



Social Network Analysis for Criminal Justice Practitioners and Analysts

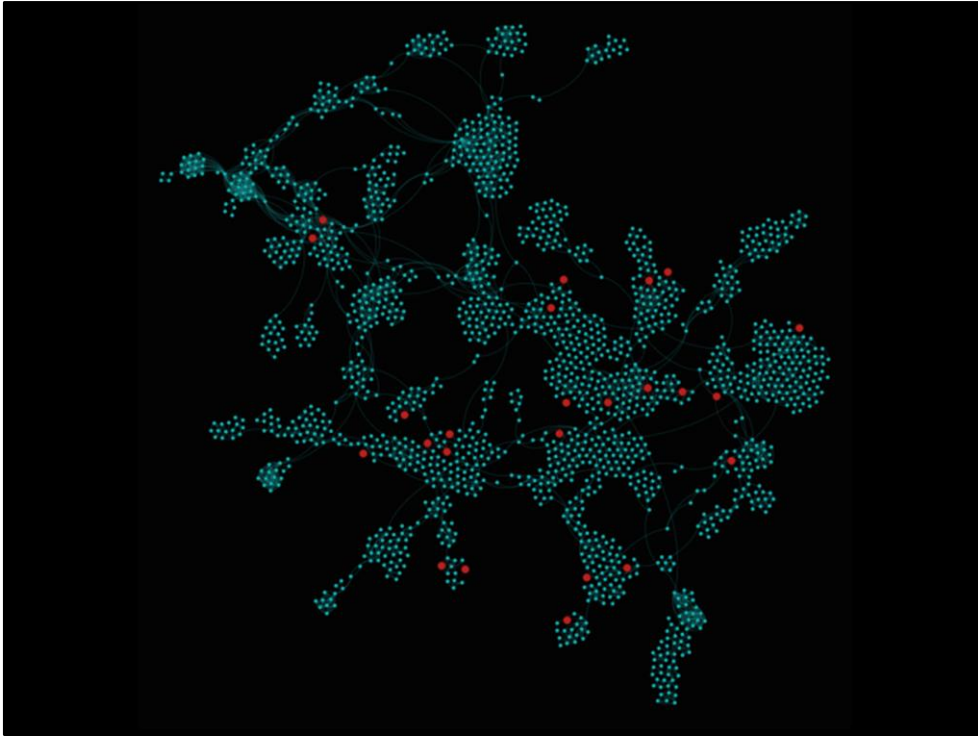
Module 3: Visualization

Andrew V. Papachristos
© 2016



Module 3: Visualization

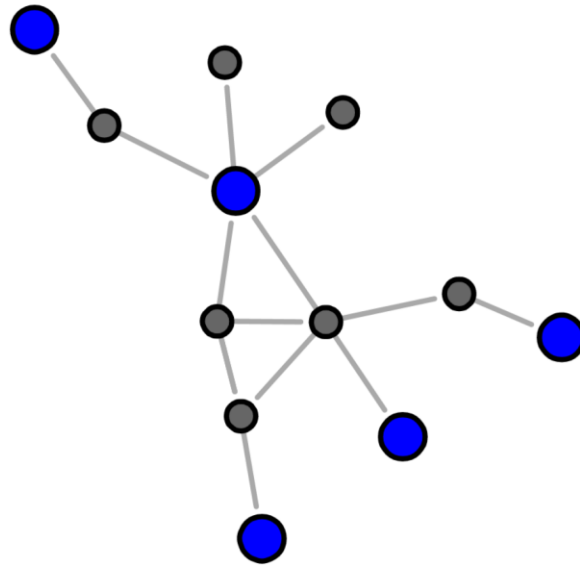
Welcome to Module 3: Visualization. The goal of this module is for participants to present social networks in visually meaningful and powerful ways. This requires learning the different parts of network visualization that can be customized and learning how to customize effectively. This module introduces tips and best practices for network visualization and presents some visualization options.



Visualization provides a bird-eye view of local structures that may have not been obvious when looking at matrices, edgelist, and two-mode datasets. Effective visualization requires some skill so that network images are meaningful and contribute to the analysis, but there is no one right way to visualize networks. I have, however, compiled some suggestions and tips for things to think about when visualizing social networks.

