



MASCHIO GASPARDO

**Strumenti statistici per l'analisi
di dati aziendali**

Anno Accademico 2023-2024

Alongi Ester
Boraso Chiara
Pisani Chiara



Indice

- Presentazione dell'azienda **03**
- Analisi esplorativa **05**
- Prima domanda di ricerca **09**
- Seconda domanda di ricerca **24**
- Terza domanda di ricerca **34**
- Conclusioni **46**





Presentazione dell'azienda



8

Stabilimenti produttivi



13

Filiali commerciali



85%

Di fatturato generato all'estero



2.000

Dipendenti nel mondo



65.000

Macchine prodotte (2023)



3.600 t

di emissioni di CO2 ridotte ogni
anno

Leader mondiale nella produzione di attrezzature agricole per lavorazione del terreno, semina, fertilizzazione, trattamento colture, manutenzione del verde e fienagione.

Dataset

AREA MANAGER 82 livelli	COUNTRY 132 livelli	REGION 582 livelli	CUSTOMER	SUPER TYPE 5 livelli	FAMILY 11 livelli	TYPE 31 livelli	MODEL 309 livelli	MEASURE 2562 livelli	PART NUMBER 15282 livelli	DATA	QUANTITY
AGA	Turkey	Burdur	2006993	Haymaking	Harvesting	Round Balers	EXTREME1	C4100009	BC0316402	2022-11	3
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx



Informazione geografica a diversi livelli di dettaglio.



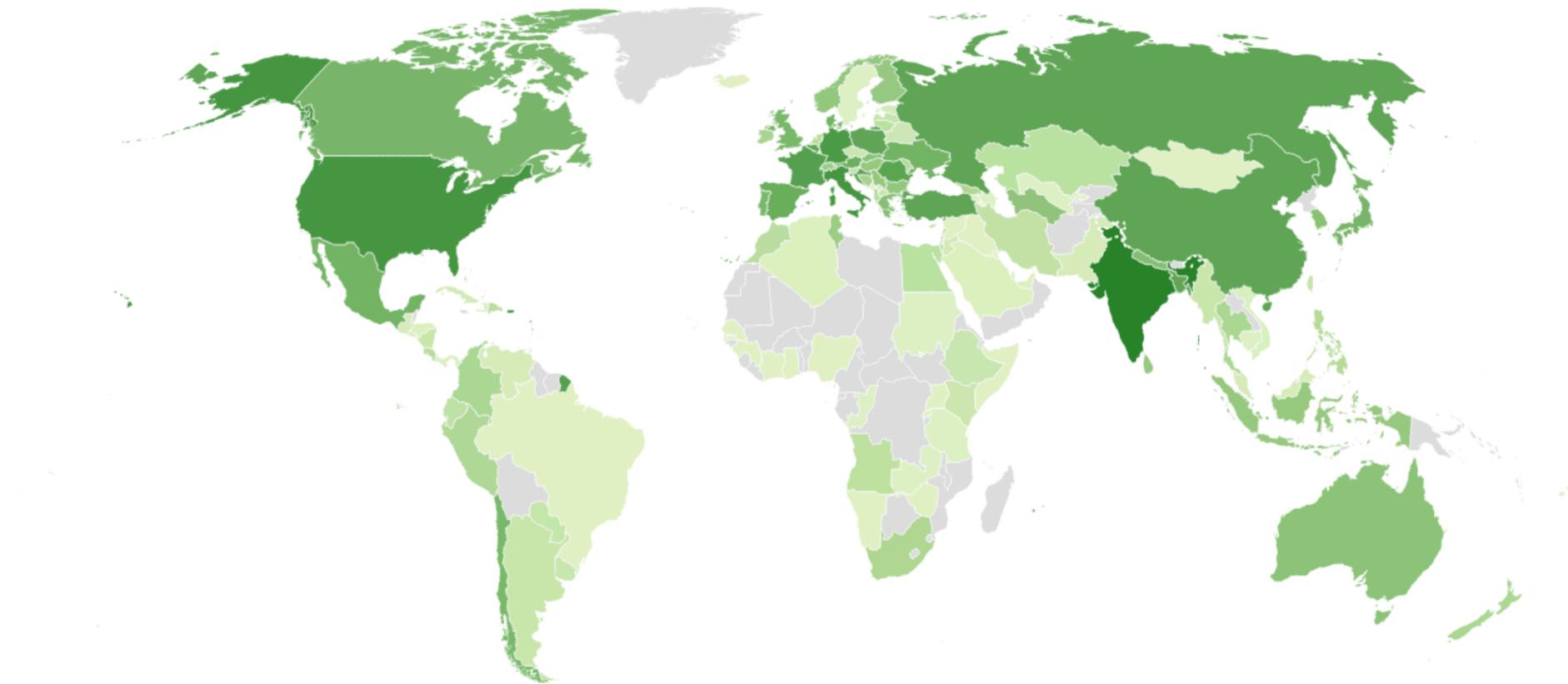
Identificazione del prodotto in diversi livelli di dettaglio.

- Il dataset completo presenta 169194 osservazioni e 12 variabili.
- **Variabile risposta:** numero prodotti venduti tra il 2016 e il 2024 (marzo)

- Avendo diversi livelli di dettaglio, si sceglie un livello per il prodotto a seconda dell'obiettivo dell'analisi.
- Avendo diversi livelli di dettaglio, si sceglie **country** o **region** a seconda del contesto.
- Non verrà utilizzata **Area Manager** in quanto una proxy dell'area geografica.

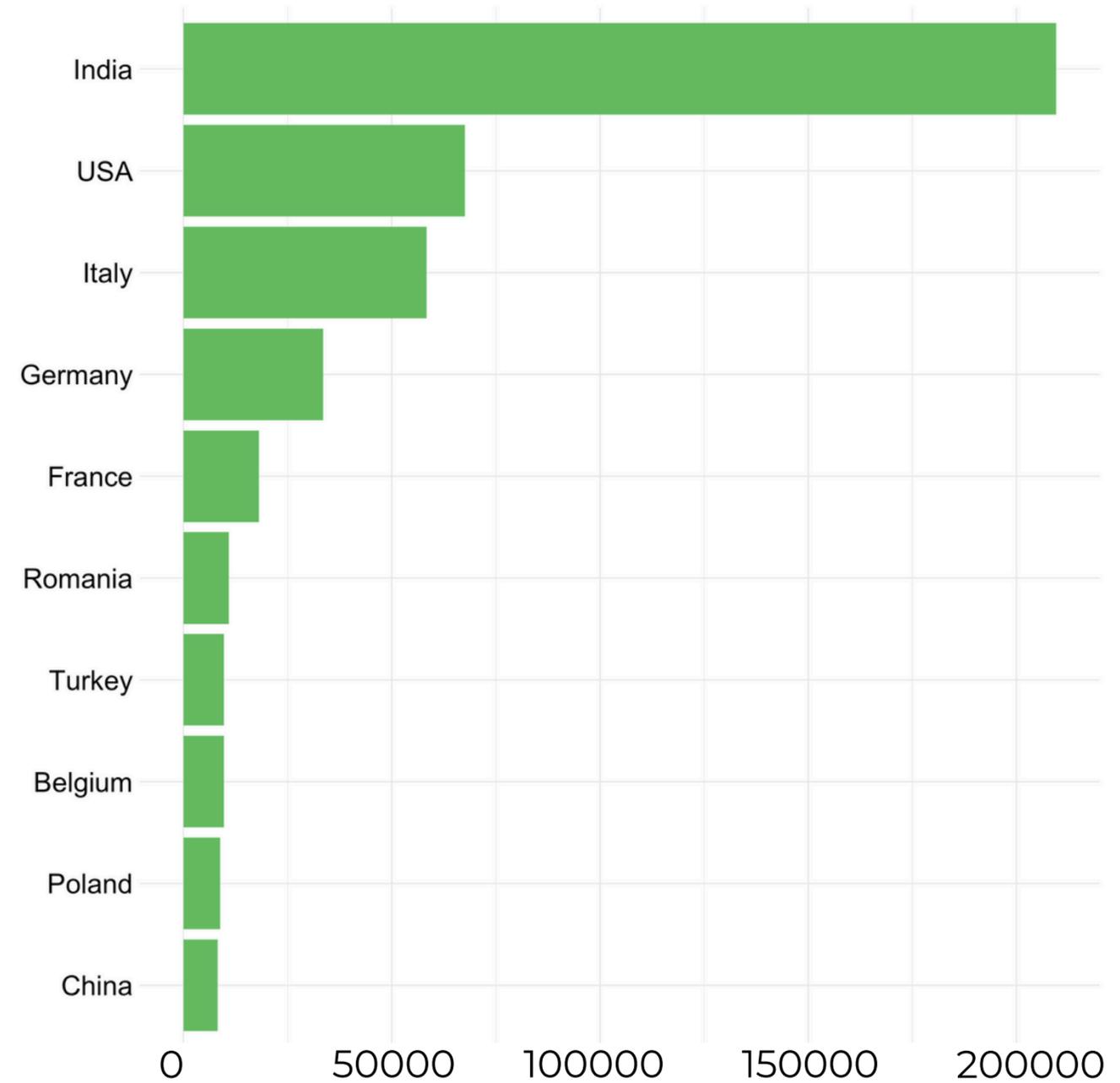
Analisi esplorativa

Numero articoli venduti



Created with Datawrapper

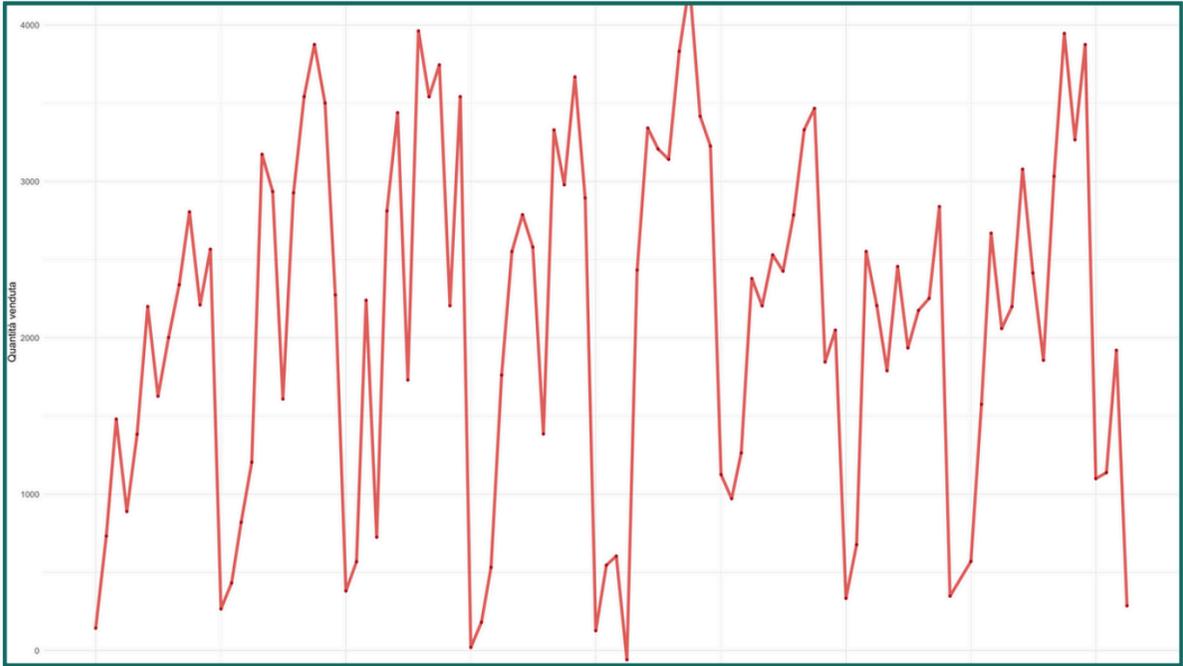
Variabile **COUNTRY** divisa in 132 livelli



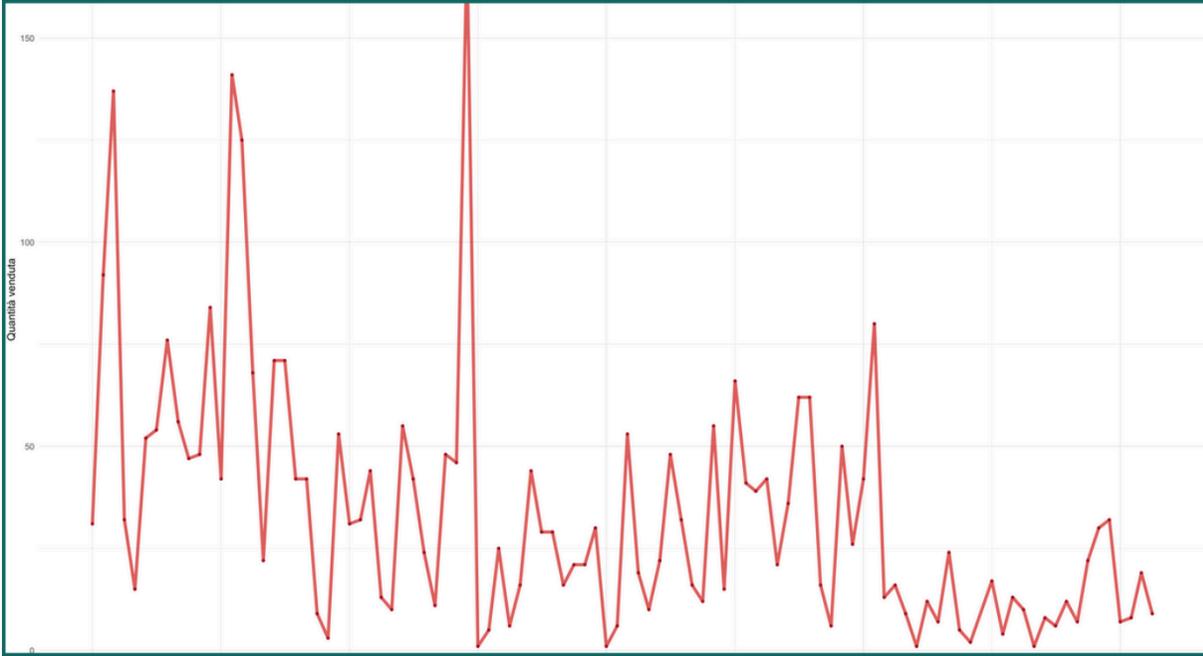
Analisi esplorativa

Andamento della quantità venduta dal 2016 al 2024

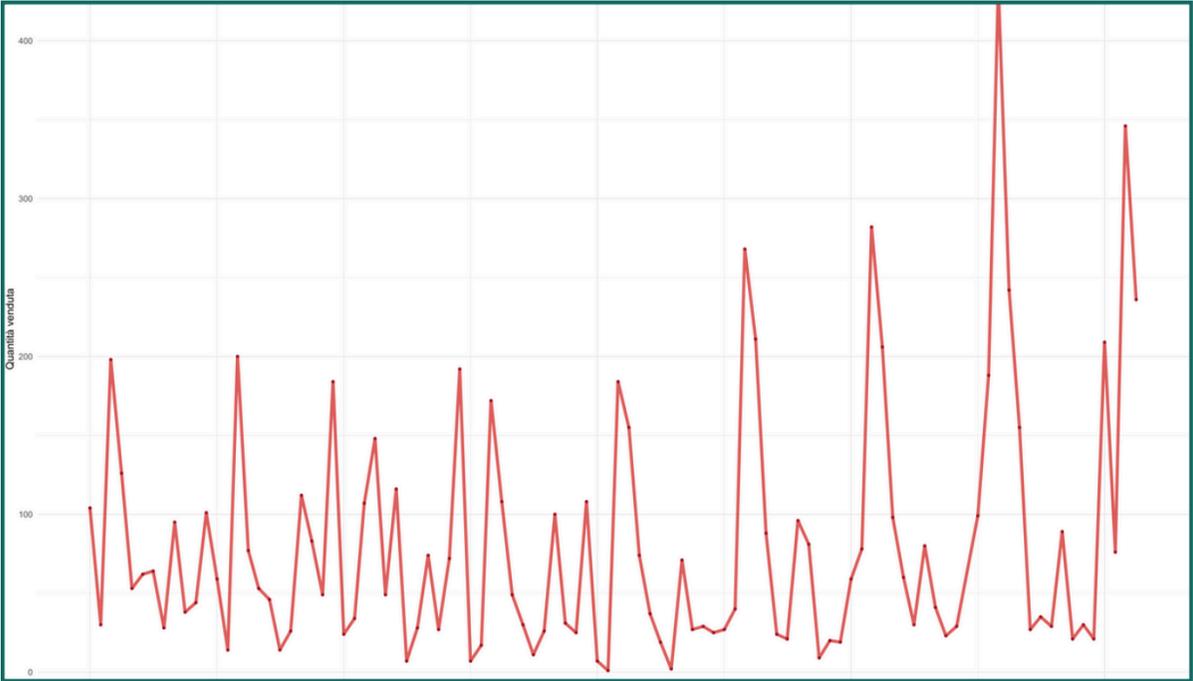
India



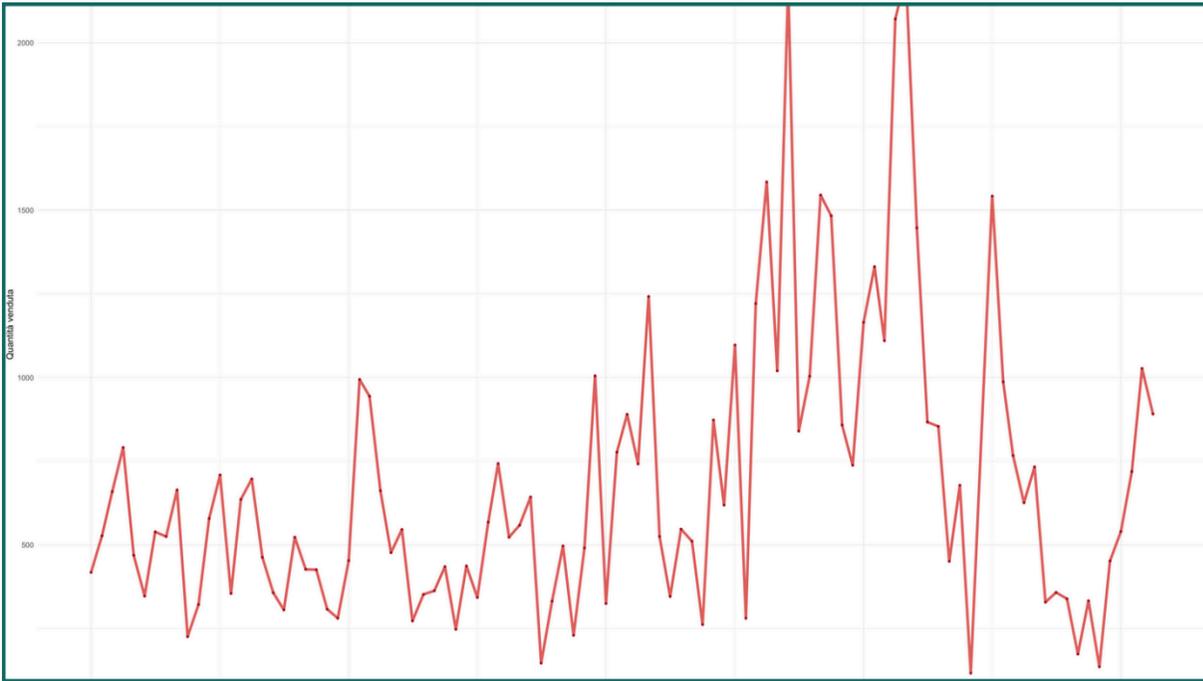
Ucraina



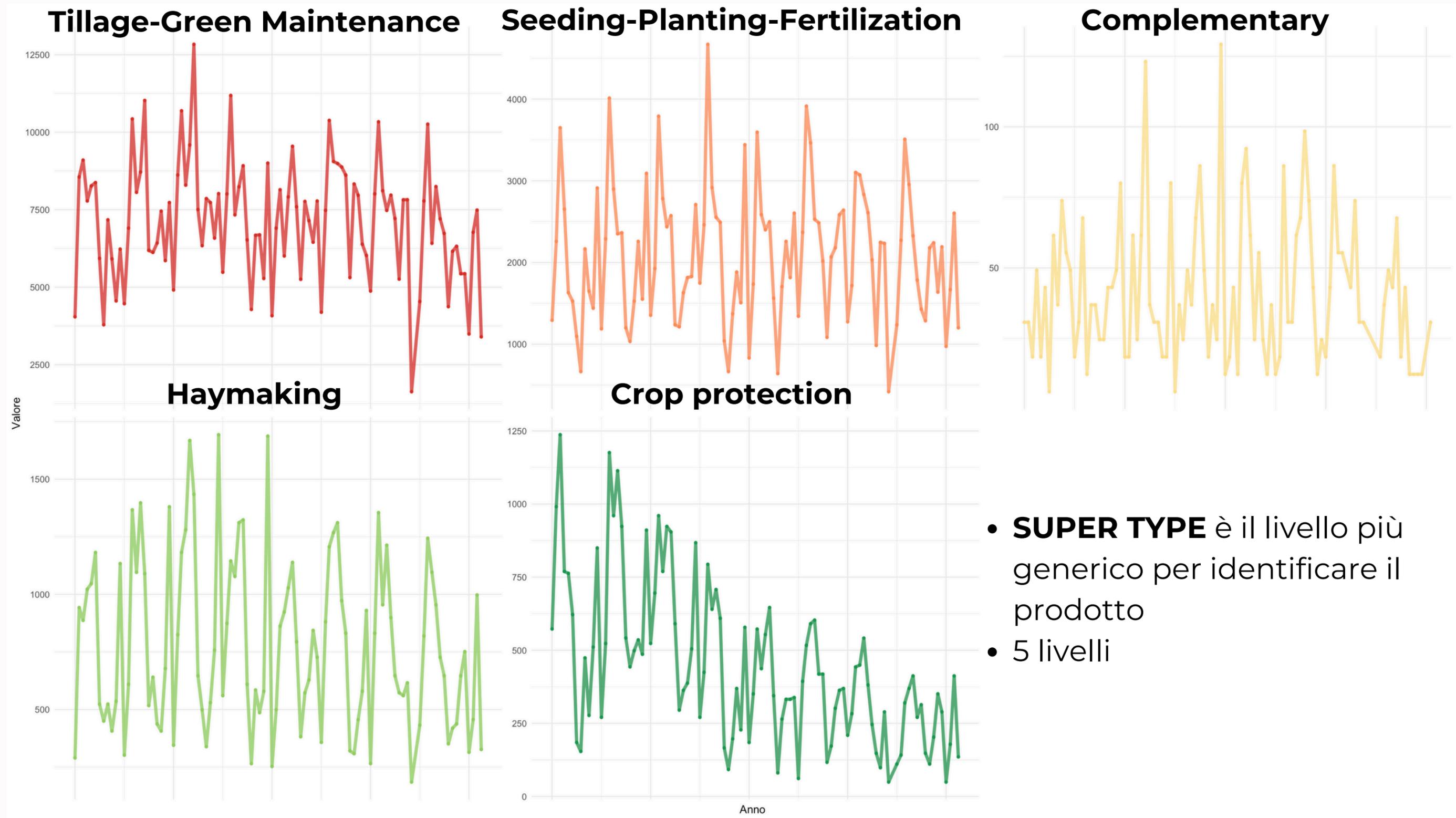
Cina



USA



Quantità totale venduta divisa per SUPER TYPE



Domande di ricerca

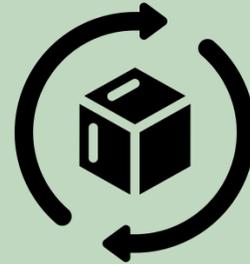
01



Quali sono i prodotti che vengono comprati dallo stesso cliente?

