Dynamic and Social Network Analysis

Lecture 3
Miray Kas
Bilkent University
Computer Engineering Department

Notations for Social Network Data

Notations

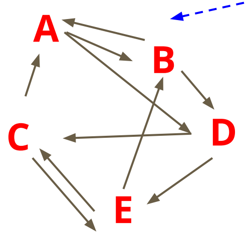
Ways for describing the social network data mathematically:

- Graph Theoretic
- Algebraic

Adjacency Matrix

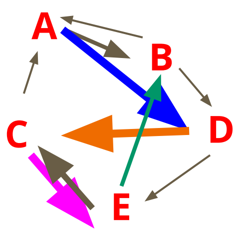
Diagram

	A	B	C	D	E
A	0	1	0	1	0
B	1	0	0	1	0
C	1	0	0	0	1
D	0	0	1	0	1
E	0	1	1	0	0



Graph

	A	B	C	D	E
A	0	2	0	5	0
B	1	0	0	1	0
C	1	0	0	0	7
D	0	0	8	0	1
E	0	2	4	0	0



Network

Graph based and algebraic notations are complementary and interchangeable!

Terminology

- Different terms to represent the *almost* same thing
  - Actor ⟺ Node ⟺ Vertex
  - Relation(ship) ⟺ Tie ⟺ Edge ⟺ Link

Graph Representation Notation

This section is partially adapted from Barabasi Network Science lecture slides.

Networks vs. Graphs

Networks* $\neq$ *Graphs

Networks vs. Graphs

network often refers to real systems

- www
- social network
- metabolic network.

Language: (Network, node, link)

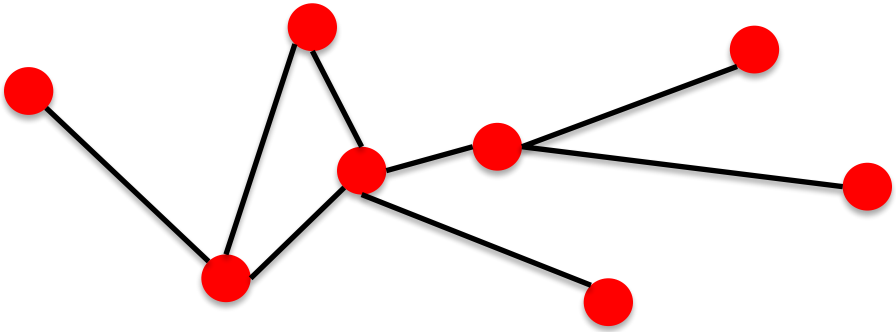
graph: mathematical representation of a network

- web graph,
- social graph (a Facebook term)

Language: (Graph, vertex, edge)

We will try to make this distinction whenever it is appropriate, but in most cases we will use the two terms interchangeably.

Components of a Complex System



- **components**: nodes, vertices   N
- **interactions**: links, edges   L
- **system**: network, graph   (N, L)

Choosing a proper representation for your question

The choice of the proper network representation determines our ability to use network theory successfully.

