WOAH SRR-SEA capacity building on risk analysis for transboundary animal disease control purposes in Southeast Asia

World Organisation for Animal Health

Australian Government
Department of Agriculture, Fisheries and Forestry

EXTRA UNIT 2
Network analysis

Department of Emerging diseases and Global health

Animal Health Research Centre (CISA)

Institute for Agronomic and Food Research (INIA)

Spain's Research Council (CSIC)

Contact: martinez.marta@inia.csic.es

CISA CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL ANIMAL HEALTH RESEARCH CENTRE

INIA CSIC

Outline

- An introduction to Network analysis
- Basic difference in Complex Networks
- Key parameters in analyzing Complex Networks
- Visualization and analysis

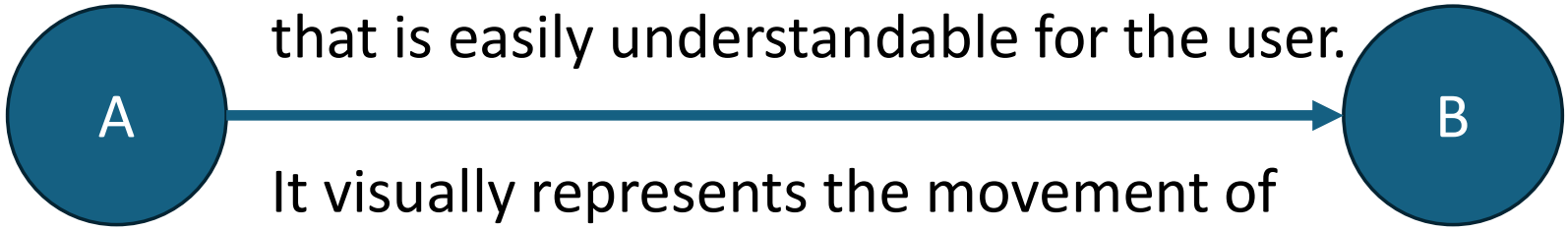
CISA
CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL
ANIMAL HEALTH RESEARCH CENTER

PART 1

An introduction to Network analysis

What is a network?

A complex network is the representation of an interconnected system in a format that is easily understandable for the user.



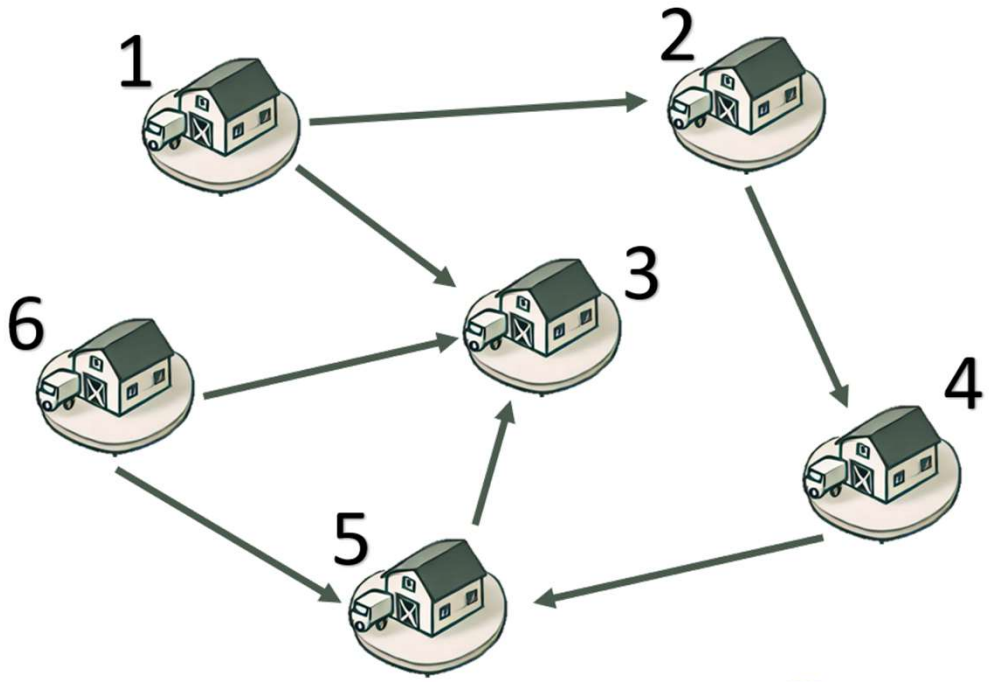
It visually represents the movement of information from one side to another and can be as complex as the user's system.

What is a network?

