PART III: BRACKENFELL
DIFFUSION ANALYSIS

Diffusion analysis is the study of how interpersonal ties amplify the spread of a product, idea, or innovation such that a change in one person's behaviour can cascade throughout a social network and produce a behavioural change in other people, communities, and populations.

Data Source

#Brackenfell Reply Network - Largest Component

Data Transformation

Using the timestamp of when a tweet is created, I define the interaction point for node A and node B to be the point in time when either node A first replies to node B or vice versa. Any time before that point, a connection between node A and node B does not exist, and any time after that, a connection between node A and node B exists. I classify an undirected dynamic network where an "infection" can be transmitted in either direction and the choice to adopt is faced by either one (sender or receiver). Therefore, at any point when node A replies to node B, node A is aware of node B's message, and at the same time, by directly replying to node B, node B is made aware of node A's message. Once exposed, both are susceptible to "infection" and their decision to adopt is a combined function of their *threshold* to resist peer pressure and their personal *profile* or preference to only adopt when their interests are aligned.

Default Parameters

Threshold $(\tau) = 60\%$

Profile $(\gamma) = 25\%$

For each simulation, I assume a passive threshold of 60% and active threshold of 25%. This means for any node to be susceptible, 60% of its neighbours should already be infected. Once this condition is met, a random number is drawn, if that number is above 25%, a node becomes infected. Initial adopters or seeds are selected from the top of each role list. Given the dynamic process, users eligible for seed selection are constrained by the date and time of their first tweet reply. An upper limit of 6 initial seeds is chosen for each simulation. Which means, depending on the diffusion period, each simulation can have up to 6 users as initial adopters.

Method

Using pythons ndlib and dynetx library, I present the dynamic **profile threshold model** and determine the diffusion process of discourse under varying seeds of adoption.


Number of Daily Tweet Replies in #Brackenfell Discourse:

#Brackenfell Interactions (User Replies): Oct 2020 - May 2021
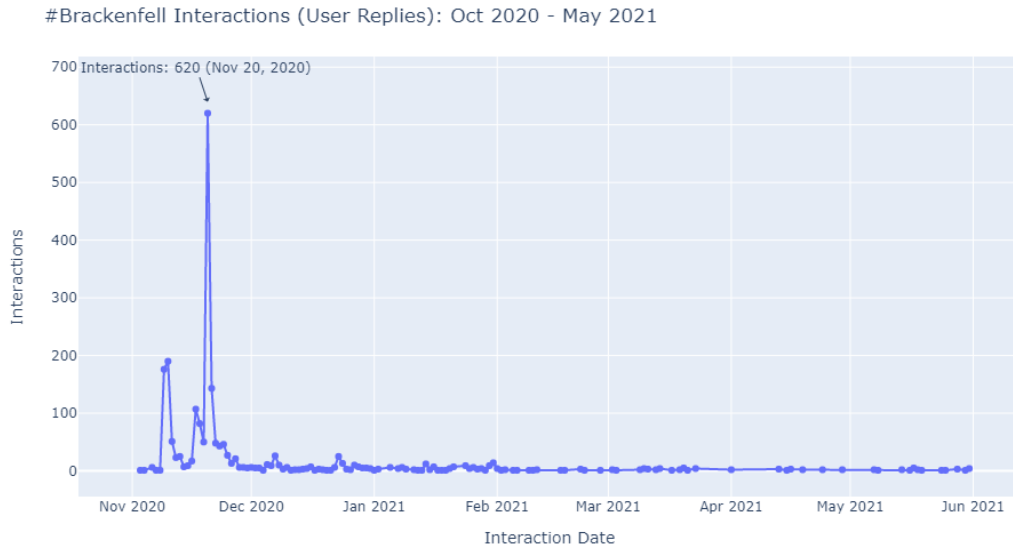
1 COMMUNITY DIFFUSION

1.1 Cohesion in Community

For each community contained in each role, I assess the level of cohesion by plotting the relationship between the average path length and the clustering coefficient of each community.

