

Predicting Missing Virus-Host Links for Avian Hosts using Graph Neural Networks

Akshath Shvetang Anna, Grace Driskill, Julia Qian

Georgia Institute of Technology
CSE 8803-EPI Fall 2024

Introduction

Zoonotic diseases are infectious illnesses transmitted from animals to humans. Birds can act as hosts to viruses, spreading the disease to humans or greatly increasing the geographic range of a disease.

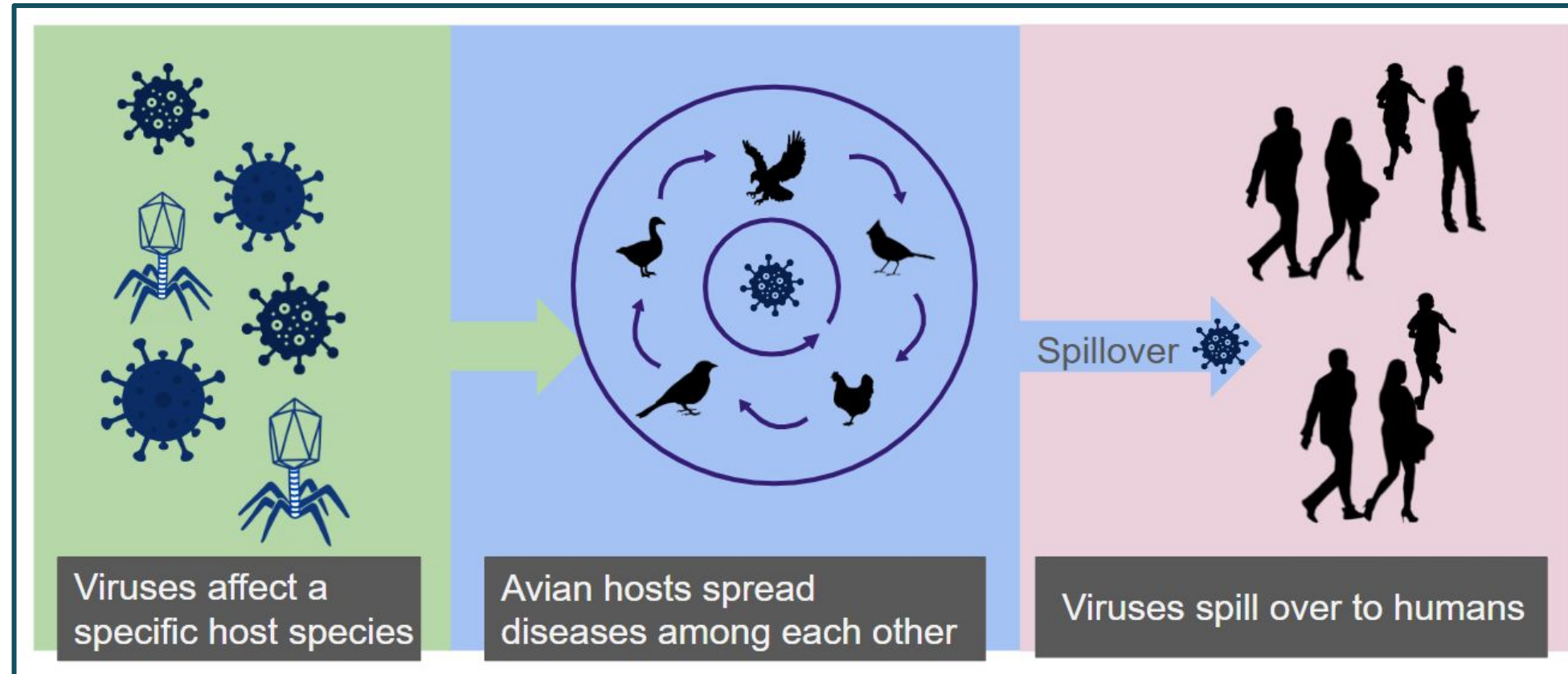


Figure 1: Spillover events.

