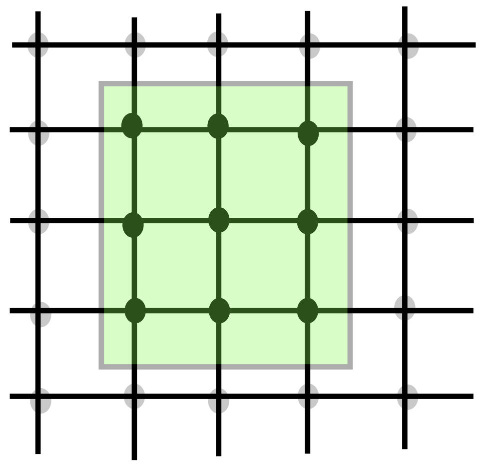
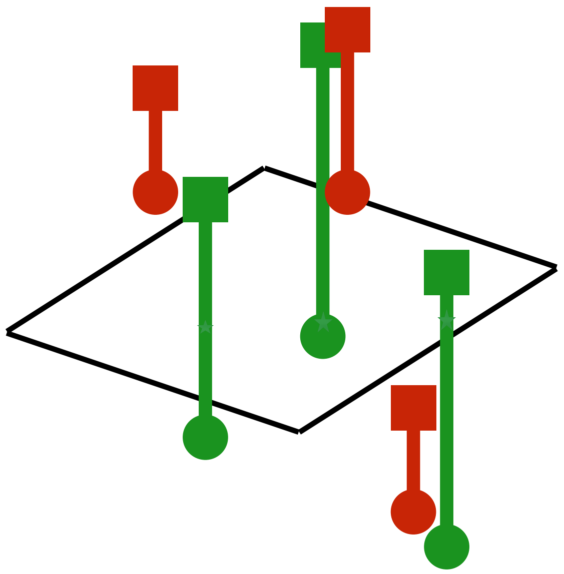
Spatial Stochastic Models for Network Analysis

Abishek Sankararaman

PhD Defense





Committee

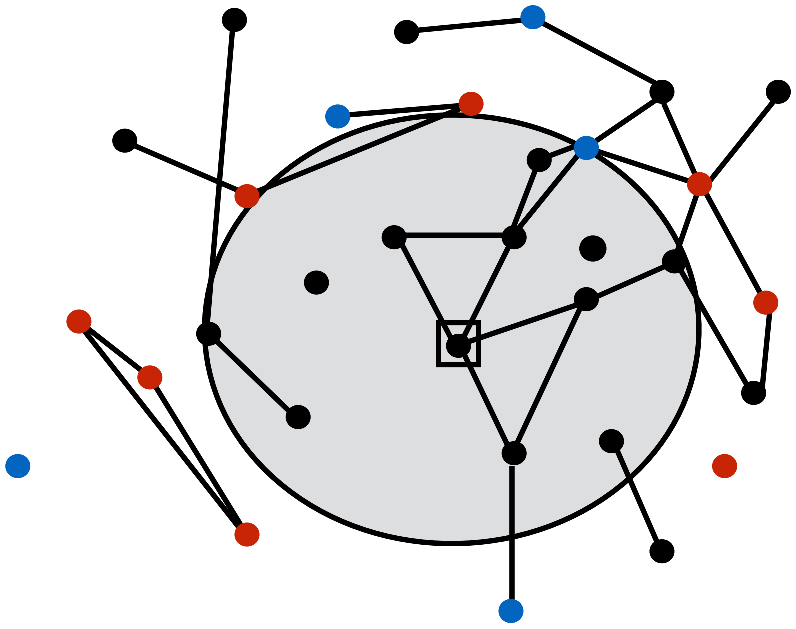
François Baccelli (Advisor)

Gustavo deVeciana

Sanjay Shakkottai

Alex Dimakis

Joe Neeman

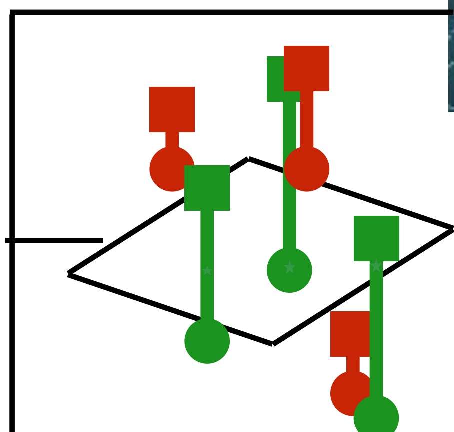


Introduction

Emerging trends in networking bring about new design challenges

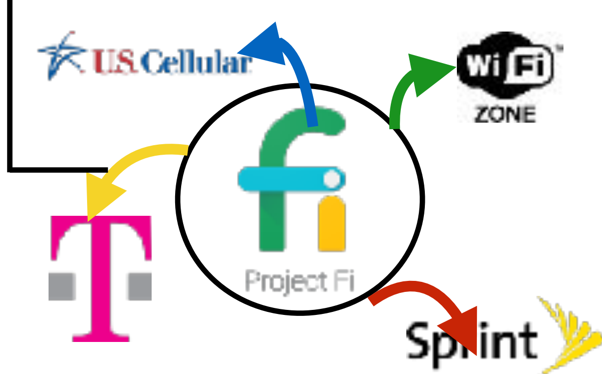
Large scale wireless networks

Data Networks



Dynamics

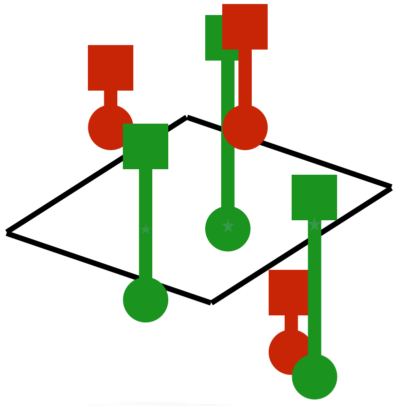
Community Detection



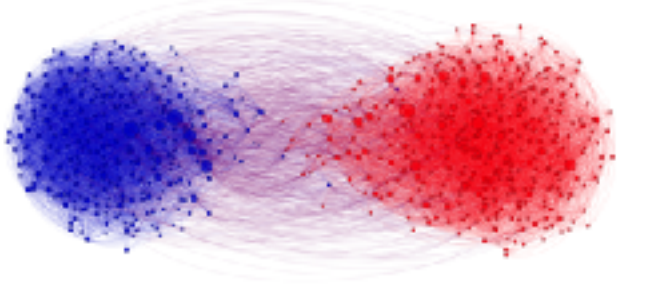
Multiple Operators Diversity

Contents of the Thesis

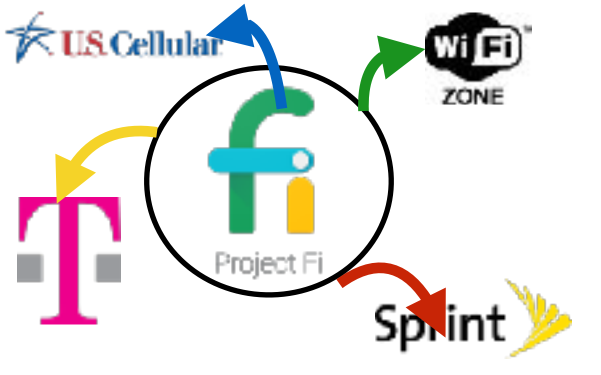
1. Dynamics on Wireless Networks



2. Graph Clustering

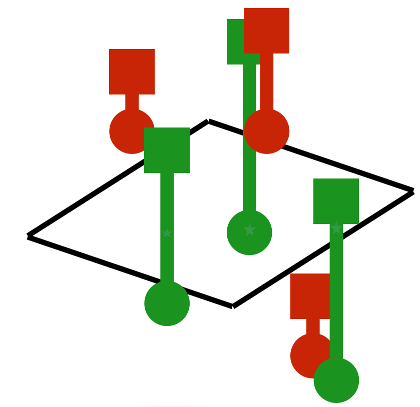


3. Diversity in multi-operator cellular networks



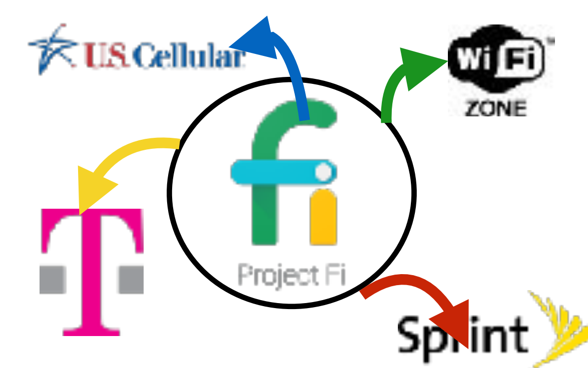
Contents of the Thesis

1. Dynamics on Wireless Networks



2. Graph Clustering

3. Diversity in multi-operator cellular networks

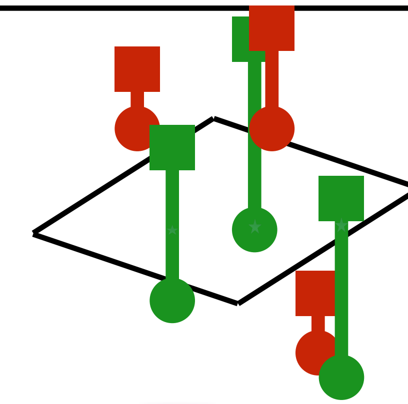


Theme -

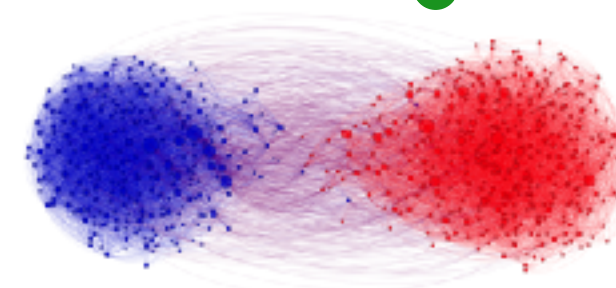
- Questions of interest
- Tractable Models
- Insights

Contents of the Thesis

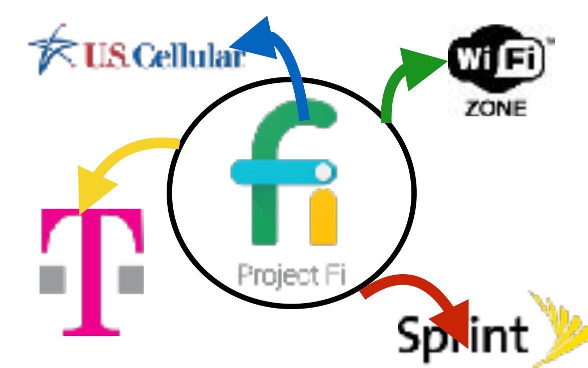
1. Dynamics on Wireless Networks



2. Graph Clustering



3. Diversity in multi-operator cellular networks



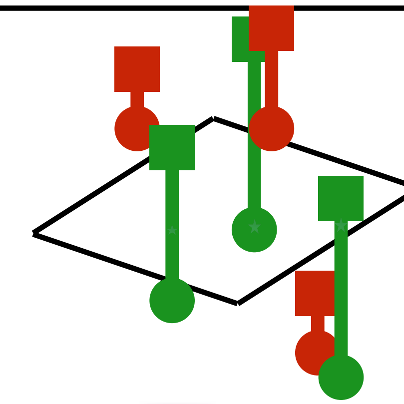
<u>Theme</u> -

- Questions of interest
- Tractable Models
- Insights

Contents of the Talk

1. Dynamics on Wireless Networks

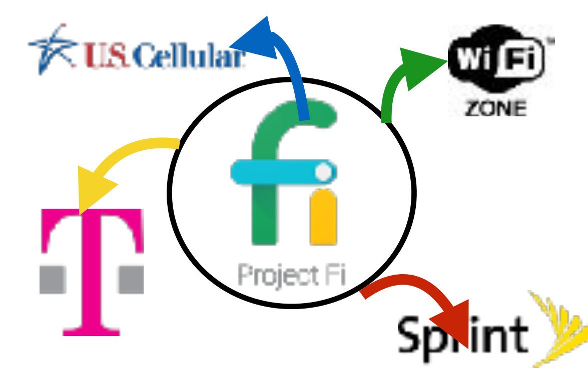
   Proposal - A Spatial Birth-Death Wireless Network Model
   Today - Interference Queueing Networks



2. Graph Clustering

   Proposal - Motivation and Broad Introduction
   Today - Community Detection on Euclidean Random Graphs

3. Diversity in multi-operator cellular networks



Theme -

- Questions of interest
- Tractable Models
- Insights

Community Detection

A.S,François Baccelli, *Community Detection on Euclidean Random Graphs,* In ACM-SIAM Symposium of Discrete Algorithms (SODA) 2018

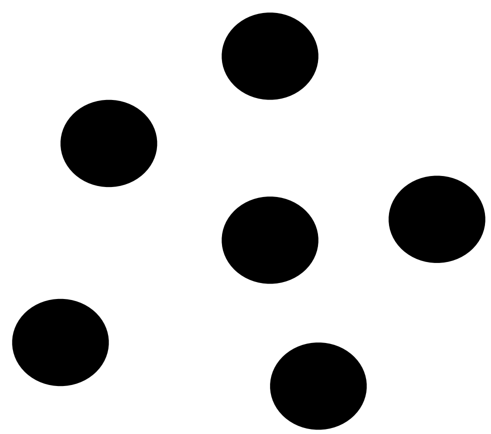
A.S, Emmanuel Abbe, François Baccelli, *Community Detection on Euclidean Random Graphs,* Full Version. Under Review at IMA Information and Inference

https://arxiv.org/abs/1706.09942

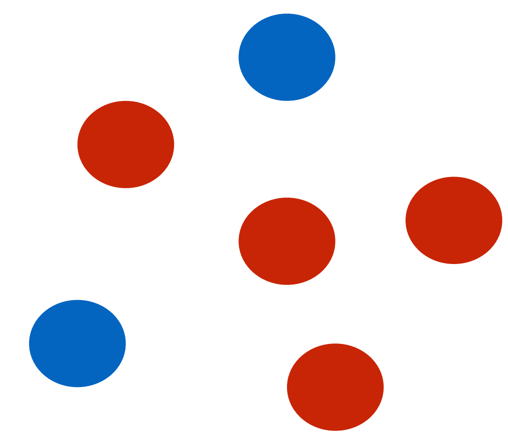
Community Detection - Abstract Definition

- Grouping objects given ***indirect information*** of memberships

A population



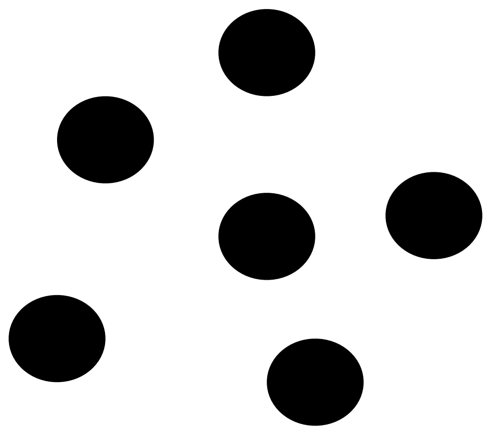
partitioned into groups



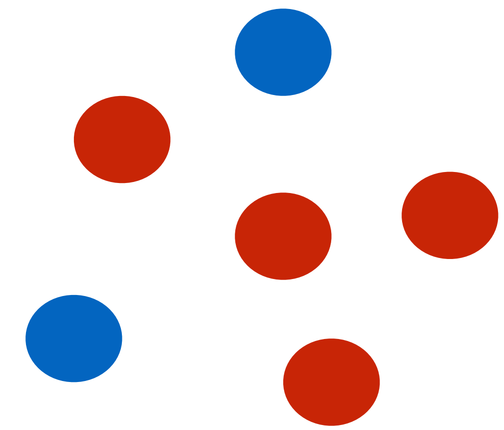
Community Detection - Examples

- Grouping objects given ***indirect information*** of memberships

A population



partitioned into groups

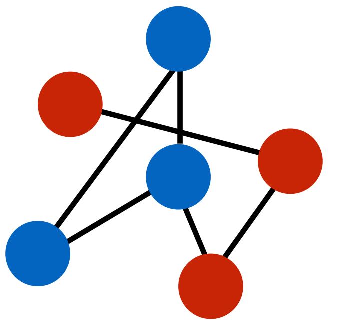


1. People on an Online Social Network

2. Proteins classified into groups based on their functional behavior
3. Grouping Base-Stations based on similarities in traffic pattern

Graph as Information

Important sub-class



Population - Represented as nodes of a graph

Membership Information - Encoded as labeled edges of the graph

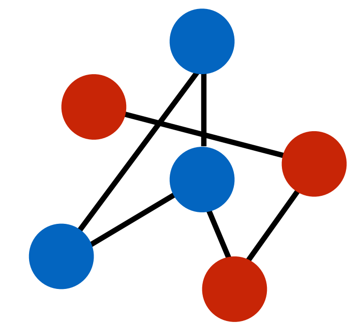
Graph Clustering Problem -

Given an unlabeled graph data, recover the partition of nodes

Graph Clustering

Fundamental theoretical problem

Statistics, CS, Physics, Information Theory, Mathematics

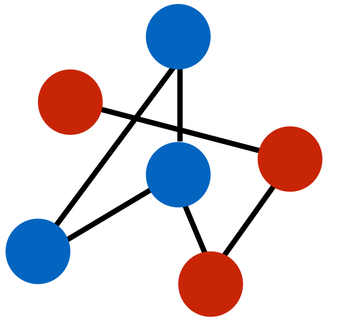


Well established applications-

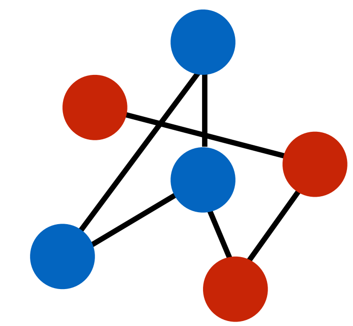
- Social Networks *(Targeted Advertising)*
- Recommendation Systems *(Users and tems)*
  [Linden,Smith and York ‘03][Sahebi and Cohen ‘11]
- Genomics *(Similar genes)*
  [Jiang, Tang and Zhang ‘04]

Graph Clustering

Given an unlabeled graph data, recover the partition of nodes



Graph Clustering



Given an unlabeled graph data, recover the partition of nodes

What if there are additional contextual information on each node ?

Web-pages, the textual content in a page

Social Networks - Personal information (age, location, income….)

Computational Biology - Metadata generated by measurements

Outline

1) Model - *The Planted Partition Random Connection Model*

2) Algorithm

3) Mathematical Results

4) Application - Haplotype Phasing

Planted Partition Random Connection Model

Planted Partition Random Connection Model

1) $N_n \sim \mathrm{Poisson}(\lambda n)$ number of nodes

On avg λ points per unit area.

Planted Partition Random Connection Model

1) $N_n \sim \mathrm{Poisson}(\lambda n)$ number of nodes

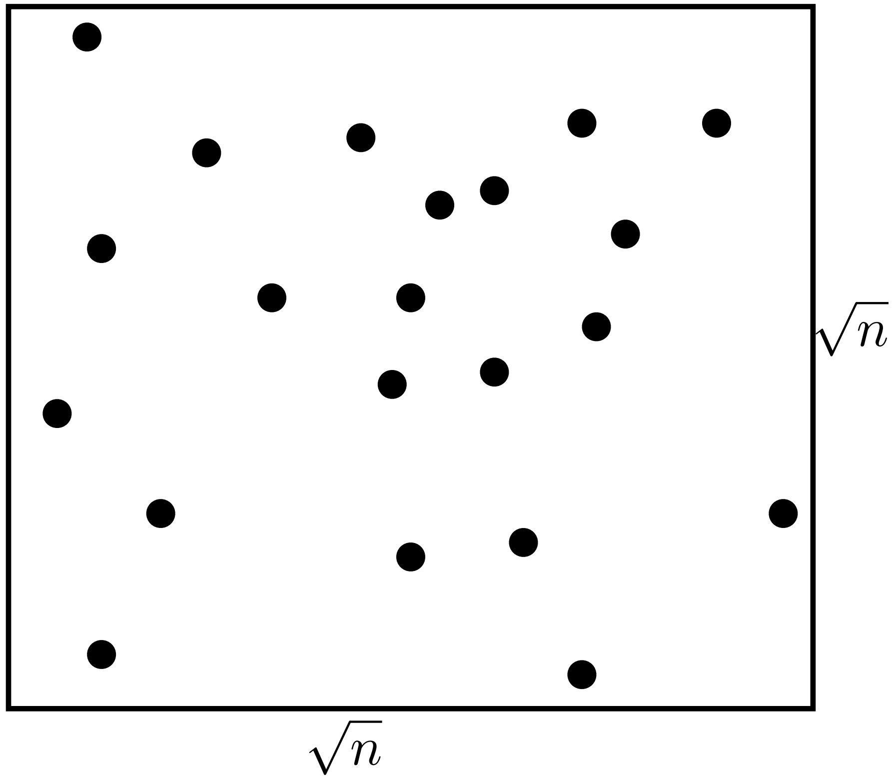
On avg λ points per unit area.

