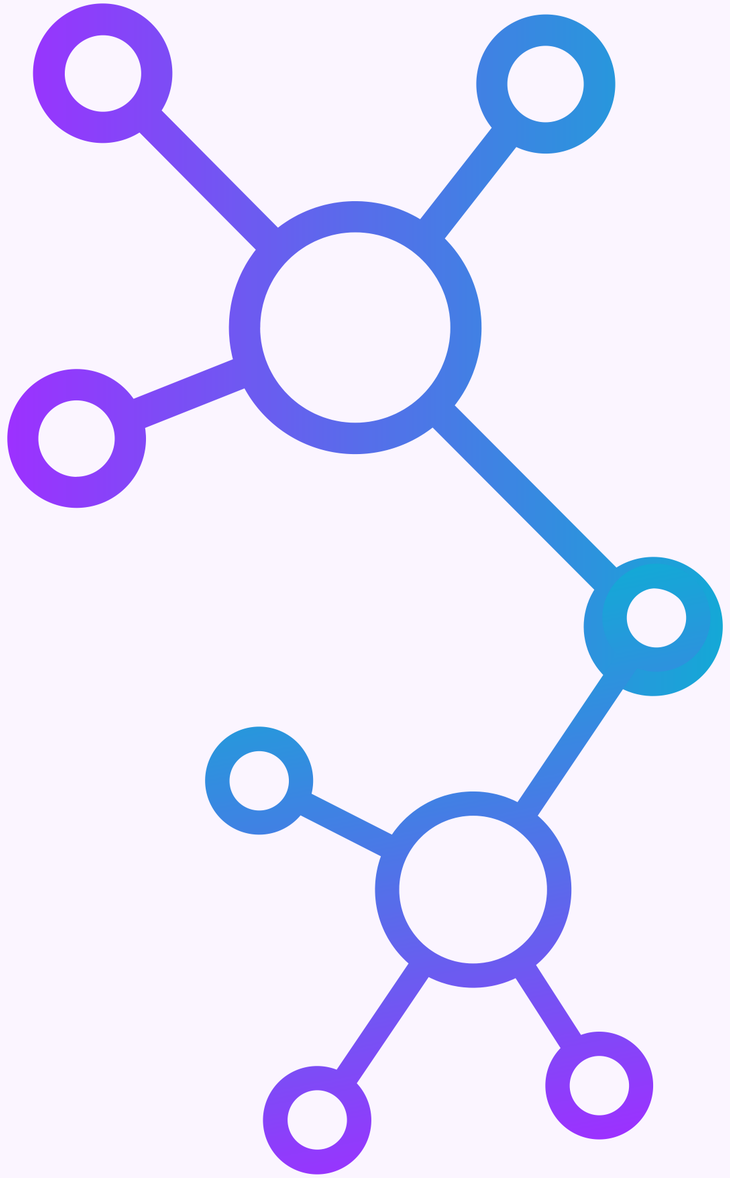
DATA VISUALIZATION PROJECT GROUP A

How much are we connected?

A graph theory study applied to Facebook



Roadmap

Datasets description

Overwiew of the datasets

Network Analysis

All useful measures to better understand the structure of the network

Research Question

1) Similarity across comunities

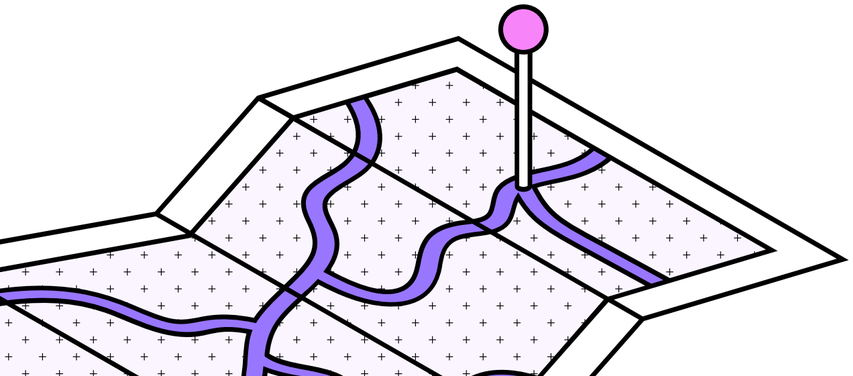
2) Information spreading in social network communities

Network Visualization

With Python and Gephi

Conclusions

Findings and possible future developments



Dataset

Ten Facebook users and their friends, the relationship between them and their anonymized features

Dataset variables

Circles:
describes the relationships between features of the vertex v's friends (circles)

Edges:
edges for the network of vertex v

Featnames:
all the anonymized features' names used in the following data set "egofeat" and "feat"

Egofeat:
features belonging to the ego node (0 stands for "feature does not belong to the node", 1 instead "feature belongs to the node"

Feat:
same format as the egofeat data set but in this case it describes the features of the nodes friends of vertex v.

Network analysis report

with **Python** and **Gephi**

Network overview

The Network is composed by 4039 **nodes** and 88234 **edges**

Network Average Clustering coefficient:
0.6055467186200876
Network Average distance:
3.6925068496963913

Degree report

Simple average degree of the Network is 44,131

Modularity report

The community identified are 16. The **modularity** of the graph is 0,722. This measure indicates that the Network have dense connections between nodes within the same community but sparse connections between nodes in different community.

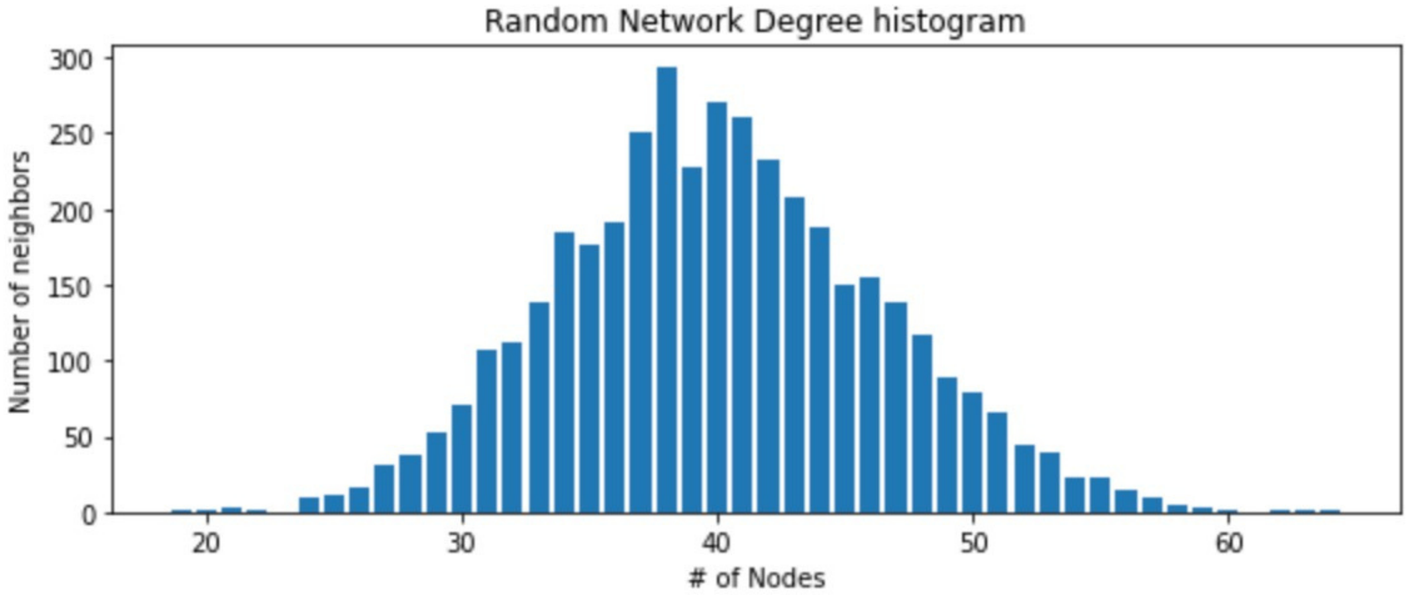
Graph distance report

The **Diameter** of the Network is 8, so the two most distant nodes are separated by 8 "moves".

