

Analisi di rete sui featuring nel rap italiano

Nicola Artuso, Gabriele Corrente

June 15, 2023

1 Introduzione

Nell'ultimo decennio la rilevanza della musica rap nella discografia italiana è cresciuta esponenzialmente. Allo stesso modo, le collaborazioni tra artisti hip hop sono diventate sempre più frequenti. Per questo abbiamo deciso di costruire e analizzare una rete indiretta pesata che riguardasse proprio queste collaborazioni, in gergo "featuring", prendendo come nodi i principali rapper italiani e come archi pesati il numero di collaborazioni tra due artisti. Abbiamo poi creato delle variabili di nodo quali zona e decade, che rappresentano rispettivamente la regione di provenienza (intesa come nord, centro e sud Italia) e la decade di debutto degli artisti, per verificare se e come una o entrambe influissero sul numero di featuring.

2 Dati¹

2.1 Raccolta e pulizia

Al fine di creare la rete, è stato necessario individuare innanzitutto gli artisti da prendere in considerazione, ovvero i nodi. Per farlo, attraverso l'API di Spotify abbiamo preso la lista dei 200 artisti hip-hop (italiani e non) più ascoltati in Italia, ordinandola per popolarità intesa come numero di followers. Questa lista è stata poi modificata, mantenendo solo gli artisti italiani con più di 100.000 followers ed eliminando tutti i gruppi (ad esempio Club Dogo e Articolo31) e i produttori. La lista finale presenta 79 rapper.

Una volta ottenuta la lista degli artisti, per poter valutare il numero di collaborazioni tra questi, abbiamo deciso di fare web scraping dal sito deezer.com, altra piattaforma musicale che, come Spotify, presenta tutte le uscite ufficiali degli artisti che abbiamo selezionato. Per ognuno dei 79 nomi presenti nella lista, tramite uno script implementato in Python abbiamo ottenuto le informazioni di tutte le canzoni pubblicate dall'artista che presentassero un featuring. Abbiamo quindi selezionato solo le canzoni che avessero una collaborazione con uno o più artisti presenti nella lista e infine abbiamo assegnato a ciascuna coppia il numero di collaborazioni.

2.2 Dataset dei nodi

Una volta ottenuta la lista degli artisti, abbiamo assegnato a ciascuno di essi un id e, manualmente, abbiamo aggiunto due covariate: zona e decade. Zona riguarda il luogo di formazione musicale dell'artista, e può assumere valore "nord", "centro" o "sud". Decade invece riguarda il decennio in cui l'artista ha pubblicato il suo primo progetto ufficiale, e nel nostro caso può assumere valore 1980, 1990, 2000, 2010 o 2020. Nella tabella 1 sono riportati come esempio le prime 5 righe del dataset dei nodi.

Nella tabella 2 è invece riportata la tabella delle frequenze per le due variabili Zona e Decade.

2.3 Dataset degli archi

Per quanto riguarda il dataset degli archi, abbiamo inserito nella prima colonna l'id del primo artista (source), nella seconda quello del secondo artista (target) e nella terza il numero di collaborazioni tra essi, che abbiamo chiamato weight, dato che useremo questo numero per dare un peso agli archi. Da notare che il ruolo di source e target è intercambiabile, e che ciascuna coppia è presente una e una sola

¹I dati ricavati sono aggiornati al 26 maggio 2023

id	Label	Zona	Decade
1	Sfera Ebbasta	Nord	2010
2	Capo Plaza	Sud	2010
3	Ghali	Nord	2010
4	Salmo	Sud	1990
5	Guè	Nord	1990

Table 1: Prime 5 righe del dataset dei nodi.

Zona/Decade	1980	1990	2000	2010	2020	Totale
Centro	0	2	4	13	0	19
Nord	0	5	9	21	8	43
Sud	1	3	2	10	1	17
Totale	1	10	15	44	9	79

Table 2: Frequenze assolute per Zona e Decade.

volta. Proprio perchè parliamo di collaborazioni, immaginiamo la rete come indiretta e pesata. Nella tabella 3 è possibile vedere le prime 5 righe del dataset degli archi.

