



UNIVERSITÀ DEGLI STUDI DI PADOVA  
CORSO DI PSICOLOGIA COGNITIVA APPLICATA

# **Il gambling attraverso la pubblicità: un'applicazione di statistica testuale**

NICCOLÒ POLO 19/07/2019

# OBIETTIVI DELLA RICERCA

Studio esplorativo dei testi pubblicitari per poter analizzare l'approccio dei distributori di giochi sul consumatore

Tramite Youtube sono state raccolte le trascrizioni di 160 testi di spot televisivi sul gioco d'azzardo trasmessi dal 2012 al 2018, suddivisi in tre macrocategorie:

- giochi d'azzardo in stile casinò (60 spot)
- scommesse sportive (45 spot)
- lotto, tombola, gratta e vinci (55 spot)

Proporre una metodologia statistica per alcune indagini psicologiche

# TEXT MINING

## COS'È?

Il **Data Mining** è una procedura che permette di esplorare grosse banche dati (Big Data) ed estrapolare informazioni e variabili tra esse

Il **Text Mining** è una tecnica particolare di Data Mining che permette di trattare dati di tipo testuale

Il **Natural Language Processing** è un processo adottato dal Text Mining per consente alla macchina di elaborare il linguaggio naturale

# METODO

Per poter operare su dati di tipo testuale è necessario procedere in una serie di step di trasformazioni del dato:

1. Lettura documenti
2. Trasformazione in corpora
3. Tokenizzazione
4. Creazione matrici DFM





## Esempio di subset del corpus di un documento

[1] "888 casinò, sempre col fiato sospeso. Inizia a giocare gratis con 88 giocate in regalo.\n888casinò.it. Il gioco è vietato ai minori e può causare dipendenza.\nDa oggi c'è un altro modo per entrare nel casinò online di 888.it. Il divertimento ti aspetta su\nnpc, ma anche su smartphone e tablet. Scarica subito la app su 888.it. Quando vuoi, dove vuoi! Avrai\noltre venti giochi e centinaia di euro in jackpot. 888.it: il tuo casinò. Condizioni su 888.it. Il gioco\npuò causare dipendenza patologica. Vietato ai minori."

## Tokenization:

processo che permette di suddividere i corpora in tokens. Questo passaggio permette di eliminare tutti i caratteri superflui al fine dell'analisi, nonché il dizionario di stopwords creato ad hoc per questo tipo di documenti

Prima parte del vettore tokens 'poker'

```
[1] "casinò" "fiato" "sospeso" "inizia"  
"giocare" "gratis" "giocate"
```

Prime quattro parole presentate in una matrice DFM comprensiva di tutti e tre i documenti

```
Document-feature matrix of: 3 documents, 4
features (58.3% sparse).
```

```
3 x 4 sparse Matrix of class "dfm"
      features
```

docs	casinò	fiato	sospeso	inizia
poker.txt	55	1	1	2
sport.txt	0	0	0	0
lotto.txt	0	0	0	2