Community Diffusion: Within-Community Cohesion
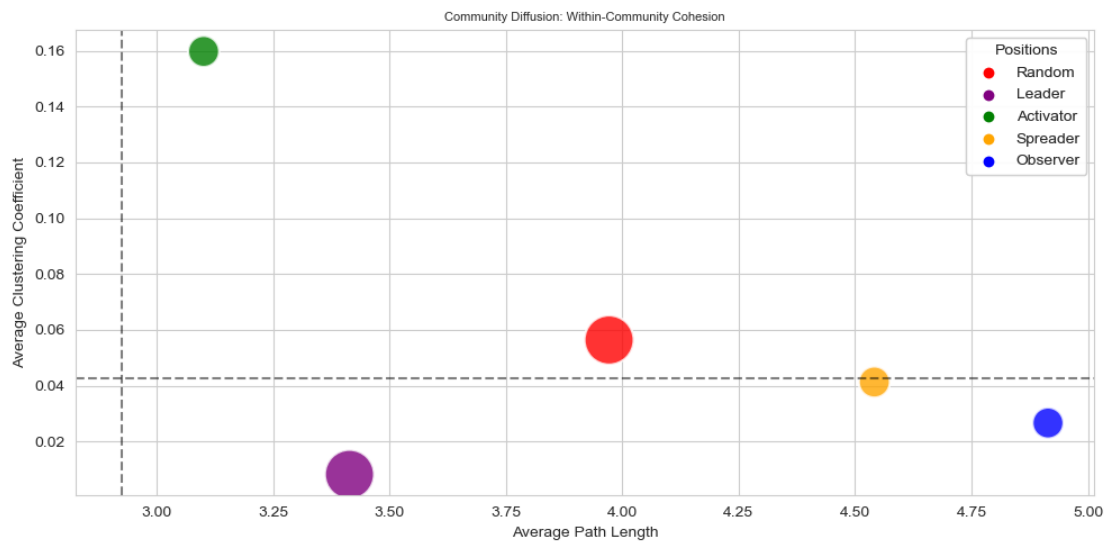
1.2   Diffusion In and Out Community

To estimate the daily spread of an infection, I simulate the diffusion process of each initial seed in each role and compare the ratio of infected nodes within community and outside community across early and late-stage diffusion. For this network, early stage refers to all infections (adoptions) before 20 November 2020 and late stage are all infections (adoptions) up to to 31 May 2021.

Period: 1 October 2020 – 19 November 2020
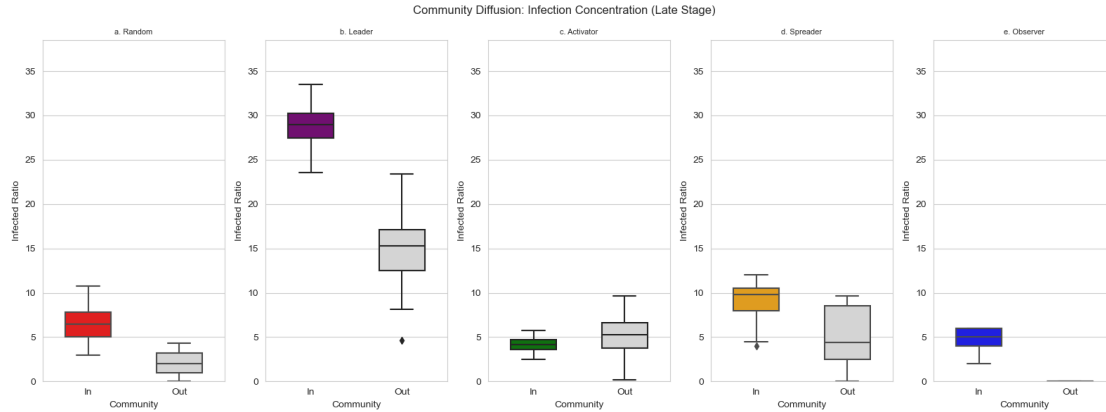
Parameters:
a. 2 Randoms Infected
b. 2 Leaders Infected
c. 0 Activators Infected
d. 2 Spreaders Infected
e. 2 Observers Infected
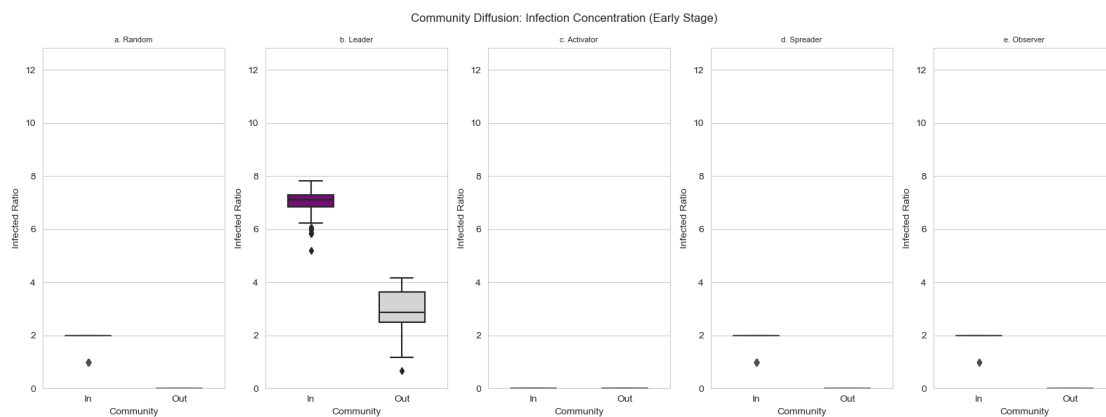
Period: 1 October 2020 – 31 May 2021

