Exact Recovery for a Family of Community-Detection Generative Models

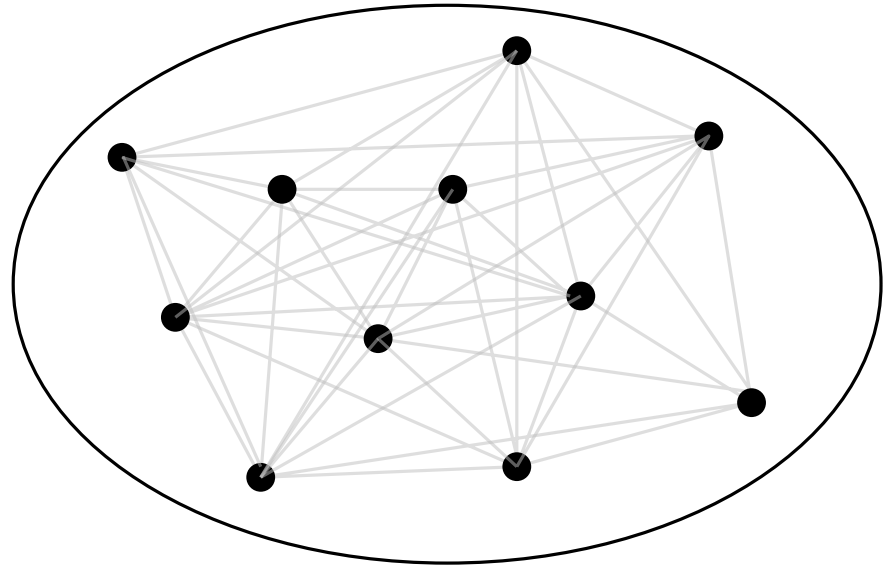
Paolo Penna

joint work with

Joachim Buhmann, Luca Corinzia, Luca Mondada

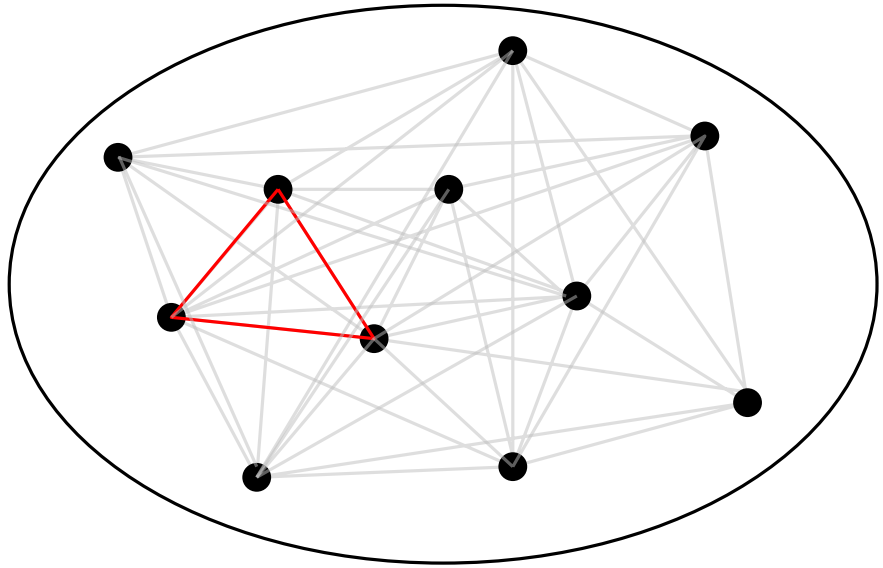
ISE Group, Inst for Machine Learning, ETH Zurich

A Toy Problem



A Toy Problem

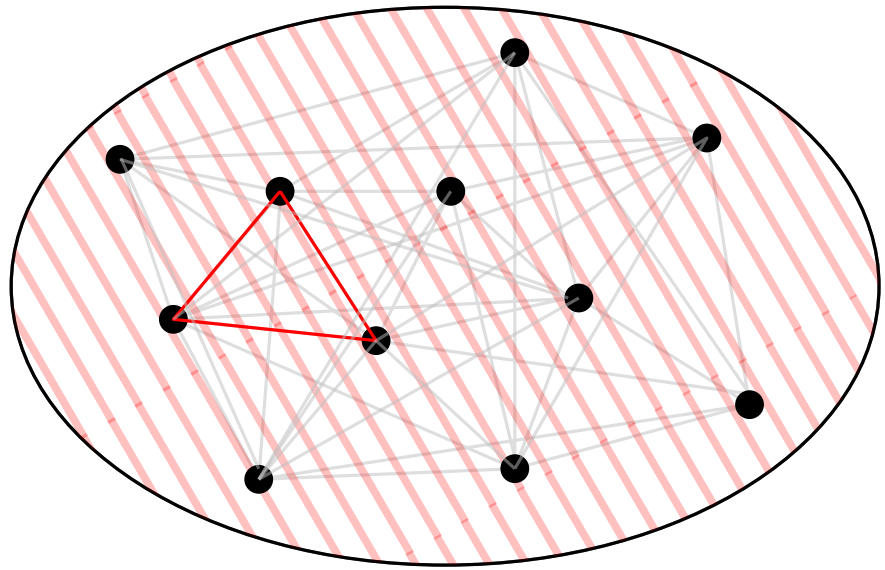
pick random triangle



A Toy Problem

pick random triangle

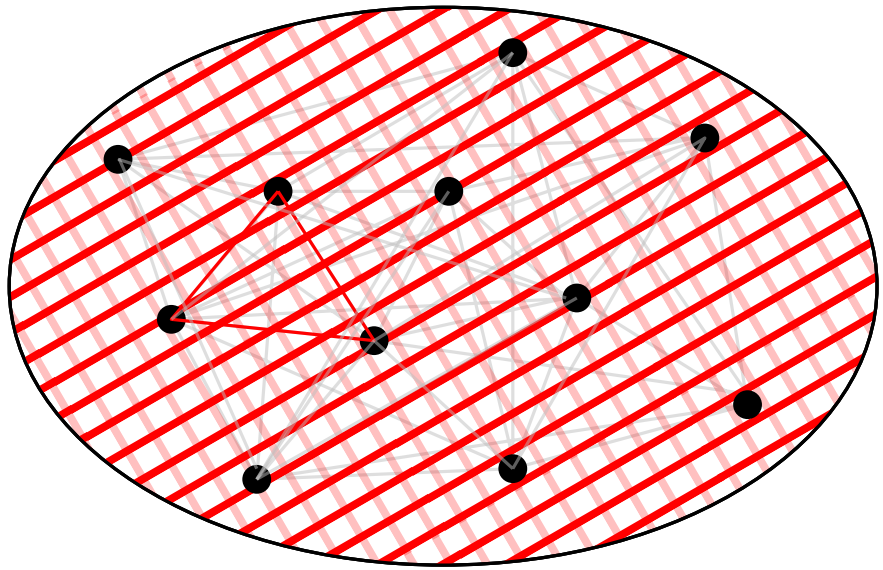
add noise



A Toy Problem

pick random triangle

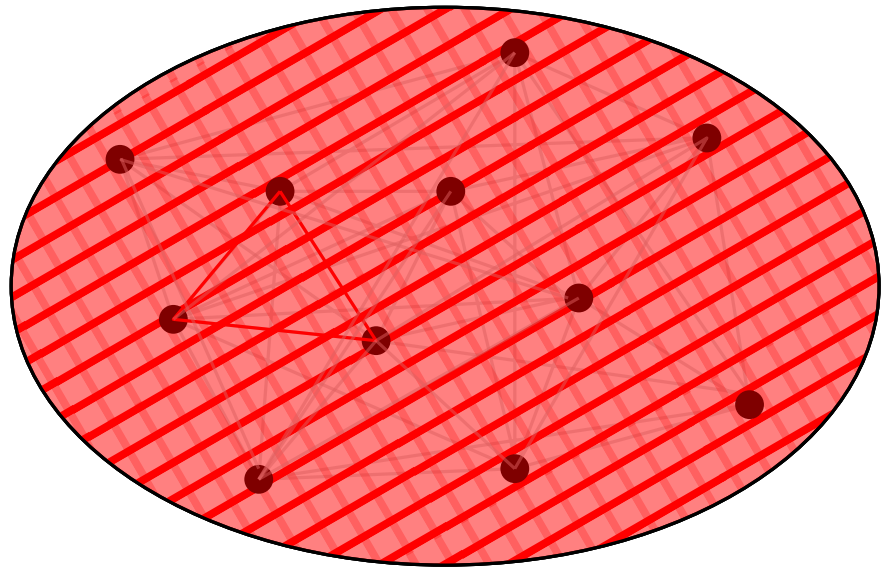
add noise



A Toy Problem

pick random triangle

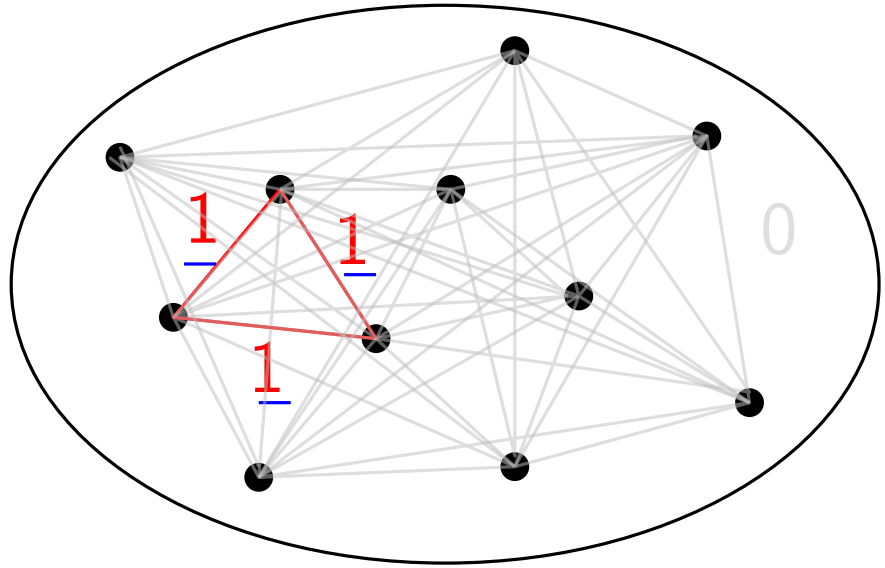
add noise



Find the triangle?

A Toy Problem

pick random triangle

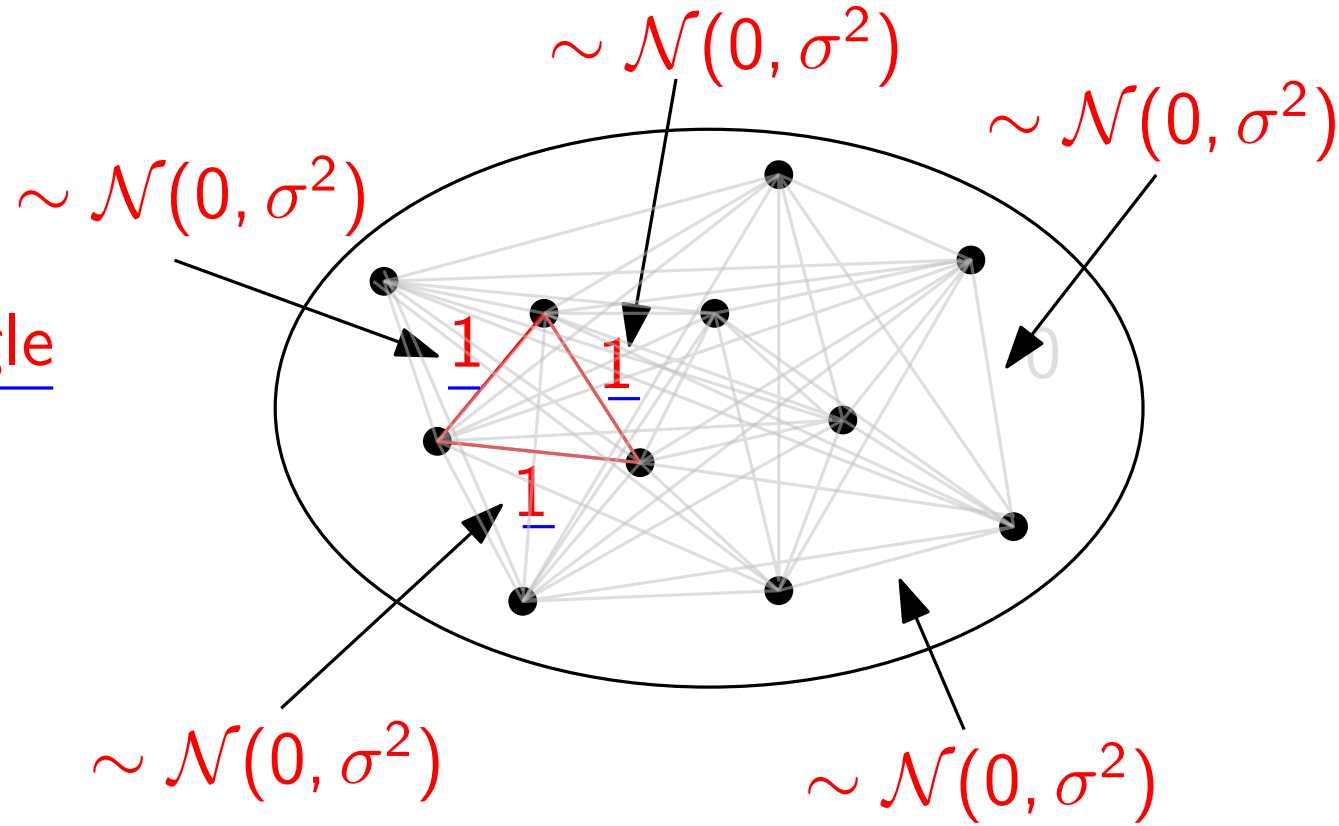


Find the triangle?

A Toy Problem

pick random triangle

add noise

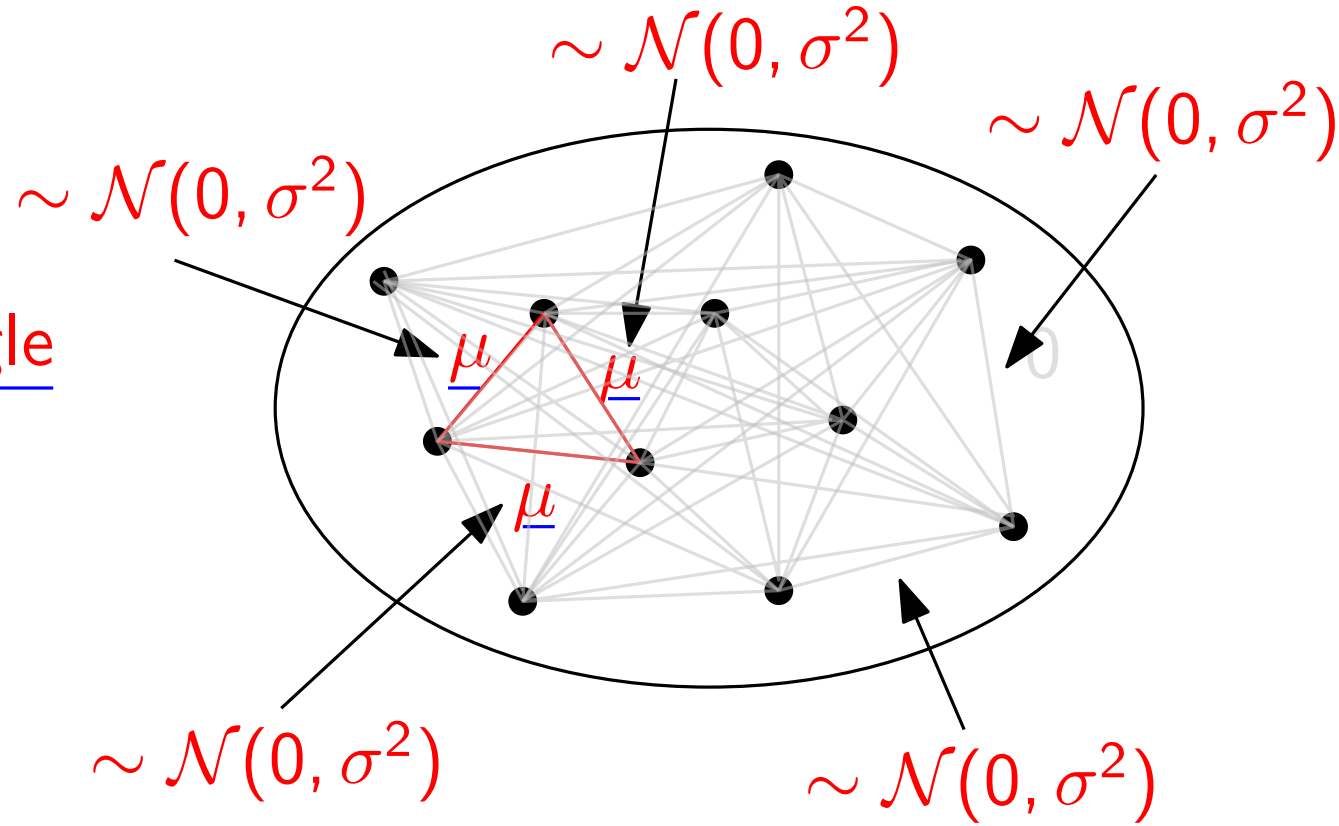


Find the triangle?

