

UNIVERSITÀ DEGLI STUDI DI PADOVA



FACOLTÀ DI INGEGNERIA

Tesi di Laurea Magistrale in Ingegneria Informatica

**Generalizzazione cartografica della rete stradale
nel progetto CARGEN:
algoritmi di armonizzazione, selezione e
tipificazione alla scala 1:50 000**

Relatore: **Prof. Sergio Congiu**

Laureando: **Elia Nicastro**

Anno Accademico 2010-2011

*A Silvia,
per avermi accompagnato fino
a questo traguardo,
che segna l'inizio di una nuova strada a
doppia corsia da percorrere assieme.*

INDICE

Introduzione.....	1
1 La cartografia e i Gis.....	3
1.1 Cenni storici.....	3
1.2 Cenni storici sulla cartografia.....	3
1.3 Tipologia e scopi di una carta geografica.....	5
1.4 I GIS.....	7
1.5 Il processo cartografico.....	10
2 La generalizzazione.....	13
2.1 Introduzione al concetto di generalizzazione.....	13
2.2 Il processo di generalizzazione.....	16
2.2.1 Generalizzazione del modello.....	16
2.2.2 Generalizzazione cartografica.....	17
2.3 La generalizzazione automatica.....	17
2.4 Modelli di generalizzazione.....	19
2.5 Gli operatori della generalizzazione	21
2.5.1 Operatori geometrici.....	22
2.5.2 Operatori per la trasformazione degli attributi.....	24
2.5.3 Operatori e algoritmi.....	24
3 Il progetto di ricerca CARGEN.....	27
3.1 La situazione cartografica in Italia.....	27
3.2 Il progetto CARGEN.....	29
3.2.1 CTRN.....	29
3.2.2 GeoDBR.....	30
3.2.3 DB25.....	31
3.2.4 La derivazione del DB25.....	32
3.2.5 La derivazione del DB50.....	33
3.3 Ambiente di lavoro.....	34
3.3.1 JTS Topology Suite.....	35
3.3.2 OpenJump.....	35
3.3.3 Cartografia per il test.....	36
4 Generalizzazione della rete stradale.....	39
4.1 Approccio alla generalizzazione stradale.....	39
4.2 Obiettivi del lavoro.....	40
4.2.1 Armonizzazione.....	41
4.2.2 Selezione.....	44
4.2.3 Tipificazione.....	45
4.3 Stato dell'arte.....	45
4.4 Data enrichment come elemento chiave della generalizzazione.....	49

4.4.1	Rappresentazione della rete stradale.....	51
4.4.2	Misure di centralità.....	51
4.4.3	Local betweenness.....	54
4.5	Fasi del processo di generalizzazione.....	57
5	Armonizzazione del grafo stradale.....	61
5.1	Obiettivi del processo.....	61
5.1.1	Risaltare gli elementi importanti e diminuire i dettagli.....	63
5.1.2	Estendere le classi stradali del modello dati.....	63
5.1.3	Aumentare la robustezza degli algoritmi di generalizzazione.....	64
5.2	Algoritmi di armonizzazione.....	64
5.2.1	Declassificazione delle vie con scarsa funzione di collegamento.....	65
5.2.2	Riclassificazione degli elementi collineari	67
5.2.3	Riclassificazione delle vie di collegamento strategiche.....	69
5.2.4	Riclassificazione delle strade inserite in contesti anomali.....	71
5.2.5	Classificazione delle strade di quartiere.....	72
5.3	Il processo di armonizzazione.....	75
6	Sfoltimento del grafo stradale.....	77
6.1	Introduzione.....	77
6.2	La selezione delle strade dangling.....	78
6.3	La selezione delle maglie.....	80
6.3.1	Sfoltimento della viabilità negli agglomerati urbani.....	81
6.3.2	Sfoltimento della viabilità principale e secondaria.....	84
6.4	Il processo di selezione.....	85
7	Tipificazione di strutture parallele.....	87
7.1	Introduzione.....	87
7.2	Individuazione dei parallelismi.....	89
7.3	Generalizzazione dei parallelismi.....	92
	Conclusion.....	97
	Riferimenti bibliografici.....	99

Introduzione

Nella realizzazione di una mappa geografica la fase di generalizzazione riveste un ruolo fondamentale; lo scopo di questa fase è quello di astrarre la realtà per rappresentarla in modo che essa risulti chiara, comprensibile e leggibile, tenendo presente la funzione che deve svolgere la mappa, e in accordo con la scala di rappresentazione scelta. La generalizzazione è necessaria perché, in una mappa, non è generalmente possibile rappresentare tutti i fenomeni geografici: lo spazio per contenerli, infatti, diminuisce con il diminuire della scala.

Spesso, nella produzione di una cartografia a bassa scala, l'astrazione della realtà avviene a partire dalle informazioni geografiche contenute in una cartografia preesistente e a scala superiore. Questa astrazione, fino a qualche anno fa, era svolta interamente a mano dal cartografo, richiedendo tempi lunghissimi; con l'avvento al digitale, però, è stata intravista la possibilità di automatizzare questo processo, e di diminuirne conseguentemente tempi e costi.

In questo ambito si va a collocare il progetto di ricerca CARGEN, che mira a sviluppare un processo di generalizzazione automatico per la produzione di un database topografico in scala 1:25 000 e 1:50 000, partendo dalle carte tecniche regionali, fornite dalla Regione Veneto, in scala 1:5000.

Il presente lavoro di tesi si inserisce nel progetto CARGEN e si pone come obiettivo specifico la realizzazione di un processo automatico per la generalizzazione della rete stradale alla scala 1:50 000.

Come già menzionato, non è possibile rappresentare nella mappa la totalità delle strade, e questo problema risulta ancora più evidente quando si opera a scale minori; si rende perciò necessario, in concomitanza al cambiamento di scala, un processo di scrematura del grafo stradale. Anche nel caso delle strade, però, ritorna il problema della mancanza di spazio, che tuttavia non sempre risulta risolvibile mediante una mera selezione delle strade; il più delle volte si rende indispensabile anche una ridefinizione delle geometrie stradali, allo scopo di trasmettere un'informazione coerente con la scala di rappresentazione.

Per garantire questa coerenza, oltre alla selezione e alla ridefinizione delle strade, è

d'obbligo attuare un'ulteriore operazione di astrazione, che permetta di adattare il livello di dettaglio con un livello compatibile con la nuova scala, in modo da mettere in luce gli elementi effettivamente importanti e mettere invece in secondo piano quelli che sono di scarso rilievo.

In questa tesi, verranno trattate in dettaglio queste operazioni di astrazione, dando particolare spazio alla loro implementazione pratica.

L'intero lavoro è strutturato in sette capitoli.

Nel primo capitolo viene introdotto il tema della cartografia, definendo, in particolare, in cosa consiste il processo cartografico per la produzione delle mappe.

Nel secondo capitolo viene definita la generalizzazione cartografica e descritti alcuni modelli concettuali relativi a questo tema proposti in letteratura; vengono trattati, inoltre, gli operatori della generalizzazione, i veri strumenti tramite cui si concretizza la generalizzazione cartografica.

Il terzo capitolo è dedicato al progetto CARGEN; vengono delineati gli obiettivi del progetto e i risultati ottenuti fino ad oggi, descrivendo anche gli strumenti di lavoro e le cartografie utilizzati.

Nel quarto capitolo viene affrontato il tema della generalizzazione della rete stradale: vengono discussi gli obiettivi del lavoro di tesi, relativi all'armonizzazione, selezione e tipificazione delle strade; viene presentato lo stato dell'arte e definiti gli strumenti utilizzati per raggiungere gli scopi del lavoro. Chiude il capitolo la proposta di un processo di generalizzazione che consideri sia gli algoritmi sviluppati in questa tesi che quelli prodotti in lavori precedenti.

Il quinto capitolo affronta dettagliatamente l'armonizzazione del grafo stradale, con lo scopo specifico di riclassificare le strade per renderle adatte al livello di dettaglio permesso dalla scala 1:50 000.

Nel sesto capitolo viene discusso il problema dello sfoltimento del grafo stradale, suddividendolo in due sotto problemi e proponendo due approcci distinti per risolverli.

Infine, nel settimo capitolo viene trattata la tipificazione delle strade parallele, descrivendo una procedura per individuarle e semplificarle.

1 La cartografia e i Gis

Per introdurre il tema della generalizzazione cartografica, in questo capitolo verranno, per prima cosa, fornite alcune definizioni d'obbligo, relative ai concetti di cartografia, carta geografica e scala; seguirà un breve resoconto sulla storia della cartografia. Successivamente, saranno illustrate in modo generale le varie classificazioni e tipologie delle carte geografiche e i sistemi informativi geografici (GIS), per concludere infine con la descrizione del processo cartografico, processo tramite il quale si realizza una mappa.

1.1 Cenni storici

La cartografia è la scienza che si occupa di riprodurre l'aspetto della Terra o di una sua parte con la minima distorsione possibile, in modo che le informazioni contenute sulla carta siano il più possibile accurate (Enc. Italiana Grolier, 1987).

La carta geografica, che quindi è il prodotto di questa riproduzione cartografica, è una rappresentazione grafica, riprodotta in scala e secondo simboli convenzionali, di una parte o di tutta la superficie terrestre, o di alcuni suoi aspetti particolari (Diz. Devoto, 1979).

La scala secondo cui la carta geografica rappresenta la realtà è il rapporto fra le distanze risultanti sulla scala stessa e le corrispondenti distanze reali sulla superficie della Terra. Così, Le carte su grande scala riproducono una piccola zona e possono contenere numerosissimi dettagli; al contrario, le carte su piccola scala coprono un'area più vasta evidenziando un ristretto e selezionato numero di informazioni.

1.2 Cenni storici sulla cartografia

Nel corso della storia, la cartografia è stata utilizzata per rappresentare una grande quantità di informazioni diverse (Enc. Italiana Grolier, 1987), comprendendo dati geografici, statistici, demografici, economici, politici, culturali, e collegandoli ad un preciso luogo geografico.

Nell'antichità, le prime carte topografiche venivano utilizzate per rappresentare

appezzamenti di terra o per prescrivere i requisiti degli edifici da costruire. Ad esempio, gli Egiziani (Enc. Italiana Grolier, 1987), intorno al sec. XIV a. C., tracciavano mappe che indicavano i confini delle proprietà ai fini fiscali; tali mappe erano divenute necessarie a causa delle inondazioni annuali del Nilo, che cancellavano i termini confinari.

Nel periodo Greco-Romano l'uso della cartografia si espanse notevolmente. I Greci, abili navigatori e colonizzatori, divennero all'inizio del VI secolo a. C. il più importante centro delle conoscenze geografiche, grazie alle numerose spedizioni militari e navali che si rendevano necessarie a fini espansionistici e commerciali. Nel III secolo a. C. il matematico Eratostene di Cirene e il contemporaneo Ipparco di Nicea gettarono le fondamenta della cartografia scientifica, elaborando un sistema di latitudine e longitudine.

I Romani, successivamente, ereditari ereditando il sapere cartografico greco, espansero notevolmente le loro conoscenze geografiche, promuovendo rilevazioni sul territorio dell'impero, al fine di rappresentare i confini delle province e la posizione di città, strade e fiumi.

Durante il periodo medievale, tra il V e il XV secolo, le scienze geografiche non fecero grandi progressi. Gli studiosi di questo periodo storico, infatti, traevano essenzialmente le loro conoscenze dai classici e dalle Sacre Scritture e rappresentavano la Terra come un disco piatto. Il mondo veniva rappresentato tramite le mappae mundi, con Gerusalemme al centro, l'Asia in alto, l'Europa in basso a sinistra e l'Africa in basso a destra.

Durante il Rinascimento si ebbe un nuovo progresso cartografico, a seguito delle scoperte compiute fra il XV e XVII secolo da esploratori come Giovanni Caboto, Cristoforo Colombo e Ferdinando Magellano. Le carte si diffusero facilmente anche grazie a innovazioni tecnologiche come la stampa e il perfezionamento della tecnica dell'incisione su rame, che fecero diminuire i costi delle mappe e permisero un maggiore dettaglio. Le mappe divennero sempre più simili a quelle in uso attualmente; ad esempio, mentre prima per descrivere l'orografia¹ si usava una rappresentazione detta "a

¹ L'orografia è la quella branca della geografia fisica che studia i rilievi della Terra, sia quelli della superficie sia quelli sottomarini.

mucchi di talpa²”.

Per quanto riguarda il periodo moderno, la prima rilevazione completa di un territorio nazionale partì in Francia nel 1773 con la famiglia Cassini, che, su commissione del Governo, allestì la prima carta topografica completa della Francia in scala 1:86 400. In Italia, nel 1872, venne fondato l'Istituto Topografico Militare (diventato nel 1882 Istituto Geografico Militare), che nello stesso anno (1872) iniziò la rilevazione del territorio italiano per la realizzazione della nuova Carta Topografica d'Italia (1921) alla scala 1:100 000.

Sul finire del XX secolo, infine, si assistette ad una vera e propria rivoluzione cartografica, animata dall'impiego massiccio dell'informatica e dall'utilizzo di nuovi mezzi come i sistemi satellitari o le rilevazioni aeree. Nascono i sistemi informativi geografici (GIS, Geographical Information System), potenti strumenti di analisi dei dati geografici, che permettono non solo la visualizzazione, ma anche l'elaborazione e la manipolazione degli elementi di una carta.

1.3 Tipologia e scopi di una carta geografica

Tramite una carta geografica è teoricamente possibile visualizzare la distribuzione geografica di qualsiasi fenomeno, come ad esempio la distribuzione delle nascite nelle diverse regioni di uno stato o la densità di popolazione, oppure di qualsiasi combinazione di elementi. Per rispondere a questa varietà di scopi, esistono moltissime tipologie di carte, classificabili in base a diverse caratteristiche, tra cui le più comuni sono i contenuti e le finalità della carta e il valore di scala adottato.

Per quanto riguarda la prima modalità di classificazione, considerando i contenuti e le finalità di una carta si possono distinguere le carte tematiche e le carte topografiche.

Le carte tematiche sono rappresentazioni di fenomeni o di concezioni astratte qualitative o quantitative, limitate ad uno o più temi specifici; un esempio potrebbe essere allora una carta tematica che illustri la frequenza dei tornado o la quantità di piogge di un certo territorio. Le carte topografiche, invece, offrono una rappresentazione il più completa possibile dell'area considerata.

2 Nelle vecchie carte i rilievi erano raffigurati in modo schematico con trattini disposti a mucchi di talpa, che indicavano il grossolano profilo delle montagne osservate di fianco.

La carta topografica è la rappresentazione della superficie fisica di un territorio, costituita dal terreno con le sue forme, dalle acque (laghi, fiumi, mari), da oggetti concreti e durevoli (edifici, strade, ponti etc.); nella stessa carta sono contenute anche altre indicazioni come i confini amministrativi, linee o variazioni di colore indicanti l'altimetria e la batimetria, nomi dei luoghi. Generalmente questo tipo di carta è derivata da rilievi fotogrammetrici e la scala con cui si rappresentarla è generalmente compresa tra 1:10 000 e 1:100 000.

Una seconda modalità di classificazione usata comunemente si basa, come già accennato, sul valore di scala adottato. In questo caso, le mappe vengono suddivise in carte geografiche a grande o a piccola scala. Le carte geografiche a grande scala sono quelle rappresentanti uno spazio ristretto, ricco di dettagli, come ad esempio la carta di una città dove anche le singole costruzioni vengono fedelmente riportate. Le carte geografiche a bassa scala, invece, rappresentano una vasta regione, uno Stato, un continente o il globo. Gli elementi rappresentati in queste carte possono quindi differire sensibilmente rispetto alla realtà, col vincolo però che le interrelazioni tra gli elementi vengano correttamente rispettate. Le informazioni comunicate da una mappa a bassa scala sono dunque generali e approssimate, ma ciò risponde perfettamente allo scopo di queste carte, che è esattamente dare una visione d'insieme della realtà rappresentata.

La scala, dunque, non è solo una caratteristica che differenzia una carta dall'altra, ma determina anche la qualità, la quantità e il tipo di informazione che la mappa può trasmettere. Si può quindi definire un'ulteriore classificazione delle mappe basata sulla scala che metta in evidenza la relazione tra valore di scala adottato e informazioni trasmesse dalla mappa, e ricavare in base a questo criterio quattro gruppi principali di mappe:

- mappe a scala maggiore di 1:10 000, che vengono usate per le rappresentazioni catastali e per le carte tecniche dei centri abitati o delle aree rurali ed industriali, come la Carta Tecnica Regionale (CTR);
- mappe a scala compresa tra 1:10 000 e 1:100 000, carte topografiche rappresentanti con molta accuratezza zone relativamente limitate della superficie terrestre (rilievi, idrografia, orografia, edificato), come le carte dell'Istituto

Geografico Militare;

- mappe a scala compresa tra 1:100 000 e 1:1 000 000, che sono carte corografiche rappresentanti, con buona ricchezza di particolari, aree abbastanza estese della superficie terrestre come regioni o intere nazioni;
- mappe a scala minore di 1:1 000 000, ovvero carte geografiche che rappresentano regioni del globo anche molto estese come i continenti.

1.4 I GIS

Con l'avvento dei GIS la cartografia subisce un cambiamento epocale.

I GIS, infatti, permettono di manipolare oggetti spaziali e di estrarre, in vari modi, una grande quantità di informazioni dalle mappe: è così che la mappa si trasforma in una sorta di oggetto dinamico, mentre sino a poco prima era vista come un elemento statico. Il singolo dato geografico diventa un'entità dotata di attributi e inserita in una gerarchia relazionale; a questa entità corrisponde, oltre che una rappresentazione grafica, un contenuto logico e semantico.

Una delle definizioni di GIS che si trovano in letteratura è la seguente:

“A GIS is a powerful set of tools for collecting, storing, retrieving at will, transforming, and displaying spatial data from the real world”

(Burrough, 1986)

(Un GIS è una potente serie di strumenti per acquisire, memorizzare, estrarre a volontà, trasformare e visualizzare dati spaziali dal mondo reale)

In altre parole, un GIS permette di integrare operazioni che tipicamente offre un database con i vantaggi offerti dalla visualizzazione della mappa. Le entità geografiche, elementi base del GIS, vengono memorizzate all'interno di un database geografico (GeoDB), detto anche database spaziale, la cui funzione è quella di memorizzare, interrogare e manipolare informazioni geografiche e dati spaziali.

Più precisamente, un sistema GIS è costituito da risorse e procedure. Le risorse coinvolte sono risorse umane, infrastrutturali e di dati; le procedure, invece, si suddividono in acquisizione, archiviazione, elaborazione, presentazione e trasmissione.

I dati sono in assoluto l'elemento più prezioso del sistema, perché hanno un costo di produzione elevato. Essi, inoltre, per mantenere il loro valore, devono costantemente essere aggiornati. Ogni dato, che viene chiamato *feature*, è costituito da due componenti: la componente spaziale, che contiene informazioni relative alla posizione (geografica), alla geometria (forma e dimensione) e alla topologia (relazioni spaziali con altri dati) del dato; e la componente non spaziale, formata da dati descrittivi, o attributi, e metadati associati alla componente spaziale che velocizzano le operazioni di query spaziale.

L'acquisizione dei dati avviene normalmente scannerizzando le mappe cartacee tradizionali; un altro metodo, invece, si basa sull'utilizzo di file contenenti le coordinate degli elementi geografici, espresse secondo un determinato sistema di riferimento. In entrambi i casi, l'acquisizione è seguita da un processo di astrazione e generalizzazione.

Ovviamente, al momento dell'acquisizione è necessario memorizzare i dati. Il modo in cui le informazioni spaziali vengono memorizzate dipende da quale modello dello spazio viene adottato dal GIS. Ci sono due modelli utilizzabili: il modello raster e il modello vettoriale.

Nel primo caso, lo spazio viene visto come un mosaico, dove ogni tassello, o pixel, contiene uno o più valori (ad esempio, il colore o l'altitudine). Il pixel rappresenta il luogo dove una grandezza assume un certo valore. Un file raster risulta allora costituito da una matrice quadrata dove ogni cella è un pixel; le dimensioni di tale file tendono ad essere molto elevate, in quanto ogni unità di spazio è sempre descritta da un pixel.

Il secondo modello, quello vettoriale, è formato da una serie di primitive geometriche che vengono istanziate all'interno di uno spazio vuoto. In questo secondo caso, a differenza di quanto avviene nel modello raster, lo spazio non viene riempito completamente, ma funge da contenitore di oggetti. Le primitive vettoriali che costituiscono il modello sono:

- il punto, definito da una coppia (x,y) o da una terna di coordinate (x, y, z) ;
- la linea, costituita da una sequenza ordinata di punti;
- il poligono, anch'esso costituito da una sequenza ordinata di punti, dove però

l'ultimo punto della sequenza corrisponde al primo.

Il punto viene utilizzato quando l'elemento che si vuole rappresentare è semanticamente un punto, oppure quando, alla scala a cui si opera, esso è di fatto un punto; la medesima considerazione vale anche per la linea. Il poligono, invece, viene utilizzato per definire una particolare area della superficie geografica, ad esempio rappresentando laghi, edifici o confini amministrativi.

Comparando il modello raster e quello vettoriale, emerge che quest'ultimo si presta molto di più alla manipolazione, all'elaborazione e all'interrogazione dei dati, che si possono integrare facilmente in un database relazionale, dove vengono trattati come valori di una colonna di una tabella (solitamente chiamata "Geometry"), e processati per mezzo di molteplici operatori.

Ciò nonostante, il GIS viene progettato in modo che possa supportare entrambi i modelli; a seconda di quale utilizzo viene fatto della carta, poi, sarà adottato un modello piuttosto che l'altro. Ad esempio, le scansioni aeree o le immagini satellitari nascono e vengono utilizzate in formato raster; inoltre, in fase di stampa la mappa viene spesso convertita dal formato vettoriale a quello raster, per consentire stampe in grande quantità con procedure tipografiche.

Una volta che i dati sono stati acquisiti e opportunamente memorizzati secondo uno dei due modelli appena illustrati, è possibile effettuare su di essi operazioni di elaborazione e trasformazione degli elementi geografici tramite degli strumenti di analisi forniti dal GIS.

Alcune di queste operazioni sono³:

- *L'overlay topologico*, in cui si effettua una sovrapposizione tra gli elementi di due temi per creare un nuovo tematismo, ad esempio per sovrapporre il tema dei confini di un parco con i confini dei comuni in modo da determinare le superfici di competenza di ogni amministrazione o la percentuale di area comunale protetta;
- *Le query spaziali*, ovvero interrogazioni dei dati a partire da criteri spaziali

³ Fonte Wikipedia: http://it.wikipedia.org/wiki/Sistema_informativo_territoriale, acceduto il 16/01/2011.

(vicinanza, inclusione, sovrapposizione etc.) ;

- Il *buffering*, che permette di creare un poligono che circonda la geometria originaria;
- La *segmentazione*, che applicato su un elemento lineare, determina il punto alla distanza specificata dall'inizio dell'elemento;
- La *network analysis*, che applicata su una rete di elementi lineari, ad esempio la rete stradale, determina i percorsi minimi tra due punti;
- La *spatial analysis*, che effettua un'analisi spaziale di varia tipologia utilizzando un modello raster, come ad esempio l'analisi di visibilità;

1.5 Il processo cartografico

Il processo cartografico, ovvero il processo di realizzazione di una carta è il frutto di una sequenza ordinata di operazioni, che inizia con l'analisi e la definizione delle caratteristiche finali della mappa, passa attraverso la raccolta dei dati, e termina con l'effettiva produzione della mappa e il suo collaudo. Tale processo resta valido sia nel caso in cui la mappa sia realizzata a mano dal cartografo sia nel caso in cui si faccia uso di tecniche e strumenti informatici.

La prima fase è quella che coinvolge la progettazione della mappa, ovvero la definizione delle caratteristiche tecniche e dei contenuti. Le caratteristiche tecniche principali sono la scala, la superficie di riferimento, la proiezione e il tipo di rappresentazione (conforme, equivalente, equidistante). I contenuti della mappa, ovvero le informazioni che si vogliono collocare nella mappa, determinano il tipo di astrazione e rappresentazione che saranno utilizzate per trattare i dati in ingresso. In base alle decisioni prese a questo livello, la carta assumerà una natura tematica, cioè focalizzata solo su alcuni aspetti, piuttosto che una natura olistica, ovvero mirata a dare una visione d'insieme.

In questa prima fase si definisce, dunque, come debba avvenire l'astrazione della realtà, e in questo modo si produce anche, per le tradizionali mappe cartacee, la legenda della carta, e per le mappe digitali, la definizione di un GeoDB.

Terminata la fase di analisi e definizione, si passa all'acquisizione dei dati. I dati possono essere raccolti in due modi, a seconda che la mappa sia il frutto di una rilevazione o di una derivazione. Nel primo caso, gli elementi del territorio vengono rilevati durante un sopralluogo con l'aiuto di sistemi di posizionamento satellitare (GPS) e fotografie aeree. Nel secondo caso, se la mappa viene derivata, vengono utilizzati dati provenienti da una cartografia preesistente, la quale dovrà avere un livello di dettaglio maggiore (e quindi un'alta scala) rispetto alla carta che si deve realizzare (questo è necessario in quanto solamente se si è in possesso di tutti i dettagli è possibile astrarre e rappresentare correttamente la realtà). In entrambi i casi, gli operatori dovranno assegnare ad ogni elemento individuato un codice diverso a seconda della natura dell'elemento (strada, ferrovia, edificio, fiume etc.), in accordo col modello già stabilito nella fase di analisi e definizione.

Una volta che il cartografo entra in possesso di tutte le informazioni (dati e specifiche) necessarie, avvia il processo di generalizzazione, fondamentale per la costruzione della mappa. Durante questa fase, il cartografo deve prendere una serie di decisioni atte a soddisfare le specifiche definite, ma anche a garantire i requisiti classici di ogni mappa, quali la leggibilità e l'usabilità. In pratica egli dovrà creare un'astrazione della realtà geografica che faciliti la comprensione e la comunicazione dell'informazione⁴.

L'ultima fase del processo cartografico consiste nel collaudo della mappa, durante il quale si verifica la correttezza e la consistenza della carta. Tra le attività più importanti di collaudo vi è il controllo finale sul terreno mediante operazioni di misura e verifica della rappresentazione cartografica: vengono confrontati sul campo i dati riportati nella mappa con rilevazioni effettuate tramite strumenti ad alta precisione, come ad esempio il GPS differenziale.

4 Come ricorda Agent (2001) "Generalisation aims to provide an abstraction of geographic reality to enhance comprehension and communication of information" (la generalizzazione mira a fornire un'astrazione della realtà geografica per migliorare la comprensione e la comunicazione delle informazioni).

2 La generalizzazione

In questo capitolo verrà trattato in modo più specifico il tema della generalizzazione cartografica. Sarà definito il concetto di generalizzazione e saranno esposte le ragioni che inducono a realizzare il relativo processo di generalizzazione, che sarà successivamente descritto; sarà introdotto il processo di generalizzazione automatica e verranno esaminati alcuni modelli concettuali della generalizzazione proposti in letteratura; infine, si tratteranno gli operatori di generalizzazione, che saranno descritti e classificati in base alla loro funzione.

2.1 Introduzione al concetto di generalizzazione

La generalizzazione è, senza alcun dubbio, uno dei passi chiave nella realizzazione di una mappa.

Osservando una carta geografica, si può notare come gli elementi della realtà vengono classificati, ridefiniti e posizionati, sempre secondo un ordine logico e grafico. Se non fossero compiute queste azioni di classificazione e ridefinizione, l'informazione veicolata tramite la carta sarebbe sensibilmente inferiore. Ad esempio, se tentassimo di riportare in una mappa topologica di una città in scala 1:100 000 l'intero edificato, otterremmo nient'altro che una nuvola di punti, creando solamente disordine e confusione; se, invece, volessimo riportare nella stessa carta tutte le strade, non riusciremmo più ad individuare con facilità le strade principali, come quelle che portano da un lato della città all'altro.

