

PART IV: PUTSAFIRST 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 cascades through a social network and produces a behavioural change in other people, communities, and populations.

Data Source

#PutSAFirst 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 (τ) = 60%

Profile (γ) = 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.

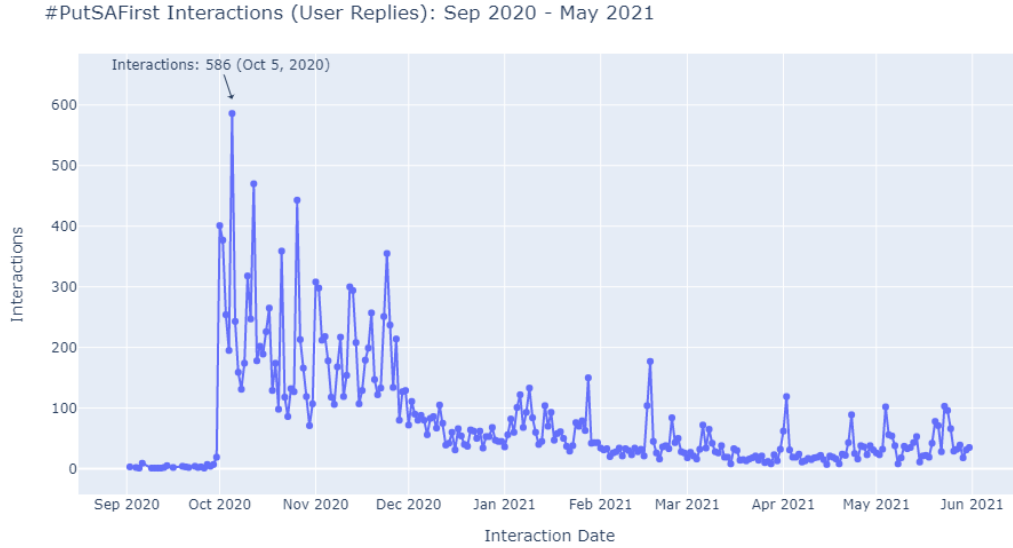
Additionally, for the purposes of the analysis below, I define:

- **Activator (1)** seeds as nodes assigned to the *Activator* cluster sorted by feature values in the order: *outdegree*, *hierarchy*, *transitivity*, *reciprocity*, *#tweeted*, *betweenness*, and *closeness*.
- **Activator (2)** seeds as nodes assigned to the *Activator* cluster sorted by feature values in the order: *betweenness*, *closeness*, *outdegree*, *hierarchy*, *transitivity*, *reciprocity*, and *#tweeted*.

Method

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

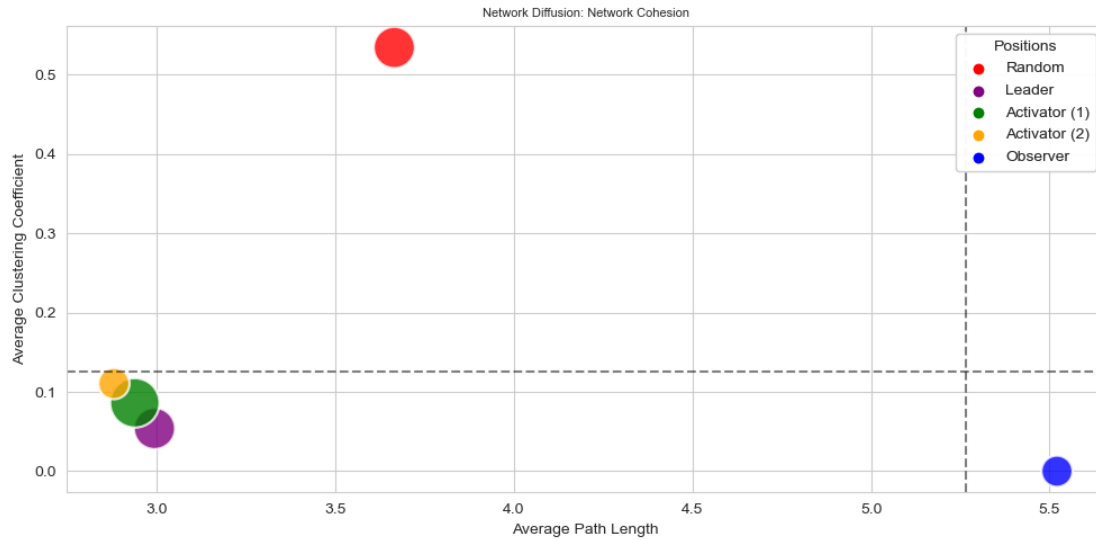
Number of Daily Interactions (Tweet Replies) in #PutSAFirst Discourse:



