Complexity, Phase Transitions, and Inference

Cristopher Moore, Santa Fe Institute

with Aurelien Decelle, Lenka Zdeborová, Florent Krzakala, Xiaoran Yan, Yaojia Zhu, Cosma Shalizi, Lise Getoor, Aaron Clauset, Mark Newman, Elchanan Mossel, Joe Neeman, Allan Sly, Pan Zhang, Jess Banks, Praneeth Netrapalli, Thibault Lesieur, Caterina de Bacco, Roman Vershynin, and Jiaming Xu

Statistical inference $\Leftrightarrow$ statistical physics

Statistical inference $\Leftrightarrow$ statistical physics

How can we find patterns in noisy data?

Statistical inference $\Leftrightarrow$ statistical physics

How can we find patterns in noisy data?
Phase transitions and fundamental limits

Statistical inference $\Leftrightarrow$ statistical physics

How can we find patterns in noisy data?
Phase transitions and fundamental limits
Optimal algorithms

Statistical inference $\Leftrightarrow$ statistical physics

How can we find patterns in noisy data?
Phase transitions and fundamental limits
Optimal algorithms
Information vs. efficient computation

Statistical inference $\Leftrightarrow$ statistical physics

How can we find patterns in noisy data?
Phase transitions and fundamental limits
Optimal algorithms
Information vs. efficient computation
Interdisciplinary exchange

Why least squares?

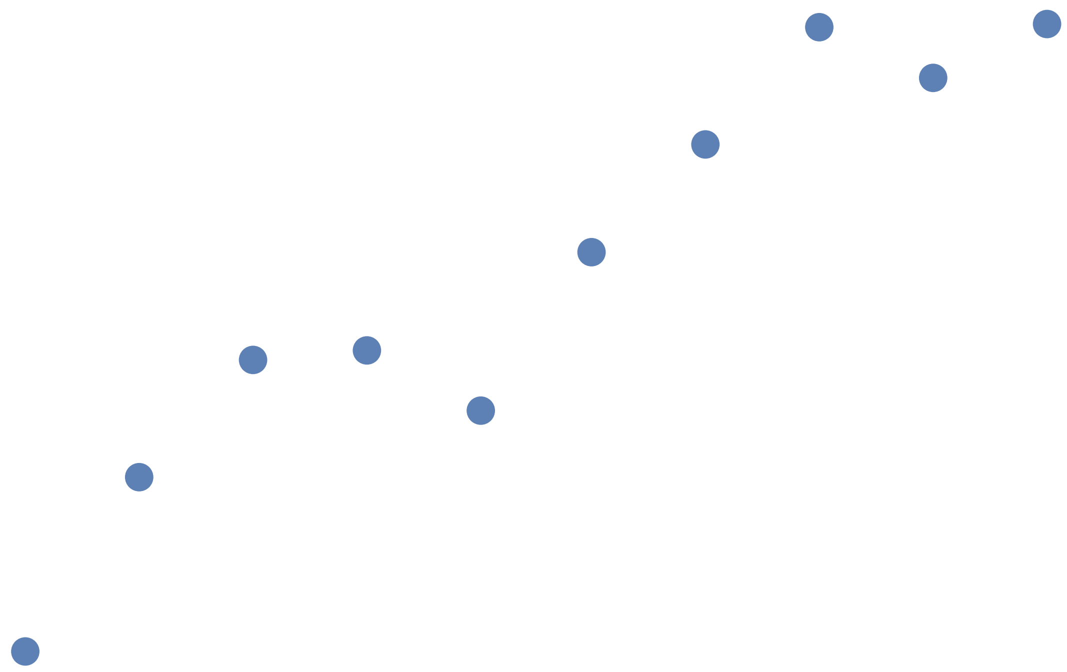
Why least squares?

the most common way to fit a line to noisy data

Why least squares?

the most common way to fit a line to noisy data

data points $Y = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$

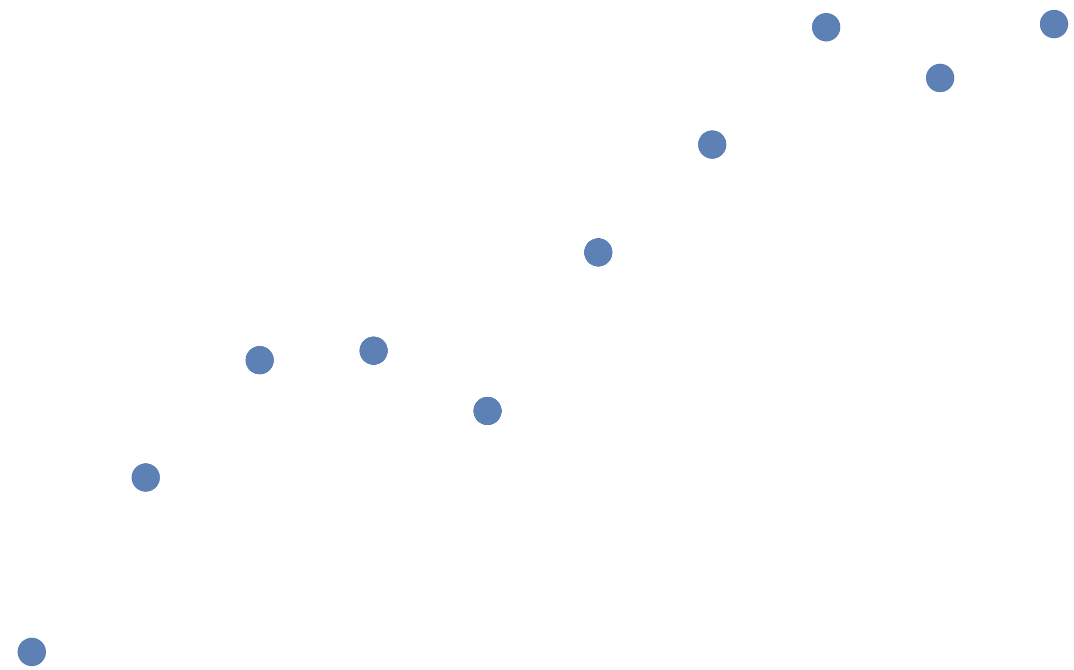


Why least squares?

the most common way to fit a line to noisy data

data points $Y = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$

model: $y_i = ax_i + b$



Why least squares?

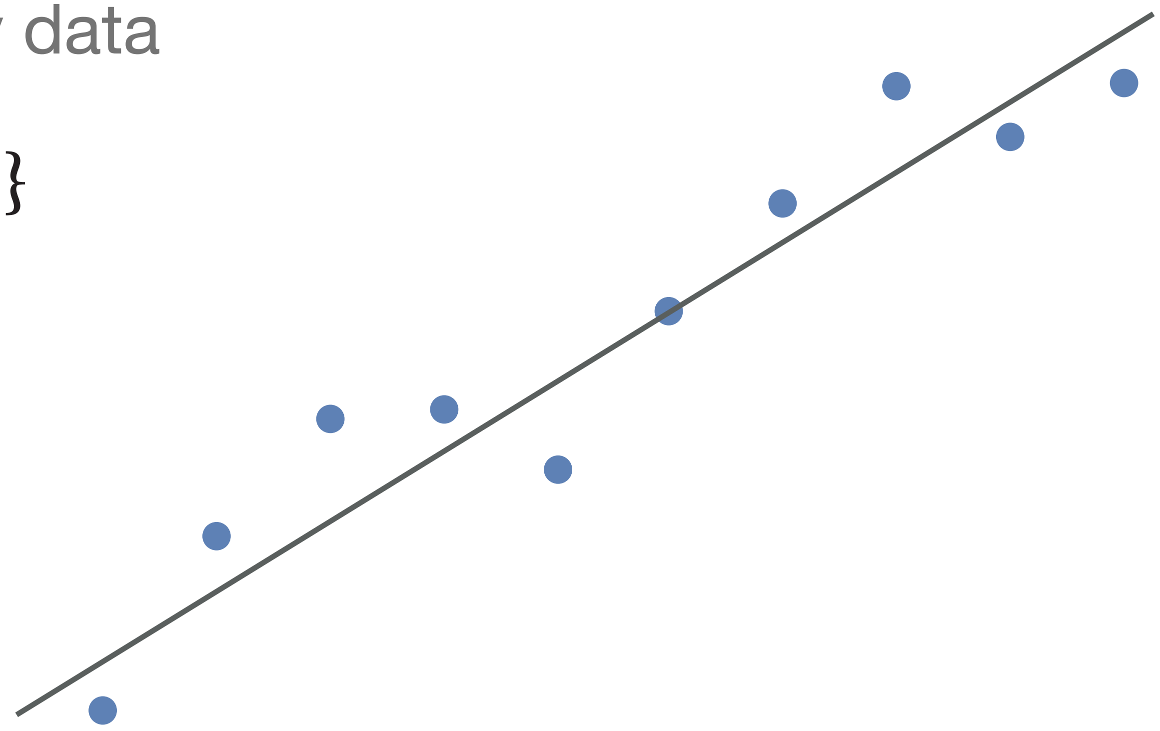
the most common way to fit a line to noisy data

data points $Y = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$

model: $y_i = a x_i + b$

find *a,b* that minimize

$$\sum_i (y_i - (a x_i + b))^2$$



Why least squares?

the most common way to fit a line to noisy data

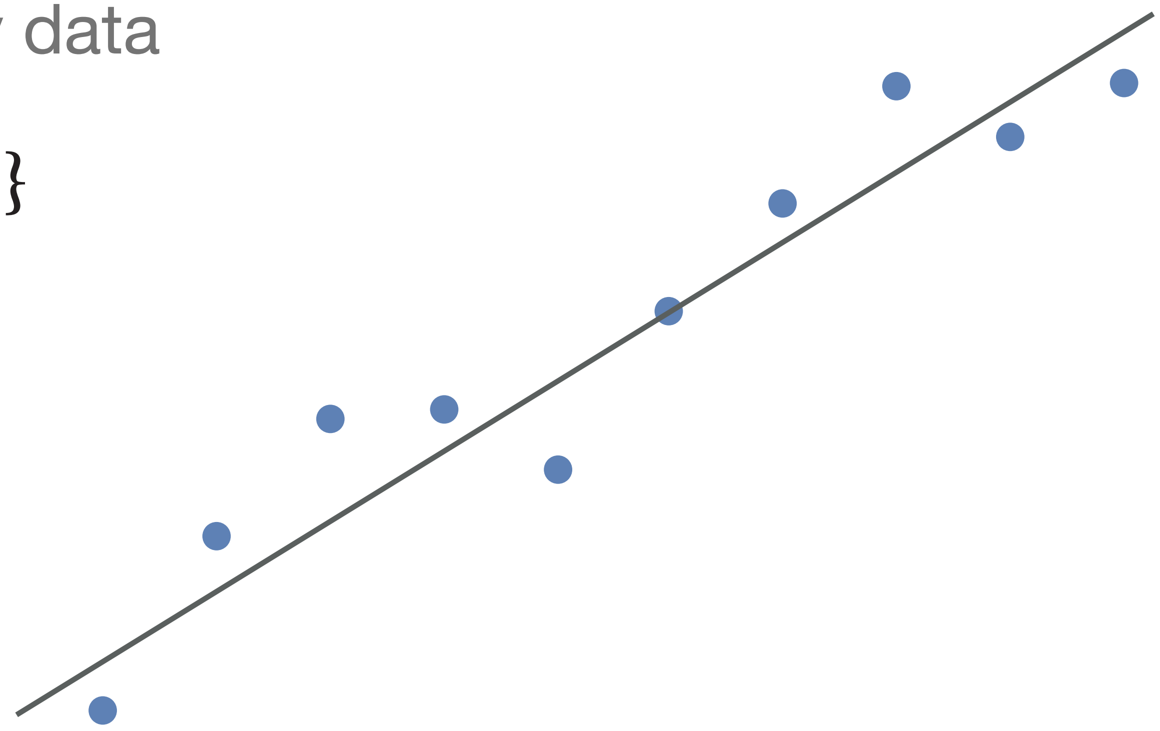
data points $Y = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$

model: $y_i = ax_i + b$

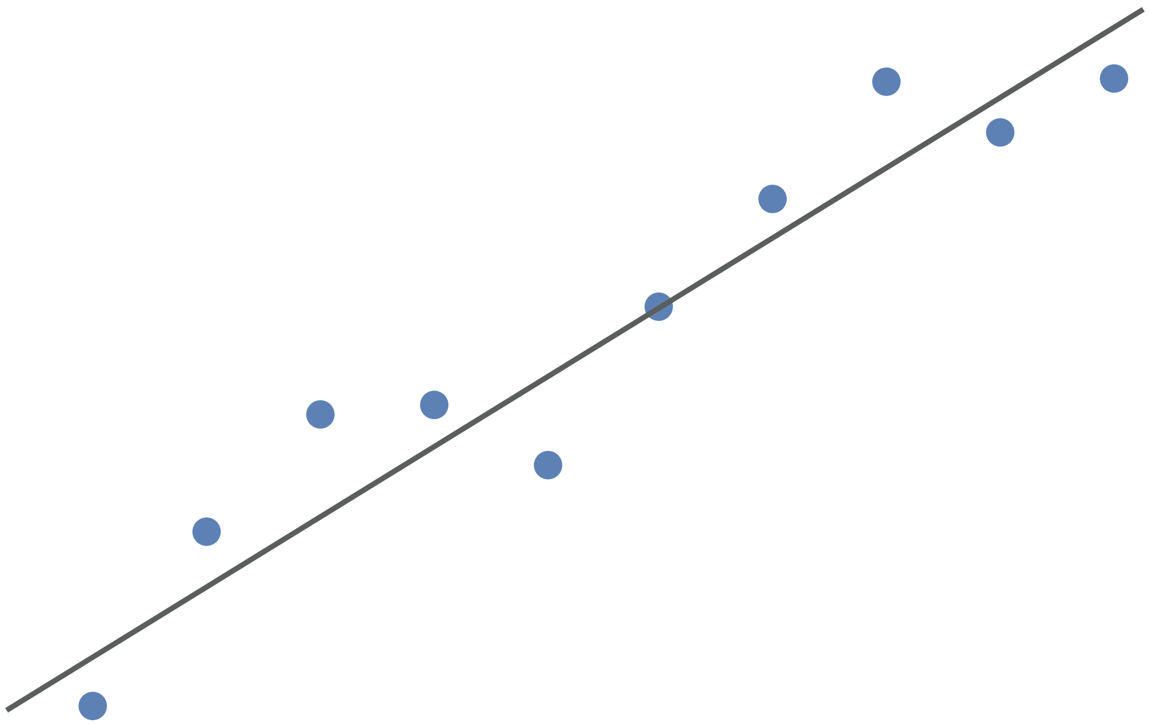
find *a,b* that minimize

$$\sum_i (y_i - (ax_i + b))^2$$

but why?

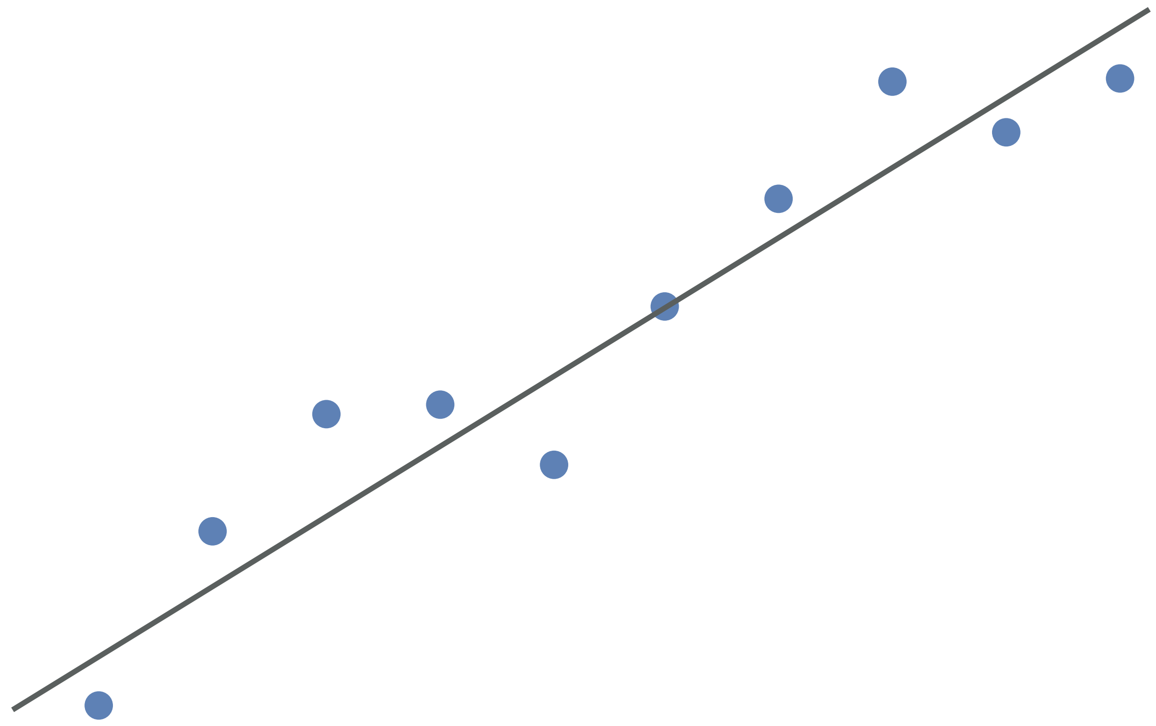


A model of noise



A model of noise

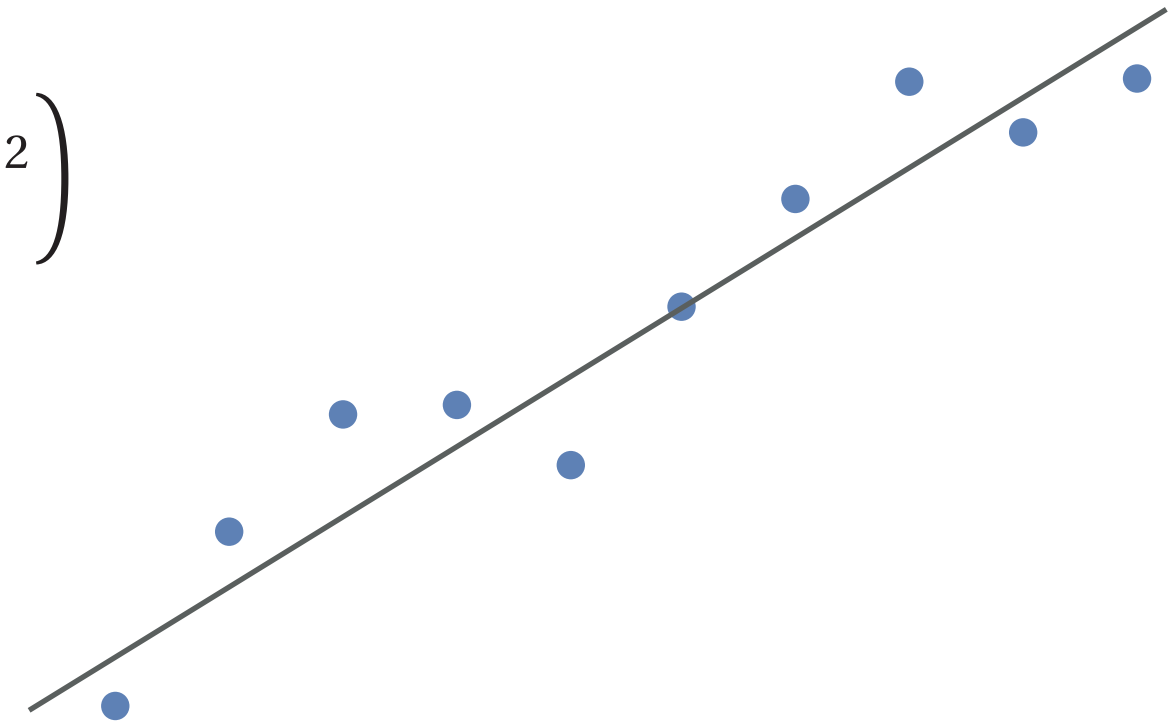
a model with noise: $y_i = ax_i + b + w$



A model of noise

a model with noise: $y_i = ax_i + b + w$

where *w* is Gaussian, $P(w) \propto \exp\left(-\frac{1}{2\sigma} w^2\right)$



A model of noise

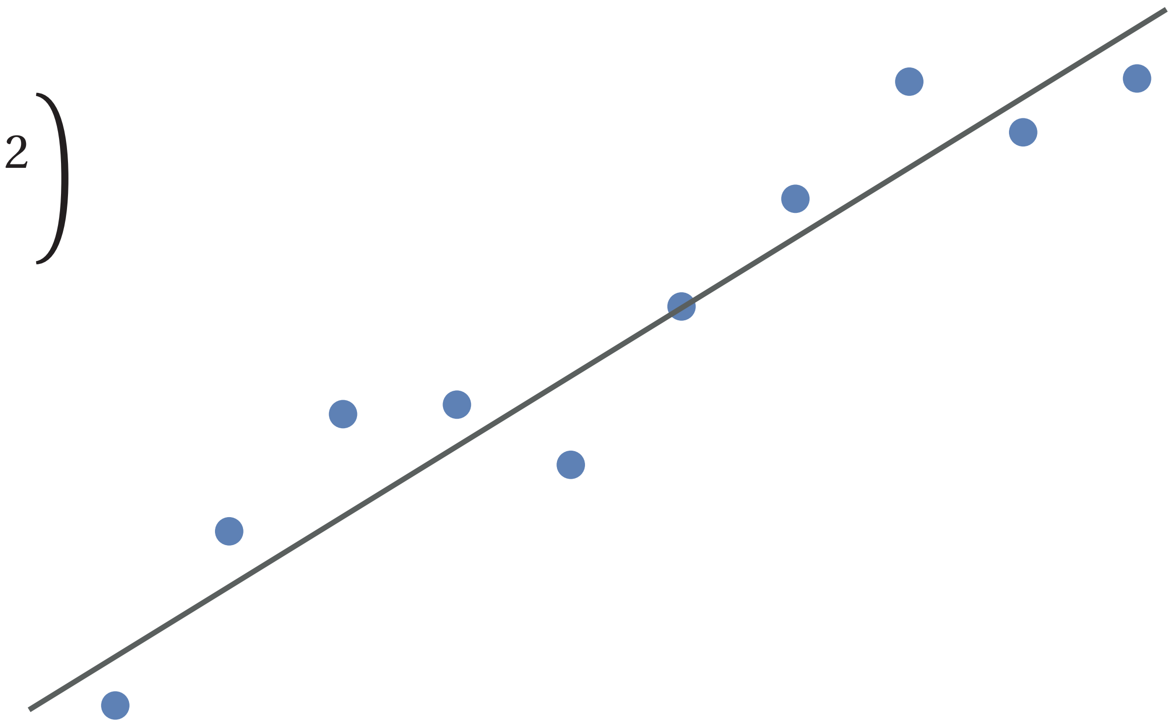
a model with noise: $y_i = ax_i + b + w$

where *w* is Gaussian, $P(w) \propto \exp\left(-\frac{1}{2\sigma} w^2\right)$

total probability of the data is

$$P(Y \mid a, b) = \prod_i P(y_i \mid a, b)$$

$$\propto \exp\left(-\frac{1}{2\sigma} \sum_i (y_i - (ax_i + b))^2\right)$$



A model of noise

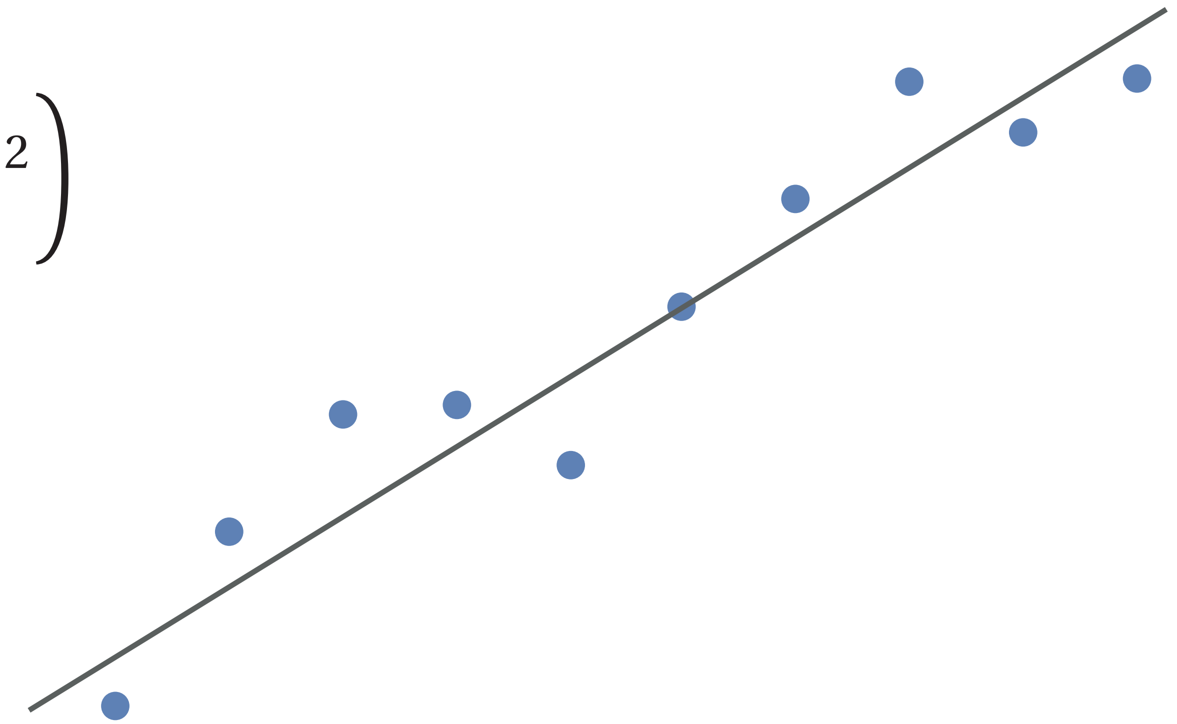
a model with noise: $y_i = a x_i + b + w$

where *w* is Gaussian, $P(w) \propto \exp\left(-\frac{1}{2\sigma} w^2\right)$

total probability of the data is

$$P(Y \mid a, b) = \prod_i P(y_i \mid a, b)$$

$$\propto \exp\left(-\frac{1}{2\sigma} \sum_i (y_i - (a x_i + b))^2\right)$$



Bayes: posterior (with flat prior) $P(a, b|Y) \propto P(Y|a, b)$

A model of noise

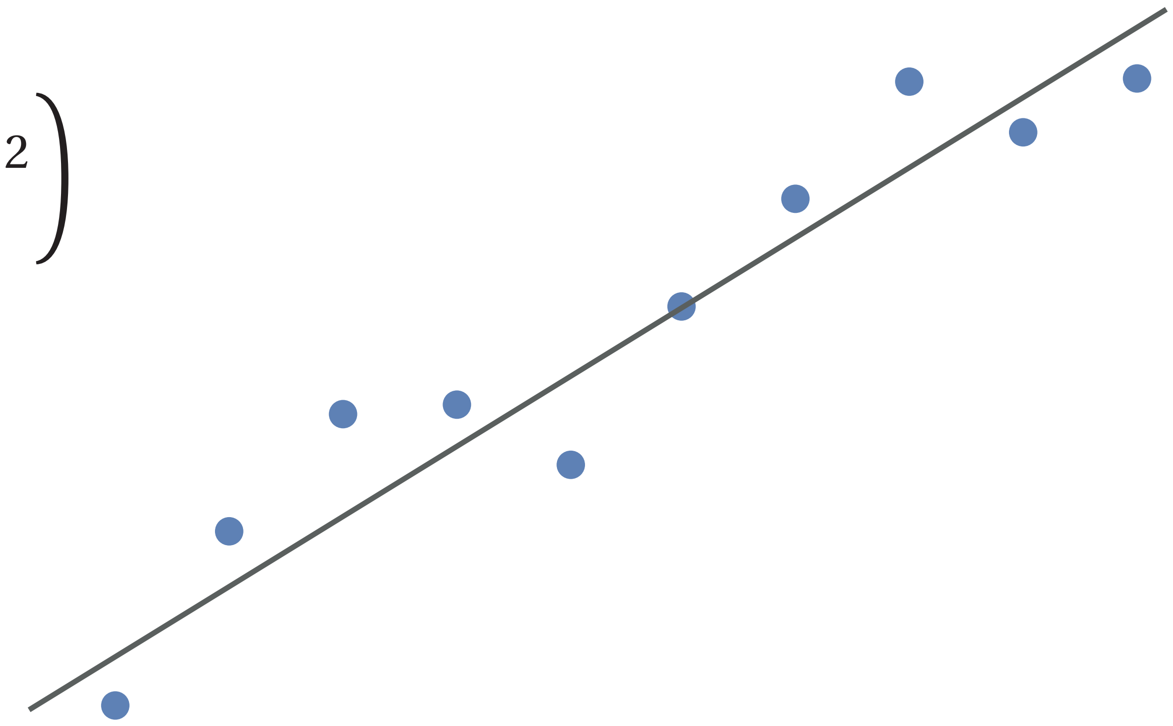
a model with noise: $y_i = a x_i + b + w$

where *w* is Gaussian, $P(w) \propto \exp\left(-\frac{1}{2\sigma} w^2\right)$

total probability of the data is

$$P(Y \mid a, b) = \prod_i P(y_i \mid a, b)$$

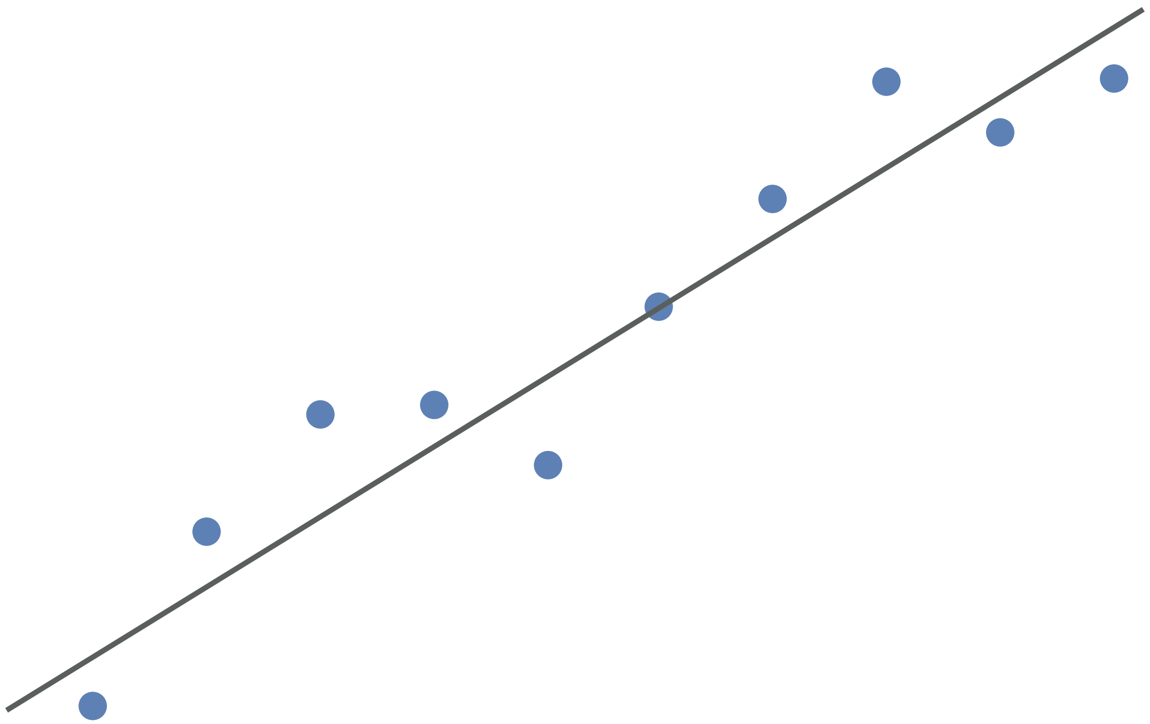
$$\propto \exp\left(-\frac{1}{2\sigma} \sum_i (y_i - (a x_i + b))^2\right)$$



Bayes: posterior (with flat prior) $P(a, b|Y) \propto P(Y|a, b)$

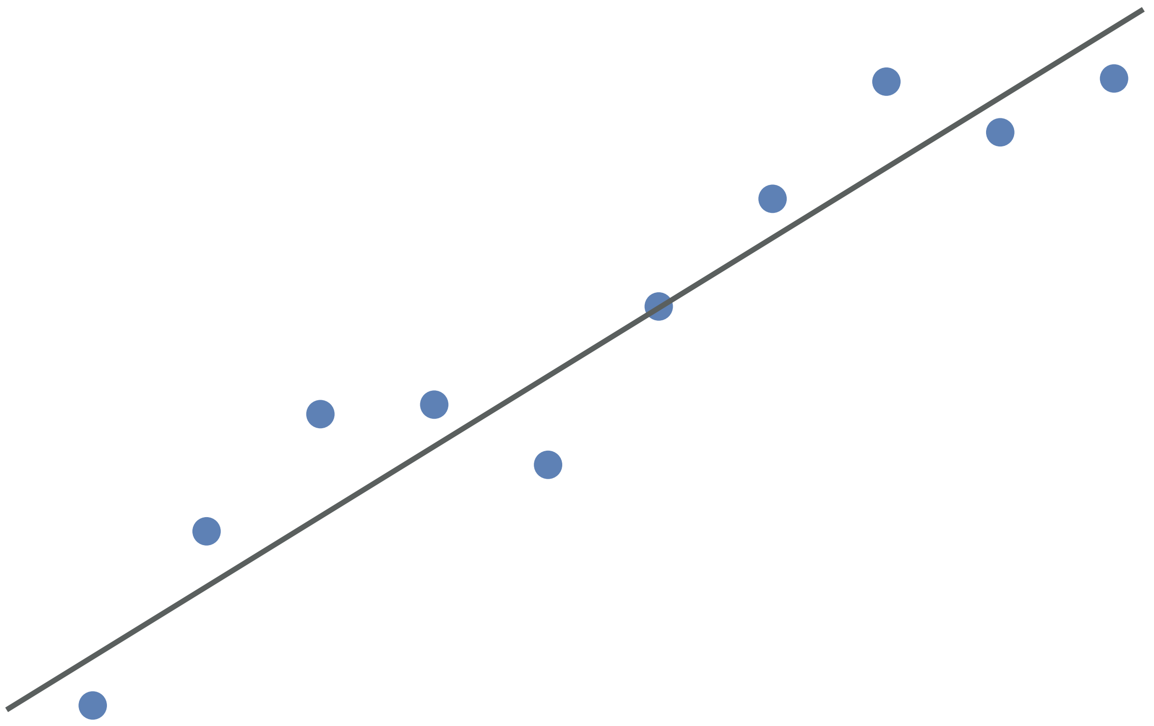
least squares = maximum likelihood estimate

From probability to energy



From probability to energy

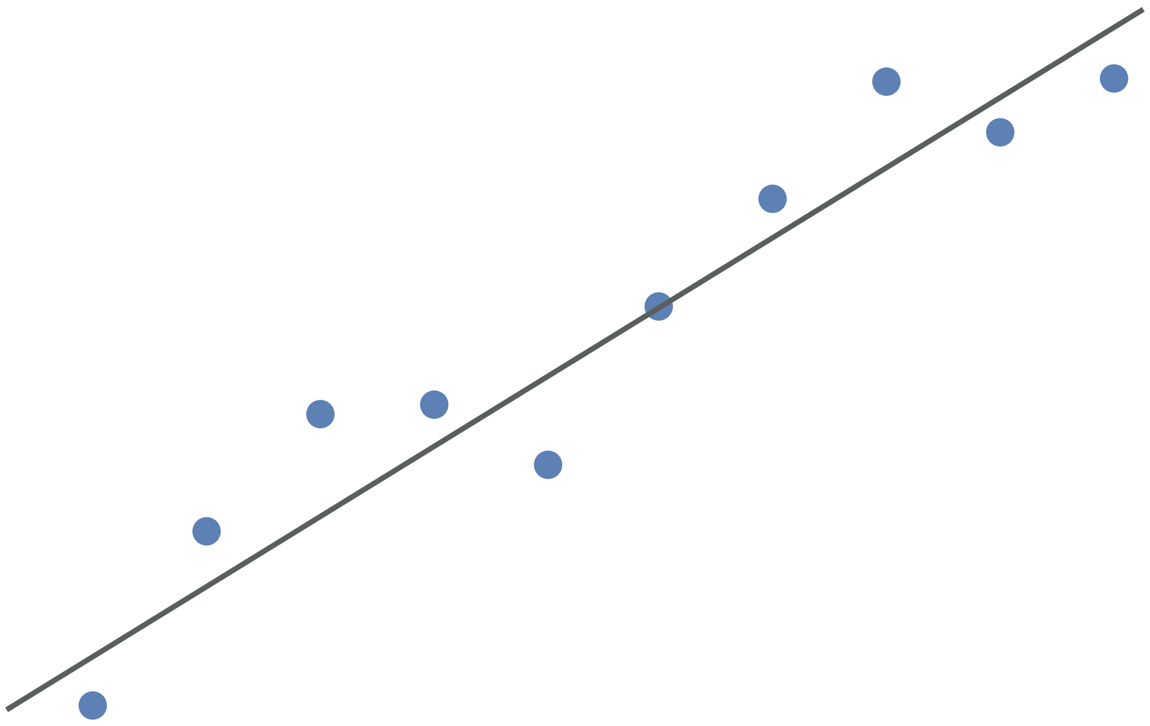
define the energy of (a,b) as $E = -\log P$



