University of Stuttgart
Germany

# Complex Network Systems

Graphs

Ilche Georgievski
ilche.georgievski@iaas.uni-stuttgart.de
Room: U38 0.353

2019/2020
Winter

# Network vs Graph

# What is a network?

A network is a collection of objects where
some pairs of objects are connected by links

Web, power grid, social network

Language: network, node, link

# Exercise

Draw a network representing your family tree

# Is that a complex network?

# Simple networks

- Regular or almost regular structure
  - except for the root, any other tree node has exactly one antecedent

- Small network does not imply simple network
  - the humankind "family tree" has billions of persons

# Complex networks

- Non-trivial structure
  - Patterns of connections are neither regular nor entirely random

- Properties

# What is a graph?

A graph is a mathematical representation of a network

Web graph, power grid graph, social graph

Language: graph, vertex, edge

# Network vs Graph

We will make this distinction whenever necessary,
but most often, we will use the two terms
interchangeably

# How to define a network?

How to build a graph?

What are the vertices?

What are the edges?

**The choice of a proper network representation of a given domain determines our ability to use networks successfully**

Unique and unambiguous representation

Links determine the question to be explored

Not so unique representation

- If we connect individuals that work with each other, we explore a _____
- If we connect scientific papers that cite each other, we explore the _____
- If we connect all papers with the same word in the title, what will we be exploring?

# Healthy nutrition domain

Foods that provide naturally occurring nutrients*:

- Vitamin D: Fatty fish, mushrooms (vitamin D is naturally formed in the body when skin is exposed to sunlight; vitamin D is added to fortified milk)
- Vitamin E: Nuts, seeds, vegetable oils, green leafy vegetables
- Magnesium: Whole grains, wheat bran and wheat germ, green leafy vegetables, legumes, nuts, seeds
- Vitamin A: Preformed vitamin A: liver, fatty fish, milk, eggs; provitamin A carotenoids: carrots, pumpkins, tomatoes, leafy green vegetables
- Calcium: Milk, yogurt, cheese, kale, broccoli
- Vitamin C: All fruits and vegetables, particularly citrus fruits and tomatoes
- Vitamin B6: Highest levels in fish, beef, poultry, potatoes and other starchy vegetables, and fruit other than citrus
- Folate: Many foods; highest levels in spinach, liver, asparagus, Brussels sprouts (mandatory, standardized addition to enriched flour and flour products)
- Zinc: Red meat, poultry, beans, nuts, some seafood, whole grains
- Thiamin: Whole grain products (mandatory, standardized addition to enriched flour and flour products)
- Copper: Shellfish, whole grains, beans, nuts, potatoes, organ meats (kidneys, liver)
- Vitamin B12: Animal products: fish, meat, poultry, eggs, milk
- Riboflavin: Milk and dairy products, eggs, meat, green leafy vegetables, legumes (mandatory, standardized addition to enriched flour and flour products)
- Niacin: Meat, fish, seeds and nuts, whole grains (mandatory, standardized addition to enriched flour and flour products)
- Selenium: Found in different plant and animal foods; highest levels in seafood and organ meats (kidneys, liver)

*pragprog.com/titles/dzcnapy/source_code

# Exercise

Draw a network representing the healthy nutrition domain

# Basic steps

1. Identify nodes
2. Identify relationships
3. Transform the list of items into a network
   - *Snowballing*

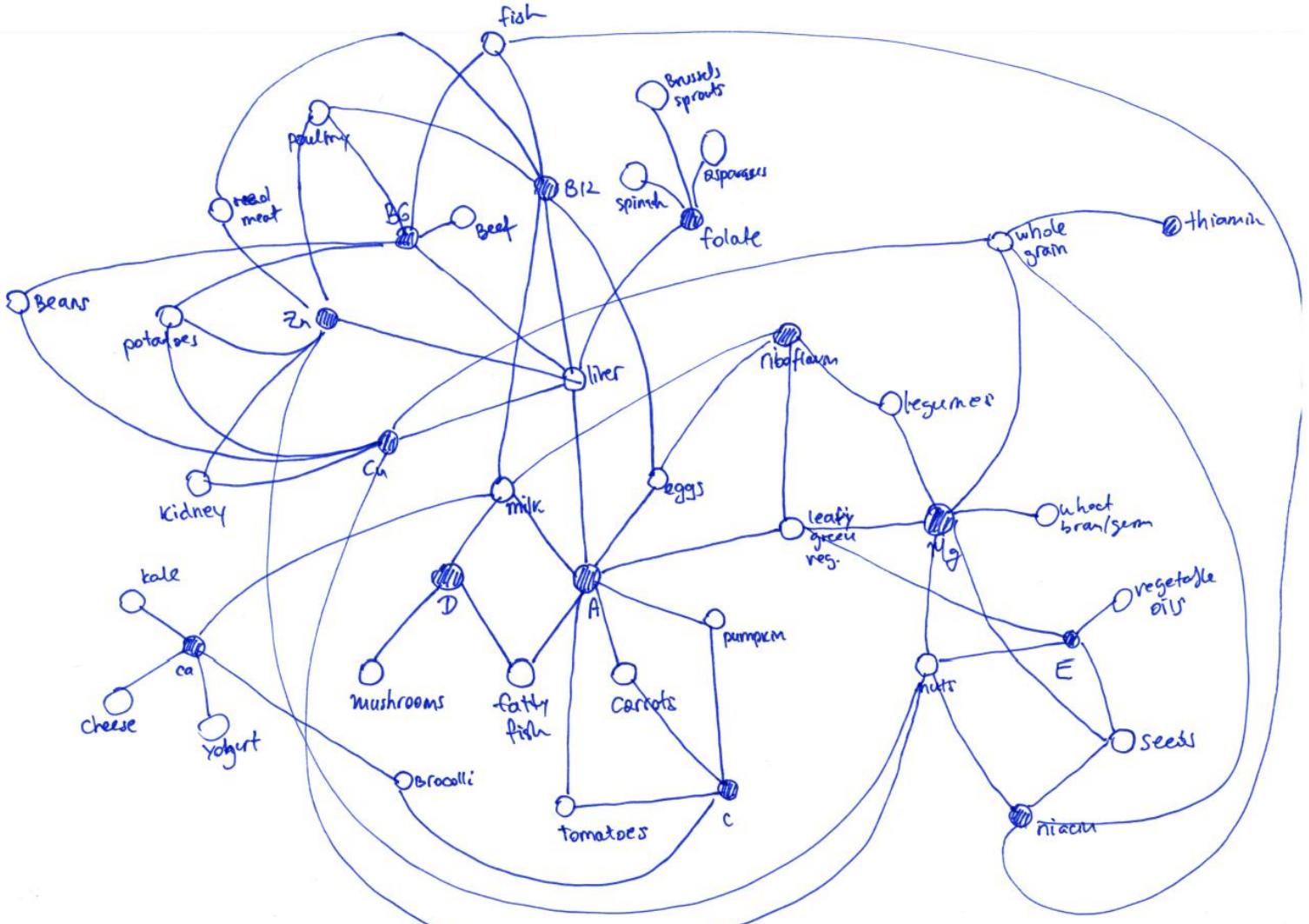
# Snowballing

- Start with a "seed" node and follow the links to discover other nodes
- Network starts with a single snowflake and grows over time until either you are satisfied with its size or there is no more "snow" to add
- Depending on the network size, may overlook some network parts if improper seed is chosen
- It might be the best to choose several seeds and follow all links originating from them