A Toy Problem

pick random triangle

add noise

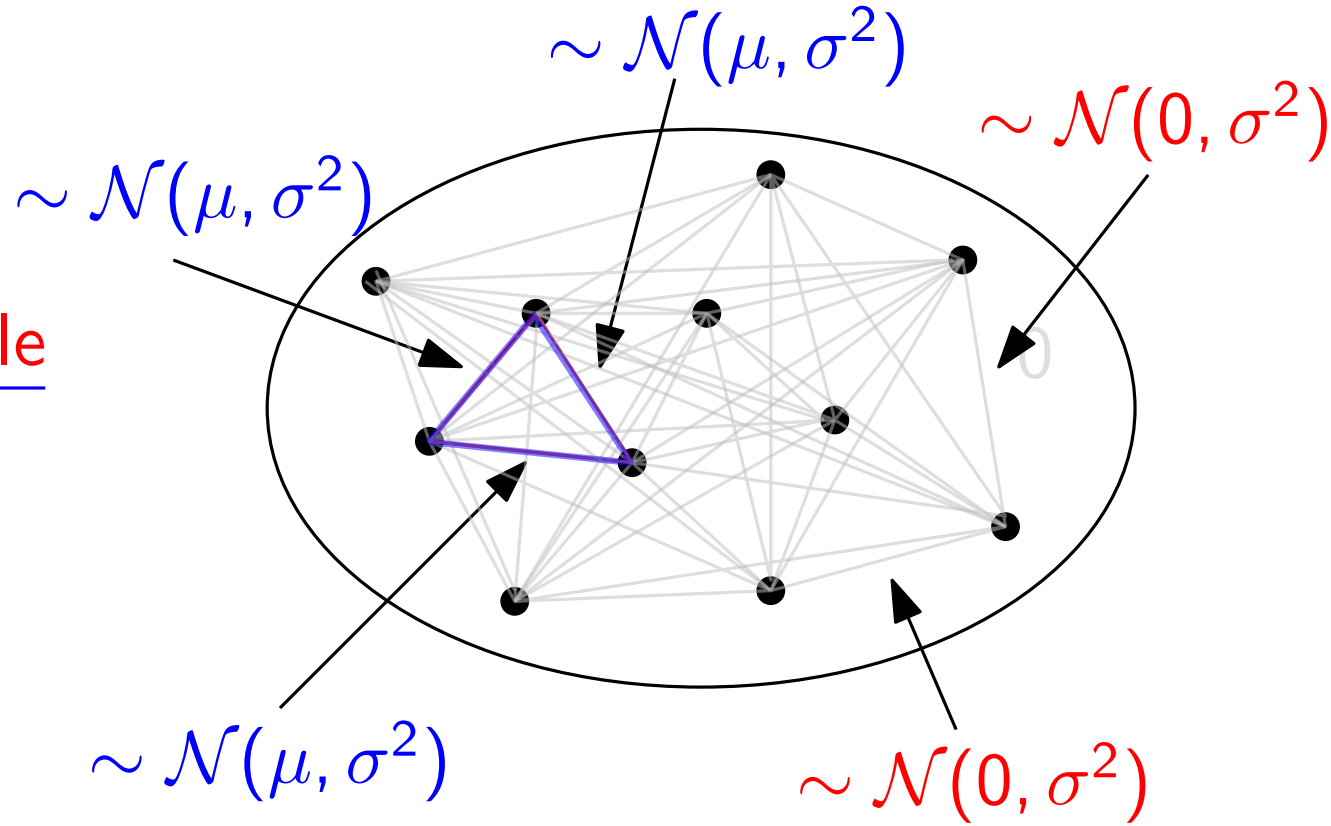


Find the triangle?

A Toy Problem

pick random triangle

add noise

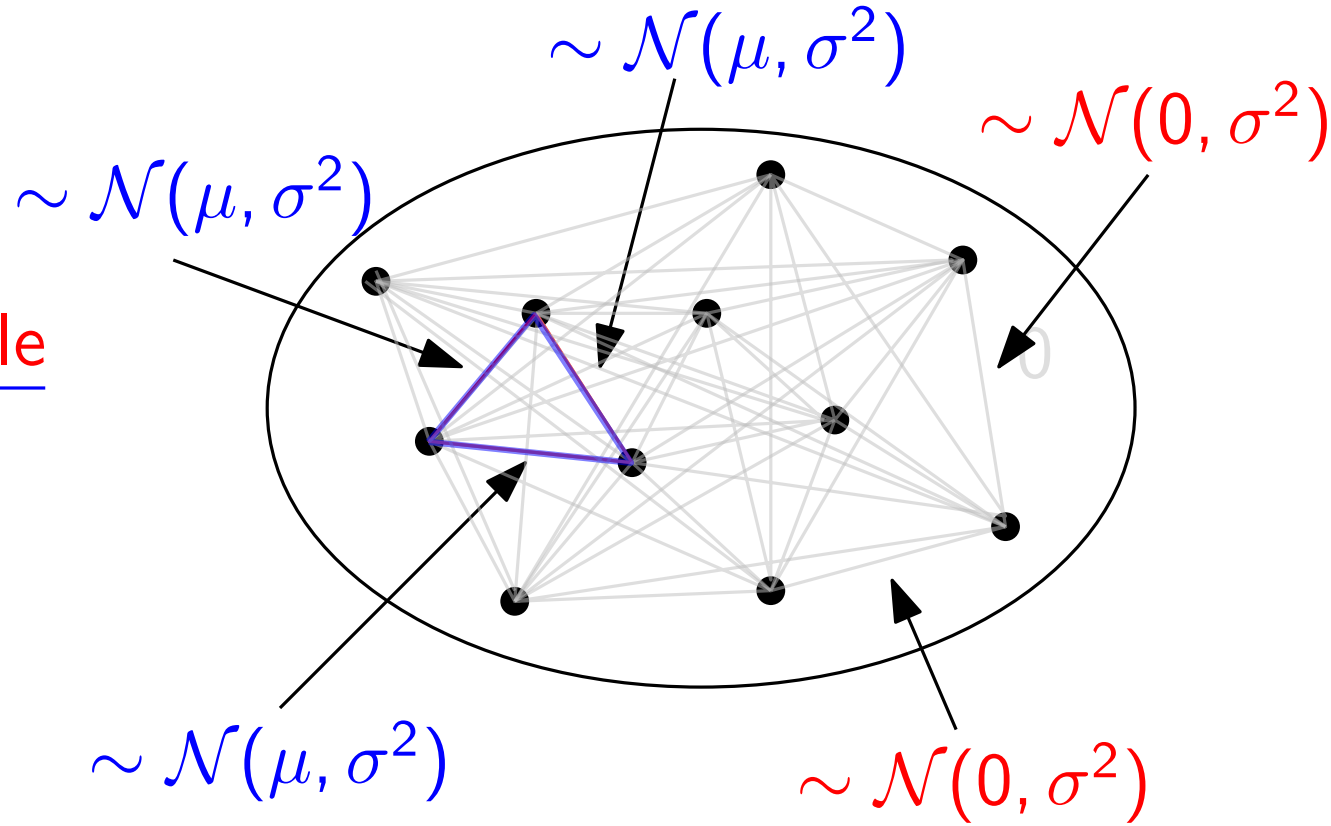


Find the triangle?

A Toy Problem

pick random triangle

add noise



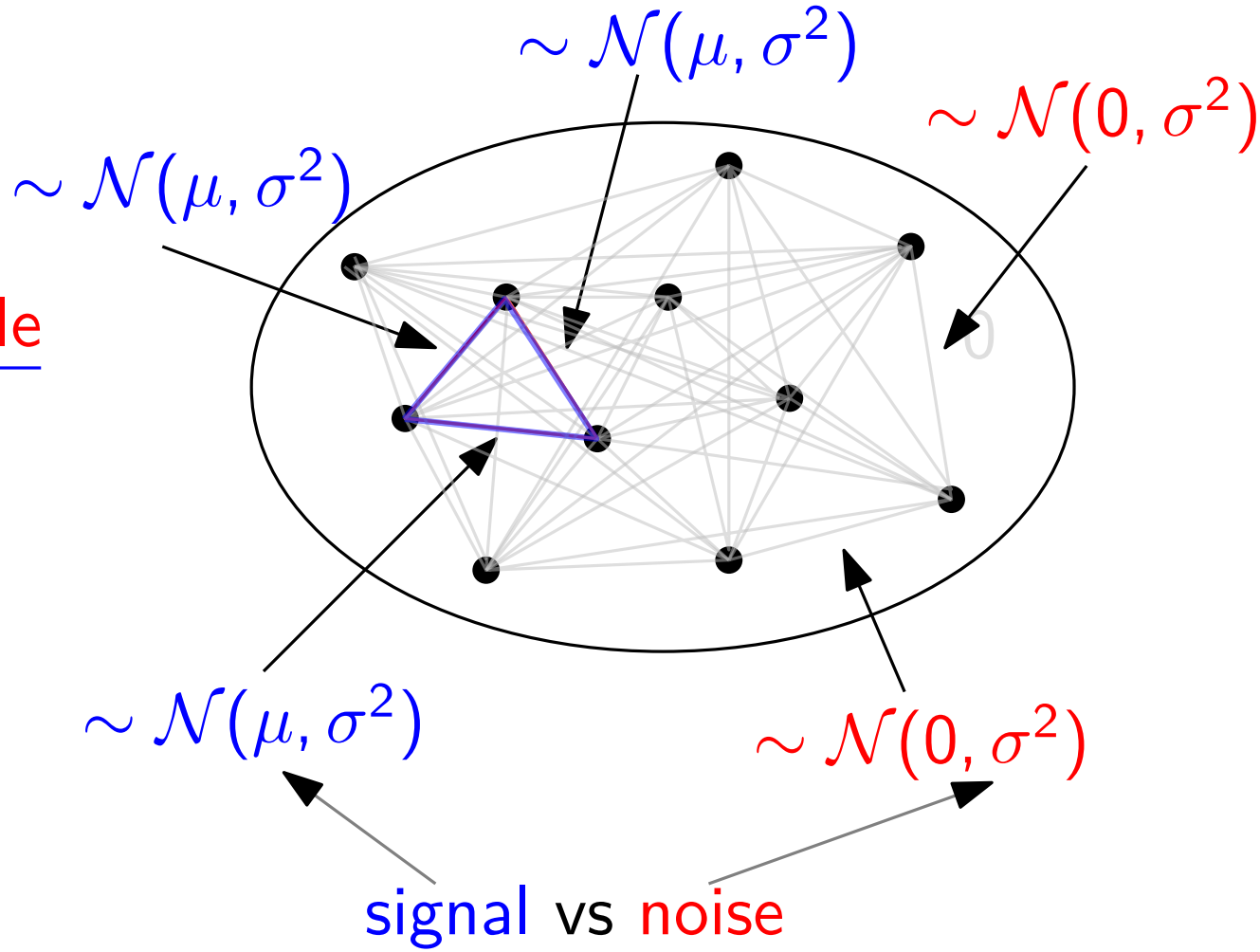
Return **heaviest** triangle

Find the triangle?

A Toy Problem

pick random triangle

add noise



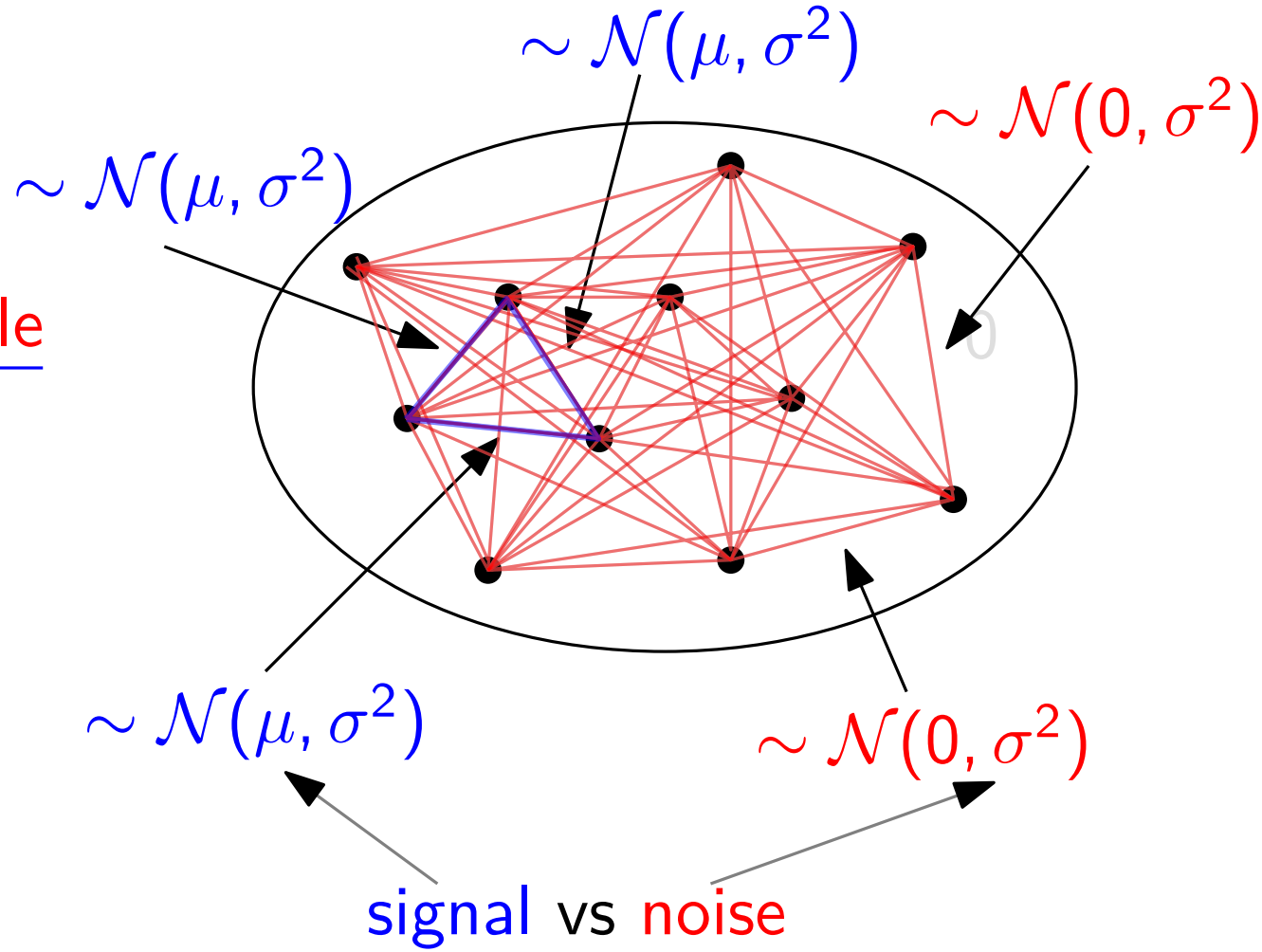
Return **heaviest** triangle

Find the triangle?

