

Uncover viral sharing through global structure of host-virus meta-network

Mathis Gheno^{1,2,‡}, [Timothée Poisot](#)^{2,3,‡}

¹ Université de Rennes; ² Université de Montréal; ³ Québec Centre for Biodiversity Sciences

[‡] Equal contributions

Correspondance to:

Timothée Poisot — timothee.poisot@umontreal.ca

Purpose: This template provides a series of scripts to render a markdown document into an interactive website and a series of PDFs.

Internals: GitHub actions and a series of python scripts. The markdown is handled with pandoc.

Motivation: It makes collaborating on text with GitHub easier, and means that we never need to think about the output.

Keywords:

pandoc
pandoc-crossref
github actions

0.1. Introduction

1. Global health needs more ecology in viral forecasting

Predicting viral spillover for future pandemic prevention + importance of ecology in the context of global health ([Belay et al. 2017](#); [Carroll et al. 2018](#)) ([Plowright et al. 2017](#)). Works have been done to predict viral hotspot according to climate change ([Carlson et al. 2022](#)), future host distribution ([Morales-Castilla et al. 2021](#)) and potential viral sharing among host ([Albery et al. 2020](#)). Bring information on potential spillover path.

2. Communicability already used in other field and context to estimate spread.

0.2. Methods Communicability quantify how well information transit between two nodes by considering all possible path in a network and penalizing longer ones. It is compute with the exponent of the adjacency matrix of the network.

$$G = \sum_{k=0}^{\infty} \frac{(A^k)}{k!} = e^A$$

where G is the communicability matrix, A the adjacency matrix and k is used as a penalizing term. It is possible to compute the exponential of a matrix with the graph spectrum :

$$G = \sum_{j=1}^n \varphi_j \varphi_j^T e^{\lambda_j}$$

where φ_j and λ_j are respectively the j^{th} eigenvectors and eigenvalues of the matrix A . Obtain cluster with Communicability is done only by removing the first dimension in the computation of the previous equation.

$$\Delta G = \sum_{j=1}^n \varphi_j \varphi_j^T e^{\lambda_j} - \varphi_1 \varphi_1^T e^{\lambda_1}$$

In this equation we are removing to global sum the dot production of the first eigenvector (and eigenvalue) to find the clustering matrix ΔG . Let's keep in mind this final result and explain why this work for clustering.

We start with the spectral form of G that can be decompose by the following way :

$$G = \varphi_1 \varphi_1^T e^{\lambda_1} + \sum_{j=2}^n \varphi_j^+ \varphi_j^{+T} e^{\lambda_j} + \sum_{j=2}^n \varphi_j^- \varphi_j^{-T} e^{\lambda_j} + \sum_{j=2}^n \varphi_j^- \varphi_j^{+T} e^{\lambda_j} \quad (1)$$

where φ_j^+ or φ_j^- indicate respectively all the positives or negatives values of the j^{th} eigenvector. A way to think about it would be that when φ_j^+ all negative values are set to 0 and when φ_j^- all positive values are set to 0. Estrada & Hatano (2008) explain that “two nodes have the same sign in an eigenvector if they can be considered as being in the same partition of the network, while those pairs having different signs correspond to nodes in different partitions.”.

$$\sum_{j=2}^{intracluster} \varphi_j \varphi_j^T e^{\lambda_j} = \sum_{j=2}^n \varphi_j^+ \varphi_j^{+T} e^{\lambda_j} + \sum_{j=2}^n \varphi_j^- \varphi_j^{-T} e^{\lambda_j}$$

and

$$\sum_{j=2}^{intercluster} \varphi_j \varphi_j^T e^{\lambda_j} = \sum_{j=2}^n \varphi_j^- \varphi_j^{+T} e^{\lambda_j}$$

so the clustering matrix is obtain with

$$\Delta G = \sum_{j=2}^{intracluster} \varphi_j \varphi_j^T e^{\lambda_j} - \left| \sum_{j=2}^{intercluster} \varphi_j \varphi_j^T e^{\lambda_j} \right|$$

The absolute operator is not useful, it is just here to remind that all inter cluster values are negatives. The dot product of $\varphi_j \varphi_j^T$ produce a matrix with positive and negative sign depending on the sign of $\varphi_j(p)$ and $\varphi_j(q)$. The first is not include because all the values the eigenvector the same sign, so it is not really informative (In fact it consider the hole network as one cluster thus it does not bring interesting information on the clustering).

In short we can rewrite ΔG as follow :

$$\Delta G = G - \varphi_1 \varphi_1^T e^{\lambda_1}$$

which was the form in which we introduce it.

