Markov Chains

Geometric Algorithms
Lecture 13

CAS CS 132

Practice Problem

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \begin{bmatrix} x_1 + x_2 \\ 2x_2 \\ x_2 + bx_3 \end{bmatrix}$$

not invertible

For what values of b is the above transformation singular? Explain your answer

Find the inverse of the matrix implementing the above transformation, given $b = 1$

Solution

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \begin{bmatrix} x_1 + x_2 \\ 2x_2 \\ x_2 + bx_3 \end{bmatrix}$$

$b = 0$

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \qquad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} 0 \\ 0 \\ b \end{bmatrix}$$

$$\vec{x} \mapsto \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & b \end{bmatrix} \vec{x} \qquad \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

-1

$$\begin{bmatrix} \boxed{1} & 1 & 0 \\ 0 & \boxed{2} & 0 \\ 0 & 0 & \boxed{b} \end{bmatrix}$$

$$\boxed{b = 0}$$

Solution

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \begin{bmatrix} x_1 + x_2 \\ 2x_2 \\ x_2 + bx_3 \end{bmatrix}$$

$b = 1$

$$\begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & \overset{-1}{1} & 0 & 1 & \overset{-1/2}{0} & 0 \\ 0 & 1 & 0 & 0 & 1/2 & 0 \\ 0 & \overset{-1}{1} & 1 & 0 & \overset{-1/2}{0} & 1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 0 & 0 & 1 & -1/2 & 0 \\ 0 & 1 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1 & 0 & -1/2 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & -1/2 & 0 \\ 0 & 1/2 & 0 \\ 0 & -1/2 & 1 \end{bmatrix}$$

Objectives

1. Motivate linear dynamical systems
2. Analyze Markov chains and their properties
3. Learn to solve for steady-states of Markov chains
4. Connect this to graphs and random walks

Keywords

linear dynamical systems

recurrence relations

linear difference equations

state vector

probability vector

stochastic matrix

Markov chain

steady-state vector

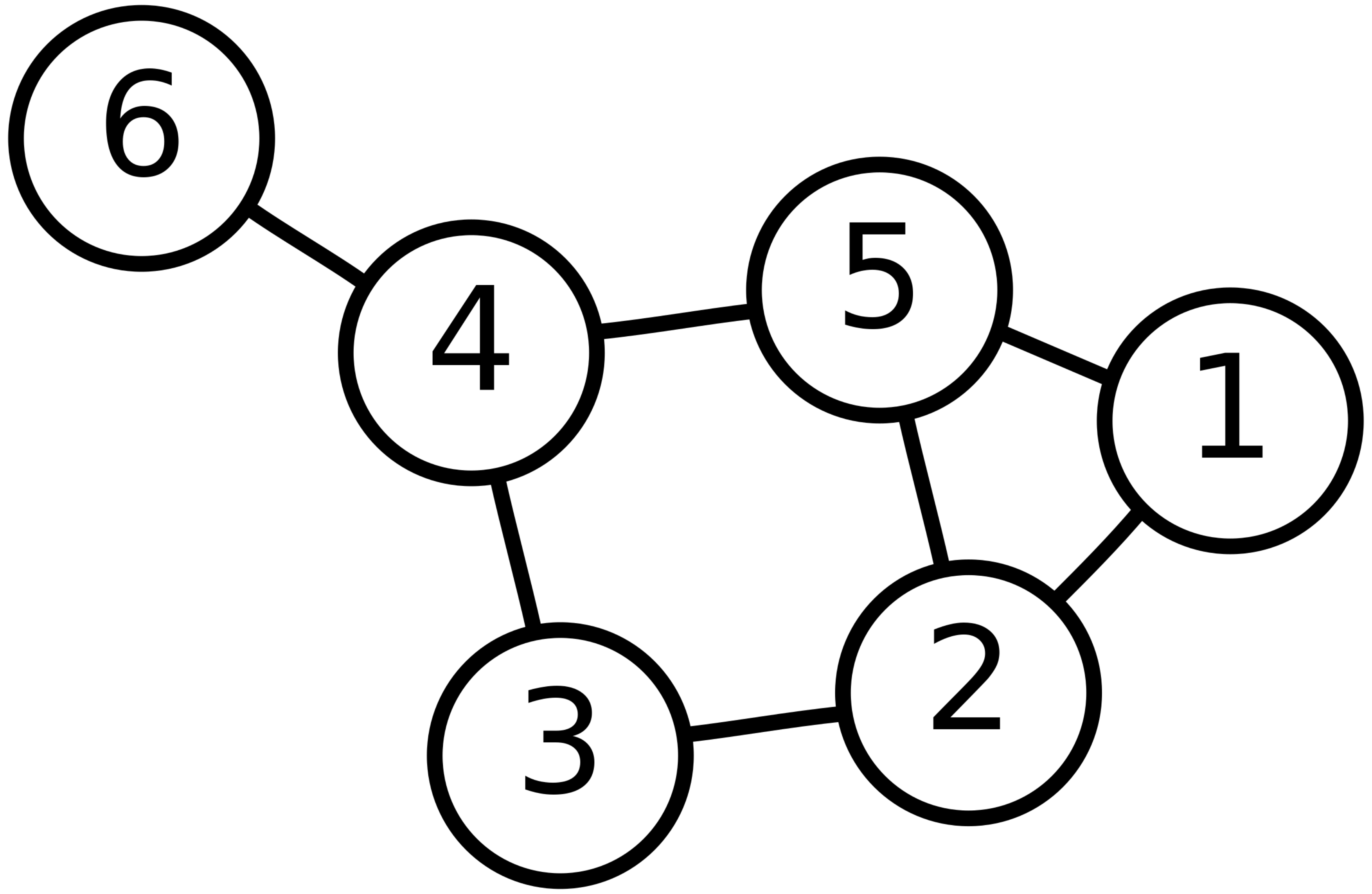
random walk

state diagram

Algebraic Graph Theory

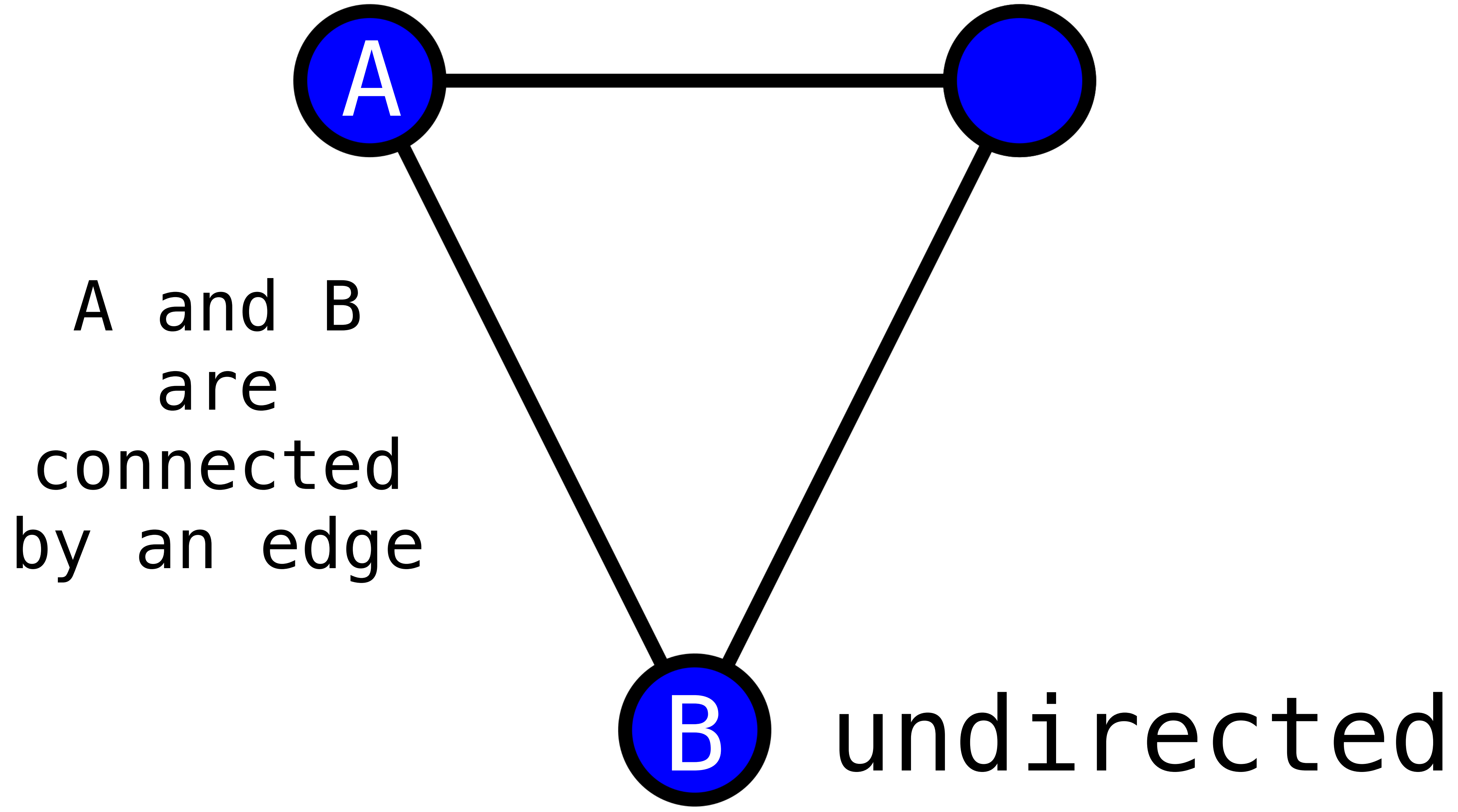
Graphs

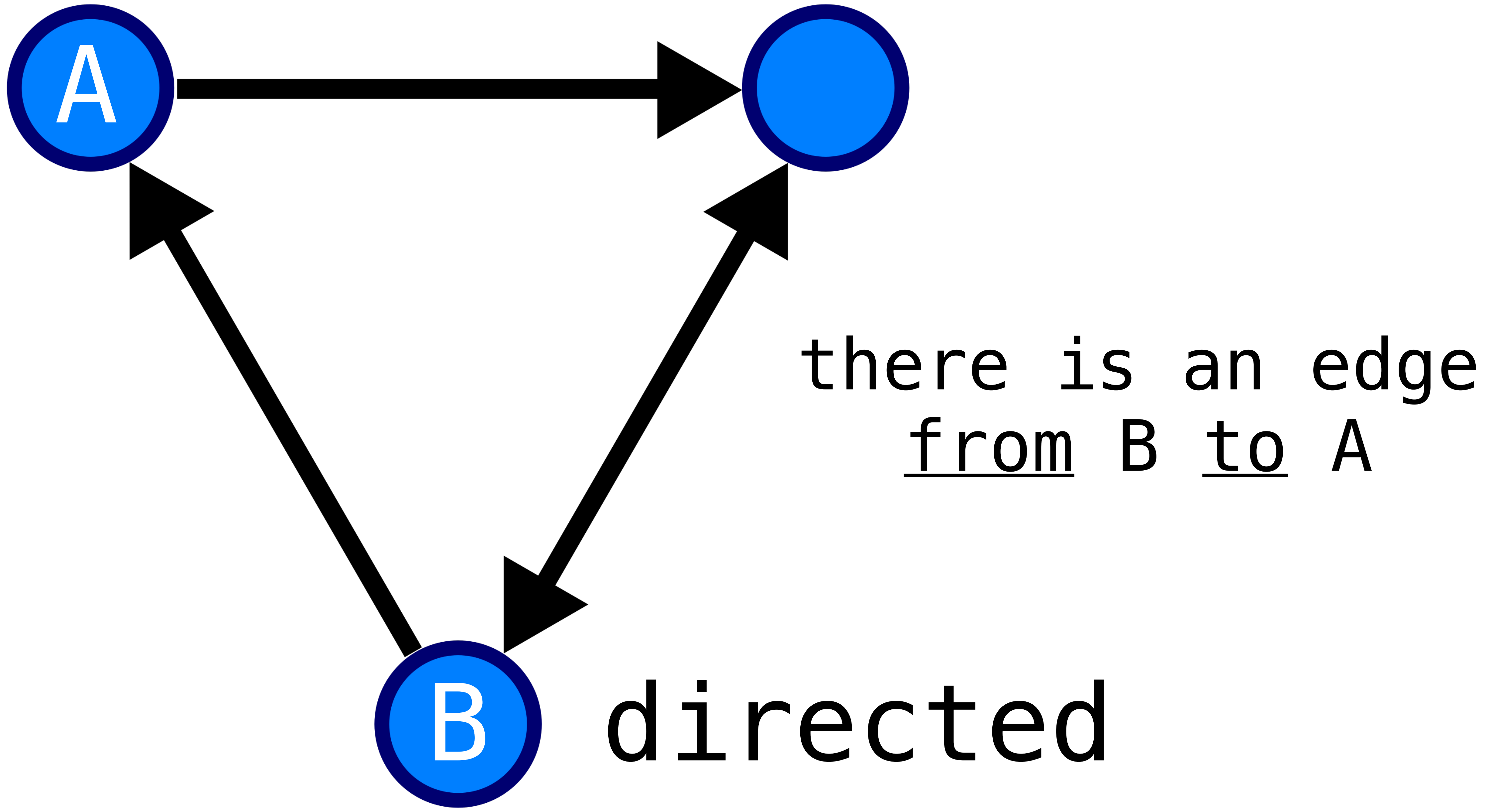
Definition (Informal). A **graph** is a collection of nodes with edges between them



