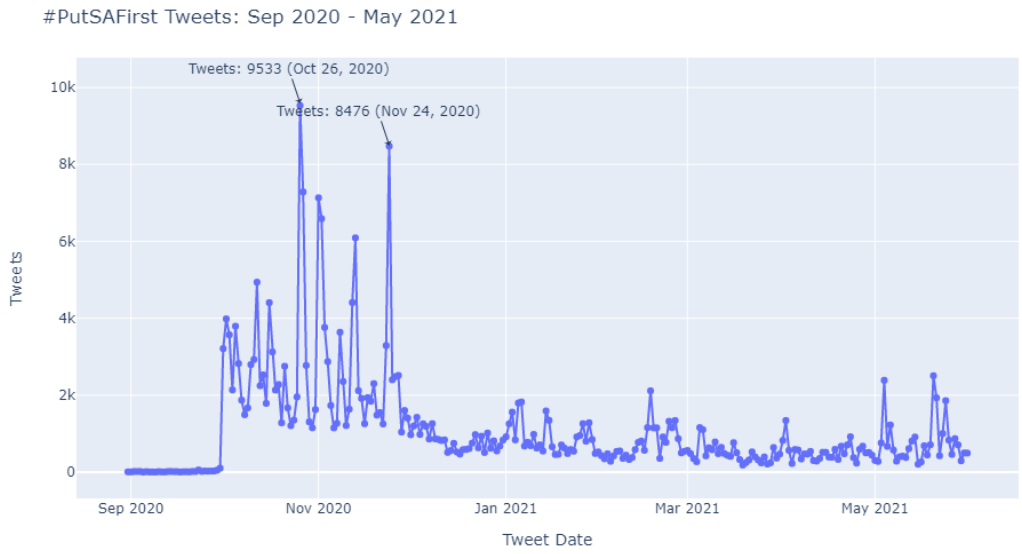
PART I: PUTSAFIRST
DESCRIPTIVE ANALYSIS

The #PutSAFirst movement, initially spearheaded by the twitter account *@uLerato_Pillay* who was later unmasked as a dismissed member of the South African Defence Force in July 2020, is a nationalist/patriotic (extremist) movement demanding the prioritisation of citizen rights over those of foreign nationals. Advocates of the movement have stood up in support of its strongly worded nationalist messaging that prides itself in its zero-tolernace for crime (such as drugs and human-trafficking), unemployment (what they call an exclusion from the labour market in favour of foreigners), and illegal immigration (that places an additional burden on public services). According to them, the South African Police Service and the state are to blame because of their lack of initiative, enforcement and proper implementation of immigration laws. Considering the nature of the tweets and anti-immigrant sentiment, and constant association with hashtags like #WeWantOurCountryBack and #ForeignersMustGo, it is often criticised for masquerading xenophobia and promulgating seeds of dissent and discord, mainly through tactics of disinformation.

Data Source

Data is extracted from Twitter's API matching the following search query: "PutSouthAfricaFirst", "PutSouthAfricansFirst", for the period: 1 September 2020 to 31 May 2021.

Number of Daily Tweets in #PutSAFirst Discourse:



Data Transformation

The extracted data set consists of tweets, retweets and replies to tweets related to this movement, however for the purposes of this analysis, I focus purely on direct engagements and interactions between users (i.e.

replies to tweets). To construct this type of network, I only include data where a user tweets a reply (sender) to a user whose tweet was replied to (receiver).

To ensure that the transformed data still reflects the full discourse and no significant information is lost as a result of the conversion from full network to reply network, I derive aggregate statistics of the excluded tweet and retweet data. In this way, the properties of the #PutSAFirst full network are still described in the the #PutSAFirst reply network. See below for a list of derived variables before transformation.

- **Derived node variables**:
  - *#tweeted* -> A measure of the total number of tweets by a user in the #PutSAFirst network.
  - *#retweeted* -> A measure of the total number of tweets by a user that are retweeted by other users in the #PutSAFirst network.
  - *#liked* -> A measure of the total number of tweets by a user that are liked by other users in the #PutSAFirst network.
- **Derived edge variables**
  - *weight* -> This captures the total number of replies from a sender to a receiver in the #PutSAFirst network. For example, replying to a user only once gives a weight = 1, but replying to a user's tweets multiple times gives a weight = number of replies to that user.

Once variables are derived, the extracted dataset is converted to a directed network, which I call the #PutSAFirst Full Network.

- **#PutSAFirst Full Network**

  The #PutSAFirst full network consists of 300443 tweets (incl. retweets, and replies to tweets) with 38772 active users behind those tweets.
  - Number of users: 38772
  - Number of tweets: 300443

After dropping all tweets and retweets that do not have a reply, the #PutSAFirst Full Network is then converted to the #PutSAFirst Reply Network.

- **#PutSAFirst Reply Network**

  The #PutSAFirst reply network includes data on users replying to tweets and users whose tweets are being replied to. This is a weighted network where weights measure the frequency of interaction, i.e. the number of times a user replies to the same users tweets.
  - Number of users: 8567
  - Number of replies: 27455

Next, I exclude all disconnected components and focus my analysis on the #PutSAFirst Reply Network - Largest Connected Component. This sub-network is found by extracting the component with the largest number of connected nodes in the network. To explain, a connected component is a network or subnetwork where every node is atleast connected to one other node, i.e. there must exist a path such that any two nodes can be connected starting from point A to point B.

- **#PutSAFirst Reply Network - Largest Connected Component**

  The #PutSAFirst reply network has 529 connected components. In this network, the largest connected component has 7606 nodes and 20951 edges, and the second largest connected component has 18 nodes and 17 edges. Going forward, I will only analyse nodes in this component.