Tips for Visualizing Networks

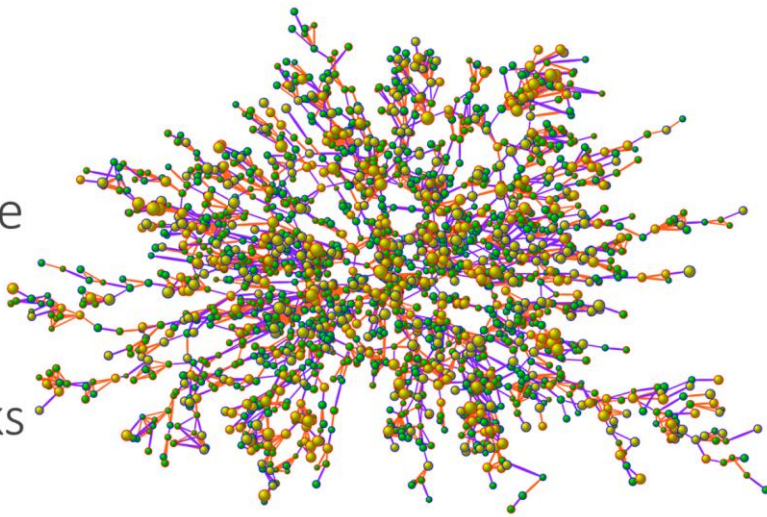
Tip #1: Aim for readability and parsimony



Network visualization tip #1: The goal of network visualization should be readability and parsimony. Social network users want visualization to add something to the analysis rather than distract from the analysis. Our brains are incredibly good at detecting patterns, so we want the visualization to contribute to cognitive recognition. This requires clear and readable networks. The most parsimonious network will use as few elements as possible while maximizing readability.

Tips for Visualizing Networks

Tip #2:
Minimize
detail in
large
networks

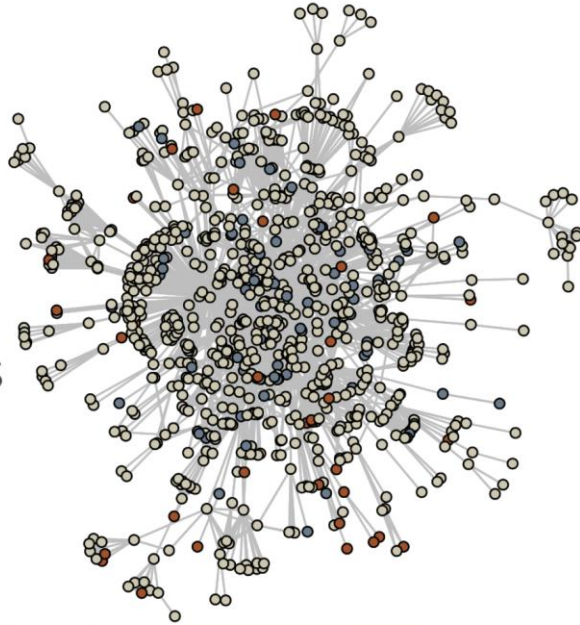


(Christakis & Fowler 2007)

Network visualization tip #2: Detail works well in smaller networks but tends to get lost in large networks. Networks with around 30 or less nodes tend to visualize well with detail. The example network in this slide is a well-known network visualization from social network health research on the social contagion of body mass index by Nicholas Christakis and James Fowler. The network image includes more than 2,000 nodes. The size of the nodes is proportionate to the person's body-mass index. Yellow nodes indicate obese people and green nodes indicate non-obese people. Nodes with blue borders are men and nodes with red border are women. Purple edges are friendship or marriage ties and orange edges are familial ties. There is an incredible amount of information contained in this single image, but it can be difficult to distinguish some of the details in such a large network especially on a small screen or page. Be conservative in the amount of detail you include in large networks, and consider the benefits of zooming in on sections of the network to emphasize detail.