A Toy Problem

pick random triangle

add noise



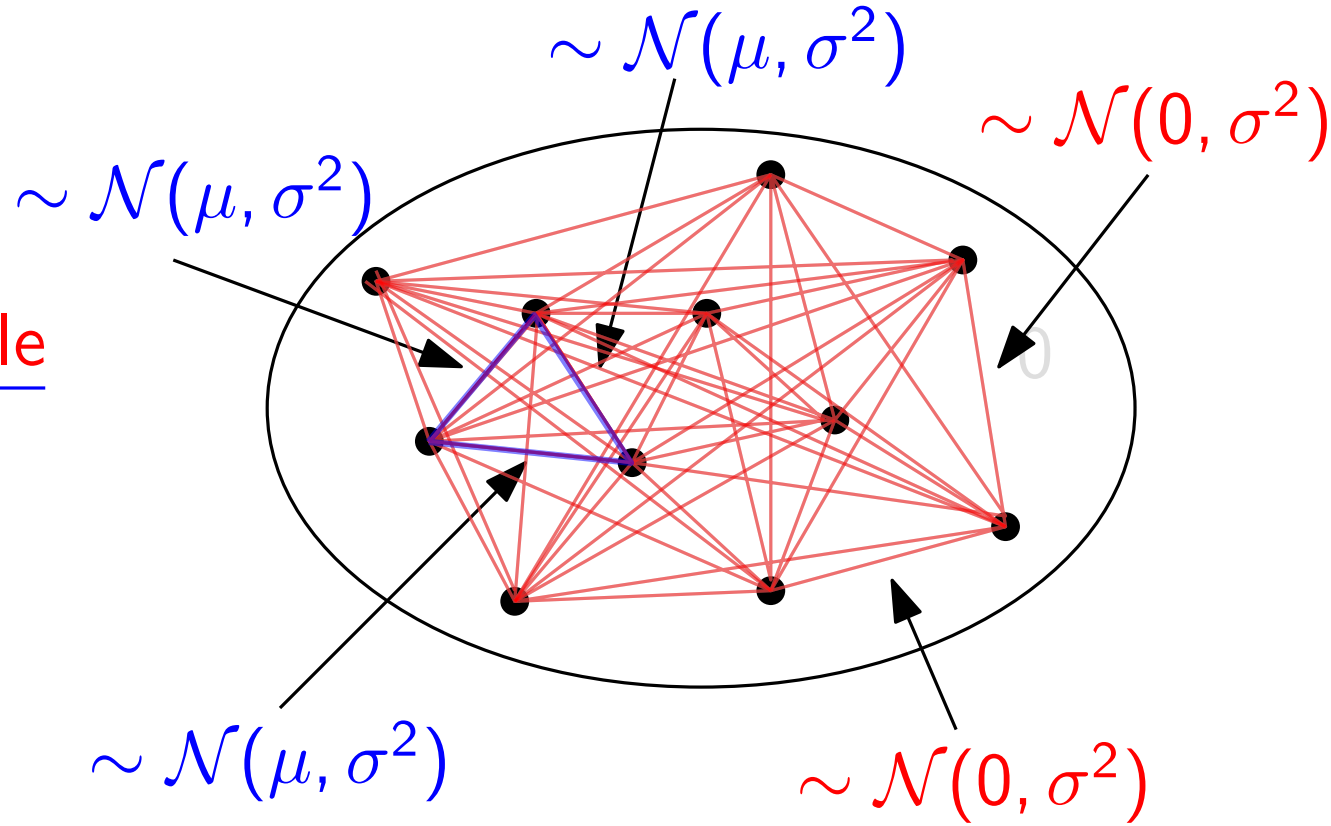
Return **heaviest** triangle

Find the triangle?

A Toy Problem

pick random triangle

add noise



random graph model

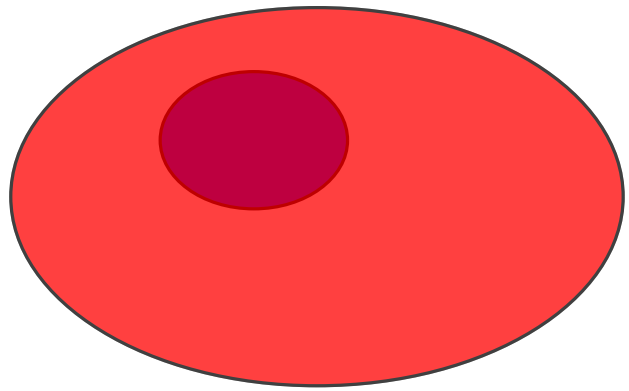
Return **heaviest** triangle

Find the triangle?

Flavor of the Problem

planted solution

random noise



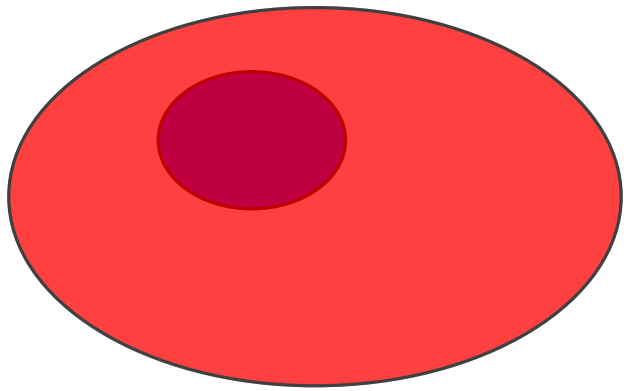
planted random model

Recover planted solution?

Flavor of the Problem

planted solution

random noise

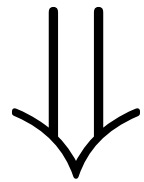


Many variants: planted clique, planted bisection, stochastic block model,...

planted random model

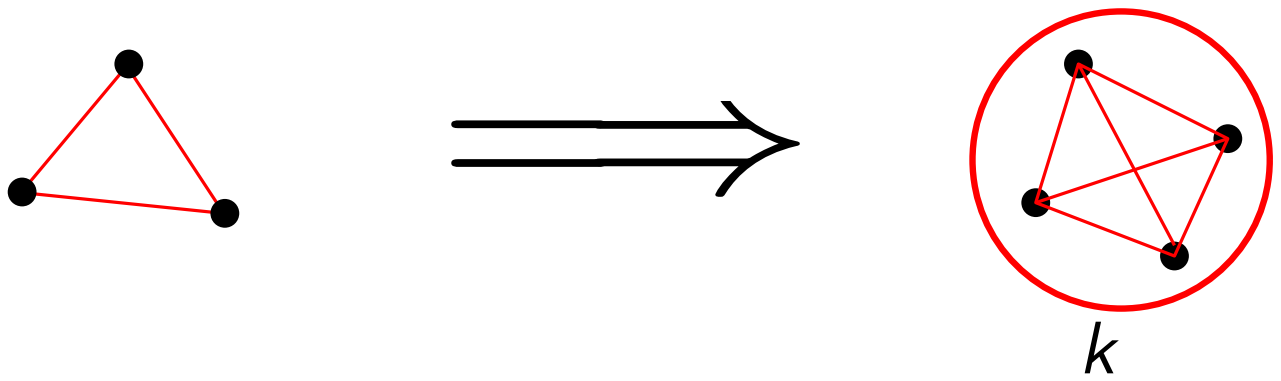
Recover planted solution?

