Universitá di Pisa

Peer to Peer Systems and Blockchains

Final Project

# Building Dynamic Overlays by Epidemic Protocols

*Author:*
Carmine Caserio

# Chapter 1

# Describing the implementation

The *Newscast protocol*[1] has been implemented by using two auxiliary data structure: the `PeerDescriptor`, that has the informations about the peer identifier and the peer view, that is the list of all neighbours that each peer has, and the `NeighbourDescriptor` that describes the item received by each neighbour, pairing to it a timestamp. In more detail, in the following there is the description of both these data structures, focusing on the implementation of the methods defined into them.

## 1.1 PeerDescriptor data structure

This data structure is composed by:

- an integer that defines the peer identifier (unique for each peer);
- a list of all the neighbours, so using the `NeighbourDescriptor` data structure because we need to know the moment in which it has been added to the list;
- an integer, defining the maximum number of neighbours the peer can have, that is an externally settled value, but it is an information that a peer needs to know.

Except for the overriding of two methods (that is, *equals()*, needed for seeing if the object passed as parameter has the same peer identifier, and so it is equal to the input object, and *compareTo()*, that does the same check of the previous method) that are used in the other ones, the methods defined in this class are:

- *selectPeer()* that gets a random peer from the list of its neighbours and the peers exchange between them the own neighbours list and their identifier, by pairing the current time as timestamp to it (since the project is just a simulation of the protocol, actually, the list exchanging is done locally, by appending to each peer the informations of the neighbour, so both for the current peer and for its neighbour); afterwards, the *updateState()* function is called, that is described below;
- *updateState()* that as it starts, it tries to remove duplicates and the information about the current peer sent by its neighbour, and then, it sorts all the elements of the cache, and removes all of them except for the latest $c$ (where, $c$ is the dimension of the cache).

[1] Newscast Computing, https://www.cs.vu.nl/pub/steen/papers/wip-newscast.pdf

## 1.2 NeighboursDescriptor data structure

This data structure is composed by:

- a long that defines the timestamp in which this entry has been inserted;
- an integer that identifies the peer that is the neighbour of the current peer (since the `NeighboursDescriptor` is the data structure used to fill the neighbours' list).

In this class there is the overloading of the *compareTo()* method which is defined both for the peer identifiers and for the timestamps. Now, we can pay attention to the implementation of the newscast protocol's main, although the heart of the protocol has actually been already discussed in the description of the *selectPeer()* and *updateState()* methods.

## 1.3 Newscast protocol implementation

In the main function, there are two calls to two different methods, in which: the first one has been defined for the *scale free network* and the other one for the *grid network*.

### 1.3.1 Scale free network

For achieving this kind of topology, the *newscastProtocolScaleFreeNetwork()* function has been used, which which in turn initialize the network, by using the *initScaleFreeNetwork()* function, described below and saves into a `.sif` file the initial topology of this network; afterwards each peer interacts one time by executing its own *selectPeer()* method starting from the first one, up to the last one that has been instantiated; at the end of such peer interaction, the function saves the topology of the current peer in another `.sif` file.

#### ***initScaleFreeNetwork()*** function

This function adds the first cache lines peers in the graph data structure; afterwards, it saves the current time for initializing all the peers with the same time and, for each peer, it takes another random peer, making sure that the selected peer is not the current one, and then it decides whether to add it or not. This decision is taken by generating a random value $\in (0, 1)$ and by seeing whether this value is less than a certain value that represent the probability of taking the random selected peer as neighbour. This initial neighbours' selection approach is known as the *Erdős-Rényi Model*. In this peers' insertion procedure the algorithm takes into account the vertices of all the inserted edges.

Afterwards, for each other peer that has to be added to the network, the algorithm chooses two random peers from the list of peers saved in the edges' list and adds this peer to these selected peers and completes the insertion procedure by adding the just constructed edges to the edges' list. This means that the more probable peers that can be selected are those that appears more frequently in the edges' list, as defined by the *Barabási Albert Model*.

Applying as first step the *Erdős-Rényi Model* guarantees that for building the scale free network, we start from a random graph.

### 1.3.2 Grid network

For achieving this other kind of topology, the *newscastProtocolGridNetwork()* function has been used, which, as for the scale free network, initialize the network, in this case, by linking:

- all the internal elements of the grid to the neighbours on their left, right, up and down;
- all the elements on the first or last row except for the first and last position, to the neighbours on their left, right and down, resp. up for the first row, resp. last row;
- all the elements on the first or last column except for the first and last position, to the neighbours on their up, down and right, resp. left for the first column, resp. last column;
- the element in position $(0, 0)$ to the neighbours on his right and down;
- the element in position $(n-1, 0)$ to the neighbours on his left and down;
- the element in position $(0, m-1)$ to the neighbours on his right and up;
- the element in position $(n-1, m-1)$ to the neighbours on his left and up.

## 1.4 Testing parameters

The parameters used in the testing part of the simulation has been defined according to:

- the number of peers that participate in the simulation for the free scale network, by using:
  - 1.000 peers;
  - 2.500 peers;
  - 5.000 peers;
  - 7.500 peers.
- the width and height of the size of peers that participate in the simulation for the grid network, by using:
  - width = 10 and height = 100;
  - width = 25 and height = 100;
  - width = 50 and height = 100;
  - width = 75 and height = 100;
- the number of cache lines that each peer has, defined for 20, 35 and 50.

Regarding the probability used in the construction of the initial graph to get a peer of the network as own neighbour, the approach used in this project has taken into account this aspect, but for generalizing the procedure, a random variable $\in (0, 1)$ has been used, such that it decides the starting value.
The program takes in input 3 integers, that correspond to: the number of peers given in input for the free scale network and the width and height for the grid network.

# Chapter 2

# Analysing the results

For all the cases and the networks taken into account, the number of cache lines that has been considered are equal to the values 20, 35 and 50, because in the original paper, the considerations that had been carried out were the ones with different values of cache lines in the interval between 20 and 50. From the analysis in the paper, indeed, with $c = 20$ (where $c$ is the number of cache lines), the variance reduction rate is acceptable, although drifts away from the theoretically predicted rate and that with $c = 40$ the theoretical prediction has an acceptable accuracy which indicates that from the point of view of this protocol the newscast layer can provide a sufficiently random sampling of available peers. Actually, the analysis taken forward in the paper shows also a case that has not been considered in here, with $c = 80$.

## 2.1 Free scale network results

Regarding the free scale network results, the analysis has been divided depending on the value of the cache line, because in this way we can see how the graph evolves taking into account that the maximum number of neighbours of the peers are the ones defined by the cache lines value. All the results shown start from an initial graph composed by a single connected component.

As we will see later in the results' analysis, the distribution found for the average clustering coefficient follows the power law distribution, that in this report has been shown by using the *log-log scale*, so that the results can be analysed more precisely; besides, the in-degree distribution has been exposed by using a *log-log scale* for a more accurate analysis of the results.

### 2.1.1 Sparse case, 20 maximum neighbours

This experiment has been carried on with different values of the number of peers parameter, as previously stated.

**Tiny number of participants: 1.000 peers**

An example of an initial scale free graph is the following one:

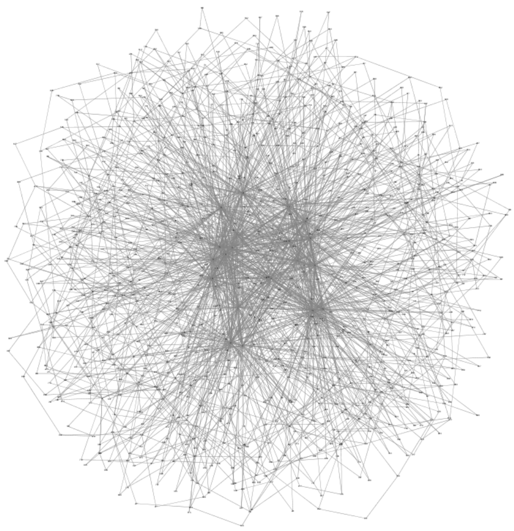


Figure 2.1: Example of initial scale free graph with 1.000 peers

From a scale free network as the previous one, a single step execution of the *Newscast protocol* for each of the peers in the network outputs the following graph:

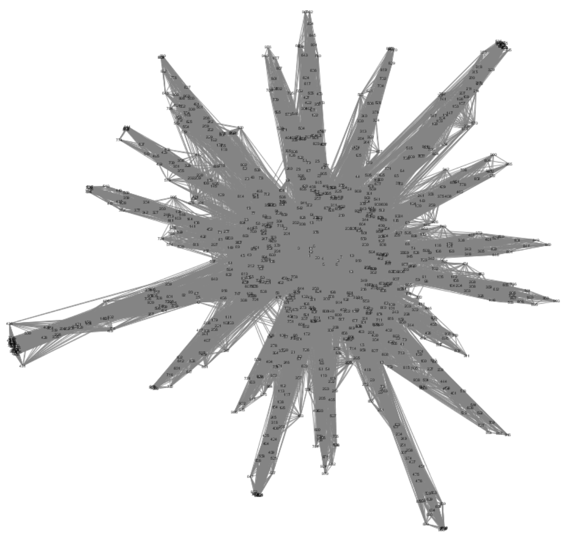


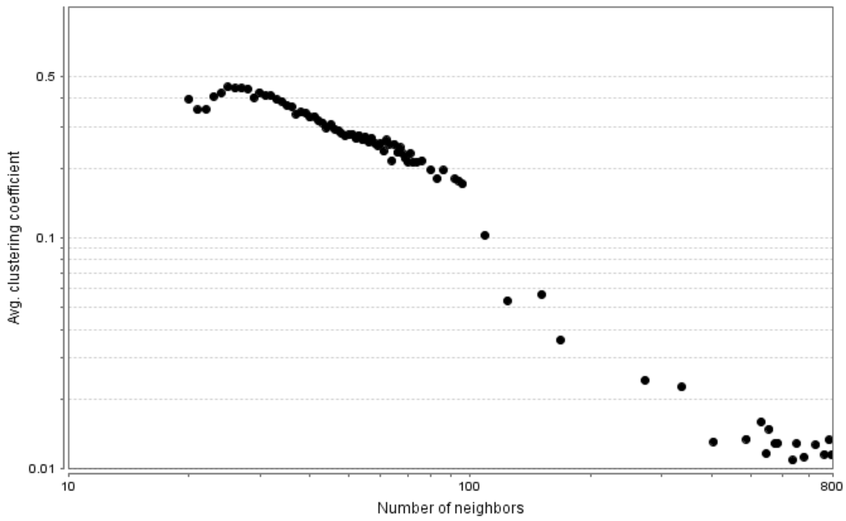
Figure 2.2: Example of final scale free graph with 1.000 peers by using a cache of size 20

The graph analysis shows us that the clustering coefficient value is 0.364 for a network diameter of 9 peers and the graph, as can be seen, is totally connected, that is, there is only a single connected component.

Other network statistics have been taken into account, such as:

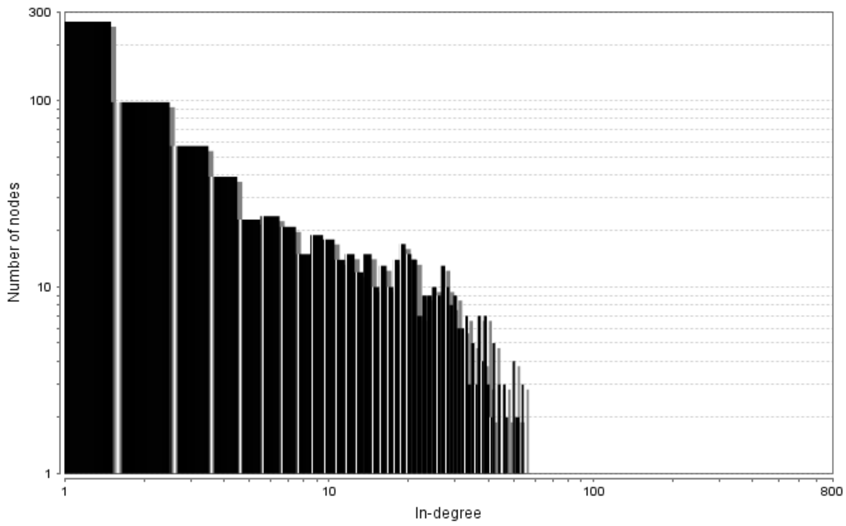
- about the out-degree distribution of the peers, we can avoid to graphically show the result, since all the nodes have an out-degree value of 20 that is equal to the cache line value;

- the average clustering coefficient of the network, that is in this case shown in the following:



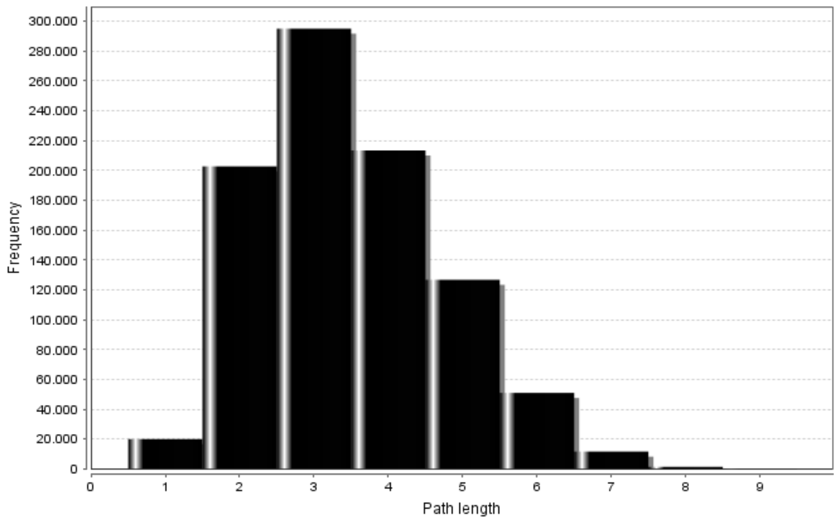

from which we can see that the average clustering coefficient tends to be near to 0.01 as the number of neighbours increases, never getting beneath that value; the highest value of clustering coefficient corresponds to 0.45, for a corresponding number of neighbours equal to $27 \sim 28$;

- the in-degree distribution of the peers, which in this case is as follows:




from which we can see that most of the nodes have an in-degree value of 1 and 2; for all the others values, the in-degree frequency tends to be near to 0; the highest in-degree value, that is one of the least common ones, is near to 800;

- the shortest path length distribution, shown in the following figure:




from which, it can be shown that the path of length 3 is the most common one, with a frequency of about 295.000 and that the longest path is the one of length 8, with a frequency near to 0.

### Small number of participants: 2.500 peers

An example of initial scale free network is shown in the following:

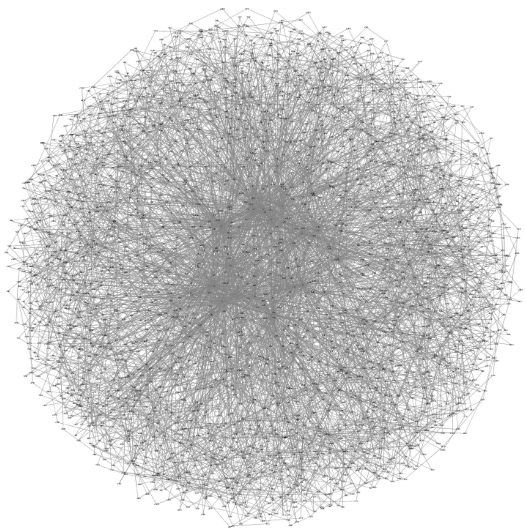


Figure 2.3: Example of initial scale free graph with 2.500 peers

After the execution of the algorithm, the resulting graph is shown in the following:

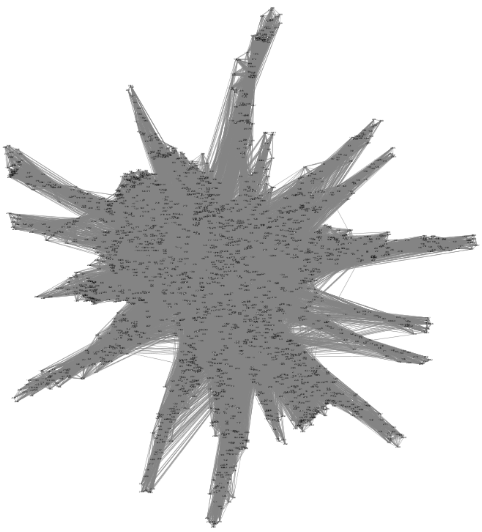
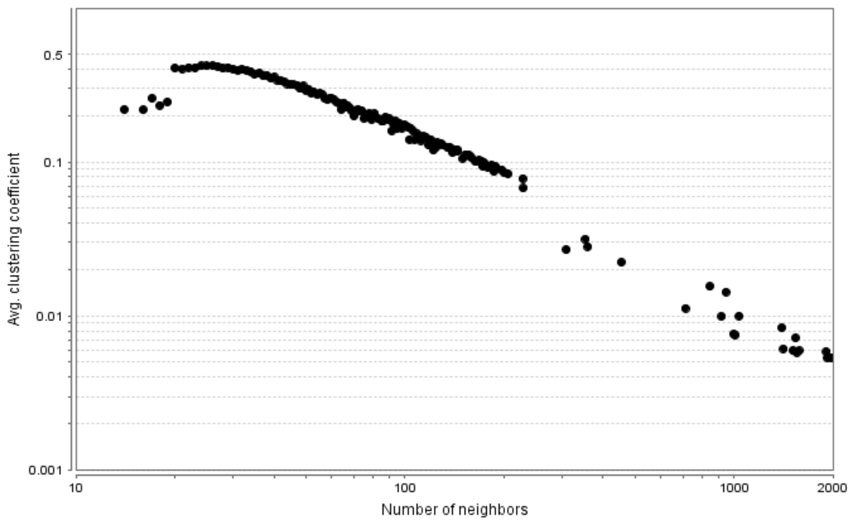


Figure 2.4: Example of final scale free graph with 2.500 peers by using a cache of size 20

For this network, the statistics taken into account have shown that:

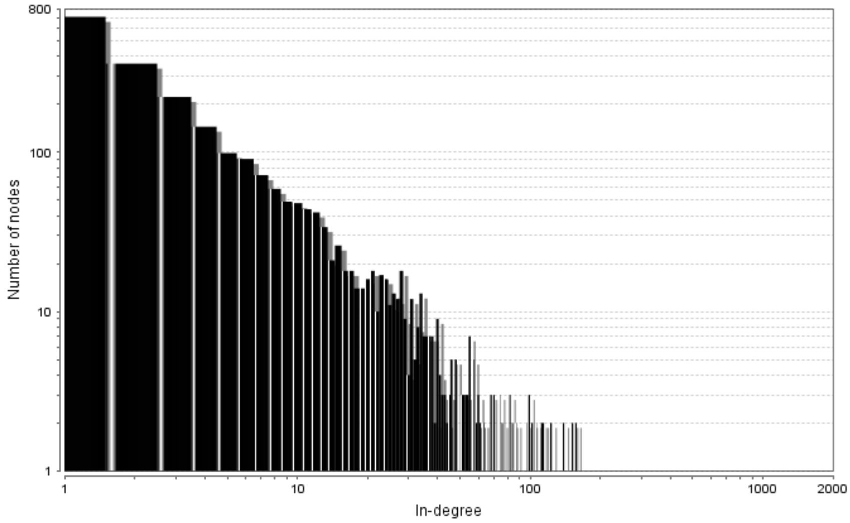
- the average clustering coefficient of the network is the following one:




  so, basically, as the number of neighbours increases, the clustering coefficient decreases up to about 0.0125. This means that when the number of neighbours increases, the nodes are much less linked among themselves;

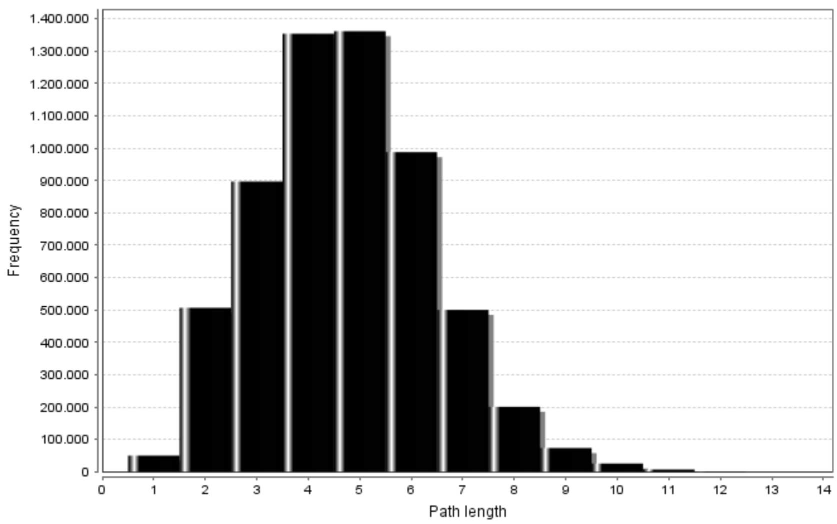
- the out-degree distribution of the peers in the network has not been shown because it is not meaningful; indeed, most of the nodes has an out-degree of 20 equal to the cache size and very few have an out-degree of 18 and 19;

- the in-degree distribution of the peers is as follows, the $x$ axis is in log scale:




we can see that the majority of the nodes, that is about 700, has an in-degree of 1 because there are many connected components; about an half of the previous class of nodes has instead an in-degree value of 2;

- the shortest path length distribution is:




it shows us that the most common shortest path lengths are the one of length 4, with a frequency of 1.345.000 $\sim$ 1.355.000 and the one of length 5, with a frequency of 1.365.000 $\sim$ 1.375.000; there are then, the paths of length 6, with a corresponding frequency of just below 1.000.000, and of length 3, with a frequency of just below the 900.000.

