Most efficient information spreaders in a network

Network Science 2021/22

Group 36
Pedro de Oliveira Rosa Alves Leitão - 90764
Tomás da Silva Faria Lampreia Gomes - 90782
João Filipe Boucinha de Sá - 93844

Instituto Superior Técnico

1 Introduction

Our work covers the topic of Opinion dynamics and information spreading in social networks. More specifically, we analyzed the characteristics of the most influential information spreaders on a given network.

In 2010, Maksim Kitsak et al. [1] presented a study about how the best spreaders of a network are the ones in its core, showing that the nodes with a higher k-shell value, generally propagate more easily through the network than the ones in the external shells, despite of their degree.

In order to know how to effectively spread information in a fast way and using as few resources as possible, we study the characteristics of the best information spreaders in a network.

This can be useful to discover where someone can strategically position himself in order to make their message reach the majority of people in the least time possible.

To carry out this study, we used two different network types, which were a LFR benchmark graph and a graph using a duplication divergence model, and we simulated the information spreading in traditional models such as the SI, SIS and SIR models, and in competitive scenarios such as the Competing Rumors model.

We found that to start disseminating information in the certer of a network helps in the spreading process, needing less time to communicate with the majority of the network.

2 Summary of gathered results

We used the k-shell decomposition to identify which nodes were present in the core of the network.

To test the results of starting the propagation in nodes with different characteristics, we created two LFR benchmark graphs, one with 10000 nodes with a minimum degree of 2 to test the results and a smaller one with 500 nodes with a minimum degree of 4 to draw the propagation of the spreading in the graph, both with a power law exponent for the degree distribution of 2.8, and 2 for the community size distribution, and with 50% of the inter-community edges being incident to each node. We also created two duplication divergence graphs, one with 10000 nodes to test the data and a smaller one with 500 nodes (for graphical demonstration purposes), both with a probability of 50% for retaining the edges of the replicated nodes.

In the graphs that we drew, the blue nodes represent the susceptible (healthy) nodes, the red nodes represent the infected ones and the green nodes represent the recovered ones.

In order to test the differences between starting disseminating information in a node with a higher k-shell value and lower degree and starting in a node with a higher degree and lower k-shell value, we computed an average of the number of iterations taken to reach a certain condition between 5 simulations for each created network.

In the SI model, we chose a contact infection rate of 0.035 and computed the number of iterations a certain node took in order to infect at least 90% of a network with a certain rumour.

In the SIS model, we chose a recovery rate of 0.025, which is lower than the contact infection rate (0.035) and so the simulation enters in an endemic state where the number of infected nodes start converging after a certain time. Because of this, to test the effectiveness of starter nodes with different attribute values, we computed the number of iterations that the simulation takes until the number of infected nodes starts to converge.

In the SIR model, we used the same contact infection and recovery rates as in the SIS model. We tested how long a rumor survives in a certain network, when starting the spreading in a certain node, until all the nodes in the network are either healthy or recovered.

For the LFR benchmark network with 10000 nodes, when we start propagating from a node with a k-shell value of 4 (this network's maximum k-shell value) and degree of 5, it takes an average of 76.80 iterations to make its information reach 90% of the network in the SI model. In the SIS model, starting the spreading in this node made the propagation need an average of 70.00 iterations to start converging to 6714.60 infected nodes on average. In the SIR model, a rumor that started propagating from this node lasted an average of 436.20 iterations before all nodes become either susceptible or recovered, and the 5 simulations run had a maximum of 4103.00 infected nodes in the network on average.

For the same network (LFR benchmark network with 10000 nodes), when we start propagating from a node with a k-shell value of 2 and degree of 6 we need around 81.40 iterations on average to spread the information to 90% of the network in the SI model. In the SIS model, the spreading took in average 90.60 iterations until converging to 6740.80. In the SIR model, starting the spreading from this node made the rumor last an average of 377.00 iterations and the maximum infected nodes in each simulation was on average 3252.40 nodes.

For the LFR benchmark network with 500 nodes, when we started propagating from a node with the highest k-shell value in the network, which is 6, and a degree of 6 in the SI model, it took an average of 27.00 iterations until reach-

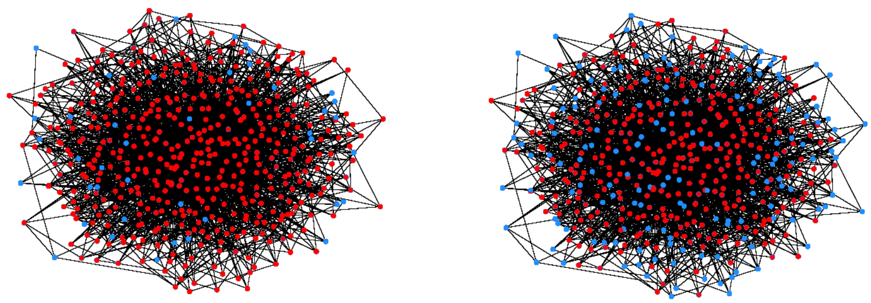

Fig. 1. State of the propagation at the 28th iteration in the LFR benchmark network with 500 nodes in the SI model when starting at the node with a k-shell value of 6 and degree 6 (left) and the propagation at the same iteration in the same network and model starting at the node with k-shell value of 4 and degree 8 (right)

ing 90% of the network. In the SIR model, the rumor that started propagating in these nodes lasted an average of 284.20 iterations and reached a maximum 348.80 infected nodes on average.