Ne consegue che non ha senso tentare di rappresentare nella mappa tutta la realtà; al contrario, è necessario selezionare le informazioni rilevanti. Il risultato sarà dunque una astratta e più o meno limitata parte della realtà (Figura 2.1).

Lo scopo della generalizzazione, così, è proprio quello di astrarre la realtà per rappresentarla in modo che essa risulti chiara, comprensibile e leggibile, tenendo presente la funzione che deve svolgere la mappa, e in accordo con la scala di rappresentazione scelta.

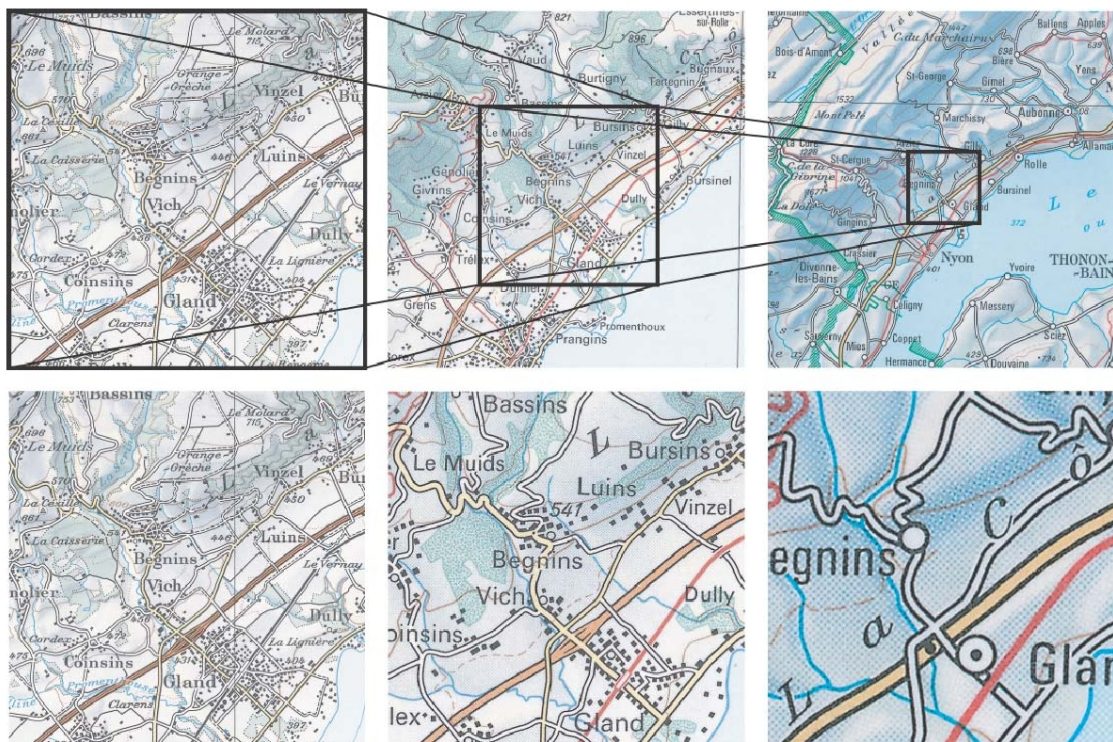


Figura 2.1: esempio di generalizzazione di una carta dalla scala 1:100.000 alla scala 1:200.000 e 1:500.000. In alto, le dimensioni delle carte sono riportate coerentemente con il valore della scala; in basso, le tre carte sono riportate secondo la stessa distanza grafica. Fonte: Cecconi (2003).

Tenendo a mente le osservazioni precedenti, la generalizzazione può essere allora definita come il processo che punta alla semplificazione delle informazioni geografiche per soddisfare i vincoli di rappresentazione e per raggiungere la finalità della carta, mettendo in evidenza gli elementi importanti e al tempo stesso rimuovendo quelli irrilevanti.

I fattori che devono essere tenuti in considerazione dal processo di generalizzazione sono molteplici; McMaster e Shea (1992) ne individuano sei:

1. Riduzione della complessità: il passaggio ad una scala inferiore porta naturalmente diversi oggetti ad entrare in conflitto per lo stesso spazio. Ridurre il numero di questi oggetti consente di far risaltare gli elementi più importanti e di mantenere la mappa leggibile.
2. Mantenimento dell'accuratezza spaziale: tanto più la scala è alta e tanto più l'accuratezza spaziale deve essere rispettata, limitando al minimo l'errore dovuto

alla diversa posizione degli oggetti nella mappa rispetto alla realtà.

3. Mantenimento dell'accuratezza degli attributi: in particolare nelle mappe tematiche, l'obiettivo è quello di minimizzare le alterazioni non intenzionali degli attributi delle *feature*.
4. Mantenimento della qualità estetica: dai colori utilizzati alla simbologia, dal bilanciamento allo stile tipografico, vari fattori influenzano l'estetica complessiva di una mappa; l'arte del cartografo sta nel mixare opportunamente questi fattori in modo da ottenere una mappa esteticamente bella.
5. Mantenimento di una logica gerarchica: elementi di una stessa categoria, ma di importanza o dimensioni differenti, devono essere differenziati, in accordo con lo scopo della mappa; ad esempio, una città grande deve risultare molto più prominente di una città piccola, mostrando una maggiore densità del grafo stradale e dell'edificato.
6. Coerente applicazione delle regole di generalizzazione: allo scopo di ottenere una generalizzazione imparziale e coerente, il cartografo deve determinare esattamente quali algoritmi applicare e in quale ordine e i parametri di input necessari per ottenere il risultato voluto quando si opera ad una data scala.

A seconda dei dati di partenza per la costruzione della mappa, si possono distinguere due processi di generalizzazione: se i dati disponibili sono il frutto di una rilevazione, cioè di un'acquisizione diretta, allora si parla di “compilazione cartografica”; se, invece, i dati sono quelli provenienti da una cartografia preesistente, si parla di “generalizzazione cartografica”.

Sia che ci si occupi di creare una mappa derivata oppure di una rilevata, il processo di generalizzazione dovrà affrontare le stesse problematiche relative al contenuto della mappa e alla sua rappresentazione; i due processi sono invece differenti per il fatto che, nella produzione di una mappa rilevata, bisogna creare una rappresentazione astratta partendo dalla realtà, mentre, nella produzione di una mappa derivata, è necessario operare questa astrazione partendo da un precedente modello.

In entrambi i casi, comunque, il risultato sarà un nuovo modello della realtà.

2.2 Il processo di generalizzazione

Il processo di generalizzazione può essere definito come “la selezione e la rappresentazione semplificata dei dettagli che meglio si adatta alla scala e o allo scopo di una mappa” (Meyen, 1973).

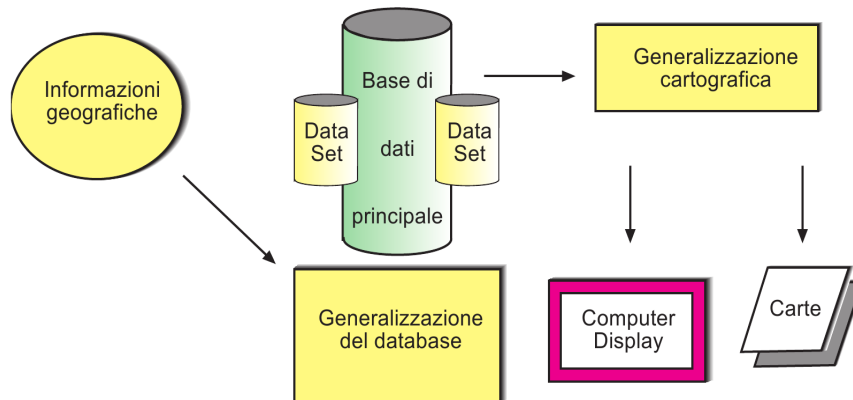


Figura 2.2: fasi della generalizzazione all'interno di un GIS: modellazione delle informazioni geografiche, che definisce il modello dati del GeoDBR, e generalizzazione delle feature. Fonte: Deruda (2005).

Due sono le componenti che caratterizzano questo processo (Figura 2.2): la generalizzazione del modello e la generalizzazione cartografica, che verranno esaminate in dettaglio nelle sezioni seguenti.

2.2.1 Generalizzazione del modello

Per mezzo della generalizzazione del modello, il risultato che si ottiene è la creazione di un modello astratto, rappresentativo della realtà; la scelta relativa a quali aspetti della realtà riportare e quali invece eliminare avviene, come già detto, in funzione degli scopi per cui viene realizzata la mappa. Ad esempio, in una carta politica verrà dato risalto ai confini amministrativi mentre elementi fisici come i rilievi o le depressioni non verranno riportati.

Questo processo porta alla definizione di una tassonomia degli elementi della realtà, cioè ad una loro suddivisione in classi. Anche i parametri che portano a distribuire gli oggetti in più classi variano a seconda delle funzioni della mappa che si vuole realizzare: così, una carta civile può classificare le strade in urbane ed extraurbane,

mentre una carta militare può enfatizzare maggiormente una distinzione delle vie di comunicazione basata sulla loro larghezza e tipo di fondo.

Nella cartografia digitale, la generalizzazione del modello corrisponde alla definizione di un GeoDatabase. In fase di definizione del database, è fondamentale ricordare che una modellazione e strutturazione efficiente della base di dati rappresenta un presupposto necessario per la buona riuscita della fase di generalizzazione cartografica, nonché per un proficuo utilizzo della cartografia digitale in ambiente GIS.

2.2.2 Generalizzazione cartografica

La seconda componente della generalizzazione è quella cartografica, che agisce sulle informazioni geografiche delle *feature*.

Una volta definito il modello astratto della realtà, si passa alla ridefinizione delle geometrie, allo scopo di fornire una coerente rappresentazione grafica al modello dei dati. Questa fase è di fondamentale importanza per garantire una mappa accurata e comprensibile. Il cartografo, per ridefinire le geometrie, si deve affidare alla sua interpretazione personale: non esistere infatti nessuno schema rigido che si possa applicare sistematicamente ad ogni singolo caso. A titolo d'esempio, un agglomerato di case potrebbe venire rappresentato come un unico blocco da un primo operatore e da due o più blocchi da un secondo operatore.

Tra i fattori che possono influenzare le scelte compiute dal cartografo troviamo in primis la scala, e successivamente il contesto. In relazione al valore di scala, un centro urbano potrebbe venire rappresentato con un punto affiancato dal nome della città oppure da un suo sottoinsieme di strade ed edifici. Il contesto invece determina il livello di importanza attribuito ad un oggetto, che può determinare la sua eliminazione o la precisione con cui lo si rappresenta. Ad esempio, un edificio posizionato in una zona densamente edificata potrebbe essere semplificato, o addirittura omesso, mentre una fattoria, posizionata in aperta campagna, potrebbe venire riportata accuratamente.

2.3 La generalizzazione automatica

La generalizzazione manuale, che un tempo era l'unica opzione praticabile, è un'operazione lunga ed onerosa; fortunatamente, le tecnologie informatiche hanno

permesso di automatizzare questa procedura, riducendo notevolmente fatica e tempo che l'operatore umano deve dedicarvi.

Per capire l'importanza che riveste l'automazione del processo di produzione delle mappe geografiche, può esser utile ricorrere ad un esempio pratico. In Francia, per la produzione di un foglio 91 x 121 cm della serie Topo100¹ con i metodi tradizionali, l'IGN² impiega mediamente 2000 ore di lavoro, di cui 1200 per la generalizzazione e 800 per il posizionamento della toponomastica. Se supponiamo che i fogli vengano realizzati in sequenza, ci vogliono circa 20 anni per realizzare le mappe relative all'intero territorio francese. Diminuendo il tempo di produzione, invece, si possono aggiornare le mappe con sempre maggior frequenza raggiungendo, naturalmente, anche un obiettivo di risparmio economico. Così, nel 2003, l'IGN ha utilizzato per la prima volta un software per la generalizzazione automatica (LAMPS2 della Laser-Scan), e ha ridotto i tempi medi per la produzione di un foglio a 150 ore per la generalizzazione e a 160 per il posizionamento dei nomi.

È evidente, dunque, il motivo per cui il campo della generalizzazione automatica desta sempre più interesse presso tutti gli enti cartografici del mondo, stimolando la produzione scientifica in questo innovativo settore.

La ricerca relativa all'automazione di questi processi si è focalizzata soprattutto sulla generalizzazione cartografica, in quanto sembra l'unica delle due fasi della generalizzazione per cui è possibile immaginare una forma di automazione; la generalizzazione del modello, infatti, presuppone una capacità di astrazione ed un livello di conoscenza difficilmente trasferibile ad un calcolatore.

La realizzazione di un processo di generalizzazione automatico si deve comunque scontrare con alcuni ostacoli (Mackaness, 2007). Un primo, comprensibile ostacolo è sicuramente la complessità del processo di progettazione, che non può prescindere dalla valutazione di molteplici fattori e vincoli: la mappa è un complesso mix di *pattern* metrici e topologici che solitamente sono molto interdipendenti. Interpretare correttamente queste forme e individuare le caratteristiche notevoli della mappa richiede una conoscenza sia cartografica che geografica. Un secondo ostacolo, inoltre, ha a che

1 Carte topografiche in scala 1:100 000.

2 Institut Géographique National, <http://www.ign.fr>.

fare con la trasformazione delle informazioni dovuta ad un cambiamento di scala: le mappe infatti, al variare della scala, mettono in luce caratteristiche geografiche diverse. Il processo di generalizzazione deve, perciò, essere in grado di estrapolare queste differenti caratteristiche in funzione della scala, a partire dalla stessa base di dati.

2.4 Modelli di generalizzazione

La generalizzazione, dagli anni '70 ad oggi, ha subito un'evoluzione che, a partire da un approccio basato su semplici algoritmi, l'ha portata ad essere definita tramite modelli matematici sempre più complessi.

Le prime attività di studio sul tema della generalizzazione, infatti, si focalizzavano sullo sviluppo di singoli algoritmi che fossero in grado di risolvere specifici problemi, come ad esempio la generalizzazione di una linea o del perimetro di un'area. Dal 1980 in poi, i ricercatori hanno cominciato a dare un'impostazione più organica ai loro studi, investigando su una possibile modellazione concettuale della generalizzazione.

Alla fine degli anni '70, Robinson (1978) sviluppò uno dei primi modelli concettuali, allo scopo di capire meglio il processo di generalizzazione. Egli suddivise il processo in due sotto-processi principali: la selezione, o preprocessing, e l'effettivo processo di generalizzazione. La prima fase coinvolge l'attività di identificazione degli elementi geografici da memorizzare nel database; la seconda interessa la manipolazione geometrica e statistica degli oggetti. Quest'ultimo sotto-processo viene suddiviso ulteriormente in tre fasi: semplificazione, ovvero eliminazione dei dettagli geografici non necessari; classificazione, cioè la categorizzazione degli oggetti; e simbolizzazione, o associazione di una codifica grafica ad ogni elemento.

Pochi anni dopo, Bertin (1983) parlò di generalizzazione concettuale e strutturale. Nella generalizzazione concettuale è possibile cambiare la natura di un elemento, definendo un nuovo livello di concettualizzazione. Ad esempio, nella generalizzazione concettuale si può decidere di trasformare un cluster di punti in un'area. Nella generalizzazione strutturale la natura del fenomeno rimane inalterata, mentre si agisce per semplificare la sua distribuzione; ciò avviene basandosi solo sulle informazioni già presenti in fase di elaborazione. Ad esempio, il cluster di punti, invece di cambiare natura, viene

generalizzato riducendo il suo numero di punti.

Grande importanza riveste anche Il lavoro di Nickerson e Freeman (1986), che schematizzarono la generalizzazione usando un modello suddiviso in cinque compiti: modifica delle *feature* per mezzo di quattro operatori (selezione, semplificazione, combinazione e conversione di tipo); ridimensionamento dei simboli; displacement delle *feature*; riduzione della scala; e posizionamento dei nomi. In questo modello, la generalizzazione della mappa alla scala finale passa per una mappa a scala intermedia, nella quale i simboli vengono prima scalati di un fattore k , per poi essere riposizionati tramite l'operazione di displacement.

Un grosso contributo deriva dal lavoro di McMaster e Shea (1992), che proposero un modello concettuale globale del processo di generalizzazione digitale. Essi suddivisero il processo in tre questioni (Figura 2.3): perché generalizzare (*why*), quando (*when*) e come (*how*). Per rispondere alla domanda *why* vengono individuati degli obiettivi filosofici (elementi teorici, elementi computazionali ed elementi specifici); in merito al *when* si specifica una valutazione cartometrica delle condizioni che motivano la generalizzazione (condizioni geometriche, misure spaziali e olistiche, controllo delle trasformazioni); alla domanda *how* viene stilato un elenco degli operatori della generalizzazione, suddivisi in relazione all'aspetto dell'oggetto che trasformano (spaziale o degli attributi).

Lamy e Ruas (1999) hanno proposto recentemente un modello di generalizzazione innovativo, basato su un sistema ad agenti, implementato poi all'interno di un progetto europeo denominato AGENT³ (Automated GEneralization New Tecnology) (AGENT, 2003). Secondo gli autori, il vantaggio principale offerto da un sistema ad agenti è quello di consentire la manipolazione degli oggetti di una mappa a diversi livelli di dettaglio. Ad ogni agente viene affidato un oggetto della mappa (micro-agent), un gruppo di oggetti (meso-agent) o l'intera mappa (macro-agent) con il compito di soddisfare i vincoli di generalizzazione imposti, utilizzando le proprie funzionalità. Gli agenti interagiscono fino a quando non viene trovata una soluzione capace di soddisfare tutti i vincoli, o se questo non dovesse essere possibile, una soluzione che soddisfi la maggior parte di essi.

3 <http://agent.ign.fr/> .

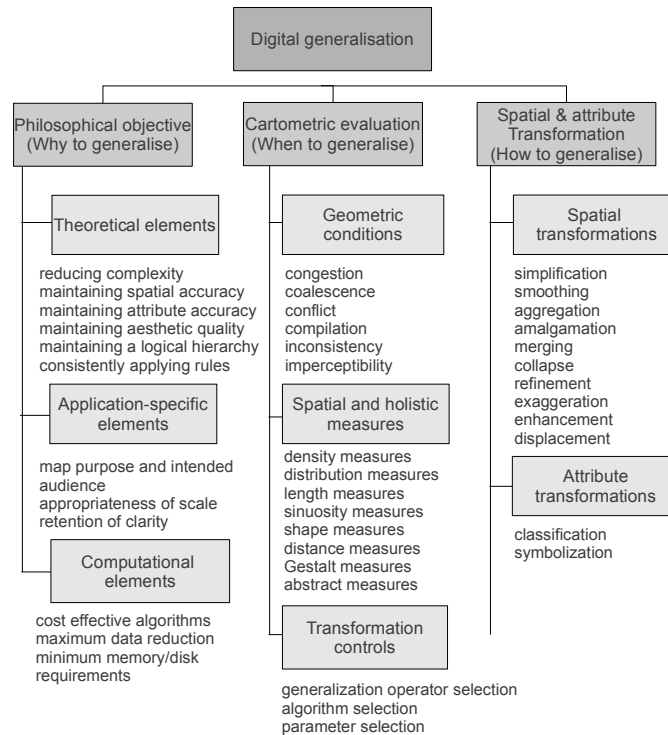


Figura 2.3: Modello concettuale sviluppato da McMaster e Shea per la generalizzazione concettuale. Fonte: McMaster e Shea (1992)

La letteratura scientifica recente si è occupata anche di studiare i vincoli che si presentano nel passaggio da una scala grande ad una piccola; di questi vincoli e degli operatori di generalizzazione, strumenti con i quali soddisfare i vincoli, si discuterà nel paragrafo seguente.

2.5 Gli operatori della generalizzazione

Quando si deve generalizzare una carta ci sono numerosi vincoli da soddisfare; il progetto, infatti, ha delle specifiche che devono essere rispettate nella soluzione del problema di generalizzazione. Un vincolo è una specifica di progetto che deve essere rispettata nella soluzione ad un problema di generalizzazione. Tra i più importanti vincoli si possono citare i seguenti:

- vincoli di tipo grafico, legati a parametri come la dimensione, la larghezza e la

distanza;

- vincoli topologici, legati a connessione, adiacenza e inclusione tra elementi;
- vincoli spaziali, legati alla conservazione delle forme e degli allineamenti;
- vincoli semantici, legati alle relazioni logiche che intercorrono tra gli oggetti e i loro insiemi;
- vincoli di disegno, legati all'aspetto grafico della carta.

Allo scopo di soddisfare questi vincoli, una grande varietà di soluzioni sono stati studiate nel corso degli anni; queste soluzioni si chiamano operatori di generalizzazione.

Un operatore di generalizzazione rappresenta un tipo di trasformazione spaziale che si vuole ottenere (Weibel and Dutton, 1999). Gli operatori sono stati sviluppati sia emulando le pratiche manuali dei cartografi, sia tramite studi puramente matematici. Ogni operatore, comunque, prende vita mediante un algoritmo di generalizzazione che lo implementa.

Diverse classificazioni sono state proposte per dare un ordine logico a questi operatori, in modo da facilitare i cartografi nella scelta dell'operatore che più si presta a soddisfare un particolare vincolo. Alcuni ricercatori suddividono gli operatori in base al tipo di geometria a cui fanno riferimento (ad esempio l'operatore di semplificazione è progettato per elementi lineari, mentre l'operatore di fusione opera su oggetti areali).

Un'importante classificazione è quella proposta da McMaster e Shea (1992), che distinguono gli operatori in operatori che trasformano le geometrie e operatori che trasformano gli attributi. Queste due classi saranno descritte in dettaglio nei seguenti due sottoparagrafi.

2.5.1 Operatori geometrici

Gli operatori geometrici sono operatori che agiscono sull'aspetto grafico e topologico di uno o più oggetti geografici.

Una possibile classificazione degli operatori geometrici è stata prodotta all'interno del progetto AGENT; questi sono stati classificati in base al numero di oggetti su cui agiscono. Nella prima classe troviamo quelli che operano su un singolo elemento, e

sono:

- *simplification operator*: riduce la granularità dei contorni di linee e aree, in pratica producendo una versione semplificata dell'oggetto grazie all'eliminazione della ridondanza di punti;
- *collapse operator*: trasforma oggetti areali in punti o linee risolvendo la progressiva mancanza di spazio.
- *Enhancement operator*: valorizza un oggetto in vari modi, ovvero ingrandendo la sua forma completa (*enlargement*) o una sua parte (*exaggeration*), addolcendo il suo contorno per migliorare l'estetica (*smoothing*), squadrando la sua geometria se si avvicina ad una forma rettangolare (*squaring*);

Nella seconda classe vi sono gli operatori che possono essere applicati indistintamente ad uno o più oggetti:

- *selection operator*: seleziona gli elementi importanti considerando lo scopo della mappa;
- *elimination operator*: elimina dalla mappa gli oggetti ritenuti non importanti o ridondanti;
- *displacement operator*: sposta di posizione un oggetto o un gruppo mantenendone inalterata la forma.

Infine, nella terza categoria ricadono gli operatori che modificano un insieme di oggetti, e che vengono definiti operatori di aggregazione:

- *combine operator*: unisce in un unico oggetto punti che precedentemente erano separati e distinti, è identificato come un operatore 0-dimensionale, in quanto agisce su geometrie zero-dimensionali;
- *merging operator*: fonde 2 o più linee in un'unica linea, che normalmente viene posizionata a mezz'ora; è un operatore 1-dimensionale, in quanto agisce su geometrie lineari;
- *Amalgamation operator*: fonde in un'unica geometria areale un gruppo di poligoni; è un operatore 2-dimensionale, in quanto agisce su geometrie areali;

- *typification operator*: riduce la complessità di un gruppo di oggetti attraverso la loro eliminazione, riposizionamento, allargamento o aggregazione mantenendo la disposizione tipica di quell'insieme di oggetti.

2.5.2 Operatori per la trasformazione degli attributi

Gli operatori di trasformazione degli attributi si differenziano da quelli geometrici in quanto non vanno a manipolare l'informazione geometrica dell'oggetto, ma agiscono sulla sua componente statistica. Ad esempio, una piantagione di latifoglie e conifere potrebbe essere trasformata in una foresta, a causa di una riduzione di scala che mette in secondo piano le caratteristiche degli alberi.

Gli operatori di questo tipo, identificati sempre da McMaster e Shea (1992), sono:

- *classification operator*: riduce i dati grezzi in un insieme di classi, operazione spesso necessaria in quanto non è sempre praticabile associare ad ogni singolo fenomeno un simbolo diverso per successivamente mapparlo;
- *symbolization operator*: codifica graficamente i dati e può essere applicato sia alla componente statistica sia a quella geografica.

2.5.3 Operatori e algoritmi

Gli operatori appena esaminati descrivono, di fatto, le singole operazioni che il cartografo compie per generalizzare la mappa. Ad ogni operatore, poi, corrisponde un algoritmo che implementa una trasformazione del dato geografico, anche se è facile imbattersi in più implementazioni dello stesso operatore, a seconda dell'oggetto su cui agisce. Ad esempio, l'operatore di semplificazione viene realizzato in modo diverso quando si applica ad una strada rispetto al caso in cui si applichi ad un edificio.

Ogni algoritmo è caratterizzato da una serie di parametri, che vanno impostati in funzione della scala, dell'oggetto a cui fanno riferimento, e in generale, del contesto. Lo stesso algoritmo di semplificazione, così, potrebbe essere applicato con parametri differenti a seconda che l'oggetto si trovi all'interno di un centro abitato, dove troppi dettagli tendono a creare confusione, o in una zona scarsamente abitata, dove non ci sono questo tipo di complicazioni. Il primo problema da affrontare è, quindi, la scelta dei parametri degli algoritmi e l'opportunità di variarli in funzione del contesto.

Il secondo problema, non meno importante, è l'ordine con cui gli algoritmi vengono eseguiti: il cartografo, per decidere la sequenza delle azioni da compiere e la loro coerenza con lo scopo della mappa o con il contesto, utilizza la sua conoscenza cartografica e la sua capacità di vedere sotto diverse astrazioni la realtà rappresentata. Gli algoritmi che implementano gli operatori di generalizzazione sono, infatti, potenti strumenti per risolvere i vincoli spaziali, semantici e grafici; essi, però, non vanno utilizzati indiscriminatamente sull'intera area da generalizzare o sempre secondo la stessa sequenza statica, ma vanno scelti, settati e applicati ogni volta diversamente, in funzione del contesto, della scala e dello scopo della mappa.

3 Il progetto di ricerca CARGEN

Il progetto CARGEN, CARtographic GENeralization, è un progetto di ricerca, partito nel 2006, che vede coinvolti il Dipartimento di Ingegneria Informatica dell'Università di Padova, la Regione Veneto, e l'Istituto Geografico Militare. L'obiettivo di questo progetto è arrivare ad un processo di generalizzazione cartografica automatizzato applicabile alla produzione di mappe derivate. I dati di ingresso del processo sono quelli della cartografia tecnica regionale (CTR), alla scala 1:5000, mentre la cartografia che si intende produrre è quella coerente con i modelli DB25 e DB50 dell'IGM.

In questo capitolo si descriverà la situazione cartografica attuale italiana, e si discuteranno gli obiettivi del progetto CARGEN e i relativi risultati raggiunti dopo i primi tre anni di progetto. Infine sarà presentato l'ambiente di lavoro, includendo le cartografie e i software utilizzati in questa tesi .

3.1 La situazione cartografica in Italia

In Italia, l'ente che si occupa di mantenere una cartografia aggiornata a livello nazionale è l'Istituto Geografico militare (IGM). La prima produzione cartografica dell'istituto è stata la *Nuova Carta Topografica d'Italia*, realizzata alla scala 1:100 000. L'IGM ha prodotto diverse carte che coprono il territorio a livello nazionale e solo una parte di queste vengono mantenute aggiornate,: per la scala 1:25 000 esistono le serie 25V, 25 e 25DB, per la scala 1:50 000 le serie 50 e 50/L e per la scala 1:100 000 le serie 100V e 100L. Tra le varie serie sussiste un preciso rapporto matematico: ogni foglio di una mappa in scala 1:100 000 è divisa in quattro settori, rappresentati in altrettante mappe della serie 1:50 000; queste mappe vengono dette “quadranti”, e ogni quadrante è a sua volta diviso in quattro parti, le “tavole”, che sono in scala 1:25 000.