Tips for Visualizing Networks

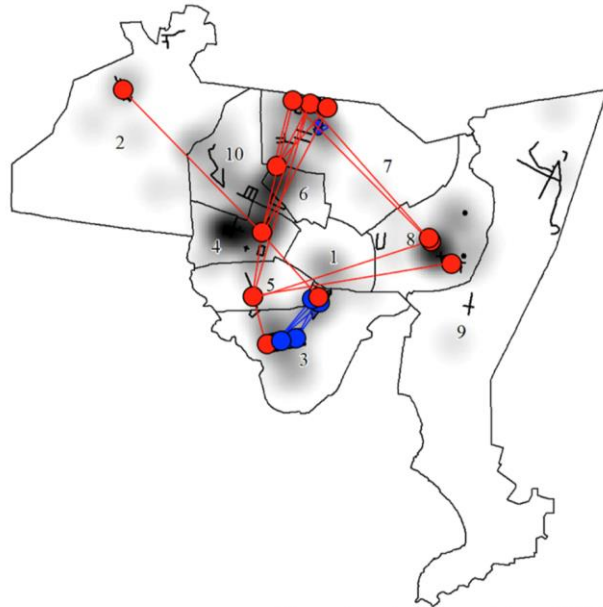
Tip #3: Minimize overlapping nodes and edges



Network visualization tip #3: Minimize overlapping nodes and edges. Again, this is easier to accomplish with smaller networks than with larger networks. However, some plotting algorithms will improve the spread of nodes and edges. In this example network of more than 900 nodes, it is difficult to see what is happening between the nodes and edges in the center because of all the overlap.

Tips for Visualizing Networks

Tip #4:
The placement
of the nodes
isn't always
meaningful

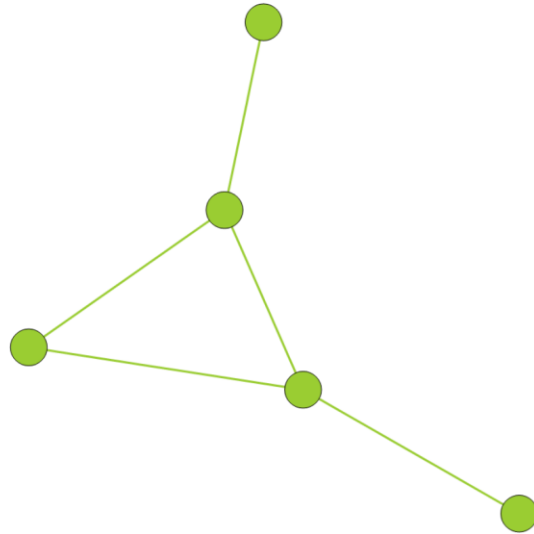


(Sierra-Arevalo & Papachristos 2015)

Network visualization tip #4: Remember that the placement of the nodes and edges isn't always meaningful. Algorithms plot nodes and edges in a graphing space usually with the most central nodes in the middle of the plot and less central nodes toward the periphery of the plots. There are a variety of algorithms to choose from to place the nodes in different ways. It is possible to map networks into geographic space, such as in the example above that maps gangs in the city of New Haven. In this example, the nodes and edges create a geographically relevant placement, situated across the city of New Haven. The most central nodes are not in the middle of the network. However, for the introductory purposes of this tutorial, location in the network does not indicate a location in space. We will mostly use default algorithms that place the most central nodes in the center of the network and the less central nodes toward the outside of the network.

Tips for Visualizing Networks

Tip #5:
The lengths of
the edges
aren't always
meaningful

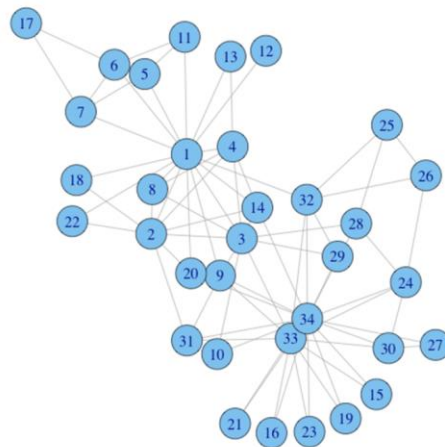


Network visualization tip #5: The same is true for the length of the edges. There are advanced applications in the visualization of social networks that can plot the length of the edges based on some value, but for the most part the varying lengths of the edges is an outcome of the plotting algorithms rather than something that is substantively meaningful. This example network is called “bull,” and it is built in to the igraph library. The default plot produced some variation in the length of the edges even though the the different lengths don’t mean anything in the 2-dimensional plotting space. In more advanced applications you could make the length or the width of an edge equal come value.

