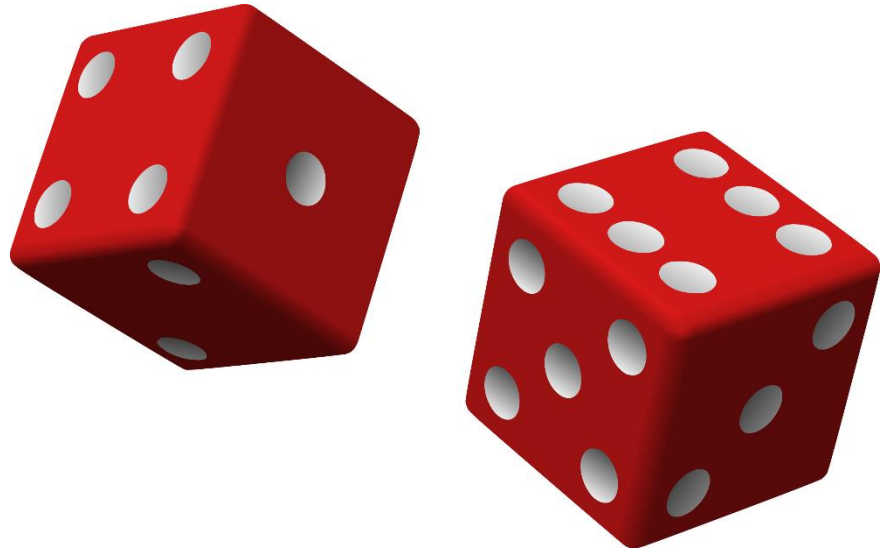
Modelling in Biology II: Stochastic processes and networks

Thomas Ouldridge
t.ouldridge@imperial.ac.uk
RSM 3.20

Course outline

- Lectures 1 & 2: What is a stochastic processes? Types of stochastic process. Analysis of discrete state stochastic processes.
- Lectures 3 & 4: Analysis of continuous state stochastic processes. Detailed balance in biological systems.
- Lectures 5 & 6: Out-of equilibrium biological systems.
- Lectures 7 & 8: Networks in biology.



Course outline

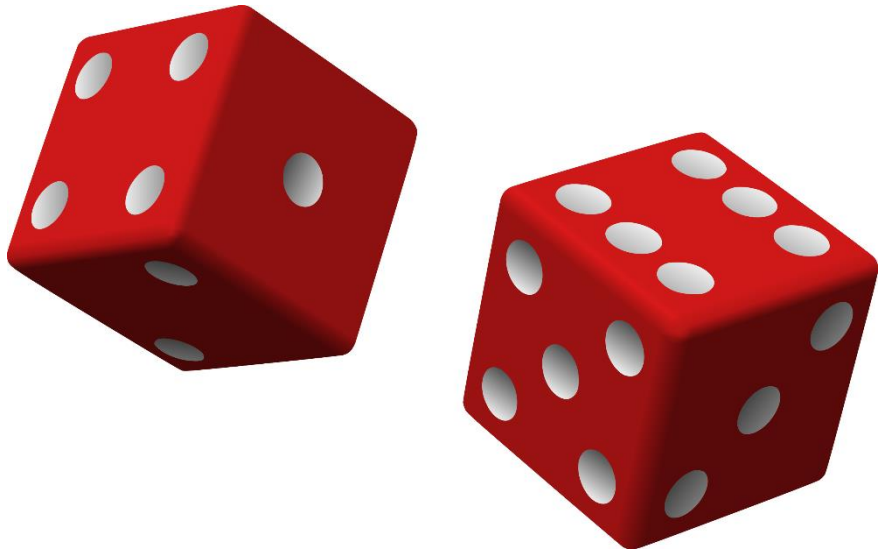
Online notes cover all the material expected (and some more).

Key Concept	Underlying ideas that you should be able to understand and explain.
Key Definition	Terminology that will get used repeatedly without being explained each time.
Key Technique	Mathematical approaches that you'll be expected to deploy under exam conditions.
Key Algorithm	Computational recipe that you need to be able to outline.
Application	Example of a technique/algorithm applied to understand a biological problem.

Lecture outline

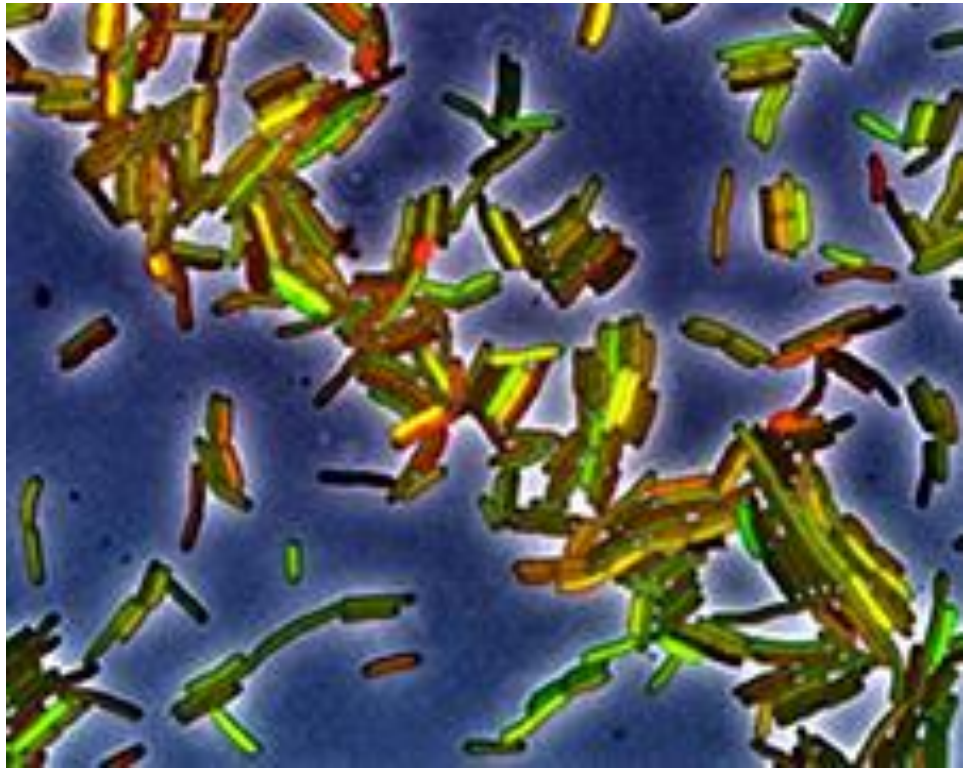
- What is a stochastic process?
- The Markov property
- Types of stochastic process

- Analysing a discrete-time, discrete-state Markov chain
- Continuous-time, discrete-state Markov processes



Randomness all around us

Behaviour in biological systems is often hard to predict.



What is a stochastic process?

Stochastic is just another word for random.

Describe the whole sequence of related random variables as a stochastic process $X(t)$.

Key Concept

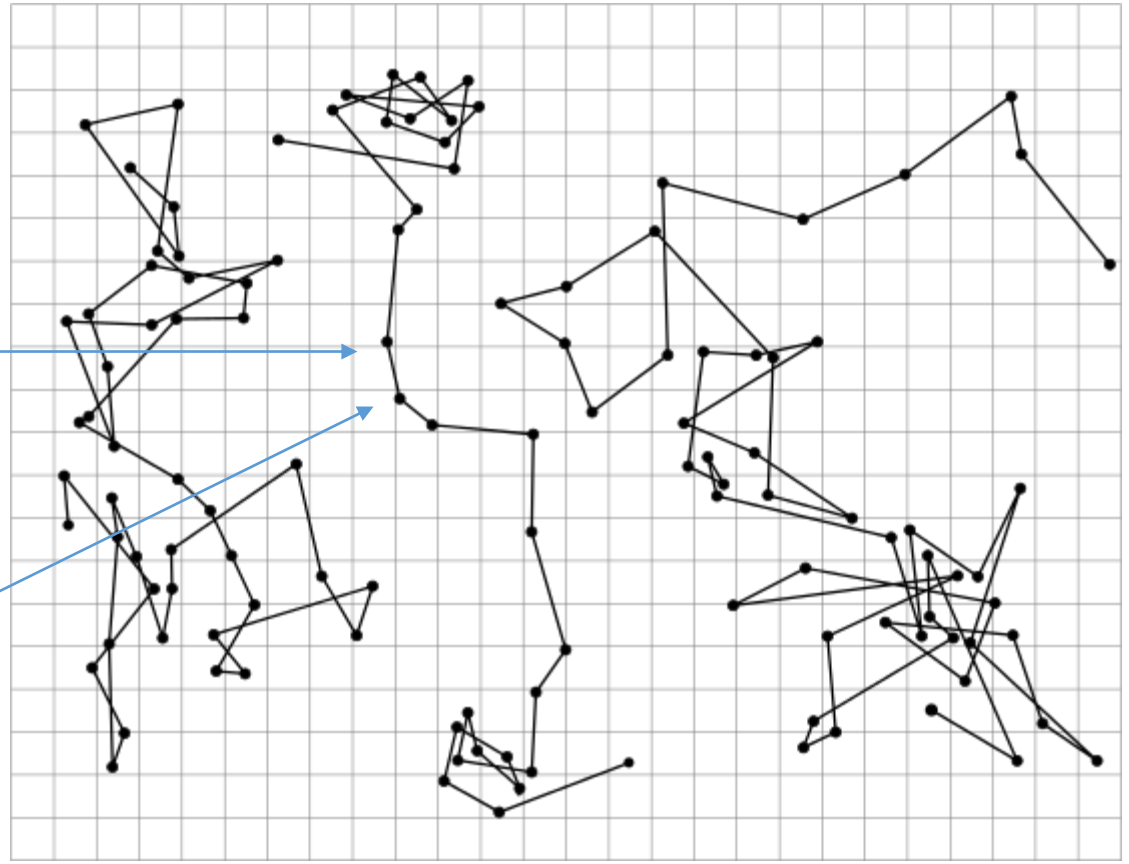
Central quantity is $p(x,t)$.

- Can we predict it?
- How does it evolve over time?
- Can we use it?

Position at time t_1 described by random variable X_1

Interrelated

Position at time t_2 described by random variable X_2



Jean Baptiste Perrin, *Les Atomes*, three tracings of the motion of colloidal particles of radius 0.53 µm.

What is a stochastic process?

We can talk about $X(t_1)$ and $X(t_2)$ just like any other pair of random variables.

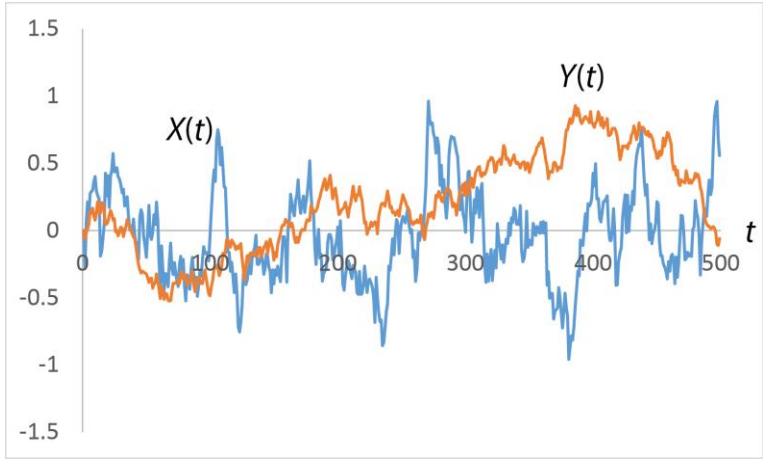
- Joint probability $p(x_1, t_1; x_2, t_2)$.
- Marginal probability $p(x_1, t_1) = \sum_{x_2} p(x_1, t_1; x_2, t_2)$.
- Conditional probability $p(x_1, t_1 | x_2, t_2)$.
- Autocorrelation is an important quantity:

$$R_X(t_1, t_2) = \frac{\mathrm{Cov}(X(t_1), X(t_2))}{\sqrt{\mathrm{Var}(X(t_1))\mathrm{Var}(X(t_2))}}$$

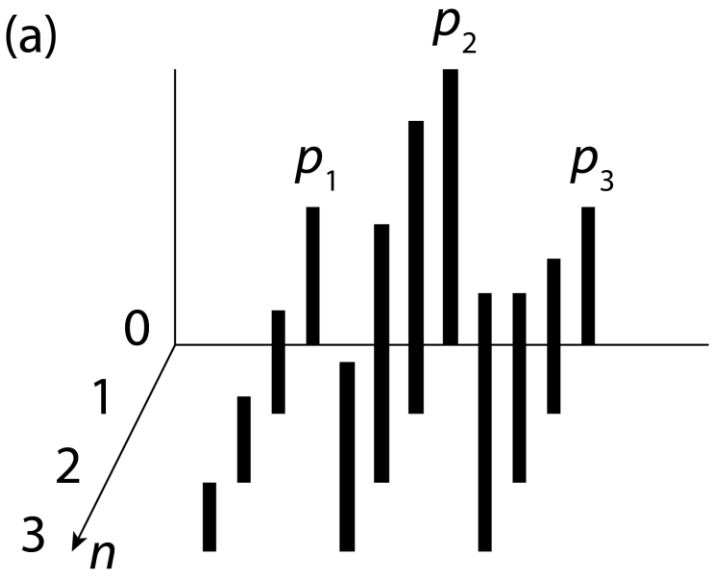
Key Concept

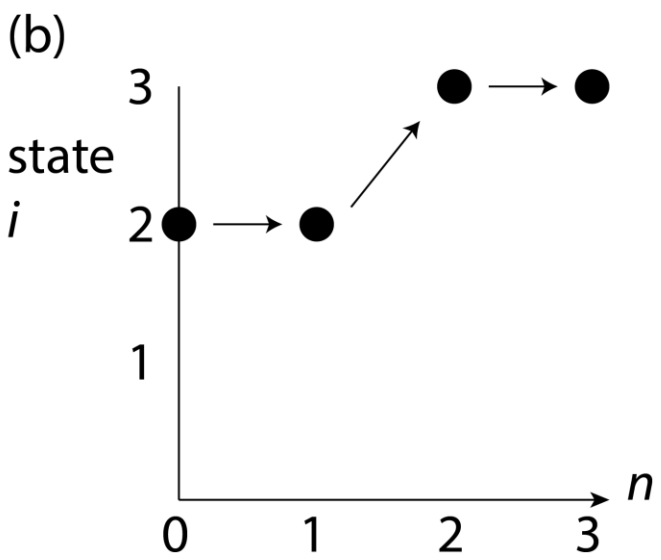
$p(x_1, t_1)$ has all the properties we expect of a distribution.

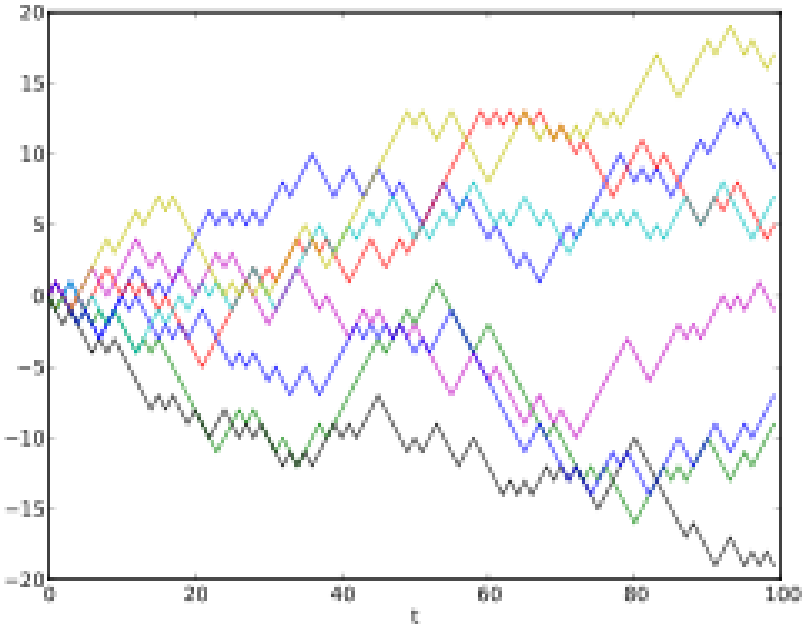
- Normalisation
- Mean
- Variance …



There are two fundamentally distinct ways to follow a stochastic process







Key Concept

Markov processes are the simplest stochastic processes

A Markov process has no memory, except through the value of the variable(s) at a given time.

$$p(x_1, t_1 | x_2, t_2; x_3, t_3; x_4, t_4 \ldots) = p(x_1, t_1 | x_2, t_2).$$

Key Concept

We're going to (largely) focus on Markov processes which are:

Type of process	Following the probability distribution	Following a sample trajectory
Discrete state, discrete time		
Discrete state, continuous time		
Continuous state, continuous time		

Discrete-time, discrete-state Markov chains

We have discrete points at which we observe the system.

Label with an index n: X_n and $p(x, n)$.

Even if the possible values of X_n are non-numeric, we can still order them into a numbered list or **vector**.

Thus $p(x,n) = \begin{pmatrix} p_1(n) \\ p_2(n) \\ \dots \end{pmatrix}$.

Useful tip:

- Columns of **T** sum to one.

The evolution of the system is entirely specified by the **transition matrix.**

$$\mathbf{T} = \begin{pmatrix} p(x_1, n+1|x_1, n) & p(x_1, n+1|x_2, n) & \dots & p(x_1, n+1|x_N, n) \\ p(x_2, n+1|x_1, n) & \dots & & \\ \dots & & & \\ p(x_N, n+1|x_1, n) & p(x_N, n+1|x_2, n) & \dots & p(x_N, n+1|x_N, n) \end{pmatrix}.$$

Key Definition

Evolution under the transition matrix

The probability distribution at step $n+1$ is simply the transition matrix applied to the probability distribution at step n.

$$p(x, n+1) = \mathbf{T}p(x, n)$$

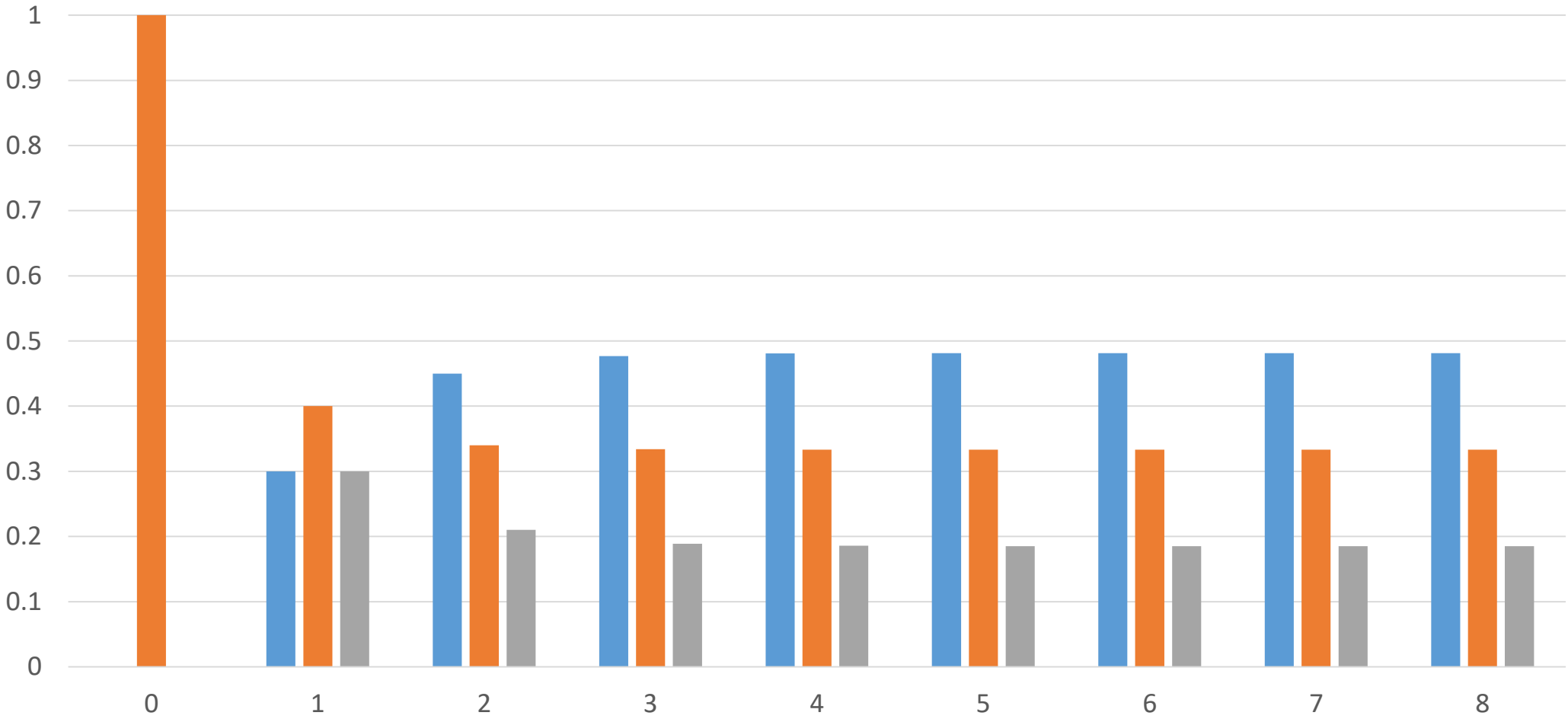
$$p_i(n+1) = \sum_j T_{ij} p_j(n)$$

Key Technique

$$p(x, n+m) = \mathbf{T}^m p(x, n)$$

$$p_i(n+m) = \sum_j T^m{}_{ij} p_j(n)$$

Evolution under the transition matrix



Stationary distributions

A stationary distribution $\pi(x)$ is such that

$$\pi(x) = \mathbf{T}\pi(x)$$

Key Concept

$$\pi_i = \sum_j T_{ij}\pi_j$$

Once in a stationary distribution, the probability distribution will not evolve. Individual trajectories will still move, though!

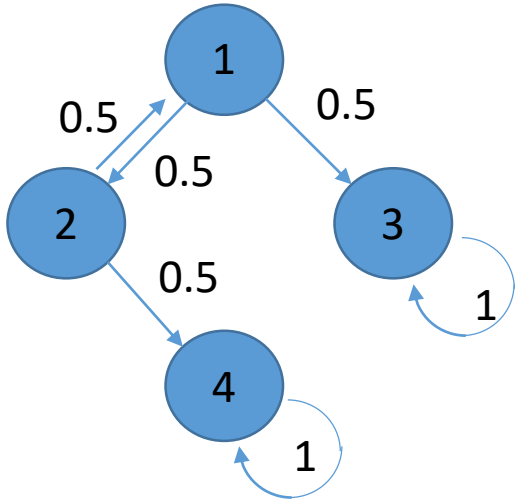
π_i is an eigenvector of T_{ij} with eigenvalue 1.

Key Technique

- At least one must exist to conserve probability.
- Other eigenvalues have a magnitude less than 1.
- Systems tend to "relax" to a stationary distribution over time.
- In general, relatively easy to check; a bit harder to find.

Half time conundrum

Which of these is a stationary distribution for the process below?



Go to www.menti.com and use code 41 99 69

(a) $$\begin{pmatrix} p(1) \\ p(2) \\ p(3) \\ p(4) \end{pmatrix} = \begin{pmatrix} 0.3 \\ 0.2 \\ 0.5 \\ 0 \end{pmatrix}$$

(b) $$\begin{pmatrix} p(1) \\ p(2) \\ p(3) \\ p(4) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0.7 \\ 0.3 \end{pmatrix}$$

(c) $$\begin{pmatrix} p(1) \\ p(2) \\ p(3) \\ p(4) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0.5 \end{pmatrix}$$

(d) $$\begin{pmatrix} p(1) \\ p(2) \\ p(3) \\ p(4) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

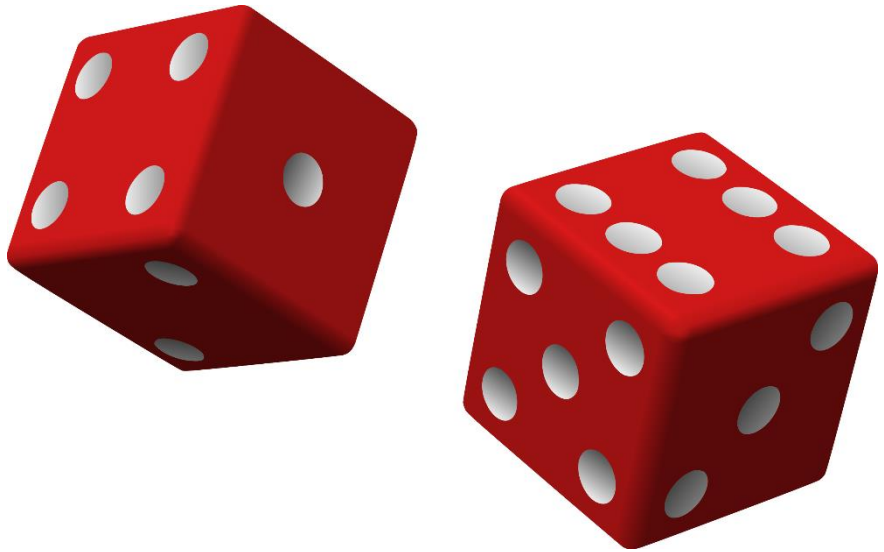
Modelling in Biology II: Stochastic processes and networks

Thomas Ouldridge
t.ouldridge@imperial.ac.uk
RSM 3.20

Lecture outline

- What is a stochastic process?
- The Markov property
- Types of stochastic process

- Analysing a discrete-time, discrete-state Markov chain
- Continuous-time, discrete-state Markov processes

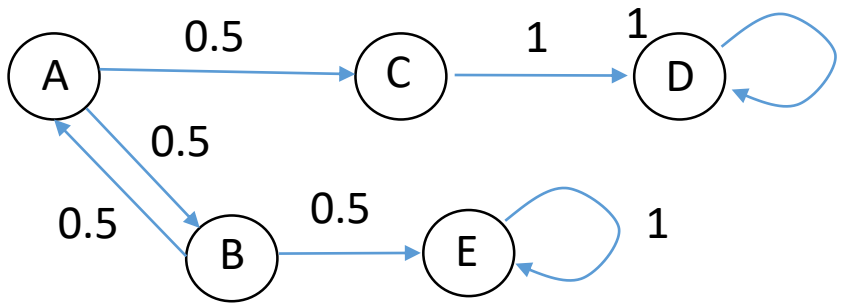


Absorption and first passage times

Some systems have absorbing states.

It is helpful to re-write such transition matrices as in the form:

$$\mathbf{T} = \begin{pmatrix} \mathbf{U} & 0 \\ \mathbf{R} & \mathbf{I} \end{pmatrix}$$



We can immediately see that

$$\mathbf{T}^m = \begin{pmatrix} \mathbf{U}^m & 0 \\ \mathbf{R} + \mathbf{RU} + \mathbf{RU}^2 + \ \ldots \mathbf{RU}^{m-1} & \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{U}^m & 0 \\ \mathbf{RW}(m-1) & \mathbf{I} \end{pmatrix}$$

With $\mathbf{W}(m) = \mathbf{I} + \ \mathbf{U} + \mathbf{U^2} + \ \ldots \mathbf{U}^m$.

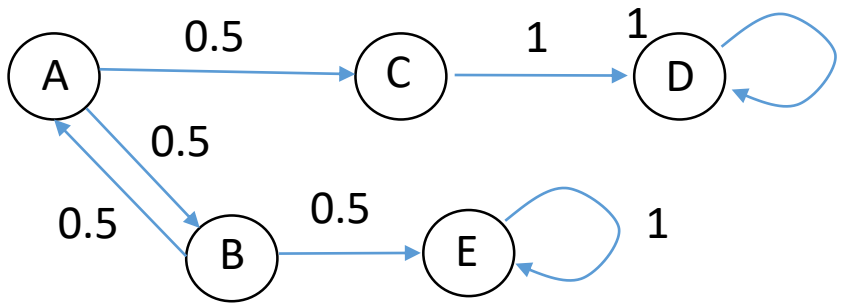
$\mathbf{W}(m)_{ij}$ is the expected number of times that non-absorbing state i is visited in the first m steps having started at j.

Absorption and first passage times

Some systems have absorbing states.

It is helpful to re-write such transition matrices as in the form:

$$\mathbf{T} = \begin{pmatrix} \mathbf{U} & 0 \\ \mathbf{R} & \mathbf{I} \end{pmatrix}$$



We can immediately see that

$$\mathbf{T}^m = \begin{pmatrix} \mathbf{U}^m & 0 \\ \mathbf{R} + \mathbf{RU} + \mathbf{RU}^2 + \ldots \mathbf{RU}^{m-1} & \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{U}^m & 0 \\ \mathbf{RW}(m-1) & \mathbf{I} \end{pmatrix}$$

With $\mathbf{W}(m) = \mathbf{I} + \mathbf{U} + \mathbf{U^2} + \ldots \mathbf{U}^m$. $\quad \mathbf{U}^m \to 0$ as $m \to \infty$

$\mathbf{W}(m)$ tends to a constant as m gets large: $\mathbf{W}(m) = \mathbf{W} = (\mathbf{I} - \mathbf{U})^{-1}$

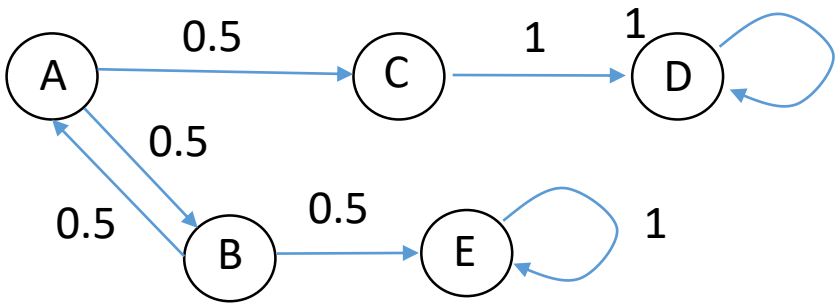
Key Definition

W is known as the **fundamental matrix**.

