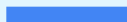


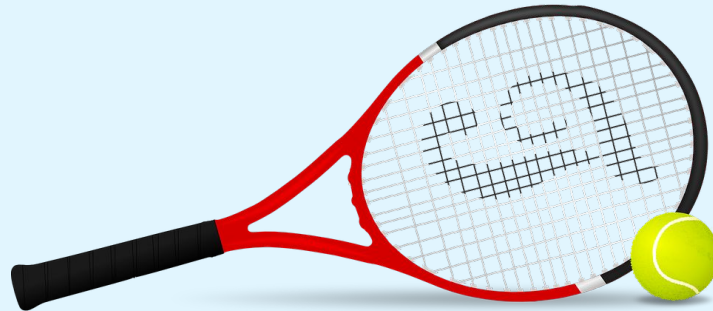
US OPEN 2021



Che cos'è lo US Open?

Lo US Open è il quarto e ultimo dei tornei di tennis del Grande Slam. Il torneo principale è composto da cinque categorie: singolare maschile e femminile, doppio maschile, femminile e misto.

Nello specifico sono andata ad analizzare l'ultima giornata di questi tornei.



SOCIAL MEDIA

SOCIAL NETWORK:

Twitter è un servizio di notizie e microblogging che consente di postare brevi messaggi di testo chiamati tweet, che vengono visualizzati nella pagina principale dell'utente.

DATA GATHERING:

Sono andata a scaricare i tweet riferiti all'hashtag #USOpenFinal attraverso le REST API messe a disposizione da Twitter dell'ultima giornata degli US Open.

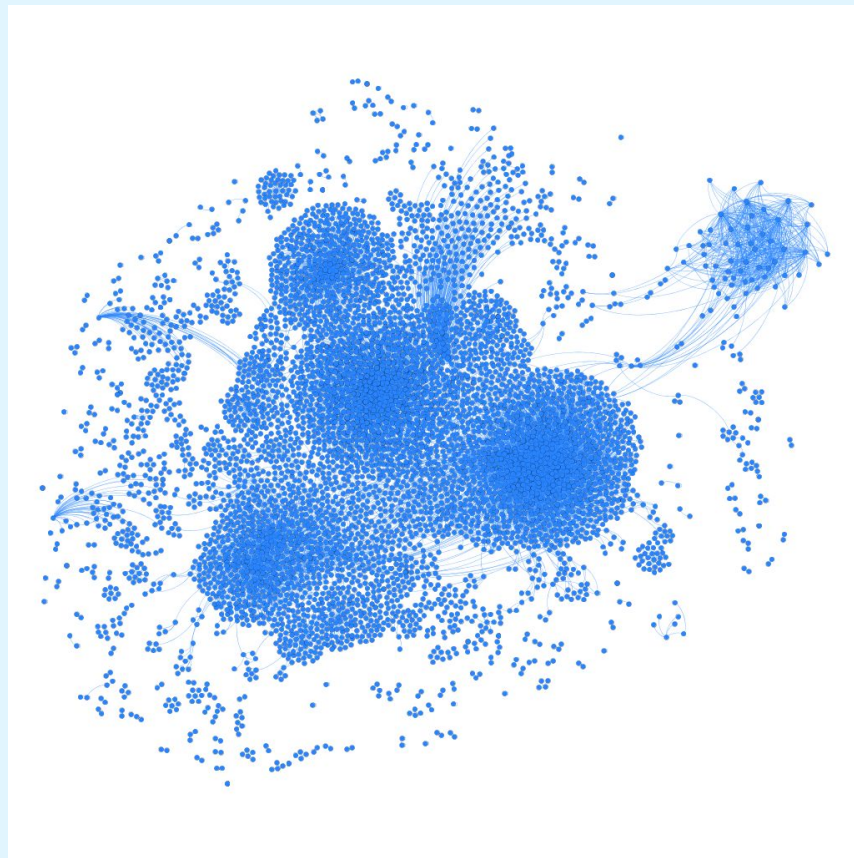
NETWORK ANALYSIS

NETWORK GRAPH

Ho un grafo orientato dove:

- I nodi della network sono rappresentati dagli utenti
- Gli out-link sono le menzioni che un utente ha fatto nei suoi tweet
- Gli in-link sono le menzioni ricevute da altri utenti

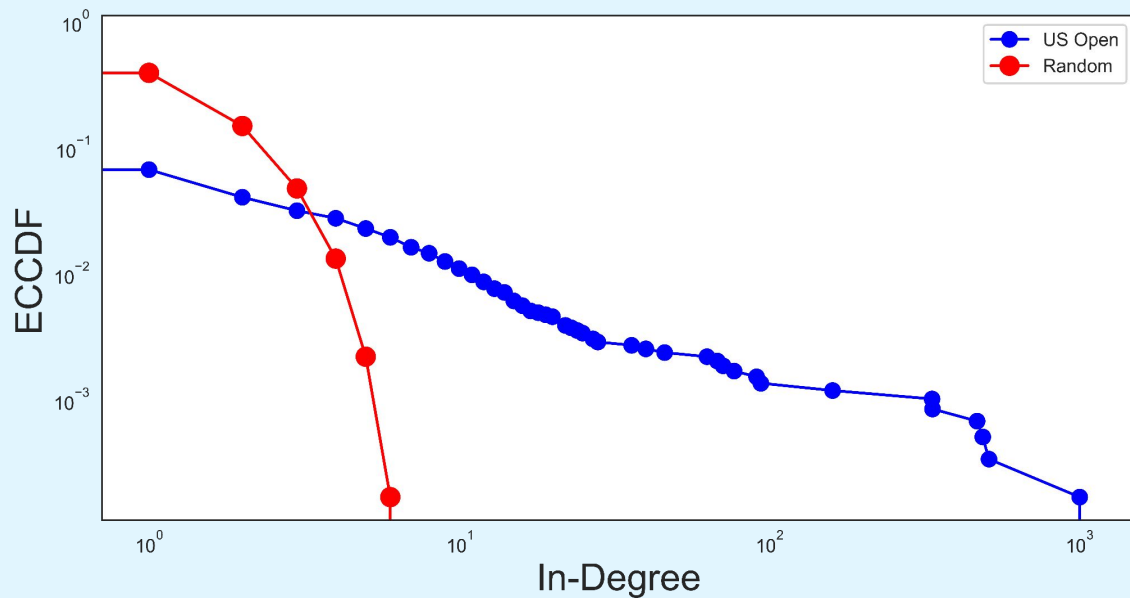
NETWORK GRAPH



Nodi: 5876

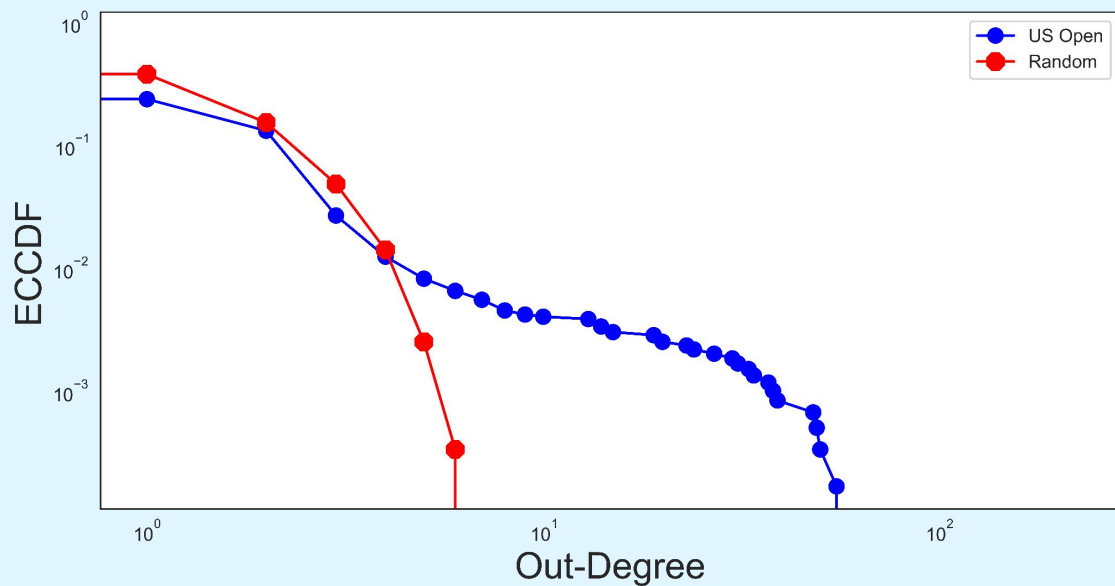
Archi: 7968

Confronto random network in-degree



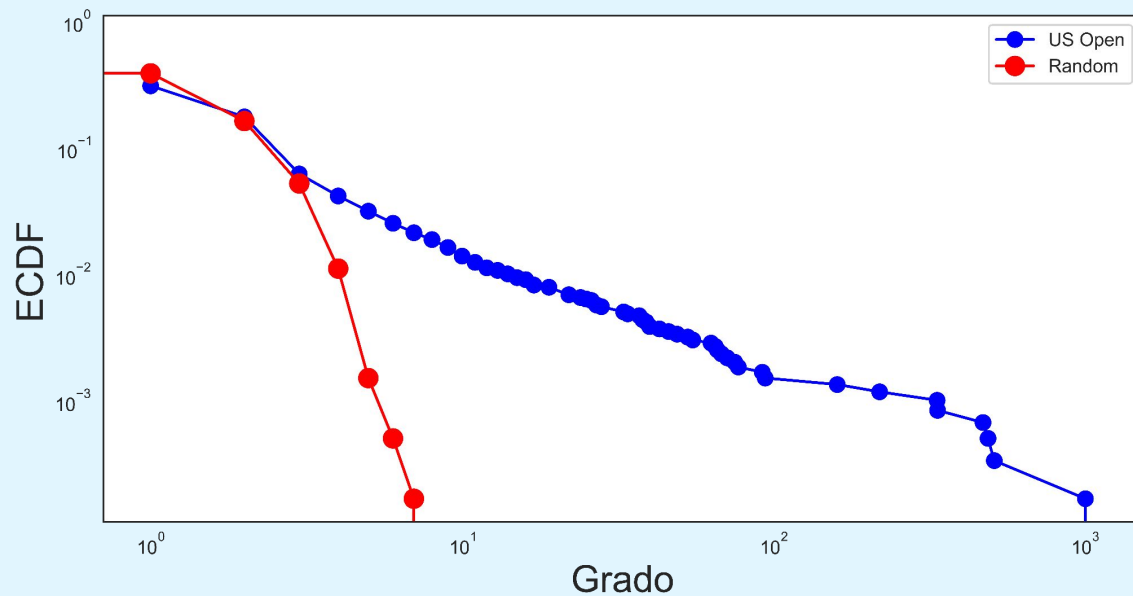
Statistiche	Random	USOpen
Grado Medio	1.347175	1.356025
Deviazione standard	1.156843	23.240929
Mediana	1.000000	0.000000
Grado Massimo	7.000000	1072.000000
Grado Minimo	0.000000	0.000000

Confronto random network out-degree



Statistiche	Random	USOpen
Grado Medio	1.347175	1.356025
Deviazione standard	1.166366	3.556679
Mediana	1.000000	1.000000
Grado Massimo	7.000000	218.000000
Grado Minimo	0.000000	0.000000

Confronto random network



Statistiche	Random	USOpen
Grado Medio	1.365214	2.712049
Deviazione standard	1.174833	23.494929
Mediana	1.000000	1.000000
Grado Massimo	8.000000	1072.000000
Grado Minimo	0.000000	1.000000
Densità	0.000232	0.000231

HUB

Ponendo il percentile a 99 otteniamo un grado di 15 e contiamo 60 Hub all'interno della nostra Network.

