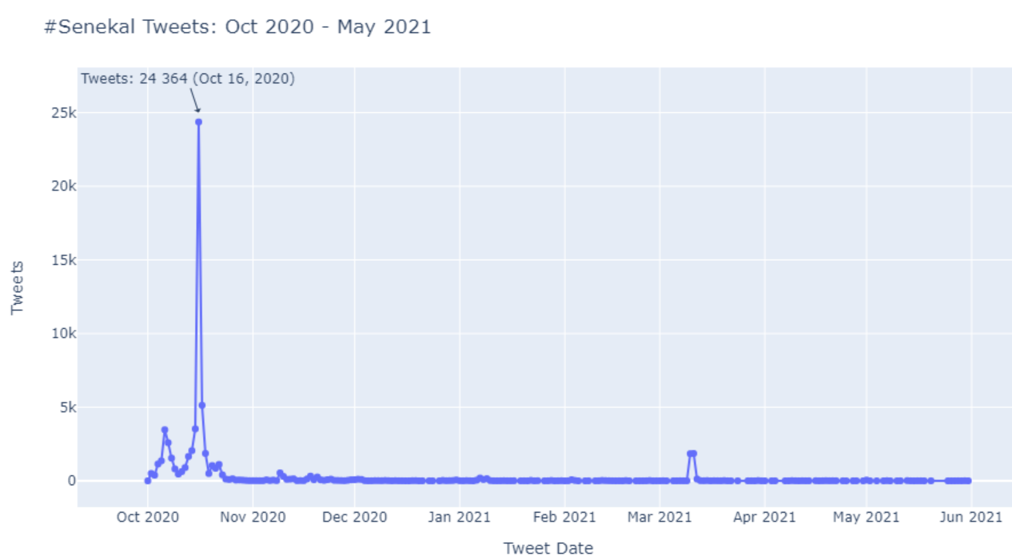
# PART I: SENEKAL
## DESCRIPTIVE ANALYSIS

The #Senekal protests erupted in the small town of Senekal, the magistracy of a rural farming district in eastern Free State, after the brutal murder of Brendin Horner, a white farm manager, whose body was found on 2 October 2020. His murder quickly fed into a pre-existing discourse about the targeted killing of white farmers by people of colour, and a perceived lack of government intervention. AfriForum, a lobby group that mobilises around white Afrikaner interests, quickly labeled Horner's murder, which came to stand as an exemplar of farm murders in general, as "an act of terrorism."

### Data Source

Data is extracted from Twitter's API matching the following search query: "Senekal eff", "Brendin Horner" and "senekal protest", for the period: 1 October 2020 to 31 May 2021.

```
--------
Number of Daily Tweets in #Senekal 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 #Senekal full network are still described in the the #Senekal 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 #Senekal network.
  - *#retweeted* -> A measure of the total number of tweets by a user that are retweeted by other users in the #Senekal network.
  - *#liked* -> A measure of the total number of tweets by a user that are liked by other users in the #Senekal network.
- **Derived edge variables**
  - *weight* -> This captures the total number of replies from a sender to a receiver in the #Senekal network. For example, replying to a user only once gives a weight = 1, but replying to a user 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 #Senekal Full Network.

- **#Senekal Full Network**

  The #Senekal full network consists of 65488 tweets (incl. retweets, and replies to tweets) with 24154 active users behind those tweets.
  - Number of users: 24154
  - Number of tweets: 65488

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

- **#Senekal Reply Network**

  The #Senekal 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: 3432
  - Number of replies: 4317

Next, I exclude all disconnected components and focus my analysis on the #Senekal 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 $B$.

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

  The #Senekal reply network has 524 connected components. In this network, the largest connected component has 2207 nodes and 3011 edges, and the second largest connected component has 14 nodes and 16 edges.
  - Number of users: 2207

- Number of replies: 3285

**Method**

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

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

IGRAPH DNW- 2207 3011
```