## Type Token Ratio

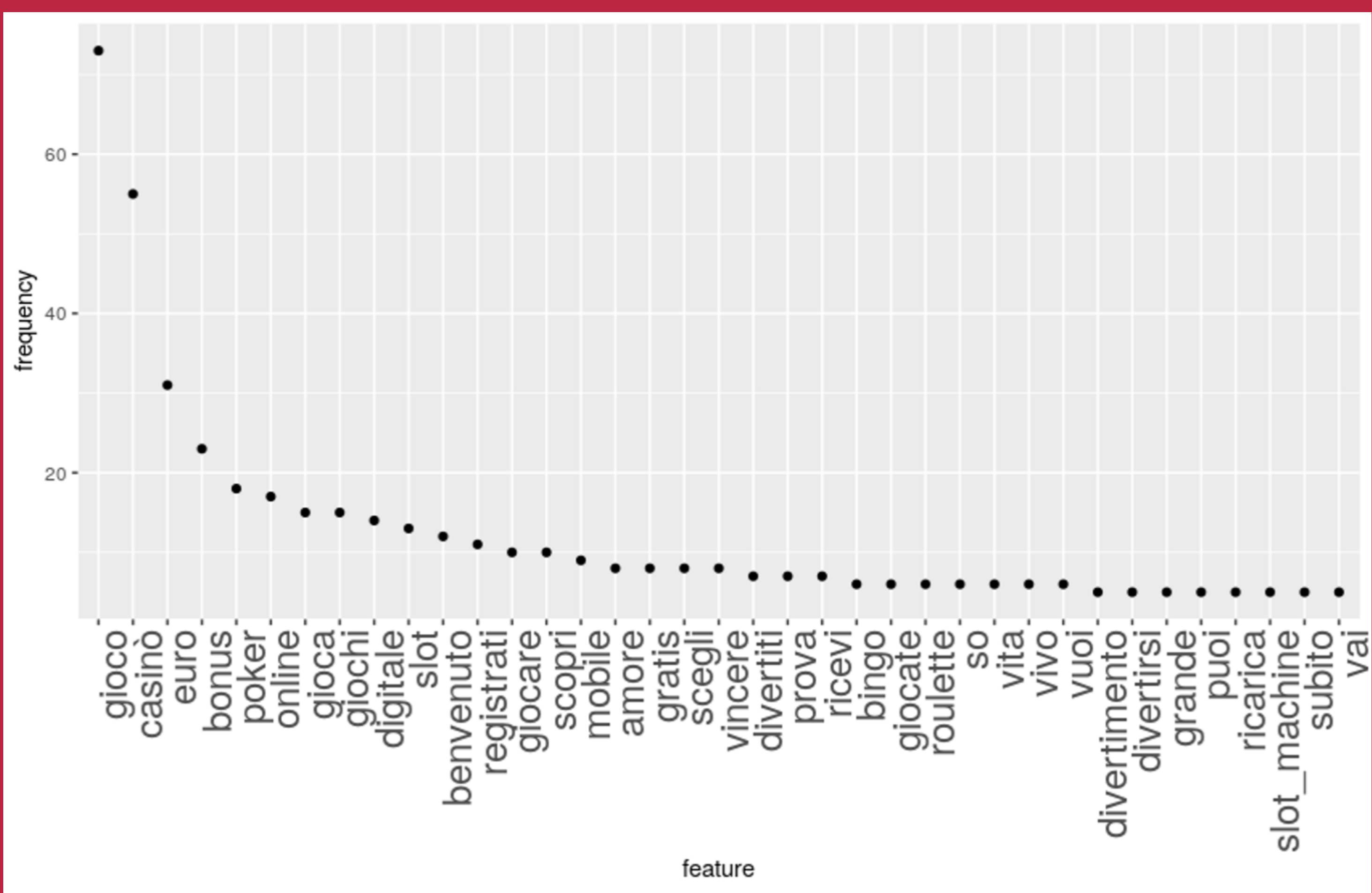
	[,1]	[,2]	[,3]
document	"poker.txt"	"sport.txt"	"lotto.txt"
TTR	0.4507317	0.5781609	0.4573477