Delle serie cartografiche alla scala 1:25 000, solo la più antica, la 25V (vecchio taglio), è stata completata. I dati di questa serie sono aggiornati mediamente al 1960, ad eccezione di alcuni elementi, che sono stati prodotti durante una campagna di aggiornamento parziale del 1984.

La serie 25 copre, dunque, solo il 36% del territorio nazionale; la sua produzione è stata

interrotta per fare spazio alla nuova serie 25DB. Questa serie introduce la cartografia digitale, e, su un totale di 2298 sezioni, ne sono state realizzate 68, ottenute tramite stereorestituzione¹ numerica o come derivazione dalla cartografia tecnica regionale numerica.

Oltre all'IGM, che lavora su cartografie con scala inferiore, da 1:25 000 in giù, vi sono altri enti che si occupano di cartografia e ai quali vengono affidate le realizzazioni di carte a media e grande scala (1:10 000 e superiore). Inizialmente era il Catasto a dover adempiere a questo compito²; successivamente tale funzione è stata trasferita alle Regioni³ che hanno potuto così gestire in modo autonomo la creazione delle carte regionali, definite tecniche in quanto create specificamente per i tecnici delle amministrazioni.

Le carte prodotte dalle Regioni sono carte ricche di particolari e vengono aggiornate frequentemente, operazione facilitata dalla ristretta porzione del territorio nazionale che viene rappresentata. La CTR costituisce la base di riferimento per la redazione degli strumenti urbanistici comunali, per i Piani di Coordinamento Provinciali, per i Piani d'Area e per i vari piani di settore della pianificazione e della programmazione regionale.

In Veneto vengono prodotte due serie di Carte Tecniche Regionali (dette CTR): una in scala 1:10 000, composta da sezioni, e una in scala 1:5 000, composta da elementi. Il taglio e l'inquadramento le rendono sovrapponibili alle carte IGM di nuova produzione⁴: in particolare, ogni foglio IGM in scala 1:50 000 è diviso in 4 quadranti della serie 25, ognuno dei quali ripartito in 4 sezioni in scala 1:10 000 della CTR, divisi a loro volta in 4 elementi in scala 1:5.000. Per ogni elemento, le dimensioni del campo cartografico sono di 2'30" in longitudine e 1'30" in latitudine, che corrispondenti a circa un'area di 3200 per 2800 metri alle latitudini medie della Regione.

Recentemente anche gli uffici cartografici regionali hanno intrapreso la strada verso la digitalizzazione cartografica, iniziando la produzione di carte tecniche regionali

1 La stereorestituzione è un'operazione che, a partire da una coppia di fotogrammi che riprendono con angolazioni diverse la medesima zona, ne ricostruisce lo spazio virtuale.

2 Legge n° 68/1960, "Norme sulla cartografia ufficiale dello Stato e sulla disciplina della produzione e dei rilevamenti terrestri e idrografici".

3 Decreto Presidente della Repubblica 24 luglio 1977, n.616.

4 Eccezion fatta, quindi, per la serie V.

numeriche (CTRN). Questo fatto, in concomitanza con il passaggio, da parte dell'IGM, dalla serie 25 alla 25DB, ha aperto un nuovo scenario nella cartografia italiana: la possibilità di usare la derivazione come mezzo per produrre la nuova serie 25DB. Questa situazione permetterebbe quindi di limitare le rilevazioni dei dati sul territorio, e accelerare i tempi di produzione delle sezioni della serie 25DB. La realizzazione di una procedura informatica per la generalizzazione cartografica permetterebbe di accelerare ulteriormente i tempi, e darebbe quindi la possibilità di automatizzare il processo di derivazione.

Il progetto CARGEN nasce proprio per questo scopo, ovvero sviluppare un processo di generalizzazione delle carte tecniche regionali per produrre una base di dati coerente con il modello 25DB dell'IGM.

3.2 Il progetto CARGEN

Il progetto CARGEN, nato da una convenzione tra la Regione Veneto ed il Dipartimento di Ingegneria dell'Informazione dell'Università di Padova, ha come obiettivo quello di individuare le metodologie informatiche di derivazione dalla cartografia tecnica numerica del DB5k e successivamente del DB25 IGM, e sperimentare le stesse con la realizzazione di un prototipo, consistente nel DB25 IGM del Parco delle Dolomiti Bellunesi. L'obiettivo del progetto è stato in seguito esteso anche alla derivazione del DB50.

La cartografia di partenza, da cui derivare i dati per il DB25, è quella della CTRN, fornita dalla Regione Veneto. Questa cartografia, viene utilizzata per popolare il GeoDBR, dal quale si derivano successivamente i dati per popolare il DB25.

Nei successivi sottoparagrafi saranno descritti i tre modelli dati coinvolti nel progetto CARGEN, ovvero il modello CTRN, GeoDBR e DB25, e saranno presentati i risultati raggiunti all'interno del progetto.

3.2.1 CTRN

La Carta Tecnica Regionale Numerica è una cartografia generale e metrica, in formato vettoriale, prodotta dalla Regione Veneto. La carta, la cui produzione trova la sua principale fonte di dati nel rilievo fotogrammetrico, gode di campagne d'aggiornamento

piuttosto frequenti ed offre quindi un dato geografico piuttosto recente e di buona qualità.

Le scale di rappresentazione adottate sono la scala 1:5 000 per la quasi totalità del territorio regionale e la scala 1:10 000 per le zone montane scarsamente urbanizzate. Gli oggetti e le informazioni territoriali contenute nella Carta Tecnica Regionale, acquisiti in forma vettoriale, sono organizzati in Livelli e Codici: i Livelli costituiscono una primaria classe di aggregazione degli oggetti, che a loro volta sono suddivisi nei Codici, relativi alle caratteristiche particolari di ciascun oggetto. In totale sono presenti 16 Livelli principali, 12 Livelli di servizio e 6 Livelli funzionali per la gestione informatica dei grafi (assi e nodi di viabilità, idrografia e ferrovia); ciò consente la codifica di 480 oggetti ed informazioni.

I dati della CTRN, però, non si prestano bene all'analisi spaziale e ad un diretto utilizzo, in quanto sono realizzati prevalentemente tramite tecniche CAD, e perciò non offrono alcuna forma di controllo di coerenza topologica. Questo fatto si ripercuote nella necessità di attuare una lunga fase di controllo e pulizia dei dati.

3.2.2 GeoDBR

Il GeoDBR è un modello dati sviluppato dalla Regione Veneto nell'ambito di un progetto per l'aggiornamento del proprio sistema informativo territoriale, realizzato secondo le specifiche definite all'interno del progetto IntesaGis⁵.

Aderendo alle specifiche IntesaGis, il GeoDBR si presta come perfetto tramite tra i dati della CTRN e quelli del DB25. Le specifiche di IntesaGis sono infatti mirate proprio alla definizione di un modello di dati che permetta una facile derivazione del DB25.

Il GeoDBR è un modello dati di moderna concezione, caratterizzato da una rappresentazione che prevede una divisione dell'informazione geografica in informazione di base e strati tematici, e la gestione di geometrie tridimensionali. Gli oggetti topografici sono rappresentati nel GeoDBR da *feature* inserite in un'organizzazione di livelli informativi a strati. Risulta così presente una forte componente gerarchica nella strutturazione del modello dati: le *feature*, prima che negli strati informativi, sono raggruppate in *feature class* e distinte tra loro tramite attributi.

⁵ IntesaGis: Intesa Stato-Regioni-Enti Locali sui Sistemi Informativi Territoriali. Per informazioni relative al progetto, si visiti <http://www.centrointerregionale-gis.it/>.

Questo tipo di organizzazione è in contrapposizione con il modello dati del DB25 e della CTRN, dove gli attributi vengono usati più per fini descrittivi che per fini tassonomici.

3.2.3 DB25

Il DB25 è il modello dati creato dall'IGM per la compilazione della cartografia della serie topografica 25DB. Il DB25 contiene un modello di rappresentazione del mondo reale basato sulle *feature*, suddivise in oggetti semplici identificati dall'attributo LAB (Label).

Ogni *feature* è caratterizzata da una tipologia di primitiva grafica, evidenziata dalla prima lettera del codice: C per la tipologia areale, L per quella lineare, P per quella puntuale e T per le *feature* testuali.

Le *feature* che popolano il DB25 si riferiscono a 291 particolari topografici, suscettibili di restituzione, di eventuale ricognizione e/o di acquisizione da Banche Dati di Enti Pubblici e Privati, e a 48 tipologie di testi, per la maggior parte legati a particolari topografici. Le *feature* in cui gli oggetti sono inquadrati derivano la loro codifica e definizione dagli oggetti descritti nello standard DIGEST, Digital Geographic Information Exchange Standard, usato in applicazioni militari, in ambito NATO, e sviluppato dal gruppo di lavoro internazionale DGIWG (Digital Geographic Information Working Group): ogni *feature* del DB25 è definita da un codice LAB, un nome, e un codice FACC (Feature and Attribute Coding Catalogue).

A differenza del GeoDBR, l'organizzazione gerarchica delle *feature* del DB25 è minima: ogni oggetto topografico è di norma descritto direttamente da un codice LAB univoco. Questa scelta si spiega nella finalità tipografica di questo modello dati: ad ogni codice LAB è infatti associata anche una vestizione grafica, che viene usata nella stampa delle carte e riportata nella legenda delle cartografie in serie 25DB. Conseguenza di questa scarsa organizzazione gerarchica è la presenza, nel DB25, di più *feature* che afferiscono allo stesso oggetto reale; la scelta di quale tra queste *feature* utilizzare per l'oggetto si basa generalmente sui limiti di acquisizione: un esempio è l'oggetto edificio industriale, che, nonostante sia descritto dal medesimo codice FACC AC000 Processing Plant/Treatment Plant, viene rappresentato dalla *feature* Opificio Generico, con codice

LAB C406 se la sua superficie è inferiore a 1500 metri quadri, oppure con la *feature* Stabilimento Industriale, LAB C405A, se la sua area risulta superiore a tale dimensione.

3.2.4 La derivazione del DB25

La parte del progetto CARGEN dedicata alla derivazione del DB25, condotta nei primi anni del progetto, è stata caratterizzata dalle seguenti tre fasi:

- creazione del GeoDBR e suo popolamento con i dati provenienti dalla CTRN in scala 1:5.000;
- analisi delle differenze tra i modelli di dati proprietari del GeoDBR e del DB25;
- la creazione del prototipo DB25.

La prima fase consisteva nel completamento delle specifiche del GeoDBR, nella creazione delle tabelle all'interno del DBMS Oracle Spatial 10g e nel loro popolamento. La migrazione dei dati dal modello della CTRN a quello del GeoDBR si è scontrata in primo luogo con la diversa rappresentazione delle geometrie usate nei due modelli, e in secondo luogo con gli errori topologici dei dati originali. All'interno del progetto sono stati perciò sviluppati algoritmi per la trasformazione delle geometrie e per il controllo topologico, in modo da migliorare la qualità dei dati facilitando la successiva fase di derivazione del DB25.

La seconda fase, che aveva come obiettivo quello di trovare una corrispondenza tra le *feature* del GeoDBR e quelle del DB25, ha permesso di evidenziare alcune discrepanze e di conseguenza suggerire alcune migliorie ai due modelli.

La terza fase, consistente nella creazione del prototipo DB25, ha visto l'implementazione di un processo di generalizzazione cartografica, allo scopo di soddisfare i vincoli dettati dal passaggio di scala. Il processo di derivazione (Savino, 2007) è stato implementato in otto passi, utilizzando il DBMS Oracle per il mantenimento dei dati, Geomedia Professional 6 per l'elaborazione e l'export dei dati, e Dynamo/Dynagen per la fase di ricostruzione, acquisizione e generalizzazione dei dati.

L'analisi e la verifica dei risultati di quest'ultima fase hanno mostrato una carenza negli strumenti messi a disposizione dai software Geomedia, ed in particolare,

Dynamo/Dynagen, che non è risultato sufficiente per attuare una generalizzazione completa e soddisfacente degli elementi geografici.

Questa analisi ha messo in luce una naturale distinzione degli elementi geografici in due gruppi: il primo, costituito da tutte le classi per cui la generalizzazione può essere ottenuta tramite le sole operazioni di selezione geometrica e/o spaziale descritte nel modello dati; il secondo, costituito da classi complesse, come la viabilità stradale, ferroviaria, l'idrografia e l'edificato, che richiedono una generalizzazione più specifica e accurata.

Tramite la procedura di popolamento si possono quindi generalizzare gran parte delle classi del modello dati, mentre, per le restanti, è richiesto lo sviluppo di algoritmi appositi per rendere queste classi complesse adatte alla rappresentazione in scala 1: 25 000 e compatibili con l'applicazione delle specifiche del modello dati. Questo fatto ha reso necessario lo sviluppo di un certo numero di algoritmi parametrici, scritti in Java, per risolvere alcuni importanti problemi di generalizzazione. In particolare, la viabilità stradale è stata semplificata mediante riconoscimento e generalizzazione degli incroci, identificazione e accorpamento delle carreggiate autostradali, e sfoltimento della viabilità secondaria; la rete fluviale tramite sfoltimento e semplificazione; l'edificato tramite amalgamento, selezione e tipificazione; infine, i binari all'interno delle stazioni ferroviarie sono stati generalizzati riducendo il loro numero.

Questi algoritmi sono stati inseriti in un processo di tipo batch, ovvero un processo che gli esegue in ordine sequenziale e predefinito. Il risultato finale del processo dipende perciò solo esclusivamente dai parametri scelti per i singoli algoritmi.

I risultati della generalizzazione del prototipo DB25 sono stati presentati in un convegno nazionale tenutosi presso l'Università di Padova nel 2009 (De Gennaro, 2009). A partire da questa data, gli obiettivi del progetto sono stati estesi alla derivazione del DB50.

3.2.5 La derivazione del DB50

La seconda fase del progetto CARGEN, quella relativa alla derivazione del DB50, ha avuto come primo obiettivo quello di realizzare un prototipo di modello dati per il DB50; per quest'ultimo, infatti, non è mai stato definito un modello dall'IGM. Nel

febbraio 2010 è stato redatto un documento⁶ contenente l'analisi della derivabilità della Carta d'Italia IGM in scala 1:50 000, a partire dal database topografico IGM DB25 in scala 1:25 000, fornendo una proposta di modello dati per un database topografico alla scala 1:50 000.

Nel frattempo sono stati avviati, all'interno del progetto CARGEN, degli studi sulla generalizzazione cartografica per la scala 1:50.000, riguardanti la generalizzazione del grafo stradale, oggetto di questa tesi, e dell'edificato (Baldin, 2010).

3.3 Ambiente di lavoro

Il modello adottato all'interno del progetto CARGEN per la gestione dei dati è stato quello client-server.

Il server è costituito da una macchina con installato un DBMS Oracle Spatial 10g, la cui funzione è quella di memorizzare e mantenere i dati spaziali, accessibili tramite query. Nel lato client, invece, sono stati installati i software Geomedia Professional 6 e Dynamo/Dynagen, entrambi di proprietà della Intergraph. Geomedia viene utilizzato principalmente come strumento d'accesso ai dati spaziali; Dynamo/Dynagen sono invece gli strumenti usati durante il processo di generalizzazione cartografica.

Questa architettura, tuttavia, è stata in parte abbandonata recentemente: infatti, sia allo scopo di semplificare lo sviluppo di nuovi algoritmi, sia allo scopo di migliorare le prestazioni temporali, è stato cambiato metodo d'accesso ai dati ed il software per visualizzarli. I dati vengono caricati in RAM e gestiti proprio come se fossero delle tabelle, grazie ad una libreria sviluppata all'interno del Progetto. Inoltre, una libreria potentissima sviluppata in Java, la JTS (JTS Topology Suite), fornisce una vasta serie di operatori spaziali, evitando così di ricorrere al DBMS di Oracle per effettuare le interrogazioni spaziali⁷.

Nei successivi sottoparagrafi saranno brevemente descritti i software più importanti usati durante l'attività di tesi, e i dati geografici su cui sono stati effettuati i relativi test.

⁶ Analisi preliminare e proposta del modello dati DB50, Laboratorio GIRTS, Università degli Studi di Padova (febbraio 2010).

⁷ Il DBMS rimane comunque importante per mantenere il sistema scalabile, in quanto, non venendo le tabelle caricate in ram, è possibile processare zone geografiche più vaste e coinvolgere più tabelle contemporaneamente.

3.3.1 JTS Topology Suite

La JTS Topology Suite⁸ è una libreria open source, scritta interamente in Java, che fornisce una modellazione ad oggetti per le geometrie lineari in uno spazio euclideo. In questa libreria sono definite tre geometrie fondamentali, *Point*, *LineString* e *Polygon*, che rappresentano rispettivamente la geometria puntuale, lineare e areale.

La JTS mette a disposizione numerose funzioni geometriche, tra le quali possiamo citare:

- gli operatori topologici, che realizzano le funzioni di intersezione, differenza, unione;
- la funzione per la creazione del buffer intorno alla geometria;
- la funzione per la costruzione dell'involucro convesso;
- alcune funzioni per la semplificazione delle geometrie, come l'algoritmo di Douglas-Peucker;
- la funzione per la costruzione del Minimum Bounding Box.

Oltre a queste funzioni, la JTS fornisce l'implementazione di indici spaziali, come il quadtree⁹, che offrono un modo veloce per la risoluzione di query spaziali.

La versione utilizzata nell'ambito di questa tesi è la JTS 1.11.

3.3.2 OpenJump

OpenJump¹⁰ è un Desktop GIS open source che permette di visualizzare, modificare e interrogare dati spaziali.

OpenJump è scritto in Java, si basa sulla JTS ed è in grado di gestire file raster, vettoriali e database (PostGis, Oracle, ArcSDE); una caratteristica degna di nota è la sua architettura modulare, che permette di estendere di molto le funzionalità di base, potendo integrare, per esempio, il proprio codice mediante la realizzazione di un plugin.

In OpenJump, la creazione di un plugin, relativo al proprio codice, offre al

⁸ <http://tsusiatsoftware.net> .

⁹ Un quadtree è una struttura dati ad albero nella quale tutti i nodi interni hanno esattamente quattro nodi figli. Esso viene utilizzato per partizionare lo spazio bidimensionale, suddividendolo ricorsivamente in quattro quadranti.

¹⁰ <http://www.openjump.org> .

programmatore il grosso vantaggio di poter visionare tramite l'interfaccia grafica gli effetti della propria applicazione. Il plugin diventa così uno strumento essenziale nello sviluppo di nuovi algoritmi che manipolano geometrie: avere una risposta grafica e istantanea è un grande aiuto per semplificare e velocizzare la fase di testing e debug.

La modalità con cui OpenJump gestisce le *feature* si basa sull'utilizzo dei *layer*, o livelli, che svolgono il ruolo di contenitori di *feature*. Ogni *layer* è in grado di contenere le *feature* relative ad uno specifico schema dati, chiamato *FeatureSchema*; quest'ultimo specifica il nome e la tipologia degli attributi che costituiscono la *feature*, simile a quanto accade nelle tabelle dei database. Un *layer* rappresenta, quindi, una vera e propria tabella, il cui schema dati è specificato dal *FeatureSchema*. Il *layer* è interrogabile per mezzo di query, che possono essere sia spaziali che non spaziali. Per migliorare le query spaziali, è possibile associare al *layer* uno degli indici spaziali forniti dalla JTS.

La versione di OpenJump utilizzata nello svolgimento di questa tesi è la 1.3.1, rilasciata nell'aprile 2009.

3.3.3 Cartografia per il test

Il progetto CARGEN ha come scopo dichiarato quello di realizzare il prototipo del DB25 IGM del Parco delle Dolomiti Bellunesi. La cartografia di questa zona è stata pertanto utilizzata nella fase di test degli algoritmi sviluppati per la generalizzazione della rete stradale. Tuttavia, sono state prese in considerazione due ulteriori cartografie per due motivi: il primo è l'impossibilità di confrontare i risultati generalizzati con quelli della carta, dato che la relativa carta geografica dell'IGM è datata, mentre i dati digitali in nostro possesso sono recenti; in secondo luogo, essendo questa una zona montuosa, il grafo stradale non presenta situazioni particolarmente complesse e di conseguenza non è facile realizzare un processo di generalizzazione valido anche per cartografie più articolate. Per questi due motivi, nella fase di test, le mappe relative alla zona del Parco delle Dolomiti Bellunesi sono state integrate con le cartografie di Monselice e Chioggia, che presentano situazioni più complesse.

L'area di Monselice consiste in un territorio di circa 150 chilometri quadrati, contenente 16253 segmenti stradali, mentre quella di Chioggia si estende su una superficie di circa

600 chilometri quadrati, e comprende 9670 elementi stradali.

In particolare, la prima cartografia, quella di Monselice, è stata utilizzata per individuare una possibile strategia di generalizzazione, che è stata successivamente implementarla, testata e raffinata. La cartografia di Chioggia, invece, è stata utilizzata in un secondo momento, per la validazione del processo di generalizzazione sviluppato.

Le strade, in queste cartografie, sono classificate secondo il modello dati definito per la CTRN, dove l'attributo LIVCOD specifica il tipo di strada (Tabella 3.1).

CODICI DELLA VIABILITÀ STRADALE
LIVCOD=A201 Asse autostrada o assimilate
LIVCOD=A202 Asse autostrada o assimilate in costruzione
LIVCOD=A205 Asse strada principale
LIVCOD=A206 Asse strada secondaria
LIVCOD=A207 Asse strada carrozzabile
LIVCOD=A208 Asse strada carreggiabile
LIVCOD=A209 Asse strada in costruzione
LIVCOD=A210 Asse strada di campagna
LIVCOD=A213 Asse ponte
LIVCOD=A216 Asse guado
LIVCOD=A217 Asse sentiero facile
LIVCOD=A218 Asse sentiero difficile
LIVCOD=A219 Mulattiera
LIVCOD=A220 Passerella pedonale
LIVCOD=A221 Asse cavalcavia
LIVCOD=A222 Sottopassaggio stradale
LIVCOD=A223 Viadotto autostradale
LIVCOD=A224 Viadotto stradale
LIVCOD=A226 Tracciato galleria stradale
LIVCOD=A226A Tracciato galleria stradale
LIVCOD=A227 Asse ferrata
LIVCOD=A235 Asse pista ciclabile

Tabella 3.1: codifica della viabilità stradale del modello dati CTRN¹¹.

¹¹ http://wiki.openstreetmap.org/wiki/Convenzioni_CTRN_5000 .

4 Generalizzazione della rete stradale

La viabilità stradale è tra gli elementi di un territorio maggiormente soggetti a modifiche; le strade, inoltre, assieme all'edificato, sono le componenti che occupano in modo capillare le zone abitate, costituendo perciò una grossa parte dell'informazione fornita da una carta. È logico quindi che, all'interno del processo di generalizzazione, la rete stradale rivesta un ruolo importante e per questo necessita di un trattamento particolare ed approfondito.

In questo capitolo sarà brevemente presentato l'approccio più usato per la generalizzazione delle strade; saranno analizzate le problematiche derivanti dal cambiamento di scala, e a partire da quest'analisi, saranno definiti gli obiettivi del lavoro di tesi; verranno successivamente discusse le soluzioni al problema di generalizzazione del grafo stradale presenti in letteratura; sarà poi discussa l'importanza che riveste il *data enrichment* nel processo, analizzate alcune misure relative alla teoria dei grafi, e proposta una misura derivata da queste. Infine sarà presentato il processo di generalizzazione, risultato finale di questa tesi.

4.1 Approccio alla generalizzazione stradale

Il problema delle strade può essere affrontato suddividendolo in due sezioni (Regnauld e McMaster, 2007): quella che riguarda la generalizzazione della singola strada, e quella che tiene conto dell'intera rete.

Nel primo caso, lo scopo è quello di derivare una strada la cui forma risulti semplice e non ambigua, in modo che rimanga chiara dopo la riduzione di scala. La generalizzazione dell'elemento stradale può avvenire tramite varie azioni:

- filtraggio, ovvero la rimozione di alcuni vertici dalla geometria; questa azione è importante anche per ridurre spazio e tempi di calcolo. L'algoritmo più famoso è il Douglas-Peucker (1973), che prende il nome dagli autori e risulta essere tra i migliori algoritmi di semplificazione (McMaster, 1987);
- smoothing, ovvero il displacement di alcuni o tutti i punti della linea per attenuarne la forma;

- caricatura, ovvero l'accentuazione di una parte o dell'intera linea, per rendere più espliciti alcuni suoi dettagli.

Per generalizzazione dell'intera rete si intende, invece, il processo che porta ad un suo sfoltimento, simile all'azione di potatura di un albero. Le azioni di intervento riguardano:

- la semplificazione degli incroci stradali;
- la rimozione di sezioni di strade, preservando le caratteristiche importanti, mantenendo la connettività e massimizzando l'informazione mantenuta alla scala derivata;
- l'identificazione di pattern, come griglie, strade parallele o anelli.

Questa tesi affronta il secondo problema, quello della generalizzazione dell'intera rete: di ciò si tratterà nei successivi paragrafi.

4.2 Obiettivi del lavoro

All'interno del progetto CARGEN, in merito alla derivazione del DB25, diversi lavori sono stati svolti per la generalizzazione del grafo stradale.

Il lavoro di Lissandron (2008) ha permesso di semplificare gli incroci stradali, prima classificandoli in base al tipo di incrocio (a T, crocevia, rotonda, etc.) e successivamente generalizzandoli. La rete autostradale è stata investigata nel lavoro di Zanon (2009), che ha avuto tra gli obiettivi quello di classificare gli elementi autostradali (tratto autostradale, area di servizio e corsia di immissione) e di accoppiare le carreggiate. Infine, con il lavoro di Biasion (2009) è stato sviluppato un algoritmo per la generalizzazione dei tornanti. Sia per il fatto che gli algoritmi sviluppati sono parametrici, sia perché le stesse problematiche di generalizzazione si riscontrano anche a scale inferiori, questi algoritmi rimangono applicabili per la derivazione cartografica alla scala 1:50 000.

Tuttavia, essi non sono sufficienti per una generalizzazione completa del grafo stradale. Così, durante lo studio iniziale del problema, sono emerse la possibilità e la necessità di intervenire in tre diverse aree, di seguito presentate.

4.2.1 Armonizzazione

La prima area di intervento tratta l'armonizzazione del grafo stradale.

Per armonizzazione del grafo stradale si intende il processo che porta a risolvere il problema della disomogeneità dei dati. Ogni dato di partenza, che identifica una strada, è costituito da una geometria, più precisamente da una linea, e da un attributo codice, che specifica la categoria della strada (strada principale, strada secondaria, strada di campagna, etc.). Rendere omogenei i dati equivale ad attribuire la medesima categoria ad un gruppo di strade che sono legate da una qualche relazione spaziale o semantica.

Più in dettaglio, l'armonizzazione si basa su due processi: il processo di *data enrichment* e il processo di riclassificazione.

Il processo di *data enrichment* ha un ruolo fondamentale nella generalizzazione, e consiste nell'esplicitare informazioni implicite dei dati; questi dati, infatti, oltre alle informazioni esplicite consistenti nella geometria e negli attributi, contengono informazioni che dipendono dal contesto, ma che non sono espresse formalmente. Ad esempio, osservando la tipologia di edificio che si trova al lato di una strada, si può determinare se una strada è di quartiere, di campagna, o magari se serve una zona industriale.

