STATISTICAL PHYSICS OF BIOLOGICAL SYSTEMS

(BIOLÓGIAI RENDSZEREK STATISZTIKUS FIZIKÁJA:

- Skálázás és fluktuációk az élővilágban

- Kollektiv viselkedés)

We shall consider systems with:

- many similar biological "units"

- in the presence of "noise"

We shall look for collective behaviour (with scaling and fluctuations):

- tansitions in global parameters

- Types: pattern formation, group motion, motion due to fluctuations, network formation, synchronization

Theoretical biology as such

Models, simulations, understanding

(versus equations with their analytical solutions)

Are fluctuations an important, inherent ingredient of life? What is their origin and impact?

Example. Animal: i) similar/different outlook, reactions, etc, ii) regular heartbeat with randomness, iii) random series of electric signals, iv) similar/different behaviour on the cells level, etc.

The distribution of fluctuations, sizes, distances, etc frequently follow a power law (scale).

In short, if a system is made of many interacting units, specific statistical features involving fluctuations and scaling emerge.

I)

Microscopic objects are subject to the so called thermal fluctuations. It is a fundamental feature of all systems that if they have a well defined temperature T, than each microscopic particle (atom or molecule) in them has in average an amount of kinetic energy $\frac{1}{2}kT$.

II)

When many similar, but not necessarily microscopic objects (biological or non-living) are present in a system there are further reasons to consider the random aspect of the behaviour (many deterministically interacting particles $->$ can be looked as random).

III)

Non-linearities are known to lead to a very complex behaviour which – especially in the presence of thermal fluctuations – can be considered as random.

Fluctuations

Noise versus fluctuations

Typically, noise is not correlated.

$$c(\mathbf{r},t) = \langle F(\mathbf{r},t)F(\mathbf{r}',t')\rangle - \langle F(\mathbf{r},t)\rangle\langle F(\mathbf{r},t)\rangle = C\delta(\mathbf{r}-\mathbf{r}',t-t')$$

where the δ function is equal to zero for any non-zero values of its arguments and C is some constant. The averaging (denoted by $\langle ... \rangle$ is made over all values of the arguments.

Fluctuations can be more complex than just white noise.

Noise and fluctuations play a central role in ordering and microscopic transport phenomena.

Scaling

$F(x)$ scales $F' = F(Ax) = g(A)F(x)$. For more arguments, analogously. This is trivially so for power law, but is not, as a rule, true for other functions. For example, $F = x^2$ scales because $F' = F(Ax) = A^2x^2 = A^2F(x)$, while $F = \log(x + B)$ does not scale since $F' = F(Ax + B) = \log(Ax + B)$ cannot be reduced.

In particular, during *second order phase transitions* the so called "critical state" (or critical phenomena) can be observed with a (power law) dependence of the relevant quantities on their parameters. A *power law dependence* of the quantity $n(s)$ (e.g., the number of schools containing s fish) is of the following form $n(s) \sim s^{-\tau}$, where $\sim$ expresses proportionality, and τ is some exponent. The power law dependence is very special.

Why are such states called critical? (sensitivity)

The important point is that scaling typically involves *universality*

Why does scaling occur during continuous transitions?

Gradual change from one state into another one without imposed symmetry breaking (left to right, checkerboard, scaling).

Role of models

particles

clusters or *aggregates*, *lattice*.

FRACTAL GEOMETRY

For such an object it can be shown that the number of particles in a circle of radius R scales as

$$N(R) \sim R^D, \tag{1}$$

where $D < d$ is a non-integer number called the *fractal dimension*. Clusters having a non-trivial D are typically *self-similar*.

Obviously, the volume of a fractal (or any object), $V(l)$, can be measured by covering it with d dimensional balls of radius l. Then the expression

$$V(l) = N(l)l^d \tag{2}$$

gives an estimate of the volume, where $N(l)$ is the *number of balls needed to cover the object completely* and l is much smaller than the linear size L of the whole structure.

Box counting versus sandbox

Definition for a biological fractal is the requirement that a power law scaling of $N(l)$ has to hold over *at least two orders of magnitude*.

Definitions

The fact that an object is a mathematical fractal then means that $N(l)$ diverges as $l \to 0$ or $L \to \infty$, respectively, according to a non-integer exponent.

Correspondingly, for fractals having a finite size and infinitely small ramifications we have

$$N(l) \sim l^{-D} \tag{3}$$

with

$$D = \lim_{l \to 0} \frac{\ln N(l)}{\ln(1/l)}, \tag{4}$$

while

$$N(L) \sim L^D \tag{5}$$

and

$$D = \lim_{L \to \infty} \frac{\ln N(L)}{\ln(L)}. \tag{6}$$

for the *growing case*, where $l = 1$.

Now we are in the position to calculate the dimension of the objects.

$$N(L) = 5^k \qquad \text{with} \qquad L = 3^k, \tag{7}$$

where k is the number of iterations completed. $D = \ln 5/\ln 3 = 1.465...$

Useful rules

a) *projection* of a fractal. $D_p = D$. For $D \geq d_s$ the projection fills the surface, $D_p = d_s$.

b) The *union* of two fractal sets A and B with $D_A > D_B$ has the dimension $D = D_A$.

c) The fractal dimension of the *intersection* of two fractals with D_A and D_B is given by

$$D_{A\cap B} = D_A + D_B - d$$

. The density of A and B particles is respectively proportional to L^{D_A}/L^d and L^{D_B}/L^d. The number of overlapping sites $N \sim L^{D_{A\cap B}}$ is proportional to these densities and to the volume of the box which leads to the above given relation.

Deterministic versus stochastic self-similarity.

For stochastic factals it is more effective to calculate the so called *density-density or pair correlation function*

$$c(\vec{\mathbf{r}}) = \frac{1}{V}\sum_{\vec{\mathbf{r}}'} \rho(\vec{\mathbf{r}} + \vec{\mathbf{r}}')\ \rho(\vec{\mathbf{r}}') \tag{8}$$

ρ is the local density, i.e., $\rho(\vec{\mathbf{r}}) = 1$ if the point $\vec{\mathbf{r}}$ belongs to the object, otherwise it is equal to zero. Ordinary fractals are typically isotropic so that $c(\vec{\mathbf{r}}) = c(r)$.

An object is non-trivially scale invariant if

$$c(br) \sim b^{-\alpha} c(r) \tag{9}$$

with α a non-integer number larger than zero and less than d. the only function which satisfies is

$$c(r) \sim r^{-\alpha} \tag{10}$$

The number of particles $N(L)$ within a sphere of radius L from their density distribution

$$N(L) \sim \int_0^L c(r) d^d r \sim L^{d-\alpha}, \tag{11}$$

$$D = d - \alpha \tag{12}$$

Self-similar and self-affine fractals

Random walks. Trace of a walker is isotropic, self-similar. Mean squared distance $R^2 = \langle R^2(t) \rangle$. $R^2 \sim t$ independently of d. This expression is equivalent to

$$N(R) \sim R^2 \tag{13}$$

for random walks $D = 2 < d$ if $d > 2$.

Fractals invariant under anisotropic rescaling are called self-affine

Single-valued, nowhere-differentiable functions. If such a function $F(x)$ has the property

$$F(x) \simeq b^{-H} F(bx) \tag{14}$$

it is self-affine, where $1 > H > 0$ is some exponent.

A definition of self-affinity equivalent to this is given by the expression for the *height correlation function* $c(\Delta x)$

$$c(\Delta x) = \langle [F(x + \Delta x) - F(x)]^2 \rangle \sim \Delta x^{2H} \tag{15}$$

A deterministic self-affine model.

A random function is, for example, the plot of the distances $X(t)$ measured from the origin as a function of time t, of a Brownian particle diffusing in one dimension.

$$\langle X_H^2(t) \rangle \sim t^{2H}, \qquad with \quad H = 1/2$$

.

Methods for determining fractal dimensions

Evaluation of numerical data

In general, such discrete sets of numbers are obtained by two main methods: i) by digitizing pictures ii) by numerical procedures used for the simulation of various biological structures.

Determine the number of particles $N(R) = R^D$ within a region of linear size R and obtain D from the slopes of the plots $\ln N(R)$ versus $\ln R$.

The roughness exponent H corresponding to *self-affine* fractals is usually determined from the definition. An alternative method is to investigate the scaling of the standard

deviation $\sigma(l) = [\langle F^2(x)\rangle_x - \langle F(x)\rangle_x^2]^{1/2}$ of the self-affine function F

$$\langle \sigma(l) \rangle \sim l^H, \tag{16}$$

where the left hand side is the average of the standard deviation of the function F calculated for regions of linear size l.

Even better: Detrended fluctuations analysis

SCALING IN CONTINUOUS PHASE TRANSITIONS

Example: a magnetic material at temperature T and magnetic field H. If this system shows a paramagnetic–ferromagnetic second order transition at T_c (for $H = 0$) then it is convenient to use the *reduced temperature*

$$t = \frac{T - T_\mathrm{c}}{T_\mathrm{c}} \tag{17}$$

as control parameter instead of T.

The order parameter in this case is the zero-field ($H = 0$) magnetisation M_0 since it is zero if $T > T_\mathrm{c}$, and non-zero below T_c. In the vicinity of T_c it behaves as

$$M_0(t) \sim |t|^\beta, \tag{18}$$

where β is the *critical exponent*. Similarly, for the specific heat (which gives the change of energy for a small change of temperature)

$$C_{H=0}(t) \sim |t|^{-\alpha}, \tag{19}$$

and the susceptibility (which is the sensitivity of the magnetisation with respect to the external field H) scales as

$$\chi(T) \sim |t|^{-\gamma}. \tag{20}$$

Describe a response of the system to some external perturbation. Close to T_c they diverge showing that there the system is extremely sensitive: it is in a critical state.

The typical size of fluid "droplets" or the so called correlation length, also scales

$$\xi \sim t^{-\nu}, \tag{21}$$

at the critical point there will be no typical length scale.

Can be derived from mean-field approximation.

PERCOLATION

Prototype of inhomogeneous media or random clusters in space

- Random occupation in space (lattice or off lattice). r_i is a number randomly selected between 0 and 1 for site i. If $r < p$ fill the site.
- control parameter: p (fraction of sites occupied)
- percolation transition at p_c. Transition is sharp, continuous. Scaling of the order parameter, $P(p)$ which is the relative weight of the "infinite" or percolating cluster

$$P(p) \sim (p - p_c)^{\beta}$$

p_c, typically > 0 is called percolation threshold. Correlation length of the infinite cluster (or the average radius of clusters) close to p_c

$$\xi \sim |p - p_c|^{-\nu_p}$$

with the pair correlation function $c(r) \sim r^{-\alpha}$ (see eq. (10)) up to ξ.

At p_c the percolating cluster is a fractal

Self-organisation

Systems slowly driven to a stationary state. No tuning. Avalanche-like series of changes close to a (critical, balanced) state. *Self-organised criticality* (SOC)

SOC model

Sandpile as an example. Critical slope, avalanches.

Model. Discrete variable $h(\mathbf{r})$ on a d dimensional lattice. Here $d = 2$, square lattice case. The rules (for the "slope")

Step 1 Choose a site (x, y) at random;

Step 2 Add one "grain" to that site: $h(x, y) \leftarrow h(x, y) + 1$;

Step 3 If $h(x, y) \leq 4$ then continue on Step 1.

Step 4 The pile is too "steep" locally so sand is redistributed among the nearest neighbours:

$$h(x, y) \leftarrow h(x, y) - 4$$

and

$$h(x', y') \leftarrow h(x', y') + 1,$$

where (x', y') denotes the nearest neighbours of site (x, y).

Step 5 The previous step is repeated until no site with $h > 4$ is found and then Step 1 is performed again.

The size of an avalanche is the number of times Step 4 was executed after adding a single grain.

$$P(\geq s) \sim s^{-\tau}, \tag{22}$$

where $P(\geq s)$ is number of avalanches larger that s and τ is around 0.25.

Applications in biology

Extinction events during evolution. Extinction rate shows peaks corresponding to major extinction events (avalanches).

