

L'UTILIZZO DELLE RETI NEURALI SOM PER L'INDIVIDUAZIONE E
L'IDENTIFICAZIONE DI CLASSI DI OMOGENEITÀ

Giovanni VIRGILIO¹, Nicola LONARDONI¹

¹ Dipartimento di Architettura e Pianificazione Territoriale. Università degli Studi di Bologna, Viale
Risorgimento 2, 40136, Bologna

SOMMARIO*

La complessità, propria dei sistemi locali, rende estremamente arduo il riconoscimento delle criticità e, per questo motivo, si dimostra ancor più difficile prevedere gli effetti delle strategie d'intervento sul territorio. Le tecniche tradizionali manifestano i propri limiti nella ricerca dei “nessi impliciti” che sottendono alle fitte interazioni che fanno di un sistema territoriale un sistema complesso. Si rende così necessario il ricorso a metodi di indagine che impiegano strumenti analitici più evoluti in grado di “percepire” l'impronta interattiva di un sistema locale e la sua capacità di auto-organizzazione. In un'ottica analitica quale quella proposta in questa sede si assume che il processo di auto-organizzazione del sistema locale sia spinto da un lato dalle necessità di effettuare scelte volte a migliorare la posizione di competitività assoluta, dall'altro, dall'esigenza di mantenere un rapporto di equilibrio con gli altri elementi del sistema. Le Reti Neurali basate sull'algoritmo SOM (Self-Organizing Map) hanno mostrato di avere i requisiti necessari al riconoscimento di sistemi territoriali interagenti, riuscendo a cogliere i rapporti tra le diverse realtà territoriali, nella fattispecie le Province Italiane, e le modalità con cui questi determinano l'assetto relazionale generale. I metodi propri della cluster analysis nel riconoscimento di classi di omogeneità vengono qui integrate con le potenzialità esplorative offerte dalle Reti Neurali.

* La stesura del presente paper, pur conseguendo ad un confronto sistematico e ad una riflessione comune, è stata curata da Giovanni Virgilio relativamente ai paragrafi 1 e 4 e da Nicola Lonardoni, relativamente al paragrafo 2; il paragrafo 3 è frutto di un lavoro unitario.

1 INTRODUZIONE

In un momento storico, quale quello attuale, caratterizzato da uno sviluppo intenso delle reti, specie quelle di comunicazione, e dal conseguente emergere delle nuove complessità (ciò che con un termine, forse anche troppo abusato, si è soliti definire *globalizzazione*), si sarebbe erroneamente portati a ritenere di essere prossimi ad una perdita di significato e di ruolo del territorio.

In realtà è vero l'esatto contrario, trovano, infatti, conferma alcune intuizioni [Sassen, 1994], che avevano delineato un'evoluzione sistemica in cui all'emergere dell'importanza del fattore *tempo*, non corrisponde, sul piano effettuale, l'annullamento della dimensione territoriale dei fenomeni, e ciò nonostante un'effettiva riduzione delle barriere spaziali, anzi, è andata crescendo l'importanza del ruolo che le variabili territoriali svolgono nell'ambito degli attuali processi di accumulazione flessibile del capitale. Da tale stato di cose deriva, come conseguenza immediata, che il territorio (da intendere qui nel suo significato più ampio) viene a svolgere un ruolo centrale nelle contraddizioni che connotano l'attuale sistema, prima fra tutte quella che vede la crisi del modello di sviluppo industriale a fronte dell'emergere di dinamiche nuove atte a soddisfare allo stesso tempo sia la necessità di risposte olistiche alla spinta prodotta dai processi di globalizzazione, sia l'esigenza di salvaguardare e valorizzare il proprio patrimonio genetico attraverso l'affermazione delle diverse identità locali. In un sistema caratterizzato da processi flessibili di accumulazione del capitale, le componenti strutturali territoriali devono evolvere verso forme in grado di cogliere l'esigenza di rapida adattabilità ai mutamenti organizzativi insiti nell'instabilità delle reti relazionali che si vengono a strutturare ai diversi livelli. In altri termini, quanto più velocemente il sistema economico converge verso una situazione caratterizzata da un'estrema uniformità dei processi produttivi (e questo discorso vale ancor di più per le produzioni immateriali), tanto più diviene necessario riuscire ad incamerare anche i differenziali minimi di vantaggio competitivo. Questo fatto sta a significare che la scelta del luogo in cui insediare la propria attività economica diviene un fattore strategico, poiché è facile comprendere come tali vantaggi crescano in termini *assoluti* in funzione delle scelte localizzative [Camagni, 2002]. Tuttavia, poiché è lecito ritenere che l'esistenza di esternalità peculiari ad uno specifico territorio sia principalmente da attribuire alla presenza di capitale e di reticoli sociali e che questi, a loro volta, costituiscano il risultato di una complessa commistione relazionale che si struttura nel tempo tra un numero estremamente ampio di elementi e di caratteristiche, si rende necessario analizzare la natura e le modalità con cui si vengono ad articolare i processi di auto-organizzazione del territorio, al fine di individuare dei fattori potenziali che ci permettano di svolgere un ruolo efficace nell'attivazione e nel monitoraggio di tali processi. Un tale proposito richiede la capacità di misurarsi con la complessità dei sistemi territoriali; una complessità che presenta una duplice natura essendo riferibile sia al contesto ambientale

in cui maturano i problemi legati alle trasformazioni e al governo del territorio sia alla natura dei problemi stessi, poiché quest'ultimi si presentano sempre con maggiore frequenza in una forma indefinita e destrutturata. E' necessario, quindi, riuscire ad isolare le eventuali relazioni di interdipendenza esistenti tra le diverse componenti; ciò permette, da un lato, di coglierne la generale articolazione permettendo in tal modo una valutazione in termini di risposta complessiva del sistema alle azioni intraprese, dall'altro, la possibilità di ricondurre i problemi di elevato grado di complessità in componenti più semplici, evidenziandone, ove possibile, le diverse relazioni causa/dipendenza. In sostanza diventa rilevante determinare l'ordine d'importanza dei singoli problemi e della proporzionalità delle risorse da dedicarvi e ciò vale tanto più quando ci si trova ad operare in presenza di risorse scarse (fatto che costituisce la condizione di normalità nell'azione territoriale) [Virgilio, 2002].

In quest'ottica, quindi, il presente studio si configura come un apporto metodologico il cui fine è quello di dimostrare come sia possibile estrarre informazioni sulla complessità dalla complessità stessa, prefiggendosi il raggiungimento di tale proposito attraverso la proposizione di un'analisi sistemica condotta con l'uso di Reti Neurali Artificiali. Lo scopo generale è quello di pervenire all'individuazione di tendenze di fondo nel comportamento, riconducibile a azioni, reazioni e ripercussioni, dei sistemi regionali, i quali, come è noto, sono composti da un vastissimo insieme di caratteristiche diversamente coinvolte nella definizione del sistema stesso. Variabili di diversa rilevanza e importanza convogliano, infatti, nella descrizione delle trasformazioni in atto o in potenza del territorio. Dall'osservazione delle dinamiche territoriali emerge che tali componenti sono variamente correlate tra loro, formando il reticolo di "nessi impliciti" che è necessario svelare al fine di conoscere i principi su cui si organizza il territorio sui suoi diversi livelli.

