



# Social Media Analytics: US Election 2020

MONITORAGGIO DELLE ELEZIONI  
AMERICANE DALLA PIATTAFORMA SOCIAL  
TWITTER

*Paparella Matteo*  
*Leuce Francesco 794407*  
*Zottola Gabriele 812363*

January 18, 2021

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Report Goal . . . . .	2
1.2	Analysis Performed . . . . .	2
<b>2</b>	<b>Data Process</b>	<b>2</b>
2.1	Data Extraction . . . . .	2
2.2	Data Exploration . . . . .	3
2.3	Data Pre-Processing . . . . .	3
2.4	Data preparation . . . . .	3
<b>3</b>	<b>Sentiment Analysis</b>	<b>4</b>
3.1	Afinn . . . . .	4
3.2	Vader . . . . .	7
3.3	Textblob . . . . .	8
3.4	OpinionLexicon . . . . .	9
<b>4</b>	<b>Social Network Analysis</b>	<b>11</b>
4.1	Introduction . . . . .	11
4.2	Metodologia . . . . .	11
4.3	Risultati Ottenuti . . . . .	12
4.4	Graph Comparison . . . . .	12
4.4.1	Nodi e Connessioni . . . . .	12
4.4.2	Average Degree . . . . .	13
4.4.3	Density . . . . .	14
4.4.4	Transitivity . . . . .	15
4.4.5	Componenti connesse . . . . .	15
4.4.6	Average Clustering . . . . .	16
4.4.7	Misure di centralità . . . . .	17
4.5	Correlazioni . . . . .	19
<b>5</b>	<b>Analisi singoli nodi</b>	<b>20</b>
<b>6</b>	<b>Conclusione</b>	<b>21</b>
<b>7</b>	<b>Sviluppi Futuri</b>	<b>22</b>

# 1 Introduction

## 1.1 Report Goal

Lo scopo dell'analisi è monitorare il comportamento degli utenti dalla piattaforma social Twitter durante il periodo delle votazioni presidenziali americane. In particolare, si cerca di individuare le differenze di pensiero espresse sui candidati, Joseph Robinette Biden e Donald John Trump, individuare gli utenti più influenti ed osservare le differenze tra i networks legati ai due politici.

## 1.2 Analysis Performed

Il progetto si divide in due sezioni: Sentiment Analysis e Social Network Analysis.

Allo scopo di osservare le differenze di opinioni sui due candidati, per ogni giorno, vengono performati 4 diversi algoritmi di Sentiment Analysis : AFINN, TextBlob, Opinion lexicon e Vader. I valori ottenuti, catalogati come positivi se maggiori di zero, negativi se minori di zero e neutri se uguali a zero, vengono quindi riportati in rapporto al numero di tweets.

Per poter avere una visione d'insieme nell'arco temporale scelto, vengono visualizzati i valori medi tramite line chart.

Si utilizzano, invece, per evidenziare le connessioni tra gli utenti più influenti sul social network, le seguenti metriche e componenti : numero di connessioni, numero di nodi, cricche, relevant community, average degree, componenti connesse, louvain partition, coverage partition, average community order, relevant community, transitivity, self loops, average clustering, density e modularity.

Nella parte di costruzione dei network sono state applicate le librerie Networkx e Graph.

# 2 Data Process

## 2.1 Data Extraction

Per il raggiungimento degli obiettivi, si scaricano i tweets in streaming dal giorno 1 Novembre 2020 al giorno 9 Novembre 2020, in una fascia oraria fissa che varia dalle 18:00 alle 23:00.

Al fine di coprire l'evento, tali tweets, salvati in lingua inglese, sono stati selezionati utilizzando i seguenti hashtags: Biden, JoeBiden, Trump, DonaldTrump, Election2020, MAGA, Vote, ElectionDay, BidenHarris.

I dati vengono scaricati utilizzando l'architettura Kafka (creazione di Kafka producer e Kafka consumer) e caricati su database MongoDB in un formato Json binario compilato.

Per un utilizzo più immediato, si sfrutta la possibilità di conversione dei file, permessa dal database non relazionale, in un formato .csv, i quali sono stati poi scaricati su Drive.

## 2.2 Data Exploration

L'esplorazione dei dati viene effettuata attraverso l'utilizzo della piattaforma Google Colab e Jupyter Notebook con linguaggio di programmazione python. I tweets scaricati si presentano in più dataset, suddivisi per giorno e per hash-tags, per un totale di 4.109.447 istanze.

Durante questa fase, è emerso che le collezioni di dati scaricati presentano tutti i seguenti attributi: ID, Screen-Name, Text, Followers, Date-time, Number-Retweet, Name, Position.

## 2.3 Data Pre-Processing

Nella fase di Data Preprocessing sono state effettuati i seguenti passaggi:

- cancellazione degli attributi considerati ridondanti(name,position,followers)
- estrapolazione degli hastags presenti nel testo dei tweets
- estrapolazione del nome utente "retweettata"
- conversione in Datetime della colonna relativa al giorno e all'orario di download
- unioni dei diversi dati scaricati in un unico DataFrame
- calcolo del numero dei tweets giornalieri scaricati

## 2.4 Data preparation

Per rendere più performanti gli algoritmi di Sentiment Analysis e migliorare i risultati, si adotta una strategia di text cleaning, dove, a seconda del modello, si utilizza il metodo ritenuto più opportuno.

Una prima metodologia, finalizzata ad applicare l'algoritmo TextBlob prevede i seguenti passaggi:

- rimozione di emoticons di diversa natura(bandiere,simboli, etc.)
- rimozione HTML tags, mentions , hashtags, urls, numbers, simboli, pictographs.
- normalizzazione del testo (conversione delle lettere in minuscolo)
- tokenization
- rimozione di "stop-words"
- rimozione della punteggiatura
- aggiunta manuale di stopwords
- lemmatizzazione del testo
- pos-tagging
- visualizzazione dei token più frequenti, allo scopo di validare la pulizia del testo effettuata

Come seconda strategia, invece, per ottimizzare il modello Vader, il testo viene rielaborato diversamente:

- rimozione urls

- rimozione di hashtags references
  - rimozione HTML tags
  - rimozione di solamente una parte delle emoticons (simboli, mappe e bandiere)
- Non si elimina in questo caso la punteggiatura ed emoticons poichè, il punteggio, viene ponderato anche secondo questi ultimi elementi.
- Per l'applicazione del modello "Afinn" si procede all'eliminazione delle seguenti componenti:
- rimozione URLS
  - rimozione hashtags e references
  - rimozione HTML tags
  - rimozione di emoticons (smiles, simboli, mappe e bandiere etc.)

