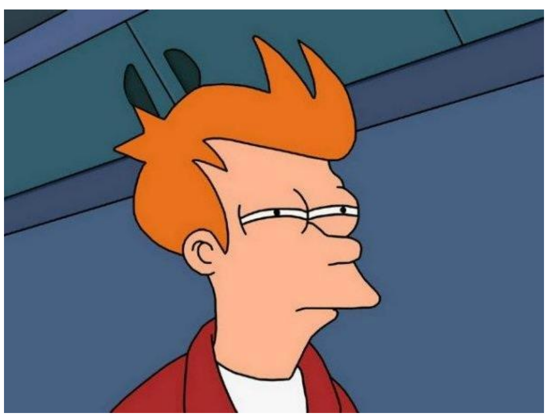
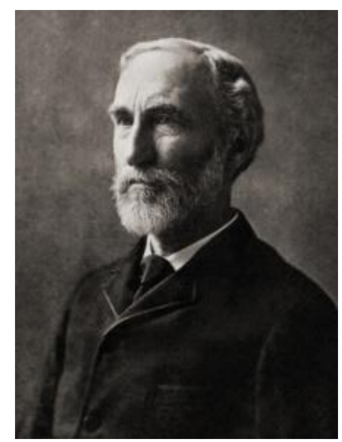
If you squint hard enough,

Gibbs distributions behave like mixtures of product measures





Renan Gross, WIS

Ronen Eldan, WIS

- Ok, so I need to tell you:
  - What Gibbs distributions are
  - What mixtures of product measures are
  - What it means to squint hard enough
  - What this is good for (maybe)

A probability measure X_n^f on $\{-1,1\}^n$ is a

Gibbs distribution with Hamiltonian f

if for $X \sim X_n^f$, we have

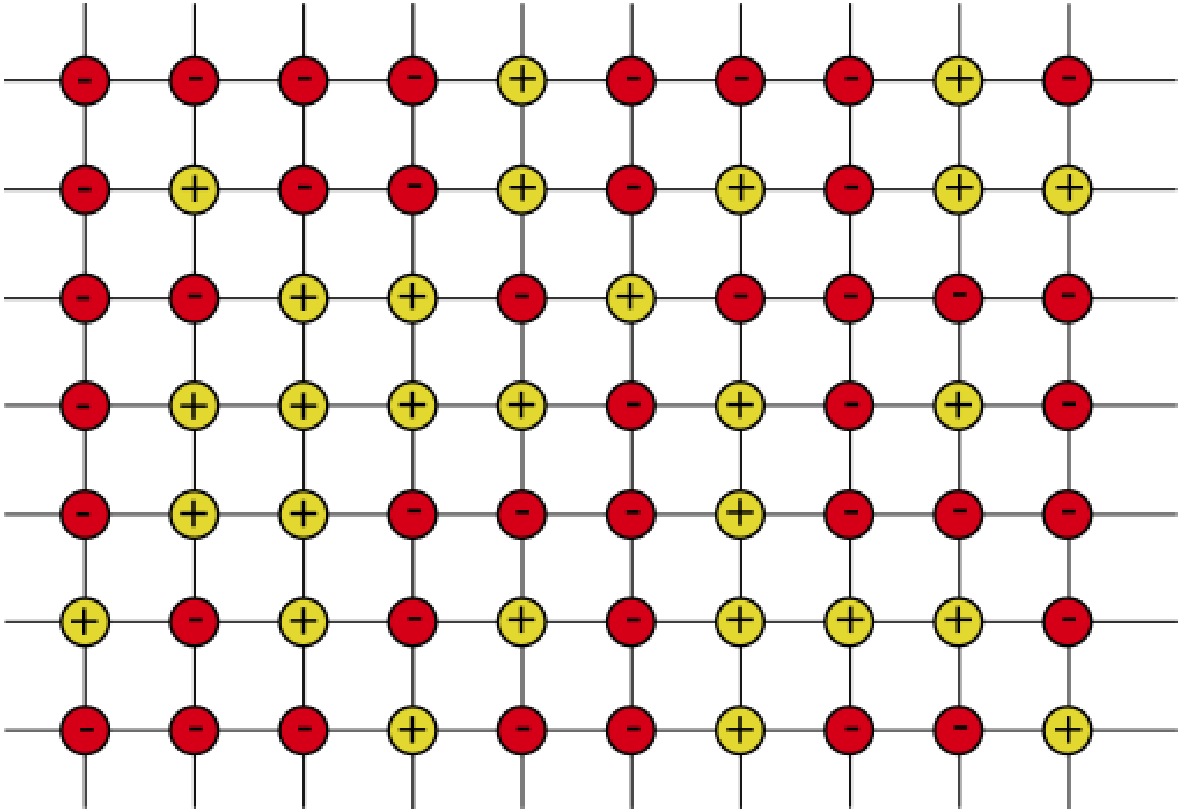
$$\boldsymbol{P[X = x] = e^{f(x)}/Z}$$

Where Z is a normalizing constant.

Notes:

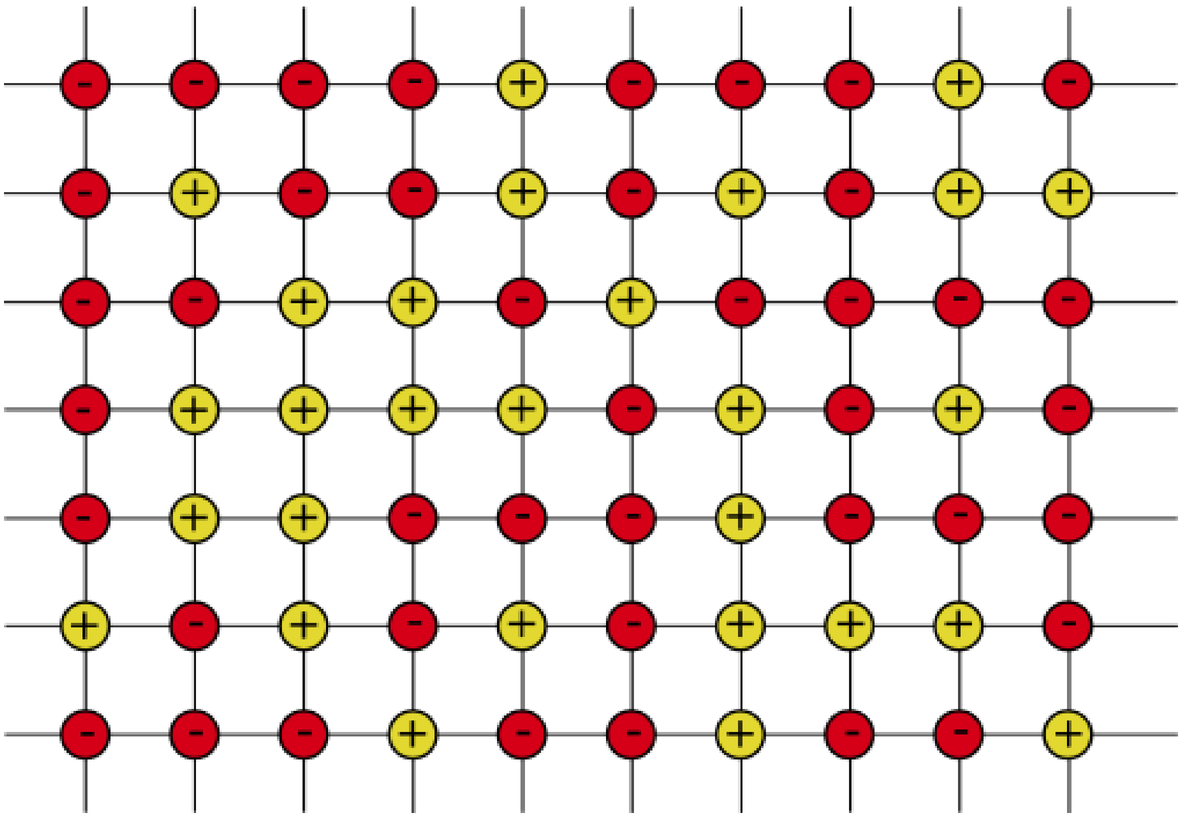
- $f: \{-1,1\}^n \to [-\infty, \infty]$
- Every distribution can be written this way…

Ising Model



$$f(x) = -\beta J \sum_{i \sim j} x_i x_j + \beta H \sum_i x_i$$

Ising Model



$$f(x) = \langle x, Ax \rangle + \langle \mu, x \rangle; \quad A \in \mathbb{R}^{n \times n}$$

$$\mu \in \mathbb{R}^n$$

Exponential random graphs

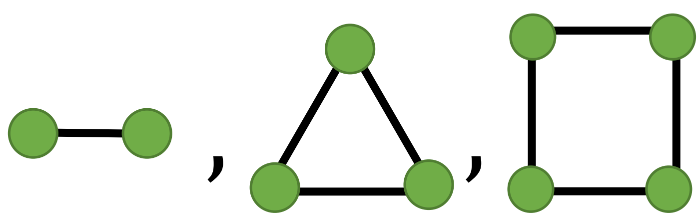
aka Graph Counting

Here, $n = \binom{N}{2}$ represents edges over graphs with N vertices

Each **edge** is an entry in the vector

$$f(G) = \sum_i \beta_i \cdot \#\{copies\ of\ H_i\ in\ G\}$$

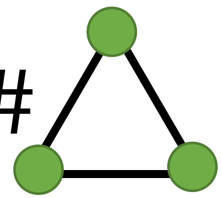
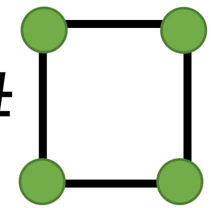
where