From probability to energy

define the energy of (a,b) as $E = -\log P$

$$E = \frac{1}{2\sigma} \sum_i (y_i - (a x_i + b))^2$$



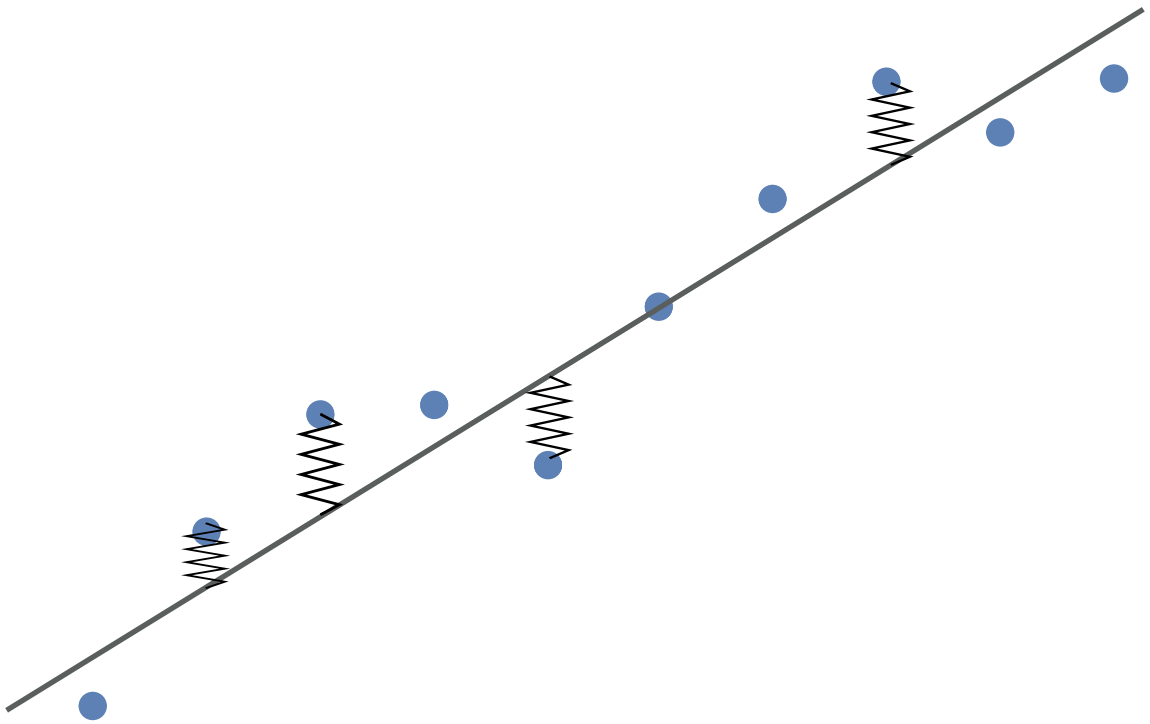
From probability to energy

define the energy of (*a*,*b*) as $E = -\log P$

$$E = \frac{1}{2\sigma} \sum_i (y_i - (a x_i + b))^2$$

springs between the model and data

$$E = \frac{1}{2} k x^2$$



From probability to energy

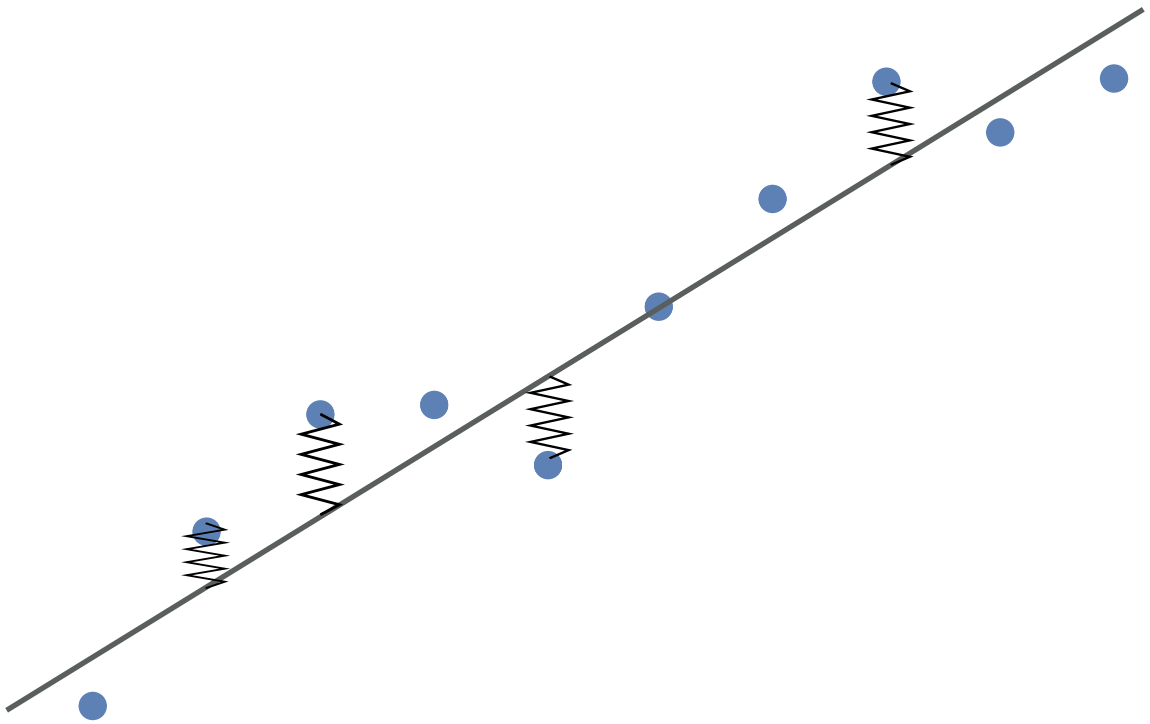
define the energy of (*a*,*b*) as $E = -\log P$

$$E = \frac{1}{2\sigma}\sum_i (y_i - (ax_i + b))^2$$

springs between the model and data

$$E = \frac{1}{2}kx^2$$

maximizing *P* = minimizing *E*



From probability to energy

define the energy of (*a*,*b*) as $E = -\log P$

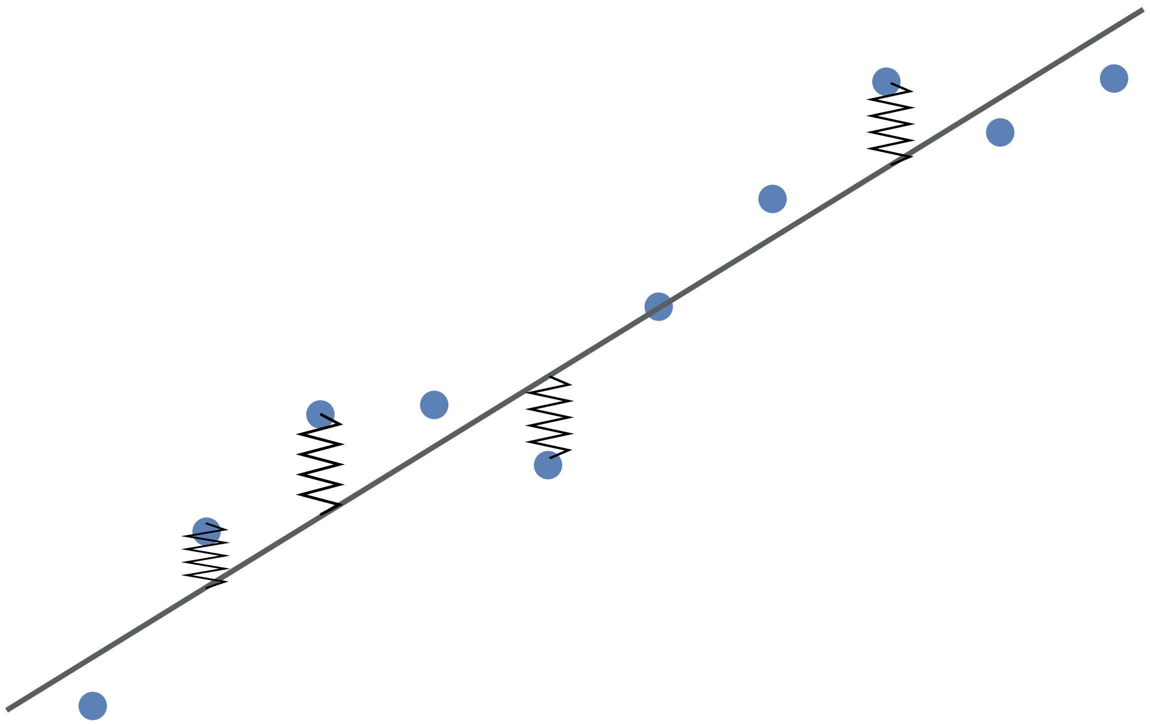
$$E = \frac{1}{2\sigma}\sum_i (y_i - (ax_i + b))^2$$

springs between the model and data

$$E = \frac{1}{2}kx^2$$

maximizing *P* = minimizing *E*

maximum likelihood estimate = ground state



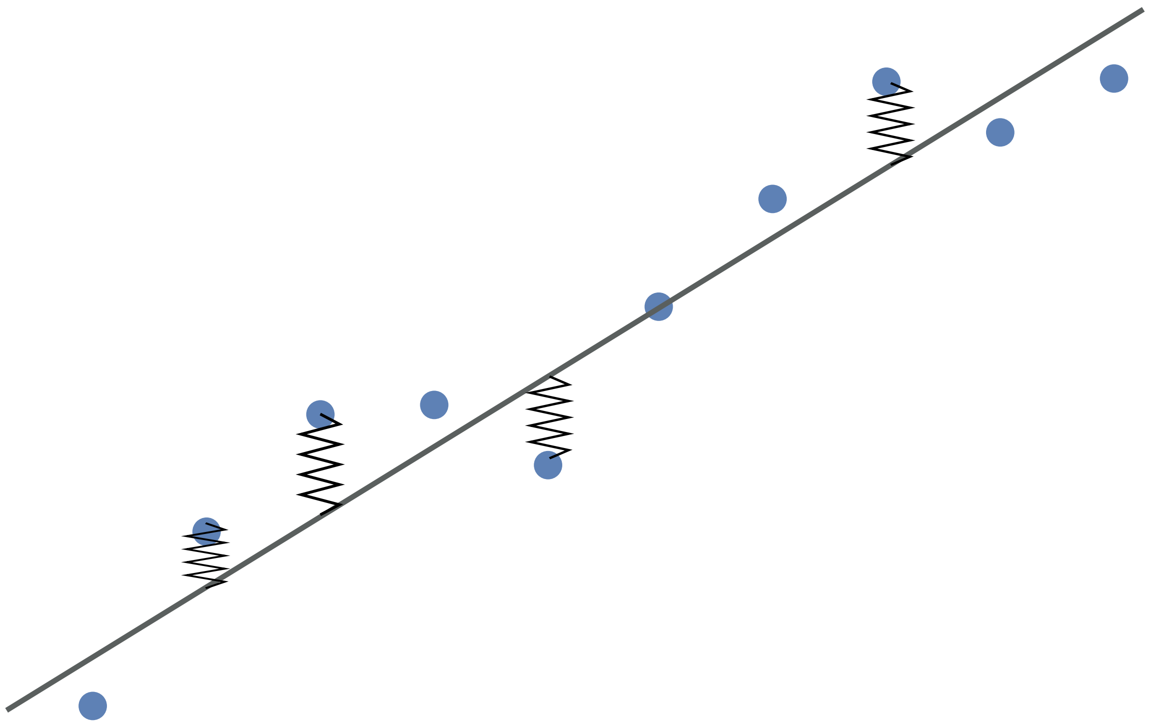
From probability to energy

define the energy of (a,b) as $E = -\log P$

$$E = \frac{1}{2\sigma} \sum_i (y_i - (a x_i + b))^2$$

springs between the model and data

$$E = \frac{1}{2} k x^2$$

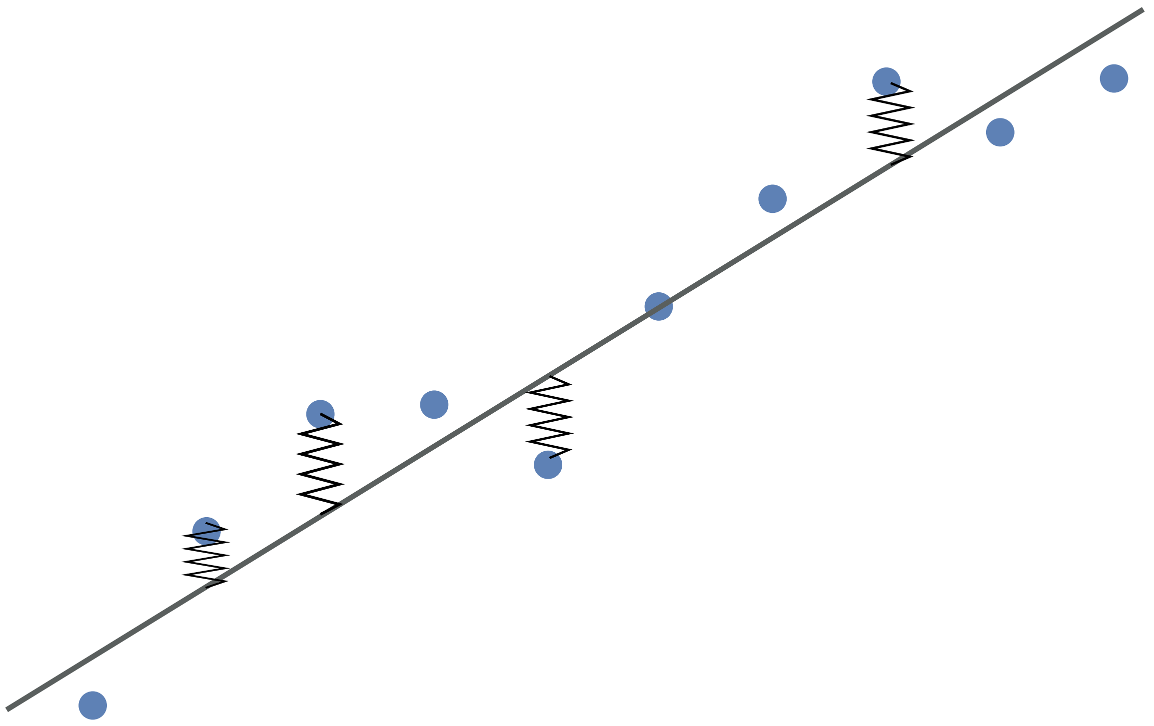


maximizing P = minimizing E

maximum likelihood estimate = ground state

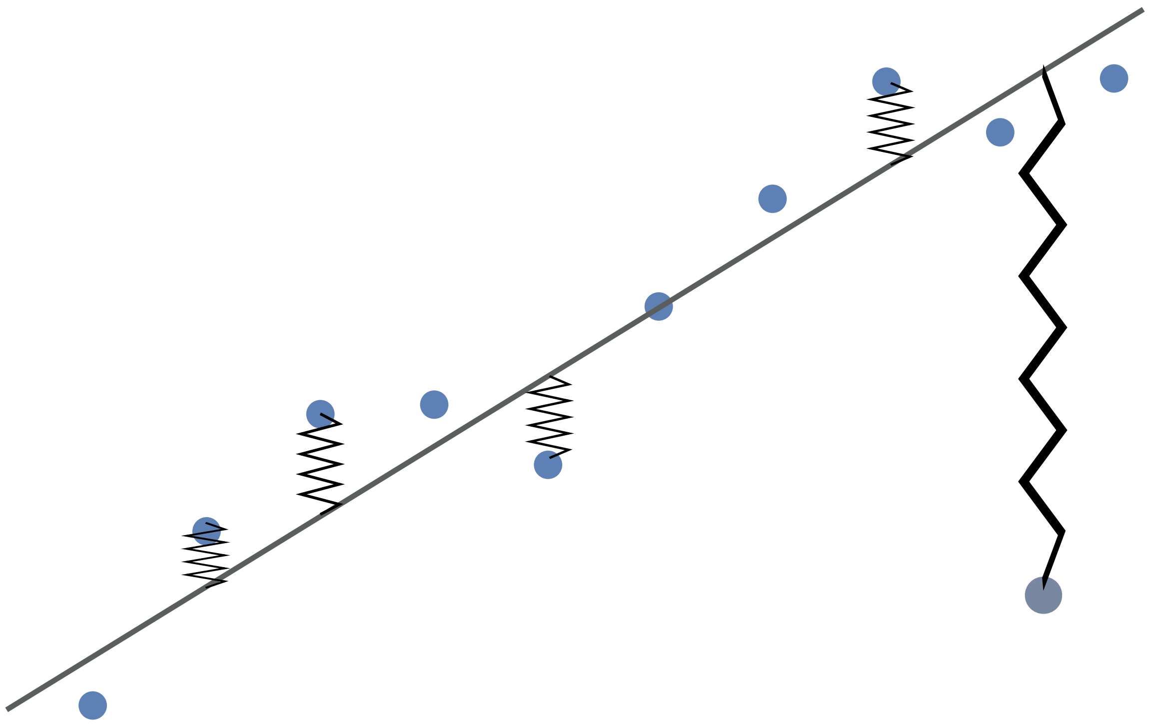
but what if the energy were different?

Changing the model



Changing the model

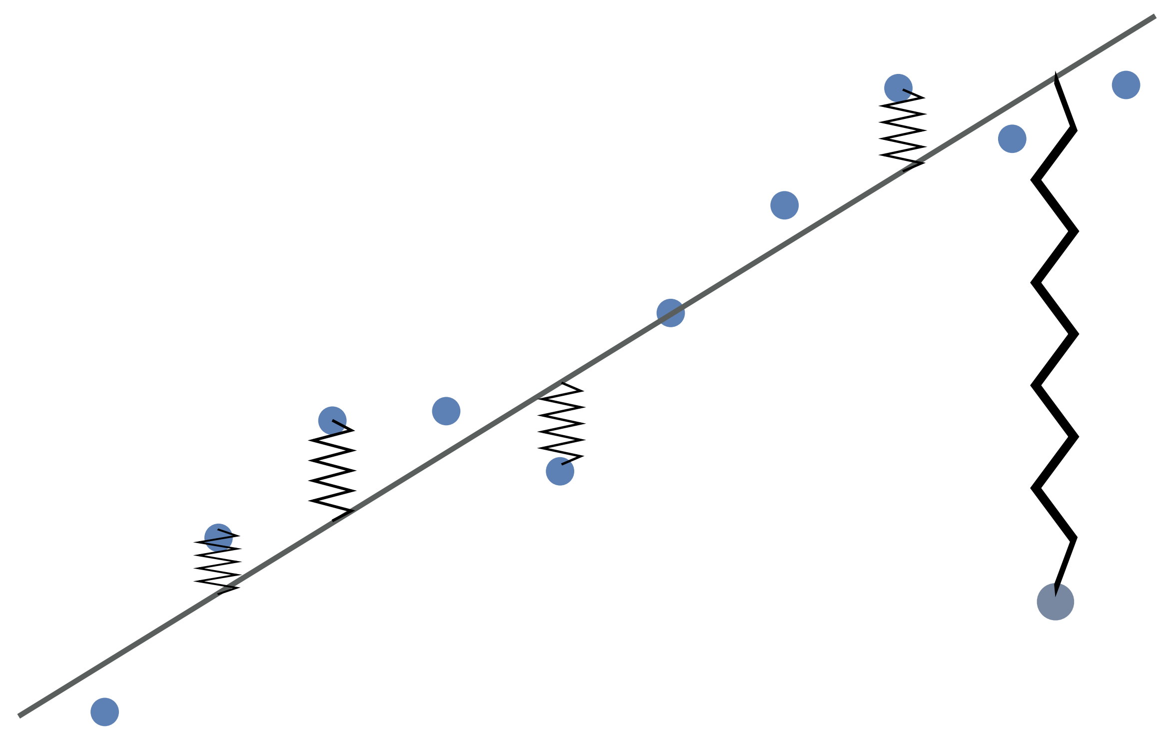
outliers skew our estimates

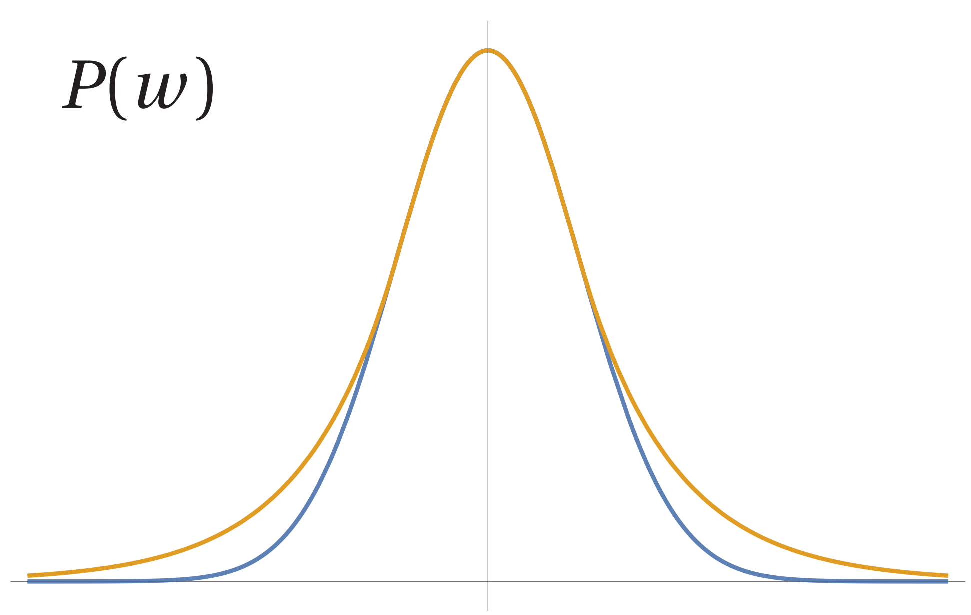


Changing the model

outliers skew our estimates

use a noise model with heavier tails



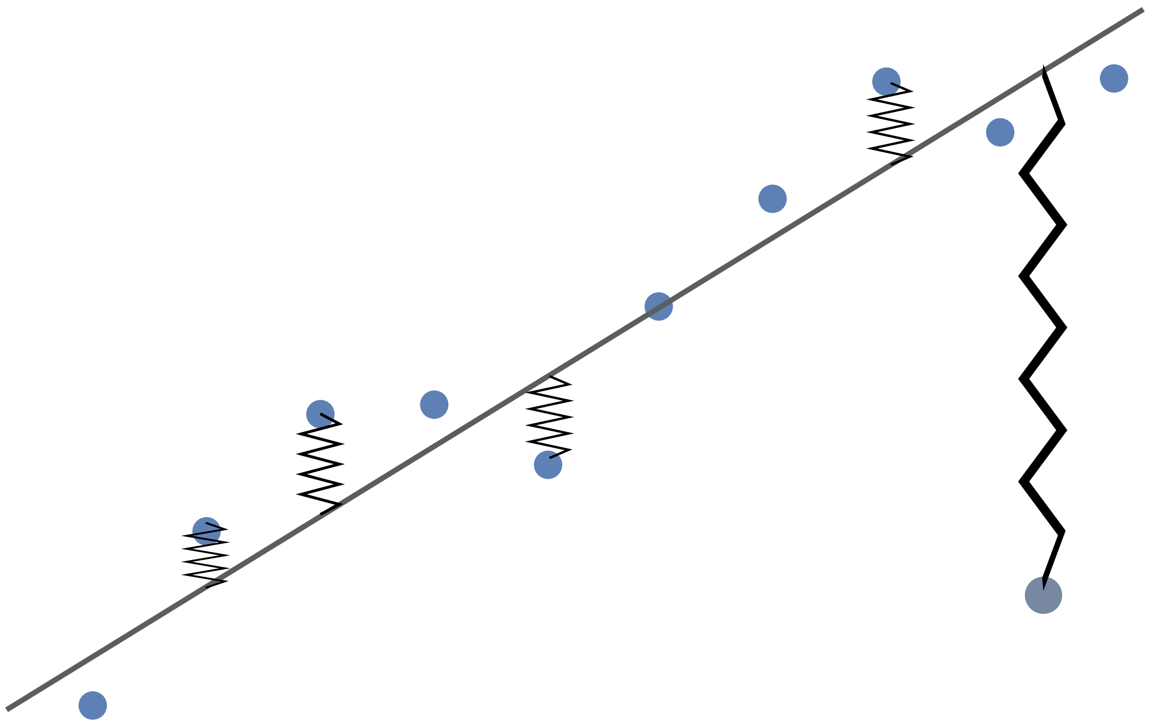


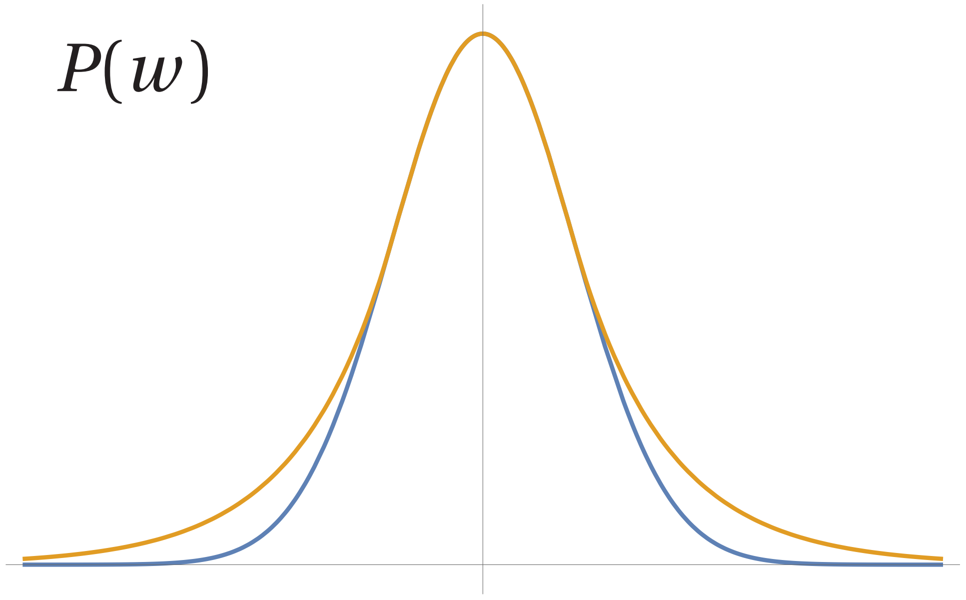
Changing the model

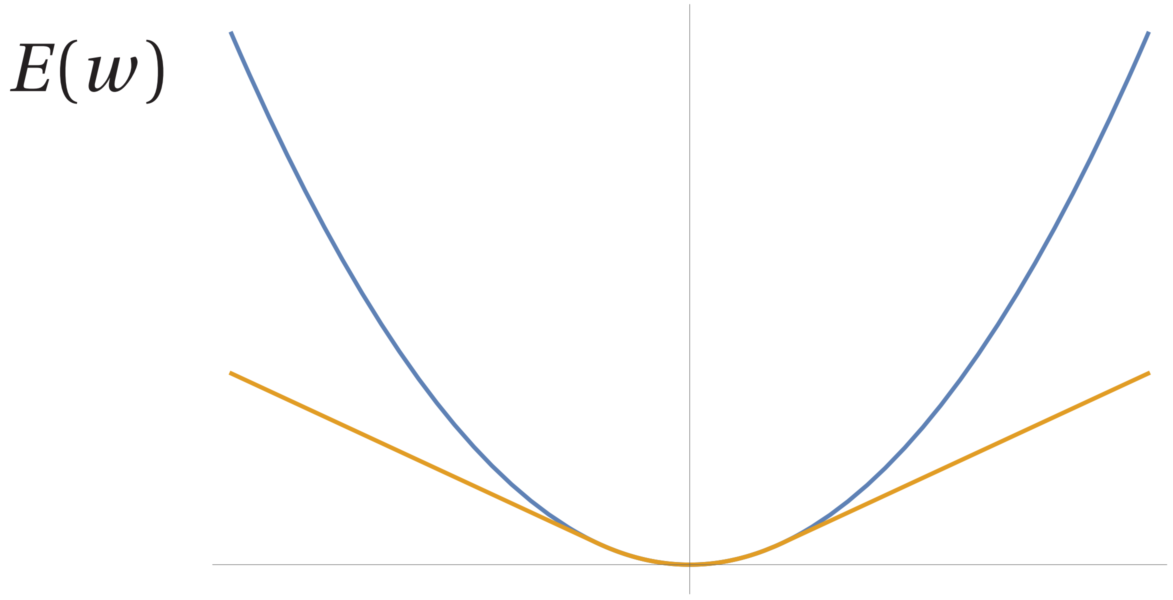
outliers skew our estimates

use a noise model with heavier tails

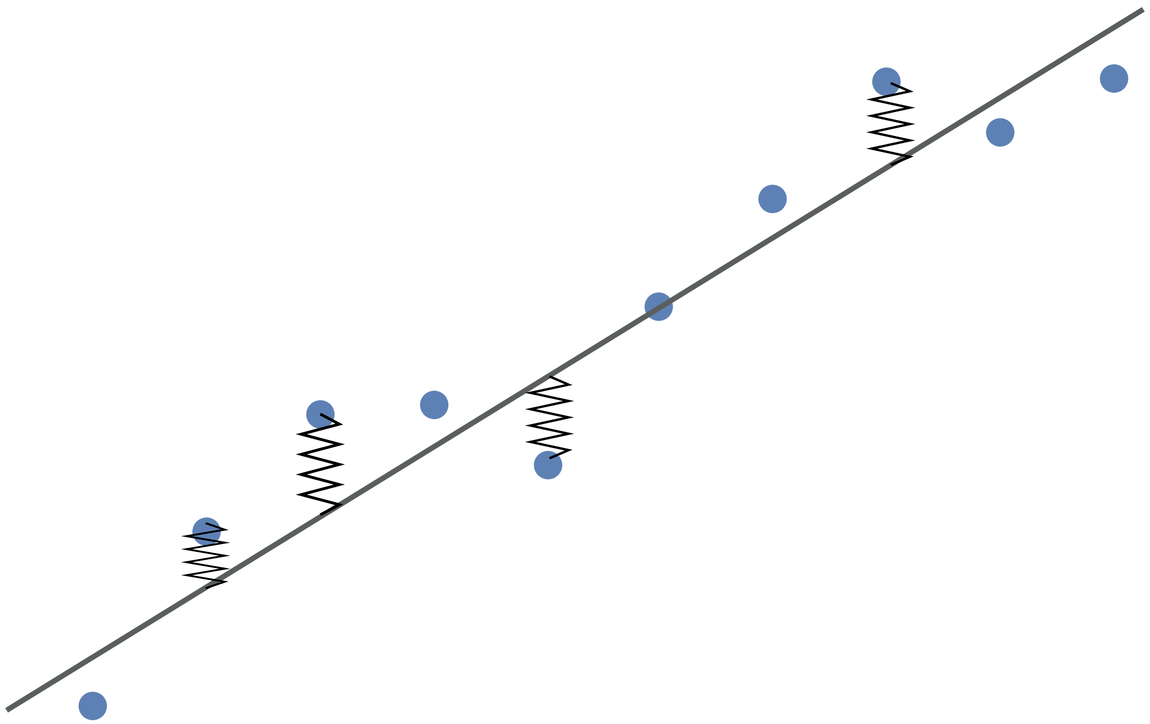
“gooey springs” that exert less force
at large distances





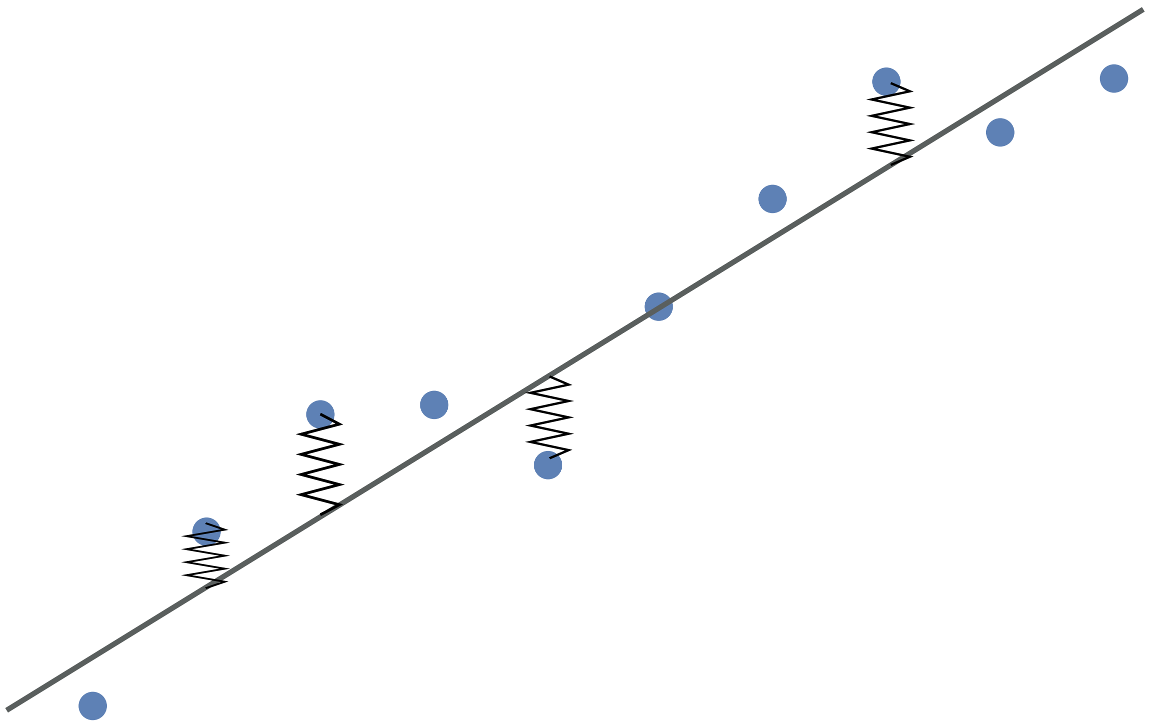


Uncertainty, equilibrium, and the energy landscape



Uncertainty, equilibrium, and the energy landscape

[Bayes] don’t just give an estimate!
what’s the posterior distribution?

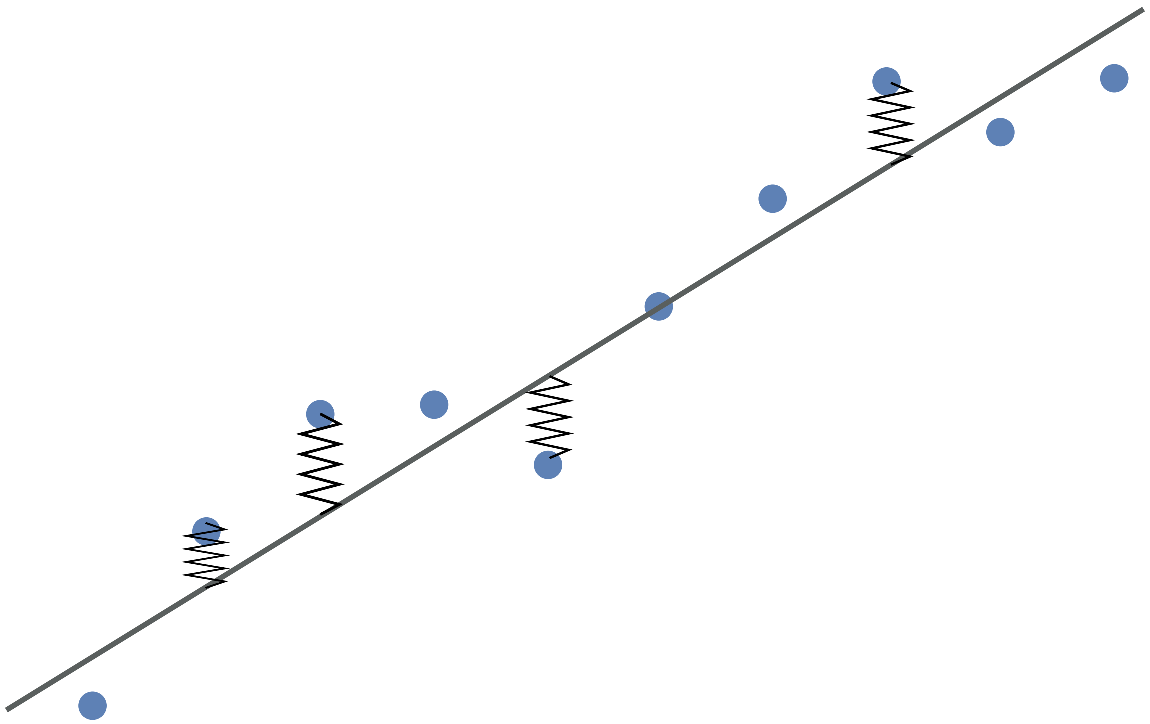


Uncertainty, equilibrium, and the energy landscape

[Bayes] don’t just give an estimate!
what’s the posterior distribution?