- Social network analysis
- Market Basket analysis

02



In che fase di ciclo di vita sono i prodotti?

- Cluster analysis

03



Qual è l'andamento delle vendite in Italia?

- Modelli di serie storiche
- Integrazione con dati Istat
- Modelli gerarchici
- Modelli di Data Mining

Domande di ricerca

01. Quali prodotti vengono comprati insieme?

Social network Analysis

Market Basket Analysis

Social Network Analysis

- Rete globale dei prodotti
- Community detection
- Reti dinamiche
- Indici di rete
- Modelli per analisi di rete

Basket Analysis

- Creazioni delle matrici di transazioni
- Individuazione delle associazioni tra prodotti.

Social Network Analysis

Rete globale

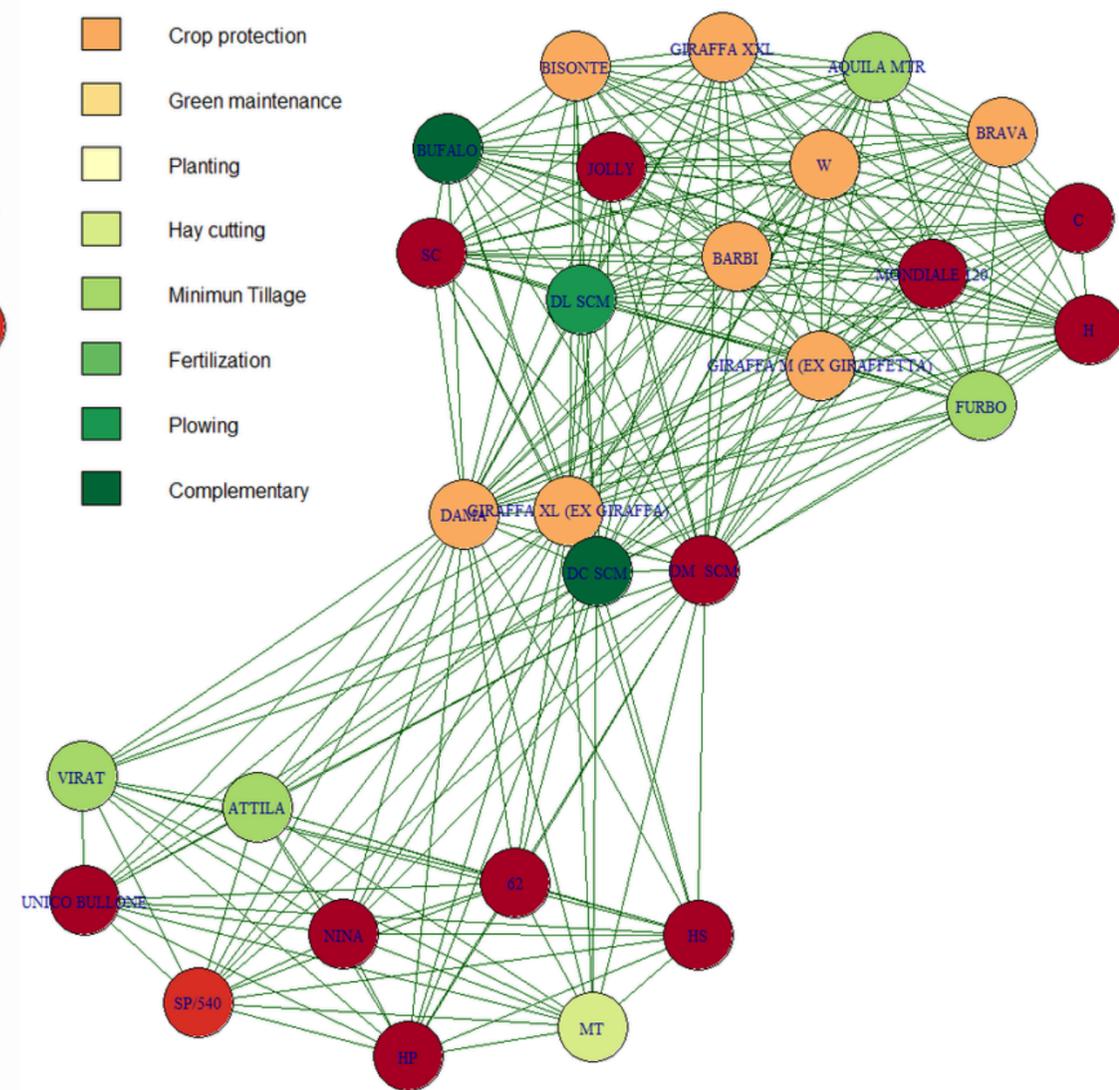
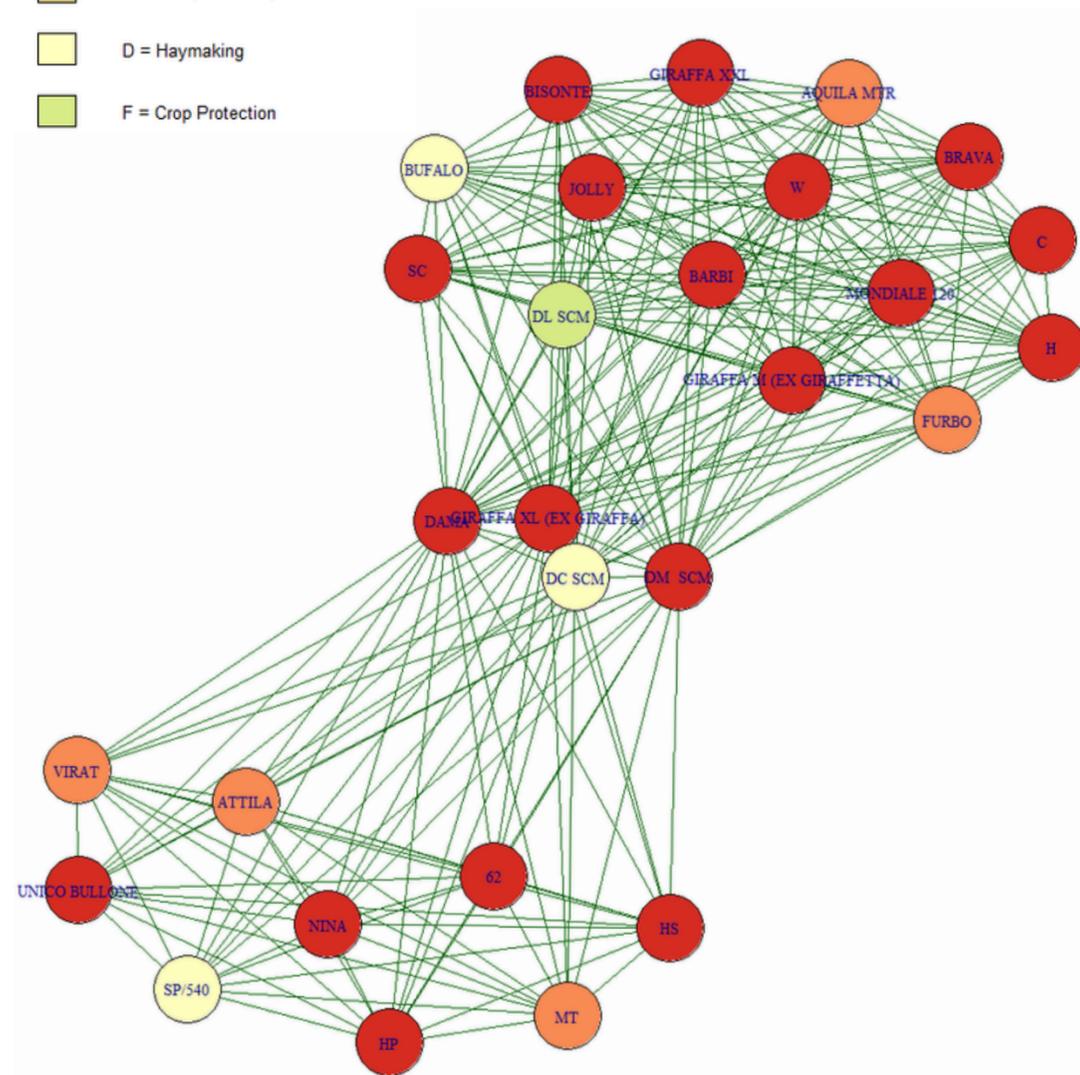
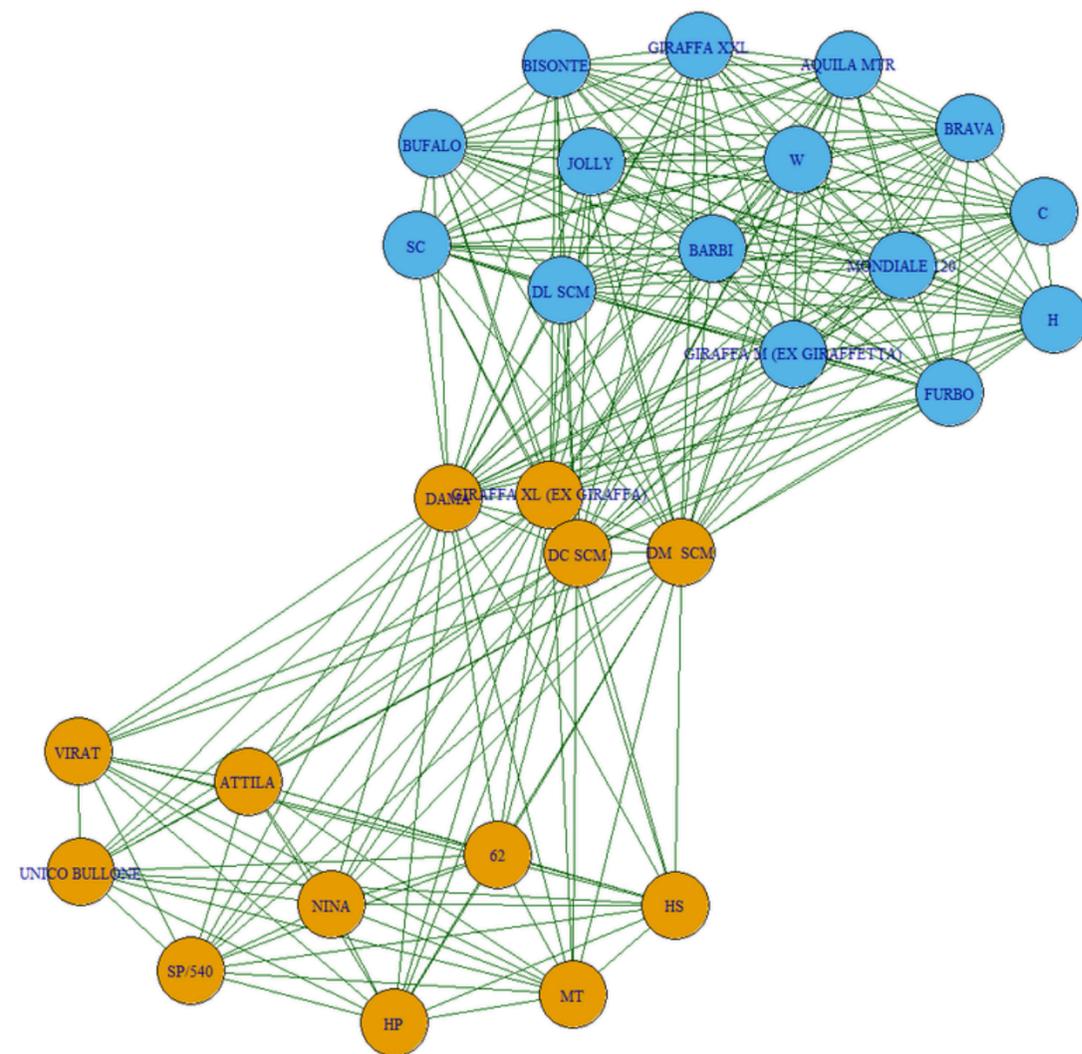


Node	Degree	Betweenness	Closeness
GIRAFFA XL	27	33.75	0.03333333
DAMA	27	33.75	0.03333333
DC SCM	27	33.75	0.03333333
DM SCM	27	33.75	0.03333333
H	18	0.24	0.02777778

- Rete **indiretta** e **pesata**
- **Nodi**: prodotti sotto forma di **MODEL** comprati almeno **150** volte dallo stesso cliente
- **Archi** pesati in base al numero di volte in cui due prodotti sono stati comprati insieme con un **numero minimo pari a 10**
- **Densità** della rete: **0.6428571**



Community Detection



- Nodi classificati tramite il metodo di **Louvain**
- Assortatività: 0.5

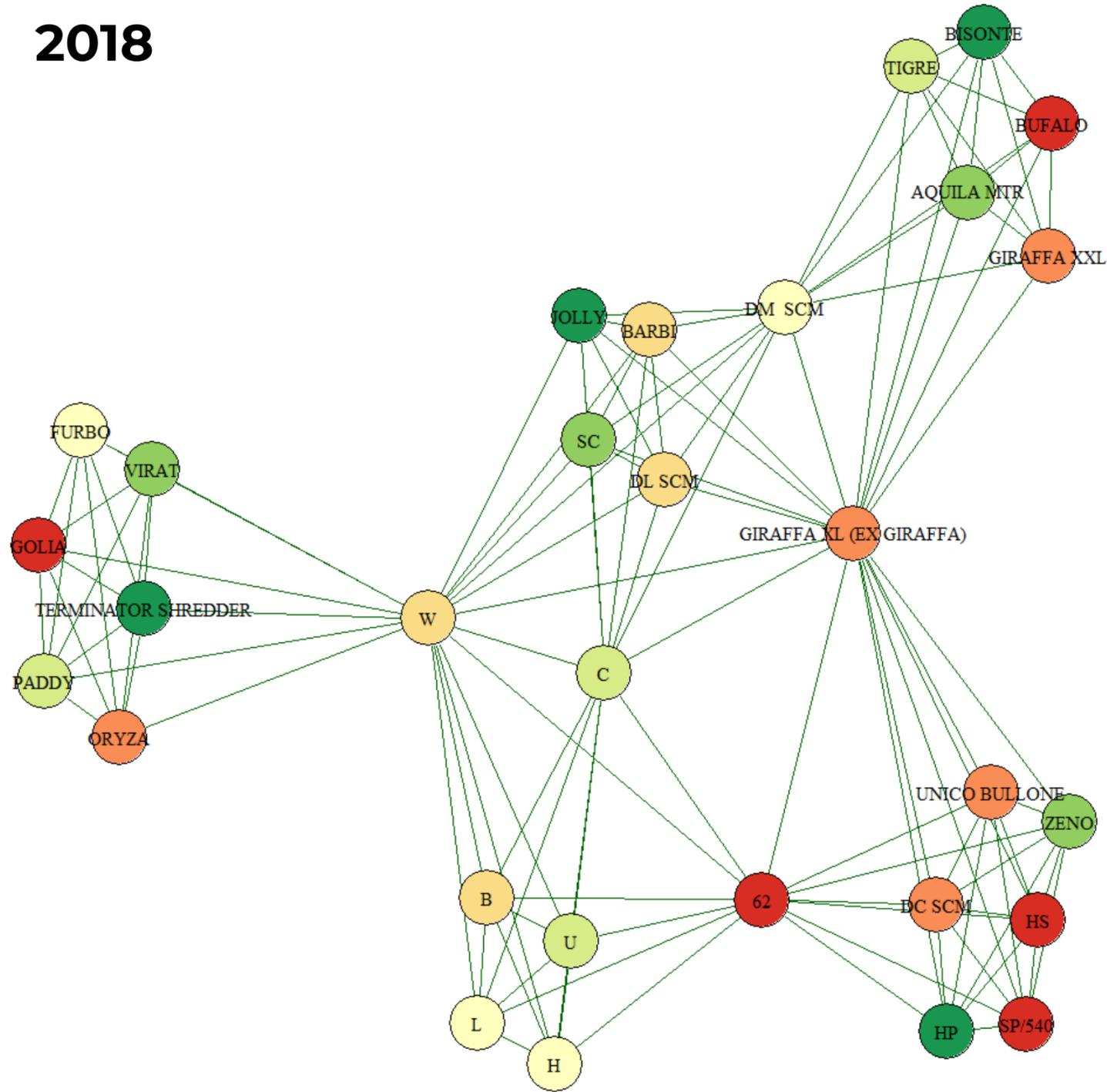
- Nodi classificati per **SUPER TYPE**

- Nodi classificati per **FAMILY**

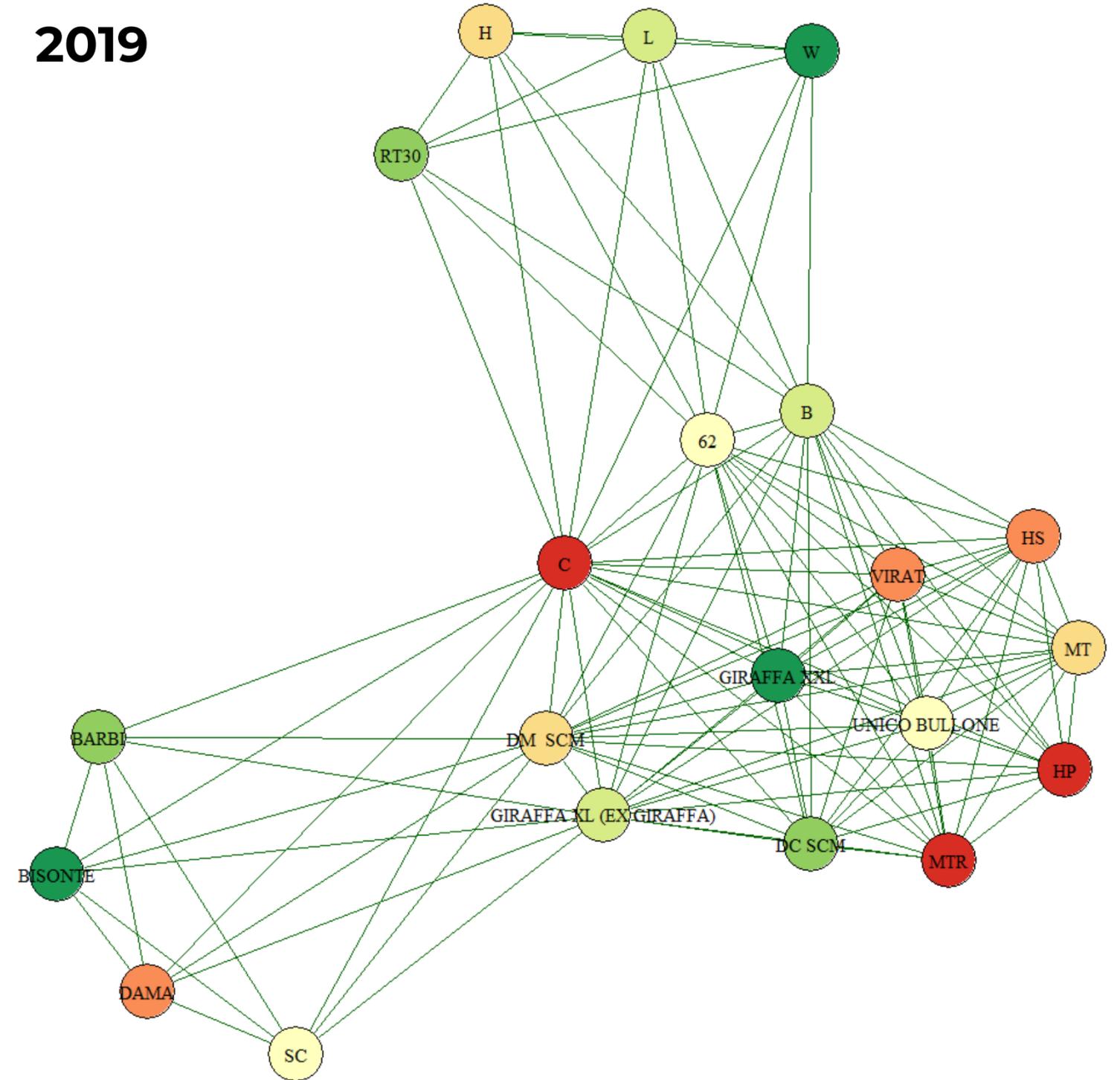
Reti di **MODEL** per anno

Prodotti venduti almeno **30 volte** nell'arco dell'anno e comprati **almeno 5 volte insieme** ad altri prodotti dallo stesso cliente