Customization

In Module 2: Data, participants explored two levels of data used in social network analysis: attributes or individual-level data and edges or relational-level data. Converting data into visual networks allows variation on those same two levels. You can customize the visualization of the nodes and/or you can customize the visualization of the edges. Customizations can include shapes, colors, labels, sizes, etc. Customization decisions will likely be shaped by your intended use of the network image. In print publications, color might not be an option in which case you will need to limit your customization to black, white, and gray. In powerpoint presentations, you might consider bold colors and large nodes with few other details. In a working paper or an online report, you might have the space for more detail and varying color palettes.

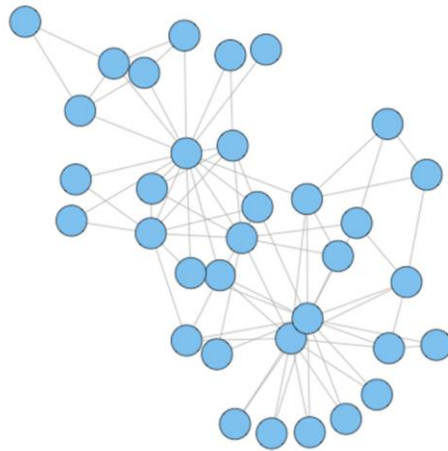
Customizing Networks



This network uses default settings.

Visualization will largely depend on the software that you use. Each software package has its own default settings for plotting social networks. These next few slides show some of the visualization options in RStudio using the igraph package. Participants not in the RStudio track might have different visualization options depending on their software preferences. This is an example network of friendships in a karate club that comes with the igraph library in RStudio. In the image above, the karate network was plotted using only the default settings in igraph. As you can see, at the time of this production the igraph defaults are light blue circles for the nodes, light gray thin lines for the edges, and darker blue labels located in the center of the nodes based on the unique identifiers from the datasets.

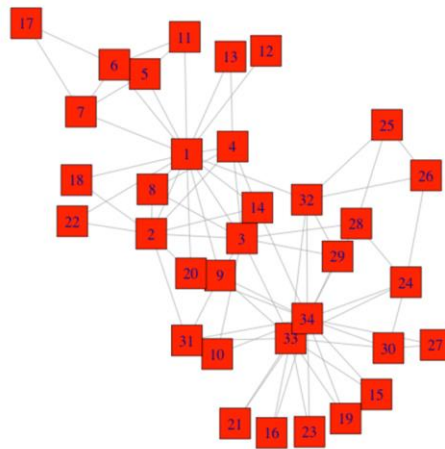
Customizing Networks



Node labels can be customized.

You can turn off the node labels in a network as in the example above or assign a special name or identifier to be the labels. In this example, the number identification labels have been turned off so that they don't appear in the middle of each node. Long names won't work well as labels in big networks as the labels will overlap. Even short names in this relatively small network might overlap because some of the nodes are so close together. You can also customize the location of the labels in relation to each node, so that the labels aren't fixed to the center of the node but rather above each node or to the side of each node.

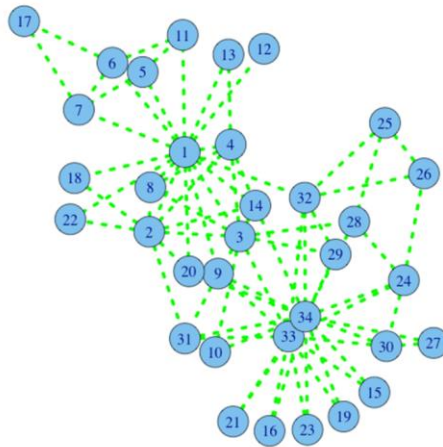
Customizing Networks



Nodes can be customized, ...

When you think about customizing your network, you will want to consider what you want the nodes to look like. In this example, the nodes are set to red square shapes.

