Network Analysis:

The Hidden Structures behind the Webs We Weave

17-338 / 17-668

Intro to Graph Theory

Thursday, August 29, 2024

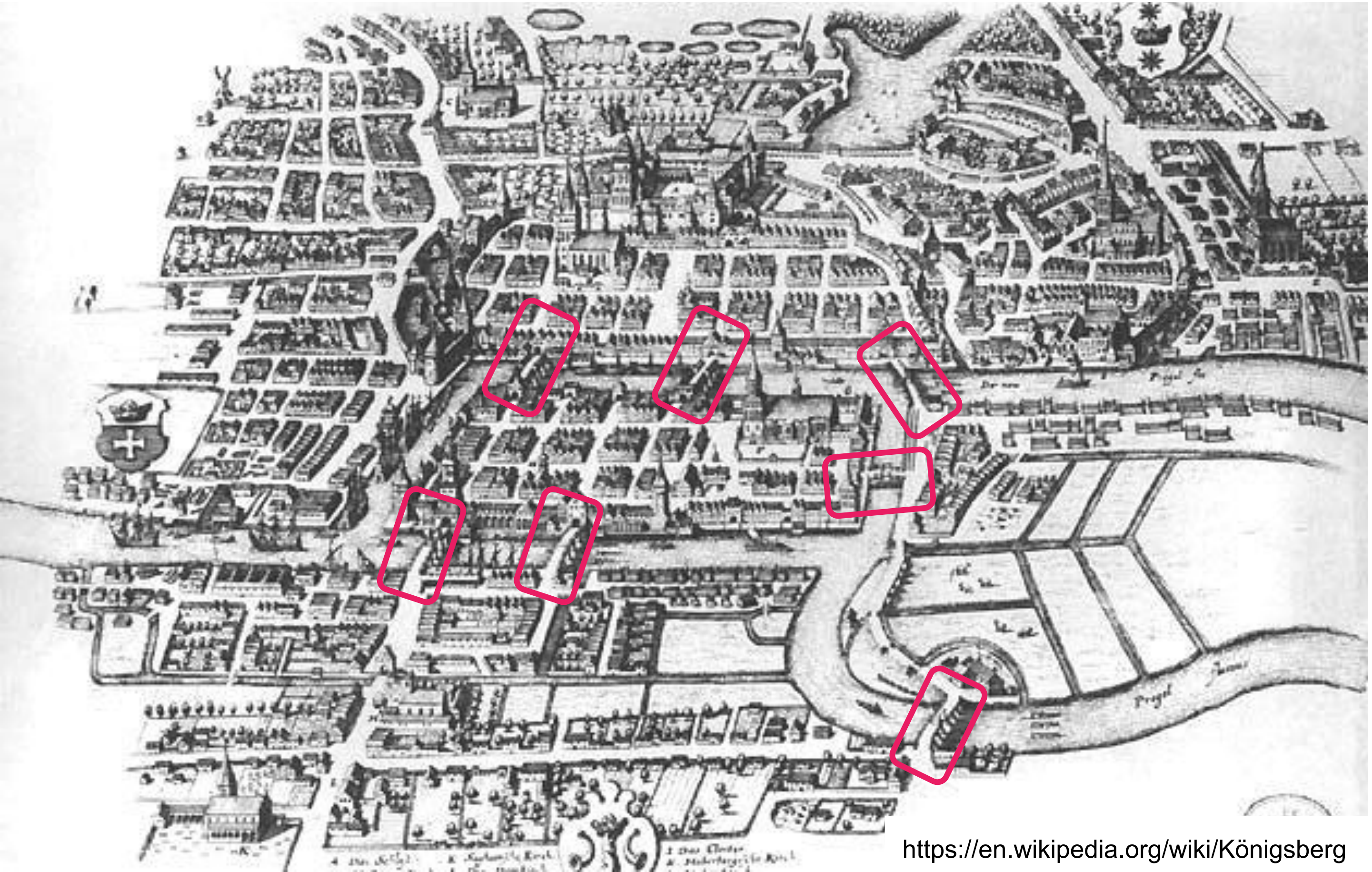
Patrick Park & Bogdan Vasilescu

Carnegie Mellon University
School of Computer Science

S3D
Software and Societal
Systems Department

2-min Quiz, on Canvas

The Seven Bridges of Königsberg (1735)



https://en.wikipedia.org/wiki/Königsberg

Königsberg was a port city on the south eastern corner of the Baltic Sea. It is today known as Kaliningrad and is part of Russia.

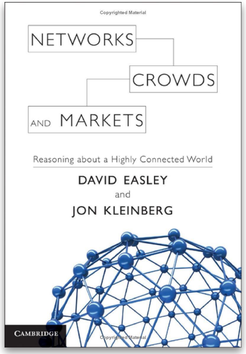
Can you walk across all seven bridges and never cross the same one twice?

Plan for Today

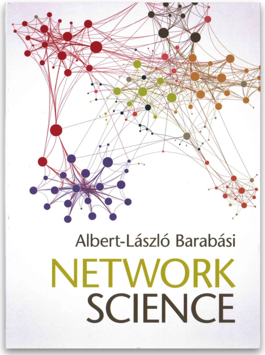
Intro to graph theory:

- In- / Out-degree & degree distribution
- Edge and Dyad
- Paths, cycles, and small-world
- Adjacency matrix representation

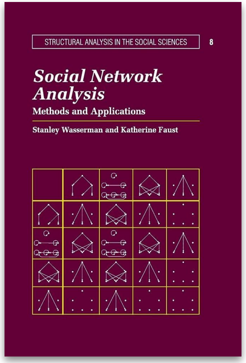
(EK Ch. 2)



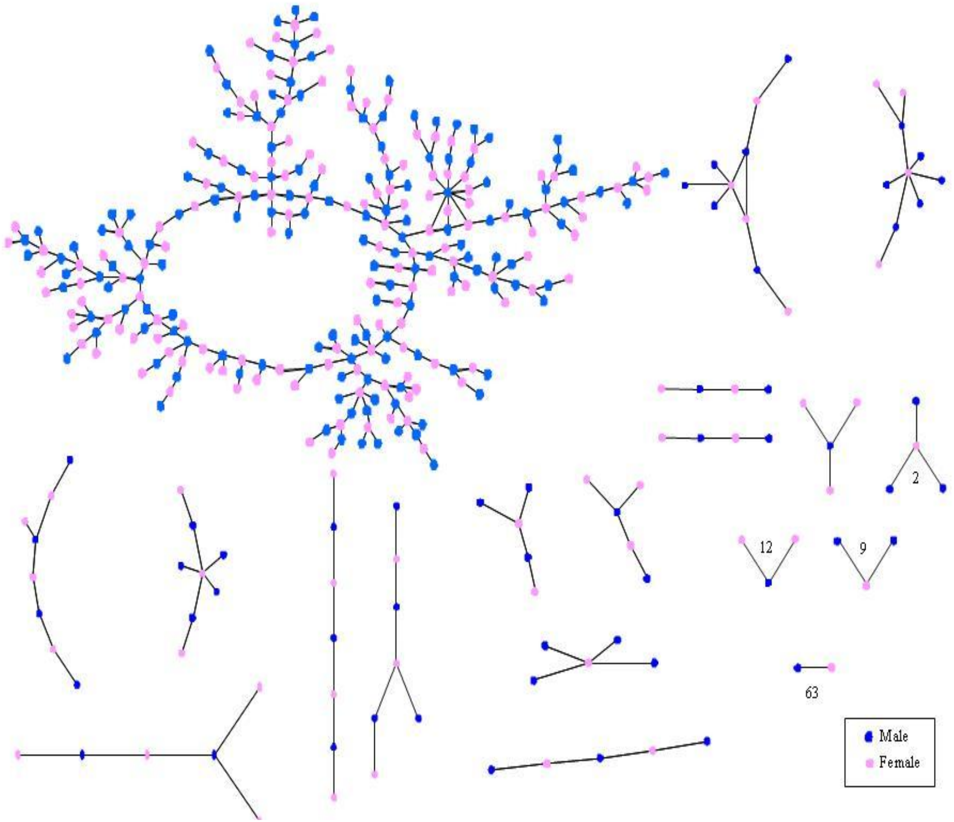
(B Ch. 2.2–2.4, 2.8–2.9)



(WF Ch. 4.1–4.3, 4.9)



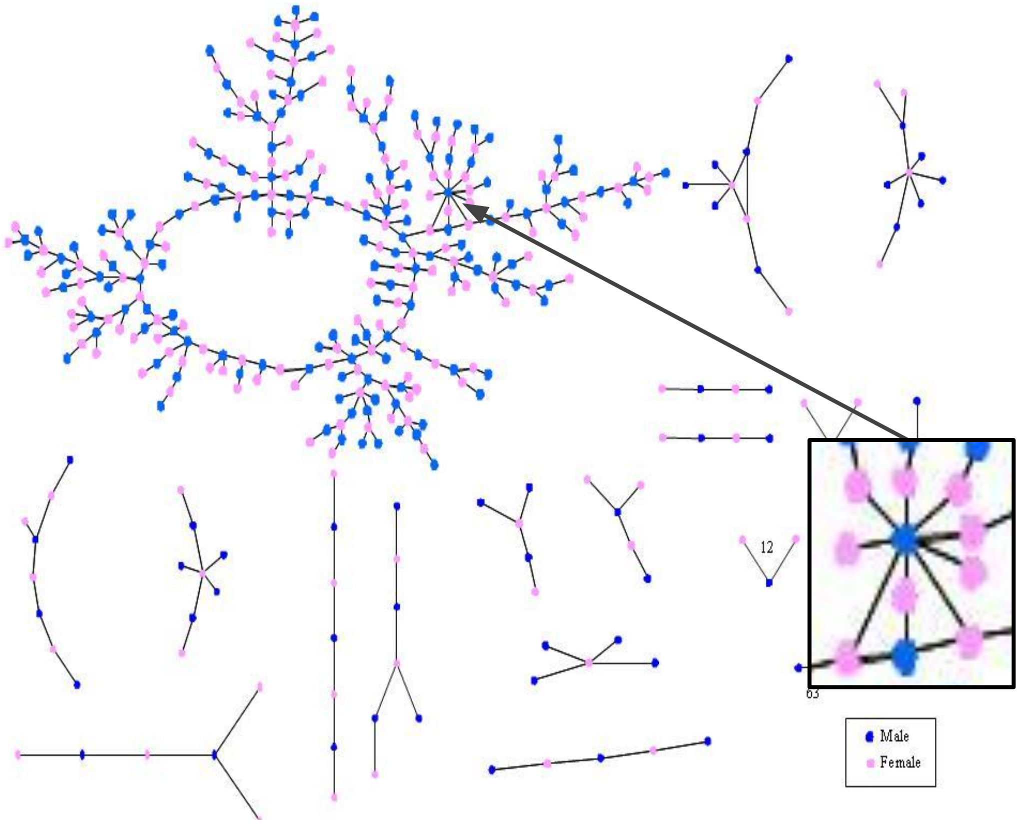
Properties of a Graph Hint at Social Aspects



1. In this graph, closed **triads** are virtually absent. Why?

2. How many groups of “branches”, or **spokes** are connected to the **cycle** in the middle? What social aspect might they represent?

