugin Fluxo* (ABREU, 2011), componente disponível

no TerraView. Para a criação das consultas SQL (*Structured Query Language*) foram utilizados os dados exportados para o Sistema Gerenciador de Bancos de Dados (SGBD) PostgreSQL versão 9.0.3 (POSTGRESQL, 2011), e consultados por meio do FindFRUG. Para a mineração de grafos, foi utilizada uma ferramenta que realiza a busca por subestruturas frequentes, uma implementação do algoritmo *FSG* encontrada em (KARYPIS, 2003), além da utilização do FindFRUG para recuperação de dados, conversão de arquivos para o formato específico do FSG e visualização das subestruturas frequentes como grafo.

A Figura 28 sintetiza o fluxo de dados entre as ferramentas que foram utilizadas em conjunto com o protótipo desenvolvido.



Fonte: Autoria própria

6.2 PROCEDIMENTOS REALIZADOS

6.2.1 Aquisição dos Dados de Desmatamento

Os dados relacionados aos padrões de desmatamento foram cedidos por (ESCADA, 2003), disponibilizados no formato Shape/ArcView, na projeção UTM, Datum SAD69 (*South American Datum 1969*) e zona 20. Arquivos neste formato também podem ser obtidos pelo site do projeto PRODES (INPE, 2011b), porém os dados disponibilizados pelo site não contemplam os polígonos do início da implantação do projeto de assentamento no município de Vale de Anari.

6.2.2 Seleção dos Dados

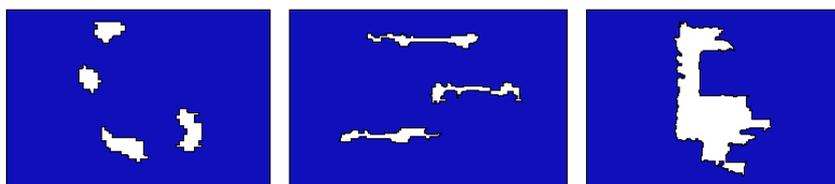
Para esta etapa foi criado um banco de dados no SIG TerraView, com os arquivos de desmatamento. Em seguida, os polígonos foram recortados de acordo com os limites do município, utilizando-se para isto a operação geográfica de intersecção, implementada no TerraView. Após esta etapa, foram selecionados apenas os objetos de polígonos que possuíam como classe principal, o atributo “desmatamento” para os anos da área de pesquisa. Este procedimento de seleção da classe de atributo “desmatamento” foi necessário, pois os dados continham outras classes, por exemplo, floresta, desnecessária para esta fase da pesquisa.

6.2.3 Definição dos Padrões de Desmatamento

Para esta etapa, foi realizada a associação de características dos padrões de desmatamento a um objeto espacial de desmatamento tropical utilizando o *software* GeoDMA, objetivando realizar a classificação dos objetos de desmatamento. Foram utilizados padrões espaciais de desmatamento para os processos em escala regional definidos por (SILVA, CÂMARA, *et al.*, 2005) com base no trabalho desenvolvido por (ESCADA, 2003).

Em (SILVA, CÂMARA, *et al.*, 2008) foram reconhecidas três estruturas elementares na análise dos dados de desmatamento no projeto de assentamento Vale do Anari (RO), como mostra a Figura 29: irregular, linear e geométrico. Por meio dessas três subestruturas, ou seja, da derivação ou combinação destas, a maioria dos padrões espaciais de desmatamento na região surgem. Por exemplo, padrões espaciais como *espinha de peixe* e *difuso* são compostos, respectivamente, por desmatamentos lineares e irregulares em diversos arranjos espaciais, em especial ortogonalidade e agregação.

Figura 29 - Tipologia de padrões espaciais de desmatamento: irregular, linear e geométrico



Fonte: (SILVA, 2006)

Assim como na tipologia descrita por (LAMBIN, GEIST e LEPERS, 2003), a tipologia proposta por (ESCADA, 2003), cada padrão está relacionado a um ator específico (SILVA, 2006):

- *Irregular* – sua distribuição espacial ocorre próxima a estradas principais e núcleos populacionais por pequenos colonos, sendo o principal uso do solo: agricultura de subsistência e/ou criação de gado, praticada por pequenos produtores, com mão de obra predominantemente familiar;
- *Linear* – a distribuição espacial se dá ao longo de estradas por pequenos colonos, sendo o principal uso do solo: agricultura de subsistência e/ou criação de gado, com mão de obra familiar;
- *Geométrico* – a distribuição espacial ocorre próxima a estradas e núcleos populacionais por meio de médios e grandes fazendeiros, cujo principal uso do solo é a criação de gado. Em que as áreas são grandes e regulares e agregam dois ou mais lotes do projeto de assentamento.

Após a classificação da área de estudo, o mapa, com os objetos de padrões de desmatamento, foi dividido em regiões homogêneas de acordo com um tamanho específico de célula. Por meio da criação de células é possível realizar o preenchimento das mesmas, uma funcionalidade do TerraView que possibilita a homogeneização das informações provenientes de diferentes fontes, que possuam formatos distintos, agregando-os em uma mesma base espaço-temporal, permitindo que atributos provenientes de tabelas sejam calculados. Para este trabalho foi realizada a criação de células da área de estudo, com o objetivo de criar regiões e assim gerar um plano de informação individual para cada região inserida dentro dos limites daquela célula. A delimitação das células foi realizada de forma empírica, após observações realizadas mediante a utilização de diferentes medidas de tamanho de célula. A escolha da dimensão se deu pelo tamanho que melhor agregasse os polígonos de padrões de desmatamento.

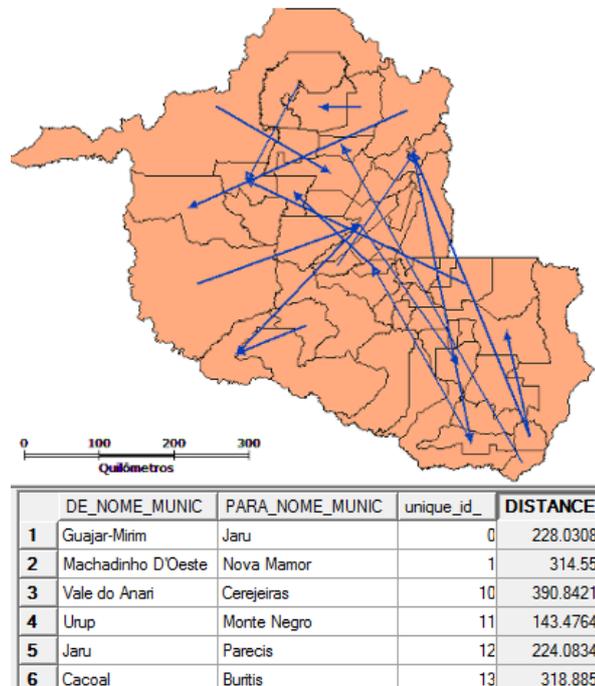
6.2.4 Criação das Ligações Entre os Objetos de Desmatamento

Após a etapa de mineração de padrões de desmatamento que foi constituída da classificação e da criação das regiões (células), e conseqüentemente, da criação de planos de informação para cada região escolhida, estes planos de informação foram exportados para o SGBD PostgreSQL, utilizando-se o *plugin* Cópia de Banco de Dados.

No PostgreSQL foi criado um arquivo denominado “Tabela”, que constitui-se basicamente de uma junção entre todos os objetos de desmatamento, obtidos por meio de consultas SQL. Em seguida, este arquivo com o resultado da consulta foi importado para o

Plugin Fluxo, do TerraView, o qual realizou a ligação entre os objetos de polígonos de desmatamento, gerando um novo plano de informação, contendo um novo atributo “*distance*” gerado após a criação do fluxo entre os objetos, com a distância em quilômetros entre todos os padrões de desmatamento. Vale ressaltar que o arquivo “Tabela” foi criado para cada uma das regiões (células) geradas. A Figura 30 exibe um exemplo de fluxo criado entre os municípios do estado de Rondônia e, na grade inferior, o novo atributo criado (DISTANCE). Utilizando o protótipo FindFRUG, nesta fase também foi criado o arquivo denominado “Vértice”, composto pelo identificador do objeto (*id*), o padrão de desmatamento e o respectivo ano.

Figura 30 - Exemplo de fluxo ligando alguns dos municípios que compõem o estado de Rondônia.



Fonte: Autoria própria

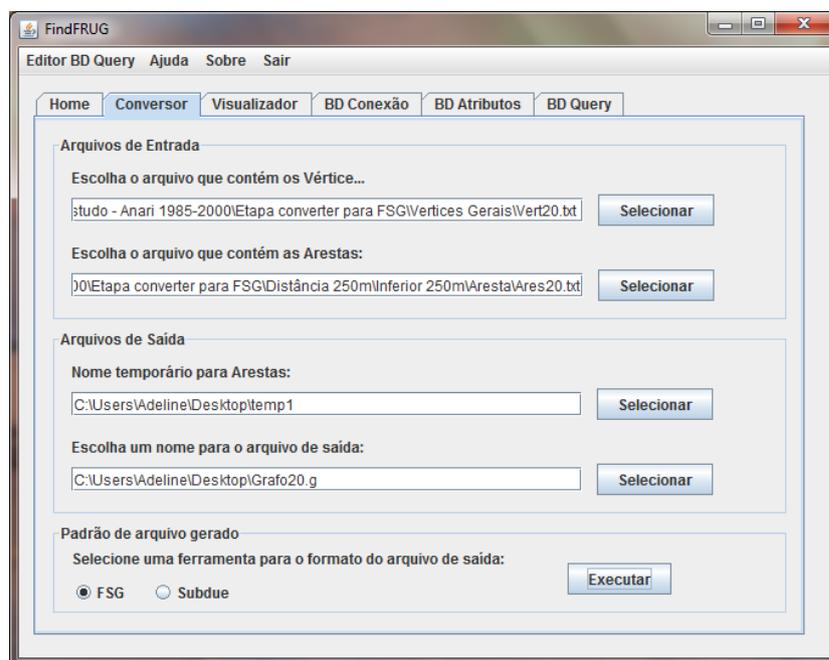
6.2.5 Consultas aos Dados

Após a etapa de criação do grafo entre os objetos de padrões de desmatamento, o banco, com o plano de informação contendo os fluxos, foi exportado para o PostgreSQL. Nesta fase foi realizada uma consulta SQL, utilizando o FindFRUG, visando selecionar uma amostra da ligação entre estes padrões. Nesta consulta foram retirados os dados duplicados e delimitada a quantidade de resultados de acordo com o atributo *distance*, obtido na fase anterior, criando-se, assim, o arquivo “Aresta”. Este arquivo continha o *id* de partida, o *id* de destino e a distância entre os objetos.

