



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

INPE-12830-PRE/8120

## CAPÍTULO 1

# REPRESENTAÇÃO COMPUTACIONAL DE DADOS GEOGRÁFICOS

Gilberto Câmara

*Bancos de dados geográficos*

INPE  
São José dos Campos  
2005

# 1 Representação computacional de dados geográficos

*Gilberto Câmara*

## 1.1 Introdução

Este capítulo examina os problemas básicos de representação computacional de dados geográficos, e esclarece questões da seguinte natureza: *Como representar os dados geográficos no computador? Como as estruturas de dados geométricas e alfanuméricas se relacionam com os dados do mundo real? Que alternativas de representação computacional existem para dados geográficos?*

Em seu livro “Olhos de Madeira”, Carlo Ginzburg nos traz um fascinante ensaio sobre a origem da palavra ‘representação’. A origem do termo remonta ao século XIII, chamando-se *représentation* aos manequins de cera exibidos junto ao cadáver dos reis franceses e ingleses durante as cerimônias funerárias (Ginzburg, 2001). Enquanto o soberano era velado, a presença do manequim era um testemunho à transcendência do rei e a sua presença futura no mundo dos mortos. O manequim tinha a função de lembrar aos presentes que o rei havia assumido uma outra forma e que uma nova vida se iniciava para o morto. Nesta nova forma, apesar de morto o rei continuaria presente para seus súditos (“*re + présentation*”).

Assim, desde a sua origem a palavra ‘representação’ está associada a uma forma abstrata de descrição do mundo. O uso do manequim como representação do soberano morto é apenas um exemplo do problema mais geral da construção de abstrações que descrevem o mundo. Para explicar como funcionam os bancos de dados geográficos, este capítulo descreve o processo de transformar aos conceitos abstratos de *espaço*

*geográfico* no referindo ao *espaço computacionalmente representado*. Para exemplificar, consideremos alguns problemas:

- Uma cientista social deseja entender e quantificar o fenômeno da *exclusão social* numa grande cidade brasileira, através de mapas de exclusão/inclusão social, gerados a partir de dados censitários (Sposati, 1996).
- Uma ecóloga pretende estudar os remanescentes florestais da Mata Atlântica, através de estudos de *fragmentação* obtidos a partir de interpretação de imagens de satélite (Pardini et al., 2005).
- Uma pedóloga pretende determinar a *distribuição* de propriedades do solo numa área de estudo, a partir de um conjunto de amostras de campo (Bönisch et al., 2004).

O que há de comum nesses casos? A especialista lida com conceitos de sua disciplina (*exclusão social, fragmentos, distribuição de propriedades do solo*) e precisa de representações que traduzam estes conceitos para o computador. Após esta tradução, ela poderá compartilhar os dados de seu estudo, inclusive com pesquisadores de outras disciplinas.

## 1.2 Descrição geral de sistemas de informação geográfica

O termo *sistemas de informação geográfica* (SIG) é aplicado para sistemas que realizam o tratamento computacional de dados geográficos. A principal diferença de um SIG para um sistema de informação convencional é sua capacidade de armazenar tanto os atributos descritivos como as geometrias dos diferentes tipos de dados geográficos. Assim, para cada lote num cadastro urbano, um SIG guarda, além de informação descritiva como proprietário e valor do IPTU, a informação geométrica com as coordenadas dos limites do lote. A partir destes conceitos, é possível indicar as principais características de SIGs:

- Inserir e integrar, numa única base de dados, informações espaciais provenientes de meio físico-biótico, de dados censitários, de cadastros urbano e rural, e outras fontes de dados como imagens de satélite, e GPS.

- Oferecer mecanismos para combinar as várias informações, através de algoritmos de manipulação e análise, bem como para consultar, recuperar e visualizar o conteúdo da base de dados geográficos.

Os componentes de um SIG estão mostrados na Figura 1.1. No nível mais próximo ao usuário, a *interface homem-máquina* define como o sistema é operado e controlado. Esta interface pode ser tanto baseada na metáfora da “mesa de trabalho” (Kuhn e Frank, 1991) (Richards e Egenhofer, 1995) (Câmara, 1999), como adaptada ao ambiente de navegação da Internet (Kraak e Brown, 2001), quanto baseada em linguagens de comando como Spatial SQL (Egenhofer, 1994) e LEGAL (Câmara, 1995). No nível intermediário, um SIG deve ter mecanismos de processamento de dados espaciais. A *entrada* de dados inclui os mecanismos de conversão de dados (Hohl, 1998). Os algoritmos de *consulta e análise espacial* incluem as operações topológicas (Egenhofer e Franzosa, 1991), álgebra de mapas (Tomlin, 1990), estatística espacial (Druck et al., 2004), modelagem numérica de terreno (Li et al., 2004) e processamento de imagens (Mather, 2004). Os mecanismos de visualização e plotagem devem oferecer suporte adequado para a apreensão cognitiva dos aspectos relevantes dos dados pesquisado (MacEachren, 2004) (Tufte, 1983) (Monmonier, 1993). No nível mais interno do sistema, um *sistema de gerência de bancos de dados geográficos* oferece armazenamento e recuperação dos dados espaciais e seus atributos. Cada sistema, em função de seus objetivos e necessidades, implementa estes componentes de forma distinta, mas todos os subsistemas citados devem estar presentes num SIG.

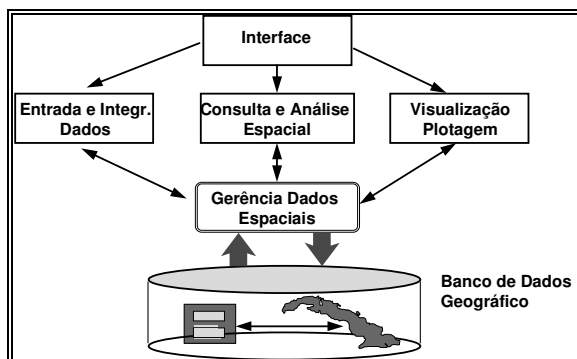


Figura 1.1 – Arquitetura de sistemas de informação geográfica.

Do ponto de vista da aplicação, o uso de sistemas de informação geográfica (SIG) implica em escolher as representações computacionais mais adequadas para capturar a semântica de seu domínio de aplicação. Do ponto de vista da tecnologia, desenvolver um SIG significa oferecer o conjunto mais amplo possível de estruturas de dados e algoritmos capazes de representar a grande diversidade de concepções do espaço. Como o presente livro está focado nos diferentes aspectos relacionados com a tecnologia de bancos de dados geográficos, discutiremos em maior detalhe a questão de gerência de dados espaciais. Leitores interessados nos demais aspectos de um SIG poderão consultar as referências acima.

### 1.3 Traduzindo a informação geográfica para o computador

Para abordar o problema fundamental da Geoinformação, que é a *produção de representações computacionais do espaço geográfico*, usamos o *paradigma dos quatro universos*, proposto inicialmente por Gomes e Velho (1995) e adaptado para a geoinformação por Câmara (1995). Este paradigma distingue quatro passos entre o mundo real e sua realização computacional (ver Figura 1.2).

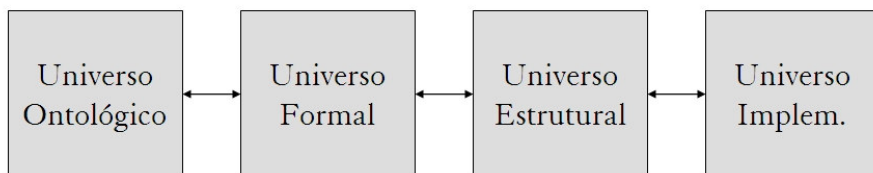


Figura 1.2 – Paradigma dos quatro universos.

No primeiro passo, nossas percepções do mundo real são materializadas em conceitos que descrevem a realidade e respondem a questões como: *Que classes de entidades são necessárias para descrever o problema que estamos estudando?* (Smith, 2003). Criamos assim o *universo ontológico*, onde incluímos os conceitos da realidade a serem representados no computador, como os tipos de solo, elementos de cadastro urbano, e caracterização das formas do terreno.

O segundo universo (*o universo formal*) inclui modelos lógicos ou construções matemáticas que generalizam os conceitos do universo ontológico e dão resposta à pergunta: *Quais são as abstrações formais necessárias para representar os conceitos de nosso universo ontológico?* Estas abstrações incluem modelos de dados e álgebras computacionais. Exemplos: o modelo entidade-relacionamento (Chen, 1976) e o modelo OMT (Rumbaugh et al., 1991). O caso específico de modelos dados geográficos é tratado em maior detalhe nos capítulos 3 e 4 deste livro e inclui o modelo OMT-G (Davis et al., 2002). A questão de linguagens é tratada no Capítulo 5.

O terceiro universo é o *universo estrutural*, onde as diversas entidades dos modelos formais são mapeadas para estruturas de dados geométricas e alfanuméricas, e algoritmos que realizam operações. Neste universo, respondemos a questões como: *Quais são os tipos de dados e algoritmos necessários para representar os modelos e as álgebras do universo formal?* As estruturas de dados são os elementos básicos de construção dos sistemas computacionais, e serão discutidas em maior detalhe neste capítulo. Aspectos do universo estrutural descritos no livro incluem arquiteturas de SGBD (Capítulo 8), conversão de dados (Capítulo 9), interoperabilidade (Capítulo 10) e disseminação de dados na Internet (Capítulo 11).

O universo de *implementação* completa o processo de representação computacional. Neste universo, realizamos a implementação dos sistemas, fazendo escolhas como arquiteturas, linguagens e paradigmas de programação. Neste livro, as questões de implementação discutidas incluem geometria computacional (Capítulo 2), métodos de acesso (Capítulo 6), processamento de consultas (Capítulo 7), além da descrição detalhada da biblioteca TerraLib (Capítulos 12 a 14).

O paradigma dos quatro universos é uma forma de compreendermos que a transposição da realidade para o computador requer uma série complexa de mediações. Primeiro, precisamos dar nomes às entidades da realidade. Depois, geramos modelos formais que as descrevem de forma precisa. A seguir, escolhemos as estruturas de dados e algoritmos que melhor se adaptam a estes modelos formais. Finalmente, fazemos a implementação num suporte computacional apropriado. Nas próximas seções, examinaremos em detalhe cada um destes universos.

#### 1.4 O universo ontológico

Ontologia é o campo da filosofia cujo objetivo é descrever os tipos e estruturas de entidades, eventos, processos e relações que existem no mundo real (Smith, 2003). Sua gênese remonta a Aristóteles, mas o interesse recente por ontologias em sistemas de informação decorre principalmente da necessidade de compartilhar informação de forma eficiente para um público cada vez mais interdisciplinar.

