Team members: Dimitar Kochev, Matei Pop, Mihael Stoyanov, Lorand Ladnai

Our data: [Email-EU](https://networkrepository.com/ia-email-EU.php?fbclid=IwY2xjawGUDyVleHRuA2FlbQIxMAABHfhin88U-90y_9Nj6tjd_D9ti7egzabl5PlBdbu_nV1lQnGrYGhFbt9n9Q_aem_z1qhvK7ifxaDvxHungKAOA)

Description: Our dataset contains a synthetic network generated from email data of a European research institution. It is considered a sparse network due to the fact that the number of edges is relatively smaller than the amount of all possible connections. Each node is an individual entity and the edge signifies a communication between two people. The vertices have no attributes. Because we assume a communication to be mutual the graph is undirected. Edges are also unweighted.

Question: How do you detect misinformation in a communication network? Can one prevent the spread?

Research scenarios:

* Case 1: Misinformation starts from a highly connected node.
* Case 2: Misinformation starts from a node with a small degree.
* Case 3: Misinformation starts from a random node.

Possible tools for the analysis: SI model; SIR model (who should do fact checks?); Random Walk;

Network stats:

|  |  |
| --- | --- |
| Density | 0.00010344855751524997 |
| Maximum Degree | 623 |
| Minimum Degree | 1 |
| Average Degree | 3 |
| Assortativity | -0.38162654735878593 |
| Number of Triangles | 146976 |
| Average Number of Triangles | 4 |
| Maximum Number of Triangles | 1615 |
| Average Clustering Coefficient | 0.11268073251806623 |
| Fraction of Closed Triangles | 0.02241862617025909 |
| Maximum K-Core | 22 |
| Lower Bound of Maximum Clique | 12 |

Questions for session 1:

1. Do the cases we want to research represent “network-aware” models?
2. Should we drop one of our cases?
3. What model would be best? SIS, SIR or SEIR? Do we want to test different models with the different cases.
4. How we can define a highly connected node in the context of our graph?
5. How long should our presentation be?

Fact-Checking Behaviors

* In 2021, 47% of people aged 16-74 years in the EU encountered untrue or doubtful information on news websites or social media during a 3-month period

[1](https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20211216-3" \t "_blank)

.

* However, only 23% of people actually verified the truthfulness of the information or content they encountered

[1](https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20211216-3" \t "_blank)

.

Upload the synthetic graph

Explain the SIR

Show the images

Parameters

Is the graph good?

Hi Michele!

Our group has some questions concerning the NA project that might be worth investigating before the next supervision on Friday. Our network is the one with the email data of a European research institution.

First, we decided to follow Luca's advice from last week and generated a synthetic network with the same number of nodes and edges. You can check the attached CSV file to make sure it is correct.

Second, we came up with conditions for our misinformation spread simulation. It is based on the SIR models however we augmented the recovering conditions according to the feedback from the supervision. The difference is that we have a threshold that does not allow any recovery before the information spreads to at least 5% of the network. Once that point is reached there is a 23% chance for an infected node to recover by itself (the number is selected according to statistics for the EU). Additionally, the recovery rate has a 100% success rate but it transmits only through the neighbors of the recovered nodes. We hope that this model sounds feasible for misinformation spread.

So, given all those modifications, we wanted to ask if we should make a comparison between the original network and the synthetic one. We also have some plots of the simulations for our different case scenarios listed in the project description.

Thank you for your time!

Dimitar

Questions for session 2:

1. We plan to research how the threshold value affects the maximum infection and the time for recovery.
2. Is our recovery rate realistic?
3. Is the difference in our cases significant or not?
4. Should we visualize the synthetic graph?

Diam in org – 9, synt-21

Tasks:

1. Test thresholds, infection rates
2. Visualize the original graph
3. Comparison between cases
4. Histograms for infection in n simulations

Presentation:

1. Show stats for the original network
2. Visualization
3. Hypothesis: we need to target a high-degree node
4. Introduce SI model on the original
   1. Show the different cases
   2. Explain the infection rate influence – we chose 0.5
5. Introduce SIR model on the original
   1. Show the different cases
   2. Explain the threshold influence – we chose 0.03
6. Show the synthetic graph
   1. Compare the spread
7. Conclusion
   1. Further discussion
   2. Limitations