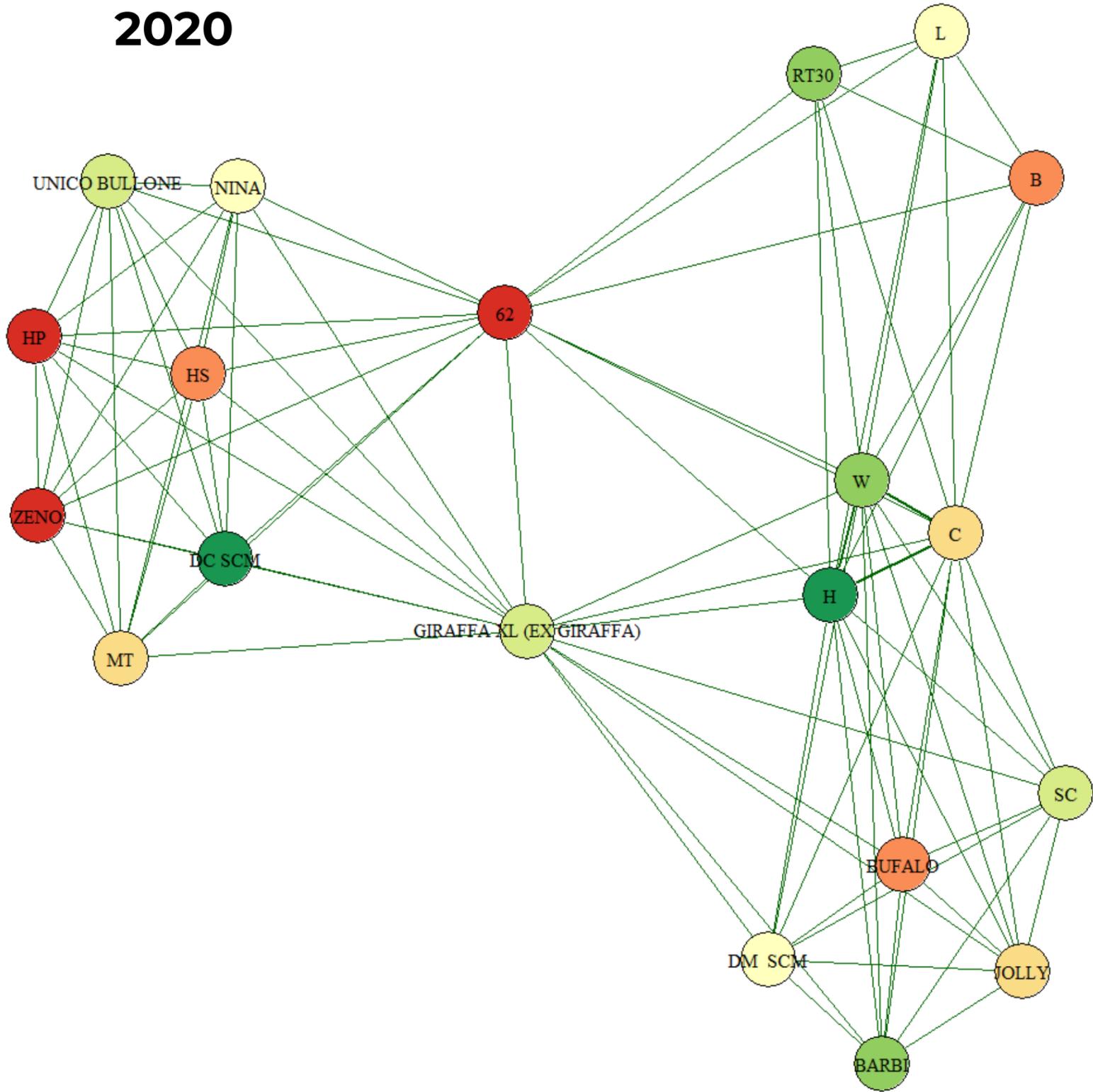
2018



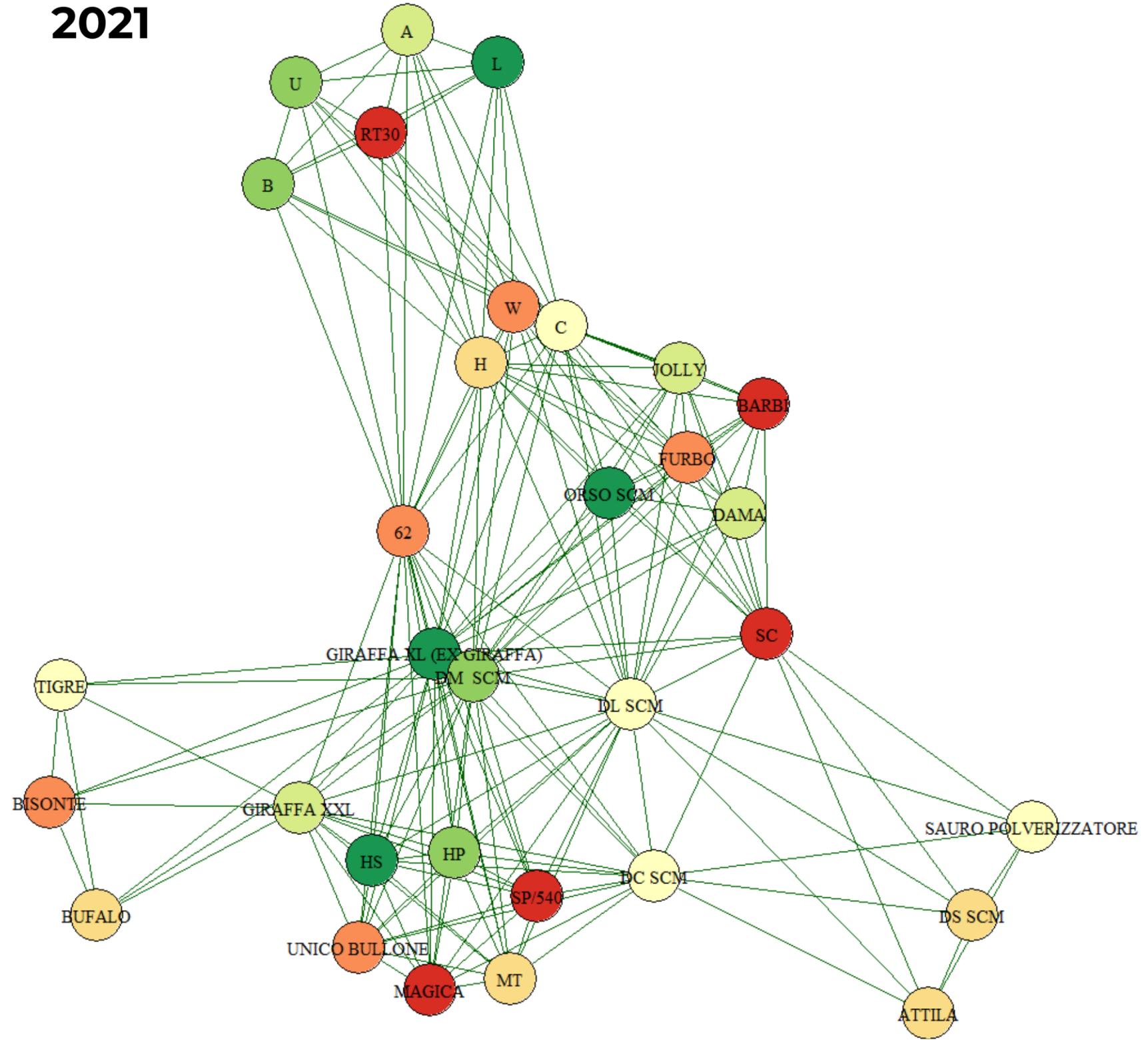
2019



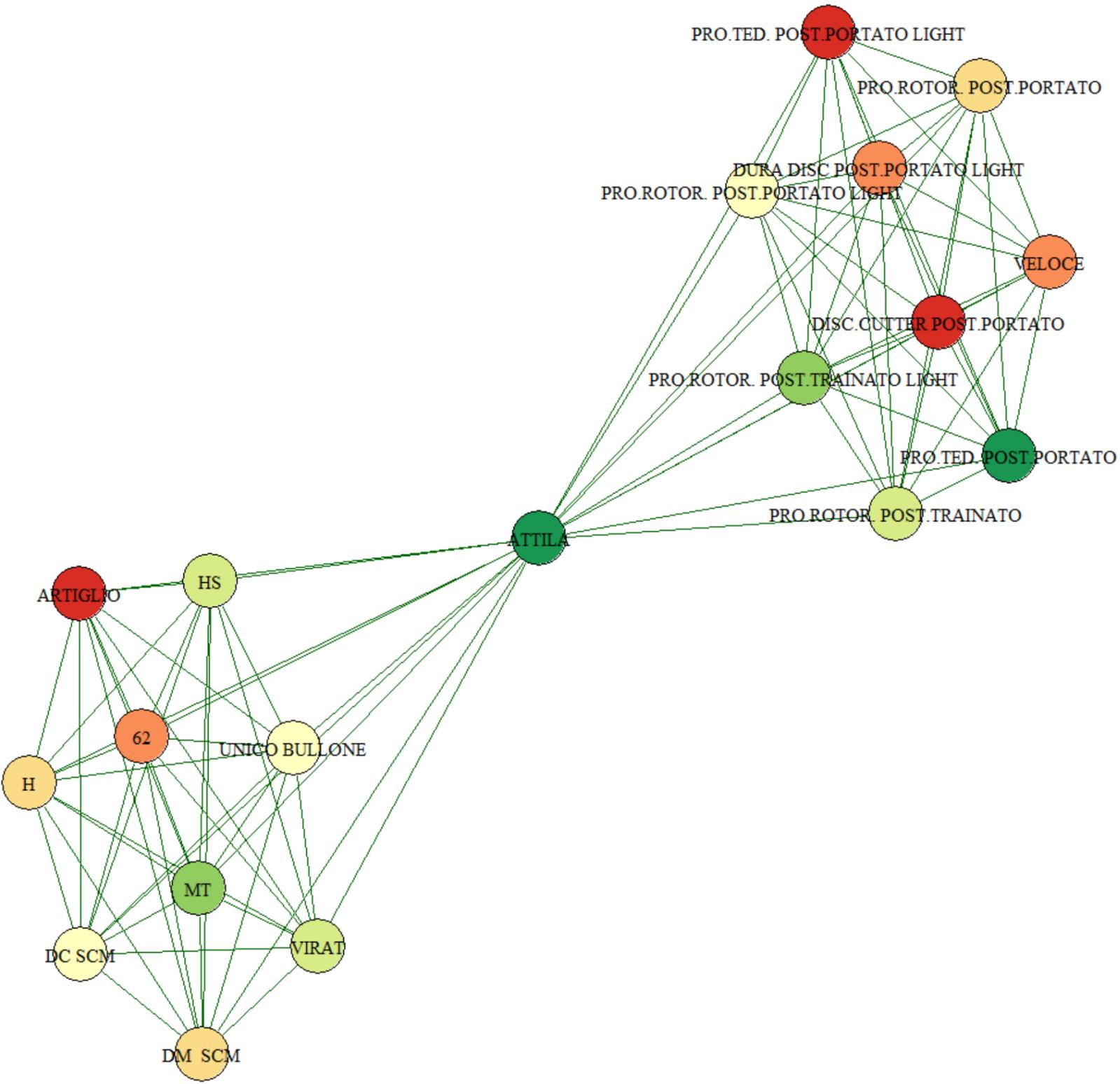
2020



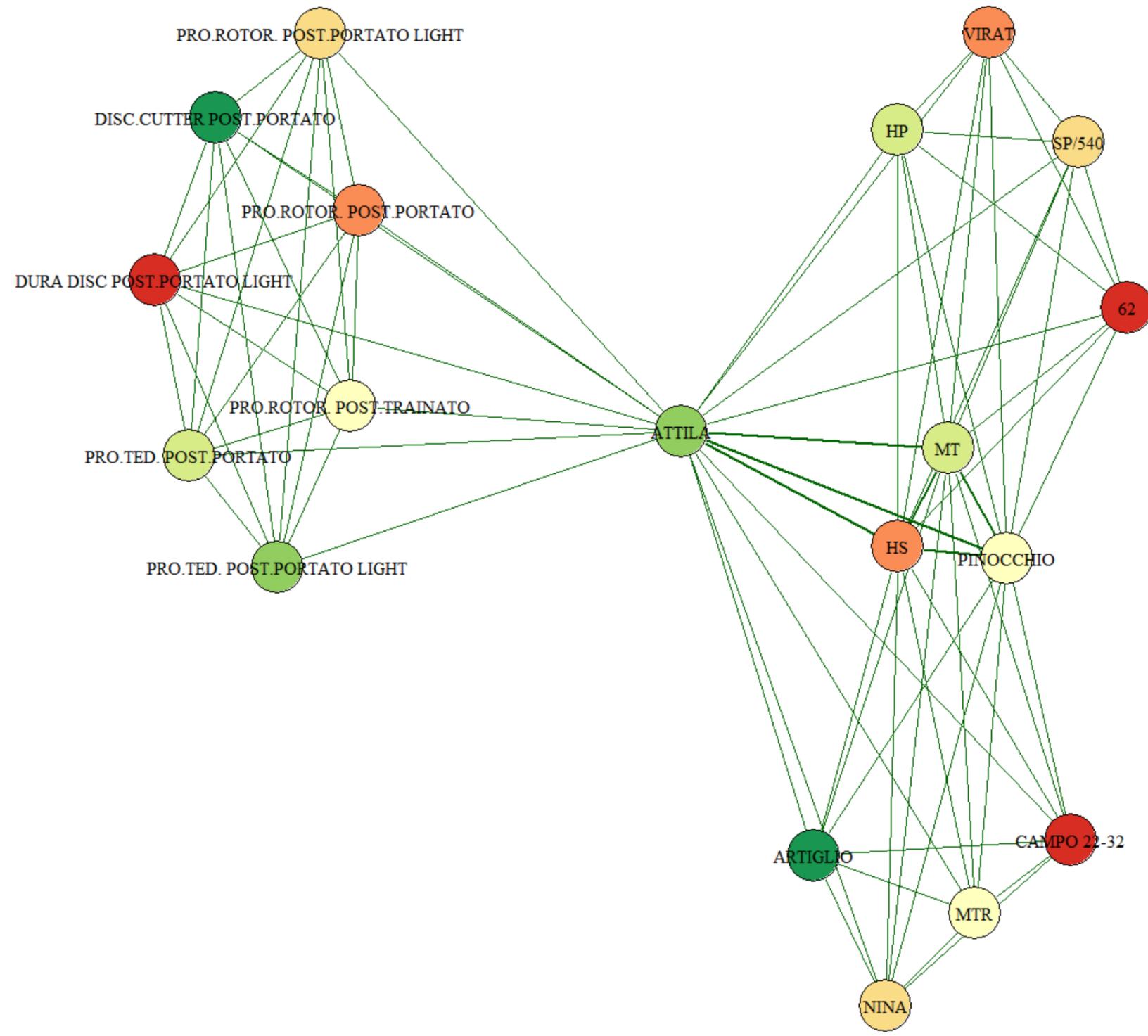
2021



2022

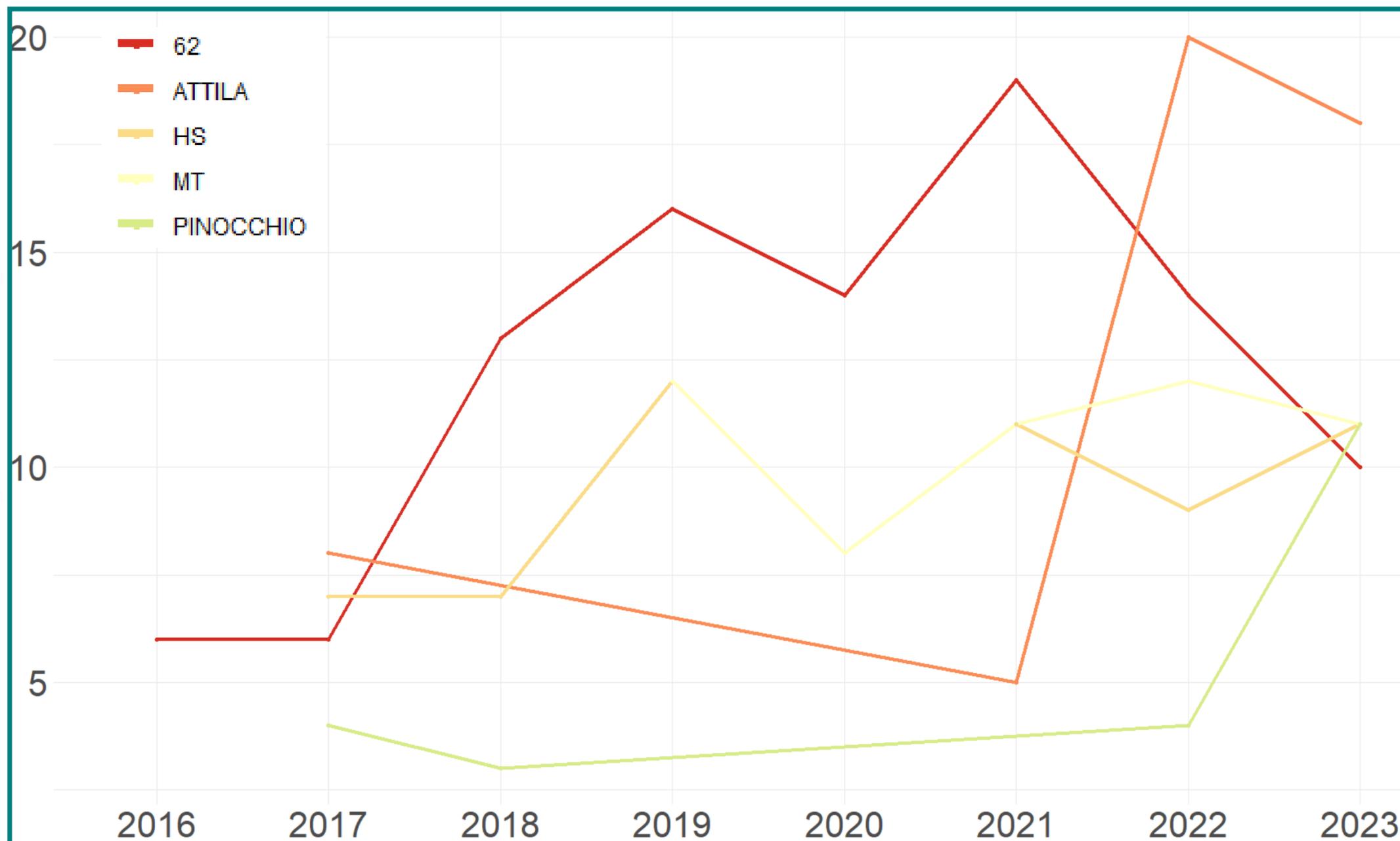


2023



Indici descrittivi della rete: grado

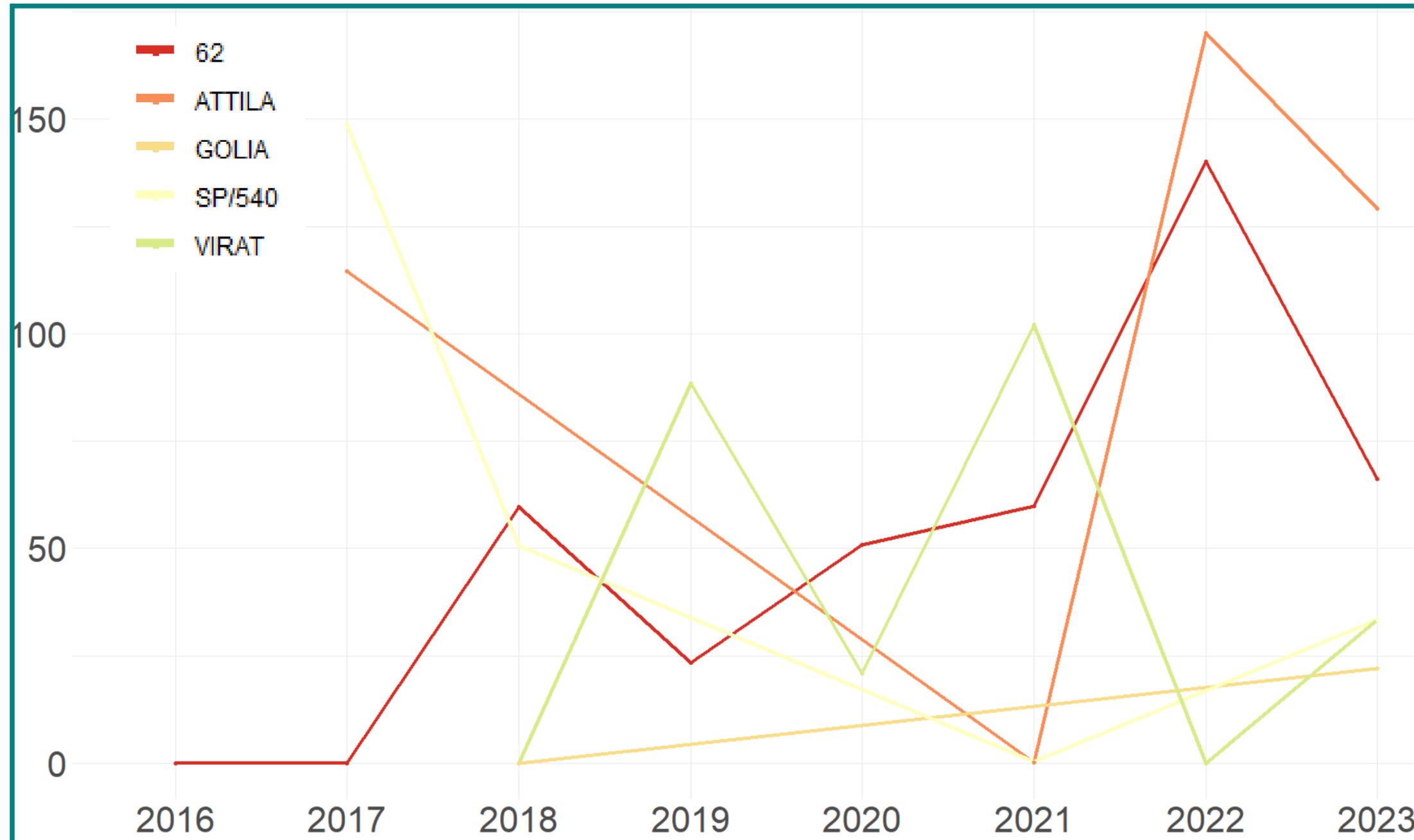
Andamento nel tempo del **grado** dei prodotti con grado più alto comprati almeno 30 volte dallo stesso cliente e almeno 3 volte insieme ad altri prodotti nell'arco del 2023



Node	Degree	Weighted Degree
ATTILA	18	21
MT	11	14
HS	11	14
PINOCCHIO	11	14
62	10	10

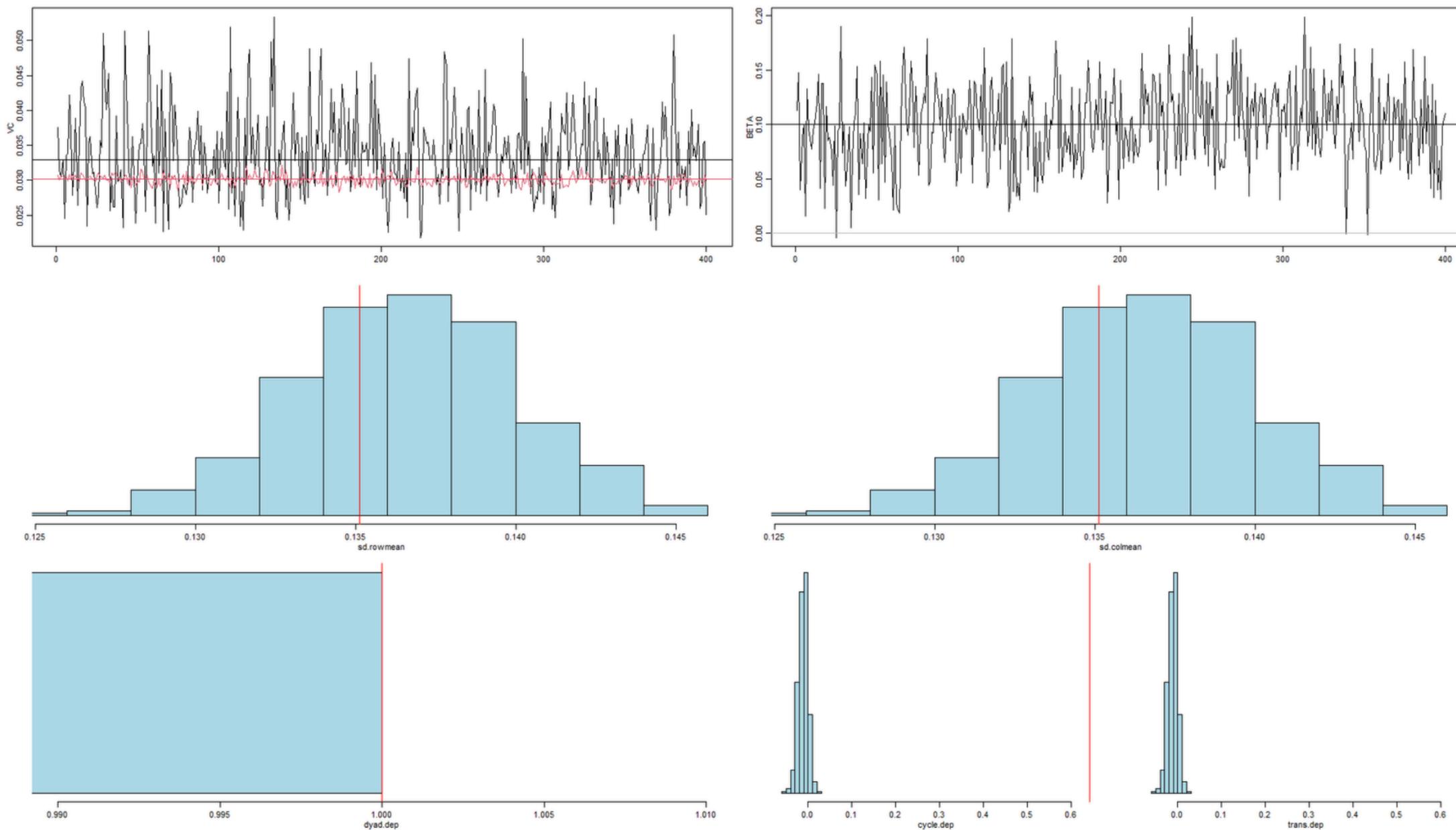
Indici descrittivi della rete: betweenness

