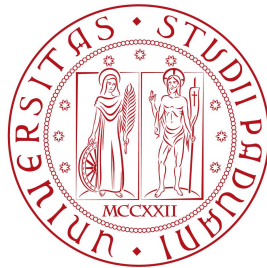


Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in

Scienze Statistiche



**Analisi delle strutture di dipendenza latenti
e previsione degli accessi al Fascicolo Sanitario Elettronico
della Provincia di Trento**

Relatore: prof. Davide Risso
Dipartimento di Scienze Statistiche

Laureanda: Chiara Ballarin
Matricola n. 2112302

Anno Accademico 2025/2026

Indice

Introduzione	1
1 Contesto dello studio e descrizione dei dati	5
1.1 Contesto e obiettivi dello studio	5
1.1.1 Il sistema TreC+ e la digitalizzazione dei servizi sanitari	5
1.1.2 Obiettivi	7
1.2 Descrizione dei dati	8
1.2.1 Dataset a livello provinciale	9
1.2.2 Dataset a livello comunale per Tesino e Val di Sole	10
1.2.3 Dataset a livello di singolo accesso	10
1.2.4 Dati ISTAT e classificazioni territoriali	11
1.3 Pre-processamento dei dati	12
1.3.1 Armonizzazione dei codici e delle classificazioni territoriali	12
1.3.2 Integrazione delle informazioni altimetriche . . .	13
1.3.3 Integrazione delle informazioni demografiche . . .	14
1.3.4 Dataset integrato	14
2 Analisi esplorative	16
2.1 Analisi dei dati aggregati provinciali	17
2.1.1 Andamento temporale degli accessi	17
2.1.2 Canali di accesso e modalità di autenticazione . .	19

2.1.3	Tipologia di operazioni svolte	21
2.1.4	Referti visualizzati e scaricati	22
2.2	Approfondimento su Tesino e Val di Sole	24
2.2.1	Finalità dell'approfondimento	24
2.2.2	Confronto con il trend provinciale	25
2.2.3	Caratteristiche demografiche e territoriali delle due aree	27
2.2.4	Analisi comunale	29
2.3	Analisi dei dati a livello di singolo accesso	30
2.3.1	Andamento temporale degli accessi	31
2.3.2	Modalità di autenticazione	33
2.3.3	Profilo demografico degli accessi	36
2.3.4	Distribuzione territoriale e tassi comunali di accesso	40
2.3.5	Standardizzazione per sesso e fascia d'età	40
2.3.6	Associazione con le caratteristiche comunali	43
3	Metodi	45
3.1	Obiettivo	45
3.2	Modelli grafici gaussiani	46
3.2.1	Modelli grafici e dipendenze condizionate	46
3.2.2	Il caso gaussiano e la matrice di precisione	46
3.3	Graphical Lasso	48
3.3.1	Interpretazione mediante regressioni condizionate	48
3.3.2	Dalla selezione dei vicinati al Graphical Lasso	48
3.3.3	Implementazione mediante il pacchetto huge	49
3.3.4	Selezione della rete	50
3.4	Costruzione dell'input per l'analisi di rete	50
3.5	Modelli additivi generalizzati	51
3.5.1	Dai modelli additivi ai modelli additivi generalizzati	51
3.5.2	Distribuzioni per dati di conteggio	53

3.5.3	Offset della popolazione	54
3.5.4	Residui di Pearson	55
3.6	Dipendenza temporale e sbiancamento	55
3.7	Trasformazione nonparanormal	57
3.8	Descrizione e interpretazione della rete stimata	59
3.9	Individuazione delle comunità nella rete	59
4	Modellazione dei conteggi e analisi di rete	61
4.1	Obiettivo	61
4.2	Dati utilizzati	61
4.3	Modelli GAM per la componente sistematica	62
4.3.1	Modello di Poisson	62
4.3.2	Modello binomiale negativo	63
4.3.3	Introduzione dell'effetto casuale comunale	64
4.3.4	Diagnostica del modello finale	66
4.3.5	Autocorrelazione dei residui e procedura di sbiancamento	67
4.3.6	Trasformazione nonparanormal e matrice finale	68
4.4	Stima della rete tramite Graphical Lasso	68
4.5	Correlazioni parziali e misure di rete	70
4.6	Coerenza territoriale degli archi	71
4.7	Clusterizzazione Louvain	72
4.8	Conclusioni	74
5	Confronto preventivo dei modelli GAM	76
5.1	Obiettivo dell'analisi	76
5.2	Stima, validazione temporale e verifica	77
5.3	Costruzione preliminare delle reti e dei cluster	78
5.4	Specificazione dei modelli previsivi	80
5.5	Criteri di valutazione	82
5.6	Selezione mediante rolling window	83

5.7	Verifica finale dei modelli	84
5.8	Interpretazione dei risultati	85
5.9	Conclusioni	86
Conclusioni		87
5.10	Limitazioni	90
5.11	Sviluppi futuri	90
A Armonizzazione dei comuni derivanti da fusione		92
B Analisi dei picchi temporali		95
B.1	Andamento giornaliero e durata dei picchi	95
B.2	Modalità di autenticazione nelle giornate di picco	97
B.3	Conclusioni	98

Elenco delle figure

2.1	Andamento mensile degli accessi alla piattaforma TreC+ tra il 6 dicembre 2025 e il 4 gennaio 2026 e relativa distribuzione degli accessi per mese.	17
2.2	Andamento giornaliero degli accessi alla piattaforma TreC+ tra il 6 dicembre 2025 e il 4 gennaio 2026 e relativa distribuzione per giorno della settimana.	18
2.3	Andamento mensile degli accessi alla piattaforma TreC+ tramite applicazione mobile, portale web e delega, tra marzo 2022 e dicembre 2025.	19
2.4	Andamento mensile degli accessi da portale tramite SPID e TEAM tra marzo 2022 e dicembre 2025.	21
2.5	Andamento mensile dei referti visualizzati e scaricati tra febbraio 2023 e agosto 2025.	23
2.6	Confronto dei tassi mensili di accesso per 1.000 abitanti tra Provincia Autonoma di Trento, Tesino e Val di Sole, distinti per accessi totali, portale web e applicazione mobile.	26
2.7	Distribuzione dell'altitudine media nei comuni della Provincia di Trento.	28
2.8	Andamento del tasso totale di accesso per 1.000 abitanti nei comuni del Tesino e della Val di Sole. La linea tratteggiata rappresenta il tasso medio mensile dell'area.	30
2.9	Andamento mensile degli accessi alla piattaforma TreC+ da agosto 2024 ad aprile 2026.	31

2.10	Distribuzione del numero giornaliero di accessi alla piattaforma TreC+ per giorno della settimana.	32
2.11	Distribuzione complessiva degli accessi alla piattaforma TreC+ per ora del giorno.	33
2.12	Andamento mensile degli accessi per modalità di autenticazione.	35
2.13	Numero complessivo di accessi e numero di accessi per 1.000 residenti, distinti per fascia d'età.	37
2.14	Distribuzione dell'età per modalità di autenticazione. . .	39
2.15	Distribuzione comunale del numero assoluto di accessi, a sinistra, e del tasso cumulato di accesso standardizzato per sesso e fascia d'età, a destra.	41
2.16	Relazione tra il tasso standardizzato di accesso e l'altitudine media del comune, a sinistra, e l'età media della popolazione residente, a destra.	43
4.1	Diagnostica del modello GAM finale.	66
4.2	Selezione del parametro di regolarizzazione mediante il criterio StARS.	69
4.3	Localizzazione geografica dei principali cluster Louvain.	73
B.1	Andamento giornaliero degli accessi alla piattaforma TreC+. Le fasce azzurre evidenziano i cinque mesi con il maggior numero complessivo di accessi, mentre la linea scura rappresenta il numero medio giornaliero di accessi calcolato per ciascuna settimana.	96
B.2	Modalità di autenticazione nelle venti giornate con il maggior numero di accessi e negli altri giorni della serie. A sinistra è riportato il numero medio giornaliero di accessi; a destra la composizione percentuale.	97

Elenco delle tabelle

2.1	Volume complessivo e distribuzione percentuale delle principali operazioni effettuate tramite TreC+ tra marzo 2022 e dicembre 2025.	22
2.2	Distribuzione dei referti visualizzati e scaricati per tipologia nel periodo compreso tra febbraio 2023 e agosto 2025.	23
2.3	Confronto della struttura per età della popolazione tra Tesino, Val di Sole e Provincia di Trento. Valori medi sul periodo 2023-2025.	27
2.4	Altitudine media ponderata per superficie di Tesino, Val di Sole e Provincia di Trento.	28
2.5	Distribuzione degli accessi per modalità di autenticazione.	34
2.6	Numero di accessi, distribuzione percentuale e tasso cumulato per sesso.	38
4.1	Coefficienti parametrici del modello binomiale negativo senza effetto casuale comunale.	63
4.2	Risultati del test di Ljung–Box prima e dopo lo sbiancamento autoregressivo.	68
4.3	Sintesi della rete stimata tramite Graphical Lasso.	70
4.4	Sintesi delle correlazioni parziali associate agli archi della rete.	70
4.5	Proporzione di archi che collegano comuni appartenenti alla stessa partizione territoriale.	71

4.6	Confronto tra proporzioni osservate di archi interni e distribuzioni simulate su reti casuali con lo stesso numero di nodi e archi.	72
4.7	Sintesi della clusterizzazione Louvain applicata alla rete stimata.	73
5.1	Finestre utilizzate nella procedura di validazione temporale a finestra mobile.	78
5.2	Caratteristiche delle reti e delle partizioni Louvain ottenute nel periodo di sviluppo.	79
5.3	Configurazioni selezionate mediante rolling window. . . .	83
5.4	Prestazioni finali sul periodo di verifica.	84
A.1	Corrispondenze tra comuni soppressi e comuni attuali utilizzate per l'armonizzazione territoriale.	93

Introduzione

La progressiva digitalizzazione dei servizi sanitari ha modificato il modo in cui i cittadini accedono alle informazioni cliniche e interagiscono con il sistema sanitario. In questo contesto, il Fascicolo Sanitario Elettronico rappresenta uno degli strumenti principali attraverso cui consultare referti, prescrizioni e altri documenti sanitari, oltre che accedere a servizi di prenotazione e gestione del proprio percorso di cura. Nella Provincia Autonoma di Trento tali funzionalità sono integrate nella piattaforma TreC+, sviluppata nell'ambito del processo di innovazione dei servizi sanitari provinciali.

La diffusione di una piattaforma di questo tipo genera una grande quantità di dati relativi alle modalità e all'intensità di utilizzo del servizio. La loro analisi consente non soltanto di descrivere l'evoluzione complessiva degli accessi, ma anche di approfondire le differenze temporali, demografiche e territoriali che caratterizzano l'utilizzo dei servizi sanitari digitali. L'utilizzo della piattaforma può infatti variare nel tempo, tra gruppi di popolazione e tra comuni, riflettendo sia la diversa composizione demografica dei territori sia caratteristiche geografiche, organizzative e tecnologiche non sempre direttamente osservabili.

Il presente lavoro di tesi, sviluppato nell'ambito del tirocinio svolto presso Deda Next, analizza i dati di accesso a TreC+ con un duplice obiettivo.

Da un lato, si intende ricostruire il contesto di utilizzo della piattaforma

e descriverne le principali dinamiche temporali, demografiche e territoriali. Dall'altro, si vuole verificare se, dopo aver tenuto conto delle componenti sistematiche degli accessi, rimangano associazioni residue tra le serie comunali e se tali informazioni possano contribuire a migliorarne la previsione.

La fase esplorativa segue un percorso coerente con il diverso livello di dettaglio delle fonti disponibili. In primo luogo, vengono analizzati i dati aggregati a livello provinciale, utilizzati per descrivere l'evoluzione complessiva degli accessi, i canali di utilizzo e le principali operazioni svolte tramite la piattaforma. Successivamente, la disponibilità di dati mensili a livello comunale per il Tesino e la Val di Sole consente un primo approfondimento dell'eterogeneità territoriale. Le differenze emerse tra le due aree e tra i comuni appartenenti alla stessa zona motivano quindi il passaggio a una base informativa più dettagliata, costituita dai dati disaggregati a livello di singolo accesso.

Su questi ultimi dati vengono approfonditi l'andamento temporale degli accessi, le modalità di autenticazione, il profilo demografico e la distribuzione territoriale.

Le evidenze emerse dall'analisi esplorativa conducono alla successiva fase modellistica, nella quale l'attenzione viene concentrata sul numero complessivo di accessi, aggregato per comune e settimana. Il dataset viene integrato con informazioni demografiche, geografiche e amministrative e utilizzato per verificare se i comuni presentino dinamiche differenti anche dopo aver tenuto conto delle principali caratteristiche osservate.

La componente sistematica dei conteggi settimanali viene inizialmente rappresentata mediante modelli additivi generalizzati (GAM). La loro flessibilità consente di descrivere un andamento temporale non lineare e di includere, oltre all'offset della popolazione, caratteristiche comunali ed effetti casuali specifici per territorio.

La componente non spiegata dal modello viene successivamente utilizzata per applicare il Graphical Lasso e stimare una struttura sparsa di dipendenze condizionate tra le serie comunali. In questa rappresentazione, i nodi corrispondono ai comuni, mentre gli archi identificano associazioni residue che rimangono dopo aver considerato simultaneamente tutte le altre serie incluse nell'analisi.

La struttura ottenuta viene quindi confrontata con le principali partizioni territoriali della provincia e suddivisa in comunità mediante l'algoritmo di Louvain, così da individuare gruppi di comuni maggiormente connessi al loro interno.

L'ultima parte del lavoro valuta se tali raggruppamenti contengano informazioni utili anche in termini previsivi.

Vengono confrontati un GAM nel quale tutti i comuni condividono lo stesso andamento stagionale, un modello che introduce deviazioni specifiche per le Comunità di Valle e una terza specificazione con deviazioni proprie delle partizioni individuate attraverso il Graphical Lasso e l'algoritmo di Louvain. Il confronto permette di verificare sia se una maggiore flessibilità nella rappresentazione delle dinamiche comunali migliori le previsioni, sia se la partizione derivata dalla rete risulti più informativa rispetto alla classificazione territoriale predefinita. La tesi è organizzata come segue. Il Capitolo 1 presenta il contesto dello studio, descrive le fonti informative e illustra le principali operazioni di pre-processamento. Il Capitolo 2 contiene le analisi esplorative condotte ai diversi livelli di aggregazione. Il Capitolo 3 introduce i riferimenti metodologici relativi al Graphical Lasso e ai modelli additivi generalizzati. Nel Capitolo 4 la metodologia viene applicata ai conteggi settimanali degli accessi, dalla specificazione del modello GAM alla costruzione e all'interpretazione della rete di dipendenze residue tra comuni.

Il Capitolo 5 è invece dedicato alla valutazione previsiva. Vengono

confrontati un modello con andamento stagionale comune, un modello con deviazioni specifiche per le Comunità di Valle e un modello basato sulle partizioni individuate mediante Graphical Lasso e algoritmo di Louvain.

Capitolo 1

Contesto dello studio e descrizione dei dati

1.1 Contesto e obiettivi dello studio

1.1.1 Il sistema TreC+ e la digitalizzazione dei servizi sanitari

Negli ultimi anni, la digitalizzazione dei servizi sanitari si è inserita nel più ampio processo di innovazione della pubblica amministrazione e dei sistemi sanitari regionali. In Italia, tale processo è stato sostenuto da diverse strategie nazionali orientate alla trasformazione digitale dei servizi pubblici, con l'obiettivo di rendere più semplice l'accesso dei cittadini alle informazioni e ai servizi online (Agenzia per l'Italia Digitale, 2016). In ambito sanitario, uno degli strumenti principali di questo percorso è rappresentato dal Fascicolo Sanitario Elettronico, che consente ai cittadini di consultare in formato digitale la propria documentazione clinica e le informazioni relative al proprio percorso sanitario, come referti, prescrizioni e altri dati clinici (Ministero della Salute, 2022; Agenzia per l'Italia Digitale, 2024).

Il progetto TreC+ si colloca all'interno delle attività sviluppate dall'Azienda Provinciale per i Servizi Sanitari della Provincia Autonoma di Trento (APSS) e, più in generale, nel percorso di digitalizzazione dei

servizi sanitari provinciali. La piattaforma rappresenta l'evoluzione della precedente Cartella Clinica del Cittadino, nata con l'obiettivo di favorire una gestione più diretta delle informazioni sanitarie da parte degli utenti. TreC+ si configura oggi come una piattaforma digitale integrata, attraverso la quale i cittadini possono accedere a diversi servizi legati alla propria salute (Azienda Provinciale per i Servizi Sanitari della Provincia Autonoma di Trento, 2026).

L'iniziativa si inserisce anche nel contesto di TrentinoSalute4.0, progetto sviluppato con il coinvolgimento dell'APSS, del Dipartimento salute e politiche sociali della Provincia Autonoma di Trento e della Fondazione Bruno Kessler. In questo ambito, Deda Next ha contribuito allo sviluppo e alla gestione del portale provinciale e dell'applicazione TreC+ (Deda Next, 2026).

Il sistema mette a disposizione dei cittadini un canale unitario per diverse funzionalità, tra cui la consultazione di referti e prescrizioni, la prenotazione di prestazioni sanitarie, la gestione di informazioni relative al proprio stato di salute e l'accesso a servizi di telemedicina. La piattaforma supporta inoltre attività di prevenzione e monitoraggio, permettendo l'aggiornamento del profilo sanitario personale e l'interazione a distanza con i professionisti (Provincia Autonoma di Trento, 2026).

L'introduzione e l'evoluzione dei servizi digitali in sanità hanno assunto particolare rilievo anche a seguito dell'emergenza sanitaria da COVID-19, che ha evidenziato la necessità di rafforzare gli strumenti di accesso e interazione a distanza tra cittadini e servizi sanitari (Deda Next, 2026). Precedenti indagini condotte dall'Università di Trento hanno mostrato che l'utilizzo della piattaforma è influenzato da caratteristiche demografiche, socioeconomiche e tecnologiche degli utenti. In particolare, gli utenti risultano maggiormente concentrati nelle fasce centrali di età e tra i soggetti con livelli di istruzione medio-alti, mentre

le classi più anziane mostrano generalmente una minore propensione all'utilizzo delle tecnologie digitali (Zanutto, 2017). Gli studi mostrano inoltre che l'utilizzo della piattaforma è strettamente legato alla familiarità con Internet e con gli strumenti digitali, evidenziando come la diffusione della piattaforma e quella degli smartphone siano andate di pari passo. Le valutazioni espresse dagli utenti risultano generalmente positive, soprattutto in relazione alla semplicità di utilizzo del sistema, al risparmio di tempo e alla possibilità di accedere più facilmente alle proprie informazioni sanitarie (Zanutto, 2016).

Un ulteriore elemento rilevante riguarda le modalità di autenticazione. L'accesso ai servizi digitali della pubblica amministrazione avviene sempre più spesso attraverso sistemi di identità digitale, tra cui SPID (*Sistema Pubblico di Identità Digitale*), che permette ai cittadini di autenticarsi e utilizzare i servizi online (Agenzia per l'Italia Digitale, 2026). Anche nel caso di TreC+, l'evoluzione delle modalità di autenticazione può modificare il modo in cui gli utenti accedono al servizio, incidendo sull'intensità e sulle forme di utilizzo della piattaforma.

La diffusione di TreC+ rende disponibili grandi quantità di dati generati dalle interazioni degli utenti con i servizi digitali. Tali informazioni consentono di descrivere i volumi di attività e di studiare le dinamiche temporali, territoriali e demografiche di utilizzo della piattaforma.

1.1.2 Obiettivi

Il presente lavoro di tesi si inserisce in questa cornice progettuale ed è stato sviluppato nell'ambito del tirocinio svolto presso Deda Next, durante il quale è stato possibile accedere ai dati operativi relativi all'utilizzo della piattaforma TreC+. L'obiettivo generale è analizzare i dati di accesso al fine di descrivere le dinamiche temporali, territoriali e demografiche di utilizzo dei servizi sanitari digitali.

In una prima fase, l'attenzione è rivolta alla costruzione e all'esplorazione dei dati disponibili, con l'obiettivo di evidenziare eventuali differenze nei livelli di utilizzo della piattaforma tra periodi temporali, aree territoriali e gruppi di popolazione. Successivamente, l'analisi viene approfondita mediante modelli volti a distinguere le componenti sistematiche degli accessi dalla parte residua non spiegata dalle covariate osservate e dalle differenze strutturali tra comuni. A partire da tale componente residua, il lavoro si propone di indagare l'eventuale presenza di strutture di dipendenza residue tra comuni, così da individuare gruppi caratterizzati da comportamenti di utilizzo simili anche dopo aver tenuto conto degli effetti osservabili inclusi nel modello. Le informazioni ottenute da questa fase vengono infine considerate in un'ottica previsiva, valutando se i gruppi individuati attraverso la rete contribuiscano a migliorare la previsione degli accessi futuri alla piattaforma.

1.2 Descrizione dei dati

L'analisi si basa su più dataset relativi all'utilizzo di TreC+, caratterizzati da livelli di dettaglio e coperture temporali differenti. Una prima parte del lavoro utilizza dati storici aggregati sull'intero territorio della Provincia Autonoma di Trento, utili per descrivere l'evoluzione complessiva della piattaforma. Una seconda parte considera i dati di accesso a livello comunale, riferiti a due aree specifiche, il Tesino e la Val di Sole. Infine, le analisi più approfondite si basano su un archivio disaggregato a livello di singolo accesso, che rappresenta la base informativa più dettagliata disponibile.

Questa articolazione permette di procedere secondo una logica progressiva: inizialmente viene ricostruito il quadro generale dell'utilizzo della piattaforma a livello provinciale; successivamente vengono

esplorate le differenze tra comuni in due aree specifiche, motivando il ricorso a una fonte più dettagliata ed estesa all'intero territorio provinciale; infine, grazie alla maggiore granularità di questi dati, è stato possibile approfondire alcune caratteristiche degli accessi.

1.2.1 Dataset a livello provinciale

La prima base informativa è costituita da quattro diversi dataset aggregati a livello provinciale, estratti dal portale Grafana utilizzato per il monitoraggio operativo di TreC+.

Il primo dataset raccoglie 43 osservazioni mensili, comprese tra marzo 2022 e dicembre 2025, relative agli accessi e all'utilizzo di alcuni servizi della piattaforma. Sono disponibili il numero totale di accessi, distinguendo tra gli accessi effettuati da portale, da applicazione mobile e tramite delega, il numero di autenticazioni tramite SPID o TEAM (Tessera Europea di Assicurazione Malattia). Il dataset comprende inoltre indicatori relativi all'utilizzo dei servizi, tra cui le prenotazioni di prelievi e appuntamenti, le ricette e i referti visualizzati o scaricati e le richieste di cambio del medico.

Un secondo dataset contiene categorie di indicatori analoghe, ma con granularità giornaliera. La copertura è limitata a 30 giorni, dal 6 dicembre 2025 al 4 gennaio 2026, e consente quindi di osservare nel dettaglio un intervallo temporale circoscritto.

Il terzo e il quarto dataset riguardano, rispettivamente, i referti visualizzati e quelli scaricati. Entrambi sono disaggregati per codice di referto e riportano conteggi distinti per diverse tipologie di documenti clinici, tra cui referti di laboratorio, specialistici e radiologici. I dati coprono il periodo compreso tra febbraio 2023 e dicembre 2025. A differenza dei primi due dataset, questi archivi non descrivono l'utilizzo complessivo della piattaforma, ma approfondiscono una specifica

dimensione del servizio, relativa alla consultazione e al download della documentazione clinica.

