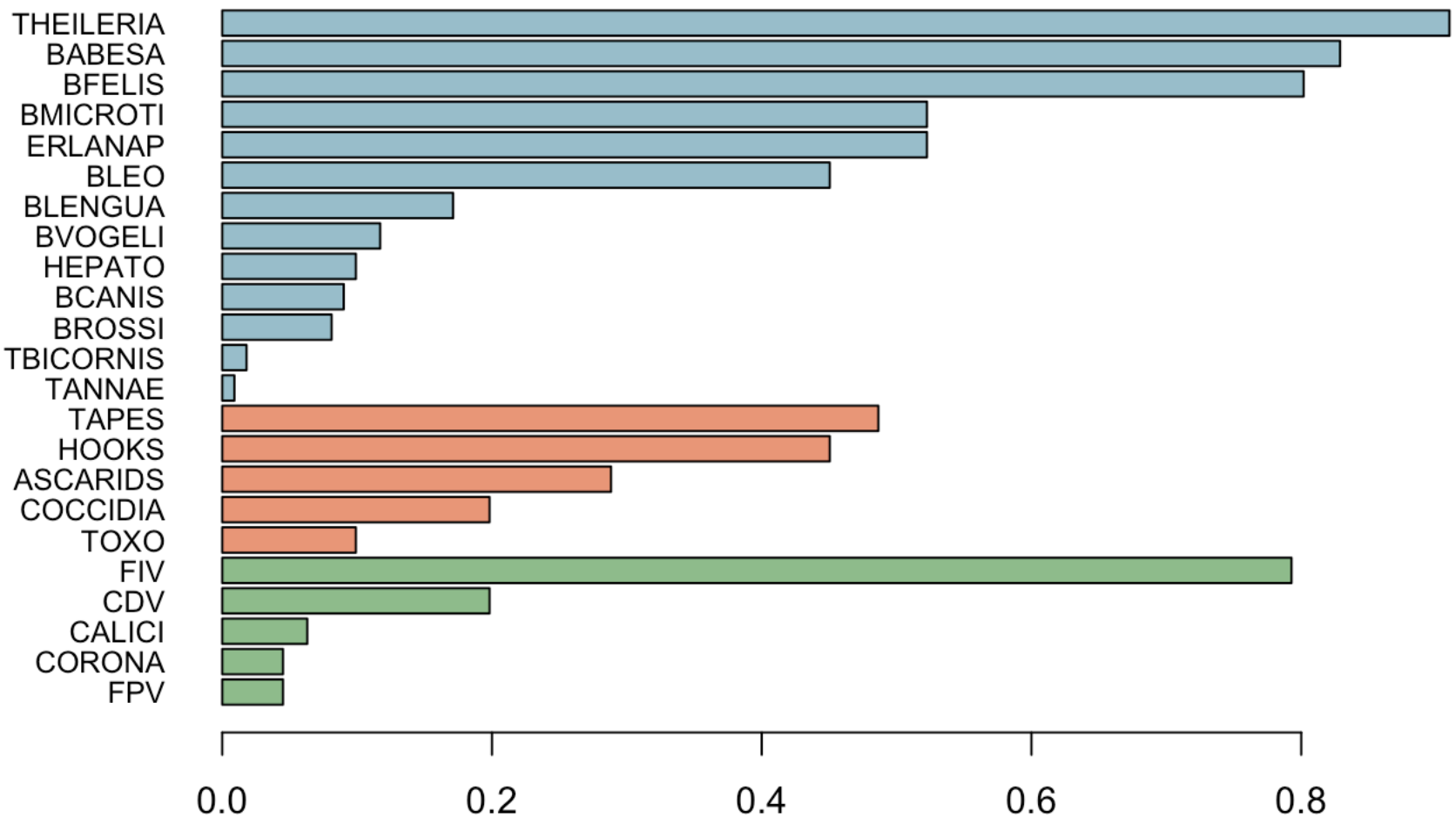
Lion Coinfection Networks

Erin Gorsich

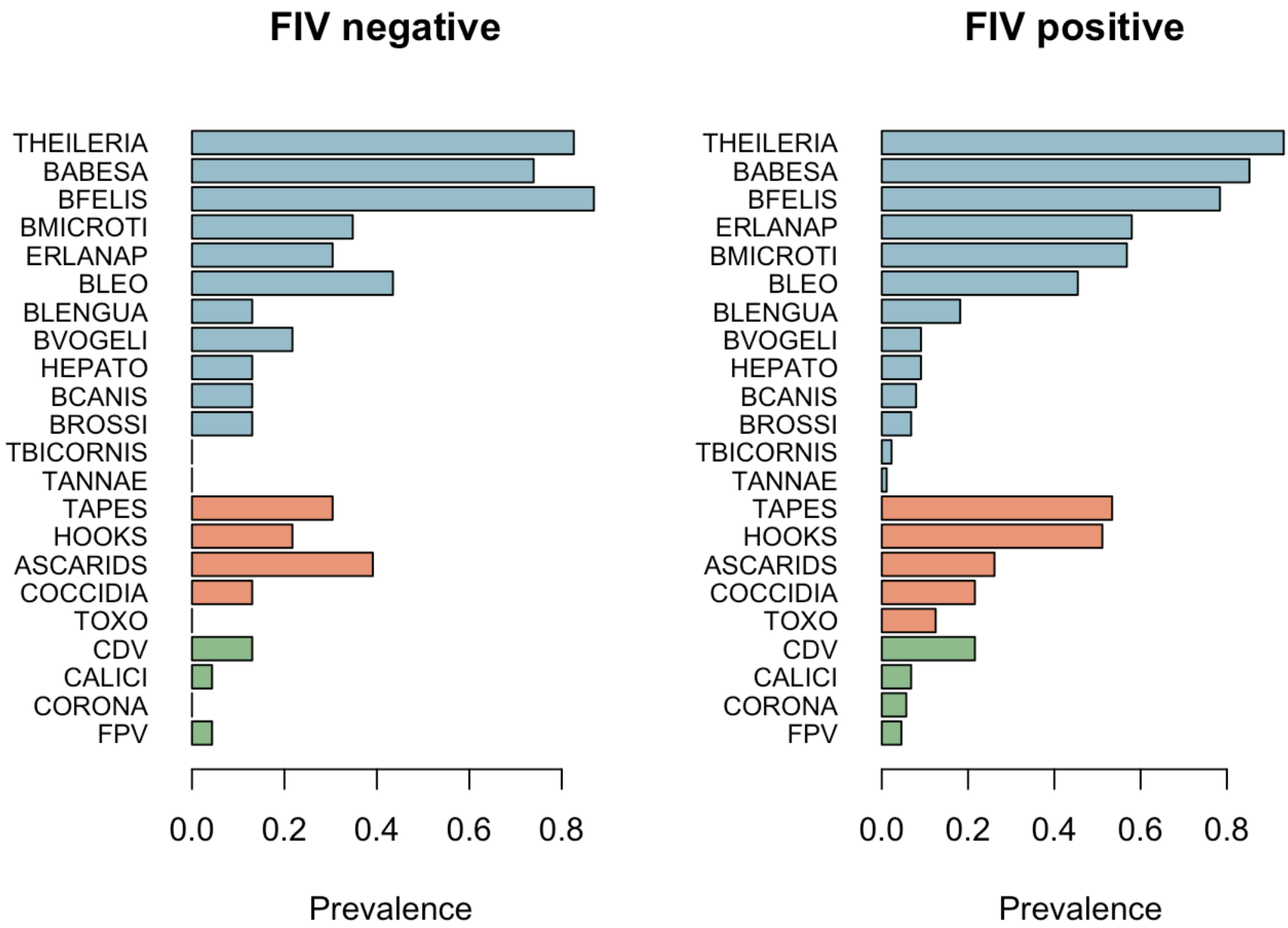
January 13, 2016

Summary statistics

Plot of overall patogen prevalence ordered by type and prevalence



Plot of prevalence in FIV+ and FIV- lions. Visually, we can see shift in pathogen community largely driven by changes in intestinal parasites (ordering consistent with above).



Network statistics

Below I consider three abstractions of the host-parastie community. A “parasite network”, a “host network”, and a bipartite network. Likely I will use the later, but they may all be useful abstractions.

Network statistics pt 1: Parasite network

For the parasite network, we define nodes as parasite/pathogen species. An edge connects two nodes if the pathogens share the same host. Edges are weighted by the number of shared hosts. All network and node statistics were calculated with igraph (for monopartite networks). In this context, we can interpret the following network statistics:

- Number of nodes: the number of parasites observed
- Mean Degree (Unweighted): the average number of pathogens each pathogen is connected to by sharing a host.
- Diameter: The maximum number of steps in the set of shortest paths between all parasite pairs. This is a measure of connectedness- if you are one parasite, how quickly can you get to all the parasites.
- Transitivity: The ratio of the number of triangles and the number of connected triples in the graph. Transitivity ranges from 0 to 1, with higher values indicating more connected subgraphs. It is a measure of clustering (high transitivity values mean you have clusters of pathogens that co-occur togeter- this will be interesting to see which ones are which!)