Andando a vedere qual è l'Hub con grado maggiore notiamo che non è uno dei partecipanti al torneo.

id_str	screen_name	grado	degree centrality
119691332	@rorybremner	1072	0.182468

Andando ad analizzare il grado dei giocatori dei tornei del singolo femminile e del singolo maschile vediamo che anch'essi sono degli Hub. In particolar modo @EmmaRaducanu risulta avere un grado molto vicino all'Hub con valore massimo.

id_str	screen_name	grado	degree centrality
1197994613665083394	@EmmaRaducanu	1002	0.170553
259925559	@DjokerNole	336	0.057191
719870195695816704	@DaniilMedwed	335	0.057021
1194627927272542209	@leylahfernandez	77	0.013106

COMPONENTI CONNESSE

Il nostro grafo orientato non risulta né essere un grafo fortemente connesso né debolmente connesso.

Esistono Giant Component?

Componenti debolmente connesse: 408

La più grande componente debolmente connessa comprende 4367 nodi e 6757 link, ovvero il 74.32% dei nodi.

Possiamo quindi affermare che è presente una Giant Component debolmente connessa molto grande che include quasi i 3/4 dei nodi della rete.

COMPONENTI CONNESSE

Componenti fortemente connesse: 5847

La più grande componente fortemente connessa comprende 19 nodi e 172 link, ovvero lo 0.32% dei nodi.

Possiamo quindi affermare che non è presente alcuna Giant Component fortemente connessa all'interno della rete, ma sono presenti un elevato numero di componenti fortemente connesse di piccole dimensioni.

RECIPROCIÀ

Calcolata come il rapporto fra il numero di coppie reciproci su il numero di coppie totali.

Reciprocità: 0.0138027

Solo l'1.3% degli archi è reciproco. Una reciprocità molto bassa è tipica delle Information Network.

TRANSITIVITÀ

Coefficiente di Clustering Globale: 0.005755

In questo caso la probabilità di trovare triadi chiuse è 0.5%, quindi molto bassa. Ciò dimostra che siamo in presenza di un grafo poco interconnesso.

Coefficiente di clustering locale medio: 0.120743

In media i nodi hanno il 12% di probabilità che i suoi vicini si connettano. Minore è l'average path length maggiore è la vicinanza dei nodi e la possibilità di raggiungere un nodo non connesso. Questo fenomeno è detto small-world phenomenon, un effetto che è visibile sia nelle social network che nelle network reali.