1.2.2 Dataset a livello comunale per Tesino e Val di Sole

Oltre ai dati aggregati provinciali, il portale Grafana rende disponibili due dataset mensili a livello comunale, relativi al Tesino e alla Val di Sole. Le variabili descrivono il numero complessivo di accessi alla piattaforma, le diverse modalità di utilizzo, tramite applicazione o portale, e i principali sistemi di autenticazione, SPID e TEAM.

Entrambi i dataset coprono il periodo compreso tra gennaio 2023 e dicembre 2025, per un totale di 36 mesi. La disponibilità di dati comunali per queste due aree ha consentito un primo approfondimento della variabilità territoriale e ha rappresentato un passaggio intermedio tra l'analisi aggregata a livello provinciale e quella successiva, estesa all'intero territorio provinciale mediante i dati relativi ai singoli accessi.

1.2.3 Dataset a livello di singolo accesso

A seguito delle prime analisi condotte sui dataset provinciali e comunali, è emersa la necessità di disporre di dati con un livello di dettaglio maggiore, in grado di descrivere non soltanto i volumi complessivi di utilizzo della piattaforma, ma anche le caratteristiche associate ai singoli accessi. È stato quindi reso disponibile un ulteriore dataset contenente informazioni disaggregate a livello di singolo accesso a TreC+. Ogni riga rappresenta quindi un accesso alla piattaforma.

I dati sono stati forniti in file distinti per Comunità di Valle, anno di riferimento e fascia di età (*under 65* e *over 65*), per un totale complessivo di 96 file. Dopo le operazioni di unione e pulizia, è stato ottenuto un dataset contenente 1.068.338 accessi, relativi al periodo compreso tra il 1° agosto 2024 e l'11 maggio 2026.

Per ciascun accesso sono disponibili informazioni relative:

- alla data e all'ora dell'accesso;
- alla modalità di autenticazione utilizzata;
- al comune associato all'utente, identificato attraverso il codice ISTAT;
- all'età dell'utente;
- al sesso dell'utente.

Rispetto ai dataset aggregati, questa struttura offre una maggiore flessibilità analitica. A differenza dei dati provinciali aggregati, tuttavia, il dataset a livello di singolo accesso non contiene informazioni sull'utilizzo dei singoli servizi della piattaforma, quali la visualizzazione e il download dei referti, le prenotazioni di prelievi e appuntamenti, le ricette o le richieste di cambio del medico, ma descrive esclusivamente le caratteristiche di ciascun accesso. La disponibilità congiunta di informazioni temporali, territoriali e demografiche consente comunque di analizzare la distribuzione degli accessi tra i diversi gruppi di popolazione, studiarne l'evoluzione nel tempo e approfondire le differenze territoriali nell'adozione dei servizi sanitari digitali.

1.2.4 Dati ISTAT e classificazioni territoriali

Oltre ai dataset relativi agli accessi a TreC+, sono state utilizzate alcune fonti informative esterne per integrare i dati con informazioni demografiche, geografiche e territoriali. I dati demografici, altimetrici e geografici sono stati ricavati da fonti ISTAT (ISTAT, 2026a,b,c, 2024), mentre le classificazioni amministrative e territoriali dei comuni sono state fornite nell'ambito del tirocinio svolto presso Deda Next.

In particolare, sono stati considerati:

- i dati sulla popolazione residente per comune, singola età e sesso relativi agli anni 2023–2026, utilizzati secondo intervalli coerenti con i diversi dataset analizzati;
- le informazioni sui codici comunali e sulle variazioni amministrative;
- i dati relativi alla superficie e all’altitudine comunale;
- il file geografico dei confini comunali;
- le classificazioni relative a Comunità di Valle, distretti sanitari, ambiti territoriali e aree di integrazione territoriale.

Queste fonti hanno permesso di integrare i dati di accesso con informazioni demografiche, amministrative e territoriali, utili per il calcolo di tassi confrontabili e per le analisi successive.

1.3 Pre-processamento dei dati

Prima di procedere con le analisi descrittive e modellistiche, è stata effettuata una fase di pre-processamento finalizzata alla costruzione di una base dati integrata e coerente con la geografia comunale corrente. Tale fase si è resa necessaria poiché le diverse fonti presentavano strutture, livelli di aggregazione e codifiche non sempre omogenei.

1.3.1 Armonizzazione dei codici e delle classificazioni territoriali

I dati di accesso sono stati integrati con le classificazioni amministrative e territoriali dei comuni della Provincia Autonoma di Trento. Per ciascun comune sono state considerate la Comunità di Valle di appartenenza, il distretto sanitario, l’ambito territoriale e l’area di integrazione territoriale.

Una particolare attenzione è stata dedicata ai comuni soppressi o nati da fusioni amministrative, poiché alcune fonti facevano riferimento a codifiche non più attive. I codici storici sono stati quindi ricondotti a quelli correnti, in modo da ottenere una base coerente con la struttura amministrativa rappresentata nello shapefile comunale. Nei log di accesso, l'armonizzazione ha riguardato in particolare il codice storico di Coredò, ricondotto al comune corrente di Predaia. Di conseguenza, i 159 codici comunali originariamente presenti nei dati di accesso corrispondono a 158 comuni nella geografia corrente.

La geografia utilizzata come riferimento comprende complessivamente 166 comuni. Al termine dell'armonizzazione, tutti i codici presenti nei log risultano associati a un comune dello shapefile e dispongono delle relative classificazioni territoriali. Tuttavia, soltanto 158 dei 166 comuni sono rappresentati nei dati di accesso. I comuni non presenti nell'estrazione sono Ala, Aldeno, Arco, Caldes, Capriana, Cembra, Lisignago, Mezzolombardo e Rovereto. Poiché tra questi figurano anche comuni di dimensione demografica rilevante, in particolare Rovereto e Arco, la loro assenza viene considerata una possibile limitazione dell'estrazione disponibile e non un'indicazione di effettivo mancato utilizzo della piattaforma.

1.3.2 Integrazione delle informazioni altimetriche

Le informazioni territoriali comprendono la superficie comunale e diverse misure altimetriche, tra cui l'altitudine minima, massima, media e mediana. Per i comuni già presenti nelle fonti con il codice corrente, i valori sono stati associati direttamente tramite il codice ISTAT. Per i comuni nati da fusione, le informazioni disponibili per i comuni originari sono state aggregate al livello della nuova unità amministrativa. In particolare, la superficie è stata ottenuta come somma delle superfici originarie, mentre l'altitudine media è stata calcolata come media

ponderata rispetto alla superficie. Il dettaglio delle corrispondenze utilizzate è riportato in Appendice A, Tabella A.1.

Questa procedura ha permesso di ottenere informazioni altimetriche complete per tutti i 166 comuni presenti nella geografia di riferimento. Nelle analisi successive, l'altitudine media viene utilizzata come proxy di alcune caratteristiche territoriali, quali la perifericità e l'accessibilità del comune, senza tuttavia essere considerata una misura esaustiva della complessità geografica del territorio.

1.3.3 Integrazione delle informazioni demografiche

I dati demografici ISTAT riportano, per ciascun comune e anno, il numero di residenti distinto per singola età e sesso. Per l'analisi dei dataset comunali relativi al Tesino e alla Val di Sole sono stati utilizzati i dati degli anni 2023, 2024 e 2025, mentre per l'integrazione del dataset a livello di singolo accesso sono stati considerati gli anni 2024, 2025 e 2026.

A partire da questi dati sono stati calcolati, per ciascun comune, la popolazione media residente nel triennio, l'età media della popolazione, e la quota di popolazione femminile. L'età media è stata ottenuta ponderando ciascuna età per il corrispondente numero di residenti.

I dati disaggregati per età e sesso sono stati inoltre conservati separatamente, così da poter essere utilizzati come denominatori per il calcolo dei tassi e degli indicatori standardizzati presentati nel capitolo successivo.

1.3.4 Dataset integrato

La base finale conserva il dettaglio del singolo accesso e associa a ciascuna osservazione diverse informazioni relative al comune di riferimento. Per l'identificazione del comune sono disponibili il codice

ISTAT e la denominazione; per l'inquadramento amministrativo sono riportati la Comunità di Valle, il distretto sanitario, l'ambito territoriale e l'area di integrazione territoriale. Le caratteristiche demografiche comprendono la popolazione media residente, l'età media e la quota di popolazione femminile, mentre quelle territoriali riguardano l'altitudine media e la superficie comunale. La base integrata contiene 1.068.338 accessi, riferiti a 158 comuni, nel periodo compreso tra il 1° agosto 2024 e l'11 maggio 2026. Essa rappresenta la base di riferimento per le analisi modellistiche.

Capitolo 2

Analisi esplorative

Le analisi esplorative sviluppate in questo capitolo seguono una logica per livelli di dettaglio progressivi. In primo luogo, vengono analizzati i dati aggregati provinciali, con l'obiettivo di descrivere l'evoluzione generale degli accessi, dei canali di utilizzo e delle principali operazioni svolte dagli utenti. Questa prima parte consente di ricostruire il contesto storico e operativo dell'utilizzo di TreC+.

In secondo luogo, l'attenzione si concentra sui dataset a livello comunale relativi alla Val di Sole e al Tesino. Questi dati permettono di introdurre il tema dell'eterogeneità territoriale, mostrando come l'aggregazione provinciale possa nascondere differenze rilevanti tra aree e tra comuni. Infine, l'analisi viene estesa al dataset individuale degli accessi, che consente di approfondire in modo più dettagliato la distribuzione degli accessi nel tempo, le caratteristiche demografiche degli utenti, le modalità di autenticazione e le differenze territoriali sull'intero territorio provinciale. Questa progressione consente di passare da una descrizione generale del fenomeno a un'analisi più fine dei comportamenti di utilizzo della piattaforma.

2.1 Analisi dei dati aggregati provinciali

La prima parte dell'analisi esplorativa riguarda i dati aggregati a livello provinciale, utilizzati per descrivere l'evoluzione complessiva dell'utilizzo della piattaforma TreC+ nel periodo osservato.

2.1.1 Andamento temporale degli accessi

La Figura 2.1 mostra una crescita marcata degli accessi nel periodo compreso tra marzo 2022 e dicembre 2025. Tra il 2022 e il 2023 i valori rimangono generalmente più contenuti, mentre all'inizio del 2024 si osserva un forte incremento, che porta gli accessi mensili a superare 1,2 milioni. Dopo i livelli più elevati raggiunti nel corso del 2024, nel 2025 emerge una lieve riduzione, seguita da una ripresa negli ultimi mesi disponibili.

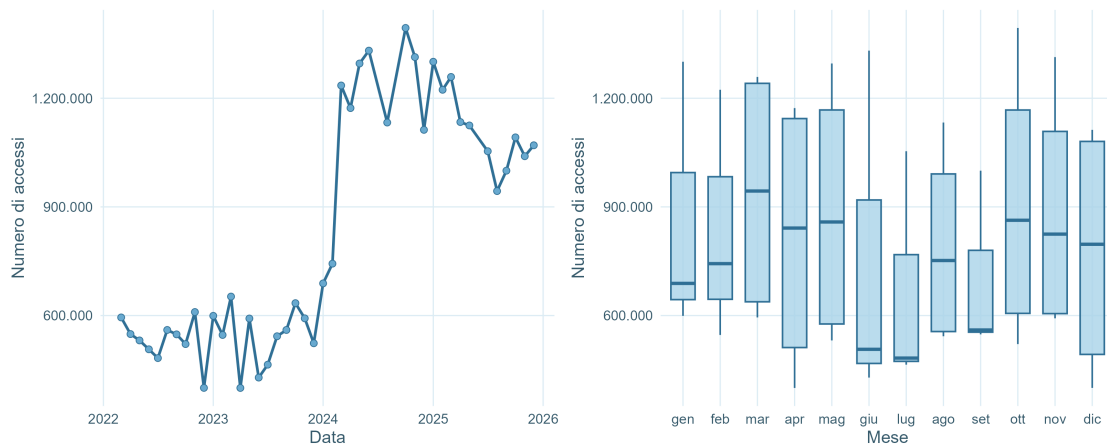


Figura 2.1: Andamento mensile degli accessi alla piattaforma TreC+ tra il 6 dicembre 2025 e il 4 gennaio 2026 e relativa distribuzione degli accessi per mese.

L'aumento osservato coincide temporalmente con l'introduzione, nel giugno 2024, della possibilità di autenticarsi tramite SPID da dispositivo mobile. L'incremento osservato potrebbe quindi riflettere anche un cambiamento nelle modalità di accesso e una maggiore facilità di utilizzo dell'applicazione, oltre a un aumento generale dell'utilizzo della

piattaforma. Questo aspetto viene approfondito nella Sezione 2.1.2, dedicata ai canali di accesso e alle modalità di autenticazione.

La distribuzione per mese dell'anno evidenzia inoltre una componente stagionale. Giugno e luglio tendono a presentare livelli inferiori rispetto agli altri mesi, mentre in alcuni periodi primaverili e autunnali si osservano valori generalmente più elevati. Questa stagionalità può riflettere una diversa intensità nell'utilizzo dei servizi sanitari digitali durante l'anno, legata sia alla normale organizzazione delle attività sanitarie sia a una riduzione di alcune attività nel periodo estivo.

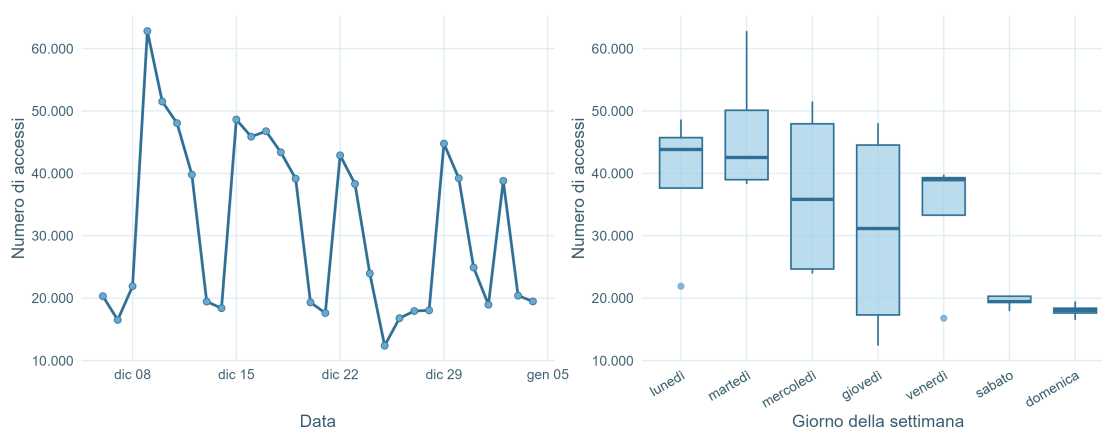


Figura 2.2: Andamento giornaliero degli accessi alla piattaforma TreC+ tra il 6 dicembre 2025 e il 4 gennaio 2026 e relativa distribuzione per giorno della settimana.

La Figura 2.2 approfondisce l'andamento giornaliero degli accessi nel periodo compreso tra il 6 dicembre 2025 e il 4 gennaio 2026. Dal grafico emerge una ciclicità settimanale: gli accessi risultano più elevati nei giorni feriali, mentre si riducono durante il fine settimana. Tale andamento appare coerente con la natura prevalentemente amministrativa e organizzativa di molte funzionalità della piattaforma, come la consultazione di referti e ricette, la gestione delle prenotazioni e l'accesso ai servizi sanitari online. Poiché il periodo considerato comprende le festività natalizie e l'inizio dell'anno, l'interpretazione dei livelli giornalieri deve tenere conto anche della possibile riduzione dell'attività nei giorni festivi e prefestivi.

2.1.2 Canali di accesso e modalità di autenticazione

Nei dati disponibili, gli accessi sono suddivisi in tre categorie: accessi tramite applicazione mobile, accessi tramite portale web e accessi tramite delega. Le prime due identificano il canale utilizzato, mentre la delega rappresenta una specifica modalità di accesso, conteggiata separatamente nei dati disponibili.

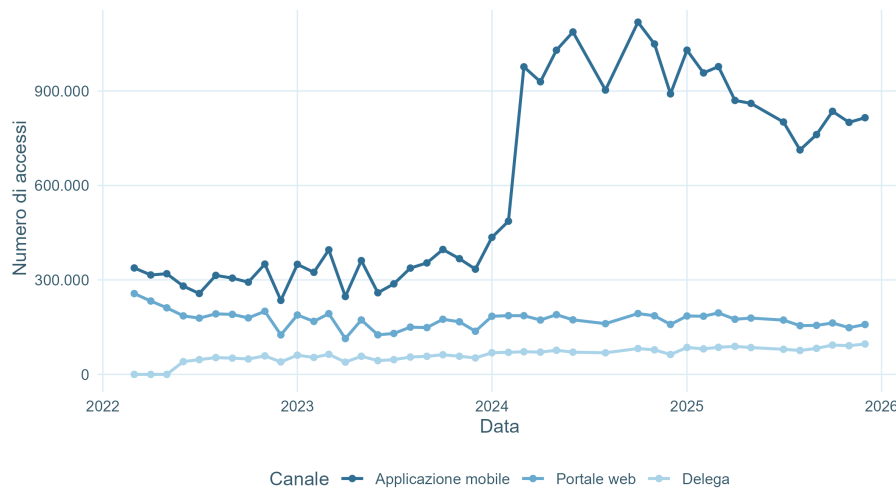


Figura 2.3: Andamento mensile degli accessi alla piattaforma TreC+ tramite applicazione mobile, portale web e delega, tra marzo 2022 e dicembre 2025.

La Figura 2.3 mostra che l'aumento complessivo osservato a partire dal 2024 è associato soprattutto alla crescita degli accessi tramite applicazione. Gli accessi da portale web e quelli tramite delega presentano invece un andamento più stabile. La crescita complessiva sembra quindi essere stata trainata principalmente dal maggiore utilizzo dell'applicazione mobile.

Accanto al canale di accesso, un ulteriore elemento rilevante riguarda le modalità di autenticazione utilizzate dagli utenti. Nei dati sono disponibili informazioni relative agli accessi tramite Tessera Sanitaria (TEAM) e SPID. La distinzione tra tali modalità è rilevante perché rimanda a differenti modalità di identificazione dell'utente e a diversi livelli di accesso ai servizi: l'autenticazione tramite SPID rientra tra le

modalità di autenticazione forte e consente l'accesso completo alle funzionalità della piattaforma, mentre l'accesso tramite TEAM permette di utilizzare un insieme più limitato di servizi. Va però considerato che, in questo dataset, le informazioni sulle modalità di autenticazione non sono rilevate nello stesso modo per tutti i canali di accesso. In particolare, la distinzione tra accessi tramite TEAM e SPID risulta solo per il portale web. La possibilità di utilizzare SPID nell'applicazione mobile è stata introdotta solo a partire dal 2024 e la relativa tracciatura non risulta disponibile nelle serie storiche aggregate. Analogamente, gli accessi tramite carta d'identità elettronica (CIE), anch'essi riconducibili a una modalità di autenticazione forte come SPID, non risultano distinguibili in questi dati. Inoltre, tra gli accessi registrati da app possono rientrare anche accessi effettuati dopo l'attivazione dell'applicazione tramite codice personale, ottenibile online o presso uno sportello abilitato. Anche tale modalità non è distinguibile in questo dataset (Azienda Provinciale per i Servizi Sanitari della Provincia Autonoma di Trento, 2026).

Questi limiti informativi riguardano esclusivamente i dati aggregati analizzati in questa prima fase. Nel dataset a livello di singolo accesso, utilizzato nelle analisi successive, le modalità di autenticazione risultano invece disponibili con un maggiore livello di dettaglio, consentendo di distinguere separatamente gli accessi tramite SPID e TEAM da portale, SPID e TEAM da applicazione mobile, CIE, accessi diretti da app e delega.

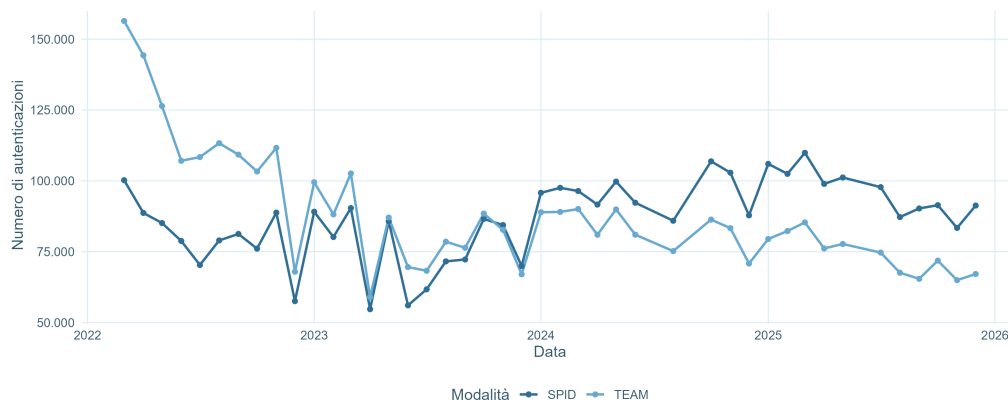


Figura 2.4: Andamento mensile degli accessi da portale tramite SPID e TEAM tra marzo 2022 e dicembre 2025.

La Figura 2.4 mostra che nel 2022 TEAM rappresentava la modalità prevalente tra le due considerate. Successivamente gli accessi tramite TEAM diminuiscono, mentre quelli tramite SPID acquistano progressivamente maggiore rilievo. A partire dal 2024, SPID si colloca stabilmente su livelli superiori rispetto a TEAM, indicando una crescente diffusione dell'identità digitale tra gli accessi effettuati dal portale.

Le evidenze mostrano quindi una trasformazione delle modalità di accesso alla piattaforma: da un lato cresce il ruolo dell'applicazione mobile come canale principale di utilizzo, dall'altro aumenta progressivamente il peso di SPID rispetto a TEAM tra le modalità di autenticazione da portale. Questi risultati suggeriscono una progressiva evoluzione dei comportamenti digitali degli utenti, che sarà ulteriormente approfondita nelle analisi basate sul dataset a livello di singolo accesso, nel quale le modalità di autenticazione sono rilevate in modo più dettagliato.

2.1.3 Tipologia di operazioni svolte

La Tabella 2.1 riporta il volume complessivo delle principali operazioni nell'intero periodo osservato.

Tabella 2.1: Volume complessivo e distribuzione percentuale delle principali operazioni effettuate tramite TreC+ tra marzo 2022 e dicembre 2025.

Operazione	Numero totale	Percentuale
Ricette visualizzate	24.610.744	50,5%
Referti visualizzati	13.393.106	27,5%
Referti scaricati da portale	6.623.240	13,6%
Ricette scaricate	2.346.986	4,8%
Prelievi prenotati	1.352.829	2,8%
Appuntamenti prenotati	246.974	0,5%
Cambio medico	146.947	0,3%

La distribuzione complessiva evidenzia una chiara prevalenza delle operazioni di consultazione della documentazione sanitaria. La visualizzazione di ricette e referti rappresenta infatti, nel complesso, circa il 78% delle operazioni considerate, con un ruolo particolarmente rilevante delle ricette visualizzate, che da sole costituiscono oltre la metà del totale.