Properties of a Graph Hint at Social Aspects

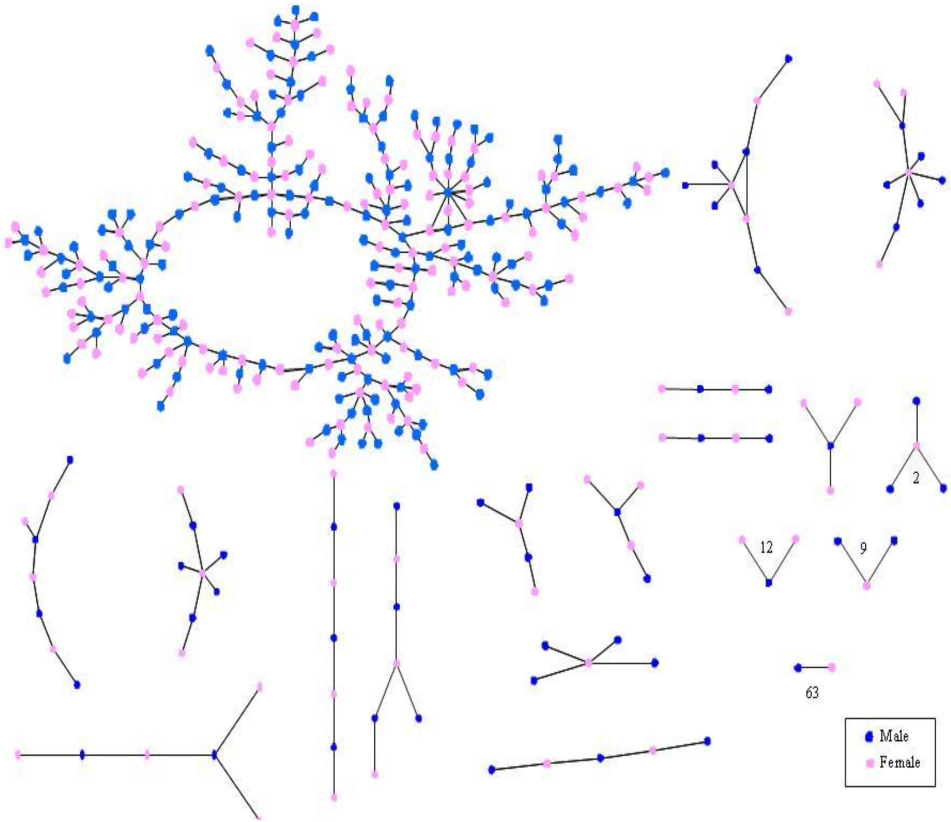


1. In this graph, closed **triads** are virtually absent. Why?

2. How many groups of “branches”, or **spokes** are connected to the **cycle** in the middle? What social aspect might they represent?

3. Fitzwilliam Darcy Esquire (“Mr. Darcy”) the romantic has nine **edges**. What does this tell us about Mr. Darcy’s social life?

Properties of a Graph Hint at Social Aspects



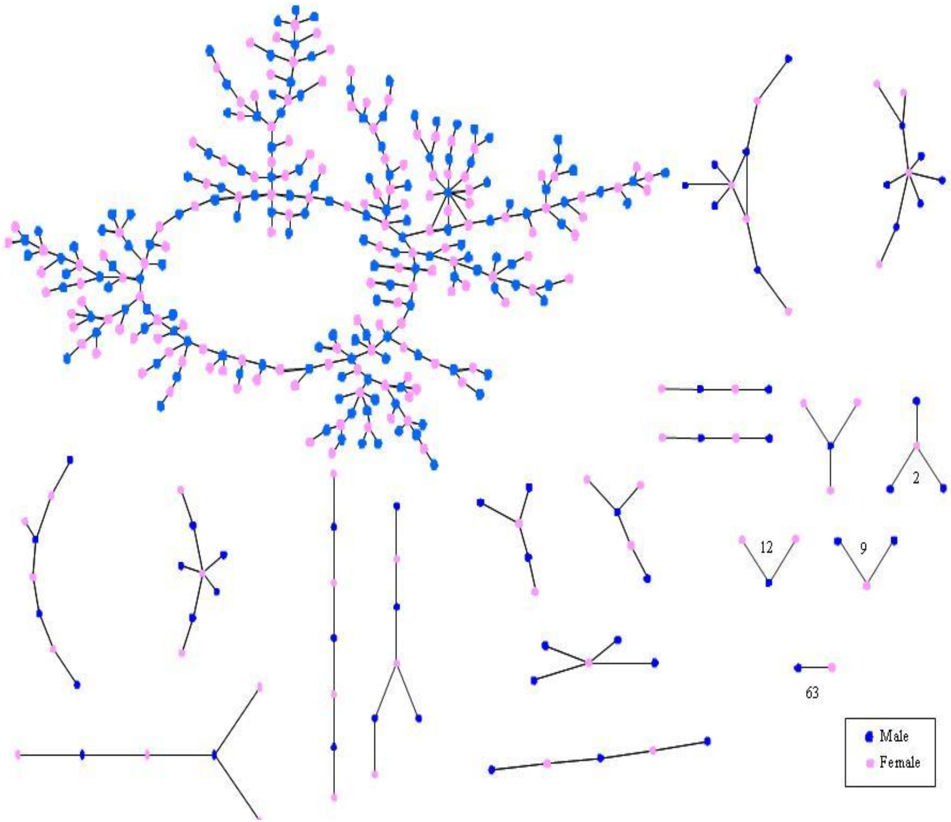
As a network analyst, your job is often:

1. To somehow map aspects of a graph onto the social situation (metrics interpretation).
2. To map some social aspect onto a quantitative characteristic in a graph (operationalizing a concept).

With this mapping, you can use the tools from **graph theory** to describe and understand the observed social phenomena and make predictions.

Node Degree

Node Degree



(including Mr. Darcy) How many romantic partners did the students have on average (i.e., average degree in this graph)?

Average Degree:

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

N: number of nodes

k: degree

L: number of edges

Node Degree

Degree is one of the most fundamental quantities in a network.

In directed networks:

In-degree: the number of links that point to a node (receiving in)

Out-degree: the number of links that a node points to other nodes (sending out)

Total degree of node *i* is the sum of in- and out-degrees

$$k_i = k_i^{in} + k_i^{out}$$

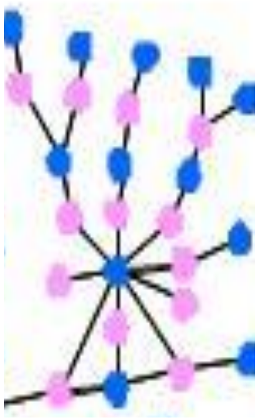
As a whole, the grand total of in-degrees equals that of their out-degrees

$$L = \sum_{i=1}^{N} k_i^{in} = \sum_{i=1}^{N} k_i^{out}$$ ← Why?

Node Degree

The average in-degree equals average out-degree

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

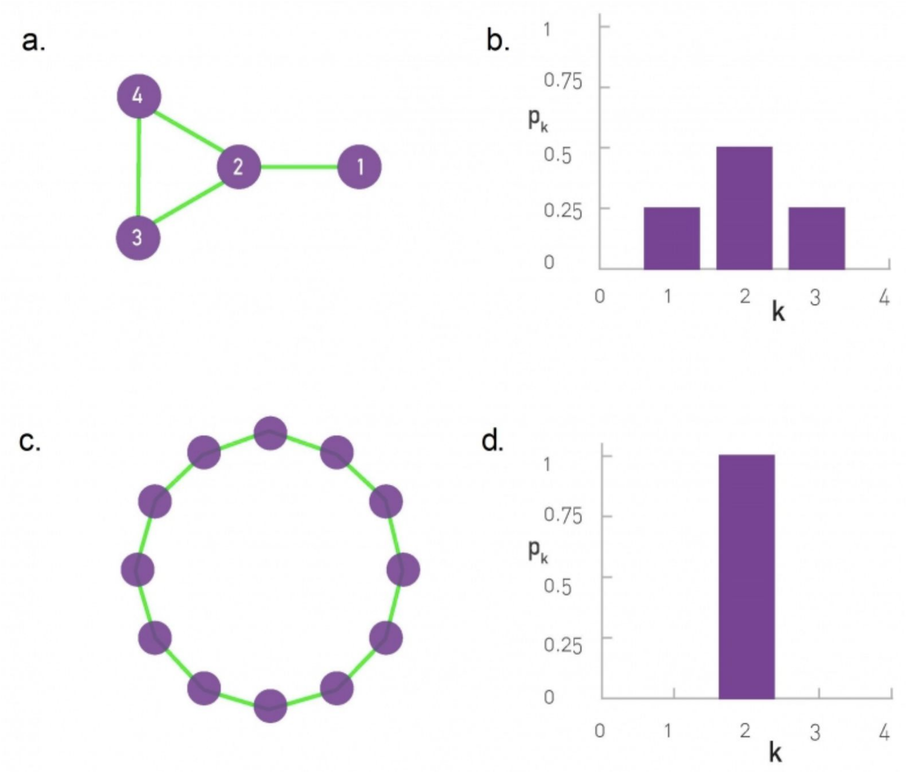


So, how equally or unequally are romantic partners distributed in the network (i.e., **degree distribution**)?