1 COMMUNITY DIFFUSION

1.1 Cohesion in Network

As a precaution, given that the communities detected in part II revealed poorly partitioned communities with modularity scores below the threshold, I choose to compare cohesion measures at the network level. Therefore, for each role, I assess the level of cohesion in the network by plotting the relationship between the average path length and the clustering coefficient of each initial seed.



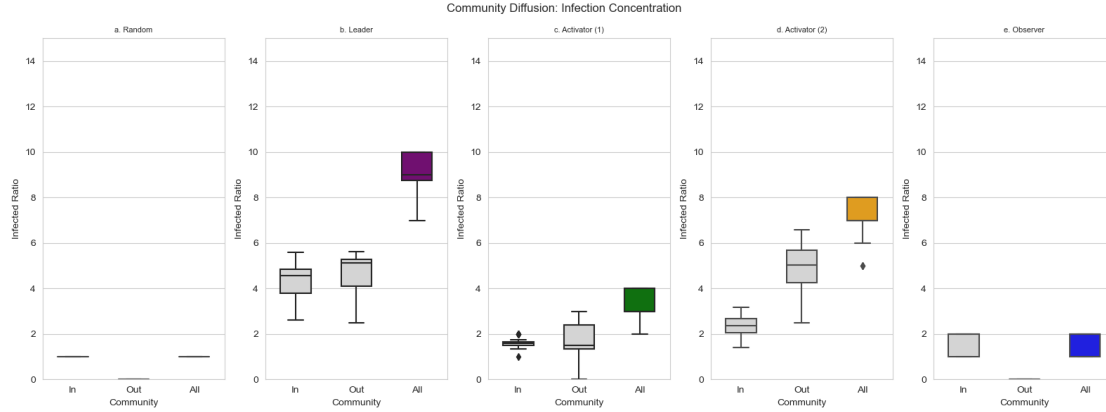
1.2 Diffusion In and Out Community

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 5 October 2020 and late stage are all infections (adoptions) up to 31 May 2021. Again, to control for low modularity obtained from the partition, I focus my analysis on the ratio of all infected nodes both within and outside community.

Period: 1 September 2020 - 4 October 2020

Parameters:

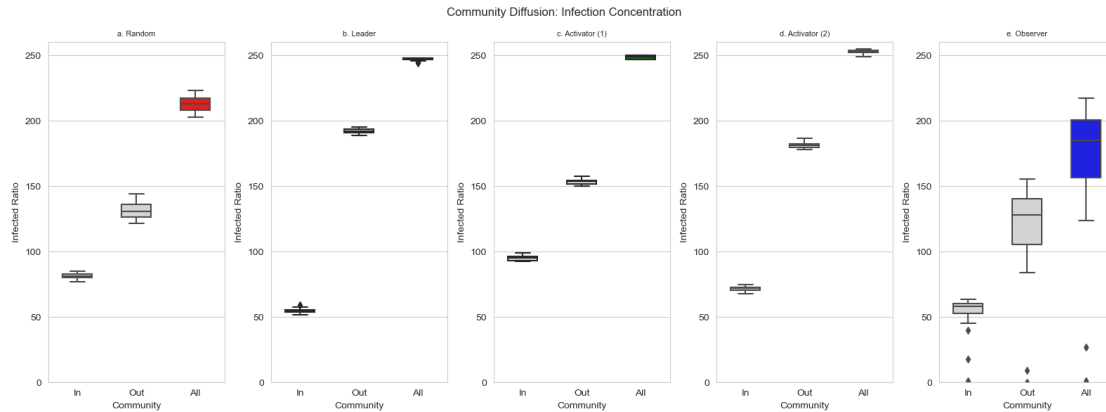
- a. 2 Randoms Infected
- b. 2 Leaders Infected
- c. 2 Activators (1) Infected
- d. 2 Activators (2) Infected
- e. 2 Observers Infected



Period: 1 September 2020 - 31 May 2021

Parameters:

- a. 6 Randoms Infected
- b. 6 Leaders Infected
- c. 6 Activators (1) Infected
- d. 6 Activators (2) Infected
- e. 4 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 5 October 2020. A community layout is applied to the snapshot graph so that users are mapped within their detected community.