Le operazioni di download presentano un peso inferiore, mentre le funzionalità legate alla prenotazione di prestazioni e al cambio del medico risultano molto meno frequenti.

Nel periodo osservato, TreC+ appare quindi utilizzata soprattutto come strumento di accesso alla documentazione clinica personale, più che per lo svolgimento di attività amministrative o gestionali.

2.1.4 Referti visualizzati e scaricati

L'analisi è stata approfondita utilizzando i dataset specifici relativi ai referti visualizzati e scaricati, nei quali le informazioni sono disaggregate per codice e tipologia di documento.

Sono state escluse alcune rilevazioni finali caratterizzate da valori anomali o non pienamente coerenti con il resto della serie. In particolare, sono stati esclusi i mesi successivi a settembre 2025: ottobre risultava mancante, tra novembre e febbraio si osservavano conteggi

molto bassi e nei mesi successivi i valori risultavano pari a zero. Tali valori sono stati quindi considerati non rappresentativi dell'andamento effettivo del servizio nel periodo osservato, verosimilmente per ragioni legate alla copertura o alla tracciatura dei dati.

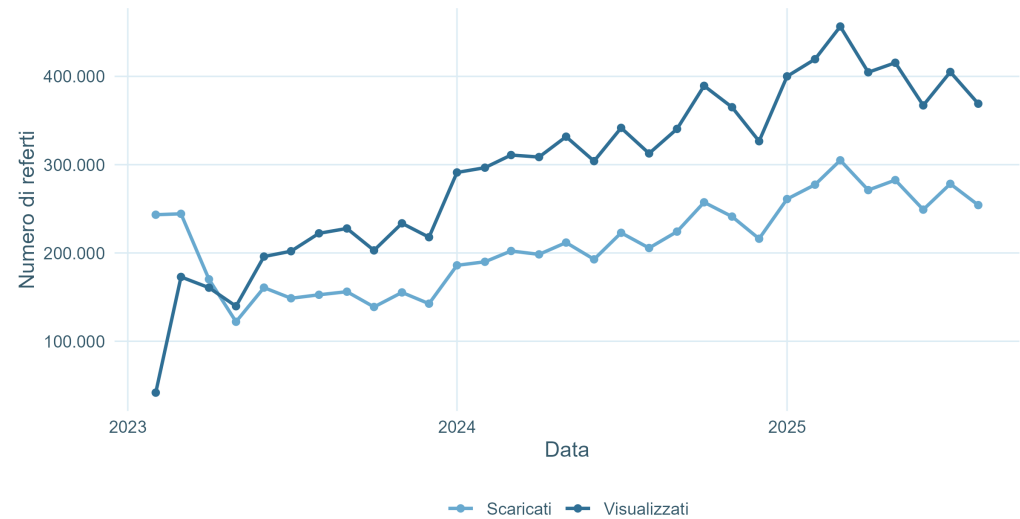


Figura 2.5: Andamento mensile dei referti visualizzati e scaricati tra febbraio 2023 e agosto 2025.

La Figura 2.5 mostra un andamento complessivamente crescente per entrambe le serie, con volumi più elevati per i referti visualizzati. La crescita delle visualizzazioni risulta particolarmente marcata, mentre i download mantengono livelli inferiori. La consultazione sembra quindi rappresentare l’operazione prevalente, mentre il download costituisce un’azione più specifica e meno frequente.

Tabella 2.2: Distribuzione dei referti visualizzati e scaricati per tipologia nel periodo compreso tra febbraio 2023 e agosto 2025.

Tipologia di referto	Referti visualizzati		Referti scaricati	
	Numero	%	Numero	%
Laboratorio (LAB)	4.855.244	50,7%	3.887.517	56,4%
Specialistica ambulatoriale (SPA)	2.033.756	21,2%	1.288.275	18,7%
Radiologia (RAD)	1.202.267	12,6%	798.280	11,6%
Pronto soccorso (PS)	373.724	3,9%	226.202	3,3%
Anatomia patologica (ANA)	237.913	2,5%	123.209	1,8%
Altro	871.324	9,1%	563.774	8,2%

La Tabella 2.2 riporta la distribuzione dei referti visualizzati e scaricati per tipologia. Poiché le categorie meno frequenti presentano volumi contenuti, sono riportate separatamente le cinque tipologie principali, mentre le restanti categorie sono aggregate nella voce “Altro”.

I referti di laboratorio rappresentano la tipologia prevalente sia tra i referti visualizzati sia tra quelli scaricati, con circa la metà delle visualizzazioni complessive e oltre la metà dei download. Seguono i referti specialistici e radiologici, che costituiscono le altre due categorie principali. Le restanti tipologie presentano invece un peso molto più contenuto.

2.2 Approfondimento su Tesino e Val di Sole

2.2.1 Finalità dell’approfondimento

I dataset relativi al Tesino e alla Val di Sole consentono di osservare gli accessi a un livello territoriale più dettagliato rispetto alle serie provinciali. Gli archivi comprendono rispettivamente 3 e 13 comuni e coprono il periodo compreso tra gennaio 2023 e dicembre 2025.

L’obiettivo è confrontare l’andamento degli accessi nelle due aree con il trend provinciale e verificare se l’aggregazione territoriale nasconda differenze rilevanti tra comuni, utili a motivare un’analisi più estesa della variabilità comunale. Le evidenze presentate in questa sezione hanno quindi un ruolo prevalentemente esplorativo: non mirano a trarre conclusioni generali sull’intera Provincia di Trento, ma a mostrare come il passaggio da dati aggregati provinciali a dati con maggiore dettaglio territoriale permetta di osservare differenze che resterebbero altrimenti nascoste.

Per rendere confrontabili comuni e territori caratterizzati da dimensioni demografiche diverse, i volumi assoluti sono stati rapportati alla

popolazione residente. Per ciascun indicatore è stato quindi calcolato un tasso di accesso per 1.000 abitanti:

$$\text{Tasso}_{it} = \frac{\text{Accessi}_{it}}{\text{Popolazione}_{it}} \times 1000,$$

dove i indica il comune o l'area territoriale e t il mese di riferimento.

L'indicatore esprime il numero di accessi registrati ogni 1.000 residenti e non la percentuale di individui che hanno utilizzato la piattaforma. Uno stesso soggetto può infatti effettuare più accessi nel medesimo periodo e contribuire più volte al numeratore.

2.2.2 Confronto con il trend provinciale

Come evidenziato nell'analisi dei dati aggregati provinciali, a partire dal 2024 si osserva una marcata crescita degli accessi, associata soprattutto al maggiore utilizzo dell'applicazione mobile. Il confronto con Val di Sole e Tesino permette quindi di verificare se la stessa dinamica sia presente anche nelle due aree considerate.

La Figura 2.6 mostra che Tesino e Val di Sole seguono solo in parte il trend provinciale. La forte accelerazione osservata a livello provinciale dal 2024 non si riproduce con la stessa intensità nelle due aree, dove l'andamento degli accessi totali appare più regolare.

La scomposizione per canale permette di comprendere meglio questa differenza. In Val di Sole e Tesino i tassi di accesso tramite applicazione rimangono generalmente inferiori rispetto al dato provinciale e mostrano una crescita più graduale, senza presentare lo stesso aumento osservato nella serie provinciale.

Per il portale web emerge invece una situazione differente. Val di Sole e Tesino presentano tassi di utilizzo superiori alla media provinciale per gran parte del periodo osservato. In queste aree il portale sembra quindi

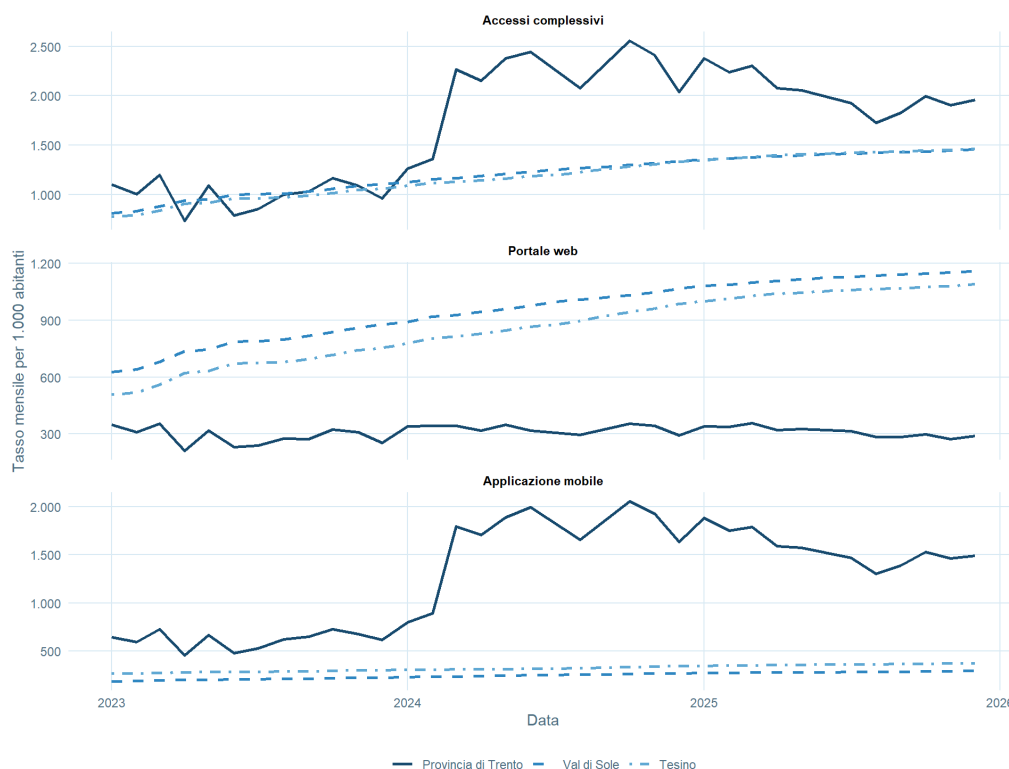


Figura 2.6: Confronto dei tassi mensili di accesso per 1.000 abitanti tra Provincia Autonoma di Trento, Tesino e Val di Sole, distinti per accessi totali, portale web e applicazione mobile.

mantenere un ruolo più rilevante, mentre il passaggio verso l'applicazione mobile appare meno marcato rispetto al quadro provinciale.

Il confronto mostra che quindi la crescita degli accessi registrata a livello provinciale non si distribuisce in modo uniforme sul territorio. Val di Sole e Tesino presentano una minore accelerazione nell'utilizzo dell'applicazione e un peso relativamente maggiore del portale web. Queste differenze rendono utile considerare anche alcune caratteristiche demografiche e territoriali delle due aree. L'obiettivo non è individuare relazioni causali, ma introdurre alcuni elementi utili a interpretare le differenze emerse nel confronto precedente.

2.2.3 Caratteristiche demografiche e territoriali delle due aree

Una prima dimensione riguarda la struttura per età della popolazione residente. Una popolazione mediamente più anziana potrebbe infatti essere caratterizzata da una minore propensione all'utilizzo dei canali mobili e da un ricorso relativamente maggiore a modalità di accesso più tradizionali, come il portale web. La Tabella 2.3 riporta l'età media e la percentuale media di residenti con più di 65 anni nelle due aree considerate e nell'intera Provincia di Trento, con riferimento al periodo 2023-2025.

Tabella 2.3: Confronto della struttura per età della popolazione tra Tesino, Val di Sole e Provincia di Trento. Valori medi sul periodo 2023-2025.

Area	Età media	% Over 65
Tesino	50,7	30,9
Val di Sole	46,5	23,9
Provincia di Trento	45,2	22,3

Dal confronto emerge che il Tesino presenta una struttura demografica più anziana rispetto alla media provinciale: l'età media è pari a 50,7 anni e la quota di popolazione con più di 65 anni raggiunge il 30,9%, contro rispettivamente 45,2 anni e 22,3% a livello provinciale. Anche la Val di Sole mostra valori superiori alla media provinciale, sebbene in misura più contenuta, con un'età media pari a 46,5 anni e una quota di popolazione over 65 pari al 23,9%. Accanto alla dimensione demografica, è stata considerata anche una caratteristica territoriale di contesto, rappresentata dall'altitudine media ponderata per superficie.

L'altitudine non viene interpretata come una causa diretta delle diverse modalità di utilizzo di TreC+, ma come una possibile proxy di caratteristiche territoriali più ampie, potenzialmente associate a dimensioni quali perifericità e accessibilità, inclusa la qualità della copertura e del segnale delle reti mobili. I risultati sono riportati nella

Tabella 2.4, mentre la Figura 2.7 mostra la distribuzione spaziale dell'altitudine media nei comuni della provincia, evidenziando le aree del Tesino e della Val di Sole.

Tabella 2.4: Altitudine media ponderata per superficie di Tesino, Val di Sole e Provincia di Trento.

Area	Altitudine media ponderata (m s.l.m.)
Provincia di Trento	1.402
Tesino	1.554
Val di Sole	2.011

Entrambe le aree considerate presentano un'altitudine media ponderata superiore a quella provinciale. In particolare, la Val di Sole registra un valore pari a 2.011 metri, mentre il Tesino presenta un valore pari a 1.554 metri, a fronte di una media provinciale pari a 1.402 metri. Questo risultato conferma il carattere più montano delle due aree analizzate, soprattutto nel caso della Val di Sole.

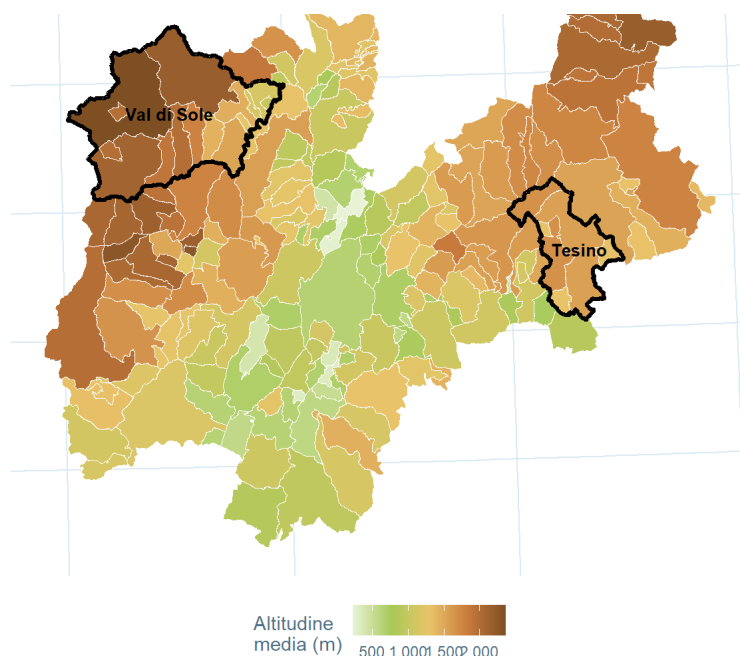


Figura 2.7: Distribuzione dell'altitudine media nei comuni della Provincia di Trento.

Le due aree considerate presentano alcune caratteristiche che le distinguono dal quadro provinciale, tra cui una popolazione mediamente

più anziana, soprattutto nel Tesino, e valori di altitudine media più elevati, in particolare nella Val di Sole. Tuttavia, il numero limitato di comuni osservati e la natura esplorativa dell'analisi non consentono di trarre conclusioni solide sul rapporto tra queste caratteristiche e le modalità di accesso alla piattaforma. Per approfondire la variabilità territoriale e il possibile ruolo dei fattori demografici e geografici, si rende quindi necessario passare al dataset a livello di singolo accesso, esteso all'intero territorio provinciale.

2.2.4 Analisi comunale

I dati comunali consentono di verificare se Val di Sole e Tesino siano internamente omogenee oppure se emergano differenze tra i comuni appartenenti alla stessa area.

La Figura 2.8 mostra l'andamento del tasso totale di accesso per 1.000 abitanti nei comuni del Tesino e della Val di Sole. La linea tratteggiata rappresenta, in ciascun pannello, il tasso medio mensile dell'area considerata, permettendo di confrontare le traiettorie dei singoli comuni con l'andamento complessivo della propria zona.

Nel Tesino, composto da tre comuni, le traiettorie risultano complessivamente simili e seguono un andamento piuttosto omogeneo. Anche i tassi medi comunali calcolati sull'intero periodo risultano molto vicini, variando tra 1.177,0 e 1.197,4 accessi per 1.000 abitanti.

In Val di Sole emerge invece una maggiore variabilità interna. Alcuni comuni si collocano stabilmente su livelli di accesso superiori rispetto alla media dell'area, mentre altri presentano valori più contenuti. I tassi medi comunali variano infatti tra 1.104,0 e 1.368,5 accessi per 1.000 abitanti, con una differenza decisamente più ampia rispetto a quella osservata nel Tesino.

Pur considerando la diversa numerosità dei comuni nelle due aree, il confronto mostra come l'aggregazione territoriale possa nascondere

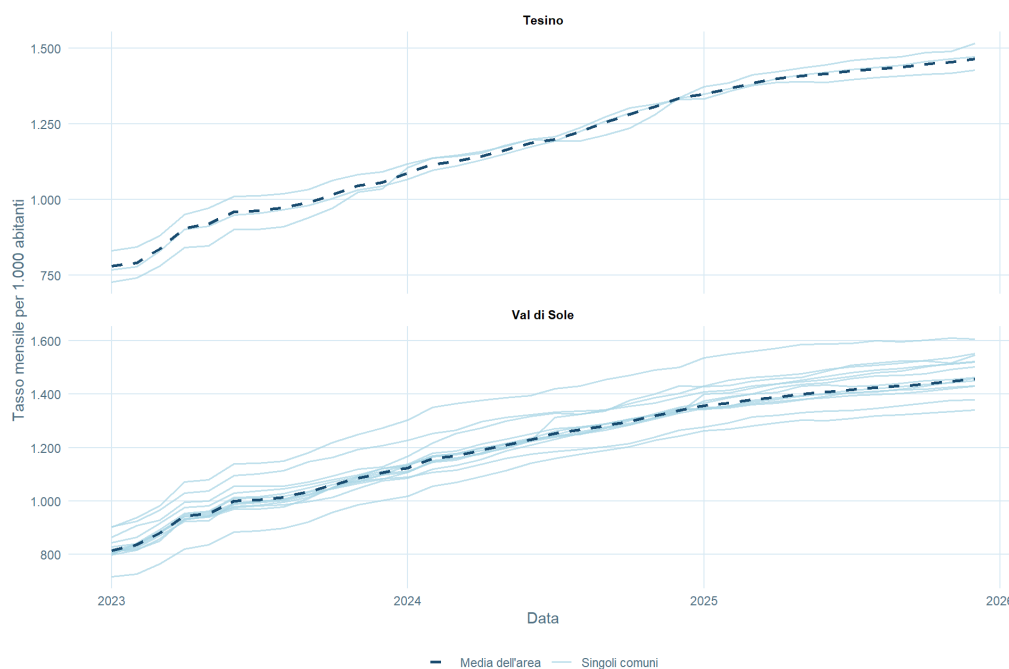


Figura 2.8: Andamento del tasso totale di accesso per 1.000 abitanti nei comuni del Tesino e della Val di Sole. La linea tratteggiata rappresenta il tasso medio mensile dell’area.

differenze rilevanti nei livelli di utilizzo della piattaforma. In particolare, la Val di Sole appare caratterizzata da una maggiore eterogeneità tra i comuni.

Il numero limitato di comuni osservati e la specificità delle due aree non consentono tuttavia di approfondire in modo robusto il possibile ruolo delle caratteristiche demografiche e territoriali. Per questo motivo, l’analisi viene successivamente estesa al dataset a livello di singolo accesso, che copre l’intero territorio provinciale e permette di analizzare con maggiore dettaglio le differenze temporali, demografiche e territoriali nell’utilizzo di TreC+.

2.3 Analisi dei dati a livello di singolo accesso

Le analisi presentate in questa sezione si basano sul dataset a livello di singolo accesso descritto nel paragrafo 1.2.3. Rispetto ai dati aggregati analizzati nelle sezioni precedenti, questi dati consentono di osservare

ogni accesso alla piattaforma TreC+ insieme alle relative caratteristiche dell'accesso e dell'utente.

L'unità di osservazione è il singolo accesso alla piattaforma e non l'utente unico. Uno stesso individuo può quindi comparire più volte nel dataset e le distribuzioni analizzate descrivono il profilo degli accessi registrati, non necessariamente quello degli utenti complessivi di TreC+.

2.3.1 Andamento temporale degli accessi

La Figura 2.9 mostra il numero mensile di accessi registrati tra agosto 2024 e aprile 2026. Il mese di maggio 2026 non è stato incluso nella rappresentazione, poiché i dati disponibili si interrompono all'11 maggio e non risultano quindi confrontabili con quelli dei mesi completi.

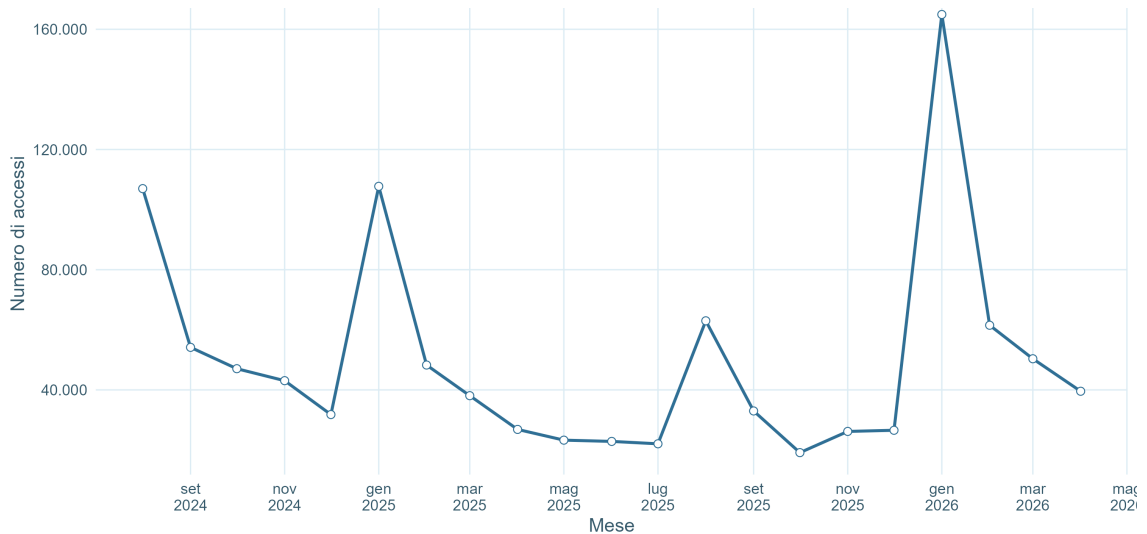


Figura 2.9: Andamento mensile degli accessi alla piattaforma TreC+ da agosto 2024 ad aprile 2026.

L'andamento non appare regolare, ma è caratterizzato da alcuni aumenti particolarmente marcati, osservabili soprattutto nei mesi di agosto 2024, gennaio 2025, agosto 2025 e gennaio 2026. Le analisi aggiuntive riportate in Appendice B mostrano che questi picchi non sono riconducibili esclusivamente a singole giornate isolate, ma coinvolgono periodi più estesi. Sebbene risultino particolarmente

associati alle modalità legate all'applicazione, gli incrementi interessano anche le altre modalità di autenticazione. I picchi sembrano quindi corrispondere a fasi di maggiore attività sulla piattaforma, pur non essendo possibile individuarne con certezza le cause.