Directed vs. Undirected Graphs

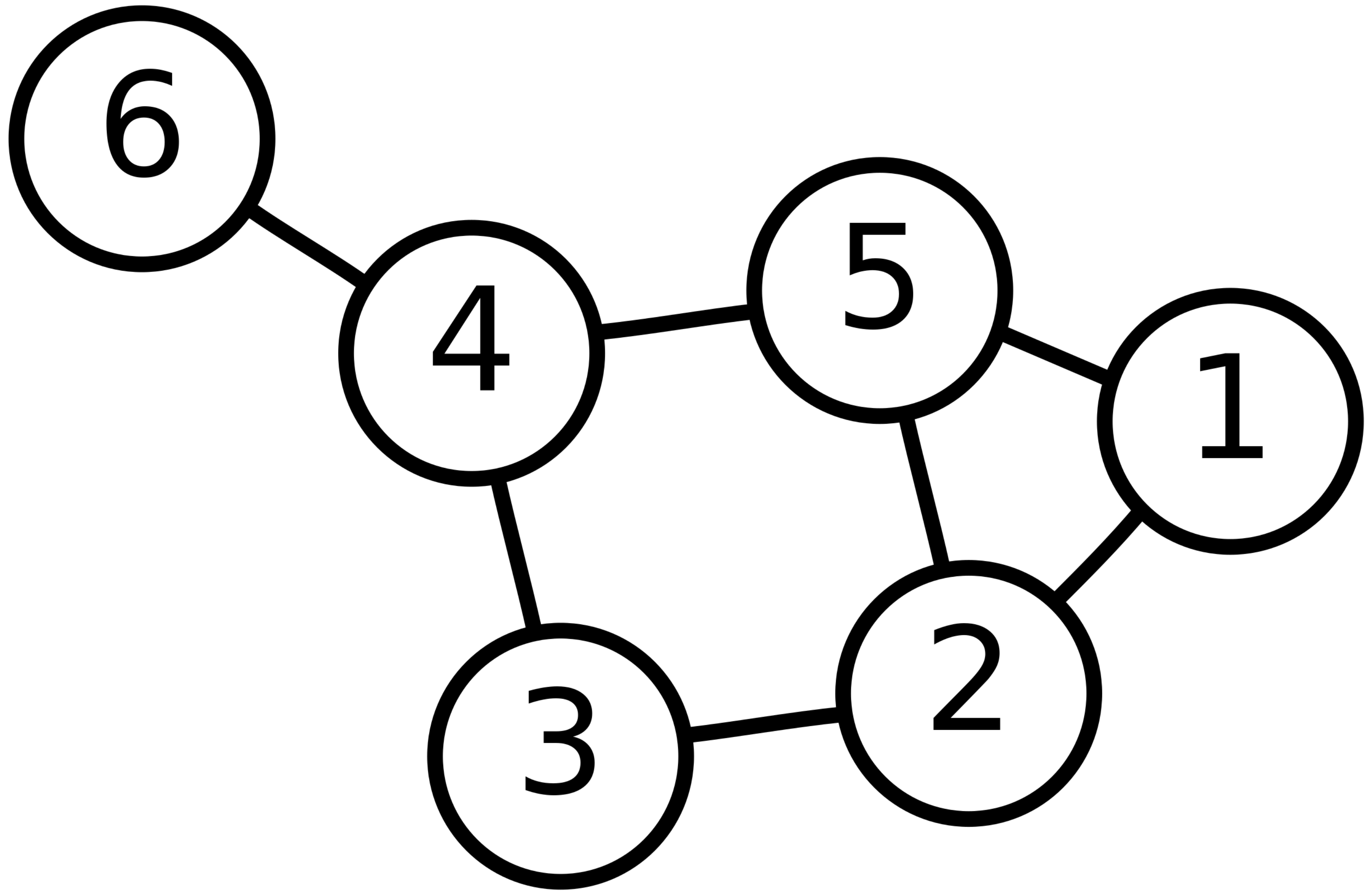
A graph is **directed** if its edges have a direction