Da un punto di vista analitico il caso italiano presenta caratteristiche di eloquente esemplarità per ciò che attiene i divari di sviluppo presenti sul territorio nazionale. Una diversità che si basa su differenti condizioni di partenza e opportunità; pertanto, comprendere quali siano le possibili leve su cui agire, al fine di provocare un mutamento che vada in una direzione individuata, costituisce di sicuro uno dei ruoli principali di chi fa pianificazione. Da qui l'esigenza di affinare capacità analitiche particolari, finalizzate ad affrontare la complessità relazionale propria dei sistemi territoriali, partendo dalla consapevolezza che: *"[...] ogni specifica realtà può essere colta come un esito, uno dei molti possibili, delle complesse interazioni di natura sociale, economica e fisico-ambientale e quindi come una delle tante configurazioni in cui il sistema si può realizzare"* [Diappi et al., 1999].

I principi che governano la realtà sono pertanto celati dietro una casualità solo apparente, avente una funzione di schermo all'effettivo manifestarsi del fenomeno dell'auto-organizzazione del territorio. Come già chiarito in precedenza nella necessità di pervenire ad una semplificazione dei sistemi e l'isolamento dei caratteri principali risiede il ruolo essenziale della classificazione, ove quest'ultima deve dare risposte congrue e coerenti al fine

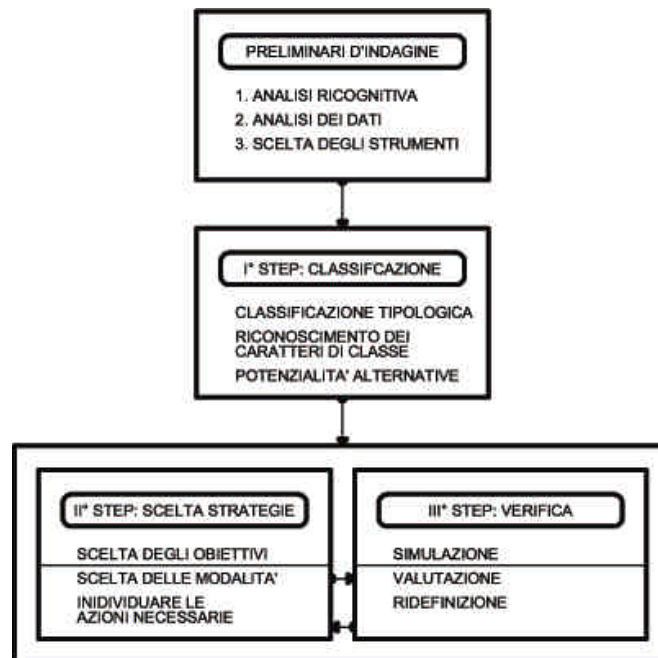
di: soddisfare l'esigenza di identificazione delle invarianti che determinano una particolare configurazione del contesto territoriale, permettere il riconoscimento di un evento e la relativa appartenenza a una precisa categoria, definire i termini in cui un sistema sia descritto in maniera esaustiva.

Da questo punto di vista l'algoritmo della *Self-Organizing Map* (*Mappa Auto-Organizzante*), che appartiene alla famiglia delle *Reti Neurali Artificiali ad apprendimento non supervisionato*, si è dimostrato particolarmente efficace nel fornire risposte alle esigenze in questione. La sua finalità principale è quella di organizzare una distribuzione di osservazioni, in generale descritte da numerose variabili, su uno spazio a dimensionalità limitata. La logica neurale permette di utilizzare dati incompleti o disturbati dispensando l'operatore da uno specifico trattamento dei dati di partenza.

A questo scopo si rendono necessari i passaggi analitici che portano alla classificazione e all'individuazione dei caratteri di classe, nonché al riconoscimento tipologico.

L'impostazione seguita è ricalca le fasi di un processo in cui il decisore pubblico è chiamato ad operare delle scelte di azione strategica e permette allo stesso tempo di testare le potenzialità espresse dalle S.O.M. come strumento analitico finalizzato al supporto del processo decisionale. In questa procedura, l'analisi dei dati, che costituisce un momento di fondamentale importanza, assume come riferimento teorico lo schema del processo proposto dalla Fraire [Fraire,1985].

Figura 1.1 – Ipotetigrafia del protocollo di indagine



Come si evince dallo schema, la fase di individuazione delle classi tipologiche assume un ruolo centrale nello sviluppo dell'intero processo analitico. Questa costituisce, infatti, un nodo fondamentale di ogni processo cognitivo, dal momento in cui dalla categorizzazione vengono rilevate le caratteristiche principali di ogni gruppo di oggetti. In riferimento ai sistemi locali, che risultano caratterizzati da entità composte e complesse regole evolutive, l'individuazione tipologica mediante metodologie classiche presenterebbe diverse difficoltà.

Infatti, questi metodi presupporrebbero non solo l'esistenza di un insieme limitato di caratteristiche tali da definire completamente ed univocamente le categorie, ma soprattutto la perfetta conoscenza del fenomeno da parte di chi indaga. Per gli stessi motivi non è possibile conoscere a priori il numero di classi, né determinarne i criteri di appartenenza. Ciò che può essere stabilito con le debite considerazioni cautelative è il criterio di somiglianza tra le osservazioni. È facile dedurre come tutte le precondizioni qui individuate sono difficilmente riscontrabili in casi di analisi di sistemi complessi. Da questo punto di vista, per quanto siano raffinate, le tecniche analitiche tradizionali non riescono a cogliere pienamente le interazioni che si stabiliscono tra le variabili sia nella loro dimensione infrarelazionale che in quella interrelazionale e che contribuiscono a definire il *campo relazionale* in cui si manifesta la capacità del territorio di auto-organizzarsi.

Di seguito verrà proposta una metodologia per l'individuazione delle classi di omogeneità, descrivendo come, a tal scopo, si integrino i risultati della S.O.M. con l'utilizzo della cluster analysis. Infine verranno presentati alcuni risultati sperimentali che avranno come oggetto le Province Italiane nel contesto della Competitività Territoriale.

2 RETI SELF - ORGANIZING MAP

Le Reti Neurali basate sull'algoritmo S.O.M. sono in grado di raggiungere buoni risultati sia in problemi di *data-clustering*, come nel riconoscimento immediato di *osservazioni estranee*¹, e ancora nella valutazione di possibili sentieri di trasformazione tra due diverse configurazioni. Questo permette, da un lato, di cogliere la struttura dei contesti analizzati dall'altro, di simulare gli effetti derivanti dalle modifiche di tale struttura. L'architettura di queste reti neurali si sviluppa principalmente su uno strato, formato da una quantità arbitraria di unità elementari dette *Processing Element* (PE), e chiamato in letteratura *Strato di Kohonen*². Generalmente i PE definiscono uno spazio a due o tre dimensioni, ma la dimensionalità dello strato è discrezionale. La particolarità della S.O.M. è quella di creare una mappa delle osservazioni *n*-dimensionali di ingresso in questo spazio costituito da un minor

¹ Si definiscono con l'appellativo "estranee" quelle osservazioni che non fanno parte del *training set* della rete neurale.