Per valutare la presenza di differenze ricorrenti nel corso della settimana, la Figura 2.10 riporta la distribuzione del numero giornaliero di accessi per giorno della settimana.

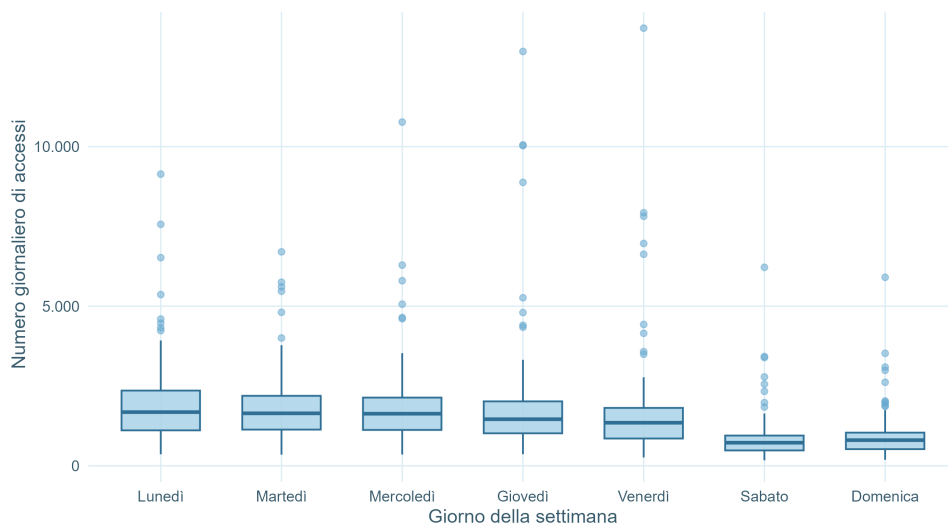


Figura 2.10: Distribuzione del numero giornaliero di accessi alla piattaforma TreC+ per giorno della settimana.

La distribuzione risulta abbastanza simile tra i giorni feriali, mentre il sabato e la domenica presentano generalmente un numero inferiore di accessi. L'utilizzo della piattaforma appare quindi maggiormente concentrato nei giorni lavorativi, anche se la variabilità osservata all'interno di ciascun giorno segnala la presenza di periodi caratterizzati da volumi particolarmente elevati.

È stata infine considerata la distribuzione degli accessi rispetto all'ora del giorno, riportata nella Figura 2.11.

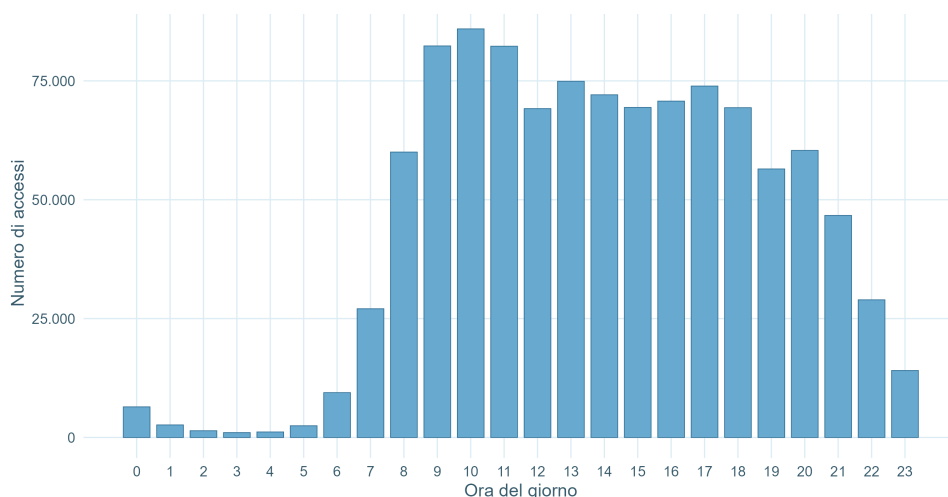


Figura 2.11: Distribuzione complessiva degli accessi alla piattaforma TreC+ per ora del giorno.

Gli accessi risultano fortemente concentrati nelle ore diurne. Il loro numero aumenta a partire dal mattino, raggiunge i valori più elevati tra le 9 e le 11 e rimane sostenuto anche nel corso del pomeriggio. Successivamente diminuisce nelle ore serali, mentre durante la notte, in particolare tra mezzanotte e le 6, i volumi risultano molto più contenuti.

2.3.2 Modalità di autenticazione

Nel dataset, la modalità di autenticazione consente di distinguere il canale attraverso cui è stato effettuato l'accesso e il sistema di identificazione utilizzato. APP e APP0 indicano gli accessi effettuati direttamente dall'applicazione, con APP0 riferita a una versione non aggiornata dell'app. SPID_APP e TEAM_APP riguardano invece gli accessi da applicazione autenticati tramite SPID o TEAM. Le modalità TEAM, SPID e CIE identificano gli accessi effettuati dal portale mediante i rispettivi sistemi di autenticazione, mentre DELEGA indica quelli eseguiti per conto di un altro utente.

La Tabella 2.5 riporta la distribuzione complessiva degli accessi. Le modalità più frequenti sono APP e TEAM, entrambe pari al 25,6% del totale, seguite da SPID, con il 20,1%, e SPID_APP, con il 13,8%.

TEAM_APP e CIE rappresentano rispettivamente il 6,8% e il 4,8% degli accessi, mentre APP0 presenta una frequenza residuale.

Tabella 2.5: Distribuzione degli accessi per modalità di autenticazione.

Modalità di autenticazione	Numero di accessi	Percentuale
APP	273.877	25,6%
TEAM	272.982	25,6%
SPID	214.617	20,1%
SPID_APP	147.301	13,8%
TEAM_APP	73.070	6,8%
CIE	50.913	4,8%
DELEGA	35.053	3,3%
APP0	525	0,05%

Considerando congiuntamente APP, APP0, SPID_APP e TEAM_APP, le modalità collegate all'applicazione rappresentano circa il 46% degli accessi. Rimane rilevante anche il ruolo del portale, soprattutto attraverso TEAM e SPID.

La percentuale relativa a DELEGA deve essere interpretata con cautela, poiché questa modalità compare soltanto nella parte finale del periodo osservato, a partire dalla fine di dicembre 2025, e non è quindi direttamente confrontabile con quelle rilevate lungo l'intera finestra temporale.

Per analizzare l'evoluzione delle diverse modalità sono stati successivamente considerati i conteggi mensili. Nel grafico temporale APP0 è stata aggregata ad APP, data la sua numerosità molto ridotta, mentre DELEGA è stata esclusa a causa della sua disponibilità limitata nel tempo.

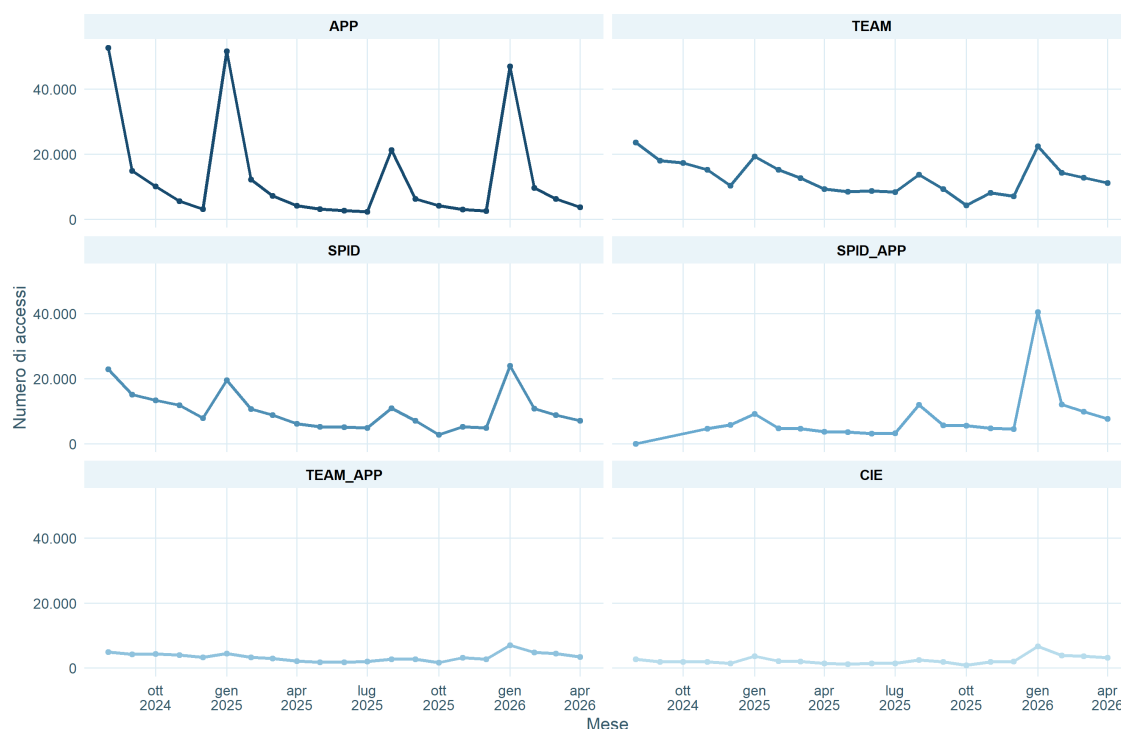


Figura 2.12: Andamento mensile degli accessi per modalità di autenticazione.

La Figura 2.12 mostra andamenti differenti tra le modalità considerate. APP presenta la maggiore variabilità e contribuisce in modo rilevante ai principali aumenti osservati nella serie. Come evidenziato anche nell'Appendice B, nelle giornate caratterizzate dai volumi più elevati aumenta soprattutto il peso di questa modalità, sebbene la crescita interessi anche gli altri sistemi di autenticazione.

TEAM e SPID mostrano un andamento più regolare e mantengono volumi elevati lungo gran parte del periodo. In particolare, TEAM rappresenta una delle principali modalità di accesso per l'intera finestra temporale, mentre SPID conserva un ruolo rilevante tra gli accessi effettuati dal portale.

SPID_APP risulta invece poco presente nella parte iniziale della serie, ma cresce progressivamente e assume un peso maggiore nei mesi più recenti. Questo andamento segnala una crescente diffusione dell'accesso tramite applicazione associato all'identità digitale SPID. Anche TEAM_APP e CIE mostrano un aumento nel tempo, pur mantenendo

volumi inferiori rispetto alle modalità principali.

2.3.3 Profilo demografico degli accessi

Le analisi demografiche si riferiscono agli accessi registrati e non agli utenti unici. Uno stesso individuo può quindi contribuire più volte ai conteggi osservati. I risultati descrivono pertanto il profilo demografico degli accessi alla piattaforma e non la distribuzione degli utenti che l'hanno utilizzata.

Accessi e tassi per fascia d'età

Per analizzare la distribuzione degli accessi rispetto all'età, la popolazione è stata suddivisa nelle fasce 0–17, 18–34, 35–49, 50–64, 65–74, 75–84 e 85 anni o più. Per ciascuna fascia sono stati considerati due indicatori: il numero complessivo di accessi osservati e il numero complessivo di accessi per 1.000 residenti nel periodo considerato. Il primo indicatore descrive il contributo di ciascuna fascia d'età al volume complessivo degli accessi. Il secondo rapporta invece il numero totale di accessi alla popolazione media residente nella stessa fascia nel periodo 2024–2026, secondo la relazione:

$$\text{Accessi per 1.000 residenti} = \frac{\text{Accessi nella fascia}}{\text{Popolazione media della fascia}} \times 1.000.$$

Poiché uno stesso soggetto può effettuare più accessi, l'indicatore non rappresenta la quota di individui che hanno utilizzato la piattaforma, ma l'intensità complessiva degli accessi rispetto alla numerosità della popolazione appartenente a ciascuna fascia d'età.

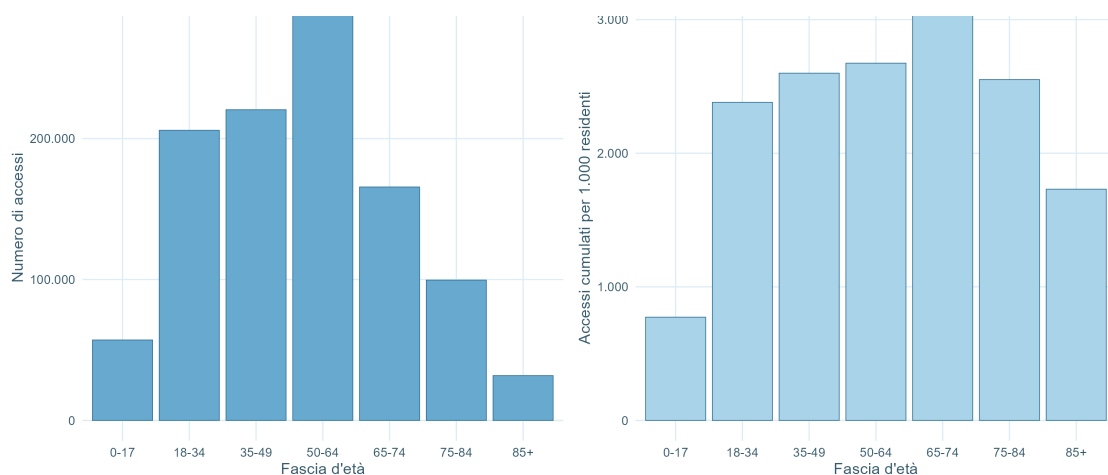


Figura 2.13: Numero complessivo di accessi e numero di accessi per 1.000 residenti, distinti per fascia d'età.

Il pannello di sinistra della Figura 2.13 mostra il numero complessivo di accessi registrati in ciascuna fascia d'età. In termini assoluti, gli accessi si concentrano soprattutto nelle fasce adulte, mentre risultano meno numerosi tra i minorenni e nelle età più avanzate. I conteggi dipendono tuttavia anche dalla struttura per età della popolazione residente: una fascia più numerosa può infatti registrare un numero elevato di accessi anche senza presentare una particolare intensità di utilizzo della piattaforma.

Il pannello di destra rapporta gli accessi alla popolazione media di ciascuna fascia, consentendo di confrontare gruppi caratterizzati da dimensioni demografiche differenti. Le due distribuzioni non coincidono completamente. Le fasce adulte mantengono livelli elevati anche dopo aver tenuto conto della rispettiva popolazione, mentre le fasce di età più avanzata assumono un rilievo maggiore rispetto a quanto emerge dai soli conteggi assoluti. Pur contribuendo meno al volume complessivo, queste fasce presentano infatti un numero relativamente elevato di accessi in rapporto ai residenti. La fascia 0–17 anni mostra invece valori contenuti sia in termini assoluti sia rispetto alla propria popolazione.

La lettura congiunta dei due pannelli consente quindi di osservare come

cambia il peso relativo delle diverse fasce d'età quando il numero di accessi viene rapportato alla popolazione residente.

Distribuzione per sesso

Per analizzare le differenze rispetto al sesso sono stati considerati sia il numero complessivo di accessi sia il tasso per 1.000 residenti separatamente per maschi e femmine. Il tasso è stato calcolato rapportando gli accessi associati ai soggetti di sesso femminile e maschile alla rispettiva popolazione media residente nel periodo 2024–2026.

Tabella 2.6: Numero di accessi, distribuzione percentuale e tasso cumulato per sesso.

Sesso	Accessi	Percentuale	Popolazione media	Tasso per 1.000
Femminile	595.541	55,7%	234.447	2.540,2
Maschile	472.797	44,3%	230.373	2.052,3

In termini assoluti, il 55,7% degli accessi è associato a soggetti di sesso femminile, mentre il restante 44,3% riguarda soggetti di sesso maschile. La popolazione femminile risulta soltanto leggermente più numerosa di quella maschile; è quindi utile confrontare i conteggi con i tassi calcolati rispetto alla popolazione dello stesso sesso.

Anche dopo aver tenuto conto della diversa numerosità della popolazione, il tasso rimane più elevato tra le donne. Nel periodo considerato sono stati registrati 2.540,2 accessi per 1.000 residenti di sesso femminile e 2.052,3 accessi per 1.000 residenti di sesso maschile. La maggiore presenza degli accessi femminili non sembra quindi dipendere esclusivamente dalla composizione per sesso della popolazione residente. È comunque importante ricordare che si tratta di tassi specifici per sesso e non di tassi standardizzati per età. Le differenze osservate potrebbero quindi risentire anche della diversa struttura per età della popolazione femminile e maschile.

Età e modalità di autenticazione

È stata infine analizzata la distribuzione dell'età in relazione alla modalità di autenticazione, per verificare se i diversi sistemi di accesso fossero associati a profili anagrafici differenti. La modalità APP0 è stata aggregata ad APP, data la sua numerosità residuale, mentre DELEGA è stata esclusa perché disponibile soltanto nella parte finale del periodo osservato.

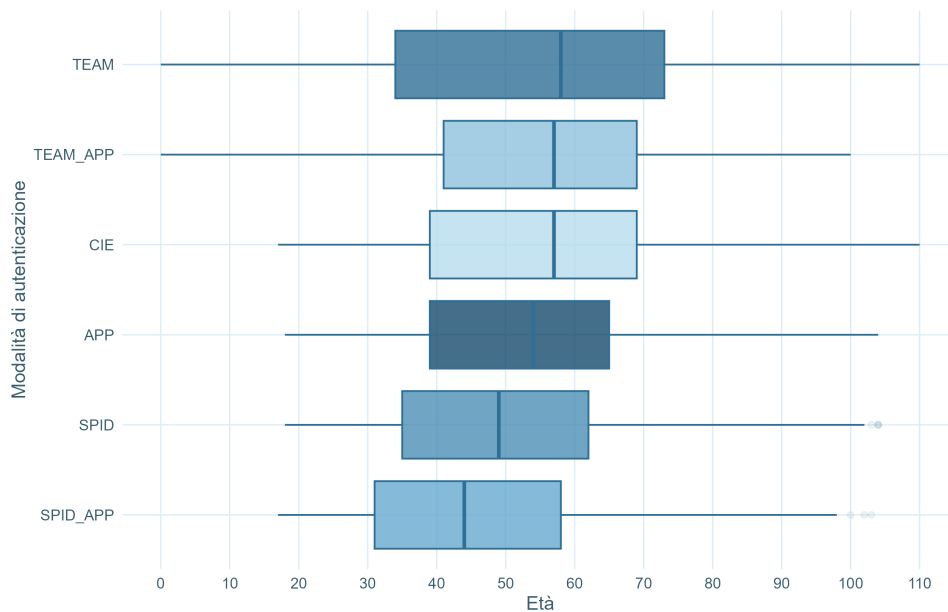


Figura 2.14: Distribuzione dell'età per modalità di autenticazione.

La Figura 2.14 mostra differenze nella distribuzione dell'età tra le modalità considerate. Gli accessi effettuati tramite SPID_APP e SPID presentano le età mediane più basse tra le modalità considerate. Ciò indica che questi sistemi vengono utilizzati da soggetti mediamente più giovani. APP e CIE occupano invece una posizione intermedia, mentre TEAM_APP e TEAM risultano associate a età mediane più elevate. Le distribuzioni di TEAM e TEAM_APP sono inoltre più ampie e comprendono accessi riferiti a un intervallo di età più esteso. Queste modalità risultano quindi utilizzate sia nelle fasce più giovani sia in

quelle adulte e anziane, anche se la parte centrale della distribuzione è spostata verso età più elevate rispetto alle modalità basate su SPID.

2.3.4 Distribuzione territoriale e tassi comunali di accesso

L'analisi territoriale è stata inizialmente condotta considerando il numero complessivo di accessi registrati in ciascun comune nel periodo osservato. I conteggi assoluti descrivono dove si concentra il maggiore volume di accessi, ma risentono fortemente della diversa dimensione demografica dei territori e non consentono quindi un confronto diretto tra comuni.

Il numero più elevato di accessi si osserva nel comune di Trento, con 225.649 accessi, pari al 21,1% del totale provinciale. Seguono Pergine Valsugana, con 59.500 accessi, Riva del Garda, con 40.623, Mori, con 27.950, e Lavis, con 23.341.

Lo stesso andamento emerge a livello di Comunità di Valle. La Val d'Adige raccoglie 227.713 accessi, pari al 21,3% del totale, seguita da Alta Valsugana e Bersntol, con 153.104 accessi, Vallagarina, con 116.272, e Val di Non, con 103.458. Complessivamente, queste quattro aree comprendono oltre la metà degli accessi osservati.

2.3.5 Standardizzazione per sesso e fascia d'età

Poiché la frequenza degli accessi varia rispetto al sesso e all'età, i confronti territoriali possono essere influenzati dalla diversa composizione demografica dei comuni. I tassi sono stati pertanto standardizzati mediante il metodo diretto.

Per ciascun comune c , sesso s e fascia d'età e , è stato inizialmente calcolato il tasso specifico:

$$\text{Tasso specifico}_{c,s,e} = \frac{\text{Accessi}_{c,s,e}}{\text{Popolazione media}_{c,s,e}} \times 1.000.$$

I tassi specifici sono stati successivamente ponderati utilizzando come popolazione standard la popolazione media complessiva dei comuni osservati, distinta per sesso e nelle fasce d'età 0–17, 18–34, 35–49, 50–64, 65–74, 75–84 e 85 anni o più. Il tasso standardizzato del comune c è quindi definito come:

$$\text{Tasso standardizzato}_c = \sum_{s,e} w_{s,e} \text{Tasso specifico}_{c,s,e},$$

dove $w_{s,e}$ rappresenta il peso dello strato definito dal sesso s e dalla fascia d'età e nella popolazione standard.

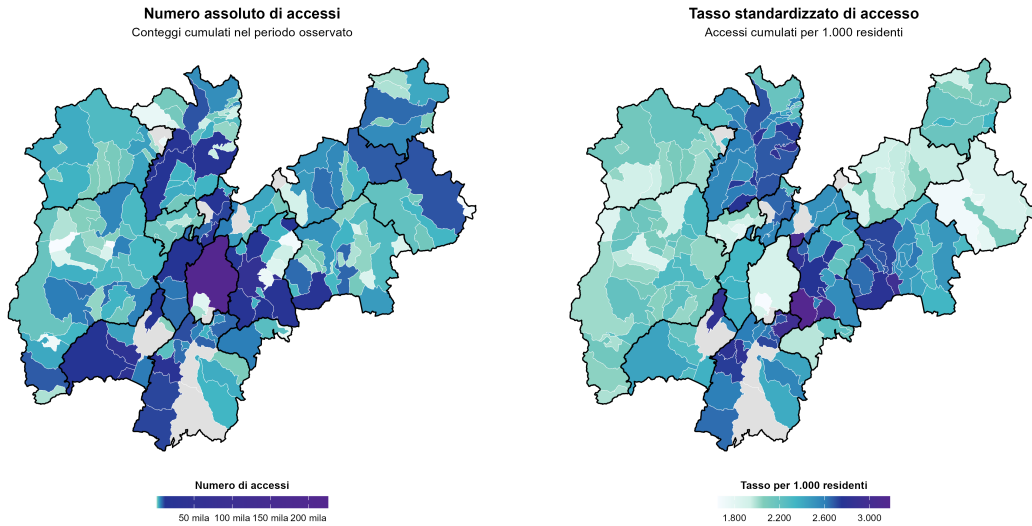


Figura 2.15: Distribuzione comunale del numero assoluto di accessi, a sinistra, e del tasso cumulato di accesso standardizzato per sesso e fascia d'età, a destra.