0.2.0.0.1 Remark 1 Each dimension of the spectra form the 2nd to the last one is a clustering configuration of the network. Cluster identified by dim 3 are not necessarily independent form those in dim 2 or 4 (or all others). Although the cluster form by dim 2 are “stronger” then 3, 4, 5 ... and so on till the last one. So the dim 2 is the one that contribute the most for ΔG (we can see it in the following example).

0.2.0.0.2 Remark 2 In a graph nodes can be consider as diffuser or receiver. The communicability matrix can be use to quantify the capacity of nodes to communicate when they act like a diffuser a receiver. For example host-host or virus-virus parts of G give information on the capacity of respectively host or virus to communicate when they are playing the same role in he network. see Benzi & Boito (2020) for better explanation

```

library(tidyverse)
library(lattice)
library(igraph)
library(colorRamps)
A = matrix(c(0,1,0,1,1,0,0,0,0,0,0,
            1,0,1,1,1,0,0,0,0,0,0,
            0,1,0,1,1,0,0,0,0,0,0,
            1,1,1,0,1,0,0,0,0,0,0,
            1,1,1,1,0,1,0,0,0,0,0,
            0,0,0,0,1,0,1,0,0,0,0,
            0,0,0,0,0,1,0,1,1,0,1,
            0,0,0,0,0,0,1,0,1,1,1,
            0,0,0,0,0,0,0,1,0,1,1,
            0,0,0,0,0,0,0,1,1,0,1,
            0,0,0,0,0,0,1,1,1,0,0), nrow =11, ncol =11)
grap = graph_from_adjacency_matrix(A, mode = "undirected")
plot(grap)

```

0.2.0.1 Example A graph with 11 nodes and 2 distinct group. First we need to compute the graph spectrum

```

spectra = eigen(A)
levelplot(spectra$vectors, ylab = "j th position" , xlab = "eigenvectors")

```

The above plot just represent the 11 eigenvectors. We can see that the first one is the only full of same sign value.

Now let's take the 2nd dimension as an example.

```

##
G_dim2 = spectra$vectors[,2]%*%t(spectra$vectors[,2])*exp(spectra$values[2])
levelplot(G_dim2, ylab = "node", xlab = "node", col.regions = rev(matlab.like(16)))

```

And that it ! The second dimension of the graph communicability identify 2 cluster (in blue).

We can compute for the third dimension

```

G_dim3 = spectra$vectors[,3]%*%t(spectra$vectors[,3])*exp(spectra$values[3])
levelplot(G_dim3, ylab = "node", xlab = "node", col.regions = rev(matlab.like(16)))

```

Which identify clusters between 5:6 and 6:7. The cluster of the third dimension are less “obvious” than those from the second dimension

Now if we want to use both dimension in the clustering, we just have to add

```

levelplot(G_dim2+G_dim3, ylab = "node", xlab = "node", col.regions = rev(matlab.like(16)))

```

We could continue like that till the last dimension (11th), but it was for the explanation. So now we can compute directly ΔG by adding dimension from 2 to 11 (or subtraction of the first dim which is exactly the same)

```

delta_G = matrix(0, nrow =11, ncol =11)
for(dim in 2:11){
  delta_G = delta_G + spectra$vectors[,dim]%*%t(spectra$vectors[,dim])*exp(spectra$values[dim])
}
levelplot(delta_G, ylab = "node", xlab = "node", col.regions = rev(matlab.like(16)))

