Spreading of computer viruses in a social network setting

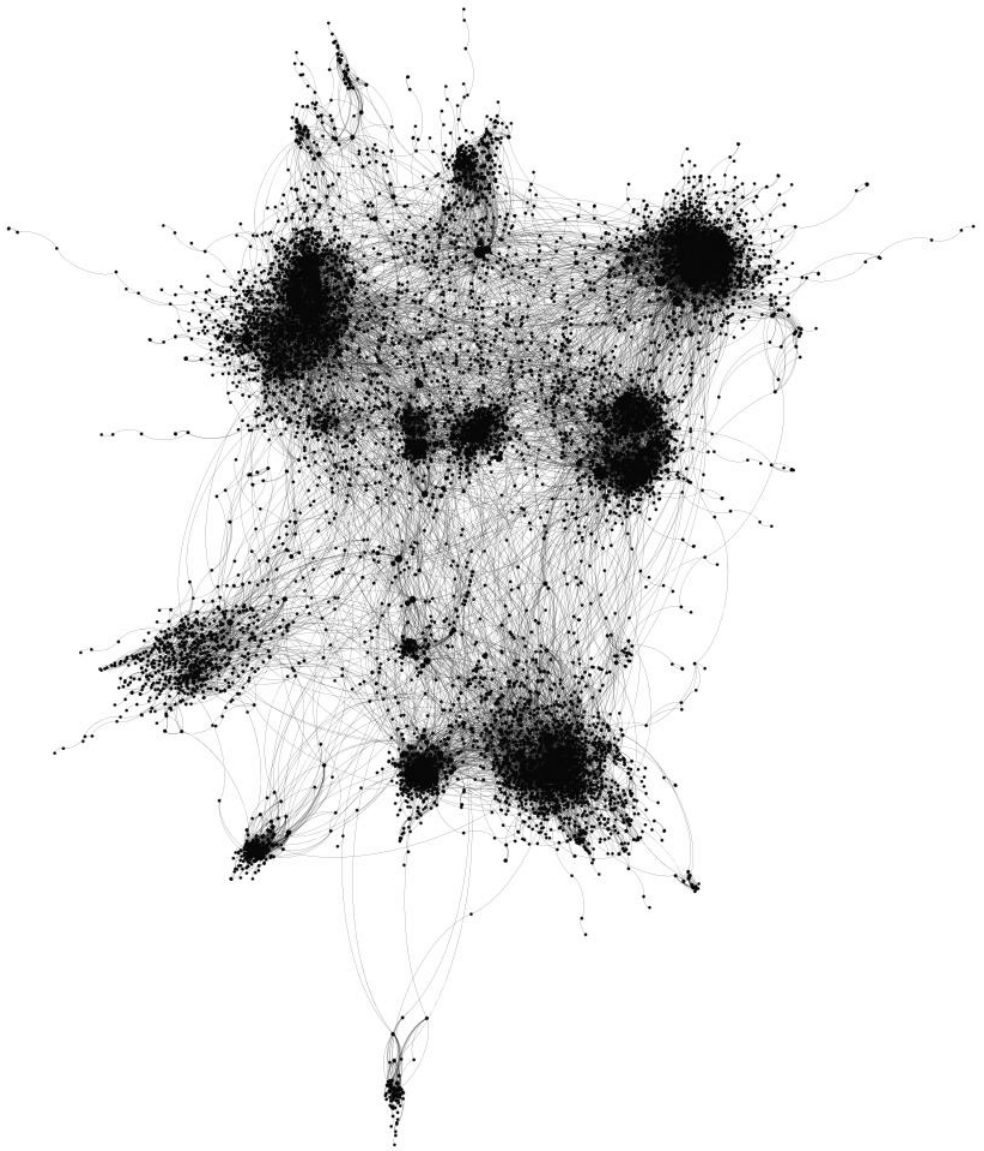
Bozhidara Pesheva

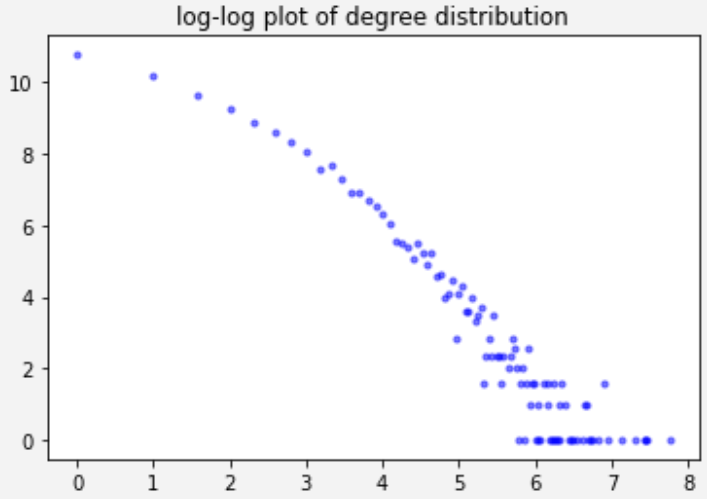
Gabriela Zhelyazkova

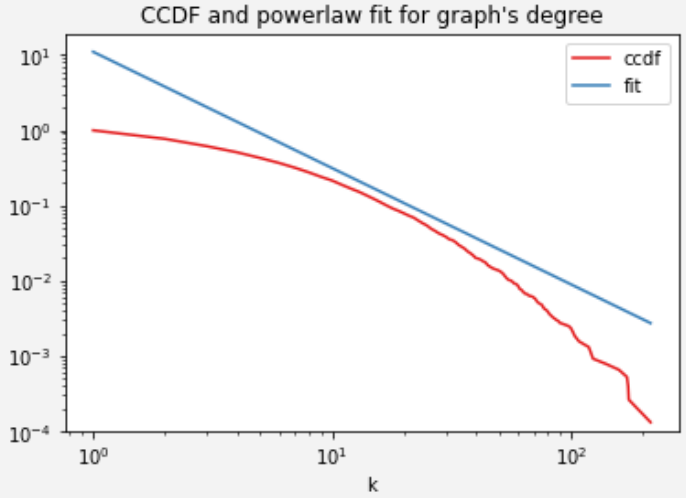
Lili Raleva

The network

- A unipartite, undirected social network of LastFM users
- Data - collected from the public API in March 2020
- Nodes (users) = 7624
- Edges (neighbour connections) = 27,806
- In the context of LastFM: *neighbours* are people with similar music tastes, and two individuals can follow each other *only if* they listen to identical artists.