[Boltzmann] at thermal equilibrium,

$$P(s) \propto \mathrm{e}^{-E(s)/T}$$



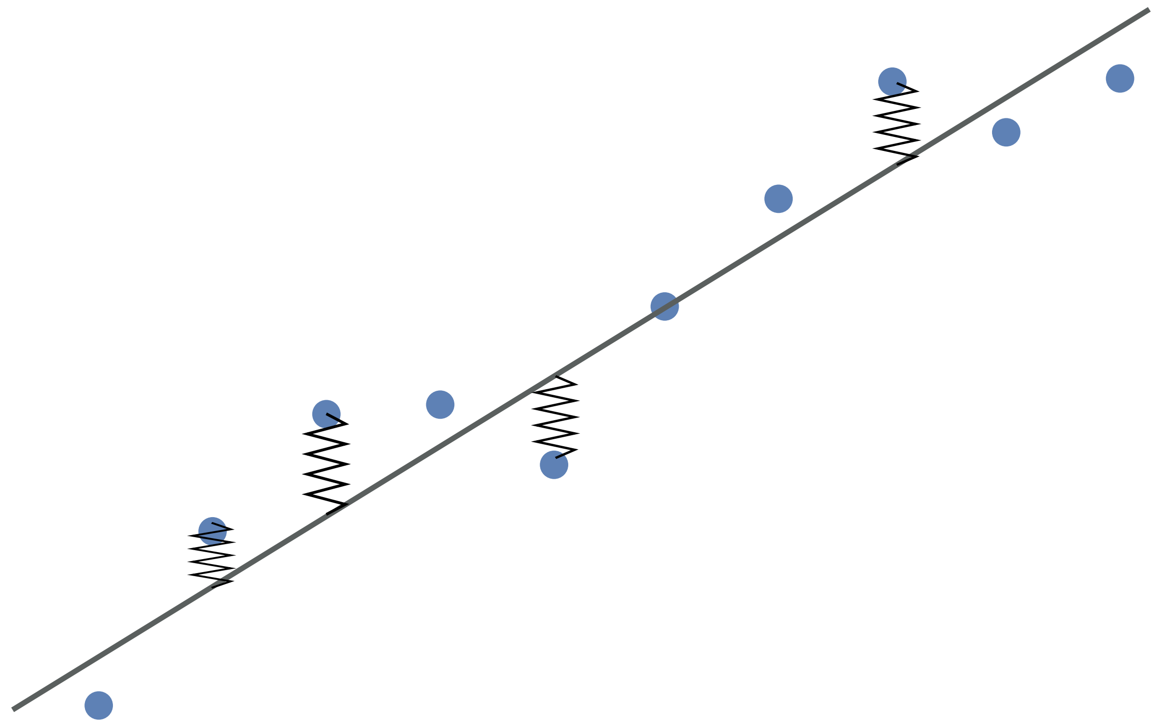
Uncertainty, equilibrium, and the energy landscape

[Bayes] don't just give an estimate!
what's the posterior distribution?

[Boltzmann] at thermal equilibrium,

$$P(s) \propto \mathrm{e}^{-E(s)/T}$$

low *T*: concentrated on ground states



Uncertainty, equilibrium, and the energy landscape

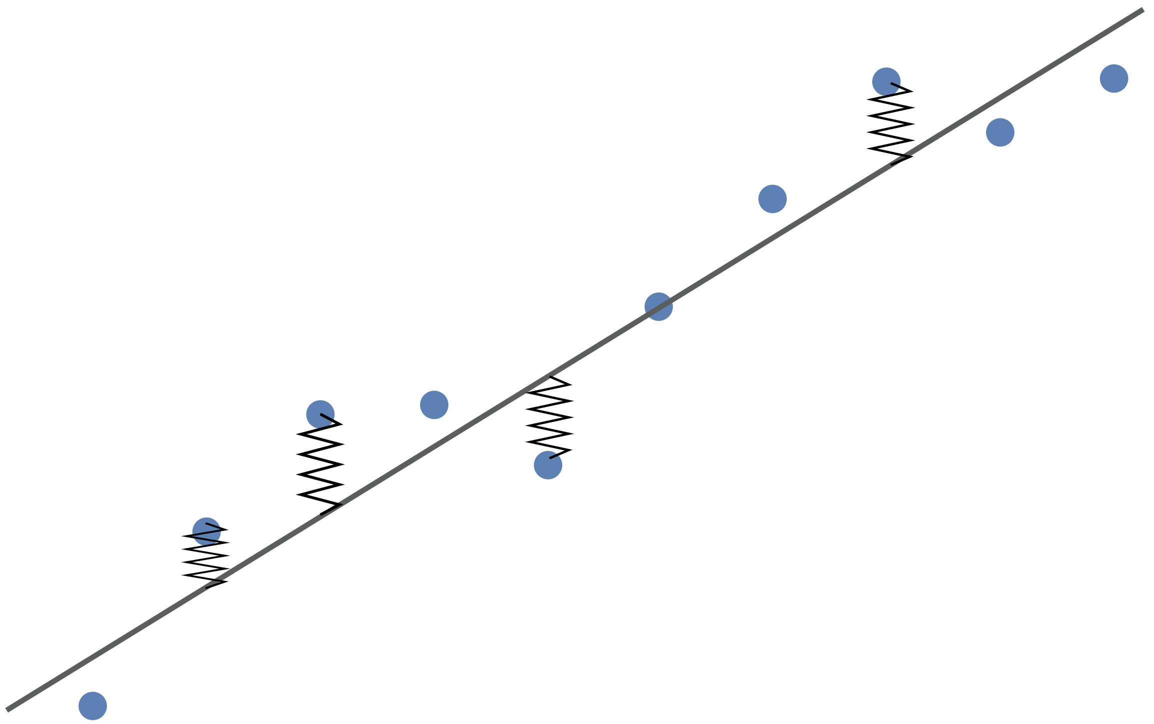
[Bayes] don't just give an estimate!
what's the posterior distribution?

[Boltzmann] at thermal equilibrium,

$$P(s) \propto \mathrm{e}^{-E(s)/T}$$

low *T*: concentrated on ground states

high *T*: uniform



Uncertainty, equilibrium, and the energy landscape

[Bayes] don't just give an estimate!
what's the posterior distribution?

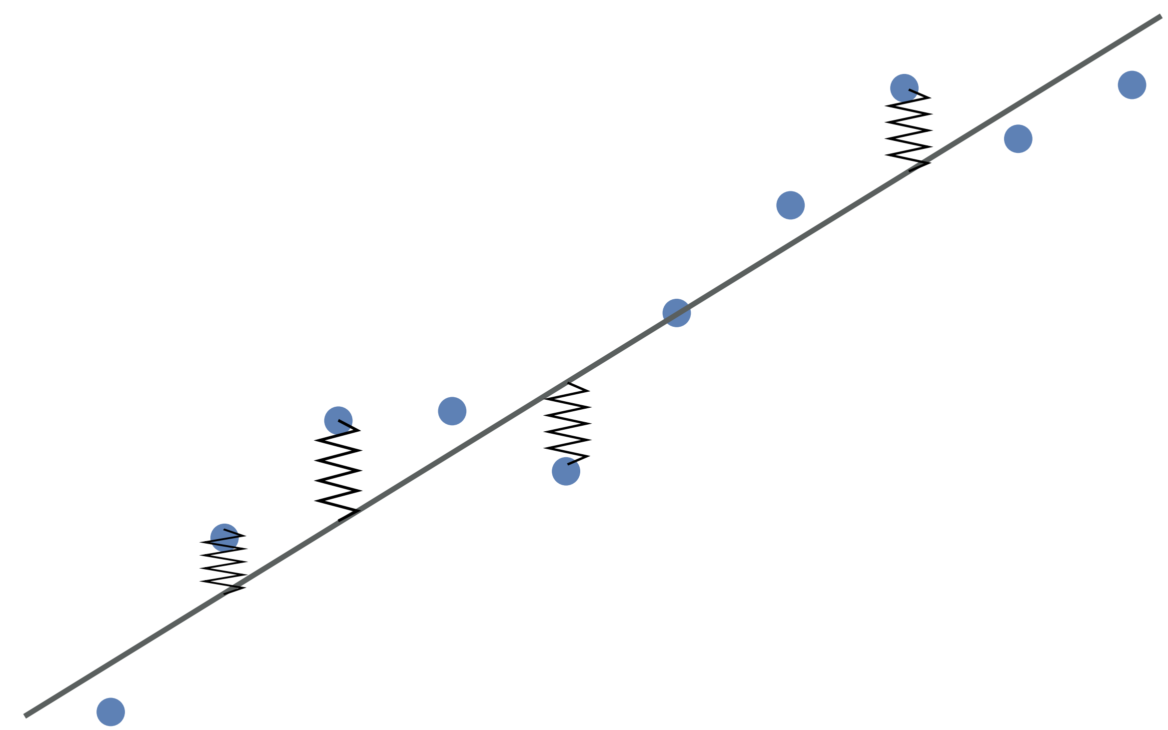
[Boltzmann] at thermal equilibrium,

$$P(s) \propto \mathrm{e}^{-E(s)/T}$$

low T: concentrated on ground states

high T: uniform

thermal noise: T=σ (or looser springs)



Uncertainty, equilibrium, and the energy landscape

[Bayes] don't just give an estimate!
what's the posterior distribution?

[Boltzmann] at thermal equilibrium,

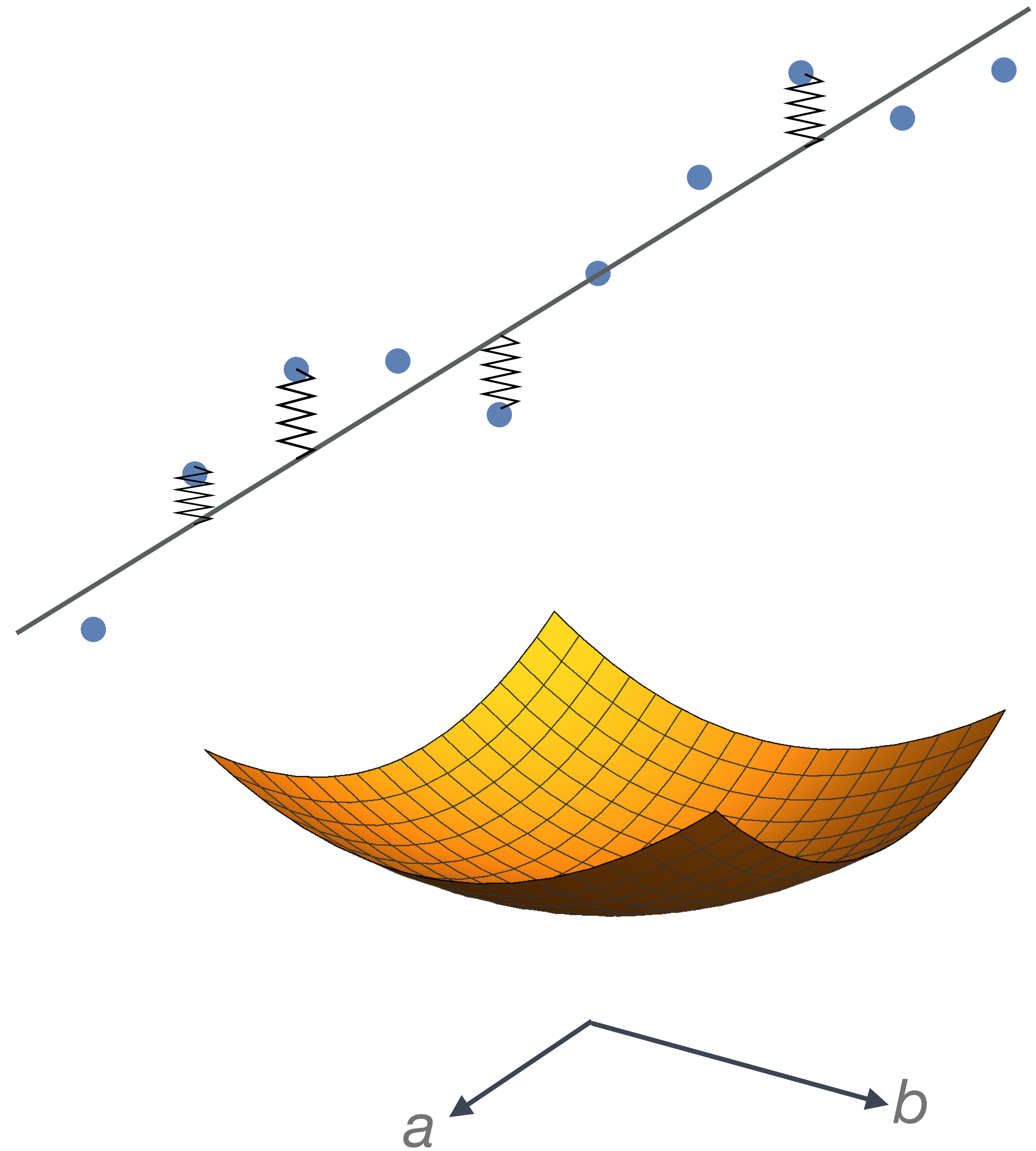
$$P(s) \propto \mathrm{e}^{-E(s)/T}$$

low *T*: concentrated on ground states

high *T*: uniform

thermal noise: *T*=*σ* (or looser springs)

E(*a*,*b*) defined by model and data



Uncertainty, equilibrium, and the energy landscape

[Bayes] don't just give an estimate!
what's the posterior distribution?

[Boltzmann] at thermal equilibrium,

$$P(s) \propto \mathrm{e}^{-E(s)/T}$$

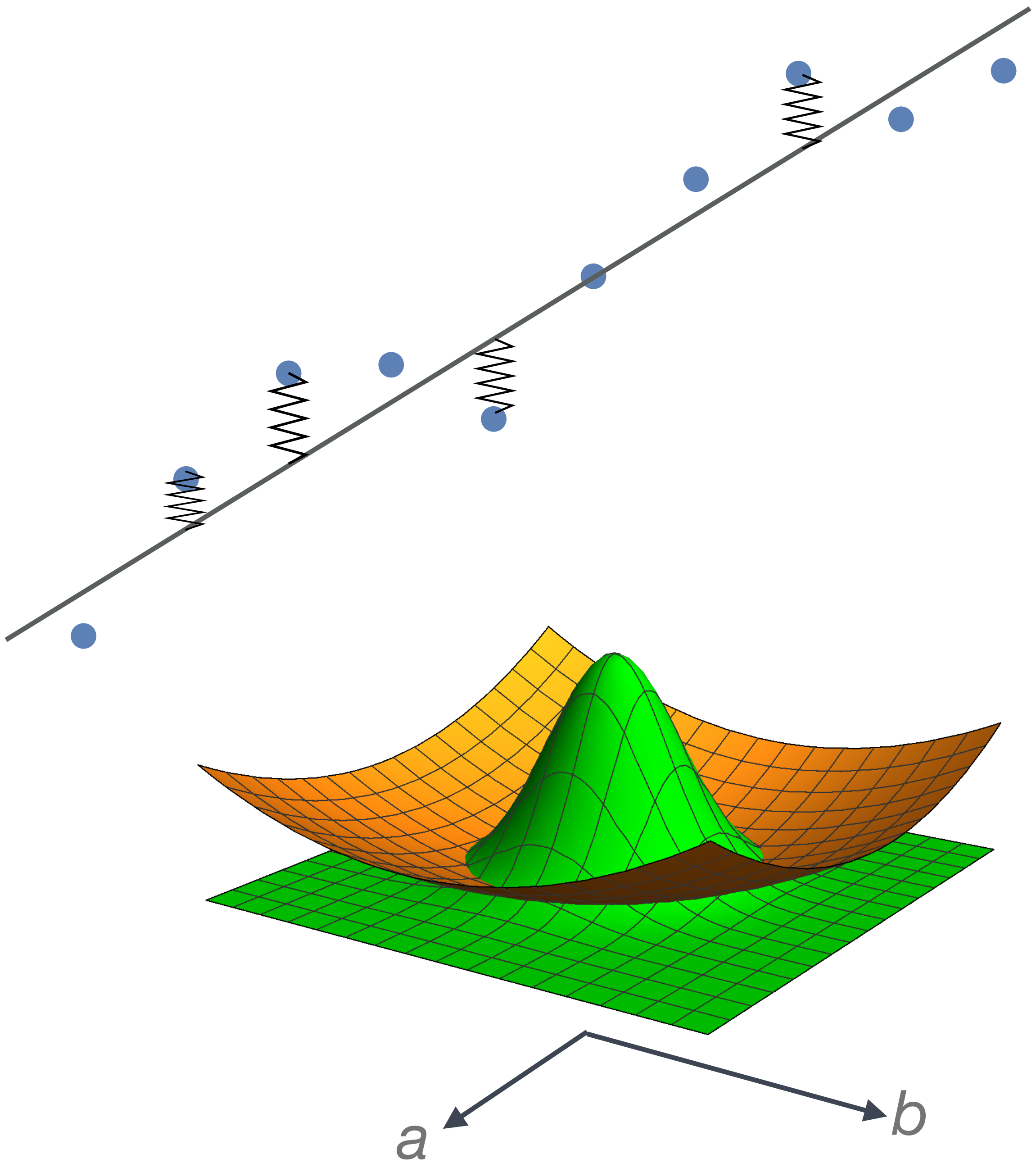
low *T*: concentrated on ground states

high *T*: uniform

thermal noise: $T=\sigma$ (or looser springs)

$E(a,b)$ defined by model and data

posterior distribution = equilibrium



Uncertainty, equilibrium, and the energy landscape

[Bayes] don't just give an estimate!
what's the posterior distribution?

[Boltzmann] at thermal equilibrium,

$$P(s) \propto \mathrm{e}^{-E(s)/T}$$

low *T*: concentrated on ground states

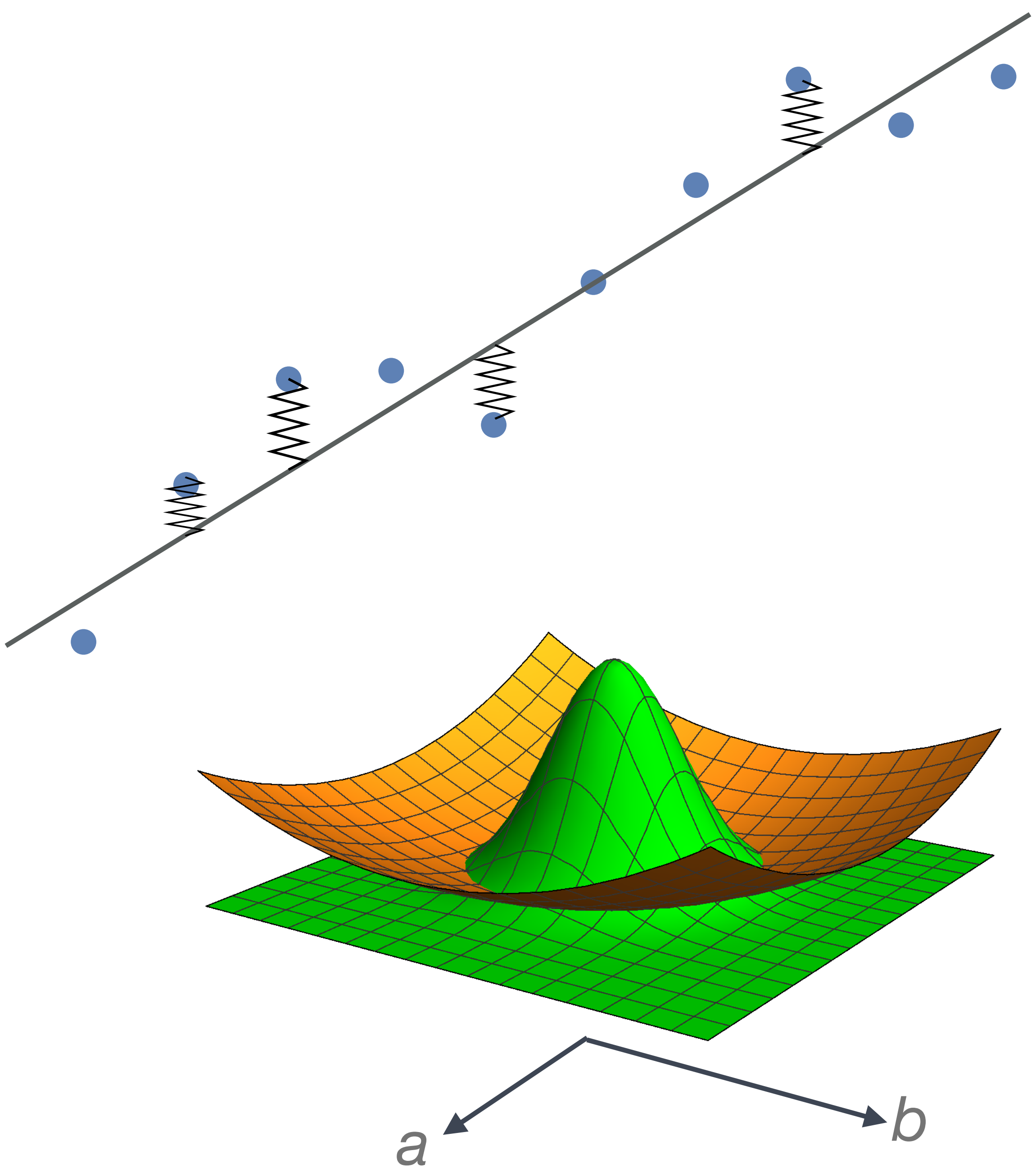
high *T*: uniform

thermal noise: $T=\sigma$ (or looser springs)

$E(a,b)$ defined by model and data

posterior distribution = equilibrium

in this case, landscape is simple and convex



The Ising model of magnetism

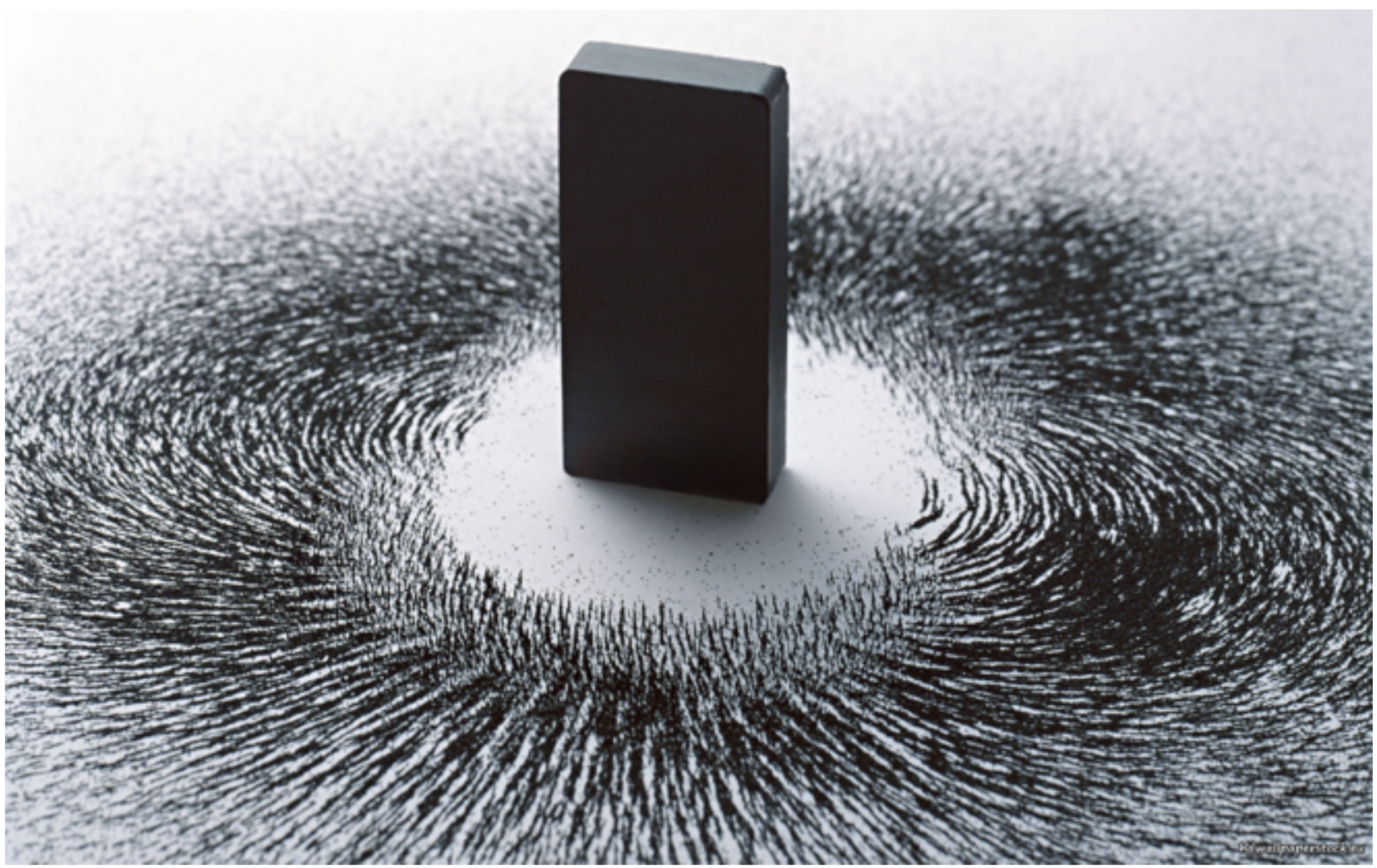
The Ising model of magnetism

the atoms of a block of iron interact with their neighbors $\uparrow\uparrow\uparrow\downarrow\downarrow\uparrow\uparrow\downarrow\uparrow\uparrow\uparrow\uparrow$

The Ising model of magnetism

the atoms of a block of iron interact with their neighbors ↑↑↑↓↓↑↑↓↑↑↑↑

when these interactions are strong enough, or the temperature is low enough, they line up and form a magnetic field



The Ising model of magnetism

the atoms of a block of iron interact with their neighbors $\uparrow\uparrow\uparrow\downarrow\downarrow\uparrow\uparrow\downarrow\uparrow\uparrow\uparrow\uparrow$

when these interactions are strong enough, or the temperature is low enough, they line up and form a magnetic field

each site has a spin $s_i = \pm 1$ and (ferromagnet) $E = -J \sum_{(i,j)} s_i s_j$

The Ising model of magnetism

the atoms of a block of iron interact with their neighbors $\uparrow\uparrow\uparrow\downarrow\downarrow\uparrow\uparrow\downarrow\uparrow\uparrow\uparrow\uparrow$

when these interactions are strong enough, or the temperature is low enough, they line up and form a magnetic field

each site has a spin $s_i = \pm 1$ and (ferromagnet) $E = -J \sum_{(i,j)} s_i s_j$

ground state: all up or all down

The Ising model of magnetism

the atoms of a block of iron interact with their neighbors $\uparrow\uparrow\uparrow\downarrow\downarrow\uparrow\uparrow\downarrow\uparrow\uparrow\uparrow\uparrow$

when these interactions are strong enough, or the temperature is low enough, they line up and form a magnetic field

each site has a spin $s_i = \pm 1$ and (ferromagnet) $E = -J \sum_{(i,j)} s_i s_j$

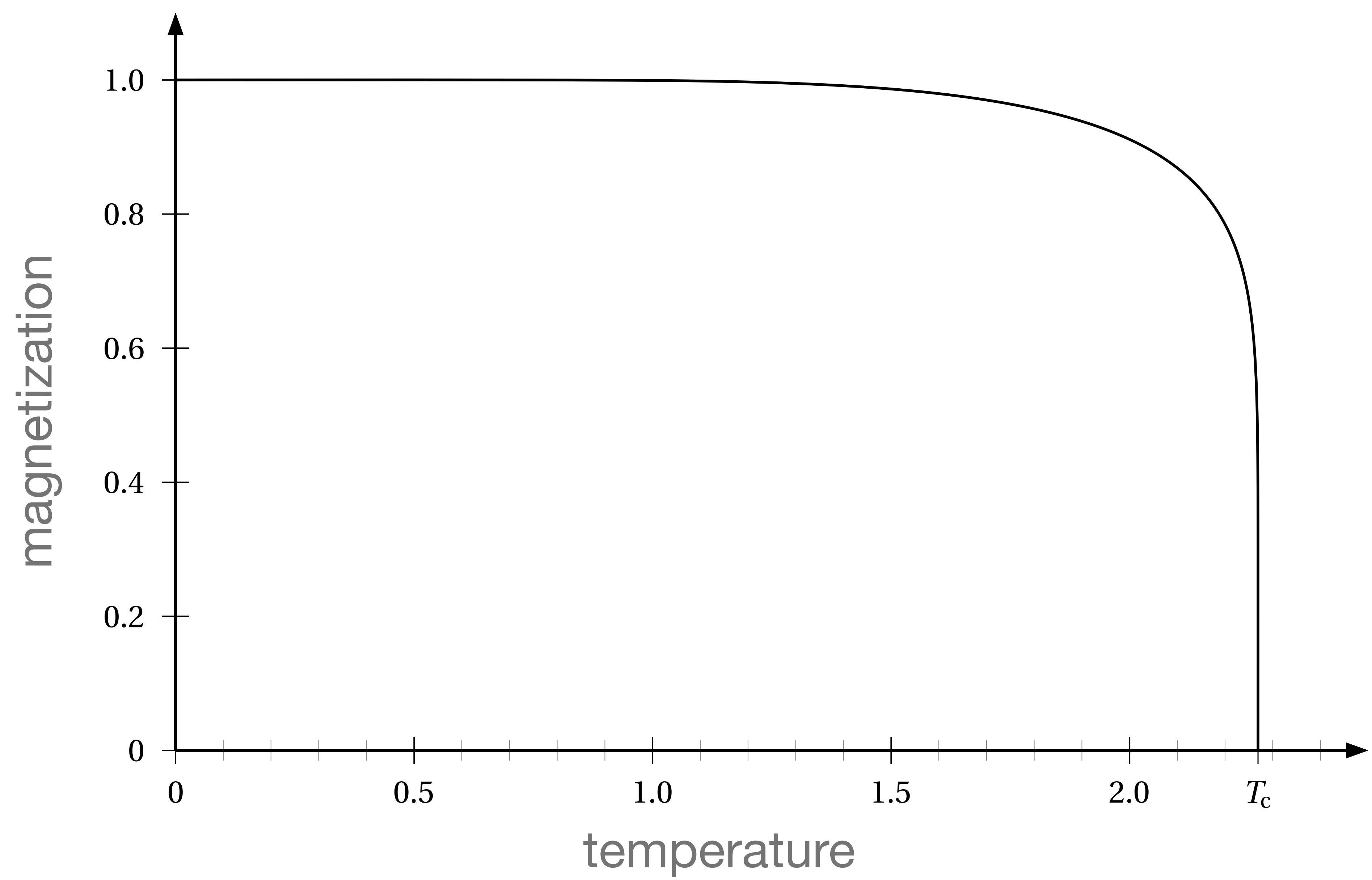
ground state: all up or all down

how does the magnetization $\left| \frac{1}{n} \sum_i s_i \right|$ vary with temperature?