2) Each node $i \in [1, N_n]$, has a

- ***Location label*** $X_i \in \left[-\frac{n^{1/d}}{2}, \frac{n^{1/d}}{2}\right]$

sampled independently and uniformly

Planted Partition Random Connection Model



1) $N_n \sim \mathrm{Poisson}(\lambda n)$ number of nodes

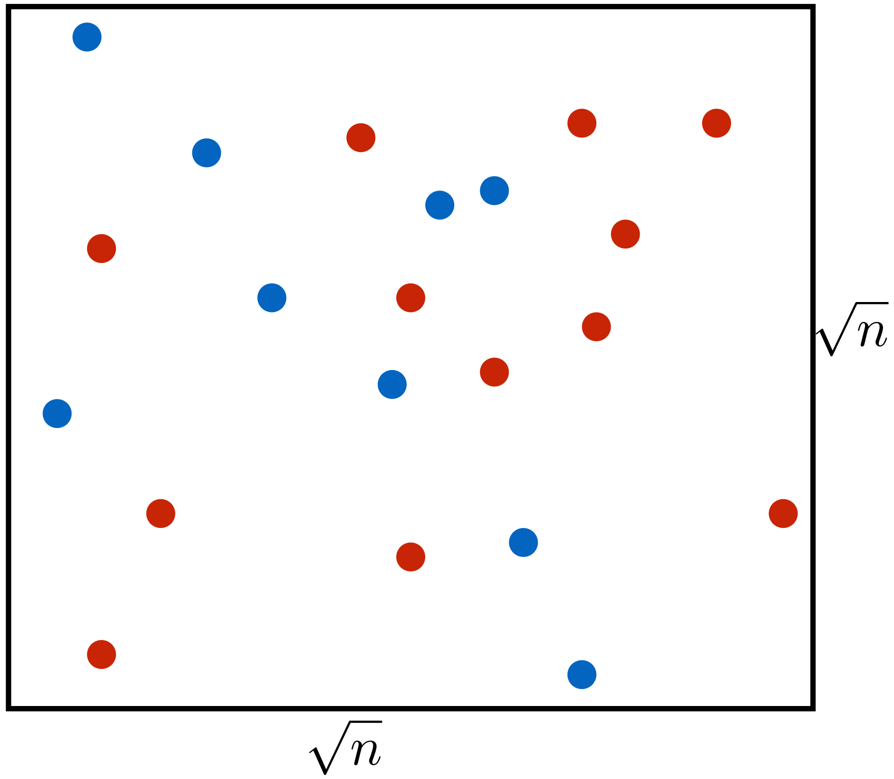
On avg λ points per unit area.

2) Each node $i \in [1, N_n]$, has a

- ***Location label*** $X_i \in \left[-\frac{n^{1/d}}{2}, \frac{n^{1/d}}{2}\right]$

sampled independently and uniformly

Planted Partition Random Connection Model



1) $N_n \sim \mathrm{Poisson}(\lambda n)$ number of nodes

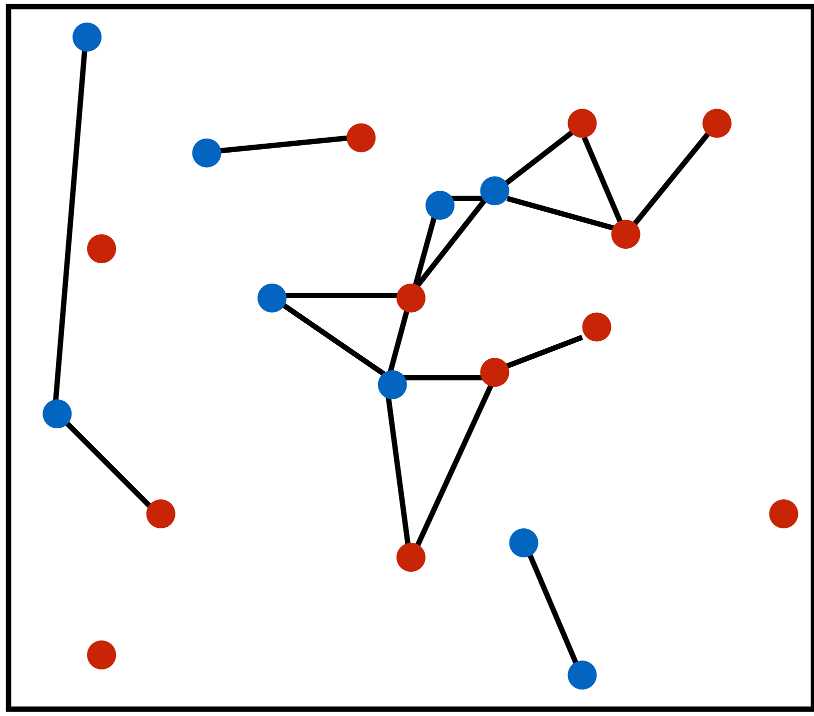
On avg λ points per unit area.

2) Each node $i \in [1, N_n]$, has a

- ***Location label*** $X_i \in \left[-\frac{n^{1/d}}{2}, \frac{n^{1/d}}{2}\right]$
- ***Community label*** $Z_i \in \{-1, +1\}$

sampled independently and uniformly

Planted Partition Random Connection Model



1) $N_n \sim \text{Poisson}(\lambda n)$ number of nodes

On avg λ points per unit area.

2) Each node $i \in [1, N_n]$, has a

- ***Location label*** $X_i \in \left[-\frac{n^{1/d}}{2}, \frac{n^{1/d}}{2}\right]$
- ***Community label*** $Z_i \in \{-1, +1\}$

sampled independently and uniformly

3) Edge between $i, j \in [1, N_n]$ with probability either

$f_{in}(||X_i - X_j||)$ - If $Z_i = Z_j$ (***same colors***)

$f_{out}(||X_i - X_j||)$ - If $Z_i \neq Z_j$ (***different colors***)

$\forall r \geq 0, f_{in}(r) \geq f_{out}(r)$

More edges within communities than across.

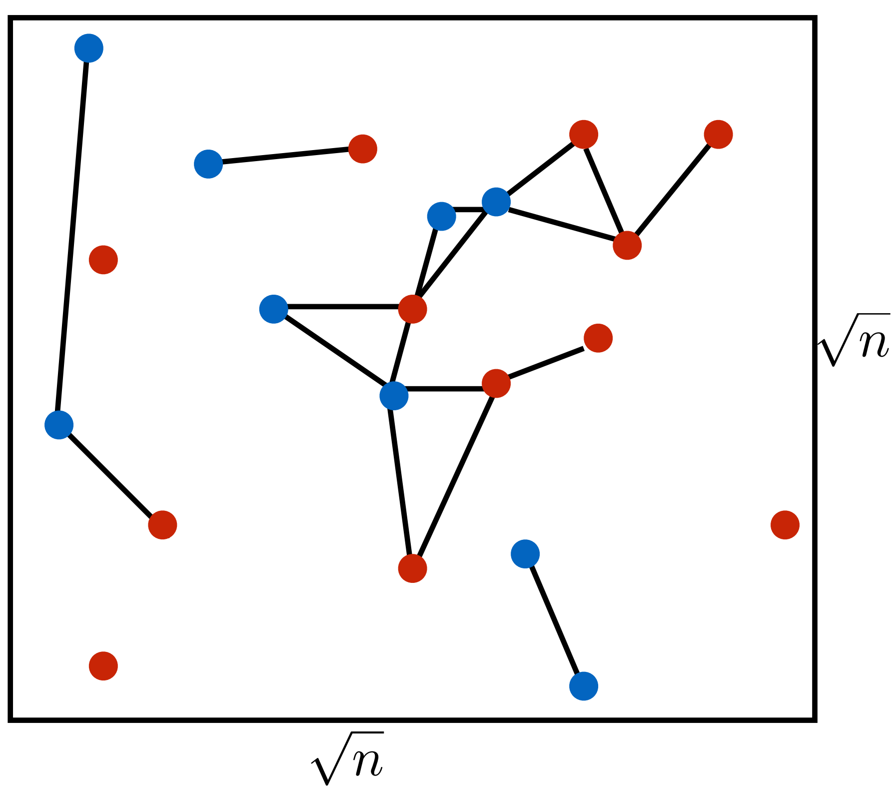
Conditional on node labels, edges are independent

Planted Partition Random Connection Model

1) $\{X_i\}_{i \in \mathbb{N}}$ - a ***Poisson Point Process*** on $\mathbb{R}^d$ with intensity λ

2) Independently ***mark*** it $\{Z_i\}_{i \in \mathbb{N}}$ each of which is uniform over $\{-1, 1\}$

3) Connect any two nodes $i \neq j \in \mathbb{N}$ with probability
$f_{in}(||X_i - X_j||)\mathbf{1}_{Z_i = Z_j} + f_{out}(||X_i - X_j||)\mathbf{1}_{Z_i \neq Z_j}$ independently for all pairs



$G_n \stackrel{d}{=} G$ restricted to $\left[-\frac{n^{1/d}}{2}, \frac{n^{1/d}}{2}\right]^d$

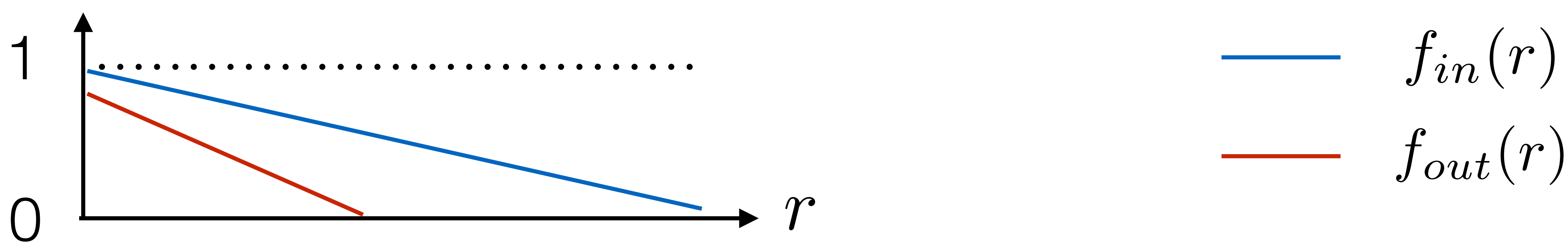
Planted Partition Random Connection Model

Model Parameters

$\lambda > 0$ Intensity

$d \geq 2$ Dimension of embedding

$$f_{in}(\cdot), f_{out}(\cdot) : \mathbb{R}_+ \to [0,1] \text{ s.t } \forall r \geq 0 \ , \ f_{in}(r) \geq f_{out}(r)$$

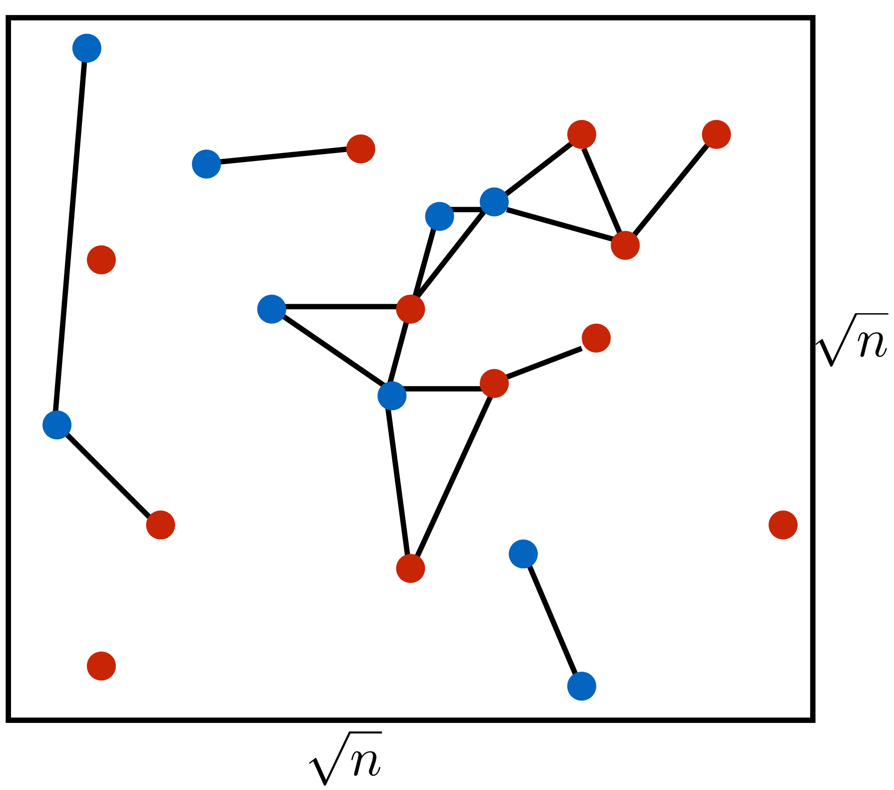


Planted Partition Random Connection Model

Assume $$\int_{x \in \mathbb{R}^d} f_{out}(||x||)dx \leq \int_{x \in \mathbb{R}^d} f_{in}(||x||)dx < \infty$$

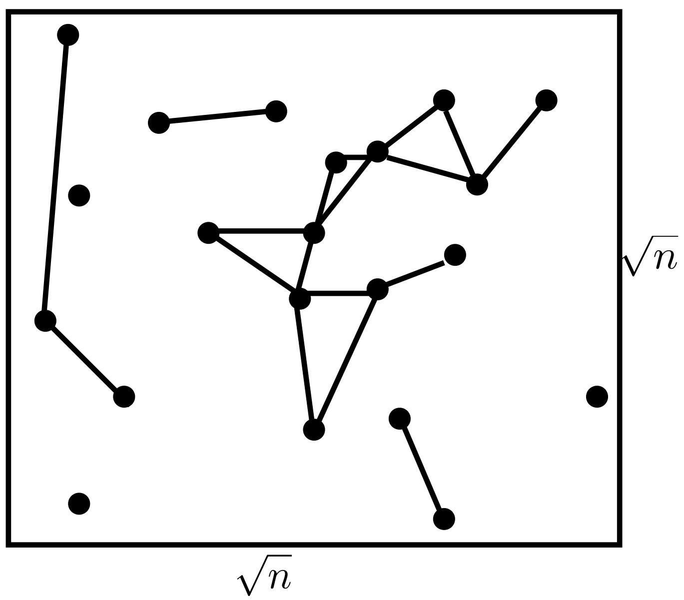
Avg # of neighbors in

- same community is - $(\lambda/2)\int_{x \in \mathbb{R}^d} f_{in}(||x||)dx - o(1)$
- opposite community is - $(\lambda/2)\int_{x \in \mathbb{R}^d} f_{out}(||x||)dx - o(1)$



Constant avg degree

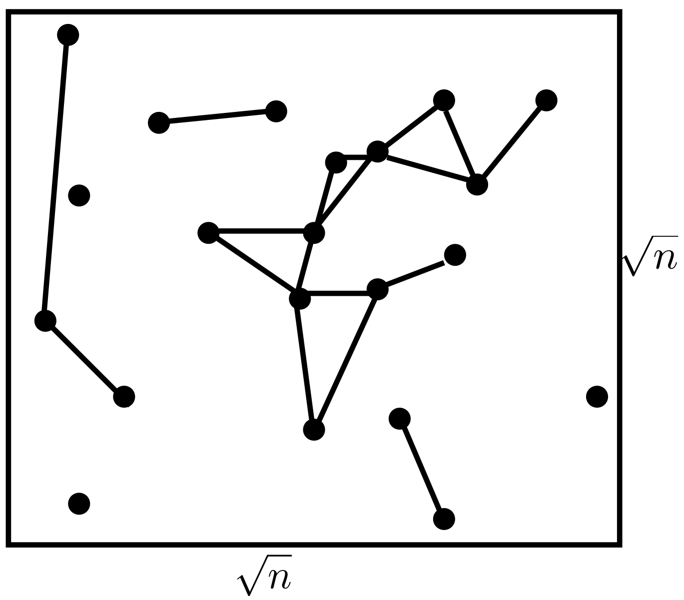
Community Detection Problem



Given G_n and $\{X_i\}_{i\in[0,N_n]}$, estimate $\{Z_i\}_{i\in[1,N_n]}$

$\{\tau_i\}_{i\in[0,N_n]}$- Community estimates

Community Detection Problem



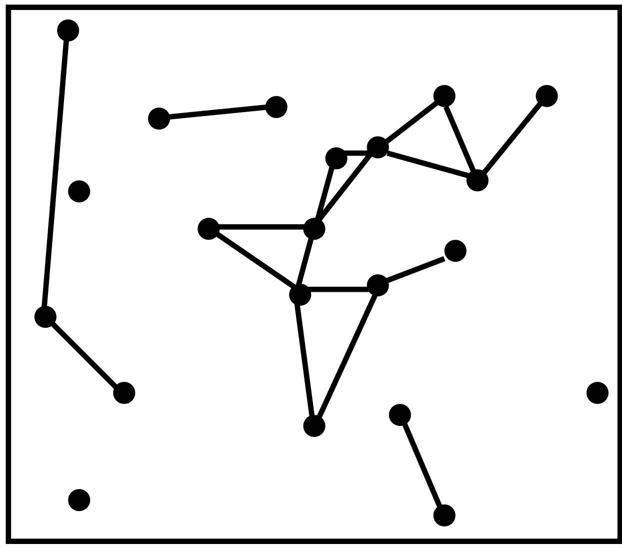
Given G_n and $\{X_i\}_{i\in[0,N_n]}$, estimate $\{Z_i\}_{i\in[1,N_n]}$

$\{\tau_i\}_{i\in[0,N_n]}$- Community estimates

$\mathcal{O}_\tau := \frac{1}{N_n}\left|\sum_{i=1}^{N_n} Z_i\tau_i\right|$ ***overlap*** of the estimator

$\mathcal{O}_\tau :=$| Fraction of correctly classified nodes - Fraction of incorrectly classified nodes |

Community Detection Problem



Given G_n and $\{X_i\}_{i \in [0, N_n]}$, estimate $\{Z_i\}_{i \in [1, N_n]}$

$\{\tau_i\}_{i \in [0, N_n]}$- Community estimates

$$\mathcal{O}_\tau := \quad \frac{1}{N_n} \left| \sum_{i=1}^{N_n} Z_i \tau_i \right|$$ ***overlap*** of the estimator

$\mathcal{O}_\tau :=$| Fraction of correctly classified nodes - Fraction of incorrectly classified nodes |

Community Detection is ***solvable*** if there exists an estimator $\{\tau_i\}_{i \in [0, N_n]}$ for every n, and some $\gamma > 0$ s.t. $\lim_{n \to \infty} \mathbb{P}[\mathcal{O}_\tau > \gamma] = 1$

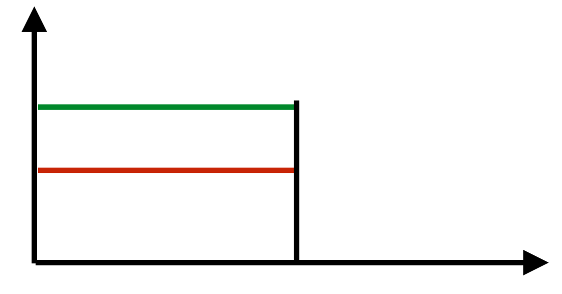
SLLN gives $\sum_{I=1}^{N_n} \frac{\tau_i Z_i}{N_n} \to 0$ *for blind guessing*

Solvability $\approx$ *asymptotically beating a random guess*

Community Detection Problem

Consider the example $f_{in}(r) = a\mathbf{1}_{r \leq R}$ $\qquad f_{out}(r) = b\mathbf{1}_{r \leq R}$

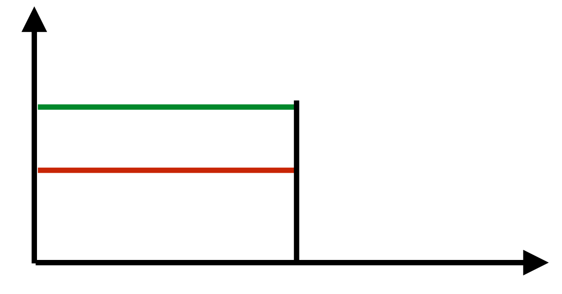
$$0 \leq b < a \leq 1$$

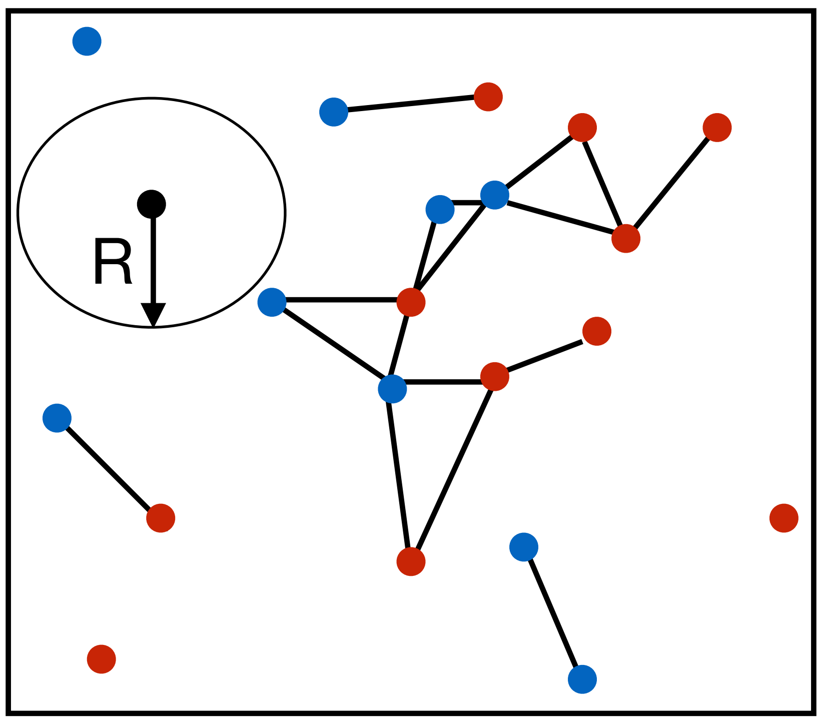


Community Detection Problem

Consider the example $f_{in}(r) = a\mathbf{1}_{r \leq R}$ $\qquad f_{out}(r) = b\mathbf{1}_{r \leq R}$

$$0 \leq b < a \leq 1$$





Isolated Nodes = No interaction with other points

Clearly $\mathcal{O}_\tau \leq 1 - e^{-\lambda \nu_d(1) R^d} < 1$

$\nu_d(1)$ Vol of unit ball in d dimensions

What was previously known ?