Parameters:
a. 6 Randoms Infected
b. 6 Leaders Infected
c. 2 Activators Infected
d. 6 Spreaders Infected
e. 6 Observers Infected

---

2   POSITIONAL DIFFUSION

To simulate the rate and extent of possible diffusion in peak discourse, I trace and plot multiple diffusion trends by assigning specific users from each role as initial seeds. I also visualise a snapshot graph showing the infection rate for each simulation from interactions on 20 November 2020. A community layout is applied to the snapshot graph so that users are mapped within their detected community.
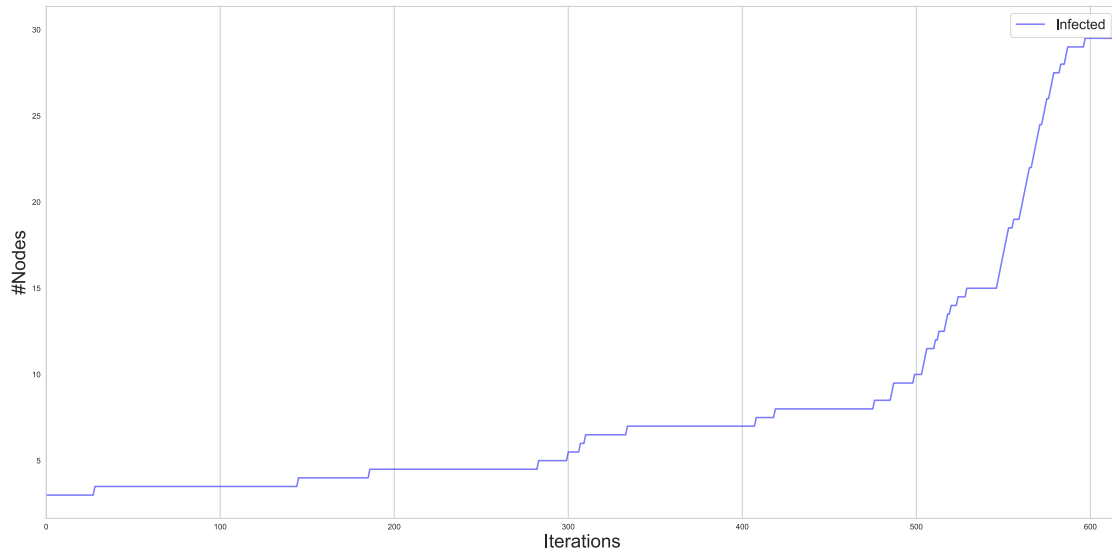
2.1   Random Seed Infection

```
--------
Period: 20 November 2020 00:00:00 - 20 November 2020 23:59:59
```

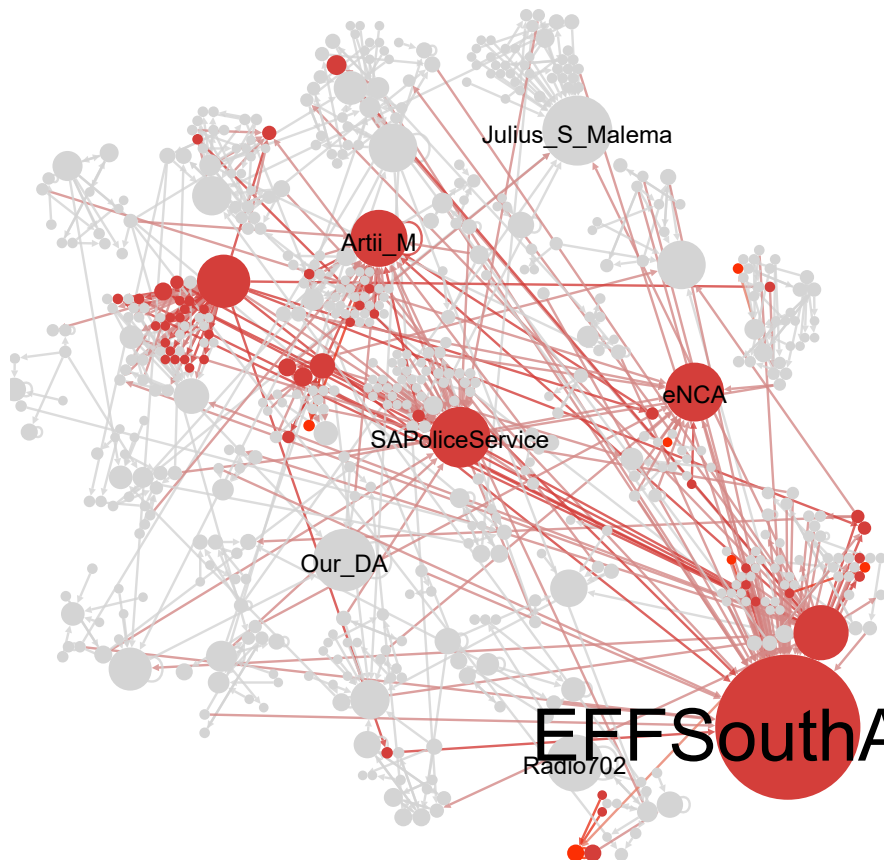**Parameters**: 1.03% Infected, 25.0% Profile, 60.0% Threshold, 0% Blocked.