  - Number of users: 7606
  - Number of replies: 26601

Method

Using pythons iGraph library, I describe the #PutSAFirst Reply Network - Largest Connected Component in terms of its **centrality** and **connectivity and cohesion** properties.

```
--------
Directed Network:

IGRAPH DNW- 7606 20951
```

---

1 CENTRALITY

1.1 Node Centrality

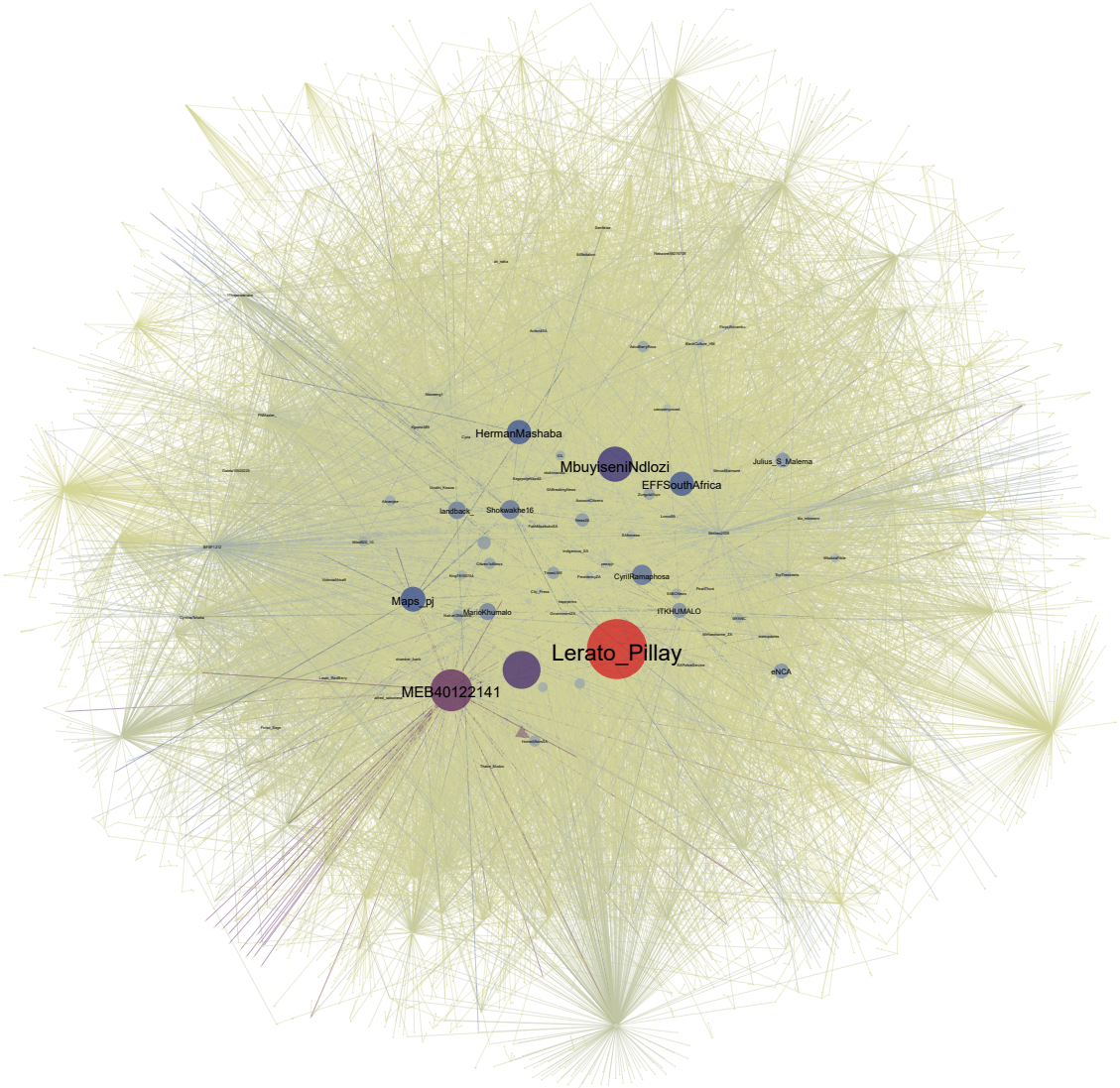
To identify key or "important" node's, I determine the centrality of a node relative to other nodes in a network.

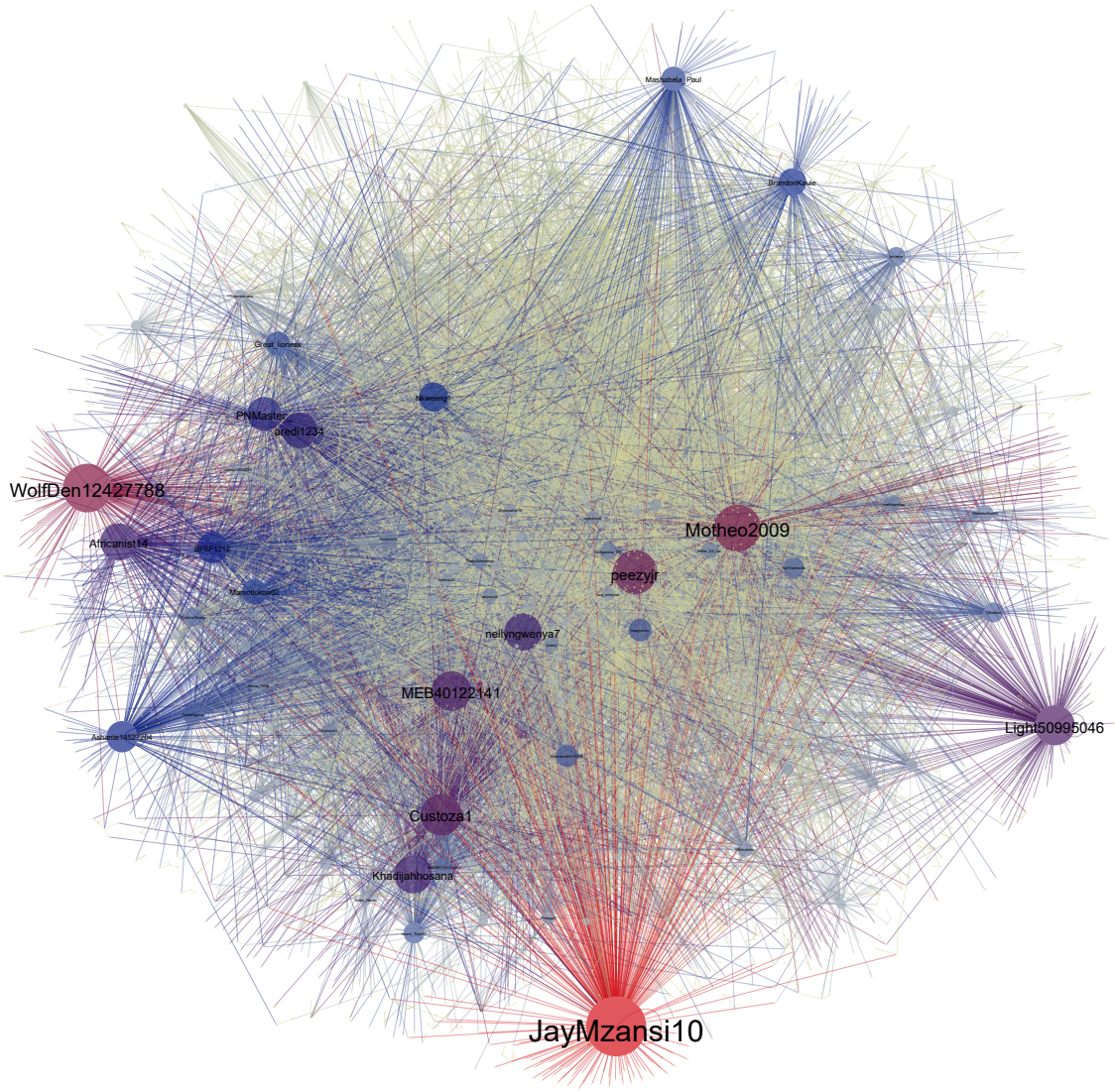
```
--------
Node Centrality: Summary Statistics
------------------------

          indegree    outdegree    closeness   betweenness      ␣
 ↪pagerank  \
count  7606.000000  7606.000000  7606.000000  7.606000e+03  7606.000000
mean      3.497370     3.497370     0.240375  8.325129e+03     0.000131
std      20.449728    15.967997     0.032569  6.255531e+04     0.001014
min       0.000000     0.000000     0.058300  0.000000e+00     0.000042
25%       0.000000     0.000000     0.220564  0.000000e+00     0.000042
50%       1.000000     1.000000     0.244526  0.000000e+00     0.000047
75%       2.000000     2.000000     0.258542  1.145871e+01     0.000078
max     889.000000   427.000000     0.354446  1.709460e+06     0.054873

       eccentricity
count   7606.000000
mean       7.627399
std        0.707994
```

```
min         6.000000
25%         7.000000
50%         8.000000
75%         8.000000
max        11.000000
```

