

PROJET FINAL

M4 SOL203 - COGNITION, SENTIMENT, OPINION

M2 langue et informatique
XIE SHILIN
21102552

Sélection de livre



Cien años de soledad (Cent ans de solitude)

L'auteur: Gabriel García Márquez

Nombre des pages: 400

Nombre des paragraphes: 856

Langue: Espagnol

Format: txt

Formation de la liste des personnages

▼ Personajes

Primera generación

José Arcadio Buendía
Úrsula Iguarán

Segunda generación

José Arcadio
Coronel Aureliano Buendía
Amaranta

Rebeca

Tercera generación

Arcadio
Aureliano José
Los diecisiete Aurelianos

Cuarta generación

Remedios la bella
José Arcadio Segundo
Aureliano Segundo

Quinta generación

Renata Remedios (Meme)
José Arcadio
Amaranta Úrsula

Sexta generación

Aureliano Babilonia

Séptima generación

Rodrigo/Aureliano

Organiser la liste de noms par ordre décroissant de longueur, faire d'abord correspondre les noms avec des noms plus longs et supprimez la chaîne après la correspondance pour éviter des correspondances répétées.

```
character_names.sort(key=len, reverse=True)

for name in character_names:
    pattern = re.compile(r'\b' + re.escape(name) + r'\b')

    matches = pattern.findall(cleaned_text)
    for match in matches:
        character_count[name] += 1
        cleaned_text = cleaned_text.replace(match, '', 1)
```

-
- 1 Coronel Aureliano Buendía
 - 2 Los Diecisiete Aurelianos
 - 3 Santa Sofía De La Piedad
 - 4 José Arcadio Buendía
 - 5 José Arcadio Segundo
 - 6 Fernanda Del Carpio
 - 7 Aureliano Babilonia
 - 8 Mauricio Babilonia
 - 9 Remedios La Bella
 - 10 Aureliano Segundo
 - 11 Prudencio Aguilar
 - 12 Gerineldo Márquez
 - 13 Remedios Moscote
 - 14 Apolinar Moscote
 - 15 Renata Remedios
 - 16 Amaranta Úrsula
 - 17 Úrsula Iguarán
 - 18 Aureliano José
 - 19 Rebeca Montiel
 - 20 Pilar Ternera
 - 21 Nicanor Ulloa
 - 22 Pietro Crespi
 - 23 José Arcadio
 - 24 Petra Cotes
 - 25 Melquíades
 - 26 Nigromanta
 - 27 Amaranta
 - 28 Arcadio
 - 29 Rodrigo
 - 30 Mr.Herbert
 - 31 Rebeca
 - 32 Gastón
 - 33 Mr.Brown
-

Extraction de relations des personnages

```
# Initialiser une matrice bidimensionnelle pour stocker le nombre total d'apparitions simultanées de chaque paire de personnages
total_matrix = [[0] * len(character_names) for _ in range(len(character_names))]
```

```
# Parcourir tous les paragraphes par groupe de 2
for i in range(0, len(paragraphs)-1):
    group_paragraphs = paragraphs[i:i+1]
    # Initialiser une matrice pour compter les apparitions simultanées de chaque paire de personnages dans le groupe actuel
    group_matrix = [[0] * len(character_names) for _ in range(len(character_names))]
```

Créer des paires par deux des 33 personnages, faire correspondre si ces deux noms apparaissent dans chacun des deux paragraphes et comptez le nombre d'occurrences.

```
sorted_character_names = sorted(character_names, key=len, reverse=True)

# Parcourir chaque paragraphe du groupe actuel
for paragraph in group_paragraphs:
    characters = []
    # Parcourir les noms de personnages triés
    for name in sorted_character_names:
        # Trouver toutes les occurrences du nom de personnage dans le paragraphe
        matches = re.findall(r'\b' + re.escape(name) + r'\b', paragraph.lower())
        characters.extend(matches)
        # Supprimer les occurrences trouvées du paragraphe pour éviter les doublons
        paragraph = paragraph.replace(name, '', len(matches))

    # Pour chaque paire de personnages trouvée dans le paragraphe, incrémenter le compteur correspondant dans la matrice du groupe
    for pair in itertools.combinations(characters, 2):
        index1 = character_to_index.get(pair[0])
        index2 = character_to_index.get(pair[1])
        if index1 is not None and index2 is not None:
            group_matrix[index1][index2] += 1
            group_matrix[index2][index1] += 1

    # Ajouter les comptages du groupe actuel à la matrice totale
    for i in range(len(character_names)):
        for j in range(len(character_names)):
            total_matrix[i][j] += group_matrix[i][j]
```