---

# 1 CENTRALITY

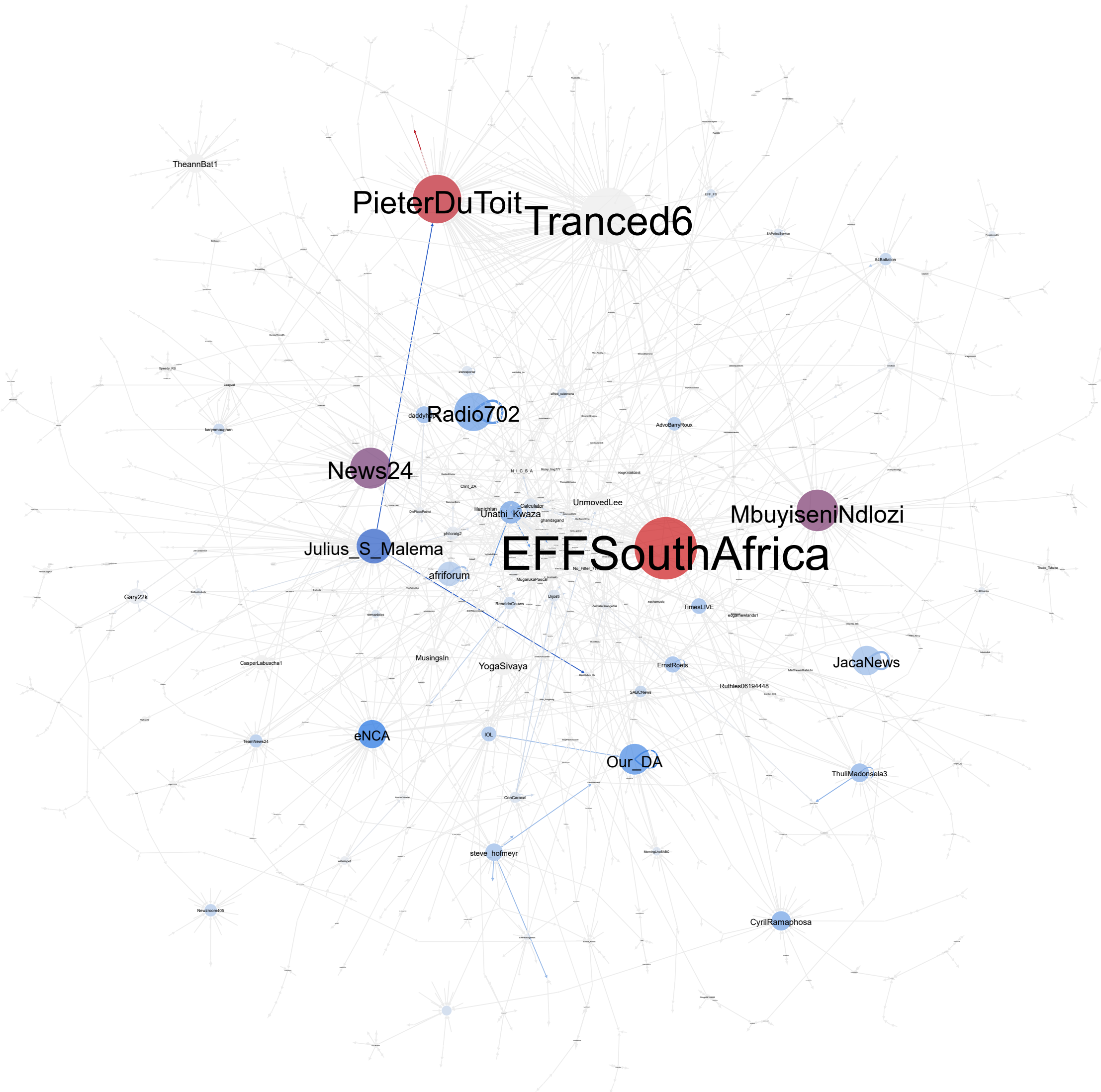
## 1.1 Node Centrality

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  2207.000000  2207.000000  2207.000000  2207.000000  2207.000000
mean      1.488446     1.488446     0.157815    64.975189     0.000453
std       4.566893     2.855500     0.026996   488.335670     0.001296
min       0.000000     0.000000     0.074709     0.000000     0.000156
25%       0.000000     1.000000     0.140242     0.000000     0.000156
50%       1.000000     1.000000     0.157865     0.000000     0.000194
75%       1.000000     1.000000     0.176128     1.000000     0.000347
max      99.000000    91.000000     0.252403  8503.500000     0.033209

       eccentricity
count   2207.000000
mean      12.830086
std        1.314416
min       10.000000
25%       12.000000
50%       13.000000
75%       13.000000
max       19.000000