² Lo stato di sviluppo attuale di questa architettura è dovuto al sostanziale contributo di Teuvo Kohonen che negli anni '90 ha concentrato gli sforzi delle sue ricerche proprio sui sistemi adattivi basati su algoritmi di quantificazione vettoriale. Un'utile approfondimento in tal senso è rappresentato dagli studi condotti dal Semeion Centro Ricerche di Scienza della Comunicazione e dal suo direttore Dr. Massimo Buscema.

numero di variabili. Questo processo di organizzazione viene comunemente definito come Feature Mapping, o Kohonen Feature Mapping nel caso particolare in cui tale mappaggio avvenga in modo non supervisionato come nel caso delle S.O.M.. Una caratteristica fondamentale di questo algoritmo è quella di preservare le relazioni topologiche tra i dati di ingresso che vengono codificati nella mappa.

Il parallelismo biologico su cui è stato sviluppato l'algoritmo delle S.O.M. riguarda la specializzazione di regioni cerebrali per stimoli di ingresso diversi. L'organizzazione di queste specializzazioni è biologicamente prodotta dalla classificazione, o meglio dalle classi diversificate, degli stimoli e delle reazioni che il sistema gestisce. Stimoli simili andranno infatti a sollecitare le stesse porzioni di tessuto delimitate da un confine di neuroni a bassa attività che rappresentano un settore di transizione tra una attività ed un'altra. Dalle fasi iniziali dell'apprendimento, il sistema nervoso tende quindi a organizzare il sistema centrale specializzando i settori su funzioni diverse. Questo processo avviene secondo i criteri di auto-organizzazione già definiti per i sistemi complessi.

2.1 L'algoritmo di base

Dato un collettivo di osservazioni predisposto in c vettori n -dimensionali $X=\{x_1, x_2, \dots, x_n\}$, si supponga di voler studiare le modalità di auto-organizzazione in uno spazio bidimensionale p per q , cioè di volerne comprimere le dimensioni in una distribuzione piana.

Lo strato di input è composto da n unità o nodi di ingresso, quello di uscita della S.O.M. viene quindi costituito dai punti del piano su cui si distribuirà il collettivo in esame ed è quindi composto da $m=(p \times q)$ Processing Elements. Ogni PE è connesso con ogni elemento dell'input tramite un vettore di pesi W che nel complesso verranno a costituire una matrice tridimensionale $[W]$.

L'inizializzazione dei pesi avviene mediante una sequenza casuale di numeri arbitrariamente piccoli in un intorno di zero, modalità in cui non possono sussistere valori identici in un insieme di numeri sufficientemente ampio. L'unico vincolo sta nella criticità di due PE associati a vettori di pesi identici.

La fase di apprendimento non supervisionato avviene tramite la presentazione ciclica o casuale di tutte le osservazioni presenti nel collettivo o in un sottoinsieme di esso scelto come *learning set*. Detto X il vettore presentato in ingresso, esso viene confrontato con ognuno degli m vettori dei pesi W . Tra gli m PE viene eletto vincitore quello associato al vettore dei pesi considerato più vicino al vettore di ingresso. La tecnica descritta è detta di Quantificazione Vettoriale ed è alla base di molti algoritmi di Reti Neurali per la classificazione (come la Learnig Vector Quantization, ad esempio). La vicinanza e la Winner Unit possono essere definite in due modi: tramite il massimo prodotto scalare; tramite la minima distanza euclidea, le cui rispettive formalizzazioni risultano essere:

$$WinnerUnit = PE_s \Leftrightarrow C_s = \max_{r \in [1..m]} \left\{ C_r = X \cdot W_r = \sum_{i=1}^n x_i \cdot w_{ri} \right\} \quad (2.1)$$

$$WinnerUnit = PE_s \Leftrightarrow D_s = \min_{r \in [1..m]} \left\{ D_r = \|X - W_r\| = \sqrt{\sum_{i=1}^n (x_i - w_{ri})^2} \right\} \quad (2.2).$$

Determinata l'unità vincitrice, viene aggiornata la matrice dei pesi sia sul vettore del PE vincitore sia per i PE compresi nel suo vicinato. Questo passaggio è il carattere distintivo delle S.O.M., quello in cui emerge la sua particolarità di conservare i caratteri topologici del collettivo preso in esame. Questo metodo permette di ridisegnare, ciclo dopo ciclo, lo spazio e la trasformazione proiettiva che permette di rappresentare un insieme di osservazioni n -dimensionali in due sole dimensioni. Il piano che verrà a formarsi sarà quindi costituito da porzioni transizionali tra le varie modalità di cui è composta la realtà. Si vedrà in seguito come ciò produca informazioni leggibili non solo dal punto di vista della classificazione, ma anche dal punto di vista dell'analisi dei fattori caratterizzanti il fenomeno e nella costruzione di potenziali sentieri di trasformazione tra una configurazione ed un'altra.

L'aggiornamento segue la formulazione:

$$\Delta W(D) = h(D) \cdot (X - W_r) \quad (2.3)$$

dove $h(D)$ rappresenta una *funzione di definizione del vicinato*. Solitamente, quest'ultima è la funzione *gaussiana*:

$$h(D) = He^{-\frac{D^2}{2\sigma^2}} \quad (2.4)$$

Sebbene nella forma originale e nel significato di tale funzione σ rappresenti la deviazione standard della variabile indipendente, che in questo caso è la distanza D , essa può anche assumere valori arbitrari assegnateli dall'esterno, al pari del parametro H , la sua importanza si manifesta ciononostante nel fatto che da essa dipende la dimensione del suo vicinato. Al variare di questi parametri viene a modificarsi l'influenza che la configurazione dell'osservazione X esercita sul vicinato. In altri casi $h(D)$ può essere definita come:

$$h(D) = \begin{cases} H \Leftrightarrow PE_r \in I \\ 0 \Leftrightarrow PE_r \notin I \end{cases} \quad (2.5)$$

dove I è l'intorno arbitrario in dimensioni, forma e vincoli.

Naturalmente, poiché la libertà di auto-organizzazione della rete è strettamente correlata con la possibilità dei PE di differenziarsi, minore sarà il vicinato influenzato dall'apprendimento e maggiore sarà la libertà di auto-organizzazione della S.O.M., ma ciò può portare a rischi d'interpretazione in quanto due osservazioni potenzialmente simili possono ritrovarsi separate a causa di un'eccessiva specializzazione che va a scapito della generalizzazione.

La fase di risposta della rete coincide con l'attivazione dei PE in corrispondenza della posizione che le è stata assegnata dopo l'auto-organizzazione. Il tutto può avvenire sia al termine dell'apprendimento restituendo i risultati finali dell'organizzazione, sia processando

un'osservazione estranea al *training set* che la S.O.M. stessa provvederà a collocare nella posizione più corretta del piano o dello spazio definito dai PE.

In ultima analisi, oltre alle informazioni sulla distribuzione in due dimensioni dei vettori n-dimensionali presentati all'ingresso, possono svilupparsi considerazioni sullo spazio auto-organizzatosi al termine dell'apprendimento.

Ad esempio, una volta terminata la fase di addestramento, è possibile reimmettere un caso del training set e valutare il grado di attivazione delle varie celle, valutando così le potenziali collocazioni alternative che uno spazio così organizzato potrebbe dare al caso in esame.

E ancora, date due posizioni differenti e ad una certa distanza è possibile valutare il sentiero di trasformazione che da una situazione porta all'altra e viceversa analizzando la composizione dei vettori di pesi associati ai punti che ne collegano il cammino.

