Raul Martinez - Ivan Ulloa

Exercise 4 Report: Visualizing a Network for Bighorn Sheep Dominance

Motivation

Networks are used in applications where complex relationships and data structures need to be organized and later visualized. They comprise two main elements referred to as nodes and edges, and both of them can hold key-value properties. The network presented in this report uses data records for source and target dominance of 28 female bighorn sheeps collected on the National Bison Range in 1984, it tells information about their social status for expression, development, and reproduction. The goal of this visualization is to store and illustrate bighorn dominance with an interactive interface that would allow users to navigate and explore data relationships.

Data Augmentation

Data at the nodes includes an identification number and age for each sheep, while data at the edges contains sheep identification for the source and target relationships, and this is described with a weight which is the number of observations for dominance for that specific record. Data augmentation is presented at the nodes by adding two additional properties named as "Sheep Dominated" and "Aggressiveness" for each sheep, they basically reveal behavioral information about dominance relative to other sheep based on historical records stored at the edges. The "Sheep Dominated" field represents the number of sheep that every one sheep has been able to dominate. The Aggressiveness score is calculated in the following way:

$$A_{score} = ln(\#dominations/age)$$

The above metric aims at expressing the total number of times a sheep was spotted dominating another sheep, this is normalized by age to better compare among all other sheep. The natural logarithm is used to smooth the color scale.

Tasks

The graph and network structure for the bighorn sheep dominance discussed aims to explore the following questions: How is dominance explained with the different properties available such as age and number of occasions? Which are the most dominant sheep? Which are the least dominant sheep? Which source and target sheep relationships are interesting? Why?

Expressiveness of design

The visualization shown in Figure 1 is expressed as a network scatter plot where the nodes which contain information about each individual sheep are represented by circular markers. The markers make use of two dimensions to represent the "Aggressiveness" as color and the "Dominance" as the size of the node. The interconnected lines between the nodes

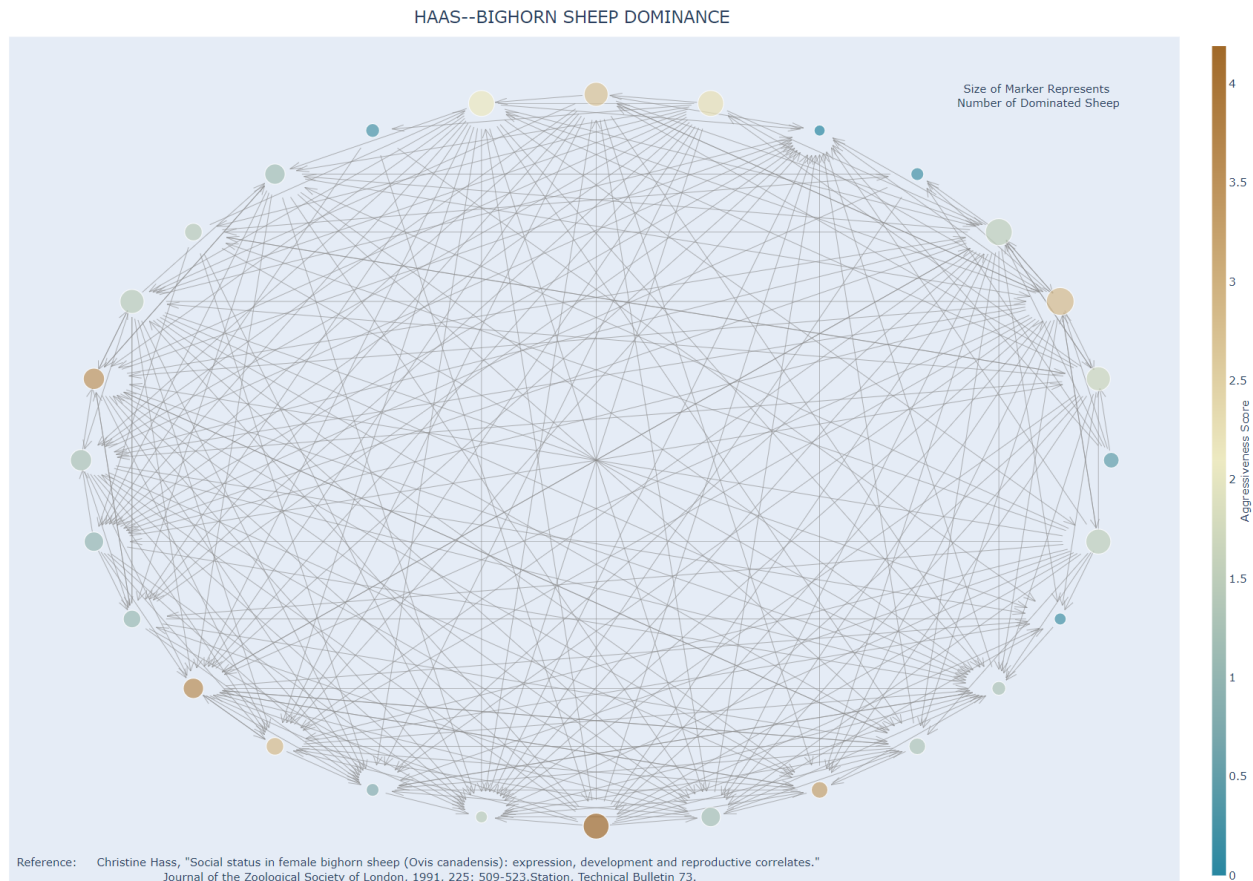represent directed edges, that is, the initial node dominates the end node in the direction of the arrow.