2.1 Random Seed Infection

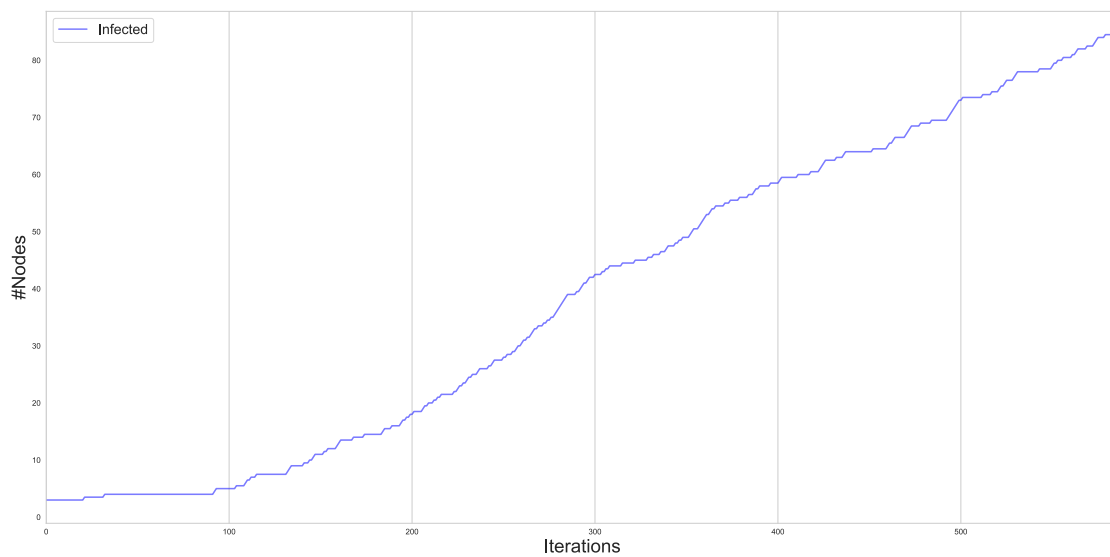
Period: 5 October 2020 00:00:00 – 5 October 2020 23:59:59

Parameters: 1.36% Infected, 25.0% Profile, 60.0% Threshold, 0% Blocked.

