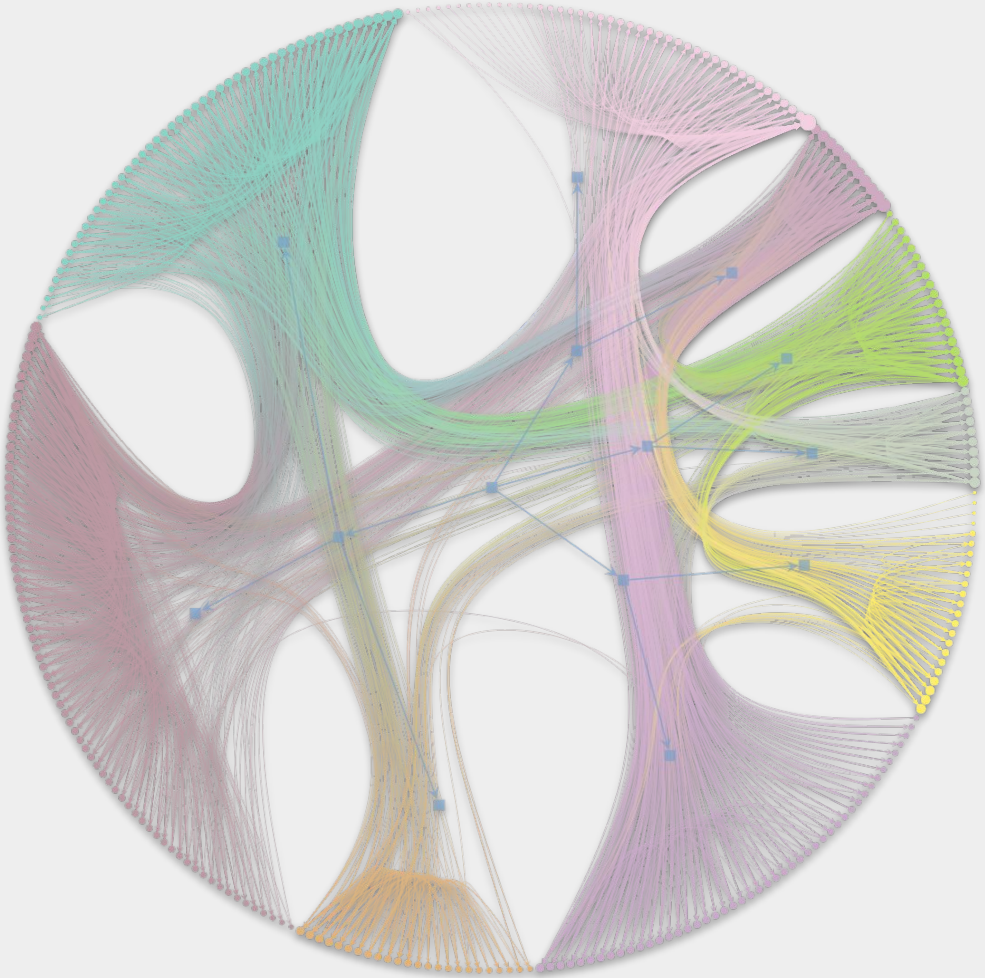
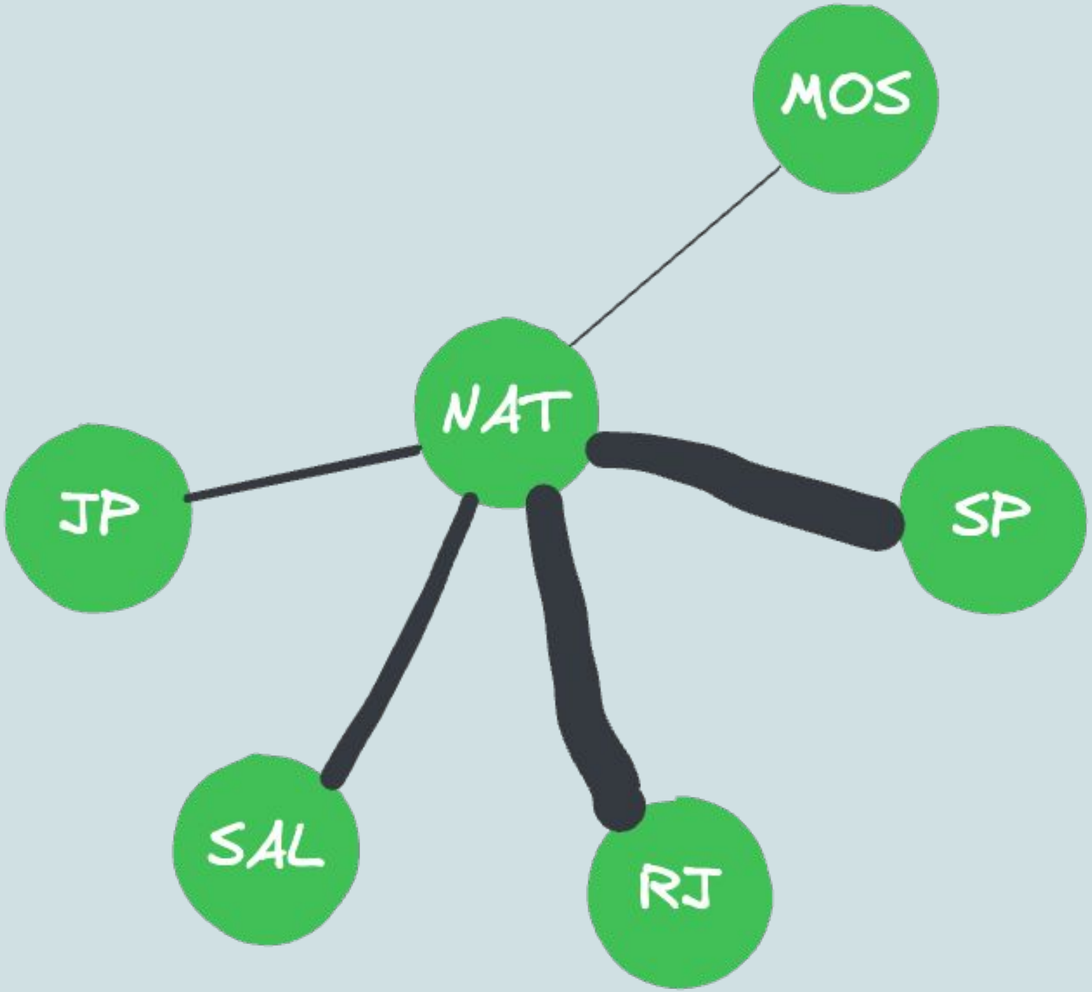
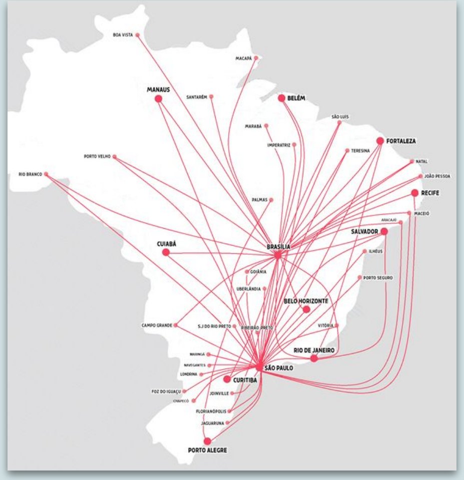
Network Elements