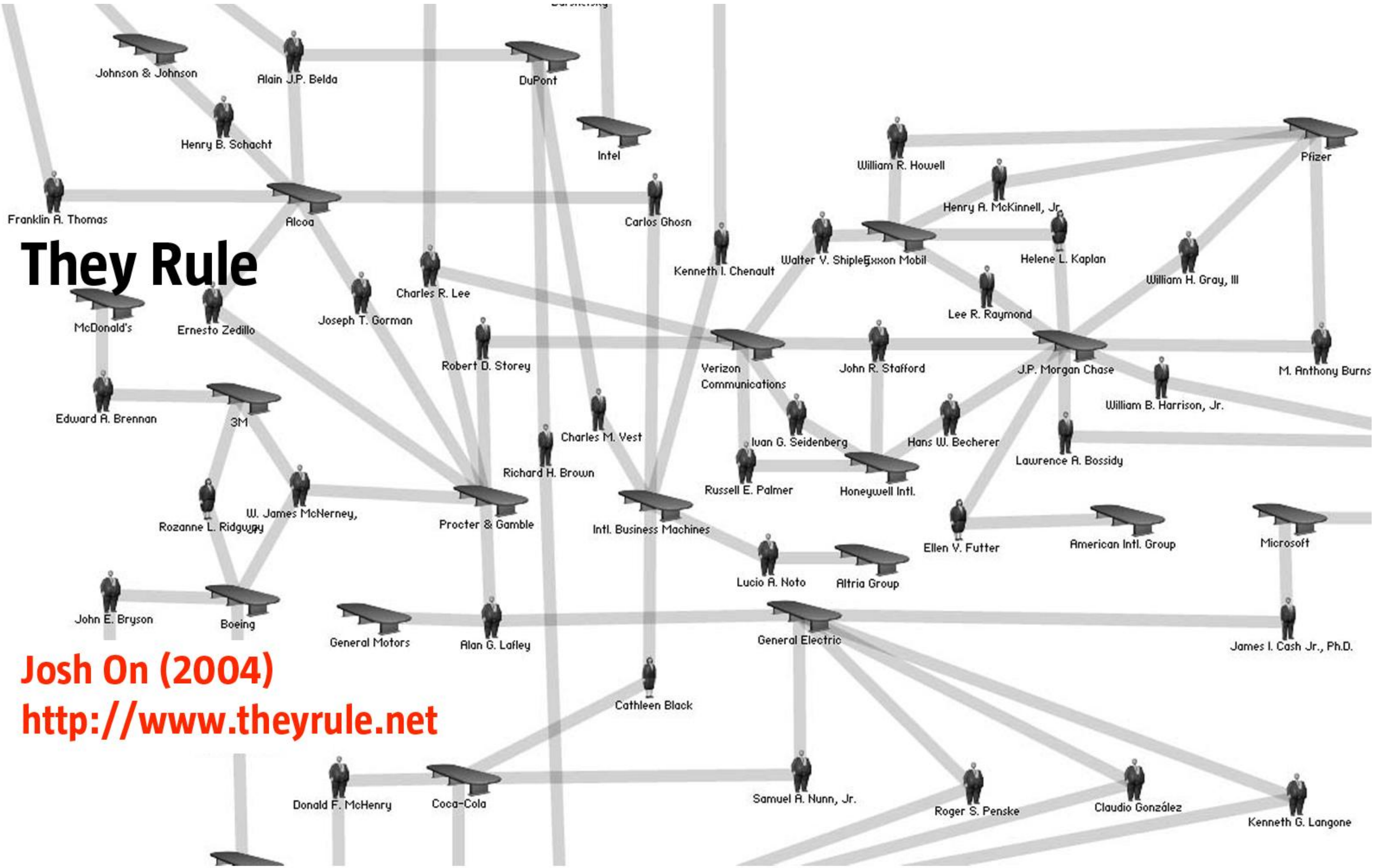
In some cases,

- There is a unique, unambiguous representation.

In other cases,

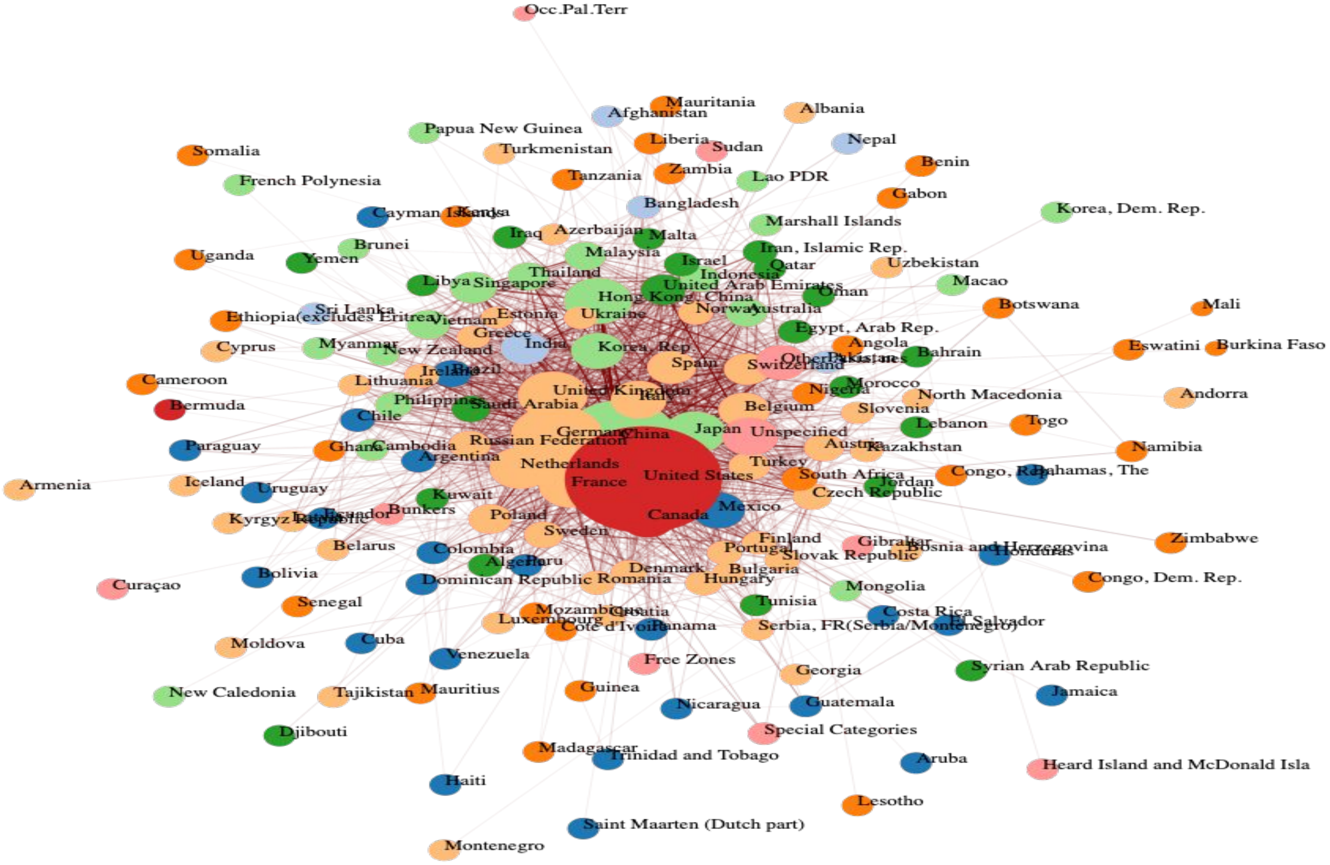
- The representation is by no means unique.

Professional Network Example



If you connect individuals that work together, you get a ***professional network***

Country Trade Network Example



If you connect countries that import/export from one another, you get a ***trade network***

(Image taken from

https://wits.worldbank.org/GlobalNetwork.aspx?lang=en)

Network of What ?!?

If you connect individuals based on their first name (*all Peters connected to each other*), you will be exploring what?

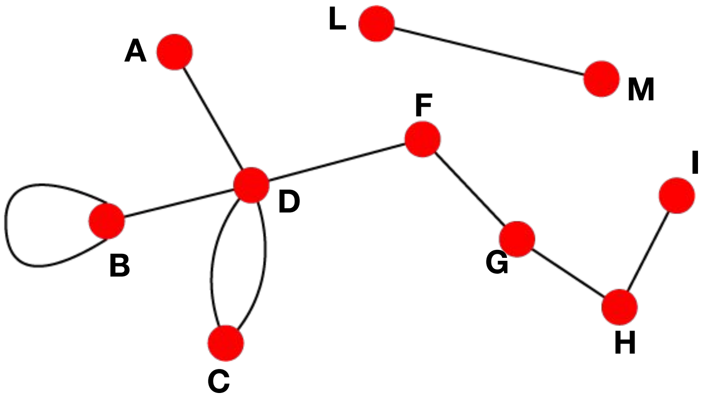
It is a network, nevertheless.