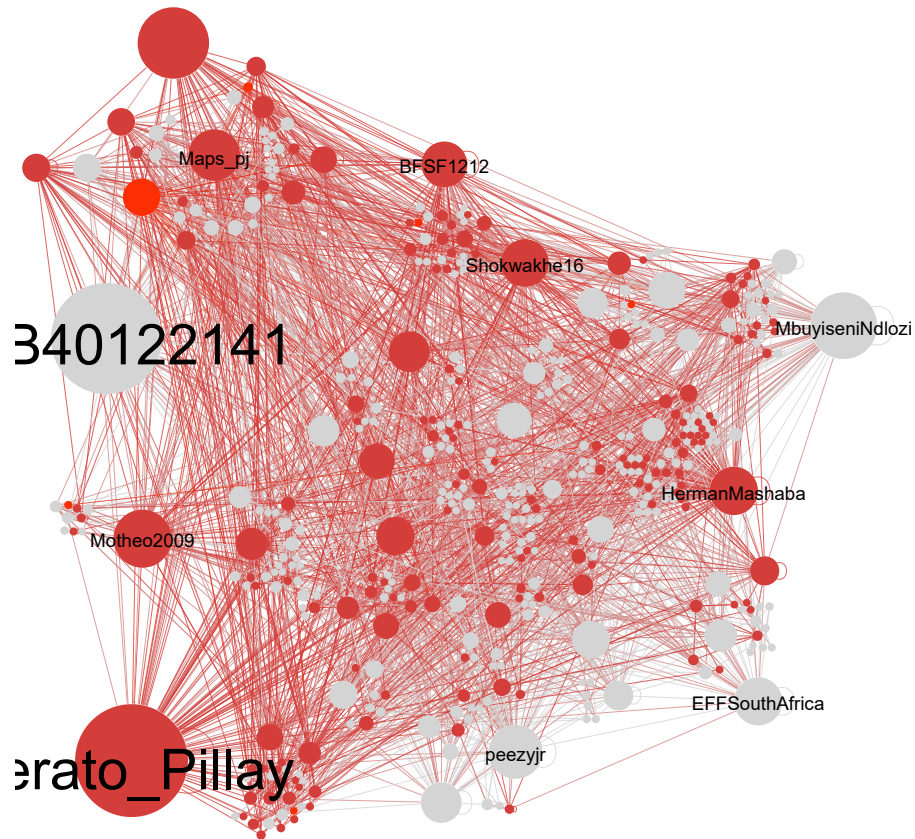
Start: 6 Randoms Infected

End: 169 Users Infected

Random Diffusion Trend



Random Diffusion Infected



2.2 Leader Seed Infection

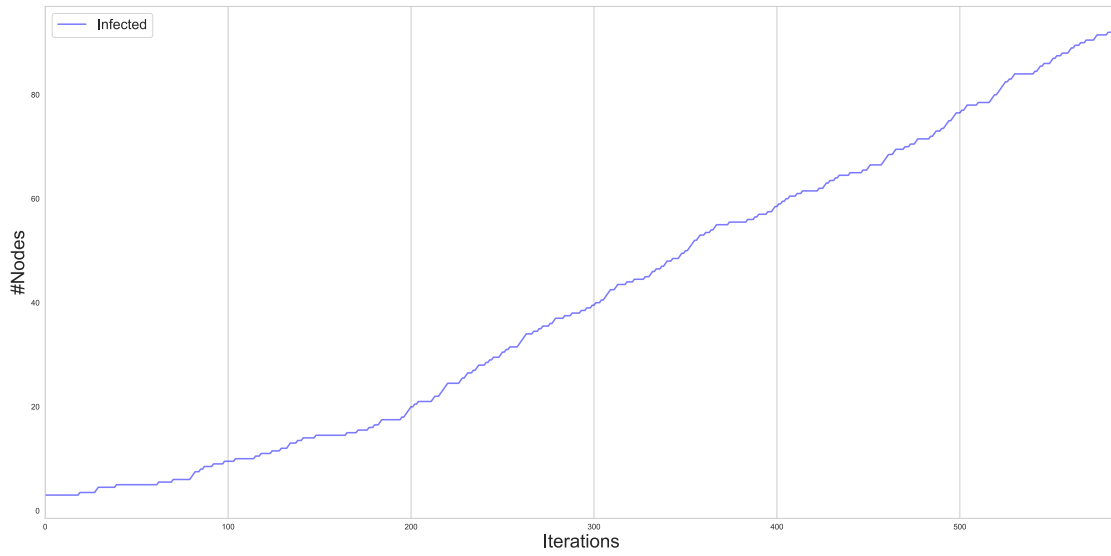
Period: 5 October 2020 00:00:00 – 5 October 2020 23:59:59

Parameters: 1.36% Infected, 25.0% Profile, 60.0% Threshold, 0% Blocked.

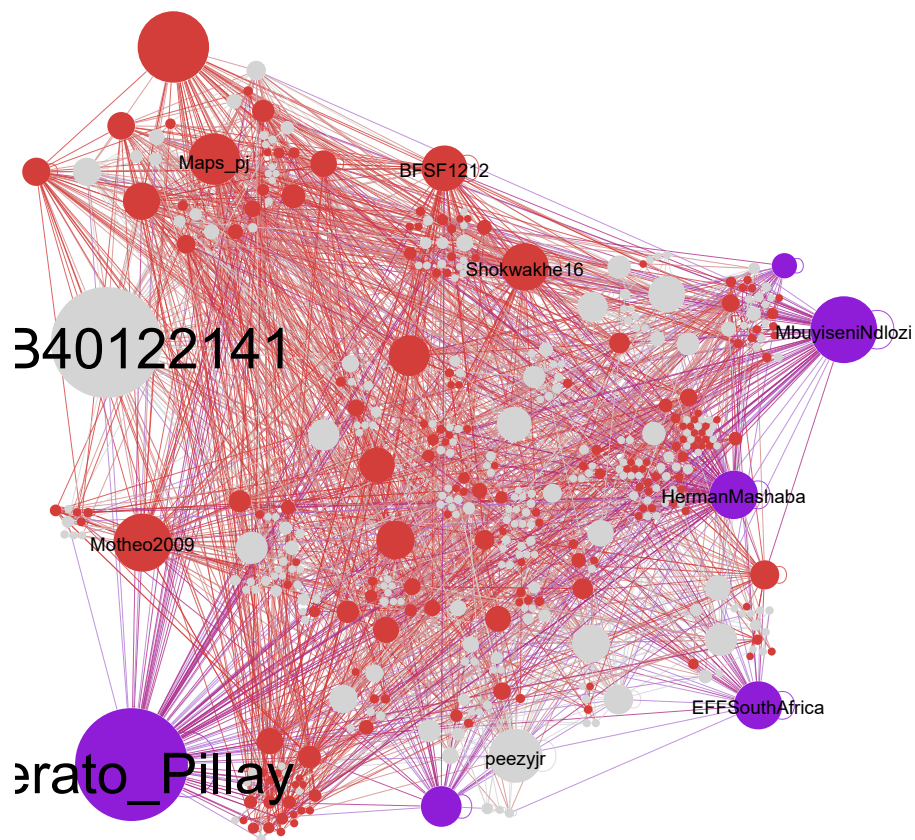
Start: 6 Leaders Infected

End: 185 Users Infected

Leader Diffusion Trend



Leader Diffusion Infected



2.3 Activator (1) Seed Infection

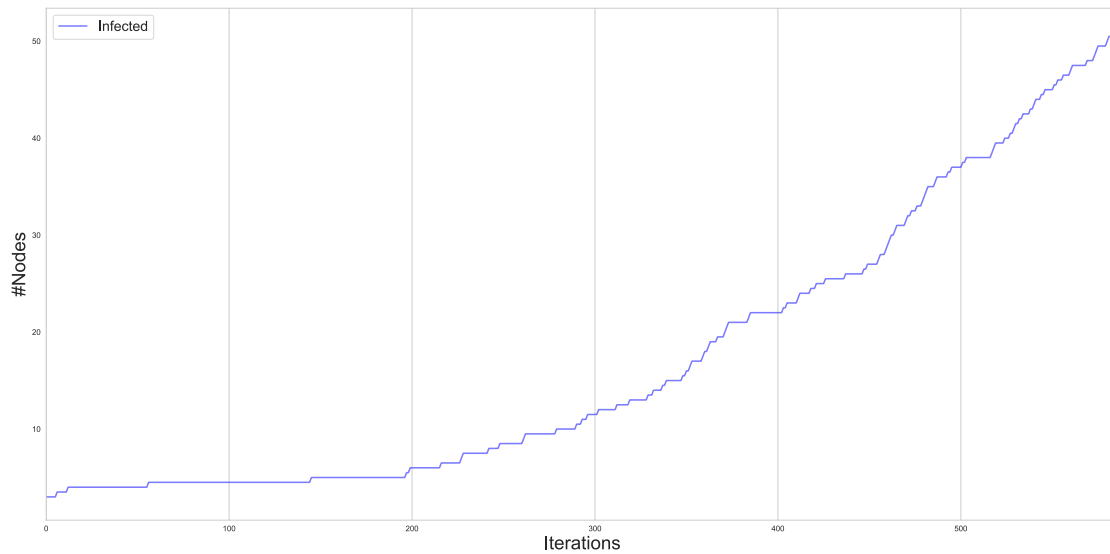
Period: 5 October 2020 00:00:00 - 5 October 2020 23:59:59

Parameters: 1.36% Infected, 25.0% Profile, 60.0% Threshold, 0% Blocked.

