

Dinamiche di prezzo e architetture di influenza: Studio sulla propagazione nel mercato retail dei carburanti

Gabriele Giorgio (925678)

20 February 2026

Sistemi Complessi: Modelli e Simulazione a.s. 2024/2025

Abstract

Questo progetto indaga le dinamiche di formazione e propagazione del prezzo dei carburanti in Lombardia, interpretando il mercato come un sistema complesso caratterizzato da interazioni locali. Attraverso l'analisi di dati spaziali e temporali, la ricerca evidenzia come i prezzi non evolvano in modo isolato, ma siano il frutto di meccanismi di imitazione tra vicini, distinguendo tra diverse tipologie di attori di mercato: dai grandi leader di volume ai «market movers» strategici.

Utilizzando quotazioni internazionali del greggio (Brent) e i prezzi alla pompa, è stata riscontrata la presenza per la benzina del fenomeno «Rocket & Feather»: un'asimmetria di trasmissione per cui i prezzi al dettaglio reagiscono con immediatezza ai rincari della materia prima (fase «razzo»), ma mostrano una significativa rigidità e lentezza nell'adeguarsi ai ribassi (fase «piuma»).

L'integrazione di un modello di regressione spaziale (SAR) e l'analisi topologica ha consentito la mappatura delle zone di influenza. Tali evidenze strutturali, unite alle dinamiche comportamentali emerse, hanno guidato lo sviluppo di un simulatore sperimentale calibrato su dati reali.

Indice

1	Introduzione	3
2	Creazione delle reti	4
2.1	Creazione reti temporali	5
2.2	Creazione reti delle distanze	6
3	Analisi reti distanze 1-5km	8
3.1	Matching	8
3.2	High matching edges e influencers	9
4	Analisi reti distanza 3km	12
4.1	Influencers nelle reti 3km	12
4.1.1	Influencers e grado	13
4.1.2	Influencers e brand	14
4.2	Cascate nelle reti 3km	17
4.2.1	Cascate e grado	17
4.2.2	Cascate e brand	19
4.2.3	Cascate e pattern temporali	22
4.2.4	Total impact, frequenza per profondità delle cascate	25
5	Analisi reti non orientate distanza 3km	29
5.1	Distribuzione del grado	29
5.2	Prezzi mediani e nuovo matching	32
5.3	Correlazione tra prezzi e centralità	32
5.4	Community detection, Louvain tramite ANOVA	35
5.5	Correlazione tra differenza di prezzo e matching	37
6	Effetto Rocket & feather	38
6.1	Prezzi medi giornalieri e movimento del greggio	39
6.2	Effetto Rocket & feather nel dataset	40
7	Regressione spaziale SAR	42
8	Simulatore	46
9	Sviluppi futuri	49
10	Conclusioni	49
	Bibliografia	51

1 Introduzione

Si precisa che la stesura della presente relazione è stata supportata da un LLM (Gemini). L'intelligenza artificiale è stata impiegata esclusivamente come strumento di supporto linguistico per l'ottimizzazione dell'esposizione, la revisione stilistica e l'affinamento del lessico, al fine di garantire una forma espositiva consona all'ambito accademico.

Ai fini della presente trattazione, si è scelto di non includere direttamente il codice sorgente, privilegiando una descrizione analitica delle procedure applicate e la discussione dei risultati ottenuti. Tuttavia, per garantire la massima trasparenza e riproducibilità della ricerca, l'intero apparato computazionale è stato reso disponibile in una repository GitHub dedicata. All'interno della repository, il codice è organizzato in notebook interattivi, dove ogni blocco logico è accompagnato da commenti esplicativi. Per agevolare la consultazione, la struttura dei notebook segue fedelmente l'indice della presente relazione: i riferimenti numerici qui riportati corrispondono esattamente alle sezioni di codice implementate, permettendo così un riscontro immediato tra il modello teorico e la sua esecuzione pratica.

Repository del Progetto

Il codice completo e le procedure di analisi sono disponibili su GitHub:

github.com/gabriele11231/fuel_price_analyses

Per secoli (da Newton in poi), la scienza è stata dominata dal Riduzionismo. L'idea era semplice: se vuoi capire come funziona un orologio, lo smonti, studi ogni ingranaggio separatamente e poi sai come funziona il tutto. Tuttavia, nel XX secolo, ci si rese conto che questo metodo falliva con sistemi come il clima, il cervello umano o l'economia. Smontare un cervello non ti spiega la coscienza; analizzare una singola goccia d'acqua non ti spiega un'onda anomala.

La moderna scienza della complessità trova uno dei suoi momenti fondativi nel 1961, quando il meteorologo Edward Lorenz scoprì fortuitamente che una minuscola troncatura decimale nei dati di input di un modello atmosferico generava scenari drasticamente divergenti. (Lorenz, 1963) Questa osservazione della dipendenza sensibile dalle condizioni iniziali (nota come Effetto Farfalla¹) segnò la fine del determinismo classico, dimostrando l'intrinseca imprevedibilità a lungo termine dei sistemi dinamici non lineari.

Nel suo lavoro fondamentale sull'architettura della complessità, Herbert Simon propone una visione strutturale dei sistemi complessi. Egli argomenta che la complessità non risiede solo nella numerosità delle parti, ma nella natura non lineare delle loro interazioni, le quali rendono difficile inferire le proprietà del sistema aggregato partendo dalla conoscenza dei singoli componenti. Individua nella gerarchia l'architettura distintiva di tali sistemi: la complessità si organizza quasi sempre in una struttura di sottosistemi annidati, dove la stabilità del tutto dipende dalla coesione delle singole parti e dalla loro articolazione in livelli di organizzazione via via superiori. (Simon, 1962).

Da questi presupposti teorici, si intende formulare una definizione sintetica che guidi l'analisi successiva:

¹Nel paper originale del 1963, Lorenz utilizzò l'esempio di un **gabbiano**. La metafora della farfalla fu introdotta solo nel 1972, come titolo di una conferenza («Does the Flap of a Butterfly's Wings in Brazil Set Off a Tornado in Texas?»), probabilmente ispirata dalla forma geometrica dell'attrattore stesso.

Definiamo sistema complesso un aggregato di componenti le cui interazioni reciproche danno luogo a fenomeni di emergenza non deducibili dall'analisi atomistica delle parti. La complessità di tale sistema è ulteriormente alimentata dal rapporto con l'ambiente esterno: i fattori esogeni non influenzano le componenti in modo isolato, ma vengono filtrati dalla rete di relazioni interne, determinando una risposta adattiva globale del sistema stesso.

Applicando la definizione appena fornita al mercato al dettaglio dei carburanti, è possibile osservare il passaggio dalla semplicità alla complessità strutturale. Un singolo distributore, analizzato isolatamente, appare come un'entità dal comportamento erratico. Tuttavia, integrando nel sistema un secondo distributore e variabili esogene (come il prezzo al barile), il sistema manifesta la sua vera architettura. Le interazioni tra i prezzi non sono semplici somme di decisioni individuali, ma mostrano schemi di reazione e adattamento temporale.

Solo osservando l'interazione tra queste parti possiamo comprendere la logica del sistema, la quale resta invisibile finché l'indagine rimane confinata al singolo elemento.

2 Creazione delle reti

Al fine di mappare e analizzare la rete dei distributori di carburante in Lombardia, è stata costruita una base dati sfruttando gli Open Data messi a disposizione dal Governo Italiano. Il dataset comprende l'anagrafica completa degli impianti attivi sul territorio nazionale e l'aggiornamento quotidiano dei relativi prezzi alla pompa, rilevati sistematicamente alle ore 8:00 di mattina.

Portale open data del [Ministero delle Imprese e del Made in Italy](#).

Struttura del Dataset Anagrafica:

idImpianto identificativo univoco del distributore.

Gestore ragione sociale o nome del responsabile.

Bandiera marchio commerciale (es. ENI, IP, No Logo).

Tipo Impianto tipologia (Stradale o Autostradale).

Nome Impianto denominazione specifica del punto vendita.

Indirizzo/Comune/Provincia dati di localizzazione amministrativa.

Latitudine/Longitudine coordinate geografiche per la mappatura GIS.

Il campo relativo alla provincia riveste un ruolo cruciale nella fase di pre elaborazione dei dati; esso funge da filtro selettivo per circoscrivere l'analisi al territorio della regione Lombardia. Attraverso questo parametro, è possibile escludere sistematicamente le unità estranee all'area di studio, garantendo la coerenza geografica del dataset rispetto agli obiettivi della ricerca.

Struttura del Dataset Prezzi:

idImpianto chiave esterna per il collegamento con l'anagrafica (permette di geolocalizzare il prezzo).

descCarburante tipologia di prodotto (es. Benzina, Gasolio, GPL).

prezzo valore unitario al litro.

isSelf indicatore booleano (1 se il prezzo è riferito al self-service, 0 per il servito).

dtComu timestamp della comunicazione del prezzo al Ministero.

La granularità del dataset comporta che per ogni singola giornata di osservazione siano presenti molteplici record associati al medesimo distributore. Tale ridondanza informativa riflette l'offerta commerciale dell'impianto: ogni stazione può infatti comunicare prezzi distinti in base alla tipologia

di carburante (es. benzina, gasolio, GPL) e alla modalità di erogazione (self service o servito). Al fine di garantire la comparabilità dei dati e ridurre il rumore statistico, l'analisi è stata circoscritta a un sottinsieme specifico di osservazioni. Dunque, sono stati filtrati i record relativi esclusivamente a benzina e diesel (gasolio), limitando l'osservazione alla sola modalità di erogazione non self-service (servito). L'intervallo temporale considerato per l'analisi dei prezzi copre un trimestre completo, estendendosi dal 1° aprile 2025 al 30 giugno 2025. Questo arco temporale di 91 giorni permette di campionare una serie storica sufficientemente ampia per identificare non solo le reazioni estemporanee tra distributori limitrofi, ma anche la propagazione di eventuali shock esogeni derivanti dalle fluttuazioni del mercato energetico nel periodo primaverile.

2.1 Creazione reti temporali

Per l'analisi delle dinamiche dei prezzi dei carburanti nella regione Lombardia, sono state implementate due reti temporali distinte, una per la benzina e una per il gasolio, avvalendosi della libreria Python Raphtory.

A differenza dei grafi statici, la struttura della rete temporale permette di catturare l'evoluzione del sistema, dove la presenza di nodi e archi, nonché i relativi attributi, sono definiti in funzione del tempo (timestamp) t .

Il modello si basa sui seguenti criteri di costruzione:

Nodi (V) ogni nodo rappresenta una singola stazione di servizio attiva sul territorio lombardo. A ciascun nodo è associato un attributo temporale corrispondente al prezzo di vendita praticato in una specifica data di osservazione.

Archi (E) un link tra due nodi u e v sussiste al tempo t se e solo se le due stazioni presentano il medesimo prezzo per la tipologia di carburante analizzata nel medesimo giorno.

Rete benzina: include esclusivamente le rilevazioni del prezzo «non servito» della benzina.

Rete diesel: include esclusivamente le rilevazioni del prezzo «non servito» del gasolio.

Se una stazione u presenta il medesimo prezzo della stazione v al tempo t , la condizione $P_{u(t)} = P_{v(t)}$ implica necessariamente una reciprocità tra i due nodi. Di conseguenza, il modello teorico di riferimento è quello di un grafo non orientato. Tuttavia, la libreria Raphtory opera nativamente su strutture di dati orientate.

Per superare questa divergenza strutturale e mappare correttamente la simmetria dei prezzi, è stata implementata una convenzione di ordinamento degli identificativi dei nodi:

Normalizzazione degli archi per ogni coppia di stazioni connesse, l'arco viene creato imponendo che il nodo sorgente (source) abbia sempre un ID numerico inferiore al nodo di destinazione (destination), ovvero $ID\{source\} < ID\{destination\}$.

Consistenza temporale tale procedura garantisce che la medesima relazione tra due stazioni venga identificata univocamente nel tempo (ad esempio, sempre come l'arco $12 \rightarrow 13$), evitando la duplicazione o l'inversione dei link in giorni differenti. Questa formalizzazione permette di trattare la rete come non orientata ai fini dell'analisi statistica.

L'architettura delle reti temporali sopra descritta funge da infrastruttura portante per l'intero progetto, operando come una struttura dati dinamica ottimizzata per l'archiviazione e la gestione integrata delle informazioni. Nello specifico, il sistema permette di convogliare in un unico modello relazionale i dati anagrafici dei benzinali e la serie storica dei prezzi. L'adozione di Raphtory è motivata dalle sue presta-

zioni superiori nella gestione di grafi temporali su larga scala, garantendo un'efficienza computazionale e una scalabilità non riscontrabili nelle librerie avversarie.

2.2 Creazione reti delle distanze

In questa sezione viene descritta la nascita delle reti delle distanze, la cui funzione è connettere coppie di distributori sulla base della loro prossimità spaziale. Sebbene l'anagrafica fornisca le coordinate geografiche precise (latitudine e longitudine), si è scelto di non adottare la distanza euclidea o «in linea d'aria». Tale metrica, infatti, ignorerebbe la morfologia del territorio e l'effettiva rete viaria: due stazioni potrebbero risultare vicine in termini di coordinate ma essere separate da ostacoli naturali o percorsi tortuosi. Ad esempio, una distanza lineare di 3 km tra una stazione in quota e una a valle potrebbe corrispondere a un percorso stradale di 15 km. Per riflettere il comportamento reale dell'utente e la reale concorrenza tra impianti, la rete è stata dunque modellata sulla distanza effettiva di percorrenza stradale.

Per superare i limiti della metrica euclidea, è stata impiegata la libreria Python OSMnx ([Boeing, 2025](#)), uno strumento specializzato nella modellazione e analisi di reti urbane basate su dati OpenStreetMap. Il framework ha permesso di caricare l'intera rete stradale della regione Lombardia, convertendola in un grafo pesato dove le intersezioni stradali fungono da nodi e i segmenti viari rappresentano gli archi che ereditano un peso corrispondente alla loro lunghezza fisica. Nonostante l'elevata richiesta di memoria RAM per il processamento di un'area geografica così estesa, questo approccio ha consentito di calcolare la distanza reale tra i distributori simulando l'effettiva percorrenza stradale tramite auto-veicolo.

Il processo di calcolo delle distanze ha previsto inizialmente la proiezione di ciascuna stazione di servizio sul grafo stradale della Lombardia: partendo dalle coordinate geografiche (latitudine e longitudine), è stato individuato il nodo della rete viaria più prossimo a ogni impianto. Una volta associata ogni stazione a un vertice del grafo, è stato possibile determinare la distanza stradale effettiva tra ogni coppia di distributori applicando l'algoritmo di Dijkstra ([Dijkstra, 1959](#)).

Il calcolo delle distanze stradali tra tutte le coppie di stazioni ha richiesto un alto numero di iterazioni dell'algoritmo di Dijkstra, un'operazione che comporta un onere computazionale elevato. Per ottimizzare i tempi di elaborazione, è stata adottata la libreria cuGraph (progetto Nvidia RAPIDS), che configurata su sistema operativo Linux (tramite Windows Subsystem for Linux - WSL) permette di sfruttare la potenza di calcolo parallelo delle schede grafiche Nvidia. Grazie all'architettura CUDA, l'algoritmo viene eseguito direttamente sulla GPU, riducendo drasticamente i tempi di esecuzione rispetto a un'implementazione tradizionale su CPU.

Nonostante l'adozione di tecniche di calcolo parallelo su GPU, l'elaborazione dell'intera matrice delle distanze ha richiesto un tempo di computazione di circa 180 minuti. Le prestazioni sono state misurate su una workstation equipaggiata con il seguente hardware: CPU Intel Core Ultra 7 265K, 32 GB di RAM DDR5 (6400 MHz) e una GPU NVIDIA RTX 5070FE. L'impiego di questa specifica configurazione hardware è stato determinante: un'implementazione equivalente basata esclusivamente su CPU avrebbe richiesto tempi di calcolo proibitivi, rendendo l'analisi dell'intero sistema regionale virtualmente impraticabile.

Maggiori informazioni sul progetto [RAPIDS](#).

Per garantire la persistenza dei dati ed evitare la reiterazione dell'oneroso calcolo computazionale, i risultati sono stati archiviati in formato CSV. È stata applicata una soglia di filtraggio spaziale cautele-
lativa pari a 30 km, sebbene le analisi successive si siano focalizzate su una distanza massima di 5 km.
Tale discrepanza è stata deliberatamente introdotta per mantenere un margine di flessibilità analitica,
senza compromettere l'efficienza, ottenendo un dataset sensibilmente più agile rispetto al salvataggio
integrale delle distanze.

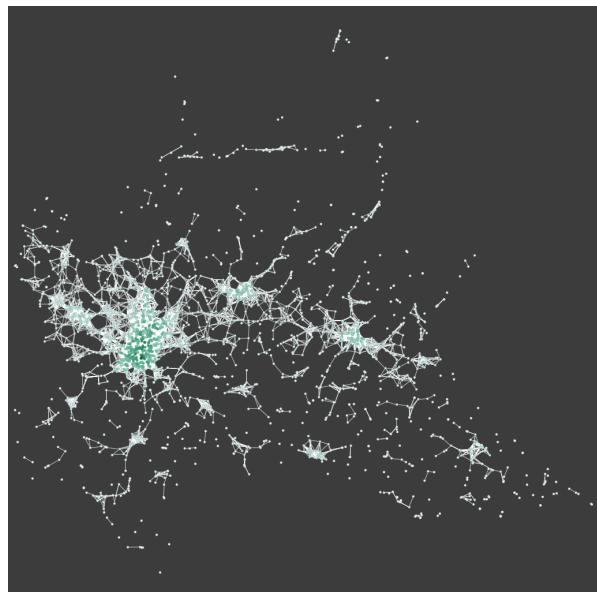
A partire dalle distanze calcolate, sono state generate cinque reti statiche di prossimità, caratterizzate
da un set di nodi invariante (le stazioni di rifornimento) e da una differente configurazione di archi
basata su soglie chilometriche discrete.

Tali reti, definite come «Reti delle distanze» (distance_network), mappano le interazioni spaziali in un
intervallo compreso tra 1 e 5 km. Nello specifico, la connettività è stata discretizzata in classi disgiunte:
la rete da 1 km connette coppie di impianti con distanza $0 < d \leq 1$ km, quella da 2 km isola i legami
nell'intervallo $1 < d \leq 2$ km, proseguendo con questa logica incrementale fino alla soglia dei 5 km.
Questa stratificazione permette di analizzare come la vicinanza fisica influenzi il coordinamento dei
prezzi, isolando l'effetto della competizione locale per fasce di distanza. Le reti che ne fuoriescono sono
tutte non orientate e prive di pesi sugli archi.

Al contrario delle reti temporali, tutte le reti statiche analizzate nel presente progetto sono state
modellate e implementate utilizzando NetworkX, una libreria di Python specializzata nello studio
di reti e grafi.



Rete 1km



Rete 5km

Figura 1: confronto strutturale tra la rete delle distanze a 1km (sinistra) e a 5km (destra). Si noti come
l'aumento del range di prossimità porti alla formazione di una «componente gigante» che connette la
quasi totalità del sistema lombardo.

3 Analisi reti distanze 1-5km

L'analisi di queste reti mira a quantificare i fenomeni di imitazione dei prezzi tra i distributori e a studiare come tale comportamento si propaghi nelle dimensioni spaziale e temporale. L'indagine è inoltre volta a verificare l'esistenza di leader di mercato (o influencer) che esercitano un impatto predominante sulle dinamiche di coordinamento locale.

Prima di procedere con l'elaborazione, è stata effettuata una fase di pulizia e allineamento dei dati per garantire la coerenza tra le diverse strutture di rete. Nello specifico, da ciascuna delle cinque reti delle distanze sono stati rimossi tutti i nodi (stazioni di servizio) per i quali non si disponeva di una serie completa di prezzi giornalieri per benzina e diesel. Le reti delle distanze conterranno ciascuna 2661 nodi (stazioni).

3.1 Matching

Trendo ispirazione da uno studio sui meccanismi di coordinamento dei prezzi tra i distributori nella città Cuneo nel 2011 ([Alderighi & Baudino, 2015](#)), è stata sviluppata un'originale metrica di matching volta a quantificare l'intensità dei comportamenti imitativi tra coppie di stazioni di servizio.

Poiché le reti delle distanze create precedentemente contengono esclusivamente informazioni topologiche e spaziali, esse fungono da base strutturale neutra alla quale vengono integrate le reti temporali dei prezzi. Attraverso l'unione di queste componenti, è possibile calcolare il matching differenziato per i due tipi di carburanti.

Questo processo porta alla generazione di 10 reti di matching distinte:

5 reti per la benzina corrispondenti alle fasce di distanza da 1 a 5 km.

5 reti per il diesel corrispondenti alle medesime fasce chilometriche.

Il fenomeno del price matching viene qui formalizzato come l'istanza in cui una stazione B, a distanza di x giorni (corrispondenti al lag temporale), replica la traiettoria di prezzo (incremento, decremento o stabilità) della stazione A ad essa connessa. Per essere classificata come imitazione, inoltre, la differenza assoluta tra il prezzo corrente di B e il prezzo di A registrato nel giorno di riferimento deve risultare inferiore alla soglia di 0,02 euro.