$H_i =$  , …

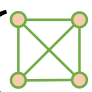
Exponential random graphs

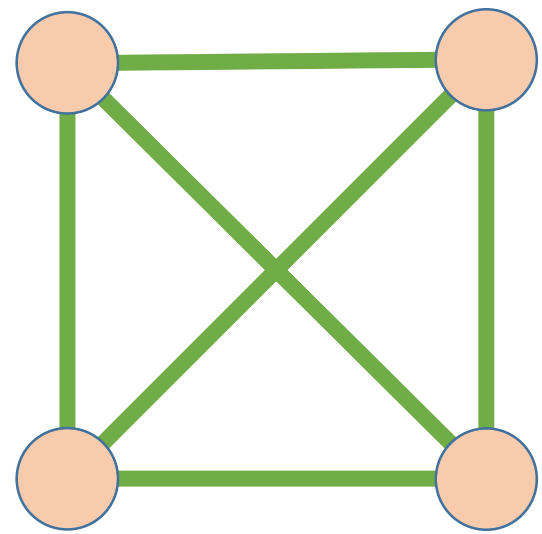
aka Graph Counting

For example, if

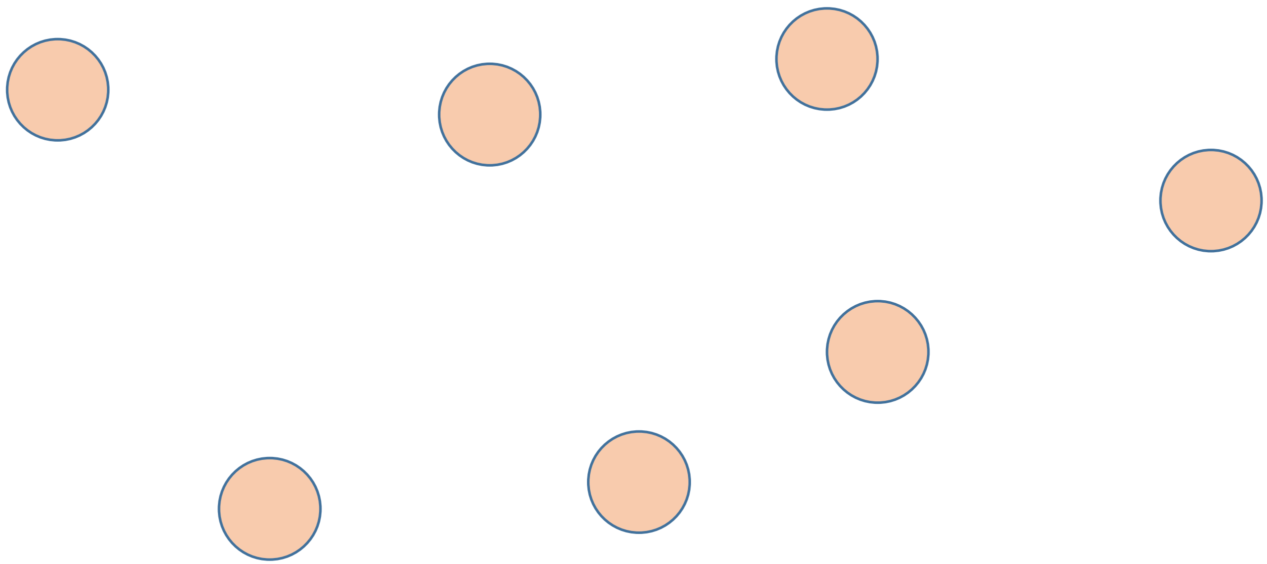
$f(G) = -\beta_1 \#$  $+ \beta_2 \#$ 

then

$f($$) = -4\beta_1 + 2\beta_2$



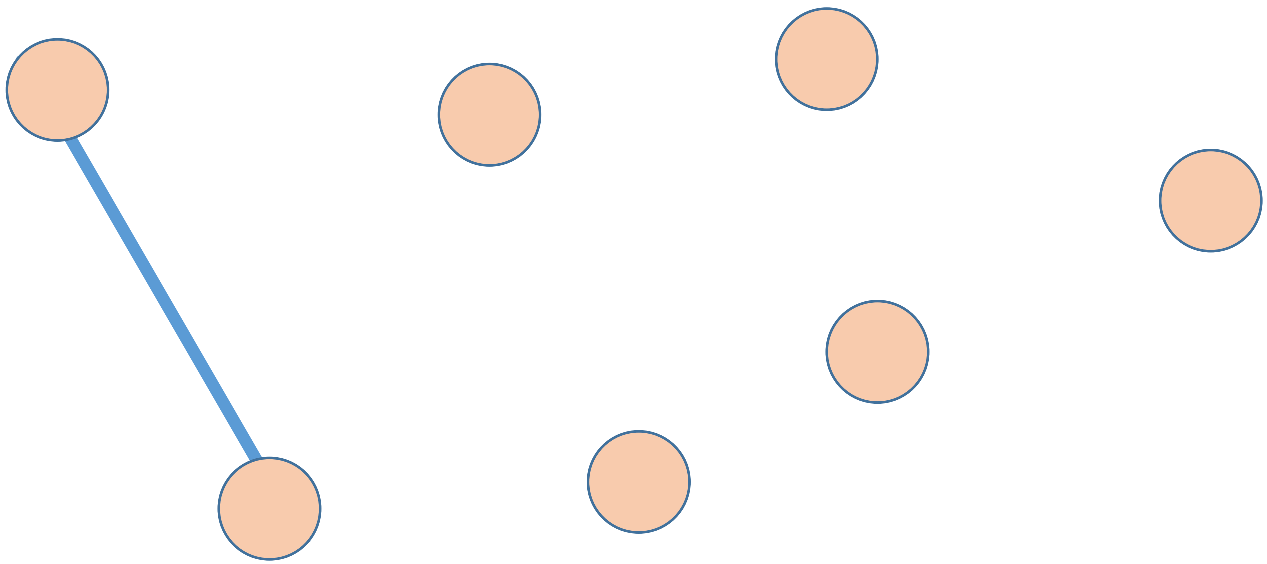
- Exponential random graphs are hard to work with.
- For example, when generating exponential random graphs with Glauber dynamics, many combinations of positive β's give an exponential mixing time.



- For other β's the mixing time is polynomial, but those correspond to uninteresting cases!

Result from "**Mixing time of exponential random graphs**" by Bhamidi, Bresler, Sly

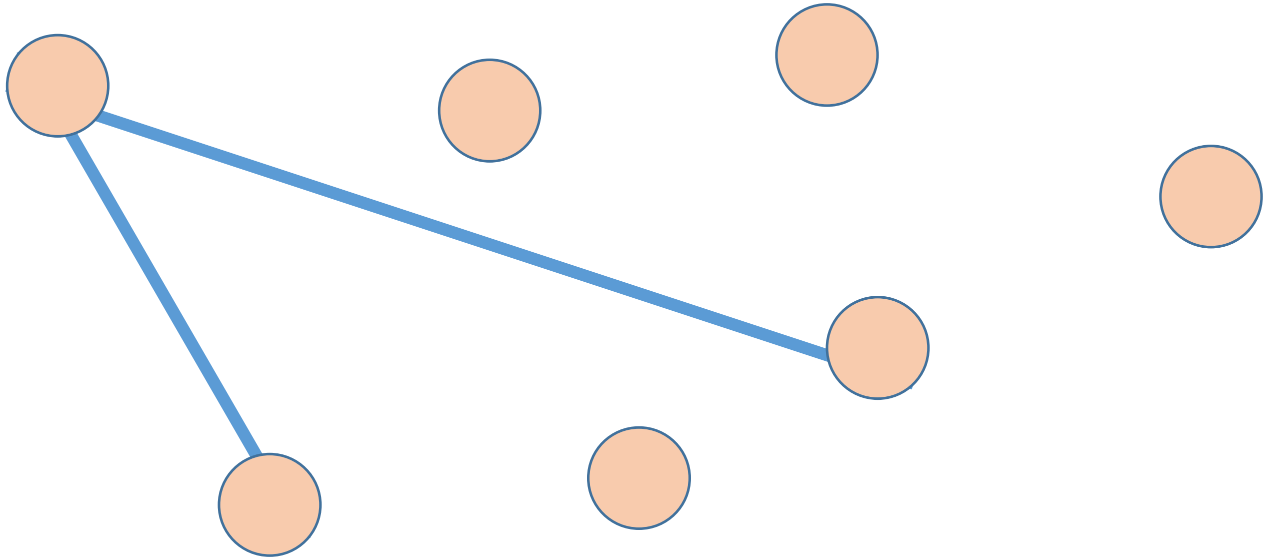
- These are hard to work with.
- For example, when generating exponential random graphs with Glauber dynamics, many combinations of positive β's give an exponential mixing time.



- For other β's the mixing time is polynomial, but those correspond to uninteresting cases!

Result from "**Mixing time of exponential random graphs**" by Bhamidi, Bresler, Sly

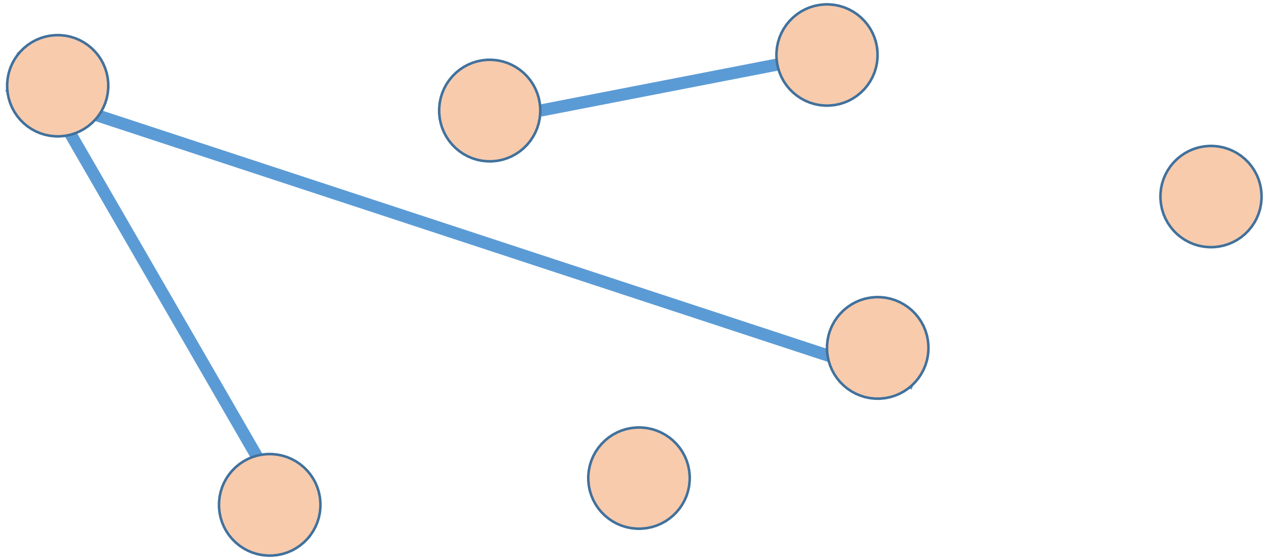
- These are hard to work with.
- For example, when generating exponential random graphs with Glauber dynamics, many combinations of positive β's give an exponential mixing time.



- For other β's the mixing time is polynomial, but those correspond to uninteresting cases!

Result from "**Mixing time of exponential random graphs**" by Bhamidi, Bresler, Sly

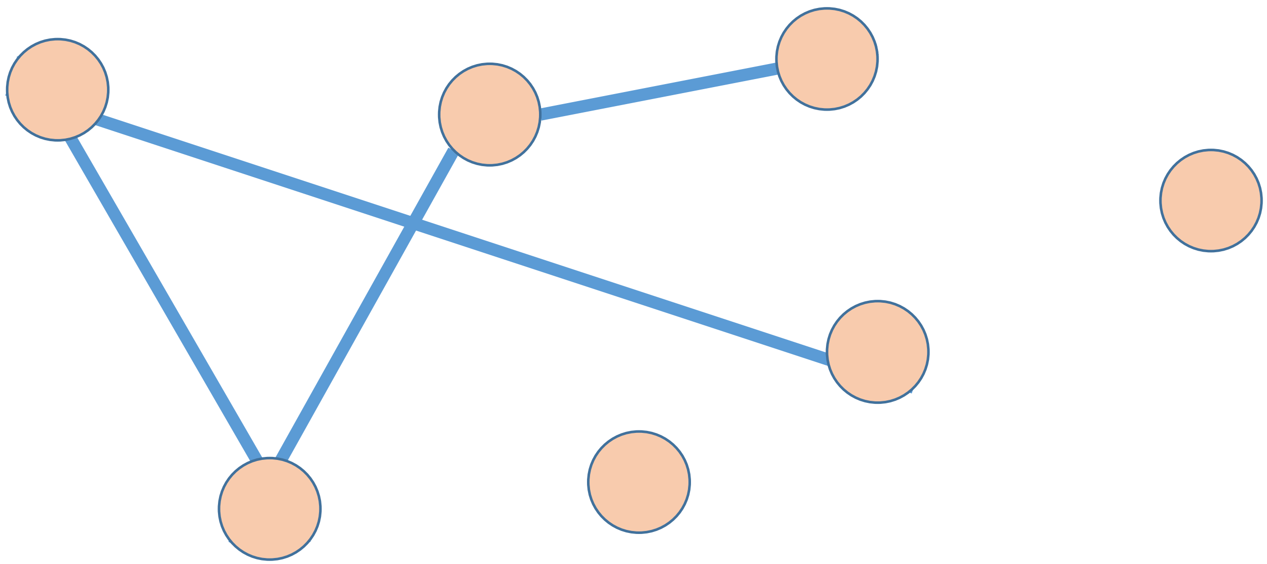
- These are hard to work with.
- For example, when generating exponential random graphs with Glauber dynamics, many combinations of positive β's give an exponential mixing time.



- For other β's the mixing time is polynomial, but those correspond to uninteresting cases!

Result from "**Mixing time of exponential random graphs**" by Bhamidi, Bresler, Sly

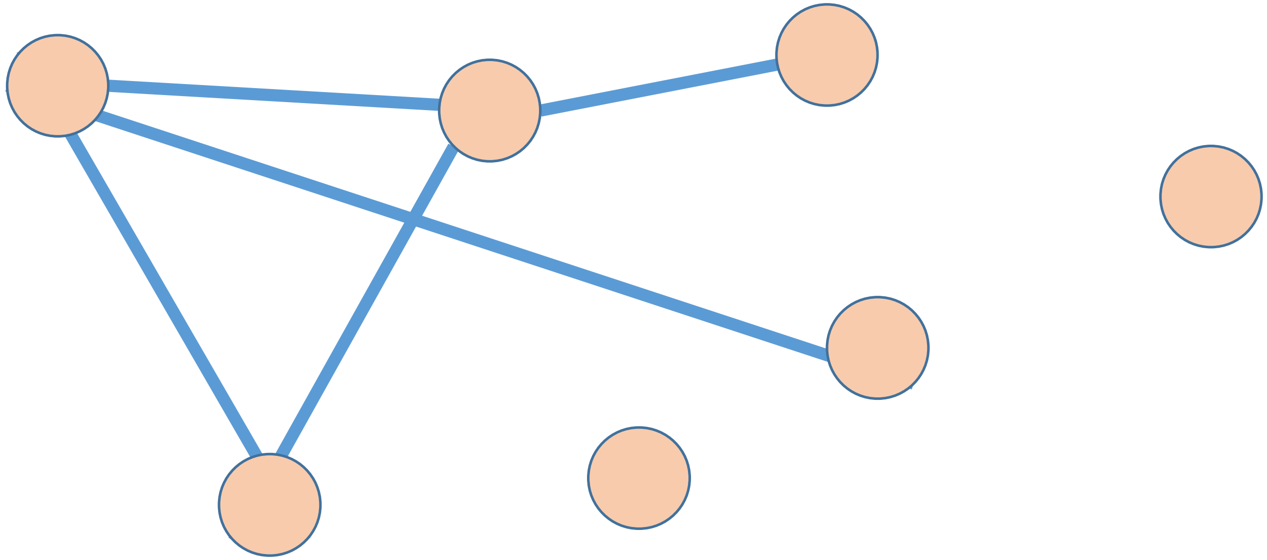
- These are hard to work with.
- For example, when generating exponential random graphs with Glauber dynamics, many combinations of positive β's give an exponential mixing time.



- For other β's the mixing time is polynomial, but those correspond to uninteresting cases!

