

In this homework, we will work with the "tweets-2017-1004" network. This is a set of 318 hashtags mined from Twitter on 10/4/2017. It is a one-mode hashtag projection of the user-hashtag bipartite network. In other words, there is a link between two hashtags if a given user used both of them within the data gathering time period. The edges have a weight that reflects how often such pairings occur. There are 837 edges; reflecting a density of 1.7%.

Part I: Simplifying the network

1. Load the twitter-20171004.graphml file.
2. Using the technique we saw in class, create a line plot of weighted degree with log-log axes.
3. Remove the nodes with weighted degree with weight of 1 or 2.
4. Using network from step 3, apply decompose and extract the giant component.

You should end up with a network of 172 nodes and 675 edges. Use this network for the rest of the assignment.

Part II: Community detection

1. Set the random seed to a fixed value, such as your birthday. This will ensure that the random aspects of the clustering will work the same way each time.
2. Run five community detection algorithms on the network (use weights)
 1. Leading eigenvector
 2. Fastgreedy
 3. Edge betweenness
 4. Walktrap (# of steps = 5)
 5. Walktrap (step = 10)
3. Compare the different clusterings based on length (the number of clusters found) and modularity using bar charts.
 1. You will find that the modularity scores are very close together and the bar chart doesn't work so well. Subtract the mean from the modularity values and plot the deviation from the mean. This will make the (small) differences easier to see.
4. Remove the result with the largest number of clusters. With the remaining four metrics, use the arandi() method from the mcclust package to compute the adjusted Rand index between all pairs of clusterings (6 pairs total). Create a similarity matrix as in the in-class example and plot using ggcorr().
5. Use the communities from the Leading Eigenvector and Walktrap 10 algorithms for the rest of the assignment.

Part III: Visualization

1. From the network, chose a hashtag that you're interested in.
2. Use `which` to find the node associated with your hashtag.
3. Find the cluster id for your chosen hashtag node in communities computed by LE and Walktrap 10.
4. Extract the clusters containing your hashtag from LE and Walktrap 10 algorithms into separate networks using `induced.subgraph`.
5. Plot each of the networks in `igraph`.
6. Export these networks to graph files and plot in Gephi, including the labels and sizing the nodes by weighted degree.
7. Save the images in two PNG files: `hashtag1.png` and `hashtag2.png`.
8. Embed the `hashtag1.png` and `hashtag2.png` files in your R markdown.
9. Question: What aspects of political communication or particular news events do the clusters represent?
10. Question: Which cluster makes the most "sense" to you as a grouping of terms? Why?
11. Combine `hwk5.Rmd`, `hwk5.R`, `hashtag1.png`, `hashtag2.png`, and `hwk5.html` into a zip archive and submit.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder