OLGA VALLS

COMPLEX NETWORKS

LAB03: Community Detection

The three **Community Detection Algorithms** used in this work are the following:

- Algorithm 1: Clauset-Newman-Moore Greedy modularity maximization
  ```
  c = list(greedy_modularity_communities(G))
  ```
  It begins with each node in its own community and joins the pair of communities that most increases modularity until no such pair exists.
  Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules.
  Modularity is often used in optimization methods for detecting community structure in networks.

- Algorithm 2: Fluid Communities
  ```
  c = list(asyn_fluidc(G, 4, max_iter=100))
  ```
  This algorithm is based on the simple idea of fluids interacting in an environment, expanding and pushing each other. It's initialization is random, so found communities may vary on different executions.
  It's able to find communities in synthetic graphs with an accuracy close to the current best alternatives, and it's the first propagation-based algorithm capable of identifying a variable number of communities in network.
  A k = 4 number of communities to be found, and a number of 100 iterations has been used.

- Algorithm 3: Louvain Method
  ```
  C = G.community_multilevel()
  ```
  Optimizing modularity theoretically results in the best possible grouping of the nodes of a given network, however going through all possible iterations of the nodes into groups is impractical so heuristic algorithms are used.
  In this Method, first small communities are found by optimizing modularity locally on all nodes, then each small community is grouped into one node and the first step is repeated.
  The method is similar to Clauset, Newman and Moore Algorithm that connects communities whose amalgamation produces the largest increase in modularity.

For the first two ones, the Python library **NetworkX** has been used. For the last one **iGraph** has been used.

To plot the Networks, **Kamada-Kawai** function has been used, except for the airports_UW, where **spring** layout have been used.

The following **measures** have been computed for the Networks with reference partitions compared to the Communities created by the different algorithms. **Radatools** software has been used.

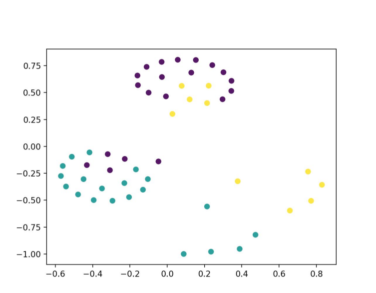
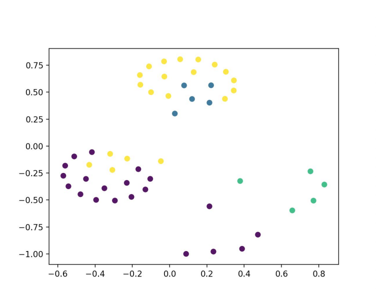
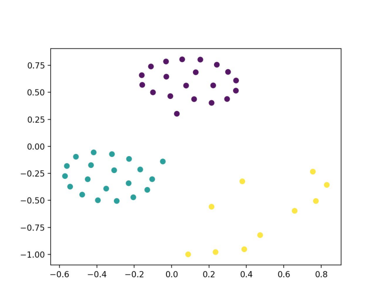
- Jaccard Distance: 1 - size of the intersection divided by the size of the union of the sample sets.
- Normalized Mutual Information Index: mutual dependence between the two variables. It quantifies the "amount of information" obtained about one random variable through observing the other random variable.
- Normalized Variation of Information Metric: distance between two clusters. It is a true metric as it obeys the triangle inequality.
- Modularity has been computed for all the partitions. It is a scale value between -1 and 1 that measures the density of edges inside communities to edges outside communities.

The **code** for the calculation of the partitions, the computing of the modularity for each algorithm, and the plotting of the different algorithms applied to each Network can be found in the `CN_Lab03_OlgaValls.py` file.

Networks

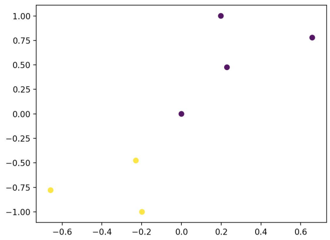
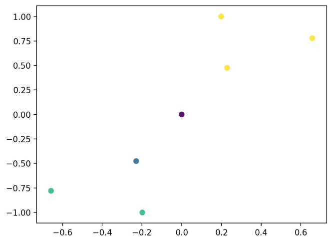
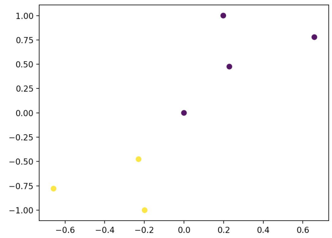
TOY NETWORKS