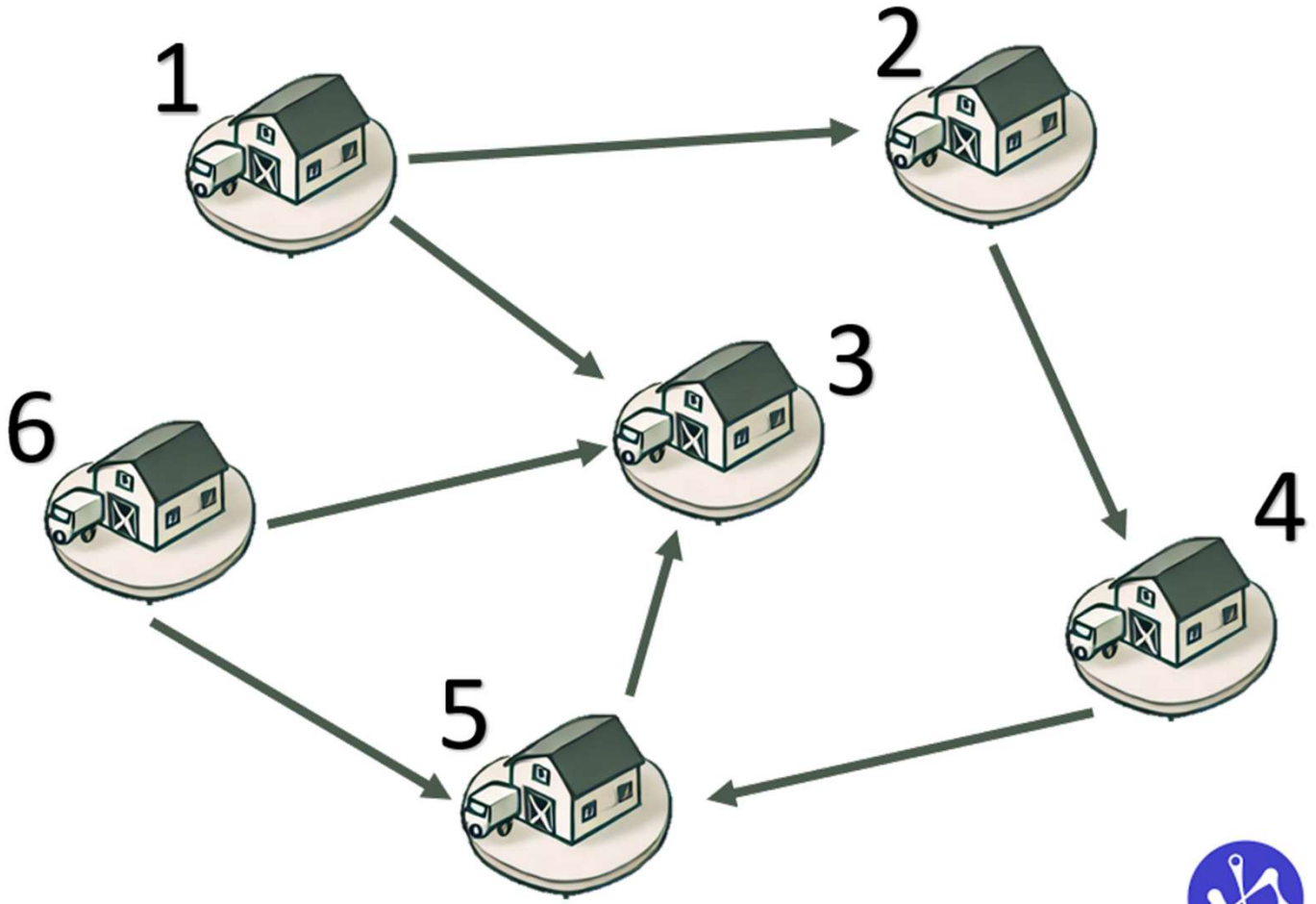
In this case, the information travels from point A (formally called a **node**) to point B, and the path is called an **edge**.

Complex Network in veterinary

From now on, and for the remainder of the module related to complex networks, these relationships will be established as exchanges of products between farms, with these products being the animals of choice.



An introduction to Network analysis



CISA
CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL
ANIMAL HEALTH RESEARCH CENTER

INIA
CSIC

What do you think?

Which of the following examples can be considered part of a complex network in the context of livestock resources?

- ❑ The routes of animal transportation connecting different farms and markets.
- ❑ The reproductive cycle of a cow on a single farm.
- ❑ The interaction between local veterinarians who monitor animal health on different farms in a region.
- ❑ The watering system within a single farm, used only by the animals residing there, with all biosecurity measures in place.
- ❑ The migratory movements of wild birds that may interact with farm animals.