Directed vs. Undirected Networks

Undirected

Links: undirected (*symmetrical*)

Graph:



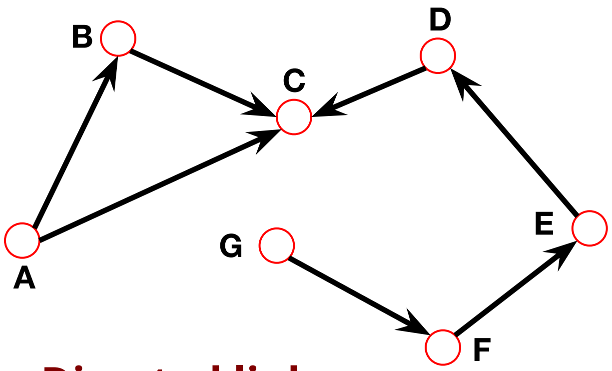
Undirected links :

- Coauthorship links
- Actor network
- Protein interactions

Directed

Links: directed (*arcs*).

Digraph = directed graph:



An undirected link is the superposition of two opposite directed links.

Directed links :

- URLs on the www
- Phone calls
- Sales networks

Directed vs. Undirected Networks

- Directed and undirected representation choices makes many things fundamentally different!

- Most of the network and node level calculations will be heavily influenced by this difference!

Let's construct a network! [Pen & Paper]

	Actor	Lives Near
n1	Allison	Ross, Sarah
n2	Drew	Elliot
n3	Eliot	Drew
n4	Keith	Ross, Sarah
n5	Ross	Allison, Keith, Sarah
n6	Sarah	Allison, Keith, Ross

Using the info you are given,

1) Get the links (e.g. edge list)
2) Draw the network

Let's construct a network! [Pen & Paper]

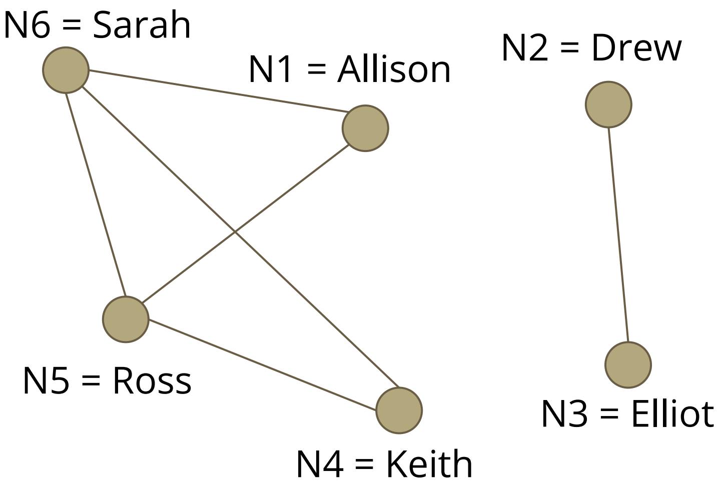
ID	Actor (Label)	Lives Near
n1	Allison	Ross, Sarah
n2	Drew	Elliot
n3	Eliot	Drew
n4	Keith	Ross, Sarah
n5	Ross	Allison, Keith, Sarah
n6	Sarah	Allison, Keith, Ross

Edge List
L1 = (n1, n5)
L2 = (n1, n6)
L3 = (n2, n3)
L4 = (n4, n5)
L5 = (n4, n6)
L6 = (n5, n6)

Let’s construct a network! [Pen & Paper]

	Actor	Lives Near
n1	Allison	Ross, Sarah
n2	Drew	Elliot
n3	Elliot	Drew
n4	Keith	Ross, Sarah
n5	Ross	Allison, Keith, Sarah
n6	Sarah	Allison, Keith, Ross

Edge List
L1 = (n1, n5)
L2 = (n1, n6)
L3 = (n2, n3)
L4 = (n4, n5)
L5 = (n4, n6)
L6 = (n5, n6)



Draw your network!

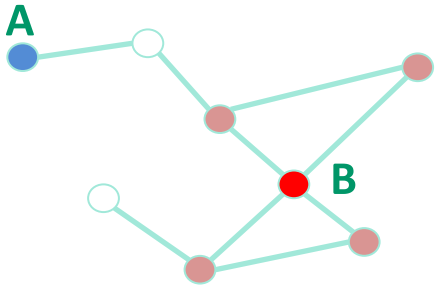
Depending on what data you have and what you want to analyze

If the question was who visited who, the network would be directed!

- All links would have an explicit direction
- Some of the links would disappear
- Some new links would be added for both directions between two nodes

Node Degrees

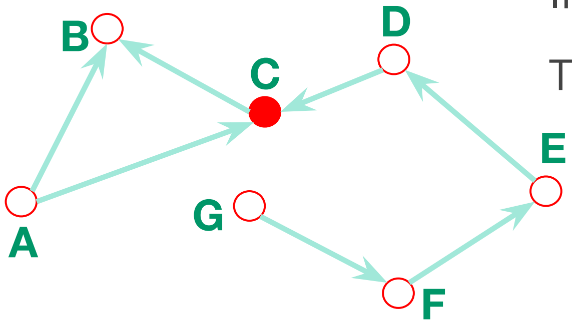
Undirected



Node degree: the number of links connected to the node.

$$k_A = 1 \qquad k_B = 4$$

Directed



In *directed networks* we can define an **in-degree** & **out-degree**.

The (total) degree is the sum of in- and out-degree.

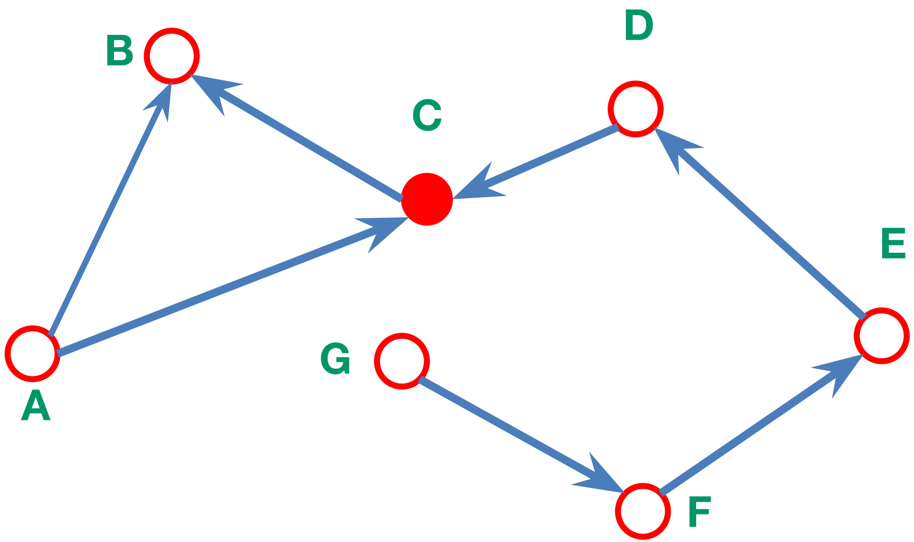
$$k_C^{in} = 2 \qquad k_C^{out} = 1 \qquad k_C = 3$$

$$k_j^{in} = \sum_{i=1}^{N} A_{ij}$$

Source: a node with $k^{in} = 0$; **Sink**: a node with $k^{out} = 0$.

Node Degrees - [Pen & Paper]

- Which nodes are sink nodes?
- Which nodes are source nodes?
- Node-B's In-degree? Out-degree? Total-degree?



Brief Statistics Review

BRIEF STATISTICS REVIEW

Four key quantities characterize a sample of N values $x_1, \ldots, x_N$:

Average (mean):

$$\langle x \rangle = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

The n^{th} moment:

$$\langle x^n \rangle = \frac{x_1^n + x_2^n + \ldots + x_N^n}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i^n$$

Standard deviation:

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \langle x \rangle\right)^2}$$

Distribution of x:

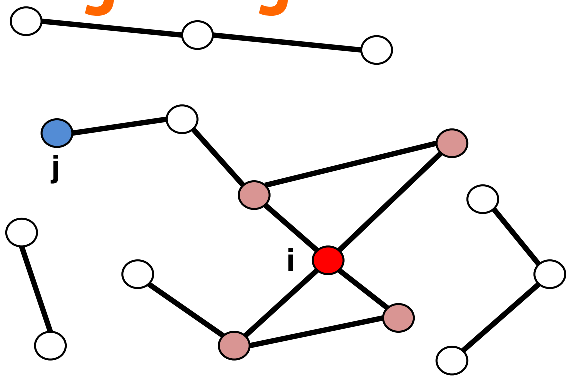
$$p_x = \frac{1}{N}\sum_i \delta_{x,x_i}$$

where p_x follows

$$\sum_i p_x = 1 \left(\int p_x \, dx = 1\right)$$

Average Degree

Undirected

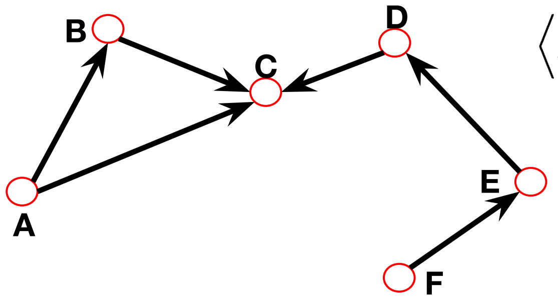


$$\langle k \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i \qquad \langle k \rangle \equiv \frac{2L}{N}$$

N – the number of nodes in the graph

L - the number of links in the graph

Directed



$$\langle k^{in} \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{in}, \quad \langle k^{out} \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{out}, \quad \langle k^{in} \rangle = \langle k^{out} \rangle$$

$$\langle k \rangle \equiv \frac{L}{N}$$

Average Degree

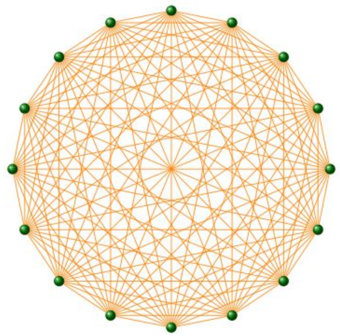
NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	⟨k⟩
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.33
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Paper	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

Multiply L/N by 2 if undirected

Real Networks are Sparse!

The maximum number of links:

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



A graph with degree L = L_{max}

is called a **complete graph**,

Its average degree is
<k> = N - 1

Network	N	L	L_{max}	K
Movie Actors	212,250	$6\ 10^6$	$1.8\ 10^{13}$	28.78
WWW (ND Sample)	325,729	$1.4\ 10^6$	10^{12}	4.51
Protein (*S. Cerevisiae*)	1,870	4,470	10^7	2.39
Coauthorship (Math)	70,975	$2\ 10^5$	$3\ 10^{10}$	3.9

(Source: Albert, Barabasi, RMP2002)

Algebraic Notation via Matrices

Representing Relations via Matrices

- Let **R** be a relation from **A** to **B** where

  A = $\{a_1, a_2, a_3, \ldots, a_m\}$ and **B** = $\{b_1, b_2, b_3, \ldots, b_n\}$
- Connection matrix **M** for **R** is defined as a **m x n** matrix where $\mathbf{M_{ij} > 0}$ if $\mathbf{\langle a_i, b_j \rangle}$ exists in **R** (e.g. there is a connection from $\mathbf{a_i}$ to $\mathbf{b_j}$).

  Otherwise, $\mathbf{M_{ij} = 0}$.
- If undirected, $a_{ij} = a_{ji}$

- The order of elements in **A** and **B** matters!
- Assume rows are **A**, columns are **B**

Representing Relations via Matrices [Pen & Paper]

A = {a, b, c, d}

B = {e, f, g}

R = { <a, e>, <c, f>, <d, g> }

Draw matrix M for R!