20x2+5x2

Clauset-N-M Community	Fluid Community	Louvain Community
		

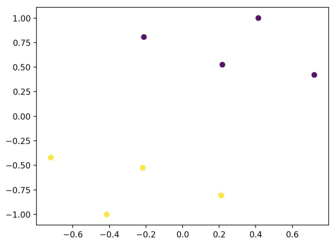
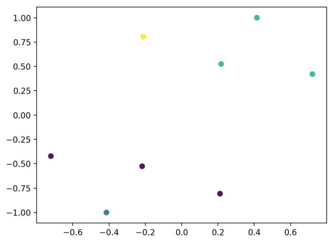
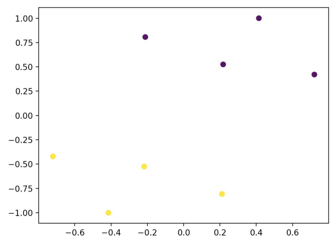
MEASURES	Clauset	Fluid	Louvain
Jaccard Index	0.4348	0.3745	**0.9412**
Normalized Mutual Information Index (arithmetic)	0.5382	0.4967	**0.9383**
Normalized Variation Of Information Metric	0.2654	0.3226	**0.0354**
Modularity	**0.5425785462209646**	0.4551728997157167	**0.5425785462209587**

graph3+1+3

Clauset-N-M Community	Fluid Community	Louvain Community
		

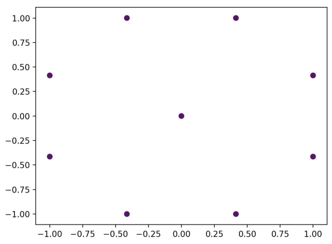
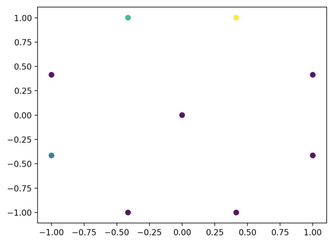
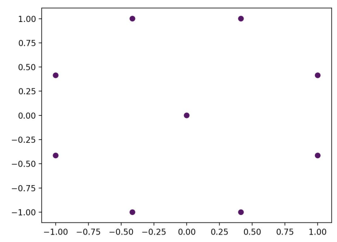
MEASURES	Clauset	Fluid	Louvain
Jaccard Index	**1.0000**	0.3333	**1.0000**
Normalized Mutual Information Index (arithmetic)	**1.0000**	0.6713	**1.0000**
Normalized Variation Of Information Metric	**0.0000**	0.3437	0.0000
Modularity	**0.3671875**	0.1015625	**0.3671875**

graph4+4

Clauset-N-M Community	Fluid Community	Louvain Community
		

MEASURES	Clauset	Fluid	Louvain
Jaccard Index	**1.0000**	0.3333	**1.0000**
Normalized Mutual Information Index (arithmetic)	**1.0000**	0.6667	**1.0000**
Normalized Variation Of Information Metric	**0.0000**	0.3333	**0.0000**
Modularity	**0.4428571428571428**	0.2346938775510204	0.42307692307692313

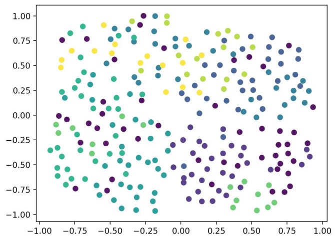
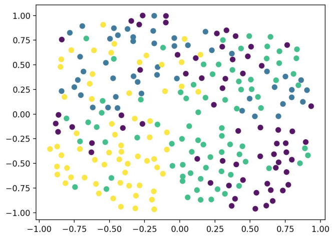
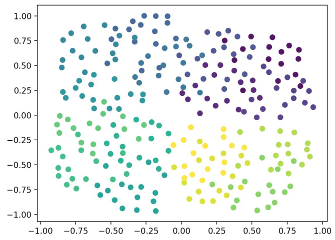
star

Clauset-N-M Community	Fluid Community	Louvain Community
		

MEASURES	Clauset	Fluid	Louvain
Jaccard Index	**1.0000**	0.4167	**1.0000**
Normalized Mutual Information Index (arithmetic)	**1.0000**	0.0000	**1.0000**
Normalized Variation Of Information Metric	**0.0000**	0.4564	**0.0000**
Modularity	**0.0**	-0.046875	**0.0**

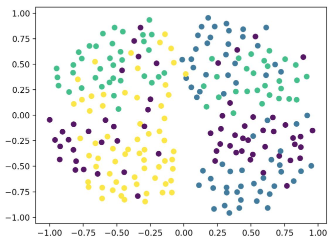
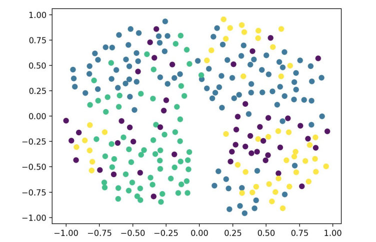
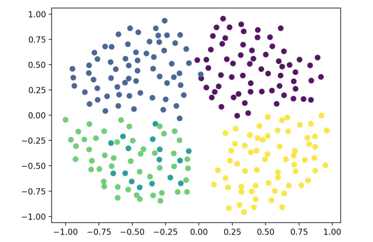
MODEL NETWORKS