# Snowballing for healthy nutrition

- Choose the first … from the list, e.g., vitamin D. Draw a circle and label it "D"
- Vitamin D is provided by fatty fish. Draw a circle that represents fatty fish, label it, and connect to the D node.
- Vitamin D is also provided by mushrooms. Draw a circle …
- Repeat the previous steps for each combination of food and nutrients
- Do not duplicate nodes

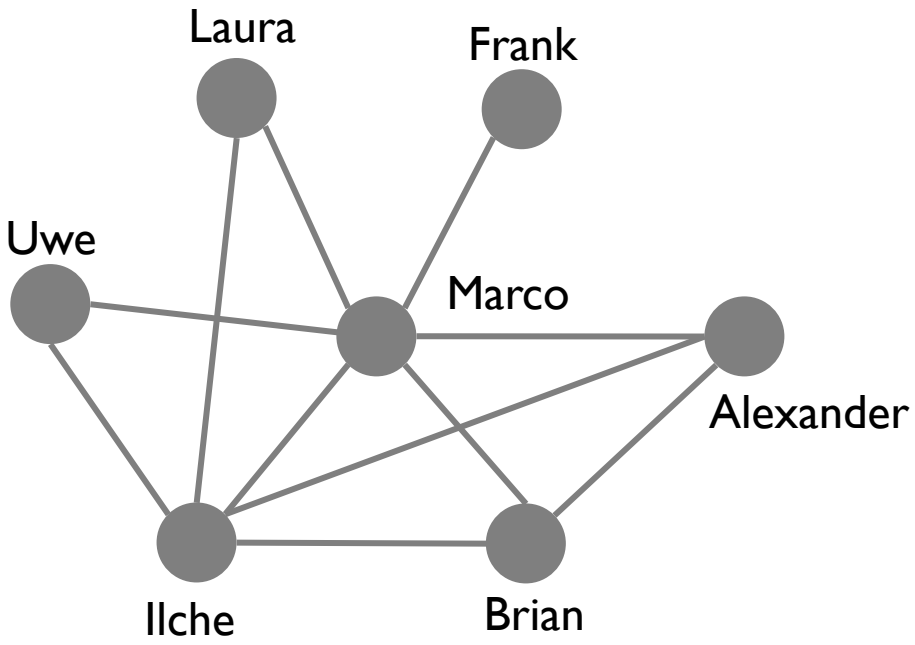
# My healthy nutrition network



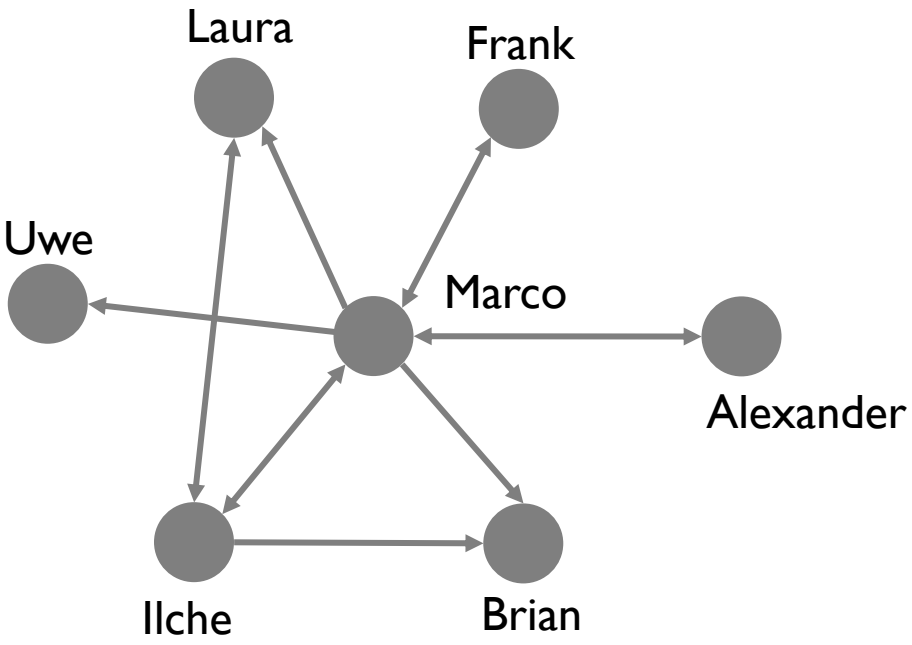
# What type of graph?

