5. Project

Bookmark this page

Project:

The last part of this assignment is an open-ended project. Choose a sociologically interesting question about either the CAVIAR network or the facebook or twitter network from the recitation notebook section (for the social network data, go to https://snap.stanford.edu/data/ego-Facebook.html and https://snap.stanford.edu/data/ego-Twitter.html, or **update (Oct 27)** any other publicly available network data set, e.g. those at https://snap.stanford.edu/data/index.html.

Try to answer your own question using the data. You can subset the data in whichever way you desire as long as it is (sociologically) meaningful. For example, in the case of a cooffending network, you could group nodes by attributes such as sex, group edges such as repeat/non-repeating cooffenses, use the weighted or unweighted co-offending networks, focus on the largest connected component, etc. Think of how you may want to subset the data in the context of the CAVIAR or the social networks, or the publicly available network data set you have chosen.

Note (October 20): The original intention of this project was to analyze a cooffending network, and the suggestions and examples are in that context. We may release the cooffending network data at a later date and you may attempt these ideas on that network then for your own interest. For the grades and upcoming due date of this analysis, please use the CAVIAR network or social network data mentioned above, or other publicly available network data set.

Project expectations/Rubric:

1. Clearly states a sociological question which is interesting and relevant to the data. The question must be sociologically motivated: for example, “Compare the network structure in 2003 vs 2009" is not a good question, without further context. If you have some reason to believe that the network structure changes in those years, then you should make that your central question: for example, “Did crimes involving youth offenders become more organized and structured over the years" is a better question, from which comparing the structure in different years becomes part of the methodology to answer the question. More examples of possible questions for cooffending networks are provided below.

2. - **(2 points)** Describes methodology for network analysis.

   - **(2 points)** Grader is convinced that the methodology makes sense for the question to be answered. Grader is convinced that no additional methodology **within the bounds of techniques taught and discussed in this module** could be applied beyond what was described. The grader should only consider additional methodology that adds meaningfully to the answer for the question: additions that simply repeat or confirm the presented results should not be considered by the grader. If a justification is provided for why a particular method was not used, the grader should be convinced by that argument.

3. - **(2 points)** Presents results, including figures and/or statistics, which address the question of interest.

   - **(2 points)** The described methodology has been applied in complete and the results shown (that is, the author did not forget to include anything they discussed in the methodology.)

4. Adequately discusses the results obtained.

   - **(2 points)** Question does not need to be successfully answered, but the grader should be convinced that the author has answered the question to the best ability of the methodology presented.

   - **(1 point)** Provides commentary on what was discovered, what were the limitations of the methods, what may have been surprising to discover, etc.

   - **(1 point)** Award this point if the question **was** successfully answered to the grader's satisfaction.

Some Possible Suggested Questions Based on CAVIAR and Social Network Data Set

Possible Project Suggestions for the CAVIAR network:

- Make use of centrality measures to identify concrete and quantifiable aspects of the criminal network that has changed over time, If possible, you could support them with a test of statistical significance. Then, provide a coherent explanation that provides examples using specific datapoints.

- How has the immediate network of central criminal figures evolved in response to the police operations?

- Consider the starting links as well as the new connections that arise throughout phases. Of the four network models discussed (Erdos-Renyi, configuration, preferential attachment, and small-world), which one(s) are the most realistic? Provide statistical tests if/when appropriate.

- Consider the clustering problem and algorithm as discussed in the “Spectral Clustering" lecture, and implement this procedure to the CAVIAR networks. What are some changes that you would notice on either the clustering quality (i.e. modularity) or the output clusters, over time? (You may feel free to perform the algorithm on an appropriate subset of the graph, say by excluding certain nodes.)

Possible Project Suggestions for the Facebook/Twitter network:

- Which attributes of the social network are assortative? Are there any disassortative attributes?

- Pick one or both of the social media network graphs, and study select features of the graph such as the degree distribution, clustering, and centrality metrics. To what extent does the power-law distribution hold, and examine how these statistics can be used to select a candidate model among the four network models discussed (Erdos-Renyi, configuration, preferential attachment, and small-world)?

- Formulate several hypotheses on how the Facebook and Twitter network graphs could differ, based on features or quantities discussed throughout the module. Then, devise a methodology to test for whether the difference is statistically significant, and carry out the analysis.

- Consider the nodes that have the most connections (in Facebook) and who are followed by the most other accounts (on Twitter). Then, consider the subgraphs that consist of each of these nodes' followers. Formulate any interesting questions based on these subgraphs or their properties.

Possible Project Suggestions for the Cooffending Network (since you will not be using this data set, this is just to give you further ideas of what is expected.):

- A common question in co-offending networks is the stability of relationships over time: do offenders commit crime with the same people over time, or do they choose new people to co-offend with as time passes? You may wish to investigate individual nodes to see how their neighborhoods change over time.

- There are other potential examples of time-based studies. For example, is there a concept of seasonality in the structure of the data? Do crime networks become more clustered during certain months or seasons?

- One can choose groups of nodes and consider homophily/assortativity, that is, how these nodes interact. For example, one can ask how females interact with males in the network: does one of the two sexes tend to have higher centrality in the network? One can also extend this to time-varying studies: for example, does the influence of females increase in more recent years?

- The CAVIAR data involved an organized crime ring. Is there organized crime in the data? How does this network compare with the organized crime in the CAVIAR network?

- How does the type of crime impact the network? Are there certain local structures, such as cliques or star graphs, that are associated with different types of crime? Can you identify different types of crime by the structure of a co-offending relationship alone?

Discussion

Hide Discussion

Topic: Module 3: Network Analysis:Networks: Written Analysis, Peer Review and Discussion / 5. Project

Add a Post

All Posts

[Staff] Statistical Significance of Degree Centrality

question posted 2 months ago by **jtourkis**

"Make use of centrality measures to identify concrete and quantifiable aspects of the criminal network that has changed over time, If possible, you could support them with a test of statistical significance. Then, provide a coherent explanation that provides examples using specific datapoints."

1) For this suggestion, if we chose a bipartite dataset, is it reasonable to test differences in statistical significance of degree centrality. ie. Create a separate treatment and control graph. Calculate the degree centralities for nodes and then perform a statistical test on the difference.

2) Can I use .05 for each test or would that require a bonferroni correction? I'm still a little confused on what counts as multiple tests. If I am comparing degree centrality of node A in control graph to node A in treatment graph and I am comparing degree centrality of node b in control to node b in treatment, am I correct in believing I can still use .05 because they are different tests? I would only need a correction if I was comparing A to A, the A to B, etc...

Thanks in advance for the clarification.

This post is visible to everyone.

0 responses

Previous

Next

edX

About
Affiliates
edX for Business
Open edX
Careers
News

Legal

Terms of Service & Honor Code
Privacy Policy
Accessibility Policy
Trademark Policy
Sitemap

Connect