Representing Relations via Matrices [Pen & Paper]

A = {a, b, c, d}

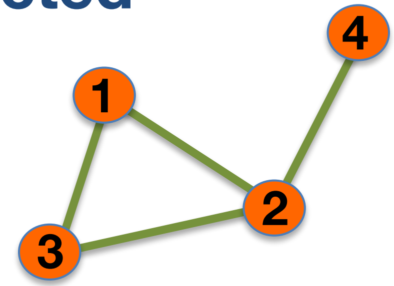
B = {e, f, g}

R = { <a, e>, <c, f>, <d, g> }

$$\mathbf{M} = \begin{array}{c} \\ \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \\ \mathbf{d} \end{array} \begin{array}{c} \begin{array}{ccc} \mathbf{e} & \mathbf{f} & \mathbf{g} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

Representing Graphs via Matrices

Undirected



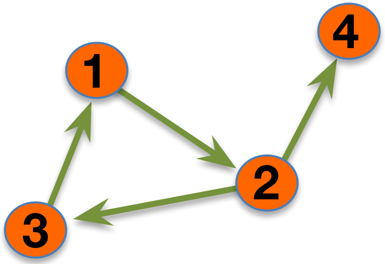
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad <k> = \frac{2L}{N}$$

Actor network, protein-protein interactions

Directed



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} \neq A_{ji}$$

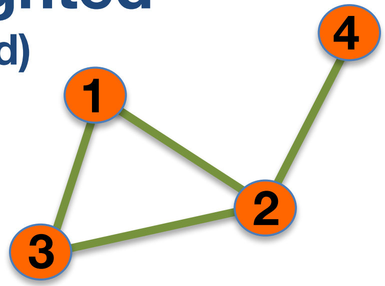
$$L = \sum_{i,j=1}^{N} A_{ij} \qquad <k> = \frac{L}{N}$$

WWW, citation networks

Representing Graphs via Matrices

Unweighted

(undirected)



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

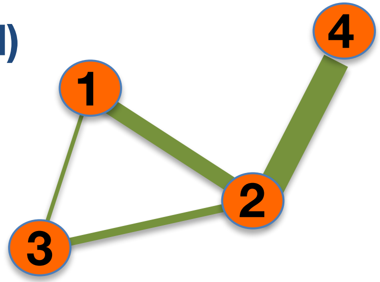
$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad <k>= \frac{2L}{N}$$

protein-protein interactions, www

Weighted

(undirected)



$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

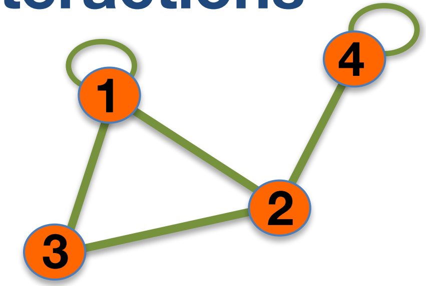
$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad <k>= \frac{2L}{N}$$

Call Graph, metabolic networks

Representing Graphs via Matrices

Self-interactions



$$A_{ij} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

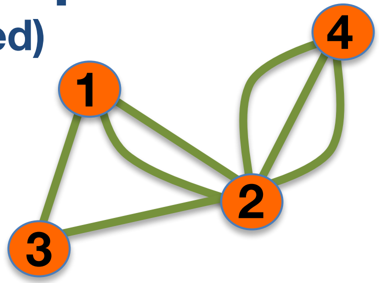
$A_{ii} \neq 0$ $\quad A_{ij} = A_{ji}$

$$L = \frac{1}{2} \sum_{i,j=1, i \neq j}^{N} A_{ij} + \sum_{i=1}^{N} A_{ii} \qquad ?$$

Protein interaction network, www

Multigraph

(undirected)



$$A_{ij} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}$$

$A_{ii} = 0$ $\quad A_{ij} = A_{ji}$

$$L = \frac{1}{2} \sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad \langle k \rangle = \frac{2L}{N}$$

Social networks, collaboration networks

Reflexive, Symmetric, Anti-Symmetric (Boolean Examples)

1			
	1		
		1	
			1

Reflexive iff $\mathbf{M_{ii} = 1}$ for all i

All 1s on diagonal

		1	0
		0	1
1	0		
0	1		

Symmetric iff $\mathbf{M = M^T}$

All identical across diagonal

			1
			0
0	1		

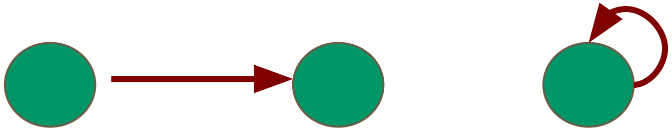
Anti-symmetric if $\mathbf{M_{ij} = 0}$ or $\mathbf{M_{ji} = 0}$ for all $\mathbf{i \neq j}$

All 1s are across 0s

Reflexive, Symmetric, Anti-Symmetric

Symmetric: Always both ways

Anti-Symmetric: Never both ways (But self loops are ok)



Symmetric
Reciprocated Love

Anti-Symmetric
Unrequited Love

Operations on Matrices: Matrix Addition & Subtraction

- You can add matrices of the same size, mxn, and get another mxn matrix as a result

Unweighted

1	0
1	1
0	0

+

1	0
0	1
1	0

=

1	0
1	1
1	0

This is like binary OR operation

Weighted

1	10
-2	5
7	8

+

3	0
5	1
2	9

=

4	10
3	6
9	17

Social Network Examples: Matrix Addition & Subtraction

- Commonly applied for
  - Different relations between same set of nodes,
  - Over time snapshots of the same network

Different Relations btw Same Set of Actors

$$\begin{array}{c|c} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \end{array} + \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Team mates? **Co-authors?** **Work Together?**