## Percentuale di hapax per documento

poker.txt	sport.txt	lotto.txt
31.90244	40.45977	29.85612

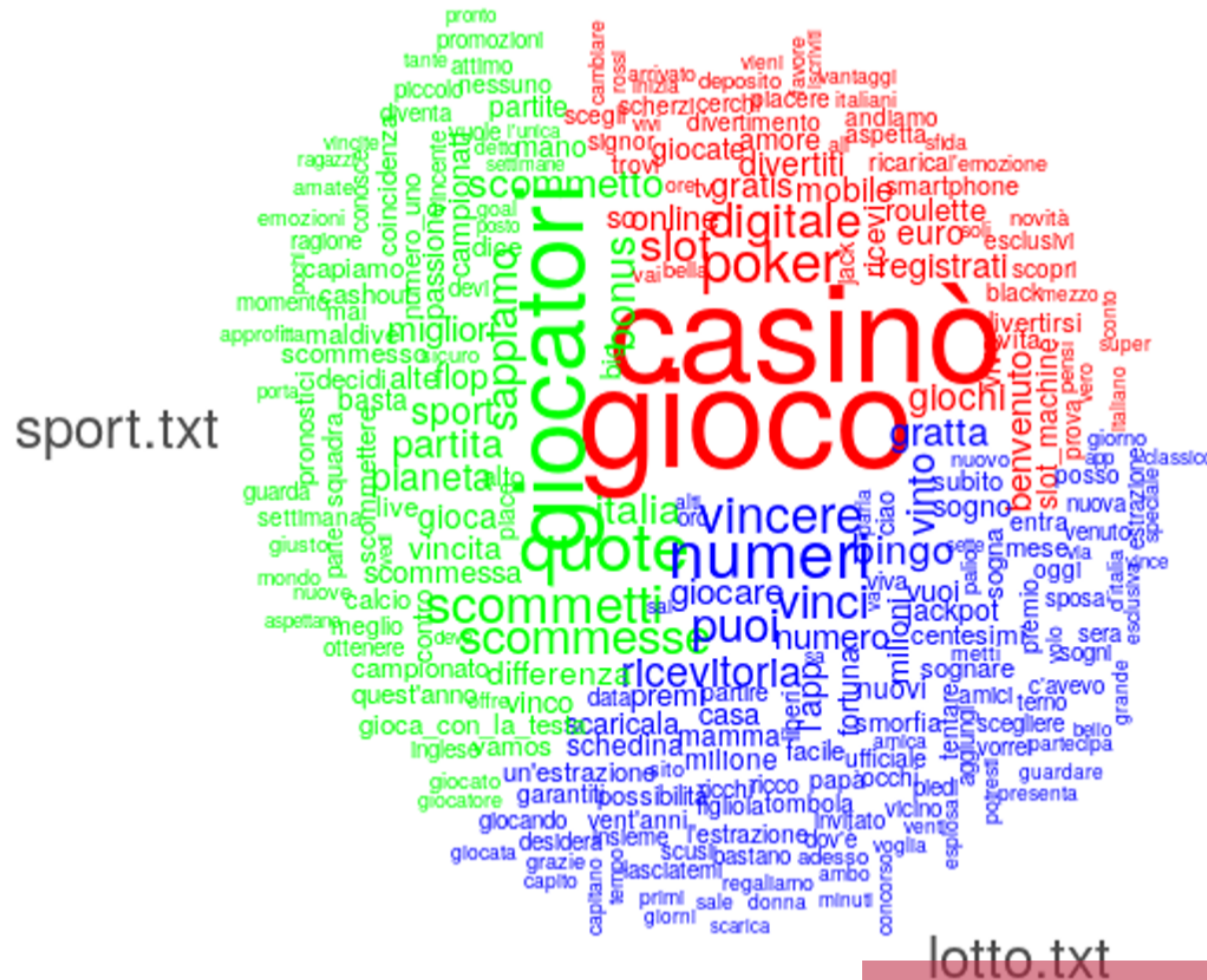


Distribuzione di  
frequenza delle  
prime 40 parole  
appartenenti al  
documento  
'poker'



