Report: Sheep fights

Michael Gallaspy and Kevin Kannappan

Motivation

We chose to visualize the dataset as a node-link diagram in order to understand relationships between sheep. The kinds of questions we hope to answer include:

- Which sheep fight the most? Which sheep win and lose the most?
- Which sheep dominate other sheep? To what extent?
- What kind of sheep relationships and behavior can we infer from the dataset?

Visualization

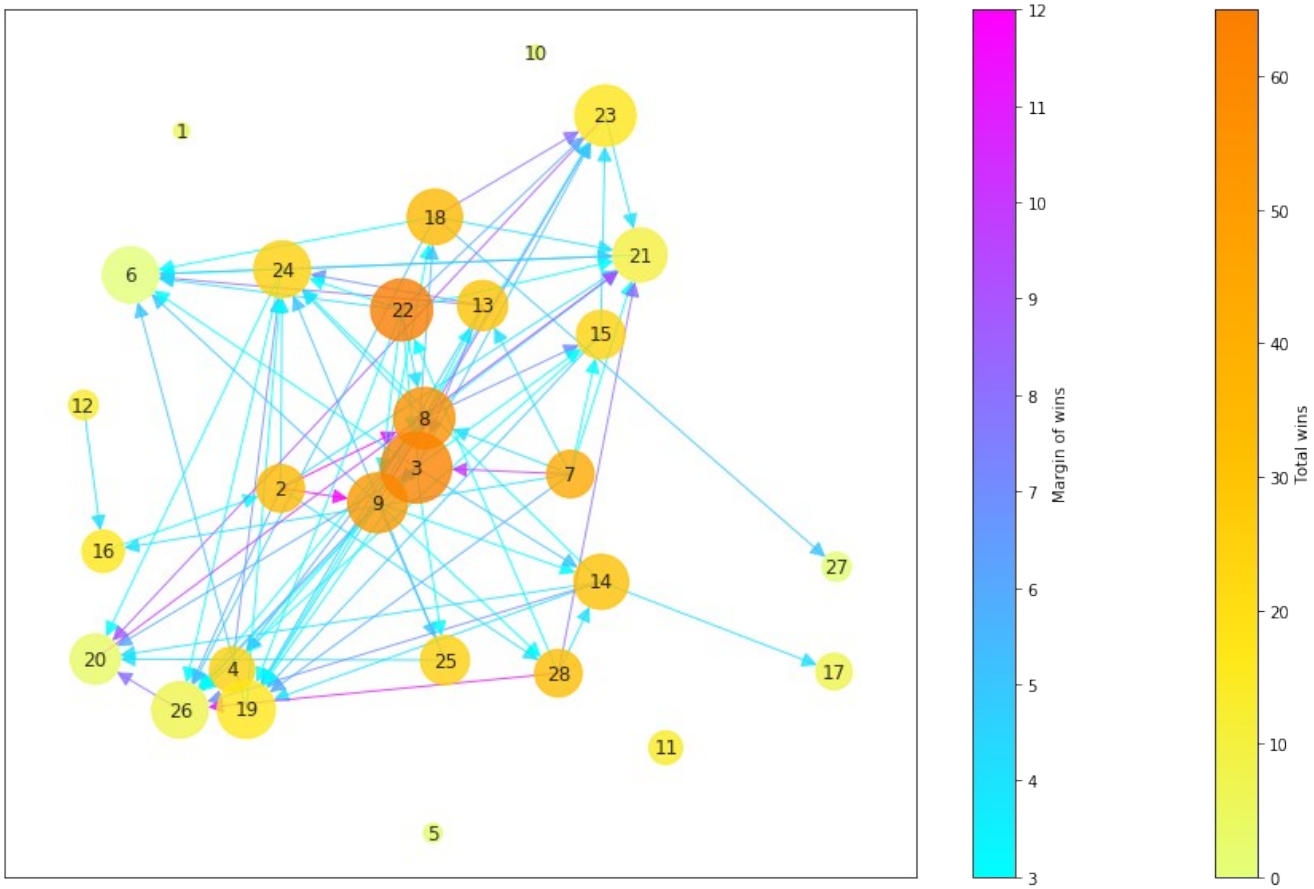

Figure 1: Sheep fights. See text for detailed description.

