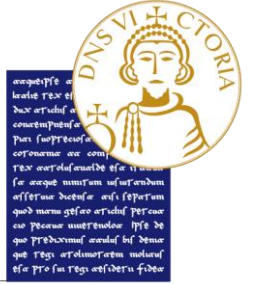


UNIVERSITÀ DEGLI STUDI DEL SANNIO

Dipartimento di Ingegneria



*Corso di Laurea Magistrale in
Ingegneria Informatica*

Valutazione di tecniche basate su grafi per l'analisi di fake news e metriche di propagazione

RELATORI:

Prof. Antonio Pecchia

Prof. Francesco Vasca

CANDIDATO:

Francesco Pio Briuolo

Matr. 399000544

Anno accademico 2022/2023

Indice

- Introduzione alle fake news
- Grafi e tecniche basate sulla propagazione
- Caso studio I: grafo bipartito
- Caso studio II: rete di propagazione gerarchica
- Conclusioni e sviluppi futuri

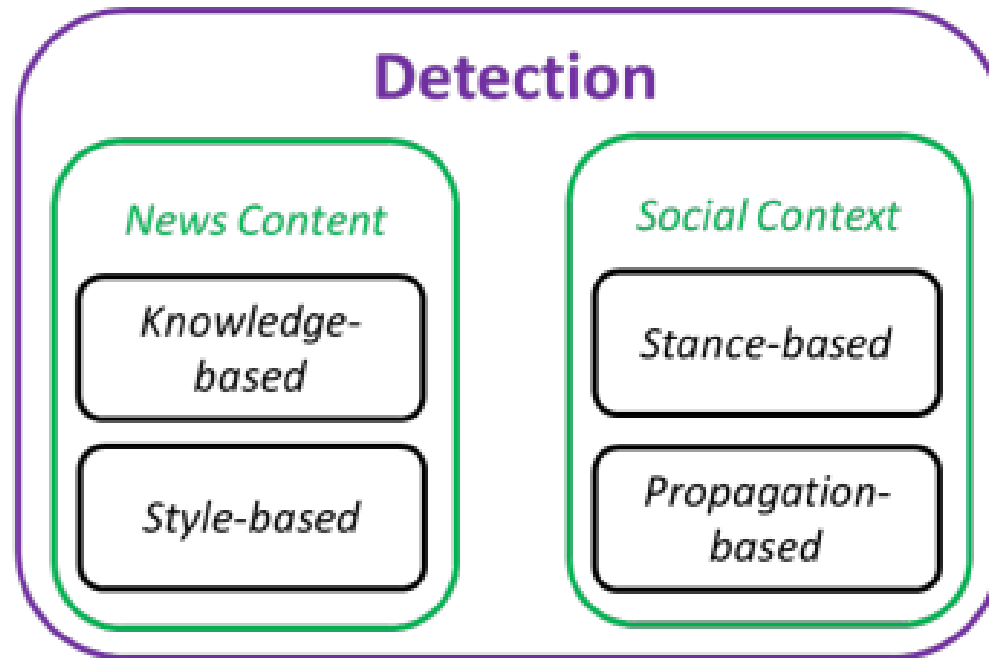
Contesto: Fake news

«Notizia intenzionalmente e verificabilmente falsa, in grado di fuorviare i lettori»

H. Allcott et al. Journal of Economic Perspectives. 2017



Problema: Rilevamento e caratterizzazione delle fake news



Tra i metodi basati sulla propagazione:

- **Analisi dei grafi di propagazione**
- Analisi delle «News Cascades»