# Undirected vs directed

## Collaborations

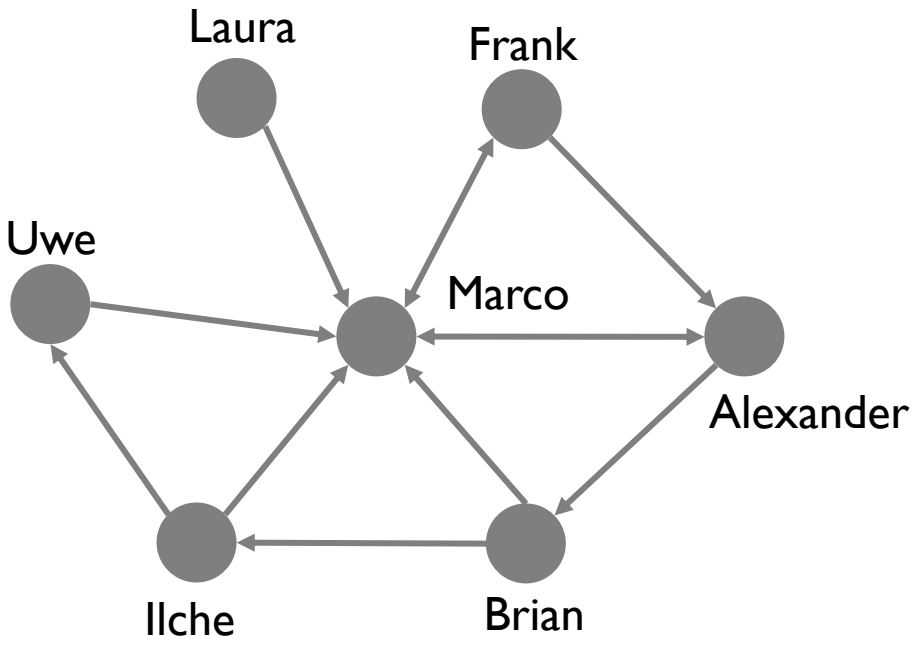



## Phone calls

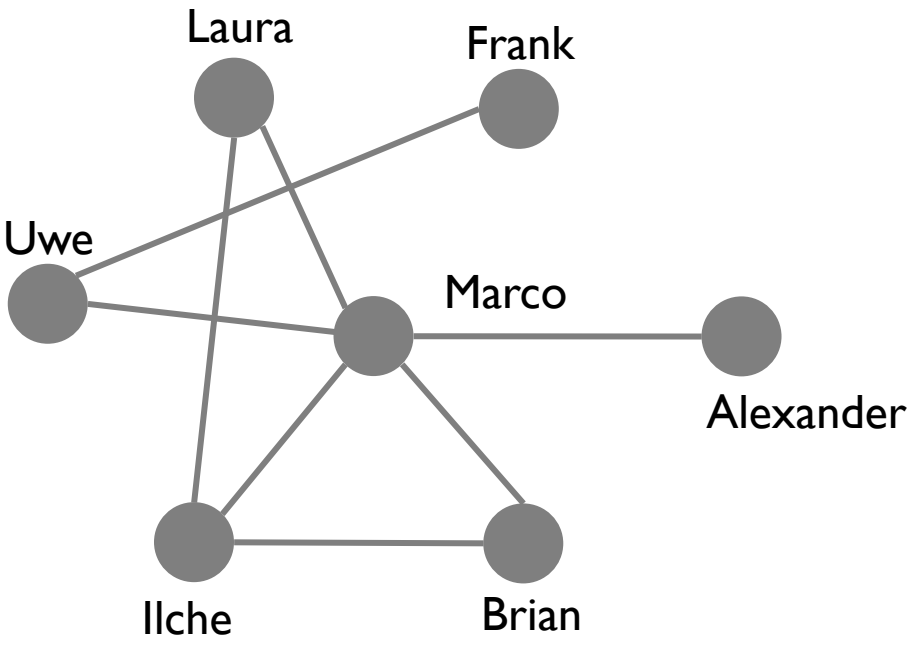


# Undirected vs directed

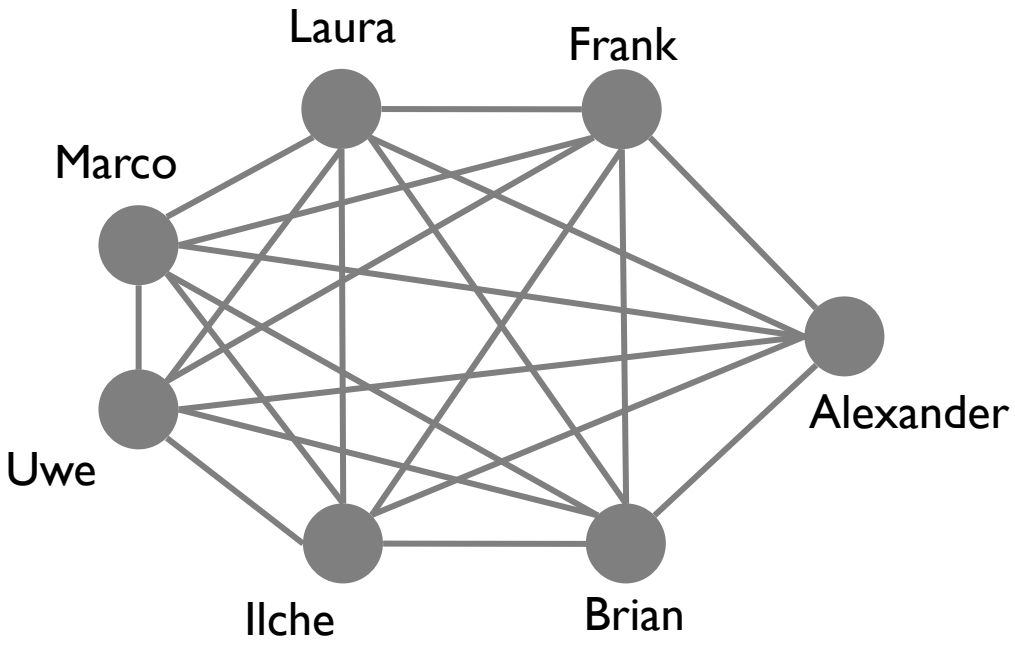
## Following on Twitter



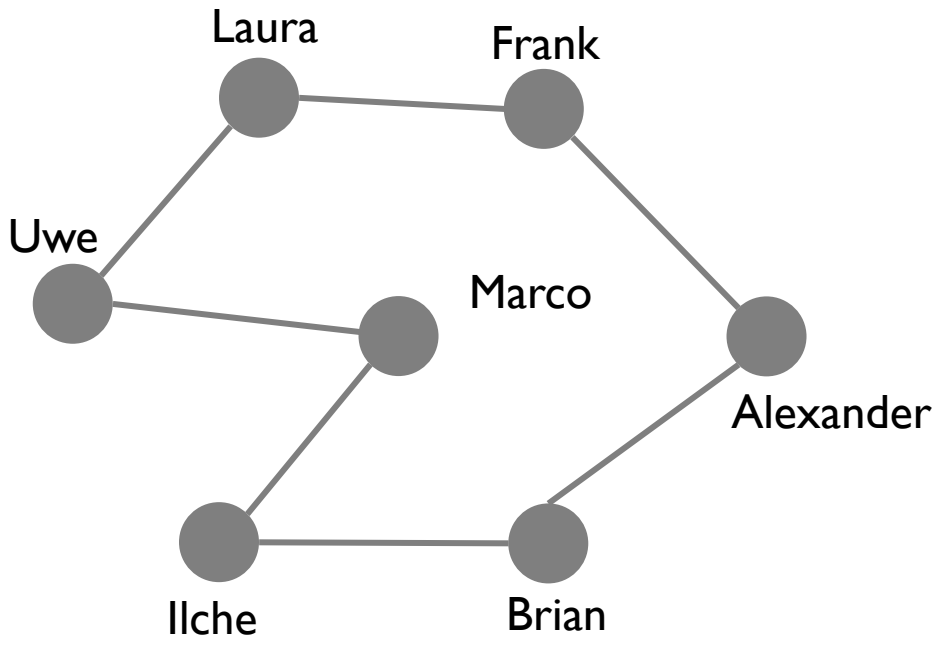

## Friendship on Facebook



# Complete graph



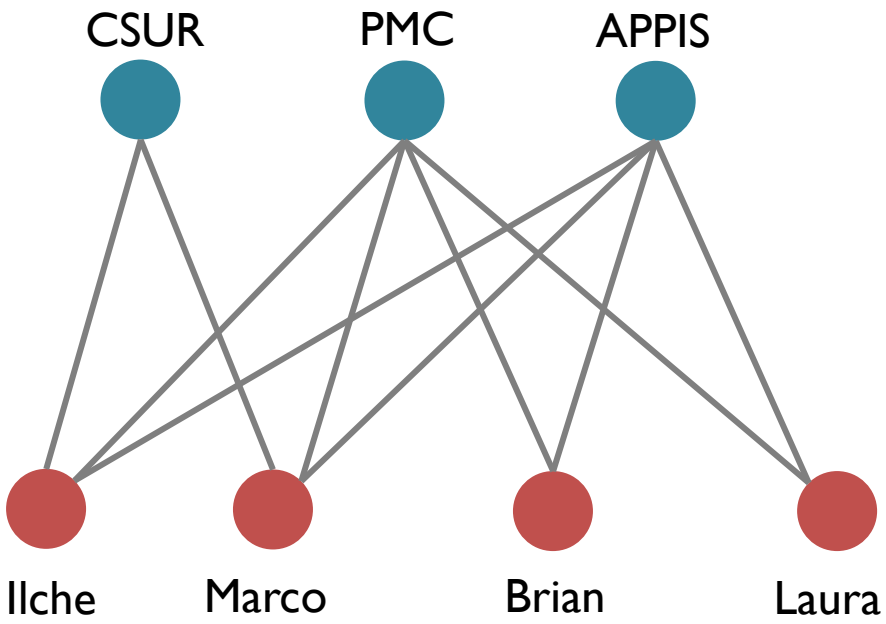

# Regular graph




What about a directed case?