```

0.3. Results

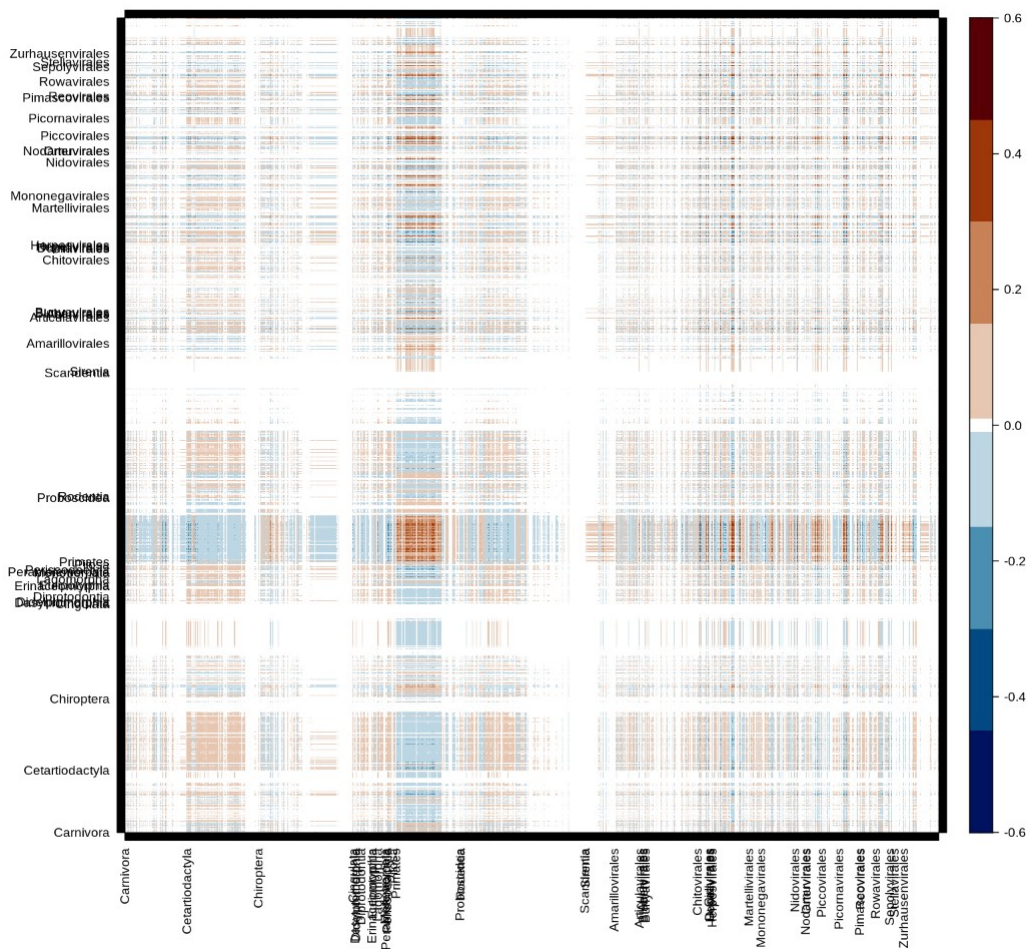


Figure 1 Figure 1 : Global matrix of clustering communicability. Positive values indicate species in same cluster, negative value species in “opposite” cluster. Only host order and virus order names are display on x and y.