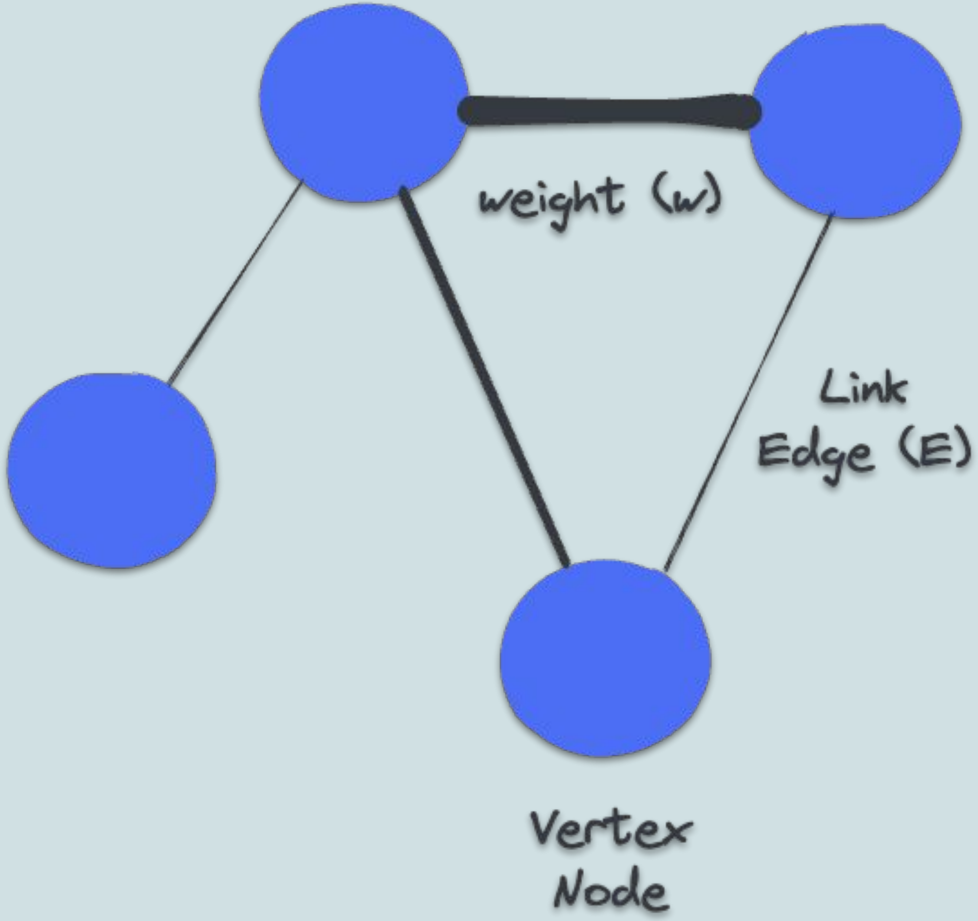
Review

ivanovitch.silva@ufrn.br
@ivanovitchm




MOS
NAT
JP
SP
SAL
RJ


BOA VISTA
MACAPÁ
MANAUS
SANTARÉM
BELÉM
SÃO LUÍS
MARABÁ
IMPERATRIZ
FORTALEZA
TERESINA
PORTO VELHO
NATAL
RIO BRANCO
JOÃO PESSOA
RECIFE
PALMAS
MACEIÓ
ARACAJU
SALVADOR
CUIABÁ
BRASÍLIA
ILHÉUS
GOIÂNIA
PORTO SEGURO
UBERLÂNDIA
BELO HORIZONTE
CAMPO GRANDE
S.J DO RIO PRETO
RIBEIRÃO PRETO
VITÓRIA
RIO DE JANEIRO
MARINGÁ
NAVEGANTES
SÃO PAULO
LONDRINA
CURITIBA
FOZ DO IGUAÇU
JOINVILLE
CHAPECÓ
FLORIANÓPOLIS
JAGUARUNA
PORTO ALEGRE