Start: 6 Randoms Infected

End: 60 Users Infected

```
--------
```
Random Diffusion Trend

Random Diffusion Infected

Julius_S_Malema

Artii_M

eNCA

SAPoliceService

Our_DA

EFFSouthA

Radio702

5

2.2 Leader Seed Infection
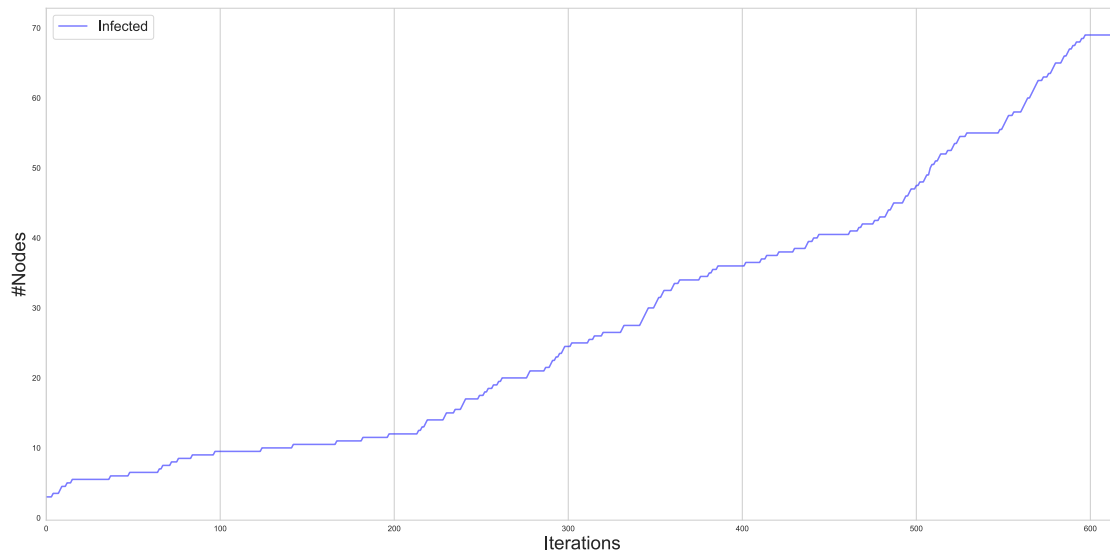

Period: 20 November 2020 00:00:00 – 20 November 2020 23:59:59

Parameters: 1.03% Infected, 25.0% Profile, 60.0% Threshold, 0% Blocked.