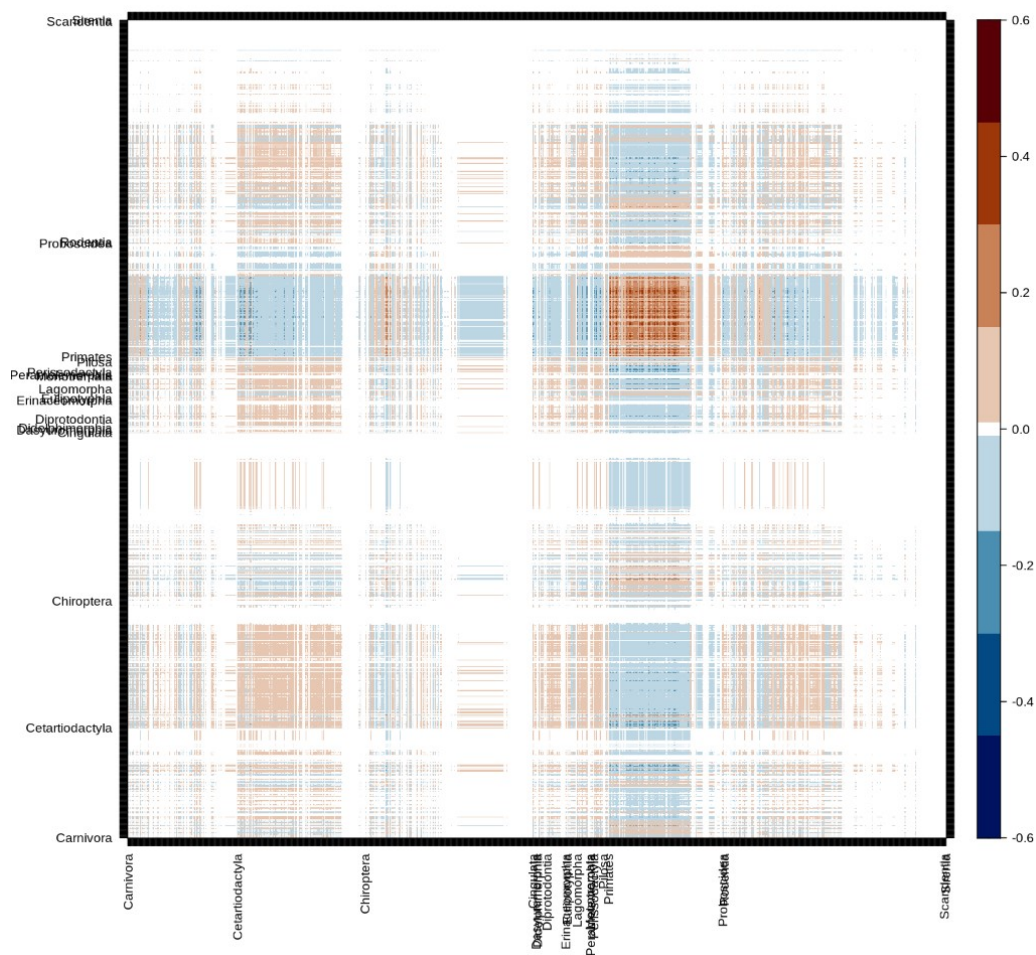


Figure 2 Figure 2 : Host-Host cluster matrix

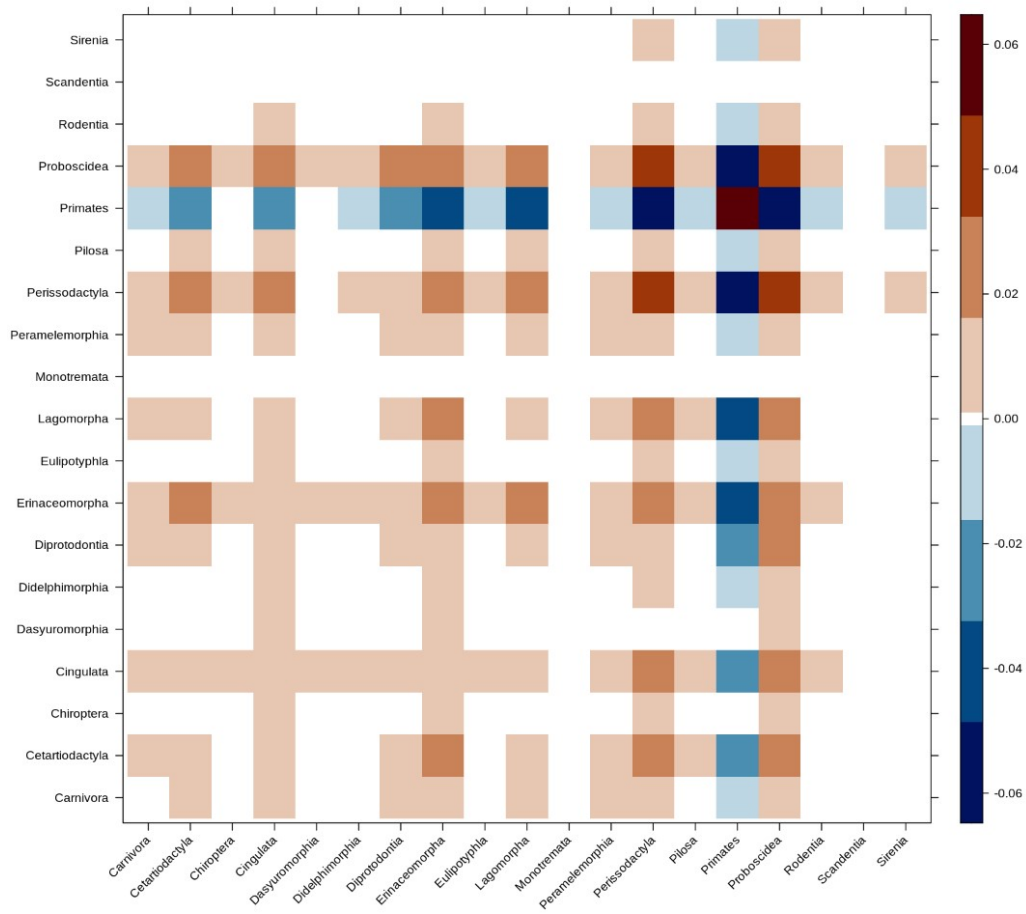


Figure 3 Figure 3 : Host Host recap (with median of clustering value for each order)