Il secondo processo su cui si basa l'armonizzazione, quello di riclassificazione, si basa sulle informazioni estrapolate durante la fase di *data enrichment*. La riclassificazione si può manifestare in due modi: tramite l'estensione delle classi esistenti, ad esempio individuando una nuova categoria di strade, come la strada di quartiere, e tramite la sostituzione dei codici. Quest'ultima modalità è una fase estremamente delicata del processo di armonizzazione, in quanto l'effetto prodotto è la modifica dei valori degli attributi dei dati iniziali; per questo motivo la riclassificazione rende quasi d'obbligo un'analisi a posteriori dei dati.

Due sono i motivi che portano ad attuare un processo di armonizzazione: il primo è in relazione al cambiamento di scala; il secondo, invece, è il relazione al fatto di dover passare da un modello dati ad un altro con caratteristiche differenti.

Con il passaggio ad una scala inferiore, i particolari presenti nella mappa ad alta scala possono mutare di importanza, oppure risultare superflui. Per capire questo fatto, si

consideri la Figura 4.1, dove è presente una strada principale (rossa) che, in prossimità della città, cambia in strada secondaria, per poi ritornare strada principale. Una distinzione di questo tipo è assolutamente lecita rappresentarla alla scala 1:5 000; mantenere però questa altalenante classificazione della strada alla scala 1:50 000 non permette di risaltare il fatto che quella è una strada importante poiché permette di oltrepassare la città; mediante l'eliminazione di questo dettaglio, attuabile riclassificando la strada secondaria in principale, si farebbe percepire all'utente della mappa la vera importanza di questa strada. L'armonizzazione ha tra gli scopi, quindi, quello di mettere in evidenza gli elementi più importanti all'interno della mappa.

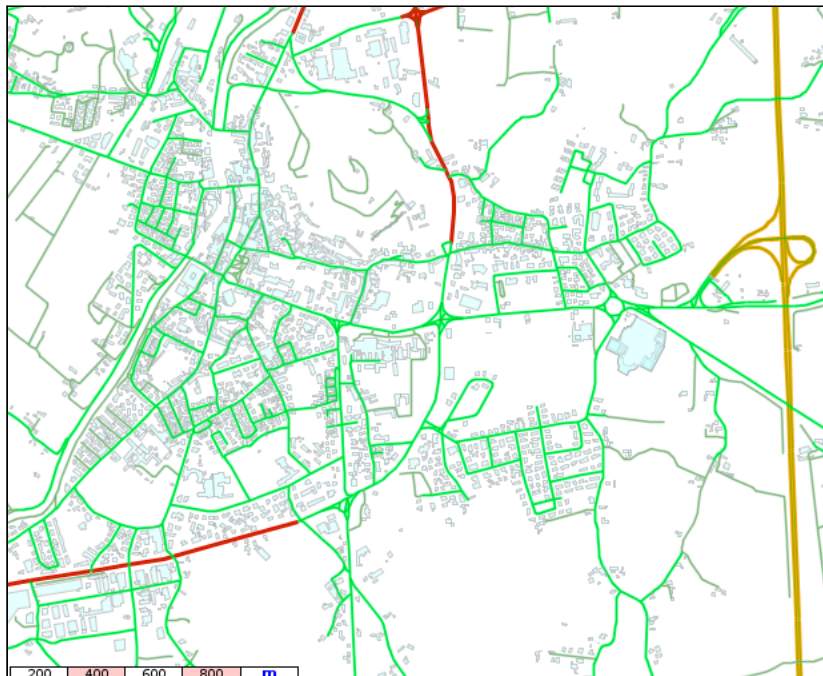


Figura 4.1: esempio di strada che, in prossimità del centro urbano, assume una classe di importanza inferiore. Dati relativi alla cartografia di Monselice (CTRN).

Il secondo obiettivo dell'armonizzazione ha a che fare con il cambiamento del modello dati, che rende necessaria l'estensione delle classi stradali esistenti. Infatti, tra due modelli dati differenti, non sempre è possibile individuare una relazione tra le classi del primo modello e quelle del secondo. In particolare, nel modello DB25 dell'IGM, è presente la classe stradale "strada urbana di quartiere", che non esiste nel modello dati CTRN. Non essendoci, in quest'ultimo modello, nessuna classe stradale con

caratteristiche simili a quella di quartiere, è necessario realizzare un processo specifico per poter individuare questo particolare tipo di strade. Quindi, un altro scopo dell'armonizzazione è quello di identificare gruppi di strade, che condividono le stesse caratteristiche, per poterli poi inserire in una nuova categoria stradale che non è definita nel modello dati di partenza.

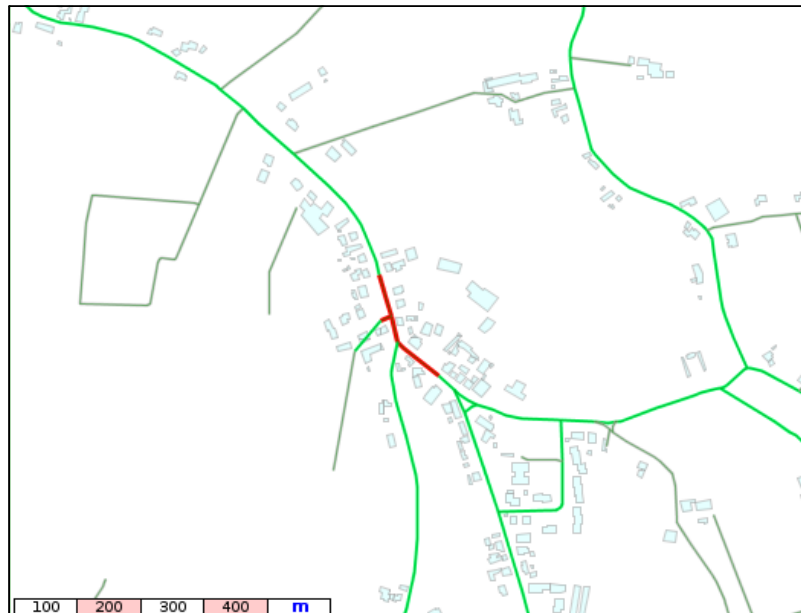


Figura 4.2: esempio di strada classificata erroneamente come strada principale (strada rossa). Dati relativi alla cartografia di Monselice (CTRN).

Attuare un processo di armonizzazione porta con se due importanti vantaggi. Innanzitutto, l'estensione delle classi esistenti rappresenta un elemento che torna utile nelle successive fasi di generalizzazione: ad esempio, caratterizzare le strade di quartiere è di aiuto al problema dello sfoltimento della viabilità nei centri urbani. In secondo luogo, l'armonizzazione ha come conseguenza quella di aumentare la robustezza degli algoritmi rispetto agli errori presenti nei dati iniziali. L'errore dei dati è, in questo campo, abbastanza frequente a causa della segmentazione stradale, caratteristica dei dati originali. Non sempre i segmenti stradali che costituiscono percettivamente la stessa strada hanno in comune la stessa categoria. In Figura 4.2 è mostrato l'esempio di una strada che, da secondaria (verde), passa per un breve tratto a strada principale (rossa), per poi ritornare secondaria. Una situazione simile rappresenta,

molto probabilmente, un errore di classificazione¹; l'armonizzazione di questa strada elimina implicitamente l'errore di classificazione presente nel dato iniziale.

Bisogna sottolineare, comunque, che lo scopo dell'armonizzazione non è quello di correggere errori, ma quello di fornire una rappresentazione dei dati che sia più omogenea e coerente rispetto a quella iniziale; la correzione è solo un effetto collaterale.

Riassumendo, in questa tesi si ritiene necessario intervenire con lo sviluppo di un processo di armonizzazione del grafo stradale, sia per la riduzione di dettaglio a cui è soggetta una mappa a bassa scala, sia per la necessità di estendere le classi esistenti per la derivazione del DB25 e sia per aumentare la coerenza tra gli elementi della mappa.

4.2.2 Selezione

La seconda area di intervento è relativa allo sfoltimento del grafo stradale.

La selezione delle strade da mantenere, o la scelta di quelle da eliminare, è una delle problematiche più evidenti che si hanno con la riduzione della scala, poiché sarebbe impensabile, nella carta derivata, rappresentare ogni singolo segmento stradale, soprattutto nelle zone dove la densità è elevata.

Una classe di strade per cui esistono regole ben precise per guidarne lo sfoltimento, è quella relativa alle strade *dangling*. Si definisce *dangling* un segmento stradale che non svolge nessuna funzione di collegamento per altre strade, ovvero ha almeno un estremo sconnesso dal resto del grafo. L'IGM regola la selezione dei *dangling* in base alla categoria e alla lunghezza (IGM, 2004)².

L'eliminazione dei *dangling* non è però sufficiente per garantire leggibilità e qualità dell'informazione. L'IGM sottolinea che le vie di comunicazione secondarie vanno opportunamente sfoltite quando la trama viaria è fitta, oppure laddove esigenze di chiarezza³ della rappresentazione grafica lo richiedano. Similmente, per la viabilità all'interno di un agglomerato urbano, si deve procedere con uno sfoltimento consistente

1 L'errore di classificazione di questa strada è confermato dalla carta geografica della serie 25DB dell'IGM, che mostra l'assenza della strada rossa in questione.

2 Una strada che non raggiunge la lunghezza minima specificata per la sua categoria può essere eliminata.

3 La chiarezza è un requisito fondamentale della rappresentazione grafica in base al quale ogni particolare deve essere rappresentato in modo tale che ne risulti agevole l'individuazione e l'identificazione sulla carta, e che sia facilitata la comprensione delle relazioni di posizione con gli altri particolari. La chiarezza è direttamente legata alla selezione dei particolari da rappresentare e al relativo convenzionalismo grafico utilizzato (IGM, 2004).

laddove vi sia la presenza di maglie particolarmente fitte (IGM, 2004).

Il processo di selezione ha dunque l'obiettivo di identificare ed eliminare i segmenti stradali che, per vari motivi, sono di scarsa utilità nella carta derivata e ne diminuiscono la chiarezza.

Anche all'interno di questo processo è inserita una fase di *data enrichment*; questa fase ha come scopo quello di individuare, tra un gruppo di strade in concorrenza per lo stesso spazio, quella che più si presta ad essere eliminata.

4.2.3 Tipificazione

L'ultima area toccata da questo lavoro di tesi è relativa alla tipificazione delle strutture parallele, o meglio, di strade ad andamento semi-parallelo, presenti nella viabilità normale.

Un tema simile è già stato trattato nel lavoro di Zanon (2009), il quale però era concentrato sull'unico obiettivo di accorpate le carreggiate autostradali: la soluzione proposta risulta dunque specifica per la rete autostradale, che è inserita in un contesto ben definito.

Per quanto riguarda la viabilità normale, invece, la presenza di incroci e di strade con andamenti irregolari e di classificazione diversa aggiunge nuove problematiche alla generalizzazione. L'obiettivo è quello di individuare coppie di strade che scorrono vicine per poi eliminare la meno importante o sostituirle entrambe con una linea di mezzera, in modo da diminuire l'area occupata dalla viabilità stradale; questo sia per dar spazio ad altri elementi geografici, sia per aumentare la chiarezza della mappa derivata.

4.3 Stato dell'arte

La ricerca sulla generalizzazione della rete stradale è stata negli anni oggetto di intensi studi, vista l'enorme importanza che la viabilità riveste nella cartografia. Il lavoro di ricerca correlato si può suddividere in due temi portanti: quello relativo al *data enrichment*, e quello relativo alla tipificazione, la quale comunque è in stretta dipendenza dal primo tema.

Il principale problema che in questa tesi viene affrontato è quello di trovare una metodologia per caratterizzare l'importanza di una strada, operazione fondamentale per la fase di selezione. È importante sottolineare che, nella cartografia digitale, le strade non vengono rappresentate come comunemente le intendiamo; Ogni elemento stradale definisce solamente il tratto stradale tra due incroci consecutivi. Per oltrepassare in parte questo problema, diventa essenziale il modello di strada proposto da Thomson e Richardson (1999), definito *stroke*, e globalmente utilizzato nell'ambito della generalizzazione delle strade. Il concetto di *stroke* si basa sulla teoria del raggruppamento percettivo (*perceptual grouping*), che descrive il modo con cui il sistema visivo umano organizza gli elementi del campo visivo. In questa teoria viene definito il principio di buona continuazione (Figura 4.3), che mette in evidenza il fatto che elementi che seguono la medesima direzione vengono percepiti dall'occhio umano come appartenenti allo stesso gruppo. Da questo principio nasce l'idea dello *stroke*, che è l'insieme di tutti quei tratti stradali che danno l'idea di formare un'unica strada. Lo *stroke* è utile laddove sia necessario individuare gli elementi più importanti di un grafo: questa operazione si può quindi svolgere basandosi sulla lunghezza dello *stroke*. Nel caso più semplice, la costruzione dello *stroke* avviene analizzando solo le geometrie dei segmenti; per ottenere risultati più sofisticati, invece, si possono considerare informazioni aggiuntive eventualmente disponibili, come la classificazione del segmento stradale o il suo nome.

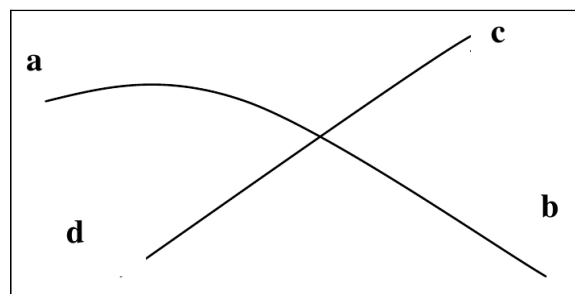


Figura 4.3: principio di buona continuazione; i segmenti a e b vengono percepiti come appartenenti alla stessa entità; lo stesso vale per i segmenti c e d.

Thomson e Brooks (2000) propongono alcuni principi relativi agli *stroke*, allo scopo di individuare per essi una gerarchia di importanza. Due di questi principi affermano che: a parità di condizioni, uno *stroke* più lungo è più importante di uno meno lungo; e che

uno stroke di classificazione superiore (ad esempio, una superstrada) è più importante di uno di categoria inferiore (ad esempio, una strada urbana). Tuttavia lo stroke, trattato esclusivamente alla luce di questi principi, non è sufficiente per individuare sistematicamente quali strade sono più importanti di altre.

Alcune soluzioni sono state proposte: Richardson e Thomson (1996) hanno identificato le strade più importanti come quelle che vengono coinvolte nei percorsi minimi da un luogo di interesse all'altro (ad esempio dall'ospedale alla stazione). Questo approccio, però, si adatta meglio a scale più basse rispetto a quelle trattate in questa tesi; inoltre, i punti di interesse devono essere chiaramente definiti e disponibili in fase di generalizzazione. Chaudhry (2004) ha invece elaborato un algoritmo per la selezione degli stroke, basato sulla loro lunghezza e sulla loro connettività: in pratica, viene identificato un gruppo iniziale di stroke importanti, e iterativamente si aggiungono a questo gruppo ulteriori stroke sino ad ottenere un grafo connesso. Per definire il numero totale di stroke da mantenere viene utilizzata la *radical law* (Topfer e Pillewizer, 1966), definita come

$$n_f = n_a \sqrt{(M_a \div M_f)}$$

dove n_f è il numero di oggetti che possono essere mostrati alla scala derivata (numero di stroke), n_a il numero di oggetti presenti nella carta iniziale, mentre M_a e M_f sono rispettivamente l'inverso della scala sorgente e l'inverso della scala di destinazione. La *radical law* risulta sicuramente utile, ma rimane comunque una misura generale, che non sempre permette di individuare correttamente gli stroke da eliminare, e deve quindi essere integrata con qualche altra misura o strategia di selezione. L'algoritmo proposto da Chaudhry, utilizzando la *radical law* come vincolo al numero massimo di stroke selezionabili, presenta il difetto di non tener conto dell'eventuale separazione visuale tra gli elementi nella mappa.

Jiang e Claramunt (2004) sviluppano un algoritmo per la selezione delle strade basato sulla teoria dei grafi. Essi modellano la rete stradale mediante un grafo, dove i nodi simboleggiano i nomi delle strade, mentre gli archi identificano le intersezioni tra coppie di strade. Per ogni strada S vengono calcolate tre misure di centralità, allo scopo di quantificare l'importanza che essa ha all'interno del grafo: la prima misura, *degree centrality*, indica il numero di strade a cui S è connessa; la seconda, *closeness centrality*,

si riferisce al numero minimo di archi che permettono a S di raggiungere tutti gli altri nodi della rete, misurando quanto, per le altre strade, S risulta vicina; la terza, *betweenness centrality*, indica se e in che misura la strada S svolge una funzione di collegamento per le altre strade. La selezione delle strade da eliminare viene fatta gerarchicamente, applicando sequenzialmente dei vincoli sui valori minimi delle tre misure. Ad esempio, prima si eliminano le strade con connettività uguale a 2, poi quelle con 3 e così via. Le misure citate sono riprese da Porta (2006), il quale ne considera ulteriori, prima fra tutte *information centrality*. In particolare, per calcolare *information centrality* relativa allo stroke S , si fa ricorso ad una misura di efficienza globale della rete, che rappresenta il valore medio dei cammini minimi nella rete; questa misura di efficienza viene calcolata due volte, una volta per l'intera rete, e una volta per la rete privata dello stroke S , e la differenza delle due efficienze fornisce *information centrality* di S . Questa misura è utilizzata per valutare le conseguenze derivanti dall'eliminazione di un nodo nella rete, e specifica in che modo diminuisce l'abilità di propagare informazioni della rete in seguito alla disattivazione della strada S .

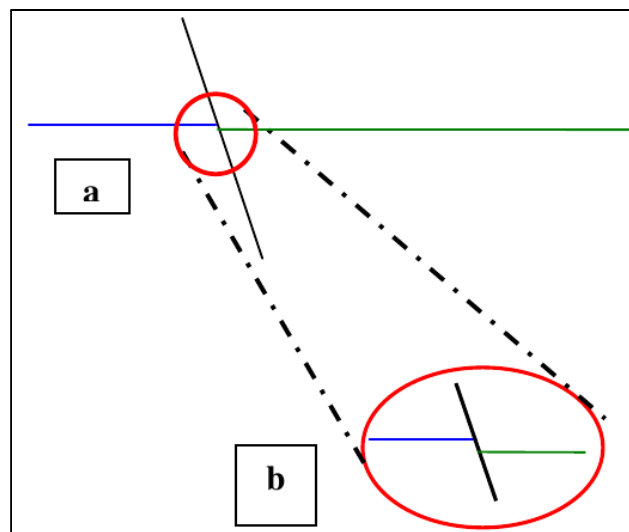


Figura 4.4: tipica situazione in cui non viene ricostruito un unico stroke. Fonte: Jiang e Claramunt (2004).

La lunghezza dello stroke, comunque, rimane una misura fondamentale per consentire di valutare la sua importanza. Bisogna sottolineare, tuttavia, che questa misura può essere falsata se il naturale proseguimento della strada è interrotta da un incrocio, oppure se i segmenti stradali presentano classificazioni differenti. Un semplice caso in

cui si materializza questo problema è mostrato da Jiang e Claramunt (2004), illustrato in Figura 4.4. È importante, perciò, semplificare le giunzioni stradali che provocano la rottura degli stroke in due parti, per poter trattare così con stroke più accurati. Questo problema, ovvero la generalizzazione degli incroci, è stato affrontato negli studi di Mackaness e Mackechnie (1999), Touya (2007) e Savino (2010).

Gli stroke e gli incroci non sono comunque gli unici pattern ad esser stati investigati. Heinzle (2006) presenta alcune strategie per l'individuazione di griglie, anelli e stelle, mentre Thom (2005) e Luan e Yang (2010) affrontano il tema dell'identificazione di strutture parallele.

Il *data enrichment*, costituito dalle informazioni ottenute dall'uso delle misure presentate in questa sezione e dal riconoscimento di particolari pattern, è l'elemento chiave per ottenere una buona generalizzazione, sia per quanto riguarda il processo di selezione delle strade, sia per quanto riguarda il processo di armonizzazione. La successiva sezione tratta in dettaglio questo tema.

4.4 Data enrichment come elemento chiave della generalizzazione

Decidere quale algoritmo o quale azione applicare ad un elemento stradale è una scelta importante che va operata avendo a disposizione il maggior numero di informazioni possibili. Le informazioni fornite dalla componente geometrica o dalle componenti statistiche del singolo dato spesso non sono sufficienti per individuare la giusta azione da compiere. Proprio per questo, la fase di *data enrichment* riveste un ruolo fondamentale nel processo, in quanto aumenta le informazioni utilizzabili per guidare verso il tipo d'azione che è più opportuno compiere. Come il cartografo utilizza durante il processo cartografico diverse informazioni che non sono propriamente scritte, ma che percepisce tramite una visione globale della mappa, il processo di generalizzazione automatico dovrebbe sfruttare tutte le informazioni nascoste tra i dati per migliorare i risultati. Sorge così il problema di come estrapolare queste informazioni dai dati.

Alcune informazioni possono essere estratte osservando il contesto in cui la strada è inserita: a quante altre strade essa è connessa e come queste ultime sono classificate, quanti edifici serve e a quale tipologia essi appartengono. Queste statistiche danno una prima caratterizzazione della strada: ad esempio, si possono identificare quali strade

appartengono ad una zona industriale verificando se gli edifici vicini sono industriali o civili, oppure quali strade non conducono a nessun particolare punto di interesse.

Anche l'analisi della morfologia è un fattore che incrementa la conoscenza relativa al fenomeno; tramite questa analisi è possibile, prima di tutto, riconoscere quali segmenti stradali danno l'idea di appartenere alla stessa strada (stroke); in secondo luogo, è possibile riconoscere particolari pattern, in modo da specificare ulteriormente il ruolo di una strada. Ad esempio, individuare mediante lo studio morfologico quali segmenti stradali appartengono ad una giunzione stradale dà la possibilità di trattare questo tipo di segmenti con una procedura specifica e adeguata.

Infine, con un'analisi più globale della rete si potrebbero caratterizzare le strade in base alla funzione che esse svolgono all'interno della rete stessa: individuare una strada che ha un ruolo chiave come via di collegamento, oppure una che ha come unico scopo quello di servire un gruppo di case, è di grande aiuto nel momento in cui si voglia far risaltare la struttura portante della rete o sfoltire zone ad alta densità stradale.

Il *data enrichment*, attraverso cui si estraggono informazioni nascoste e implicite, è quindi il fattore chiave di una buona generalizzazione. Come illustrato nella sezione 4.3, il primo modo per estrarre queste informazioni implicite si attua tramite il modello dello stroke, con il quale si individuano gruppi di segmenti appartenenti ad un'unica entità. Il passaggio da una sequenza di segmenti ad un'unica strada porta ad estrarre nuove misure, come la lunghezza della strada o il numero di strade connesse ad essa.

Per ottenere un'ulteriore informazione sullo stroke è necessario individuare il ruolo che esso svolge all'interno della rete; questa informazione è strategica per svolgere al meglio le operazioni di armonizzazione e selezione. A tal proposito, in questa tesi si è deciso di puntare sulle misure di centralità per estrarre informazioni sullo stroke che non sono basate solamente sul contesto in cui è inserito. Le quattro misure di centralità considerate, ovvero la *degree centrality*, *closeness centrality*, *betweenness centrality* e l'*information centrality*, come si vedrà nel paragrafo 4.4.2, non hanno però dato i risultati sperati. È stata perciò sviluppata una misura derivata, per risolvere le carenze dimostrate delle misure di centralità analizzate.

Prima di descrivere queste misure di centralità, però, è opportuno specificare in che modo la rete stradale può essere formalmente modellata. Questo è l'obiettivo della

sezione che segue.

4.4.1 Rappresentazione della rete stradale

Una rete stradale può essere modellata mediante l'utilizzo di un grafo $G(V, N)$, dove V è l'insieme dei vertici e N l'insieme degli archi. Mediante un grafo, è possibile rappresentare la rete stradale in due differenti modi.

Nel primo caso, ogni arco dell'insieme N identifica un segmento stradale e ogni vertice dell'insieme V simboleggia l'intersezione tra due o più segmenti. Associando come peso ad ogni arco la lunghezza del relativo tratto stradale, questo tipo di grafo⁴ si adatta benissimo per il calcolo dei cammini minimi tra due punti geografici.

Se, invece, al posto del segmento stradale si utilizza lo stroke come elemento primitivo, la rete stradale viene modellata sotto un altro punto di vista. All'insieme dei vertici si fa corrispondere l'insieme degli stroke, mentre l'intersezione tra coppie di stroke (n_1, n_2) viene rappresentata tramite un arco che va da n_1 a n_2 . È d'obbligo notare che questo tipo di grafo⁵ non funziona per il calcolo dei cammini minimi, semplicemente perché non ha nessun significato attribuire un peso all'arco.

4.4.2 Misure di centralità

Sulla base dei due modelli appena descritti, sono state implementate le quattro misure di centralità, a cui si era accennato nella sezione 4.3, e che di seguito vengono descritte e analizzate in modo più completo.

Si consideri il grafo degli stroke $G(V, E)$, con $n = |V|$, e lo stroke v_i appartenente a V . Le quattro misure di centralità, relative allo stroke v_i , sono così definite:

1. *Degree Centrality*:

$$C_d(v_i) = \frac{1}{n-1} \cdot \sum_{k=1}^n r(v_i, v_k)$$

dove $r(v_i, v_k) = 1$ se i due stroke si intersecano, 0 altrimenti; questa misura indica il numero di stroke a cui v_i è connesso;

⁴ Verrà chiamato, da qui in poi, grafo planare.

⁵ Verrà chiamato, da qui in poi, grafo degli stroke.

2. *Closeness Centrality*:

$$C_c(v_i) = \frac{(n-1)}{\sum_{k=1 \text{ to } n} d(v_i, v_k)}$$

dove $d(v_i, v_k)$ è la distanza minima tra lo stroke v_i e v_k ; questa misura indica quanto v_i è vicino a tutti gli altri stroke;

3. *Betweenness Centrality*:

$$C_b(v_i) = \frac{1}{(n-1)(n-2)} \cdot \sum_{j=1}^n \sum_{k=1}^{j-1} \frac{p_{kj}(i)}{p_{kj}}$$

dove $p_{kj}(i)$ è il numero di percorsi minimi da v_k a v_j che passano per v_i e p_{kj} il numero di percorsi minimi da v_k a v_j ; questa misura indica quanto v_i è presente nei percorsi minimi tra tutti gli altri nodi della rete;

4. *Information Centrality*:

$$C_i(v_i) = \frac{\Delta E_{glob}}{\Delta E_{glob}} = \frac{E_{glob}(G) - E_{glob}(G')}{E_{glob}(G)}$$

dove $E_{glob}(G)$, $E_{glob}(G')$ sono rispettivamente il valore di efficienza globale⁶ dell'intera rete e della rete privata del nodo v_i ; questa misura indica la capacità della rete a diffondere informazioni senza il nodo v_i .

Analizzando queste quattro misure, per prima cosa si osserva che solo la *degree centrality* è una misura locale: il suo valore dipende infatti da un limitato numero di nodi, vicini al nodo v_i . Le restanti tre misure, invece, sono riferite alla struttura dell'intero grafo. Quest'ultimo fatto costituisce, di per sé, una prima forma di gerarchia, ma non si adatta perfettamente con scale medie (1:25 000 e 1:50 000); quando si opera a questi livelli di scala, infatti, è necessario individuare anche localmente una gerarchia di importanza. Ad esempio, se pensiamo a due quartieri che si trovano in zone opposte della città, i valori di *closeness* delle strade appartenenti ad un quartiere sono influenzati dalla presenza delle strade nell'altro quartiere. In questa situazione, poiché la misura di *closeness* dipende da tutta la rete stradale rappresentata nel grafo, non è possibile identificare la strada centrale del quartiere. La dipendenza delle misure dall'intera rete,

6 L'efficienza globale è definita come $E_{glob}(G) = \frac{1}{n(n-1)} \sum_{i,j \in N, i \neq j} \frac{1}{d_{ij}}$.

dunque, diminuisce la loro utilità nel produrre un'informazione valida per la generalizzazione.