2.2 L'algoritmo di Classificazione Automatica

Allo scopo di avere un metro di confronto tra gli strumenti della Statistica Multidimensionale Classica e le nuove tecniche di analisi proprie dell'Intelligenza Artificiale è stato implementato un algoritmo di Cluster Analysis rendendo così possibile valutare potenzialità e limiti di entrambi gli approcci. Per la Classificazione Automatica si è utilizzato un algoritmo gerarchico aggregativi che andiamo a descrivere con maggiore dettaglio.

La formulazione generalizzata della metrica attraverso la quale viene definita la *Matrice delle distanze o di similarità*, ossia la matrice delle distanze reciproche tra gli n vettori di variabili, è nota dalla letteratura in materia (Hardy et al., 1964; Leti 1979; Rizzi, 1985; Bolasco 1999;) come la metrica di Minkowski e può essere assunta come espressione:

$$D_M(X_i, X_j) = \left[\sum_{s=1}^N |x_{is} - x_{js}|^I \right]^{\frac{1}{I}} \quad (2.8)$$

che per $I=1$ assume la forma di:

$$D(X_i, X_j) = \sum_{s=1}^N |x_{is} - x_{js}| \quad (2.9)$$

nota come distanza City Block o Manhattan, che presenta la caratteristica di porre "[...] *sullo stesso livello tutte le differenze*" [Rizzi, 1985], mentre per $I=2$ assume la forma della distanza euclidea:

$$D_e(X_i, X_j) = \sqrt{\sum_{s=1}^N (x_{is} - x_{js})^2} \quad (2.10)$$

che contrariamente alla distanza di Manhattan attribuisce un peso maggiore alle caratteristiche principali in senso statistico, ossia le variabili con maggiore varianza.

Per il presente studio si è optato per l'uso della distanza di Manhattan in quanto al contrario delle altre metriche è in grado di definire uno spazio caratterizzato da distanze minime, questo

si traduce sul piano analitico su una maggiore efficacia nell'individuazione dei Clusters e sul piano operativo riduce i tempi di elaborazione dei dati.

La scelta dell'algoritmo di classificazione automatica e, soprattutto, dei criteri da adottare nel processo, si rivelano determinanti per il raggiungimento di una suddivisione ottimale del collettivo in gruppi che presentino un adeguato livello di omogeneità interna. Questo perché si assume che una suddivisione possa essere considerata soddisfacente nel momento in cui la classificazione ottenuta sia in grado di cogliere tutte le concentrazioni locali di osservazioni, il che implica la possibilità di disporre, ad analisi conclusa, di un elenco di tutte le tipologie presenti nel collettivo.

Tuttavia, poiché non si dispone di un metodo generale che garantisca, a priori, l'esatta determinazione della soglia entro la quale un insieme di elementi si possa definire concentrato o disperso, omogeneo o viceversa eterogeneo, risulta necessario fare ricorso ad alcuni indici di riferimento capaci di cogliere le caratteristiche globali del sistema analizzato e sulla base di queste è possibile estrarre informazioni che risultano preziose in fase di selezione e implementazione dell'algoritmo di processo. Sotto questo profilo, infatti, se le osservazioni si rivelassero equidistribuite la ricerca di una classificazione tipologica risulterebbe inutile in quanto ogni caso si identificherebbe con una tipologia a sé, in accordo con l'eterogeneità espressa da una distribuzione uniforme. La concentrazione assoluta equivarrebbe ad isolare una sola classe alla quale appartenerebbero tutti i casi in esame. Lo studio della classificazione, quindi, si rende utile nei casi che manifestano un grado di concentrazione intermedio, laddove esistono dei gruppi, ma non un solo gruppo. Il riscontro di un'alta concentrazione restringerebbe la scelta dell'algoritmo di classificazione più opportuno. In questo caso, infatti, il metodo migliore per l'identificazione delle cluster potrebbe rivelarsi un algoritmo gerarchico di tipo scissorio o *top-down*. Tale procedura manifesta i suoi limiti in corrispondenza di collettivi numerosi. Infatti, anche con le moderne tecnologie, il tempo di calcolo di un algoritmo gerarchico scissorio per un'insieme di 100 elementi risulterebbe proibitivo [Rizzi, 1985, *op cit.*].

I risultati ottenuti dimostrano che si è di fronte ad una situazione prossima all'equidistribuzione, ciò implica che si dovrà procedere ad una forzatura dei gruppi, ottenibile mediante un processo aggregativo di tipo gerarchico, e che, inoltre, è lecito attendersi la presenza di un gruppo di casi che presentino caratteri di eterogeneità, la cui numerosità dipenderà dal livello di soglia ritenuto accettabile.

Il metodo aggregativo basato sull'algoritmo gerarchico si traduce nella formazione sequenziale di cluster il cui baricentro costituisce l'elemento su cui innescare il processo di aggregazione successivo. Si noti che al fine di impedire che il rapporto di dissomiglianza tra i singoli punti possa essere in un qualche modo condizionato dalla quantità di elementi presenti nel cluster, ad ogni ciclo viene assegnato sia alle osservazioni ancora aggregate che ai baricentri dei cluster già formati una massa unitaria.

Il metodo proposto, noto in letteratura come *Metodo del legame medio*, presenta la proprietà non trascurabile di garantire sempre il soddisfacimento delle proprietà di *partizione ben strutturata minimale* (Rizzi, 1985, op cit.).

Il passo successivo è quello di determinare, attraverso un processo iterativo, il numero dei clusters ritenuto soddisfacente. L'iterazione aggrega ad ogni ciclo le coppie di punti (osservazioni o baricentri) tali per cui è valida la seguente relazione:

$$D_{x_{G1}, x_{G2}} \leq L_a \quad (2.11)$$

dove L_a è il *Livello di dissomiglianza accettabile*. Il processo si interrompe quando non vi sono più coppie di punti o baricentri che distino meno di questo valore.

3 UN ESERCIZIO CON LE PROVINCIE ITALIANE

Il database su cui si è scelto di condurre lo studio è costituito dall'Atlante della Competitività delle Province, prodotto dall'Istituto Tagliacarne per conto di Unioncamere, sul quale si è operata una selezione secondo i criteri che verranno tra breve illustrati. La banca dati consta di più di 500 variabili, che derivano da diverse fonti istituzionali, riferite a ciascuna delle 103 province italiane.

La scelta della dimensione provinciale come unità di riferimento trova la sua principale giustificazione nel fatto che, da un punto di vista amministrativo e politico, è proprio a questo livello che dalla legge 8 giugno 1990, n.142, è demandata la sintesi delle linee strategiche per mezzo dei Piani Territoriali di Coordinamento Provinciale. D'altronde, come è stato giustamente rilevato “[...] il PTCP nasce da un lato come risposta alla domanda di scelte strategiche, strutturali, di lunga prospettiva e dall'altro a quelle realmente ipotizzabili e concretamente operabili da selezionare tra quelle proposte o rivendicate, sulla base delle potenzialità finanziarie effettivamente spendibili e/o prefigurabili nel tempo, intorno alle quali far confluire la discussione e poi la decisione di enti e istituzioni pubbliche e degli stessi operatori privati” (Campos Venuti et al., 2000). Ciò avvala questa scelta, che sembra allo stato attuale coerente con la necessità di un generale coordinamento delle politiche del territorio mirate ad attivare processi di competitività.