Degree distribution is the probability that a randomly selected node will have degree *k*:

$$p_k = \frac{N_k}{N}$$

Node Degree



The degree distribution is central to studying various social and physical phenomena.

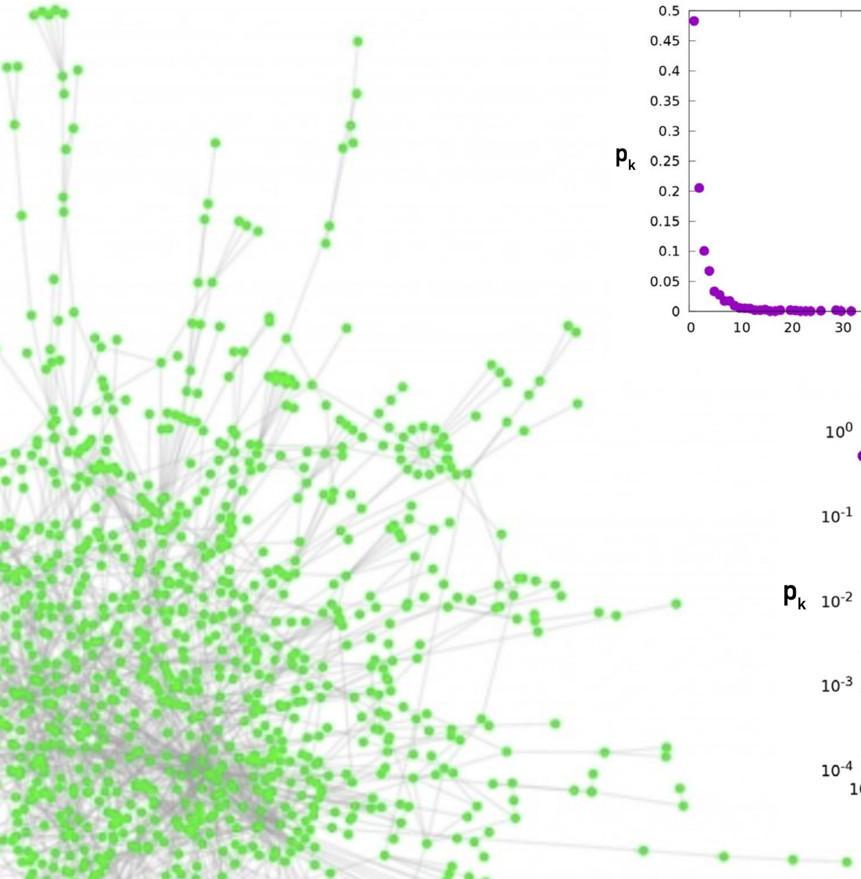
Examples:

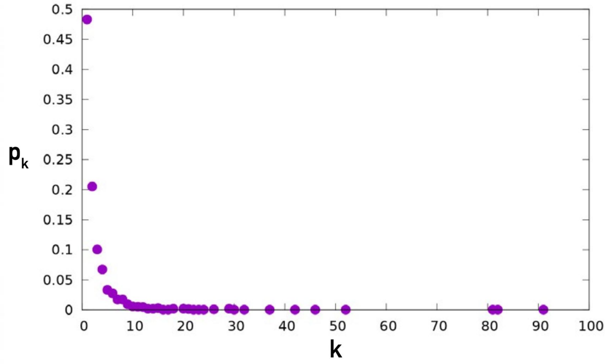
Powergrid failure (or graph robustness)

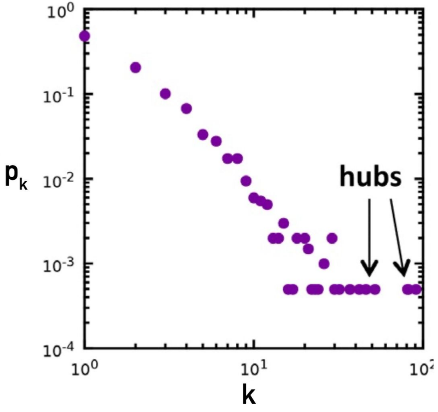
Spread of viruses (contagion dynamics)

Social inequality (long-tail distributions)

Node Degree







Real world network degree distributions tend to be heavily skewed.

In later lectures, we will revisit degree distributions and the models that offer explanations for the skew.

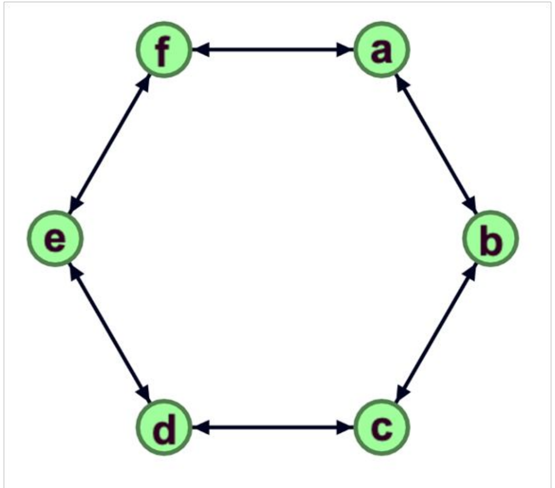
Edges and Dyads

Edges and Dyads

An edge is a direct connection between nodes

Dyad is any pair of nodes

So, how many edges and dyads does this ring network have?

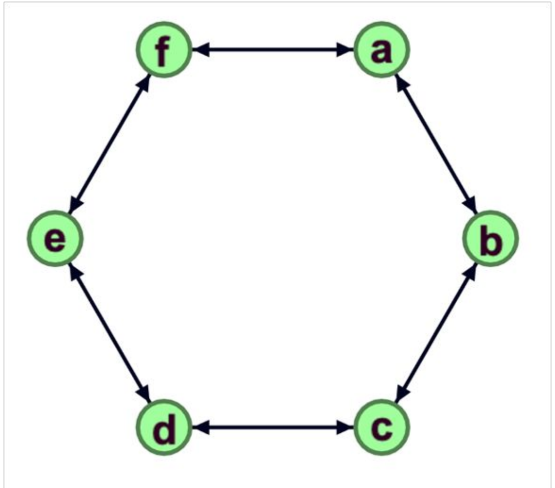


Edges and Dyads

An edge is a direct connection between nodes

Dyad is any pair of nodes

So, how many edges and dyads does this ring network have?



Edges: 6

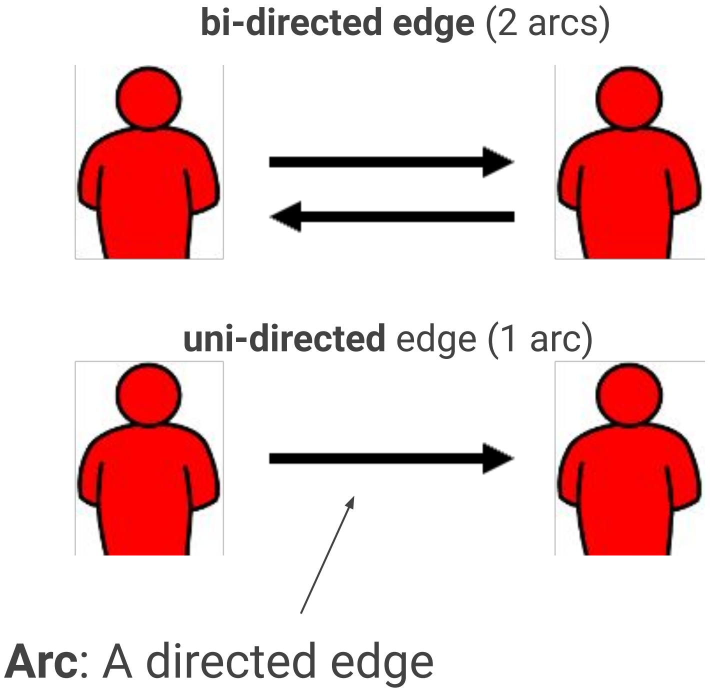
Dyads: 15

Edges and Dyads

Types of edges:

Uni-directed edge can imply **hierarchical** relationships (status, dominance…)

Bi-directed edge can imply **reciprocity** in a relationship



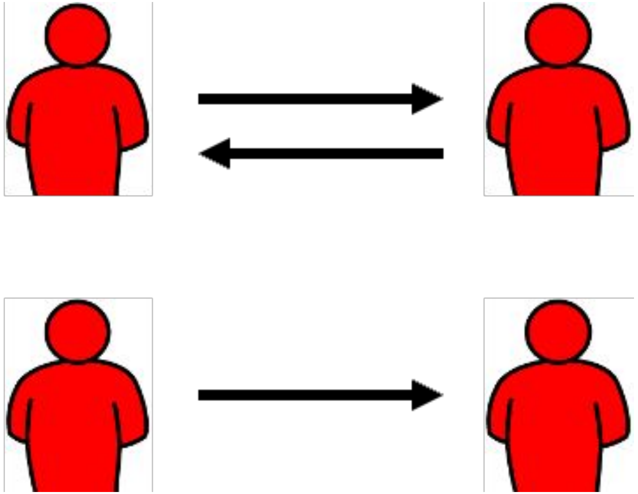
Edges and Dyads

Norms of Reciprocity

- One of the universal norms across cultures
- Mutual obligations to each other
- Balance: The amount one gives is roughly equivalent to what they receive
- Variation in this norm exists and can be seen as characteristics of societies

"I scratch your back, you scratch mine"

"Eye for an eye, tooth for a tooth"



Edges and Dyads

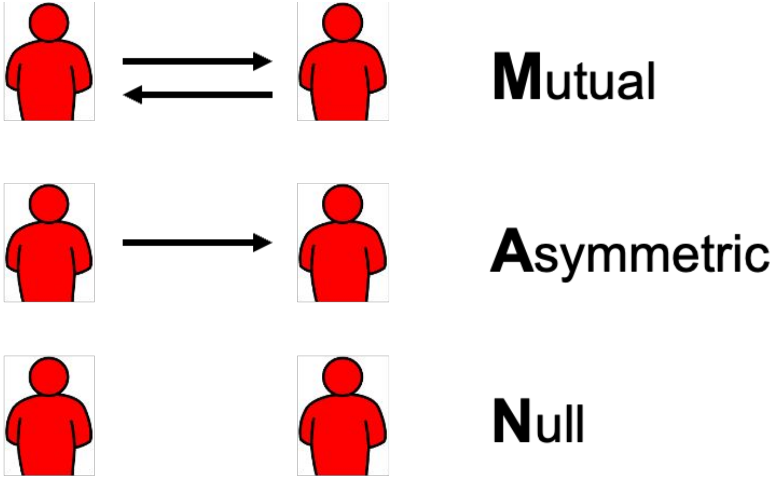
A **dyad** is any unordered pair of nodes.