SOC model of evolution

Fitness of a genotype or phenotype is its reproductive contribution to subsequent generations *relative* to the other (geno or pheno) types. (Another def: Fitness of a species represents its ability to survive environmental changes as a function of its genetic code).

Mutation and selection. Simple models lead to smoothly growing fitness.

Ecosystem of N species, fitness of the ith species ($i = 1, 2, \ldots, N$) is a real number

$$0 \leq B_i \leq 1. \tag{23}$$

Species with low fitness has larger selection pressure

1. We choose the species with the smallest fitness

$$B_j = \min_{i=1,\ldots,N} B_i. \tag{24}$$

2. mutation

$$B_j = \texttt{RND}, \tag{25}$$

where `RND` is equally distributed in [0,1].

3. Finally,

$$B_{j+1} = \texttt{RND}, \quad \text{and} \quad B_{j-1} = \texttt{RND}. \tag{26}$$

Leads to a non-trivial stable distribution with a threshold $B_{\rm c} \sim 0.67$.

Activity of a the system, of a site, correlations (power law)

The model exhibits punctuated equilibrium: avalanche size (avalanches with B less than the critical B) ($P(\geq s) \sim s^{-\tau}$, where $\tau \simeq 1.07$

Mean-field version (with K randomly chosen sites for update) can be solved, gives scaling

SOC in lung inflation

The bronchial airways have a fractal tree–like structure, 35 generations, airways closure for the last 10-14 generations

Emptied dog lung inflated by a constant flow and the terminal airway resistance (R_t) measured. R_t decreases in power law distributed discrete jumps.

The airways open in bursts.

Model

$t = 0$, airways are closed. Slowly growing external P_E. Opening threshold pressure P_{ij}, uniformly distributed on $[0, 1]$ (i - generation, j - column number).

Opening until no $P_{ij} < P_E$

Size distribution of *first* avalanches $n(s)$, s is the number of airways opening

We fix $P_{00} = p$ and open all airways with thresholds less than p. Average over realizations. This is percolation on a Caley tree. Can be solved exactly, leading to

$$n(s) \sim s^{-\tau} f(s^{\sigma}|p - p_c|)$$

where τ=3/2, σ=1/2, p_c=1/2

$f(u)$=const for small u and $f(u \gg 1) \to 0$.

Average over randomly chosen $P_{00} = p$: Integrate over p.

with $z = s^{\sigma}|p - p_c|$ and $dp = s^{-\sigma}\ dz$ we get

$$n(s) \sim s^{-\tau-\sigma}$$

.

BACTERIAL COLONIES

Collective behaviour of bacteria (colony, motion)

Microscopic versus macroscopic levels, self-organization

Bacteria in colonies

Experimental conditions

(agar, humidity, temperature, chemotaxis)

Microbiological background

Proliferation.

Motility.

Procaryotes move in aquatic environments by rotating their flagella, which are rather rigid membrane-bound helical protein polymers.

Run and tumble, chemotaxis

gliding

Morphology diagram

Variations in temperature, humidity, chemical composition of the substrate, etc. can give rise to different morphologies.

Morphology diagram of the strain *Bacillus subtilis* OG-01 as a function of peptone and agar concentration. The dashed line in indicates the boundary of the active movement of bacterial cells inside the colonies. The morphologies are classified as follows: fractal (A), compact with rough boundary (B), branching with periodic growth phases (C), compact with diffuse boundary (D) and dense branching (E).

Statistical measures of morphology

Compact morphology

Spreading of bacteria: The Fisher-Kolmogorov equation

Simplest bacterial spreading, nutrients are abundant and the substrate is wet, cells proliferate and migrate freely (randomly diffuse). Bacterial density, ϱ

Fisher-Kolmogorov equation

$$\partial_t \varrho = D_\varrho \nabla^2 \varrho + f(\varrho). \tag{27}$$

diffusion coefficient D_ϱ, mean squared displacement $d^2(t)$

$$\overline{d^2(t)} = 2D_\varrho t, \tag{28}$$

bacterial *multiplication* is in f. $f(\varrho) = r(c)\varrho$ for small ϱ, where r depends on the local nutrient concentration c. $r \sim c$ holds for small values of c. Max. cell density ϱ_*, because of accumulation of toxic metabolites. Units such that $\varrho_* = 1$. Growth rate must decrease as $\varrho \to 1$, $f(1) = 0$. Logistic function $f(\varrho) = r\varrho(1-\varrho)$ is used.

Numerical integration results in a growing domain, with $v \approx v_*$ where

$$v_* = 2\sqrt{rD_\varrho} \tag{29}$$

to calculate v, the expanding domain of $\varrho(x,t) \sim 1$ is written in a moving frame of reference as $\tilde{\varrho}(u,t) = \varrho(x,t)$ where $u = x - vt$ and without loss of generality $v > 0$, $\varrho(-\infty) = 1$ and $\varrho(\infty) = 0$ is assumed one obtains

$$\partial_t \tilde{\varrho} = D_\varrho \tilde{\varrho}'' + v\tilde{\varrho}' + f(\tilde{\varrho}), \tag{30}$$

Eq. (30) has a *stationary* solution for *any* value of $v \geq v_*$, i.e. v can not be derived by the stationarity condition imposed on $\tilde{\varrho}$. "Velocity selection problem".

Experiments lead to agreement

Self-affine interface roughening

Eden model: assembly of "cells" is grown on a lattice. Rules: In each step of the process one of the lattice sites next to the "populated" areas is chosen randomly and occupied.

Leads to self-affine surface.

The Kardar-Parisi-Zhang (KPZ) equation.

$$\partial_t h = \nu\partial_x^2 h + \frac{\lambda}{2}(\partial_x h)^2 + v + \eta, \tag{31}$$

$h(x,t)$ is the position of the surface. The surface tension term, $\nu\partial_x^2 h$, tends to smoothen the interface

The nonlinear $\lambda(\partial_x h)^2 + v$ terms reflects the *isotropy* of the growth. If the displacement is *perpendicular* to the interface,

$$\Delta h \approx v\Delta t\sqrt{1 + tg^2\phi} \approx v\Delta t + \frac{v\Delta t}{2}(\partial_x h)^2 \tag{32}$$

holds for small slopes.

Finally, $\eta(x,t)$ represents some sort of *noise*.

$C_\eta(x,t)$ wite noise.

The $H = 1/2$ result (Eden, KPZ) is not compatible with the experimental findings.

Quenched noise.

Assumption of an uncorrelated noise is not realistic. In fact, η is a function of the position of the interface

$$\eta(x,t) = 2D\tilde{\eta}(x, h(x,t)) \tag{33}$$

with

$$C_{\tilde{\eta}}(x,y) = \Delta(x)\Delta(y), \tag{34}$$

where $\Delta(u) \sim 1$ holds for some finite interval of u around zero.

The parameters λ, ν and v can be transformed out, i.e., it is sufficient to investigate the $\lambda = \nu = v = 1$ case

$$\partial_t h = \partial_x^2 h + \sqrt{1 + (\partial_x h)^2} + \eta, \tag{35}$$

D in (33), determines the average magnitude of the noise as $\sqrt{C_{\tilde{\eta}}(0,0)} = \sqrt{2D}$.

If $D > D_* \sim 1$ then the interface encounters points where it becomes pinned. If the density of such pinning sites is high enough then eventually the propagation of the whole surface can be blocked. Due to the single-valuedness of h, the cluster is *directed*.

Such a *directed percolation cluster* is characterised by two correlation lengths, being parallel ($\xi_\parallel$) and perpendicular ($\xi_\perp$) to the interface. At threshold density of the pinning sites p_c as

$$\xi_\parallel \sim |p - p_c|^{-\nu_\parallel}, \qquad \xi_\perp \sim |p - p_c|^{-\nu_\perp} \tag{36}$$

with

$$\nu_\parallel = 1.733 \quad \text{and} \quad \nu_\perp = 1.097. \tag{37}$$

Complete blocking appears when $\xi_\parallel$ becomes equal to the system size L. The width w of such a blocked surface is in the order of $\xi_\perp$, hence

$$L^H \sim w \approx \xi_\perp \sim |p - p_c|^{-\nu_\perp} \sim \xi_\parallel^{\nu_\perp/\nu_\parallel} \approx L^{\nu_\perp/\nu_\parallel}, \tag{38}$$

Thus,

$$H = \frac{\nu_\perp}{\nu_\parallel} = 0.633 \tag{39}$$

in accord with the numerical simulation of Eq. (35) with $D \gg D_*$.

In the $D \ll D_* \sim 1$ case (standard) KPZ equation can be applied resulting in interfaces with $H = 1/2$.

In the moving, but noise dominated regime the interface consists of both blocked and freely moving segments resulting in a *higher* roughness exponent.

Branching morphology

Mullins-Sekerka instability in diffusion-limited systems

Growth determined by the field u (e.g., the concentration) as

$$v_n = \vec{n}\nabla u, \tag{40}$$

where v_n and $\vec{n}$ is the dimensionless velocity and the normal vector of the surface.

$$\partial_t u = D\nabla^2 u \tag{41}$$

with boundary conditions $u(x \to \infty) = u_\infty = const$ and $u(x) = u_\Gamma(x)$ at the interface.

$$u_\Gamma = d_0\kappa, \tag{42}$$

where d_0 is proportional to the capillary tension and κ is the curvature of the interface.

If the growth is slow (Laplace equation)

$$0 = \nabla^2 u. \tag{43}$$

To show that the interface is *unstable* we consider a growing circle of radius R a boundary condition (42) and $u(R_\infty) = 1$, where $R_\infty \gg R$. The field $u(r, \phi) = u_0(r)$ satisfies the Laplace equation for slow growth which reads in polar coordinates as

$$u_0'' + \frac{u_0'}{r} = 0 \tag{44}$$

yielding

$$u_0 = 1 + B \ln \frac{r}{R_\infty}, \quad \text{with} \quad B = \frac{d_0/R - 1}{\ln R/R_\infty} > 0, \tag{45}$$

and the velocity of the interface is

$$\frac{dR}{dt} \equiv V(R) = u_0'(R). \tag{46}$$

If the surface is perturbed as $R(t) = R_0(t) + R_1(t, \phi)$, then u changes. We solve the problem approximately by, first, solving the Laplace equation and then insert this solution into (42) to obtain R_1.

We write the perturbation in the form of

$$R_1(\phi, t) = \hat{R}_1(t) \cos m\phi, \tag{47}$$

$$u(r, \phi, t) = u_0(r, t) + \hat{u}_m(t) u_m(r, \phi) \tag{48}$$

with $\hat{R}_1 \ll R_0$, $\hat{u}_m \ll 1$ and $u_m = r^{-m} \cos m\phi$. Both u_0 and u_m is a solution of (43)

and with this particular form the boundary condition (42) can be also satisfied. The curvature κ in polar coordinates

$$\kappa = \frac{R^2 + 2R'^2 - RR''}{(R^2 + R'^2)^{3/2}} \approx \frac{1}{R_0} - \frac{R_1''}{R_0^2}, \tag{49}$$

where prime denotes derivation in respect to ϕ. Thus in a linear approximation (42) reads as

$$d_0 \left(\frac{1}{R_0} + \frac{\hat{R}_1 m^2 \cos m\phi}{R_0^2} \right) = u(R(\phi), \phi), \tag{50}$$

where

$$u(R(\phi), \phi) \approx u(R_0, \phi) + (\partial_r u)(R_0, \phi)\hat{R}_1 \cos m\phi \tag{51}$$

resulting in the

$$\left(\frac{m^2 d_0}{R_0^2} - V(R_0) \right) \hat{R}_1 = \frac{\hat{u}_m}{R_0^m} \tag{52}$$

relation between $\hat{R}_1$ and $\hat{u}_m$. Substituting (52) back into (40) and keeping the linear terms of $\hat{R}_1$ one finds

$$\frac{d\hat{R}_1}{dt} = \Lambda_m(R_0)\hat{R}_1 \tag{53}$$

with

$$\Lambda_m(R_0) = \frac{m}{R_0} V(R_0) + V'(R_0) - \frac{m^3 d_0}{R_0^3}. \tag{54}$$

Λ_m is the amplification rate of the perturbation. For $\Lambda > 0$ the perturbation grows exponentially in time, while it dies out quickly if $\Lambda < 0$. The upper cutoff is due to the surface tension d_0. If $d_0 = 0$ the problem is ill-posed.