Il testo, per l'utilizzo del modello 'Opinion Lexicon', si processa nel seguente modo:

- rimozione URLS
- rimozione hashtags e references
- rimozione HTML tags
- rimozione di emoticons (smiles, simboli, mappe e bandiere etc.)
- tokenization
- lemmatization

Dopo aver elaborato il corpo dei tweets, il dataset viene suddiviso giornalmente ottenendo così 9 dataframe, ognuno dei quali relativo ad un giorno di streaming. Si definisce in seguito una funzione in grado di assegnare ad ogni tweet il candidato politico a cui si riferisce il testo, identificando quindi, per ogni giorno, due differenti insiemi di dati e di utenti: un primo relativo a Donald Trump ed un secondo relativo a Joseph Biden.

### 3 Sentiment Analysis

La Sentiment Analysis viene effettuata giornalmente con i 4 modelli precedentemente citati.

Per ogni giorno e per ogni dataframe relativo al candidato politico, vengono calcolati i valori dei tweets, catalogati come positivi (valore maggiore di 0), negativi (valore minore di 0) e neutri (valore uguale 0) e calcolate le proporzioni sul totale dei dati.

Al fine di rendere chiare le differenze tra giorni e candidati si visualizzano i risultati finali tramite line chart, costruiti sul software Tableau.

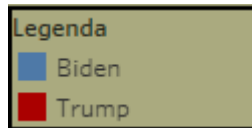
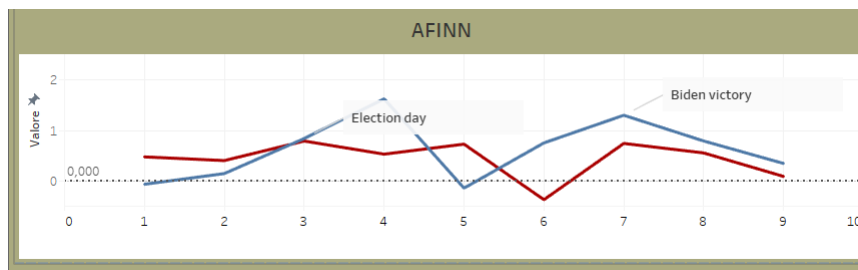
#### 3.1 Afinn

Dall'utilizzo di questo algoritmo emerge quanto segue:

- per il periodo che va dal 1 al 3 novembre, le opinioni espresse dagli utenti Twitter risultano essere maggiormente positive per il candidato repubblicano a discapito dell'avversario democratico che risulta avere un valore medio inferiore a zero

- il giorno di inizio delle elezioni, il 3 novembre, i valori medi risultano essere, per entrambi i candidati, estremamente simili e positivi.
- il 4 novembre (inizio spoglio dei seggi) e il 5 Novembre, il trend, per entrambi i candidati, risulta calare raggiungendo valori negativi di sentiment
- rileviamo un picco per Biden il 7 novembre, data in cui viene dichiarato vincente.

Il grafico viene costruito calcolando la media dei valori di ogni tweet, per ogni giorno e candidato.



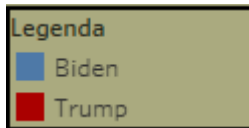
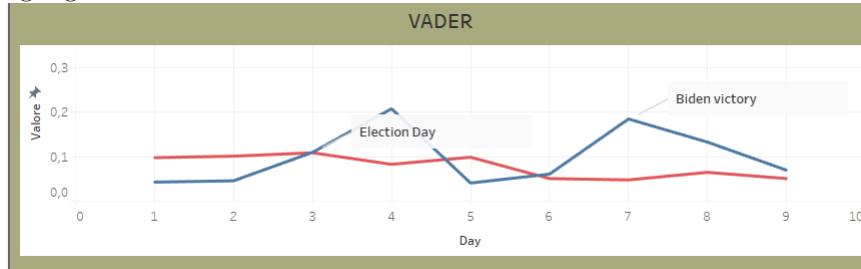
Trump			
Giorni	Positivi	Negativi	Neutrali
1 Novembre	0.503	0.225	0.279
2 Novembre	0.424	0.376	0.2
3 Novembre	0.422	0.350	0.228
4 Novembre	0.382	0.253	0.365
5 Novembre	0.441	0.279	0.280
6 Novembre	0.259	0.411	0.331
7 Novembre	0.382	0.220	0.398
8 Novembre	0.324	0.210	0.466
9 Novembre	0.335	0.242	0.424

Biden			
Giorni	Positivi	Negativi	Neutrali
1 Novembre	0.283	0.304	0.413
2 Novembre	0.336	0.279	0.386
3 Novembre	0.394	0.166	0.440
4 Novembre	0.516	0.149	0.335
5 Novembre	0.360	0.273	0.367
6 Novembre	0.381	0.174	0.445
7 Novembre	0.489	0.107	0.404
8 Novembre	0.364	0.165	0.474
9 Novembre	0.349	0.266	0.385

### 3.2 Vader

Dall'applicazione di questo algoritmo, emergono le stesse conclusioni precisate per il primo metodo utilizzato. L'unica differenza viene osservata nei valori che non scendono mai sotto la soglia della neutralità (zero).

Il grafico viene costruito calcolando la media dei "compound" dei tweet, per ogni giorno e candidat.



Trump				
Giorni	Positivi	Negativi	Neutrali	Compound
1 Novembre	0.097	0.053	0.848	0.105
2 Novembre	0.39	0.15	0.48	0.101
3 Novembre	0.33	0.12	0.5	0.109
4 Novembre	0.47	0.08	0.5	0.083
5 Novembre	0.33	0.1	0.56	0.099
6 Novembre	0.18	0.07	0.51	0.051
7 Novembre	0.014	0.0056	0.3	0.048
8 Novembre	0.28	0.11	0.59	0.065
9 Novembre	0.33	0.12	0.54	0.051

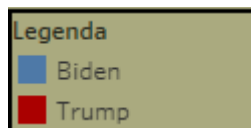
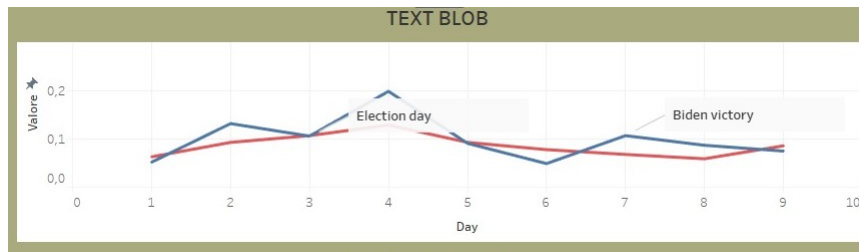