Um sistema de informação pode ser concebido como um mecanismo de comunicação entre duas partes: o produtor e o usuário. Para que funcione, é necessário que haja uma concordância entre os conceitos das partes. Numa perspectiva mais geral, seu sucesso depende da existência de uma comunidade que compartilhe as definições utilizadas para construí-lo. Por exemplo, considere o caso de um estudo sobre *segregação* em áreas urbanas. Existem diferentes conceitos de segregação na literatura sociológica (Caldeira, 2000) (Massey e Denton, 1993) (Torres, 2004) (White, 1983). Para construir um sistema de informação que permita o estudo da segregação urbana, é preciso que o produtor de informação defina qual dos diferentes conceitos estará sendo

representado, como esta representação será construída, e como o usuário pode compreender as características e limitações desta representação.

Deste modo, o problema fundamental de um sistema de informação é definir o conjunto de conceitos a ser representado. Se quisermos que estes conceitos sejam compartilhados por uma comunidade interdisciplinar, é fundamental que os conceitos utilizados sejam devidamente explicitados. Assim, surge a pergunta: “*Qual o papel dos conceitos na representação do mundo?*” A melhor forma de responder é baseando-se na perspectiva realista (Searle, 1998):

1. A realidade existe independentemente das representações humanas.
2. Nós temos acesso ao mundo através de nossos sentidos e de nossos instrumentos de medida.
3. As palavras em nossa linguagem podem ser usadas para referir-se a objetos do mundo real.
4. Nossas afirmações são verdadeiras ou falsas dependendo de sua correspondência aos fatos do mundo.
5. Algumas afirmações em nossa linguagem dizem respeito a uma realidade externa e independente (“*há neve no topo do Monte Everest*”). Outras afirmações dizem respeito a convenções socialmente construídas (“*este papel é uma certidão de nascimento*”).

Como nos ensina Searle (1993), esta perspectiva tem conseqüências importantes sobre nossa concepção do mundo:

*“Apesar de termos representações mentais e lingüísticas do mundo sob a forma de crenças, experiências, afirmações, teorias, etc., há um mundo, ‘lá fora’, totalmente independente destas representações. A órbita elíptica dos planetas relativamente ao Sol e a estrutura do átomo de hidrogênio são inteiramente independentes das representações que os seres humanos têm de tais fenômenos. Já coisas como o dinheiro, a propriedade, o casamento e os governos são criados e sustentados pelo comportamento cooperativo humano.”*

*“Na sua maior parte, o mundo existe independentemente da linguagem (princípio 1) e uma das funções da linguagem é representar como são as coisas no mundo (princípio 3). Um aspecto crucial no qual a realidade e a linguagem entram em contato é marcado pela noção de verdade. Em*



geral, as afirmações são verdadeiras na medida em que representam com precisão uma característica da realidade que existe independentemente da afirmação (princípio 4).”.

O projeto de um sistema de informação requer, como passo inicial, a escolha das entidades a ser representados e, se possível, a descrição organizada destas entidades por meio de conceitos. Esta descrição forma uma *ontologia de aplicação*, definida como “um conjunto de conceitos compartilhados por uma comunidade” (Gruber, 1995). Para os dados geográficos, uma geo-ontologia tem dois tipos básicos de conceitos: (a) conceitos que correspondem a fenômenos físicos do mundo real; (b) conceitos que criamos para representar entidades sociais e institucionais (Smith e Mark, 1998) (Fonseca et al., 2003). Chamamos o primeiro tipo de *conceitos físicos* e o segundo de *conceitos sociais* (Tabela 1.1). Embora todos os conceitos resultem do uso compartilhado da linguagem, há uma diferença entre conceitos que se referem ao mundo físico (“*A Amazônia possui uma floresta tropical*”) e aqueles que resultam de convenções humanas (“*Esta é uma reserva indígena*”).

Nossa geo-ontologia diferencia entre conceitos associados a entidades que pode ser individualizadas e identificadas nominalmente (caso de *lagos e lotes*) e aquelas que variam de forma contínua no espaço (caso de *poluição*).

Tabela 1.1 – Tipos de conceitos associados a entidades geográficas

	<i>Realidade física</i>	<i>Realidade social</i>
<i>Entidades individualizáveis</i>	<i>indivíduos bona fide</i> (e.g., montanha)	<i>indivíduos fiat</i> (e.g., lote)
<i>Entidades com variação contínua</i>	<i>topografias físicas</i> (e.g., poluição)	<i>topografias sociais</i> (e.g., segregação)

Os conceitos físicos podem ser subdivididos em:

- Conceitos associados a entidades individualizáveis, que possuem uma fronteira bem definida a partir de diferenciações qualitativas ou descontinuidades na natureza. Designados como *indivíduos bona fide*

(do latim “boa fé”), sua existência decorre de nossa necessidade de dar nomes aos elementos do mundo natural. Por exemplo, embora a superfície da Terra apresente uma variação contínua no espaço, nossa percepção do espaço depende da associação de nomes especiais a variações bem definidas no terreno. Daí nascem conceitos como *montanha, vale e desfiladeiro*.

- Conceitos associados a entidades que tem variação contínua no espaço, associadas aos fenômenos do mundo natural, não estando a princípio limitadas por fronteiras. Chamamos estes conceitos de topografias físicas, onde o termo “topografia” está associado a qualquer grandeza que varia continuamente. Exemplos incluem *temperatura, altimetria, declividade e poluição*.
- Os conceitos sociais podem ser subdivididos em:
- Conceitos que descrevem entidades individuais criadas por leis e por ações humanas. Estas entidades possuem uma fronteira que as distingue do seu entorno e tem uma identidade única. Sua existência depende usualmente de um registro legal. Designadas como *indivíduos fiat* (do latim “fazer”), incluem conceitos como *lotes, municípios e países*.
- Conceitos descrevendo entidades que têm variação contínua no espaço, associadas a convenções sociais. Tome-se o caso de *pobreza*, conceito socialmente definido que ocorre no espaço de forma ininterrupta (“*em cada lugar há algum tipo diferente de pobreza*”). Chamamos estes conceitos de topografias sociais. Exemplos incluem: *exclusão social, segregação urbana, desenvolvimento humano*.

Uma geo-ontologia é um conjunto de conceitos e um conjunto de relações semânticas e espaciais entre estes termos. Cada conceito tem um nome, uma definição e um conjunto de atributos. O conjunto das relações semânticas inclui as relações de sinonímia, similaridade, e hiponímia (também dito especialização: “*hospital é um tipo de prédio*”). Por exemplo:

- **rio**: “Curso de água natural, de extensão mais ou menos considerável, que se desloca de um nível mais elevado para outro mais baixo,

aumentando progressivamente seu volume até desaguar no mar, num lago, ou noutro rio”.

- **riacho:** “rio pequeno, mais volumoso que o regato e menos que a ribeira”
- relação semântica: um riacho *é um* rio. (hiponímia).

O conjunto de relações espaciais inclui as relações topológicas como pertinência e adjacência, relações direcionais como “ao norte de”, e relações informais como “no coração de” ou “perto de”. Por exemplo:

- **afluente:** “curso de água que deságua em outro curso de água, considerado principal”.
- relação espacial: um afluente *está conectado a* um rio.

Na maior parte dos sistemas de informação atuais, as ontologias de aplicação não estão explicitadas, o que reduz o potencial de compartilhamento da informação. Com o advento da Internet, que permite a disseminação de dados forma ampla e para um público heterogêneo, a necessidade de explicitar as ontologias utilizadas tornou-se ainda mais premente. A explicitação das ontologias de aplicação está na base das propostas recentes da “Web Semântica” (Berners-Lee et al., 2001) e de propostas de padrões como OWL. Como resultado de pesquisas recentes, já temos vários sistemas disponíveis na Internet para criação e gestão de ontologias, como o Protegé (Noy et al., 2001). Para dados geográficos, o consórcio OGC (“Open Geospatial Consortium”) propôs o formato GML como mecanismo de descrição de ontologias geográficas. Fazemos uma descrição mais detalhada do tema nos Capítulos 9 e 10 deste livro.

## 1.5 O universo formal

O universo formal representa um componente intermediário entre os conceitos do universo ontológico e as estruturas de dados e algoritmos computacionais. Como os computadores trabalham com estruturas matemáticas, a passagem direta de conceitos informais da ontologia de aplicação para estruturas de dados poderia gerar decisões inconsistentes. No universo formal, buscamos estabelecer um conjunto de entidades lógicas que agrupem os diferentes conceitos da ontologia de aplicação da

forma mais abrangente possível. Adicionalmente, neste universo definimos ainda como serão associados valores aos diferentes conceitos; ou seja, como podemos medir o mundo real. Deste modo, o universo formal tem duas partes: (a) como medir o mundo real (teoria da medida); (b) como generalizar os conceitos da ontologia em entidades formais abrangentes. Estas duas partes serão discutidas a seguir.

### 1.5.1 Atributos de dados geográficos: teoria da medida

Para representar dados geográficos no computador, temos de descrever sua variação no espaço e no tempo. Em outras palavras, precisamos poder a perguntas como: “qual é o valor deste dado aqui e agora?”. Isto requer uma compreensão dos processos de mensuração da realidade, de forma consistente com os dois primeiros princípios de Searle (1998): “a realidade existe independentemente das representações humanas” e “nós temos acesso ao mundo através de nossos sentidos e de nossos instrumentos de medida”. O processo de medida consiste em associar números ou símbolos a diferentes ocorrências de um mesmo atributo, para que a relação dos números ou símbolos reflita as relações entre as ocorrências mensuradas. Por exemplo, podemos medir a poluição numa cidade através de sensores localizados em diferentes locais. Cada um destes sensores nos dará uma medida diferente. Esta atribuição é denominada *escala de medida*. A referência geral mais importante sobre escalas de medidas é o trabalho de Stevens (1946), que propõe quatro escalas de mensuração: *nominal*, *ordinal*, *intervalo* e *razão*.

Os níveis nominal e ordinal são *temáticos*, pois a cada medida é atribuído um número ou nome associando a observação a um tema ou classe. A *escala nominal* classifica objetos em classes distintas sem ordem inerente, como rótulos que podem ser quaisquer símbolos. As possíveis relações entre os valores são identidade ( $a = b$ ) e dessemelhança ( $a \neq b$ ). Um exemplo é a cobertura do solo, com rótulos como “*floresta*”, “*área urbana*” e “*área agrícola*”.

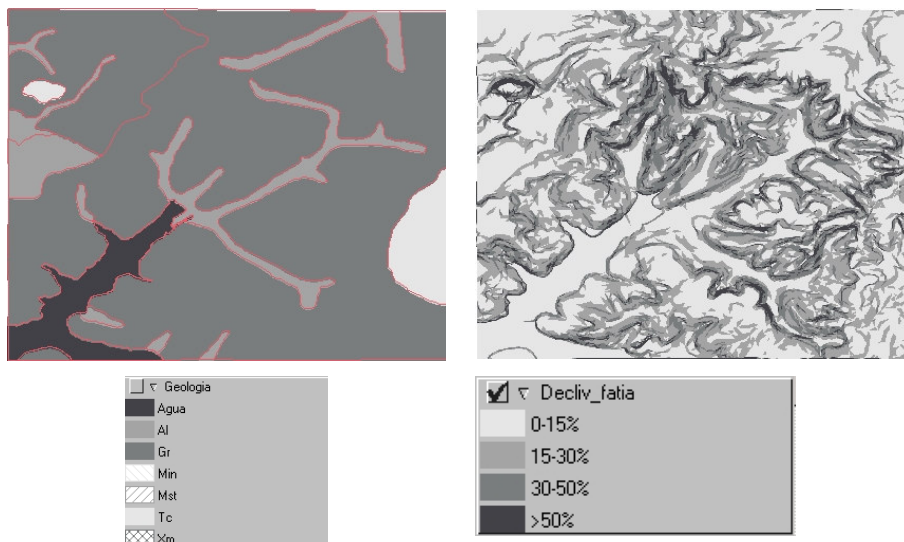


Figura 1.3 – Exemplos de medida nominal (mapa geológico) e medida ordinal (mapa de classes de declividade) .