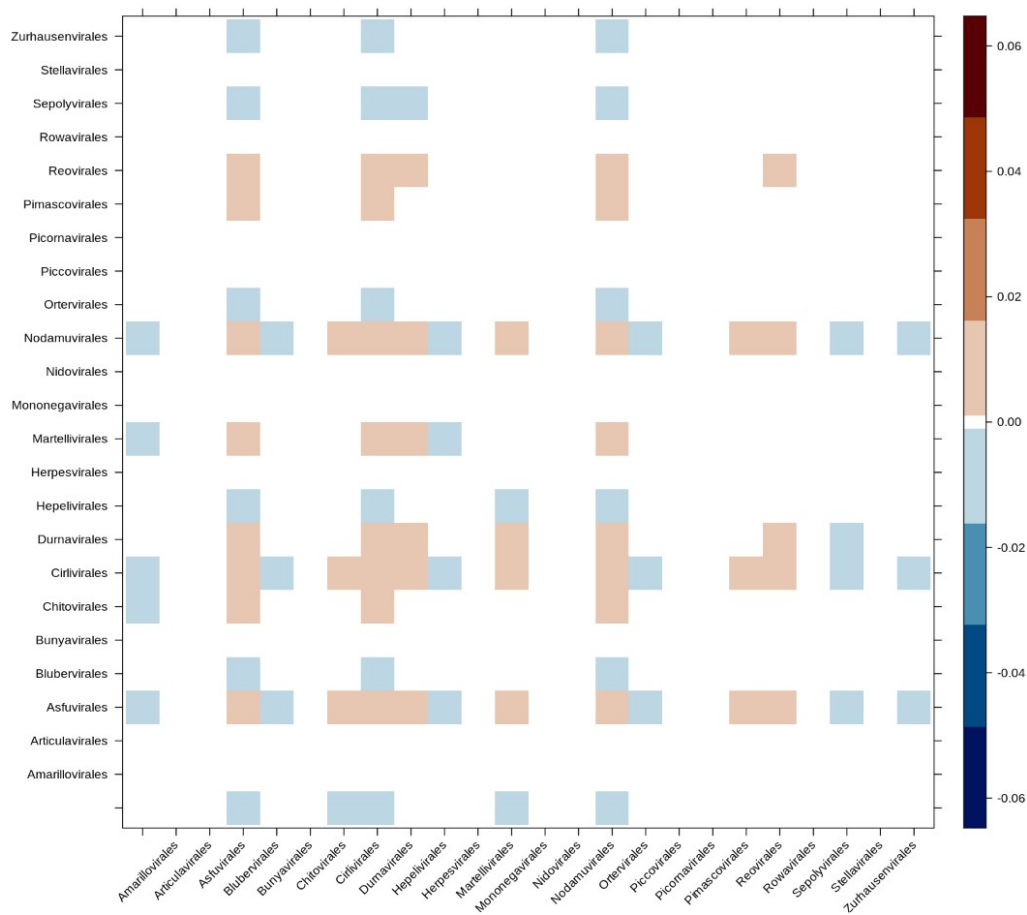


Figure 5 Figure 5 : Virus-Virus recap (with median of cluster value for each order)