Tale relazione viene quantificata rapportando il numero totale di match rilevati al numero complessivo di osservazioni nel periodo considerato, esprimendo il risultato in valore percentuale.

L'analisi del matching è stata applicata a tutte le reti delle distanze (da 1 km a 5 km), e per ogni rete è stato valutato il comportamento imitativo considerando sei diversi intervalli temporali, con un ritardo (lag) compreso tra 0 e 5 giorni.

Si definisce l'indicatore di **match** I_{match} tra la stazione A e la vicina B al tempo t con un ritardo temporale τ come:

$$\Delta P_{A,t} = P_{A,t+1} - P_{A,t}$$

$$\Delta P_{B,t+\tau} = P_{B,t+\tau+1} - P_{B,t+\tau}$$

$$I_{\text{match}}(t, \tau) = \begin{cases} 1 & \text{se } \text{sgn}(\Delta P_{A,t}) = \text{sgn}(\Delta P_{B,t+\tau}) \text{ e } |P_{A,t+1} - P_{B,t+\tau+1}| \leq 0.02 \\ 0 & \text{altrimenti} \end{cases}$$

Dove:

- t è il giorno corrente.
- τ è il ritardo in giorni (lag).
- $\text{sgn}(x)$ è la funzione segno (che restituisce 1, -1 o 0).
- La prima condizione verifica che la direzione del cambiamento sia identica (o che entrambe siano stazionarie).
- La seconda condizione verifica che la differenza assoluta di prezzo finale rientri nella soglia di 0,02€.

Per analizzare correttamente le dinamiche di influenza, la struttura delle reti è stata trasformata da non orientata a orientata. Di conseguenza, ogni legame tra due nodi non è più rappresentato da un singolo arco generico, ma da due archi diretti contrapposti.

L'orientamento degli archi riflette la gerarchia di influenza: un arco uscente dal nodo i verso il nodo j indica che il distributore i tende a «farsi influenzare» o a replicare le variazioni di prezzo proposte dal distributore j . Questa distinzione è superflua per il «lag 0», in cui il matching tra la stazione A e la stazione B è speculare (le variazioni avvengono simultaneamente e la percentuale di aderenza coincide in entrambe le direzioni). Tuttavia, l'orientamento diventa cruciale per i lag superiori a 0: in questi casi, la capacità di B di seguire A dopo n giorni è intrinsecamente diversa dalla capacità di A di seguire B nello stesso intervallo. Tale approccio permette di discernere la gerarchia del mercato, identificando se sia la stazione A a conformarsi alle strategie di B o viceversa. Mentre nel lag 0 le percentuali di matching sulle due direzioni saranno identiche, nei lag successivi esse divergeranno sensibilmente, consentendo di mappare i flussi di influenza.

Ogni arco è, quindi, stato arricchito con sei attributi scalari, definiti «matching_x», dove l'indice x denota l'ampiezza del lag temporale espresso in giorni. Questi attributi rappresentano la frequenza percentuale dei match calcolata per ogni specifica finestra temporale, permettendo di mappare non solo l'intensità del legame tra le stazioni, ma anche la latenza della risposta competitiva.

$$G = (V, E) \rightarrow G' = (V, A)$$

Dove per ogni arco non orientato $\{i, j\} \in E$, si generano due archi orientati:

- $(i \rightarrow j)$ con peso $W_{i \rightarrow j} = \text{Match}_{i \rightarrow j}(t, \tau)$
- $(j \rightarrow i)$ con peso $W_{j \rightarrow i} = \text{Match}_{j \rightarrow i}(t, \tau)$

Per $\tau = 0$, $W_{i \rightarrow j} = W_{j \rightarrow i}$. Per $\tau > 0$, tipicamente $W_{i \rightarrow j} \neq W_{j \rightarrow i}$.

Qualora il numero di osservazioni temporali disponibili per una coppia di stazioni non raggiunga la soglia minima richiesta per un determinato lag, l'attributo «matching_x» viene impostato al valore convenzionale di -1. Tale parametro funge da filtro per la pulizia topologica della rete: ogni arco che presenta un valore «matching_5» pari a -1 viene rimosso dal grafo finale.

3.2 High matching edges e influencers

Una volta generate le reti di matching, è stata condotta un'analisi granulare per identificare gli «high matching edges e gli «influencers».

Un arco viene classificato come «high matching edge» qualora l'attributo «matching_x» presenti un valore superiore a 50; tale soglia indica che la stazione «follower» ha replicato la traiettoria di prezzo della stazione «leader» in più della metà delle osservazioni disponibili per quel determinato lag e km. Parallelamente, introduciamo il concetto di «influencer» per identificare i nodi con la più alta capacità

di coordinamento spaziale. Un nodo è definito «influencer» se almeno il 50% dei suoi vicini che lo imitano presenta un punteggio di matching superiore a 50. In altri termini, un «influencer» è un nodo in cui almeno la metà degli archi incidenti entranti (ovvero gli archi di chi «si fa influenzare» da esso) è costituita da «high matching edges».

Entrambe le metriche sono state calcolate ed estratte sistematicamente per ogni combinazione di raggio chilometrico (1-5 km), intervallo di lag (0-5 giorni) e tipologia di carburante (benzina e diesel), permettendo di mappare l'evoluzione nel tempo e nello spazio.

$$\text{High matching edge}_{A \rightarrow B}(x) \Leftrightarrow \text{matching}_x > 50$$

$$\text{Influencer}(A, x) \Leftrightarrow \frac{|\{B \in \text{Neighbors}(A) : \text{Hme}_{B \rightarrow A}(x)\}|}{|\text{In-Degree}(A)|} \geq 0.5$$

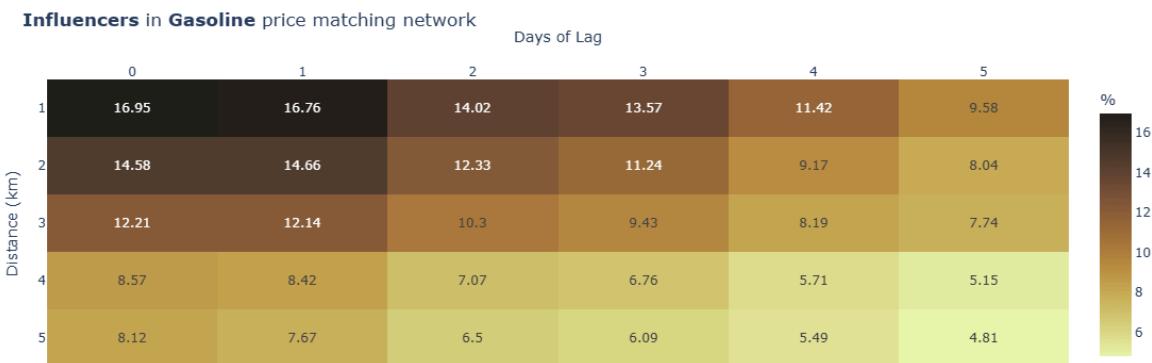
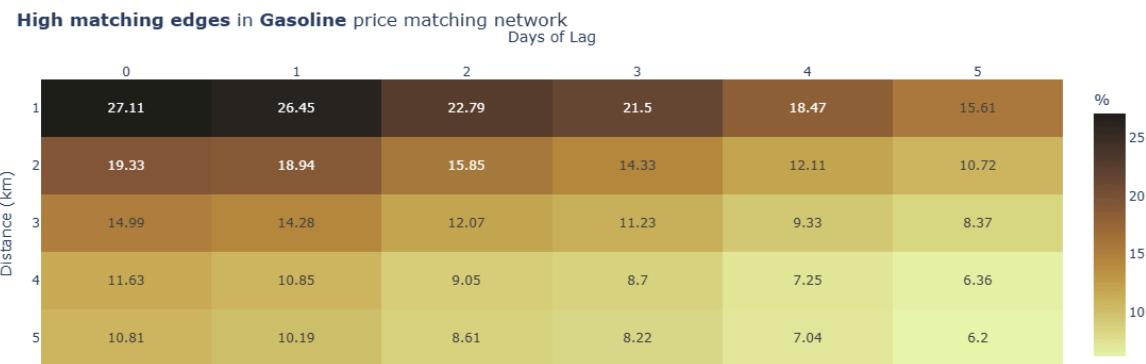


Figura 2: benzina, «high matching edges» (sopra) e «influencers» (sotto).



Figura 3: diesel, «high matching edges» (sopra) e «influencers» (sotto).

Le quattro heatmap sopra riportate illustrano la distribuzione densimetrica degli «high matching edges» e degli «influencers» in funzione delle due variabili cardine dell’analisi: il ritardo temporale (lag) e la prossimità spaziale. Nello specifico, l’asse delle ascisse (*x*) riporta i giorni di lag (da 0 a 5), mentre l’asse delle ordinate (*y*) categorizza le stazioni in base ai raggruppamenti chilometrici di distanza stradale (da 1 a 5 km). L’intensità cromatica delle mappe rappresenta la variazione percentuale di questi elementi all’interno della rete, permettendo di identificare visivamente come la propensione al coordinamento dei prezzi decada o si trasformi all’aumentare della distanza e della latenza temporale.

I dati mostrano che il mercato della benzina è significativamente più reattivo e interconnesso rispetto a quello del diesel. Le percentuali di matching più elevate suggeriscono che i gestori monitorano con maggiore attenzione il prezzo della «verde». Questo è coerente con la natura della domanda: la benzina è il prodotto primario per i consumatori privati (B2C), che sono molto sensibili alle differenze di prezzo esposte sul tabellone. Il diesel, avendo una forte componente business/trasporti (spesso regolata da carte carburante o contratti aziendali), subisce una pressione competitiva visiva leggermente inferiore.

L’analisi dei dati evidenzia che l’incremento del raggio chilometrico (km) e del ritardo temporale (lag) agisce come fattore limitante sulla percentuale di «high matching edges» e di «influencers» all’interno delle reti. Coerentemente con la letteratura, all’aumentare della distanza fisica tra le stazioni diminuisce

la probabilità di un'influenza reciproca significativa. Parallelamente, maggiore è l'intervallo temporale che intercorre da una variazione di prezzo, minore è la probabilità che le stazioni limitrofe si adattino alla nuova quotazione proposta.

In tutti e quattro i grafici emerge un pattern universale: la concentrazione massima di attività (i colori più scuri) si trova sempre nell'angolo in alto a sinistra (distanza 1km, Lag 0-1), questo conferma la legge della prossimità: la competizione nel mercato dei carburanti è un fenomeno iper locale. Un benzinaio è ossessionato dal prezzo del vicino immediato entro 1 km (coerentemente con quanto affermato da ([Alderighi & Baudino, 2015](#))) e reagisce quasi istantaneamente (nello stesso giorno o il giorno dopo).

È fondamentale precisare una distinzione metodologica riguardo al «lag 0»: in questo scenario, ogni «high matching edge» è intrinsecamente bidirezionale poiché la variazione di prezzo è simultanea (il match da $A \rightarrow B$ implica necessariamente quello da $B \rightarrow A$). Per una comparazione rigorosa con i lag successivi, il numero di «high matching edges» a «lag 0» andrebbe idealmente dimezzato. Sebbene esistano archi bidirezionali anche per lag superiori a 0, il significato informativo è differente: essi non indicano simultaneità, bensì un adattamento reciproco e continuo tra due stazioni che rispondono l'una all'altra sistematicamente dopo un determinato numero di giorni.

4 Analisi reti distanza 3km

Sulla base delle evidenze emerse dalle heatmap riportate in [Figura 2](#) e [Figura 3](#), si è proceduto alla sintesi di una rete unificata per ciascun carburante (benzina e diesel). Questa nuova struttura integra le prime tre fasce di prossimità, connettendo tutti i distributori caratterizzati da una distanza stradale reciproca compresa tra 0 e 3 km.

Tale consolidamento è stato dettato da una duplice esigenza: da un lato, l'osservazione empirica ha mostrato come l'influenza spaziale sia massima nel primo chilometro per poi decadere progressivamente; dall'altro, la gestione di dieci reti distinte rendeva l'analisi d'insieme eccessivamente frammentaria. Riducendo il dataset a due reti principali, una per il diesel e una per la benzina è stato possibile condensare tutte le analisi spaziali e temporali precedentemente svolte in un modello più robusto e leggibile. Questo approccio permette di mantenere una visione sistematica della competizione locale, preservando al contempo la significatività statistica dei legami identificati tra 0 e 3 km.

Sebbene l'analisi visiva delle heatmap confermi un decadimento della capacità di influenza già oltre la soglia del primo chilometro, si è optato per un'estensione del raggio a 3km. Tale scelta rappresenta un compromesso metodologico volto a bilanciare la precisione della dipendenza spaziale con la necessità di preservare un volume di dati e di archi sufficiente a garantire la significatività statistica delle analisi successive. La rete risultante permette quindi di catturare non solo le interazioni dirette di prossimità immediata, ma anche le dinamiche di coordinamento di secondo ordine che potrebbero emergere su una scala urbana più ampia.

4.1 Influencers nelle reti 3km

All'interno della rete consolidata (0-3 km), il calcolo degli «influencers» è stato eseguito per ogni intervallo temporale, assegnando a ciascun nodo un set di attributi definiti « $x_influencer$ » (con $x \in \{0, \dots, 5\}$). Tuttavia, per le fasi successive dell'analisi, sono stati considerati esclusivamente i nodi identificati come influencer al «lag 1». Tale restrizione è stata imposta per ridurre fattori esterni nella dinamica di formazione dei prezzi.

Un nodo classificato come «influencer» al «lag 1» esercita un potere condizionante immediato: la

sua variazione di prezzo viene recepita e replicata dalla maggioranza dei suoi vicini entro le 24 ore successive. Al contrario, l'estensione dell'analisi a lag superiori (fino a 5 giorni) introdurrebbe un rumore statistico eccessivo; in un intervallo di tempo così ampio, l'allineamento dei prezzi potrebbe essere influenzato da fattori macroeconomici esterni o da variazioni della rete distributiva globale, rendendo difficile isolare l'effettivo potere di leadership del singolo distributore.

Si precisa che, per coerenza metodologica, la definizione di «influencer» (basata sulla quota di «high matching edges» incidenti) è rimasta invariata rispetto a quanto stabilito nelle fasi precedenti [Sezione 3.2](#).

4.1.1 Influencers e grado

Una volta isolati i nodi classificati come «influencer» al «lag 1», l'analisi si è focalizzata sulla caratterizzazione della loro importanza strutturale all'interno della rete dei 3 km. A tal fine, è stato effettuato un confronto tra il grado medio degli «influencer» e quello dei «non influencer».

L'obiettivo di questa comparazione è determinare se la capacità di orientare il mercato sia correlata alla connettività fisica del distributore: in altre parole, se i leader di prezzo siano anche i nodi topologicamente più centrali, ovvero quelli che vantano il maggior numero di vicini nel raggio di prossimità definito. Di seguito vengono presentati i risultati ottenuti per i due tipi di carburante, seguiti da una discussione sulle evidenze emerse.

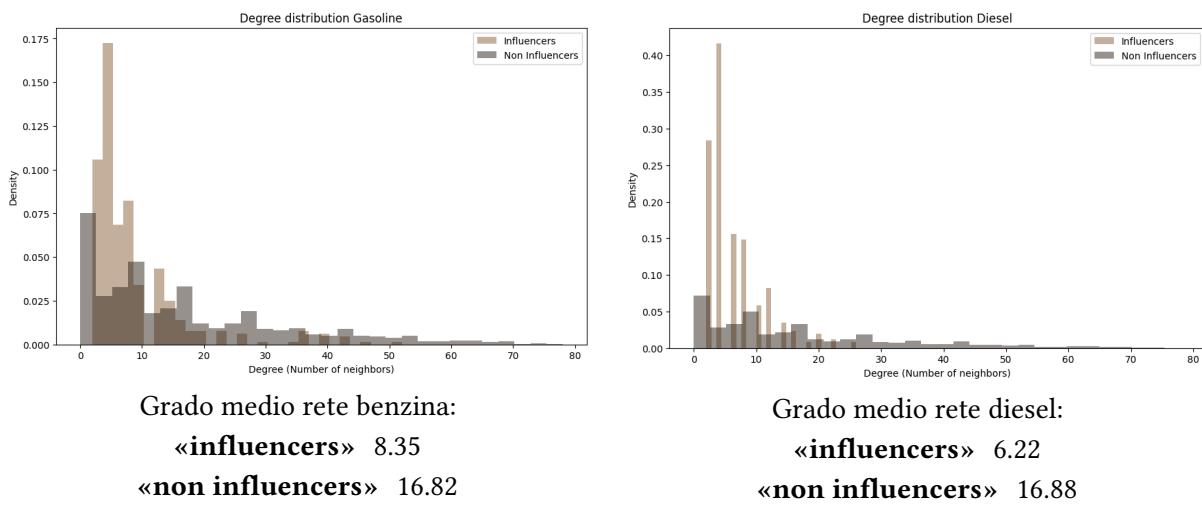


Figura 4: confronto del grado medio tra «influencer» e «non-influencer» (lag 1) nella rete 3km.

Contrariamente a quanto ipotizzato dalla teoria classica delle reti, dove la centralità di grado (degree centrality) è spesso sinonimo di potere e influenza, i risultati ottenuti per entrambi i carburanti (mostrati in [Figura 4](#)) indicano che gli «influencer» al «lag 1» possiedono un grado medio inferiore rispetto ai «non influencer». Questa evidenza suggerisce che un'elevata integrazione nella rete non si traduca automaticamente in capacità di leadership; al contrario, una connettività eccessiva sembra «diluire» l'efficacia del segnale di prezzo.

Mentre i nodi ad alto grado si concentrano tipicamente in aree urbane dense come Milano, in questi contesti la leadership di prezzo risulta frammentata: è statisticamente improbabile che un distributore riesca a influenzare simultaneamente una vasta platea di competitor (talvolta superiori a 50). In ambienti così saturi, il «rumore di fondo» generato dalle molteplici strategie di prezzo impedisce

l'emergere di un leader chiaro. Al contrario, stazioni situate in aree meno sature (come la periferia), con un numero ridotto di vicini (ad esempio quattro), esercitano un'influenza molto più potente. In questi cluster ristretti, i competitor possono monitorare e reagire ai segnali di prezzo in modo diretto e univoco, senza le interferenze tipiche di un mercato affollato.

Questi risultati indicano che il mercato lombardo non è guidato da grandi «direttori d'orchestra» capaci di comandare vaste aree, ma piuttosto da una moltitudine di team leader locali. La struttura del mercato appare dunque come un arcipelago di piccole «isole di influenza» indipendenti, piuttosto che un impero centralizzato. Tale frammentazione suggerisce che la coordinazione dei prezzi sia un fenomeno micro territoriale, dove la chiarezza del segnale prevale sulla portata della connettività.

4.1.2 Influencers e brand

Il mercato petrolifero italiano è storicamente dominato da grandi marchi nazionali e internazionali come Eni/Agip, Q8, IP, Esso e Tamoil. Tuttavia, dopo aver dimostrato che una centralità topologica elevata (grado alto) non è una condizione necessaria per esercitare influenza, è diventato prioritario analizzare come il ruolo di «influencer» sia distribuito tra i diversi attori del mercato.

Un'ampia porzione dell'offerta è infatti rappresentata dai rivenditori indipendenti, noti localmente come «Pompe Bianche». Questa categoria è estremamente eterogenea e spazia dalle catene della Grande Distribuzione Organizzata (GDO) (come Coop) a piccole stazioni a conduzione familiare.

Sebbene il dataset anagrafico descritto nella [Sezione 2](#) includa la dicitura «Pompe Bianche», la sua applicazione risulta parziale e disomogenea. Molti operatori indipendenti o minori, come ad esempio Keropetrol, vengono infatti catalogati singolarmente anziché essere aggregati sotto un'unica etichetta, generando una significativa frammentazione del campione. Per ovviare a tale dispersione e garantire la robustezza statistica dell'analisi, si è resa necessaria una riclassificazione secondaria che potesse in caso di necessità unificare questi attori sotto una comune categoria analitica.

Viene quindi definito come «Pompa Bianca» qualsiasi operatore con una quota di mercato inferiore al 5%. Tale soglia permette di aggregare i piccoli marchi indipendenti.

Brand minori inclusi nell'analisi: 4.0, AF Petroli, Auchan, BC, Beyfin, Bianco Carburanti, bpetrol, Carrefour, COIL, Costantin, Ego, EKOpoint, Enercoop, Ernesto Rondini, Eurfuel, Europam, GABOGAS2, Gamma Petroli, Garage Roberto, Giap, GNP, Iper Station, Itala Petroli, Italiana Carburanti, KEROPETROL, Kerotris, LEONE, Loro, LP Carburanti, Moro, Noaloil, OIL ONE, Oil Italia, OutletCarburanti, PAN OIL, Pit stop, Restiani, Retitalia, San Marco Petroli, Sarni Oil, Shell, Sia fuel, Socogas, Sommese Petroli, STOM, Tam service, Total Erg, Vega.

Al fine di determinare quali brand esercitino la maggiore pressione competitiva sul mercato, è stata analizzata la distribuzione degli «influencer» per ciascun marchio. I risultati sono presentati attraverso dei grafici a barre che riportano la percentuale di stazioni classificate come influencer (lag 1) sul totale dei punti vendita di ogni brand.