Diffusion-limited aggregation

Particles released from distant points diffuse and aggreagte. DLA clusters are scale invariant and are characterised by the fractal dimension

$$D_{DLA} = 1.715. \tag{55}$$

Models resolving individual bacteria

We model the *individual* bacteria by some idealised units, defining their behaviour and interactions, and obtain the behaviour computer simulations. Testing against experiments.

Modelling non-motile bacteria.

Each particle i is characterised by its *position* x_i and *cell cycle state* E_i.

i) Changes in E_i control the *sporulation* and *division*. If E_i decays below a threshold value (0) then the particle becomes inactive (sporulates). Above $E_i = 1$, the model bacterium divides, and both of the daughter cells receive an initial value of $0 < E_* < 1$. Changes in E_i are coupled to the nutrient consumption rate ω_i as

$$\frac{dE_i}{dt} = \kappa\omega_i - \epsilon \tag{56}$$

where κ is a factor relating the maximal nutrient consumption rate with the shortest cell cycle time, and ϵ is the generic "maintenance" term incorporating all other free energy expenditures.

ii) The *nutrient uptake* is limited by both the local concentration c and enzymatic rates. It is approximated by

$$\varrho(x_i)\omega_i = \min\left[\omega_{max}\varrho(x_i), \omega_0 c(x_i)\right]. \tag{57}$$

where ϱ is the local cell density, ω_{max} is the maximal uptake rate of the cells, and $\omega_0 c$ is the maximal diffusive transport from the substrate to the cells.

iii) Changes in c are given by (diffusion, sinks)

$$\frac{dc}{dt} = D_c\nabla^2 c - \sum_i \omega_i \delta(x - x_i). \tag{58}$$

$\kappa = 1$ and $\omega_{max} = 1$ can be assumed.

Gives reasonable agreement with some experiments.

Modelling motile bacteria.

We keep the rules i – iii and include new ones describing the motion.

iv) The active particles move randomly within a *boundary* described by

$$\frac{dx_i}{dt} = v_0\vec{e}, \tag{59}$$

where $\vec{e}$ denotes a random unit vector.

v) collisions of the particles with the boundary are counted and when a threshold value (N_c) is reached the new neighbouring cell can be occupied (slime, surfactant).

The resulting morphology diagram shows agreement with experiments. To explain further features: repulsive chemotaxis.

Sporulated bacteria emit diffusive chemicals acting as *chemorepellents*, i.e., the random swimming becomes biased: the cells move less toward the chemical gradient.

Chiral and rotating colonies

When bacterial motion is strongly influenced by cell-cell interactions, new phenomena occur.

Chiral patterns

- for a distinct phenotype of the same strain (longer).

- their orientation/motion is parallel to each other

Orientation ϕ_i is calculated. ϕ_i is changed according to the intrinsic chirality of the motion and the alignment to a local orientation field Φ

$$\phi_i(t + \Delta t) = P(\phi_i(t), \Phi(x_i(t))) + \varphi + \xi_i(t), \tag{60}$$

where φ represents the intrinsic chirality of the motion of the bacteria, ξ is noise and the interaction with the local orientation field is described by the projector function P as

$$P(\alpha, \beta) = \alpha + (\beta - \alpha) \mathrm{mod} \pi$$

. Φ represents the tracks in the agarose gel, its initial value was chosen to be the orientation of the first walker entering that field.

SYNCRONIZATION

The spontaneous phase ordering ("harmonization") of many, periodically changing units

Empirical observations

Neural networks (pacemaker in the heart) singing and life cicle of cicada, fire flies, walking, breathing, synchronized clapping, menstrual cycle, etc

Common: oscillators coupled through non-linear interaction

Two types of signals:

I) Delta function-like "firing" II) continuous

two types of mechanisms

Kuramoto-model (mean field, continuous)

Interaction free oscillator (rotator):

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \omega, \tag{61}$$

where ϕ is the phase, in the sense mod 2π, and ω is he so called natural frequency of the oscillator. N interacting oscillators:

$$\frac{\mathrm{d}\phi_i}{\mathrm{d}t} = \omega_i + \sum_{j=0}^{N-1} \Gamma_{ij}(\phi_j - \phi_i), \quad i = 0, 1, \ldots, N-1. \tag{62}$$

Here the $\Gamma_{ij}(\phi)$ interaction is a 2π-periodic function. A In the simplest case

$$\Gamma_{ij}(\phi) = \frac{K}{N}\sin(\phi), \quad i, j = 0, 1, \ldots, N-1. \tag{63}$$

leads to minimizing the phase difference between two oscillators, if $K > 0$. Mean field.

$$\frac{\mathrm{d}\phi_i}{\mathrm{d}t} = \omega_i + \frac{K}{N}\sum_{j=0}^{N-1} \sin(\phi_j - \phi_i) \quad i, j = 0, 1, \ldots, N-1. \tag{64}$$

ω_i are from a given $g(\omega)$ distribution. We assume, that $g(\omega)$ is such that

$$g(\omega) = \frac{1}{N}\sum_{i=0}^{N-1} \delta(\omega_i - \omega) \tag{65}$$

where $N(\omega_0, \sigma)$ Gaussian, with expectation value ω_0, and width σ. If $g(\omega) = \delta(\omega - \omega_0)$, than the Kuramoto model is equivalent to the ferromagnetic XY model.

Order parameter

Let us use the transformation

$$\psi_i = \phi_i - \omega_0 t, \qquad \omega_i \leftarrow \omega_i - \omega_0 \tag{66}$$

The Kuramoto-model is invariant under this transformation

$$\frac{\mathrm{d}\psi_i}{\mathrm{d}t} = \omega_i + \frac{K}{N}\sum_{j=0}^{N-1}\sin(\psi_j - \psi_i). \tag{67}$$

The states of the above equations can be described by the order parameter

$$z(t) = Z(t)e^{\mathrm{i}\theta(\mathrm{t})} = \frac{1}{N}\sum_{j=0}^{N-1} e^{\mathrm{i}\psi_\mathrm{j}(\mathrm{t})} \tag{68}$$

Here $z(t)$ or $Z(t)$ are the *complex* or *real order parameter*, while $\theta(t)$ is the average phase.

Clearly, $Z = |z|$.

$Z \simeq 1$, if ψ are almost equal, and $Z \simeq 0$, if ψ are random.

After rearranging it can be seen that

$$\frac{\mathrm{d}\psi_i}{\mathrm{d}t} = \omega_i + KZ\sin(\theta - \psi_i). \tag{69}$$

Formally looks like a non-interacting system

From rotational symmetry z behaves as a non-changing rotation:

$$z(t) = Ze^{\mathrm{i}(\Omega\mathrm{t}+\theta_0)}, \tag{70}$$

where Z, Ω and θ_0 are constants. Like a single big rotator.

Kuramoto has shown the existence of a bifurcation:

for $K < K_c$ the order parameter $Z = 0$. However, for small $\varepsilon = (K - K_c)/K_c$ the solution is not zero:

$$z = \sqrt{\varepsilon/\beta}e^{\mathrm{i}\theta}, \tag{71}$$

where θ is some constant. $K_{c,\mathrm{G}}$ is the critical coupling strength:

$$K_{c,\mathrm{G}} = \sqrt{\frac{8}{\pi}}\sigma. \tag{72}$$

Noisy oscillators

Identical oscillators plus independent white noise for each. Phases may become independent. Coupled Langevin equations with noisy forces $\xi_i(t)$. With $\psi_i = \phi_i - \omega_0 t$ we obtain (each of the natural frequencies is equal to ω_0)

$$\frac{\mathrm{d}\psi_i}{\mathrm{d}t} = \frac{K}{N}\sum_{j=0}^{N-1}\sin(\psi_j - \psi_i) + \xi_i(t)$$

$$< \xi_i(t) >_t = 0 \qquad < \xi_i(t)\xi_j(t') >_t = 2\sigma^2\delta(t-t')\delta_{ij}$$

Synchronization transition for an intermediate K/σ^2. As before we use the order parameter (or mean field):

$$Z = X + iY = \frac{1}{N}\sum_i e^{i\psi_i}$$

from here (after some algebra)

$$\frac{d\psi_i}{dt} = K(-X\sin\psi_i + Y\cos\psi_i) + \xi_i(t)$$

Goal is to write a self-consistent equation for the distribution of phases. Z is slow compared to noise (and then the above is like a Langevin eq. for an individual oscillator. The corresponding Fokker-Planck equation for the probability density of the phases

$$\frac{\partial P(\psi,t)}{dt} = K\frac{\partial}{\partial\psi}[(X\sin\psi - Y\cos\psi)P] + \sigma^2\frac{\partial^2 P}{\partial\psi^2}$$

For $N \to \infty$

$$Z = \int_0^{2\pi} d\psi P(\psi,t)e^{i\psi}.$$

To analize this non-linear system, we expand the density into Fourier series

$$P(\psi,t) = \frac{1}{2\pi}\sum_l P_l(t)e^{il\psi}$$

Note that $Z = X = iY = P_1^* = P_{-1}$ is equal to the first mode, and because of normalization $P_0 = 1$. Substituting the expansion into the F-P equation and separating the Fourier harmonics leads to a system of ordinary diff. equations

$$\frac{dP_l}{dt} = -\sigma^2 l^2 P_l + lK(P_{l-1}P_1 - P_{l+1}P_1^*)/2$$

The first three equations

$$\dot{P}_1 = \frac{K}{2}(P_1 - P_2 P_1^*) - \sigma^2 P_1,$$

$$\dot{P}_2 = K(P_1^2 - P_3 P_1^*) - 4\sigma^2 P_2,$$

$$\dot{P}_3 = \frac{3K}{2}(P_2 P_1 - P_4 P_1^*) - 9\sigma^2 P_3,$$

$P_0 = 1$ and all other P_i equal to zero (homogeneous distribution) is a solution. Linearizing around this the only unstable mode is the first one. It is stable if $K < 2\sigma^2$ and unstable if $K > 2\sigma^2$. This is the critical coupling. Since around the transition $K \simeq 2\sigma^2$ we estimate $|P_2| \sim |P_1|^2$ and $|P_3| \sim |P_1|^3$ and $\dot{P}_2 \simeq 0, P_3 \simeq 0$ (see the above equations). If we express P_2 through P_1 we get

$$\dot{Z} = (K/2 - \sigma^2)Z - \frac{K^2}{8\sigma^2}|Z|^2 Z$$

Equation for Hopf bifurcation. Stationary solution:

$$|Z|^2 = (K - 2\sigma^2)4\sigma^2/K^2$$

The order parameter grows as a square root of the distance from the critical point ($|Z| \sim \sqrt{\Delta K}$).

More realistic:

Interaction is distance dependent, *there is no "normalization"*

Simulations on a lattice

$$\dot{\phi}_i = \omega_i + \frac{K}{\eta} \sum_{j \neq i} \frac{1}{r_{ij}^{\alpha}} \sin(\phi_j - \phi_i), \tag{73}$$

where r_{ij} is the distance between the i-th and j-th oscillator and η is a normalization coefficient. Mean-field case when $\alpha = 0$. When $\alpha \to \infty$, nearest neighbour interaction case.

K versus α. Critical α? "Physical" is usually $\alpha = d - 1$.

If $\alpha \geq d$ the coupling term is bounded, for every N:

$$\left| \sum_{j \neq i}^{N} \frac{1}{r_{ij}^{\alpha}} \sin(\phi_j - \phi_i) \right| \leq \sum_{j \neq i}^{\infty} \frac{1}{r_{ij}^{\alpha}} < \infty. \tag{74}$$

However, for $\alpha \leq d$, in the limit of large N the coupling term may diverge depending on ω_i.

For $\alpha \leq d$ synchronization is enhanced as $N \to \infty$ for every (positive) value of the coupling constant K.

In 1d for $\alpha \lessapprox 1$ the system is fully locked. Three regimes.