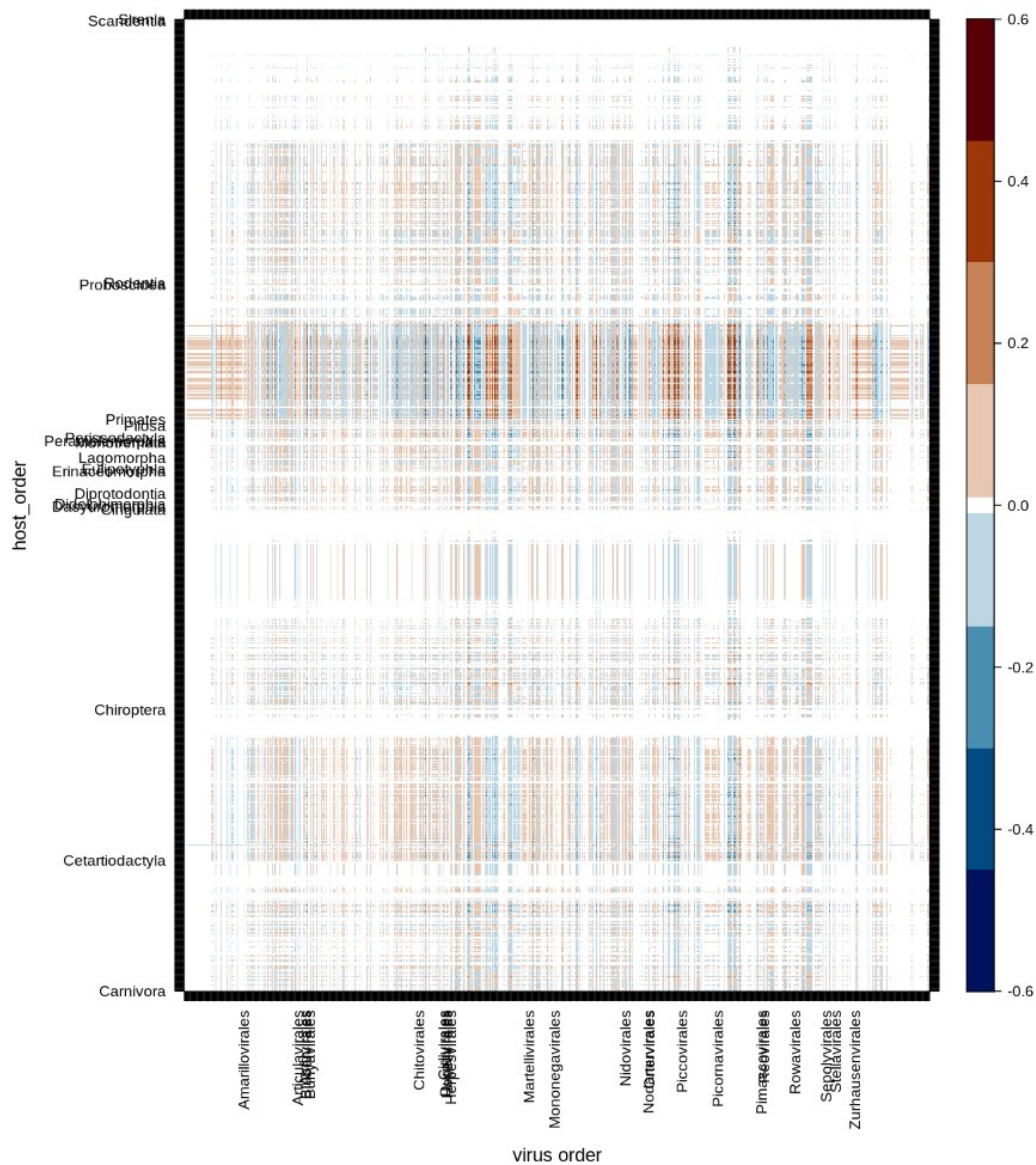


Figure 6 Figure 6 : Host-Virus cluster matrix

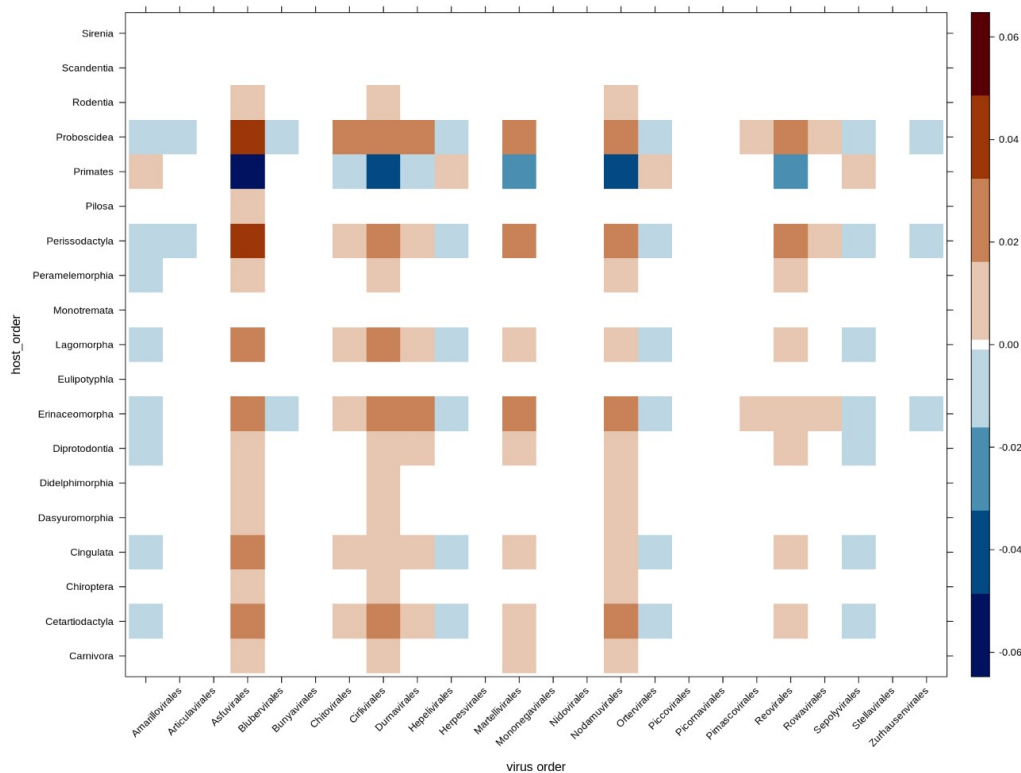


Figure 7 Figure 7 : Host-Virus recap (with median of cluster value for each order)