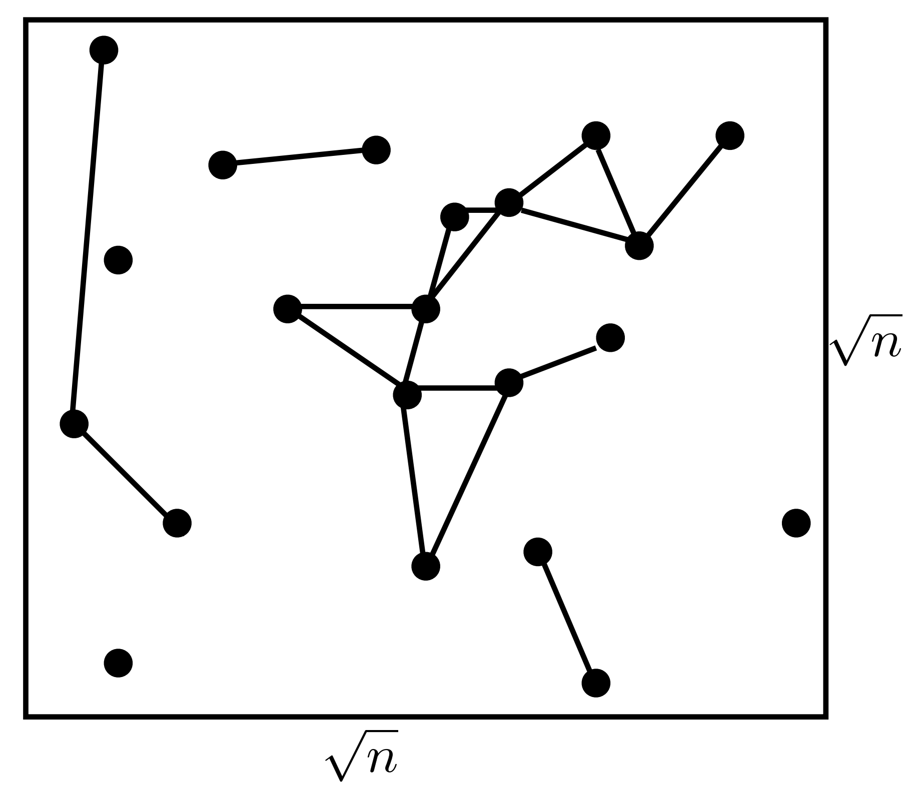
A lot of work on the Stochastic Block Model (SBM)

2 symmetric communities -
[Mossel, Neeman, Sly ’15][Mossel, Neeman, Sly ’13][Massoulie´’ 14]

Efficient algorithm, whenever it is information theoretically possible
Explicit closed form formulas for such an threshold

Unfortunately their techniques do not work,
- our model not locally tree like

How to exploit the knowledge of spatial locations



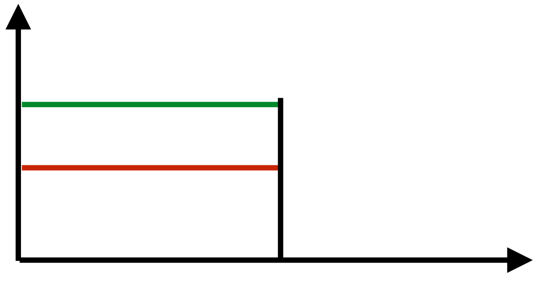
Algorithm Idea

Algorithm Idea

Spatial graph - *Locally dense* but globally *sparse*

Algorithm Idea

Spatial graph - *Locally dense* but globally *sparse*

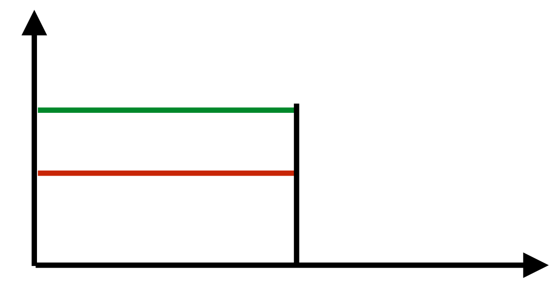


Consider the example $f_{in}(r) = a\mathbf{1}_{r \leq R}$, $f_{out}(r) = b\mathbf{1}_{r \leq R}$

$$0 \leq b < a \leq 1$$

Algorithm Idea

Spatial graph - *Locally dense* but globally *sparse*

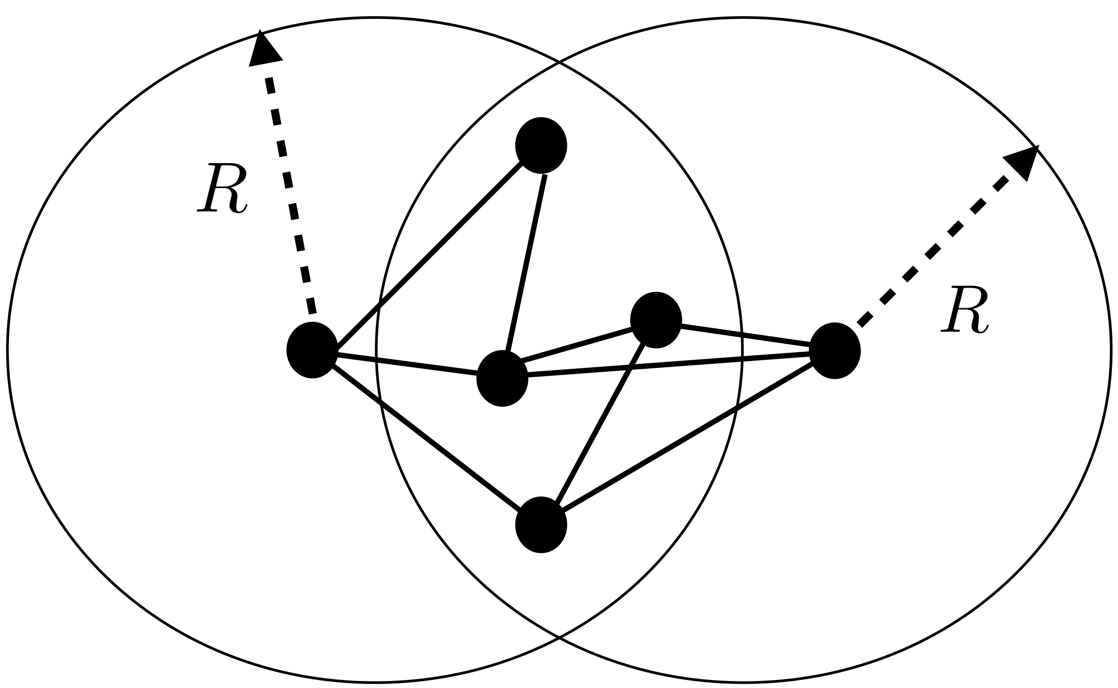


Consider the example $f_{in}(r) = a\mathbf{1}_{r \le R}$, $f_{out}(r) = b\mathbf{1}_{r \le R}$

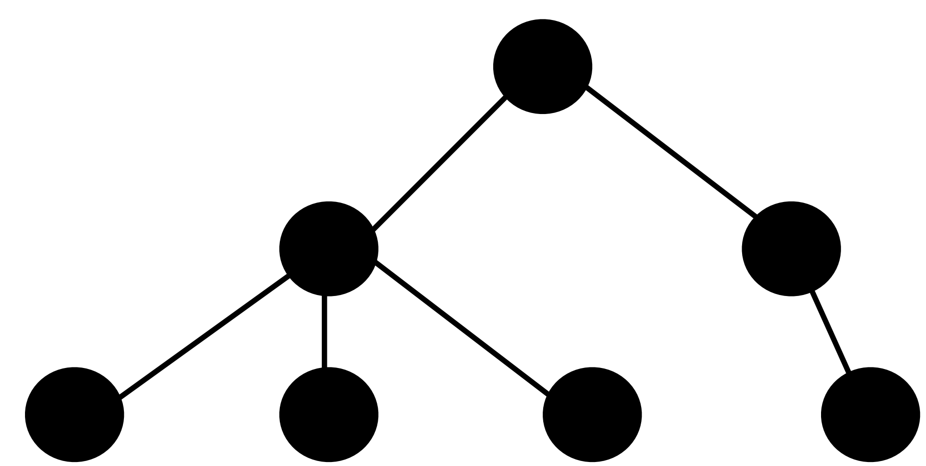
$$0 \le b < a \le 1$$

<u>*Locally Dense*</u> - '*Nearby*' nodes connect with *constant probability* independent of n .

<u>*Globally Sparse*</u> - Order n edges in total

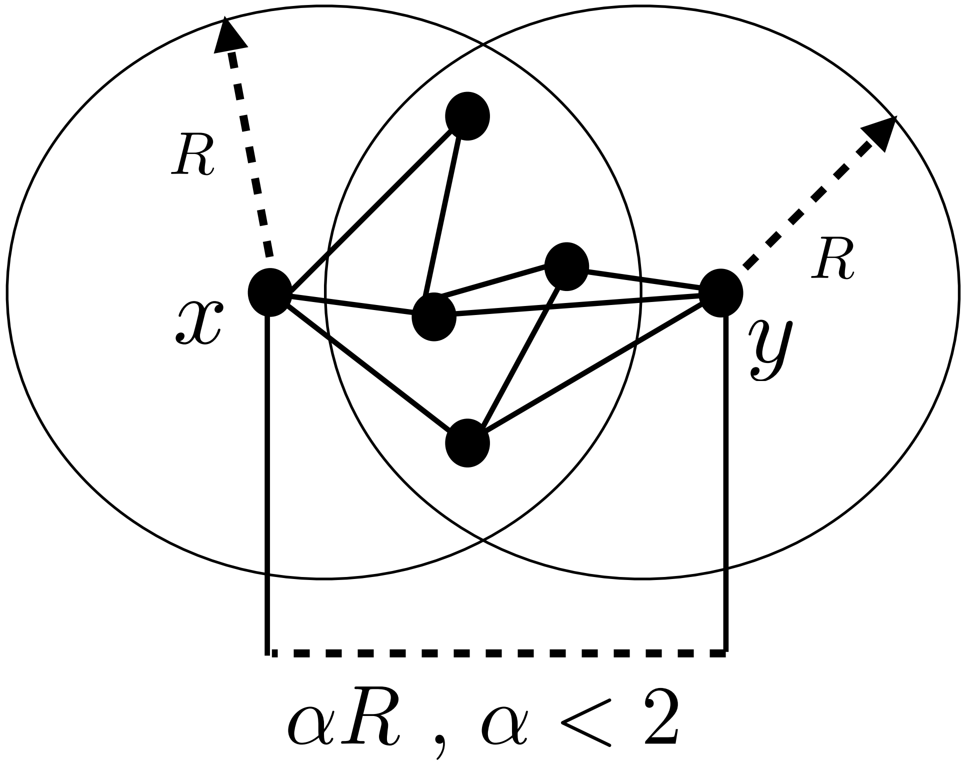


Spatial Graph



SBM

Algorithm Idea

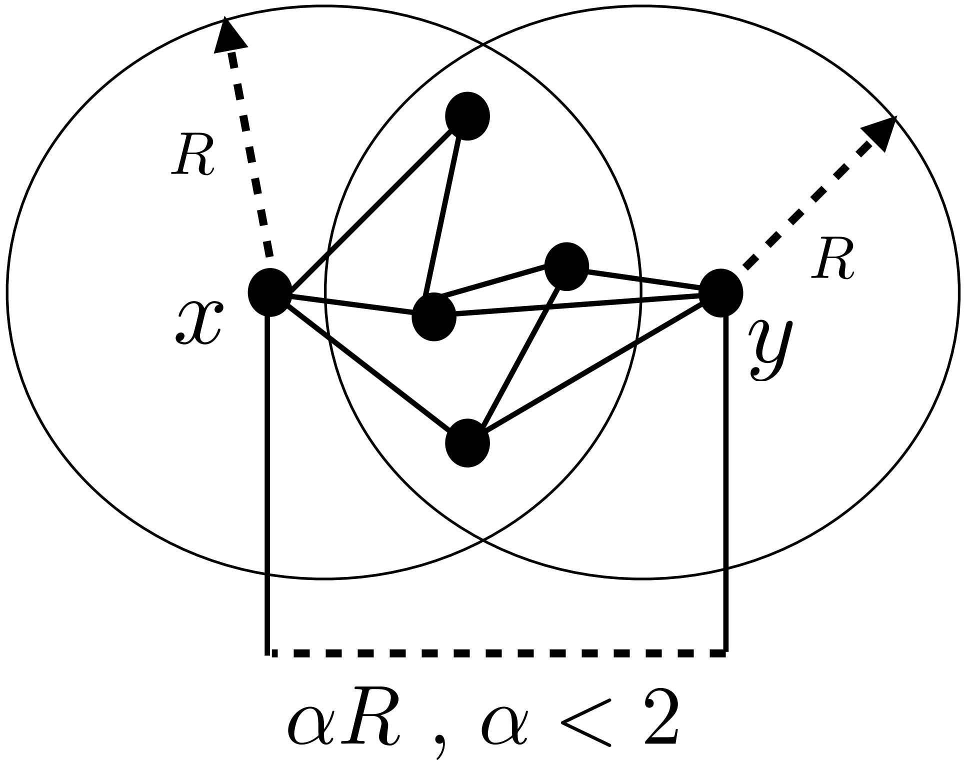


common neighbors is Poisson with mean

Same community - $\lambda c(\alpha) R^d \left(\frac{a^2 + b^2}{2} \right)$

Opposite communities - $\lambda c(\alpha) R^d ab$

Algorithm Idea



common neighbors is Poisson with mean

Same community - $\lambda c(\alpha) R^d \left(\frac{a^2+b^2}{2}\right)$

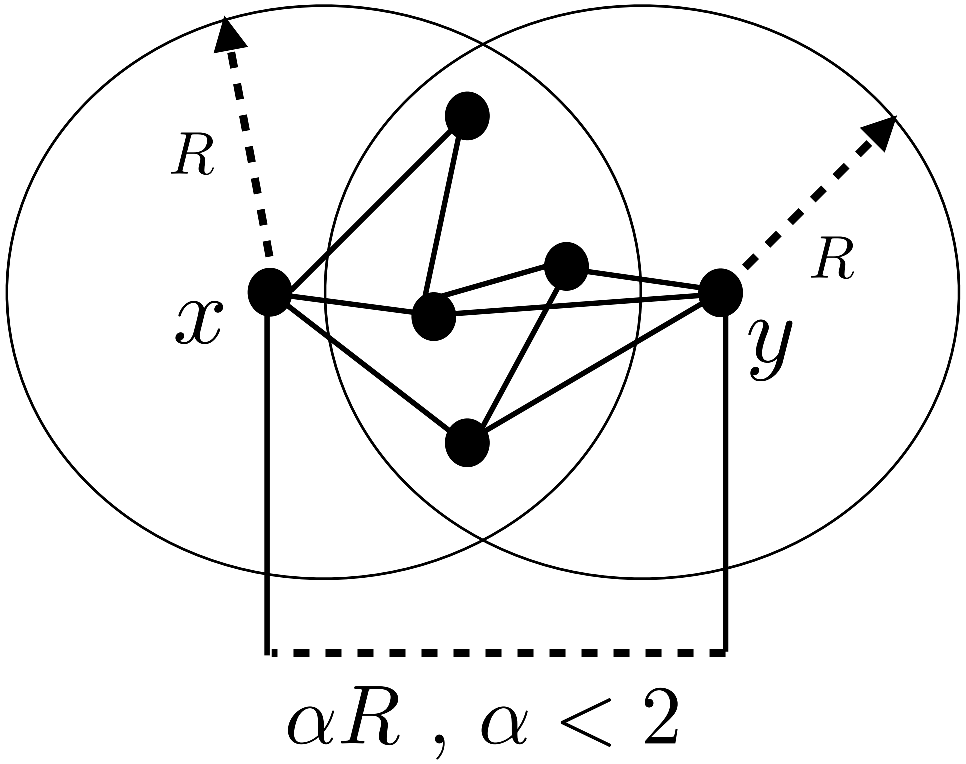
Opposite communities - $\lambda c(\alpha) R^d ab$

Set threshold - $T(\alpha) = c(\alpha) R^d \lambda \left(\frac{a+b}{2}\right)^2$

Pairwise-Classify(x,y)

- IF # (common neighbors) < $T(\alpha)$, *DECLARE* community(x) = community(y)
- ELSE *DECLARE* community(x) $\neq$ community(y)

Algorithm Idea



common neighbors is Poisson with mean

Same community - $\lambda c(\alpha) R^d \left(\frac{a^2 + b^2}{2} \right)$

Opposite communities - $\lambda c(\alpha) R^d ab$

Set threshold - $T(\alpha) = c(\alpha) R^d \lambda \left(\frac{a+b}{2} \right)^2$

__Pairwise-Classify(x,y)__

- IF # (common neighbors) < $T(\alpha)$, *DECLARE* community(x) = community(y)
- ELSE *DECLARE* community(x) $\neq$ community(y)

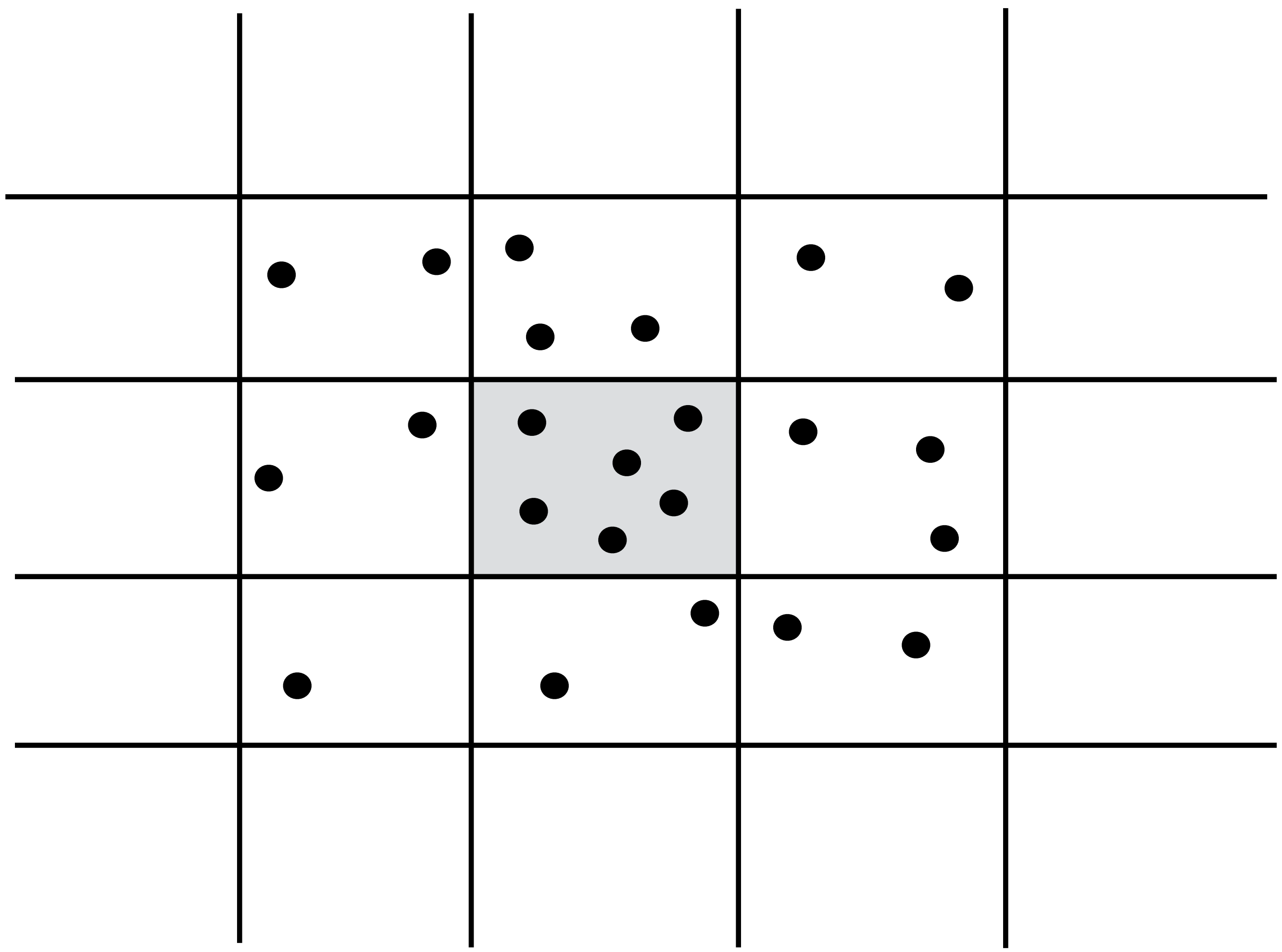
Chernoff bound -

P(Mis-classifying a given pair of nodes at distance αR) $\leq e^{-\lambda c'(\alpha) R}$

