Exercise 04 Report: *BigHorn Dominance Network Visualization*

Adelle Driker, Bo Yan, Yuan Hu

Motivation: In order to understand and interpret the dominant relationships between female bighorns and their behaviors, a visualization of node-link network graphs was created. Attributes such as ages and observed dominance during a fight were considered and integrated in visualization.

Data: Bighorn sheep dominance http://moreno.ss.uci.edu/data.html#sheep

Node: id, age Edge: source, target, weight (observed occurrence of dominance)

Tasks: Given the sheep datasets, this report aims to interpret the following via node-link visualization using networkX:

1. Which bighorn sheep dominates the most during fights?
2. How is the age attribute related to these dominance relationships?

From a visualization prospective, the following actions will be taken:

- EDA on nodes and edges (non-visualization)
- Explain the deficiency of raw node-link visualization in networkX with 4 layouts
- Explain the effectiveness and expressiveness of improved node-link visualization
  - Edge width (indicated by normalized weights)
  - Circle size (indicated by degree of centrality or weighted degree of centrality)
  - Circle color (indicated by age groups)
  - Miscellaneous (color map, font size, layouts)

Data Preparation/Augmentation:

Firstly, basic EDA was conducted to investigate the degree centrality of all of the sheep nodes, and the count of different edges that each sheep node has was calculated.

Data processing and augmentation was mainly performed in accordance with the need of making improvements on the raw NetworkX implementation. The process includes:

- Normalize weights for edges:
  Weights for each edge were normalized so that edges can be differentiated by applying different widths.
- Calculated total count of edges for each node:
  This is also known as degree of centrality, this metric will be applied directly as a strategy to indicate node size.
- Calculated weighted degree of centrality for each node:
  Total observed occurrence of dominance (weight attributes), as a main indicator of sheep dominance, was divided by the degree of centrality that each sheep node has, so that node size on a graph can be differentiated by the weighted degree of centrality of each node.
- Ages were binned into 3 groups:
  Ages from 0-9 were grouped into (0, 3], (3, 6] and (6, 9]

Sample processed node and edge data:

	id	age	total_weight	avg_weight	age_group
0	1	8	5.0	1.000000	(6, 9]
1	2	8	35.0	2.500000	(6, 9]

	source	target	weight	norm_weight
0	1	2	1	-0.747111
1	1	3	1	-0.747111

Visualization:

Raw node-link visualization:

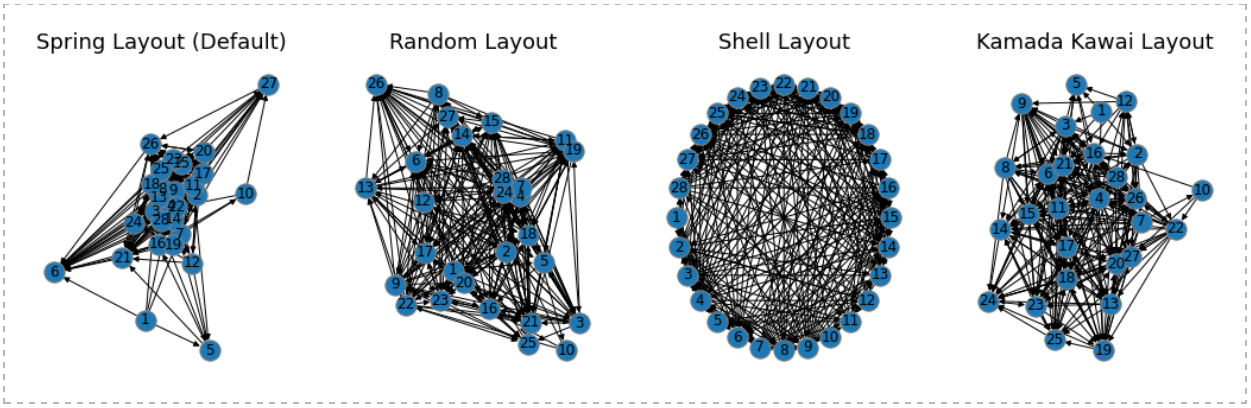

Fig. 1 Raw node-link implementation in 4 layouts

Deficiency of Fig. 1 implementation:

- Sheep nodes were overlapped in certain layouts (default and random layout).
- Very limited information was represented and conveyed because the capability of differentiation for each graph element (circle size, line width, color map) is restricted.