A *escala ordinal* introduz a idéia de ordenação, caracterizando os objetos em classes distintas que possuem uma ordem natural (por exemplo 1 – ruim, 2 – bom, 3 – ótimo ou “0-10%”, “11-20%”, “mais que 20%”). A distância definida entre os elementos não é significativa. Nesta escala são evidenciadas as relações “<” ou “>”, isto implica que para todo  $a$  e  $b$ , as relações  $a < b$ ,  $a > b$  ou  $a = b$  são possíveis. Um exemplo é a aptidão agrícola de solos, com rótulos como “muito apto”, “apto”, “pouco apto”, e “inapto” (ver Figura 1.3).

As medidas temáticas não estão associadas à magnitude do fenômeno. Quando o estudo necessita de uma descrição mais detalhada, que permita comparar intervalo e ordem de grandeza entre eventos, recorre-se aos níveis de medidas denominados de *numéricos*, onde as regras de atribuição de valores baseiam-se em uma escala de números reais.

Existem dois níveis de medidas baseados em escalas de números reais: *escala por intervalo* e o *escala por razão*. A *escala por intervalo* possui um ponto zero arbitrário, uma distância proporcional entre os intervalos e uma faixa de medidas entre  $[-\infty, \infty]$ . A temperatura em graus Celsius é exemplo de medida por intervalo, onde o ponto zero corresponde a uma

convenção (a fusão do gelo em água). Por ter uma referência zero arbitrária, valores medidos no nível por intervalo não podem ser usados para estimar proporções. Operações aritméticas elementares (adição e subtração) são válidas, porém multiplicação e divisão não são apropriadas. Por exemplo, dados  $a$  e  $b$ , pode-se ter  $a = b + c$ , onde  $c$  é a diferença entre  $a$  e  $b$  em alguma unidade padrão. Assim, a temperatura em São Paulo pode ser  $c$  graus mais baixa do que a temperatura em Campos de Jordão.

A *escala de razão* permite um tratamento mais analítico da informação, pois nela o ponto de referência zero não é arbitrário, mas determinado por alguma condição natural. Sua faixa de valores é limitada entre  $[0, \infty]$ . Nesta escala existe um ponto zero absoluto que não pode ser alterado e um intervalo arbitrário com distâncias proporcionais entre os intervalos. Números negativos não são permitidos, pois o número zero representa ausência total daquilo que está sendo medido. Por exemplo, na descrição de atributos como peso e volume de objetos não há valores negativos. No caso de temperatura em graus Kelvin, a condição natural é o ponto de repouso dos átomos da matéria, a partir do qual não se consegue temperaturas menores. Este ponto é o zero absoluto para temperatura, zero graus Kelvin. O fato de ponto de referência zero ser absoluto permite afirmações tais como  $a$  é duas vezes mais pesado que  $b$ . Desta forma, dado  $a$  e  $b$  pode-se ter  $a = c \times b$ , onde  $c$  indica o número de vezes que  $b$  vai até  $a$ , a relação de  $a$  para  $b$ . Operações matemáticas de adição, subtração, multiplicação e divisão são suportadas nesta escala.

A Tabela 1.2 apresenta um resumo das escalas de medidas, destaca a característica principal, apresenta algumas operações admitidas e exemplos para cada uma delas.

Tabela 1.2 – Tipos de medidas de dados geográficos

<i>Escala</i>	<i>Características</i>	<i>Exemplos</i>	<i>Operações possíveis</i>
<i>Nominal</i>	Descrição	Tipo de solo, vegetação, uso do solo	Seleção, Comparação

---

<i>Ordinal</i>	Ordem	Classes de declividade, aptidão de uso	Mediana, Máximo, Mínimo
<i>Intervalo</i>	Distância	Altimetria	Diferença, Soma
<i>Razão</i>	Valores absolutos	Renda, população, taxa de natalidade	Operações aritméticas

---

### 1.5.2 Espaço absoluto e espaço relativo

Antes de considerar os diferentes modelos formais para dados geográficos, é necessário analisarmos brevemente os conceitos de *espaço absoluto* e *espaço relativo*. Esta distinção decorre da possibilidade de representarmos no computador a localização dos objetos no espaço ou apenas o posicionamento relativo entre eles, como ilustrado na Figura 1.4. Nesta figura, mostramos à esquerda os distritos da cidade de São Paulo, identificados por suas fronteiras. Neste caso, trata-se de uma representação no espaço absoluto, na qual as coordenadas das fronteiras devem corresponder às estabelecidas na legislação. Do lado direito, mostramos um grafo com as conexões dos distritos, que formam uma rede (repetimos a imagem dos distritos por razões de melhor legibilidade da figura). No modelo de redes, a localização exata de cada distrito não é armazenada, pois a rede só captura as relações de adjacência. Dizemos então que a rede de conexões dos distritos é um modelo de *espaço relativo*.

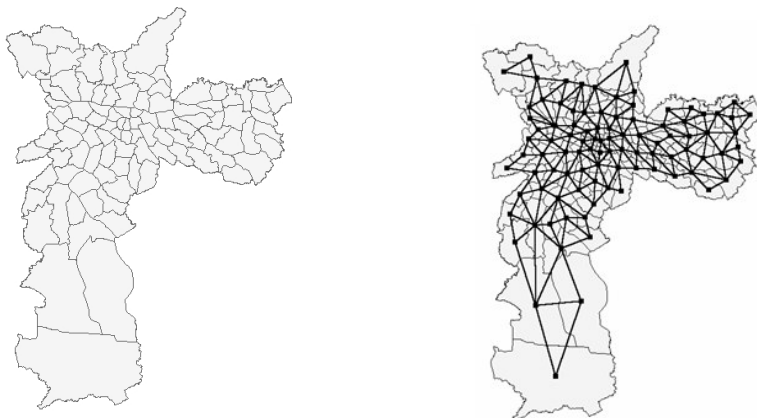


Figura 1.4 – Dualidade entre espaço absoluto e espaço relativo. À esquerda, distritos de São Paulo com suas fronteiras. À direita, grafo mostrando a rede de conectividade entre os distritos (espaço relativo). O mapa da esquerda foi repetido por razões de melhor legibilidade.

A distinção entre espaço absoluto e espaço relativo é de grande importância para a Geografia. Milton Santos (Santos, 1985) refere-se ao “espaço dos fixos” e ao “espaço dos fluxos”. Castells (1999) fala em “espaço de lugares” e “espaços de fluxos”. Vejam o que Helen Couclelis comenta a respeito do tema:

*“Espaço absoluto, também chamado cartesiano, é um container de coisas e eventos, uma estrutura para localizar pontos, trajetórias e objetos. Espaço relativo, ou leibnitziano, é o espaço constituído pelas relações espaciais entre coisas” (Couclelis, 1997).*

Uma das escolhas básicas que fazemos na modelagem dos fenômenos geográficos é definir se utilizaremos representações no espaço absoluto ou no espaço relativo. Esta escolha depende primordialmente do tipo de análise que queremos realizar. Usualmente, consultas espaciais que envolvem dois tipos de entidades (“*quais os rios que cruzam esta estação ecológica?*”) requerem a representação no espaço absoluto. O mesmo vale para questões de álgebra de mapas (“*áreas inaptas tem declividade maior que 15% ou solos arenosos?*”). Quando os procedimentos de análise



envolvem apenas as relações de conectividade (“como chegar na estação de metrô Clínicas, partindo da estação Liberdade?” ou “qual é a média da mortalidade infantil de meus vizinhos?”) podemos utilizar representações no espaço relativo. Quando falamos em entidades como estradas, linhas de transmissão, conexões de água e esgoto, cadeias de mercado e linhas de comunicação, o espaço relativo é na maioria das vezes plenamente adequado.

### 1.5.3 Modelos no espaço absoluto: geo-campos e geo-objetos

Existem dois modelos formais para entidades geográficas no espaço absoluto: *geo-campos* e *geo-objetos*. O modelo de *geo-campos* enxerga o espaço geográfico como uma superfície contínua, sobre a qual variam os fenômenos a serem observados. Por exemplo, um mapa de vegetação associa a cada ponto do mapa um tipo específico de cobertura vegetal, enquanto um mapa geoquímico associa o teor de um mineral a cada ponto. O modelo de *geo-objetos* representa o espaço geográfico como uma coleção de entidades distintas e identificáveis, onde cada entidade é definida por uma fronteira fechada. Por exemplo, um cadastro urbano identifica cada lote como um dado individual, com atributos que o distinguem dos demais.

*Definição 1.1. Geo-Campo.* Um geo-campo representa um atributo que possui valores em todos os pontos pertencentes a uma região geográfica. Um *geo-campo*  $gc$  é uma relação  $gc = [R, A, f]$ , onde  $R \subset \mathfrak{R}^2$  é uma partição conexa do espaço,  $A$  é um atributo cujo domínio é  $D(A)$ , e a função de atributo  $f: R \rightarrow A$  é tal que, dado  $p \in R$ ,  $f(p) = a$ , onde  $a \in D(A)$ .

A noção de geo-campo decorre da definição física associada (segundo o Aurélio, “campo é um conjunto de valores de uma grandeza física que, numa região do espaço, dependem só das coordenadas dos pontos pertencentes a essa região”). Em outras palavras, para cada ponto do espaço, um campo terá um valor diferente.

*Definição 1.2 Geo-Objeto.* Um geo-objeto é uma entidade geográfica singular e indivisível, caracterizada por sua identidade, suas fronteiras, e seus atributos. Um *geo-objeto* é uma relação  $go = [id, a_1, \dots, a_n, G]$ , onde  $id$  é um identificador único,  $G$  é um conjunto de partições 2D conexas e

distintas  $\{R_p, \dots, R_n\}$  do espaço  $\mathfrak{R}^2$ , e  $a_i$  são os valores dos atributos  $A_p, \dots, A_n$ . Note-se que um geo-objeto pode ser composto por diferentes geometrias, onde cada geometria tem uma fronteira fechada (e.g., o Japão com suas diferentes ilhas).

Um exemplo de geo-campo (uma imagem IKONOS da cidade do Rio de Janeiro) e de um conjunto de geo-objetos (os distritos dessa cidade) é apresentado na Figura 1.5. A variável associada à imagem é a reflectância do solo, medida pelo sensor óptico do satélite. Os geo-objetos associados aos distritos de São Paulo são mostrados numa gradação de tons de cinza, cuja intensidade é proporcional ao índice de exclusão social (Sposati, 1996); quanto mais escuro, mais o distrito possui moradores em situação de exclusão social. Os dados na Figura 1.3 acima (geologia e declividade) também são exemplos de geo-campos.

