Step 1: Create Journals, Authors, Papers
Do this for every Journal XML file. Change the GitHub url in the first line of code.

1. https://raw.githubusercontent.com/KatBCN/SDMLab1/main/BigDataMiningAndAnalytics.xml
2. https://raw.githubusercontent.com/KatBCN/SDMLab1/main/IEEETransactionsOnBigData.xml
3. https://raw.githubusercontent.com/KatBCN/SDMLab1/main/DataIntelligence.xml

Creating Graph from Journal XML Data - Includes All Authors per Article and generic WROTE relationship:

```
        WITH
"https://raw.githubusercontent.com/KatBCN/SDMLab1/main/BigDataMiningAndAnalytics.xml"
AS uri
        CALL apoc.load.xml(uri, '', {}, true)
        YIELD value
        UNWIND value._result as result
        UNWIND result._hits as hits
        WITH [x in hits._hit WHERE x._info] as articles
        UNWIND articles as article
        WITH [item in article._info WHERE item._type="title"][0]._text as title,
        [item in article._info WHERE item._type="venue"][0]._text as journal,
        [item in article._info WHERE item._type="volume"][0]._text as volume,
        [item in article._info WHERE item._type="number"][0]._text as number,
        ([item in article._info WHERE item._type="venue"][0]._text +
        ",vol." + [item in article._info WHERE item._type="volume"][0]._text +
        "-" + [item in article._info WHERE item._type="number"][0]._text) as
journalVolume,
        [item in article._info WHERE item._type="pages"][0]._text as pages,
        [item in article._info WHERE item._type="year"][0]._text as year,
        [item in article._info WHERE item._type="type"][0]._text as type,
        [item in article._info WHERE item._type="access"][0]._text as access,
        [item in article._info WHERE item._type="ee"][0]._text as doiLink,
        [item in article._info WHERE item._type = "authors"] AS authorList

        MERGE (p:Paper {title: title})
        SET p.access = access, p.type = type, p.pages = pages, p.doiLink = doiLink

        MERGE (jv:JournalVolume {title: journalVolume})
        SET jv.year = year, jv.volume = volume, jv.number = number

        MERGE (j:Journal {title: journal})

        MERGE (p)-[:PUBLISHED_IN]->(jv)

        MERGE (jv)-[:VOLUME_OF]->(j)

        WITH p, authorList
        UNWIND authorList AS authors
        WITH [x in authors._authors WHERE x._type = "author"] AS individuals, p
        UNWIND individuals as individual
        MERGE (a:Author {name:individual._text})
        MERGE (a)-[:WROTE]->(p);
```

Step 2: Create Conferences, Authors, Papers
Do this for every Conference XML file. Change the url link in the first line of code.

1. https://raw.githubusercontent.com/KatBCN/SDMLab1/main/IEEE_ACM_BDCAT.xml
2. https://raw.githubusercontent.com/KatBCN/SDMLab1/main/IEEE_BigComp.xml

Creating Graph from Conference XML Data - Includes All Authors per Article and generic WROTE relationship:

```
WITH "https://raw.githubusercontent.com/KatBCN/SDMLab1/main/IEEE_ACM_BDCAT.xml"
AS uri
CALL apoc.load.xml(uri, '', {}, true)
YIELD value
UNWIND value._result as result
UNWIND result._hits as hits
WITH [x in hits._hit WHERE x._info] as articles
UNWIND articles as article
WITH [item in article._info WHERE item._type="title"][0]._text as title,
[item in article._info WHERE item._type="venue"][0]._text as conference,
[item in article._info WHERE item._type="year"][0]._text as year,
([item in article._info WHERE item._type="venue"][0]._text +
"-" + [item in article._info WHERE item._type="year"][0]._text) as
conferenceEdition,
[item in article._info WHERE item._type="pages"][0]._text as pages,
[item in article._info WHERE item._type="type"][0]._text as type,
[item in article._info WHERE item._type="ee"][0]._text as doiLink,
[item in article._info WHERE item._type="access"][0]._text as access,
[item in article._info WHERE item._type = "authors"] AS authorList

MERGE (p:Paper {title: title})
SET p.access = access, p.type = type, p.pages = pages

MERGE (ce:conferenceEdition {title: conferenceEdition})
SET ce.year = year, ce.doiLink = doiLink

MERGE (c:Conference {title: conference})

MERGE (p)-[:PUBLISHED_IN]->(ce)

MERGE (ce)-[:EDITION_OF]->(c)

WITH p, authorList
UNWIND authorList AS authors
WITH [x in authors._authors WHERE x._type = "author"] AS individuals, p
UNWIND individuals as individual
MERGE (a:Author {name:individual._text})
MERGE (a)-[:WROTE]->(p);
```

Step 3: Set properties for “corresponding” author (although not essential to solve exercises B, C, and D)
Do this for every Journal & Conference XML file. Change the url link in the first line of code.