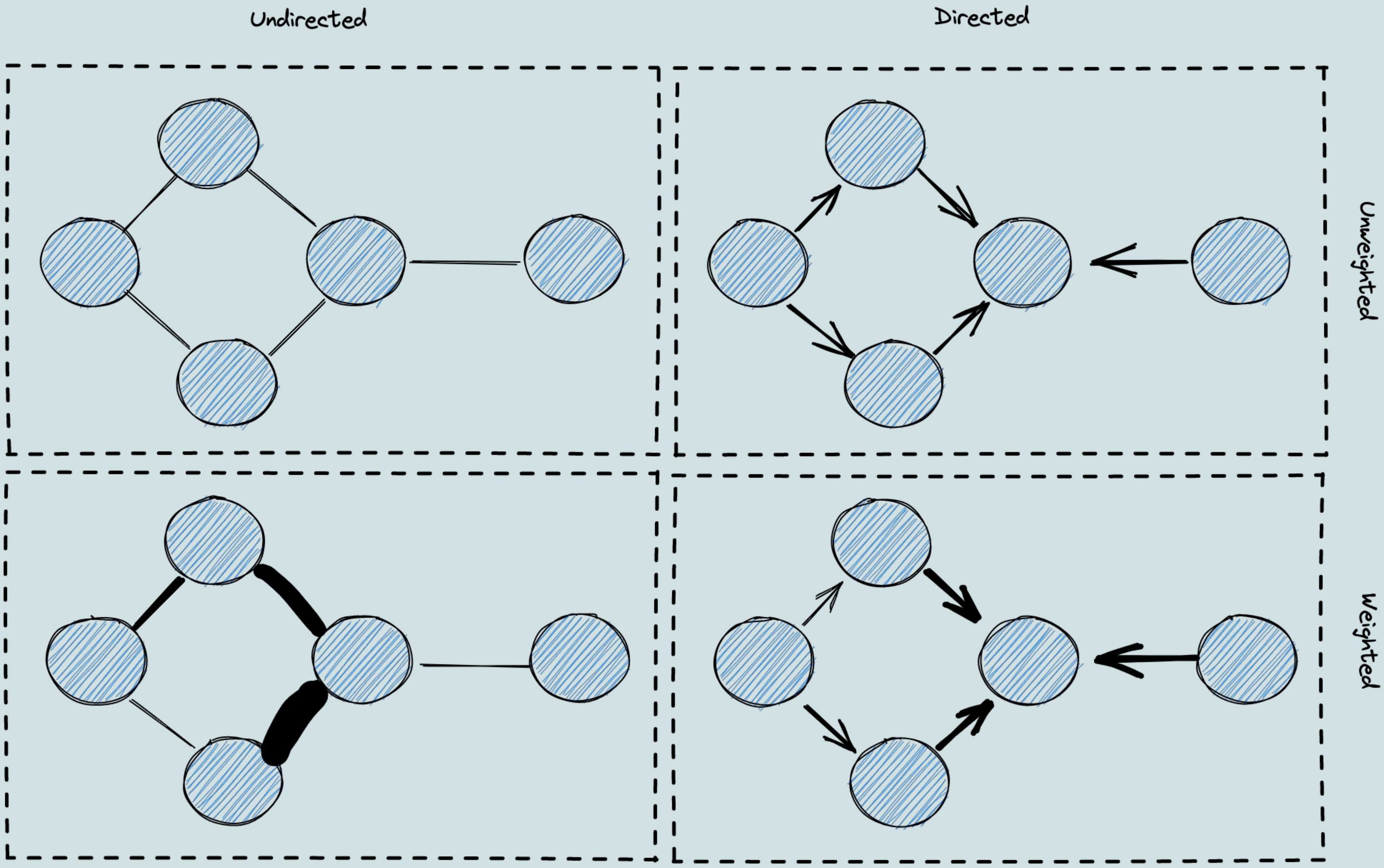


Basic Definitions

$$G = (V, E, W)$$

$$E \subseteq V \times V$$

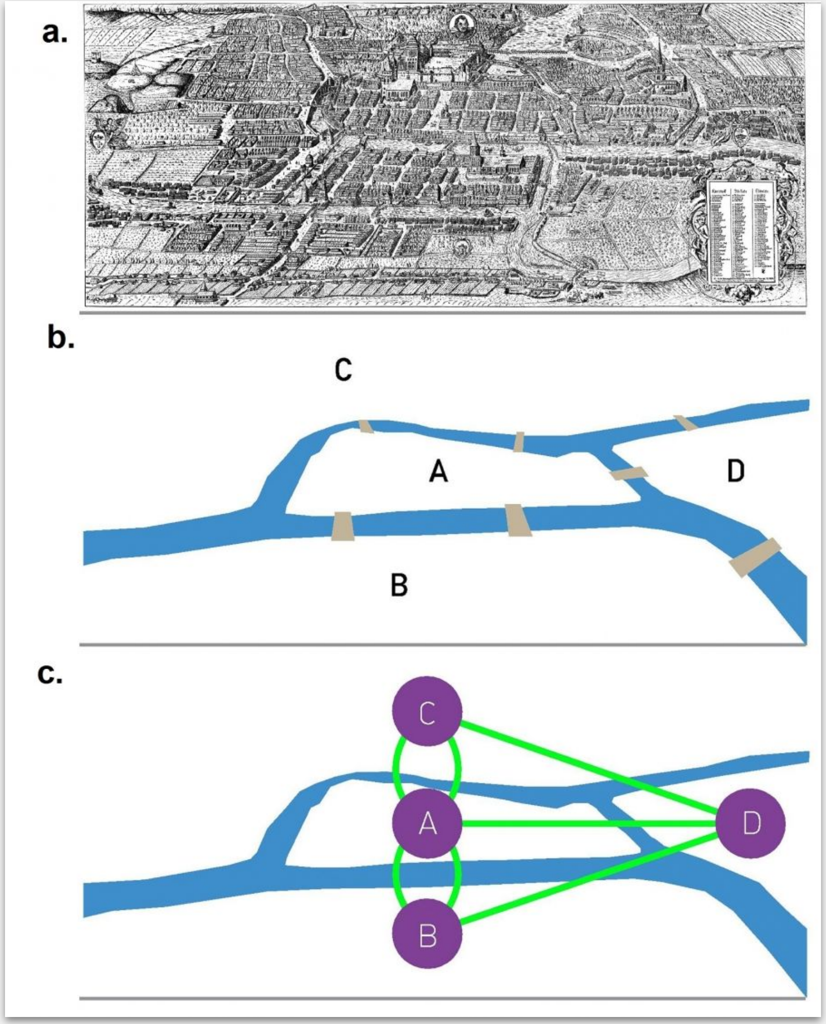
$$W \subseteq R^{+}$$


Undirected
Directed
Unweighted
Weighted

Seven Bridges of Königsberg

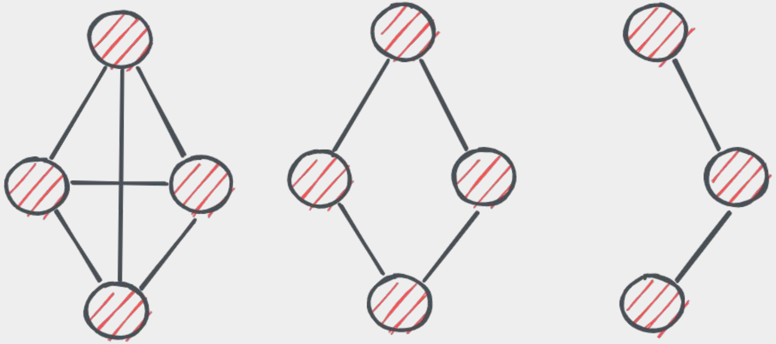
Can one walk across all seven bridges and never cross the same one twice?

The rigorous language for the description of networks is found in graph theory, a field of mathematics that can be traced back to the pioneering work of Leonhard Euler in the eighteenth century.



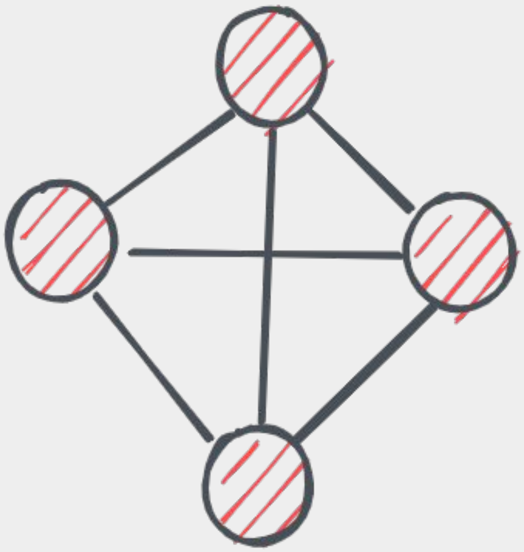
Despite many attempts, no one could find such path. The problem remained unsolved until 1735, when Leonhard Euler, offered a rigorous mathematical proof that such path does not exist.

Density and Sparsity, Subnetworks



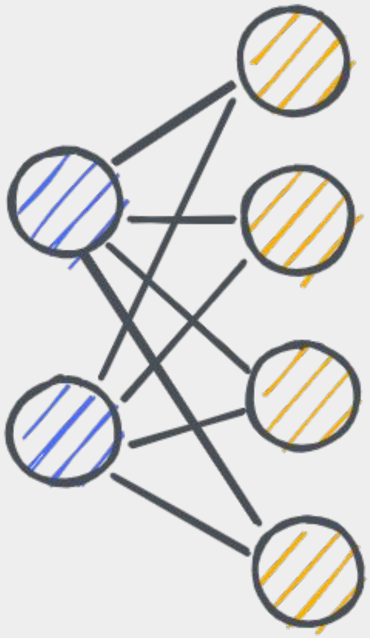
Density and Sparsity

The maximum number of links in a network is bounded by the possible number of distinct connections among the nodes of the system. The maximum number of links is therefore given by the number of pairs of nodes. A network with the maximum number of links, in which all possible pairs of nodes are connected by links, is called a **complete network.**



$$L_{max} = \binom{N}{2} = \frac{N(N-1)}{2}$$

$$L_{max}^{directed} = N(N-1)$$



$$L_{max} = N_1 \times N_2$$

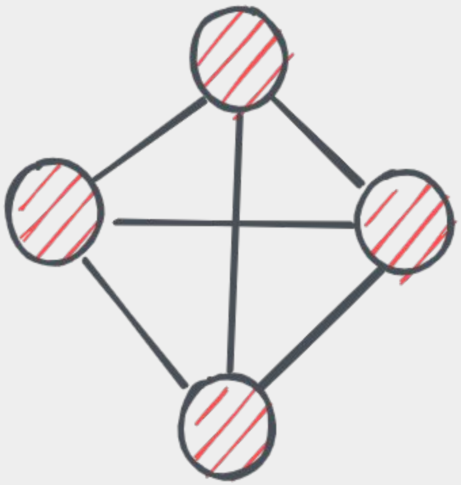
Density and Sparsity

$$d = \frac{L}{L_{max}} = \frac{2L}{N(N-1)}$$

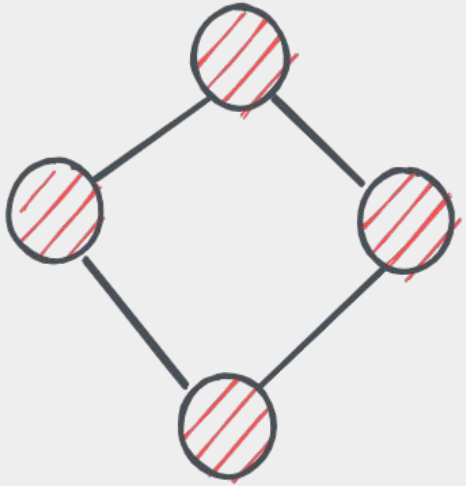
$$d = \frac{L}{L_{max}^{directed}} = \frac{L}{N(N-1)}$$

