

Il ruolo della formazione universitaria ed extrauniversitaria: la profilazione degli studenti partecipanti ad un esperimento di mappatura

Gian Pietro Zaccomer¹

¹ Dipartimento di Lingue e letterature, comunicazione, formazione e società (DILL),
Università degli Studi di Udine, gianpietro.zaccomer@uniud.it

Abstract. All'interno una batteria di esperimenti controllati – condotta presso l'Università di Udine tramite un geoquestionario predisposto online – si è chiesto di individuare alcuni luoghi della regione Friuli Venezia Giulia. In questo contributo viene proposta una profilazione di poco più di 200 studenti coinvolti ottenuta grazie all'impiego di una procedura di *clustering* per variabili miste e costruita al fine di capire come la formazione acquisita influisca sulla loro capacità di portare a termine il compito di mappatura assegnato.

Parole chiave: Geoquestionari, Friuli Venezia Giulia, Cluster Analysis.

1 Il quadro di riferimento

L'informazione geografica, ossia quella accompagnata da una sua precisa posizione sulla superficie terrestre, ha assunto sempre più importanza non solo a livello di pubblica amministrazione, sia a supporto delle decisioni – integrando spesso il sistema informativo preesistente – sia per una più oculata pianificazione territoriale, ma anche nella sfera privata grazie alla larghissima diffusione dei moderni strumenti digitali che accompagnano le persone nella vita di tutti i giorni. Mai come oggi vi è stata una così ampia conoscenza delle geografie del mondo reale, siano esse anche semplicemente circoscritte ai ristoranti, alle stazioni di rifornimento o ai propri itinerari di viaggio da percorrere con forme di mobilità lenta. Infatti, non di rado tali strumenti implementano funzioni di segnalazione volontaria atti a creare mappe personali per poter organizzare le proprie attività ludiche o professionali. Ma tutti gli utenti sono capaci di sfruttare in ugual modo queste possibilità o il loro impiego è condizionato in una certa misura dalle proprie conoscenze, in particolare da quelle geografiche e informatiche?

Rispetto ad una domanda di ricerca così ampia, questo lavoro focalizza l'attenzione sullo studio del ruolo esercitato dalla formazione, sia universitaria che extrauniversitaria, nella capacità di portare a termine degli esercizi di mappatura proposti a studenti di differenti insegnamenti impartiti presso l'Università di Udine: l'obiettivo è quindi quello di verificare empiricamente se e come la formazione posseduta si inserisca nel processo di composizione dei profili dei partecipanti agli esperimenti e, più in generale, come influisca sulla loro capacità di portare a termine

operazioni di mappatura, azioni che oggi stanno alla base delle funzioni implementate nei moderni strumenti digitali a partire dagli onnipresenti smartphone.

1.1 La linea di ricerca sul rischio e sul degrado del paesaggio

Prima di indirizzare l'attenzione sulla specifica applicazione oggetto di questo lavoro, appare necessario qualificarla come ultimo approfondimento di un più ampio percorso di ricerca derivato dalla diretta partecipazione dell'Università di Udine – tra il 2014 e il 2018 – alla stesura del Piano Paesaggistico Regionale (PPR) del Friuli Venezia Giulia (FVG)¹ il quale ha permesso ai ricercatori coinvolti di intraprendere diverse linee di ricerca [1].

Per quanto riguarda il gruppo dei geografi udinesi, questo ha dato il via, terminato l'impegno del PPR, al “progetto dipartimentale PaRiDe” dedicato a *I paesaggi del rischio e del degrado: dalla percezione, alla rappresentazione e alla territorializzazione. Saperi interdisciplinari e consapevolezze a supporto delle politiche di governo del territorio*. Una delle caratteristiche peculiari di questo progetto era quella di ricorrere a diversi approcci abbracciando una prospettiva multidisciplinare che, oltre a coinvolgere le competenze più strettamente geografiche, faceva anche ricorso a quelle tecnologico-informatiche e a quelle socio-economiche. Se da questo ultimo punto di vista, l'attività di ricerca si è prima di tutto concentrata sull'applicazione degli indicatori compositi in ambito paesaggistico (lavoro raccolto nel volume [2] e, con una chiave di lettura più specifica, in [3]), successivamente l'analisi statistica ha supportato tutte le fasi della ricerca geografica che, con pieno ricorso alle nuove tecnologie, ha coperto un ampio spettro di tematiche riguardanti il rischio e il degrado regionale.