# Bipartite graph

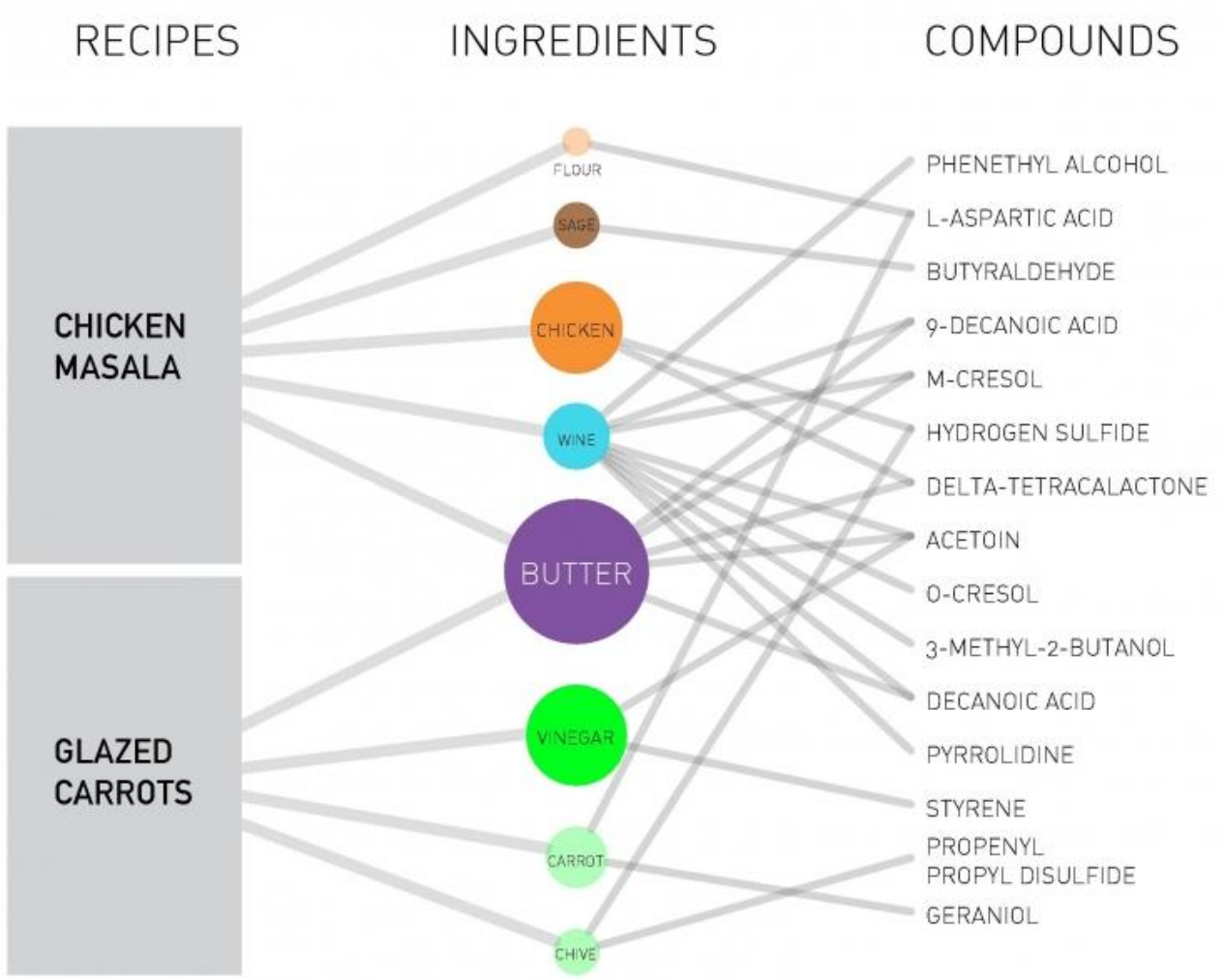
Persons that authored papers




Authors-to-papers network

# NYTIMES Human Disease Network

# Tripartite graph



# Unweighted vs weighted

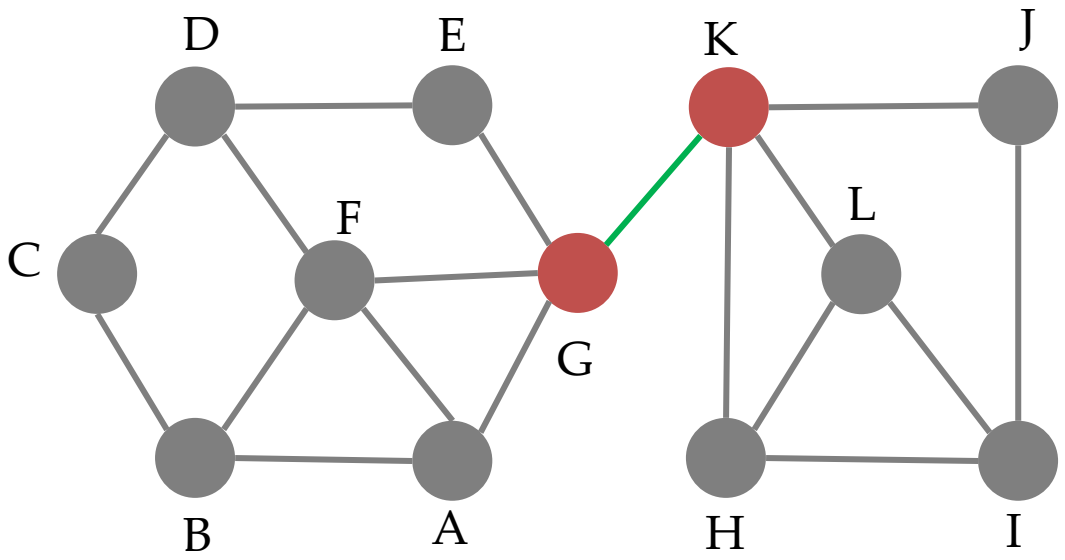
- Which of the following networks are weighted graphs?
  - Friendships on Facebook
  - Internet
  - Collaborations
  - Hyperlink
  - Roads

