



TRAPSCAPE

UN'ANALISI LINGUISTICA E VISUALE COMPARATIVA DEI GENERI MUSICALI TRAP E RAP

A. MAISTO, A. PAONE, S. PELOSI, P. VITALE

UNIVERSITÀ DEGLI STUDI DI SALERNO

DIPARTIMENTO DI SCIENZE POLITICHE E DELLA COMUNICAZIONE

OBIETTIVI

- Costruzione di una risorsa terminologica e di strumenti interattivi di consultazione per l'analisi dell'universo linguistico e semantico di un fenomeno in crescita, in Italia, come la musica **Trap**
- Elaborazione, verifica e applicazione di una pipeline per l'estrazione, l'analisi e il trattamento degli **User Generated Content**
- Comparazione tra musica Trap e Rap
- Approccio ibrido:
 - Web Mining
 - Linguistica Computazionale
 - Data Visualization

LA MUSICA TRAP

- Rappresenta un fenomeno relativamente recente nel panorama musicale internazionale.
- Presenta diverse analogie con la musica Rap:
 - Caso di interesse in materia di strutture metriche
 - Aspetti morfosintattici e lessicali unici
- Una prospettiva comparativa tra questi due generi rappresenta una novità nel panorama della ricerca sociolinguistica

OUTLINE

- L'analisi dei testi si sviluppa attraverso la seguente pipeline:
 1. Estrazione automatica dei testi
 2. Preprocessing e processing automatico dei testi
 3. Vettorializzazione dei testi e calcolo della *similarity* tra canzoni
 4. Estrazione delle strutture metriche
 5. Emotion Analysis
 6. Taboo Language Analysis
 7. Calcolo della distanza sintattica media
 8. Data Visualization

ESTRAZIONE

Scraping:

- Sono state raccolte le prime 10 playlist italiane di musica trap da Spotify
- Per ciascuno degli autori sono stati raccolti gli ultimi 50 testi pubblicati su Genius.com

Refinement:

- Tutti i testi sono stati analizzati in automatico attraverso le API di Google al fine di rimuovere eventuali testi in inglese
- Dai file json relativi alle singole lyrics sono stati estratti testi e metadati funzionali all'analisi

Format:

- I dati sono stati organizzati in formati tabulari per analisi testuali e visualizzazioni ed edgelist per analisi di rete.

PREPROCESSING E PROCESSING DEI TESTI

LG-Starship
(Maisto, A. 2020)

POS Tagging e Lemmatizzazione
Estrazione del TF-IDF

TINT
(Palmero Aprosio
& Moretti, 2016)

Parsing Sintattico

VETTORIALIZZAZIONE E CALCOLO DELLA SIMILARITY

- Ogni canzone è stata trasformata in un vettore numerico sulla base dei punteggi di TF-IDF:
 - È stata costruita una matrice di canzoni per parole
 - Ogni canzone è associata ad una parola contenuta in essa tramite il punteggio di TF-IDF. Se la parola è assente, il valore corrispondente è zero.
 - I vettori sono stati confrontati tramite un algoritmo di *Cosine Similarity*

ESTRAZIONE DELLE STRUTTURE METRICHE

- Per l'estrazione delle Strutture Metriche (rime e assonanze) è stato sviluppato un algoritmo *ad-hoc*:
 - L'input dell'algoritmo è il testo suddiviso in righe
 - Per ogni riga vengono estratti gli ultimi 4 caratteri
 - Caratteri accentati e suoni assimilabili sono stati unificati
 - Le consonanti sono state tutte uniformate
 - Se almeno gli ultimi 3 caratteri coincidono, alle due righe viene attribuito lo stesso codice (le righe sono state identificate con lettere maiuscole dalla A alla E)

ESTRAZIONE DELLE STRUTTURE METRICHE

- Il risultato è un file in cui la canzone è sostituita da una serie di codici:
 - ABCCADBE CCACCCBBB AAACAAAC BBDBACE DDDDDADBCEAD