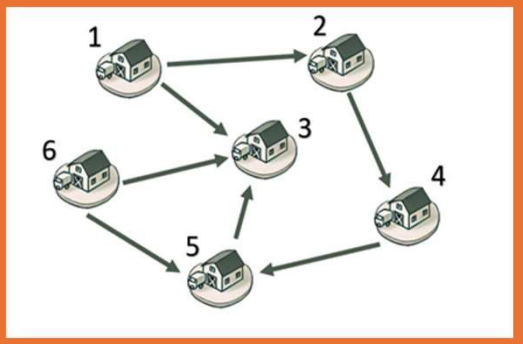
PART 2

Basic difference in Complex Networks

Directed vs Undirected Networks

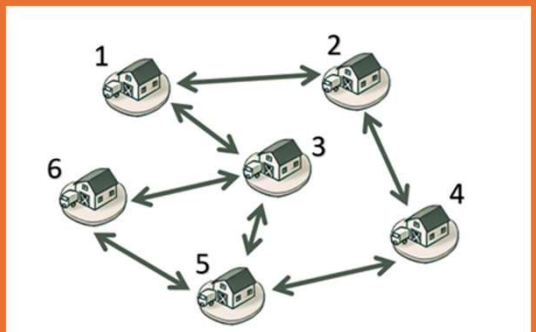
Directed networks

The livestock product exchanges between farms have a **specific direction**, meaning the flow of animals moves from one farm to another but not necessarily in reverse



Undirected networks

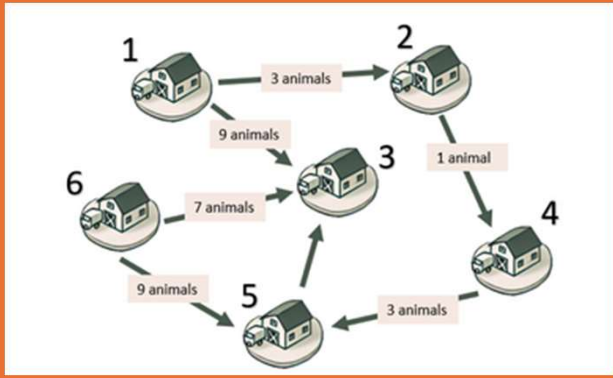
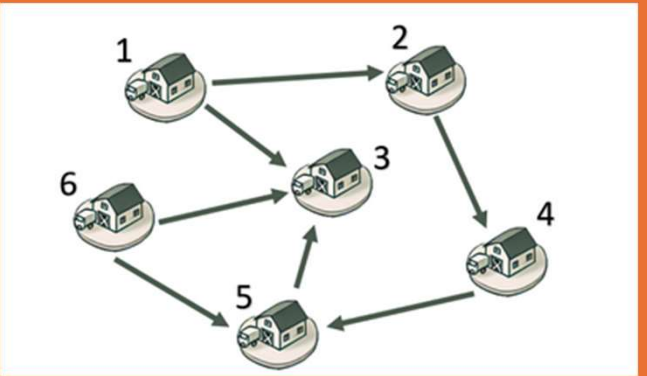
Represent exchanges where livestock products move freely in **both directions**, implying mutual trade between farms



CISA
CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL
ANIMAL HEALTH RESEARCH CENTER

INIA
CSIC

Weighted vs Unweighted Networks

Weighted networks	Each connection between farms is assigned a value or weight, indicating the volume, frequency, or importance of the livestock product exchanges between them.	
Unweighted networks	Only show whether or not an exchange exists between farms	

CISA CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL

INIA CSIC

PART 3

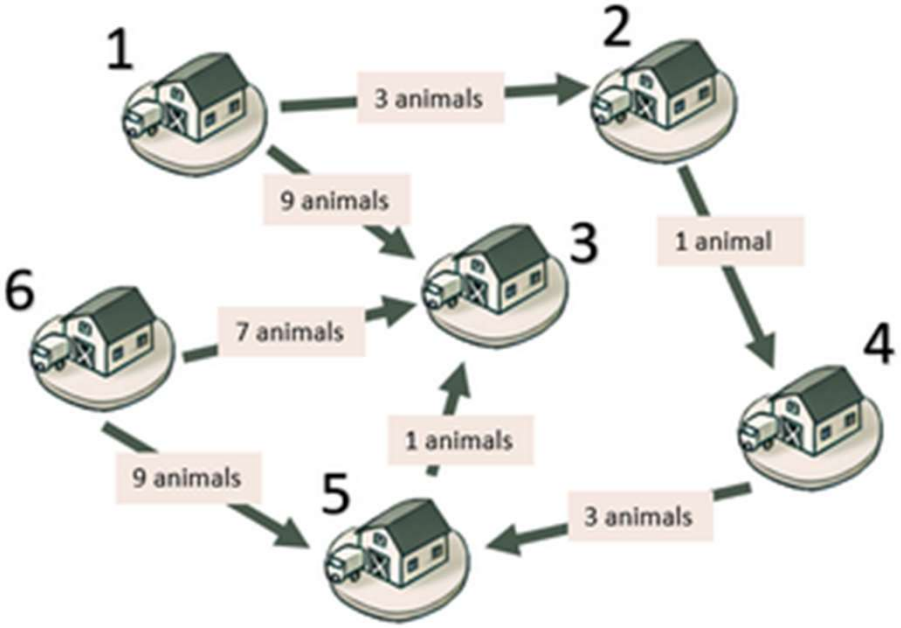
Key parameters in analyzing Complex Networks