# Unweighted vs weighted

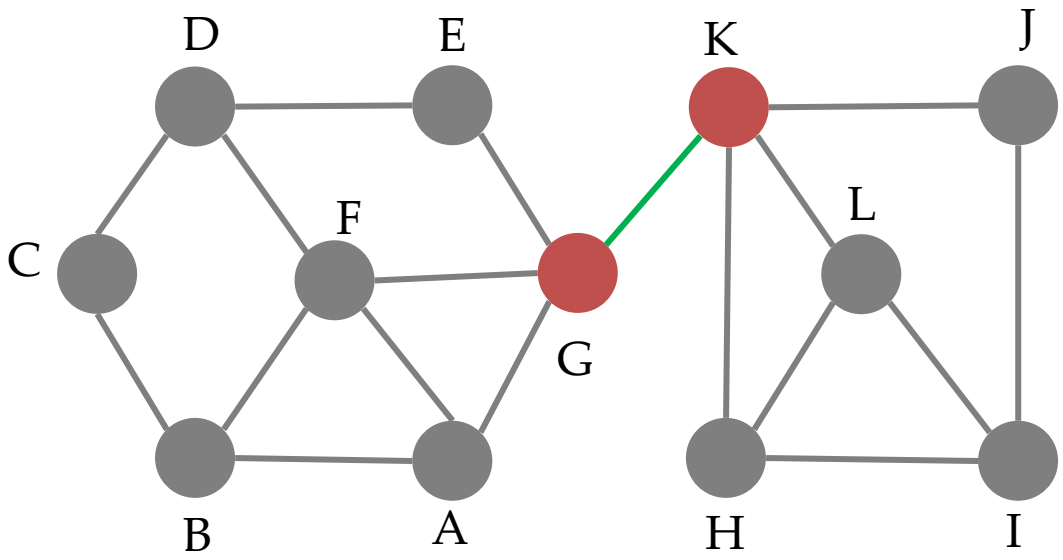
- Which of the following networks are weighted graphs?
  - Friendships on Facebook
  - Internet
  - Collaborations
  - Hyperlink
  - Roads

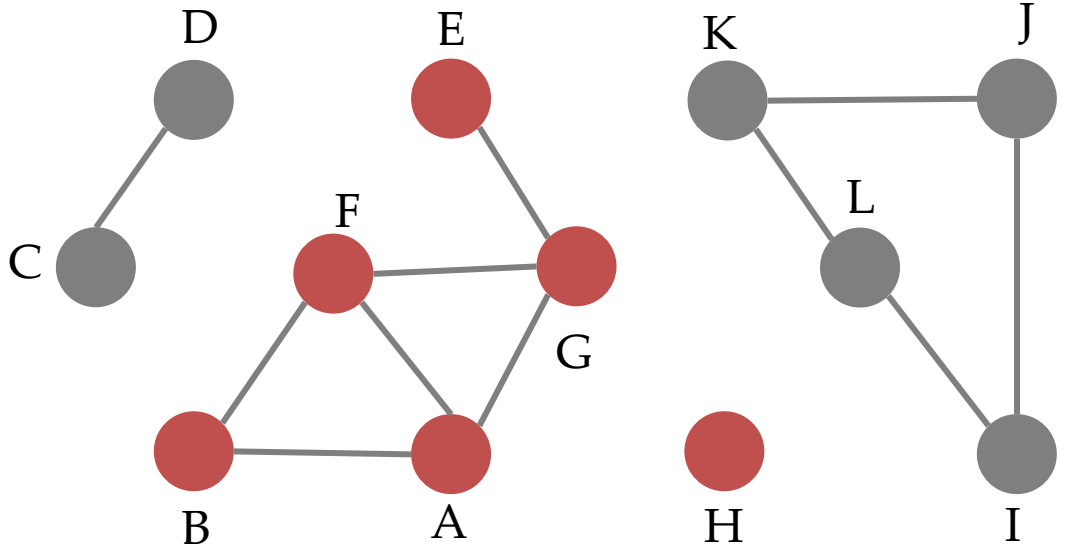
- What could their weights be?