I lavori iniziali hanno riguardato le prime riflessioni sul legame tra la creazione di nuova informazione e la comunicazione della conoscenza prodotta attraverso la cartografia [4], l'analisi preliminare delle potenzialità e dei limiti dell'impiego di mappe integrate in questionari online (da ora in poi richiamati, più brevemente, con il termine di *geoquestionari*) [5], e le possibili azioni che i soggetti, a diverse scale, possono intraprendere per contrastare il degrado [6]. Il quadro d'insieme del progetto PaRiDe è stato restituito nel volume [7] che ha approfondito gli argomenti nell'intento di produrre nuove informazioni e mappe dei luoghi del FVG interessati da situazioni di rischio e degrado. In questo saggio, dopo alcune riflessioni sui concetti teorici di base e una disamina dell'attenzione al degrado nelle politiche regionali, per supportare la fase empirica della ricerca sono state progettate, e portate a termine, quattro diverse indagini che hanno fatto sia uso di *geoquestionari*, ma anche di uno specifico concorso fotografico a cui è seguita una serie di interviste agli autori degli scatti in modo da approfondire alcuni aspetti soggettivi, più qualitativi rispetto a quanto ottenuto dalla mera compilazione dei *geoquestionari*, sulla percezione del degrado.

¹ Tutti i documenti ufficiali del PPR possono essere scaricati dal sito ufficiale: www.regione.fvg.it/rafvfg/cms/RAFVG/ambiente-territorio/pianificazione-gestione-territorio/FOGLIA21/. Ultima visita: 2 maggio 2022.

Alla fine del 2019, il progetto PaRiDe ha visto una sua propaggine attraverso un'ulteriore ricerca dedicata a *La Volunteered Geographic Information (VGI) come strumento per la mappatura e l'analisi dei paesaggi del rischio e del degrado del Friuli Venezia Giulia* che ha permesso di completare e approfondire alcuni aspetti lasciati in sospeso anche se lo scoppio della pandemia ha costretto ad una parziale rimodulazione dei suoi obiettivi. Quindi, sempre nell'intento di studiare i paesaggi del rischio e del degrado, l'attenzione è stata rivolta alle potenzialità della VGI ([8], [9]) e, soprattutto, alle caratteristiche qualitative dei dati da essa prodotti [10]. Infatti, analizzando quelli precedentemente raccolti durante la ricerca, ci si è sempre preoccupati della loro qualità intrinseca chiedendosi se gli utenti siano sempre in grado di fornire segnalazioni di buona qualità utilizzando le mappe interattive online: un primo lavoro di valutazione della loro accuratezza posizionale si trova in [11], ma su questo aspetto si tornerà nel prossimo paragrafo poiché collegato a quanto qui proposto.

Per completare la rassegna dell'ampio lavoro portato a termine in questa linea di ricerca, si segnala tra il 2020 e il 2022 sono stati conclusi ulteriori approfondimenti, alcuni dei quali a maggiore orientamento quantitativo come lo studio dell'influenza della residenza del rispondente sugli elementi mappati sfruttando un approccio statistico non parametrico [12] e l'analisi della percezione del rischio attraverso una tecnica multivariata non lineare [13]. Altri lavori hanno preferito privilegiare un taglio più qualitativo come l'esame degli aspetti geografici del confronto tra il sapere esperto e quello contestuale con particolare riferimento al degrado delle strutture e delle aree militari, all'abbandono e al dissesto idraulico nonché ai siti di cava [14] e, più recentemente, il vaglio delle potenzialità della produzione fotografica personale, soprattutto amatoriale, quale strumento per raccogliere e studiare le percezioni dei cittadini sul degrado del paesaggio [15].

1.2 La batteria di esperimenti controllati

All'interno della ricerca dedicata alla VGI, sulla base delle considerazioni espresse in [11] e seguendo le idee proposte in [16], durante l'anno accademico 2019-20 presso l'Università degli Studi di Udine è stata condotta una batteria di esperimenti controllati in cui si richiedeva ai partecipanti in forma di esercizio – sempre attraverso mappe interattive integrate in un questionario predisposto sfruttando il servizio online Enketo che supporta la cartografia di OpenStreetMap – sia di individuare la posizione di tre diversi luoghi presenti in regione predeterminati dai ricercatori (nella fattispecie: il Castello di Miramare vicino a Trieste, la laguna di Grado e di Marano e il fiume Tagliamento), sia di indicare su base volontaria altri due luoghi in FVG, uno degradato e uno di particolare bellezza, per un totale di cinque esercizi.

Il fatto di aver imposto a priori agli studenti quali luoghi individuare sulla mappa interattiva (attraverso un punto, una linea o un poligono a propri scelta) ha permesso di studiare, oltre alla composizione demografica e alle caratteristiche del loro percorso di studi, l'accuratezza posizionale degli esercizi proposti. I risultati dell'analisi univariata, comprensiva della valutazione dell'accuratezza, sono stati pubblicati in [17] a cui si rimanda direttamente per ogni approfondimento.

Tabella 1. Lista delle variabili contenute nel *dataset*.