Absorption and first passage times

Armed with $\mathbf{W}$ we can calculate expected "time" (number of steps) to absorption for a system that starts in j:

$$\langle t_j^{\text{abs}} \rangle = \sum_i W_{ij}$$



We can also work out the probability of ending up in each potential absorbing state: by definition, $(\mathbf{RW})_{lj}$ is probability of ending in absorbing state l having started in state j.

$$\mathbf{T}^m = \begin{pmatrix} \mathbf{U}^m & 0 \\ \mathbf{R} + \mathbf{RU} + \mathbf{RU}^2 + \ \ldots \mathbf{RU}^{m-1} & \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{U}^m & 0 \\ \mathbf{RW}(m-1) & \mathbf{I} \end{pmatrix}$$

Key Technique

Useful tip:
We can find **first passage times** to non-absorbing states too; just adjust the transition matrix to make them absorbing!

Simulating a discrete-time Markov chain

Method 1. Evaluate the full probability vector as a function of n.

- Conceptually simple; just keep applying $\mathbf{T}$ to your probability vector (as we did before).
- No random numbers required!

Key Algorithm

Method 2. Follow a single sample trajectory.

- At each time step we chose a new state using the probabilities encoded in $\mathbf{T}$.
- Random number required! Must choose between a number of options, each with a different probability.
- Generating many trajectories allows you to calculate averages.

Why would you ever use method 2?

- Some questions make more sense at the trajectory level.
- For large (or infinite) systems, manipulating the whole matrix is a pain.

Key Algorithm

Review of where we stand

Type of process	Following the probability distribution	Following a sample trajectory
Discrete state, discrete time	Transition matrix equation	Single-step simulation
Discrete state, continuous time		
Continuous state, continuous time		

Discrete-state, continuous-time Markov processes

Not much has changed in terms of the state space. But now the dynamics is specified by transition rates K_{ij}.

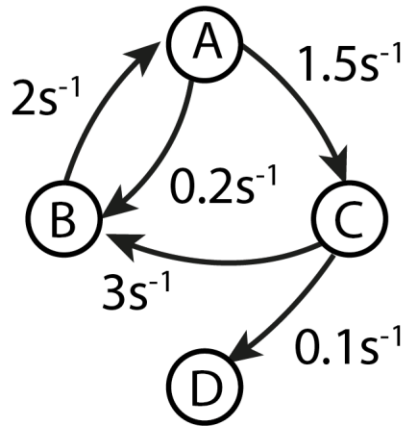
$$\frac{dp_i(t)}{dt} = \sum_{j \neq i} K_{ij} p_j(t) - K_{ji}\, p_i(t)$$

This is the **master equation**.

We can define a **rate matrix K.**

- K_{ii} defined by $K_{ii} = -\sum_{j \neq i} K_{ji}$.
- With this definition, $\frac{dp_i(t)}{dt} = \sum_j K_{ij} p_j(t)$

(a)



(b)

$$k \begin{pmatrix} -1.7 & 2 & 0 & 0 \\ 0.2 & -2 & 3 & 0 \\ 1.5 & 0 & -3.1 & 0 \\ 0 & 0 & 0.1 & 0 \end{pmatrix}$$

Key Definition

Discrete-state, continuous-time Markov processes

Not much has changed in terms of the state space. But now the dynamics is specified by transition rates K_{ij}.

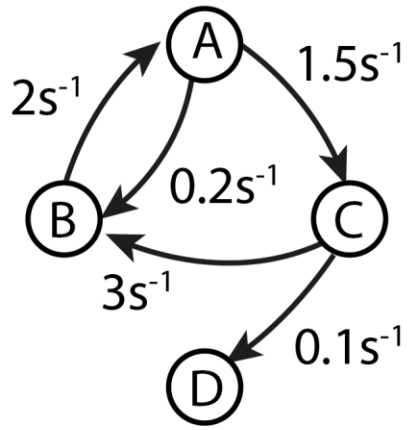
$$\frac{dp_i(t)}{dt} = \sum_{j \neq i} K_{ij} p_j(t) - K_{ji}\, p_i(t)$$

This is the **master equation**.

We can define a **rate matrix K.**

- K_{ii} defined by $K_{ii} = -\sum_{j \neq i} K_{ji}$.
- With this definition, $\frac{dp_i(t)}{dt} = \sum_j K_{ij} p_j(t)$

(a)



(b)

$$k \begin{pmatrix} -1.7 & 2 & 0 & 0 \\ 0.2 & -2 & 3 & 0 \\ 1.5 & 0 & -3.1 & 0 \\ 0 & 0 & 0.1 & 0 \end{pmatrix}$$

Note:

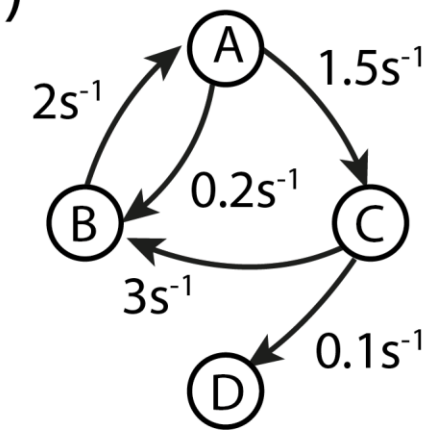
- Rate matrix has units of 1/time.
- Each column sums to zero.
- Not the same as the transition matrix **T.**

Evolution of the probability distribution under the rate matrix

$$\frac{dp_i(t)}{dt} = \sum_j K_{ij} p_j(t)$$

$$p_i(t) = \left(e^{t\mathbf{K}}\right)_{ij} p_j(0).$$

(a)



(b)

$$k \begin{pmatrix} -1.7 & 2 & 0 & 0 \\ 0.2 & -2 & 3 & 0 \\ 1.5 & 0 & -3.1 & 0 \\ 0 & 0 & 0.1 & 0 \end{pmatrix}$$

Stationary distribution is an eigenvector of the rate matrix with eigenvalue 0.

Key Technique

For short times,

$$p_i(\Delta t) \approx \Delta t K_{ij} p_j(0) + p_j(0).$$

$$p_i(t + \Delta t) = T_{ij} p_j(t)$$

Possible simulation method?

Review of where we stand

Type of process	Following the probability distribution	Following a sample trajectory
Discrete state, discrete time	Transition matrix equation	Step-by-step simulation
Discrete state, continuous time	Master equation	Discretised step-by-step?
Continuous state, continuous time		

Summary of lectures 1 & 2

- We can use random variables to capture the unpredictability of natural and human-constructed systems.
- A stochastic process is a series of related random variables, $X(t)$.
- Markov processes are simple and ubiquitous.
- We can deal with discrete-space, discrete-time processes, and have started to look at discrete-space, continuous time; we will go further next week.
- We can look at processes from the trajectory-level or at the level of the whole distribution.
- We can often make statements about the long-term behaviour.
- Simulation is an important tool for probing stochastic processes.

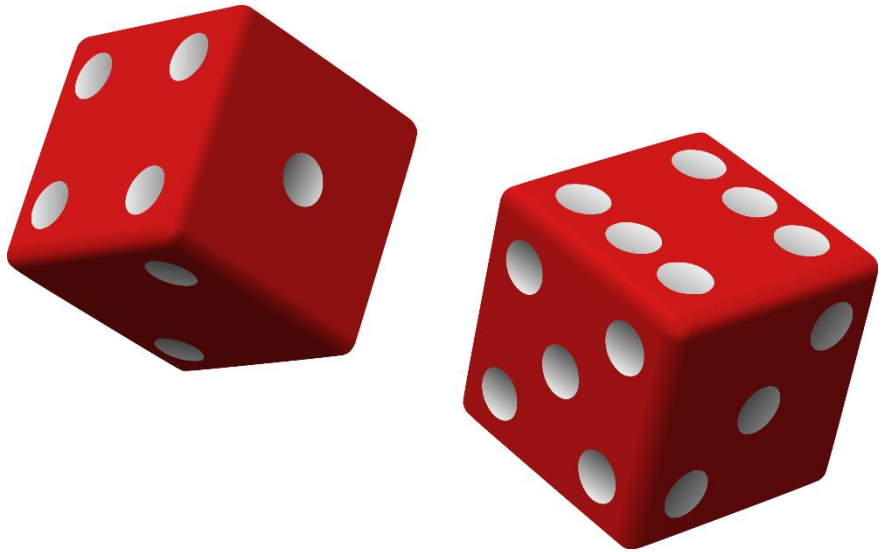
Modelling in Biology II: Stochastic processes and networks

Thomas Ouldridge
t.ouldridge@imperial.ac.uk

Lecture outline

- Finishing of continuous time, discrete state Markov processes
- Continuous time and continuous state Markov processes

- Different types of stationary distribution: in and out of detailed balance
- Examples of biological systems in equilibrium.



Last lecture

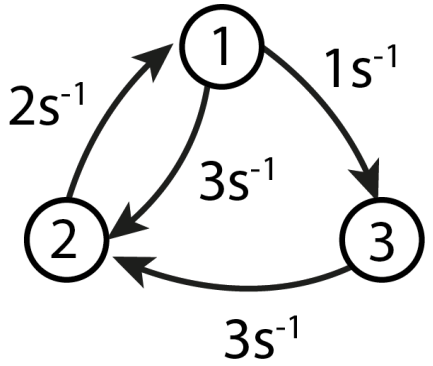
- A stochastic process is a set of random variables connected to each other (over time).

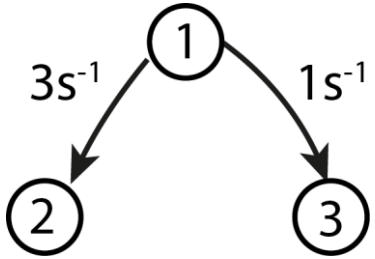
	Discrete state, discrete time	Discrete state, continuous time
Evolution	$p_i(n+1) = \sum_j T_{ij} p_j(n)$	$\frac{dp_i(t)}{dt} = \sum_j K_{ij} p_j(t)$
Stationary distribution	$\pi_i = \sum_j T_{ij}\pi_j$ – e/vector with e/value 1	$0 = \sum_j K_{ij}\pi_j$ – e/vector with e/value 0
Follow the probability distribution	Iterate the evolution equation	Discretise ODE into units of Δt; analyse approximate discrete time evolution
Follow a sample trajectory	Use random number to pick state based on probabilities encoded in $\mathbf{T}$	Discretise ODE into units of Δt; analyse approximate discrete time evolution?
First passage/time to absorption	Fundamental matrix methods	?

Discrete-state, continuous time processes from the trajectory perspective

Imagine we've just landed in state 1. What happens next?

Analyse the process of leaving state 1 only.





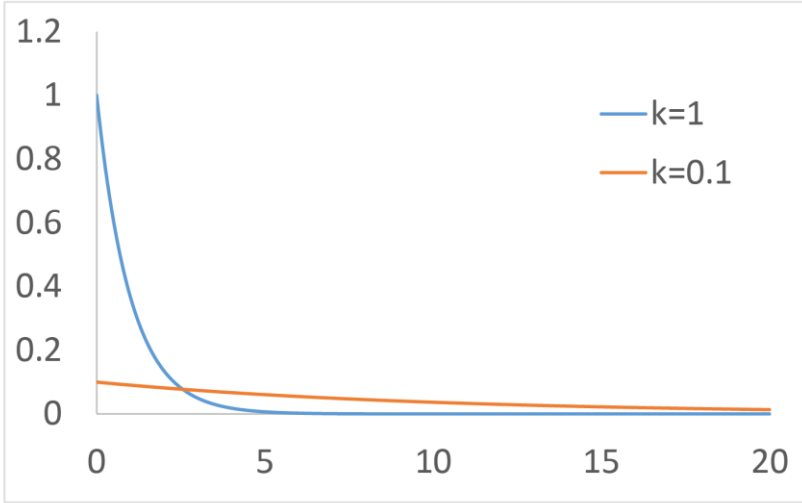
$$\frac{dp_1}{dt} = -(K_{21} + K_{31})p_1.$$

$$p_1 = \mathrm{e}^{-k_{\mathrm{tot}}t}, \quad k_{\mathrm{tot}} = K_{21} + K_{31}.$$

$$p_2 = \frac{K_{21}}{K_{21} + K_{31}}(1 - \mathrm{e}^{-k_{\mathrm{tot}}t}), \; p_3 = \frac{K_{31}}{K_{21} + K_{31}}(1 - \mathrm{e}^{-k_{\mathrm{tot}}t}).$$

$$P(t_1) = k_{\mathrm{tot}}\mathrm{e}^{-k_{\mathrm{tot}}t_1}$$

$$\langle t_1 \rangle = 1/k_{\mathrm{tot}}$$

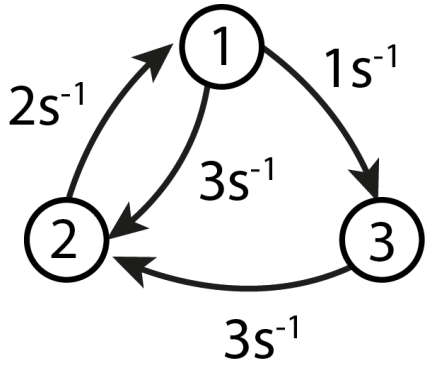


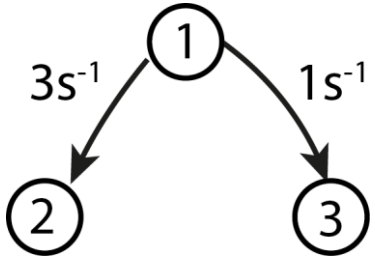
Discrete-state, continuous time processes from the trajectory perspective

Simulation is therefore easy:

Pick a new state with the appropriate probability.

$$p_{1\to2} = \frac{K_{21}}{K_{21}+K_{31}},\ p_{1\to3} = \frac{K_{31}}{K_{21}+K_{31}}.$$



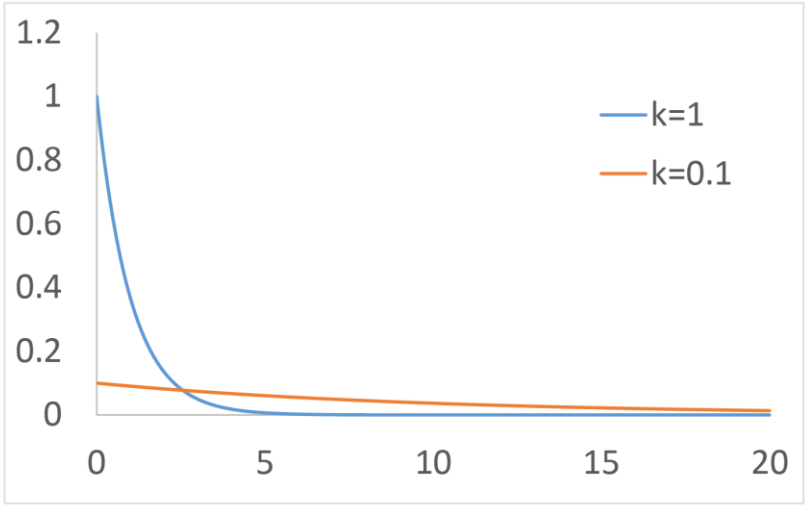


Sample a time from the exponential distribution.

$$P(t_1) = k_{\text{tot}} \mathrm{e}^{-k_{\text{tot}} t_1}$$

Repeat for new state.

Key Algorithm

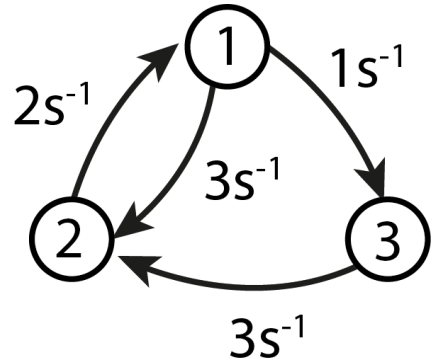


Discrete-state, continuous time processes from the trajectory perspective

If we ignore the time, the sequence of states visited is equivalent to a discrete time model.

- We can use the fundamental matrix method to analyse the sequence of states visited.

- Actual times can be recovered since the time spent in one state prior to transition is $P(t) = k_{\mathrm{tot}} \mathrm{e}^{-k_{\mathrm{tot}} t}$.



?

Continuous-space, continuous-time Markov processes

The Master equation now involves an integral.

$$\frac{dp_i(t)}{dt} = \sum_{j \neq i} K_{ij} p_j(t) - K_{ji}\, p_i(t) \longrightarrow \frac{\partial p(x,t)}{\partial t} = \int dx' K(x|x') p(x',t) - K(x'|x) p(x,t).$$

The **Fokker-Planck** equation:

Taylor expansion

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\big(A(x)p(x,t)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(B(x)p(x,t)\big).$$

The meaning of the Fokker-Planck equation

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\big(A(x)p(x,t)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(B(x)p(x,t)\big).$$

Drift term

Diffusion term

Key Concept

Trial solutions: check by substitution

$$\frac{\partial p(x,t)}{\partial t} = D\frac{\partial^2}{\partial x^2}p(x,t)$$

Key Technique

$$0 = -\frac{\partial}{\partial x}\big(A(x)p(x,t)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(B(x)p(x,t)\big).$$

Stationary distribution

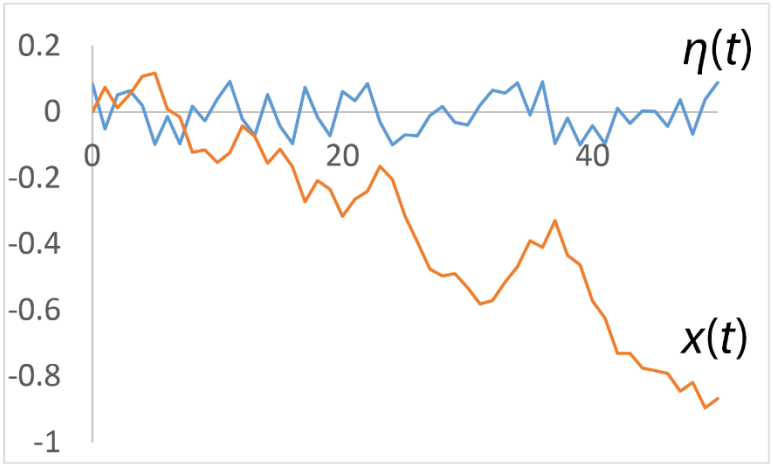
Continuous-state, continuous time processes from the trajectory perspective

Key Concept

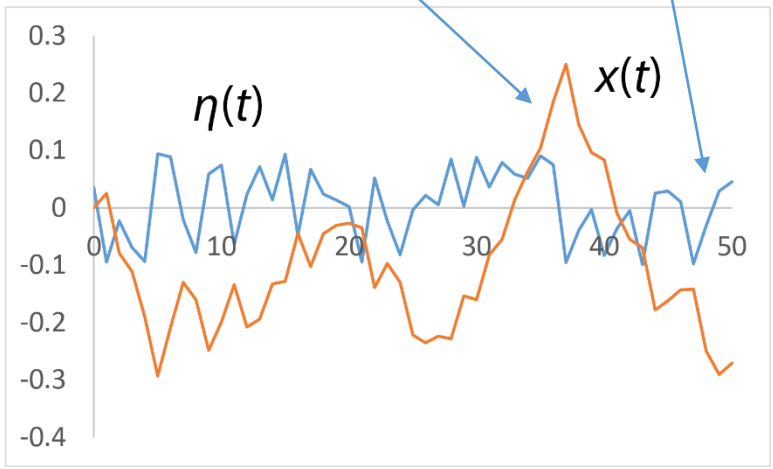
$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\big(A(x)p(x,t)\big) + D\frac{\partial^2}{\partial x^2}p(x,t). \quad \longrightarrow \quad \frac{dX(t)}{dt} = A(X) + \zeta(t)$$

Specific realisations

(a)



(b)



Continuous-state, continuous time processes from the trajectory perspective

Key Concept

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\big(A(x)p(x,t)\big) + D\frac{\partial^2}{\partial x^2}p(x,t). \longrightarrow \frac{dX(t)}{dt} = A(X) + \zeta(t)$$

$A(X)$ is a deterministic force.

$\zeta(t)$ is a Gaussian random force.

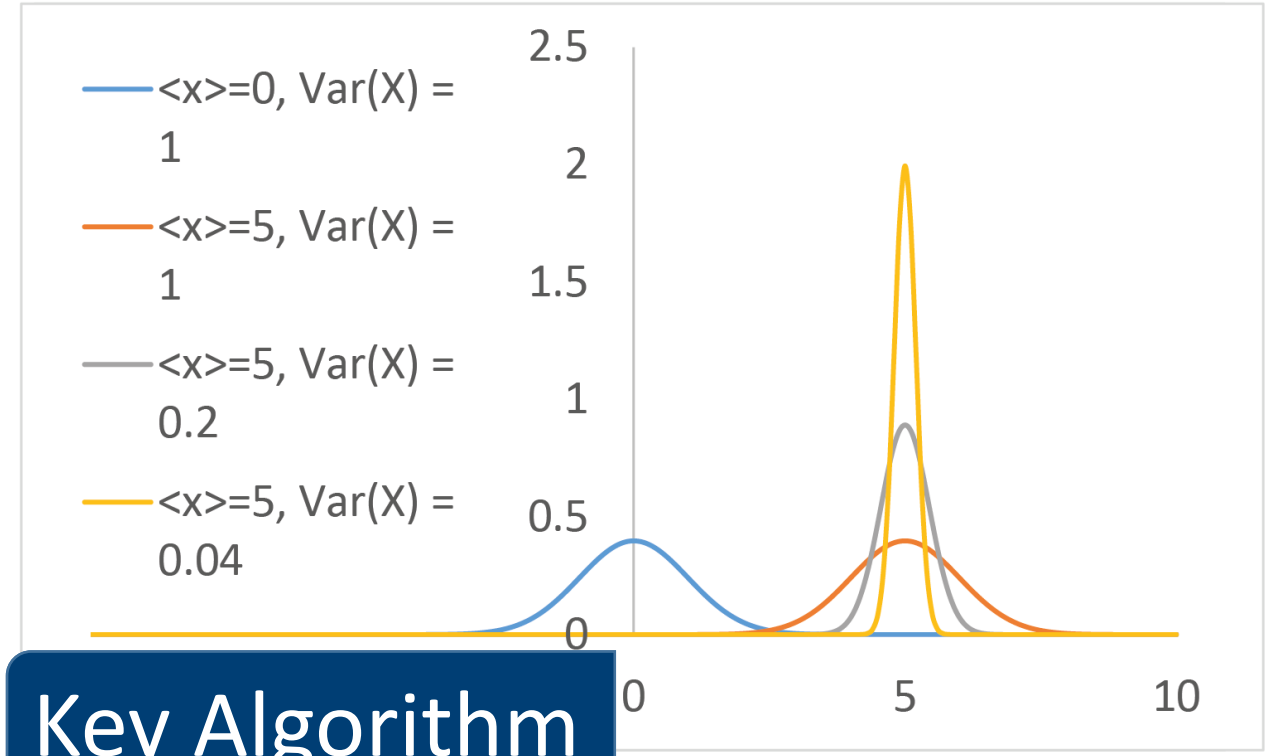
- $\langle\zeta(t)\rangle = 0.$
- $\langle\zeta(t)\zeta(t')\rangle = 2D\delta(t-t').$

This is the **Langevin** equation

Net effect over Δt:

$$\Delta x = A(x)\Delta t + \int_0^{\Delta t} dt\,\eta(t)$$

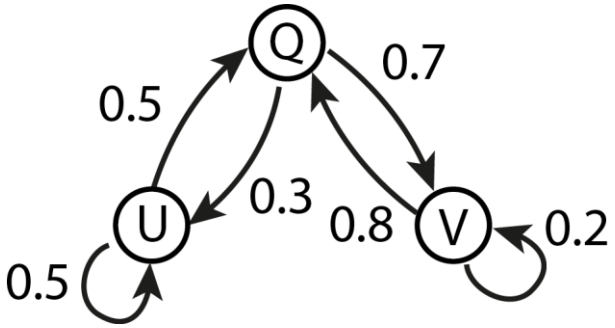
Also Gaussian with mean 0 and variance $2D\Delta t$.

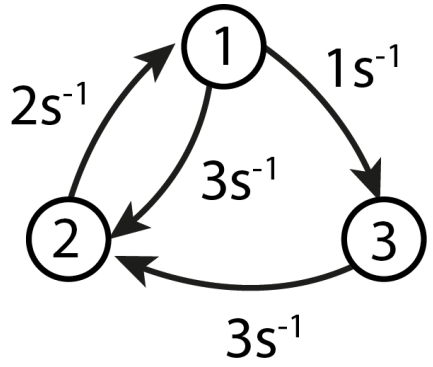


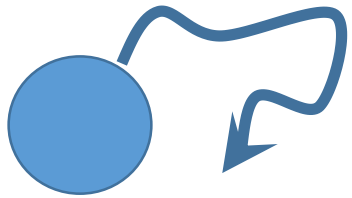
Key Algorithm

Summary

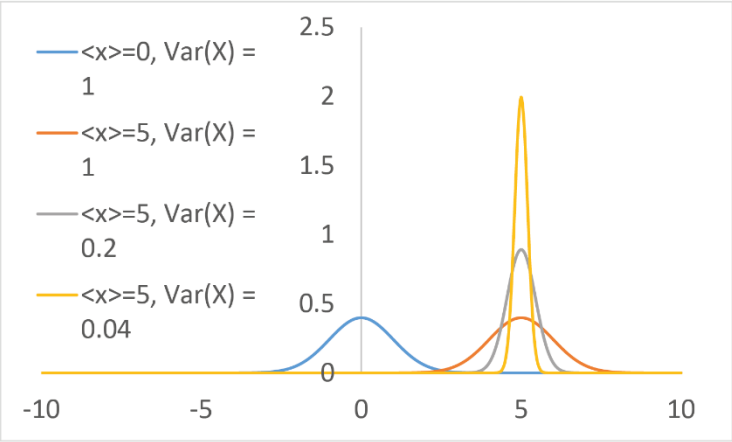
Type of process	Following the probability distribution	Following a sample trajectory
Discrete state, discrete time	Transition matrix equation	Simulation: Step-by-step
Discrete state, continuous time	Master equation. Discretised step-by step integrator	Simulation: Gillespie algorithm
Continuous state, continuous time	Fokker-Planck equation	Langevin equation







Conundrum



Consider the Fokker-Planck Equation:

$$\frac{\partial p(x,t)}{\partial t} = 3p(x,t) + \frac{\partial^2}{\partial x^2} p(x,t).$$

By thinking qualitatively about how the two terms would cause a Gaussian $p(x,t)$ to evolve, Identify the subsequent behaviour:

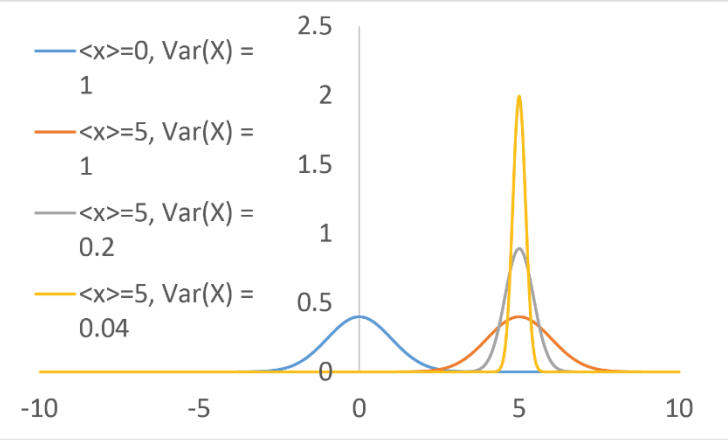
- (a) $p(x,t)$ moves to the left (negative x) and gets broader.
- (b) $p(x,t)$ moves to the left (negative x) and gets narrower.
- (c) $p(x,t)$ moves to the right (positive x) and gets broader.
- (d) $p(x,t)$ moves to the right (positive x) and gets narrower

Go to www.menti.com and use code 79 31 70

Modelling in Biology II: Stochastic processes and networks

Thomas Ouldridge
t.ouldridge@imperial.ac.uk

Conundrum



Consider the Fokker-Planck Equation:

$$\frac{\partial p(x,t)}{\partial t} = 3p(x,t) + \frac{\partial^2}{\partial x^2} p(x,t).$$

By thinking qualitatively about how the two terms would cause a Gaussian $p(x,t)$ to evolve, Identify the subsequent behaviour:

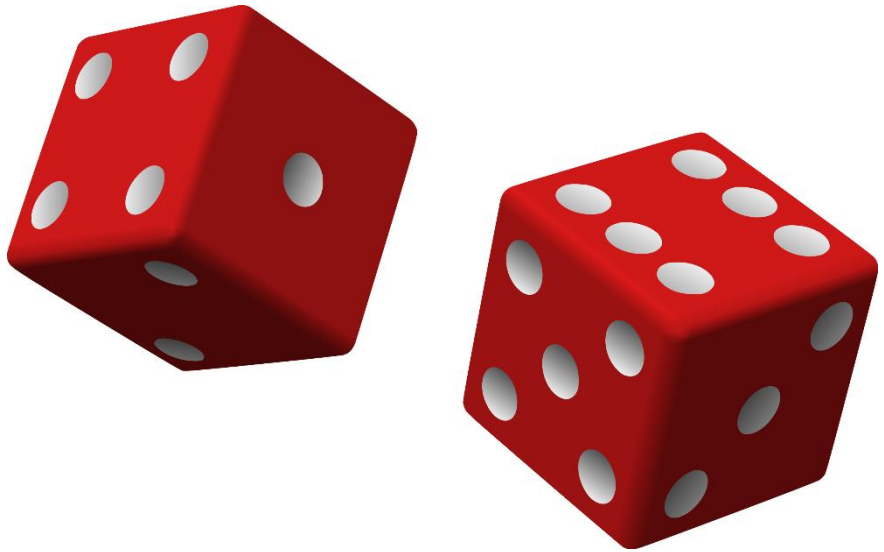
- (a) $p(x,t)$ moves to the left (negative x) and gets broader.
- (b) $p(x,t)$ moves to the left (negative x) and gets narrower.
- (c) $p(x,t)$ moves to the right (positive x) and gets broader.
- (d) $p(x,t)$ moves to the right (positive x) and gets narrower

Go to www.menti.com and use code 79 31 70

Lecture outline

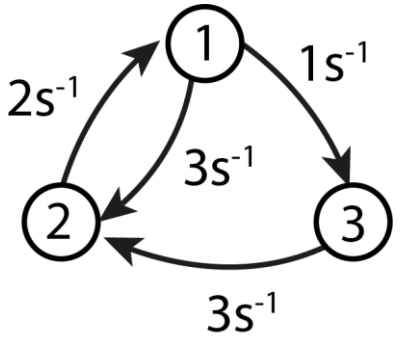
- Finishing of continuous time, discrete state Markov processes.
- Continuous time and continuous state Markov processes

- Different types of stationary distribution: in and out of detailed balance
- Examples of biological systems in equilibrium.