Extraction de relations des personnages

josé arcadio buendía et úrsula iguarán apparaissent ensemble 3 fois.
 josé arcadio buendía et coronel aureliano buendía apparaissent ensemble 38 fois.
 josé arcadio buendía et amaranta apparaissent ensemble 97 fois.
 josé arcadio buendía et rebeca apparaissent ensemble 77 fois.
 josé arcadio buendía et arcadio apparaissent ensemble 61 fois.
 josé arcadio buendía et aureliano josé apparaissent ensemble 2 fois.
 josé arcadio buendía et santa sofía de la piedad apparaissent ensemble 9 fois.
 josé arcadio buendía et josé arcadio segundo apparaissent ensemble 25 fois.
 josé arcadio buendía et aureliano segundo apparaissent ensemble 30 fois.
 josé arcadio buendía et fernanda del carpio apparaissent ensemble 2 fois.
 josé arcadio buendía et josé arcadio apparaissent ensemble 42 fois.
 josé arcadio buendía et amaranta úrsula apparaissent ensemble 8 fois.
 josé arcadio buendía et remedios moscote apparaissent ensemble 4 fois.
 josé arcadio buendía et prudencio aguilar apparaissent ensemble 59 fois.
 josé arcadio buendía et melquíades apparaissent ensemble 151 fois.
 josé arcadio buendía et pilar ternera apparaissent ensemble 20 fois.
 josé arcadio buendía et nicanor ulloa apparaissent ensemble 4 fois.
 josé arcadio buendía et rebeca montiel apparaissent ensemble 4 fois.
 josé arcadio buendía et pietro crespi apparaissent ensemble 40 fois.
 josé arcadio buendía et gerineldo márquez apparaissent ensemble 7 fois.
 josé arcadio buendía et petra cotes apparaissent ensemble 6 fois.
 josé arcadio buendía et herbert apparaissent ensemble 2 fois.
 josé arcadio buendía et brown apparaissent ensemble 5 fois.
 josé arcadio buendía et mauricio babilonia apparaissent ensemble 2 fois.
 josé arcadio buendía et mème apparaissent ensemble 17 fois.
 josé arcadio buendía et apolinar moscote apparaissent ensemble 24 fois.
 úrsula iguarán et coronel aureliano buendía apparaissent ensemble 1 fois.
 úrsula iguarán et josé arcadio segundo apparaissent ensemble 1 fois.
 úrsula iguarán et melquíades apparaissent ensemble 4 fois.
 coronel aureliano buendía et amaranta apparaissent ensemble 116 fois.
 coronel aureliano buendía et rebeca apparaissent ensemble 49 fois.
 coronel aureliano buendía et arcadio apparaissent ensemble 18 fois.
 coronel aureliano buendía et aureliano josé apparaissent ensemble 20 fois.
 coronel aureliano buendía et santa sofía de la piedad apparaissent ensemble 38 fois.
 coronel aureliano buendía et josé arcadio segundo apparaissent ensemble 54 fois.

```

import csv
with open('pair_counts.csv', 'w', newline='') as csvfile:
    fieldnames = ['source', 'destination', 'weight']
    writer = csv.DictWriter(csvfile, fieldnames=fieldnames)
    writer.writeheader()
    for i in range(len(character_names)):
        for j in range(i+1, len(character_names)):
            pair_count = total_matrix[i][j]
            if pair_count > 0:
                character1 = character_names[i]
                character2 = character_names[j]
                writer.writerow({'source': character1, 'destination': character2, 'weight': pair_count})
  
