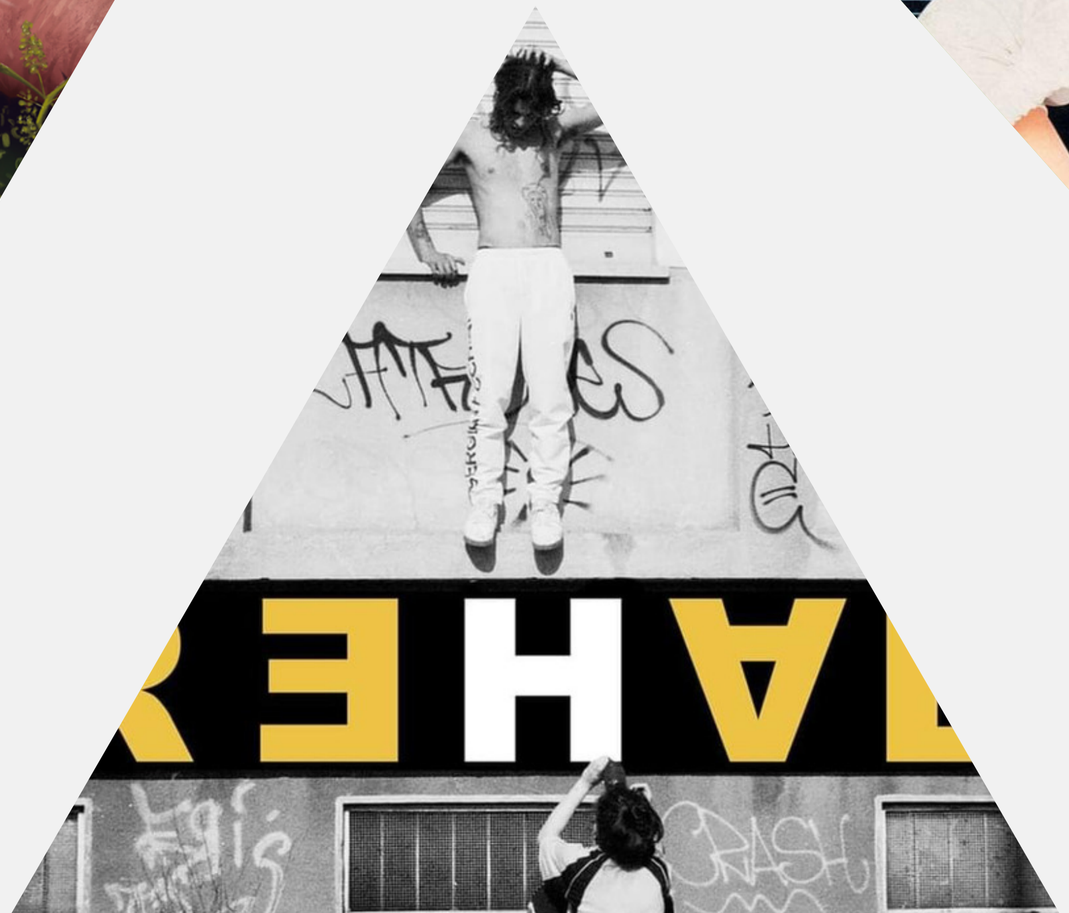
The inner workings of the italian music scene



The idea:

- Build a network of collaborations between artists of different genres, capable of representing the current state of the italian music scene, which can be easily decomposed to analyze single genres

The questions:

- Is there any difference in the behaviours of artists of different genres? Any significant commistion between genres?
- Does the network show any assortative behaviour like most social networks? If so, is it structural? What about the single genres?

