Giacomo Usai - S4390761 giacomo.usai.iic97@gmail.com UniGe

Introduction

Once built the graph, I **compute the metrics** we discussed in class and **draw some conclusions on the type of the underlying network**. This is the first paper of three. In the second I will analyze the robustness and in the last one the social contagion scenario.

From: Stanford Large Network Dataset Collection (http://snap.stanford.edu/data/ego-Facebook.html)

Social networks : online social networks, edges represent interactions between people

My real Network Analysis

Name: Facebook **Type**: Graph

Number of nodes: 4039 (named from 0 to 4038) **Number of edges:** 88234

Average degree: 43.6910

Density: 0.010820, so we can see that our graph is not dense: it's **sparse** because ρ → 0.

is directed: False **is complete**: False

Diameter (it is the maximum eccentricity): 8

top 7 nodes with highest-betweenness_centrality (runn. time : ~3min.):

1. **107**: 0.48077531149557645
2. **1684**: 0.33812535393929544,
3. **3437**: 0.23649361170042005,
4. **1912**: 0.22967697101070242,
5. **1085**: 0.14943647607698152,
6. **0:** 0.14672864694039878,
7. **698**: 0.115768513859876

top 7 nodes with highest-closeness_centrality (runn. time : ~3min.):

1. **107**: 0.45969945355191255
2. **58**: 0.3974018305284913,
3. **428**: 0.3948371956585509,
4. **563**: 0.3939127889961955,
5. **1684**: 0.39360561458231796,
6. **171**: 0.37049270575282134,