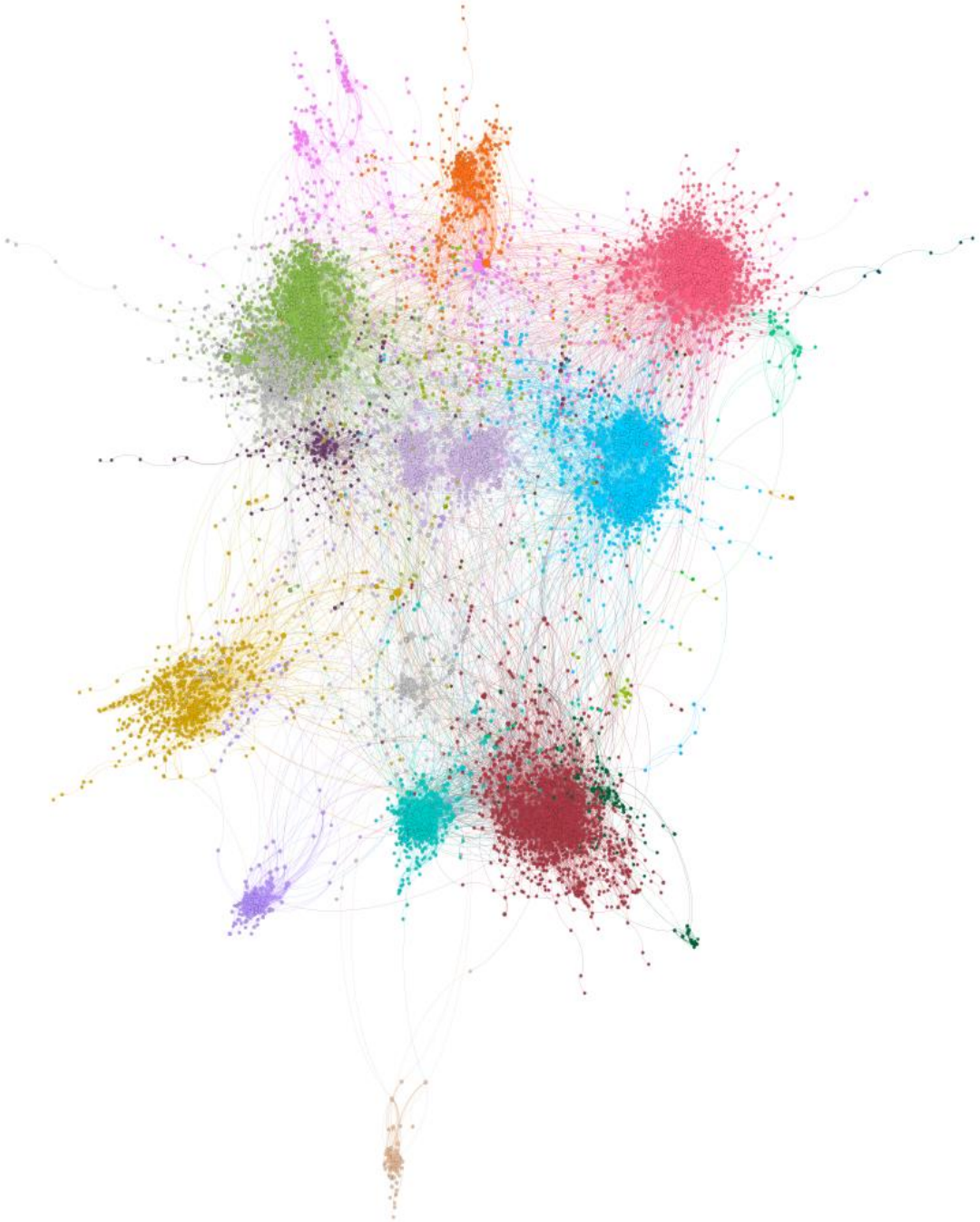


Exploratory Data Analysis Figures

- Average path length = 5,232
- Average degree = 7,294
- Network diameter = 15
- Graph density = 0,001
- Average clustering coefficient = 0,285