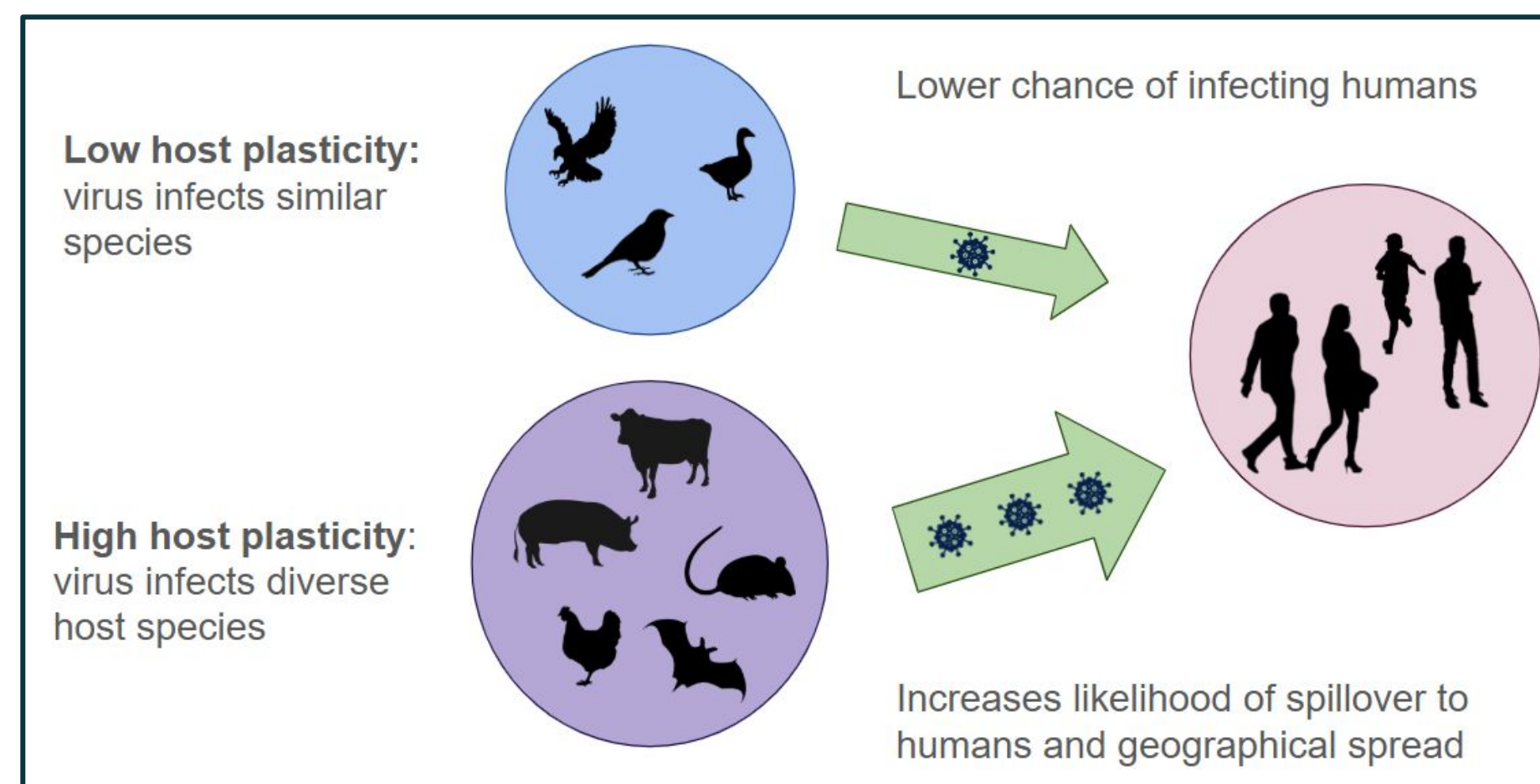


Figure 2: Host plasticity affects chance of spillover.