Result from "**Mixing time of exponential random graphs**" by Bhamidi, Bresler, Sly

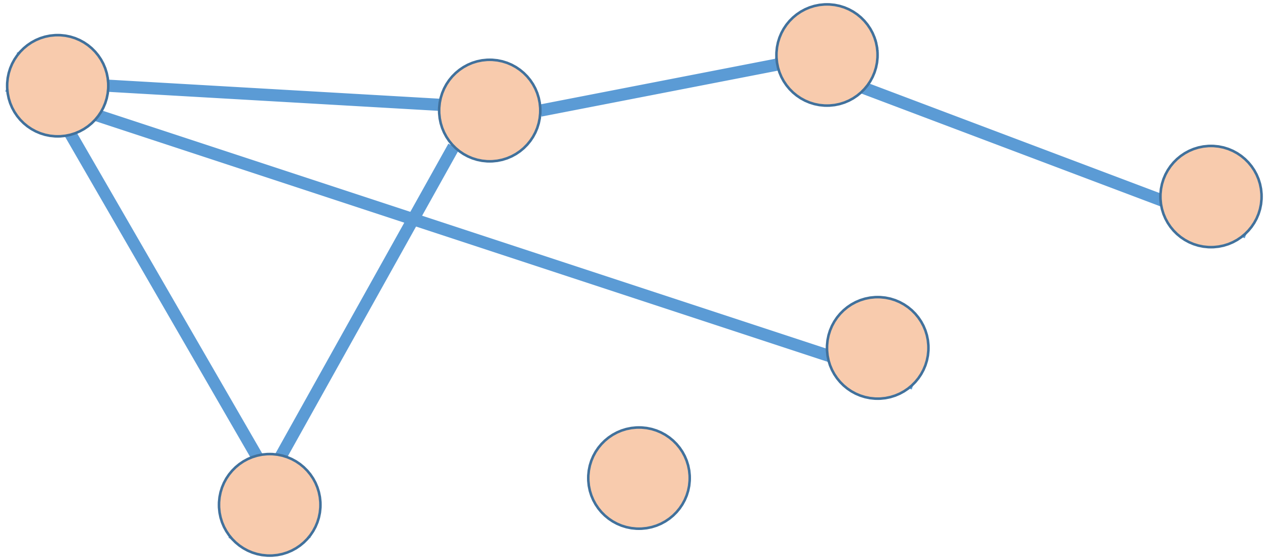
- These are hard to work with.
- For example, when generating exponential random graphs with Glauber dynamics, many combinations of positive β's give an exponential mixing time.



- For other β's the mixing time is polynomial, but those correspond to uninteresting cases!

Result from "**Mixing time of exponential random graphs**" by Bhamidi, Bresler, Sly

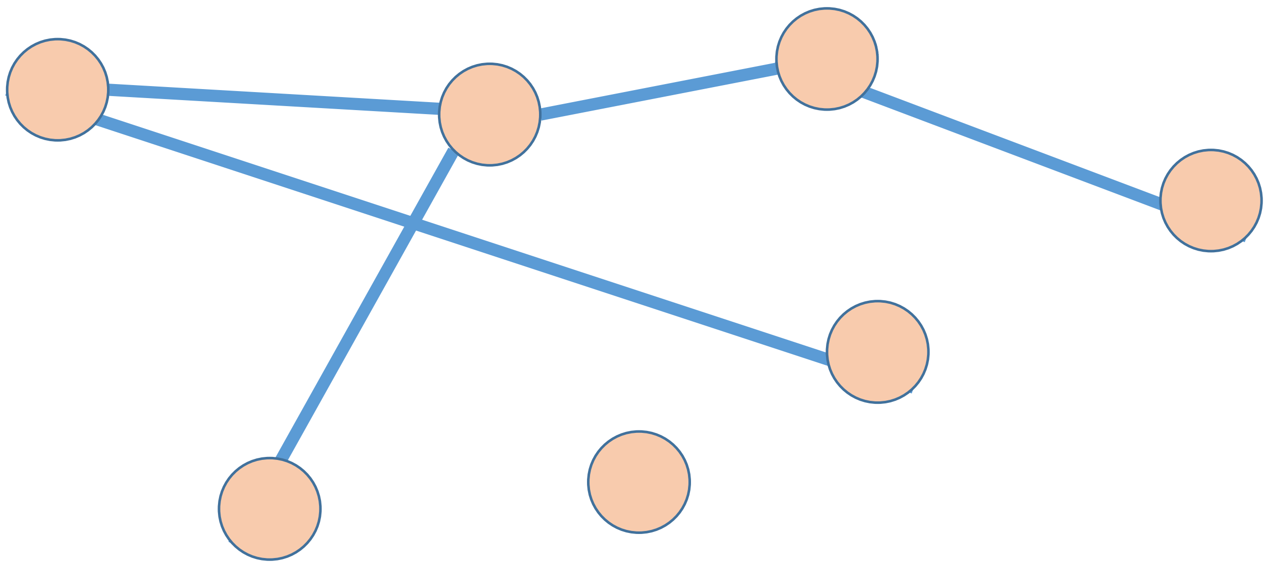
- These are hard to work with.
- For example, when generating exponential random graphs with Glauber dynamics, many combinations of positive β's give an exponential mixing time.



- For other β's the mixing time is polynomial, but those correspond to uninteresting cases!

Result from "**Mixing time of exponential random graphs**" by Bhamidi, Bresler, Sly

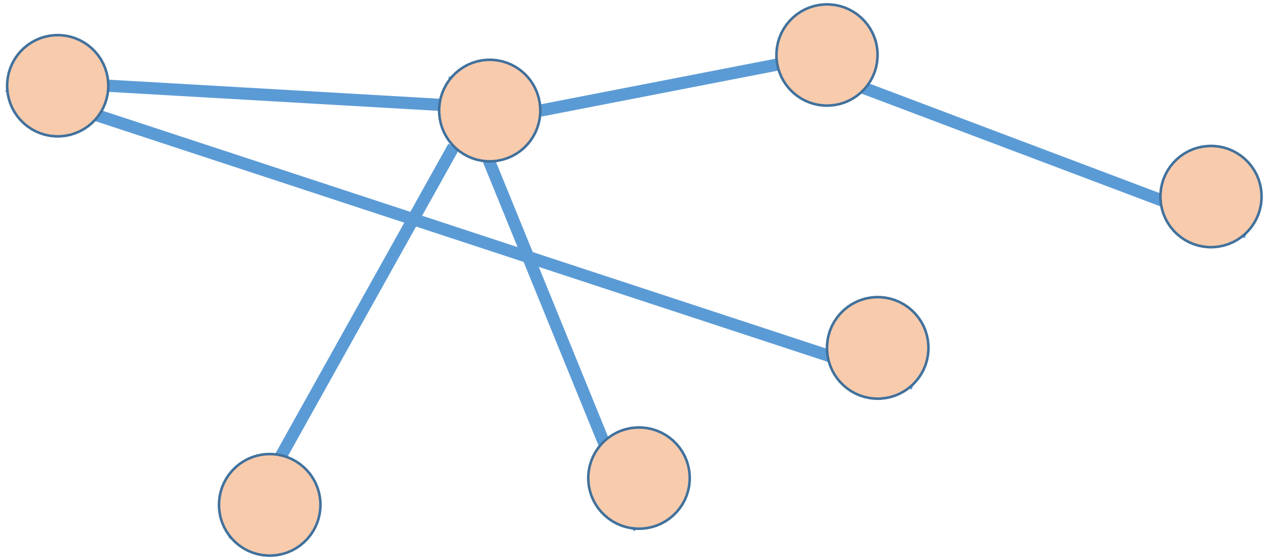
- These are hard to work with.
- For example, when generating exponential random graphs with Glauber dynamics, many combinations of positive β's give an exponential mixing time.



- For other β's the mixing time is polynomial, but those correspond to uninteresting cases!

Result from "**Mixing time of exponential random graphs**" by Bhamidi, Bresler, Sly

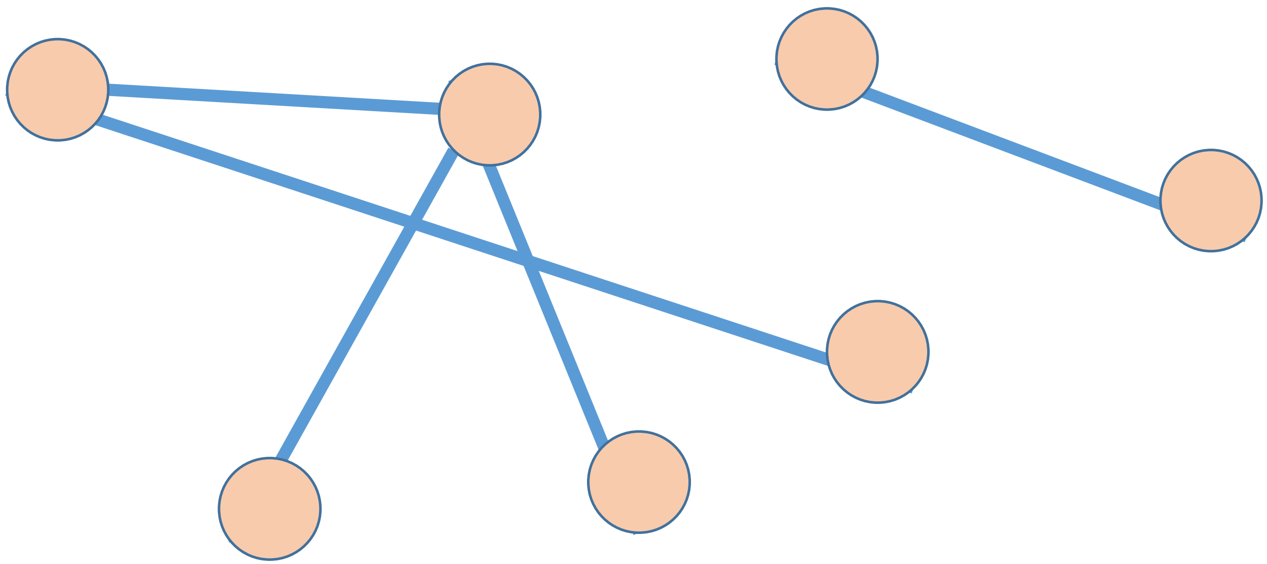
- These are hard to work with.
- For example, when generating exponential random graphs with Glauber dynamics, many combinations of positive β's give an exponential mixing time.



- For other β's the mixing time is polynomial, but those correspond to uninteresting cases!

Result from “**Mixing time of exponential random graphs**” by Bhamidi, Bresler, Sly

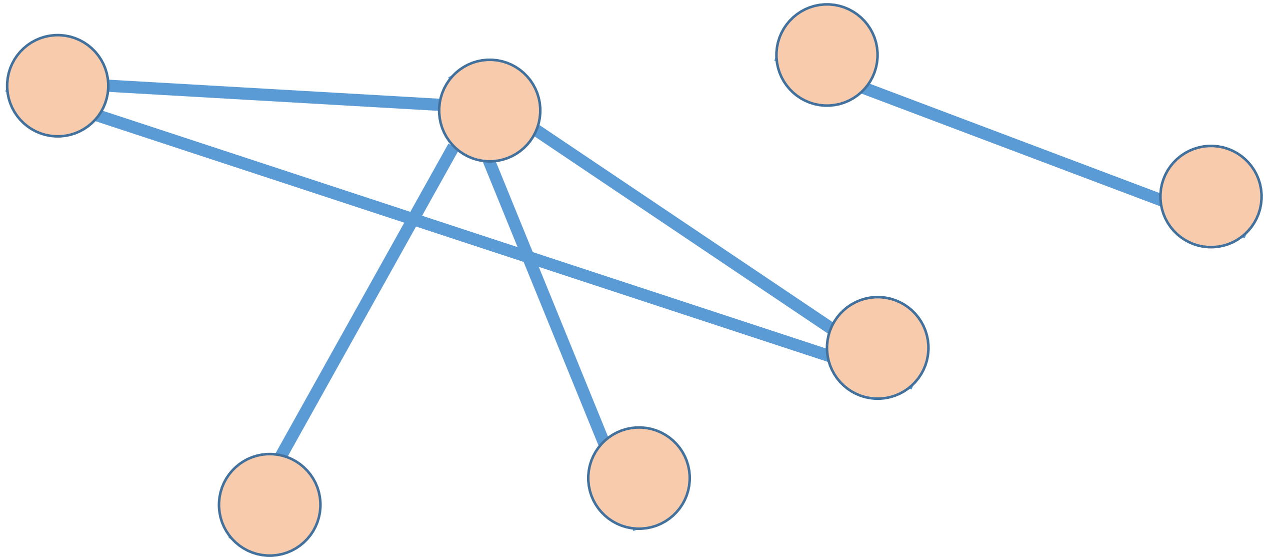
- These are hard to work with.
- For example, when generating exponential random graphs with Glauber dynamics, many combinations of positive β's give an exponential mixing time.



- For other β's the mixing time is polynomial, but those correspond to uninteresting cases!

Result from "**Mixing time of exponential random graphs**" by Bhamidi, Bresler, Sly

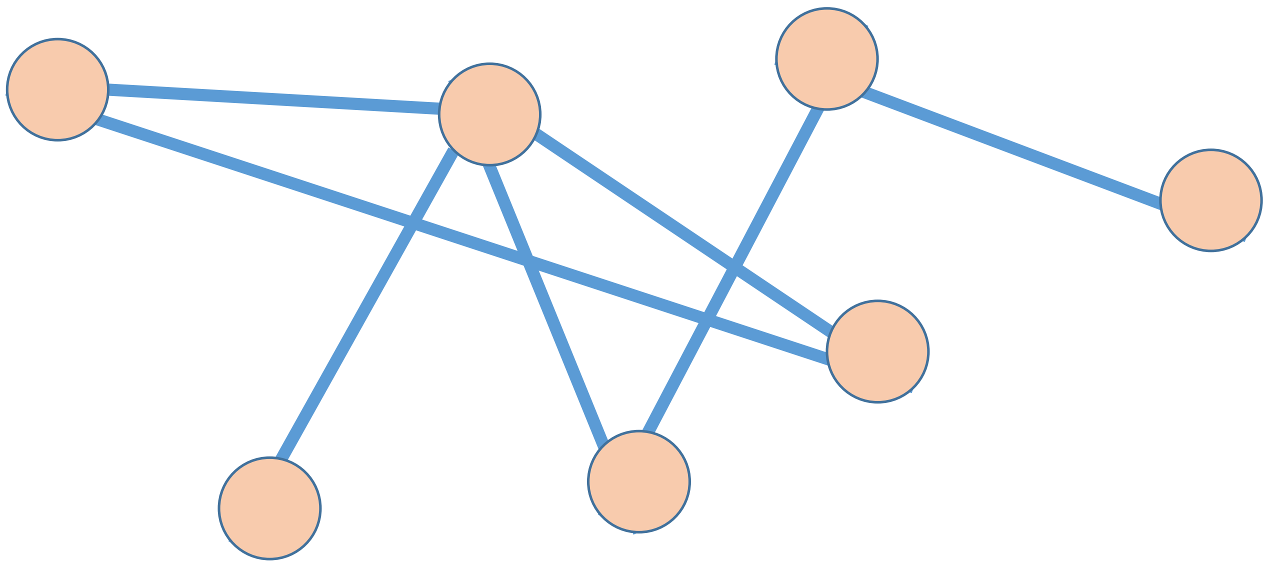
- These are hard to work with.
- For example, when generating exponential random graphs with Glauber dynamics, many combinations of positive β's give an exponential mixing time.