The Ising model of magnetism

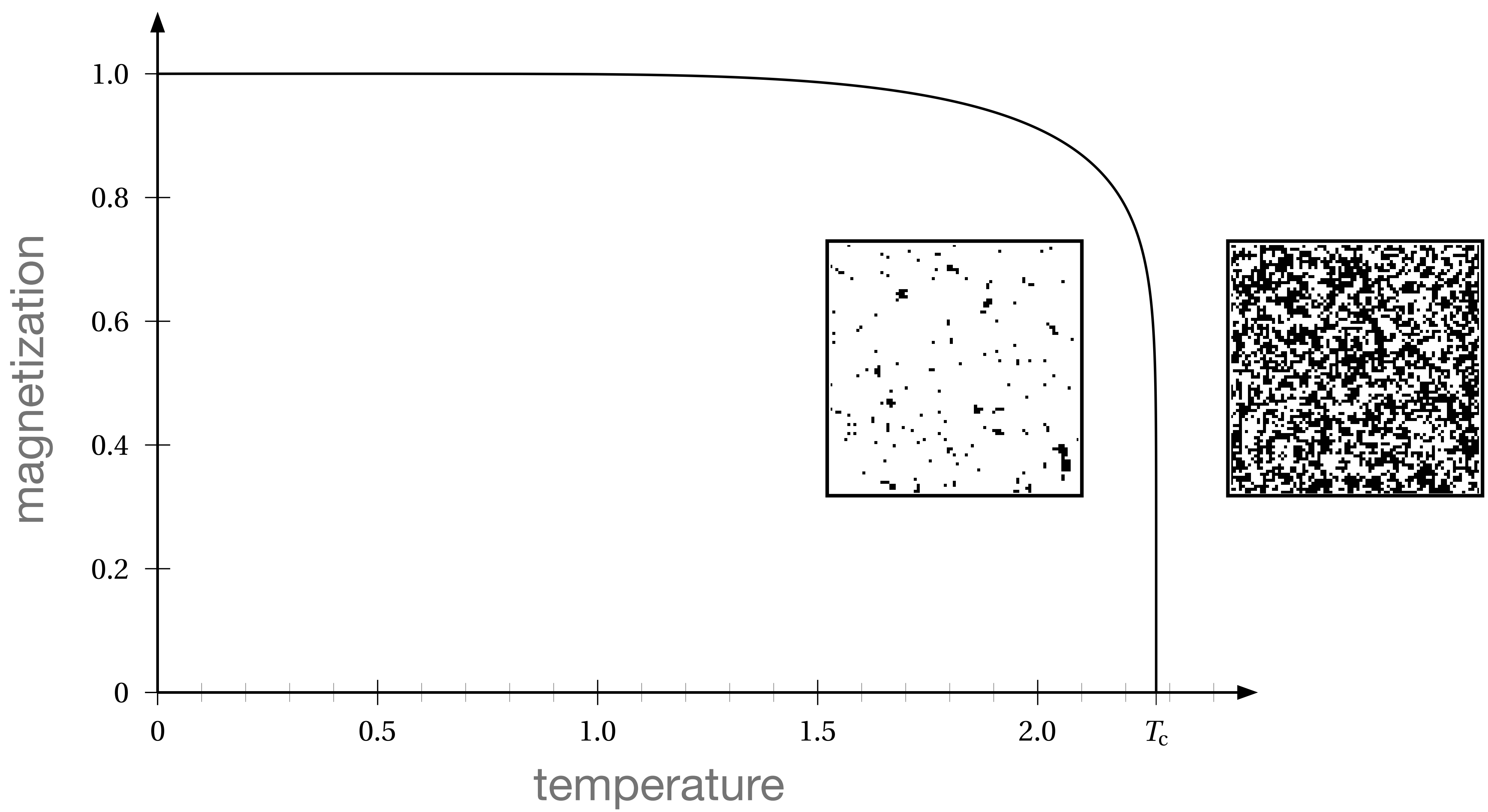
The Ising model of magnetism

at a critical temperature, the iron suddenly loses its magnetic field



The Ising model of magnetism

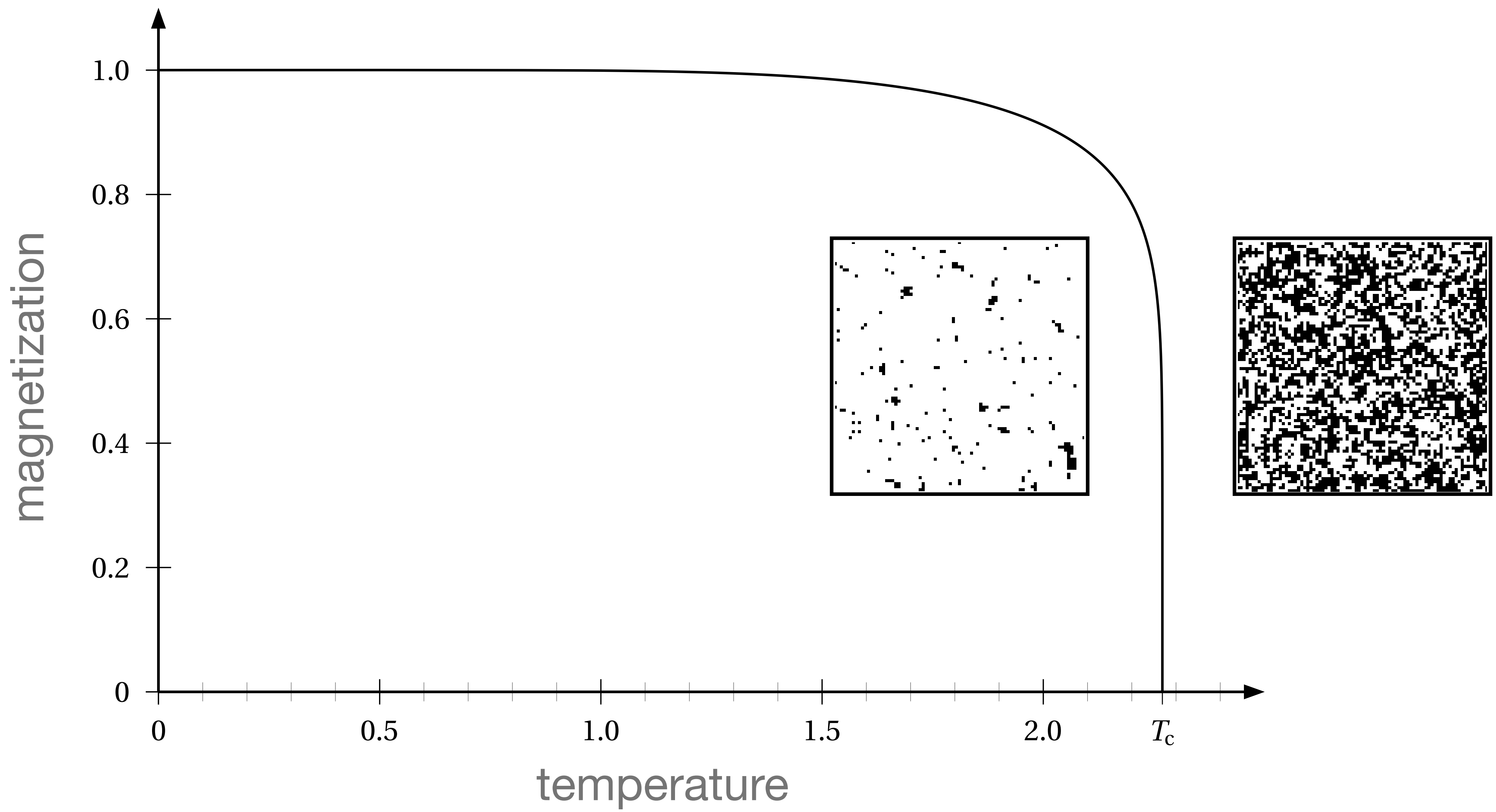
at a critical temperature, the iron suddenly loses its magnetic field



The Ising model of magnetism

at a critical temperature, the iron suddenly loses its magnetic field

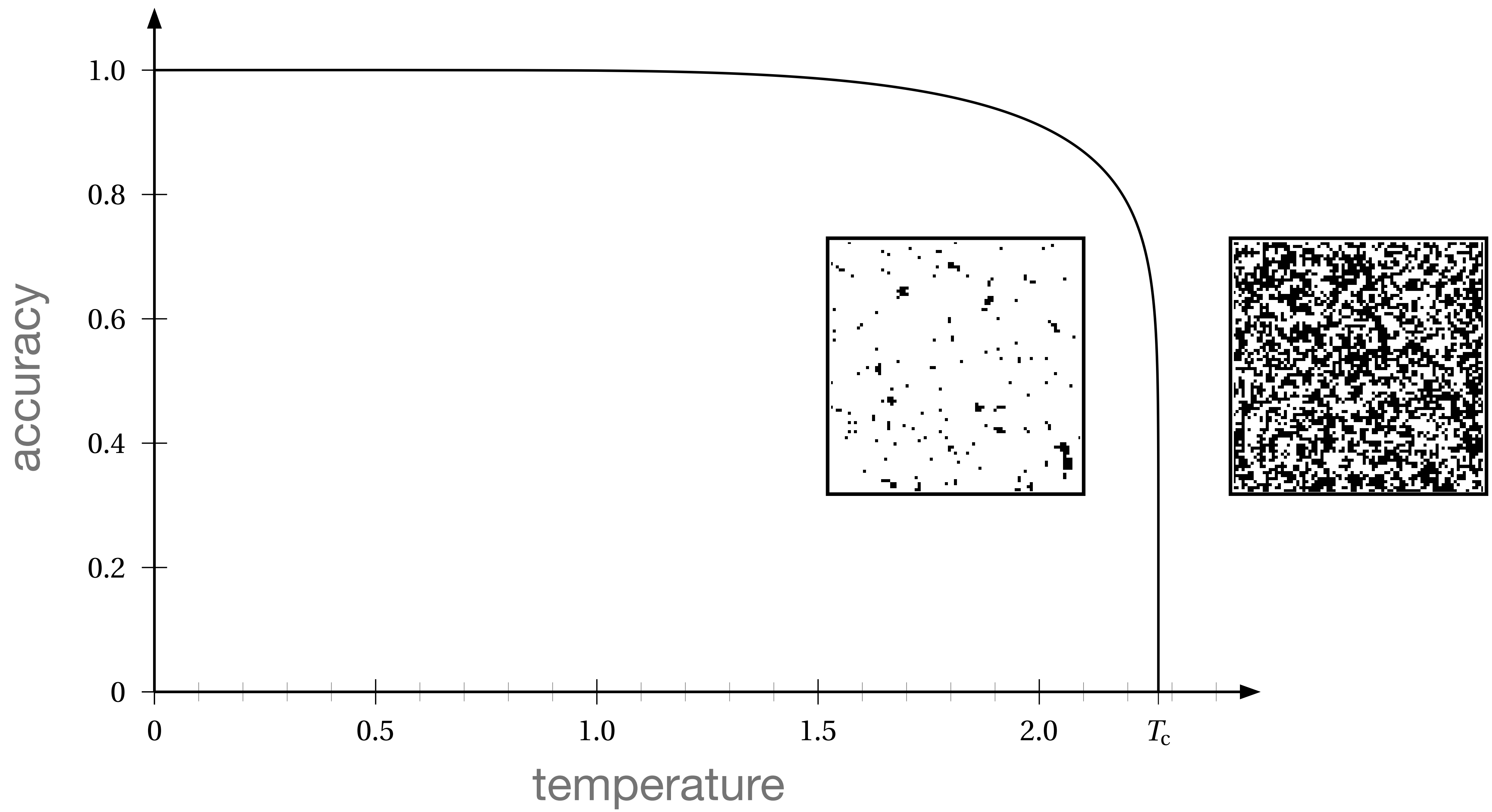
atoms become uncorrelated: $\langle s_i s_j \rangle \sim \mathrm{e}^{-r/\ell}$, no long-range information



The Ising model of magnetism

at a critical temperature, the iron suddenly loses its magnetic field

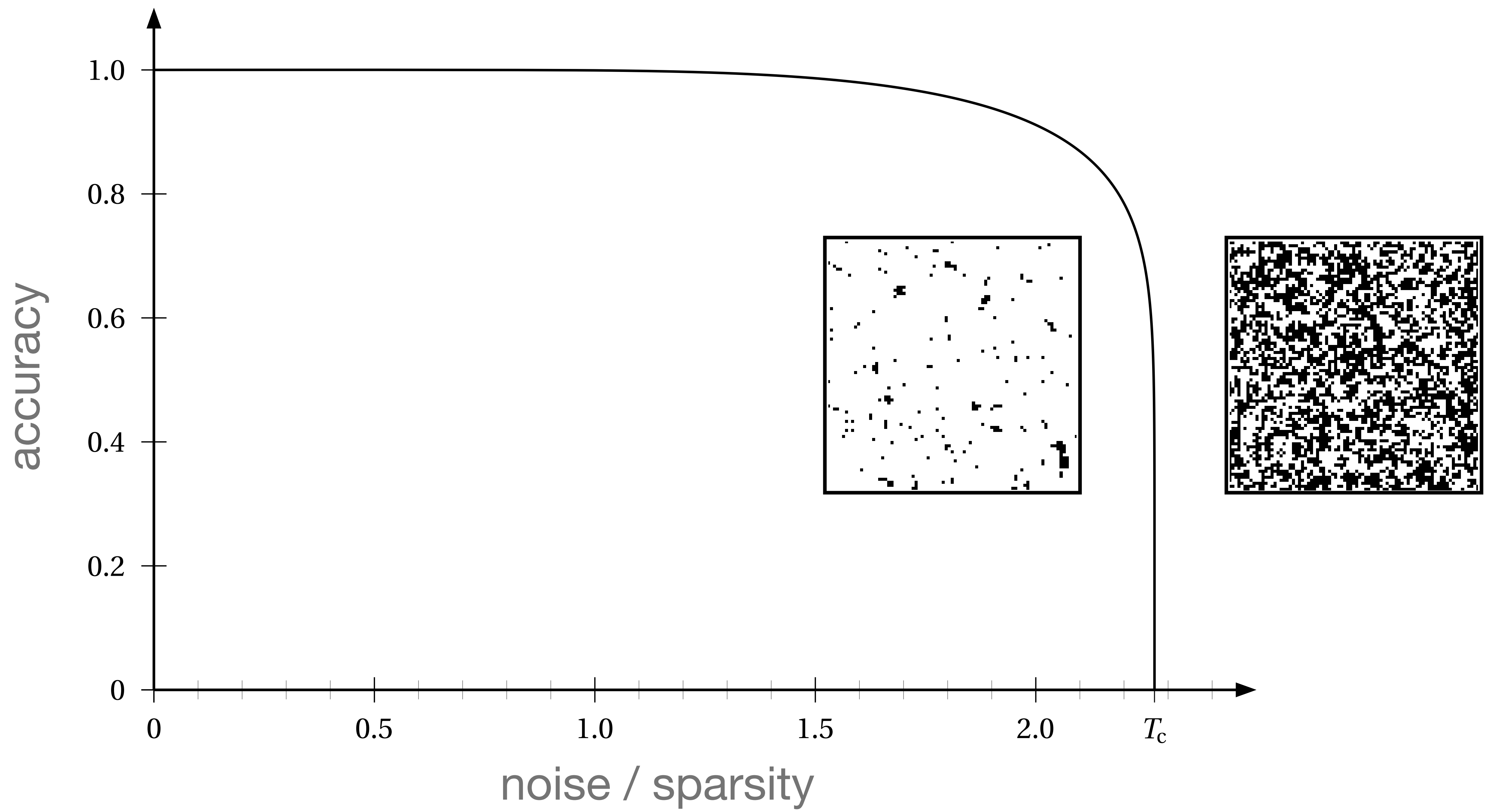
atoms become uncorrelated: $\langle s_i s_j \rangle \sim \mathrm{e}^{-r/\ell}$, no long-range information



The Ising model of magnetism

at a critical temperature, the iron suddenly loses its magnetic field

atoms become uncorrelated: $\langle s_i s_j \rangle \sim \mathrm{e}^{-r/\ell}$, no long-range information



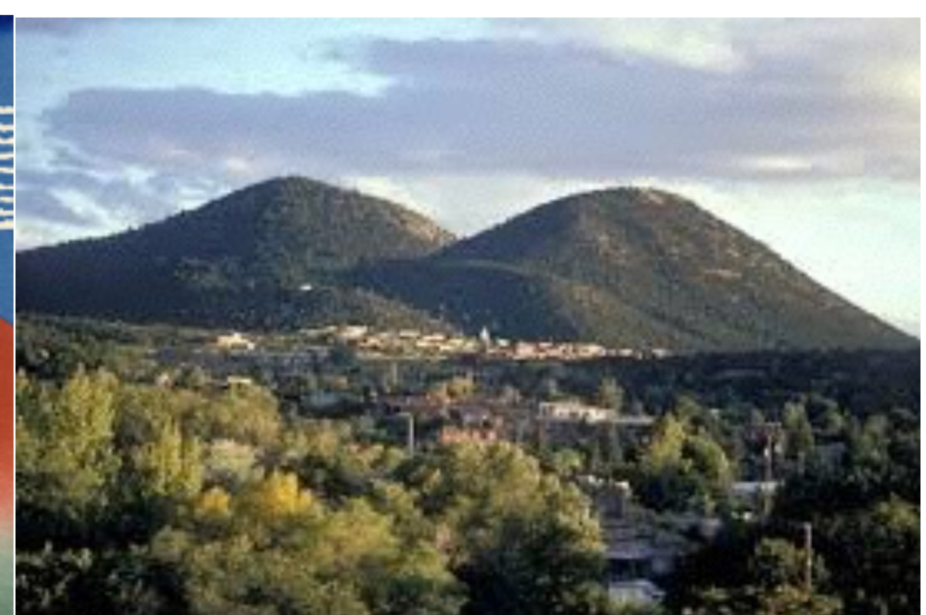
Bumpy landscapes

Bumpy landscapes

least squares has a landscape with one optimum, and the Ising model has two

Bumpy landscapes

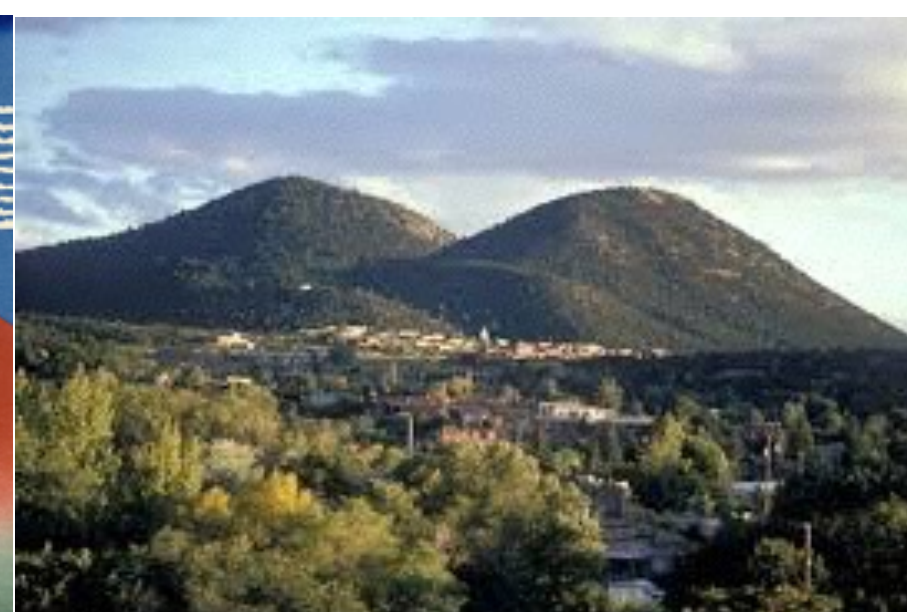
least squares has a landscape with one optimum, and the Ising model has two



Bumpy landscapes

least squares has a landscape with one optimum, and the Ising model has two

but a “spin glass” with energy $E = -\sum_{(i,j)} J_{ij} s_i s_j$ can have exponentially many

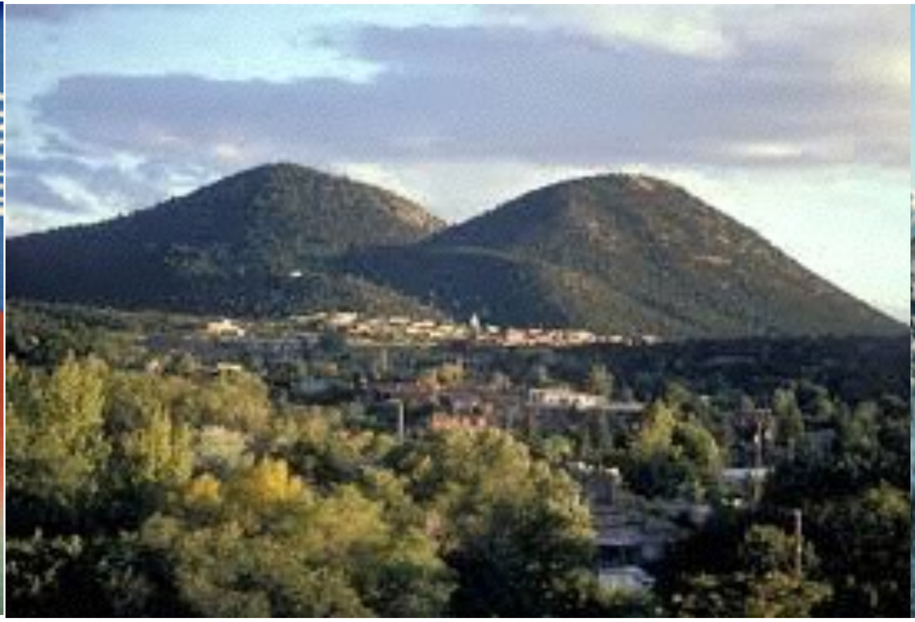




Bumpy landscapes

least squares has a landscape with one optimum, and the Ising model has two

but a “spin glass” with energy $E = -\sum_{(i,j)} J_{ij} s_i s_j$ can have exponentially many

suppose the interactions J_{ij} depend on the data and the model

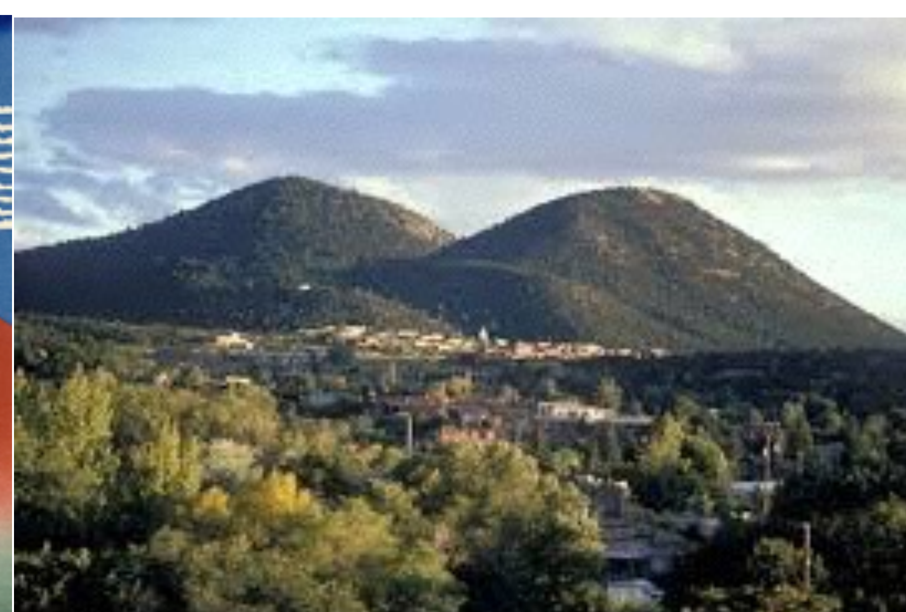

Bumpy landscapes

least squares has a landscape with one optimum, and the Ising model has two

but a “spin glass” with energy $E = -\sum_{(i,j)} J_{ij} s_i s_j$ can have exponentially many

suppose the interactions J_{ij} depend on the data and the model

which local optimum is the true one?


Bumpy landscapes

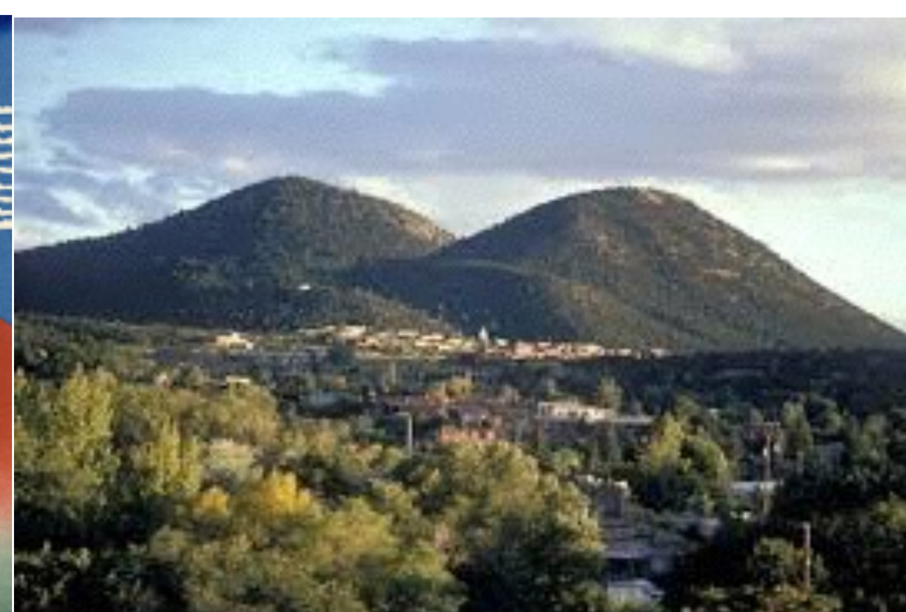
least squares has a landscape with one optimum, and the Ising model has two

but a “spin glass” with energy $E = -\sum_{(i,j)} J_{ij} s_i s_j$ can have exponentially many

suppose the interactions J_{ij} depend on the data and the model

which local optimum is the true one?

can we find it efficiently? can we find it at all, given the posterior distribution?


Bumpy landscapes

least squares has a landscape with one optimum, and the Ising model has two

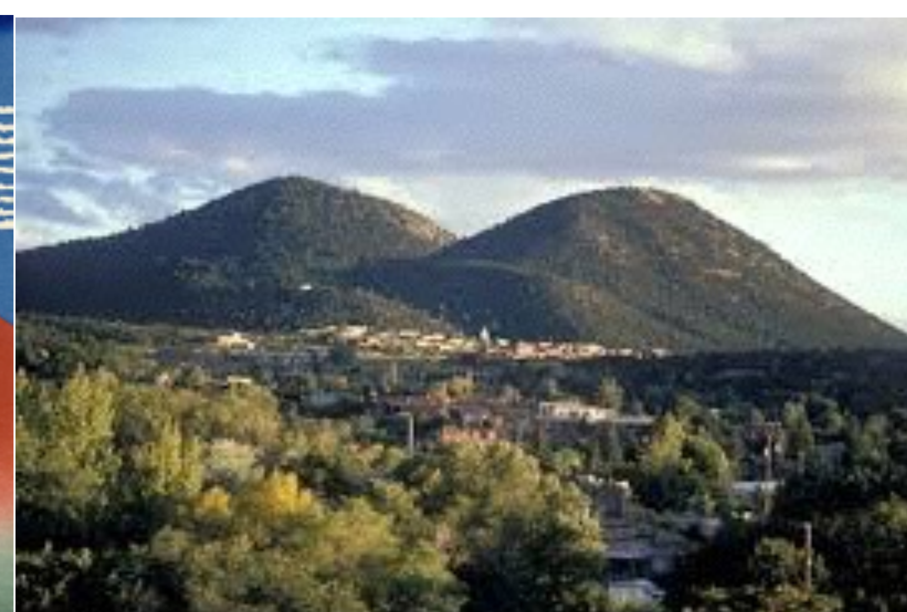
but a “spin glass” with energy $E = -\sum_{(i,j)} J_{ij} s_i s_j$ can have exponentially many

suppose the interactions J_{ij} depend on the data and the model

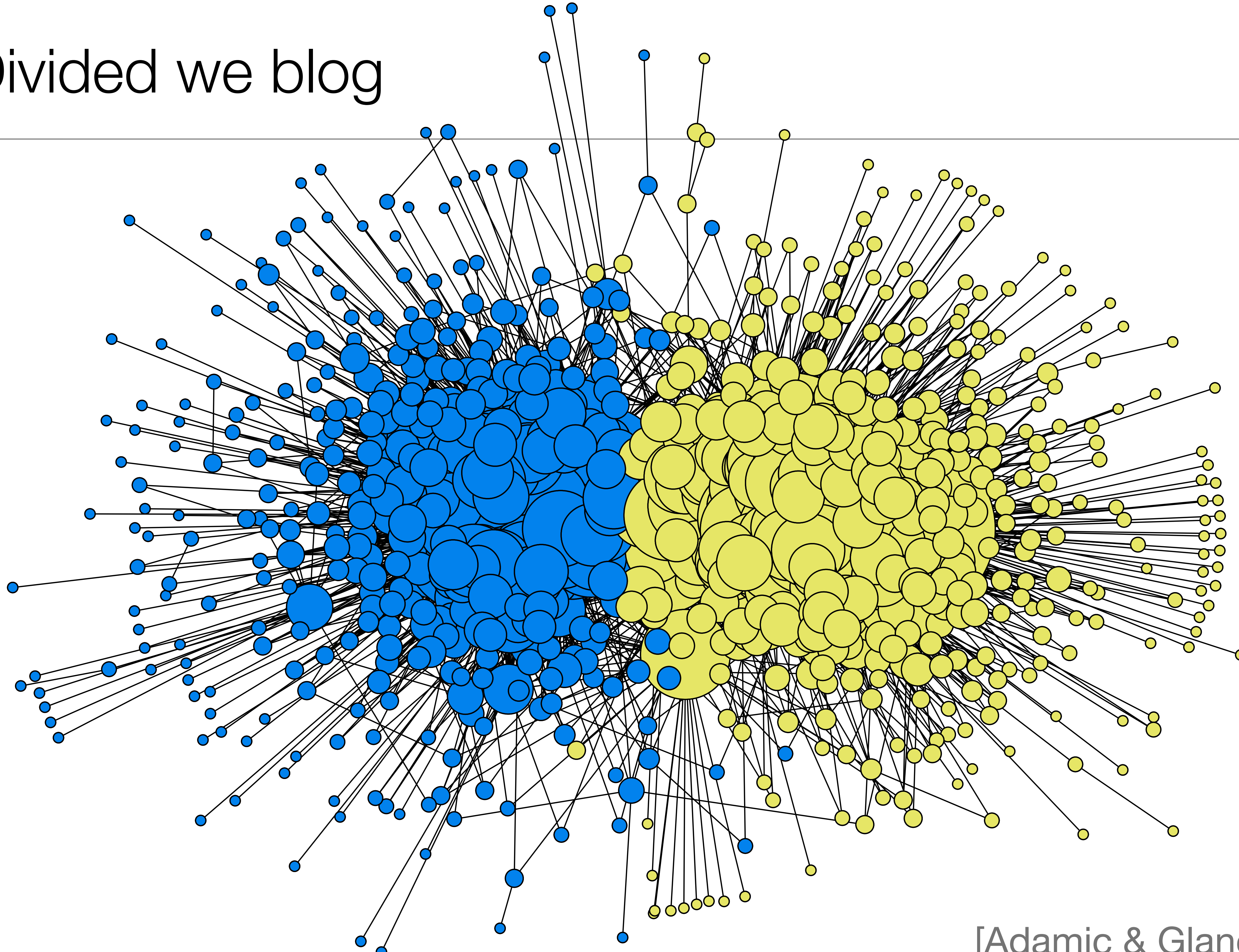
which local optimum is the true one?

can we find it efficiently? can we find it at all, given the posterior distribution?

let’s look at a classic problem in social networks…

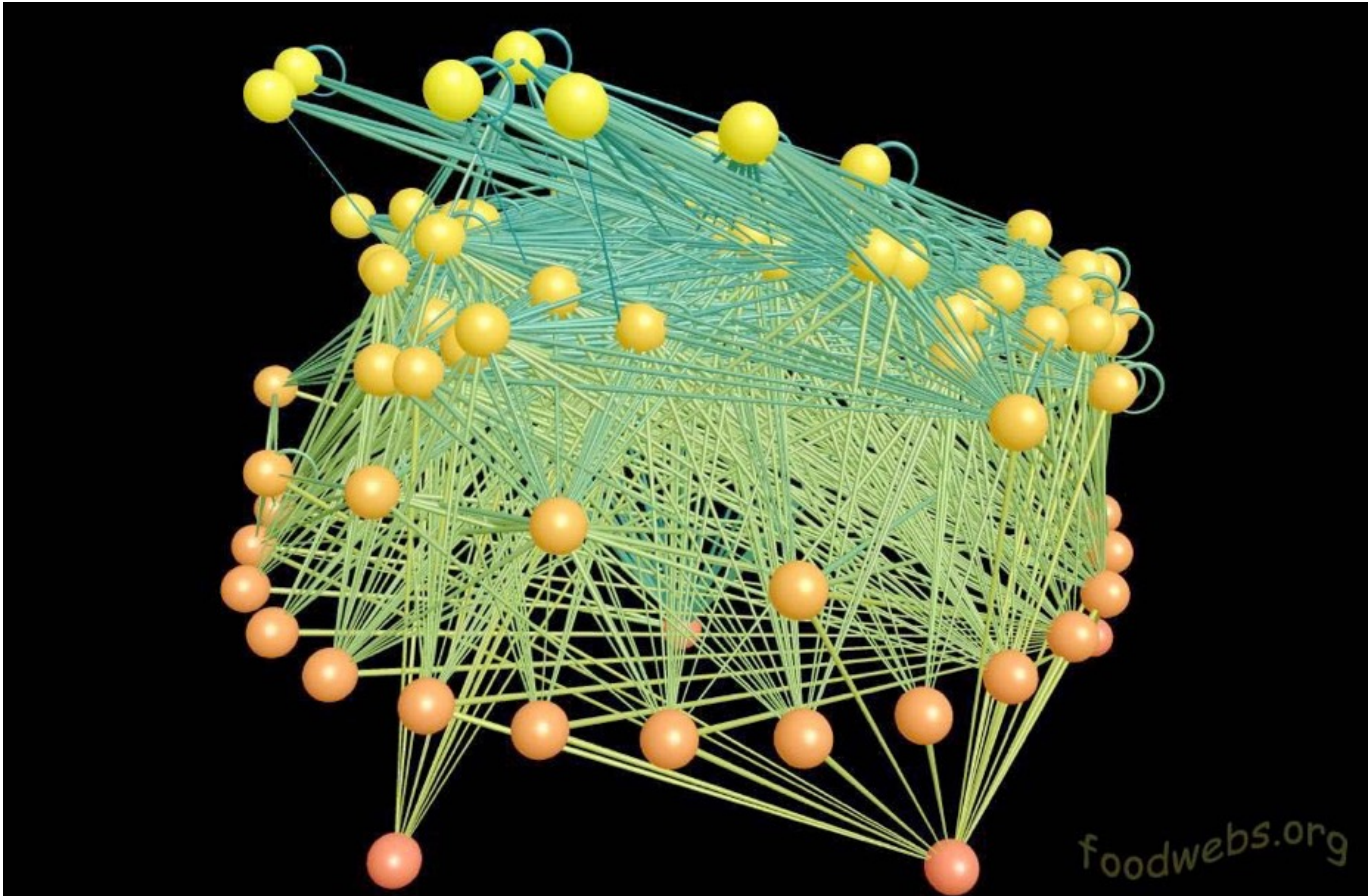




Divided we blog

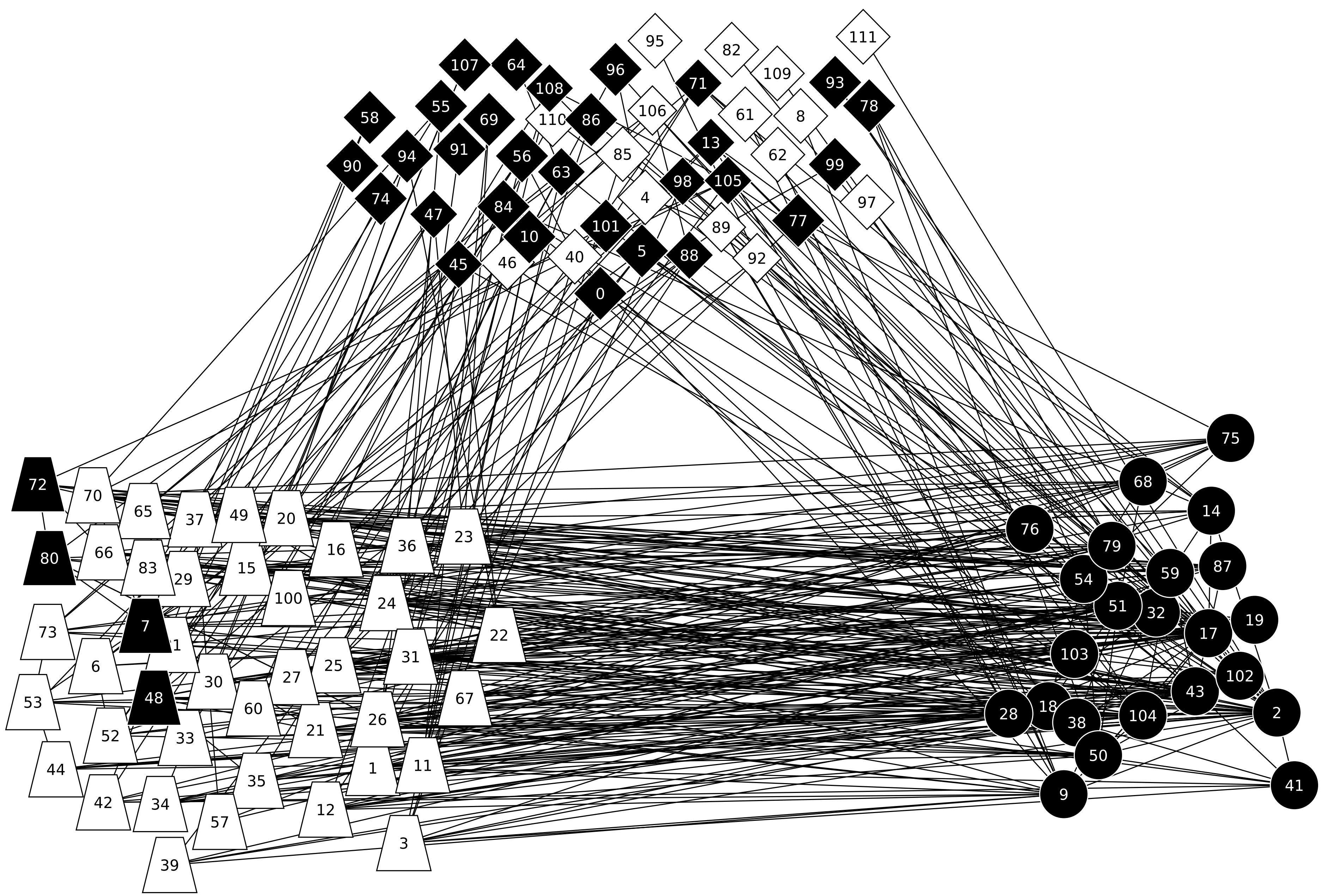


[Adamic & Glance]