For the same network (LFR benchmark network with 500 nodes), when we started propagating from a node with a k-shell value of 4 and a degree of 8 in the SI model, it took an average of 29.00 iterations until reaching 90% of the network. In the SIR model, the rumor that started propagating in these nodes lasted an average of 253.60 iterations, reaching a maximum 269.60 infected nodes on average.

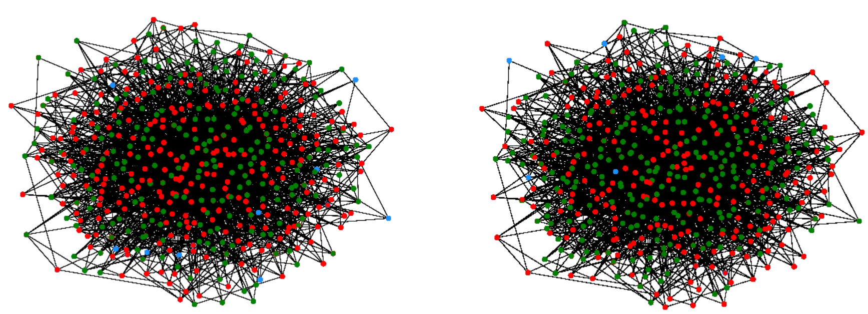

Fig. 2. The LFR benchmark network with 500 nodes in the SIR model at the 49th iteration when starting at the node with a k-shell value of 6 and degree 6 (left) and the network at the same iteration when starting at the node with k-shell value of 4 and degree 8 (right)

For the duplication divergence network with 10000 nodes, starting propagating in a node with a k-shell value of 15 (this network's maximum k-shell value) and a degree of 16 took an average of 36.00 iterations to reach 90% of the network. In the SIS model, it took an average of 34.00 iterations until converging to 7527.80 infected nodes on average. In the SIR model, the rumor that started propagating from this node lasted 414.20 iterations on average before all nodes become susceptible or recovered, and reached a maximum of 5777.20 infected nodes on average.

For the same network (duplication divergence network with 10000 nodes), when we start propagating from a node with a k-shell value of 7 and a degree of 39 it took an average of 39.40 iterations to reach 90% of the network in the SI model. In the SIS model, starting propagating from this node took an average of 34.40 iterations until converging to 7718.20 infected nodes on average. In the SIR model, when we started the spreading from this node, the rumor lasted 403.20 iterations on average before having no infected nodes remaining and reached a maximum of 5762.40 infected nodes.

For the duplication divergence network with 500 nodes, when we started propagating from a node with the highest k-shell value in the network, which is 8, and a degree of 9 in the SI model, it took an average of 42.80 iterations until reaching 90% of the network. In the SIS model, starting propagating in this node took an average of 35.60 iterations until converging to 360.00 infected nodes on average.

For the same network (duplication divergence network with 500 nodes), when we started propagating from a node with a k-shell value of 4 and a degree of 11 in the SI model, it took an average of 50.80 iterations until reaching 90% of the network. In the SIS model, starting propagating in this node took an average of 43.80 iterations until converging to 353.80 infected nodes on average.

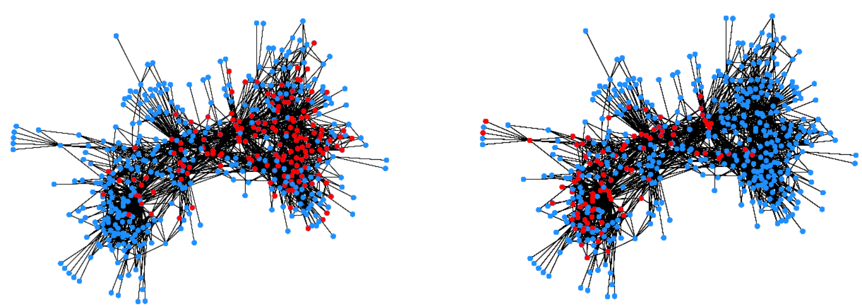

Fig. 3. The duplication divergence network with 500 nodes in the SIS model at the 21st iteration when starting at the node with a k-shell value of 8 and degree 9 (left) and the network at the same iteration when starting at the node with k-shell value of 4 and degree 11 (right)

We also simulated the information propagation with the same networks in the same models, but starting the spreading from two nodes instead of just one.