La prima visualizzazione adotta la riclassificazione precedentemente discussa, aggregando tutti i marchi minori sotto la categoria unificata delle «Pompe Bianche». La seconda, invece, offre un maggior grado di dettaglio mantenendo i singoli brand; in quest'ultimo caso, per evitare distorsioni statistiche dovute a campioni troppo esigui, sono stati esclusi dall'analisi i marchi che presentano un'unica stazione nel territorio analizzato.

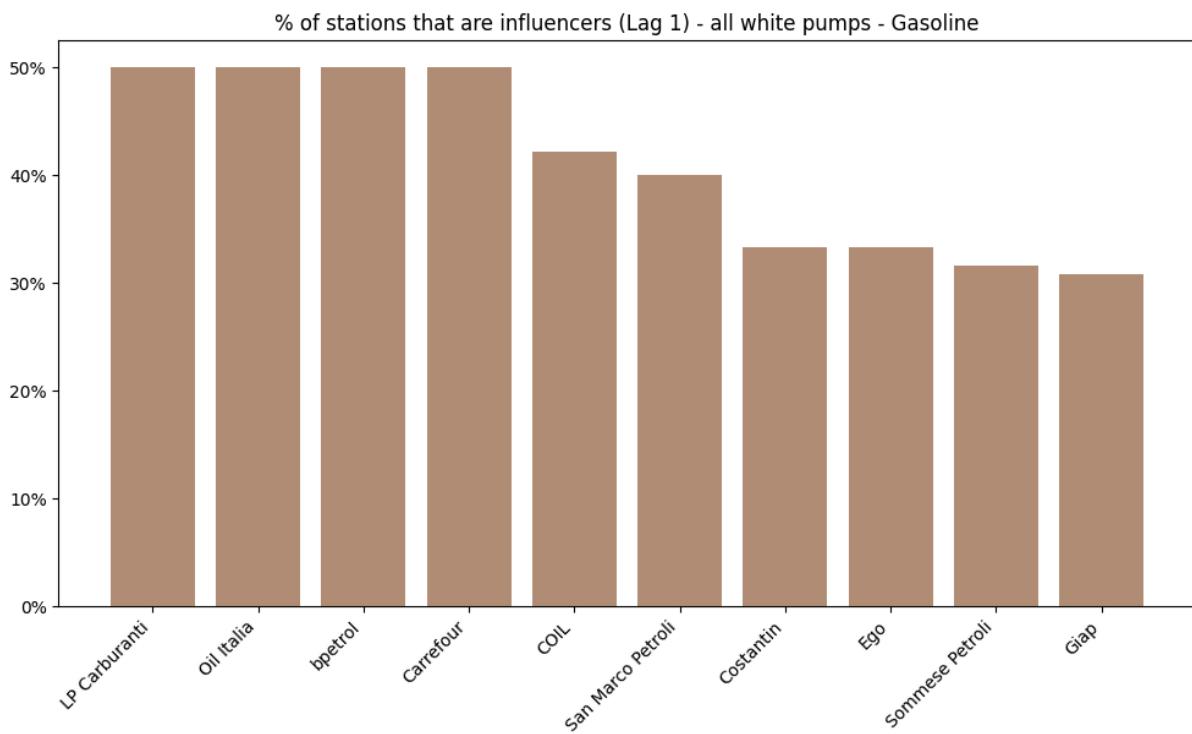
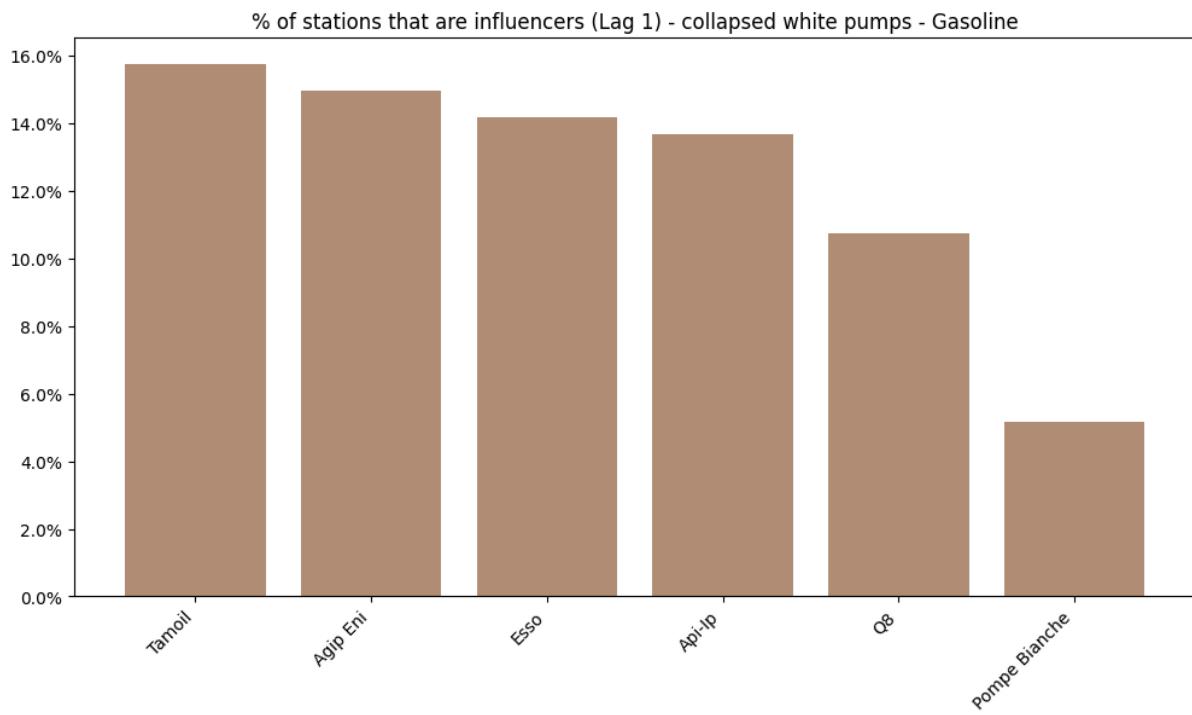


Figura 5: benzina, comparazione presenza «influencers» nei brand con pompe bianche unite e divise.

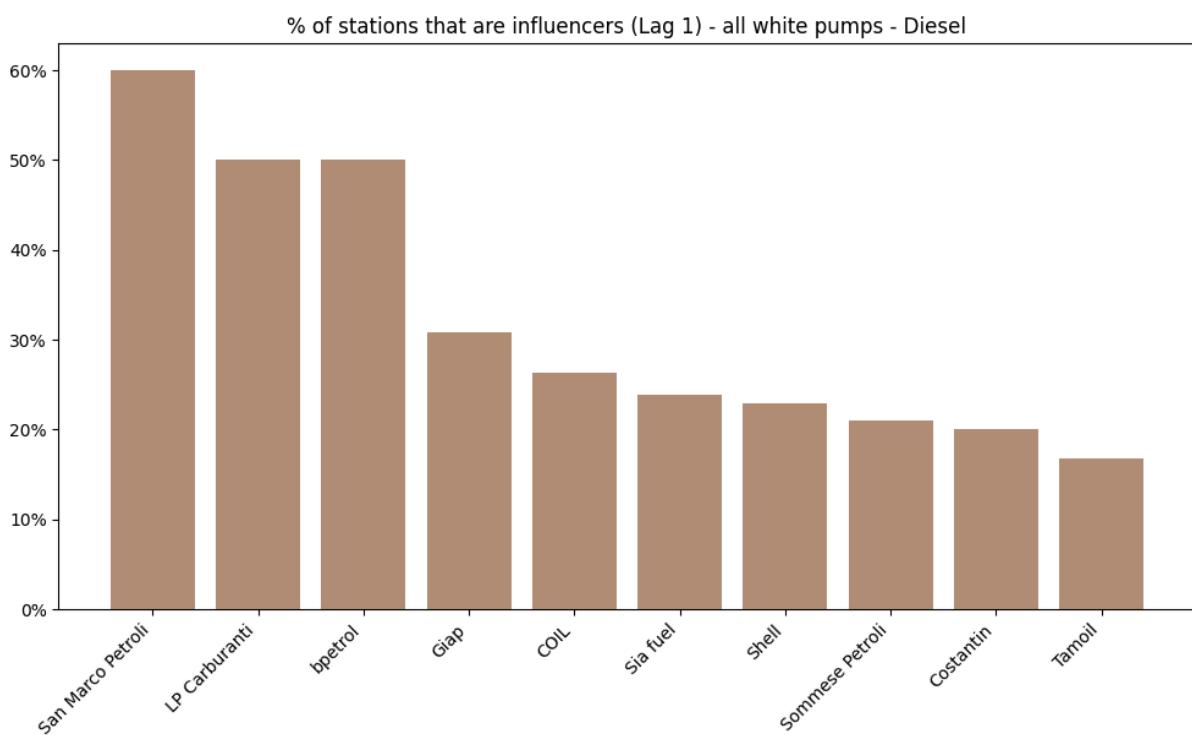
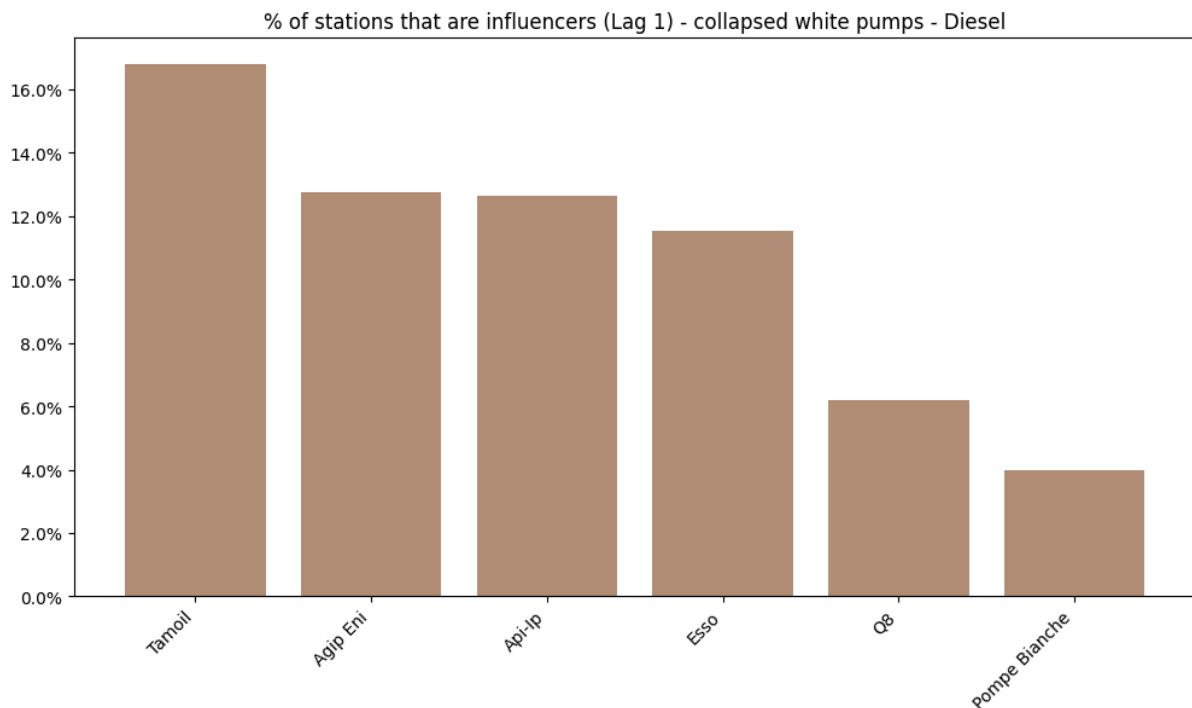


Figura 6: diesel, comparazione presenza «influencers» nei brand con pompe bianche unite e divise.

Sebbene nel primo grafico per entrambi i carburanti le Pompe Bianche mostrino collettivamente una percentuale di «influencer» tra le più basse rispetto ai grandi brand, tale evidenza è riconducibile all'elevata eterogeneità interna della categoria. L'accorpamento, infatti, somma realtà estremamente diverse, diluendo l'impatto di quei singoli operatori che agiscono come veri e propri «price disruptor».

Questa ambiguità viene risolta nel secondo grafico, dove l'analisi disaggregata rivela la reale potenza dei marchi indipendenti. Nel mercato della benzina, ad esempio, emerge il caso emblematico di LP Carburanti, dove il 50% delle stazioni è classificato come «influencer». In termini previsionali, ciò implica che l'apertura di un nuovo punto vendita di questo marchio avrebbe una probabilità del 50% di influenzare la micro rete circostante.

Tuttavia, coerentemente con quanto discusso nella [Sezione 4.1.1](#), tale capacità di influenza resta vincolata alla topologia della rete: la leadership si manifesta con successo solo laddove il numero di vicini non sia eccessivamente elevato. In contesti di iper concorrenza urbana, anche un operatore aggressivo vede il proprio segnale di prezzo saturato dal rumore degli altri competitor. In sintesi, mentre nel dato aggregato le «Pompe Bianche» non spiccano poiché molte di esse operano in zone ad alta competitività, l'analisi puntuale dimostra come, la capacità dei singoli brand di «Pompe Bianche» di dettare il prezzo sia superiore a quella di qualsiasi Major.

4.2 Cascate nelle reti 3km

Mentre l'analisi precedente si è focalizzata sull'influenza spaziale immediata all'interno del vicinato locale, in questa sezione ampliamo l'orizzonte analitico per esaminare la propagazione temporale. L'obiettivo è identificare i nodi «macro influencer»: stazioni capaci di innescare vere e proprie «cascate di influenza» attraverso la rete nel corso di più giorni, agendo come catalizzatori per spostamenti di mercato su scala più ampia. Nello studio sul mercato di Cuneo, gli autori ([Alderighi & Baudino, 2015](#)) introducono questo concetto per descrivere come una variazione di prezzo isolata possa innescare una reazione a catena tra i competitor chiamandolo «Domino effect».

Definiamo una cascata di prezzo come una sequenza di eventi di matching singolo a 1 giorno di lag, ordinata cronologicamente. Ad esempio, se la stazione *A* varia il proprio prezzo al tempo *t*, la stazione *B* risponde al tempo *t + 1* e la stazione *C* segue *B* al tempo *t + 2*, identifichiamo un percorso di propagazione diretto: *A → B → C*.

Significativamente, il modello permette l'esistenza di cicli temporali (ad esempio, *A → B → C → A* o *A → B → A*). Poiché ogni arco diretto viene trattato come un evento marcato temporalmente (time stamped), un collegamento da *A* verso *B* al tempo *t* è considerato un'unità analitica distinta rispetto allo stesso collegamento spaziale al tempo *t + 1*. Questo approccio dinamico permette di mappare non solo la gerarchia statica, ma il flusso continuo di informazioni e reazioni che attraversa il mercato.

Si precisa che, per coerenza metodologica, la definizione di matching (basata sullo stesso movimento di prezzo e sulla similarità del prezzo finale) è rimasta invariata rispetto a quanto stabilito nelle fasi precedenti, vedi [Sezione 3.1](#).

Una volta identificate tutte le cascate sono state prese in considerazione solo quelle caratterizzate da una profondità superiore ai due passaggi (ad esempio, *A → B → C → D*). Questo approccio permette di mappare la persistenza del segnale di prezzo nel tempo.

Per ciascun nodo della rete, viene calcolata la frequenza delle cascate con profondità superiore a due, misurando così la capacità di ogni stazione di agire come motore primario di una propagazione estesa nel tempo. Da qui in avanti verranno considerate esclusivamente le cascate con profondità superiore a due passaggi.

4.2.1 Cascate e grado

Come eseguito nella [Sezione 4.1.1](#), è interessante testare la correlazione tra il grado del nodo e l'innesto delle cascate. L'obiettivo primario è determinare se l'elevata connettività spaziale rappresenti un

preditore affidabile per la propagazione profonda dei prezzi.

Nello specifico, si intende indagare se le stazioni caratterizzate dal maggior numero di vicini siano anche sistematicamente responsabili dell'attivazione di cascate temporali. Questo passaggio è cruciale per verificare se il potere di mercato, inteso come leadership temporale, sia una proprietà emergente dei nodi più centrali della rete o se, coerentemente con quanto osservato per gli «influencer» di primo ordine, la capacità di innescare l'effetto domino risponda a logiche differenti dalla mera densità di connessioni locali.

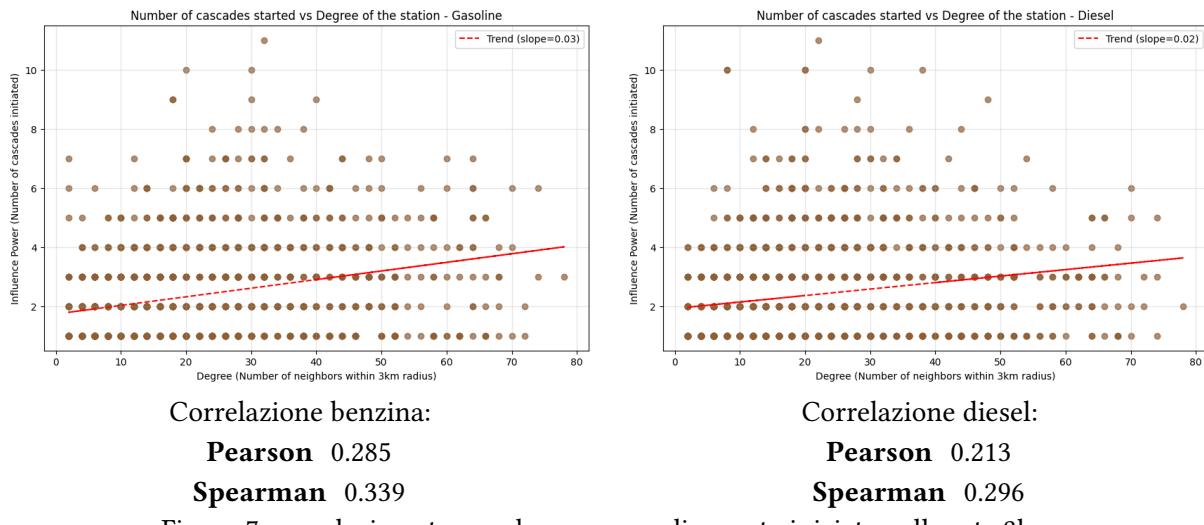


Figura 7: correlazione tra grado e numero di cascate iniziate nella rete 3km.

L'analisi della correlazione tra il grado del nodo e la capacità di innescare cascate temporali rivela una relazione positiva ma debole (pendenza della linea di tendenza pari a 0,03 per la benzina e 0,02 per il diesel).

Sebbene un numero maggiore di vicini offra, da un punto di vista puramente statistico, più opportunità di propagazione del prezzo, tale connettività non garantisce affatto la leadership di mercato. I coefficienti calcolati confermano questa dinamica: la correlazione di Pearson indica un legame lineare molto basso, mentre il coefficiente di Spearman, leggermente superiore, suggerisce una relazione monotonica. In altri termini, all'aumentare dei vicini, l'influenza tende generalmente a crescere, ma in modo non proporzionale e altamente imprevedibile.

I dati permettono di identificare due profili opposti che mettono in discussione la centralità geografica come unico driver:

I «giganti silenziosi» nodi caratterizzati da un grado estremamente elevato (70 – 80 vicini, tipici delle aree ad alta densità come il centro di Milano), che tuttavia mostrano un potere di influenza quasi nullo (1 o 2 cascate innescate). Questi distributori sono circondati da competitor, ma il loro segnale di prezzo viene ignorato dal mercato.

I «piccoli leader» nodi con grado medio basso (20 – 30), capaci però di innescare un numero elevato di cascate (10 – 11). Nonostante la platea ridotta, questi distributori godono di una leadership indiscussa nel proprio vicinato.

Questa divergenza, come già discusso nella [Sezione 4.1.1](#), è riconducibile al fenomeno del rumore competitivo. In aree ad alta densità (grado elevato), la presenza di troppi competitor diluisce l'attenzione: è improbabile che un numero elevato di stazioni segua simultaneamente lo stesso leader. Al contrario,

in aree meno sature, le relazioni competitive sono più dirette ed esclusive; un distributore con una strategia aggressiva in un'area periferica agisce come un punto di riferimento univoco, costringendo i pochi vicini a una reazione immediata.

In conclusione, la centralità spaziale (grado) è un predittore necessario ma non sufficiente per la leadership. Essere visibili a molti non equivale a essere generatrori di cascate.

4.2.2 Cascate e brand

Una volta accertato che la leadership temporale non è una mera funzione del grado del nodo, l'analisi si sposta sulla caratterizzazione dei brand responsabili della genesi delle cascate. Seguendo la metodologia applicata nella [Sezione 4.1.2](#), l'obiettivo è verificare se la capacità di innescare un effetto domino esteso sia prerogativa dei grandi player nazionali o se, al contrario, confermi la predominanza dei rivenditori indipendenti già osservata per l'influenza locale.

In questa fase, valuteremo se le cascate di prezzo abbiano origine prevalentemente da operatori indipendenti o se le Major conservino un ruolo di coordinamento nella propagazione dei segnali.

I grafici presentati sotto introducono una metrica inedita: l'efficienza.