Queste misure, inoltre, presentano lo svantaggio di dipendere dalla posizione all'interno del grafo dello stroke a cui si riferiscono: un stroke, così, avrà sicuramente valori di centralità superiori se si trova a centro mappa, e, proprio per queste ragioni, la *closeness centrality* è stata reputata inadatta per valutare i diversi livelli d'importanza degli stroke. L'*information centrality* soffre le stesse limitazioni, cui si aggiunge la complessità computazionale; quest'ultima, in particolare, è dell'ordine di $O(n^4)$, se la si calcola per tutti gli n stroke della rete, poiché il calcolo di $E_{glob}(G')$ comporta il ricalcolo delle distanze minime. Si è notato, comunque, che queste distanze non variano sempre con la disattivazione di un nodo della rete, per esempio nel caso degli stroke dangling che non si trovano mai nel cammino minimo tra due qualsiasi nodi della rete; tenendo conto di ciò, la complessità computazionale diminuirebbe. Tuttavia, anche considerando questo fatto nell'implementazione della misura, i tempi di calcolo sono risultati inaccettabili, e l'*information centrality* è stata quindi abbandonata.

L'ultima misura di centralità, la *betweenness centrality*, è stata quella che più si è avvicinata all'obiettivo di dare una gerarchia di importanza alle strade. Questa misura riesce spesso a distinguere, anche localmente, le strade che svolgono importanti funzioni di comunicazione per il resto della rete da quelle che invece non sono così critiche. Essendo questa una misura che richiede il calcolo dei percorsi minimi tra tutti i nodi della rete, il costo computazionale è, applicando l'algoritmo di Dijkstra, dell'ordine di $O(n^3)$. Per quanto riguarda invece l'utilizzo pratico della *betweenness centrality*, i valori più alti si trovano per lo più negli stroke a centro mappa, come d'altra parte ci si poteva aspettare: uno stroke che si trova fuori dalla zona centrale sarà presente, infatti, molte meno volte nel cammino minimo tra gli altri nodi della rete. Per questa caratteristica della misura, se la selezione delle strade si attuasse imponendo un valore di soglia minimo di *betweenness*, il centro mappa risulterebbe molto più denso di strade rispetto al resto. A questa proprietà negativa si aggiunge anche l'elevato tempo computazionale, che pur essendo di un ordine di grandezza inferiore rispetto a quello della *information centrality*, rimane ancora troppo elevato⁷. Tutte queste considerazioni hanno portato a

⁷ Il tempo di calcolo della *betweenness* sulla cartografia di Monselice, costituita da circa 4500 stroke, in una macchina con processore 1.86GHz, è stato di 22 minuti.

sviluppare una misura derivata dalla *betweenness*, definita *local betweenness*.

4.4.3 Local betweenness

L'idea della *local betweenness* nasce dall'osservazione che, per distinguere quali strade in un gruppo più o meno limitato di esse rappresentano importanti vie di comunicazione, non serve coinvolgere l'intero grafo: se uno stroke svolge il ruolo di “ponte” per gli altri stroke della rete, a maggior ragione esso svolgerà tale funzione per quelli che lo circondano. Ciò significa che solamente considerando le strade in un intorno dello stroke si può dedurre se e in che misura lo stroke svolge una funzione di collegamento.

Per definire la *local betweenness*, essendo essa basata su n sottografi diversi, bisogna prima definire in cosa consiste il sottografo relativo al singolo stroke.

Il grafo $G_i(V_i, E_i)$ relativo allo stroke v_i , sottografo di $G(V, E)$, è costituito da tutti gli stroke v_j appartenenti a V per i quali vale la relazione

$$d_{ij} < d_{max}$$

dove d_{ij} è la distanza minima tra i due stroke, e d_{max} è una soglia massima prefissata.

La distanza minima tra due stroke è intesa come la lunghezza del cammino minimo da v_i a v_j ; se i due stroke si intersecano la loro distanza è zero. E_i , invece, è l'insieme di tutti gli archi utilizzati nei percorsi minimi verso v_i . La dimensione di ogni sottografo varia quindi in base alle caratteristiche dell'intorno dello stroke.

La *local betweenness* viene definita come

$$C_{lb}(v_i) = \frac{1}{(|G_i| - 1)(|G_i| - 2)} \sum_{v_j, v_k \in G_i, j > k} \frac{p_{kj}(i)}{p_{kj}}$$

dove $p_{kj}(i)$ è il numero di percorsi minimi da v_k a v_j che passano per v_i e p_{kj} il numero di percorsi minimi da v_k a v_j .

Il tempo computazionale della *local betweenness* per l'intero grafo dipende dal parametro d_{max} e dalla densità stradale. Più stroke si trovano nella stessa zona e più cammini minimi devono essere calcolati. Per d_{max} che tende a infinito, la *local betweenness* corrisponde alla *betweenness*⁸.

⁸ Il tempo di calcolo della *local betweenness* sulla cartografia di Monselice (4500 stroke), con parametro $d_{max} = 3000$, è stato di 1 minuto e 10 secondi, effettuando il calcolo sulla stessa macchina

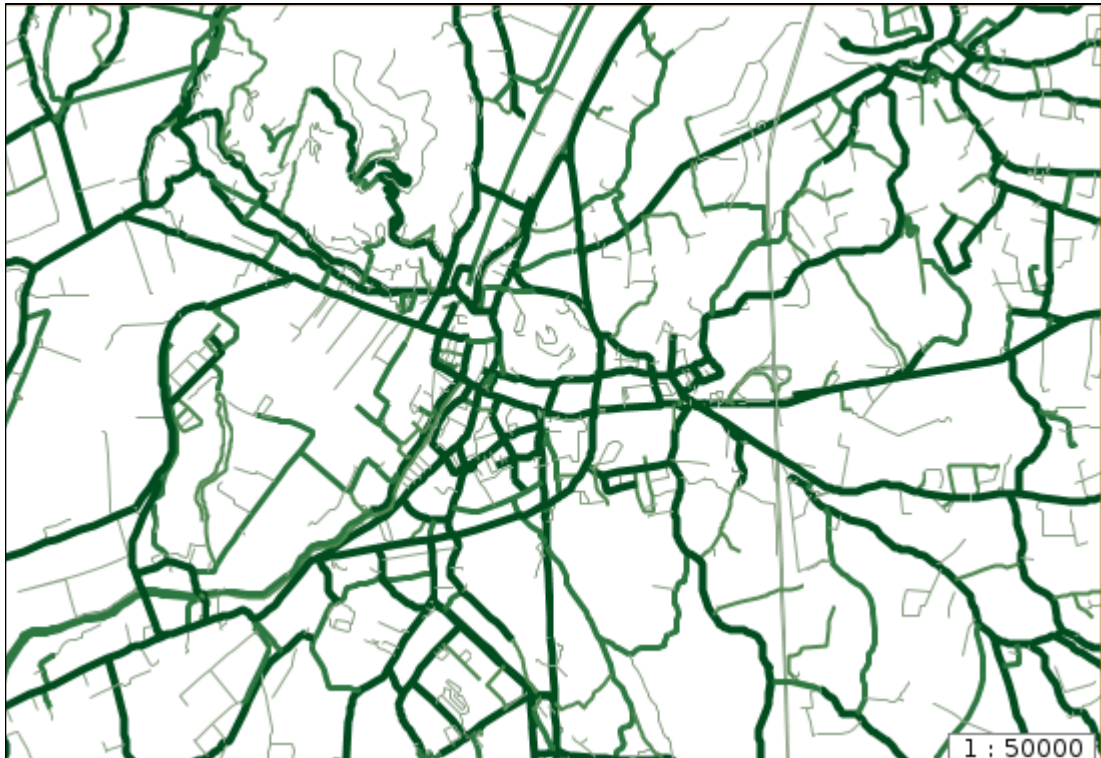


Figura 4.5: *Local betweenness* nell'area di Monselice, con $d_{\max} = 3000$. Lo spessore della linea dello stroke è proporzionale al valore di *local betweenness*. L'immagine rappresenta solo una parte dell'area in cui la *local betweenness* è stata calcolata.

Il calcolo della *local betweenness* implica il calcolo dei cammini minimi tra tutte le coppie di stroke che si trovano a distanza inferiore di d_{\max} ; è stato perciò implementato l'algoritmo di Dijkstra, che permette di calcolare tutti i percorsi minimi da un nodo a tutti gli altri nodi del grafo, in tempo $O(|V|^2)$. Il grafo $G(V, E)$ è definito dall'insieme dei vertici V , che simboleggiano le intersezioni stradali, e dall'insieme E che rappresenta i segmenti stradali. Ogni stroke s_i è identificato dall'insieme di vertici V_i per mezzo dei quali si interseca con gli altri stroke; l'algoritmo eseguito sullo stroke s_i calcola le distanze tra i vertici V_i e tutti gli altri vertici di V a distanza massima d_{\max} . Durante il calcolo delle distanze, per ogni vertice v appartenente a V viene memorizzato anche il suo predecessore nel percorso minimo verso il nodo più vicino di V_i . Le distanze tra tutte le coppie di vertici in V_i sono impostate a 0, in quanto tali vertici rappresentano lo stesso stroke s_i .

utilizzata per il calcolo della *betweenness*.

Dijkstra($G(V, E), V_i, d_{max}$):	
<pre> 1 for each $v \in V$: // Initializations 2 $dist[v] := infinity$; // Unknown distance function from V_i to v 3 $pred[v] := undefined$; // Previous node in optimal path from V_i 4 end for ; 5 for each $v \in V_i$: 6 $dist[v] := 0$; 7 end for ; 8 $Q := V$; 9 while Q is not empty : 10 $u :=$ vertex in Q with smallest $dist[]$; 11 if $dist[u] > d_{max}$: 12 break ; 13 end if ; 14 remove u from Q ; 15 for each neighbor v of u: 16 $d := dist[u] + E(u, v).weight$; 17 if $d < dist[v]$: 18 $dist[v] := d$; 19 $pred[v] := u$; 20 end if ; 21 end for ; 22 end while ; 23 return $pred[]$; </pre>	

Tabella 4.1: algoritmo di Dijkstra, adattato al calcolo dei percorsi minimi tra stroke.

L'algoritmo di Dijkstra, adattato per il calcolo dei percorsi minimi tra stroke, è descritto in tabella 4.1; esso ha come output l'array $pred_i[]$ dei predecessori, che indica, per ogni nodo, il relativo nodo predecessore nel cammino minimo verso s_i . Il calcolo della *local betweenness* non utilizza i valori delle distanze minime, ma solo dei cammini. L'informazione necessaria viene estrapolata dalla struttura $pred_i[]$, dalla quale si ricavano gli stroke appartenenti al sottografo $G_i(V_i, E_i)$. Al contrario, il numero di volte n_i che v_i partecipa nel percorso minimo tra due stroke appartenenti a G_i , non viene ricavato da $pred_i[]$, ma da tutti gli array $pred_j[]$, con v_j appartenente a V_i . n_i viene aggiornato dopo ogni esecuzione dell'algoritmo di Dijkstra relativa ad uno stroke di $G_i(V_i, E_i)$; il valore finale di n_i si potrà ottenere solamente dopo aver calcolato tutti i percorsi minimi tra tutti gli stroke appartenenti a $G_i(V_i, E_i)$.

4.5 Fasi del processo di generalizzazione

Una volta specificati gli obiettivi di questo lavoro e introdotte le tecniche e le metodologie per raggiungerli, in particolare quelle relative al *data enrichment*, è possibile definire la struttura del processo di generalizzazione del grafo stradale.

Il processo di generalizzazione è il risultato di tre decisioni: quali algoritmi utilizzare e in che contesto, con quali parametri, e in che ordine eseguirli. Nel caso specifico del progetto CARGEN, trattato in questa tesi, è stato chiaramente necessario considerare anche gli algoritmi relativi alle strade già prodotti all'interno del progetto stesso.

La struttura che è stata definita vede come prima fase l'armonizzazione del grafo stradale, costituita da una prima parte di *data enrichment* e da una seconda parte di riclassificazione. Questa scelta è motivata dal fatto che la fase di selezione e quella di tipificazione sono in stretta dipendenza dalla fase di armonizzazione: la riclassificazione delle strade, infatti, si ripercuote sulla costruzione degli stroke, le cui lunghezze sono uno dei fattori determinanti nel processo di selezione; inoltre, la riclassificazione migliora il processo di tipificazione, sia delle giunzioni stradali che delle strade parallele. Ad esempio, rendere omogenea una strada, parallela ad un'altra strada, che nei dati iniziali è costituita da segmenti non tutti di stessa classificazione, ottimizza la semplificazione di strutture parallele, essendo l'algoritmo di tipificazione basato sugli stroke, che sono a loro volta costruiti in funzione della categoria del segmento stradale. Allo stesso modo, la tipificazione degli incroci stradali, utilizzando anch'essa il modello degli stroke, può beneficiare di una classificazione più omogenea delle strade.

Bisogna osservare, comunque, che già durante l'armonizzazione si fa un importante uso degli stroke. Proprio per questo motivo, è stato deciso di implementare un algoritmo per risolvere il problema dei segmenti disallineati da un incrocio (Figura 4.4); questo algoritmo permette di costruire un unico stroke invece di due, esattamente come verrebbe percepito visualizzando la strada a bassa scala.

Una volta risolti i disallineamenti e armonizzato il grafo, si passa allo sfoltimento della viabilità relativa agli stroke dangling della rete. In questa fase l'eliminazione dei dangling è infatti necessaria: permette di alleggerire i successivi algoritmi di tipificazione, che così si troveranno a processare una mole di dati decisamente inferiore.

Successivamente, viene attuata una prima fase di tipificazione, durante la quale vengono risolti i parallelismi tra strade abbastanza lunghe e generalizzati gli incroci. Risolvere i problemi di parallelismo ha una certa priorità rispetto alla tipificazione degli incroci: osservando infatti il comportamento degli algoritmi per la tipificazione degli incroci, sviluppati da Lissandron (2008), si è notato che, nel caso in cui sono coinvolte due strade parallele, accorpare le due strade prima di risolvere l'incrocio permette di ottenere risultati migliori.

È d'obbligo sottolineare, tuttavia, che in questa fase non è conveniente risolvere tutti i parallelismi presenti nei dati: si possono sicuramente tralasciare le strade parallele di minor lunghezza, le quali potrebbero essere eliminate dal successivo processo di selezione. Ad esempio, è frequente che all'interno di un quartiere si abbiano più strade tra loro parallele e vicine: la selezione, nel caso di due strade parallele abbastanza corte, potrebbe facilmente eliminarne una, oppure entrambe. Così, sfoltire il grafo stradale prima di semplificarlo evita di compiere operazioni inutili, come generalizzare due strade parallele che poi in fase di eliminazione verrebbero scartate. Lo stesso non vale, invece, per gli incroci: eliminare alcuni segmenti stradali prima di generalizzare le giunzioni potrebbe causare una classificazione errata degli incroci da parte del relativo algoritmo. Ad esempio, se si sfoltiscono alcuni dei segmenti stradali appartenenti ad una rotonda, sarà molto più probabile nella fase di classificazione degli incroci non riconoscere la rotonda, avendo essa non più una forma caratteristica. Risolti dunque i parallelismi tra strade parallele sufficientemente lunghe, generalizzati gli incroci e sfoltito il grafo, si potrà procedere con la tipificazione di tutte le restanti strade parallele.

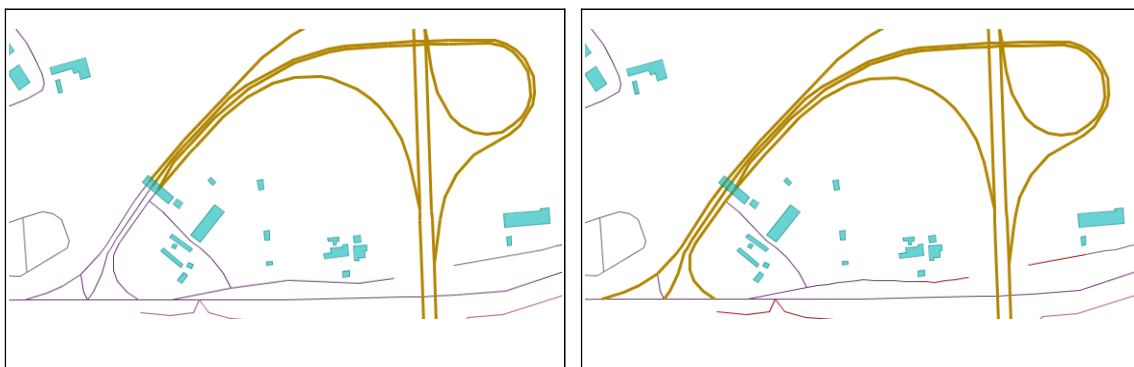


Figura 4.6: Omogeneizzazione dei segmenti stradali e autostradali, separati dal casello autostradale. A sinistra l'accesso al casello prima dell'armonizzazione, a destra le stesse strade armonizzate. L'operazione facilita e migliora la successiva tipificazione dei tratti paralleli. (cartografia di Monselice CTRN).

Per quanto riguarda la generalizzazione della rete autostradale, si è osservato che, rendendo omogenei i segmenti stradali che collegano la viabilità normale all'autostrada, si ottiene un risultato migliore nella fase di tipificazione di questo tipo di raccordi (Figura 4.6). Per questo stesso motivo, è conveniente posizionare questo algoritmo successivamente alla fase di armonizzazione.

Infine, all'ultimo passo del processo di generalizzazione, si colloca l'algoritmo sviluppato da Biasion (2010) per la tipificazione dei tornanti. Questo algoritmo troverà come dati di partenza una situazione semplificata rispetto a quelli della cartografia originale, con un grafo sfoltito e armonizzato, e non andrà quindi a modificare strade che potrebbero poi essere eliminate in fase di selezione.

I restanti capitoli descriveranno le fasi relative all'armonizzazione del grafo stradale (capitolo 5), alla selezione (capitolo 6) e alla tipificazione delle strutture parallele (capitolo 7).

In Tabella 4.2 viene riassunta la sequenza di passi del processo di generalizzazione.

PROCESSO DI GENERALIZZAZIONE: SEQUENZA DEGLI ALGORITMI	
N°	ALGORITMO
1	Risoluzione dei segmenti stradali disallineati in prossimità di un incrocio
2	Armonizzazione del grafo stradale (<i>data enrichment</i> + riclassificazione)
3	Sfoltimento del grafo stradale: stroke dangling
4	Tipificazione delle strutture parallele aventi lunghezza > 300 mt
5	Generalizzazione degli incroci stradali
6	Sfoltimento del grafo stradale: maglie + segmenti stradali dangling
7	Tipificazione delle strutture parallele, di qualsiasi lunghezza
8	Generalizzazione della rete autostradale
9	Generalizzazione dei tornanti

Tabella 4.2: elenco ordinato degli algoritmi utilizzati nel processo di generalizzazione.

5 Armonizzazione del grafo stradale

Questo capitolo tratta il tema dell'armonizzazione del grafo stradale, ovvero il processo tramite il quale si vuole rendere più omogenea la classificazione delle strade, allo scopo di far risaltare alcuni particolari e di eliminarne altri che creano più disordine che informazione nella mappa derivata. Verrà quindi introdotto il problema di come quantificare l'importanza di una strada; saranno poi analizzati gli obiettivi del processo di armonizzazione e descritti gli algoritmi sviluppati per raggiungerli.

5.1 Obiettivi del processo

Per armonizzazione del grafo stradale si intende quel processo che porta a risolvere il problema della disomogeneità dei dati, attribuendo la medesima categoria ad un determinato gruppo di strade, tra le quali sussiste una qualche relazione spaziale o semantica.

Gli obiettivi dell'armonizzazione possono essere sintetizzati nei tre seguenti punti:

- rendere il livello di dettaglio coerente con il valore di scala specificato per il processo di generalizzazione, facendo risaltare gli elementi importanti della rete stradale, e ponendo in secondo piano quelli di scarsa rilevanza;
- estendere le classi stradali esistenti, per facilitare il passaggio da un modello dati all'altro;
- aumentare la robustezza degli algoritmi in presenza di errori semantici dei dati.

Tale processo si suddivide in due sotto-processi, quello di *data enrichment* e quello di riclassificazione. La fase di *data enrichment* ha come scopo quello di estrapolare informazioni che non sono esplicitamente presenti nei dati, informazioni che successivamente vengono usate per pilotare la fase di riclassificazione. Quest'ultima fase si concretizza mediante due modalità: tramite l'estensione delle classi esistenti del modello dati, ad esempio individuando una nuova categoria di strade, come la strada di quartiere; e tramite la sostituzione dei codici, ad esempio portando una strada secondaria ad assumere il ruolo di strada principale (o viceversa).

Il processo di armonizzazione è in stretta relazione con l'aspetto grafico della mappa: per esempio, una strada declassificata verrà rappresentata graficamente con un segno meno spesso, occuperà per questo meno spazio e passerà in secondo piano rispetto ad altre strade.

Per valutare l'importanza di una strada, sia a livello globale che rispetto ad un'altra strada, sono stati individuati tre fattori: categoria stradale, lunghezza e ruolo come via di comunicazione. È da notare, tuttavia, che questo punto del processo risulta spesso molto delicato, poiché le regole o i fattori tramite cui si stabilisce quantitativamente e qualitativamente l'importanza di una strada influenzano pesantemente l'output finale.

La categoria della strada è sicuramente il primo elemento usato per distinguere il valore di una strada: un'autostrada sarà più importante di una strada principale, che a sua volta sarà più importante di una strada secondaria.

Una secondo fattore di distinzione è dato dalla lunghezza della strada: una strada lunga mezzo chilometro è molto probabilmente meno importante di una lunga dieci.

Il contesto geografico è un ulteriore elemento che caratterizza una strada: se pensiamo ad una località di montagna che è attraversata da una sola strada, quest'ultima acquista valore solo per il fatto di essere l'unica a servire il paese; invece, una strada che porta ad una fattoria sarà fondamentale per la fattoria stessa, ma a livello globale sarà considerata di scarso interesse. Una strada quindi è tanto più importante quanto è alta la sua utilità nel collegare il resto della rete.

Questi tre fattori individuati non devono per forza coesistere per determinare un buon punteggio alla strada: se questa è corta, o se è di bassa categoria, ma è l'unica a connettere due sottografi della rete, tale strada riveste comunque un ruolo rilevante nella rete; se invece non è l'unica via di comunicazione che mette in contatto due zone, ma è l'unica ad avere una categoria elevata, la strada resta comunque la più importante nell'area in cui è inserita.

Di seguito sono descritti in modo dettagliato gli obiettivi del processo di armonizzazione.

5.1.1 Risaltare gli elementi importanti e diminuire i dettagli

Quando si osserva una mappa a bassa scala, si può notare che alcuni elementi vengono messi in risalto per il ruolo che svolgono.

Questo implica, naturalmente, alcune modifiche. È così che una strada che attraversa un centro urbano manterrà lo stesso segno grafico dall'inizio alla fine, anche se in realtà tale strada all'interno della città è classificata con una categoria inferiore; in tal modo, questa strada sarà facilmente individuabile in mezzo a tutte le strade di quartiere. Cambiare la categoria di questa strada è, quindi, il presupposto per poter comunicare un'informazione migliore all'utente della carta.

Allo stesso scopo, un breve tratto stradale, che da un lato si incrocia alla strada principale e dall'altro prosegue come strada secondaria, non avrebbe motivo di essere classificato come principale, poiché ad una scala bassa questo dettaglio creerebbe solo una fastidiosa segmentazione nella mappa; viceversa, un'immagine migliore si otterrebbe prolungando la strada secondaria fino all'incrocio con la strada principale.

Se, invece, consideriamo una strada connessa ad una strada principale, che attraversa un gruppo di case e termina oltre queste, è evidente che il tratto che prosegue oltre le case ha un ruolo molto meno importante del primo tratto. Declassificare quest'ultima parte di strada, ad esempio trasformandola in sentiero, permette di evidenziare la funzione svolta dal primo tratto di strada.

Infine, ci sono strade che all'interno della rete hanno il fondamentale compito di connettere due sottografi, e che perciò meritano di essere messe in risalto. La riclassificazione di queste strade, strategiche per la comunicazione, è un'operazione necessaria ma estremamente delicata: il rischio che si corre è quello di far risaltare nella mappa strade che nella realtà sono invece di infima importanza. Questa operazione, d'altra parte, è comune nella derivazione di mappe a scale inferiori; infatti, comparando la mappa 25DB dell'IGM relativa a Monselice con i dati digitali della CTRN, si osserva la presenza di strade, rilevanti per la funzione di collegamento, che aumentano di importanza con il cambiamento di scala, da 1:5 000 a 1:25 000.

5.1.2 Estendere le classi stradali del modello dati

Il passaggio da un modello dati A ad un modello dati B risulta difficoltoso nel momento

in cui non si riescano ad associare, ad ogni classe C_i del modello B, una o più classi del modello A. Per risolvere un problema simile, l'unica soluzione diventa quella di individuare e successivamente estrarre, dalle classi del modello A, le feature che hanno caratteristiche simili a quelle di classe C_i ; questo significa identificare una particolare relazione semantica e spaziale, che risulti efficace per estrapolare tutte e sole le feature conformi alle caratteristiche della classe C_i .

In particolare, in questo lavoro di tesi, è stato posto come obiettivo definire una metodologia per individuare, tra le varie strade della mappa, quelle che possono essere associate alla classe “strada urbana di quartiere”, classe presente nel modello DB25 e nel prototipo DB50, ma assente nel modello della CTR.

5.1.3 Aumentare la robustezza degli algoritmi di generalizzazione

Analizzando i dati utilizzati per i test, relativi alla zona di Monselice e di Chioggia, si evidenziano numerose situazioni in cui le classificazioni delle strade non sono coerenti con il contesto in cui sono inserite. Ad esempio, si possono notare casi di strade che per brevi tratti cambiano classificazione, oppure strade dangling che sono connesse a strade di classificazione inferiore. Queste incongruenze determinano:

- errori in fase di ricostruzione degli stroke, che si ripercuotono poi sugli altri algoritmi di generalizzazione;
- errori nella rappresentazione grafica della mappa, che hanno come conseguenza una certa sconnessione e incoerenza percepite del grafo stradale.

Perciò, avere una classificazione delle strade più omogenea ha come effetto collaterale quello di aumentare la robustezza degli algoritmi di generalizzazione, nel caso i dati presentino errori semantici.

5.2 Algoritmi di armonizzazione

Il processo di armonizzazione sviluppato si articola in più fasi, ognuna delle quali mira a raggiungere un preciso sotto-obiettivo. Dall'esame degli obiettivi presentati nelle sezioni precedenti, si possono estrapolare le azioni da attuare nel processo, elencate di seguito:

- declassificazione degli elementi che hanno una scarsa funzione di collegamento;

- riclassificazione dei tratti stradali che vengono percepiti come un'unica strada;
- riclassificazione degli elementi che rappresentano l'unica via di comunicazione;
- correzione delle incoerenze semantiche presenti nei dati;
- classificazione delle strade di quartiere.

Nei successivi paragrafi, verranno discusse le azioni appena elencate, attuate nel processo di armonizzazione.

5.2.1 Declassificazione delle vie con scarsa funzione di collegamento

Per elementi stradali con scarsa funzione di collegamento si intendono due tipologie di strade: quelle che non svolgono alcuna funzione di collegamento, i.e. le strade dangling, e quelle che non vengono mai considerate nei percorsi minimi tra due qualsiasi punti della mappa. In realtà, il primo caso corrisponde ad un sottocaso del secondo, in quanto una strada dangling non sarà mai inclusa in un percorso minimo che non abbia la stessa strada come sorgente o destinazione.