1. https://raw.githubusercontent.com/KatBCN/SDMLab1/main/BigDataMiningAndAnalytics.xml
2. https://raw.githubusercontent.com/KatBCN/SDMLab1/main/IEEETransactionsOnBigData.xml
3. https://raw.githubusercontent.com/KatBCN/SDMLab1/main/DataIntelligence.xml
4. https://raw.githubusercontent.com/KatBCN/SDMLab1/main/IEEE_ACM_BDCAT.xml
5. https://raw.githubusercontent.com/KatBCN/SDMLab1/main/IEEE_BigComp.xml

This must be done after defining nodes and relationships. Set property of role:”corresponding” in the WROTE relationship for first Authors per Article:

```
WITH
"https://raw.githubusercontent.com/KatBCN/SDMLab1/main/BigDataMiningAndAnalytic
s.xml" AS uri
CALL apoc.load.xml(uri, '', {}, true)
YIELD value
UNWIND value._result as result
UNWIND result._hits as hits
WITH [x in hits._hit WHERE x._info] as articles
```

```
UNWIND articles as article
WITH [x in article._info WHERE x._type = "authors"] AS authorList, article,
[x in article._info WHERE x._type = "title"] AS title
UNWIND authorList as authors
WITH [x in authors._authors WHERE x._type = "author"] AS individual, title
WITH individual[0]._text as author, title[0]._text as title
MATCH (a:Author {name:author})-[w:WROTE]-(p:Paper {title:title})
SET w.role = "corresponding"
```

Step 4: Verify Statistics - just to confirm we are working with same data.

```
MATCH (n)
WITH labels(n) as labels, size(keys(n)) as props, size((n)--()) as degree
RETURN
DISTINCT labels,
count(*) AS NumofNodes,
avg(props) AS AvgNumOfPropPerNode,
min(props) AS MinNumPropPerNode,
max(props) AS MaxNumPropPerNode,
avg(degree) AS AvgNumOfRelationships,
min(degree) AS MinNumOfRelationships,
max(degree) AS MaxNumOfRelationships
```

labels	Numof Nodes	AvgNumOfPropPerNode	MinNumPropPerNode	MaxNumPropPerNode	AvgNumOfRelationships	MinNumOfRelationships	MaxNumOfRelationships
[Paper]	1515	4.351155115511548	2	5	4.932673267326732	1	45
[JournalVolume]	61	4.0	4	4	10.344262295081966	3	30
[Journal]	3	1.0	1	1	20.333333333333332	12	31
[Author]	4647	1.0	1	1	1.280826339573918	1	54
[conferenceEdition]	16	3.0	3	3	60.4375	15	141
[Conference]	3	1.0	1	1	5.333333333333333	2	8

Papers with only one relationship seems suspicious…could be a data quality issue…how to check?

Problems to Solve:

1. Insert Citations - most important for part B & C

Some kind of random relationship generator between papers, but a paper cannot cite itself.

Query for papers.csv:

```
MATCH(author:Author)-[:WROTE
{role:"corresponding"}]->(paper:Paper)-[:PUBLISHED_IN]->(volumeEdition)-[]->(journalConference)
RETURN author.name as author, paper.title as title, volumeEdition.title as volumeEdition, volumeEdition.year as year,
journalConference.title as journalConference
```

2. Insert Keywords - important for part C & D

Some kind of random keyword assignment from the keywords given in the laboratory instructions: data management, indexing, data modeling, big data, data processing, data storage and data querying.

3. Configure all cypher code into Python code

Requirement of assignment even though most of these things can or should be solved by experimenting with Cypher in the browser first.

4. Parse :Paper nodes of type:"Editorship" for City and Date information of the Conference Edition - not high priority

This might be quickest to do by creating a .csv file or pandas data frame to assign the relationships and properties.

The information that is showing up in the :Paper node should be transferred to the :conferenceEdition node. In some cases, these "papers" also have authors associated with them, which are the organizers of the :conferenceEdition.

```
        MATCH(c:Conference)-[]-(ce:conferenceEdition)-[]-(p:Paper {type:'Editorship'})
        RETURN c,ce,p
```

Results:

c	ce	p
{"title":BDCAT}	{"year":2017,"title":BDCAT-2017,"doiLink":https://doi.org/10.1145/3148055}	{"title":Proceedings of the Fourth IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, BDCAT 2017, Austin, TX, USA, December 05 - 08, 2017,"type":Editorship}
{"title":BDCAT}	{"year":2020,"doiLink":https://doi.org/10.1109/BDCAT50828.2020,"title":BDCAT-2020}	{"title":7th IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, BDCAT 2020, Leicester, United Kingdom, December 7-10, 2020,"type":Editorship}
{"title":BDCAT}	{"year":2018,"doiLink":https://ieeexplore.ieee.org/xpl/conhome/8603671/proceeding,"title":BDCAT-2018}	{"title":5th IEEE/ACM International Conference on Big Data Computing Applications and Technologies, BDCAT 2018, Zurich, Switzerland, December 17-20, 2018,"type":Editorship}
{"title":BDCAT}	{"year":2019,"title":BDCAT-2019,"doiLink":https://doi.org/10.1145/3365109}	{"title":Proceedings of the 6th IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, BDCAT 2019, Auckland, New Zealand, December 2-5, 2019.,"type":Editorship}
{"title":BDCAT}	{"year":2021,"title":BDCAT-2021,"doiLink":https://doi.org/10.1145/3492324}	{"title":BDCAT '21 - 2021 IEEE/ACM 8th International Conference on Big Data Computing, Applications and Technologies, Leicester, United Kingdom, December 6 - 9, 2021,"type":Editorship}
{"title":BDCAT}	{"year":2016,"title":BDCAT-2016,"doiLink":https://doi.org/10.1145/3006299}	{"title":Proceedings of the 3rd IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, BDCAT 2016, Shanghai, China, December 6-9, 2016,"type":Editorship}
{"title":BDC}	{"year":2014,"doiLink":https://ieeexplore.ieee.org/xpl/conhome/7312651/proceeding,"title":BDC-2014}	{"title":1st IEEE/ACM International Symposium on Big Data Computing, BDC 2014, London, UK, December 8-11, 2014,"type":Editorship}
{"title":BDC}	{"year":2015,"doiLink":https://ieeexplore.ieee.org/xpl/conhome/7406204/proceeding,"title":BDC-2015}	{"title":2nd IEEE/ACM International Symposium on Big Data Computing, BDC 2015, Limassol, Cyprus, December 7-10, 2015,"type":Editorship}
{"title":BigComp}	{"year":2017,"doiLink":https://ieeexplore.ieee.org/xpl/conhome/7877084/proceeding,"title":BigComp-2017}	{"title":2017 IEEE International Conference on Big Data and Smart Computing, BigComp 2017, Jeju Island, South Korea, February 13-16, 2017,"type":Editorship}
{"title":BigComp}	{"year":2015,"doiLink":https://ieeexplore.ieee.org/xpl/conhome/7062153/proceeding,"title":BigComp-2015}	{"title":2015 International Conference on Big Data and Smart Computing, BIGCOMP 2015, Jeju, South Korea, February 9-11, 2015,"type":Editorship}
{"title":BigComp}	{"year":2020,"doiLink":https://ieeexplore.ieee.org/xpl/conhome/9050588/proceeding,"title":BigComp-2020}	{"title":2020 IEEE International Conference on Big Data and Smart Computing, BigComp 2020, Busan, Korea (South), February 19-22, 2020,"type":Editorship}
{"title":BigComp}	{"year":2016,"doiLink":https://ieeexplore.ieee.org/xpl/conhome/7422342/proceeding,"title":BigComp-2016}	{"title":2016 International Conference on Big Data and Smart Computing, BigComp 2016, Hong Kong, China, January 18-20, 2016,"type":Editorship}
{"title":BigComp}	{"year":2021,"doiLink":https://doi.org/10.1109/BigComp51126.2021,"title":BigComp-2021}	{"title":IEEE International Conference on Big Data and Smart Computing, BigComp 2021, Jeju Island, South Korea, January 17-20, 2021,"type":Editorship}

{"title":BigComp}	{"year":2019,"doiLink":https://ieeexplore.ieee.org/xpl/conhome/8671661/proceeding,"title":BigComp-2019}	{"title":IEEE International Conference on Big Data and Smart Computing, BigComp 2019, Kyoto, Japan, February 27 - March 2, 2019,"type":Editorship}
{"title":BigComp}	{"year":2014,"doiLink":https://ieeexplore.ieee.org/xpl/conhome/6731712/proceeding,"title":BigComp-2014}	{"title":International Conference on Big Data and Smart Computing, BIGCOMP 2014, Bangkok, Thailand, January 15-17, 2014,"type":Editorship}
{"title":BigComp}	{"year":2018,"doiLink":https://ieeexplore.ieee.org/xpl/conhome/8316805/proceeding,"title":BigComp-2018}	{"title":2018 IEEE International Conference on Big Data and Smart Computing, BigComp 2018, Shanghai, China, January 15-17, 2018,"type":Editorship}

Example of one hit in our XML structure for an editorship:

```
<hit score="1" id="397046">
<info><authors><author pid="09/2505">Herwig Unger</author><author pid="65/1664">Jinho
Kim</author><author pid="13/7122">U Kang</author><author pid="49/6689">Chakchai
So-In</author><author pid="13/1151">Junping Du</author><author pid="41/6237">Walid
Saad</author><author pid="25/6540">Young-Guk Ha</author><author pid="01/2903-2">Christian
Wagner 0002</author><author pid="51/1752">Julien Bourgeois</author><author
pid="55/6568">Chanboon Sathitwiriyawong</author><author pid="117/9378">Hyuk-Yoon
Kwon</author><author pid="29/654">Carson K. Leung</author></authors><title>IEEE
International Conference on Big Data and Smart Computing, BigComp 2021, Jeju Island,
South Korea, January 17-20,
2021</title><venue>BigComp</venue><publisher>IEEE</publisher><year>2021</year><type>Edito
rship</type><key>conf/bigcomp/2021</key><doi>10.1109/BIGCOMP51126.2021</doi><ee>https://d
oi.org/10.1109/BigComp51126.2021</ee><url>https://dblp.org/rec/conf/bigcomp/2021</url></i
nfo>
<url>URL#397046</url>
</hit>
```

B ATEEMPT AT GETTING QUERY