Lyricognizer

Outil d'attribution de paroles de musique à leur auteur

Aymane Hamdaoui & Charles Kayssieh

9 février 2024

Télécom Paris

Table des matières

Introduction

Introduction

Problématique

Est-il possible de créer un outil qui attribue à des paroles de chansons leur auteur uniquement à l'aide d'une base de données de lyrics de chanteurs choisis préalablement et de la *Normalized compression distance* (NCD) ?

Fonctionnement

Fonctionnement général

Fonctionnement général du système :

1. Collecte des données
2. Traitement du texte
3. Tri de la BDD
4. Calcul de la NCD
5. Renvoi des résultats

Collecte des données

Création d'un script *lyrics_grabber.py*

Utilisation de l'API Genius

Récupération des paroles des 100 musiques les plus populaires de chaque artiste

Artistes

- Charles Aznavour
- B.B. Jacques
- Damso
- Drake
- Freeze Corleone
- Gazo
- Lomepal
- Mylène Farmer
- Nekfeu
- Soolking

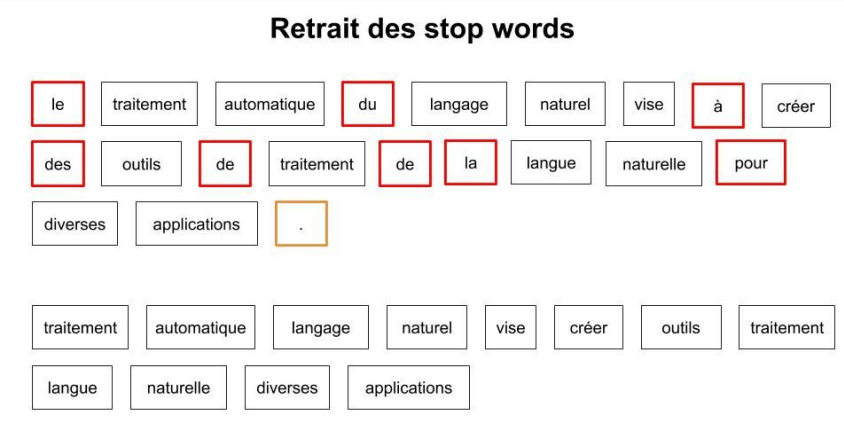
Traitement du texte

Trois niveaux de traitement avec le script *cleaner.py* :

1. Suppression de la ponctuation, des sauts de ligne, des annotations et passage en minuscules
2. Suppression des mots vides
3. Lemmatisation

} Détection de la langue

Mots vides et lemmatisation



Fonctionnement des stop words

Fonctionnement de la lemmatisation

Tri de la BDD

Organisation de la BDD :

1. Par artiste, chaque musique est aléatoirement numerotée entre 0 et 99
    - 0-19 = utilisation pour les données test
    - 20-99 = utilisation pour les données d'entraînement
2. Traitement des 1000 musiques des 3 manières différentes

Calcul de la NCD

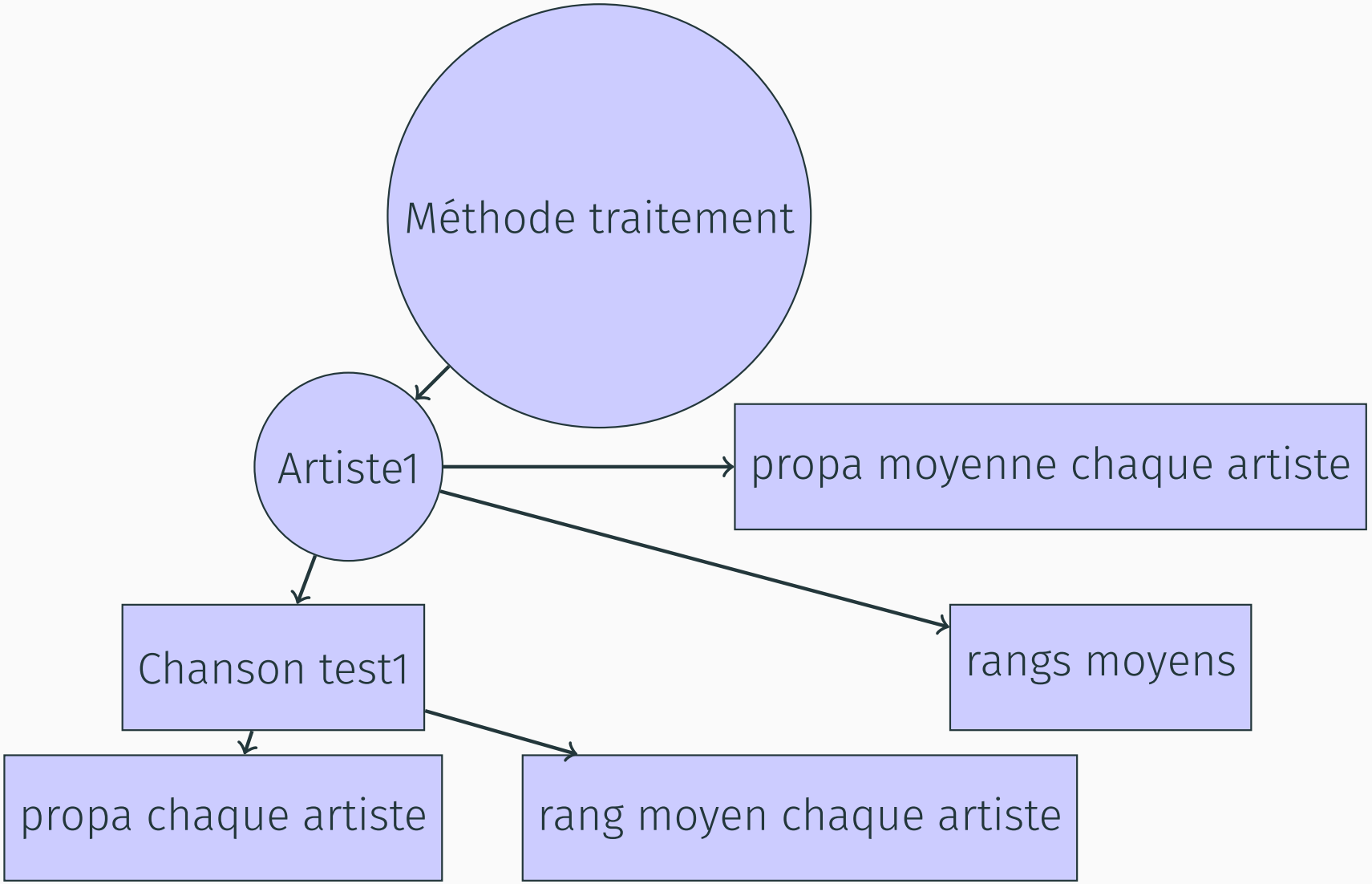
Formule théorique :