Maximum Indegree:

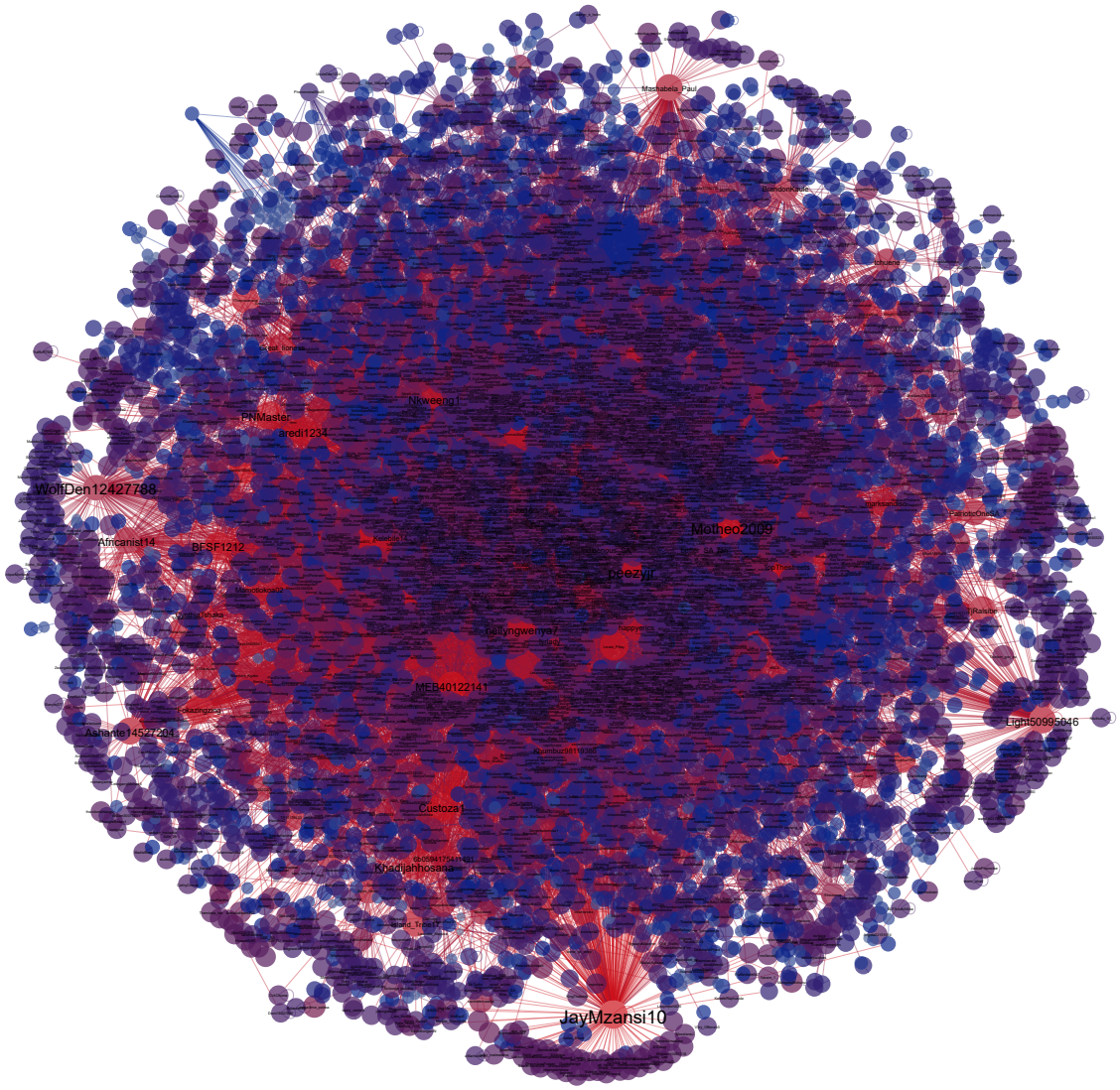
Lerato_Pillay





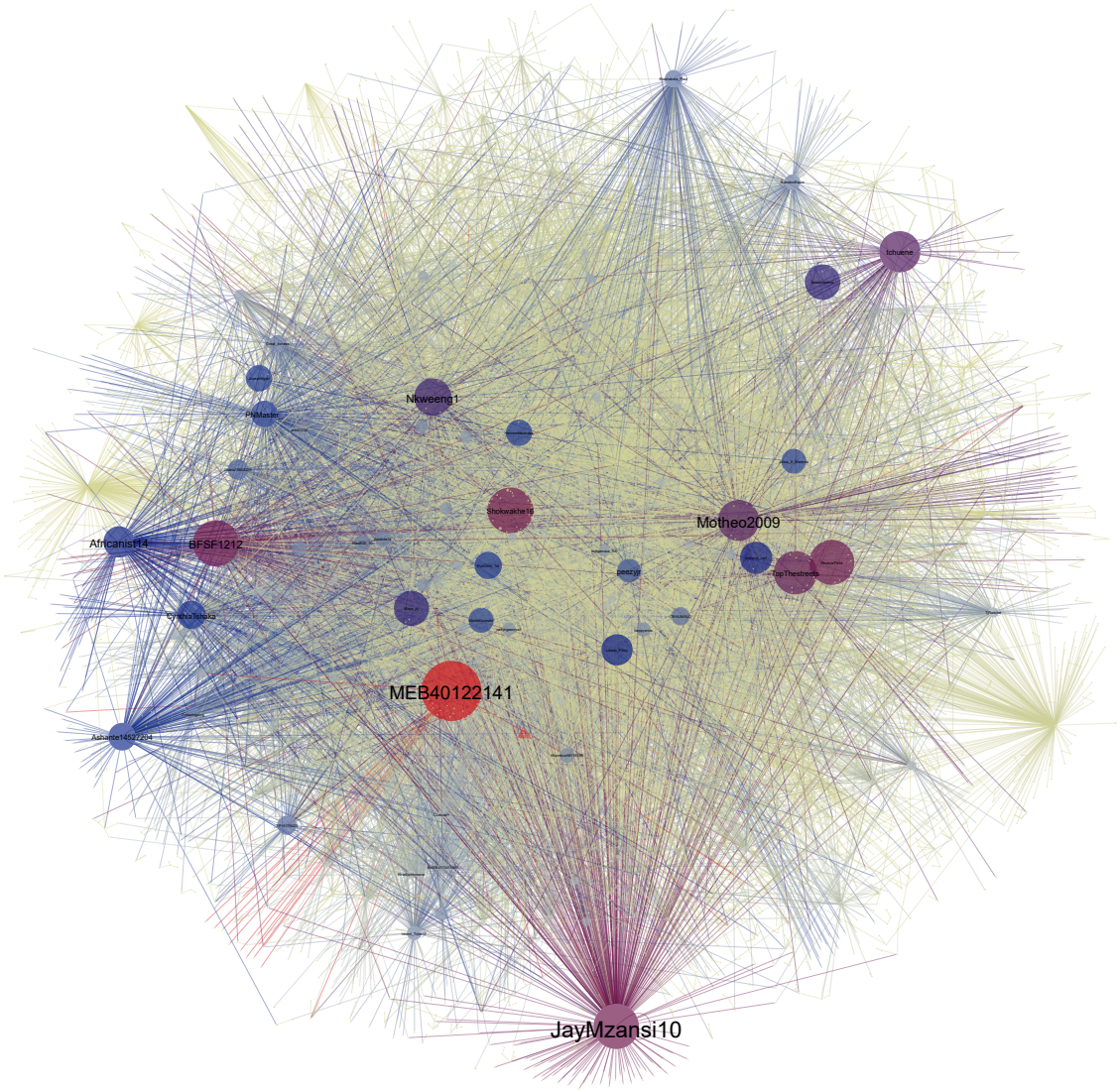
