(i) possible reasons for the differences in the two communities detected by the two algorithms

For both the dataset, Karate Club data and email data, the two algorithms are applied to detect the communities. The package Communities provides a function of using Girvan Newman algorithm to detect the communities. However, it does not directly generate the best number of communities of the dataset but provides the communities of all possible values of K (number of Communities). Therefore, to find the best number of communities for using Girvan Newman Algorithm, we calculated the value of modularity of possible communities detected for the corresponding number of communities. On the other hand, Louvain community package automatically returns the best number of communities, while it also uses modularity as a criterion.

For Karate club dataset, applying Girvan Newman algorithm, 5 communities are detected in the Karate Club dataset, so the modularity is maximized. With a modularity of 0.401, the detected communities are:

[0, 1, 3, 7, 10, 11, 12, 13, 14, 15],

[2, 16, 19, 20, 30, 32],

[4, 5, 6, 9, 22],

[8, 17, 21, 23, 24, 25, 26, 27, 28, 29, 31, 33],

[18]

Applying Louvain algorithm to Karate club dataset, we found that the best number of communities is 4, at which the modularity of the communities detected is maximized. With a modularity of 0.416, the detected communities are:

[0, 1, 2, 3, 7, 10, 11, 12, 13, 14, 15],

[16, 20, 30, 32],

[4, 5, 6, 9, 22],

[8, 17, 18, 19, 21, 23, 24, 25, 26, 27, 28, 29, 31, 33]

Even though the two algorithms detect different numbers of communities, the two algorithms reach similar result. The major difference is node 18, which is an independent community while detected by the Girvan Newman algorithm, but belongs to a large community while detected by Louvain algorithm. Other than the node 18, the communities detected by the two algorithms are very similar, and the corresponding modularity are also similar

For email dataset, applying Girvan Newman algorithm, 8 communities are detected in the Karate Club dataset, so the modularity is maximized. With a modularity of 0.201, the detected communities are:

[1, 3, 4, 6, 8, 9, 11, 13, 14, 15, 17, 18, 20, 21, 22, 31, 32, 33, 36, 37, 40, 41, 42, 44, 45, 47, 49, 50, 52, 53, 54, 55, 56, 57, 59, 62, 63, 64, 65, 66, 67, 68, 71, 72, 75, 76, 77, 78, 79, 80, 81],

[2, 5, 7, 10, 12, 19, 24, 25, 26, 27, 28, 29, 35, 38, 39, 46, 48, 58, 60, 61, 69, 70, 73, 74],

[16],[51],[23],[43],[34],[30]

Applying Louvain algorithm to Karate club dataset, we found that the best number of communities is 3, at which the modularity of the communities detected is maximized. With a modularity of 0.276, the detected communities are:

[25, 2, 5, 7, 12, 19, 24, 26, 27, 28, 29, 32, 35, 37, 38, 39, 46, 48, 53, 58, 60, 61, 68, 69, 73, 74, 51, 10, 23, 43, 70, 34, 30],

[1, 13, 18, 22, 36, 40, 41, 42, 47, 49, 52, 57, 64, 77, 78, 79, 80, 81, 3, 6, 9, 11, 16, 17, 21, 31, 45, 54, 59, 62, 63, 67, 72, 76, 71, 65],

[14, 33, 50, 55, 75, 4, 56, 15, 20, 44, 66, 8]

Comparing the communities in Email dataset detected by the two algorithms to the communities in Karate club dataset detected by the same algorithms, we found that the difference between the communities in email dataset detected by the two algorithms are larger. While Louvain detected 3 communities in the Email dataset, Girvan Newman detects 8 communities, while 5 of them only contains a single node. Moreover, while Girvan Newman detects 8 communities, it reaches the highest modularity of 0.201. However, Louvain detects 3 communities, and the corresponding modularity is 0.276, which is much higher than 0.201.

Based on the results of the two datasets detected by the two algorithms, it appears that Girvan Newman is more like to place a single node into its own community to reach higher modularity for overall dataset. The reason might be derived to the largest difference between Girvan Newman and Louvain, which is how they approached to detect the communities. Girvan Newman is a top-down algorithm, which detects communities by progressively removing edges from the original network. It first calculates the betweenness of each edge, then remove the edge with the highest betweenness. The Girvan Newman algorithm repeat this step until each node is in its own community.

On the other hand, Louvain is a bottom-up algorithm, which is consisted of two steps. For the first step, the algorithm placed each node in its own community. For each node, the algorithm calculated the modularity gains obtained by moving it into each one of its neighbors’ community. Then, each node is moved move the node to the community associated to the largest gain or stays in its original community if no gain is possible. The Louvain algorithm applies this procedure repeatedly and sequentially for all nodes until no further improvement can be achieved, which leads to the end of this first step. For the second step, the algorithm builds a new network whose nodes are the communities estimated during the first step. The inter- and intra-community links are represented in the new network by weighted regular links and self-loops, respectively. The first step is then applied to this network, and both steps are repeated until stable communities are reached.

(ii) and possible explanation and insight into the communities based on the two algorithms