SENTIMENT ANALYSIS

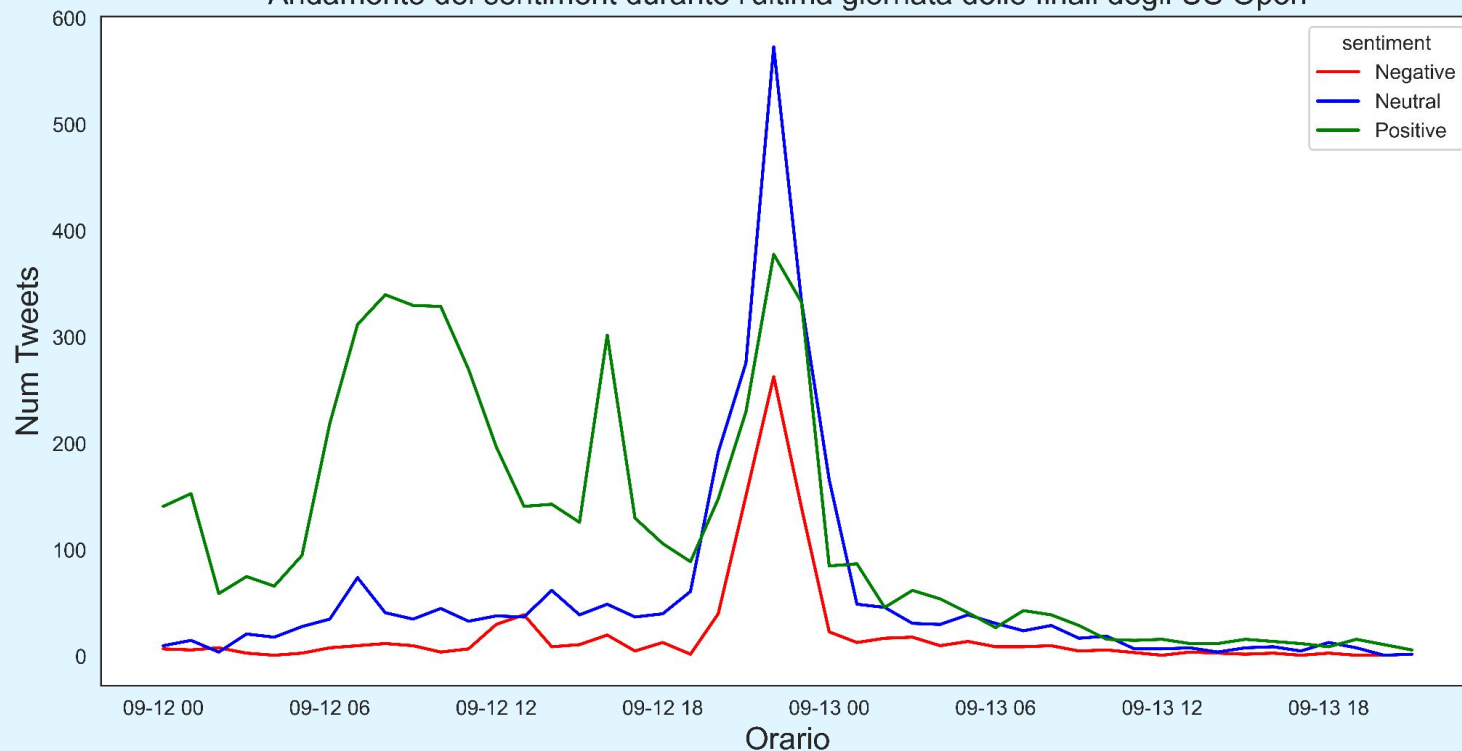
DOMANDA DI RICERCA

Mi sono chiesta come potesse essere il sentimento generale durante l'ultima giornata degli US Open.

Nello specifico come fosse il sentiment del pre-partita nei confronti dei giocatori della finale del singolo maschile per vedere quale potesse essere il favorito e come questo poi sia cambiato o meno durante e dopo la partita.

SENTIMENT ANALYSIS

Andamento del sentiment durante l'ultima giornata delle finali degli US Open



Positive: 59.88%

Neutral: 29.49%

Negative: 10.63%

SENTIMENT ANALYSIS

Sono andata ad analizzare il Sentiment nei confronti dei giocatori del pre-partita, della partita e del post-partita.