Over-time Snapshots of Trade Network

$$\begin{array}{c|c} & \begin{array}{cc} \text{Import} & \text{Export} \end{array} \\ \begin{array}{c} \text{APAC} \\ \text{EU} \\ \text{LATAM} \end{array} & \begin{bmatrix} 1 & 10 \\ -2 & 5 \\ 7 & 8 \end{bmatrix} \end{array} + \begin{bmatrix} 3 & 0 \\ 5 & 1 \\ 2 & 9 \end{bmatrix} = \begin{bmatrix} 4 & 10 \\ 3 & 6 \\ 9 & 17 \end{bmatrix}$$

2020 **2021** **Total**

Operations on Matrices: Scalar Multiplication

- Commonly applied to give weights to different relations or snapshots.

$$3 \begin{bmatrix} 1 & 10 & 4 \\ -2 & 5 & 3 \\ 7 & 8 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 30 & 12 \\ -6 & 15 & 9 \\ 21 & 24 & 0 \end{bmatrix}$$

Operations on Matrices: Matrix Multiplication

- Multiplying two matrices in different orders will not give the same results!
- Inner dimensions should match!

2-by-~~3~~ x ~~3~~-by-2 ⇒ 2-by-2

(5*5)+(6*5)+(3*5)

$$\begin{bmatrix} 5 & 6 & 3 \\ 3 & 2 & 2 \end{bmatrix} \times \begin{bmatrix} 5 & 6 \\ 5 & 6 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 70 & 88 \\ 35 & 42 \end{bmatrix}$$

(3*6)+(2*6)+(2*6)

Left to Right

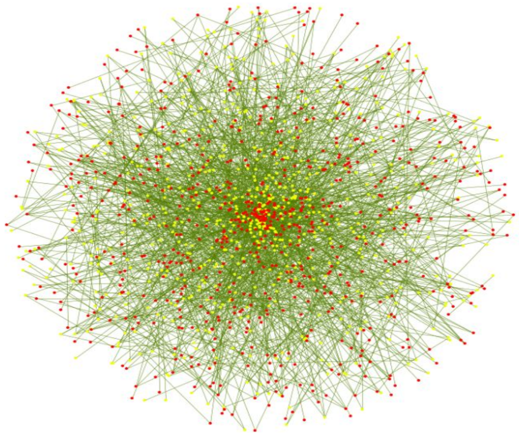
Down Right

3-by-~~2~~ x ~~2~~-by-3 ⇒ 3-by-3

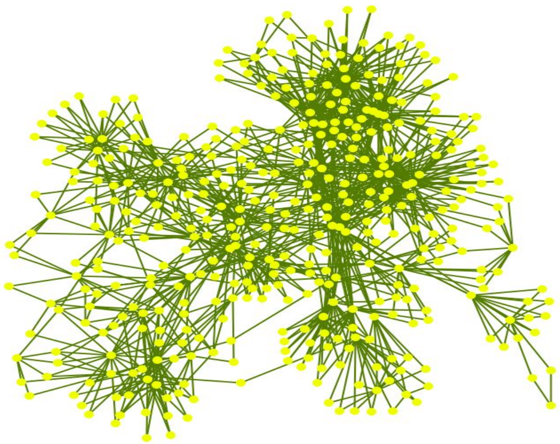
$$\begin{bmatrix} 5 & 6 \\ 5 & 6 \\ 5 & 6 \end{bmatrix} \times \begin{bmatrix} 5 & 6 & 3 \\ 3 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 43 & 42 & 27 \\ 43 & 42 & 27 \\ 43 & 42 & 27 \end{bmatrix}$$

Modes

- **One mode network:** All actors come from one set.
- **Two mode network:** Actors come from 2 distinct set.
- We will work with all types of networks: one, two, or higher modes of networks

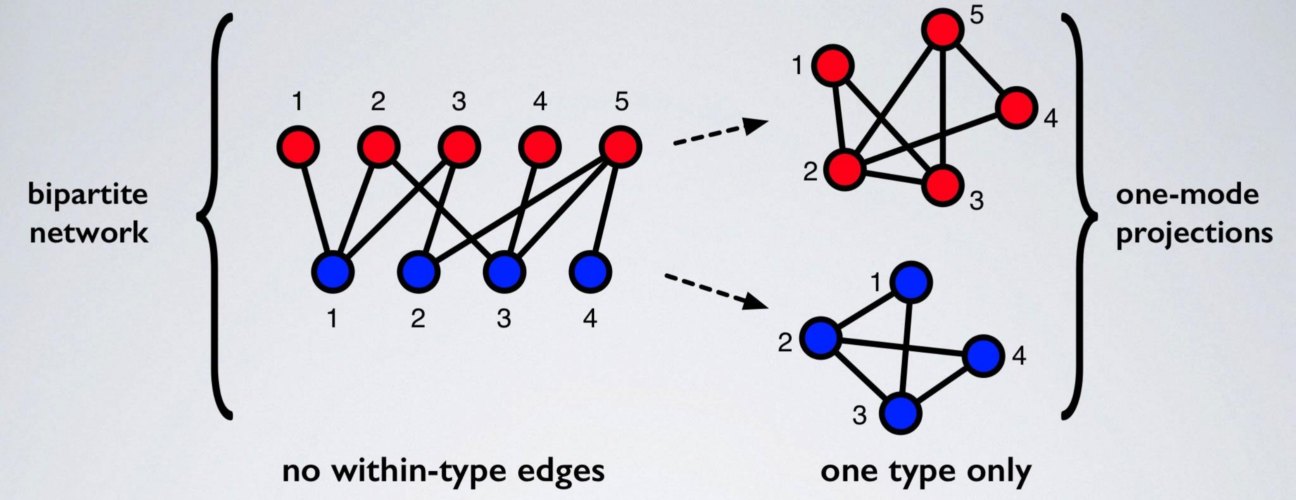


Two Mode: author-to-paper



One Mode: paper-to-paper

Modes



authors & papers

actors & movies/scenes

musicians & albums

people & online groups

people & corporate boards

people & locations (checkins)