Step 1:

Scraping the scrapeable

Scraping

- I tried getting data that could be intended as a heterogeneous snaphot of the italian music scene as of today:

Get first 1k artists in Italy
for 8 different genres

- Rap
- Hip-hop
- Indie
- Indie-pop
- Pop
- Emo
- Punk
- Punk-rock

Artists' attributes:

- id
- Name
- Total followers
- Popularity
- Genre (metadata)

Scraping

Episode two

First:

- Filter italian artists

(Check if in the "genres" metadata there is at least one between "ita-", "roma-", "neap-", "naple-", "milan-")

- Remove duplicates

Then:

Identifying feats

- Select each artist
- Find on which albums and singles they appear on
- Store name of the artist on which album/single they appear on

Final preprocessing steps

Aggregation of genres:

- Rap / Hip hop ⟶ Rap
- Indie / Indie-pop ⟶ Indie
- Emo / Punk / Punk-rock ⟶ Alt

(Pop didn't need it)

Change missclassified artists:

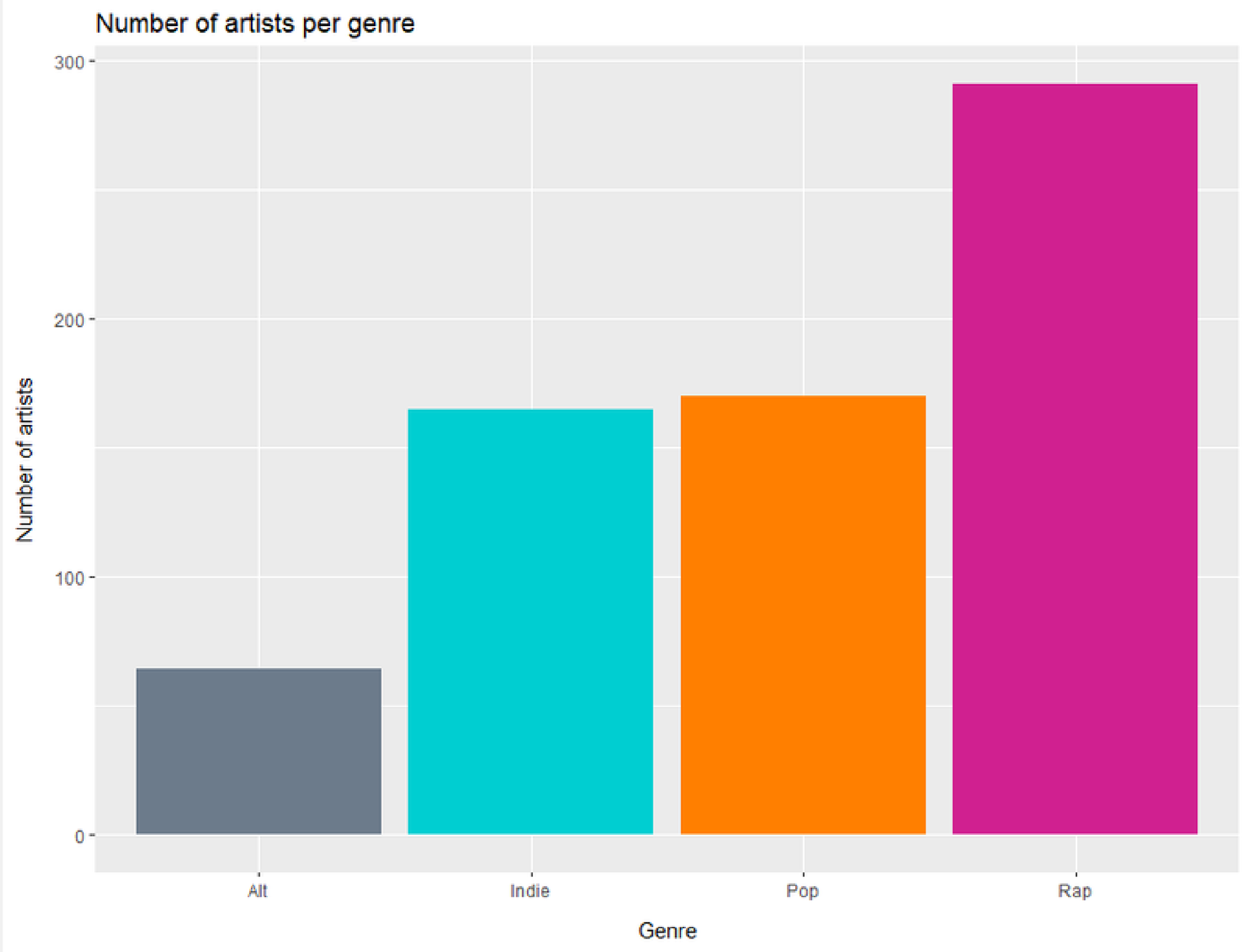
- Some artists are present in more than one genre
- When removing duplicates, some of them ended up in a less pertinent category
- 21 artists missclassified, most of them rappers in the pop category