Queste sperimentazioni sono state condotte con l'obiettivo generale di identificare dei criteri di competitività territoriale. Tale contesto ha avviato una ricerca di strumenti il cui requisito essenziale fosse quello di poter affrontare la complessità territoriale che trova una definizione appropriata nel concetto di “*Mileiu Innovateur*” introdotto dai lavori che Roberto Camagni ha pubblicato, negli ultimi anni, nel panorama dell'analisi della competitività locale. Il milieu è definito dallo stesso Camagni [Camagni, 1999] come “[...] un insieme di relazioni che insistono su uno spazio limitato e che portano ad unità un sistema locale di produzione, una cultura produttiva, un insieme di attori e di rappresentazioni, generando un processo localizzato di apprendimento collettivo.”. In queste poche parole vengono sinteticamente

descritte le caratteristiche principali di un sistema locale, quali l'auto-organizzazione, i fattori di relazione e la componente della crescita collettiva sociale.

Un primo passo per identificarne i fattori costitutivi dei sistemi territoriali in competizione consiste nell'individuare un collettivo di unità tali non solo da rappresentare un milieu, ma dalla caratteristica di essere confrontabili tra loro. Questo elemento è forse alla base del concetto stesso di competitività, in quanto l'esistenza stessa della medesima può sussistere solamente tra attori che coprono posizioni marginali di un intorno prossimo a all'equilibrio, ossia a configurazioni che si riconducano ad una perfetta sovrapposizione. Lo scopo di identificare gruppi di entità tra loro omogenee è quindi duplice: da un lato si tendono a definire i risultati generali di una classificazione (riduzione delle variabili significative, identificazione dei parametri di appartenenza ad una classe, riconoscibilità delle osservazioni), dall'altro si cerca di valutare le potenzialità competitive di contesti diversi attraverso lo studio di una metrica comparativa. Ciò equivale, in generale, a definire dei parametri di somiglianza che si costruiscono sulla composizione del sistema osservato.

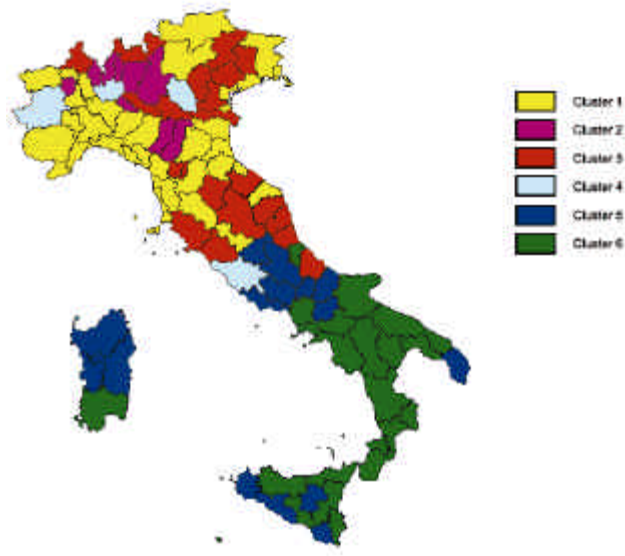
3.1 Individuazione e identificazione tipologica: le classi omogenee

Come si è detto, nel caso di sistemi territoriali, l'omogeneità, o la similitudine, può essere espressa attraverso la misurazione delle variabili che ne descrivono la composizione oppure attraverso uno studio che tenga conto dei fattori relazionali e auto-organizzativi.

Figura 3.1 – Risultati della Cluster Analysis sulle 160 variabili preselezionate



Figura 3.2 – Risultati della sequenza S.O.M. – Cluster Analysis sulle medesime variabili



Allo scopo di confrontare i due criteri è stata effettuata la ricerca delle classi attraverso l'uso di una Cluster Analysis e attraverso le reti S.O.M., che sono strumenti dedicati al secondo approccio nonostante sia necessario uno strumento in *font-end* che realizzi fisicamente l'ultima fase di calcolo per le aggregazioni. Per quest'ultimo passaggio è stato applicato un algoritmo gerarchico aggregativo con esattamente gli stessi criteri di quello usato singolarmente. Il livello di dissomiglianza accettabile, al pari delle altre impostazioni, è stato assegnato con criteri comuni al preciso scopo di aumentare la confrontabilità dei due approcci. Il valore di L_a è stato calcolato come il quartile del range di variabilità della distanza tra le osservazioni. Questa scelta non è finalizzata alla individuazione del miglior partizionamento del collettivo, ma a forzare l'espressione delle caratteristiche proprie dei due metodi.

Nelle Figure 3.1 e 3.2 sono rappresentati i risultati ottenuti rispettivamente con la sola cluster analysis e con la sequenza S.O.M.-cluster.

La cluster analysis ha prodotto l'aggregazione di due soli insiemi che separano nettamente la penisola in due macroregioni che comprendono la prima il Centro-Nord e la seconda il Mezzogiorno. Vengono poi isolati sette casi accreditati come eterogenei. Si tratta di Province la cui particolarità è plausibile sulla base delle variabili considerate.

La S.O.M.-cluster, invece, individua ben sei raggruppamenti e non prevede la presenza di casi così particolari da dover essere considerati singolarmente. Come si può notare dalla figura, la lettura dell'Italia da Nord a Sud evidenzia la presenza di regioni di transizione. L'aggregazione di un numero maggiore di classi è indice del fatto che le S.O.M. tendono a cogliere aspetti più particolari e specializzati rispetto al metodo precedentemente descritto. L'assenza dell'individuazione di casi isolati manifesta la capacità di mantenere un certo grado

di generalizzazione. La superficie definita dalla S.O.M. tende a definire la costruzione di un'area in cui si concentrano configurazioni anomale, andamenti stimati inconsueti nella variabilità delle caratteristiche principali del collettivo.

Ciò che emerge evidente da queste considerazioni è che il principale contributo delle S.O.M. alla classificazione non deriva dalla mera individuazione di classi, fatto sicuramente degno di approfondimenti ulteriori, ma soprattutto dalla capacità esplorativa che questi strumenti manifestano nei risultati ottenuti nell'ottica della riduzione delle variabili. Ciò sottolinea nuovamente la potenzialità dell'analisi neurale nell'individuazione di elementi determinanti al fine della conoscenza dei sistemi territoriali su qualsiasi piano: economico, politico o sociale. I caratteri del luogo vengono esaminati nel contesto, nell'intreccio di relazioni che sono coinvolte nelle trasformazioni e che descrivono le posizioni assunte dalle osservazioni. Su questi assunti, si costruiscono altri aspetti delle capacità esplorative delle S.O.M., vediamo alcuni.