Andamento nel tempo della **betweenness** dei prodotti con betweenness più alta comprati almeno 30 volte dallo stesso cliente e almeno 3 volte insieme ad altri prodotti nell'arco del 2023



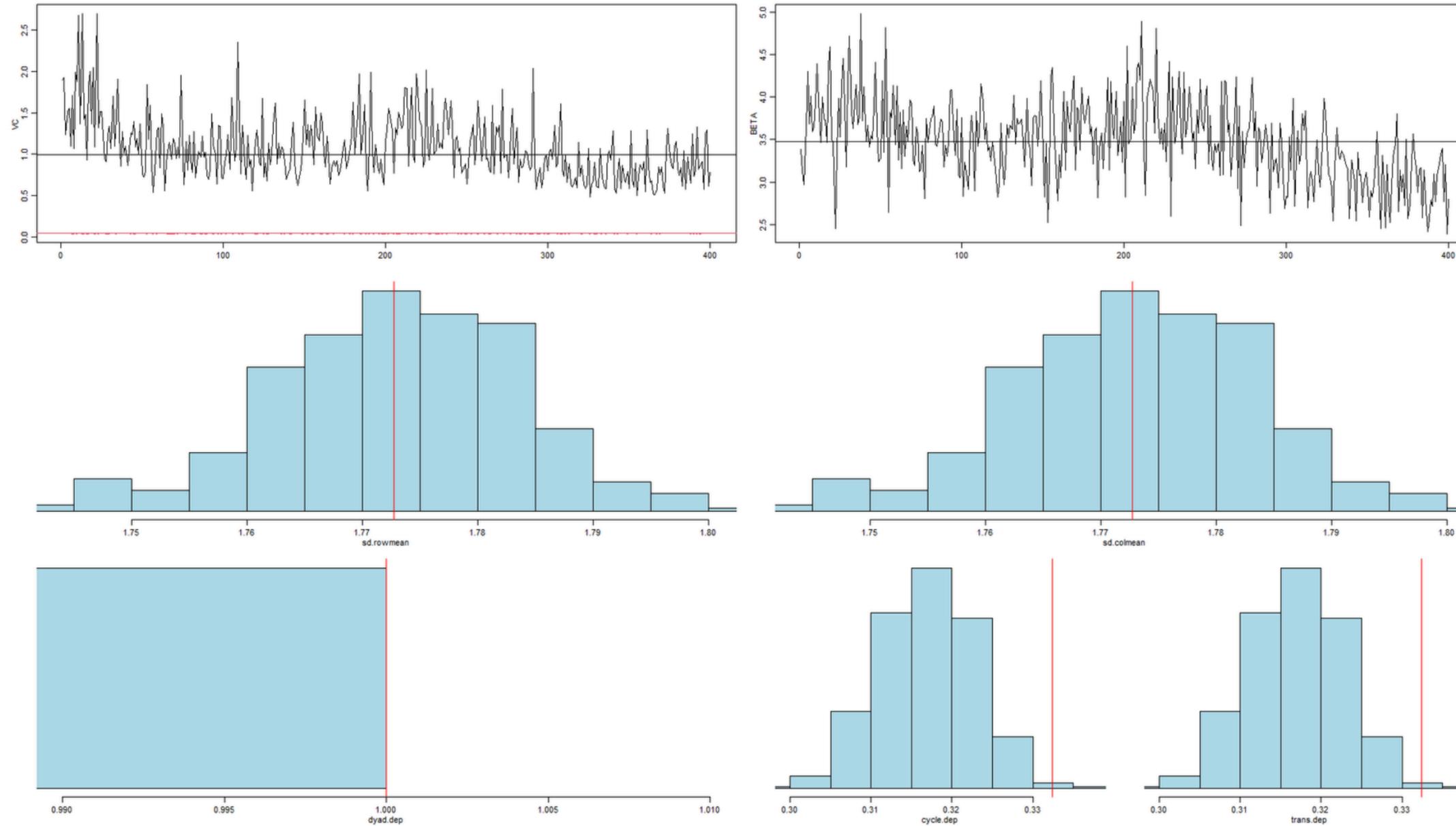
Node	Betweenness
ATTILA	129.1428571
62	66.1428571
VIRAT	33.33333333
SP/540	33.33333333
GOLIA	22.1428571

Social Relations Model



$$\begin{aligned}
 y_{i,j} &= \mu + a_i + b_j + \varepsilon_{i,j} \\
 \{(a_1, b_1), \dots, (a_n, b_n)\} &\sim N(0, \Sigma_{ab}) \quad i.i.d. \\
 \{(\varepsilon_{i,j}, \varepsilon_{j,i})\} &\sim N(0, \Sigma_\varepsilon) \quad i.i.d.
 \end{aligned}
 \quad
 \Sigma_{ab} = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}
 \quad
 \Sigma_\varepsilon = \sigma_\varepsilon^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Additive and multiplicative effects model, AME



$$y_{i,j} = \beta^{\top} \mathbf{x}_{i,j} + a_i + b_j + \mathbf{u}_i^{\top} \mathbf{v}_j + \varepsilon_{i,j}$$

Domande di ricerca

01. Quali prodotti vengono comprati insieme?

Social network Analysis

Market Basket Analysis

Social Network Analysis

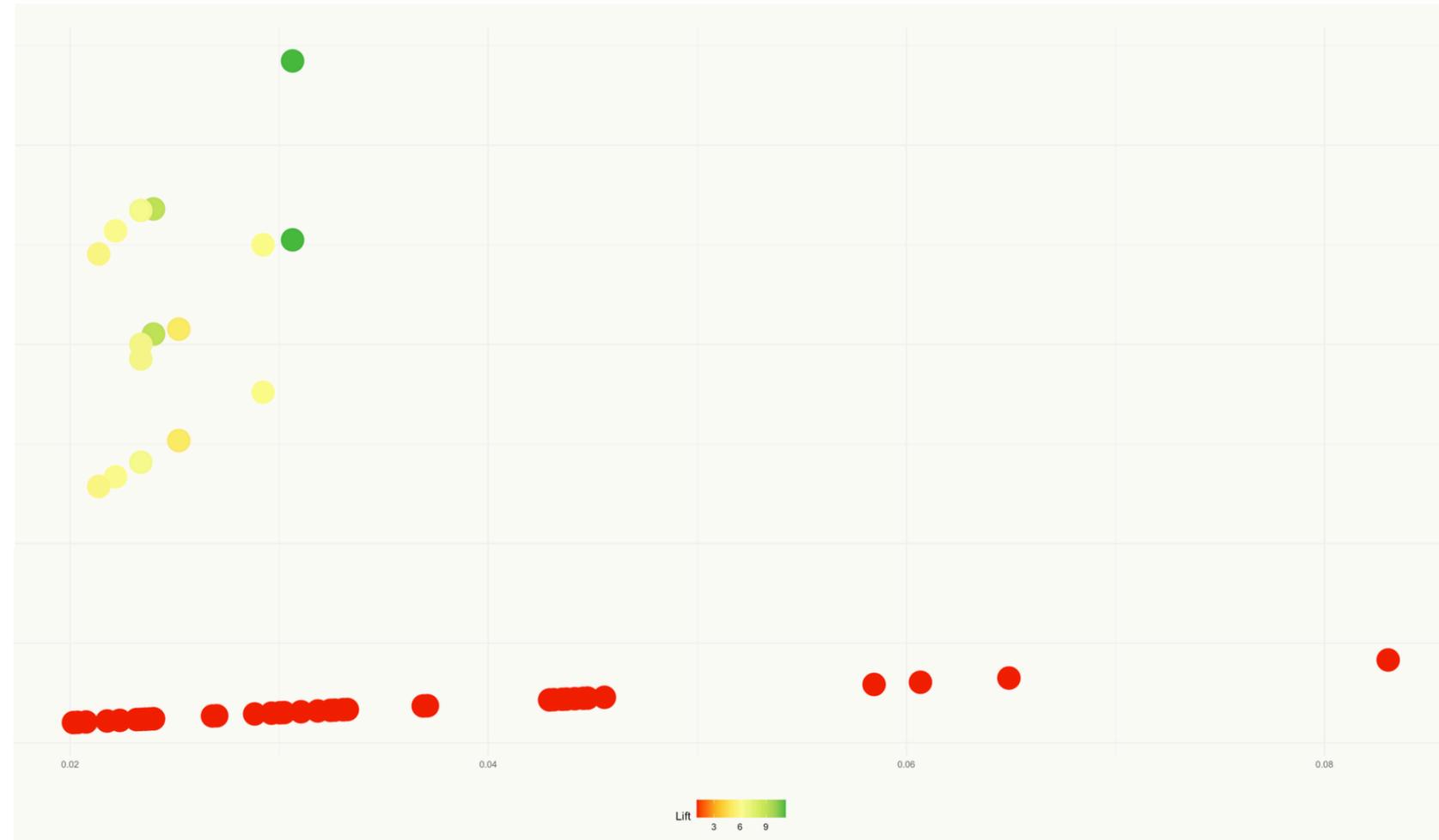
- Motivazioni
- Rete globale dei prodotti
- Community detection
- Reti dinamiche
- Indici di rete
- Modelli per analisi di rete

Basket Analysis

- Motivazioni
- Creazioni delle matrici di transizioni
- Individuazione delle associazioni tra prodotti.

Market Basket Analysis

2020



supporto=0.02
confidenza=0.01
MODEL

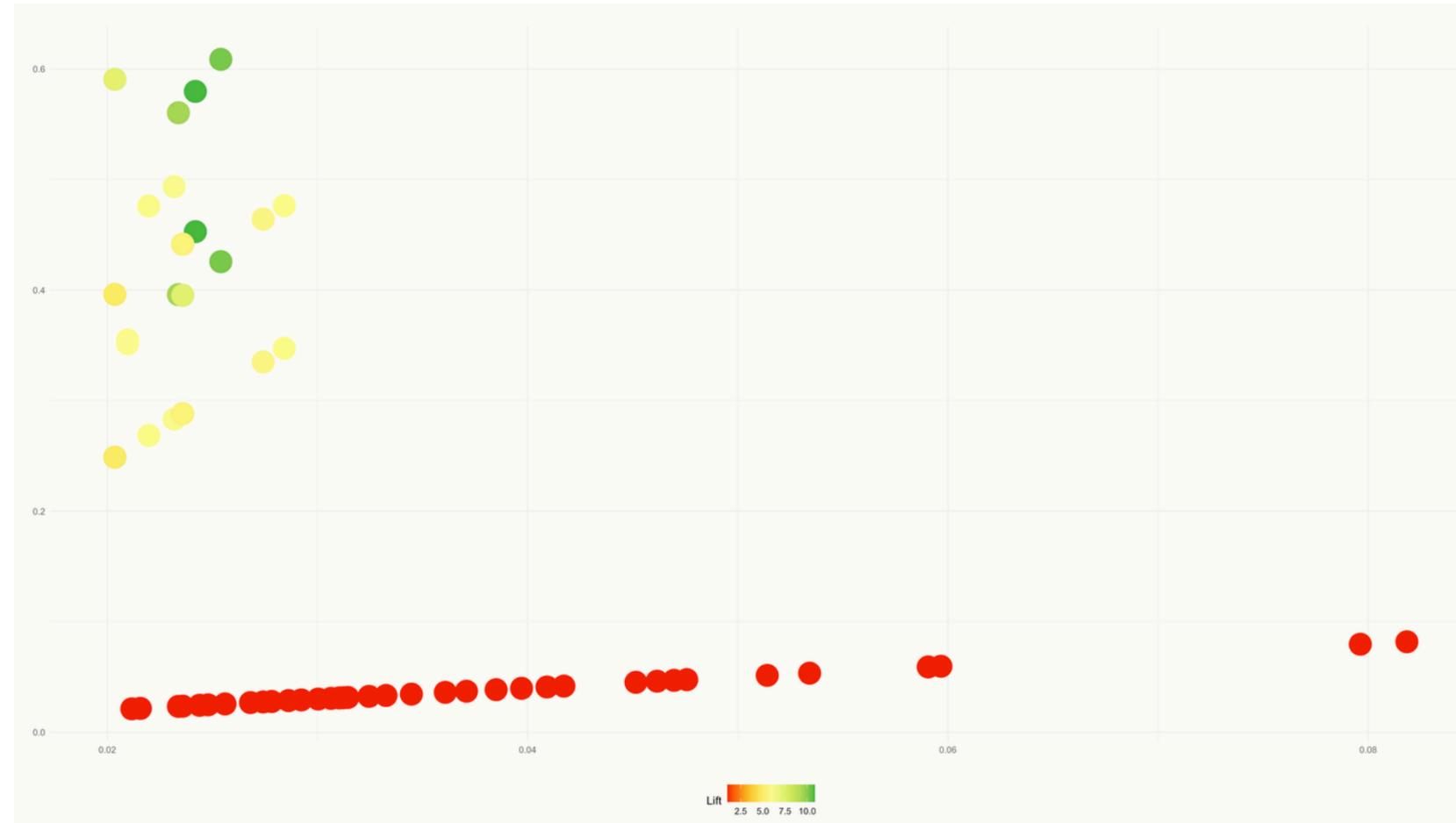
55 associazioni
con lift > 1

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{DAMA}	=> {DC SCM}	0.03063898	0.6846847	0.04474904	11.284786	152
[2]	{DAMA}	=> {DM SCM}	0.02398710	0.5360360	0.04474904	9.169913	119
[3]	{DC SCM}	=> {DAMA}	0.03063898	0.5049834	0.06067325	11.284786	152
[4]	{DM SCM}	=> {DAMA}	0.02398710	0.4103448	0.05845596	9.169913	119

lift>9

Market Basket Analysis

2021



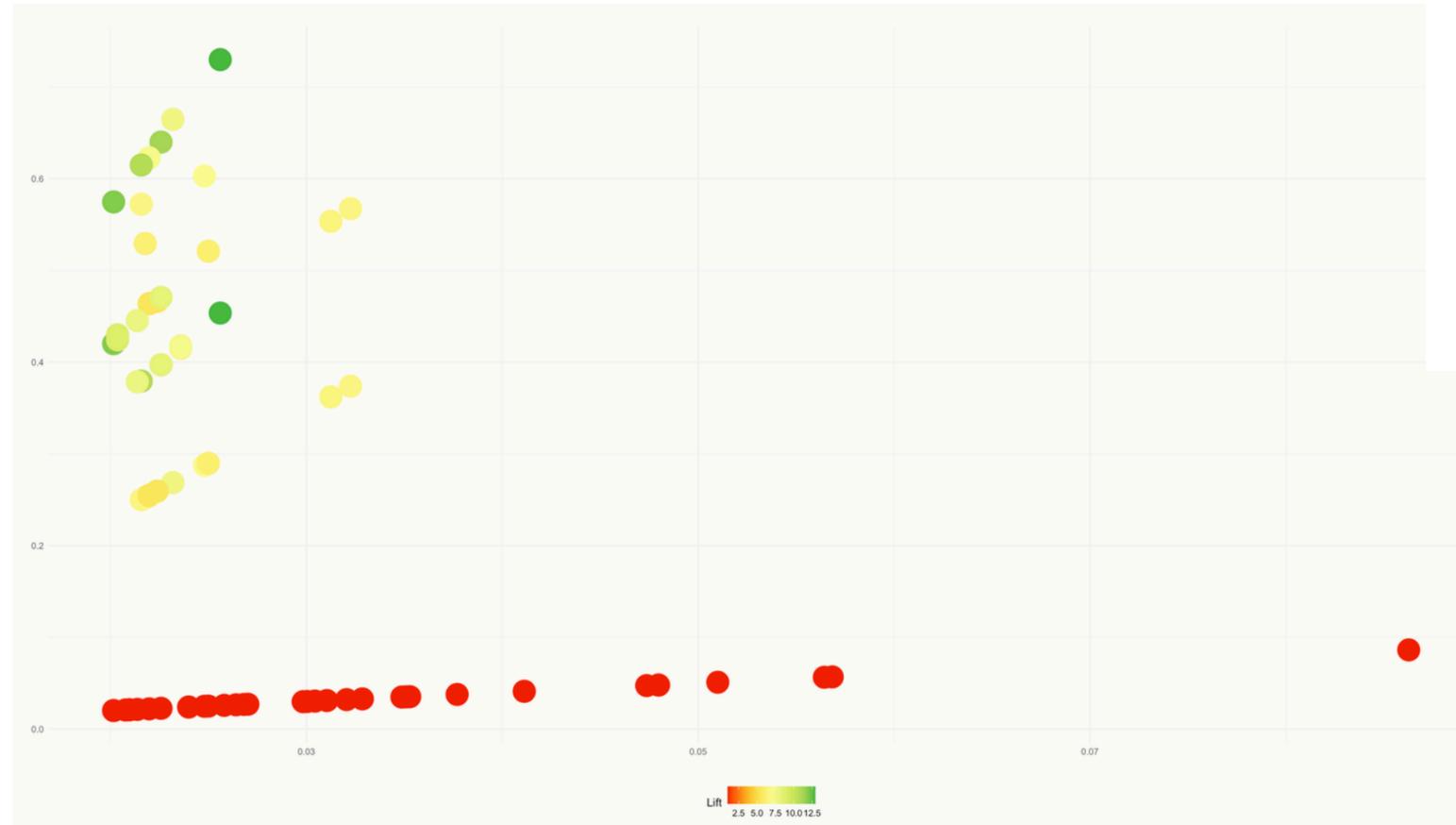
64 associazioni con lift > 1

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{DAMA}	=> {BISONTE}	0.02539811	0.6086957	0.04172546	10.201821	126
[2]	{DAMA}	=> {DC SCM}	0.02418867	0.5797101	0.04172546	10.852611	120
[3]	{DAMA}	=> {DM SCM}	0.02338238	0.5603865	0.04172546	9.488318	116
[4]	{DC SCM}	=> {DAMA}	0.02418867	0.4528302	0.05341665	10.852611	120
[5]	{BISONTE}	=> {DAMA}	0.02539811	0.4256757	0.05966539	10.201821	126
[6]	{DM SCM}	=> {DAMA}	0.02338238	0.3959044	0.05906067	9.488318	116

lift>9

Market Basket Analysis

2022

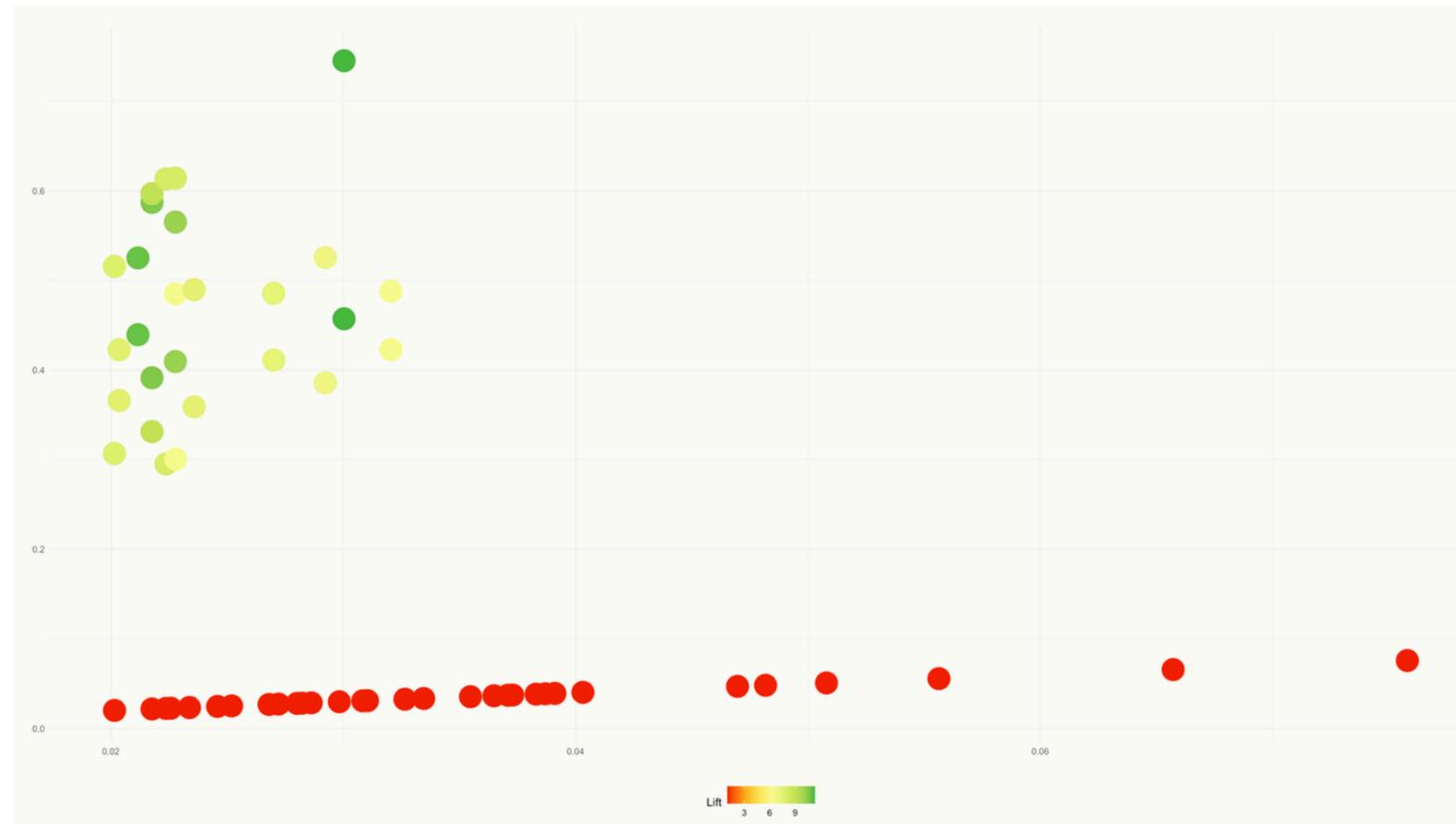


71 associazioni con lift > 1

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{DAMA}	=> {DM SCM}	0.02559968	0.7298851	0.03507357	12.93200	127
[2]	{BRAVA LARGE}	=> {BISONTE}	0.02257609	0.6400000	0.03527515	11.25901	112
[3]	{DAMA}	=> {BISONTE}	0.02156823	0.6149425	0.03507357	10.81819	107
[4]	{DAMA}	=> {DC SCM}	0.02015723	0.5747126	0.03507357	11.97962	100
[5]	{DM SCM}	=> {DAMA}	0.02559968	0.4535714	0.05644023	12.93200	127
[6]	{DC SCM}	=> {DAMA}	0.02015723	0.4201681	0.04797420	11.97962	100
[7]	{BISONTE}	=> {BRAVA LARGE}	0.02257609	0.3971631	0.05684338	11.25901	112
[8]	{BISONTE}	=> {DAMA}	0.02156823	0.3794326	0.05684338	10.81819	107

lift>9

Market Basket Analysis 2023



61 associazioni con lift > 1

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{DAMA}	=> {DM SCM}	0.03003427	0.7450000	0.04031445	11.337255	149
[2]	{AQUILA MTR}	=> {DM SCM}	0.02176980	0.5966851	0.03648458	9.080229	108
[3]	{GIRAFFA XXL}	=> {BISONTE}	0.02176980	0.5869565	0.03708930	10.550331	108
[4]	{DAMA}	=> {BISONTE}	0.02277767	0.5650000	0.04031445	10.155670	113
[5]	{DAMA}	=> {DC SCM}	0.02116509	0.5250000	0.04031445	10.897594	105
[6]	{DM SCM}	=> {DAMA}	0.03003427	0.4570552	0.06571256	11.337255	149
[7]	{DC SCM}	=> {DAMA}	0.02116509	0.4393305	0.04817577	10.897594	105
[8]	{BISONTE}	=> {DAMA}	0.02277767	0.4094203	0.05563394	10.155670	113
[9]	{BISONTE}	=> {GIRAFFA XXL}	0.02176980	0.3913043	0.05563394	10.550331	108
[10]	{DM SCM}	=> {AQUILA MTR}	0.02176980	0.3312883	0.06571256	9.080229	108

lift>9

Domande di ricerca

02.