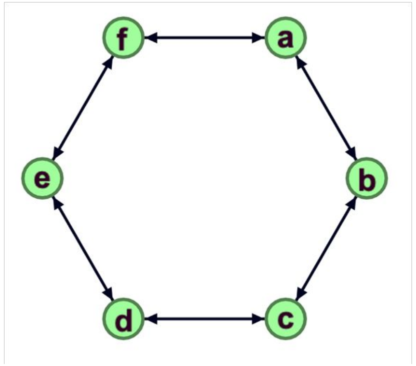
You can count the number of mutual, asymmetric, and null dyads in a graph.

This complete enumeration is called a **dyad census.**

How many mutual, asymmetric, and null dyads are in this ring lattice?

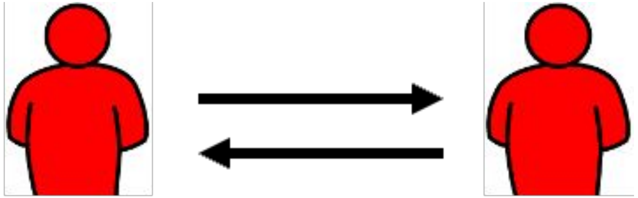
Three Types of Dyads





How to measure the level of reciprocity in a social network?

“You bought me coffee, now it’s my turn”



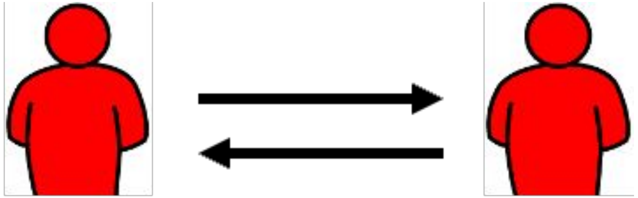
How to measure the level of reciprocity in a social network?

Prevalence of *M*, relative to *A* or *N* is one descriptive measure of reciprocity

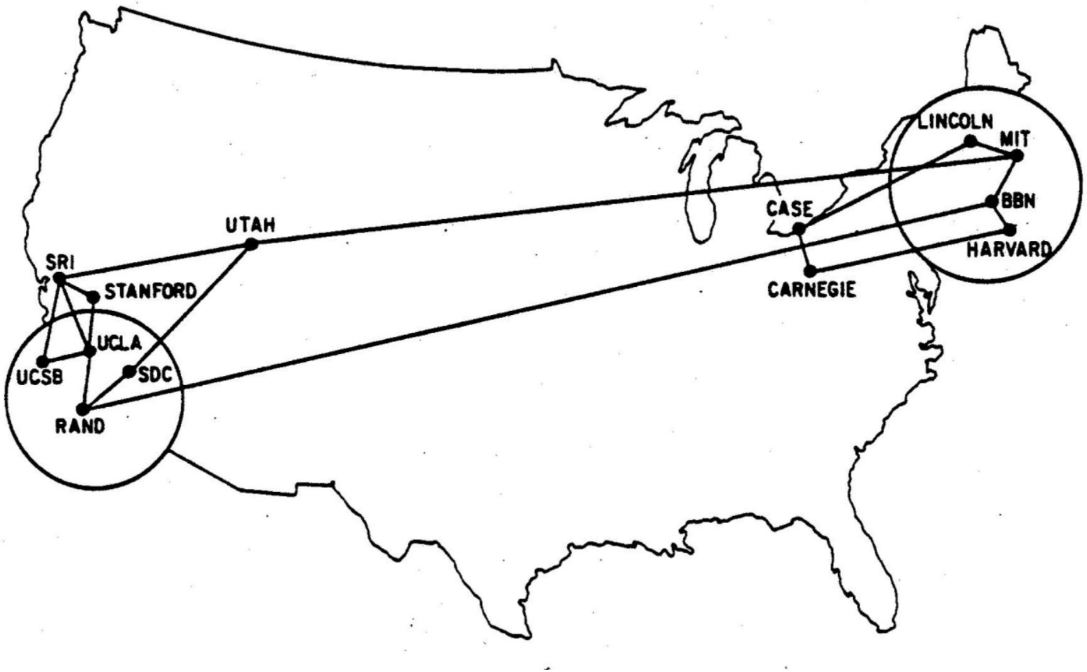
For comparing two networks of similar size and density, this approach is an informative first step

However, problematic for networks that differ in size and density. Why?

"You bought me coffee, now it's my turn"

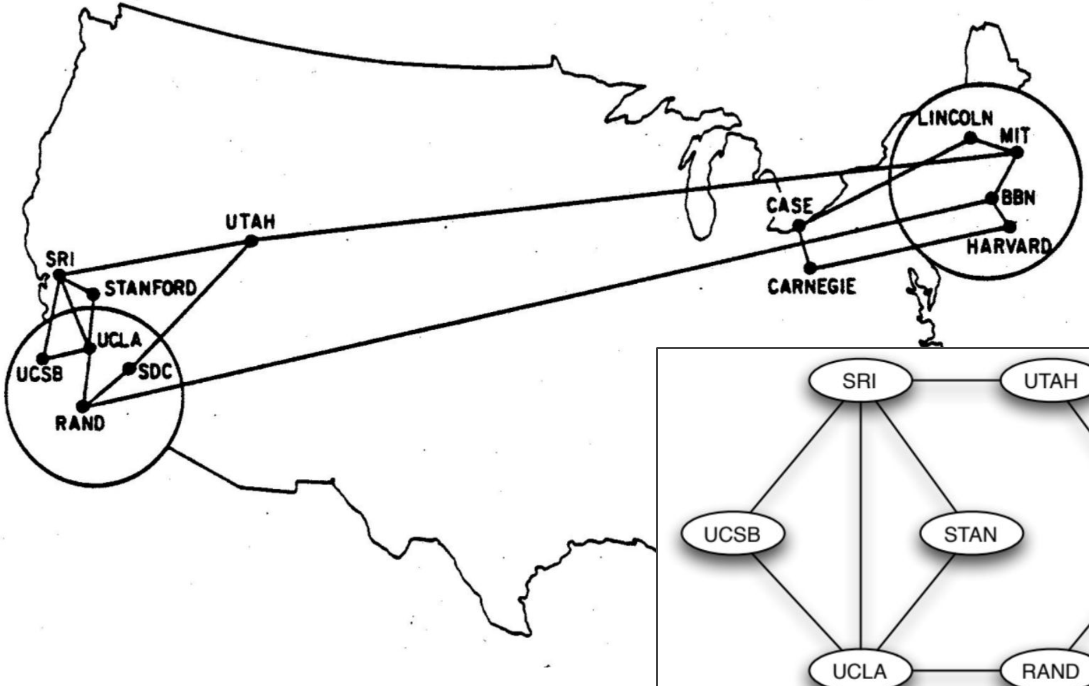


Paths and Distances



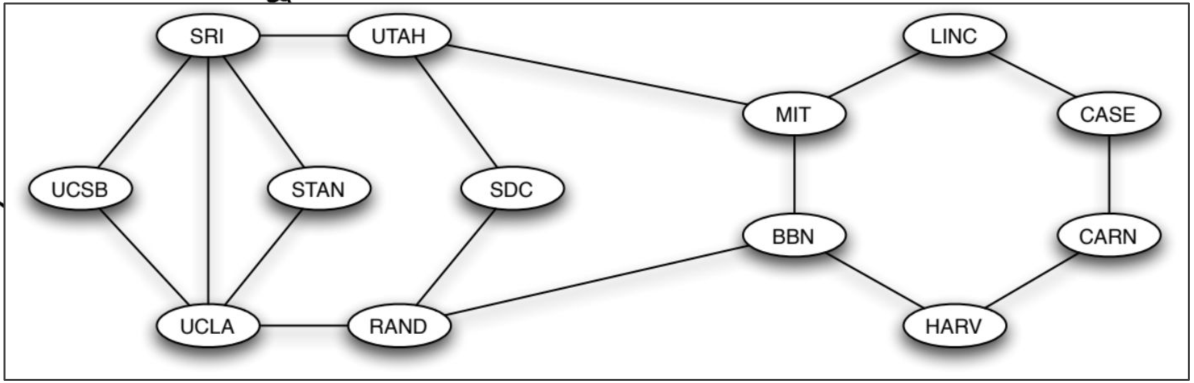
Does anyone know what this network is?

