Assignment 2: Creating Nodes and Edges

Step 1: Reviewing the .csv data to assure everything is there

In [47]:
```
import pandas as pd
df = pd.read_csv("Assignment_2.csv")
```

In [48]:
```
df.head()
```

Out[48]:

	Book_ID	Title	Extended Title	Place of Publication	Name of Publisher
0	9918687950101480	De gildestem.	NaN	S.l.	ACW
1	9918748890101480	Onze lieve vrouw van Kortenbos.	NaN	Sint-Truiden	Onze lieve vrouw van Kortenbos
2	9918753570101480	Nieuws stichting Chili.	NaN	Tongerlo	Norbertijnen Tongerlo
3	9918754260101480	Toets	informatieblad van de syndicale werking LBC-NV...	Leuven	Unknown
4	9918761110101480	Pausbezoek.	NaN	Unknown	Unknown

In [49]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1119 entries, 0 to 1118
Data columns (total 5 columns):
 #   Column                Non-Null Count  Dtype 
---  ------                --------------  ----- 
 0   Book_ID               1119 non-null   int64 
 1   Title                 1119 non-null   object
 2   Extended Title        313 non-null    object
 3   Place of Publication  1119 non-null   object
 4   Name of Publisher     1119 non-null   object
dtypes: int64(1), object(4)
memory usage: 43.8+ KB
```

Step 2: Mapping the nodes/edges and creating new csv files

In [52]:
```
# Create a DataFrame for nodes
nodes_df = pd.DataFrame()
nodes_df['Label'] = df['Book_ID'].tolist() + df['Place of Publication'].tolist() + df[

# Drop duplicate values to keep only unique nodes
nodes_df = nodes_df.drop_duplicates()
nodes_df['ID'] = range(1, len(nodes_df) + 1)

# Create a DataFrame for edges
edges_df = pd.DataFrame()
edges_df['Source'] = df['Book_ID'].map(nodes_df.set_index('Label')['ID'])
edges_df['Target'] = df['Place of Publication'].map(nodes_df.set_index('Label')['ID'])
edges_df['Source_2'] = df['Place of Publication'].map(nodes_df.set_index('Label')['ID'
```

```
edges_df['Target_2'] = df['Name of Publisher'].map(nodes_df.set_index('Label')['ID'])
edges_df['Source_3'] = df['Book_ID'].map(nodes_df.set_index('Label')['ID'])
edges_df['Target_3'] = df['Name of Publisher'].map(nodes_df.set_index('Label')['ID'])


# Save nodes and edges to separate CSV files
nodes_df.to_csv('nodes.csv', index=False)
edges_df.to_csv('edges.csv', index=False)
```