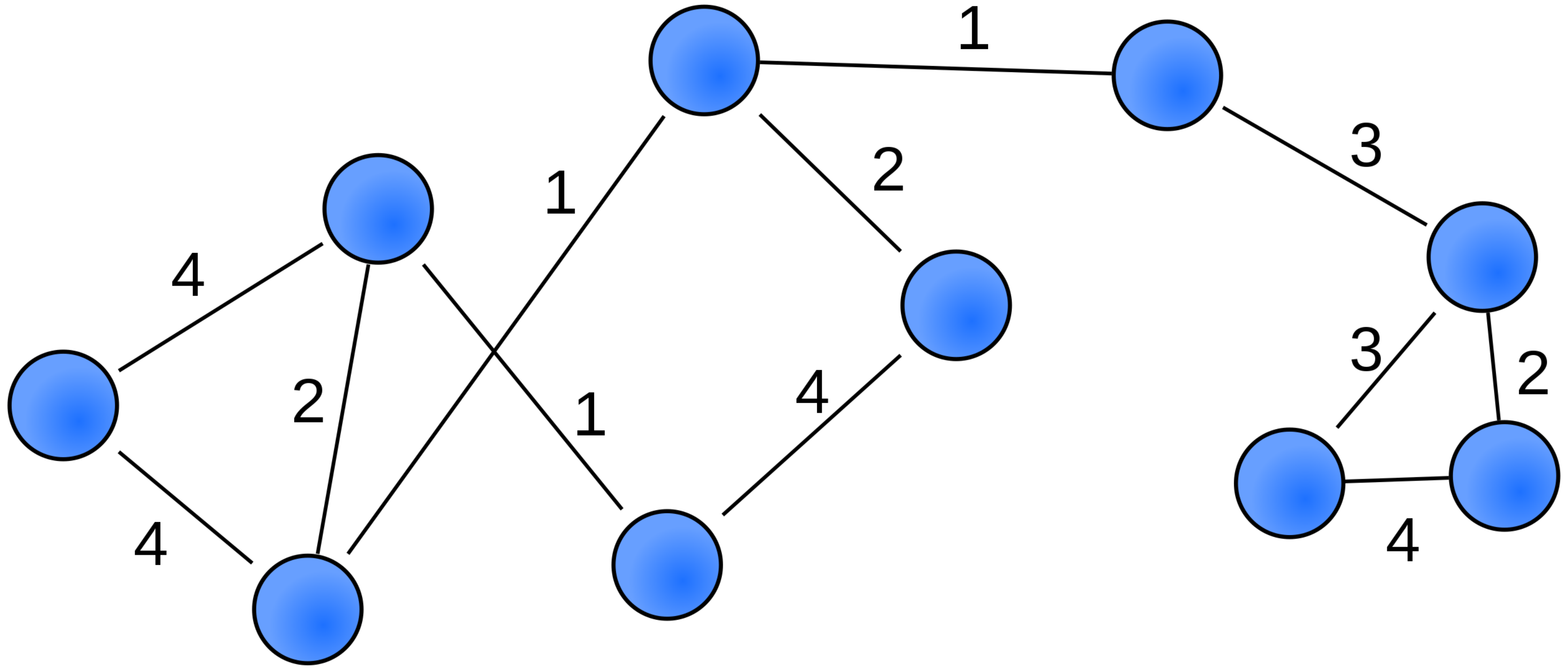


Weighted vs Unweighted graphs

A graph is **weighted** if its edges have associated values



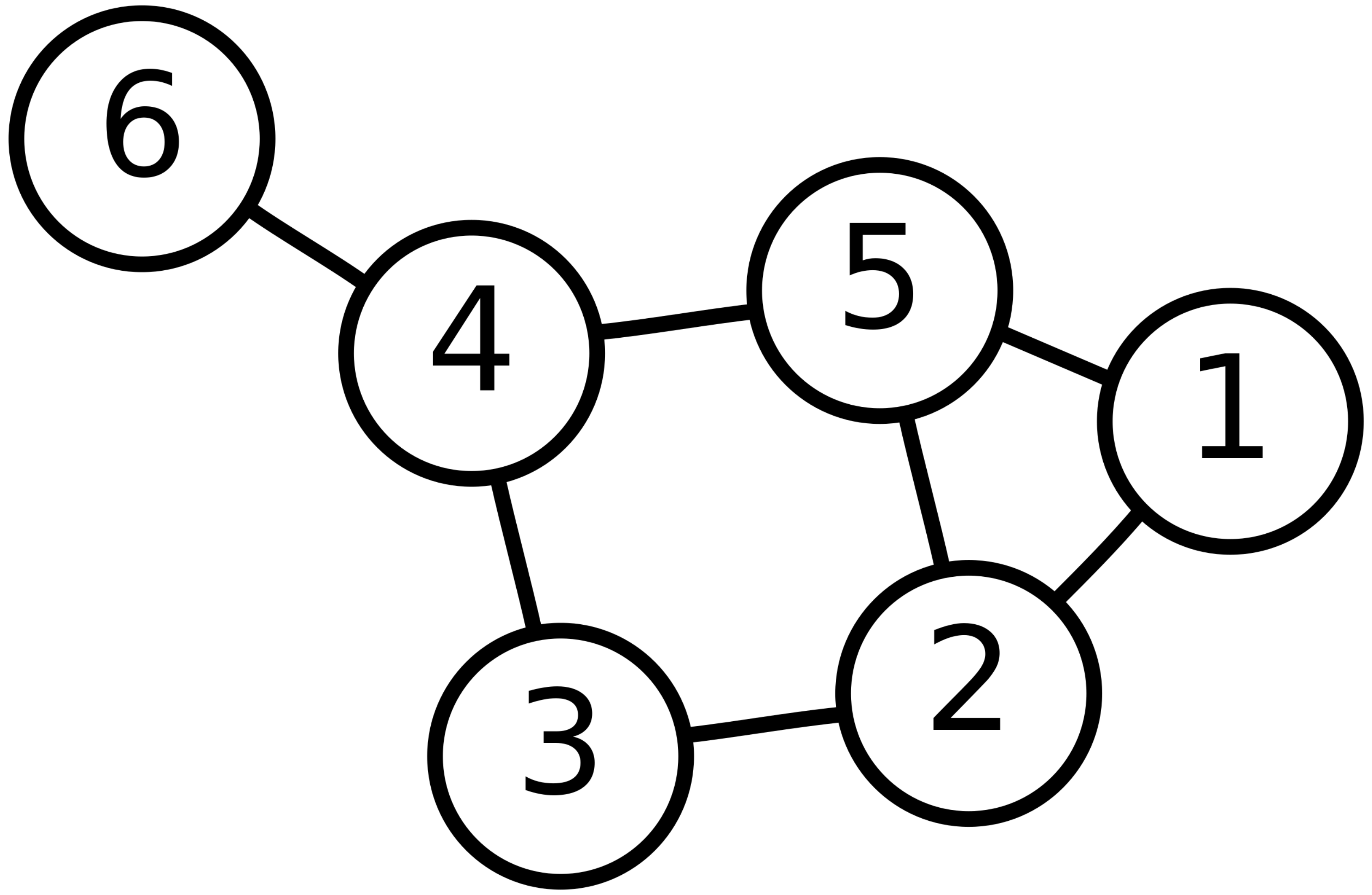
unweighted



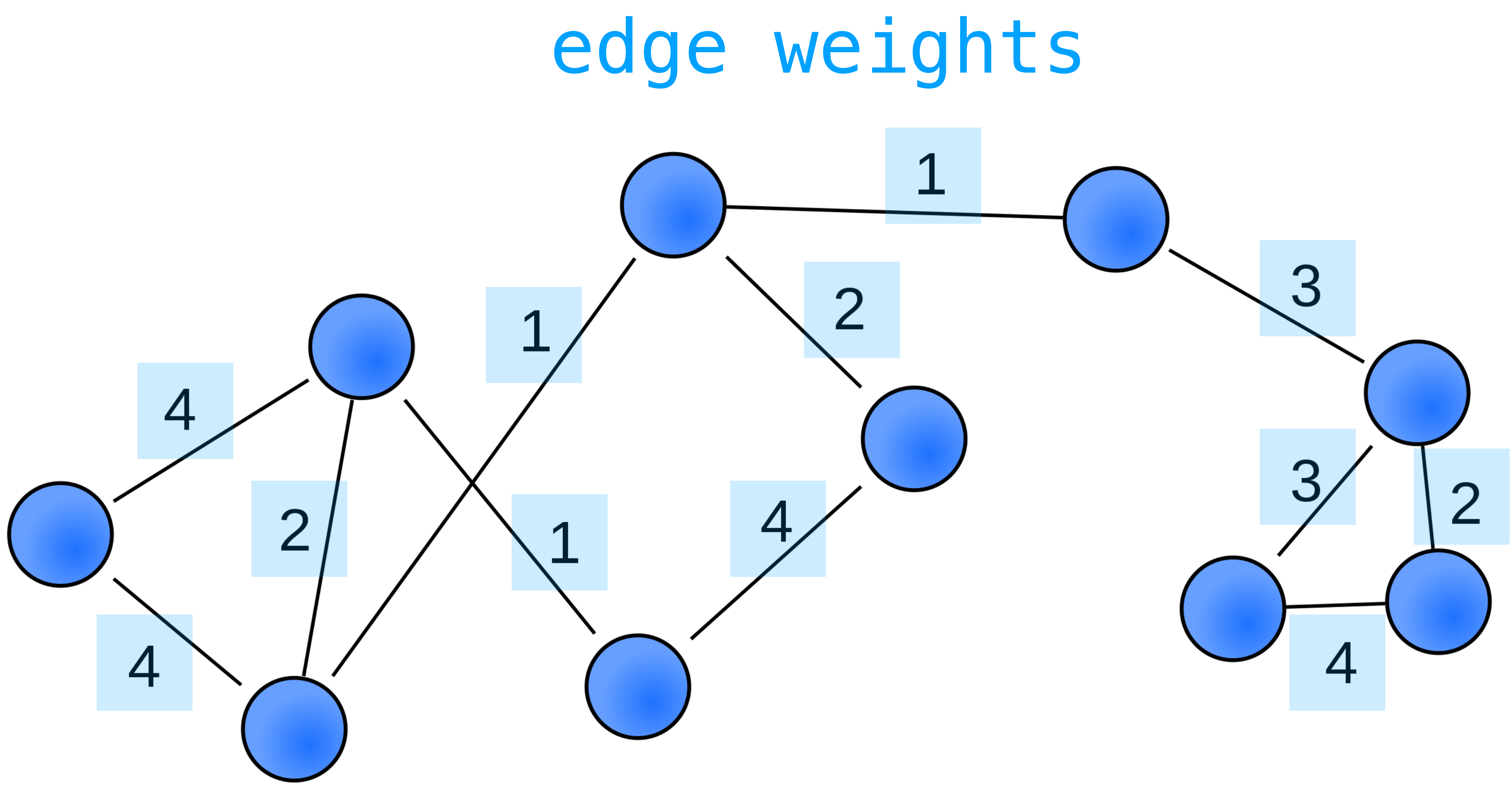
weighted

Weighted vs Unweighted graphs

A graph is **weighted** if its edges have associated values



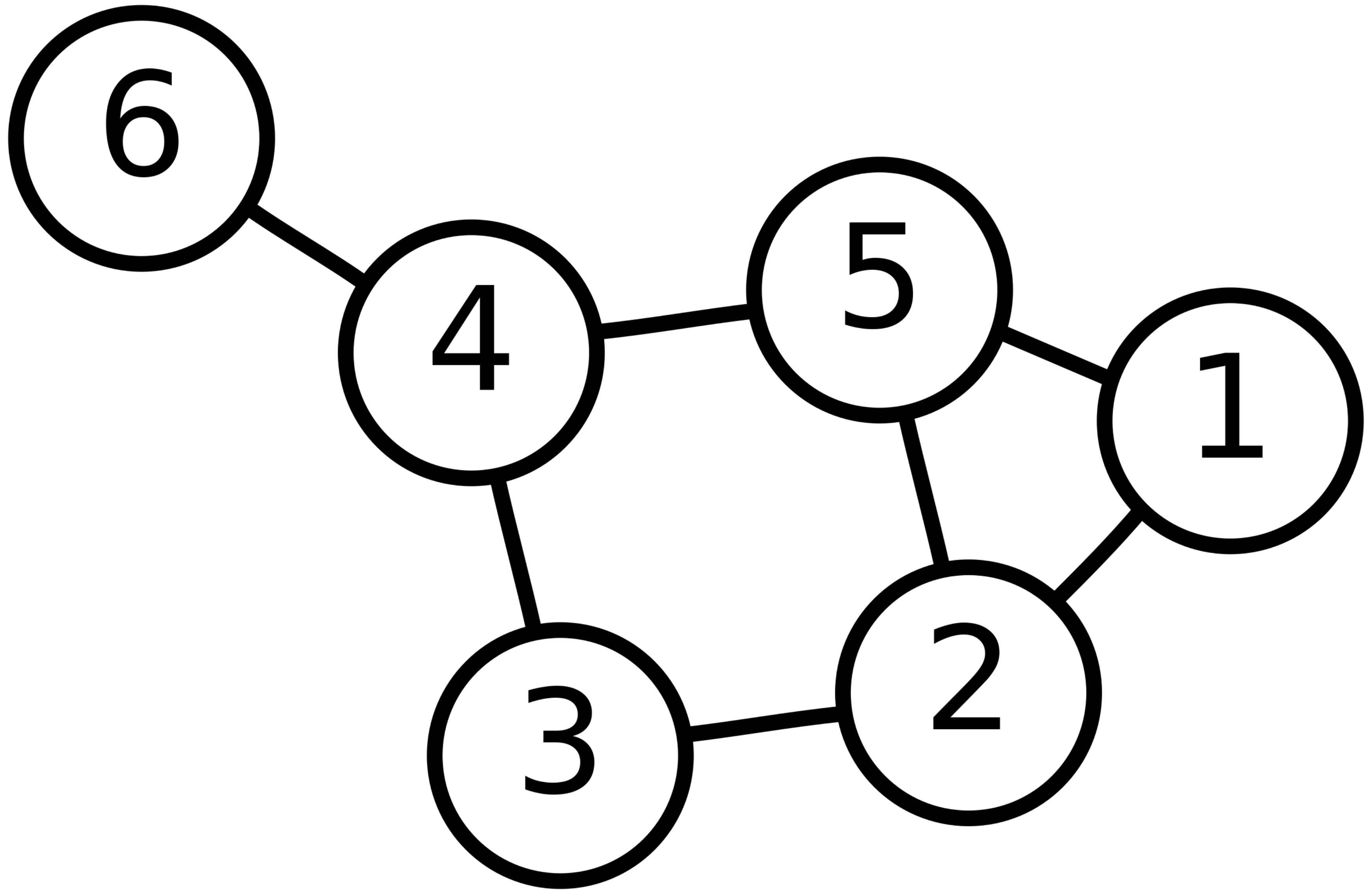
unweighted



weighted

Simple Graphs

A graph is **simple** if it is undirected, has no self loops, and no multi-edges



Four Kinds of Graphs

	directed	undirected
weighted	nodes are traffic lights edges are streets weights are number of lanes	nodes are musicians edges are collaborations weights are number of collaborations
unweighted	nodes are instagram users edges are follows	nodes are bodies of land edges are pedestrian bridges

Four Kinds of Graphs

	directed	undirected
weighted	nodes are traffic lights edges are streets weights are number of lanes	nodes are musicians edges are collaborations weights are number of collaborations
unweighted	nodes are instagram users edges are follows	nodes are bodies of land edges are pedestrian bridges Today

Four Kinds of Graphs

	directed	undirected
weighted	nodes are traffic lights edges are streets weights are number of lanes Markov Chains	nodes are musicians edges are collaborations weights are number of collaborations
unweighted	nodes are instagram users edges are follows	nodes are bodies of land edges are pedestrian bridges Today

Fundamental Question

Fundamental Question

How do we represent a graph formally in a computer?

Fundamental Question

How do we represent a graph formally in a computer?

There are a couple ways, but one way is to use matrices

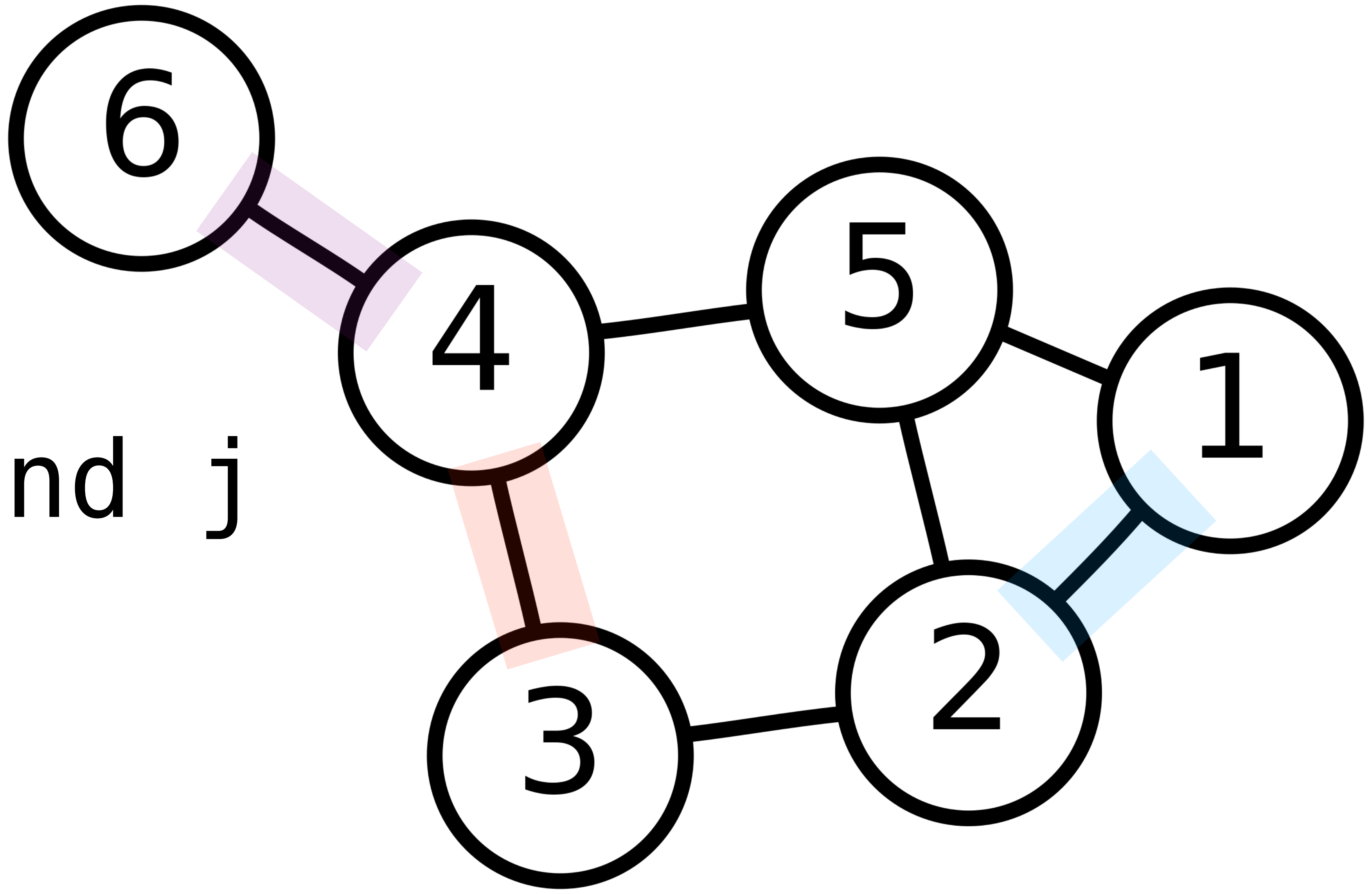
Adjacency Matrices

$$
\begin{array}{c}
\begin{array}{cccccc} & \color{deepskyblue}{A_{12}} & & \color{red}{A_{34}} & & \color{purple}{A_{46}} \end{array} \\
\begin{array}{c} \\ \color{deepskyblue}{A_{21}} \\ \\ \color{red}{A_{43}} \\ \\ \color{purple}{A_{64}} \end{array}
\begin{bmatrix}
0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0
\end{bmatrix}
\end{array}
$$

Let G be an simple graph with its nodes labeled by numbers 1 through n

We can create the **adjacency matrix** A for G as follows

$$A_{ij} = \begin{cases} 1 & \text{there is an edge between i and j} \\ 0 & \text{otherwise} \end{cases}$$



Symmetric Matrices

Definition. A $n \times n$ matrix is **symmetric** if

$$A^T = A$$

Example.

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

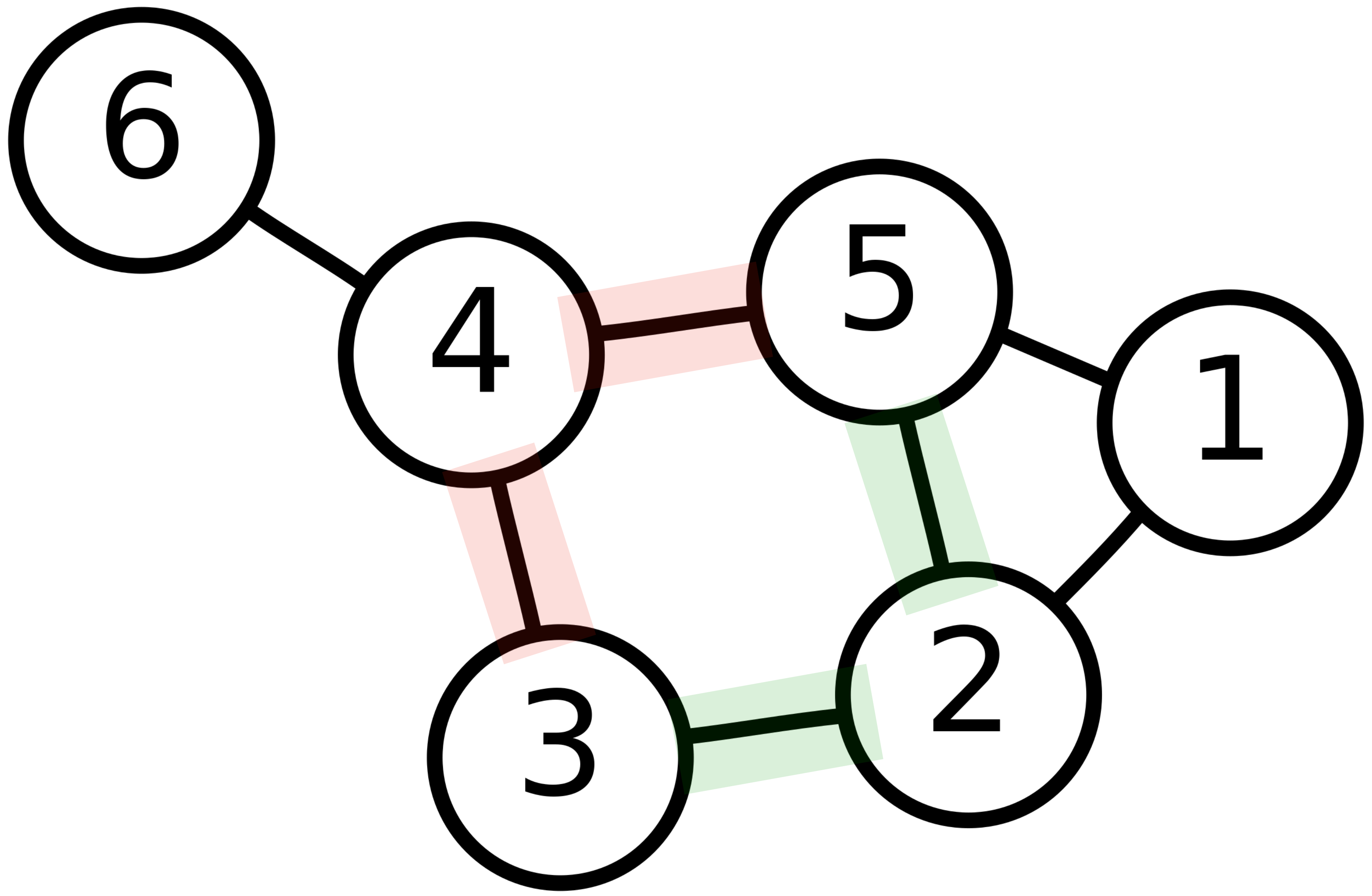
Once we have an adjacency matrix, we can do linear algebra on graphs