In che fase di ciclo di vita sono i prodotti?

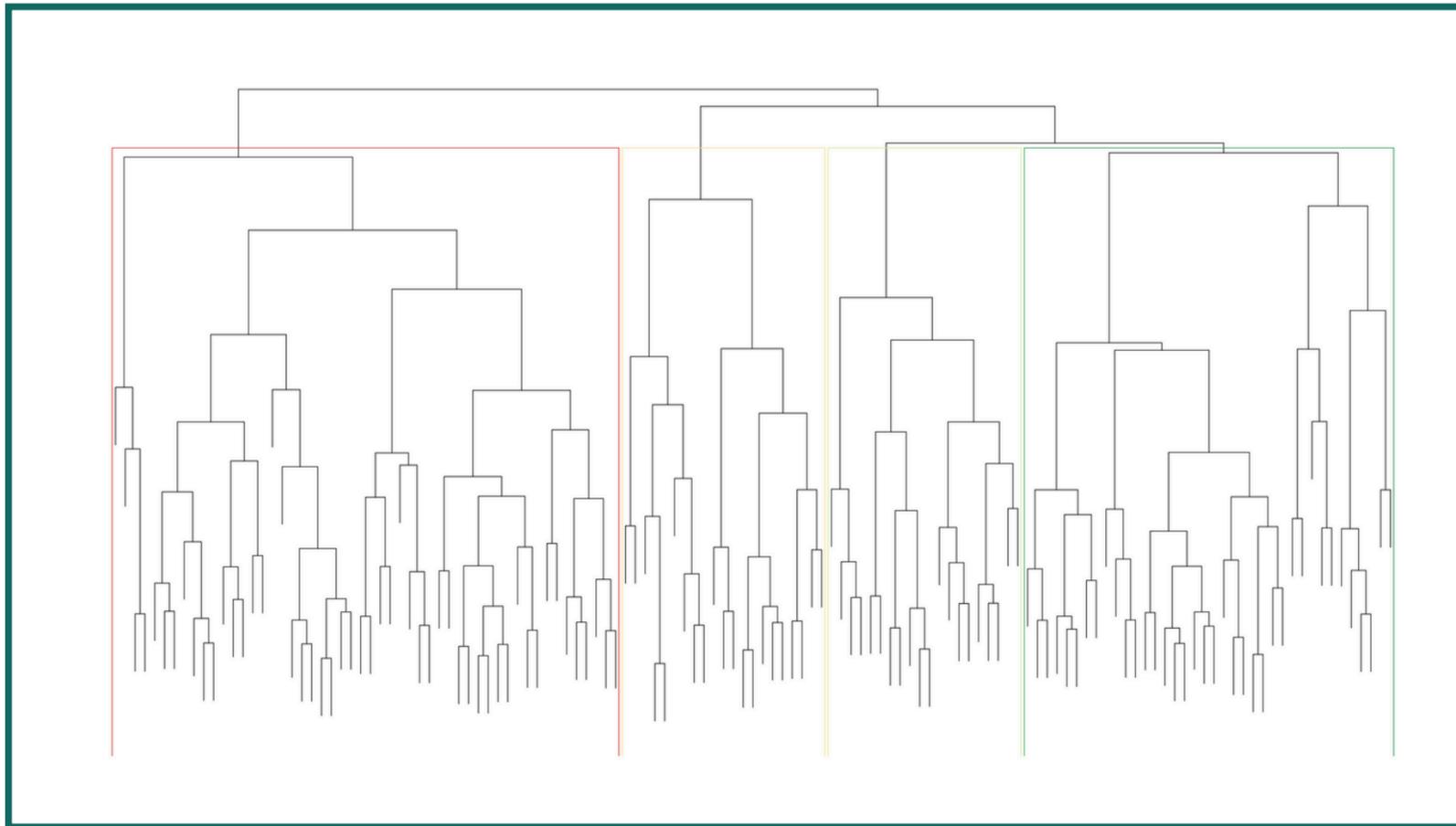
Cluster Analysis

Cluster analysis

- Motivazioni
- Modello di Bass
- Modello GGM
- Descrizione dei cluster

Cluster con il modello di Bass

$$z(t) = m(t) \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}}$$



Dendrogramma per 4 cluster usando il legame completo

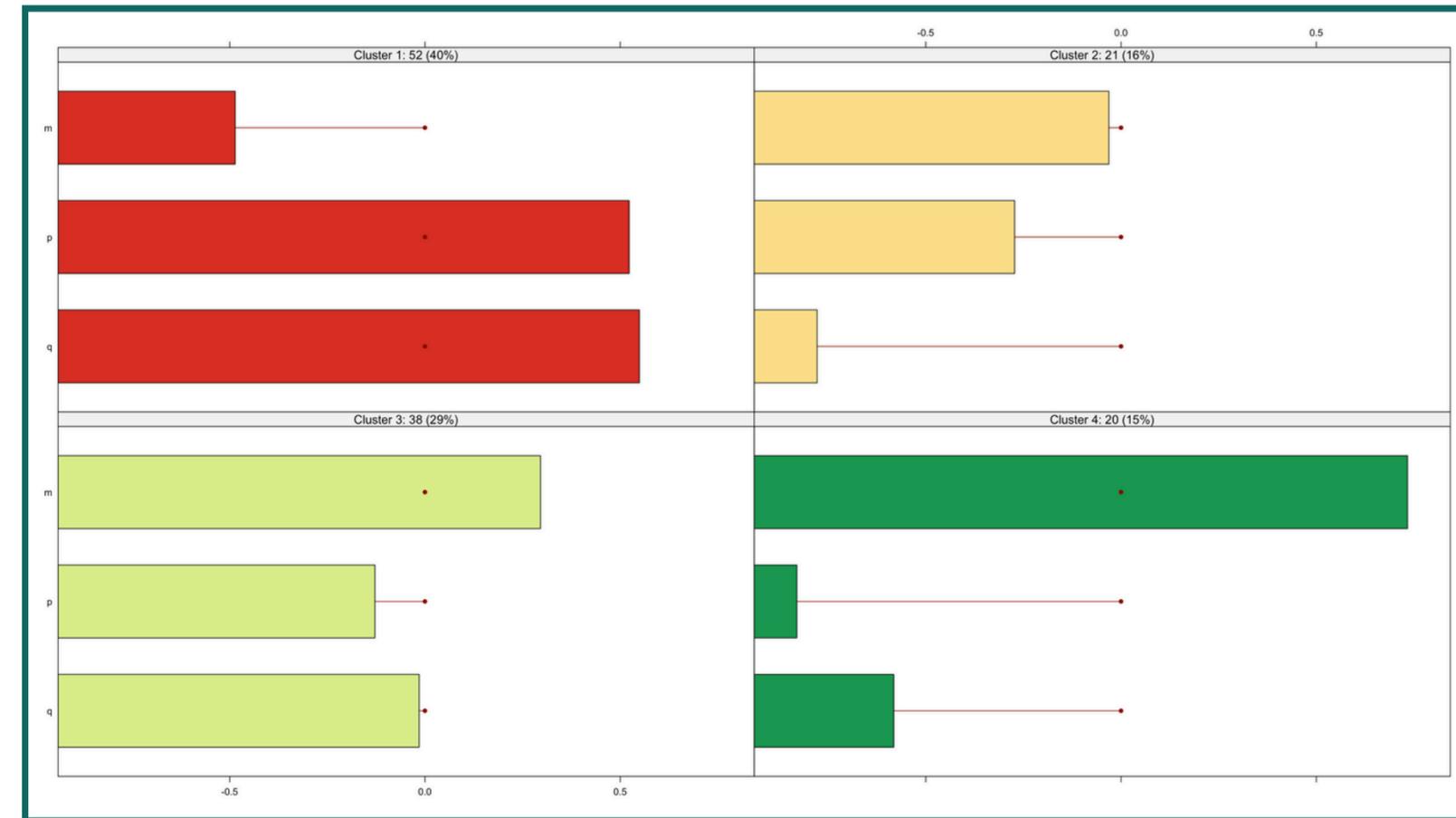
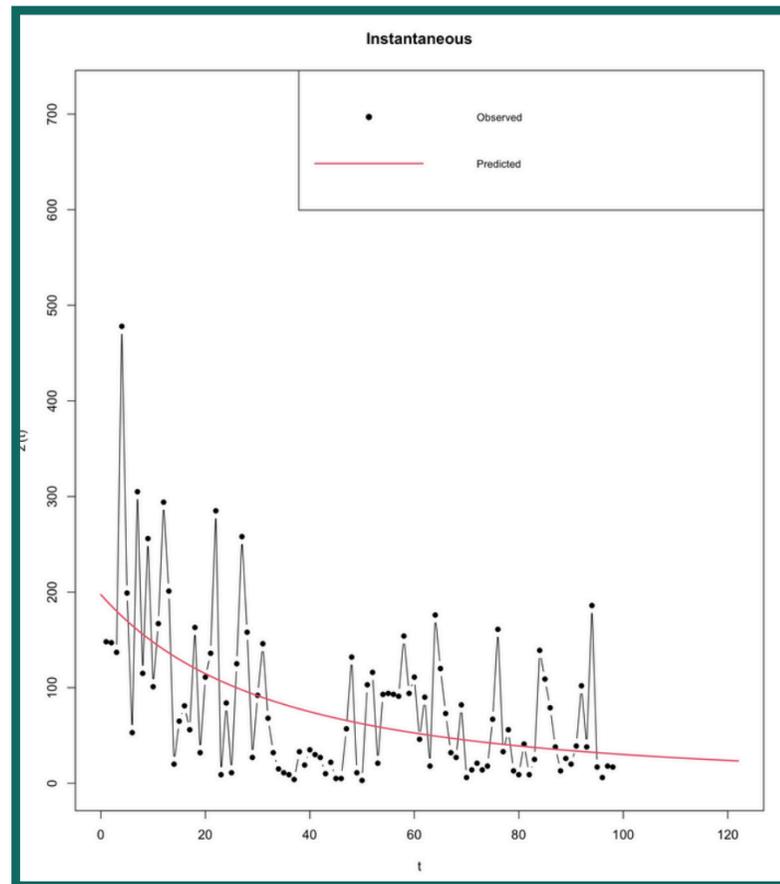


Grafico a barre per i 3 parametri (m, p, q) del modello di Bass scalati

Cluster 1



JOLLY

Caratteristiche:

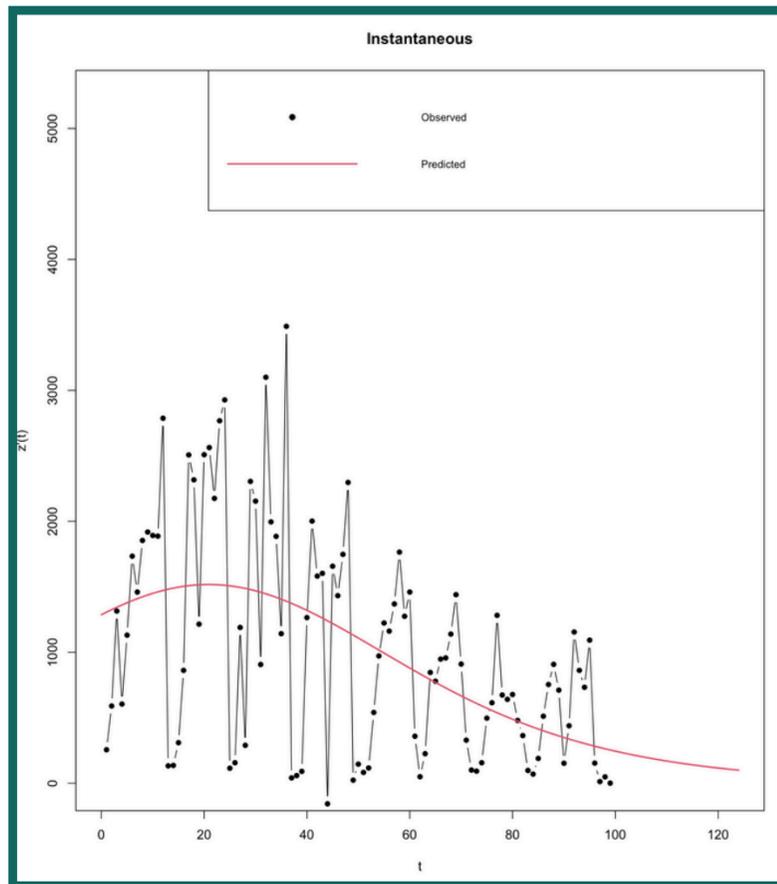
- m basso → Mercato potenziale limitato
- p, q alti → Forte attrattività iniziale, significativa influenza sociale

Strategie:

- Massimizzare adozione iniziale
- Sfruttare effetto dominio del passaparola
- Mercato che potrebbe raggiungere la saturazione

m	p	q
1307.823	0.0092	0.0399

Cluster 2



ROTARY TRILLERS H

Caratteristiche:

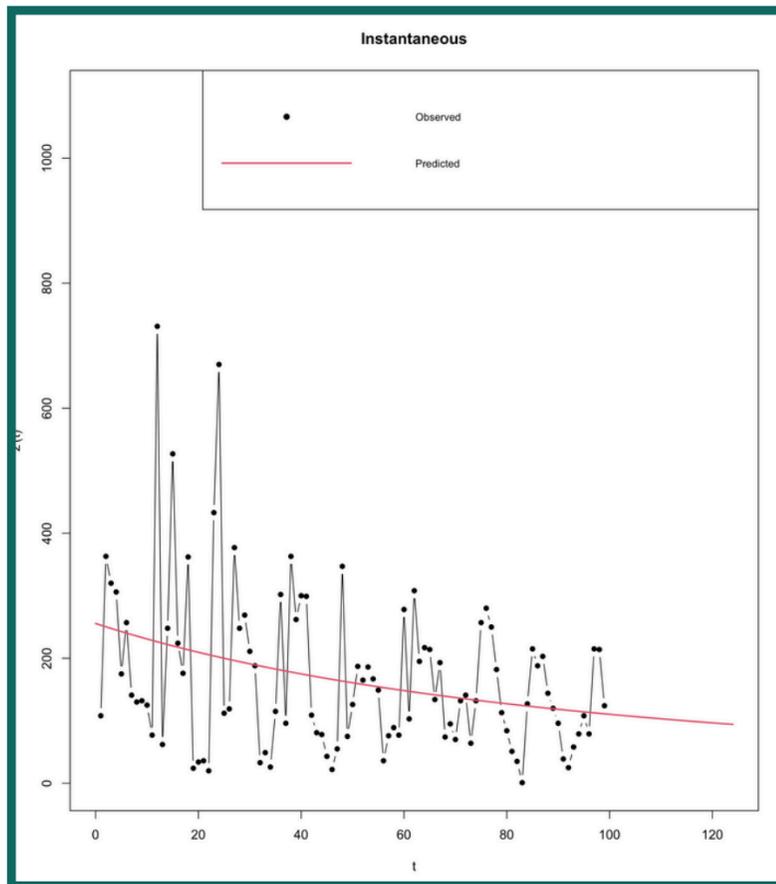
- m alto → Mercato potenziale ampio
- p medio → No adozione immediata su larga scala
- q basso → Adozione poco dipendente dalle interazioni tra consumatori.

Strategie:

- Attirare clientela tramite promozione continua.
- Potenziale a lungo termine significativo

m	p	q
1.1391.5	0.0043	-0.0035

Cluster 3



62

Caratteristiche:

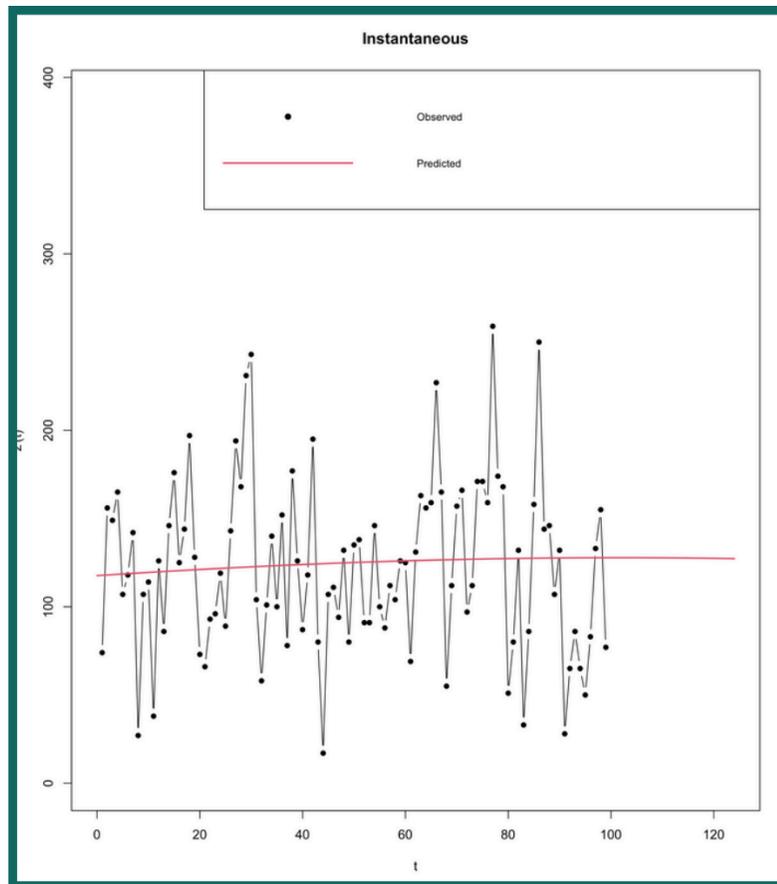
- m alto \longrightarrow Mercato potenziale ampio
- p, q medi \longrightarrow Bilanciamento tra attrattiva iniziale e influenza sociale.

Strategie:

- Combinazione di strategie che risaltino innovazione e incentivino il passaparola.

m	p	q
18644.6	0.0525	0.0214

Cluster 4



GIRAFFA XL

Caratteristiche:

- m molto elevato → Mercato potenziale molto grande
- p molto basso → Innovazione percepita poco significativa.
- q basso → Poco beneficio dal passaparola

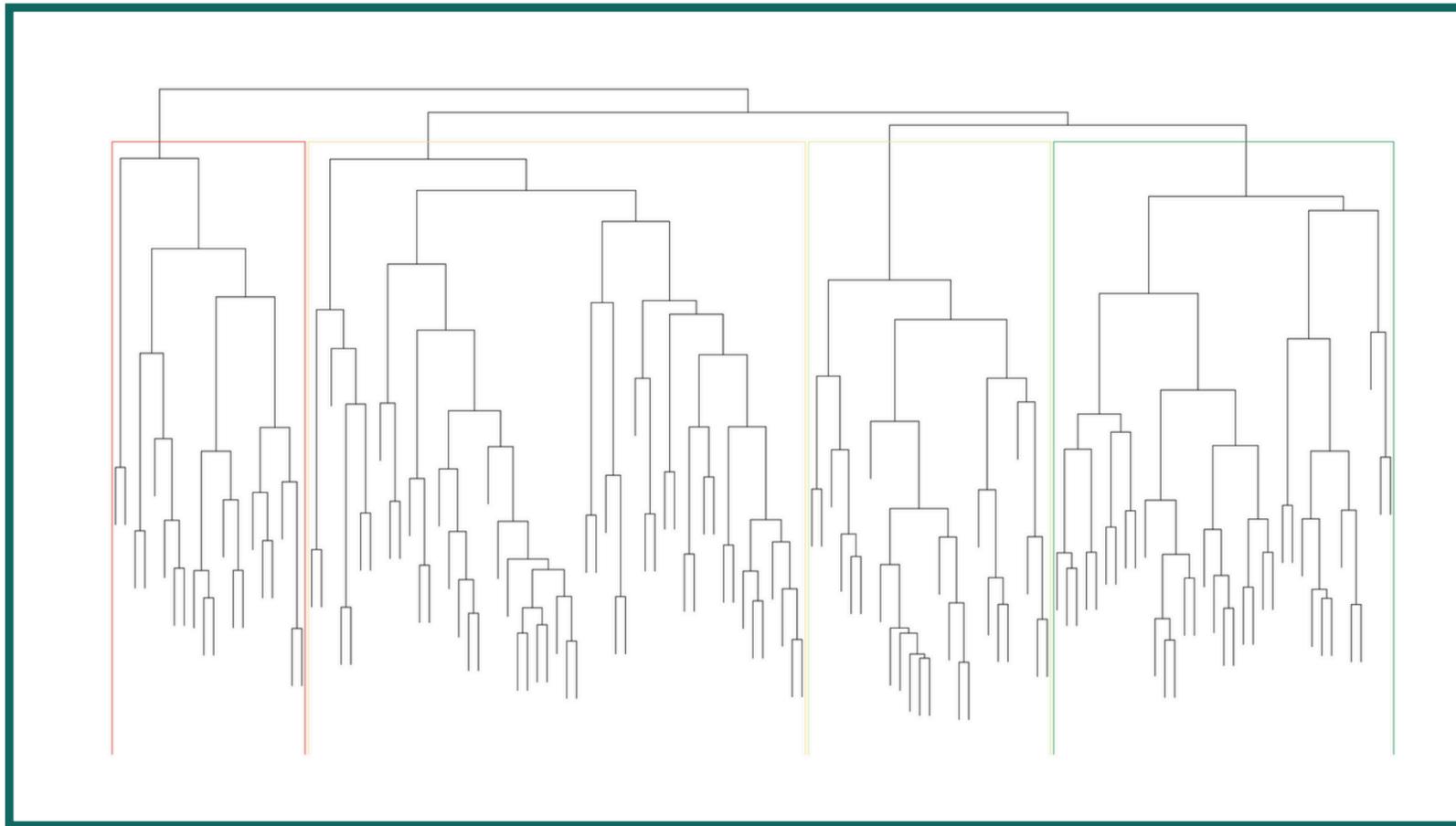
Strategie:

- Maggiore consapevolezza del prodotto.