Maximum Closeness:

Motheo2009



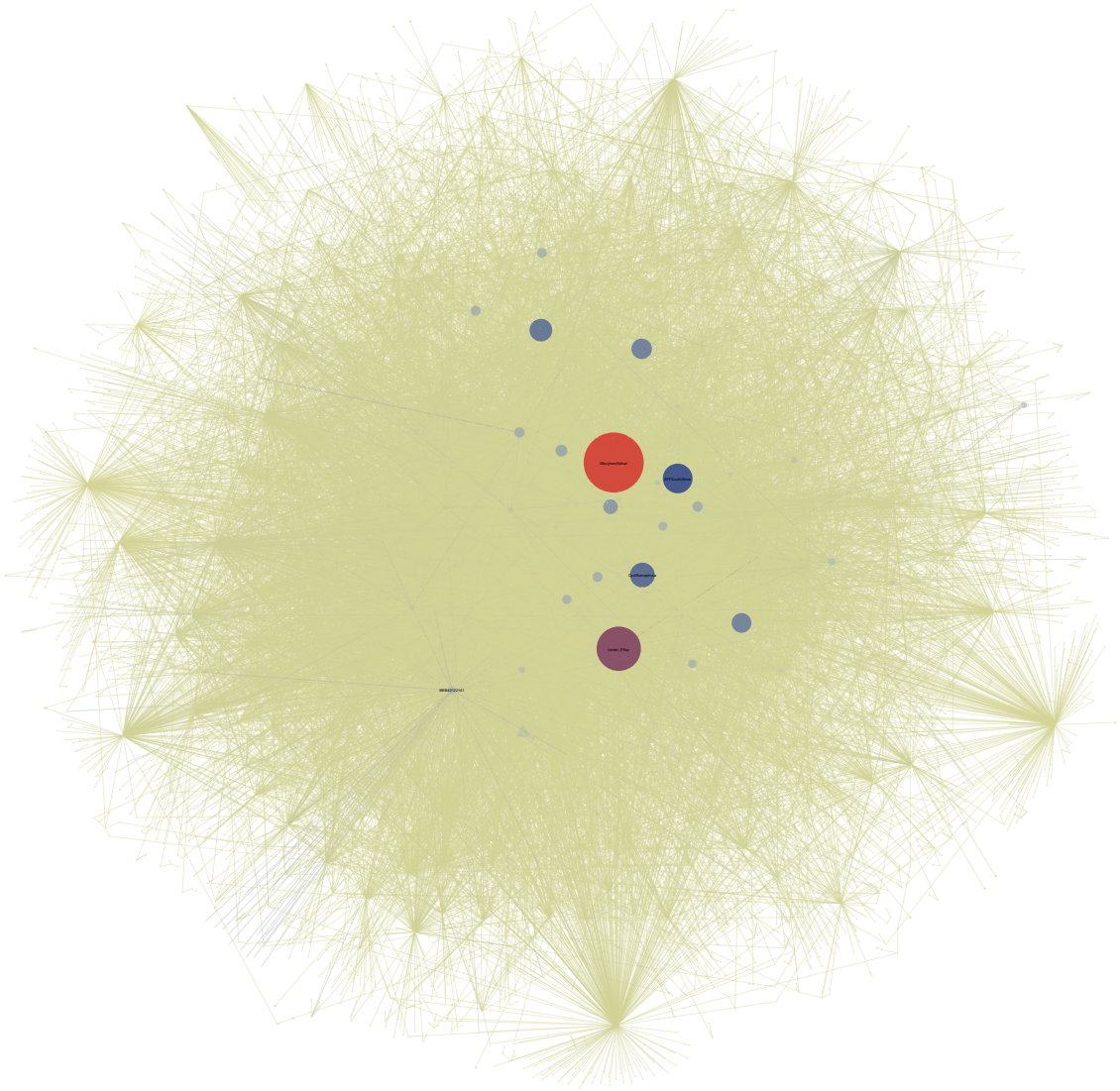
Maximum Betweeness:

MEB40122141



Maximum PageRank:

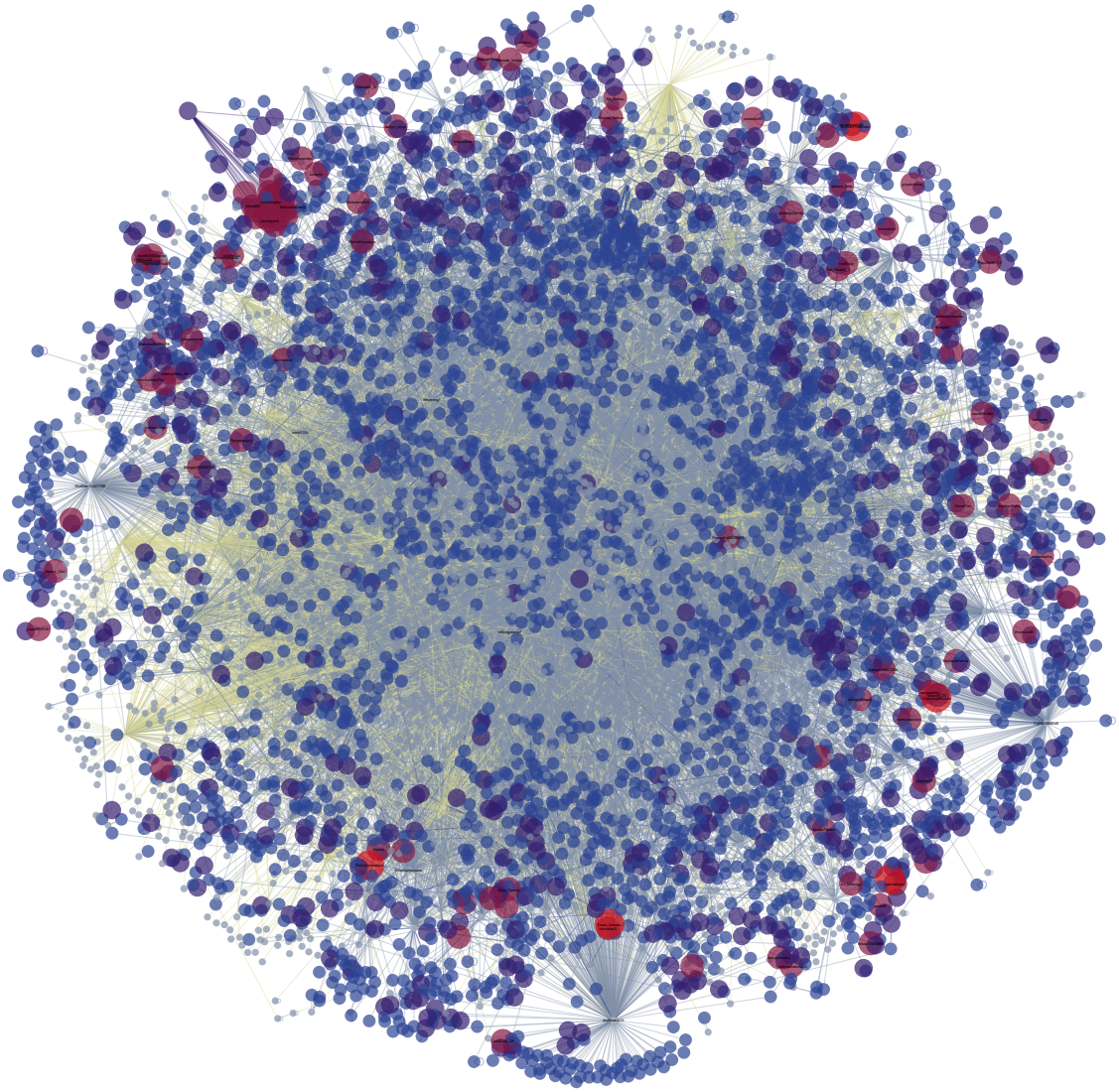
MbuyiseniNdlozi



Minimum Eccentricity:

	id_	label_
0	1309217630599999488	Ashante14527204
1	302713191	BFSF1212
2	313952950	peezyjr
3	14697575	News24
4	4692100098	mduduzzee_dube
..	...	...
89	1270404607773224966	NaN
90	1033132883811205127	NaN
91	1283039096135462916	mothol96901389

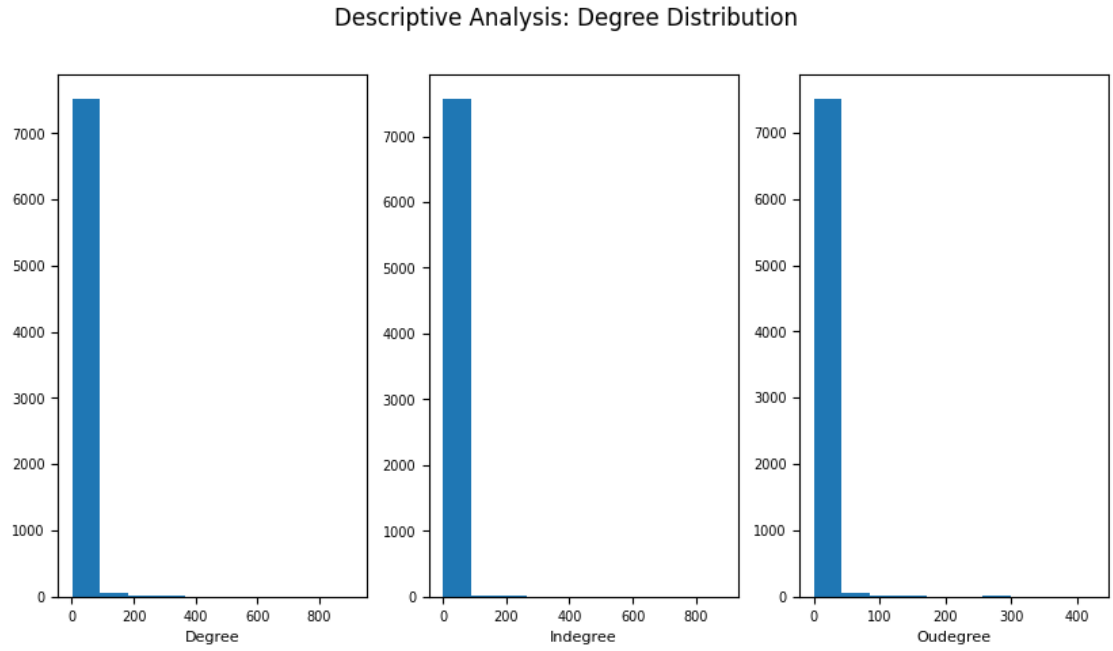
```
92   1257264937845604354                NaN
93   1312299508047523841  daise_scientifi

[94 rows x 2 columns]
```