```

```

import pandas as pd

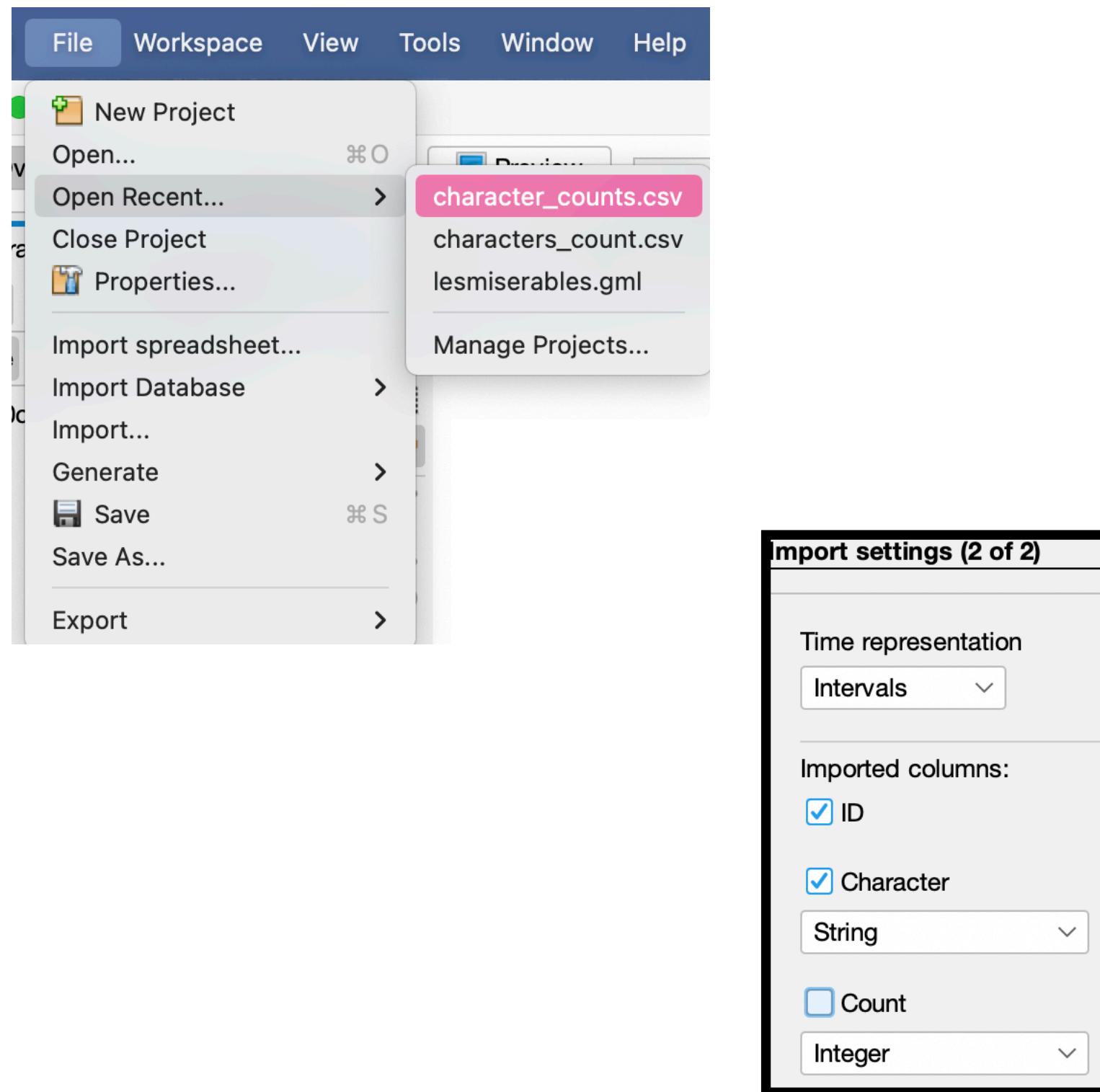
# Lire le premier fichier CSV et créer un dictionnaire de correspondance des noms vers les ID
df1 = pd.read_csv("/Users/shailynnxiie/Documents/M2/Cognition/character_counts.csv")
name_to_id = {name: idx for idx, name in zip(df1['ID'], df1['Character'])}