### Medium number of participants: 5.000 peers

An example of initial free scale graph is shown in the following:

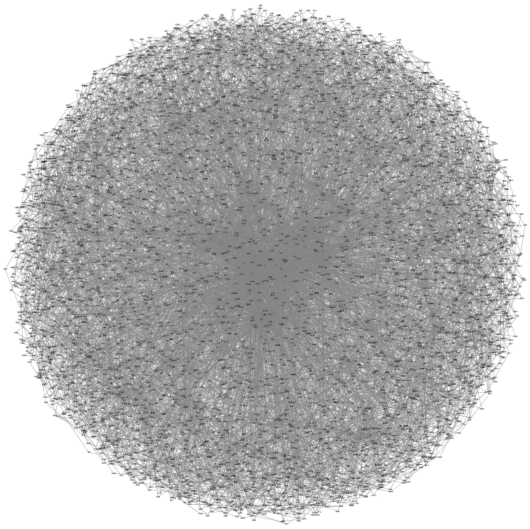


Figure 2.5: Example of initial free scale graph with 5.000 peers

The execution of the algorithm returns a single connected component; the clustering coefficient is of 0.383 and the diameter is of 15 peers. From the graph of the example shown above, the resulting graph is shown below.

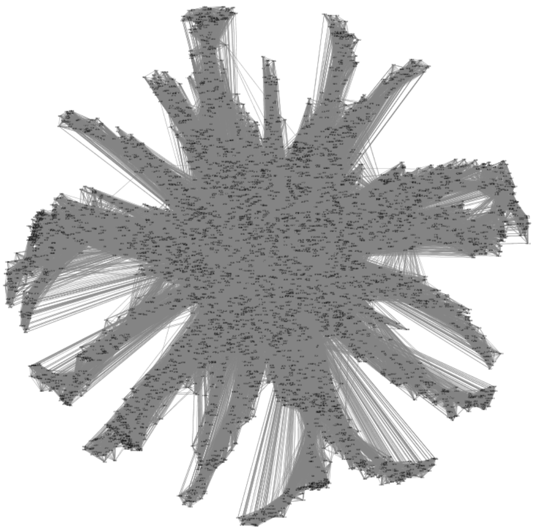
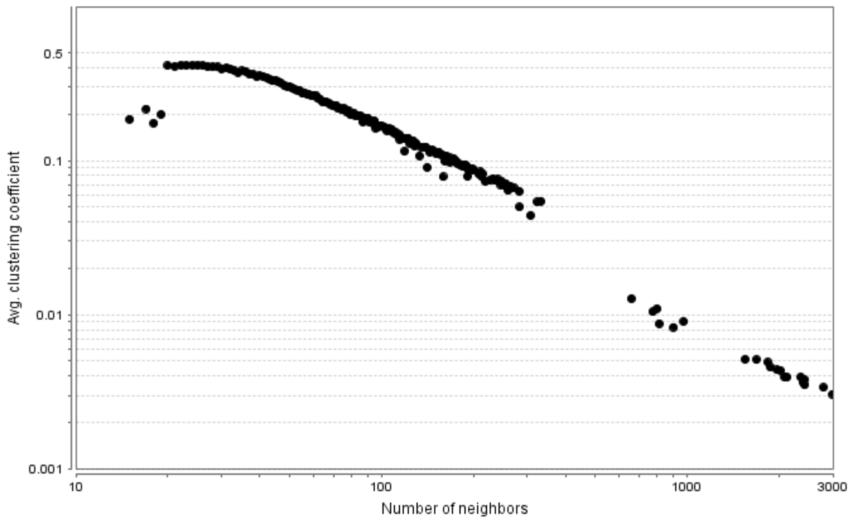


Figure 2.6: Example of free scale graph with 5.000 peers

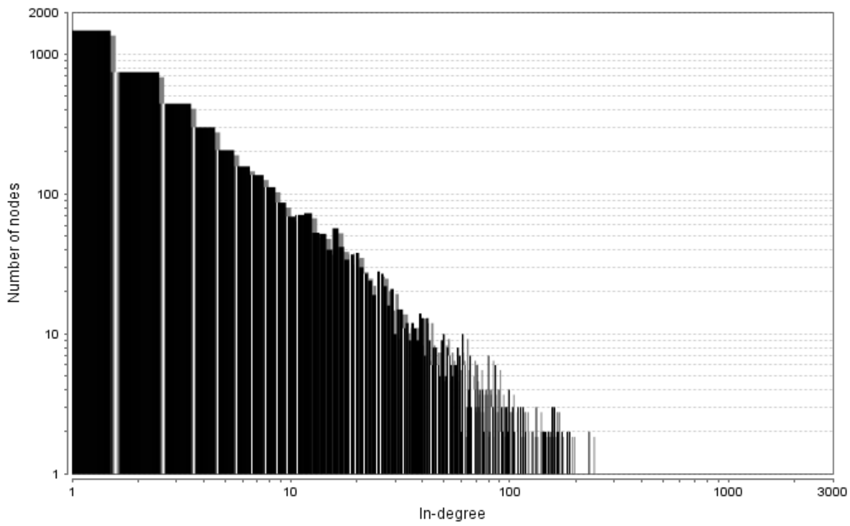
In this case, the statistics have shown that:

- the average clustering coefficient is the one in the next figure:




in which the $x$ axis is logarithmic and as the number of neighbours increases, the value of the average clustering coefficient decreases from a value of about 0.415 up to $0.01 \sim 0.005$.

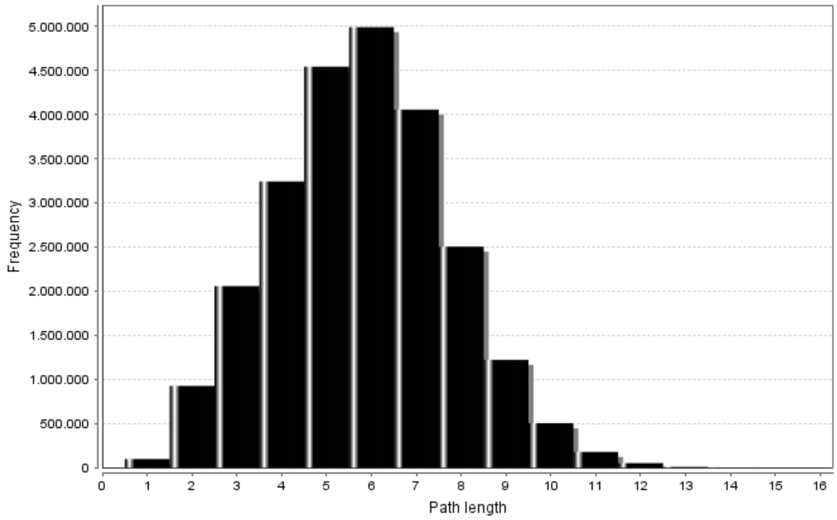
- the in-degree distribution is




and we can see that the most common in-degree values are 1 and 2 with approximate values of respectively, 1.480 and 740, so overall the number of nodes in the that have these low in-degree values are 2.220;

- the out-degree distribution, in this case too, is not exposed since all the peers have an out-degree of 20, except for a very few number of nodes that have an out-degree of 19;

- about the shortest path length distribution, we have a situation like the following:




in which, the most common path length is the one of length 6, with a frequency of about 5.000.000; the second most common path is the one of length 5 with a frequency of about 4.510.000 and the third one is the one of length 7, with a frequency of 4.000.000; these path lengths together with the length 4 and 8 are the most common ones.

**Large number of participants: 7.500 peers**

In this experiment, the initial free scale graph has 7.500 peers and the execution of the algorithm returns a single connected component, with clustering coefficient equal to 0.395 and diameter size of 18 peers. The output of the algorithm is shown in the following:

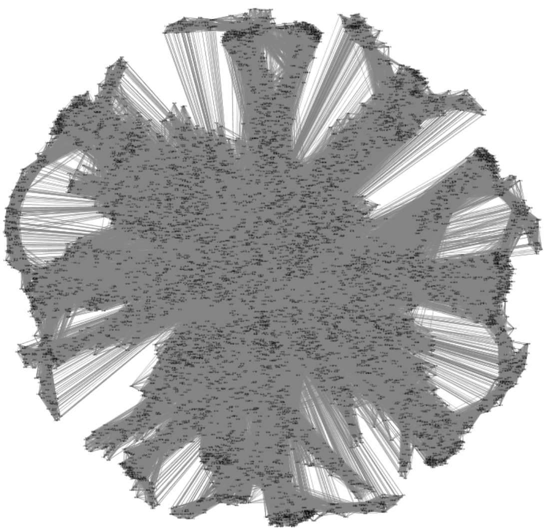
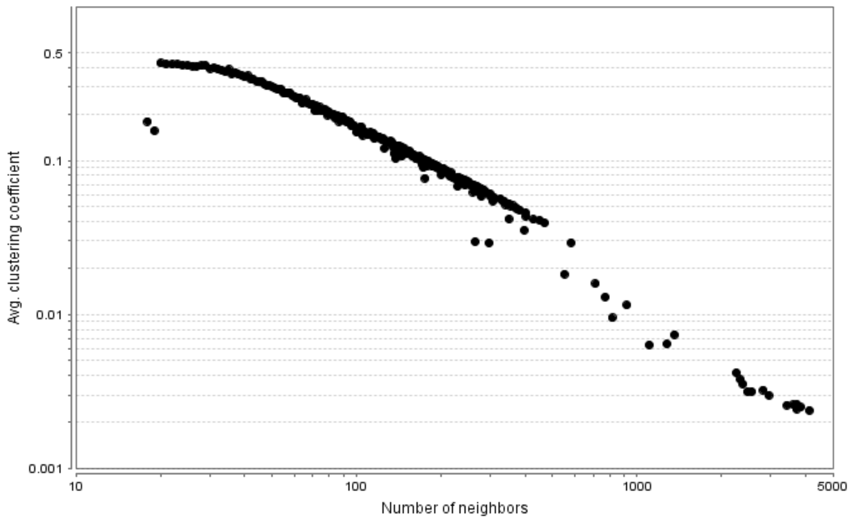


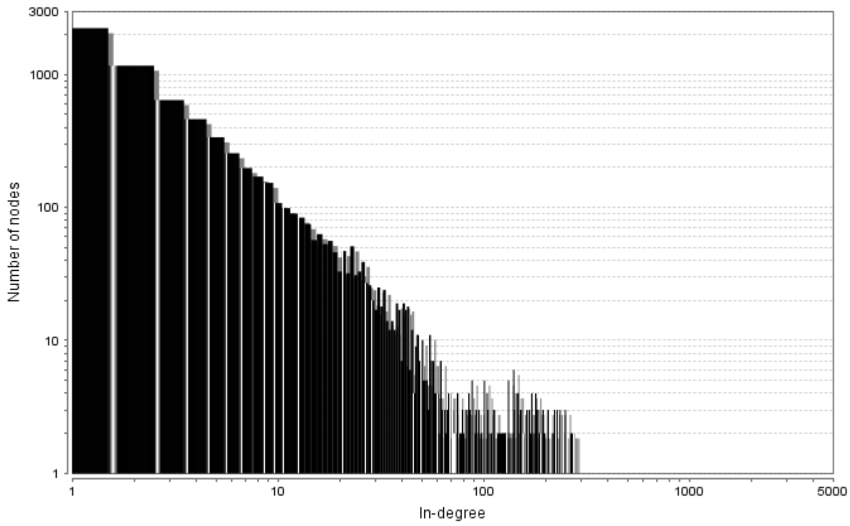
Figure 2.7: Example of final scale free graph with 7500 peers by using a cache of size 20

The statistics collected have shown that:

- the average clustering coefficient, except for a number of neighbours of less than 20, for which the corresponding average clustering coefficient values are about 0.18 $\sim$ 0.155, it has the highest value equal to 0.43 for a corresponding value of neighbours equal to 20; afterwards, the values descends by following the behaviour shown below, up to arriving to the minimum value that corresponds to 0.010 $\sim$ 0.005, with a corresponding number of neighbours equal to 3.000 $\sim$ 4.100;

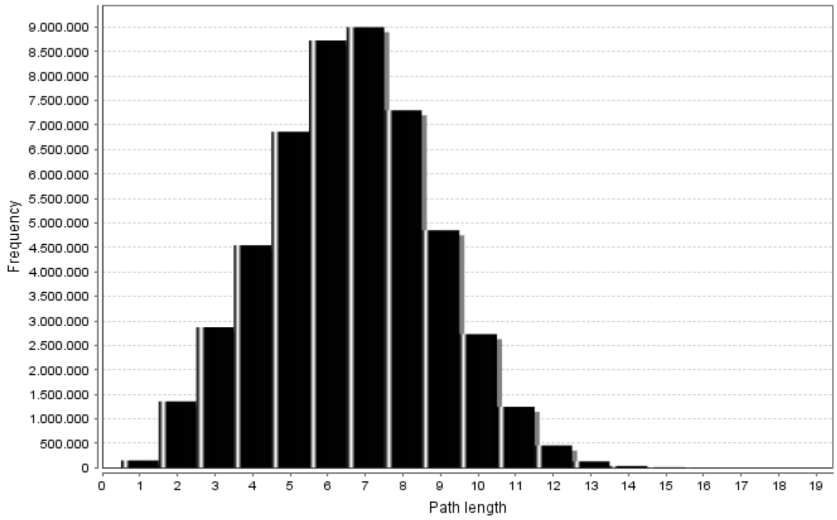



- the in-degree distribution for the most part of nodes is in the range 1 up to 3 (that is, the nodes with this in-degree are the most common ones) and the corresponding value varies between 2.240 $\sim$ 2.250; the highest value corresponds to 1 and the corresponding value is just above the 2.200 value, as can be seen by the following:




- about the out-degree distribution, the unique value reported is 20, since all the peers have full cache;

- the shortest path length distribution is exposed as in the following:




from which it is shown that in the range $4 \rightarrow 9$ there are the most common path lengths, with a peak on the length 7, for which corresponds 9.000.000 paths; the second most common one is the one of length 6, with a corresponding frequency of about 8.750.000; in this range the less common path is the one of length 4, with a frequency of 4.500.000.

### 2.1.2 Middle case: 35 maximum neighbours

The experiments taken forward in this case fall in the same ones taken into account before.

#### Tiny number of participants: 1.000 peers

In this case, an example of initial free scale graph is the following one:

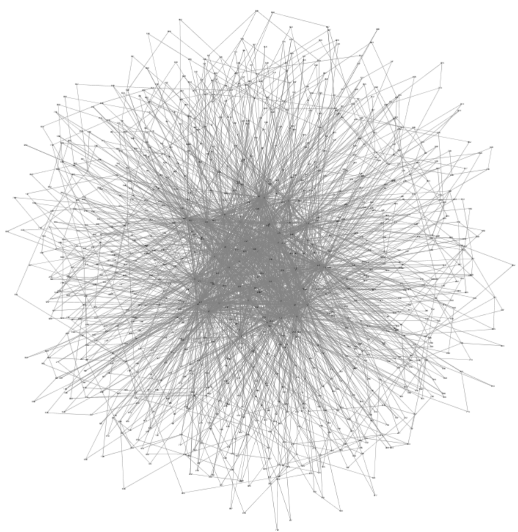


Figure 2.8: Example of scale free graph with 1.000 peers

After the algorithm has been executed, the graph obtained is the following one:

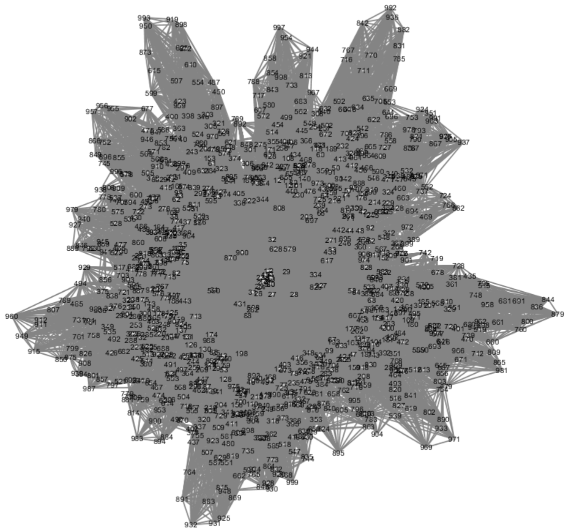
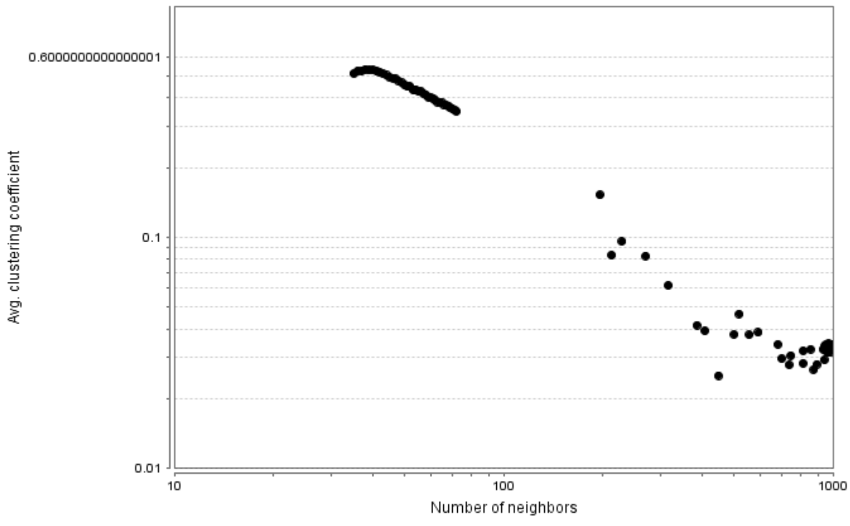


Figure 2.9: Example of final scale free graph with 1.000 peers

The statistics gathered have shown that the clustering coefficient is 0.471, since the graph is a single connected component, the network diameter is 7 and the out-degree is mostly equal to 35 (for about 975 nodes), except for very few other nodes that have an out-degree value of 33 or 34.

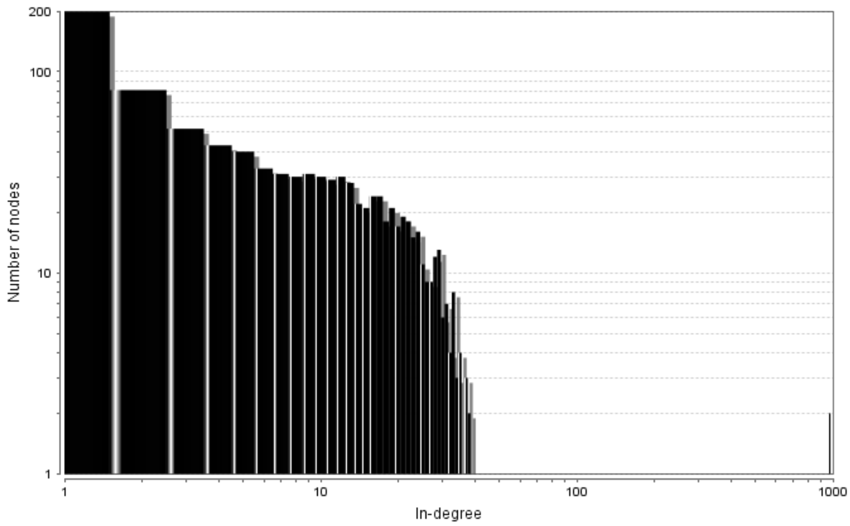
For the others statistics, we show briefly the results obtained in the following figures:

- the average clustering coefficient is shown in the following:




it illustrates that the average clustering coefficient does never go above the 0.55; from the picture above, we can see that for a number of neighbours that does not exceed the value 80, the clustering coefficient values are very near each other, starting from a value just above the 0.5 up to arriving to 0.35; while, instead, for the last ones, they don't go above the 0.05;

- the in-degree distribution as follows:




in which we can see that the most common in-degree values are the 1 and 2 with corresponding values of about, respectively, 198 and 82; the highest in-degree values don't exceed the value 1.000;

- the shortest path length distribution, that for this case is highest for the value 2 with a corresponding frequency of 510.000 and for the value 3 with a corresponding frequency of about 340.000; all the others values are beneath the 80.000 and the lowest one is 6 with a very low corresponding frequency.