Contributo della tesi

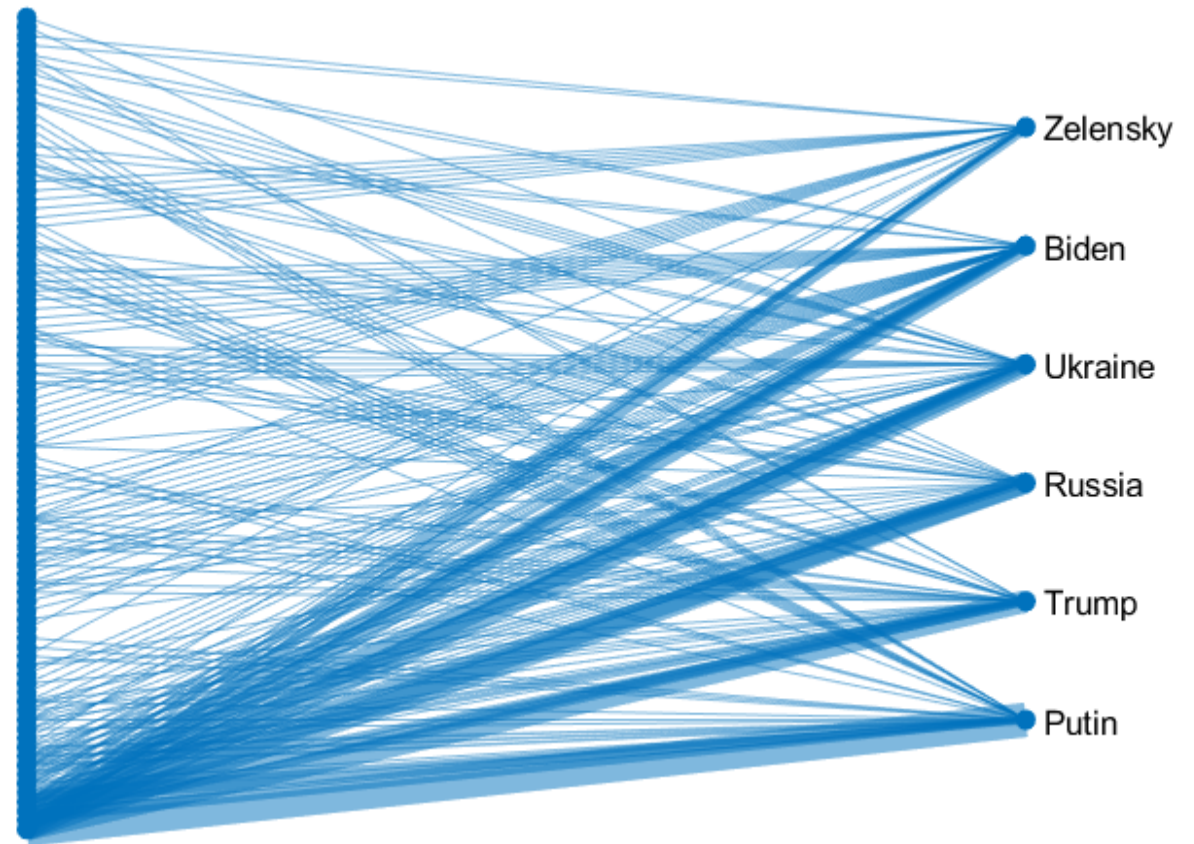
Presentazione di due approcci basati sui grafi per capire in che modo possono contribuire alla caratterizzazione e alla detection delle fake news:

- **Grafo bipartito**
- **Rete di propagazione gerarchica**

Caso Studio I: Grafo Bipartito

Grafo che comprende una rete di «speakers» e una di «keywords»

