Manipulating networks and data with python

Suzana Santos
FGV-EMAp

July 5th 2022

Networkx

Features (according to wikipedia)

- Classes for graphs and digraphs.
- Conversion of graphs to and from several formats.
- Ability to construct random graphs or construct them incrementally.
- Ability to find subgraphs, cliques, k-cores.
- Explore adjacency, degree, diameter, radius, center, betweenness, etc.
- Draw networks in 2D and 3D.

Building a network

- **From file:**

```
import networkx as nx
nx.read_edgelist(path_to_file)
```

- **From edge list:**

```
nx.from_edgelist(list_of_edges)
```

- **From adjancecy matrix:**

```
nx.from_numpy_array(adjacency_matrix)
```

These methods will return an object of the networkx class Graph.

Adding/removing nodes or edges

- **Nodes:**

```
G.add_nodes_from(list_of_nodes)
```

```
G.remove_nodes_from(list_of_nodes)
```

- **Edges**:

```
G.add_edges_from(list_of_edges)
```

```
G.remove_edges_from(list_of_edges)
```

Listing nodes, edges, and features

- **Nodes**

`G.nodes`

Set of node labels

- **Edges**

`G.edges`

Set of tuples corresponding to edges

- **Degrees**

`G.degree`

Set of tuples with the node label and the corresponding degree

Writing a graph to a file

```
nx.write_edgelist(graph, path_to_file)
```

Pandas

Library for data manipulation and analysis.

Some features (according to wikipedia):

- DataFrame object
- Data alignment and integrated handling of missing data.
- Label-based slicing, fancy indexing, and subsetting of large data sets.
- Data structure column insertion and deletion.
- Data set merging and joining.
- Time series-functionality: Date range generation[6] and frequency conversions, moving window statistics, moving window linear regressions, date shifting and lagging.
- Provides data filtration.