6.2.6 Criação de Arquivos para o FSG

Utilizando uma das funcionalidades implementadas no FindFRUG, neste caso a conversão de arquivos, foram gerados os devidos arquivos para serem processados pelo FSG. Desta forma, foi realizada a transformação dos dois arquivos, gerados nas etapas anteriores (Vértice e Aresta), para o formato de entrada compatível com o arquivo padrão do FSG. A Figura 31 exibe a tela de apresentação deste conversor.

Figura 31 - Interface desenvolvida para converter os arquivos para o tipo padrão de entrada do FSG.



Fonte: Autoria própria

A Figura 32 mostra um modelo de arquivo padrão de entrada para o FSG, após a criação deste arquivo para cada uma das regiões que foram escolhidas nas etapas antecedentes. Neste ponto foi gerado um arquivo único com o ajuntamento de todos os arquivos que foram criados para cada região (célula), ou seja, foi gerado um arquivo (*.g) com a quantidade de arquivos de grafos (células selecionadas) para verificar a frequência de subestruturas entre estas ligações. As linhas iniciadas com o caractere “v” indicam a definição dos vértices, com o *id* do vértice seguido de seu respectivo rótulo. O *id* representa o identificador do objeto padrão de desmatamento e o rótulo o tipo de seu padrão e ano de ocorrência. As linhas iniciadas com o caractere “u” indicam a definição das arestas, no qual são passados os dois vértices conectados pelos seus *ids* juntamente com a rotulação da aresta, que neste estudo representa a distância entre os objetos de desmatamento.

Figura 32 - Exemplo de um arquivo padrão do FSG.

```

1 t # 56 vertices and 51 edges
2 v 1810 irregular_df_9497
3 v 1821 irregular_df_9700
4 v 1827 irregular_df_9700
5 v 1831 irregular_df_9700
6 v 1849 irregular_df_8891
7 v 4119 irregular_df_8588
8 v 4163 irregular_df_9497
52 v 4711 irregular_df_9700
53 v 4712 irregular_df_9194
54 v 4715 irregular_df_9194
55 v 4738 linear_df_9700
56 v 4743 irregular_df_9497
57 v 4788 irregular_df_9700
58 u 4304 4340 2
59 u 4304 4461 2
60 u 4304 4601 2
61 u 4304 4629 2
62 u 4304 4788 2
63 u 1821 4715 2
64 u 4310 4715 2
65 u 4316 4337 2
66 u 4316 4340 2
67 u 4316 4393 2
68 u 4316 4601 2

```

Fonte: Autoria própria

6.2.7 Mineração de Grafos e Visualização

Nesta seção ocorre a busca por subestruturas frequentes, a qual é obtida por meio da implementação do FSG. O arquivo gerado na etapa anterior é utilizado como entrada para a ferramenta, então é escolhido um suporte mínimo e outros parâmetros de acordo com a necessidade e especificidade da mineração. Após esta mineração, a ferramenta gera um arquivo de saída (*.fp) apresentando todas as subestruturas frequentes encontradas e a quantidade de ocorrência das subestruturas.

A Figura 33 exibe um trecho do arquivo gerado como saída, o qual apresenta um padrão encontrado entre as ligações dos padrões de desmatamento. A primeira linha da figura contém a letra “t”, a qual caracteriza cada subgrafo encontrado. Nessa mesma linha, em forma de comentário (precedido por “#”), são passadas duas informações relevantes: a primeira apresenta o identificador para o padrão, que neste caso constituiu-se “3-142”, em que o número 3 representa o número de arestas que o grafo contém (denominado tamanho) e 142 indica que foi o 142º grafo deste tamanho a ser encontrado; a segunda informação importante está relacionada à frequência do padrão minerado, ou seja, o valor “19”, que informa que foram encontrados 19 padrões desse mesmo tipo dentre as células consideradas para análise.

Figura 33 - Exibição do padrão de grafo com 3 arestas.

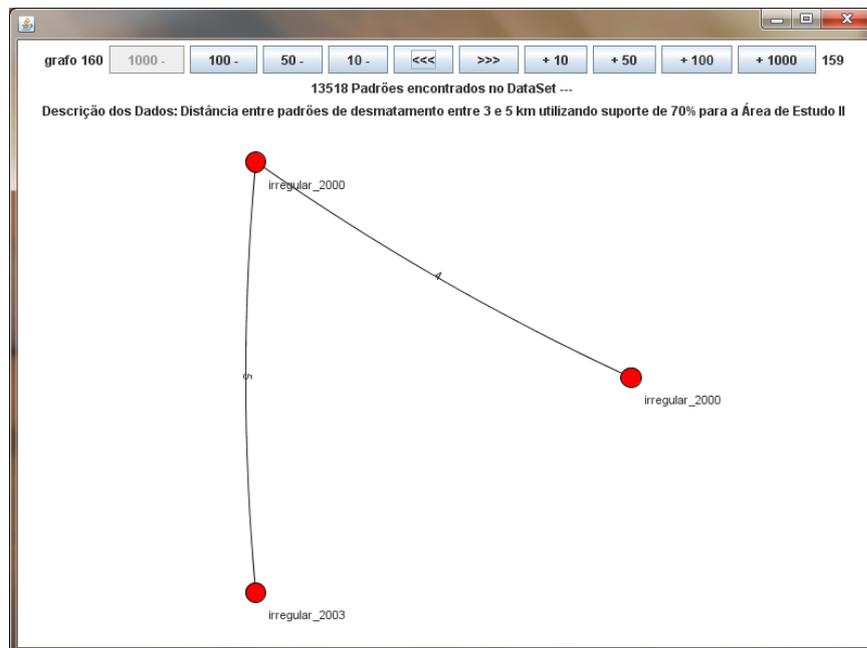
```

1811 t # 3-142, 19
1812 v 0 irregular_df_8891
1813 v 1 irregular_df_9194
1814 v 2 irregular_df_9194
1815 v 3 irregular_df_9700
1816 u 0 1 2
1817 u 0 2 2
1818 u 0 3 2
1819 t # 3-143, 20
1820 v 0 irregular_df_8891
1821 v 1 irregular_df_9194
1822 v 2 irregular_df_9194
1823 v 3 irregular_df_9497
1824 u 0 1 2
1825 u 0 2 2
1826 u 0 3 2

```

Fonte: Autoria própria

Este arquivo gerado foi utilizado como entrada para o FindFRUG, especificamente o visualizador implementado, o qual lê e realiza o mapeamento dos dados para um formato passível de interpretação por meio da API JUNG (BERNSTEIN, 2011) apresentando, por meio de uma visualização gráfica, todas as subestruturas frequentes descobertas contidas no arquivo gerado após a extração dos padrões de subestruturas frequentes (FRANÇA, 2011). Um exemplo de visualização pode ser visto na Figura 34.

Figura 34 - Visualização de um dos padrões frequentes encontrados.

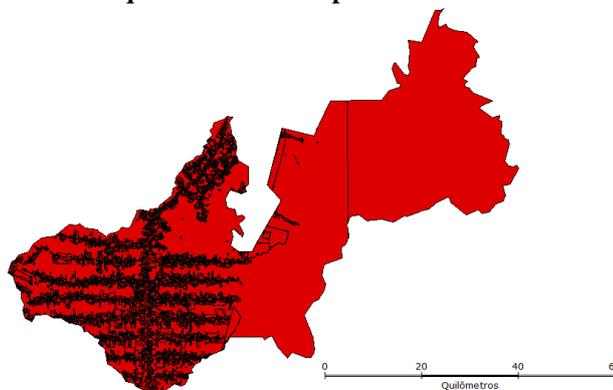
Fonte: Autoria própria

6.3 ESTUDO DE CASO

6.3.1 Área de Estudo

Foi realizada a aquisição de polígonos de desmatamento de 1985 a 2000 para o município de Vale do Anari. Para isto foi necessária a seleção das cenas (231/66 e 231/67) que cobrem o município. A Figura 35 exibe os polígonos contendo os dados relacionados ao projeto de assentamento do Vale do Anari.

Figura 35 - Arquivo formato Shape/ArcView do Vale do Anari.

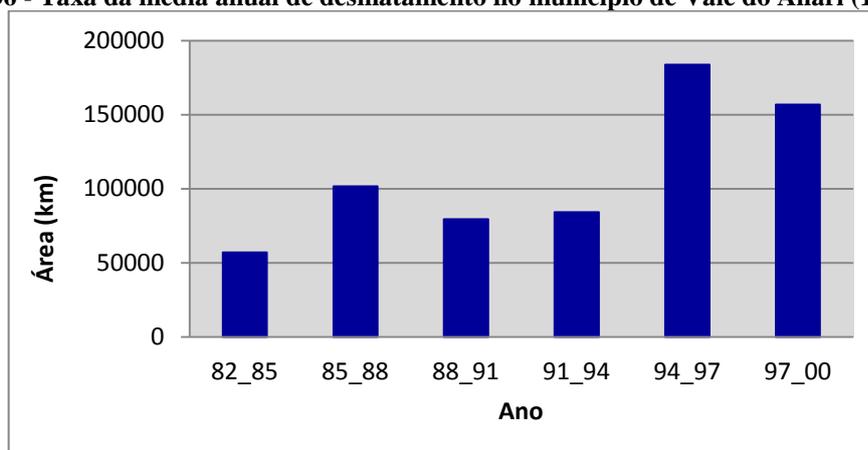


Fonte: (INPE, 2011b)

6.3.2 Evolução do Desmatamento

O processo de desmatamento do lote ocorre quando inicialmente clareiras são abertas na frente do lote, próxima dos ramais e estradas, de forma a possibilitar a ocupação sucessiva de lotes, da frente para o fundo, à medida que ocorre a acessibilidade e melhoria da infraestrutura (ESCADA, 2003).

A Figura 36 apresenta a média anual de desmatamento para períodos de três anos entre 1982 e 2000. Pode-se observar que a área de estudo sofreu um intenso processo de alteração da cobertura de terra nos períodos analisados, mostrando uma evolução de desmatamento ocorrida no município. Em 1982_1985, caracterizado como o período de implantação dos projetos de assentamento, pode-se observar que poucas áreas de floresta foram desmatadas, enquanto que em 1985_1988, com a evolução das ocupações, mais áreas de terra sofreram desmatamento. O período de maior incidência de desmatamento foi em 1994_1997, com o desmatamento de mais hectares em relação ao período anterior. No período de 1988 a 1997, as taxas de desmatamento do município apresentaram evolução similar a da Amazônia Legal, apresentando menores índices entre 1988_1991 e alto crescimento em 1994_1997.