- For other β's the mixing time is polynomial, but those correspond to uninteresting cases!

Result from "**Mixing time of exponential random graphs**" by Bhamidi, Bresler, Sly

- These are hard to work with.
- For example, when generating exponential random graphs with Glauber dynamics, many combinations of positive β's give an exponential mixing time.



- For other β's the mixing time is polynomial, but those correspond to uninteresting cases!

Result from "**Mixing time of exponential random graphs**" by Bhamidi, Bresler, Sly

A product measure is a measure where every entry is independent.

In a mixture of product measures, first we choose probabilities, then we choose from that probability.

Let ρ be a measure on $[-1,1]^n$. A random variable $X(\rho)$ is a ρ-mixture if

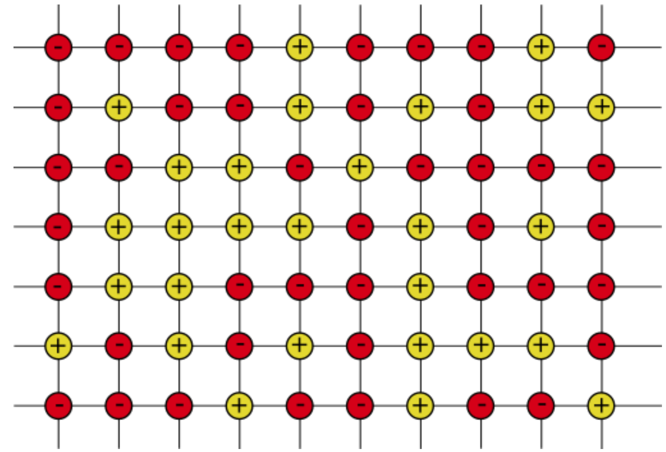
- $\Pr[X(\rho) = x] = \int \Pr[X(z) = x] d\rho(z)$

  Where $X(z)$ is the random vector with independent entries whose expectation is Z.

(Every distribution can be written as a mixture…)

For example, for Ising models:

$$f(x) = -\beta J \sum_{i \sim j} x_i x_j$$



We expect that:

- For small β, spins will be iid Bernoulli with probability ½
- For large β, spins tend to point either up or down.

  - We could expect that w.p ½, spins are iid Bernoulli w.p p, and w.p ½ , spins are iid Bernoulli w.p (1-p).

Theorem: $\boldsymbol{P[X=x]=e^{f(x)}/Z}$

If the Hamiltonian f is "nice enough", then

- The Gibbs distribution can be approximated by a ρ-mixture
- Further, ρ is concentrated on vectors $X \in [-1,1]^n$ which satisfy

$$\left\|X-\tanh\left(\nabla f(X)\right)\right\|_1 = o(n)$$

- Here, $\partial_i f(y) = \frac{f(y^{i+})-f(y^{i-})}{2}$, i.e energy difference in flipping one coordinate for discrete vectors.
- Harmonic extension for continuous vectors.

- What is a nice Hamiltonian?

- “A nice Hamiltonian is a smooth Hamiltonian”

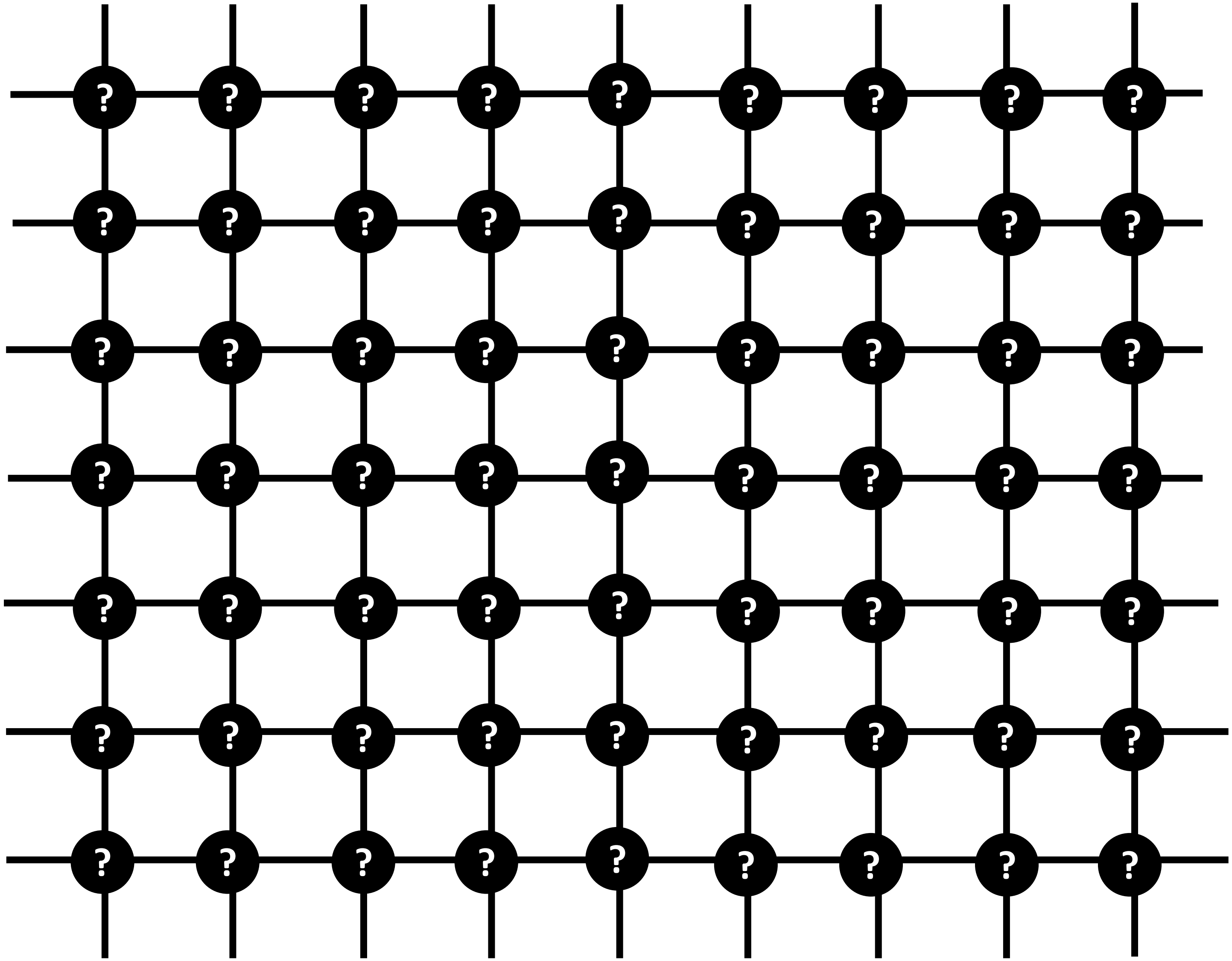
- What is a nice Hamiltonian?

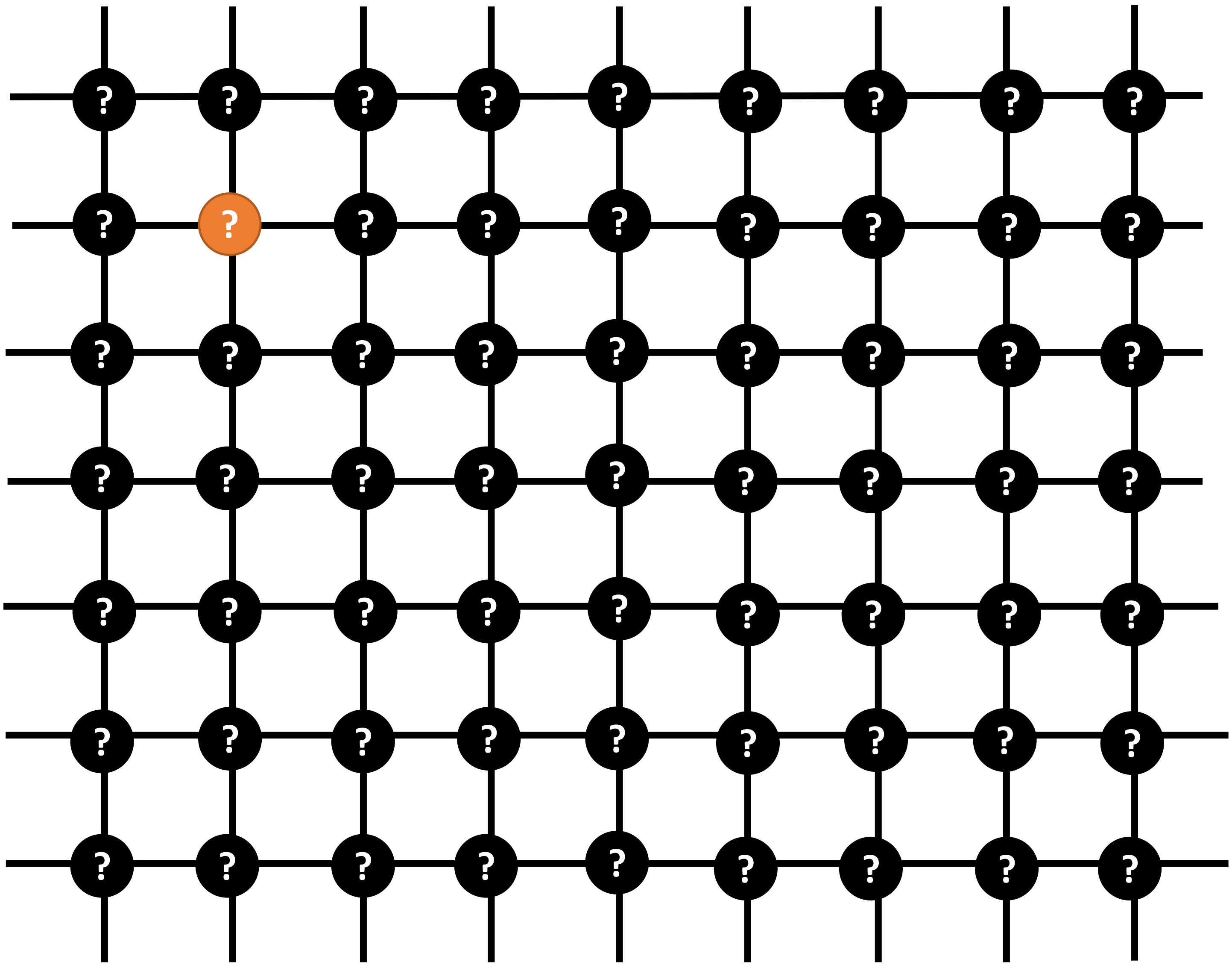
- "A nice Hamiltonian is a smooth Hamiltonian"
- Borrowing from physics, you might call this a "mean-field" Hamiltonian
- A sufficient condition: the Gaussian width of the gradient of the Hamiltonian is low.