**Small number of participants: 2.500 peers**

In the experiment pursued for this case, the graph obtained at the end of the execution of the algorithm has been the following one:

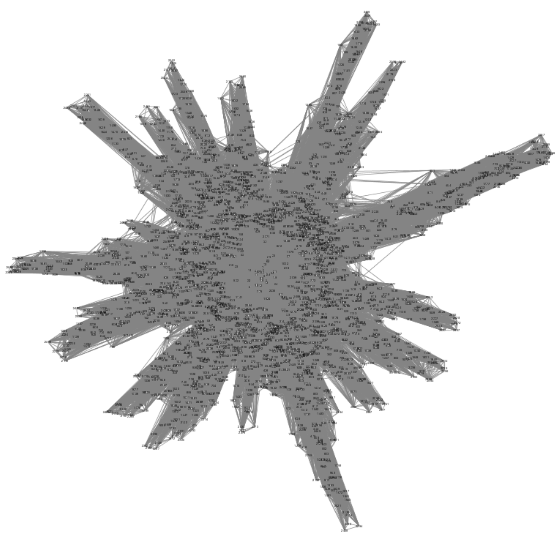
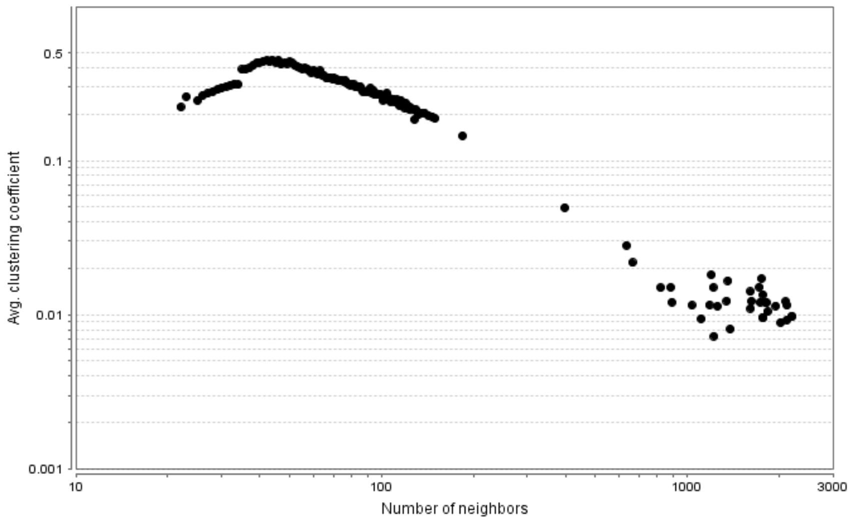


Figure 2.10: Example of final free scale graph with 2.500 peers

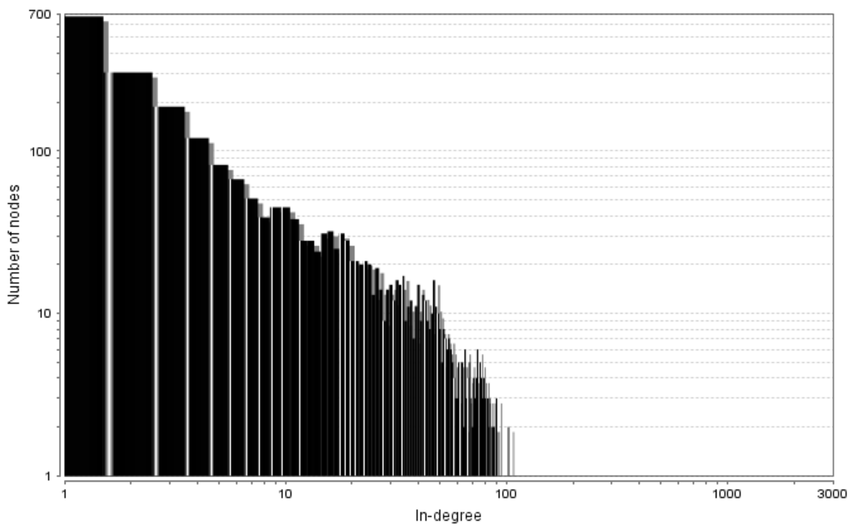
it is a single connected component with a network diameter equal to 10, a clustering coefficient equal to 0.377 and the following statistics:

- the average clustering coefficient, shown in the next picture:




it shows us that the clustering coefficient does never go above the 0.450 value; it follows an exponential distribution since it decays exponentially up to arriving to the 0.010 $\sim$ 0.015;

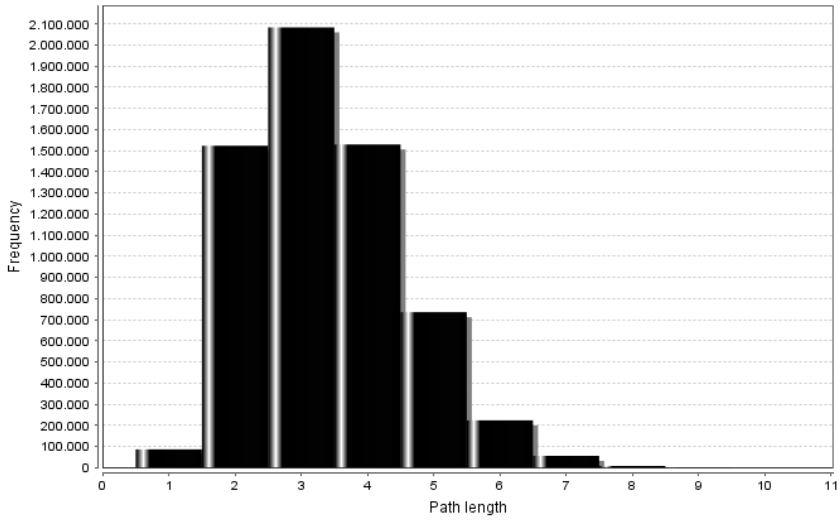
- the in-degree distribution of the network, shown in the following:




it shows that, for the most part, the nodes have 1 or 2 incoming edges. Overall, the number of nodes that have 1 or 2 incoming edges are about 970; the maximum in-degree value that some node can have is below the 3.000;

- for the most part of the nodes, the out-degree distribution, in this case, is equal to 35; there are very few nodes that have a non-full cache and the less full cache has 23 cache lines; the number of nodes that don't have a full cache are less than 300;

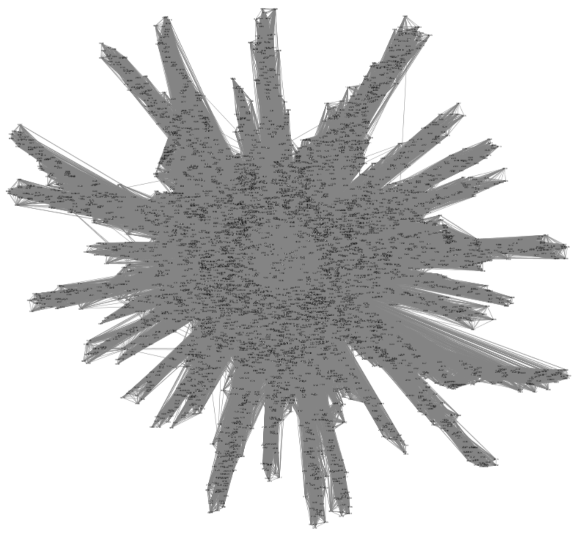
- the shortest path length distribution is shown in the following:




from this, we can see that there are few paths with length greater than 6 and that, mostly, the lengths of the paths ranges between 2 and 5, with a peak for the 3 value, with a corresponding frequency of about 2.080.000.

### Medium number of participants: 5.000 peers

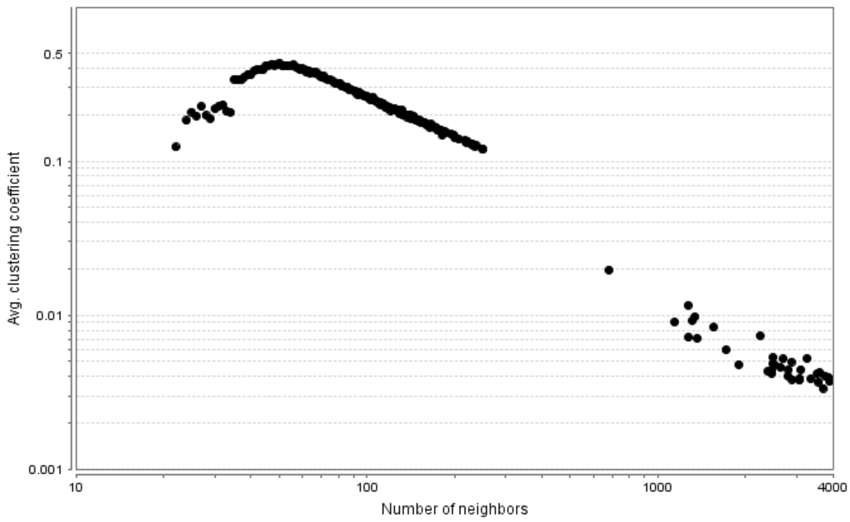
In the experiment taken forward in this case, the resulting graph is shown in the following:



The clustering coefficient obtained by the algorithm is equal to 0.335, by considering that the resulting graph is a single connected component with diameter equal to 12.

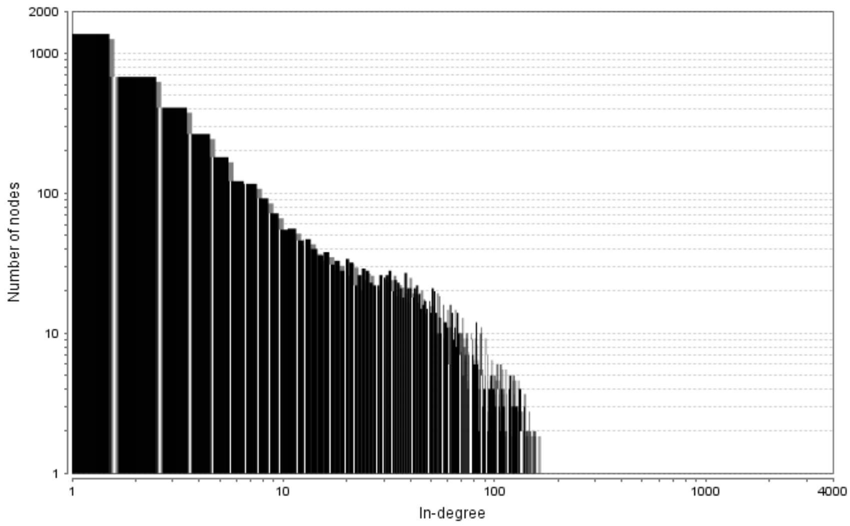
The following statistics have been analysed:

- the average clustering coefficient, shown in the following picture:




  it shows us that the clustering coefficient follows an exponential decay after the peak on about 0.435 $\sim$ 0.440, with a corresponding number of neighbours equal to about 50 $\sim$ 55;

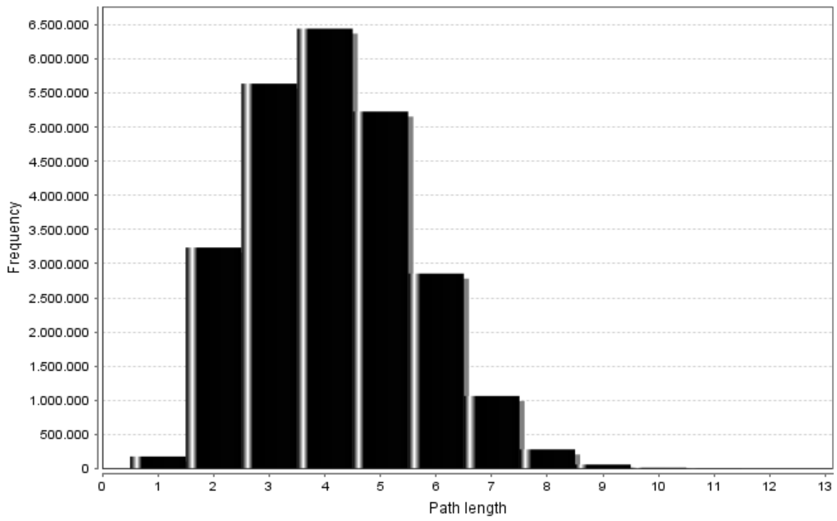
- the in-degree distribution as shown below:




  in which the number of nodes that have an in-degree of 10 $\sim$ 20 are the most common ones, instead of the few that have an in-degree of over 100;

- the out-degree distribution that shows us that the most common outgoing number of edges from a single node is the 35 value, and there are very few other nodes that have a smaller degree; the least full cache is filled with 24 peers;

- the shortest path length distribution that is obtained from the execution of the algorithm, is the following one:




from which, we can see that the most common length of a path is 4, with a frequency just below of 6.500.000; the most frequent paths are of length that ranges between 2 and 6, with a frequency of about, respectively, 3.200.000 and 2.800.000.

**Large number of participants: 7.500 peers**

In this case, the execution of the algorithm returns a single connected component with a clustering coefficient equals to 0.335; the network diameter in this graph is 14. The resulting graph is exposed below:

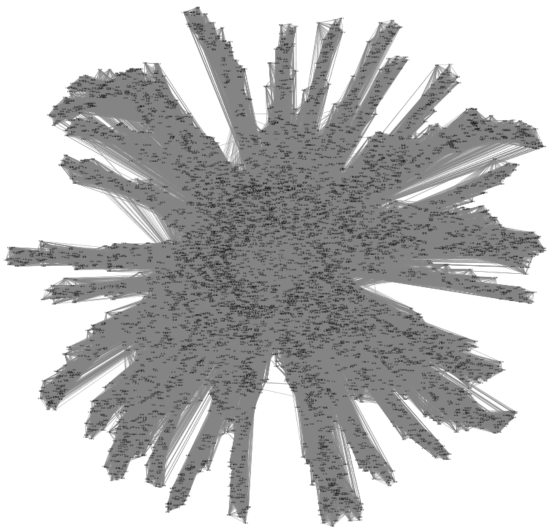
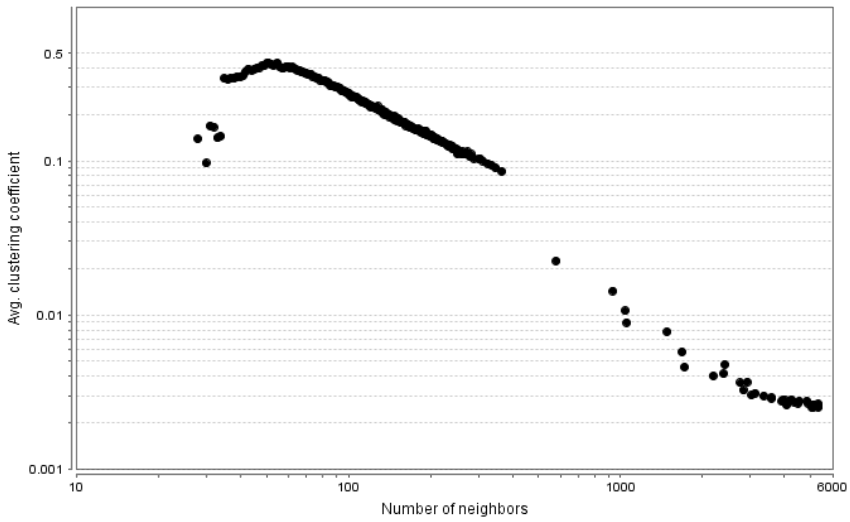


Figure 2.11: Example of final graph with 7.500 peers by using a cache of size 35

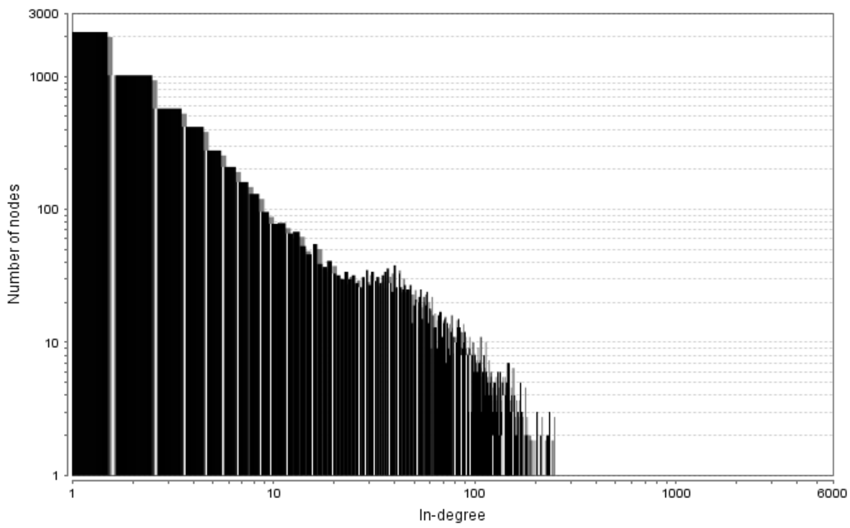
The statistics that have been analysed are shown below:

- the average clustering coefficient is as follows:




from this we can see that the distribution that this statistic follows is the exponential one, since the decay is exponential; the highest value reached is about 0.430 for a corresponding number of neighbours equal to about 55 $\sim$ 60; afterwards, the values decay as the number of neighbours increases up to arriving to just above the 0.000 value, where the corresponding of neighbours is about 5.600;

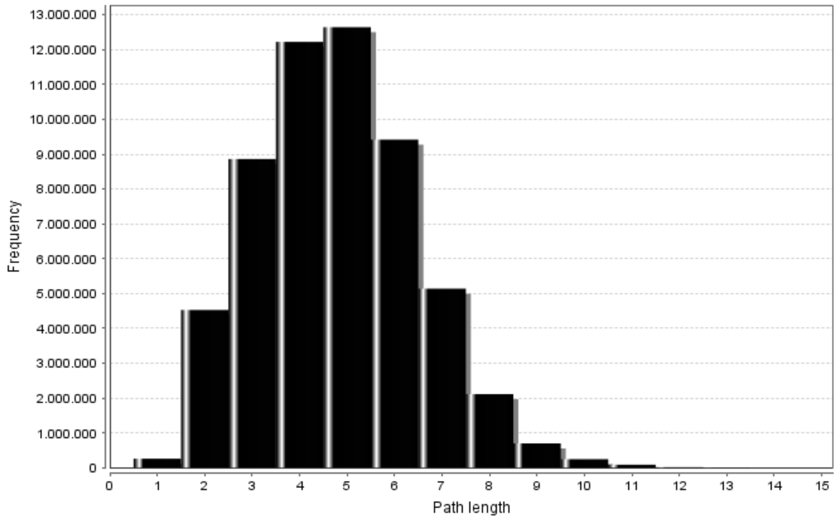
- the in-degree distribution is shown in the following:




from which, we can see that the most common in-degree value is 1, with a corresponding number of nodes of about 2.160 and that there are few nodes that have an in-degree value less than 10;

- the out-degree distribution statistic, which is not shown for this case, gives us the same out-degree value for all the nodes, that is 35, because almost all the nodes are connected to exactly 35 neighbours; with the exception of very few that have a cache at least filled with 31 neighbours;

- the shortest path length distribution which is shown in the following:




shows us that the most common paths range from 2 up to 7, with a frequency, for these extremes of about, respectively, 4.500.000 and 5.000.000; the most frequent path is the one of length 5, with a corresponding frequency of about 12.700.000.

### 2.1.3 Dense case: 50 maximum neighbours

The experiments taken forward in this case are the same of the previous considered cases.

#### Tiny number of participants: 1.000 peers

In the experiment taken into account, the algorithm returns a single connected component graph with a clustering coefficient equals to 0.621 and a diameter of 6. Since the number of peers is quite low if compared with the cache size of a single peer of the network, the graph results in a picture, as the following one:

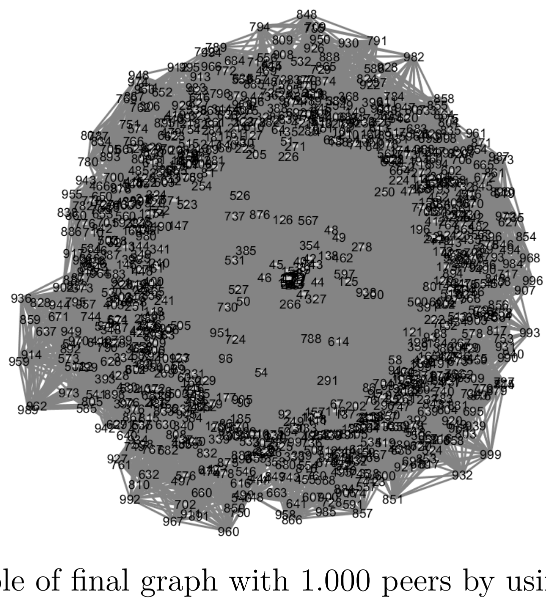
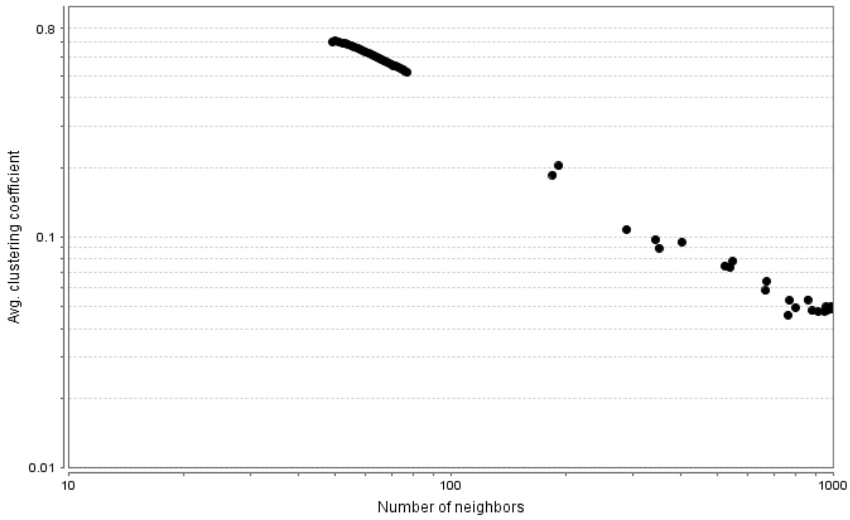


Figure 2.12: Example of final graph with 1.000 peers by using a cache of size 50

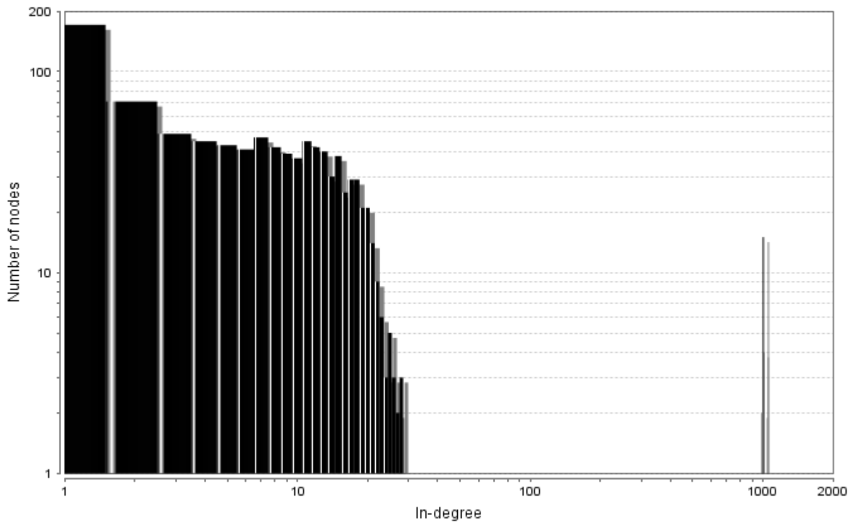
The statistics that have been considered for this situation are the following ones:

- the average clustering coefficient, shown in the following:




  that, from the peak of just above 0.70, it decays as the number of neighbours increases to about 0.05; from the picture above, the clustering coefficient of 0.621, reported earlier, finds its justification;

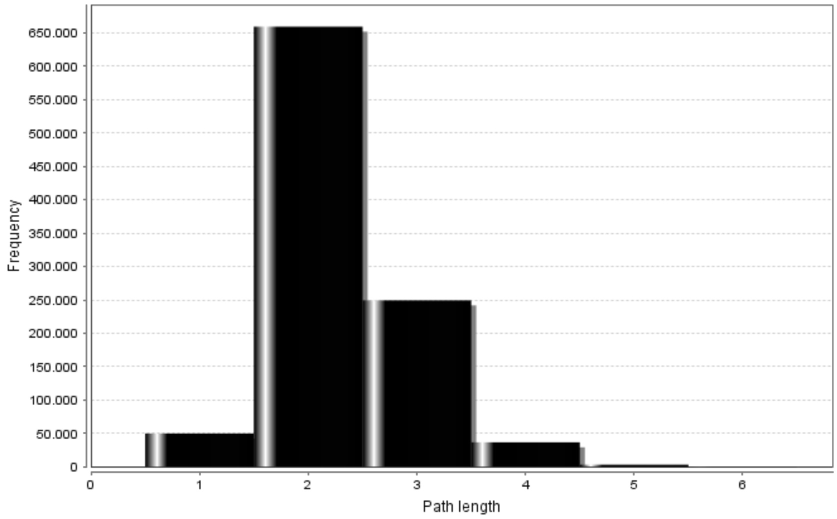
- the in-degree distribution, shown as follows:




  which shows us that for most of the nodes, the in-degree is 1 and that there exist a not so little part of nodes with an in-degree up to 1.000; in the range from 1.000 up to 2.000, instead, the number of peers that have an in-degree in that interval are very few;

- the out-degree is not shown, since, for most of the nodes, this value is equal to 50 and very few other nodes have instead the value 49;

- the shortest path length distribution is shown in the following:



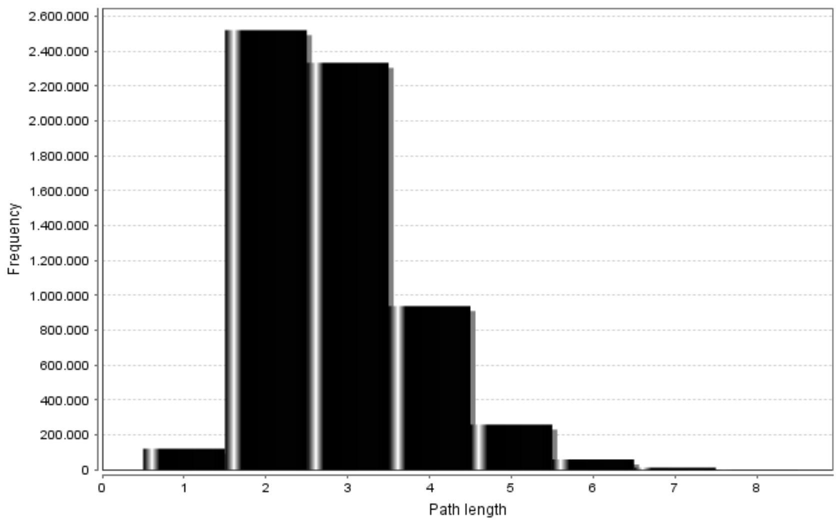

it let us evince that the most common shortest path length is the one of length 2, with a frequency of about 660.000 and the second most common one is the one with value 3 and a corresponding frequency of about 250.000.

### Small number of participants: 2.500 peers

In this experiment, the algorithm returns a single connected component graph with a clustering coefficient equals to 0.466 and a network diameter equal to 8. In this case, since the resulting graph is a single connected component, it's not significant to show the initial and final graph, given the high number of peers and edges between them.

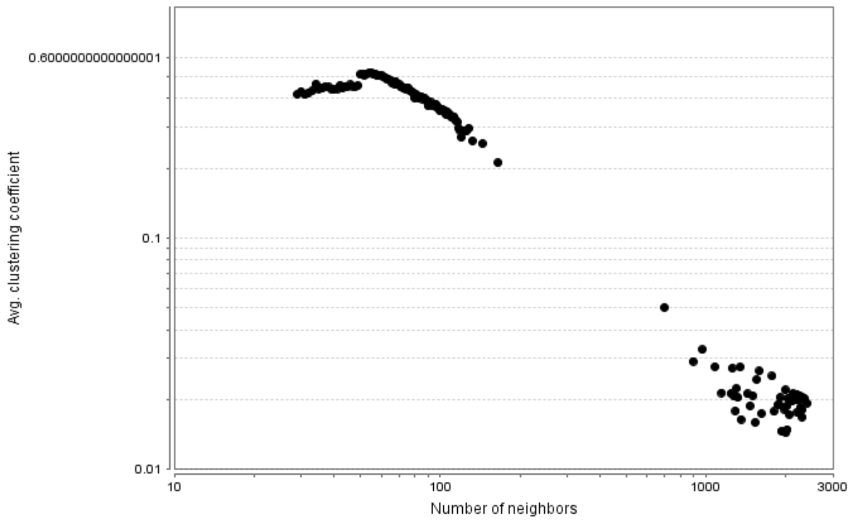
The statistics taken into account show us that:

- the shortest path length distribution, for this experiment is shown in the following:



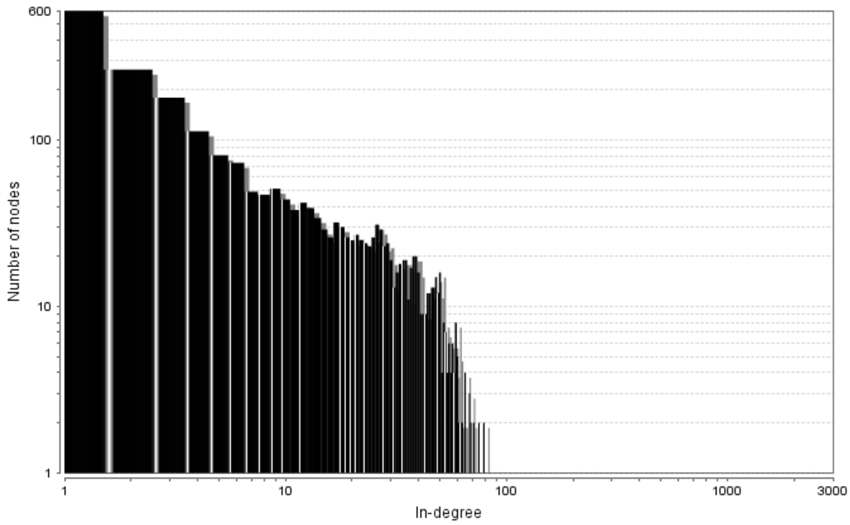

and it exhibit two most common path length values that are 2 and 3 with a frequency, respectively, of $\sim 2.550.000$ and of $\sim 2.370.000$;

- the average clustering coefficient, obtained from the network analysis is justified by the following:




  that shows us the behaviour of this coefficient as the number of neighbours varies: we have that the highest clustering coefficient obtained is just below the 0.525, that corresponds to $56 \sim 57$ corresponding value for the number of neighbours; it decays up to arriving to values just beneath the 0.025 value, that corresponds to number of neighbours that ranges from about 1.000 to 2.600;

- the in-degree distribution is the following one:




  from which we can see that the most common in-degree is found for values such as $1, 2$ or 3, and, overall, the number of nodes that have these values are about 1.050; the frequency of in-degree values gets always lower up to about 100; from 100 up to 3.000 there are sporadic nodes that have a so high value of in-degree;

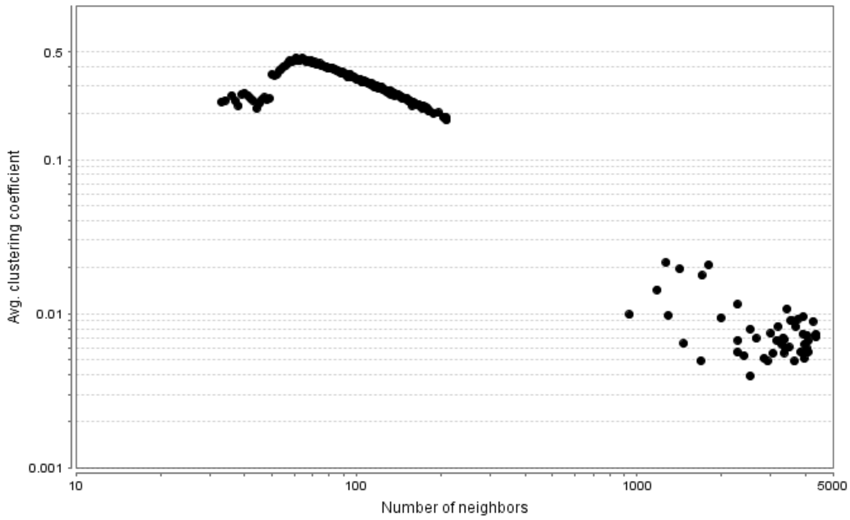
- the out-degree distribution is not shown, since mostly, the nodes have an out-degree value of 50; for the about 550 peers that doesn't have the cache totally filled, the minimum number of busy lines in the cache is 29, and the number of peers that have an out-degree that gets closer to the 50 value, increases always more.

**Medium number of participants: 5.000 peers**

In this experiment, as before, the algorithm returns a single connected component where the network diameter is 10 and the clustering coefficient is 0.355.

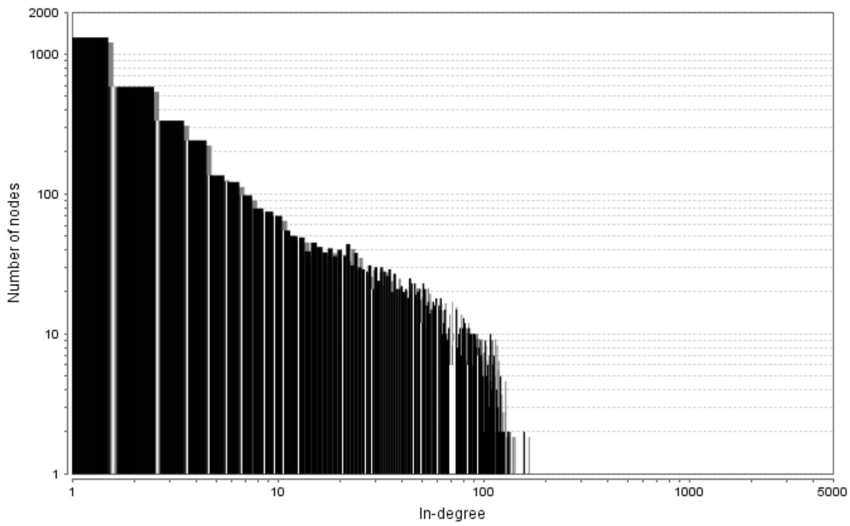
The analysed statistics have shown that:

- the average clustering coefficient, shown in the following:



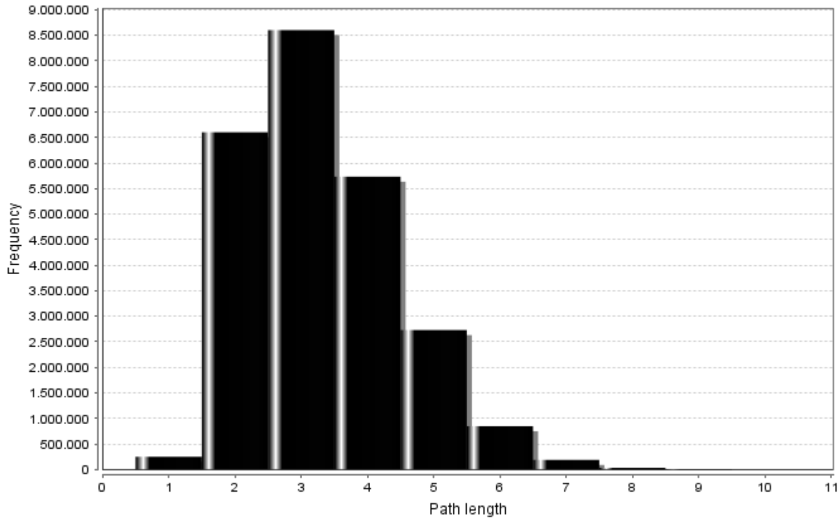

  justifies the clustering coefficient obtained from the network analysis; the highest value is just above the 0.45 and then the clustering coefficient decays when the number of neighbours increases, up to arriving to the values in the range from 2.000 to 5.000, where the corresponding clustering coefficient is very near to 0;

- the in-degree distribution is the following one:




  and it shows us that for this kind of graph, the more common values are the first two, with a corresponding overall frequency of about 1.900, the maximum in-degree value that can be reach is lower than 5.000, with a frequency very near to 0;

- the out-degree, that is not shown, results in a value equal to 50 for less than 4.500 nodes, but the remaining nodes have an out-degree in the range from 36 to 49;
- the shortest path length distribution is shown in the following:

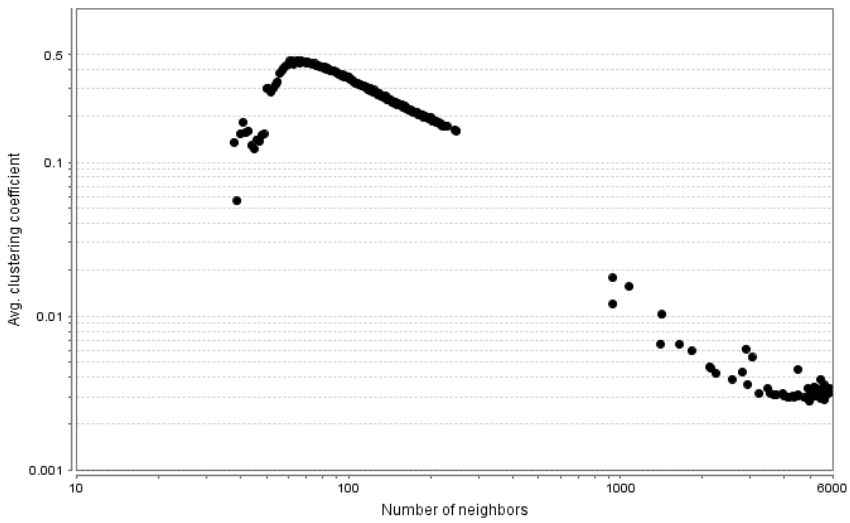



and it shows us that the most common path is of length 3 and the second one is the one with value 2, with respectively, a frequency of about 8.570.000 and 6.550.000.

**Large number of participants: 7.500 peers**

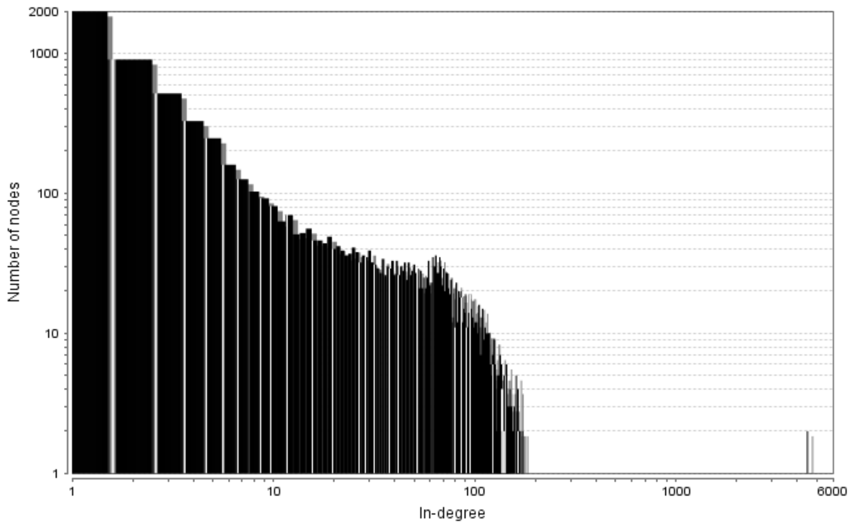
In this experiment, still, the result gives us a single connected component graph with a diameter of 11 and a clustering coefficient of 0.318. The analysis of the statistics is shown below:

- the average clustering coefficient is shown in the following:




in this case, the decaying curve involved ranges from about 0.46 for about 70 neighbours up to 0.16 starts for about 260; we can see also that starting from a number of about 900 neighbours up to 6.000 neighbours, this coefficient is very near to 0;

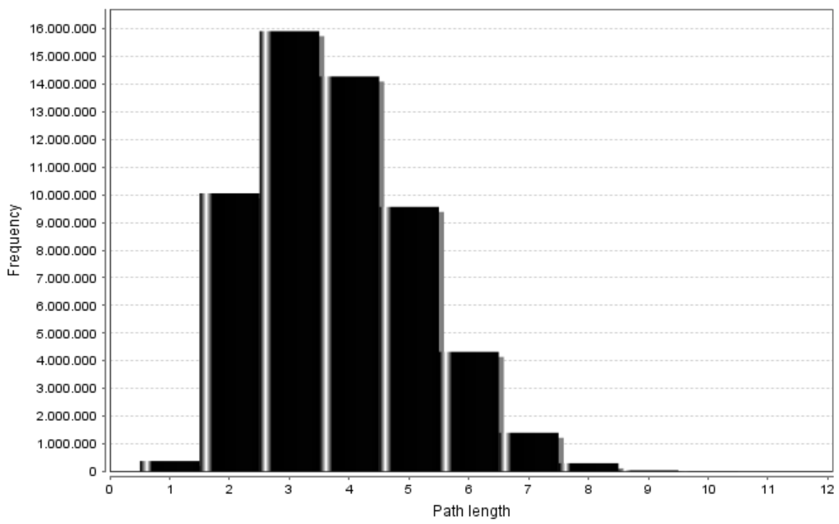
- the in-degree distribution shows us that the nodes that have an in-degree value equal to 1, 2 or 3 are the most common ones, as can be seen by the following picture:




while, instead, as the in-degree increases, the number of nodes that have that in-degree decay up to about 170; the number of nodes that have an in-degree that ranges from about 170 up to 6.000 are so few that are not rendered in the figure above;

- about the out-degree distribution, except for less than 300 nodes, for which their value ranges from 42 to 49, the most common value for the nodes is equal to 50;

- the shortest path length distribution can be analysed by the following histogram:




It shows us that the most common path length is the one of value 3, with a frequency of about 41.8 millions; the second most common shortest path is of length 2 and has a frequency of about 13.5 millions.

## 2.2 Grid network results

As before, about the grid network results, the analysis has been divided depending on the value of the cache line, because in this way we can see how the graph evolves taking into account that the maximum number of neighbours of the peers are the ones defined by the cache line value. All the results shown start from an initial graph composed by a single connected component.

In this case, instead of splitting the analysis by cache lines, it is divided by grid network.

As before, the results about the average clustering coefficient and the in-degree distribution has been shown in *log-log scale*, for a better analysis of the results.

### 2.2.1 Tiny number of participants: 10x100 grid network

In this case, we have a network composed by peers placed in a rectangle fashion, in which for each of the 10 rows, each row is composed of 100 peers interconnected as explained in the implementation section.