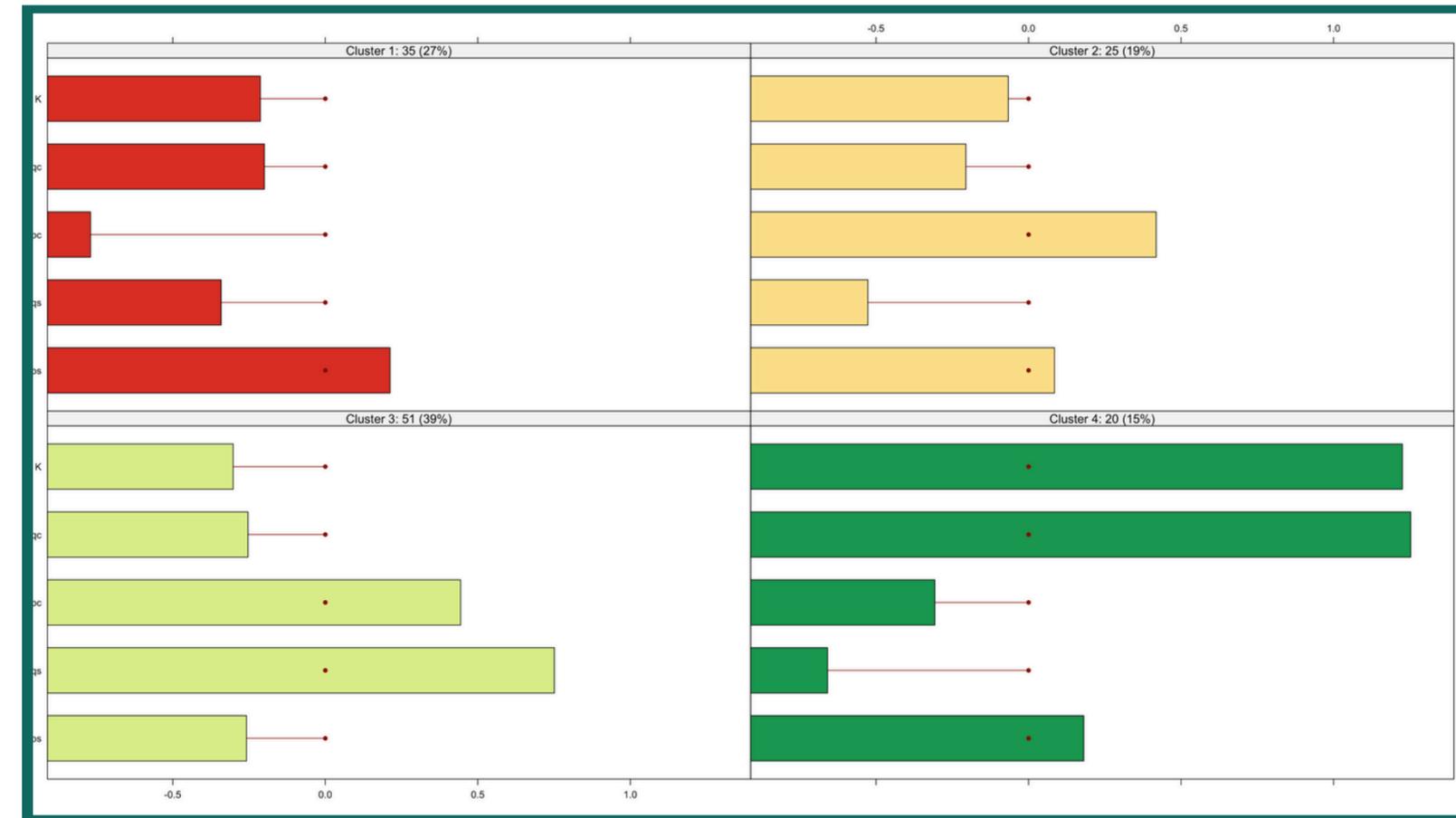
m	p	q
28355.3	0.0009	0.0029

Cluster con il modello GGM

$$z(t) = K \sqrt{\frac{1 - e^{-(p_c + q_c)t}}{1 + \frac{q_c}{p_c} e^{-(p_c + q_c)t}} \frac{1 - e^{-(p_s + q_s)t}}{1 + \frac{q_s}{p_s} e^{-(p_s + q_s)t}}$$



Dendrogramma per 4 cluster usando il legame completo



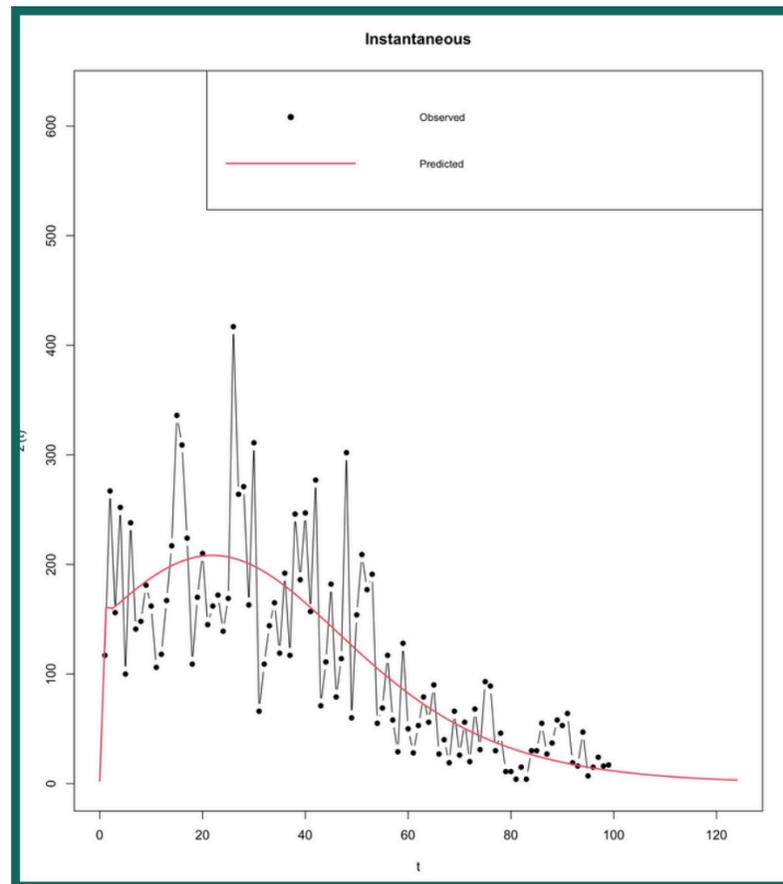
i Grafico a barre dei 5 (K, qc, pc, qs, qp) parametri del modello GGM scalati

Cluster 1 (n=35)

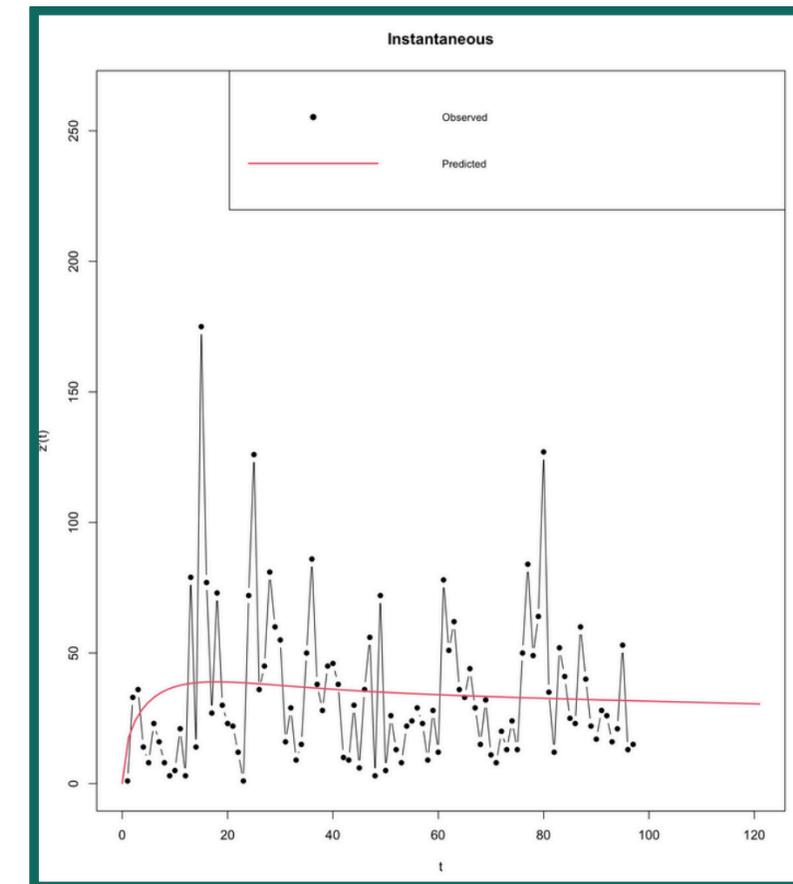
K	qc	pc	qs	ps
81671	0.0016	-1.8578	0.0057	0.0228

Cluster 2 (n=25)

K	qc	pc	qs	ps
110656	0.0016	-0.0769	0.0029	-0.0227



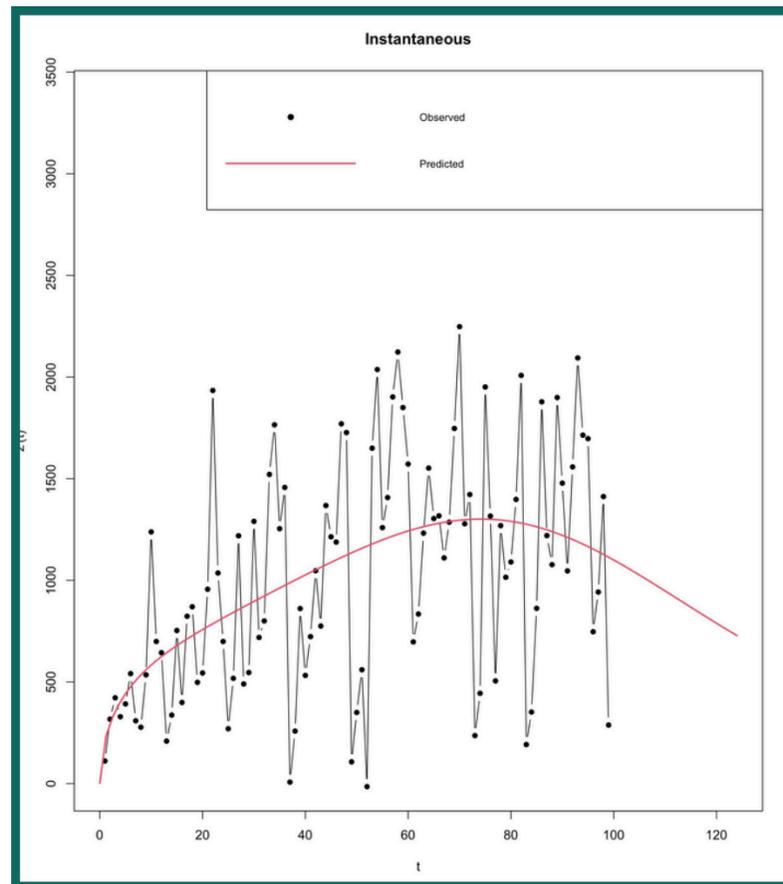
ROTARY TRILLERS W



BRAVA LARGE

Cluster 3 (n=51)

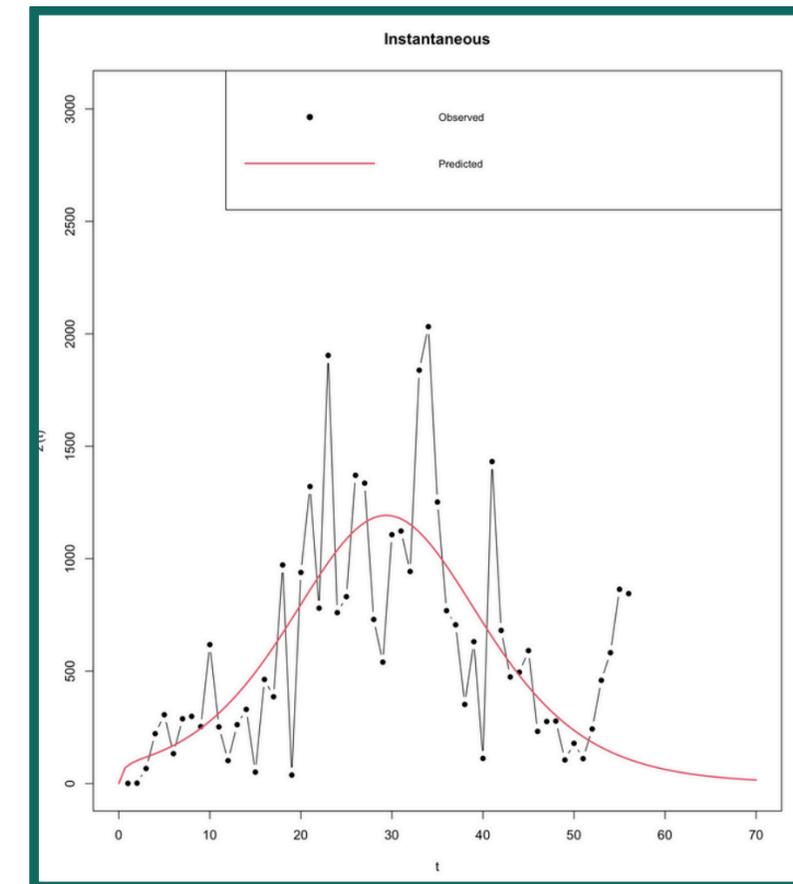
K	qc	pc	qs	ps
64061	0.0013	-0.0389	0.0218	-0.1449



VIRAT

Cluster 4 (n=20)

K	qc	pc	qs	ps
36657 1	0.0136	-1.1645	0.0010	0.0147



RT30

Prodotti maturi



Prodotti nella fase iniziale della crescita



Prodotti in fase di crescita

Prodotti "dormienti"

Domande di ricerca

03. Qual è l'andamento delle vendite in Italia?

Modelli di serie storiche

- SARIMA (p,d,q)x(P,D,Q)
- ETS

Integrazione con dati Istat

- Motivazioni
- Unificazione del dataset

Modelli gerarchici

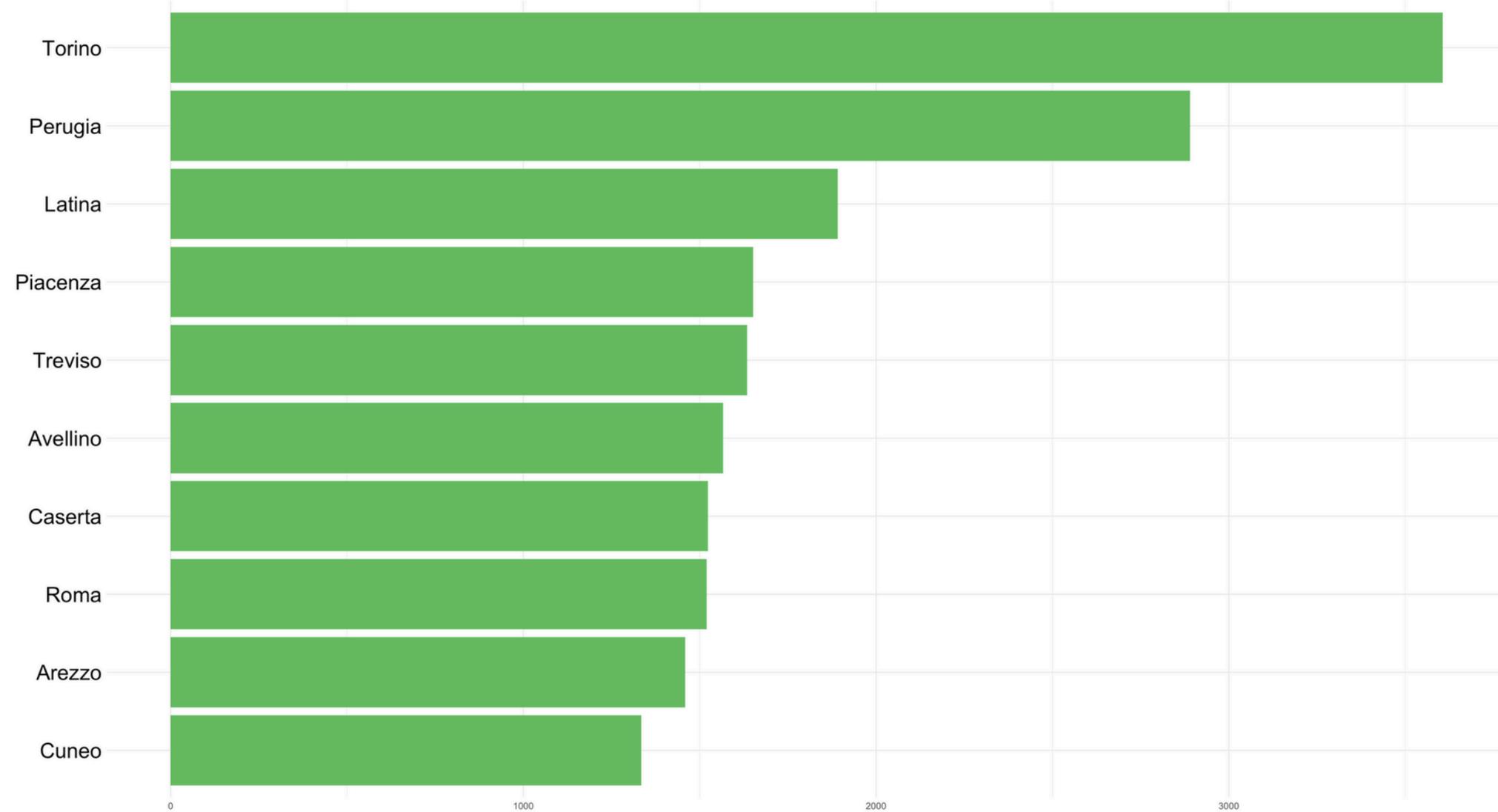
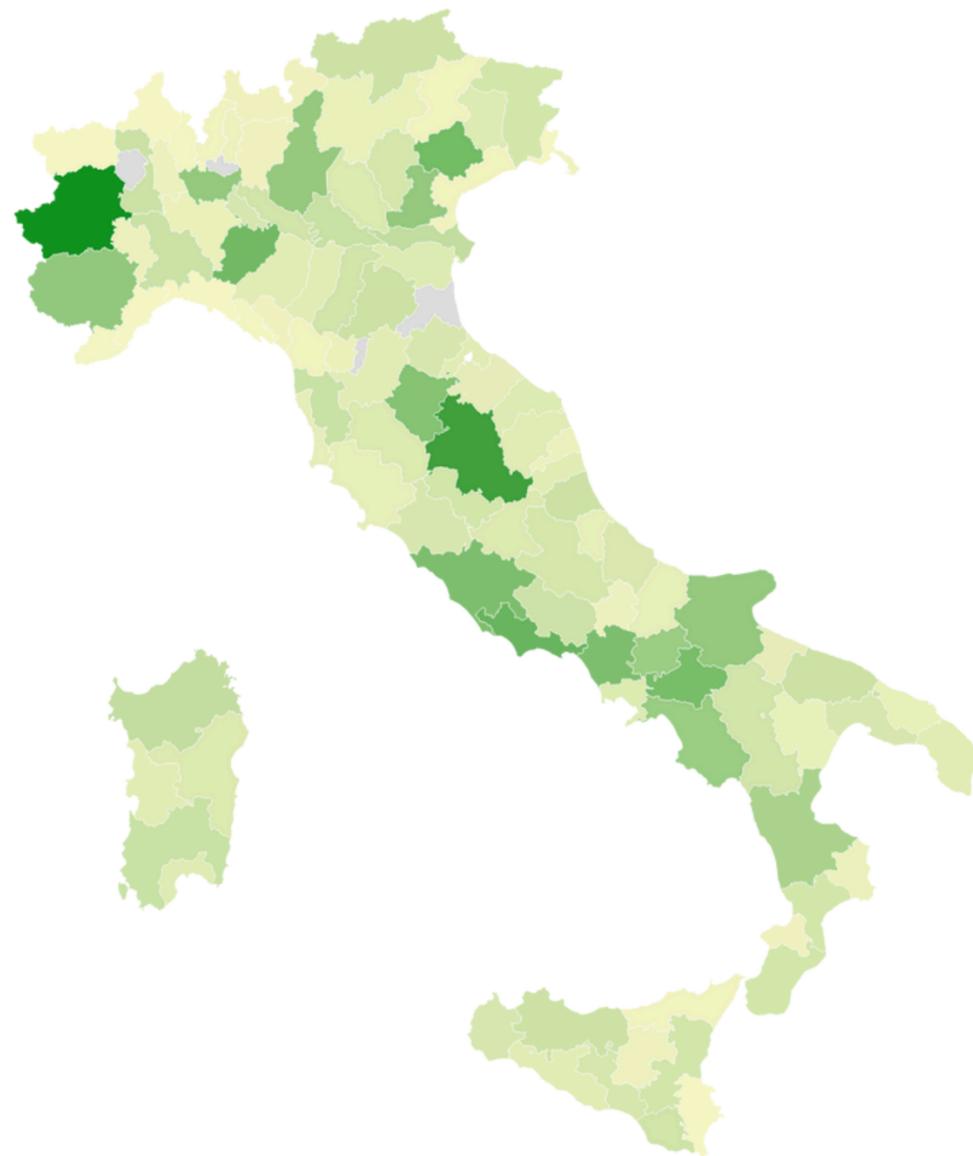
- Pooled, unpooled o gerarchico?

Modelli di Data mining

- Motivazioni
- Convalida Incrociata
- Confronto tra XGBoost e Random Forests

Analisi esplorativa

Numero articoli venduti



Solo quattro province sono escluse dalle vendite

Previsione

- Previsioni di vendita a livello regionale
- Eliminazione Liguria e Valle D'Aosta
- Dataset diviso in **stima e verifica**.
 1. Stima: 01-2016 / 12-2022
 2. Verifica: 01-2023/ 03-2024
- Previsione sui 6 periodi successivi.