$$\mathrm{NCD}(x, y) = \frac{C(x + y) - \min(C(x), C(y))}{\max(C(x), C(y))}$$

Utilisation de 3 méthodes de compression

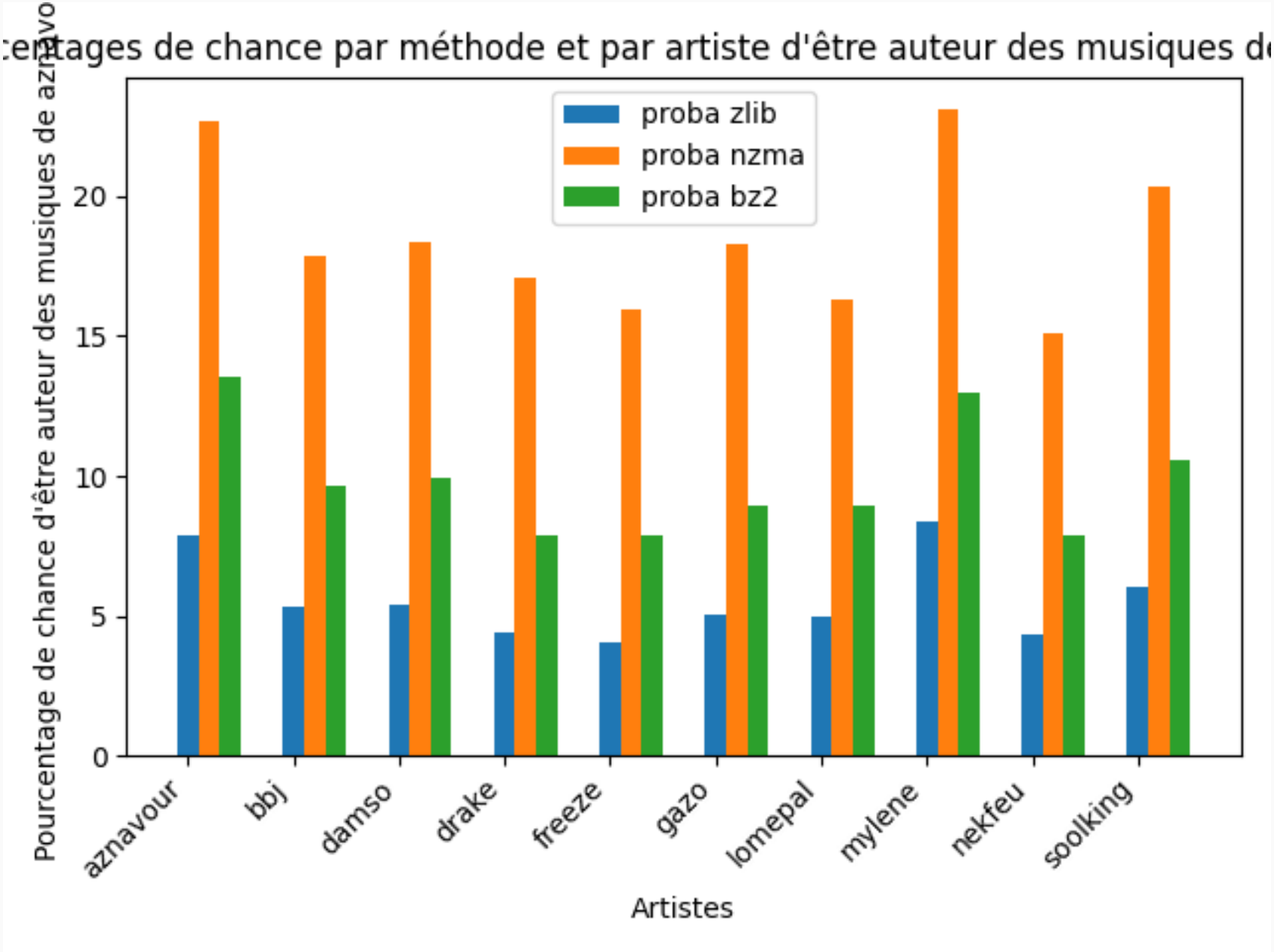
- zlib : la plus populaire (utilisée par Linux, macOS, iOS, ...) - compression par dictionnaire & encodage de Huffman
- lzma : utilisée par le programme 7zip - compression par dictionnaire
- bz2 : la plus efficace - réorganisation de donnée & encodage de Huffman