Biden				
Giorni	Positivi	Negativi	Neutrali	Compound
1 Novembre	0.091	0.07	0.836	0.043
2 Novembre	0.882	0.048	0.068	0.046
3 Novembre	0.855	0.046	0.097	0.109
4 Novembre	0.114	0.38	0.516	0.208
5 Novembre	0.076	0.071	0.857	0.041
6 Novembre	0.045	0.022	0.931	0.061
7 Novembre	0.103	0.029	0.867	0.177
8 Novembre	0.093	0.036	0.87	0.133
9 Novembre	0.08	0.05	0.868	0.069

### 3.3 Textblob

L'andamento medio di sentiment, utilizzando la libreria Text Blob, risulta essere molto simile alle due precedenti, con picchi diversi ed una variazione nel 2 novembre. Inoltre, si nota che l'ultimo giorno di osservazione, la sentiment media dei tweets, risulta essere maggiormente favorevole per il candidato repubblicano.

Il grafico viene costruito calcolando la media del valore di polarity per ogni tweet, per ogni giorno e per ogni candidato.



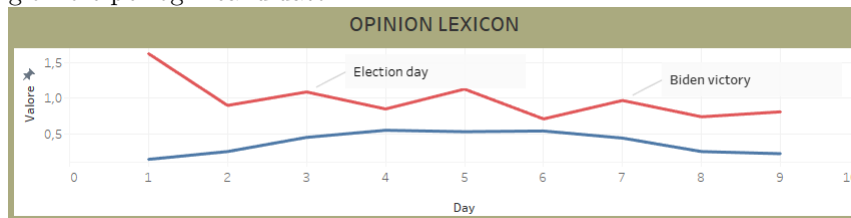
Trump			
Giorni	Positivi	Negativi	Neutrali
1 Novembre	0.31	0.21	0.47
2 Novembre	0.35	0.12	0.48
3 Novembre	0.33	0.15	0.5
4 Novembre	0.37	0.11	0.5
5 Novembre	0.3	0.12	0.56
6 Novembre	0.33	0.15	0.51
7 Novembre	0.019	0.009	0.3
8 Novembre	0.27	0.15	0.57
9 Novembre	0.34	0.12	0.52

Biden			
Giorni	Positivi	Negativi	Neutrali
1 Novembre	0.26	0.15	0.47
2 Novembre	0.39	0.15	0.48
3 Novembre	0.33	0.12	0.5
4 Novembre	0.47	0.08	0.5
5 Novembre	0.33	0.1	0.56
6 Novembre	0.18	0.07	0.51
7 Novembre	0.014	0.0056	0.3
8 Novembre	0.28	0.11	0.59
9 Novembre	0.33	0.12	0.54

### 3.4 OpinionLexicon

I risultati di Sentiment Analysis utilizzando la libreria Opinion Lexicon risultano essere differenti rispetto alle precedenti. Infatti, sia i trend che i picchi si discostano molto. Il trend per Trump risulta essere "altanelante", mentre l'andamento di sentiment di Biden in salita, con picco massimo il 4 Novembre. La funzione di Opinion Lexicon, scritta in modo tale che inverta il valore del tweet nel caso di una negazione ("not") nella frase, risulta avere un forte impatto sull'analisi.

Il grafico viene costruito calcolando la media dei valori dei tweets, per ogni giorno e per ogni candidato.



Legenda	
<span style="color: blue;">■</span>	Biden
<span style="color: red;">■</span>	Trump

Trump			
Giorni	Positivi	Negativi	Neutrali
1 Novembre	0.689	0.115	0.196
2 Novembre	0.648	0.091	0.261
3 Novembre	0.734	0.076	0.186
4 Novembre	0.637	0.149	0.214
5 Novembre	0.763	0.079	0.157
6 Novembre	0.613	0.122	0.265
7 Novembre	0.719	0.088	0.192
8 Novembre	0.609	0.101	0.289
9 Novembre	0.654	0.134	0.211

Biden			
Giorni	Positivi	Negativi	Neutrali
1 Novembre	0.339	0.279	0.383
2 Novembre	0.352	0.173	0.475
3 Novembre	0.418	0.128	0.454
4 Novembre	0.484	0.122	0.395
5 Novembre	0.513	0.189	0.298
6 Novembre	0.510	0.130	0.360
7 Novembre	0.436	0.133	0.431
8 Novembre	0.349	0.155	0.496
9 Novembre	0.370	0.204	0.426

## 4 Social Network Analysis

### 4.1 Introduction

La Social Network Analysis è stata sviluppata tramite l'utilizzo di metriche di grafo e algoritmi di community detection. Come oggetto di confronto, si sono scelti i giorni 3 e 7 Novembre per i seguenti motivi:

- il 3 Novembre è stato il giorno "ufficiale" di inizio delle votazioni, con il maggior numero di americani alle urne e con i primi exit pool e sondaggi di voto
- il 7 novembre viene considerato il giorno in cui il candidato democratico Biden viene dato vincitore.

### 4.2 Metodologia

Si considerano le relazioni tra gli utenti che hanno effettuato i tweets osservando unicamente il fenomeno di "retweet". Per ogni tweet, è stato estrapolato il nome utente di coloro che hanno scritto il messaggio ed il nome utente della persona che ha retweettato.