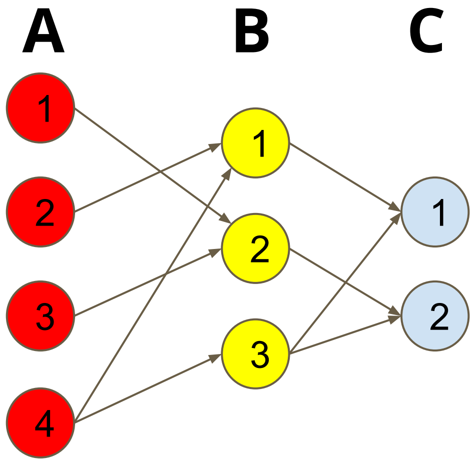
metabolites & reactions

genes & substrings

words & documents

plants & pollinators

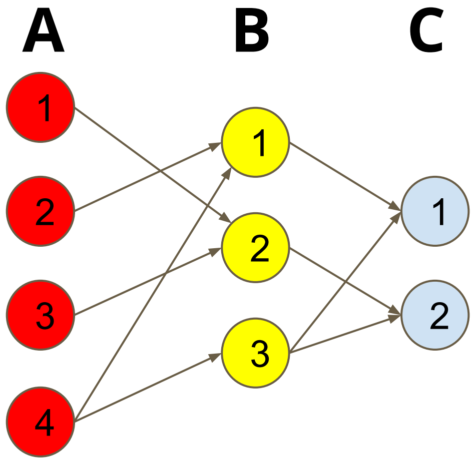
Example: Bi-Partite Networks & Matrix Multiplication



- From which element of A can I reach to which element of C?
- Which elements of A have a path to which elements of C?
- Which elements of A and C are (indirectly) connected?

	C1	**C2**
A1	No	Yes
A2	Yes	No
A3	No	Yes
A4	Yes	Yes

Example: Bi-Partite Networks & Matrix Multiplication



A-by-B

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

B-by-C

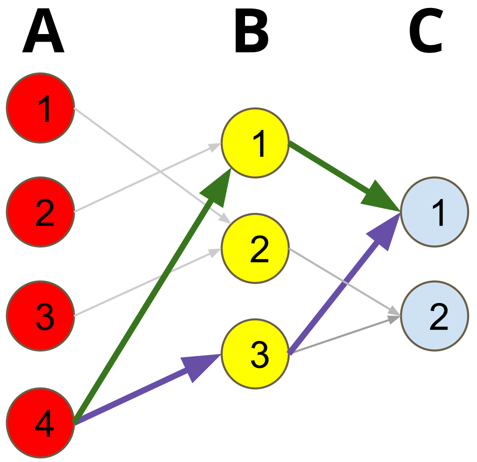
1	0
0	1
1	1

A-by-C

0	1
1	0
0	1
2	1

What is the meaning of the resulting matrix?

Example: Bi-Partite Networks & Matrix Multiplication



A-by-B

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

×

B-by-C

1	0
0	1
1	1

=

A-by-C

0	1
1	0
0	1
2	1

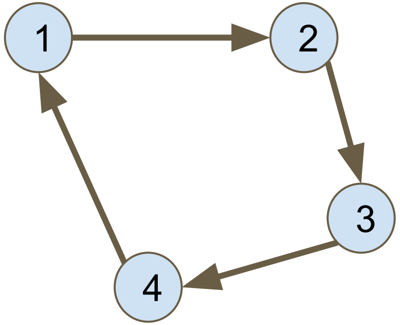
No path from A2 to C2

Two paths from A4 to C1:

- A4 → B1 → C1
- A4 → B3 → C1

Operations on Matrices: Powers of Adjacency Matrix

- The powers of the adjacency matrix counts things.
- Entry i, j in A^s gives the number of walks from i to j of length s.



A

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

×

A

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

=

$\mathbf{A^2}$

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

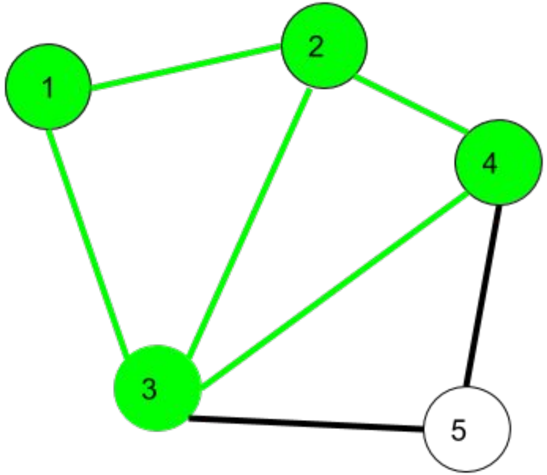
Meaning of values in A^2 : Which node can reach to which node(s) in 2 steps? And how many ways does it have to achieve that?