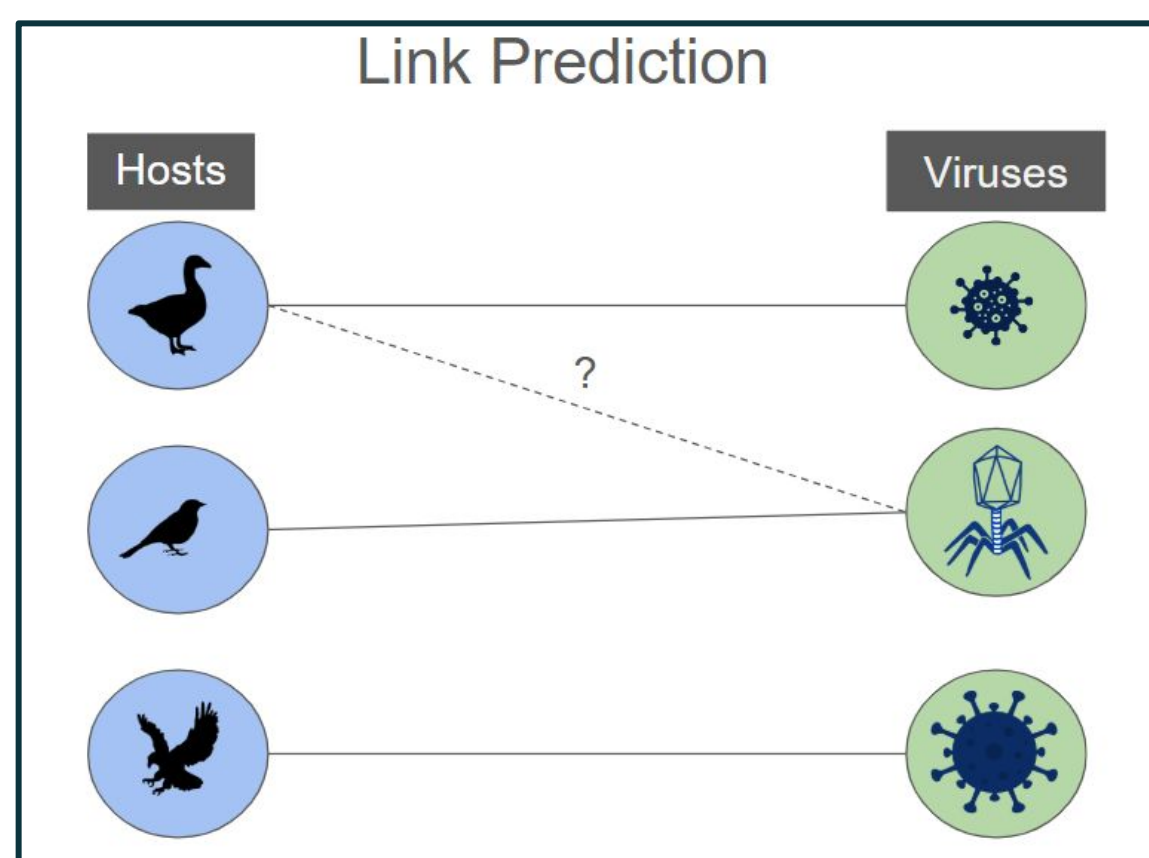


Figure 3: Host-virus interactions as a link-prediction problem

Host-virus interactions can be represented through biological networks. Graph Neural Networks can be applied to predict undiscovered links between hosts and viruses.

Our project predicts previously unknown relationships between bird hosts and known viruses.