In these experiments we start from a network as follows:

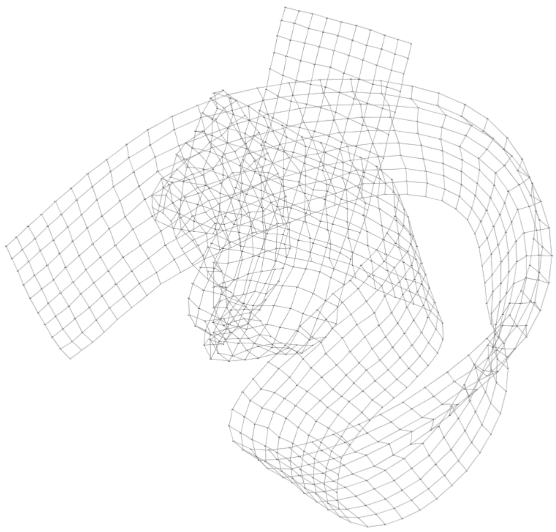


Figure 2.1: Starting $10 \times 100$ grid network

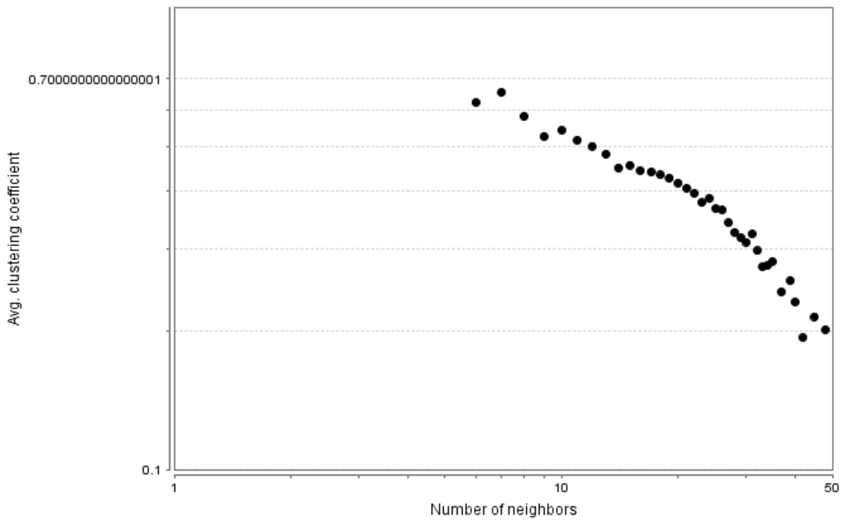
### Sparse case: 20 maximum neighbours

After the execution of the algorithm, for this case, we have that the resulting graph is the following one:

Figure 2.2: Final 10 × 100 grid network

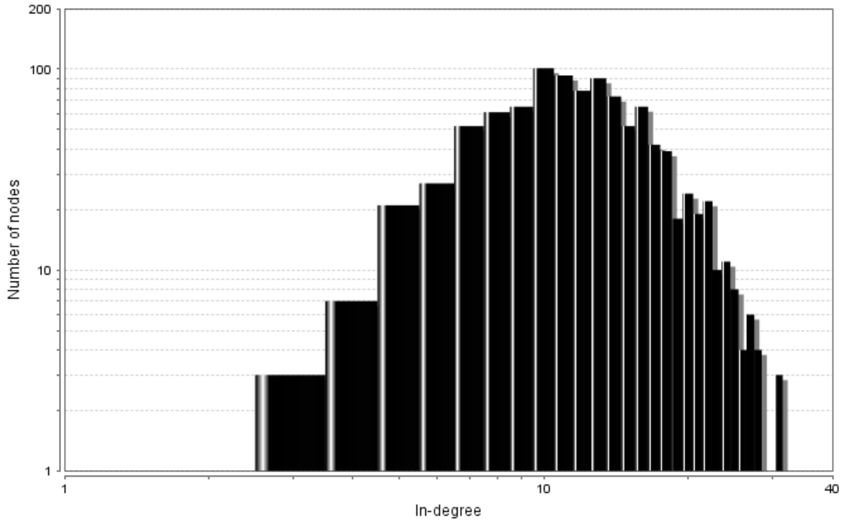
The results of the analysis shows us that the clustering coefficient is 0.430 and that the resulting graph is a single connected component; the network diameter of this graph is 53, since it develops on the width of the graph. Besides, we have that:

- the average clustering coefficient, that can be shown in the following:



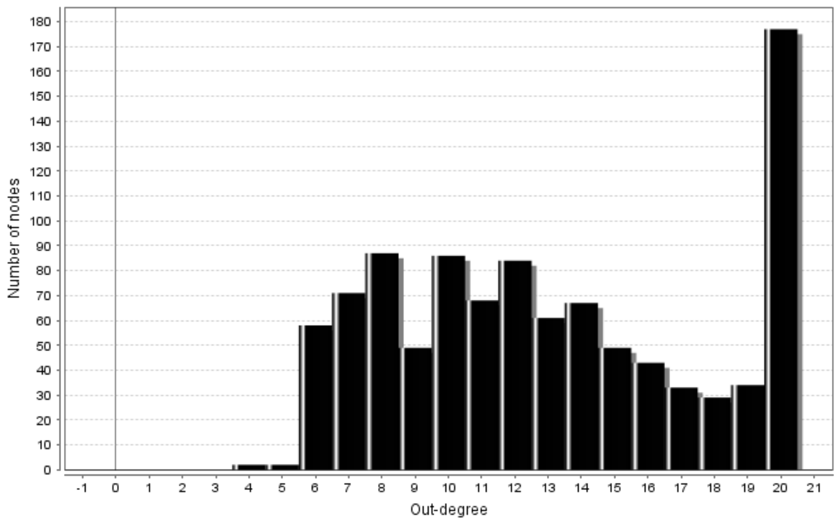

has as maximum value 0.65 and it decreases up to arriving at just below than 0.20; from this, we can see that the clustering coefficient mentioned earlier makes sense;

- the in-degree distribution, that can be shown as follows:



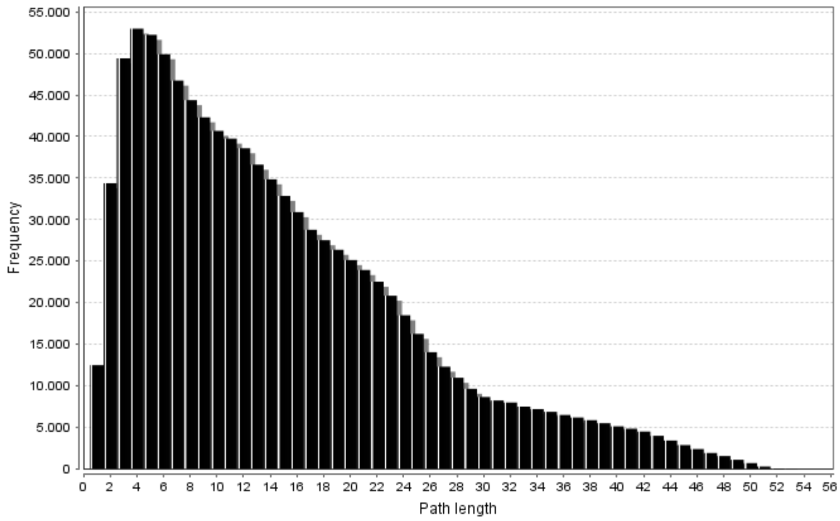

has as most common in-degree values the ones in the interval $10 \sim 13$;

- the out-degree distribution, that can be shown in the following:




has very much nodes that do not have an out-degree equal to 20, since the network is organized as a grid;

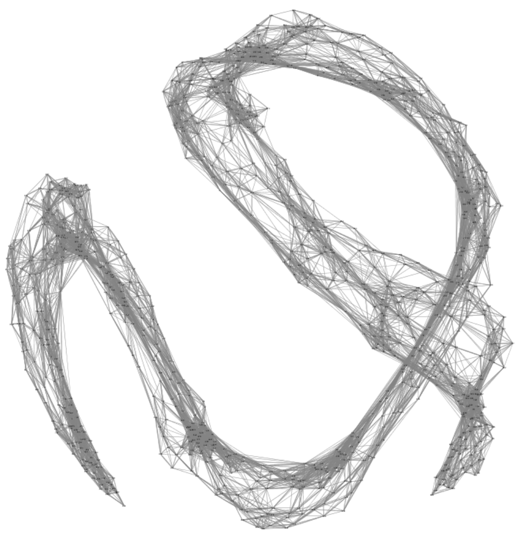
- the shortest path length distribution, that can be shown as follows:




  the results obtained in this kind of graph are very much different from the ones in the free scale networks, because, since the network is organized as a grid, the paths that can be taken into account are very much longer; in the picture this corresponds to the range from 20 $\sim$ 22 up to the ending value, 56.

**Middle case: 35 maximum neighbours**

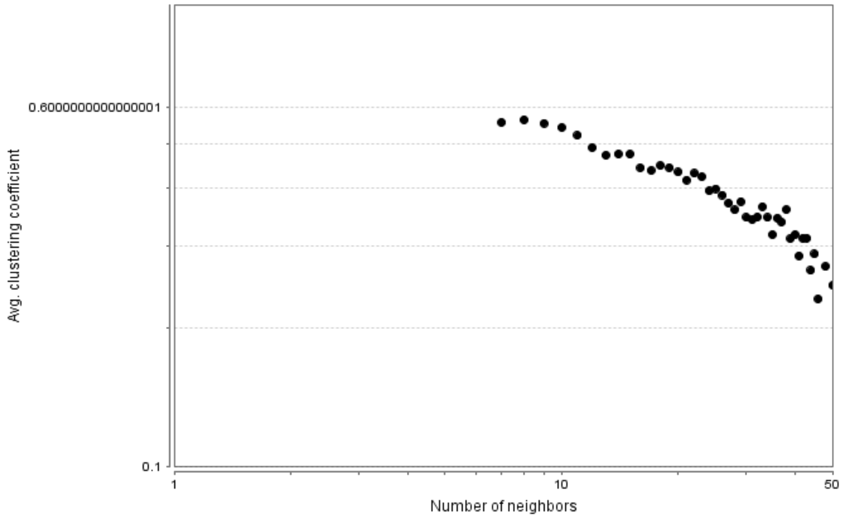
This time, the graph obtained is more dense as can be seen in the following:



It has a clustering coefficient value of 0.438 and the network diameter measures 51 peers.

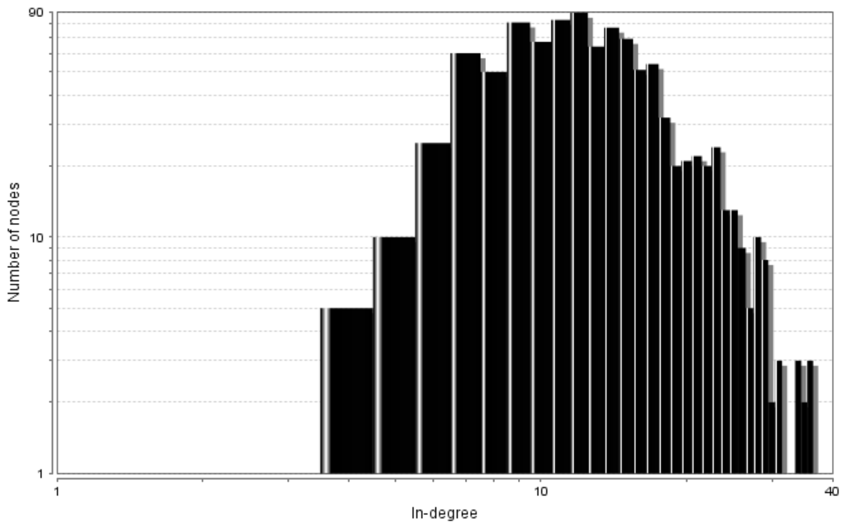
The other statistics have shown that:

- the average clustering coefficient, shown in the following:



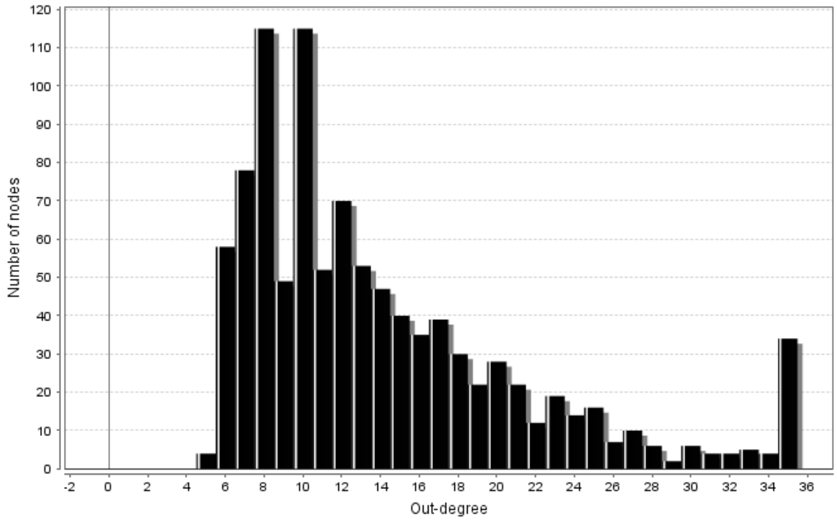

  justify the values mentioned before as clustering coefficient value. It results to be quite high, since the graph is totally connected;

- the in-degree distribution, that we show below:



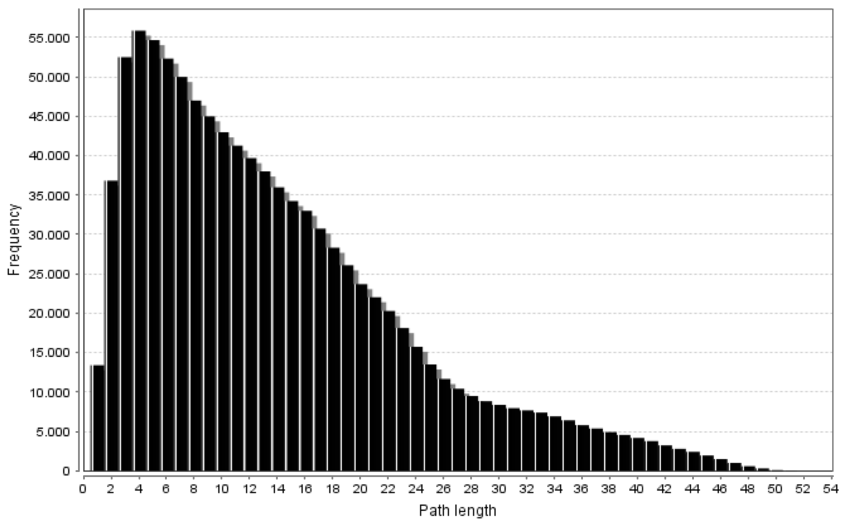

  finds its most common values for an in-degree value in the range from 7 to 17;

- the out-degree distribution, that is shown in the following:




does exhibit a totally different behaviour of the previous kind of graphs; indeed, we have that very few nodes have a totally full cache if compared to the situations for the free scale network, while instead there so many other nodes that have a total of 8 or 10 neighbours (this is due to the grid composition);

- the shortest path length distribution, shown as follows:



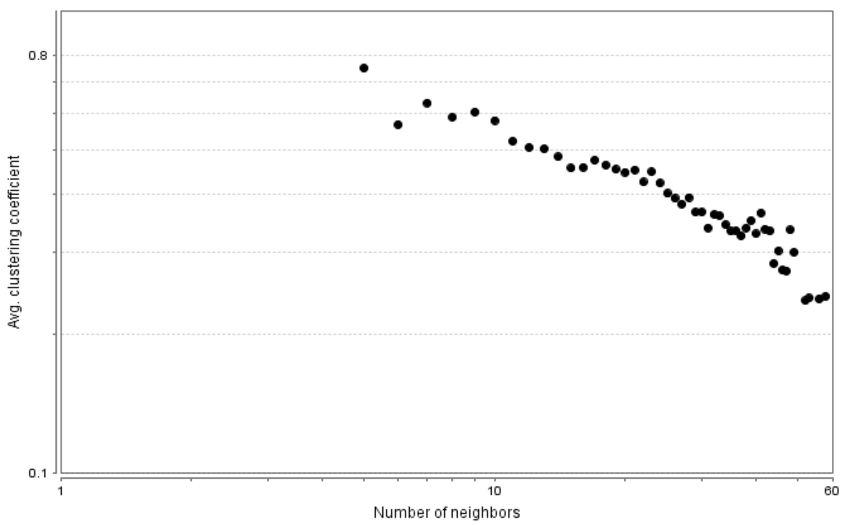

that exhibit the same behaviour of the *20-lines-per-cache*; indeed, we have that we may have possible quite long paths, but the most frequent ones are those of short length, so in the range $3 \sim 10$, with a frequency of at most 55.200.

**Dense case: 50 maximum neighbours**

When the number of lines of cache increases, we should obtain a graph more dense of the previous one. The resulting graph for this situation is shown below:

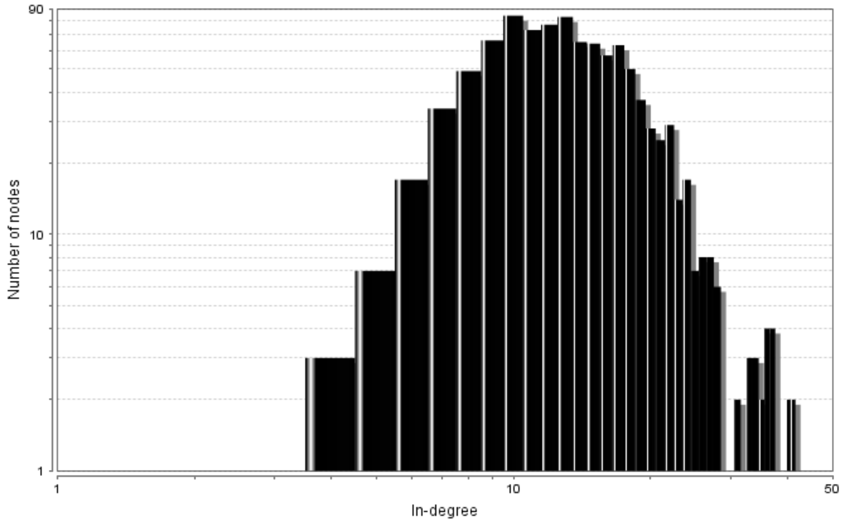
The statistics gathered have shown that it has a clustering coefficient value of 0.443 and the network diameter measures 51 peers as before. For the others statistics, we briefly show them in the following:

- the average clustering coefficient is shown as follows:



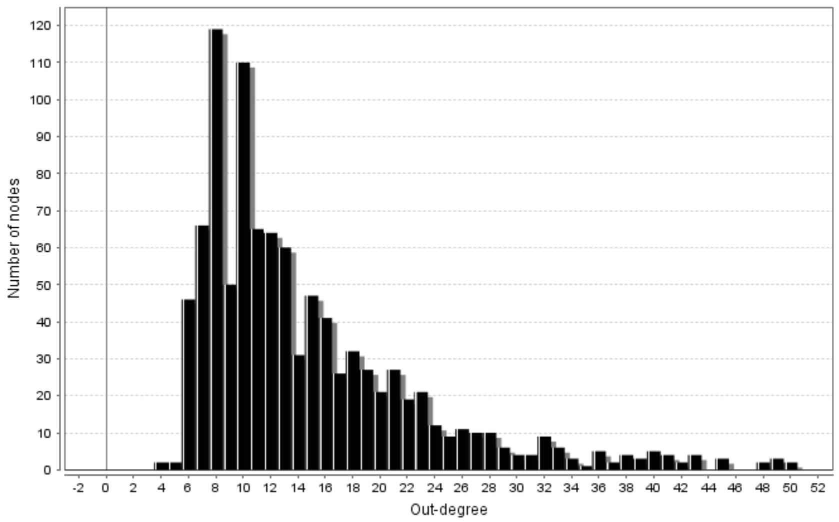

It shows that the highest clustering coefficient value gathered is 0.75, since there exists a peer that has three quarters of the maximum of edges existing between he and his neighbours, and it decreases as the number of neighbours increases, up to arriving to just below 0.25;

- the in-degree distribution, shown as follows:



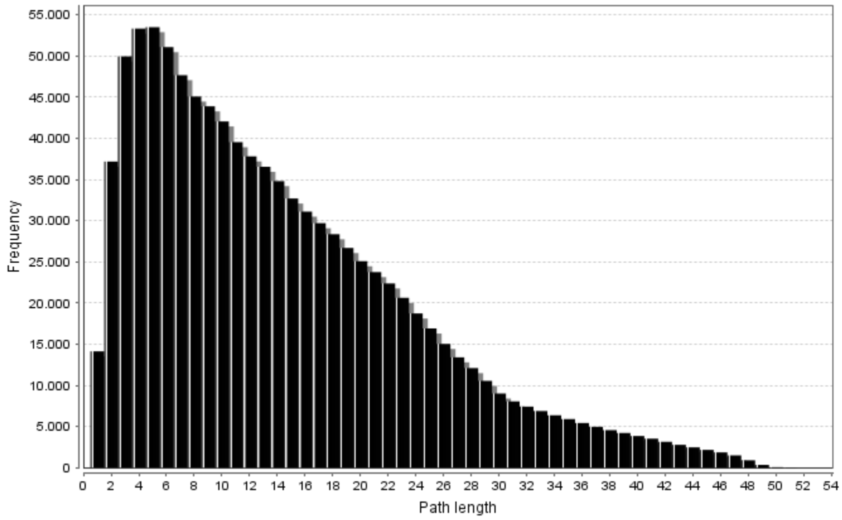

it shows us that the most common in-degree value is between the values $8 \sim 18$;

- the out-degree distribution is shown in the following:




It shows us that the caches of the peers are averagely composed of $7 \sim 13$ neighbours, instead of being totally full and the value to be focused on 50;

- the shortest path length distribution, illustrated as follows:




that shows us the range of the most common path length values, that is, for this case from 3 $\sim$ 11 up to 24 and with a corresponding frequency of at most 52.800 $\sim$ 53.000.