1.2 Network Centrality

To determine relationships between nodes and their position in the network, I measure a node's ties relative to the ties present in the network and the distribution of ties throughout the network.

```
--------
Network Centrality: Summary Statitics
------------------------
```



Average Degree:

6.994740993952143

Density:

0.0003622006433551975

Average Path Length:

5.264935643592867

---

2 CONNECTIVITY AND COHESION

To examine how tightly connected or clustered the network is, I examine the direction, frequency and consistency of relations between nodes and the nodes in their neighbourhood.

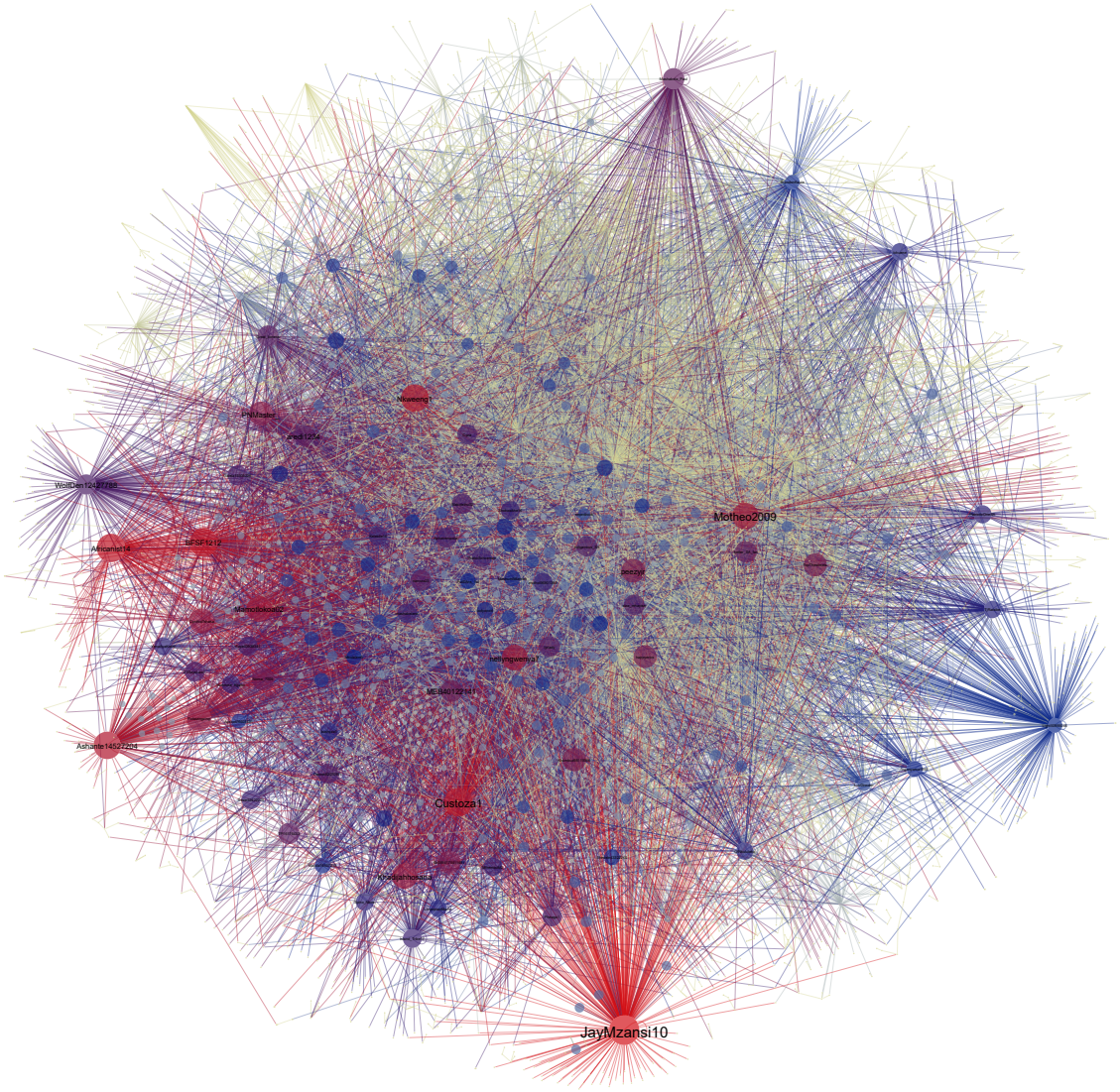
Connectivity and Cohesion: Summary Statistics

---

	reciprocity	transitivity	hierarchy
count	7.606000e+03	7606.000000	7606.000000
mean	8.938600e+04	1597.596108	29026.853800
std	3.200461e+05	3394.844552	71379.822977
min	0.000000e+00	0.000000	0.000000
25%	0.000000e+00	0.000000	0.000000
50%	0.000000e+00	0.000000	0.000000
75%	4.080000e+03	497.250000	5367.000000
max	3.869550e+06	13610.000000	444240.000000