--------
```

**Maximum Indegree:**

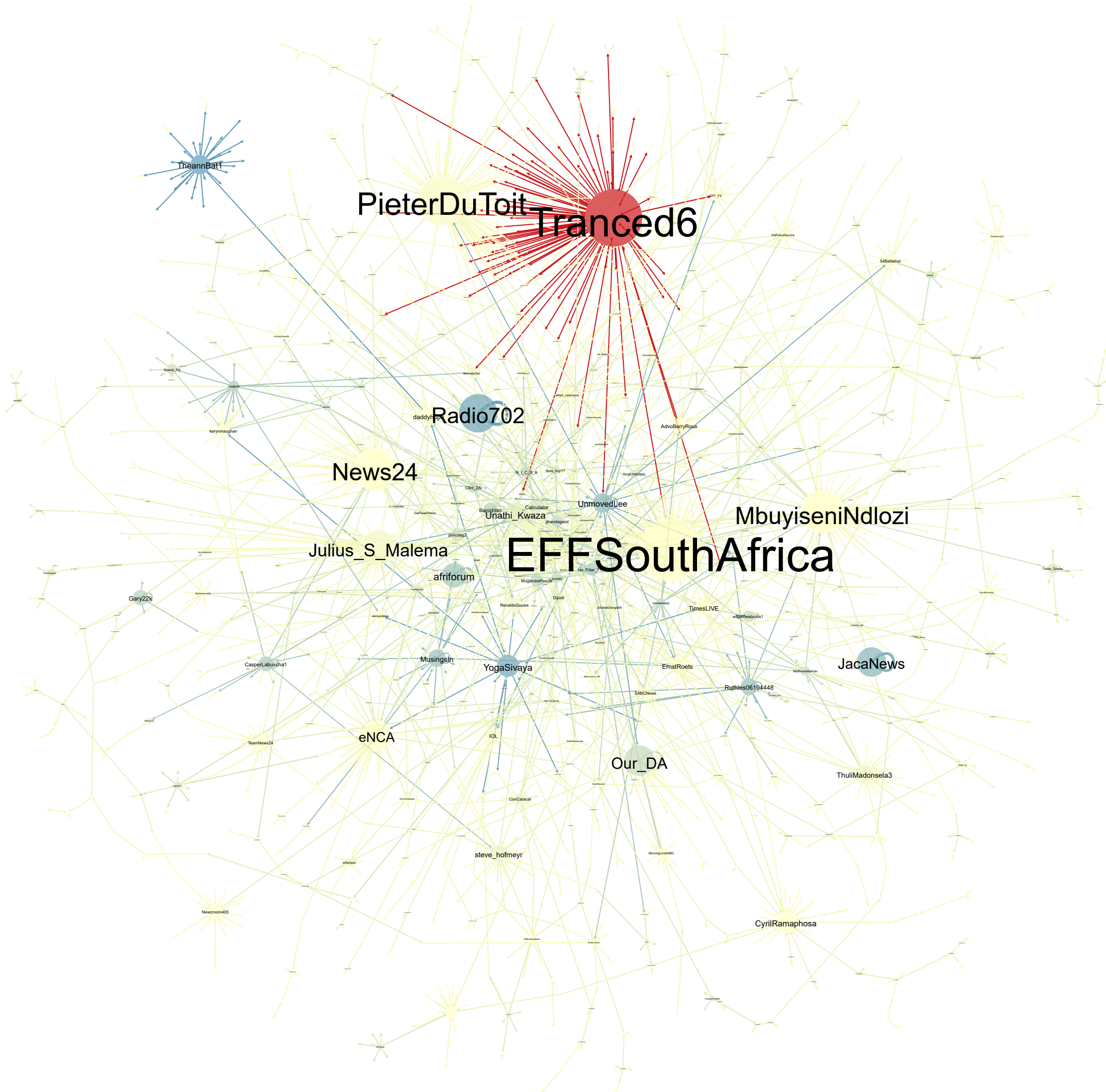
EFFSouthAfrica




--------

**Maximum Outdegree:**

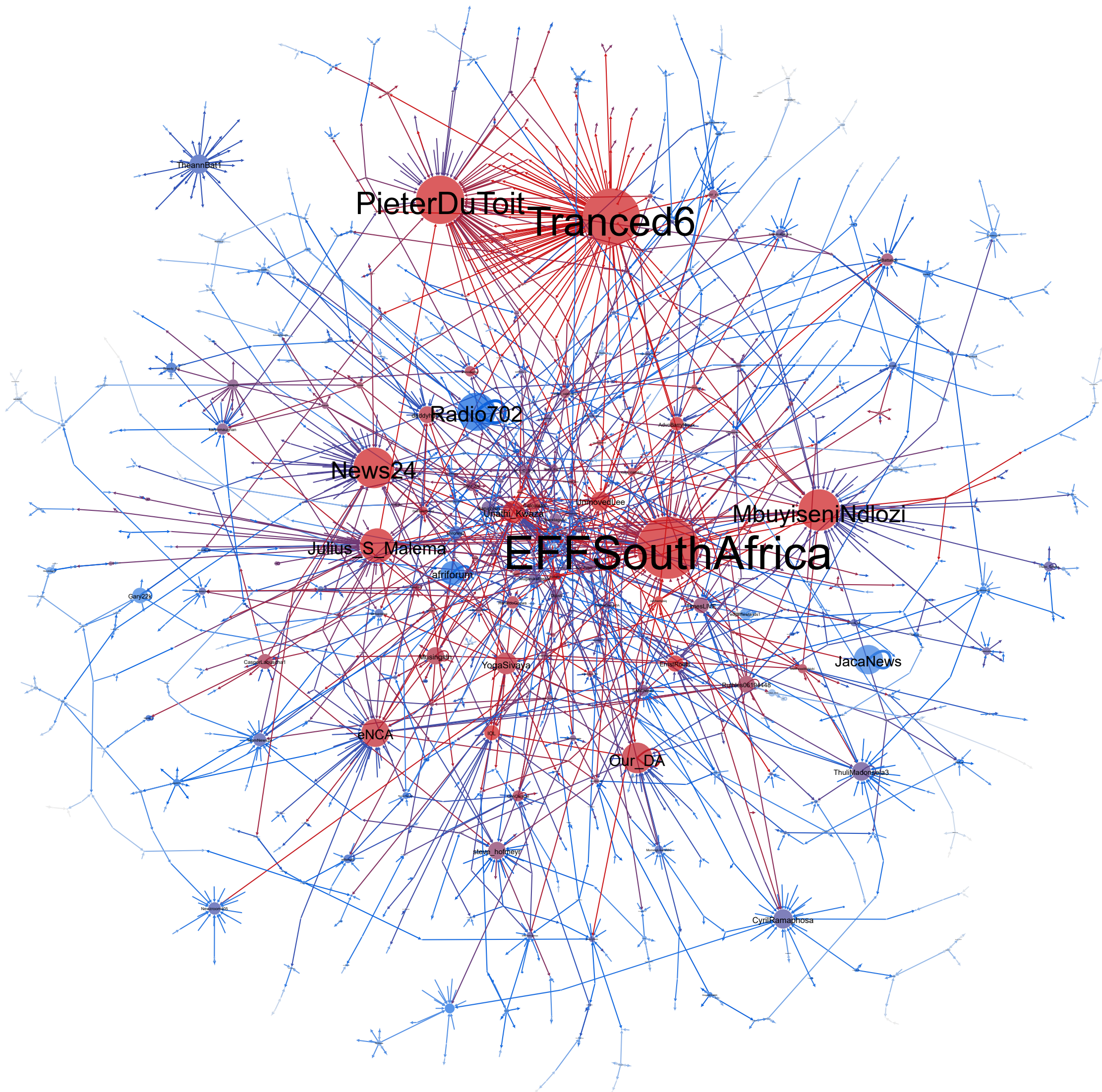
Tranced6




--------

**Maximum Closeness:**

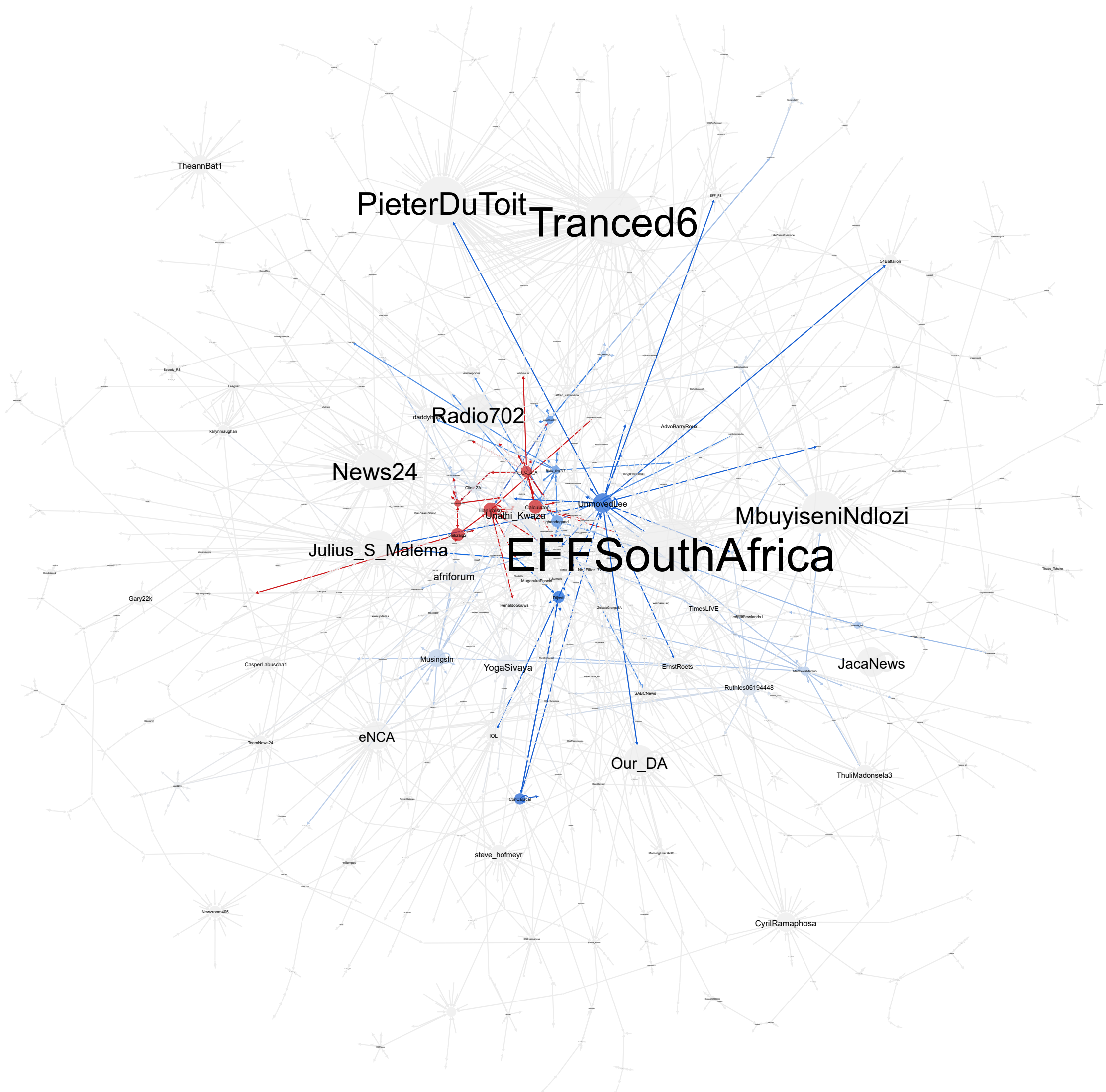