Triangles

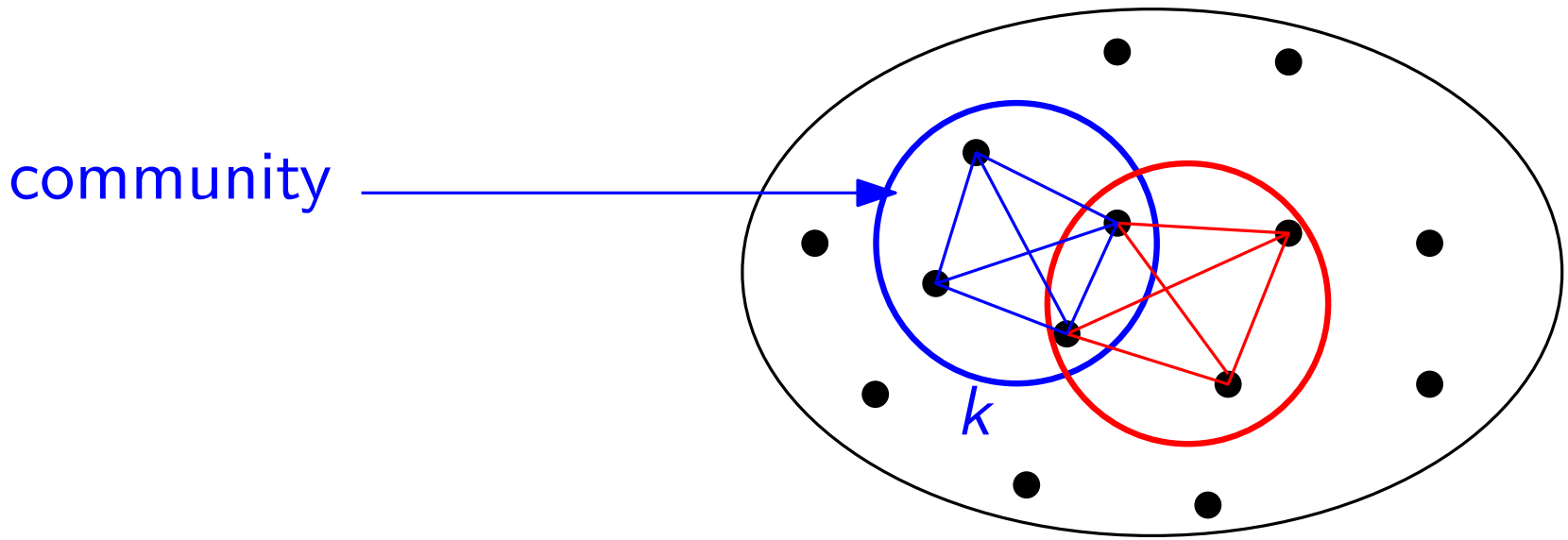


Our General Model

Generalization #1

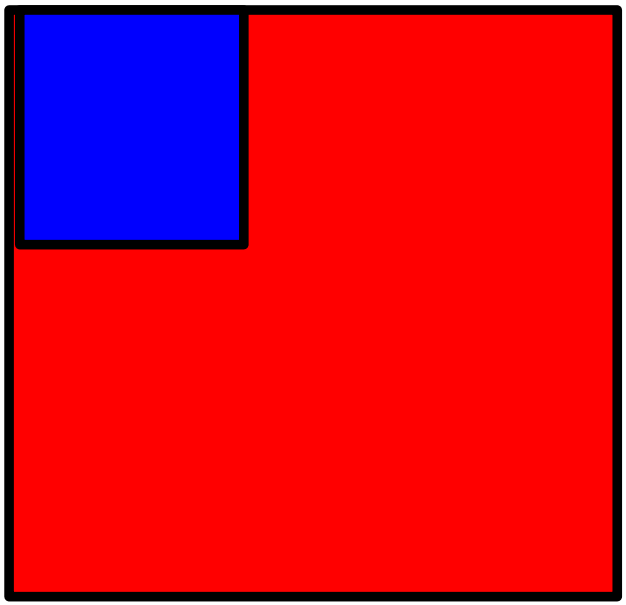


Planted Random Models

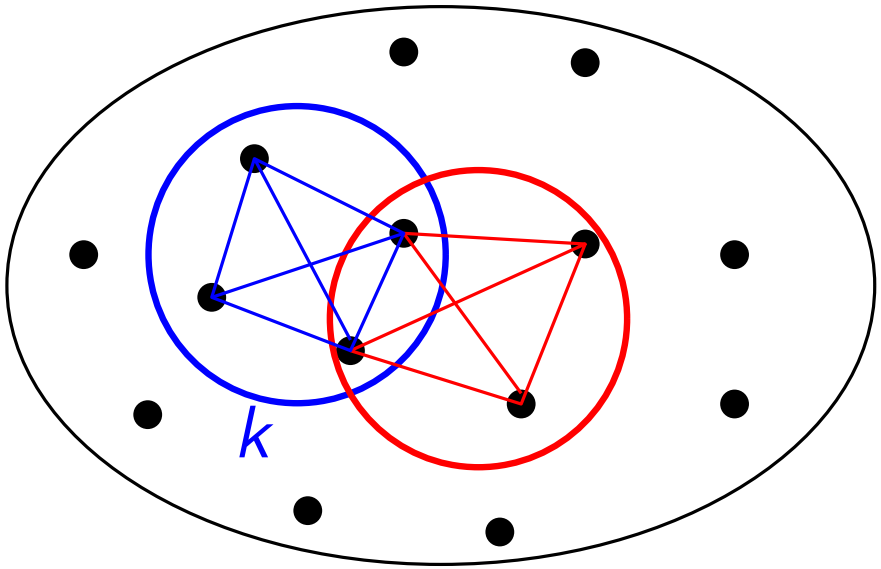


Random Graph

Planted Random Models

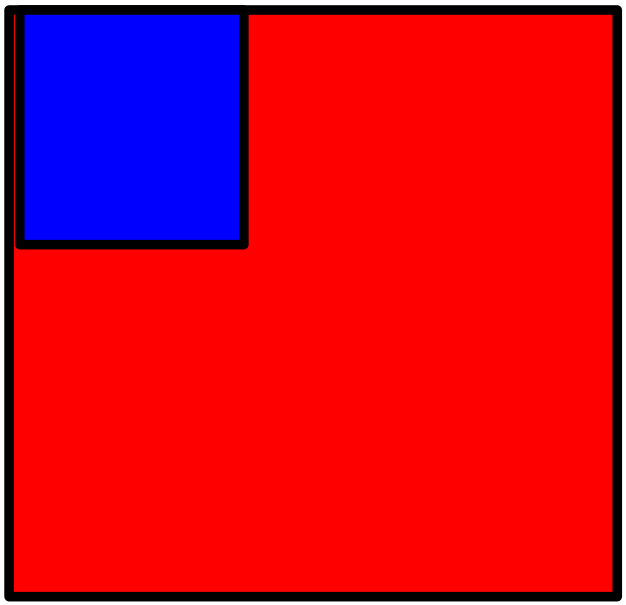


Weighted
Stochastic Block
Model

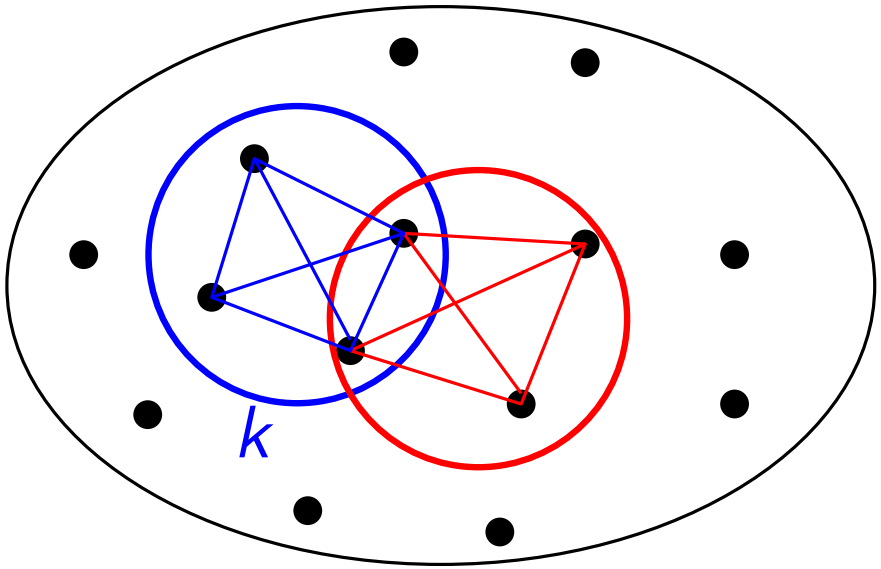


Random Graph

Planted Random Models



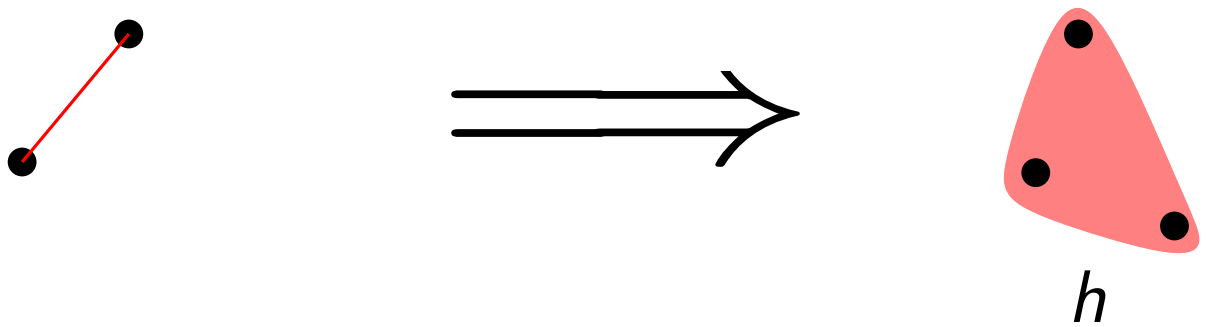
Weighted
Stochastic Block
Model



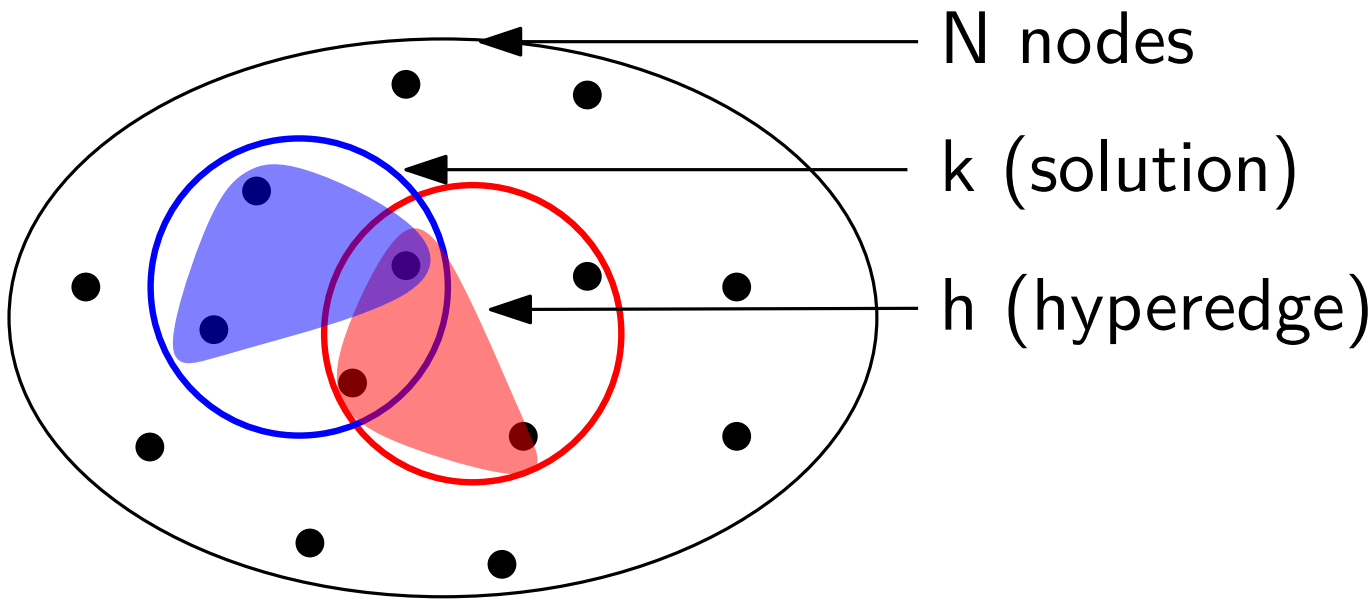
Random Graph

Densest k-Subgraph Problem

Generalization #2



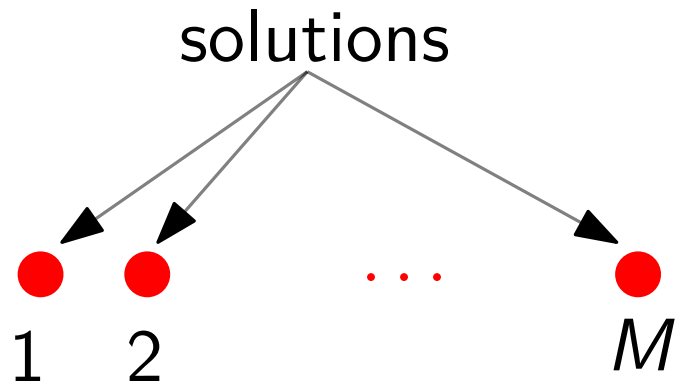
Our Model



The Simplest Model

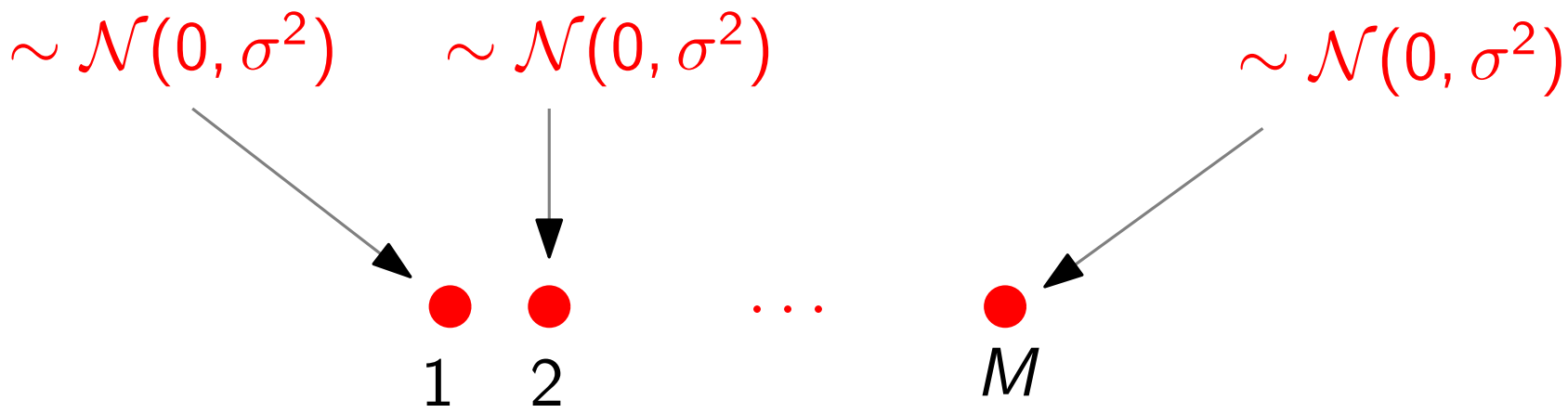
Random Energy Model (REM)

The Simplest Model



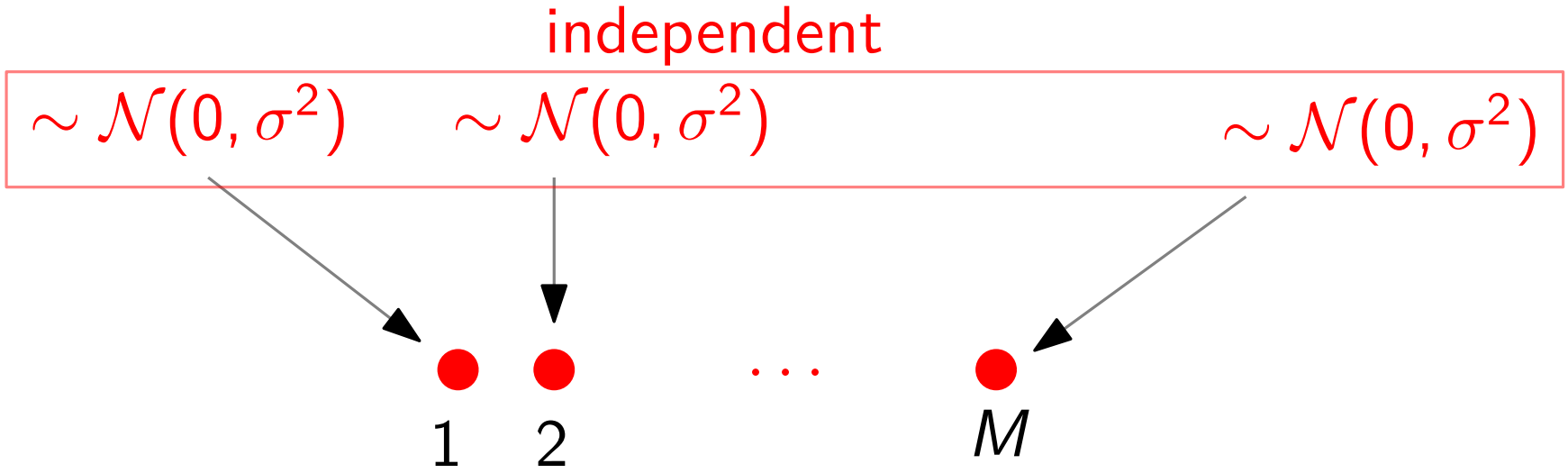
Random Energy Model (REM)

The Simplest Model



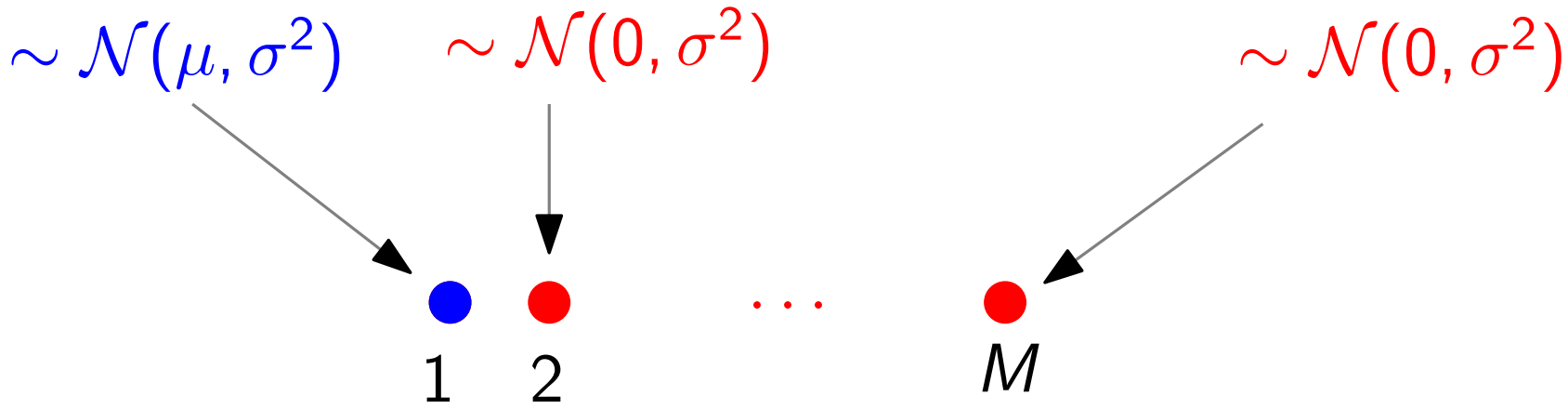
Random Energy Model (REM)

The Simplest Model



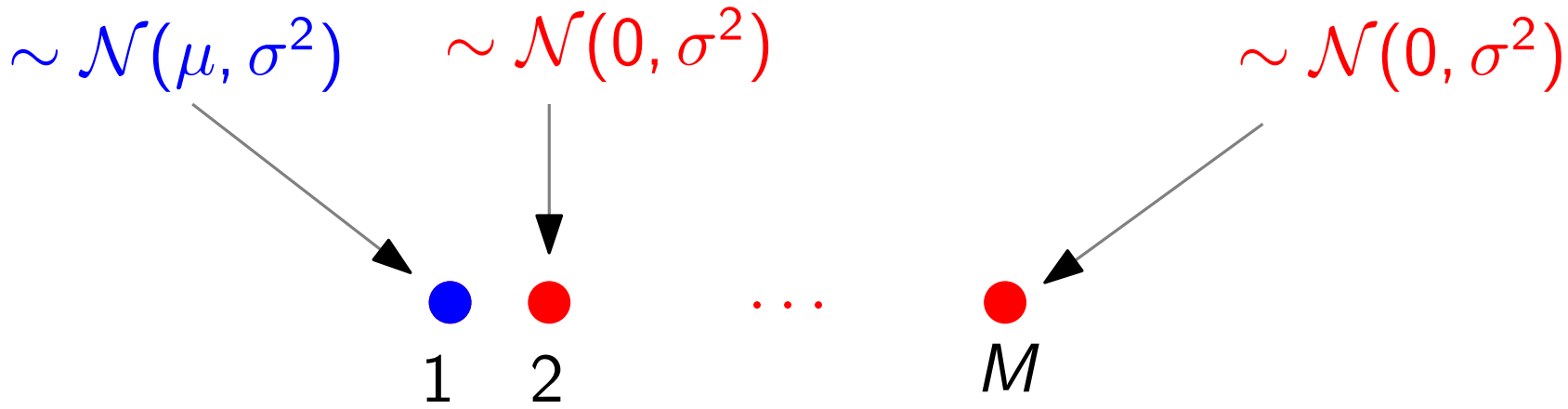
Random Energy Model (REM)

The Simplest Model



Planted Random Energy Model (P-REM)

The Simplest Model

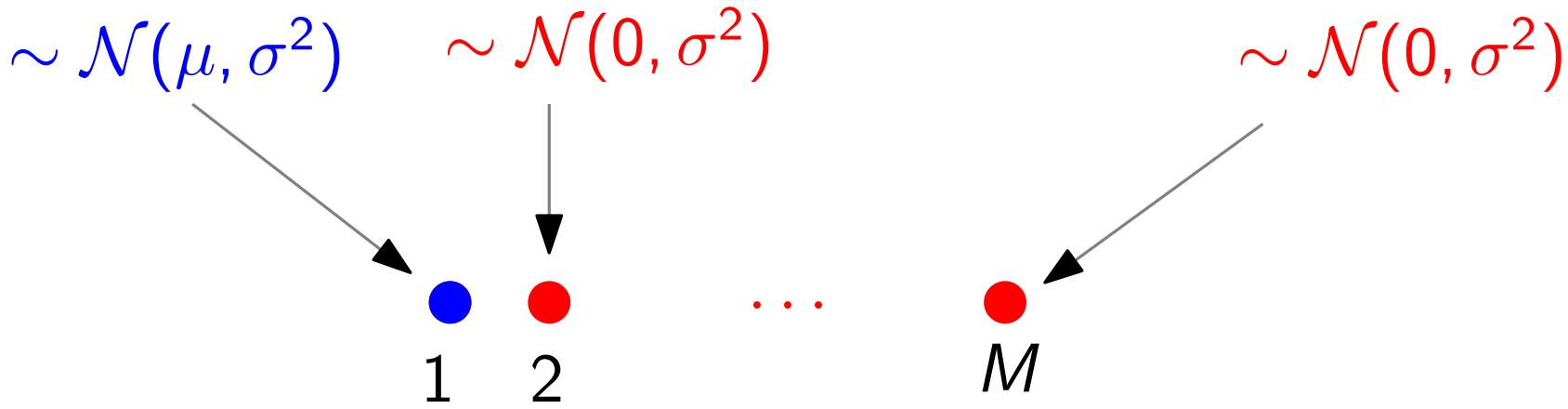


Planted Random Energy Model (P-REM)

Return **max weight** one

Recover planted solution?

The Simplest Model



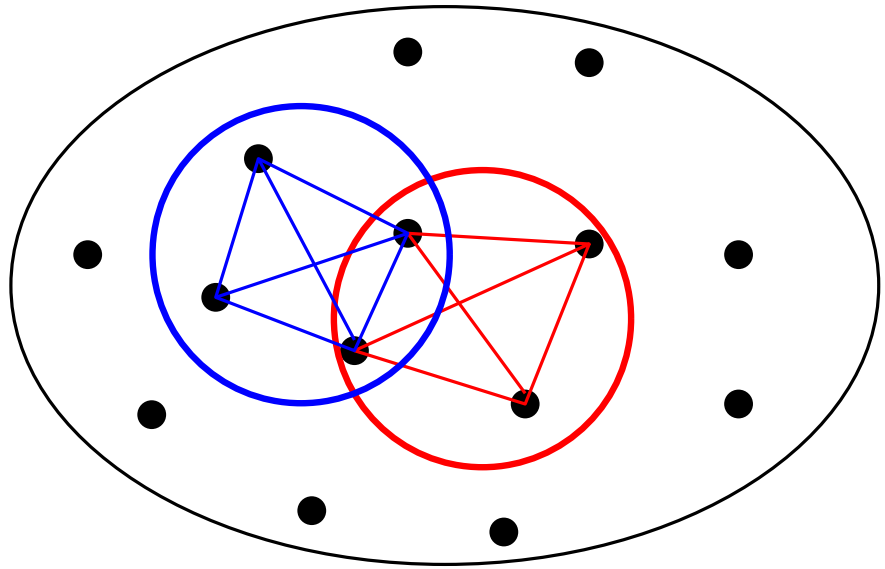
Planted Random Energy Model (P-REM)

Maximum Likelihood (ML)

Return **max weight** one

Recover planted solution?