Source	Target	Weight
23	3	3
23	8	14
23	9	6
23	10	2
23	19	2

Table 3: Prime 5 righe del dataset degli archi.

3 Analisi

3.1 Adattamento del dataset e creazione del grafo

Conseguentemente alla creazione (illustrata nelle sezioni 2.2 e 2.3) e all'importazione dei dataset in *RStudio*, è stato creato il grafo in R mediante l'utilizzo dell'apposita funzione `graph from data frame`, utilizzando come archi le colonne Source e Target del dataset degli archi. Sono stati poi aggiunti come etichette per i nodi i valori della colonna Label del dataframe dei nodi, cioè i nomi degli artisti. Il grafo è stato reso indiretto mediante la funzione `simplify()`, in modo da eliminare tutte le coppie del tipo (Source, Target) = (Target, Source).

3.2 Rappresentazioni grafiche

Mediante l'ausilio del software *Gephi* la rete è stata rappresentata graficamente utilizzando l'algoritmo *ForceAtlas*, quindi facendo in modo che le coppie di artisti con numerose collaborazioni fossero più vicini tra loro rispetto agli altri. Abbiamo quindi prodotto tre grafi basati sul grado e sulle covariate dei nodi. Di seguito sono rappresentati:

- grafo con nodi con grandezza e intensità di colore proporzionali al grado del nodo stesso (figura 1),

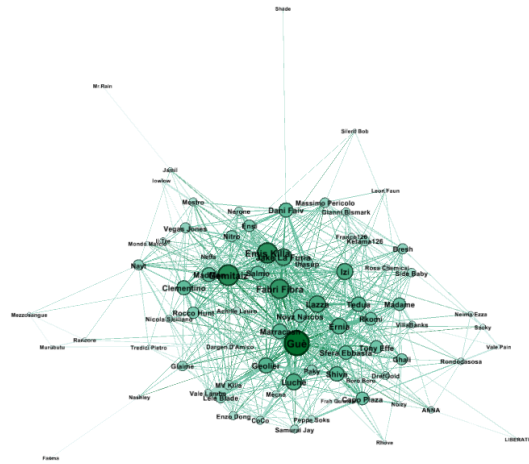


Figure 1: grafo con nodi con grandezza e intensità di colore proporzionali al grado del nodo stesso.

- grafo con nodi colorati in base alla zona di formazione dell'artista (figura 3),
 - in viola il Nord
 - in arancione il Centro
 - in verde il Sud

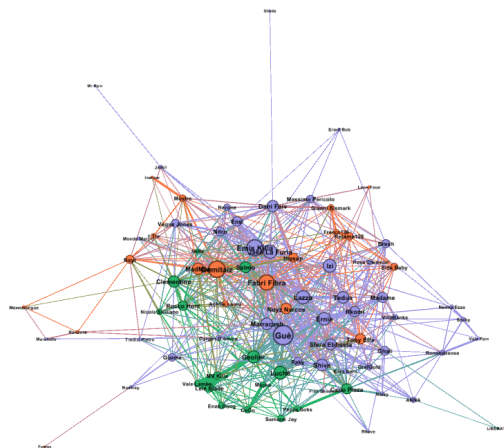


Figure 2: grafo con nodi colorati in base alla zona di formazione dell'artista

- grafo con nodi colorati in base alla decade di debutto dell'artista (figura 2).
 - in verde scuro il 1980 (unico nodo corrispondente a Neffa)
 - in azzurro il 1990
 - in verde il 2000
 - in viola il 2010
 - in arancione il 2020

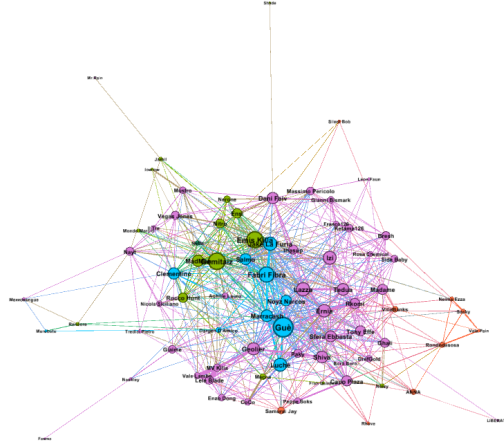


Figure 3: grafo con nodi colorati in base alla decade di debutto dell'artista.