Data

Feature	Description
Species	Species name
NCBI Taxon ID	Taxid as given by NCBI Taxonomy, used as ID for virus
Family	Virus family
Genus	Virus Genus
Envelope	1 if virus has an envelope, 0 if not
Circular	1 if the viral genome is circular, 0 if it is linear
Double Stranded	1 if the genome is double-stranded, 0 if not
RNA	1 if the viral genome is RNA, 0 if it is DNA
Segmented	1 if the genome is segmented, 0 if it is monopartite
Positive Sense	1 if the RNA is positive-sense, 0 otherwise or if genome is DNA
Negative Sense	1 if the RNA is negative-sense, 0 otherwise or if genome is DNA
Size	Size of genome in KB
GC	Guanine-Cytosine content of genome
Genes	Number of genes
Budding	Virus is released through budding
Lysis	Virus is released through lysis
Release Other	Virus is released through other means or is unspecified
Cytoplasm	1 if gene transcription occurs in the cytoplasm, 0 if it occurs in the nucleus
CE Clathrin	Viral entry through clathrin-mediated endocytosis
CE Receptor	Viral entry using receptors
CE Glycoprotein	Viral entry using glycoproteins
CE Other	Viral entry through other or unspecified means

Table 1: Virus traits.

A set of known virus-host interactions, along with virus and avian traits were gathered from multiple sources.

Feature	Description
Scientific Name	Host's scientific name
NCBI Taxon ID	Taxid as given by NCBI Taxonomy, used as ID for host
Family	Host family
Genus	Host Genus
Mass	Species average body mass in grams
Trophic Level	The position the species occupies in a food web. Herbivore, Carnivore, Scavenger or Omnivore
Trophic Niche	A more specific categorization of the species' diet. Frugivore, Granivore, Vertivore, Aquatic Predator, etc.
Primary Lifestyle	Where the species spends the majority of its time. Aerial, Terrestrial, Insectorial, Aquatic, or Generalist
Range Size	The total area of the mapped range of the species in square kilometers
Habitat	The environment of the species. Desert, Grassland, Forest, Human modified, Marine, etc.
Habitat Density	The species density of this species' habitat. Discrete measure 1 to 3.
Migration	Categorizes this species' migration habits as 1 = Sedentary, 2= Partially migratory or 3 = Migratory
Evolutionary Distinctiveness	Measures how isolated a species is on a phylogenetic tree

Table 2: Avian traits.