Similar in 2d for $\alpha \lessapprox 2$

Analogous to continuous phase transitions (but control parameter is now an exponent!!)

$$\lim_{N \to \infty} \lim_{K \to 0^+} Z(\alpha) = \begin{cases} 1 \text{ if } \alpha \leq d \\ 0 \text{ if } \alpha > d \ , \end{cases} \tag{75}$$

Integrate-and-fire type oscillators

Definition/model:

for a given oscillator: i) x state variable, monotonously increasing until the threshold $x = 1$. ii) At this point x immediately jumps back to zero and the cycle starts from the beginning. iii) the evolution of x: let $x = f(\phi)$, where $f : [0, 1] \to [0, 1]$ smooth, monotonously increasing function (at this point we also assume it is convex), i.e., $\mathrm{d}f/\mathrm{d}\phi > 0$ and $\mathrm{d}^2 f/\mathrm{d}\phi^2 < 0$. Here $\phi \in [0, 1]$ is a phase variable, such that, $\mathrm{d}\phi/\mathrm{d}t = 1/T$, where T is the period, and $x = 0 \Leftrightarrow \phi = 0$ and $x = 1 \Leftrightarrow \phi = 1$.

From these it follows trivially that $f(0) = 0$ and $f(1) = 1$. Since f invertible, let $g = f^{-1}$, i.e.., $g(x) = \phi$).

Interaction: when an oscillator ,,fires" ($x_i = 1$), then it increases the state variable of the other oscillators by ε or increases their x up to the threshold (if $x_j + \varepsilon$ would be larger than 1)

$$x_i(t) = 1 \Rightarrow x_j(t^+) = \min(1, x_j(t) + \varepsilon) \quad \forall j \neq i. \tag{76}$$

Let us first consider two oscillators (here always $\omega_i = \omega = 1/T_i$)

It is easy to see (also graphically) that they tend to synchronize continuously.

If it is true for two, than it is true for many.

Absorption

Definition: an avalanche is a series of firings induced by a single firing, so that all of the firings are resulted from the intermediate effect of the previous one within the same avalanche

There is no "recovery" stage (an oscillator which fired in a given avalanche does not react to the rest of the firings within the same avalanche

question of time scales

Mechanism of synchronization

Valid for any monotonously increasing mapping

If there are groups of oscillators already synchronized they "move" with different phase velocity!

It starts if there are at least two oscillators for which $x_{i+1} - x_i < \varepsilon$

The probability of this can be estimated from a Poisson distribution of the initial states and the result agrees with the intuitively expected one: there is such a pair of oscillators with a probability close to 1 if $1/N < \varepsilon$.

MICROSCOPIC MECHANISMS OF BIOLOGICAL MOTION

The physics of microscopic objects

Processes involving microscopic objects (proteins) are inherently stochastic.

The motion of any object in a thermal environment can be described by the *Langevin equation* :

$$m\ddot{x}(t) = -\gamma\dot{x}(t) + \gamma\sqrt{2D}\xi(t) + F(x,t) \ , \tag{77}$$

where x, m, γ, and D denote the position, mass, viscous friction coefficient, and diffusion coefficient of the object, respectively. $\xi(t)$ is white noise.

Viscous friction and the thermal noise are not independent. Their magnitudes are connected by the fluctuation-dissipation theorem (or Einstein relation):

$$D = \frac{k_{\mathrm{B}}T}{\gamma} \ . \tag{78}$$

$\tau_{\mathrm{relax}} = m/\gamma$ - characteristic time for loosing velocity. $\lambda_{\mathrm{relax}} = vm/\gamma$ - characteristic distance. If $\lambda_{\mathrm{relax}}/a \ll 1$ the damping is strong, because the particle stops on a much shorter distance than its size a. Supposing that m is proportional to $a^3\rho$ and γ is proportional to $a\eta$ (cf. Stokes law), where ρ and η are the density and dynamic viscosity of the medium respectively, $\lambda_{\mathrm{relax}}/a$ becomes proportional to

$$R = \frac{va}{\eta/\rho} = \frac{va}{\nu} \ , \tag{79}$$

which is called the *Reynolds number* ($\nu = \eta/\rho$ is the kinematic viscosity).

Reynolds number for biomolecules is in the order of 10^{-3} (substituting actual numbers). It means that the viscous friction can stop a protein on a distance ($\sim 10^{-3}$ nm) much shorter than the size of the atoms/molecules.

Thus, the acceleration term of the Langevin equation (77) can be neglected:

$$\dot{x}(t) = F(x,t)/\gamma + \sqrt{2D}\xi(t) \ . \tag{80}$$

From this stochastic ordinary differential equation one can derive a deterministic partial differential equation, the *Fokker-Planck equation* for the probability density $P(x,t)$ of the position of the particle:

$$\partial_t P(x,t) = -\partial_x J(x,t) \ , \tag{81}$$

where

$$J(x,t) = \frac{F(x,t)}{\gamma} P(x,t) - \frac{k_{\rm B}T}{\gamma} \partial_x P(x,t) \tag{82}$$

is the probability current of the particle. One particle, stochastic $->$ many particles, deterministic.

If the force field $F(x,t)$ is the negative gradient of a potential: $F(x,t) = -\partial_x V(x,t)$, the probability current can be written in the form

$$J(x,t) = -\frac{k_{\rm B}T}{\gamma} {\rm e}^{-V(x,t)/k_{\rm B}T} \partial_x \left[{\rm e}^{V(x,t)/k_{\rm B}T} P(x,t) \right] \ . \tag{83}$$

because $\partial[a^{f(x)}]/\partial x = a^{f(x)} \partial f(x)/\partial x$

Kramers rate and Arrhenius law

For Brownian particles wiggling in deep potential wells (compared to $k_{\rm B}T$), rarely jumping out into one of the neighbouring wells.

the escape (or jumping) rate constants, k, follow the Arrhenius law

$$k = \nu \, {\rm e}^{-\Delta E/k_{\rm B}T}, \tag{84}$$

where ΔE denotes the threshold energy for activation and ν is a frequency prefactor.

For *deep* potential well probability density near the bottom can be well *approximated* by its equilibrium value

$$P_{\rm eq}(x) = P_0 {\rm e}^{-V(x)/k_{\rm B}T} \ , \tag{85}$$

which can be derived from Eq. (83) by setting its right-hand side to zero:

$$\partial_x \left[{\rm e}^{V(x,t)/k_{\rm B}T} P(x,t) \right] = 0 \ , \tag{86}$$

$$ {\rm e}^{V(x,t)/k_{\rm B}T} P(x,t) = Const \ , \tag{87}$$

The normalisation factor is approximately

$$P_0 = \frac{1}{\int_b^c {\rm e}^{-V(x)/k_{\rm B}T} {\rm d}x} \ , \tag{88}$$

Molecular motors

The origin of biological motion: *motor proteins* operating at the *molecular scale*. Numerous different families of these molecular motors have been discovered, and the newest experimental techniques developed in recent years have allowed to study *in vitro* the operation of *individual* motor proteins.

The motors move in discrete, unidirectional steps with step sizes in the range of a few nanometers, and exert piconewton forces. The motors use the *chemical energy* stored in adenosine triphosphate (ATP) molecules or in proton gradient as fuel and convert it into *mechanical work*.

Characterisation of motor proteins

Two important kinds of filamentous structures: the *microtubules*, the *actin filaments*. Along these polymeric filaments molecular motor proteins can move. Three different motor protein families have been identified by now: *kinesins*, *dyneins* which move along the microtubules, and *myosins* which move along the actin filaments.

Kinesin

Native kinesin is a dimeric molecule with a tail and two globular ($\sim$9×3×3 nm) head domains. The heads are highly conserved, and each contains an ATP and a tubulin binding site.

- kinesin moves unidirectionally parallel to the protofilaments towards the plus end of the microtubule;
- under an increasing load the speed of the kinesin decreases almost linearly;
- under its stall force (about 5 pN) kinesin still consumes ATP at an elevated rate;
- in the absence of ATP (in rigor state) kinesin binds to the microtubule very strongly: it supports forces in excess of 10 pN
- the step size ($\sim$ 8 nm) is identical with the periodicity of the protofilaments
- kinesin hydrolyses one ATP per each 8-nm step;
- occasionally kinesin takes backward steps;
- one step consists of two sequential subprocesses with

Moreover, changing a few amino acids in the neck can reverse the directionality of the motor

Muscle contraction

Myosin cross-bridges extending from the thick myosin filaments attach to the binding sites of the thin actin filaments and exert force on them, inducing a relative sliding of the actin and myosin filaments.

Rotary motors

Bacterial flagellar motor, DNA motors move unidirectionally along their polymeric track, while they are also rotating around it.

ATP synthase: The function of the ATP synthase is to couple the synthesis of ATP with a proton flux across the membrane down the proton gradient. *binding change mechanism*

Motility assay

Optical trap uses a highly focused laser light to grab and manipulate microscopic dielectric objects. The dielectric energy density in the object is

$$\rho(\boldsymbol{x}) = -\frac{1}{2}\Delta\alpha E(\boldsymbol{x})^2 \ , \tag{89}$$

where $E(\boldsymbol{x})$ is the electric field of the light, and $\Delta\alpha$ is the polarisability of the object relative to the suspending solution. Energy density is minimal in the center of the focal region.

Fluctuation driven transport

Thermal ratchets

non-equilibrium fluctuations

Sawtooth shaped periodic potential. In his *Lectures*, Feynman used the "ratchet and pawl" engine to illustrate some implications of the second law of thermodynamics.

The two basic types of nonequilibrium fluctuations are the i) *fluctuating potential (or "flashing ratchet")*, where the potential is time dependent and fluctuates between two or more different states; and the ii) *fluctuating force (or "rocking ratchet")*, where the potential is static but a fluctuating external force with zero time average is applied.

The Langevin equation (80) for the motion of a particle in such systems can be written as

$$\gamma\dot{x} = -\partial_x V(x,t) + F(t) + F_{\rm load} + \gamma\sqrt{2D}\xi(t) \ , \tag{90}$$

where $V(x,t)$ is the periodic ratchet potential, $F(t)$ is the fluctuating external force (if there is any), and $F_{\rm load}$ is the load force.

Basic ratchet models

1. Flashing

Parrondo's game

Game A: win with $p - \epsilon$, (loose with $1 - p + \epsilon$)

Game B: if capital is multiple of M win with $p_1 - epsilon$, otherwise win with $p_2 - \epsilon$

Two separately loosing games played together win! (e.g., for $p = 1/2$, $p_1 = 1/10$ and $p_2 = 3/4$

2. Rocking

If an external force alternates between $+F$ and $-F$, than when the force points to the right ($F(t) = +F$), the energy barrier for jumping to the right becomes smaller by F-times λ_1. In the opposite case the barrier becomes smaller but only by F-times the shorter distance λ_2.

Frequency response is very different

$v(x,t) = V(x)$. In the steady state, when the probability density $P(x,t)$ and the current $J(x,t)$ have their stationary values $P_{\rm st}(x)$ and $J_{\rm st}(x)$, this leads to $\partial_x J_{\rm st}(x) = 0$, i.e., the stationary current is constant in space: $J_{\rm st}(x) = J_{\rm st}$. Integrating Eq. (83) from 0 to λ

$$J_{\rm st}\frac{\gamma}{k_{\rm B}T}\int_0^\lambda {\rm e}^{V(x)/k_{\rm B}T}{\rm d}x = -\left[{\rm e}^{V(x)/k_{\rm B}T}P_{\rm st}(x)\right]_{x=0}^{\lambda} \ . \tag{91}$$

The period is λ, thus, the right-hand side of this equation is zero. The integral on the LHS is always positive,

$$J_{\rm st} = 0 \ , \tag{92}$$

Note that the stationary (or equilibrium) solution of the probability density is the Boltzmann distribution:

$$P_{\rm st}(x) = \frac{{\rm e}^{-V(x)/k_{\rm B}T}}{\displaystyle\int_0^\lambda {\rm e}^{-V(x)/k_{\rm B}T}{\rm d}x} \ , \tag{93}$$

which follows from Eq. (83) with zero current as was discussed in the case of the Arrhenius law.

