Problem 2

2a Correct answers:

- It is based on a null model
- It is applicable to weighted networks

2b Modularity score assigns a number to a given partition of the network into different clusters. A partition with a higher modularity score is considered to be a better split of the network into different communities.

This score is higher when the number of connection in each cluster is more than what is expected based on a given null model. Such groups are deemed to be good communities according to modularity.

The term $(D_m/2E)^2$ comes from the fact that in this formula the configuration model has been used as a null model.

2c The notes are ordered randomly.

Instead of considering all communities only those corresponding to direct neighbors of the given node v are considered.

The new network has as its nodes the communities and there is an edge between two new nodes if there was at least one edge between two vertices in the different communities. The edges are then weighted by the total number of edges between the communities in the previous graph.

2d As research groups are defined as working on similar topics, we can expect that they collaborate a lot withing the group. This would indicate that research groups form groups of nodes that have more connection with each other than other nodes.

Modularity optimization would yield a partition where nodes are more likely to connect then based on their degrees. Which would make sense from the idea about research groups.

However, we have to be careful as the network is sparse and the number of links inside the expected communities is very small. So we have to make use of the resolution parameter because else all small groups will be arbitrarily joint into a group with approximately 1000 edges.

Problem 3

3a Correct answers:

- It always puts isolated components into their own communities
- It requires knowledge about the number of communities

3b Define the $n \times n$ matrix W as $W_{ij} = w_{ij}$ and let $s_i := \sum_{j=1}^{n} A_{ij} w_{ij}$. Then the weighted version of the Laplacian matrix is

$$L_{ij} = \begin{cases} s_i - W_{ij} & \text{if } i = j, \\ -W_{ij} & \text{else.} \end{cases}$$

3c The algorithm describe a method for solving the min-cut problem. In this problem, a graph has to be divided into k groups by cutting away the edges between different groups. The problem is then to find the partition into k groups that minimizes this so-called ratio-cut score.

It turns out that the solution to this problem can be expressed in terms of the Laplacian matrix.

However, finding the solution is still hard. But by relaxing the problem an explicit solution can be obtained by considering the first k eigenvectors of the Laplacian.

This solution then has to be cast back into the a partition into groups, which is done by the clustering algorithm in the last step. In the end, we obtain an approximate solution to the ratio-cut problem, which gives us a good community assignment as the idea is that communities have more links between members than to the outside. So cutting between communities should be cheap.

3d In the plot we see three dense collections of points separated by empty space. Since each point corresponds to a a node we would guess that these three clusters indicate the three different communities in our network.

3e From this plot we clearly see 4 separate groups of points (nodes). So then we would guess that there are 4 communities instead of 3.

The problem here was that the plots we are shown are 2-dimensional, while the data points are 4-dimensional. Hence we formally need to do clustering in 4D. To get the 2D plot we had to project the data onto 2 axis. Apparently, the once that where chosen for the previous figure did not show the separation into 4 groups.

Problem 4

4a Correct answers:

- There are only spurious links
- It is more likely that G^O differs a lot from G than a little

4b Note that $R(v,w) = \mathbb{P}(\{v,w\} \in E | G^O)$. Using the total law of probability we then write

$$\mathbb{P}(\{v,w\} \in E | G^O) = \sum_G \mathbb{P}(\{v,w\} \in E | G^O, M_n = G)\mathbb{P}(M_n = G). \tag{1}$$

Note that

$$\mathbb{P}(\{v,w\} \in E | G^O, M_n = G) = \mathbf{1}_{\{v,w\} \in E}. \tag{2}$$

In addition we have (by Bayes'rule)

$$\mathbb{P}(M_n = G | G^O) = \frac{\mathbb{P}(G^O | M_n = G)\mathbb{P}(M_n = G)}{\mathbb{P}(G^O)}. \tag{3}$$