Structure des résultats

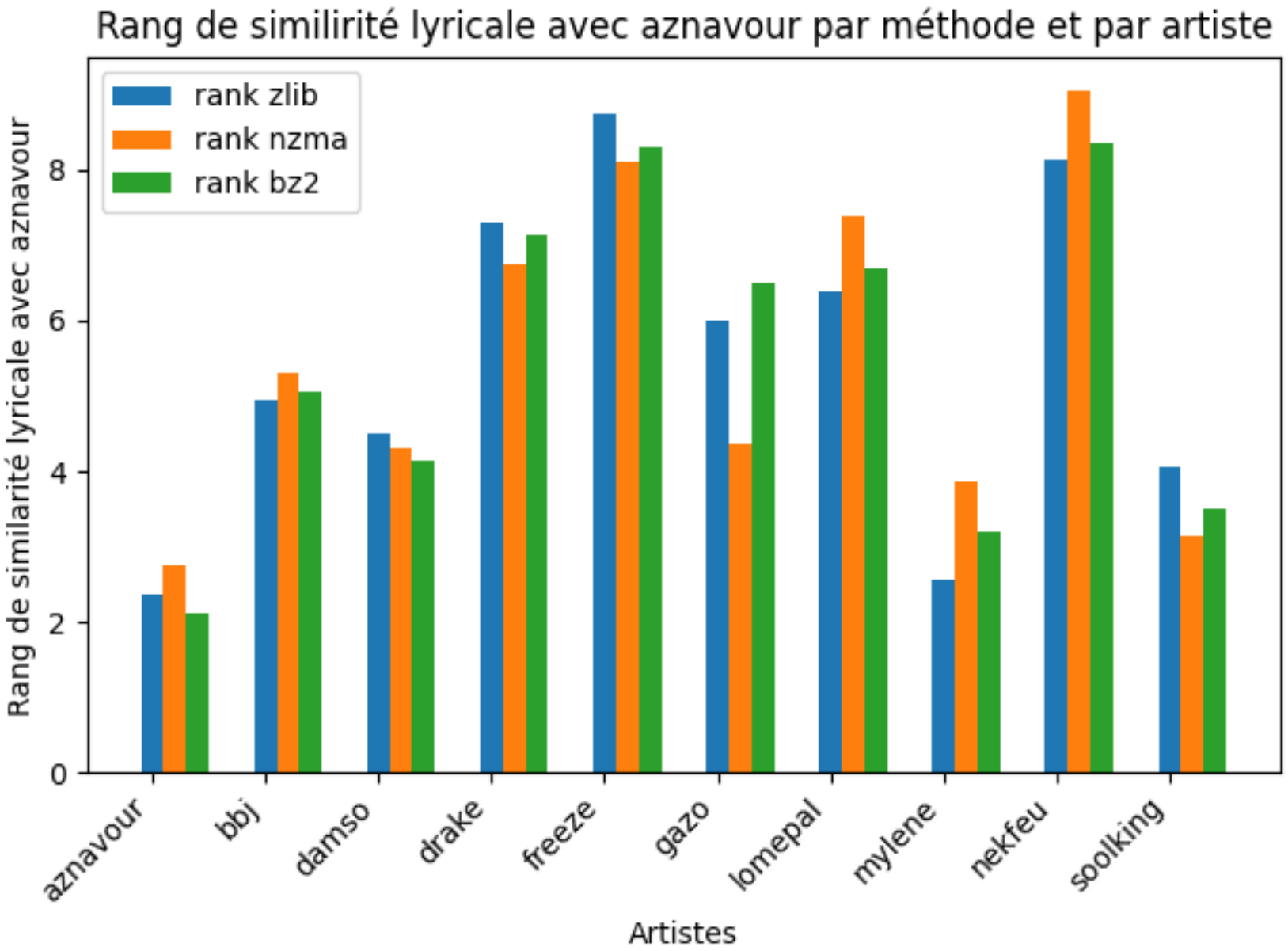


Résultats

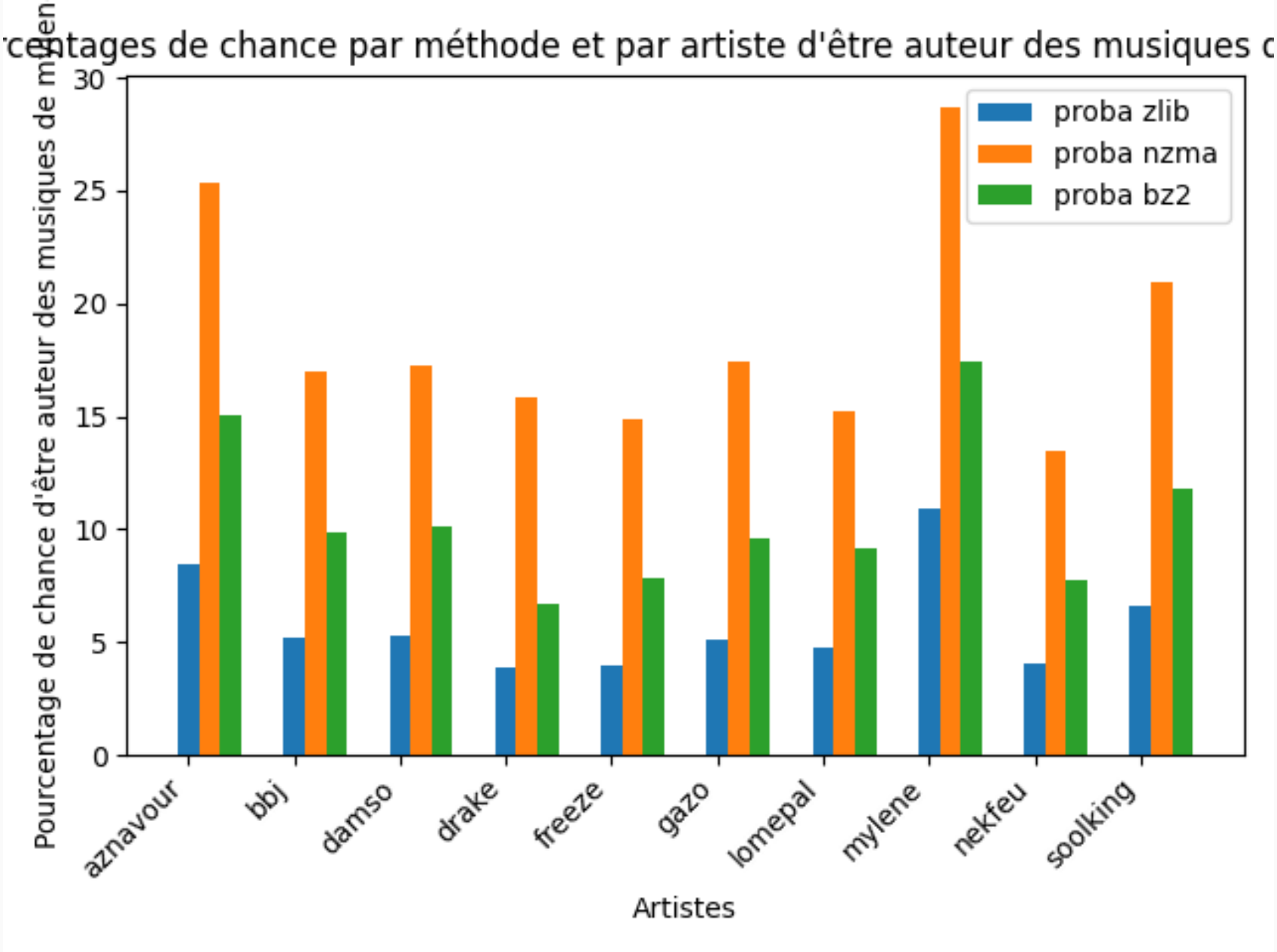
Charles Aznavour



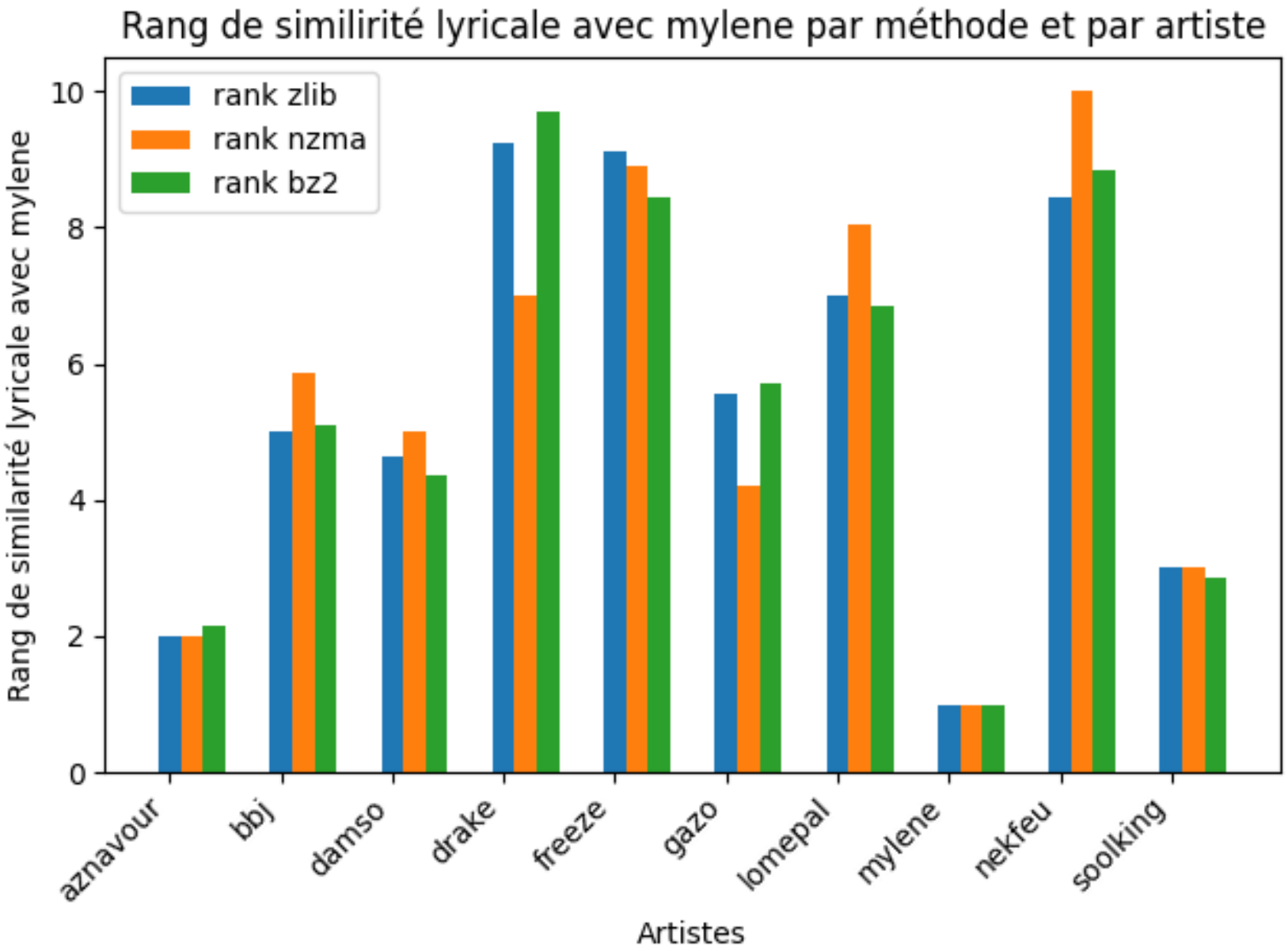
Charles Aznavour



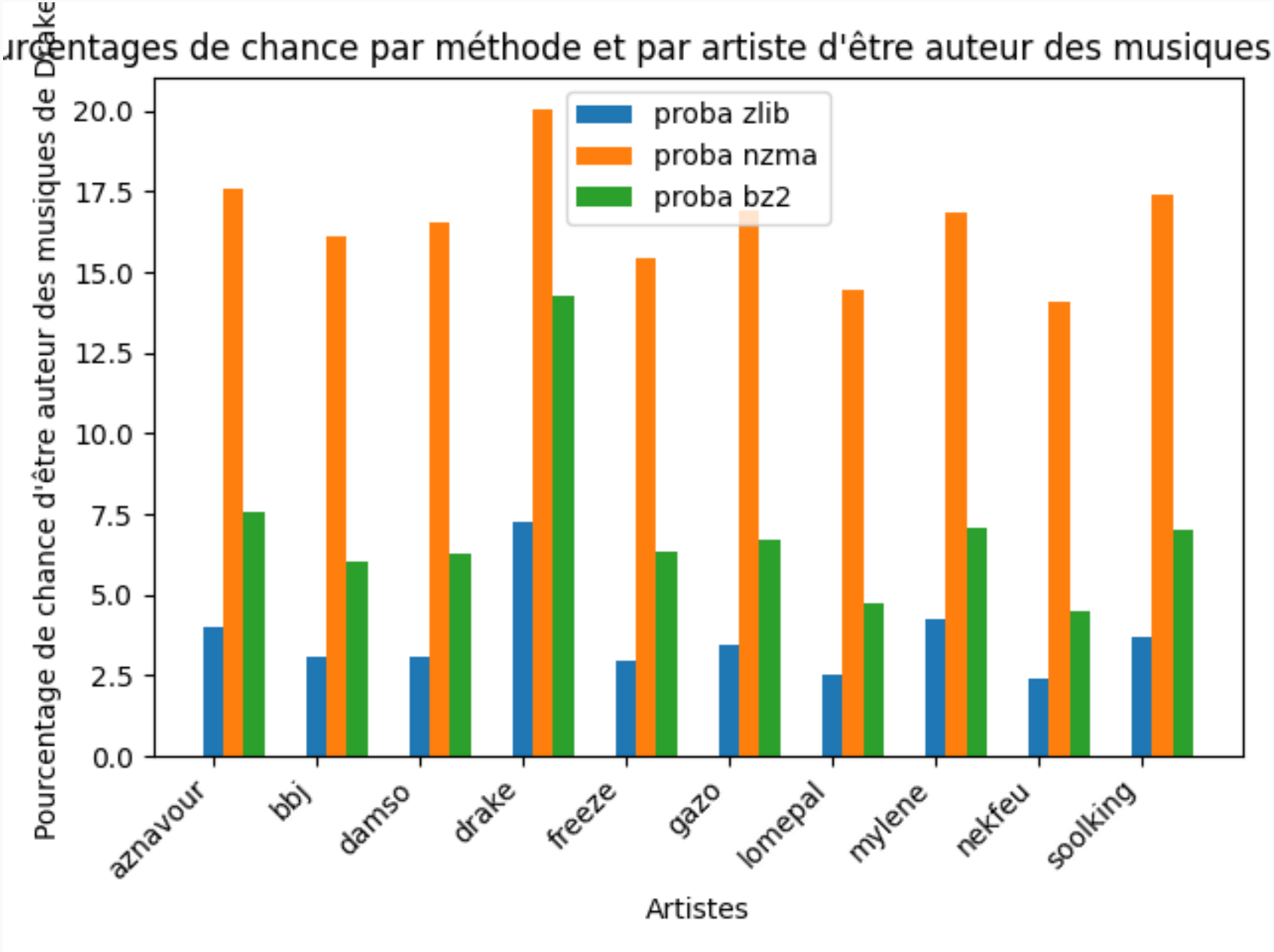