Il metodo mediante il quale si possono individuare questo tipo di strade si basa sulla *local betweenness*, che è stata definita come (4.4.3)

$$C_{lb}(v_i) = \frac{1}{(|G_i| - 1)(|G_i| - 2)} \sum_{v_j, v_k \in G_i, j > k} \frac{p_{kj}(i)}{p_{kj}}$$

L'elemento stradale v_i ha un valore di *local betweenness* pari a zero se

$$p_{kj}(i) = 0, \forall v_k, v_j \in G_i, k \neq j,$$

ovvero se non esistono percorsi minimi tra due qualsiasi nodi che attraversano V_i . Perciò le vie con scarsa funzione di comunicazione, per come sono state definite, sono caratterizzate dall'aver un valore nullo di *local betweenness*.

L'algoritmo sviluppato per declassificare le strade con *local betweenness* uguale a zero, è specifico per le strade principali; non ha molto senso, infatti, declassificare strade che già sono considerate di scarsa importanza. Le strade principali con *local betweenness* nulla, una volta individuate, verranno declassificate alla categoria strettamente inferiore, ovvero a strade secondarie. Se, comunque, la lunghezza di questi segmenti risulta superiore ad una certa soglia (esempio 400 mt), tali segmenti non subiranno modifiche.

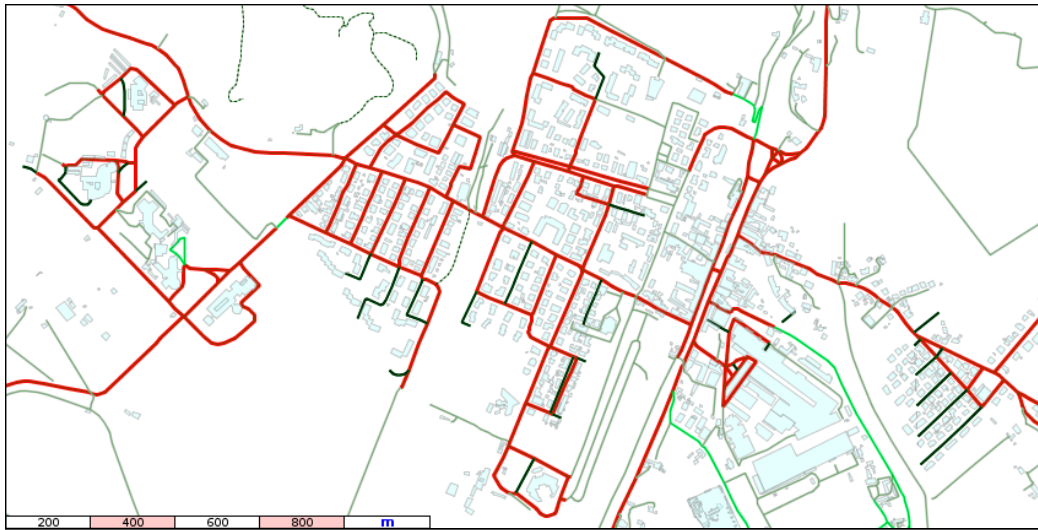


Figura 5.1: Esempio di strade principali declassificate; in rosso le strade principali, in nero le strade principali con *local betweenness* uguale a zero, declassificate a strade secondarie. Fonte: cartografia di Monselice (CTRN).

L'algoritmo che realizza questa operazione viene applicato due volte in sequenza: una prima volta sul grafo degli stroke, e una seconda volta sul grafo planare, in quanto la sola applicazione sul primo tipo di grafo presenta il difetto di non individuare alcuni dei segmenti che andrebbero declassificati.

Un ulteriore algoritmo è stato realizzato per trattare le strade dangling che forniscono l'accesso ad abitazioni, ma che però terminano in una zona senza edificato (Figura 5.2). Per quanto riguarda questo tipo di strade, è stata declassificata a sentiero la parte che non fornisce accesso ad alcun edificio.

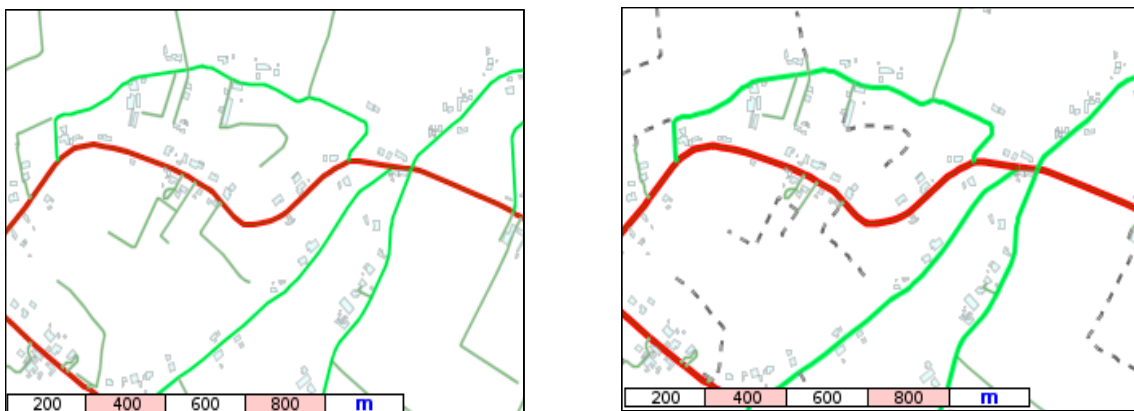


Figura 5.2: declassificazione a sentiero (linea tratteggiata) delle strade dangling che non servono alcun edificio.

5.2.2 Riclassificazione degli elementi collineari

In questa fase, i tratti stradali che vengono percepiti come un'unica strada vengono armonizzati. Questa operazione avviene tramite la riclassificazione del tratto, che acquista la categoria della strada di cui è il naturale proseguimento. Cambiando la categoria, la strada può aumentare di importanza, ad esempio passare da strada secondaria a strada principale, oppure essere declassificata. È stato scelto di separare i due casi, prima tentando di riclassificare le strade verso categorie superiori, e successivamente verso categorie inferiori.

Il grafo utilizzato per la riclassificazione degli elementi collineari è quello degli stroke. Questa scelta è d'obbligo se non si vuole correre il rischio di riclassificare un segmento stradale A che geometricamente è il migliore proseguimento di una certa strada S, ma che in realtà rappresenta solo una laterale di S (ad esempio se S presenta una curvatura in prossimità dell'intersezione con A).

Dall'analisi dei dati sono state evidenziate due situazioni ove intervenire con questo tipo di armonizzazione, che hanno portato alla definizione di due algoritmi.

Il primo algoritmo ha l'obiettivo di individuare una sequenza di tre strade allineate, S_1 , S_2 e S_3 , con la condizione che la strada in posizione centrale abbia una classificazione diversa rispetto alle altre due e che sia di lunghezza inferiore ad una soglia massima prestabilita (ad esempio 1000 mt); si attribuirà, quindi, la stessa categoria alle tre strade. Per ogni strada S_2 , di classe C_a , si trovano, se esistono, le due strade S_1 e S_3 che rappresentano il proseguimento naturale di S_2 . Una volta trovate S_1 e S_3 , si verifica che esse abbiano la stessa classificazione C_b ; in caso affermativo, a seconda della relazione che sussiste tra C_a e C_b , vengono distinti due casi:

- $C_a < C_b$, ovvero si deve aumentare la classe di S_2 ; in questo caso, S_2 assume la categoria C_b ;
- $C_a > C_b$, ovvero si deve declassificare la stroke S_2 ; in questo caso, è necessario verificare che S_2 non sia connesso a due (o più) strade, S_3 e S_4 , di categoria superiore a C_b , con S_3 e S_4 non direttamente connesse. Infatti, se ciò dovesse capitare, declassificando la strada S_1 il collegamento tra S_3 e S_4 avverrebbe, nella nuova situazione, per mezzo di una strada di categoria inferiore rispetto alla

situazione precedente.

Tramite questa procedura vengono comunemente escluse le strade di categoria C_a che si trovano tra due strade di categoria rispettivamente C_b e C_c . Nell'algoritmo appena descritto, invece, si è deciso di inglobare anche questo caso, con il vincolo $C_a < C_b, C_c$ e limitando la lunghezza della strada da riclassificare ad un valore decisamente inferiore (200 mt) per non rischiare di effettuare modifiche dove non sia strettamente necessario. In questo caso, adottando la nomenclatura utilizzata precedentemente, S_2 assumerà la categoria più bassa tra le due categorie C_b e C_c .

In Figura 5.3 è mostrato l'effetto di questo algoritmo sul centro di Monselice.

Il secondo algoritmo sviluppato tratta quei tipi di strade che, partendo da un incrocio a T, dopo poche centinaia di metri cambiano classificazione. L'algoritmo, per ogni strada S_2 che sia di classe C_a e che inizi da un incrocio a T, individua, se esistono, la strada S_1 di categoria C_a che attraversa l'incrocio, e la strada S_3 di categoria C_b , naturale proseguimento di S_2 .

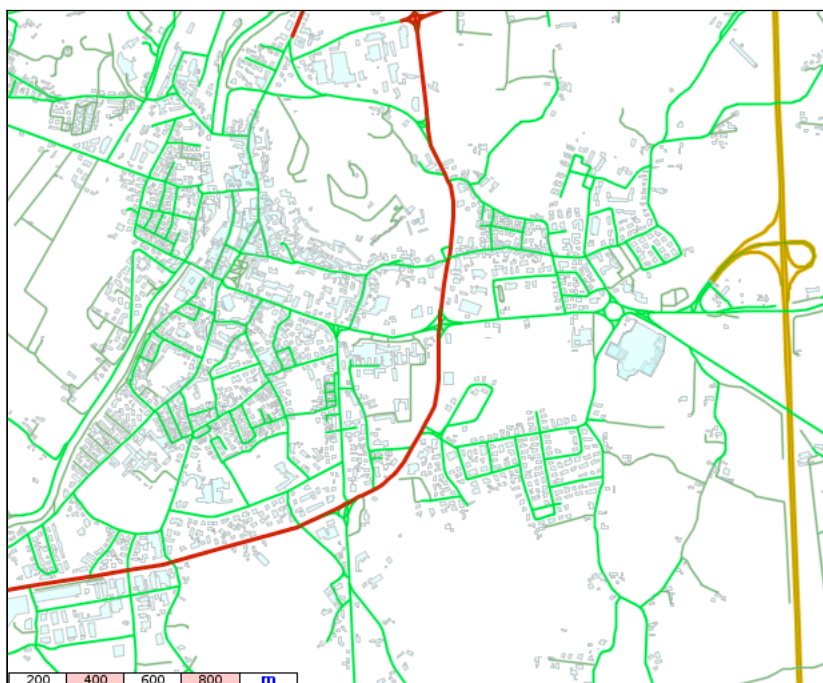


Figura 5.3: riclassificazione della strada secondaria passante per Monselice da secondaria a principale. Si confronti con la situazione iniziale di Figura 4.1.

Anche nel caso di questo algoritmo si distinguono due situazioni, a seconda della relazione che sussiste tra C_a e C_b :

- $C_a < C_b$, ovvero si deve aumentare la classe di S_2 ; in questo caso, S_2 assume la categoria C_b ;
- $C_a > C_b$, ovvero si deve declassificare la stroke S_2 ; in questo caso è necessario effettuare un controllo sulle altre strade connesse a S_2 . Se esiste una strada S_4 , di classe superiore a C_b , connessa a S_2 , ma non a S_1 , la declassificazione di S_2 non viene effettuata allo scopo di mantenere alto il collegamento tra S_4 e S_1 .

In Figura 4.6 è mostrata un'applicazione di questo algoritmo, in cui le strade coinvolte in un incrocio a T vengono riclassificate.

5.2.3 Riclassificazione delle vie di collegamento strategiche

Una via di comunicazione è una strada che collega due zone geografiche. Quando questa via è l'unica ad espletare questa funzione, essa assume un ruolo centrale per le zone che mette in collegamento, e merita dunque di essere enfatizzata.

Una strada S_1 di classificazione C_a , per essere riclassificata in base al suo ruolo di via di comunicazione strategica, deve esibire le due seguenti caratteristiche:

1. avere gli estremi connessi a due strade S_2 e S_3 di categoria superiore;
2. essere l'unica strada che permette alle stesse strade di comunicare. Più precisamente, non deve esistere nessun percorso, formato da sole strade di categoria superiore a C_a , che metta in collegamento S_2 a S_3 .

L'algoritmo sviluppato per risolvere questo problema di riclassificazione tratta singolarmente le categorie della viabilità stradale. Nell'ordine, si riclassificano le vie di comunicazione di categoria bassa, per poi esaminare quelle di categoria superiore. In generale, per individuare le strade candidate ad essere riclassificate, si fa uso della *local betweenness*.

L'obiettivo dell'algoritmo è quello di ottenere un sottografo connesso della viabilità principale, in modo che non ci siano, ad esempio, due grossi sottografi di strade principali connessi da una o più strade di livello inferiore.

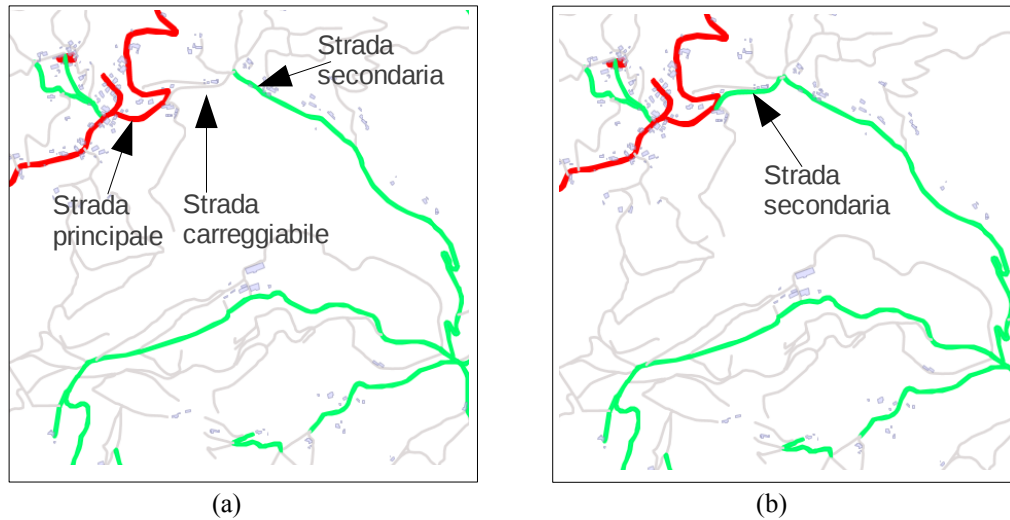


Figura 5.4: Esempio di riclassificazione delle vie di collegamento strategiche. La strada carreggiabile (a) viene riclassificata come strada secondaria (b); essa rappresenta l'unica strada che mette in comunicazione due strade di categoria superiore.

Per descrivere come opera l'algoritmo, si consideri il caso di un grafo stradale planare G con due classi di strade, C_a e C_b , con C_a superiore a C_b . Si vogliono individuare le strade di classe C_b che hanno un ruolo chiave nel mantenere connesso il sottografo G_a , costituito dalle strade di classe C_a . Le fasi dell'algoritmo sono le seguenti:

1. prima selezione

Per ogni elemento s_b appartenente a G_b , si controlla se esiste un segmento s_a appartenente a G_a , connesso a s_b , per il quale vale la relazione $C_{lb}(s_a) < C_{lb}(s_b)$. In caso positivo, a partire dal punto estremo p di s_b non connesso a S_a , si trova, se esiste, una sequenza di segmenti di classe C_b che porta ad un segmento di classe C_a . Il primo elemento della sequenza è costituito da s_b ; se p è già connesso ad un segmento di G_a , la ricerca termina; altrimenti, viene impostato come secondo elemento della sequenza il segmento connesso a p con valore di *local betweenness* maggiore. La costruzione della sequenza termina quando si raggiunge un segmento di classe C_a , oppure quando la lunghezza totale dei segmenti individuati supera una soglia massima prefissata.

2. Scematura

Una volta individuate le sequenze di segmenti candidate ad essere riclassificate,

bisogna verificare che tali sequenze svolgano veramente una funzione di collegamento tra due sottografi di G_a . A tale scopo, per ogni sequenza di segmenti che ha come estremi i punti p_1 e p_2 , si calcola la lunghezza del percorso da p_1 a p_2 nel sottografo G_a . Nel caso in cui questa lunghezza dovesse risultare infinita, o maggiore di una lunghezza limite (ad esempio 10 km), i segmenti della sequenza verrebbero riclassificati con classe C_a e inseriti nel sottografo G_a .

In generale, in presenza di più classi, il ruolo che aveva C_a nella descrizione precedente sarà rappresentato da tutte le classi superiori a C_b . Per la classe da associare alla sequenza di segmenti individuata, viene scelta, tra tutte le classi dei segmenti connessi agli estremi della sequenza, quella di importanza minore.

In Figura 5.4 è mostrato l'effetto di questo algoritmo sulla zona di Monselice.

5.2.4 Riclassificazione delle strade inserite in contesti anomali

Il caso in cui una strada S_1 , di classe C_a , inserita in un contesto anomalo dove non esiste nessun'altra strada connessa a S_1 di categoria uguale o superiore a C_a , è una situazione che normalmente non si manifesta in una mappa geografica. Se S_1 è la strada con la categoria più alta tra tutte le strade a cui è connessa, è molto probabile che una simile strada sia conseguenza di un errore dei dati. Le strade inserite in contesti anomali rappresentano, dunque, un caso limite.

A prescindere, comunque, dalla presenza o meno di errori nella classificazione di S_1 , una strada di questo genere non ha motivo di mantenere una classificazione superiore rispetto alle strade che si trovano vicino. Risulta sicuramente più opportuno declassificarla, assegnandole la classe della strada di categoria più elevata a cui si connette. Si noti, inoltre, che la declassificazione viene applicata solamente quando la lunghezza della strada è inferiore ad una soglia prestabilita L_{max} .

L'algoritmo sviluppato per trattare questo genere di strade "anomale" agisce anche su un insieme di strade connesse di stessa classificazione che si trovano isolate. Considerando un gruppo di strade, allora, se la somma delle loro lunghezze è inferiore ad L_{max} , il gruppo di strade verrà declassificato.



Figura 5.5: Esempio di declassificazione delle strade inserite in contesti anomali. A sinistra le strade originali; a destra le strade declassificate.

Un esempio del risultato prodotto da questo algoritmo è riportato in Figura 5.5.

5.2.5 Classificazione delle strade di quartiere

Individuare le strade di quartiere è un'operazione importante, soprattutto nel contesto di questa attività di tesi, dato che i modelli dati di destinazione (il DB25 dell'IGM e il prototipo DB50) includono questa tipologia di strada.

Le caratteristiche di una strada di quartiere possono essere individuate analizzando una carta della serie 25DB dell'IGM. Si possono osservare in particolare tre elementi che

aiutano ad identificare una strada di quartiere:

- la lunghezza: una strada di quartiere presenta infatti una lunghezza abbastanza contenuta (< 800 mt);
- scarse funzioni tipiche di una via di comunicazione: una strada di quartiere, per definizione, non è un'importante via di comunicazione per il resto del grafo stradale;
- numero di edifici serviti: una strada di quartiere serve un buon numero di abitazioni.

Ai fini operativi, si può testare numericamente la prima ipotesi semplicemente definendo una lunghezza L_{\max} che una strada di quartiere può avere. La seconda osservazione, invece, può essere quantificata ponendo un limite superiore C_{lb_max} al valore di *local betweenness* della strada. Infine, per tradurre numericamente in che misura una strada è di servizio ad edifici, si può procedere nel seguente modo: si identificano tutti gli edifici a distanza massima d_{\max} dalla strada; se il rapporto R tra l'area totale occupata da questi edifici e l'area del buffer della strada, di dimensione d_{\max} , è maggiore di una soglia specificata R_{\min} , la strada viene considerata come una strada che serve un buon numero di abitazioni.

Una volta quantificate le osservazioni sulle caratteristiche tipiche che presenta una strada di quartiere, è possibile realizzare una procedura sistematica per identificare tutte quelle strade che soddisfano le tre ipotesi precedenti.

L'algoritmo sviluppato si suddivide in due fasi, di cui la prima mira a definire un insieme iniziale di strade di quartiere, e la seconda ad aggiungere a questo insieme ulteriori segmenti stradali che soddisfano particolari vincoli.

Fase 1: prima selezione

In questa fase, considerando il grafo degli stroke, vengono identificate come strade di quartiere tutte le strade per cui valgono contemporaneamente le tre seguenti condizioni: la lunghezza della strada è inferiore a L_{\max} ; il rapporto R tra l'area degli edifici limitrofi e l'area del buffer della strada soddisfa la disuguaglianza $R > R_{\min}$; e il valore di *local betweenness* C_{lb} a sua volta soddisfa la disuguaglianza $C_{lb} < C_{lb_max}$.

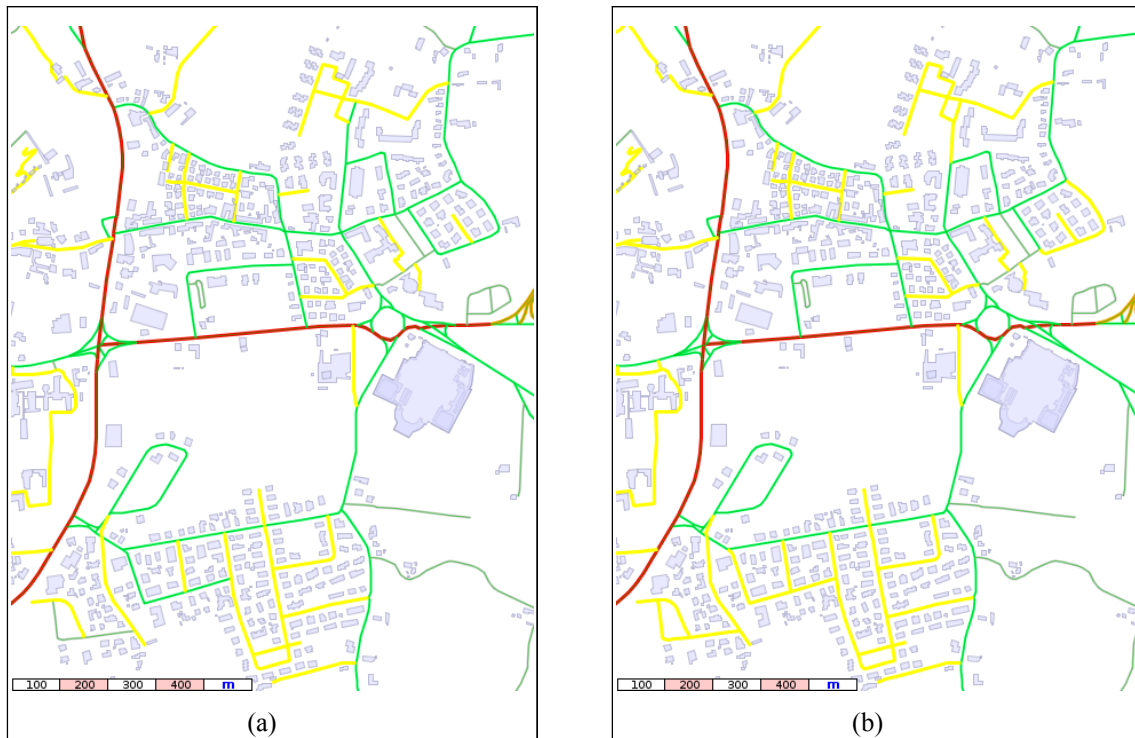


Figura 5.6: Classificazione delle strade di quartiere, indicate con colore giallo, dopo la prima fase (a) e dopo la seconda (b).

Fase 2: seconda selezione

In questa fase viene considerato il grafo planare delle strade. Non sarebbe possibile né conveniente, infatti, basare la ricerca delle strade di quartiere solo sugli stroke, poiché ciò non permette di individuare alcuni tratti stradali che in realtà soddisfano le condizioni definite precedentemente; in particolare, se uno stroke offre accesso a numerose abitazioni, ma ha una lunghezza troppo elevata, l'intera strada viene scartata. Processare i singoli tratti stradali, al contrario, permette di scavalcare questo problema. Va sottolineato, comunque, che la ricerca delle strade di quartiere applicata al solo grafo planare non porta a buoni risultati, in quanto, essendo la lunghezza di un segmento molto inferiore rispetto a quella dello stroke, molti più segmenti soddisfaranno i tre vincoli discussi in precedenza. Il grafo planare, perciò, va trattato con molta cautela.

Utilizzando il grafo planare si eseguono i seguenti passi per ogni segmento stradale S :

- a) si trovano gli insiemi I_1 , che rappresenta i segmenti connessi all'estremo iniziale di S e I_2 , che rappresenta i segmenti connessi all'estremo finale di S ;
- b) se I_1 o I_2 contengono solo segmenti di quartiere e $C_{lb}(S) < C_{lb_max}$, si classifica S

come strada di quartiere;

- c) se I_1 e I_2 contengono, oltre che segmenti di strade di quartiere, anche altri tipi di strade, se $C_{lb}(s) < 0.001$ e se $R < R_{min}$, S viene classificata come strada di quartiere. Imporre il vincolo sul valore di $C_{lb}(S)$ molto stringente può essere importante poiché garantisce di non cadere nell'errore di classificare come strada di quartiere una strada che, pur essendo connessa ad altre strade di quartiere, svolge in realtà un ruolo differente in quella particolare zona.

In Figura 5.6 sono mostrati i risultati delle due fasi appena descritte.

5.3 Il processo di armonizzazione

Dopo aver esposto gli algoritmi di armonizzazione e analizzato come essi agiscono, è ora possibile definire il processo di armonizzazione, costituito dall'ordine con cui vengono eseguiti gli algoritmi esaminati. La successione degli algoritmi e la loro identificazione tramite progressivi numerici sono riportate in Tabella 5.1.

In prima posizione è stato inserito l'algoritmo per declassificare le strade inserite in contesti anomali; l'applicazione di questo algoritmo avviene in modo limitato, con lo scopo di eliminare piccoli errori di classificazione che possono condizionare negativamente i successivi algoritmi. Lo stesso algoritmo viene nuovamente applicato in forma più forte dopo la declassificazione delle strade principali con scarsa funzione di comunicazione (2) e la riclassificazione degli elementi collineari (3), allo scopo di eliminare le situazioni anomale rimaste. Successivamente, vengono riclassificati i tratti stradali fondamentali per la comunicazione (5) e viene riapplicato l'algoritmo per la riclassificazione degli elementi collineari (6); questa ripetuta esecuzione è dovuta al fatto che l'algoritmo di riclassificazione delle importanti vie di collegamento potrebbe aver generato nuovi casi da trattare per l'algoritmo di riclassificazione degli elementi collineari. In settima posizione vi è l'algoritmo di declassificazione in sentiero delle strade dangling che non servono alcun edificio. Infine (8) viene eseguito l'algoritmo per la classificazione delle strade di quartiere, che avrà come input stroke più lunghi (fattore determinante per l'algoritmo) rispetto ai dati originali, grazie, in particolare, agli algoritmi per la declassificazione delle strade in contesti anomali e per la

riclassificazione degli elementi collineari.

PROCESSO DI ARMONIZZAZIONE: SEQUENZA DEGLI ALGORITMI	
N°	ALGORITMO
1	Declassificazione strade in contesti anomali, con $L_{\max} = 150$ mt
2	Declassificazione strade principali con <i>local betweenness</i> = 0
3	Riclassificazione degli elementi collineari
4	Declassificazione strade in contesti anomali, con $L_{\max} = 1000$ mt
5	Riclassificazione delle importanti vie di collegamento
6	Riclassificazione degli elementi collineari
7	Declassificazione in sentiero delle strade dangling che non servono alcun edificio
8	Classificazione delle strade di quartiere

Tabella 5.1: elenco ordinato degli algoritmi utilizzati nel processo di armonizzazione.

6 Sfoltimento del grafo stradale

Questo capitolo tratta il processo di selezione degli elementi del grafo stradale, operazione necessaria e fondamentale quando si attua una riduzione di scala su una cartografia preesistente. Sarà descritto il problema dello sfoltimento stradale suddividendolo in due sotto-problemi, il primo relativo alle strade *dangling* e il secondo relativo alle maglie. Sarà infine definito il processo di selezione, individuando la sequenza di esecuzione degli algoritmi presentati.