Variabile	Descrizione
X_1	Corso in presenza o online
X_2	Genere maschile o femminile
X_3	Studente a tempo pieno o lavoratore
X_4	Residenza in o fuori il Friuli Venezia Giulia
X_5	Conoscenza geografica del Friuli Venezia Giulia: comuni ben conosciuti (numero)
X_6	Frequenza (attuale o passata) di Corsi Informatici Universitari (CIU)
X_7	Frequenza (attuale o passata) di Corsi Informatici Extrauniversitari (CIE)
X_8	Ore al giorno su computer o smartphone
X_9	Ore al giorno su Internet
X_{10}	Utente GIS di qualsiasi tipo: sì o no
X_{11}	Sviluppo di sito Internet: sì o no
X_{12}	Gestione di un blog: sì o no
X_{13}	Contribuzione a piattaforme collaborative a contenuto libero: sì o no
X_{14}	Abilità GIS (numero)
X_{15}	Abilità informatiche (numero)
X_{16}	Condivisione della posizione geografica delle proprie fotografie: sì o no
X_{17}	Condivisione propria posizione geografica: sì o no
X_{18}	Utilità della condivisione delle informazioni geografiche via rete: accordo o disaccordo
X_{19}	Conoscenze universitarie GIS: Livello I (nessuna conoscenza)
X_{20}	Conoscenze universitarie GIS: Livello II (conoscenza solo teorica)
X_{21}	Conoscenze universitarie GIS: Livello III (conoscenza teorica e pratica)
X_{22}	Numero di esercizi completati (escluse le segnalazioni libere): da 0 a 3
X_{23}	Numero di esercizi completati (comprese le segnalazioni libere): da 0 a 5

Sulla base della valutazione proposta in [17], si è poi cercato di verificare empiricamente se sulla capacità di portare a termine gli esercizi proposti avesse influito non solo la volontà degli studenti di cimentarsi con l'utilizzo delle mappe interattive, ma anche le loro conoscenze pregresse, sia teoriche che pratiche, sui *Geographic Information Systems* (GIS). Infatti, questi strumenti presentano difficoltà intrinseche che, per ipotesi, vengono più facilmente superate disponendo almeno di una conoscenza di base dei GIS appresa nei corsi universitari. I risultati ottenuti sono stati presentati nel successivo lavoro [18] che rappresenta il primo tentativo del progetto PaRiDe che mira ad indagare la capacità degli studenti universitari nel portare a termine esercizi che implicano la mappatura dei luoghi attraverso strumenti interattivi.

Ricapitolando, se il *dataset* di Tabella 1 è già stato studiato soprattutto dal punto di vista univariato, focalizzando maggiormente l'attenzione sull'accuratezza posizionale, e dal punto di vista della dipendenza lineare, assumendo come variabile dipendente il numero di esercizi complessivi portati a termine X_{23} , in questo lavoro si intende concludere la sua fase di analisi quantitativa mantenendo ancora un'ottica multivariata, ma impiegando questa volta tecniche utili per classificare, sempre dal punto di vista statistico, gli studenti in specifici *gruppi*, o *cluster*, a cui corrispondono precisi *profili* in termini di caratteristiche, in particolar modo di quelle relative alle loro conoscenze geografiche e informatiche.

2 La profilazione degli studenti

2.1 Un problema di natura multivariata

Per portare a termine un compito di profilazione è prassi ricorrere alla *Cluster Analysis*, ossia ad un insieme piuttosto ampio di tecniche di analisi multivariata volte alla selezione e al raggruppamento delle unità statistiche (in questo caso gli studenti universitari), appartenenti a un campione o a una popolazione, che risultano omogenee – secondo uno specifico criterio – rispetto ad un determinato *dataset* o a una parte di esso [19].

Nel caso qui affrontato esiste una caratteristica dell'insieme dei dati considerato che limita notevolmente le possibilità di scelte tra le diverse tecniche. Infatti, come si osserva facilmente dalla Tabella 1, il *dataset* costruito presenta variabili di natura diversa, sia qualitative nominali e ordinali riferite a modalità di risposta testuali, sia quantitative discrete derivanti da conteggi.

Per risolvere questo tipo di problema, piuttosto comune soprattutto per le necessità di *customer segmentation* del marketing, sulla base del *framework* sviluppato per l'algoritmo BIRCH [20], ideato per ampi *dataset*, in [21] è stato proposto un approccio integrato poiché sia la metrica utilizzata sia l'algoritmo di classificazione fanno riferimento ad un modello probabilistico che permette di considerare simultaneamente variabili di diversa natura. Per ottenere le migliori prestazioni da questo approccio è necessario introdurre alcune ipotesi sulle distribuzioni delle variabili di diversa natura, ma soprattutto quella di indipendenza tra le stesse.

Tale proposta è stata implementata nel software *Statistical Package for Social Science* (SPSS), successivamente acquisito da IBM, attraverso una procedura denominata *TwoStep Cluster Analysis* (TSCA) poiché questa si suddivide in due passi sequenziali [22]. Il primo, definito di *pre-cluster*, riguarda le dimensioni del *dataset* da trattare in quanto serve esclusivamente per scalare il problema di classificazione qualora il *dataset* risulti molto ampio come quelli derivanti dal mondo dei *big data*. Il secondo passo, quello di *clustering* vero e proprio, sfrutta una procedura gerarchica ascendente che, a differenza di quelle classiche, presenta la comodità di individuare automaticamente il numero "ideale" di gruppi; infatti, se nelle procedure gerarchiche classiche la scelta del numero dei gruppi è lasciata all'esperienza del ricercatore, la seconda fase della TSCA permette di individuare il numero ideale di gruppi sfruttando inizialmente l'*Akaike's* (AIC) o, a scelta, il *Bayesian Information Criterion* (BIC) e, successivamente, l'entità della variazione della distanza tra gruppi vicini nell'ipotesi di una loro aggregazione.