Mylène Farmer



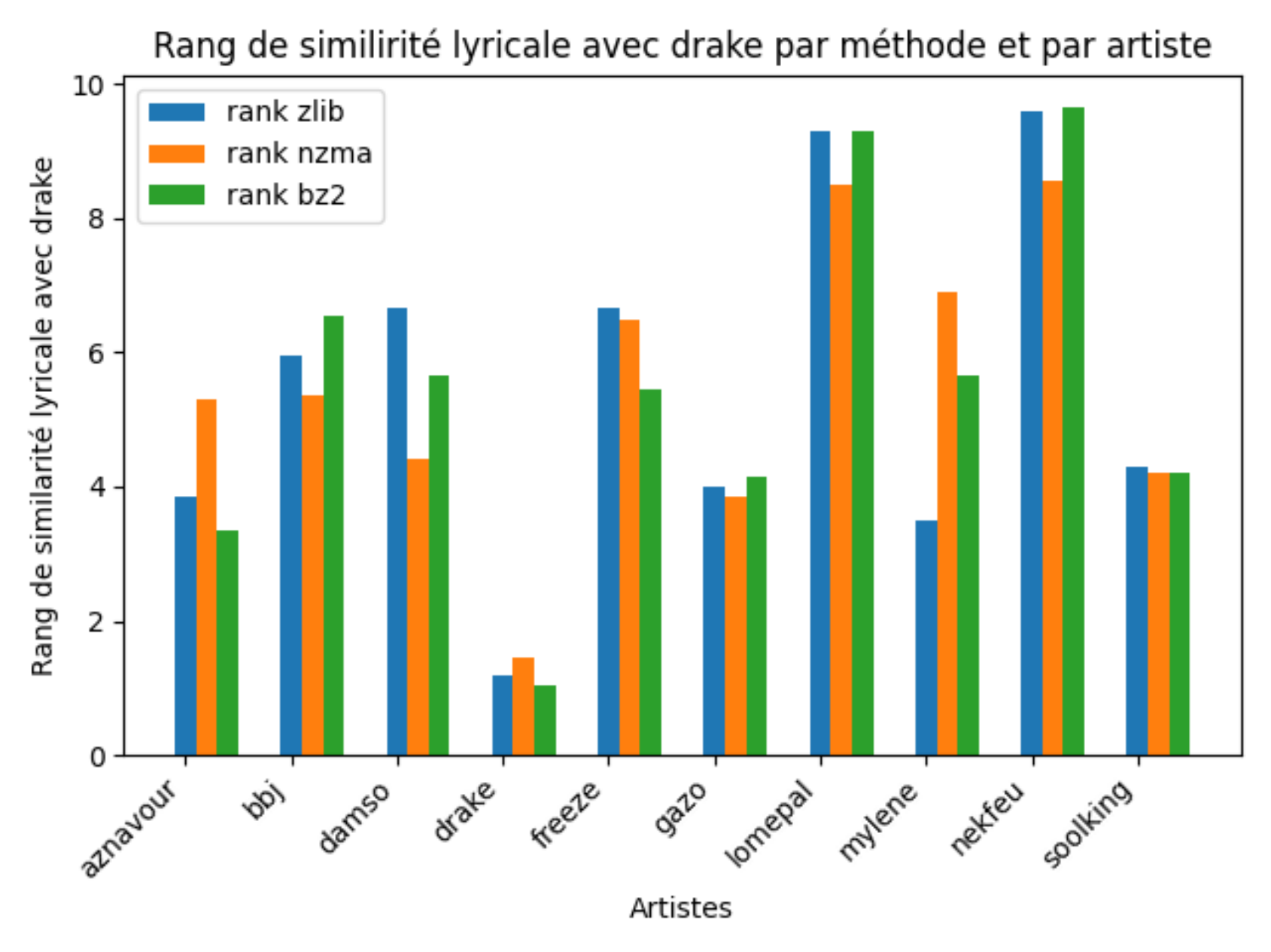
Mylène Farmer



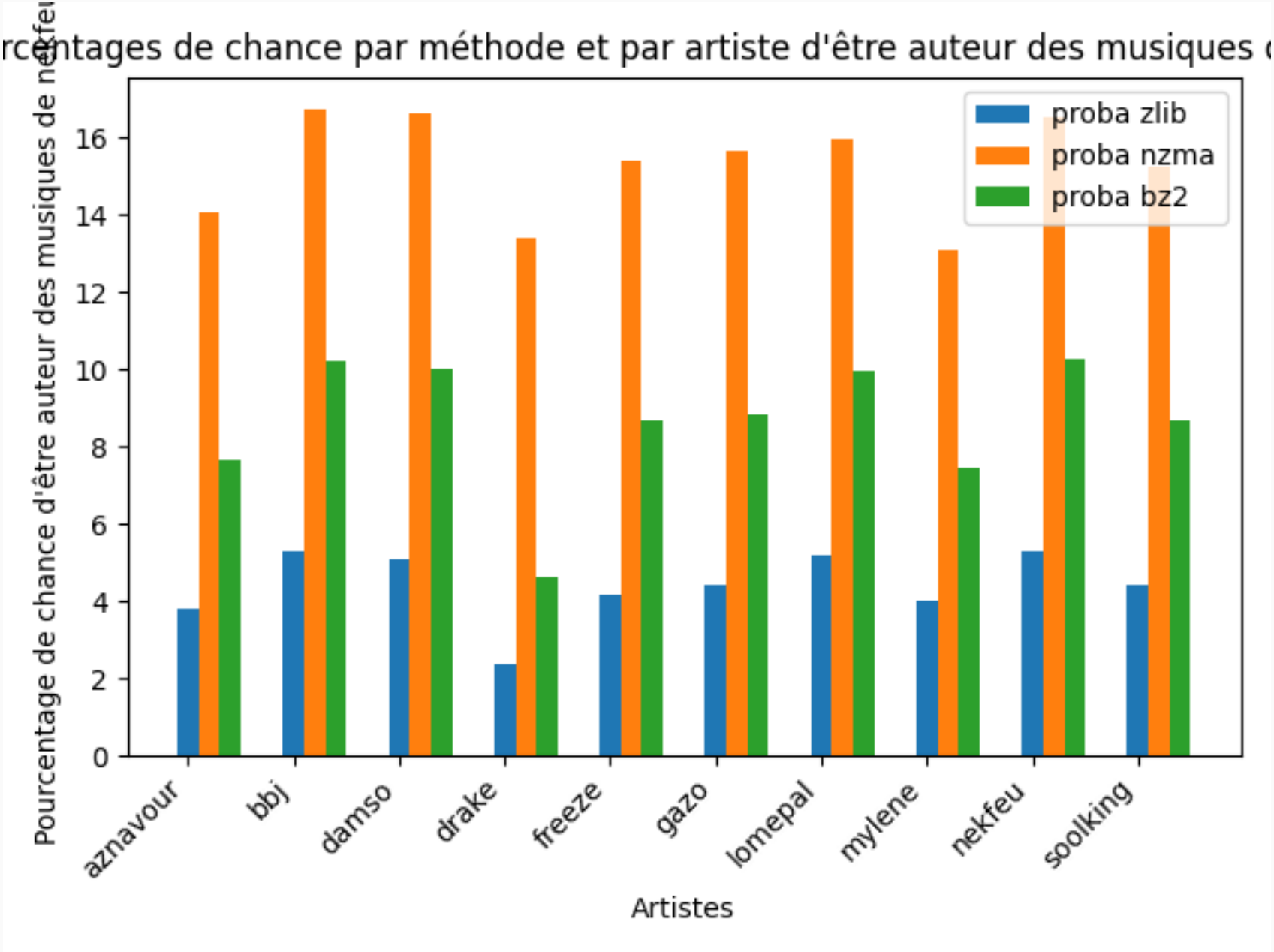
Drake


urcentages de chance par méthode et par artiste d'être auteur des musiques