EFFSouthAfrica




--------

**Maximum Betweeness:**

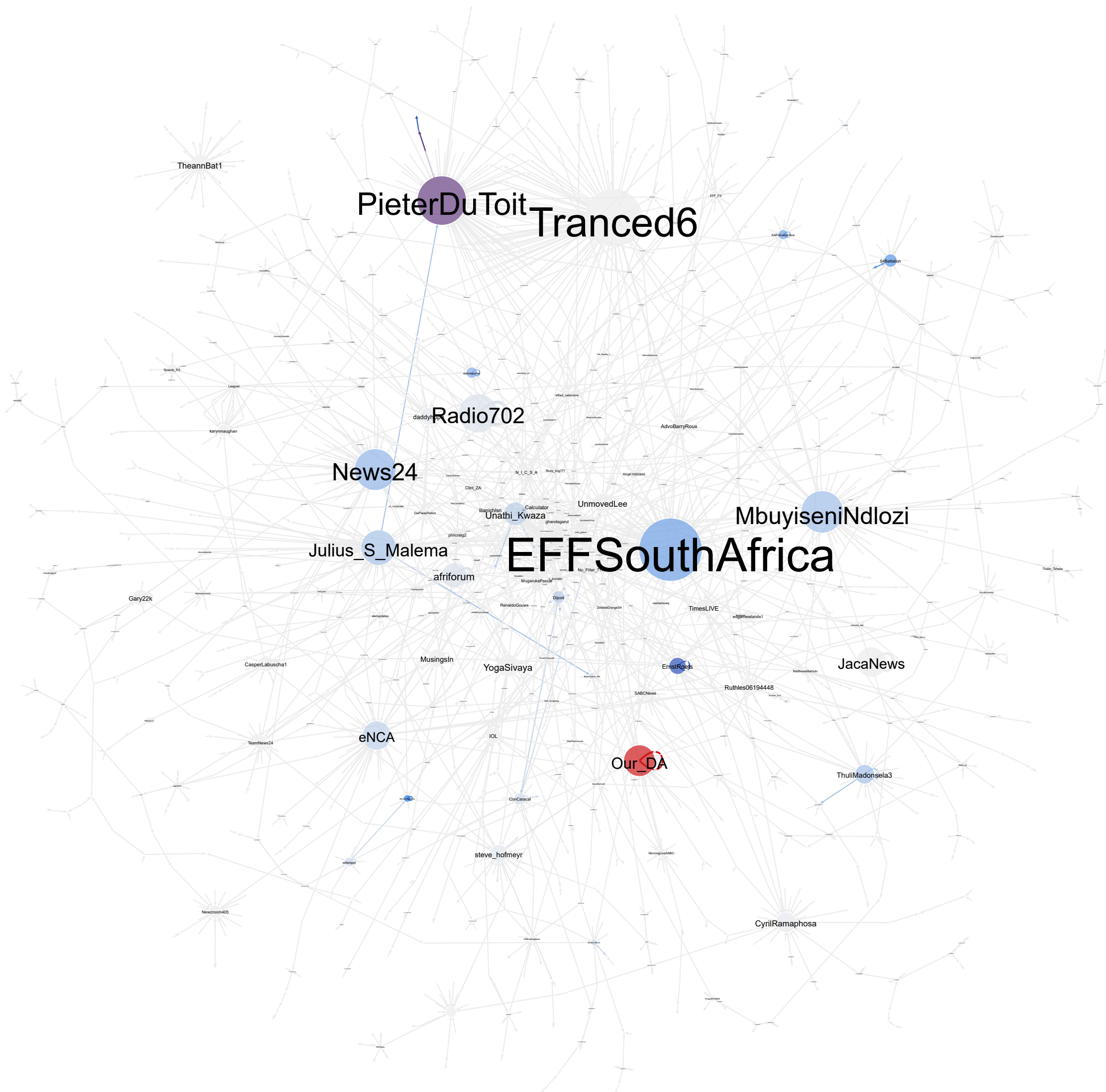
N_I_C_S_A




--------

**Maximum PageRank:**

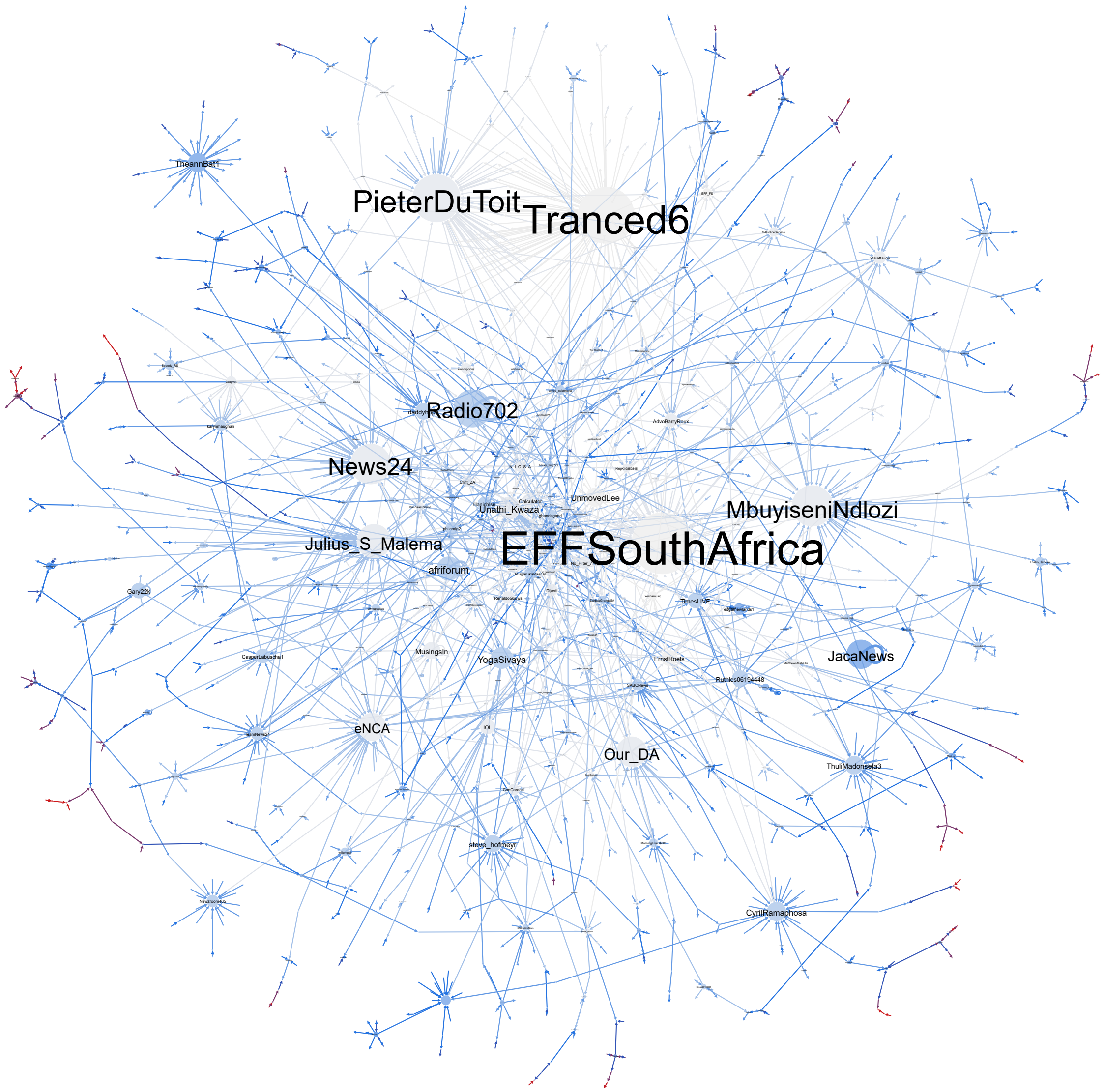
Our_DA




--------

**Minimum Eccentricity:**

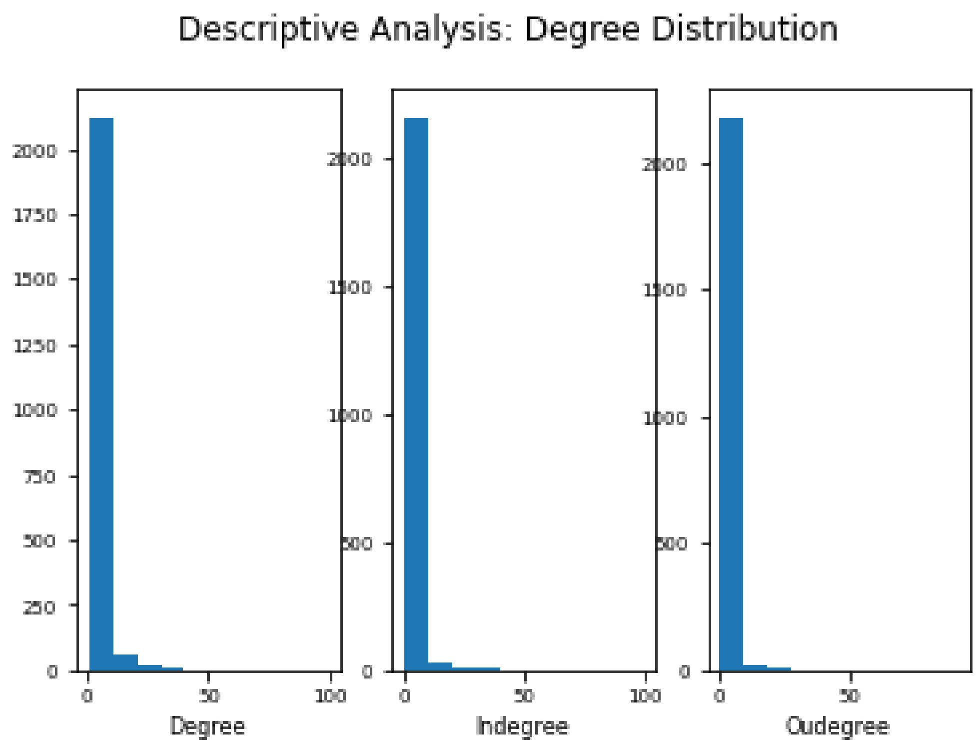
| | label |
|---|---|
| 0 | EFFSouthAfrica |
| 1 | UnmovedLee |
| 2 | ErnstRoets |
| 3 | Tranced6 |
| 4 | citrusramaphosa |
| 5 | EFF_FS |
| 6 | xeshamusiq |




## 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:**