Node Degree

- **In-degree**: The number of incoming connections to a node, indicating how many other nodes are sending products to it
- **Out-degree**: The number of outgoing connections from a node, showing how many other nodes it sends products to.
- **Total degree (unweighted)**: The sum of both in-degree and out-degree, showing the total number of connections a node has, regardless of direction.
- **Weighted in-degree**: The total weight of incoming connections, representing the volume or significance of products being received by the node.
- **Weighted out-degree**: The total weight of outgoing connections, reflecting the volume or significance of products being sent from the node.
- **Total degree (weighted)**: The sum of weighted in-degree and weighted out-degree, indicating the total volume or significance of product exchanges a node participates in.

CISA
CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL
INIA
CSIC

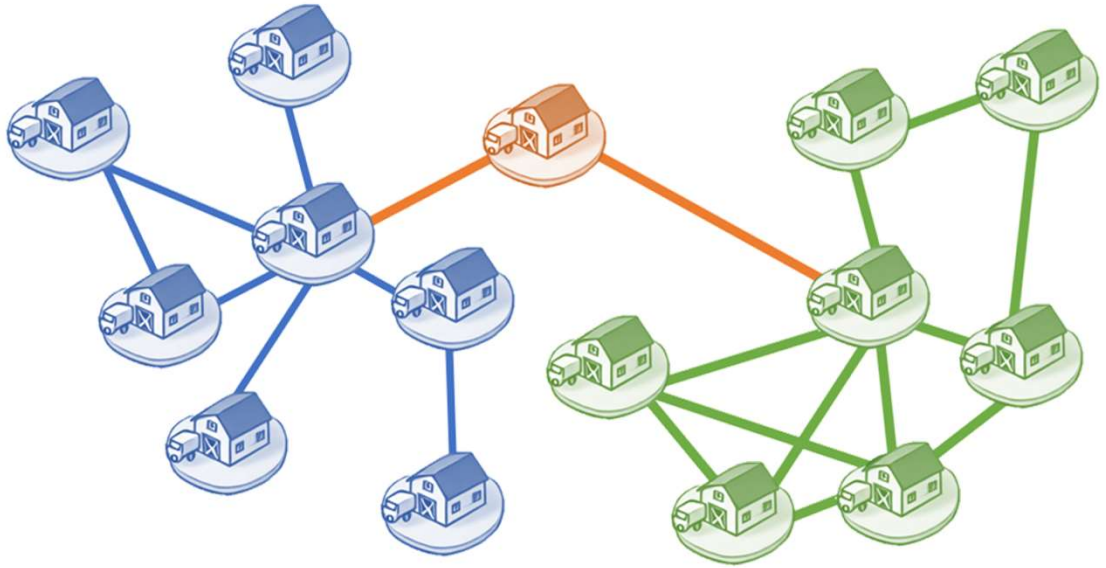
Node Degree



NODE 5	
Parameter	Values
In-degree	2
Out – degree	1
Total – degree	2 + 1 = 3
Weighted in -degree	9 + 3 = 12
Weighted out -degree	1
Total weighted degree	12 + 1 = 13

Betweenness centrality

It indicates the node's role as a bridge or critical connector between different parts of the network



Global properties

Average degree: Indicates overall network interconnectedness, where a higher value means faster disease spread.

Degree distribution: Shows the variety of node connections, with scale-free networks relying on a few highly connected nodes.

Clustering coefficient: Measures how closely nodes group together, helping predict localized outbreaks.

Path length: Describes the average number of steps between nodes, with shorter paths enabling rapid disease spread.

Network density: Reflects the ratio of edges to possible connections, affecting the difficulty of controlling disease outbreaks.

Component size: Refers to the size of the largest interconnected subnetwork, influencing the potential for widespread infection.