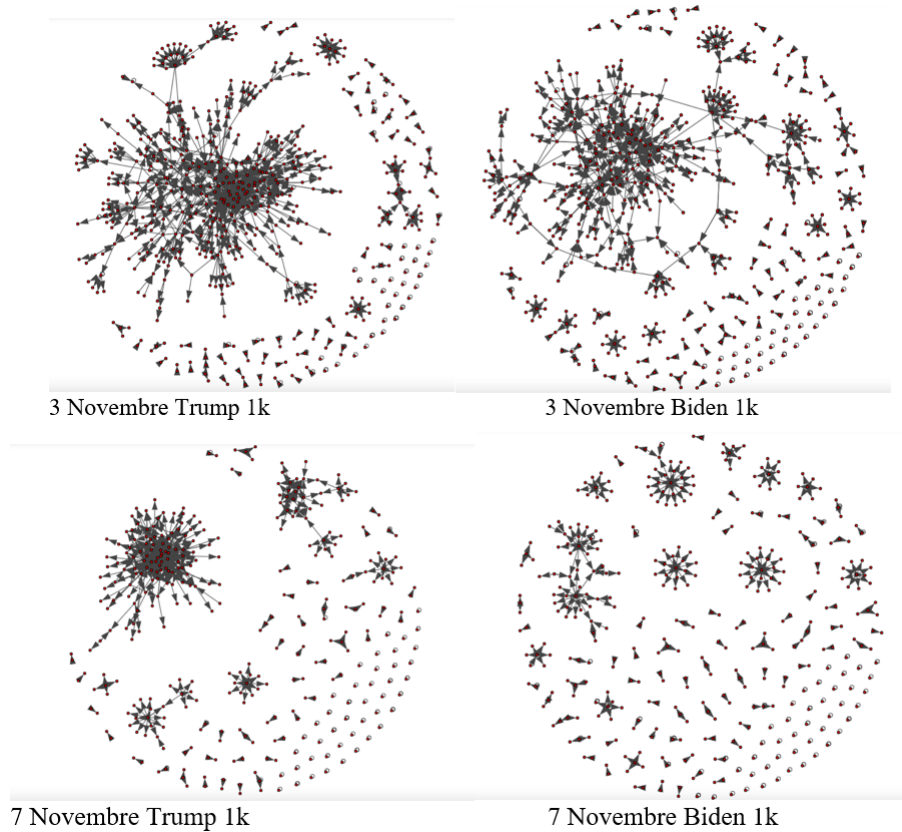
In questo modo, vengono costruite relazioni più solide delle semplici mention.

All'interno di ogni network, sono stati selezionati i primi 1000 utenti che possiedono il maggior numero di connessioni rispetto al network originale. Così facendo, vengono isolati le persone più "influenti" vicine ai due candidati politici.

In seguito all'utilizzo e al confronto di metriche di Social Network Analysis, si è effettuata una fase di community detection, sviluppata applicando l'algoritmo di partizionamento Louvain, più indicato nel caso di network di grandi dimensioni.

### 4.3 Risultati Ottenuti

Si visualizzano i risultati ottenuti a livello grafico.



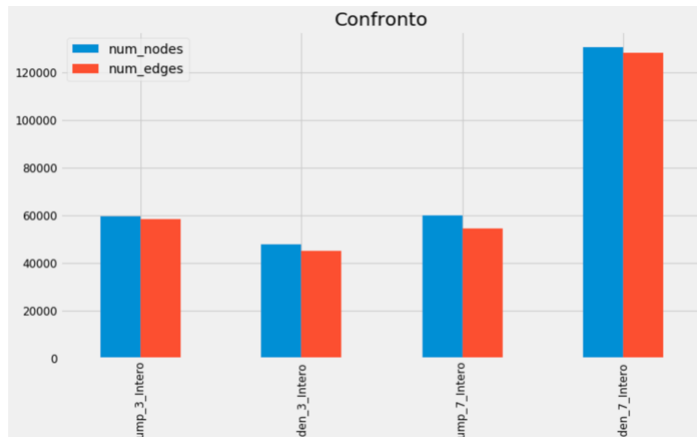
### 4.4 Graph Comparison

Vengono effettuati i confronti, per le metriche utilizzate, sfruttando la visualizzazione a barplot

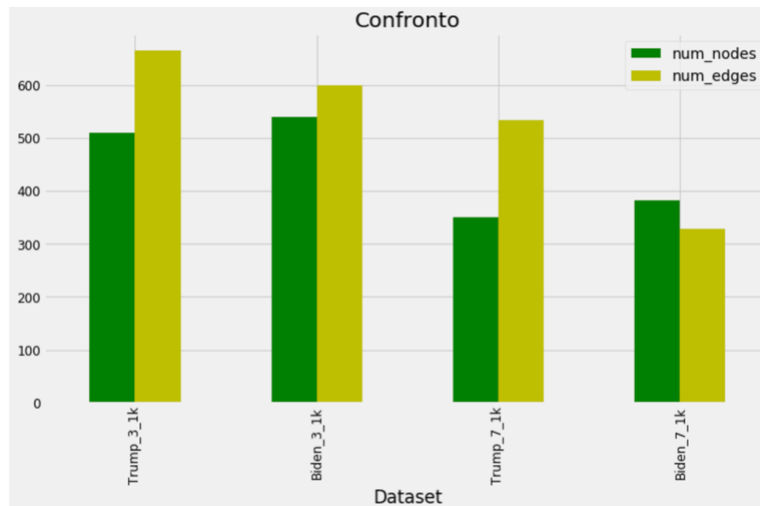
#### 4.4.1 Nodi e Connessioni

Vengono visualizzati i confronti sia per i network interi, che per i network con le prime 1000 persone maggiormente connesse.

Si può osservare il numero di edges e nodi che sono presenti. Il numero di edges è leggermente inferiore al numero di nodi in tutti e quattro i casi. Il dataset con il maggior numero di nodi (e di conseguenza di edges) è il dataset riferito a Biden del 7 novembre.

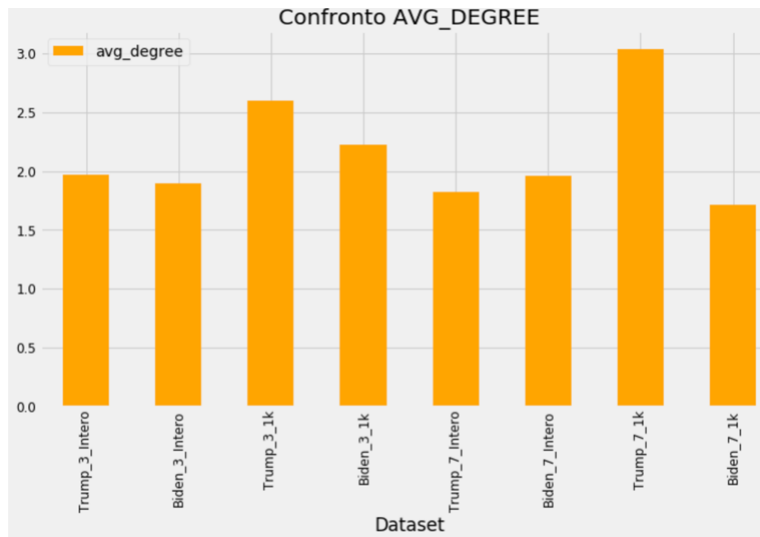


In particolare, si nota come il numero di edges di Trump superi nettamente il numero di nodi, mentre con Biden questa differenza è meno accentuata. Nell'insieme di nodi ed edges del 7 novembre è possibile osservare come il numero di edges è inferiore al numero di nodi.