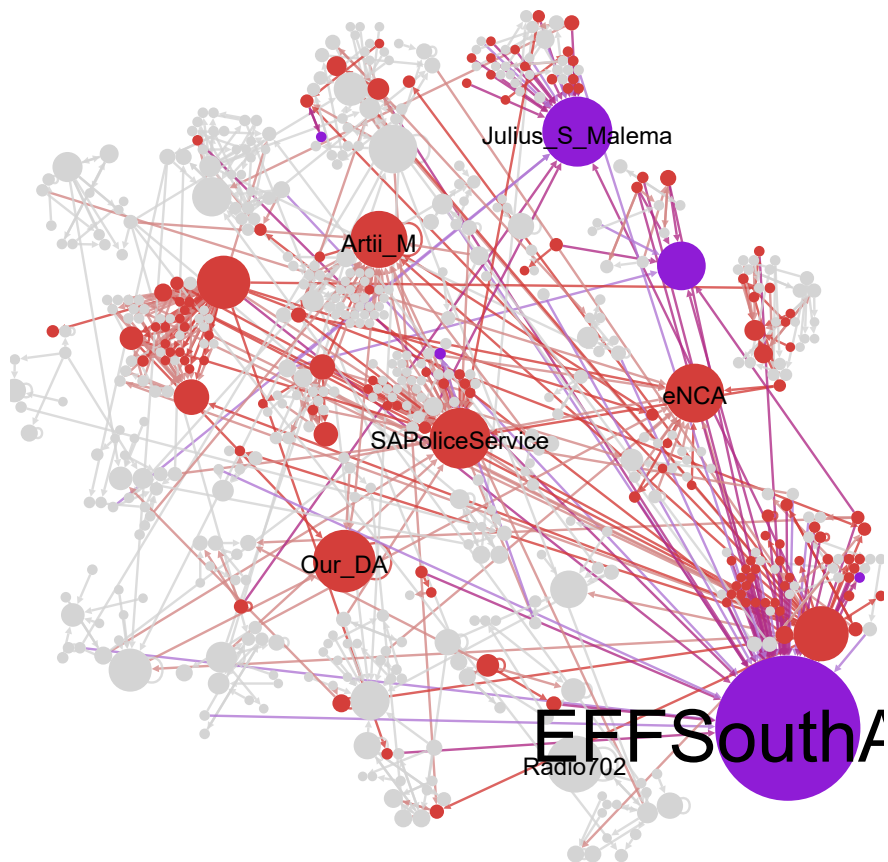**Start**: 6 Leaders Infected

End: 139 Users Infected

Leader Diffusion Trend

Leader Diffusion Infected

2.3 Activator Seed Infection
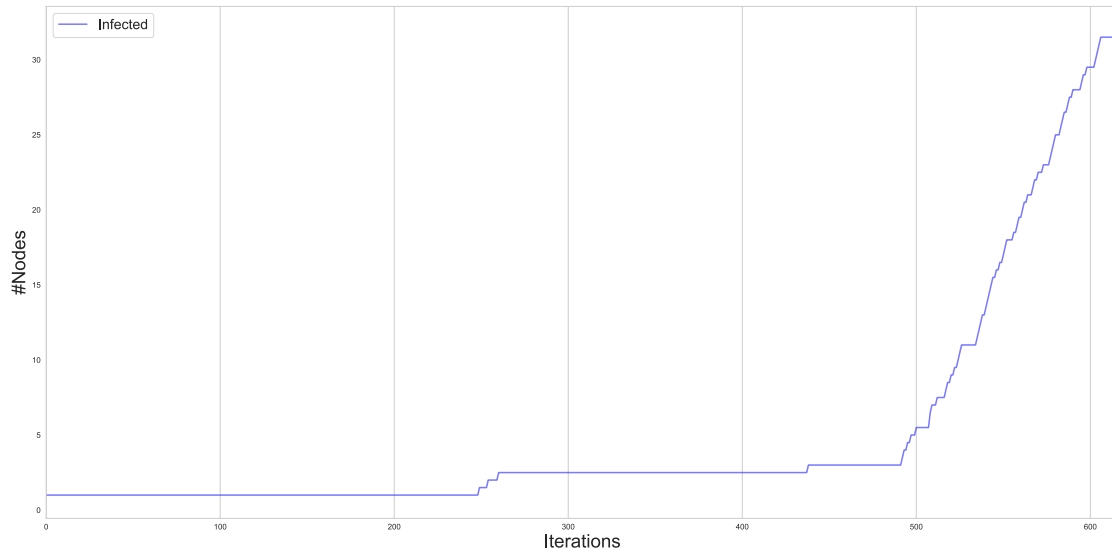

Period: 20 November 2020 00:00:00 - 20 November 2020 23:59:59

Parameters: 0.34% Infected, 25.0% Profile, 60.0% Threshold, 0% Blocked.

Start: 2 Activators Infected

End: 64 Users Infected

Activator Diffusion Trend

Activator Diffusion Infected

2.4   Spreader Seed Infection

Period: 20 November 2020 00:00:00 – 20 November 2020 23:59:59

Parameters: 1.03% Infected, 25.0% Profile, 60.0% Threshold, 0% Blocked.