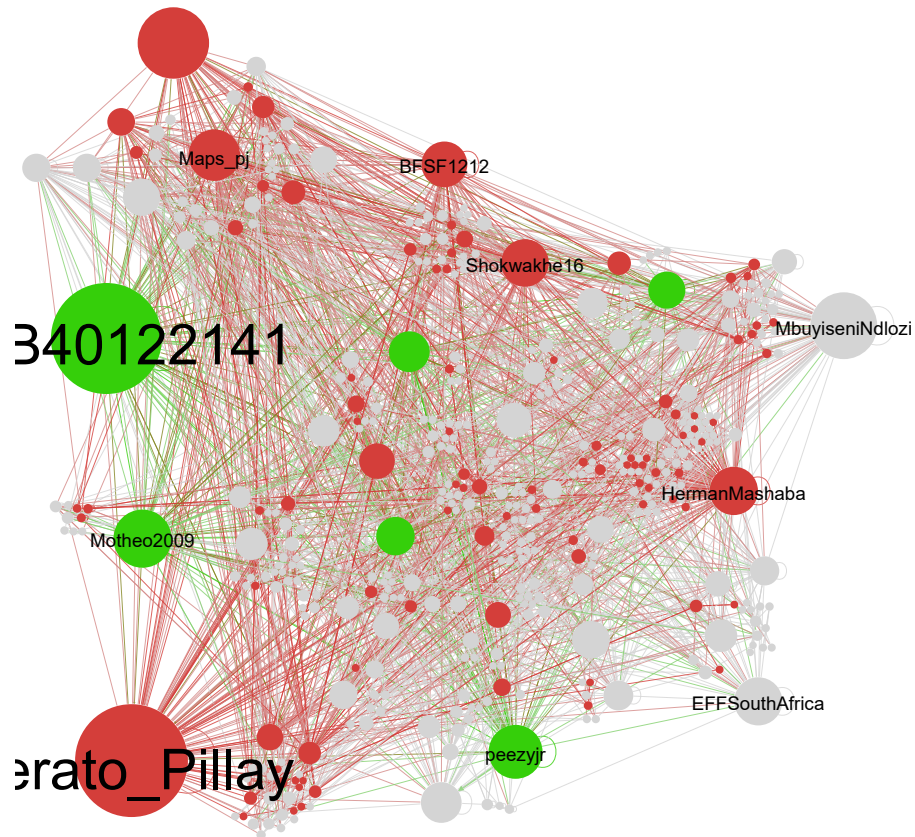
Start: 6 Activators (1) Infected

End: 102 Users Infected

Activator (1) Diffusion Trend



Activator (1) Diffusion Infected



2.4 Activator (2) Seed Infection

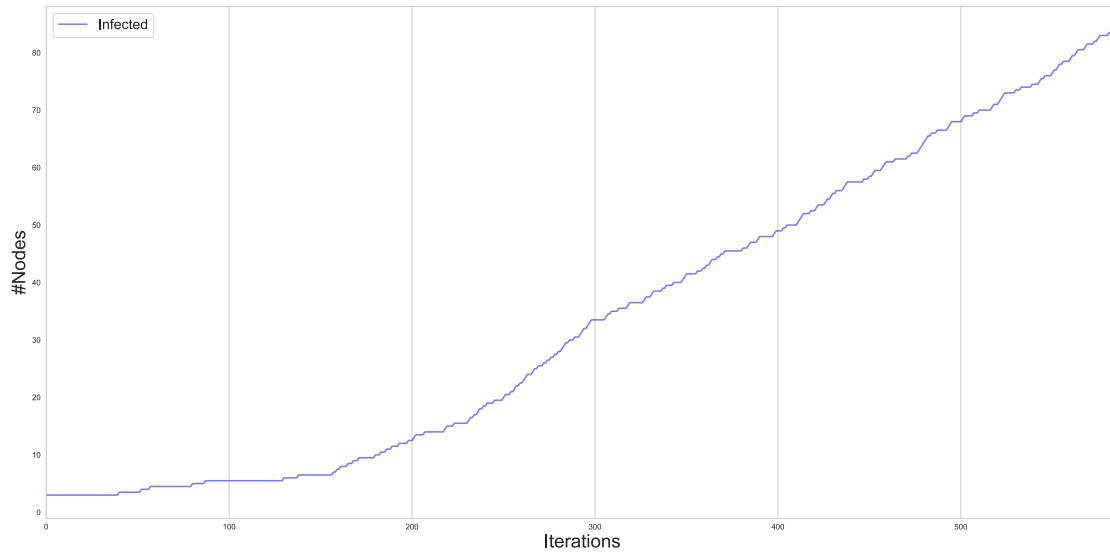
Period: 5 October 2020 00:00:00 - 5 October 2020 23:59:59

Parameters: 1.36% Infected, 25.0% Profile, 60.0% Threshold, 0% Blocked.