- Graph transitivity = 0,179
- Global clustering coefficient = 0.22
- Assortativity= 0.017

The RQ

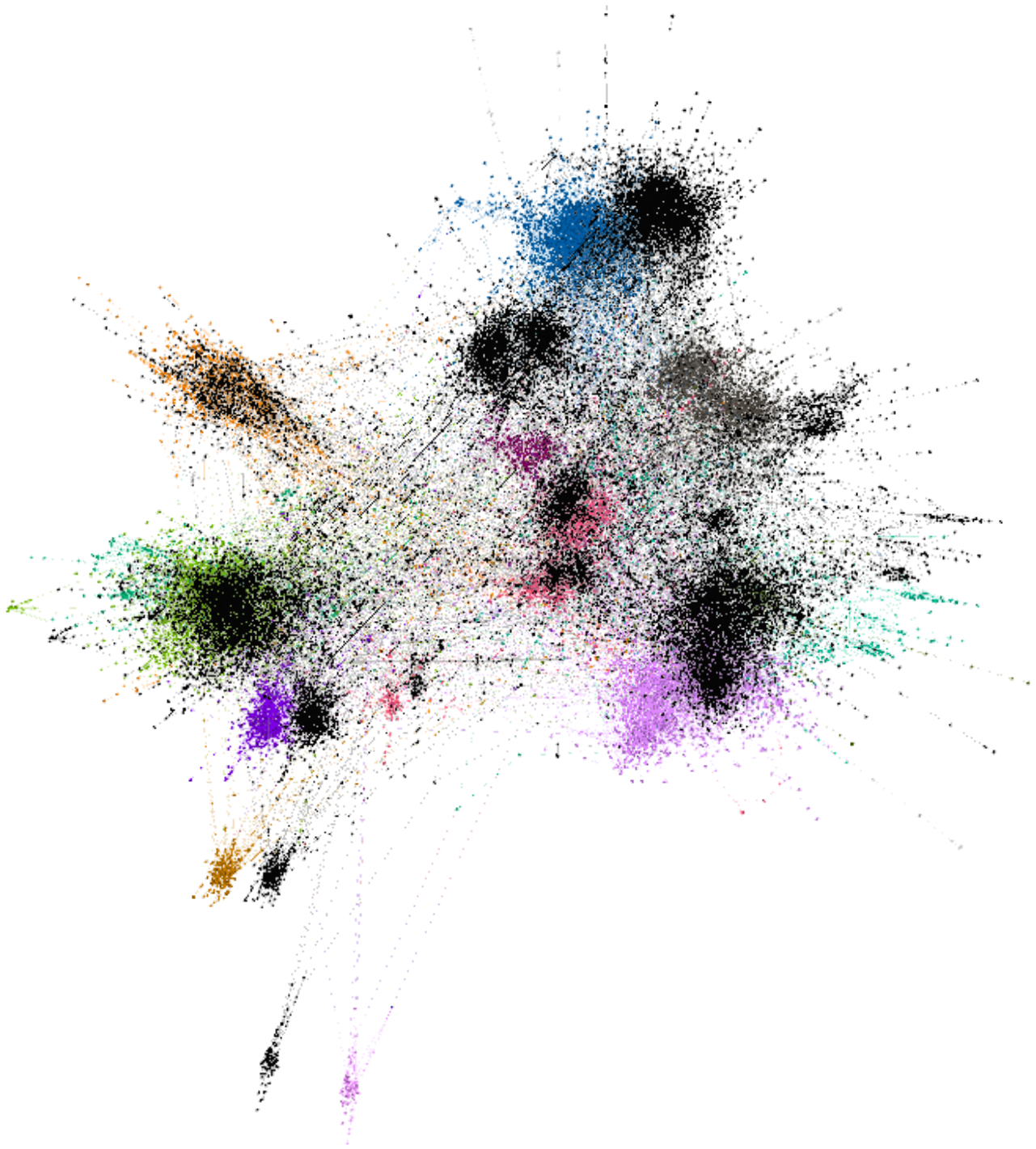
Does the size of the starting node's community affect the speed of the computer virus spread?