Example: Squared Adjacency Matrices

Given an adjacency matrix A*, can we interpret anything meaningful from* A^2*?*

Example: Squared Adjacency Matrices

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

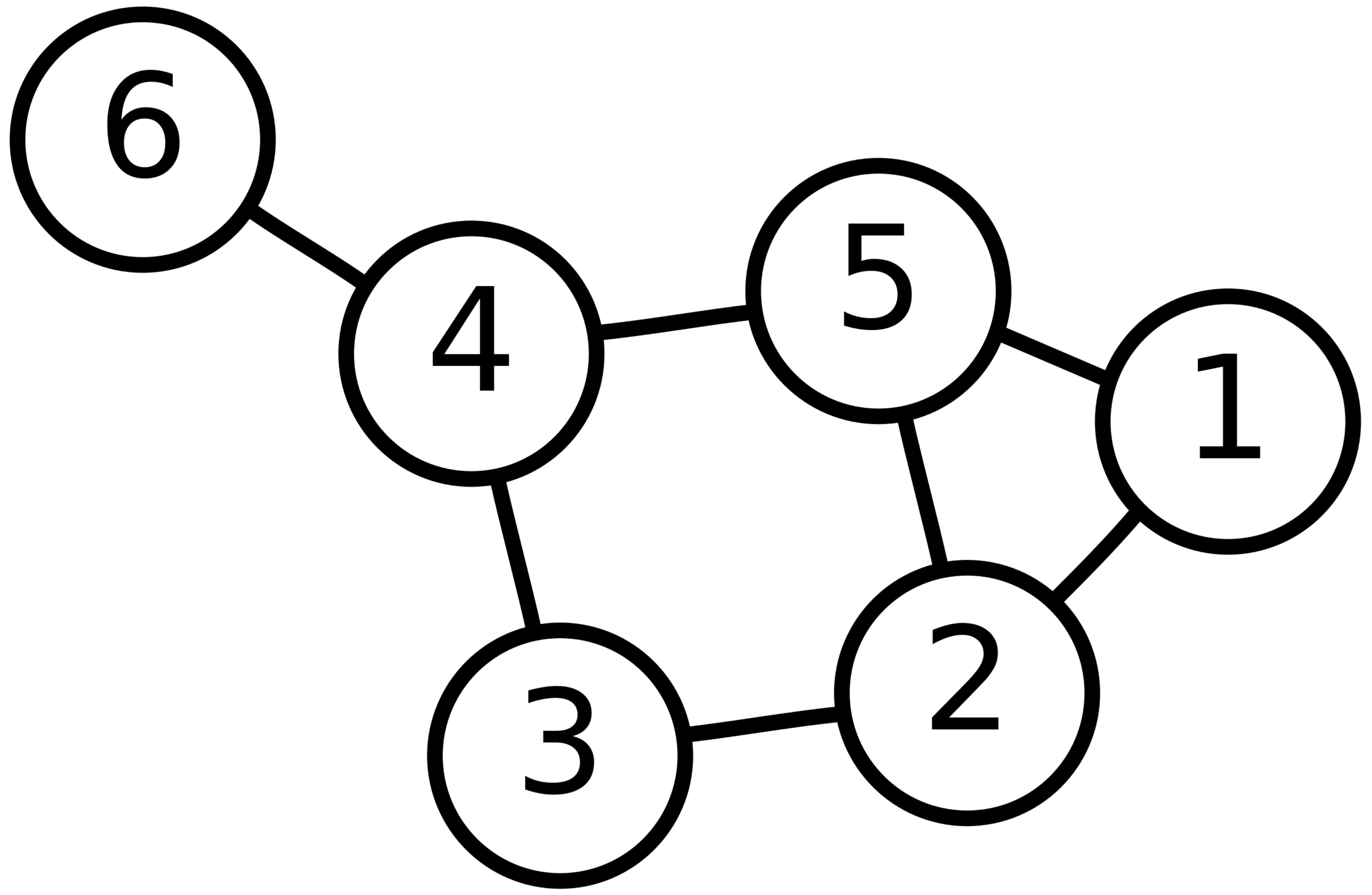


$A \; A$

$$(A^2)_{53} = 1(0) + 1(1) + 0(0) + 1(1) + 0(0) + 0(0) = 2$$

Example: Squared Adjacency Matrices

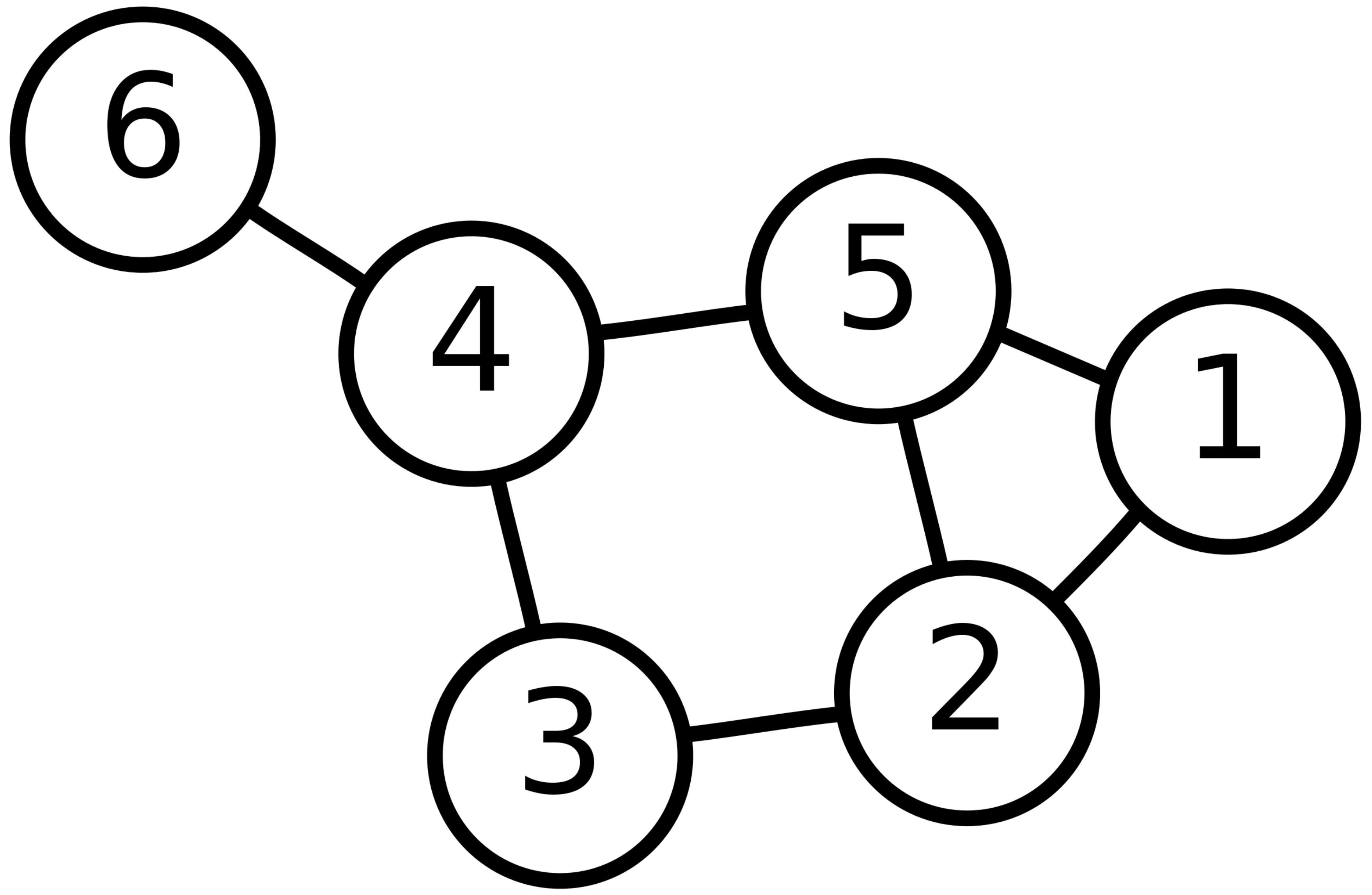
$$(A^2)_{ij} = A_{i1}A_{1j} + A_{i2}A_{2j} + \ldots + A_{in}A_{nj}$$



Example: Squared Adjacency Matrices

$$(A^2)_{ij} = A_{i1}A_{1j} + A_{i2}A_{2j} + \ldots + A_{in}A_{nj}$$

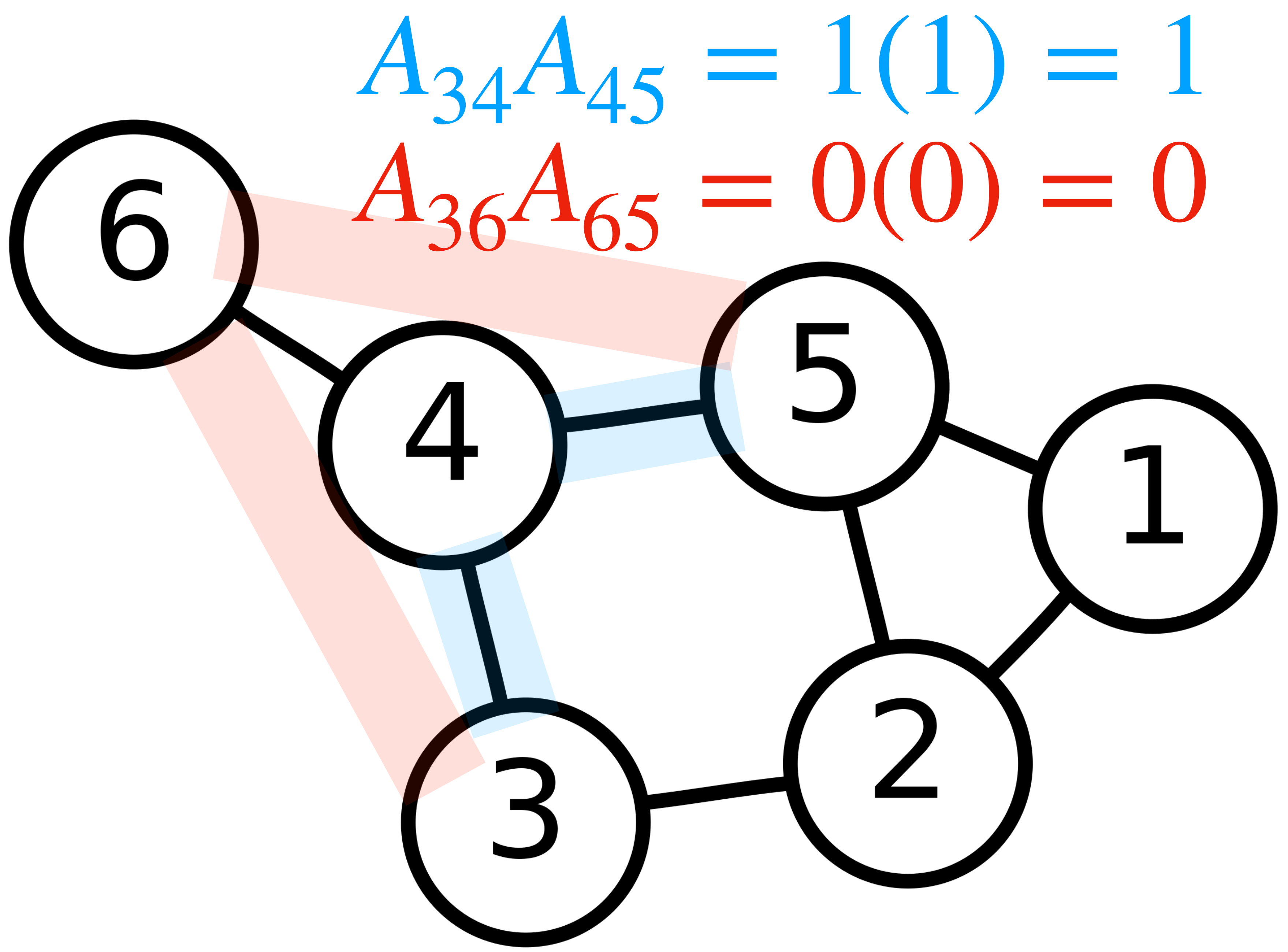
$$A_{ik}A_{kj} = \begin{cases} 1 & \texttt{there are edges i to k and k to j} \\ 0 & \texttt{otherwise} \end{cases}$$



Example: Squared Adjacency Matrices

$$(A^2)_{ij} = A_{i1}A_{1j} + A_{i2}A_{2j} + \ldots + A_{in}A_{nj}$$

$$A_{ik}A_{kj} = \begin{cases} 1 & \texttt{there are edges i to k and k to j} \\ 0 & \texttt{otherwise} \end{cases}$$