La Figura 2.15 mette a confronto due diverse letture della distribuzione territoriale. La mappa di sinistra evidenzia soprattutto la concentrazione degli accessi nei comuni più popolosi: Trento emerge nettamente rispetto al resto del territorio, seguito dai comuni nominati sopra. La mappa di destra mostra invece il tasso standardizzato, che tiene conto sia della diversa numerosità della popolazione residente sia della sua composizione per sesso e fascia d'età. La lettura territoriale cambia: Trento, pur concentrando oltre un quinto degli accessi assoluti, presenta

un tasso standardizzato pari a 1.897,9 accessi per 1.000 residenti, tra i valori relativamente più contenuti. Emergono invece alcuni comuni caratterizzati da un volume assoluto inferiore ma da una maggiore intensità relativa di accesso.

Il tasso standardizzato più elevato si osserva nell'Altopiano della Vigolana, con 3.180,2 accessi per 1.000 residenti. Seguono Albiano, con 3.049,3, Caldonazzo, con 3.024,1, Besenello, con 2.953,0, e Castelnuovo, con 2.947,2.

Anche il confronto tra Comunità di Valle cambia rispetto a quello basato sui soli conteggi. La Val d'Adige, che raccoglie il maggior numero assoluto di accessi, presenta un tasso standardizzato pari a 1.896,9 accessi per 1.000 residenti. I valori più elevati si osservano invece in Alta Valsugana e Bersntol, con 2.731,7 accessi per 1.000 residenti, Vallagarina, con 2.679,3, Valle di Cembra, con 2.628,2, e Valsugana e Tesino, con 2.627,0. La maggiore concentrazione assoluta degli accessi nella Val d'Adige sembra quindi dipendere soprattutto dalla sua dimensione demografica, mentre altre aree presentano livelli relativi di accesso più elevati.

Per le Comunità di Valle, i tassi sono stati standardizzati direttamente a tale livello territoriale, aggregando gli accessi e la popolazione per sesso e fascia d'età e applicando la stessa popolazione standard utilizzata per i comuni.

I tassi standardizzati comunali variano da 1.655 a 3.180 accessi per 1.000 residenti. La mediana è pari a 2.341 e il 50% dei comuni presenta valori compresi tra 2.060 e 2.611 accessi per 1.000 residenti. La variabilità territoriale rimane quindi presente anche dopo aver tenuto conto della composizione demografica.

2.3.6 Associazione con le caratteristiche comunali

È stata infine esplorata la relazione tra il tasso di accesso e due caratteristiche comunali: l'altitudine media e l'età media della popolazione residente. Le associazioni sono state valutate mediante il coefficiente di correlazione di Pearson, che misura la direzione e l'intensità della relazione lineare tra due variabili quantitative. Le relazioni sono state inoltre rappresentate mediante diagrammi di dispersione. Ciascun punto corrisponde a un comune, mentre la retta di regressione sintetizza l'andamento medio della relazione osservata.

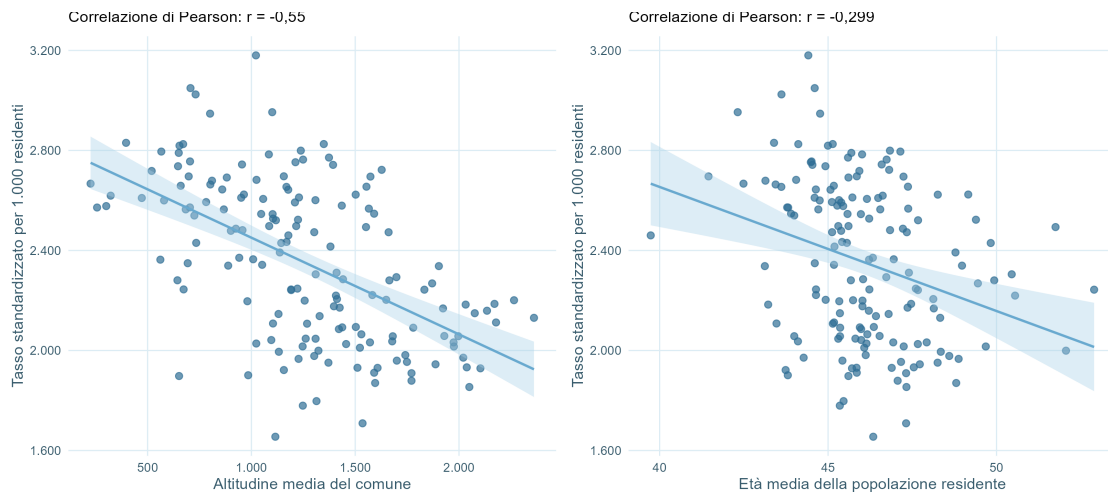


Figura 2.16: Relazione tra il tasso standardizzato di accesso e l'altitudine media del comune, a sinistra, e l'età media della popolazione residente, a destra.

Come mostrato nella Figura 2.16, entrambe le associazioni risultano negative. La relazione con l'altitudine media è la più marcata: il coefficiente di Pearson è pari a $-0,550$, indicando un'associazione lineare negativa. I comuni situati ad altitudini maggiori tendono quindi a presentare tassi standardizzati di accesso più bassi.

Anche l'età media della popolazione residente è negativamente associata al tasso standardizzato, ma la relazione è più debole. Il coefficiente di Pearson è pari a $-0,299$, indicando che i comuni con una popolazione mediamente più anziana tendono a mostrare livelli relativi di accesso

più contenuti, sebbene la dispersione dei punti intorno alla retta sia maggiore rispetto a quella osservata per l'altitudine.

Le relazioni osservate hanno natura esplorativa. I risultati indicano che i comuni con maggiore altitudine e con una popolazione mediamente più anziana tendono a presentare tassi standardizzati di accesso più bassi, ma non consentono di stabilire un rapporto di causa-effetto. Le associazioni potrebbero infatti dipendere anche da altre caratteristiche territoriali e demografiche non considerate nell'analisi.

Capitolo 3

Metodi

3.1 Obiettivo

L'obiettivo dell'analisi è individuare una rete di associazioni residue tra le serie comunali degli accessi al Fascicolo Sanitario Elettronico, verificando se alcuni comuni presentino comportamenti simili nell'utilizzo del servizio anche dopo aver tenuto conto delle informazioni disponibili.

A questo scopo, viene inizialmente valutato il contributo delle principali caratteristiche demografiche e geografiche del territorio, dell'eterogeneità tra i comuni e dell'andamento temporale. La componente sistematica dei conteggi viene modellata mediante un modello additivo generalizzato, dal quale si ricavano i residui, che rappresentano la parte di variabilità non spiegata dagli effetti considerati.

Su tali residui viene applicato il Graphical Lasso, un metodo che consente di stimare una rete sparsa di dipendenze condizionate. I nodi identificano i comuni, mentre la presenza di un arco indica che due serie mantengono un'associazione dopo aver considerato sia le componenti già modellate sia le relazioni con tutti gli altri comuni inclusi nell'analisi. La rete permette quindi di evidenziare una possibile struttura di dipendenza residua non direttamente spiegata dalle informazioni disponibili.

3.2 Modelli grafici gaussiani

3.2.1 Modelli grafici e dipendenze condizionate

I modelli grafici consentono di rappresentare la struttura di dipendenza tra un insieme di variabili attraverso un grafo. Un grafo non orientato può essere indicato come

$$G = (V, E),$$

dove V è l'insieme dei nodi ed E è l'insieme degli archi che collegano alcune coppie di nodi. Il carattere non orientato del grafo implica che gli archi non possiedano una direzione: il collegamento tra due nodi descrive una relazione simmetrica e non stabilisce un rapporto causale tra le variabili corrispondenti (Lauritzen, 1996).

Una prima possibilità per studiare la relazione tra due variabili consiste nel considerarne la correlazione marginale. Tale associazione può tuttavia essere indiretta, poiché può dipendere dal legame di entrambe le variabili con una terza componente.

I modelli grafici considerano invece le dipendenze condizionate, valutando la relazione tra due variabili dopo aver tenuto conto simultaneamente di tutte le altre incluse nell'analisi. In un grafo non orientato, l'assenza di un arco tra i nodi i e j rappresenta tale condizione di indipendenza. La presenza di un arco indica invece che tra le due variabili rimane un'associazione anche dopo aver considerato il contributo delle restanti variabili (Lauritzen, 1996).

3.2.2 Il caso gaussiano e la matrice di precisione

Nei modelli grafici gaussiani si assume che il vettore delle p variabili considerate,

$$\mathbf{X} = (X_1, \dots, X_p)^\top,$$

segua una distribuzione normale multivariata:

$$\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma),$$

dove $\boldsymbol{\mu}$ è il vettore delle medie e Σ è la matrice di covarianza.

Gli elementi fuori diagonale della matrice di covarianza,

$$\Sigma_{ij} = \text{Cov}(X_i, X_j),$$

descrivono le associazioni marginali tra le variabili. Per rappresentare le dipendenze condizionate si considera invece la matrice di precisione, definita come l'inversa della matrice di covarianza:

$$\Omega = \Sigma^{-1}.$$

Nel caso gaussiano, la struttura degli zeri degli elementi fuori diagonale di Ω determina la struttura del grafo. In particolare, un elemento nullo della matrice di precisione corrisponde quindi all'assenza di un arco tra i nodi i e j e indica che le due variabili sono condizionatamente indipendenti, date tutte le altre. Un elemento diverso da zero segnala invece la presenza di un'associazione condizionata e, di conseguenza, di un collegamento tra i due nodi (Zhao *et al.*, 2012).

La stima del grafo può pertanto essere ricondotta all'individuazione della struttura degli zeri della matrice di precisione. Stabilire quali elementi fuori diagonale di Ω sono nulli equivale infatti a determinare quali coppie di comuni non risultano direttamente collegate, una volta considerate simultaneamente tutte le altre serie incluse nell'analisi (Meinshausen e Bühlmann, 2006).

3.3 Graphical Lasso

3.3.1 Interpretazione mediante regressioni condizionate

Il legame tra la matrice di precisione e la struttura del grafo può essere interpretato anche attraverso regressioni lineari condizionate. Per ciascuna variabile X_i , si considera la regressione su tutte le altre componenti del vettore:

$$X_i = \sum_{j \neq i} \beta_{ij} X_j + \varepsilon_i.$$

Nel caso gaussiano, i coefficienti di tale regressione sono legati agli elementi della matrice di precisione attraverso la relazione

$$\beta_{ij} = -\frac{\Omega_{ij}}{\Omega_{ii}}.$$

Ne consegue che

$$\beta_{ij} = 0 \quad \Longleftrightarrow \quad \Omega_{ij} = 0.$$

Il vicinato del nodo i può quindi essere identificato con l'insieme delle variabili che presentano un coefficiente non nullo nella regressione di X_i su tutte le restanti variabili. La ricostruzione del grafo può pertanto essere ricondotta a un problema di selezione delle variabili, nel quale si cerca di individuare, per ciascun nodo, quali componenti appartengano al suo vicinato (Meinshausen e Bühlmann, 2006).

3.3.2 Dalla selezione dei vicinati al Graphical Lasso

Su questa interpretazione si basa la *neighborhood selection* proposta da Meinshausen e Bühlmann (2006). Per ciascun nodo viene stimata una regressione Lasso della variabile corrispondente su tutte le altre e i coefficienti non nulli forniscono una stima del suo vicinato.

La penalizzazione L_1 è particolarmente adatta a questo scopo, poiché può produrre coefficienti esattamente nulli e, al tempo stesso, mantiene convesso il problema di ottimizzazione. Essa consente quindi di effettuare la selezione delle variabili anche in contesti ad alta dimensionalità.

L'approccio basato sulla selezione dei vicini può essere interpretato come un'approssimazione del problema esatto di massimizzazione della verosimiglianza gaussiana penalizzata. Il Graphical Lasso invece, affronta direttamente tale problema, permettendo di stimare congiuntamente una matrice di precisione sparsa (Friedman *et al.*, 2008). Indicando con S la matrice di covarianza campionaria e con Ω la matrice di precisione, la stima è definita come

$$\hat{\Omega}_\lambda = \arg \max_{\Omega \succ 0} \{ \log \det(\Omega) - \text{tr}(S\Omega) - \lambda \|\Omega\|_1 \},$$

dove il vincolo $\Omega \succ 0$ richiede che la matrice sia definita positiva, mentre $\lambda > 0$ controlla l'intensità della penalizzazione.

La penalizzazione L_1 favorisce una stima sparsa della matrice di precisione. Poiché i suoi elementi nulli fuori diagonale corrispondono ad assenze di archi, la soluzione del problema determina anche la struttura del grafo stimato.

3.3.3 Implementazione mediante il pacchetto `huge`

Per le analisi presentate nel Capitolo 4, il Graphical Lasso è stato applicato mediante il pacchetto `huge`, sviluppato per la stima di grafi non orientati in contesti ad alta dimensionalità (Zhao *et al.*, 2012).

La funzione `huge` consente di stimare un percorso di regolarizzazione corrispondente a una sequenza decrescente di valori di λ . A ciascun valore del parametro è associata una diversa matrice di precisione e, di conseguenza, una diversa struttura del grafo.

Per valori più elevati di λ , la penalizzazione è più intensa e la soluzione presenta un numero maggiore di elementi nulli. La rete risulta quindi più sparsa. Al diminuire di λ , viene ammesso un numero crescente di elementi non nulli e il grafo diventa progressivamente più denso. L'output della procedura è pertanto costituito da una successione di reti, tra le quali deve essere selezionata quella da utilizzare nelle analisi successive (Zhao *et al.*, 2012).

3.3.4 Selezione della rete

La rete lungo il percorso di regolarizzazione viene selezionata mediante StARS, *Stability Approach to Regularization Selection*, un metodo sviluppato per la scelta del parametro di regolarizzazione nei modelli grafici ad alta dimensionalità (Liu *et al.*, 2010). L'idea alla base del metodo è selezionare la rete che risulta sufficientemente stabile rispetto a piccole variazioni del campione osservato. Per ciascun valore di λ vengono estratti ripetutamente sottocampioni dei dati e, su ciascuno di essi, viene stimato un grafo. La stabilità è quindi valutata in base alla frequenza con cui gli archi compaiono nelle diverse reti stimate: una forte variabilità nella presenza degli archi indica una rete instabile. StARS seleziona il più piccolo valore di λ per cui l'instabilità complessiva della rete rimane al di sotto di una soglia prefissata.

3.4 Costruzione dell'input per l'analisi di rete

Il Graphical Lasso viene applicato a una matrice nella quale le righe rappresentano le osservazioni e le colonne le variabili. Nel presente contesto, le osservazioni corrispondono alle settimane, mentre le variabili sono associate ai comuni.

L'utilizzo diretto dei conteggi osservati non permetterebbe tuttavia di interpretare la rete come una struttura di dipendenze residue. Le serie

possono infatti riflettere componenti sistematiche già note, come l'andamento temporale generale, le differenze nella dimensione demografica e l'eterogeneità tra i territori. Tali elementi potrebbero generare associazioni tra le serie che non corrispondono alla struttura di dipendenza che si intende analizzare.

Prima della stima della rete viene quindi costruita una matrice residuale attraverso una sequenza di passaggi. In primo luogo, la componente sistematica dei conteggi viene modellata mediante un modello additivo generalizzato. Dai valori osservati e attesi vengono poi ricavati residui standardizzati. Successivamente, l'eventuale dipendenza temporale ancora presente nelle singole serie viene trattata mediante una procedura di sbiancamento. Infine, viene applicata una trasformazione nonparanormal (Liu *et al.*, 2009), finalizzata a rendere le distribuzioni marginali maggiormente compatibili con il quadro gaussiano.

Le sezioni successive descrivono il ruolo teorico di ciascun passaggio. Le specifiche scelte adottate nell'analisi, comprese la formulazione del modello e i parametri utilizzati, vengono invece presentate nel Capitolo 4.

3.5 Modelli additivi generalizzati

3.5.1 Dai modelli additivi ai modelli additivi generalizzati

Per rimuovere dai conteggi le componenti sistematiche è possibile ricorrere ai modelli additivi generalizzati, GAM. Questi modelli permettono di rappresentare relazioni non necessariamente lineari tra la variabile risposta e le covariate, senza doverne specificare a priori una forma parametrica rigida.

Il punto di partenza è il modello additivo, nel quale il valore atteso della

risposta viene espresso come somma degli effetti delle diverse covariate:

$$E(Y_i | \mathbf{x}_i) = \beta_0 + \sum_{j=1}^q f_j(x_{ij}),$$

dove β_0 rappresenta l'intercetta e $f_j(\cdot)$ è una funzione liscia associata alla j -esima covariata.

Le funzioni lisce vengono generalmente rappresentate mediante spline. Una spline è costruita a partire da funzioni polinomiali definite su intervalli successivi e raccordate in modo regolare in corrispondenza di alcuni punti, detti nodi. La funzione può essere espressa attraverso una combinazione di funzioni base:

$$f_j(x) = \sum_{k=1}^{K_j} b_{jk}(x) \gamma_{jk},$$

dove $b_{jk}(x)$ sono le funzioni base e γ_{jk} i relativi coefficienti.

La dimensione della base determina il livello massimo di flessibilità della funzione stimata. In fase di stima, i coefficienti γ_{jk} vengono determinati massimizzando un criterio penalizzato che bilancia l'adattamento ai dati e la regolarità della soluzione. L'intensità di tale penalizzazione è controllata dal parametro di smoothing λ : valori elevati di λ inducono soluzioni più regolari e quindi più lisce, mentre valori ridotti di λ consentono una maggiore flessibilità della funzione stimata.

I parametri di smoothing possono essere stimati mediante il criterio della verosimiglianza ristretta, o REML. (Wood, 2017).

I modelli additivi generalizzati estendono questa impostazione a variabili risposta che non seguono necessariamente una distribuzione normale. Indicando con

$$\mu_i = E(Y_i | \mathbf{x}_i),$$

la struttura generale del modello è

$$g(\mu_i) = \beta_0 + \sum_{j=1}^q f_j(x_{ij}),$$

dove $g(\cdot)$ è una funzione di collegamento e la distribuzione della risposta viene scelta in relazione alla natura dei dati. I GAM costituiscono pertanto un'estensione dei modelli lineari generalizzati, nella quale alcuni effetti lineari sono sostituiti da funzioni lisce stimate a partire dai dati (Wood, 2017).

3.5.2 Distribuzioni per dati di conteggio

Quando la variabile risposta è un conteggio, una prima possibilità è rappresentata dalla distribuzione di Poisson:

$$Y_i \mid \mathbf{x}_i \sim \text{Poisson}(\mu_i).$$

Nel modello di Poisson, la media e la varianza condizionate coincidono:

$$E(Y_i \mid \mathbf{x}_i) = \text{Var}(Y_i \mid \mathbf{x}_i) = \mu_i.$$

Per garantire che il valore atteso sia positivo, viene generalmente utilizzata la funzione di collegamento logaritmica:

$$\log(\mu_i) = \beta_0 + \sum_{j=1}^q f_j(x_{ij}).$$

L'ipotesi di uguaglianza tra media e varianza può tuttavia risultare troppo restrittiva. Si parla di sovradisersione quando la variabilità dei conteggi, condizionatamente alle covariate considerate, è maggiore di quella prevista dal modello di Poisson. In presenza di sovradisersione, il modello di Poisson può quindi sottostimare la variabilità della risposta.

Una possibile alternativa è costituita dalla distribuzione binomiale negativa:

$$Y_i \mid \mathbf{x}_i \sim \text{NegBin}(\mu_i, \theta),$$

la cui varianza può essere espressa come

$$\text{Var}(Y_i \mid \mathbf{x}_i) = \mu_i + \frac{\mu_i^2}{\theta}.$$

Il parametro θ regola la dispersione della distribuzione. Per valori elevati di θ , la componente quadratica si riduce e la binomiale negativa si avvicina al modello di Poisson; per valori più contenuti, la varianza può invece superare la media (Ver Hoef e Boveng, 2007).

La distribuzione di Poisson rappresenta quindi un possibile punto di partenza per la modellazione dei conteggi. Nel corso dell'analisi verrà valutata la presenza di sovradisersione e, sulla base delle verifiche effettuate, verrà scelta la distribuzione maggiormente adeguata tra quella di Poisson e la binomiale negativa.

3.5.3 Offset della popolazione

Nei dati territoriali, il numero atteso di eventi dipende anche dalla dimensione della popolazione esposta. A parità di comportamento nell'utilizzo del servizio, un comune più popoloso tende infatti a registrare un numero maggiore di accessi.

Nei modelli per conteggi con funzione di collegamento logaritmica, la popolazione può essere inclusa mediante un *offset*. Indicando con P_i la popolazione associata all'osservazione i , il predittore assume la forma

$$\log(\mu_i) = \log(P_i) + \eta_i,$$

dove η_i contiene gli effetti delle covariate e delle eventuali funzioni lisce. L'offset entra nel modello con coefficiente fissato pari a uno. Il modello continua quindi ad avere come risposta il numero di eventi osservati, ma

tiene conto della diversa esposizione delle unità considerate. In questo modo, gli effetti inclusi nel predittore descrivono variazioni nel numero atteso di accessi a parità di popolazione (Wood, 2017).

3.5.4 Residui di Pearson

Dopo la stima del modello, la parte non spiegata dalle componenti considerate viene rappresentata attraverso i residui. Nei modelli per conteggi, la semplice differenza tra valore osservato e valore atteso,

$$Y_i - \hat{\mu}_i,$$

non è direttamente confrontabile tra osservazioni, poiché la variabilità della risposta dipende dal livello della media.

Per questo motivo vengono utilizzati i residui di Pearson, definiti come

$$r_i^P = \frac{Y_i - \hat{\mu}_i}{\sqrt{\widehat{\text{Var}}(Y_i)}}.$$

Il numeratore rappresenta lo scostamento tra il conteggio osservato e quello previsto dal modello, mentre il denominatore lo rapporta alla variabilità attesa. I residui di Pearson esprimono quindi uno scostamento standardizzato e risultano maggiormente confrontabili tra osservazioni caratterizzate da valori attesi differenti (Wood, 2017). Nel presente lavoro, tali residui costituiscono il punto di partenza per la costruzione della matrice utilizzata nell'analisi di rete.

3.6 Dipendenza temporale e sbiancamento

I residui ottenuti dal modello additivo generalizzato rappresentano la componente dei conteggi non spiegata dagli effetti inclusi nel modello. La rimozione della componente sistematica non garantisce tuttavia che tali residui siano indipendenti nel tempo. Nel caso di osservazioni

raccolte in settimane consecutive, uno scostamento registrato in un determinato periodo può infatti essere associato agli scostamenti osservati nelle settimane precedenti.

La presenza di autocorrelazione residua risulta rilevante per l'analisi di rete, poiché la formulazione classica dei modelli grafici gaussiani considera osservazioni indipendenti del medesimo vettore multivariato (Meinshausen e Bühlmann, 2006). Nel presente contesto, tale assunzione riguarda le settimane, che costituiscono le osservazioni della matrice fornita al Graphical Lasso.