A Figura 1.5 também ilustra uma questão importante: *existem diferenças fundamentais entre geo-campos e geo-objetos? Ou seriam apenas duas maneiras de ver o mesmo tipo de dado?* Considere os retângulos desenhados no interior das duas representações mostradas. Na figura à esquerda, o interior do retângulo tem as mesmas propriedades do geo-campo que o contém. Para cada ponto interior ao retângulo, podemos recuperar o valor do atributo (neste caso, a reflectância da imagem). Verificamos que *uma partição espacial genérica de um geo-campo compõe outro geo-campo com as mesmas propriedades.*

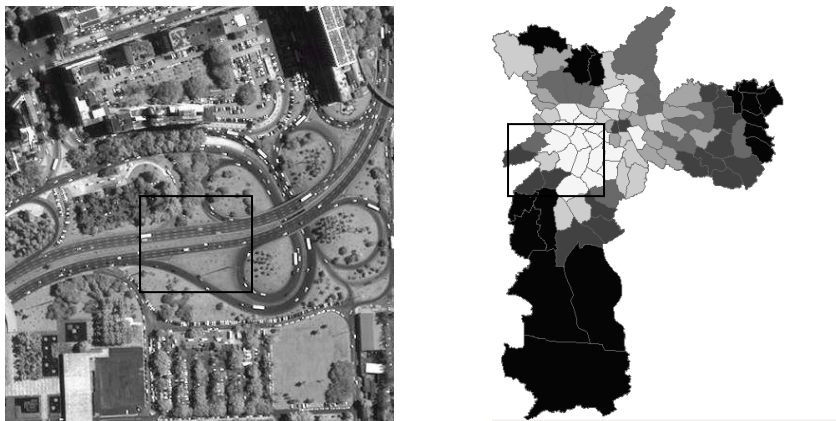


Figura 1.5 – Exemplo de geo-campo (imagem IKONOS do Rio de Janeiro) e de conjunto de geo-objetos (distritos da cidade de São Paulo).

Considere agora a figura da direita (distritos de São Paulo). O interior do retângulo mostrado não define mais um conjunto de geo-objetos com as mesmas propriedades do conjunto completo. O retângulo intercepta parcialmente alguns objetos. Como cada objeto é único e não pode ser dividido sem perder suas características originais, verificamos que *uma partição espacial genérica de um conjunto de geo-objetos não compõe outro conjunto de geo-objetos com as mesmas propriedades*.

A diferença essencial entre um geo-campo e um geo-objeto é o papel da fronteira. A fronteira de um geo-campo é uma divisão arbitrária relacionada apenas com nossa capacidade de medida. Na Figura 1.5, os limites da imagem correspondem apenas a eventuais limitações do instrumento sensor e não do fenômeno medido. Assim, o geo-campo pode ser dividido em partes e ainda assim manter sua propriedade essencial (que é sua função de atributo).

Por contraste, um geo-objeto é essencialmente definido por sua fronteira, que o separa do mundo exterior; ele não pode ser dividido e manter suas propriedades essenciais. Dentro da fronteira, todas as propriedades do objeto são constantes. Tomemos um distrito de São Paulo, como a Sé, que tem um código único de identificação no censo do

IBGE. Se dividirmos a Sé em duas partes, precisamos de dois novos códigos de identificação para caracterizar os dois novos distritos.

O exame da Figura 1.5 ilustra outra propriedade dos geo-objetos. É bastante comum lidarmos com um conjunto de geo-objetos que representam uma partição consistente do espaço; isto é, os recobrimentos espaciais destes objetos não se interceptam e eles possuem o mesmo conjunto de atributos. Estas características fazem com que possamos agrupar estes objetos numa coleção.

*Definição 1.3 Coleção de geo-objetos.* Uma coleção de geo-objetos é relação  $cgo = [id, o_1, \dots, o_n, A_1, \dots, A_n]$ , onde  $id$  é um identificador único, e  $o_1, \dots, o_n$  são geo-objetos que possuem os atributos  $A_1, \dots, A_n$ . Usualmente, se  $R_i$  for a região geográfica associada a  $o_i$ , temos  $R_i \cap R_j = \emptyset, \forall i \neq j$ . Deste modo, uma coleção reúne geo-objetos cujas fronteiras não se interceptam, e têm o mesmo conjunto de atributos.

O uso de coleções de geo-objetos é bastante freqüente em bancos de dados geográficos, pois é muito conveniente tratar geo-objetos similares de forma consistente. Por exemplo, falamos dos distritos da cidade de São Paulo, dos municípios do estado do Ceará, e das reservas indígenas da Amazônia. A idéia de coleções de geo-objetos é ainda útil para propormos um modelo orientado-a-objetos para dados geográficos, discutido a seguir.

#### 1.5.4 Modelos no espaço relativo: redes

O modelo de redes concebe o espaço geográfico como um conjunto de pontos no espaço (chamados de nós), conectados por linhas (chamados arcos), onde tanto os nós quanto os arcos possuem atributos. Os fenômenos modelados por redes incluem fluxo de pessoas ou materiais, conexões de influência, linhas de comunicação e acessibilidade. Um dos atrativos do modelo de redes é que o suporte matemático para este modelo (*a teoria de grafos*) é uma área de pesquisa consolidada (Bondy e Murty, 1976) (Gross e Yelen, 1998).

O problema que deu início à teoria dos grafos foi uma questão espacial. Em 1736, o matemático Leonard Euler vivia na cidade de Königsberg (na época parte da Prússia; hoje chamada Kaliningrad e pertencente à Rússia) onde haviam duas ilhas próximas no meio da

cidade, cruzadas por sete pontes (ver Figura 1.6 à esquerda). Euler se perguntou se havia uma maneira de fazer um circuito fechado (sair e voltar para um mesmo lugar), cruzando cada uma das pontes apenas uma vez. Ele construiu um grafo equivalente (ver Figura 1.6 à direita) e demonstrou que o problema era insolúvel.

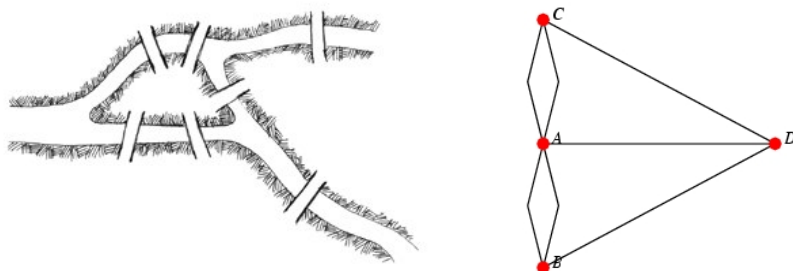


Figura 1.6 – As sete pontes de Königsberg e o grafo equivalente.

*Definição 1.4 Redes.* Uma rede é uma estrutura geográfica que tem como suporte um grafo  $G = [N, A, \varphi]$ , onde  $N$  é um conjunto de nós,  $A$  é um conjunto de arcos (arestas), e  $\varphi(a) = (u, v)$  é uma função de incidência que associa cada arco  $a \in A$  a um par de nós  $(u, v) \in N$ . No caso geográfico, os nós podem estar associados a uma localização  $(x, y)$  do espaço para fins de referência.

Como os nós de uma rede são abstrações de entidades existentes no espaço, eles podem estar associados aos seus atributos descritivos. Por exemplo, na rede mostrada na Figura 1.4, cada nó está associado a um distrito de São Paulo, e poderia ter diferentes atributos que descrevem este distrito. Também os arcos de uma rede podem ter propriedades, como o custo de percorrimento de um nó a outro. As propriedades mensuráveis das redes incluem operações diretas computáveis sobre a topologia do grafo, como qual o caminho ótimo entre dois nós. Também podemos computar operações matemáticas que envolvem apenas as relações de conectividade, como os indicadores locais de autocorrelação espacial (veja-se a respeito, Druck et al., 2004).

A definição de redes pode ser estendida para considerar o caso de conexões bidirecionais, como no caso de redes de transporte, onde as relações entre os nós não são simétricas, pois os fluxos em sentidos opostos podem ser diferentes. A Figura 1.7 ilustra uma rede simples e uma rede com conexões bidirecionais.

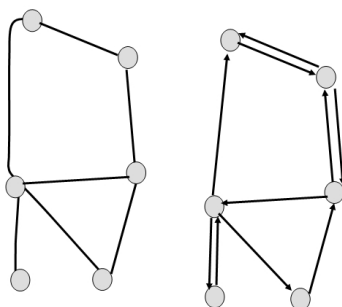


Figura 1.7 – Exemplos de redes simples e de redes com conexões bidirecionais.

Os modelos de rede têm grande utilidade em problemas de geoinformação, incluindo assuntos como gerenciamento de serviços como água, esgoto, eletricidade e telefonia. Para maiores referências, deve-se consultar Birkin et al (1996) e Godin (2001).

### 1.5.5 Um modelo orientado-a-objetos para dados geográficos

As seções anteriores nos permitem apresentar um modelo orientado-a-objetos que apresenta uma versão unificada dos dados geográficos, com base nos conceitos básicos de *geo-campo*, *coleção de geo-objetos* e *rede*. Para fins de organização lógica, o modelo considera a existência de uma classe genérica, chamada de *plano de informação* (ou *layer*), que é uma generalização destes dois conceitos. O conceito de *plano de informação* captura uma característica comum essencial dos três conceitos básicos: cada instância deles é referente a uma localização no espaço e tem um identificador único. Assim, o uso do conceito de *plano de informação* permite organizar o banco de dados geográfico e responder a perguntas como: “*Quais são os dados presentes no banco, qual o modelo associado a cada um e qual a região geográfica associada?*” Adicionalmente, como cada

geo-campo está associado a uma única função de atributo, ele pode ser especializado em *geo-campo temático* (associado a medidas nominais ou ordinais) e *geo-campo numérico* (associados a medidas por intervalo ou por razão). Com estes seis conceitos, construímos um modelo formal básico para dados geográficos, mostrado na Figura 1.8.

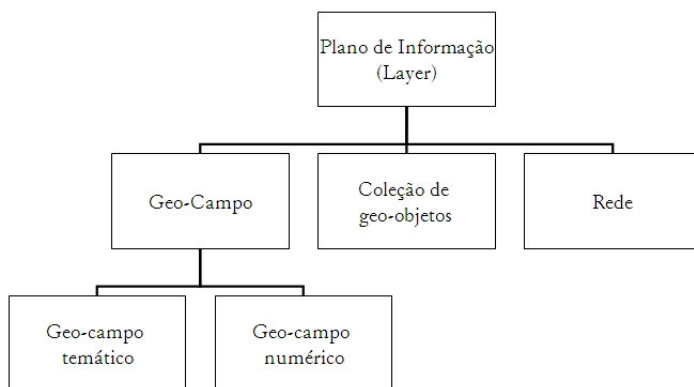


Figura 1.8 – Modelo OO básico para dados geográficos.