Who eats whom



I record that I was born on a Friday



The stochastic block model

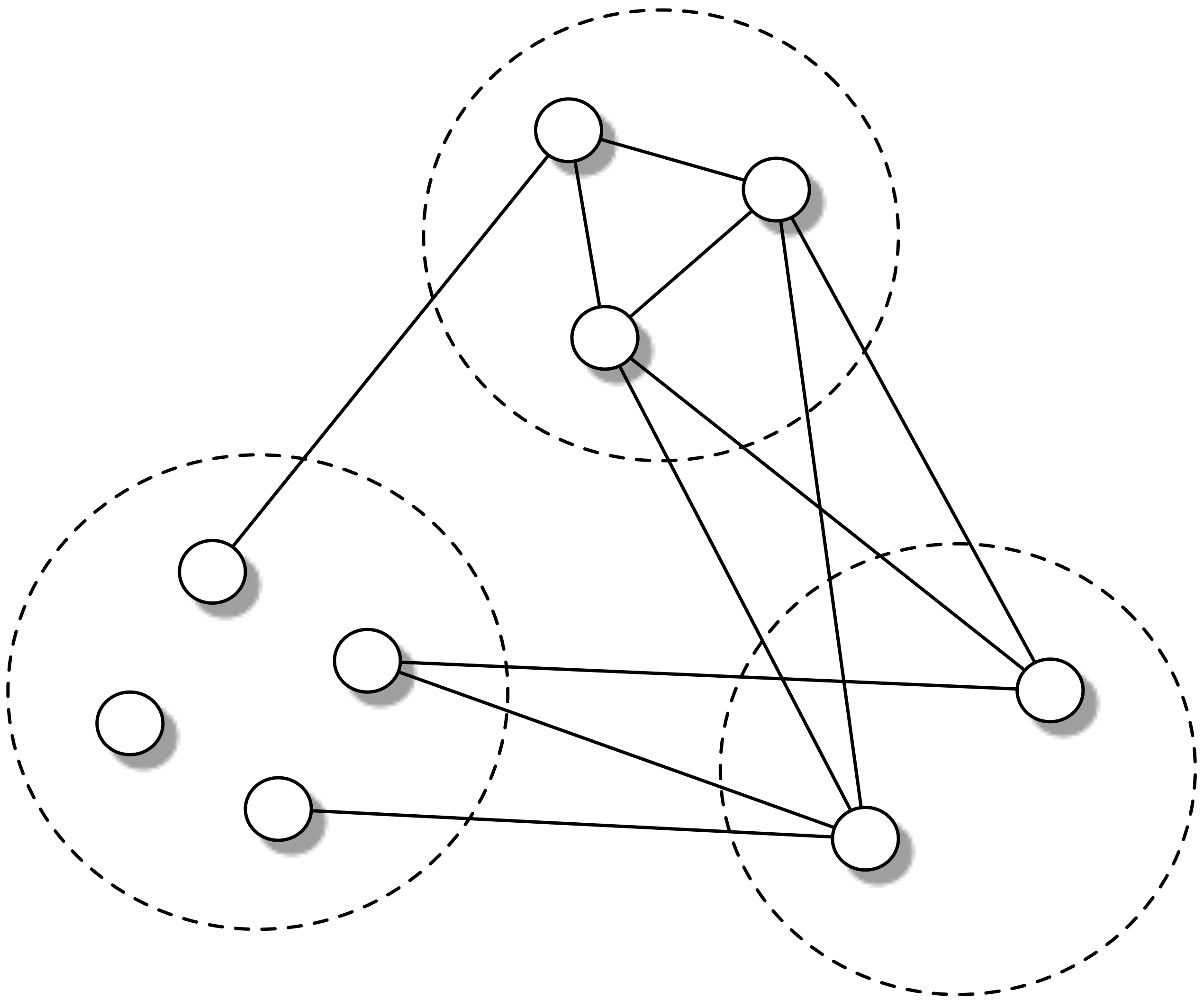
The stochastic block model

nodes have discrete labels: k “groups” or types of nodes

The stochastic block model

nodes have discrete labels: *k* “groups” or types of nodes

k×*k* matrix *p* of connection probabilities

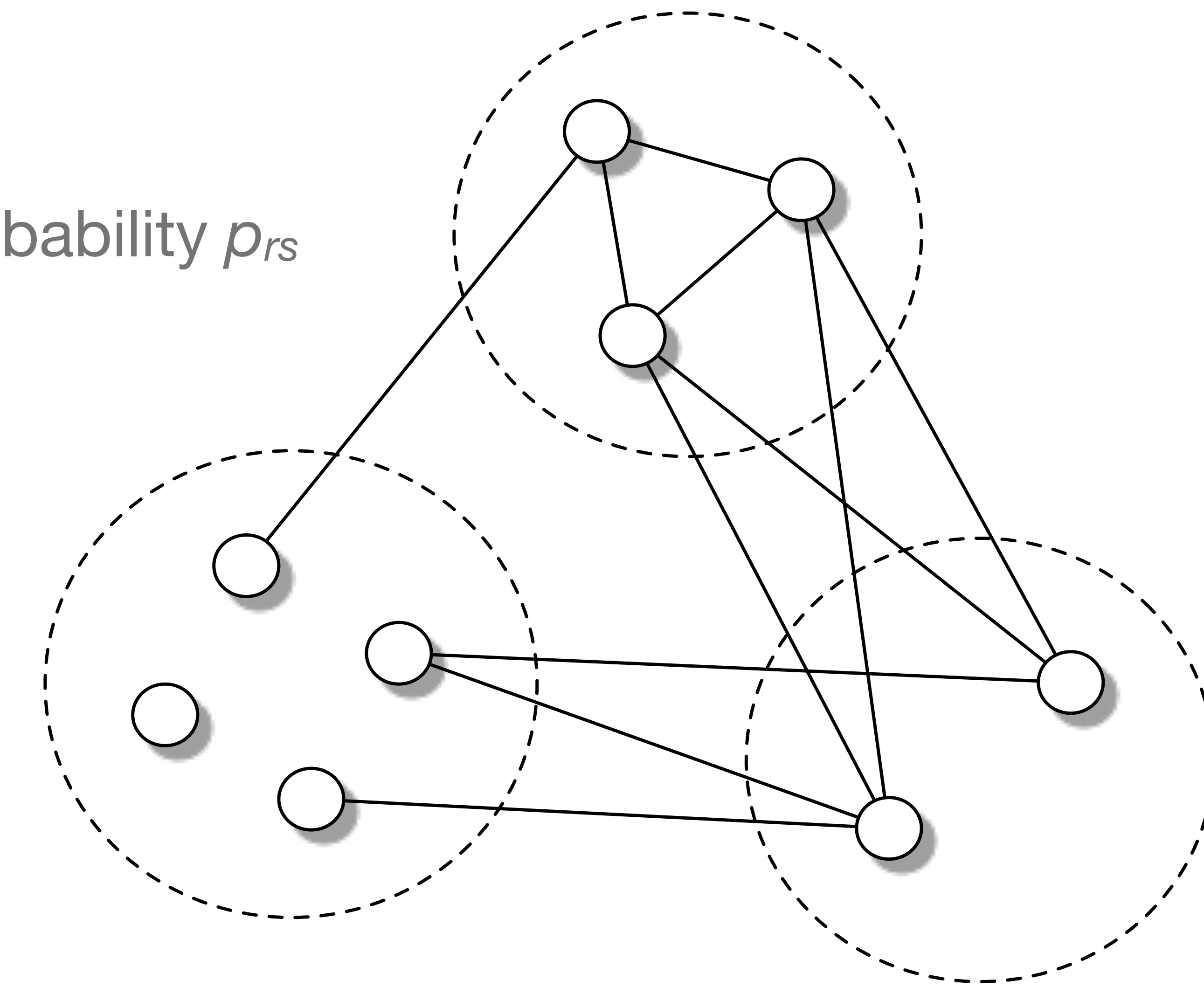


The stochastic block model

nodes have discrete labels: k "groups" or types of nodes

$k \times k$ matrix p of connection probabilities

if $t_i = r$ and $t_j = s$, there is a link $i \to j$ with probability p_{rs}



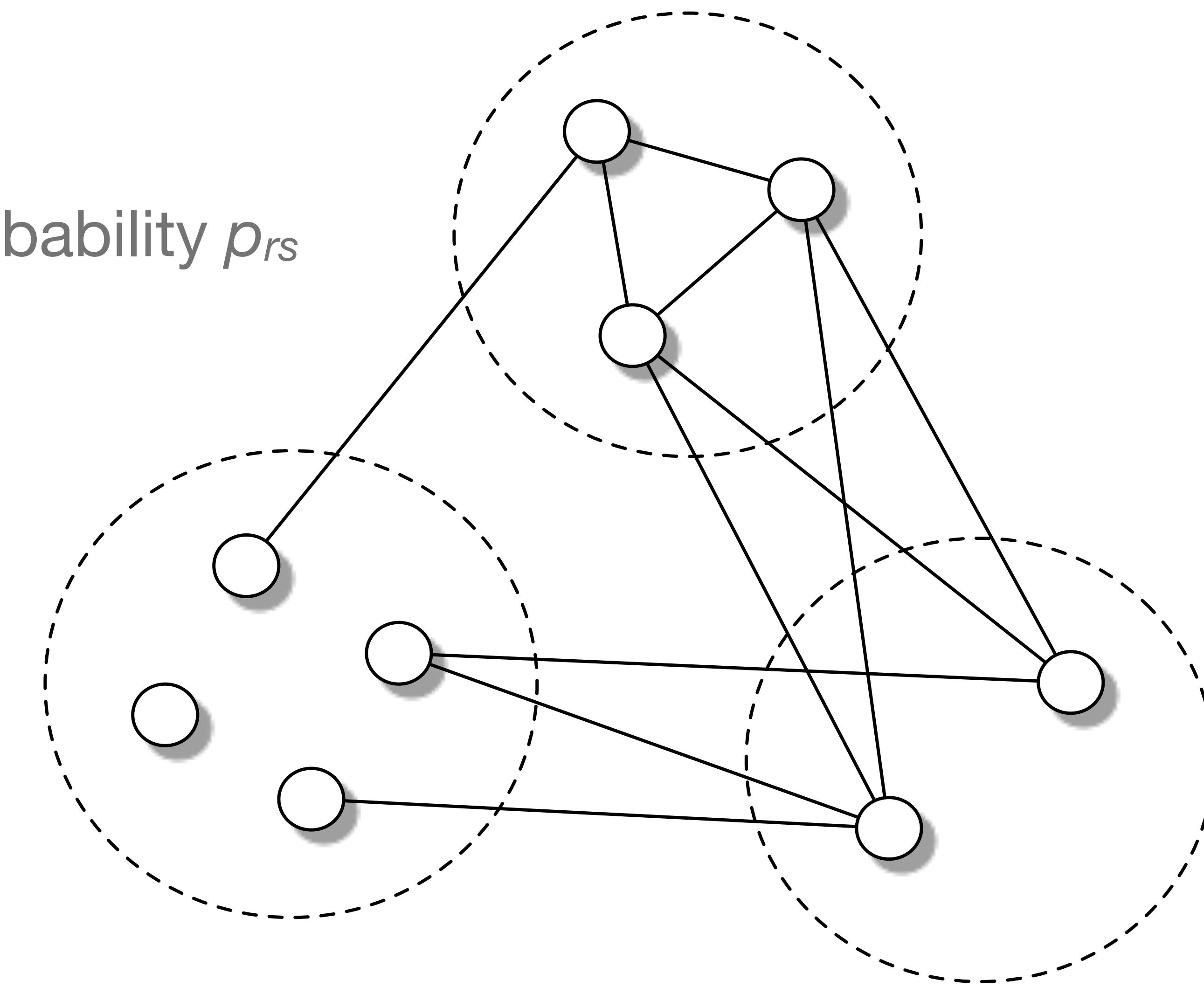
The stochastic block model

nodes have discrete labels: k "groups" or types of nodes

$k \times k$ matrix p of connection probabilities

if $t_i=r$ and $t_j=s$, there is a link $i \to j$ with probability p_{rs}

sparse: $p=O(1/n)$



The stochastic block model

nodes have discrete labels: *k* "groups" or types of nodes

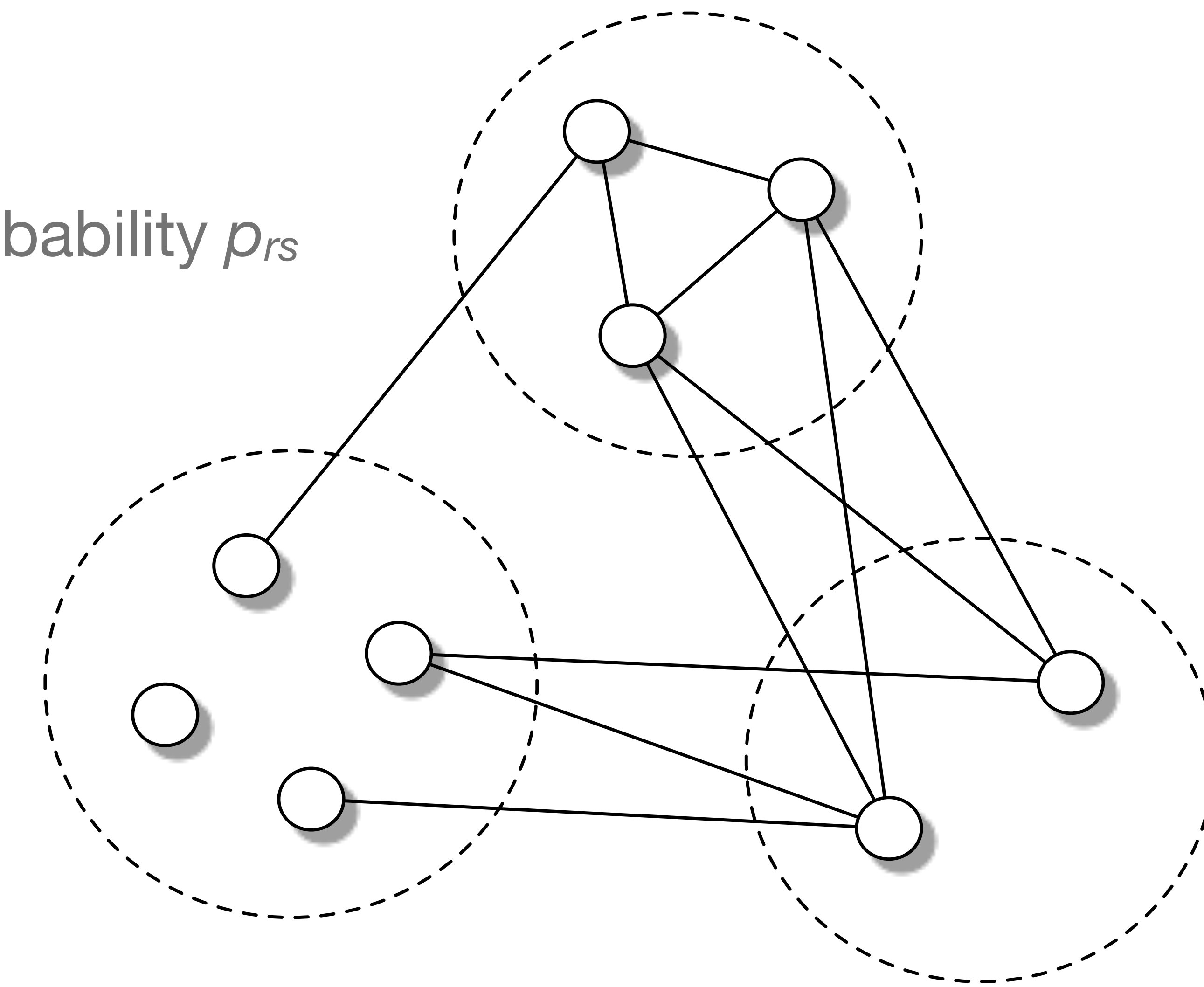
$k \times k$ matrix p of connection probabilities

if $t_i = r$ and $t_j = s$, there is a link $i \rightarrow j$ with probability p_{rs}

sparse: $p = O(1/n)$

popular special case:

$$p = \frac{1}{n} \begin{pmatrix} c_{\mathrm{in}} & \cdots & c_{\mathrm{out}} \\ \vdots & \ddots & \\ c_{\mathrm{out}} & & c_{\mathrm{in}} \end{pmatrix}$$



The stochastic block model

nodes have discrete labels: *k* “groups” or types of nodes

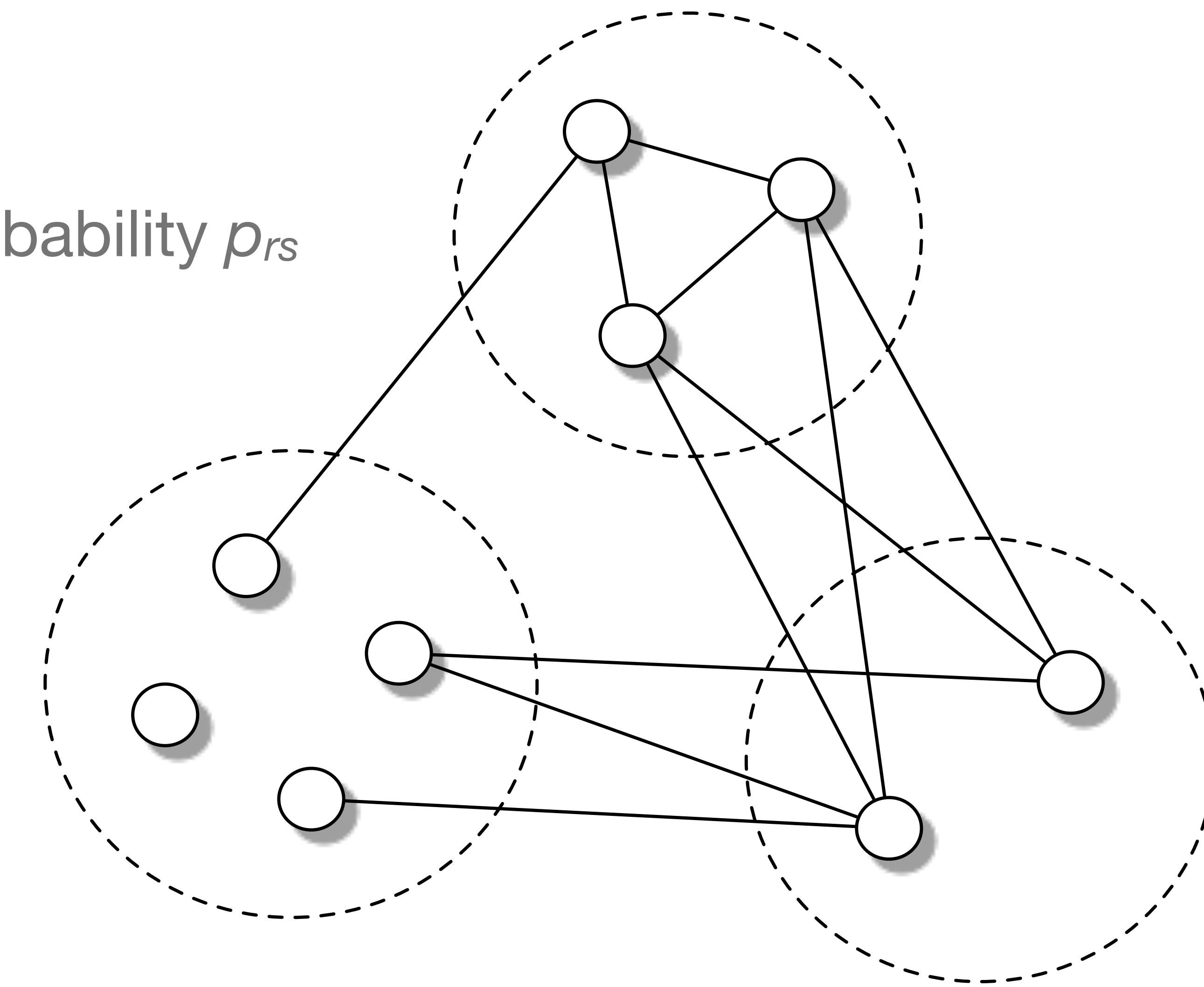
$k \times k$ matrix *p* of connection probabilities

if $t_i = r$ and $t_j = s$, there is a link $i \to j$ with probability p_{rs}

sparse: $p = O(1/n)$

popular special case:

$$p = \frac{1}{n} \begin{pmatrix} c_{\text{in}} & \cdots & c_{\text{out}} \\ \vdots & \ddots & \\ c_{\text{out}} & & c_{\text{in}} \end{pmatrix}$$



ferromagnetic (assortative, homophilic) if $c_{\text{in}} > c_{\text{out}}$

Likelihood and energy

Likelihood and energy

the probability of *G* given the types *t* is a product over edges and non-edges:

$$P(G \mid t) = \prod_{(i,j) \in E} p_{t_i, t_j} \prod_{(i,j) \notin E} (1 - p_{t_i, t_j})$$

Likelihood and energy

the probability of *G* given the types *t* is a product over edges and non-edges:

$$P(G \mid t) = \prod_{(i,j)\in E} p_{t_i,t_j} \prod_{(i,j)\notin E} (1 - p_{t_i,t_j})$$

the corresponding energy is

$$E(t) = -\log P(G \mid t) = -\sum_{(i,j)\in E} \log p_{t_i,t_j} - \sum_{(i,j)\notin E} \log(1 - p_{t_i,t_j})$$

Likelihood and energy

the probability of G given the types t is a product over edges and non-edges:

$$P(G \mid t) = \prod_{(i,j) \in E} p_{t_i,t_j} \prod_{(i,j) \notin E} (1 - p_{t_i,t_j})$$

the corresponding energy is

$$E(t) = -\log P(G \mid t) = -\sum_{(i,j) \in E} \log p_{t_i,t_j} - \sum_{(i,j) \notin E} \log(1 - p_{t_i,t_j})$$

like Ising model, but with weak antiferromagnetic interactions on non-edges

Likelihood and energy

the probability of *G* given the types *t* is a product over edges and non-edges:

$$P(G \mid t) = \prod_{(i,j) \in E} p_{t_i,t_j} \prod_{(i,j) \notin E} (1 - p_{t_i,t_j})$$

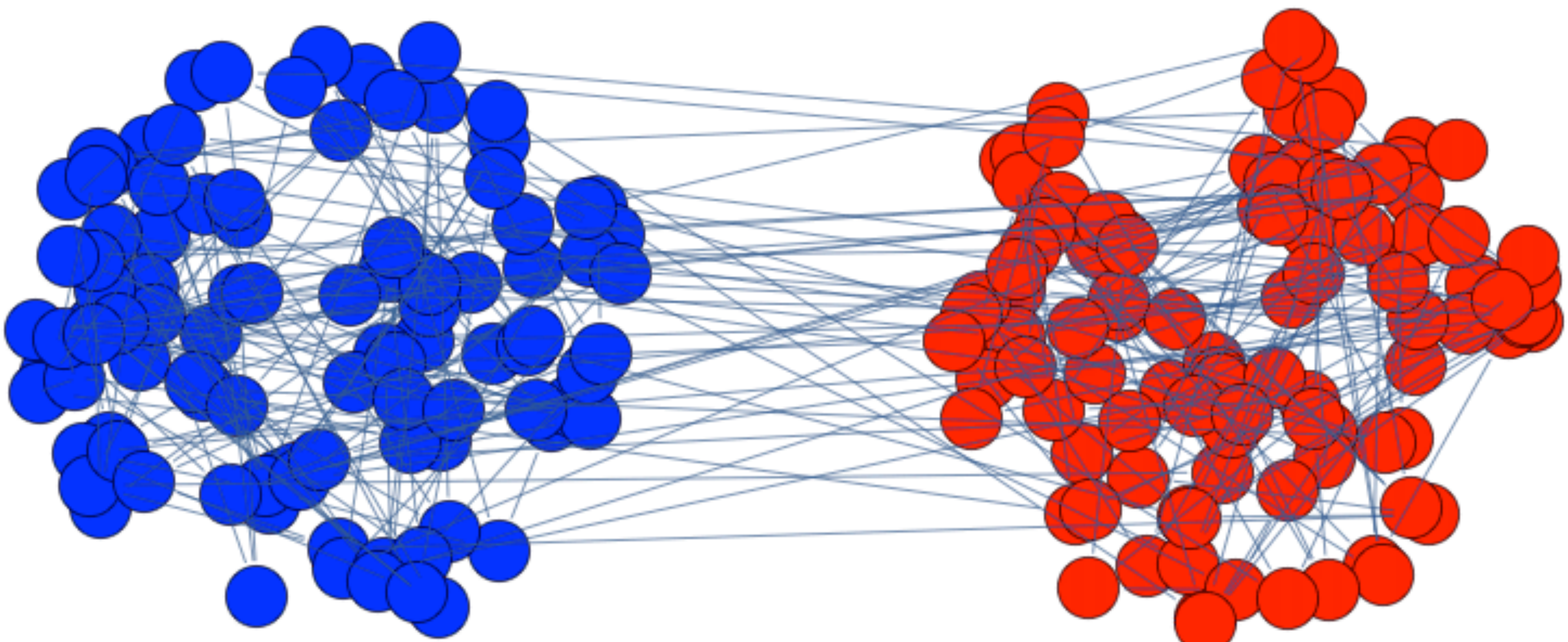
the corresponding energy is

$$E(t) = -\log P(G \mid t) = -\sum_{(i,j) \in E} \log p_{t_i,t_j} - \sum_{(i,j) \notin E} \log(1 - p_{t_i,t_j})$$

like Ising model, but with weak antiferromagnetic interactions on non-edges

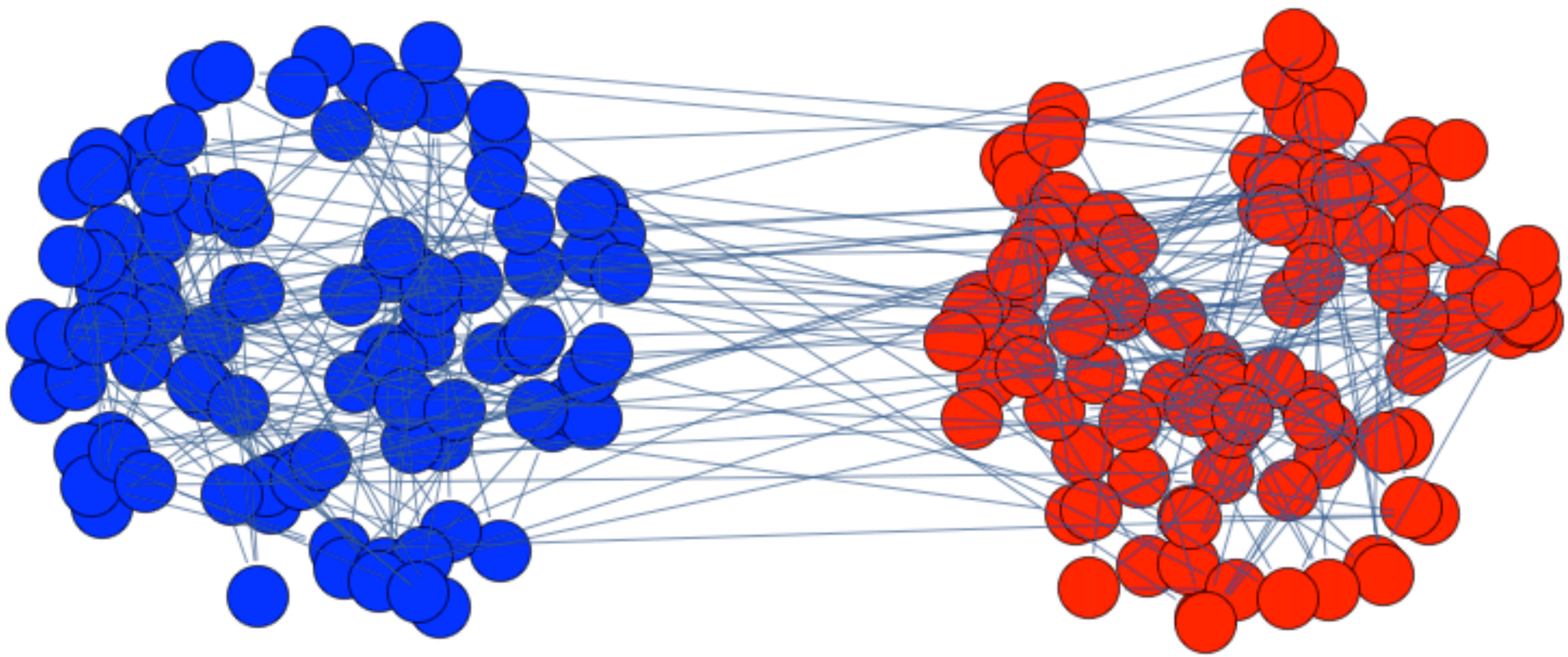
what can we learn from the “physics” of the block model?

Ground states vs. the landscape



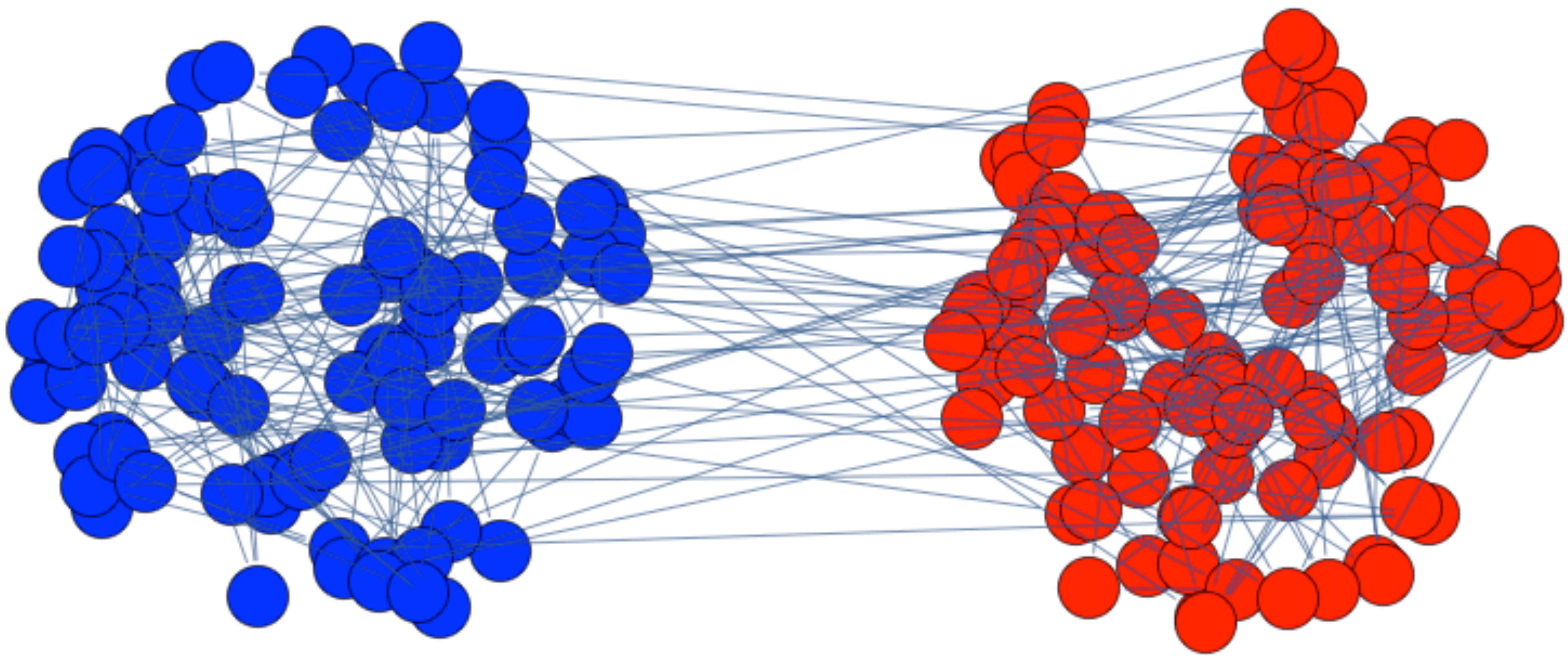
Ground states vs. the landscape

even random graphs have good-looking communities: only 11% of edges cross!



Ground states vs. the landscape

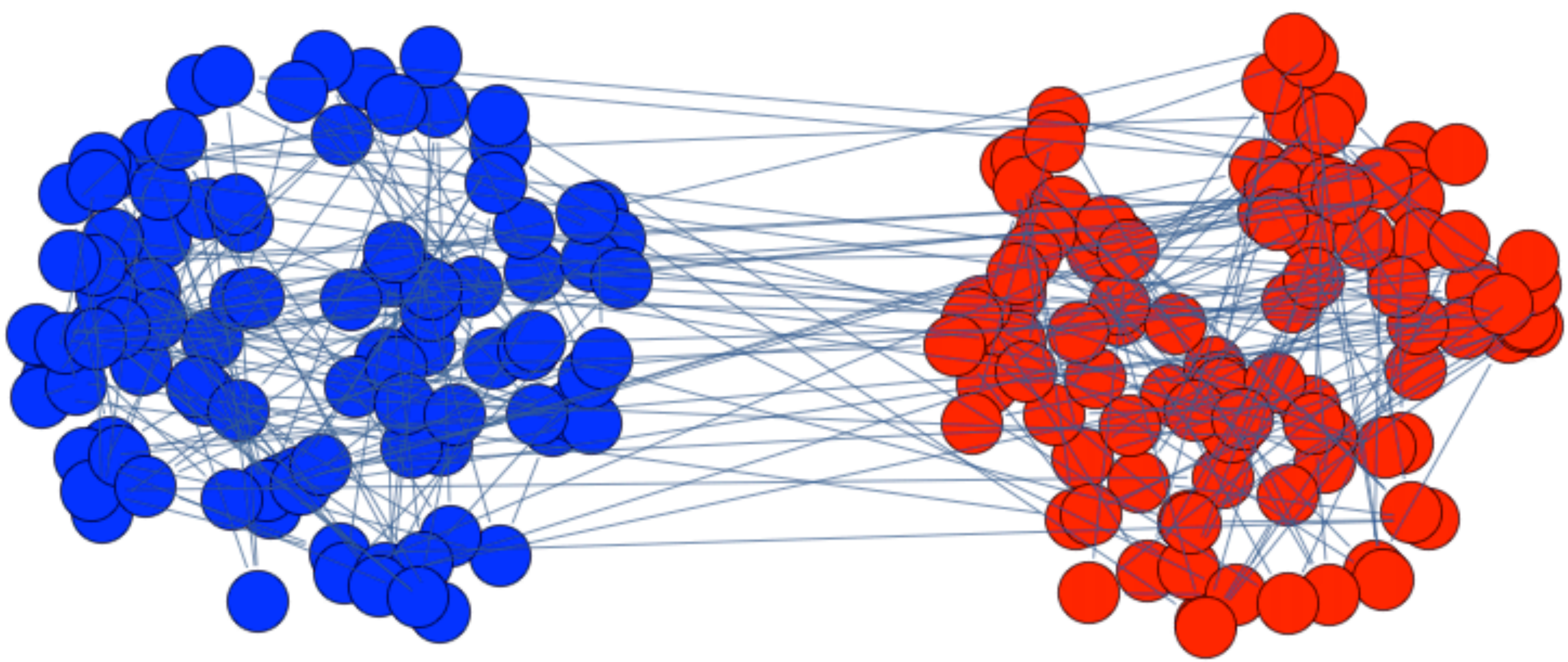
even random graphs have good-looking communities: only 11% of edges cross!