Improved node-link visualization:

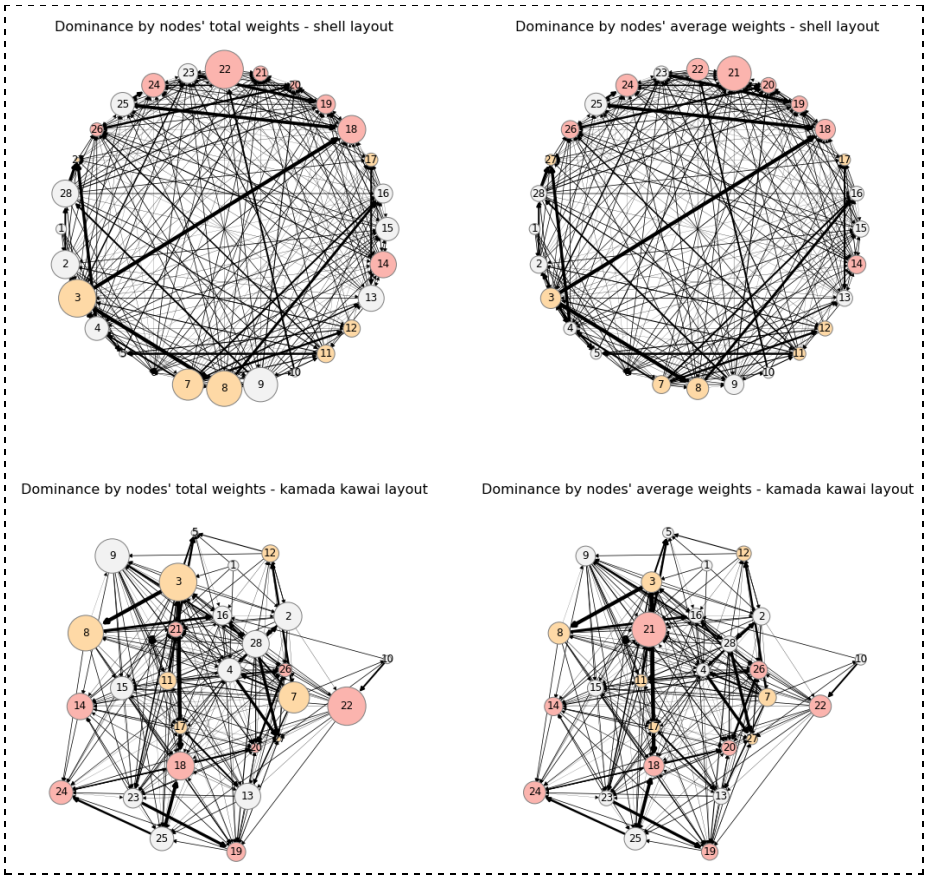

Fig. 2 degree of centrality VS weighted degree of centrality

Fig. 2 displays two sets of selected layouts: Shell and Kamada Kawai. Under each set, two strategies of weights were applied to indicate a node size.

- Strategy 1: Degree of centrality, count the number of dominance that a node has
- Strategy 2: Sum of total weights for a node then divided by degree of centrality

In Fig.2, sheep nodes were also represented in circles with different colors, which indicates the sheep node's age groups. The width of the edges represents the total occurrence of observed dominance (normalized weights) between two sheep, and the arrow indicates a dominating relationship during a fight.

When comparing the two graphs in the same layout, it renders distinct dominance results: in Kamada Kawai layout, the top 3 nodes with the most degrees of centrality are the sheep that dominate fights the most (sheep 22, 8, 3); however, when the degrees of centrality were divided by total observed occurrences, the top 3 nodes became 21, 24, 8, meaning even a sheep that may fight with mostly the same sheep, the number of fights between this sheep with same opponent may have occurred many times.

Meanwhile, ages are positively correlated to the results of fights, older sheep (red-ish circles) tend to have more dominating relationships, which also has been observed more, since there are thicker edges between an older sheep and other sheep.

Conclusions:

To conclude, in terms of which bighorn sheep dominates the most: it depends - the sheep that wins the fights with the largest number of different sheep could be the one that dominates the most, also the sheep that wins more fights with fewer sheep could also be the one that dominates the most because the fights between the same opponents may occur a significant number of times.

The age attribute can be a strong indicator of dominance relationships; more specifically, an old sheep tends to dominate more fights or has been observed more when dominating a fight with the same opponent.

From a visualization perspective, improved node-link graphs convey more effectively and expressively than the raw implementation by integrating the information to node size, node color, edge width.