Figura 36 - Taxa da média anual de desmatamento no município de Vale do Anari (1982-2000)

Fonte: Autoria própria

6.3.3 Projetos de Assentamento

O município de Vale do Anari possui dois projetos de assentamento planejado pelo INCRA, que iniciaram em 1982, denominados PA Vale do Anari e PA Machadinho, os quais possuem configuração espacial espinha de peixe e dendrítico, respectivamente, com lotes em média de 50 ha. Esse modelo de ocupação de assentamento planejado e gerenciado pelo INCRA é o modelo predominante no estado de Rondônia (ESCADA, 2003).

O PA Vale do Anari é um projeto de assentamento contíguo, com configuração espacial ortogonal (espinha de peixe). O processo de assentamento é caracterizado por assentamentos de pequenos produtores rurais, onde as agregações são dispersas e ocupam uniformemente toda a extensão da Unidade de Ocupação⁸, estando presentes em quase todos os ramais de assentamento (ESCADA, 2003; ESCADA, MONTEIRO, *et al.*, 2005).

O PA Machadinho tem maior extensão no município de Machadinho D'Oeste, porém existe uma pequena porção na região norte do município de Vale do Anari. Esse PA possui padrões multidirecionais que acompanham a linha de drenagem (dendrítico), com reservas em bloco. Neste processo de assentamento ocorreu a criação de bolsões onde se concentraram pequenas fazendas, com padrões definidos, permitindo que fosse diferenciado e desmembrado dos assentamentos dos quais teve origem. De acordo com pesquisas de campo realizadas, nesta área predominam médios produtores rurais, os quais utilizam estratégias diversificadas de uso de terra, mão de obra e capital (MATOS, ROCHA, *et al.*, 2011; ESCADA, 2003).

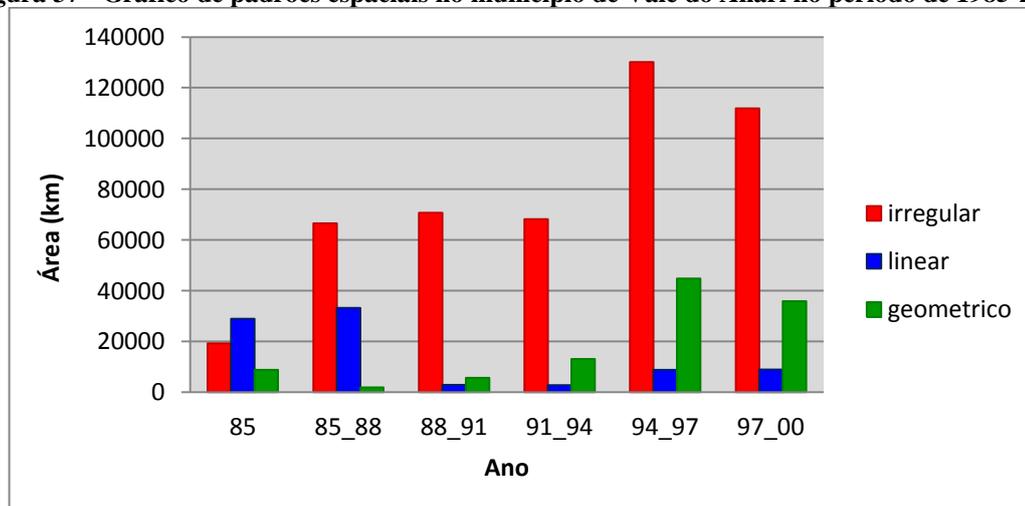
⁸ Unidades de Ocupação (UOP) - áreas irregulares que representam padrões espaço-temporais distintos, diferenciando-as de seu entorno. Por meio de métodos visuais são delimitadas, sobre imagens de satélite e constituídas pela repetição de elementos de textura, definidos como parcelas de desmatamento que possuem propriedades semelhantes, mesma estrutura e configuração espacial (ESCADA, MONTEIRO, *et al.*, 2005).

6.3.4 Classificação Realizada no Município

Com base no estudo realizado por (SILVA, CÂMARA, *et al.*, 2008) para a área do PA Vale do Anari, foi realizado nesta pesquisa uma nova classificação utilizando o *software* GeoDMA. A classificação realizada abordou os dois projetos de assentamento presentes no município, o que resultou em duas áreas de análises.

Para proporcionar uma melhor compreensão da quantidade de polígonos que foram classificados em cada tipo de padrão de desmatamento, os dados foram dispostos no gráfico da Figura 37. Analisando o gráfico pode-se observar que no início da implantação dos projetos de assentamentos, o padrão de desmatamento predominante foi o “*linear*”, caracterizado pela abertura de terras realizadas por colonos ao longo de estradas. A partir do período de 88_91, o padrão “*irregular*” começou a prevalecer no município, ao caracterizar o processo de desmatamento, em que se inicia pela ocupação da frente do lote, que são conectados a estradas, evoluindo até o fundo do lote. Ambos os padrões possuem os mesmos atores em comum: pequenos colonos com uso da terra semelhante. O padrão “*geométrico*” inicia um crescimento progressivo a partir de 1988 até 2000. Este gráfico assemelha-se com o gráfico obtido por (SILVA, 2006) em sua pesquisa, uma vez que apresenta uma mesma evolução temporal de desmatamento no município de Vale do Anari. No entanto, é importante salientar que apesar de tomarmos como base a classificação realizada em sua pesquisa, neste trabalho consideramos os dois projetos de assentamento presentes no município, PA Machadinho e PA Vale do Anari, e não apenas o PA Vale do Anari como abordado na pesquisa citada.

Figura 37 - Gráfico de padrões espaciais no município de Vale do Anari no período de 1985-2000.



Fonte: Autoria própria com base em (SILVA, 2006)

A matriz de confusão é representada por uma matriz quadrada que indica as classificações corretas e erradas, em que as classificações corretas encontram-se na diagonal principal da matriz. Também foi utilizado o estimador de Kappa, o mais adequado por realizar comparações entre classificações, possibilitando a eliminação da subjetividade incluída durante a avaliação. A Figura 38 exibe a matriz de confusão resultante da classificação, o modelo obtido foi comparado a outro modelo, tido como a melhor classificação realizada, e constatou-se que o modelo encontrado realizou satisfatoriamente a classificação dos padrões de desmatamento, indicando um desempenho de 98% na classificação dos dados.

Figura 38 - Matriz de confusão

	0	1	2
0	35	9	0
1	0	5805	0
2	0	0	220

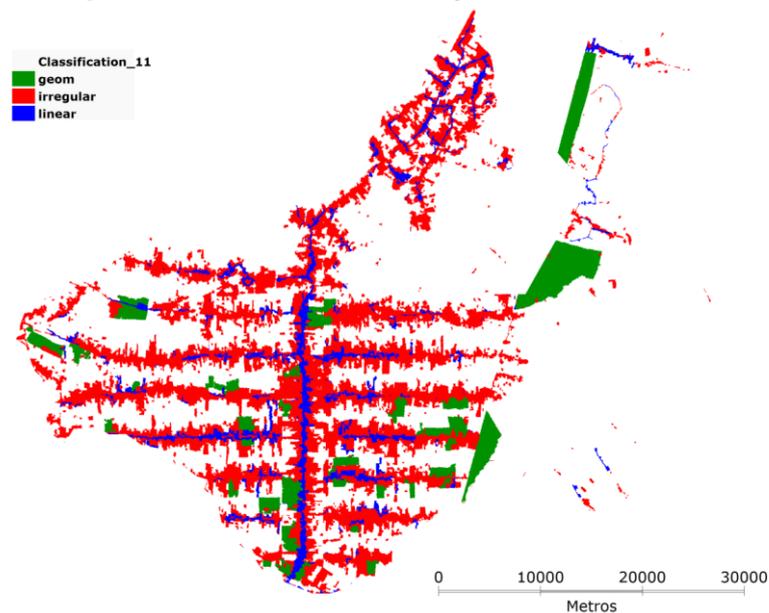
0 -> geom
1 -> irregular
2 -> linear

kappa: 0.981988

Fonte: Autoria própria

A Figura 39 apresenta o resultado dessa classificação em um mapa temático. A diferença do padrão espacial de desmatamento em áreas que possuem diferentes tamanhos de propriedade reflete na diferença que existe entre as estratégias de ocupação e de uso da terra dos diversos atores presentes na região (ESCADA, 2003).

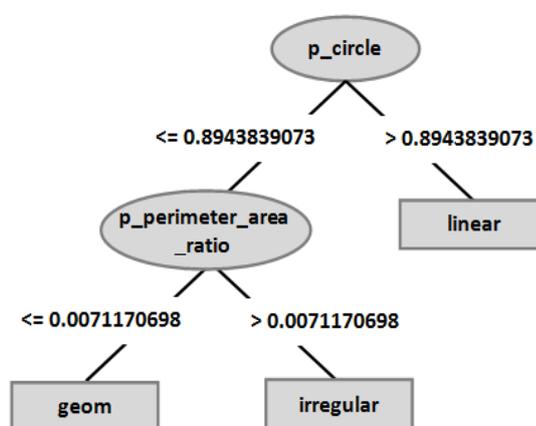
Figura 39 - Resultado da classificação usando o GeoDMA



Fonte: Autoria própria

Na Figura 40 é apresentada a árvore de decisão, não normalizada, gerada pelo algoritmo C4.5 para a classificação dos padrões de desmatamento em todos os períodos que foram analisados. Observou-se que para discriminar os padrões do tipo “geométrico” e “irregular” foram necessários apenas duas métricas *CIRCLE* e *PERIMETER AREA RATIO*. A árvore de decisão gerada é parecida com a árvore de decisão encontrada na pesquisa realizada por (SILVA, 2006), uma vez que as mesmas métricas utilizadas para discriminar os padrões de desmatamento foram idênticas, corroborando, assim, esta pesquisa.