256_4_4_2_15_18_p

Clauset-N-M Community	Fluid Community	Louvain Community
		

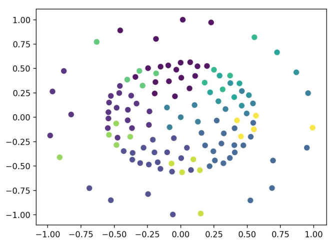
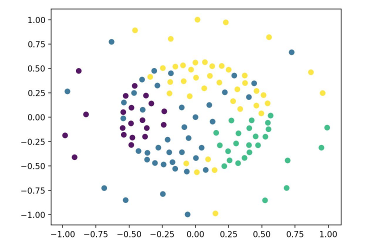
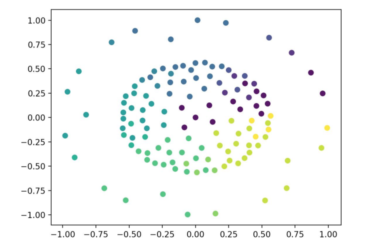
MEASURES	Clauset	Fluid	Louvain
Jaccard Index	0.2498	0.1284	**1.0000**
Normalized Mutual Information Index (arithmetic)	0.6233	0.3849	**1.0000**
Normalized Variation Of Information Metric	0.3332	0.4523	**0.0000**
Modularity	0.7656597969165392	0.6305960306906689	**0.7818042125081897**

256_4_4_4_13_18_p

Clauset-N-M Community	Fluid Community	Louvain Community
		

MEASURES	Clauset	Fluid	Louvain
Jaccard Index	0.3080	0.2716	**0.9048**
Normalized Mutual Information Index (arithmetic)	0.4240	0.3128	**0.9517**
Normalized Variation Of Information Metric	0.2880	0.3381	**0.0254**
Modularity	0.6967733384846492	0.619992693080448	**0.6974188902601712**

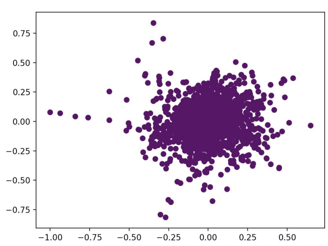
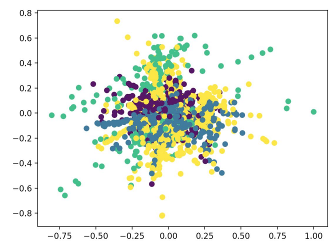
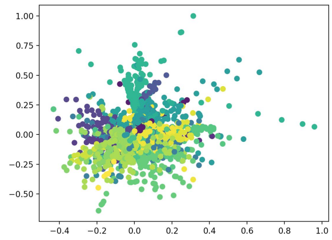
rb125

Clauset-N-M Community	Fluid Community	Louvain Community
		

MEASURES *(with rb125-1 partition)*	Clauset	Fluid	Louvain
Jaccard Index	0.5833	0.2667	**0.6167**
Normalized Mutual Information Index (arithmetic)	0.8284	0.4721	**0.8450**
Normalized Variation Of Information Metric	0.1381	0.2590	**0.1223**
Modularity	0.6087328970850735	0.4631052944675827	**0.6168970001542905**

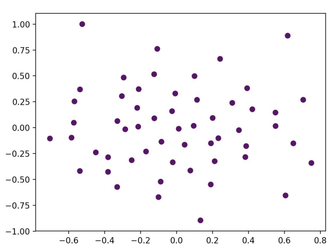
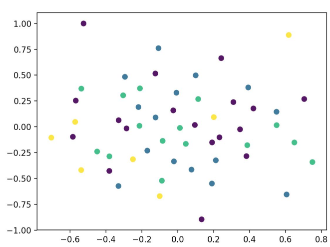
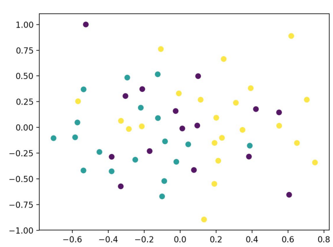
REAL NETWORKS