For this, in the LFR benchmark network with 10000 nodes, we tested starting the spreading from two nodes with the maximum k-shell value in the network, which is 4, one with degree 5 and the other with degree 6, needing an average of 70.80 iterations to reach 90% of the network in the SI model. In the SIS model, when starting the spreading in these two nodes, it took an average of 66.60 iterations to start converging to an average of 6745.60 infected nodes. In the SIR model, when starting propagating in these two nodes, the propagation lasted in average 427.60 iterations and had a maximum of 4088.80 infected nodes on average. Meanwhile, when we started the propagation in this network from two connected nodes, both with the highest k-shell value and degrees 6 and 106 in the SIR model, the rumor spreaded lasted 411.20 iterations on average and reached a maximum of 4093.60 infected nodes on average.

For the same network (LFR benchmark network with 10000 nodes), when we started the spreading from two nodes with the same degrees (5 and 6) but with a smaller k-shell value of 2, the rumor needed an average of 79.00 iterations to reach 90% of the network in the SI model. In the SIS model, a rumor that started propagating from these nodes took an average of 69.40 iterations to converge to an average of 6553.60 infected nodes. In the SIR model, the spreading that started in these two nodes lasted in average 498.40 iterations and had a maximum of 4050.80 infected nodes on average.

With the duplication divergence network with 10000 nodes, when we started the spreading from two nodes with the highest k-shell value, which is 15, and degrees 16 and 18, it took an average of 35.60 iterations to reach 90% of the network in the SI model. In the SIS model, it takes an average of 30.00 iterations until converging to 7560.40 infected nodes on average. In the SIR model, when starting propagating in these two nodes, the rumor lasted 410.20 iterations on average, having a maximum of 5729.20 infected nodes on average. Meanwhile, when we started the propagation in this network from two connected nodes, both with the highest k-shell value and degrees 16 and 36 it was needed 35.80 iterations on average to spread to 90% of the network in the SI model.

For the same network (duplication divergence network with 10000 nodes), while starting from two nodes with a k-shell value of 7 and degrees 15 and 39 it takes an average of 38.80 iterations to reach 90% of the network in the SI model. In the SIS model, when we started spreading from these two nodes, it took an average of 33.00 iterations to start converging to 7657.40 infected nodes on average. In the SIR model, when we started propagating in these two nodes, the rumor lasted 380.80 iterations on average and had a maximum of 5769.80 infected nodes on average.

For the duplication divergence network with 500 nodes, when we started the spreading from two nodes with the highest k-shell value, which is 8, and degrees 9 and 10, it took an average of 38.80 iterations to reach 90% of the network in the SI model. In the SIR model, when starting propagating in these two nodes, the rumor lasted 267.20 iterations on average, having a maximum of 274.60 infected nodes on average.

For the same network (duplication divergence network with 500 nodes), when we started the spreading from two nodes with 4 as their k-shell value, and degrees 4 and 11, it took an average of 39.40 iterations to reach 90% of the network in the SI model. In the SIR model, when starting propagating in these two nodes, the rumor lasted 195.80 iterations on average, having a maximum of 225.40 infected nodes on average.

We also simulated a model for Competing Rumors, in which a node is labeled as recovered to represent a well informed person, and is a node with the highest k-shell value in the network and lowest degree (in that k-shell), while a node labeled as infected, to represent someone spreading Fake News, starts as a node with a lower k-shell value and higher degree. The two different nodes start spreading information to their neighbors and the simulation ends when there are only 5% nodes labeled as healthy (which represents that they didn't receive any information). We used the same contact infection rate for both the recovered and the infected nodes, which was 0.035.

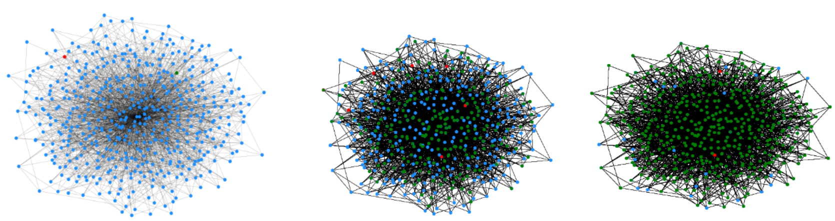

Fig. 4. The Competing Rumors simulation in the LFR benchmark network with 500 nodes of the green node with a k-shell value of 6 and a degree of 6 against the red node with a k-shell value of 4 and a degree of 8 in the initial state (left) after 13 iterations (middle) and after 26 iterations (right)

By simulating the Competing Rumors model in the LFR network with 10000 nodes, by having a node with the highest k-shell value, which is 4, and a degree of 5 as the starter recovered node and having a node with a k-shell value of 2 and a degree of 6 as the starter infected node, the simulation ended after 83 iterations, with 9282 nodes labeled as recovered and 226 nodes labeled as infected. In the same network, by having a node with the highest k-shell value and degree 174 as the starter recovered node and a node with the same k-shell value and degree 5 as the starter infected node, the simulation ended after 82 iterations with 9429 recovered nodes and 85 infected nodes.