Model

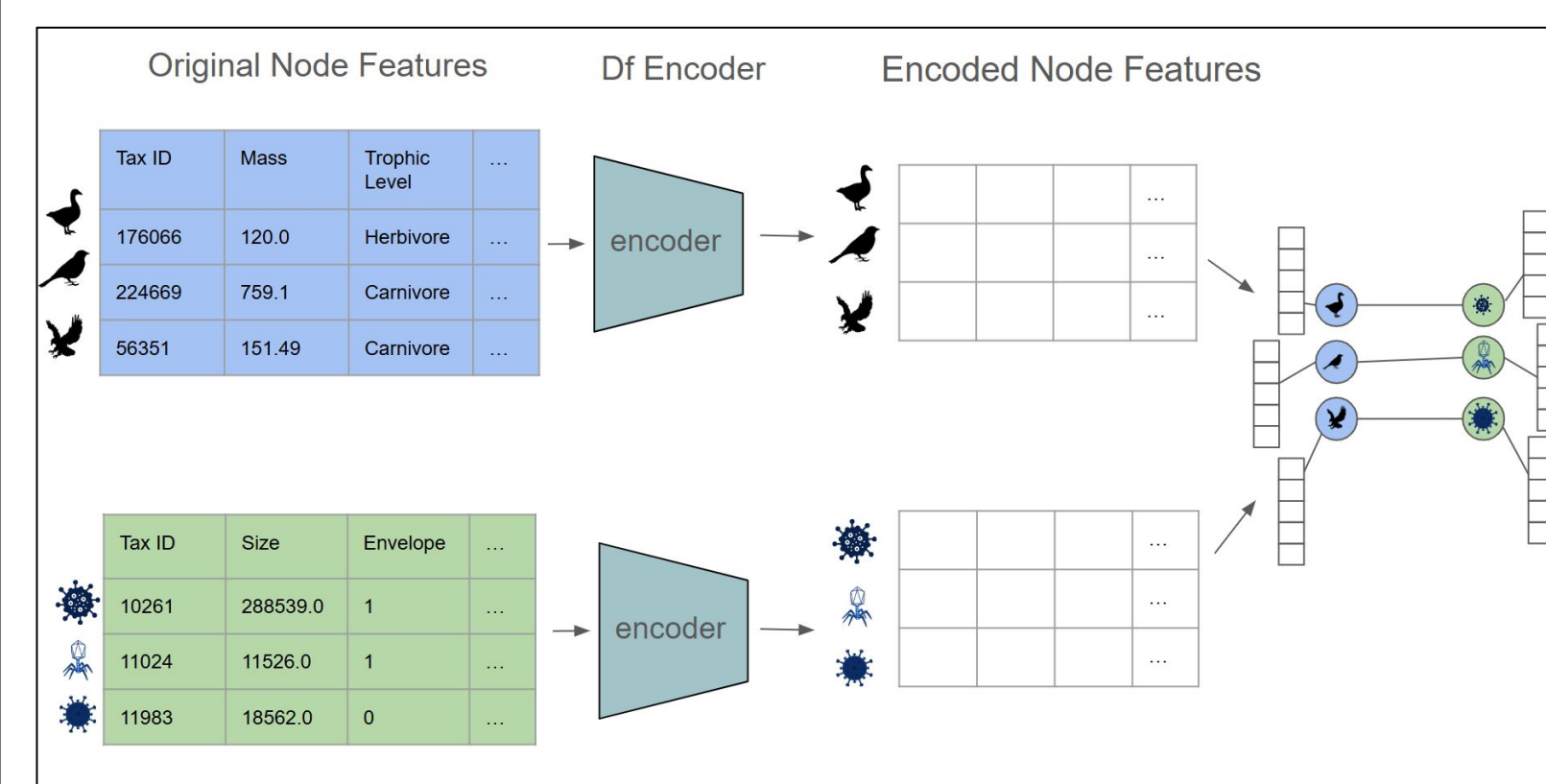


Figure 4: Data encoding.

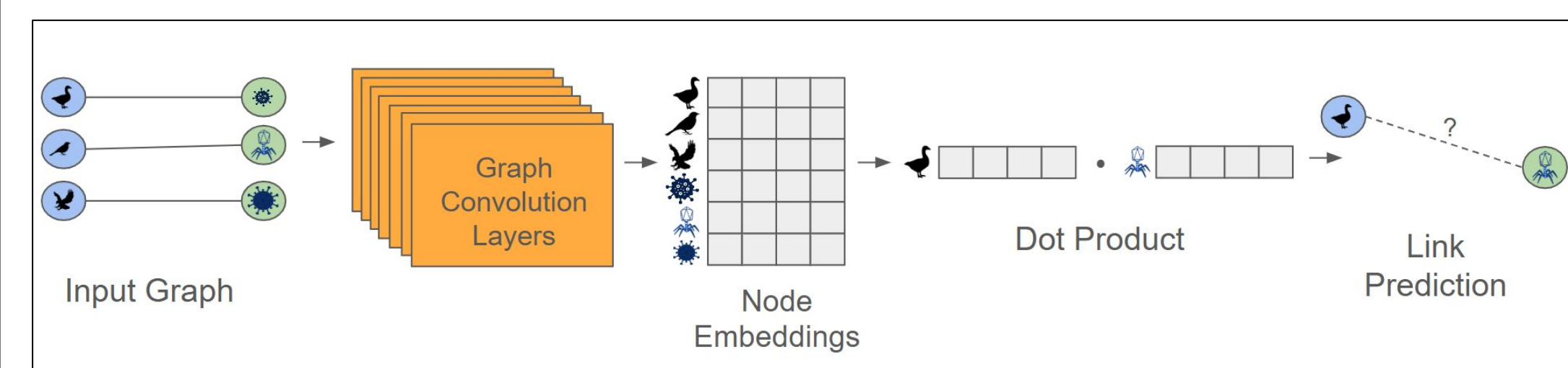


Figure 4: GNN model for link prediction.