Per ridurre la dipendenza seriale interna alle singole serie comunali viene utilizzata una procedura di sbiancamento. Indicando con r_{it} il residuo relativo al comune i nella settimana t , la sua dinamica può essere rappresentata mediante un modello autoregressivo di ordine p :

$$r_{it} = \phi_{i1}r_{i,t-1} + \phi_{i2}r_{i,t-2} + \cdots + \phi_{ip}r_{i,t-p} + \varepsilon_{it},$$

dove $\phi_{i1}, \dots, \phi_{ip}$ sono i coefficienti autoregressivi ed ε_{it} rappresenta l'innovazione al tempo t .

L'innovazione può essere espressa come

$$\varepsilon_{it} = r_{it} - \sum_{k=1}^p \phi_{ik}r_{i,t-k}.$$

Essa rappresenta la parte del residuo corrente che non può essere prevista a partire dai valori precedenti della stessa serie. Le innovazioni vengono quindi utilizzate come residui sbiancati e sostituiscono i residui originali nella costruzione della matrice destinata all'analisi di rete.

L'obiettivo dello sbiancamento non è garantire in senso assoluto l'indipendenza delle osservazioni, ma ridurre la dipendenza seriale lineare, rendendo più plausibile l'assunzione di osservazioni temporalmente indipendenti. L'efficacia della procedura può essere valutata attraverso la funzione di autocorrelazione dei residui sbiancati e

mediante test congiunti, come il test di Ljung–Box, che verifica la presenza di autocorrelazioni residue fino a un determinato ritardo (Box *et al.*, 2015).

3.7 Trasformazione nonparanormal

Il modello grafico gaussiano assume che il vettore delle variabili considerate segua una distribuzione normale multivariata. I residui di Pearson, anche dopo la rimozione della dipendenza temporale, non sono tuttavia necessariamente distribuiti normalmente.

Per rendere meno restrittiva l’assunzione gaussiana è possibile ricorrere ai modelli a copula gaussiana. Questi estendono i modelli grafici gaussiani trasformando marginalmente le variabili mediante funzioni monotone lisce. Indicando con

$$\mathbf{X} = (X_1, \dots, X_p)^\top$$

il vettore delle variabili osservate, si assume che esistano funzioni monotone $f_1, \dots, f_p$ tali che

$$(f_1(X_1), \dots, f_p(X_p))^\top \sim \mathcal{N}_p(\mathbf{0}, \Sigma).$$

L’ipotesi di normalità non viene quindi formulata direttamente sulle variabili originarie, ma sul vettore ottenuto dopo aver trasformato separatamente ciascuna componente. L’obiettivo della trasformazione nonparanormal è stimare la matrice di covarianza del vettore gaussiano sottostante, così da migliorare la ricostruzione del grafo non orientato (Liu *et al.*, 2009; Zhao *et al.*, 2012).

La trasformazione viene stimata a partire dai ranghi delle osservazioni. Considerando n osservazioni della j -esima variabile,

$$x_{1j}, \dots, x_{nj},$$

si ordinano i valori e si indica con u_{ij} il rango dell'osservazione x_{ij} . Il rango viene quindi riscalato come

$$\hat{u}_{ij} = \frac{u_{ij}}{n+1}.$$

Una prima possibilità, denominata *normal score*, consiste nel trasformare ciascuna osservazione mediante

$$\hat{f}_j(x_{ij}) = \Phi^{-1}(\hat{u}_{ij}),$$

dove Φ^{-1} rappresenta la funzione quantile della distribuzione normale standard.

Poiché i ranghi più estremi possono produrre valori trasformati molto elevati in valore assoluto, è possibile utilizzare una versione troncata della trasformazione:

$$\hat{f}_j(x_{ij}) = \begin{cases} \Phi^{-1}(\delta), & \hat{u}_{ij} \leq \delta, \\ \Phi^{-1}(\hat{u}_{ij}), & \delta < \hat{u}_{ij} \leq 1 - \delta, \\ \Phi^{-1}(1 - \delta), & \hat{u}_{ij} > 1 - \delta, \end{cases}$$

dove

$$\delta = \frac{1}{4n^{1/4}\sqrt{\pi \log n}}.$$

Il troncamento limita quindi l'influenza delle osservazioni collocate nelle code estreme della distribuzione empirica. Nel pacchetto **huge**, le trasformazioni nonparanormal sono implementate mediante la funzione `huge.npn`, applicata separatamente alle colonne della matrice dei dati prima della stima della rete (Zhao *et al.*, 2012).

3.8 Descrizione e interpretazione della rete stimata

La struttura della rete stimata può essere rappresentata mediante una matrice di adiacenza, i cui elementi indicano la presenza o l'assenza di un arco tra ciascuna coppia di nodi.

L'intensità e il segno dell'associazione possono essere descritti mediante le correlazioni parziali, ricavate dalla matrice di precisione come

$$\rho_{ij|-ij} = -\frac{\Omega_{ij}}{\sqrt{\Omega_{ii}\Omega_{jj}}}.$$

Un valore positivo indica che i due comuni tendono a presentare scostamenti residui concordi, mentre un valore negativo segnala andamenti residui in direzioni opposte. Il valore assoluto esprime invece l'intensità dell'associazione condizionata (Lauritzen, 1996).

Una misura sintetica della posizione di ciascun nodo nella rete è il grado, definito come il numero di archi incidenti sul nodo stesso. I comuni con grado più elevato risultano collegati a un numero maggiore di altre unità, mentre un grado pari a zero identifica un comune isolato nella rete stimata.

3.9 Individuazione delle comunità nella rete

All'interno di una rete, una comunità è costituita da un insieme di nodi maggiormente collegati tra loro rispetto a quanto avviene con il resto del grafo. L'individuazione di tali gruppi consente di verificare se la struttura stimata presenti aggregazioni di comuni caratterizzate da associazioni residue simili.

Una misura utilizzata per valutare una partizione della rete è la modularità, che confronta il numero di collegamenti osservati all'interno delle comunità con quello atteso in una rete casuale avente una struttura dei gradi comparabile. Nel caso di una rete non pesata, essa

può essere espressa come

$$Q = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \mathbb{I}(c_i = c_j),$$

dove A_{ij} è l'elemento della matrice di adiacenza, k_i e k_j sono i gradi dei nodi, m è il numero complessivo di archi e $\mathbb{I}(c_i = c_j)$ indica che i nodi i e j appartengono alla stessa comunità.

Per individuare la partizione viene utilizzato l'algoritmo di Louvain, una procedura iterativa che ricerca una configurazione caratterizzata da elevata modularità. In una prima fase, ciascun nodo viene spostato nella comunità che produce il maggiore incremento della modularità.

Successivamente, le comunità ottenute vengono aggregate in nuovi nodi e la procedura viene ripetuta finché non si osservano ulteriori miglioramenti (Blondel *et al.*, 2008).

I gruppi così individuati rappresentano quindi comunità definite esclusivamente sulla base della struttura delle connessioni stimate, senza imporre a priori una classificazione geografica o amministrativa.

Capitolo 4

Modellazione dei conteggi e analisi di rete

4.1 Obiettivo

In questo capitolo viene applicata agli accessi settimanali al Fascicolo Sanitario Elettronico la procedura metodologica descritta nel capitolo precedente. In una prima fase viene modellata la componente sistematica dei conteggi, tenendo conto dell'andamento temporale, delle caratteristiche osservate dei comuni e dell'eterogeneità persistente tra territori.

I residui ottenuti vengono successivamente utilizzati per stimare, mediante Graphical Lasso, una rete di associazioni condizionate tra le serie comunali. La struttura risultante viene infine analizzata e confrontata con le principali suddivisioni territoriali della provincia.

4.2 Dati utilizzati

I dati utilizzati comprendono 158 comuni osservati per 92 settimane, per un totale di 14536 osservazioni. Il periodo considerato va dalla settimana del 5 agosto 2024 alla settimana del 4 maggio 2026, mentre il numero complessivo di accessi registrati è pari a 1039746.

La variabile risposta è il numero settimanale di accessi al FSE osservato nel comune i durante la settimana t , indicato con Y_{it} . Nella specificazione iniziale vengono considerate quattro caratteristiche comunali: l'altitudine media, la superficie comunale, la quota femminile e l'età media della popolazione nel periodo d'interesse. Per tenere conto della diversa dimensione demografica dei comuni, viene incluso nel modello un offset pari al logaritmo della popolazione residente. La componente temporale viene invece rappresentata mediante una funzione liscia del tempo.

4.3 Modelli GAM per la componente sistematica

La componente sistematica degli accessi viene modellata mediante modelli additivi generalizzati. La costruzione del modello avviene progressivamente, valutando la distribuzione della risposta, il contributo delle covariate e la presenza di eterogeneità persistente tra comuni.

4.3.1 Modello di Poisson

La prima specificazione assume per la risposta una distribuzione di Poisson:

$$Y_{it} \sim \text{Poisson}(\mu_{it}),$$

con predittore

$$\begin{aligned} \log(\mu_{it}) = & \log(P_{it}) + \beta_0 + \beta_1 \text{altitudine}_i + \beta_2 \text{età media}_i \\ & + \beta_3 \text{quota femminile}_i + \beta_4 \text{superficie}_i + f(t). \end{aligned}$$

In questa formulazione, P_{it} indica la popolazione residente nel comune i durante la settimana t , mentre $f(t)$ rappresenta la funzione liscia del tempo.

L'indice di dispersione calcolato sui residui di Pearson è pari a 9.68, valore molto superiore a uno. La variabilità osservata risulta quindi maggiore di quella prevista dal modello di Poisson, rendendo necessario considerare una distribuzione in grado di rappresentare la sovradisersione.

4.3.2 Modello binomiale negativo

Il modello viene quindi ristimato assumendo per la risposta una distribuzione binomiale negativa e mantenendo invariato il predittore:

$$Y_{it} \sim \text{NegBin}(\mu_{it}, \theta).$$

Il passaggio alla binomiale negativa determina un netto miglioramento dell'adattamento. L'indice di dispersione dei residui di Pearson scende da 9.68 a 1.13, risultando molto più vicino a uno. Anche l'AIC diminuisce considerevolmente, passando da 207576 a 111641. La distribuzione binomiale negativa viene pertanto adottata nelle specifiche successive.

Nel modello tutte le covariate considerate risultano statisticamente significative e presentano coefficienti negativi.

Covariata	$\hat{\beta}$	Errore standard	p-value	$\exp(\hat{\beta})$
Altitudine media	-0.00011	0.000008	< 0.001	0.99989
Età media della popolazione	-0.00583	0.00181	0.001	0.99419
Quota femminile	-0.73112	0.230751	0.002	0.48137
Superficie comunale	-0.00058	0.00009	< 0.001	0.99942

Tabella 4.1: Coefficienti parametrici del modello binomiale negativo senza effetto casuale comunale.

Poiché il modello utilizza una funzione di collegamento logaritmica, la variazione percentuale associata a un incremento Δx di una covariata è calcolata come

$$\left[\exp(\hat{\beta}\Delta x) - 1 \right] \times 100.$$

A parità delle altre componenti del modello, un aumento di 100 metri dell'altitudine media è associato a una riduzione del numero atteso di accessi pari a circa l'1.1%. Un incremento di un anno dell'età media della popolazione corrisponde a una riduzione attesa di circa lo 0.6%, mentre un aumento di un punto percentuale della quota femminile è associato a una diminuzione di circa lo 0.7%. Infine, un incremento di 10 chilometri quadrati della superficie comunale corrisponde a una riduzione attesa di circa lo 0.6%.

Le associazioni stimate risultano quindi di entità contenuta, ma indicano che le caratteristiche demografiche e territoriali considerate spiegano una parte delle differenze osservate tra i comuni. Rimane tuttavia possibile la presenza di eterogeneità persistente non adeguatamente rappresentata dalle sole covariate disponibili.

4.3.3 Introduzione dell'effetto casuale comunale

Per tenere conto delle differenze stabili tra comuni non completamente spiegate dalle covariate osservate, viene introdotto un effetto casuale comunale:

$$\log(\mu_{it}) = \log(P_{it}) + \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + f(t) + u_i,$$

con

$$u_i \sim \mathcal{N}(0, \sigma_u^2).$$

Il termine u_i rappresenta una deviazione persistente propria del comune i , non spiegata dalla popolazione residente, dalle covariate osservate e dall'andamento temporale comune.

L'introduzione dell'effetto casuale determina un ulteriore miglioramento: l'AIC diminuisce da 111641 a 110478, mentre la devianza spiegata aumenta dal 69.3% al 72.4%.

Dopo l’inserimento dell’effetto casuale comunale, l’età media della popolazione e la quota femminile non risultano più statisticamente significative. Tali covariate presentano infatti una variabilità tra i comuni e l’intercetta casuale comunale assorbe parte delle differenze sistematiche nei livelli medi di accesso tra territori, riducendo la possibilità di identificarne separatamente il contributo. Poiché l’età media della popolazione e la quota femminile non risultano più statisticamente significative dopo l’introduzione dell’effetto casuale comunale, il modello viene ristimato escludendo progressivamente tali covariate. Il modello finale assume una distribuzione binomiale negativa:

$$Y_{it} \sim \text{NegBin}(\mu_{it}, \theta),$$

con predittore

$$\begin{aligned} \log(\mu_{it}) = & \log(P_{it}) + \beta_0 + \beta_1 \text{altitudine}_i + \beta_2 \text{superficie}_i \\ & + f(t) + u_i. \end{aligned}$$

Si sottolinea che le associazioni osservate per l’altitudine e la superficie non devono essere interpretate come effetti diretti sul numero di accessi. Tali variabili possono infatti rappresentare anche caratteristiche territoriali non direttamente osservate, come la dispersione degli abitati, l’accessibilità ai servizi o il grado di ruralità del comune.

Componente temporale

La funzione liscia del tempo assoluto risulta fortemente rilevante e presenta 87.6 gradi di libertà effettivi. Poiché l’obiettivo del modello è rimuovere in modo accurato le variazioni temporali condivise dai comuni, viene utilizzata una base spline flessibile con $k = 90$, valore prossimo al numero complessivo di settimane osservate.

La flessibilità effettivamente utilizzata viene regolata dalla

penalizzazione interna al GAM, con parametro di smoothing stimato mediante il criterio REML. La curva risultante descrive un andamento temporale articolato, caratterizzato da variazioni rapide e picchi localizzati.

Poiché, in questa fase, il modello non viene utilizzato per estrapolare l'andamento degli accessi al di fuori del periodo osservato, ma per rimuovere le variazioni temporali comuni che, se lasciate nei residui, potrebbero generare associazioni diffuse tra le serie e influenzare la rete successivamente stimata, l'elevata flessibilità della componente temporale risulta giustificata.

4.3.4 Diagnostica del modello finale

Prima di utilizzare i residui nelle fasi successive, viene valutata l'adeguatezza del modello finale. La diagnostica considera i residui di Pearson rispetto al predittore lineare, il relativo QQ-plot e la loro distribuzione, riportati nella Figura 4.1.

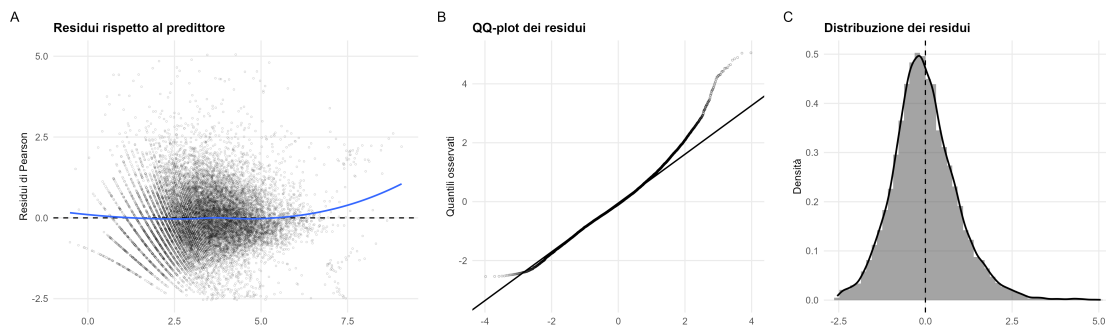


Figura 4.1: Diagnostica del modello GAM finale.

I residui risultano complessivamente concentrati intorno allo zero. Nel grafico rispetto al predittore lineare emerge tuttavia una lieve struttura in corrispondenza dei valori più elevati, dove la curva tende ad aumentare. Il modello rappresenta quindi adeguatamente la parte centrale delle osservazioni, mostrando una minore capacità di descrivere alcuni livelli particolarmente elevati degli accessi.

Il QQ-plot evidenzia deviazioni dalla normalità soprattutto nella coda destra, mentre la distribuzione dei residui risulta leggermente asimmetrica e caratterizzata da alcuni valori positivi elevati. La normalità dei residui non costituisce un'assunzione del modello binomiale negativo; tali deviazioni sono tuttavia rilevanti per la successiva applicazione del Graphical Lasso e motivano l'impiego della trasformazione nonparanormal.

Nel complesso, la diagnostica indica che il modello rappresenta adeguatamente la componente media e la dispersione dei conteggi. Rimane da valutare la presenza di dipendenza temporale residua all'interno delle singole serie comunali.

4.3.5 Autocorrelazione dei residui e procedura di sbiancamento

La presenza di autocorrelazione nelle serie residue comunali viene valutata mediante il test di Ljung–Box, applicato separatamente a ciascun comune considerando i primi quattro ritardi temporali. Sui residui di Pearson originari, il test risulta significativo per 135 comuni su 158, pari all'85.4% del totale. La modellazione della componente media non risulta quindi sufficiente a rimuovere la dipendenza seriale presente nelle serie comunali.

Viene inizialmente applicato a ciascuna serie comunale un modello autoregressivo di ordine uno. Dopo lo sbiancamento AR(1), il numero di comuni con test significativo scende a 58, pari al 36.7%. La struttura AR(1) riduce quindi l'autocorrelazione, ma lascia una dipendenza residua rilevante in una quota consistente delle serie.

In un secondo passaggio viene applicato un modello autoregressivo di ordine due. Dopo lo sbiancamento AR(2), soltanto 4 comuni su 158, pari al 2.5%, presentano ancora un test significativo.

Fase	Test significativi	Percentuale significativa
Residui di Pearson	135	85.4%
Dopo AR(1)	58	36.7%
Dopo AR(2)	4	2.5%

Tabella 4.2: Risultati del test di Ljung–Box prima e dopo lo sbiancamento autoregressivo.

Il modello AR(2) risulta quindi più efficace nel ridurre la dipendenza seriale. I residui ottenuti dopo tale procedura, approssimativamente privi di autocorrelazione lineare per la quasi totalità dei comuni, vengono utilizzati nelle fasi successive.

4.3.6 Trasformazione nonparanormal e matrice finale

Ai residui sbiancati viene applicata la trasformazione nonparanormal troncata mediante la funzione `huge.npn()`. Le colonne della matrice trasformata vengono successivamente centrate e riscalate.

La matrice finale ha dimensione 92×158 , con righe corrispondenti alle settimane e colonne associate ai comuni. Dopo la standardizzazione, ciascuna colonna presenta media nulla e deviazione standard unitaria. I valori ottenuti rappresentano quindi scostamenti residui rispetto al comportamento atteso dei comuni, depurati dalla dipendenza temporale interna alle singole serie e ricondotti a una scala maggiormente compatibile con il quadro dei modelli grafici gaussiani.

4.4 Stima della rete tramite Graphical Lasso

Il Graphical Lasso viene applicato alla matrice dei residui sbiancati, trasformati e standardizzati. Il percorso di regolarizzazione viene stimato considerando 40 valori del parametro λ , compresi tra 0.214 e 0.534.

La soluzione finale viene selezionata mediante il criterio StARS, utilizzando una soglia di stabilità pari a 0.05. Il valore selezionato è

$$\lambda = 0.358.$$

La Figura 4.2 mostra l'andamento dell'instabilità lungo il percorso di regolarizzazione.

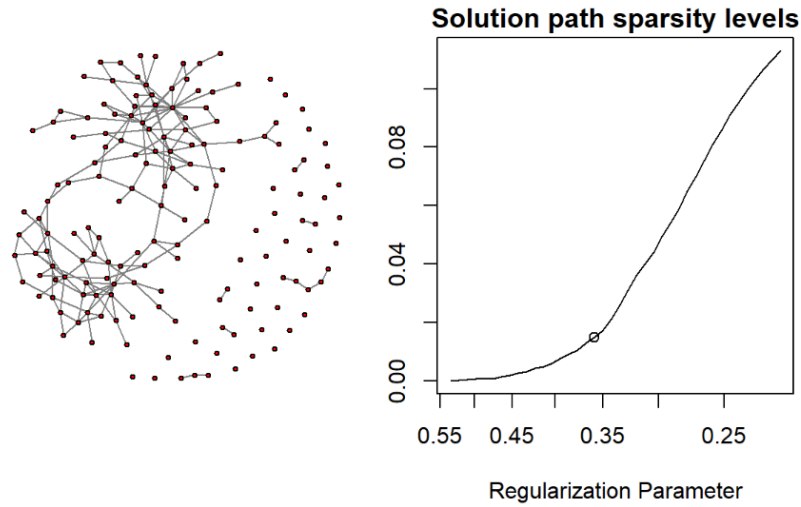


Figura 4.2: Selezione del parametro di regolarizzazione mediante il criterio StARS.

La rete selezionata contiene 185 archi su 12403 possibili coppie di comuni, con una densità pari a

$$\frac{185}{12403} = 0.0149.$$

Viene quindi selezionato circa l'1.5% delle connessioni possibili, evidenziando una struttura complessivamente molto sparsa. Tale sparsità dipende anche dalla soglia di instabilità StARS adottata. Il valore 0.05 rappresenta infatti una scelta relativamente restrittiva, effettuata per limitare la rete alle connessioni che risultano maggiormente stabili rispetto alle variazioni del campione. Una soglia

più elevata avrebbe prodotto una rete più densa, ma avrebbe comportato l'inclusione di associazioni meno stabili.

Indicatore	Valore
Numero di comuni	158
Numero di archi	185
Archi possibili	12403
Densità	0.0149

Tabella 4.3: Sintesi della rete stimata tramite Graphical Lasso.

4.5 Correlazioni parziali e misure di rete

Tra i 185 archi selezionati, 172 presentano una correlazione parziale positiva e 13 una correlazione parziale negativa. La netta prevalenza di archi positivi indica che, tra i comuni collegati, gli scostamenti residui tendono più frequentemente a muoversi nella stessa direzione.

Le associazioni presentano comunque intensità contenuta. La correlazione parziale minima è pari a -0.073 , quella massima a 0.127 , mentre la mediana del valore assoluto è pari a 0.027 .

Indicatore	Valore
Numero di archi	185
Archi positivi	172
Archi negativi	13
Correlazione parziale minima	-0.073
Correlazione parziale massima	0.127
Mediana di $ \rho $	0.027

Tabella 4.4: Sintesi delle correlazioni parziali associate agli archi della rete.

Dal punto di vista topologico, la rete presenta 39 componenti connesse e 32 comuni isolati. Il grado medio è pari a circa 2.34, mentre il grado massimo è pari a 15.

La rete non forma quindi un'unica struttura compatta, ma risulta frammentata in sottogruppi separati. Anche i comuni maggiormente connessi presentano un numero di collegamenti contenuto rispetto al

numero totale di nodi. La struttura stimata seleziona pertanto un insieme limitato di associazioni condizionate, anziché una dipendenza diffusa tra tutti i territori.

4.6 Coerenza territoriale degli archi

Per valutare la coerenza territoriale della rete viene calcolata la proporzione di archi che collegano comuni appartenenti alla stessa Comunità di Valle, allo stesso Distretto 2022, allo stesso Ambito di integrazione o allo stesso Ambito territoriale.