In termini semplici, questa metrica risponde alla domanda: quanto è probabile che una singola stazione di un determinato marchio scateni un «terremoto» dei prezzi nel proprio vicinato?

$$\text{Efficienza} = \frac{\text{Numero totale di cascate innescate}}{\text{Numero di stazioni del brand}}$$

In altre parole, l'efficienza misura **quanto in media una stazione di un brand generi cascate**. Si tratta di un indicatore normalizzato che esprime la **propensione media** del singolo impianto a fungere da «market mover», depurando il dato assoluto dalla grandezza della rete. Questo permette di confrontare la pervasività strategica di un gigante come Eni con l'aggressività tattica di un piccolo operatore locale.

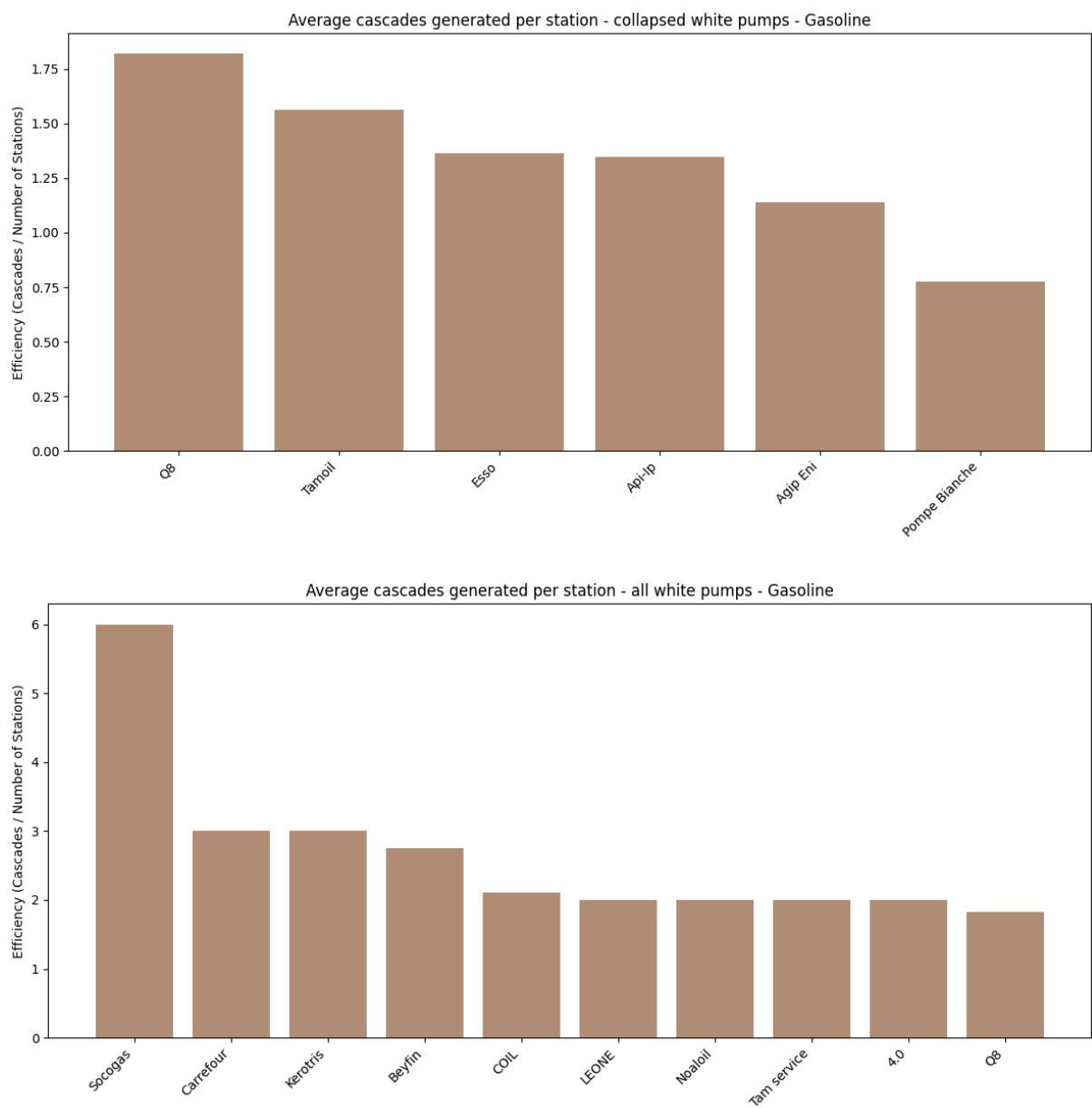


Figura 8: benzina, numero medio cascate iniziare per brand con pompe bianche unite e divise.

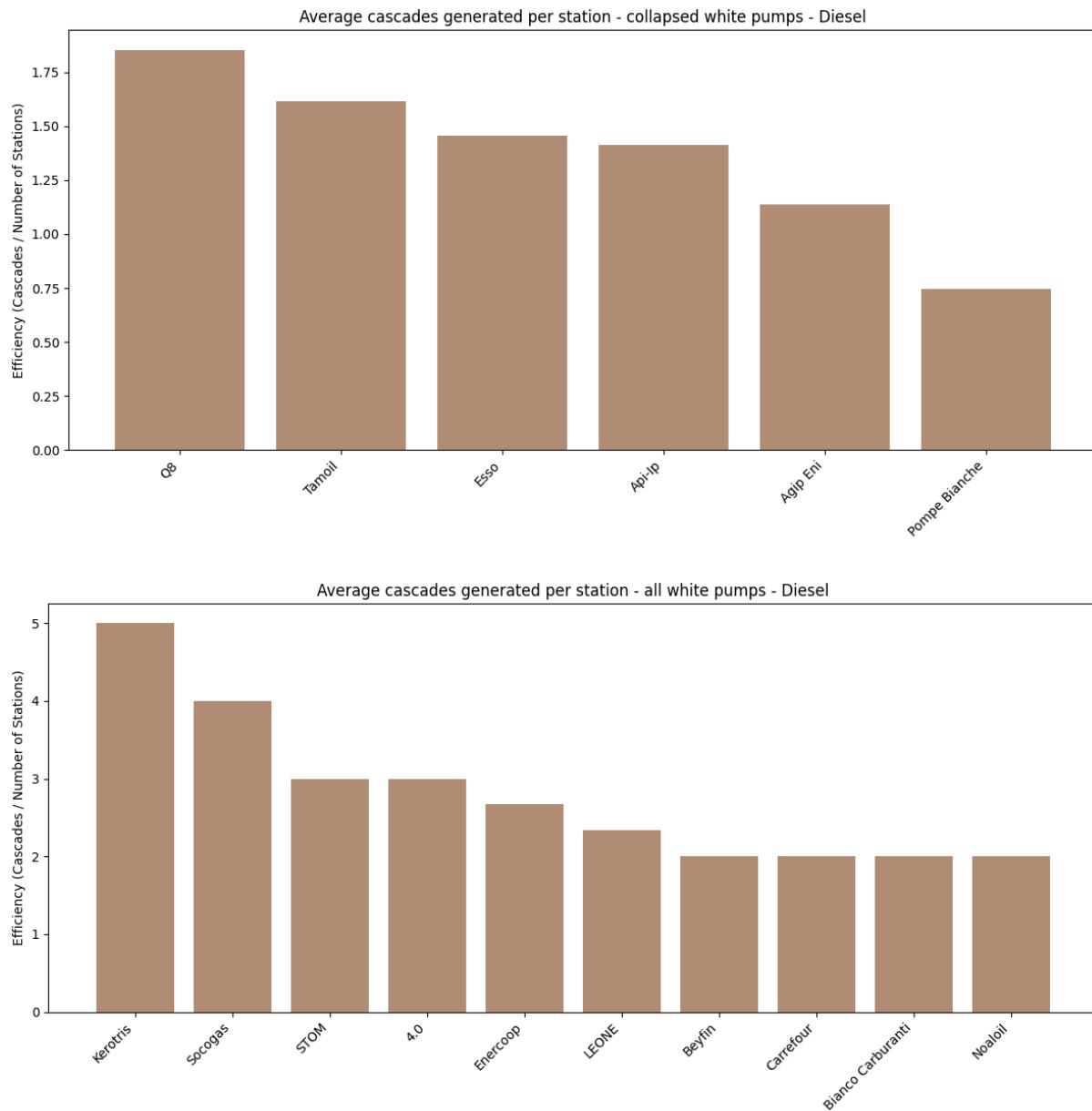


Figura 9: diesel, numero medio cascate iniziate per brand con pompe bianche unite e divise.

In linea con quanto emerso dall’analisi sugli «influencer», il primo grafico per entrambi i carburanti sembra confermare una netta predominanza dei grandi marchi nazionali. Nella visualizzazione aggregata, player come Q8 e Tamoil appaiono come i più efficienti del mercato, registrando un indice di circa 1,6 - 1,8 cascate innescate per singola stazione. Al contrario, la macro-categoria delle «Pompe Bianche» scivola in coda alla classifica, con un punteggio medio inferiore a 0,8.

Limitarsi a questa prima lettura porterebbe tuttavia a una conclusione fuorviante: si rischierebbe di dipingere le major come macchine da guerra perfettamente ottimizzate e gli indipendenti come attori numerosi ma singolarmente inefficaci. Come ritrovato anche nella Sezione 4.1.2 questa distorsione statistica è causata dall’intrinseca eterogeneità del gruppo «Pompe Bianche», nel quale le performance dei rivenditori più aggressivi vengono diluite e nascoste.

È nel secondo grafico, attraverso la disaggregazione dei dati, che emerge la reale dinamica di potere del mercato locale:

L'impatto di Socogas questo brand si rivela un vero e proprio «gigante» in termini di efficienza, raggiungendo un valore di 6,0. Significa che, mediamente, una singola stazione Socogas è in grado di innescare un numero di cascate tre volte superiore rispetto al «migliore» dei grandi marchi (Q8).

Il ruolo della GDO (Carrefour) e Kerotris entrambi i marchi si attestano su un valore di efficienza prossimo a 3,0, doppiando le performance delle compagnie petrolifere tradizionali. Il posizionamento di Carrefour, in particolare, conferma la teoria secondo cui le stazioni della grande distribuzione agiscono come ancora di prezzo: utilizzando il carburante come prodotto civetta per attrarre clienti nel punto vendita, adottano strategie estremamente aggressive che costringono l'intero vicinato all'adeguamento forzato.

Il ridimensionamento delle major significativo notare come Q8, che nel grafico aggregato appariva leader, scivoli nelle retrovie quando confrontato con i veri market movers locali.

Il confronto tra questi due scenari narra, in definitiva, una storia di qualità contro quantità. L'efficienza non è garantita dalla mera estensione della rete: i grandi brand come Eni o Q8 vedono il loro punteggio «diluito» da migliaia di stazioni che agiscono come follower passivi.

Sebbene le «Pompe Bianche» appaiano deboli come gruppo, i loro «campioni locali» sono i veri predatori del mercato. In termini di influenza dinamica, un singolo impianto Socogas esercita una pressione competitiva equivalente a quella di tre stazioni Q8.

4.2.3 Cascate e pattern temporali

Procediamo ora all'esame dell'eventuale presenza di pattern temporali nella generazione delle cascate. Si ricorda che, coerentemente con i criteri metodologici precedentemente stabiliti, questa analisi si focalizza esclusivamente sulle cascate con profondità superiore a due passaggi. Tale restrizione è fondamentale per garantire l'osservazione di eventi di diffusione dei prezzi realmente significativi, distinguendoli da semplici reazioni isolate o bidirezionali. L'obiettivo è verificare se la nascita di cascate segua una ciclicità specifica o sia influenzata da particolari finestre temporali durante il periodo di analisi.

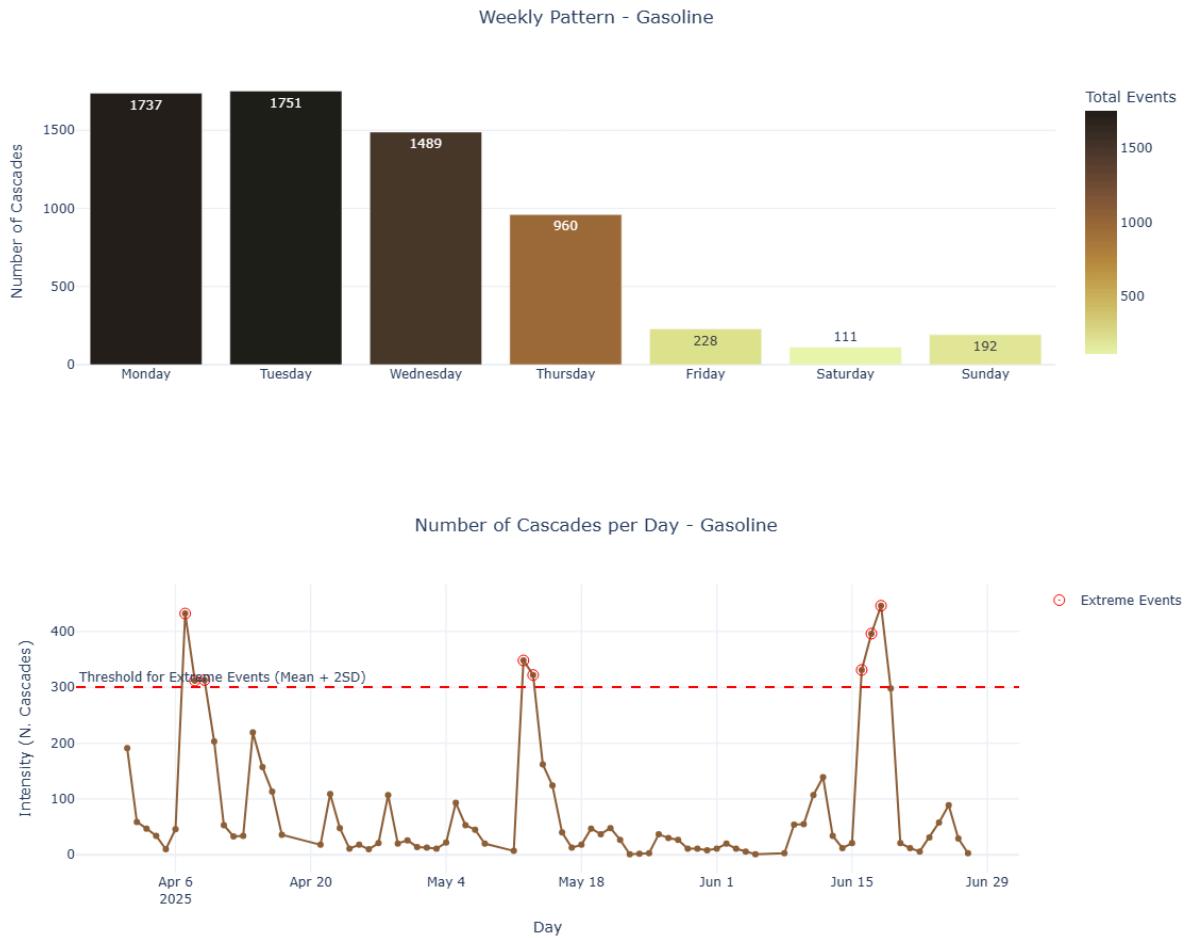


Figura 10: benzina, distribuzione settimanale e su singoli giorni nascita cascate

Il grafico a barre relativo alla distribuzione settimanale rivela una strategia di prezzo altamente sincronizzata e trasparente, tipica dei mercati con forte dipendenza dai segnali centrali.

Le cascate non si verificano in modo casuale. La stragrande maggioranza degli eventi di diffusione è concentrata tra lunedì (1.737) e martedì (1.751). Questo dato conferma una prassi consolidata del settore: i gestori ricevono le nuove quotazioni o le direttive centrali il lunedì mattina, procedendo all'adeguamento dei prezzi tra le prime 24 e 48 ore della settimana.

L'attività decresce progressivamente con il passare dei giorni. Se il mercoledì mantiene una discreta reattività (1.489), il giovedì segna l'inizio di un declino (960), fino a giungere a una fase di «ibernazione» nel fine settimana (venerdì-domenica), dove le cascate scendono sotto le 200 unità.

L'analisi della timeline descrive un mercato caratterizzato da periodi di relativa stabilità interrotti da improvvise accelerazioni. Abbiamo identificato tre momenti specifici in cui il sistema è entrato in una fase di forte agitazione (gli extreme events evidenziati in rosso):

Prima ondata (7-9 aprile) in picco massiccio (fino a 432 cascate in un solo giorno). Rappresenta probabilmente una reazione corale a uno shock rialzista dei prezzi del greggio o alla transizione stagionale primaverile. In questa finestra, quasi metà della rete si è mossa all'unisono.

Seconda ondata (12–13 maggio) un secondo picco (348 cascate). È interessante notare come avvenga a circa un mese di distanza dal primo, suggerendo la presenza di cicli di aggiustamento mensili o una risposta ritardata a precedenti fluttuazioni di mercato.

Terza ondata (16–18 giugno) il picco più alto e sostenuto (fino a 446 cascate). Collocandosi a metà giugno, questo evento coincide tipicamente con l'adeguamento delle tariffe in previsione dell'aumento della domanda estiva.

L'evidenza empirica suggerisce che il mercato non si trovi in uno stato di costante ebollizione. Al contrario, esso attraversa lunghi periodi di calma apparente interrotti da brevi esplosioni di attività sincronizzata. Quando una cascata ha inizio in questi periodi critici, raramente rimane un incidente isolato; si trasforma rapidamente in un fenomeno di «contagio di massa». Questo comportamento indica che i market mover identificati in precedenza non agiscono solo nello spazio, ma scelgono finestre temporali precise per emettere i loro segnali, massimizzando l'effetto domino.

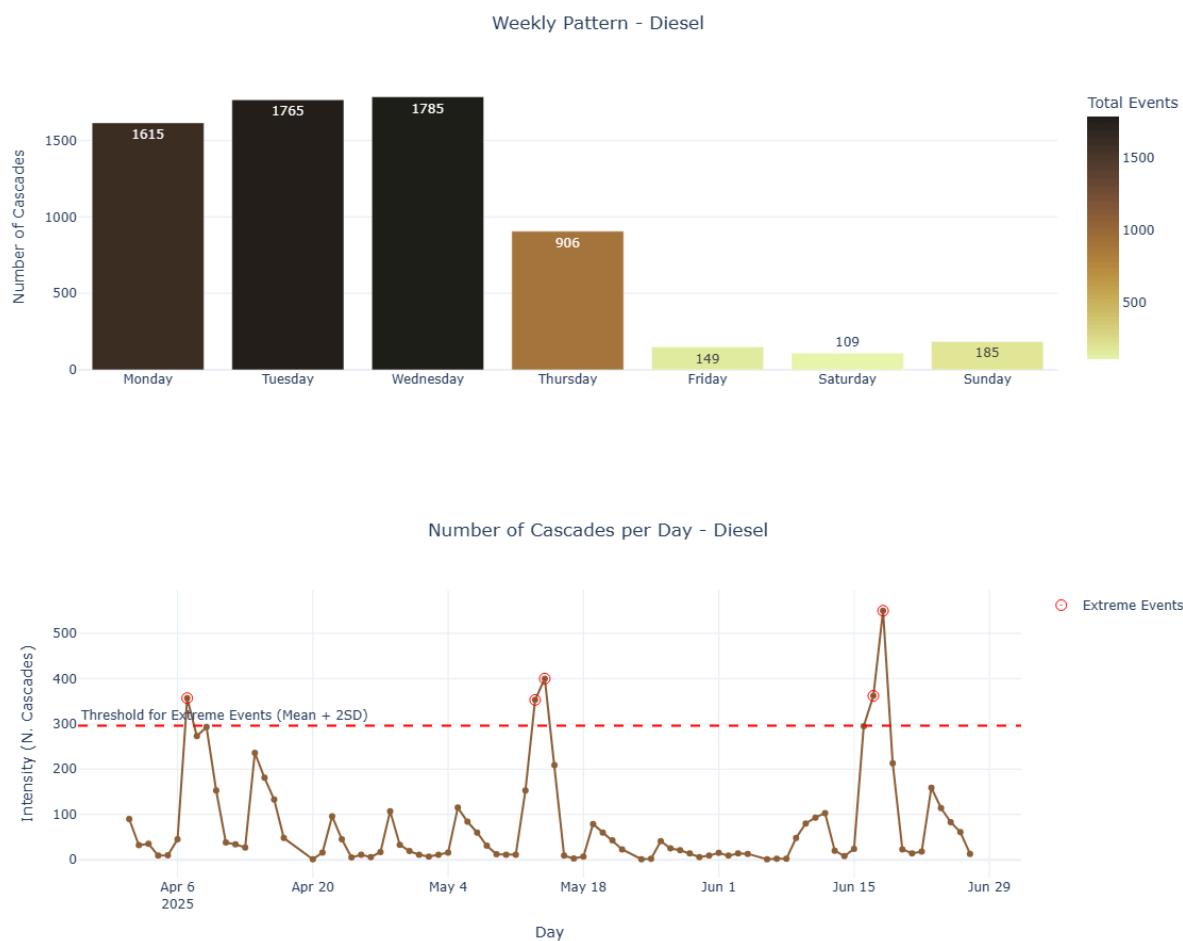


Figura 11: diesel, distribuzione settimanale e su singoli giorni nascita cascate

L'analisi temporale delle cascate profonde (profondità > 2) per il mercato del diesel rivela schemi straordinariamente simili a quelli osservati per la benzina. Tale sovrapposizione conferma che il mercato dei carburanti in Lombardia opera secondo una logica di sincronizzazione unificata, guidata da impulsi discreti piuttosto che da fluttuazioni continue.