6.1 Introduzione

Con il passaggio ad una scala inferiore, gli elementi del grafo stradale, naturalmente, si ritroveranno nella carta derivata molto più vicini tra loro rispetto alla situazione di partenza. Questo effetto di collasso dello spazio ha come risultato quello di rendere la mappa molto meno leggibile.

Non è pensabile, quindi, rappresentare nella carta derivata ogni singolo segmento stradale presente nella cartografia di partenza, soprattutto nel caso di zone dove la densità delle strade è molto elevata. La selezione delle strade da mantenere, o similmente di quelle da eliminare, risulta essere dunque un'azione necessaria che, se mal compiuta, può compromettere la qualità dell'informazione veicolata dalla mappa. È importante, perciò, guidare l'azione di selezione secondo principi che garantiscano il mantenimento della struttura portante della rete stradale: ovvero, devono essere mantenute la connettività e la forma caratteristica della rete, rispettando l'ordine gerarchico definito dalla classificazione delle strade.

Il problema dello sfoltimento stradale richiede, innanzitutto, l'individuazione delle caratteristiche che una strada deve esibire per poter essere eliminata. È conveniente, a tale scopo, suddividere l'insieme delle strade in due categorie: nella prima categoria rientrano tutte quelle strade che sono connesse al grafo stradale tramite un unico punto, ovvero le strade *dangling*; nella seconda categoria, invece, tutte le restanti. Risulta opportuno applicare per le strade di queste due categorie strategie di selezione diversificate ai fini della selezione, poiché esse presentano caratteristiche intrinseche molto differenti.

Così, nella selezione delle strade *dangling*, i parametri che vengono reputati importanti ai fini di un'eventuale eliminazione sono: la lunghezza della strada, la classe di appartenenza, e il fatto di servire o meno edifici di vario genere.

Per le altre strade, invece, e più precisamente per le strade che fanno parte di una maglia, il problema è molto più complesso e necessita in primis di un'analisi del contesto.

Nelle due sezioni seguenti verrà analizzato il problema della selezione strade più in dettaglio, differenziandolo secondo le due categorie appena introdotte.

6.2 La selezione delle strade *dangling*

Secondo quanto specificato dall'IGM (2004), i limiti di acquisizione per le strade *dangling* alla scala 1:50 000 sono definiti in base alla lunghezza della strada e alla tipologia. Per quanto riguarda la viabilità principale, risultano eliminabili i tratti di lunghezza inferiore a 200 metri. Per la viabilità secondaria, invece, risultano eliminabili le carrarecce di lunghezza inferiore a 400 metri e le mulattiere e i sentieri lunghi meno di 600 metri.

La differenza, rispetto alla selezione delle strade relativa alla scala 1:25 000, è che non vengono mantenute le strade *dangling* di lunghezza inferiore ai limiti di acquisizione, anche se sono le uniche a servire un gruppo di edifici.

Essendo la lunghezza il parametro principale usato nella selezione delle strade *dangling*, la scelta di quale tipo di grafo adottare, se planare o degli *stroke*, ha un ruolo determinante sul risultato finale.

Utilizzare il grafo planare offre la possibilità di eliminare tutti i segmenti inferiori alla soglia di acquisizione; lo svantaggio di questa scelta, però, è legato al fatto che, una volta eliminati i *dangling*, segmenti che prima erano connessi ad uno di questi diventano a loro volta *dangling*. Se i nuovi *dangling* sono sotto i limiti di acquisizione, vengono anch'essi eliminati generando nuovi *dangling*; questo effetto ricorsivo, dunque, porta all'eliminazione di strade sufficientemente lunghe, ma che hanno connesse una o più strade laterali. In Figura 6.1 (a), (b) e (c) viene presentato un esempio di questo fenomeno.

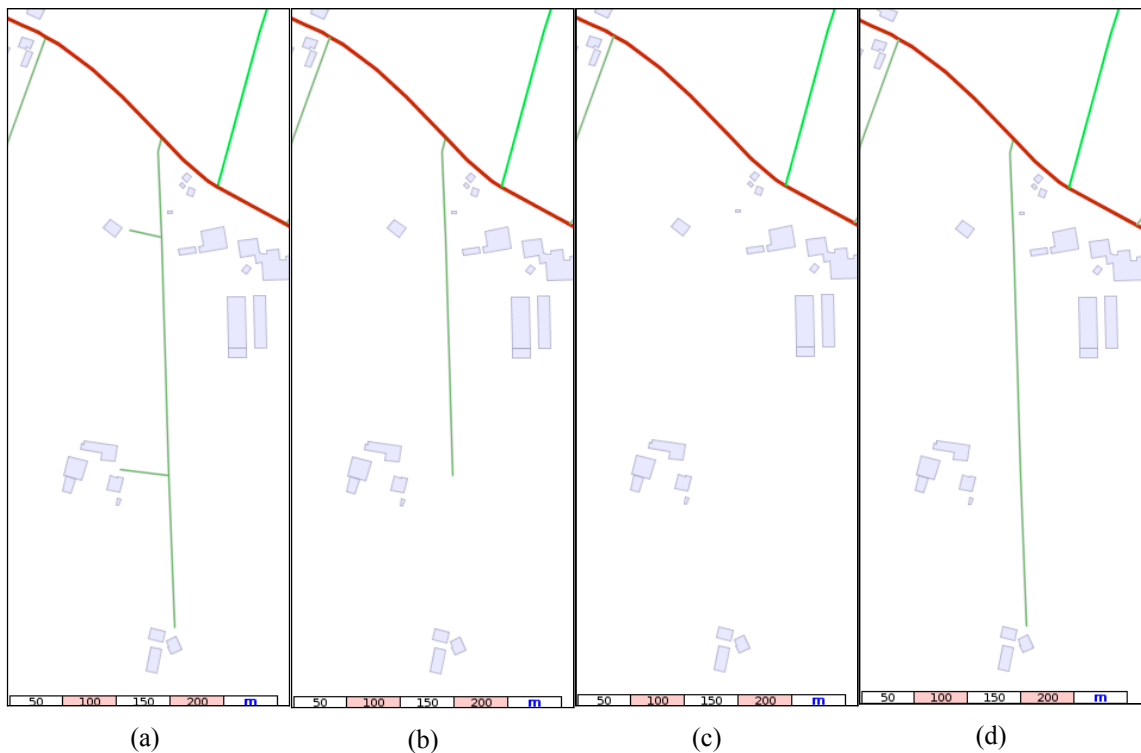


Figura 6.1: (a), (b) e (c) mostrano in successione l'eliminazione ricorsiva dei segmenti dangling sotto i limiti di acquisizione. Le strade dell'esempio sono classificate come strade di campagna; si noti l'eliminazione della strada più lunga (450 metri), che però soddisfa i limiti di acquisizione (400 metri). (d) illustra invece l'eliminazione ricorsiva degli stroke dangling; in questo caso la strada di 450 metri viene mantenuta.

Per risolvere il problema che si manifesta nell'esempio di Figura 6.1 è necessario utilizzare il grafo degli stroke. Mediante l'uso di questo secondo tipo di grafo, infatti, si evita il progressivo smantellamento della strada S (Figura 6.1 (d)).

L'algoritmo sviluppato si scompone in due fasi, che vengono sequenzialmente eseguite fino a quando non si verificano più ulteriori eliminazioni:

1. individuazione ed eliminazione di tutti gli stroke dangling sotto i limiti di acquisizione: in questa fase viene considerato dangling ogni stroke connesso al massimo con un solo punto al resto della rete, che può comunque differire dal punto iniziale o finale dello stroke; questa fase viene reiterata fino a quando non ci sono più nuovi stroke dangling da eliminare;
2. riclassificazione delle sezioni finali degli stroke dangling che non servono alcun

edificio: in questa seconda fase vengono prodotti nuovi stroke dangling, classificati come sentieri, che verranno poi analizzati nella successiva iterazione della prima fase.

Una volta terminata l'eliminazione ricorsiva degli stroke, restano ancora i segmenti dangling del grafo planare che non soddisfano i limiti di acquisizione. L'eliminazione di questi mediante il procedimento ricorsivo appena presentato, però, ripropone nuovamente il problema dello smantellamento delle strade. Da un'analisi del comportamento dell'algorithm, se applicato sul grafo planare, si è osservato che le cause che provocano l'eliminazione parziale di uno stroke sono gli anelli e le maglie, che vengono trattati nella sezione successiva. La semplificazione degli anelli e delle maglie, effettuata in via preventiva all'eliminazione dei segmenti dangling del grafo planare, può migliorare il processo di selezione, evitando, dove possibile, l'eliminazione parziale delle strade dangling (Figura 6.2).

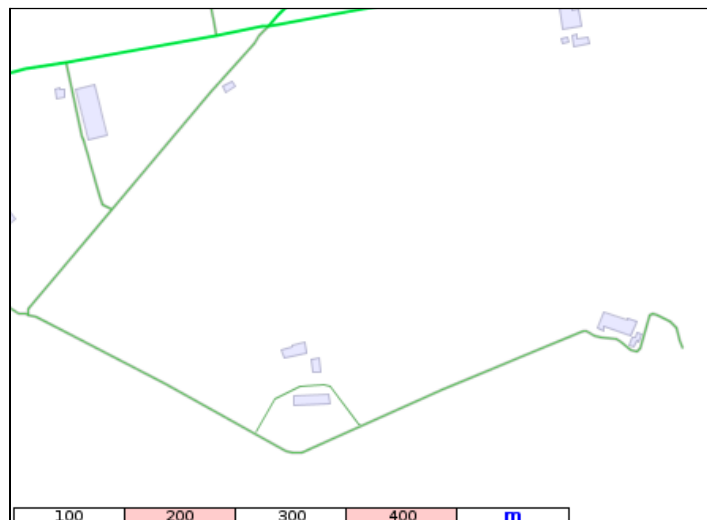


Figura 6.2: esempio di situazione in cui la presenza di un anello produce un segmento dangling sotto i limiti di acquisizione. La preventiva eliminazione dell'anello permette al segmento dangling di superare i vincoli di acquisizione.

6.3 La selezione delle maglie

Le strade che attraversano una o più maglie, e che non ricadono nella categoria delle strade dangling, rappresentano il fenomeno più critico all'interno del processo di

selezione.

Per quanto concerne la viabilità principale e secondaria, l'IGM scrive che

“sono opportunamente sfoltiti i tratti di lunghezza maggiore di quella indicata quando la trama viaria è fitta, oppure quando particolari esigenze di rappresentazione grafica lo richiedano” (IGM, 2004).

Invece, in merito alla viabilità all'interno degli agglomerati urbani, scrive:

“vengono mantenuti tutti gli attraversamenti principali mentre possono essere sfoltite, seppure in modo contenuto, le strade secondarie. Tuttavia, quando quest'ultime formano una maglia particolarmente fitta, tale da condizionare in modo marcato il disegno dei fabbricati all'interno degli isolati delimitati, si procede ad uno sfoltimento più consistente mediante l'eliminazione di ulteriori tratti di limitata lunghezza grafica o di scarsa utilità alla funzione di collegamento”(IGM, 2004).

Questo significa che, nei centri urbani, lo sfoltimento interessa principalmente le strade di quartiere. Individuare, quindi, questo tipo di strade dà la possibilità di sfruttare un'importante informazione aggiuntiva per condurre al meglio il processo di selezione della viabilità urbana. Per quanto riguarda la restante viabilità, la scelta delle strade da mantenere si dovrà totalmente basare sul loro livello di utilità per la funzione di collegamento.

6.3.1 Sfoltimento della viabilità negli agglomerati urbani

Una volta individuate le strade di quartiere mediante l'algoritmo presentato nel paragrafo 5.2.5, è possibile procedere alla sfoltimento della viabilità all'interno degli agglomerati urbani.

Grazie al *data enrichment*, che introduce una nuova classe di strade, si può suddividere in due momenti questa fase di eliminazione. Per prima cosa, verranno sfoltite solamente le strade classificate come strade di quartiere, e in un secondo momento tutte le restanti. L'eliminazione anticipata delle strade di quartiere, avendo queste per definizione¹ un ruolo irrilevante alla funzione di collegamento, garantisce che, in zone ad alta densità viaria, non vengano eliminate erroneamente strade che in realtà svolgono un ruolo importante all'interno dell'agglomerato urbano.

¹ Si veda paragrafo 5.2.5.

L'algoritmo per l'eliminazione delle strade di quartiere si basa sull'analisi delle maglie del grafo stradale planare. Una maglia è rappresentata da un poligono, che ha come lati due o più segmenti stradali. L'area della maglia costituisce l'informazione chiave per l'eliminazione delle strade: più l'area sarà piccola, e più le strade saranno graficamente vicine. Quando ciò si verifica, per le ragioni discusse all'inizio del capitolo, almeno una di queste strade dovrà essere eliminata, così da garantire una rappresentazione chiara delle restanti strade della maglia.

L'eliminazione procede, quindi, identificando tutte le maglie di area inferiore ad una soglia prefissata. Per ognuna di queste maglie si individua, se presente, il segmento stradale, classificato come strada di quartiere, che più si presta ad essere eliminato. Se nella maglia M_i vi fosse un unico segmento di quartiere, esso verrebbe dunque eliminato. In caso contrario, una prima selezione dei candidati ad essere eliminati avviene secondo la seguente regola:

- a) se esistono segmenti che non hanno un naturale proseguimento né da un estremo né dall'altro, ovvero sono di fatto degli stroke, questi vengono selezionati come candidati;
- b) come prima alternativa, se esistono segmenti che rappresentano l'elemento iniziale o finale del rispettivo stroke, questi vengono selezionati come candidati;
- c) come seconda alternativa, tutti i segmenti di quartiere vengono selezionati come candidati all'eliminazione.

Una volta individuati i candidati, si individua, tra tutte le maglie direttamente connesse alla maglia, M_i quella di area minore, che ha come lato in comune con M_i uno dei candidati; quest'ultimo lato sarà il segmento che verrà eliminato.

In Figura 6.3 (b) (c) sono mostrati gli effetti di questo algoritmo.

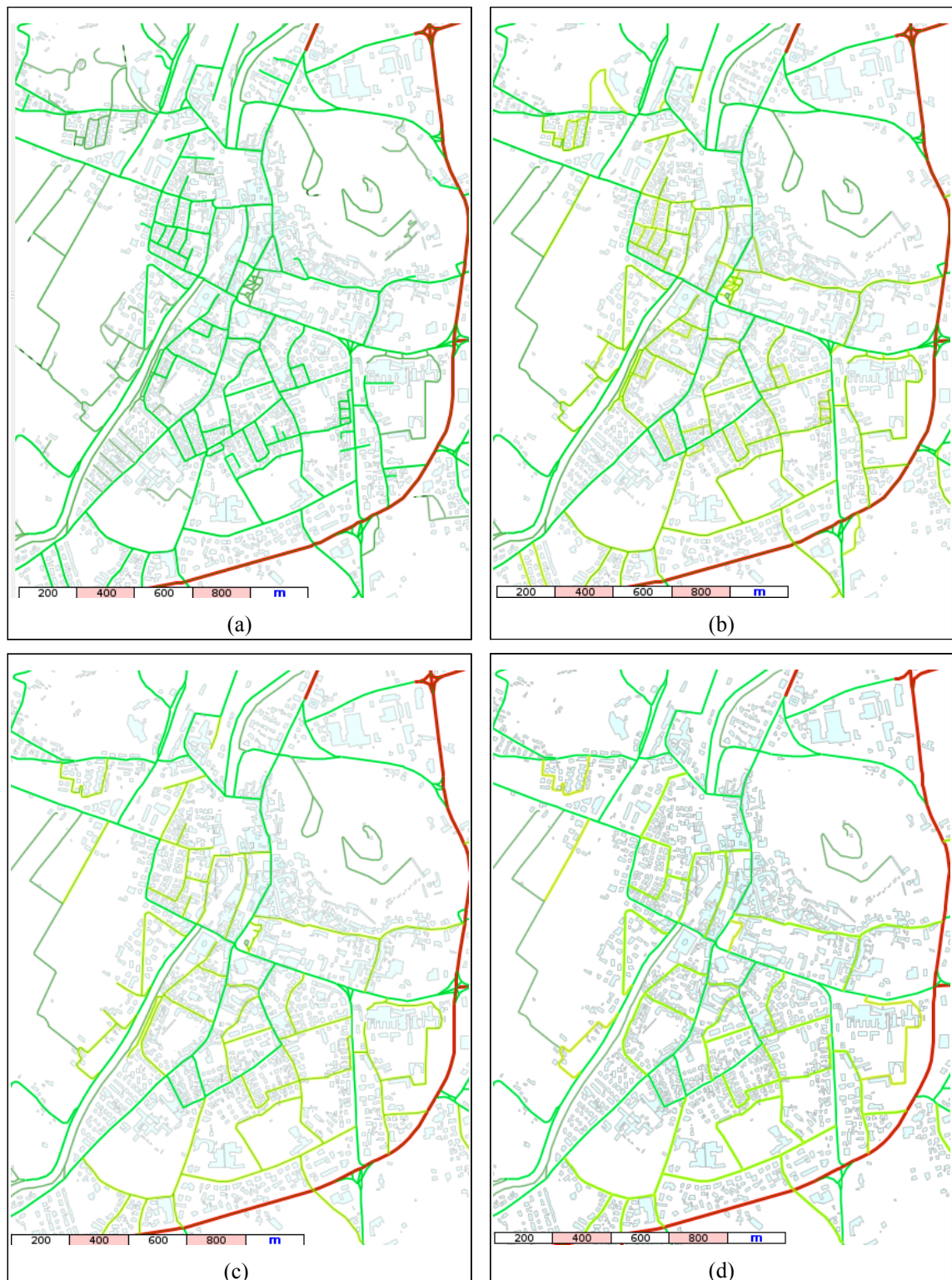


Figura 6.3: (a) situazione iniziale; (b) eliminazione degli stroke dangling e classificazione delle strade di quartiere (in giallo); (c) sfoltimento delle strade di quartiere; (d) sfoltimento maglie alta viabilità ed eliminazione segmenti dangling.

6.3.2 Sfoltimento della viabilità principale e secondaria

Terminato lo sfoltimento della viabilità di quartiere, è possibile procedere con l'analisi della restante viabilità.

A differenza dell'eliminazione di una strada di quartiere, che non genera ripercussioni negative sulla struttura della rete per definizione, l'eliminazione di una strada della viabilità principale o secondaria potrebbe cambiare di molto l'armonia e l'aspetto del grafo stradale. Il punto delicato dello sfoltimento di queste strade è che non sempre l'eliminazione di una strada che concorre con altre strade per lo stesso spazio è la soluzione corretta: ci sono situazioni complesse, infatti, in cui non basta eliminare una strada, ma è necessario ridisegnare le geometrie. È il caso, questo, delle strade parallele e degli incroci, che non possono essere risolti mediante una semplice eliminazione di segmenti stradali.

Oltre alle difficoltà nell'individuazione della strada candidata ad essere eliminata, per questo tipo di strade si aggiunge l'ulteriore problema di capire se l'eliminazione sia veramente l'azione corretta. È necessario, perciò, definire un metodo che sia in grado di distinguere una strada che può essere eliminata da una che invece deve essere ridisegnata, per evitare di alterare negativamente la struttura del grafo stradale.

L'algoritmo per lo sfoltimento delle strade di quartiere, analizzato nel paragrafo precedente, ha il difetto di non verificare le eventuali ripercussioni negative derivanti dall'eliminazione della strada, e dunque il suo utilizzo per lo sfoltimento della restante viabilità andrebbe bene solamente per le strade candidate che soddisfano il punto a); se invece le strade candidate sono quelle del punto b) o c), l'eliminazione di una di esse può provocare l'interruzione di una strada che potrebbe avere un ruolo importante nella rete. Per evitare di eliminare erroneamente questo tipo di strade, è stato realizzato un altro algoritmo, sempre basato sull'area delle maglie, che opera con la seguente sequenza:

1. per ogni maglia di area inferiore ad una soglia massima viene individuata la categoria più bassa C_a , tra le categorie di tutti i lati che formano la maglia; l'insieme dei candidati iniziali S sarà formato da tutti i lati della maglia di classe C_a ;

2. tra i candidati S , si calcolano i sottoinsiemi S_a , S_b e S_c che soddisfano rispettivamente i vincoli del punto a), b) e c) dell' algoritmo descritto precedentemente;
3. se $S_a \neq \emptyset$, viene eliminato il lato di S_a che ha il più basso valore di *local betweenness*;
4. altrimenti, se $S_b \neq \emptyset$, viene individuato il lato s di S_b che ha il più basso valore di *local betweenness*; s viene eliminato se e solo se $C_{lb}(S) < C_{lb_max}$, dove C_{lb_max} è una soglia massima prefissata (0.001);
5. in alternativa, se $S_c \neq \emptyset$, viene individuato il lato s di S_c che ha il più basso valore di *local betweenness*; s viene eliminato se e solo se $C_{lb}(S) < C_{lb_max}$.

Nell'algoritmo appena descritto, le situazioni critiche di incertezza sono quelle relative ai punti 4) e 5), che identificano quei casi in cui il segmento della maglia da eliminare fa parte di uno stroke che è formato da altri segmenti. Per questi due casi, ponendo un limite superiore al valore di *local betweenness*, si riesce ad evitare di eliminare quei segmenti che hanno un ruolo critico nel grafo stradale, e allo stesso tempo individuare alcuni segmenti che invece hanno tutte i requisiti per essere eliminati (in particolare le strade che formano anelli).

In Figura 6.3 (d) è presentato l'effetto dell'algoritmo nel centro di Monselice, mentre in Figura 6.4 viene mostrato un caso di semplificazione che coinvolge maglie aventi come lati solo parti di stroke.

6.4 Il processo di selezione

Come evidenziato nelle sezioni precedenti, il problema della selezione è suddivisibile in due sotto-problemi: quello che riguarda lo sfoltimento delle strade dangling e quello relativo allo sfoltimento delle maglie.

È stato inoltre osservato che utilizzare il grafo planare o quello degli stroke per l'eliminazione dei *dangling* porta a risultati diversi. L'uso del grafo planare produce uno sfoltimento molto più accentuato, mentre l'uso degli stroke limita l'eliminazione ai soli stroke che sono costituiti da un unico segmento stradale.

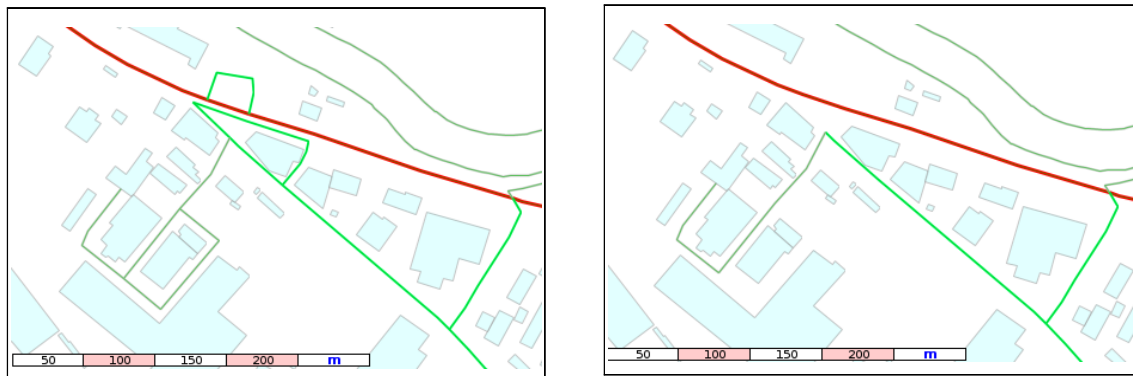


Figura 6.4: eliminazione delle maglie che soddisfano i vincoli relativi ai punti 3 e 4 dell'algoritmo per la selezione delle maglie della viabilità principale e secondaria.

Per i motivi esemplificati in Figura 6.1 e Figura 6.2, si è quindi deciso di sfoltere, prima di tutto, solo gli stroke *dangling*, eliminare successivamente le maglie di area piccola, ed infine sfoltere tutti i rimanenti segmenti *dangling* sotto i limiti di acquisizione.

Per la selezione delle maglie, invece, è stata data priorità alle maglie formate da almeno una strada di quartiere. La motivazione di questa scelta è da cercare nella definizione data di strada di quartiere: l'eliminazione di queste strade, essendo esse di scarsa utilità alla funzione di comunicazione, non altera sensibilmente la struttura della rete.

Il processo di selezione può quindi essere definito secondo la sequenza di algoritmi elencata in Tabella 6.1

PROCESSO DI SELEZIONE: SEQUENZA DEGLI ALGORITMI	
N°	ALGORITMO
1	Sfoltimento stroke <i>dangling</i> sotto i limiti di acquisizione
2	Sfoltimento maglie di piccola area formate da almeno una strada di quartiere
3	Sfoltimento maglie di piccola area della viabilità principale e secondaria
4	Sfoltimento segmenti stradali <i>dangling</i> sotto i limiti di acquisizione

Tabella 6.1: elenco ordinato degli algoritmi utilizzati nel processo di selezione.

7 Tipificazione di strutture parallele

In questo capitolo sarà trattato il problema della tipificazione delle strutture parallele della viabilità normale. Si analizzerà quando risulta preferibile costruire una linea di mezzeria tra due strade parallele, o quando invece è preferibile lasciare intatta una delle due strade e modificare l'altra. Sarà presentato successivamente un metodo per l'individuazione dei parallelismi tra due strade; infine, si discuterà la vera e propria azione di tipificazione, descrivendo il modo con cui ricavare le strade semplificate, e il modo con cui gestire tutte le strade connesse alle due strade originali.

7.1 Introduzione

La presenza di elementi stradali spazialmente allineati è un fenomeno che si verifica tipicamente all'interno del grafo stradale. Le autostrade e le strade a scorrimento veloce, per esempio, sono spesso formate da due carreggiate separate e parallele; oltre a questo tipo di strade, caratterizzate da una struttura ben regolare, all'interno degli agglomerati urbani si possono incontrare forme di parallelismi più o meno marcate.

Nel passaggio ad una scala di rappresentazione inferiore, la mancanza di spazio nella mappa è un problema che emerge con forza nel caso di strade parallele. Queste strade necessitano di una trattazione specifica: le tecniche di selezione delineate nel capitolo 6, infatti, possono risolvere solo situazioni estremamente semplici, e l'operatore di selezione non risulta adatto a risolvere il problema dei parallelismi.

Lo strumento per risolvere questo tipo di situazioni è l'operatore di *merging*, che fonde due o più linee in un'unica linea. Bisogna sottolineare comunque che l'operazione di fusione di una coppia di strade parallele in una linea di mezzeria, pur essendo una scelta abbastanza automatica nel caso di strade a scorrimento veloce o autostrade, non è sempre la soluzione corretta quando si opera sulla viabilità urbana.

Oltretutto, i parallelismi che si verificano in un agglomerato urbano possono coinvolgere strade di classificazione differente: supponendo quindi che la strada rappresentata nella mappa di classe superiore sia anche nella realtà più importante dell'altra, per risolvere il problema dell'assenza di spazio è più corretto intervenire sulla

strada di importanza inferiore. In (IGM, 2004) viene specificato che la strada di classe superiore deve restare inalterata, mentre quella di classe inferiore deve essere spostata; il collassamento in un'unica linea, però, in questo documento non viene trattato.



Figura 7.1: esempi di tipificazione tra strade di diversa e uguale importanza.