3.3 Statistiche e commenti sui risultati grafici

Grazie all'adattamento in R, è stato possibile ricavare alcune informazioni importanti sulle caratteristiche del grafo. Il diametro del grafo risulta 5. Questo risultato indica il più lungo *shortest path* e conseguentemente afferma che per collegare un nodo ad un altro sono necessari al massimo 5 archi. Lo *shortest path* medio è pari a circa 2.02, il che indica che mediamente due artisti possono essere collegati tra di loro mediante 2 archi, per cui è possibile assumere ragionevolmente che il nostro grafo sia *small world*. Il *grado* medio è pari a 15.29 (ovvero mediamente un nodo è connesso con circa 15 altri nodi, il che sottolinea la massiccia presenza di collaborazioni tra artisti di questo genere), mentre il *grado* massimo è pari a 46 e corrisponde a Guè. Per quanto riguarda la *betweenness*, si ha un valore medio di 40.62 (ogni nodo è presente mediamente in circa 40 *shortest paths*) e un valore massimo di 394.52, sempre corrispondente a Guè. Si hanno quindi elementi sufficienti, grazie anche ai grafici riportati sopra, per affermare che Guè sia un *hub*.

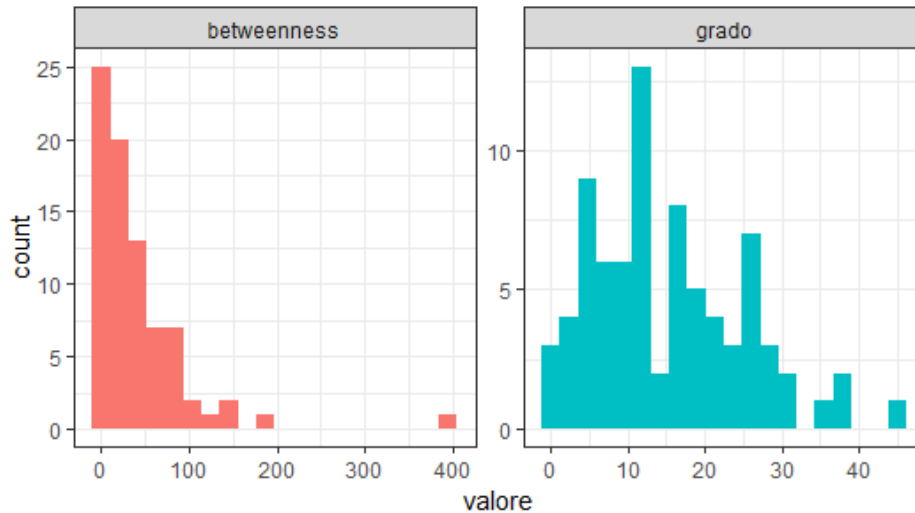


Figure 4: istogrammi per betweenness e grado

Relativamente alla suddivisione in decadi si sono ottenuti i seguenti risultati (figure 5 e 6):

- per il 1990, grado massimo 46 e betweenness massima 394.55 corrispondenti a Guè

- per il 2000, grado massimo 38 corrispondente a Gemitaiz e betweenness massima 186.91 corrispondente a Emis Killa
- per il 2010, grado massimo 30 corrispondente a Lazza e betweenness massima 137.59 corrispondente a Dani Faiv.
- per il 2020, grado massimo 29 corrispondente a Vale Pain e betweenness massima 54.41 corrispondente a Rhove.