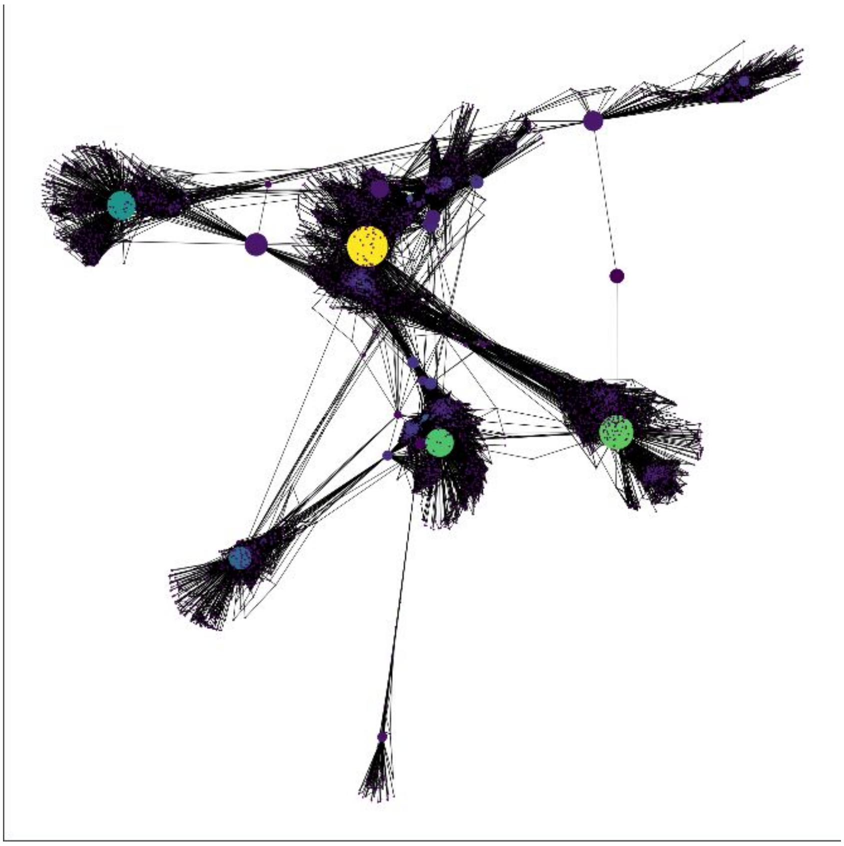
7. **348**: 0.36991572004397216

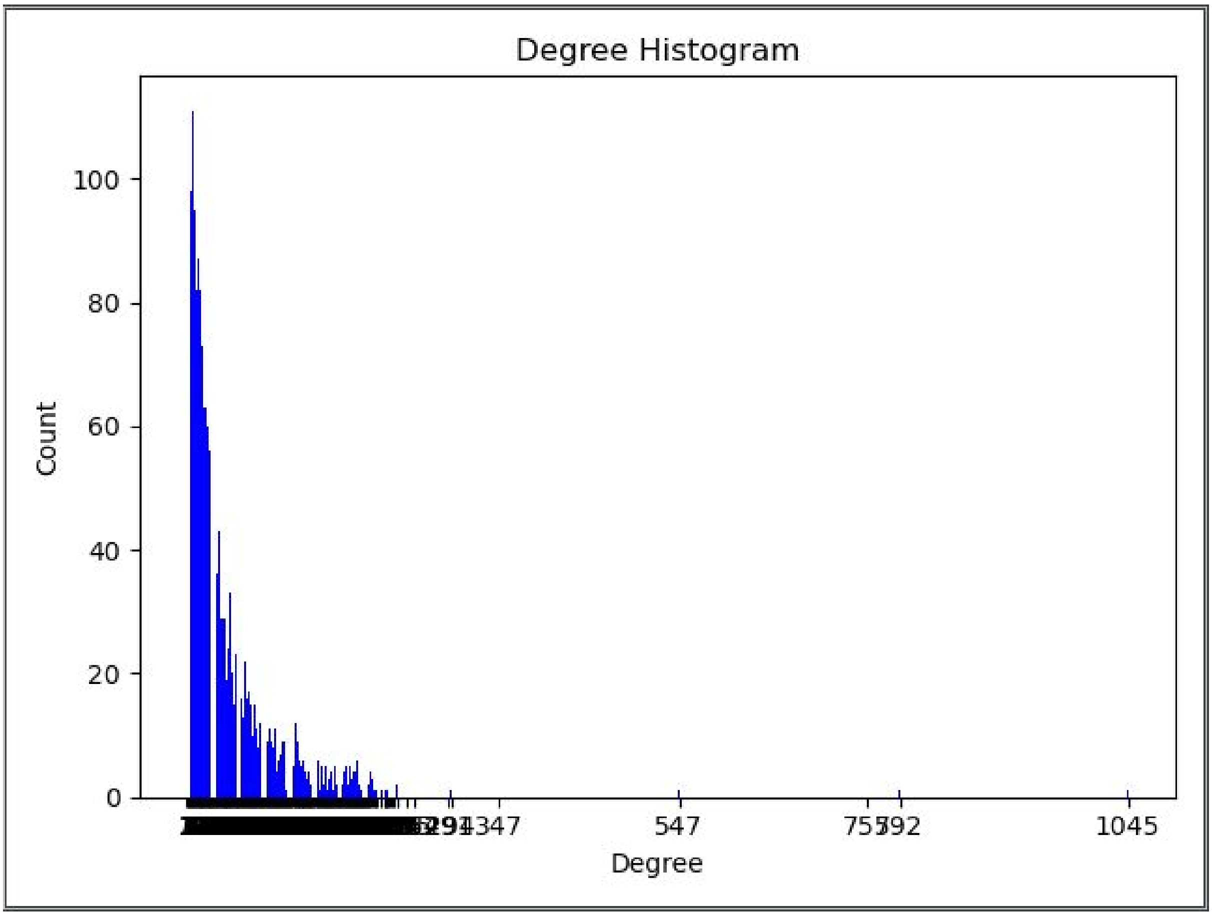
top 7 nodes with highest-degree (runn. time : ~15sec.):

