

PART II: SENEKAL COMMUNITY ANALYSIS

Community detection in network theory is a graph partitioning problem where a group of nodes is identified such that the connections a node has within a group are more than the connections a node has outside a group.

Data Source

#Senekal Reply Network - Largest Component

Data Transformation

Community detection techniques are typically applied to undirected networks. The direction in which a tie is sent or received is irrelevant; a connection is established as long as there is an interaction between users. The current network is therefore converted to an undirected network for the purposes of this analysis.

Method

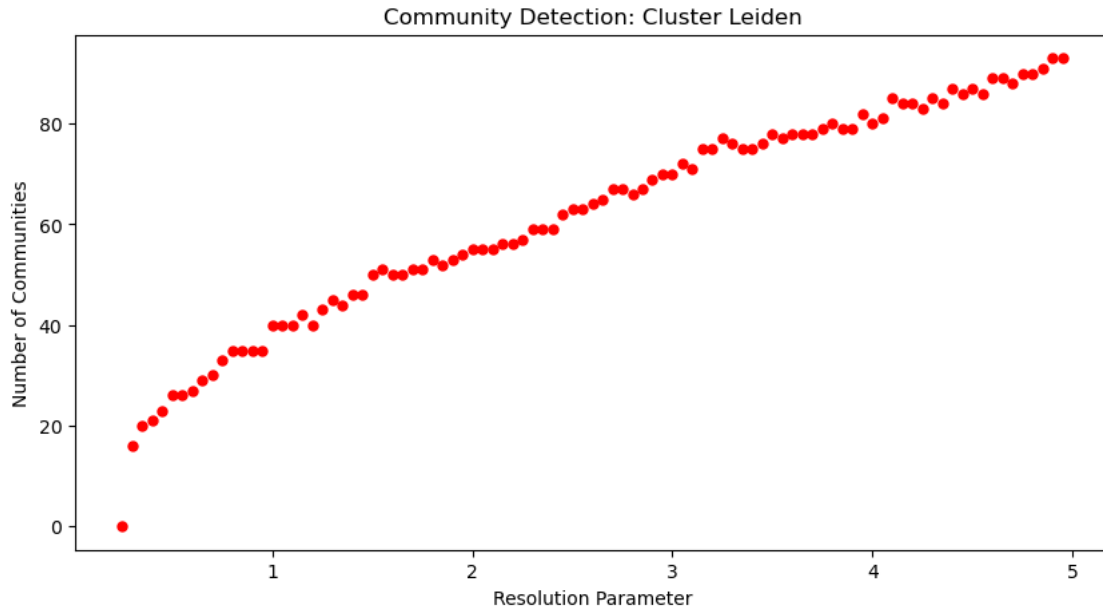
Using python's iGraph library, I apply the **Leiden** method to this community detection problem.

Undirected Graph:

IGRAPH U-W- 2207 2622

1 OPTIMISATION

To choose an optimal resolution parameter (γ), I estimate the leiden algorithm and explore community outcomes across different values of γ . Letting the data tell me which points are more optimal than others, I select γ at the points in the graph where the number of communities detected reach a constant plateau or flatness.



2 DETECTION

Setting $\gamma = 0.8$, I detect 35 communities with a modularity of 0.85. The detected communities are visualised below along with descriptions of the top 3 largest communities (by weighted degree).

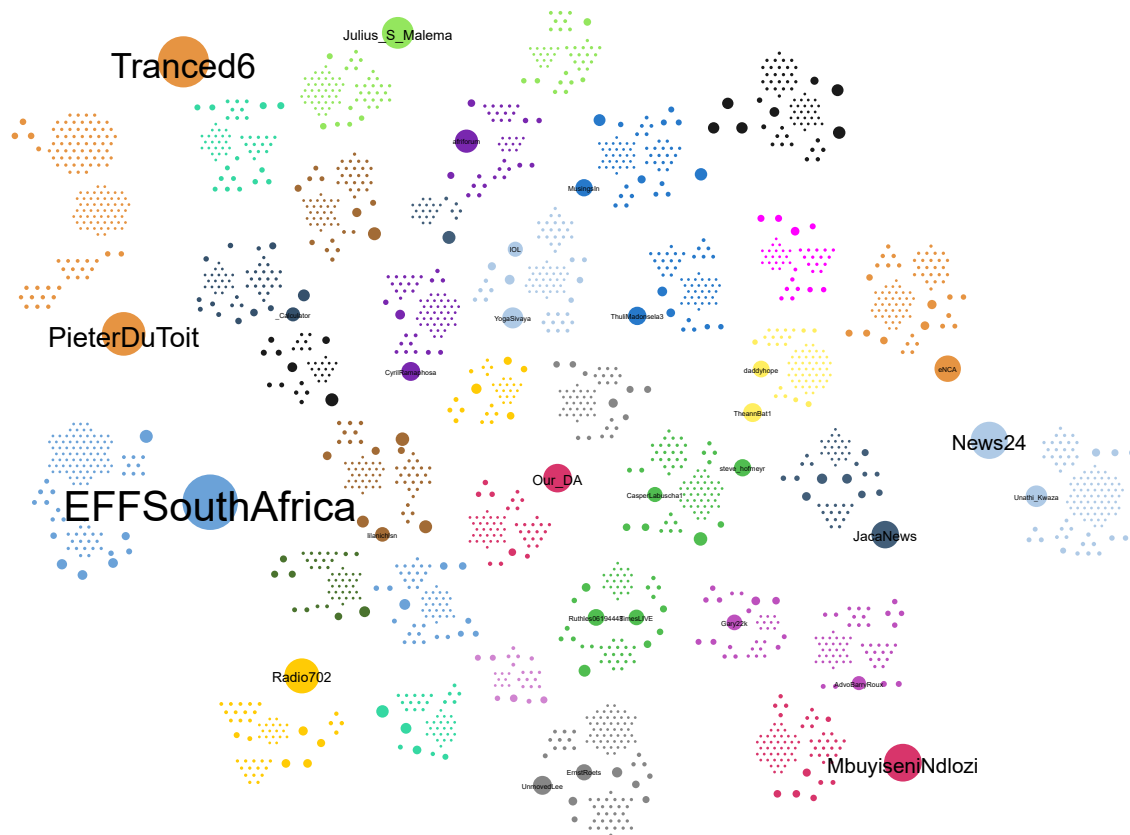
Communities:

35

Modularity:

0.853981873076067

Senekal Communities:



Top 3 Weighted Degree Communities:

Community 1:

Size:

150

Density:

0.01888143176733781

Average Degree:

5.63

Total Degree:

844

Members (10*) :

0 Tranced6
1 PieterDuToit
2 MightiJamie
3 AbutiJosef
4 VelempiniN
5 carol_zeph3
6 numbaONEhustla
7 miyiref16
8 SAfmRadio
9 SiyaMfundisi

Community 2:

Size:

122

Density:

0.01910310256062864

Average Degree:

4.62

Total Degree:

564

Members (10*) :

0 EFFSouthAfrica
1 No_Filter_7777
2 c_kumalo
3 KingK10850845
4 xeshamusiq
5 RhodieEx
6 IvorChalmers_
7 BukhosiNome
8 ntsikimazwai
9 4_South_Africa

Community 3:

Size:
109

Density:
0.02123683316343867

Average Degree:
4.59

Total Degree:
500

Members (10*) :

0 News24
1 Unathi_Kwaza
2 chatnark
3 SundayTimesZA
4 IviksaF
5 kalliekriel
6 MilowoMazwane
7 CharlotteKhuma3
8 e744890fafb84a4
9 Collen82241585

*Top 10 highest degree in community.

3 EVALUATION

Using various pair counting scores, I compare community outcomes of the leiden model at $\gamma=0.8$ and $\gamma=2.5$, and compare the leiden model at $\gamma=0.8$ and the louvain model.

3.1 Leiden ($\gamma=0.8$) vs Leiden ($\gamma=2.5$)

Leiden ($\gamma=2.5$)

Communities:
76

Modularity:

0.8466383135298357

Comparison Scores:

Variation of Information:

0.8299987545465558

Adjusted Rand:

0.5619120596338842

Normalized Mutual Information:

0.8299987545465558

3.2 Leiden (Gamma=0.8) vs Louvain

Louvain

Communities:

41

Modularity:

0.851687930991534

Comparison Scores:

Variation of Information:

0.8458184186306478

Adjusted Rand:

0.6837633035629342

Normalized Mutual Information:

0.8458184186306478

4 SUMMARY

In this section, I use community detection methods to identify densely connected groups and communities present in the #Senekal twitter reply network.

When searching for optimum modularity, I find the resolution parameter at $\gamma=0.8$ and $\gamma=2.5$ robust. I detect 35 and 76 communities with a modularity score of 0.854 and 0.847 at these points, respectively. Similarly, using the louvain method, 41 communities are detected with a modularity score of 0.851. The coincidence scores from the evaluation and the modularity scores, which are above the 0.7 threshold, are all a sign of a well partitioned network. However, given the large number of communities detected, a reasonable or meaningful interpretation is difficult. In other words, attempting to characterise all 35 communities or predict each of their behaviour and influence in the discourse based on their memberships will unlikely result in an intuitive outcome.

To atleast get some sense of the communities detected, I examine the sizes, densities, and degrees of the top 3 largest communities as well as the top 10 members (sorted by *weighted degree*) assigned to each of the 3 communities. Going by the users and tweet content, each community appears balanced in terms of political or other affiliations as well as members in support or opposition of the protest. I fail to find evidence of echo chambers (atleast in the top 3 largest communities). Ofcourse, this may be different in smaller, less prominent communities.

In the next section, I classify users by the positions or roles they occupy within the discourse based on their patterns of interactions and other network features.