Il peso degli archi indica il grado di fakeness delle notizie



Caso Studio I: Dataset

Dati ottenuti tramite tecniche di
Web Scraping

Fonte: politifact.com

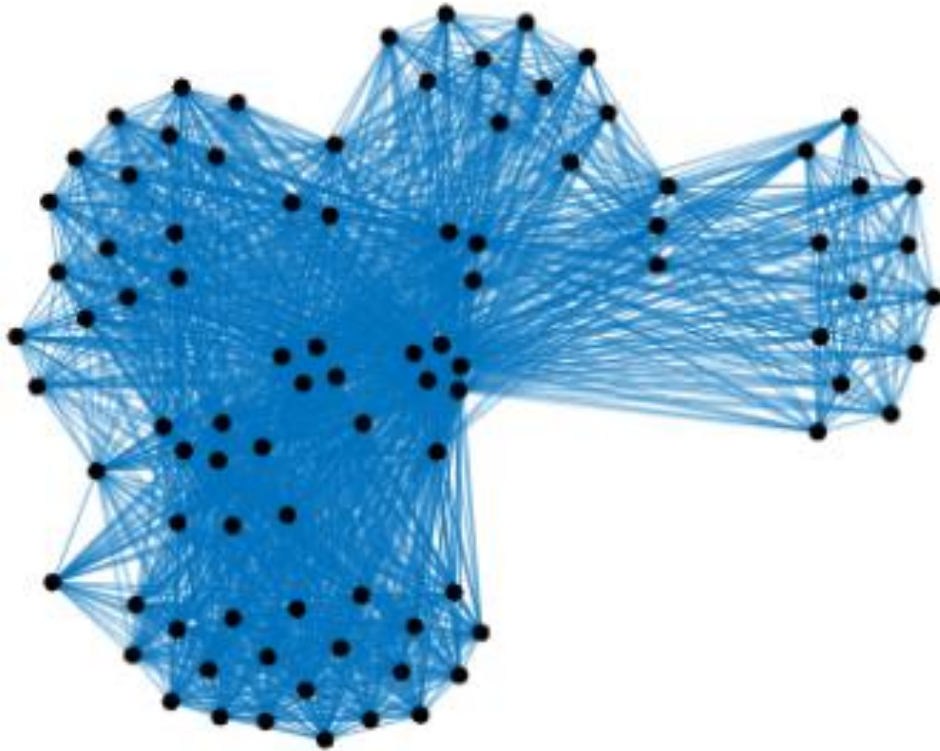


BeautifulSoup

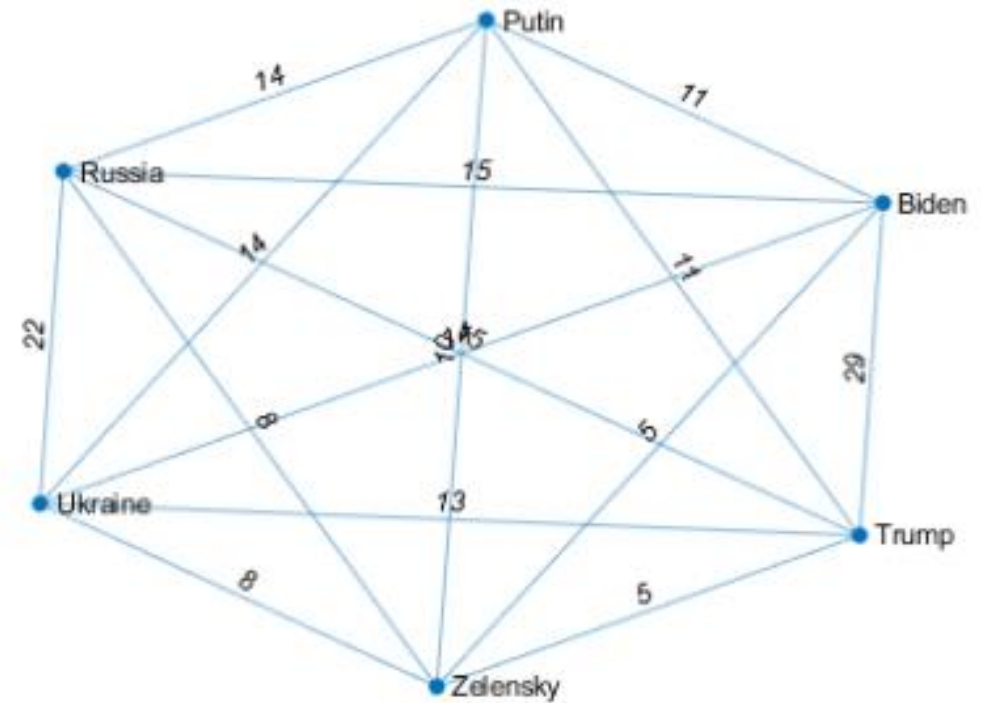
Statistica	PolitiFact
#Notizie	652
#Speakers	90
#Keywords	6