# Connected graph

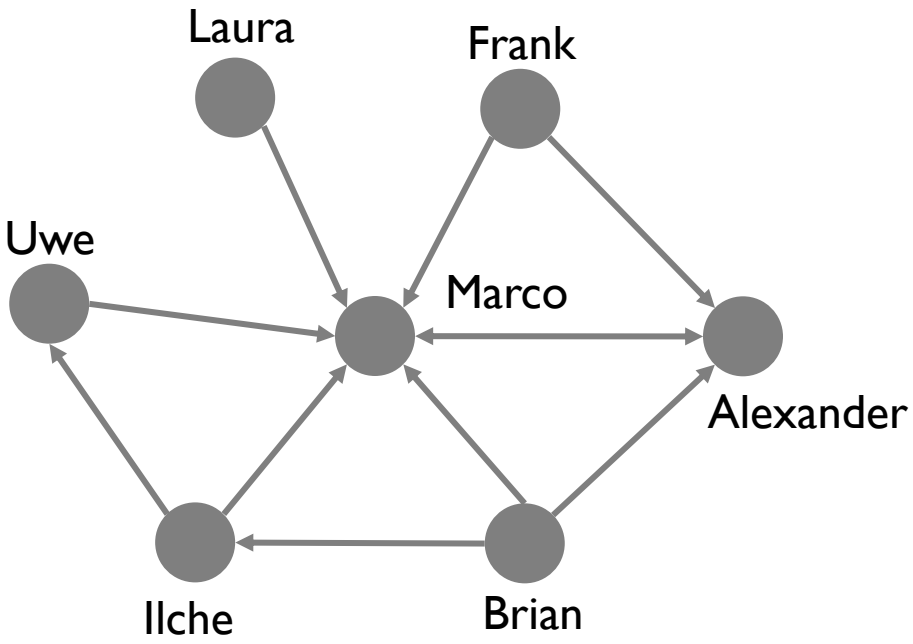


# Connected graph

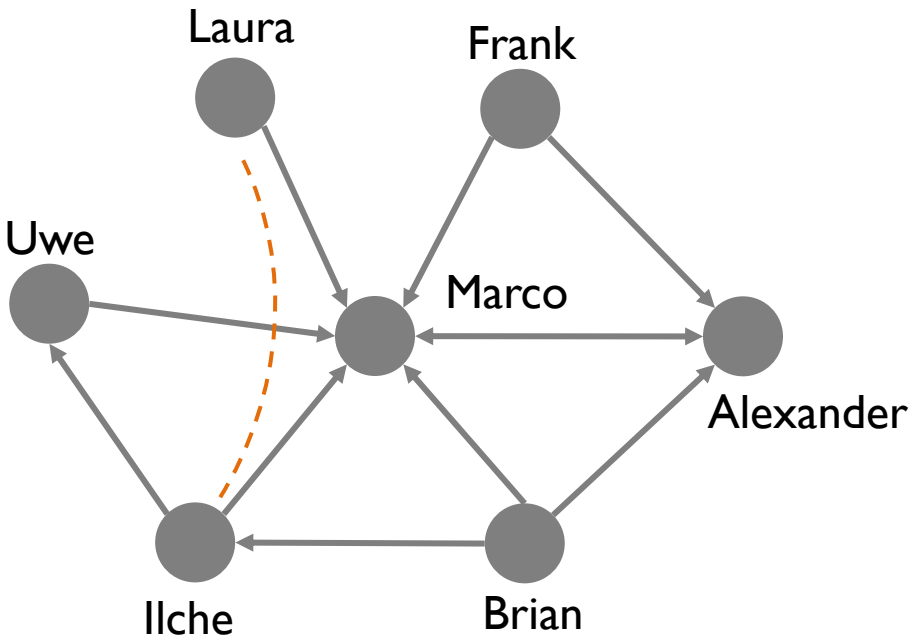


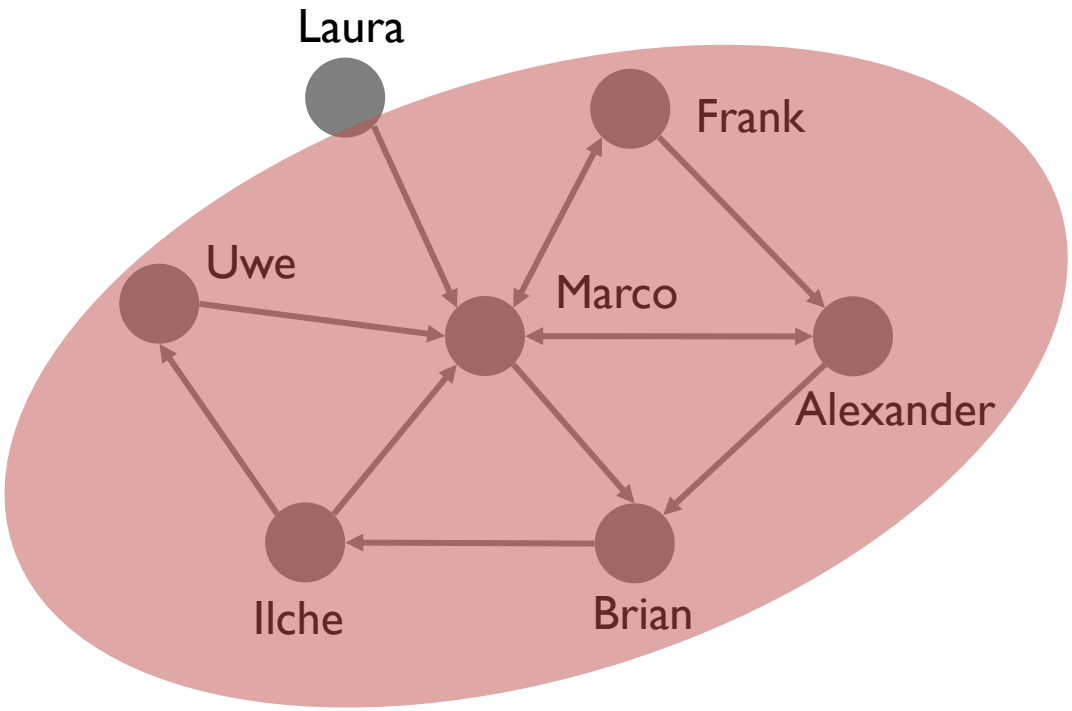
# Strongly vs weakly connected directed graph



# Strongly vs weakly connected directed graph



# Strongly connected components



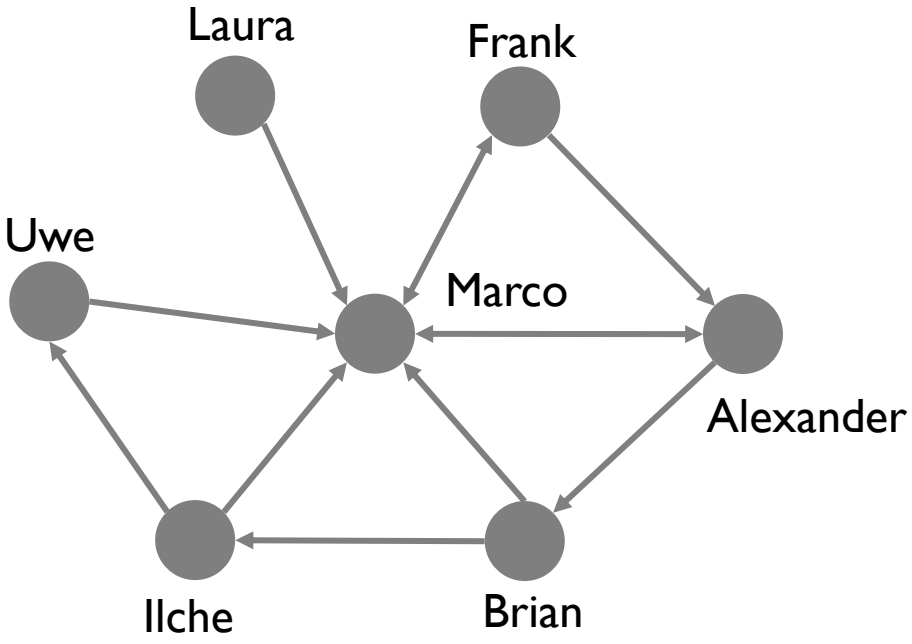
# What type of representation?

- Dots and lines
- Incidence matrix
- Adjacency matrix
- Edge list
- Adjacency list

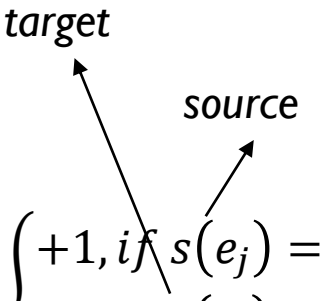
# Incidence matrix

(directed)

$G$




$$B = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix}$$

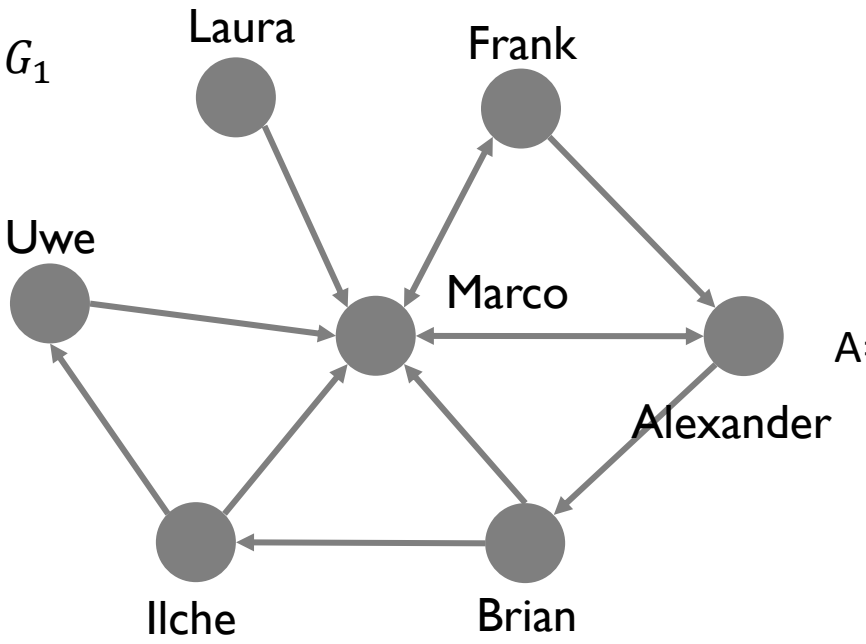



$$b_{ij} = \begin{cases} +1, if\ s(e_j) = v_i \\ -1, if\ t(e_j) = v_i \\ 0, otherwise \end{cases}$$