airports_UW

Clauset-N-M Community	Fluid Community	Louvain Community
		

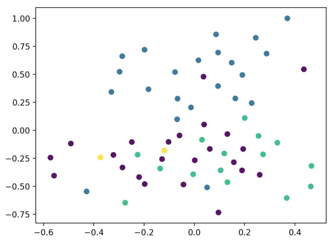
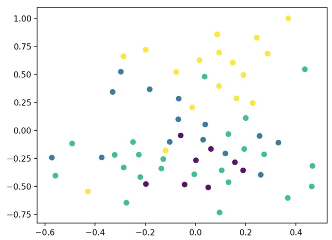
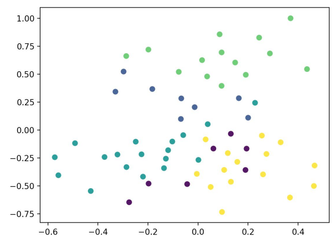
MEASURES	Clauset	Fluid	Louvain
Modularity	-8.355951990733535e-17	0.301141049976736	**0.6919049657372429**

cat_cortex_sim

Clauset-N-M Community	Fluid Community	Louvain Community
		

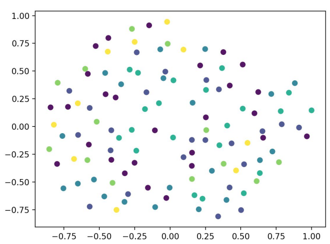
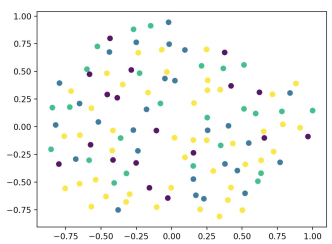
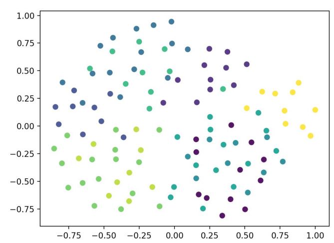
MEASURES	Clauset	Fluid	Louvain
Jaccard Index	0.2572	0.1685	**0.5894**
Normalized Mutual Information Index (arithmetic)	0.0000	0.1429	**0.6726**
Normalized Variation Of Information Metric	0.3341	0.5738	**0.1981**
Modularity	-3.545599775302562e-17	**0.3384118872407228**	0.266097341683014

dolphins

Clauset-N-M Community	Fluid Community	Louvain Community
		

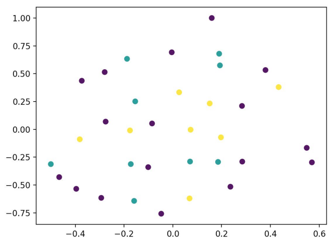
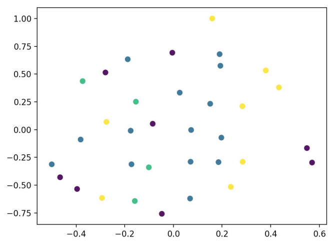
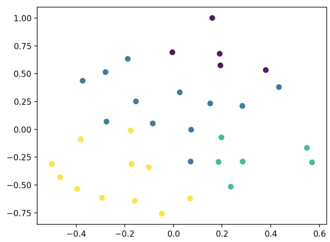
MEASURES	Clauset	Fluid	Louvain
Jaccard Index	0.4086	**0.4903**	0.3631
Normalized Mutual Information Index (arithmetic)	0.3860	0.2878	**0.5109**
Normalized Variation Of Information Metric	0.2705	0.2975	**0.2581**
Modularity	0.49549068470392743	0.43588070092163944	**0.5185317036509631**

football

Clauset-N-M Community	Fluid Community	Louvain Community
		

MEASURES	Clauset	Fluid	Louvain
Jaccard Index	0.0530	0.0682	**0.7004**
Normalized Mutual Information Index (arithmetic)	0.1199	0.1052	**0.8903**
Normalized Variation Of Information Metric	0.7788	0.7223	**0.1095**
Modularity	0.549740665142672	0.4959043987130234	**0.6045695626834572**

zachary_unwh

Clauset-N-M Community	Fluid Community	Louvain Community
		

MEASURES	Clauset	Fluid	Louvain
Jaccard Index	0.2415	0.2051	**0.4754**
Normalized Mutual Information Index (arithmetic)	0.0009	0.0194	**0.5866**
Normalized Variation Of Information Metric	0.4902	0.5534	**0.2359**
Modularity	0.3806706114398416	0.33966140696909886	**0.41880341880341887**

Conclusions

As seen in the previous tables, for most of the Networks, the Louvain Method is the algorithm that performs the best clustering.

The plots show a clear color separation for each of the communities for this algorithm, while for the other two the colors are more mixed (mixed that some nodes are considered from one community while they should belong to another community).

The values for the modularity are also strong for the Clauset-Newman-Moore algorithm, which indicates the relation between the density of edges inside communities to edges outside communities.

Needed to say that for the Fluid algorithm, the number of communities is fixed a priori, and this same k is used for all the Networks of this work. Modifying the k and the number of iterations that this algorithm perform, for a specific Network could result in a better performance than Louvain Method, although some tests should be performed.