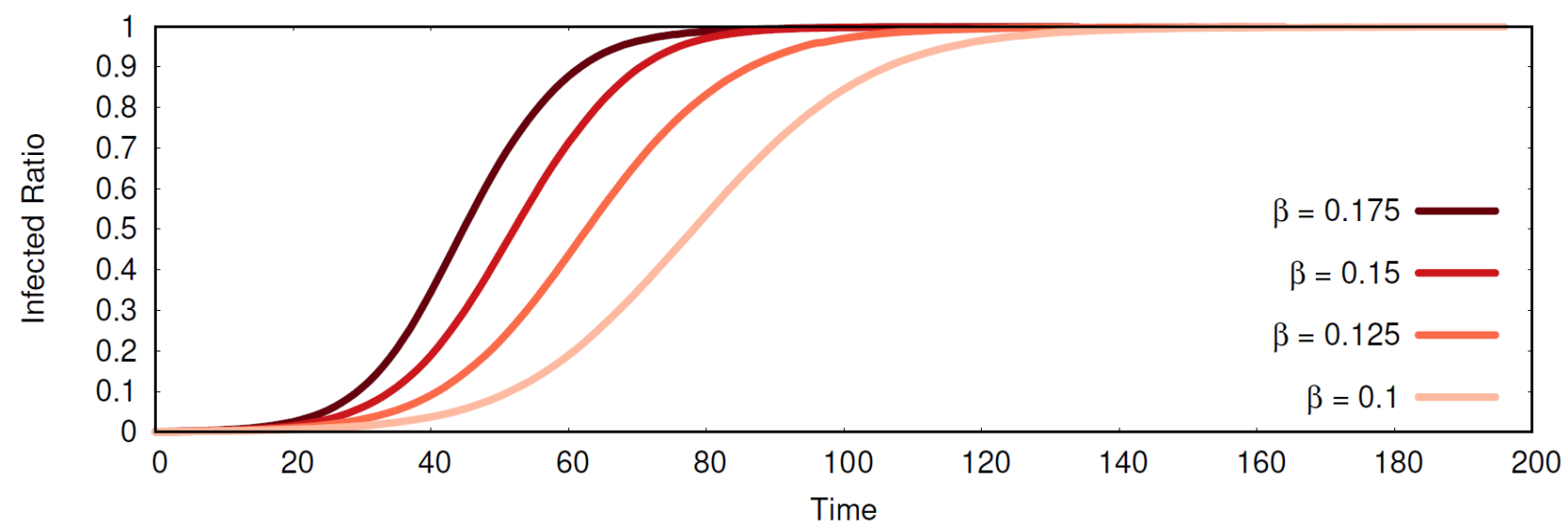
Starting nodes and how we choose them

- Communities – by definition and how they are in our case
- Graph modularity ≈ 0.8
- Choice of the starting node community
- Degree or betweenness centrality

The same nodes in our case



SI



- SI models diseases with **no recovery.**
- Always **in the end the whole network will be infected.**
- Parameter of the model is **infection rate - β**

https://www.networkatlas.eu/

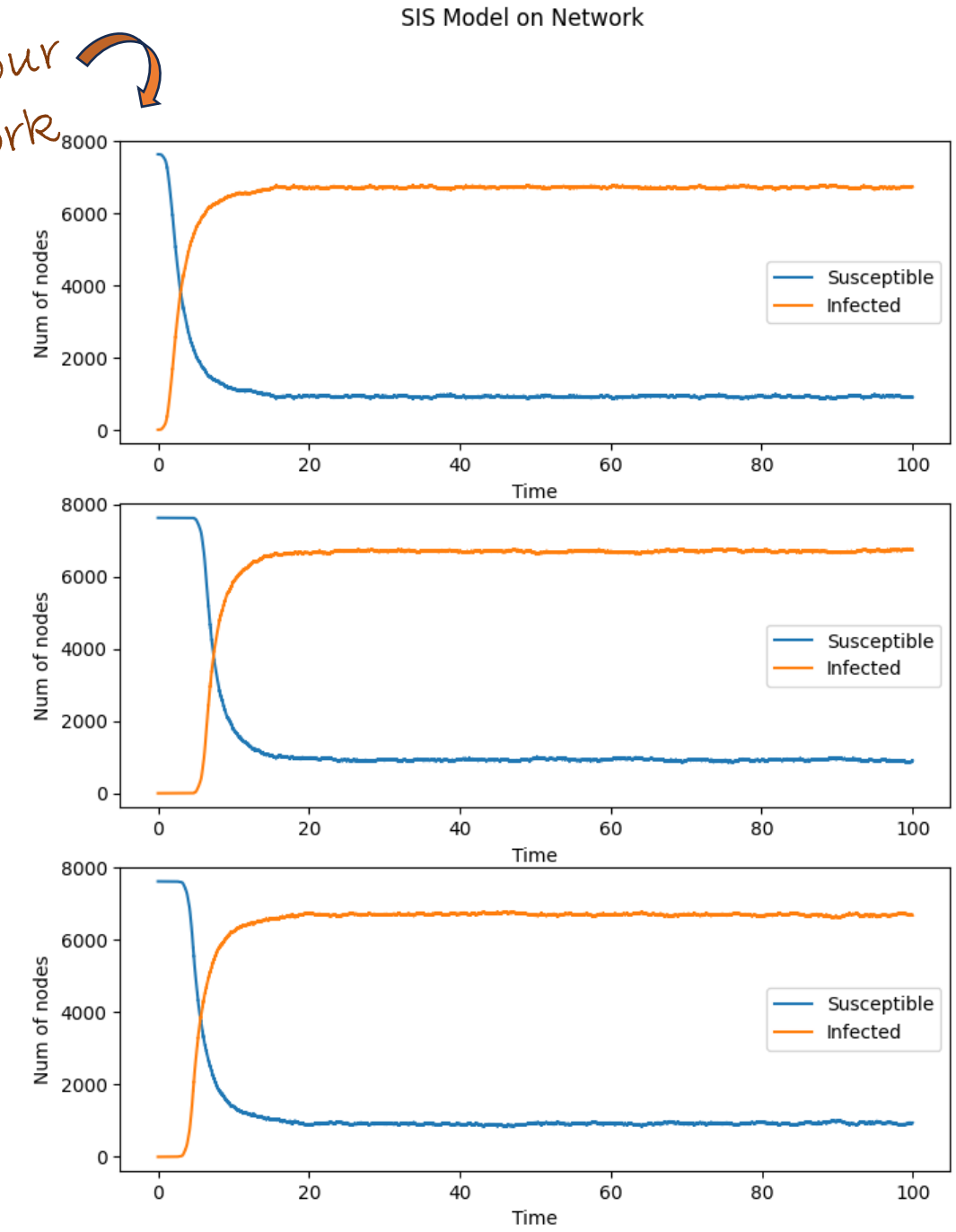
SIS

- Parameters:
- **β** - infection rate
- **μ**- recovery rate

- S→I transition
- I→S transition

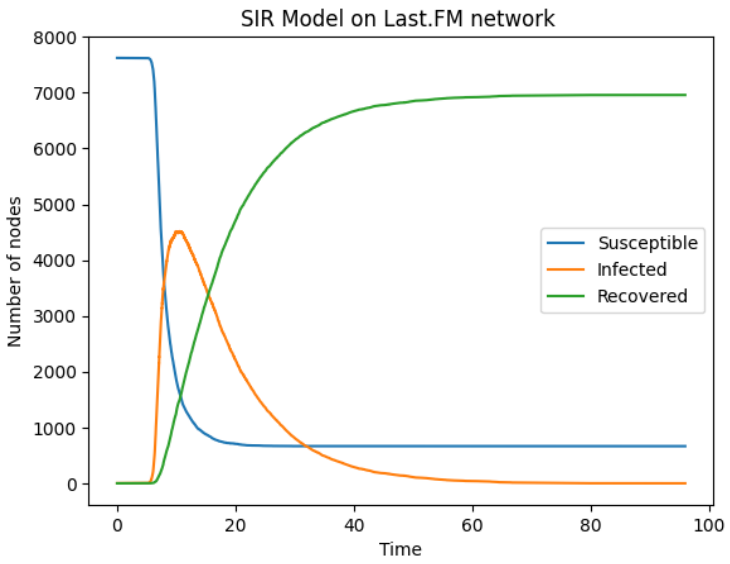
- There is possibility of recovering and becoming infected again.
- **Endemic state -** number of people recovering is perfectly balanced by the new infected.
- Possibility of eradicating the disease.