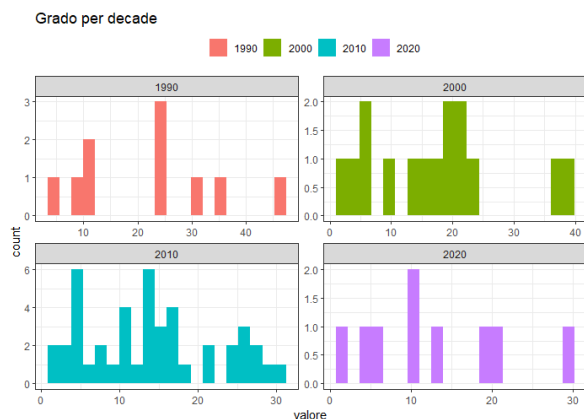


Figure 5: Istogrammi per il grado relativamente alla decade



Figure 6: Istogrammi per la betweenness relativamente alla decade

Calcoli equivalenti sono stati eseguiti per grado e betweenness in base alla zona (figure 7 e 8) e si è ottenuto:

- per il Nord, grado massimo 46 e betweenness massima 394.55 corrispondenti a Guè
- per il Centro, grado massimo 38 e betweenness massima 154.23 corrispondenti a Gemitaiz
- per il Sud, grado massimo 27 corrispondente a Geolier e betweenness massima 93.17 corrispondente a Nayt.

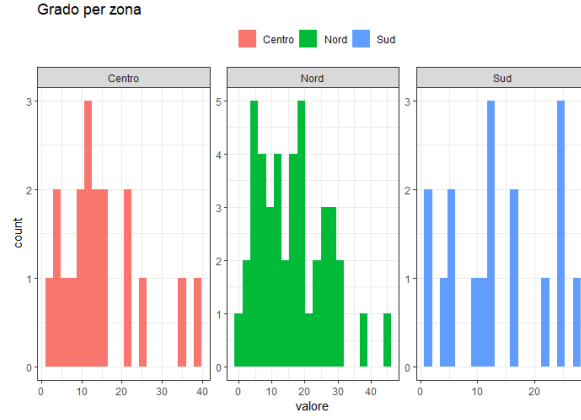


Figure 7: Istogrammi per il grado relativamente alla zona

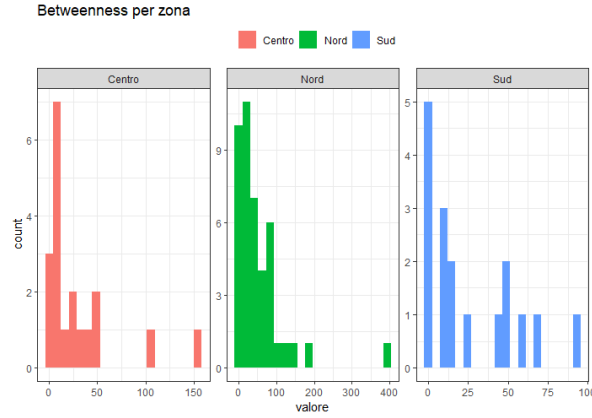


Figure 8: Istogrammi per la betweenness relativamente alla zona

4 Conclusioni

Dall'analisi effettuata sono emersi alcuni pattern notevoli nella distribuzione delle collaborazioni tra gli artisti considerati. Innanzitutto sembra, dai diagrammi del grafo, che gli artisti che hanno debuttato negli anni 90 siano i più centrali all'interno della rete, in netto contrasto rispetto a quelli esorditi negli anni 2020 (che tuttavia nel grafo sono molto vicini tra loro, segno che le "nuove leve" stanno collaborando molto); la più probabile causa di tale differenza è data dal fatto che indubbiamente gli artisti con più anni di carriera abbiano più notorietà e più possibilità di collaborazioni, oltre al fatto che essendo nella scena da più tempo è chiaro che abbiano avuto modo di fare più featuring. Tale conclusione è evidenziata anche dai valori di grado e betweenness, entrambi decrescenti con l'avanzare del tempo (per gli anni 2010 e 2020 la betweenness minima è addirittura pari a 0). Per quanto riguarda la zona di provenienza, la divisione sembra essere meno netta relativamente alla zona tra Centro e Nord, mentre è evidente una sorta di aggregazione per quanto riguarda gli artisti del Sud. La centralità di Guè inoltre, conferma la diffusa sensazione che sia il rapper più prolifico e presente nel rap italiano. Altro elemento convincente della rete è il fatto che artisti che hanno pubblicato dei joint-album insieme siano effettivamente molto vicini tra loro, come per esempio Marracash e Guè o come Gemitaiz e Madman, due delle coppie più consolidate della scena.