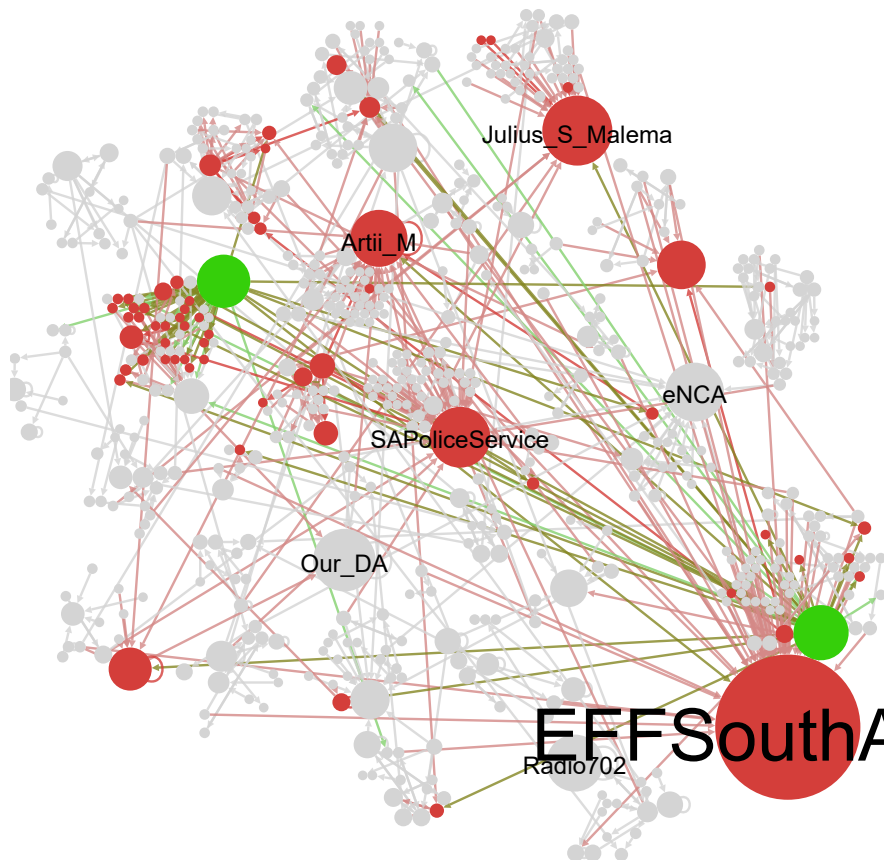**Start**: 6 Spreaders Infected

End: 107 Users Infected

Spreader Diffusion Trend

Spreader Diffusion Infected
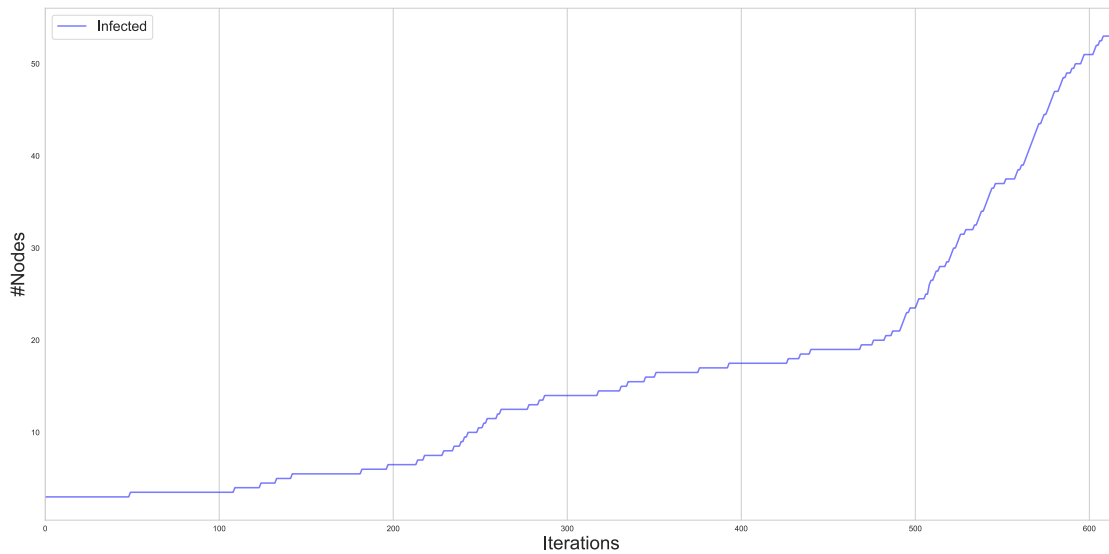
2.5 Observer Seed Infection

Period: 20 November 2020 00:00:00 – 20 November 2020 23:59:59

Parameters: 1.03% Infected, 25.0% Profile, 60.0% Threshold, 0% Blocked.