SARIMA(p,d,q)x(P,D,Q)

- Sono potenti per modellare una vasta gamma di serie storiche, compresi i dati con stagionalità.
- I modelli ARIMA sono più complessi e richiedono una comprensione più profonda dei dati e dei loro pattern.
- Richiedono che la serie sia stazionaria.

ETS

- Lisciamento esponenziale triplo: **Holt-Winters**
- Medie ponderate di osservazioni passate, in cui i **pesi decrescono esponenzialmente** quanto più le osservazioni diventano vecchie.
- Cambiamenti recenti nelle serie storiche
- Non richiedono che la serie sia stazionaria

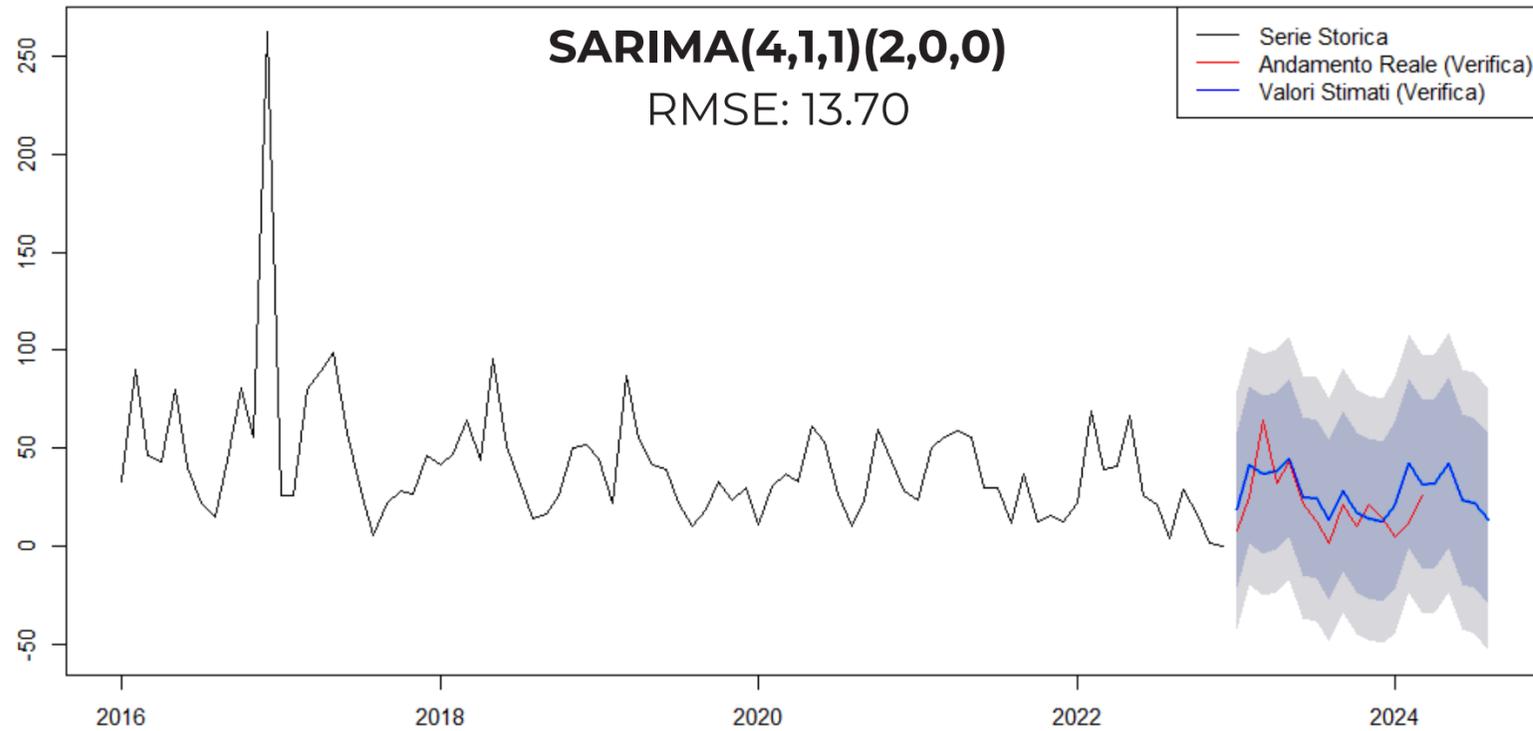
Previsione delle vendite

SARIMA (p, d, q)x(P, D, Q)

Toscana

SARIMA(4,1,1)(2,0,0)

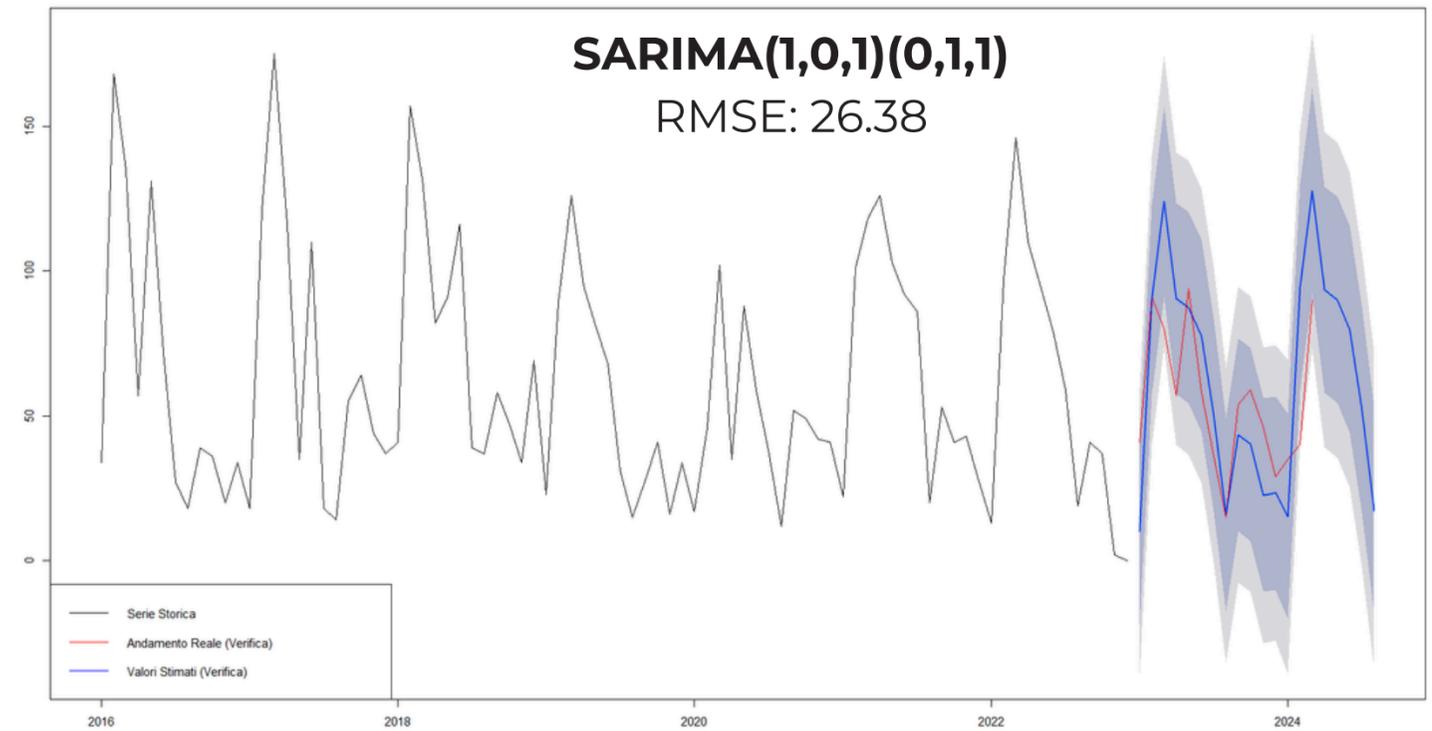
RMSE: 13.70



Campania

SARIMA(1,0,1)(0,1,1)

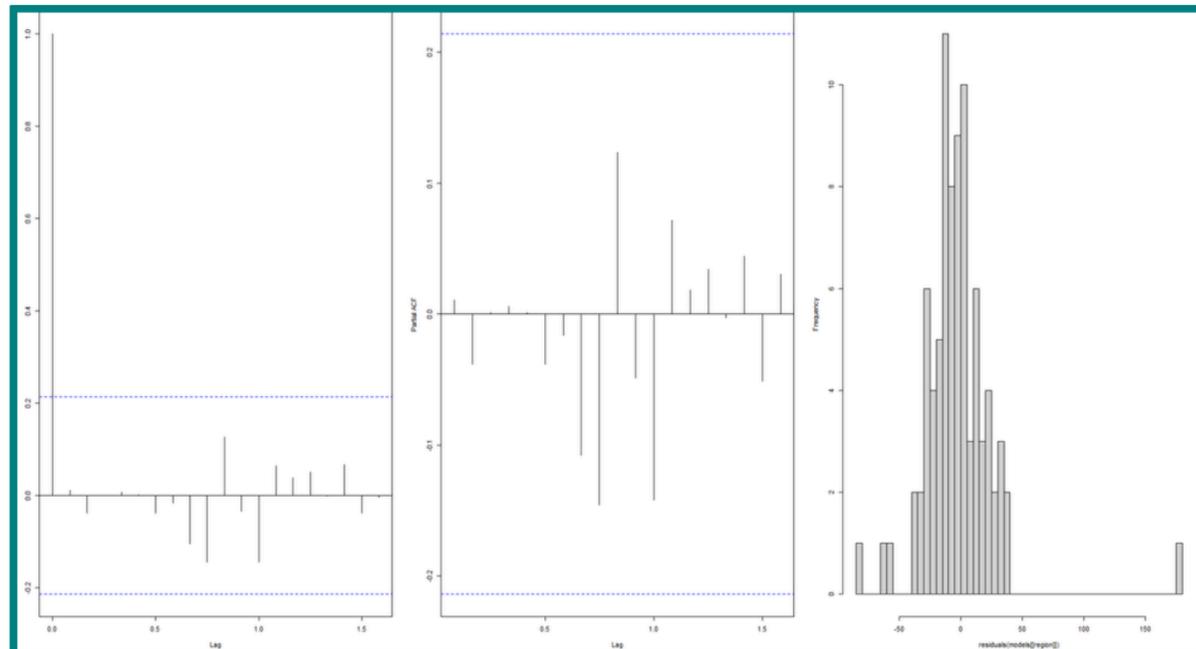
RMSE: 26.38



ACF

PACF

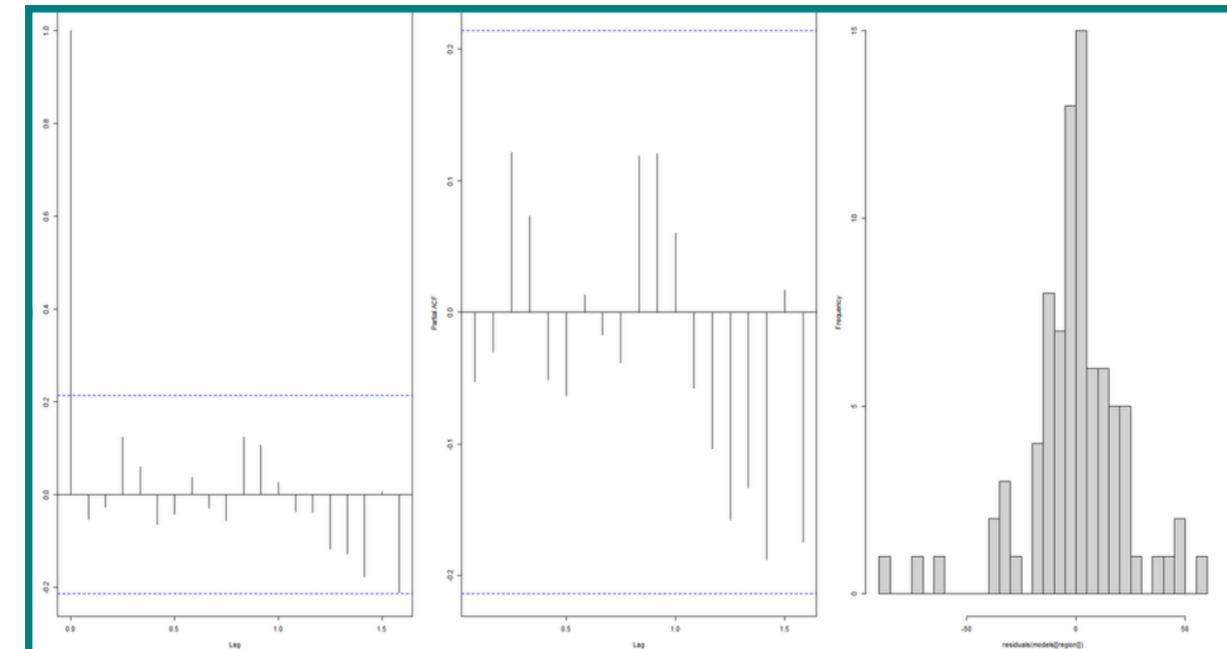
Residui



ACF

PACF

Residui

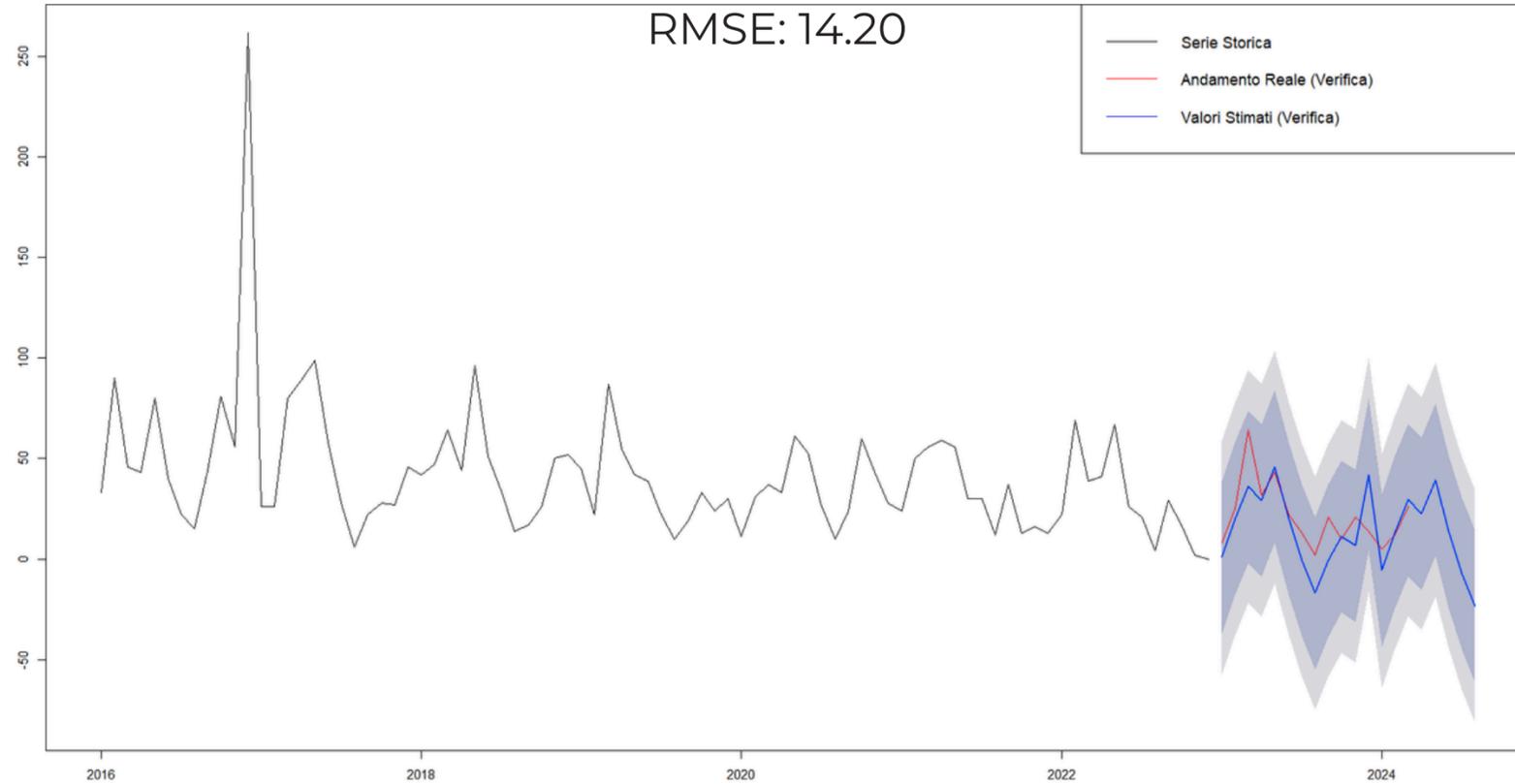


Previsione delle vendite

Lisciamento Esponenziale (EST)

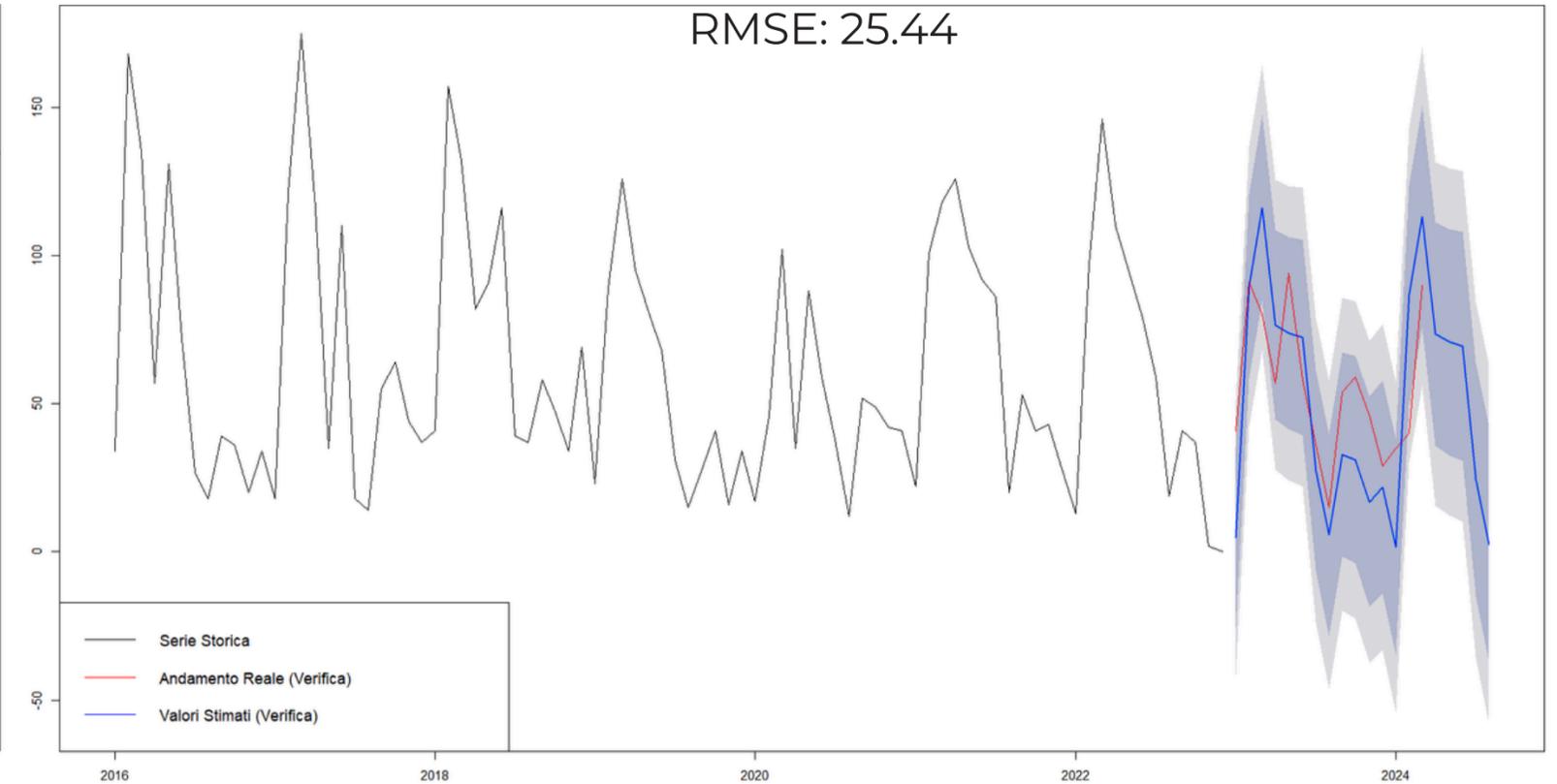
Toscana

RMSE: 14.20



Campania

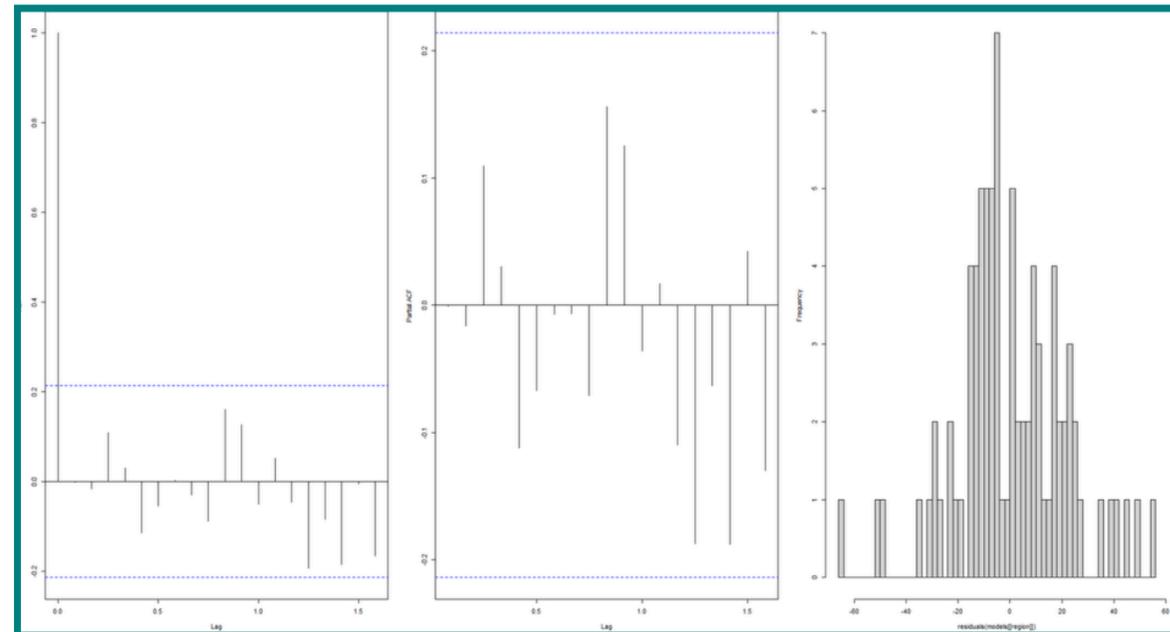
RMSE: 25.44



ACF

PACF

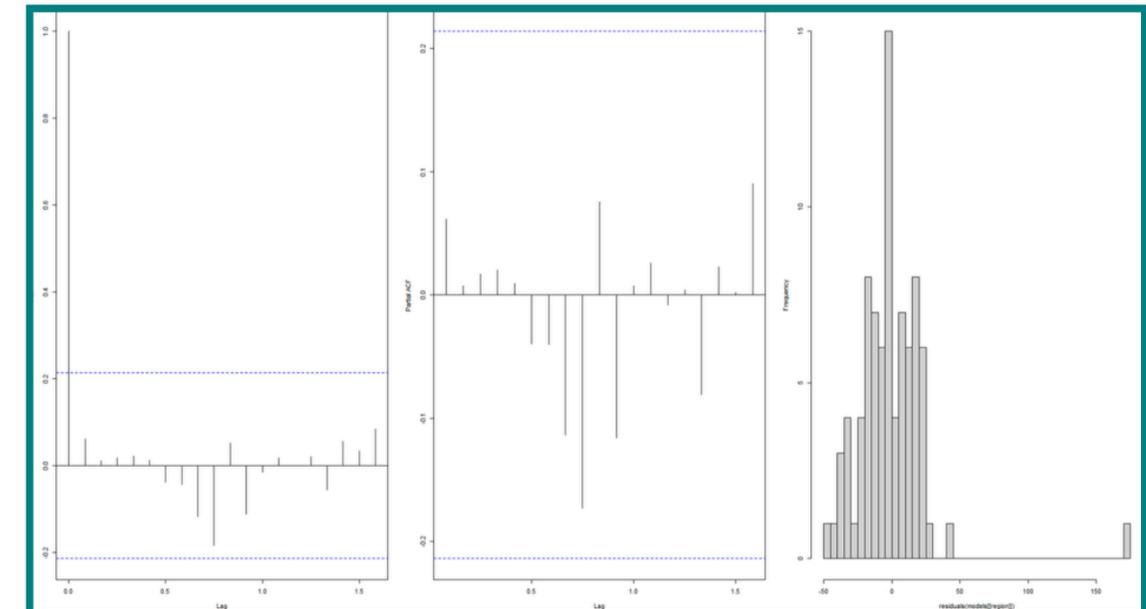
Residui



ACF

PACF

Residui



Previsione

Confronto performance

- I modelli SARIMA hanno tutti messo in evidenza una forte **componente stagionale**.
- Modelli di Holt-Winters sono modelli più semplici.
- Modelli non confrontabili con il criterio dell'AIC.
- Modelli confrontati con **RMSE** sul dataset di verifica.