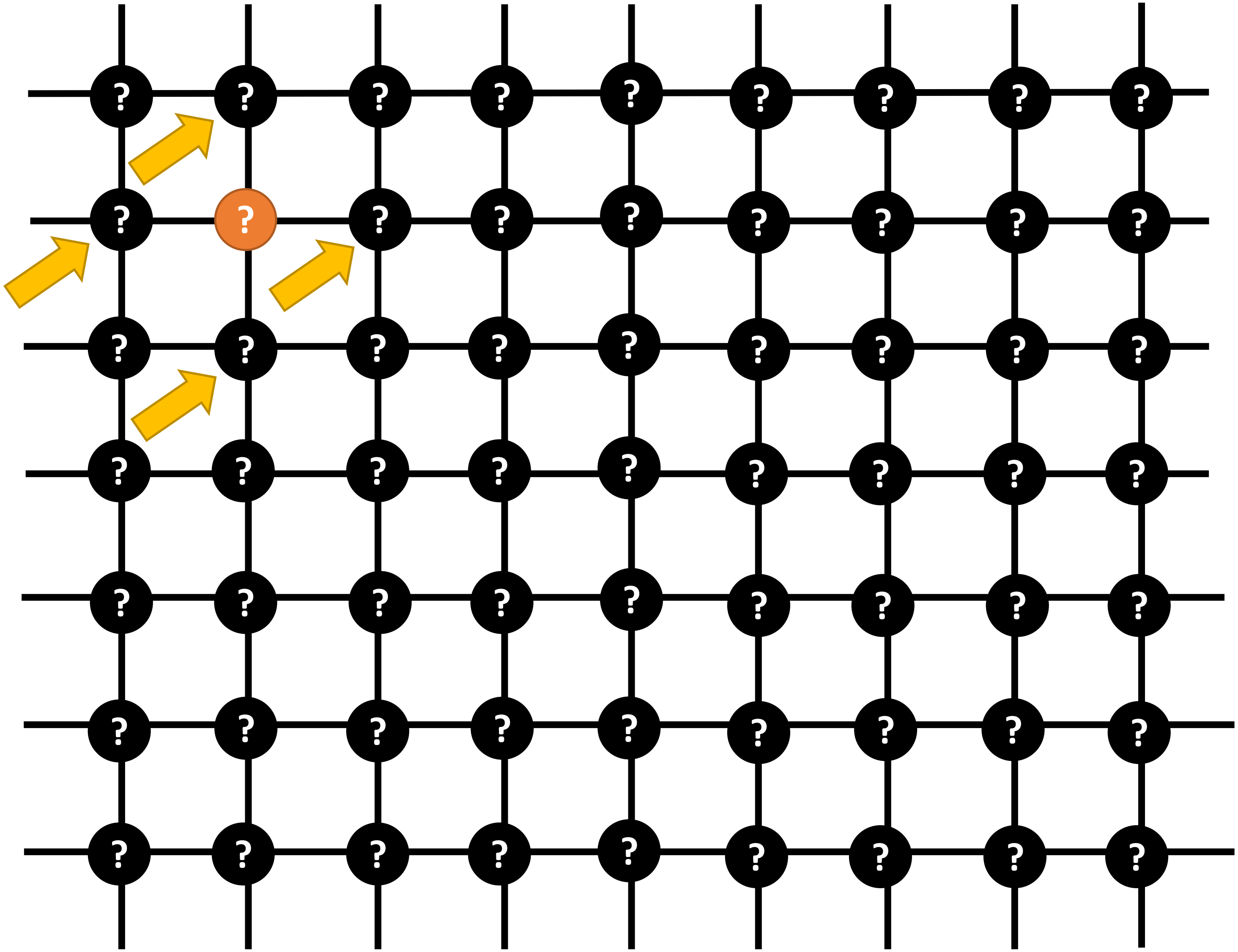
$$GW(K) = \mathbb{E}\left[\sup_{x \in K} \langle x, \Gamma \rangle\right]$$

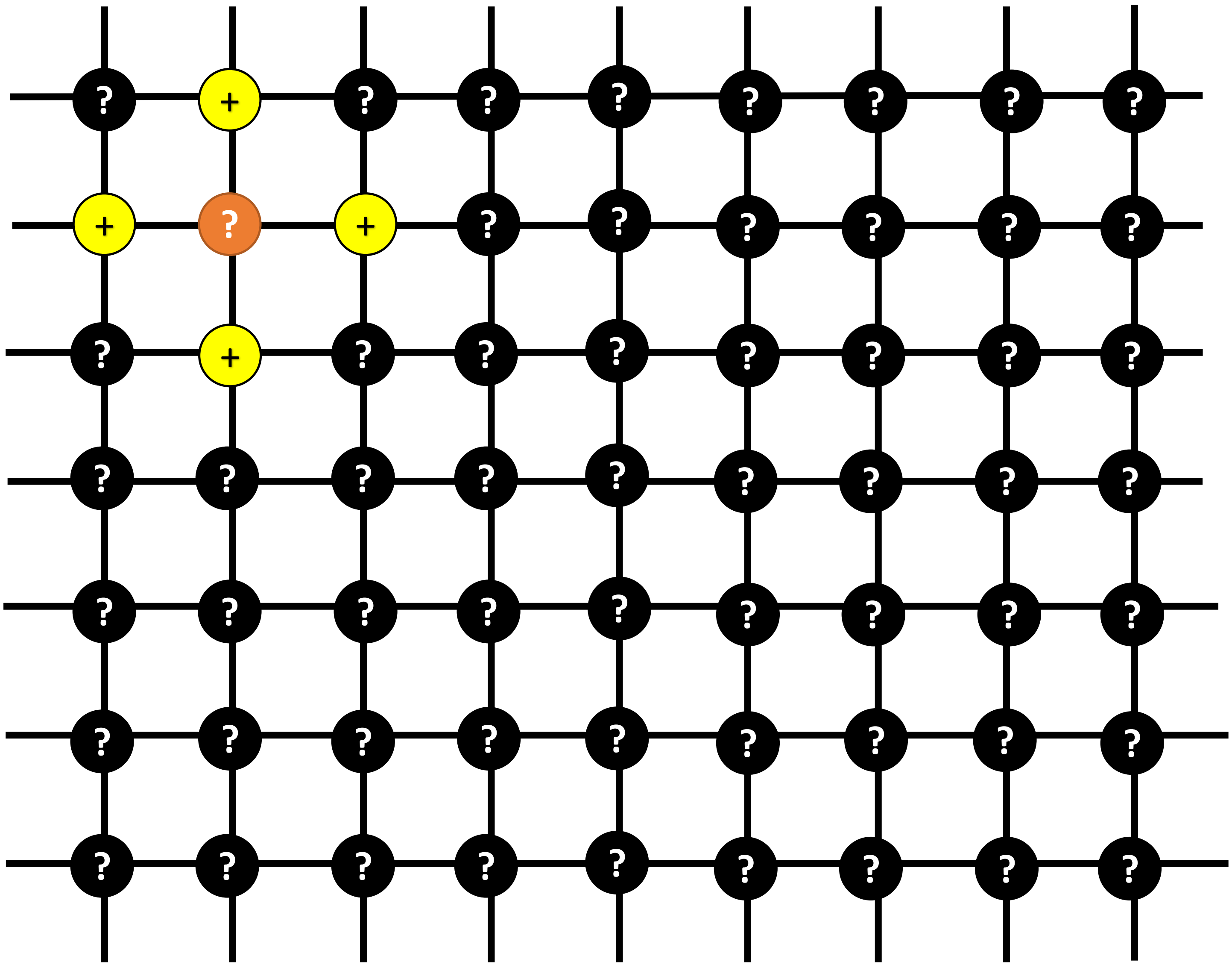
where $\Gamma \sim \mathrm{N}(0, \mathrm{Id})$ is a standard Gaussian in $\mathbb{R}^n$.