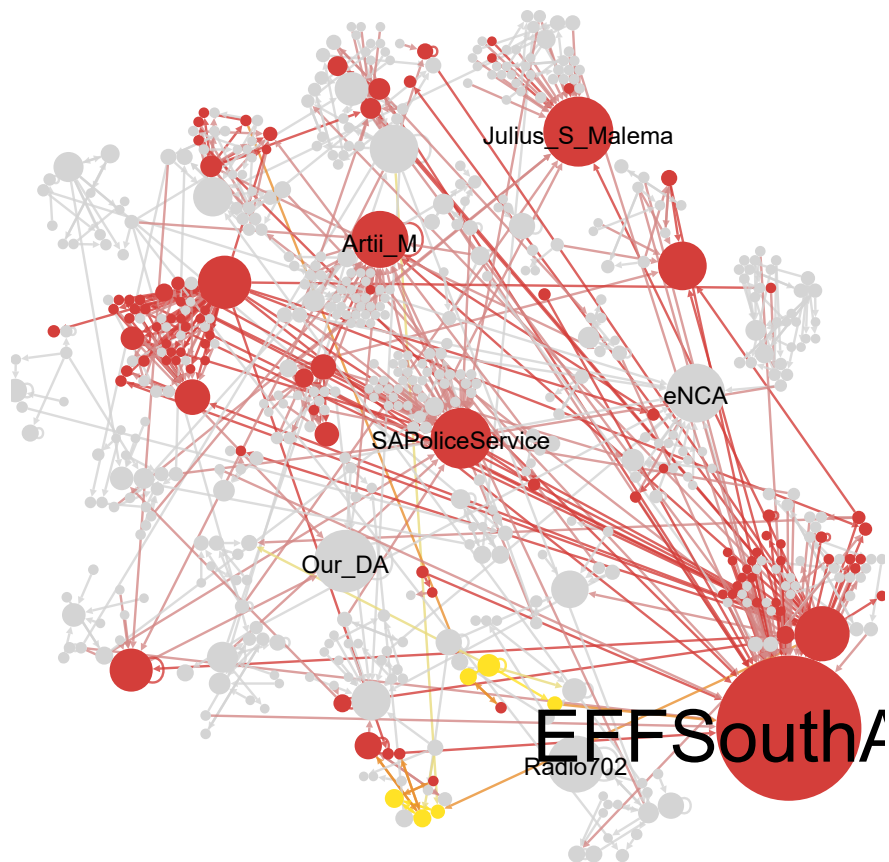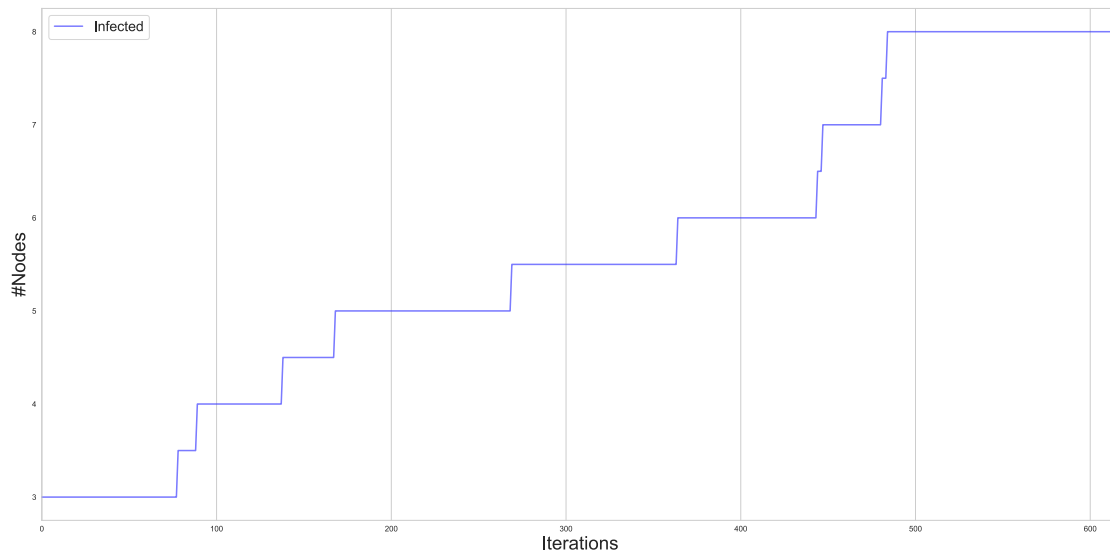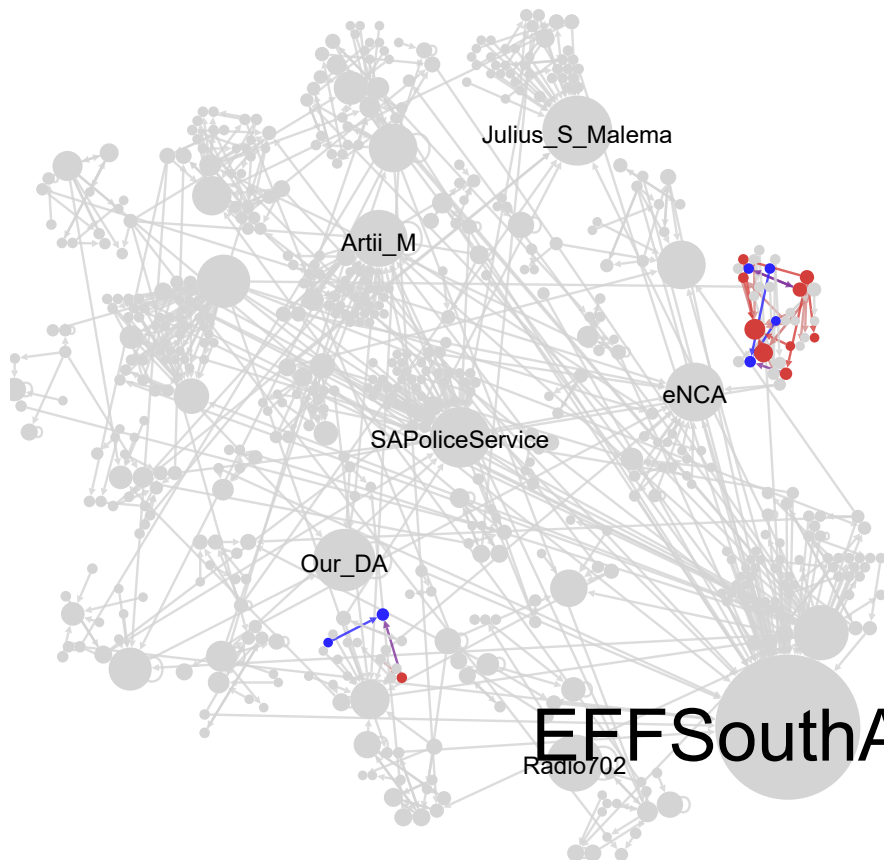**Start:** 6 Observers Infected