ESTRAZIONE DELLE STRUTTURE METRICHE

Chiudi quella porta se non vuoi casino	B
Chiudi quella porta che qua dietro c'è un casinò	B
La gente fa avanti e indietro come fosse alla stazione	C
Beh, fatti un giro alla stazione	C
Purifico la mia anima, un bagno dentro al Gange	A
La mia gente ha gli occhi rossi	D
Come se ha smesso di piangere	B
Il latte versato come il sangue di Dio	E
Tovagliolo, chiazze d'olio, tiro su quel petrolio	C
Questi rapper parlano di strada come avessero un portfolio	C
Chiudi quella porta, già sento odore di vomito	A
Non me ne frega un cazzo, voglio un grosso portafoglio	C
Per comprarmi il tuo orgoglio e il tuo orologio	C
Tony ti stacca la testa e ci fa l'orologio	C
Sto con mio fratello Side e fumo tutto il giorno	B
E chiudi quella porta se non vuoi tutto il branco attorno	B
Tutto il brano attorno	B

ESTRAZIONE DELLE STRUTTURE METRICHE

- In seguito è stata calcolata la presenza di determinate sequenze:
 - Rima Baciata (AA)
 - Rima Alternata (ABAB)
 - Rima Incrociata (ABBA)
 - Rima Ripetuta (ABC ABC)
 - Rima Invertita (ABC CBA)

- Sono state escluse le rime Incatenate (ABA BCB CDC)

EMOTION DETECTION

Risorsa lessicale applicate: NRC Affect Intensity Lexicon

- <https://saifmohammad.com/WebPages/AffectIntensity.htm>
- Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3), 436-465.
- Mohammad, S. M. (2017). Word affect intensities. *arXiv preprint arXiv:1704.08798*.

EMOTION DETECTION

- Emozioni considerate: Anger, Fear, Anticipation, Trust, Surprise, Sadness, Joy, Disgust
- Entrate Totali: 20.255
 - Aggettivi: 2.932
 - Nomi: 16.117
 - Avverbi: 133
 - Verbi: 1.018
- 9.921 EN
- 10.334 ITA
- Modifiche apportate alla risorsa:
 - lemmatizzazione e flessione automatica
 - integrazione con le risorse del modulo italiano di Nooj

TABOO LANGUAGE ANALYSIS

Risorse Lessicale applicata:

- Maisto, A., Pelosi, S., Vietri, S., Vitale, P., & Paolo II, V. G. (2017). Mining offensive language on social media. *CLiC-it 2017 11-12 December 2017, Rome*, 252.

	Trap	Rap
Matches	336	239
Lemmi	49	57
Ratio	6,9	4,1

CALCOLO DELLA DISTANZA SINTATTICA MEDIA

- L'idea è quella di calcolare i livelli di profondità generati durante il Parsing da una frase:
 - Partendo dal livello 0 (ROOT) si calcola a che distanza sintattica (Liu et al., 2017) si trova l'elemento più lontano dal ROOT
 - Una volta calcolato il numero di livelli presenti nella frase, il punteggio viene diviso per il numero totale di elementi, in modo da avere un indice compreso tra 0 (solo il ROOT) a 1 (ogni parola corrisponde ad un livello)

CALCOLO DELLA DISTANZA SINTATTICA MEDIA

- L'indice totale del testo si ottiene dalla media delle distanze sintattiche medie dei singoli enunciati contenuti nel testo.
- Nella tabella l'indice di dipendenza di alcuni testi:

Titolo/argomento del testo	Punteggio
DEF 2020 – Sezione I	0,39
Ulisse (Joyce)	0,38
Il pendolo di Foucault (Eco)	0,37
Marcovaldo (Calvino)	0,36
Il Piccolo Principe (de Saint-Exupéry)	0,35
Articolo divulgativo (tema World Beer Awards)	0,34
Lo Hobbit (Tolkien)	0,34

CALCOLO DELLA DISTANZA SINTATTICA MEDIA

TRAP	RAP
0,35	0,37

Autori (primi 7)	Punteggio
Marracash (RAP)	0,4059
Dorian Nox (TRAP)	0,3968
Rocco Hunt (RAP)	0,389
Filippo Ruggieri (RAP)	0,3865
Edo Fendy (RAP)	0,3814
Dark Polo Gang (TRAP)	0,3813
Machete (TRAP)	0,3804