Pourcentage de chance d'être auteur des musiques de Drake
proba zlib
proba nzma
proba bz2
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
aznavour
bbj
damso
drake
freeze
gazo
lomepal
mylene
nekfeu
soolking
Artistes

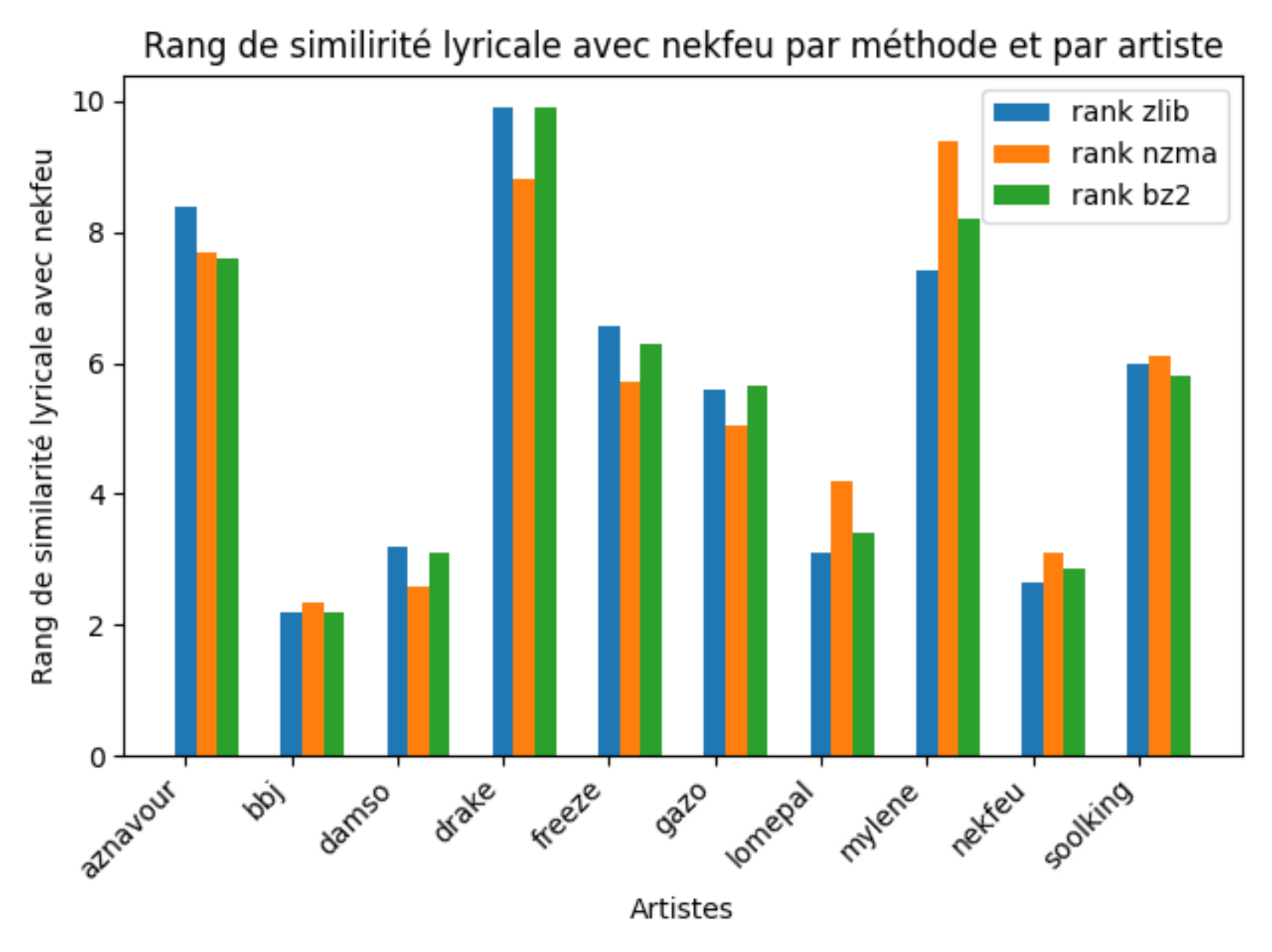
Drake


Rang de similirité lyricale avec drake par méthode et par artiste
rank zlib
rank nzma
rank bz2
Rang de similarité lyricale avec drake
0
2
4
6
8
10
aznavour
bbj
damso
drake
freeze
gazo
lomepal
mylene
nekfeu
soolking
Artistes