Example: Squared Adjacency Matrices

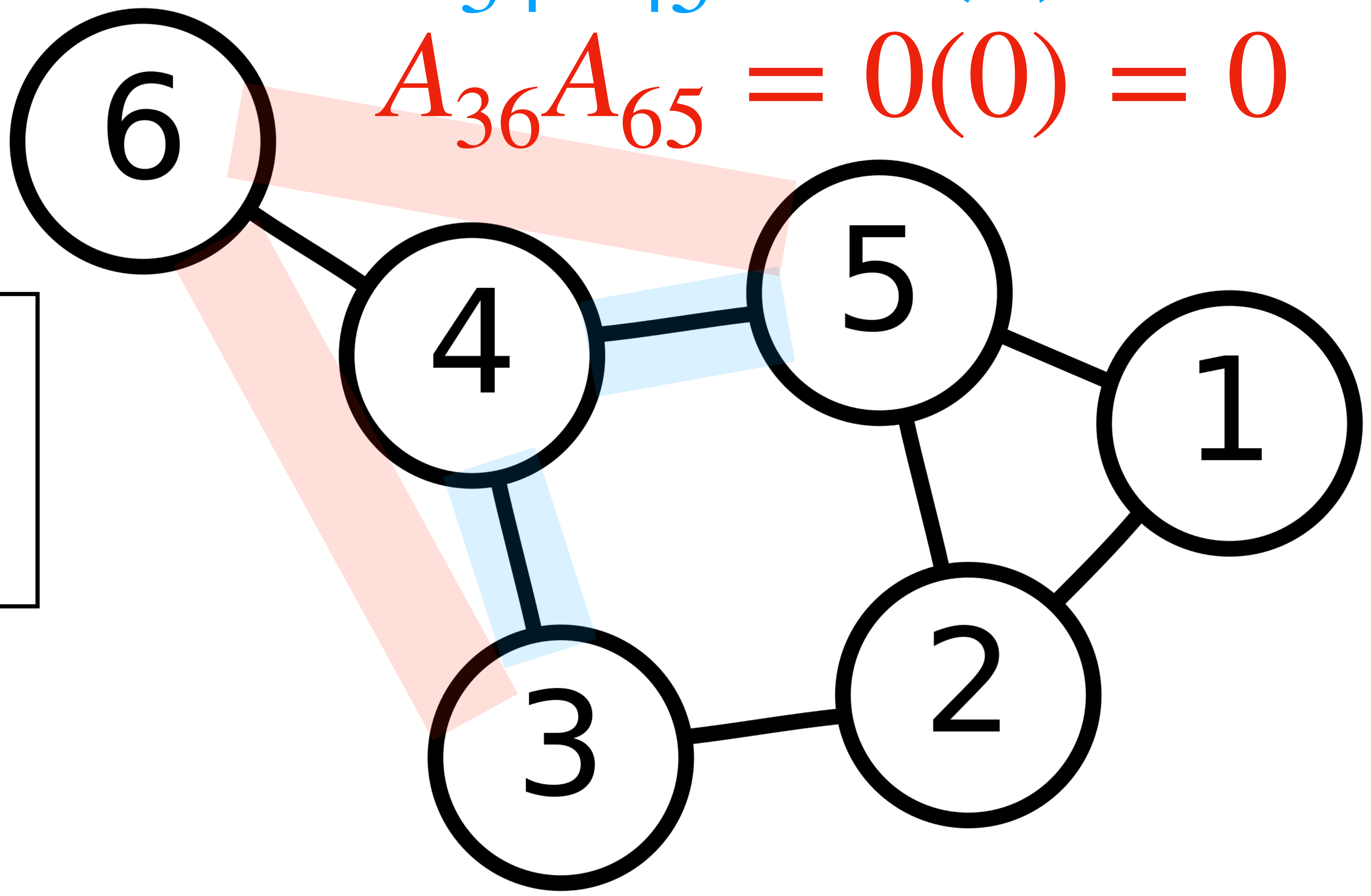
$$(A^2)_{ij} = A_{i1}A_{1j} + A_{i2}A_{2j} + \ldots + A_{in}A_{nj}$$

$$A_{ik}A_{kj} = \begin{cases} 1 & \texttt{there are edges i to k and k to j} \\ 0 & \texttt{otherwise} \end{cases}$$

$$(A^2)_{ij} = \boxed{\textbf{number of 2-step paths from i to j}}$$

$A_{34}A_{45} = 1(1) = 1$

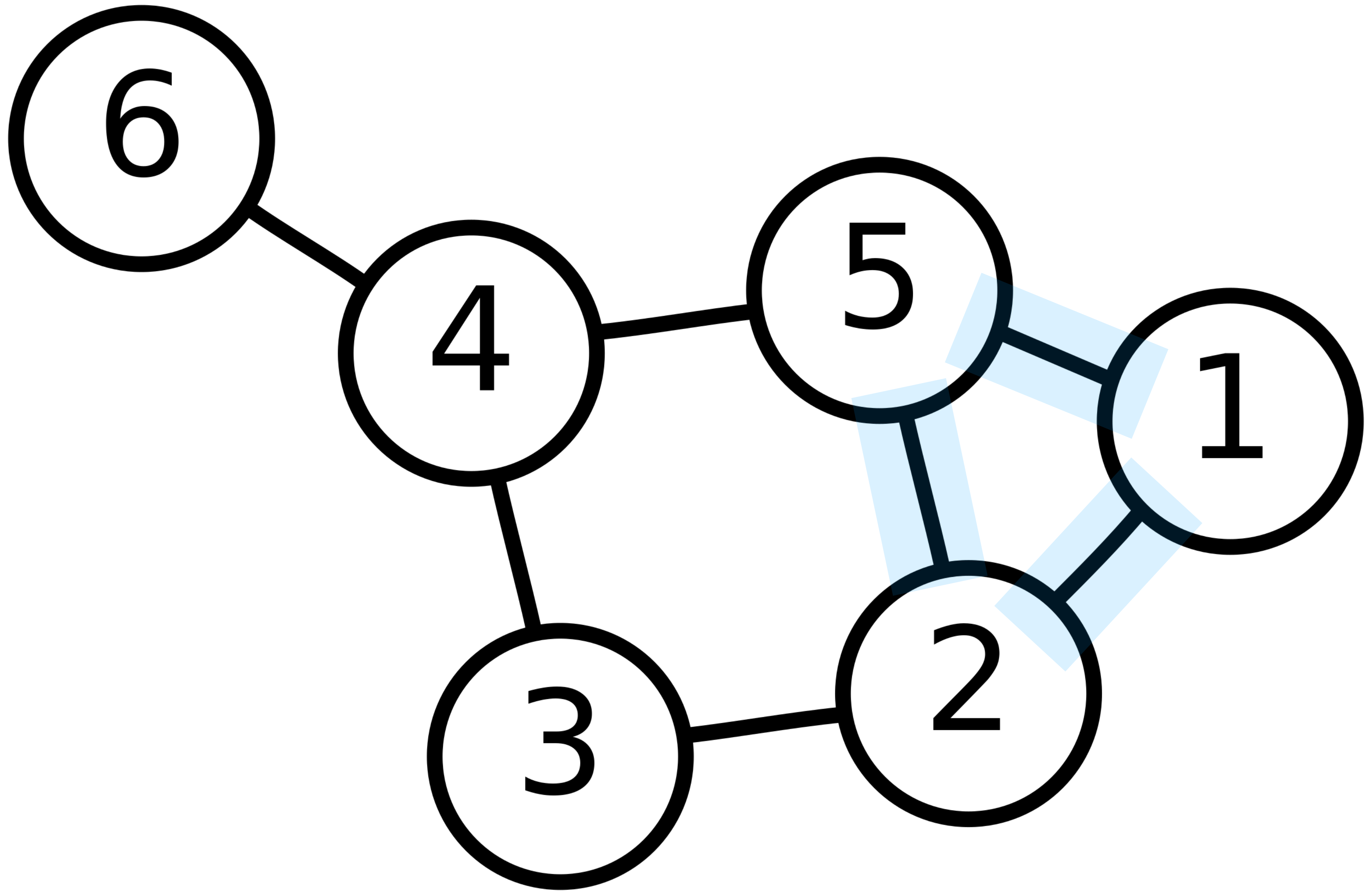
$A_{36}A_{65} = 0(0) = 0$



Application: Triangle Counting

A **triangle** in an undirected graph is a set of three distinct nodes with edges between every pair of nodes

Triangles in a social network represent mutual friends and tight cohesion (among other things)



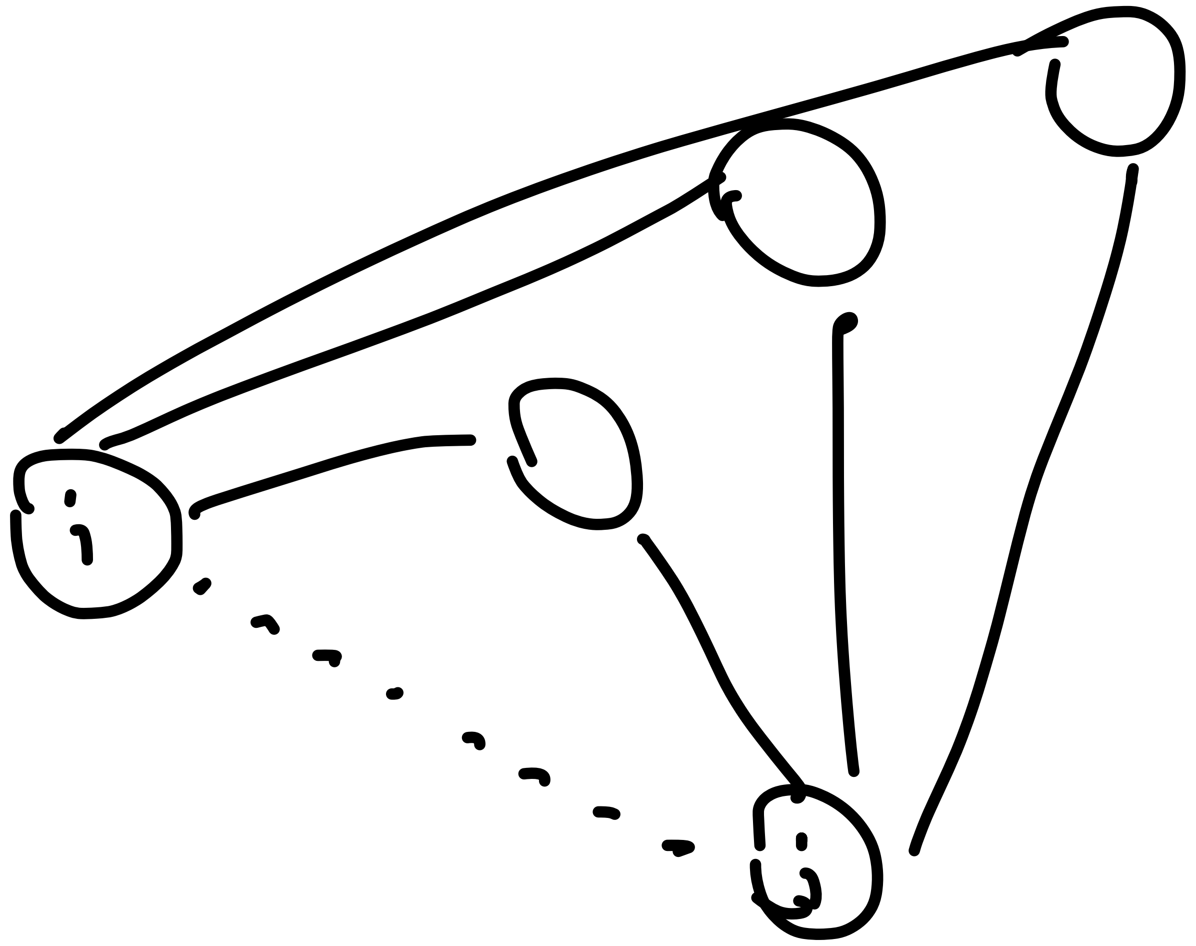
Application: Triangle Counting (Naive)

```
FUNCTION tri_count_naive($A$):
  count = 0
  for i from 1 to n:
    for j from i + 1 to n:
      for k from j + 1 to n:
        if $A_{ij} = 1$ and $A_{jk} = 1$ and $A_{ki} = 1$: # an edge between each pair
          count += 1:
  RETURN count
```

Application: Triangle Counting

Theorem. For an adjacency matrix A, the number of triangle containing the edge (i,j) is

$$(A^2)_{ij} * A_{ij}$$



Verify:

Application: Triangle Counting

```
FUNCTION tri_count(A):
    compute A^2
    count ← sum of (A^2)_ij * A_ij for all distinct i and j
    RETURN count / 6        # why divided by 6?
```

Application: Triangle Counting

```
FUNCTION tri_count(A):
  # in NumPy '*' is entry-wise multiplication
  #      also called the HADAMARD PRODUCT
  count ← sum of the entries of A^2 * A
  RETURN count / 6
```

Application: Triangle Counting

```
FUNCTION tri_count(A):
  # in NumPy '*' is entry-wise multiplication
  #      also called the HADAMARD PRODUCT
  # and 'np.sum' sums the entry of a matrix
  RETURN np.sum((A @ A) * A) / 6
```

demo

Dynamical Systems

Change

Change

```
Things change
```

Change

Things change

Things change from one state of affairs to another state of affairs

Change

Things change

Things change from one state of affairs to another state of affairs

Things change often in unpredictable ways

Change

Things change

Things change from one state of affairs to another state of affairs

Things change often in unpredictable ways

If something changes unpredictably, what can we say about it?

Dynamical Systems

Dynamical Systems

Definition (Informal). A **dynamical system** is a thing (typically with interacting parts) that changes over time

Dynamical Systems

Definition (Informal). A **dynamical system** is a thing (typically with interacting parts) that changes over time

A dynamical system has *possible states* which it can be in as time elapses and its behavior is defined by a *evolution function*

Dynamical Systems

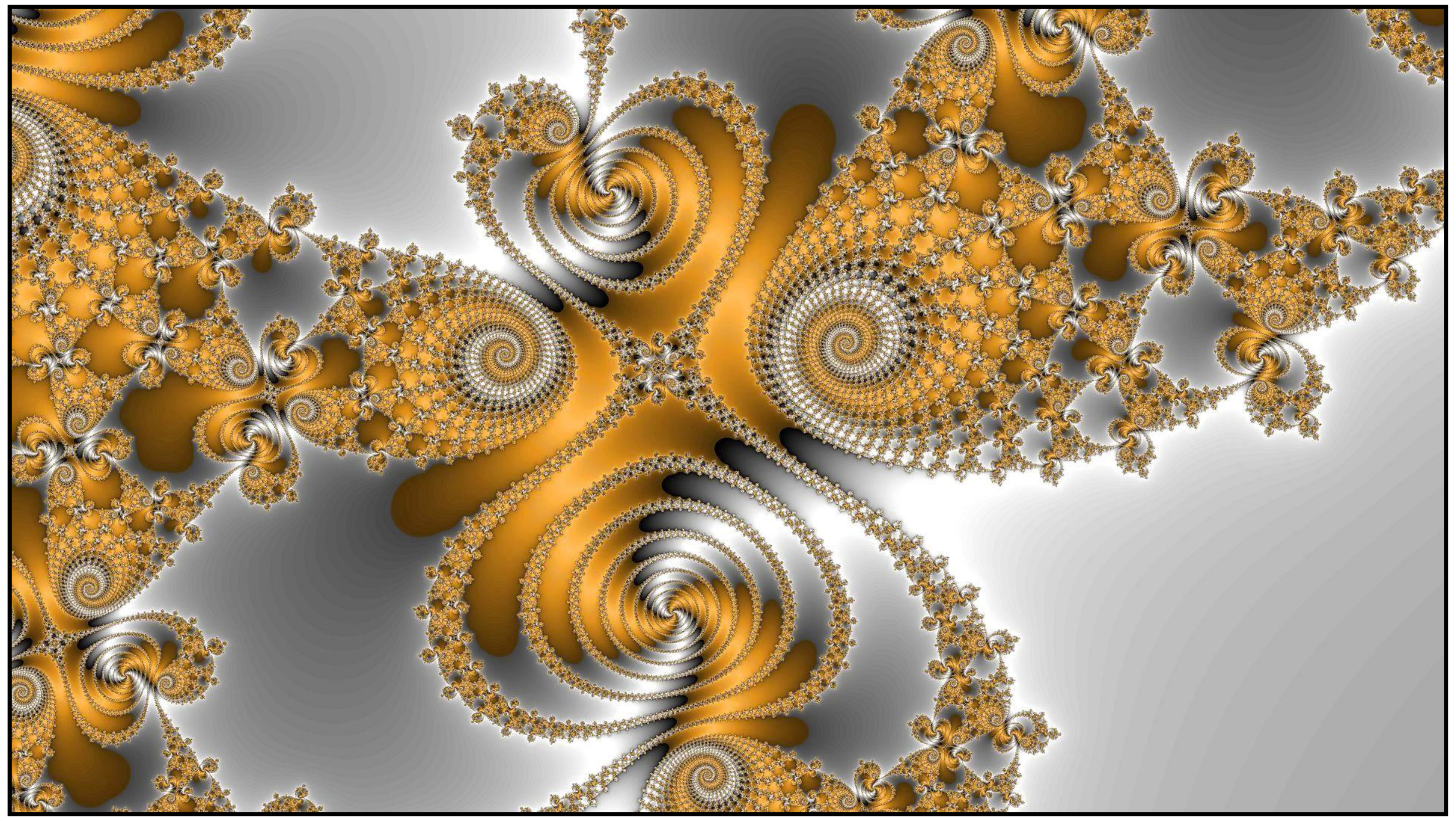
Definition (Informal). A **dynamical system** is a thing (typically with interacting parts) that changes over time

A dynamical system has *possible states* which it can be in as time elapses and its behavior is defined by a *evolution function*

Examples.