3.2 L'analisi dello spazio definito dalla S.O.M.

Si è accennato, in precedenza, che le capacità di Feature-Mapping della S.O.M. si rivelano nella costruzione di uno spazio bidimensionale in cui è possibile collocare le osservazioni a dimensionalità più elevata oggetto di studio (Figura 3.3). La fase di interrogazione standard della S.O.M. associa ogni elemento del collettivo al punto che ne rappresenta meglio la posizione all'interno del piano organizzato. Ma questo non significa che non esistano altri punti del piano che si avvicinino molto alla possibilità di essere associati a quel sistema pur non essendo eletti come miglior rappresentare. Ad ogni punto può essere infatti associato una certa *“capacità di rappresentazione”* legata alla propria posizione e inversamente proporzionale alla distanza tra il vettore caratteristico dell'osservazione e il vettore dei pesi associato al Processing Element che è legato alla posizione del punto. Infatti potremo definire *“incapacità rappresentativa”*, del punto r in relazione all'osservazione associata al vettore X , il valore

$$D_r = \|X - W_r\| = \sqrt{\sum_{i=1}^n (x_i - w_{ri})^2} \quad (3.1).$$

Tale misura indica quindi in quali altre zone del piano una certa osservazione potrebbe tendere a collocarsi. Nella Tabella 3.1 possiamo vedere nello spazio (40 punti) x (40 punti) quali possono essere le zone alternative di collocazione della Provincia di Caltanissetta e in Figura 3.3 il relativo grafico.

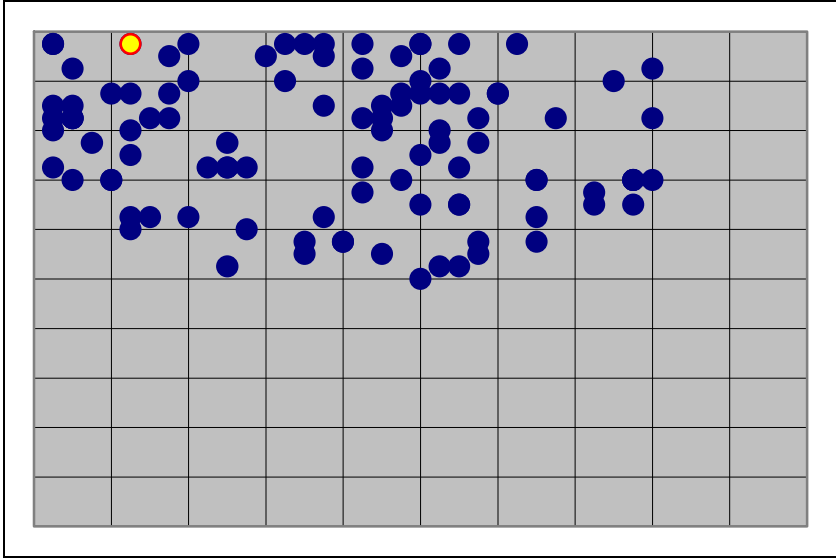
Si noti l'esistenza di minimi locali e zone intere con diversi livelli di rappresentatività per la Provincia in esame. Nel caso della Provincia di Caltanissetta, emerge una zona particolarmente buona sotto questo aspetto in prossimità della propria posizione, come era

lecito presupporre, ed in colore chiaro un'ampia porzione del piano a capacità rappresentativa costante.

Tabella 3.1 – Tabella della misura dell'incapacità rappresentativa di una porzione di punti

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1,93	2,05	2,06	1,97	1,87	1,85	1,88	1,94	2,02	2,13	2,27	2,38	2,33	2,29	2,32	2,44	2,63	2,86	3,12	3,39
2	1,93	2,06	2,26	2,01	1,90	1,88	1,92	1,96	2,03	2,15	2,34	2,62	2,39	2,29	2,28	2,36	2,50	2,70	2,94	3,20
3	1,89	1,97	2,01	1,95	1,91	1,91	1,94	1,98	2,04	2,14	2,26	2,36	2,32	2,28	2,28	2,33	2,43	2,59	2,79	3,03
4	1,86	1,90	1,93	1,95	1,97	1,98	1,99	2,01	2,05	2,11	2,18	2,23	2,25	2,27	2,30	2,35	2,42	2,54	2,70	2,90
5	1,87	1,92	1,98	2,06	2,11	2,10	2,06	2,05	2,07	2,11	2,16	2,21	2,26	2,31	2,36	2,41	2,47	2,55	2,66	2,82
6	1,93	1,99	2,09	2,23	2,34	2,25	2,15	2,11	2,11	2,14	2,20	2,26	2,32	2,39	2,45	2,50	2,55	2,61	2,69	2,79
7	2,01	2,06	2,16	2,35	2,65	2,36	2,20	2,15	2,16	2,21	2,27	2,34	2,42	2,50	2,57	2,61	2,67	2,73	2,78	2,83
8	2,10	2,12	2,18	2,29	2,37	2,29	2,21	2,20	2,23	2,30	2,37	2,45	2,52	2,60	2,69	2,72	2,79	2,88	2,94	2,93
9	2,22	2,21	2,21	2,23	2,25	2,24	2,24	2,28	2,35	2,43	2,51	2,58	2,63	2,69	2,74	2,81	2,92	3,07	3,17	3,09
10	2,42	2,37	2,33	2,31	2,30	2,32	2,37	2,44	2,52	2,61	2,68	2,73	2,77	2,80	2,84	2,91	3,01	3,19	3,50	3,20
11	2,69	2,63	2,57	2,53	2,52	2,54	2,60	2,67	2,75	2,82	2,88	2,92	2,93	2,94	2,96	3,00	3,06	3,16	3,22	3,15
12	3,02	2,97	2,90	2,85	2,83	2,85	2,89	2,95	3,01	3,06	3,09	3,10	3,09	3,08	3,08	3,10	3,11	3,13	3,13	3,09
13	3,38	3,31	3,23	3,17	3,14	3,15	3,19	3,23	3,27	3,30	3,30	3,28	3,26	3,24	3,21	3,21	3,20	3,18	3,14	3,11
14	3,72	3,64	3,53	3,44	3,40	3,41	3,44	3,48	3,52	3,54	3,52	3,49	3,45	3,39	3,34	3,31	3,30	3,25	3,21	3,17
15	4,09	3,96	3,79	3,66	3,60	3,60	3,65	3,71	3,75	3,76	3,73	3,68	3,62	3,55	3,48	3,41	3,37	3,34	3,31	3,27

Figura 3.3 – Grafico relativo alla distribuzione effettuata dalla S.O.M.



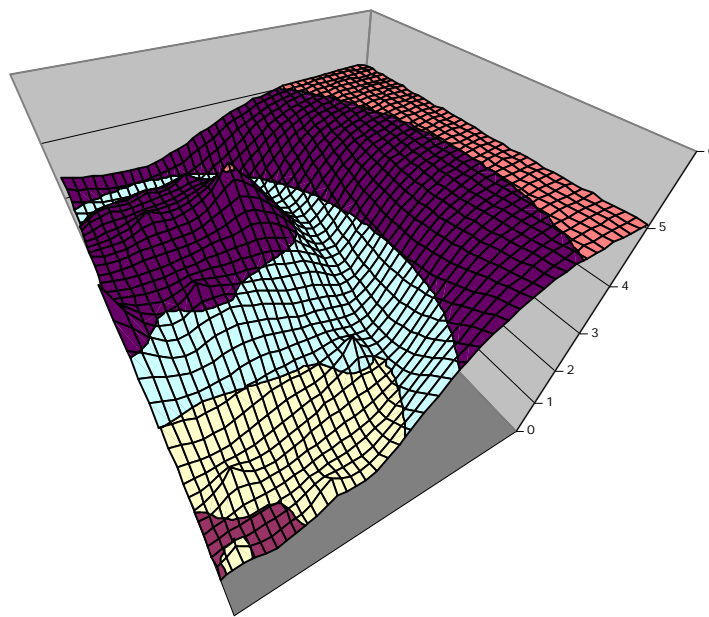
Tra le Province presenti nella prima area, troviamo che le più distanti in termini di posizione sono le Province di Matera e Potenza, casi di Province che si prestano ad un confronto con quella in esame. L'area in colorazione più chiara rappresenta, invece, un'insieme di punti che hanno lo stesso grado di rappresentatività nei confronti di Caltanissetta, e ciò indica che, se

Caltanissetta riuscisse a raggiungere una configurazione presente in quell'area, probabilmente si troverebbe poi in una situazione di equilibrio dalla quale sarebbe difficile spostarla.