Caso Studio I: Metodi e tecniche

Grafo delle proiezioni nello spazio degli speakers



Grafo delle proiezioni nello spazio delle keywords



Caso Studio I: Risultati

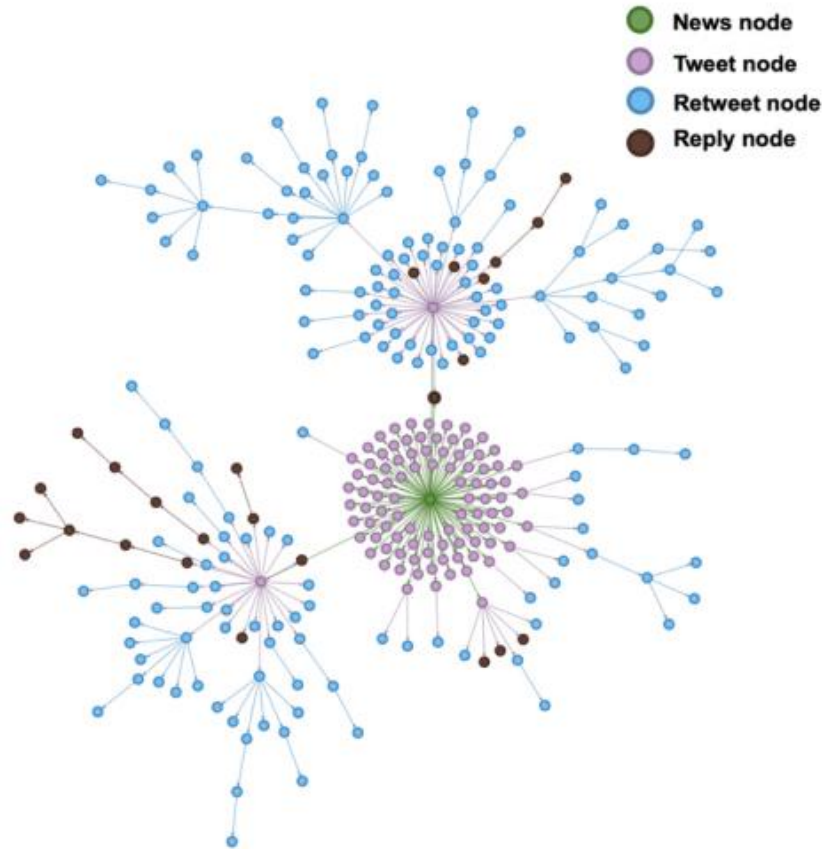
Matrice delle distanze tra coppie di keywords

	Biden	Putin	Russia	Trump	Ukraine	Zelensky
Biden		0,00088	0,001005	0,001063	0,001071	0,002322
Putin	0,00088		0,000681	0,000747	0,000708	0,00146
Russia	0,001005	0,000681		0,000899	0,000756	0,001952
Trump	0,001063	0,000747	0,000899		0,000939	0,001908
Ukraine	0,001071	0,000708	0,000756	0,000939		0,001992
Zelensky	0,002322	0,00146	0,001952	0,001908	0,001992	

Caso Studio II: Dataset FakeNewsNet

Statistica	PolitiFact	GossipCop
#Notizie reali	277	6945
#Fake news	351	3684
#Utenti	384,813	739,166
#Tweets	275,058	1,058,330
#Retweets	293,438	530,833
#Risposte	125,654	232,923

Caso Studio II: Rete di propagazione gerarchica



Struttura della rete:

- Macro-livello
- Micro-livello



Caso Studio II: Caratteristiche di rete

Caratteristiche strutturali: modellano i pattern di diffusione delle notizie.

Caratteristiche temporali: modellano la frequenza di interazioni dell'utente.

Caratteristiche linguistiche: valutate in termini di sentiment a partire dal contenuto delle notizie originali.

Caso Studio II: Metodologia e Implementazione

Algoritmi di Machine Learning per determinare l'importanza delle features:

- Decision Tree
- Random Forest
- Logistic Regression
- Gaussian Naive Bayes
- Support Vector Machine
- Feedforward Neural Network