The fraction of possible links that actually exist, which is the same as the fraction of pairs of nodes that are actually connected, is called the **density of the network.**

$$d = \frac{L}{L_{max}}$$



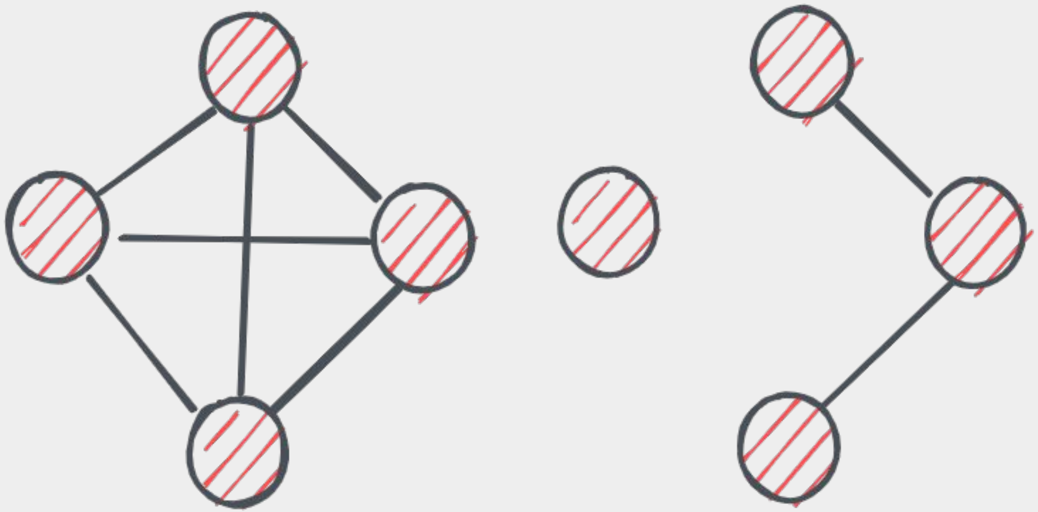
$$d = 1$$



$$d < 1$$

Density and Sparsity

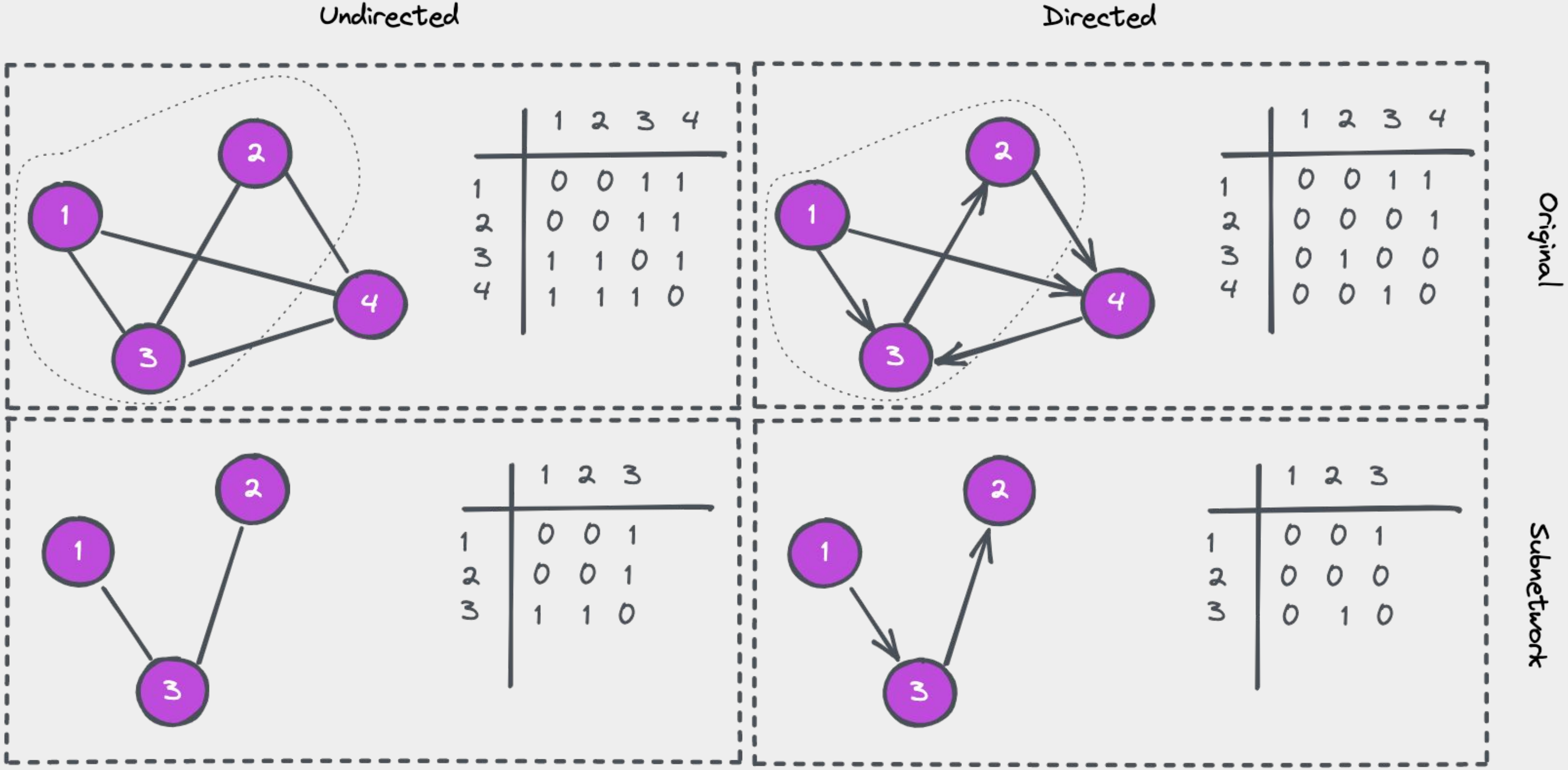
We say that the **network is sparse** if the number of links grows proportionally to the number of nodes (L ~ N), or even slower. If instead the number of links grows faster, e.g. quadratically with network size (L ~ N^2), then we say that the **network is dense.**



$$L \ll L_{max} \to d \ll 1$$

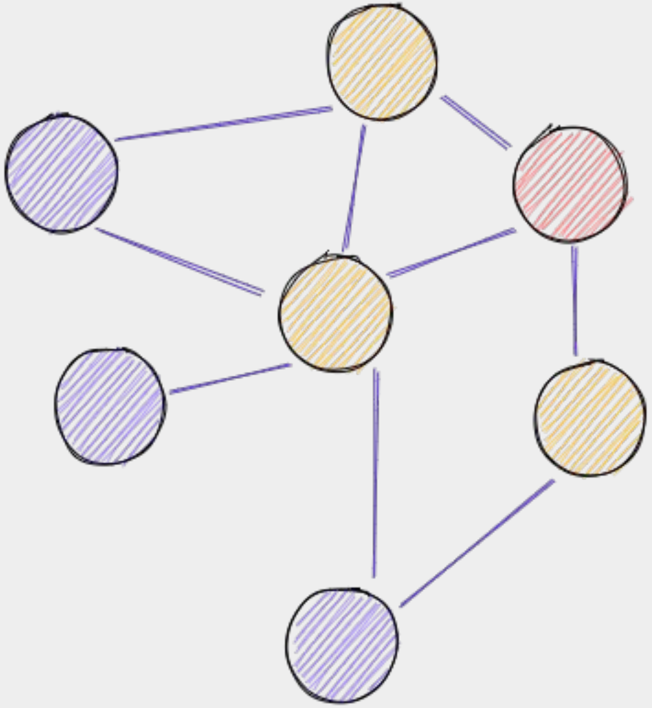
Network	Type	Nodes (N)	Links (L)	Density (d)	Average degree ($\langle k \rangle$)
Facebook Northwestern Univ.		10,567	488,337	0.009	92.4
IMDB movies and stars		563,443	921,160	0.000006	3.3
IMDB co-stars	W	252,999	1,015,187	0.00003	8.0
Twitter US politics	DW	18,470	48,365	0.0001	2.6
Enron email	DW	87,273	321,918	0.00004	3.7
Wikipedia math	D	15,220	194,103	0.0008	12.8
Internet routers		190,914	607,610	0.00003	6.4
US air transportation		546	2,781	0.02	10.2
World air transportation		3,179	18,617	0.004	11.7
Yeast protein interactions		1,870	2,277	0.001	2.4
C. elegans brain	DW	297	2,345	0.03	7.9
Everglades ecological food web	DW	69	916	0.2	13.3