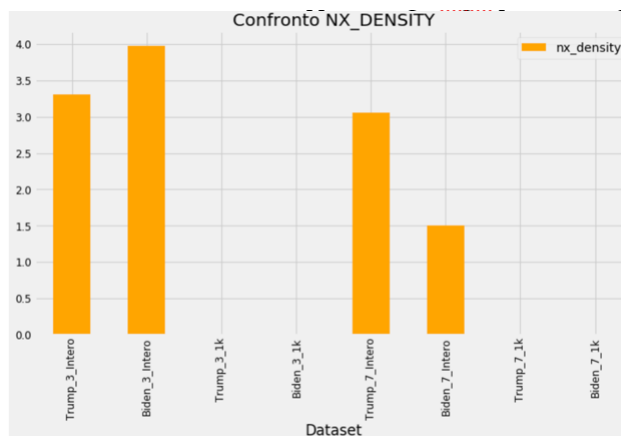
#### 4.4.2 Average Degree

Si osserva che la media di degree per nodo è maggiore per i dataset di Trump riferiti alle sue prime mille connessioni. Il giorno 3 novembre ed il giorno 7 novembre, si nota un picco superiore a 2.5 degree per nodo.



#### 4.4.3 Density

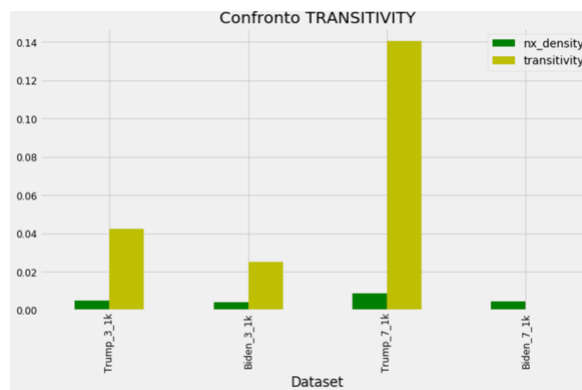
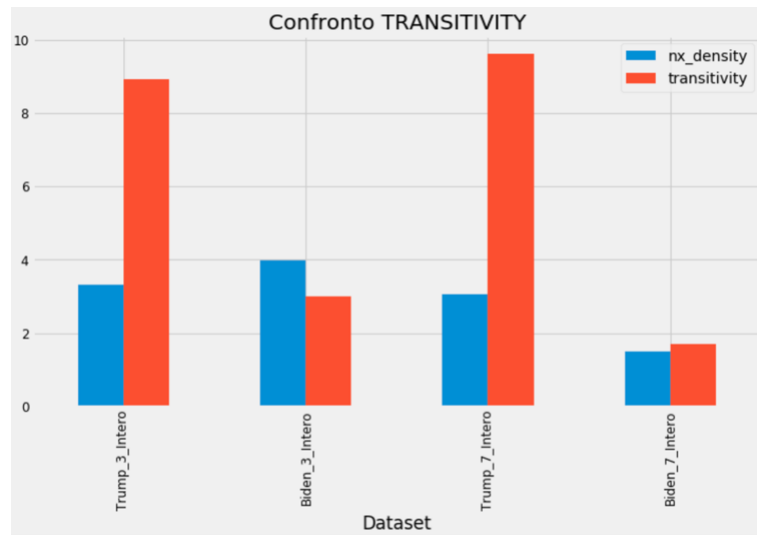
È possibile osservare come la densità per i networks con le prime 1000 componenti con più connessioni sia prossima allo zero. Ciò significa, che il numero di edge possibili sia di gran lunga maggiore rispetto al numero di edge effettivamente presente. Al contrario, si può notare come la densità del grafo relativo a Biden relativo al 3 Novembre maggiore rispetto a tutti.



#### 4.4.4 Transitivity

La transitivity appare rilevante per i networks legati a Trump. Si può ipotizzare che siano presenti delle relazioni maggiori all'interno di chi ha effettuato i tweets legati a trump per i due giorni di riferimento.

Il valore della transitivity è stato confrontato con la metrica density allo scopo di osservare un'eventuale relazione.

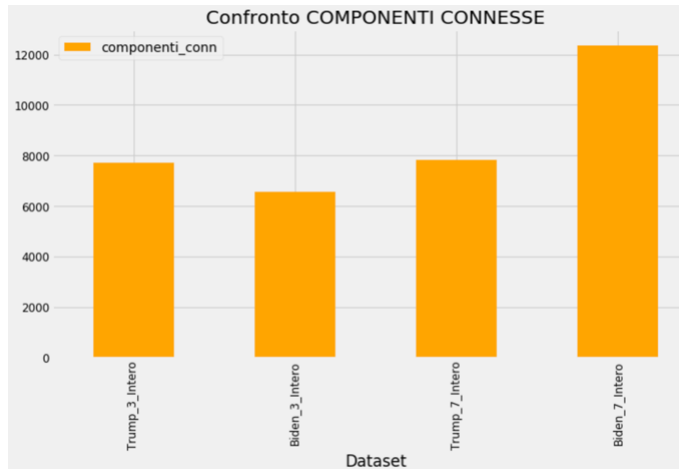


#### 4.4.5 Componenti connesse

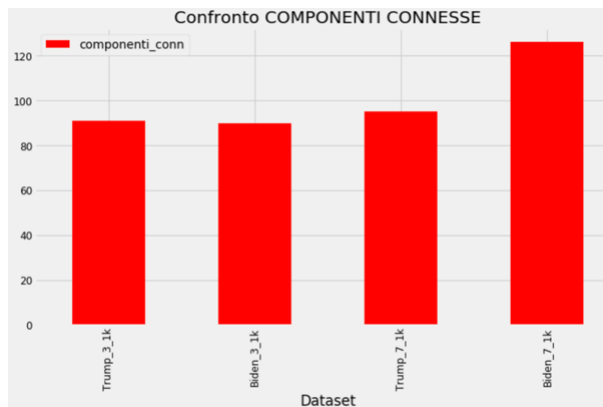
Il maggior numero di componenti connesse sono presenti nel grafo legato a Biden del 7 novembre. Quest'ultimo risulta anche essere il grafo con maggior numero



di nodi ed edges, dimostrando una correlazione tra questi due elementi. Al contrario, i grafi relativi a Trump presentano gli stessi valori.