# Lire le deuxième fichier CSV
df2 = pd.read_csv("/Users/shailynnxiie/Documents/M2/Cognition/relation.csv")

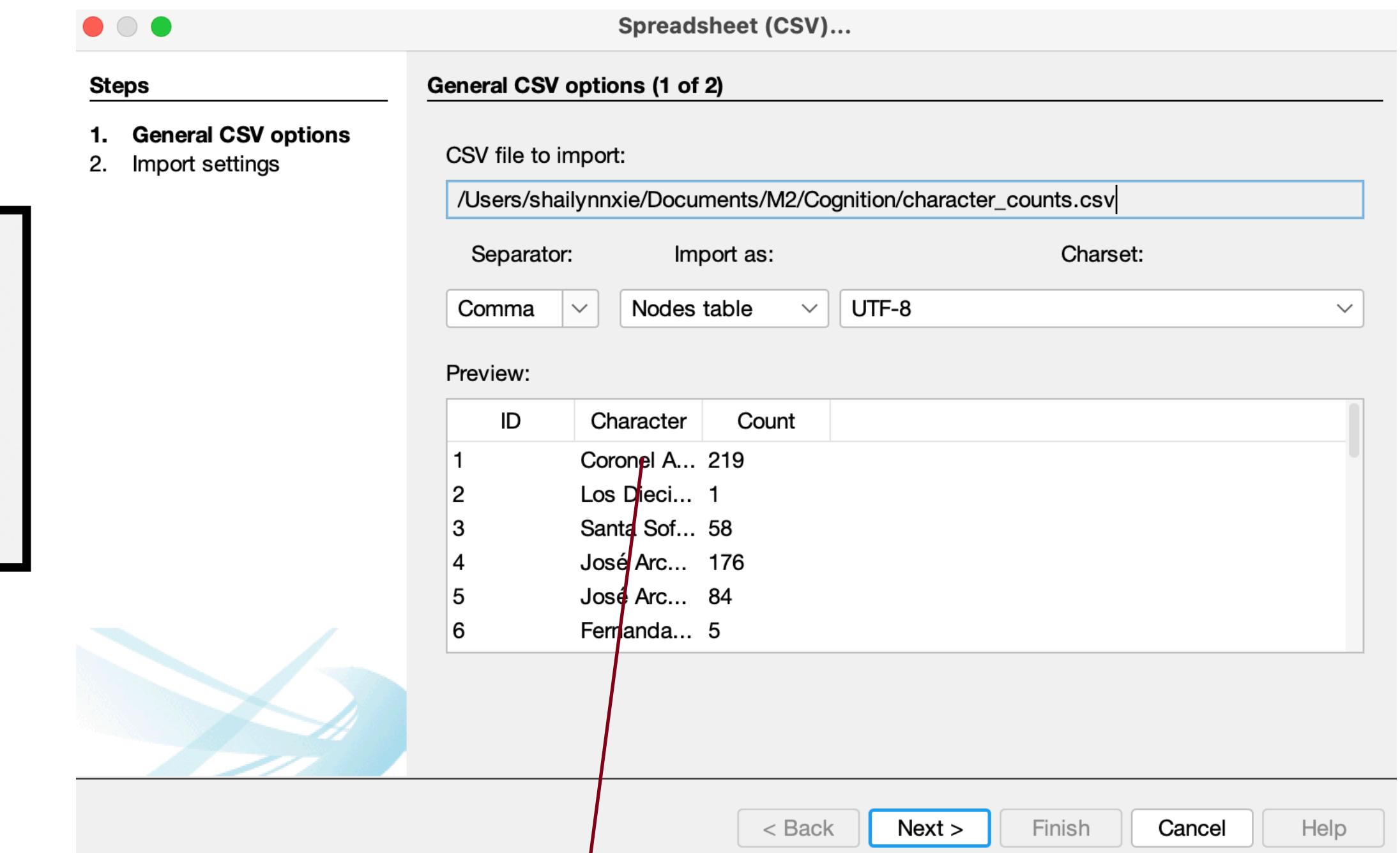
# Convertir les noms en ID et les ajouter au DataFrame
df2['source_id'] = df2['source'].map(name_to_id)
df2['destination_id'] = df2['destination'].map(name_to_id)

# Sauvegarder le résultat dans un nouveau fichier CSV
df2.to_csv("second_file_updated.csv", index=False)
  