poker.txt

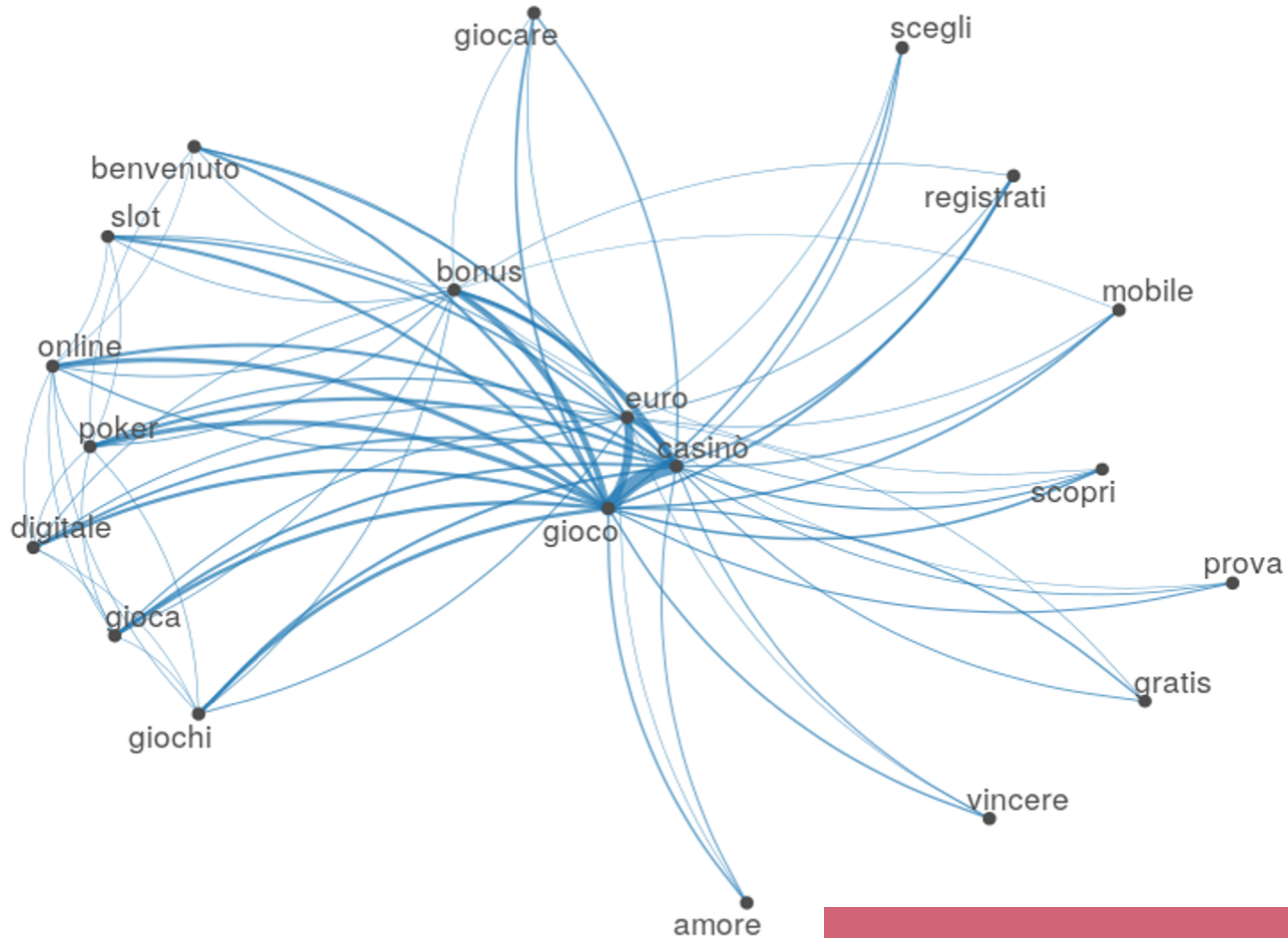
# Wordcloud comparativa dei tre documenti