### 2.2.2 Small number of participants: 25x100 grid network

In this case, the network that we'll use as basis for the experiments by using 20, 35 and 50 cache lines, is the following one:

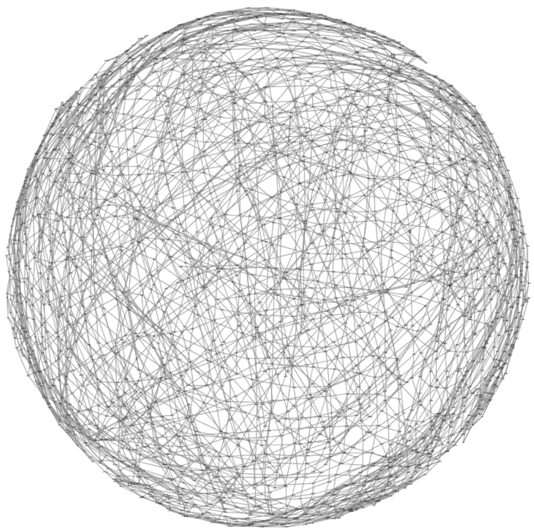


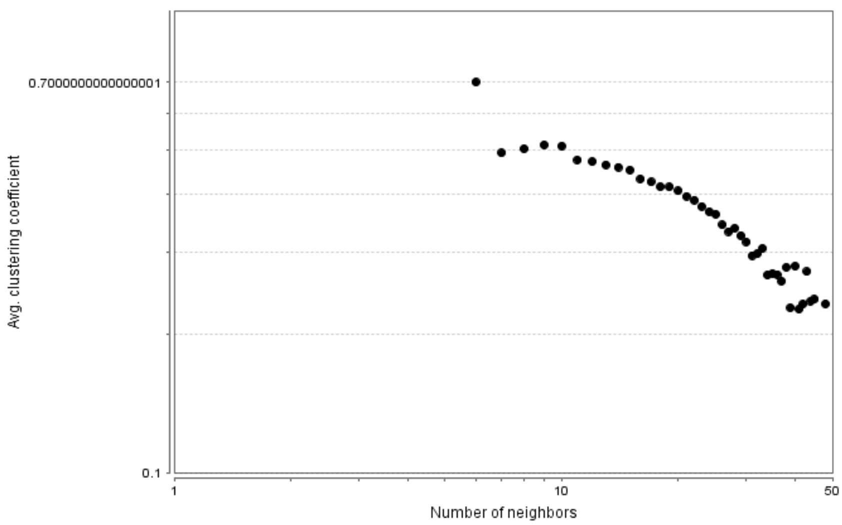
Figure 2.3: Example of initial graph with 2.500 peers

**Sparse case: 20 maximum neighbours**

After the execution of the algorithm, in this case, we have that the resulting graph is the following one:

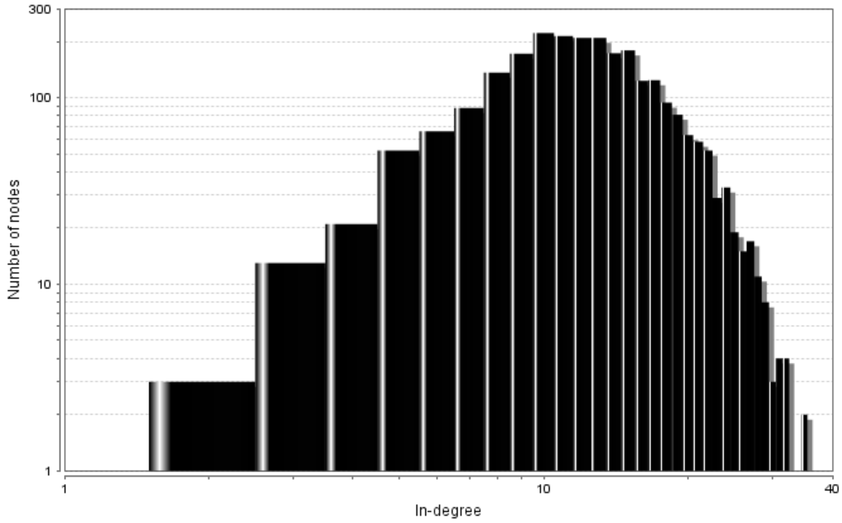
The resulting analysis shows us that the clustering coefficient is 0.407 and that the resulting graph is a single connected component; the network diameter of this graph is 58, since for this measurement it follows the width of the graph. Moreover, we have that:

- the average clustering coefficient, that can be shown in the following:



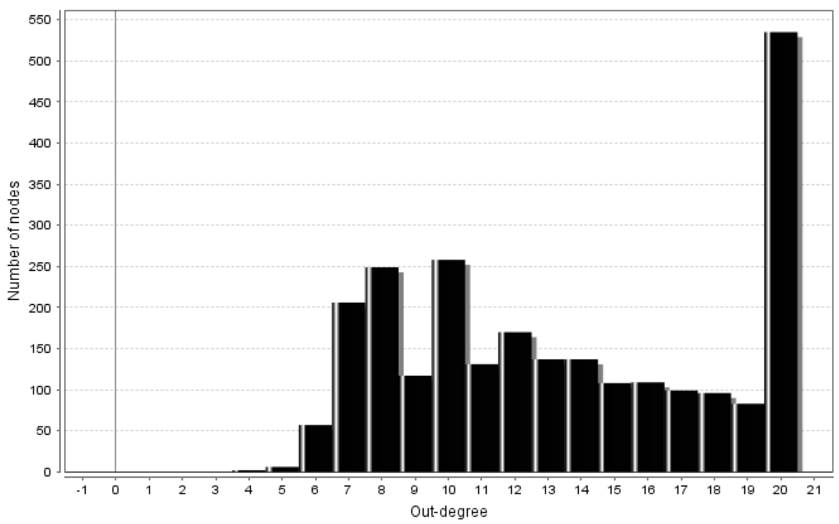

It shows us that except for a number of neighbours equal to 6, in which the average clustering coefficient is pretty high (with a value of 0.7), all the others starts from a value near to 0.5 and decrease up to just above 20. This behaviour is justified by the shape of the network;

- the in-degree distribution, shown as follows:



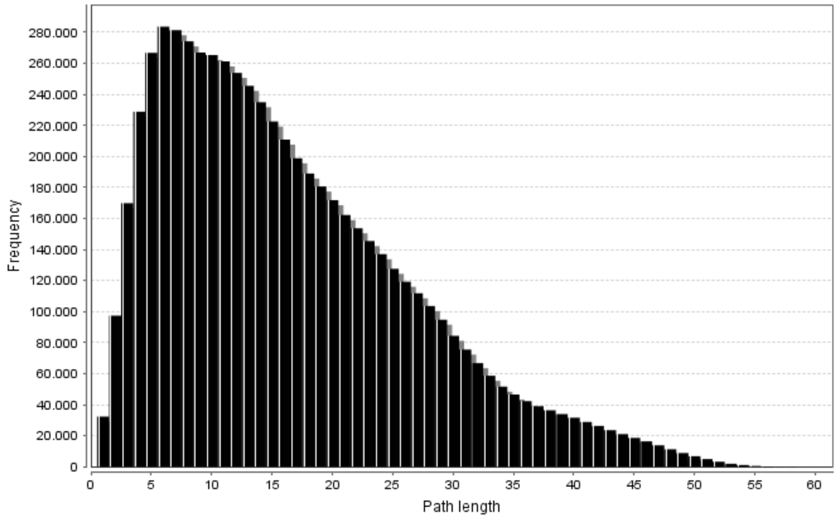

that is very common for values that ranges from 8 to 17;

- the out-degree distribution, that is shown below:




that illustrates that the most common out-degree value is the 20 value, since it is the value of the cache lines involved in this experiment;

- the shortest path length distribution, as below:




which shows us that there exists very long paths that can be achieved, up to arriving to a length of 55; this histogram shows us also that there is a inverse correlation between the shortest path length distribution and the frequency in which it occurs, regardless of the initial $4 \sim 5$ values of the path length.

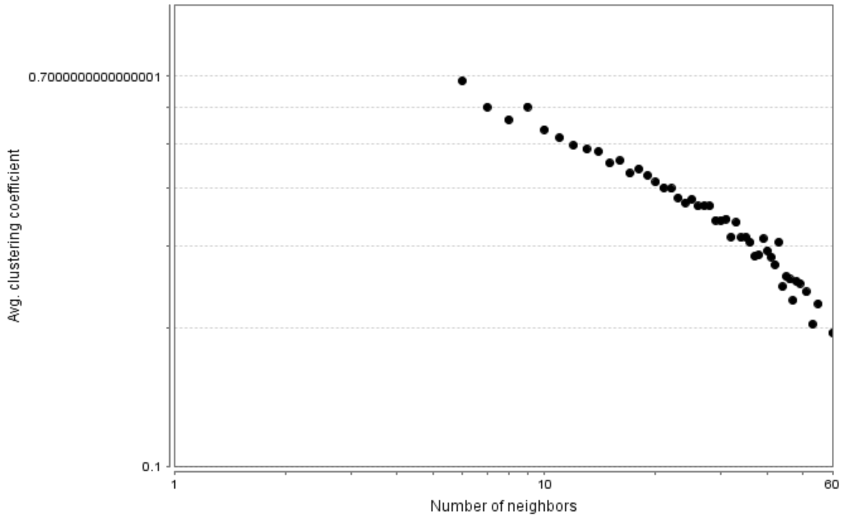
### Middle case: 35 maximum neighbours

In this other case, the resulting graph obtained after the execution of the algorithm is the following one:

From the statistic analysis, we have that it has a clustering coefficient value of 0.422 and the network diameter is of 51 peers.

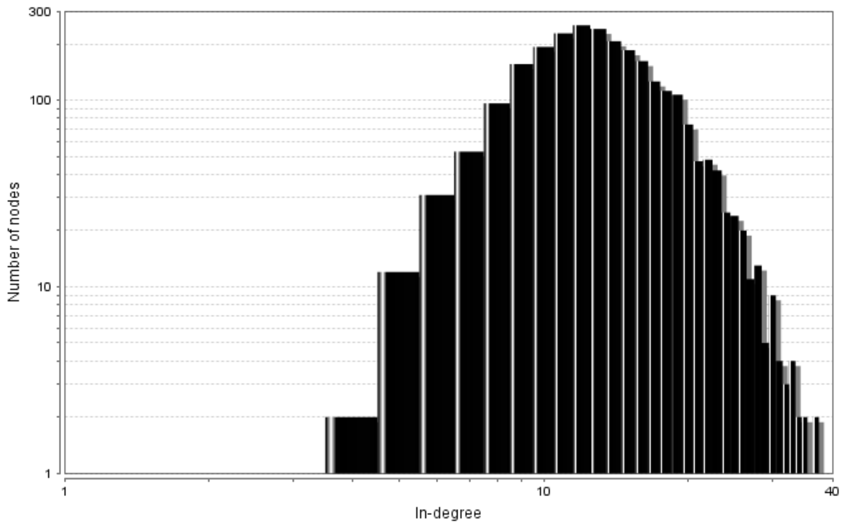
For the other statistics, they are analysed in the following:

- the average clustering coefficient is the following one:



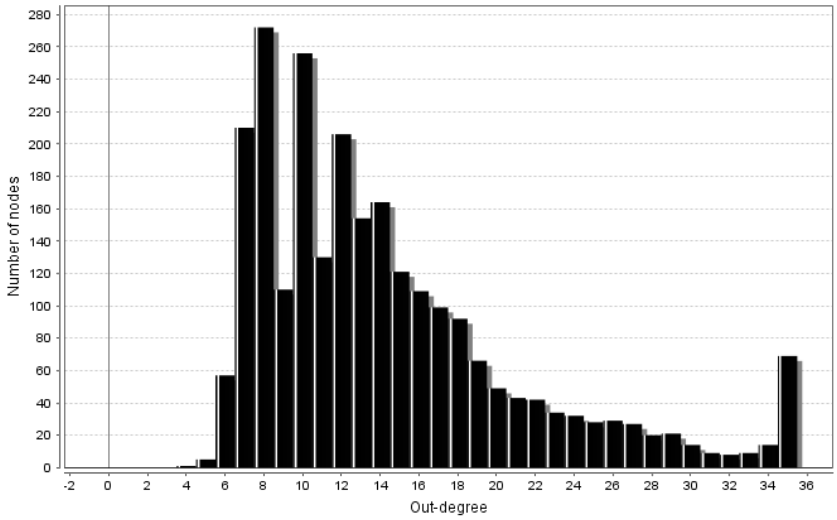

  It illustrates as the clustering coefficient decreases when the number of neighbours starts growing, from the initial greatest value of $0.68 \sim 0.69$ to the final lowest value of 0.20;

- the in-degree distribution is shown as:



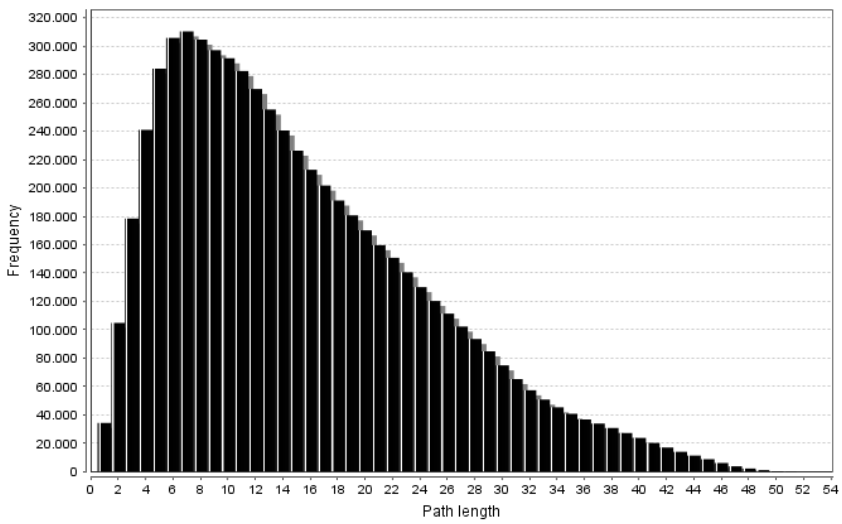

  This distribution shows that the most common in-degree values of the peers are those in the range from 9 to 13;

- the out-degree distribution is shown below:




We can see from this, that the number of neighbours with a full size cache are very few if compared to the ones that have an out-degree value of 8 or 10; the reason behind is given by the shape of the graph composed by the peers;

- the shortest path length distribution is shown in the following:



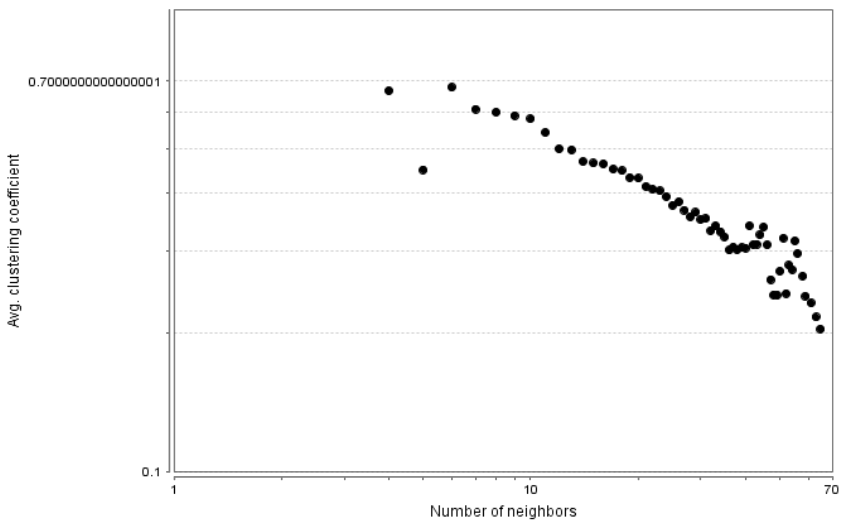

It shows that, as it's safe to assume, the number of paths decreases as the length of the path increases.

**Dense case: 50 maximum neighbours**

In this experiment, we have obtained a graph as the following:

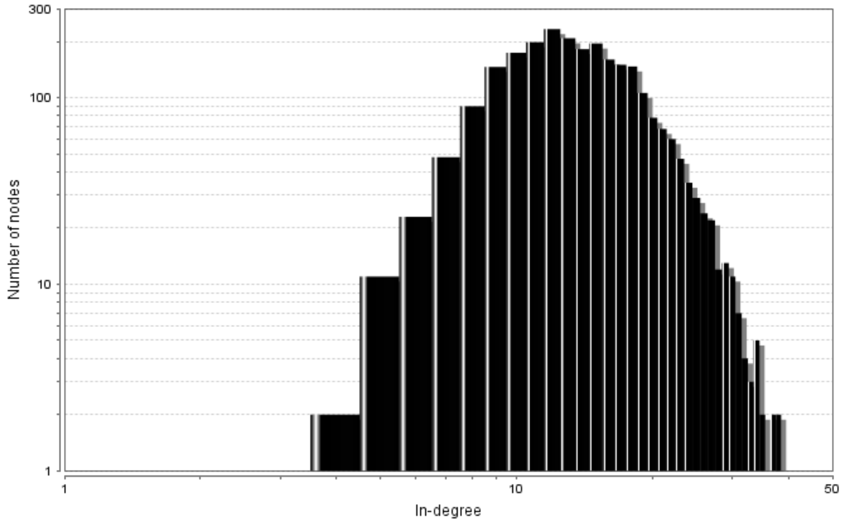
This graph has a clustering coefficient of 0.425, it's composed by a single connected component and it has a diameter equals to 51 peers. The statistics have shown that:

- the average clustering coefficient, shown in the figure below:



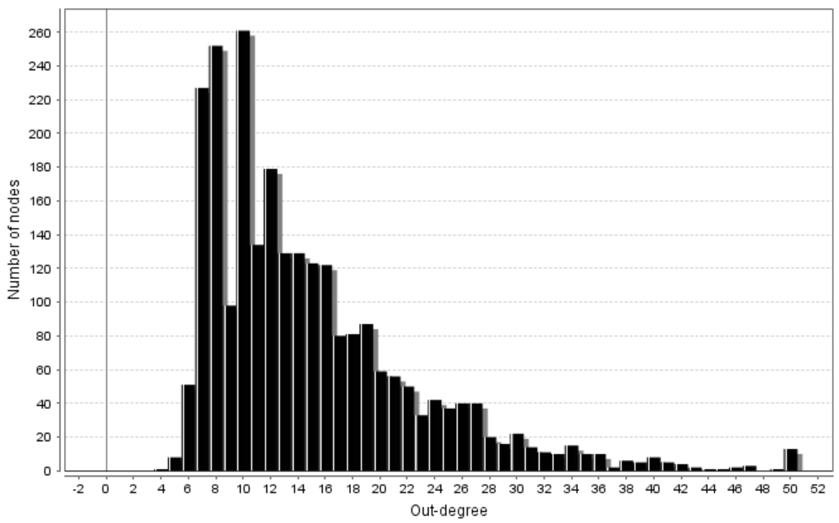

  justifies the value mentioned earlier, starting from the highest one of 0.67 until we get the lowest one of 0.20;

- the in-degree distribution, shown as follows:



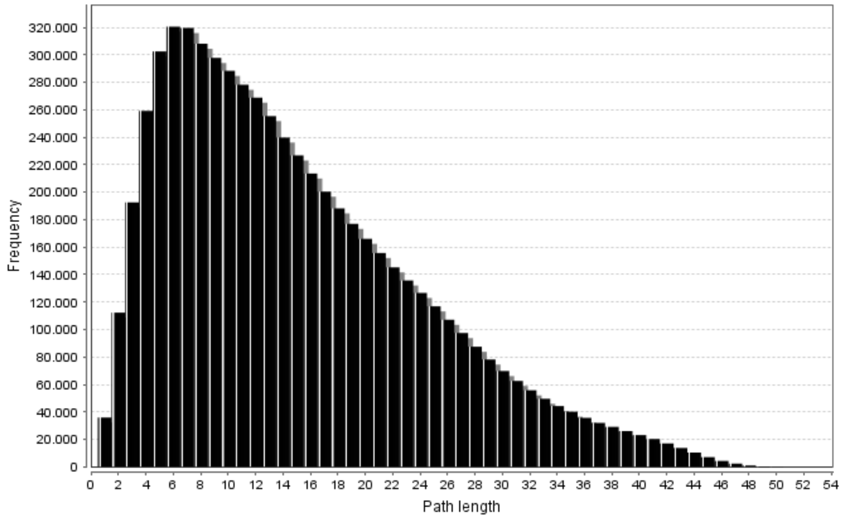

allows us to analyse the most common in-degree value in the nodes; this value oscillates between 9 and 18, with a frequency, respectively, of $\sim 145$ both;

- the out-degree distribution, as in the following:




that allows us to deduce that the most common out-degrees of the peer in the network belong to the interval $7 \sim 10$, with an overall number of nodes equal to 830 nodes;

- the shortest path length distribution, as follows:




and, as told before, there is an inverse correlation between the frequency and the length of the paths, with possible paths of length up to 48.