Ground states vs. the landscape

even random graphs have good-looking communities: only 11% of edges cross!

many local optima, with nothing in common

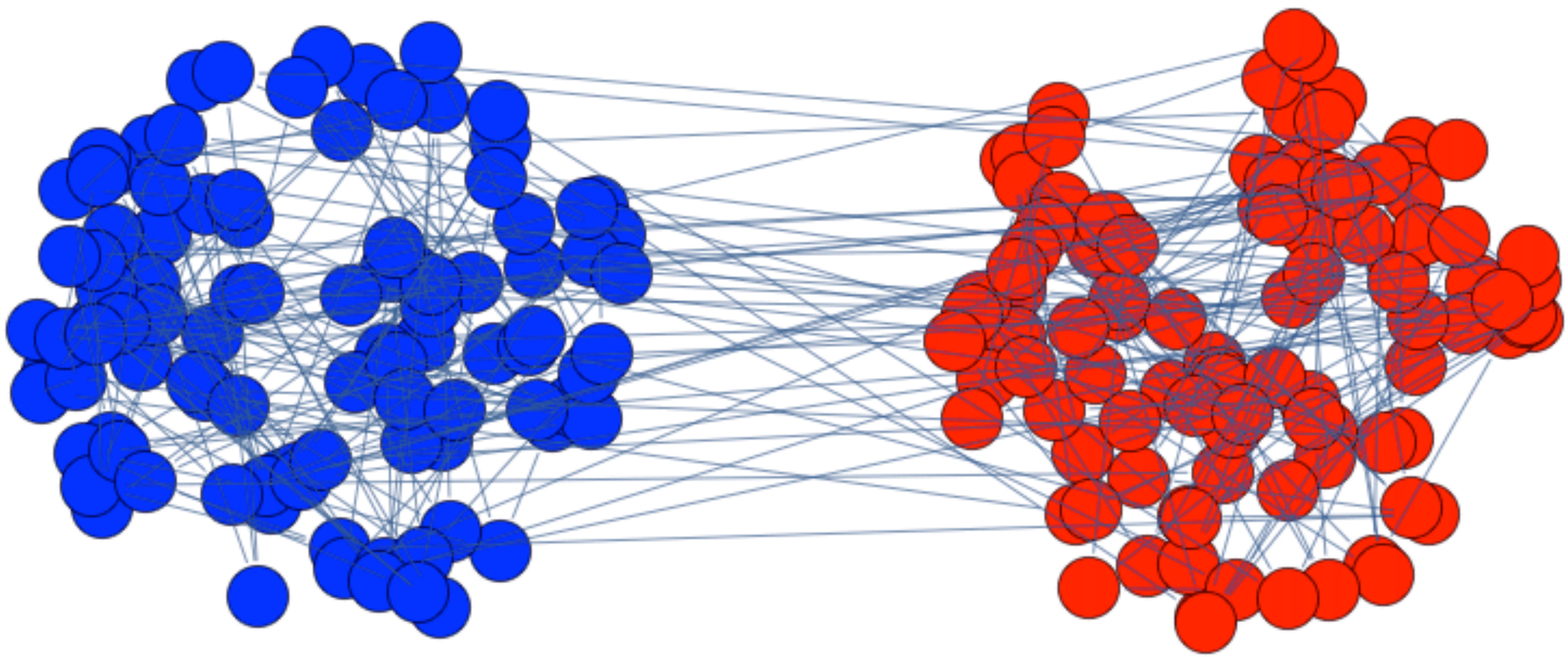


Ground states vs. the landscape

even random graphs have good-looking communities: only 11% of edges cross!

many local optima, with nothing in common

we need to understand the entire landscape, not just the optimum



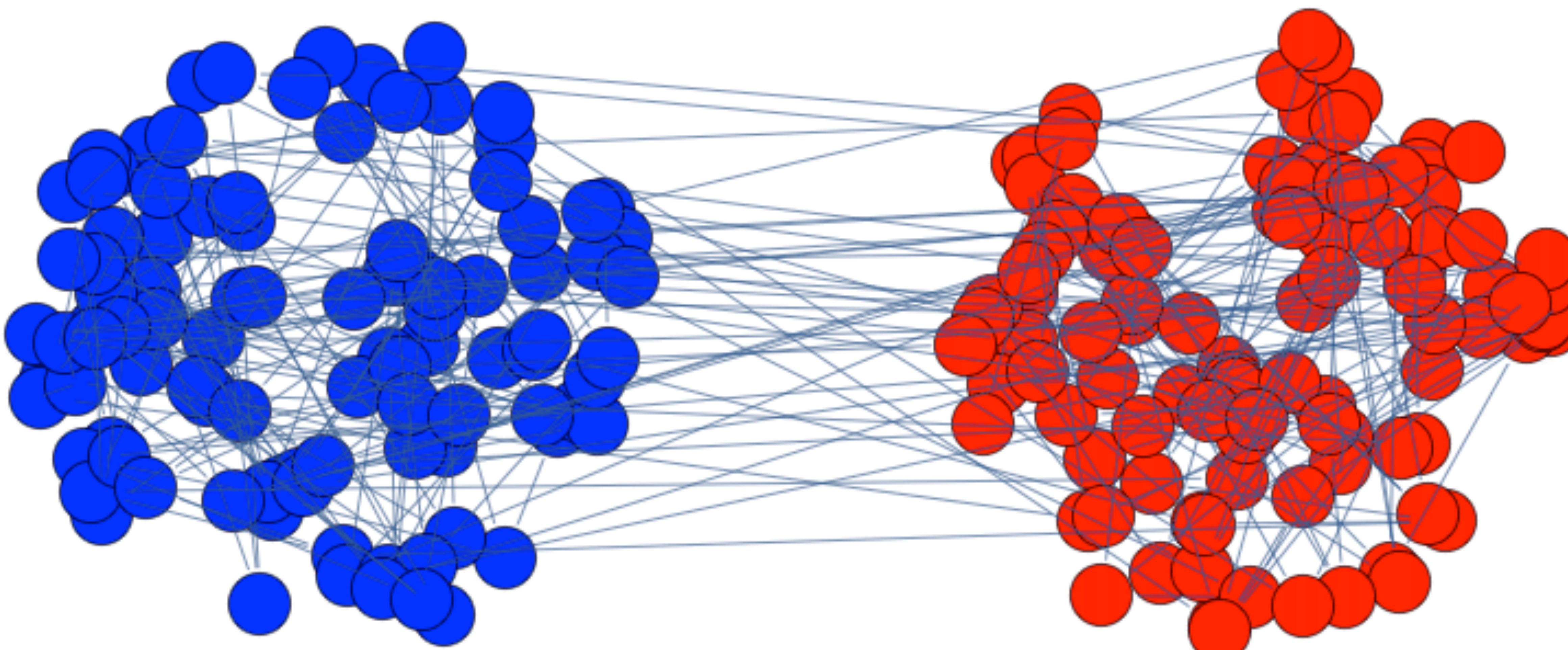
Ground states vs. the landscape

even random graphs have good-looking communities: only 11% of edges cross!

many local optima, with nothing in common

we need to understand the entire landscape, not just the optimum

otherwise, we could be *overfitting…*



Overfitting

Overfitting

we, and our algorithms, are prone to false positives

Overfitting

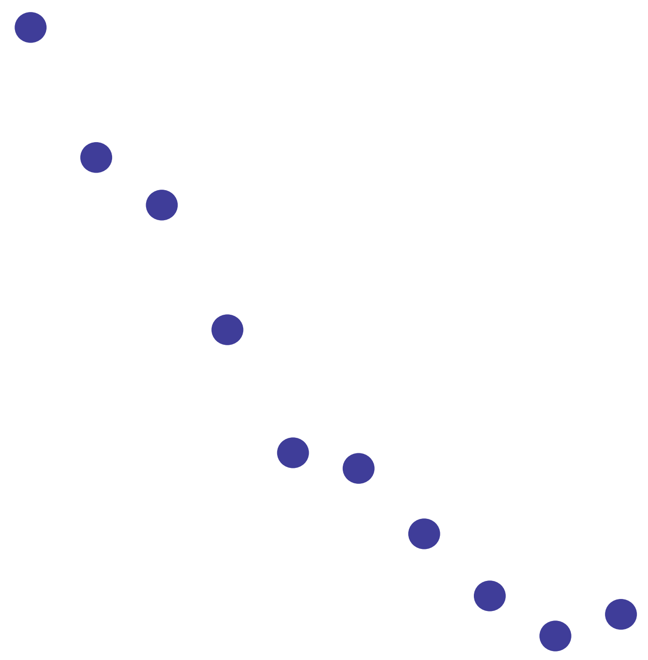
we, and our algorithms, are prone to false positives

fitting the data with fancy models is tempting…

Overfitting

we, and our algorithms, are prone to false positives

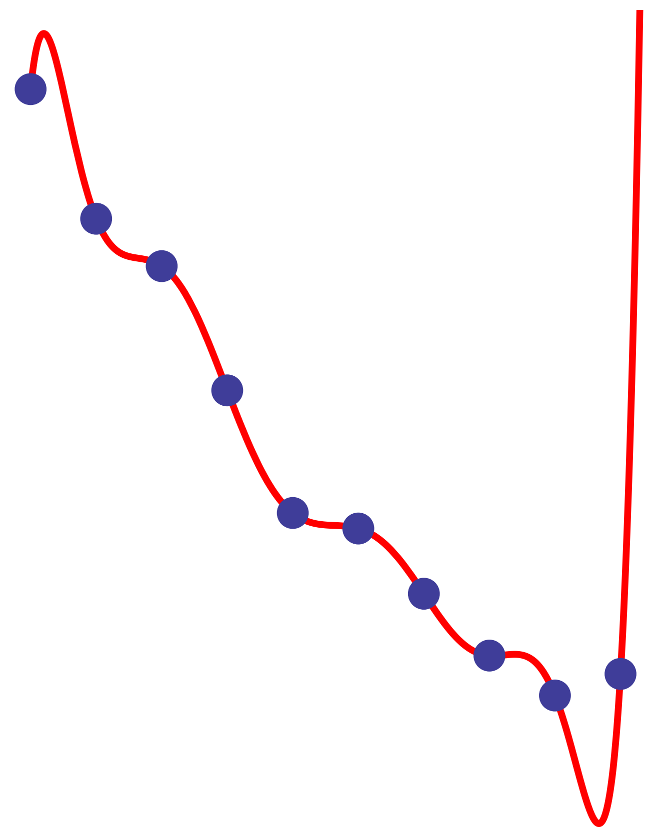
fitting the data with fancy models is tempting…



Overfitting

we, and our algorithms, are prone to false positives

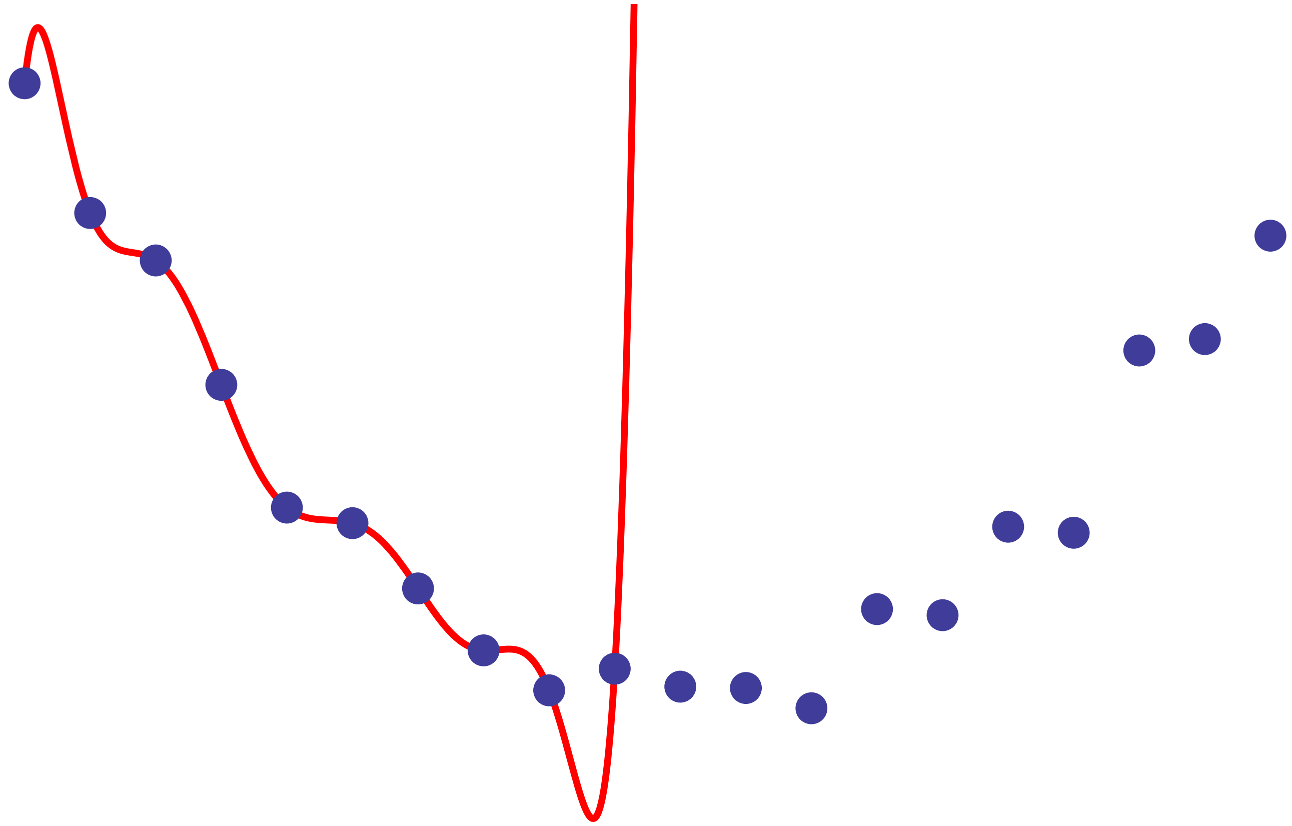
fitting the data with fancy models is tempting…



Overfitting

we, and our algorithms, are prone to false positives

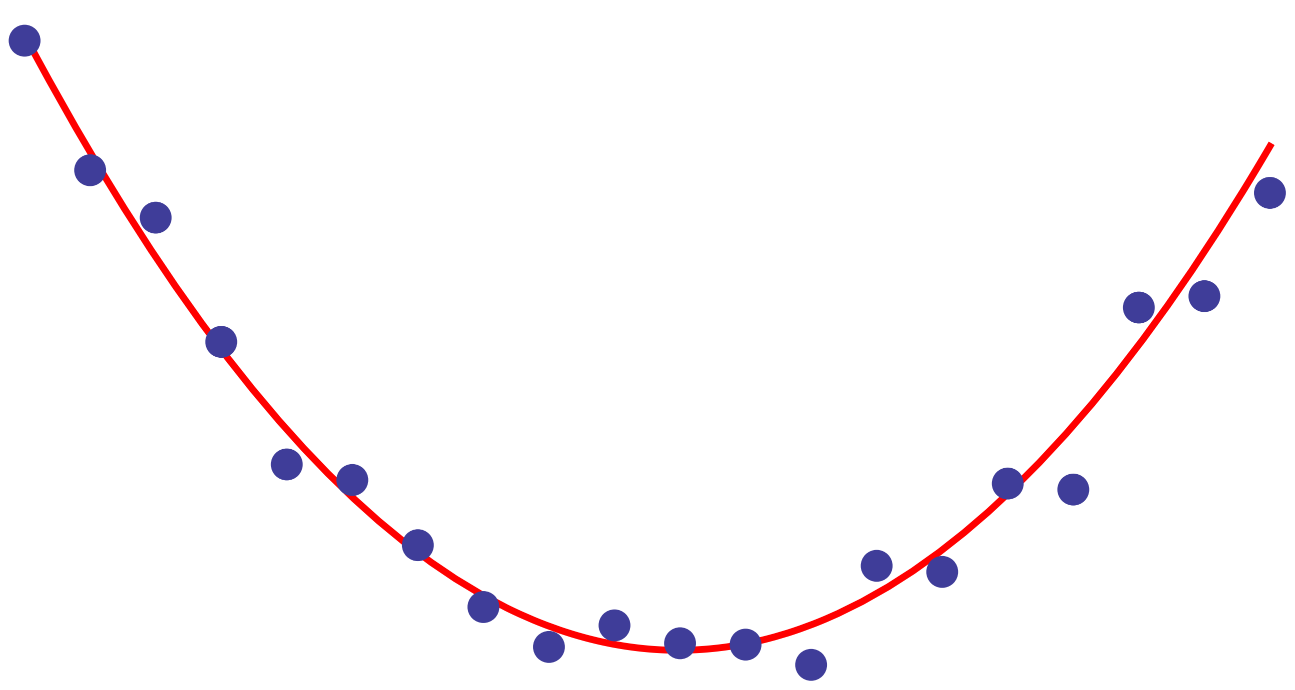
fitting the data with fancy models is tempting…



Overfitting

we, and our algorithms, are prone to false positives

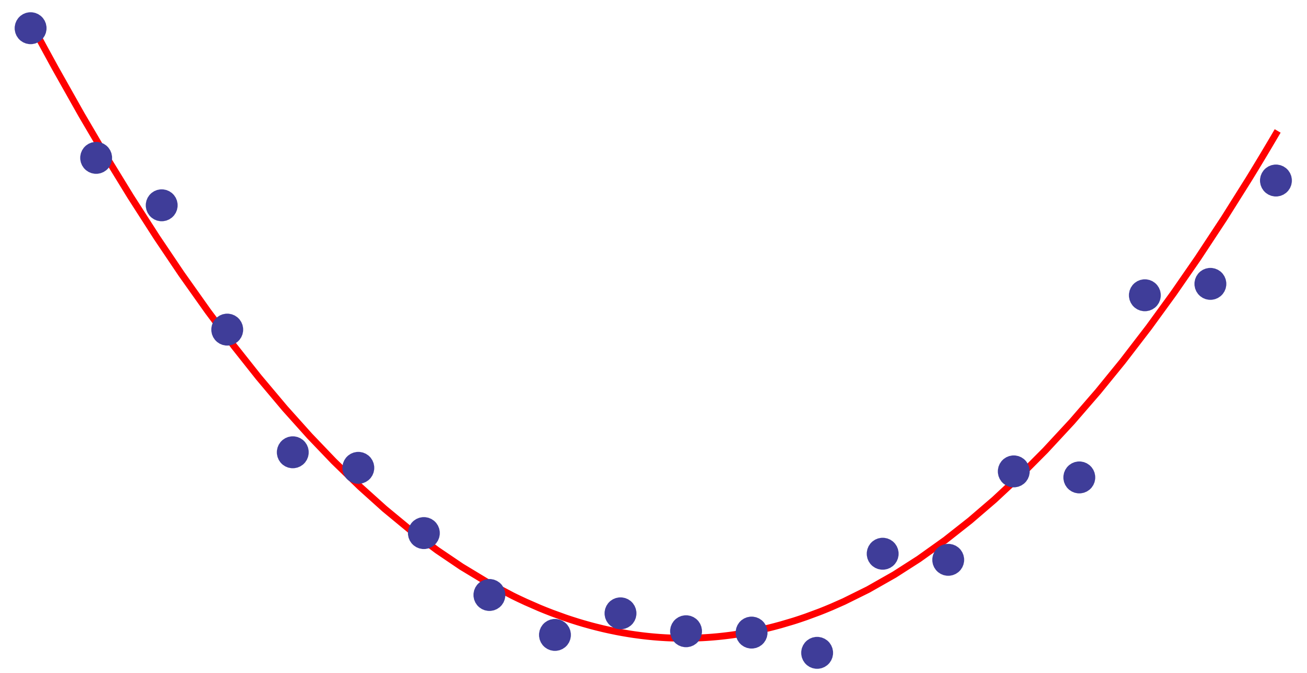
fitting the data with fancy models is tempting…



Overfitting

we, and our algorithms, are prone to false positives

fitting the data with fancy models is tempting…

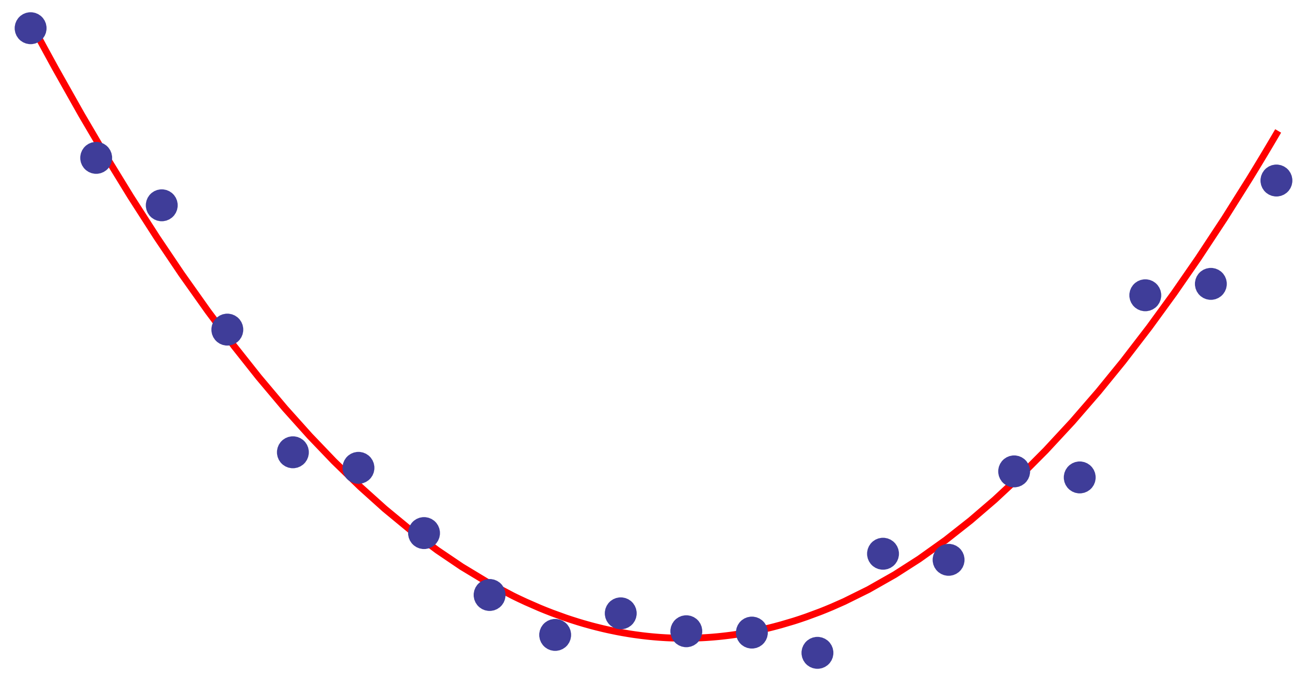


but often we’re really fitting the noise, not the underlying process

Overfitting

we, and our algorithms, are prone to false positives

fitting the data with fancy models is tempting…



but often we're really fitting the noise, not the underlying process

we want to understand the coin, not the coin flips

Information in the block model: the effect of a link

Information in the block model: the effect of a link

k equal groups, $p = \frac{1}{n} \begin{pmatrix} c_{\text{in}} & \cdots & c_{\text{out}} \\ \vdots & \ddots & \\ c_{\text{out}} & & c_{\text{in}} \end{pmatrix}$: average degree $c = \frac{c_{\text{in}} + (k-1)c_{\text{out}}}{k}$

Information in the block model: the effect of a link

k equal groups, $p = \frac{1}{n}\begin{pmatrix} c_{\text{in}} & \cdots & c_{\text{out}} \\ \vdots & \ddots & \\ c_{\text{out}} & & c_{\text{in}} \end{pmatrix}$: average degree $c = \frac{c_{\text{in}} + (k-1)c_{\text{out}}}{k}$

if there is a link $i \to j$, the probability distribution of t_j is related to that of t_i by a transition matrix

Information in the block model: the effect of a link

k equal groups, $p = \frac{1}{n} \begin{pmatrix} c_{\text{in}} & \cdots & c_{\text{out}} \\ \vdots & \ddots & \\ c_{\text{out}} & & c_{\text{in}} \end{pmatrix}$: average degree $c = \frac{c_{\text{in}} + (k-1)c_{\text{out}}}{k}$

if there is a link $i \to j$, the probability distribution of t_j is related to that of t_i by a transition matrix

$$\frac{1}{kc} \begin{pmatrix} c_{\text{in}} & \cdots & c_{\text{out}} \\ \vdots & \ddots & \\ c_{\text{out}} & & c_{\text{in}} \end{pmatrix} = \lambda \mathbb{1} + (1-\lambda) \begin{pmatrix} 1/k & \cdots & 1/k \\ \vdots & \ddots & \\ 1/k & & 1/k \end{pmatrix}$$

where $\lambda = \frac{c_{\text{in}} - c_{\text{out}}}{kc}$

Information in the block model: the effect of a link

k equal groups, $p = \frac{1}{n}\begin{pmatrix} c_{\text{in}} & \cdots & c_{\text{out}} \\ \vdots & \ddots & \\ c_{\text{out}} & & c_{\text{in}} \end{pmatrix}$: average degree $c = \dfrac{c_{\text{in}} + (k-1)c_{\text{out}}}{k}$

if there is a link $i \to j$, the probability distribution of t_j is related to that of t_i by a transition matrix

$$\frac{1}{kc}\begin{pmatrix} c_{\text{in}} & \cdots & c_{\text{out}} \\ \vdots & \ddots & \\ c_{\text{out}} & & c_{\text{in}} \end{pmatrix} = \lambda \mathbb{1} + (1-\lambda)\begin{pmatrix} 1/k & \cdots & 1/k \\ \vdots & \ddots & \\ 1/k & & 1/k \end{pmatrix}$$

where $\lambda = \dfrac{c_{\text{in}} - c_{\text{out}}}{kc}$

with probability λ, copy from *i* to *j;* with probability $1-\lambda$, set *j*'s type randomly

Information in the block model: the effect of a link

k equal groups, $p = \frac{1}{n}\begin{pmatrix} c_{\text{in}} & \cdots & c_{\text{out}} \\ \vdots & \ddots & \\ c_{\text{out}} & & c_{\text{in}} \end{pmatrix}$: average degree $c = \frac{c_{\text{in}} + (k-1)c_{\text{out}}}{k}$

if there is a link $i \to j$, the probability distribution of t_j is related to that of t_i by a transition matrix

$$\frac{1}{kc}\begin{pmatrix} c_{\text{in}} & \cdots & c_{\text{out}} \\ \vdots & \ddots & \\ c_{\text{out}} & & c_{\text{in}} \end{pmatrix} = \lambda \mathbb{1} + (1-\lambda)\begin{pmatrix} 1/k & \cdots & 1/k \\ \vdots & \ddots & \\ 1/k & & 1/k \end{pmatrix}$$

where $\lambda = \frac{c_{\text{in}} - c_{\text{out}}}{kc}$

with probability λ, copy from *i* to *j;* with probability $1-\lambda$, set *j*'s type randomly

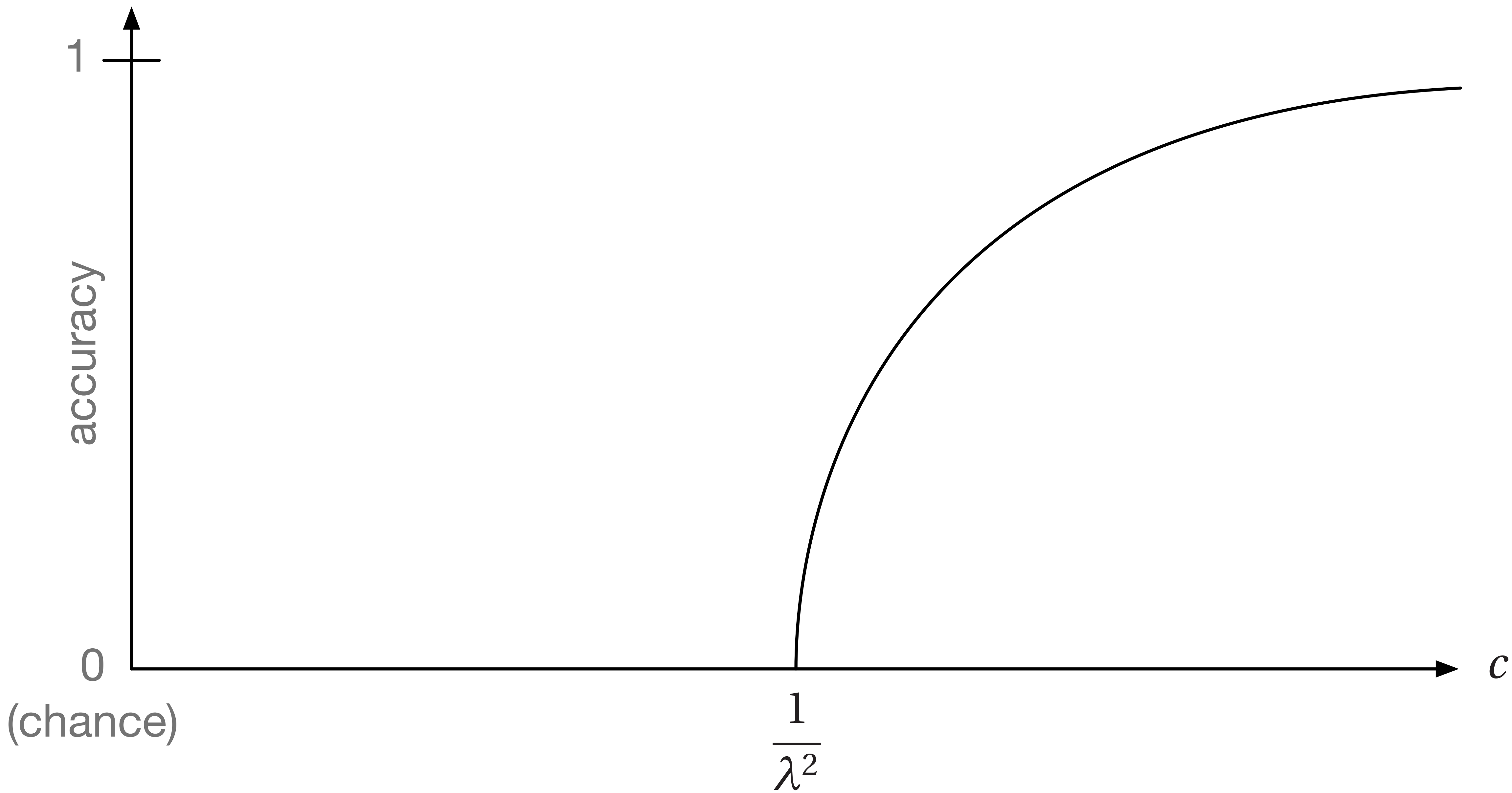
if λ is fixed, community detection gets easier as *c* increases…

Detectability thresholds

For two groups of equal size [DKMZ, MNS, M, KMMNSSZ, BLM]:

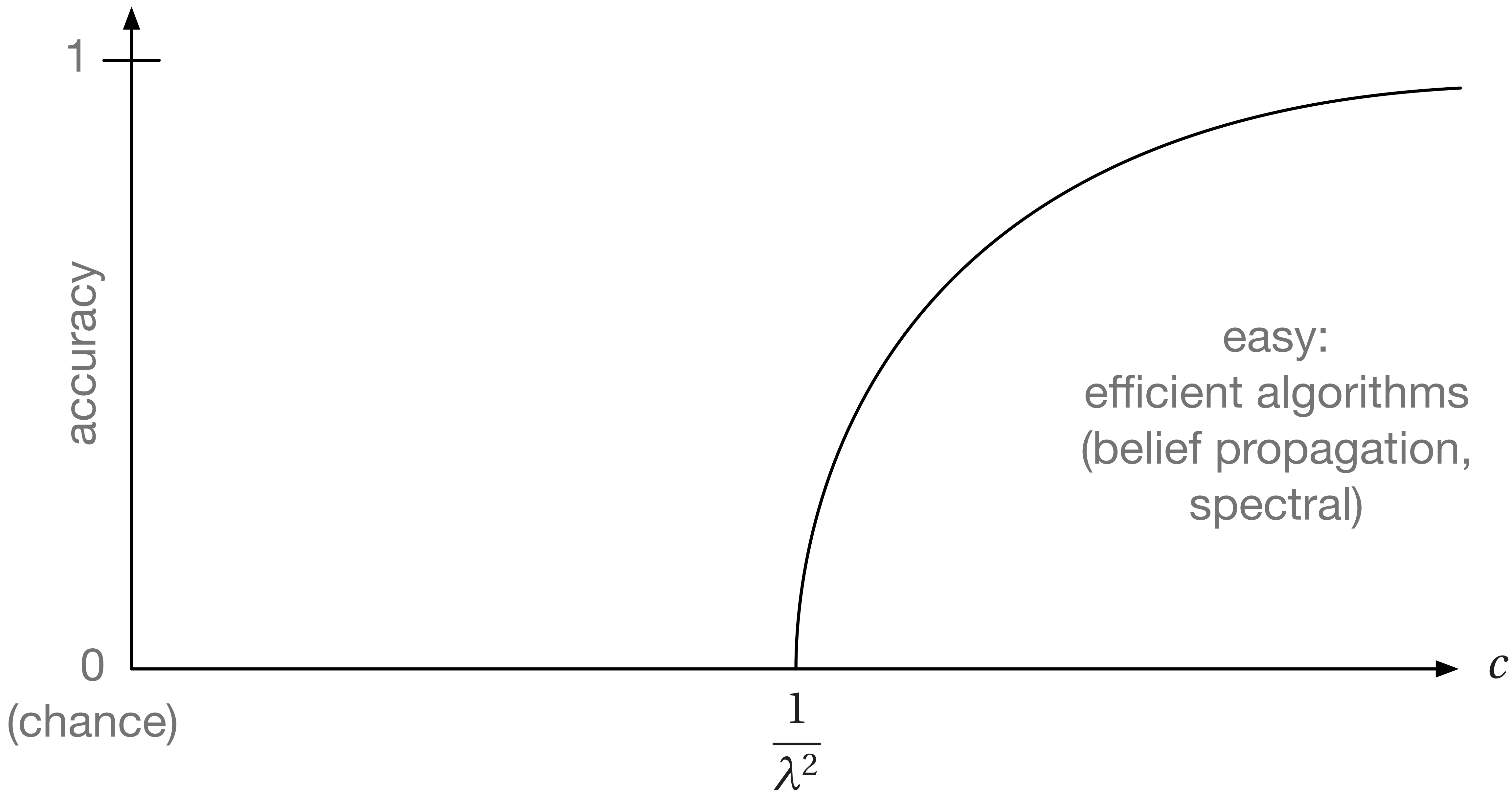
Detectability thresholds

For two groups of equal size [DKMZ, MNS, M, KMMNSSZ, BLM]:



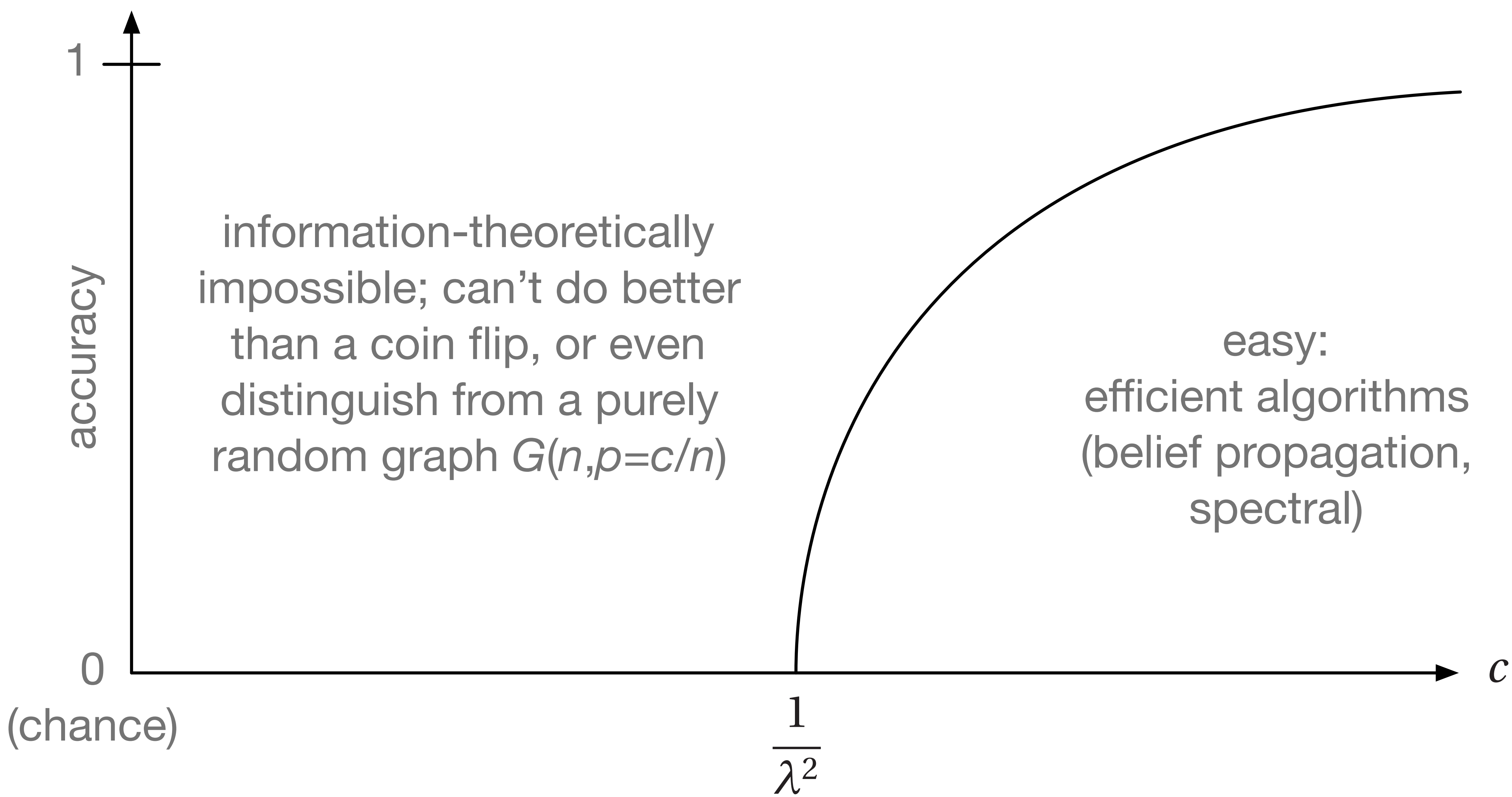
Detectability thresholds

For two groups of equal size [DKMZ, MNS, M, KMMNSSZ, BLM]:



Detectability thresholds

For two groups of equal size [DKMZ, MNS, M, KMMNSSZ, BLM]:

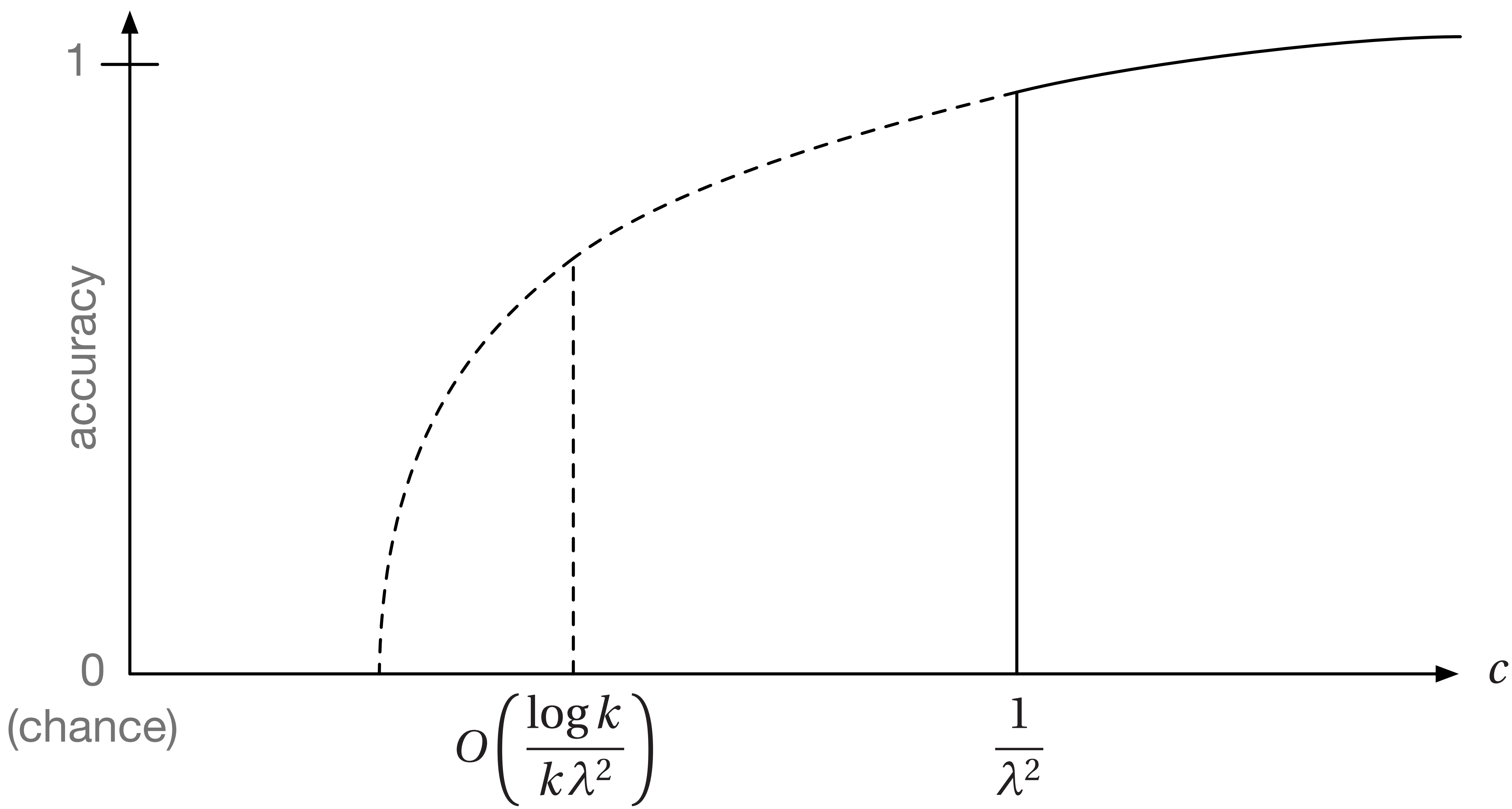


Detectability thresholds

For $k \geq 4$ groups [DKMZ, KMMNSSZ, BLM, BMNN, AS]:

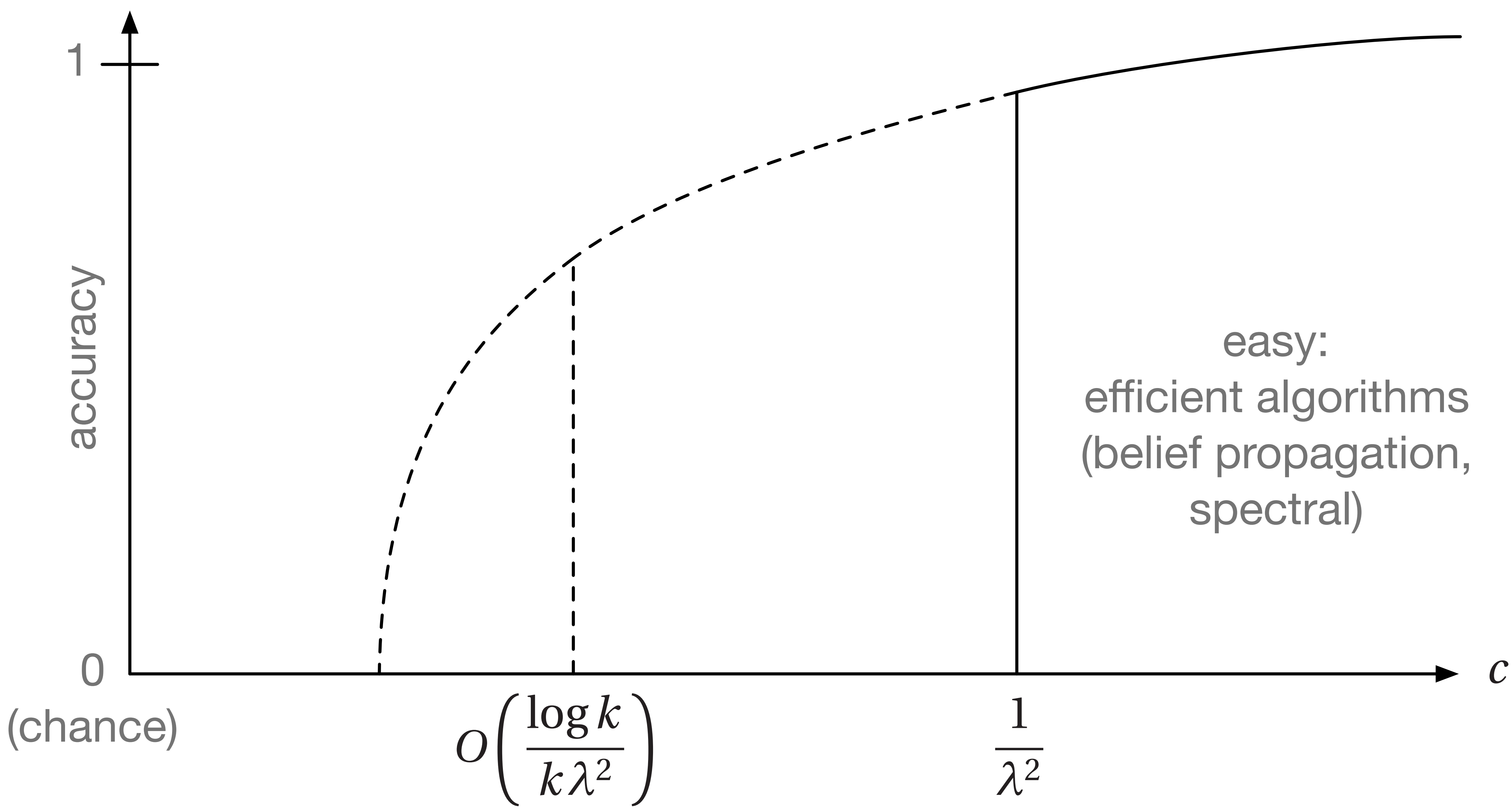
Detectability thresholds

For k≥4 groups [DKMZ, KMMNSSZ, BLM, BMNN, AS]:



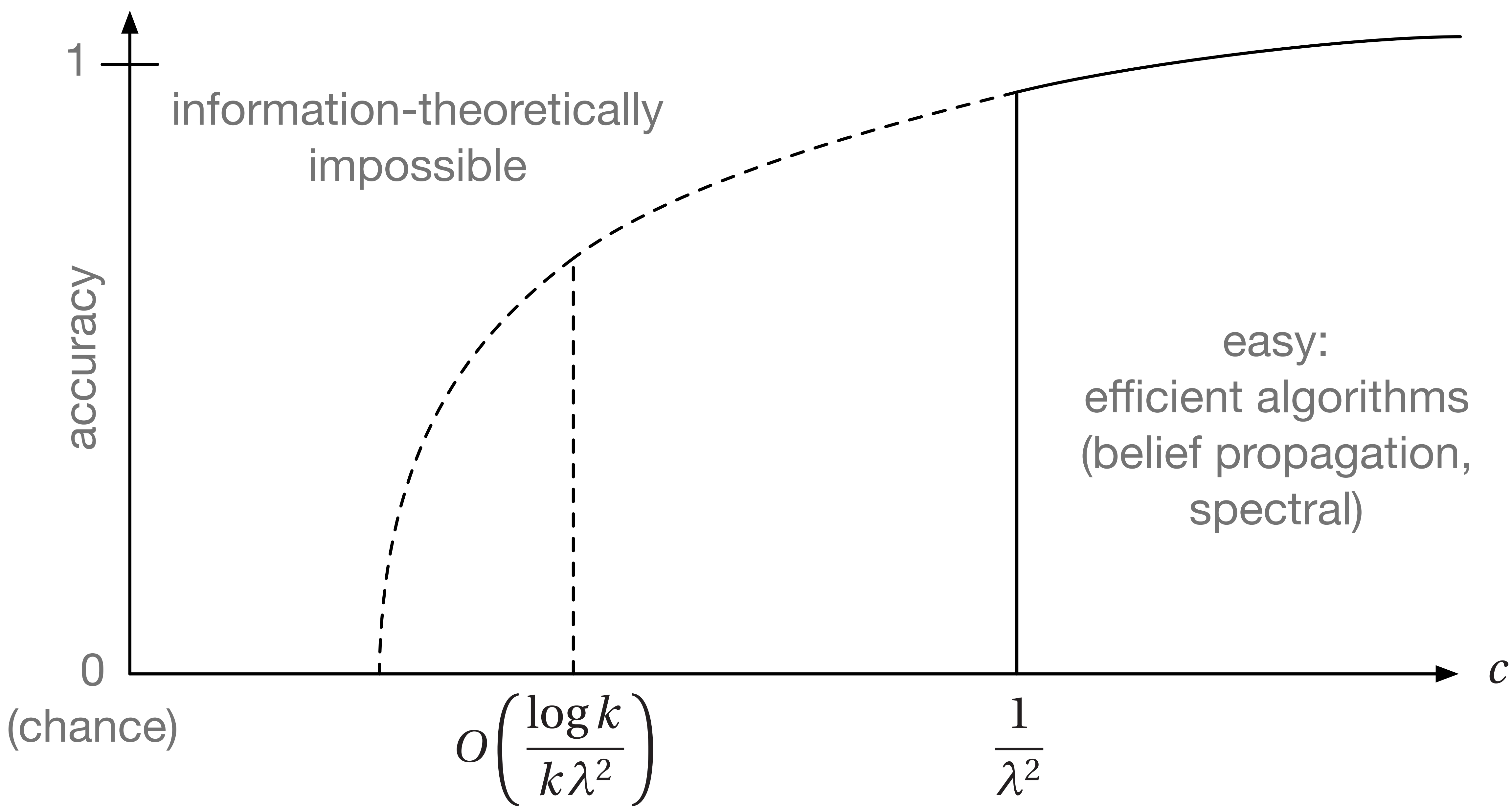
Detectability thresholds

For k≥4 groups [DKMZ, KMMNSSZ, BLM, BMNN, AS]:



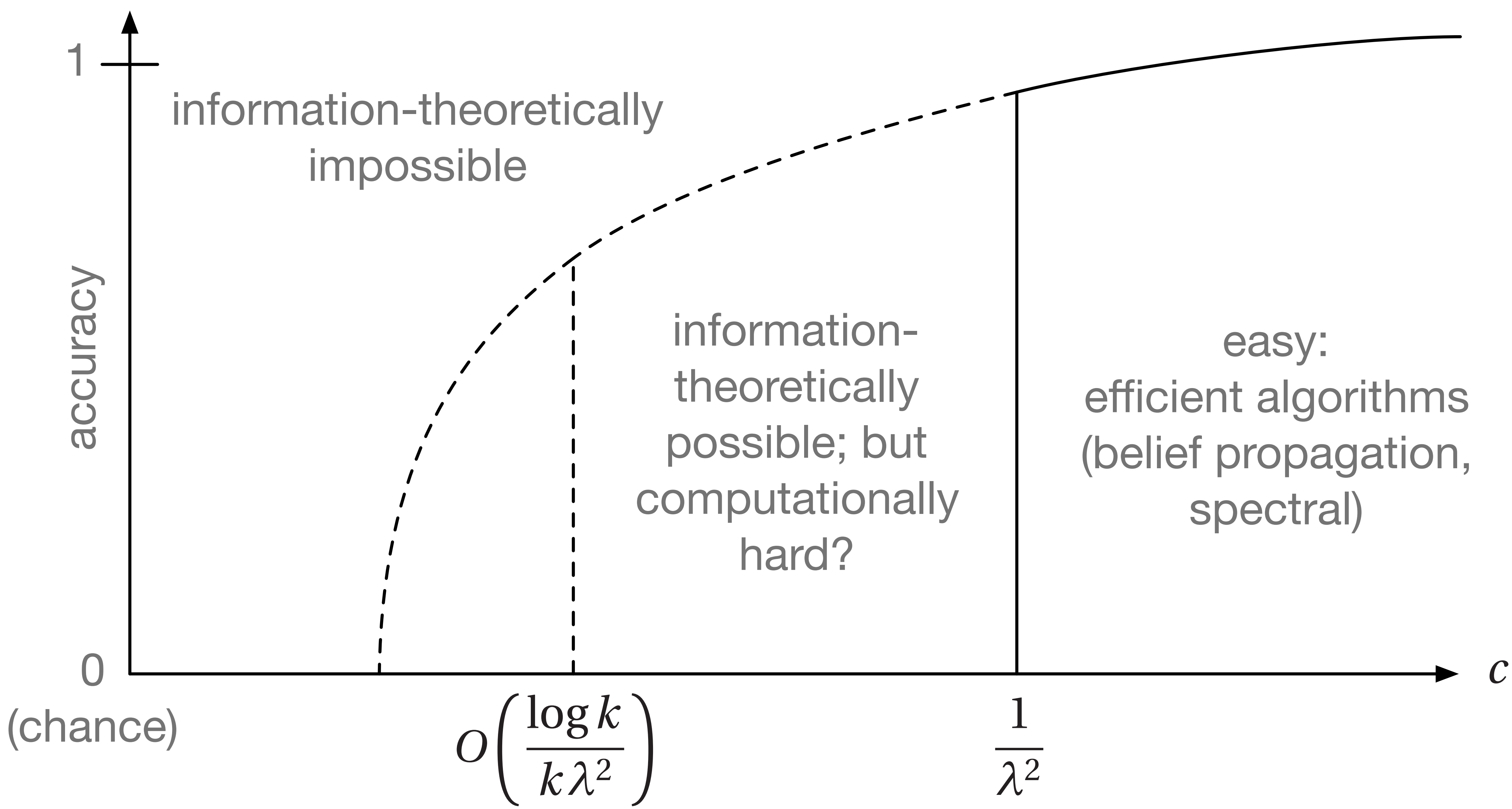
Detectability thresholds

For k≥4 groups [DKMZ, KMMNSSZ, BLM, BMNN, AS]:



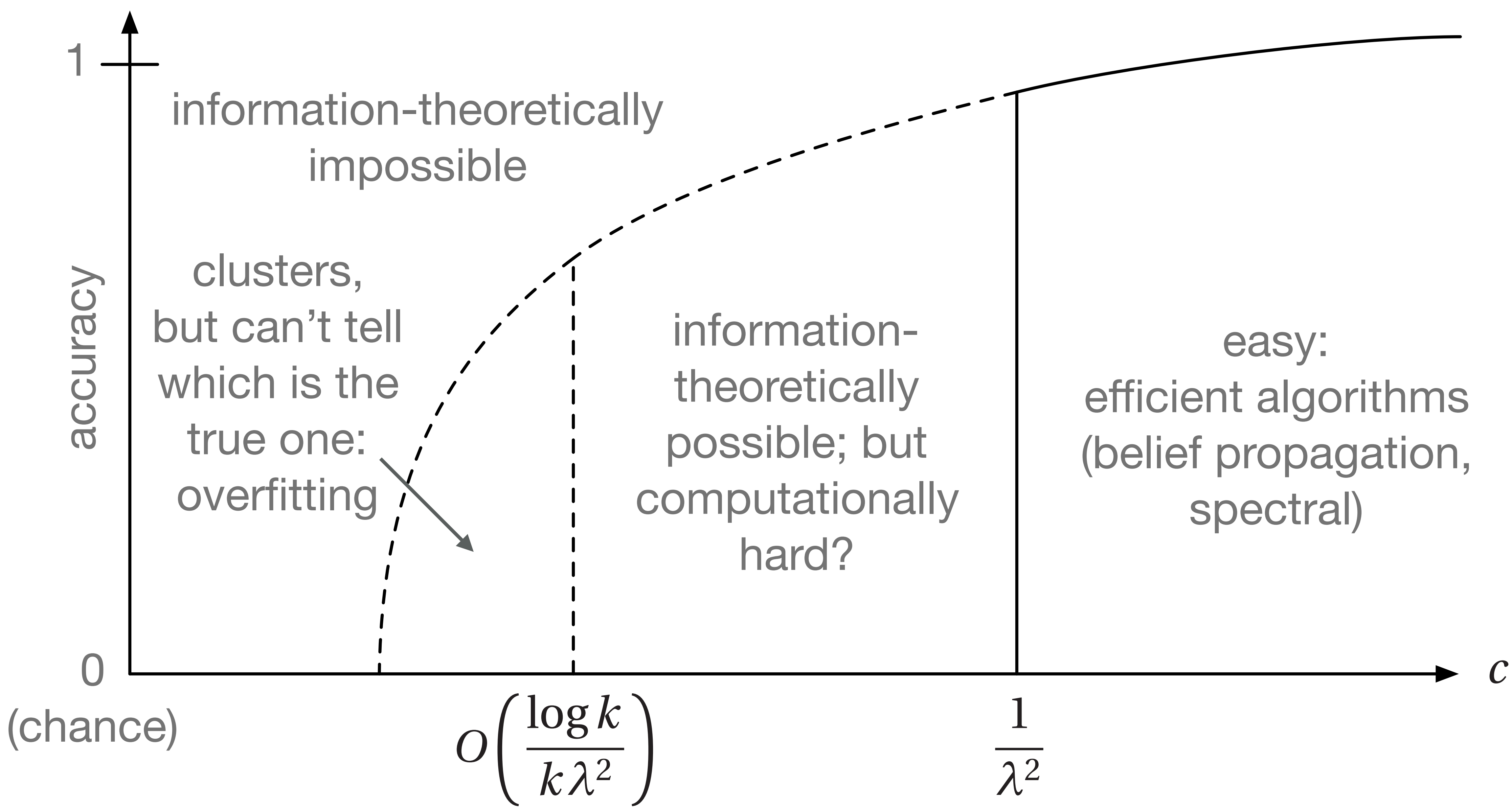
Detectability thresholds

For k≥4 groups [DKMZ, KMMNSSZ, BLM, BMNN, AS]:

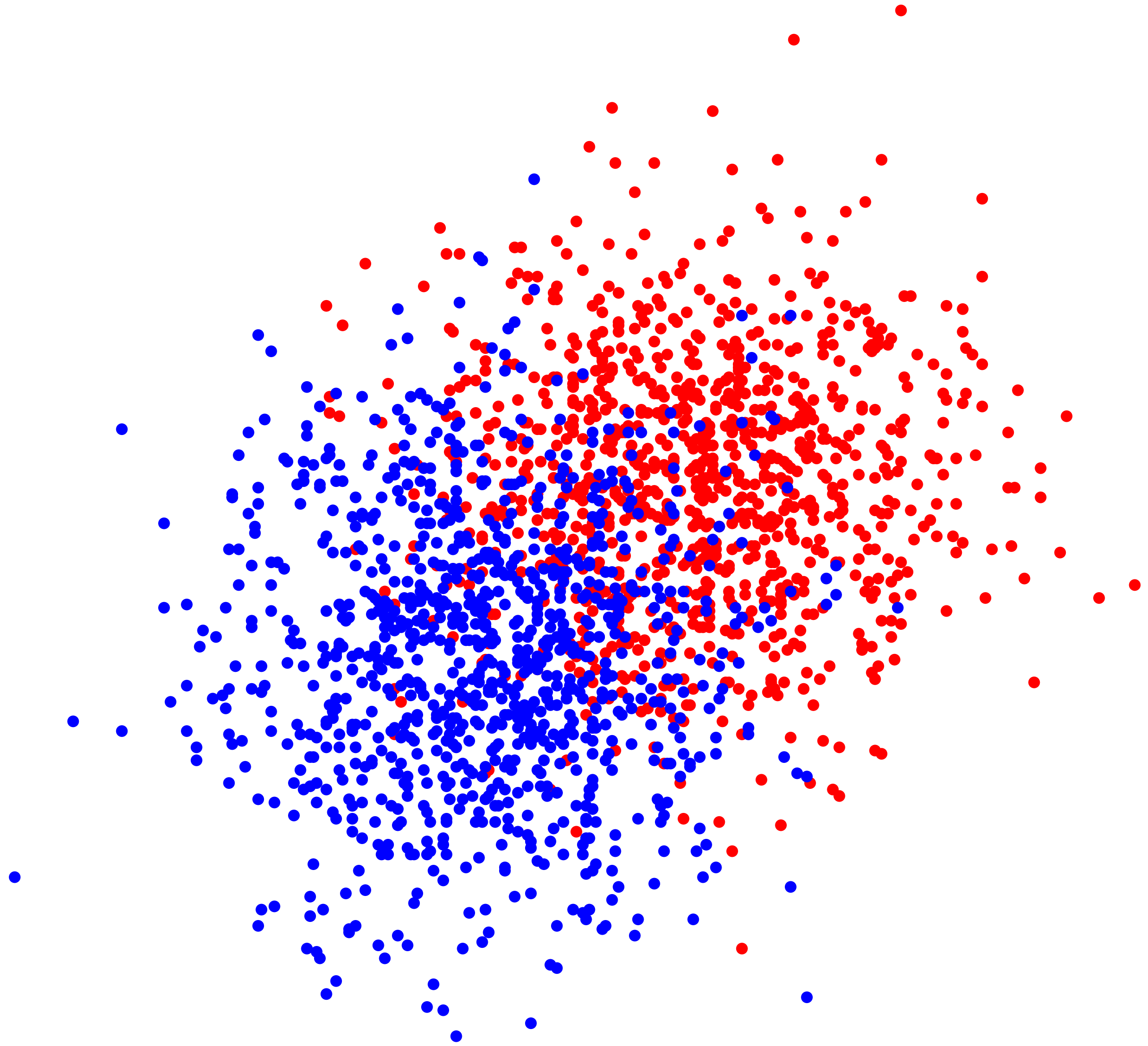


Detectability thresholds

For k≥4 groups [DKMZ, KMMNSSZ, BLM, BMNN, AS]:

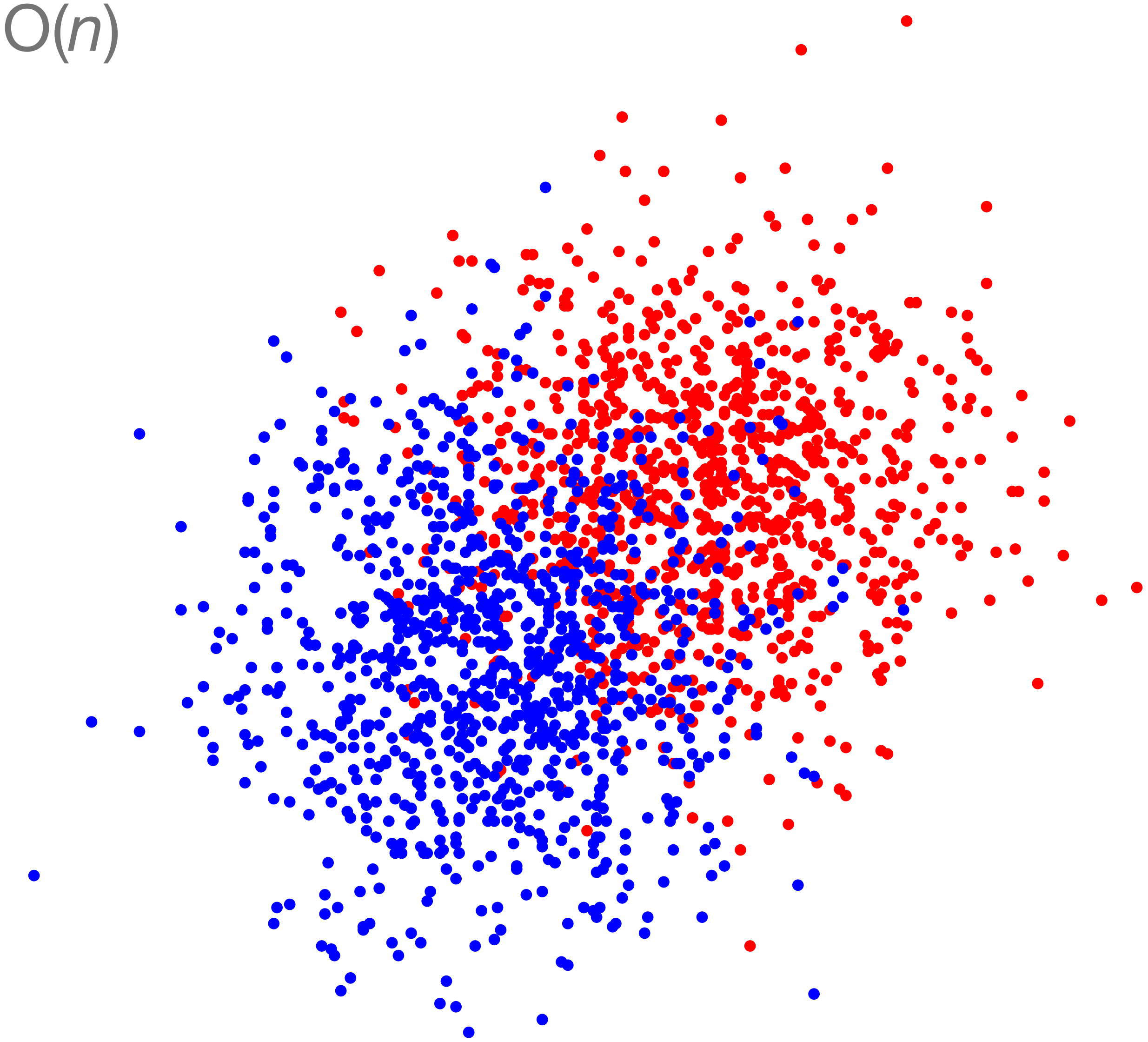


Clustering high-dimensional data



Clustering high-dimensional data

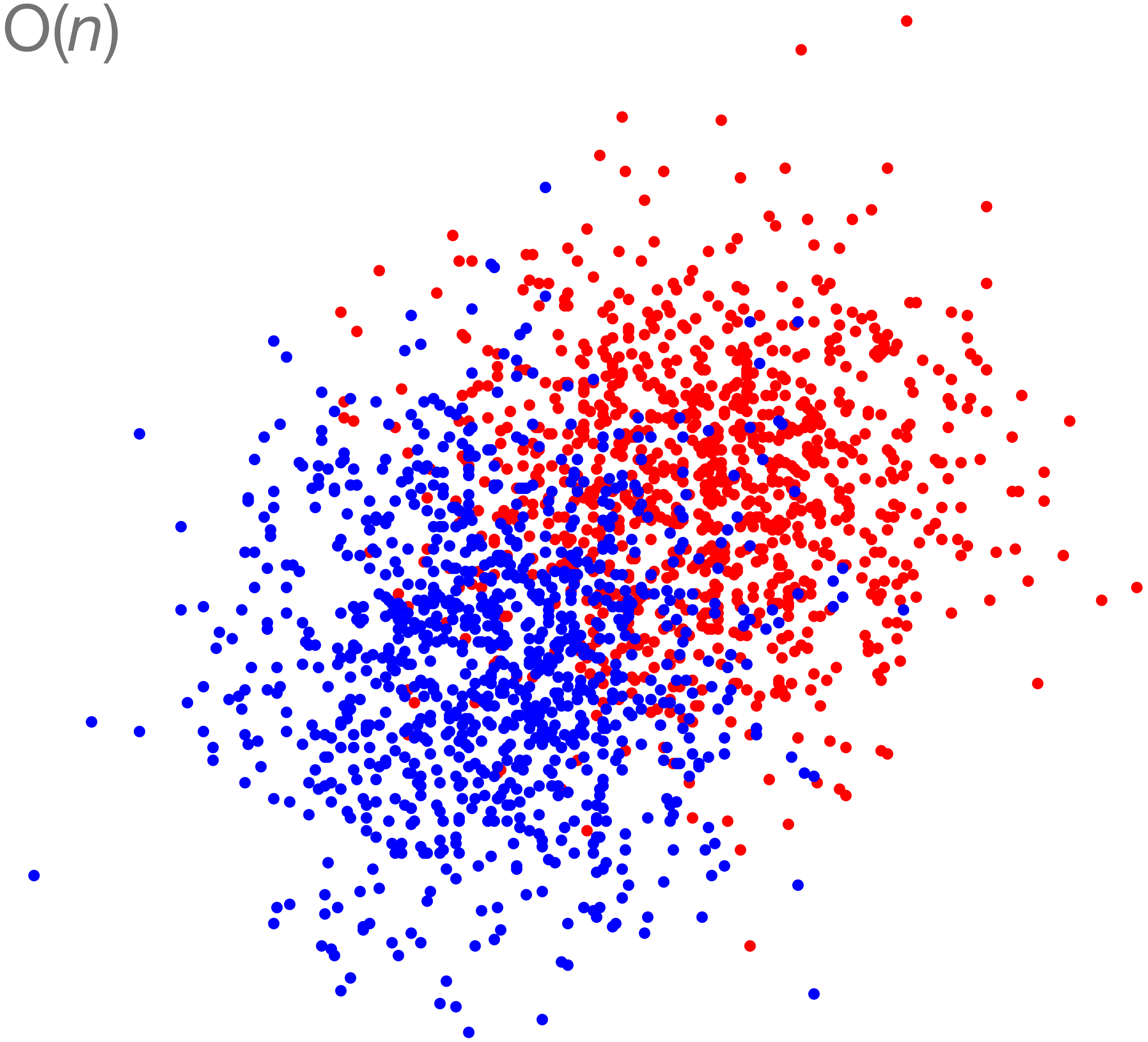
m points in *n*-dimensional space, where $m = O(n)$



Clustering high-dimensional data

m points in n-dimensional space, where m=O(n)

k clusters with Gaussian noise

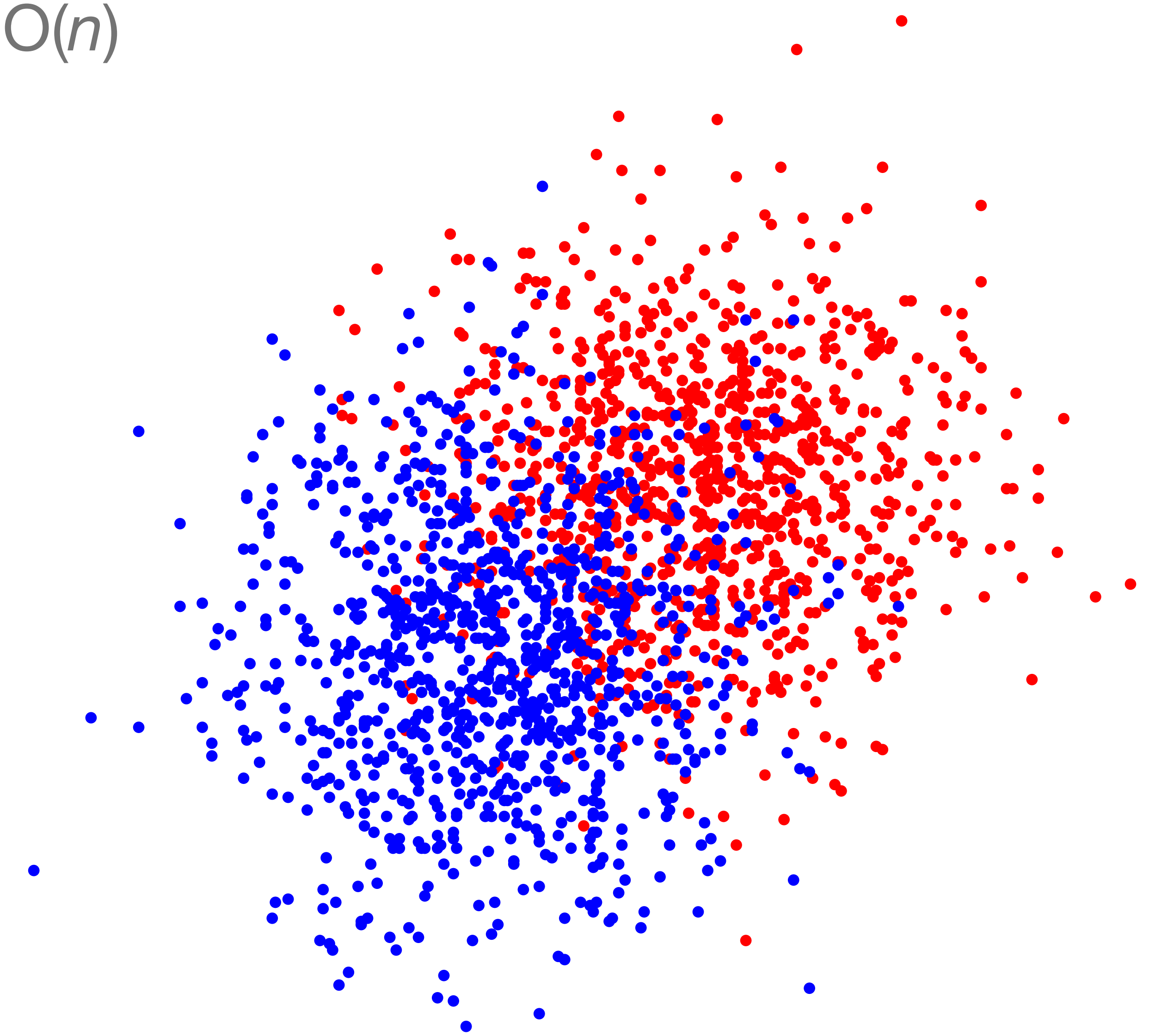


Clustering high-dimensional data

m points in n-dimensional space, where m=O(n)

k clusters with Gaussian noise

when can we...



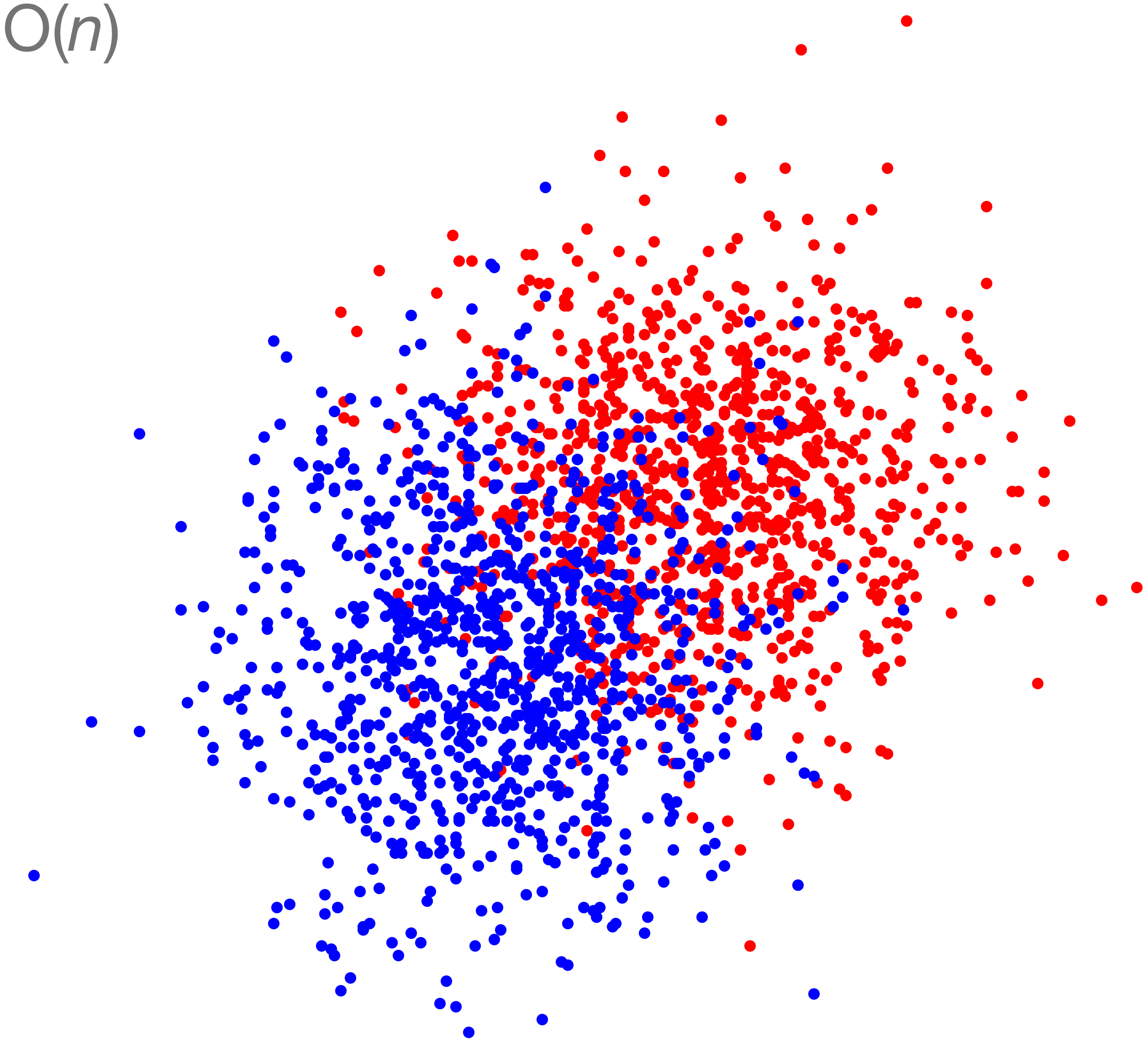
Clustering high-dimensional data

m points in *n*-dimensional space, where *m*=O(*n*)

k clusters with Gaussian noise

when can we...

find the cluster centers?