O modelo mostrado na Figura 1.8 serve de base para a maioria dos modelos de dados orientados-a-objetos adotados atualmente em geoinformação:

- O software SPRING (Câmara et al., 1996) inclui os conceitos de *rede*, *geo-campo numérico* e *geo-campo temático*, *coleção de geo-objetos* (chamada de *mapa cadastral*). Os geo-campos numéricos admitem as *imagens* como caso particular.
- No ArcGIS (ESRI, 2000), a coleção de geo-objetos é chamada de *features* (*feições*). Os geo-campos numéricos são chamados de *surfaces* (*superfícies*), e as imagens também são modeladas como caso particular de geo-campos numéricos. As redes (*networks*) também são incluídas.
- No modelo OpenGIS (OGC, 1998), os geo-campos são chamados de *coverage*, e a coleção de objetos é chamada de *feature collection*. O modelo OpenGIS não tem o conceito explícito de *layer*, mas

considera que as visões de *feature collection* e *coverage* são complementares.

- Na TerraLib (vide Capítulo 12 deste livro), o conceito de plano de informação (*layer*) é um conceito usado para organizar a informação no banco de dados. Os conceitos de *geo-campos* e de coleções de *geo-objetos* são implícitos. Como se trata de uma biblioteca, os designers da TerraLib quiseram permitir diferentes alternativas de projeto de sistema.

### 1.6 Do universo ontológico ao universo formal

Para passar do universo ontológico para o universo formal, precisamos responder à pergunta: *como os conceitos da ontologia de aplicação são formalizados?* Colocando o problema de forma mais geral: *Que critérios deve satisfazer um conceito para que seja utilizável em estudos quantitativos associados à geoinformação?* Tais critérios são:

- O conceito deve ser passível de ser associado a propriedades mensuráveis.
- Estas propriedades devem ser medidas no território e devem permitir diferenciar as diferentes localizações.
- Os resultados quantitativos e os modelos matemáticos utilizados devem ser validados em estudos de campo, que devem incluir dimensões objetivas e subjetivas do fenômeno em questão.

Para representar um conceito genérico como “exclusão social”, precisamos definir precisamente quais atributos caracterizam a exclusão social e como podemos medi-los no território. Esta caracterização realiza a passagem do universo ontológico para o universo formal. Com base em conceitos bem estabelecidos e associados a medidas quantitativas no espaço, podemos construir *territórios digitais*. O processo pode ser resumido na Figura 1.9.



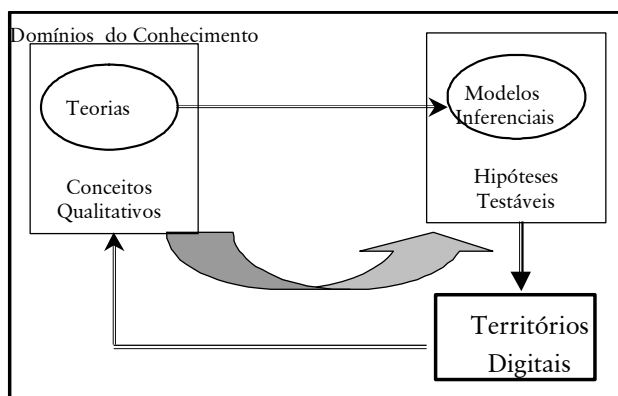


Figura 1.9 – Relação entre a construção dos territórios digitais e as teorias disciplinares (cortesia de Silvana Amaral Kampel).

Os especialistas desenvolvem teorias gerais sobre os fenômenos, que incluem o estabelecimento de conceitos organizadores de sua pesquisa (como “exclusão” ou “vulnerabilidade”). Para passar destas teorias para a construção computacional, é necessário que o especialista formule modelos inferenciais quantitativos. Estes modelos devem ser submetidos a testes de validação e de corroboração, através dos procedimentos de análise quantitativa. Os resultados numéricos podem então dar suporte ou ajudar a rejeitar conceitos qualitativos.

Após definir como que atributos mensuráveis serão associados ao conceito, o projetista do sistema de informação deverá decidir se este conceito será modelado no *espaço absoluto* ou no *espaço relativo*. A decisão deve-se dar essencialmente em função das propriedades que queremos medir. Se a localização exata é fundamental, ou se precisamos saber o valor do fenômeno em todos os pontos da região de estudo, então é necessário usar os modelos de espaço absoluto. Se o fluxo e as conexões são essenciais, então podemos usar o modelo de rede.

Se precisamos dos dados expressos no espaço absoluto, então devemos escolher ainda qual o modelo apropriado (geo-campo ou geo-objeto). Para isto, a decisão depende essencialmente do papel da fronteira. Se as fronteiras são parte essencial das entidades modeladas, estamos tratando com *indivíduos* e não com *topografias* (vide Tabela 1.1) e o modelo de

geo-objetos é o mais adequado. Senão, usaremos os modelos de geo-campos.

## 1.7 Universo estrutural

As estruturas de dados utilizadas em bancos de dados geográficos podem ser divididas em duas grandes classes: *estruturas vetoriais* e estruturas matriciais.

### 1.7.1 Estruturas de dados vetoriais

As estruturas vetoriais são utilizadas para representar as coordenadas das fronteiras de cada entidade geográfica, através de três formas básicas: pontos, linhas, e áreas (ou polígonos), definidas por suas coordenadas cartesianas, como mostrado na Figura 1.10.

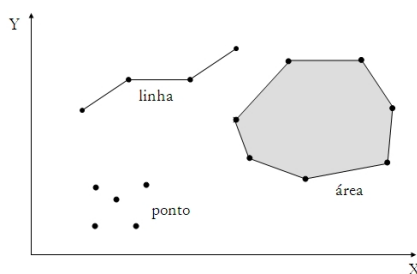


Figura 1.10 – Representações vetoriais em duas dimensões.

Um *ponto* é um par ordenado  $(x, y)$  de coordenadas espaciais. O ponto pode ser utilizado para identificar localizações ou ocorrências no espaço. São exemplos: localização de crimes, ocorrências de doenças, e localização de espécies vegetais. Uma *linha* é um conjunto de pontos conectados. A linha é utilizada para guardar feições unidimensionais. De uma forma geral, as linhas estão associadas a uma topologia arco-nó, descrita a seguir. Uma *área* (ou *polígono*) é a região do plano limitada por uma ou mais linhas poligonais conectadas de tal forma que o último ponto de uma linha seja idêntico ao primeiro da próxima. Observe-se também que a fronteira do *polígono* divide o plano em duas regiões: o interior e o exterior. Os polígonos são usados para representar unidades

espaciais individuais (setores censitários, distritos, zonas de endereçamento postal, municípios). Para cada unidade, são associados dados oriundos de levantamentos como censos e estatísticas de saúde.

### 1.7.2 Vetores e topologia: o caso dos geo-objetos

A topologia é a parte da matemática na qual se investigam as propriedades das configurações que permanecem invariantes nas transformações de rotação, translação e escala. No caso de dados geográficos, é útil ser capaz de determinar relações como adjacência (“vizinho de”), pertinência (“vizinho de”), intersecção, e cruzamento.

Objetos de área podem ter duas formas diferentes de utilização: como objetos *isolados* ou objetos *adjacentes*. O caso de objetos isolados é bastante comum em SIG urbanos, e ocorre no caso em que os objetos da mesma classe em geral não se tocam. Por exemplo, edificações, piscinas, e mesmo as quadras das aplicações cadastrais ocorrem isoladamente, não existindo segmentos poligonais compartilhados entre os objetos. Finalmente, temos objetos adjacentes, e os exemplos típicos são todas as modalidades de divisão territorial: bairros, setores censitários, municípios e outros. Neste caso, pode-se ter o compartilhamento de fronteiras entre objetos adjacentes, gerando a necessidade por estruturas topológicas. Estes também são os casos em que recursos de representação de buracos e ilhas são mais necessários.

Quando queremos armazenar as estruturas de dados do tipo polígono no caso de *objetos adjacentes*, temos uma decisão básica a tomar: guardamos as coordenadas de cada objeto isoladamente, e assim duplicamos as fronteiras em comum com outros objetos, ou armazenamos cada fronteira comum uma única vez, indicando a que objetos elas estão associadas? No primeiro caso é chamado de *polígonos sem topologia* e o segundo, de *topologia arco-nó-polígono*, comparados na Figura 1.11.

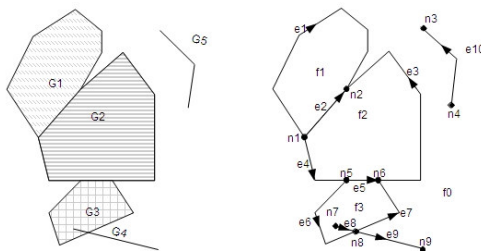


Figura 1.11 – Polígonos sem topologia (à esquerda) e topologia arco-nó-polígono (à direita). (Fonte: Ravada, 2003).

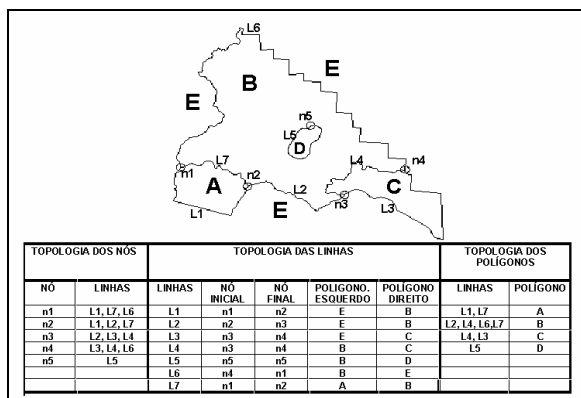


Figura 1.12 – Topologia arco-nó-polígono.

A topologia arco-nó-polígono, como mostrado na Figura 1.12, requer três listas separadas. Os pontos inicial e final de cada linha são chamados de nós. Para cada nó, armazenamos as linhas nele incidentes. Para cada linha, armazenamos os nós inicial e final, permitindo assim que a linha esteja associada a um sentido de percorrimento; guardamos ainda os dois polígonos separados por cada linha (à esquerda e à direita, considerando o sentido de percorrimento). Para cada polígono, guardamos as linhas que definem sua fronteira.

### 1.7.3 Vetores e topologia: o caso das redes

Objetos de linha podem ter variadas formas de utilização. Analogamente aos objetos de área, podemos ter objetos de linha isolados, em árvore e em rede. Objetos de linha isolados ocorrem, por exemplo, na representação de muros e cercas em mapas urbanos. Objetos de linha organizados em uma árvore podem ser encontrados nas representações de rios e seus afluentes, e também em redes de esgotos e drenagem pluvial. E podem ser organizados em rede, nos casos de redes elétricas, telefônicas, de água ou mesmo na malha viária urbana e nas malhas rodoviária e ferroviária.

