

# Rete bipartita: Analisi dei voti elettorali

## Introduzione

Il nostro report analizza la struttura di una rete bipartita costruita a partire dai dati di un'elezione anonima. Lo scopo è identificare i legami significativi tra i candidati in base alle co-occorrenze nei voti degli elettori. I nostri dati di partenza comprendono una tabella anonima di voti che è stata successivamente pulita e processata.

Le nostre analisi includono l'applicazione di due tecniche di correzione statistica (Bonferroni e FDR) per determinare la significatività dei legami tra i candidati.

## Metodologia

### Pulizia e preparazione del dataset

I dati sono stati preprocessati in, sostanzialmente, 3 passaggi:

- Sostituzione dei valori non validi ("[NA]") con valori mancanti effettivi
- Rimozione delle righe interamente vuote
- Creazione di liste di voti validi per ciascun elettore

### Creazione della rete bipartita

La rete bipartita è stata costruita considerando:

- Nodi degli elettori: un nodo per ciascun votante
- Nodi dei candidati: un nodo per ciascun candidato votato
- Link: un link tra un elettore e un candidato per ogni voto espresso

### Analisi delle co-occorrenze

Le co-occorrenze tra candidati sono state calcolate analizzando le combinazioni di voti espressi da ciascun elettore. Considerando quindi il conteggio dei voti totali che sarebbe il numero di voti ricevuti da ciascun candidato abbiamo le co-occorrenze che sarebbero il numero di volte in cui due candidati sono stati votati insieme.

## Correzioni

Per determinare la significatività delle co-occorrenze sono stati utilizzati due approcci:

- **Correzione di Bonferroni:** è stata applicata per controllare il livello di significatività

$\alpha$

rispetto al numero totale di test. Essa è molto conservativa, richiede p-valore molto piccoli per dichiarare un risultato significativo rischiando anche di scartare risultati verosimili.

$$p - value_{(k)} < \frac{\alpha}{T}$$

- **False Discovery Rate (FDR):** applicando questa correzione si tende ad accettare un numero maggiore di risultati significativi rispetto alla correzione di Bonferroni, riducendo però la specificità. La FDR controlla la proporzione di falsi positivi tra tutti i risultati dichiarati significativi.

$$p - value_{(k)} < \frac{\alpha K}{T}$$

dove:

- $\alpha$  = livello di significatività scelto
- $K$  = posizione del p-value nell'ordine crescente
- $T$  = numero totale di test

## Distribuzione

Per calcolare il p-valore delle co-occorrenze utilizziamo la **distribuzione ipergeometrica**. Essa permette di calcolare la probabilità che X co-occorrenze avvengano dato N, NA e NB. I parametri chiave della distribuzione sono:

- N: numero totale di elettori
- NA e NB: numero totale di voti ricevuti rispettivamente dai due candidati (CA) e (CB)

$$p = \sum_{i=X}^{\min(NA, NB)} \frac{\binom{NA}{i} \binom{N-NA}{NB-i}}{\binom{N}{NB}}$$

La somma calcola la probabilità cumulativa di osservare almeno X co-occorrenze, data l'ipotesi nulla.

*Come interpretiamo questo p-valore?*

Esso rappresenta la probabilità che CA e CB abbiano X o più co-occorrenze semplicemente per caso. Se esso è molto piccolo si può rifiutare l'ipotesi nulla di indipendenza.