### 2.2.3 Medium number of participants: 50x100 grid network

In this case, the network used as input of the algorithm, by using 20, 35 and 50 cache lines, is the following one:

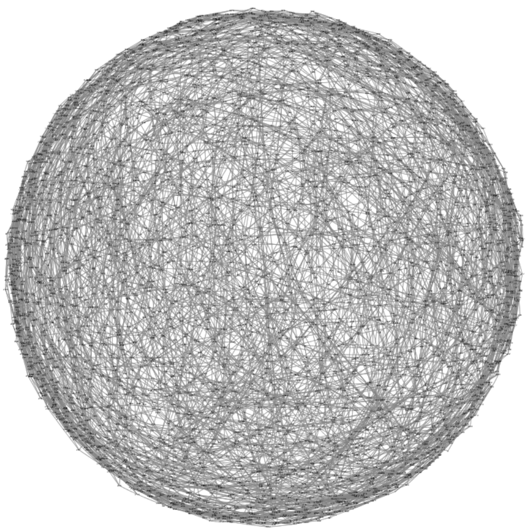
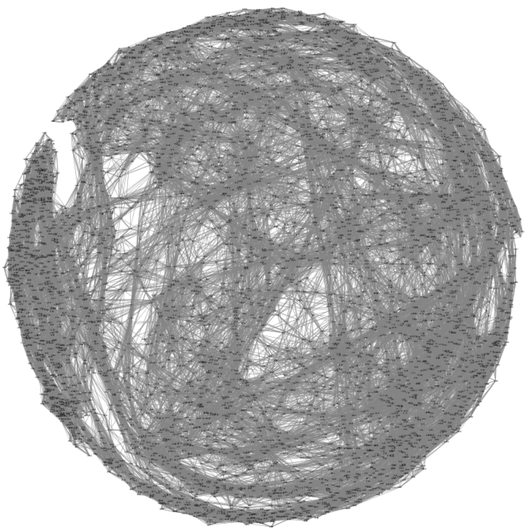


Figure 2.4: Example of initial graph with 5.000 peers

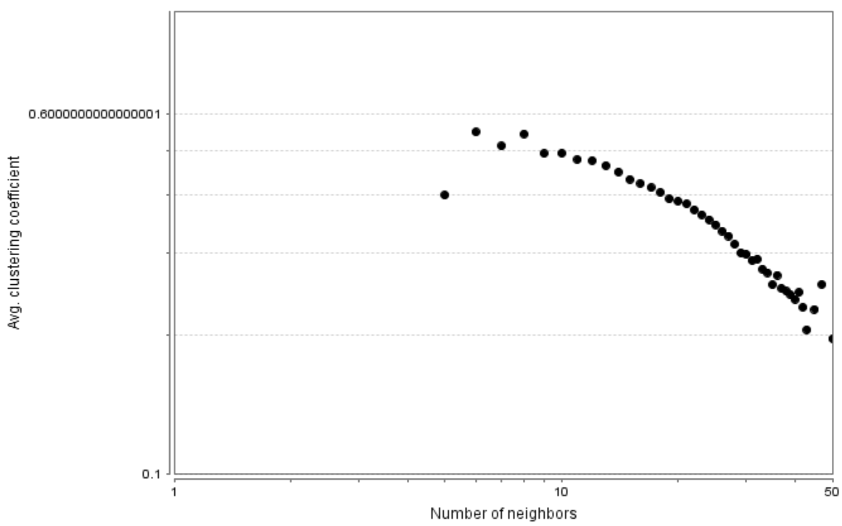
### Sparse case: 20 maximum neighbours

The output of the algorithm, in this case, is the network below:



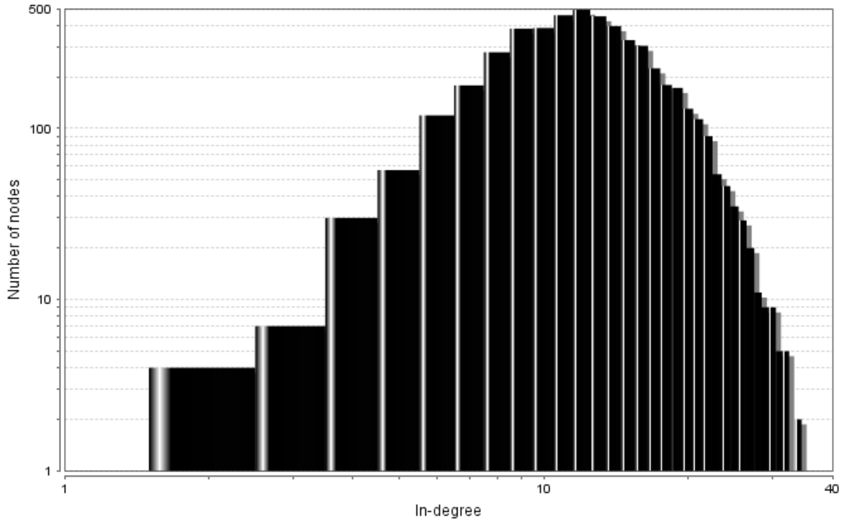
The analysis on the result have shown that the clustering coefficient is 0.397; the resulting graph is a single connected component of 61 peers of diameter. Other statistics have shown that:

- the average clustering coefficient can be shown as the following:



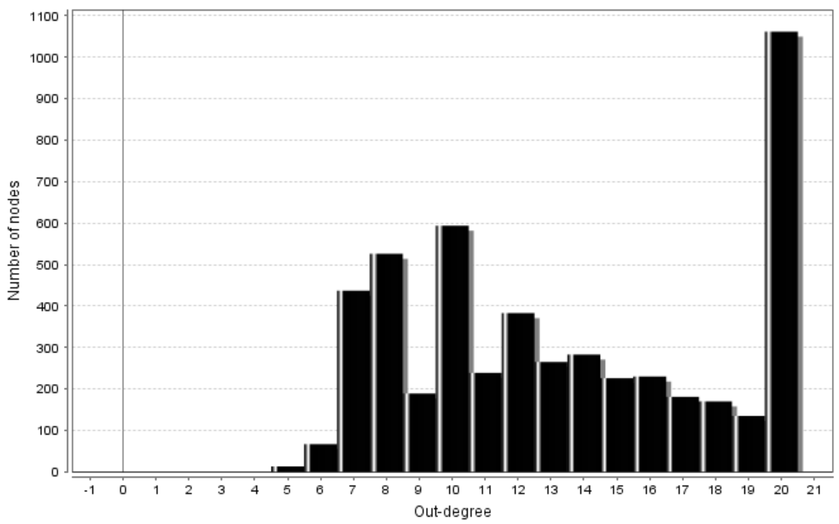

It shows us that the maximum average clustering coefficient is of 0.55 (obtained with a number of neighbours equal to 5) and it decreases until it reaches the minimum value, that is, 0.20 (when the number of neighbours is 50);

- the in-degree distribution, shown as follows:



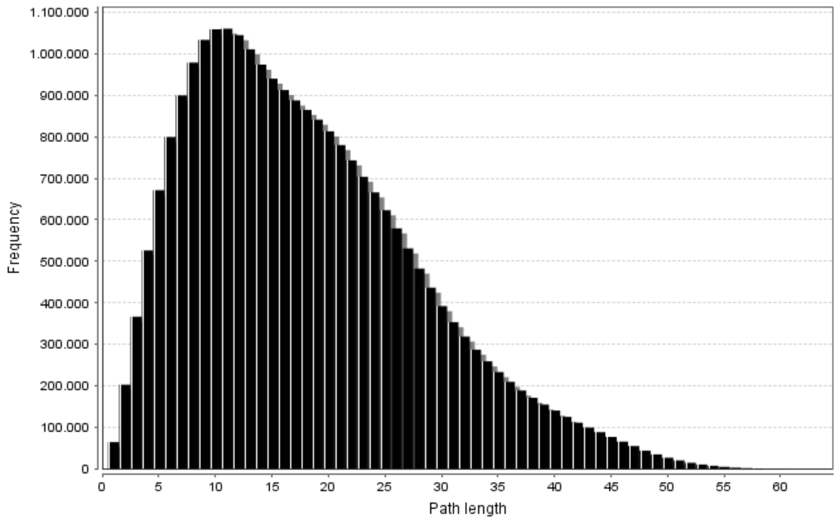

allows us to analyse the most common in-degree value in the nodes; this value is in the range $9 \sim 14$, with a maximum frequency of 490;

- the out-degree distribution, shown in the following:




that illustrates that the most common out-degree value is the 20 value, since it is the value of the cache lines involved;

- the shortest path length distribution, as illustrated below:



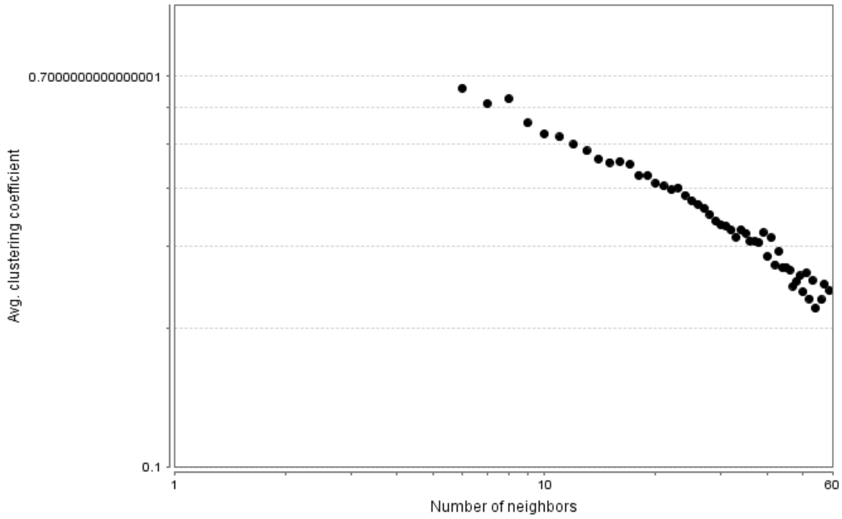

that can reach a path of length 58 (with a frequency near to 0); the most common path length are the ones with values in the range $4 \sim 27$.

**Middle case: 35 maximum neighbours**

The other situation taken into account is the following one, in which the number of lines in the cache is equal to 35 and the resulting graph after the execution of the algorithm is the following one:

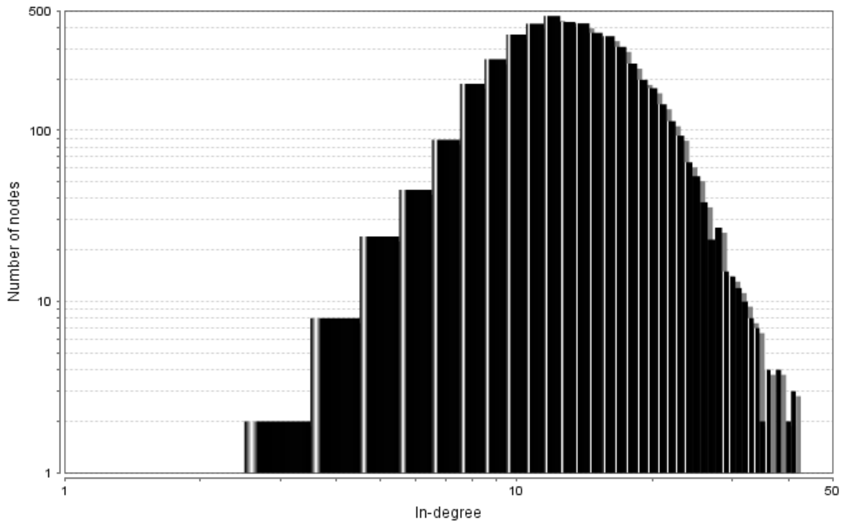
The analysis of the statistics of this graph has shown a clustering coefficient equal to 0.414, for a single connected component with a network diameter of 61 peers. The other statistics are presented below:

- the average clustering coefficient, as can be seen by the next picture:



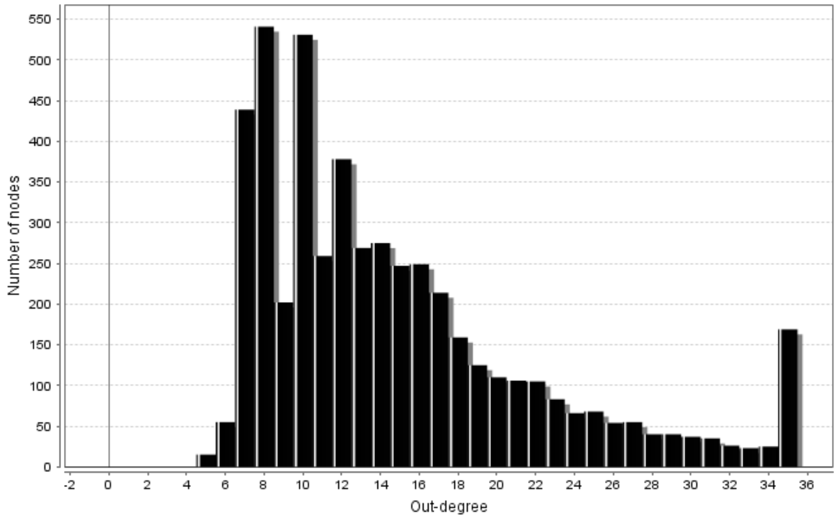

is quite high for the nodes that have a $6 \sim 8$ neighbours, with the highest value in 6 with a corresponding clustering coefficient of 0.67; as the number of neighbours increases, the clustering coefficient decreases up to arriving to the minimum for a number of neighbours of $55 \sim 57$, with a corresponding coefficient of about 0.22;

- the in-degree distribution that is illustrated in the following:



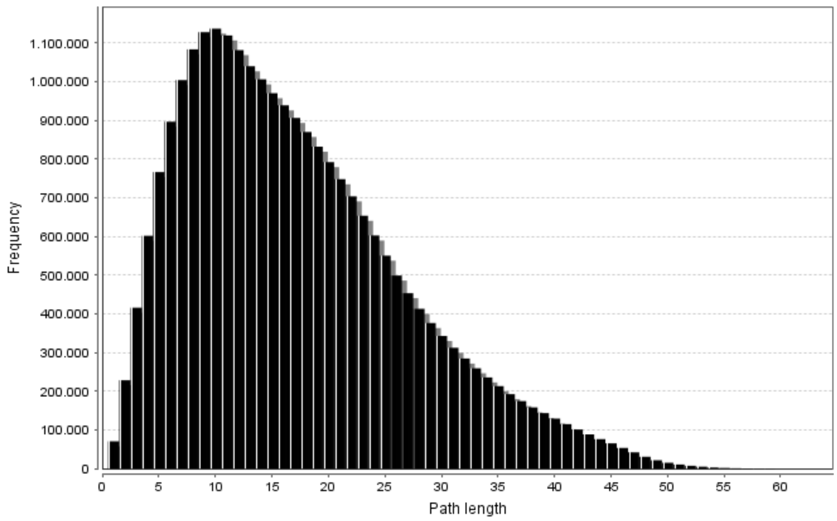

that shows us that the most common number of incoming edges for the nodes of the network is between the range: $[9, 18]$, where the most common is 12, value hold by about 460 peers; from the picture above, we can see that there exist also very few nodes with an high number of incoming edges, and so well linked by the other peers, with a maximum in-degree value of 42;

- the out-degree distribution, as shown below:




that allows us to observe that the out-degree value equal to 35 is not so common and this means that most of the nodes have a not full cache at their disposal, while instead the most common values are in the interval: $[7, 17]$ with an overall number of nodes of about 2.340;

- the shortest path length distribution, as illustrated below:



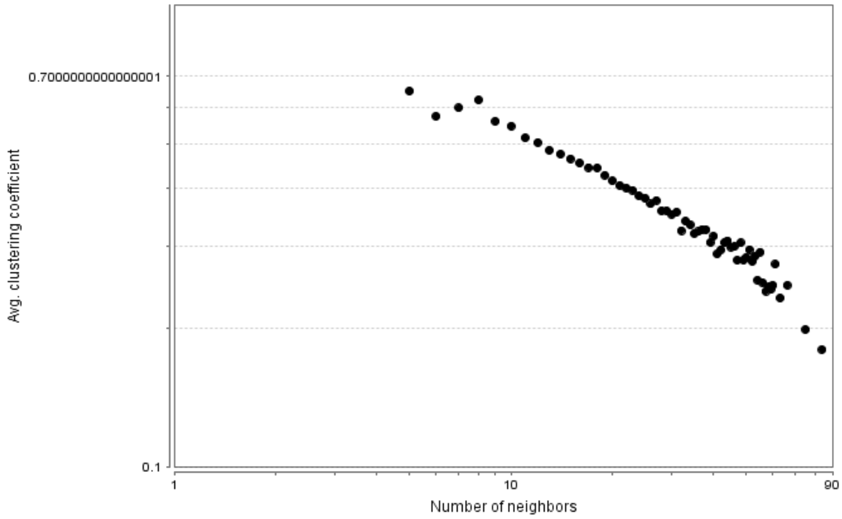

that allows us to see that it reaches the highest peak on the path of length 10, where the frequency is about 1.110.000 and that the frequency gets lower when the length of the path is above the value 10, otherwise, it increases. The maximum shortest path length that can be reached is 55 peers long.

**Dense case: 50 maximum neighbours**

For the following experiment, the graph used for the statistics analysis is not shown, since the rendering of the network is not significant.

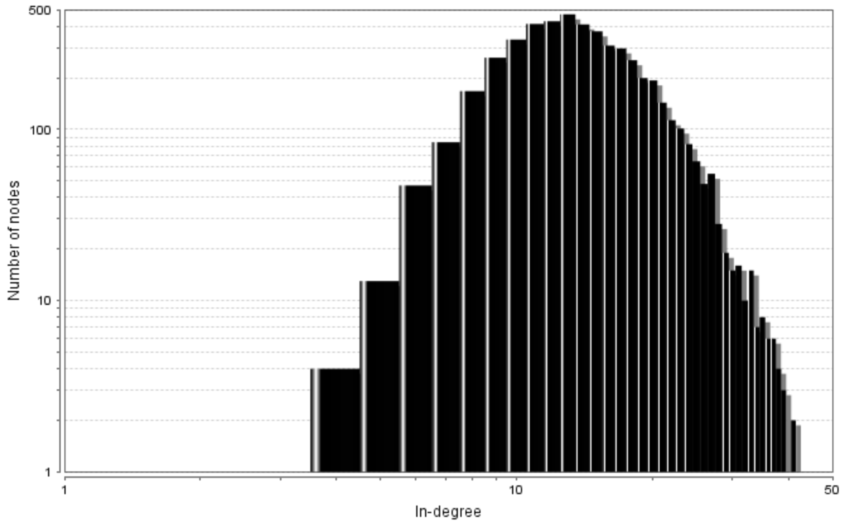
The analysis of the network above has shown that the clustering coefficient of the graph is the value 0.417 and the graph is a single connected component with a diameter equal to 58. The other statistics for this graph have shown that:

- the average clustering coefficient, presented below:



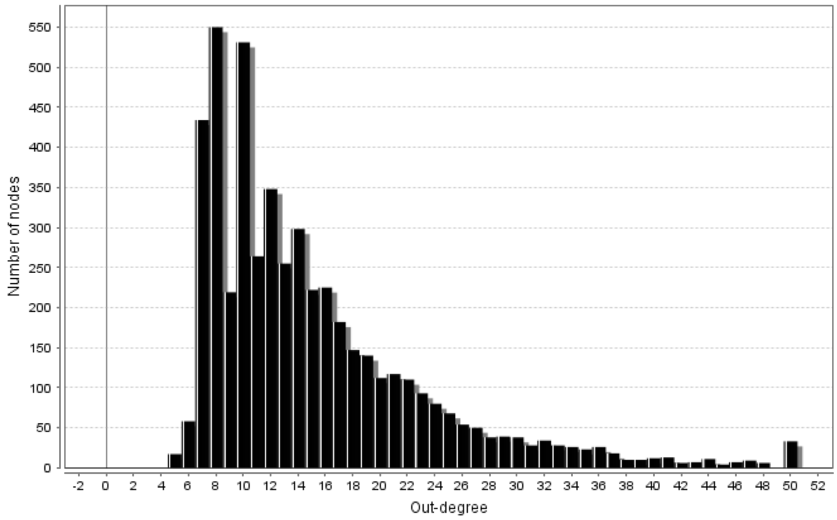

  that shows us the behaviour similar to all the other already presented cases, in which the highest value is for 5 neighbours and is equal to 0.65, and it decreases until it reaches the value $0.17 \sim 0.18$ for the corresponding value $85 \sim 87$ of the number of neighbours;

- the in-degree distribution, shown in the following:



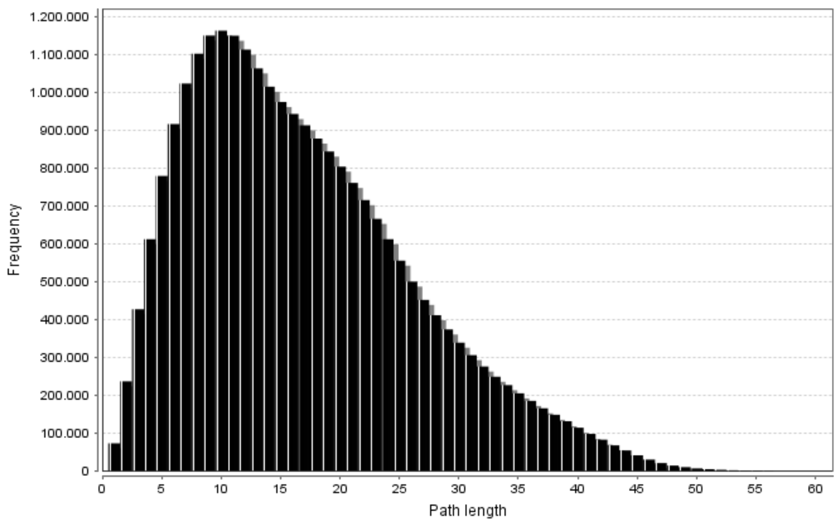

  allows us to deduce that the most common in-degree values are the ones in the range $[9, 18]$ in which the most common value is the 13, with corresponding frequency of 465;

- the out-degree distribution as well, is presented by the picture below.




From the picture above, we can see that the highest out-degree value equals to 50 is hold by very few number of nodes, while instead the most common value belong to the range $[7, 10]$, with an highest pick of out-degree for the value 8, and a corresponding number of nodes equal to 550;

- the shortest path length distribution, as follows:




that has the same behaviour of the distribution of the shortest path length seen earlier for the previous cases, in which the highest peak is obtained for the value 10 that has the corresponding frequency of about 1.160.000.

### 2.2.4 Large number of participants: 75x100 grid network

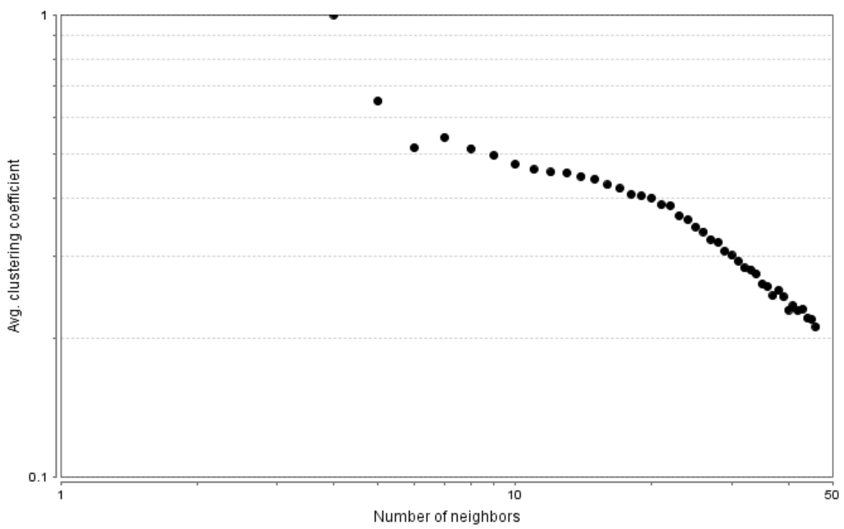
For this last experiments taken forward in this case, the network used as input of the algorithm, by using 20, 35 and 50 cache lines, has not been shown, since the rendering procedure does not achieve a significant graphical result.