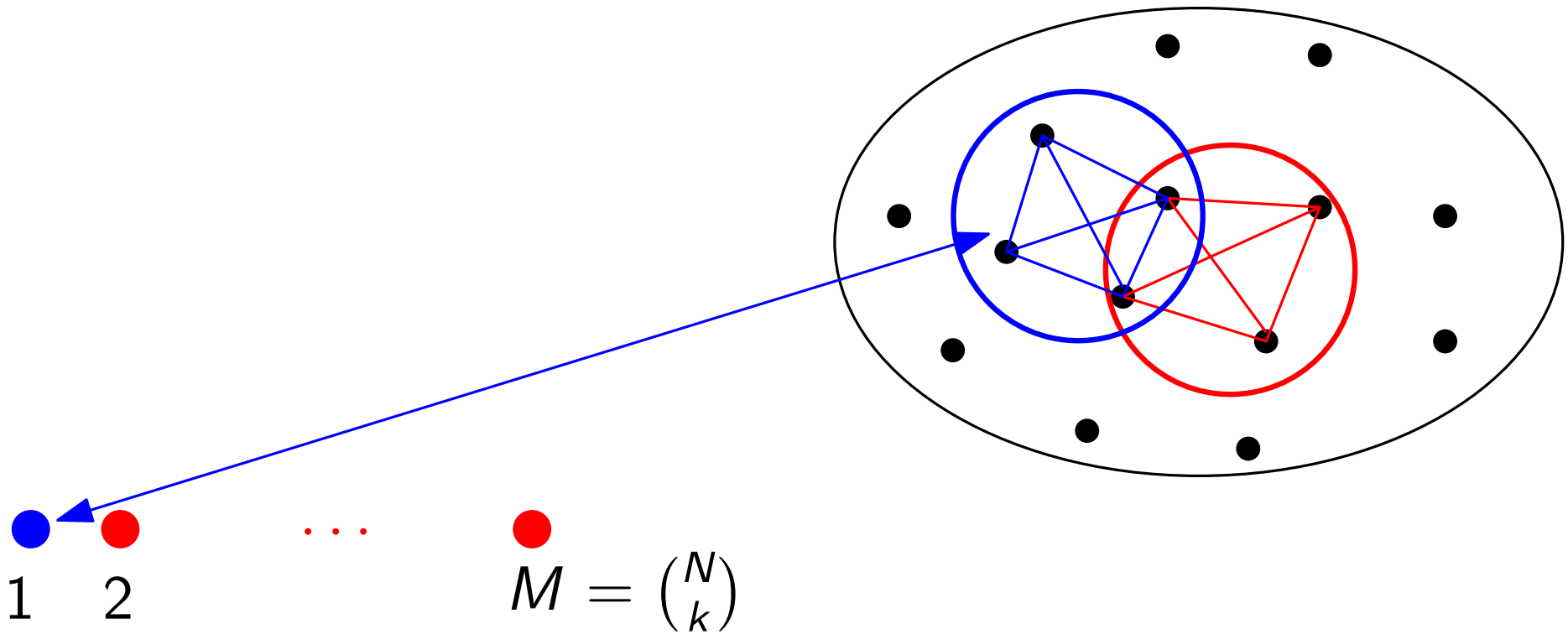
"Simple vs Hard"



P-REM

Random Graph

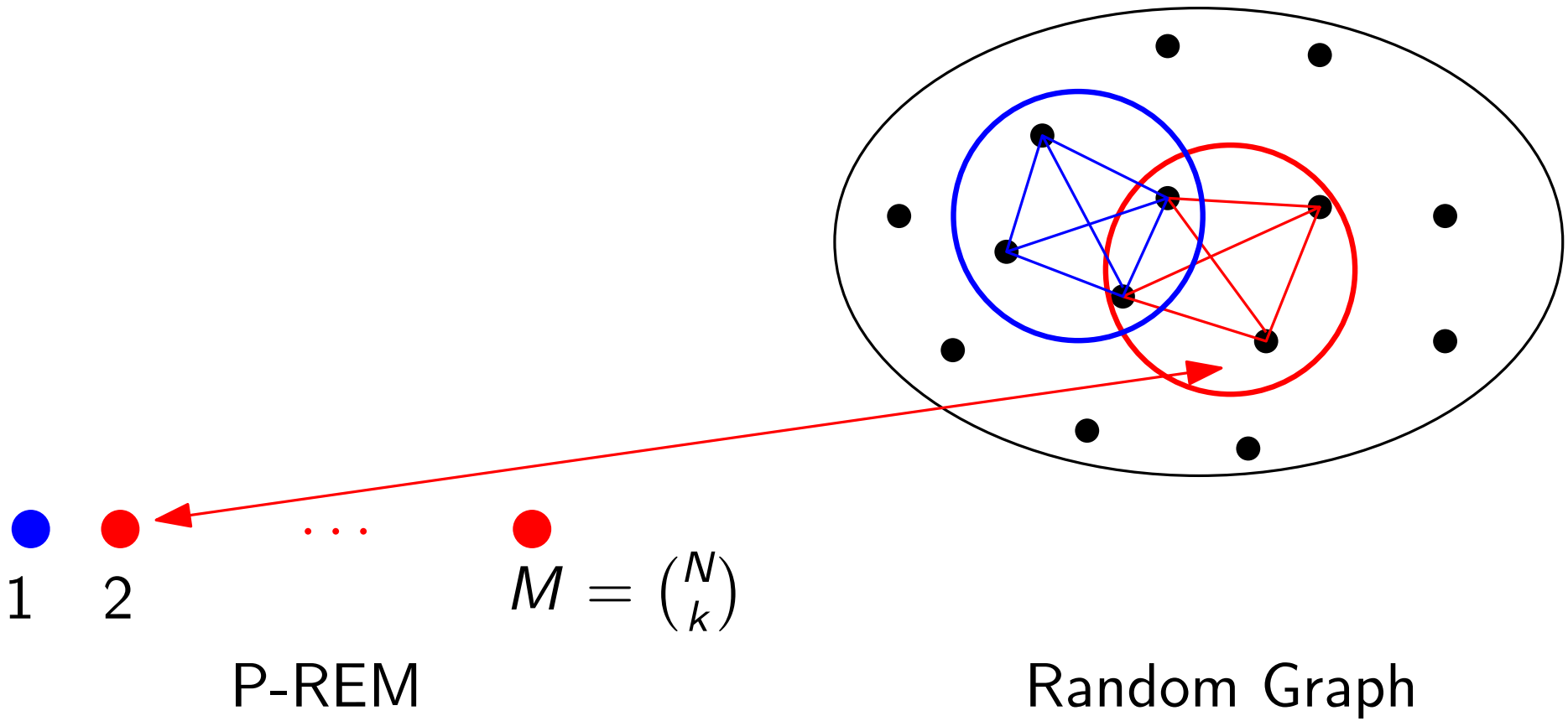
"Simple vs Hard"



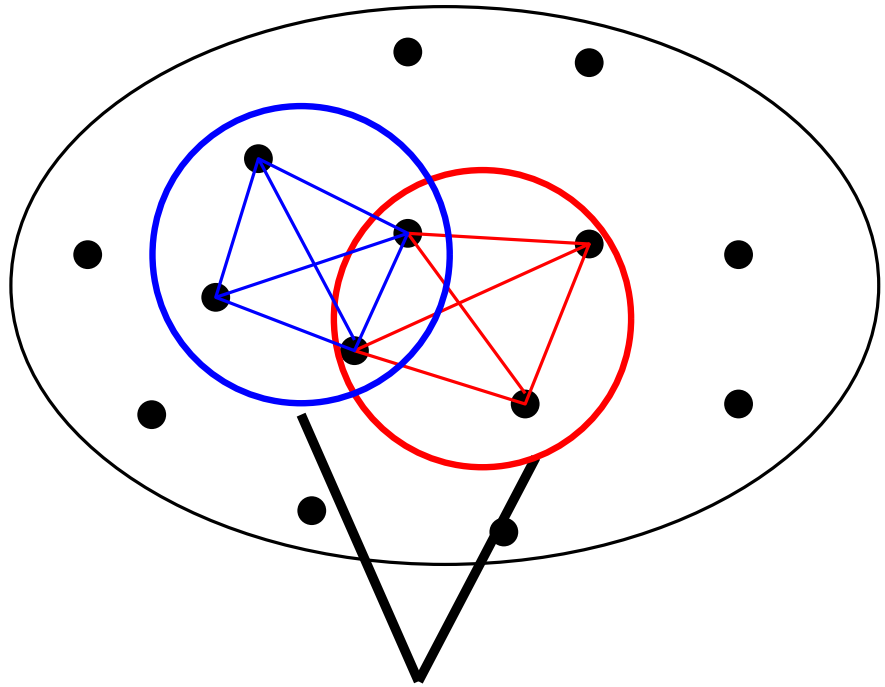
P-REM

Random Graph

"Simple vs Hard"



"Simple vs Hard"



independent

dependent

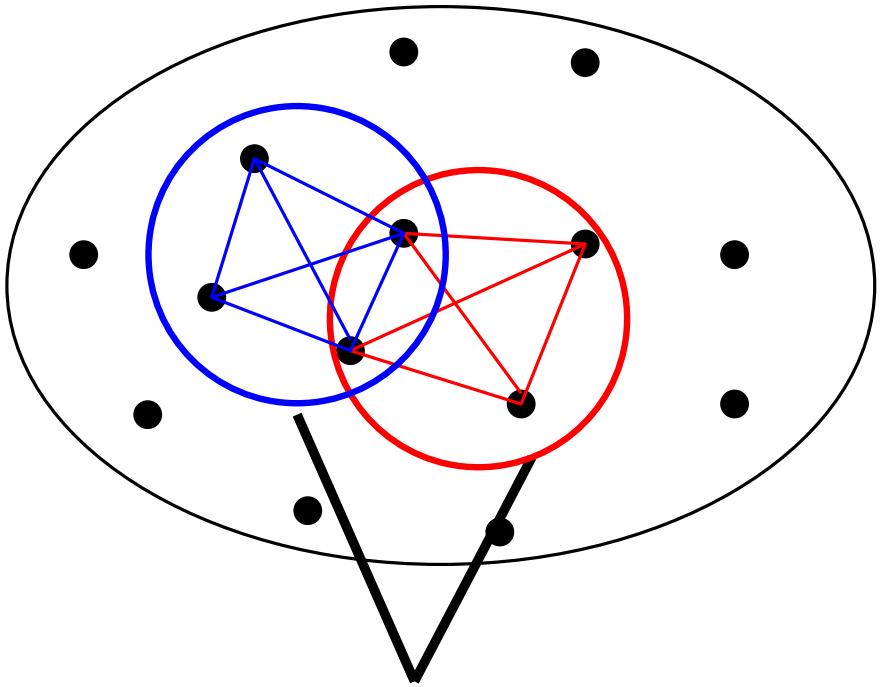
P-REM

Random Graph

"Simple vs Hard"