## Analisi e risultati

Rete Bipartita: Elettori e Candidati

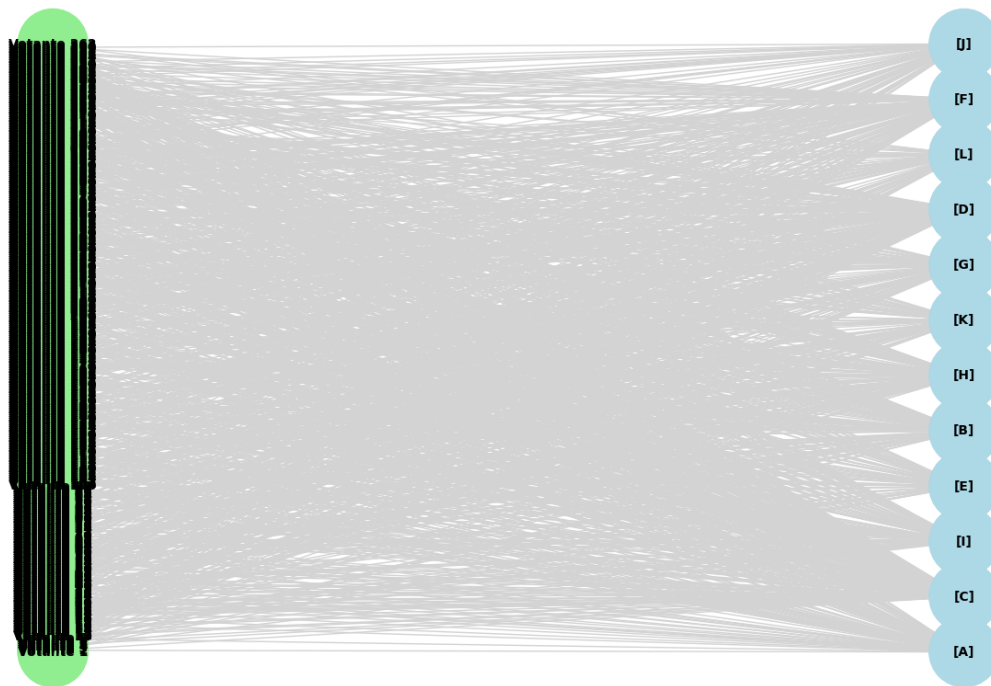


Figura 1: Rete bibartita elettori e candidati

Questo grafico mostra i nodi rappresentanti elettori e candidati disposti in due insiemi separati con i link che rappresentano i voti.

### Legami significativi

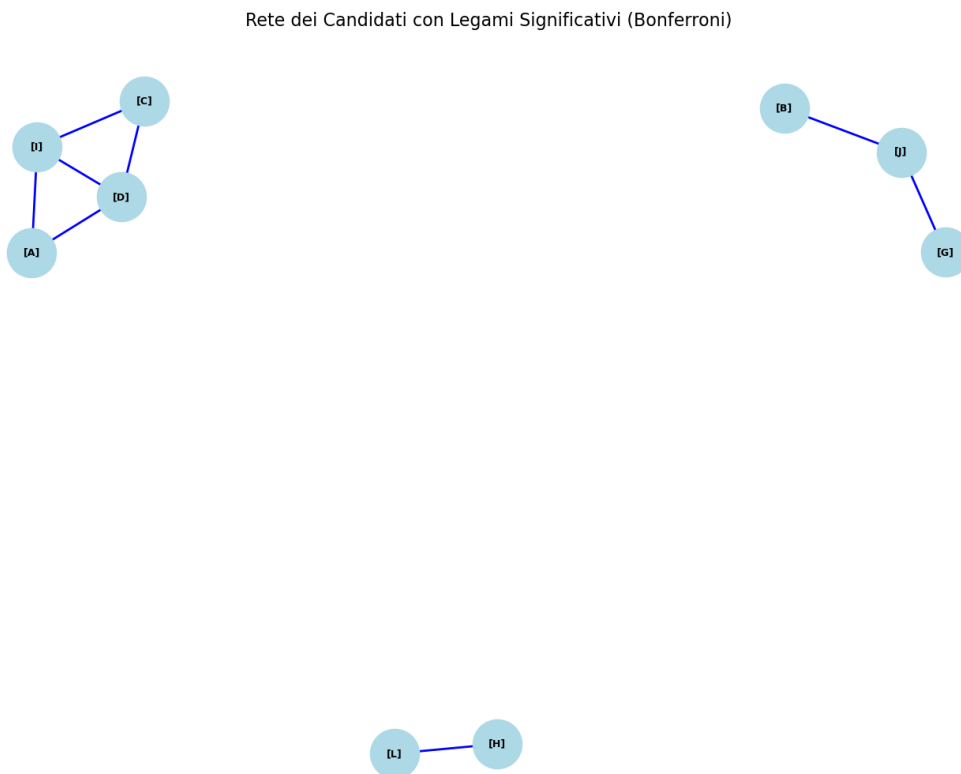
I legami significativi identificati con la **correzione di Bonferroni** sono:

- [A] e [D] votati insieme 66 volte (p-valore:  $4,97 \times 10^{-12}$ )
- [B] e [J] votati insieme 35 volte (p-valore:  $2,59 \times 10^{-7}$ )
- [G] e [J] votati insieme 27 volte (p-valore:  $8,5 \times 10^{-5}$ )
- [C] e [D] votati insieme 83 volte (p-valore:  $8,76 \times 10^{-7}$ )

- [D] e [I] votati insieme 61 volte (p-valore:  $3,94 \times 10^{-8}$ )
- [H] e [L] votati insieme 35 volte (p-valore:  $6,52 \times 10^{-8}$ )
- [C] e [I] votati insieme 60 volte (p-valore: 0,0001)
- [A] e [I] votati insieme 44 volte (p-valore:  $1,11 \times 10^{-5}$ )

I legami significativi identificati con la **correzione FDR** includono quelli indicati con la correzione di Bonferroni con l'aggiunta di:

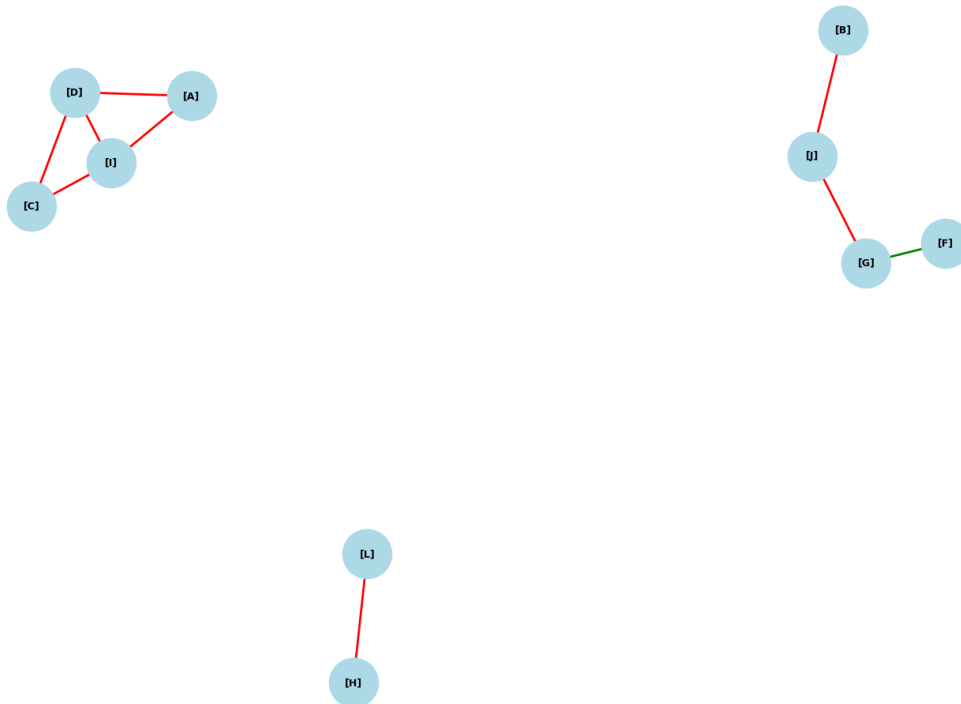
- [F] e [G] votati insieme 37 volte (p-valore: 0,0008)



**Figura 2: Comunità rilevate con correzione di Bonferroni**

Questo grafico evidenzia solo i link più stringenti tra i candidati offrendo una visualizzazione chiara delle relazioni più forti.

Rete dei Candidati con Legami Significativi (FDR)



**Figura 3: Comunità rilevate con correzione FDR**

In questo grafico si include anche un legame con significatività meno stringente (quello evidenziato in verde piuttosto che in rosso) rispetto a Bonferroni, mostrando una rete più ampia. Questa rete permette di osservare relazioni potenzialmente rilevanti che potrebbero invece sfuggire con criteri più restrittivi.

## Conclusioni

Notiamo come ci sono delle alleanze implicite tra candidati. Esse potrebbero essere spiegate da diversi fattori tra cui:

- Affinità politiche o ideologiche: i candidati potrebbero appartenere a gruppi con prospettive o programmi simili che, di conseguenza, attirano voti dagli stessi elettori
- Strategie elettorali di coordinamento: alcune coppie di candidati potrebbero aver collaborato in modo informale per ottenere supporto reciproco

Si potrebbe anche pensare che i candidati più noti o con maggiore visibilità potrebbero essere associati più frequentemente dagli elettori. Inoltre, le relazioni personali o di comunità tra candidati ed elettori potrebbero influire sulla scelta dei voti. Le co-occorrenze possono anche riflettere il tentativo di alcuni elettori di sostenere combinazioni di candidati ritenuti complementari in termini di competenze e rappresentanza. Altra ipotesi che si potrebbe fare è che, potrebbero esserci dei candidati più influenti in determinate comunità o in determinate aree geografiche.

***L'effetto trascinamento*** è emerso. Esso si manifesta in particolare nelle coppie con una disparità di notorietà dove un candidato meno conosciuto sembra beneficiare della presenza di uno più affermato. Ad esempio, il candidato C che ha ottenuto 159 voti ha probabilmente trascinato con sé il candidato I, il quale è stato votato ben 60 volte insieme a C, su un totale di 100 voti ricevuti. Anche il candidato L sembra aver subito l'effetto trascinamento da parte di H, avendo in comune con quest'ultimo 35 voti su 60 ricevuti. Inoltre, anche il candidato J potrebbe aver beneficiato dell'effetto trascinamento, avendo 35 voti in comune con il candidato B, su un totale di 64 voti ricevuti. Abbiamo citato i 3 effetti più vistosi di trascinamento presenti all'interno delle comunità rilevate, ognuno di questi avviene in una comunità diversa, quindi tutte le comunità sembrano coinvolte. Infine, vorremmo evidenziare che in ognuno dei casi sopracitati, il numero di voti in comune con il compagno di comunità, del candidato che sembra beneficiare dell'effetto trascinamento, è superiore al 50% dei voti ricevuti dal candidato.