```

Importer les fichiers .csv dans Gephi



Graph Type: Mixed
of Nodes: 33
of Edges: 0
Dynamic Graph: no
Dynamic Attributes: no
Multi Graph: no



The screenshot shows the 'Spreadsheet (CSV)...' import dialog. It has two tabs: 'General CSV options (1 of 2)' and 'Import settings (2 of 2)'. The first tab shows the file path '/Users/shailynnxié/Documents/M2/Cognition/character_counts.csv', separator 'Comma', import as 'Nodes table', and charset 'UTF-8'. The second tab shows a preview of the data with columns 'ID', 'Character', and 'Count'. The data listed is:

ID	Character	Count
1	Coronel A...	219
2	Los Dieci...	1
3	Santa Sof...	58
4	José Arc...	176
5	José Arc...	84
6	Fernanda...	5



05

Importer les fichiers .csv dans Gephi

Graph x Data Table x

Nodes Edges Configuration Add node Add edge Search/Replace Import Spreadsheet

Spreadsheet (CSV)...

Steps

1. General CSV options
2. Import settings

General CSV options (1 of 2)

CSV file to import:
/Users/shailynnxe/Documents/M2/Cognition/Edges.csv

Separator: Comma Import as: Edges table Charset: UTF-8

Preview:

Source	Target	Type	Weight
4	17	non dirigé	3
4	1	non dirigé	38
4	27	non dirigé	97
4	31	non dirigé	77
4	28	non dirigé	61
4	18	non dirigé	2

< Back Next > Finish Cancel Help

Spreadsheet (CSV)...

Steps

1. General CSV options
2. Import settings

Import settings (2 of 2)

Time representation: Intervals

Imported columns:

Source
 Target
 Type

Import report

Source: Edges.csv

Issues Report

Nodes Issues

Parallel edges detected, remember to choose a merge strategy INFO

Graph Type: Undirected More options...

of Nodes: 32
of Edges: 224

Dynamic Graph: no
Dynamic Attributes: no
Multi Graph: no

New workspace
Append to existing workspace

OK Cancel

Importer les fichiers .csv dans Gephi

Source	Target	Type	Id	Label	Interval
32	26	Undirected	447		
8	15	Undirected	446		
33	26	Undirected	445		
33	15	Undirected	444		
33	8	Undirected	443		
30	33	Undirected	442		
24	15	Undirected	441		
24	8	Undirected	440		
24	33	Undirected	439		
12	26	Undirected	438		
12	14	Undirected	437		
12	15	Undirected	436		
22	14	Undirected	435		
22	15	Undirected	434		
22	12	Undirected	433		
21	19	Undirected	432		
20	26	Undirected	431		
20	14	Undirected	430		
20	15	Undirected	429		
20	32	Undirected	428		
20	8	Undirected	427		
20	24	Undirected	426		
20	12	Undirected	425		
20	22	Undirected	424		
25	26	Undirected	423		
25	14	Undirected	422		
25	15	Undirected	421		
			4	José Arcadio Buendía	
			17	Úrsula Igurán	
			1	Coronel Aureliano Buendía	
			27	Amaranta	
			31	Rebeca	
			28	Arcadio	
			18	Aureliano José	
			3	Santa Sofía De La Piedad	
			5	José Arcadio Segundo	
			10	Aureliano Segundo	
			6	Fernanda Del Carpio	
			23	José Arcadio	
			16	Amaranta Úrsula	
			13	Remedios Moscote	
			11	Prudencio Aguilar	
			25	Melquíades	
			20	Pilar Ternera	
			21	Nicanor Ulloa	
			19	Rebeca Montiel	
			22	Pietro Crespi	
			12	Gerineldo Márquez	
			24	Petra Cotes	
			30	Mr.Herbert	
			33	Mr.Brown	
			8	Mauricio Babilonia	
			15	Renata Remedios	
			14	Apolinar Moscote	