Algorithm Idea

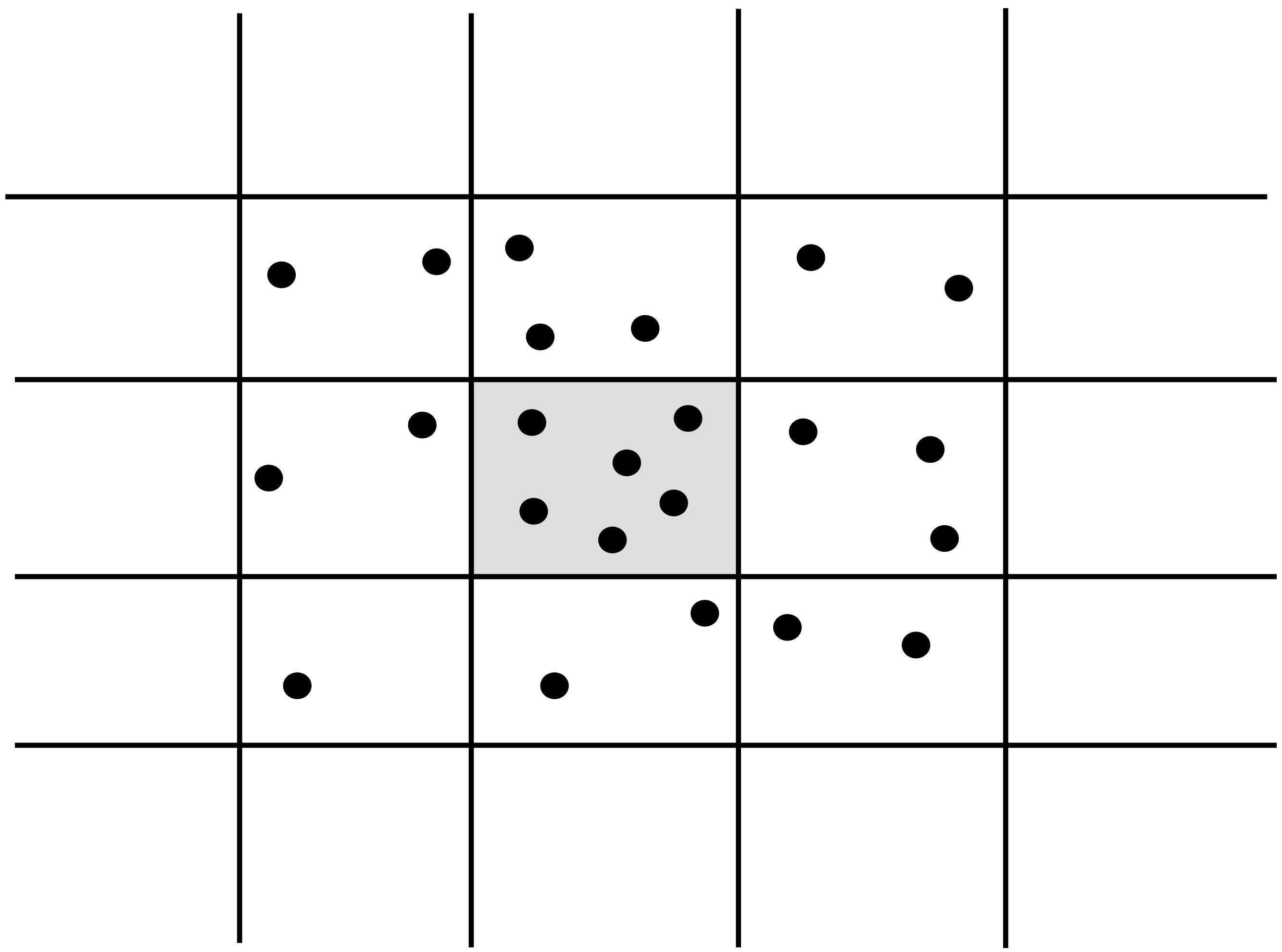
Tesselate $\mathbb{R}^d$ into grids of side $R/4$

Classify cells to be Good or Bad



Algorithm Idea

Tesselate $\mathbb{R}^d$ into grids of side $R/4$

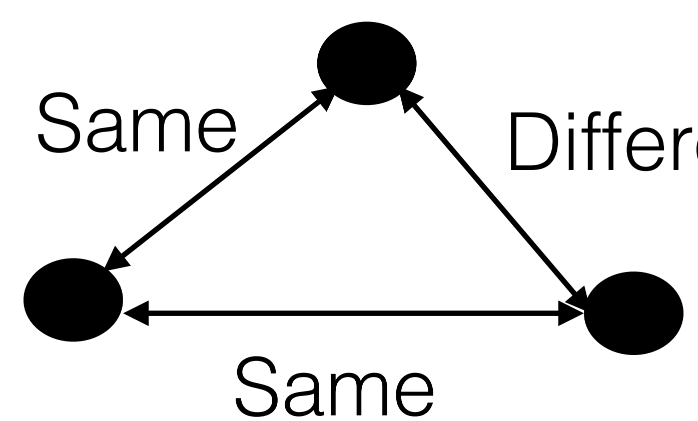


Classify cells to be Good or Bad

Cell ***Good*** if

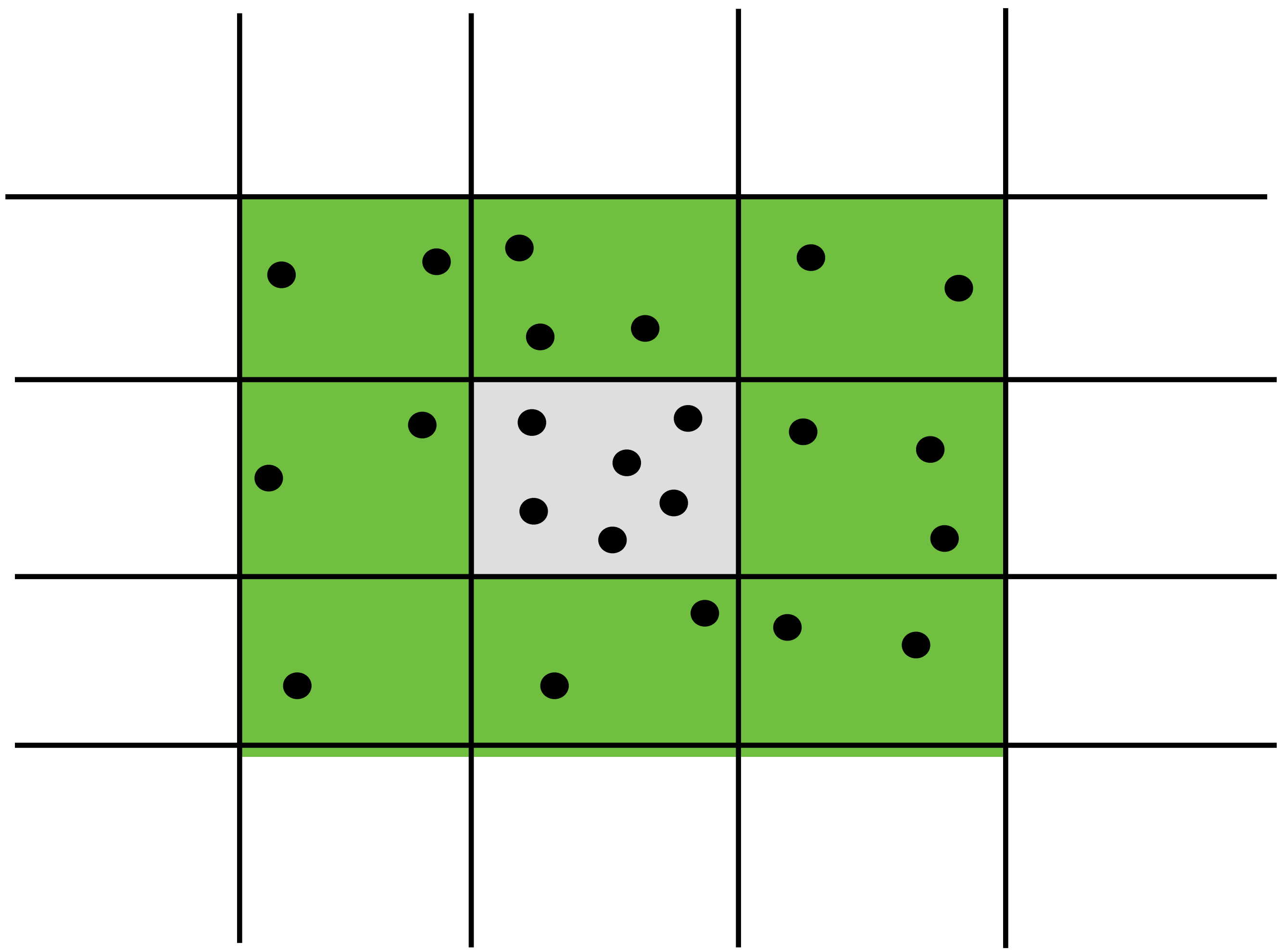
1. At-least $(1-\epsilon)$ Mean # of nodes
2. No ***inconsistencies*** in pairwise checks *with all neighboring cells*

Example of Inconsistent output



Algorithm Idea

Tesselate $\mathbb{R}^d$ into grids of side $R/4$

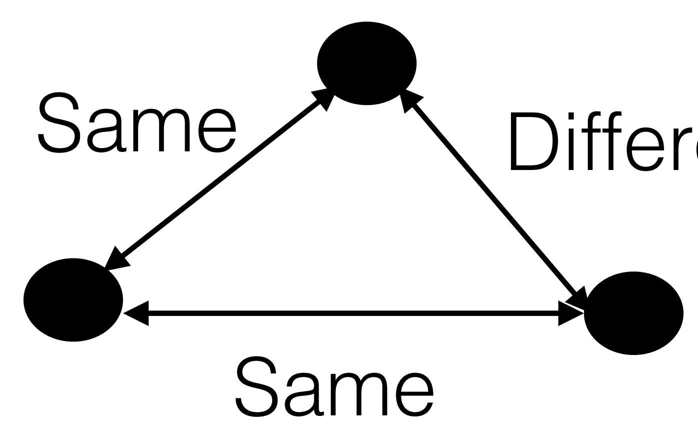


Classify cells to be Good or Bad

Cell ***Good*** if

1. At-least $(1-\epsilon)$ Mean # of nodes
2. No ***inconsistencies*** in pairwise checks *with all neighboring cells*

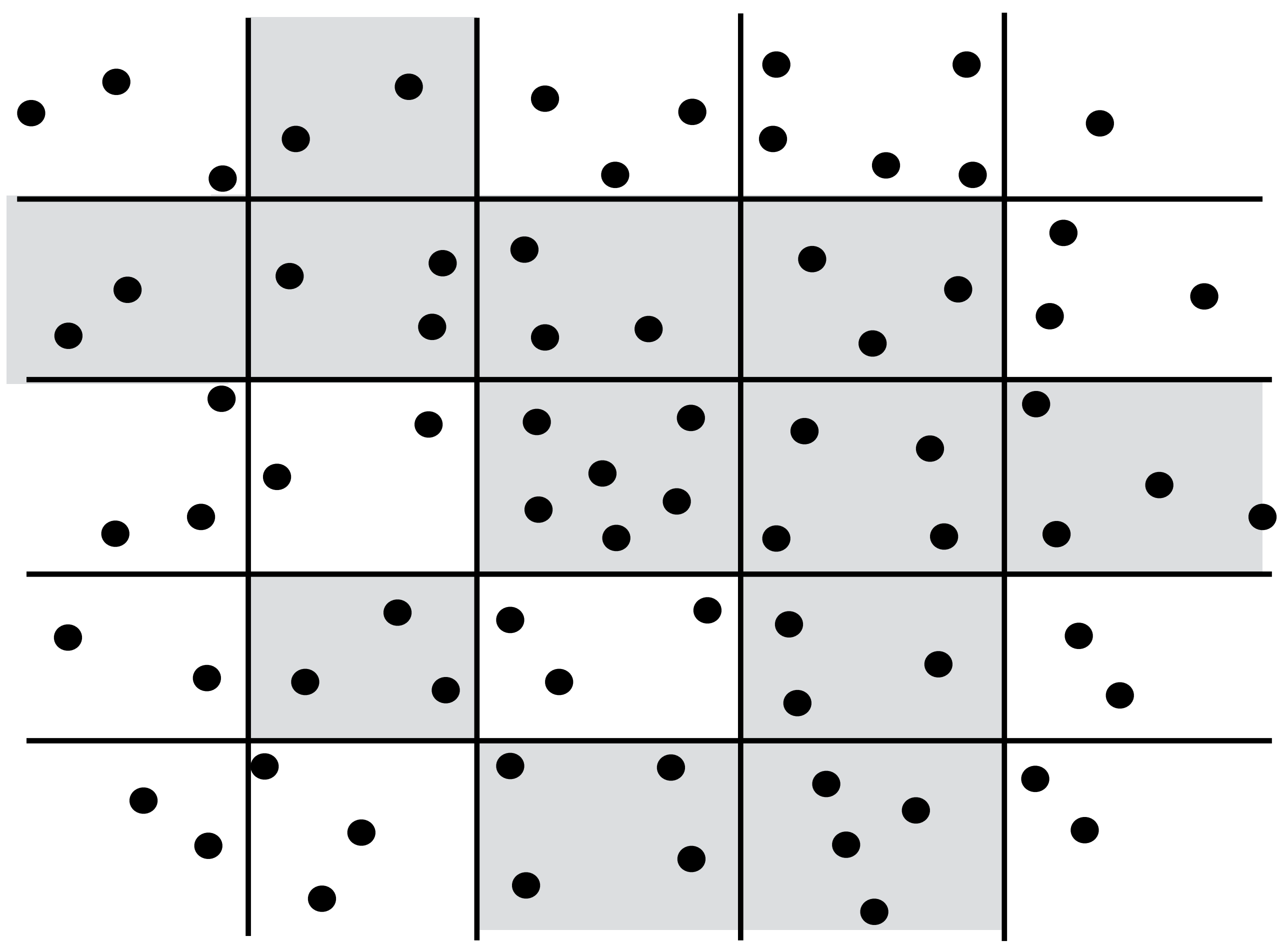
Example of Inconsistent output



Algorithm Idea

<u>Main Routine</u>

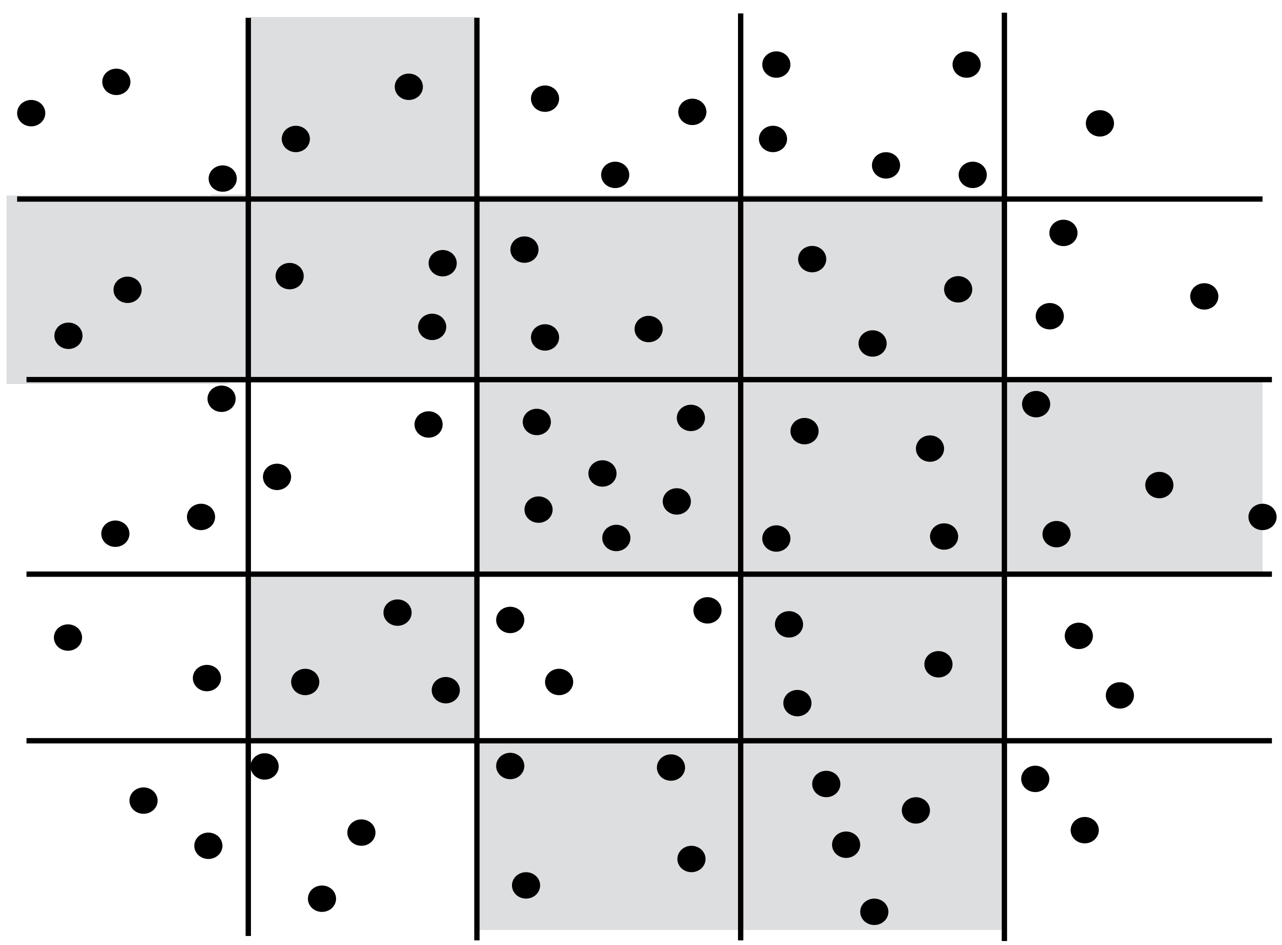
- Partition each good component with Pairwise-Classify
- Output +1 estimate to all nodes in bad cells



Algorithm Idea

Main Routine

- Partition each good component with Pairwise-Classify
- Output +1 estimate to all nodes in bad cells



Algorithm succeeds if a "large" connected component of "gray" cells is present

A k-Dependent Percolation Process. [Liggett, Schonmann, Stacey, '97]

Solvability Phase Transition

An overlap of γ is ***achievable*** if there exists an estimator $\{\tau_i\}_{i=1}^{N_n}$ such that $\lim_{n\to\infty} \mathbb{P}[\mathcal{O}_\tau > \gamma] = 1$

Solvability Phase Transition

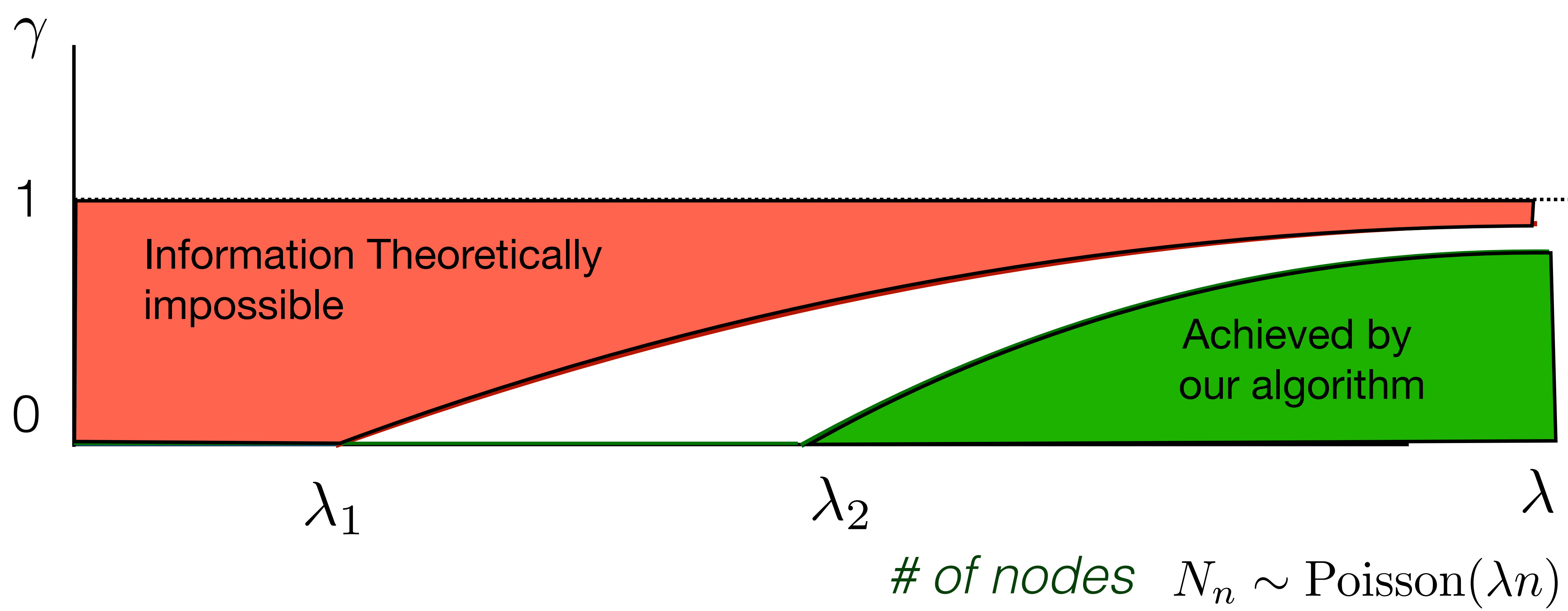
An overlap of γ is ***achievable*** if there exists an estimator $\{\tau_i\}_{i=1}^{N_n}$ such that $\lim_{n\to\infty} \mathbb{P}[\mathcal{O}_\tau > \gamma] = 1$

Solvability iff any $\gamma > 0$ is achievable

Solvability Phase Transition

An overlap of γ is ***achievable*** if there exists an estimator $\{\tau_i\}_{i=1}^{N_n}$ such that $\lim_{n\to\infty} \mathbb{P}[\mathcal{O}_\tau > \gamma] = 1$

Solvability iff any $\gamma > 0$ is achievable

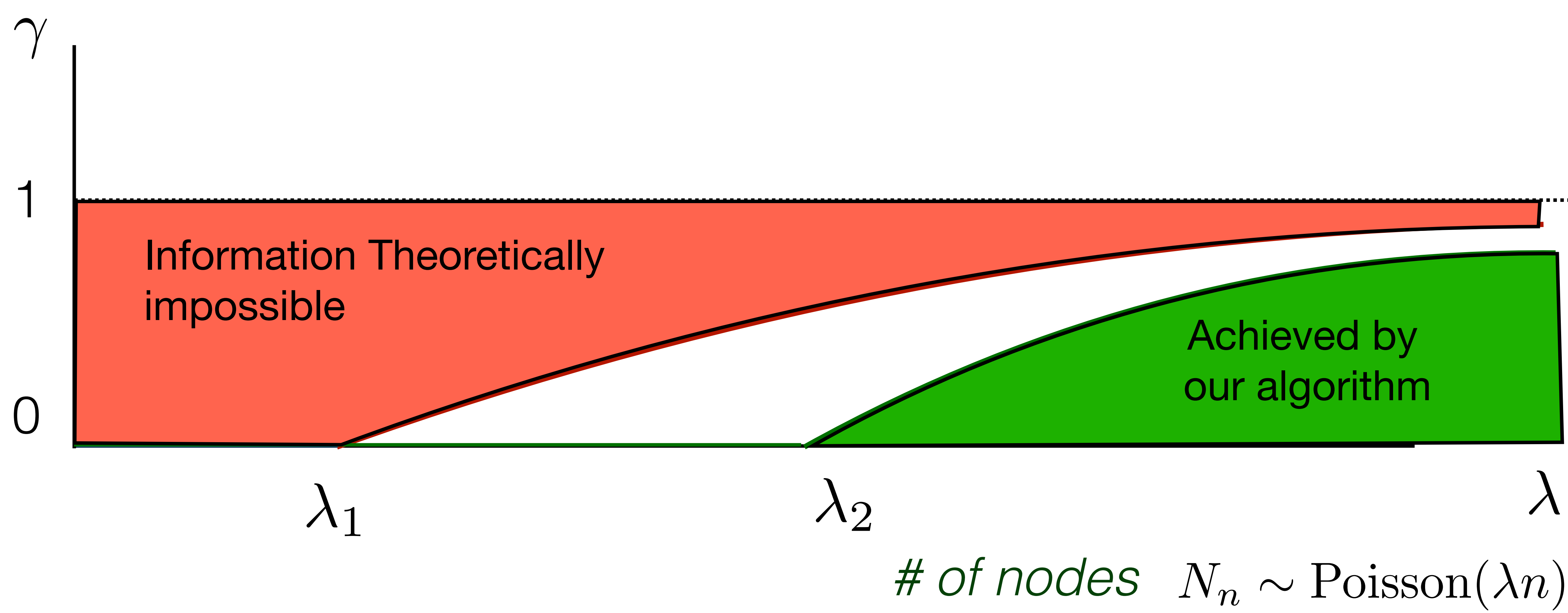


Solvability Phase Transition

<u>Theorem</u> - $\forall f_{in}(\cdot), f_{out}(\cdot), d \geq 2,$ $\exists\, 0 < \lambda_1 \leq \lambda_2 < \infty$ such that -

$\lambda < \lambda_1 \implies$ Community Detection is not solvable

$\lambda > \lambda_2 \implies$ Our algorithm solves Community Detection efficiently



Our algorithm is *asymptotically optimal*.