Finally, another application of the total law of probability yields

$$\mathbb{G}^{\mathbb{O}} = \sum_{G'} \mathbb{P}(G^O | M_n = G')\mathbb{P}(M_n = G').$$

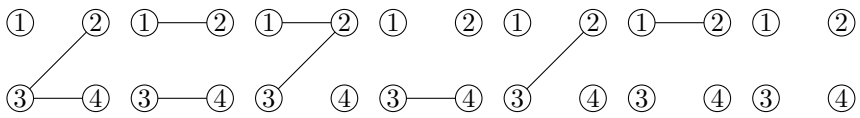
Substituting this into (3) give

$$\begin{aligned}\mathbb{P}(M_n = G | G^O) &= \frac{\mathbb{P}(G^O | M_n = G)\mathbb{P}(M_n = G')}{\sum_{G'} \mathbb{P}(G^O | M_n = G')\mathbb{P}(M_n = G')} \\ &= \frac{(|E^O| - |E|)\mathbf{1}_{G \subset G^O}}{\sum_{G'}(|E^O| - |E'|)\mathbf{1}_{G' \subset G^O}}\end{aligned}$$

And substituting this and (2) into (1) yield the required result.

4c We can remove the term $\sum_{G'}(|E^O| - |E'|)\mathbf{1}_{G' \subset G}$ since this is the same for every edge. Moreover, it only scales the reliability and hence does not influence the ordering.

4d The only graphs for which $G \subset G^O$ are given below.



So the value in the denominator is given by $1 \times 3 + 2 \times 3 + 3 \times 1 = 12$.

All the edges that are currently missing have a reliability of 0

All other edges occur twice in in graphs with two edges and once in graphs with one edge. So the reliability of any such edge is $(1 \times 2 + 2 \times 1)/12 = 1/3$.

All in all this metric does not seem very useful.

4e Correct answer: B

4f The variantion would start with a random graph and add/delete a random edge. Then it computes $(|E^O| - |E'|)\mathbf{1}_{G' \subset G^O}$ and we accept the change if this value grows, compared to the previous value. If it shrinks we only accept it with probability

$$\frac{(|E^O| - |E'|)\mathbf{1}_{G' \subset G^O}}{(|E^O| - |E^{\text{old}}|)\mathbf{1}_{G^{\text{old}} \subset G^O}}.$$

After a while we see that the value of $(|E^O| - |E|)\mathbf{1}_{G \subset G^O}$ converges. From that moment on we compute $(|E^O| - |E|)\mathbf{1}_{G \subset G^O}\mathbf{1}_{\{v,w\} \in E}$ for all edges and add it to the reliability matrix.

Problem 5

5a Correct answers:

- It is a local similarity metric
- You can use it directly as a reliability score in a link prediction problem

5b The intuition is that in social networks two people are likely connected if they are connected to many of the same people. Connected people usually have the same friends.

The denominator tries to solve the issue that vertices with high degree gave a natrual tendency to have many connections in common from the basic fact they have many connections. Thus by divifing by the total number of connections we penalize vertices with more connections.

5c Correct answer: C

5d $\Gamma(4) = \{1, 3, 8\}$, $\Gamma(6) = \{2, 3, 5, 8, 9\}$ so $S(4, 6) = 2/6 = 1/3$.

5e If the ROC curve is above the black line, then the predictor performs better than guessing, while it performs wore poorly than guessing if it is below.

A reliability metric that does follow the black line would flip a coin for each edge to decide whether it is missing.

5f AUC literally means *area under curve*. So we need to compute this area. Two estimates are:

- Les miserables: AUC ≈ 0.89
- Karate club: AUC ≈ 0.60

5g There are 2926 possible links, thus 2672 could be missing. Since we select 250 and only one is missing we have a FPR of at least $249/2672 \approx 0.093$ and at most $250/2672 \approx 0.094$.

Corresponding to these values there is a TPR of approximately 0.78. This is the probability that the one missing link is among the 250 chosen ones.