Subnetwork

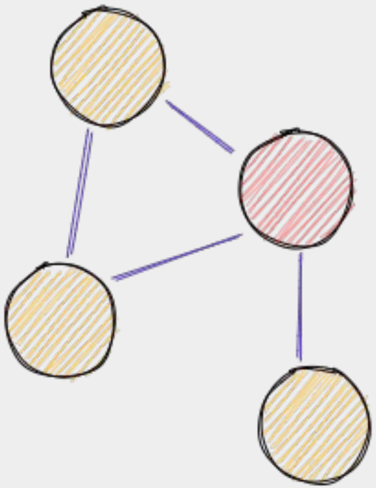

Undirected
Directed
Original
Subnetwork
1 2 3 4
1 0 0 1 1
2 0 0 1 1
3 1 1 0 1
4 1 1 1 0
1 2 3 4
1 0 0 1 1
2 0 0 0 1
3 0 1 0 0
4 0 0 1 0
1 2 3
1 0 0 1
2 0 0 1
3 1 1 0
1 2 3
1 0 0 1
2 0 0 0
3 0 1 0

Subnetwork

A special type of subnetwork is the ego network of a node, which is the subnetwork consisting of the chosen node — called the ego — and its neighbors. Ego networks are often studied in social network analysis.

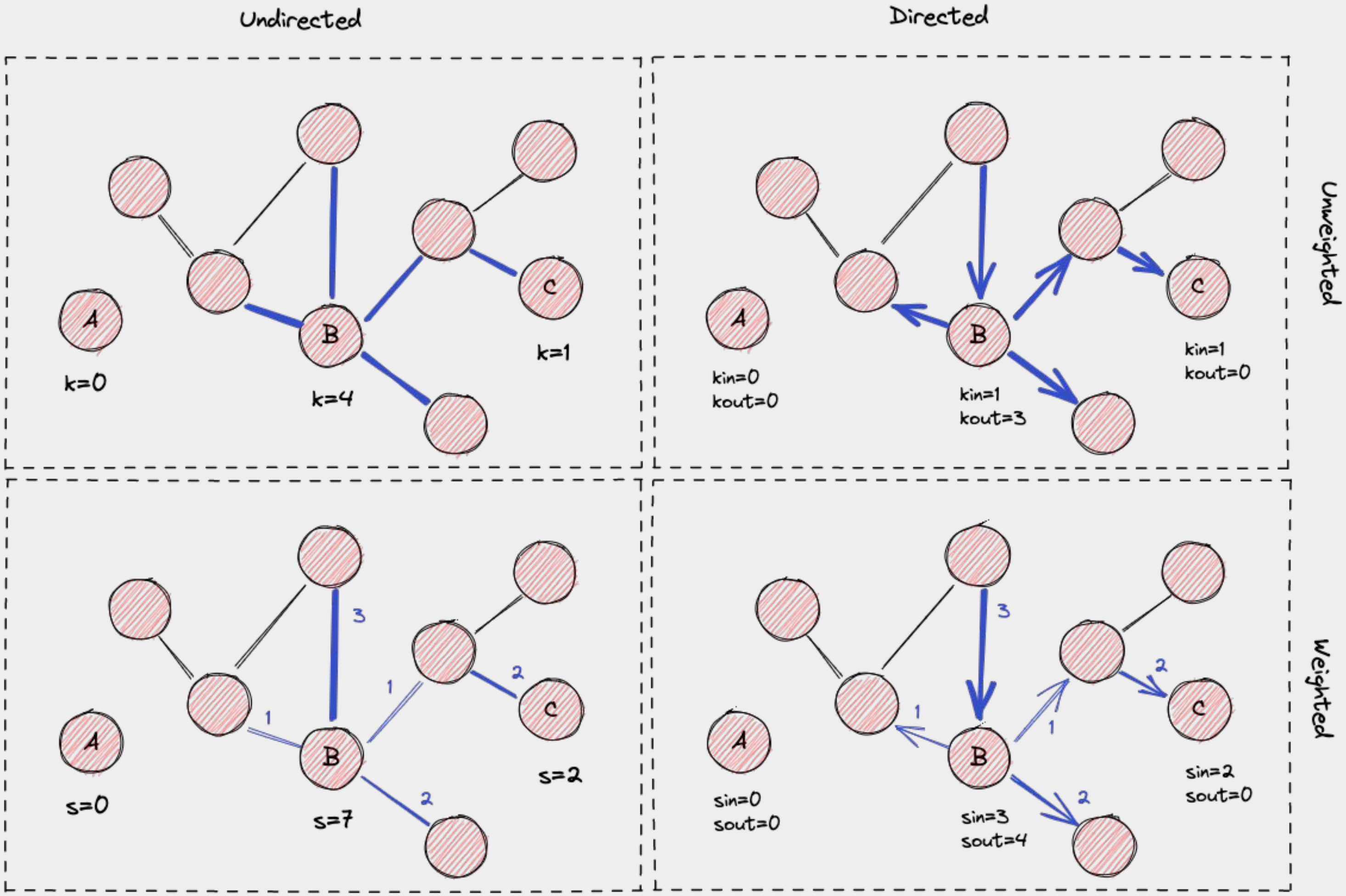


Original Network



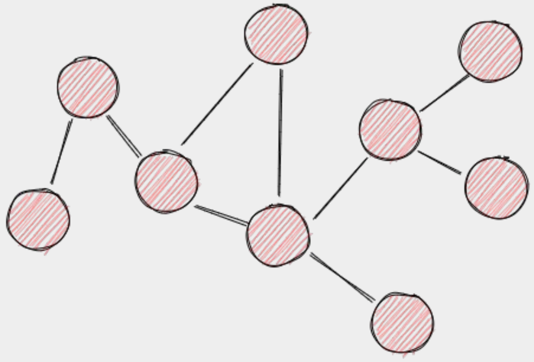
Ego Network

Degree and Network Representation



Undirected
Directed
Unweighted
Weighted
A
B
C
k=0
k=4
k=1
kin=0
kout=0
kin=1
kout=3
kin=1
kout=0
3
1
1
2
2
s=0
s=7
s=2
sin=0
sout=0
sin=3
sout=4
sin=2
sout=0

The average degree of a network is defined as:

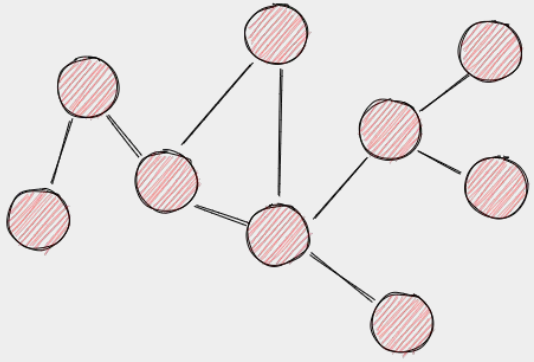


$$\langle \boldsymbol{k} \rangle = \frac{\sum_i \boldsymbol{k}_i}{N}$$

Since each link contributes to the degree of two nodes in an undirected network:

$$\langle k \rangle = \frac{2L}{N} = d(N-1) \qquad d = \frac{2L}{N(N-1)}$$

The average degree of a network is defined as:

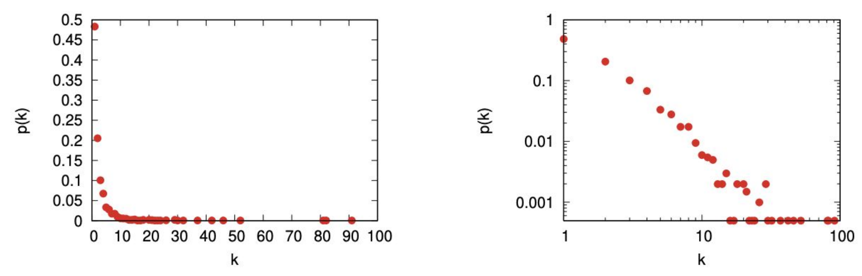


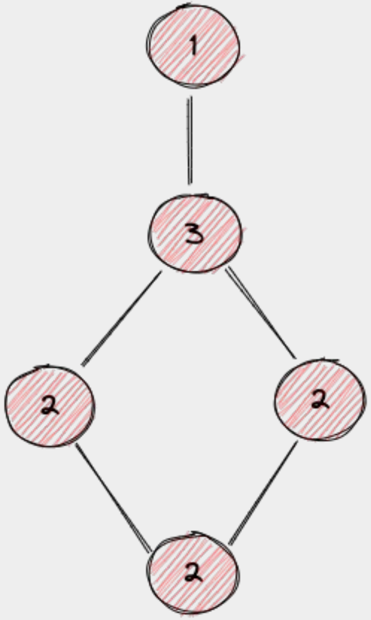
$$\langle k \rangle = \frac{2L}{N} = d(N-1)$$

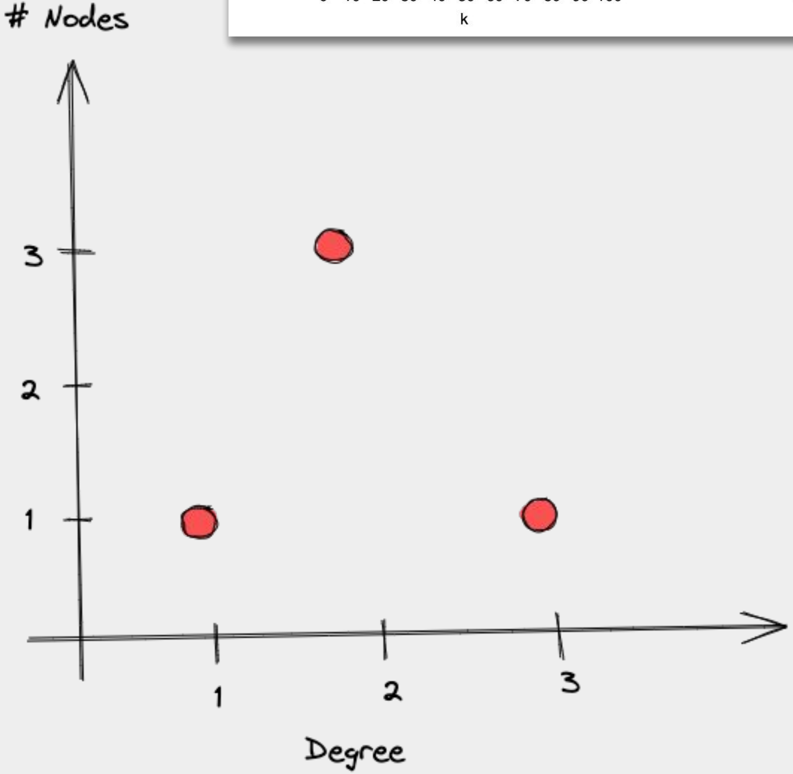
$$d = \frac{\langle k \rangle}{N-1}$$

This makes sense because the **maximum possible degree** of a node is k_{max} = *N - 1*, obtained when the node is connected to every other node. Intuitively, the **density is the ratio between the average and maximum degree**.

Degree Distribution

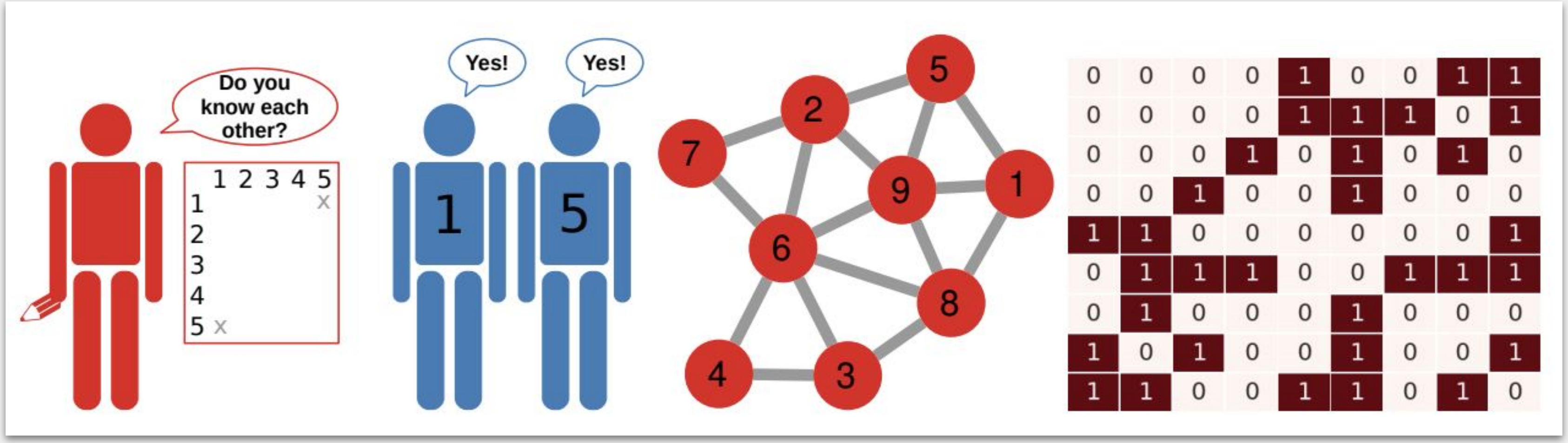






Network Representation

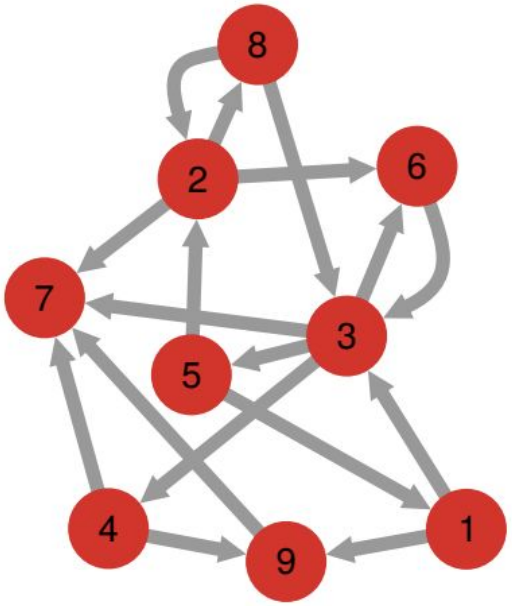
The **adjacency matrix** is the basic representation of a graph as a **matrix** . Each row/column corresponds to a node. Each cell represents an edge, set to one if the edge exists, and zero otherwise.