A bipartite graph is constructed with autoencoded trait features. This is then used as input to the GNN to learn node embeddings for predicting links.

Results

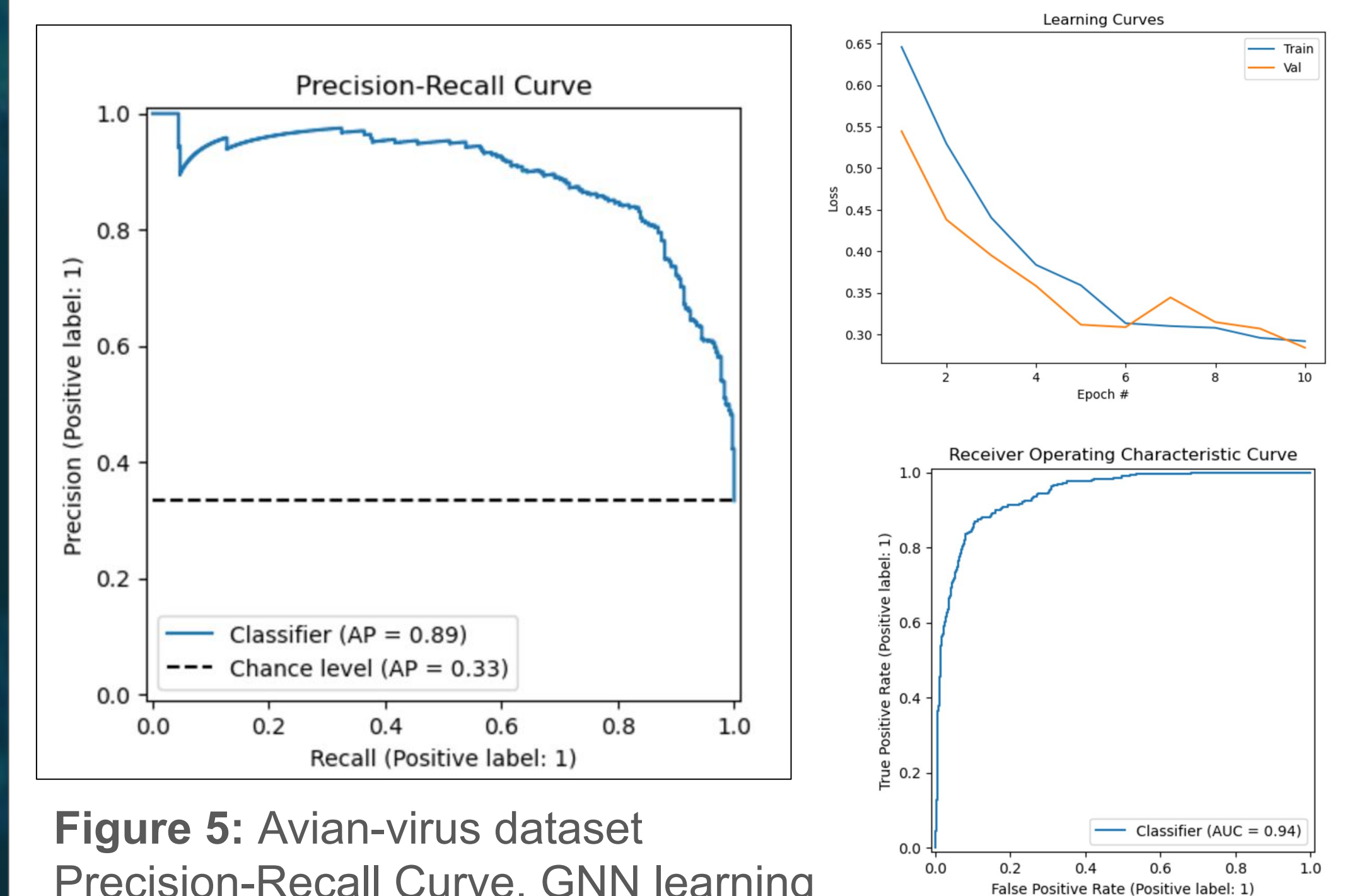


Figure 5: Avian-virus dataset Precision-Recall Curve, GNN learning curve, and ROC Curve