Figura 40 - Árvore de decisão do município de Vale do Anari



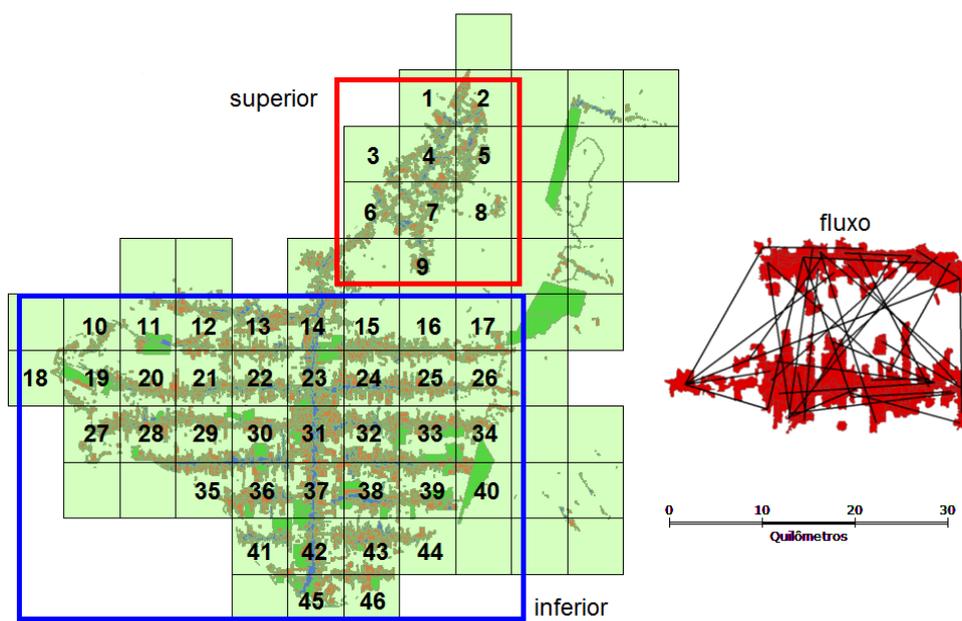
Fonte: Autoria própria

Após a classificação, a área foi dividida em regiões homogêneas que agregassem uma maior quantidade de objetos de desmatamento. Esta divisão foi necessária, pois a mineração de grafos, utilizada neste trabalho, tem por objetivo descobrir subestruturas frequentes presentes em um conjunto de grafos. Dessa forma, cada região representará um grafo e o conjunto dessas regiões representará o conjunto de grafos a ser minerado, em busca de relacionamentos frequentes entre as diversas regiões que compõem o município de Vale do Anari.

Para proporcionar a criação destas regiões foi escolhido, de maneira empírica, um tamanho de célula que melhor agregasse os polígonos de desmatamento. Após exaustivos testes, o tamanho de célula selecionado foi relativo a uma área de 36 km² (células de 6 x 6 km), que melhor atendeu tanto em quantidade de polígonos como na divisão da estrutura espacial da área estudo. Assim, foi possível criar regiões que contemplassem mais elementos, os quais posteriormente foram modelados como grafos e realizou-se a ligação entre os objetos de padrões de desmatamento. A Figura 41 apresenta a divisão do município em regiões (células), a seleção das regiões que foram utilizadas para a criação dos grafos, além de um

exemplo da criação de fluxos entre os padrões de desmatamento. Foram definidas duas áreas para análise, com base na configuração espacial dos dois projetos de assentamento presentes no município: a região superior, em destaque na cor vermelha, está localizado o PA Machadinho e a região inferior, destacada na cor azul, o PA Vale do Anari.

Figura 41 - Seleção de polígonos de desmatamento, após a classificação, delimitados em células e criação do fluxo entre todos os objetos de desmatamento de uma região.



Fonte: Autoria própria

Após a delimitação das células foi criado um plano de informação para cada região utilizando o *software* TerraView e, por conseguinte, foi criado o fluxo entre cada objeto de padrão de desmatamento contido nos limites das regiões consideradas. Como resultado da ligação entre os objetos de desmatamento, foi gerado um atributo contendo as distâncias entre os padrões, o qual foi utilizado como dado preponderante na mineração de grafos.

Relações envolvendo os diferentes tipos de padrões de desmatamento foram encontradas junto com o ano de ocorrência e a distância entre os objetos. As distâncias empregadas foram escolhidas com base na estrutura fundiária de cada lote, que tem 250 x 2000 m de extensão. Dessa forma as distâncias apresentadas mostram relações de vizinhança entre os lotes, além de apresentar a evolução temporal da ocupação dos mesmos. O arquivo contendo as ligações entre os objetos de desmatamento foi minerado com diferentes tipos de suporte mínimo variando entre um percentual de 10% a 100%, visando com isto descobrir uma quantidade relevante de subestruturas frequentes entre as diferentes regiões (células) selecionadas.

7 RESULTADOS

Este capítulo apresenta os resultados alcançados com a aplicação do método proposto. Os resultados são apresentados separadamente para cada projeto de assentamento presente no município, PA Machadinho e PA Vale do Anari, respectivamente, região superior e inferior.

7.1 SUBESTRUTURAS FREQUENTES – PA MACHADINHO

7.1.1 Padrão Irregular

Após análises realizadas com os resultados das diferentes distâncias que foram testadas, verificou-se que as distâncias de 250 m e 500 m melhor representavam as relações de vizinhança entre os padrões de desmatamento, por causa do tamanho dos lotes.

Com a distância de 250 m foi visto que os padrões irregulares do período de 94_97 sempre estavam conectados a outros padrões do início da ocupação como 85, 85_88, 88_91 e 94_97. A frequência desse tipo de relacionamento ocorreu porque existe um limiar de desmatamento muito alto no período de 94_97, modificando a lógica da ocupação do lote que é realizada da frente do lote para o fundo. O padrão irregular do período de 94_97 era esperado, uma vez que houve uma grande quantidade de desmatamento neste período em todo o município e, conseqüentemente, no PA Machadinho.

Como pode ser observado na Tabela 4, utilizando o suporte de 40%, o qual proporcionou maior presença de subestruturas frequentes de padrões irregulares para o PA Machadinho, pode-se observar uma grande presença de padrões de períodos anteriores ligados ao de 94_97. O padrão “*irregular_85_88*” conectado ao “*irregular_88_91*” apresenta um processo recorrente, mostrando a evolução temporal da ocupação de um lote que, ocorre em anos consecutivos. A mineração de grafos pode confirmar este padrão temporal utilizando para isto as subestruturas frequentes.

Utilizando a distância de 500 m observa-se na Tabela 5, com suporte de 90%, uma ampla presença de padrões do tipo irregular entre 85_88 e 94_97 ligados aos irregulares de 94_97 e 97_00. Isto indica uma característica de ocupação do lote, pois com uma distância de 500 m entre um objeto e outro, padrões mais novos podem ser detectados já que estes, muitas vezes, surgem após a frente do lote estar totalmente ocupada. O padrão “*irregular_85_88*” conectado ao “*irregular_88_91*” e “*irregular_88_91*” ligado ao “*irregular_91_94*”,

semelhantemente ao que ocorreu com a distância de 250 m, apresentam uma evolução temporal de maneira contínua.

Tabela 4 - Distância de 250 m para detecção de padrões Irregulares ligados a Irregulares
Número de subestruturas frequentes com distância de 250 m e
suporte de 40% - 9 células

Ano		Irregular					
		85	85_88	88_91	91_94	94_97	97_00
Irregular	85					4	
	85_88			4		4	
	88_91					5	4
	91_94						
	94_97					6	
	97_00						

Fonte: Autoria própria

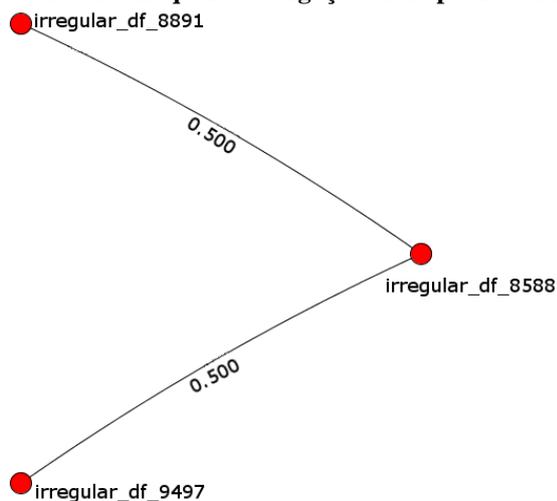
Tabela 5 - Distância de 500 m para detecção de padrões Irregulares ligados a Irregulares
Número de subestruturas frequentes com distância de 500 m e
suporte de 90% - 9 células

Ano		Irregular					
		85	85_88	88_91	91_94	94_97	97_00
Irregular	85						
	85_88			9		9	
	88_91				9	8	8
	91_94					9	8
	94_97					9	9
	97_00						

Fonte: Autoria própria

As distâncias de 1, 2 e 4 km não foram consideradas para a parte superior do município por esta apresentar subestruturas heterogêneas ou mesmo não conter subestruturas frequentes relevantes, uma vez que a ocupação do lote se dá de forma contínua. Porém, em alguns casos, principalmente nas extremidades do projeto de assentamento, esta ocupação ocorre de forma desordenada, devido a fatores ambientais e de acessibilidade. A Figura 42 apresenta um dos padrões mais frequentes para a região superior com presença em 8 das 9 células selecionadas. Esta subestrutura frequente mostra duas relações existentes: primeiro, um padrão temporal na evolução do desmatamento do tipo irregular nos períodos de 85_88 e 88_91; e segundo, os irregulares de 85_88 e 94_97, mostram que o período de 94_97 teve altas taxas de desmatamento e que áreas que não tinham sido desmatadas anteriormente, foram desmatadas no período de 94_97.

Figura 42 - Subestrutura frequente de ligação entre padrões irregulares



Fonte: Autoria própria

7.1.2 Padrão Linear

Para os padrões do tipo linear, caracterizados como estradas abertas para acesso a lotes, houve a existência de subestruturas frequentes bastante heterogêneas. Após diferentes testes, a distância de 500 m foi considerada a mais adequada para uma análise do fenômeno na região, conforme Tabela 6. A presença destas subestruturas frequentes reflete o tipo de ocupação existente no local, ou seja, em todos os períodos com o padrão linear, este padrão sempre se liga a um padrão irregular mais recente, principalmente 94_97 e 97_00. Este resultado ocorreu devido à caracterização do padrão linear, ou seja, o padrão linear está ligado aos ramais e surge antes dos padrões irregulares, pois este irá dar acesso aos lotes e proporcionar o surgimento de outros padrões, auxiliando no desenvolvimento e ocupação dos lotes. Devido a esta característica as subestruturas frequentes descobertas mostraram-se heterogêneas quanto à sua distribuição espacial, ou seja, não há dominância de relações entre padrões de desmatamento. As subestruturas frequentes descobertas apresentaram relação temporal quanto ao surgimento do padrão irregular ligado a padrões lineares anteriores ou atuais. Além disso, foi observado que existe uma trajetória temporal da frente para o fundo de forma mais linear no tempo.

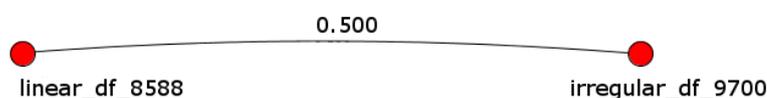
A Figura 43 mostra a subestrutura frequente descoberta com maior ocorrência em 5 células das 9 analisadas, apresentando uma relação temporal entre um padrão linear e o posterior surgimento do padrão irregular.