Network Representation

A non-symmetric adjacency matrix corresponding to a directed graph

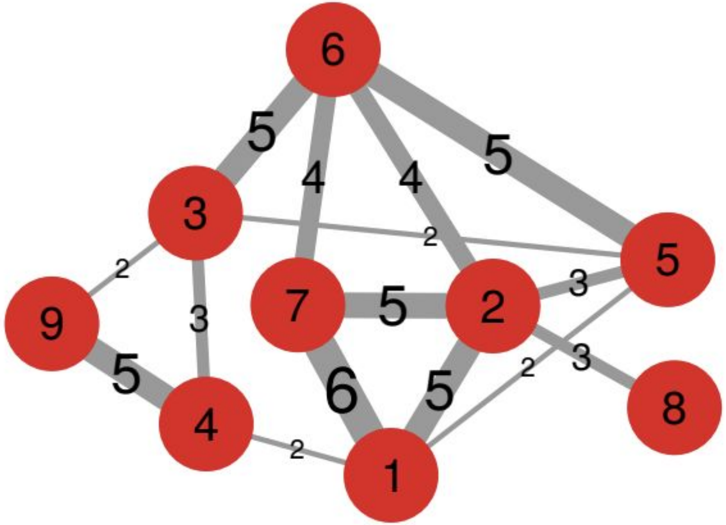
0	0	1	0	0	0	0	0	1
0	0	0	0	0	1	1	1	0
0	0	0	1	1	1	1	0	0
0	0	0	0	0	0	1	0	1
1	1	0	0	0	0	0	0	1
0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0



Network Representation

A non-binary adjacency matrix corresponding a weighted graph

0	5	0	2	2	0	6	0	0
5	0	0	0	3	4	5	3	0
0	0	0	3	2	5	0	0	2
2	0	3	0	0	0	0	0	5
2	3	2	0	0	5	0	0	0
0	4	5	0	5	0	4	0	0
6	5	0	0	0	4	0	0	0
0	3	0	0	0	0	0	0	0
0	0	2	5	0	0	0	0	0

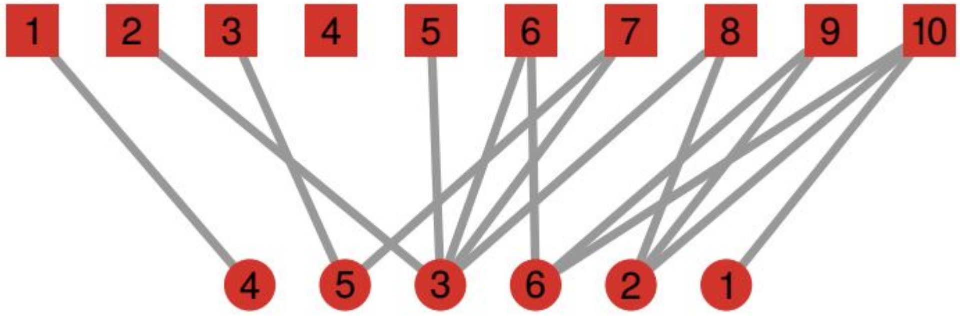


Network Representation

$$\begin{array}{c|cc} & |V_1| & |V_2| \\ \hline |V_1| & 0 & A \\ |V_2| & A^T & 0 \end{array}$$

A non-square adjacency matrix corresponding a bipartite graph

0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	1	1	1
0	1	0	0	1	1	1	1	0	0
1	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	1	0	0	0
0	0	0	0	0	1	0	0	1	1



Contact

Mailing list
Issue tracker
Source

Releases

Stable (notes)

3.3 — April 2024
Documentation

Latest (notes)

3.4 development
Documentation

Archive

NetworkX

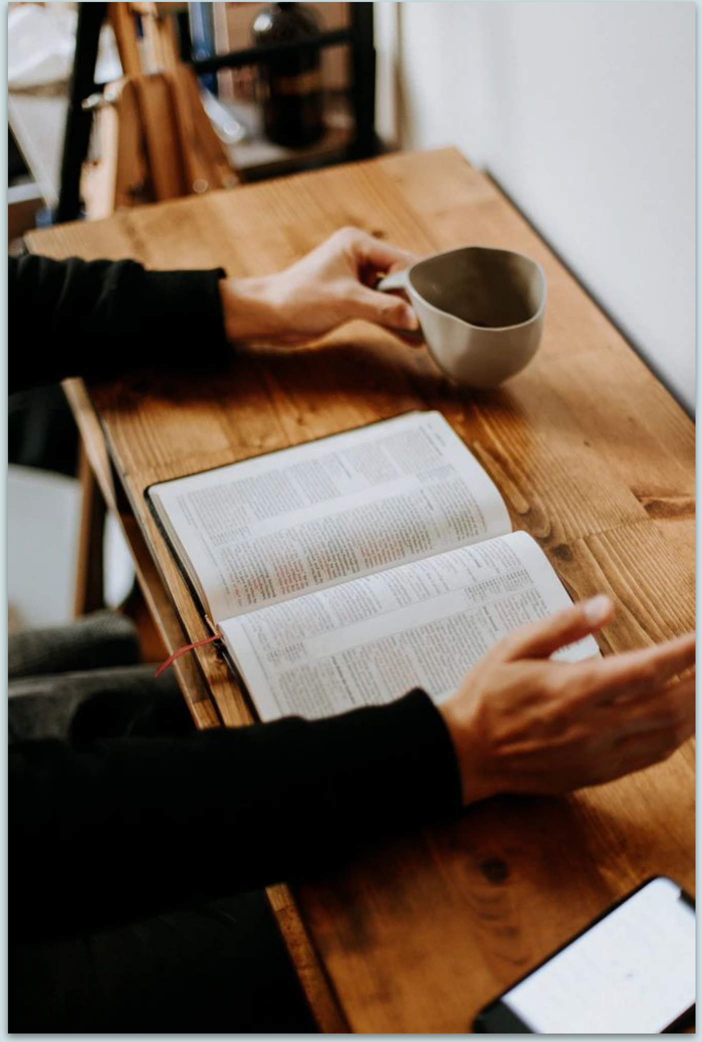
Network Analysis in Python

NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.

Software for complex networks

- Data structures for graphs, digraphs, and multigraphs
- Many standard graph algorithms
- Network structure and analysis measures
- Generators for classic graphs, random graphs, and synthetic networks
- Nodes can be "anything" (e.g., text, images, XML records)
- Edges can hold arbitrary data (e.g., weights, time-series)
- Open source 3-clause BSD license
- Well tested with over 90% code coverage
- Additional benefits from Python include fast prototyping, easy to teach, and multi-platform