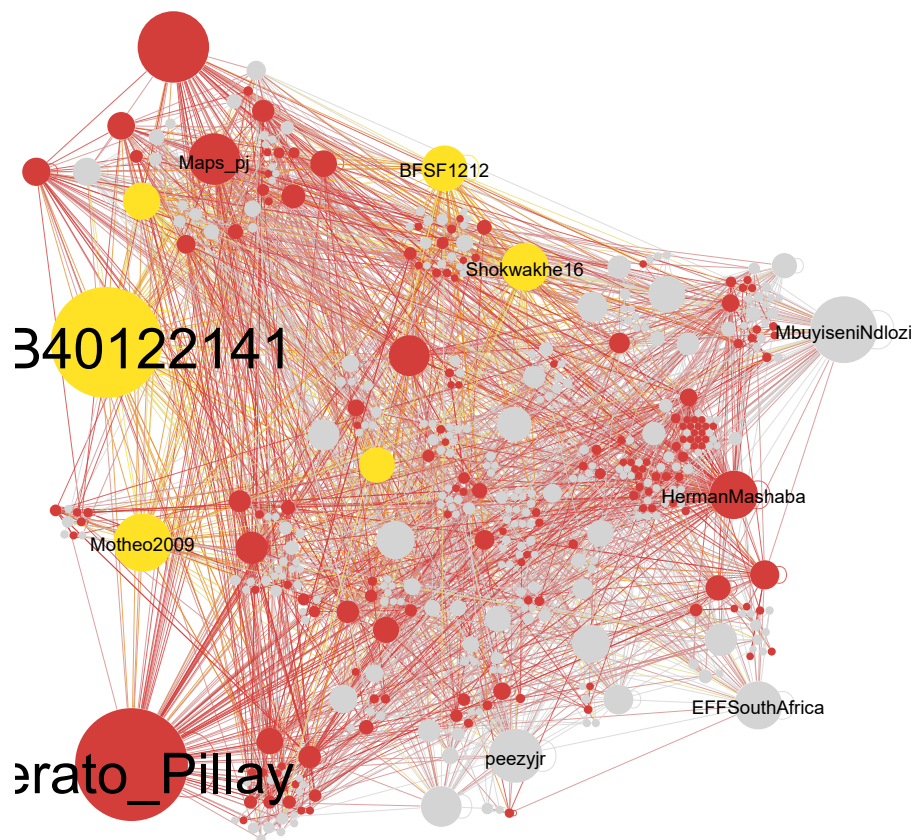
Start: 6 Activators (2) Infected

End: 168 Users Infected

Activator (2) Diffusion Trend



Activator (2) Diffusion Infected



2.5 Observer Seed Infection

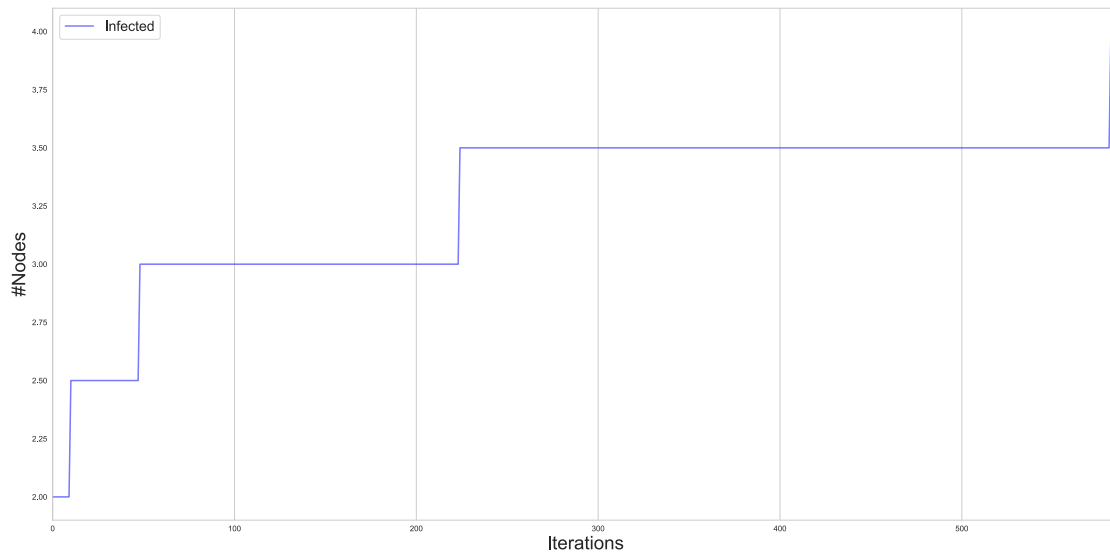
Period: 5 October 2020 00:00:00 - 5 October 2020 23:59:59

Parameters: 0.91% Infected, 25.0% Profile, 60.0% Threshold, 0% Blocked.