Haplotype Assembly
An Application of Euclidean Community Detection

A.S, Haris Vikalo, François Baccelli, *Haplotype Phasing and Community Detection,* In Préparation

Haplotype Assembly - Problem

Reconstruct the string from noisy measurements

Haplotype Assembly - Problem

Reconstruct the string from noisy measurements

0 1 1 0 1 1 1 0 1 1 0 0 0 1 1 0 1	$\mathbf{s}$
1 0 0 1 0 0 0 1 0 0 1 1 1 0 0 1 0	$\mathbf{s}^c$

Haplotype Assembly - Problem

Reconstruct the string from noisy measurements

0 1 1 0 1 1 1 0 1 1 0 0 0 1 1 0 1 $\mathbf{s}$

1 0 0 1 0 0 0 1 0 0 1 1 1 0 0 1 0 $\mathbf{s}^c$

Each *paired-read* consists of

- The underlying string $\mathbf{s}$ or $\mathbf{s}^c$ that is unknown

Haplotype Assembly - Problem

Reconstruct the string from noisy measurements

0 1 1 0 1 1 1 0 1 1 0 0 0 1 1 0 1 $\mathbf{s}$

1 0 0 1 0 0 0 1 0 0 1 1 1 0 0 1 0 $\mathbf{s}^c$

Each *paired-read* consists of

- The underlying string $\mathbf{s}$ or $\mathbf{s}^c$ that is unknown
- A set of locations that is known

Haplotype Assembly - Problem

Reconstruct the string from noisy measurements

0 1 1 0 1 1 1 0 1 1 0 0 0 1 1 0 1 $\quad \mathbf{s}$

1 0 0 1 0 0 0 1 0 0 1 1 1 0 0 1 0 $\quad \mathbf{s}^c$

Each *paired-read* consists of

- The underlying string $\mathbf{s}$ or $\mathbf{s}^c$ that is unknown
- A set of locations that is known
- Noisy measurement of the unknown chosen string at the known chosen locations

Haplotype Assembly - Problem

Reconstruct the string from noisy measurements

0 1 1 0 1 1 1 0 1 1 0 0 0 1 1 0 1 $\mathbf{s}$

1 0 0 1 0 0 0 1 0 0 1 1 1 0 0 1 0 $\mathbf{s}^c$

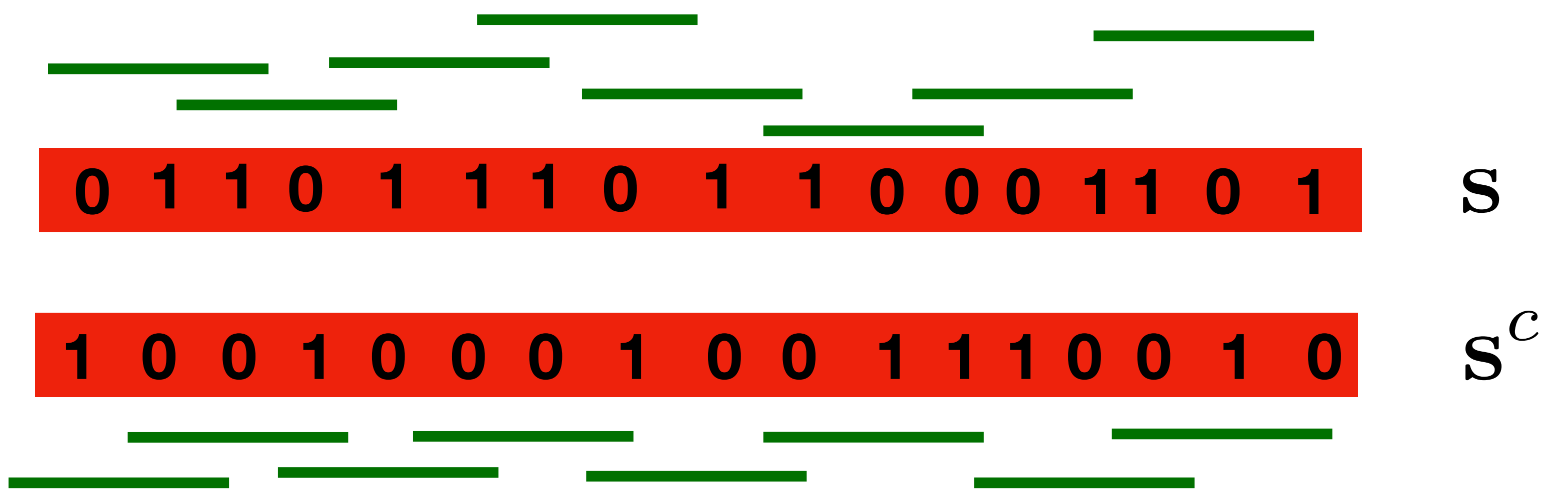
Each *paired-read* consists of

- The underlying string $\mathbf{s}$ or $\mathbf{s}^c$ that is unknown
- A set of locations that is known
- Noisy measurement of the unknown chosen string at the known chosen locations

Read 1 - Positions - 2,10 Values:000,011

Haplotype Assembly - Problem

Reconstruct the string from noisy measurements



Read 1 - Positions - 2,10 Values:000,011
Read 2 - Positions - 4,11 Values:00,01
Read 3 - Positions - 1,9   Values:0010,01101
Read 4 - Positions - 2,11 Values:00011,01

.

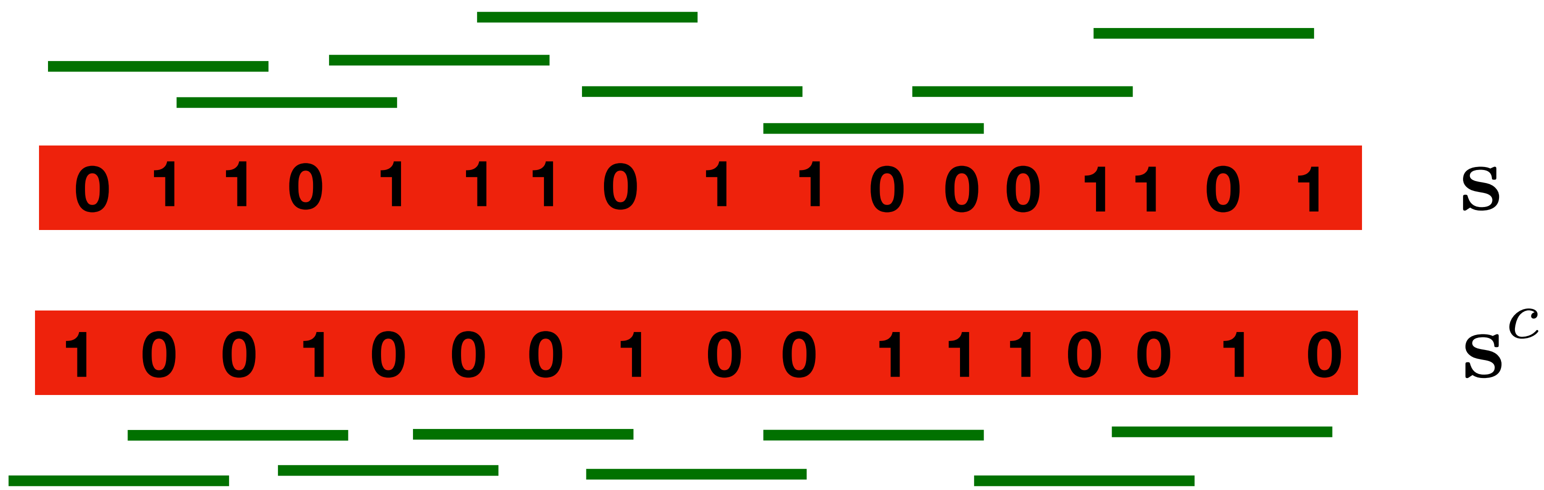
.

.

Read m - Positions - 21,40. Values:0,01100

Haplotype Assembly - Problem

Reconstruct the string from noisy measurements



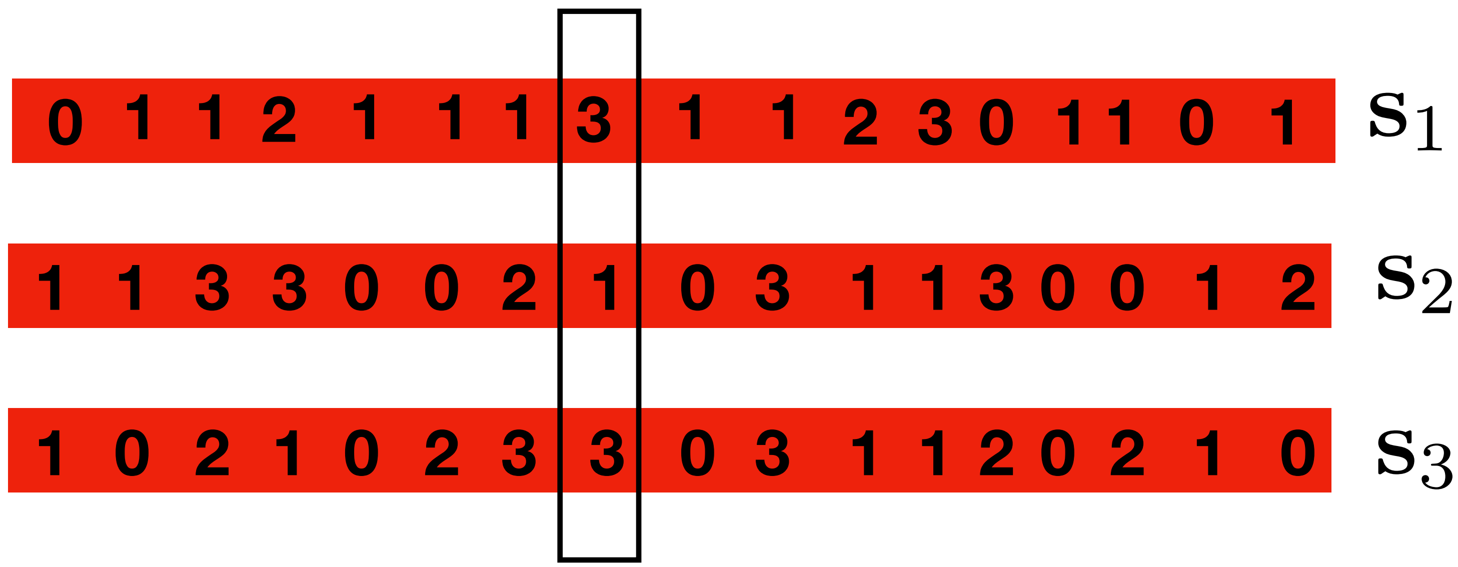
Fundamental and *challenging* problem in computational genomics

Binary alphabet - long history

We consider the general case of multiple strings and multiple alphabets

Haplotype Assembly - Problem

Reconstruct the string from noisy measurements



At all positions, not all strings are identical

Fundamental and *challenging* problem in computational genomics

Binary alphabet - long history

We consider the general case of multiple strings and multiple alphabets

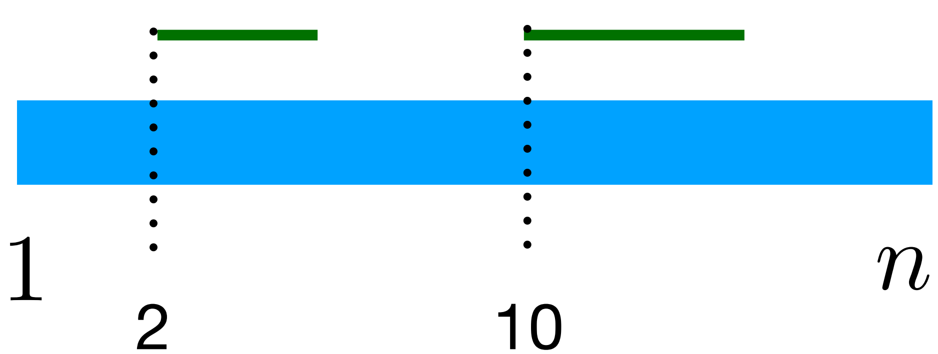
Proposed Algorithm

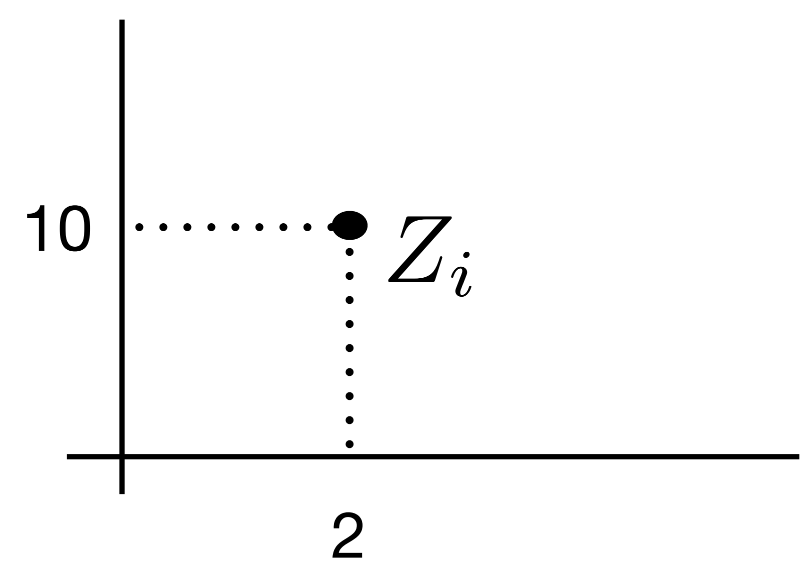
Each read is a node of a weighted spatial graph

Read i - Positions - 2,10 Values:000,011

The unknown string - ‘community label’

The set of positions - ‘location label’





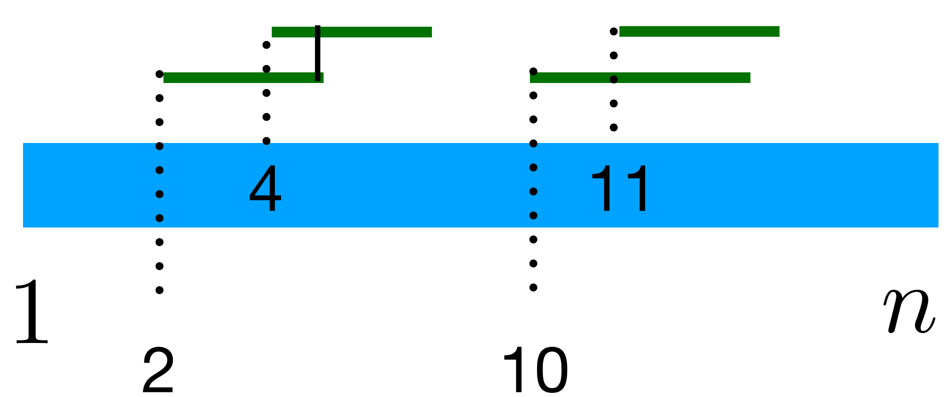
Proposed Algorithm

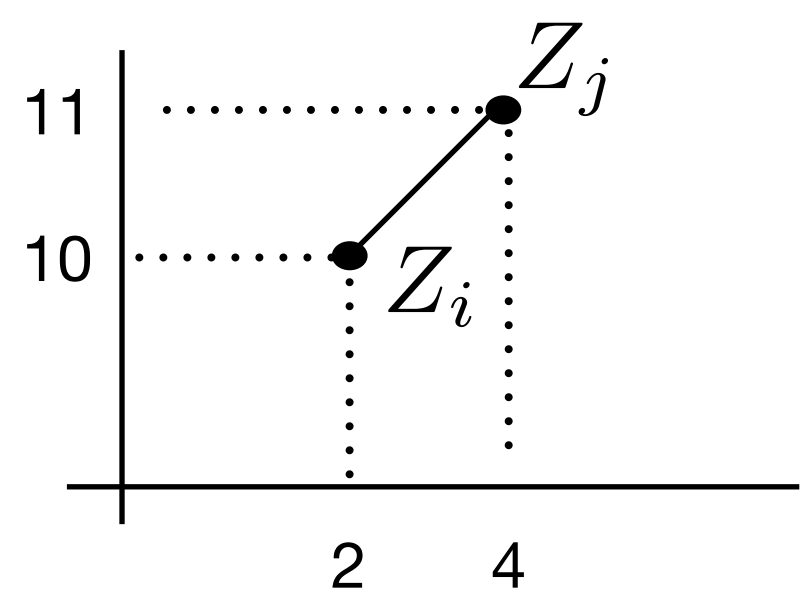
Each read is a node of a weighted spatial graph

```
Read i - Positions - 2,10  Values:000,011
Read j - Positions - 4,11  Values:01,101
```

The unknown string - 'community label'

The set of positions - 'location label'





Proposed Algorithm

Each read is a node of a weighted spatial graph

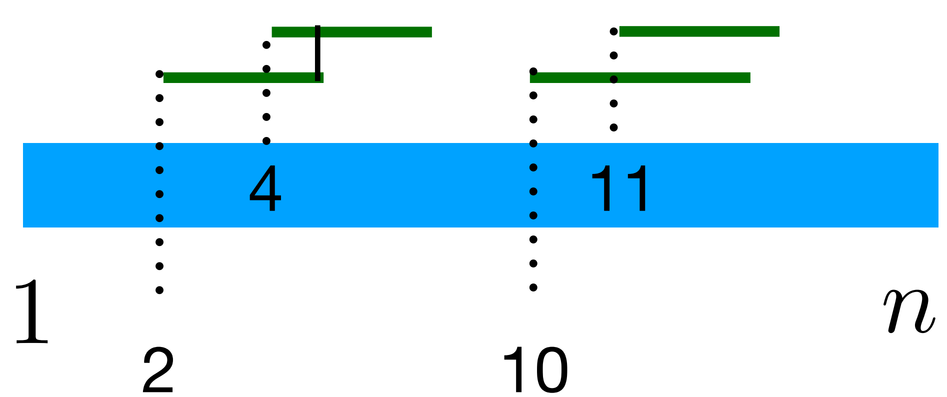
```
Read i - Positions - 2,10  Values:000,011
Read j - Positions - 4,11  Values:01,101
```

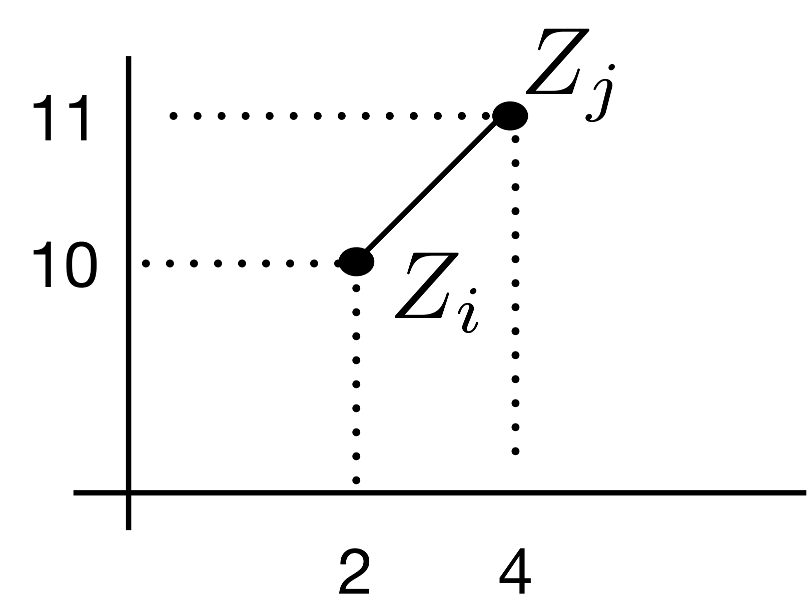
The unknown string - ‘community label’