07

Importer les fichiers .csv dans Gephi

Graph x Data Table x

Nodes Edges Configuration Add node Add edge Search/Replace Import Spreadsheet Export table More

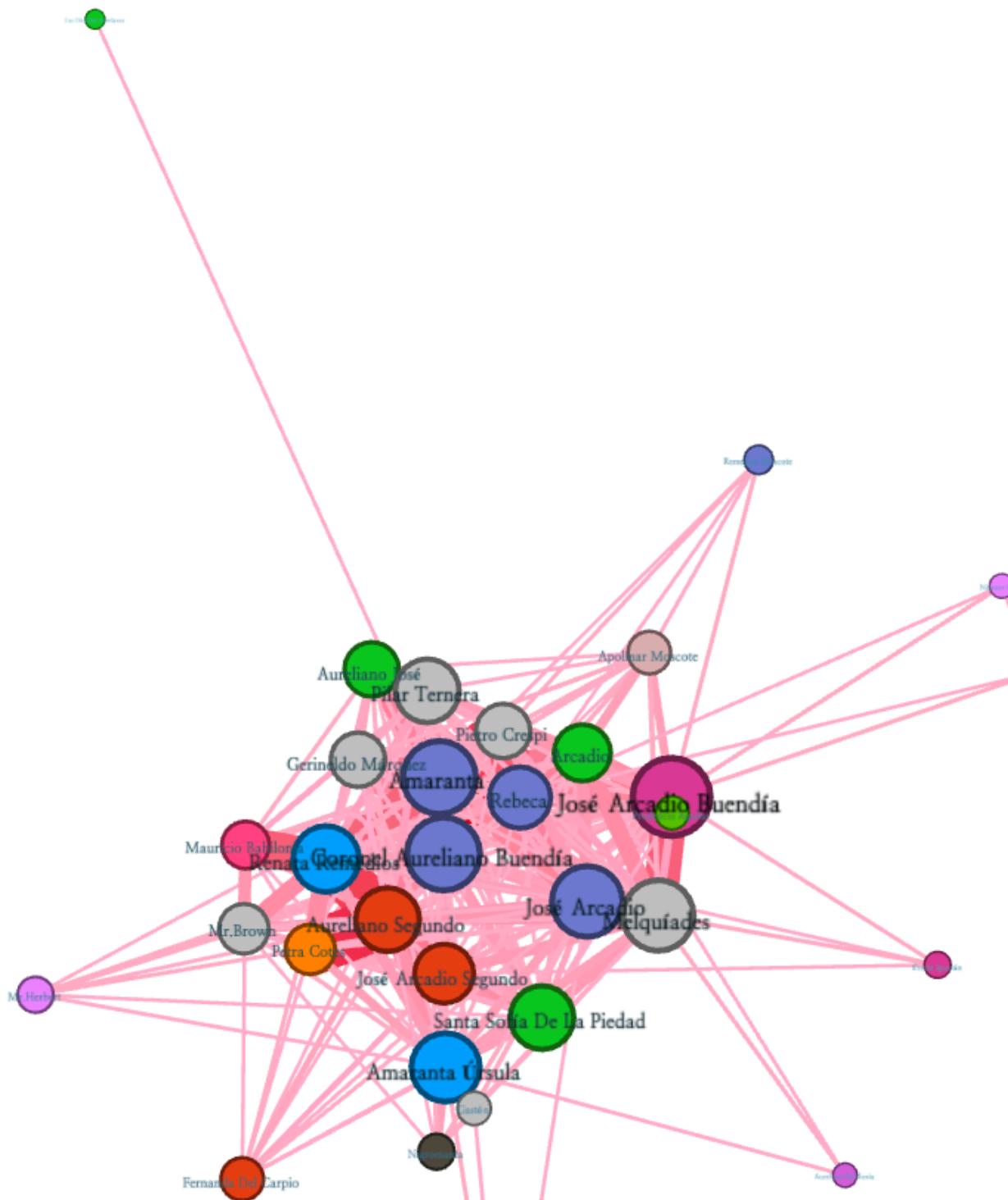
Id	Label	Interval	Eccentricity	Closeness...	Harmonic Closeness ...	Betwe...	Generation
4	José Arcadio Buendía	2.0	0.861111	0.919355	58.82...	1	
17	Úrsula Iguarán	3.0	0.516667	0.553763	0.0	1	
1	Coronel Aureliano Buendía	2.0	0.815789	0.887097	23.76...	2	
27	Amaranta	2.0	0.815789	0.887097	46.21...	2	
31	Rebeca	2.0	0.72093	0.806452	27.13...	2	
28	Arcadio	2.0	0.688889	0.774194	8.420...	3	
18	Aureliano José	2.0	0.673913	0.758065	1.552...	3	
3	Santa Sofía De La Piedad	2.0	0.738095	0.822581	6.445...	3	
5	José Arcadio Segundo	2.0	0.704545	0.790323	5.933...	4	
10	Aureliano Segundo	2.0	0.738095	0.822581	4.604...	4	
6	Fernanda Del Carpio	2.0	0.607843	0.677419	0.153...	4	
23	José Arcadio	2.0	0.794872	0.870968	26.01...	2	
16	Amaranta Úrsula	2.0	0.775	0.854839	31.89...	5	
13	Remedios Moscote	3.0	0.508197	0.55914	0.0	2	
11	Prudencio Aguilar	3.0	0.553571	0.607527	0.071...		
25	Melquíades	2.0	0.794872	0.870968	28.78...		
20	Pilar Ternera	2.0	0.738095	0.822581	9.477...		
21	Nicanor Ulloa	3.0	0.484375	0.521505	0.0		