The visualization is a node-link diagram where nodes represent sheep, and directed edges between sheep encode information about their fights. Several channels have been used to encode information, including position, color, and size of various elements. Because nearly every pair of sheep

has fought, it is impractical to encode information in link channels while simultaneously naively drawing a link for every pairing – the number of links makes visual differentiation of edges impossible. For that reason a derived value from the data set is used to determine a new set of edges. The age attribute of sheep is ignored as no compelling connection was discovered between age and other attributes in the data set.

The nodes of the diagram represent individual sheep from the data set, with a centered label identifying individuals. The radius of the node indicates the total number of fights a sheep has been in. No legend is provided for this channel, but the values range from 4 to 99. The node color indicates the number of fights it has won, with a legend on the right providing a numerical value corresponding to the color. The color scheme has been chosen so as to be robust when simulating colorblindness, with an example included in the report.

The links of the diagram encode the "margin of wins", a derived value that is how many *more* times a sheep has won a fight than lost against another sheep. In other words, if sheep A fought sheep B 5 times and lost 3 times, its margin of wins against B would be 2, whereas B's margin of wins against A would be -2 since it lost more than it won. By considering only positive margins of wins, the potential number of links that can be drawn is reduced, however the number of potential links is still too large and so is further reduced in the visualization by considering only significant margins of wins (greater than the median). The direction of the link encodes which sheep in a pair is dominant. The color of the link encodes the margin of wins, with a corresponding legend. This color scheme is chosen so as not to interfere with the color scheme for total number of wins, and is also robust under color blindness.

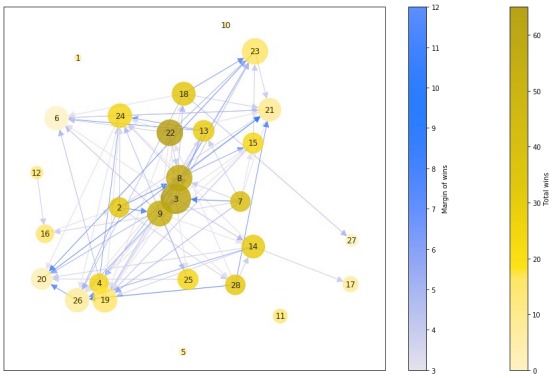

Figure 2: Simulated red-blindness, which aside from monochromacy has the most severe impact on the visualization.

The nodes are positioned with using NetworkX's spring layout, with parameters chosen to balance proximity of the nodes and legibility of the links. Nodes with links tend to be closer together under this force-based semi-random positioning scheme. In other words, the position channel encodes via proximity those sheep who tend to fight among one another frequently.

Findings

The rich encoding of information helps us to discover many findings. For one, the most victorious sheep such as 3, 8 and 22 are easy to identify through the node color, and one can see immediately from node size that they are also among the sheep who fight the most. One is also able to pick out frequent losers like 6 and 26, as they are large nodes that are more green-yellow colored. The small size and lack of connections from 1, 5, 10 and 11 to other nodes leads us to conclude both that they don't fight frequently, and moreover that when they do fight they neither dominate nor are dominated by other sheep compared to what is typical.

Proximity also shows us that 3, 8, and 9 can be “socially” associated. The links between them are obscured, but by following the links leading from them we can see they tend to fight the same sheep. Even after greatly reducing the number of links, it is difficult to draw conclusions from most of them because of the amount of overlap, however reducing the links further imperils the meaningfulness both of conclusions that can be drawn from links, and conclusions that can be drawn from position, since the links are a factor in positioning.

The link colors also lead us to interesting observations. For example one might be tempted to think that frequent winners like 3, 8, and 9 do not lose easily against other sheep. In fact we can see that the margin of wins in their fights against other ship tend to be small, and that they also lose relatively frequently against 2 and 7. The most frequent winners tend to be those sheep who fight many other sheep, however those sheep who have relatively uncontested records tend to fight less. This observation is suggestive of two hypotheses of sheep behavior: 1) those who *wander* the most and encounter other sheep tend to *win* the most, but conversely 2) that the most *dominant* sheep who lose infrequently are those with smaller “social” circles.