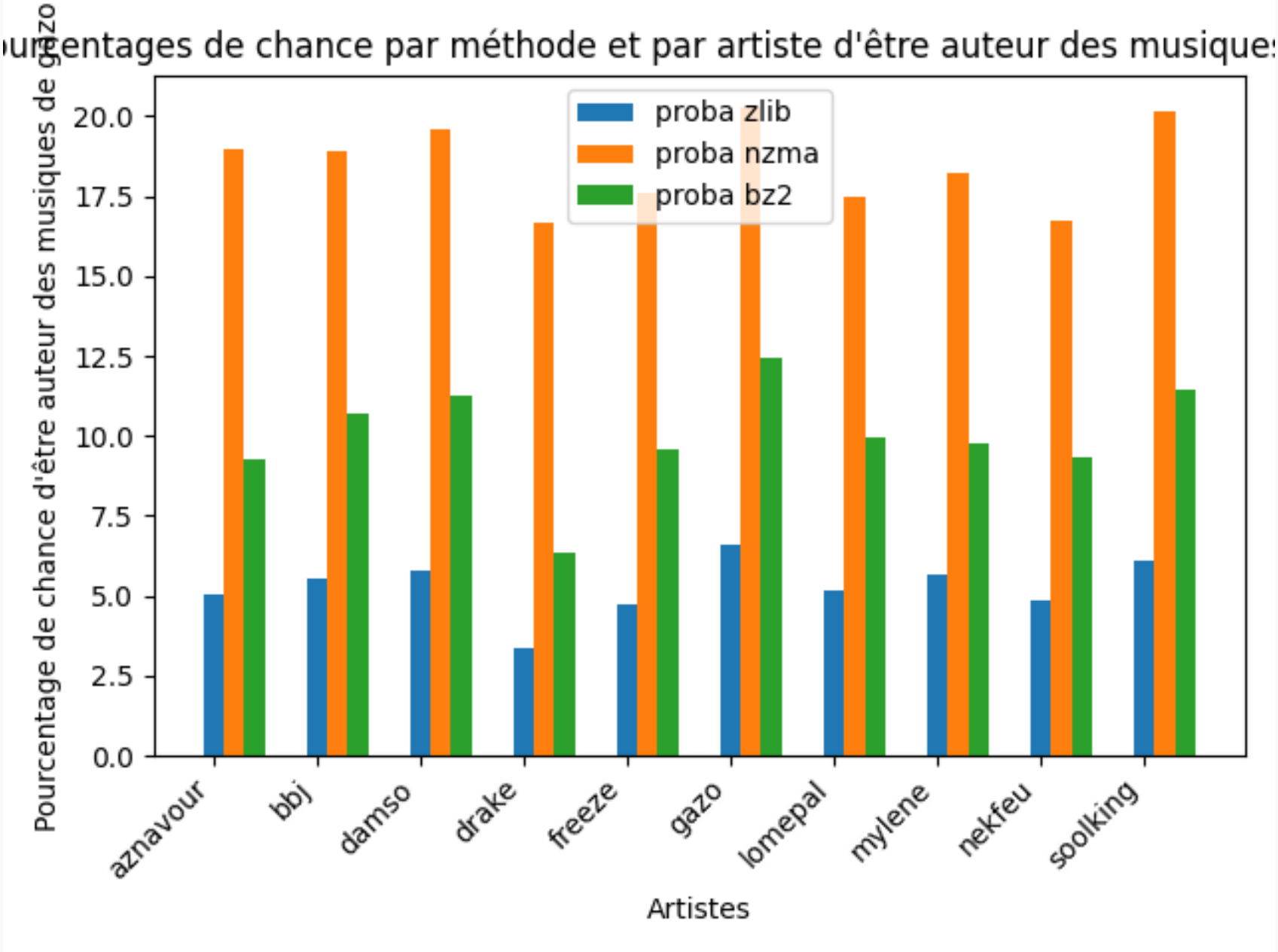
Nekfeu


rcentages de chance par méthode et par artiste d'être auteur des musiques
Pourcentage de chance d'être auteur des musiques de nekfeu
0
2
4
6
8
10
12
14
16
proba zlib
proba nzma
proba bz2
aznavour
bbj
damso
drake
freeze
gazo
lomepal
mylene
nekfeu
soolking
Artistes