No caso das redes, é fundamental armazenar explicitamente as relações de adjacência, utilizamos a *topologia arco-nó*. Um *nó* pode ser definido como o ponto de intersecção entre duas ou mais linhas, correspondente ao ponto inicial ou final de cada linha. Nenhuma linha poderá estar desconectada das demais para que a topologia da rede possa ficar totalmente definida. Para exemplificar, considere-se a Figura 1.13, que mostra um exemplo de como a topologia arco-nó pode ser armazenada.

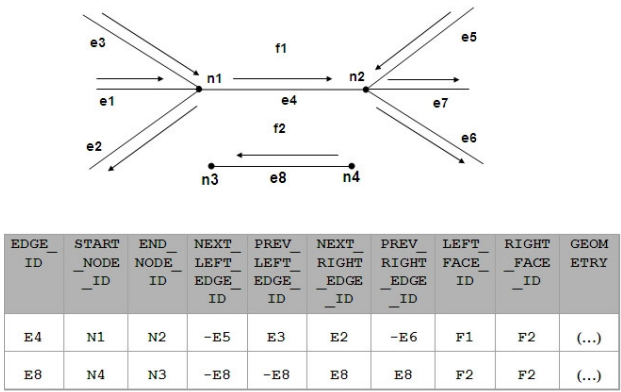


Figura 1.13 – Estrutura de dados para topologia arco-nó no Oracle Spatial SGBD (Fonte: Ravada, 2003).

### 1.7.4 Vetores e topologia: o caso dos dados 2,5 D

Uma das possibilidades associadas a dados vetoriais é a associação de valores que denotem a variação espacial de uma grandeza numérica. No caso mais simples, associamos a cada localização no espaço um valor numérico de atributo. Neste caso, como os valores de localização estão no plano e o valor adicional descreve uma superfície sobre este plano. Os dados resultantes são chamados de dimensão “dois e meio”, pois não se tratam estritamente de dados tridimensionais, pois o suporte espacial ainda são localizações 2D. A Figura 1.14 ilustra exemplo de dados de dimensão 2,5.

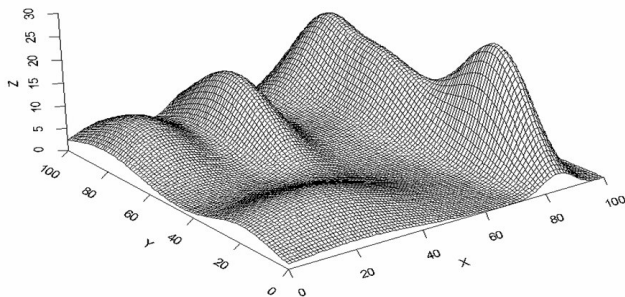


Figura 1.14 – Exemplo de dado com dimensão 2,5 (cortesia de Renato Assunção).

A maneira mais comum de armazenar estes dados é através de estruturas matriciais (vide próxima seção). Temos três alternativas que usam estruturas vetoriais:

- Conjunto de amostras esparsas 2,5D, constituído de pares ordenados  $(x,y,z)$ , onde  $(x,y)$  é uma localização no plano e  $z$  um valor numérico de atributo.
- Conjunto de isolinhas (curvas de nível), que são linhas às quais estão associados valores numéricos. As isolinhas não se cruzam, e são entendidas como estando “empilhadas” umas sobre as outras.
- A *malha triangular* ou TIN (do inglês “triangular irregular network”) é uma estrutura do tipo vetorial com topologia do tipo *nó-arco* e

representa uma superfície através de um conjunto de faces triangulares interligadas.

A malha triangular é a estrutura vetorial mais utilizada para armazenar dados 2,5D. Cada um dos três vértices da face do triângulo armazenados as coordenadas de localização  $(x, y)$  e o atributo  $z$ , com o valor de elevação ou altitude. Em geral, nos SIGs que possuem pacotes para MNT, os algoritmos para geração da malha triangular baseiam-se na triangulação de Delaunay com restrição de região. Quanto mais equiláteras forem as faces triangulares, maior a exatidão com que se descreve a superfície. O valor de elevação em qualquer ponto dentro da superfície pode ser estimado a partir das faces triangulares, utilizando-se interpoladores. A Figura 1.15 mostra uma superfície tridimensional e a grade triangular correspondente.

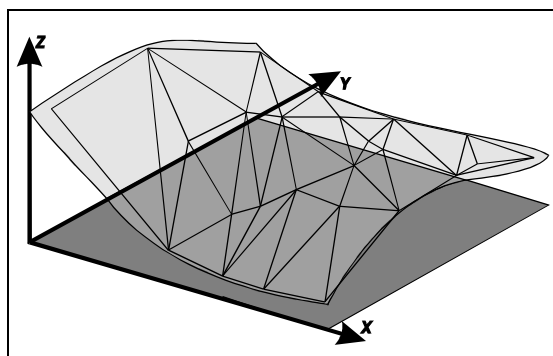


Figura 1.15 – Superfície e malha triangular correspondente. (cortesia de Laércio Namikawa ).

### 1.7.5 Hierarquia de representações vetoriais

Para um entendimento mais detalhado das representações vetoriais em GIS, deve-se inicialmente precisar o que se entende por primitivas geométricas: *coordenadas 2D*, *coordenadas 2,5D*, *nó 2D*, *nó 2,5D*, *nó de rede*, *arcos*, *arcos orientados*, *isolinhas* e *polígonos*. Dada uma região  $R \subset \mathfrak{R}^2$ , pode-se definir:

- COORDENADA\_2D – Uma coordenada 2D é um objeto composto por uma localização singular  $(x_p, y_j) \in R$ ;
- COORDENADA\_2,5D – Uma coordenada 2,5D é um objeto composto por uma localização singular  $(x_p, y_j, z)$ , onde  $(x_p, y_j) \in R$ ;
- PONTO2D – Um ponto 2D é um objeto que possui atributos descritivos e uma coordenada 2D;
- LINHA2D – Uma linha 2D possui atributos e inclui um conjunto de coordenadas 2D;
- ISOLINHA – uma isolinha contém uma linha 2D associada a um valor real (cota);
- ARCO ORIENTADO – um arco orientado contém uma linha 2D associada a uma orientação de percorrimto;
- NÓ2D – um nó 2D inclui uma coordenada2D  $(x_p, y_i) \in R$  e uma lista  $L$  de linhas 2D (trata-se da conexão entre duas ou mais linhas, utilizada para manter a topologia da estrutura);
- NÓ REDE – um nó de rede contém um nó 2D e uma lista de arcos orientados;
- NÓ 2,5D – um nó 2,5D instância desta classe contém uma coordenada 2,5D  $(x_p, y_p, z_i)$  e um lista  $L$  de linhas 2D (trata-se da conexão entre três ou mais linhas de uma grade triangular);
- POLÍGONO – um polígono pode ser armazenado como uma lista de coordenadas 2D (caso dos geo-objetos sem topologia) ou por uma uma lista de linhas 2D e uma lista de nós 2D (caso de topologia *arco-nó-polígono*).

Uma vez definidas as primitivas geométricas vetoriais, pode ser estabelecida a hierarquia de representações geométricas vetoriais, como mostrado na Figura 1.16, onde distinguem-se os relacionamentos de especialização *é-um* (“is-a”), inclusão de uma instância *parte-de* (“part-of”), inclusão de um conjunto de instâncias *conjunto-de* (“set-of”) e inclusão de uma lista de identificadores de instâncias *lista-de* (“list-of”).



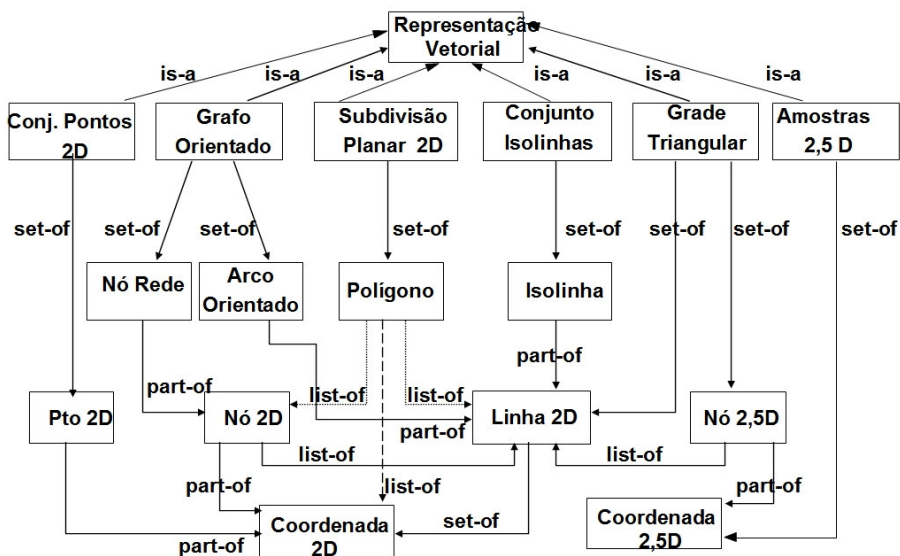


Figura 1.16 – Hierarquia de classes para estruturas vetoriais.

Distinguímos os seguintes tipos de estruturas de dados vetoriais:

- CONJUNTO DE PONTOS 2D – uma instância desta classe é um conjunto de pontos 2D utilizados para guardar localizações isoladas no espaço (p.ex. no caso de poços de petróleo);
- CONJUNTO DE ISOLINHAS – uma instância desta classe é um conjunto de linhas, onde cada linha possui uma cota e as linhas não se interceptam;
- SUBDIVISÃO PLANAR – para uma região geográfica  $R$  qualquer, uma subdivisão planar contém um conjunto  $P_g$  de polígonos que não se sobrepõem;
- GRAFO ORIENTADO – uma instância desta classe é uma representação composta de um conjunto de nó de rede e de um conjunto de arco orientado 2D;

- MALHA TRIANGULAR – uma instância desta classe contém um conjunto de nós 2,5D e um conjunto L de linhas 2D tal que todas as linhas se intersectam, mas apenas em seus pontos iniciais e finais;
- MAPA PONTOS 2,5D – uma instância desta classe é um conjunto de coordenadas 2,5D. Trata-se de um conjunto de amostras 2,5D.

### 1.7.6 Representação matricial

As estruturas matriciais usam uma grade regular sobre a qual se representa, célula a célula, o elemento que está sendo representado. A cada célula, atribui-se um código referente ao atributo estudado, de tal forma que o computador saiba a que elemento ou objeto pertence determinada célula. Nesta representação, o espaço é representado como uma matriz  $P(m, n)$  composto de  $m$  colunas e  $n$  linhas, onde cada célula possui um número de linha, um número de coluna e um valor correspondente ao atributo estudado e cada célula é individualmente acessada pelas suas coordenadas.

A representação matricial supõe que o espaço pode ser tratado como uma superfície plana, onde cada célula está associada a uma porção do terreno. A resolução do sistema é dada pela relação entre o tamanho da célula no mapa ou documento e a área por ela coberta no terreno, como mostrado na Figura 1.17.