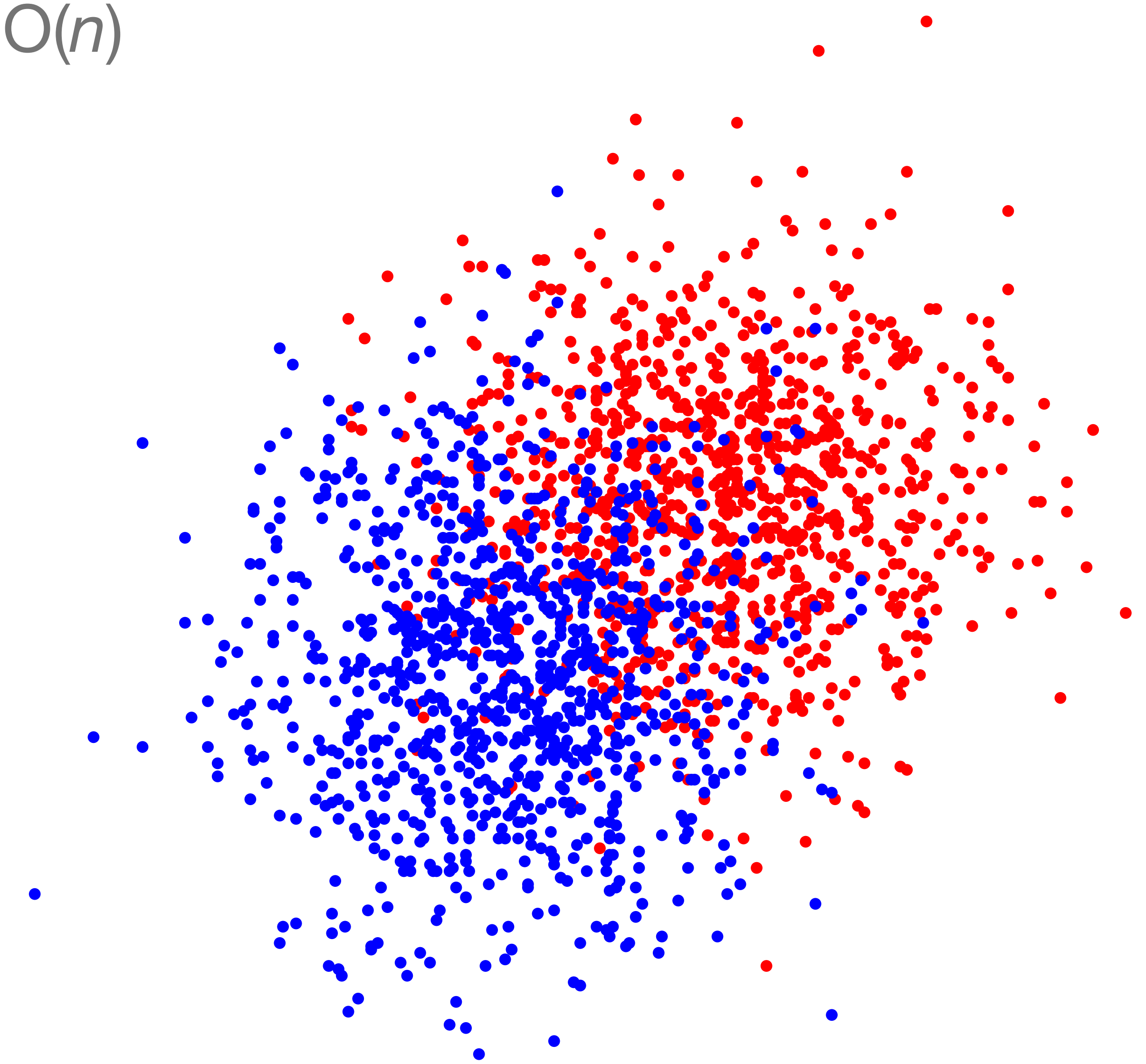
Clustering high-dimensional data

m points in n-dimensional space, where m=O(n)

k clusters with Gaussian noise

when can we...

- find the cluster centers?
- label the points better than chance?



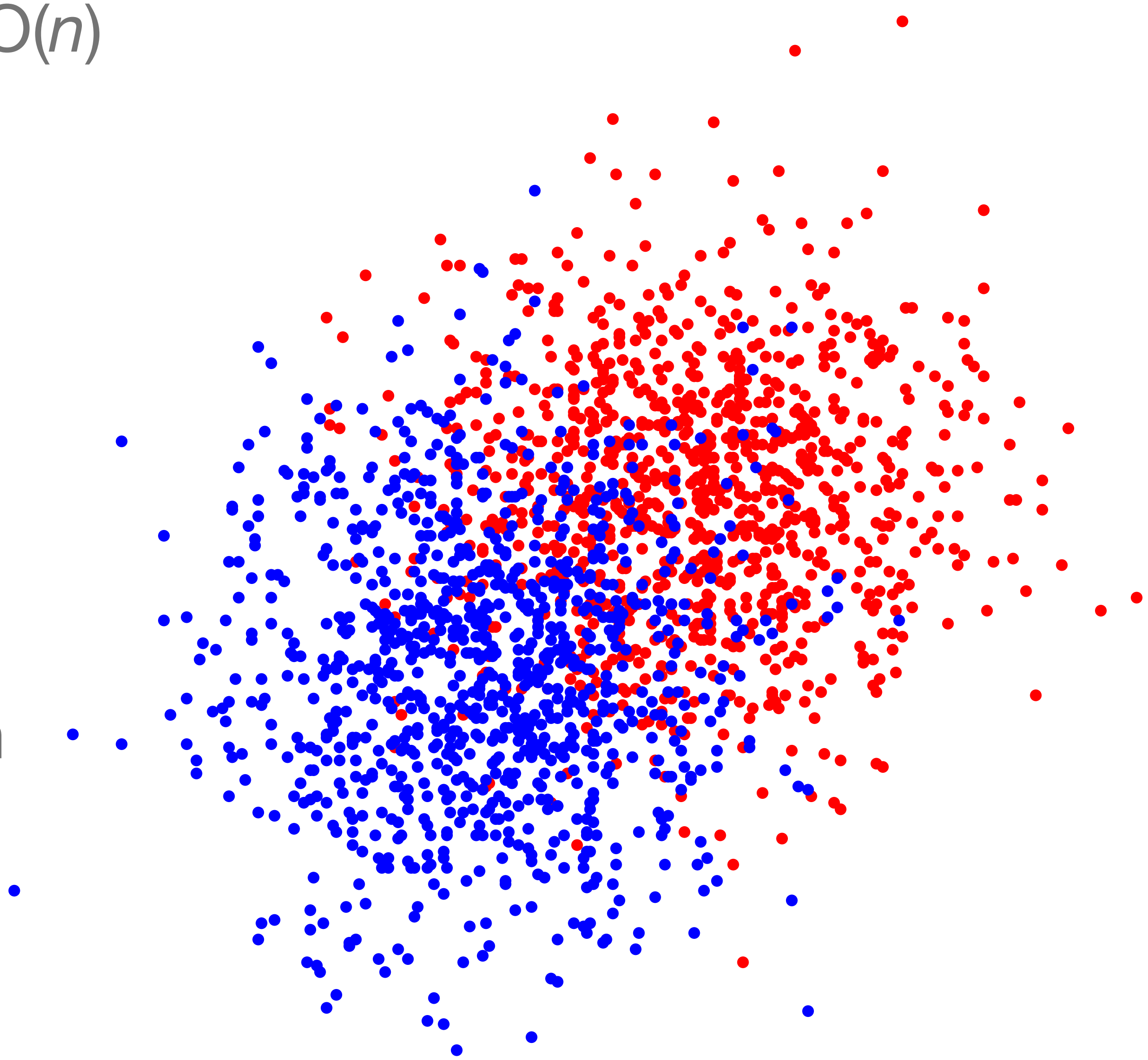
Clustering high-dimensional data

m points in n-dimensional space, where $m=O(n)$

k clusters with Gaussian noise

when can we...

- find the cluster centers?
- label the points better than chance?
- tell that there are clusters, i.e., distinguish from a null model with one big cluster?



Clustering high-dimensional data

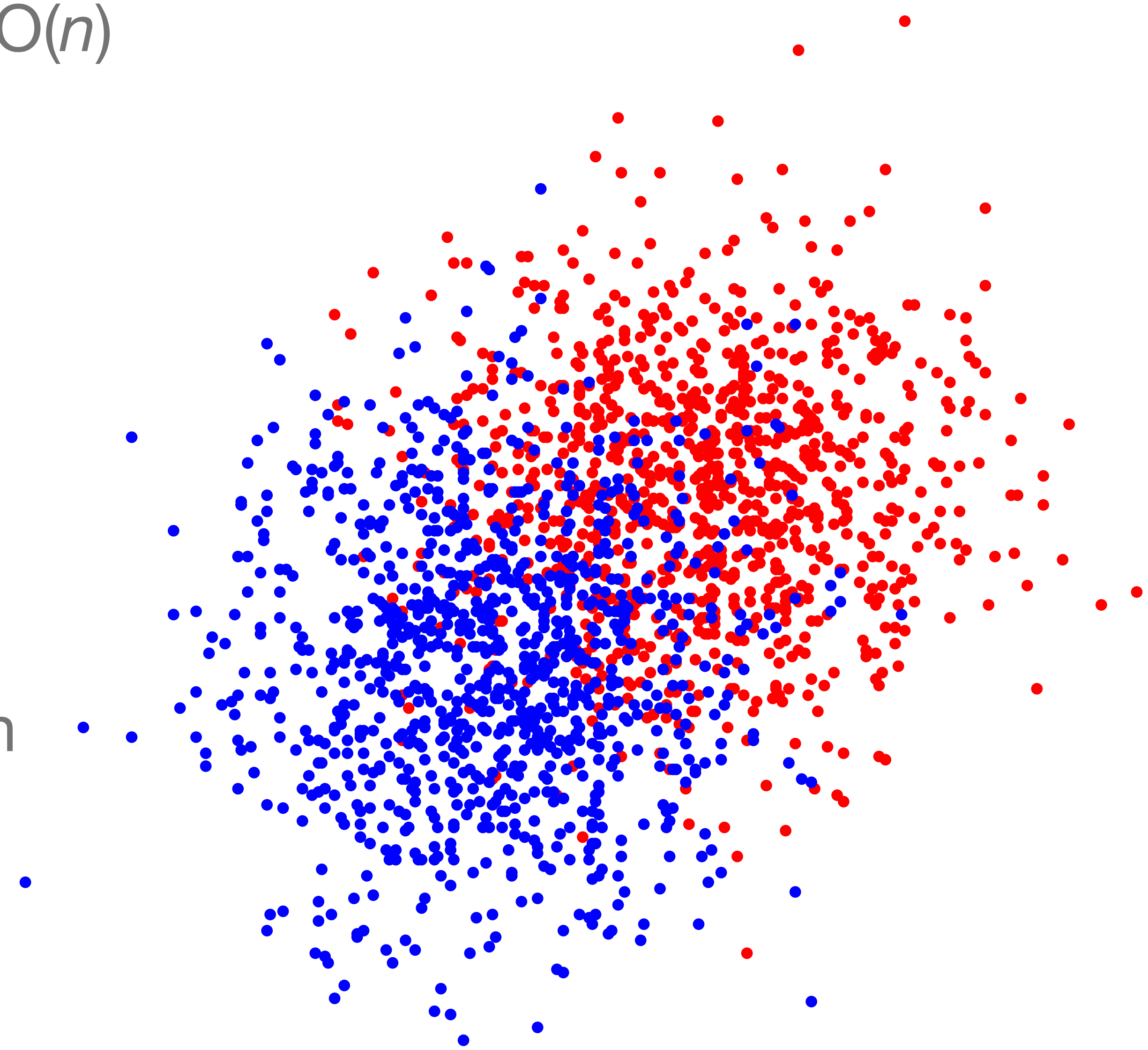
m points in n-dimensional space, where $m=O(n)$

k clusters with Gaussian noise

when can we...

- find the cluster centers?
- label the points better than chance?
- tell that there are clusters, i.e., distinguish from a null model with one big cluster?

phase transitions as a function of noise vs. cluster distances, and m/n



Clustering high-dimensional data

m points in n-dimensional space, where m=O(n)

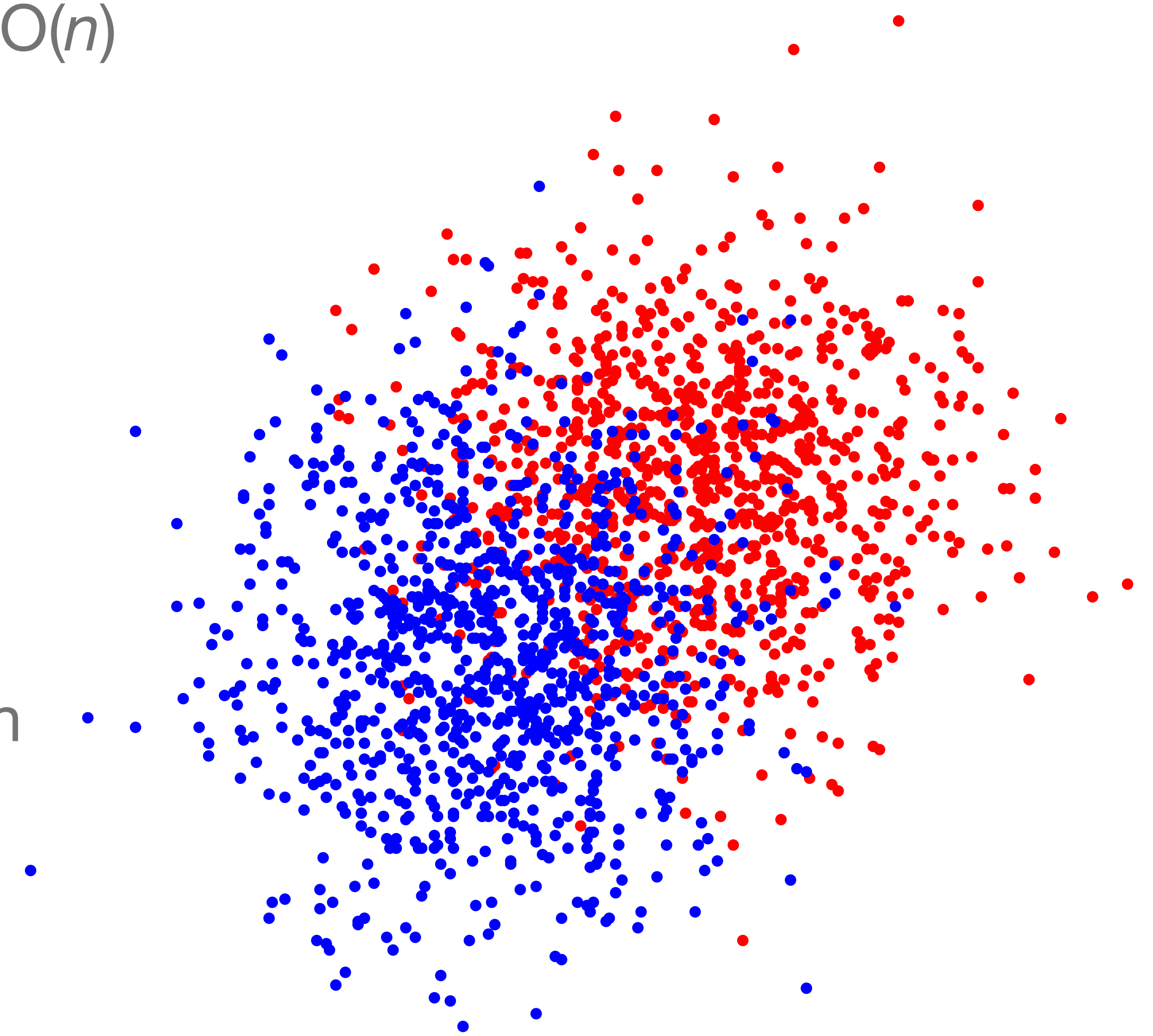
k clusters with Gaussian noise

when can we...

- find the cluster centers?
- label the points better than chance?
- tell that there are clusters, i.e., distinguish from a null model with one big cluster?

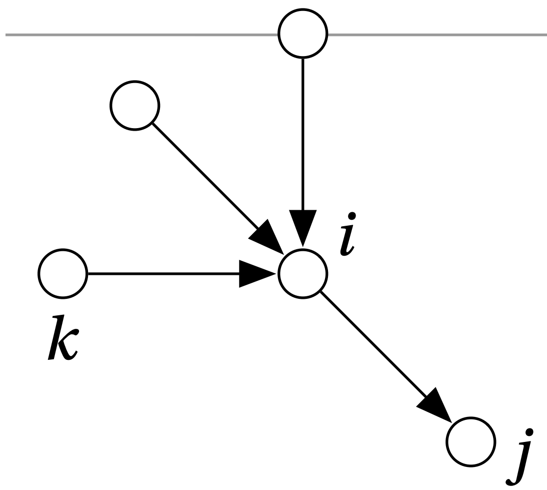
phase transitions as a function of noise vs. cluster distances, and m/n

when k is large enough, we can do better (information-theoretically) than PCA



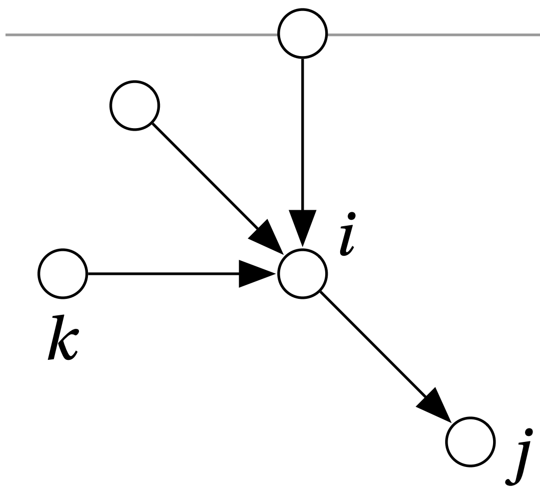
Techniques

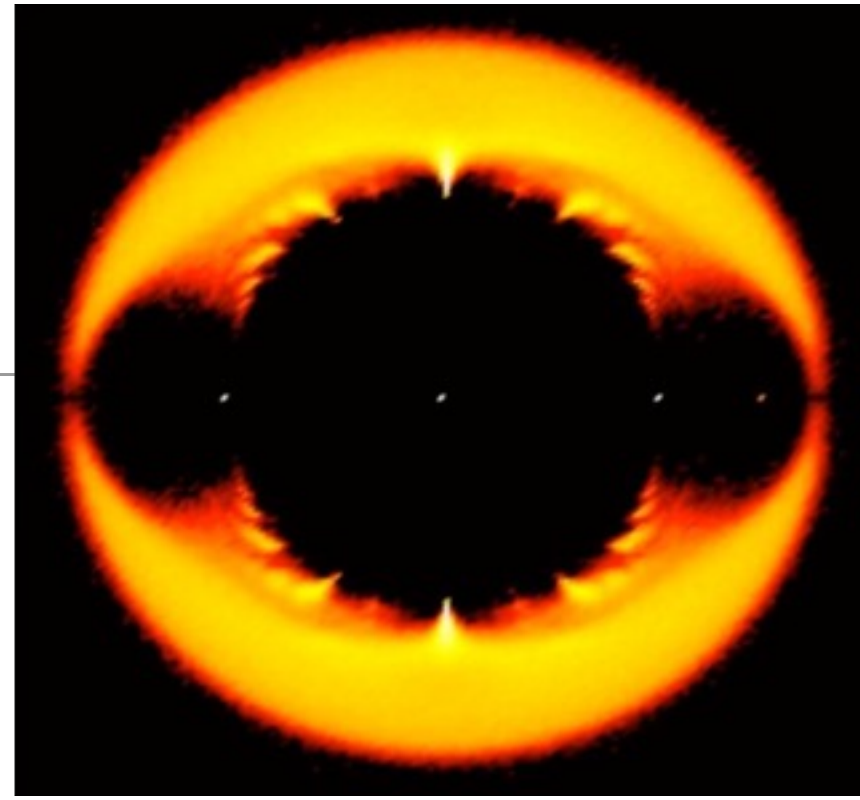
Techniques



If we iteratively estimate the probabilities with which nodes belong to groups, can we avoid a fixed point where each node is equally likely to be in each group? What can we learn about the ancestor of a family tree from its descendants?

Techniques

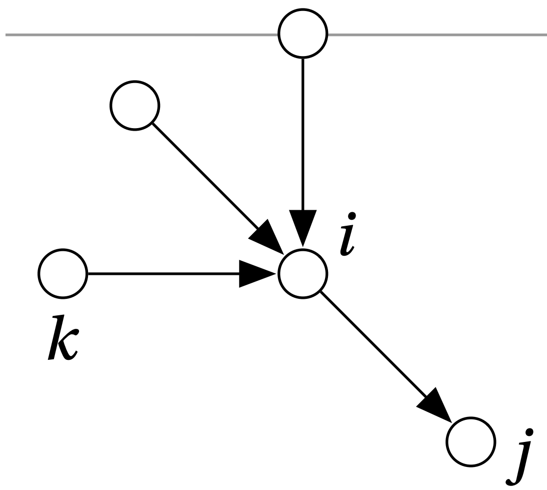


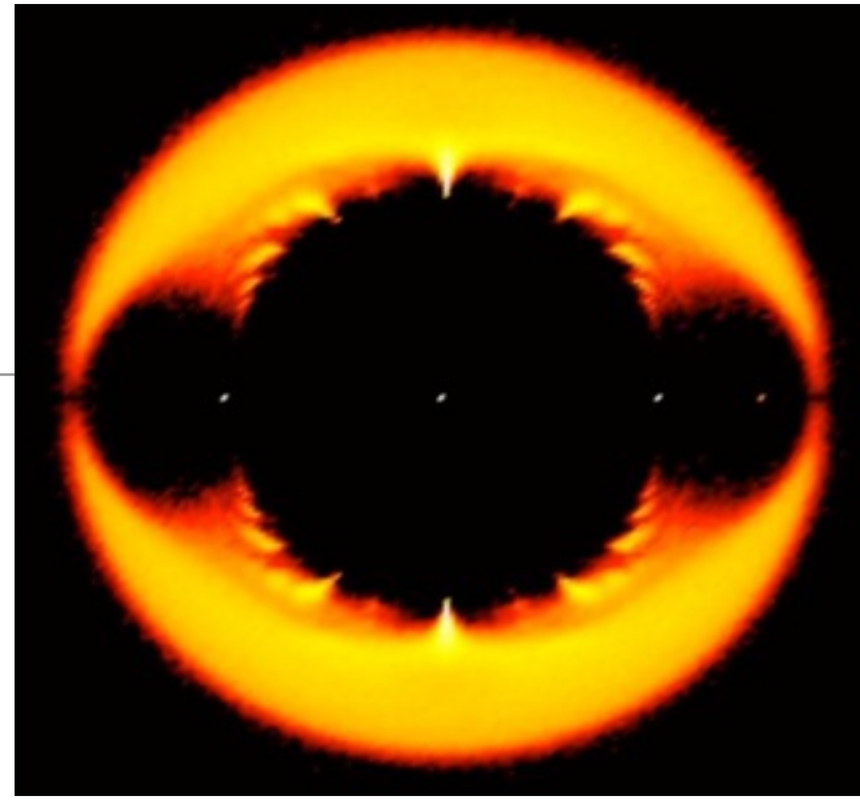


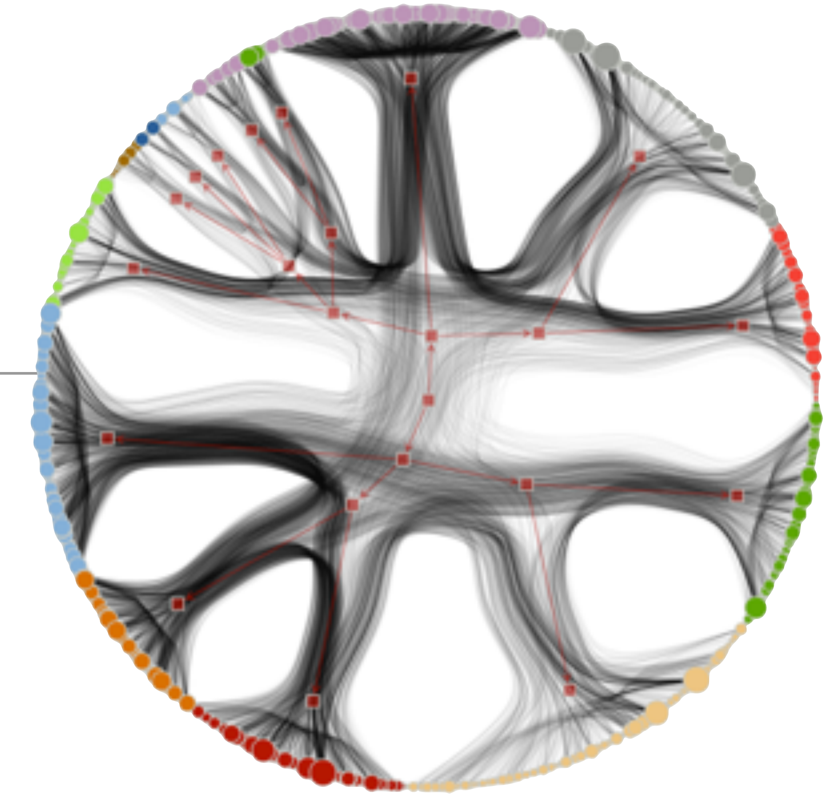
If we iteratively estimate the probabilities with which nodes belong to groups, can we avoid a fixed point where each node is equally likely to be in each group? What can we learn about the ancestor of a family tree from its descendants?

How does community structure affect random walks (or epidemics) on networks? When does it show up in the spectrum of the adjacency matrix? When is it dominated by the randomness in the graph?

Techniques





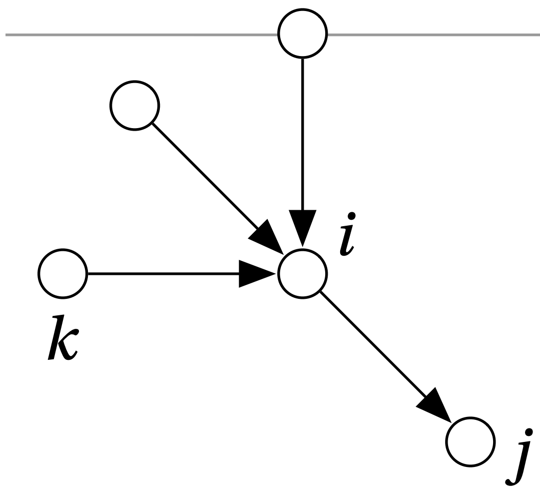


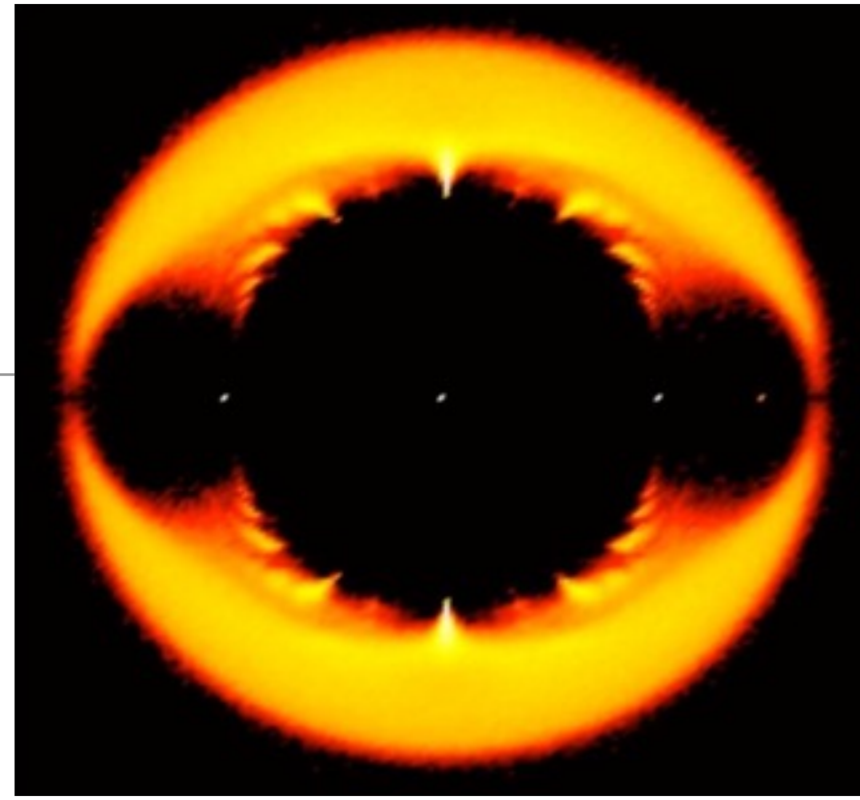
If we iteratively estimate the probabilities with which nodes belong to groups, can we avoid a fixed point where each node is equally likely to be in each group? What can we learn about the ancestor of a family tree from its descendants?

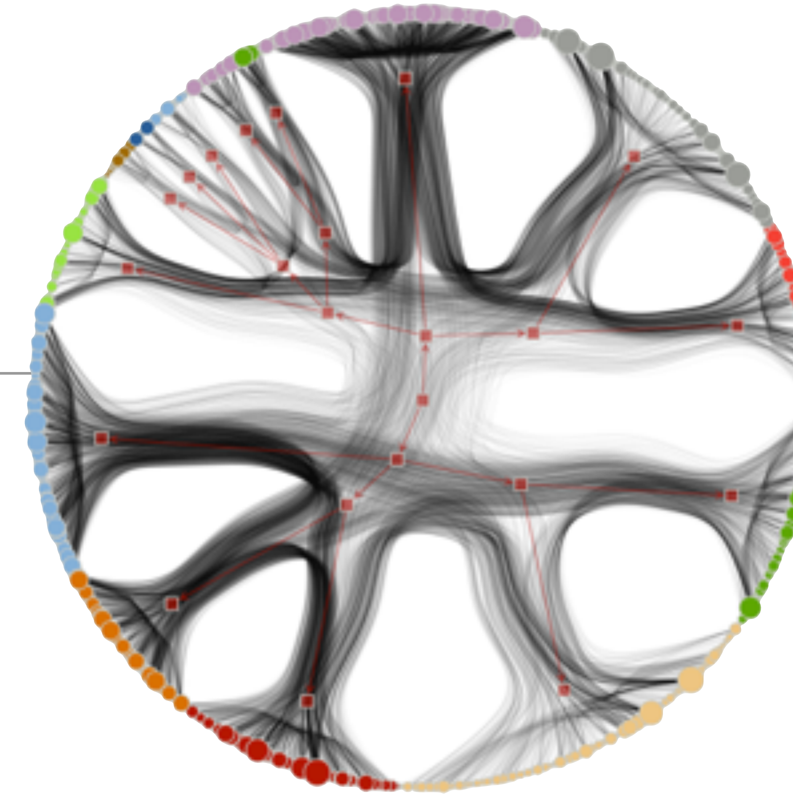
How does community structure affect random walks (or epidemics) on networks? When does it show up in the spectrum of the adjacency matrix? When is it dominated by the randomness in the graph?

How can we tell the difference between the block model and a null model with no community structure? Can we bound the likelihood ratio between them? How can we tell when an apparent community is real, instead of overfitting?

Techniques







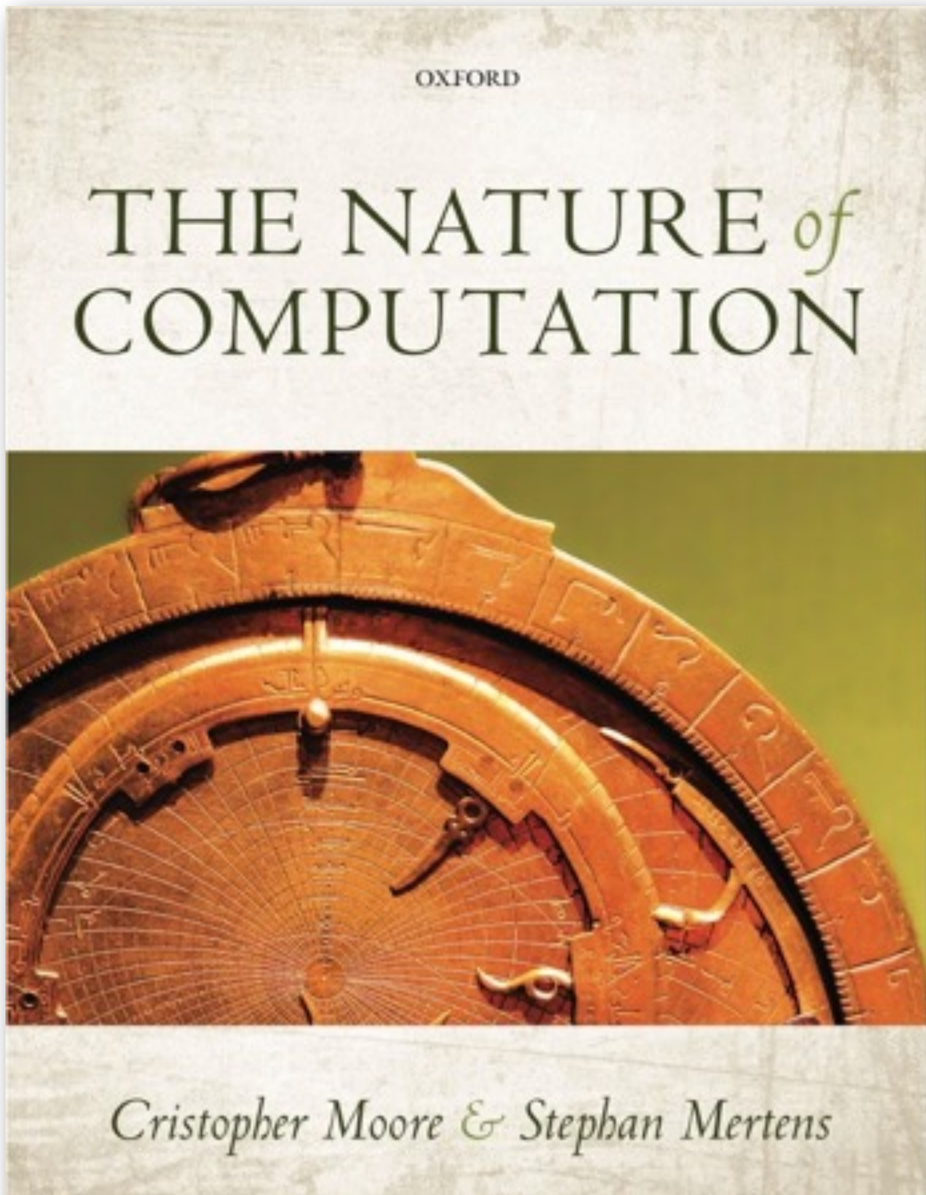
If we iteratively estimate the probabilities with which nodes belong to groups, can we avoid a fixed point where each node is equally likely to be in each group? What can we learn about the ancestor of a family tree from its descendants?

How does community structure affect random walks (or epidemics) on networks? When does it show up in the spectrum of the adjacency matrix? When is it dominated by the randomness in the graph?

How can we tell the difference between the block model and a null model with no community structure? Can we bound the likelihood ratio between them? How can we tell when an apparent community is real, instead of overfitting?

Next two lectures!

A little light reading



www.nature-of-computation.org

To put it bluntly: this book rocks! It somehow manages to combine the fun of a popular book with the intellectual heft of a textbook.

Scott Aaronson, MIT

This is, simply put, the best-written book on the theory of computation I have ever read; one of the best-written mathematical books I have ever read, period.

Cosma Shalizi, Carnegie Mellon