Rimane sempre valida la possibilità di impostare manualmente un numero diverso di gruppi, rispetto a quello ideale, poiché spesso diventa necessario considerare, ai fini della ricerca dei profili tipo, un differente livello di aggregazione: in questo caso la soluzione ideale può essere fruttata come punto di partenza per la ricerca di una soluzione pragmatica più facilmente interpretabile dal punto di vista del profilo ricercato, ma che peggiora la qualità dell'operazione di classificazione.

Per completare, dal punto di vista teorico, l'introduzione alle problematiche legate all'impiego della TSCA è necessario tornare al problema delle ipotesi richieste

all'inizio: la manualistica del software [23] afferma che, dato che tali condizioni si presentano difficilmente nella realtà, la procedura implementata nel software si comporta "ragionevolmente bene" anche nel caso di violazioni delle ipotesi, ma viene sempre lasciata alla sensibilità del ricercatore la valutazione dei risultati, ossia il giudicare se quelli ottenuti sono soddisfacenti rispetto alle esigenze di classificazione.

Dal punto di vista pratico è sempre meglio cercare di porsi il più vicino possibile ai dettami teorici: a riguardo del mancato rispetto dell'assunzione sulle distribuzioni poco si può fare poiché i dati sono quelli rilevati con il geoquestionario, invece per quanto riguarda l'ipotesi di indipendenza, a quest'ultima ci può più facilmente avvicinare attraverso una buona selezione delle variabili da considerare soprattutto qualora il *dataset* presenti, al suo interno, un elevato grado di multicollinearità.

2.2 La selezione delle variabili

Quanto appena detto pone il problema di quali variabili individuare per la profilazione, decisione che non può essere delegata all'algoritmo di *clustering* e che, quindi, deve essere basata su un'approfondita analisi della connessione e della correlazione tra le variabili del *dataset*.

Questo tipo di analisi è già stata condotta in [18] che ha messo in evidenza la forte presenza di una ampia multicollinearità all'interno del *dataset* qui utilizzato, ma ora è necessario riconsiderare singolarmente quali sono le variabili fortemente legate tra loro e, quindi, quali devono essere escluse e quali invece andranno a costituire il *subset* da sottoporre alla TSCA.

Proprio nei precedenti lavori [17] e [18] si era già considerato l'influenza dei singoli livelli di conoscenza universitari in ambito GIS sul numero di esercizi svolti, analisi che aveva soprattutto messo in rilievo l'importante ruolo del terzo livello di conoscenza (Liv. III) di tipo sia teorico che pratico. Date le caratteristiche della TSCA, per questo lavoro le variabili dicotomiche X_{19} , X_{20} e X_{21} sono state collasate in un'unica variabile ordinale *LEVEL* che permette di ridurre la dimensione del *dataset* da 23 a 21 variabili, ma le sue relazioni con le altre variabili vanno riconsiderate alla luce di tale trasformazione.

L'analisi delle correlazioni tra variabili quantitative e di connessione tra le variabili qualitative ha messo in relazione il complesso sistema di relazioni tra le stesse. Sulla base di questo sistema sono state selezionate cinque variabili non solo slegate tra loro, ma anche ritenute significative ai fini di profilazione qui cercata.

Prima di tutto, per quanto riguarda quelle qualitative, sono state selezionate: *LEVEL* la variabile di livello delle conoscenze universitarie GIS, X_4 il cui valore unitario indica se lo studente è residente in FVG, mentre per X_7 segnala la frequenza, attuale o pregressa, di *Corsi Informatici Extrauniversitari* (CIE). Queste tre variabili qualitative sono risultate non connesse tra loro al 5% secondo il test del chi-quadrato a due code. Guardando al significato di queste tre variabili, si osserva che se la prima e la terza sono legate alla *formazione universitaria e extrauniversitaria* dello studente, la seconda condiziona fortemente la *conoscenza geografica* del FVG del rispondente al geoquestionario poiché essa risulta direttamente legata alla X_5 . Infatti, come già evidenziato in [17], gli studenti di fuori regione conoscono prevalentemente solo il

comune dove è situata la sede universitaria (in questo caso Udine o Gorizia), mentre quelli residenti ne conoscono in media circa 5,5 in più.