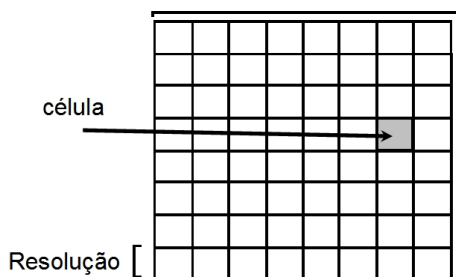


Figura 1.17 – Estrutura matricial.

A estrutura matricial pode ser utilizada para representar diferentes tipos de dados:

- *Grade regular*: representação matricial de dimensão “dois e meio” na qual cada elemento da matriz está associado a um valor numérico, como mostra a Figura 1.18 à esquerda.
- *Matriz temática*: representação matricial 2D na qual cada valor da matriz é um código correspondente à uma classe do fenômeno estudado, como mostra a Figura 1.18 à direita.

15	17	20	21	23	1	1	2	2	2
16	18	22	23	24	1	1	2	2	2
19	19	22	23	24	1	1	2	2	2
23	23	27	28	28	2	2	3	3	3
22	22	31	32	33	2	2	3	3	3

Figura 1.18 – À esquerda, grade regular com valores de temperatura em graus Celsius e, à direita, matriz temática com dados classificados (1 = “15-20 graus”, 2 = “20-25 graus”, 3 = “25-35 graus”).

### 1.7.7 Espaços celulares: generalização de estruturas matriciais

Um *espaço celular* é uma estrutura matricial generalizada onde cada célula está associada a vários tipos de atributos. Os espaços celulares têm várias vantagens sobre estruturas matriciais simples. Usando matrizes com um único atributo (como o caso dos dados mostrados na Figura 1.18), um fenômeno espaço-temporal complexo precisa de várias matrizes separadas para ser representado, o que resulta em maior dificuldade de gerência e de interface. Num espaço celular, a mesma célula está associada a diferentes informações, com ganhos significativos de manuseio dos dados.

Os espaços celulares são muito convenientes para armazenamento em bancos de dados objeto-relacionais. Toda a estrutura de um espaço

celular pode ser armazenada numa única tabela, o que faz o manuseio dos dados ser bem mais simples que os dados vetoriais ou mesmo que os dados matriciais indexados. Aplicações como álgebra de mapas e modelagem dinâmica ficam mais simples de implementar e operar. Um exemplo de espaço celular é mostrado na Figura 1.19, onde mostramos uma parte de um banco de dados onde há um espaço celular onde a Amazônia foi dividida em células de 25 x 25 km<sup>2</sup>; cada uma delas está associada a diferentes atributos socioeconômicos e ambientais (na Figura 1.19, o atributo visualizado é umidade média nos três meses mais secos do ano). Os espaços celulares ainda não são estruturas de dados comuns nos bancos de dados geográficos, e atualmente apenas a TerraLib tem suporte para este tipo de estrutura (vide Capítulo 12). Com a ênfase crescente dos SIG em modelos dinâmicos, podemos prever que esta estrutura será futuramente amplamente disponível nas diferentes implementações de bancos de dados geográficos.

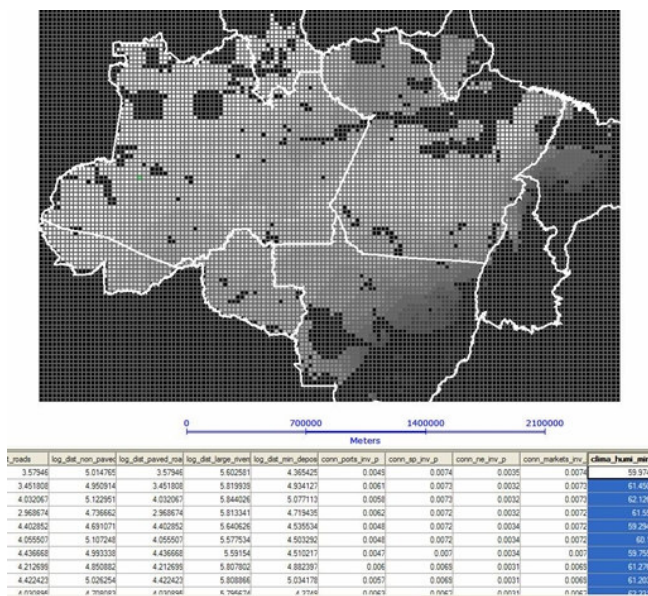


Figura 1.19 – Espaço celular com a Amazônia dividida em células de 25 x 25 km<sup>2</sup>; o atributo visualizado é umidade média nos três meses mais secos do ano (cortesia: Ana Paula Dutra de Aguiar).

## 1.8 Do universo formal para o universo estrutural

A passagem do universo formal (geo-campos, geo-objetos e redes) para o universo estrutural não é unívoca. Para cada tipo de entidade do modelo formal, há diferentes possibilidades de uso de estruturas de dados, a saber:

- *Geo-objetos*: como as fronteiras são elementos essenciais, são usualmente armazenados em estruturas poligonais, com as opções *polígonos sem topologia* ou *topologia arco-nó-polígono*.
- *Redes*: como a topologia é parte essencial, as redes devem ser armazenadas como um grafo orientado.
- *Geo-campos numéricos*: podem ser armazenados como amostras 2,5D, malhas triangulares ou grades regulares.
- *Geo-campos temáticos*: admitem o armazenamento como estruturas vetoriais (polígonos) ou matriciais (matrizes temáticas).

Os diferentes compromissos de armazenamento para as entidades do modelo formal são discutidos a seguir. Note-se que um espaço celular (discutido na Seção 1.7.7) pode guardar uma combinação arbitrária de geo-campos numéricos e temáticos.

### 1.8.1 Estruturas de dados para geo-objetos

A escolha entre estruturas topológicas ou não-topológicas para geo-objetos em bancos de dados geográficos depende também do suporte oferecido pelo SGBD. Nos SIG cujas estruturas de dados geométricas são manuseadas fora do SGBD (como o SPRING e o Arc/Info), é comum a escolha da topologia arco-nó-polígono. No caso dos bancos de dados geográficos, a maneira mais simples de armazenar geo-objetos é guardando cada um deles separadamente, o que implica em estruturas não-topológicas. Esta forma de trabalho foi sancionada pelo consórcio Open GIS e é suportada pelos diferentes SGBDs (Oracle, PostgreSQL, MySQL). No entanto, várias aplicações requerem o uso da topologia arco-nó-polígono, e alguns SGBDs com suporte espacial já estão incluindo esta opção, com o Oracle Spatial (Ravada, 2003).

### 1.8.2 Estruturas de dados para geo-campos temáticos

Geo-campos temáticos admitem tanto a representação matricial quanto a vetorial. Para a produção de cartas e em operações onde se requer maior precisão, a representação vetorial é mais adequada. As operações de álgebra de mapas são mais facilmente realizadas no formato matricial. No entanto, para um mesmo grau de precisão, o espaço de armazenamento requerido por uma representação matricial é substancialmente maior. Isto é ilustrado na Figura 1.20.

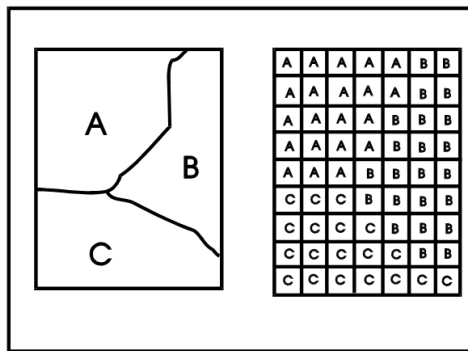


Figura 1.20 – Geo-campo temático em estruturas vetorial e matricial.

A Tabela 1.3 apresenta uma comparação entre as vantagens e desvantagens de armazenamento matricial e vetorial para geo-campos temáticos. Esta comparação leva em conta os vários aspectos: relacionamentos espaciais, análise, armazenamento. Nesta tabela, o formato mais vantajoso para cada caso é apresentado em destaque.

O armazenamento de geo-campos temáticos em estruturas vetoriais é uma herança da cartografia, onde limites entre classes temáticas eram desenhados com precisão em mapas. No entanto, sabemos que estes limites são imprecisos, na grande maioria dos casos. Assim, como nos ensina Peter Burrough, as estruturas matriciais são mais adequadas:

*“Os limites desenhados em mapas temáticos (como solo, vegetação, ou geologia) raramente são precisos e desenhá-los como linhas finas muitas vezes não representa adequadamente seu caráter. Assim, talvez*

*não nos devemos preocupar tanto com localizações exatas e representações gráficas elegantes. Se pudermos aceitar que limites precisos entre padrões de vegetação e solo raramente ocorrem, nós estaríamos livres dos problemas de erros topológicos associados como superposição e interseção de mapas” (Burrough, 1986).*

Tabela 1.3 – Comparação entre estruturas vetoriais e matriciais para mapas temáticos

<i>Aspecto</i>	<i>Vetorial</i>	<i>Matricial</i>
<i>Armazenamento</i>	Por coordenadas (mais eficiente)	Requer mais espaço de armazenamento
<i>Algoritmos</i>	Problemas com erros geométricos	Processamento mais rápido e eficiente.
<i>Escalas de trabalho</i>	Adequado tanto a grandes quanto a pequenas escalas	Mais adequado para pequenas escalas (1:25.000 e menores)
<i>Análise, Simulação e Modelagem</i>	Representação indireta de fenômenos contínuos Álgebra de mapas é limitada	Representa melhor fenômenos com variação contínua no espaço Simulação e modelagem mais fáceis

### 1.8.3 Estruturas de dados para geo-campos numéricos

Para geo-campos numéricos, a escolha básica se dá entre malhas triangulares e grades regulares. As demais estruturas de dados (amostras 2,5D e isolinhas) são formatos intermediários, utilizados para entrada ou saída de dados, mas não adequadas para análise.

As malhas triangulares são normalmente melhores para representar a variação do terreno, pois capturam a complexidade do relevo sem a necessidade de grande quantidade de dados redundantes. As grades regulares têm grande redundância em terrenos uniformes e dificuldade de adaptação a relevos de natureza distinta no mesmo mapa, por causa da grade de amostragem fixa.

Para o caso de variáveis geofísicas e para operações como visualização 3D, as grades regulares são preferíveis, principalmente pela maior facilidade de manuseio computacional. A Tabela 1.4 resume as principais vantagens e desvantagens de grades regulares e malhas triangulares.

Tabela 1.4 – Estruturas para geo-campos numéricos

	Malha triangular	Grade regular
<i>Vantagens</i>	Melhor representação de relevo complexo	Facilita manuseio e conversão
	Incorporação de restrições como linhas de crista	Adequada para dados não-altimétricos
<i>Problemas</i>	Complexidade de manuseio	Representação relevo complexo Cálculo de declividade

#### 1.8.4 Representações computacionais de atributos de objetos

Entende-se por atributo qualquer informação descritiva (nomes, números, tabelas e textos) relacionada com um único objeto, elemento, entidade gráfica ou um conjunto deles, que caracteriza um dado fenômeno geográfico. Nos bancos de dados geográficos, os atributos de objetos geográficos são armazenados em relações convencionais. As representações geométricas destes objetos podem ser armazenadas na mesma tabela que os atributos ou em tabelas separadas, mas ligadas por identificadores únicos. Estes aspectos são discutidos em maior detalhe nos capítulos 3, 5, e 8 deste livro.