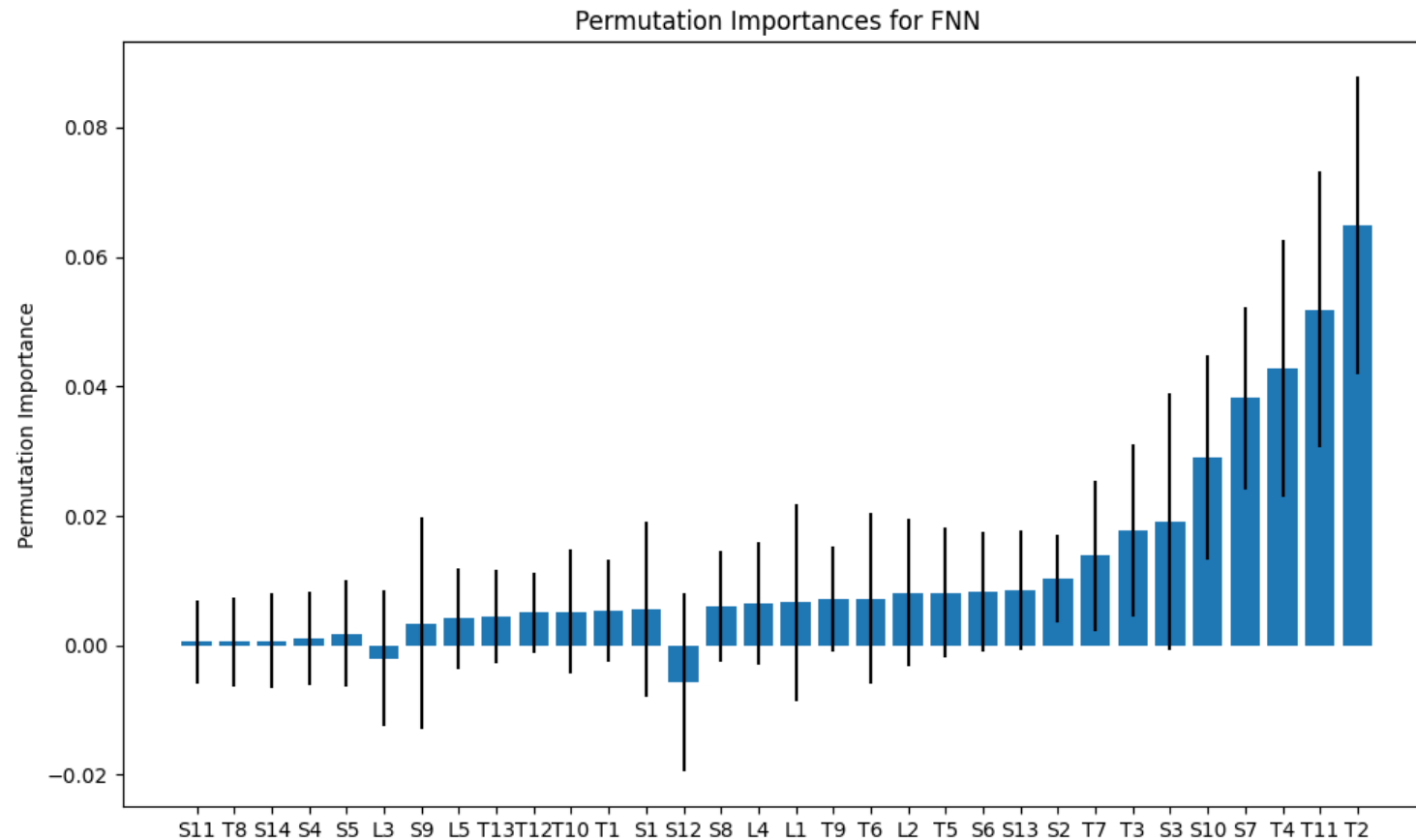
Caso Studio II: Risultati (PolitiFact)

Classifier	Accuracy	Precision	Recall	F1-score
RF	0.818 ± 0.011	0.811 ± 0.021	0.825 ± 0.009	0.818 ± 0.008
FNN	0.796 ± 0.011	0.740 ± 0.021	0.909 ± 0.047	0.816 ± 0.011
SVM	0.802 ± 0.000	0.754 ± 0.000	0.891 ± 0.000	0.817 ± 0.000
DT	0.750 ± 0.035	0.737 ± 0.043	0.775 ± 0.044	0.754 ± 0.032
LR	0.748 ± 0.000	0.690 ± 0.000	0.891 ± 0.000	0.778 ± 0.000
GNB	0.766 ± 0.000	0.699 ± 0.000	0.927 ± 0.000	0.797 ± 0.000