P-REM

search is hard

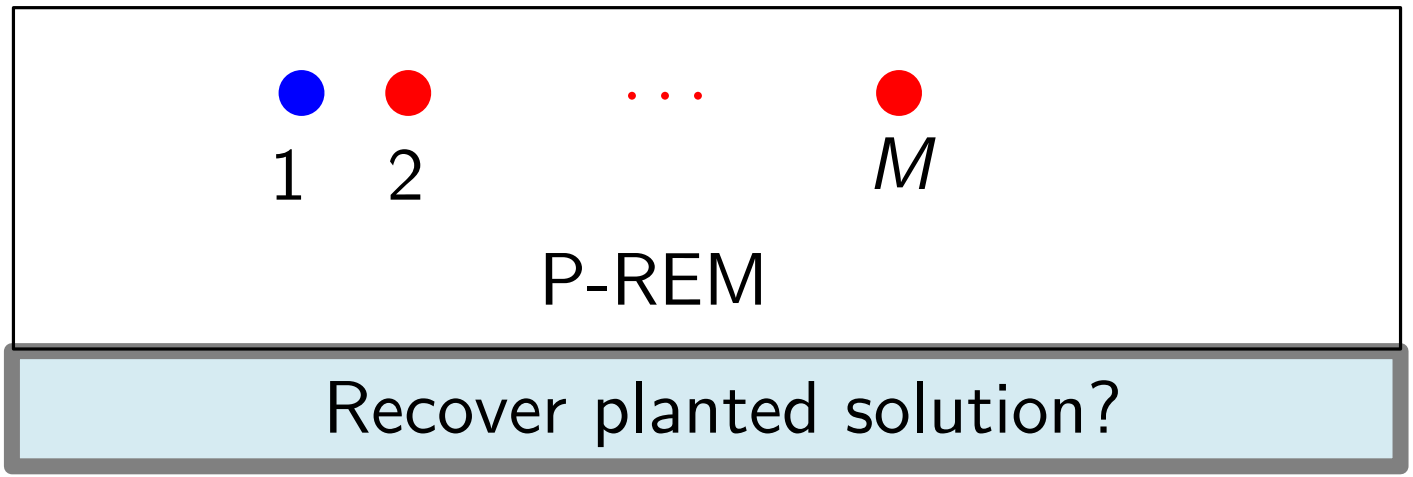


dependent

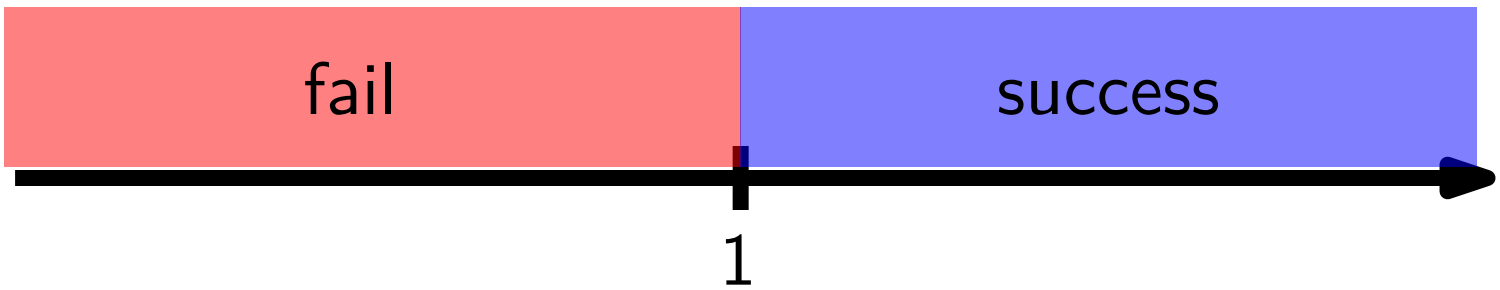
Random Graph

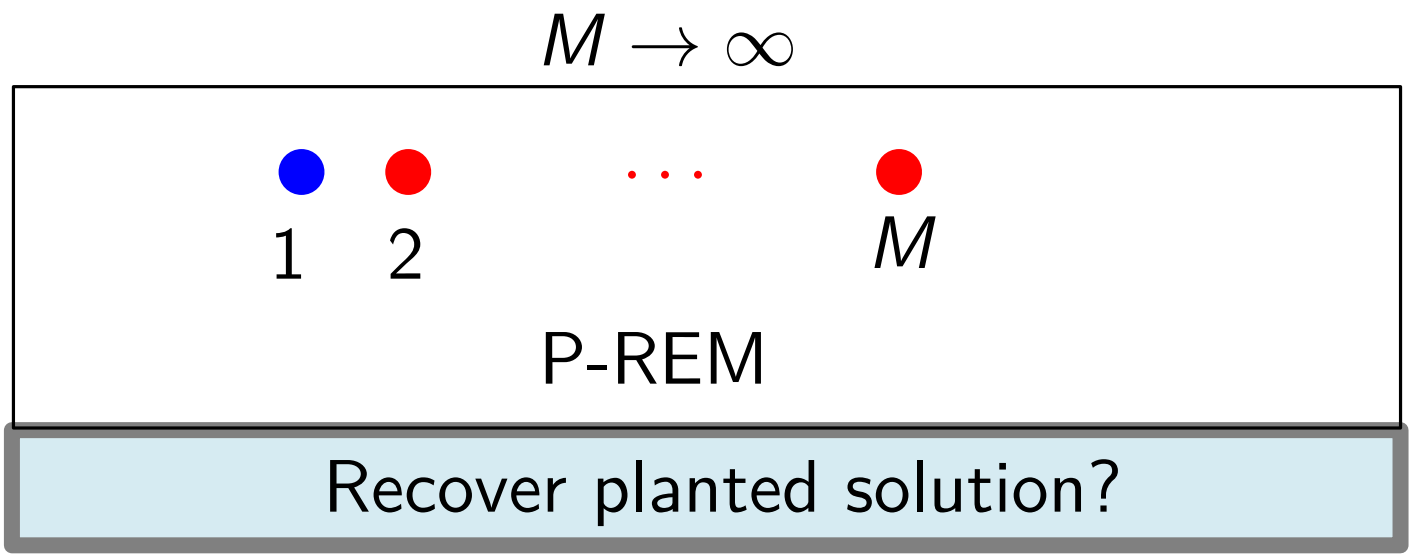
search maybe easier

Our Contribution

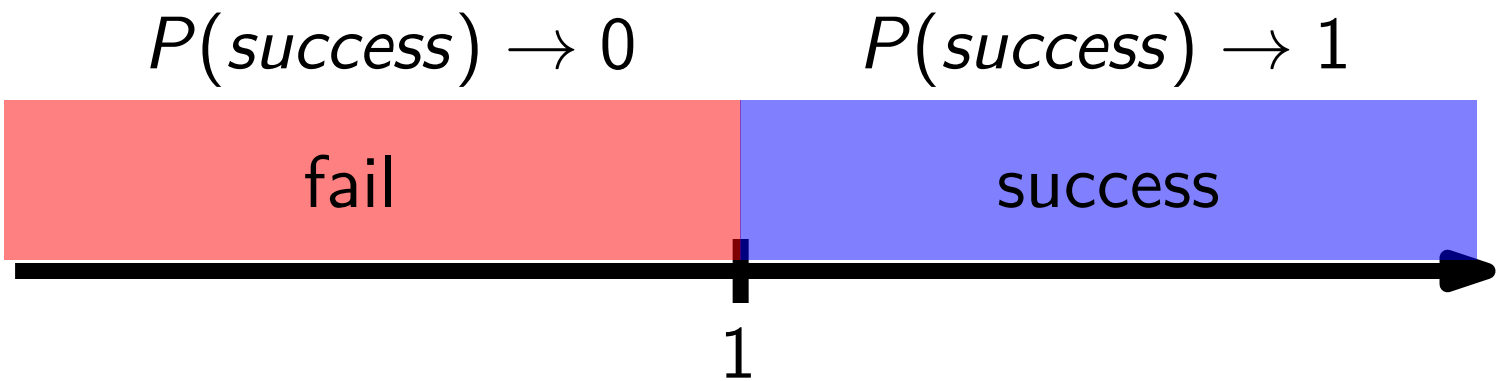


Signal to Noise Ratio $= \frac{\widehat{\mu}}{\widehat{\sigma}}$

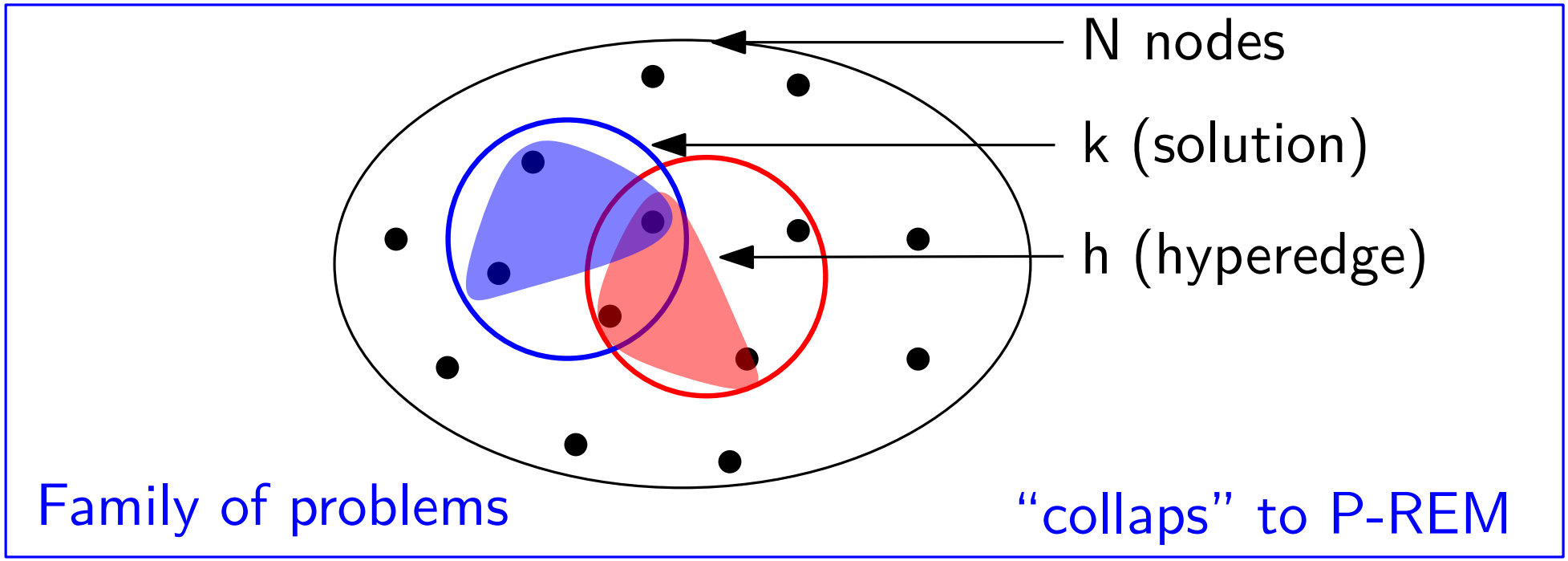




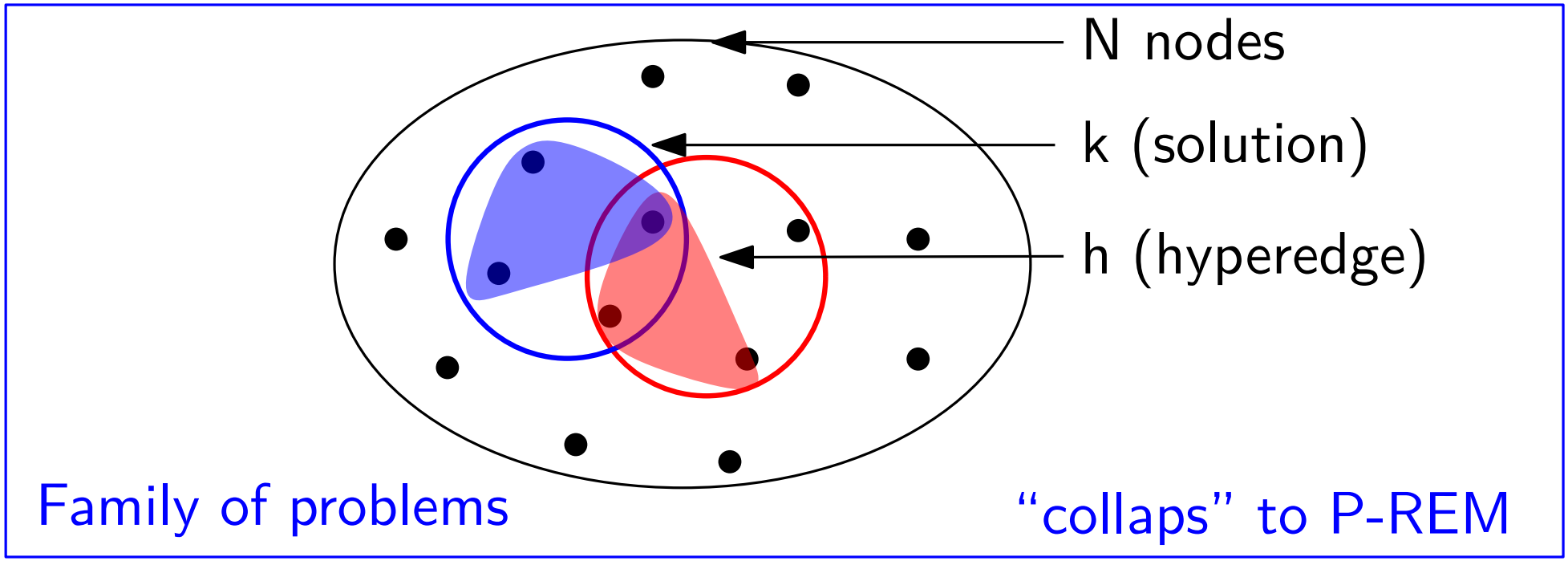
Signal to Noise Ratio $= \frac{\widehat{\mu}}{\widehat{\sigma}}$

