Who Eats Whom: Unraveling Ecological Patterns

Presented by Lavanya Middha & Surabhi Ravindran Nair

Project Goal:

> **Analyze Feeding Interactions**
>
> Uncover Ecological patterns across species and help ecologists make more data-driven informed decisions.

Data Exploration & Pre-Processing

Data Source: iNaturalist

Extracted observational feeding interaction data.

- Each interaction: a pair of linked observations.
  - One from predator → prey
  - One from prey → predator, connected by a partner observation ID.

Data Cleaning

From 7,661 research grade observations, 2013 problematic entries were identified:

- 1,950 empty values.
- Invalid formats & unclear taxon names.
- All NA values were removed,

Kept only expert-verified entries (quality_grade = "research").

Feature Engineering

```
df1.columns
```

```
Index(['id', 'uuid', 'observed_on', 'time_observed_at', 'quality_grade', 'url',
       'image_url', 'description', 'place_guess', 'latitude', 'longitude',
       'place_town_name', 'place_county_name', 'place_state_name',
       'place_country_name', 'scientific_name', 'common_name',
       'iconic_taxon_name', 'taxon_id', 'taxon_kingdom_name',
       'field:id meant for "eater" or organism being eaten?',
       'field:url for "partner" observation',
       'field:is observation one of these special types of feeding?'],
      dtype='object')
```

We identified the most important attributes relevant to the dataset that can be modeled in Neo4j

Reduced number of columns from 71 to 23.

Data Structuring

1

Cleaned Dataframe

Each row represents one valid predator-prey interaction.

2

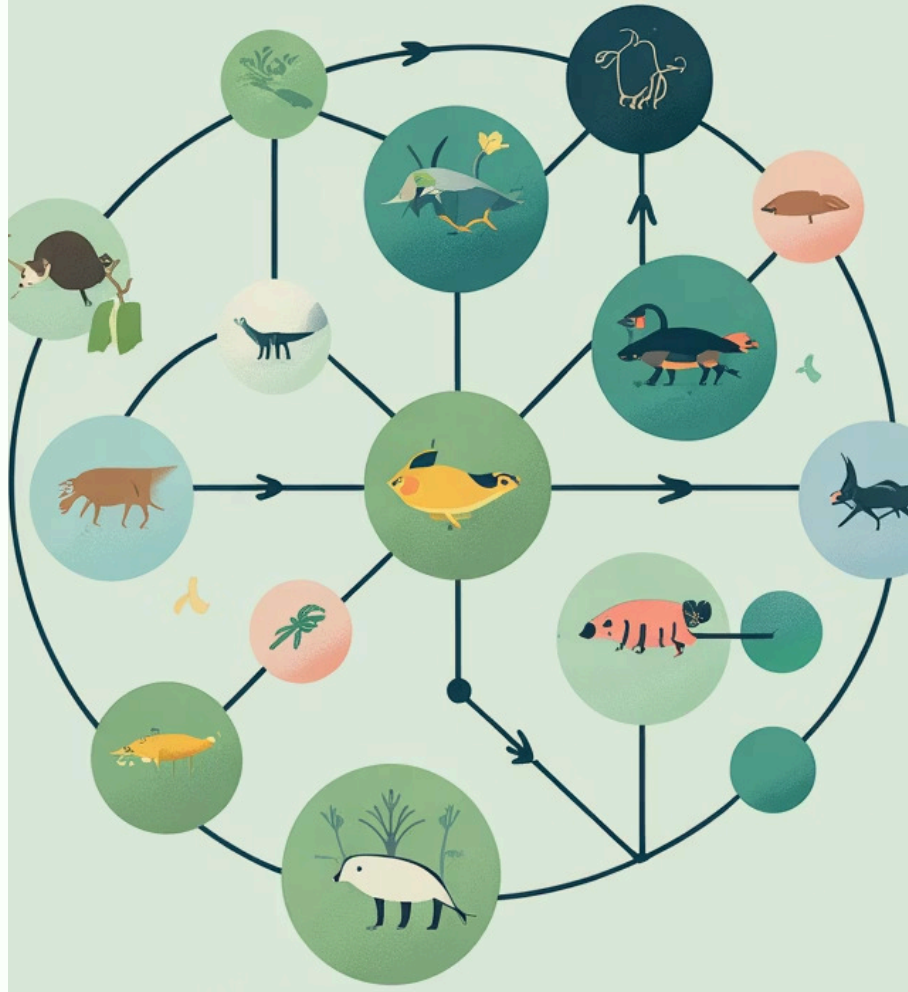
Created a data dictionary

To validate Taxa and observation fields

3

Split Dataset

Taxa dictionary (scientific name, common name, iconic taxon, taxon ID) and EATS (predator, prey, location, feeding type).



Graph Modeling

1

Neo4j Preparation

Completed foundational courses offered by Neo4j and other online resources.

2

Graph Construction

Imported Taxa nodes and EATEN_BY relationships.