lotto.txt

# Analisi

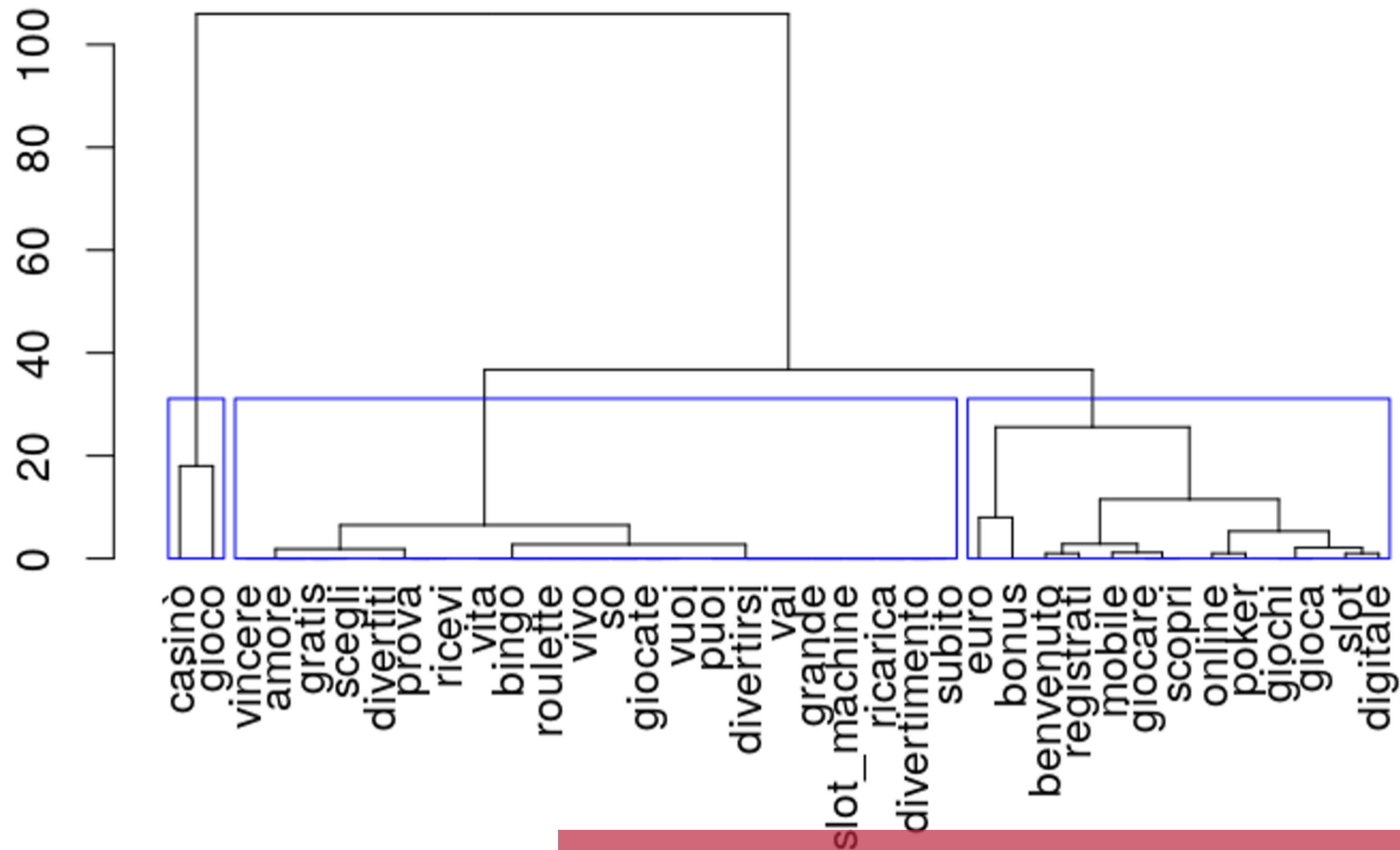
Rete  
semantica  
documento  
'poker'