**Tabela 6 - Distância de 500 m para detecção de padrões Lineares ligados a Irregulares
Número de subestruturas frequentes com distância de 500 m e
suporte de 40% - 9 células**

Ano	Irregular					
	85	85_88	88_91	91_94	94_97	97_00
Linear	85			3	3	
	85_88			3	4	5
	88_91		3		3	3
	91_94			3		
	94_97					3
	97_00		3	3	3	

Fonte: Autoria própria

Figura 43 - Subestrutura frequente relacionando a ligação entre um padrão linear a um irregular



Fonte: Autoria própria

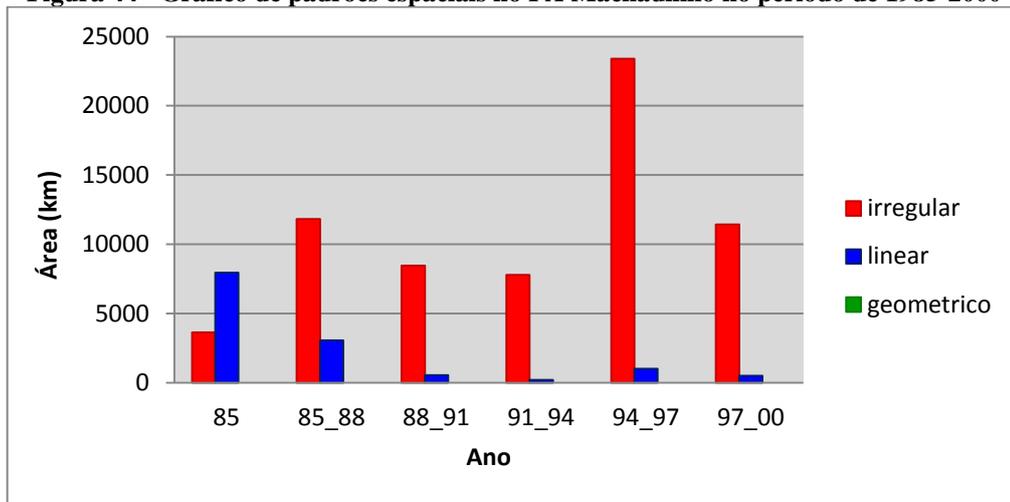
O padrão geométrico não foi analisado para o PA Machadinho do município, pois não existem padrões deste tipo nesta região.

7.1.3 Análise dos Resultados PA Machadinho

No início da implantação do projeto de assentamento, o padrão de desmatamento dominante foi o “*linear*”, que se caracteriza pela abertura de terras ao longo de estradas por colonos. A partir do período de 85_88, o padrão “*irregular*” teve um crescimento no desmatamento, porém houve uma queda nos períodos seguintes, 88_91 e 91_94. No entanto, no período de 94_97 ocorreu uma grande incidência de desmatamento no projeto de assentamento, reflexo da alta taxa de desmatamento ocorrida em toda Amazônia Legal, porém no período de 97_00 teve uma queda de mais 40% de desmatamento em relação ao período anterior, como mostra a Figura 44.

Esse processo de desmatamento no PA Machadinho pode ser observado nas subestruturas frequentes descobertas, uma vez que estas apresentaram mais dados irregulares do período de 94_97, além de apresentarem padrões lineares do início do projeto de assentamento ligados a padrões irregulares mais recentes, mostrando com isso que regiões que não foram ocupadas em períodos anteriores, foram ocupadas posteriormente.

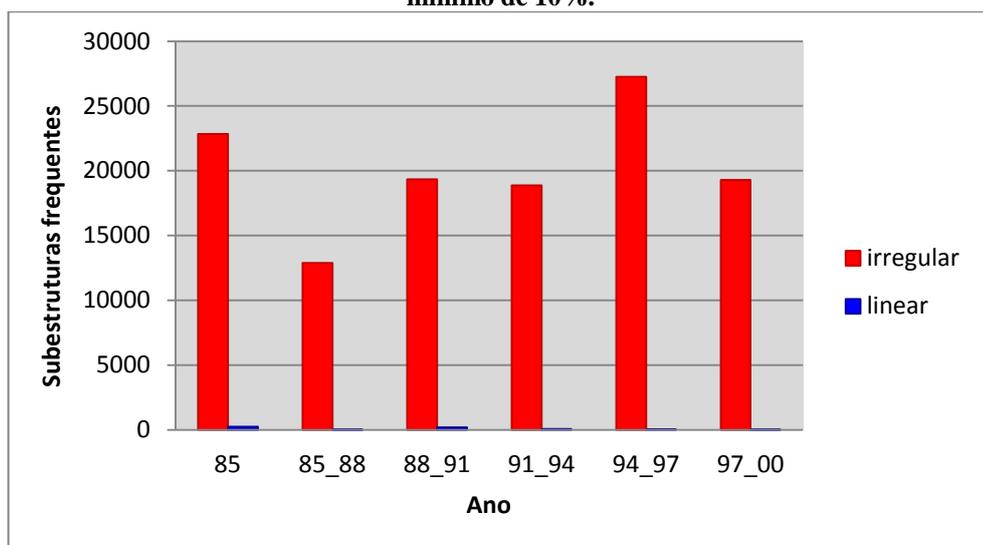
Figura 44 - Gráfico de padrões espaciais no PA Machadinho no período de 1985-2000



Fonte: Autoria própria

A Figura 45 apresenta um gráfico de frequência das subestruturas frequente descobertas. Os dados foram minerados utilizando um suporte mínimo de 10% para a distância de 500 m, no PA Machadinho. Pode-se observar que apesar dos períodos de 85, 88_91, 91_94 e 97_00 apresentarem grande quantidade de subestruturas descobertas, essas subestruturas frequentes não indicam padrões relevantes, ou seja, são subestruturas frequentes que não ocorrem em mais de uma região do PA Machadinho. Quanto menor o suporte mínimo utilizado na mineração, menos subestruturas frequentes relevantes serão descobertas.

Figura 45 - Gráfico de frequência de subestruturas frequentes para distância de 500 m, utilizando suporte mínimo de 10%.



Fonte: Autoria própria

7.2 SUBESTRUTURAS FREQUENTES – PA VALE DO ANARI

7.2.1 Padrão Irregular

Análises foram realizadas com diferentes distâncias para o PA Vale do Anari. Observou-se que as distâncias de 250 m, 500 m e 1 km eram as que melhor apresentavam a relação entre os objetos de desmatamento, levando em consideração os padrões irregulares conectados entre si.

Analisando a distância de 250 m foi verificado que os padrões irregulares do período de 94_97 e 97_00 possuíam alta frequência de relações que surgiram com padrões após a década de 90. Este fato não significa que a região passou a ser ocupada após o período da década de 90, mas que houve uma maior ocupação de terra posterior ou no decorrer deste período, processo que foi refletido pela quantidade de subestruturas frequentes descobertas, conforme Tabela 7. A quantidade de padrões também reflete a alta incidência de desmatamentos neste período.

Tabela 7 - Distância de 250 m para detecção de padrões Irregulares ligados a Irregulares

Número de subestruturas frequentes com distância de 250 m e suporte de 30% - 36 células							
Ano		Irregular					
		85	85_88	88_91	91_94	94_97	97_00
Irregular	85						
	85_88						
	88_91				17	22	
	91_94					14	16
	94_97						21
	97_00						11

Fonte: Autoria própria

As distâncias de 500 m e 1 km mostram uma grande quantidade de subestruturas frequentes nos anos de 94_97 e 97_00 ligadas a todos os períodos posteriores a 85, o que indica uma maior ocupação nesta época, uma vez que para esta análise não é levada em conta apenas uma parte do PA Vale do Anari e sim todas as células pertencentes a essa região (Tabela 8 e Tabela 9). Desta forma, os lotes localizados nas extremidades do “*padrão espinha de peixe*” (configuração espacial formada pela totalidade dos padrões irregular, linear e geométrico), influenciam na ocorrência de subestruturas frequentes, pois nas pontas desta

estrutura a ocupação foi mais lenta devido as más condições de estradas e até mesmo a maior distância do núcleo urbano, refletindo assim no tipo de subestruturas frequentes.

Tabela 8 - Distância de 500 m para detecção de padrões Irregulares ligados a Irregulares
Número de subestruturas frequentes com distância de 500 m e
suporte de 90% - 37 células

Ano		Irregular					
		85	85_88	88_91	91_94	94_97	97_00
Irregular	85						
	85_88			33		34	33
	88_91					33	35
	91_94					36	
	94_97					34	37
	97_00						35

Fonte: Autoria própria

Tabela 9 - Distância de 1 km para detecção de padrões Irregulares ligados a Irregulares
Número de subestruturas frequentes com distância de 1 km e
suporte de 80% - 37 células

Ano		Irregular					
		85	85_88	88_91	91_94	94_97	97_00
Irregular	85						
	85_88						31
	88_91				32	33	32
	91_94					33	32
	94_97					32	33
	97_00						33

Fonte: Autoria própria

7.2.2 Padrão Linear

Considerando os padrões lineares ligados a padrões irregulares, verifica-se que a melhor distância que caracteriza o relacionamento entre estes padrões foi a de 500 m, pois as subestruturas frequentes descobertas apresentam um arranjo que corresponde ao fenômeno que ocorre na região, ou seja, os padrões lineares de 94_97 ligados a padrões irregulares de 91_94, 94_97 e 97_00, mostrando um processo de desmatamento à medida que a ocupação é realizada nas proximidades do lote e também da frente do lote para o seu limite posterior, como mostra a Tabela 10.

Tabela 10 - Distância de 500 m para detecção de padrões Lineares ligados a Irregulares
Número de subestruturas frequentes com distância de 500 m e
suporte de 30% - 37 células

Ano		Irregular					
		85	85_88	88_91	91_94	94_97	97_00
Linear	85						
	85_88				12		
	88_91						
	91_94						
	94_97		12		12	15	14
	97_00						

Fonte: Autoria própria

7.2.3 Padrão Geométrico

O padrão geométrico apresentou resultados melhores quando aplicado à distâncias maiores. Na criação de fluxo entre os padrões de desmatamento é gerada uma linha que conecta os centroides de dois polígonos. Como o padrão geométrico, possui uma área de polígono extensa, então distâncias pequenas não conseguem realizar a ligação deste padrão com padrões vizinhos, sendo necessário utilizar distâncias maiores.

Para o PA Vale do Anari foram consideradas distâncias de 2 e 4 km, resultando em subestruturas frequentes com padrões que mostram um desenvolvimento estrutural da área, apresentando ligações entre padrões geométricos e irregulares. Em ambas as distâncias pode-se observar que o padrão geométrico de 97_00 sempre estava relacionado a padrões irregulares de 88_91, 91_94, 94_97 e 97_00 (Tabela 11). O padrão “geométrico_97_00” teve forte presença devido ao fato de que este possuiu uma maior expansão no triênio de 94_00 e 97_00.