The set of positions - ‘location label’

$$w_{ij} = \frac{\#\text{Sites the reads agrees on} - \#\text{Sites the reads differs}}{\#\text{Total number of overlapping sites}}$$

$$w_{ij} = \frac{2-1}{2+1} \qquad \text{Overlapping Sites} = \{4, 10, 11\}$$





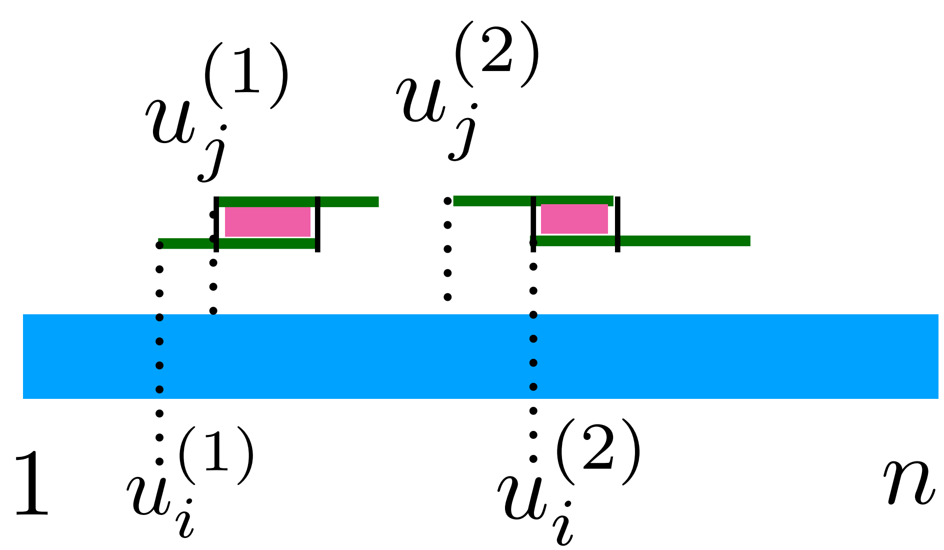
Proposed Algorithm

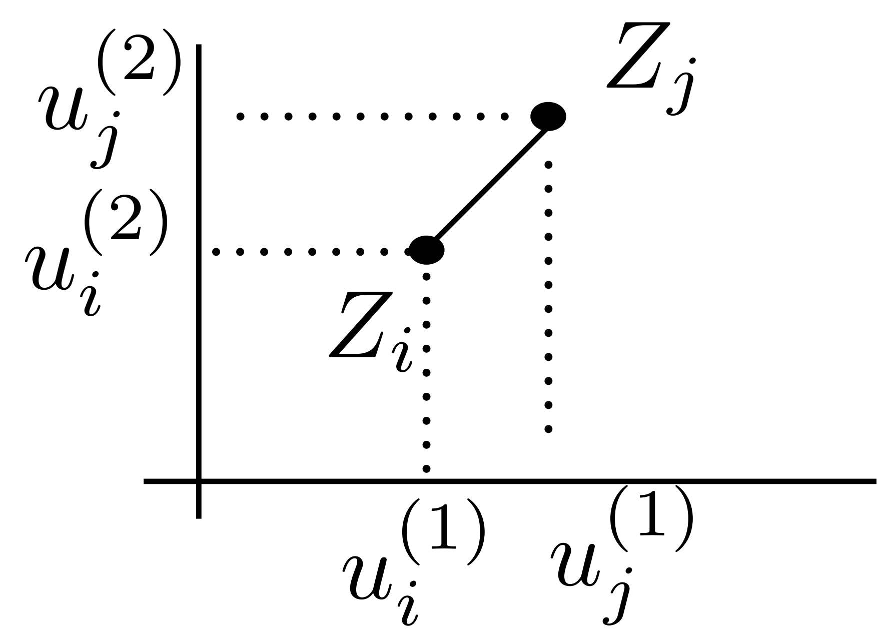
1) Create the weighted spatial graph

2) Euclidean Community Detection

3) Each position in each string estimated by a majority of all reads estimated to be originating from the string in consideration and covering the position

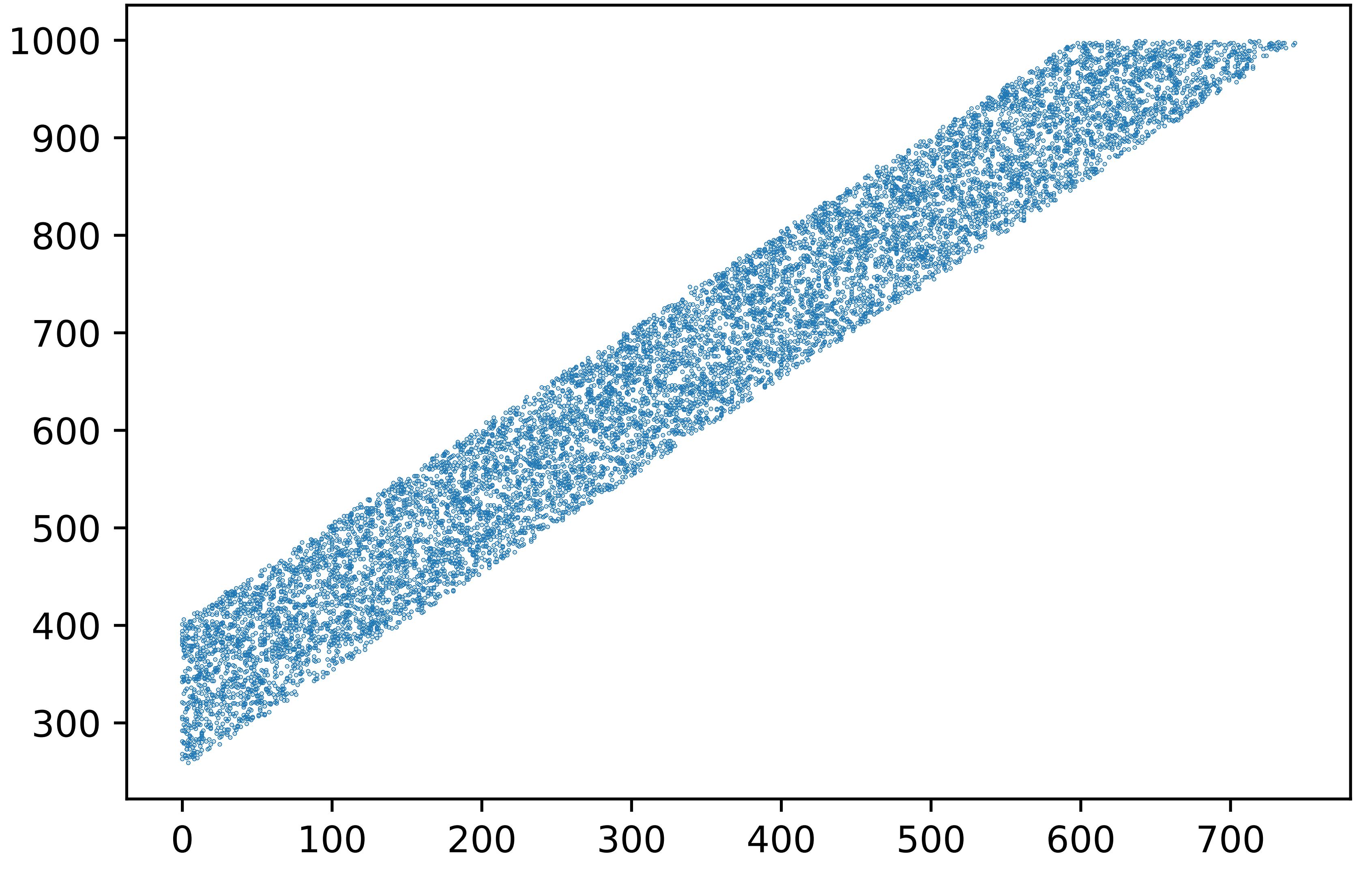
$$w_{ij} = \frac{\#\text{Sites the reads agrees on} - \#\text{Sites the reads differs}}{\#\text{Total number of overlapping sites}}$$





A paired-end read measurement

Proposed Algorithm



Benchmark simulation data with 4 strings and string length 700.

Prior Work

AltHaP - [Hashemi, Zhu and Vikalo '18]

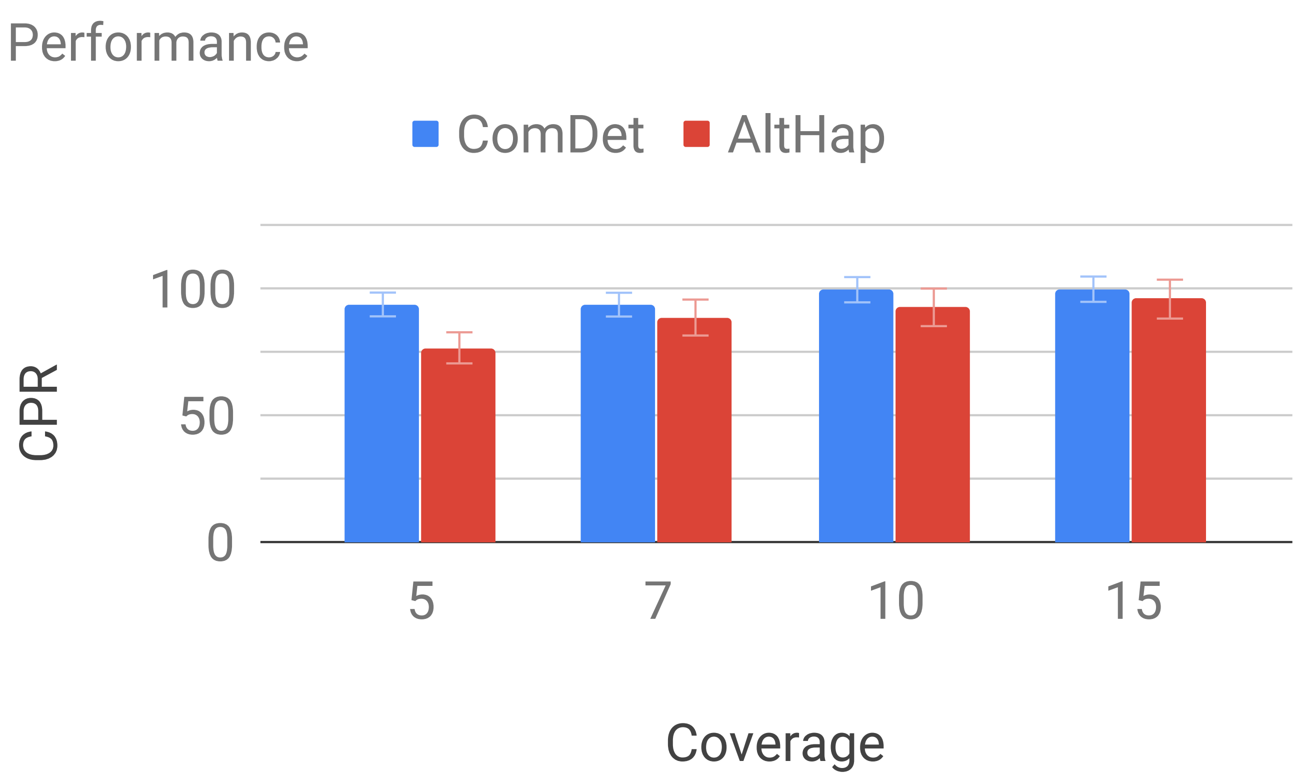
State of art and the first algorithm to handle multiple strings and alphabet sizes

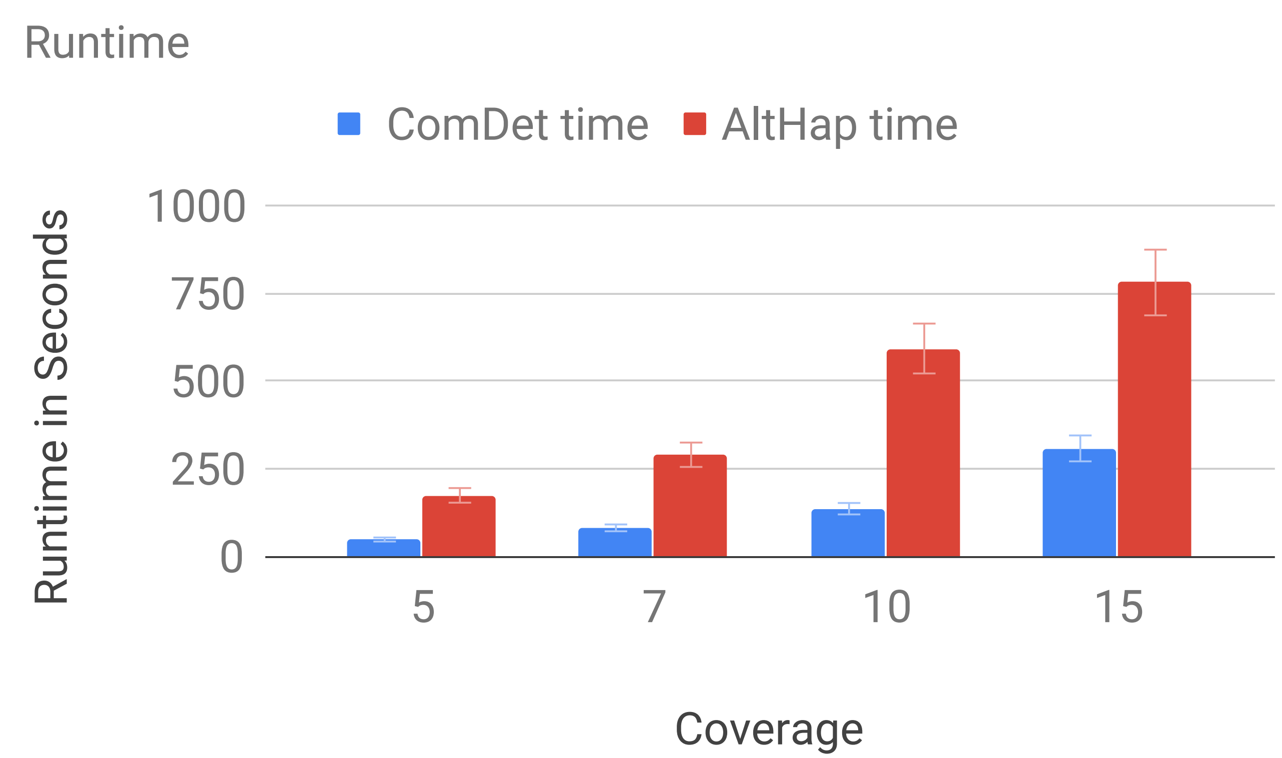
Poses the problem as noisy tensor completion

Ignores the spatial representation of data. Thus, computationally expensive !

Our algorithm has an Implicit 'regularizer' to force that all strings at all locations are uniformly sampled by reads