Path: A sequence of connected edges where all nodes *and* edges are distinct



SRI−UCLA−RAND−BBN **is** a path

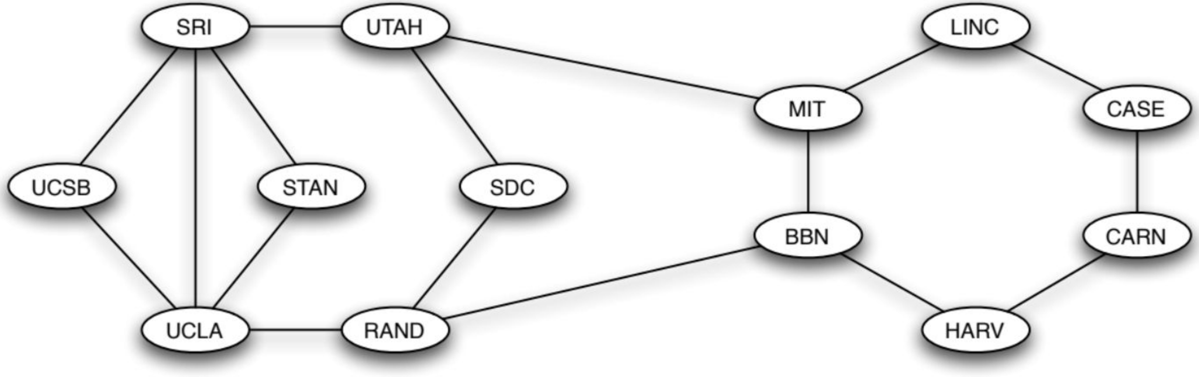
SRI−UCLA−STAN−SRI−UTAH **is NOT** a path



Arpanet 1970: The very first internet
https://en.wikipedia.org/wiki/ARPANET

Shortest Path: A path with the minimum number of edges between two nodes

CARN-CASE-LINC-MIT **is shorter than** CARN-HARV–BBN-RAND-SDC-UTAH-MIT

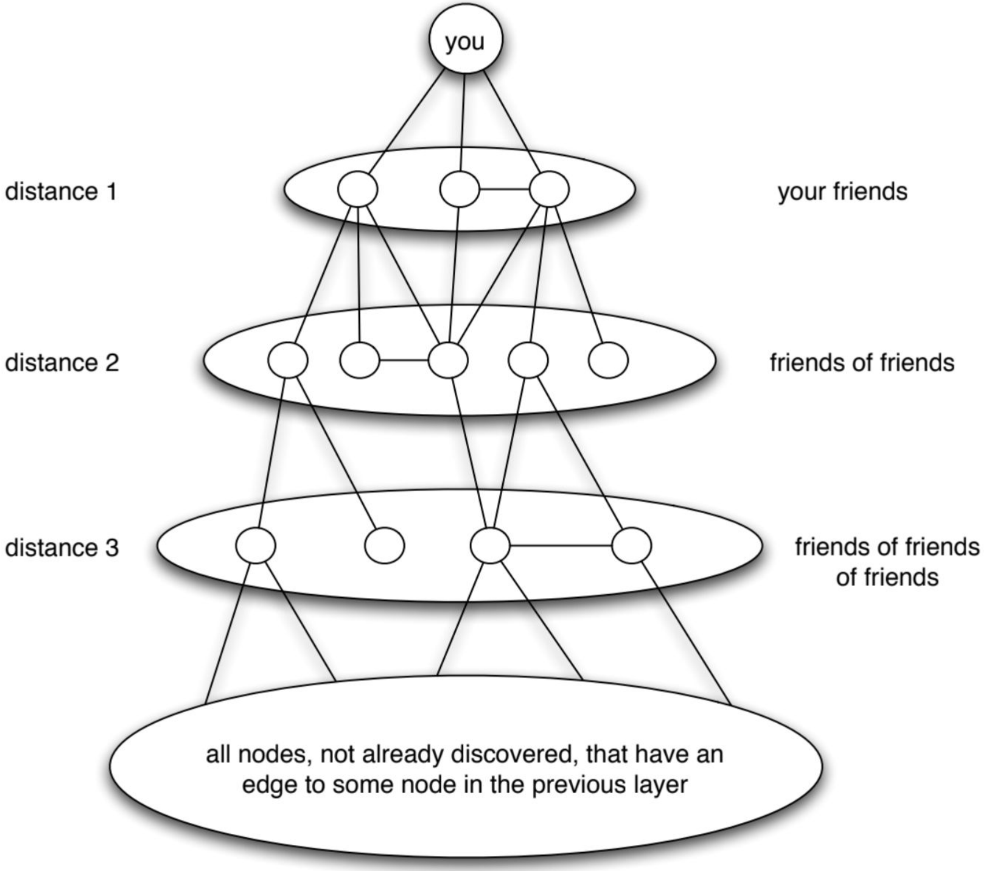


Distance between two nodes: length of the shortest path between them (also called geodesic).

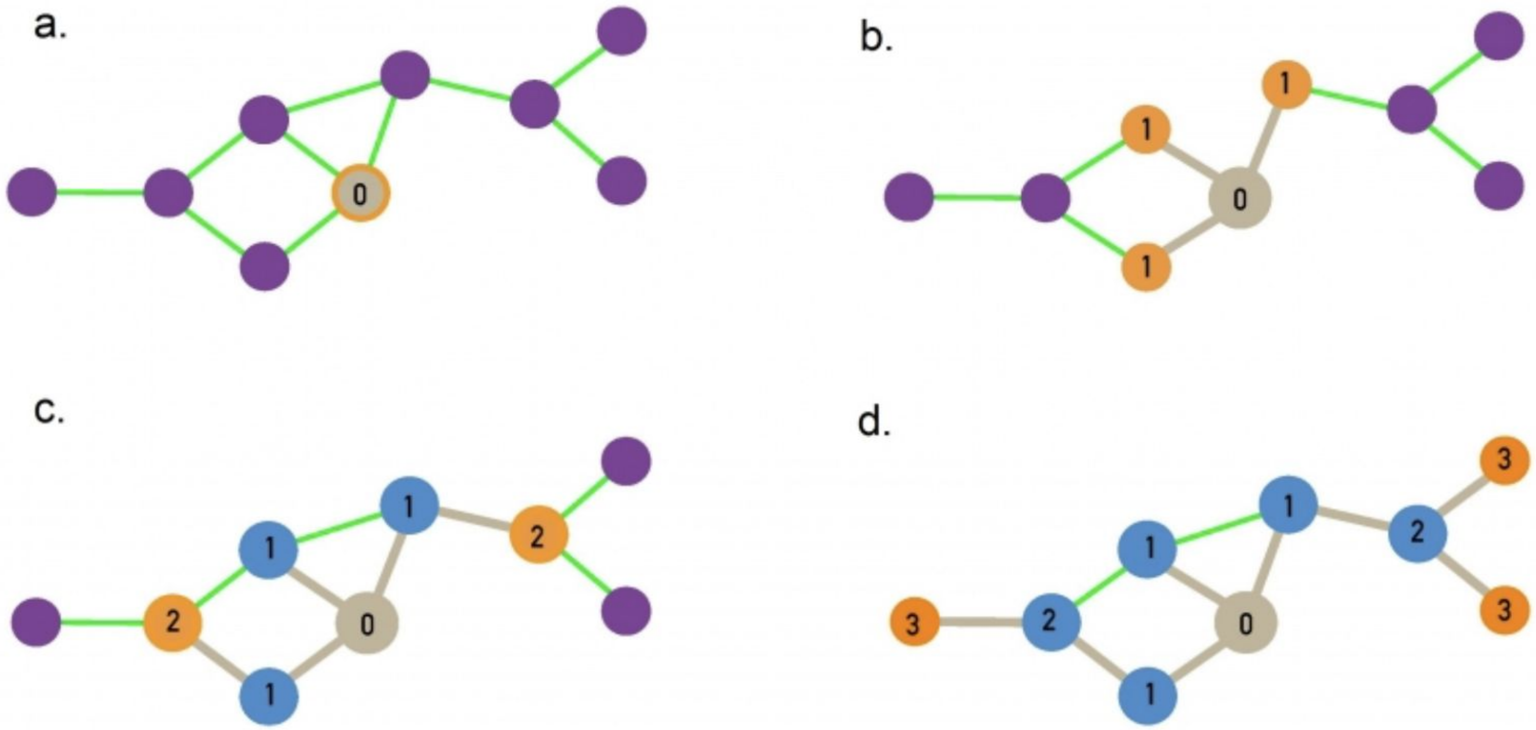
Average (shortest) path length, <*d*>

$$< d >= \frac{1}{N(N-1)} \sum_{i,j=1, i \neq j}^{N} d_{ij}$$

Calculating Distances: The Breadth-First Search Algorithm



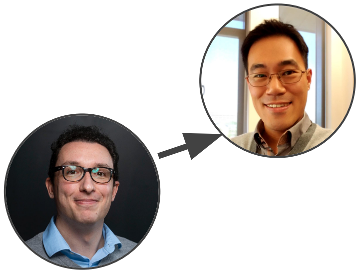
Calculating Distances: The Breadth-First Search Algorithm



You probably have paths connecting you to a large fraction of the world's population!

You probably have friends who grew up in other countries.

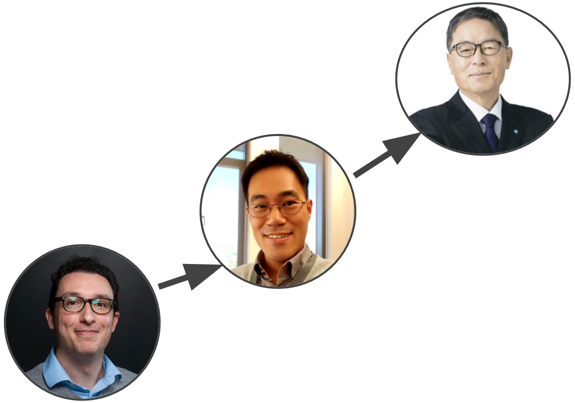
↳ You have a path distance of 1 (containing a single edge) to each of them.



You probably have paths connecting you to a large fraction of the world's population!

You probably have friends who grew up in other countries.
↳ You have a path distance of 1 (containing a single edge) to each of them.

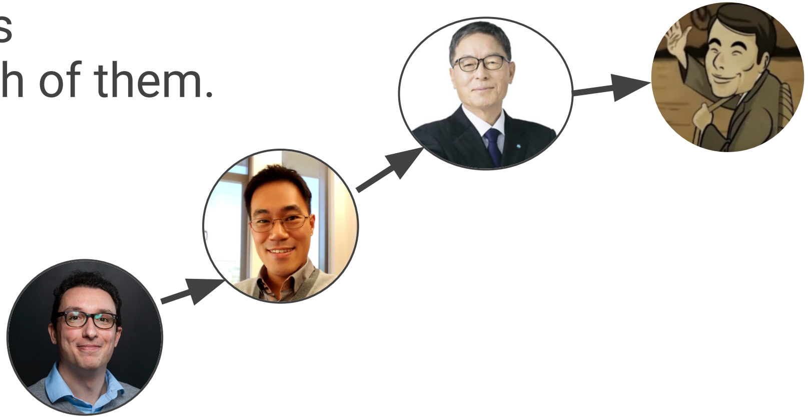
Now consider, say, the parents of these friends
↳ You have a path of distance 2 to each of them.