Detailed balance – a type of stationary distribution

Define the flux $\varphi_{ji} = K_{ji}p_i - K_{ij}p_j$

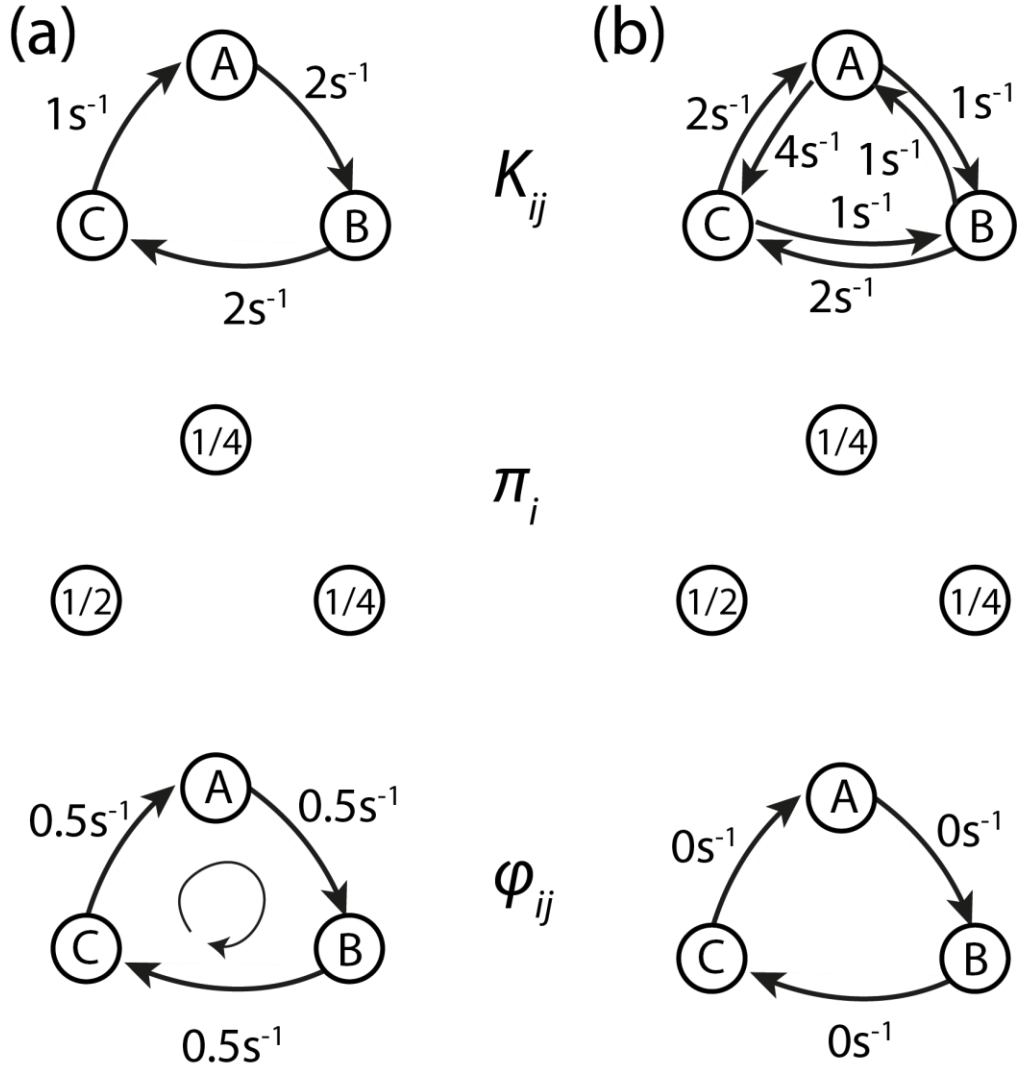


Key Concept

Net flow from i to j.

Not the same as $K_{ji} - K_{ij}$; probabilities also count.

For a stationary distribution in detailed balance, $\varphi_{ji} = 0$ for all pairs.



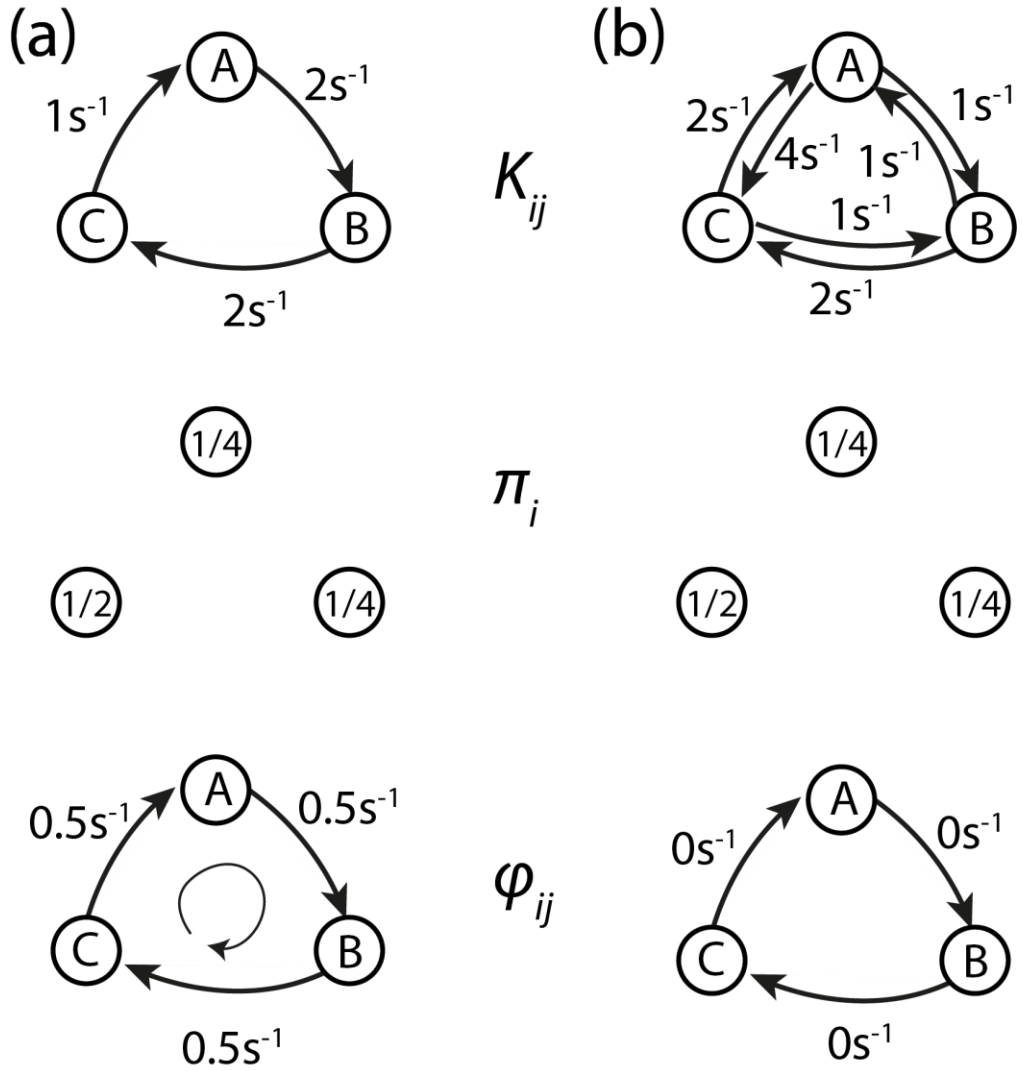
Detailed balance – a type of stationary distribution

Whether a system has detailed balance in the stationary distribution depends on K_{ij} and π_j.

$$\varphi_{ji} = K_{ji}\pi_i - K_{ij}\pi_j$$

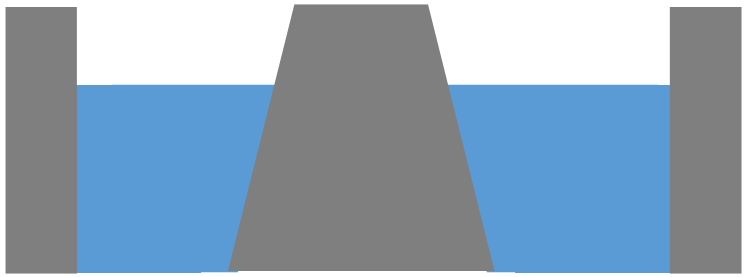
π_i is set by K_{ji}. So detailed balance is a property of the rate matrix (or equivalent).

A randomly generated K_{ji} will almost certainly not show detailed balance. But (bio)physical systems often do, and this helps a lot!

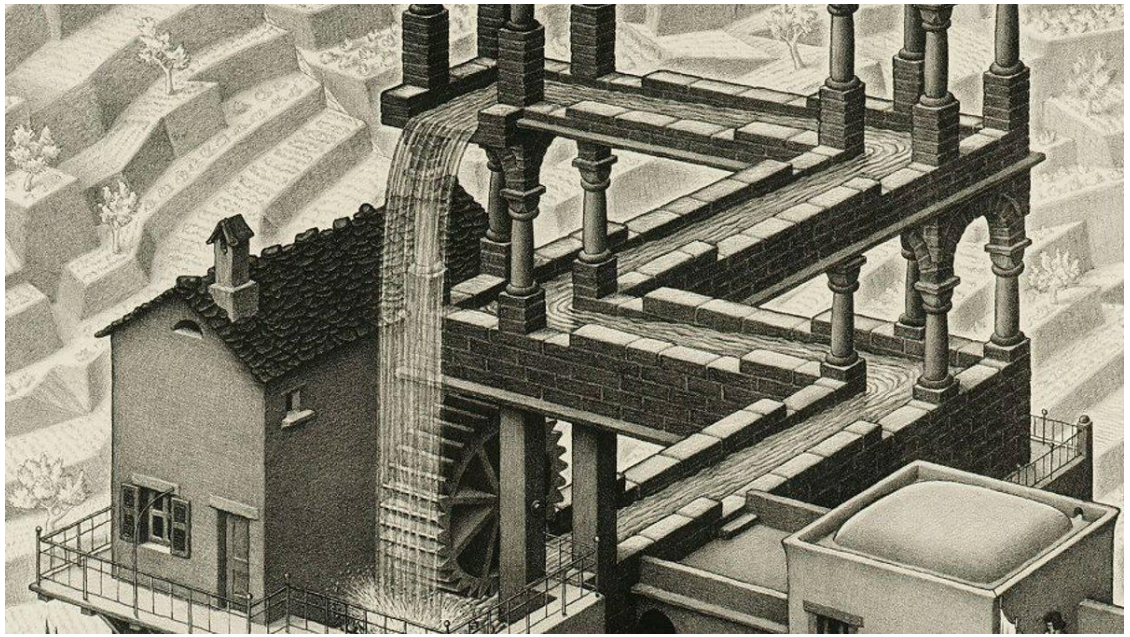


Systems that can reach equilibrium show detailed balance

Systems will tend towards thermodynamic equilibrium if they're not constantly driven by work or fuel.



Once a state has reached equilibrium, we can no longer extract useful work from it.



There are no net fluxes in equilibrium.

Systems that eventually reach equilibrium obey detailed balance.

Key Concept

Consequences of detailed balance

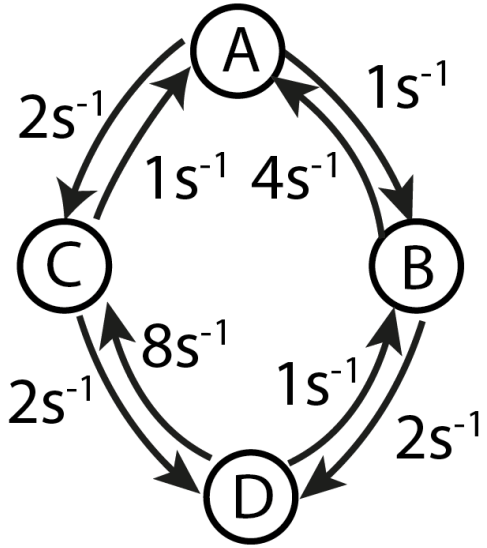
Relative probabilities are easy to calculate.

$$\varphi_{ji} = K_{ji}\pi_i - K_{ij}\pi_j = 0$$

$$\Rightarrow \frac{\pi_i}{\pi_j} = \frac{K_{ij}}{K_{ji}}$$

Holds regardless of the rest of all other K_{ji}.

Can even make progress if two states are not directly connected.



Key Technique

The Boltzmann distribution

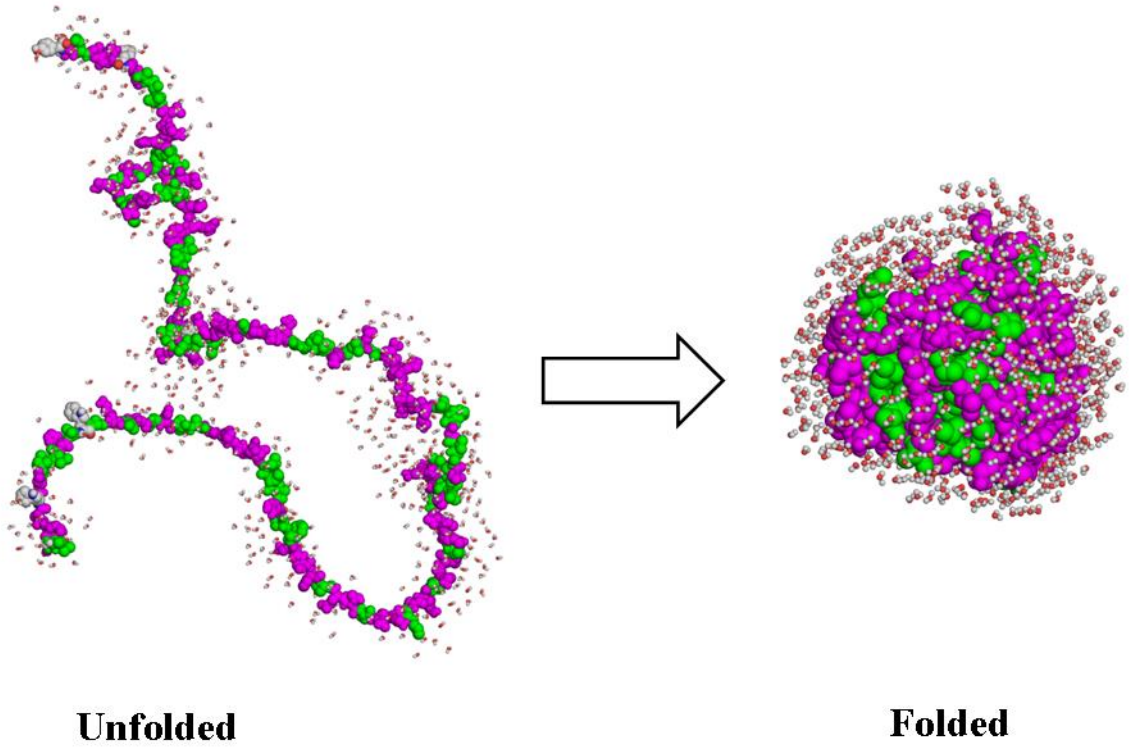
Not only do equilibrium systems obey detailed balance, the stationary distribution is given by:

$$\pi_i = \frac{\mathrm{e}^{-F_i/k_BT}}{\sum_i \mathrm{e}^{-F_i/k_BT}}.$$

F_i is the free-energy of state i. It's like an energy, but also includes a contribution from the number of microstates in the macrostate i (the entropy of state i).

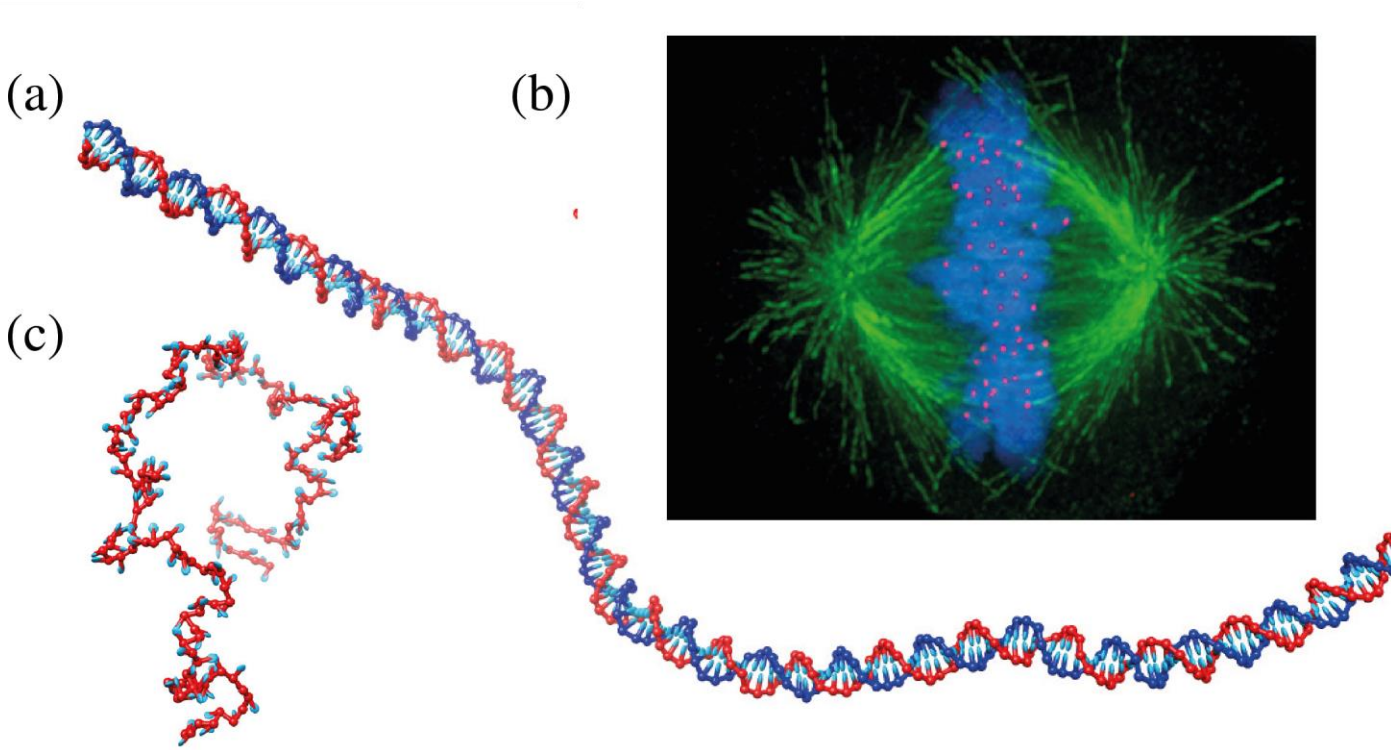
$$F_i = E_i - TS_i.$$

Relative rates are determined by differences in free energy



Key Concept

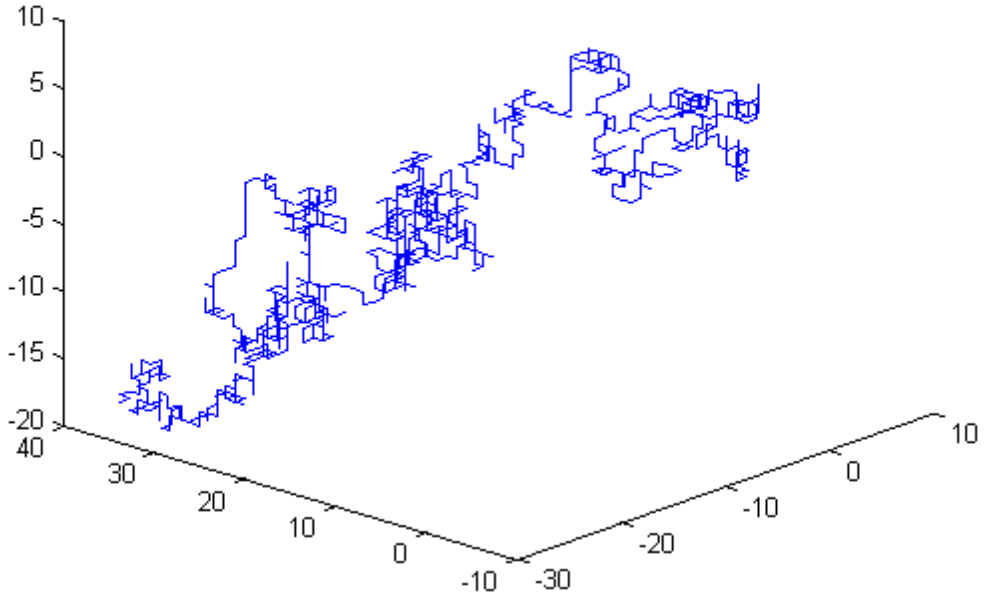
Example: Lattice polymer elasticity



Application

Lattice freely-jointed chain:

- N units.
- Each length b.
- Each points in one of the six directions on a cubic lattice.
- No self-avoidance.



Example: Lattice polymer elasticity

- Each monomer independent.
- All configurations have the same free energy.

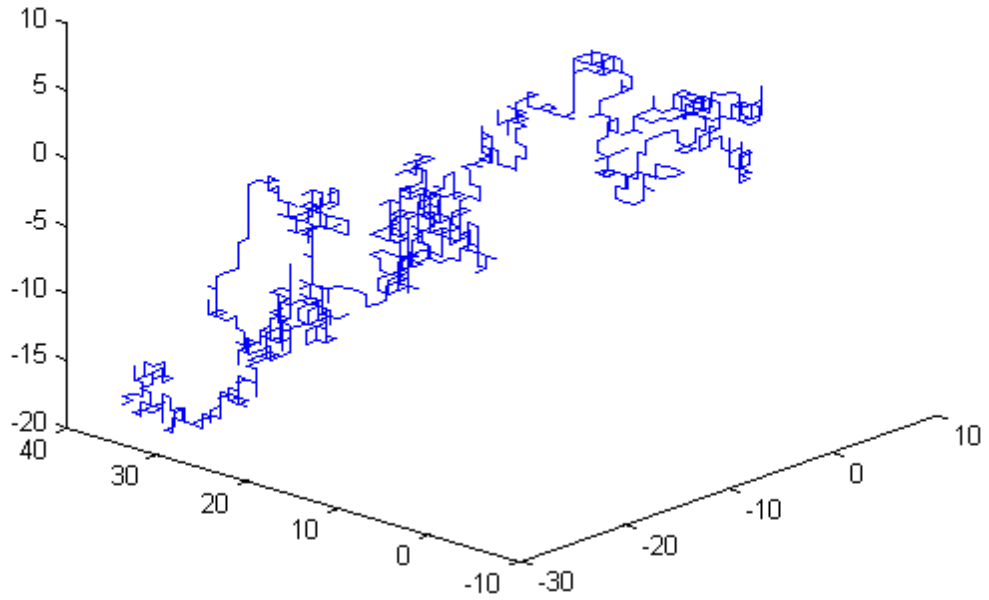
Consider behaviour of:

$$\langle \boldsymbol{R} \rangle = \sum_{i=1}^{N} \langle \boldsymbol{R}_i \rangle$$

$$\langle \boldsymbol{R}^2 \rangle = \left\langle \left(\sum_{i=1}^{N} \boldsymbol{R}_i \right)^2 \right\rangle$$

Lattice freely-jointed chain:

- N units.
- Each length b.
- Each points in one of the six directions on a cubic lattice.
- No self-avoidance.



Example: Lattice polymer elasticity

Now apply a tension f along x:

- Each monomer independent.
- Free energy is orientation-dependent.

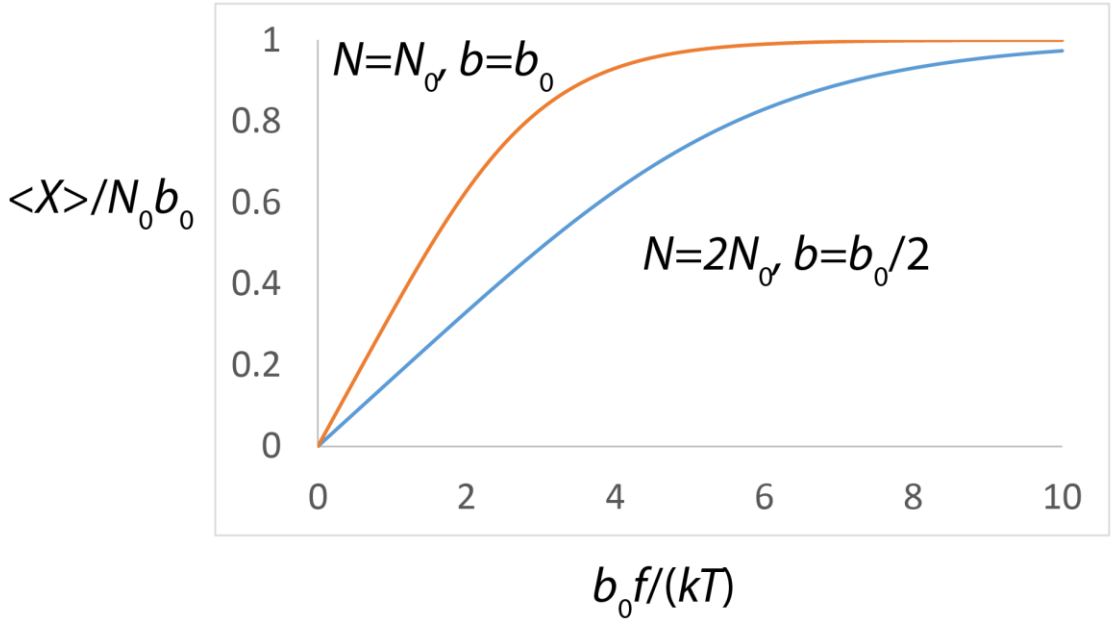
$$P(x_n) \propto \begin{cases} \exp(bf/k_BT) & \text{for } x_n = b, \\ \exp(-bf/k_BT) & \text{for } x_n = -b, \\ 4 & \text{for } x_n = 0. \end{cases}$$

Consider behaviour of:

$$\langle X \rangle = \sum_{i=1}^{N} \langle X_i \rangle = Nb \frac{\sinh\left(\frac{bf}{k_BT}\right)}{2 + \cosh\left(\frac{bf}{k_BT}\right)}$$

Lattice freely-jointed chain:

- N units.
- Each length b.
- Each points in one of the six directions on a cubic lattice.
- No self-avoidance.



