

## 1 Question 1

Let  $G$  consist of two components: a  $K_{20}$  on vertex set  $A$  and a  $K_{10,10}$  on parts  $B$  and  $C$ . In the complement  $\overline{G}$ :

$A$  is independent,  $B \cong K_{10}$ ,  $C \cong K_{10}$ , no edges between  $B$  and  $C$ ,

and all edges between  $A$  and  $B \cup C$  are present.

The only possible triangles are:

$$\binom{10}{3} \text{ in } B, \quad \binom{10}{3} \text{ in } C, \quad 20 \binom{10}{2} \text{ using } A \text{ and } B, \quad 20 \binom{10}{2} \text{ using } A \text{ and } C.$$

Hence

$$T(\overline{G}) = \binom{10}{3} + \binom{10}{3} + 20 \binom{10}{2} + 20 \binom{10}{2} = 120 + 120 + 900 + 900 = \boxed{2040}.$$

## 2 Question 2

Let  $A \in \mathbb{R}^{n \times n}$  be the adjacency matrix of an undirected graph, so  $A$  is symmetric. Define the Rayleigh quotient

$$R(A, x) = \frac{x^\top A x}{x^\top x}, \quad x \in \mathbb{R}^n \setminus \{0\}.$$

**Step 1: Reduction to the unit sphere.** For any  $\alpha \neq 0$ ,

$$R(A, \alpha x) = \frac{(\alpha x)^\top A (\alpha x)}{(\alpha x)^\top (\alpha x)} = \frac{\alpha^2 x^\top A x}{\alpha^2 x^\top x} = R(A, x),$$

so  $R(A, x)$  is homogeneous of degree zero and is constant on rays. Thus, the stationary points of  $R(A, \cdot)$  on  $\mathbb{R}^n \setminus \{0\}$  are exactly the stationary points of  $R(A, \cdot)$  restricted to the unit sphere

$$S^{n-1} = \{x \in \mathbb{R}^n : x^\top x = 1\}.$$

On  $S^{n-1}$  we simply have

$$R(A, x) = x^\top A x \quad \text{subject to } x^\top x = 1.$$

**Step 2: Lagrange multipliers.** Consider the constrained optimization of

$$f(x) = x^\top A x \quad \text{subject to } g(x) = x^\top x - 1 = 0.$$

By the method of Lagrange multipliers, a stationary point  $x \in S^{n-1}$  satisfies

$$\nabla f(x) = \lambda \nabla g(x)$$

for some scalar  $\lambda \in \mathbb{R}$ . Since  $A$  is symmetric,

$$\nabla f(x) = 2Ax, \quad \nabla g(x) = 2x,$$

so the stationarity condition becomes

$$2Ax = \lambda \cdot 2x \iff Ax = \lambda x.$$

Thus, any stationary point  $x$  of  $R(A, \cdot)$  (equivalently of  $f$  on  $S^{n-1}$ ) is an eigenvector of  $A$ .

**Converse.** Conversely, if  $x \neq 0$  is an eigenvector of  $A$  with  $Ax = \lambda x$ , then  $y = x/\|x\|$  lies on  $S^{n-1}$  and satisfies

$$\nabla f(y) = 2Ay = 2\lambda y \quad \text{and} \quad \nabla g(y) = 2y,$$

so  $\nabla f(y)$  is proportional to  $\nabla g(y)$  and  $y$  is a stationary point of  $f$  under the constraint  $g(y) = 0$ , hence a stationary point of  $R(A, \cdot)$ .

Therefore, a non-zero vector  $x$  is a stationary point of the Rayleigh quotient  $R(A, \cdot)$  if and only if  $x$  is an eigenvector of  $A$ .

### 3 Question 3

Using this formula, we obtain numerically:

$$Q_{50 \text{ clusters}} \approx 0.2004, \quad Q_{\text{random partition}} \approx -0.0001.$$

A random partition has modularity very close to 0 (slightly negative here), which indicates that the density of edges inside the clusters is no better (or even slightly worse) than what would be expected at random under the null model. In contrast, the partition into 50 clusters yields a positive modularity  $Q \approx 0.20$ , which means that there is a significantly higher edge density inside clusters than expected at random.