19	Rebeca Montiel	3.0	0.484375	0.521505	0.0		
22	Pietro Crespi	2.0	0.673913	0.758065	5.181...		
12	Gerineldo Márquez	2.0	0.673913	0.758065	3.336...		
24	Petra Cotes	2.0	0.645833	0.725806	0.595...		
30	Mr.Herbert	2.0	0.574074	0.629032	0.0		
33	Mr.Brown	2.0	0.645833	0.725806	2.769...		
8	Mauricio Babilonia	2.0	0.632653	0.709677	0.906...		
15	Renata Remedios	2.0	0.72093	0.806452	2.880...	5	
14	Apolinar Moscote	3.0	0.596154	0.672043	1.694...		

Graph x Data Table x

Nodes Edges Configuration Add node Add edge Search/Replace Import Spr

Source	Target	Type	Id	Weight
32	26	Undirected	447	9.0
8	15	Undirected	446	159.0
33	26	Undirected	445	1.0
33	15	Undirected	444	67.0
33	8	Undirected	443	44.0
30	33	Undirected	442	9.0
24	15	Undirected	441	69.0
24	8	Undirected	440	2.0
24	33	Undirected	439	12.0
12	26	Undirected	438	1.0
12	14	Undirected	437	6.0
12	15	Undirected	436	6.0
22	14	Undirected	435	7.0
22	15	Undirected	434	14.0
22	12	Undirected	433	18.0
21	19	Undirected	432	1.0
20	26	Undirected	431	6.0
20	14	Undirected	430	2.0
20	15	Undirected	429	21.0
20	32	Undirected	428	4.0
20	8	Undirected	427	12.0
20	24	Undirected	426	2.0
20	12	Undirected	425	6.0
20	22	Undirected	424	12.0
25	26	Undirected	423	1.0
25	14	Undirected	422	9.0
25	15	Undirected	421	4.0