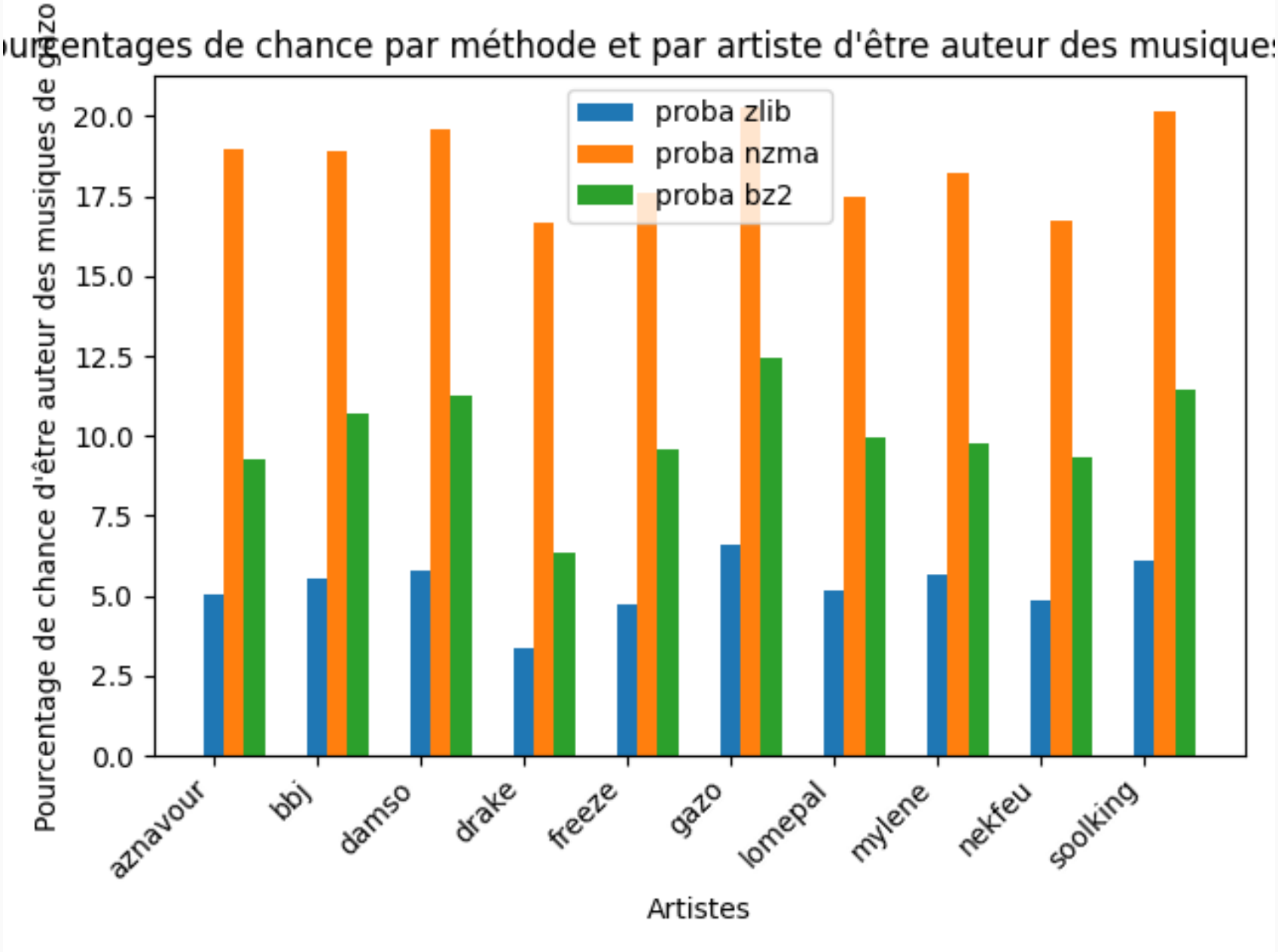
Nekfeu


Rang de similirité lyricale avec nekfeu par méthode et par artiste
rank zlib
rank nzma
rank bz2
Rang de similarité lyricale avec nekfeu
0
2
4
6
8
10
aznavour
bbj
damso
drake
freeze
gazo
lomepal
mylene
nekfeu
soolking
Artistes

Gazo


ourcentages de chance par méthode et par artiste d'être auteur des musiques
Pourcentage de chance d'être auteur des musiques de gazo
proba zlib
proba nzma
proba bz2
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
aznavour
bbj
damso
drake
freeze
gazo
lomepal
mylene
nekfeu
soolking
Artistes

Gazo


Pourcentages de chance par méthode et par artiste d'être auteur des musiques
Pourcentage de chance d'être auteur des musiques de gazo
proba zlib
proba nzma
proba bz2
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
aznavour
bbj
damso
drake
freeze
gazo
lomepal
mylene
nekfeu
soolking
Artistes

Conclusion

Conclusion et ouverture

- Utilisation de la NGD

$$\mathrm{NGD}(A,B) = \frac{\max(\log(f(A)), \log(f(B))) - \log(f(A \cap B))}{\log(N) - \min(\log(f(A)), \log(f(B)))}$$

- Prévision des résultats possibles ?

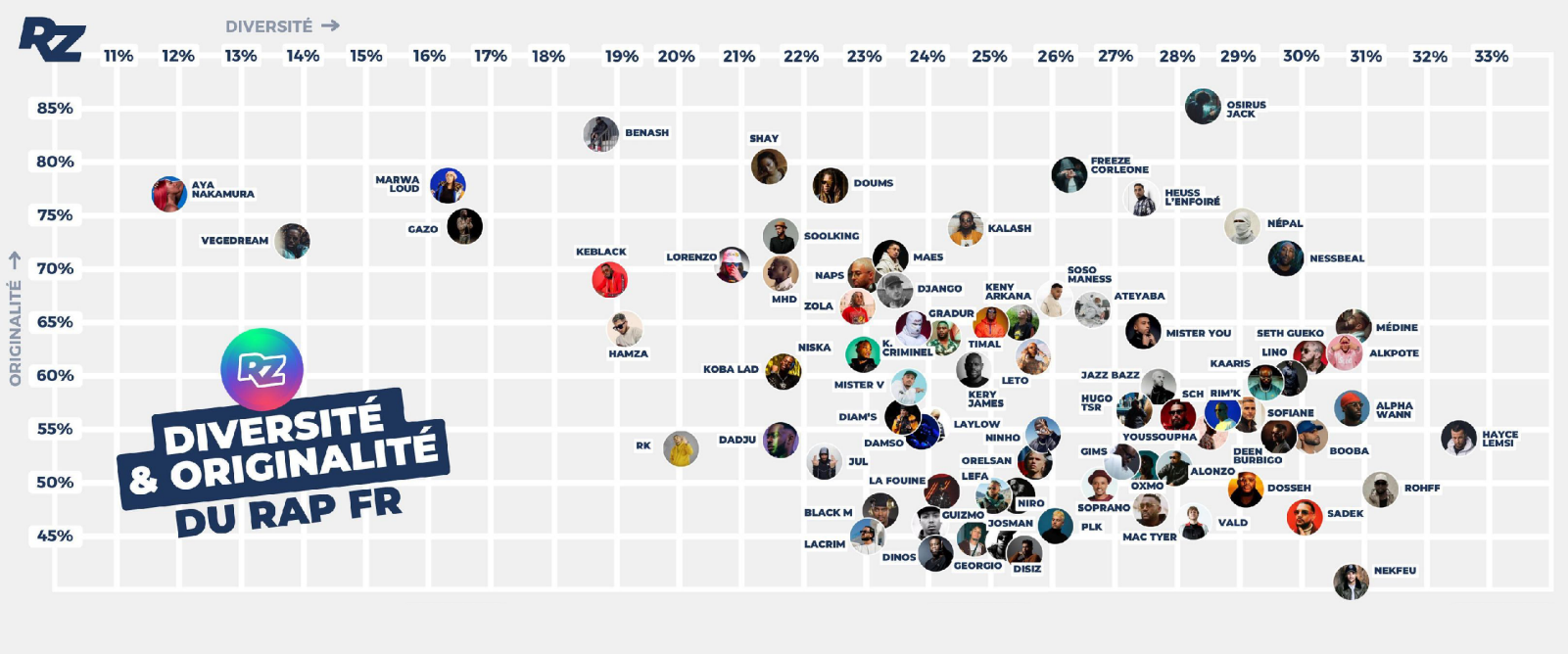


Figure 1 : Graphique diversité & originalité des auteurs de rap français

Questions?