From our network

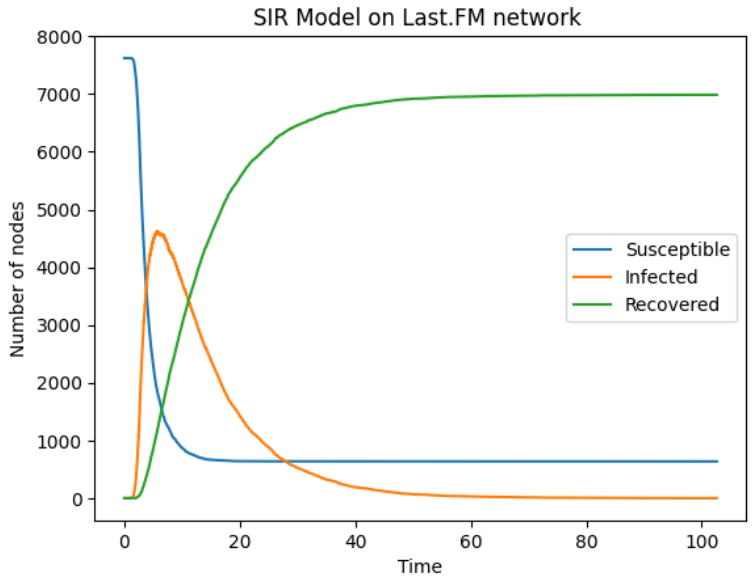


SIR

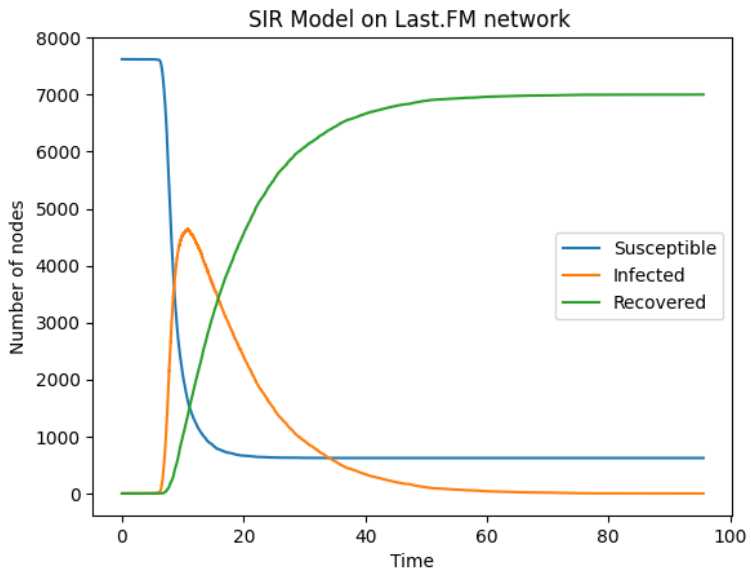
Our choice



#7237 – largest community



#3571 – middle-sized community



#5851 – small community

Parameters:

- $\beta = 0.3$ S→I transition
- $\mu = 0.1$ I→R transition

But why are the plots so similar...?

But how did we chose our parameters...?

Node 7237

% of removed notes after the infection has died down			
	Infection rate		
Recovery rate	0.2	0.3	0.4
0.05	0.93494	0.94937	0.96642
0.1	0.87933	0.91973	0.93048
0.15	0.82581	0.88103	0.90805

# of steps needed for the spread to stabilize (die down)			
	Infection rate		
Recovery rate	0.2	0.3	0.4
0.05	160	171	210
0.1	102	105	107
0.15	63	68	69

Node 3571

% of removed notes after the infection has died down			
	Infection rate		
Recovery rate	0,2	0,3	0,4
0,05	0.86	0,89	0,91
0,1	0,87	0,91	0,92
0,15	0.82	0,87	0,91

# of steps needed for the spread to stabilize (die down)			
	Infection rate		
Recovery rate	0,2	0,3	0,4
0,05	190	200	210
0,1	100	94	100
0,15	65	67	70