Accanto a queste prime tre variabili qualitative ritenute importanti per la profilazione, è stata selezionata, poiché non connessa alle precedenti, la X_{16} che riguarda invece un comportamento dello studente, ossia quello di condividere la posizione geografica delle proprie fotografie postate sui social. Questo tipo di informazione è sicuramente influenzata da altre considerazioni non indagate esplicitamente dal geoquestionario, come quelle legate al rispetto della *privacy*, ma potrebbe tuttavia esprimere una minima capacità nel saper sfruttare le possibilità offerte dalle nuove tecnologie. Pur non essendo questa variabile così rilevante come le precedenti, essa viene comunque considerata ma andrà rivalutata alla luce dei risultati ottenuti dalla classificazione.

Infine, le ultime variabili selezionate sono X_8 , che rappresenta il numero di ore passate al giorno sui dispositivi digitali quali computer o smartphone, e X_9 , che invece rappresenta il numero di ore al giorno che lo studente trascorre navigando in rete: esse risultano decisamente correlate tra loro (0,710). La scelta è quindi ricaduta sulla sola X_9 poiché, pur non essendo stato specificatamente richiesto la motivazione dell'utilizzo di queste tecnologie (per lavoro, studio o svago), questa sembra essere portatrice di un'informazione più circoscritta rispetto alla precedente.

In estrema sintesi, il *subset* selezionato è composto da cinque variabili: *LEVEL*, X_4 , X_7 , X_9 , e X_{16} . Quattro di queste sono qualitative, di cui tre dicotomiche e una ordinale, mentre una è quantitativa: anche la composizione del *subset* è caratterizzata da una presenza di variabili miste e, quindi, è necessario ricorrere alla TSCA.

2.3 La profilazione statistica

Prima di iniziare la descrizione dei profili ottenuti, va subito messo in evidenza che la procedura di TSCA è stata completata progressivamente in modo tale da contemperare il mantenimento della sua qualità con la ricerca del numero ottimale dei profili ottenuti: l'idea di base è quella di non sacrificare troppo questa qualità cercando di ridurre al minimo il numero dei *cluster* ottimali in modo da non doverli poi aggregare ai fini di una loro più facile interpretazione. Seguendo le specifiche tecniche relative alla versione del *software* utilizzato [24], si è selezionata una misura di distanza basata sul logaritmo della verosimiglianza, prescritta nel caso di variabili miste, e l'individuazione automatica del numero ottimale di *cluster* attraverso il criterio basato sul BIC.

Come punto di partenza sono state considerate le tre variabili qualitative che esprimono le conoscenze geografiche e informatiche (*subset* I: *LEVEL*, X_4 e X_7). Successivamente, a questo primo *subset* è stata subito aggiunta l'ultima variabile qualitativa relativa alla condivisione delle coordinate geografiche delle fotografie postate sui social (*subset* II: *LEVEL*, X_4 , X_7 e X_{16}) e, infine, l'unica quantitativa legata alla navigazione in rete (*subset* III: *LEVEL*, X_4 e X_7 , X_9 , e X_{16}) completando l'analisi preliminare: il riepilogo dei risultati ottenuti è mostrato in Fig. 2. In questa figura, oltre al numero di variabili di input e quello ideale dei *cluster* determinato dalla procedura, viene fornita una misura della qualità che si fonda sempre sui criteri di *coesione* (massima similarità tra le unità statistiche all'interno di un gruppo) e *separazione*

(massima diversità tra i gruppi) dei *cluster* ottenuti. Per effettuare la valutazione della qualità si ricorre al *coefficiente di silhouette* che sintetizza, in un'unica misura normalizzata, entrambi i criteri: il suo *range* di variazione è compreso tra -1, che indica la qualità peggiore, e +1 che indica quella migliore [23]. Per una sua agile interpretazione, lo stesso grafico di output segnala con il colore verde che solo per valori superiori a 0,5 la qualità della profilazione può essere ritenuta “buona”.



Fig. 2. Riepiloghi delle analisi per il subset I (in alto a sinistra), per il subset II (in alto a destra), per il subset III (in basso a sinistra) e per il subset IV (in basso a destra).

Sempre dalla Fig. 2 è facile osservare che tutti i *subset* esaminati presentano buone misure di silhouette (rispettivamente I: 0,96, II: 0,86 e III: 0,62), ma il numero ideale di *cluster* da interpretare sale da 8 a 10. Per questo si è provato a rimuovere la variabile X_{16} relativa alla condivisione delle coordinate geografiche delle fotografie postate (*subset* IV: $LEVEL, X_4, X_7$ e X_9) per capire se rimuovendo la variabile meno significativa dal punto di vista dell'interpretazione dei profili si ottenesse un'apprezzabile riduzione del numero ideale dei *cluster*: come si vede questo diminuisce a 6 senza compromettere la qualità del cluster poiché il coefficiente di silhouette si mantiene a 0,64.

Prima di iniziare l'analisi vera e propria, è stata praticata una ricerca di eventuali valori anomali per il *subset* IV poiché questi possono condizionare i risultati finali: ne sono stati individuati solo alcuni che, tolti, hanno ridotto la numerosità complessiva a

208 studenti, ma hanno permesso di restringere a 5 i cluster ottimali mantenendo praticamente stabile il coefficiente di qualità (0,63).