You probably have paths connecting you to a large fraction of the world's population!

You probably have friends who grew up in other countries.
↳ You have a path distance of 1 (containing a single edge) to each of them.

Now consider, say, the parents of these friends
↳ You have a path of distance 2 to each of them.
Then consider your friend's parents' friends
↳ You have a path of distance 3 to each of them.

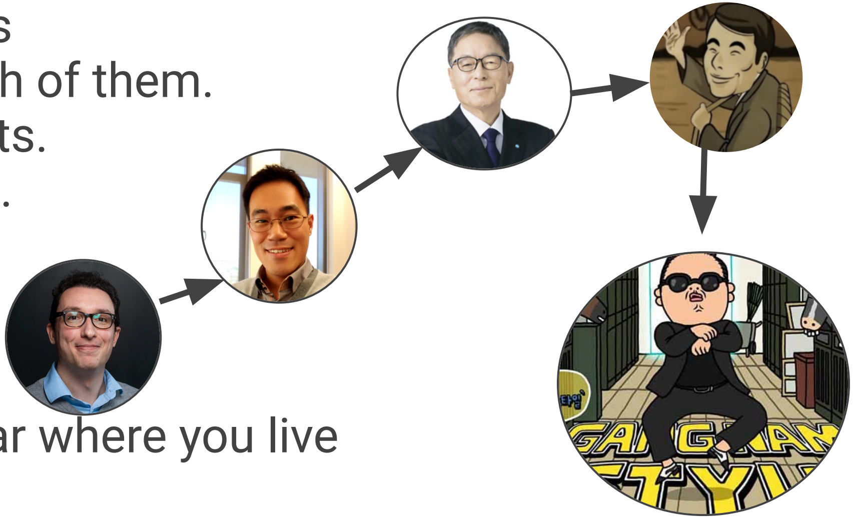


You probably have paths connecting you to a large fraction of the world's population!

You probably have friends who grew up in other countries.
↳ You have a path distance of 1 (containing a single edge) to each of them.

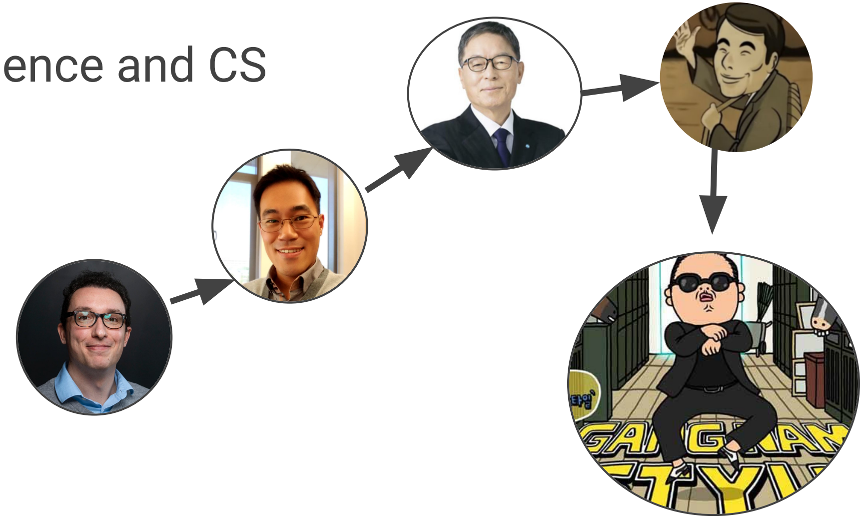
Now consider, say, the parents of these friends
↳ You have a path of distance 2 to each of them.
Then consider your friend's parents' friends
↳ You have a path of distance 3 to each of them.
Next consider their friends and descendants.
↳ You still have a path to each of them.

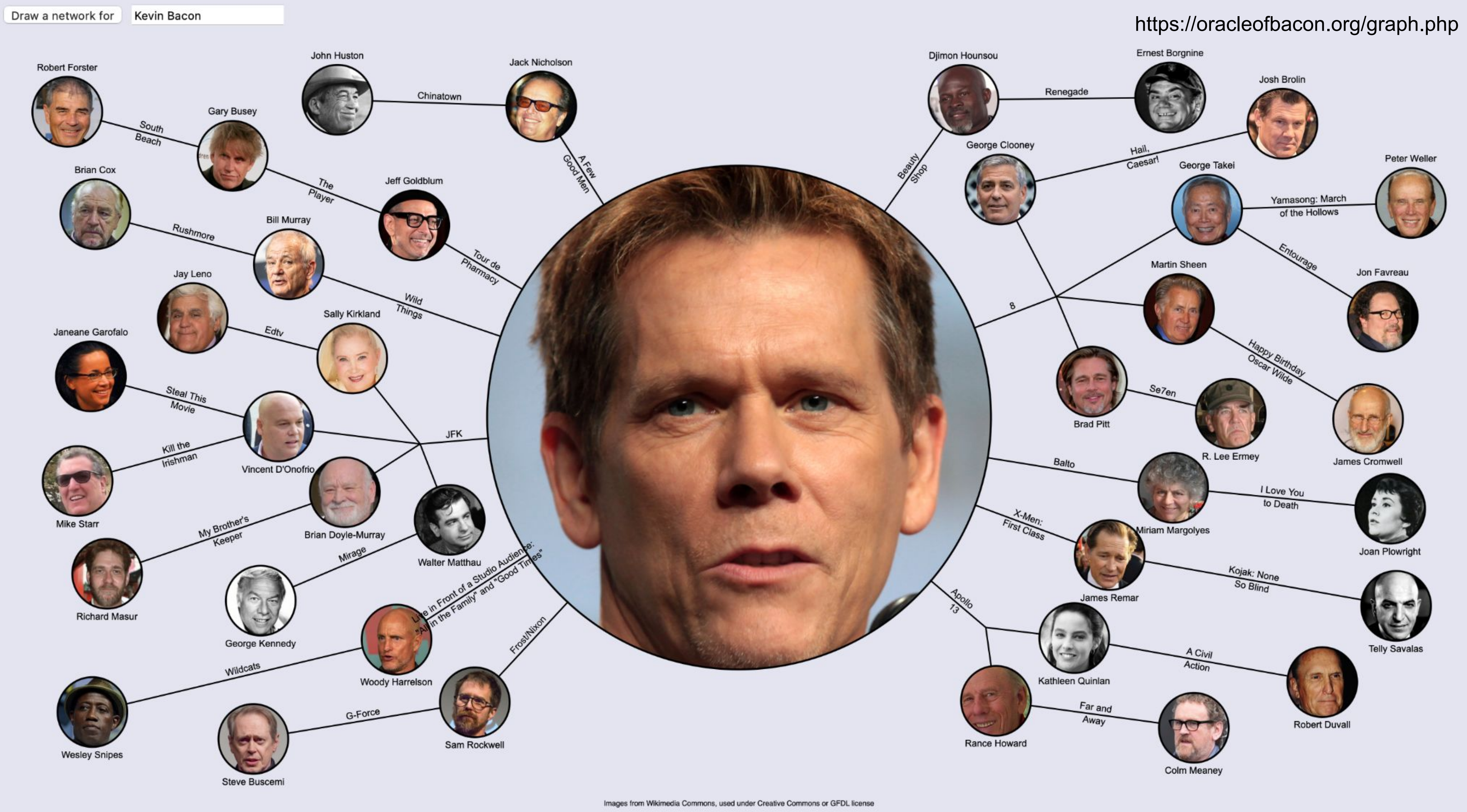
By now, we're talking about people who:
- Have never heard of you
- May speak a different language
- May have never traveled anywhere near where you live



You probably have paths connecting you to a large fraction of the world's population!

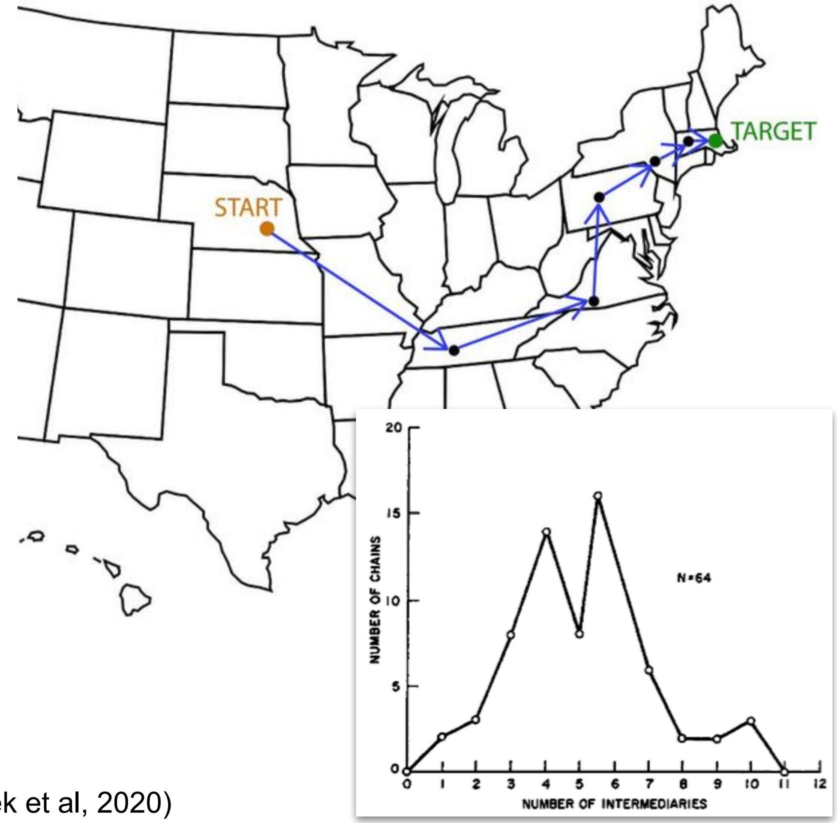
- Is this path objectively the shortest?
- Could Bogdan possibly be only 2 hops away from Psy?
- How can he search the shortest path without doing a BFS on a network map?
- This is a research topic in network science and CS
  → Decentralized search problem