Il grafico relativo alla distribuzione settimanale evidenzia una strategia di prezzo altamente coordinata, con una lieve ma significativa sfumatura rispetto alla benzina.

Analogamente alla benzina, l'attività di repricing è concentrata nei primi giorni della settimana. Tuttavia, il diesel mostra i volumi più alti il martedì (1.765 cascate) e il mercoledì (1.785), superando leggermente il lunedì (1.615). Questo spostamento verso il mercoledì suggerisce un potenziale «ritardo di adeguamento» per il diesel. I rivenditori potrebbero dare priorità all'aggiornamento dei prezzi della benzina (prodotto principale per il consumatore privato) per poi allineare il Diesel nelle 24 ore successive, oppure attendere che i segnali di prezzo per i trasporti commerciali si consolidino a metà settimana.

La timeline del diesel conferma anche'essa che il mercato non evolve in modo fluido, ma segue una dinamica di equilibrio con picchi: lunghi periodi di fluttuazione minima, interrotti da tre massicce onde di riposizionamento coordinato.

Prima ondata (6–8 aprile) un picco improvviso che raggiunge circa 360 cascate. Questo evento segna il primo grande aggiustamento sincronizzato del periodo analizzato.

Seconda ondata (12–14 maggio) una seconda ondata, più intensa, che sfiora le 400 cascate. La perfetta coincidenza temporale con il ciclo della benzina rafforza l'ipotesi di un comportamento di «reset» mensile.

Terza ondata (16–18 giugno) l'evento più violento dell'intero periodo, con un picco superiore alle 550 cascate. L'entità di questo «spike» suggerisce un effetto di «corsa all'adeguamento» che ha coinvolto simultaneamente quasi l'intera rete, probabilmente spinto dalle aspettative sulla domanda estiva o da una variazione critica nei costi del Platts diesel.

Il Diesel mostra una reattività ancor più esplosiva della benzina nei momenti di picco (550 vs 446 cascate nell'evento di giugno). Questo suggerisce che, sebbene il Diesel possa reagire con un leggero ritardo all'inizio della settimana, quando il segnale di prezzo viene finalmente trasmesso, la sua propagazione attraverso la rete è più virulenta. In conclusione, la leadership di prezzo nel comparto Diesel appare più concentrata e reattiva ai grandi volumi, confermando la natura del carburante come pilastro del settore logistico e commerciale.

4.2.4 Total impact, frequenza per profondità delle cascate

La prossima fase dell'indagine consiste nel mappare geograficamente le stazioni che esercitano la maggiore capacità di propagazione all'interno della rete. Per quantificare questo potere d'influenza, è stata sviluppata una metrica di sintesi denominata «Total Impact Score».

Tale punteggio viene calcolato sulla base della somma delle profondità di tutte le cascate valide (ovvero quelle con profondità superiore a 2) innescate da una singola stazione.

$$\text{Total Impact}_i = \sum_{j=1}^{N_i} D_{i,j} \mid D_{i,j} > 2$$

In questa formulazione, N_i rappresenta il numero totale di cascate iniziata dalla stazione i , mentre $D_{i,j}$ indica la profondità (ovvero il numero di step temporali) della j -esima cascata.

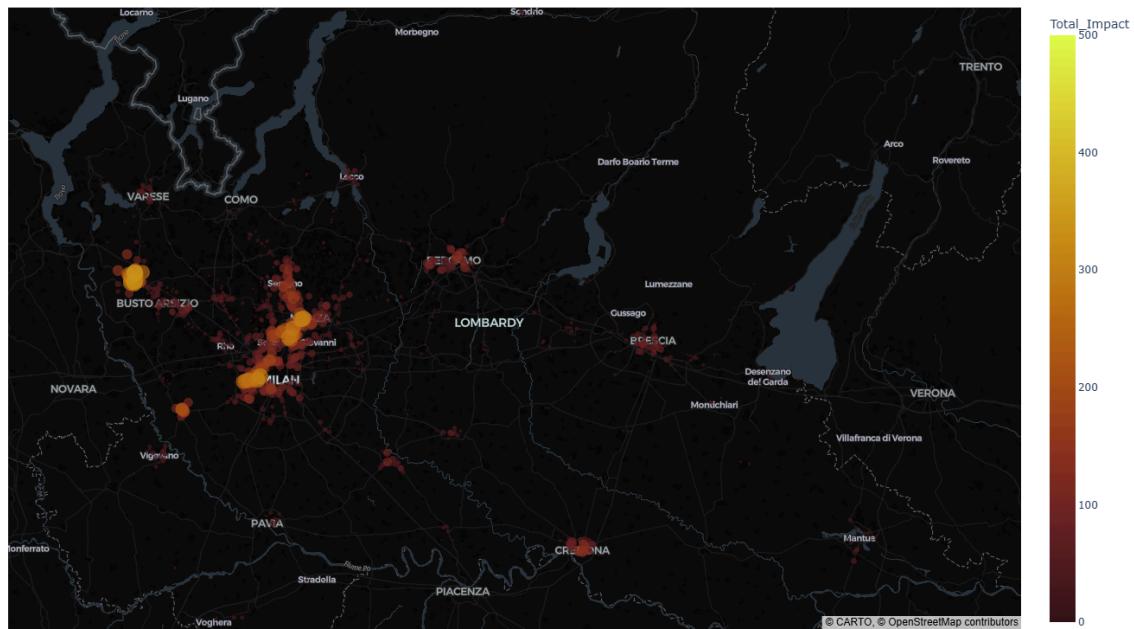
L'efficacia di questa metrica risiede nella sua capacità di integrare due dimensioni critiche:

- **La frequenza:** rappresentata dal limite superiore della sommatoria (N_i), ovvero quante volte la stazione agisce come catalizzatore.

- **La portata:** rappresentata dal valore di $D_{i,j}$, che misura quanto lontano si propaga la tendenza lungo la catena dei competitor prima di esaurirsi.

Combinando questi due fattori, il «Total Impact Score» ci permette di identificare le stazioni che non si limitano a influenzare il vicinato immediato, ma che danno il via a movimenti di mercato su vasta scala. I risultati sono presentati in una mappa dove la dimensione e il colore di ciascun punto sono proporzionali al punteggio cumulativo ottenuto. Questa visualizzazione consente di individuare immediatamente i cluster di influenza e di verificare se la leadership sia concentrata nei nodi urbani ad alta densità o se, come suggerito dalle analisi precedenti, esistano «centri di potere» isolati.

Total Impact Map: Frequency × depth of cascades initiated by stations - Gasoline



Total Impact Map: Frequency × depth of cascades initiated by stations - Diesel

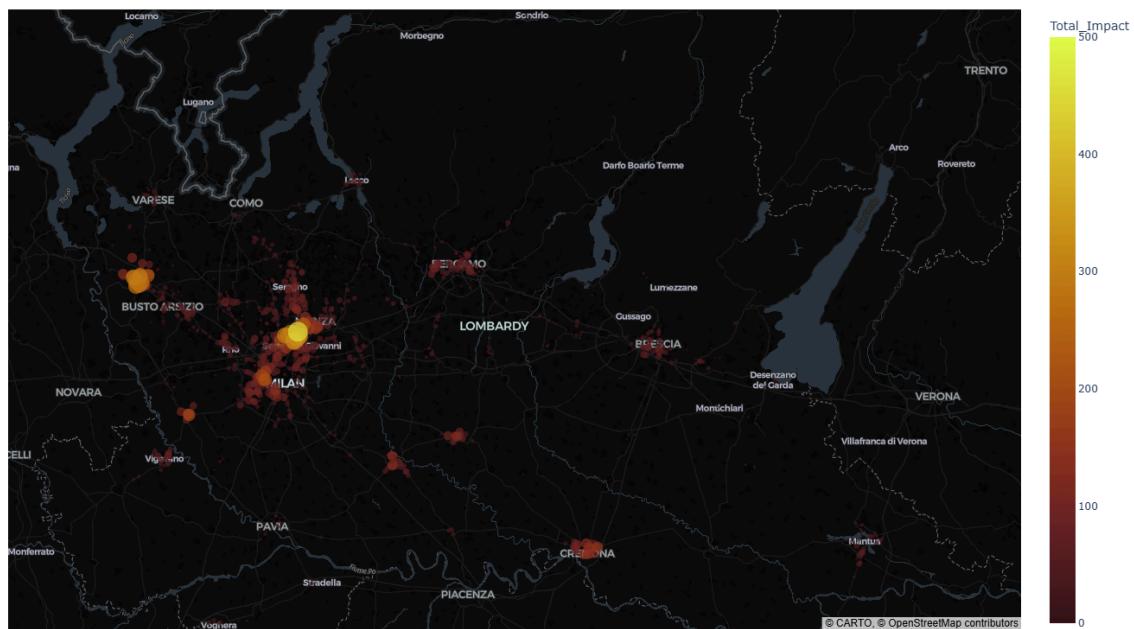


Figura 12: benzina sopra, diesel sotto, mappa raffigurante il «Total Impact».

Le mappe presentate illustrano la distribuzione geografica del «Total Impact Score» per le stazioni di rifornimento di benzina e diesel in Lombardia. Ogni punto rappresenta un distributore, e il colore (dal

rosso scuro al giallo brillante) unito alla dimensione del marcatore ne evidenzia la capacità di generare ampi effetti a catena sui prezzi.

Il dato più evidente in entrambe le mappe è che l'influenza di mercato non è distribuita in modo omogeneo sul territorio, ma è fortemente polarizzata. Gli epicentri assoluti del contagio (i cluster gialli di massima intensità) si concentrano in due macro-aree: il cuore dell'area metropolitana di Milano e il suo hinterland Nord-Ovest e l'asse strategico del Varesotto, in particolare il polo di Busto Arsizio / Gallarate.

Queste zone corrispondono ai bacini a massima densità abitativa, industriale e di pendolarismo della regione. L'elevata concentrazione di stazioni in questi spazi ristretti crea un ambiente molto competitivo: un abbassamento di prezzo in questi snodi nevralgici innesca una reazione a catena che si propaga rapidamente alle stazioni limitrofe. Al contrario, le vaste aree periferiche (sud della Lombardia e valli alpine) appaiono scure, indicando stazioni che operano come «price takers» o in regimi di micro monopolio locale, incapaci di innescare cascate profonde.

La mappa della benzina restituisce l'immagine tipica di un mercato orientato al consumatore finale e guidato da logiche di vicinato.

Non vi è un singolo nodo che raggiunge il picco estremo del diesel (i valori massimi si attestano tra i 300 e i 350, rimanendo su cromie arancio gialle). Tuttavia, i nodi ad alto impatto sono numericamente molto superiori e più distribuiti.

L'intera area di Milano, il suo vasto hinterland e tutto il corridoio verso Varese sono accessi da molteplici epicentri di media-alta intensità. La competizione qui si combatte «strada per strada»: ci sono numerosi leader locali in grado di innescare guerre di prezzo di quartiere o di provincia, creando una rete di contagio molto più frammentata, reattiva e complessa da prevedere rispetto a quella del diesel.

Mentre osservando la mappa del diesel, il mercato appare governato da una logica altamente centralizzata. Il singolo nodo con la massima capacità di innesco in assoluto (il grande marcatore giallo brillante che sfiora la quota 450-500) si trova a nord dell'area metropolitana di Milano, precisamente a Cinisello Balsamo.

Al di fuori di questo «super influencer» centrale e di un cluster secondario sull'asse nord-ovest (Varese/ Busto Arsizio), il resto del territorio appare molto scuro. Questo indica che la stragrande maggioranza dei distributori subisce il prezzo in modo passivo. Essendo un mercato fortemente legato alla logistica e alle flotte commerciali, bastano pochissimi snodi nevralgici ad altissimo volume per dettare la linea di prezzo a decine di stazioni subordinate.

Per concludere questa sezione è stata eseguita un'analisi comparativa tra i primi 100 nodi classificati per «Total Impact» e i leader identificati nella [Sezione 4.1](#) precedente «1_Influencer» rivela una sovrapposizione minima tra i due gruppi, 16 su 100 tra i principali generatori di cascate sono «influencer» per il mercato della benzina mentre il numero si abbassa a 3 su 100 per il diesel . Questa divergenza suggerisce che la leadership di prezzo risponda a meccanismi strutturali distinti all'interno della rete: mentre alcuni nodi sono specializzati nel facilitare una portata globale innescando cascate, altri esercitano il proprio potere esclusivamente attraverso un coinvolgimento locale diretto.

La mancanza di una correlazione significativa indica che queste due forme di potere sistematico sono funzionalmente indipendenti. In altre parole, la capacità di condizionare immediatamente i propri vicini non garantisce la forza necessaria per agire come catalizzatore di un «effetto domino» più ampio. Tale evidenza empirica suggerisce che il mercato lombardo sia governato da una doppia gerarchia:

una tattica, basata sulla reattività spaziale, e una strategica, definita dalla capacità di propagazione temporale.

5 Analisi reti non orientate distanza 3km

Dopo aver analizzato i flussi direzionali dell'influenza e le gerarchie di innesto, l'indagine procede con la trasformazione della rete dei 3km da diretta a non diretta, perdendo momentaneamente attributi relativi alle dinamiche di matching (frequenza e lag temporale). Questo passaggio metodologico è fondamentale per spostare il focus dai singoli rapporti di causalità («chi influenza chi») verso una comprensione più ampia della connettività strutturale sottostante e dei pattern comunitari che definiscono il mercato.

Rimuovendo la direzionalità, è possibile far emergere l'architettura latente delle relazioni competitive, identificando cluster di stazioni che operano in regimi di interdipendenza reciproca. Tale approccio permette di mappare le «comunità di prezzo» non solo sulla base della vicinanza geografica, ma come veri e propri ecosistemi coesi, dove la fitta rete di scambi riflette la solidità dei legami economici e la stabilità delle dinamiche di confronto nel mercato lombardo. Per il momento, dato che in precedenza, ogni coppia di nodi aveva due archi che andavano in entrambe le direzioni e su ogni arco vi erano gli attributi di matching, trasformando la rete in non diretta queste informazioni vengono momentaneamente perse e gli archi non avranno peso.

5.1 Distribuzione del grado

Per caratterizzare la natura strutturale della rete (la quale non integrando: prezzi, influencer e matching risulta la medesima per benzina e diesel), è stata analizzata la distribuzione del grado dei nodi, confrontando i dati empirici con un modello teorico basato sulla legge di potenza. Come illustrato nel lavoro di ([Barabási & Bonabeau, 2003](#)), in una rete scale free, la maggior parte dei nodi presenta pochi collegamenti, mentre una piccola frazione (gli hub) ne possiede un numero estremamente elevato.

Nello specifico, è stato testato l'adattamento dei dati empirici a una distribuzione a legge di potenza (power law), avvalendosi della libreria analitica powerlaw in Python.

L'algoritmo di stima è stato configurato per il trattamento di variabili discrete (discrete=True), impostando la soglia minima di valutazione a un singolo grado di connessione (xmin=1). La validazione del modello è stata effettuata sia a livello statistico che visivo, tracciando la Funzione di Densità di Probabilità (PDF) su un piano cartesiano a scala bi-logaritmica (log-log plot). In tale rappresentazione, la sovrapposizione tra la curva dei dati empirici e la retta teorica del modello permette di verificare la presenza di un'architettura di tipo scale-free (a invarianza di scala).

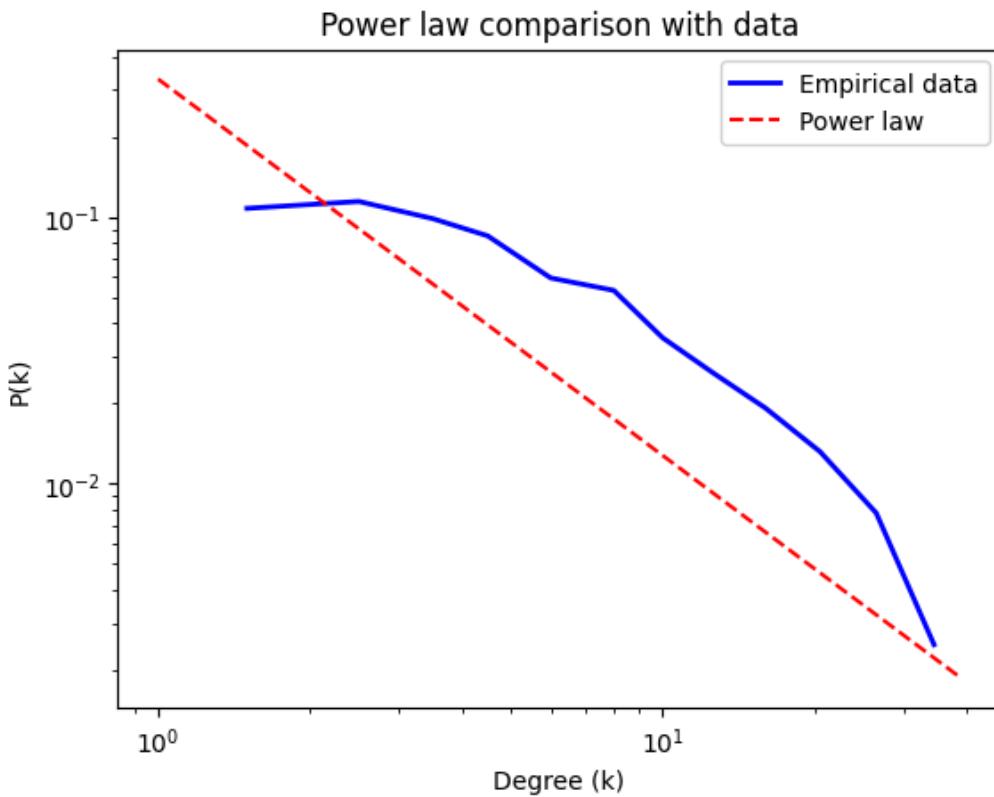


Figura 13: analisi della distribuzione del grado

L’osservazione del grafico rivela una chiara deviazione dal modello teorico. La presenza di una marcata «gobba» nella distribuzione empirica suggerisce che i dati non seguano una legge di potenza.

Va specificato che la semplice osservazione visiva di una retta su un grafico bi-logaritmico non costituisce una prova sufficiente per decretare l’appartenenza di una rete alla classe scale-free, in quanto altre distribuzioni a coda pesante (heavy-tailed) possono generare pattern grafici affini. Per confermare la robustezza del modello, l’adattamento della legge di potenza è stato sottoposto a un test di validazione statistica mediante il calcolo del rapporto di verosimiglianza (Log-Likelihood Ratio test, R), affiancato dal relativo p-value per testarne la significatività.

La legge di potenza è stata comparata a coppie contro due distribuzioni alternative di riferimento: la distribuzione Esponenziale (tipica delle reti spaziali e fisiche, dove la probabilità di trovare nodi con un grado molto elevato decresce molto rapidamente in modo più che lineare) e la distribuzione Lognormale (spesso osservata in sistemi biologici e sociali dove i processi di crescita sono moltiplicativi, ma i valori estremi sono più rari rispetto a una power law.). L’analisi ha valutato analiticamente quale funzione matematica minimizzasse la perdita di informazione nel descrivere i dati empirici del mercato lombardo. Oltre all’esito numerico del test statistico, le funzioni di densità di probabilità (PDF) delle tre distribuzioni teoriche sono state sovrapposte alla distribuzione empirica dei gradi, permettendo una valutazione ispettiva diretta.

Power law comparison with data and lognormal and exponential

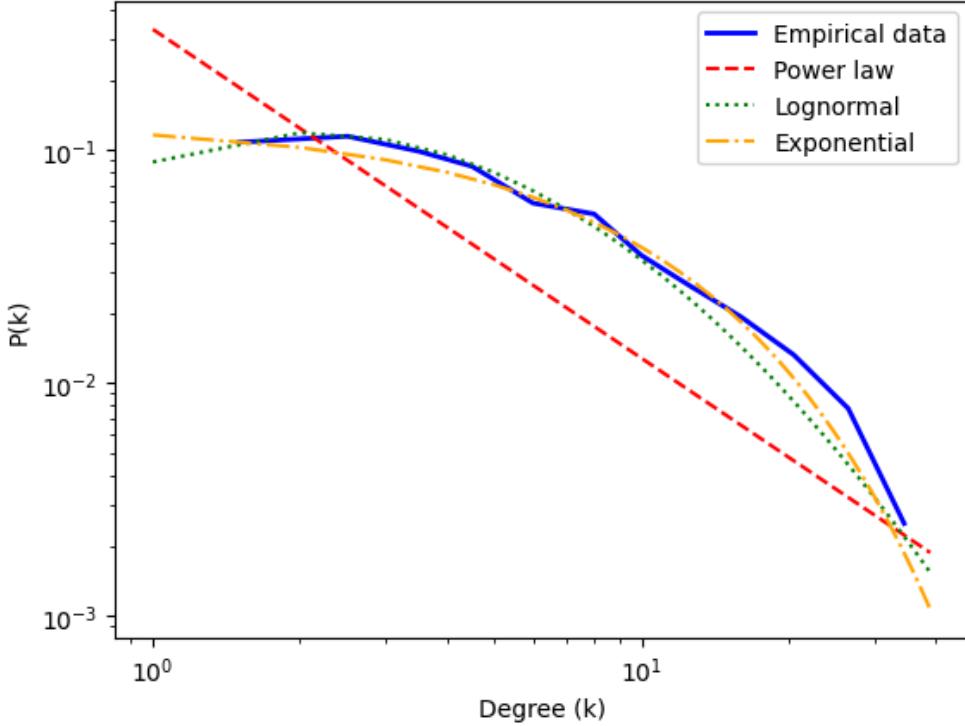


Figura 14: comparazione, dati empirici con varie distribuzioni:

power law vs exponential R=-1119.113, p=0.000

power law vs lognormal R=-1051.005, p=0.000

lognormal vs exponential R=-68.108, p=0.000

La distribuzione esponenziale sembra essere quella che meglio descrive i dati empirici.