#### 1.9 Universo de implementação

No universo de implementação, são tomadas as decisões concretas de programação e que podem admitir número muito grande de variações. Estas decisões podem levar em conta as aplicações às quais o sistema é voltado, a disponibilidade de algoritmos para tratamento de dados geográficos e o desempenho do hardware. Neste livro, aspectos do universo de implementação são tratados em diferentes capítulos:



- Os algoritmos de geometria computacional para problemas como ponto-em-polígono, simplificação de linhas e intersecção de linhas e polígonos são tratados no Capítulo 2.
- Os problemas de indexação espacial, que representam um componente determinante no desempenho total do sistema, são abordados no Capítulo 6.
- As questões de processamento e otimização de consultas espaciais são discutidas no Capítulo 7.
- Uma discussão detalhada dos SGBDs com suporte espacial é apresentada no Capítulo 8.
- Os Capítulos 12 a 14 apresentam a biblioteca TerraLib, um ambiente para construção de aplicativos geográficos.

### **1.10 Leituras suplementares**

Este Capítulo apresentou uma visão geral dos diferentes aspectos envolvidos com a representação computacional dos dados geográficos, que é o grande objetivo dos bancos de dados geográficos. Para uma visão geral de geoinformação sob o ponto de vista da Ciência da Computação, a referência mais atualizada é Worboys e Duckham (2004). O livro de Druck et al (2004) apresenta uma discussão sobre as questões de análise espacial de dados geográficos. Sobre o tema de bancos de dados geográficos, os livros de Rigaux et al (2002) e Shekar e Chawla (2002) são leituras complementares a este livro. A coletânea editada por Sellis et al (2003) apresenta um conjunto de artigos excelentes sobre os problemas emergentes de bancos de dados espaço-temporais.

## Referências

- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. **Scientific American**, v. May, 2001.
- BIRKIN, M.; CLARKE, G.; CLARKE, M. P.; WILSON, A. **Intelligent GIS : Location Decisions and Strategic Planning**. New York: John Wiley, 1996.
- BONDY, J. A.; MURTY, U. S. R. **Graph Theory with Applications**. London: The Macmillan Press LTD, 1977.
- BÖNISCH, S.; ASSAD, M. L.; CÂMARA, G.; MONTEIRO, A. M. Representação e Propagação de Incertezas em Dados de Solos: I - Atributos Numéricos. **Revista Brasileira de Ciência do Solo**, v. 28, n.1, p. 33-47, 2004.
- BURROUGH, P. **Principles of Geographical Information Systems for Land Resources Assessment**. Oxford, England, Oxford University Press, 1986.
- CALDEIRA, T. **Cidade de Muros: Crime, Segregação e Cidadania em São Paulo (City of Walls: Crime, Segregation and Citizenship in Sao Paulo)**. São Paulo: Edusp, 2000.
- CÂMARA, G. **Modelos, Linguagens e Arquiteturas para Bancos de Dados Geográficos**. São José dos Campos, SP: Instituto Nacional de Pesquisas Espaciais (INPE), 1995.Ph.D., 1995.
- CÂMARA, G.; SOUZA, R.; FREITAS, U.; GARRIDO, J. SPRING: Integrating Remote Sensing and GIS with Object-Oriented Data Modelling. **Computers and Graphics**, v. 15, n.6, p. 13-22, 1996.
- CÂMARA, G. S., R.C.M.; MONTEIRO, A.M.V.; PAIVA, J.A.C; GARRIDO, J. Handling Complexity in GIS Interface Design. In: I Brazilian Workshop on Geoinformatics. SBC, Campinas, SP, 1999.
- CASTELLS, M. **A Sociedade em Rede**. São Paulo: Paz e Terra, 1999.
- CHEN, P. S. S. The Entity-Relationship Model: Towards a Unified View of Data. **ACM Transactions on Database Systems**, v. 1, n.1, p. 9-36, 1976.
- COUCLELIS, H. From Cellular Automata to Urban Models: New Principles for Model Development and Implementation. **Environment and Planning B: Planning and Design**, v. 24, p. 165-174, 1997.
- DAVIS, C.; BORGES, K.; LAENDER, A. OMT-G: An Object-Oriented Data Model for Geographic Applications. **GeoInformatica**, v. 3, n.1, 2002.

- DRUCK, S.; CARVALHO, M. S.; CÂMARA, G.; MONTEIRO, A. M. V. **Análise Espacial de Dados Geográficos**. Brasília: EMBRAPA (ISBN 85-7383-260-6), 2004.
- EGENHOFER, M. Spatial SQL: A Query and Presentation Language. **IEEE Transactions on Knowledge and Data Engineering**, v. 6, n.1, p. 86-95, 1994.
- EGENHOFER, M.; FRANZOSA, R. Point-Set Topological Spatial Relations. **International Journal of Geographical Information Systems**, v. 5, n.2, p. 161-174, 1991.
- ESRI, 2000, *Modelling Our World : The ESRI Guide to Geodatabase Design*, Redlands, CA.
- FONSECA, F.; DAVIS, C.; CAMARA, G. Bridging Ontologies and Conceptual Schemas in Geographic Applications Development. **Geoinformatica**, v. 7, n.4, p. 355-378, 2003.
- GODIN, L. **GIS in Telecommunications**. Redlands, CA: ESRI Press, 2001.
- GINZBURG, C. **Olhos de Madeira: Nove Reflexões sobre a Distância**. São Paulo: Companhia das Letras, 2001.
- GOMES, J.; VELHO, L. Abstraction Paradigms for Computer Graphics. **The Visual Computer**, v. 11, n.5, p. 227-239, 1995.
- GROSS, J.; YELLEN, J. **Graph Theory and Its Applications**. Boca Raton, FL: CRC Press, 1998.
- GRUBER, T. R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. **Int. Journal of Human-Computer Studies**, v. 43, p. 907-928, 1995.
- HOHL, P. **GIS Data Conversion: Strategies, Techniques, and Management**. Clifton Park, NY: OnWorld Press, 1998.
- KRAAK, M.-J.; BROWN, A., eds., 2001, **Web Cartography**. London, Taylor and Francis.
- KUHN, W.; FRANK, A., 1991. A Formalization of Metaphors and Image-Schemas in User Interfaces. In: MARK, D.; FRANK, A., eds., **Cognitive and Linguistic Aspects of Geographic Space**: Dordrecht, Kluwer Academic Publishers, p. 419-434.
- LI, Z.; ZHU, Q.; GOLD, C. **Digital Terrain Modeling: Principles and Methodology**. London: Taylor and Francis, 2004.
- MACEACHREN, A. M. **How Maps Work : Representation, Visualization, and Design**. New York: Guilford Press, 2004.

- MASSEY, D. S.; DENTON, N. A. **American Apartheid: Segregation and the Making of the Underclass**. Cambridge: Harvard University Press, 1993.
- MATHER, P. M. **Computer Processing of Remotely-Sensed Images : An Introduction (3rd ed)**. New York: John Wiley, 2004.
- MONMONIER, M. **Mapping It Out: Expository Cartography for the Humanities and Social Sciences**. Chicago: University of Chicago Press, 1993.
- NOY, N. F.; SINTEK, M.; DECKER, S.; CRUBEZY, M.; FERGERSON, R. W.; MUSEN, M. A. Creating Semantic Web Contents with Protege-2000. **IEEE Intelligent Systems**, v. 16, n.2, p. 60-71, 2001.
- OGC, 1998, The OpenGIS Specification Model: The Coverage Type and Its Subtypes, Wayland, MA, Open Geospatial Consortium.
- PARDINI, R.; SOUZA, S.; BRAGANETO, R.; METZGER, J.-P. The role of forest structure, fragment size and corridors in maintaining small mammal abundance and diversity in an Atlantic forest landscape. **Biological Conservation**, v. 124, p. 253-266, 2005.
- RAVADA, S. **Topology Management in Oracle Spatial 10g**. Dagstuhl Seminar on Computational Cartography and Spatial Modelling, 2003. <http://www.dagstuhl.de/03401/Materials/>.
- RICHARDS, J.; EGENHOFER, M. A Comparison of Two Direct-Manipulation GIS User Interfaces for Map Overlay. **Geographical Systems**, v. 2, n.4, p. 267-290, 1995.
- RIGAUX, P.; SCHOLL, M.; VOISARD, A. **Spatial Databases with Application to GIS**. San Francisco: Morgan Kaufman, 2002.
- RUMBAUGH, J.; BLAHA, M.; PREMERLANI, W.; EDDY, F.; LORENSEN, W. **Object-Oriented Modeling and Design**. Englewood Cliffs, NJ: Prentice-Hall, 1991.
- SANTOS, M. **Espaço e Método**. São Paulo: Nobel, 1985.
- SEARLE, J. R. Rationality and Realism, What is at Stake? **Dædalus**, v. 122, n.4, 1993.
- SEARLE, J. R. **Mind, Language and Society**. New York: Basic Books, 1998.
- SELLIS, T.; FRANK, A. U.; GRUMBACH, S.; GUTING, R. H.; KOUBARAKIS, M., eds., 2003, **Spatio-Temporal Databases: The Chorochronos Approach**: Berlin, Springer.
- SHEKHAR, S.; CHAWLA, S. **Spatial Databases: A Tour**. New York: Prentice-Hall, 2002.

- SMITH, B., 2003. Ontology and Information Systems. In: ZALTA, E. N., ed., **The Stanford Encyclopedia of Philosophy. The Metaphysics Research Lab, Center for the Study of Language and Information.**: Stanford, Stanford University.
- SMITH, B.; MARK, D. Ontology and Geographic Kinds. In: International Symposium on Spatial Data Handling. Vancouver, Canada, 1998. p. 308-320.
- SPOSATI, A. **Mapa de Exclusão/Inclusão Social de São Paulo.** São Paulo: EDUC, 1996.
- STEVENS, S. S. On the theory of scales of measurement. **Science**, v. 103, n.2684, p. 677-680, 1946.
- TOMLIN, C. D. **Geographic Information Systems and Cartographic Modeling.** Englewood Cliffs, NJ: Prentice-Hall, 1990.
- TORRES, H. Segregação residencial e políticas públicas: São Paulo na década de 1990 (Spatial segregation and public policies: São Paulo in the 1990s). **Revista Brasileira de Ciências Sociais (Brazilian Social Sciences Journal)**, v. 54, p. 41-56, 2004.
- TUFTE, E. **The Visual Display of Quantitative Information.** Cheshire, CT: Graphics Press, 1983.
- WHITE, M. J. The measurement of spatial segregation. **American Journal of Sociology**, v. 88, p. 1008-1018, 1983.
- WORBOYS, M. F.; DUCKHAM, M. **GIS - A Computing Perspective (2nd edition).** Boca Raton: CRC Press, 2004.