Draw a network for
Kevin Bacon
https://oracleofbacon.org/graph.php
Robert Forster
South Beach
Gary Busey
John Huston
Chinatown
Jack Nicholson
A Few Good Men
Brian Cox
The Player
Jeff Goldblum
Rushmore
Bill Murray
Tour de Pharmacy
Jay Leno
Wild Things
Edtv
Sally Kirkland
Janeane Garofalo
Steal This Movie
Vincent D'Onofrio
JFK
Kill the Irishman
Mike Starr
Brian Doyle-Murray
My Brother's Keeper
Richard Masur
Walter Matthau
Mirage
George Kennedy
Live in Front of a Studio Audience: "All in the Family" and "Good Times"
Woody Harrelson
Wildcats
Wesley Snipes
Frost/Nixon
Sam Rockwell
G-Force
Steve Buscemi
Djimon Hounsou
Renegade
Ernest Borgnine
Beauty Shop
Josh Brolin
George Clooney
Hail, Caesar!
George Takei
Peter Weller
Yamasong: March of the Hollows
Entourage
Jon Favreau
Martin Sheen
8
Happy Birthday Oscar Wilde
Brad Pitt
Se7en
R. Lee Ermey
James Cromwell
Balto
Miriam Margolyes
I Love You to Death
Joan Plowright
X-Men: First Class
James Remar
Kojak: None So Blind
Telly Savalas
Apollo 13
Kathleen Quinlan
A Civil Action
Robert Duvall
Rance Howard
Far and Away
Colm Meaney
Images from Wikimedia Commons, used under Creative Commons or GFDL license

The Small-World Phenomenon / Six Degrees of Separation



(Baek et al, 2020)

Milgram and colleagues', 1960s:

Asked 296 random "starters" to try forwarding a letter to a "target" person, a stockbroker in Boston, via someone they knew on a first-name basis, with the same instructions.

Among the 64 chains that succeeded in reaching the target; the median length was six!

Can every node in a graph reach every other node by a path?

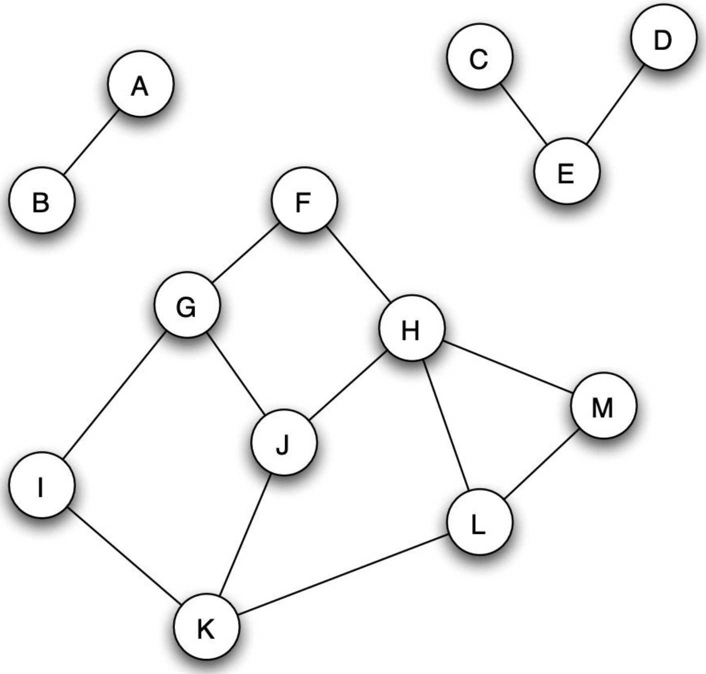
“Yes” in the 13-node 1970 Arpanet graph.

We say the graph is **connected** if for every pair of nodes, there is a path between them.

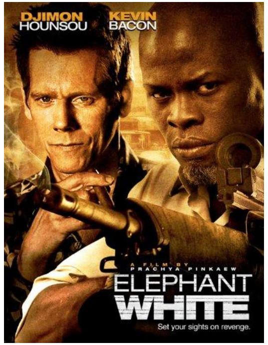
(In fact, every edge in the 1970 Arpanet belongs to a cycle, to increase robustness, e.g., a construction crew accidentally cutting through one cable.)

“No” in general. “No” for Milgram. Probably “no” in your world friendship network.

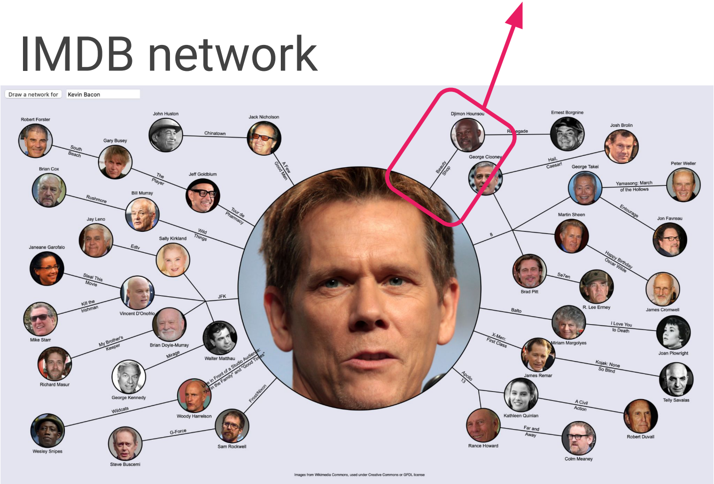
But, we get **connected components.**

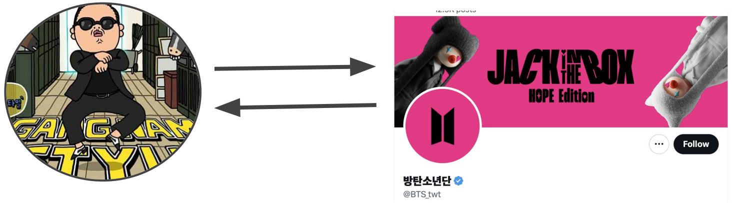


Homework 1: What is the average path length?



IMDB network





S. Korean Twitter
@mention network

Homework 1: What is the average path length?

A path length can be measured between any two nodes in the same component.

In this homework, you will compute the path lengths of all the dyads that include **Kevin Bacon** in the IMDB dataset (and the dyads that include **Psy** in the Twitter dataset) and study the path length distribution.

Is Kevin Bacon reachable within six hops from anywhere in the network? What about Psy?

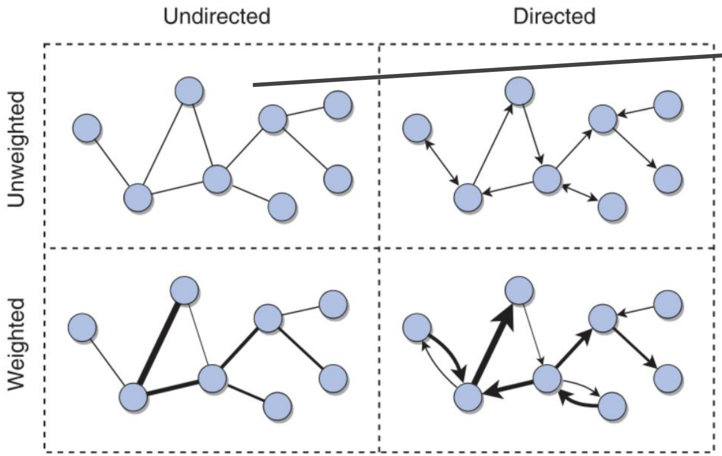
Do you think these two superstars will have significantly shorter or longer average path lengths than the global averages in their respective networks? **Why?**

Adjacency Matrices

A Graph Can Be Represented as a N by N Square Adjacency Matrix

The rows represent the “source” node (the senders) and columns represent the “target” nodes (the receivers).

The cells represent the presence (=1) or absence (=0) of an edge (i.e., whether they are **adjacent**).



	a	b	c	d	e	f	g	h
a		1						
b	1		1	1				
c		1		1				
d		1	1		1	1		
e				1			1	1
f				1				
g					1			
h					1			

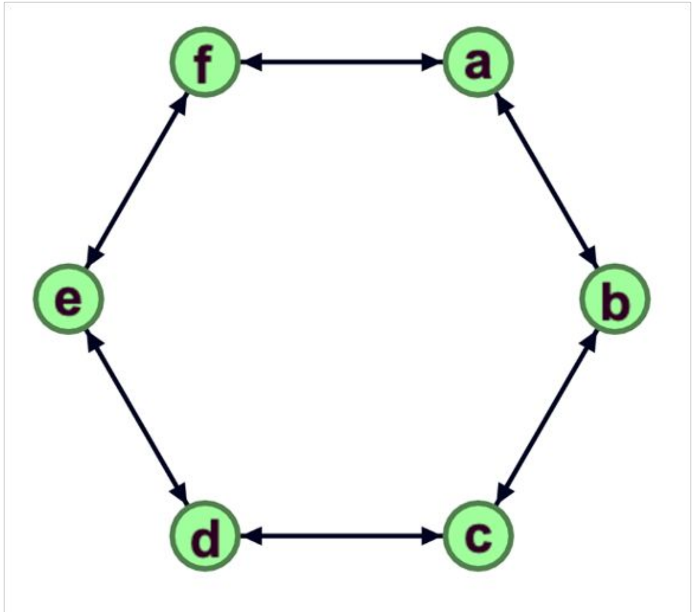
Matrix Representation of a Graph

	a	b	c	d	e	f
a	x	1				1
b	1	x	1			
c		1	x	1		
d			1	x	1	
e				1	x	1
f	1				1	x

Q: How does this matrix representation look like when visualized?