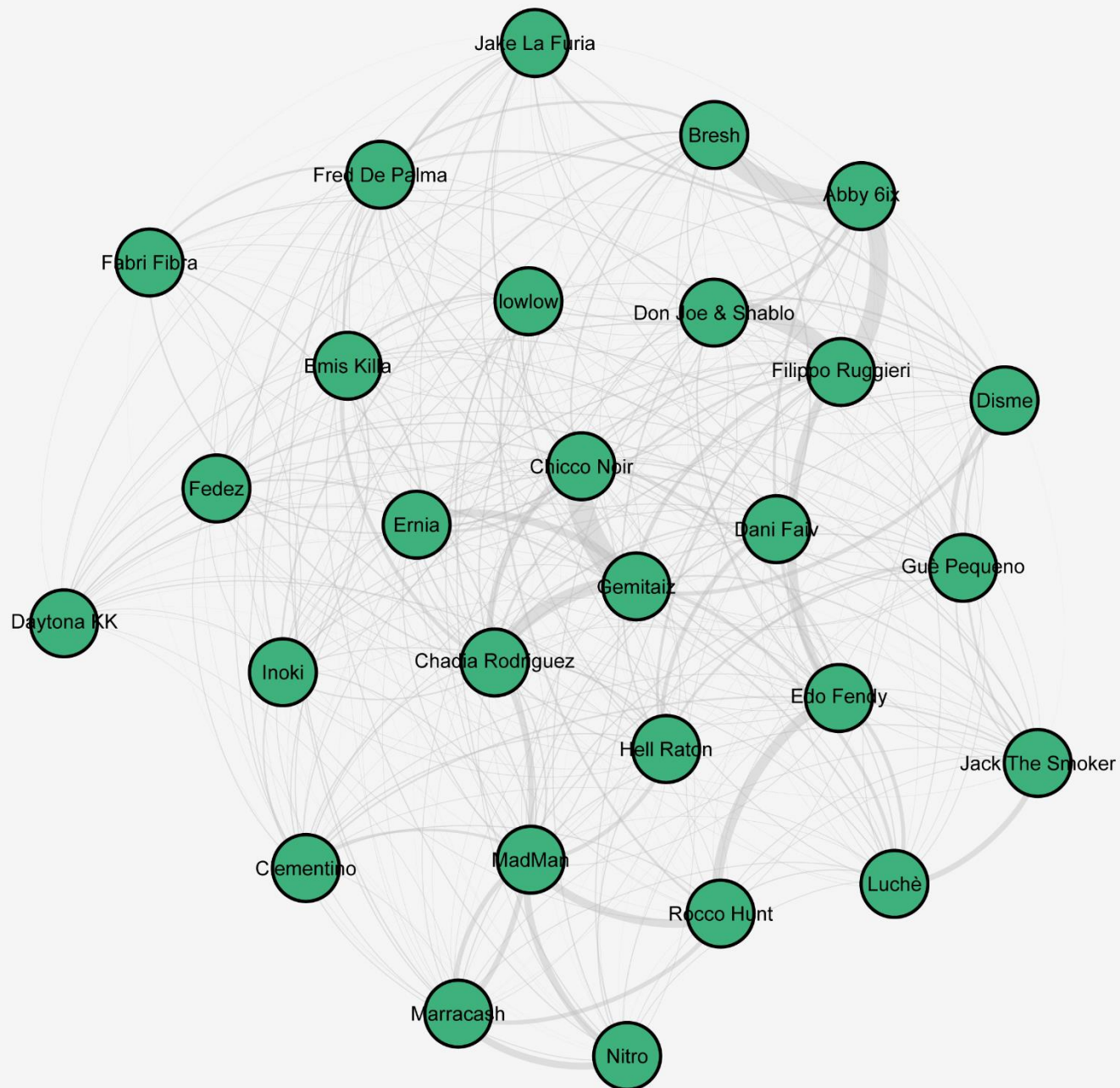
Autori (ultimi 7)	Punteggio
Low Kidd (TRAP)	0,3068
Lortex (TRAP)	0,3084
Chiello (TRAP)	0,3143
DrefGold (TRAP)	0,3312
Lele Blade (TRAP)	0,3334
CapoPlaza (TRAP)	0,3354
Fabri Fibra (RAP)	0,3361

DATA VISUALIZATION

- I dati generati dall'applicazione delle differenti tipologie di analisi sono stati organizzati in tabelle e Network per ottenere differenti visualizzazioni
- Lo scopo finale è quello di offrire ad appassionati o studiosi una panoramica visuale del fenomeno Trap, anche in comparazione con il Rap

