Question 1

Calculate the highest possible modularity score obtained in the karate club example with the given cluster membership. Explain the calculation.

Selected Answer: .

Correct Answer: The highest value for this network is 0.64

As only the cells in the diagonal are non-zero, a=e where e is the share of edges within a cluster or module.

a	a^2	$e-a^2$
40.7%	17%	24%
10.2%	1%	9%
6.8%	0%	6%
42.4%	18%	24%
100.0%	36%	**0.64**

In case of cluster 1, there are 48 of the 156 edges, which is 40.7%

✓

Response Feedback: [None Given]

Question 2

Explain when the highest/lowest theoretical score (-1, 1) can be obtained.

Selected Answer: .

Correct Answer: The maximum value of 1 can be obtained when the number of clusters is near infinite.
In that case a will be very small and a^2 will come close to zero. Consequently $e-a^2$ is close to e

✓ As modularity is the sum over $e-a^2$ this will be close to 1 being the sum over e

Response Feedback: [None Given]

Question 3

Try to explain the meaning of a negative modularity score

Selected Answer: .

Correct Answer: ✓
The modularity will have a negative value when there are more edges between modules than within the modules.
This would mean that $e-a^2$ is negative for most modules.

This would mean that the clustering algorithm is in some strange way sensitive to the community structure but is reacting in an opposite way to the present structure.

Response
Feedback: [None Given]

Question 1

What are the basic assumptions often underlying modularity maximization approaches discussed in this paper

Correct ✓

Answer: The authors put these three assumption forward:

1. empirical networks with modular structure tend to exhibit a clear optimal partition
   This implies that there is an unique partition that is clearly distinct from other suboptimal partitions
2. high-modularity partitions of an empirical network are structurally similar to this optimal partition
   This means that the optimal partition can be identified by comparing modularity scores within one network across different partitions
3. the estimated Qmax can be meaningfully compared across networks.
   Consequently, a general rule of thumb can be obtained indicating when a partitioning is good or close to the optimal partition

Question 2

The authors discuss some cases how cleverly designed algorithms can provide a solution to the resolution limit. Explain if and how the Louvain method proposed by Blondel fits in this discussion.

Correct ✓

Answer: Two approaches described on page 3 are relevant to the Louvain method:

1. The original procedure provides a layered agglomerative approach where the results from the first round are merged together and the reduced network is input to a next step.

2. An extended version of the Louvain method (eg as implemented in Gephi or Pajek) uses a target resolution parameter to identify the modules at a particular scale.

Question 3

What random graph model is often assumed to be underlying the calculation of modularity, especially related to the inter-module connections. Explain

Correct ✔
Answer: The calculation of modularity assumes a Erdös-Renyi-like distribution of edges among modules. This means that every edge has the same probability whatever source and target module.

This is clear from the formula used to calculate the expected share of edges within a module as the square of the sum of the row shares.

Question 4

Figure 2, 3 and 4 exhibit clear plateaus in the modularity function. Try to explain the meaning of this plateau and the consequences for obtained cluster solutions .

Correct ✔
Answer: These figures refer to the second assumption described by the authors: that a high-modularity partition of a network is structurally close to the optimal solution.

The figures indicate that this assumption does not hold. May high-modularity partitions can be obtained with each clear structural differences. The solution one obtains from a modularity based partitioning does not have to be close to this implicit optimal clustering of the network.

Question 5

Just like the possible scenarios to circumvent the resolution limit, the authors also present some possible approaches to cope with the degeneracy problem. DIscuss the proposed solutions and discuss the short comings

Correct Answer: The authors suggest 4 different approaches

1. Combination of high modularity approaches

2. Estimation of statistical significance of a high-modularity partition

3. Use of local methods like Map Equation

✓ 4. Generative models and likelihood functions. These come with