https://networkx.org/



Further Reading

1 *Introduction* 9

I Basics 21

2 *Probability Theory* 22

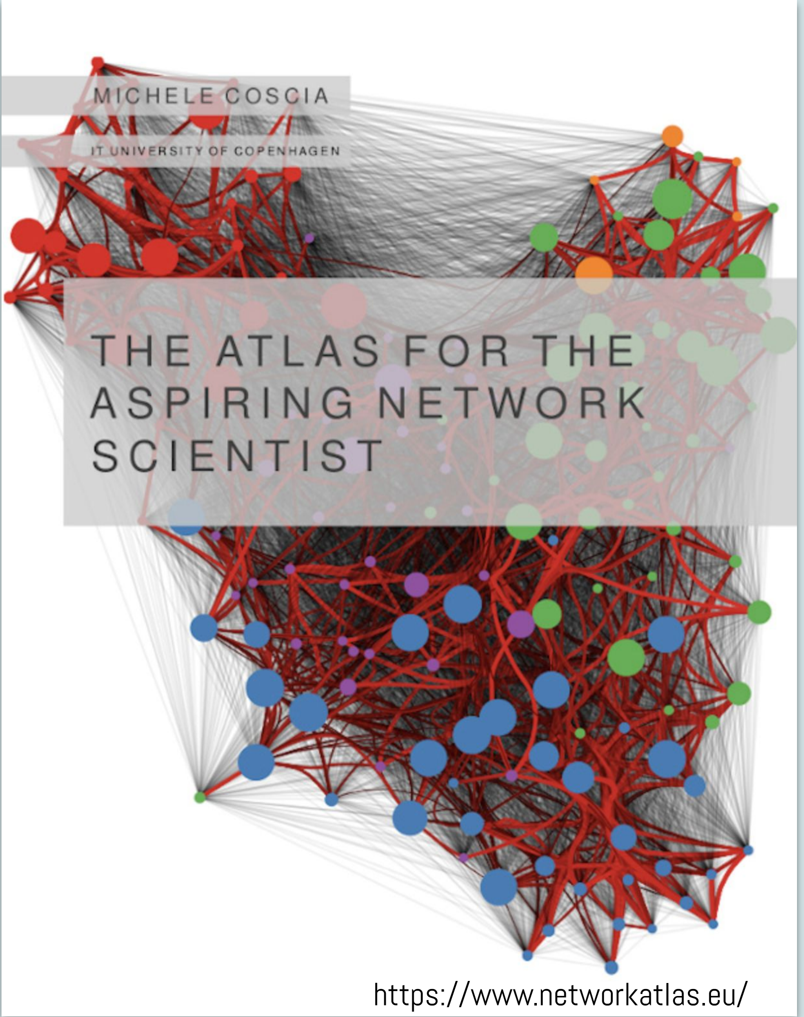
3 *Basic Graphs* 40

4 *Extended Graphs* 48

5 *Matrices* 68

II Simple Properties 89

6 *Degree* 90



Case Study

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

COVID-19: A scholarly production dataset report for research analysis

Breno Santana Santos[a,b,*], Ivanovitch Silva[a], Marcel da Câmara Ribeiro-Dantas[c], Gisliany Alves[a], Patricia Takako Endo[d], Luciana Lima[a]

[a] Universidade Federal do Rio Grande do Norte (UFRN), Rio Grande do Norte, Brazil
[b] Núcleo de Pesquisa e Prática em Inteligência Competitiva (NUPIC), Universidade Federal de Sergipe (UFS), Itabaiana, SE, Brazil
[c] Institut Curie (UMR168), Sorbonne Université (EDITE), Paris, France
[d] Universidade de Pernambuco (UPE), Pernambuco, Brazil

ARTICLE INFO

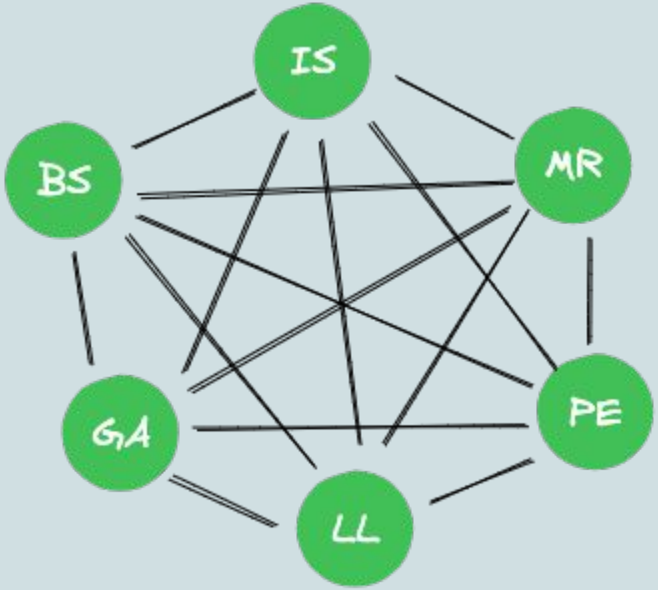
Article history:
Received 7 July 2020
Revised 6 August 2020
Accepted 12 August 2020
Available online 19 August 2020

Keywords:
COVID-19
SARS-CoV-2
Pandemic
Data Science
Bibliometrics
Scientometrics

ABSTRACT

COVID-2019 has been recognized as a global threat, and several studies are being conducted in order to contribute to the fight and prevention of this pandemic. This work presents a scholarly production dataset focused on COVID-19, providing an overview of scientific research activities, making it possible to identify countries, scientists and research groups most active in this task force to combat the coronavirus disease. The dataset is composed of 40,212 records of articles' metadata collected from Scopus, PubMed, arXiv and bioRxiv databases from January 2019 to July 2020. Those data were extracted by using the techniques of Python Web Scraping and preprocessed with Pandas Data Wrangling. In addition, the pipeline to preprocess and generate the dataset are versioned with the Data Version Control tool (DVC) and are thus easily reproducible and auditable.

Coauthorship networks



Universidade Federal do Rio Grande do Norte

Campus Universitário Lagoa Nova, Caixa postal 1524, Natal, RN, Brazil 60023857

25,807 Documents

14,665 Authors

Give feedback

Documents | Structure | Collaborators | Sustainable Development Goals 2023 New

New: See at one glance Sustainable Development Goals mapped to this organisation

Sustainable Development Goals (SDGs) are specific research areas that are helping to solve real-world problems. Elsevier data science teams have built extensive keyword queries, supplemented with machine learning, to map documents to SDGs with very high precision. Times Higher Education (THE) is using Elsevier SDG data mapping as part of its Impact Rankings. More about SDGs

SDG contributions

No poverty
Goal 1
View 105 documents

Zero hunger
Goal 2
View 292 documents

Good health and well-being
Goal 3
View 3,169 documents

Quality education
Goal 4
View 379 documents

Gender equality
Goal 5
View 171 documents

Clean water and sanitation
Goal 6
View 612 documents

Affordable and clean energy
Goal 7
View 1,050 documents

Decent work and economic growth
Goal 8
View 263 documents

Industry, innovation and infrastructure
Goal 9
View 472 documents

Reduced inequalities
Goal 10
View 285 documents

Sustainable cities and communities
Goal 11
View 447 documents

Responsible consumption and production
Goal 12
View 377 documents

Climate action
Goal 13
View 270 documents

Life below water
Goal 14
View 572 documents

Life on land
Goal 15
View 623 documents

Peace, justice and strong institutions
Goal 16
View 205 documents

Partnership for the goals
Goal 17
View 3,099 documents

Scopus®

ELSEVIER

Elsevier Data Repository

Elsevier 2023 Sustainable Development Goals (SDGs) Mapping

Published: 13 July 2023 | Version 1 | DOI: 10.17632/y2zyy9vwzy.1
Contributors: Alexandre Bedard-Vallee, Chris James, Guillaume Roberge

Description

The United Nations Sustainable Development Goals (SDGs) challenge the global community to build a world where no one is left behind.

Since 2018, Elsevier has generated SDG search queries to help researchers and institutions track and demonstrate progress toward the SDG targets. In the past 5 years, these queries, along with the university's own data and evidence supporting progress and contributions to the particular SDG outside of research-based metrics, are used for the THE Impact Rankings.

For 2023, the SDGs use the exact same search query and ML algorithm as the Elsevier 2022 SDG mappings, with only minor modifications to five SDGs, namely SDG 1, 4, 5, 7 and 14. In these cases, the queries were shortened by removing exclusion lists based on journal identifiers. These exclusion lists often contained thousands of items to filter out content in journals that were not core to the SDGs.

To replicate the effect of these journal exclusions, sets of keywords were used to closely mimic the effects the journal exclusions had on the SDG content, while greatly reducing the overall query size and complexity. By following this approach, we were able to limit the changes to the publications in each SDG by less than 2 percent for most SDGs, while reducing the query size by 50 percent or more.

These shortened queries also have the added benefit of running faster in Scopus, allowing further analysis of the SDG data to be done more easily.

Files

SDG 2023 Queries

SDG Descriptions

Institutions

Elsevier BV

Dataset metrics

Latest version

Version 1
Published: 13 Jul 2023
DOI: 10.17632/y2zyy9vwzy.1

https://elsevier.digitalcommonsdata.com/datasets/y2zyy9vwzy/1