Therefore, among the two shown partitions, the 50-cluster partition is clearly more meaningful from the modularity point of view and should be considered the better (more “optimal”) clustering.

### 4 Question 4

A simple example is given by the following two non-isomorphic graphs.

$$G_1 : \quad 1 - 2 - 3 - 4$$

$$G_2 : \quad \begin{array}{c} 2 \\ | \\ 1 - 0 - 3 \end{array}$$

The shortest-path distances in both graphs produce the same multiset:

$$\{1, 1, 1, 2, 2, 3\}.$$

Therefore, the shortest-path kernel maps  $G_1$  and  $G_2$  to exactly the same feature representation, even though they are clearly non-isomorphic.

### 5 Question 5

We denote by  $\ell^{(t)}(v)$  the label of node  $v$  after  $t$  WL iterations.

**Initial labels (iteration  $t = 0$ ).** Both graphs  $G$  and  $G'$  have the same multiset of node labels

$$\{1, 1, 2, 3, 4, 4, 5\}.$$

Hence the feature vectors  $\phi_0(G)$  and  $\phi_0(G')$  (counts of each label) are

$$\phi_0(G) = \phi_0(G') = (\#1, \#2, \#3, \#4, \#5) = (2, 1, 1, 2, 1),$$

and their inner product is

$$\langle \phi_0(G), \phi_0(G') \rangle = 2^2 + 1^2 + 1^2 + 2^2 + 1^2 = 11.$$

**One WL refinement step ( $t = 1$ ).** For each node we form the new (temporary) label

$$\ell^{(1)}(v) = (\ell^{(0)}(v), \text{sorted multiset of neighbour labels}).$$

For  $G$  (graph (a)) we obtain:

$v$	$\ell^{(0)}(v)$	neighbour labels
2	2	$\{1, 3, 4, 5\}$
5	5	$\{1, 2, 3\}$
3	3	$\{1, 2, 4, 5\}$
4	4	$\{1, 2, 3\}$
$1_{\text{top}}$	1	$\{2, 5\}$
$1_{\text{bottom}}$	1	$\{3, 4\}$

so the six new labels are

$$(2 \mid 1, 3, 4, 5), (5 \mid 1, 2, 3), (3 \mid 1, 2, 4, 5), (4 \mid 1, 2, 3), (1 \mid 2, 5), (1 \mid 3, 4).$$

For  $G'$  (graph (b)) we obtain:

$v$	$\ell^{(0)}(v)$	neighbour labels
2	2	$\{1, 4, 5\}$
1	1	$\{2, 5\}$
5	5	$\{1, 2, 4\}$
$4_{\text{left}}$	4	$\{2, 3\}$
3	3	$\{4, 4\}$
$4_{\text{bottom}}$	4	$\{3, 5\}$

so the six new labels are

$$(2 \mid 1, 4, 5), (1 \mid 2, 5), (5 \mid 1, 2, 4), (4 \mid 2, 3), (3 \mid 4, 4), (4 \mid 3, 5).$$

Across both graphs, these temporary labels are distinct *except* for  $(1 \mid 2, 5)$ , which appears once in  $G$  and once in  $G'$ . After re-encoding to integers, the feature vectors  $\phi_1(G)$  and  $\phi_1(G')$  (these are just counts of the new labels) have only one common nonzero coordinate, corresponding to  $(1 \mid 2, 5)$ , giving

$$\langle \phi_1(G), \phi_1(G') \rangle = 1.$$

**WL kernel after one iteration.** The WL subtree kernel with  $h = 1$  is

$$k^{(1)}(G, G') = \langle \phi_0(G), \phi_0(G') \rangle + \langle \phi_1(G), \phi_1(G') \rangle = 11 + 1 = 12.$$

**Interpretation.** At iteration  $t = 0$  the graphs look identical to the kernel (same multiset of node labels). After one WL refinement, only one node in  $G$  and one node in  $G'$  still share the same label (the node labeled 1 whose neighbours are labeled 2 and 5). All other nodes have different refined labels. Thus the WL kernel after one iteration is relatively small, indicating that, although the graphs have the same label distribution, their local neighbourhood structure differs significantly. The WL procedure has therefore started to distinguish the two graphs based on their structure.

## References