3

Added necessary libraries

Installed Graph Data Science & APOC NEO4j specific libraries.

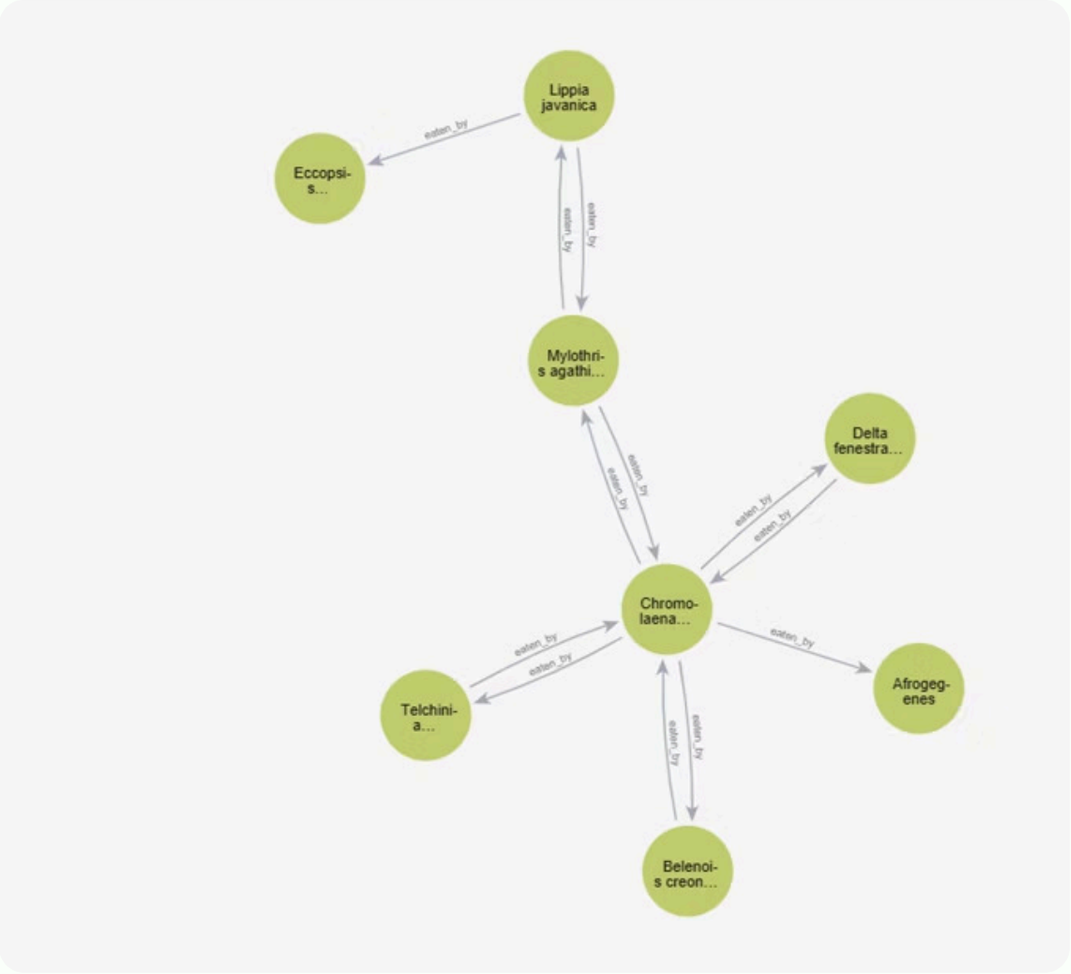
Key Findings

Number of Unique Species	5315, but only 3320 are research grade
Number of Relationships	2890 unique relationships
Country with highest observations	United States (followed by South Africa)

Preliminary Analysis: Key Findings

Strongly Connected Cycles

Visual graph layouts revealed large strongly connected components, suggesting circular or misreported feeding links.



Max Length : 6

Eccopsis incultana -> Lippia javanica -> Mylothris agathina agathina -> Chromolaena odorata -> Belenois creona severina -> Cynanchum viminale

Preliminary Analysis: Key Findings

Critical Species: Western Honey Bee

Apis mellifera (betweenness 649.5) acts as a major connector in the network.

Predator with Most Unique Prey

Ardea herodias (Great Blue Heron).

Top five:

common_name	scientific_name	preyCount
Great Blue Heron	Ardea herodias	47
Western Honey Bee	Apis mellifera	41
Tan Jumping Spider	Platycryptus undatus	20
Southern Brown-hooded Kingfisher	Halcyon albiventris albiventris	20
Osprey	Pandion haliaetus	18

Preliminary Analysis: Key Findings

Most Eaten Prey

Lepidoptera (Butterflies & Moths) - eaten by 38 predators.