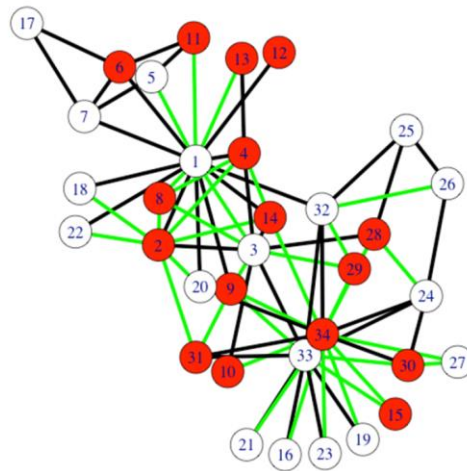
Customizing Networks



...and edges can be customized.

The lines between the nodes can also be customized. In this example the nodes have the default setting, and the edges have been customized to be thicker, dotted, and green.

Customizing Networks



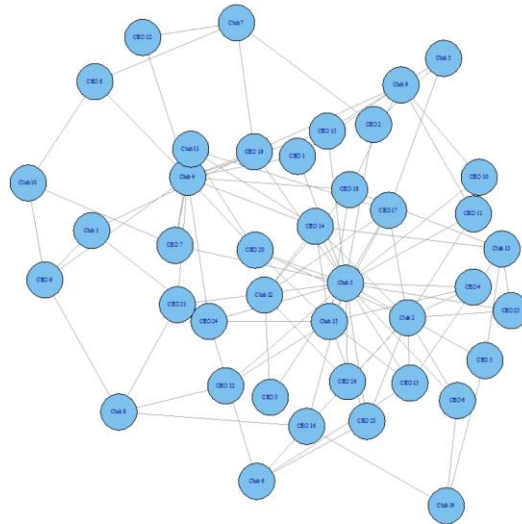
Vary customization within the set of nodes and edges.

Often there are differences among the nodes and between the edges that are worth presenting in network visualization. Though it requires some extra steps, it is possible to customize nodes and edges differently within the same network based on some sort of attribute. In the example above some of the nodes are red while others are white, and some of the edges are green while others are black. These were just randomly assigned for a visualization exercise, but perhaps the red nodes are boys in the karate club, white nodes are girls in the karate club, green edges are friendships, and black edges are sparring during karate lessons. Even considering the meaning of hypothetical colors adds important information to our interpretation of the visualized network.

Visualizing Two-Mode Networks

Recall from Module 2: Data that two-mode networks are networks that contain two different types of nodes. These are also called affiliation or bipartite networks. It is worth being familiar with two-mode networks because they are often compatible with the format of official arrest data.

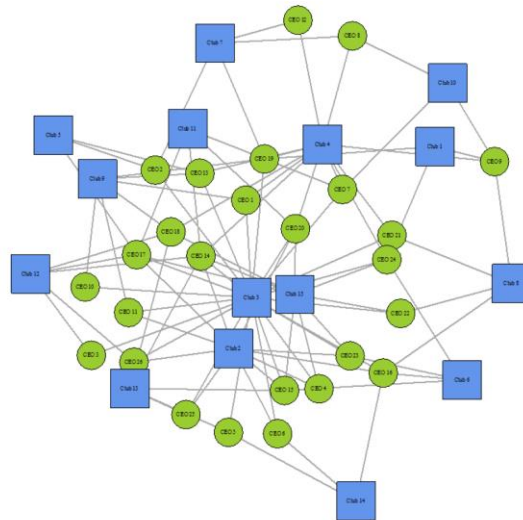
Two-Mode Networks



A two-mode network with default plot settings.

Two-mode networks especially require some attention to visualization customization. Without specification, `igraph` plots a two-mode network with the same default settings as it plots a one-mode network. There is no differentiation in this example above between the two different types of nodes. At first glance, especially with the labels being so small, it is difficult to see that this is even a two-mode network.