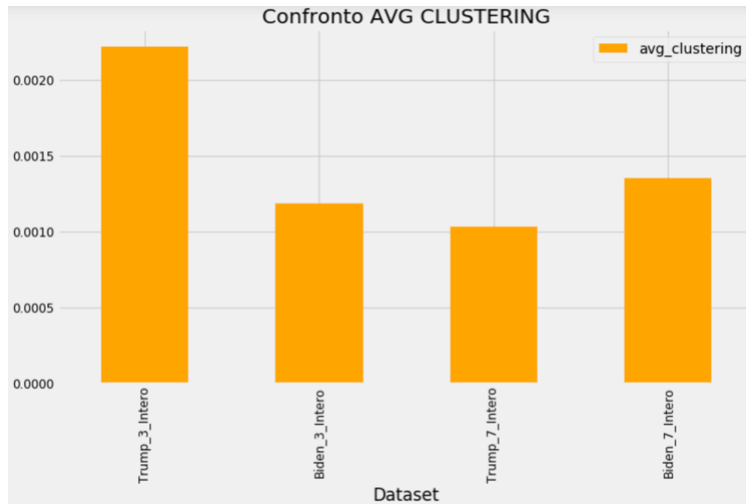


Anche per quanto riguarda i grafi relativi ai primi 1000 utenti con maggior connessioni, si osserva una somiglianza di andamenti con i grafi del network interi. Di conseguenza si nota come Biden presenta un maggior numero di componenti connesse, nonostante il numero minore di nodi ed edges.

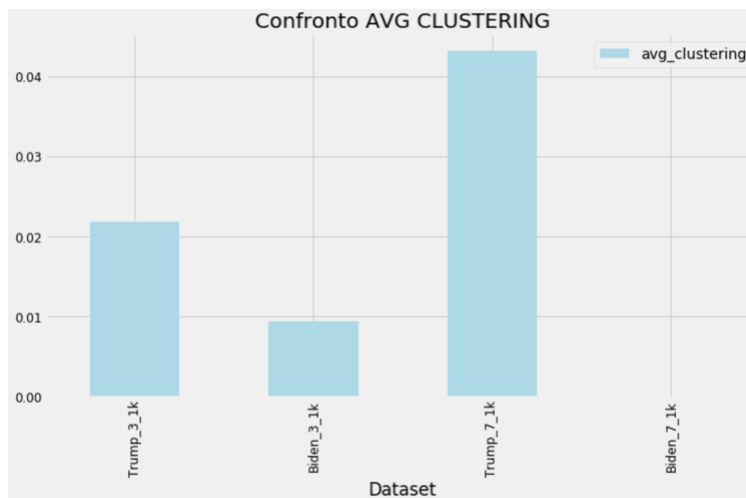


#### 4.4.6 Average Clustering

Analizzando la metrica di "average clustering" si può notare una maggiore, anche se leggera, tendenza a formare cluster nel grafo legato a Trump del 3 novembre rispetto al giorno del 7 novembre.



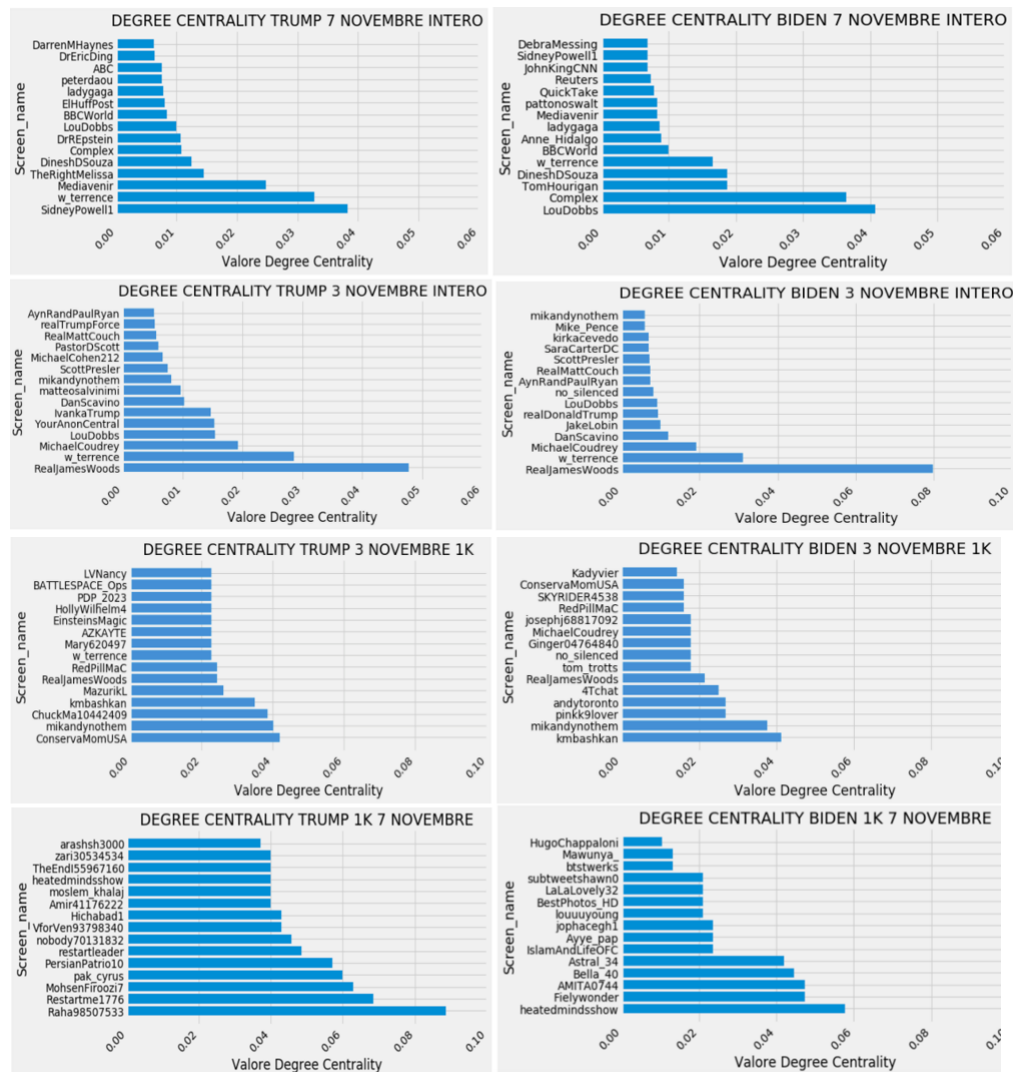
Invece, si osserva come l'average clustering sia maggiore nel dataset di trump del 7 novembre per 1k, nonostante da grafo intero si poteva ipotizzare un maggior valore di average clustering per il 3 novembre.