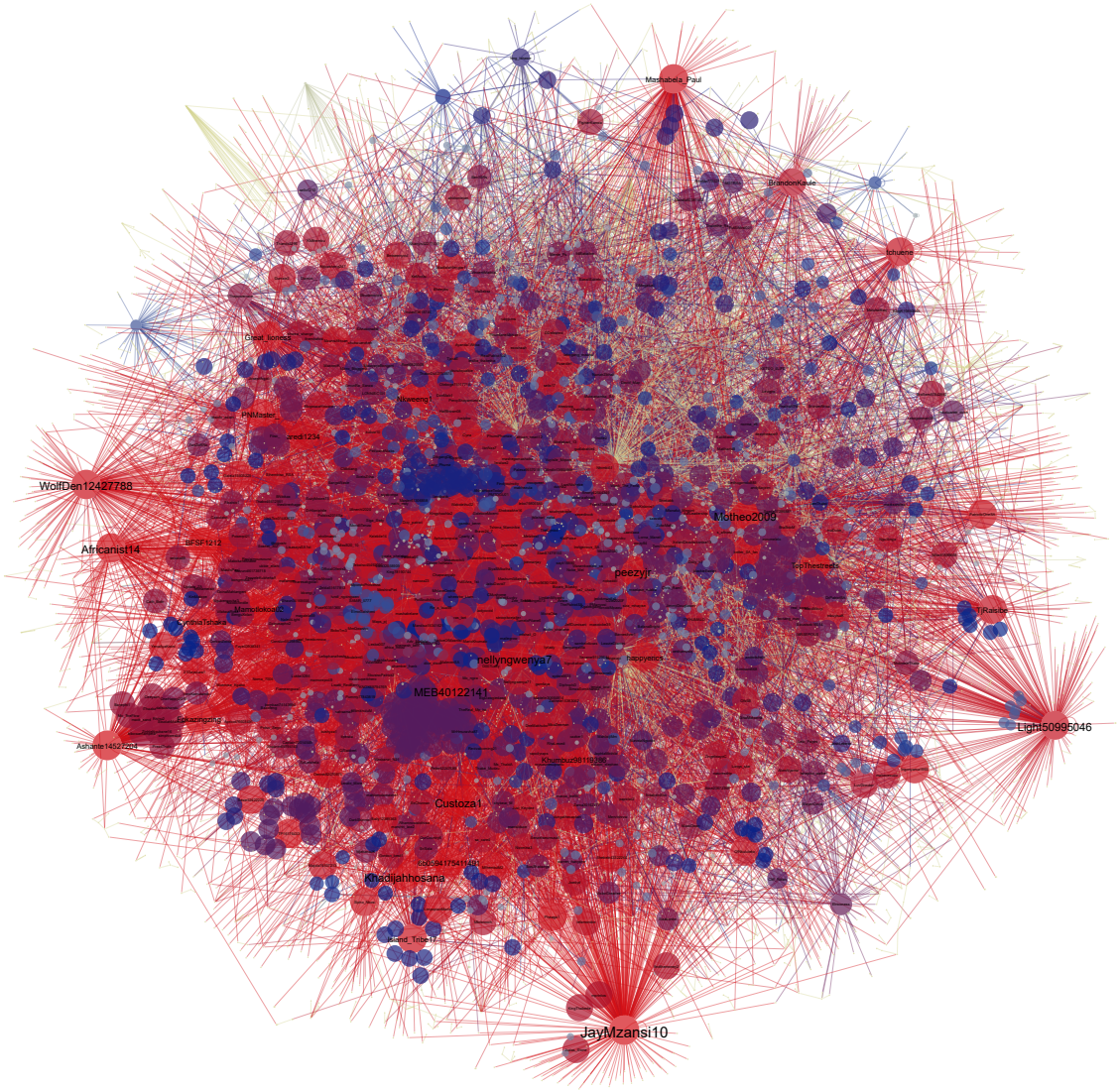
Maximum Reciprocity:

JayMzansi10



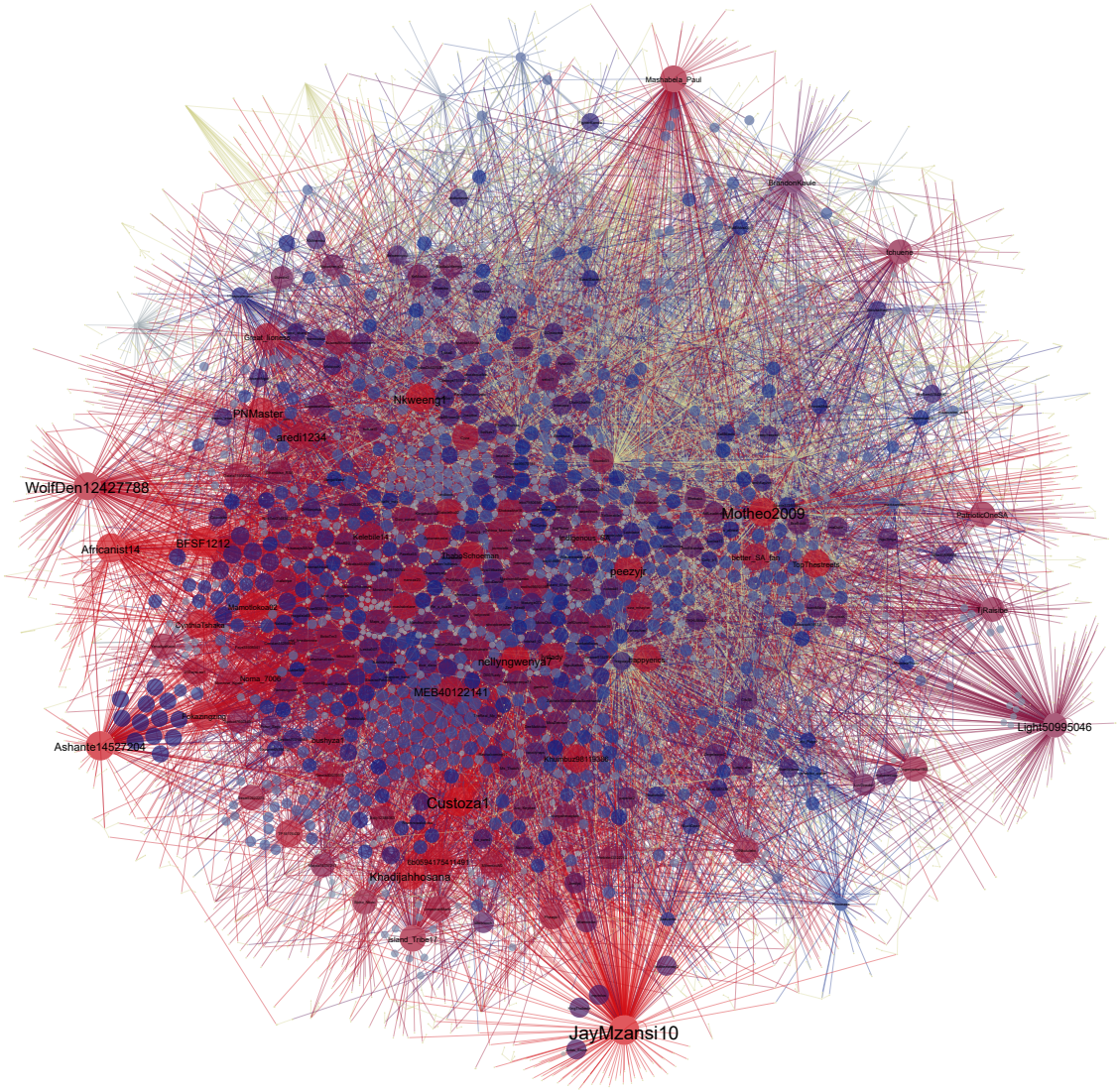
Maximum Transitivity:

JayMzansi10



Maximum Hierarchy:

JayMzansi10



Average Clustering Coefficient:

```
0.12535597226951034
```

---

3 EVALUATION

To determine if the current network exhibits small world properties, I simulate multiple random networks and compare the average distances and clustering features of either network.

```
------------------------
Random Network:
------------------------

--------
Average Path Length:

 8.679281082242662


--------
Average Clustering Coefficient:

 0.0007513296495444674
```

---

4 SUMMARY

In this section, I perform descriptive analysis on the #PutSAFirst twitter reply network to determine certain node features and network properties that identify significant players in the discourse. I summarise a few findings below.

In addition to skewed or unequal degree distributions, which indicates a tendency of preferential attachment toward popular users, this network exhibits high clustering (163.7x higher) and small average distances (1.6x smaller) than randomly simulated networks of the same size. Consistent with the impression of small worlds.

Twitter accounts found to rank highly on certain network measures are listed below:

- **Lerato_Pillay** (Indegree)
- **Motheo2009** (Closeness)
- **MEB40122141** (Betweenness)
- **MbuyiseniNdlozi** (PageRank)
- **JayMzansi10** (Outdegree, Reciprocity, Transitivity, Hierarchy)

I find **Lerato_Pillay**, a copy account of the original *@uLerato_Pillay* handle, to share a similar notoriety (*indegree*) as the previous account within the discourse. Currently the user behind this account has been suspended off twitter, likely due to the racially discriminitary understones of their tweets such as:

> "#ForeignersMustleaveSA nothing else will stop us. This #SayNoToXeno is nonsense, those who feel South Africans are Xenophobic they must Voetsek back to their Countries"
>
> "#Hillbrow Even Fikile Mbalula is aware that we have a problem of foreigners. Especially Zimbabweans who come here to kill and take… #ForeignersMustGo"

Given **JayMzansi10**'s high rank in measures of *reciprocity*, *transitivity* and *hierarchy*, it is likely the case that users forming part of **JayMzansi10**'s neighbourhood tend to (strictly) communicate with members of

the same neighbourhood. These discussions are potentially “echo chamber’s” where opposite or contradicting views never enter the discussion or are easily dismissed. Below is an example of their most retweeted tweet.

Another account that appears to be close to the conversations of #PutSAFirst is **Motheo2009** who tweets:

← Thread

MOtheoWaMOtheo
@Motheo2009

Alex residents are currently busy destroying shacks owned by Zimbabwean illegal immigrants who are behind the spate of violent crimes in the area. @landback_ @victoriaafrica8 @PutSACitizens1 @Lerato_Pillay @mudzu_thabe @Maps_pj #WeWantOurCountryBack #PutSouthAficansFirst

6:06 PM · Jan 3, 2021 · Twitter for Android

101 Retweets 23 Quote Tweets 209 Likes

← Tweet

MOtheoWaMOtheo
@Motheo2009

Replying to @bonglez

President Mbeki was right Nigerians are diarrhoea to the world. #HumanTrafficking #23SeptemberCleanSA #humantraffickingawareness #Checkpoint #ForeignersMustGo #PutSouthAfricansFirst

9:56 AM · Sep 23, 2020 · Twitter for Android

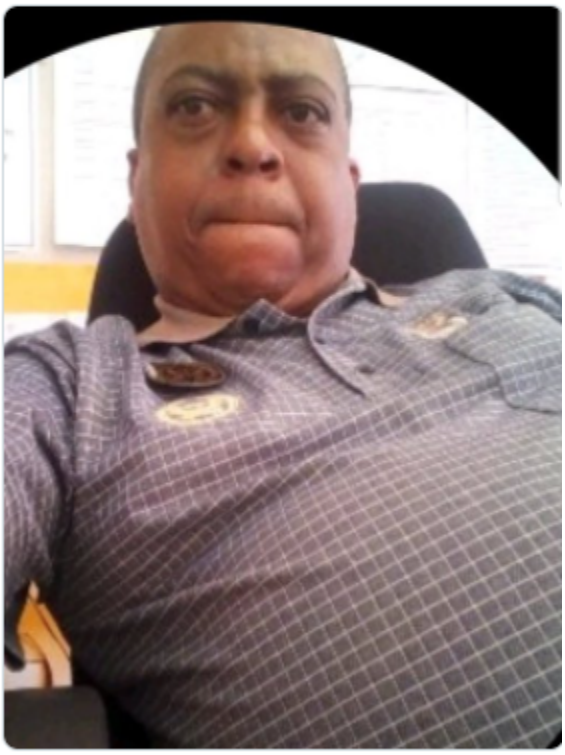
213 Retweets 15 Quote Tweets 467 Likes

Finally, I find the user **MEB40122141**, who appears to be running his account as the official #PutSAFirst page, to have the widest reach (*betweeness*) in the network. They also overtake **Lerato_Pillay** in having the highest *degree* measure for each month after November 2020 (see dynamic network below). Considering their high *betweenness*, high *indegree* and high *outdegree*, it is fair to assume that this player has some influence in the network.

← Thread

PUTSAFIRST 👁 🔥 💯
@MEB40122141

Meet Jock Kennedy...senior immigration official who runs between Ellisras...Thabazimbi and Vryheid to assist Pakistanis to obtain their visas and Permanent Residency permits with fraudulent papers 😡let's make him famous😡
@HomeAffairsSA
#WeWantOurCountryBack



1:02 PM · Apr 21, 2021 · Twitter for Android

323 Retweets 18 Quote Tweets 241 Likes

In the next section, I explore the network further by using community detection methods to identify groups (communities) of users that are densely connected and have high levels of interaction within the discourse.