Tabela 11 - Distância de 4 km para detecção de padrões Geométricos ligados a Irregulares
Número de subestruturas frequentes com distância de 4 km e
suporte de 20% - 35 células

Ano		Irregular					
		85	85_88	88_91	91_94	94_97	97_00
Geométrico	85						
	85_88						
	88_91						
	91_94					7	
	94_97			7	7		8
	97_00			8	7	8	11

Fonte: Autoria própria

7.3 DISCUSSÃO SOBRE OS RESULTADOS OBTIDOS

Para este trabalho foram utilizados dados do município de Vale do Anari que possui dois projetos de assentamento planejado pelo INCRA, PA Machadinho e PA Vale do Anari. Este município foi escolhido por conter características relacionadas ao desmatamento e também por ser uma área com estudos relevantes, possibilitando assim proporcionar uma interpretação e validação da mineração realizada.

Conforme discutido por (PEREIRA, ESCADA e RENNO, 2007), os desmatamentos influenciam a dinâmica da paisagem, pois podem gerar ou ampliar áreas adjacentes de desmatamento. Considerando ainda que (SILVA, CÂMARA, *et al.*, 2008) realiza uma associação dos padrões de desmatamento a atores específicos, é possível verificar a relevância da investigação do processo que determina a ampliação do desmatamento a partir da influência de atores e suas interações. Tal análise, até o momento, dispunha apenas de recursos convencionais de análise de dados e imagens, exceto pelos meios providos pelo GeoDMA para a mineração de imagens de sensoriamento remoto. Entretanto, mesmo o GeoDMA não dispõe de tecnologia suficiente para explicitar e minerar desmatamentos e suas ligações para determinar padrões, especialmente numa perspectiva estrutural (grafos) conforme abordado neste trabalho. Diante deste contexto, esta pesquisa avança nesta área provendo recursos para identificação de padrões de grafos em dados de sensoriamento remoto.

Por meio da análise de especialistas em desmatamento, as subestruturas descobertas puderam contar um pouco da história da ocupação realizada no PA Vale do Anari, podendo confirmar a relação temporal na evolução dos padrões ao mostrar a relação de um padrão anterior a um ano posterior, ou de mesmo ano, que estejam presentes nas imediações. Estes resultados puderam mostrar que a ocupação de um lote ocorre de maneira contínua, quando apresentou relações entre padrões temporalmente próximos.

Um fator observado ao se utilizar a metodologia desenvolvida está relacionado ao tamanho da área de análise, pois o emprego deste método em áreas específicas proporciona uma maior probabilidade de encontrar subestruturas frequentes que corroborem com a realidade existente no local. Por exemplo, neste trabalho foi utilizada toda a região inferior do município para uma das análises, porém devido à grande heterogeneidade do desenvolvimento da ocupação na região, muitos padrões foram influenciados pela grande quantidade de objetos de certo período de tempo. Entretanto quando aplicado, por exemplo, a um conjunto de regiões (células) específicas, como o centro do município (que possui uma

evolução de ocupação mais ordenada) poderia apresentar através desta metodologia o desenvolvimento temporal entre os diferentes padrões existentes, mostrando assim subestruturas frequentes com períodos mais próximos.

Outro fator observado com os resultados alcançados e analisados é referente ao período de 85 para a região inferior do município de Vale do Anari. Foi realizada uma quantificação do número de subestruturas frequentes que continham padrões do tipo irregular, linear e geométrico para cada período, utilizando as distâncias de 250 m e 500 m, minerado com o menor suporte mínimo possível (30% para 250 m e 10% para 500 m). Como resultado da análise foi verificado que o padrão “*irregular_85*” possuía a maior quantidade de elementos presentes nas subestruturas frequentes descobertas. No entanto, não houve subestruturas frequentes com maior frequência (utilizando um suporte mínimo elevado), entre as regiões (células) selecionadas, que contivessem este padrão como elemento principal, uma vez que os padrões “*irregular_94_97*” e “*irregular_97_00*” também apresentaram uma forte presença entre as subestruturas descobertas. Utilizando o menor suporte, os mesmos padrões apresentaram forte frequência utilizando um suporte mínimo elevado. A não ocorrência desse padrão entre as subestruturas frequentes é porque ele funciona como o linear, é um artifício do dado (ESCADA, 2003).

8 CONCLUSÕES E TRABALHOS FUTUROS

Representar dados de sensoriamento remoto por intermédio de grafos constitui-se em uma estratégia atraente e inovadora, uma vez que a utilização de formalismos matemáticos (grafos) faz parte de uma área de pesquisa bem fundamentada e com diversas ferramentas computacionais e aplicações relevantes. Assim, testar e desenvolver uma abordagem, na qual grafos possam ser associados aos dados de sensoriamento remoto promove uma nova perspectiva para a descoberta de relações entre padrões.

Este trabalho desenvolveu uma abordagem utilizando mineração de grafos para a busca de subestruturas frequentes em dados de sensoriamento remoto, nos quais os padrões de desmatamento sejam relacionados com outros objetos de desmatamento em conjunto com o ano de ocorrência do padrão. Os resultados obtidos puderam confirmar fenômenos recorrentes da região, dentre eles podemos citar: a grande presença de subestruturas frequentes no período 1994_1997, ano em que houve uma grande quantidade de áreas desmatadas no município de Vale do Anari; a relação temporal entre os padrões em pequenas distâncias, revelando, assim, as ocupações em lotes em que desmatamentos são realizados de maneira contínua.

O método proposto é relevante para a análise e diagnóstico de relações temporais entre padrões de desmatamento. A aplicação desta proposta em dados reais de sensoriamento remoto foi realizada utilizando uma área para estudo de caso, cujos dados foram gentilmente fornecidos pela pesquisadora Isabel Escada, a qual também subsidiou o entendimento das subestruturas frequentes descobertas devido ao seu conhecimento e pesquisas realizadas na região de estudo.

Como trabalhos futuros espera-se realizar a aplicação deste método em áreas específicas do município de Vale do Anari e assim descobrir subestruturas frequentes que mostrem uma relação de maior proximidade temporal entre os padrões de desmatamento, uma vez que esta pesquisa abordou todo município, sendo fortemente influenciada por padrões que ocorreram no período de 94_97. Além disso, espera-se que padrões e resultados obtidos nesta pesquisa possam indicar processos relevantes em estudos futuros.

A partir da metodologia desenvolvida e dos resultados alcançados, foram aceitos dois artigos em conferências internacionais: “The Fourth International Conference on Geographic Object-Based Image Analysis” (GeoBIA) e “IEEE International Geoscience and Remote Sensing Symposium” (IGARSS).

REFERÊNCIAS BIBLIOGRÁFICAS

- ABREU, E. S. Divisão de Processamento de Imagens - DPI/INPE. **TerraView**, 2011. Disponível em: <<http://www.dpi.inpe.br/terraview/php/dow.php?body=plgFlow>>. Acesso em: 14 ago 2011.
- AGGARWAL, C. C.; WANG, H. **Managing and Mining Graph Data**. New York: Spring, 2010. 620 p.
- AGRAWAL, R.; SRIKANT, R. Fast Algorithms for Mining Association Rules in Large Databases. **Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94)**, San Francisco, CA, USA, p. 13, 1994.
- APPEL, A. P. **Métodos para o pré-processamento e mineração de grandes volumes de dados multidimensionais e redes complexas**. Tese (Doutorado em Ciências - Ciências de Computação e Matemática Computacional) - Universidade de São Paulo - USP. São Carlos, p. 180. 2010.
- BEKALO, M. T. **Spatial Metrics and Landsat Data for Urban Landuse Change Detection in Addis Ababa, Ethiopia**. Dissertação (Mestrado de Ciência em Tecnologias Geoespaciais) - Universitat Jaume I. Castellon, Spain, p. 89. 2009.
- BERNSTEIN, G. **JUNG 2.0 Tutorial**, 2011. Disponível em: <<http://www.grottonetworking.com/JUNG/JUNG2-Tutorial.pdf>>. Acesso em: 21 jun. 2011.
- CÂMARA, G.; MONTEIRO, A. M.; FUNCKS, S. D.; CARVALHO, M. S. Análise Espacial e Geoprocessamento. **Análise Espacial de Dados Geográficos**, 2004. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/analise/cap1-intro.pdf>>. Acesso em: 6 set. 2011. cap. 1.
- CÂMARA, G.; DAVIS, C. Introdução. **Introdução à Ciência da Geoinformação**, 2001. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/introd/cap1-introducao.pdf>>. Acesso em: 6 set. 2011.
- CÂMARA, G.; MONTEIRO, A. M. V. Conceitos Básicos em Ciência da Geoinformação. **Introdução à Ciência da Geoinformação**, 2001. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/introd/cap2-conceitos.pdf>>. Acesso em: 6 set. 2011. cap. 2.
- CÂMARA, G.; QUEIROZ, G. R. Arquiteturas de Sistemas de Informação Geográfica. **Introdução à Ciência da Geoinformação**, 2001. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/introd/cap3-arquitetura.pdf>>. Acesso em: 6 set 2011. cap. 3.
- CARVALHO, M. A. G. **Teoria dos Grafos: uma introdução**, 2005. Disponível em: <http://http://www.4shared.com/document/mgX5KE3E/TEORIA_DOS_GRAFOS_UMA_INTRODUC.html>. Acesso em: 21 nov. 2009.

- CCRS. Tutorial: Fundamentals of Remote Sensing. **Natural Resources Canada - Canada Centre for Remote Sensing**, 2007. Disponível em: <http://www.nrcan.gc.ca/sites/www.nrcan.gc.ca/earth-sciences/files/pdf/resource/tutor/fundam/pdf/fundamentals_e.pdf>. Acesso em: 20 nov. 2011.
- COOK, D. J.; HOLDER, L. B. Graph-Based Data Mining. **IEEE Intelligent Systems**, Piscataway, NJ, USA, v. 15, p. 32-41, March 2000.
- COOK, D. J.; HOLDER, L. B. **Mining Graph Data**. Hoboken, New Jersey: Wiley-Interscience, 2007. 502 p.
- CORMEN, T. H.; LEISERSON, C. E.; STEIN, C.; RIVEST, R. L. **Algoritmos: Teoria e Prática**. Rio de Janeiro: Elsevier, 2002. 936 p.
- DIESTEL, R. **Graph Theory**. 3. ed. New York: Springer, 2005.
- DOGRUSOZ, E.; AKSOY, S. Modeling urban structures using graph-based spatial patterns. **IEEE International Geoscience & Remote Sensing Symposium (IGARSS '07)**, p. 4826-4829, 2007.
- ERSAHIN, K.; CUMMING, I. G.; WARD, R. K. Segmentation and Classification of Polarimetric SAR Data Using Spectral Graph Partitioning. **IEEE Transactions on Geoscience and Remote Sensing**, v. 48, p. 164 -174, jan. 2010.
- ESCADA, M. I. S. **Evolução de Padrões da Terra na Região Centro-Norte de Rondônia**. Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais - INPE. São José dos Campos. 2003.
- ESCADA, M. I. S.; MONTEIRO, A. M. V.; AGUIAR, A. P. D.; CARNEIRO, T. G. S.; CÂMARA, G. Análise de padrões e processos de ocupação para a construção de modelos na Amazônia: Experimentos em Rondônia. **XII Simpósio Brasileiro de Sensoriamento Remoto (SBSR)**, Goiânia, GO, Brasil, p. 2973-2983, 2005.
- FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery. **AI Magazine**, Cambridge, v. 17, p. 37-54, 1996. Disponível em: <<http://www.aaai.org/aitopics/assets/PDF/AIMag17-03-2-article.pdf>>. Acesso em: 20 out. 2011.
- FELZENSZWALB, P. F.; HUTTENLOCHER, D. P. Efficient Graph-Based Image Segmentation. **International Journal of Computer Vision**, Hingham, MA, USA , 2004.
- FEOFILOFF, P. Análise de Algoritmos. **Complexidade computacional e problemas NP-completos**, 2011. Disponível em: <http://www.ime.usp.br/~pf/analise_de_algoritmos/aulas/NPcompleto2.html>. Acesso em: 4 out. 2011.
- FEOFILOFF, P.; KOHAYAKAWA, Y.; WAKABAYASHI, Y. **Uma Introdução Sucinta à Teoria dos Grafos**, 12 jul. 2011. Disponível em:

<<http://www.ime.usp.br/~pf/teoriadosgrafos/texto/TeoriaDosGrafos.pdf>>. Acesso em: 20 out. 2011.

