Q4

(1.) Depending on the value of k, we have a certain amount of clusters where we assign the data points in.
The number of cluster k determines the number of the k eigenvectors of L.
Hence, it also determines how many eigenvectors are being used to run the k-means algorithm: $Z = (v^1, v^2 \dots v^k)$

(2.) See Jupyter NoteBook in .zip

(3.)

k	2	3	4
Mismatch 0	0.34%	0.34%	0.34%
Mismatch 1	0.31%	0.31%	0.31%

The mismatch rates remain the same over $k = 2, 3, 4$

However, due to the random start point (k-means) the results might fluctuate.

(4.) Since the mismatch rate is already reasonably small there is no tuning of k needed.
However, the common approach to increase k led to worse results than $k = 2, 3, 4$.
This indicates that there might be a miscalculation somewhere. Generally, the accuracy should increase with a larger number of k.

(5.) Spectral Clustering is an appropriate method to cluster the political blog dataset. Blogs within the same cluster certainly tend to share the same political views. More ~~eor~~ tightly connected blogs ~~be~~ compared to other political blogs were not visible from the data. For further research more information of the political blogs is necessary and can be used to investigate the connection between certain blogs and their political view more detailed.