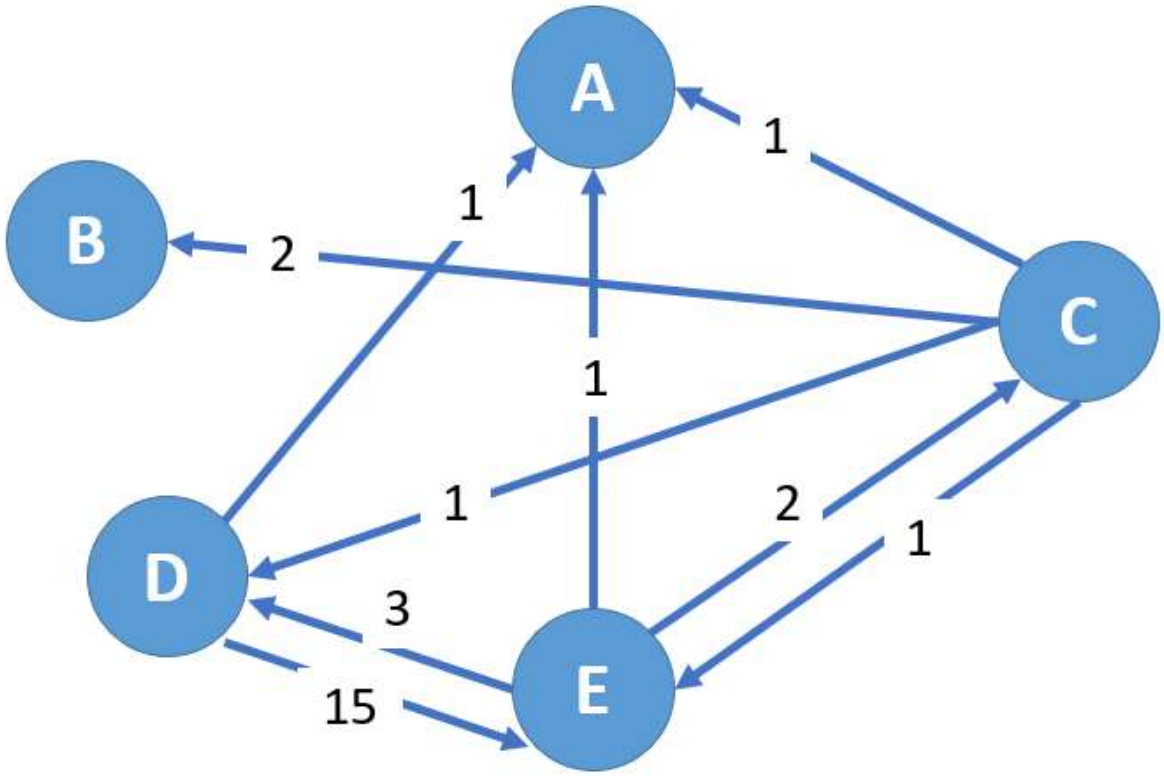
Network topology: The overall structure (random, scale-free, or small-world) impacts disease dynamics and spread patterns.

Exercise 1

Key parameters in analyzing Complex Networks

Exercise 1

Create a table indicating the corresponding parameters for each node in the network:



Exercise 1

Create a table indicating the corresponding parameters for each node in the network:

Node	In-Degree	Weighted In-Degree	Out-Degree	Weighted Out-Degree	Total Degree	Weighted Total Degree
A						
B						
C						
D						
E						

Exercise 1

Assuming that this example represents a complex network showing the exchange of a certain animal species intended for consumption between different farms, indicate which node corresponds to each key parameter about complex networks previously studied.

Question	Answer
Node with the highest indegree (unweighted)	
Node with the highest degree (unweighted)	
Node with the highest weighted indegree	
Node with the highest weighted outdegree	
Node with the highest total weighted degree	
Node with the highest outdegree (unweighted)	

CISA
CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL
ANIMAL HEALTH RESEARCH CENTER

PART 4

Visualization and analysis

Tutorial: Loading data into Gephi

Step 1: Download and install Gephi

Visit the Gephi website and download the latest version of the software suitable for your operating system.
Follow the installation instructions provided on the website.

Step 2: Prepare your Data

Gephi supports various data formats. Ensure your data is formatted correctly, with nodes and edges clearly defined.

- **Nodes**: Typically, a list of entities (e.g., individuals, locations) in your network.
- **Edges**: Connections between these entities (e.g., interactions, transmissions).

Step 3: Import data into Gephi

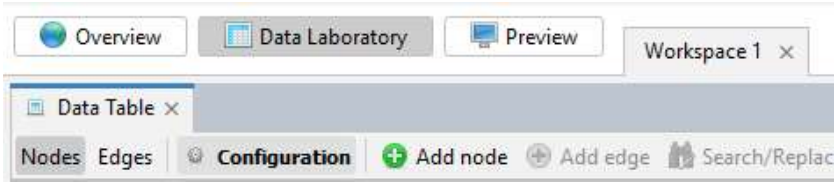
Open Gephi: Launch the software to access the main interface.
Create a new project: Click on File > New Project to start a new project.