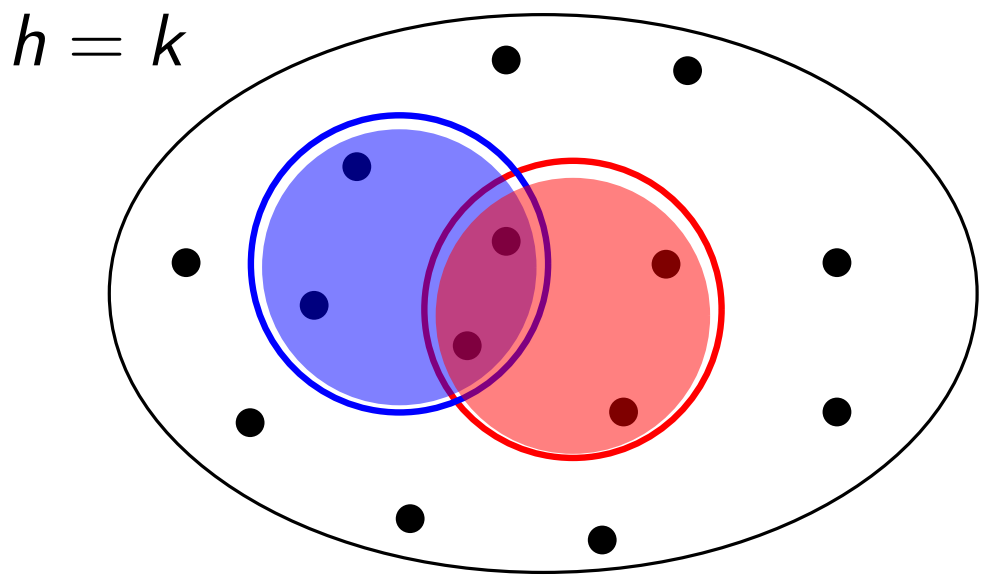


Overview of Contribution

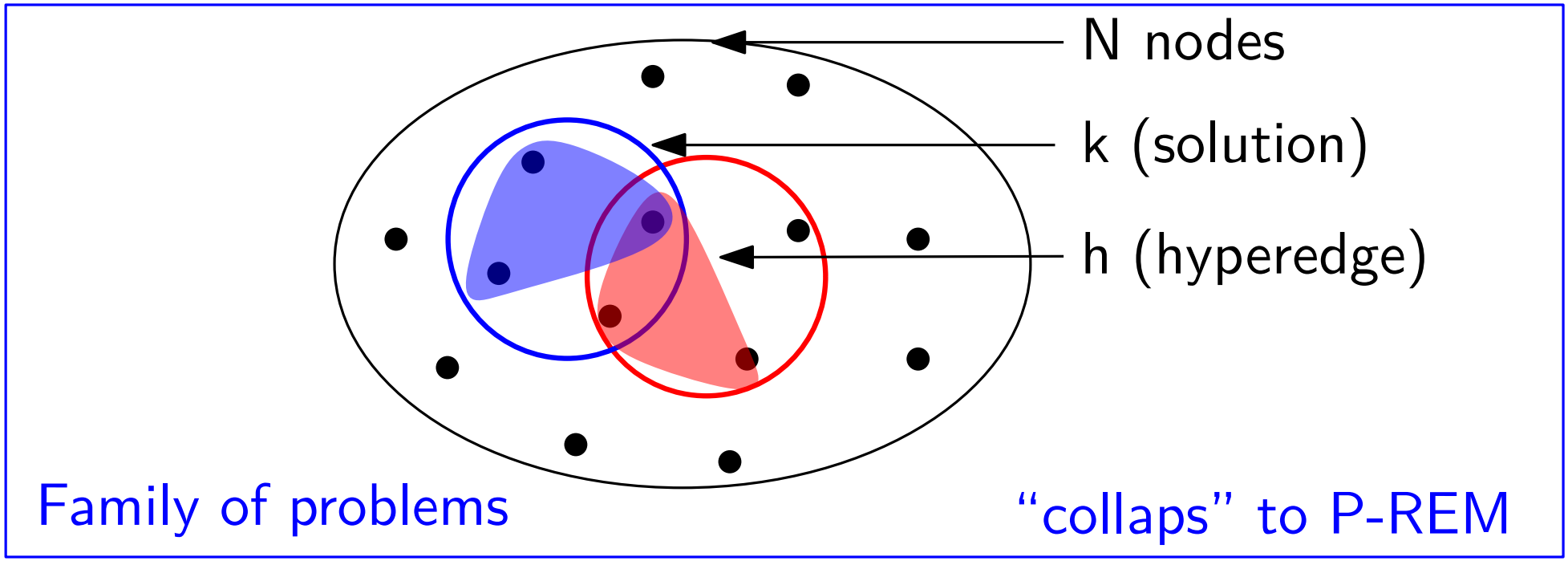


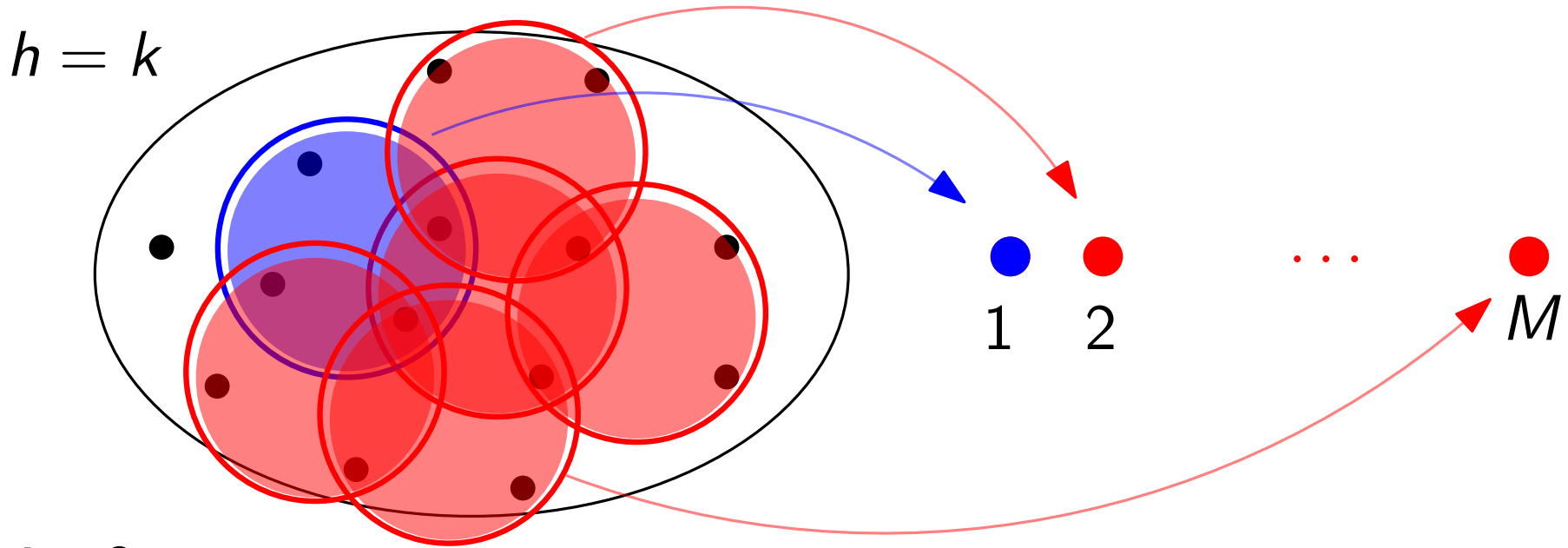
Overview of Contribution



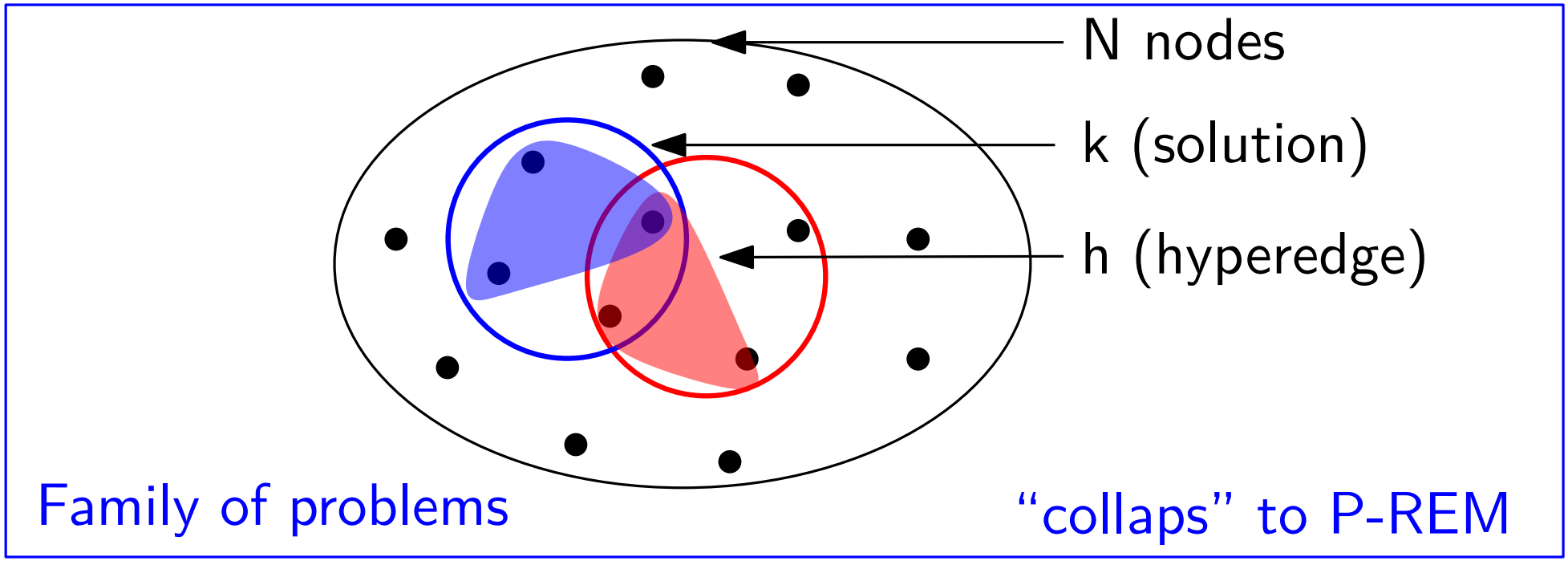


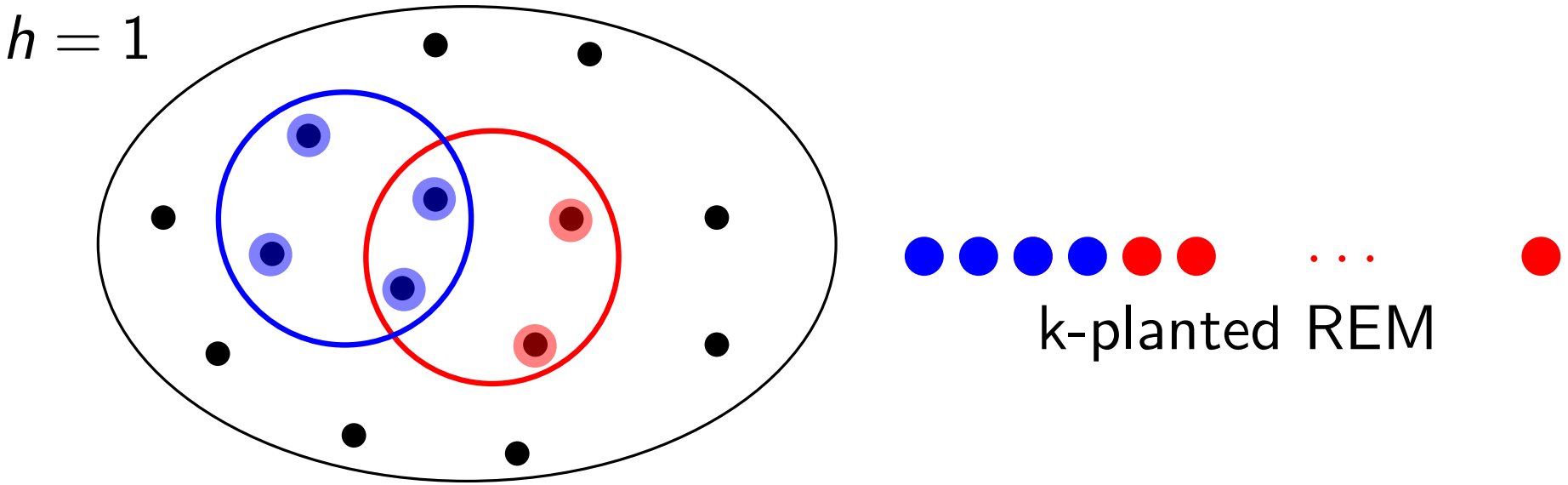
Overview of Contribution



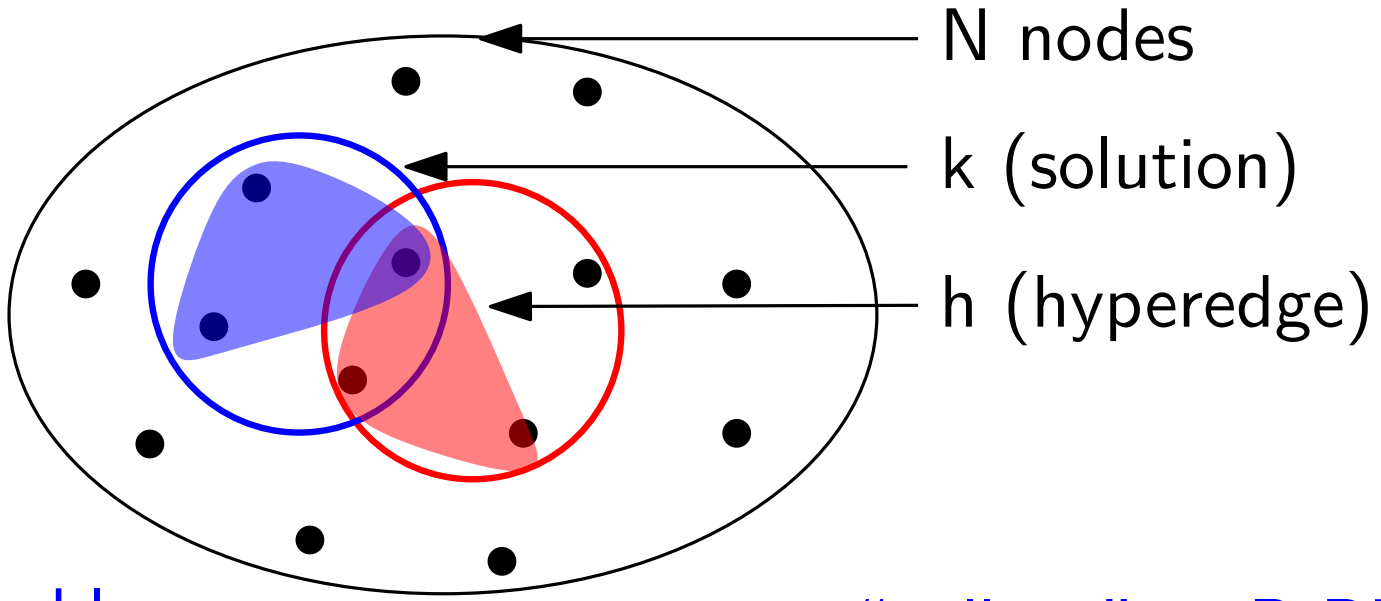


Overview of Contribution



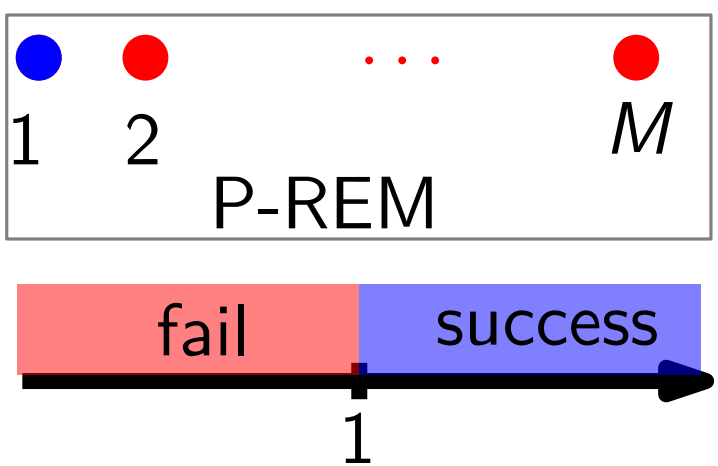


Overview of Contribution

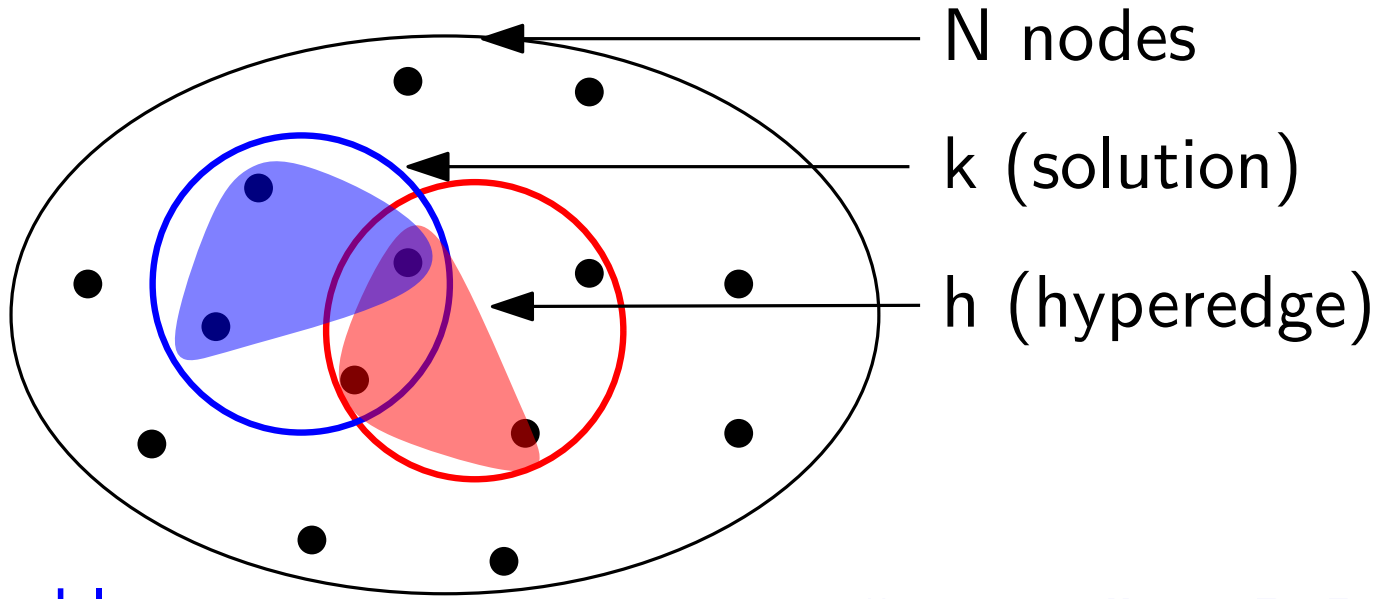


Family of problems

"collaps" to P-REM

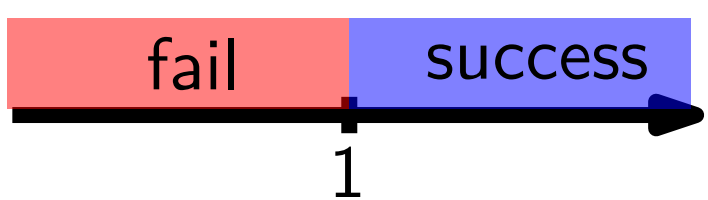


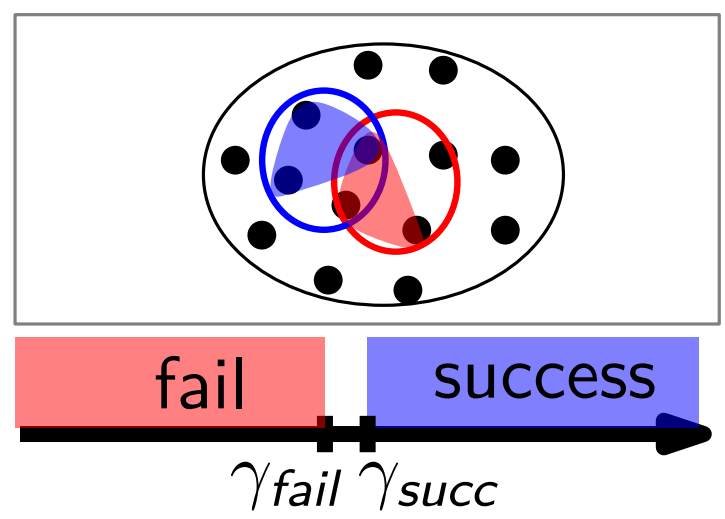
Overview of Contribution



Family of problems

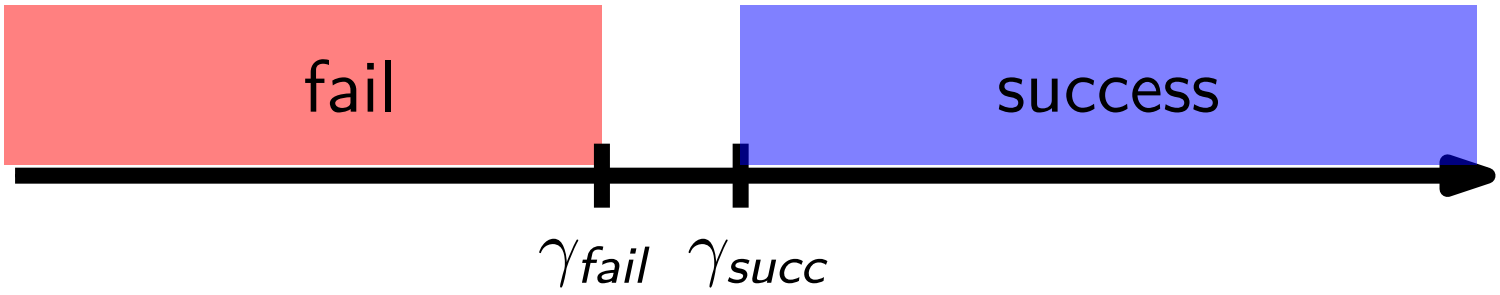
"collaps" to P-REM





Technique ("reduction")

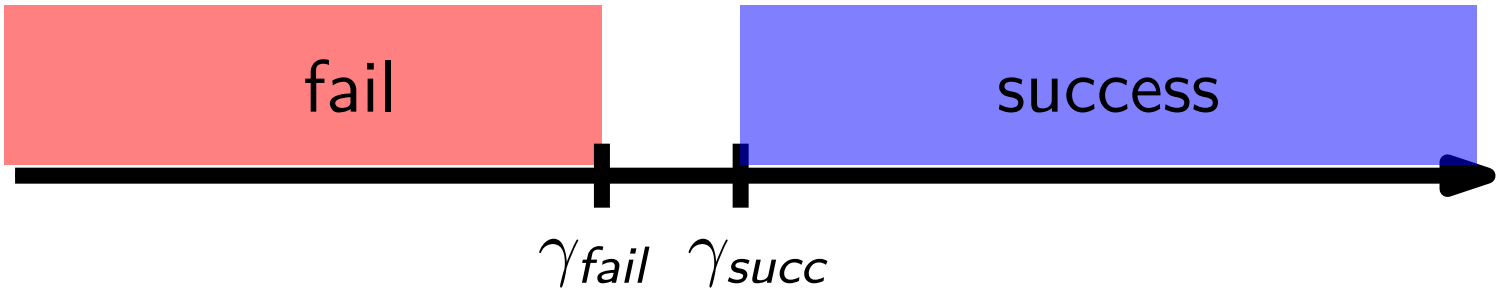
Bounds



h	Model	γ_{fail}	γ_{succ}
1	k-P-REM	1	1
2	Graph	$\sqrt{\frac{1}{k-1}}$	$2\sqrt{\frac{1}{k-1}}$
$2 < h < k$	Hypergraph	$\sqrt{\frac{1}{\binom{k-1}{h-1}}}$	$2\sqrt{\frac{h}{\binom{k-1}{h-1}}}$
k	P-REM	1	1

$$k = o(\log N)$$

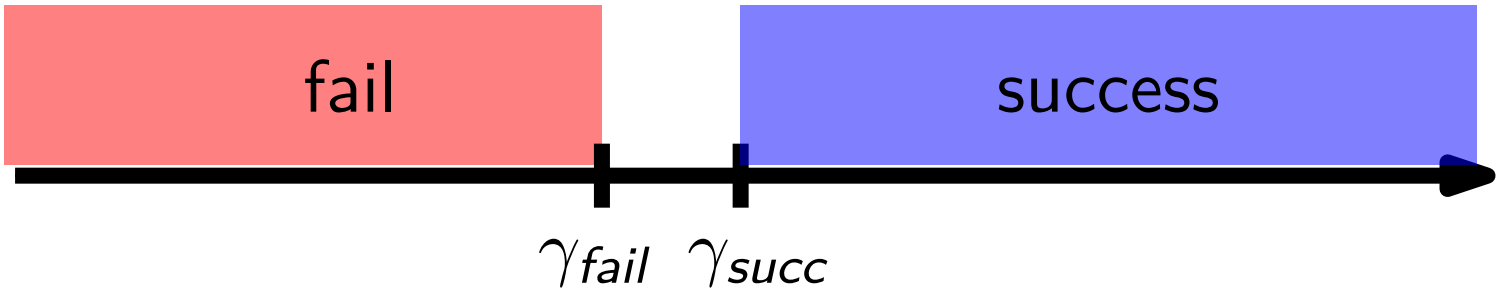
Bounds



h	Model	γ_{fail}	γ_{succ}
1	k-P-REM	1	1
2	Graph	$\sqrt{\frac{1}{k-1}}$	$2\sqrt{\frac{1}{k-1}}$
$2 < h < k$	Hypergraph	$\sqrt{\frac{1}{\binom{k-1}{h-1}}}$	$2\sqrt{\frac{h}{\binom{k-1}{h-1}}}$
k	P-REM	1	1

$k = o(\log N)$

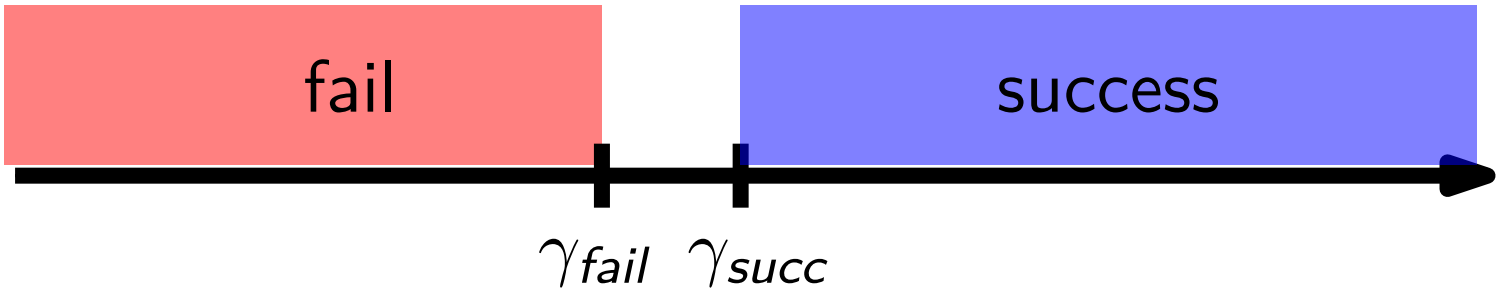
Bounds



h	Model	γ_{fail}	γ_{succ}
1	k-P-REM	1	1
2	Graph	$\sqrt{\frac{1}{k-1}}$	$2\sqrt{\frac{1}{k-1}}$
$2 < h < k$	Hypergraph	$\sqrt{\frac{1}{\binom{k-1}{h-1}}}$	$2\sqrt{\frac{h}{\binom{k-1}{h-1}}}$
k	P-REM	1	1