Analisi

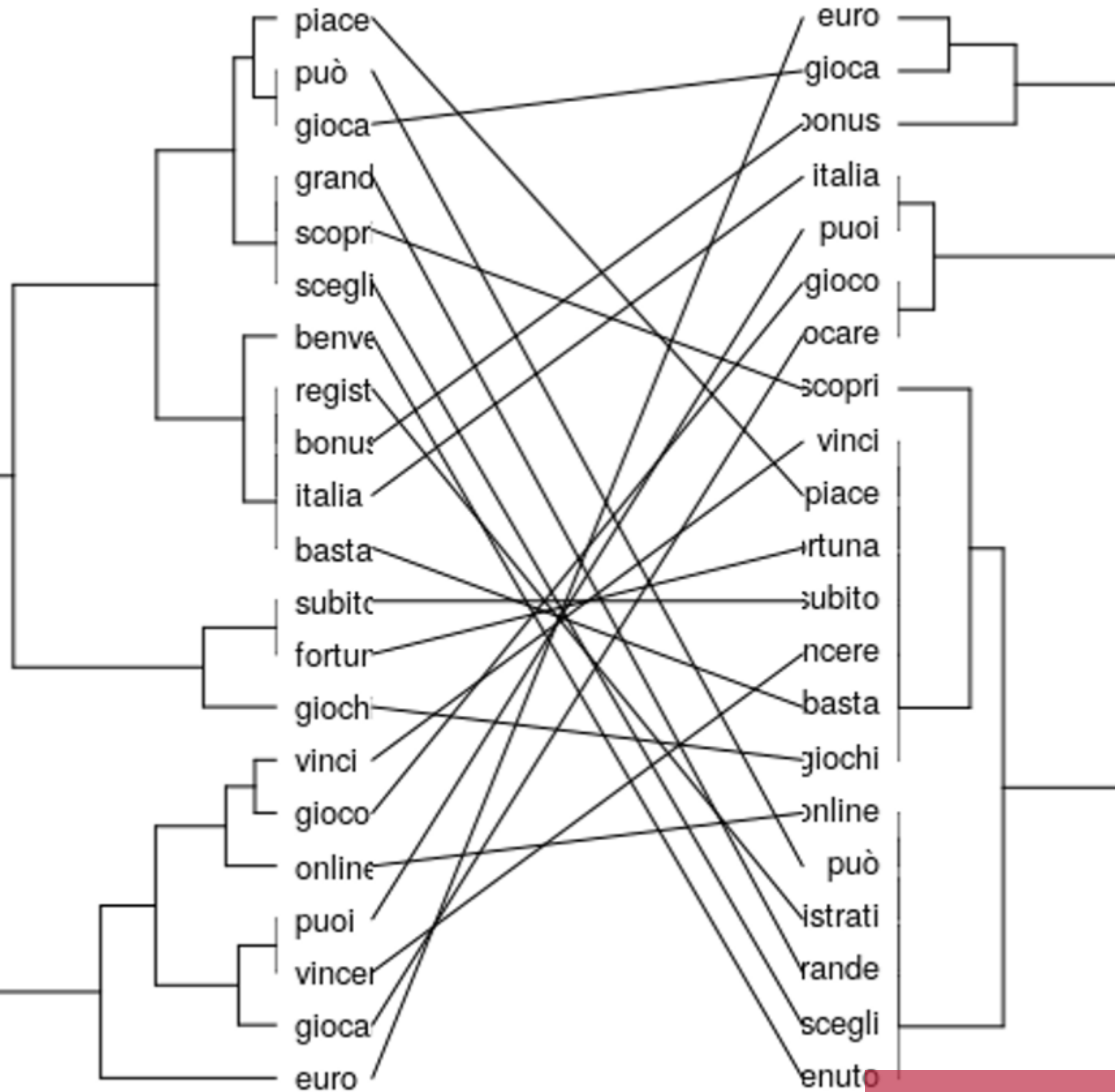


Hierarchical  
clustering  
documento  
'poker'



lotto

sport



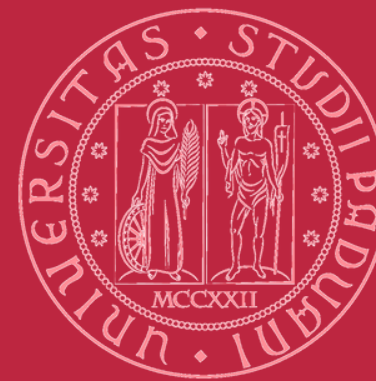
Entanglement  
documento 'lotto' e  
documento 'sport'

Analisi



# CONCLUSIONI

- Messaggi pubblicitari molto semplici
- **'poker'**: messaggio basato su triade gioco-casinò-euro
- **'sport'**; forte leva sul cliente presentato come giocatore/atleta
- **'lotto'**: tema chiave la vincita in denaro



# SPUNTI FUTURI

- Ampliare i dati: raccolta pubblicità presenti nel web
- Studio dei messaggi dei clienti