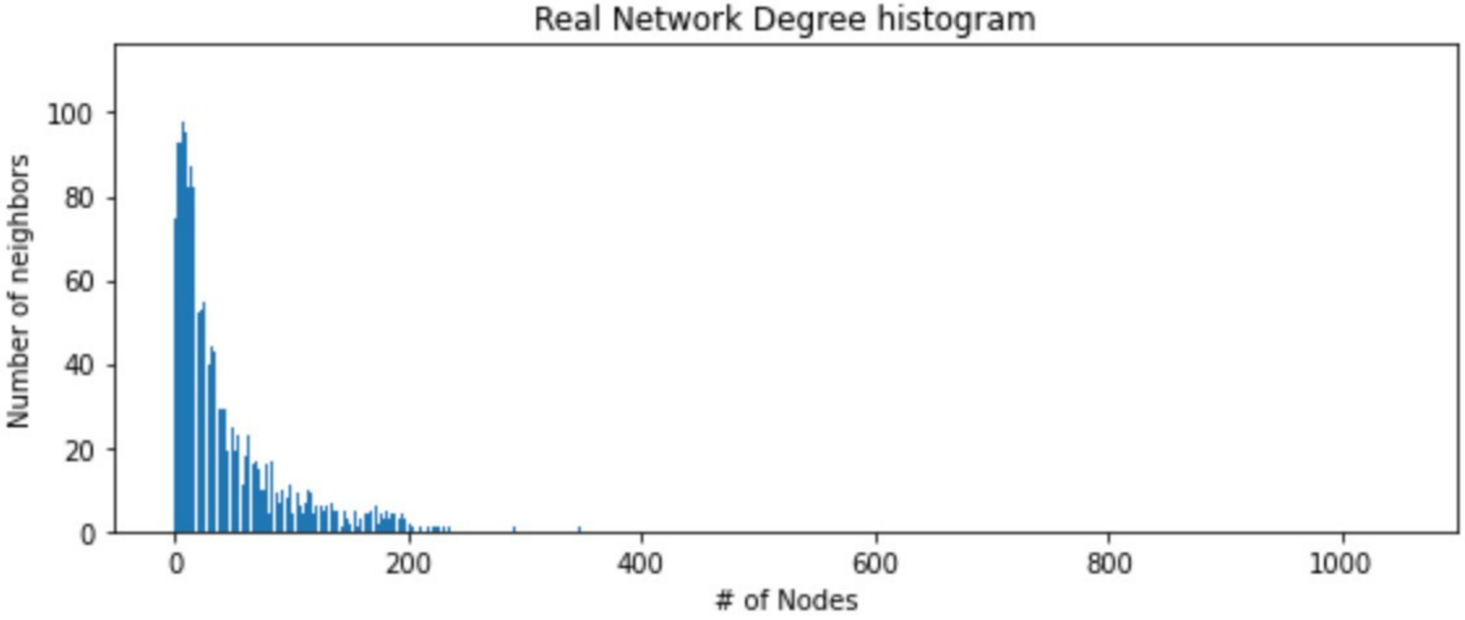
Graph density report

The **density** of the Network is 0,011. The network is consequently not very dense, making its network effects weaker.

Random Graphs
vs.
Real Graphs



Normal Distribution



Power-law Distribution

Communities
in a Social Network

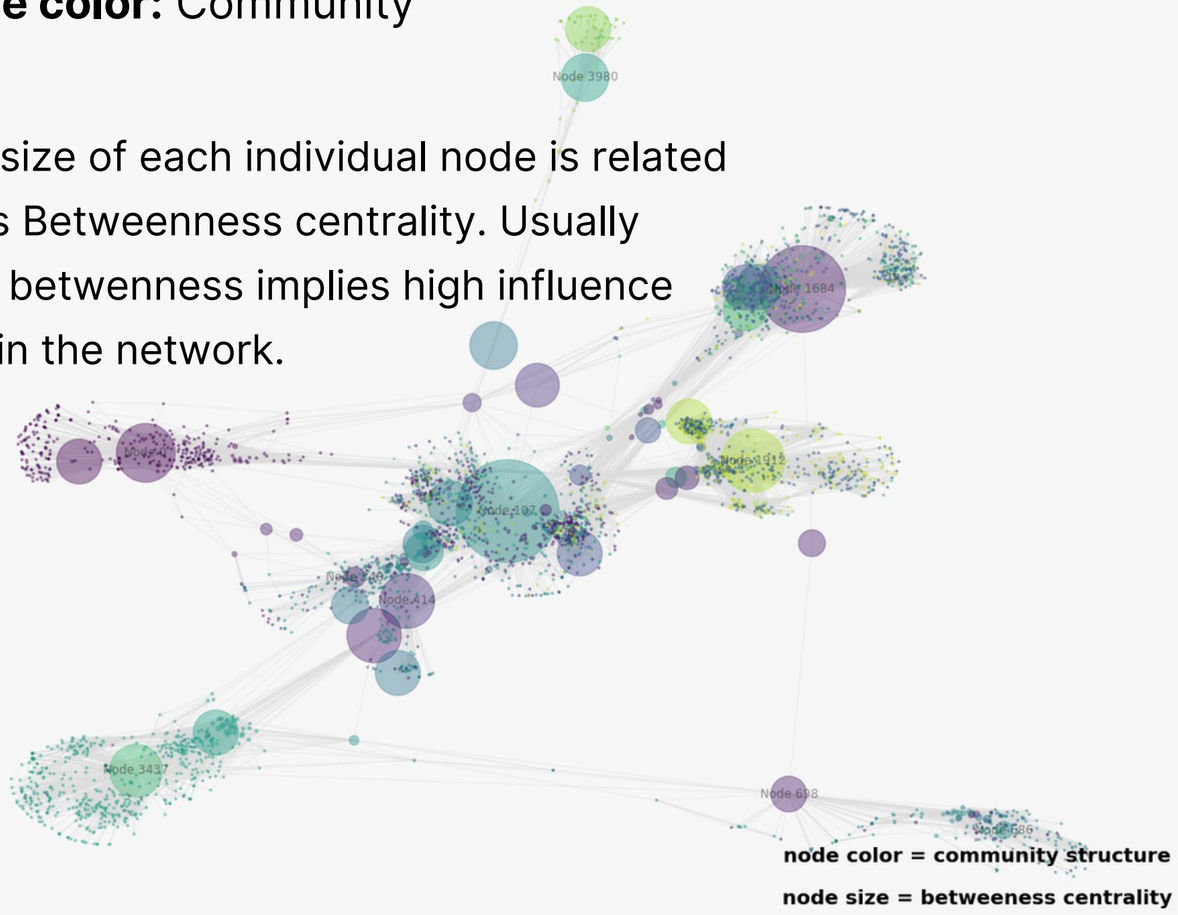
By optimizing modularity, a measure which measures the strength of division of a network into modules, it is possible to group nodes into different sets such that each set of nodes is densely connected internally.