#### 4.4.7 Misure di centralità

##### Degree centrality

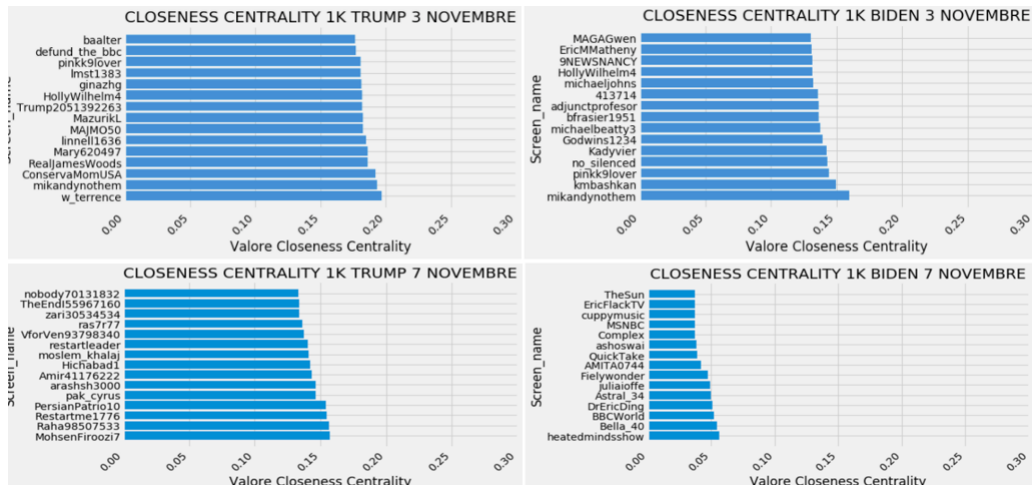
Viene calcolata la degree centrality per ogni tipologia di network allo scopo di osservare quali siano i nodi con maggiori connessioni. Il giorno 3 Novembre presenta valori di Degree Centrality maggiormente elevati. Inoltre, per lo stesso giorno vengono riscontrate delle somiglianze nei nodi, mentre per il giorno 7 Novembre vi sono nodi diversi con degree centrality elevate.



### Closeness centrality

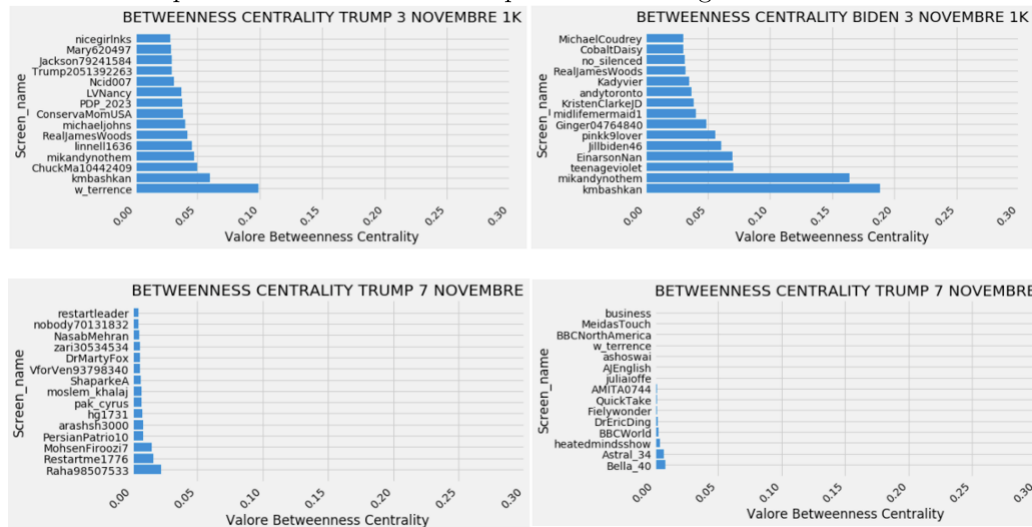
Si calcola la Closeness centrality, metrica grazie alla quale si può notare la vicinanza tra i vari nodi, studiando quindi quali nodi influenzano maggiormente gli altri all'interno del network.

Questa metrica viene applicata per il network delle mille persone maggiormente influenti e vengono visualizzati i nodi con closeness centrality elevata. Si nota come si abbiano valori tendenzialmente più elevati per i nodi nel giorno 3 Novembre, fenomeno che indica vicinanza tra di essi, contrariamente al giorno 7.



### Betweenness centrality

La betweenness centrality si basa sul calcolo di path più brevi che passano per il nodo scelto ad gni rilevazione. Vengono rese uguali le scale dei valori così da permettere i confronti. Da tali confronti, si nota come il giorno 3 Novembre presenti valori di Betweenness centrality maggiori rispetto al 7. Risultano quindi esserci relazioni più strette tra i nodi del 3 rispetto al secondo giorno di confronto.



### 4.5 Correlazioni

Viene effettuato un confronto tra le metriche, valutando il livello di correlazione utilizzando il metodo di Pearson. Al fine di ottenere una miglior visualizzazione,

viene utilizzata una scala a colori, con un blu scuro per una correlazione positiva tendente a 1, mentre un blu chiaro per indicare una correlazione negativa tendente a -1.

	num_nodes	num_edges	avg_degree	nx_density	transitivity	component_conn	self_loops	avg_clustering	cliques	modularity	louvain_partition	coverage_partition	avg_community_order	community_relevanti
num_nodes	1	0.999468	-0.459398	0.570036	0.446974	0.977884	0.990198	-0.508699	0.999478	0.544535	0.977975	0.411359	0.948581	0.84399
num_edges	0.999468	1	-0.448765	0.554531	0.428466	0.972918	0.989887	-0.503377	0.999999	0.532744	0.973	0.386424	0.945477	0.827352
avg_degree	-0.459398	-0.448765	1	-0.53959	-0.452462	-0.522761	-0.484752	0.980213	-0.449063	-0.96705	-0.522567	-0.879471	-0.423067	-0.575413
nx_density	0.570036	0.554531	-0.53959	1	0.788081	0.724392	0.651472	-0.54887	0.554063	0.530148	0.724075	0.328213	0.617164	0.828727
transitivity	0.446974	0.428466	-0.452462	0.788081	1	0.596028	0.512768	-0.44244	0.42772	0.47225	0.586752	0.334868	0.470664	0.785903
component_conn	0.977884	0.972918	-0.522761	0.724392	0.596028	1	0.986282	-0.564024	0.97279	0.584408	0.999999	0.436323	0.944997	0.923781
self_loops	0.990198	0.989887	-0.484752	0.651472	0.512768	0.986282	1	-0.53738	0.989735	0.585232	0.98926	0.388344	0.945114	0.859293
avg_clustering	-0.508699	-0.503377	0.980213	-0.54887	-0.44244	-0.564024	-0.53738	1	-0.500819	-0.967491	-0.563931	-0.782636	-0.517893	-0.586824
cliques	0.999478	0.999999	-0.449063	0.554063	0.42772	0.97279	0.989735	-0.500819	1	0.533117	0.972873	0.389049	0.945477	0.827267
modularity	0.544535	0.532744	-0.96705	0.530148	0.47225	0.584408	0.585232	-0.967491	0.533117	1	0.584314	0.896188	0.499756	0.647015
louvain_partition	0.977975	0.973	-0.522567	0.724075	0.586752	0.999999	0.98926	-0.563931	0.972873	0.584314	1	0.436194	0.945299	0.923855
coverage_partition	0.411359	0.386424	-0.879471	0.328213	0.334868	0.436323	0.388344	-0.782636	0.389049	0.896188	0.436194	1	0.285907	0.510008
avg_community_order	0.948581	0.945477	-0.423067	0.617164	0.470664	0.944997	0.945114	-0.517893	0.945477	0.499756	0.945299	0.285907	1	0.848852
community_relevanti	0.84399	0.827352	-0.575413	0.828727	0.785903	0.923781	0.859293	-0.586824	0.827267	0.647015	0.923855	0.510008	0.848852	1