Step 2:

The network, finally

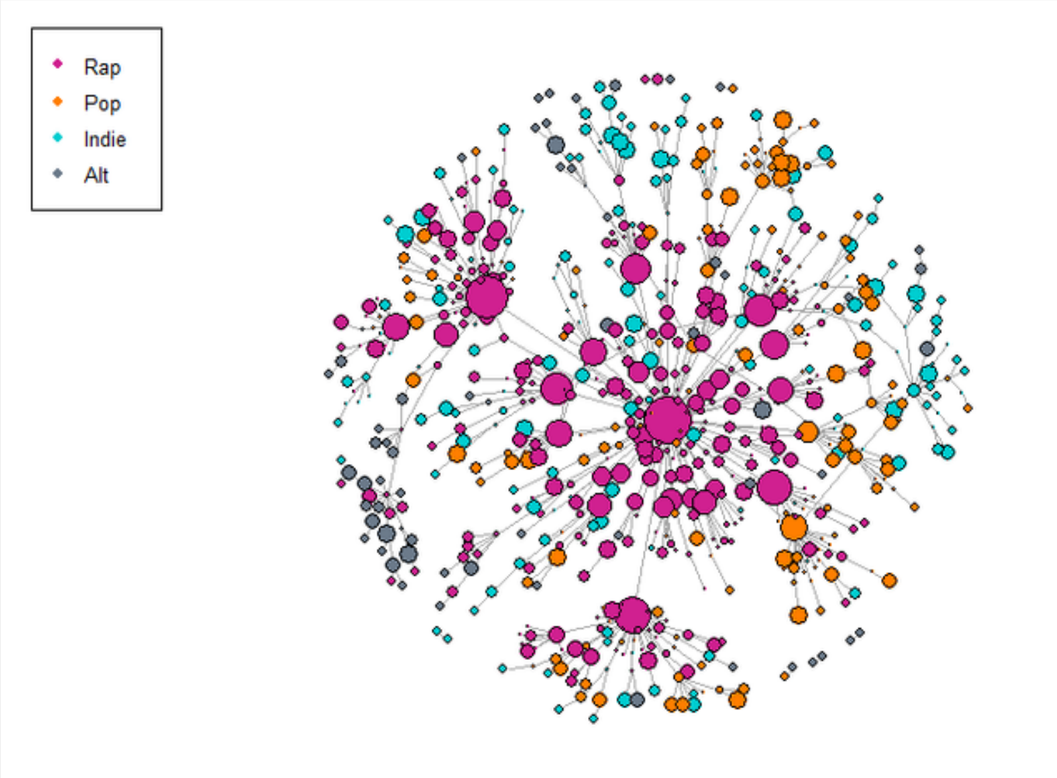
Artists' collaboration network:

- 690 nodes
- 3297 edges
- No multi-edges
- <k> = 9.56

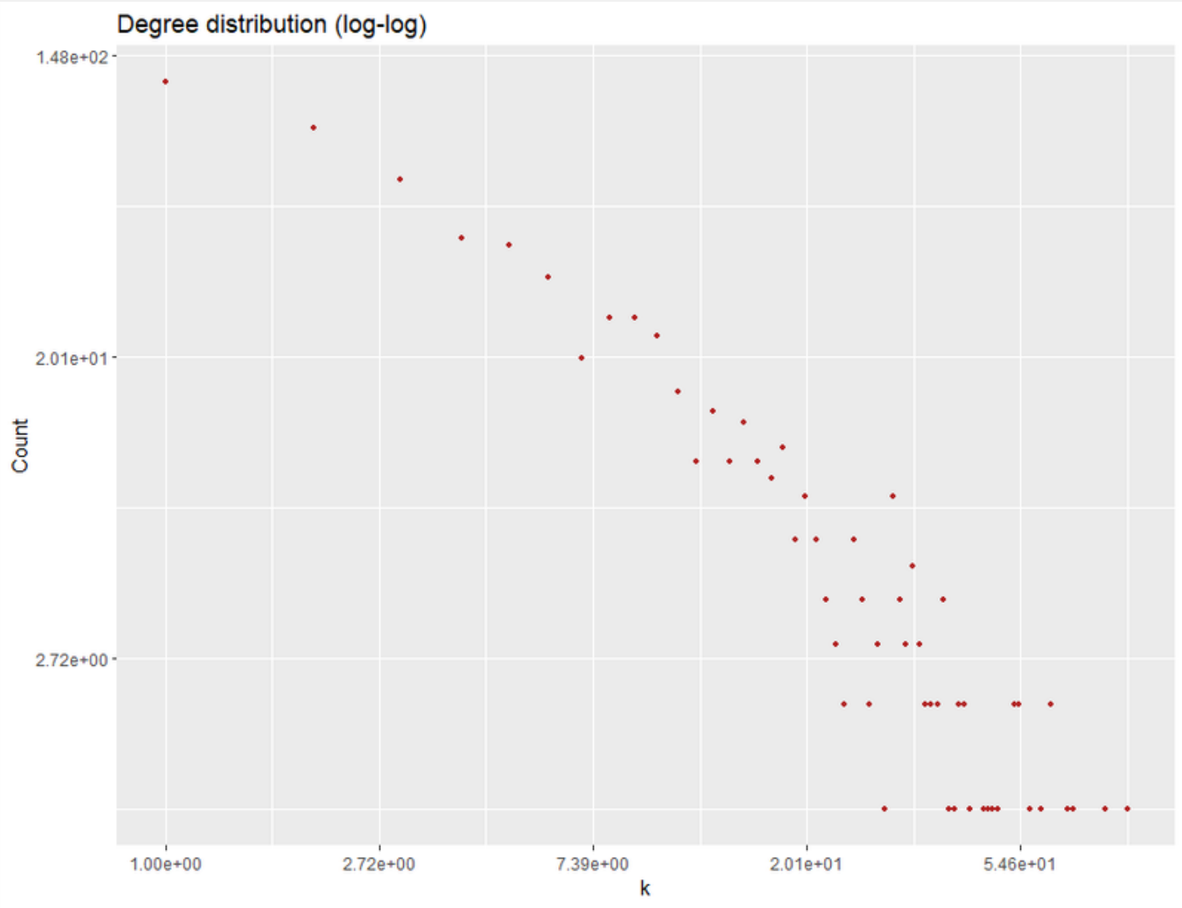


Artists' collaboration network:

- Network visualization



- Degree distribution

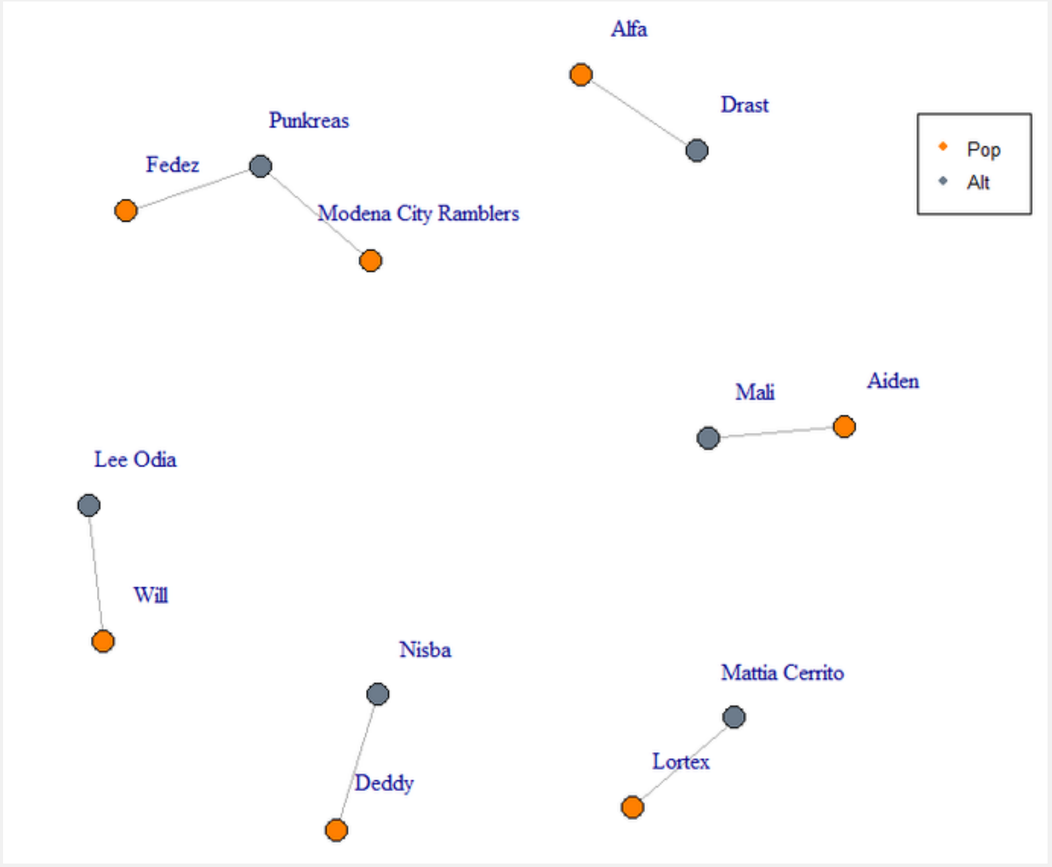


Are artists from different genres collaborating?

Contact layer of the mixing matrix:

	Alt	Indie	Pop	Rap
Alt	26	24	7	58
Indie	0	196	128	327
Pop	0	0	371	363
Rap	0	0	0	1797

- 55% of the network's edges are between rap artists
- Rap artists "appear" in 77% of the edges



Other segregation metrics:

Global segregation:

E-I index:

- Normalized difference between between-group ties and within group ties
- E-I index value = -0.45

Freeman's segregation index:

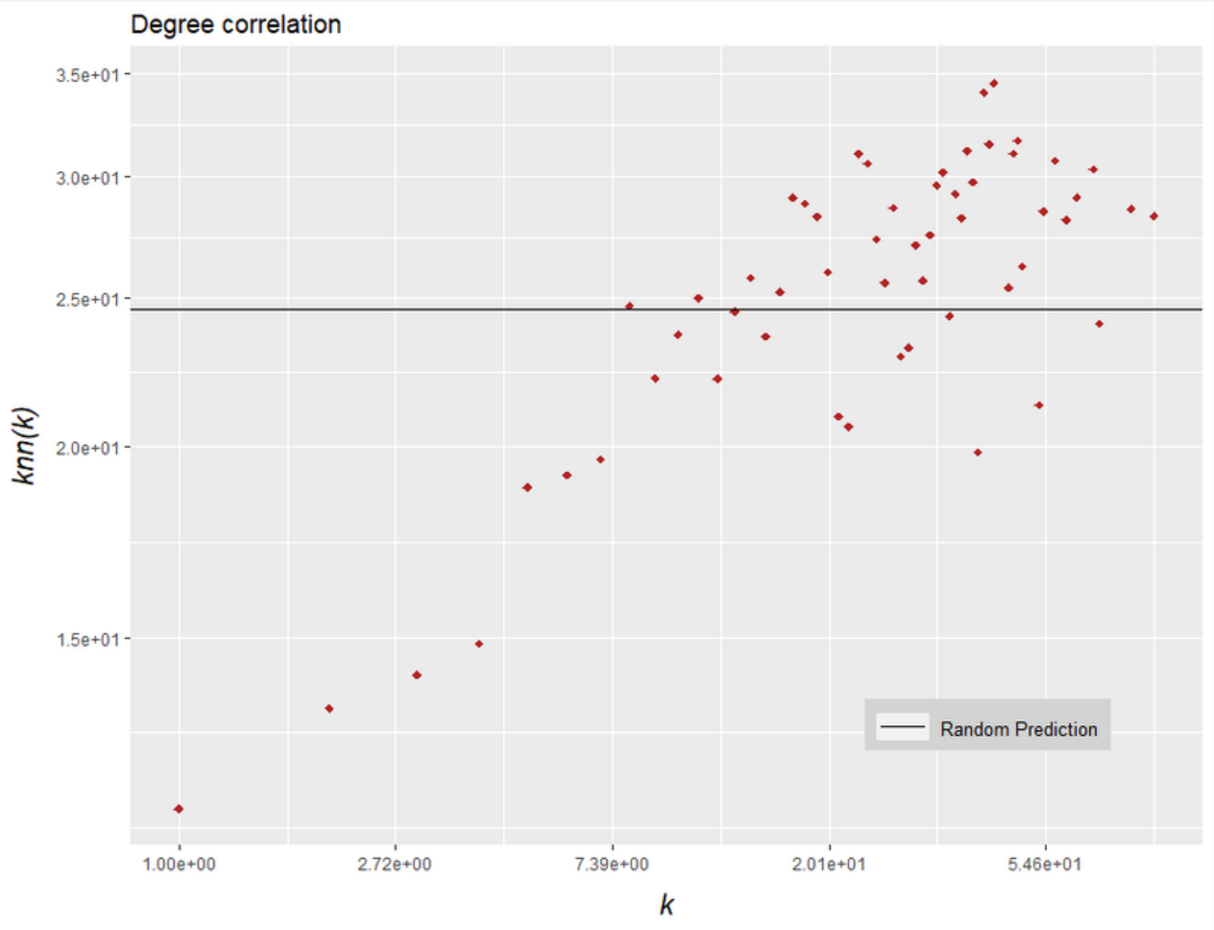
- Generalized version which allows the presence of more than two groups
- Freeman's index value = 0.6

Segregation between specific genres:

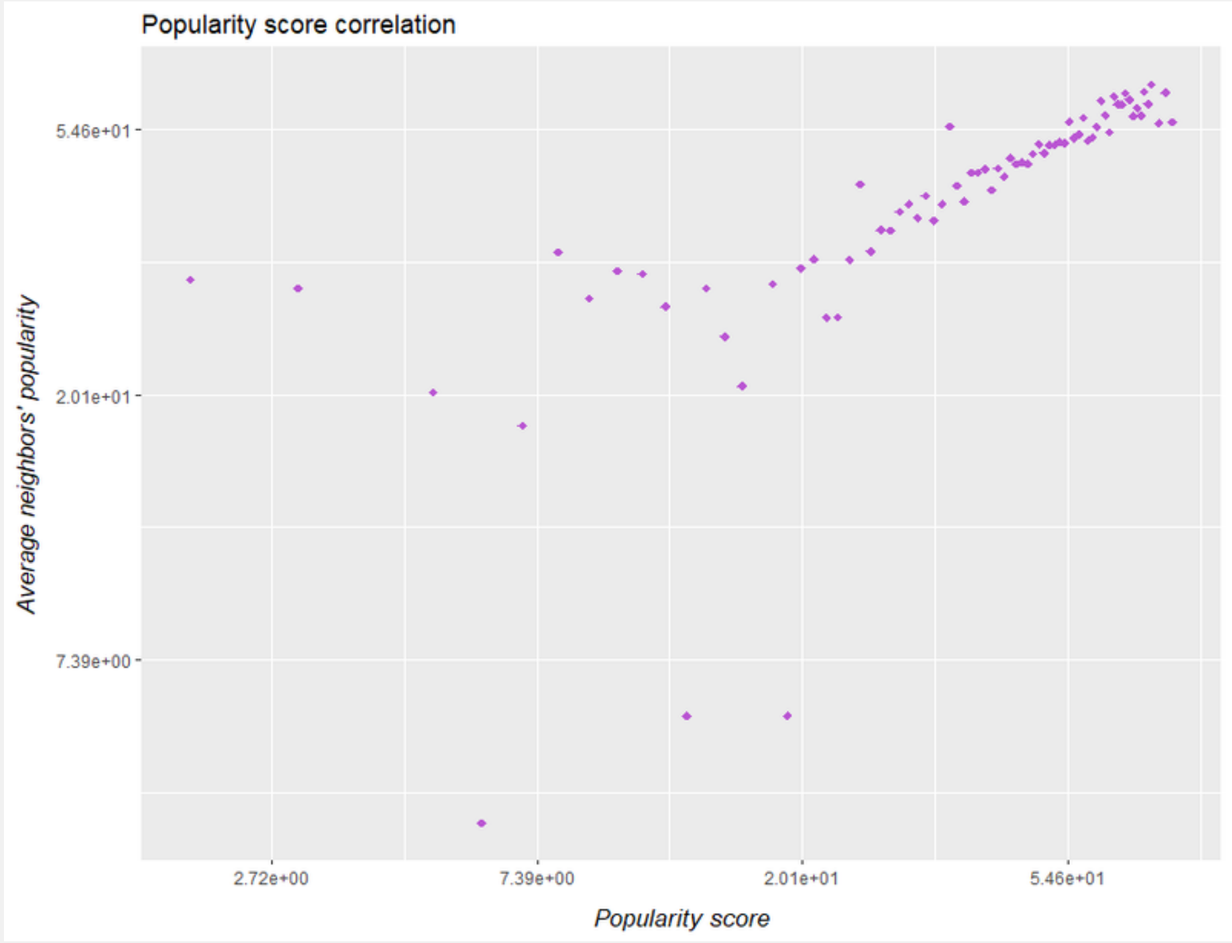
	Rap	Pop	Indie	Alt
Rap	0	0.69	0.7	0.9
Pop	0.69	0	0.63	0.96
Indie	0.7	0.63	0	0.76
Alt	0.9	0.96	0.76	0

Assortativity:

- Assortativity (degree): 0.16



Assortativity (popularity): 0.44



Are hubs not necessarily the most popular nodes?

- Let's consider as hubs the nodes with a degree higher than the 95th percentile of the degree distribution
- 36 hubs with k > 34
- 95th percentile of the popularity scores distribution: 66
- Hubs average popularity score: 60
- 60% percent of the hubs have a popularity score lower than 66
- Furthermore, only 36% of the hubs would also be hubs by followers count

Then, who are the hubs?

- Either very famous and relevant artists...
- ...or producers (which of course get a lot of features) and rap artists with very long careers, but which are not as relevant today