Realistic models

Kinesin

i) We suppose that each of the two Brownian heads moves along its own one-dimensional periodic potential with period $L = 8$ nm. ii) Each period contains a deep potential well corresponding to the binding site of the β-tubulin and the potentials everywhere else are flat. iii) Each well has an asymmetric "V" shape reflecting the polarity of the protofilaments iv) the heads are connected at a hinge, and a spring acts between them. At the beginning the spring is unstrained. v) After one of the heads binds an ATP, its hydrolysis causes a conformational change in this head; this means that the rest length of the spring changes from zero to 8 nm right after the hydrolysis. Reaching its new 8 nm rest length another conformational change occurs in the head as a consequence of the ADP release: the rest length of the spring changes back to zero.

We assume a constant force F applied to the hinge.

The motion of the heads can be described by two coupled Langevin equations

$$\gamma \dot{x}_1 = -\partial_x V(x_1) - F_1^{\text{load}} - K\left[x_1 - x_2 - l(t)\right] + \gamma\sqrt{2D}\xi_1(t) \ ,$$
$$\gamma \dot{x}_2 = -\partial_x V(x_2) - F_2^{\text{load}} + K\left[x_1 - x_2 - l(t)\right] + \gamma\sqrt{2D}\xi_2(t) \ ,$$

where x_1 and x_2 denote the positions of the front and back heads respectively, $V(x)$ is the periodic ratchet potential, K is the stiffness of the spring, $F_1^{\text{load}} = F_2^{\text{load}} = F/2$ are the load forces acting on the heads, and the two thermal noise terms $\xi_1(t)$, $\xi_2(t)$ are uncorrelated.

$l(t)$ represents the rest length of the spring, which is not an explicit function of time, but alternates stochastically between 0 and 8 nm as ATP is hydrolysed or ADP is released. Depends also on x_1 and x_2 because ATP can be hydrolysed only in the relaxed state of the molecule.

The parameters can be estimated including the depth Q of the potential valley.

We assume that at the beginning both heads are in the same well of the potential. When the spring stretches the it has a large probability $p_{0L}^+ = k_{0L}^+/(k_{0L}^+ + k_{0L}^-)$ that the front head jumps to the forward direction, and a backward step has only a small probability $p_{0L}^- = k_{0L}^-/(k_{0L}^+ + k_{0L}^-)$ where k_{0L}^+ and k_{0L}^- denote the corresponding jumping rate constants.

The average lifetime $t_{0L} = 1/(k_{0L}^+ + k_{0L}^-)$ of the stressed state for larger load increases. Completing this subprocess the second conformational change occurs: the spring is trying to contract. Now for low load force the probabilities are: $p_{L0}^+ = k_{L0}^+/(k_{L0}^+ + k_{L0}^-)$ (is close to 1), while $p_{L0}^- = k_{L0}^-/(k_{L0}^+ + k_{L0}^-) << 1$

Since the potential wells are deep compared to $k_{\mathrm{B}}T$, the jumping rate constants can be calculated from the Kramers rate with a potential that is the sum of the ratchet potential, the quadratic spring potential, and the linear potential of the load force. From the jumping rate constants one can calculate the average displacement

$$d = (p_{0L}^{+} p_{L0}^{+} - p_{0L}^{-} p_{L0}^{-}) L =$$

$$(k_{0L}^{+} k_{L0}^{+} - k_{0L}^{-} k_{L0}^{-})[1/(k_{0L}^{+} + k_{0L}^{-})][1/(k_{L0}^{+} + k_{L0}^{-})] L$$

and duration

$$t = t_{0L} + t_{L0} = 1/(k_{0L}^{+} + k_{0L}^{-}) + 1/(k_{L0}^{+} + k_{L0}^{-})$$

of a step.

At large ATP concentration dividing the average displacement by the average time gives the average velocity $v = d/t$ of the kinesin. At low ATP concentration the rate-limiting factor is the diffusion of the ATP to the kinesin, thus, the average velocity is proportional to the average displacement during one cycle multiplied by the ATP consumption rate

$$\nu(c_{\mathrm{ATP}}) = \frac{\nu_{\mathrm{sat}}\, c_{\mathrm{ATP}}}{K_{\mathrm{m}} + c_{\mathrm{ATP}}} \approx \frac{\nu_{\mathrm{sat}}}{K_{\mathrm{m}}}\, c_{\mathrm{ATP}} \approx \text{const } c_{\mathrm{ATP}} \ , \tag{94}$$

where c_{ATP} denotes here the ATP concentration, $\nu_{\mathrm{sat}} = 1/t$ is the inverse of the cycle time, and K_{m} is the mechanochemical Michaelis-Menten constant.

Long-Stride model

Chemically reversible model

the direction of motion is controlled by subtle differences in the chemical mechanism of ATP hydrolysis

In this model the direction of motion changes when the relative rates for releasing $\mathrm{P_i}$ and ADP change.

Single head?

Collective effects

muscle tissues, eukaryotic cilia and flagella rigidly attached dynein molecules drive the microtubules sliding, large group of motor proteins can carry the same cargo, motility assays

Interesting collective effects, reversal of current, etc for the non-connected

Collective behaviour of rigidly attached particles

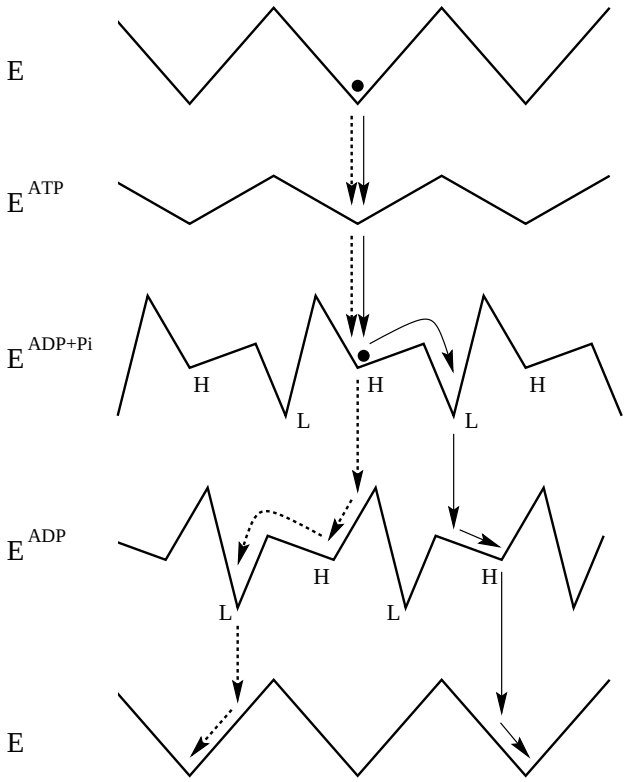

Figure 1: Ratchet mechanism for chemically reversible motion. The motor concomitantly cycles through its chemical states while catalysing ATP hydrolysis (in the vertical direction) and translocates through space along a microtubule (horizontally). (After [**?**])

periodic, asymmetric potential, is turned "on" and "off" independently for each particle, and the particles are attached to a rigid backbone (incommensurate with the period of the potential)

the particles are uniformly distributed in the periods. $P_{\rm on}(x,t)$ and $P_{\rm off}(x,t) = 1 - P_{\rm on}(x,t)$ show which fraction of the particles is in the "on" and "off" states of the potential respectively at time t and position x ($0 \le x < 1$). The equations of motion

$$\begin{aligned} \partial_t P_{\rm on} + v\partial_x P_{\rm on} &= -\omega_{\rm on}(x) P_{\rm on} + \omega_{\rm off}(x) P_{\rm off} \ , \\ \partial_t P_{\rm off} + v\partial_x P_{\rm off} &= +\omega_{\rm on}(x) P_{\rm on} - \omega_{\rm off}(x) P_{\rm off} \ , \end{aligned} \tag{95}$$

where $\omega_{\rm on}(x)$ and $\omega_{\rm off}(x)$ denote the transition rate constants between the to states of the potential, and the velocity of the backbone, v, is determined by $v = f_{\rm ext} + f$. The external force $f_{\rm ext}$ and the average force

$$f = -\int_0^1 {\rm d}x \, (P_{\rm on}\partial_x V_{\rm on}(x) + P_{\rm off}\partial_x V_{\rm off}(x)) \tag{96}$$

exerted by the potentials (which should be zero for the flat potential) are normalised per particle.

In the steady state, using the relation $P_{\rm off} = 1 - P_{\rm on}$,

$$v\partial_x P_{\rm on} = -(\omega_{\rm on}(x) + \omega_{\rm off}(x))P_{\rm on} + \omega_{\rm off}(x) \ , \tag{97}$$

$$f_{\rm ext} = v + \int_0^1 {\rm d}x \, P_{\rm on}\partial_x(V_{\rm on}(x) - V_{\rm off}(x)) \ , \tag{98}$$

allowing the determination of the external force $f_{\rm ext}(v)$ as a function of the velocity v of the backbone. Eq. (97) can be solved either analytically for some potential shapes or in a power expansion as a function of v.

Collective effects for symmetric periodic potentials

Let the perturbation $\Omega\Theta(x) > 0$ around the equilibrium state for the given parameters, be also symmetric, and different from zero close to the minimum only. Then (it can be shown) if Ω is smaller than $\Omega_{\rm c}$, the velocity of the backbone is zero, but for $\Omega > \Omega_{\rm c}$ the solution bifurcates to $v = \pm v(\Omega)$, while the $v = 0$ solution becomes unstable!

Qualitatively: For $v = 0$ the localised excitation $\Omega\Theta(x)$ leads to a depletion of $P_{\rm on}(x)$ near the minimum. Since $\Theta(x)$ and the potentials are spatially symmetric, the force f vanishes. If due to perturbation v becomes > 0, the depletion of $P_{\rm on}(x)$ is being transported to the right. Population along the positive potential slope is depleted, while along the negative increases. The average force f pulls the backbone to the right (and increases the effect).

COLLECTIVE MOTION

Flocking: collective motion of self-propelled particles

Models and simulations

A generic model for two dimensional SPP systems

Particles with location $\vec{x}_i$ and velocity $\vec{v}_i$ pointing in direction ϑ_i. Magnitude of the velocity fixed to v_0, local neighbourhood is $S(i)$

$$\vartheta_i(t+\Delta t) = \langle \vartheta(t) \rangle_{S(i)} + \xi, \tag{99}$$

where the noise ξ is a random variable with a uniform distribution in the interval $[-\eta/2, \eta/2]$. Updated as

$$\vec{x}_i(t+\Delta t) = \vec{x}_i(t) + \vec{v}_i(t)\Delta t \tag{100}$$

with $|\vec{v}_i| = v_0 = const$

Ferromagnetic analogy, but momentum is not conserved.
$v \to 0$ (XY modell) and $v \to \infty$ (mean-field) limits.

Galilean invariance violated

Scaling properties

Results from simulations

Order parameter

$$\phi \equiv \frac{1}{N} \left| \sum_j \vec{v}_j \right|. \tag{101}$$

Simulations started from a random, disordered configuration with $\phi(t=0) \approx 0$. Phase transition:

$$\phi(\eta) \sim \begin{cases} (\frac{\eta_c(\varrho)-\eta}{\eta_c(\varrho)})^\beta & \text{for } \eta < \eta_c(\varrho) \\ 0 & \text{for } \eta > \eta_c(\varrho) \end{cases}, \tag{102}$$

where $\eta_c(\varrho)$ is the critical noise amplitude

$$\beta_{2d} = 0.42 \pm 0.03, \tag{103}$$

found to be *different* from the the mean-field value 1/2.

Role of density. All of the $\phi(\eta, \varrho)$ functions can be collapsed to the same function $\tilde{\phi}(x)$ by rescaling η with $\eta_c(\varrho)$:

$$\phi(\eta, \varrho) = \tilde{\phi}(\eta/\eta_c(\varrho)), \tag{104}$$

where $\tilde{\phi}(x) \sim (1-x)^\beta$ for $x < 1$, and $\tilde{\phi}(x) \approx 0$ for $x > 1$. The critical line $\eta_c(\varrho)$ in the

$\eta - \varrho$ parameter space was found to follow

$$\eta_c(\varrho) \sim \varrho^{\kappa}, \tag{105}$$

with $\kappa = 0.45 \pm 0.05$.