Si può osservare come l'average clustering ed il numero di nodi presentino una correlazione negativa. Il numero di cliques aumenta nettamente in base all'aumentare del numero di nodi e di edges. L'average clustering presenta delle correlazioni negative con gli altri indicatori, così come anche l'average degree. Invece tra average degree ed average clustering vi è comunque una forte correlazione positiva.

## 5 Analisi singoli nodi

Per rilevare i nodi maggiormente influenti presenti nei networks si utilizza la metrica pagerank, passaggio fondamentale per poter visualizzare i nomi degli utenti più "influenti".

Nell'elenco a sinistra si mostrano gli influencer legati a Donald Trump, a destra gli influencer legati a Joe Biden nel giorno 3 novembre.

Screen_name	pagerank	Screen_name	pagerank
RealJamesWoods	0.016776	RealJamesWoods	0.032392
w_terrence	0.009842	w_terrence	0.011890
MichaelCoudrey	0.006674	MichaelCoudrey	0.007626
YourAnonCentral	0.006358	JakeLobin	0.004750
LouDobbs	0.005166	DanScavino	0.004359
matteosalvinimi	0.004190	realDonaldTrump	0.003432
DanScavino	0.003477	LouDobbs	0.003336
MichaelCohen212	0.002892	AynRandPaulRyan	0.003167
mikandynothem	0.002630	RealMattCouch	0.002899
ScottPresler	0.002408	no_silenced	0.002887

Nell'elenco a sinistra si mostrano gli influencer legati a Donald Trump, a destra gli influencer legati a Joe Biden nel giorno 7 novembre.

Screen_name	pagerank	Screen_name	pagerank
SidneyPowell1	0.015182	LouDobbs	0.017296
w_terrence	0.013146	Complex	0.017183
Mediavenir	0.010325	TomHourigan	0.008823
TheRightMelissa	0.005603	DineshDSouza	0.007710
DineshDSouza	0.004886	w_terrence	0.006981
Complex	0.004580	BBCWorld	0.004157
DrREpstein	0.004249	ladygaga	0.004058
LouDobbs	0.003864	Anne_Hidalgo	0.003895
ElIHuffPost	0.003371	Mediavenir	0.003815
BBCWorld	0.003361	pattonoswalt	0.003808

*RealJamesWoods* = attore e doppiatore statunitense.

*W-terrence* = attore e comico, legato a Trump, ha scritto un libro legato alla casa bianca

*MichaelCoudrey* = attivista e conservatore (Trump) che gestisce social media per i politici.

*SidneyPowell1* = scrittrice libro, sostenitrice Trump

*Mediavenir* = Sito di notizie francesi

*YourAnonCentral* = sito degli anonymous, sito principale.

*LouDobbs* = sostenitore di Trump con anche un libro su di lui, business network.

*Complex* = Pagina orientata al Pop e community persone Pop.

*ChuckMa10442409* = Conservatore, a favore di Trump.

*Abdytoronto* = attivista per i diritti umani sotto vari fronti.

*Heatedmindsshow* = Giornalista senior di BBCnews

*Heatedmindsshow* = international Thought Provoking Entertainment

## 6 Conclusione

Confrontando i risultati degli algoritmi di Sentiment Analysis Afinn, Vader e Text blob, con i fatti di cronaca riportati dai "Mass Media", durante il periodo di osservazione, si può notare una coerenza tra i due fenomeni. Infatti, laddove si sono verificati eventi particolarmente favorevoli per uno dei due candidati politici (vittoria in uno stato/acquisizione di seggi), si ottengono dei picchi riconducibili ad un numero di tweet fortemente positivi.

Meno netta invece è la corrispondenza "sentiment-cronaca" per Opinion lexicon dove, l'inversione del valore in presenza di negazione, cambia l'andamento delle rette. Si nota infatti un trend crescente per Biden durante i giorni di spoglio ma con risultati sempre sotto quelli relativi a Trump.

Lo svolgimento della social network analysis consente di avere una visione maggiormente dettagliata per quanto riguarda i giorni del 3 novembre e del 7 novembre, nonché di Trump e Biden. Da una prima osservazione grafica, si nota che i giorni presentavano delle differenze, in particolare il giorno del 3 novembre è

stato caratterizzato da un maggiore grado di connessione tra gli utenti, dato confermato dalle metriche di centralità con valori tendenzialmente sempre maggiori rispetto al 7 novembre. Inoltre, Trump ha presentato in entrambi i due giorni una tendenza ad avere utenti maggiormente connessi rispetto a Biden, fatto dimostrato sia dalle metriche di centralità, ma anche da average degree, density, average clustering e transitivity. Nonostante per il giorno 7 novembre vi siano a disposizione più tweets e di conseguenza un maggior numero di nodi e archi per il dataset intero, i valori per alcune metriche, si sono dimostrati in linea anche per i network con i primi mille utenti con più connessioni. Grazie alle metriche di centralità si sono osservati gli utenti più “influenti”, mostrando una tendenza dichiaratamente a favore di Trump, rispetto a quasi nessun utente dichiaratamente a favore di Biden. In conclusione i sostenitori di Trump si dimostrano più “coinvolti” a livello di network.

## 7 Sviluppi Futuri

Al fine di poter compiere un’analisi più accurata, si ipotizzano le seguenti ottimizzazioni:

- arco di tempo di osservazione maggiore (i.e. un mese)
- analisi di community detection e individuazione giornaliera delle persone maggiormente “influenti”, con lo scopo di osservazione più dettagliate.