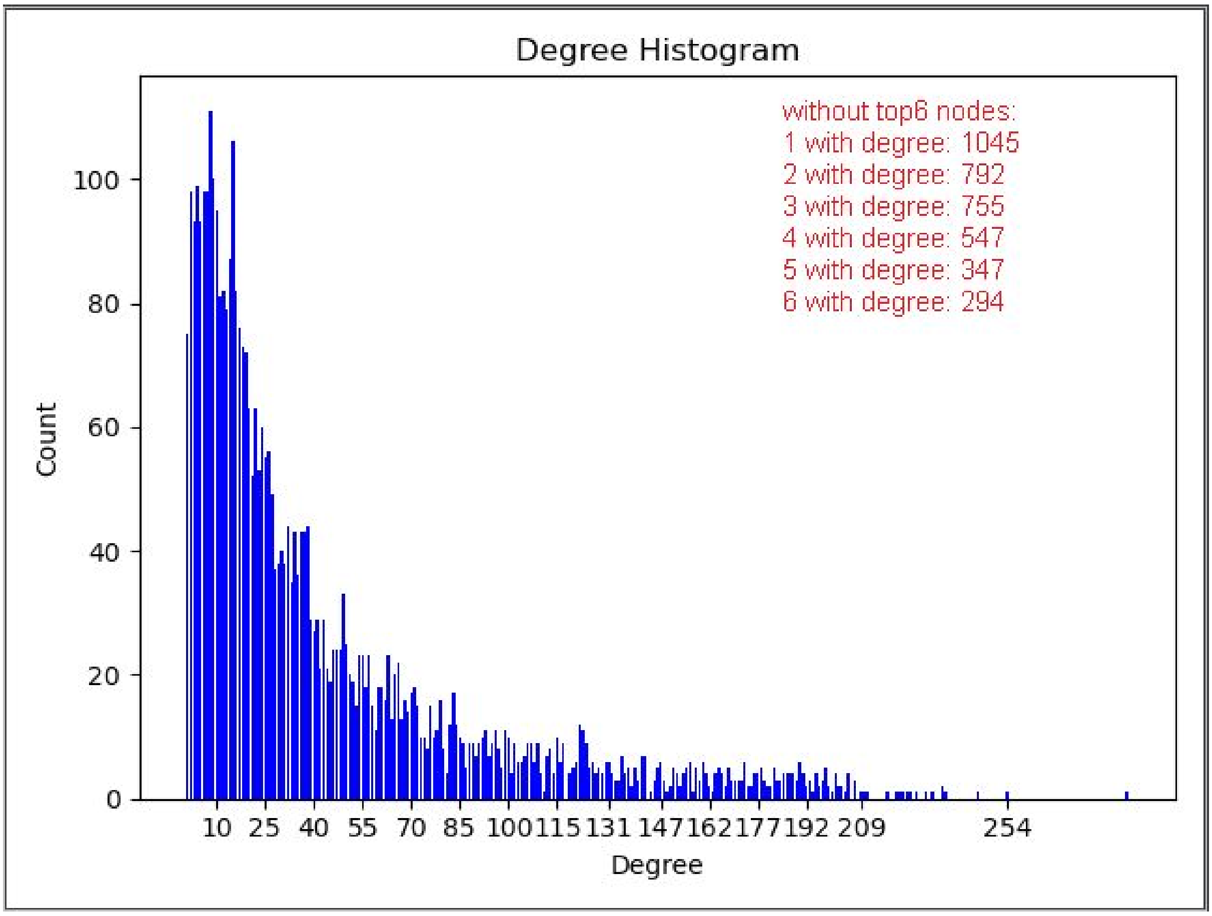
1. node: **107** degree:1045
2. node: **1684** degree:792
3. node: **1912** degree:755
4. node: **3437** degree:547
5. node: **0** degree:347
6. node: **2543** degree:294
7. node: **2347** degree:291

node with max-degree is 107 with degree=1045 and clustering=0.049038 that is under the average_clustering!

My graph have the same characteristics of a **scale-free** network, in particular we can see that it follows a **power-law** distribution. The most notable characteristic in a scale-free network is the relative commonness of vertices with a degree that greatly exceeds the average.
In addition, I found some **hubs**: people with degree >> Average degree (=43.6910).




Degree Histogram
Count
Degree
0
20
40
60
80
100
347
547
792
1045


Degree Histogram
without top6 nodes:
1 with degree: 1045
2 with degree: 792
3 with degree: 755
4 with degree: 547
5 with degree: 347
6 with degree: 294
Count
Degree
0
20
40
60
80
100
10 25 40 55 70 85 100 115 131 147 162 177 192 209 254

In graph theory, a **clustering coefficient** is a measure of the degree to which nodes in a graph tend to cluster together. Evidence suggests that in most real-world networks, and in particular social networks, **nodes tend to create tightly knit groups characterised by a relatively high density of ties**; this likelihood tends to be greater than the average probability of a tie randomly established between two nodes (Holland and Leinhardt, 1971;[1] Watts and Strogatz, 1998[2]).
Two versions of this measure exist: the global and the local. We calculated the **local** that gives an indication of the embeddedness of single nodes.

average_clustering: 0.605547

top 5 nodes with Low-clustering coefficient:

1. **3437**: 0.032230414314509376,
2. **0**: 0.04196165314587463,
3. **1684**: 0.044774546986936364,
4. **107**: 0.049038479165520905,
5. **3980**: 0.0853302162478083.

top 5 nodes with High-clustering coefficient:

1. **595**: 0.9883040935672515,
2. **3919**: 0.9848484848484849,
3. **3639**: 0.9848484848484849,
4. **3668**: 0.9818181818181818,
5. **576**: 0.978021978021978.

In conclusion, we can say that the **node 107** is the most important node followed by node **1684.**

Giacomo Usai

NOTE

Examples of questions you can answer are the following:

1. Does the graph have the same characteristics of a random or a power-law network?
2. Which are the most important nodes, with respect to a given centrality measure?
3. Are the paths short with respect to the size of the network?
4. Is the network **dense**?
5. And so on

Some measures provide indicators to know.