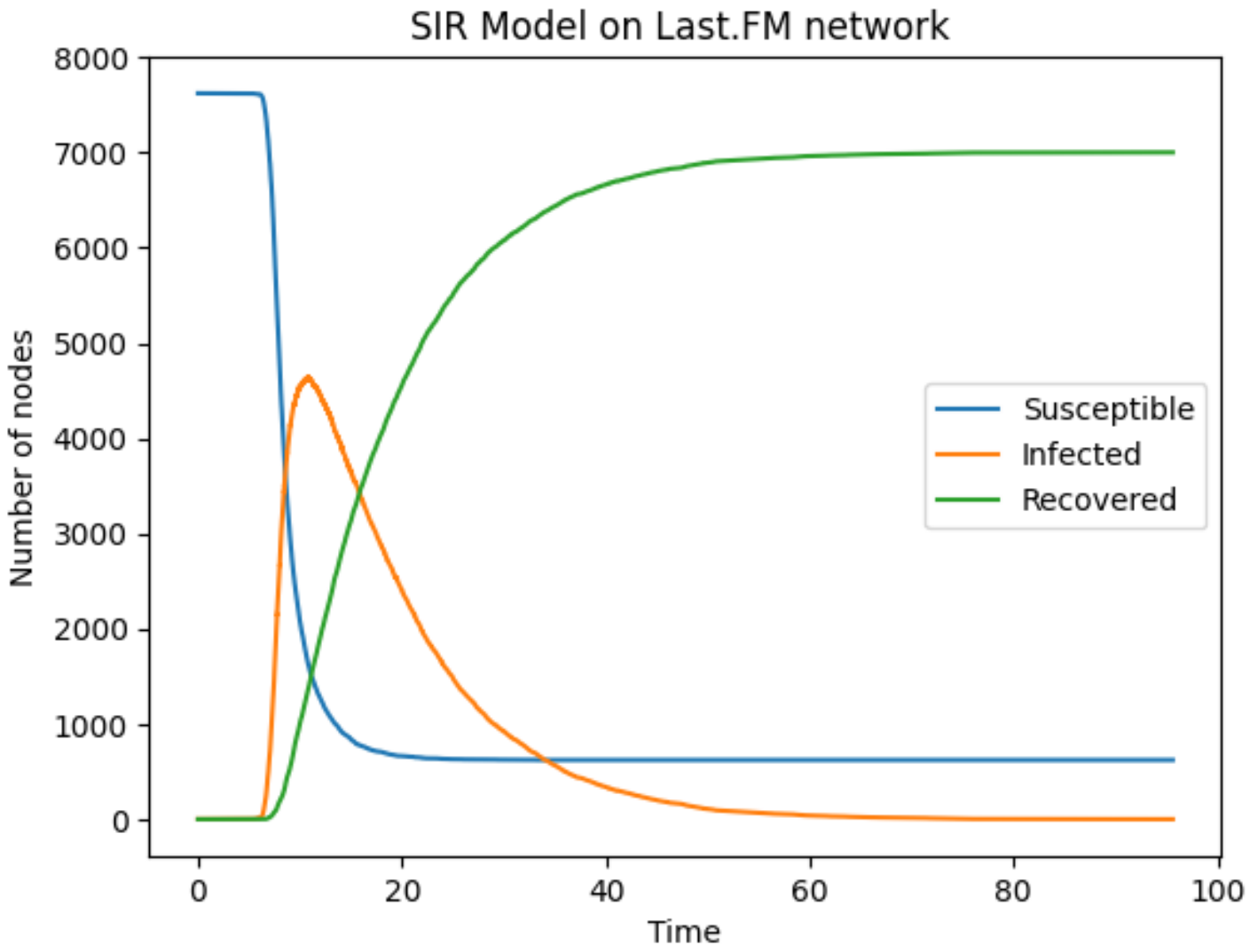
Node 5851

% of removed notes after the infection has died down			
	Infection rate		
Recovery rate	0.2	0.3	0.4
0.05	0.93652	0.94898	0.96459
0.1	0.88156	0.90766	0.9297
0.15	0.81584	0.87684	0.90989

# of steps needed for the spread to stabilize (die down)			
	Infection rate		
Recovery rate	0.2	0.3	0.4
0.05	186	216	209
0.1	109	112	92
0.15	83	69	69

Analysis of the results

- Disease dies out before everyone has been infected.
- Very fast spread. Highest number of infected is more than half of the nodes in around 3-4 time steps.
- Always more Recovered than Susceptible in the end.