Different from diluted ferromagnets.

3d SPP system

Generalisation of (99) for the 3d case

$$\vec{v}_i(t + \Delta t) = v_0 \ \mathbf{N}(\ \mathbf{N}(\langle \vec{v}(t) \rangle_{S(i)}) + \vec{\xi}), \tag{106}$$

where $\mathbf{N}(\vec{u}) = \vec{u}/|\vec{u}|$ and the noise $\vec{\xi}$ is uniformly distributed in a sphere of radius η.

Phase diagram: different from 3d disordered ferromagnets

1d SPP system

In 1d the dynamics is special (specific crowding effects).

A model of people, moving in a narrow channel.

N off-lattice particles along a line of length L with coordinates x_i and dimensionless velocity u_i updated as

$$x_i(t + \Delta t) = x_i(t) + v_0 u_i(t) \Delta t, \tag{107}$$

$$u_i(t + \Delta t) = G(\langle u(t) \rangle_{S(i)}) + \xi_i. \tag{108}$$

$\langle u \rangle_{S(i)}$ for the ith particle is calculated over the interval $[x_i - \Delta, x_i + \Delta]$, where $\Delta = 1/2$. The function G incorporates both the "propulsion" and "friction" preferring v_0 on the average: $G(u) > u$ for $u < 1$ and $G(u) < u$ for $u > 1$.

$$G(u) = \begin{cases} (u+1)/2 & \text{for } u > 0 \\ (u-1)/2 & \text{for } u < 0, \end{cases} \tag{109}$$

Phase transition with

$$\beta_{1d} = 0.60 \pm 0.05, \tag{110}$$

which is different from both the the mean-field value $1/2$ and $\beta_{2d} \approx 0.4$ found in 2d.

Further variants of SPP models

Effect of boundary conditions.

Short range "hard-core" repulsion

$$\vartheta_i(t+\Delta t) = \Phi\Big(-\sum_{\substack{j\neq i \\ |\vec{x}_j-\vec{x}_i|<\epsilon^*}} \mathbf{N}(\vec{x}_j(t)-\vec{x}_i(t))\Big), \tag{111}$$

Rotating phase within a circular region

Theoretical results

Continuum equations for the 1d system

We denote by $n(u,x,t)dudx$ the number of particles moving with a velocity between $[v_0u, v_0(u+du)]$ at time t in the $[x, x+dx]$. The particle density $\rho(x,t)$

$$\rho = \int n du, \tag{112}$$

local dimensionless average velocity $U(x,t)$ as

$$\rho U = \int n u du. \tag{113}$$

It can be shown that

$$\partial_t \rho = -v_0 \partial_x(\rho U) + D\partial_x^2 \rho \tag{114}$$

with

$$D = v_0^2 \tau \sigma^2/2. \tag{115}$$

Corresponds to conservation of particles; the connection between its "macroscopical" parameters v_0, D and the parameters of the underlying microscopical dynamics is established (σ^2 is the average standard deviation of the local velocity distribution). The diffusion term is a consequence of the non-vanishing correlation time τ.

For the local average velocity U

$$\partial_t U = f(U) + \mu^2 \partial_x^2 U + \alpha \frac{(\partial_x U)(\partial_x \rho)}{\rho} + \zeta, \tag{116}$$

where $f(U) = (G(U)-U)/\tau$, $\mu^2 = (dG/dU)/(6\tau)$, $\alpha = 2\mu^2$ and

$$\zeta = \frac{1}{\rho\tau}\int \nu u du. \tag{117}$$

Note, that $f(U)$ is an antisymmetric function with $f(U) > 0$ for $0 < U < 1$ and

$f(U) < 0$ for $U > 1$. The noise ζ is *not* negligible in this case, and satisfies

$$\overline{\zeta} = 0 \tag{118}$$

$$\overline{\zeta^2} = \sigma^2/\rho\tau^2. \tag{119}$$

Eqs. (114) and (116) with the coefficients μ, α, σ, v_0 and D represents the continuum theory of a large class of SPP models in 1d.

Non-linear coupling term $(\partial_x U)(\partial_x \rho)/\rho$!

For $\alpha = 0$ the dynamics of the velocity field U is independent of ρ, and Eq.(116) is equivalent to the time dependent Ginsburg-Landau Φ^4 model of spin chains, where domains of opposite magnetisation develop at finite temperatures.

Numerics shows tansition

The existence of long-range order

Navier-Stokes equation for the continuum, long wavelength description.

Continuum equations of motion for $\vec{v}$ and ρ consistent with the symmetries and conservation laws of the problem.

A simplified version is:

$$\begin{aligned} \partial_t \mathbf{v} + (\mathbf{v}\nabla)\mathbf{v} &= (\alpha - \beta|\mathbf{v}|^2)\mathbf{v} - \nabla P + \\ &\quad D\nabla^2\mathbf{v} + \vec{\xi} \\ \partial_t \rho + \nabla(\rho\mathbf{v}) &= 0, \end{aligned} \tag{120}$$

where the $\alpha, \beta > 0$ make $\mathbf{v}$ have a nonzero magnitude, D is diffusion constant and $\vec{\xi}$ is noise.

The pressure P depends on the local density

$$P(\rho) = \sum_n \sigma_n (\rho - \rho_0)^n. \tag{121}$$

Dynamic renormalisation group gives for the measure of the fluctuations, $C_C(\vec{R})$, defined as:

$$C_C(\vec{R}) = C\left(\vec{R}\right) - \lim_{|\vec{R}'|\to\infty} C\left(\vec{R}'\right) \quad , \tag{122}$$

with

$$C(\vec{R}) \equiv \left\langle \vec{v}(\vec{R} + \vec{r}, t) \cdot \vec{v}(\vec{r}, t) \right\rangle \tag{123}$$

and

$$C(\vec{R} \to \infty) = \phi^2, \tag{124}$$

decays to zero much more rapidly, as $|\vec{R}| \to \infty$, than the analogous correlation function in magnets. Quantitatively, for points whose separation $\vec{R} \equiv \vec{R}_\perp$ lies perpendicular to the mean direction of motion

$$C_C(\vec{R}) \propto R_\perp^{2\chi} \tag{125}$$

holds, where the universal exponent is

$$\chi = -\frac{1}{5} \tag{126}$$

exactly, in $d = 2$, and is smaller than its equilibrium value, $1 - \frac{d}{2}$, in magnetic systems for all $d < 4$.

Increased fluctuations in different parts of the system actually *enhance* the exchange of information between those different parts (leading to supression of fluctuations).

Anisotropy: particles make small errors $\delta\theta$ in their direction of motion, their random motion *perpendicular* to the mean direction of motion $\langle\vec{v}\rangle$ is much larger than that *along* $\langle\vec{v}\rangle$; the former being $\propto \delta\theta$, while the later is proportional to $1 - \cos\delta\theta \sim \delta\theta^2$.

"Purposeful" collective motion

N organisms (pedestrians, particles) i of mass m_i tends to move with a certain desired speed v_i^0 into a certain direction $\vec{e}_i^0$, and adapts its actual velocity $\vec{v}_i$ with a certain characteristic time τ_i.

Keeping a velocity-dependent distance from other particles j and walls W. "Interaction forces" $\vec{f}_{ij}$ and $\vec{f}_{iW}$. Acceleration equation

$$m_i \frac{d\vec{v}_i}{dt} = m_i \frac{v_i^0(t)\vec{e}_i^0(t) - \vec{v}_i(t)}{\tau_i} + \sum_{j(\neq i)} \vec{f}_{ij} + \vec{f}_{iW}\,, \tag{127}$$

while the change of position $\vec{r}_i(t)$ is given by the velocity $\vec{v}_i(t) = d\vec{r}_i/dt$. The tendency of two organisms i and j to stay away from each other by a repulsive interaction force

$$A_i \exp[(r_{ij} - d_{ij})/B_i]\,\vec{n}_{ij}$$

where A_i and B_i are constants. $d_{ij} = \|\vec{r}_i - \vec{r}_j\|$ denotes the distance between the centers of mass, and

$$\vec{n}_{ij} = (n_{ij}^1, n_{ij}^2) = (\vec{r}_i - \vec{r}_j)/d_{ij}$$

is the normalised vector pointing from organism j to i.

If their distance d_{ij} is smaller than the sum $r_{ij} = (r_i + r_j)$ of their radii r_i and r_j, the "pedestrians" touch each other. Then, two additional forces a "body force"

$$k(r_{ij} - d_{ij})\,\vec{n}_{ij}$$

counteracting body compression and a "sliding friction force"

$$\kappa(r_{ij} - d_{ij})\,\Delta v_{ji}^t\,\vec{t}_{ij}$$

impeding *relative* tangential motion, if particle i comes close to j. Herein, $\vec{t}_{ij} = (-n_{ij}^2, n_{ij}^1)$ means the tangential direction and $\Delta v_{ji}^t = (\vec{v}_j - \vec{v}_i) \cdot \vec{t}_{ij}$ the tangential velocity difference, while k and κ represent large constants. Thus,

$$\vec{f}_{ij} = \{A_i \exp[(r_{ij} - d_{ij})/B_i] + kg(r_{ij} - d_{ij})\}\,\vec{n}_{ij} + \kappa g(r_{ij} - d_{ij})\Delta v_{ji}^t\,\vec{t}_{ij}\,,$$

where the function $g(x)$ is zero, if the pedestrians do not touch each other ($d_{ij} > r_{ij}$), otherwise equal to the argument x.

The corresponding interaction force with the wall reads

$$\vec{f}_{iW} = \{A_i \exp[(r_i - d_{iW})/B_i] + kg(r_i - d_{iW})\}\,\vec{n}_{iW} \\ -\kappa g(r_i - d_{iW})(\vec{v}_i \cdot \vec{t}_{iW})\,\vec{t}_{iW}\,.$$

Simulations (two lanes in a "corridor", through a "door")

Herding versus individual motion

Each organism i has an individual direction $\vec{e}_i$ and tries to follow the average direction $\langle \vec{e}_j^{\,0}(t)\rangle_i$ of his neighbours j in a certain radius R_i. These options are weighted with some parameter p_i:

$$\vec{e}_i(t+\delta t) = N\left[(1-p_i)\,\vec{e}_i + p_i\,\langle \vec{e}_j^{\,0}(t)\rangle_i\right]\,, \tag{128}$$

where $N(\vec{z}) = \vec{z}/\|\vec{z}\|$ denotes normalization of a vector $\vec{z}$. Individualistic behaviour if p_i is low, but herding behaviour if p_i is high.

Simulations show optimum for intermediate p

NETWORKS

Topology: the simplest aspect of complex systems. Still rich

(neural networks, food chains, genetic networks, friendship, scientific collaborators, etc)

Real networks are self-organized

The Erdős-Rényi model

Start with N vertices and no bonds. With probability $p_{\rm ER}$, connect each pair of vertices with an edge.

A tree of order k is a connected graph with k vertices and $k-1$ edges (links), while a cycle of order k is a cyclic sequence of k edges such that every two consecutive edges and only these have a common vertex. If $p_{\rm ER} \sim c/N$ with $c < 1$, then almost all vertices belong to isolated trees, but at $p_{\rm ER} \sim 1/N$, (i.e. $c = 1$), cycles of all orders appear.

$p_c \sim 1/N$ is the percolation threshold of the (mean field percolation-like) system. At p_c a large cluster forms, that in the asymptotic limit contains infinitely many (proportional to N) vertices (*giant connected component*).

Connectivity distribution.

Probability that a vertex has k edges follows the Poisson distribution

$$P(k) = \binom{N-1}{k} p_{\rm ER}^k (1-p_{\rm ER})^{N-1-k}, \tag{129}$$

and its expectation value being $\lambda = \bar{k} = (N-1)p_{\rm ER}$. For large N

$$P(k) = e^{-\lambda}\lambda^k/k!, \tag{130}$$

Diameter: $L = \langle L_{(i,j)} \rangle_{i,j}$ (average of the number of edges along the shortest path between randomly selected nodes i and j)

$$L_{ER}; \qquad N \sim \bar{k}^{L_{ER}}, \qquad L_{ER} \sim \ln N / \ln \bar{k}$$

The small-world (Watts and Strogats) model.
Crossover from regular lattice to random.

Each link is rewired with probability $p_{\rm WS}$. For $p << 1$, $L = \langle L_{i,j} \rangle \sim N$; for $p_{\rm WS} = 1$, L grows logarithmically with N. For $0.01 < p_{\rm WS}$ the model is small-world, ($L \simeq L_{ER}$), but its clustering coefficient C remains high.

C is an average over the clustering coefficients of nodes; n_i = number of (nearest) neighbours of i. $n_i(n_i-1)/2$ is the total number of possible connections between

them. l_i = number of connections actually present

$$C_i = l_i/[n_i(n_i - 1)/2]$$

ratio of the existing loops to the all possible ones.
Connectivity distribution depends on $p_{\rm WS}$: for $p_{\rm WS} = 0$ we have $P(k) = \delta(k-z)$, where z is the coordination number of the lattice, while for finite $p_{\rm WS}$, $P(k)$ is still peaked around z, but it gets broader.

Equilibrium networks

Real networks typically exhibit growth and rearrangement. Rearrangement: e.g., from fully connected to star (as a function of conditions, hot or cold environment). Classes of topologies as a function of fluctuations

Statistical ensembles for graphs.

We shall consider a set, $\{g_a\}$, of *undirected graphs*, containing N nodes and M links.

In a heat bath at temperature T, the *canonical ensemble* is defined by the partition function

$$Z(T) = \sum_{\{g_a\}} e^{-E_a/T}, \tag{131}$$

where E_a is the *energy* assigned to the different configurations.

Restructuring: "diffusion" of edges or removing the given edge and connecting two randomly selected nodes. Then, the energy difference $\Delta E_{ab} = E_b - E_a$ between the original g_a and the new g_b configurations is calculated and the relocation is carried out following the Metropolis algorithm. If $\Delta E_{ab} = E_b - E_a < 0$ relocation is accepted; otherwise relocation with probability $e^{-\Delta E_{ab}/T}$ only. For $T \to \infty$, the ER random graphs.

The resulting dynamics, by construction, satisfies the *detailed balance condition* (absence of currents).

$$\mathbf{W}_{ab}\mathbf{P}_a = \mathbf{W}_{ba}\mathbf{P}_b$$

where $\mathbf{W}_{ab}$ gives the transition rate from g_a to g_b, and $\mathbf{P}_a$ is the probability to find the system in configuration g_a.

Formal equivalence to a *Kawasaki type lattice gas* on a special lattice, which is the edge-dual graph of the fully connected network. Interacting, diffusing "particles".

$Z(T)$ contains many terms corresponding to *topologically equivalent* graphs (indexing is different only). We consider E_α that depend only on the topology t_α,

$$Z(T) = \sum_{\{t_\alpha\}} \mathcal{N}_\alpha e^{-E_\alpha/T}. \tag{132}$$

Here we introduced $\mathcal{N}_\alpha$ to count the number of configurations belonging to topology

t_α. Expression (132) can be rewritten as

$$Z(T) = \sum_{\{t_\alpha\}} e^{-E_\alpha/T + \ln(\mathcal{N}_\alpha)} = \sum_{\{t_\alpha\}} e^{-F_\alpha/T}, \tag{133}$$

$$F_\alpha = E_\alpha - TS_\alpha, \tag{134}$$

$$S_\alpha = \ln(\mathcal{N}_\alpha), \tag{135}$$

where F_α is the free energy and S_α is the entropy of the topology t_α.

Singularities in the thermodynamic functions derived from $Z(T)$ correspond to *phase transitions in the topology!*.

Order parameters:

i) Either $\Phi = \Phi_s = s_{\max}/M$, the number of edges of the largest connected component of the graph $s_{\max}$ normalized by the total number of edges M,

ii) or $\Phi = \Phi_k = k_{\max}/M$, the highest degree in the graph $k_{\max}$ divided by M. The corresponding conditional free energy $F(\Phi, T)$

$$e^{-F(\Phi,T)/T} = Z(\Phi, T) = \sum_{\{g_a\}_\Phi} e^{-E_a/T}, \tag{136}$$

where $\{g_a\}_\Phi$ is a subset of $\{g_a\}$, consisting of all the graphs with order parameter Φ.

Sudden change in the position of the global minimum of $F(\Phi, T)$ indicates discontinuous (first order) phase transition, gradual shift: continuous phase transition.

Based on the lattice gas analogy we expect that if $\langle k \rangle < 1$, then for a suitable choice of the energy (one that rewards clustering) an ER type dispersed-compact or percolation phase transition occurs at a finite temperature $T(\langle k \rangle)$. The simplest choice is

$$f(s_{\max}) = -s_{\max} = -\Phi_s M. \tag{137}$$

In this case it can be shown that for

$$T > T_{\rm c}(\langle k \rangle) = \frac{1}{\langle k \rangle - 1 - \ln(\langle k \rangle)}, \tag{138}$$

the free energy has a minimum at $\Phi_s = \Phi_s^*(T) = 0$, *i.e.*, the configuration is dispersed. When the temperature drops below $T_{\rm c}(\langle k \rangle)$, the minimum moves away from $\Phi_s = 0$ and a giant component appears. Near the critical temperature $T_{\rm c}(\langle k \rangle)$,

$$\Phi_s^*(T) = 2\frac{T^{-1} - T_{\rm c}^{-1}(\langle k \rangle)}{\langle k \rangle^2 - 3\langle k \rangle + 2}, \tag{139}$$

indicating that we are dealing with a *continuous topological phase transition.*

Next, *energies* are assigned to the *vertices (local).*

$$E = \sum_{i=1}^{N} f(k_i), \tag{140}$$

where k_i denotes the degree (number of neighbors) of vertex i. The "fitness" of an individual vertex depends on its connectivity. Order parameter is $\Phi = \Phi_k = k_{\max}/M$.

The energy $E = -\sum k_i^2$: mapping to the Ising-model
Assign the negative energy $-J$ to all pairs of edges that share a common vertex.

$$E = -\frac{J}{2}\sum_{i=1}^{N} k_i(k_i - 1) = -\frac{J}{2}\sum_{i=1}^{N} k_i^2 + \frac{1}{2}JM, \tag{141}$$

This is like $f(k_i) = -(J/2)k_i^2$. The constant term can be omitted. This form of the energy is in full analogy with the usual definition of the energy

$$E = -J\sum_{<i,j>} n_i n_j \tag{142}$$

of a lattice gas on the edge-dual graph of the fully connected network with nearest neighbor attraction. By measuring the energies (and temperature) in units of J we can set $J = 1$.

Introducing the $s_i \in [-1, 1]$ spin-like variables n_i as $n_i = (1 + s_i)/2$. The energy with the help of the spins is expressed as

$$E = -\frac{1}{4}\sum_{<i,j>} s_i s_j - \frac{1}{2}\sum_{i=1}^{N} s_i - \frac{1}{8}N(N-1). \tag{143}$$

This is similar to a ferromagnetic Ising-model in an external magnetic field. If $n_{\text{empty}} = n_{\text{filled}}$, then $\sum s_i = 0$, no external filed like term.

For the particular form of $f(k_i)$ chosen, the topology with the lowest overall energy is a "star" (for simplicity, we consider $M < N$), where all the M edges are connected to single node.

Energy $E = -\sum k_i \ln k_i$: continuous phase-transition

Another application-motivated choice for the single vertex energy is $f(k_i) = -k_i \ln(k_i)$, inspired, in part, by the logarithmic law of sensation.

In this case the configuration with the lowest energy is a fully connected subgraph [or

almost fully connected if M cannot be expressed as $n(n-1)/2$]. The star configuration is also quite favorable, since both maximal energies scale as $-M \ln M$ to leading order. Amongst the sub-dominant terms in the energy, there is a difference in the order of $\sqrt{M} \ln \sqrt{M}$ between the two, in favor of the fully connected subgraph. As before, we choose the order parameter to be $\Phi = \Phi_k = k_{\max}/M$, since this can easily distinguish between these two configurations: $k_{\max} \approx \sqrt{2M}$ for a fully connected subgraph counting M edges, and $k_{\max} \approx M$ for a star.

Due to the entropic part changing differently, there are two transitions: from disordered to stars and from stars to "full subgraphs".

Growing networks

The scale-free model

Many systems in nature have $P(k)$ following a power-law. Origin: Real world networks *i) grow ii) preferentially.* (e.g. WWW, genetic, friendship networks)

Rules:

(1) *Growth*: Starting with a small number (m_0) of vertices, at every timestep we add a new vertex with $m(\leq m_0)$

(2) *Preferential attachment*: the probability Π that a new vertex will be connected to vertex i depends on k_i

$$\Pi(k_i) = k_i / \sum_j k_j. \tag{144}$$

m is only parameter. $P(k) \sim k^\gamma$, and $\gamma \simeq 3$.

Calculation of $P(k)$

We assume that k is continuous

$$\Pi(k_i) = k_i / \sum_j k_j$$

can be interpreted as a continuous rate of change of k_i (degree of vertex i). Consequently, for a vertex i

$$\frac{\partial k_i}{\partial t} = A\Pi(k_i) = A\frac{k_i}{\sum_{j=1}^{m_0+t-1} k_j}. \tag{145}$$

Since $\sum_j k_j = 2mt$ and that at a time step is $\Delta k = m$, we obtain $A = m$, resulting in

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t}. \tag{146}$$

$$\ln k = 1/2 \ln t + \text{Const}, \qquad k_i = \text{Const } t^{1/2}$$

with the initial condition that vertex i was added to the system at time t_i with connectivity $k_i(t_i) = m$, is

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{0.5}. \tag{147}$$

"rich-gets-richer" phenomenon Using (147), the probability that a vertex has a connec-

tivity $k_i(t)$ smaller than k, $P(k_i(t) < k)$,

$$P(k_i(t) < k) = P(t_i > \frac{m^2 t}{k^2}). \tag{148}$$

We add the vertices at *equal time intervals*,

$$P_i(t_i) = \frac{1}{m_0 + t}. \tag{149}$$

Substituting this

$$P(t_i > \frac{m^2 t}{k^2}) = 1 - P(t_i \leq \frac{m^2 t}{k^2}) = 1 - \frac{m^2 t}{k^2(t + m_0)}. \tag{150}$$

The probability density for $P(k)$ can be obtained using

$$P(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{2m^2 t}{m_0 + t} \frac{1}{k^3}, \tag{151}$$

predicting

$$\gamma = 3, \tag{152}$$

independent of m.

Deterministic scale free models

Deterministic trees. Tree topology is typical and allows analytic calculations.

Step 0: Start with n edges going out radially from a centre.

Step 1: substitute each edge with n new edges "starting" from the centre and "ending" on a circle (circle for visualization only).

Step 2: From every m-th node out of the the n^2 nodes on the circle, draw n new edges, so that each new edge ends on a new concentric circle.

Next, Step 1 and Step 2 are repeated many times so that each time an edge is substituted with n edges stating from an edge's inner end, and Step 2 is carried out only for edges in the outmost layer.

After q iterations, there are

number of nodes	with number of edges
j^0	n^{q+1}
j^1	n^q
.....	.....

j^q n

with $j = n^2/m$, q is the number of iterations, single edges leading "back" not counted. For $n = 4, m = 2, j = 8$. The ratio of nodes having k edges as $P'(k)$,

$$P'(4k) = P'(k)/8$$

If $P'(k)$ scales as $k^{-\delta}$, $\delta = \ln 8/\ln 4 = 3/2$. The corresponding "smooth" distribution:

$$P(k) \sim k^{-\gamma} \sim k^{-(3/2+1)}$$

It directly follows from the above, that in general, $\gamma = \ln j/\ln n + 1$.

Flux distribution assuming that an amount of flux $\nu = 1$ is entering the network at its "outmost" edges and the outflow takes place at the central node.

number of edges	with amount of flux
j^0	nj^q
j^1	nj^{q-1}
.....	.....
$j^q n$	1

In the above example $n = 4, m = 2, j = 8$. Ration of nodes having flux ν is $P'(\nu)$, we have

$$P'(8\nu) = P'(\nu)/8$$

(there are 8 times less edges having 8 times more flux). $P'(\nu) \sim \nu^{-\Delta}$, with $\Delta = \ln 8/\ln 8 = 1$. The corresponding continuous distribution and as above

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

so that $\alpha = 2$.

Here we obtained for the flux distribution an exponent $\alpha = 2$ independent of the two parameters. Seems to be universal. Independent of the form of boundary conditions. Two classes one with exponent 2 one with exponent 2.2.

Metabolic fluxes. Internet traffic.

Hierarchical, no loops, zero C

Only loops (triangles), $C = 1$

for both:

$i = 1 \qquad N(k) = 1 \qquad k = 14$

$i = 1 \qquad N(k) = 2 \qquad k = 6$

$i = 1 \qquad N(k) = 6 \qquad k = 2$

after n iterations $(2/3)3^{n-1}$ vertexes with degree $2^{i+1} - 2$.

$$N(k/2) = 3N(k), \qquad N(k) \sim k^{-\gamma}$$

$$k/2^{-\gamma} = 3k^{-\gamma}, \qquad 2^{\gamma} = 3$$

Delta function type distribution with peaks separated by a distance proportional to k.

For the density distribution (assuming that the delta functions are spread over intervals of length k ($N(k)$ values should be renormalized with $1/k$

$$P(k) \sim k^{-(1+\ln 3/\ln 2)}$$

Most advanced: scaling of $C(k) \sim k^{-\omega}$ (hierarchically connected clusters!)

Stability against attacks: random, intentional

stable against random, unstable against intentional

Spreading of infection:

Model: i) healthy vertex is infected with rate ν if has infected neighbour, ii) sick vertexes become healthy with rate δ.

Effective spreading rate: $\lambda = \nu/\delta$. It was shown that for arbitrary $P(k)$

$$\lambda_c = \bar{k}/\bar{k}^2$$

This is finite for ordinary networks. For scale free nets with $\gamma < 3$ the standard deviation $\bar{k}^2$ diverges, epidemics spreads for arbitrarily small λ (see viruses on the WWW, or HIV epidemics - sexual networks are scale free...(Nature, 2001))

Network communites/clusters/modules

Large networks are bound to be structured

Internal organization is hierarchical

Finding dense sub-graphs: i) divisive/agglomerative methods (has serious problems), ii) clique percolation

Clique percolation

Clique: maximal subgraph

k-cliques, the central objects of our investigation, are defined as complete (fully connected) subgraphs of k vertices

We also introduce a few new notions specific to our problem. (i) *k-clique adjacency*: two k-cliques are adjacent if they share $k-1$ vertices, i.e., if they differ only in a single vertex. (ii) *k-clique chain*: a subgraph, which is the union of a sequence of adjacent k-cliques. (iii) *k-clique connectedness*: two k-cliques are k-clique-connected if they are parts of a k-clique chain. (iv) *k-clique percolation cluster (or component)*: it is a maximal k-clique-connected subgraph, i.e., it is the union of all k-cliques that are k-clique-connected to a particular k-clique.

A giant k-clique component appears in an E-R graph at $p = p_c(k)$, where

$$p_c(k) = \frac{1}{[(k-1)N]^{\frac{1}{k-1}}}. \tag{153}$$

Obviously, for $k = 2$ this result agrees with the known percolation threshold ($p_c = 1/N$) for E-R graphs, because 2-clique connectedness is equivalent to regular (edge) connectedness.

Expression (153) can be obtained by requiring that after rolling a k-clique template from a k-clique to an adjacent one (by relocating one of its vertices), the expectation value of the number of adjacent k-cliques, where the template can roll further (by relocating another of its vertices), be equal to 1 at the percolation threshold.

The above expectation value can be estimated as $(k-1)(N-k-1)p^{k-1}$, where the first term $(k-1)$ counts the number of vertices of the template that can be selected for the next relocation, the second term $(N-k-1)$ counts the number of potential destinations for this relocation, out of which only the fraction p^{k-1} is acceptable, because each of the new $k-1$ edges (associated with the relocation) must exist in order to obtain a new k-clique. For large N, it becomes $(k-1)Np_{\rm c}^{k-1} = 1$

Thus, in the random graph there is a threshold, at which an "infinite" community appears and, as in ordinary percolation, around this point clusters of all sizes occur.

N^*, is the number of vertices belonging to the largest cluster. *Order parameter* associated with this choice as the relative size of the largest cluster:

$$\Phi = N^*/N. \tag{154}$$

The other choice is the number $\mathcal{N}^*$ of k-cliques of the largest percolation cluster (or equivalently, the number of vertices of the largest component in the k-clique adjacency graph). The associated order parameter

$$\Psi = \mathcal{N}^*/\mathcal{N}, \tag{155}$$

where $\mathcal{N}$ denotes the total number of k-cliques in the graph (or the total number of vertices in the adjacency graph). $\mathcal{N}$ can be estimated as

$$\mathcal{N} \approx \binom{N}{k} p^{k(k-1)/2} \approx \frac{N^k}{k!} p^{k(k-1)/2}, \tag{156}$$

because k different vertices can be selected in $\binom{N}{k}$ different ways, and any such selection makes a k-clique only if all the $k(k-1)/2$ edges between these k vertices exist, each with probability p.

The largest component at the percolation threshold, becomes infinitely large (becomes a giant component) in the $N \to \infty$ limit. Erdős and Rényi showed that for the random graphs they introduced, the size of the largest component $\mathcal{N}^*$ at $p = p_c \equiv 1/N$ diverges with the system size as $N^{2/3}$, or equivalently, the order parameter Ψ scales as $N^{-1/3}$.

Since the giant component at the threshold has a tree-like structure, its number of vertices, N^*, also diverges as $N^{2/3}$.

If we assume, that the k-clique adjacency graph is like an E-R graph, then at the threshold the size of its giant component $\mathcal{N}_c^*$ scales as $\mathcal{N}_c^{2/3}$. Plugging $p = p_c$ from Expression (153) into Eq. (156) and omitting the N-independent factors we get the scaling

$$\mathcal{N}_c \sim N^{k/2} \tag{157}$$

for the total number of k-cliques. Thus, the size of the giant component $\mathcal{N}_c^*$ is expected to scale as $\mathcal{N}_c^{2/3} \sim N^{k/3}$ and the order parameter Ψ_c as $\mathcal{N}_c^{2/3}/\mathcal{N}_c \sim N^{-k/6}$.

This is valid, however, only if $k \leq 3$. For $k > 3$ it predicts that the number of k-cliques of the giant percolation cluster, *i.e.*, the number of vertices of the giant component in the k-clique adjacency graph, $\mathcal{N}_c^{2/3} \sim N^{k/3}$, grows faster than N. Problem.

On the other hand, in analogy with the structure of the giant component of the classical E-R problem, we expect that the giant component in the adjacency graph also has a tree-like structure at the threshold.

As a consequence, almost every vertex of the adjacency graph corresponds to a vertex of the original graph. Thus, in the adjacency graph the giant component should not grow faster than N at the threshold. Therefore, for $k > 3$ we expect that $\mathcal{N}_c^* \sim N$, and using Eq. (157), $\Psi_c = \mathcal{N}_c^*/\mathcal{N}_c \sim N^{1-k/2}$. In summary:

$$\Psi_c \sim \begin{cases} N^{-k/6} & \text{for } k \leq 3 \\ N^{1-k/2} & \text{for } k \geq 3 \end{cases} . \tag{158}$$

Finding overlapping communities

An order k community is a k-clique percolation cluster. This expresses the observation that more densely connected parts have many fully connected small subgraphs in them.

Overlaps are important!

Each node i of a network can be characterised by a *membership number* m_i, which is the number of communities the node belongs to. In turn, any two communities α and β can share $n_{\alpha,\beta}^{ov}$ nodes, which we define as the *overlap size* between these communities. Naturally, the communities also constitute a *network* with the overlaps being their links. The number of such links of community α can be called as its *community degree*, d_α^{comm}. Finally, the size n_α^s of any community α can most naturally be defined as the number of its nodes.

Corresponding distributions

Weighted networks with w_i being the "strength" of the $i-th$ link (connecting two given nodes).

Find the right k and w_{crit} from the condition that the resulting network of communities should be maximally structured. In practice, we decrease the threshold w^* (below which the links are neglected) until the largest community becomes twice as big as the second largest one.

Results show scaling, indicating hierarchy. Community degree distribution has an exponential decay for small degree values.

Collaboration networks

$P(k) \sim k^{\gamma}, \qquad \gamma \simeq 2,3$

$\Pi(k_1, k_2)$: probability of a new link between authors with k_1 and k_2 collaborators

A natural hypothesis is to assume that $\Pi(k_1, k_2) \sim k_1 k_2$

Modelling the web of science

$k_i(t)$ is the number of links node i has at time t; $T(t)$ and $N(t)$ are the total number of links and total number of nodes at time t. New researchers join the field at a constant rate

$$N(t) = \beta t. \tag{159}$$

The average number of links per node in the system at time t is thus given by:

$$< k >= \frac{T(t)}{N(t)}. \tag{160}$$

Probability that between node i and j a new internal link is created as

$$\Pi_{ij} = \frac{k_i k_j}{\sum'_{s,m} k_s k_m} \, 2N(t) \, a, \tag{161}$$

where the prime sign indicates that the summation is done for $s \neq m$ values, a is number of newly created internal links per node in unit time.

If node i has k_i links, the probability that an incoming node will connect to it

$$\Pi_i = b \, \frac{k_i}{\sum_j k_j}, \tag{162}$$

where b is the average number of new links that an incoming node creates.

(Here the number of authors on a paper is constant; this is an approximation).

New links join the system with a constant rate, β,

$$\frac{dk_i}{dt} = \frac{b\beta k_i}{\sum_j k_j} + 2 \, N(t) \, a \sum_j{}' \frac{k_i k_j}{\sum'_{s,m} k_s k_m}. \tag{163}$$

internal and external preferential attachment rules:

$$\sum_i k_i = T(t) = \int_0^t 2 \left[N(t')a + b\beta \right] dt' = t\beta(at + 2b). \tag{164}$$

$$< k >= \frac{T(t)}{N(t)} = at + 2b, \tag{165}$$

$$\sum_{s,m}{}' k_s k_m = \sum_s k_s \sum_m k_m - \sum_m k_m^2 \approx \left(\sum_i k_i \right)^2 . \tag{166}$$

We have used here the fact that $T(t)^2$ depends on N^2, while $\sum_i k_i^2$ depends only linearly on N (we investigate the $N \to \infty$ limit). From here:

$$\frac{dk_i}{dt} = \frac{bk_i}{t(at + 2b)} + \frac{2k_i a}{at + 2b}. \tag{167}$$

Introducing the notation $\alpha = a/b$, we obtain:

$$\frac{dk_i}{dt} = \frac{k_i}{t} \frac{2t\alpha + 1}{t\alpha + 2}. \tag{168}$$

$$k_i(t) \sim (t/t_i)^{1/2} \qquad \text{if} \quad t \ll \alpha$$

$$k_i(t) \sim (t/t_i)^2 \qquad \text{if} \quad t \gg \alpha$$

The distribution function for the t_i in the $[0, t]$ interval is simply $\rho(t) = 1/t$. $P(k)$ can be obtained after determining the $t_i(k_i)$ dependence

$$P(k) = -\rho(t) \left. \frac{dt_i}{dk_i} \right|_k =$$

$$= \frac{1}{t}(tk^{-3}) \qquad \text{if} \quad t \ll \alpha$$

$$= \frac{1}{t}(tk^{-3/2}) \qquad \text{if} \quad t \gg \alpha$$

Number of authors with $N(k)$ has two scaling regimes, crossover k_c scales with time