FERRAZ, S. F. B.; VETTORAZZI, C. A.; THEOBALD, D. M.; BALLESTER, M. V. R.. Landscape dynamics of Amazonian deforestation between 1984 and 2002 in central Rondônia, Brazil: assessment and future scenarios. **Forest Ecology and Management**, v. 204, p. 67-83, 2005.

FOLDOC. Application Program Interface. **Free On-Line Dictionary of Computing**, 2010. Disponível em: <<http://foldoc.org/Application+Program+Interface>>. Acesso em: 8 set. 2011.

FORMAN, R. T. T. **Land mosaics: the ecology of landscapes and regions**. Cambridge. Cambridge University Press, p. 632. 1995. ISBN (0 521 47462 0).

FORTIN, J.; ZIELÍNSKI, P.; DUBOIS, D.; FARGIER, H. Criticality analysis of activity networks under interval uncertainty. **Journal of Scheduling**, Kluwer Academic Publishers Hingham, MA, USA, v. 13, p. 609--627, Dez. 2010.

FRANÇA, F. S. D. **Enriquecimento Semântico de Padrões Minerados em Grafos**. Dissertação (Mestrado em Ciência da Computação) - Universidade do Estado do Rio Grande do Norte e Universidade Federal Rural do Semiárido. Mossoró. 2011.

GEOSERVER. **GeoServer**, 2010. Disponível em: <<http://geoserver.org/display/GEOS/Welcome>>. Acesso em: 21 set. 2010.

GOLDSCHMIDT, R.; PASSOS, E. **Data Mining: Um Guia Prático: Conceitos, Técnicas, Ferramentas, Orientações e Aplicações**. Rio de Janeiro: Elsevier, 2005. 255 p.

GRACIANO, A. B. V. **Rastreamento de objetos baseado em reconhecimento estrutural de padrões**. Dissertação (Mestrado em Ciências - Ciência da Computação) - Instituto de Matemática e Estatística, Universidade de São Paulo - USP. São Paulo, p. 130. 2007.

GRAPHVIZ. Graph Visualization. **Graphviz - Graph Visualization Software**, 2011. Disponível em: <<http://www.graphviz.org/About.php>>. Acesso em: 6 out. 2011.

GUO, D. Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP). **International Journal of Geographical Information Science**, Bristol, PA, USA, v. 22, p. 801-823, January 2008.

GUO, D.; GAHEGAN, M.; MACEACHREN, A. M.; ZHOU, B. Multivariate Analysis and Geovisualization with an Integrated Geographic Knowledge Discovery Approach. **Cartography and Geographical Information Science**, v. 32, p. 113-132, 2005.

GUO, D.; CHEN, J.; MACEACHREN, A. M.; LIAO, K.. A Visualization System for Space-Time and Multivariate Patterns (VIS-STAMP). **IEEE Transactions on Visualization and Computer Graphics**, Piscataway, NJ, USA, v. 12, p. 1461-1474, November 2006.

HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. 2. ed. San Francisco: Morgan Kaufmann, 2006. 772 p.

- HAN, J.; KOPERSKI, K.; STEFANOVIC, N. GeoMiner: a system prototype for spatial data mining. **Proceedings of the 1997 ACM SIGMOD international conference on Management of data (SIGMOD '97)**, New York, NY, USA, p. 553-556, 1997.
- HOLDER, L.; COOK, D.; GONZALEZ, J.; JONYER, I. Structural Pattern Recognition in Graphs. In: DECHANG CHEN, X. C. **Pattern recognition and string matching**. Netherlands: Springer, v. 13, 2002. p. 763.
- HUAN, J.; WANG, W.; PRINS, J. Efficient Mining of Frequent Subgraphs in the Presence of Isomorphism. **Proceedings of the Third IEEE International Conference on Data Mining (ICDM'03)**, Washington, DC, USA, p. 549-552, 2003.
- IAP. Corte raso. **IAP - Instituto Ambiental do Paraná**, 2010. Disponível em: <<http://www.iap.pr.gov.br/modules/conteudo/conteudo.php?conteudo=5>>. Acesso em: 26 out. 2011.
- IBGE. Instituto Brasileiro de Geografia e Estatística. **IBGE**, 2011a. Disponível em: <<http://www.ibge.gov.br>>. Acesso em: 17 jul. 2001.
- IBGE. Instituto Brasileiro de Geografia e Estatística (IBGE). **Censo 2010**, 2011b. Disponível em: <<http://www.censo2010.ibge.gov.br>>. Acesso em: 17 jul. 2011.
- INOKUCHI, A.; WASHIO, T.; MOTODA, H. **An Apriori-Based Algorithm for Mining Frequent Substructures from Graph Data**. London, UK: Springer-Verlag. 2000. p. 13-23.
- INPE. Divisão de Geração de Imagens (DGI). **Instituto Nacional de Pesquisas Espaciais (INPE)**, 2011. Disponível em: <<http://www.dgi.inpe.br>>. Acesso em: 21 fev 2012.
- INPE. Projeto PRODES. **Monitoramento da Floresta Amazônica Brasileira por Satélite**, 2011b. Disponível em: <<http://www.obt.inpe.br/prodes>>. Acesso em: 26 jun. 2011.
- JIANG, C.; COENEN, F.; ZITO, M. Frequent sub-graph mining on edge weighted graphs. **Proceedings of the 12th international conference on Data warehousing and knowledge discovery (DaWaK'10)**, Bilbao, Spain, p. 77--88, 2010.
- JUNG. **JUNG - Java Universal Network/Graph Framework**, 2010. Disponível em: <<http://jung.sourceforge.net/index.html>>. Acesso em: 14 abr 2011.
- KARWATH, A.; RAEDT, L. D. **Predictive Graph Mining**. Berlin, Heidelberg: Springer-Verlag. 2004. p. 1-15.
- KARYPIS, G. PAFI Software Package for Finding Frequent Patterns in Diverse Datasets. **Karypis Lab**, 2003. Disponível em: <<http://glaros.dtc.umn.edu/gkhome/pafi/overview>>. Acesso em: 04 abr. 2011.
- KORTING, T. S. GeoDMA - Geographical Data Mining Analyst. **Divisão de Processamento de Imagens - DPI/INPE**, 2007. Disponível em: <<http://www.dpi.inpe.br/geodma>>. Acesso em: 10 set. 2010.

KORTING, T. S.; FONSECA, L. M. G.; ESCADA, M. I. S.; SILVA, F. C.; SILVA, M. P. S. GeoDMA - A Novel System for Spatial Data Mining. **Proceedings of the 2008 IEEE International Conference on Data Mining Workshops (ICDM'08)**, Pisa, Italy, p. 975-978, dec. 2008.

KURAMOCHI, M.; KARYPIS, G. Frequent Subgraph Discovery. **Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM '01)**, Los Alamitos, CA, USA, v. 0, p. 313-320, 2001.

KURAMOCHI, M.; KARYPIS, G. An Efficient Algorithm for Discovering Frequent Subgraphs. **Journal IEEE Transactions on Knowledge and Data Engineering**, Piscataway, NJ, USA, v. 16, p. 1038 - 1051, September 2004.

LAM, W. W. M.; CHAN, K. Analyzing web layout structures using graph mining. **IEEE International Conference on Granular Computing (GrC '08)**, Hangzhou, p. 361 - 366, 2008a.

LAM, W. W. M.; CHAN, K. C. C. A Graph Mining Algorithm for Classifying Chemical Compounds. **IEEE International Conference on Bioinformatics and Biomedicine (BIBM '08)**, Philadelphia, PA, p. 321-324, nov. 2008b.

LAMBIN, E. F.; GEIST, H. J.; LEPERS, E. Dynamics of land-use and land-cover change in Tropical Regions. **Annual Review of Environment and Resources**, v. 28, n. 1, p. 205-241, 2003.

LAST, M.; KANDEL, A.; BUNKE, H. **Data Mining in Time Series Databases**. Singapore: World Scientific Printers, v. 57, 2004.

LOPES, A. V. **Estrutura de Dados para a construção de software**. Canoas: ULBRA, v. 1, 1999. 440 p.

LORENA, R. B. **Linking spatial patterns of land-use to agents of deforestation in the Brazilian Amazon**. Tese (Doutorado em Ciências) - Université catholique de Louvain. Louvain-La-Neuve, Belgique. 2008.

LORENA, R. B.; LAMBIN, E. F. Linking spatial patterns of deforestation to land use using satellite and field data. **IEEE International Geoscience and Remote Sensing Symposium (IGARSS '07)**, Barcelona, p. 3357-3361, July 2007.

MANOCH, P.; TONGSIRI, N.; CHEEVADHANARAK, S.; CHAIJARUWANICH, J. Improvement of Short Read Genome Assembly with Graph Mining Technique. **International Conference on Future Information Technology**, Singapore, v. 13, 2011.

MAPSERVER. **MapServer - open source web mapping**, 2010. Disponível em: <<http://mapserver.org>>. Acesso em: 21 set. 2010.

MARTINS, C. B. **Apostila da disciplina análise de algoritmos**. Faculdade de Filosofia, Ciências e Letras de Mandaguari. Mandaguari. 2004.