Matrix Representation of a Graph

	a	b	c	d	e	f
a	x	1				1
b	1	x	1			
c		1	x	1		
d			1	x	1	
e				1	x	1
f	1				1	x



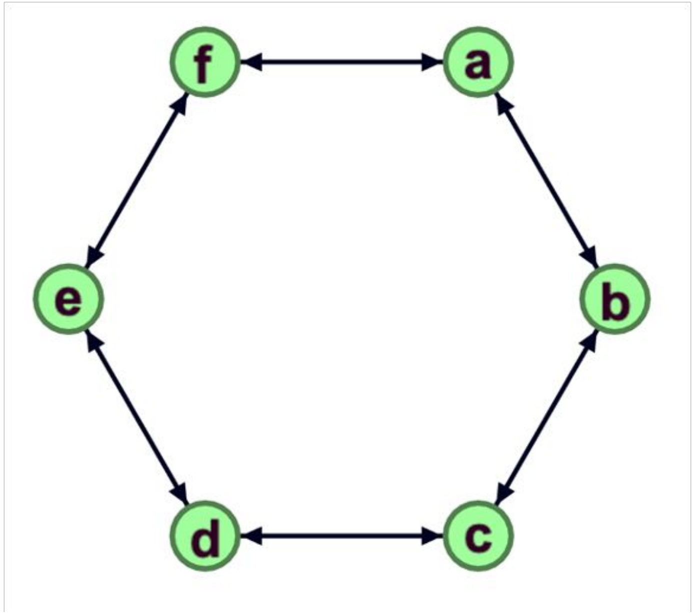
Graph Properties through Matrix Operations

Properties of a graph can be computed through matrix operations

Example:

Node degree: Number of edges connected to a node

- Out-degree: Number of edges where a node is the source
- In-degree: Number of edges where a node is the target



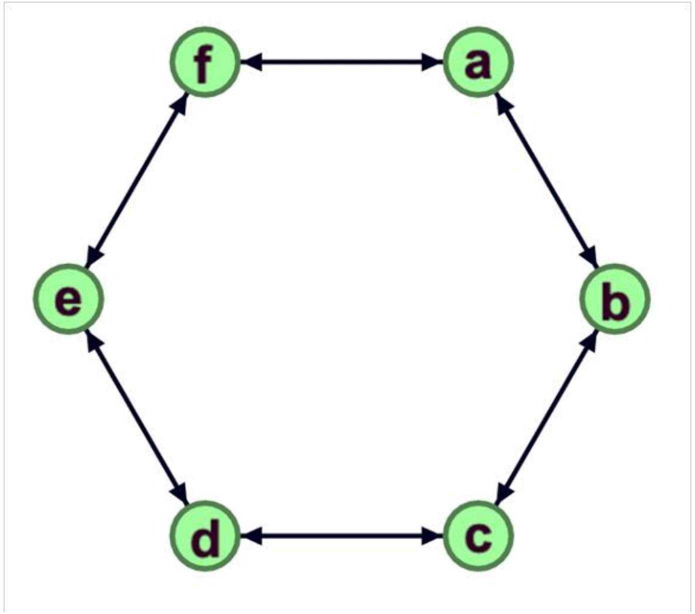
Graph Properties through Matrix Operations

Properties of a graph can be computed through matrix operations

Example:

Out-degree of every node is 2

In-degree of every node is also 2



Graph Properties through Matrix Operations

Properties of a graph can be computed through matrix operations

Example: Node Degree

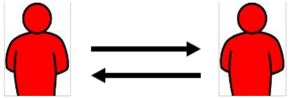
	a	b	c	d	e	f
a	x	1				1
b	1	x	1			
c		1	x	1		
d			1	x	1	
e				1	x	1
f	1				1	x

→ Sum all the 1's of node *a*'s row in the adjacency matrix, *A* → $k_i = A_{i+}$

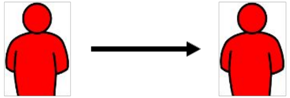
Note, in an undirected network, row sum and column sum → $k_i = A_{i+} = A_{+i}$

Graph Properties through Matrix Operations

One can compute the number of mutual (*M*), asymmetric (*A*), and null (*N*) dyads with the adjacency matrix, *X*:



Mutual $$M = \sum_{i<j} X_{ij}X_{ji}$$



Asymmetric $$A = X_{++} - 2M$$

Null $$N = \binom{g}{2} - A - M$$

X_{++}: Total number of directed edges in the graph

g: Total number of nodes in the graph

	a	b	c	d	e	f
a	0	1	0	0	0	1
b	1	0	1	0	0	0
c	0	1	0	1	0	0
d	0	0	1	0	1	0
e	0	0	0	1	0	1
f	1	0	0	0	1	0

Graph Properties through Matrix Operations

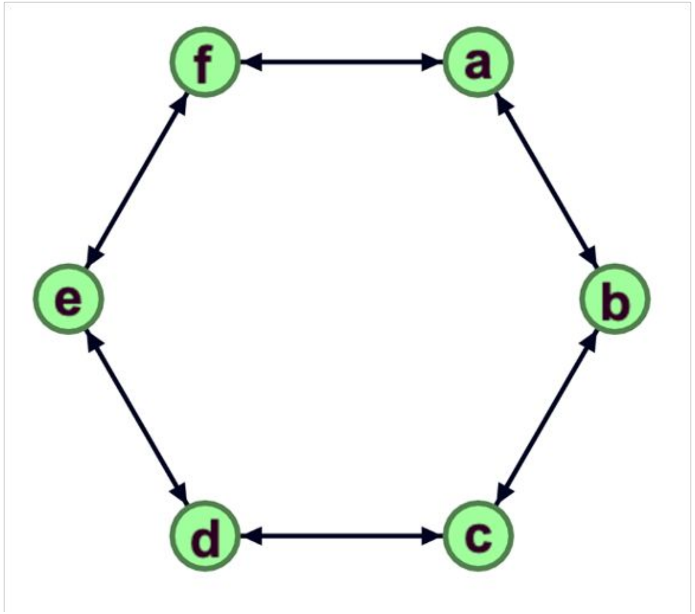
Properties of a graph can be computed through matrix operations

Another Example:

Path length: Minimum number of hops from one node to another node

Path length from *a* to *d* is 3

(a → f → e → d) or (a → b → c → d)



Graph Properties through Matrix Operations

Another example:

	a	b	c	d	e	f
a	0	1	0	0	0	1
b	1	0	1	0	0	0
c	0	1	0	1	0	0
d	0	0	1	0	1	0
e	0	0	0	1	0	1
f	1	0	0	0	1	0

Matrix multiplication for adjacency matrix *X*:

All *PL*=2 can be found by multiplying *X* to itself

Graph Properties through Matrix Operations

Properties of a graph can be computed through matrix operations

Cell value is the number of shortest paths at path length = 2

$$XX = X \times X =$$

X

	a	b	c	d	e	f
a	0	1	0	0	0	1
b	1	0	1	0	0	0
c	0	1	0	1	0	0
d	0	0	1	0	1	0
e	0	0	0	1	0	1
f	1	0	0	0	1	0

$\times$

X

	a	b	c	d	e	f
a	0	1	0	0	0	1
b	1	0	1	0	0	0
c	0	1	0	1	0	0
d	0	0	1	0	1	0
e	0	0	0	1	0	1
f	1	0	0	0	1	0

=

	a	b	c	d	e	f
a	2	0	1	**0**	1	0
b	0	2	0	1	0	1
c	1	0	2	0	1	0
d	0	1	0	2	0	1
e	1	0	1	0	2	0
f	0	1	0	1	0	2

Graph Properties through Matrix Operations

Properties of a graph can be computed through matrix operations

Cell value is the number of shortest paths at path length = 2

There is 1 path of length=2 from a to c

$XX =$

X

	a	b	c	d	e	f
a	0	1	0	0	0	1
b	1	0	1	0	0	0
c	0	1	0	1	0	0
d	0	0	1	0	1	0
e	0	0	0	1	0	1
f	1	0	0	0	1	0

$\times$

X

	a	b	c	d	e	f
a	0	1	0	0	0	1
b	1	0	1	0	0	0
c	0	1	0	1	0	0
d	0	0	1	0	1	0
e	0	0	0	1	0	1
f	1	0	0	0	1	0

$=$

	a	b	c	d	e	f
a	2	0	1	**0**	1	0
b	0	2	0	1	0	1
c	1	0	2	0	1	0
d	0	1	0	2	0	1
e	1	0	1	0	2	0
f	0	1	0	1	0	2

Graph Properties through Matrix Operations

Properties of a graph can be computed through matrix operations

The shortest path from a to d is found when the (a, d) cell becomes larger than zero

Cell value is the number of shortest paths at path length = 3

XXX =

	a	b	c	d	e	f
a	2	0	1	0	1	0
b	0	2	0	1	0	1
c	1	0	2	0	1	0
d	0	1	0	2	0	1
e	1	0	1	0	2	0
f	0	1	0	1	0	2

×

	a	b	c	d	e	f
a	0	1	0	0	0	1
b	1	0	1	0	0	0
c	0	1	0	1	0	0
d	0	0	1	0	1	0
e	0	0	0	1	0	1
f	1	0	0	0	1	0

=

	a	b	c	d	e	f
a	0	3	0	**2**	0	3
b	3	0	3	0	2	0
c	0	3	0	3	0	2
d	**2**	0	3	0	3	0
e	0	2	0	3	0	3
f	3	0	2	0	3	0

Graph Properties through Matrix Operations

As you successively multiply, you are sort of "walking" one successive "step" or "hop" at a time.

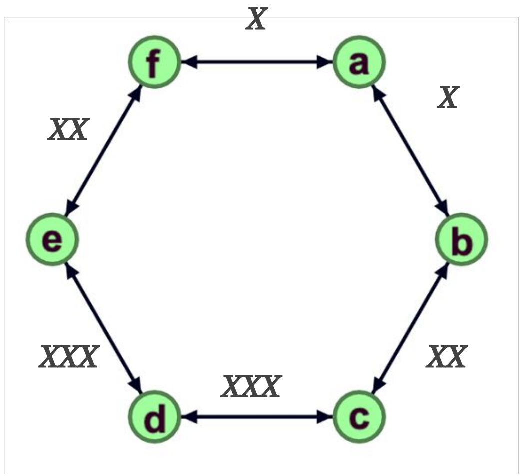
- Walks and Reachability:

  Powers of a matrix give the number of walks from one node to another node.

  X (adjacency matrix): length 1 walk

  X^2 : Number of length 2 walks

  X^{g-1} : Number of length g-1 walks. (g=number of nodes)



Summary

Graph theory as our basic framework

Degree, degree distributions

Edges and dyads

Paths, shortest paths

Connectivity and connected components

Six degrees of separation

Adjacency matrices, definition and operations