» economics (stocks)
» physical/chemical systems
» populations
» weather

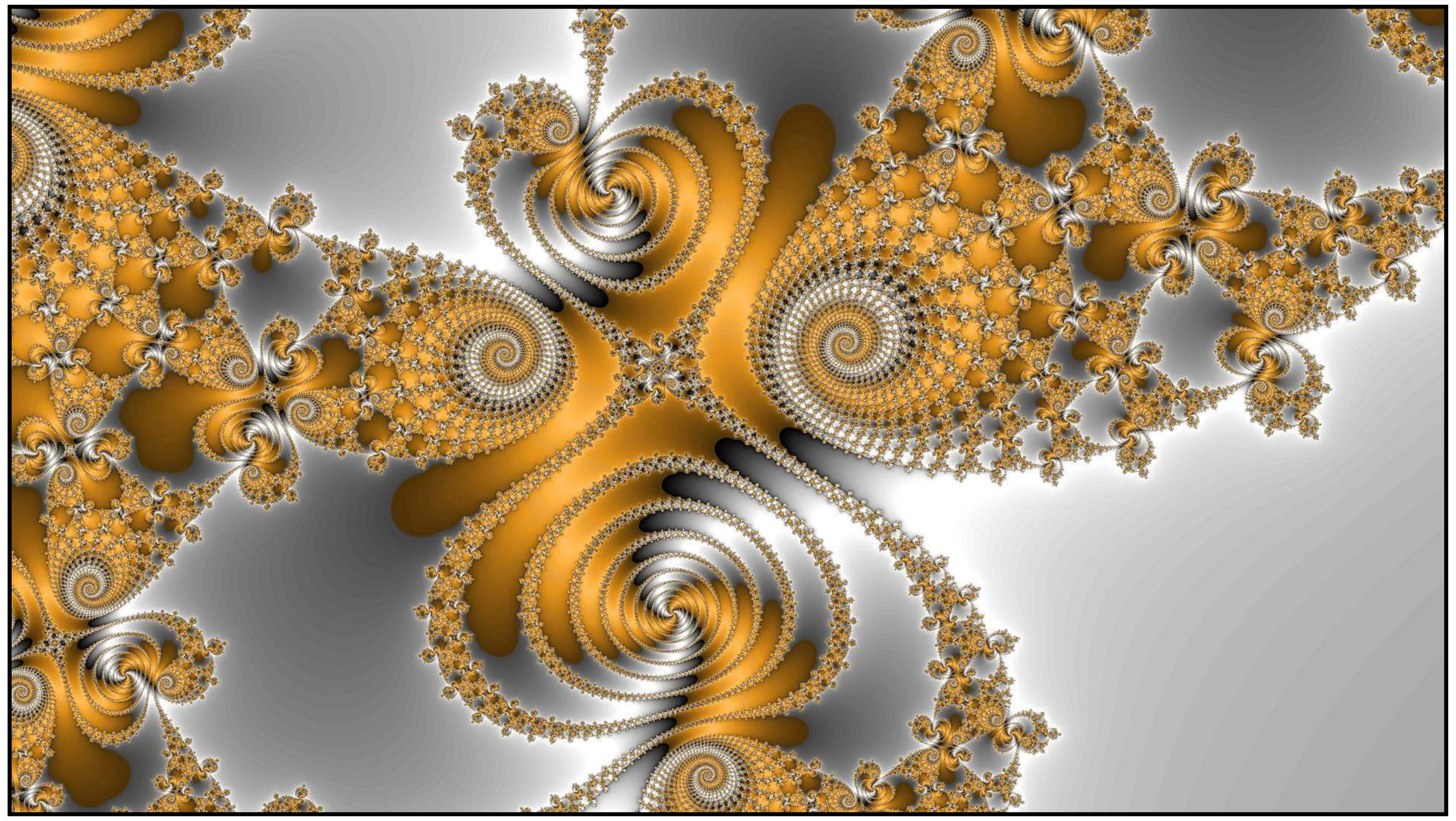
An Aside: Chaos Theory



https://commons.wikimedia.org/wiki/File:Fr137.jpg

An Aside: Chaos Theory

Complex systems like the weather
or the economy look nearly random

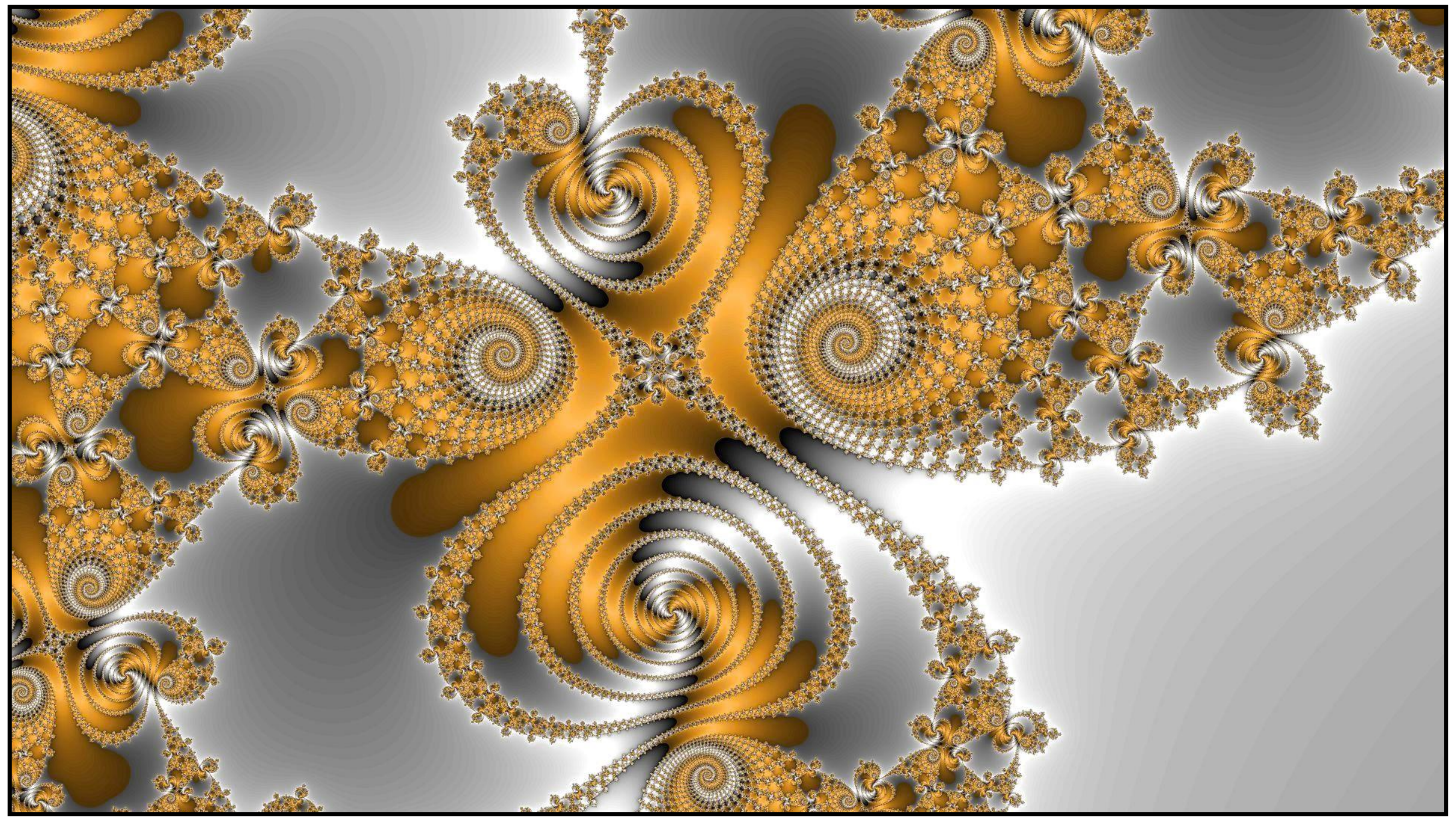


https://commons.wikimedia.org/wiki/File:Fr137.jpg

An Aside: Chaos Theory

Complex systems like the weather or the economy look nearly random

But even in chaotic systems there are *underlying patterns* and *repeated structures*



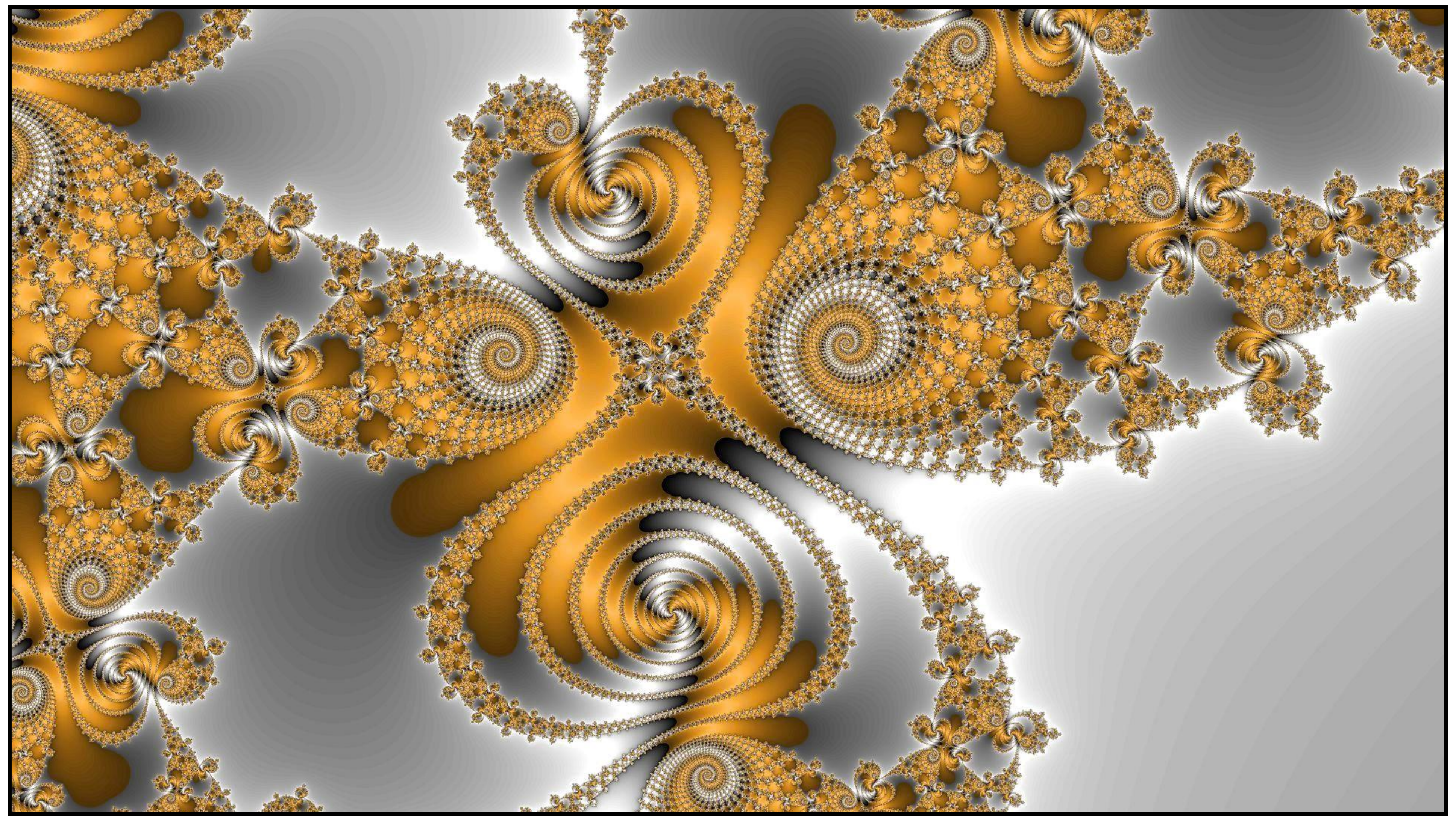
https://commons.wikimedia.org/wiki/File:Fr137.jpg

An Aside: Chaos Theory

Complex systems like the weather or the economy look nearly random

But even in chaotic systems there are *underlying patterns* and *repeated structures*

Often it's useful to consider chaotic systems in terms of <u>global</u> properties



https://commons.wikimedia.org/wiki/File:Fr137.jpg

Motivating Questions

Motivating Questions

What does a dynamical system look like "in the long view?"

Motivating Questions

What does a dynamical system look like "in the long view?"