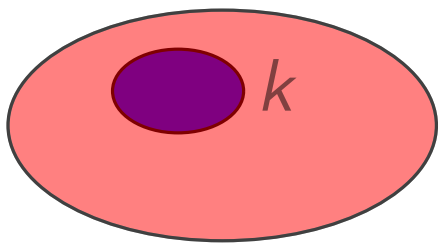
$k = o(\log N)$

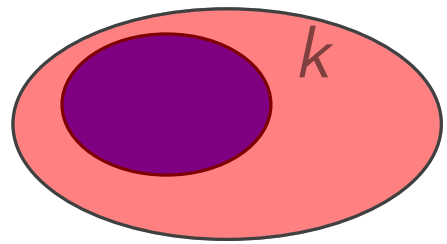
Bounds



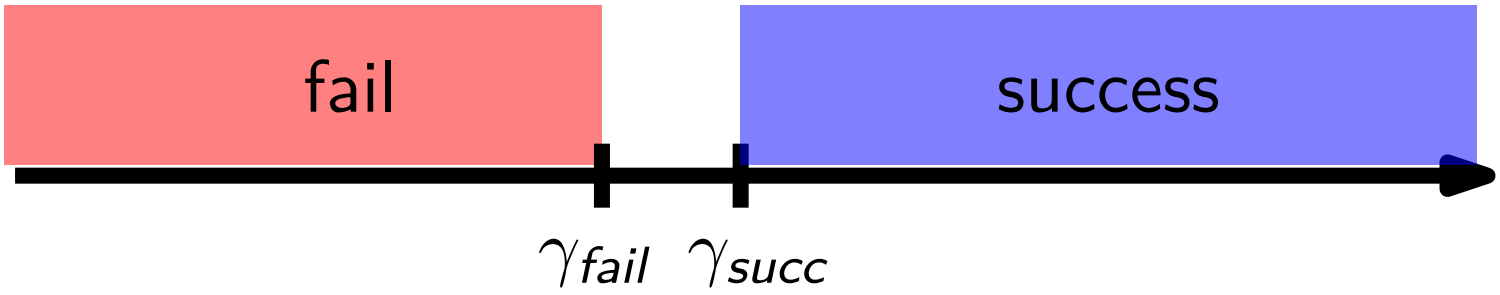
h	Model	γ_{fail}	γ_{succ}
1	k-P-REM	1	1
2	Graph	$\sqrt{\frac{1}{k-1}}$	$2\sqrt{\frac{1}{k-1}}$
$2 < h < k$	Hypergraph	$\sqrt{\frac{1}{\binom{k-1}{h-1}}}$	$2\sqrt{\frac{h}{\binom{k-1}{h-1}}}$
k	P-REM	1	1

$$k = o(\log N)$$





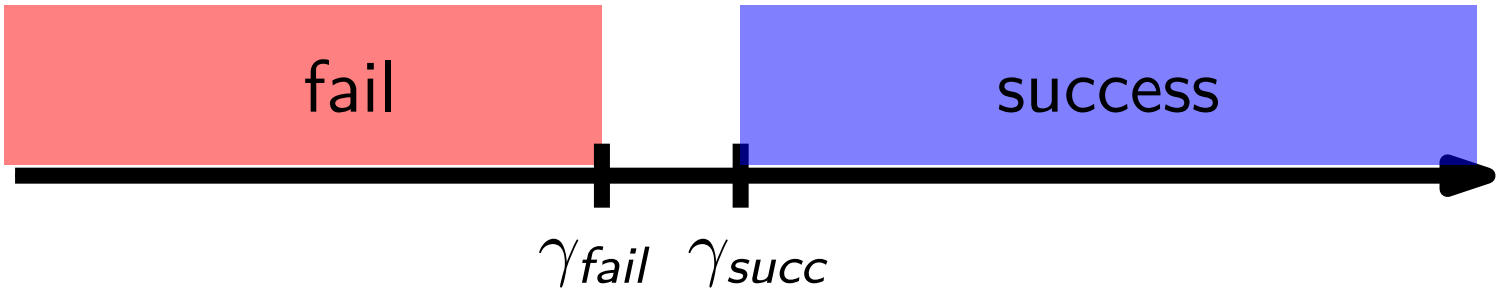
Bounds



h	Model	γ_{fail}	γ_{succ}
1	k-P-REM	1	1
2	Graph	$\sqrt{\frac{1}{k-1}}$	$2\sqrt{\frac{1}{k-1}}$
$2 < h < k$	Hypergraph	$\sqrt{\frac{1}{\binom{k-1}{h-1}}}$	$2\sqrt{\frac{h}{\binom{k-1}{h-1}}}$
k	P-REM	1	1

$$k = o(\log N)$$

Bounds



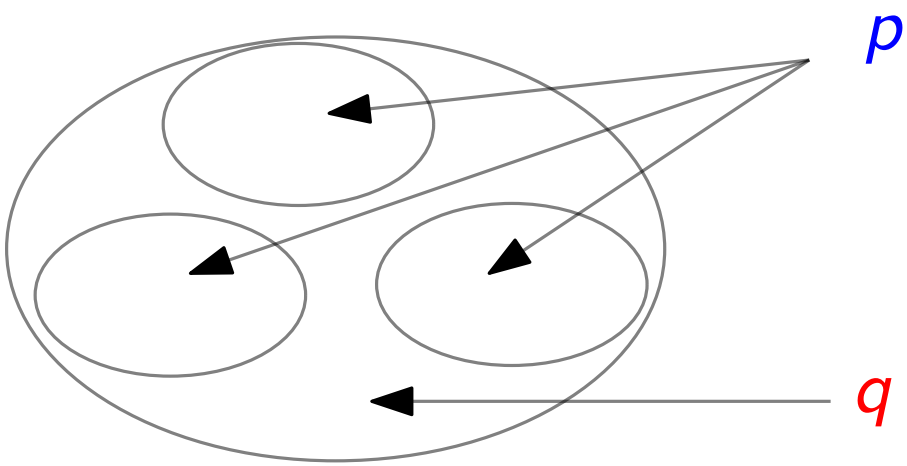
h	Model	γ_{fail}	γ_{succ}
1	k-P-REM	1	1
2	Graph	$\sqrt{\frac{1}{k-1}}$	$2\sqrt{\frac{1}{k-1}}$
$2 < h < k$	Hypergraph	$\sqrt{\frac{1}{\binom{k-1}{h-1}}}$	$2\sqrt{\frac{h}{\binom{k-1}{h-1}}}$
k	P-REM	1	1

$k = o(\log N)$

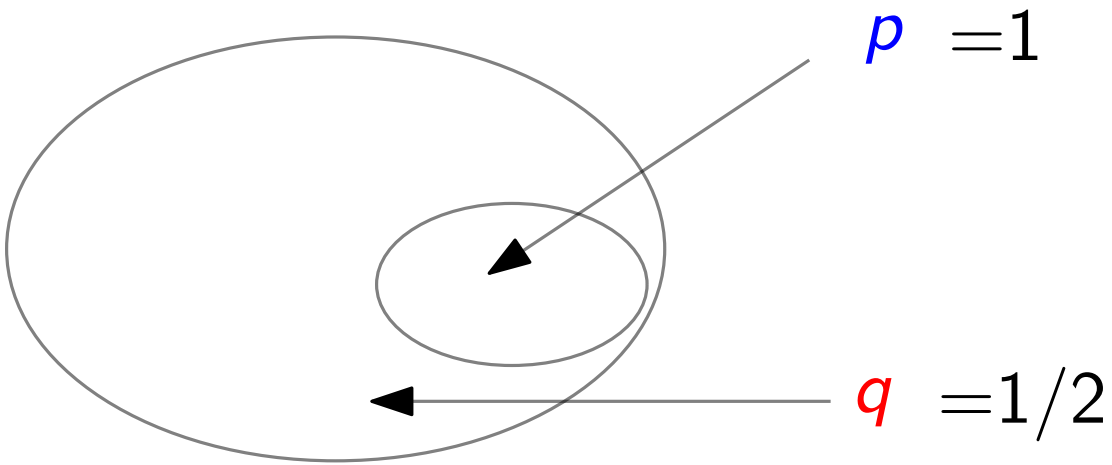
maximum likelihood (ML)

Recover planted solution?

Connections to Other Works



Connections to Other Works



Bernulli

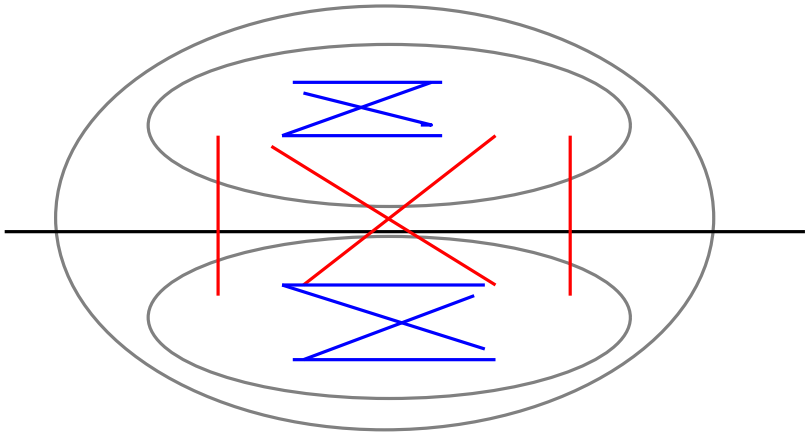
Planted Clique

$k = \Theta(\log N)$ $\qquad$ $k = \sqrt{N}$

A nearly tight sum-of-squares lower bound for the planted clique problem (Barak et al., FOCS'16).

$\vdots$

Connections to Other Works



p

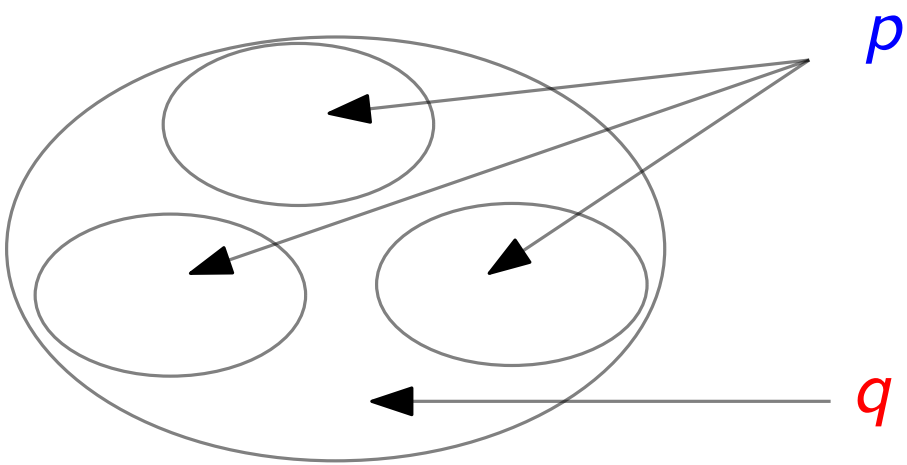
q

Bernulli

$k = N/2$

Bisection

Connections to Other Works



Bernulli

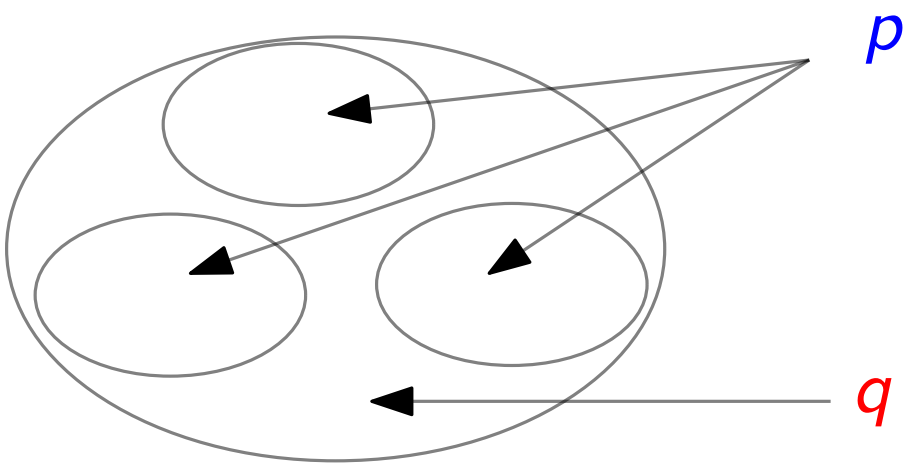
Stochastic Block Model

"Exact recovery in the stochastic block model" (Abbe, , Bandeira, and Hall, IEEE Transactions on Information Theory '16) ,

"Consistency thresholds for the planted bisection model" (Mossel, Neeman, and Sly, STOC 15)

"Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery," (E. Abbe and C. Sandon, FOCS '15)

Connections to Other Works



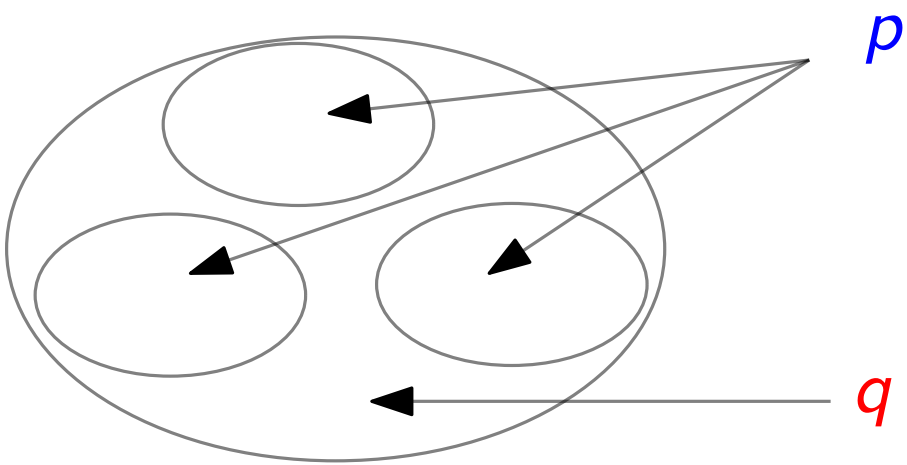
Weighted Stochastic Block Model

Generic distr.

$$k = N/c$$

"Information-theoretic bounds for exact recovery in weighted stochastic block models using the renyi divergence" (Jog and Loh, arXiv 2015)

Connections to Other Works

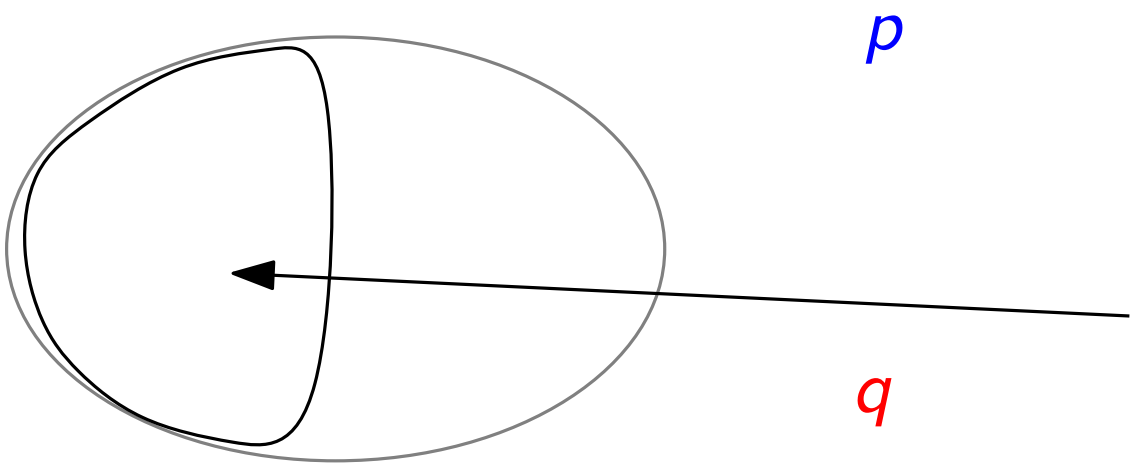


Bernulli

Hypergraph
Stochastic Block Model

"Consistency of spectral partitioning of uniform hypergraphs under planted partition model"
(Ghoshdastidar and Dukkipati, NIPS '14)
"Consistency of spectral hypergraph partitioning under planted partition model,"
(Ghoshdastidar et al, The Annals of Statistics '17)

Connections to Other Works



Gaussian

$k = N/2$

Weighted Hypergraph
Stochastic Block Model

"Community detection in hypergraphs, spiked tensor models, and Sum-of-Squares"
(Chiheon Bandeira, Goemans, Int. Conf. on Sampling Theory and Applications, '17)

Computational Aspects

- trade off (dependency, recoverability, hardness)

Other Problems

- our technique ("reduce" to REM)

Thank You!!