L'ipotesi di una distribuzione esponenziale è stata verificata analiticamente confrontando i momenti della distribuzione. Dato un campione di gradi k , è stato stimato il parametro λ attraverso la media campionaria, essendo quest'ultima uno stimatore non distorto del valore atteso:

$$E[X] = \bar{x} = \frac{1}{\lambda} \Rightarrow \lambda = \frac{1}{\bar{x}}$$

In una distribuzione esponenziale, la varianza teorica σ_{th}^2 è strettamente legata al parametro λ :

$$\sigma_{\text{th}}^2 = \frac{1}{\lambda^2}$$

Il test di conferma consiste nel verificare lo scostamento tra la varianza teorica σ_{th}^2 e la varianza empirica calcolata sui dati σ_{emp}^2 :

$$\sigma_{\text{emp}}^2 = \frac{1}{n-1} \sum_{i=1}^n (k_i - \bar{x})^2$$

La quasi totale coincidenza tra σ_{th}^2 e σ_{emp}^2 osservata nei dati analizzati valida definitivamente il modello esponenziale per la topologia della rete ($\sigma_{\text{emp}}^2 - \sigma_{\text{th}}^2 \approx 1$).

La validazione del modello esponenziale rispetto a quello scale-free permette di trarre ulteriori conclusioni sulla natura del mercato dei carburanti in Lombardia:

Rigetto dell'ipotesi scale free la distribuzione del grado conferma che la rete non è dominata da pochi hub iper connessi. Al contrario, la connettività decade rapidamente, rendendo statisticamente improbabile l'esistenza di nodi con un numero di vicini fuori scala.

Determinismo geografico questo risultato è coerente con la natura spaziale della rete, governata dal vincolo geometrico dei 3 km. La topologia del mercato non è il risultato di una crescita preferenziale astratta, ma è rigidamente limitata dalla densità fisica dei punti vendita sul territorio.

Assenza di hub globali La struttura esponenziale conferma l'assenza di stazioni capaci di influenzare l'intero sistema regionale in un unico passaggio. L'influenza è fisicamente limitata, nessuna stazione può raggiungere l'intera regione; il segnale di prezzo deve necessariamente propagarsi attraverso una catena di intermediari locali.

Leadership Distribuita la topologia suggerisce un mercato composto da cluster locali omogenei.

Questo supporta ulteriormente l'evidenza che la leadership di prezzo è un fenomeno distribuito e atomizzato, piuttosto che centralizzato. La competizione si configura quindi come una serie di ecosistemi locali interconnessi, dove il potere di mercato è una funzione della posizione strategica all'interno del proprio cluster e non di una centralità assoluta nella rete.

5.2 Prezzi mediani e nuovo matching

Al fine di ottenere un valore sintetico rappresentativo per ogni nodo all'interno del dataset, è stato calcolato il prezzo mediano basato sullo storico delle osservazioni.

La scelta della mediana, in luogo della media aritmetica, è motivata dalla natura stessa della serie temporale analizzata (composta da 91 osservazioni per ciascun punto vendita). In un dataset di tale ampiezza, la media risulta eccessivamente sensibile a eventuali fluttuazioni isolate, errori di inserimento o shock di prezzo temporanei che non riflettono la strategia di posizionamento ordinaria della stazione.

La mediana permette di:

Mitigare l'effetto degli outlier neutralizzando l'impatto di picchi di prezzo anomali.

Identificare il «prezzo tipico» estraendo il valore che meglio rappresenta il comportamento di mercato del distributore nel periodo osservato, garantendo una base stabile.

Partendo dalla [Sezione 5](#) fino a questo punto, l'analisi sulla rete dei 3km sia per benzina che per diesel ha considerato il grafo come non diretto e non pesato, focalizzandosi esclusivamente sulla topologia della connettività. Tuttavia, per calcolare metriche di rete più avanzate e riflettere l'intensità delle interazioni, è stato reintrodotto un sistema di pesatura degli archi basato sulla frequenza di matching illustrata nella [Sezione 3.1](#). Il sistema di pesatura adottato integra la dinamica di reazione osservata nella rete orientata dei 3km. Per ogni coppia di nodi, il peso dell'arco non diretto viene calcolato come la media bidirezionale dell'attributo «*x*-matching», dove l'indice *x* rappresenta il lag temporale (da 0 a 5 giorni), questa procedura permette di condensare le informazioni del grafo diretto in una struttura non orientata, mantenendo però l'informazione sull'intensità relazionale.

5.3 Correlazione tra prezzi e centralità

Prendendo come riferimento il recente studio di ([Fronzetti Colladon et al., 2024](#)), nel quale le metriche di centralità proprie della network science (inclusa la misura di Distinctiveness) vengono impiegate per spiegare le variazioni dei prezzi al consumo, si è proceduto a un'analisi analoga applicata al mercato

lombardo.

Dato che alcune metriche di centralità necessitano di avere archi pesati, tramite la metodologia descritta nella [Sezione 5.2](#), ad ogni arco è stato assegnato un peso basato sull'attributo «`1_matching`» derivato dalla rete orientata dei 3km. Tale scelta permette di ponderare le connessioni in base all'intensità della risposta competitiva a breve termine (lag di un giorno).

Analizzando la correlazione tra prezzo mediano e centralità ci si pone come obiettivo di determinare se i «leader» strutturali della rete (ovvero i nodi più centrali o meglio connessi) tendano a mantenere prezzi più alti, sfruttando una posizione di dominio locale, o se al contrario la centralità sia associata a una maggiore pressione competitiva e, di conseguenza, a prezzi più contenuti.

Metrica	Spearman r	p -value	Pearson r	p -value
Benzina				
Degree	-0.175	0.000	-0.022	0.249
Betweenness	-0.161	0.000	-0.024	0.209
Closeness	-0.174	0.000	-0.031	0.115
Eigenvector (non pesata)	-0.180	0.000	-0.077	0.000
Eigenvector (pesata)	-0.157	0.000	0.011	0.580
Laplacian	-0.275	0.000	-0.170	0.000
Distinctiveness	-0.283	0.000	-0.252	0.000
Diesel				
Degree	-0.060	0.002	0.067	0.001
Betweenness	-0.036	0.060	0.017	0.383
Closeness	-0.029	0.141	0.075	0.000
Eigenvector (non pesata)	-0.035	0.073	-0.045	0.020
Eigenvector (pesata)	-0.033	0.088	-0.090	0.000
Laplacian	-0.157	0.000	-0.109	0.000
Distinctiveness	-0.178	0.000	-0.184	0.000

Tabella 1: Correlazione tra metriche di centralità e prezzo mediano per benzina e diesel.

I risultati mostrati in tabella fanno emergere un pattern strutturale coerente e di grande interesse analitico: esiste una relazione inversamente proporzionale tra la centralità topologica di una stazione e il suo livello di prezzo. In entrambi i mercati (benzina e diesel), i coefficienti di correlazione di Spearman risultano costantemente negativi. Ciò significa che all'aumentare della centralità di un nodo indipendentemente dal fatto che venga misurata tramite Degree, Closeness o Eigenvector il suo prezzo mediano tende a diminuire.

Da questa evidenza empirica derivano tre considerazioni fondamentali sulle dinamiche del mercato lombardo:

1. **L'effetto «sconto competitivo»:** la correlazione negativa quantifica quello che potremmo definire uno «sconto competitivo». Le stazioni situate in aree ad alta densità di connessioni (quartieri ad alto Degree) sono sottoposte a una severa pressione concorrenziale. Per sopravvivere in un ecosistema in cui il consumatore ha a disposizione numerose alternative a distanze minime, i distributori sono costretti ad abbattere i margini e ridurre i prezzi. Al contrario, i nodi periferici o topologicamente

isolati (a bassa centralità) godono di rendite di posizione assimilabili a micro monopoli locali, che consentono loro di mantenere prezzi sistematicamente più elevati.

2. **Il ruolo chiave della Laplacian e Distinctiveness centrality:** il dato più rilevante risiede nell'entità delle correlazioni, i valori negativi più forti si registrano in corrispondenza della Laplacian centrality ($r \approx -0.275$ per la benzina) e della Distinctiveness centrality. A differenza del semplice Degree (che conta solo il numero di vicini), queste metriche avanzate premiano i nodi che fungono da «ponti» tra diverse comunità o che occupano posizioni strutturalmente uniche nel grafo. Questo implica che le stazioni con il posizionamento spaziale più strategico i veri e propri hub di snodo del mercato sono anche i più aggressivi nelle politiche di pricing.
3. **Divergenza tra i mercati benzina vs diesel:** l'effetto di svalutazione competitiva è significativamente più marcato nel mercato della benzina (r compreso tra -0.17 e -0.28) rispetto a quello del Diesel (r compreso tra -0.06 e -0.18). Questa divergenza conferma strutturalmente la diversa natura dei due carburanti, il mercato della benzina è orientato al consumatore finale ed è estremamente sensibile alla concorrenza spaziale e di vicinato. Il mercato del diesel, al contrario, appare meno vincolato alla geografia di prossimità, essendo guidato in gran parte da logiche di logistica pesante e contrattualistica aziendale (flotte commerciali, accordi B2B), dove il prezzo alla pompa esposto al pubblico gioca un ruolo marginale.

Il confronto tra l'evidenza empirica qui emersa e il lavoro di (Fronzetti Colladon et al., 2024), rivela un'interessante inversione di paradigma.

Nel loro studio, metriche di centralità come la Betweenness o la Distinctiveness sono state associate a livelli di prezzo più elevati. La spiegazione teorica risiedeva nel ruolo di «ponte» strutturale di tali nodi, stazioni con alta centralità venivano identificate come punti di raccordo verso aree periferiche o segmenti di mercato isolati. In quel contesto, la posizione strategica garantiva una sorta di monopolio o duopolio naturale, permettendo ai distributori di sfruttare la scarsità di alternative locali per applicare prezzi premium.

Tuttavia, l'analisi del mercato lombardo basata su un vincolo spaziale di 3km, sensibilmente più stringente rispetto ai 10 km adottati (Fronzetti Colladon et al., 2024) delinea un ecosistema radicalmente differente:

1. **Dinamica dei ponti iper competitivi:** In una regione ad elevata densità abitativa e saturazione di punti vendita, i nodi con elevata Laplacian e Distinctiveness centrality non collegano zone isolate, ma fungono da interfacce critiche tra cluster già iper competitivi.
2. **Il «fuoco incrociato» del prezzo:** invece di godere di una rendita di posizione, il «leader strutturale» della rete lombarda si trova al centro di molteplici bacini di utenza sovrapposti. La sua centralità lo espone contemporaneamente alle strategie aggressive di diversi gruppi di competitor, trasformando la posizione di rilievo topologico in una vulnerabilità commerciale.
3. **Lo sconto competitivo:** questo fenomeno giustifica la correlazione negativa osservata: per mantenere la propria rilevanza in un punto di snodo così trafficato e monitorato, il distributore è costretto ad agire come price-taker o ad adottare politiche di prezzo aggressive, subendo quello che abbiamo definito «sconto competitivo».

In sintesi, mentre nella letteratura considerata la centralità è sinonimo di potere di mercato, nella fitta rete dei distributori lombardi essa si configura come un onere competitivo, dove l'iper connessione agisce da catalizzatore per il ribasso dei prezzi.

5.4 Community detection, Louvain tramite ANOVA

Questa fase della ricerca mira a identificare comunità di prezzo distinte all'interno della rete, superando la semplificazione dei confini amministrativi (comuni o province) come definizione predefinita di mercato. L'obiettivo è permettere ai dati stessi di rivelare i reali confini economici attraverso un ciclo di ottimizzazione statistica che combina la teoria delle reti con l'analisi della varianza (ANOVA).

Il raggruppamento dei punti vendita è affidato all'algoritmo di Louvain ([Blondel et al., 2008](#)), una tecnica di ottimizzazione della modularità che identifica cluster di nodi densamente interconnessi. L'algoritmo include un parametro cruciale denominato risoluzione:

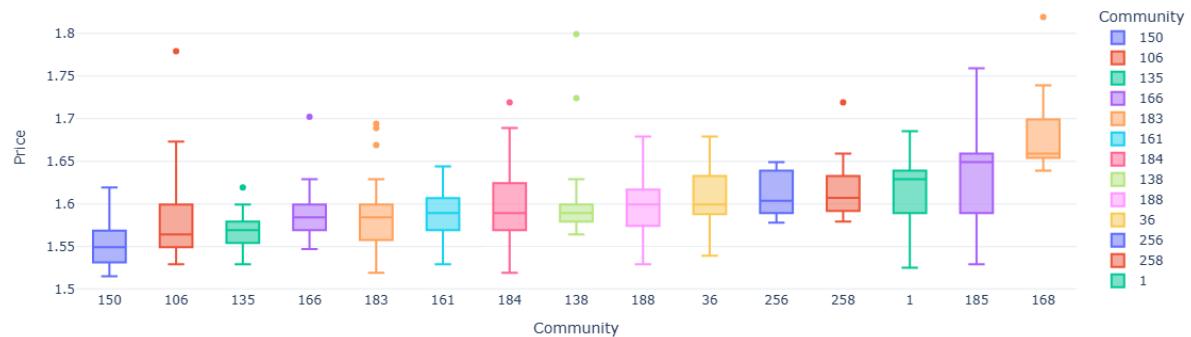
- Una **bassa risoluzione** produce poche comunità di grandi dimensioni (visione macroscopica).
- Una **alta risoluzione** frammenta la rete in numerosi piccoli distretti locali (visione microscopica).

Per determinare il livello di granularità ottimale per il mercato lombardo, è stata introdotta una metrica di controllo statistico: la statistica F di ANOVA a una via. Per ogni possibile configurazione delle comunità, l'algoritmo valuta la coerenza interna dei prezzi:

- Un valore **F basso** indica comunità statisticamente irrilevanti, dove i prezzi sono distribuiti in modo eterogeneo e sovrapponibile tra i gruppi.
- Un valore **F elevato** segnala l'identificazione di vere e proprie «Zone di Prezzo» distinte, caratterizzate da una bassa varianza interna e un'alta varianza tra i gruppi (es. un cluster economico contrapposto a uno premium).

Il codice esegue una simulazione iterativa, modulando gradualmente la risoluzione alla ricerca del valore che massimizzi il segnale statistico (F-statistic). Per evitare l'overfitting e garantire l'efficienza computazionale, è stato implementato un meccanismo di early stopping che interrompe la ricerca una volta raggiunto il picco di separazione tra le zone di prezzo. Una volta terminato l'algoritmo assegna come attributo a ogni nodo della rete un numero che corrisponde alla community per lui individuata. I risultati sono stati validati visivamente tramite l'analisi dei boxplot. Questi grafici ordinano le comunità identificate dalla più economica alla più costosa prendendo solo le per 15, confermando empiricamente la capacità dell'algoritmo di isolare cluster con comportamenti di prezzo divergenti. Ciò dimostra che la posizione topologica e l'appartenenza a una specifica comunità di rete sono determinanti significativi del prezzo finale al consumo.

ANOVA Validation: price zone (F-Score: 14.36) - Diesel



ANOVA Validation: price zone (F-Score: 14.36) - Diesel

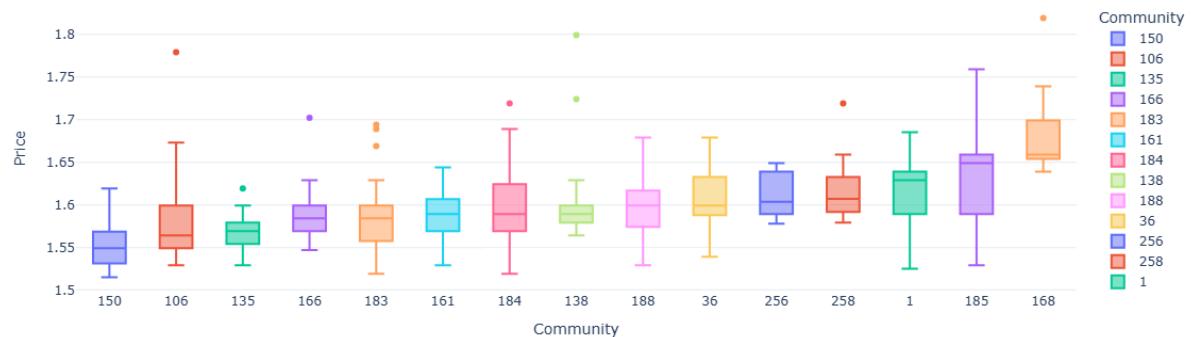


Figura 15: benzina sopra, diesel sotto, confronto boxplot community identificate.

L’analisi comparativa dei box plot relativi alle comunità ottimali rivela una marcata divergenza strutturale tra i due segmenti del mercato dei carburanti. L’ampiezza delle distribuzioni di prezzo all’interno dei cluster identificati evidenzia comportamenti competitivi profondamente diversi:

Mercato della benzina (bassa varianza) Le comunità della benzina sono caratterizzate da una varianza interna quasi nulla, rappresentata graficamente da box estremamente compresi. Questo dato statistico indica un allineamento di prezzo quasi perfetto all’interno di ciascun distretto competitivo: il segnale di prezzo si propaga con rapidità e viene replicato in modo rigoroso dai distributori appartenenti allo stesso ecosistema locale.

Mercato del diesel (alta varianza) Al contrario, il diesel mostra una varianza intra-comunitaria significativamente maggiore, evidenziata da box molto più estesi verticalmente. Questa dispersione visiva e statistica suggerisce che i prezzi del diesel sono meno reattivi e meno rigidamente vincolati alle dinamiche topologiche di vicinato rispetto a quelli della benzina.

Questa asimmetria strutturale trova una solida giustificazione economica nella diversa natura della domanda che caratterizza i due prodotti.

Mentre la benzina è un bene tipicamente orientato al consumatore privato (Business-to-Consumer, B2C), dove l’elasticità al prezzo e l’attenzione al totem esposto su strada sono massime, il mercato del diesel è intrinsecamente legato alla logistica e alle flotte aziendali (Business-to-Business, B2B). Nel

settore dei trasporti professionali, il «prezzo al tabellone» assume un’importanza secondaria rispetto alle condizioni stipulate tramite contratti privati, fuel cards aziendali o sconti sui grandi volumi. Di conseguenza, il prezzo pubblico del diesel risulta essere una variabile meno rappresentativa della reale transazione economica e, pertanto, meno soggetta all’allineamento spaziale immediato che caratterizza la spietata guerra dei prezzi al dettaglio della benzina.

5.5 Correlazione tra differenza di prezzo e matching

In questa sezione si è proceduto alla verifica dell’esistenza di una correlazione tra il grado di matching di due nodi e la loro divergenza di prezzo mediano. L’obiettivo è confermare se una forte affinità nelle dinamiche temporali di prezzo si traduca effettivamente in un posizionamento di mercato simile tra i distributori.

Per ottenere una visione d’insieme, l’analisi è stata reiterata per ogni lag temporale da 0 a 5 giorni. Coerentemente con la metodologia esposta nella [Sezione 5.2](#), il peso degli archi del grafo è stato ricalcolato per ogni scenario, riflettendo l’intensità del matching allo specifico lag considerato.

Nello specifico, la procedura ha inteso verificare se:

1. Un elevato valore di matching sia associato a una differenza di prezzo mediano tendente allo zero.
2. Se tale correlazione si mantenga significativa al variare del lag, o se la «vicinanza economica» sia una prerogativa esclusiva dei distributori che reagiscono con massima rapidità (lag 0 e 1).

Questa verifica non è puramente formale: essa serve a dimostrare che gli archi del nostro grafo non rappresentano semplici connessioni spaziali, ma sono veri e propri canali di trasmissione di una strategia commerciale comune.

Lag Temporale	Spearman r	p -value	Pearson r	p -value
Benzina				
Lag 0	−0.852	0.000	−0.451	0.000
Lag 1	−0.853	0.000	−0.464	0.000
Lag 2	−0.857	0.000	−0.466	0.000
Lag 3	−0.860	0.000	−0.471	0.000
Lag 4	−0.861	0.000	−0.475	0.000
Lag 5	−0.862	0.000	−0.478	0.000
Diesel				
Lag 0	−0.837	0.000	−0.412	0.000
Lag 1	−0.842	0.000	−0.427	0.000
Lag 2	−0.851	0.000	−0.430	0.000
Lag 3	−0.857	0.000	−0.436	0.000
Lag 4	−0.860	0.000	−0.441	0.000
Lag 5	−0.862	0.000	−0.445	0.000

Tabella 2: Correlazione tra frequenza di matching e differenza di prezzo mediano per diversi lag temporali.