Graph Theory Basics Reminder: Walk, Trail, and Path

Walk

- Vertices can be repeated
- Edges can be repeated

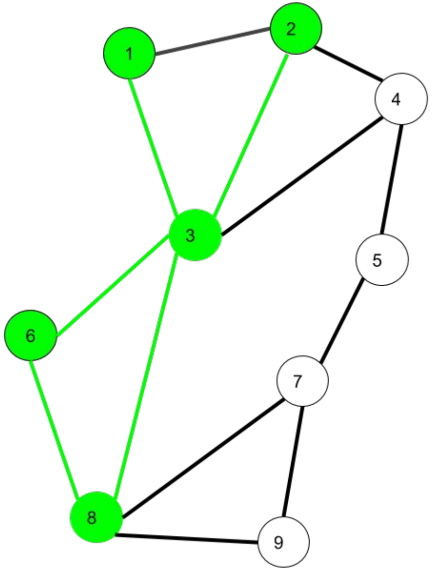
1 → 2 → 3 → 4→ 2→ 1→ 3 is a walk



Trail

- Vertices can be repeated
- Edges cannot be repeated

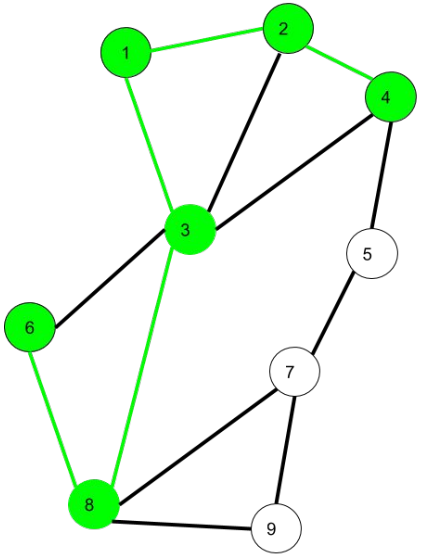
1→ 3→ 8→ 6→ 3→ 2 is a trail



Path

- Vertices cannot be repeated
- Edges cannot be repeated

6 → 8 → 3 → 1→ 2 → 4 is a Path



https://www.geeksforgeeks.org/mathematics-walks-trails-paths-cycles-and-circuits-in-graph/

[Pen & Paper] Real Life Questions via Matrix Operations

Purchase Relation Set:

<Buyer1, House1>,

<Buyer2, House1>,

<Buyer3, House 3>,

<Buyer4, House4>,

<Buyer5, House5>

Marketing Relation Set:

<Realtor1, House1>,

<Realtor2, House1>,

<Realtor2, House 2>,

<Realtor3, House3>,

<Realtor3, House4>,

<Realtor4, House5>,

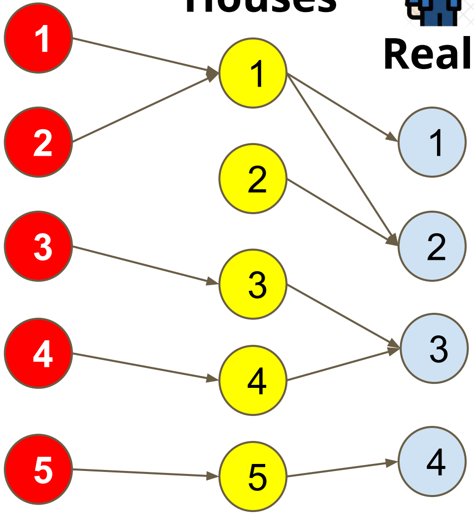
- **Which buyer paid which realtor(s)?**
- **Which realtor(s) had most transactions?**
- **Which realtor made the most money?**

[Pen & Paper] Real Life Questions via Matrix Operations

Buyers

Houses

Realtors



B x H

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

×

H x R

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

=

B x R

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

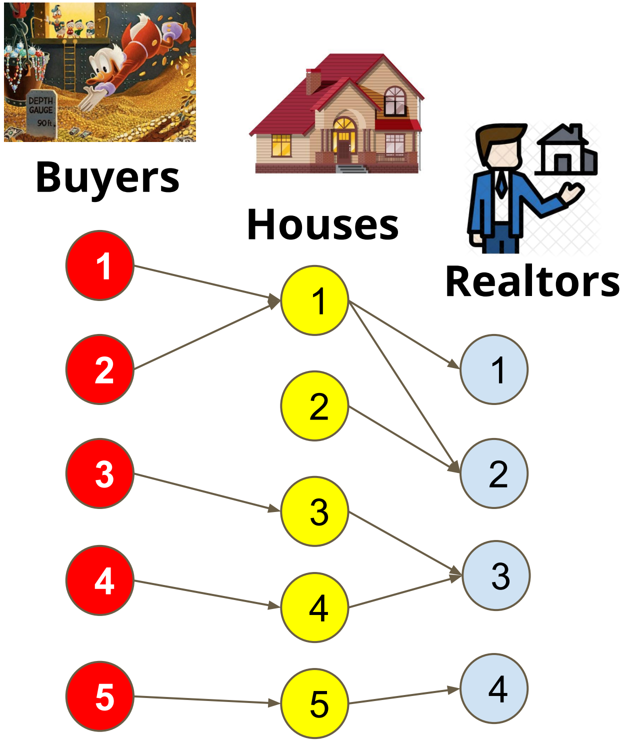
Marketing Relation Set:

<R1, H1>, <R2, H1>, <R2, H2>, <R3, H3>, <R3, H4>,<R4, H5>

Purchase Relation Set:

<B1, H1>, <B2, H1>, <B3, H3>, <B4, H4>, <B5, H5>

[Pen & Paper] Real Life Questions via Matrix Operations



B x H

1/2	0	0	0	0
1/2	0	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

x

H x R

1/2	1/2	0	0
0	1	0	0
0	0	1	0
0	0	1	0
0	0	0	1

=

B x R

1/4	1/4	0	0
1/4	1/4	0	0
0	0	1	0
0	0	1	0
0	0	0	1

Marketing Relation Set:

<R1, H1>, <R2, H1>, <R2, H2>, <R3, H3>, <R3, H4>,<R4, H5>

Purchase Relation Set:

<B1, H1>, <B2, H1>, <B3, H3>, <B4, H4>, <B5, H5>

Answers to 3 questions

B x R

1/4	**1/4**	0	0
1/4	**1/4**	0	0
0	0	1	0
0	0	1	0
0	0	0	1

Hidden Assumption: Buyers share the house 50/50, realtors share the commission 50/50.

- **Which buyer paid which realtor(s)?**
  - B1 paid R1, R2
  - B2 paid R1, R2
  - B3, B4 paid R3
  - B5 paid R4
- **Which realtor(s) had most transactions?**
  - R1, R2, and R3 each had 3 transactions
- **Which realtor made the most money?**
  - Don't know!
  - We don’t know if payments to the realtors are fixed or based on house prices.
  - We don’t know house prices.

Next Lecture:

Network Topology, Metrics, and Centrality