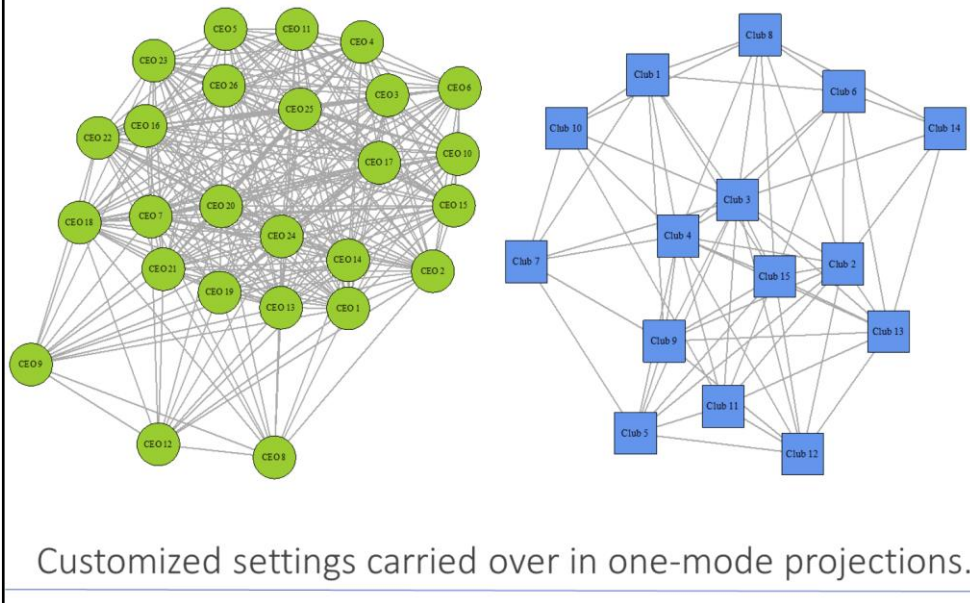
Two-Mode Networks



The same two-mode network with customized plot settings.

This is the same network as the previous slide, a group of Minneapolis CEOs and their associations in various clubs. These data come from Joseph Galaskiewicz's 1985 book, *Social Organization of an Urban Grants Economy: A Study of Business Philanthropy and Nonprofit Organizations*. In this visualization, the two different types of nodes are distinct by customizing the shape, size, and color of the two types of nodes. It is clearer in this visualization that the blue squares are the clubs and the green circles are the CEOs.

Two-Mode Networks



Recall from Module 2: Data, the projections show how nodes of a single type are connected through the second type of nodes in two-mode networks. They project the two-mode network into two one-mode networks. If you are interested in demonstrating the one-mode projections from a two-mode network, it might be useful to carry over the same visualization customization settings to continue to differentiate between the two different types of nodes. Above are the two one-mode projects from the Minneapolis CEO affiliation networks. Maintaining the same green circles for the CEOs and the same blue squares for the clubs shows the continuity between the two-mode version and the one-mode projections. We see clearly tight social network between CEOs based on their shared clubs and the many connections between clubs based on their shared CEOs.

REFERENCES

Christakis, Nicholas A. and James H. Fowler. 2007. "The Spread of Obesity in a Large Social Network over 32 Years." *The New England Journal of Medicine* 357:370-9.
(<http://www.nejm.org/doi/full/10.1056/nejmsa066082>).

Galaskiewicz, Joseph. 1985. *Social Organization of an Urban Grants Economy: A Study of Business Philanthropy and Nonprofit Organizations*. Orlando, FL: Academic Press, Inc.

Sierra-Arevalo, Michael and Andrew V. Papachristos. 2015. "Applying Group Audits to Problem Oriented Policing in New Haven." In *Preventing Crime with Network Analysis*, edited by A. Malm and G. Bichler. Boulder, CO: Lynne Rienner.



Social Network Analysis for Criminal Justice Practitioners and Analysts

Module 3: Visualization

Andrew V. Papachristos
© 2016

Module 3: Visualization covered a few best practices when generating network visualizations. Several customization options for nodes and edges were also introduced as well as some things to think about in two-mode networks. Next participants should turn to the Module 3: Visualization Lab for a short Google activity. Participants in the NAVCAP track will explore some of the visualization options in the NAVCAP application. Participants in the RStudio track will import social network data into RStudio, convert the data into networks, and iteratively build customization arguments to the basic plotting command. Module 4: Analytics introduces some of the calculations and statistics that can be analyzed in social networks.