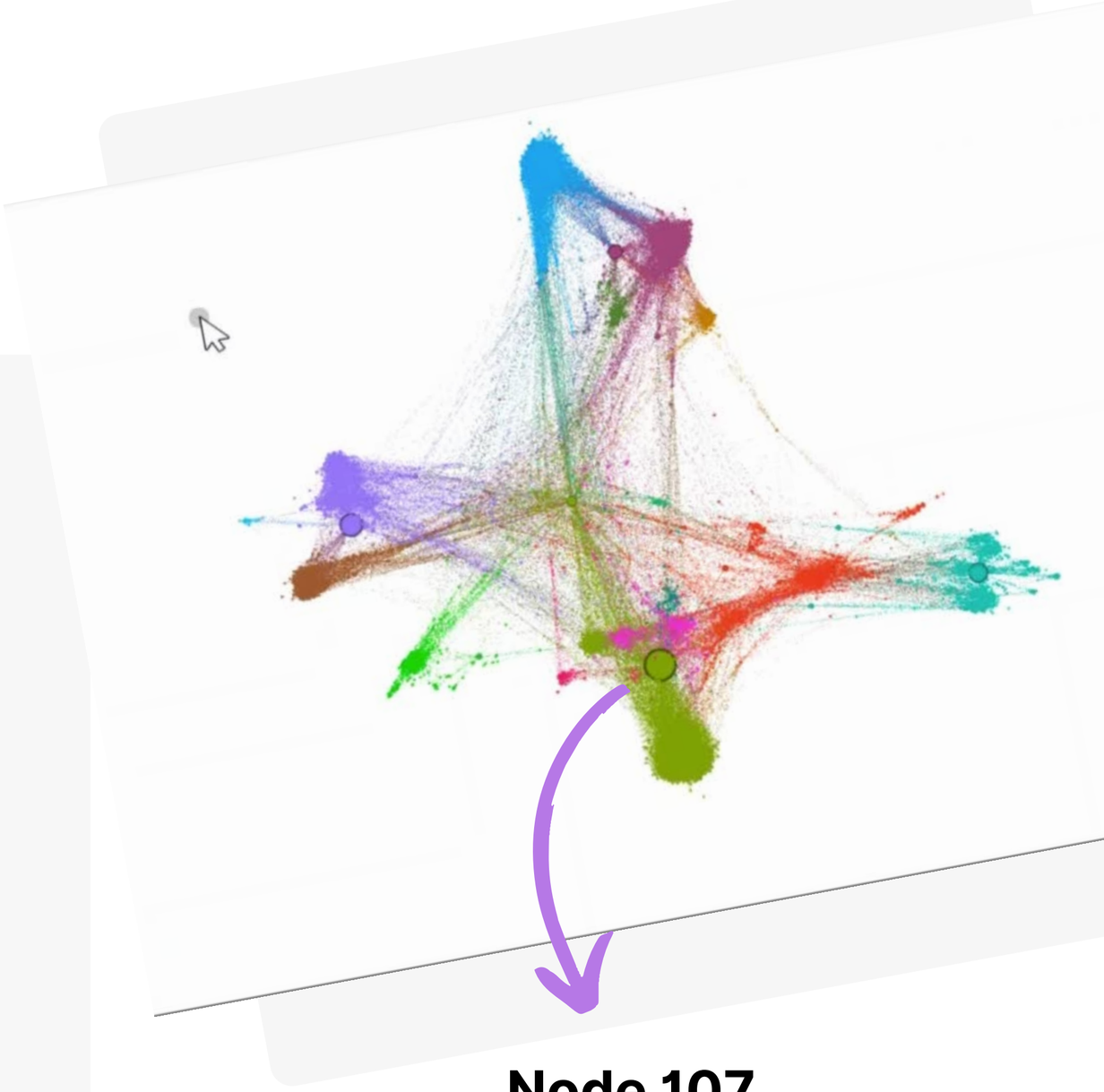
Network analysis

Gephi

Node size: Betweenness centrality
Node color: Community

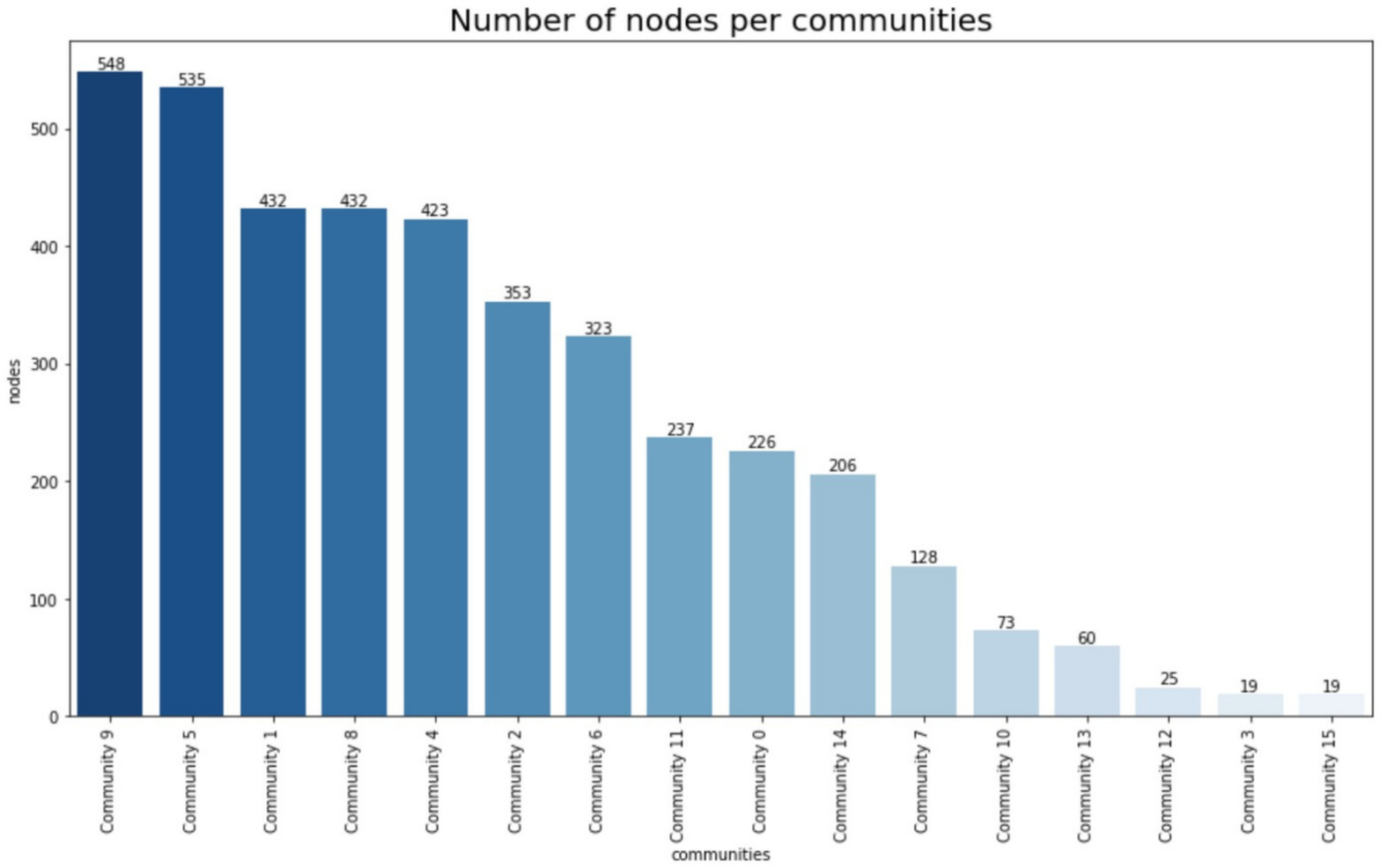
The size of each individual node is related to its Betweenness centrality. Usually high betwenness implies high influence within the network.





Node 107

How many nodes per community?



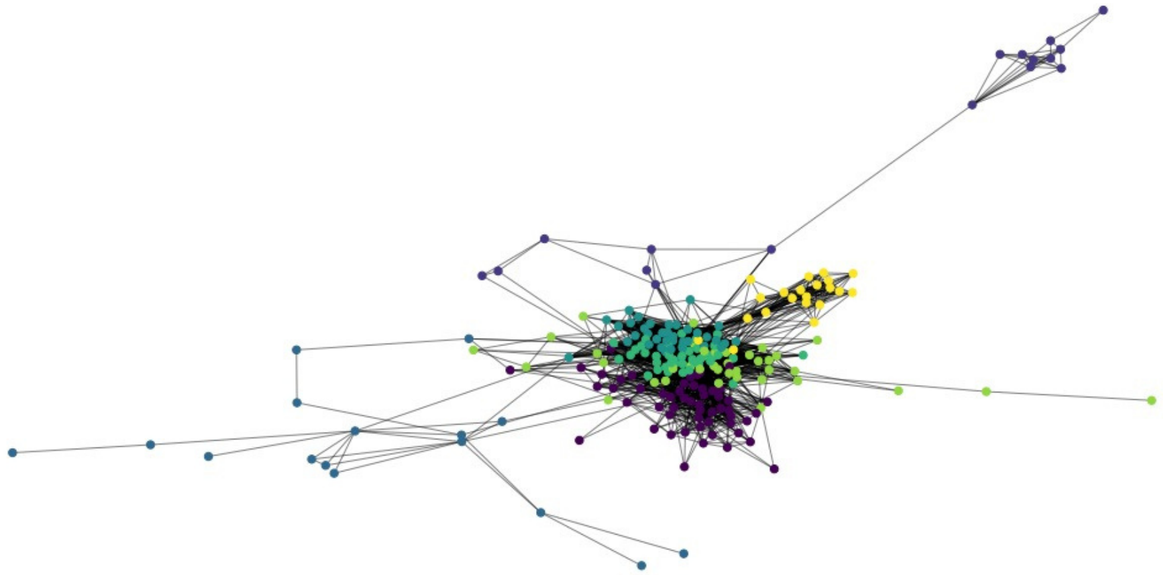

```
0 is in community number 0
107 is in community number 9
3980 is in community number 13
3437 is in community number 10
686 is in community number 14
1684 is in community number 2
1912 is in community number 4
698 is in community number 14
348 is in community number 1
414 is in community number 1
```

Both node 348 network and 414's one belong to the same community (number 1), the same occurs for 698 and 686

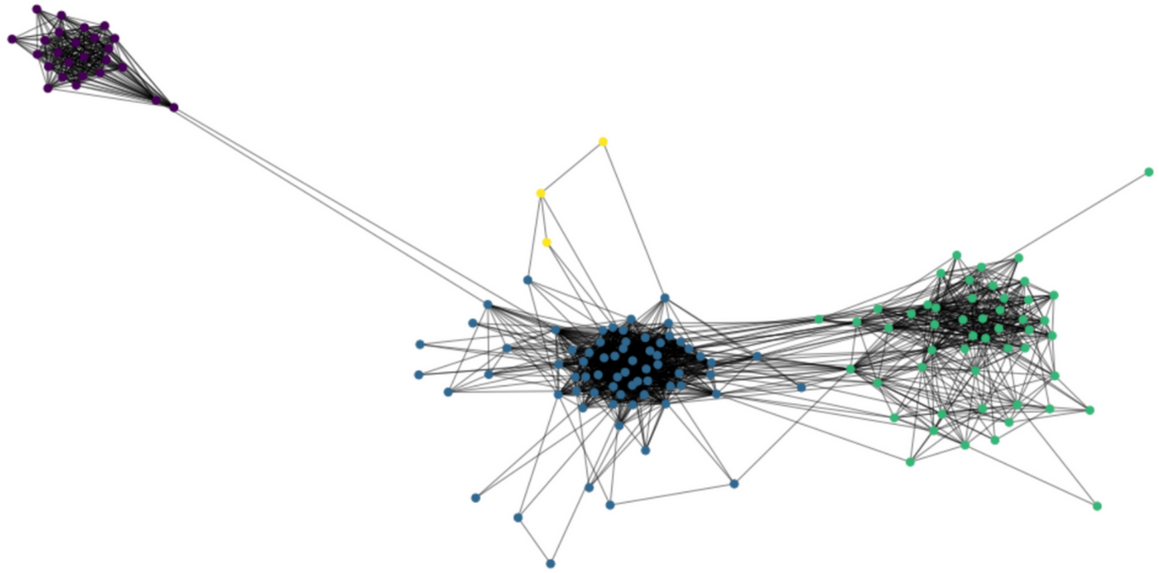
Interesting Insight - Part 1

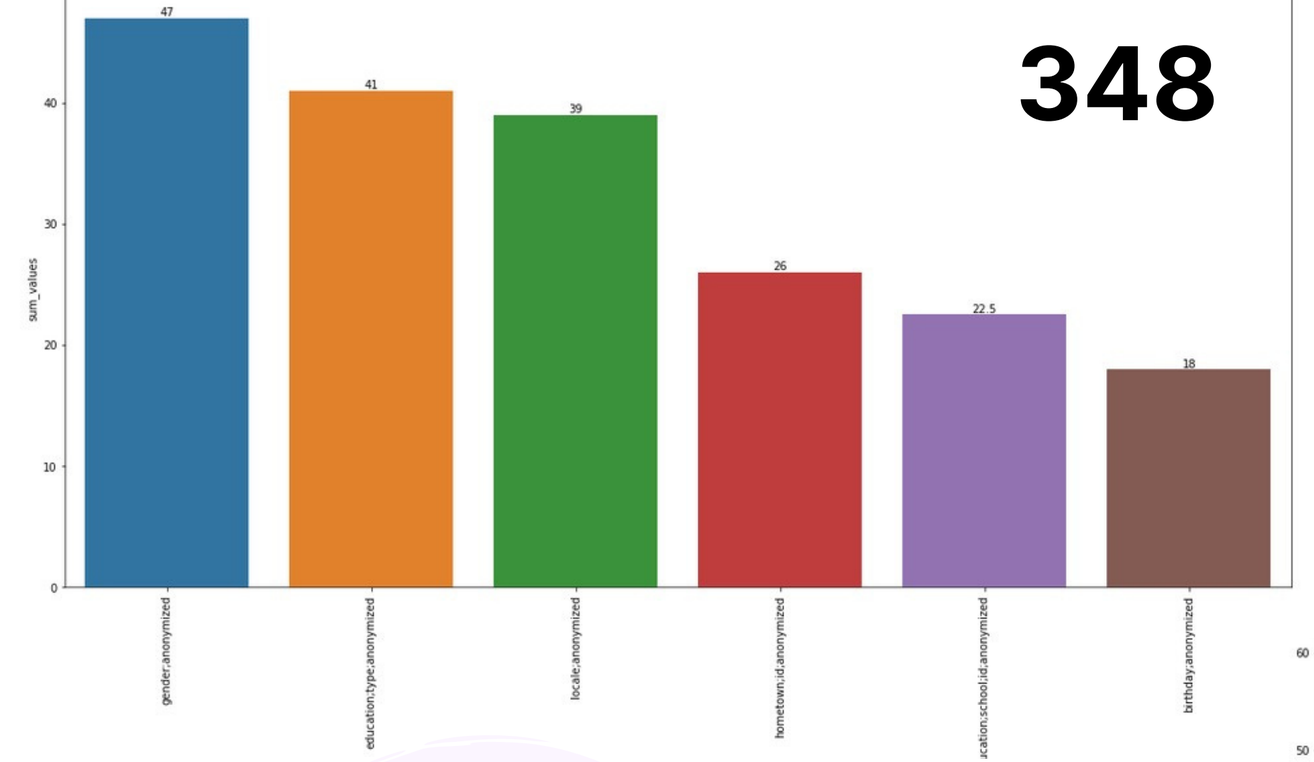
Feature Analysis

Node 348 Network



Node 414 Network



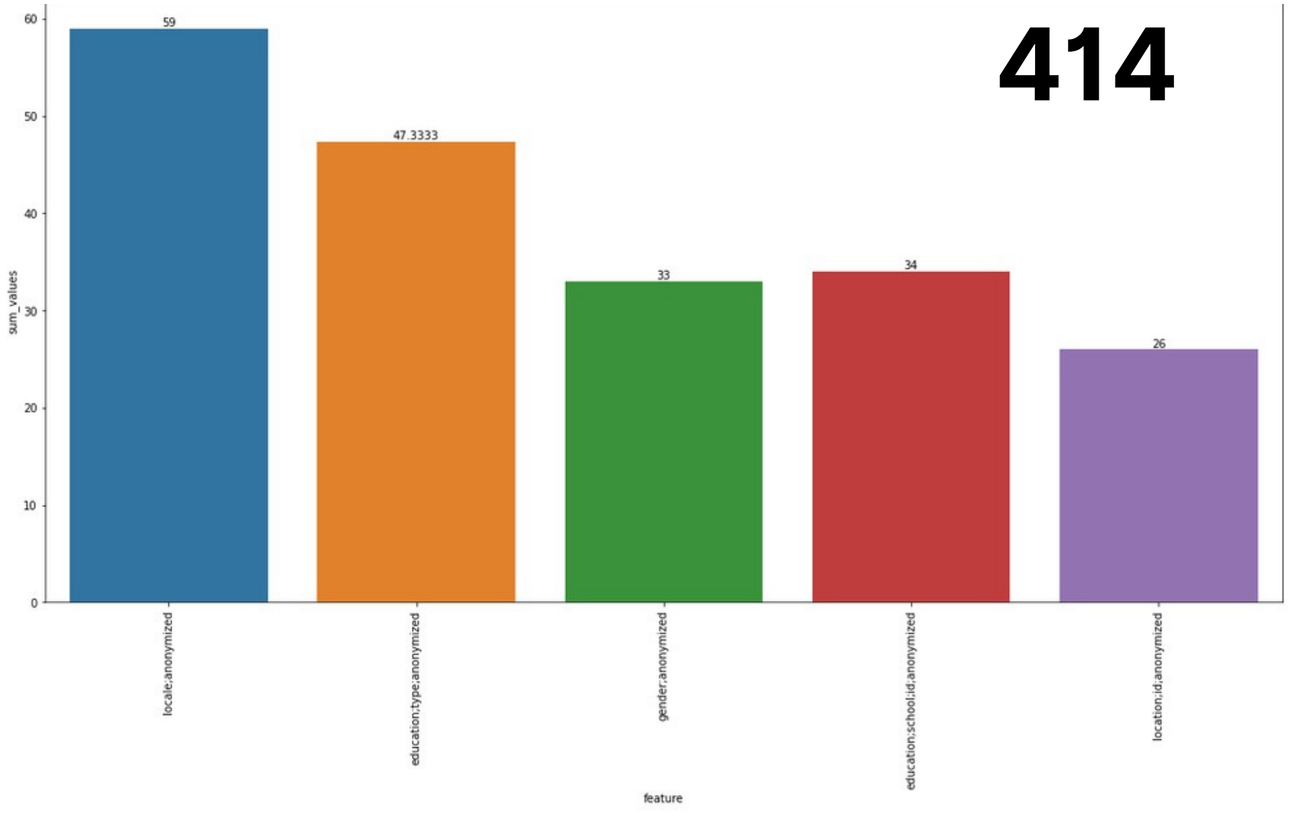


348

1. Gender
2. Education type
3. Locale
4. Hometown id
5. Education School
6. Birthday

HYP:
nodes' network belonging to the same community have the same recurring features

