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?