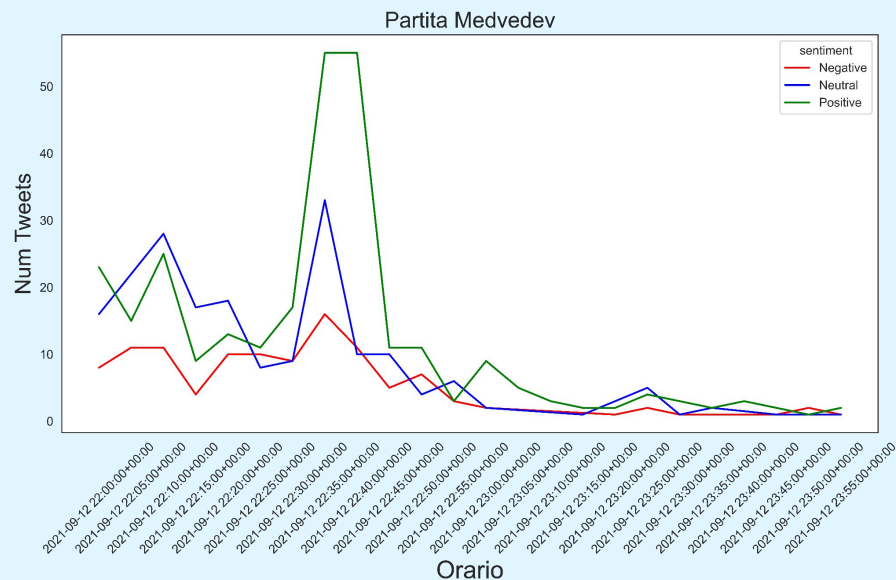
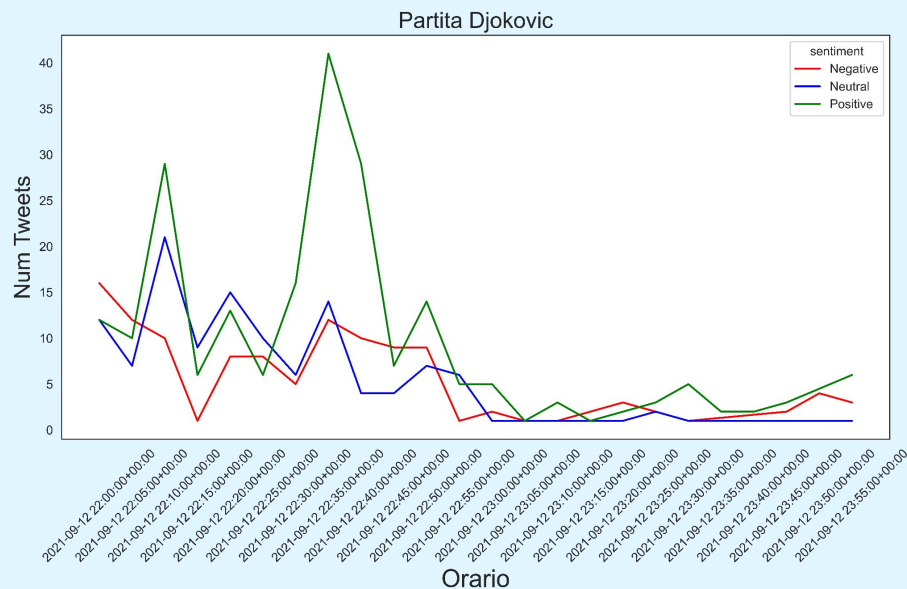
Quello che volevo fare era:

- Dedurre quale fosse il favorito prima della finale.
- Vedere se durante la partita il sentimento generale è cambiato.
- Vedere in base a chi vinceva se ciò era deducibile dall'analisi del Sentiment.

Giocatore	Pre Partita	Partita	Post Partita
@DaniilMedwed	0.67353	0.6955	0.8316
@DjokerNole	0.60878	0.5994	0.8072

SENTIMENT ANALYSIS

Possiamo notare l'andamento generale del sentiment dei tweet durante la partita per entrambi i giocatori.



POSSIBILI SVILUPPI FUTURI

- ➔ Si potrebbe effettuare Community Detection per vedere se si vengono a creare delle comunità significative attorno ai giocatori
- ➔ Estendendo il dataset anche alla finale dei tornei del singolo femminile si potrebbe andare a vedere quali sono i tornei più seguiti/che portano maggiori interazioni su Twitter tra singolo femminile e singolo maschile