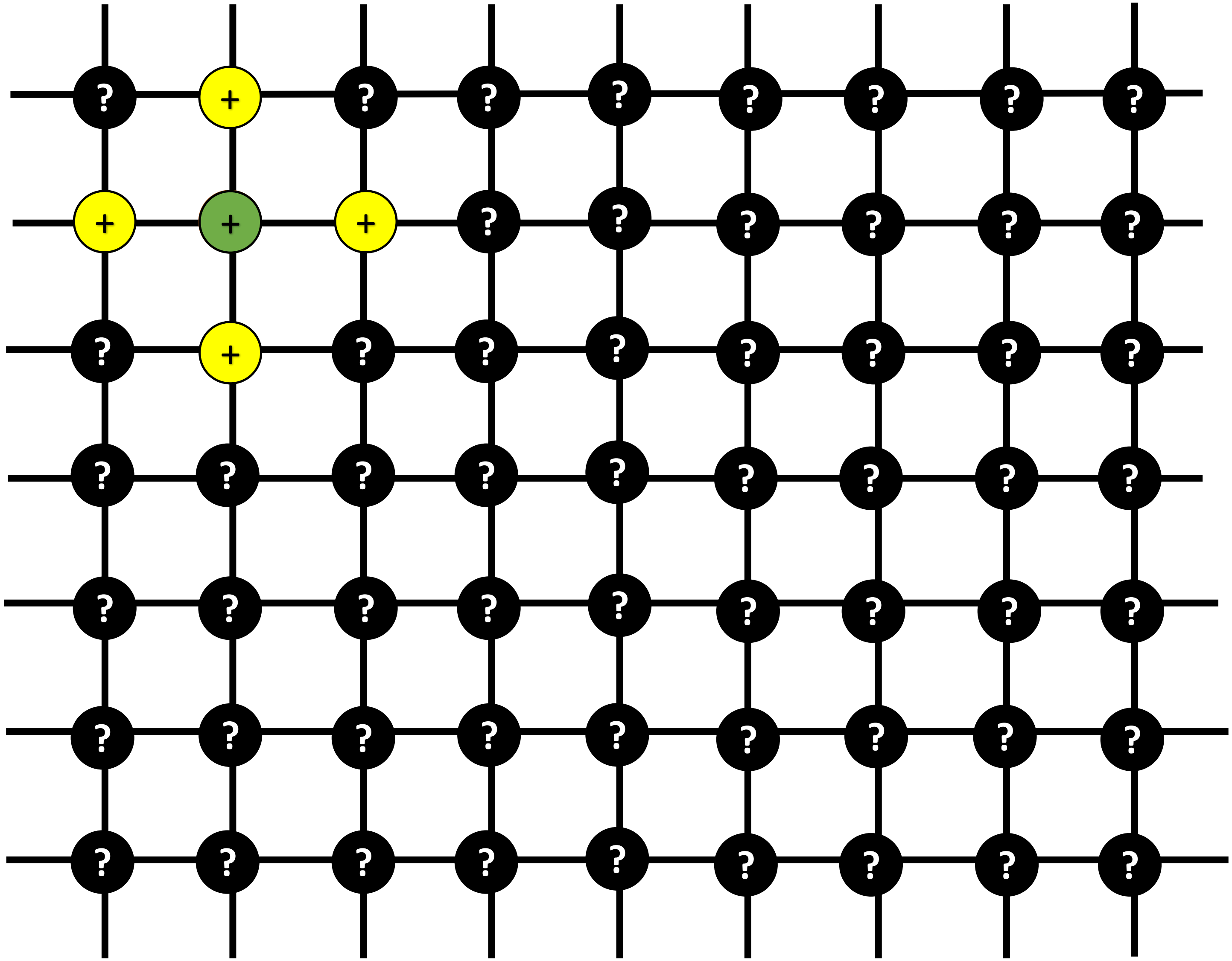
- Correlation with random noise

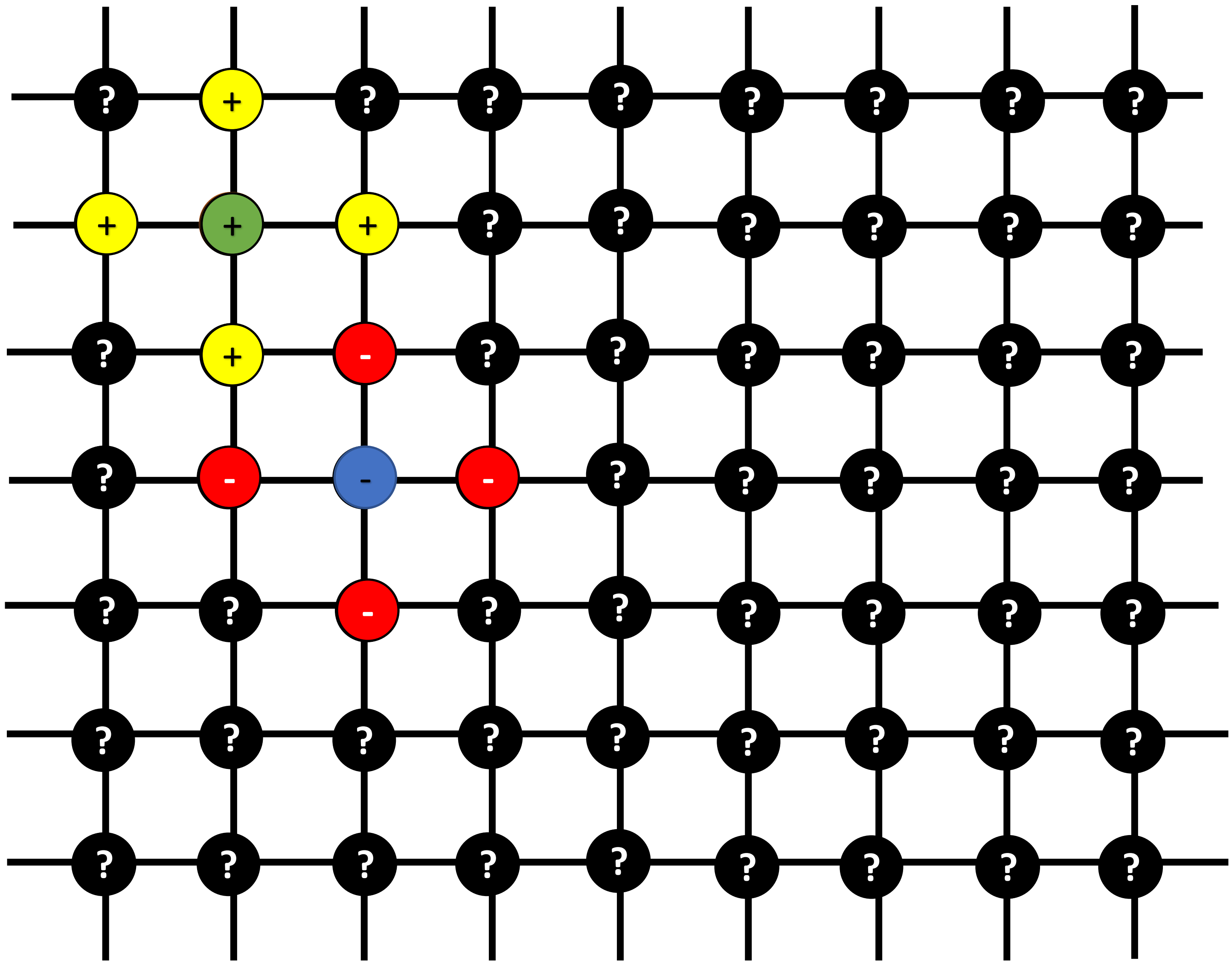


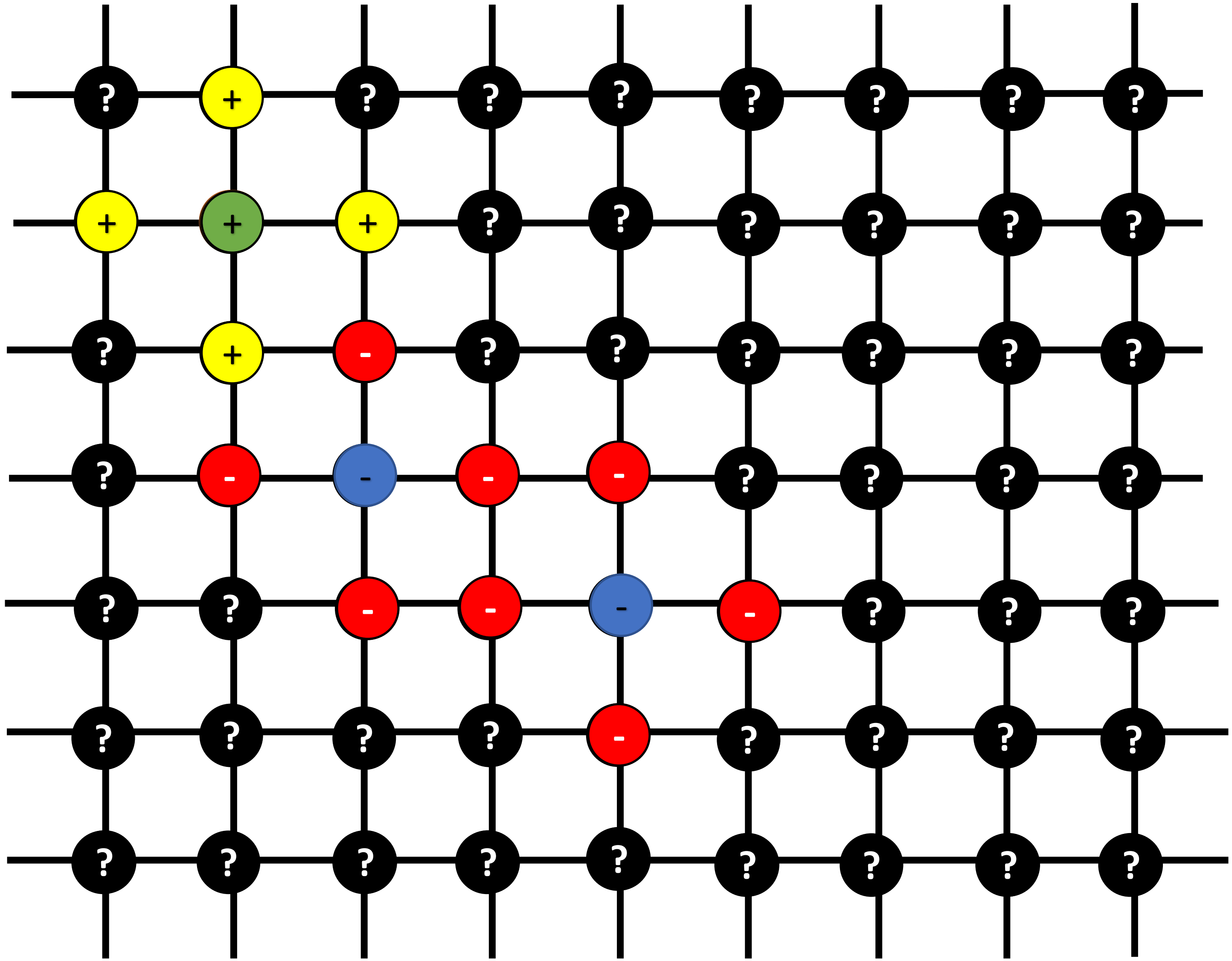


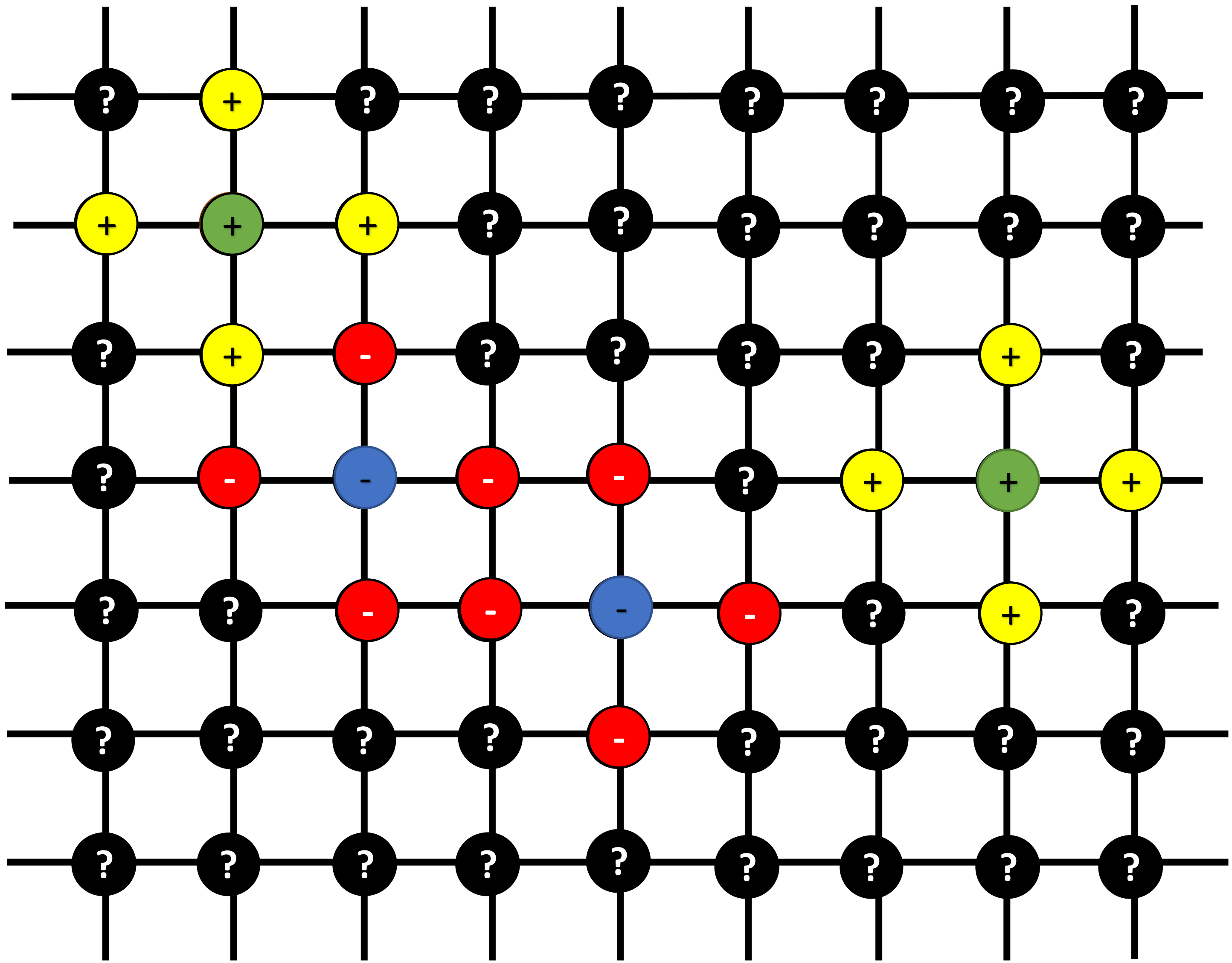


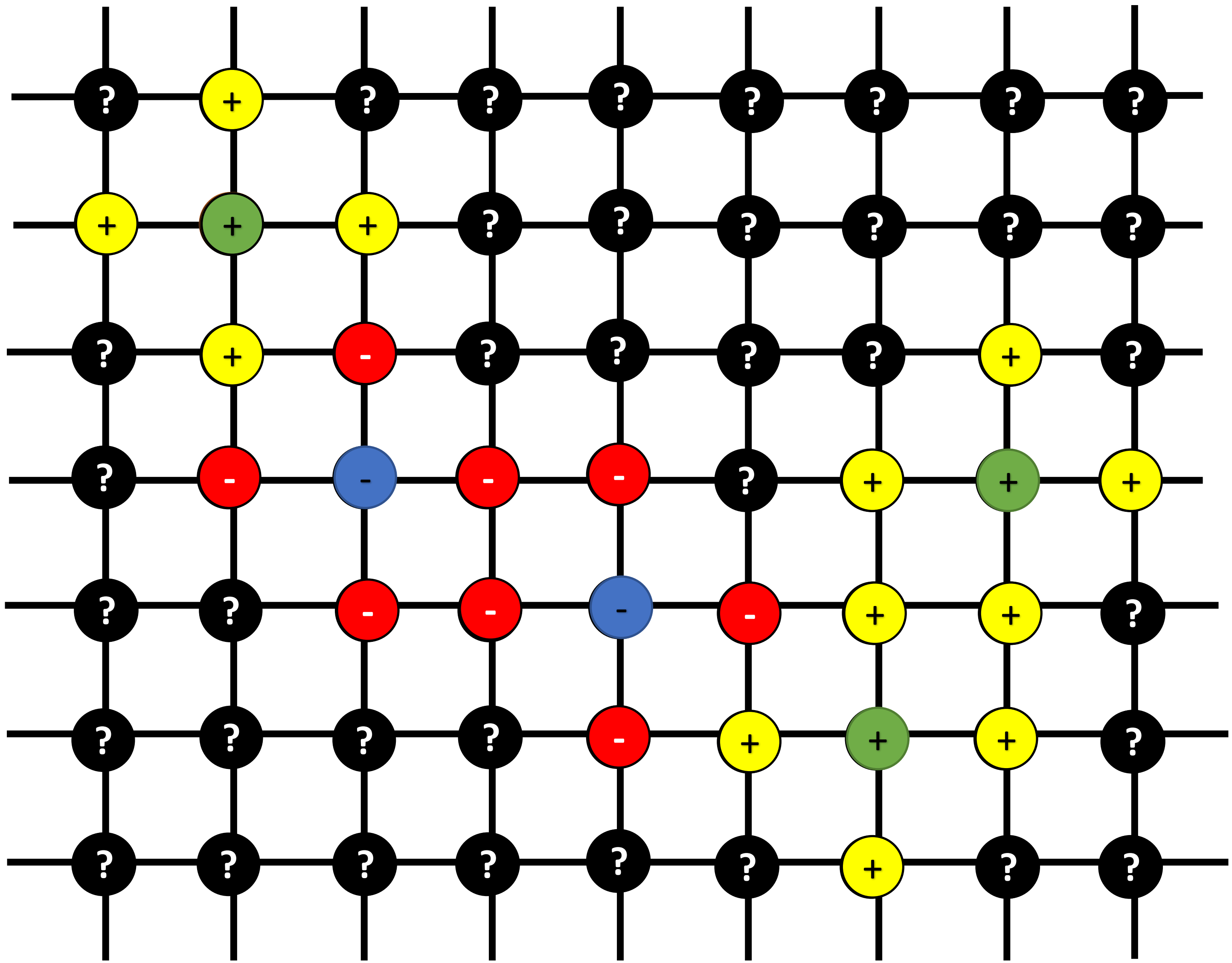


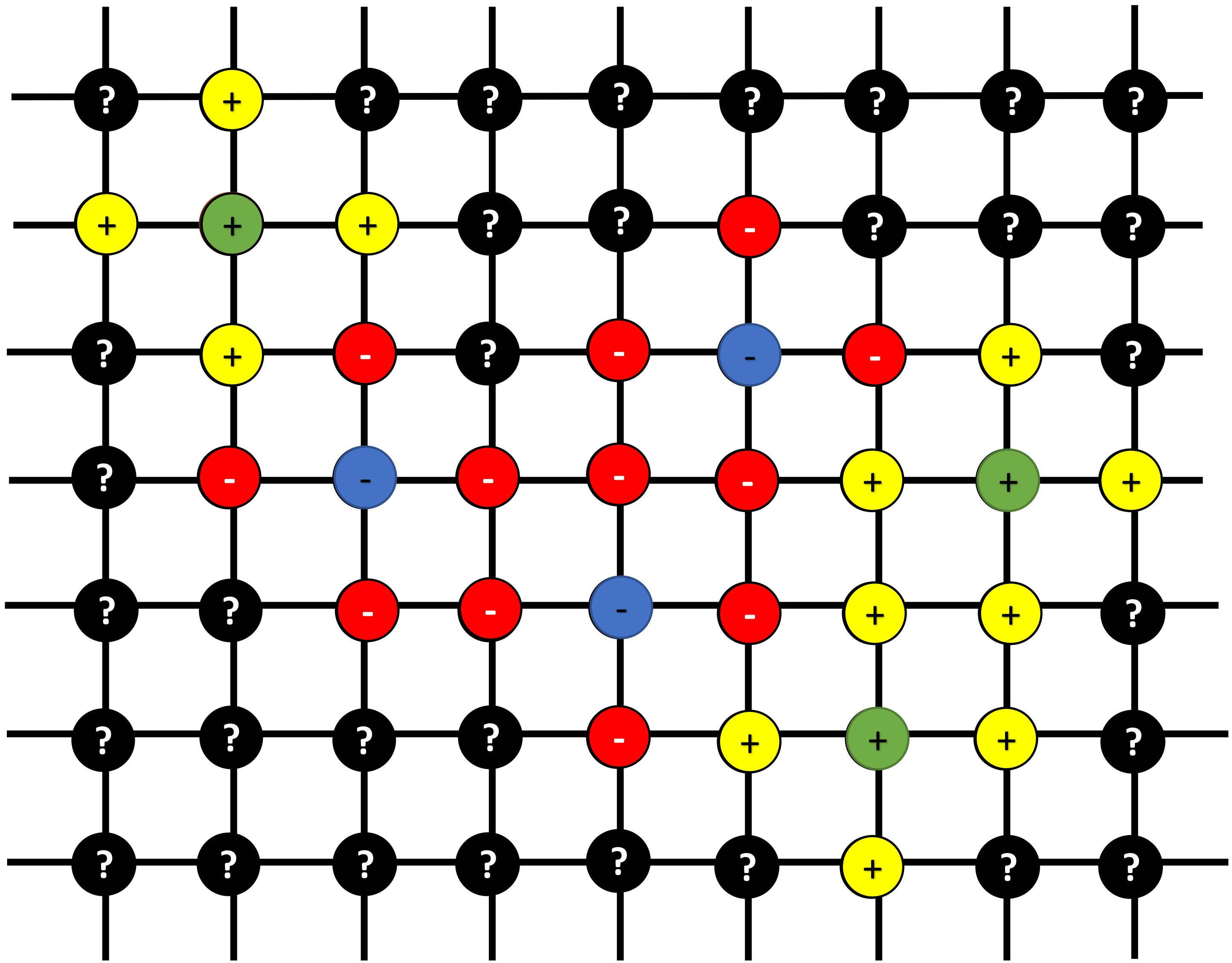


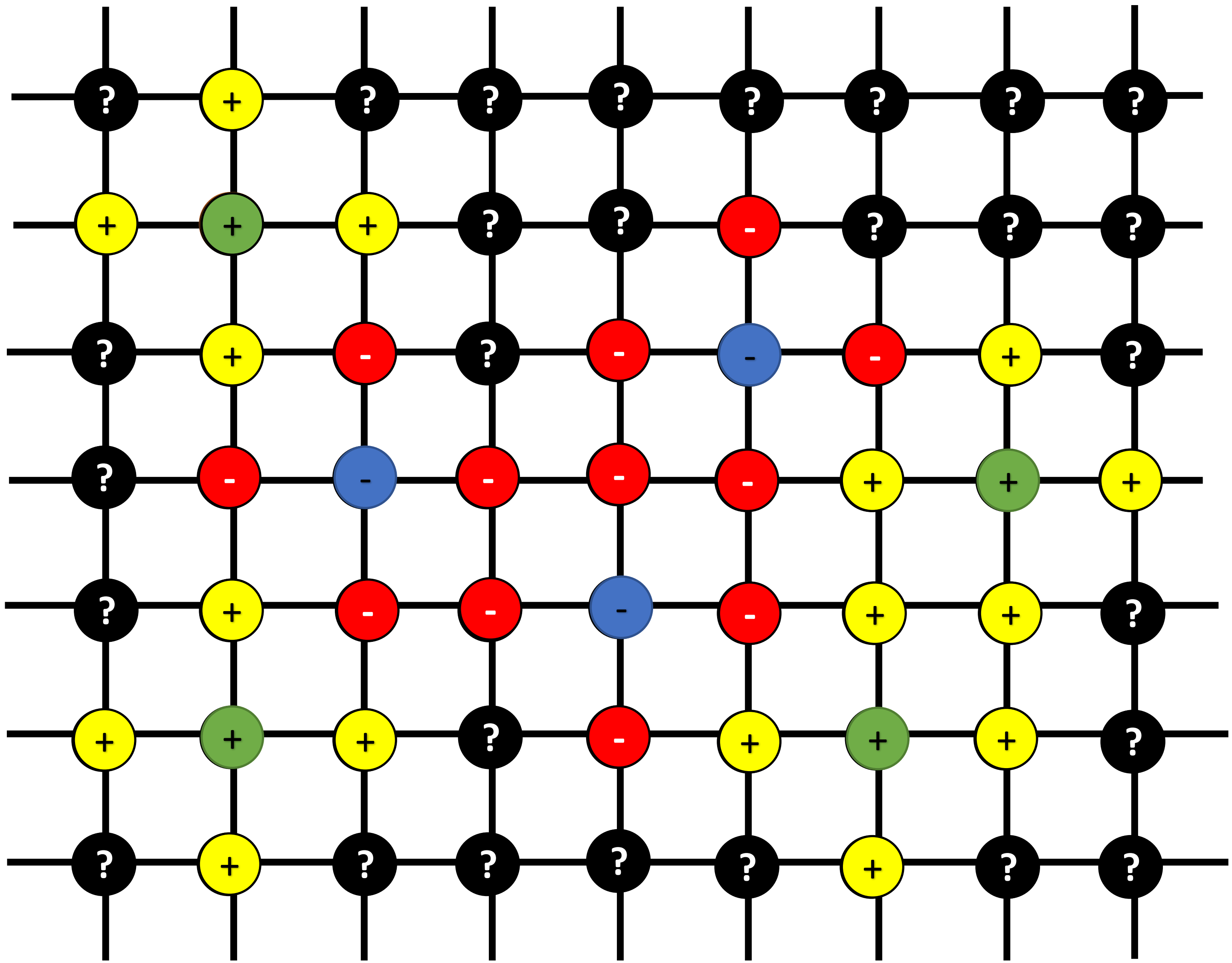


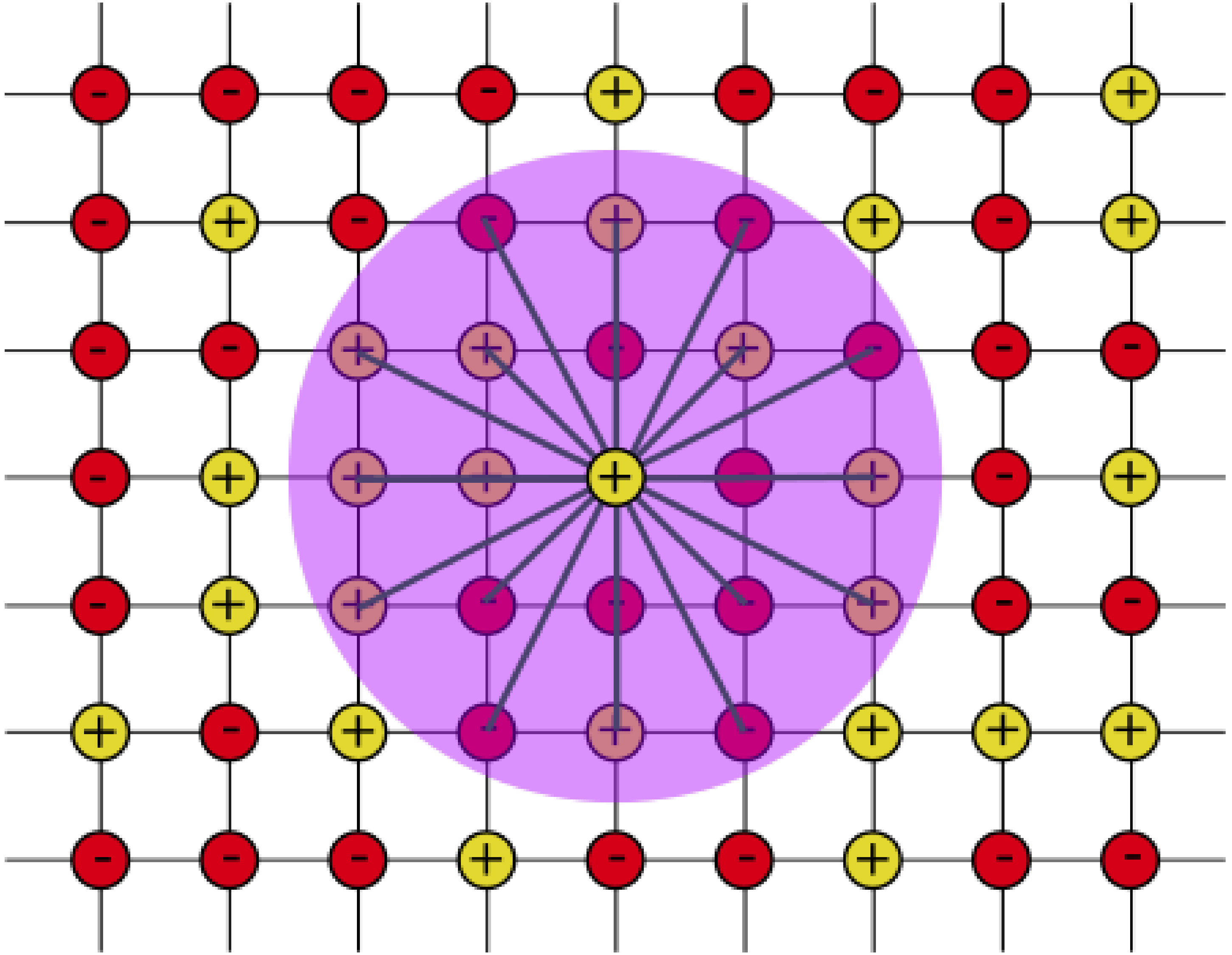


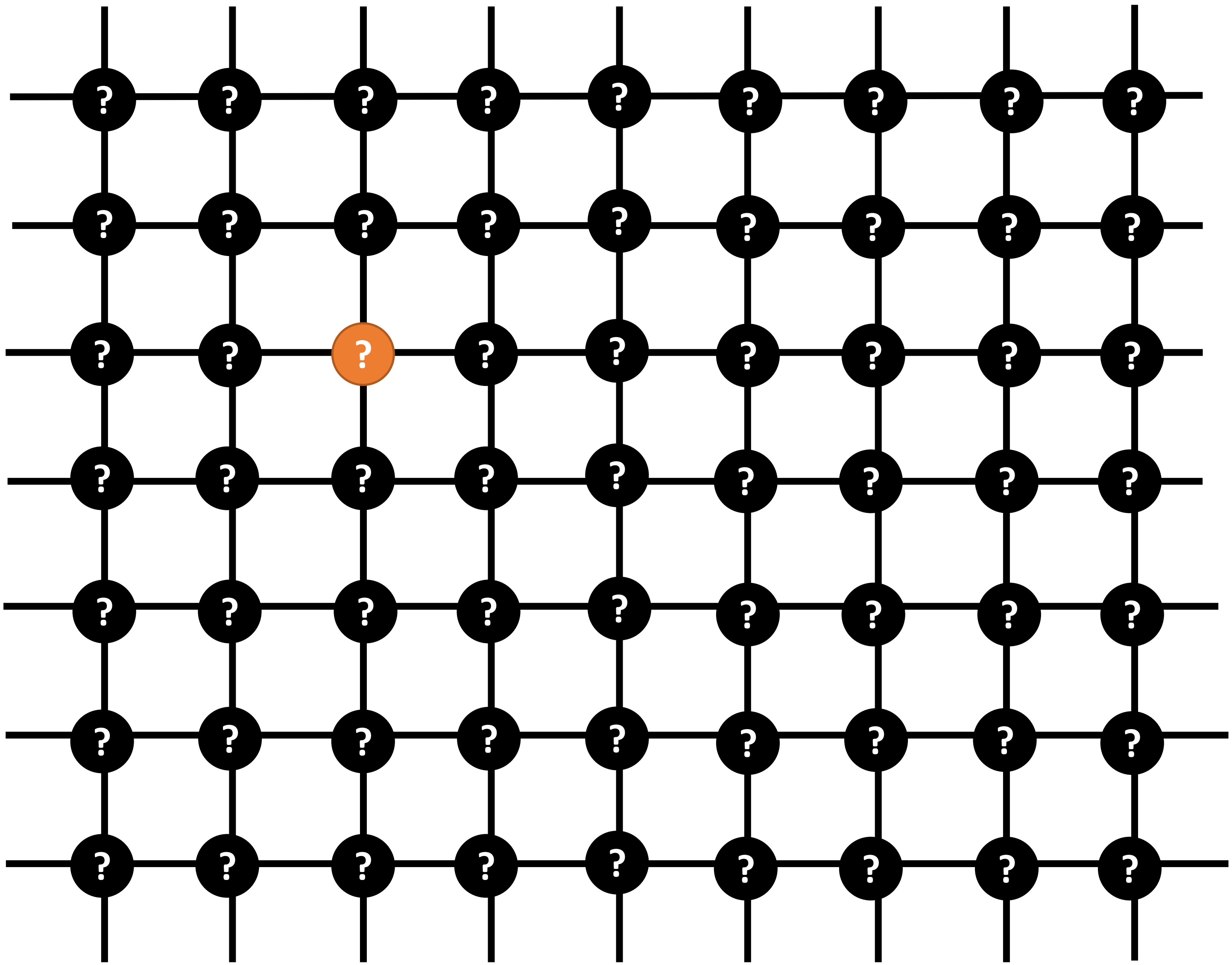


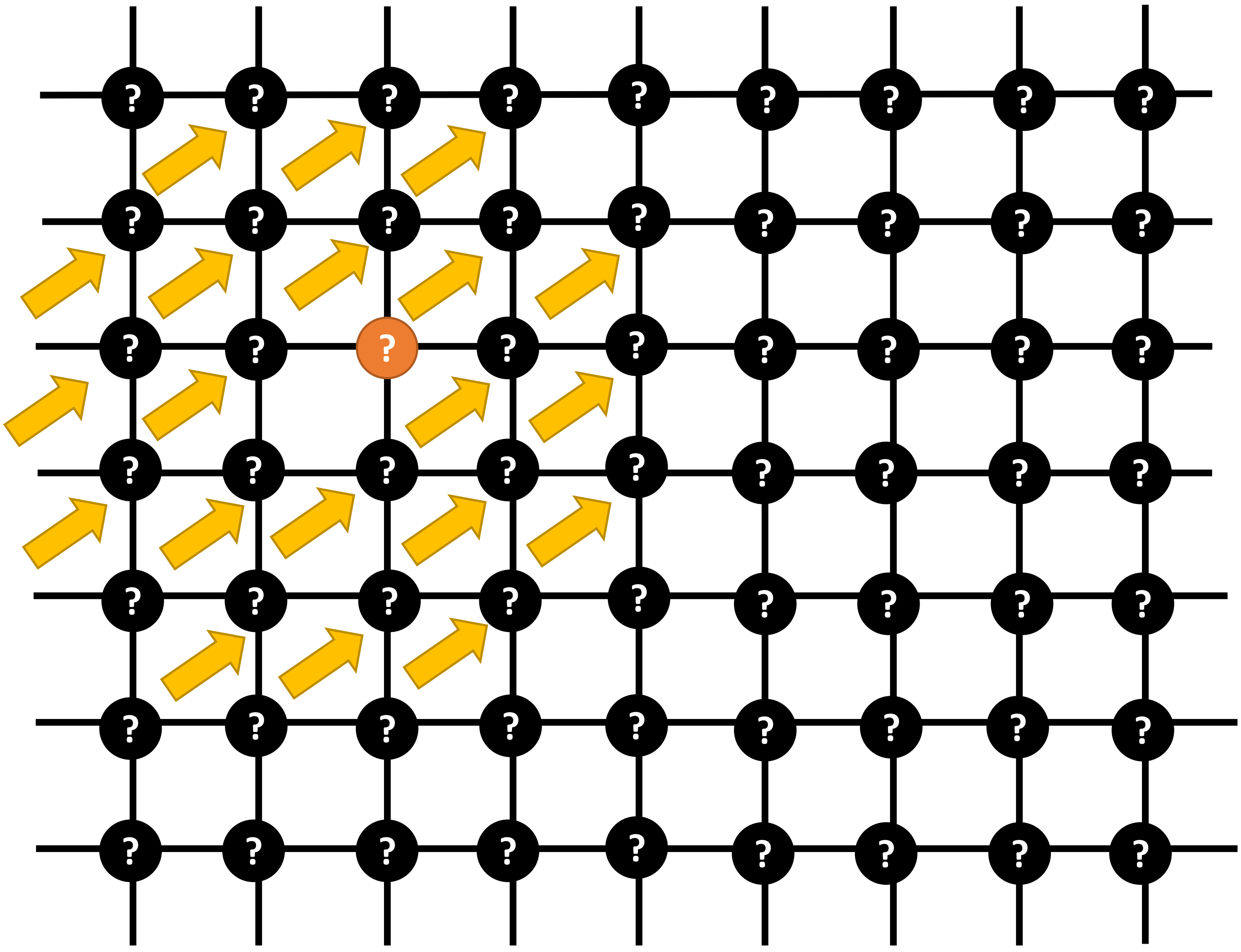


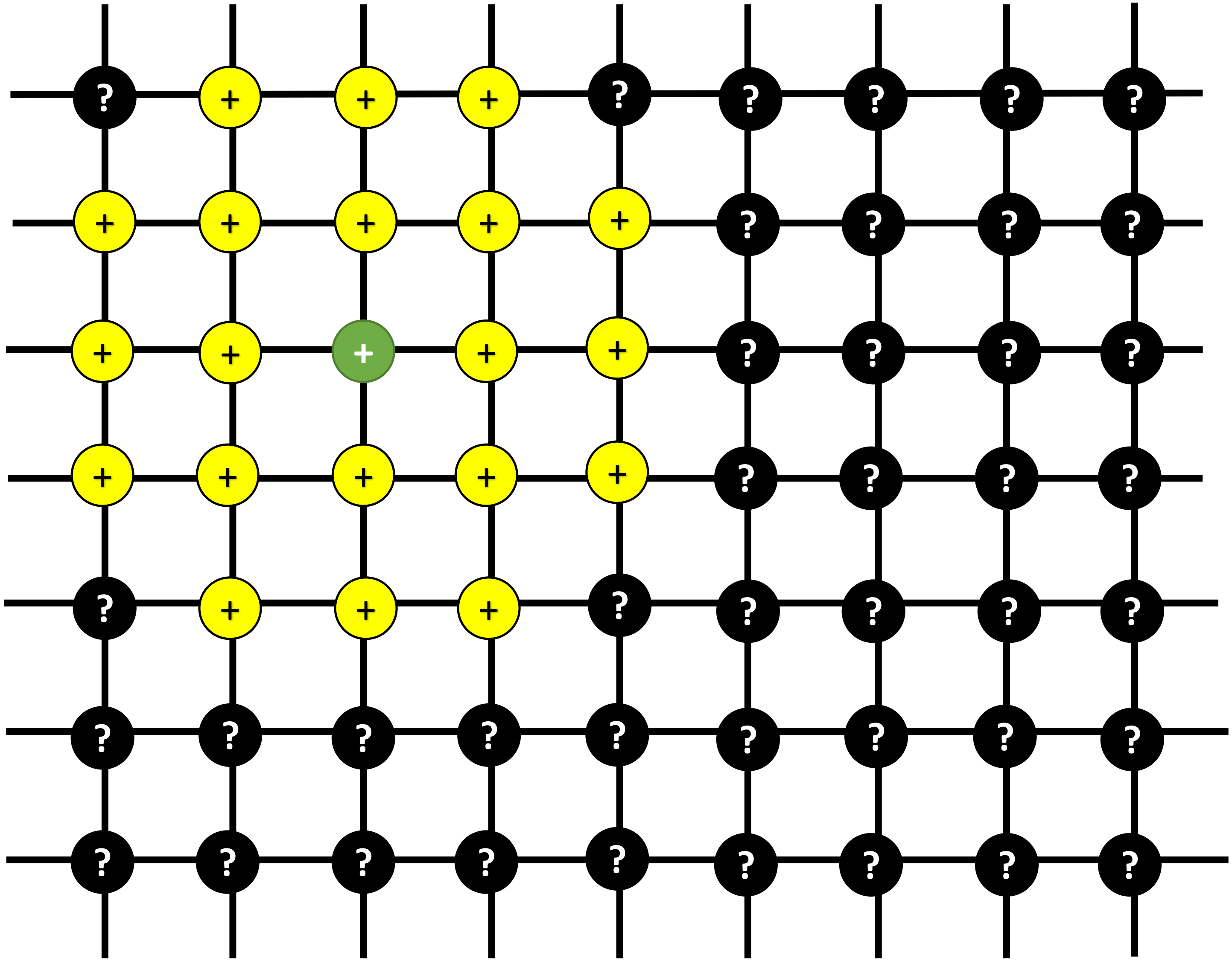












- This equation

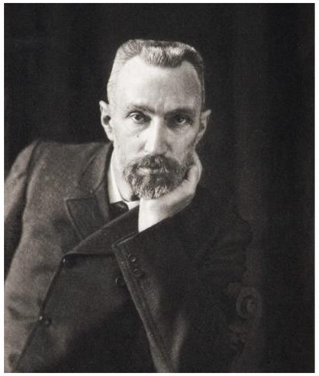
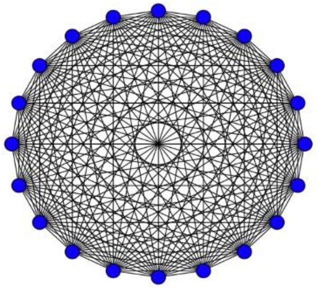
$$\left\|X - \tanh\left(\nabla f(X)\right)\right\|_1 = o(n)$$

tells you how to break the symmetry of the system.

