# Understanding social graphs

**Exercise I**

We denote $G(N,E)$ the undirected social graph of interactions of a brand on the Facebook social networks. The graph is composed of $N=450$ user profiles who are inter-connected to an average number of 20 people: $<k>=20$. They have all interacted at least once with the brand in the last month.

1. Estimate the number of edges L of the social Graph E.

Step 1: Express mathematically the definition of $<k>$, we denote it (1)

Step 2: Express $\sum\_{i=1}^{N}k\_{i}$ as a function of L, we denote it (2)

Step 3: Replace (2) in (1) to get L as a function of $<k>$ and $N$.

1. Estimate the density of the graph $d$ and the probability that two given nodes are connected?

Hint: $d=\frac{L}{L\_{max}}$

**Exercise 2 – Toward the concept of social capital**

Let’s consider the following undirected graph representing the interactions between the CEO’s of big size enterprises and industries.



An interactive version of this graph is available online at : <https://graphcommons.com/selections/f440e480-479e-47ae-a3f0-90d73a001472>

Short url: <http://bit.ly/2oJmrg7>

1. Calculate the probability $p\_{k=i}=\frac{N\_{k=i}}{N}$, of a node to get the exact degree $i$ for $i=1$ up to the maximum. Check that $\sum\_{i=1}^{max}pk\_{i}=1$.
2. Display the degree distribution of the nodes of the graph.
3. Express the average degree of a graph as a function of $k$ and $p\_{k}$. Calculate the average degree of the graph using the values calculated in (2).
4. Calculate the local clustering coefficient of each node and fill in the table below. Use the symmetry of the nodes to avoid redundant calculations. The formula for calculating the local clustering coefficient of a node $n$ with degree $k\_{n}$ is as follows:

$$C\_{n}=\frac{2L\_{n}}{k\_{n}(k\_{n}-1)}$$

$$L\_{n} is the number of links observed between the neighbors of node n.$$

|  |  |  |
| --- | --- | --- |
| Node | Degree | Local Clustering Coefficient |
| Sergei |  |  |
| Anna |  |  |
| Lary |  |  |
| Michelle |  |  |
| Marc |  |  |
| Jen |  |  |
| John |  |  |
| Jack |  |  |
| Tom |  |  |
| Guy |  |  |
| Thomas |  |  |
| Franck |  |  |
| Martin |  |  |
| Bernard |  |  |
| Kevin |  |  |
| Trevis |  |  |
| Elon |  |  |

1. Place the CEO’s in the two-dimensional plot below and display the average degree and average LCC to create a quadrant on the chart.

 

1. Discuss the advantages and inconvenient of being in the four sections of the quadrant. Illustrate your discussion based on the visual representation of the CEO’s graph and their position observed in each section.

*For more information consult the theory of social capital of Burt.*

*Ref:* [*http://snap.stanford.edu/class/cs224w-readings/burt00capital.pdf*](http://snap.stanford.edu/class/cs224w-readings/burt00capital.pdf)

|  |  |  |
| --- | --- | --- |
|  | Degree | LCC |
| Section 1 | Below average | Below average |
| Section 2 | Below average | Above average |
| Section 3 | Above average | Below average |
| Section 4 | Above average | Above average |

**Exercise 3 – It’s a small world after all**

We consider a social random graph of average degree $<k>$.

1. Estimate what is the average number of people, denoted $N(d=1)$, at an **exact** distance 1 of a given node $n$ of the graph.

$N(d=1)$=

1. Now, estimate the number of people at an **exact** distance $dϵ\left⟦2,3…,d\_{max}\right⟧$ of a given node $n$ of the graph.

$N(d=2)$=

$N(d=3)$=

$N(d=4)$=

 …

$N(d=d\_{max})$=

1. Using your results of (2), estimate the number of people at an exact distance $dϵ\left⟦2,3,4,5,6,7\right⟧$ of a given individual belonging to the society (supposing that society is a random graph). Note that according to the sociologist Dunbar, the average number of acquaintance of a given people equals $<k>=150$. This number is known as the Dunbar Number.

$N(d\leq d\_{max})$=

1. From what threshold of $d$ do the number of nodes surpass the number of people on earth. What do you think of this result?
2. What is the number of people estimated at a distance **up to or equal to** $dϵ\left⟦2,3…,d\_{max}\right⟧$ of a given node. We denote this value $N(d<i)$. Express this value as a function of $<k>$.

|  |  |
| --- | --- |
| Up to  | Formula |
| 1 | $$N(d\leq 1)≈$$ |
| 2 | $$N(d\leq 2)≈$$ |
| 3 | $$N(d\leq 3)≈$$ |
| 4 | $$N(d\leq 4)≈$$ |
| d | $$N\left(d\right)≈ (eq.1)$$ |

1. Simplify the expression (eq.1) of $N\left(d\right)$ provided in previous section. Multiply it with <k> (eq.2) and substract (eq.2) with (eq.1)
2. Given that N(d) cannot exceed the number of nodes of the graph, there exist a value $d$ such that $N\left(d\right)≈N$. Assuming that $<k>\gg 1$, express $d\_{max}$ as a function of $N$ and $<k>$.
3. The last equation indicates the scaling of the graph diameter with the size of the system. In real-graphs, it is indeed more often corresponding to the average distance $d$. This is due to the existence of a very few large shortest paths such as the diameter. Finally, the small-world property in networks is characterized by equation 1.

$$d\_{ }≈\frac{ln⁡(N^{ })}{ln⁡(<k>)}$$

Given that the average distance between Facebook users equals 3.5, that number of Facebook contacts are in average 160 and that the platform is composed of 2.2 Billion users. Using equation 1, test if Facebook matches with the small-world criterion.

In the next laboratory class, we will test the small-world nature of real large complex systems.