Conclusion

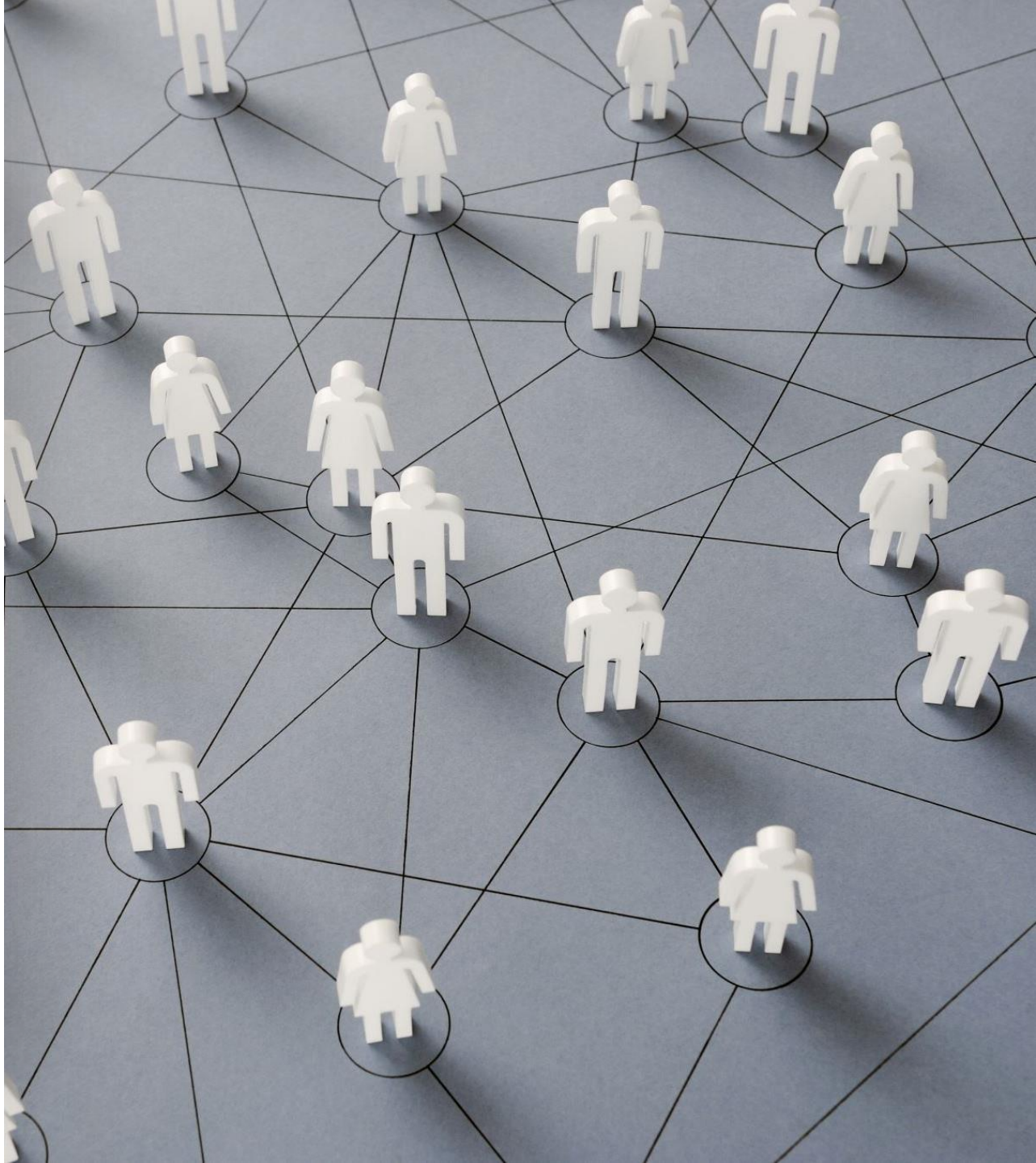
- No significant change in the speed starting from different sized communities.
- Constant ≈ 91% Infection ratio => the starting node is not an influential factor when it comes to the scale or speed of the virus' spread.
- General conclusion:

Based on the results, we accept that the size of community of the starting node is not strongly affecting the speed or the spread of the virus.

Did Last.FM make the correct decision to hide all and any URLs on the platform?

Limitations and further research

- Since friendships networks are constantly evolving, the drawn conclusion is case-specific for the network screencap at our disposal.
- The chosen model parameters could be tuned in accordance with a specific type of virus.
- Use of overlapping communities.



Thank you for your attention!

BACKUP SLIDE
Starting nodes and how we choose them

- Communities = in our case every community detected by the Louvain community detection algorithm and consist of at least 1% of all nodes

  => everything else considered as node not part of a community.
- Considering a node from the largest and the smallest communities, and a node outside a community.
- Degree and betweenness centrality.
- The same nodes with highest degree are with highest betweenness centrality in our case.