i.e. If there is a phase transition, you should find it in

the solutions of the equation

Easy example:


The Curie Weiss Ising model:

$$f(X) = \frac{\beta}{n} \sum_{i \neq j} x_i x_j$$

$$= x^T A x \quad \text{for} \quad \mathrm{A} = \begin{bmatrix} 0 & \cdots & \frac{\beta}{n} \\ \vdots & \ddots & \vdots \\ \frac{\beta}{n} & \cdots & 0 \end{bmatrix}$$

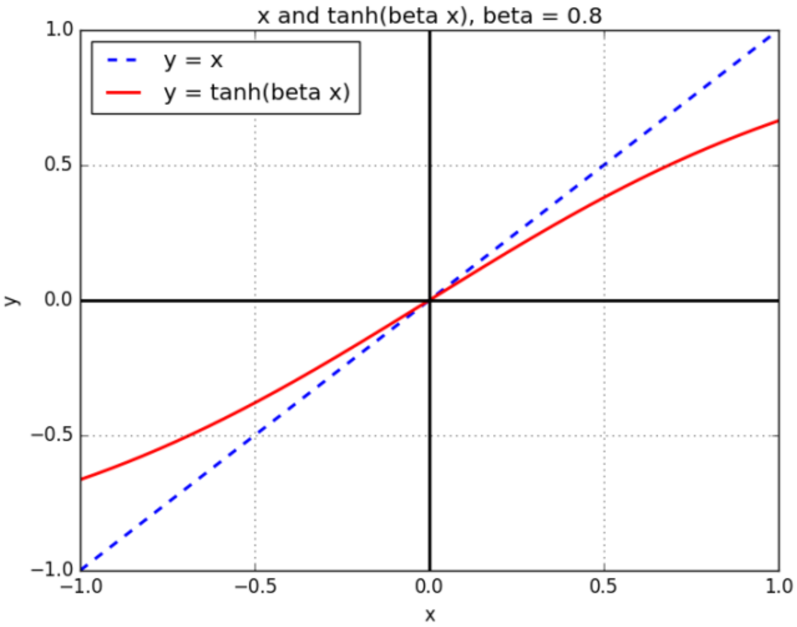
This gives $\nabla f(x) = Ax$, and so

$$\|x - \tanh(Ax)\|_1 = o(n)$$

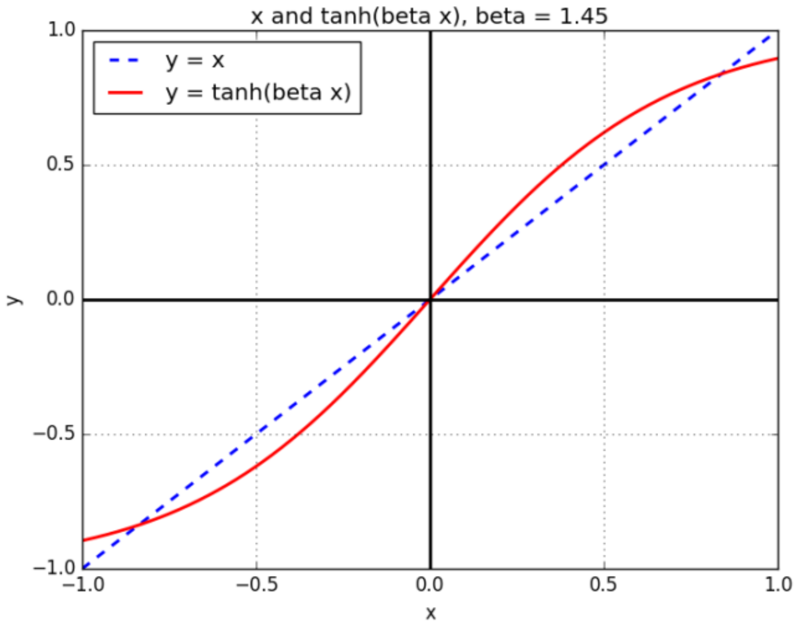
$$\|x - \tanh(Ax)\|_1 = o(n)$$

If $x = \mathbf{1} \cdot \bar{x}$ is a scalar vector, we get a scalar equation

$$\bar{x} = \tanh(\beta \bar{x})$$



$\beta < 1$