#### Sparse case: 20 maximum neighbours

The output of the algorithm, in this case, is not shown since the rendering of the graph is not significant.

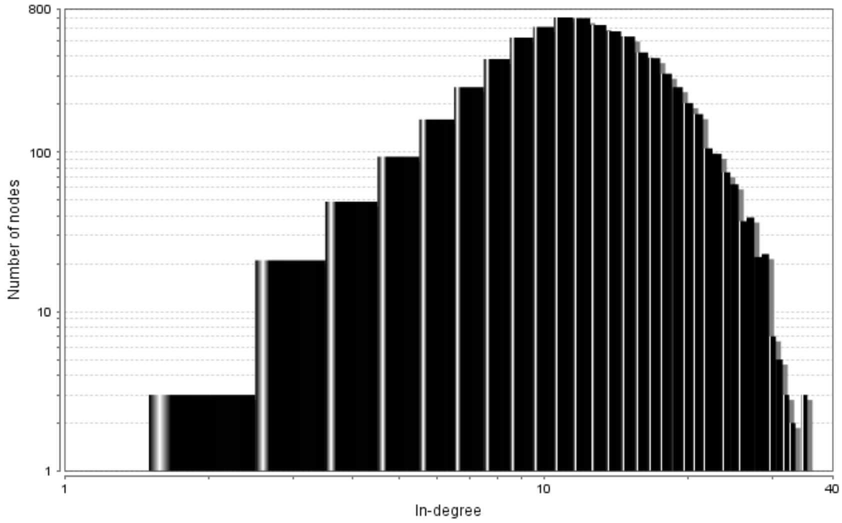
The analysis of the result have shown that the clustering coefficient is 0.396 and that the resulting graph is a single connected component with a diameter of 72 peers. Other statistics have shown that:

- the average clustering coefficient can be shown in the following:



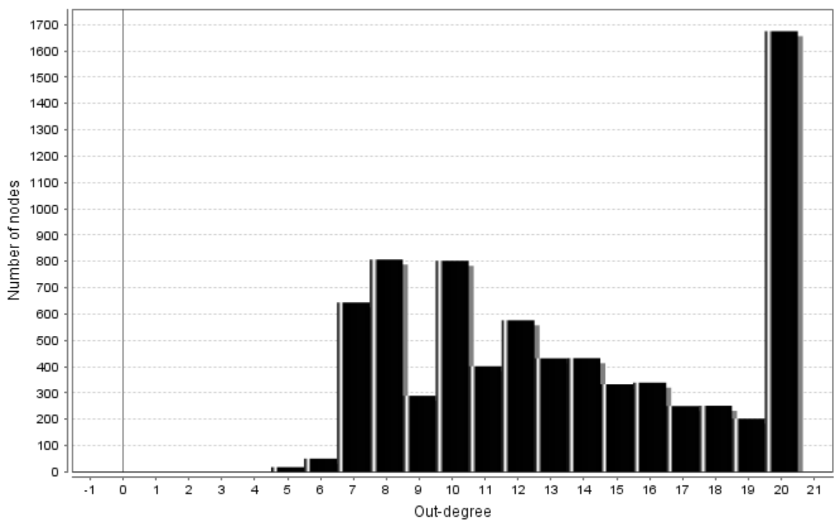

  allows us to see that the behaviour of the other already shown cases is marked in this one, since the highest clustering coefficient value is equal to 1, when there are 4 neighbours, after this, as the number of neighbours increases, the clustering coefficient decreases up to the value just above 0.20;

- the in-degree distribution, shown as follows:



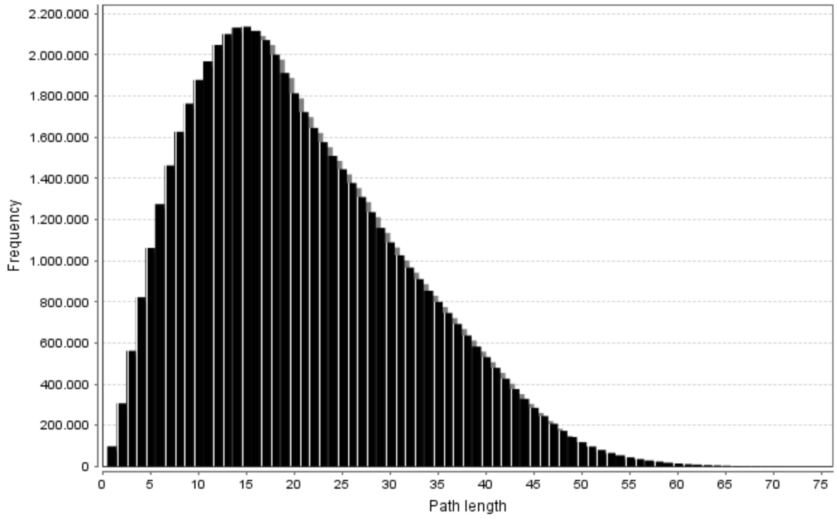

that let us denote that the most common in-degree values is in the range $8 \sim 17$, with the most common in-degree value of 11 where the corresponding number of nodes is near to 700;

- the out-degree distribution, in the following:




in this case, the most common out-degree distribution is the value 20, that is much nodes have a full cache at their disposal, anyway, there are many others that have a non full cache with only 7, 8 or 10 entries; the number of nodes that have a full cache is equal to about 1.680, while the overall of the nodes with just 7, 8 or 10 entries is equal to 2.250, this means that the full cache nodes are not so much if compared to the other ones;

- the shortest path length distribution, illustrated below:



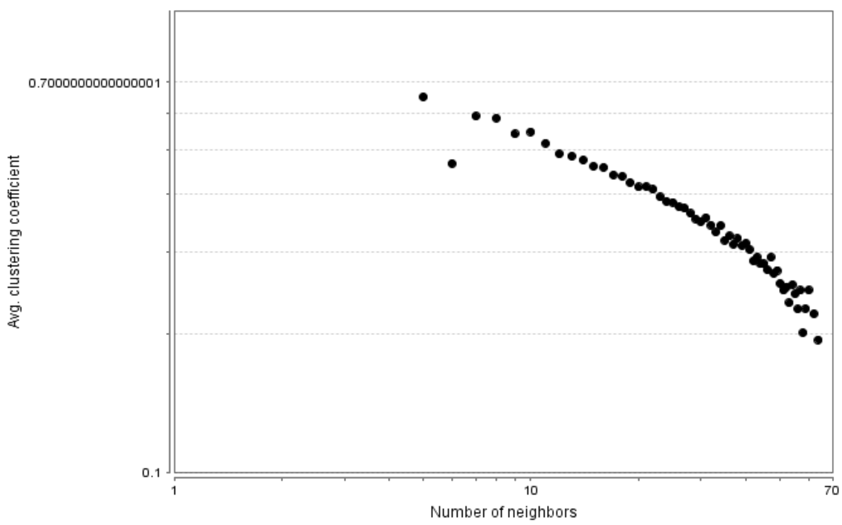

allows us to deduce that the most common path length ranges from 5 to 30, since the frequency of these two extremes exceed 1.000.000; the longest path arrives to 65 peers and the most frequent one is the one of length 15.

### Middle case: 35 maximum neighbours

The next situation taken into account is the one with 35 lines of cache for each peer, in which the resulting graph after the execution of the algorithm is not shown since it is not significant.

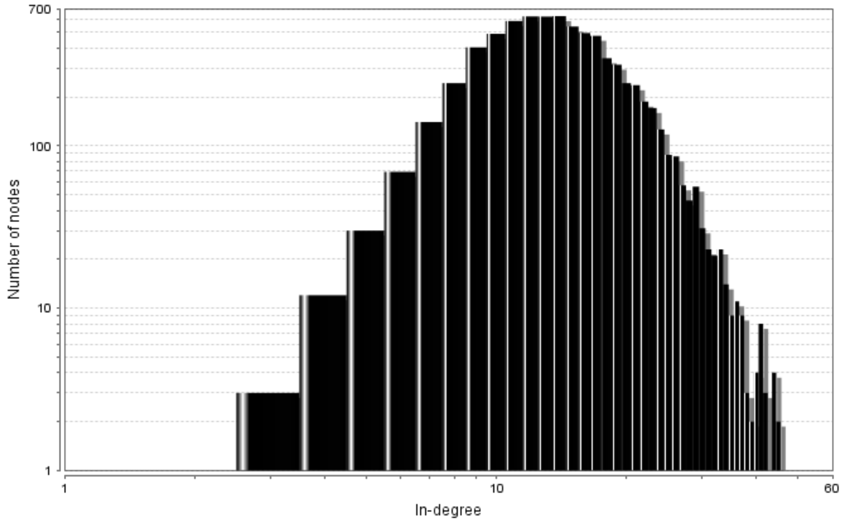
The analysis of the statistics of this graph has shown a clustering coefficient equal to 0.412, for a single connected component with a network diameter of 70 peers. The other statistics are presented below:

- the average clustering coefficient is shown in the following picture:



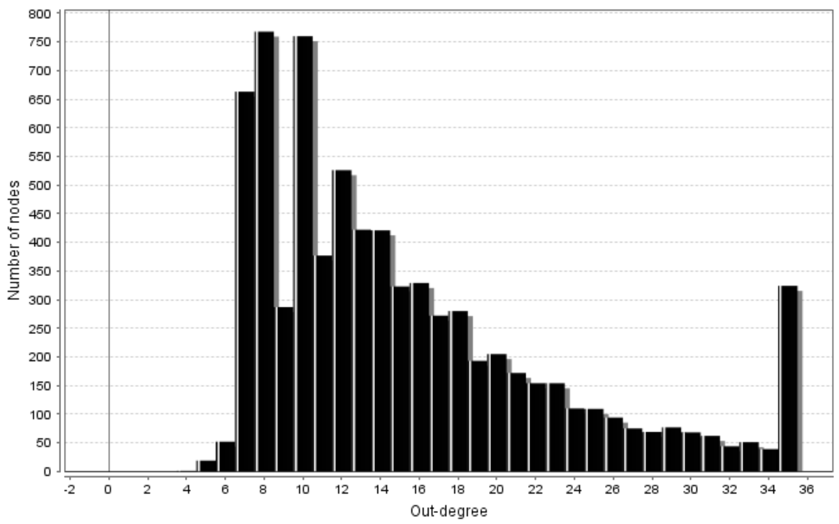

from which, we can see that we obtain the maximum value of 0.65 when the number of neighbours is equal to 5 and after this, all the other values decrease up to arriving to the minimum one of $\sim 0.19$ with the number of neighbours equal to about $64 \sim 66$;

- the in-degree distribution is illustrated in the next picture:



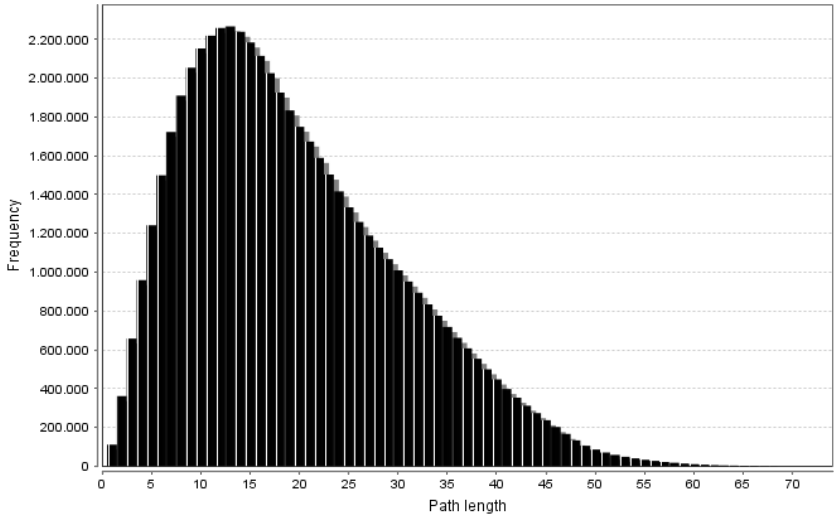

where we can see that the most common in-degree values ranges in the interval $[9, 17]$, indeed, the overall number of nodes that have an in-degree value in that interval are about 4.830; the least common in-degree values ranges from 38 to 54; this means that there exist very few nodes that compare often in the list of neighbours of some other peers;

- the out-degree distribution is illustrated in the following:




and from this, we see that having a cache totally full is a property hold by few nodes, if we compare this value, that is $\sim 325$, with the out-degree values equal to 7, 8 or 10, since the overall number of nodes that have a cache of 7, 8 or 10 entries is about 2.180;

- the shortest path length distribution is shown as follows:



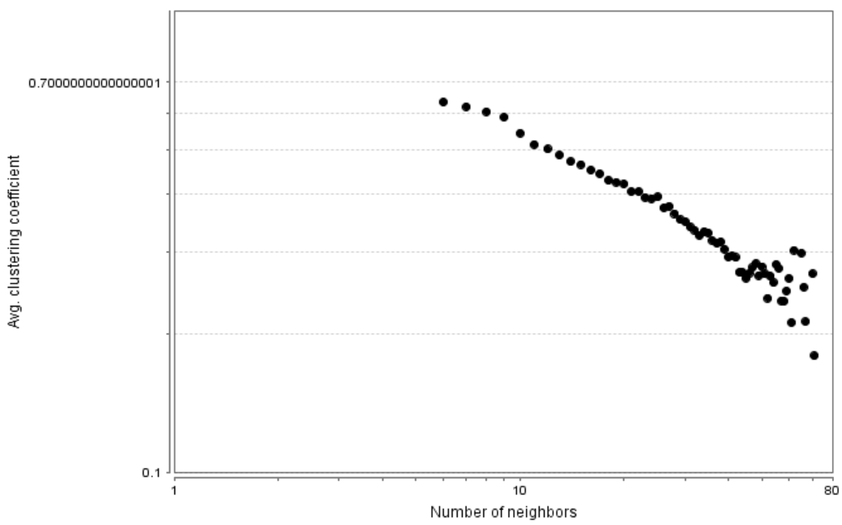

from the picture above, we can deduce that the most frequent paths are the ones with length that ranges in the interval $[5, 26]$, in which the highest frequency is given when the length is equal to 13; the longest path is the one with length equal to 64.

**Dense case: 50 maximum neighbours**

The output of the algorithm, in this case, is not shown, since the rendering of the network does not achieve a good graphical result.

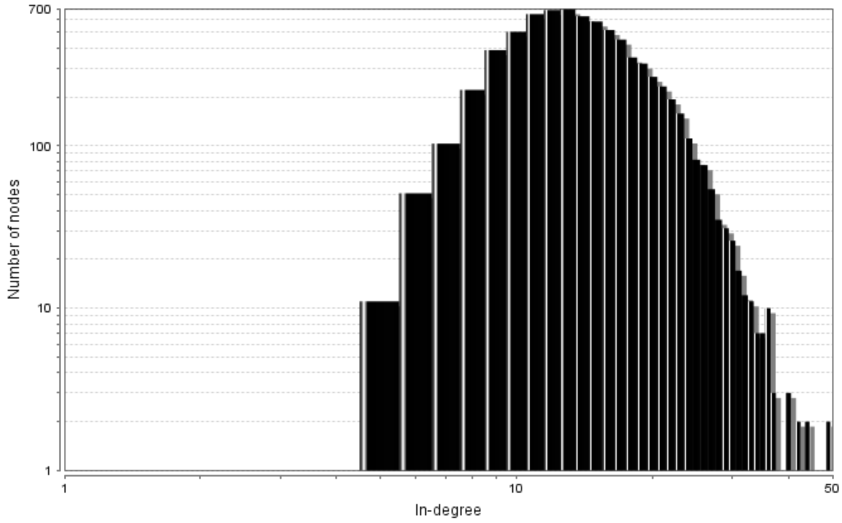
The analysis of the result have shown that the clustering coefficient is 0.415 and that the resulting graph is a single connected component with a diameter of 67 peers. Other statistics have shown that:

- the average clustering coefficient is the one in the following:



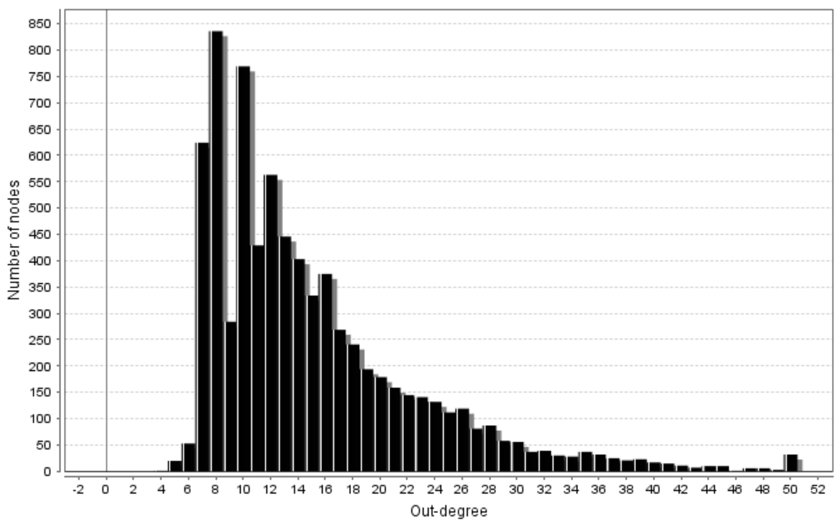

from which it shows that we have in this case a result where the maximum value of beneath 0.65 and the value decreases as the number of neighbours increase, as in everything else, up to arriving to the minimum value equal to 0.18, for the corresponding value of neighbours equal to 72;

- the in-degree distribution is presented below:



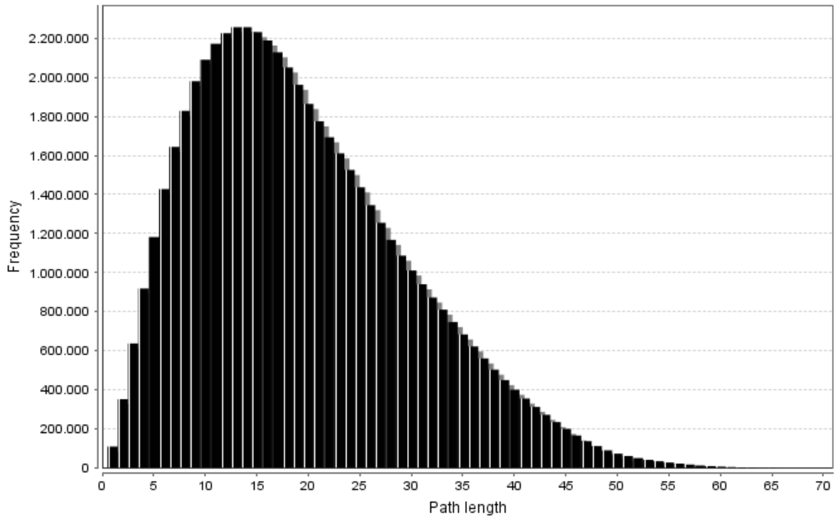

that shows us that the most common in-degree values are the ones that belong to the range $[9, 18]$, where for the extremes of this interval, the corresponding number of nodes are respectively, 390 for the left and 350 for the right boundaries; the most common in-degree value is 13, reached by almost 700 nodes and there exist nodes that have a pretty high in-degree value, with the highest one equal to 49, with a frequency near to 0;

- the out-degree distribution is shown as follows:




from this we can see that most of the nodes have a cache with less than half saved elements, with the most common values ranging in the interval $[7, 12]$, where the most common one is equal to 8 which is held by $830 \sim 840$ peers. The overall number of nodes that are in the specified range is about 3.500;

- the shortest path length distribution is shown in the following:




  it allows us to deduce that the length of a path becomes less and less frequent as its length increases. The most common lengths of the paths belong to the interval between 5 and 28, where the corresponding frequencies are respectively, 1.180.000 for the left and 1.150.000 for the right boundaries of the range; the two most common path lengths are 13 and 14 with a frequency of about 2.250.000 both.

## 2.3 Conclusions

All the results on the free scale networks and all the grid networks that have been analysed, have shown that:

- for the free scale networks, the graphs built have a low clustering coefficient, taking into account the size of the network. This kind of graphs follows the properties defined in the *Barabási-Albert Model*, since in the graphs that follow this kind of model there are highly clustered hubs that behaves like connectors for the nodes with low degree; more specifically, the average clustering coefficient is not so high because there are highly clustered hubs and many low-degree nodes. The network diameter is usually low because, for this kind of networks, it usually grows as the logarithm of the network's size divided by the logarithm of the logarithm of the network's size;

- for the grid networks, since they're built by using a regular grid, they present a high aggregation; besides, they present an high clustering coefficient and since the nodes should be linked cluster by cluster, they should present an high average distance between two nodes and we've seen in the analysis of the results that this property is held; moreover, in this kind of graphs, we've obtained networks with high clustering coefficient together with a common behaviour of the average path length that quite frequently rapidly decreases. In this kind of graphs, most of the properties of the *Watts–Strogatz Model* are held.