

---

## Homework3 : Clustering Analysis for Complex Networks

---

|             |          |
|-------------|----------|
| Yuze Liu    | 50207903 |
| Luting Chen | 50133507 |
| Vicky Zheng | 50037709 |

11/29/2016

### 1 INTRODUCTION OF MCL ALGORITHM

MCL Algorithm is The Markov Cluster Algorithm. It is a graph clustering algorithm. In the graph, each vertex is connected to others by weighted or unweighted edges. MCL is based on Random Walks, start at a arbitrary node and then randomly travel to connected node, it will be more likely to stay within a cluster than travel between them. So by doing random walks upon the graph, it is possible to discover the clusters in the graph.

### 2 IMPLEMENTATION

The MCL I implemented is based on the following algorithm which is provided in the class:

1. Input is an un-directed graph, power parameter  $k$ , and inflation parameter  $r$ .
2. Create the associated matrix.
3. Add self loops to each matrix.
4. Normalize the matrix.
5. Expand by taking the  $k^{th}$  power of the matrix.
6. Inflate by taking inflation of the resulting matrix with parameter  $r$ .
7. Repeat steps 5 and 6 until a steady state is reached.

### 8. Interpret resulting matrix to discover clusters.

In my implementation, I used python 2.7, sklearn packag has been used to normalize the matrix. The iterations is set as 100 times, which turns out that for the given data, after 100 times iteration, the steady state will always be reached, and it won't take more than 5 seconds.

In order to show the best result, the Pajek is used to visualize the clusters. The clusters are dependent on our power parameter (k) and inflation parameter (r). As we increased k, the number of clusters decreased whereas when we increased r, the number of clusters increased.

## 3 VISUALIZATION OF THE 3 GIVEN DATASET

Three datasets were given for testing our implementation of MCL Algorithm. After tuning our parameters, the most reasonable result and the related parameters are shown below.

Dataset1 attweb\_net :

k = 3, r = 2, clusters : 6

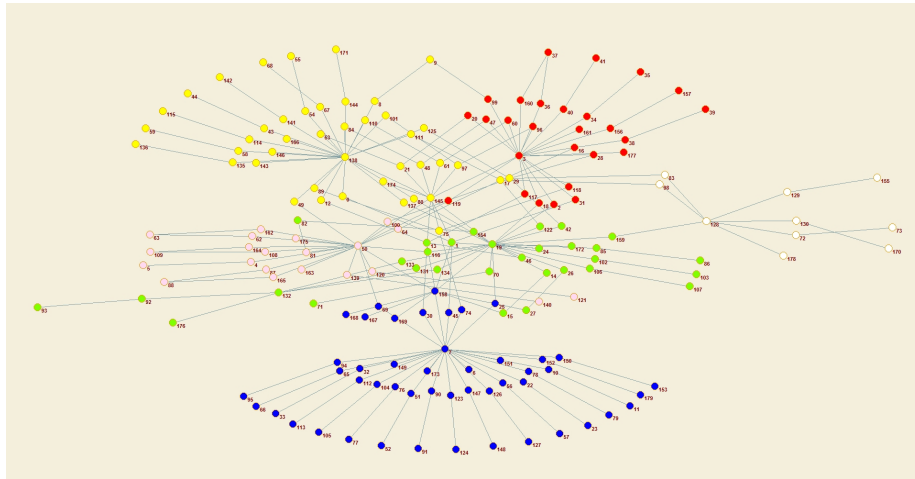


Figure 3.1: Result of attweb\_net graph 2D Version,k = 3, r = 2, clusters = 6

Dataset2 physics\_collaboration\_net :

k = 7, r = 7, clusters : 5

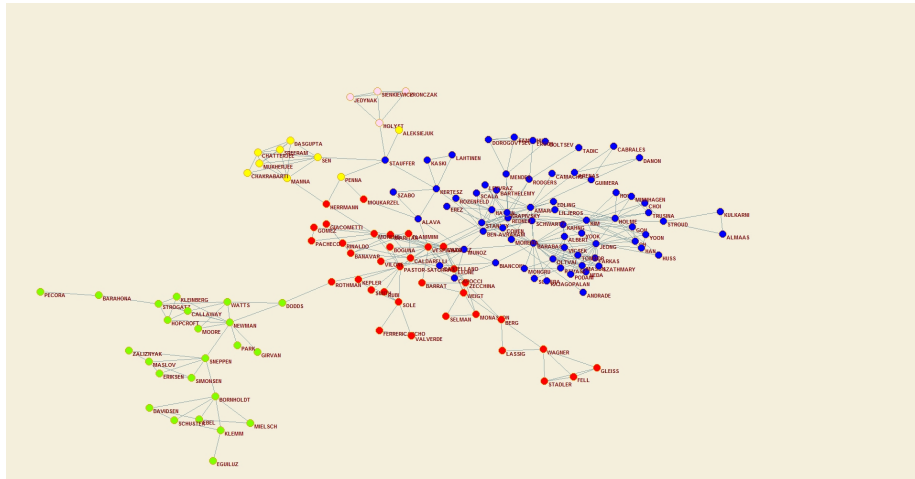


Figure 3.2: Result of physics\_collaboration\_net 2D Version,  $k = 7$ ,  $r = 7$ , clusters = 5

Dataset3 yeast\_undirected\_metabolic :  
 $k = 10$ ,  $r = 3$ , clusters : 9

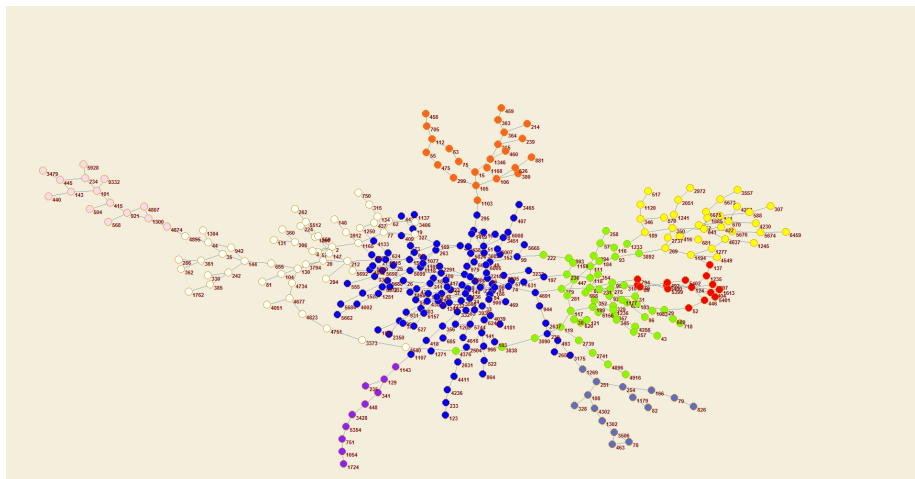


Figure 3.3: Result of yeast\_undirected\_metabolic graph 2D Version,  $k = 10$ ,  $r = 3$ , clusters = 9