Using Networks to Study Zoonotic Pathogens

Networks are tools to explore species interactions

Components:

Vertex or node

Some kind of relationship (edges)

Can be directed or undirected. E.g. predator-prey networks use directed relationships, sharing same water sources undirected relationships

Often uses bolded edges to show *weights*. Interactions w/ diff imptance

e.g. France receives less trade from Spain than Spain receives from France

V useful for social dynamics, economics

Basic types are bipartite and unipartite networks

Bipartite: e.g. predator-prey

Unipartite: e.g. all the same species competing over same food resources

Can break up bipartite networks into unipartite “layers”

Can represent with a matrix!

‘Edge list’ matrix.

See igraph package in R

Can force big, over-complicated networks down into unipartite networks that make things clear. Gives example of using a big species interaction network and extracting just the virus interactions from it to make West Nile behavior clear

Usutu virus – flavivirus similar to West Nile

There are a lot of different layouts you can get from the same edge list. Same network, but diff visualizations, depend on what algo you use

‘Nicely’ layout is generally the best. Distributes vertices evenly in the frame, minimizes edge crossing, makes edge links uniform, reflects inherent symmetries, and conforms to the frame of your plotting field.

Contact-based pathogen network was crucial for examining AIDS.

Social interactions b/w people – here, sexual contact

Harder to make for fomites; harder still for density-dependent pathogens like mumps

Stats in networks: **degree, centrality, betweenness**

Degree = how many nodes a pathogen has.

Shows a network of diff viruses and diff bat host species. Rabies has the highest *degree* bc has the most known hosts.

Centrality = how central a node is in a network. Are several metrics

Most rigorous = eigenvector centrality

Proportional to centrality of the whole network

Good for identifying spillover hosts!

Can be skewed by disproportionately studied species, like macaques. Control for sampling and testing bias!

One way: network of shared parasites by hosts, then divide centrality by a measure of research effort.

Components = betweenness

Good for finding amplifying hosts

e.g. immunocompromised hosts tend to shed more virus, as do pregnant or nursing

Network of MERS coronvavirus from when it spread to South Korea in 2014.

2 superspreaders in hospital – high betweenness in the network!

Finding hosts that act as superspreaders, by looking for high betweenness, is v helpful for figuring out where pathogen will have opportunity to spill over into humans

Shows a host unipartite network from Ecohealth’s PREDICT database.

Domestic animals tend to have high centrality! (partly artifact of being more studied)

Many viruses shared b/w many species

Domestic animals, rodents, bats, and primates have high centrality

Virus interface bipartite network

Good for IDing risk areas

**Overall network structure metrics** – stats of network as a whole, not of individual nodes

Modularity

Barber’s bipartite value is one metric. Goes from 0 to 1. 0 = everything connected at random, 1 = no connections.

Bat network at 0.68 – there’s some modularity, certain nodes associate with each other distinct from the rest of the network

In this case, it’s bc of geo isolation and unique lifestyle – large Australian fruit bats

There are generalist viruses like rabies, and specialists that are often too small to see when color the map by module

Network metrics are not static! Totally dependent on the layout and algorithm you use to make your network. So, iterate your network a lot to figure out how static these values are.

Above, pretty consistent 0.5-0.65 range for that modularity value whenever make the network, and pretty consistently get ~45 species in that module

Simulating bird flu in farms

Depending on where an outbreak starts, how much of the farming community will it prop thru?

Certain network structures are more vulnerable!

Can amplify quite badly from factory farms, less so from backyard farms UNLESS reaches a factory farm

**Future horizons:**

Bigger data. Shows a Shiny app – github.com/abhik1368/Shiny\_NetPredictor

Research effort is biased and incomplete. Use algos and sims to figure out what we’re missing from our surveillance efforts?

**Questions**

First thing to do with a metric is count number of nodes and vertices

Structure metrics other than modularity

(Density?) – metric of how many possible edges can exist in the network vs how many *do* exist in the network

Can use this to build a null model to see if you got the network structure you see by random chance

**Null hyp testing is crucial in network work!**

Nestedness – are nodes off on their own branches, or is there a parent node that gives rise to them?

It’s not uncommon for viruses to not group with rest of their genus when you look at a virus/ host network.

For birds, flock structure can be v impt. And v diff from domestic than for wild birds – worth separating them out rather than making a birds-in-general network

Often, there are viruses that exploit diff body areas and so don’t compete w/ each other much. Viruses that target same tissues in host tend to compete.

Underresearched species? Many. There’s strong surveillance in wealth countries (cf West Nile and Usutu), but of course there are *more* in tropical species-rich areas.

Fruit bats are well-studied cause they’re big and easy to catch. (Can take more blood from a larger bat! Can miss viruses in small bats more easily, bc can’t take as much blood if you want to do catch-and-release. Plus they’re harder to find).

Charismatic species get more research, particularly bigger species. Gorillas are v well-studied. Nocturnal species, solitary species like pangolin, less so.

Need to do strategic surveillance to study less-studied species.

Can use networks to study climate change impacts?

Can def model some temporal aspects, but haven’t seen them used for climate change. If could make a predictive algo to change the weight for a virus being prevalent or not as climate changes (e.g. greater range or more vectors bc of climate change), abundance will go up, new hosts will come into contact, network might become more connected!

Networks have been helpful for studying effects of urbanization. Urbanization and air travel have made networks more connected. Cf HIV and West Nile.

How to handle reporting bias?

Count number of publications; e.g. 2,000 on rabies, 2 on Ntembe bat virus. Use that publication bias as a proxy for testing.

Idealling you want actual testing bias. Include even negatives in the reporting – where species have been tested but not found positive. But that’s hard.