```
 2.976891708201178
```

--------
**Density:**

```
 0.0006184476081831443
```

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

```
 6.034344711335862
```

---

# 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.

```
--------
```

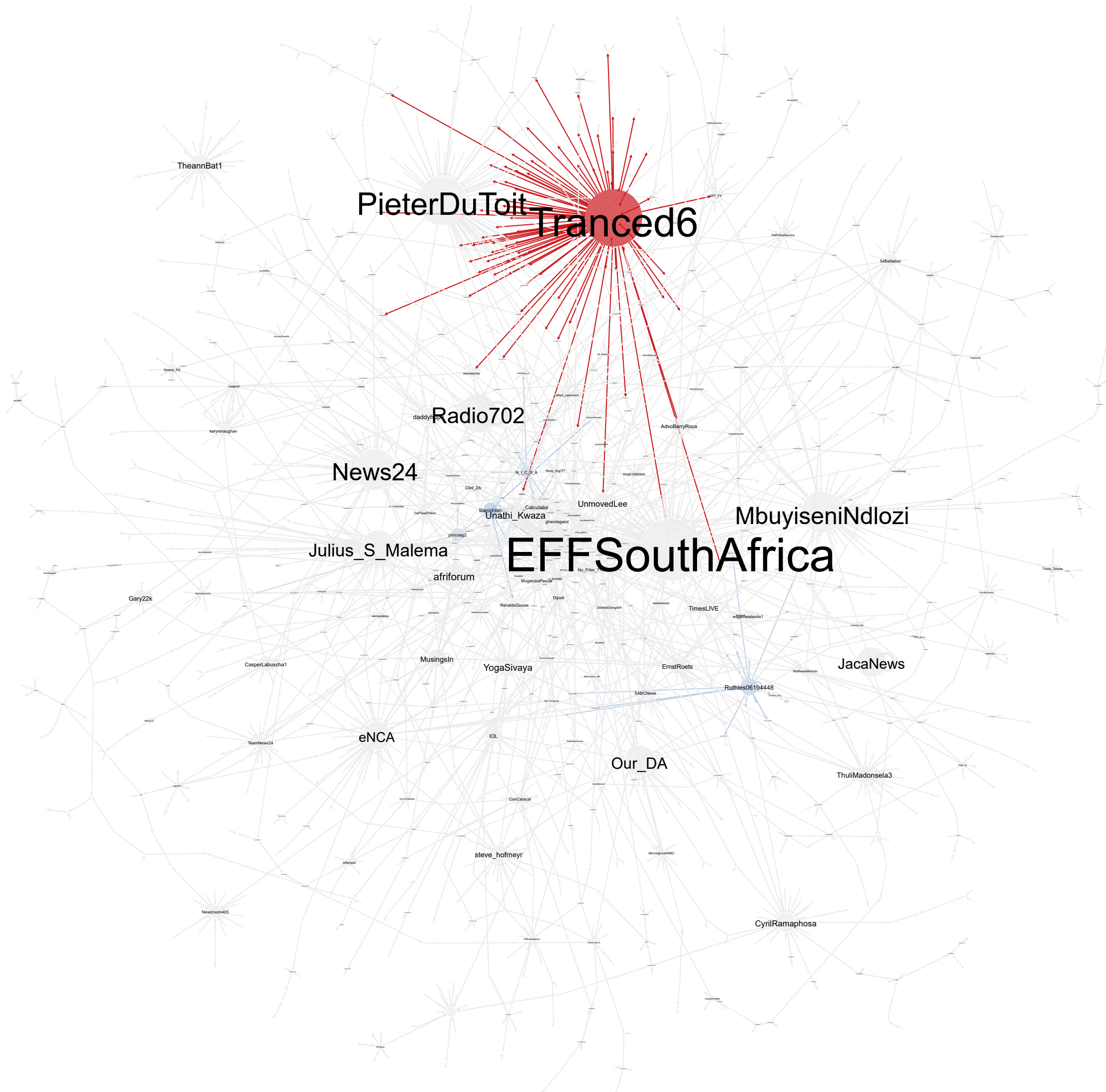
**Connectivity and Cohesion: Summary Statistics**

```
------------------------

       reciprocity  transitivity    hierarchy
count  2207.000000   2207.000000  2207.000000
mean     16.019937      0.043951     7.169461
std     110.518569      0.360528   117.516320
min       0.000000      0.000000     0.000000
25%       0.000000      0.000000     0.000000
50%       0.000000      0.000000     0.000000
75%       2.000000      0.000000     1.000000
max    4463.000000     10.000000  5441.000000
```