$$v_i \in V$$
$$e_j \in E$$
$$\mathrm{G} = (\mathrm{V}, \mathrm{E})$$

# Adjacency matrix

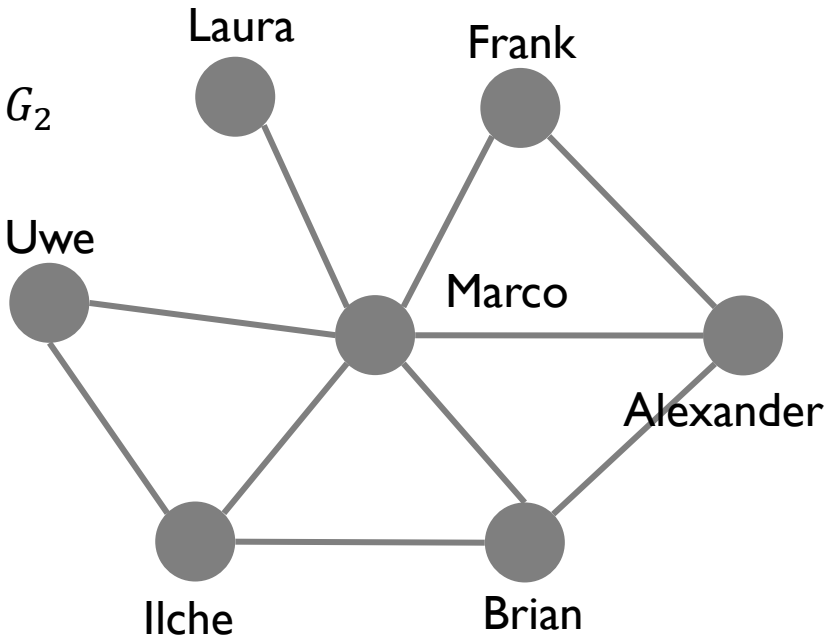
(directed and undirected)




$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$k_i^{out} = \sum_{j=1}^{n} A_{ij} \qquad k_j^{in} = \sum_{i=1}^{n} A_{ij}$$

$$a_{ij} = \begin{cases} 1, & \text{if there is some edge between } v_i \text{ and } v_j \\ 0, & \text{otherwise} \end{cases}$$

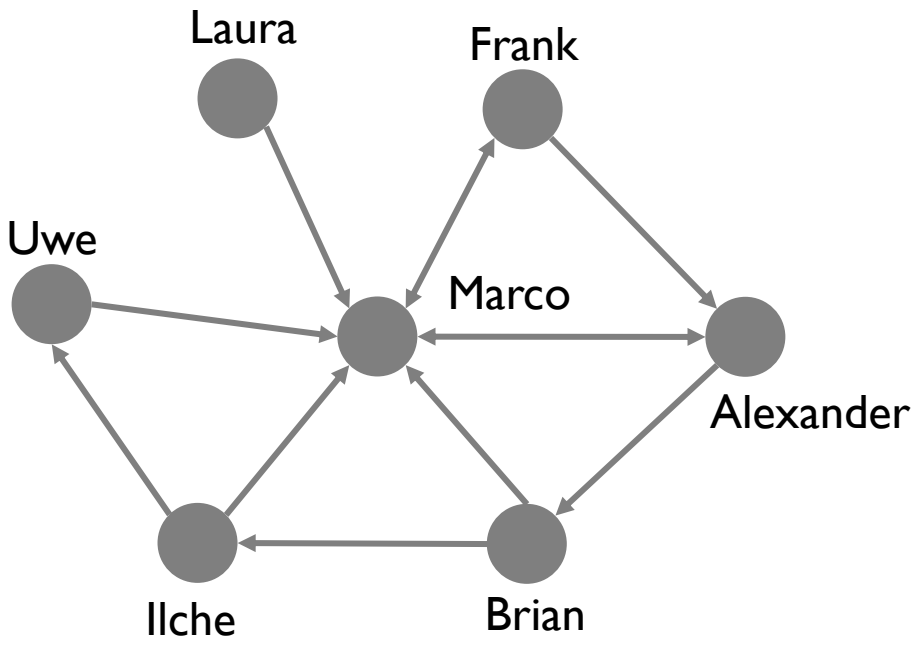



$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

$$v_i, v_j \in V$$
$$(v_i, v_j), (v_j, v_i) \in E$$
$$\mathrm{G} = (\mathrm{V}, \mathrm{E})$$

$$k_i = \sum_{j=1}^{n} A_{ij} \qquad k_j = \sum_{i=1}^{n} A_{ij}$$

# Edge list vs adjacency list




- (L,M)
- (F,M)
- (F,A)
- (A,M)
- (A,B)
- (B,M)
- (B,I)
- ….