Le proporzioni osservate di archi interni alle diverse partizioni territoriali sono riportate nella Tabella 4.5. Il valore più elevato si osserva per il Distretto 2022, mentre la proporzione più contenuta riguarda le Comunità di Valle.

Partizione territoriale	Proporzione osservata di archi interni
Comunità di Valle	0.292
Distretto 2022	0.454
Ambito di integrazione	0.405
Ambito territoriale	0.405

Tabella 4.5: Proporzione di archi che collegano comuni appartenenti alla stessa partizione territoriale.

Le proporzioni osservate non sono tuttavia direttamente confrontabili tra le diverse partizioni territoriali, poiché dipendono anche dal numero e dalla dimensione dei gruppi che le compongono. Una suddivisione formata da pochi gruppi di grandi dimensioni, come quella dei distretti, tende infatti a presentare una quota più elevata di archi interni anche in assenza di una specifica struttura territoriale. Al contrario, in una partizione più frammentata, come quella delle Comunità di Valle, la probabilità che due comuni collegati appartengano casualmente allo stesso gruppo risulta inferiore.

Per valutare se le proporzioni osservate siano superiori a quelle attese per effetto della sola struttura delle partizioni, vengono generate 1000 reti casuali con lo stesso numero di nodi e di archi della rete osservata, mantenendo invariate le assegnazioni territoriali dei comuni. Per tutte le partizioni considerate, la proporzione osservata di archi interni supera il quantile 97.5% della corrispondente distribuzione simulata.

Partizione	Osservato	Mediana simulata	Quantile 97.5%
Distretto 2022	0.454	0.330	0.400
Ambito di integrazione	0.405	0.211	0.270
Ambito territoriale	0.405	0.200	0.259
Comunità di Valle	0.292	0.087	0.130

Tabella 4.6: Confronto tra proporzioni osservate di archi interni e distribuzioni simulate su reti casuali con lo stesso numero di nodi e archi.

Nonostante la bassa densità e la debole intensità delle correlazioni parziali, gli archi non risultano quindi distribuiti casualmente rispetto alle suddivisioni considerate. Le connessioni residue tendono a concentrarsi all'interno delle partizioni territoriali più frequentemente di quanto atteso nelle reti simulate.

4.7 Clusterizzazione Louvain

L'algoritmo Louvain viene applicato al sottografo formato dai 126 comuni con almeno un collegamento. I 32 comuni isolati vengono esclusi dalla clusterizzazione, poiché non dispongono di archi utili alla definizione di una comunità di rete.

Come peso degli archi viene utilizzato il valore assoluto delle correlazioni parziali. La formazione dei gruppi dipende quindi dall'intensità delle associazioni residue, indipendentemente dal loro segno.

L'algoritmo individua 19 cluster tra i comuni connessi, con una modularità pari a 0.757.

Indicatore	Valore
Comuni totali	158
Comuni connessi	126
Comuni isolati	32
Cluster Louvain tra comuni connessi	19
Modularità	0.757

Tabella 4.7: Sintesi della clusterizzazione Louvain applicata alla rete stimata.

I cluster presentano dimensioni variabili. I quattro gruppi più grandi comprendono rispettivamente 18, 16, 16 e 13 comuni, mentre sono presenti diversi gruppi più piccoli, formati da due, tre o quattro unità. La partizione risulta quindi articolata in numerose comunità di dimensione contenuta, coerentemente con la natura sparsa e frammentata della rete.

La distribuzione geografica dei sei cluster di maggiore dimensione è riportata nella Figura 4.3. In ciascun pannello vengono evidenziati i comuni appartenenti a uno specifico cluster, mentre gli altri territori sono rappresentati in grigio.



Figura 4.3: Localizzazione geografica dei principali cluster Louvain.

La rappresentazione cartografica evidenzia una parziale struttura territoriale. Alcuni cluster comprendono infatti gruppi di comuni geograficamente vicini o contigui. Altri risultano invece composti da più nuclei distinti oppure includono comuni collocati in aree lontane della provincia.

La prossimità geografica contribuisce quindi alla formazione di alcune comunità, ma non ne spiega interamente la composizione. La rete individua anche associazioni tra territori distanti caratterizzati da dinamiche residue simili.

4.8 Conclusioni

La modellazione dei conteggi ha evidenziato una marcata sovradisersione, rendendo più adeguata la distribuzione binomiale negativa rispetto a quella di Poisson. L'introduzione dell'effetto casuale comunale ha inoltre permesso di rappresentare una parte dell'eterogeneità persistente tra territori non spiegata dalle caratteristiche osservate.

Nel modello finale, l'altitudine media e la superficie comunale risultano statisticamente significative e presentano entrambe un'associazione negativa con il numero atteso di accessi. A parità delle altre componenti del modello, i comuni caratterizzati da un'altitudine media o da una superficie maggiore tendono quindi a registrare un numero inferiore di accessi rispetto alla popolazione residente.

Tali associazioni non devono tuttavia essere interpretate come effetti causali diretti. L'altitudine e la superficie possono infatti rappresentare caratteristiche territoriali più ampie, come la perifericità, l'accessibilità ai servizi, la dispersione degli abitati o il grado di ruralità del comune. Dopo la rimozione della componente sistematica, lo sbiancamento AR(2) ha ridotto l'autocorrelazione presente nelle singole serie residue. I

residui sbiancati, trasformati e standardizzati sono stati quindi utilizzati per stimare la rete mediante Graphical Lasso.

La rete risultante è molto sparsa, frammentata in diverse componenti e caratterizzata da correlazioni parziali generalmente deboli. Le connessioni mostrano tuttavia una concentrazione territoriale superiore a quella osservata nelle reti simulate.

La clusterizzazione Louvain individua comunità solo parzialmente sovrapponibili alle partizioni amministrative considerate. La struttura della rete sembra quindi riflettere sia relazioni locali tra comuni vicini sia associazioni tra territori distanti accomunati da dinamiche residue simili. I raggruppamenti ottenuti verranno utilizzati nella successiva analisi previsiva, per valutare se la condivisione di dinamiche residue possa contribuire a migliorare la previsione degli accessi.

Capitolo 5

Confronto previsivo dei modelli GAM

5.1 Obiettivo dell'analisi

L'ultima parte dell'analisi è dedicata al confronto previsivo tra modelli GAM caratterizzati da una diversa rappresentazione della componente stagionale. L'obiettivo è verificare se la struttura individuata mediante Graphical Lasso possa fornire un'informazione utile anche in termini previsivi, oltre che descrittivi.

Sono state considerate tre famiglie di modelli:

1. un GAM base, nel quale tutti i comuni condividono la stessa curva stagionale;
2. un GAM con curva stagionale globale e deviazioni specifiche per Comunità di Valle;
3. un GAM con curva stagionale globale e deviazioni specifiche per i cluster ottenuti dal Graphical Lasso.

La maggiore complessità introdotta dalla presenza di numerosi gruppi è controllata mediante factor-smooth penalizzate, che consentono deviazioni dalla curva comune solo quando supportate dai dati. Il meccanismo si basa su una penalizzazione che contrae le curve di gruppo verso un effetto medio condiviso, con intensità determinata durante la

stima attraverso i parametri di smoothing. In presenza di scarsa informazione nei dati, la penalizzazione prevale e le curve di gruppo rimangono prossime a quella globale; al contrario, quando emergono pattern sistematici ben supportati dai dati, il modello consente deviazioni più marcate. Questa formulazione risulta particolarmente adatta in contesti con numerosi gruppi e dimensioni campionarie eterogenee, poiché bilancia flessibilità del modello e rischio di overfitting.

5.2 Stima, validazione temporale e verifica

Il pannello comprende 158 comuni osservati per 92 settimane. Le prime 74 settimane costituiscono il periodo di sviluppo dei modelli, mentre le ultime 18 settimane, corrispondenti alle settimane 75–92, vengono riservate alla verifica finale. Tale suddivisione consente di mantenere il periodo di verifica completamente separato dalle fasi di selezione e confronto delle configurazioni.

All'interno del periodo di sviluppo, la selezione dei modelli viene effettuata mediante una procedura di validazione temporale a finestra mobile di ampiezza fissa. A differenza di una suddivisione casuale dei dati, questa procedura mantiene invariato l'ordine cronologico delle osservazioni: in ogni partizione, il modello viene stimato utilizzando esclusivamente settimane precedenti rispetto a quelle impiegate per la valutazione (Tashman, 2000).

In ciascuna partizione, la finestra di stima comprende 52 settimane consecutive, mentre le 4 settimane immediatamente successive costituiscono il periodo di validazione. Dopo ogni iterazione, sia l'inizio sia la fine della finestra di stima vengono avanzati di 4 settimane. La sua ampiezza rimane quindi costante, mentre cambia il periodo temporale sul quale il modello viene stimato. Anche la finestra di validazione avanza di 4 settimane e non si sovrappone a quella della

partizione precedente.

La procedura produce cinque valutazioni successive, riportate nella Tabella 5.1. In questo modo, ciascuna configurazione viene valutata più volte su periodi futuri distinti, riducendo la dipendenza della scelta da un'unica suddivisione temporale.

Tabella 5.1: Finestre utilizzate nella procedura di validazione temporale a finestra mobile.

Fold	Periodo di stima	Periodo di validazione
1	settimane 1–52	settimane 53–56
2	settimane 5–56	settimane 57–60
3	settimane 9–60	settimane 61–64
4	settimane 13–64	settimane 65–68
5	settimane 17–68	settimane 69–72

Le prestazioni ottenute nelle cinque partizioni vengono aggregate e utilizzate per selezionare le configurazioni da sottoporre alla verifica conclusiva. Una volta completata la selezione, i modelli scelti vengono ristimati utilizzando l'intero periodo di sviluppo, costituito dalle settimane 1–74.

Le settimane 75–92 rimangono escluse da tutte le decisioni relative alla scelta della struttura del modello e dei relativi parametri. Esse vengono utilizzate una sola volta, al termine della procedura, per confrontare le prestazioni previsive finali dei modelli selezionati.

5.3 Costruzione preliminare delle reti e dei cluster

Prima di avviare la procedura di validazione temporale, le reti e i relativi cluster comunali vengono costruiti utilizzando l'intero periodo di stima, costituito dalle prime 74 settimane. Le ultime 18 settimane, riservate alla verifica finale, rimangono invece escluse sia dalla stima della rete sia dalla selezione delle configurazioni previsive.

La costruzione della rete segue la procedura descritta nel Capitolo 4. Sull'intero periodo di stima viene inizialmente stimato un GAM

binomiale negativo con offset della popolazione, funzione liscia del tempo, effetto casuale comunale e covariate territoriali e demografiche candidate. Anche in questo caso, il modello finale selezionato mantiene l'altitudine media e la superficie comunale, mentre l'età media della popolazione e la quota femminile non vengono incluse. Il modello viene quindi utilizzato per ricavare i residui di Pearson.

La presenza di dipendenza seriale viene successivamente valutata mediante il test di Ljung–Box. Il confronto tra lo sbiancamento autoregressivo di ordine uno e quello di ordine due mostra che la struttura AR(1) lascia, anche in questo caso, una quota ancora rilevante di serie autocorrelate, pari al 27% dei comuni, mentre l'AR(2) riduce ulteriormente l'autocorrelazione residua, portando tale quota al 9%. Per questo motivo, nelle analisi successive vengono utilizzati i residui sbiancati mediante un modello AR(2).

La matrice dei residui sbiancati viene quindi sottoposta alla trasformazione nonparanormal, centrata e riscalata. Su tale matrice viene stimato il Graphical Lasso e le comunità della rete vengono individuate mediante l'algoritmo di Louvain. Sono considerate tre soglie del criterio StARS:

$$0.05, \quad 0.10, \quad 0.15.$$

Per ciascuna soglia viene ottenuta una diversa rete e, di conseguenza, una diversa partizione dei comuni in cluster. Le principali caratteristiche delle tre soluzioni sono riportate nella Tabella 5.2.

Tabella 5.2: Caratteristiche delle reti e delle partizioni Louvain ottenute nel periodo di sviluppo.

Soglia StARS	λ	Archi	Densità	Componenti	Isolati	Cluster connessi	Cluster totali
0.05	0.399	105	0.0085	59	46	25	71
0.10	0.342	353	0.0285	13	11	14	25
0.15	0.293	798	0.0643	1	0	11	11

All'aumentare della soglia StARS, la rete diventa progressivamente più densa e meno frammentata. La soluzione con soglia 0.05 presenta molti comuni isolati e un numero elevato di gruppi, mentre con soglia 0.15 tutti i comuni appartengono a un'unica componente connessa. La soglia 0.10 rappresenta una soluzione intermedia. La scelta definitiva non viene tuttavia effettuata sulla base delle sole caratteristiche della rete, ma mediante la successiva validazione temporale delle prestazioni previsive. Le partizioni così costruite vengono mantenute fisse durante la validazione temporale. In ciascuna finestra temporale vengono ristimati e valutati i modelli previsivi, ma non vengono ripetute la stima del GAM residuale, la ricostruzione della rete e la clusterizzazione Louvain. La procedura rolling valuta pertanto le prestazioni previsive dei modelli condizionatamente alle strutture di rete definite preliminarmente sull'intero periodo di sviluppo.

5.4 Specificazione dei modelli previsivi

Tutti i modelli previsivi assumono una distribuzione binomiale negativa con funzione di collegamento logaritmica e includono:

- l'offset pari al logaritmo della popolazione comunale;
- l'altitudine media e la superficie comunale;
- un effetto casuale specifico per comune;
- una struttura di autocorrelazione residua $AR(1)$ all'interno di ciascun comune.

Nella fase previsiva non viene ripetuta la selezione dell'intero insieme di covariate demografiche e territoriali. Vengono mantenute l'altitudine media e la superficie comunale, al fine di garantire una specificazione parsimoniosa e uniforme tra le diverse famiglie di modelli confrontate.

Il modello base è definito da

$$\log(\mu_{it}) = \log(P_i) + f(\text{settimana ISO}_t) + \beta_1 \text{Altitudine}_i + \beta_2 \text{Superficie}_i + u_i,$$

dove settimana ISO_t indica la posizione della settimana t all'interno dell'anno, con valori compresi tra 1 e 52, o 53 negli anni che la prevedono, mentre $f(\cdot)$ è una spline ciclica comune a tutti i comuni. La spline ciclica viene utilizzata per rappresentare la stagionalità annuale, imponendo continuità tra la fine e l'inizio dell'anno: il valore stimato per le ultime settimane deve quindi raccordarsi con quello delle prime settimane dell'anno successivo. In questo modo, per prevedere periodi futuri il modello riutilizza la struttura stagionale già appresa per le corrispondenti settimane dell'anno.

Una spline costruita sull'indice temporale assoluto, come il numero progressivo della settimana, dovrebbe invece essere prolungata oltre l'intervallo osservato. Tale estrapolazione può risultare instabile e dipendere eccessivamente dall'andamento stimato nella parte finale della serie.

Nei modelli con struttura di gruppo, alla curva globale viene aggiunta una deviazione factor-smooth:

$$\begin{aligned} \log(\mu_{it}) = & \log(P_i) + f(\text{settimana ISO}_t) + g_{G_i}(\text{settimana ISO}_t) \\ & + \beta_1 \text{Altitudine}_i + \beta_2 \text{Superficie}_i + u_i. \end{aligned}$$

dove G_i identifica alternativamente la Comunità di Valle o il cluster Graphical Lasso del comune i . Le funzioni $g_{G_i}(\cdot)$ rappresentano deviazioni penalizzate rispetto alla curva globale.

La dimensione della base della spline globale è stata selezionata tra

$$k \in \{10, 20, 30, 40\},$$

mentre per le deviazioni factor-smooth sono stati confrontati

$$k_{\text{fs}} \in \{4, 6, 8\}.$$

Per il modello basato sui cluster, la selezione ha riguardato congiuntamente k , k_{fs} e la soglia StARS.

5.5 Criteri di valutazione

Le prestazioni previsive sono state confrontate attraverso quattro misure:

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|,$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2},$$

$$\text{Bias} = \frac{1}{n} \sum_{j=1}^n (\hat{y}_j - y_j),$$

e

$$\text{RMSLE} = \sqrt{\frac{1}{n} \sum_{j=1}^n [\log(1 + \hat{y}_j) - \log(1 + y_j)]^2}.$$

La configurazione migliore all'interno di ciascuna famiglia è stata scelta minimizzando principalmente l'RMSE aggregato sulle previsioni fuori campione prodotte dalle cinque finestre temporali. Tale criterio attribuisce un peso maggiore agli errori di entità più elevata e consente quindi di penalizzare le configurazioni che, pur mostrando un errore medio contenuto, producono previsioni particolarmente inaccurate in alcune settimane o per alcuni comuni.

Le altre misure vengono comunque considerate come indicatori complementari per valutare rispettivamente l'errore medio assoluto, la

tendenza alla sovrastima o alla sottostima e la qualità relativa delle previsioni su scale differenti.

5.6 Selezione mediante rolling window

La procedura rolling ha selezionato per tutte le famiglie il valore $k = 10$. Nei modelli con deviazioni factor-smooth è stato inoltre scelto $k_{fs} = 4$, cioè il valore più parsimonioso tra quelli considerati. Per il modello basato sui cluster è stata selezionata la soglia StARS pari a 0.10. Le configurazioni selezionate e le relative prestazioni rolling sono riportate nella Tabella 5.3.

Tabella 5.3: Configurazioni selezionate mediante rolling window.

Modello	StARS	k	k_{fs}	MAE	RMSE	RMSLE
Cluster Graphical Lasso	0.10	10	4	28.4	135	0.746
GAM base	–	10	–	31.6	140	0.775
Comunità di Valle	–	10	4	30.2	154	0.723

Il modello con cluster Graphical Lasso ottiene il valore più basso per tutte le misure. Il vantaggio rispetto al GAM base è già visibile nel periodo di selezione e non emerge quindi soltanto nella verifica finale. La procedura sceglie inoltre una complessità contenuta sia per la curva globale sia per le deviazioni di gruppo. Questo risultato suggerisce che l'informazione utile dei cluster non dipende da una flessibilità eccessiva, ma da correzioni relativamente semplici rispetto alla stagionalità comune.

Il modello per Comunità di Valle presenta un MAE inferiore a quello del GAM base, ma un RMSE più elevato. Ciò indica che l'introduzione di deviazioni territoriali può migliorare l'errore medio, ma non riduce altrettanto efficacemente gli errori di maggiore entità.

5.7 Verifica finale dei modelli

Una volta selezionata la configurazione migliore di ciascuna famiglia, i modelli sono stati ristimati sull'intero periodo di sviluppo, costituito dalle prime 74 settimane.

I parametri di dispersione della distribuzione binomiale negativa e il coefficiente di autocorrelazione AR(1) sono stati stimati globalmente sul set di sviluppo e mantenuti costanti per tutte le specificazioni considerate, al fine di garantire la confrontabilità tra i modelli e concentrare la fase di selezione sulla struttura della componente media. La stima finale del parametro di dispersione della binomiale negativa è risultata pari a

$$\hat{\theta} = 6.937,$$

mentre il coefficiente autoregressivo utilizzato nei modelli previsivi è

$$\hat{\rho} = 0.252.$$

Le prestazioni ottenute sulle 18 settimane di verifica sono riportate nella Tabella 5.4.

Tabella 5.4: Prestazioni finali sul periodo di verifica.

Modello	StARS	k	k_{fs}	MAE	RMSE	Bias	RMSLE
Cluster Graphical Lasso	0.10	10	4	29.3	228	-26.4	0.325
Comunità di Valle	–	10	4	32.3	252	-29.6	0.326
GAM base	–	10	–	34.4	263	-26.6	0.381

Il modello basato sui cluster Graphical Lasso conferma le prestazioni migliori nel periodo di verifica. Rispetto al GAM base, il MAE si riduce da 34.4 a 29.3, con un miglioramento relativo pari a circa il 14.8%.

L'RMSE passa da 263 a 228, corrispondente a una riduzione di circa il 13.3%. Anche l'RMSLE diminuisce da 0.381 a 0.325, con un miglioramento di circa il 14.7%.

Il modello per Comunità di Valle migliora rispetto al GAM base, ma rimane meno accurato del modello basato sulla rete. In particolare, il suo RMSE è pari a 252, mentre quello del modello Graphical Lasso è pari a 228. Questo confronto mostra che il miglioramento non deriva soltanto dall'introduzione di curve specifiche per gruppi, ma anche dalla capacità dei cluster della rete di individuare insiemi di comuni con dinamiche temporali più omogenee dal punto di vista previsivo. Tutti i modelli presentano un bias negativo, indicando una tendenza generale alla sottostima degli accessi nel periodo finale. Il bias del modello Graphical Lasso è tuttavia molto vicino a quello del GAM base. La riduzione dell'errore non sembra quindi derivare da una semplice correzione del livello medio delle previsioni, ma da una migliore rappresentazione delle differenze nella dinamica stagionale tra gruppi di comuni.

5.8 Interpretazione dei risultati

I risultati ottenuti supportano l'utilità previsiva della struttura derivata dal Graphical Lasso. Il modello di rete supera il GAM base sia durante la selezione rolling sia sul periodo di verifica indipendente. Tale coerenza è importante, poiché riduce il rischio che il vantaggio osservato sia dovuto a caratteristiche specifiche delle sole settimane finali.

La selezione di $k_{fs} = 4$ per entrambe le famiglie con deviazioni di gruppo suggerisce che le differenze tra gruppi possono essere rappresentate con funzioni relativamente semplici. La curva globale descrive la parte principale della stagionalità annuale, mentre le factor-smooth apportano correzioni limitate e regolarizzate. In questo modo si evita di stimare una curva completamente indipendente per ciascun gruppo, soluzione che avrebbe prodotto modelli più instabili e maggiormente esposti al rischio di sovradattamento.

Il confronto con le Comunità di Valle fornisce inoltre un riferimento utile. Anche la partizione amministrativa migliora le previsioni rispetto alla specificazione completamente comune, ma il vantaggio è inferiore a quello ottenuto mediante i cluster della rete. Le Comunità di Valle rappresentano infatti una classificazione territoriale predefinita, mentre i cluster Graphical Lasso sono costruiti direttamente a partire dalle dipendenze residue tra le serie comunali. Essi possono pertanto riunire comuni non necessariamente contigui, ma caratterizzati da andamenti simili dopo avere controllato la componente temporale, le covariate comunali, la popolazione e l'eterogeneità specifica dei comuni.

5.9 Conclusioni

L'analisi previsiva mostra che la struttura ricavata dal Graphical Lasso contiene informazione utile per descrivere le differenze tra comuni. Il modello con curva globale e deviazioni factor-smooth per tutti i cluster ottiene le migliori prestazioni in termini di MAE, RMSE e RMSLE, sia nella selezione rolling sia nella verifica finale.

La procedura adottata evita l'introduzione di soglie arbitrarie sulla dimensione o sul numero dei cluster. Tutti i gruppi vengono mantenuti distinti e la loro complessità viene controllata attraverso la penalizzazione delle factor-smooth. Inoltre, k , k_{fs} e la soglia StARS vengono selezionati esclusivamente mediante previsioni fuori campione, mentre il periodo di verifica rimane separato fino alla valutazione conclusiva.

In conclusione, i risultati indicano che i cluster individuati dalla rete non costituiscono soltanto una sintesi descrittiva delle dipendenze residue, ma possono essere utilizzati per migliorare concretamente la previsione degli accessi settimanali al Fascicolo Sanitario Elettronico.