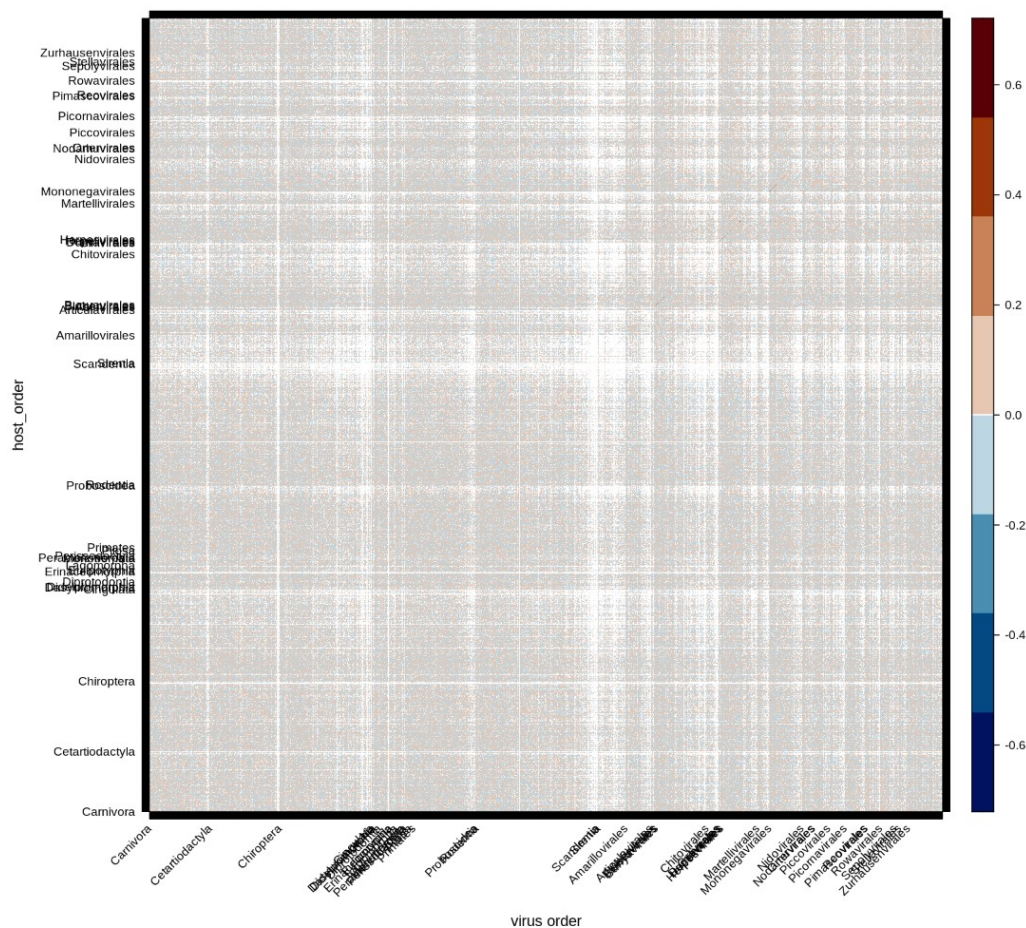


Figure 8 Figure 8 : Global matrix of clustering communicability, with random edges connection. Number of edges for each nodes (Host and virus) is conserved. X and Y are ordered the same way as fig. 1