The importance of a node or an area in the network.

The distance among nodes or areas in the network Cohesion degree of an area in the network.

A **geodesic path** (**shortest path**) is a path between two vertices such that no shorter paths exist.

The **diameter of a network** is the length of the longest shortes-path between any pairs of vertices.

The simplest centrality measure is node degree and it can be applied to directed and undirected graphs:

1. **indegree**: number of links entering node i;
2. **outdegree**: number of links leaving node i;
3. **degree**: number of links of node i. Provides an indication of the ability of a node engaging in a direct relationship with the other nodes;

node degree=The total number of links of the node;

The maximun number of possible edges in a simple undirected network is ½ (N)(N-1)

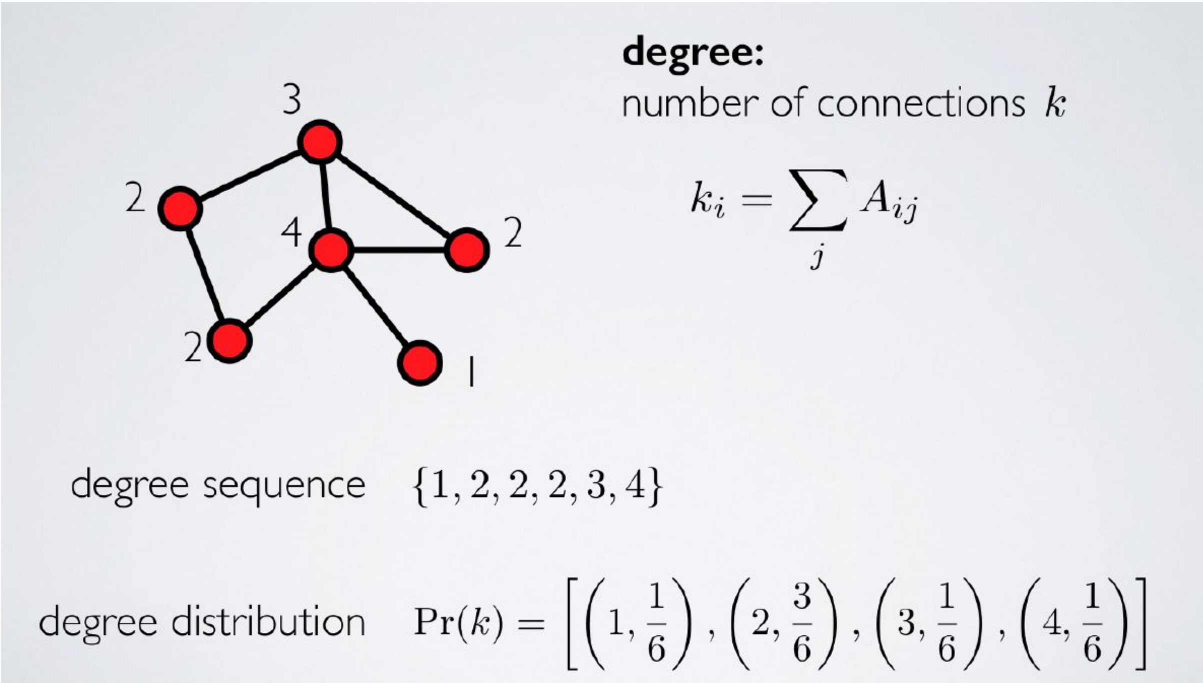
The density ρ of a network is the fraction of these edges really present in the network. So, it's simply: #edges/#max_number of possible edges. **0 <= ρ <= 1 always.**

If (ρ → costant when N → ∞) **dense** network;

else If (ρ → 0 when N → ∞) **sparse** network.

Degree distribution pk: Provides the probability that a randomly selected node

in the network has degree k. Nodes'link/maxLink.



Betweenness_centrality of a node v
is the sum of the fraction of all-pairs shortest paths that pass through v.
So, the node's size it's related with the node's betweenness_centrality. In practice: how many pairs of individuals would have to go through you in order to reach one another in the minimum number of hops? Their removal may disrupt communication.

- Nodes with high betweenness have more important.

Closeness measures the mean distance of a vertex to other vertices:

- Nodes with high closeness have better access to information or more direct influence on
  other vertices.

Closeness centrality of a node u is the reciprocal of the sum of the shortest path distances from u to all $n-1$ other nodes. Since the sum of distances depends on the number of nodes in the graph, closeness is normalized by the sum of minimum possible distances $n-1$.