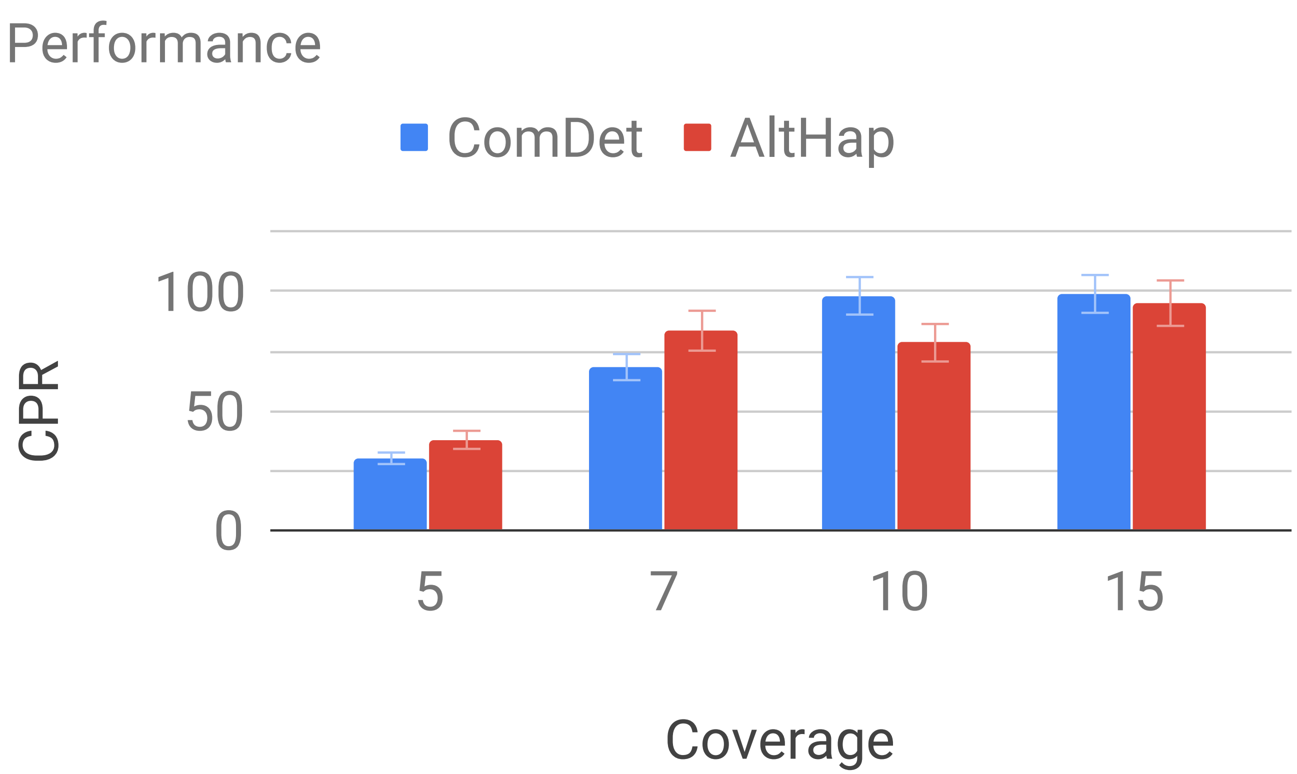
Haplotype Assembly - Results

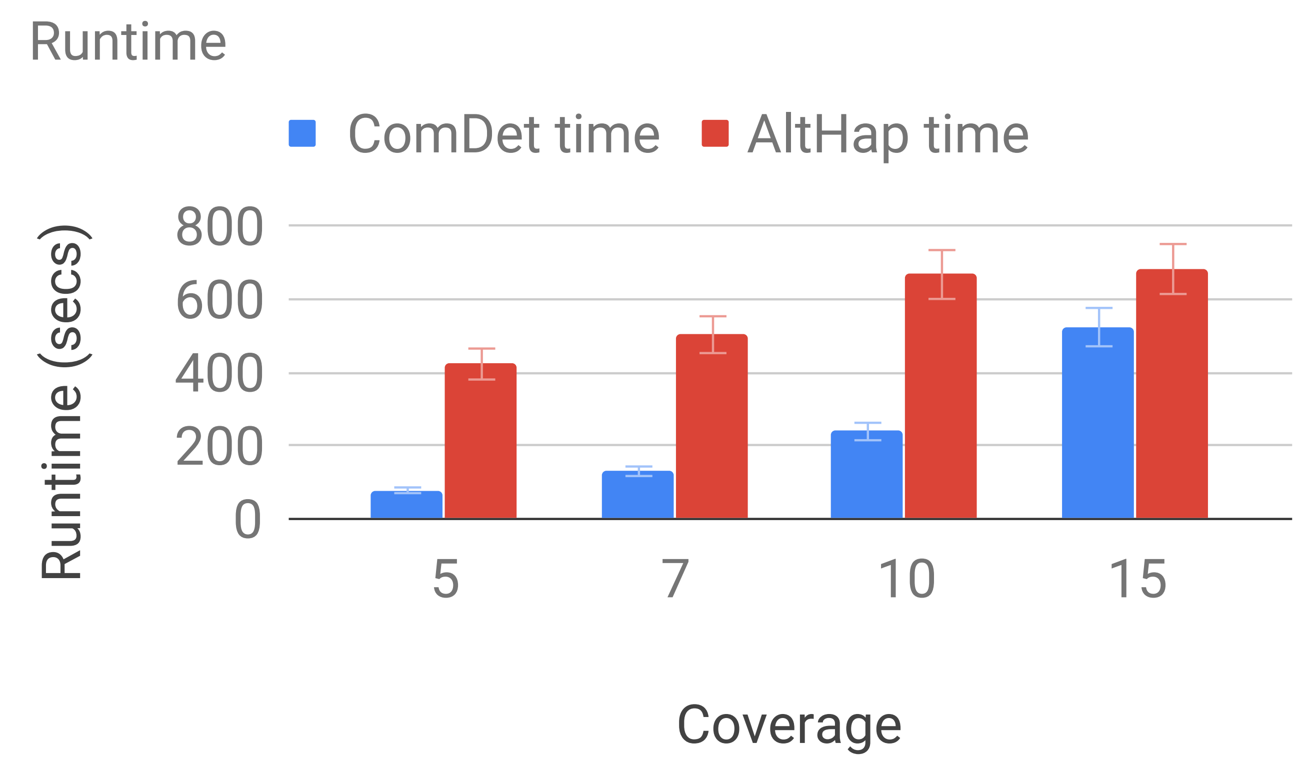




Ploidy = 3, Alphabet size = 4, Error Rate in reads = 0.01

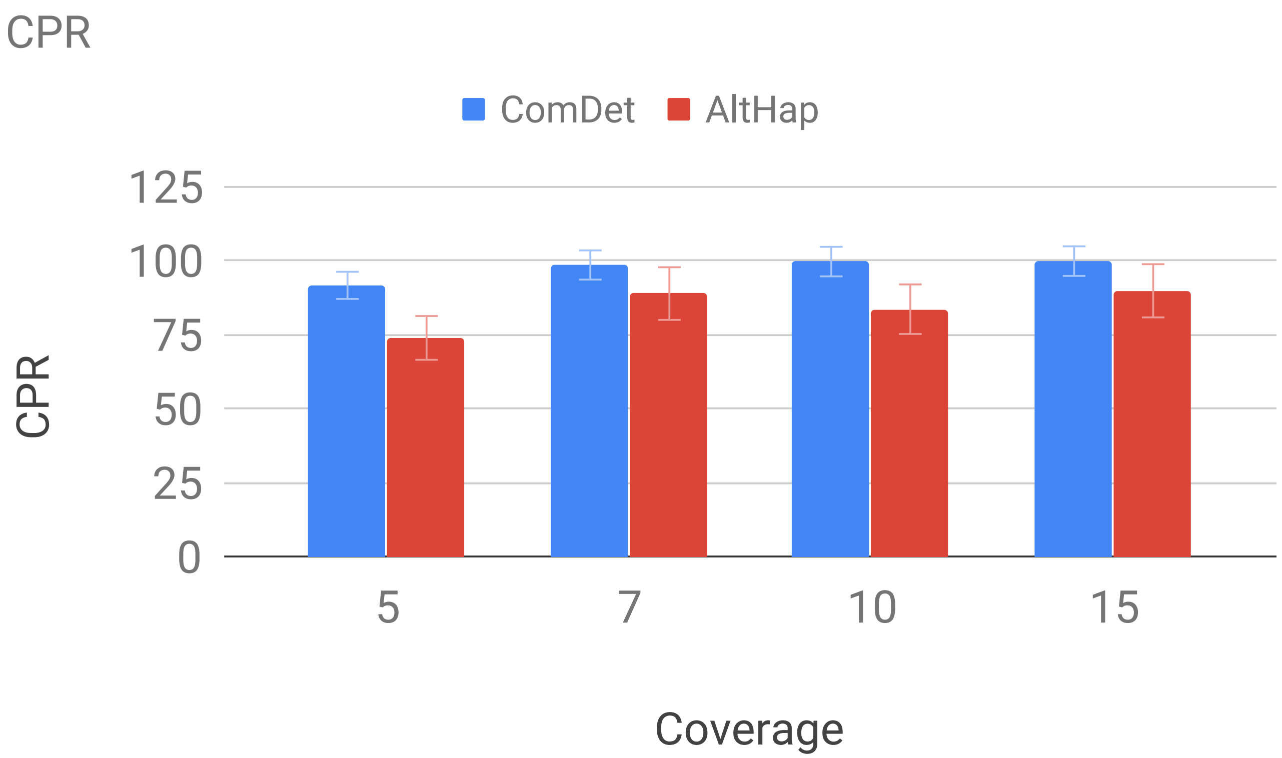
Haplotype Assembly - Results

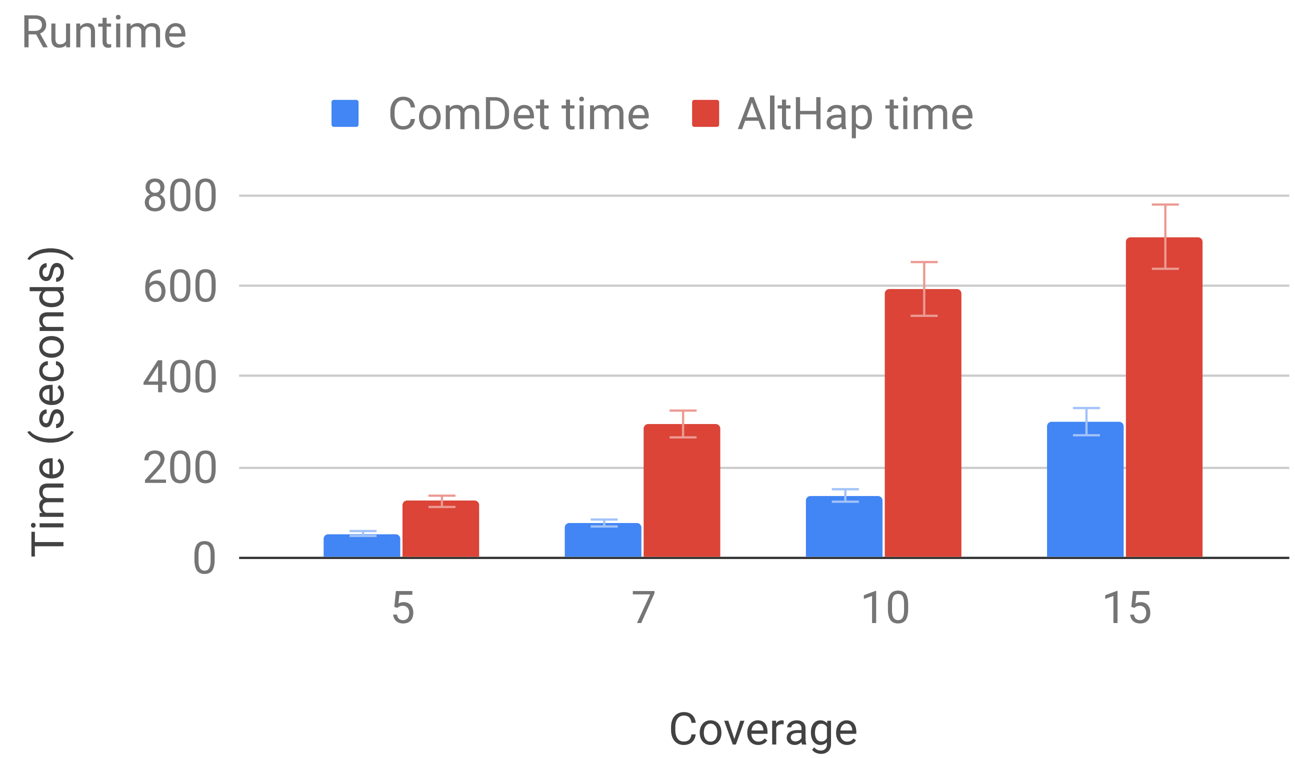




Ploidy = 4, Alphabet size = 4, Error Rate in reads = 0.05

Haplotype Assembly - Results





Ploidy = 3, Alphabet size = 4, Error Rate in reads = 0.002

Conclusions

First step towards Euclidean Community Detection

Some recent improvements
[Abbe and Boix ’18], [Polyanskiy and Wu ’18], [Alaoui and Montanari ‘19]

Open Mathematical Problems -

1) Unknown number of communities
2) Heterogeneous densities for the various communities
3) Characterization of the hardness of the problem

   (Phase Transitions and Statistical/Computational *gaps*)

4) Apply the method to other problems involving paired-end reads

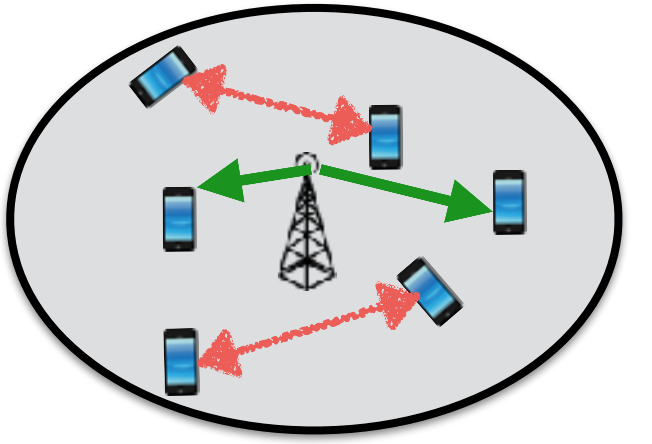
Spatial Dynamics for Wireless Networks

A.S,François Baccelli and Sergey Foss, *Interference Queueing Networks on Grids,*
In Annals of Applied Probability (To Appear)

A.S and François Baccelli, *Spatial Birth-Death Wireless Networks,*
In IEEE Transactions on Information Theory, 2017

Ad-Hoc Wireless Networks

Networks without a centralized infrastructure

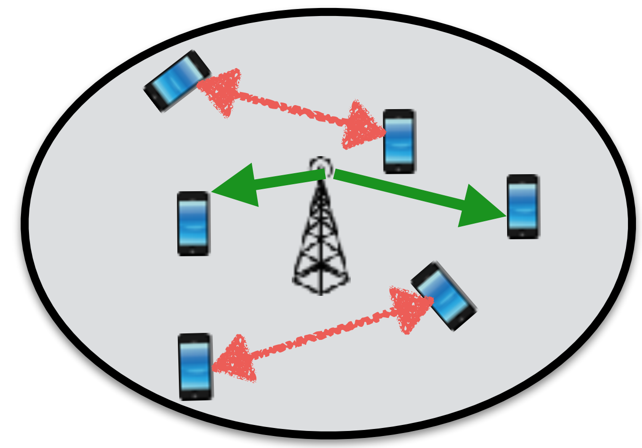


Examples -

1) Overlaid Device-to-Device (D2D) Networks

Ad-Hoc Wireless Networks

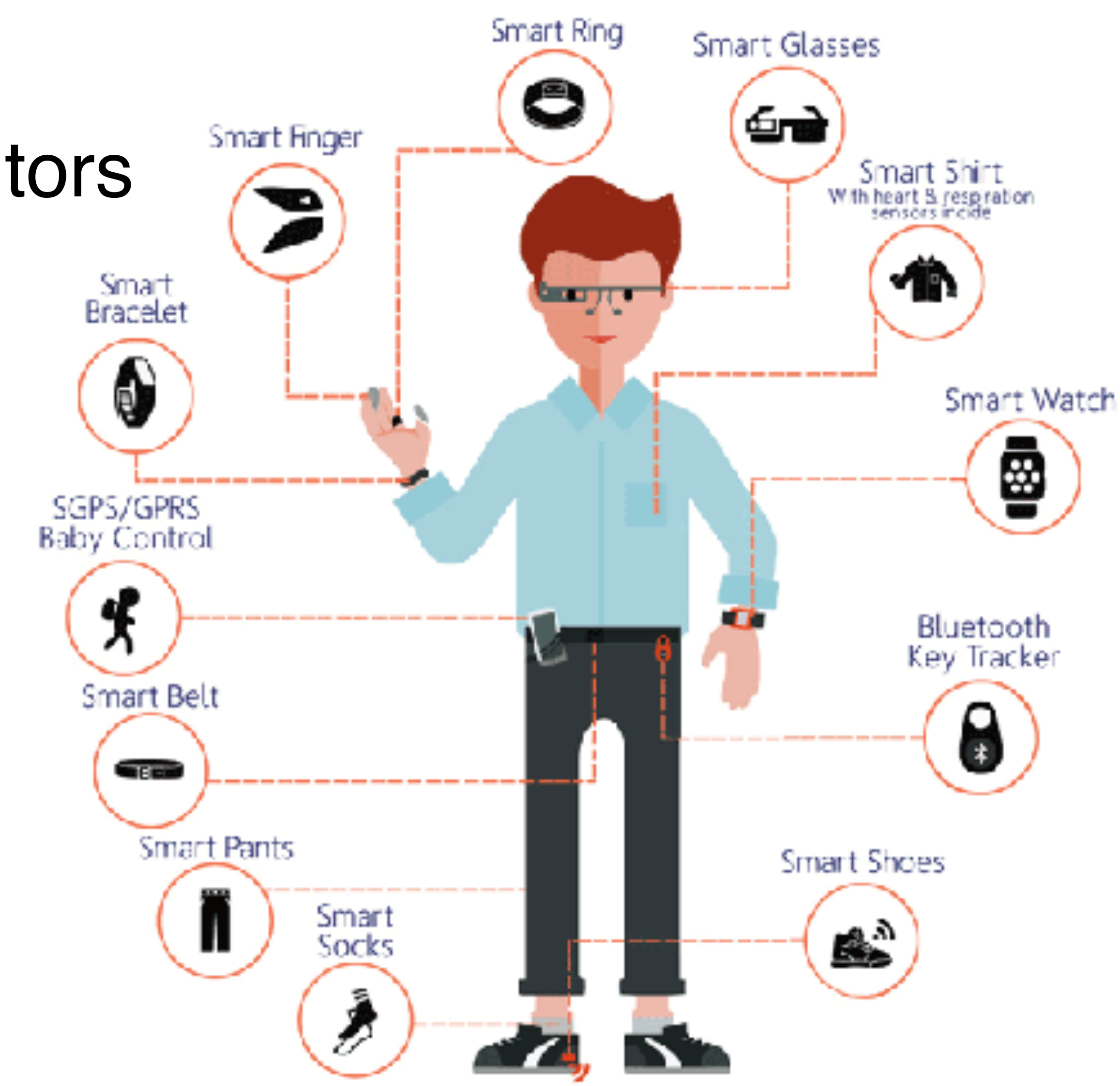
Networks without a centralized infrastructure



Examples -

1) Overlaid Device-to-Device (D2D) Networks

2) Internet of Things - Sensors and monitors

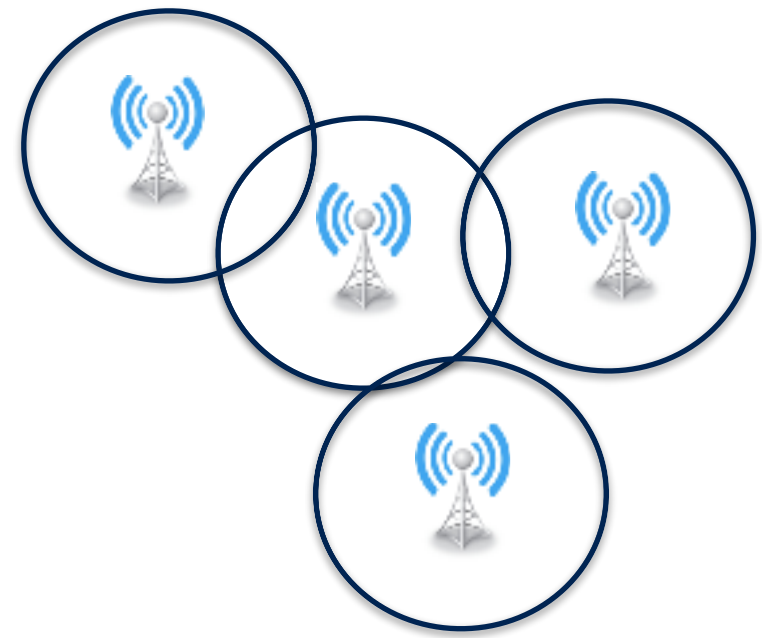


Wireless devices everywhere !

Spatio-Temporal Dynamics

Wireless Spectrum is a space-time shared resource

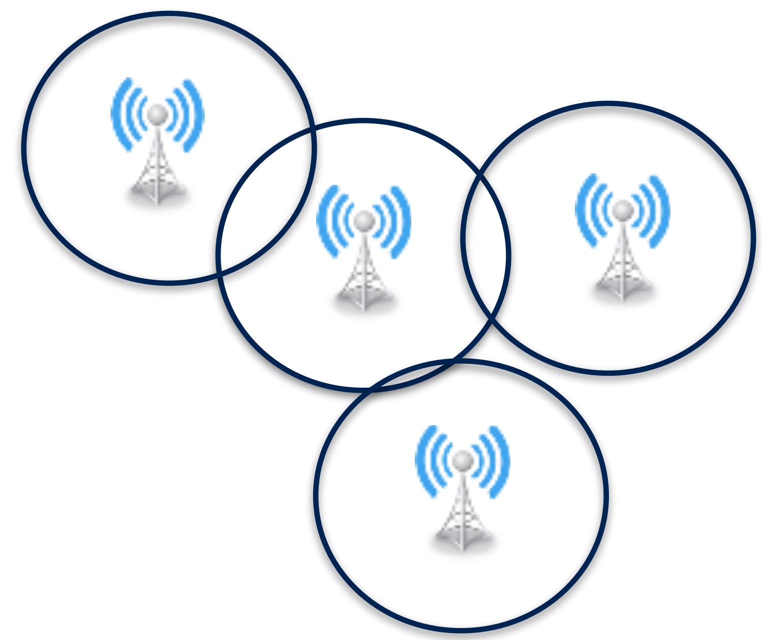
Spatial Component - Interference



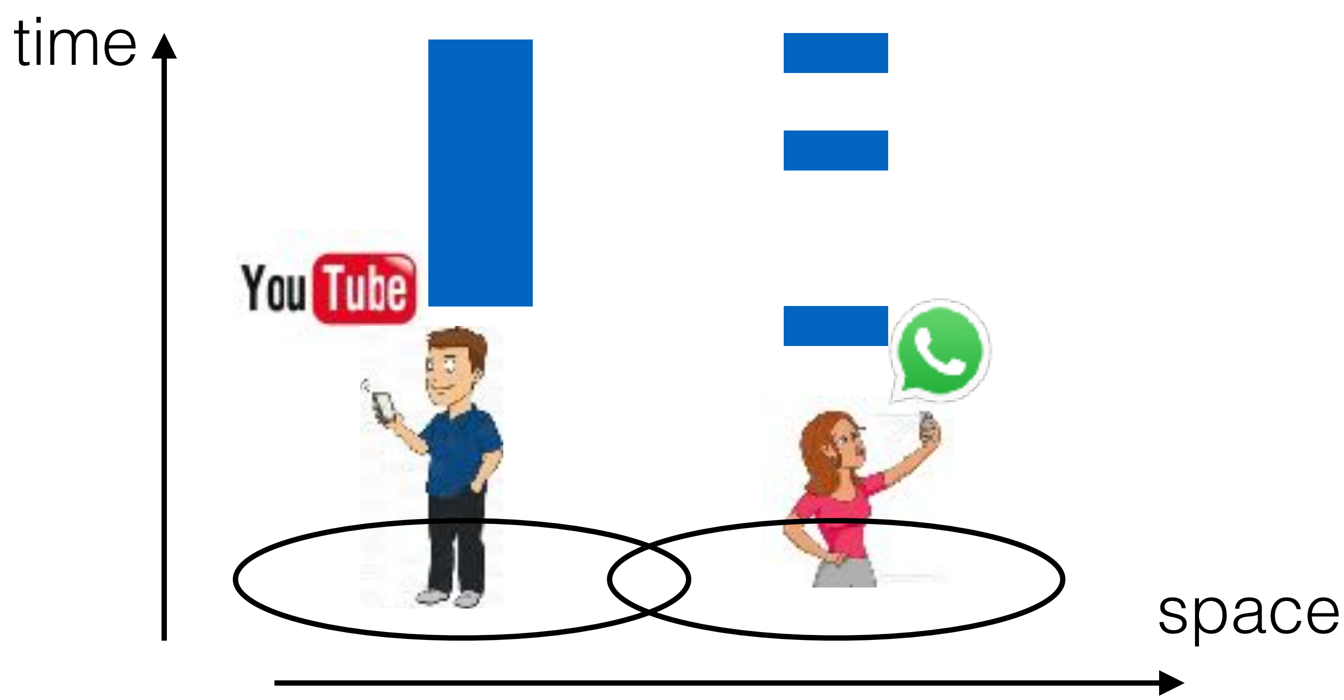
Spatio-Temporal Dynamics

Wireless Spectrum is a space-time shared resource

Spatial Component - Interference



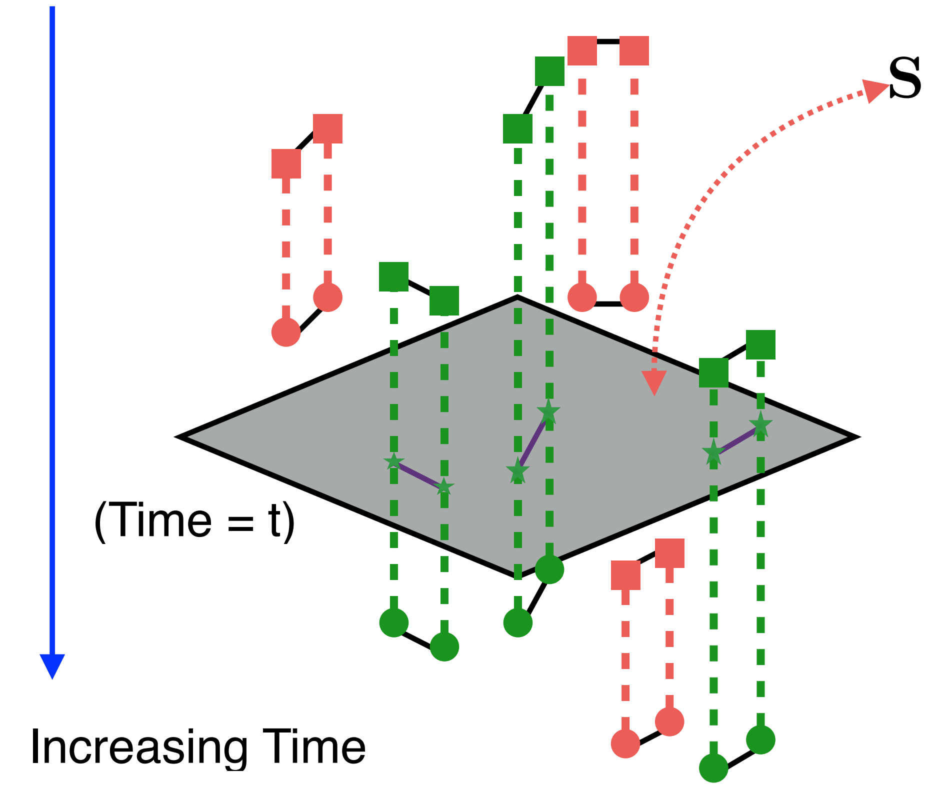
Temporal Component - Traffic Patterns



Understanding ***scalability*** properties of simple protocols

Scalability

Protocol - *A link transmits whenever they have a file by treating interference as noise*

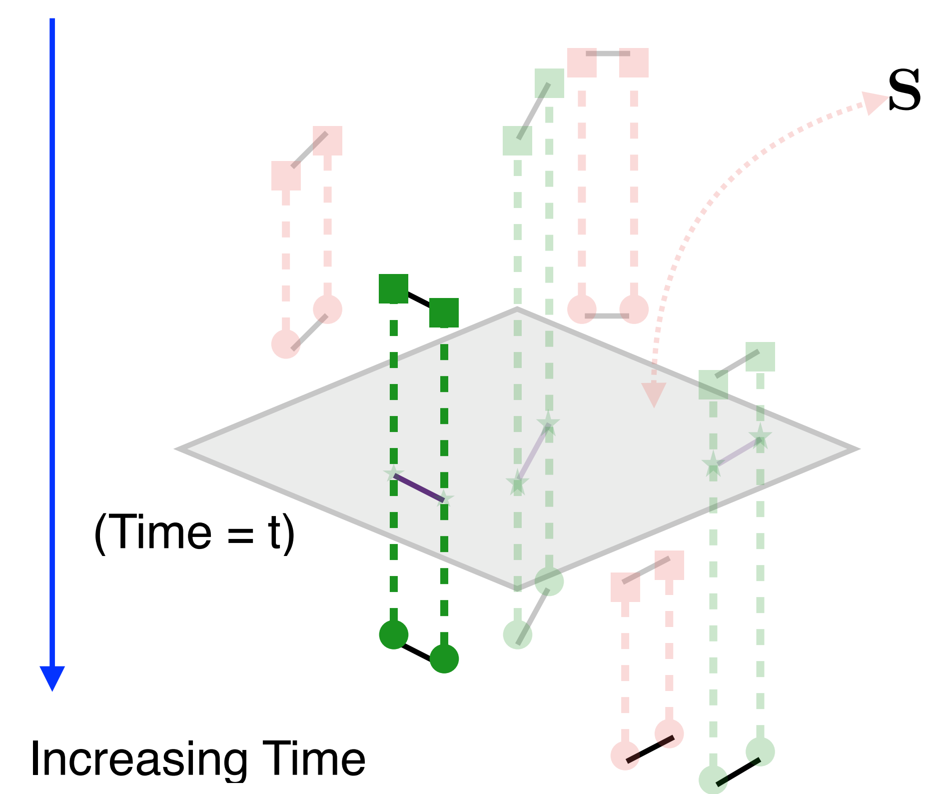


Scalability

Protocol - *A link transmits whenever they have a file by treating interference as noise*

Does performance of a typical link deteriorate with increasing network size $|\mathbf{S}|$?