1. Locale
2. Education type
3. Gender
4. Education School
5. Location id



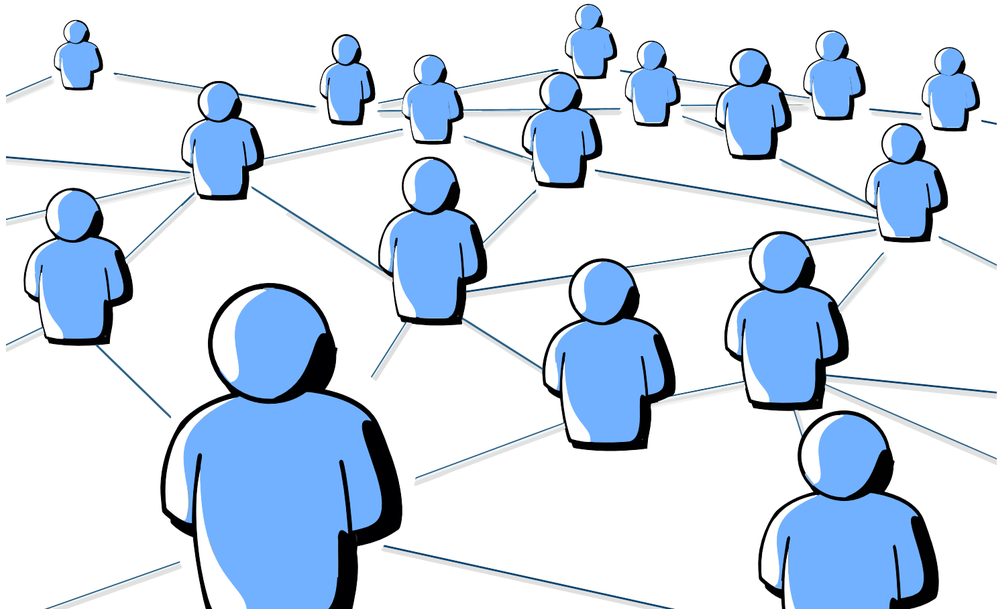
414

Interesting insights

- Part 2

The speed of information in social networks

Nowadays misinformation is one of the most important negative externalities of social networks.
Some assumption have been made for this study:



- Since the network was unweighted it was assigned a weight to each edge obtained by the average between source node's one centrality and node's two one
- The information spread in this case will be defined as ”secret”.
- Only two people know the secret at the beginning of the infection process

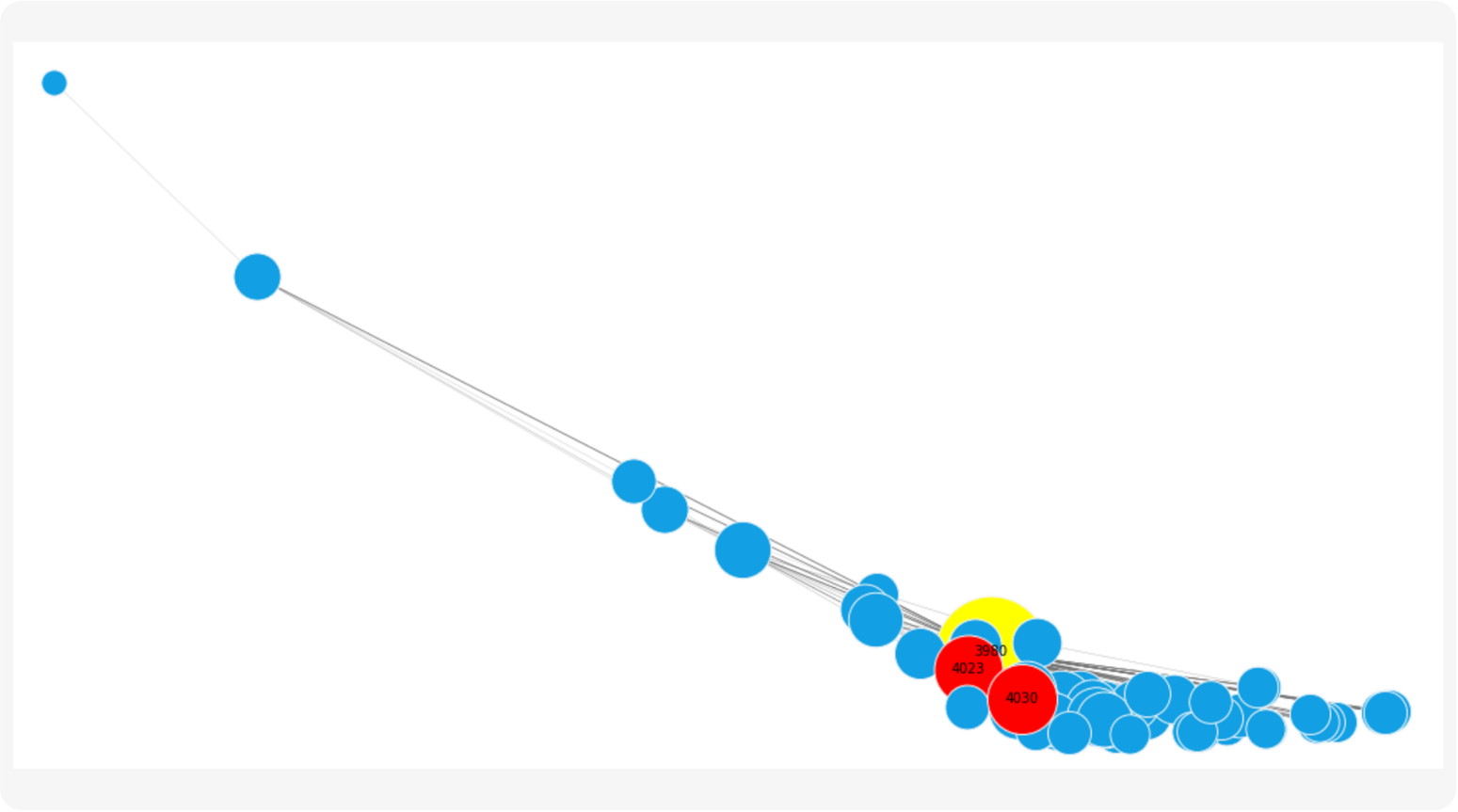
Main actors

In the information spread process

The secret keeper: Actors that initially know the secret. the ***red nodes*****.**

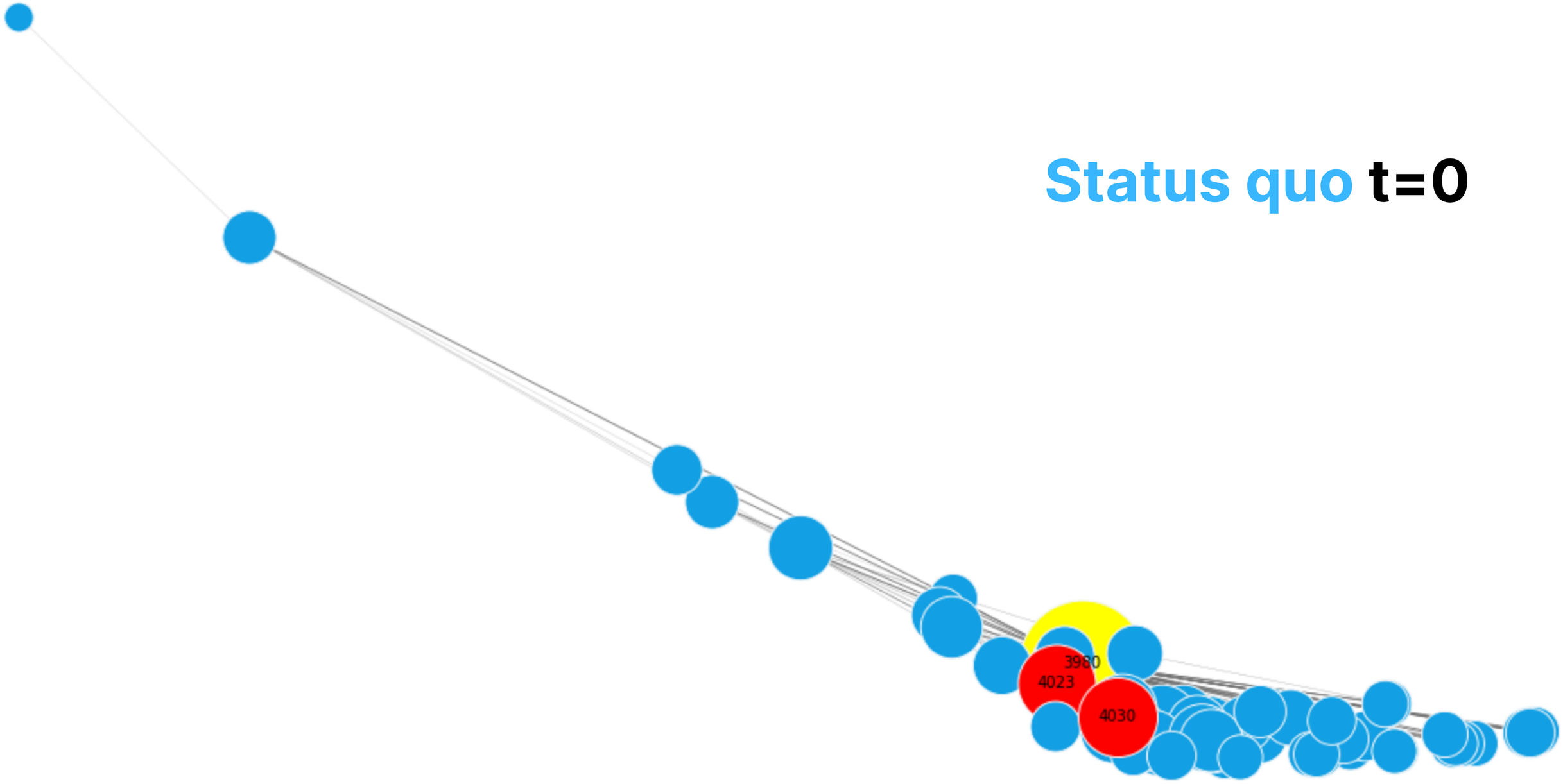
The confidant: the first node to which the secret keeper passes the information. Is represented by the ***yellow node***. Once the information spread through other nodes, they also become yellow.