3.3 *Cammino di trasformazione, interrogazione virtuale e simulazione*

E' lecito supporre che lo spazio definito dalla S.O.M. coincida con la proiezione di una realtà n -dimensionale sul piano; ossia che ad ogni punto del piano sia associato un vettore n -dimensionale che rappresenta uno scenario, una configurazione di un sistema locale.

Figura 3.3 – Grafico relativo all'incapacità rappresentativa del caso di Caltanissetta

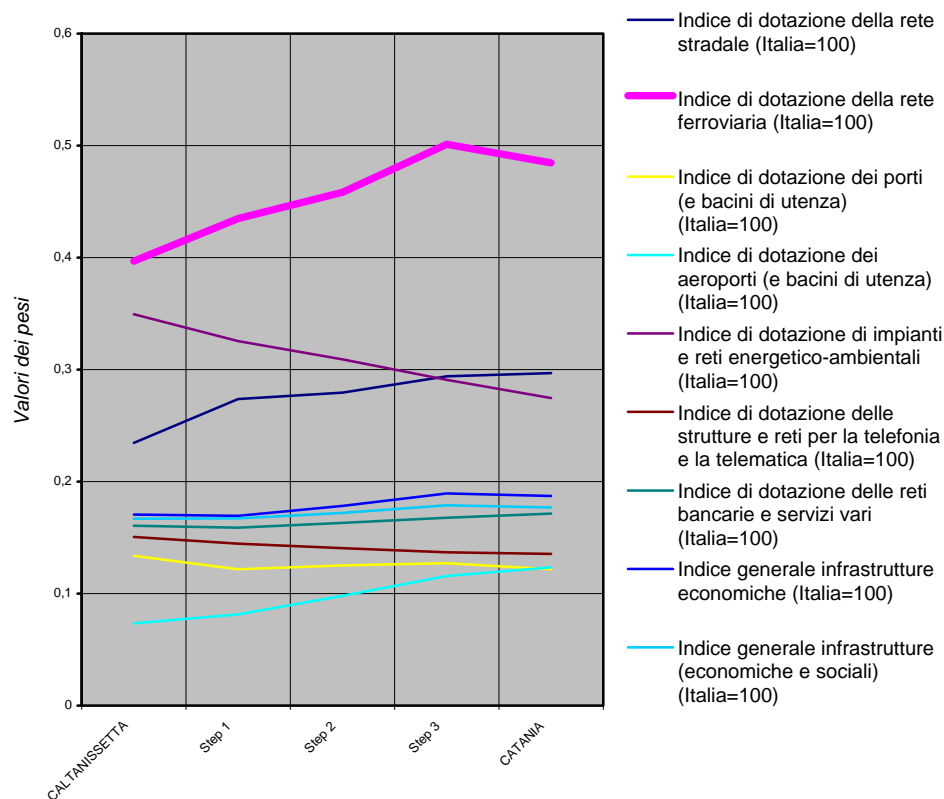


Punti contigui, per il principio dell'auto-organizzazione su cui si fonda l'algoritmo, sono associati a sistemi territoriali simili tra loro. Sulla base di queste considerazioni, l'insieme di punti che giacciono sul segmento congiungente due sistemi dati individuati in due posizioni distinte nel piano definito dalla S.O.M., rappresentano una successione di passaggi che indicano il cambiamento graduale tra un assetto ed un altro. Consideriamo il caso del passaggio tra la Provincia di Caltanissetta e la Provincia di Catania, caso che si presenta nel cammino verso la zona di equilibrio. Per ragioni di leggibilità sono state riportate nel grafico di Figura 3.4 solamente le variabili riguardanti il settore delle dotazioni infrastrutturali.

Come è possibile vedere in questo grafico, evidenziato dalla linea ispessita è rappresentato l'andamento Step by Step dell'Indice di dotazione della rete ferroviaria, scelto, a titolo esemplificativo del metodo di lettura del grafico. Dal confronto tra i valori del punto di partenza e di quello di arrivo, si può constatare come la Provincia di Caltanissetta si trovi in

una condizione di sottodotazione rispetto al caso Catanese. Dalla lettura del cammino di trasformazione di Caltanissetta possiamo leggere, però che tale indice necessita di un incremento forte per l'avvio nei primi steps, ma successivamente, per l'ultimo passaggio, manifesta un decremento. Il significato di tale andamento è associato all'importanza che il fattore trasporto e comunicazione su strade ferrate assume per una prima fase di rilancio della Provincia in esame, ma sottolinea anche che la dotazione di rete ferroviaria non è sufficiente al raggiungimento di una posizione di competitività rispetto alla Provincia di Catania. Si rendono infatti necessari altri interventi che nell'ultimo passaggio di transizione assumono un peso maggiore.

Figura 3.4 – Grafico relativo al cammino di trasformazione tra Caltanissetta e Catania



Individuati i margini di intervento sul sistema, è possibile definire delle linee strategiche atte all'avvio del sistema oggetto di intervento sul percorso di sviluppo scelto. Al fine di verificare la bontà delle scelte effettuate, è possibile formulare la descrizione di una osservazione virtualmente modificata in termini vettoriali e proporla in ingresso alla S.O.M. al fine di verificare che il riassetto programmato per il sistema locale giaccia sul cammino prestabilito. Ciò permette di effettuare valutazioni preventive, e quindi simulazioni, e consente di attestare e validare obiettivi teorici con possibile attinenza alla realtà.

4 CONCLUSIONI

Il fatto che la competitività si esprima in termini di vantaggi assoluti anziché relativi, implica la persistenza del rischio concreto di vedere condannate al declino intere porzioni del territorio, a causa dell'esclusione dai processi di globalizzazione (Camagni, 2002). Si rende quindi necessaria una rielaborazione di gran parte degli approcci operativi che riguardano l'azione politica locale allo scopo interagire sulle *precondizioni* di natura sociale, economica e infrastrutturale. In sostanza, gli sforzi della pianificazione vanno concentrati su di un innalzamento complessivo della qualità territoriale favorendo in tal modo la crescita del livello di competitività locale.

Migliorare la posizione di competitività di un territorio implica quindi un'azione strategica strutturata su un insieme di variabili che si pongono tra loro in un rapporto di interagenzia, che il più delle volte resta a livello latente. Il risultato finale di tali processi interattivi assume forme sinergiche che si vengono a stabilire tra elementi diversi quali: capitale umano e sociale, conoscenza tacita, reticoli relazionali, rapporti di complementarietà e cooperazione.