CISA
CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL
ANIMAL HEALTH RESEARCH CENTER
INIA
CSIC

Tutorial: Loading data into Gephi

Import node data:

- In the Data Laboratory tab, select the node Data Table.
- Go to File > Import Spreadsheet.
- Select your nodes CSV file. Ensure your file contains at least an ID column (an unique identificator of the nodes) and any other relevant attributes (e.g., Label, Type). The next image shows an example:
- Click Next, then Finish to import the nodes. At the last window, select the option Append to existing workspace.



	A	B	C	D
1	id	label	latitude	longitude
2	0	Node_0	41.5411997	-8.93743985
3	1	Node_1	41.663961	1.0229159
4	2	Node_2	40.8711601	-8.56615534
5	3	Node_3	37.3796718	-5.79457309
6	4	Node_4	38.8040242	-1.54462607
7	5	Node_5	41.0014787	-1.41414322
8	6	Node_6	42.9797983	-6.09715669
9	7	Node_7	41.4492079	-8.92236045
10	8	Node_8	40.0607828	0.20784107
11	9	Node_9	42.4692864	-5.35817567
12	10	Node_10	36.9976108	-0.08356238

Tutorial: Loading data into Gephi

Import edge data:

- Similarly, click on Edge Data Table and go to File > Import Spreadsheet.
- Select your edges CSV file. Ensure it contains Source, Target, and any other relevant attributes (e.g., Weight). It is necessary that the source and target values are related to the IDs assigned to the nodes in the previous step. The next image shows an example:
- Click Next. At the last window, select the option Append to existing workspace and then Finish to import the edges.

	A	B
1	source	target
2	0	1
3	0	2
4	0	3
5	0	4
6	0	5
7	0	6
8	0	7
9	0	8

Tutorial: Loading data into Gephi

Step 4: Visualize and analyze your network

1.Overview: Switch to the Overview tab to start visualizing your network.

2.Layout: Choose a layout algorithm from the Layout panel (e.g., Force Atlas, Yifan Hu) to arrange your network graph.

3.Appearance: Customize the appearance of nodes and edges using the Appearance panel. You can adjust size, color, and other visual properties based on attributes.

4.Filters: Apply filters to focus on specific parts of your network (e.g., filter by node degree, edge weight).

5.Statistics: Use the Statistics panel to run network analysis (e.g., calculate centrality measures, identify clusters).

Advise: To learn how to use the different tools to visualize and analyze graphs in Gephi, please visit https://gephi.org/users/ where you can find a full tutorial made by Gephi distributors.

CISA
CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL
ANIMAL HEALTH RESEARCH CENTER

INIA
CSIC

Tutorial: Loading data into Gephi

Step 5: Save and export

Save your project regularly to avoid data loss (File > Save).

Export your visualizations and data as needed for reports or further analysis (File > Export).

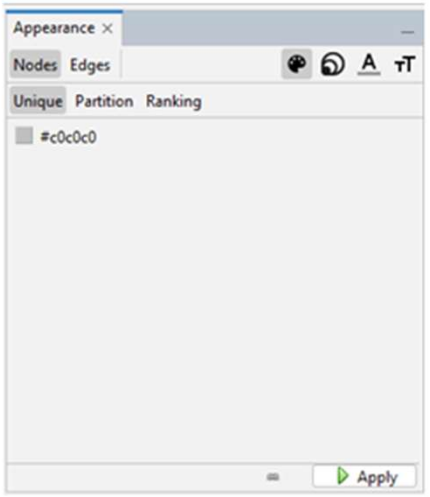
Tutorial: Loading data into Gephi

By following these steps, epidemiologists can leverage Gephi to gain valuable insights into disease transmission networks, identify critical points of intervention, and effectively communicate findings through clear and impactful visualizations.

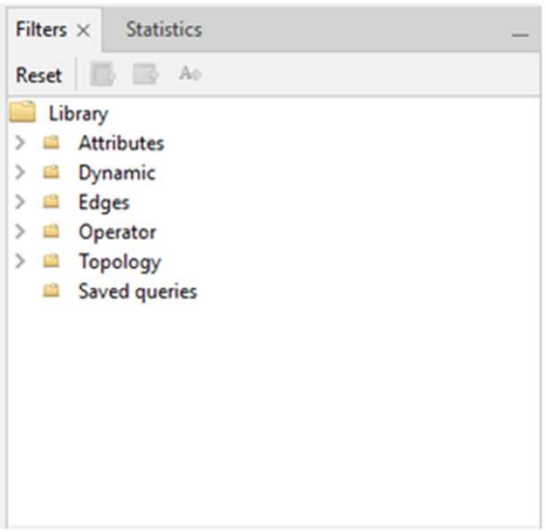
CISA
CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL
ANIMAL HEALTH RESEARCH CENTER