```
--------
```

**Maximum Reciprocity:**

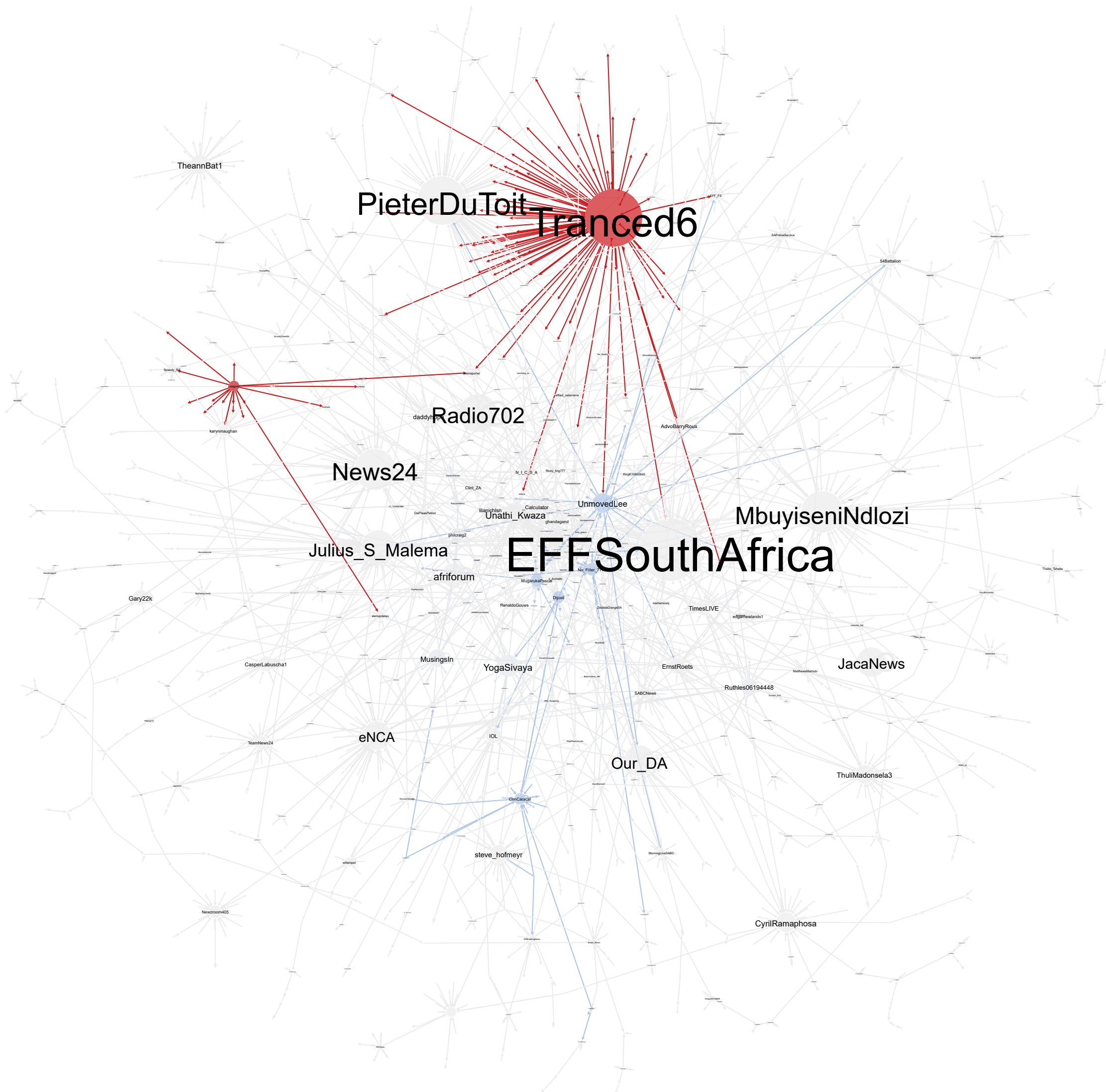
```
Tranced6
```




--------

**Maximum Transitivity:**

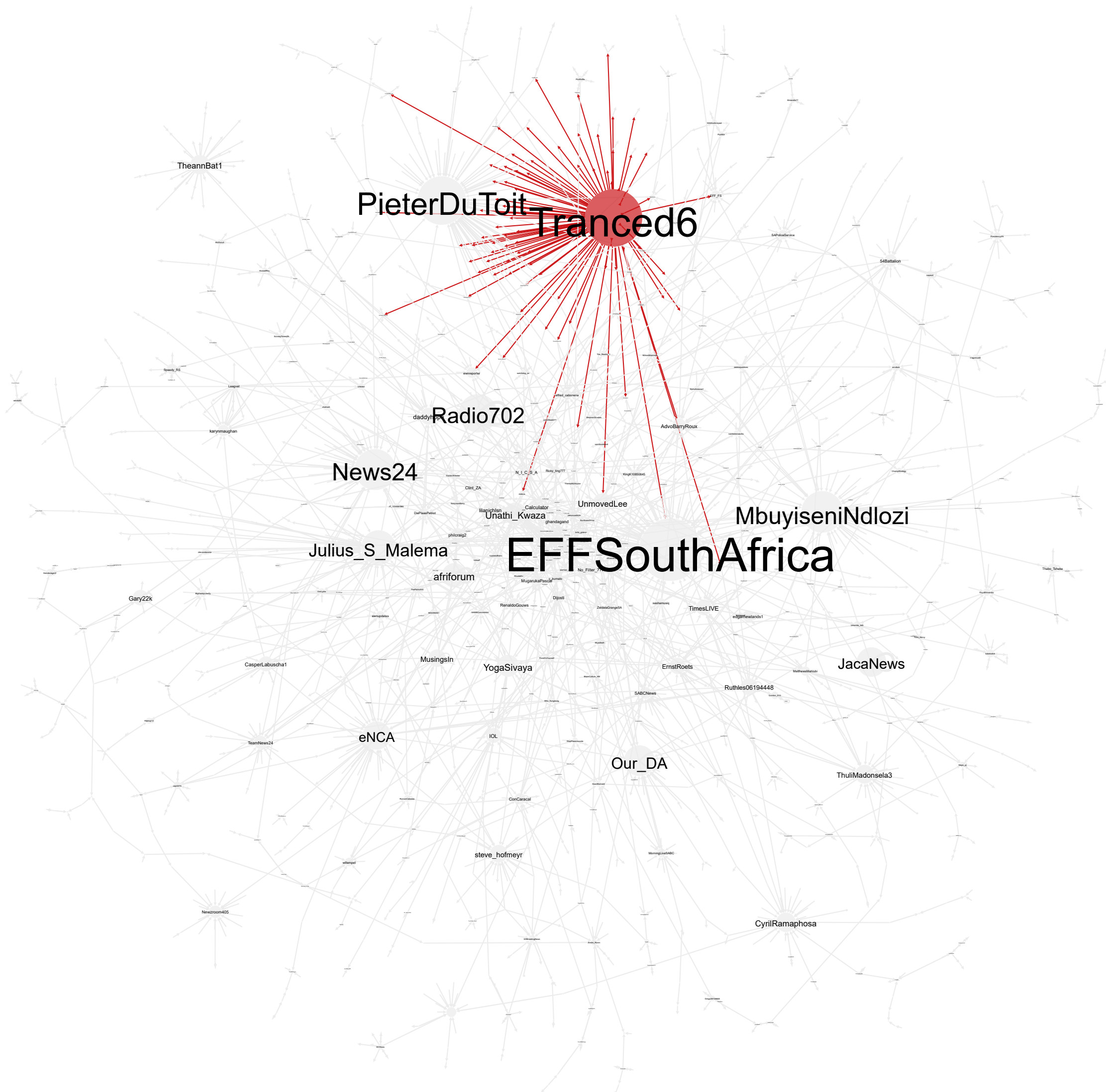
Laagvat




--------

**Maximum Hierarchy:**

Tranced6




--------

**Average Clustering Coefficient:**