Example: Overdamped particles

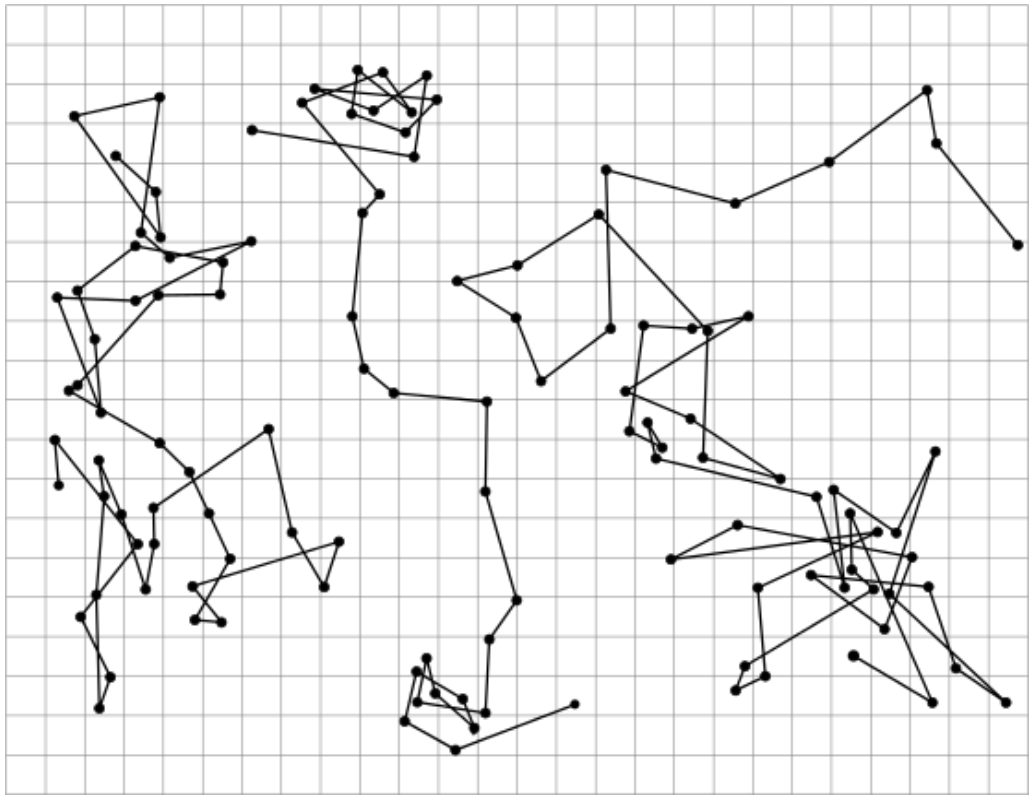
Model as a continuous-time, continuous-state Markov process.

Application

Forces on the particle:

- Drag force $-\gamma v$.
- Force due to potential energy $-\frac{dV(X)}{dX}$.
- Random force.

In the overdamped limit, these must be equal (no ballistic motion).



$$\frac{dX}{dt} = -\frac{1}{\gamma}\frac{dV(X)}{dx} + \zeta(t) \longleftarrow \langle \zeta(t)\zeta(t') \rangle = 2D\delta(t-t'), D = k_B T/\gamma$$

Example: Overdamped particles

$$\frac{dx}{dt} = -\frac{1}{\gamma}\frac{dV(X)}{dx} + \zeta(t)$$

$$\langle \zeta(t)\zeta(t') \rangle = 2D\delta(t-t'), D = k_B T/\gamma$$

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left(-\frac{1}{\gamma}\frac{dV(x)}{dx}p(x,t)\right) + \frac{k_B T}{\gamma}\frac{\partial^2}{\partial x^2}p(x,t).$$

Example: Overdamped particles

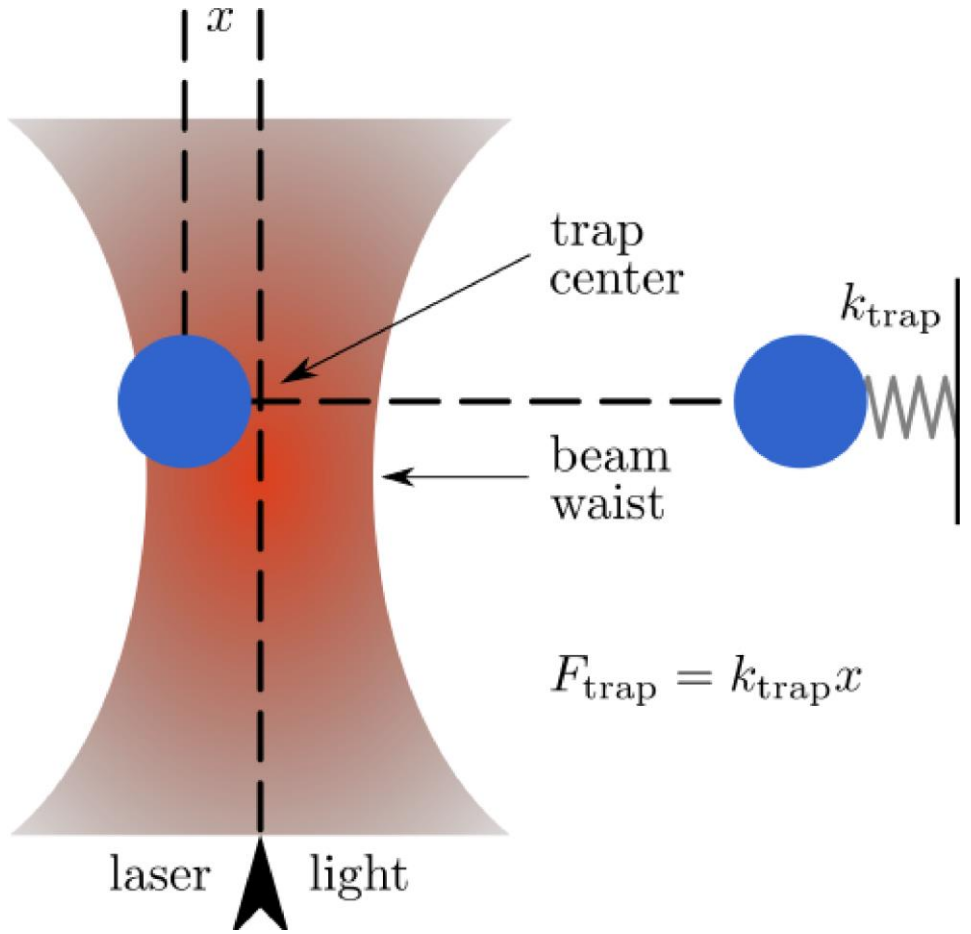
$$0 = \frac{1}{\gamma}\frac{\partial}{\partial x}\left(k_{\text{trap}} x \pi(x)\right) + \frac{k_B T}{\gamma}\frac{\partial^2}{\partial x^2}\pi(x).$$

Trial solution:

$$\pi(x) \propto \mathrm{e}^{-\left(k_{\text{trap}}/2k_B T\right)x^2}$$

Note:

- Matches what we expect from Boltzmann.
- Solution is Gaussian; width depends on trap stiffness
- No dependence on γ.



$$V(x) = \frac{1}{2}k_{\text{trap}}x^2$$

Summary

- There are two types of stationary distribution: those with and without detailed balance.
- Equilibrium systems – sometimes a good model in biologically relevant contexts – obey detailed balance and the Boltzmann distribution. Equilibrium modelling is often a first port-of-call for biological molecules with heterogeneous states.
- It is easy to work out the relative probability of two states in a detailed-balanced system in the stationary distribution.
- This is *particularly true* if we know the free energy. Don't even need to bother with the rates!

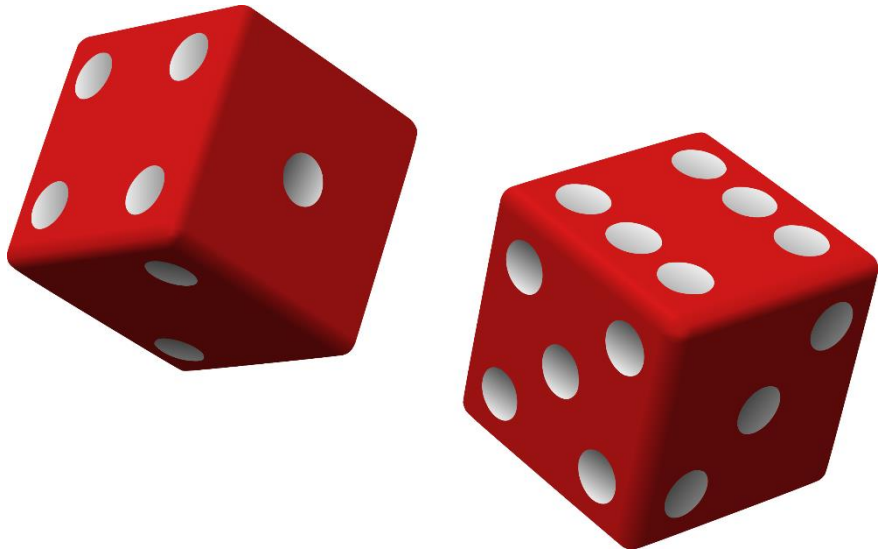
Modelling in Biology II: Stochastic processes and networks

Thomas Ouldridge
t.ouldridge@imperial.ac.uk

Lecture outline

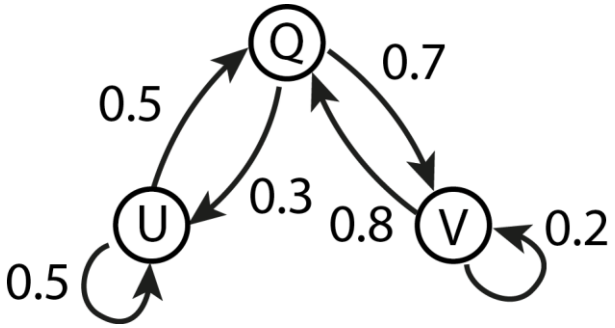
- Non-equilibrium stationary states

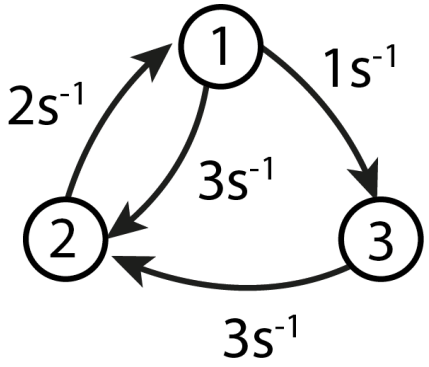
- Time-dependence of stochastic processes

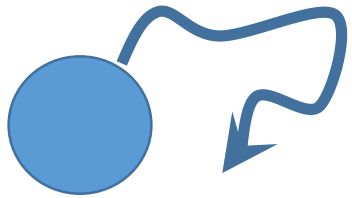


Recap

Type of process	Following the probability distribution	Following a sample trajectory
Discrete state, discrete time	Transition matrix equation	Simulation: Step-by-step
Discrete state, continuous time	Master equation. Discretised step-by step integrator	Simulation: Gillespie algorithm
Continuous state, continuous time	Fokker-Planck equation	Langevin equation

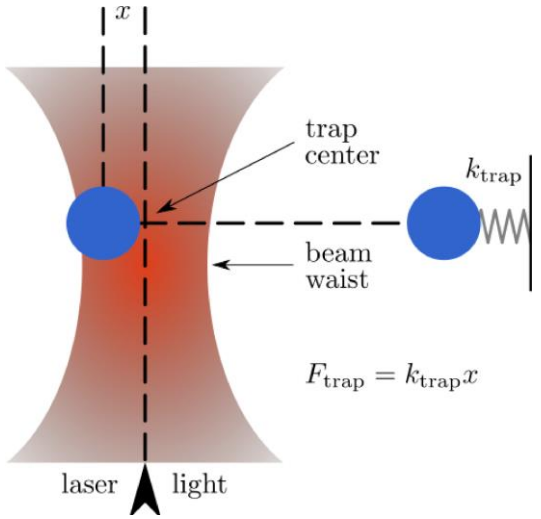


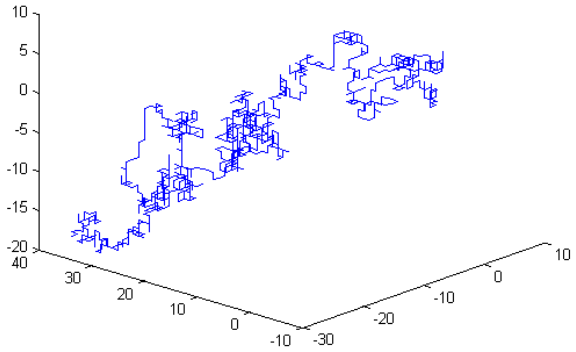


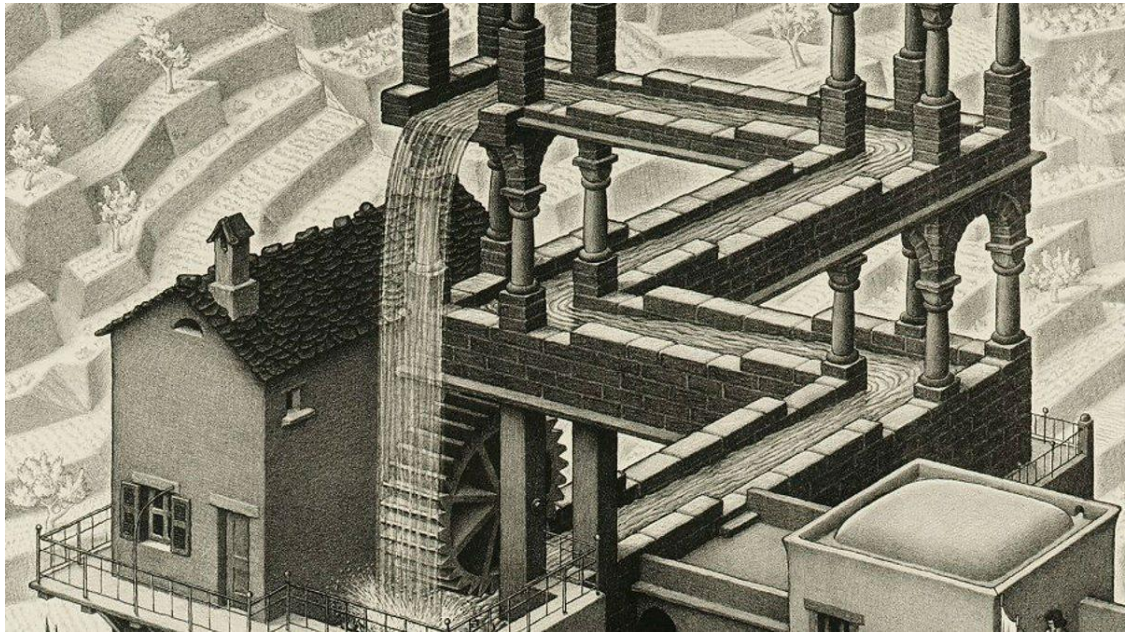


Recap

- Detailed Balance
- Boltzmann Distribution for equilibrium systems





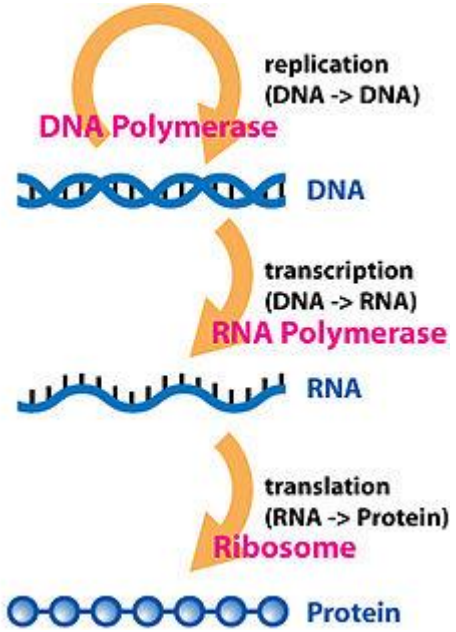


$$\pi_i = \frac{e^{-F_i/k_B T}}{\sum_i e^{-F_i/k_B T}}.$$

Non-equilibrium stationary distributions: population processes

Population $X(t)$:

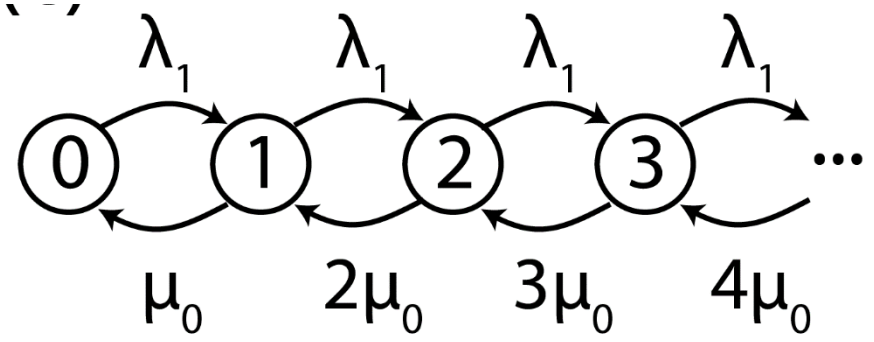
- Increase by one at a rate $\lambda(x)$.
- Decrease by one at a rate $\mu(x)$.



Key Concept

The immigration-death model

$$\frac{dp_x}{dt} = -(\lambda_1 + \mu_0 x)p_x + \lambda_1 p_{x-1} + \mu_0(x+1)p_{x+1}.$$

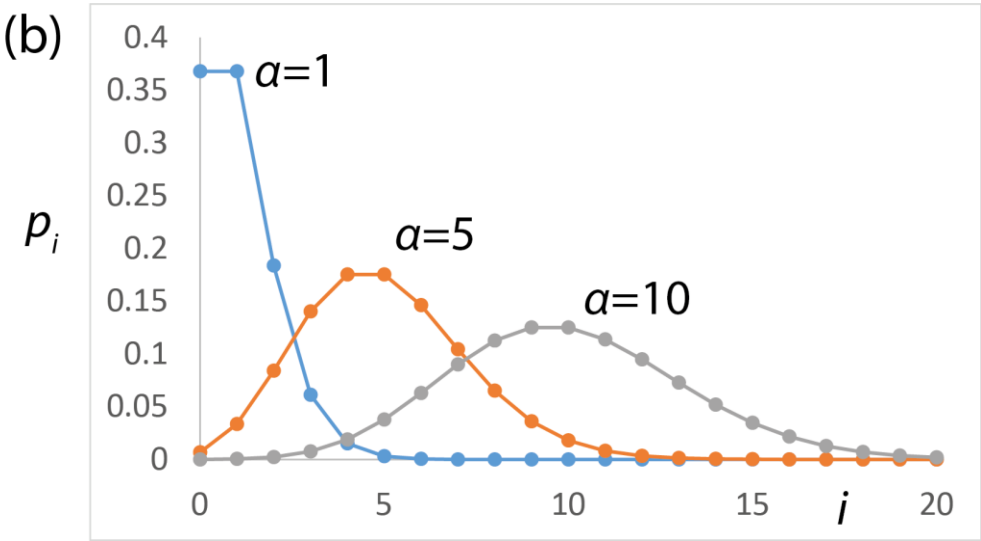


Neat trick: multiply by x and sum over x.

$$\frac{d\langle X\rangle}{dt} = \lambda_1 - \mu_0\langle X\rangle \longrightarrow \lambda_1 = \mu_0\langle X\rangle \text{ in stationary state.}$$

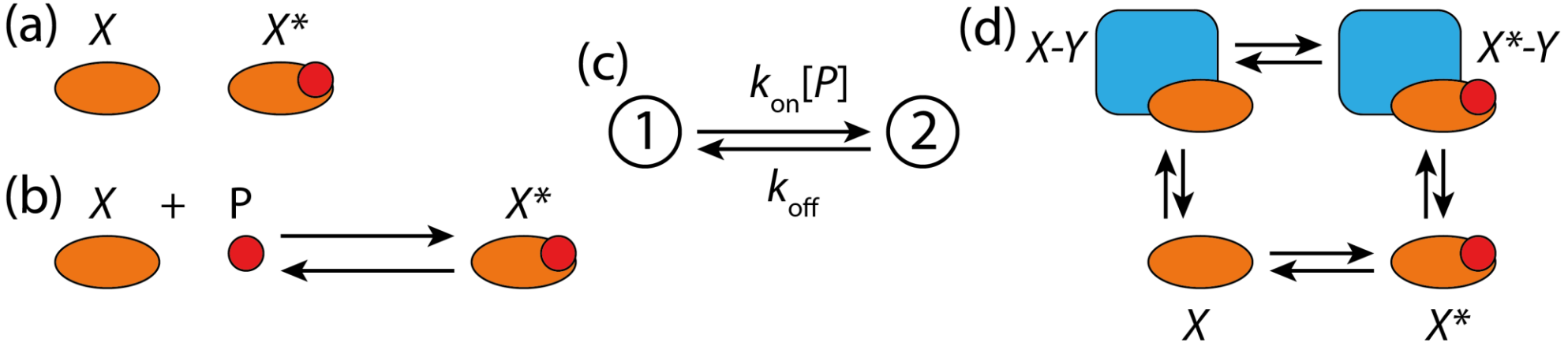
In fact, we can find the whole distribution using detailed balance (!!):

$$\lambda_1 \pi_x = \mu_0(x+1)\pi_{x+1} \qquad \pi_x = \frac{\left(\frac{\lambda_1}{\mu_0}\right)^x}{x!}\pi_0$$



Biasing stationary distributions through driving

X is a protein that can be phosphorylated (X*) or unphosphorylated.



In equilibrium, the proportion of X and X* is determined by the free energy of phosphate binding:

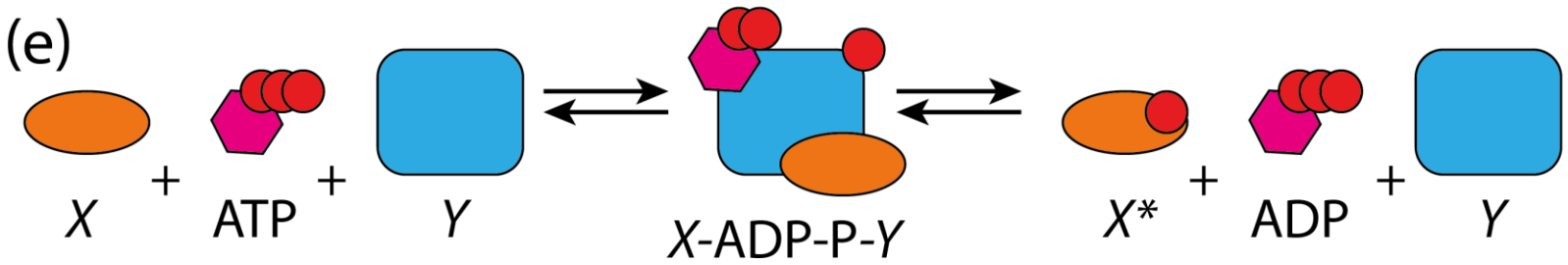
Application

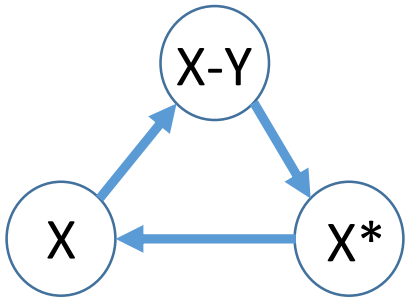
$$\frac{p(X^*)}{p(X)} = \frac{k_{\text{on}}[P]}{k_{\text{off}}} = e^{-\Delta F_{\text{phos}}/k_B T}.$$

Biasing stationary distributions through driving

We can change the bias by driving the system through an input of chemical fuel

Allow phosphorylation to occur by exchange of phosphate with a nucleotide, facilitated by an enzyme.

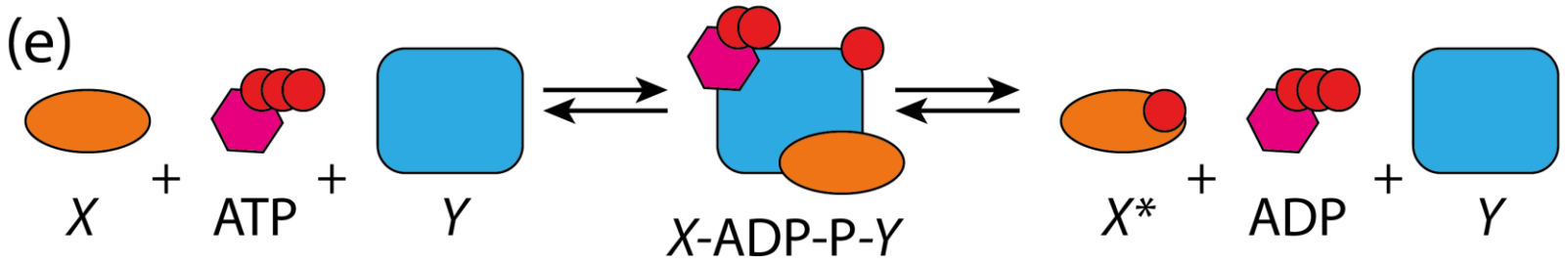




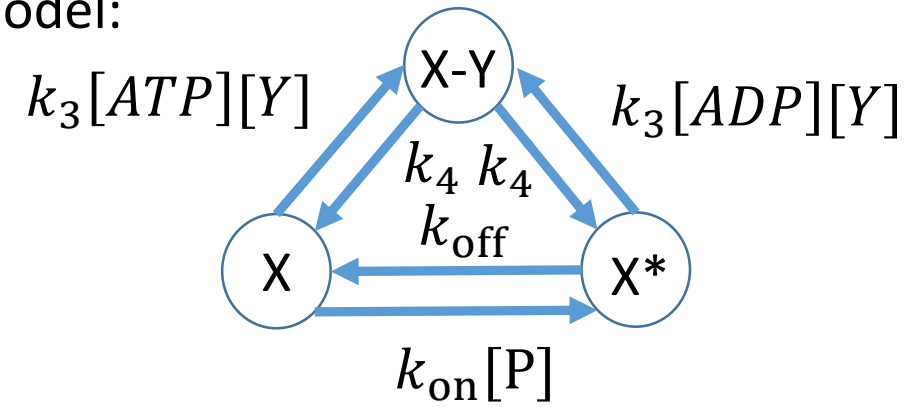
From the perspective of X/X*, fuel is constantly fed in.

No need to obey detailed balance.

Biasing stationary distributions through driving



A very simple model:

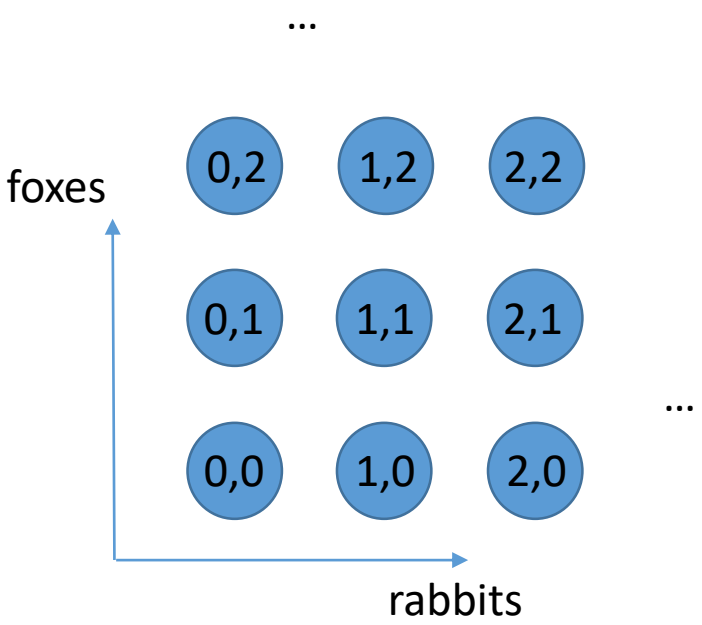


In general, violates detailed balance.

Allows stationary distribution to deviate from $\frac{p(X^*)}{p(X)} = \frac{k_{\text{on}}[P]}{k_{\text{off}}} = e^{-\Delta F_{\text{phos}}/k_B T}$.

Yield is sensitive to amount of Y – good for passing on a signal.

Conundrum



Rabbit population R, fox population F

Model 1 (vegetarian foxes):

- $R \to R + 1$ at rate $\lambda_R R$.
- $R \to R - 1$ at rate $\mu_R R$.
- $F \to F + 1$ at rate $\lambda_F F$.
- $F \to F - 1$ at rate $\mu_F F$.

Model 2 (non-vegetarian foxes)

- $R \to R + 1$ at rate $\lambda_R R$.
- $R \to R - 1$ at rate $\mu_R FR$.
- $F \to F + 1$ at rate $\lambda_F RF$.
- $F \to F - 1$ at rate $\mu_F F$.

Which of the models obeys detailed balance?

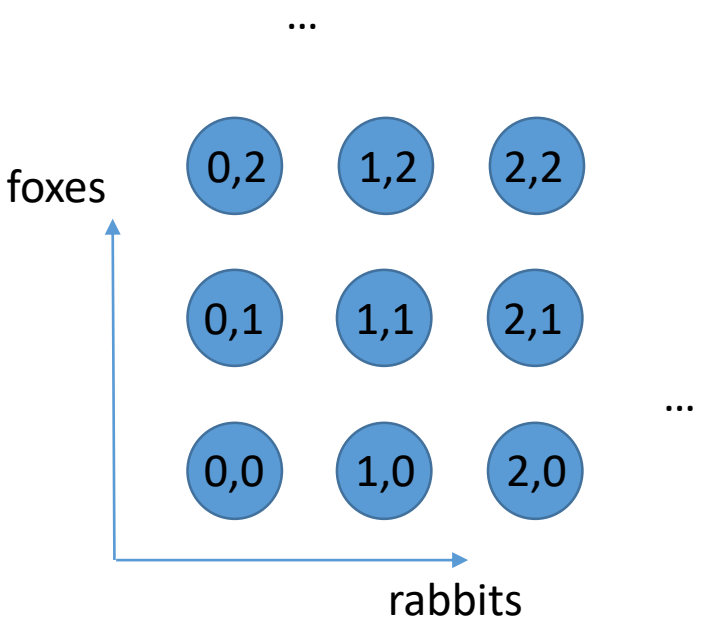
(a) 1 and 2.
(b) 1 only.
(c) 2 only.
(d) Neither.

Go to www.menti.com and input code **59 46 45** to suggest an answer.

Modelling in Biology II: Stochastic processes and networks

Thomas Ouldridge
t.ouldridge@imperial.ac.uk

Conundrum



Rabbit population R, fox population F

Model 1 (vegetarian foxes):

- $R \to R + 1$ at rate $\lambda_R R$.
- $R \to R - 1$ at rate $\mu_R R$.
- $F \to F + 1$ at rate $\lambda_F F$.
- $F \to F - 1$ at rate $\mu_F F$.

Model 2 (non-vegetarian foxes)

- $R \to R + 1$ at rate $\lambda_R R$.
- $R \to R - 1$ at rate $\mu_R FR$.
- $F \to F + 1$ at rate $\lambda_F RF$.
- $F \to F - 1$ at rate $\mu_F F$.

Which of the models obeys detailed balance?

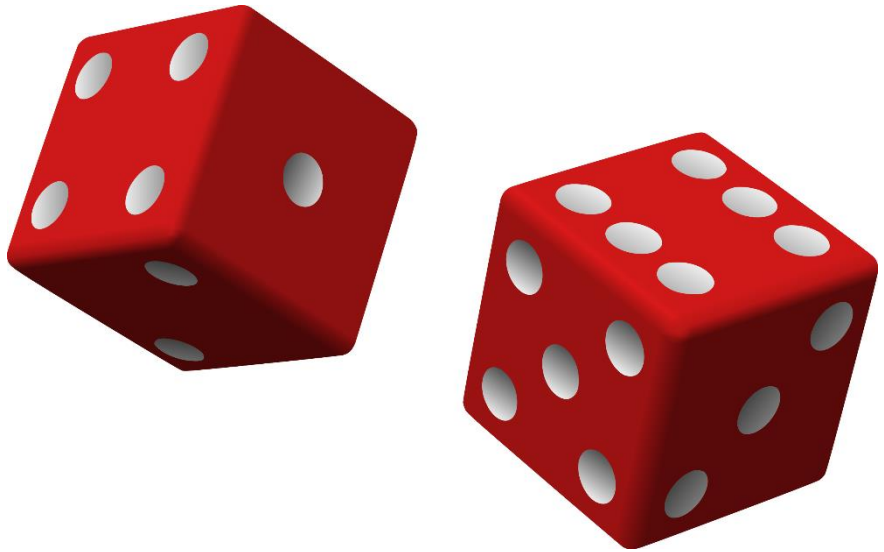
(a) 1 and 2.
(b) 1 only.
(c) 2 only.
(d) Neither.

Go to www.menti.com and input code **59 46 45** to suggest an answer.

Lecture outline

- Non-equilibrium stationary states

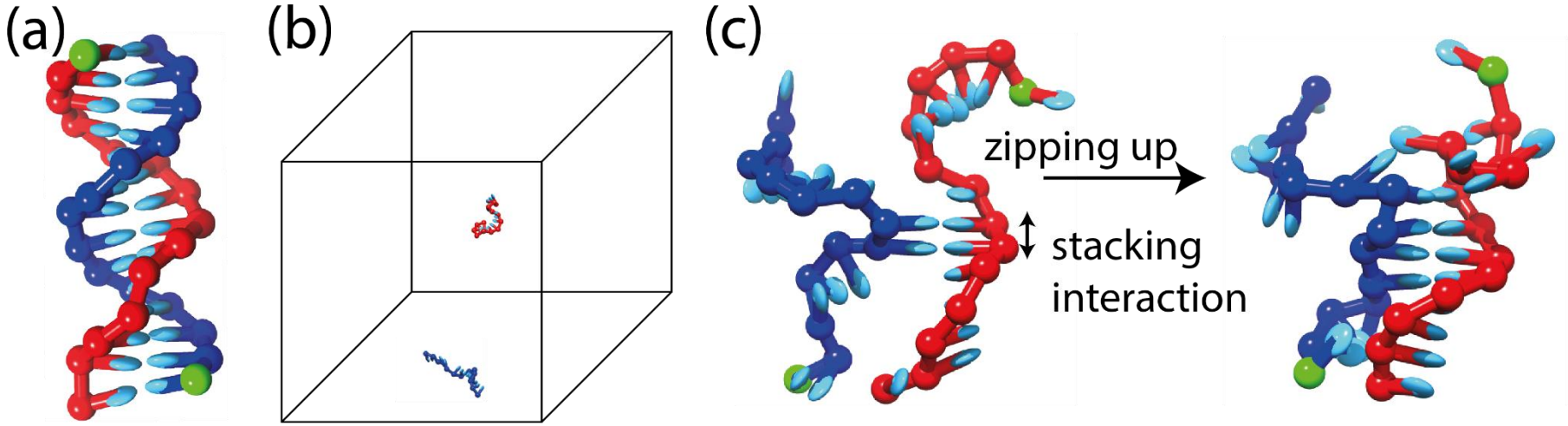
- Time dependence of stochastic processes



Two contexts for time-dependence

- Relaxation of a probability distribution over time towards the stationary distribution.
- The properties of individual trajectories, regardless of whether the distribution as a whole is stationary.

Example: DNA dissociation

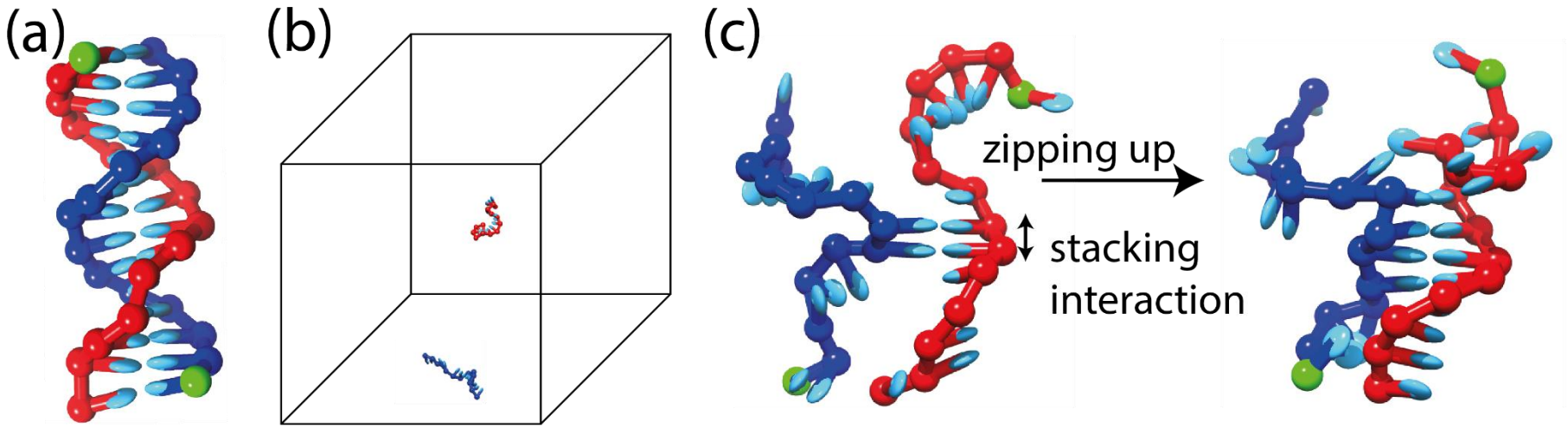


Application

$$\frac{k_{\text{on}}}{k_{\text{off}}} = \sigma \ll 1$$

$$\frac{k_{\text{zip}}}{k_{\text{unzip}}} = t > 1$$

Example: DNA dissociation



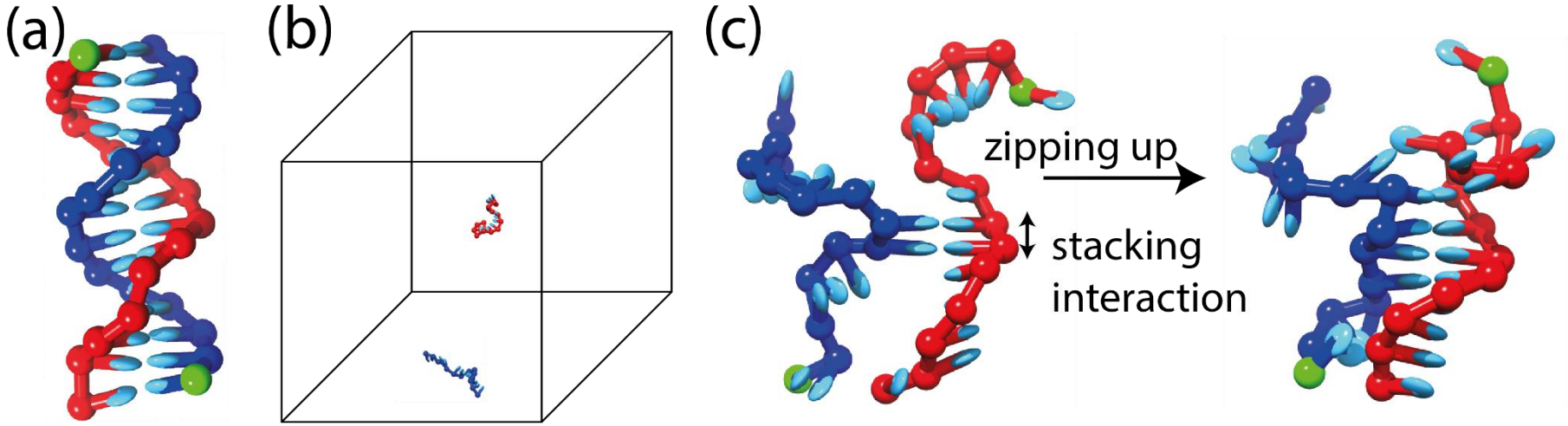
Assumption: system reaches stationary state within the bound states prior to escape.

$$\frac{k_{\text{on}}}{k_{\text{off}}} = \sigma \ll 1 \qquad \frac{k_{\text{zip}}}{k_{\text{unzip}}} = t > 1$$

Overall dissociation rate $\sim k_{\text{off}}\pi(X = 1|X > 0)$

[This is called transition state theory].

Example: DNA dissociation



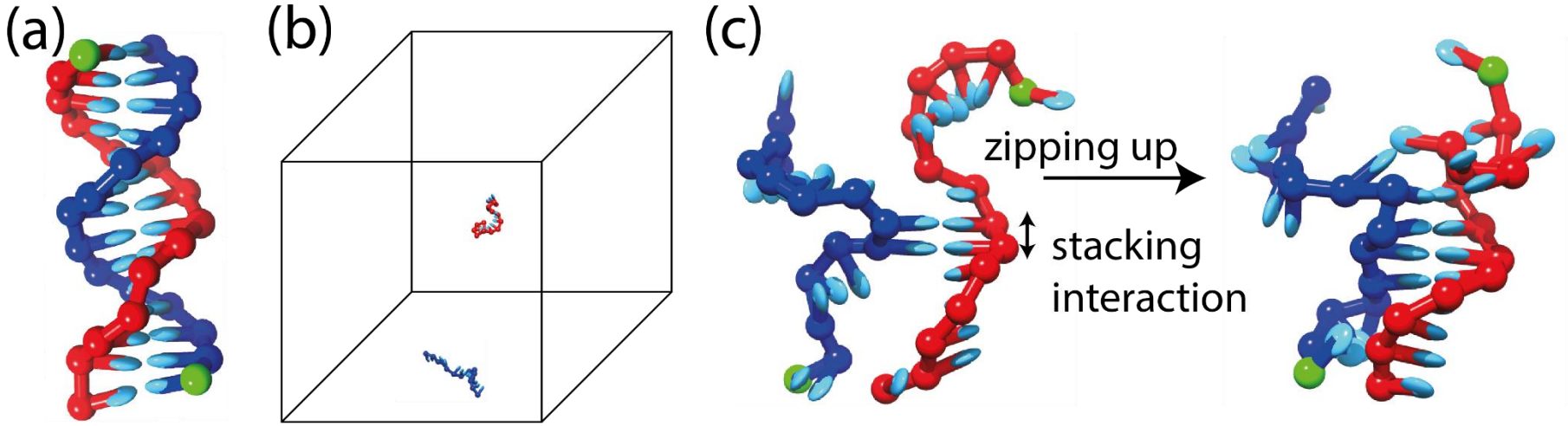
System obeys detailed balance.

$$\frac{k_{\text{on}}}{k_{\text{off}}} = \sigma \ll 1 \qquad \frac{k_{\text{zip}}}{k_{\text{unzip}}} = t > 1$$

$$\pi_{\text{eq}}(x) = \pi_{\text{eq}}(x-1)\frac{k_{\text{zip}}}{k_{\text{unzip}}} = t\pi_{\text{eq}}(x-1)$$

for $x > 1$.

Example: DNA dissociation



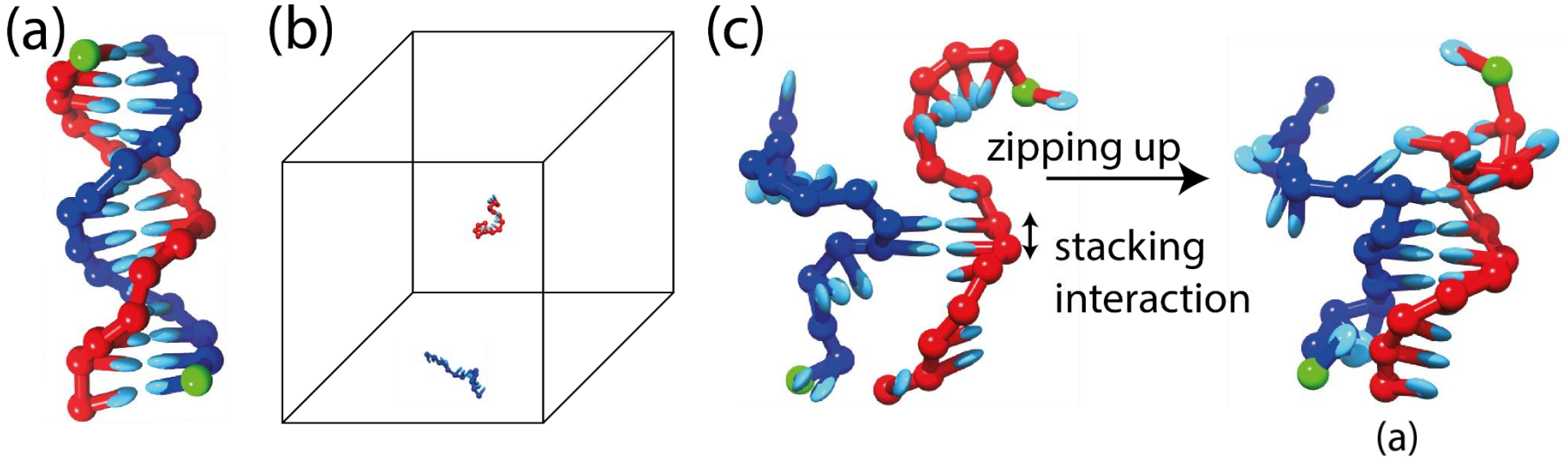
So, for a duplex of length N,

$$\pi_{\mathrm{eq}}(X = 1 | X > 0) = \frac{1}{\sum_{n=1}^{N} t^{n-1}}.$$

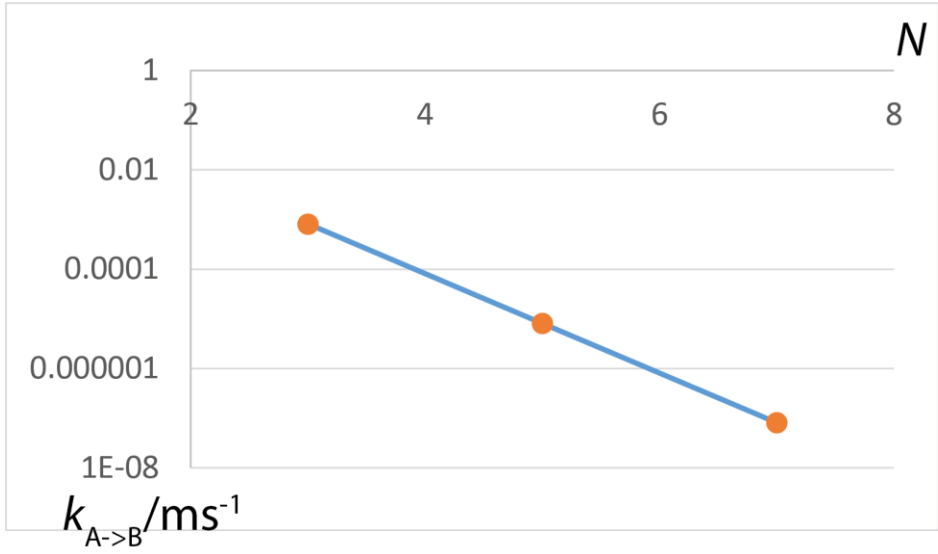
$$\frac{k_{\mathrm{on}}}{k_{\mathrm{off}}} = \sigma \ll 1 \qquad \frac{k_{\mathrm{zip}}}{k_{\mathrm{unzip}}} = t > 1$$

Dissociation rate $\sim \frac{k_{\mathrm{off}}}{\sum_{n=1}^{N} t^{n-1}}$

Example: DNA dissociation



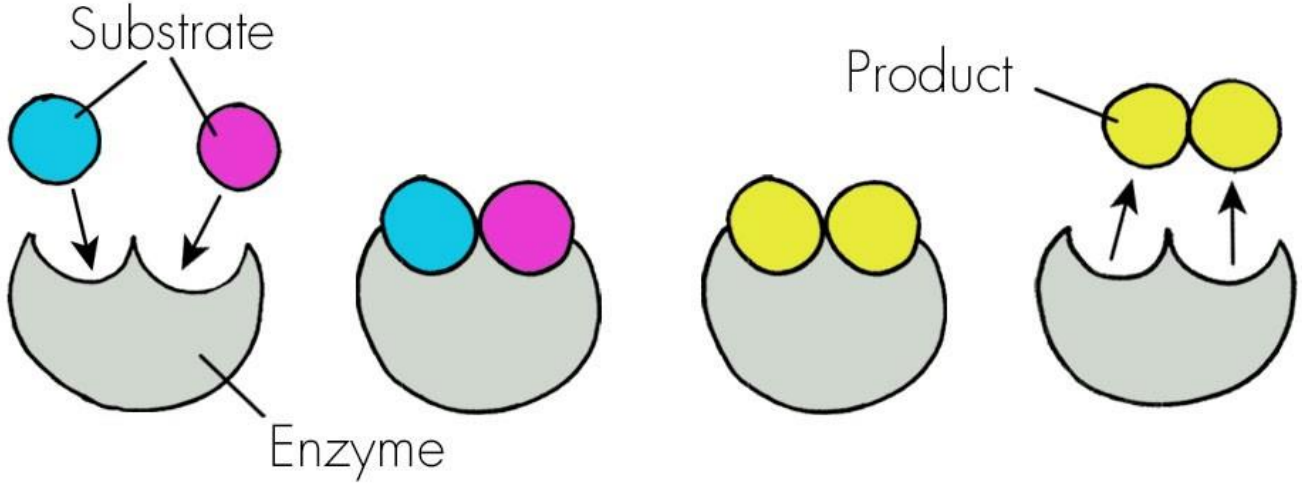
$$\frac{k_{\mathrm{dis}}(N_1)}{k_{\mathrm{dis}}(N_2)} \approx t^{N_2-N_1} = \left(\frac{k_{\mathrm{zip}}}{k_{\mathrm{unzip}}}\right)^{N_2-N_1}$$



Competing outcomes of molecular reactions

When we look a the details of a molecular reaction, we might want to know:

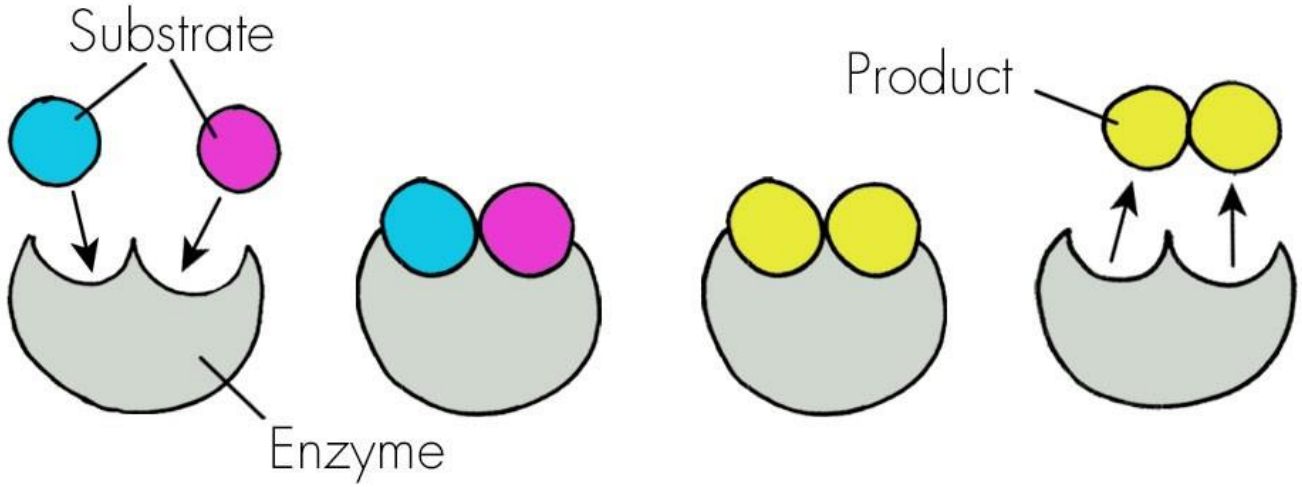
- Given that reactants are bound, does the reaction proceed to completion?
- Given multiple substrates, what is the probability of acting on the intended one?
- Given multiple reaction outcomes, which ones are more likely?



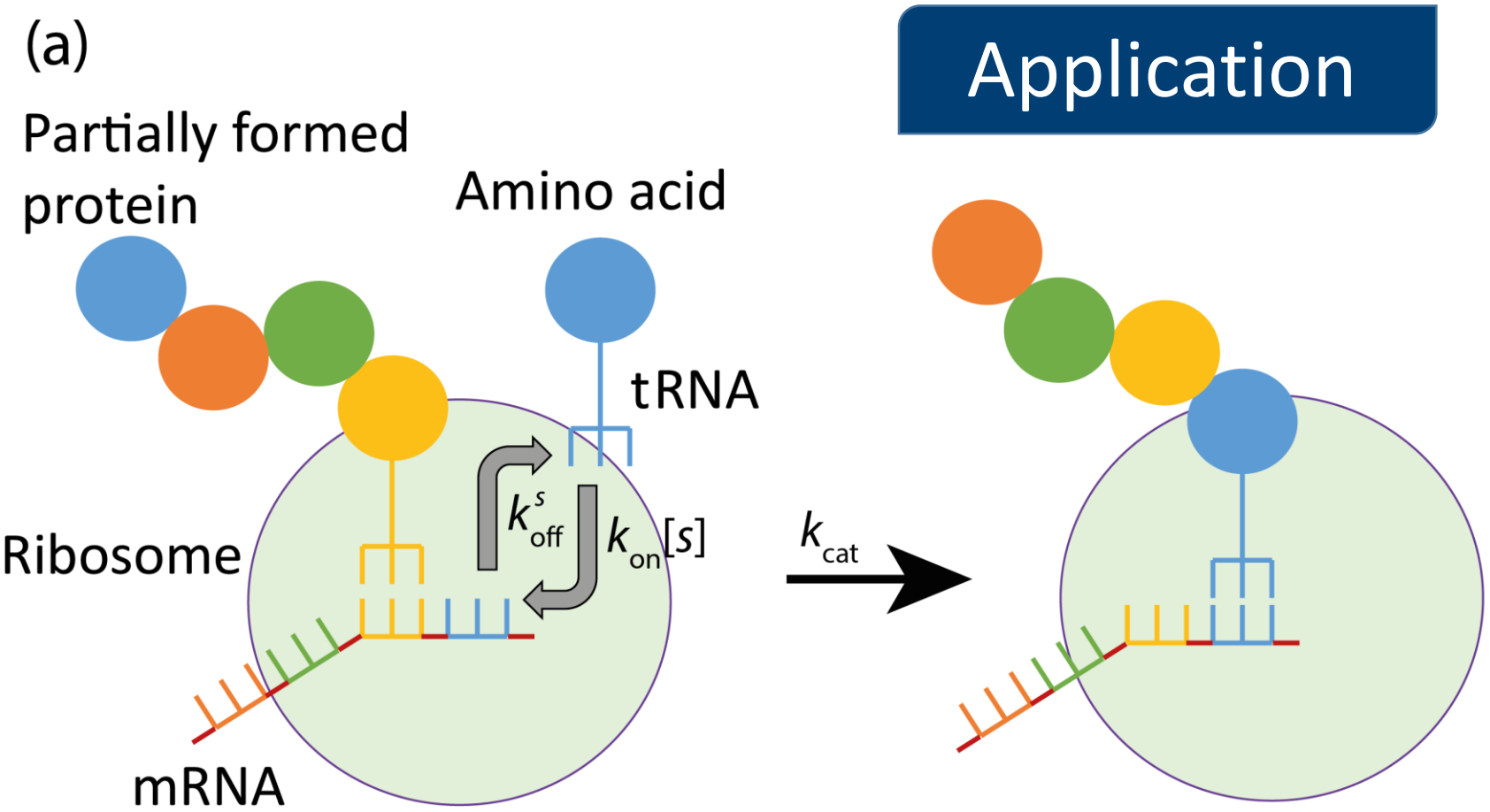
Competing outcomes of molecular reactions

Systems tend naturally to be describable using a discrete-state, continuous time Markov process.

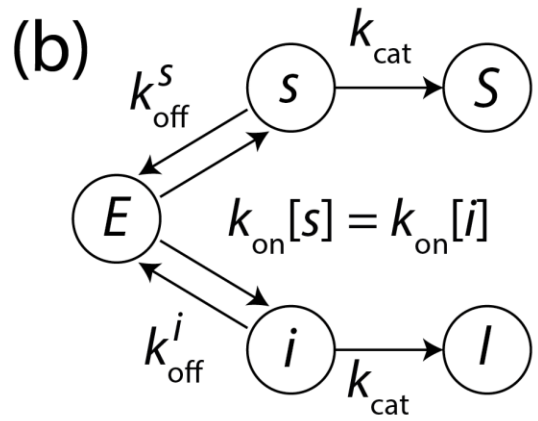
Often helpful to construct a discrete-time process by thinking only about the sequence of states visited (ignoring time).

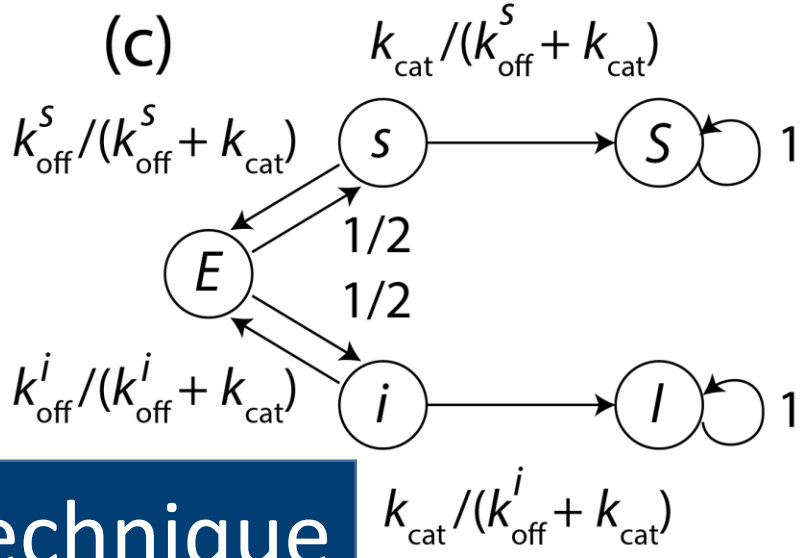


Example: protein translation



Application





Key Technique

Example: protein translation

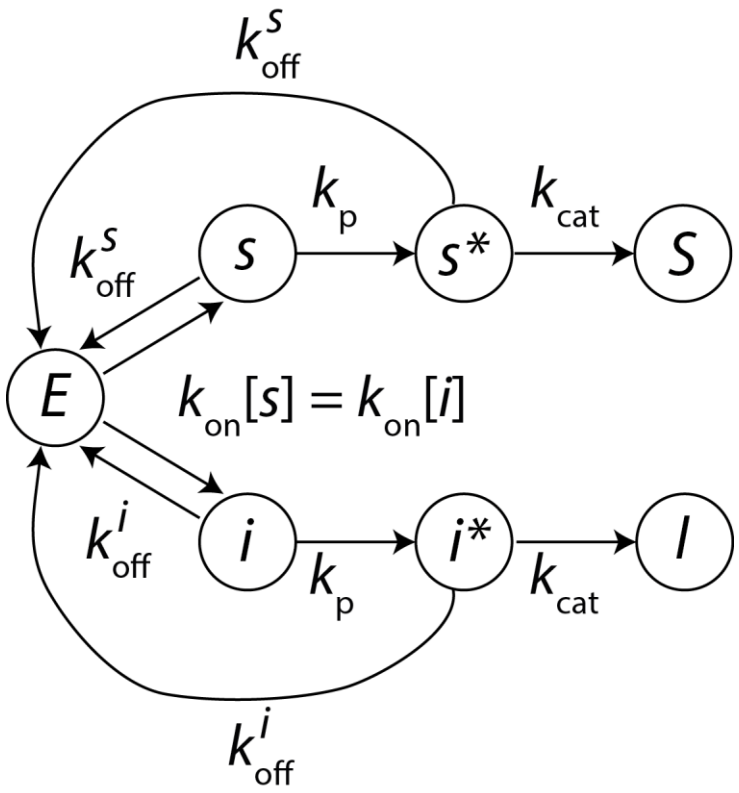
$$W = (\mathrm{I} - U)^{-1} = \frac{1}{1 - \frac{k_{\mathrm{off}}^{s}}{2\left(k_{\mathrm{off}}^{s} + k_{\mathrm{cat}}\right)} - \frac{k_{\mathrm{off}}^{i}}{2\left(k_{\mathrm{off}}^{i} + k_{\mathrm{cat}}\right)}} \begin{pmatrix} 1 & \frac{k_{\mathrm{off}}^{s}}{\left(k_{\mathrm{off}}^{s} + k_{\mathrm{cat}}\right)} & \frac{k_{\mathrm{off}}^{i}}{\left(k_{\mathrm{off}}^{i} + k_{\mathrm{cat}}\right)} \\ 1/2 & 1 - \frac{k_{\mathrm{off}}^{i}}{2\left(k_{\mathrm{off}}^{i} + k_{\mathrm{cat}}\right)} & \frac{k_{\mathrm{off}}^{i}}{2\left(k_{\mathrm{off}}^{i} + k_{\mathrm{cat}}\right)} \\ 1/2 & \frac{k_{\mathrm{off}}^{s}}{2\left(k_{\mathrm{off}}^{s} + k_{\mathrm{cat}}\right)} & 1 - \frac{k_{\mathrm{off}}^{s}}{2\left(k_{\mathrm{off}}^{s} + k_{\mathrm{cat}}\right)} \end{pmatrix}$$

$$R = \begin{pmatrix} 0 & \frac{k_{\mathrm{cat}}}{\left(k_{\mathrm{off}}^{s} + k_{\mathrm{cat}}\right)} & 0 \\ 0 & 0 & \frac{k_{\mathrm{cat}}}{\left(k_{\mathrm{off}}^{i} + k_{\mathrm{cat}}\right)} \end{pmatrix}$$

Ratio of correct/incorrect:

$$\frac{p_S}{p_I} = \frac{k_{\mathrm{off}}^{i} + k_{\mathrm{cat}}}{\left(k_{\mathrm{off}}^{s} + k_{\mathrm{cat}}\right)}$$

Example: protein translation – kinetic proofreading



$$\frac{p_S}{p_I} = \frac{P(s \text{ before } i|\mathrm{E})P(S \text{ before } E|s)}{P(i \text{ before } s|\mathrm{E})P(I \text{ before } E|i)}$$

Thus:

$$\frac{p_S}{p_I} = \frac{k^i_{\mathrm{off}} + k_p}{\left(k^s_{\mathrm{off}} + k_p\right)} \frac{k^i_{\mathrm{off}} + k_{\mathrm{cat}}}{\left(k^s_{\mathrm{off}} + k_{\mathrm{cat}}\right)}$$

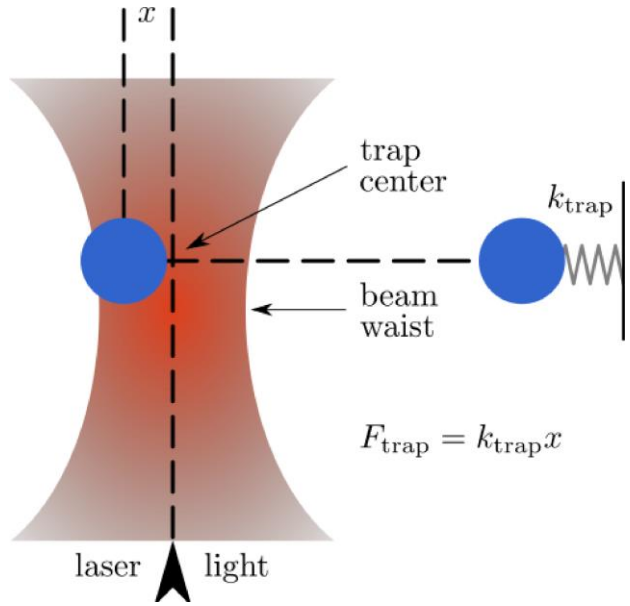
Time dependence in overdamped diffusion

Motion of an overdamped particle obeys the equations:

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left(-\frac{1}{\gamma}\frac{dV(x)}{dx}p(x,t)\right) + \frac{k_B T}{\gamma}\frac{\partial^2}{\partial x^2}p(x,t).$$

$$\frac{dX}{dt} = -\frac{1}{\gamma}\frac{dV(X)}{dX} + \zeta(t),$$

$$\langle \zeta(t)\zeta(t')\rangle = 2D\delta(t-t'), D = \frac{k_B T}{\gamma}.$$

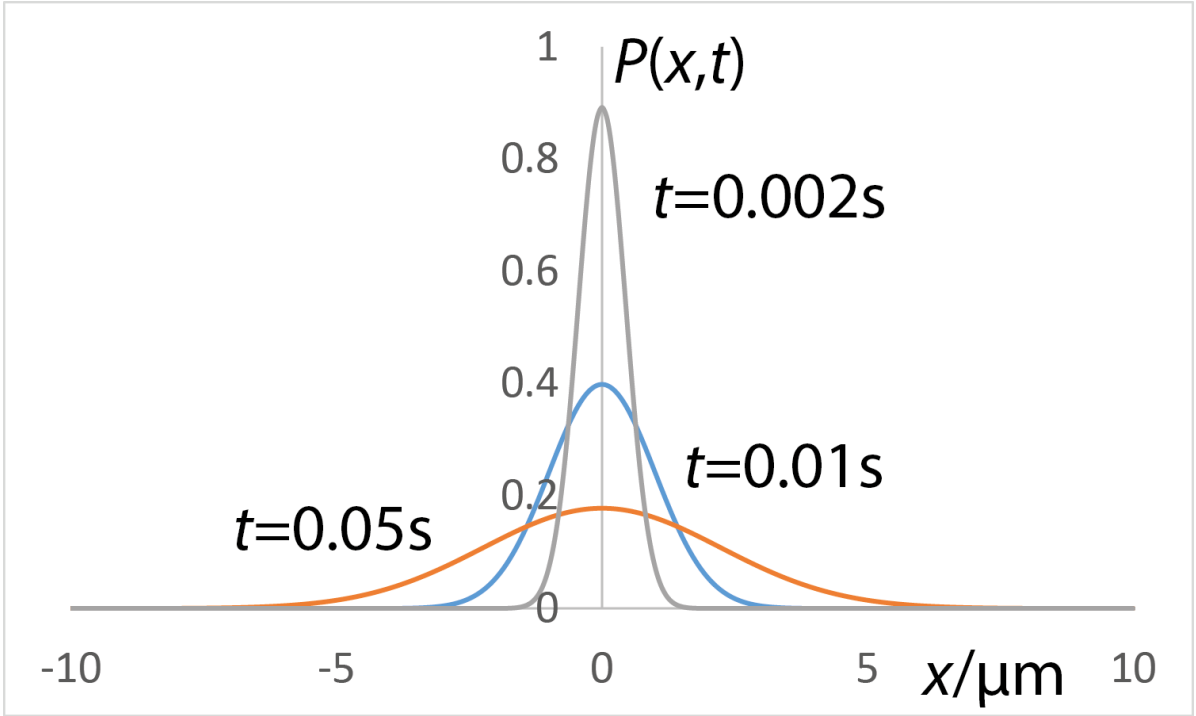


Application

Example: Freely diffusing particle

$$\frac{\partial p(x,t)}{\partial t} = \frac{k_B T}{\gamma}\frac{\partial^2}{\partial x^2}p(x,t) = D\frac{\partial^2}{\partial x^2}p(x,t).$$

$$\pi(x,t) = \sqrt{\frac{1}{4\pi D t}}\,\mathrm{e}^{-\frac{x^2}{4Dt}}.$$



Using $D = 10^{-10}\mathrm{m}^2\mathrm{s}^{-1}$

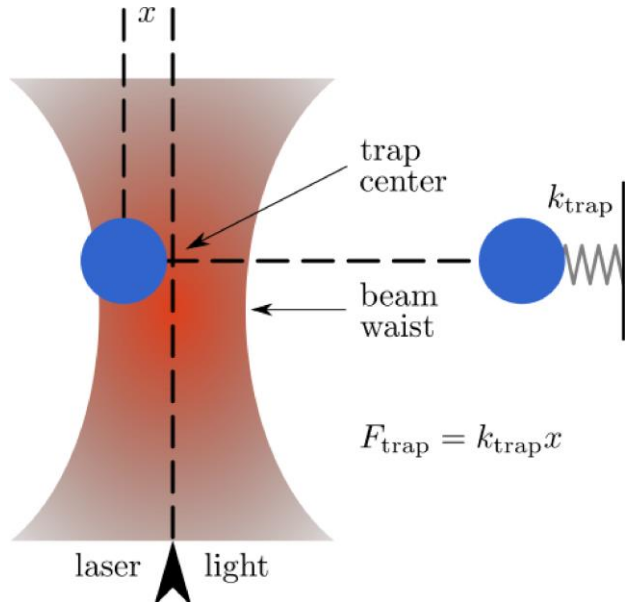
Example: time-dependent correlations in quadratic well

$$\frac{dX}{dt} = -\frac{1}{\gamma}\frac{dV(X)}{dx} + \zeta(t) = -\frac{\kappa}{\gamma}x + \zeta(t),$$

$$\langle \zeta(t)\zeta(t') \rangle = 2D\delta(t-t'), D = \frac{k_B T}{\gamma}.$$

$$X(t) = X(0)\mathrm{e}^{-\frac{\kappa}{\gamma}t} + \mathrm{e}^{-\frac{\kappa}{\gamma}t}\int_0^t dt'\,\zeta(t')\mathrm{e}^{\frac{\kappa}{\gamma}t'}.$$

We can find the average behaviour of X, and the autocorrelation over time.



$$R_X(t,0) = \frac{\mathrm{Cov}(X(t),X(0))}{\sqrt{\mathrm{Var}(X(t))\mathrm{Var}(X(0))}}$$

In steady state, $\langle X(t) \rangle = \langle X(0) \rangle = 0$ and $\mathrm{Var}(X(t)) = \langle X(t)^2 \rangle = \langle X(0)^2 \rangle = k_B T/\kappa$.

Summary

- We introduced biological systems that show interesting stochastic dynamics.
  - Sometimes we're interested in a system out of the stationary distribution.
  - Sometimes we're interested in the timescale over which individual trajectories move around the stationary distribution.

Modelling in Biology II: Stochastic processes and networks

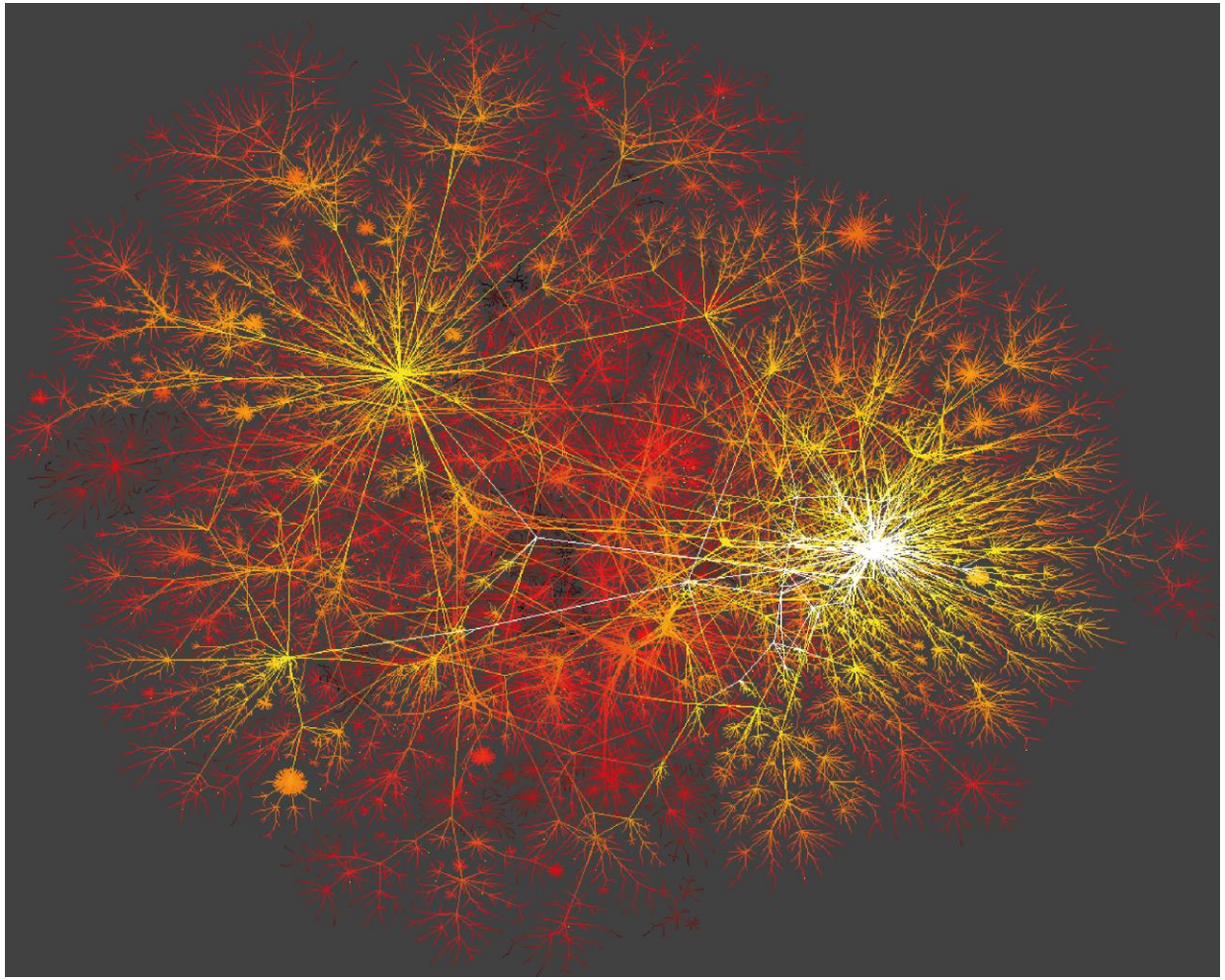
Thomas Ouldridge
t.ouldridge@imperial.ac.uk

Lecture outline

Point processes

Networks

- What is a network?
- Network properties
- A null model of a network
- Networks in biology



Spike trains

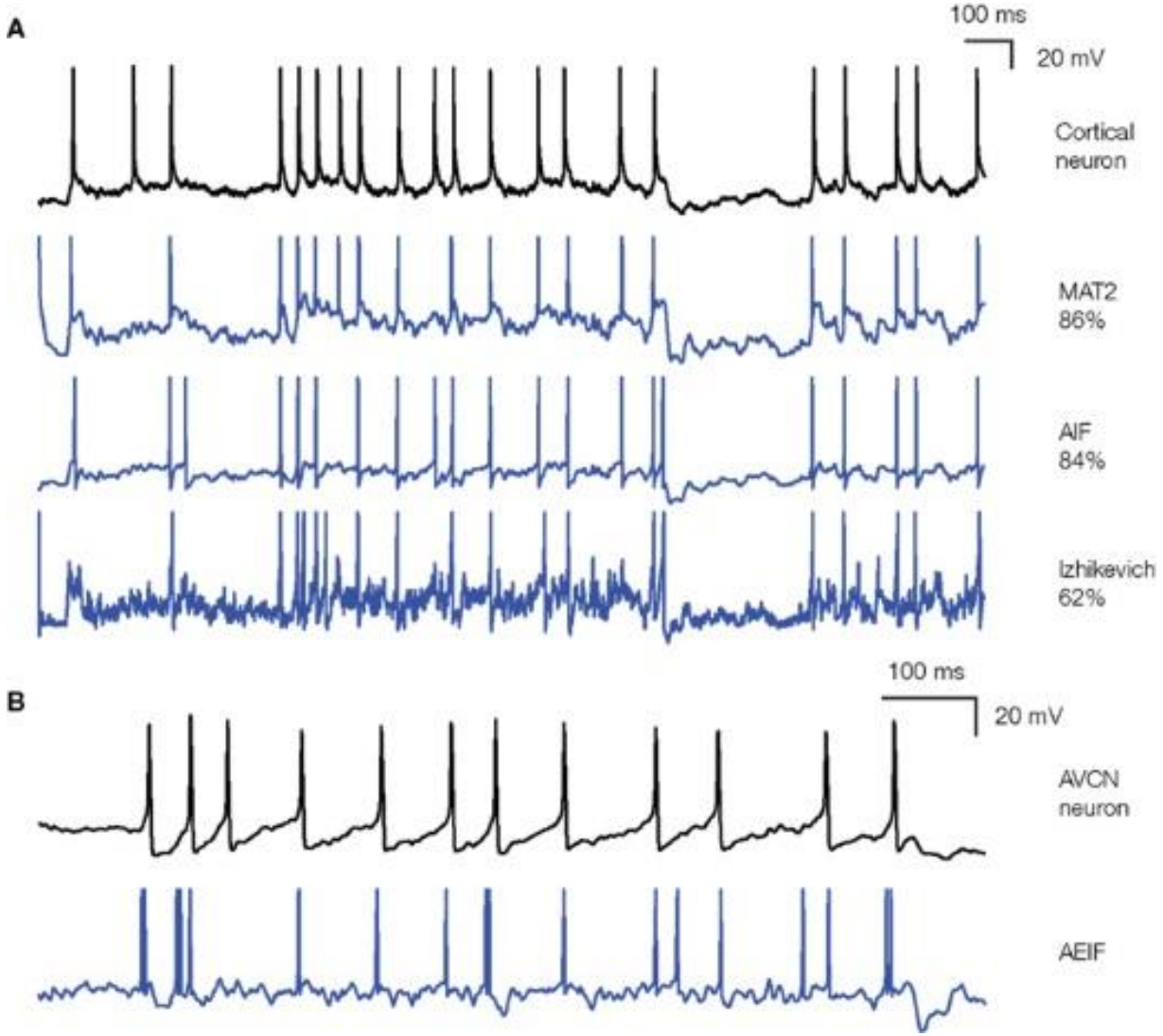
Nerve impulses are an example of a *point process*.

Can represent them as a series of isolated points along a (time) axis.

Gaps between points form a stochastic process $\Theta(n)$.

Simplest possible process: Poisson point process:

- Each gap is independent.
- Spikes arrive at a constant rate ν.



Spike trains: Point Poisson process

Let $p(\theta)$ be the probability density of an interval having length θ.

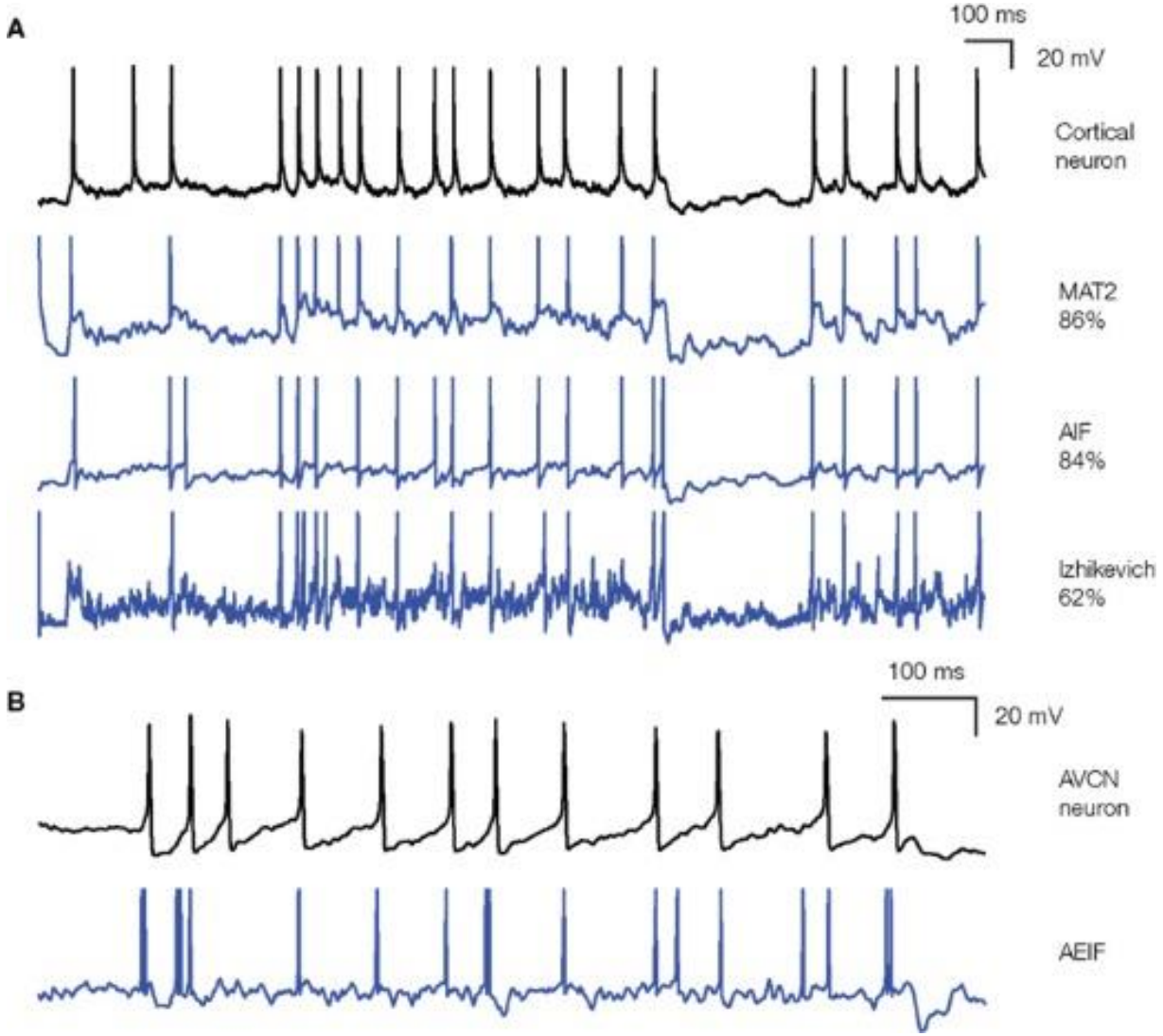
$p(\theta)\delta\theta$ is the probability that nothing happens for a time θ, and then a spike arrives in $\delta\theta$.

Define survival probability $S(\theta)$:

$$\frac{dS(\theta)}{d\theta} = -\nu S(\theta)$$

$$S(\theta) = \mathrm{e}^{-\nu\theta}$$

$$p(\theta) = \nu\mathrm{e}^{-\nu\theta}$$



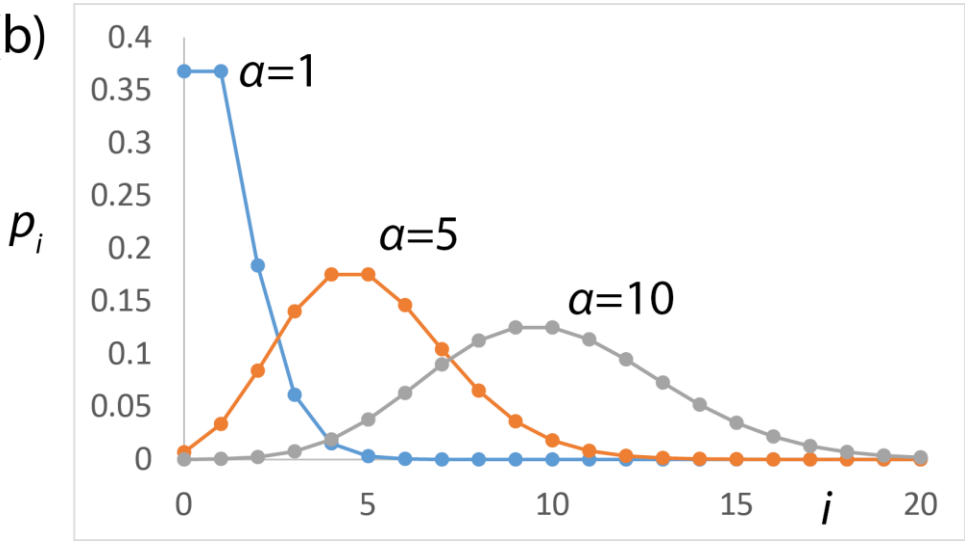
Spike trains: Point Poisson process

How many spikes do we expect during a time window t, $p_n(t, \nu)$?

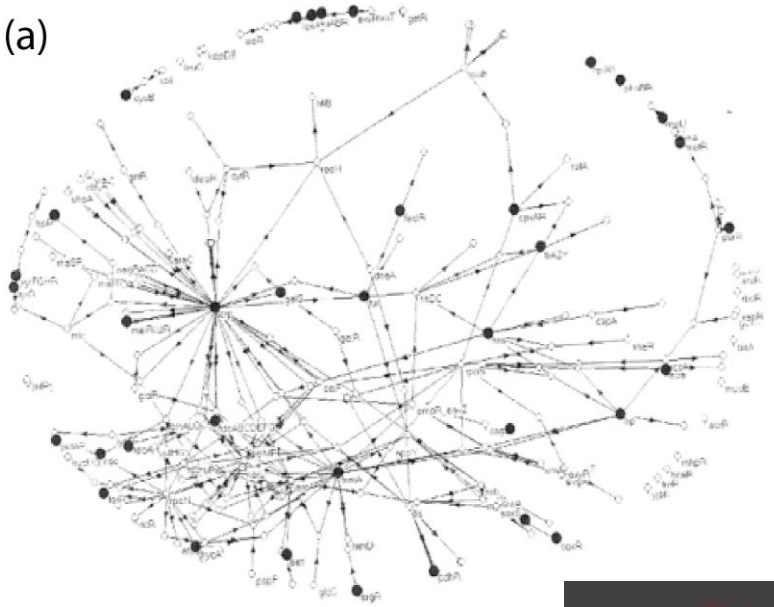
- Probability of no events is $p_0(t) = S(t) = \mathrm{e}^{-\nu t}$.
- Probability of having $n+1$ events is probability of having a first event at $\theta < t$, $p(\theta) = \nu \mathrm{e}^{-\nu\theta}$ and then n more events before t, integrated over all possible θ.
- Guess solution of Poisson with mean νt:
  - Average clearly makes sense.
  - Matches $p_0(t, \nu)$.
  - Matches $p_{n+1}(t, \nu)$ if it matches $p_n(t, \nu)$.

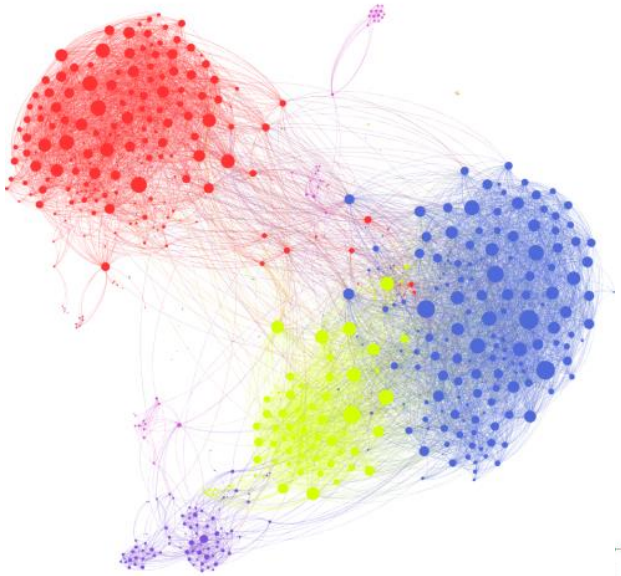
This is enough!

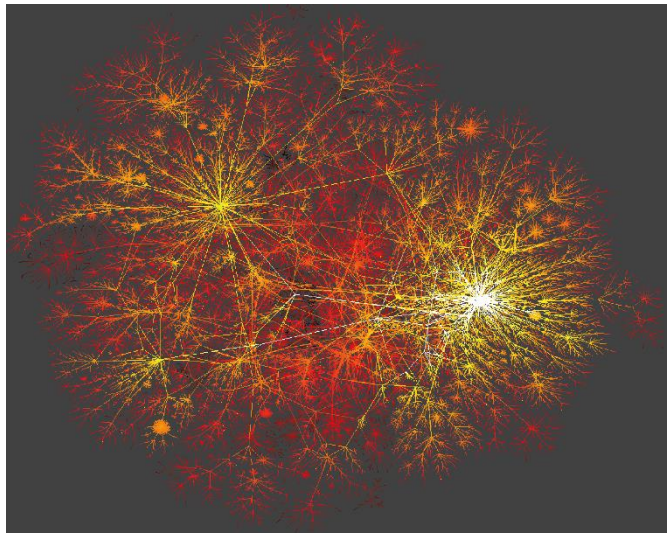
$$p_n(t, \nu) = \frac{(\nu t)^n}{n!} \mathrm{e}^{-\nu t}$$

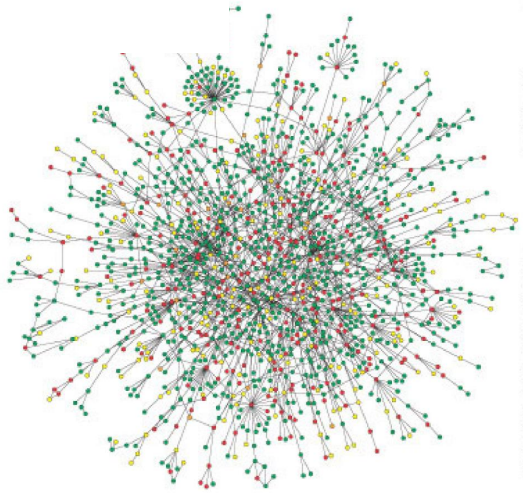


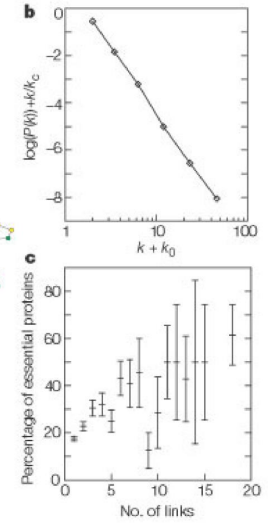
What is a network and why do we care?



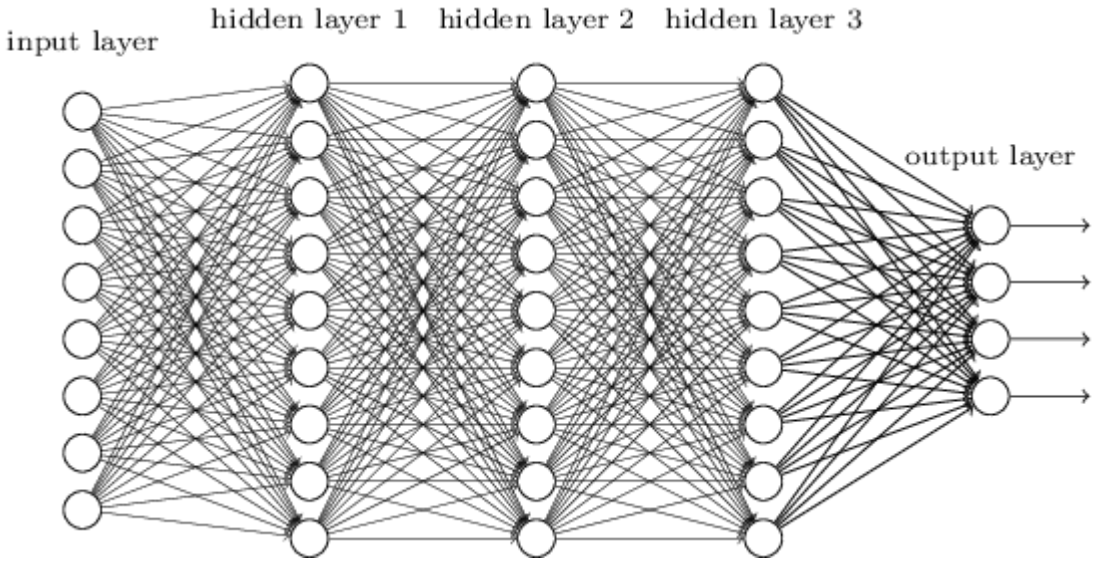








What is a network and why do we care?



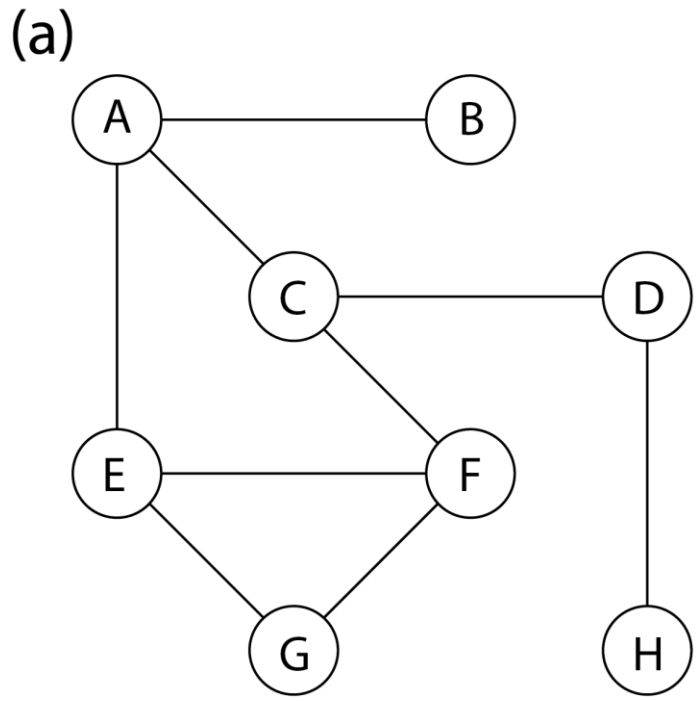
Graphs – a mathematical description

- Nodes/vertices
- Edges/arcs

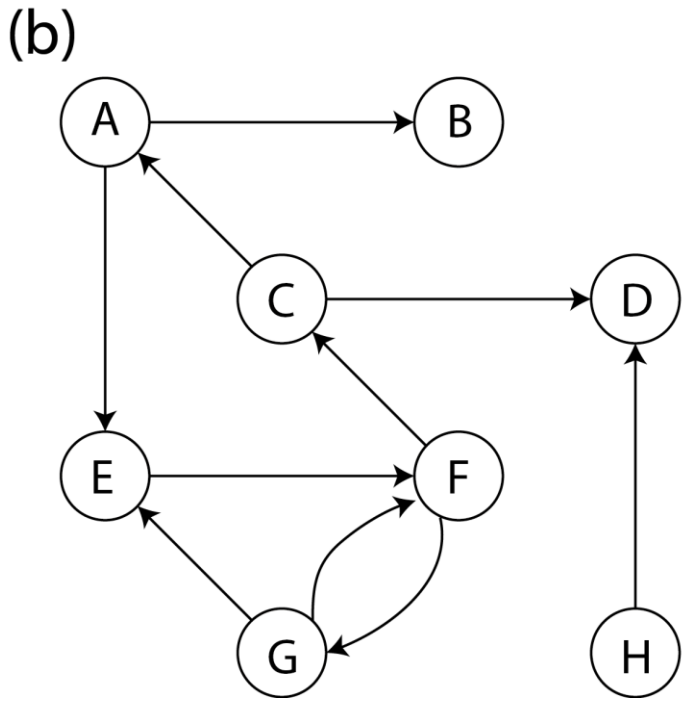
Key Concept

Key Definition

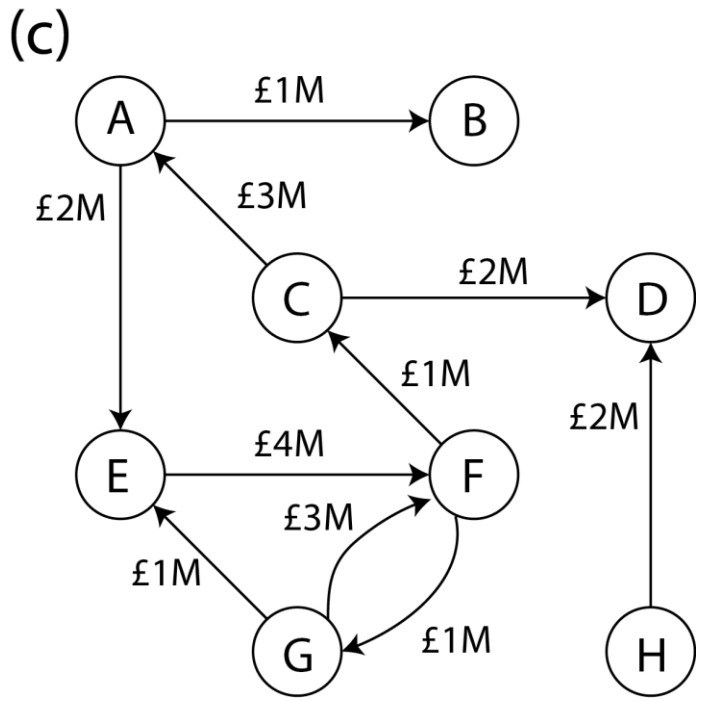
Undirected, unweighted



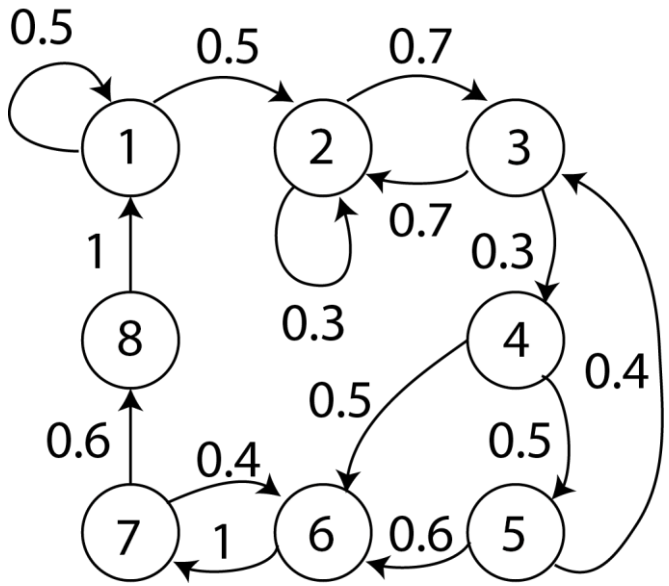
Directed, unweighted



Directed, weighted



Graphs – a mathematical description

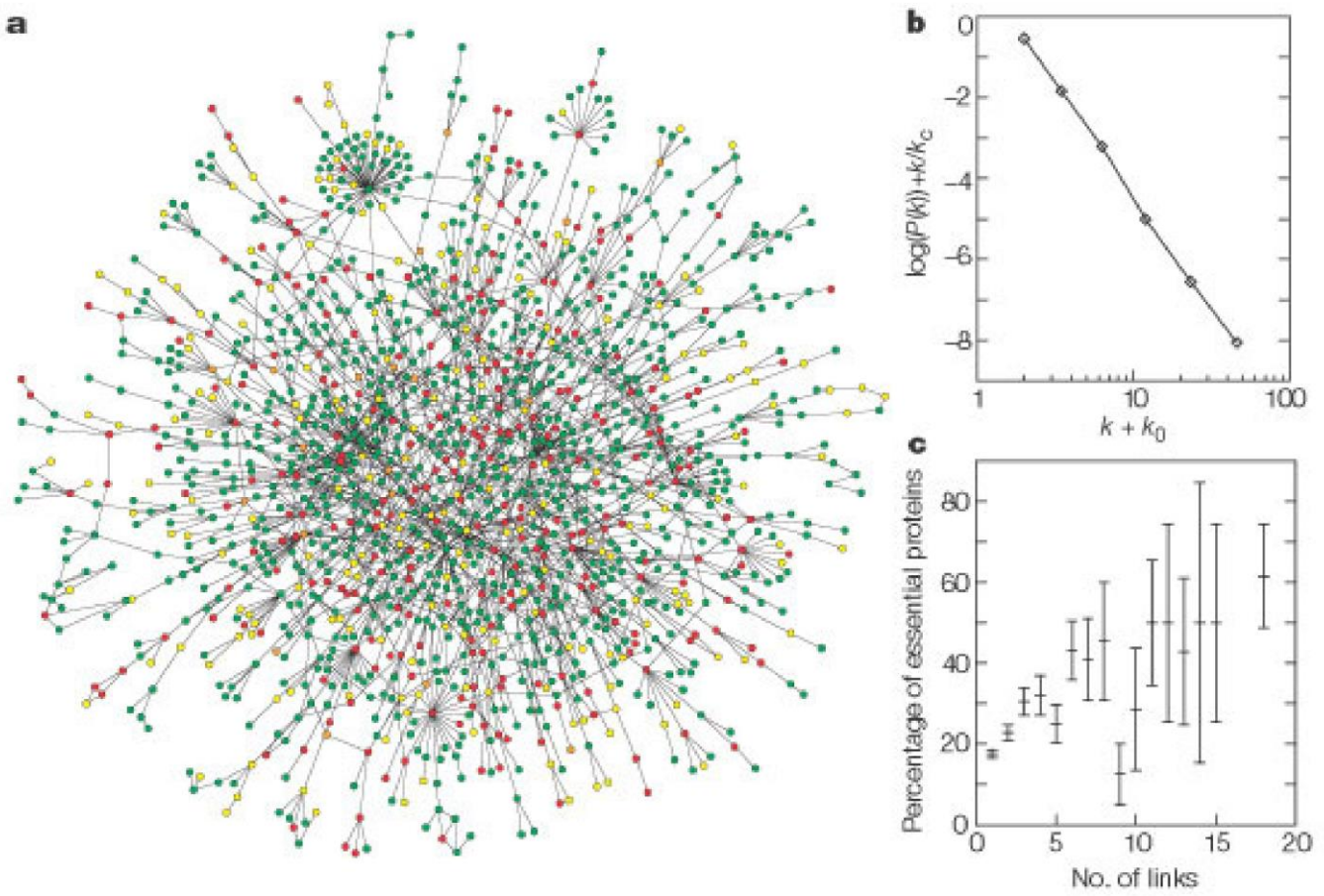


What sort of properties are interesting?

Statistical properties of graphs can be very informative even if we don't know all the details.

- High/low connectivity on average?
- Clustered?
- Hubs?
- Path lengths between nodes?

Network **topology**



Basic properties of graphs

At the level of nodes/pairs of nodes

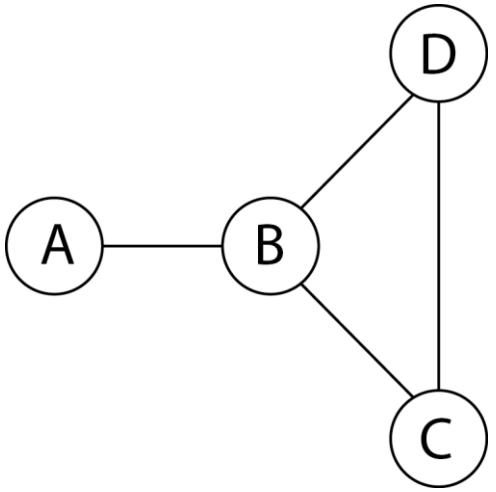
- Degree
- Adjacency of two nodes
- Shortest path length
- Local clustering coefficient:

$$C_i = \frac{2n_i}{k_i(k_i - 1)}$$

Edges between these nodes

Nodes to which i is connected

Key Definition

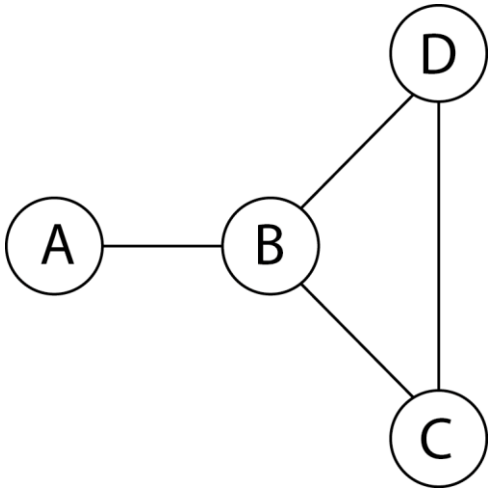


Basic properties of graphs

At the level of the whole graph

- Connectivity
- Average degree
- Adjacency matrix
- Network diameter
- Global clustering coefficient

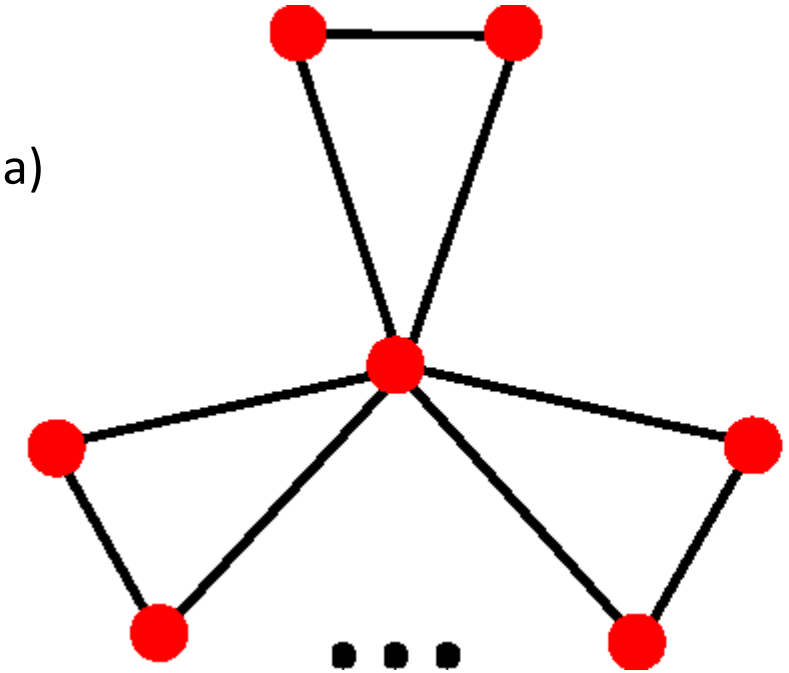
Key Definition



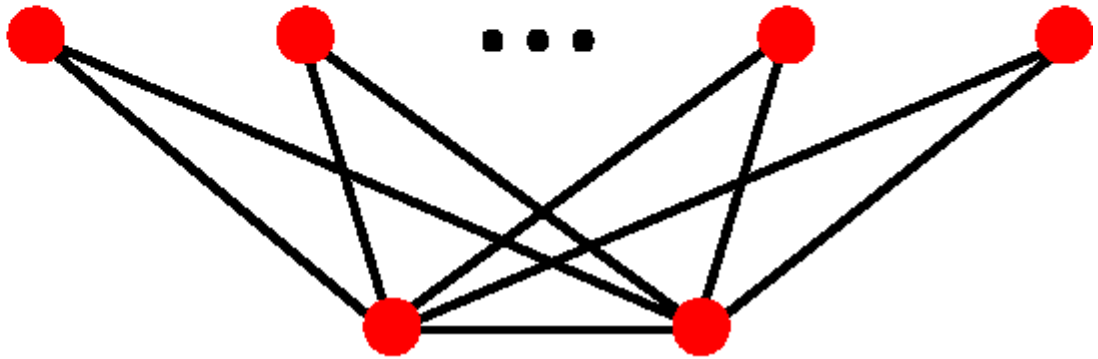
As well as averages, we're often interested in distributions of node properties, eg. $p(k)$.

Conundrum

(a)



(b)



Which one of these graphs has the higher global clustering coefficient?

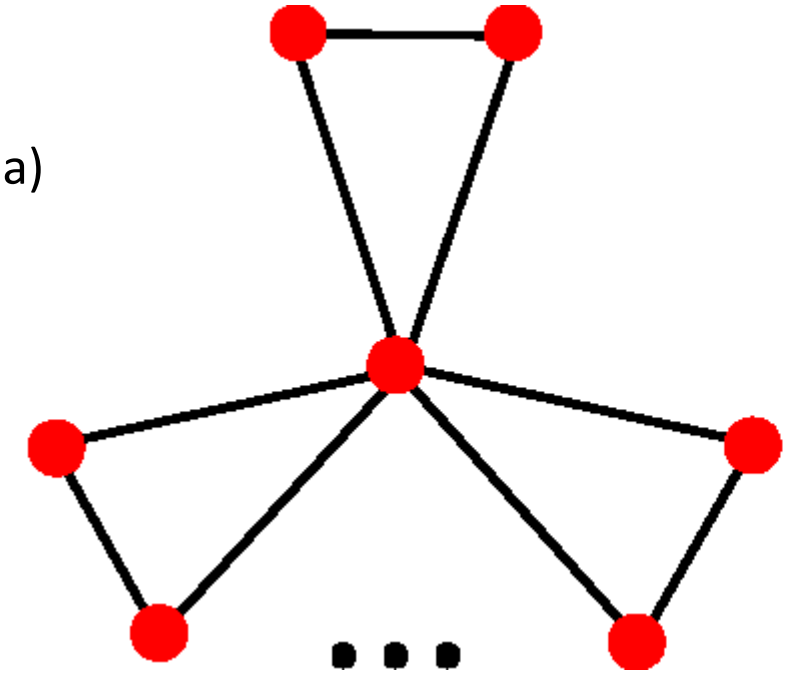
Go to www.menti.com and enter code **99 50 91** to submit your answer/

Modelling in Biology II: Stochastic processes and networks

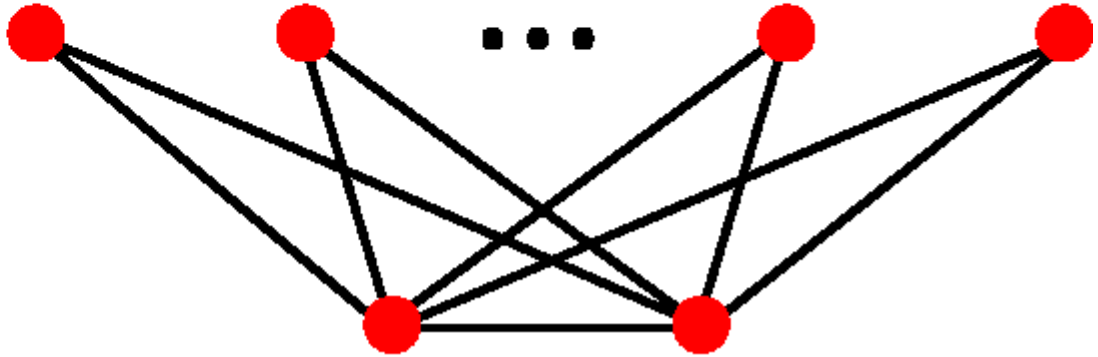
Thomas Ouldridge
t.ouldridge@imperial.ac.uk

Conundrum

(a)



(b)



Which one of these graphs has the higher global clustering coefficient?

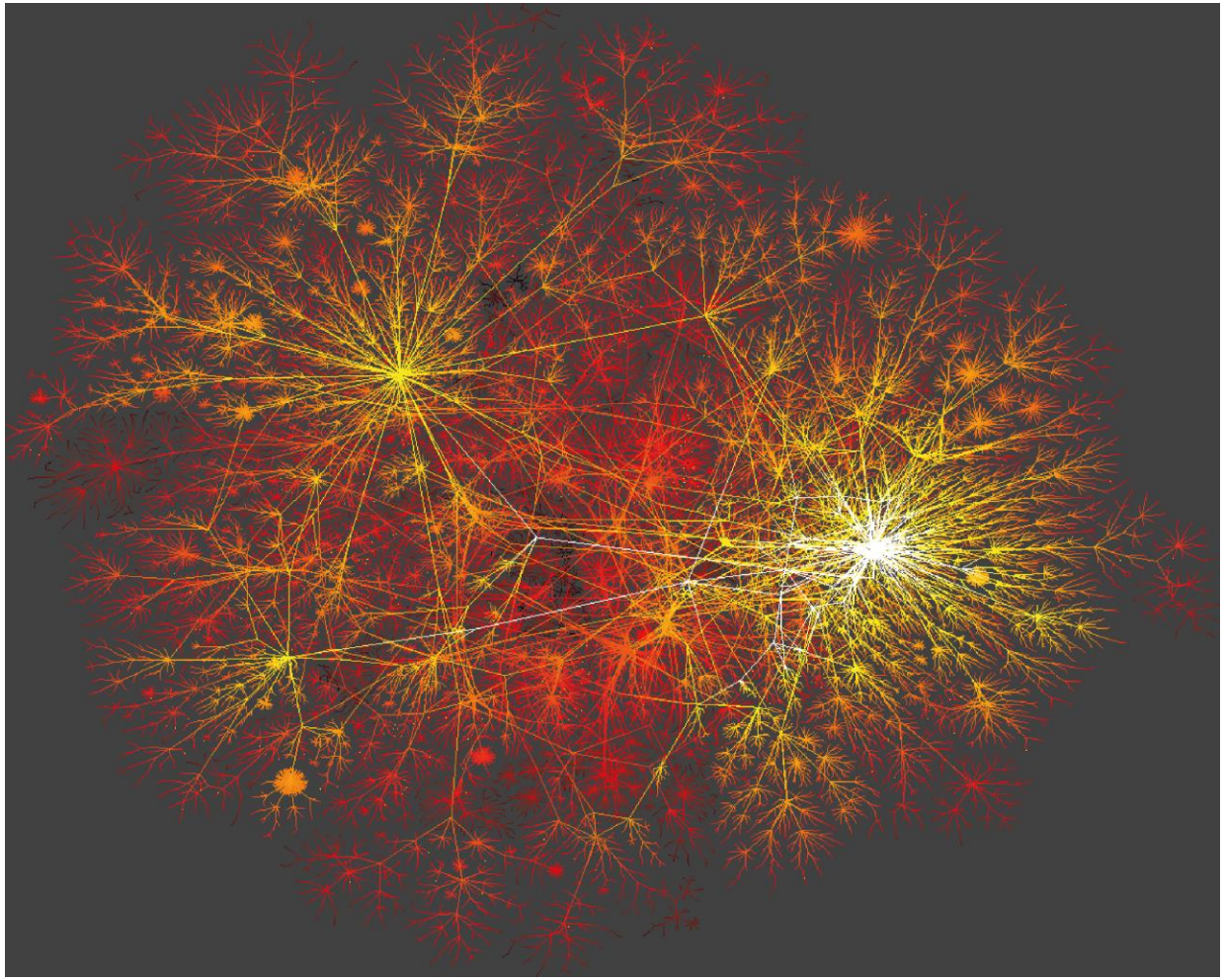
Go to www.menti.com and enter code **99 50 91** to submit your answer/

Lecture outline

Point processes

Networks

- What is a network?
- Network properties
- A null model of a network
- Networks in biology

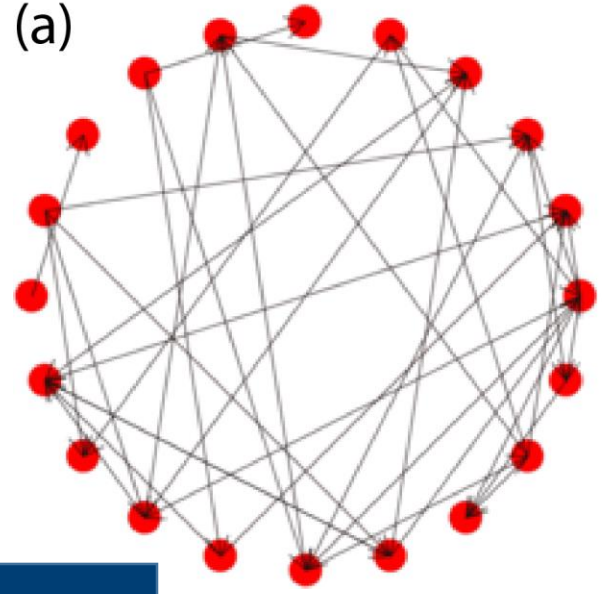


Erdős-Rényi graphs: the basic null model

What does a truly random graph look like, so that we can compare it to real data?

Two (almost equivalent) algorithms for building a random graph: both called Erdős-Rényi graphs.

- Start with N nodes.
- Chose desired average degree $\bar{k}$.
- Consider each possible edge in turn, and include with a probability $p = \bar{k}/(N-1)$.



Key Concept

Key Algorithm

Key Definition

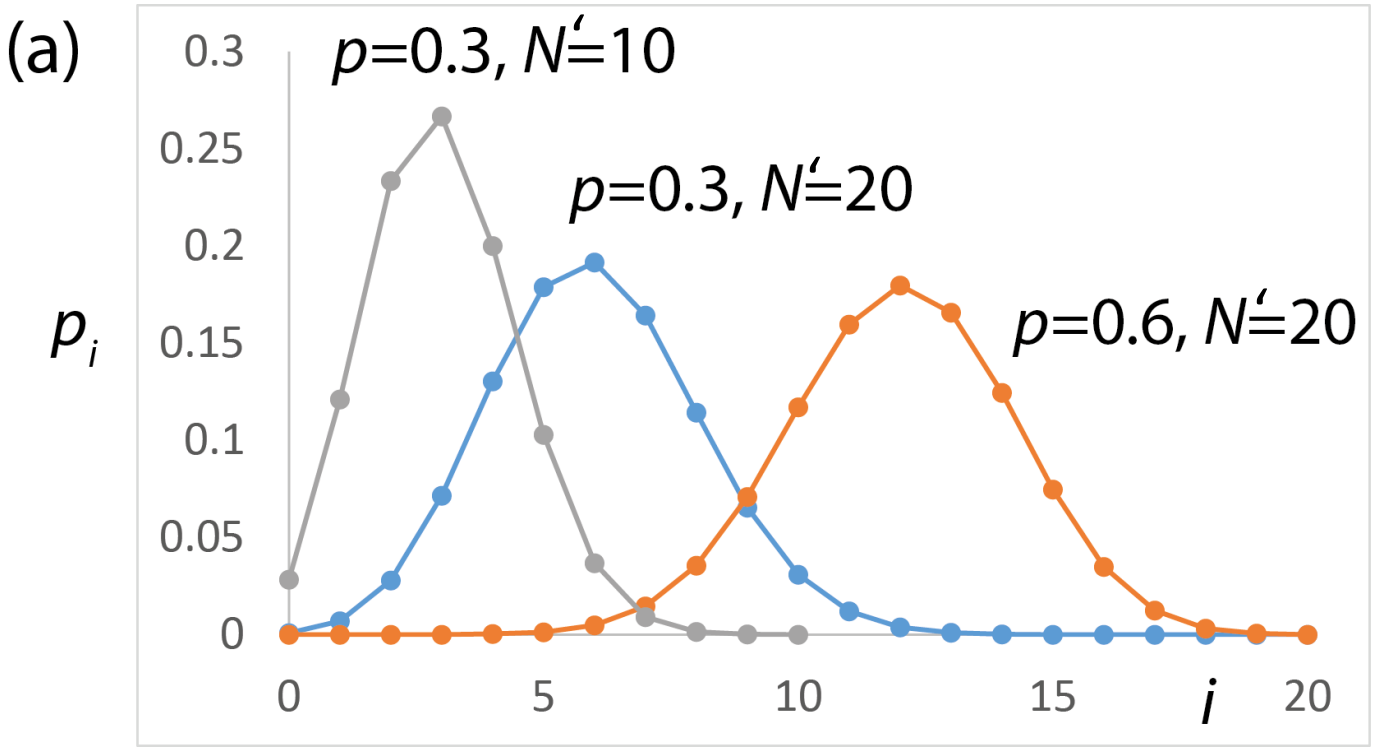
Erdős-Rényi graphs: properties

- Average degree is $\overline{k}$.
- Degree distribution is binomial ($N' = N - 1$):

$$p(k) = p^k (1-p)^{N'-k} \frac{(N')!}{(N'-k)!\,k!}.$$

$$\langle K \rangle = \overline{k} = (N')p$$

$$\mathrm{Var}(K) = (N')p(1-p)$$



Erdős-Rényi graphs: properties

- In the limit $N' \to \infty$, $N'p$ finite:

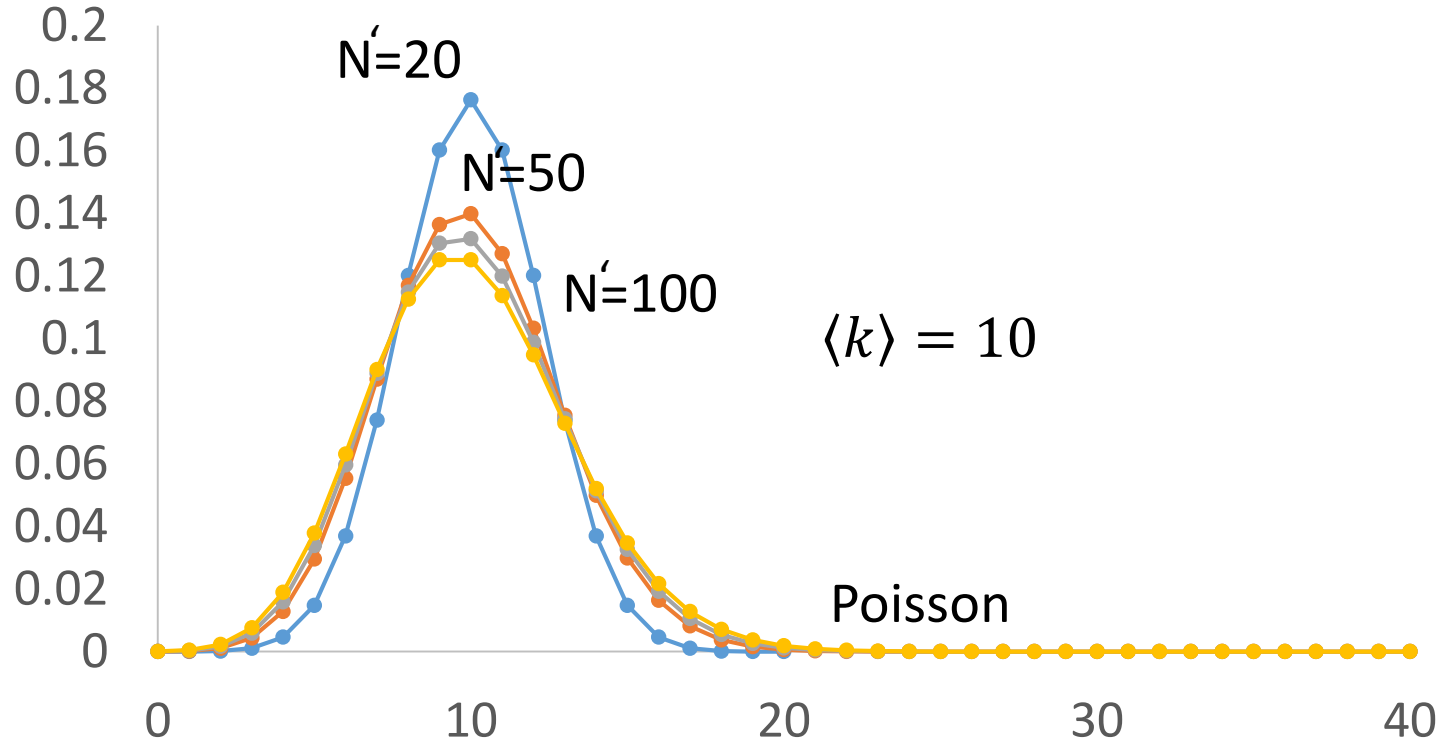
$$\langle K \rangle = N'p$$

$$\mathrm{Var}(K) \approx N'p$$

Support now 0 to ∞.

Well approximated by Poisson!

$$p(k) \approx \frac{(N'p)^k \mathrm{e}^{-N'p}}{k!}$$



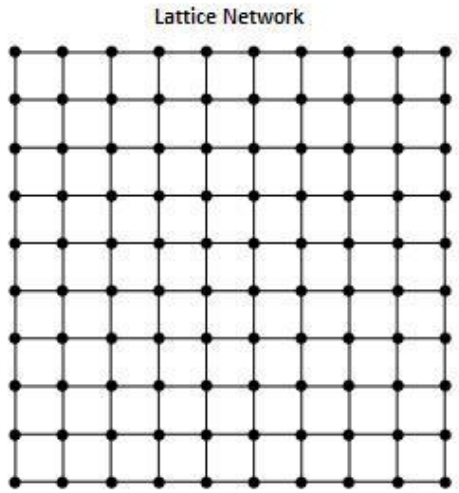
Erdős-Rényi graphs: properties

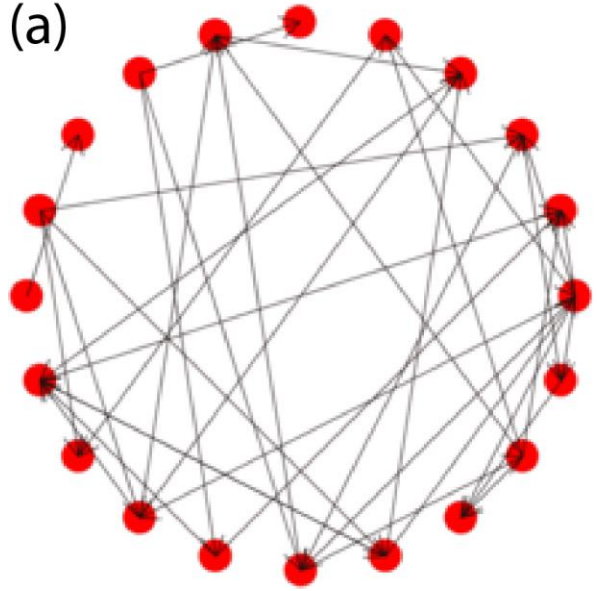
- Clustering coefficient: p.

=> Low clustering in the limit $N' \to \infty$, $N'p$ finite.

Diameter $D \sim \log N$.

=> Easy to get from one node to another.





Erdős-Rényi vs Real World

- Each connection included with a random probability $p=\frac{\langle k \rangle}{N-1} = \frac{\bar{k}}{N-1}$.
- Degree distribution is binomial with a well-defined peak.
- Clustering is low for large N and fixed $\bar{k}$ (sparse matrix).
- Diameter $D \sim \log N$.

Real-world networks? Often end to be:

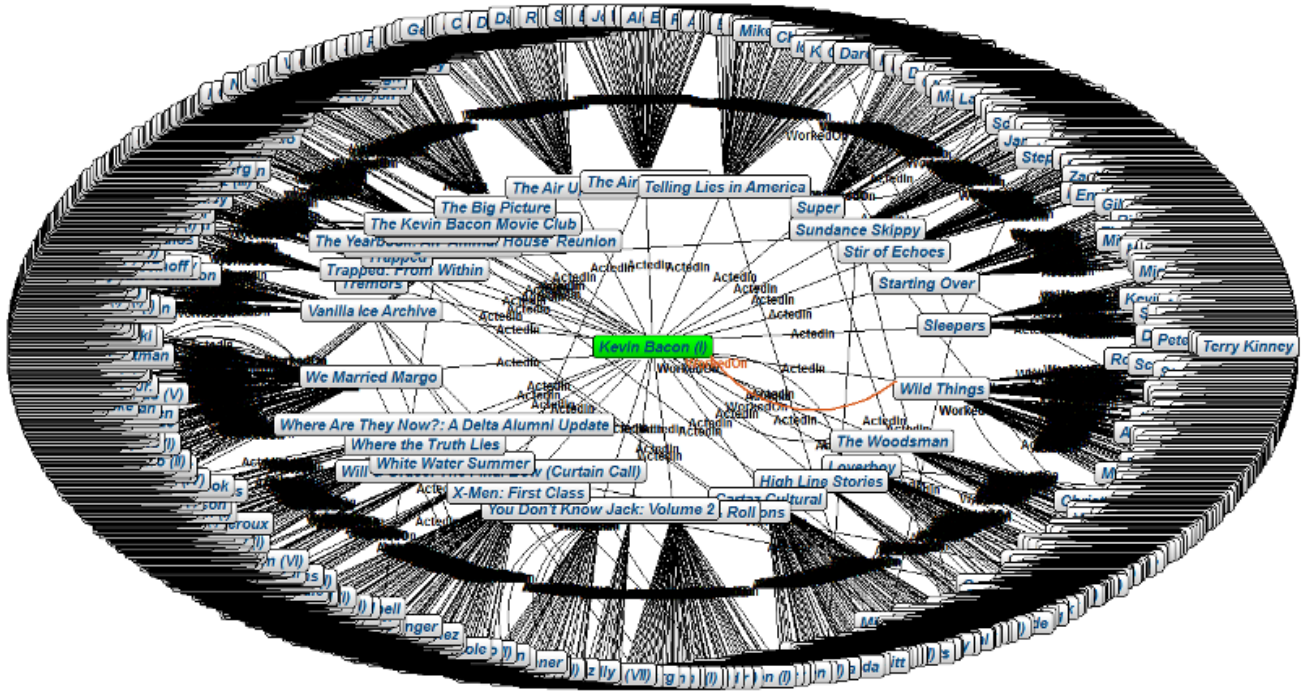
- Sparse
- $D \sim \log N$ (six degrees of separation)
- Much higher clustering.
- Degree distributions look very different.

Key Concept

Small-world networks

- Despite sparseness and high clustering, diameter of (eg. social) networks is surprisingly small. This is the **small world** phenomenon.

- Also seen in eg. Regulatory networks.



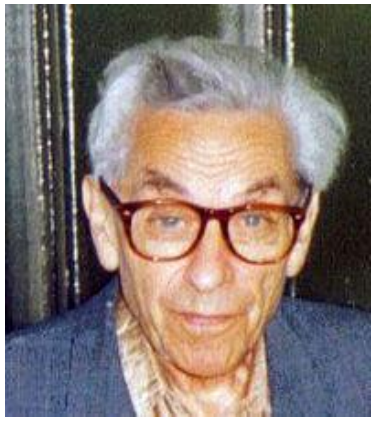
Small-world networks

T. E. Ouldridge

J. P. K. Doye

R. S. Berry

P. Salamon



P. Erdos

T.E. Ouldridge, A.A. Louis and J.P.K. Doye, J. Chem. Phys. 134, 085101 (2011)

J.P.K. Doye, D.J. Wales and R.S. Berry, J. Chem. Phys. 103, 4234-4249 (1995)

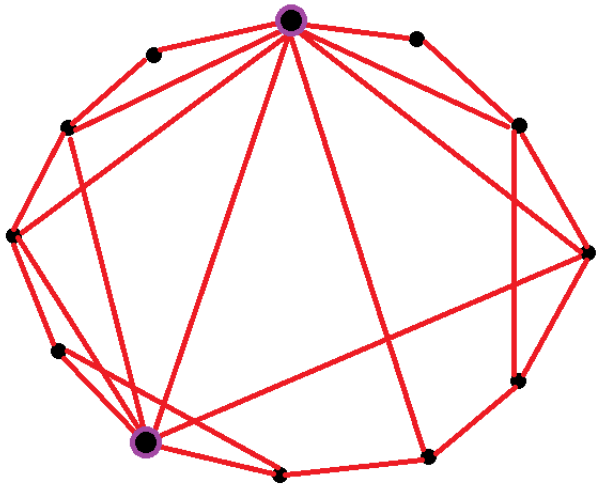
B. Andresen, R. S. Berry, A. Nitzan, and P. Salamon, Phys. Rev. A, 15, 2086-2093 (1977)

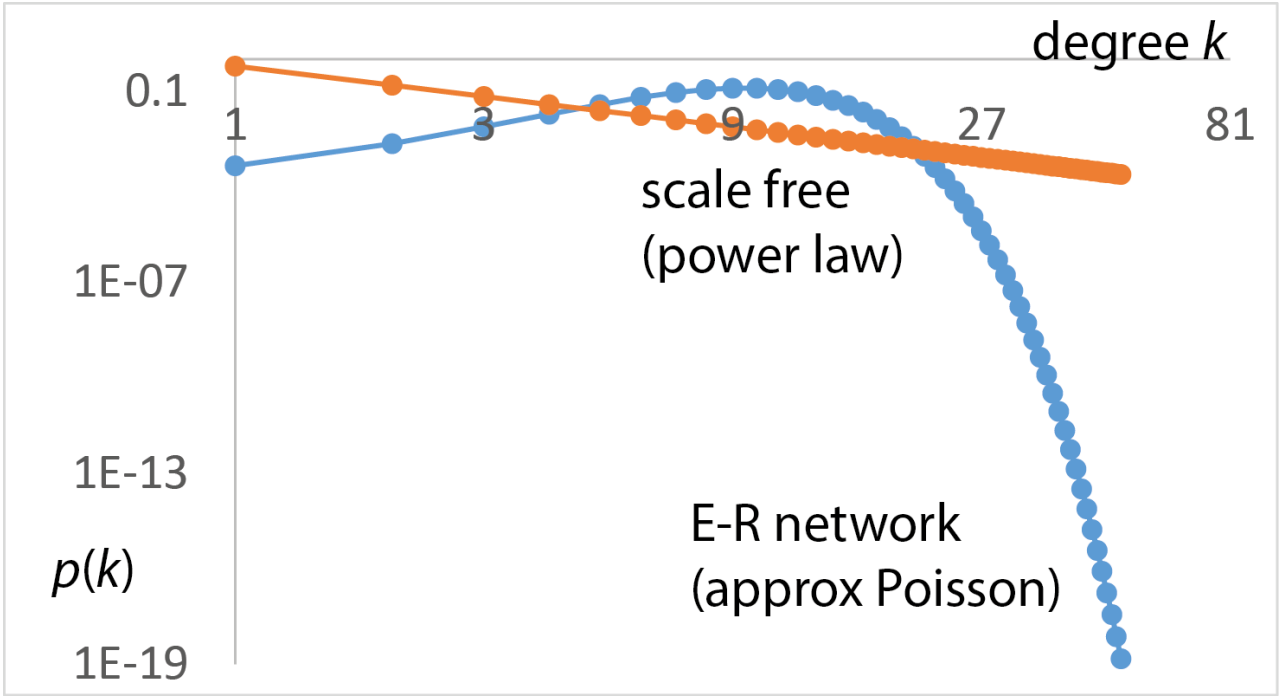
P. Salamon and P. Erdos, Can. Math. Bulletin, 31, 129-138 (1988)

Hubs and degree distributions

Contributing to the small world phenomenon: $P(k)$ much broader than predicted by ER.

We often see more "hubs" with a higher degree of connectivity than expected.



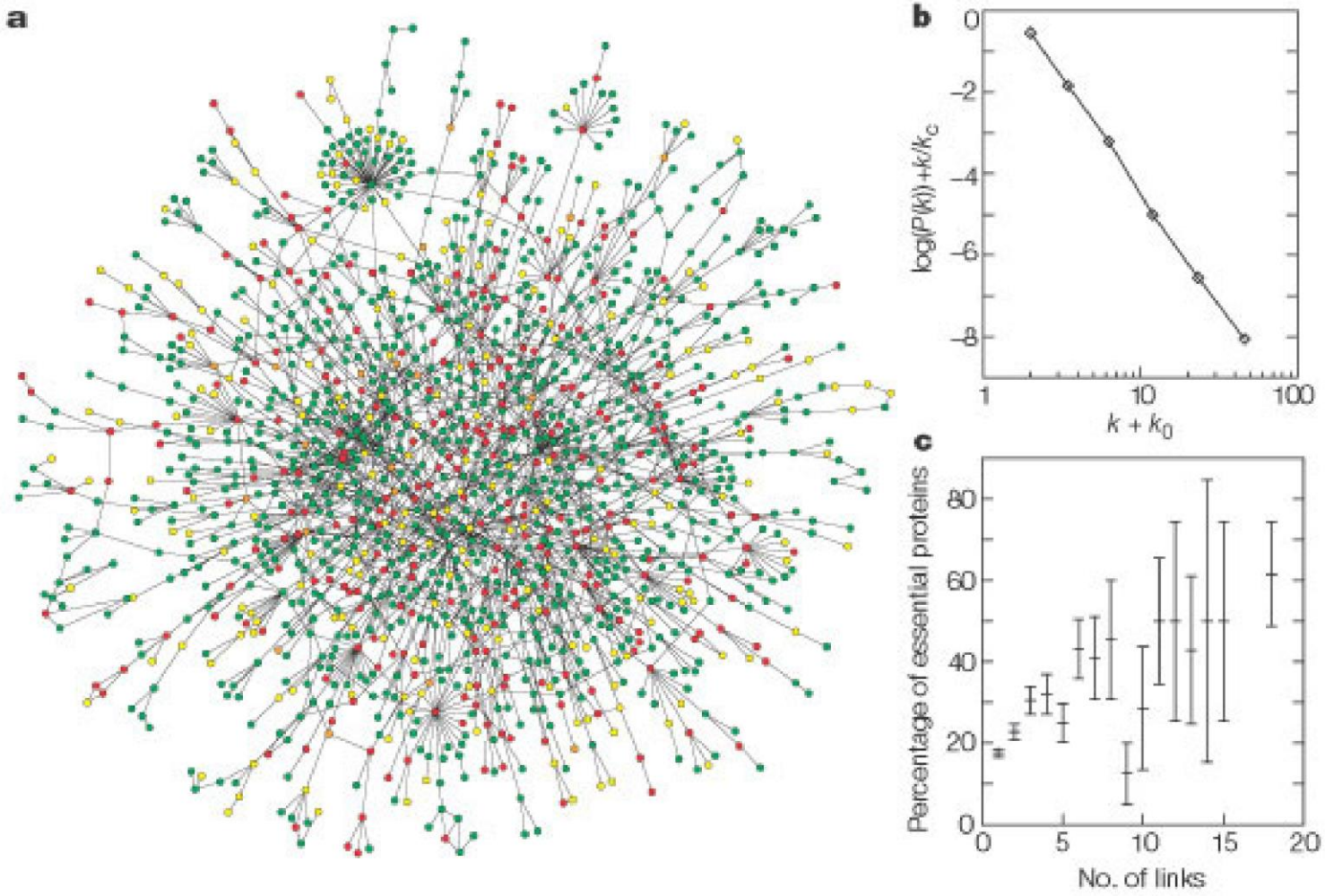


Many networks show "scale free" behaviour.

Key Definition

$$P(k) \sim k^{\alpha}$$

Hubs and degree distributions



Generating a scale free network with the Barbarási-Albert model

- Start with an initial network with at least one edge.
- Add new nodes one at a time; connect to m existing nodes. Chose nodes with a probability proportional to number of connections.
- Heavily-linked nodes get more connections; gives you $P(k) \sim k^{-3}$.

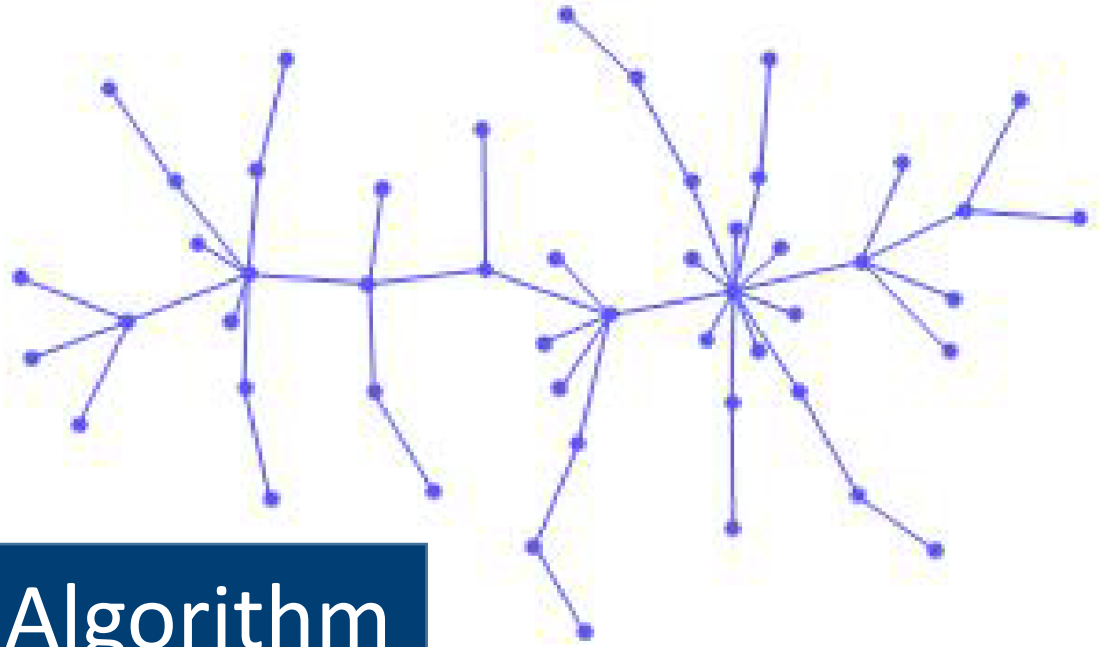
Does this give a mechanistic explanation of why scale-free networks emerge?

Key Algorithm

Generating a scale free network with the Barbarási-Albert model

- Start with an initial network with at least one edge.
- Add new nodes one at a time; connect to m existing nodes. Chose nodes with a probability proportional to number of connections.
- Heavily-linked nodes get more connections; gives you $P(k) \sim k^{-3}$.

Does this give a mechanistic explanation of why scale-free networks emerge?



Key Algorithm

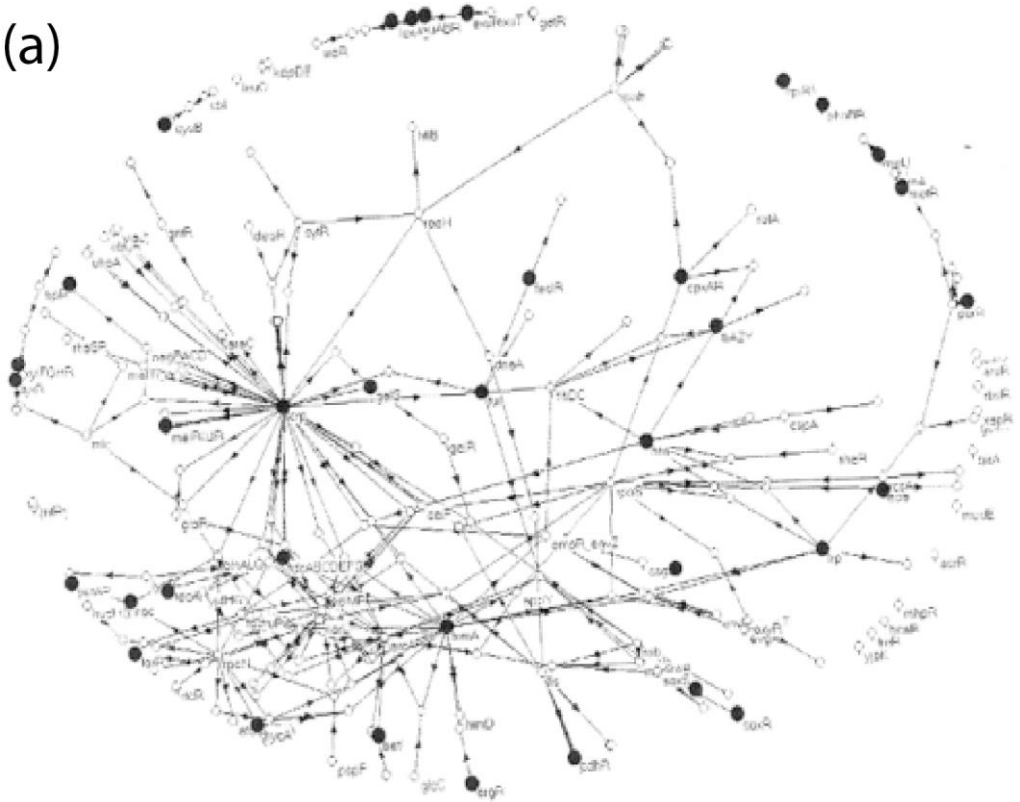
No clustering – we'll be exploring the importance of clustering for a process on a network in the practical.

Identifying over-represented motifs

Transcription network of *E. coli*.:

- Proteins that influence transcription of each other.
- Directed.
- Self links possible (bold dots).

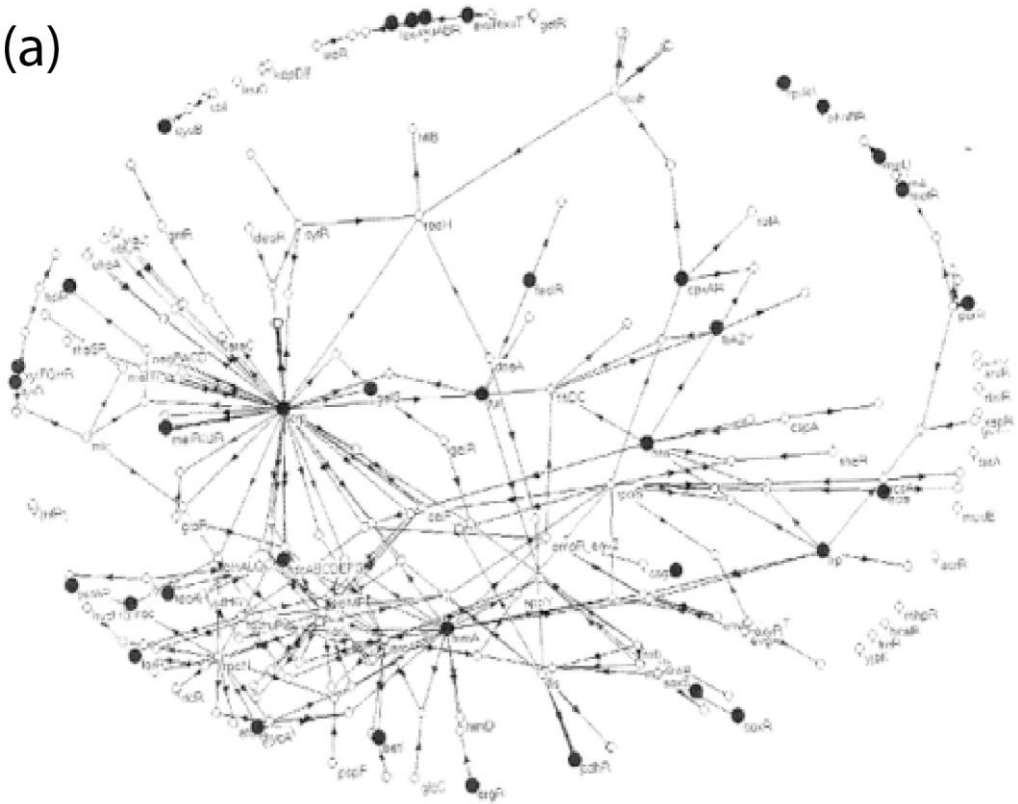
The network is functional. Are some subgraphs helpful in achieving function? We'd expect these **motifs** to be more numerous than in a random graph.



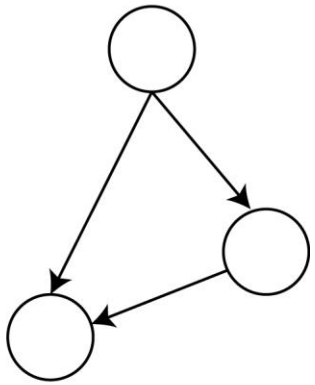
Application

Identifying over-represented motifs

(a)



(b)



There are 420 nodes and 520 edges in this graph.

Would we expect to see 42 feed-forward loops in an ER graph with the same size and average degree?

We would expect $\sim \bar{k}_{\text{out}}^3 \approx 2$.

Highly improbable that these have occurred by accident! Useful/easy to evolve?

Key Technique

Summary

- Networks arise when we have multiple entities interacting in a complex way.
- We see them on subcellular, cellular and organismal level.
- Even if we don’t know everything, statistical properties can tell us a lot about function and the underlying mechanism by which they were created.
- The Erdős-Rényi network is a null, random network to compare real-life networks to.
- Real life networks are often more clustered, with a very different degree distribution and an over-representation of “motifs”.
- Still often have small diameters: the small world phenomenon.