Figure 1. Visualization shows the network graph where the circular markers represent nodes or individual bighorn sheep and the lines-arrows represent sheep spotted dominating another sheep.

Effectiveness of the solution

There is an emphasis on using the hover feature to show the metadata and explore each of the sheep nodes. The divergent "earth" color scale used to express the aggressiveness metric of sheeps is the most distinctive feature as it allows for a quick comparison between the nodes via color differences. Using the size of nodes allows to distinguish between the most dominant and the least dominant nodes, here the middle range values are slightly more difficult to identify due to the markers being spread across the figure which in turn makes side to side comparisons difficult. There was careful consideration in scaling the size of the nodes as the values in some cases could have been zero, to avoid this and offset value was used on all of the nodes to increase the minimum size and allow for the markers to be visible.

Interaction

Interactive features consist of a zoom in/out box function as well as hovering metadata information for each of the nodes. As previously mentioned, the use of the hovering tool is critical in order to better compare values as each of the node sizes could have a slight difference between them making it difficult to compare by just looking at their size. Figure 2 shows an example of the metadata presented when hovering over a node. The information is used to first identify the Sheep with its ID, the Sheep Dominated field represents the number of sheep that this sheep has dominated at least once, next is a field for the age of sheep and lastly the Aggressiveness metric.
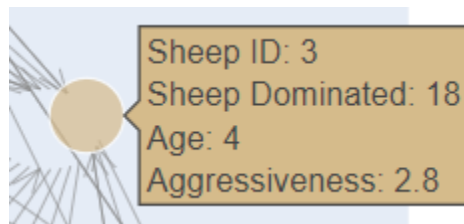

Figure 2. Metadata shown with a hover tool at each node. Depicted information contains sheep ID, number of dominated sheep, age, and the aggressiveness score metric.

Conclusion

The network and visual elements presented above allows to explore bighorn sheep dominance for data collected in 28 females. It was noted that most of the more aggressive bighorn sheep (darker color) are also the youngest, reporting 1 year of age in most cases. Sheep ID #3 is reported as the most dominant with dominance over 18 different sheep as well as an aggressiveness score of 2.8. The most aggressive sheep was ID #22 with a score of 4.2. The older sheep 4+ years tend to have average dominance counts but lower aggressiveness scores compared to their younger counterparts. Overall the network visualization provides enough information to answer the questions presented here, that said, additional information could be extracted from the node to node link using weights where sheep-sheep dominance information is contained.