- MATOS, J. M. S.; ROCHA, V.M. S.; CARVALHO, A. P. F.; CARNEIRO, G. M. G.; ROSA, M.; RIBEIRO, R. M.; COSTA, R. C.; SCHWARZ, C. O. Avaliação da evolução do desmatamento em assentamentos do Incra a partir dos dados do Prodes e Deter para os anos 1997-2010. **XV Brazilian Remote Sensing Symposium (SBSR)**, Curitiba, PR, Brasil, p. 5833-5840, 2011.
- MAXWELL, E. K.; BACK, G.; RAMAKRISHNAN, N. Diagnosing memory leaks using graph mining on heap dumps. **Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '10)**, Washington, DC, USA, p. 115--124, 2010.
- MCGARIGAL, K.; MARKS, B. J. FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure. **Pacific Northwest Research Station General Technical**, Portland, OR: U.S., p. 132, 1995.
- MEDEIROS, C. B. Grand Research Challenges in Computer Science in Brazil. **Computer**, Los Alamitos, CA, USA, v. 41, p. 59-65, 2008.
- MENDES, P. Desmatamento na Amazônia Legal cai 11%, diz Inpe. **G1 Natureza**, 05 dez. 2011. Disponível em: <<http://g1.globo.com/natureza/noticia/2011/12/desmatamento-na-amazonia-legal-cai-11-diz-inpe.html>>. Acesso em: 05 dez. 2011.
- MICHALAK, K.; KORCZAK, J. Graph Mining Approach to Suspicious Transaction Detection. **6th International Symposium Advances in Artificial Intelligence and Applications**, Szczecin, Poland, p. 69–75, 2011.
- MICHEL, J.; INGLADA, J. Multi-Scale Segmentation and Optimized Computation of Spatial Reasoning Graphs for Object Detection in Remote Sensing Images. **IEEE International Geoscience and Remote Sensing Symposium (IGARSS '08)**, Boston, MA , p. 431 - 434, 2008.
- MILLER, H. J.; HAN, J. **Geographic Data Mining and Knowledge Discovery**. 2. ed. [S.l.]: CRC Press, 2009. 486 p.
- MORAES, E. C. **Fundamentos de Sensoriamento Remoto**. São José dos Campos: Instituto Nacional de Pesquisas Espaciais - INPE, 2002. 281 p.
- MOREIRA, M. A. **Fundamentos do Sensoriamento Remoto e Metodologias de Aplicação**. 3. ed. São José dos Campos: UFV - Universidade Federal de Viçosa, 2007. 315 p.
- MOTA, J. S.; CÂMARA, G.; ESCADA, M. I. S.; BITTENCOURT, O.; FONSECA, L. M. G.; VINAS, L. **Case-based reasoning for eliciting the evolution of geospatial objects**. Berlin, Heidelberg: Springer-Verlag. 2009. p. 405-420.
- NETTO, P. O. B. **Grafos: teoria, modelos e algoritmos**. Norwell, Massachusetts: Edgard Blücher, 2003. 314 p.

NIJSSSEN, S.; KOK, J. N. A Quickstart in Frequent Structure Mining Can Make a Difference. **Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '04)**, New York, NY, USA, p. 647-652, 2004.

PARSAYE, K. **Intelligent Databases: object-oriented, deductive hypermedia technologies**. New York: John Wiley, 1989.

PEREIRA, L. M.; ESCADA, M. I. S.; RENNÓ, C. D. Análise da evolução do desmatamento em áreas de pequenas, médias e grandes propriedades na região centro-norte de Rondônia, entre 1985 e 2000. **XIII Brazilian Remote Sensing Symposium (SBSR)**, Florianópolis, SC, Brazil, p. 6905-6912, 2007. Disponível em: <<http://urlib.net/rep/dpi.inpe.br/sbsr@80/2006/11.10.12.31>>.

POSTGRESQL. PostgreSQL. **PostgreSQL - The world's most advanced open source database**, 2011. Disponível em: <<http://www.postgresql.org>>. Acesso em: 17 jul. 2011.

PRADHAN, S. K. **A relational database approach for frequente subgraph mining**. Dissertação (Mestrado em Ciência da Computação e Engenharia) - The University of Texas at Arlington. Arlington. 2006.

QUEIROZ, G. R.; CÂMARA, G. Arquitetura de Sistemas de Informação Geográfica. **Introdução à Ciência da Geoinformação**, 2001. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/introd/cap3-arquitetura.pdf>>. cap. 3.

SAITO, É. A. **Caracterização de trajetórias de padrões de ocupação humana na Amazônia Legal por meio de mineração de dados**. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais - INPE. São José dos Campos. 2011.

SAITO, É. A.; KORTING, T. S.; FONSECA, L. M. G.; ESCADA, M. I. S. Mineração de dados espaciais de desmatamento do PRODES utilizando métricas da paisagem caso de estudo município de Novo Progresso - PA. **III Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação**, Recife - PE, p. 001-009, jul. 2010.

SAITO, É. A.; ESCADA, M. I. S.; FONSECA, L. M. G.; KORTING, T. S. Análise de padrões de desmatamento e trajetória de padrões de ocupação humana na Amazônia usando técnicas de mineração de dados. **XV Brazilian Remote Sensing Symposium (SBSR)**, Curitiba, PR, Brasil, p. 2833-2840, 2011.

SAUSEN, T. Sensoriamento Remoto - Tópicos em meio ambiente e ciências atmosféricas. **Instituto Nacional de Pesquisas Espaciais - INPE**, São José dos Campos, 2005. Disponível em: <<http://mtc-m15.sid.inpe.br/sid.inpe.br/iris@1915/2005/11.08.13.16>>.

SCHOWENGERDT, R. A. **Remote Sensing: Models and Methods for Image Processing**. 3. ed. [S.l.]: Elsevier, 2007. 560 p.

SHRIVASTAVA, S.; KULSHRESTHA, K.; SINGH, P.; PAL, S. N. Pruthak: mining and analyzing graph substructures. **Proceedings of the Eighth Workshop on Mining and Learning with Graphs (MLG '10)**, Washington, D.C., p. 110-118, 2010.

SILBERCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. **Sistema de Banco de Dados**. 5. ed. Rio de Janeiro: Elsevier, 2006. 781 p.

SILVA, A.; JR., W. M.; ZAKI, M. J. Structural correlation pattern mining for large graphs. **Proceedings of the Eighth Workshop on Mining and Learning with Graphs (MLG '10)**, Washington, D.C., p. 119-126, 2010.

SILVA, M. P. S. Mineração de Dados: Conceitos, Aplicações e Experimentos com Weka. **IV Escola Regional de Informática RJ/ES (IV ERI RJ/ES)**, Nov. 2004a.

SILVA, M. P. S. **Mineração de Imagens usando Ontologias**. Monografia da proposta de Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais. São José dos Campos. 2004b.

SILVA, M. P. S. **Mineração de Padrões de Mudança em Imagens de Sensoriamento Remoto**. Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais - INPE. São José dos Campos. 2006.

SILVA, M. P. S.; CÂMARA, G.; SOUZA, R. C. M.; VALERIANO, D. M.; ESCADA, M. I. S.. Mining patterns of change in remote sensing image databases. **Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM '05)**, Houston, Texas, USA, v. 1, p. 362-369, Nov. 2005.

SILVA, M. P. S.; CÂMARA, G.; ESCADA, M. I. S.; MODESTO, R. C. Remote-sensing image mining: detecting agents of land-use change in tropical forest areas. **International Journal of Remote Sensing**, Bristol, PA, USA, v. 29, p. 4803-4822, January 2008.

SPRING. Sistema de Processamento de Informações Georeferenciadas. **SPRING**, 2010. Disponível em: <<http://www.dpi.inpe.br/spring>>. Acesso em: 21 set. 2010.

SUDAM. Superintendência de Desenvolvimento da Amazônia. **SUDAM**, 2011. Disponível em: <<http://www.sudam.gov.br>>. Acesso em: 20 dez. 2011.

TERRAVIEW. TerraView 4.1.0. **Instituto Nacional de Pesquisas Espaciais -INPE**, São José dos Campos, SP, 2011. Disponível em: <<http://www.dpi.inpe.br/terraview/index.php>>. Acesso em: 10 fev. 2011.

TILTON, J. C.; COOK, D. J.; KETKAR, N. S. The Integraton of Graph-Based Knowledge Discovery with Image Segmentation Hierarchies for Data Analysis, Data Mining and Knowledge Discovery. **IEEE International Geoscience & Remote Sensing Symposium (IGARSS '08)**, p. 491-494, 2008.

TURNER, M. G.; GARDNER, R. H.; O'NEILL, R. V. **Landscape ecology: in the theory and practice, pattern and process**. New York: Springer-Verlag, 2001. 401 p.

VASCONCELOS, C. H.; NOVO, E. Influence of precipitation, deforestation and Tucuruí reservoir operation on malaria incidence rates in southeast Para, Brazil. **IEEE International Geoscience and Remote Sensing Symposium (IGARSS '03)**, Toulouse, France, v. 7, p. 4567 - 4569, 2003.

VILLALOBOS, A. R. Grafos. **Grafos - software para la construcción, edición y análisis de grafos**, 2010. Disponível em: <<http://arodrigu.webs.upv.es/grafos/doku.php?id=inicio>>. Acesso em: 5 out. 2011.

WASHIO, T.; MOTODA, H. State of the art of graph-based data mining. **ACM SIGKDD Explorations Newsletter**, New York, NY, USA, v. 5, n. 1, p. 59-68, 2003.

WITTEN, I. H.; FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques**. 2. ed. San Francisco, CA: Morgan Kaufmann, 2005.

YAN, X.; HAN, J. gSpan: Graph-Based Substructure Pattern Mining. **Proceeding of the 2002 International Conference on Data Mining (ICDM '02)**, Washington, DC, USA, p. 721-724, 2002.

YEH, J.-B.; WU, C.-H. Extraction of robust visual phrases using graph mining for image retrieval. **Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS '10)**, Paris, p. 3681 - 3684 , 2010.

YOU, C. H.; HOLDER, L. B.; COOK, D. J. Graph-Based Data Mining in Dynamic Networks: Empirical Comparison of Compression-Based and Frequency-Based Subgraph Mining. **IEEE International Conference on Data Mining Workshops (ICDMW '08)** , Pisa, p. 929 - 938, 2008.

ZAMALIEVA, D.; AKSOY, S.; TILTON, J. C. Finding Compound Structures in Images using Image Segmentation and Graph-based Knowledge Discovery. **IEEE International Geoscience & Remote Sensing Symposium (IGARSS '09)**, p. 252-255, 2009.

ZOU, R.; HOLDER, L. B. Frequent subgraph mining on a single large graph using sampling techniques. **Proceedings of the Eighth Workshop on Mining and Learning with Graphs (MLG '10)**, Washington, D.C., p. 171-178, 2010.