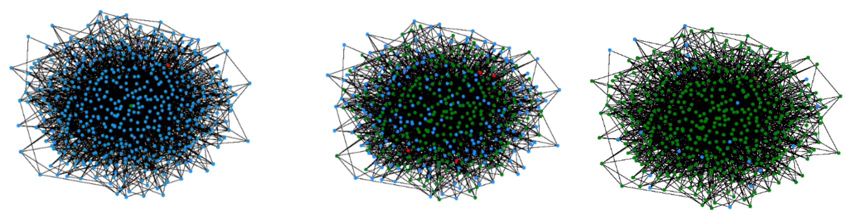

Fig. 5. The Competing Rumors simulation in the LFR benchmark network with 500 nodes of the green node with a k-shell value of 6 and a degree of 350 against the red node with the same k-shell value and a degree of 6 in the initial state (left) after 10 iterations (middle) and after 20 iterations (right)

3 Discussion

From all the simulations, we were able to take many conclusions, but the most blatant one was that indeed k-shell value of a node is a far more significant attribute than that of its degree, for most cases. Both graphs revealed that information disseminated from a node at the centre of the network is spread at a higher pace than if it started from a node at the periphery, regardless of the degree of either nodes. The simulations of the SIR model showed that information will have more persistence when also starting in a high k-shell, and that its position also affects the maximum number of nodes it is able to reach. When we used 2 spreaders instead of one, the k-shell value was still able to create as much of an impact to both the spreading velocity and the persistence on either graph. However, we didn't manage to reproduce the findings that Maksim Kitsak et al. did, regarding the use of 2 starter nodes that are not connected, as the results didn't significally change. It is worth noting that the type of graph also manipulated the results of the simulations for the 3 models (SI, SIS, SIR). We believe it has to do with the existence of communities in the LFR benchmark graph, and that it led to a slower rate of spreading, but a higher persistence, which is rather interesting. Finally, the Competing Rumors model revealed that both the position of a node in the network and its degree have a say on which spreader will get to have a higher impact on the network.

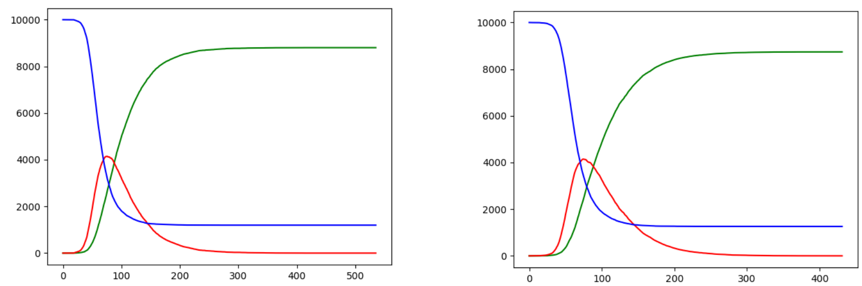

Fig. 6. The evolution of the LFR benchmark network with 10000 nodes when starting the spreading in a node with k-shell value of 4 and degree of 5 (left) and the evolution of the same network when starting in a node with k-shell value of 2 and degree of 6 (right)

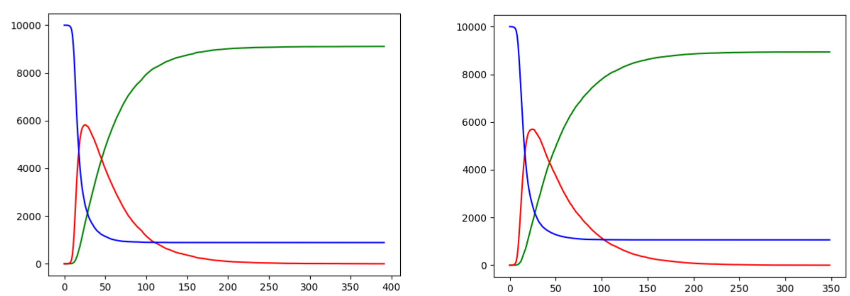

Fig. 7. The evolution of the duplication divergence network with 10000 nodes when starting the spreading in a node with k-shell value of 15 and degree of 16 (left) and the evolution of the same network when starting in a node with k-shell value of 7 and degree of 39 (right)

4 Supporting references

References

1. Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Hernán A. Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6:888–893, 2010.