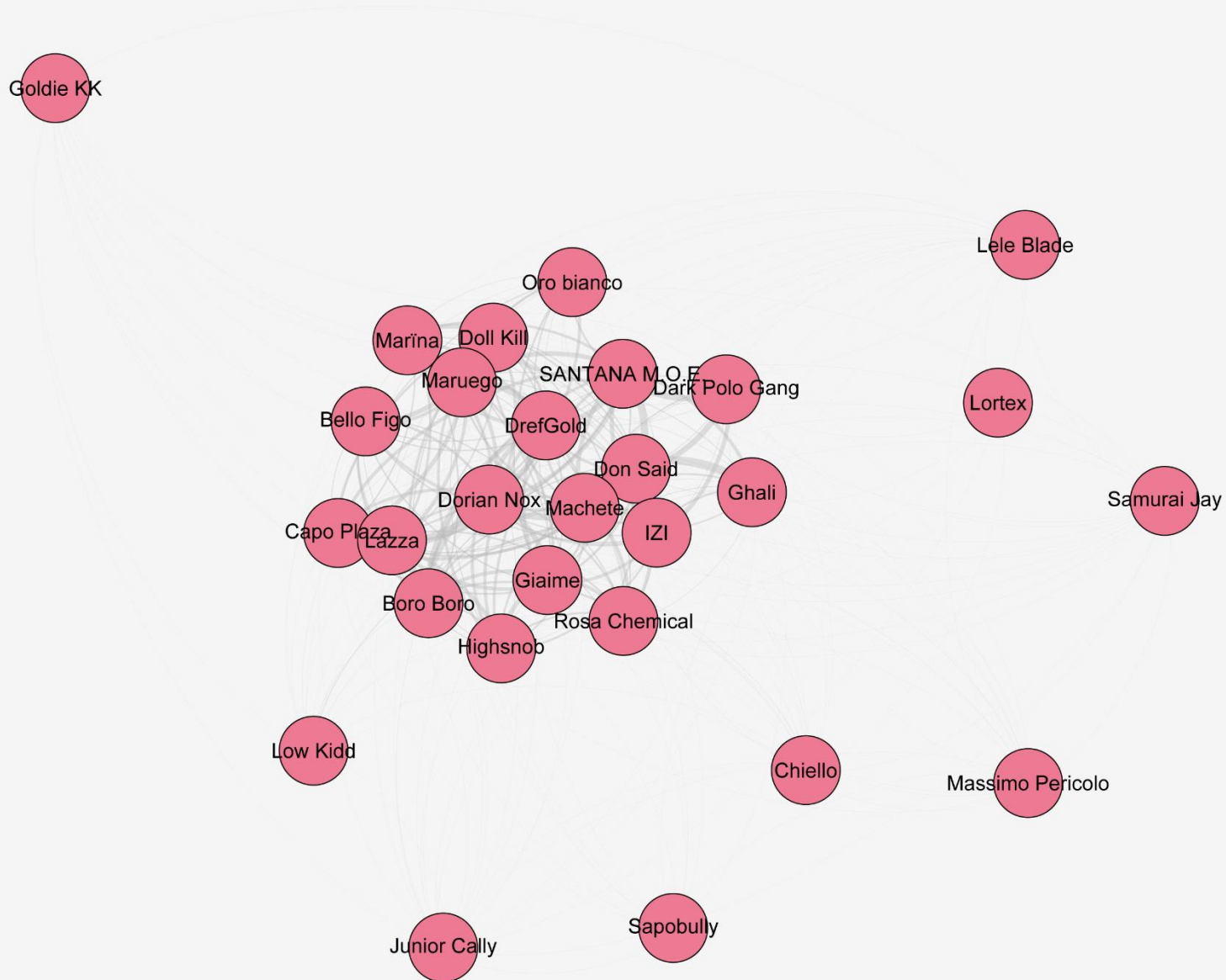
Network Visualization

Rap Similarity



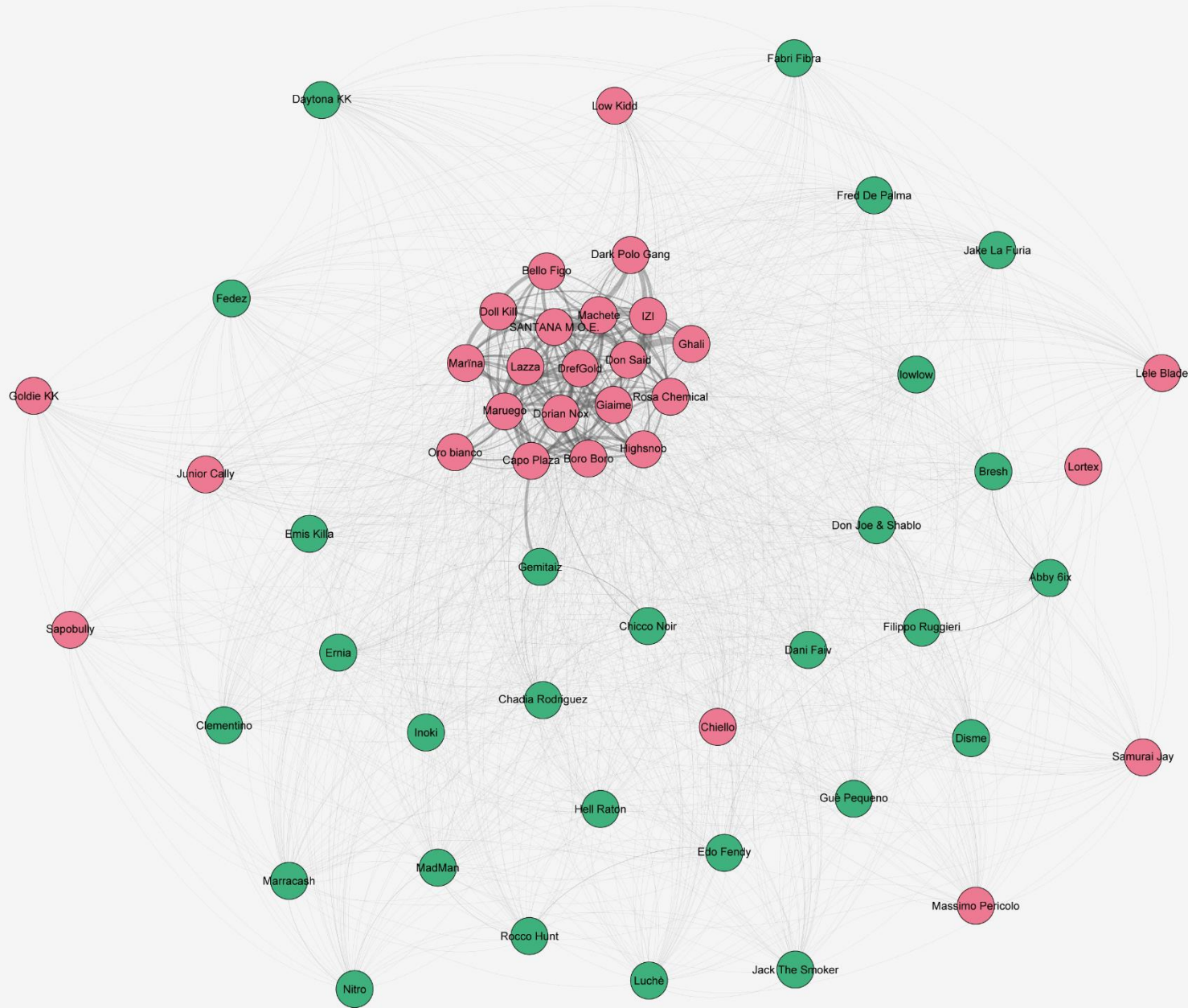
Network Visualization

Trap Similarity



Network Visualization

Rap & Trap Similarity



DATA VISUALIZATION

Tutti i dati analizzati sono stati sistematizzati in una dashboard interattiva che consente di esplorare il fenomeno da più prospettive, sia a livello micro, osservando i risultati a livello di singoli artisti, sia a livello macro, definendo dei bilanci tra i due generi in termini di emozioni, parole sconosciute e complessità delle strutture metriche.

[Link alla Dashboard](#)

CONCLUSIONI

- In un prossimo futuro il concetto di *distanza sintattica media* sarà arricchito e confluirà, assieme ad altre metriche, in un valore unico di comprensibilità del testo
- È, inoltre, previsto lo sviluppo di dizionari elettronici dello *slang*, a partire dalla rielaborazione degli *item* lessicali non presenti nelle risorse linguistiche
- La piattaforma, che è al momento solo in una fase sperimentale, sarà arricchita con estrazione e l'analisi di nuovi testi in relazione ai due generi selezionati
- La pipeline può essere sperimentata e verificata anche su altri generi musicali



UNIVERSITÀ DEGLI STUDI
DI SALERNO



DIPARTIMENTO DI
SCIENZE **POLITICHE** E
DELLA **COMUNICAZIONE**

amaisto@unisa.it

spelosi@unisa.it

pvitale@unisa.it

Antonietta.paone@outlook.it