The unaware: all the nodes that do not know the secret. They are colored ***blue.***

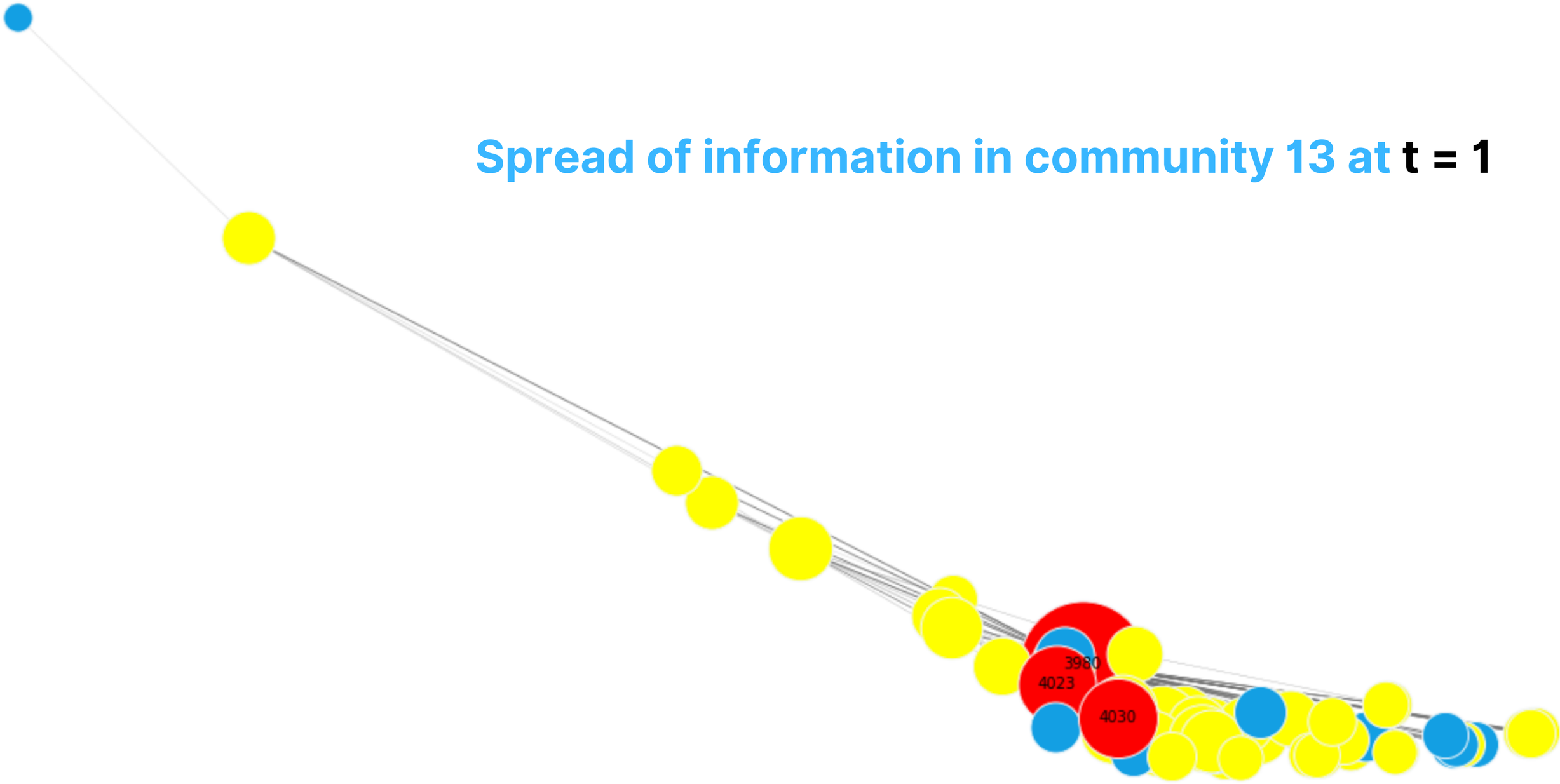


Spread of information in community 13 at t = 0

Status quo t=0

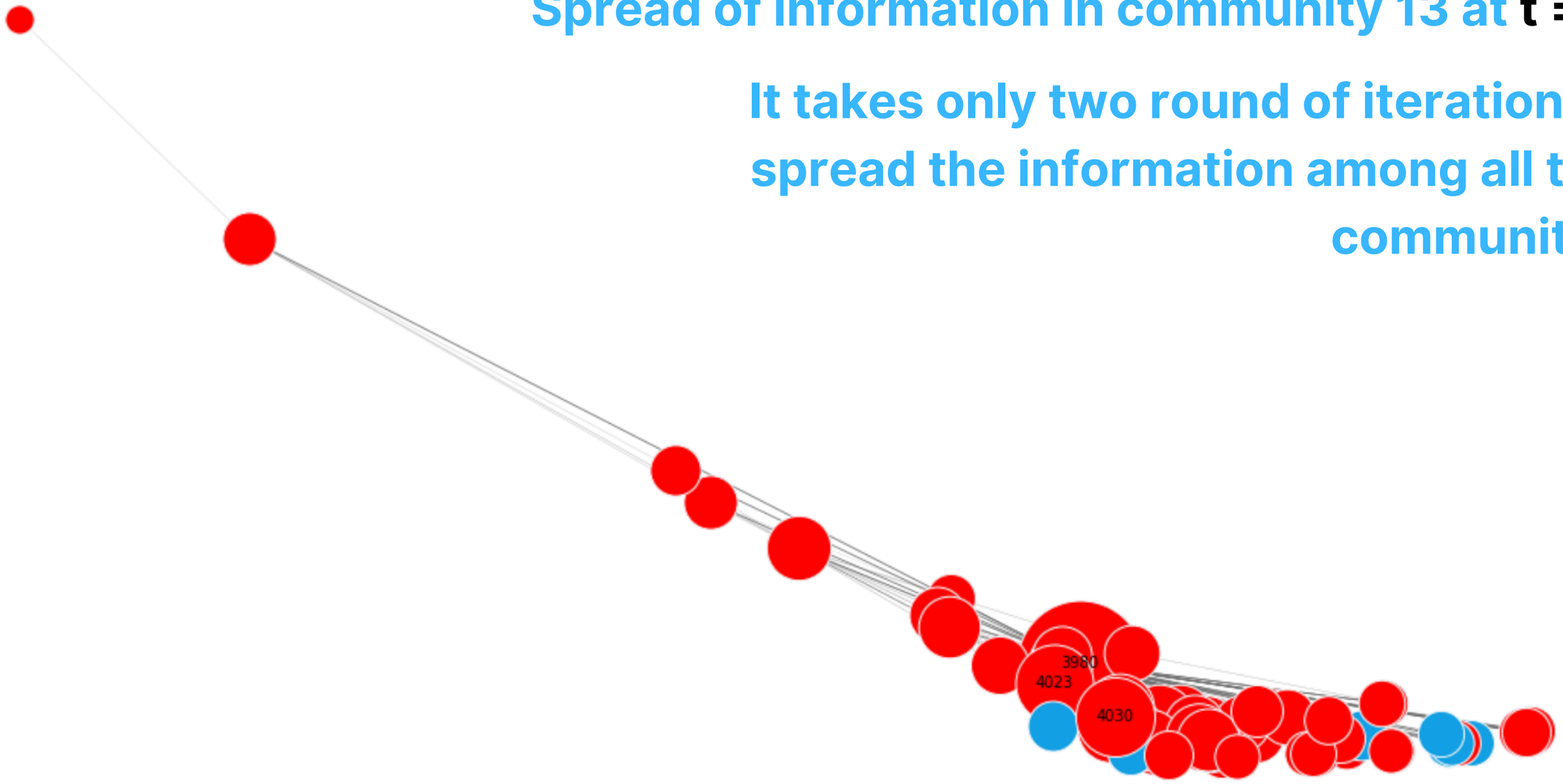


Spread of information in community 13 at t = 1



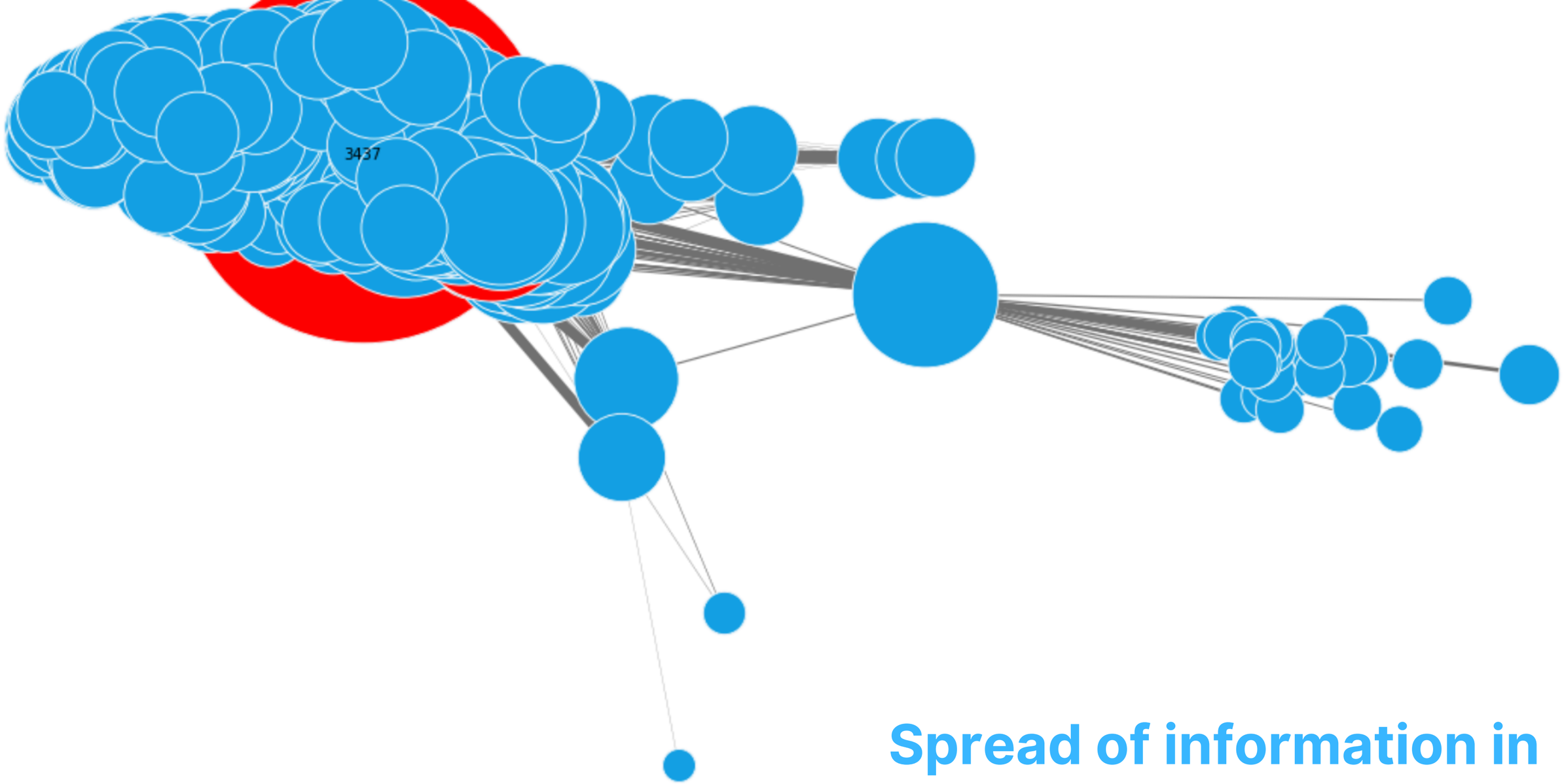
Spread of information in community 13 at t = 2

It takes only two round of iteration to spread the information among all the community.