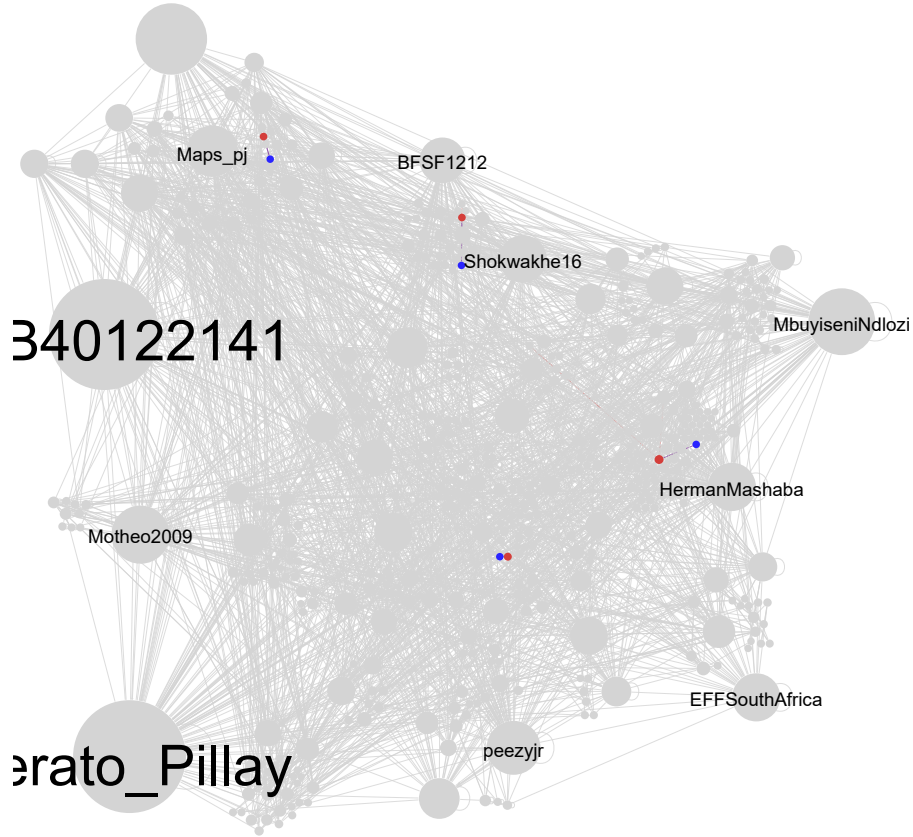
Start: 4 Observers Infected

End: 8 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 #PutSAFirst twitter reply network.

To explore the link between community cohesion and diffusion, I map the average path length and clustering coefficient for initial seeds in each role. The cohesion plot shows that *Observer* seeds have the highest average path length, while *Activator (2)*, *Activator (1)*, and *Leader* seeds have the lowest. *Observer* seeds are found to have the lowest clustering coefficient, and *Random* seeds to have the highest. Assuming redundant ties are reinforcing, *Random* seeds are best placed to outperform.

The diffusion performance of seeds in each role are represented by boxplots. I find *Leader* seeds to have the advantage in early-stages, and *Activator (2)* seeds in late-stages, albeit by a marginal difference. In fact, the late-stage results for *Leader*, *Activator (1)*, and *Activator (2)* seeds are very similar, much like their cohesion attributes. *Random* and *Observer* seed performance are not too far behind either. The diffusion performance in this network does not appear to be driven by the type of user or role initiating the spread. In other words, unlike the other networks, I fail to find a significant relationship between position and diffusion. This

could be attributed to the clustering coefficient measure described in part I. This network has a substantially higher clustering coefficient than would be expected in a random network (163.7x higher compared to 20.9x higher and 25.1x higher found in the #Senekal and #Brackenfell networks, respectively). Inline with most predictions, this confirms that redundant ties act as a social reinforcement mechanism or multiplier effect, so much so that they compensate for the lack of popularity and centrality of nodes and positions. This is 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 *Random*, *Activator (2)*, *Activator (1)*, and lastly, *Observer* seeds. The network snapshots show *Leader* seeds to achieve higher and faster penetration throughout the network relative to seeds in other roles. *Observer* seeds either take too long to gain momentum or fail to break out of the trap entirely.

This exercise in diffusion mechanics shows that highly clustered networks benefit from a social reinforcement mechanism or multiplier effect that accelerates the spread of an idea or social behaviour, even if the user spreading the idea is not central or popular.