In un contesto quale quello appena illustrato, occorre allora dotarsi di strumenti analitici in grado di fare emergere "i nessi impliciti" che si vengono a stabilire tra i diversi elementi e che contribuiscono ad innalzare la soglia di complessità del sistema territoriale.

In questa logica il presente studio ha cercato di testare le potenzialità analitiche offerte dalle reti neurali, ed in modo particolare da quelle S.O.M., mettendo in luce l'aspetto applicativo e proponendo metodi di indagine operativa. Dallo studio condotto emergono elementi di sicuro interesse, ma al contempo anche alcune criticità sulle quali occorre condurre alcune riflessioni.

Non avendo a disposizione serie storiche di dati sufficienti a definire un processo evolutivo e auto-organizzativo del territorio, non è ancora stato possibile riprodurre in modo efficace il percorso dinamico di apprendimento collettivo, di organizzazione politica e culturale che si colloca alla base della differenziazione qualitativa dei territori. Ciononostante, chi scrive è dell'opinione che le S.O.M. potrebbero avere un ruolo determinante nella simulazione dei processi auto-organizzativi del territorio.

Una questione aperta rimane quella sulla confrontabilità delle regioni oggetto di analisi. Se da un lato le S.O.M. ridefiniscono una distribuzione in uno spazio a dimensionalità ridotta (rispetto a quello identificato dal numero delle variabili prese in considerazione) concedendo la possibilità di avanzare nuove considerazioni sulle osservazioni studiate, dall'altro non risolvono i problemi che nell'aggregazione effettuata con la Cluster sono emersi quali: la scelta del Livello di Dissomiglianza Accettabile o la scelta, in alternativa, del numero di classi, le aggregazioni sufficienti a definire i membri del collettivo.

Si noti inoltre che esiste un legame tra il numero dei gradi di libertà che il sistema studiato possiede e la dimensionalità dello spazio definito dalla S.O.M.. Questo fattore, riferito al

territorio, rappresenta un limite che va considerato nel momento in cui l'analista si accinge alla lettura dei risultati. Si pensi al territorio come ad un sistema pluridimensionale e ai gradi di libertà, vincolati oppure no, cui è concesso di auto-organizzarsi. Ridurne la dimensionalità può produrre un modello affine e attinente con la realtà, ma il livello di approssimazione è necessariamente alto (anche se decisamente superiore a quello ottenibile con l'applicazione di tecniche tradizionali di analisi). Questo dimostra che lo strumento pur dimostrandosi utile e flessibile, non permette di prescindere dal livello euristico dell'analisi; livello che può essere garantito esclusivamente dall'esperienza e dalla professionalità del Planner.

Bibliografia

- Bolasco S., (1999); *Analisi Multidimensionale dei Dati. Metodi, strategie e criteri*, Carocci, Roma.
- Breda M., (1999); *Self-Organizing Maps* in Buscema M. (a cura di), (1999a); *Reti Neurali Artificiali e Sistemi Sociali Complessi. Teoria – Modelli - Applicazioni*; Volume I: - *Teoria e Modelli*; Franco Angeli, Milano.
- Buscema M. (a cura di), (1999a); *Reti Neurali Artificiali e Sistemi Sociali Complessi. Teoria – Modelli - Applicazioni*; Volume I: - *Teoria e Modelli*; Franco Angeli, Milano.
- Buscema M. (a cura di), (1999b); *Reti Neurali Artificiali e Sistemi Sociali Complessi. Teoria – Modelli - Applicazioni*; Volume II: - *Applicazioni*; Franco Angeli, Milano.
- Buscema M., Diappi L., Stabilini S., (1995), The neural Vision of a competitive City: an exploration in the european Context, *Proceedings of the IX Colloquium of Theoretical Geography*, Spa
- Camagni R. (2002), *On the concept of territorial competitiveness: sound or misleading?* Urban Studies, Vol. 39, n° 13, Pagg. 2395-2411, Dic. 2002.
- Campos Venuti G., Federico Oliva F. (2000); Relazione introduttiva, in (a cura di) Servizio urbanistica e Pianificazione Territoriale dell'Amministrazione Provinciale di Pesaro Urbino, *Note Introduttive al P.T.C. e regole e criteri per la copianificazione*; Piano Territoriale di Coordinamento della Provincia di Pesaro e Urbino.
- Diappi L., Bolchi P., Franzini L., (1999), GIS e complessità urbana: un approccio neurale, in Besio M., Monti C. (a cura di); *Dal cannocchiale alle stelle, strumenti per il nuovo piano*; Franco Angeli, Milano; pp 180-200.
- Fischer, M.M., Gopal S. (1994). Artificial Neural Networks: A New Approach to Modelling Interregional Telecommunication Flows. *Journal of Regional Science* 34(4), 503-27.
- Fraire M., (1985), Le fasi dell'analisi dei dati, in Rizzi A., *Analisi dei dati. Applicazioni dell'informatica alla statistica*; NIS, Roma, pp. 211- 227
- Kohonen, T. (1982), Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43: pp.59-69.
- Kohonen, T. (1991), Self-organizing maps: optimization approaches. In Kohonen, T., Mäkisara, K., Simula, O., and Kangas, J., (Eds.), *Artificial Neural Networks. Proceedings of ICANN'91, International Conference on Artificial Neural Networks*, Volume II, pp. 981-990, North-Holland, Amsterdam.
- Kohonen, T. (1995), *Self-Organizing Maps*. Springer, Berlin.
- Leti G., (1979), *Distanze e indici statistici*, La Goliardica, Roma.
- Ocelli S., Novelli E. (1999), *Profili Descrittivi di Distribuzioni Spaziali: alcune Misure di Diversificazione*. Atti della XIX Conferenza Nazionale dell' Associazione di Italiana di

Scienze Regionali (AISRE). A.I.S.Re., Milano. L'articolo è presente anche in: *Cybergeo*, 108, 1999.

Rizzi A., (1985), *Analisi dei dati. Applicazioni dell'informatica alla statistica*; NIS, Roma,

Virgilio G., (2002), *Nuovi strumenti per i processi di piano: vantaggi e criticità dei sistemi spaziali di supporto alle decisioni*. In (a cura di I.Pi.Ge.T. – C.N.R. Napoli) Atti della XXIII Conferenza Nazionale dell' Associazione di Italiana di Scienze Regionali (AISRE). A.I.S.Re., Milano.

ABSTRACT

Complexity, which is a property of local systems, makes extremely difficult the recognition of the critical elements and, for this matter, it shows the limits in provisioning effects of strategic choices in land planning. Traditional methods are inadequate in researching of “implicit connections”. This concept is on the basis of the really close interactions which associate the complexity to spatial-planning. It needs new models in which developed analytical tools are used “to capture” the interactivity of a local system and his ability in self-organizing. This analysis assumes that this process of self-organization is moved by the choices done to improve local absolute competitiveness, and is promoted by the need of maintaining an equilibrate position in relation with the other components of the system. The Neural Networks based on S.O.M. (Self-Organizing Map) algorithm has shown good results in pattern recognition and in the study of the relationships between the different regional realities, specifically the Italy’s Provinces. They revealed able to identify homogeneous classes applying a cluster analysis hierarchical algorithm as a front-end of the process. This fact underlines the importance of the integration between traditional statistical methods and the relatively new tools developed in neuro-science.