Invece, per quanto riguarda coppie di strade di pari classificazione¹, risulta necessario stabilire se tra esse ve ne sia una di importanza maggiore. Per valutare questo fatto, è stato deciso di utilizzare come parametro di confronto la lunghezza delle due strade: se, quindi, una delle due strade è molto più lunga dell'altra, è altamente probabile che all'interno della rete essa rivesta un ruolo più importante. Perciò, tra due strade S_1 ed S_2 di lunghezza rispettivamente l_1 e l_2 , con $l_1 > l_2$, la strada S_1 sarà reputata di importanza superiore rispetto a S_2 se $l_1 > l_2 \cdot k$, con $k > 1$. Un'eccezione a questa regola si ha nel caso in cui la strada più lunga è dangling; in questa situazione, sarà considerata più importante la strada di lunghezza inferiore. In Figura 7.1 sono mostrati due esempi di tipificazione dove sono coinvolte strade di diversa e uguale importanza.

Una volta chiarito in che modo viene scelta, se c'è, la strada di importanza maggiore, si può passare a descrivere in che modo sia possibile individuare due strade parallele; questa operazione viene trattata nella seguente sezione.

7.2 Individuazione dei parallelismi

Visivamente, due strade vengono percepite parallele dall'occhio umano se la distanza che le separa è sempre, all'incirca, la stessa. Tradurre questa considerazione in un algoritmo, però, non è così semplice, e si scontra in particolare col problema di come definire la distanza tra le due strade.

L'idea più semplice è quella di verificare se, ad ogni punto della prima strada, il punto più vicino appartenente alla seconda strada è a distanza inferiore ad una certa soglia: questa prima soluzione si adatta bene al caso in cui le due strade abbiano un andamento semplice e regolare, come avviene ad esempio per le autostrade.

Non è scontato, tuttavia, che a due punti consecutivi della prima strada corrispondano due punti consecutivi della seconda; inoltre, la distanza tra due punti consecutivi della stessa strada potrebbe variare, influenzando il processo di identificazione del parallelismo. Per gestire quest'ultima incertezza, è sufficiente aumentare il numero di punti della geometria di ogni strada, in modo da garantire una distanza massima prefissata tra due punti consecutivi. Per quanto riguarda l'altro problema, è opportuno prima spiegare come viene costruita la linea di mezzeria $S_{a,b}^m$ tra due strade parallele, S_a

1 È bene notare che appartenere alla stessa classe non significa anche avere la stessa importanza.

e S_b . L'idea è quella di trovare una sequenza di coppie di punti

$$seq_m = \langle (p_a^1, p_b^1), (p_a^2, p_b^2), \dots \rangle,$$

dalla quale poi ricavare la sequenza di punti della linea di mezzeria $\langle p_{ab}^1, p_{ab}^2, \dots \rangle$, dove p_{ab}^i è il punto medio del segmento con estremi p_a^i e p_b^i . Per trovare la sequenza di coppie, bisogna tener presente che il parallelismo tra due strade non si verifica necessariamente a partire dagli estremi di queste, ma potrebbe manifestarsi solamente nei tratti centrali delle due strade. Precisato questo, si può passare a descrivere come individuare, se esistono, i tratti paralleli di due strade.

Sia L_{min} la lunghezza minima dei tratti paralleli che si vogliono individuare, e d_{max} la distanza massima tollerata tra le strade parallele. Consideriamo la strada S_a , definita dalla sequenza di punti $\langle p_a^1, p_a^2, \dots, p_a^n \rangle$. Per individuare le strade candidate ad essere considerate parallele a S_a , si effettua una prima scrematura tra tutte le strade nell'intorno di S_a , tramite il calcolo del buffer attorno a S_a . Verranno selezionate solo le strade S_b che, all'interno del buffer, scorrono per una lunghezza superiore alla soglia L_{min} .

Una seconda scrematura viene effettuata controllando le lunghezze dei cammini minimi tra queste strade e S_a , in modo da evitare di fondere in un'unica strada due strade che nel grafo sono troppo lontane. Se per la strada S_b questa lunghezza risulta inferiore a max_{path} , allora sarà effettuato il calcolo della sequenza $\langle (p_a^1, p_b^1), (p_a^2, p_b^2), \dots \rangle$.

Supponiamo che la strada S_b , definita dalla sequenza di punti $\langle p_b^1, p_b^2, \dots, p_b^m \rangle$, soddisfi questi due vincoli. La sequenza seq_m delle coppie di punti viene così individuata:

1. si trova, tra tutte le possibili coppie di punti, la coppia (p_a^i, p_b^j) a distanza minima;
2. a partire dalla coppia (p_a^i, p_b^j) , si individua, tra le coppie (p_a^{i+1}, p_b^j) e (p_a^i, p_b^{j+1}) , quella per cui la distanza tra i due punti è minima; questa coppia sarà inserita nella sequenza seq_m , in posizione successiva a (p_a^i, p_b^j) , e questa procedura sarà ripetuta fino a quando non si giunge alla coppia (p_a^n, p_b^m) ;
3. a partire dalla coppia (p_a^i, p_b^j) , si individua, tra le coppie (p_a^{i-1}, p_b^j) e (p_a^i, p_b^{j-1}) ,

quella per cui la distanza tra i due punti è minima; questa coppia sarà inserita nella sequenza seq_m , in posizione precedente a (p^i_a, p^j_b) ; questa procedura viene ripetuta fino a quando non si giunge alla coppia (p^1_a, p^1_b) .

Il risultato di questi tre passi sarà la sequenza

$$Seq_m = \langle (p^1_a, p^1_b), \dots, best\{(p^{i-1}_a, p^j_b), (p^i_a, p^{j-1}_b)\}, (p^i_a, p^j_b), best\{(p^{i+1}_a, p^j_b), (p^i_a, p^{j+1}_b)\}, \dots, (p^n_a, p^m_b) \rangle,$$

dove $best\{(p^{i-1}_a, p^j_b), (p^i_a, p^{j-1}_b)\}$ è la funzione che seleziona la coppia di punti a distanza minima.

La sequenza viene successivamente partizionata in blocchi di sottosequenze parallele e non parallele: in ogni blocco ci saranno tutte le coppie consecutive, per cui la distanza tra i relativi due punti è sempre inferiore a d_{max} (blocco parallelo), oppure è sempre maggiore a d_{max} (blocco non parallelo).

Per ogni blocco parallelo b_i , viene calcolata la linea di mezzeria S^i_m e la relativa lunghezza L^i_m ; se questa lunghezza risulta minore di L_{min} , il blocco parallelo b_i viene convertito in blocco non parallelo, per evitare di unire le strade solamente per un breve tratto.

La corretta individuazione di questi blocchi è legata all'ordine con cui le sequenze di punti delle due strade sono memorizzate; a priori, infatti, non è possibile riconoscere se queste sequenze sono orientate secondo la stessa direzione, e potrebbe capitare che vicino al primo punto di una strada corrisponda l'ultimo punto dell'altra. Per superare questo problema, la procedura descritta per il calcolo della linea di mezzeria viene applicata quattro volte: a partire dal punto (p^i_a, p^j_b) , si può procedere incrementando o decrementando entrambi gli indici i e j , oppure incrementando uno e decrementando l'altro. Alla fine di queste quattro esecuzioni, verranno considerati i blocchi per i quali la somma totale L_m delle relative linee di mezzeria è massima. Tra questi blocchi, verranno scartati tutti i blocchi b_i per cui la relativa lunghezza di mezzeria è minore di L_{min} : i rimanenti blocchi genereranno le effettive linee di mezzeria tra la strada S_a e S_b .

Per le sottosequenze di coppie che formano questi blocchi, vengono effettuate alcune verifiche e modifiche, allo scopo di gestire particolari situazioni che si possono verificare nella viabilità stradale. In particolare, vengono trattati i problemi che sorgono quando le due strade parallele si incrociano, o una delle due strade ha una deviazione

improvvisa, o quando una delle due strade termina all'incrocio con una strada, mentre l'altra strada prosegue.

7.3 Generalizzazione dei parallelismi

Una volta individuati i blocchi paralleli e ricavate le relative linee di mezzeria, si può procedere con l'effettiva semplificazione del parallelismo tra la strada S_a e S_b .

Il primo passo consiste nell'individuare l'area che deve essere collassata o in una linea di mezzeria, o in una strada di importanza maggiore. Il calcolo di quest'area, intesa come poligono, è fondamentale per individuare e gestire tutti gli altri elementi stradali che vengono coinvolti nella generalizzazione delle due strade S_a e S_b . Per il calcolo di questo poligono, si utilizzano i blocchi individuati: per il blocco

$$b_k = \langle (p_{k,1}^a, p_{k,1}^b), \dots, (p_{k,t}^a, p_{k,t}^b) \rangle$$

il relativo poligono sarà costituito dalla sequenza di punti

$$\langle p_{k,1}^a, p_{k,2}^a, \dots, p_{k,t}^a, p_{k,t}^b, \dots, p_{k,1}^b, p_{k,1}^a \rangle$$

Su queste aree vengono poi effettuati dei controlli topologici incrociati con altri fenomeni rappresentati nella mappa, per verificare che la linea di mezzeria non generi inconsistenze. Se tramite queste verifiche si dovesse individuare la presenza di altri elementi geografici all'interno dell'area del poligono costruito, la tipificazione delle due strade viene annullata. Ad esempio, può essere verificata e opportunamente trattata la presenza all'interno dell'area di un corso d'acqua, di un edificio industriale o di uno civile.

Se la coppia di strade soddisfa i controlli topologici, si passa alla fase di tipificazione delle due geometrie. Si consideri, per descrivere questa fase, la tipica situazione di due strade che mostrano un parallelismo nella parte centrale delle rispettive geometrie. In questo caso, vengono identificati tre blocchi, b_1 , b_2 e b_3 , con b_2 blocco parallelo e b_1 e b_3 non paralleli. Siano i tre blocchi così definiti:

$$\begin{aligned} b_1 &= \langle (p_{1,1}^a, p_{1,1}^b), \dots, (p_{1,t}^a, p_{1,t}^b) \rangle \\ b_2 &= \langle (p_{2,1}^a, p_{2,1}^b), \dots, (p_{2,r}^a, p_{2,r}^b) \rangle \\ b_3 &= \langle (p_{3,1}^a, p_{3,1}^b), \dots, (p_{3,s}^a, p_{3,s}^b) \rangle \end{aligned}$$

e sia la linea di mezzeria del blocco b_2

$$S_m^2 = \langle p_1^m, p_2^m, \dots, p_r^m \rangle,$$

dove p_i^m è il punto medio del segmento avente come vertici i punti $p_{2,i}^a$ e $p_{2,i}^b$.

La costruzione della geometria tipificata di S_a avviene nel seguente modo: si individua il primo punto p_i^m di S_m^2 per il quale vale la relazione

$$\text{dist}(p_i^m, p_1^m) \geq L_{\text{conn}};$$

allo stesso modo si individua il primo punto p_j^m di S_m^2 per cui vale

$$\text{dist}(p_j^m, p_r^m) \geq L_{\text{conn}}.$$

L_{conn} è una lunghezza prestabilita che ha come scopo quello di rendere dolce il passaggio dalla strada originale alla linea di mezzeria.

La strada risultante S_a^f sarà costituita dalla sequenza di punti

$$S_a^f = \langle p_{1,1}^a, \dots, p_{1,t}^a, p_i^m, p_{i+1}^m, \dots, p_j^m, p_{3,1}^a, \dots, p_{3,s}^a \rangle.$$

In maniera analoga, viene costruita la strada risultante S_b^f :

$$S_b^f = \langle p_{1,1}^b, \dots, p_{1,t}^b, p_i^m, p_{i+1}^m, \dots, p_j^m, p_{3,1}^b, \dots, p_{3,s}^b \rangle.$$

Poiché le due strade finali condividono il tratto relativo alla linea di mezzeria, una di queste strade deve per forza essere divisa in due parti: quella che precede il tratto condiviso e quella che lo succede. La decisione di quale strada mantenere unita viene presa in base alla lunghezza di S_a^f e S_b^f : la strada di lunghezza minore sarà spezzata in due parti. Ad esempio, se tra le due, la strada più corta è S_b^f , questa verrà sostituita con le due strade

$$\begin{aligned} S_{b1}^f &= \langle p_{1,1}^b, \dots, p_{1,t}^b, p_i^m \rangle \\ S_{b2}^f &= \langle p_j^m, p_{3,1}^b, \dots, p_{3,s}^b \rangle. \end{aligned}$$

Nel caso generale, il numero di blocchi può essere maggiore o uguale a uno; è uguale a uno quando le strade sono parallele dall'inizio alla fine, e maggiore di uno quando c'è almeno un tratto di strada che non risulta parallelo all'altra. La costruzione, con un numero di blocchi generici, delle strade tipificate avviene similmente a quanto descritto

per il caso con due blocchi non paralleli e uno parallelo.

Una volta costruite le due strade tipificate, rimangono da controllare tutte le strade che precedentemente erano connesse al tratto parallelo. Per semplicità, consideriamo le strade connesse a S_a ; per S_b la trattazione è analoga. In generale, si possono distinguere tre diverse situazioni, riassunte nei seguenti punti:

1. esiste una strada S_c che inizia e termina all'interno dell'area parallela: in questo caso, S_c viene eliminata, in quanto lo spazio in cui era posizionata è stato collassato;
2. esiste una strada S_c che attraversa l'area parallela: in questo caso, S_c non subisce alcuna modifica;
3. esiste una strada S_c che termina esternamente sul bordo dell'area parallela: in questo caso, S_c viene prolungata fino ad intersecare la linea di mezzeria.

In Figura 7.2 sono mostrati i risultati prodotti dall'algoritmo di tipificazione descritto relativi alla zona di Monselice.

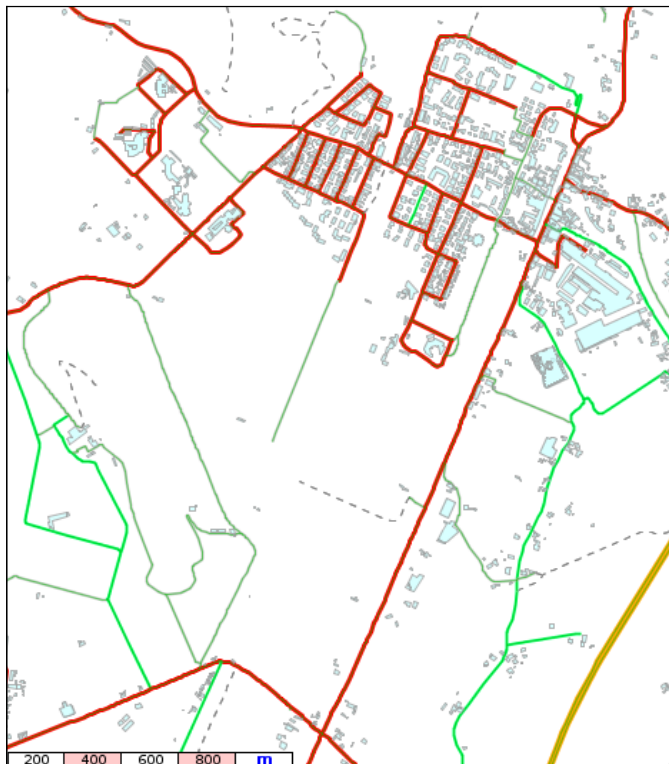
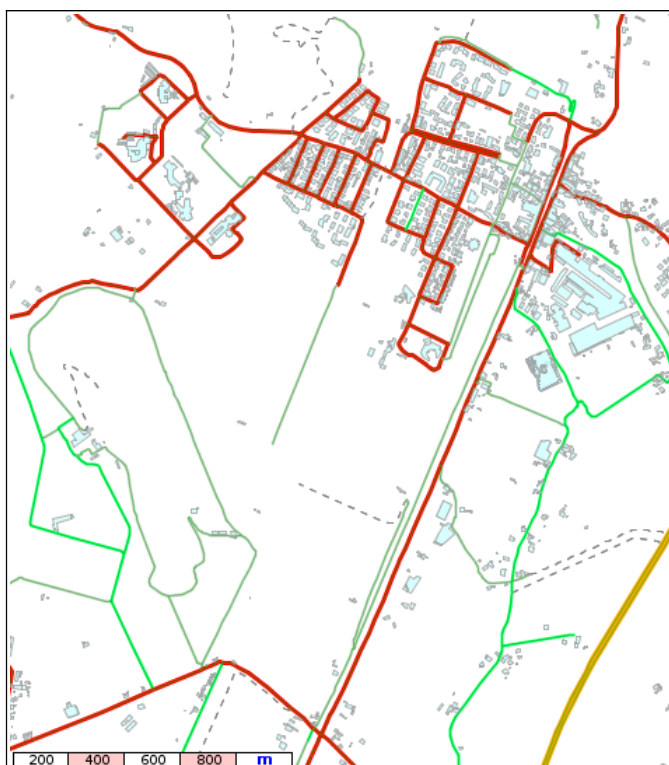


Figura 7.2: esempio di semplificazione di strade parallele.

Conclusione

In questo lavoro di tesi il problema della generalizzazione della rete stradale è stato scomposto in tre processi: armonizzazione, selezione e tipificazione.

Ognuno di questi processi è stato caratterizzato da una fase di *data enrichment* allo scopo di estrapolare delle informazioni necessarie al processo stesso; in particolare, allo scopo di definire una gerarchia di importanza tra le strade, è stata sviluppata una nuova misura, definita *local betweenness*, la quale si è dimostrata un valido ed efficace strumento per supportare i processi di armonizzazione e selezione. Tramite la *local betweenness*, così, è stato possibile mettere in risalto la struttura portante del grafo stradale, individuare le strade da inserire nella classe “strada urbana di quartiere” e infine usare convenientemente il processo di selezione per eliminare tutte quelle strade di scarsa rilevanza nel contesto della mappa.

I risultati prodotti dal processo di armonizzazione si sono dimostrati essenziali, poiché hanno permesso di adattare il livello di dettaglio della mappa alla nuova scala, nonché di fornire ai restanti algoritmi di generalizzazione un insieme di dati più omogeneo e capace di migliorarne le prestazioni. Va sottolineato, tuttavia, che tale processo può avere anche effetti negativi, ovvero può mettere in risalto strade che nella realtà non hanno alcuna importanza; infatti, nel caso di una forte presenza di errori nella classificazione delle strade iniziali, gli algoritmi di armonizzazione possono generare dati che risultano in contrasto con la realtà. Diventa perciò d'obbligo, terminato il processo di armonizzazione, effettuare un controllo per verificare la coerenza tra le strade armonizzate e la realtà.

Relativamente agli algoritmi per lo sfoltimento stradale, dai test effettuati si è potuto osservare che i tratti stradali coinvolti in questo processo di eliminazione sono stati sempre quelli di scarsa utilità per il collegamento; questo risultato positivo è stato ottenuto grazie alla caratterizzazione preventiva delle strade urbane di quartiere, e alla loro selezione tramite un algoritmo specifico appositamente realizzato.

L'eliminazione delle restanti strade è stata invece attuata in modo più conservativo, per evitare di compromettere la struttura portante della rete stradale; per questo motivo, non sempre sono state eliminate le strade per le quali l'eliminazione sarebbe stata l'azione più appropriata e automatica da compiere, tralasciando così la soluzione di alcune

situazioni critiche. Queste situazioni, però, spesso vengono risolte con il processo di semplificazione delle strutture parallele: in diversi casi, infatti, l'algoritmo di tipificazione, piuttosto che tipificare una coppia di strade, elimina quella meno importante. Il motivo per cui queste strade non vengono eliminate subito dal processo di selezione è il loro coinvolgimento in più maglie; in questi casi, infatti, l'algoritmo di selezione delle maglie limita l'eliminazione delle strade alle sole aventi un valore di *local betweenness* molto basso, e questo in alcuni casi risulta un vincolo troppo stringente.

Tutti gli algoritmi sviluppati per i processi di armonizzazione, selezione e tipificazione, sono stati realizzati in forma parametrica; questa caratteristica li rende particolarmente versatili e applicabili per la generalizzazione, oltre che alla scala 1:50.000, anche alla scala 1:25 000. Anche grazie ai risultati positivi ottenuti in questa tesi, dunque, il traguardo verso un completo processo di generalizzazione della rete stradale appare sempre più vicino.

È d'obbligo comunque precisare che un'ultima e grossa operazione rimane ancora da sviluppare: il displacement. Questa operazione risulta l'unica in grado di risolvere quelle situazioni in cui sono presenti più elementi ravvicinati che non possono essere eliminati, siano essi strade, fiumi o qualsiasi altro particolare di interesse, e per questo l'operatore di displacement specifico per la rete stradale è la prossima frontiera della ricerca all'interno del progetto CARGEN.

Riferimenti bibliografici

AGENT (2001) *Strategic Algorithms Using Organisations*, Public report, 2001, pp 6.

AGENT (2003) *Generalisation Modelling using an agent paradigm* Public report, 2003.

Baldin R. (2010) *Context awareness and typification in building generalisation*, Tesi di Laurea Magistrale, Università degli Studi di Padova, 2010.

Bertin J. (1983) *Semiology of graphics: Diagrams, Networks, Maps* Madison, WI: University of Wisconsin Press, pp 300.

Biasion S. (2009) *Sviluppo di algoritmi per l'automazione della generalizzazione cartografica dei tornanti*, Tesi di Laurea Specialistica, Università degli Studi di Padova, 2009.

Cecconi A. (2003) *Integration of cartographic generalization and multi-scale databases for enhanced web mapping* PhD. Thesis, University of Zurich.

Chaudhry O. (2004) *Rural and Urban Road Network Generalization Deriving 1:250,000 from 1:1250: Part 1 - Research paper* Masters in Geographical Information Science, University of Edinburgh.

Dizionario Devoto-Oli (1979) *Vocabolario illustrato della lingua Italiana* Selezione dal Reader's Digest, Milano.

De Gennaro M., Rumor M., Savino S. (2009) *Le procedure per la derivazione del DB25 dal DBT della Regione del Veneto: risultati del progetto CARGEN* Bollettino della Associazione Italiana di Cartografia, 135, Aprile 2009.

Deruda G, Falchi E, Falchi U. e Vacca G. (2005) *La generalizzazione cartografica automatica in ambiente GIS* Bollettino SIFET 2/2005.

Douglas D. e Peucker T. (1978) *Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature*, The Canadian Cartographer.

Enciclopedia Italiana Grolier (1987) Vol. 4, *Cartografia*, pp 293 – 296.

Heinzle F., Sester M., Anders K.H. (2005) *Graph-based Approach for recognition of*

patterns and implicit information in road networks, Proceedings of 22nd International Cartographic Conference, 9-16 July 2005, La Coruna/Spain.

IGM (2004) *Norme e Segni convenzionali per la realizzazione dei fogli della Carta d'Italia alla scala 1:50 000* Istituto Geografico Militare, Firenze.

Jiang B. e Claramunt C. (2004) *A structural approach to the model generalization of an urban street network*, GeoInformatica 8 pp 157-171.

Lamy S., Ruas A., Demazeau Y., Jackson M., Mackaness W. e Weibel, R. (1999) *The application of Agents in Automated Map Generalisation* Proceedings of 19th ICA meeting. Ottawa pp 160-169.

Lissandron I. (2008) *Sviluppo di algoritmi per la generalizzazione cartografica della viabilità stradale*, Tesi di laurea Specialistica, Università degli Studi di Padova, 2008.

Luan X. e Yang B. (2010) *Generating Strokes of Road Networks Based on Pattern Recognition*, ICA Workshop, 2010, Zurich.

Mackaness W. A. (2007) *Understanding Geographic Space*, in *Generalisation of Geographic Information: Cartographic Modelling and Applications*, Mackaness, W. A., Ruas, A. and Sarjakoski, L. T. (eds.), pp. 1–10, Elsevier, Oxford.

Mackaness W. e Mackechnie G. (1999) *Automating the Detection and Simplification of Junctions in Road Networks*, GeoInformatica, Vol. 3, Issue 2, 1999. pp. 185-200.

McMaster R. B. (1987) *Automated line generalization*, cartographia, 24(2), pag. 74-111.

McMaster R. B. e Shea K. S. (1992) *Generalization in Digital Cartography*, Association of American Geographers, Washington D.C.

Meyen E. (1973) *Multilingual dictionary of technical terms in cartography*, International Cartographic Association, Steiner, Wiesbaden.

Nickerson B. G. e Freeman H. R. (1986) *Development of a rule-based system for automatic map generalization* Proceedings of the Second International Symposium on spatial data Handling, Seattle, Washington.

Porta S., Crucitti P. e Latora V. (2006) *The network analysis of urban streets: a*

primal approach, Environment and Planning B: Planning and Design 33(5) pp 705–725.

Regnauld N. e McMaster R. (2007) *A synoptic view of generalisation operators in Generalisation of Geographic Information: Cartographic Modelling and Applications*, Mackaness, W. A., Ruas, A. and Sarjakoski, L. T. (eds.), pp. X–XX, Elsevier, Oxford.

Richardson D. E. e Thomson R. C. (1996) *Integrating Thematic, Geometric and Topologic Information in the Generalization of Road Networks*. Cartographica Volume 33 (1) 1996, pp 75-83.

Robinson A. H., Sale, R. e Morrison J. L. (1978) *Elements of Cartography*. New York: Wiley & Sons.

Savino S. (2007) *Il processo di generalizzazione cartografica: dalla Carta Tecnica Regionale al DB25 IGM*, Tesi di laurea Specialistica, Università degli Studi di Padova, 2007.

Savino S., Rumor M, Zanon M. e Lissandron I. (2010) *Data enrichment for road generalization through analysis of morphology in the CARGEN project*, ICA Workshop, 2010, Zurich.

Thom S. (2005) *A Strategy for Collapsing OS Integrated Transport Network(tm) dual carriageways*. In: Proceedings of the 8th ICA Workshop on Generalisation and Multiple Representation, A Coruña.

Thomson R. C. e Richardson D. E. (1999) *The 'Good Continuation' Principle of Perceptual Organization Applied to the Generalization of Road Networks* Proceedings of the ICA 19th ICC, Ottawa, pp 1215-1223.

Topfer F. e Pillewizer W. (1966) *The Principles of Selection, A Means of Cartographic Generalisation*, Cartographic Journal, 3(1) pp 10-16.

Touya G. (2007) *A Road Network Selection Process Based on Data Enrichment and Structure Detection* in Proceedings of the 10th ICA Workshop, Moscow, 2 - 3 August 2007.

Weibel R. e Dutton G. (1999) *Generalising spatial data and dealing with multiple representations, in Geographic information systems: Principles, techniques, applications and management*, (eds), Second Edition, Cambridge, GeoInformation

Interanational, pp. 125-155.

Zanon M. (2009) *Algoritmi per la generalizzazione cartografica: tipificazione di strutture parallele e trattamento delle autostrade e dei corsi d'acqua artificiali*, Tesi di Laurea Specialistica, Università degli Studi di Padova, 2009.

RINGRAZIAMENTI

Con questa tesi termina un percorso di studi durato quasi sette anni, ricco di emozioni e di tanti sacrifici.

Con le sole mie forze, però, a questa meta non ci sarei mai arrivato.

Ringrazio, prima di tutto, la mia famiglia, per avermi sempre voluto bene e sopportato anche nei miei momenti più difficili.

Ringrazio Silvia, per aver ritoccato il mio italiano maccheronico, che da leggenda è passato a legenda...e soprattutto per avermi portato a vivere esperienze indimenticabili che da solo non avrei mai nemmeno immaginato.

Ringrazio tutti i ragazzi del laboratorio GIRTS, per tutto il tempo passato insieme lavorando, ma in allegria.

Un grazie a parte va a Sandro, disponibile come un amico e saggio come un professore.

Grazie a tutti i miei amici di Castelfranco, Nico, Fabio, Joe, Andrea e Davide, per essermi stati sempre vicini in questi ultimi anni, a *o amigo mais embebido do mundo*, Icaro, e agli amici dell'università, Kappa, Francè, Aurelicchio, Greg, Skimo e Gianni, per il bellissimo tempo passato assieme a parlare di muri di fuoco (firewall) e tanto altre cose.

Ringrazio infine i Bambini dal Mondo, per aver colorato i miei pomeriggi con tanti sorrisi multiculturali.

Oggi, finalmente e purtroppo, termina il mio status di studente universitario...

xè ora de lavorar!

Elia Nicastro, Padova, marzo 2011