- Weighted Assortivity: The correlation between a parasites degree and the degree of the parasites connected to it. Assortivity ranges from -1 to 1, with positive values indicating that nodes with high or low numbers of hosts interact mostly with nodes with similar degree values. I interpret negative assortivity values as generalists pathogens are found with specialists.

```
parasitenet
```

```
##                 NumNodes MeanDegree MeanDegreeUnWt  VarDegree   Diameter
## Overall_Network       22  213.09091      15.181818 39390.4675 0.06984479
## FIV_negative          21   23.71429       5.809524   641.9143 0.20192308
## FIV_positive          21  148.76190      13.809524 19268.6905 0.09798967
##                 Reciprocity Transitivity VertexConnectivity
## Overall_Network           1    0.8027505                 NA
## FIV_negative              1    0.6322314                 NA
## FIV_positive              1    0.7852548                 NA
##                 WeightedAssortativity
## Overall_Network            -0.6129795
## FIV_negative                       NA
## FIV_positive               -0.6623247
```

Network statistics pt 2: Lion network

For the lion network, we define nodes as lions and edges as sharing the same pathogen/parasite. Edges are weighted by the number of shared parasites. The same definitions apply:

- Number of nodes: the number of lions observed. By comparing FIV+ and FIV- networks we can evaluate if the number of lions with FIV is greater than the number of lions without FIV.
- Mean Degree (Weighted): the average number of pathogens per host. FIV negative lions share fewer parasites/pathogens with other lions than FIV positive hosts. This gets more exciting when you compare their distributions!
- Diameter: The maximum number of steps in the set of shortest paths between all lion pairs. This is a measure of connectedness- if you are one lion, how quickly can you get to all the other lions.
- Transitivity: The ratio of the number of triangles and the number of connected triples in the graph. Transitivity ranges from 0 to 1, with higher values indicating more connected subgraphs. This measure shows that both FIV+ and FIV- networks are consistently really highly connected, so just reflects the fact that co-infections are really common.
- Weighted Assortivity: The correlation between a parasites degree and the degree of the parasites connected to it. Assortivity ranges from -1 to 1, with positive values indicating that nodes with high or low numbers of hosts interact mostly with nodes with similar degree values. Here FIV negative lions with high numbers of parasites/pathogens interact mostly wiht lions with few numbers of pathogens.

```
lionnet
```

```
##                  NumNodes MeanDegree MeanDegreeUnWt  VarDegree   Diameter
## Overall_Network       111  306.72072      103.80180 13408.9667 0.01391042
## FIV_negative           23   40.86957       18.86957   208.4822 0.06673407
## FIV_positive           88  206.00000       74.61364  8203.0805 0.02153209
##                  Reciprocity Transitivity VertexConnectivity
## Overall_Network            1    0.9614247                 NA
## FIV_negative               1    0.9238166                 NA
## FIV_positive               1    0.9297232                 NA
##                  WeightedAssortativity
## Overall_Network             -0.9312783
## FIV_negative                -0.9709883
## FIV_positive                        NA
```

Parasite and lion node statistics

Node statistics are properties of each node. For example, node statistics from the lion network tell you about how well connected or not each lion is based on its shared pathogens. Node statistics are great because they can be regressed againsed age, sex, condition. I calcuated the following monopartite statistics and provide definitions based on the lion network:

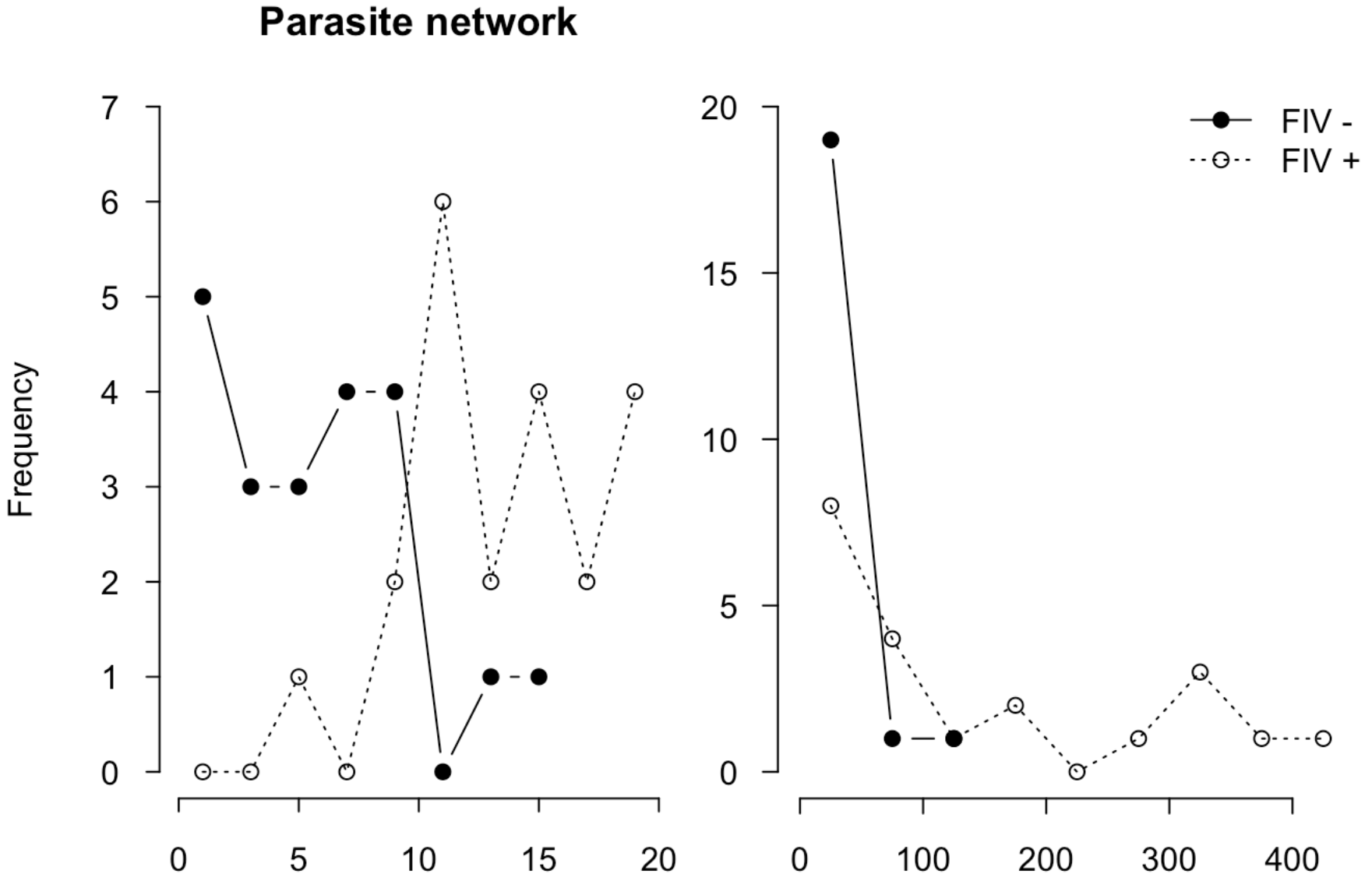
- Weighted Degree: The number of lions connected to the focal lion by sharing a parasite. If lions share 5 parasites, this value is 5.
- Unwieghted Degree: The number of lions connected to the focal lion (unweighted)
- Betweenness: The number of shortest paths between any two lions that go through the focal lion. This is measure of how central the lion is in the network.
- Closeness: The number of steps required to reach every other lion. This is another measure of centrality.
- Transitivity: For the focal lion, this is the ratio of the triangles connected to the vertex and the triples centered on the vertex. I find this confusing and prefer the global measure.
- Average nearest neighbor degree: The average number of lions connected to the immediate neighbors of the focal lion.