- L: M
- F: M, A
- A: M, B
- B: M, I
- I: M, U
- U: M
- M: A, F

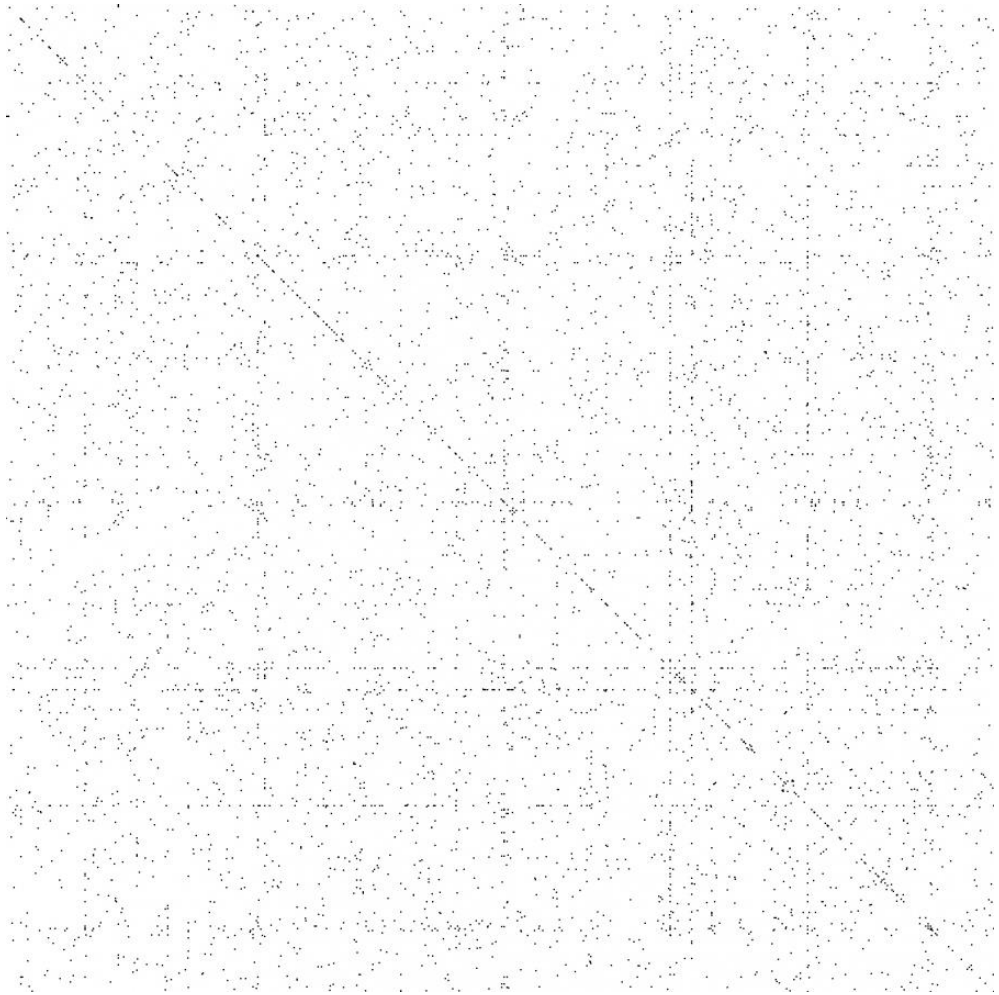
Easier to work with when the network is large and/or sparse

# Most real-world networks are sparse

| Network | Number of nodes | Average degree |
|---|---|---|
| Social networks (LinkedIn) | 6 946 668 | 8.78 |
| Citation networks (US patents) | 3 764 105 | 8.77 |
| WWW (Stanford and Berkley) | 319 717 | 9.65 |
| Coauthorships (DBLP) | 317 080 | 6.62 |
| IMDB (German movies) | 21 258 | 3.97 |
| Roads (California) | 1 957 027 | 2.82 |
| Proteins (Yeast) | 4 626 | 6.40 |

Jure Leskovec , Kevin J. Lang , Anirban Dasgupta & Michael W. Mahoney (2009) Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters, Internet Mathematics, 6:1, 29-123.

# Yeast protein-protein interaction network

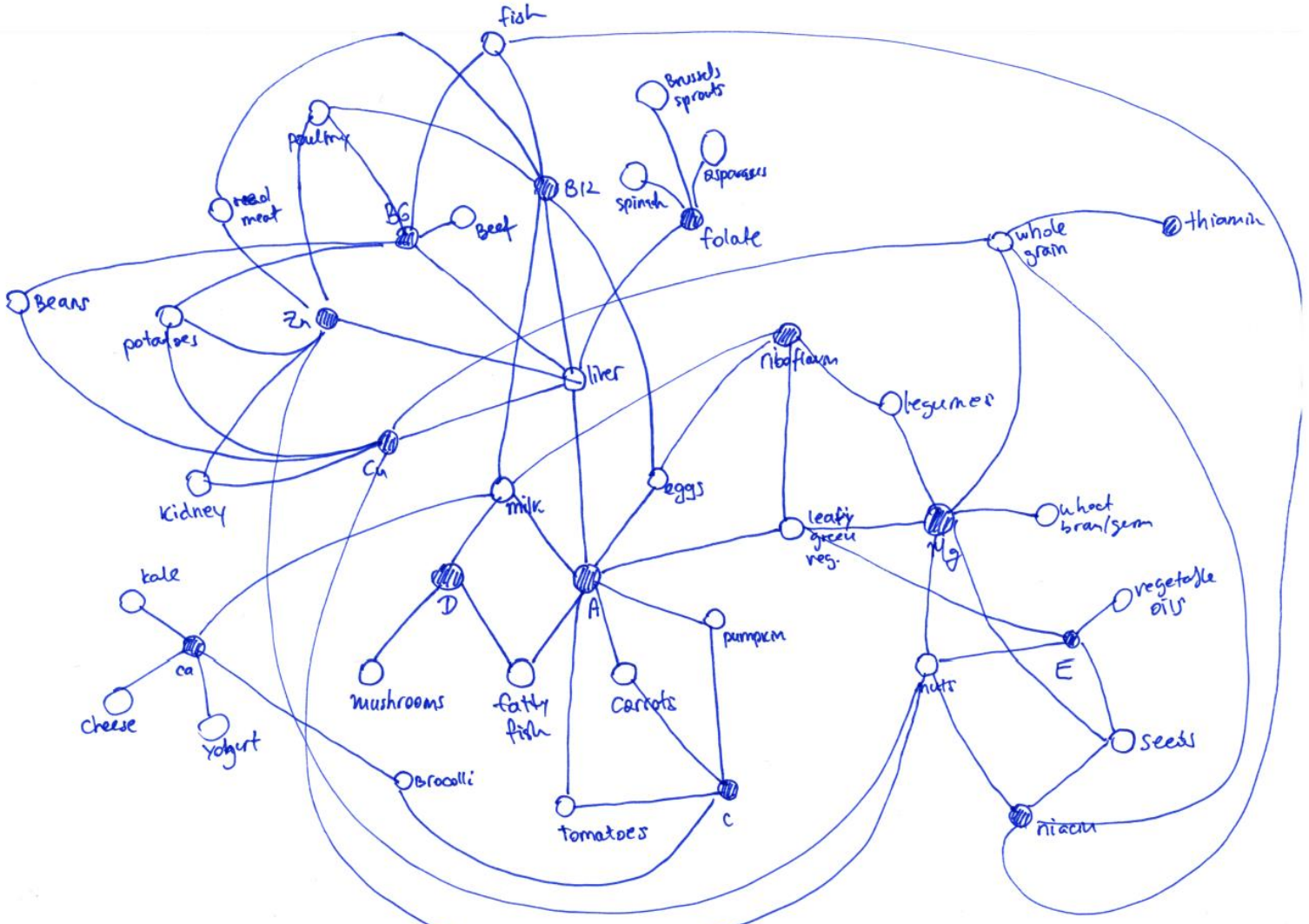


# Network representation

| | | |
|---|---|---|
| Email networks | directed | ? |
| Facebook friendships | undirected | unweighted |
| Citation networks | directed | unweighted |
| Collaboration networks | undirected | weighted |
| Mobile phone calls | directed | weighted |
| Protein interactions | undirected | unweighted |

# My healthy nutrition network

Bipartite
Directed
Weighted



# Credit

- Leskovec, J. Analysis of Networks, CS224W, Stanford University (2018), http://web.stanford.edu/class/cs224w/
- Barabási, A. Network Science, http://networksciencebook.com/
- Zinoviev, D.. Complex Network Analysis in Python: Recognize, Construct, Visualize, Analyze, Interpret, The Pragmatic Bookshelf, 2018.