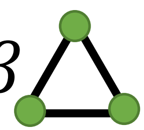


$\beta > 1$

- This equation

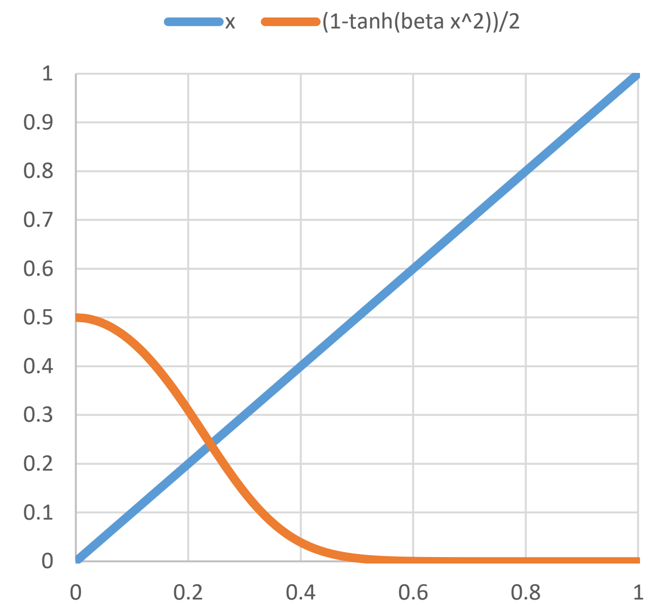
$$\left\| X - \tanh\left(\nabla f(X)\right) \right\|_1 = o(n)$$

tells you how to break symmetry.

Harder example: -β  counts in ERGM,

$$X = \frac{1 - \tanh\left(\frac{\beta}{n} X^2\right)}{2}$$

$$X = \frac{1 - \tanh\left(\frac{\beta}{n} X^2\right)}{2}$$



- For small β, only the trivial solution exists:

$$X = \mathbf{1} \cdot c$$

- For large β there exists a bipartite solution
  - Corresponds to cuts in graphs!
  - As $\beta \to \infty$, can tend to $G(\frac{n}{2}, \frac{n}{2}, \frac{1}{2})$
- What else?

For more information, call

1-800-https://arxiv.org/abs/1708.05859

Thanks!