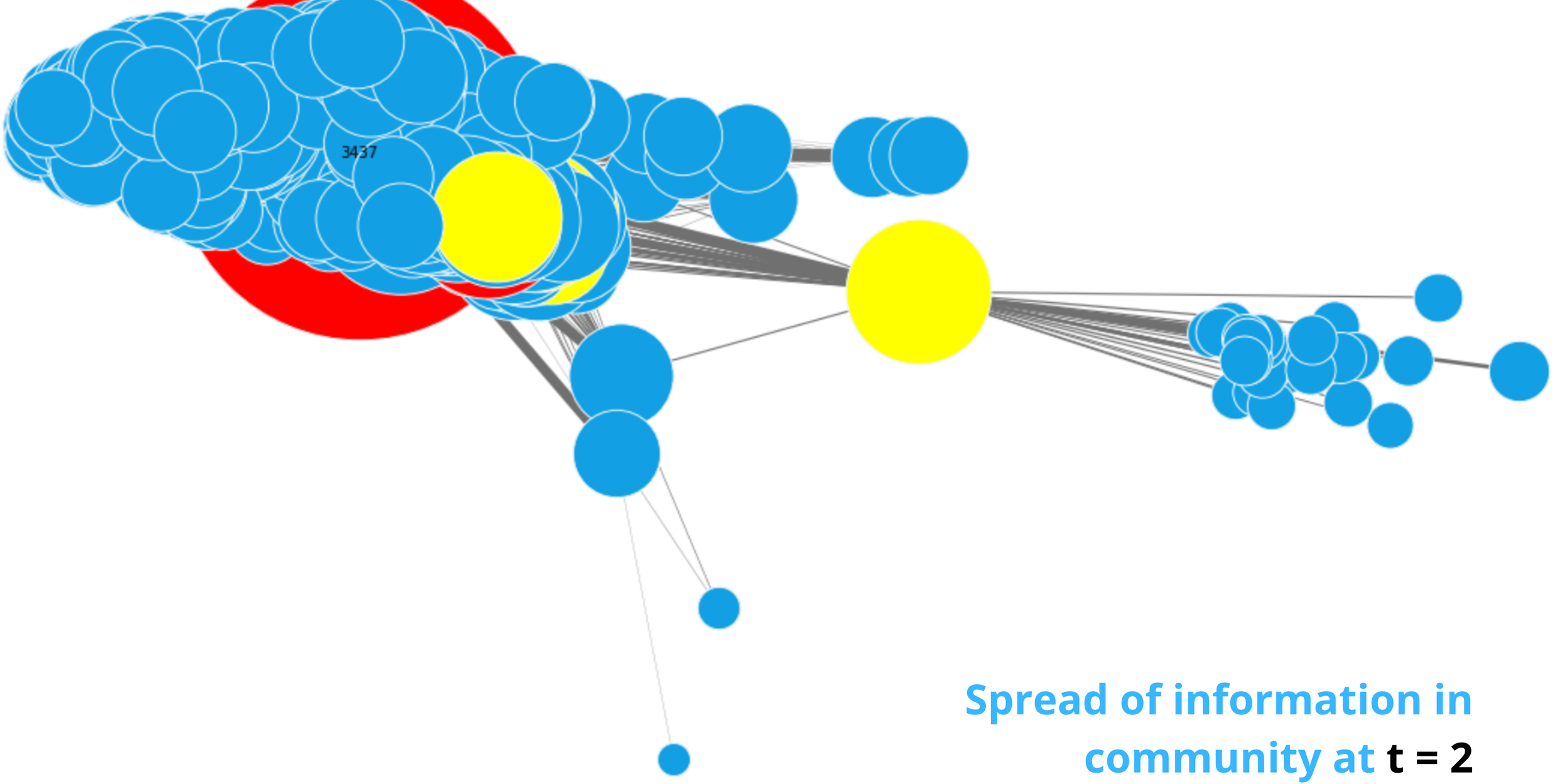


community 15- STATUS QUO

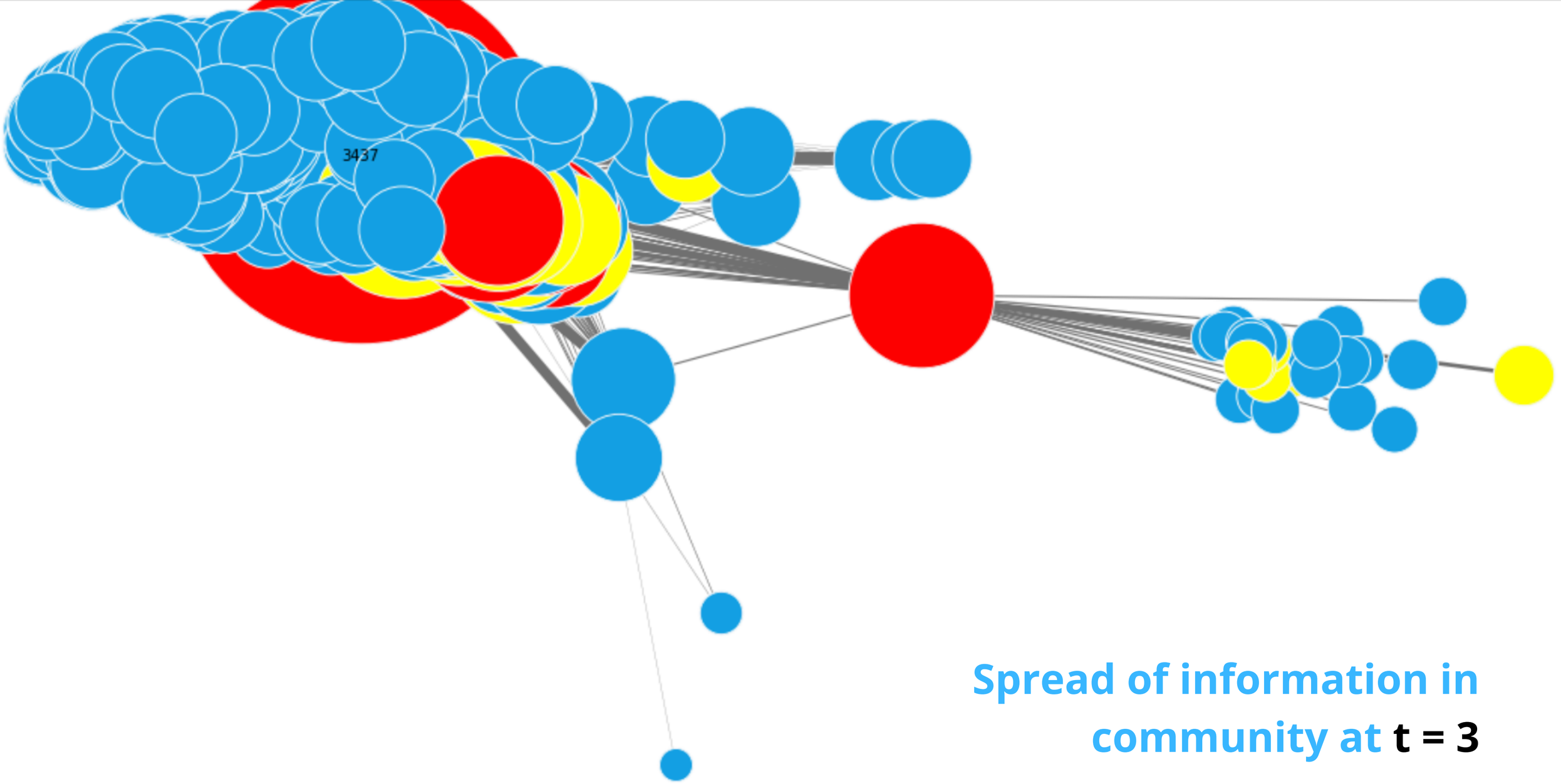
But what happens when we choose a bigger network, how fast does information travels?



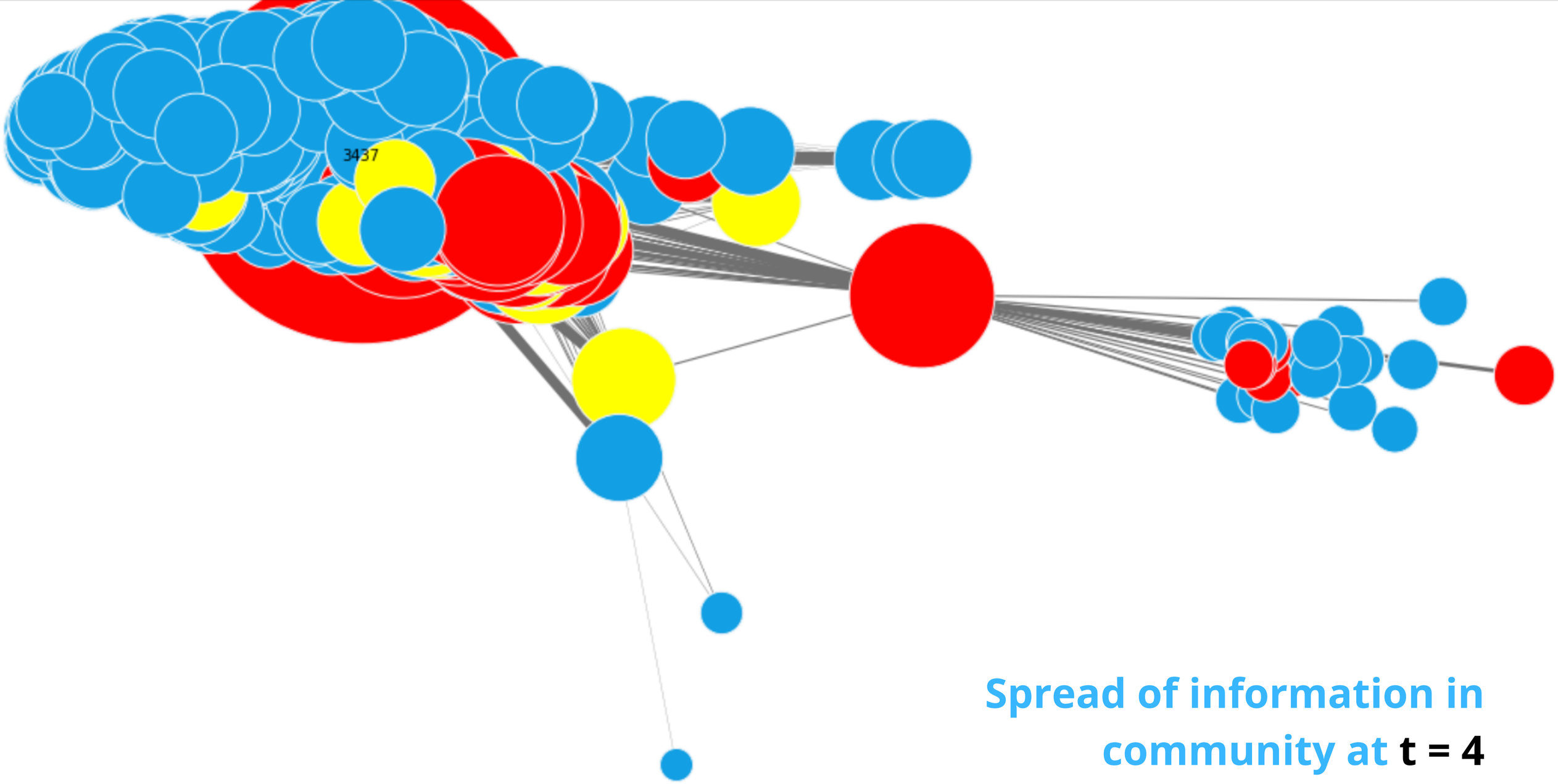
Spread of information in community 13 at t = 1



Spread of information in community at t = 2



Spread of information in community at t = 3



Spread of information in community at t = 4

Results & findings

Even testing **different communities** with **different metrics**, still the spread of information is very fast.

Information spreads much more faster in small networks: in just two rounds the information seems to "infect" all of its components

Of course in bigger networks this takes longer but it takes **just 4 rounds** to infect most of the nodes in the network

This is only a sample of Facebook network, just imagine how fast information spreads in real networks

Echo chamber effects represent the major **negative externalities** of this high interconnectedness

Thank you for your attention!

Enrico Grandi
748701

Martina Crisafulli
756241

Lorenzo Antolini
756861