```
0.025795904313428337
```

---

# 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:

 17.40918285362087


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

 0.0011544533569588272
```

---

# 4 SUMMARY

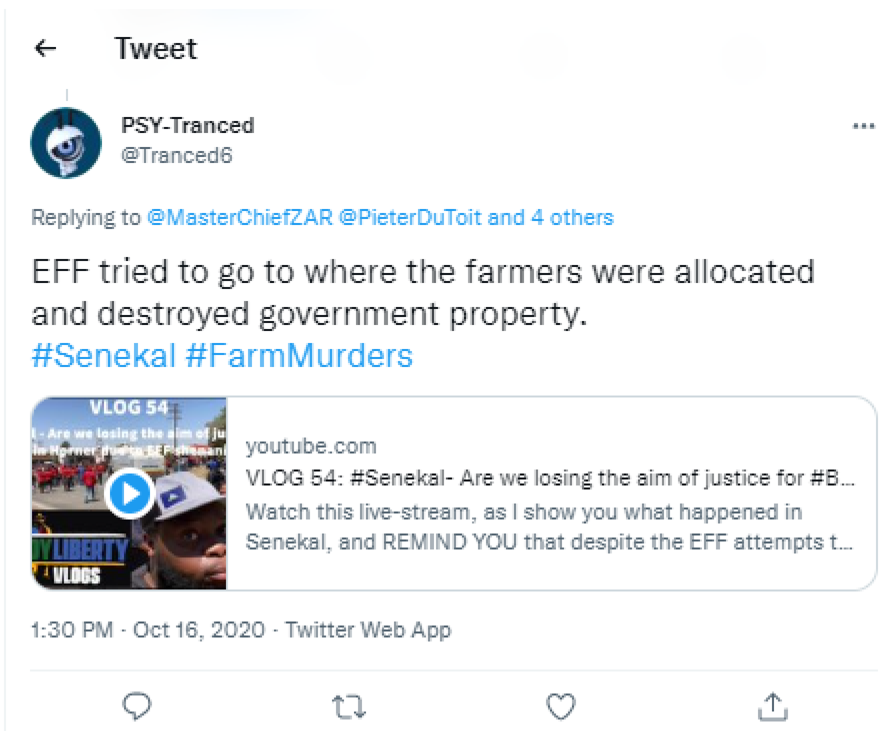
In this section, I perform descriptive analysis on the #Senekal twitter reply network to determine certain node features and network properties that identify significant players in the discourse. I summarise a few findings below.

In addition to skewed or unequal degree distributions, which indicates a tendency of preferential attachment toward popular users, this network exhibits high clustering (20.9x higher) and small average distances (2.9x 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:

- **EFFSouthAfrica** (Indegree, Closeness)
- **Our_DA** (PageRank)
- **N_I_C_S_A** (Betweenness)
- **Tranced6** (Outdegree, Reciprocity, Hierarchy)
- **Laagvat** (Transitivity)

Given **Tranced6**'s high rank in measures of *outdegree*, *reciprocity*, *transitivity*, and *hierarchy*, it is likely the case that users forming part of **Tranced6**'s neighbourhood strongly prefer communicating with members of the same neighbourhood as compared to others. Usually this is a sign that discussions are potentially "echo chamber's" where opposite or contradicting views never enter the discussion or are easily dismissed. Surprisingly, upon further inspection into the real targets of **Tranced6**'s replies, the theory of the "echo chamber" weakens. This is because majority of replies are in response to tweets sent by users with opposing views, aiming to counter and discredit them. Replies to these users are often a copy and paste of the same message: "EFF tried to go to where the farmers were allocated and destroyed government property." Still, it is plausible that although **Tranced6**'s replies target tweets outside their views, other users in their neighbourhood may choose to only communicate in a close knit clique or "echo chamber" where similar messaging in support of **Tranced6**'s tweets are circulated.


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.