La prima valutazione ha riguardato l'importanza di ogni singola variabile nella costruzione dei gruppi: reso unitario il peso della variabile più importante, che secondo la procedura è *LEVEL*, le altre sono state, in ordine decrescente, X_7 (0,77), X_4 (0,75), mentre X_9 è risultata di peso decisamente inferiore (0,09). Per quanto riguarda l'interpretazione dei *cluster*, il software sintetizza i risultati numerici nella Fig. 3.

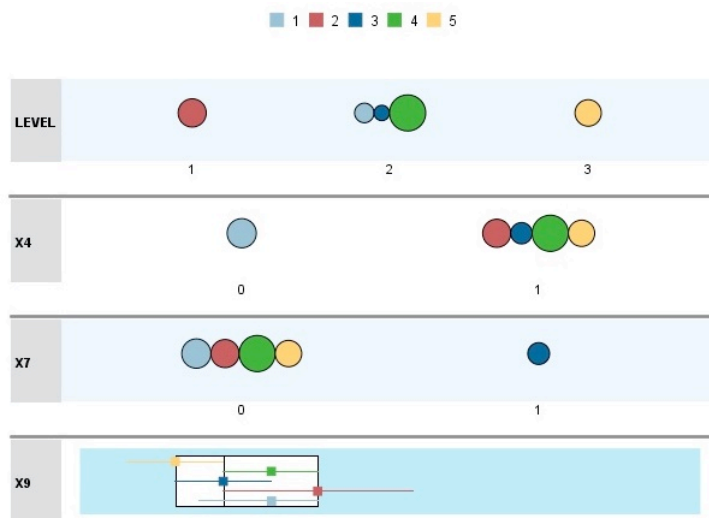


Fig. 2. Output grafico di SPSS per il confronto dei cinque *cluster* individuati dalla TSCA.

In Fig. 3, la concentrazione sulle modalità delle variabili qualitative (dove, si ricorda, per X_4 il valore unitario indica i residenti in FVG, mentre per X_7 coloro che hanno frequentato CIE) sono indicate con dei simboli grafici circolari. Per quanto riguarda l'unica variabile quantitativa X_9 vengono invece riportati i singoli *box-plot*. Dal grafico di output, e dalla lettura di tutte le tabelle *pivot* (i cui valori più rilevanti saranno di seguito proposti nella descrizione dei profili), è possibile profilare i singoli gruppi tenendo conto sin da subito che dall'osservazione dei *box-plot* non sovrapposti della variabile quantitativa, che si rammenta pesare ben poco nella costruzione dei gruppi, si osserva come questa differenzi soprattutto il secondo gruppo dal quinto.

È finalmente possibile individuare i profili dei singoli gruppi. Il *cluster* 5 ricomprende 35 studenti residenti con conoscenze universitarie GIS sia pratiche che teoriche (Liv. III), che non hanno mai frequentato CIE e con un minore tempo trascorso in rete (mediana di 2 ore). All'opposto, il *cluster* 2 è composto da 40 studenti che non hanno nessuna conoscenza universitaria GIS (Liv. I), che non hanno mai seguito CIE e con un maggiore tempo speso navigando in rete (mediana di quasi 5 ore). In questi due primi gruppi è evidente il ruolo esercitato dal livello di conoscenza GIS e dal tempo di navigazione in rete. Se è vero quanto dimostrato in [17] e [18] sul fondamentale ruolo della formazione universitaria, è possibile supporre che se il *cluster* 5 dovrebbe portare

alle migliori performance rispetto agli esercizi completati, il *cluster 2* dovrebbe ottenere risultati decisamente peggiori: la verifica di questa ipotesi sarà presto fornita.

I rimanenti studenti si ripartiscono nei tre gruppi restanti: questi possiedono in gran maggioranza una conoscenza solo teorica dei GIS (Liv. II). Infatti, il *cluster 3* è composto da 24 studenti residenti con una maggioranza di Liv. II (ce ne sono alcuni degli altri livelli), ma che a differenza degli altri studenti partecipanti agli esperimenti hanno tutti seguito CIE: a questo punto ci si può chiedere se questa formazione extrauniversitaria ha avuto un effetto sulla loro capacità di portare a termine gli esercizi proposti. Il *cluster 4* è ancora relativo a 66 studenti tutti residenti del Liv. II che però non hanno seguito CIE, mentre nel restante *cluster 1* sono ricompresi i 43 studenti tutti non residenti con prevalente conoscenza GIS sempre del secondo livello (ma anche in questo caso ci sono alcuni studenti di altri livelli) e che non hanno seguito CIE.