Degree distributions:

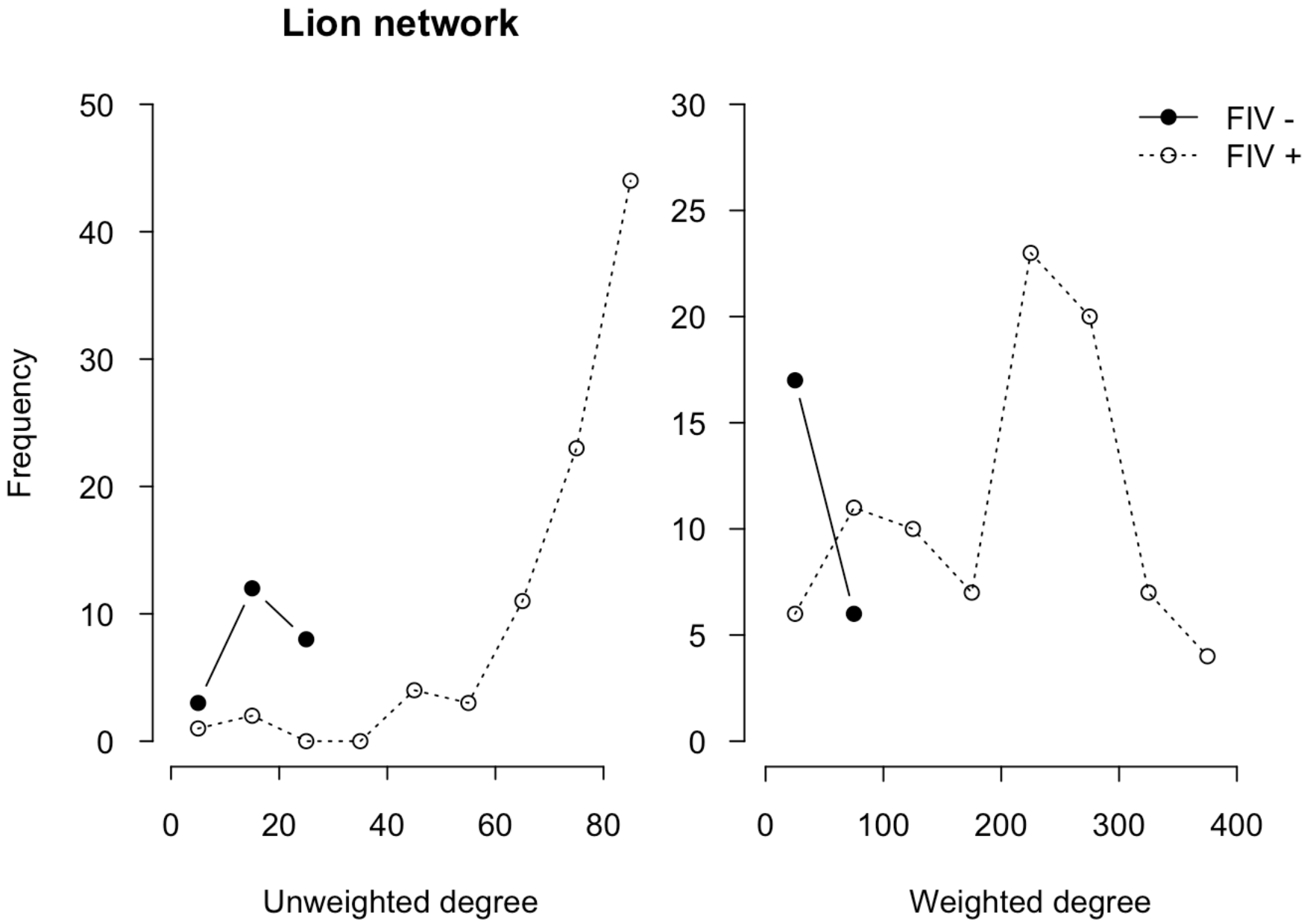
For each parasite, unweighted degree reflects the number of other parasites the focal parasite co-occurs with in at least one hosts. In FIV + lions, the unweighted degree distribution (left) is shifted right, suggesting that the pathogen community is overall more connected.

For each parasite, weighted degree reflects the number of other parasites the focal parasite co-occurs with summed over all hosts. The weighted degree distribution (right) follows similar patterns as the unweighted degree distribution.

This reflects the fact that in FIV + lions, 90% (19/21) of parasites co-occur with ten or more other parasites compared to 10% (2/21) in FIV - lions. (NOTE: This could be because there are fewer FIV - lions but we can bootstrap test that!)



For each lion in the lion network, unweighted degree reflects the number of other lions the focal lion is connected to by sharing any parasite while weighted degree is summed over each parasite. In FIV + lions, ?



Network statistics pt 3: Bipartite network

Visualization

Virus nodes are green, gut parasite nodes are salmon, blue nodes are blood borne pathogens