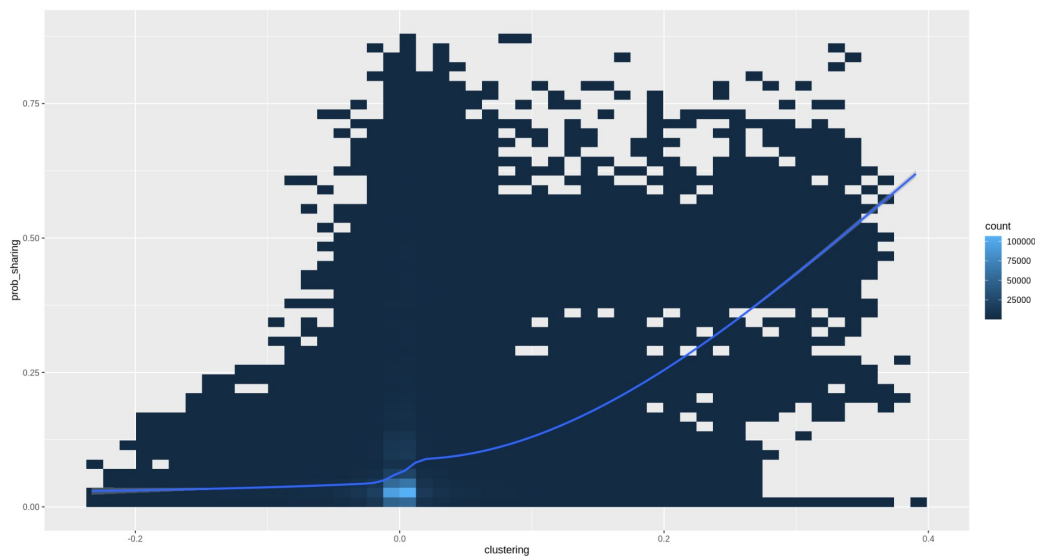


Figure 9 Figure 9 : comparison with Albery sharing metric.

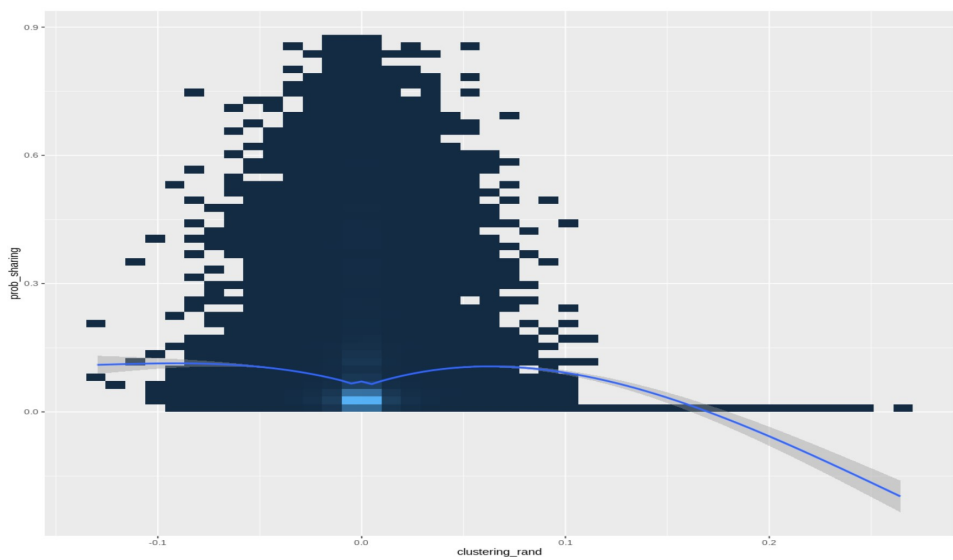


Figure 10 Figure 10 : comparison with Albery sharing metric for a random network

- Albery, G.F., Eskew, E.A., Ross, N. & Olival, K.J. (2020). [Predicting the global mammalian viral sharing network using phylogeography](#). *Nature Communications*, 11, 2260.
- Belay, E.D., Kile, J.C., Hall, A.J., Barton-Behravesh, C., Parsons, M.B., Salyer, S., *et al.* (2017). [Zoonotic Disease Programs for Enhancing Global Health Security](#). *Emerging Infectious Diseases*, 23, S65–70.
- Benzi, M. & Boito, P. (2020). [Matrix functions in network analysis](#). *GAMM-Mitteilungen*, 43, e202000012.
- Carlson, C.J., Albery, G.F., Merow, C., Trisos, C.H., Zipfel, C.M., Eskew, E.A., *et al.* (2022). [Climate change increases cross-species viral transmission risk](#). *Nature*, 607, 555–562.
- Carroll, D., Daszak, P., Wolfe, N.D., Gao, G.F., Morel, C.M., Morzaria, S., *et al.* (2018). [The Global Virome Project](#). *Science*, 359, 872–874.
- Estrada, E. & Hatano, N. (2008). [Communicability in complex networks](#). *Physical Review E*, 77, 036111.
- Morales-Castilla, I., Pappalardo, P., Farrell, M.J., Aguirre, A.A., Huang, S., Gehman, A.-L.M., *et al.* (2021). [Forecasting parasite sharing under climate change](#). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 376, 20200360.
- Plowright, R.K., Parrish, C.R., McCallum, H., Hudson, P.J., Ko, A.I., Graham, A.L., *et al.* (2017). [Pathways to zoonotic spillover](#). *Nature Reviews Microbiology*, 15, 502–510.

Conclusion