Per risolvere le questioni appena sollevate, si devono ora considerare le medie condizionate al gruppo degli esercizi completati al netto (X_{22}) e al lordo (X_{23}) delle segnalazioni libere. Come atteso si osserva subito che per il *cluster 5*, con conoscenze universitarie GIS sia pratiche sia teoriche, le medie sono 2,43 e 3,34 pro capite, mentre per il *cluster 2*, privo di tali conoscenze, sono 1,45 e 2,20. I valori medi del *cluster 5* rappresentano i massimi per tutti i gruppi individuati e quindi lo qualificano, unicamente rispetto allo specifico aspetto qui studiato, come il “migliore” nel portare a termine gli esercizi di mappatura proposti durante gli esperimenti.

L'aspetto più interessante di questi risultati riguarda il fatto che le medie del *cluster 3*, che identifica unicamente gli studenti che hanno seguito CIE, sono rispettivamente 2,21 e 3,25: queste risultano le migliori prestazioni dopo quelle del *cluster 5*. In altre parole, rimane empiricamente dimostrato il ruolo della formazione informatica extrauniversitaria che permette agli studenti coinvolti negli esperimenti di non farsi intimidire nell'interagire con le mappe interattive pur non avendo frequentato un insegnamento universitario impartito presso un laboratorio GIS.

Infine, i *cluster 4* e *1* presentano risultati non troppo diversi tra loro poiché entrambi composti da studenti del secondo livello, ma dove la migliore conoscenza geografica del FVG degli studenti residenti permette loro di ottenere una performance leggermente più alta in media: rispettivamente 1,55 e 2,39 esercizi pro capite dei residenti del *cluster 4*, contro 1,42 e 2,23 dei non residenti del *cluster 1*. Quest'ultimo gruppo possiede performance molto simili a quelle del *cluster 2*: questi possono essere considerati come i “peggiori” profili rispetto al portare a conclusione gli esercizi proposti.

3 Conclusioni

Nel presente contributo è stato verificato empiricamente, grazie ai dati raccolti attraverso una batteria di esperimenti di mappatura, come la formazione geografica e informatica sia stata inclusa all'interno dei profili generati tramite *cluster analysis* e come questa abbia influito sulla capacità degli studenti coinvolti nel portare a termine gli esercizi proposti dai ricercatori. Più precisamente, il ruolo della formazione universitaria in ambito GIS è risultata di fondamentale importanza anche per la formazione dei gruppi, ma l'aver seguito corsi informatici extrauniversitari è risultato

di particolare rilievo soprattutto per quegli studenti che possiedono una conoscenza solo teorica dei GIS. Non sorprende quindi che gli studenti che non vantano nessuna delle due formazioni appena citate hanno dimostrato una minor capacità nel saper interagire con la mappa interattiva integrata nel geoquestionario.

Le informazioni raccolte durante questa batteria di esperimenti, considerata sin da subito come un'esperienza pilota, possono risultare utili anche per un futuro miglioramento e allargamento della rilevazione a più vasta scala. A tal proposito si possono sottolineare quanto meno due aspetti: il piano di campionamento e il contenuto conoscitivo del geoquestionario. Per quanto riguarda il piano di campionamento – nonostante gli insegnamenti universitari siano stati selezionati in modo da ricomprendere diversi livelli di conoscenza in ambito GIS e diversi corsi di laurea triennale in modo tale da non aver studenti in comune –, in futuro sarà necessario ampliare il numero di insegnamenti, coinvolgendo anche le altre sedi locali della stessa università, per rendere ancora più popolato il *dataset* da sottoporre ad analisi. Relativamente al geoquestionario, si ravvisa la necessità di raccogliere maggiori informazioni. Solo per fare un esempio, per quanto riguarda le ore trascorse davanti agli strumenti digitali sarà fondamentale differenziare non solo la loro tipologia, ma anche la ragione del loro utilizzo.

Ringraziamenti

L'autore desidera ringraziare i colleghi che, a diverso titolo, hanno collaborato a partire dal 2018 al progetto PaRiDe, ossia Salvatore Amaduzzi, Giorgia Bressan, Andrea Guaran, Mauro Pascolini e, più recentemente, Francesco Visentin.

Riferimenti bibliografici

1. Guaran A., Pascolini M. (a cura di): Pianificazione e governo del paesaggio: analisi, strategie, strumenti. L'apporto pluridisciplinare dell'Università di Udine al Piano paesaggistico regionale del Friuli Venezia Giulia, Forum, Udine (2019).
2. Zaccomer G.P.: L'analisi territoriale socio-economica in ambito paesaggistico. Gli indicatori compositi per la zonizzazione territoriale del Friuli Venezia Giulia, Forum, Udine (2018).
3. Zaccomer G.P.: L'analisi territoriale socio-economica in ambito paesaggistico. Metodi multivariati e indicatori compositi a confronto. In: Federazione italiana delle Associazioni Scientifiche per le Informazioni Territoriali e Ambientali, Atti della Conferenza Nazionale di Geomatica e Informazione Territoriale ASITA 2019, ASITA, Milano, 1033-1040 (2019).
4. Bressan G., Pascolini M.: Dalle percezioni della popolazione al dato georeferito: studio quali-quantitativo del paesaggio del Friuli V.G. In: Federazione italiana delle Associazioni Scientifiche per le Informazioni Territoriali e Ambientali, Atti della Conferenza Nazionale di Geomatica e Informazione Territoriale ASITA 2019, ASITA, Milano, 115-122 (2019).
5. Bressan G., Amaduzzi S.: Map based surveys for mapping high-quality and degraded sites. In: Gallia A. (a cura di), Territorio: rischio / risorsa, Labgeo Caraci, Roma, 197-213 (2020).
6. Amaduzzi S., Bressan G., Guaran A., Pascolini M., Zaccomer G.P.: I paesaggi che cambiano: quali risposte contro il degrado? In: Cerutti S., Tadini M. (a cura di),