Does it reach a kind of equilibrium? (think heat diffusion)

Motivating Questions

What does a dynamical system look like "in the long view?"

Does it reach a kind of equilibrium? (think heat diffusion)

Or does some part of the system dominate over time? (think the population of rabbits without a predator)

Linear Dynamical Systems

Linear Dynamical Systems

Linear Dynamical Systems

Definition. A **(discrete time) linear dynamical system** is a described a $n \times n$ matrix A. It's **evolution function** is the matrix transformation $\mathbf{x} \mapsto A\mathbf{x}$

Linear Dynamical Systems

Definition. A **(discrete time) linear dynamical system** is a described a $n \times n$ matrix A. It's **evolution function** is the matrix transformation $\mathbf{x} \mapsto A\mathbf{x}$

The possible states of the system are vectors in $\mathbb{R}^n$

Linear Dynamical Systems

Definition. A **(discrete time) linear dynamical system** is a described a $n \times n$ matrix A. It's **evolution function** is the matrix transformation $\mathbf{x} \mapsto A\mathbf{x}$

The possible states of the system are vectors in $\mathbb{R}^n$

Given an **initial state vector** $\mathbf{v}_0$, we can determine the **state vector** of the system after i time steps:

$$\mathbf{v}_i = A\mathbf{v}_{i-1}$$

Linear Dynamical Systems

Definition. A **(discrete time) linear dynamical system** is a described a $n \times n$ matrix A. It's **evolution function** is the matrix transformation $\mathbf{x} \mapsto A\mathbf{x}$

T

A tells us how our system evolves over time

Given an **initial state vector** $\mathbf{v}_0$, we can determine the **state vector** of the system after i time steps:

$$\mathbf{v}_i = A\mathbf{v}_{i-1}$$

State Vectors

$\vec{v}_{i+1} = A\vec{v}_i$

$$
\begin{aligned}
\mathbf{v}_1 &= A\mathbf{v}_0 \\
\mathbf{v}_2 &= A\mathbf{v}_1 = A(A\mathbf{v}_0) \\
\mathbf{v}_3 &= A\mathbf{v}_2 = A(AA\mathbf{v}_0) \\
\mathbf{v}_4 &= A\mathbf{v}_3 = A(AAA\mathbf{v}_0) \\
\mathbf{v}_5 &= A\mathbf{v}_4 = A(AAAA\mathbf{v}_0) \\
&\vdots
\end{aligned}
$$

The state vector $\mathbf{v}_k$ tells us what the system looks like after a number k time steps

This is also called a *recurrence relation* or a *linear difference function*

How to: Determining State Vectors

Question. Determine the state vector $\mathbf{v}_i$ for the linear dynamical system with matrix A given the initial state vector $\mathbf{v}_0$

Solution. Compute

$$\mathbf{v}_i = A^i \mathbf{v}_0$$

Warm up:
Population Dynamics

The Setup

The Setup

We're working for the census. We have population measurements for a <u>city</u> and a <u>suburb</u> which are geographically coincident

The Setup

We're working for the census. We have population measurements for a city and a suburb which are geographically coincident

We find by analyzing previous data that each year:

» 5% of the population moves from city → suburb

» 3% of the population moves from suburb → city

Fundamental Question

Can we make any predictions about the population of the city and suburb in 20 years?

Assumptions: No immigration, emigration, birth, death, etc. **The overall population stays fixed.**

Setting up Linear Equations

Setting up Linear Equations

If city_0 = city pop. = 600,000

and suburb_0 = suburb pop. = 400,000

Setting up Linear Equations

If city_0 = city pop. = 600,000
and suburb_0 = suburb pop. = 400,000

then the populations next year are given by:

$$\text{city}_1 = (0.95)\text{city}_0 + (0.03)\text{suburb}_0$$
$$\text{suburb}_1 = (0.05)\text{city}_0 + (0.97)\text{suburb}_0$$

Setting up Linear Equations

If city_0 = city pop. = 600,000
and suburb_0 = suburb pop. = 400,000

then the populations next year are given by:

$$\text{city}_1 = (0.95)\text{city}_0 + (0.03)\text{suburb}_0$$
$$\text{suburb}_1 = (0.05)\text{city}_0 + (0.97)\text{suburb}_0$$

people who stayed

people who left

Setting up a Matrix

$$\begin{bmatrix} \text{city}_1 \\ \text{suburb}_1 \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} \text{city}_0 \\ \text{suburb}_0 \end{bmatrix} = \begin{bmatrix} 582{,}000 \\ 418{,}000 \end{bmatrix}$$

We expect the population of the city to decrease in a year

Setting up a Matrix

$$\begin{bmatrix} \text{city}_2 \\ \text{suburb}_2 \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} \text{city}_1 \\ \text{suburb}_1 \end{bmatrix} = \begin{bmatrix} 565{,}440 \\ 434{,}560 \end{bmatrix}$$

The next year, we expect the population of the city to *continue* to decrease

Will it decrease indefinitely?

Setting up a Matrix

$$\begin{bmatrix} \text{city}_k \\ \text{suburb}_k \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} \text{city}_{k-1} \\ \text{suburb}_{k-1} \end{bmatrix}$$

This is a *linear dynamical system*

So we want to guess what the population will look like in 20 years, we need to compute

$$\begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix}^{20} \begin{bmatrix} \text{city}_0 \\ \text{suburb}_0 \end{bmatrix}$$

demo

Markov Chains

Stochastic Matrices

$$\begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix}$$

What's special about this matrix?

- » Its entries are nonnegative
- » Its columns sum to 1

This should make us think probability

Stochastic Matrices

Definition. A $n \times n$ matrix is **stochastic** if its entries are nonnegative and its $\boxed{\text{columns}}$ sum to 1

Example.

$$\begin{bmatrix} 0.7 & 0.1 & 0.3 \\ 0.2 & 0.8 & 0.3 \\ 0.1 & 0.1 & 0.4 \end{bmatrix}$$

Markov Chains

Definition. A **Markov chain** is a linear dynamical system whose evolution function is given by a <u>stochastic</u> matrix

(We can construct a "chain" of state vectors, where each state vector only depends on the one before it)

Key Property of Stochastic Matrices

Key Property of Stochastic Matrices

Stochastic matrices redistribute the "stuff" in a vector.

Key Property of Stochastic Matrices

Stochastic matrices redistribute the "stuff" in a vector.

Theorem. For a stochastic matrix A and a vector $\mathbf{v}$,

$$\text{sum of entries of } \mathbf{v} \;\; = \;\; \text{sum of entries of } A\mathbf{v}$$

Key Property of Stochastic Matrices

The sum of the entries of $\mathbf{v}$ can be computed as

$$\text{np.sum}(v) = \mathbf{1}^T\mathbf{v} = \langle \mathbf{1}, \mathbf{v} \rangle \qquad \underline{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

So the previous statement can be written

$$\mathbf{1}^T(A\mathbf{v}) = \mathbf{1}^T\mathbf{v}$$

Key Property of Stochastic Matrices

$$\mathbf{1}^T(A\mathbf{v}) = \mathbf{1}^T\mathbf{v}$$

A is stochastic

Let's verify this:

More General Solutions

More General Solutions

In our example, we analyzed the dynamics of a *particular* population

More General Solutions

In our example, we analyzed the dynamics of a *particular* population

What if we're interested more generally in the behavior of the process for *any* population?

More General Solutions

In our example, we analyzed the dynamics of a *particular* population

What if we're interested more generally in the behavior of the process for *any* population?

We need to shift from a population vector to a population ***distribution*** vector

Returning to the Example

$$\begin{bmatrix} \text{city}_k \\ \text{suburb}_k \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} \text{city}_{k-1} \\ \text{suburb}_{k-1} \end{bmatrix}$$

Returning to the Example

$$\begin{bmatrix} \text{city}_k \\ \text{suburb}_k \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix}^k \begin{bmatrix} \text{city}_0 \\ \text{suburb}_0 \end{bmatrix}$$

Returning to the Example

$$\begin{bmatrix} \text{city}_k \\ \text{suburb}_k \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix}^k \begin{bmatrix} 600{,}000 \\ 400{,}000 \end{bmatrix}$$

Returning to the Example

$$\begin{bmatrix} \text{city}_k \\ \text{suburb}_k \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix}^k \begin{bmatrix} 600{,}000 \\ 400{,}000 \end{bmatrix}$$

But what if we start of with a different population?

Returning to the Example

$$\begin{bmatrix} \text{city}_k \\ \text{suburb}_k \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix}^k \begin{bmatrix} 600{,}000 \\ 400{,}000 \end{bmatrix}$$

But what if we start of with a different population?

Do we have to do all our work over again?

Returning to the Example

$$\begin{bmatrix} \text{city}_k \\ \text{suburb}_k \end{bmatrix} = \begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix}^k \boxed{\begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix}}$$

60% of pop. in city

40% of pop. in suburb

Not really

But rather than thinking in terms of populations, we need to think about **how the population is distributed**

Probability Vectors

Probability Vectors

Definition. A **probability vector** is a vector of nonnegative values that sum to 1

Probability Vectors

Definition. A **probability vector** is a vector of nonnegative values that sum to 1

They represent

» discrete probability distributions

» distributions of collections of things

Probability Vectors

Definition. A **probability vector** is a vector of nonnegative values that sum to 1

They represent

» discrete probability distributions

» distributions of collections of things

These are really the same thing

Probability Vectors (Example)

The vector $\begin{bmatrix} 1/3 \\ 1/6 \\ 1/2 \end{bmatrix}$ represents the distribution where we choose:

$$\frac{4}{12} + \frac{2}{12} + \frac{6}{12} = 1$$

1 with probability $1/3$

2 with probability $1/6$

3 with probability $1/2$

Probability Vectors (Example)

The vector $\begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix}$ represented the distribution of the population, but we can also think of this as:

If we choose a random person from the population we'll get someone:

in the city with probability 0.6

in the suburbs with probability 0.4

The point

The point

We'll be interested in the dynamics of Markov chains on <u>probability vectors</u>

The point

We'll be interested in the dynamics of Markov chains on <u>probability vectors</u>

Since stochastic matrices preserve $\mathbf{1}^T\mathbf{v}$, they *transform* one distribution into another

The point

np.sum($(\vec{v})$)

$1 \cdot v$

$(1, v)$

We'll be interested in the dynamics of Markov chains on probability vectors

"dot" product $1, v$

Since stochastic matrices preserve $\mathbf{1}^T\mathbf{v}$, they *transform* one distribution into another

Can we say something about how the distribution changes in the long run?

Steady-State Vectors

Steady-State Vectors

Definition. A **steady-state vector** for a stochastic matrix A is a probability vector $\mathbf{q}$ such that

$$A\mathbf{q} = \mathbf{q}$$

A steady-state vector is *not changed* by the stochastic matrix. They describe equilibrium distributions

Returning to the Example

Returning to the Example

How do we interpret a steady–state vector for our example?

Returning to the Example

How do we interpret a steady–state vector for our example?

The populations in the city and the suburb stay the same over time

Returning to the Example

How do we interpret a steady-state vector for our example?

The populations in the city and the suburb stay the same over time

The same number of people are moving into and out of the city each year

Fundamental Questions

Do steady states exist?

Are they unique?

How do we find them?

Finding Steady-State Vectors

$$A\mathbf{q} = \mathbf{q}$$

Let's solve this equation for $\mathbf{q}$:

$$A\vec{q} - \vec{q} = \vec{0}$$

$$(A - I)\vec{q} = \vec{0}$$

Finding Steady-State Vectors

$$A\mathbf{q} - \mathbf{q} = \mathbf{0}$$

Finding Steady-State Vectors

$$A\mathbf{q} - I\mathbf{q} = \mathbf{0}$$

Finding Steady-State Vectors

$$(A - I)\mathbf{q} = \mathbf{0}$$

Finding Steady-State Vectors

$$(A - I)\mathbf{q} = \mathbf{0}$$

This is a matrix equation so
we know how to solve it

How to: Steady-State Vectors

How to: Steady-State Vectors

Question. Determine if the Markov chain with stochastic matrix A has a steady-state vector. If it does, find such a vector

How to: Steady-State Vectors

Question. Determine if the Markov chain with stochastic matrix A has a steady-state vector. If it does, find such a vector

Solution. Solve the equation $(A - I)\mathbf{x} = \mathbf{0}$ and find a solution whose entries sum to 1 (this will be possible given a free variable)

How to: Steady-State Vectors

Question. Determine if the Markov chain with stochastic matrix A has a steady-state vector. If it does, find such a vector

Solution. Solve the equation $(A - I)\mathbf{x} = \mathbf{0}$ and find a solution whose entries sum to 1 (this will be possible given a free variable)

If there is no such solution, the system does not have a steady state

$$\begin{bmatrix} 0.95 & 0.3 \\ 0.05 & 0.97 \end{bmatrix}$$

Example

$$\begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} -0.05 & 0.03 \\ 0.05 & -0.03 \end{bmatrix} \sim \begin{bmatrix} -5 & 3 \\ 5^{\,-5} & -3^{\,+3} \end{bmatrix} \sim \begin{bmatrix} -5 & 3 \\ 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -3/5 \\ 0 & 0 \end{bmatrix}$$

$x_1 = 3/5\, x_2$

x_2 is free

$x_1 + x_2 = 1$

$3/5\, x_2 + x_2 = 1$

$8/5\, x_2 = 1$

$x_1 = 3/8$

$x_2 = 5/8$

$$\begin{bmatrix} 3/8 \\ 5/8 \end{bmatrix}$$

demo

Existence vs Convergence

If $(A - I)\mathbf{x} = \mathbf{0}$ infinitely many solutions, then it has a stable state

This **does not** mean:

» the stable state is unique
» the system converges to this state

Convergence

Convergence

Definition. For a Markov chain with stochastic matrix A, an initial state $\mathbf{v}_0$ **converges** to the state $\mathbf{v}$ if $\lim_{k\to\infty} A^k\mathbf{v}_0 = \mathbf{v}$

Convergence

Definition. For a Markov chain with stochastic matrix A, an initial state $\mathbf{v}_0$ **converges** to the state $\mathbf{v}$ if $\lim_{k\to\infty} A^k\mathbf{v}_0 = \mathbf{v}$