I risultati delineano quello che possiamo definire un «effetto specchio». Indipendentemente dal lag temporale considerato (da 0 a 5 giorni) e dalla tipologia di carburante, si osservano coefficienti di correlazione di Spearman estremamente elevati, compresi tra -0,85 e -0,86. Questa relazione inversamente proporzionale, prossima alla perfezione statistica, indica che:

Alto matching = alta convergenza le coppie di stazioni che coordinano frequentemente i propri movimenti di prezzo tendono a convergere verso livelli di prezzo pressoché identici. Non si limitano a copiare la «tempistica» del cambiamento, ma si allineano sul medesimo valore economico.

Basso matching = divergenza economica al contrario, i distributori che non mostrano sincronizzazione nelle variazioni temporali presentano ampi divari nei prezzi mediani, segnale di un'appartenenza a segmenti di mercato o cluster competitivi differenti.

La discrepanza sistematica tra i due indici (Spearman e Pearson) rivela la natura non lineare delle dinamiche competitive osservate.

L'aspetto più rilevante è la stabilità del coefficiente dal «lag 0» al «lag 5». Questa costanza suggerisce che l'allineamento tra competitor non è un fenomeno effimero o casuale, ma una caratteristica strutturale del mercato lombardo. Se due stazioni sono identificate dal modello come «coppie strategiche», esse mantengono un allineamento rigoroso e costante durante l'intero arco temporale analizzato.

La definizione matematica di matching adottata in questa ricerca cattura con estrema precisione la realtà della competizione economica. I nodi che il grafo identifica come interconnessi sono effettivamente i protagonisti delle «guerre di prezzo» locali, operando per mantenere il proprio differenziale entro margini di pochi centesimi rispetto ai rivali diretti.

6 Effetto Rocket & feather

Fino a questo punto, l'indagine si è concentrata sulle dinamiche endogene del sistema, analizzando le interazioni spaziali e temporali tra i nodi senza considerare le sollecitazioni esogene. Tuttavia, come discusso nella [Sezione 1](#) e teorizzato da ([Simon, 1962](#)), i sistemi complessi possiedono una natura intrinsecamente gerarchica. Nel nostro caso, il livello superiore alla distribuzione al dettaglio è occupato dal mercato globale del greggio, un sistema complesso regolato da dinamiche macroeconomiche che dettano il prezzo del petrolio al barile.

Il trasferimento di valore da questo livello superiore alla «pompa» non è tuttavia immediato né simmetrico. Tale asincronia è nota in letteratura come fenomeno «Rockets and Feathers» (razzi e piume), termine coniato nel lavoro di ([Bacon, 1991](#)). La metafora descrive un'asimmetria comportamentale: i prezzi al dettaglio tendono a riflettere i rialzi del costo del greggio con la rapidità di un razzo, mentre reagiscono ai ribassi con la lentezza di una piuma in caduta. Sebbene le osservazioni originali di Bacon riguardassero il mercato del Regno Unito, studi più recenti, tra cui quello di ([Bakhat et al., 2022](#)), confermano che questa pratica è strutturalmente radicata anche nel mercato italiano, con tempi di latenza significativi per entrambi i carburanti.

L'obiettivo di questa sezione è dunque quello di quantificare, all'interno del nostro dataset, l'inerzia temporale del sistema: analizzeremo in quanto tempo i prezzi della benzina e del diesel si adattano alle fluttuazioni del prezzo del petrolio, verificando se la struttura topologica identificata influenzi o subisca questa asimmetria di trasmissione

6.1 Prezzi medi giornalieri e movimento del greggio

Sfruttando le reti temporali descritte nella [Sezione 2.1](#), è stata condotta un'analisi sistematica dei prezzi medi giornalieri per ciascun nodo e per entrambe le tipologie di carburante. Tale procedura ha l'obiettivo di verificare in prima istanza attraverso un'ispezione visiva se i segmenti del diesel e della benzina seguano trend evolutivi coerenti. Questa analisi preliminare è fondamentale per identificare eventuali scostamenti sistematici nella dinamica di prezzo dei due prodotti all'interno della rete, fornendo una base empirica per le successive valutazioni statistiche sulla sincronicità dei mercati.

L'analisi è stata estesa ai prezzi del greggio Brent, presi come riferimento per il mercato europeo. I dati, relativi al medesimo arco temporale del dataset principale (1° aprile – 30 giugno 2025), sono stati acquisiti con l'obiettivo di condurre una validazione visiva della coerenza tra le fluttuazioni del mercato dei capitali e le risposte dei prezzi al dettaglio.

I dati relativi ai prezzi giornalieri del petrolio greggio Brent sono stati ottenuti tramite questa [piattaforma](#).

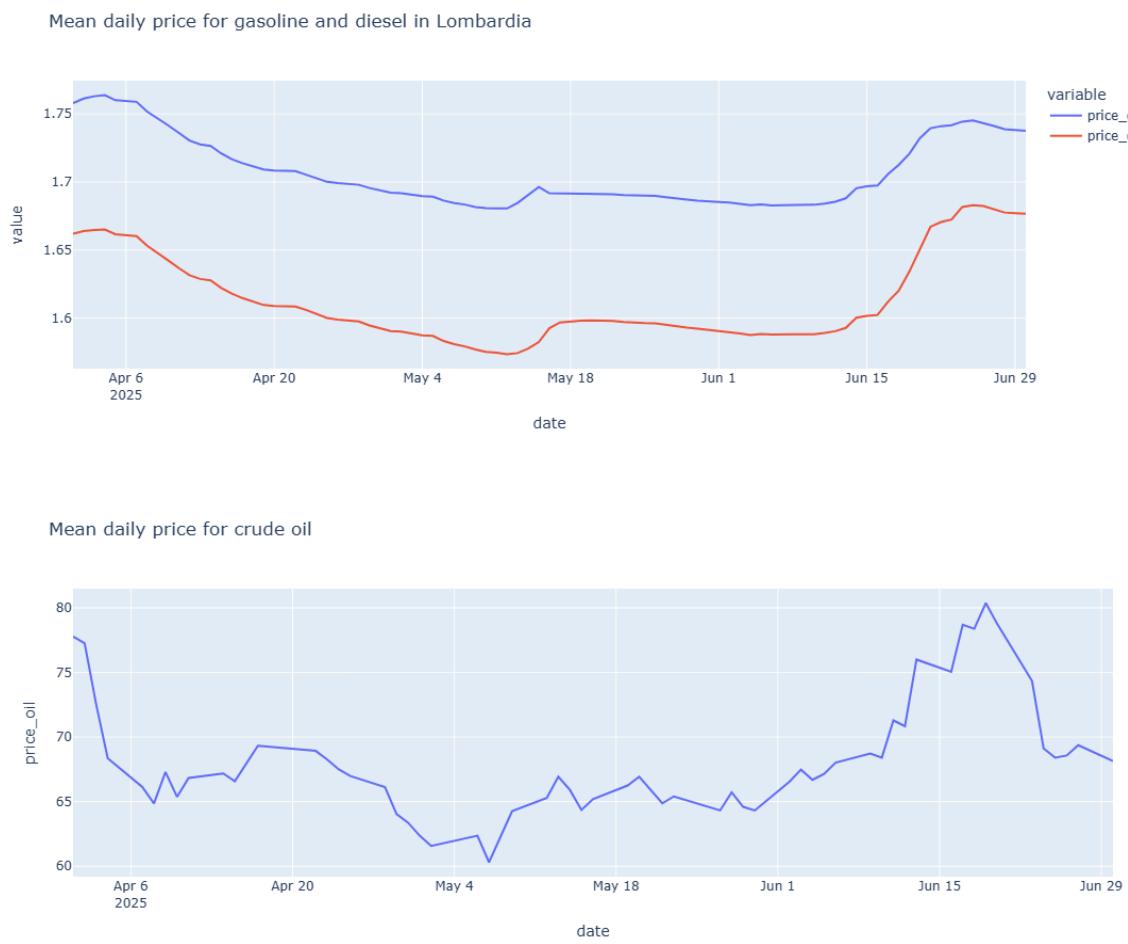


Figura 16: Comparazione tra prezzi alla pompa (sopra) e prezzo al barile (sotto).

Dall'analisi grafica emerge come le quotazioni della benzina e del diesel seguano trend evolutivi pressoché speculari. Parallelamente, l'ispezione visiva evidenzia una marcata co-movimentazione tra

i benchmark del petrolio greggio (Brent) e i prezzi finali alla pompa. Tale allineamento preliminare indica che le fluttuazioni del mercato retail non sono isolate, ma rispondono in modo sistematico alle sollecitazioni del mercato energetico globale. La quasi totale sovrapponibilità dei trend supporta l'ipotesi che il costo della materia prima rappresenti il principale driver esogeno del sistema, pur lasciando spazio all'indagine sulle asimmetrie temporali di trasmissione che verranno quantificate nei paragrafi successivi.

Osservando il prezzo del Brent emerge un'anomalia significativa in corrispondenza dell'inizio del mese di aprile, caratterizzata da una decisa flessione delle quotazioni del greggio. Tale shock esogeno può essere ricondotto alle tensioni commerciali internazionali e, nello specifico, all'annuncio dei nuovi dazi doganiali da parte dell'amministrazione Trump proprio all'inizio di aprile 2025. Come analizzato in un report di [Assolombarda](#).

6.2 Effetto Rocket & feather nel dataset

Al fine di validare empiricamente le evidenze emerse dall'ispezione grafica, si procederà al calcolo dei coefficienti di correlazione tra le variabili in esame. Tale analisi permetterà di determinare con precisione il grado di interdipendenza tra i prezzi al dettaglio e del greggio. Successivamente, verranno analizzate le velocità di trasmissione dei prezzi per rilevare l'eventuale presenza dell'effetto «Rockets and Feathers». L'obiettivo è valutare se il sistema manifesti un'asimmetria di risposta, ovvero se gli incrementi dei prezzi al dettaglio in reazione ai picchi del greggio avvengano con una rapidità superiore rispetto alla velocità di aggiustamento durante le fasi di flessione del mercato.

In una prima fase di pre-processing, le serie storiche giornaliere dei prezzi alla pompa e delle quotazioni del petrolio sono state allineate mediante un'operazione di intersezione temporale (inner join), al fine di escludere i giorni di chiusura dei mercati finanziari ed evitare distorsioni. Successivamente, i dati sono stati trasformati calcolando le differenze prime giornaliere (variazioni assolute). Per filtrare il rumore di fondo causato dalla volatilità quotidiana e isolare la componente strutturale del mercato, è stata applicata una media mobile a 3 giorni sulle quotazioni del greggio. La derivata di tale indicatore ha permesso di classificare ogni singola osservazione in un regime macroeconomico di rincaro (uptrend) o di ribasso (downtrend).

Il nucleo della verifica empirica si basa su un'analisi di cross correlazione iterativa, condotta su una finestra temporale di 20 giorni (lag 0 - lag 20). La serie delle variazioni del prezzo del carburante è stata traslata nel tempo per misurare la magnitudo della risposta ritardata agli shock originari del petrolio. Per testare l'asimmetria, la correlazione lineare è stata calcolata in modo condizionato: isolando prima le sole fasi di uptrend (per misurare la rapidità della fase Rocket) e successivamente le fasi di downtrend (per la fase Feather).

L'estrazione algoritmica dei picchi di massima correlazione per le due curve risultanti permette di quantificare oggettivamente il fenomeno: l'effetto Rockets & Feathers viene considerato validato qualora il numero di giorni necessari per registrare la reazione massimale agli aumenti del greggio risulti strettamente inferiore al tempo di assorbimento necessario durante i cali di prezzo.

È necessario precisare che l'applicazione della correlazione incrociata con sfasamento temporale (time-lagged cross-correlation) comporta una riduzione della dimensione effettiva del campione ($N_{\text{eff}} = N - k$) all'aumentare del ritardo k considerato. Tuttavia, dato l'orizzonte temporale della presente analisi ($T \approx 90$ giorni) e la finestra di sfasamento massima impostata ($k_{\max} = 20$), la numerosità campionaria residua si mantiene ampiamente entro i margini necessari a garantire la significatività statistica dei

coefficienti calcolati. Tale ampiezza permette di evitare distorsioni dovute a campioni eccessivamente ridotti, assicurando che le stime sulla velocità di trasmissione dei prezzi rimangano robuste e affidabili anche per i ritardi più elevati.



Figura 17: Comparazione tra prezzi alla pompa (sopra) e prezzo al barile (sotto).

L'analisi rivela una profonda dicotomia strutturale tra i due carburanti, confermando solo parzialmente l'ipotesi «Rockets and Feathers» a seconda della tipologia di utenza:

Mercato della benzina (B2C) i dati confermano pienamente l'ipotesi di asimmetria. La risposta del prezzo alla pompa agli shock positivi (rialzi del greggio) avviene con un ritardo medio di soli 3 giorni (corr: 0.314), mentre l'aggiustamento a fronte di shock negativi (ribassi) richiede 7 giorni (corr: 0.393). Questo differenziale di 4 giorni suggerisce l'adozione di strategie di pricing opportunistiche, volte a massimizzare i margini di breve periodo sfruttando la flessibilità della domanda privata.

Mercato del diesel (B2B) al contrario, non emerge alcuna evidenza statistica di asimmetria temporale, mostrando un tempo di reazione speculare di 6 giorni sia per gli incrementi che per i decrementi. Tale comportamento, unito alla maggiore varianza intra comunitaria identificata nei test ANOVA, delinea un mercato strutturalmente diverso. La domanda professionale, meno sensibile al prezzo esposto e spesso regolata da accordi contrattuali, disincentiva le fluttuazioni speculative rapide tipiche della benzina. Di conseguenza, il mercato del Diesel appare più

«viscoso», con una latenza di 6 giorni (corr: 0.379) contro i 3 della benzina in fase di rialzo, ma significativamente più simmetrico con latenza sempre di 6 giorni (corr: 0.356) in fase di discesa.

7 Regressione spaziale SAR

Per analizzare le interdipendenze geografiche tra i punti vendita, viene implementato un modello di autoregressione spaziale (SAR), spesso definito Spatial Lag Model. Questo framework econometrico è progettato specificamente per gestire la dipendenza spaziale in dati cross-sectional, superando i limiti dei modelli lineari classici che assumono l'indipendenza delle osservazioni.

Nel contesto del mercato dei carburanti, il modello SAR assume che il prezzo fissato da una specifica stazione di servizio non sia un valore isolato, ma risulti direttamente influenzato dalle strategie di prezzo adottate dai competitor limitrofi. L'approccio metodologico e la selezione delle variabili locali necessarie per l'esecuzione del modello sono stati mutuati dal lavoro di ([Bergantino et al., 2020](#)), nonostante loro avessero molti più dati inerenti all'analisi del territorio come ad esempio i prezzi delle abitazioni e lavorassero su un territorio meno vasto.

Il modello è stato applicato solo al mercato della benzina e configurato utilizzando come variabile dipendente il prezzo mediano di ogni nodo. Per isolare l'effetto dell'interazione spaziale, il campione è stato raffinato escludendo i nodi isolati (ovvero privi di vicini entro il raggio d'azione definito di 3km), concentrando l'analisi esclusivamente sul mercato della benzina, dove l'evidenza di asimmetria è risultata più marcata.

Le variabili indipendenti sono state raggruppate in tre categorie analitiche, seguendo l'approccio di ([Bergantino et al., 2020](#)):

Concorrenza locale misurano la struttura del mercato nel vicinato spaziale di ogni stazione:

- **Brand market share:** rappresenta la capillarità locale di un marchio, calcolata come il rapporto tra le stazioni appartenenti allo stesso brand e il totale degli impianti nell'area.
- **Concentration Ratio (CR3):** indica la quota di mercato aggregata dei tre principali operatori nell'area, fungendo da proxy per il potere oligopolistico locale.
- **Herfindahl-Hirschman Index (HHI):** misura il livello di concentrazione del mercato (somma dei quadrati delle quote di mercato), permettendo di distinguere tra mercati competitivi e mercati altamente concentrati (metrica simile a CR3 si deciderà poi in fase di esecuzione quale utilizzare.)
- **Station density:** il numero totale di stazioni nell'area, utilizzato come misura della densità dell'offerta.
- **Low-cost presence:** il conteggio dei distributori indipendenti o della grande distribuzione (es. «Pompe Bianche», Auchan, Carrefour), fondamentali per valutare la pressione competitiva sui prezzi dei brand premium.

Variabili di controllo del punto vendita:

- **Tipologia stradale:** variabile categorica per distinguere tra stazioni stradali e autostradali.
- **Brand dummies:** variabili indicatrici per intercettare gli effetti fissi legati al posizionamento di mercato e alle politiche di prezzo specifiche di ogni marchio.
- **Influencer:** se costituisce un influencer a 1 giorno di lag nella rete 3km orientata.

Variabili territoriali e strutturali:

- **Community:** identifica l'appartenenza alle community individuata precedentemente nella [Sezione 5.4](#), per verificare se l'integrazione in un macro-cluster influenzi il prezzo, considerando solo le community con una dimensione pari o superiore a 10 elementi.
- **Matrice dei pesi spaziali (W):** Costruita sulla base della distanza geografica tra le stazioni entro i 3km, pesata per il valore di matching temporale (1-matching), per riflettere non solo la vicinanza fisica ma anche la solidità del legame relazionale identificato nella rete.

Per l'inserimento delle variabili categoriche (Brand e Community) nel modello di regressione, è stata adottata la tecnica della codifica one-hot con l'esclusione di una categoria di riferimento. Tale procedura è indispensabile per evitare la cosiddetta «trappola delle variabili dummy», una condizione di perfetta multicollinearità che renderebbe la matrice dei dati singolare e impedirebbe la stima dei coefficienti.

Nello specifico, è stato aggiunto il prefisso «AAA» alle categorie «Pompe Bianche» e «Zone_Other» (community di cui fanno parte tutti i nodi non appartenenti a una community presa in considerazione) prima della generazione delle variabili dummy. Poiché gli algoritmi di calcolo escludono automaticamente la prima categoria in ordine alfabetico come baseline, questa operazione garantisce che i distributori indipendenti e le zone non classificate fungano da termine di paragone. In questo modo, i coefficienti stimati per gli altri brand e zone rappresentano direttamente lo scostamento di prezzo (differenziale) rispetto alle pompe bianche e alla zona che raggruppa tutti i nodi con community inferiori ai 10 membri, facilitando l'interpretazione economica dei risultati.

Quando si trasforma una variabile qualitativa (es. il Brand: Eni, Esso, IP) in numeri per una regressione, si creano delle colonne «0 o 1».

Il problema (multicollinearità perfetta): Se hai 3 brand e crei 3 colonne, la loro somma sarà sempre 1 per ogni riga (perché ogni stazione appartiene a un brand). Se la regressione include anche l'intercetta (il termine costante), il modello troverà due modi diversi per calcolare la stessa cosa. Matematicamente, questo significa che le colonne sono linearmente dipendenti:

$$\text{Dummy}_A + \text{Dummy}_B + \text{Dummy}_C = 1 \quad (\text{Intercetta})$$

La soluzione: si elimina una colonna. Se eliminiamo «Brand A», il modello userà quella categoria come punto zero. Se il coefficiente di «Brand B» è +0.05, significa che il Brand B costa 5 centesimi in più rispetto al Brand A.

Le librerie come Pandas o Scikit-Learn di solito eliminano la prima colonna seguendo l'ordine alfabetico. Rinominando «Pompe Bianche» in «AAA_Pompe Bianche», ci assicuriamo che sia lei a sparire e a diventare la base. È un'ottima scelta perché ha senso confrontare i prezzi dei grandi marchi (Eni, Shell) rispetto ai distributori indipendenti.

Per la stima del modello SAR, è stata definita una matrice di pesi spaziali (W) personalizzata, volta a catturare l'intensità delle interazioni tra i punti vendita non solo in termini di prossimità fisica, ma anche di affinità comportamentale. La costruzione del peso $w_{\{i,j\}}$ tra due nodi i e j si basa sull'integrazione di due componenti fondamentali:

1. **Affinità (matching):** è stato utilizzato l'attributo di `1_matching` estratto dalla rete 3km orientata. Poiché tale valore indica la frequenza con cui due stazioni aggiornano i prezzi tra loro, esso funge da proxy per l'interdipendenza strategica tra i retailer.
2. **Prossimità geografica:** sfruttando le distanze sempre fino a 3km calcolate nella [Sezione 2.2](#), è stata applicata la funzione dell'inverso della distanza. Questo assicura che i distributori spazialmente più

vicini esercitino un'influenza superiore sulla formazione del prezzo del nodo target rispetto a quelli remoti.

Il peso finale è ottenuto dal prodotto tra il valore di matching e il reciproco della distanza. Tale approccio permette di modellare una rete in cui il legame competitivo è massimo tra stazioni vicine che mostrano una forte sincronia nei prezzi, fornendo una rappresentazione più granulare e realistica delle dinamiche di mercato rispetto a una semplice matrice basata solo sulla distanza.