End: 16 Users Infected

Observer Diffusion Trend

Observer Diffusion Infected

3   SUMMARY

In this section, I simulate multiple diffusion processes and make inferences about the people, positions, and groups that have the potential to effectively spread information and incite widespread discourse in the #Brackenfell twitter reply network.

To explore the link between community cohesion and diffusion, I match eligible seeds to the communities they are contained in and map the average path length and clustering coefficient of each role. I find 5 seperate communities containing seeds from the *Leader* and *Random* roles, respectively, and 2 communities matching seeds from either *Spreader*, *Observer*, and *Activator* roles, respectively. The cohesion plot shows that communities that contain *Observer* seeds have the highest average path length, while those that contain *Activator* seeds have the lowest. *Leader* communities are found to have the lowest clustering coefficient and *Activator* seeds have the highest. Assuming redundant ties are reinforcing, *Activator* are best placed to outperform.

The diffusion performance of seeds in each role are represented by boxplots. I find *Leader* seeds successful at permeating both within and outside communities in both early-stage and late-stage diffusion, achieving the highest infection rates in both stages of the simulation. *Spreader* seeds are second highest in performance, with *Random* seed performance falling slightly behind. The worst performing diffusers in this network are *Activator* and *Observer* seeds. There are a few notable observations that come out of these simulations. *Activator* seeds, although recording high clustering and low average distance features, are amongst the worst performers. This can however be attributed to the fact that only two nodes are assigned to the *Activator* role, and therefore lack the power to infect many users outside their community and across the network. Furthermore, *Spreader* seeds maintain a slight advantage over *Random* seeds even though *Random* seed communities measure relatively lower average path lengths. Also, *Leader* seeds are contained in communities that have the lowest average clustering levels, yet manage to perform significantly better than seeds in every other role. This suggests, more than anything, when deciding which role is most effective in inciting a diffusion process, the type of features exhibited by specific users in a particular role are the most important factors to consider. In this case leader properties are key. The overall diffusion process shows that higher levels of infection are achieved within-community than outside community. These results are consistent with complex contagion where redundant ties are leveraged to achieve greater virality.

As an additional step, I examine the outcome of diffusion during peak discourse (i.e. the day with the highest interaction of tweet replies). In these simulations, users are assumed to be tweeting in real time. From the first tweet response to the last, I trace and plot the infection trend and visualise the spread at specific points during the day. The results are consistent with those found earlier. *Leader* diffusion is the most expansive in reach, followed by *Spreader*, *Activator*, *Random*, and lastly, *Observer* seeds. The network snapshots show *Leader* seeds to achieve higher and faster penetration of outside communities relative to seeds in other roles. The diffusion trends show *Activator* and *Random* seeds starting off slow before attaining enough support (probably as more users come online to reply to tweets)., while *Observer* seeds either take too long to gain momentum or fail to break out of the trap entirely.

This excercise in diffusion mechanics shows three main things: (i) the identification of key users and roles is critical in promoting the adoption of a new idea or behaviour within and across communities, (ii) early-stage diffusion is a strong predictor of late-stage diffusion, and (iii) close knit communities act as a social

reinforcement mechanism that can accelerate the spread of a new idea or behaviour.