Figure 6: Myiopsitta monachus (Monk parakeet)

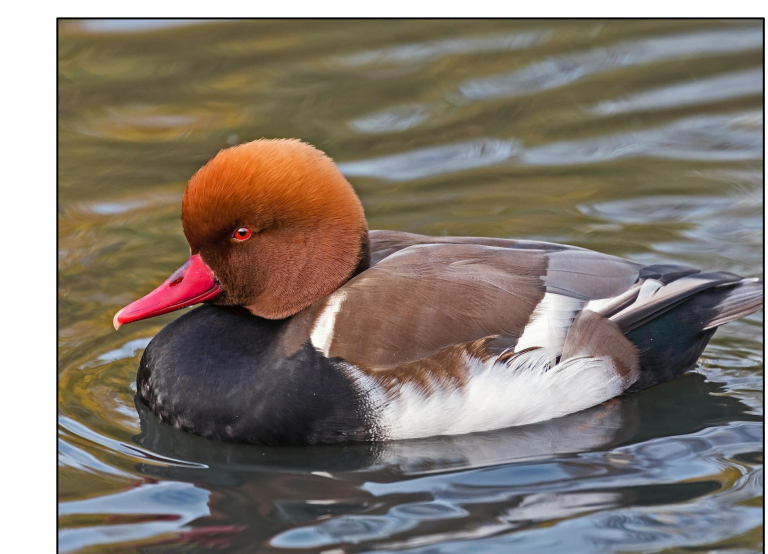


Figure 7: Netta rufina (Red-crested pochard)

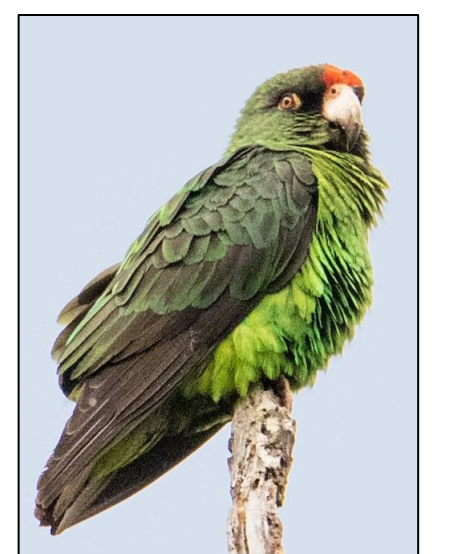


Figure 8: Poicephalus gulelmi (Red-fronted parrot)

Edge Score	Virus (taxid)	Host (taxid)	Avg Distance to known Hosts	Avg Distance to Other Species
0.9731	Avipoxvirus isolate PM9 (468380)	Myiopsitta monachus (176066)	159.33	156.40
0.9726	Fowl adenovirus HR2 (911324)	Netta rufina (30387)	161.44	185.58
0.9724	Chicken parvovirus 399/HRV/2010 (1081806)	Poicephalus gulelmi (241588)	207.11	156.79
0.9720	Avipoxvirus isolate Pennsylvania (1294111)	Diomedea sanfordi (2811340)	169.90	171.41
0.9719	Melegrivirus A (1330070)	Gracula religiosa (116992)	207.11	150.39

Table 3: Top 5 predicted previously unobserved virus-host interactions



Figure 9: Diomedea sanfordi (Northern royal albatross)



Figure 10: Gracula religiosa (Common hill myna)