Variabile	Coefficiente	Err. Std.	Stat-z	p-value
Costante	4.4253	0.2626	16.851	< 0.001 ***
N. Low-Cost	0.0041	0.0011	3.799	< 0.001 ***
CR3 (Concentrazione)	0.0002	0.0001	2.060	0.039 **
Is_Highway (Autostrada)	0.1100	0.0076	14.527	< 0.001 ***
Is_Influencer	-0.0182	0.0030	-6.111	< 0.001 ***
Brand: Agip/Eni	0.0163	0.0031	5.321	< 0.001 ***
Brand: API/IP	0.0088	0.0031	2.850	0.004 ***
Brand: Tamoil	-0.0091	0.0038	-2.414	0.016 **
Brand: Q8	0.0053	0.0033	1.641	0.101
Brand: Esso	0.0016	0.0036	0.451	0.652
Latitudine	-0.0584	0.0055	-10.572	< 0.001 ***
Longitudine	-0.0103	0.0029	-3.538	< 0.001 ***
Zone_174 (Max Rincaro)	0.1798	0.0153	11.758	< 0.001 ***
Zone_48 (Rincaro)	0.0701	0.0146	4.790	< 0.001 ***
Zone_251 (Sconto)	-0.0658	0.0129	-5.109	< 0.001 ***
Zone_153 (Sconto)	-0.0722	0.0146	-4.956	< 0.001 ***
Zone_159 (Max Sconto)	-0.0910	0.0119	-7.624	< 0.001 ***
W_Price (ρ)	0.0061	0.0205	0.296	0.767
Altre Dummy di Zona (26 sign.)			Sì	

Tabella 3: Risultati della stima Maximum Likelihood per il modello SAR. Sono riportate in dettaglio le coordinate spaziali, i brand principali e le zone caratterizzate dai coefficienti di maggiore magnitudo.

Livelli di significatività: *** p < 0.01, ** p < 0.05.

*Pseudo R*²: 0.306 | Osservazioni: 2409 | Log-Likelihood: 3904.9

Lo Pseudo *R*² si attesta a 0.306, indicando che il modello è in grado di spiegare circa il 30,6% della variabilità nei prezzi dei carburanti. Sebbene tale valore rappresenti una base statisticamente accettabile per l'analisi di dati cross-sectional, esso suggerisce implicitamente che la restante quota di varianza (quasi il 70%) è guidata da fattori esogeni non formalizzati in questa equazione.

Il risultato più sorprendente e controintuitivo dell'output risiede nel coefficiente di ritardo spaziale (Spatial Lag, W_{Price}), che risulta prossimo allo zero (0.006) e statisticamente non significativo ($p = 0.767$). Contrariamente agli assunti della teoria economica spaziale standard, questo modello suggerisce che la strategia di pricing di una stazione non subisca un'influenza statistica diretta e misurabile dai

prezzi praticati dai vicini immediati (definiti entro la rete dei 3km). Questa anomalia è verosimilmente imputabile a due fattori congiunti: in primo luogo, la frammentazione della matrice dei pesi spaziali (che presenta numerose componenti disconnesse); in secondo luogo, la massiccia inclusione di variabili dummy zonali. Queste ultime agiscono come «spugne», assorbendo interamente la variabilità spaziale locale e rendendo di fatto ridondante il termine di interazione spaziale pura.

Determinati strutturali e fisiche:

Rete autostradale (Is_Highway) con un coefficiente di +0.110 ($p < 0.001$), risulta essere la variabile binaria di maggiore impatto. La localizzazione sull'infrastruttura autostradale genera un rincaro di circa 11 centesimi rispetto alla viabilità ordinaria.

Nodi influencer (Is_Influencer) le stazioni identificate topologicamente come «influencer» applicano prezzi significativamente più bassi (-1.8 centesimi). Tale evidenza suggerisce che i market leader locali stimolino la competizione attraverso logiche di discounting per massimizzare i volumi di vendita, costringendo i follower ad adeguarsi, piuttosto che guidare speculazioni al rialzo.

Densità Low-Cost (N_LowCost) in modo apparentemente controiduitivo, una maggiore densità di distributori indipendenti è correlata a un lieve incremento dei prezzi (+0.4 centesimi). Questa dinamica potrebbe indicare un bias di selezione spaziale: gli operatori low-cost tendono a concentrarsi strategicamente in aree ad alta domanda, dove il livello base dei prezzi (fisiologicamente più alto) permette loro di operare comunque in modo competitivo pur mantenendo margini interessanti.

Effetti legati al brand, il modello definisce una chiara gerarchia di posizionamento rispetto alla baseline indipendente («Pompe Bianche»):

Fascia premium marchi storici come Agip/Eni (+1.6 centesimi) e API/IP (+0.8 centesimi) impongono un sovrapprezzo statisticamente significativo, giustificabile dalla percezione del brand, dalla fidelizzazione e dai servizi ancillari.

Fascia discount Tamoil si posiziona significativamente al di sotto della baseline (-0.9 centesimi), delineando una chiara strategia di aggressione sul prezzo.

Fascia neutrale operatori come Esso e Q8 non presentano differenze statisticamente significative rispetto ai distributori indipendenti ($p > 0.10$).

Effetti geografici e territoriali

Coordinate sia la latitudine che la longitudine presentano coefficienti negativi e significativi, evidenziando un gradiente geografico regionale: all'interno del dominio di studio, i prezzi tendono a decrescere man mano che ci si sposta verso Nord e verso Est.

Macro aree di prezzo il modello isola specifici cluster territoriali con dinamiche estreme. Ad esempio, la Zone_174 si configura come un outlier di massimo, con rincari di quasi 18 centesimi, seguita dalla Zone_48 (+7 centesimi). All'opposto, le Zone_159 (-9 centesimi), Zone_153 (-7.2 centesimi) e Zone_251 (-6.5 centesimi) si caratterizzano come le aree strutturalmente più competitive ed economiche della rete. L'elevata significatività statistica di questi effetti fissi giustifica ulteriormente la neutralità del termine W_{Price} : i cosiddetti neighborhood effects (effetti di vicinato) sono di fatto già inglobati e spiegati dall'appartenenza a queste specifiche macro-zone.

8 Simulatore

In coerenza con gli obiettivi del corso «Sistemi Complessi: Modelli e Simulazione», il presente elaborato compie ora il naturale passaggio dalla fase di modellizzazione analitica a quella operativa, procedendo allo sviluppo di un simulatore specificamente concepito per il mercato della benzina.

Capitalizzando le evidenze empiriche emerse nelle precedenti fasi di indagine, l'architettura del simulatore integra tre pilastri fondamentali: i risultati dell'autoregressione spaziale (modello SAR [Sezione 7](#)), le interdipendenze topologiche individuate all'interno delle community ottimizzate [Sezione 5.4](#), e le dinamiche temporali asimmetriche tipiche dell'effetto «Rockets and Feathers» [Sezione 6](#).

La rete su cui viene applicata la simulazione è quella orientata a 3km.

La fase di inizializzazione dell'algoritmo prevede un rigoroso filtraggio dei coefficienti estratti dalla regressione spaziale: per garantire la massima robustezza e stabilità delle proiezioni, vengono isolate e implementate esclusivamente le variabili che hanno dimostrato una solida significatività statistica, depurando il sistema dal rumore di fondo.

Per garantire un adeguato livello di realismo al simulatore, il passaggio successivo prevede il calcolo delle deviazioni standard all'interno delle singole community. L'estrazione di questa metrica di dispersione è fondamentale per mappare l'eterogeneità locale e inizializzare le fluttuazioni di prezzo, evitando un comportamento deterministico irrealistico da parte dei nodi appartenenti al medesimo raggruppamento.

La fase di setup del simulatore procede con l'inizializzazione del prezzo per ciascun nodo (stazione di servizio), definendo così lo stato di partenza della rete al tempo $t = 0$. Per garantire che la simulazione rifletta fedelmente l'eterogeneità del mercato reale, il valore iniziale non viene assegnato in modo omogeneo, ma è calcolato dinamicamente a partire da un prezzo baseline di riferimento globale (pari a 1.600). A questa base di partenza vengono applicati due ordini di correzioni, corrispondenti alle due anime del simulatore:

Componente strutturale vengono sommati algebricamente i coefficienti estratti dal modello SAR (i rincari per la presenza in autostrada o gli sconti legati a un determinato brand o community). Questa operazione ancora il prezzo di partenza alle caratteristiche intrinseche e spaziali del singolo punto vendita.

Componente locale viene introdotta una perturbazione basata sulla deviazione standard della community topologica di appartenenza. L'integrazione di questa metrica di dispersione permette di simulare il rumore statistico e le micro-fluttuazioni locali, dotando la rete di un realismo comportamentale che previene la convergenza artificiale verso valori puramente teorici.

Al fine di incrementare il realismo della simulazione, il prezzo risultante viene sottoposto a un processo di arrotondamento. Nello specifico, i valori vengono approssimati applicando la strategia del 9-ending pricing, allineando così gli output del simulatore alle convenzioni di pricing psicologico osservate quotidianamente nel mercato retail dei carburanti.

Dopo aver inizializzato i prezzi si passa all'architettura della simulazione si fonda su una logica a doppio livello progettata per modellare l'evoluzione dei prezzi attraverso l'interazione tra shock macroeconomici e risposte micro locali.

1. Generazione della volatilità esogena (shock del greggio)

Per emulare i cicli di mercato, la variazione del prezzo del greggio è generata attraverso una funzione sinusoidale periodica con ciclo di 30 giorni, con un prezzo base di 65 che oscilla tra 55 e 75. Questa scelta metodologica permette di simulare una volatilità controllata ma realistica, rappresentando le fluttuazioni dei costi di approvvigionamento a cui i distributori devono reagire.

Per garantire la coerenza dimensionale tra gli input del mercato energetico globale e gli output della simulazione retail, è stato implementato un fattore di conversione volumetrico. Poiché le quotazioni del greggio sono convenzionalmente espresse in dollari per barile (USD/bbl), mentre i prezzi alla pompa sono calcolati in euro per litro (EUR/L), il modello integra il seguente rapporto di conversione: un barile di petrolio corrisponde approssimativamente a 159 litri, per traslare le variazioni di prezzo del greggio direttamente sull'unità di misura finale, viene applicato l'inverso di tale valore ($\frac{1}{159}$). Questa normalizzazione permette al simulatore di calcolare in modo lineare come uno shock di prezzo sul mercato delle commodity si riflette sul costo unitario al litro.

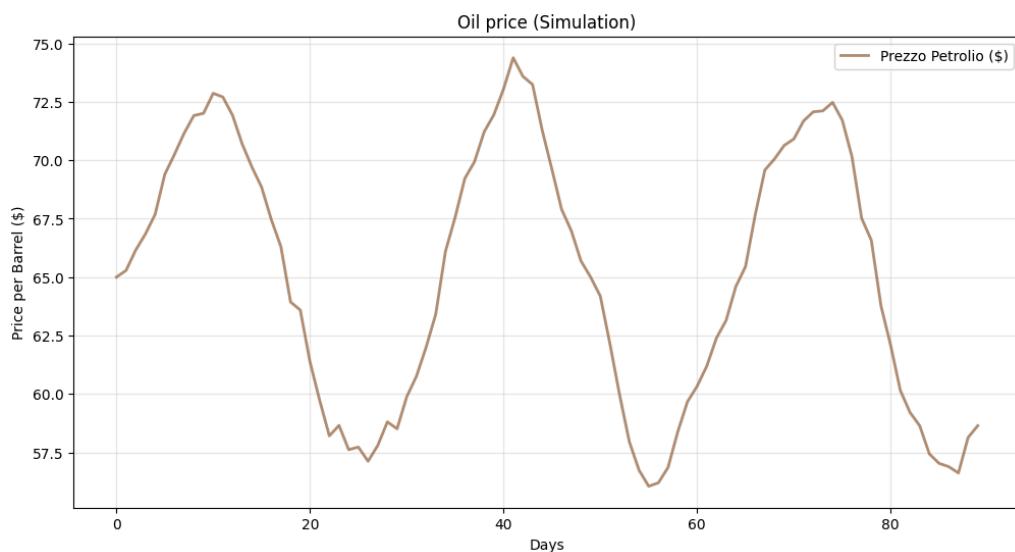


Figura 18: simulazione prezzo petrolio su 90 giorni

2. Meccanismi di aggiustamento del prezzo retail

Il comportamento dei singoli punti vendita è governato da due punti decisionali:

Asimmetria temporale (Rockets and Feathers) i nodi reagiscono con velocità differenziata alle variazioni del greggio, traslando i rincari con immediatezza (3 giorni) e i ribassi con una latenza calcolata (7 giorni), replicando così l'inefficienza del mercato reale trovata nella [Sezione 6](#).

Interazione topologica (matching) ogni stazione calibra il proprio prezzo in base alla pressione competitiva dei vicini, pesata tramite il coefficiente di matching della rete temporale precedentemente identificato.

3. Vincoli di equilibrio

Il fenomeno dei «Rockets and Feathers» introduce una distorsione temporale nel sistema: la reattività ai rincari (latenza di 3 giorni) è significativamente superiore alla velocità di recepimento dei ribassi (latenza di 7 giorni). All'interno di un modello iterativo, questa asincronia genera un effetto di accumulo esponenziale che porterebbe i prezzi verso soglie economicamente irrealistiche (4 €/L). Per mitigare questa deriva e garantire la stabilità del sistema, è stato introdotto un meccanismo di corre-

zione giornaliera pari al 2% del gap di variazione tra un prezzo considerato equo e il prezzo calcolato al giorno. Questa operazione agisce come un fattore di smorzamento, riducendo progressivamente lo scarto tra il prezzo equo e la nuova condizione di equilibrio. Tale correzione non annulla l'asimmetria comportamentale dei nodi, ma impedisce che la divergenza temporale si trasformi in una divergenza di valore, mantenendo il simulatore entro i confini di plausibilità del mercato reale.

Il prezzo equo rappresenta un prezzo teorico di equilibrio:

$$P_{f,t} = (O_t \times C) + M$$

Dove i parametri sono così definiti:

- $P_{f,t}$: Prezzo equo teorico al tempo t (EUR/L).
- O_t : Quotazione del greggio al tempo t (EUR/bbl).
- C : Fattore di conversione barile/litro ($\frac{1}{159}$).
- M : Margine strutturale fisso, comprensivo di costi operativi e accise (1.20 EUR/L).

Infine, come applicato anche per l'inizializzazione, ogni output viene processato con un arrotondamento a tre cifre decimali con terminazione fissa a .009, riflettendo accuratamente le convenzioni di pricing dell'industria petrolifera.



Figura 19: simulazione prezzo benzina 10 stazioni differenti 90 giorni

Sebbene il simulatore sviluppato serva come proof of concept, esso costituisce un modello che richiede ulteriori fasi di calibrazione per rifinarne la capacità predittiva.

Le iterazioni future dovranno focalizzarsi sull'ottimizzazione dell'architettura del codice e sull'integrazione di variabili di ordine superiore per incrementarne la precisione granulare. È tuttavia essenziale sottolineare che, poiché la determinazione dei prezzi retail è intrinsecamente legata alla volatilità dei mercati globali del greggio, il raggiungimento di una replica perfetta della realtà è limitato da fattori macroeconomici esterni e stocastici. Tali dinamiche, per loro natura, non possono essere interamente confinate o modellate all'interno di una simulazione a circuito chiuso, definendo così il confine tra l'approssimazione modellistica e la complessità imprevedibile dei sistemi reali.

9 Sviluppi futuri

Ricollegandoci all’impianto metodologico iniziale, le reti temporali illustrate in [Sezione 2.1](#) all’interno delle quali gli archi modellano l’esatta uguaglianza giornaliera dei prezzi tra i punti vendita, fa emergere un fenomeno empirico di estremo rilievo: oltre 160 coppie di stazioni per entrambi i carburanti (165 per la benzina e 164 per il diesel) hanno mantenuto una sincronizzazione assoluta delle tariffe per l’intero orizzonte di osservazione di 91 giorni.

Questa correlazione strutturale e persistente richiede un doveroso approfondimento analitico. Risulta infatti fondamentale determinare se tale allineamento derivi da assetti proprietari condivisi, dall’impiego di algoritmi di pricing automatizzati che replicano istantaneamente le mosse dei competitor, oppure se si tratti della manifestazione di dinamiche di collusione tacita localizzata.

10 Conclusioni

L’analisi condotta sul mercato retail dei carburanti in Lombardia ha permesso di decostruire la visione tradizionale che considera i punti vendita come entità isolate, confermando come la formazione del prezzo sia l’espressione emergente di un sistema complesso e fortemente interconnesso. Attraverso l’integrazione di reti spaziali, dinamiche temporali e modelli econometrici, la ricerca ha mappato non solo la struttura fisica del mercato, ma le architetture invisibili dell’influenza competitiva.

Dall’indagine sono emerse evidenze strutturali e comportamentali che ridefiniscono le logiche di leadership e propagazione del prezzo. La competizione si gioca su scala micro territoriale, con la massima intensità di imitazione concentrata nel raggio del primo chilometro. Il mercato non è governato da un vertice centralizzato, ma si configura come un arcipelago di «isole di influenza», dove la leadership è frammentata e distribuita.

Contrariamente alla teoria classica delle reti, un’elevata connettività topologica non garantisce potere di mercato. I nodi iper-connessi nei centri urbani subiscono un «fuoco incrociato» che diluisce il loro segnale di prezzo. I veri market movers capaci di innescare profonde cascate di prezzo sono spesso attori indipendenti (le cosiddette «Pompe Bianche», come il caso emblematico di Socogas) situati in posizioni meno saturate, la cui aggressività tattica domina l’ecosistema locale.

Lo studio ha rivelato una netta divergenza tra il mercato della benzina e quello del diesel. La benzina, rivolta al consumatore privato, mostra una varianza di prezzo intra comunitaria quasi nulla. In questo segmento si manifesta palesemente l’asimmetria del fenomeno «Rocket & Feather», con prezzi che reagiscono ai rincari del greggio in soli 3 giorni, ma impiegano 7 giorni per recepire i ribassi.

Il diesel, legato a logiche aziendali e logistiche, appare invece più simmetrico nei tempi di reazione e meno vincolato all’allineamento spaziale immediato.

La propagazione dei prezzi non è un flusso continuo, ma procede per ondate sincronizzate concentrate nei primi giorni della settimana, intervallate da fasi di stabilità. Questa profonda comprensione delle regole microscopiche e spaziali, depurata dal rumore di fondo tramite la regressione SAR, ha reso possibile la costruzione di un simulatore sperimentale capace di replicare l’evoluzione macroscopica del mercato della benzina con un certo grado di realismo.

In sintesi, la ricerca dimostra che il prezzo esposto sui tabelloni non è una semplice equazione basata sui costi di approvvigionamento, ma il risultato di una continua negoziazione spaziale e topologica.

Le prospettive future di indagine, come il miglioramento del simulatore e l’analisi delle oltre 160 coppie di stazioni che hanno mantenuto una perfetta sincronizzazione per 91 giorni consecutivi, potranno

chiarire se questa complessa danza competitiva sia guidata da algoritmi di pricing automatizzati, da assetti proprietari condivisi o da forme di collusione locale.

Bibliografia

- Alderighi, M., & Baudino, M. (2015). The pricing behavior of Italian gas stations: Some evidence from the Cuneo retail fuel market. *Energy Economics*, 50, 33–46. <https://doi.org/10.1016/j.eneco.2015.04.017>
- Bacon, R. W. (1991). Rockets and feathers: the asymmetric speed of adjustment of UK retail gasoline prices to cost changes. *Energy Economics*, 13(3), 211–218. [https://doi.org/10.1016/0140-9883\(91\)90022-R](https://doi.org/10.1016/0140-9883(91)90022-R)
- Bakhat, M., Rosselló, J., & Sansó, A. (2022). Price transmission between oil and gasoline and diesel: A new measure for evaluating time asymmetries. *Energy Economics*, 106, 105766. <https://doi.org/10.1016/j.eneco.2021.105766>
- Barabási, A.-L., & Bonabeau, E. (2003). Scale-Free Networks. *Scientific American*, 288(5), 60–69. <https://www.jstor.org/stable/26060284>
- Bergantino, A. S., Capozza, C., & Intini, M. (2020). Empirical investigation of retail fuel pricing: The impact of spatial interaction, competition and territorial factors. *Energy Economics*, 90, 104876. <https://doi.org/10.1016/j.eneco.2020.104876>
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10), P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Boeing, G. (2025). Modeling and Analyzing Urban Networks and Amenities With OSMnx. *Geographical Analysis*. <https://doi.org/10.1111/gean.70009>
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1), 269–271. <https://doi.org/10.1007/BF01386390>
- Fronzetti Colladon, A., Verdoliva, G., Segneri, L., & Vitali, A. G. (2024). Analyzing gasoline prices in five Italian cities: Insights from social network analysis. *Cities*, 150, 105075. <https://doi.org/10.1016/j.cities.2024.105075>
- Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20(2), 130–141.
- Simon, H. A. (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society*, 106(6), 467–482.

*Questo progetto è dedicato al mio amico Danilo,
venuto a mancare in sella alla sua moto.*