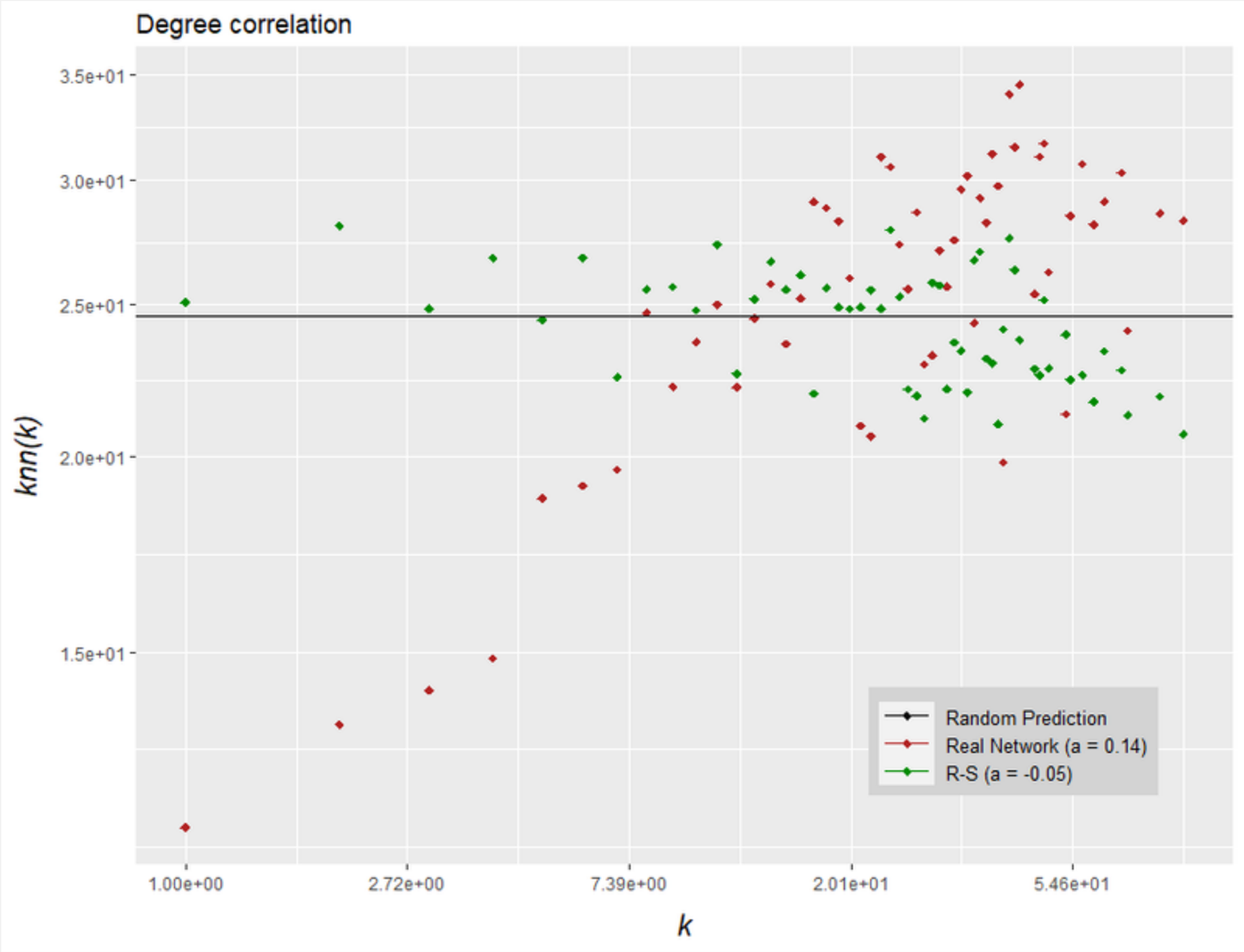
Is the assortativity structural?

We can address this by rewiring the network while preserving the degree distribution

- Randomization with Simple Links (R-S)
- Randomization with Multiple Links (R-M)

RESULTS:

- Assortativity (R-S) = -0.06
- Assortativity (R-M) = -0.05



Assortativity in genre sub-networks:

	Degree	Degree (R-S)	Degree (R-M)	Popularity
Rap	0.08	-0.09	-0.07	0.41
Pop	-0.14	-0.1	-0.07	0.16
Indie	0.01	-0.04	-0.02	0.3
Alt	-0.16	-0.14	-0.06	0.43

There are some differences from the original network:

- While all the sub-networks are assortative by popularity...
- ...the indie network is neutral by degree
- The pop and alt networks are disassortative by degree
- They also seem to be structurally disassortative, but the alt's network very low number of edges makes making assumptions difficult

Thank you and keep listening to the music you enjoy :)