Représentation de graph



Nodes Edges A T

Unique Partition Ranking

Generation

- 2 (26.32%)
- 3 (21.05%)
- 4 (21.05%)
- 1 (10.53%)
- 5 (10.53%)
- 6 (5.26%)

Palette...

Nodes Edges A T

Unique Partition Ranking

Weight

Color:

Layout x

Force Atlas

Run

Force Atlas

Inertia: 0.1
Repulsion strength: 2000.0
Attraction strength: 10.0
Maximum displacement: 10.0
Auto stabilize function:

Autostab Strength: 80.0
Autostab sensibility: 0.2
Gravity: 30.0
Attraction Distrib.:
Adjust by Size:

Nodes Edges A T

Unique Partition Ranking

Degree

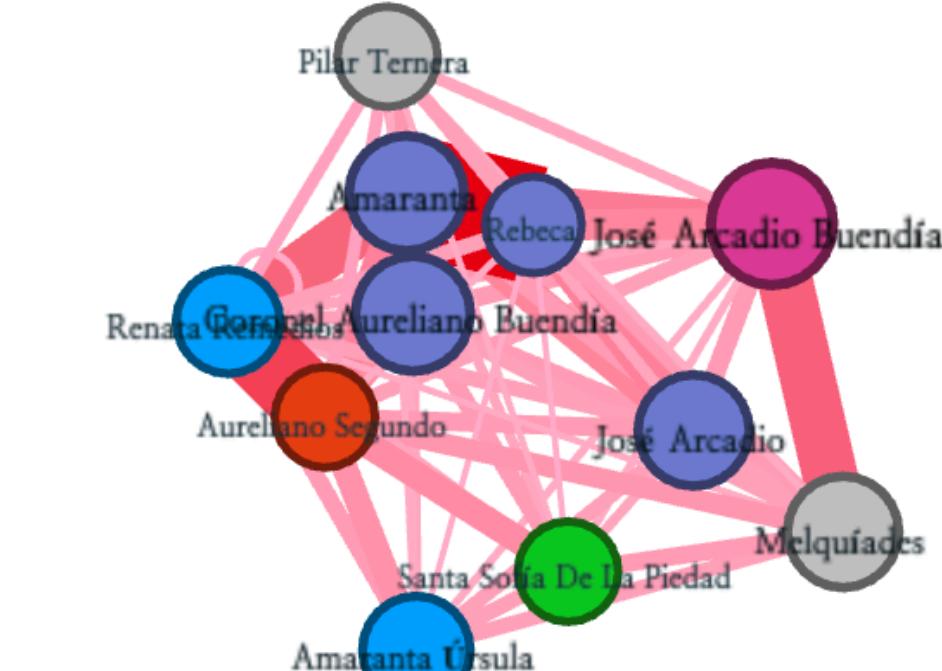
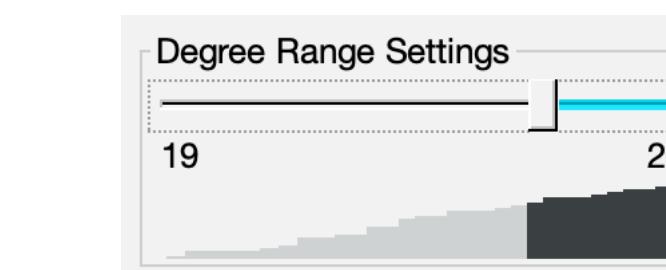
Color:

Nodes Edges A T

Unique Ranking

Degree

Min size: 0.5 Max size: 2.5



09

Démonstration sur Gephi

Merci de votre attention