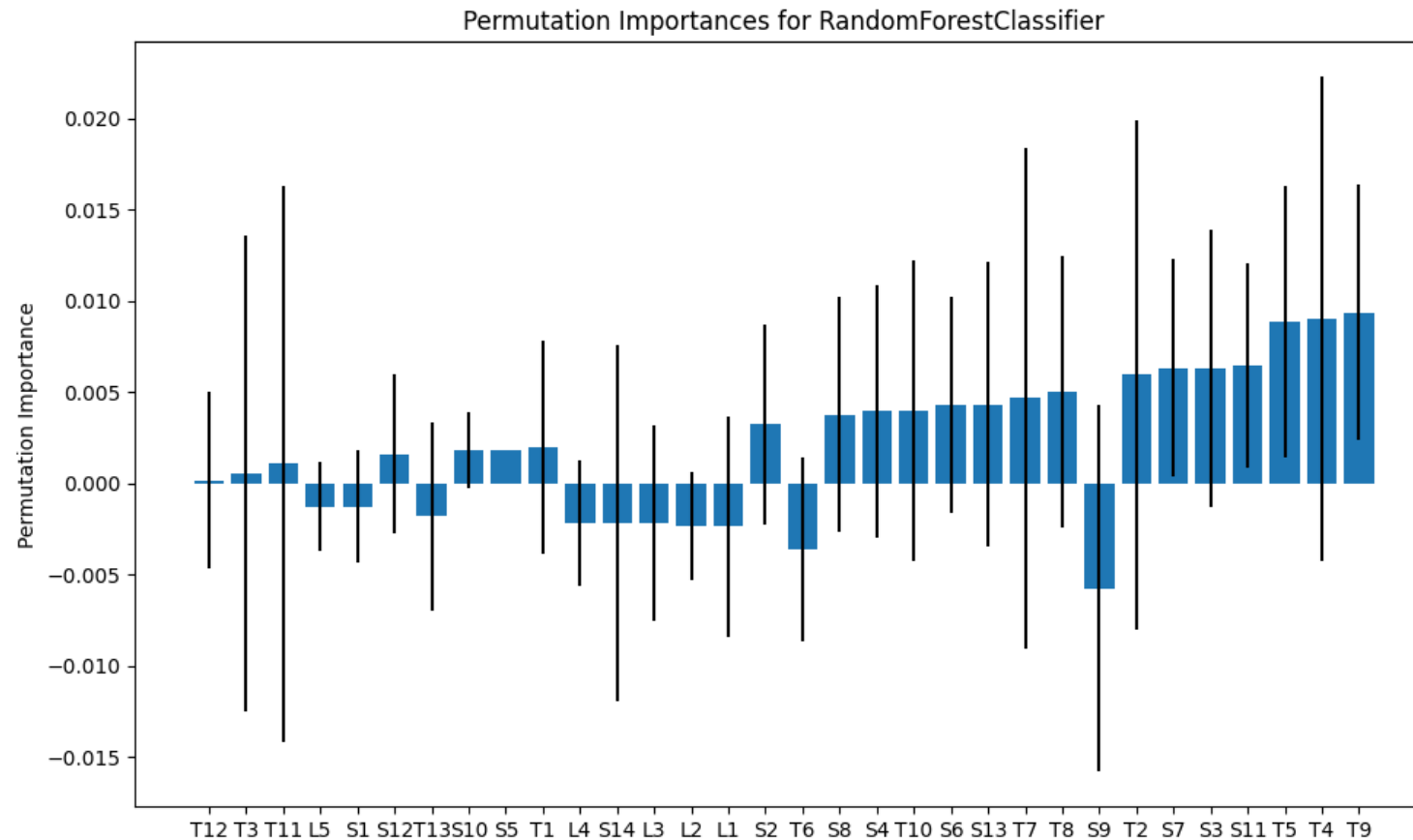
Caso Studio II: Risultati (GossipCop)

Classifier	Accuracy	Precision	Recall	F1-score
RF	0.879 \pm 0.003	0.866 \pm 0.003	0.898 \pm 0.004	0.881 \pm 0.003
FNN	0.874 \pm 0.003	0.861 \pm 0.006	0.891 \pm 0.010	0.875 \pm 0.004
SVM	0.854 \pm 0.000	0.877 \pm 0.000	0.824 \pm 0.000	0.850 \pm 0.000
DT	0.840 \pm 0.005	0.843 \pm 0.002	0.837 \pm 0.012	0.840 \pm 0.006
LR	0.843 \pm 0.000	0.882 \pm 0.000	0.791 \pm 0.000	0.834 \pm 0.000
GNB	0.655 \pm 0.000	0.858 \pm 0.000	0.370 \pm 0.000	0.518 \pm 0.000

Caso Studio II: Feature Importances (1)



Caso Studio II: Feature Importances (2)



Conclusioni

- Importanza di avere a disposizione un dataset costantemente aggiornato.
- Individuare caratteristiche di rete utili per l'analisi delle fake news.
- Tecniche basate sul contenuto e tecniche basate sul contesto sono tra loro complementari.

Sviluppi futuri

- Analisi di rete sui grafi delle proiezioni.
- Approfondire l'analisi del grafo k-partito.
- Utilizzare le features più importanti su dati accessibili.