- Mosaico/Mosaic, Memorie geografiche NS17, Società di Studi Geografici, Firenze, 493-502 (2019).
7. Amaduzzi S., Bressan G., Guaran A., Pascolini M., Zaccomer G.P.: Paesaggi del degrado. Indagini ed esperienze in Friuli Venezia Giulia tra rischi e degradi, Forum, Udine (2021).
 8. Goodchild, M.: Citizens as Sensors: The World of Volunteered Geography. *GeoJournal* 69 (4), 211-221 (2007).
 9. Borruso G.: La ‘nuova cartografia’ creata dagli utenti. Problemi, prospettive, scenari, *Bollettino Associazione Italiana di Cartografia* 138, 241-252 (2010).
 10. Fonte C.C., Antoniou V., Bastin L., Estima J., Arsanjani J. J., Bayas, J.C. L., See L., Vatsava R.: Assessing VGI Data Quality. In: Foody G., See L., Fritz S., Mooney P., Olteanu-Raimond A.M., Fonte C.C., Antoniou V. (eds.), *Mapping and the Citizen Sensor*, Ubiquity Press: London, UK, 137-163 (2017).
 11. Bressan G.; Assessing the positional accuracy of perceptual landscape data: A study from Friuli Venezia Giulia, Italy. *Transactions in GIS* 25, 642–671 (2021).
 12. Bressan G., Zaccomer G.P., Grassetti L.; Studying the spatial distribution of volunteered geographic data through a non-parametric approach. In: Gervasi O., Murgante B., Misra S., Garau C., Blečić I., Taniar D., Apduhan B.O., Rocha A.M., Tarantino E., Torre C.M., Karaca Y. (eds.), *Computational Science and Its Applications – ICCSA 2020*, Springer International Publishing, Cham (CH), 488-504 (2020).
 13. Bressan G., Guaran A., Zaccomer G.P.: Insights on risk perception: The case of Friuli Venezia Giulia. *Sustainable Mediterranean Construction*, Special issue 5, 13-15 (2021).
 14. Bressan G., Guaran A., Visentin F., Zaccomer G.P.: Aspetti geografici del confronto fra sapere esperto e contestuale: un’analisi regionale nei paesaggi del “degrado”, *Bollettino della Società Geografica Italiana* 4(2), 2021, 3-18 (2021).
 15. Bressan G., Guaran. Using photographs to collect and study citizens’ perceptions of landscape degradation: a regional case in Italy, *Journal of Cultural Geography*, 1-24 (2022). DOI: 10.1080/08873631.2022.2041273.
 16. Poplin, A.: How user-friendly are online interactive maps? Survey based on experiments with heterogeneous user. *Cartography and Geographic Information Science* 42 (4), 358-376 (2015).
 17. Zaccomer G.P., Bressan, G.: Can university students provide high-quality geoinformation? First results of an empirical study. *Bollettino della Associazione Italiana di Cartografia* 168, 94-106 (2020).
 18. Zaccomer G.P., Sigura, M.: Georeferencing skills and student profile. Results from a multivariate regression model. In: *Proceedings of the International Cartographic Association* 4, 1-6 (2021).
 19. Everitt B.S., Landau S., Leese M., Stahl D., *Cluster Analysis*. 5th ed., Wiley & Sons, Chichester (2011).
 20. Zhang T., Ramakrishnan R., Livny M.: BIRCH: A New Data Clustering Algorithm and Its Applications, *Data Mining and Knowledge Discovery* 1, 141-182 (1997).
 21. Chiu T., Fang D.P., Chen J., Wang Y., Jeris C.: A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment. In: *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 263-268 (2001). <https://doi.org/10.1145/502512.502549>
 22. SPSS Inc.: The SPSS TwoStep Cluster Component. A scalable component enabling more efficient customer segmentation. White paper – technical report, SPSS Inc., Chicago (2001).
 23. Norušis M.: *IBM SPSS Statistics 19. Statistical Procedures Companion*. Prentice Hall - Pearson Education, Upper Saddle River, New Jersey (2012).
 24. IBM: *IBM SPSS Statistics Base 25*. IBM Corporation, Armonk, New York (2017).