Tutorial: Loading data into Gephi

Appearance window

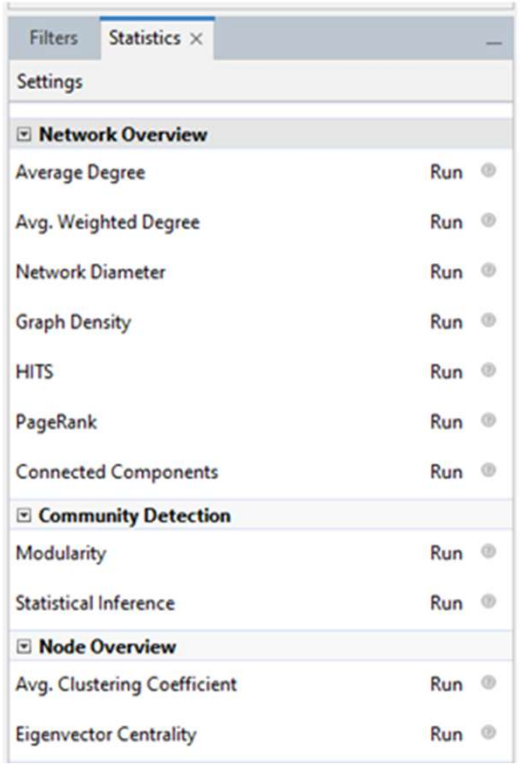


Filter window

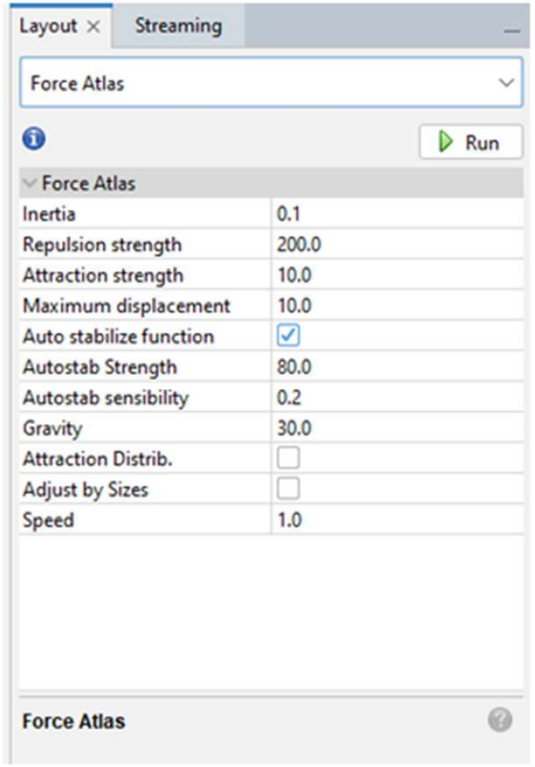


Tutorial: Loading data into Gephi

Statistics window



Layout window



Exercise 2

Analyzing a farm network for epidemic spread

Exercise 2

As an exercise, you will analyze a network of farms **where nodes represent farms and edges represent connections or interactions between them**. The nodes are geolocated, providing valuable spatial information. Your task is to study the network, identify its main characteristics, and suggest strategies for controlling an epidemic if it were to spread among the farms.

This is a **free exercise**, in which each student is expected to show the knowledge he/she has learned about complex networks. Each student must insert an image of the original complex network (without filters) resulting from its integration in Gephi and carry out a study according to the properties of the networks that have been discussed during the module, as well as other characteristics that the student considers.

Exercise 2

First, you will import the provided CSV files (at the end of the text) into Gephi and examine the network's structure.

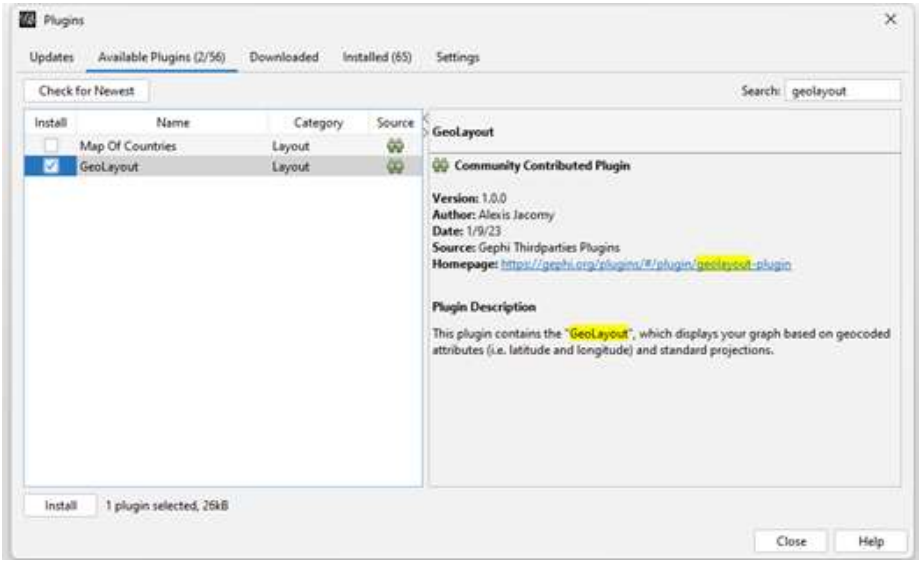
For example, some key characteristics to analyze include:

- **Degree distribution**: Determine the number of connections each farm has to understand the network's connectivity.
- **Centrality measures**: Identify the most influential farms using metrics such as degree centrality, betweenness centrality, and closeness centrality.
- **Clusters**: Detect clusters or communities within the network to understand how farms group together and identify potential points for intervention.
- **Spatial distribution**: Use geolocation data to visualize the spatial layout of the network, helping you understand the geographic spread and proximity of farms (see the tutorial at the end of the text).

Exercise 2: Installing and using the Gephi plugin for geolocated coordinates

Step 1: Install GeoLayout plugin

- Open Gephi and navigate to Tools > Plugins.
- In the Available Plugins tab, search for GeoLayout.
- Select the plugin and click Install. Follow the prompts to complete the installation.



Exercise 2: Installing and using the Gephi plugin for geolocated coordinates

Step 2: Import and visualize geolocated data

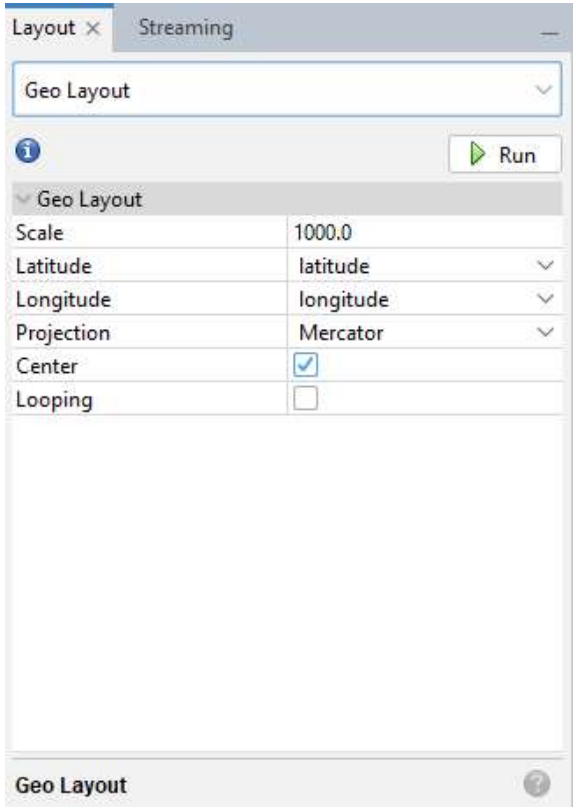
Import nodes and edges:

- Open Gephi and create a new project.
- Go to File > Import Spreadsheet and select your nodes CSV file. Ensure it includes columns for ID, Label, and Coordinates (latitude and longitude).
- Click Next, then Finish to import the nodes.
- Repeat the process for the edges CSV file, ensuring it includes Source, Target, and any relevant attributes.

Configure GeoLayout plugin:

- Switch to the Overview tab.
- Select the Layout panel and choose the GeoLayout algorithm.
- Enter the appropriate columns for latitude and longitude in the plugin settings.
- Run the layout algorithm to position the nodes based on their geospatial coordinates.
- You can repeat the process as many times as needed, varying the scale value to obtain a good visualization of your geospatial network.

Exercise 2: Installing and using the Gephi plugin for geolocated coordinates



Exercise 2: Installing and using the Gephi plugin for geolocated coordinates

Analyze and interpret the network:

- Use the Statistics panel to calculate centrality measures, cluster detection, and other metrics.
- Visualize the network spatially to understand the geographic spread and proximity of farms.
- Apply filters and customize the appearance to highlight key nodes and clusters.

Discussion:

- Identify critical nodes (hubs) that play significant roles in transmission.
- Suggest containment strategies, such as isolating key farms, increasing surveillance in highly connected farms, and enhancing biosecurity measures.

Exercise 2: Installing and using the Gephi plugin for geolocated coordinates

By following these steps, you will gain insights into the farm network's structure and develop strategies to control an epidemic, utilizing both network analysis and geospatial visualization in Gephi.