As we repeatedly multiply $\mathbf{v}_0$ by A, we get closer and closer to $\mathbf{v}$ (in the limit)

Convergence

Definition. For a Markov chain with stochastic matrix A, an initial state $\mathbf{v}_0$ **converges** to the state $\mathbf{v}$ if $\lim_{k \to \infty} A^k \mathbf{v}_0 = \mathbf{v}$

As we repeatedly multiply $\mathbf{v}_0$ by A, we get closer and closer to $\mathbf{v}$ (in the limit)

Example of Non-Convergence

Example of Non-Convergence

Non–Example. I is a stochastic matrix and

$$I\mathbf{v} = \mathbf{v}$$

for any choice of $\mathbf{v}$

Example of Non-Convergence

Non–Example. I is a stochastic matrix and

$$I\mathbf{v} = \mathbf{v}$$

for any choice of $\mathbf{v}$

So this system does not have a unique steady state

Example of Non-Convergence

Non–Example. I is a stochastic matrix and

$$I\mathbf{v} = \mathbf{v}$$

for any choice of $\mathbf{v}$

So this system does not have a unique steady state

And no vectors converge to the same stable state

Regular Stochastic Matrices

Regular Stochastic Matrices

Definition. A stochastic matrix A is **regular** if A^k has all positive entries for *some nonnegative* k

Regular Stochastic Matrices

Definition. A stochastic matrix A is **regular** if A^k has all positive entries for *some nonnegative* k

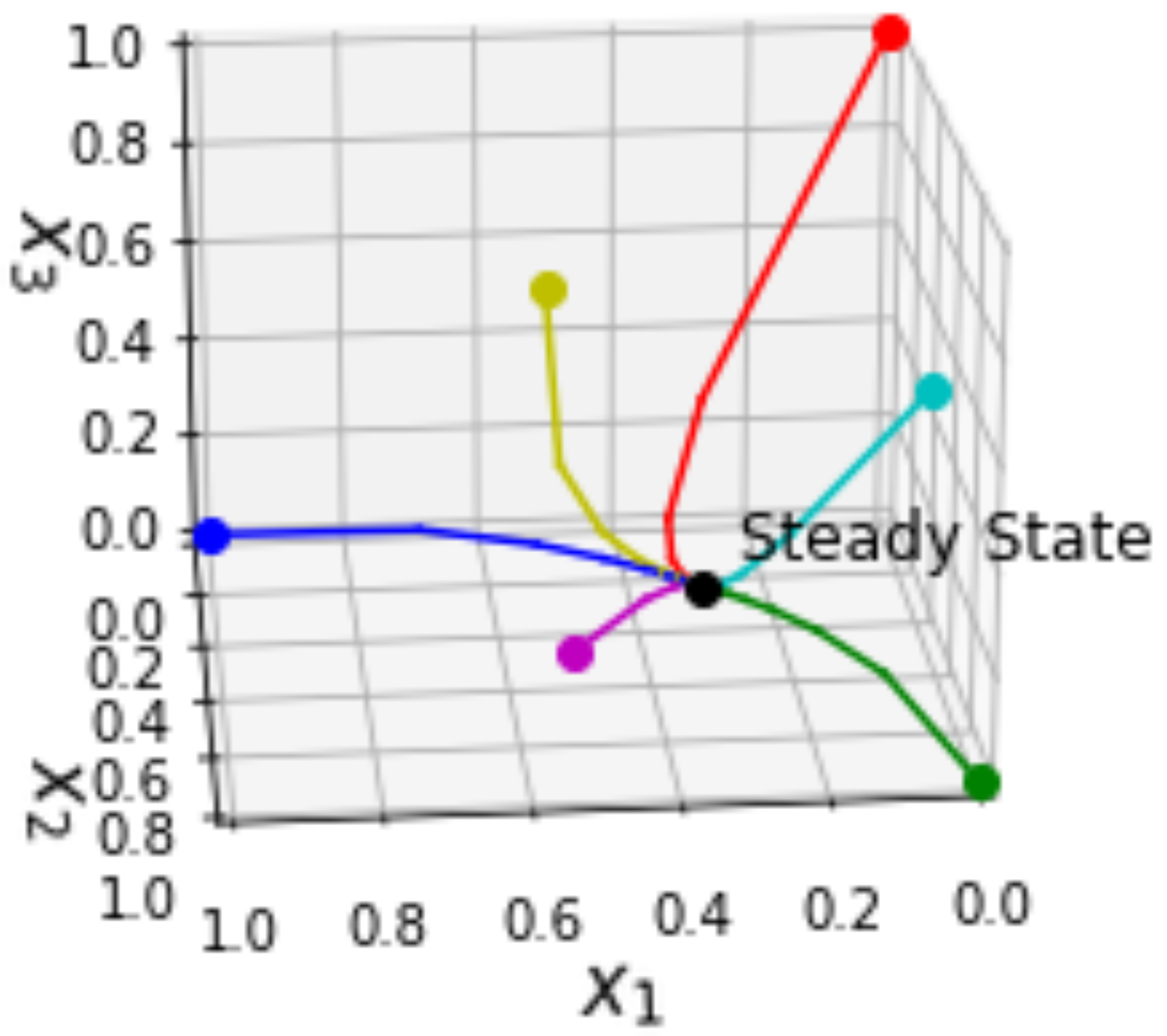
Theorem. A regular stochastic matrix P has a unique steady state, and

<u>every</u> probability vector converges to it

Mixing

This process of converging to a unique steady state is called "mixing"

This theorem says, after some amount of mixing, we'll be close to the stable state, **no matter where we started**



How to: Regular Stochastic Matrices

Question. Show that A is regular, and then find it's unique steady state

Solution. Find a power of A which has all positive entries, then solve the equation $(A - I)\mathbf{x} = \mathbf{0}$ as before

Example

$$\begin{bmatrix} 0.5 & 0.4 & 0 \\ 0.5 & 0.4 & 0.5 \\ 0 & 0.2 & 0.5 \end{bmatrix}$$

State Diagrams

Definition. A **state diagram** is a directed weighted graph whose adjacency matrix is stochastic.

Example.

Naming Convention Clash

The nodes of a state diagram are often called <u>states</u>

The vectors which are dynamically updated according to a linear dynamical system are called <u>state vectors</u>

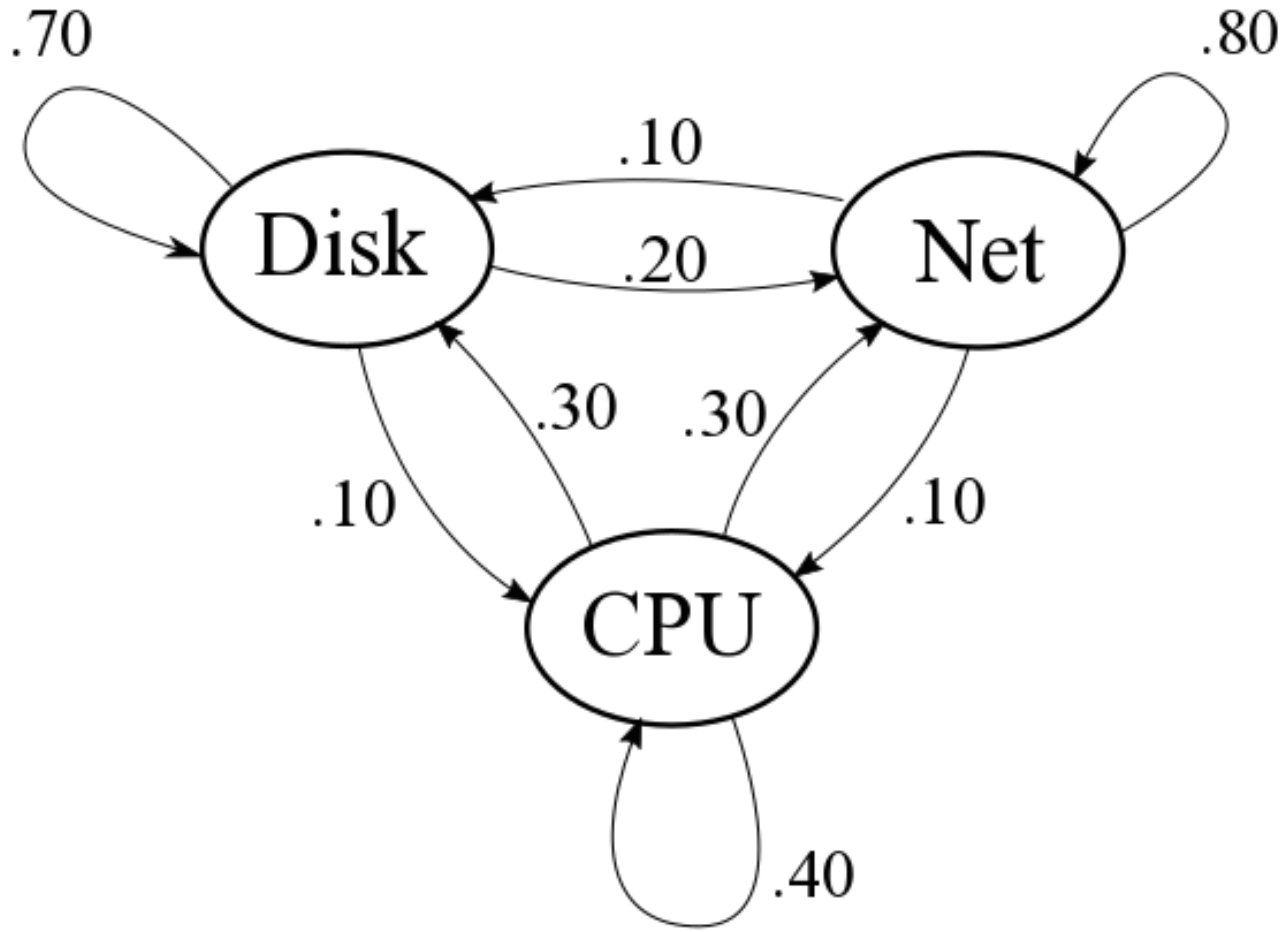
This is an unfortunate naming clash

Example: Computer System

Imagine a computer system in which tasks request service from disk, network or CPU

In the long term, which device is busiest?

This is about finding a stable state

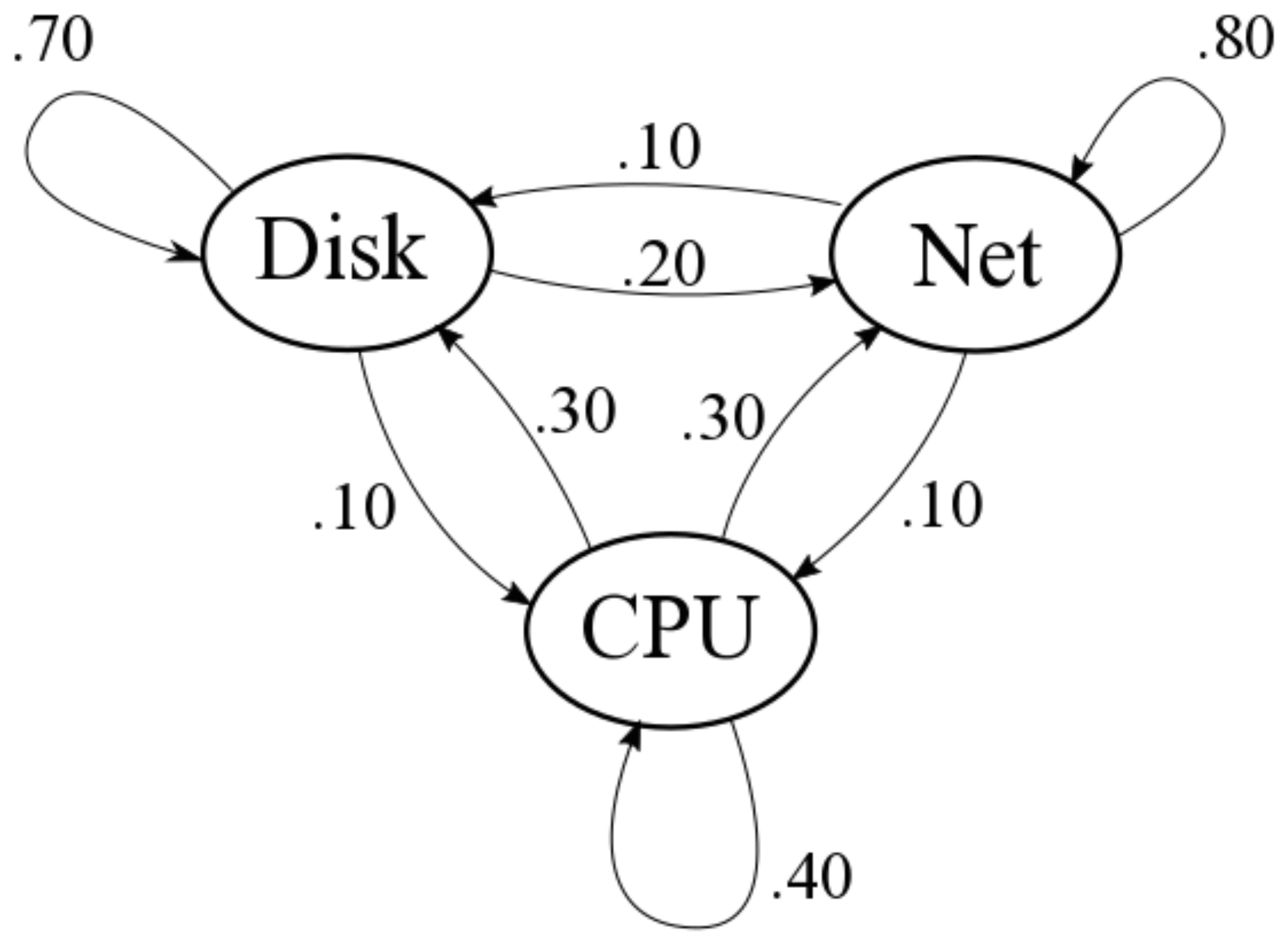


How To: State Diagram

Question. Given a state diagram, find the stable state for the corresponding linear dynamical system

Solution. Find the adjacency matrix for the state diagram and go from there

Example



Summary

Markov chains allow us to reason about dynamical systems that are dictated by some amount of randomness

Stable states represent global equilibrium