Top Five

common_name	scientific_name	predatorCount
Butterflies and Moths	Lepidoptera	38
Ray-finned Fishes	Actinopterygii	21
Common Wild Fig	Ficus burkei	21
Yellow Justicia	Justicia flava	15
Winged and Once-winged Insects	Pterygota	15

Isolated Species

Species with only self-links, indicating observation noise or taxonomy issues.

common_name	scientific_name
Common Kingfisher	Alcedo atthis
Brush Rabbit	Sylvilagus bachmani
Northern highbush blueberry	Vaccinium corymbosum
Napoleon Spider	Synema globosum
Adanson's House Jumper	Hasarius adansoni
Highveld Lesser-Thicktail Scorpion	Uroplectes triangulifer

Identified Data Issues & Discrepancies

Biological Inconsistencies

- Incorrect labels (e.g., butterflies as "bottom feeders")
- Reversed predator/prey
- Plant-eating misclassified.

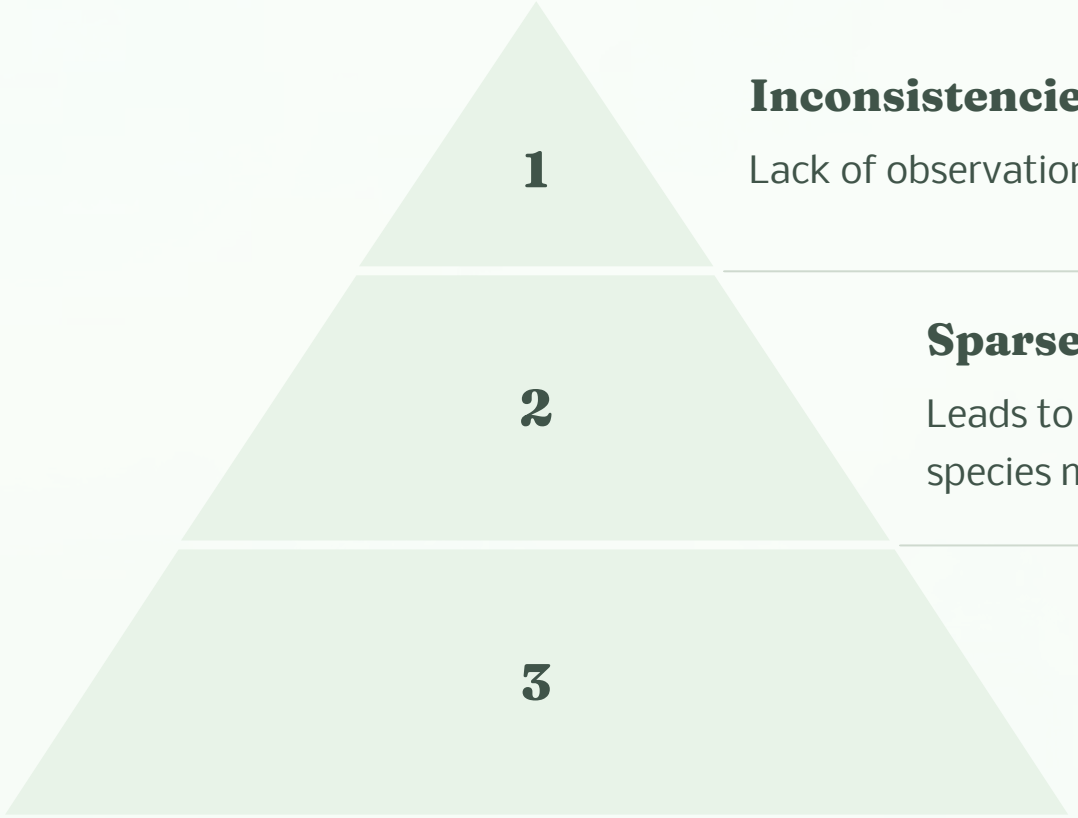
Interaction-Type Errors

Observations mis-tagged as predator-prey (e.g., pollination, parasitism).

Structural Problems

- Species "eating themselves
- Misaligned chains
- Data sparsity leading to distorted interpretations.

Current Assumptions & Interpretation



Inconsistencies

Lack of observations for certain groups (e.g., decomposers).

Sparse Data

Leads to inaccurate trophic-level inferences and misleading critical species metrics.

Validation Needed

Additional validation required before fully trusting the network.

Next Steps: Enhancing Data Reliability

Expert Review

Validate Plantae, Fungi, Chromista datasets with Bradley.

Seek guidance on non-predatory interactions such as pollination

LLM-Based Validation

Build an AI pipeline to check biological plausibility, cross-verify with sources, and flag suspicious interactions.

Geographical Filtering

Restrict dataset to U.S. observations for a more reliable, coherent subset.

Next Steps: Structural Improvements

Trophic Attributes

Define producers, primary consumers, secondary consumers.

Functional Feeding Groups

Add herbivore, carnivore, omnivore, etc.

Scalable Data Dictionary

Create a robust schema before repopulating the full dataset.

Thank You

Feel free to share any suggestions, recommendations, or questions you might have.