RMSE	
SARIMA	ETS
20.54901	22.17591

Integrazione dati Istat (<http://dati.istat.it/>)

- Domanda: Ci sono altri fattori che possono influenzare l'andamento delle vendite?

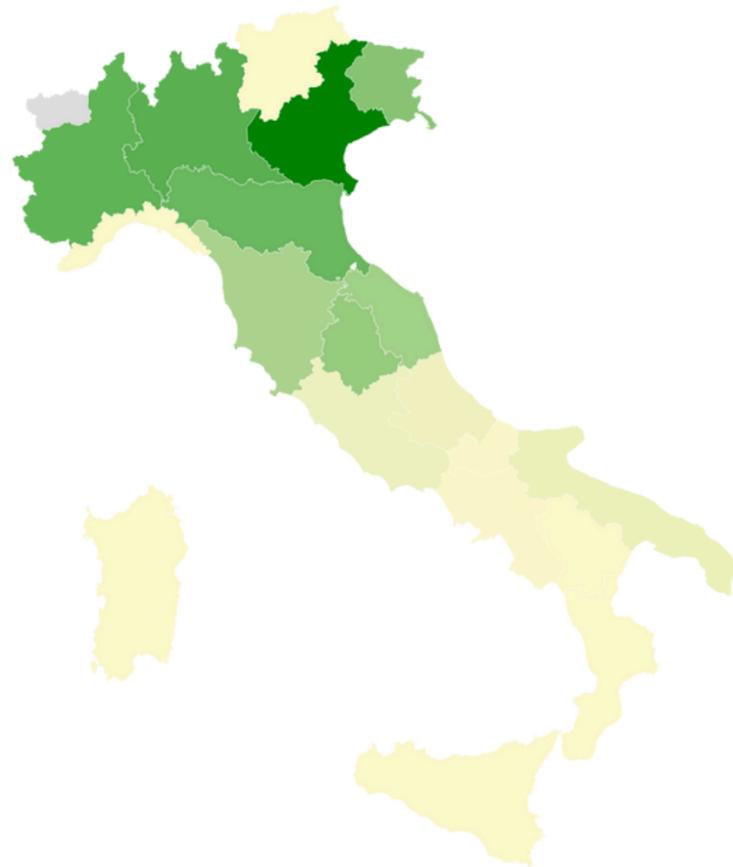
└───▶ **Produzione agricola** di ciascun territorio

└───▶ **Conformazione** del territorio (montagna, collina, pianura)

- Problematiche:
 - a. Dati disaggregati
 - b. Incoerenze tra definizioni delle province nel tempo e tra dataset
 - c. Mancanza di serie storiche mensili.

Produzione agricola italiana

Produzione di semi (anno 2023)



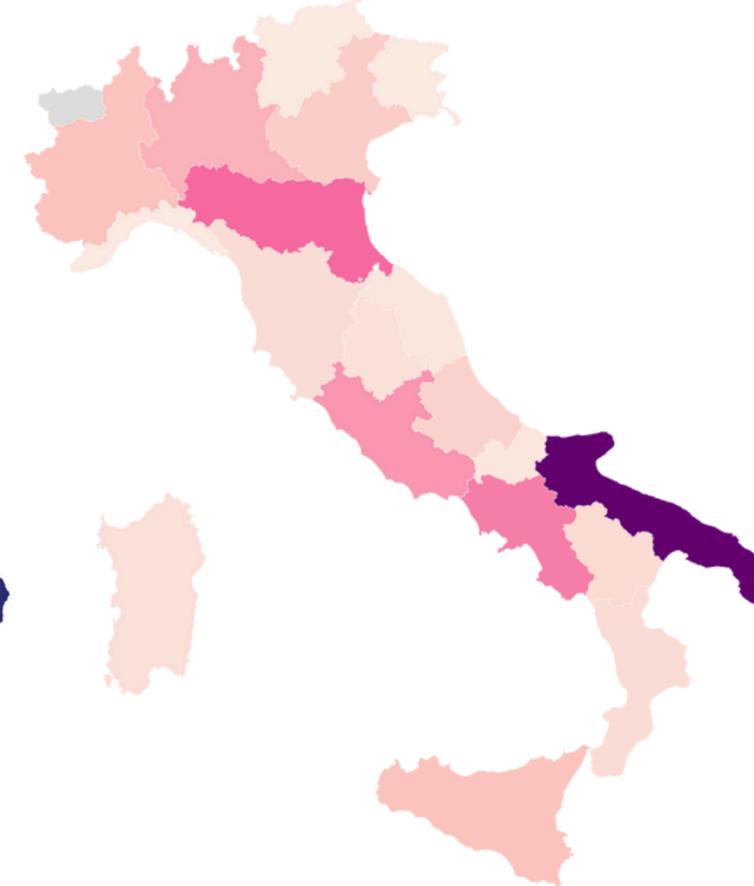
Created with Datawrapper

Produzione di frumento (anno 2023)



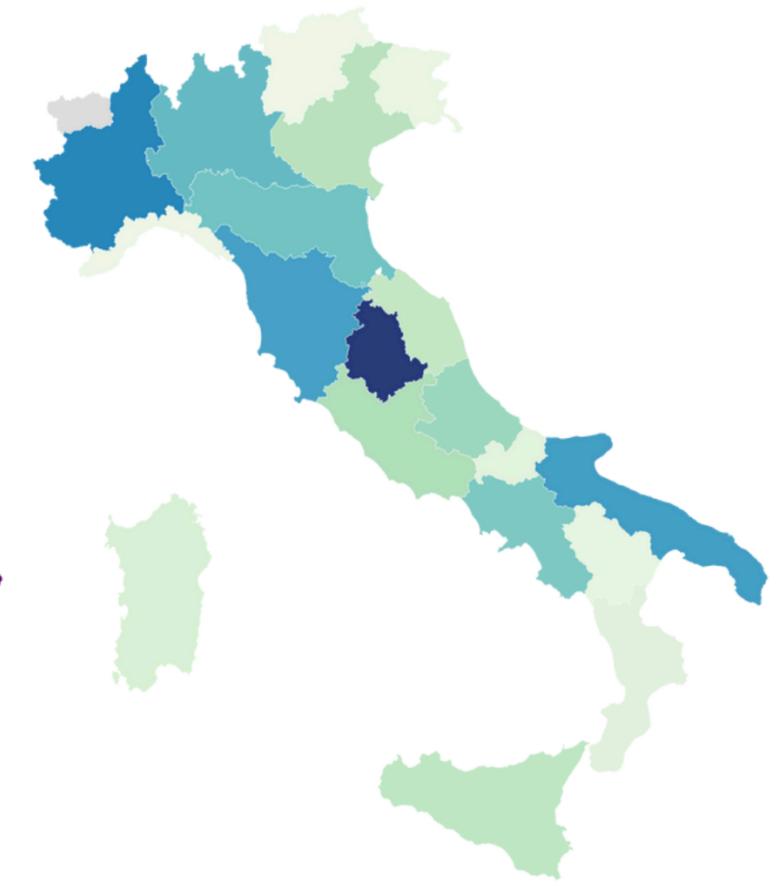
Created with Datawrapper

Produzione di ortaggi (anno 2023)



Created with Datawrapper

Produzione di legumi (anno 2023)



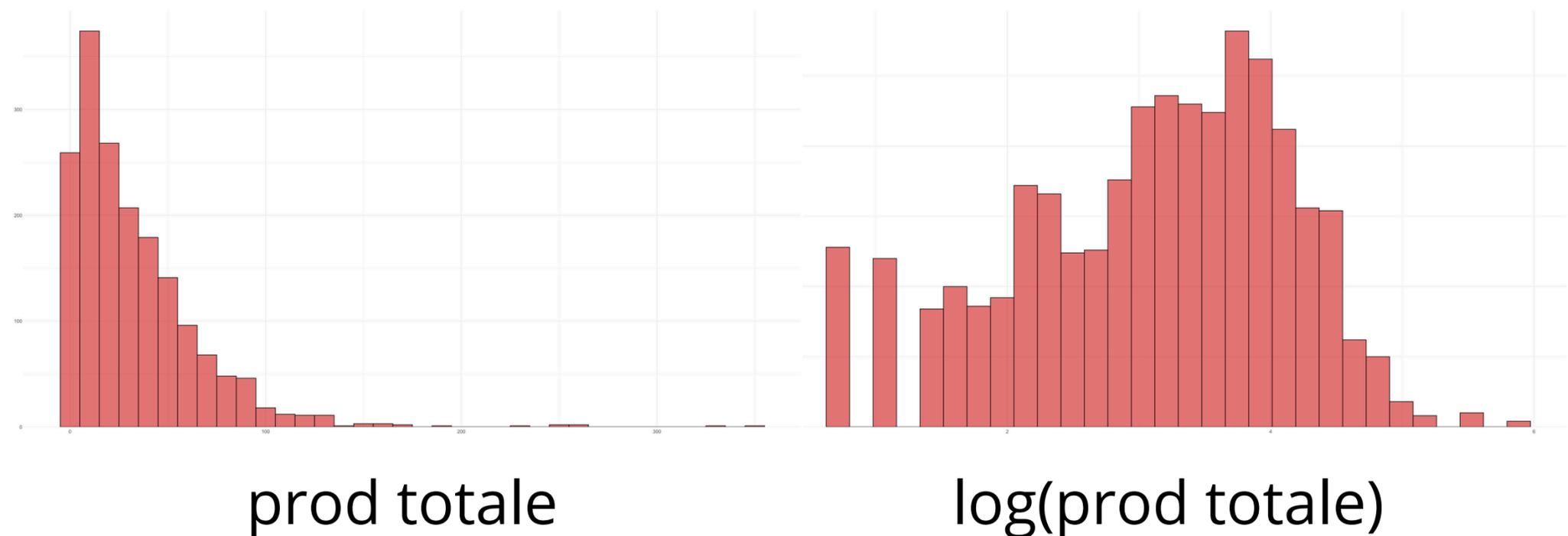
Created with Datawrapper

- **Produzione** in quintali di: Semi, Frumento, Ortaggi, Legumi dal 2016 al 2023
- **Superficie agricola utilizzata** (in ettari) per **zona altimetrica**.

Dataset

Regione	Semi	Legumi	Ortaggi	Frumento	Montagna	Pianura	Collina	Mese	Ann o	Prod totale
Abruzzo	5024737	10387122	364022196	129775855	234884.28	0	188086.38	1	2016	31
xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx

- 11 variabili
- 1762 osservazioni
- variabile risposta: produzione totale
- Sostituzione valori mancanti con 0
- Eliminazione resi



Modelli gerarchici

Modello Unpooled

Modello Pooled

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.849e+00  4.769e-02  59.745 < 2e-16 ***
Semi         3.865e-01  4.416e-02   8.752 < 2e-16 ***
legumi       4.566e-01  2.887e-02  15.819 < 2e-16 ***
Ortaggi      3.695e-01  6.120e-02   6.038 1.92e-09 ***
Frumento    -2.932e-01  6.425e-02  -4.563 5.40e-06 ***
Montagna    2.362e-01  2.767e-02   8.535 < 2e-16 ***
collina      2.518e-01  2.751e-02   9.155 < 2e-16 ***
Pianura      7.419e-07  1.894e-07   3.916 9.36e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.948 on 1681 degrees of freedom
(70 osservazioni eliminate a causa di valori mancanti)
Multiple R-squared:  0.343,    Adjusted R-squared:  0.3403
F-statistic: 125.4 on 7 and 1681 DF,  p-value: < 2.2e-16
```

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
RegioneAbruzzo  4.05960  1.06056  3.828 0.000134 ***
RegioneBasilicata  2.60180  0.34855  7.465 1.34e-13 ***
RegioneCalabria  3.38798  0.16662 20.334 < 2e-16 ***
RegioneCampania  4.25084  0.70260  6.050 1.78e-09 ***
RegioneEmilia   2.79793  0.50807  5.507 4.22e-08 ***
RegioneFriuli   5.74643  0.60002  9.577 < 2e-16 ***
RegioneLazio    3.41054  0.53892  6.329 3.18e-10 ***
RegioneLiguria  4.45107  0.56264  7.911 4.62e-15 ***
RegioneLombardia  5.62999  0.76206  7.388 2.35e-13 ***
RegioneMarche   2.40691  0.65894  3.653 0.000268 ***
RegioneMolise   4.20191  0.34670 12.120 < 2e-16 ***
RegionePiemonte  4.03551  0.53783  7.503 1.01e-13 ***
RegionePuglia   -1.49765  1.79081  -0.836 0.403108
RegioneSardegna  0.37053  0.83250  0.445 0.656323
RegioneSicilia  -3.06864  1.07058  -2.866 0.004205 **
RegioneToscana  1.45543  1.23299  1.180 0.238007
RegioneTrentino  5.42880  2.19937  2.468 0.013674 *
RegioneUmbria   4.25185  0.39525 10.757 < 2e-16 ***
RegioneValle    4.15960  0.74985  5.547 3.37e-08 ***
RegioneVeneto   4.88835  0.58002  8.428 < 2e-16 ***
Semi            0.30347  0.12863  2.359 0.018423 *
legumi          0.08228  0.04420  1.861 0.062866 .
Ortaggi         0.53123  0.23236  2.286 0.022367 *
Frumento        0.39510  0.11503  3.435 0.000608 ***
Montagna        0.29147  0.74687  0.390 0.696402
collina         2.39047  0.67698  3.531 0.000425 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7988 on 1663 degrees of freedom
(70 osservazioni eliminate a causa di valori mancanti)
Multiple R-squared:  0.9398,    Adjusted R-squared:  0.9389
F-statistic: 998.7 on 26 and 1663 DF,  p-value: < 2.2e-16
```

Modello gerarchico a intercetta casuale

REML criterion at convergence: 4174.6

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.9798	-0.5233	0.1193	0.6719	3.1621

Random effects:

Groups	Name	Variance	Std.Dev.
Regione	(Intercept)	1.0496	1.0245
	Residual	0.6427	0.8017

Number of obs: 1689, groups: Regione, 20

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)	
(Intercept)	3.099e+00	3.172e-01	9.744e+00	9.769	2.4e-06	***
Semi	4.315e-01	1.191e-01	4.459e+02	3.623	0.000325	***
legumi	1.046e-01	4.378e-02	1.620e+03	2.388	0.017050	*
Ortaggi	4.872e-01	1.986e-01	1.323e+02	2.453	0.015478	*
Fumento	3.780e-01	1.139e-01	1.582e+03	3.319	0.000924	***
Montagna	6.540e-01	2.316e-01	1.062e+01	2.824	0.017081	*
collina	7.100e-01	2.295e-01	1.203e+01	3.094	0.009269	**
Pianura	-5.319e-07	1.066e-06	1.136e+01	-0.499	0.627259	

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

	(Intr)	Semi	legumi	Ortagg	Frumnt	Montgn	Collin
Semi	0.155						
legumi	0.013	0.094					
Ortaggi	0.235	0.097	-0.062				
Fumento	0.024	-0.335	-0.144	-0.406			
Montagna	-0.049	-0.020	0.000	0.069	0.008		
collina	0.047	0.106	-0.022	-0.196	-0.078	-0.403	
Pianura	-0.680	-0.201	-0.009	-0.326	-0.036	0.127	-0.006

fit warnings:

Some predictor variables are on very different scales: consider rescaling

Anova tra gerarchico e Pooled

refitting model(s) with ML (instead of REML)

Data: finale3

Models:

lm_pool: logp ~ . - Regione - Data - YEAR - prod_tot

lm1: logp ~ . - prod_tot - Regione - Data - YEAR + (1 | Regione)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
lm_pool	9	4622.6	4671.5	-2302.3	4604.6			
lm1	10	4153.1	4207.4	-2066.5	4133.1	471.53	1	< 2.2e-16 ***

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova tra gerarchico e Unpooled

refitting model(s) with ML (instead of REML)

Data: finale3

Models:

lm1: logp ~ . - prod_tot - Regione - Data - YEAR + (1 | Regione)

lm_no_pool: logp ~ . - Data - YEAR - prod_tot - 1 - Pianura

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
lm1	10	4153.1	4207.4	-2066.5	4133.1			
lm_no_pool	27	4062.3	4208.9	-2004.1	4008.3	124.83	17	< 2.2e-16 ***

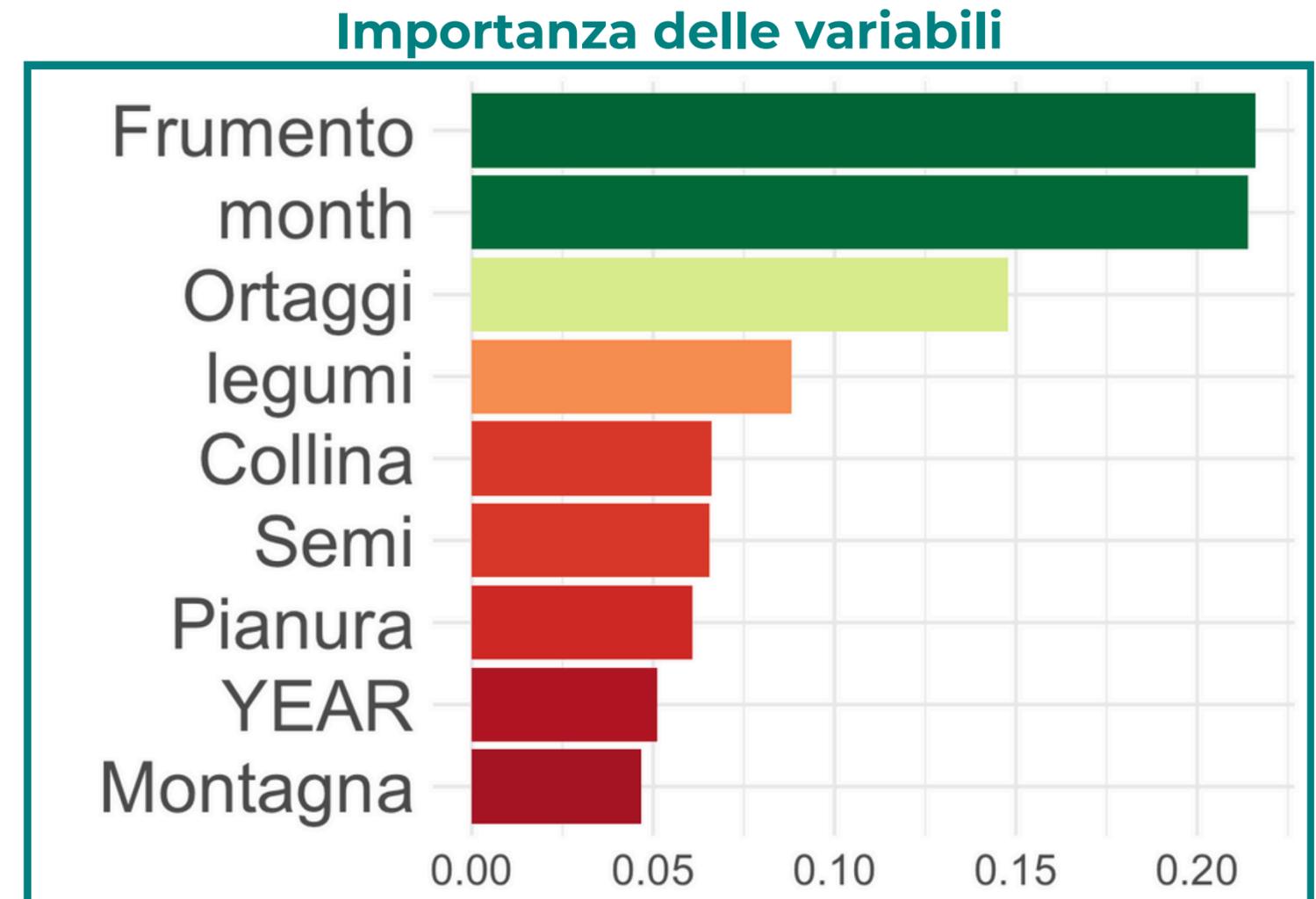
signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Modelli di Data Mining

Modelli realizzati tramite convalida incrociata con 5 fold

RMSE	
random forest	22.3382
albero	25.72648
bagging	22.662
xgboost	21.28941



Limiti

```
graph LR; A[Limiti] --> B[Bassa numerosità dei dati e numero ridotto di variabili quantitative]; A --> C[Bassa qualità dei dati ottenuti da Istat];
```

Bassa numerosità dei dati e numero ridotto di variabili quantitative

Bassa qualità dei dati ottenuti da Istat

Sviluppi futuri

Unire i vantaggi forniti dai modelli di Data Mining e di serie storiche, ad esempio **LSTM**



**GRAZIE PER
L'ATTENZIONE!**