Conclusioni

Il presente lavoro ha analizzato gli accessi alla piattaforma TreC+ con l'obiettivo di descriverne le principali dinamiche temporali, demografiche e territoriali e di valutare l'esistenza di strutture residue di dipendenza tra i comuni della Provincia Autonoma di Trento. L'analisi è stata sviluppata secondo un percorso progressivo, partendo dai dati aggregati provinciali e comunali e arrivando alla modellazione dei conteggi settimanali ricavati dal dataset a livello di singolo accesso.

Le analisi esplorative hanno mostrato una crescita rilevante dell'utilizzo della piattaforma nel periodo osservato, associata soprattutto al maggiore utilizzo dell'applicazione mobile. Tra gli accessi da portale si osserva inoltre una progressiva crescita del ruolo di SPID rispetto a TEAM. Le principali operazioni effettuate tramite TreC+ riguardano la consultazione di ricette e referti, confermando il ruolo della piattaforma come strumento di accesso alla documentazione sanitaria personale.

L'andamento temporale non risulta regolare. Accanto alla ciclicità settimanale, caratterizzata da livelli più elevati nei giorni feriali, emergono alcuni picchi concentrati soprattutto nei mesi di gennaio e agosto.

Dal punto di vista demografico, gli accessi si concentrano soprattutto nelle fasce adulte, mentre i tassi rapportati alla popolazione attribuiscono un peso maggiore anche ad alcune fasce più anziane. Gli accessi associati a soggetti di sesso femminile risultano più frequenti sia in termini assoluti sia rispetto alla popolazione residente dello stesso

semplice. Le modalità di autenticazione presentano inoltre profili anagrafici differenti: gli accessi tramite SPID sono associati a età mediane più basse, mentre TEAM risulta maggiormente utilizzato nelle fasce di età più elevate. Il confronto territoriale mostra che la distribuzione degli accessi cambia quando si tiene conto della dimensione e della composizione demografica dei comuni. Alcuni territori caratterizzati da volumi assoluti contenuti presentano infatti un'intensità di utilizzo relativamente elevata, mentre per altri comuni emerge la situazione opposta.

La modellazione settimanale dei conteggi tramite GAM binomiale negativo ha evidenziato associazioni negative tra il numero atteso di accessi e le caratteristiche comunali considerate: altitudine media, superficie territoriale, quota di popolazione femminile ed età media. In altri termini, i comuni più estesi, più elevati, con una popolazione più anziana o a maggioranza femminile tendono a registrare un numero maggiore di accessi.

Dopo l'introduzione dell'effetto casuale comunale, l'età media della popolazione e la quota femminile non hanno tuttavia mantenuto un contributo separatamente identificabile, mentre altitudine e superficie sono rimaste nella specificazione finale. Queste ultime non devono essere interpretate come determinanti dirette degli accessi, ma possono rappresentare caratteristiche territoriali non osservate, come la dispersione degli abitati, l'accessibilità ai servizi o il grado di ruralità del comune. L'effetto casuale ha inoltre permesso di rappresentare una parte dell'eterogeneità persistente tra territori non spiegata dalle covariate disponibili.

Il Graphical Lasso applicato ai residui sbiancati, trasformati e standardizzati ha prodotto una rete molto sparsa, caratterizzata da correlazioni parziali generalmente deboli, più componenti separate e un numero non trascurabile di comuni isolati. Dopo aver rimosso la

componente temporale comune, le caratteristiche osservate e l'eterogeneità comunale, non emerge quindi una dipendenza diffusa tra tutti i territori, ma un insieme limitato di associazioni residue. Nonostante la bassa densità della rete, gli archi mostrano una concentrazione territoriale superiore a quella osservata nelle reti simulate con lo stesso numero di nodi e connessioni. La clusterizzazione Louvain individua tuttavia gruppi che non coincidono completamente con le suddivisioni amministrative esistenti. Alcune comunità comprendono comuni vicini o contigui, mentre altre riuniscono territori più distanti, suggerendo che la prossimità geografica rappresenti soltanto una delle componenti della struttura individuata.

L'analisi previsiva ha fornito l'evidenza principale sull'utilità dei cluster ricavati dalla rete. Il modello con curva stagionale globale e deviazioni factor-smooth specifiche per i cluster Graphical Lasso ha ottenuto le prestazioni migliori nelle ultime 18 settimane riservate alla verifica finale. Rispetto al GAM base, il MAE si riduce da 34.4 a 29.3, l'RMSE da 263 a 228 e l'RMSLE da 0.381 a 0.325. Anche il modello basato sulle Comunità di Valle migliora rispetto alla specificazione comune, ma rimane meno accurato del modello costruito sui cluster della rete.

Il vantaggio previsivo non sembra dipendere da una maggiore complessità delle curve. La procedura rolling seleziona infatti valori contenuti sia per la spline globale sia per le deviazioni di gruppo. I cluster Graphical Lasso forniscono quindi una classificazione utile perché riuniscono comuni caratterizzati da dinamiche residue simili, anche quando non appartengono alla stessa area amministrativa. La rete non rappresenta quindi soltanto una sintesi descrittiva delle dipendenze condizionate, ma produce un'informazione utilizzabile nella previsione degli accessi futuri.

5.10 Limitazioni

I risultati devono essere interpretati alla luce di alcune limitazioni. Il dataset a livello di singolo accesso descrive gli accessi registrati e non gli utenti unici, per cui uno stesso individuo può comparire più volte.

L'estrazione non comprende inoltre otto comuni presenti nella geografia provinciale, tra cui alcuni territori demograficamente rilevanti. Le cause dei principali picchi temporali non sono direttamente osservabili nei dati disponibili. Infine, il periodo settimanale analizzato è relativamente breve e la struttura della rete può dipendere dalle scelte effettuate nella costruzione dei residui, nella selezione del parametro di regolarizzazione e nella definizione del periodo di stima.

5.11 Sviluppi futuri

Possibili sviluppi futuri riguardano l'integrazione di informazioni più complete sui servizi utilizzati, sulle caratteristiche socioeconomiche dei territori e sugli eventuali cambiamenti organizzativi della piattaforma.

Un ulteriore approfondimento potrebbe concentrarsi separatamente sull'utilizzo dell'applicazione e del portale e sulle diverse modalità di autenticazione, così da verificare se le dinamiche temporali, territoriali e demografiche osservate cambino in relazione al canale utilizzato.

Una serie temporale più lunga permetterebbe inoltre di valutare la stabilità della rete nel tempo. Sarebbe infine utile verificare la capacità previsiva delle partizioni su periodi successivi o su dati provenienti da altri contesti territoriali, così da valutarne la generalizzabilità.

In conclusione, i risultati mostrano che, anche dopo aver tenuto conto delle principali componenti sistematiche, rimangono associazioni tra comuni non pienamente riconducibili alle suddivisioni amministrative. I cluster individuati dalla rete risultano inoltre utili in previsione, migliorando la capacità predittiva rispetto sia al modello comune sia a

quello basato sulle Comunità di Valle. La rete assume quindi un valore sia descrittivo sia previsivo.

Appendice A

Armonizzazione dei comuni derivanti da fusione

Per garantire la coerenza tra le diverse basi dati utilizzate nell'analisi, è stata costruita una tabella di corrispondenza tra i comuni soppressi a seguito di fusione e i comuni attualmente esistenti. Tale operazione si è resa necessaria perché alcune fonti territoriali e altimetriche facevano riferimento a un assetto amministrativo precedente rispetto a quello utilizzato nei dati di accesso e nello shapefile comunale corrente.

La Tabella A.1 riporta le corrispondenze utilizzate per ricondurre i comuni precedenti al rispettivo comune attuale. Le corrispondenze sono state verificate manualmente sulla base della documentazione ISTAT relativa ai codici statistici delle unità amministrative territoriali e alle variazioni amministrative e territoriali dei comuni (ISTAT, 2026a).

Il caso di Soraga non è riportato nella tabella, poiché non si tratta di una fusione comunale, ma di una variazione di denominazione: il comune di Soraga ha assunto la denominazione di Soraga di Fassa. Tale variazione è stata comunque considerata nella fase di armonizzazione dei nomi comunali.

Tabella A.1: Corrispondenze tra comuni soppressi e comuni attuali utilizzate per l'armonizzazione territoriale.

Comune attuale	Comuni precedenti confluiti
Altavalle	Faver; Grauno; Grumes; Valda
Altopiano della Vigolana	Bosentino; Centa San Nicolò; Vattaro; Vigolo Vattaro
Amblar-Don	Amblar; Don
Borgo Chiese	Brione; Cimego; Condino
Borgo d'Anaunia	Castelfondo; Fondo; Malosco
Borgo Lares	Bolbeno; Zuclo
Castel Ivano	Ivano Fracena; Samone; Spera; Strigno; Villa Agnedo
Cembra Lisignago	Cembra; Lisignago
Contà	Cunevo; Flavon; Terres
Dimaro Folgarida	Dimaro; Monclassico
Madruzzo	Calavino; Lasino
Novella	Brez; Cagnò; Cloz; Revò; Romallo
Pieve di Bono-Prezzo	Pieve di Bono; Prezzo
Porte di Rendena	Darè; Vigo Rendena; Villa Rendena
Predaia	Coredo; Smarano; Taio; Tres; Vervò
Primiero San Martino di Castrozza	Fiera di Primiero; Siror; Tonadico; Transacqua
San Giovanni di Fassa-Sèn Jan	Pozza di Fassa; Vigo di Fassa
San Lorenzo Dorsino	Dorsino; San Lorenzo in Banale
Sella Giudicarie	Bondo; Breguzzo; Lardaro; Roncone
Terre d'Adige	Nave San Rocco; Zambana
Tre Ville	Montagne; Preore; Ragoli

Continua nella pagina successiva

Tabella A.1: Corrispondenze tra comuni soppressi e comuni attuali utilizzate per l'armonizzazione territoriale. (*continua*)

Comune attuale	Comuni precedenti confluiti
Valdaone	Bersone; Daone; Praso
Vallelaghi	Padergnone; Terlago; Vezzano
Ville d'Anaunia	Nanno; Tassullo; Tuenno
Ville di Fiemme	Carano; Daiano; Varena

Per i comuni derivanti da fusione, l'altitudine media del comune attuale è stata ottenuta come media ponderata per la superficie dei comuni originari:

$$\bar{h}_A = \frac{\sum_{i \in \mathcal{P}(A)} h_i S_i}{\sum_{i \in \mathcal{P}(A)} S_i},$$

dove A indica il comune attuale, $\mathcal{P}(A)$ l'insieme dei comuni originari confluiti in esso, h_i l'altitudine media del comune originario i e S_i la relativa superficie territoriale.

Appendice B

Analisi dei picchi temporali

A integrazione dell'analisi temporale presentata nel Capitolo 2.3, è stato svolto un approfondimento sui principali picchi osservati nella serie degli accessi. L'obiettivo è verificare se i valori più elevati siano riconducibili a singole giornate isolate, circostanza che potrebbe suggerire anomalie puntuali di registrazione, oppure se coinvolgano periodi più estesi. È stata inoltre analizzata la composizione degli accessi nelle giornate di maggiore intensità, per valutare il contributo delle diverse modalità di autenticazione.

B.1 Andamento giornaliero e durata dei picchi

La Figura B.1 mostra il numero giornaliero di accessi osservato nell'intero periodo. Le fasce evidenziate corrispondono ai cinque mesi con il maggior numero complessivo di accessi, mentre la linea più scura rappresenta il numero medio giornaliero di accessi calcolato per ciascuna settimana. Il mese di maggio 2026 è stato escluso dall'analisi perché disponibile soltanto fino all'11 maggio.

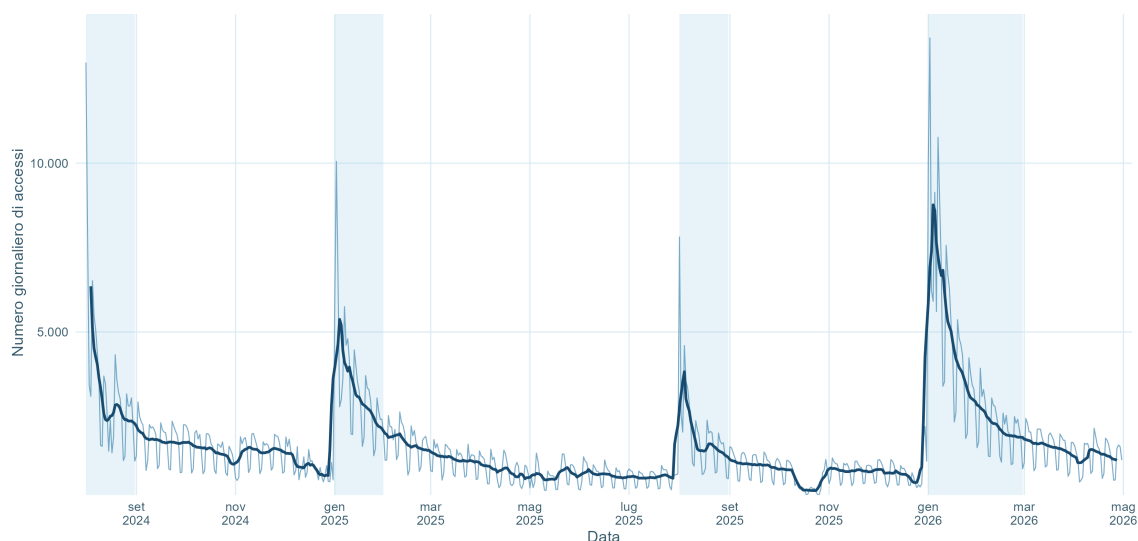


Figura B.1: Andamento giornaliero degli accessi alla piattaforma TreC+. Le fasce azzurre evidenziano i cinque mesi con il maggior numero complessivo di accessi, mentre la linea scura rappresenta il numero medio giornaliero di accessi calcolato per ciascuna settimana.

I mesi con i volumi più elevati sono gennaio 2026, gennaio 2025, agosto 2024, agosto 2025 e febbraio 2026. I principali aumenti si concentrano quindi soprattutto nei mesi di gennaio e agosto, anche se con intensità differenti nei diversi anni.

L'andamento giornaliero mostra che i valori elevati non sono generalmente limitati a un'unica osservazione. In corrispondenza dei principali picchi, anche la media settimanale degli accessi giornalieri aumenta in modo evidente e rimane elevata per più settimane consecutive. Gli incrementi sembrano quindi interessare periodi più prolungati e non dipendere esclusivamente da anomalie isolate.

L'evidenza è particolarmente chiara all'inizio del 2026. La settimana compresa tra il 5 e l'11 gennaio registra 48.272 accessi, mentre quella tra il 29 dicembre 2025 e il 4 gennaio ne registra 41.050. Anche la settimana successiva, dal 12 al 18 gennaio, mantiene un volume elevato, pari a 35.198 accessi. La presenza di più settimane consecutive caratterizzate da valori particolarmente alti conferma quindi che il picco di gennaio 2026 coinvolge un intervallo temporale più esteso.

A livello giornaliero, il massimo si osserva il 2 gennaio 2026, con 13.709 accessi. Seguono il 1° agosto 2024, con 12.980 accessi, e il 7 gennaio 2026, con 10.768 accessi. Anche le giornate immediatamente vicine presentano tuttavia valori elevati, confermando che i principali aumenti non sono limitati a singoli giorni isolati.

B.2 Modalità di autenticazione nelle giornate di picco

Per approfondire la composizione dei picchi, sono state individuate le venti giornate con il maggior numero complessivo di accessi. Queste giornate sono state confrontate con tutti gli altri giorni della serie, considerando sia il numero medio giornaliero di accessi per modalità sia la loro composizione percentuale.

La modalità APP0 è stata aggregata ad APP, data la sua numerosità residuale, mentre DELEGA è stata esclusa dal confronto perché disponibile soltanto nella parte finale del periodo osservato.

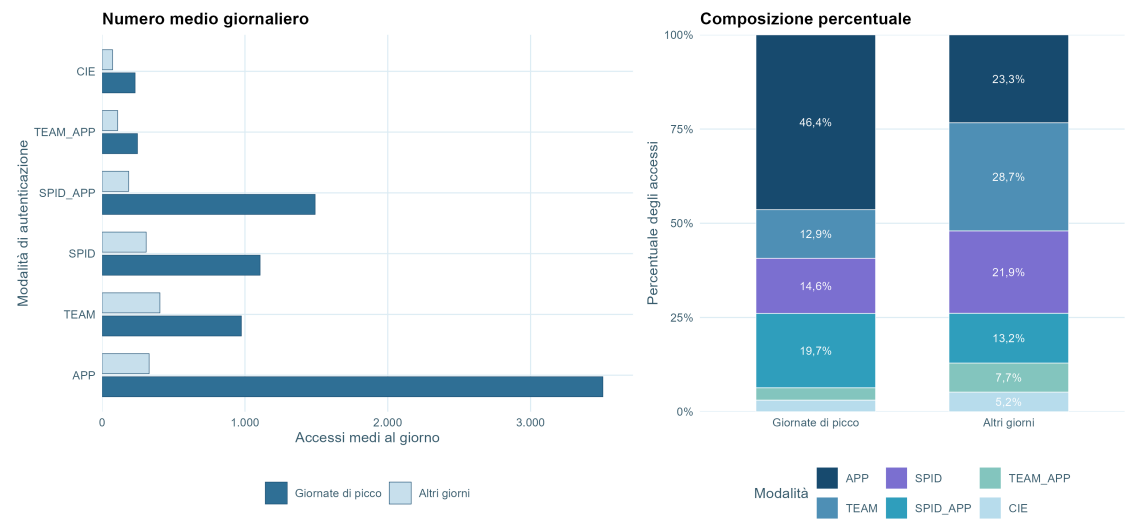


Figura B.2: Modalità di autenticazione nelle venti giornate con il maggior numero di accessi e negli altri giorni della serie. A sinistra è riportato il numero medio giornaliero di accessi; a destra la composizione percentuale.

Il confronto in termini percentuali mostra che nelle giornate di picco aumenta soprattutto il peso della modalità APP, mentre TEAM e SPID rappresentano una quota relativamente più contenuta rispetto agli altri giorni. Anche SPID_APP mantiene un contributo rilevante. La composizione dei picchi risulta quindi maggiormente orientata verso le modalità collegate all'applicazione.

Il confronto dei valori assoluti fornisce tuttavia un'informazione complementare. Nelle giornate di picco non aumentano soltanto gli accessi effettuati tramite APP e SPID_APP, ma cresce anche il numero medio giornaliero di accessi associato alle altre modalità considerate, tra cui TEAM, SPID, TEAM_APP e CIE.

I picchi non sembrano quindi essere prodotti esclusivamente da una singola modalità di autenticazione. Le modalità legate all'applicazione, e in particolare APP, forniscono il contributo più evidente e aumentano anche il proprio peso relativo, ma l'incremento interessa più in generale l'attività di accesso alla piattaforma.

B.3 Conclusioni

L'analisi mostra che i principali picchi temporali non sono limitati a singole giornate isolate, ma coinvolgono più giorni e, nei casi più evidenti, diverse settimane consecutive. L'aumento della media settimanale degli accessi giornalieri nei periodi evidenziati conferma che si tratta di fasi caratterizzate da un livello di attività complessivamente più elevato.

La composizione per modalità di autenticazione evidenzia un ruolo particolarmente importante degli accessi tramite APP e SPID_APP.

Allo stesso tempo, nelle giornate di picco aumentano in valore assoluto anche le altre modalità, indicando che la crescita non riguarda esclusivamente l'utilizzo dell'applicazione.

I dati disponibili non consentono tuttavia di individuare con certezza le cause di questi incrementi.

Bibliografia

Agenzia per l'Italia Digitale (2016). Strategia per la crescita digitale 2014-2020. https://www.agid.gov.it/sites/default/files/repository_files/documenti_indirizzo/strategia_crescita_digitale_ver_def_21062016.pdf.

Agenzia per l'Italia Digitale (2024). Sanità digitale. <https://www.agid.gov.it/it/piattaforme/sanita-digitale>.

Agenzia per l'Italia Digitale (2026). Spid - sistema pubblico di identità digitale. <https://www.agid.gov.it/it/piattaforme/spid>.

Azienda Provinciale per i Servizi Sanitari della Provincia Autonoma di Trento (2026). Trec+. <https://www.asuit.tn.it/servizi/trec>.

Blondel V. D.; Guillaume J.-L.; Lambiotte R.; Lefebvre É. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, **2008**(10), P10008.

Box G. E. P.; Jenkins G. M.; Reinsel G. C.; Ljung G. M. (2015). *Time Series Analysis: Forecasting and Control*. Wiley, Hoboken, 5 edizione.

Deda Next (2026). Apss trento – trec+. <https://www.dedanext.it/portfolio/case-study/apss-trento-trec/>.

Friedman J.; Hastie T.; Tibshirani R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, **9**(3), 432–441.

- ISTAT (2024). Confini delle unità amministrative a fini statistici: confini amministrativi al 1° gennaio 2023.
<https://www.istat.it/notizia/confini-delle-unita-amministrative-a-fini-statistici-al-1-gennaio-2018-2/>.
- ISTAT (2026a). Codici statistici delle unità amministrative territoriali: comuni, città metropolitane, province e regioni.
<https://www.istat.it/classificazione/codici-dei-comuni-delle-province-e-delle-regioni/>.
- ISTAT (2026b). Popolazione residente.
<https://demo.istat.it/app/?i=POS&l=it>.
- ISTAT (2026c). Principali statistiche geografiche sui comuni.
<https://www.istat.it/classificazione/principali-statistiche-geografiche-sui-comuni/>.
- Lauritzen S. L. (1996). *Graphical Models*, volume 17 di *Oxford Statistical Science Series*. Clarendon Press, Oxford.
- Liu H.; Lafferty J.; Wasserman L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research*, **10**, 2295–2328.
- Liu H.; Roeder K.; Wasserman L. (2010). Stability approach to regularization selection for high-dimensional graphical models. In *Advances in Neural Information Processing Systems*, volume 23.
- Meinshausen N.; Bühlmann P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, **34**(3), 1436–1462.
- Ministero della Salute (2022). Fascicolo sanitario elettronico 2.0 entra nella fase operativa.

<https://www.salute.gov.it/new/it/news-e-media/notizie/fascicolo-sanitario-elettronico-20-entra-fase-operativa/>.

Provincia Autonoma di Trento (2026). *TreC+ Guida agli utenti. Edizione 2026*. Dipartimento Salute e Politiche sociali, Trento.

Tashman L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting*, **16**(4), 437–450.

Ver Hoef J. M.; Boveng P. L. (2007). Quasi-poisson vs. negative binomial regression: How should we model overdispersed count data? *Ecology*, **88**(11), 2766–2772.

Wood S. N. (2017). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, Boca Raton, 2 edizione.

Zanutto A. (2016). Seconda indagine sugli esiti della sperimentazione del sistema trec. Relazione tecnica, Università degli Studi di Trento, Trento.

Zanutto A. (2017). Report indagine valutativa cawi sugli iscritti al sistema trec al 31-05-2016. Relazione tecnica, Università degli Studi di Trento, Dipartimento di Economia e Management, Trento.

Zhao T.; Liu H.; Roeder K.; Lafferty J.; Wasserman L. (2012). The huge package for high-dimensional undirected graph estimation in R. *Journal of Machine Learning Research*, **13**, 1059–1062.