No := Scalability



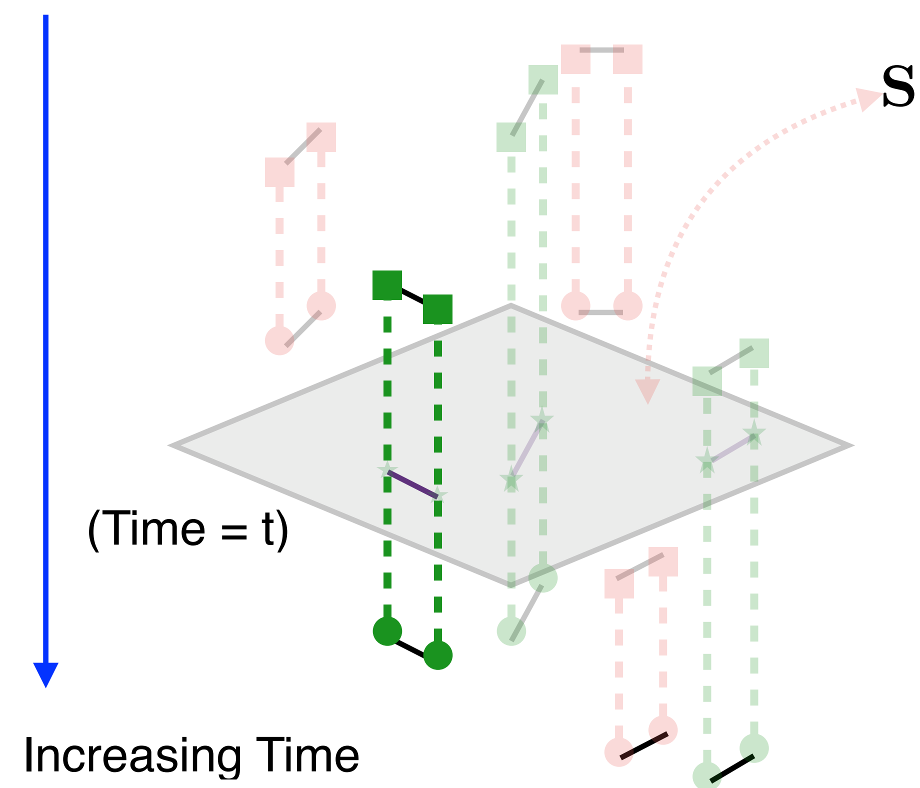
Scalability

Protocol - *A link transmits whenever they have a file by treating interference as noise*

Does performance of a typical link deteriorate with increasing network size $|\mathbf{S}|$?

No := Scalability

Long Range Space-Time correlations can disrupt scalability



Scalability

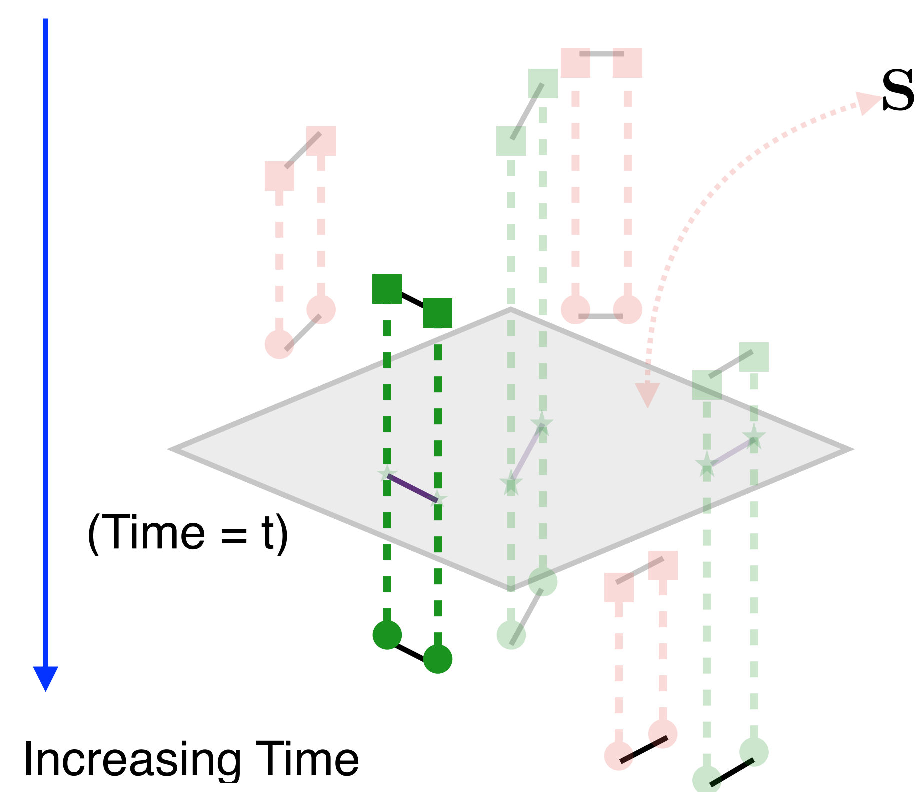
Protocol - *A link transmits whenever they have a file by treating interference as noise*

Does performance of a typical link deteriorate with increasing network size $|\mathbf{S}|$?

No := Scalability

Long Range Space-Time correlations can disrupt scalability

Scalability hard to infer from 'brute force' ray tracing simulations



Scalability

Protocol - *A link transmits whenever they have a file by treating interference as noise*

Does performance of a typical link deteriorate with increasing network size $|\mathbf{S}|$?

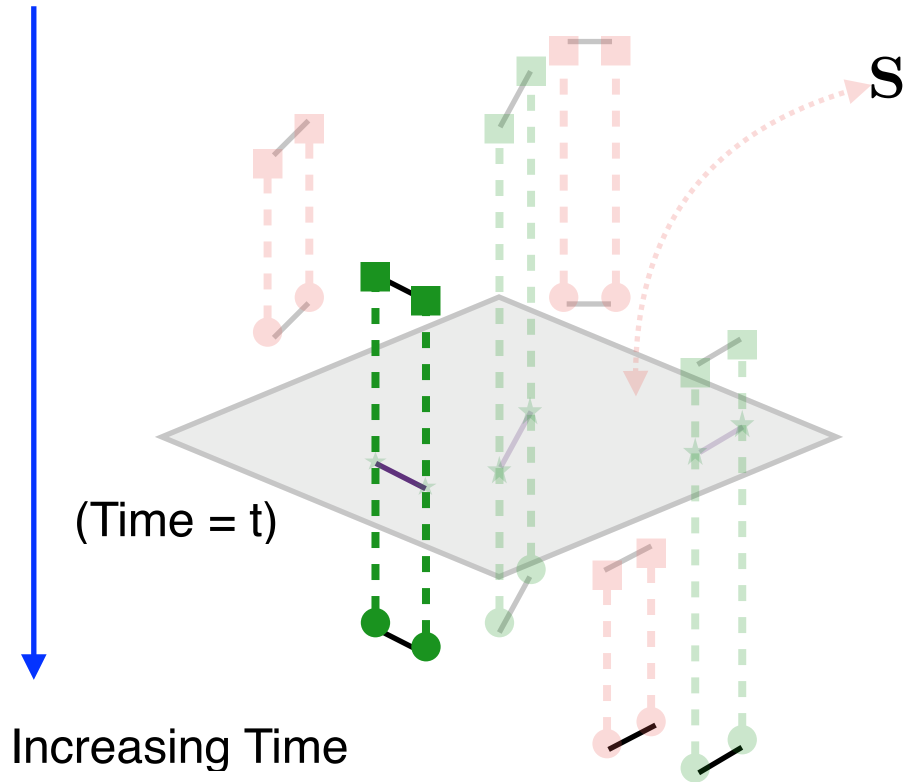
No := Scalability

Long Range Space-Time correlations can disrupt scalability

Scalability hard to infer from 'brute force' ray tracing simulations

Infinite Network - A tractable model to address such questions

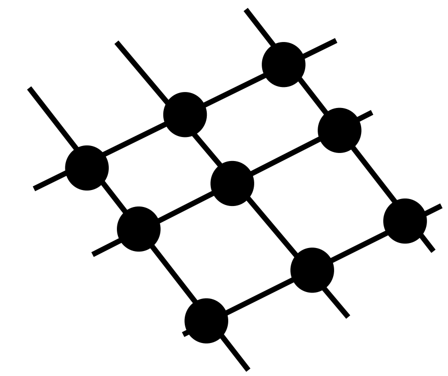
analogous to the Ising Model



Wireless Dynamics on Grids

Protocol - *A link transmits whenever they have a file by treating interference as noise*

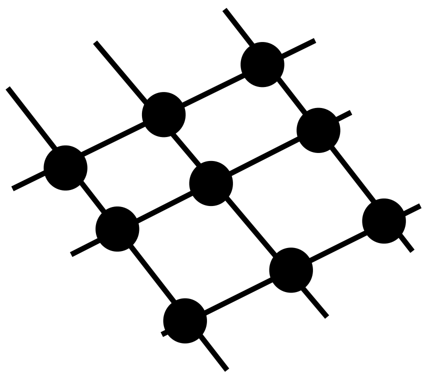
Discrete Space - d dimensional grid



Wireless Dynamics on Grids

Protocol - *A link transmits whenever they have a file by treating interference as noise*

Discrete Space - d dimensional grid



Each wireless link (Tx-Rx pair) is abstracted as a point

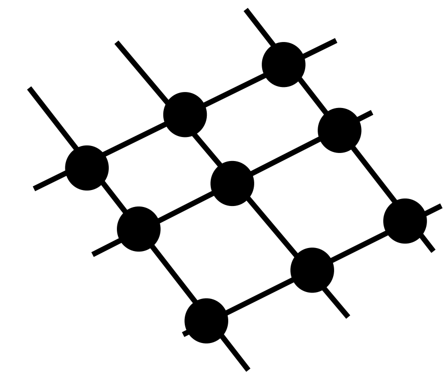
Links (points) ‘arrive’ uniformly in space and transmit

Links exit after completion of a file transfer

Wireless Dynamics on Grids

Protocol - *A link transmits whenever they have a file by treating interference as noise*

Discrete Space - d dimensional grid



Each wireless link (Tx-Rx pair) is abstracted as a point

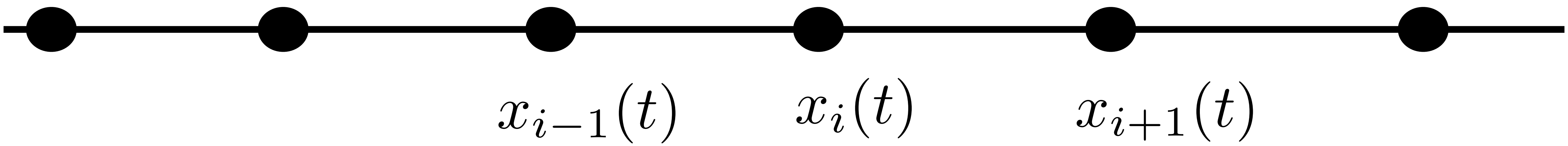
Links (points) ‘arrive’ uniformly in space and transmit

Links exit after completion of a file transfer

Instantaneous rate of transfer - Linearization of Shannon capacity formula

Interference as Noise

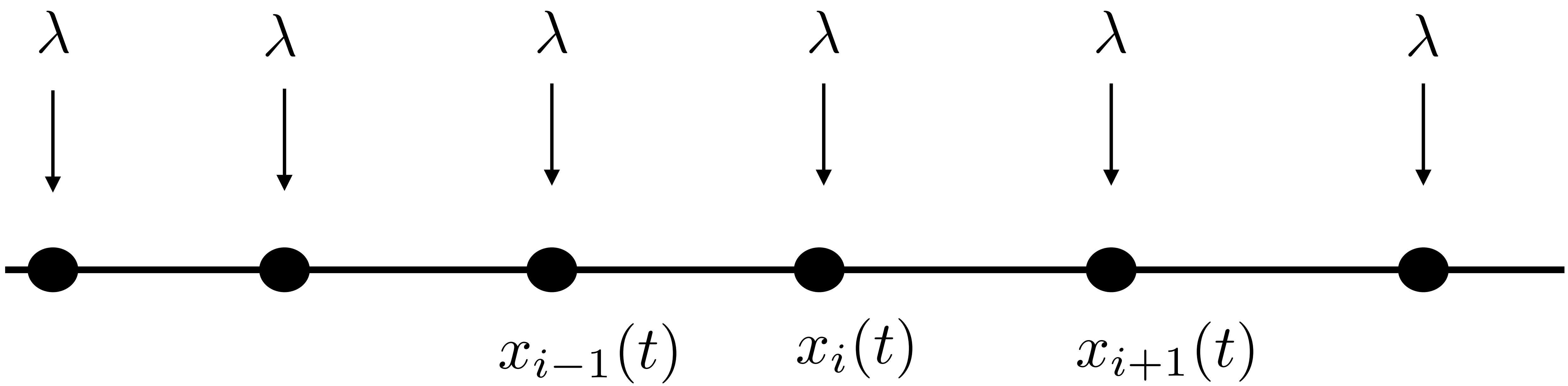
A warm up to the Model



$x_i(t) \in \mathbb{N}$ Number of links in cell $i \in \mathbb{Z}$ at time $t \geq 0$

$\{x_i(t)\}_{i \in \mathbb{Z}}$ Queue lengths at time $t \geq 0$

A warm up to the Model

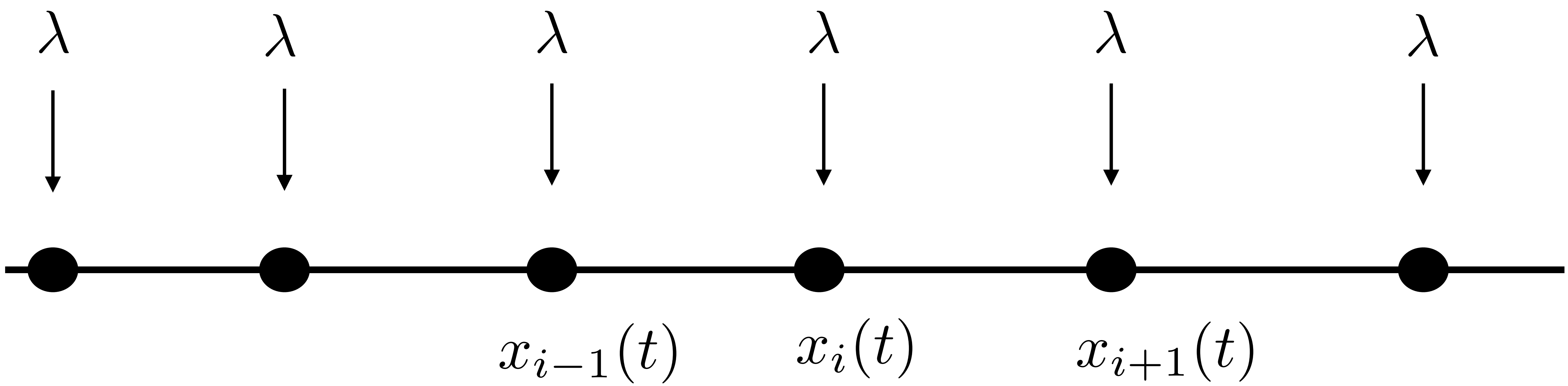


$x_i(t) \in \mathbb{N}$ Number of links in cell $i \in \mathbb{Z}$ at time $t \geq 0$

$\{x_i(t)\}_{i\in\mathbb{Z}}$ Queue lengths at time $t \geq 0$

Independent Poisson Arrivals

A warm up to the Model



$x_i(t) \in \mathbb{N}$ Number of links in cell $i \in \mathbb{Z}$ at time $t \geq 0$

$\{x_i(t)\}_{i \in \mathbb{Z}}$ Queue lengths at time $t \geq 0$

Independent Poisson Arrivals

Rate of departure from queue $i \in \mathbb{Z}$ at time t $\dfrac{x_i(t)}{x_{i-1}(t) + x_i(t) + x_{i+1}(t)}$

If 'neighboring' queues are large, instantaneous departure rate is small.

Rate of Departure - SIR

$\{x_i(t)\}_{i\in\mathbb{Z}^d} \in \mathbb{N}^{\mathbb{Z}^d}$ Queue Lengths

Rate of Departure - SIR

$\{x_i(t)\}_{i \in \mathbb{Z}^d} \in \mathbb{N}^{\mathbb{Z}^d}$ Queue Lengths

Interference Sequence $\{a_i\}_{i \in \mathbb{Z}^d}$

Rate of Departure - SIR

$\{x_i(t)\}_{i \in \mathbb{Z}^d} \in \mathbb{N}^{\mathbb{Z}^d}$ Queue Lengths

Interference Sequence $\{a_i\}_{i \in \mathbb{Z}^d}$

$a_i \geq 0 \; \forall i \in \mathbb{Z}^d$
$a_0 = 1$

Positivity

$a_i = a_{-i} \; \forall i \in \mathbb{Z}^d$

Symmetry

$L = \sup\{||i||_\infty : a_i > 0\} < \infty$

Finite Support

Interference at queue i - $\sum_{j \in \mathbb{Z}^d} a_j x_{i-j}(t)$

SIR at a customer in queue i at time t $\frac{1}{\sum_{j \in \mathbb{Z}^d} a_j x_{i-j}(t)}$

Rate of Departure - SIR

$\{x_i(t)\}_{i\in\mathbb{Z}^d} \in \mathbb{N}^{\mathbb{Z}^d}$ Queue Lengths

Interference Sequence $\{a_i\}_{i\in\mathbb{Z}^d}$

$a_i \geq 0 \ \forall i \in \mathbb{Z}^d$
$a_0 = 1$

Positivity

$a_i = a_{-i} \ \forall i \in \mathbb{Z}^d$

Symmetry

$L = \sup\{||i||_\infty : a_i > 0\} < \infty$

Finite Support

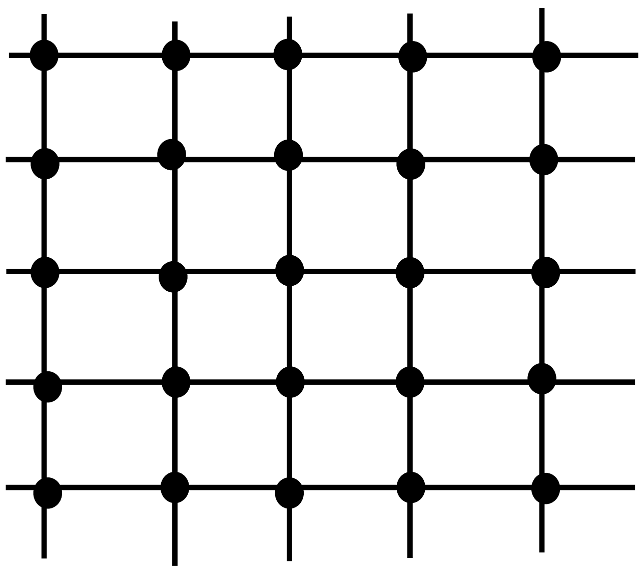
Interference at queue i - $\sum_{j\in\mathbb{Z}^d} a_j x_{i-j}(t)$

SIR at a customer in queue i at time t $\dfrac{1}{\sum_{j\in\mathbb{Z}^d} a_j x_{i-j}(t)}$

$$\boxed{\text{Rate of departure from any queue } i \text{ at time } t \quad \frac{x_i(t)}{\sum_{j\in\mathbb{Z}^d} a_j x_{i-j}(t)}}$$

Translation Invariant in Space

Interference Queueing Dynamics



$\{x_i(t)\}_{i\in\mathbb{Z}^d} \in \mathbb{N}^{\mathbb{Z}^d}$ Queue lengths at time $t \geq 0$

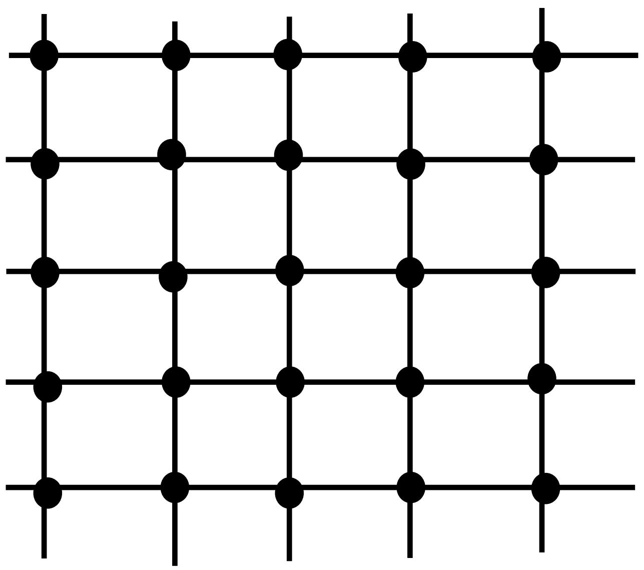
Independent rate λ Poisson arrivals

Rate of departure from queue $i \in \mathbb{Z}$ at time t $\dfrac{x_i(t)}{\sum_{j\in\mathbb{Z}^d} a_j x_{i-j}(t)}$

If 'neighboring' queues are large, instantaneous departure rate is small.

In the toy example, $a_i = 1$ if $|i| \leq 1$ and $a_i = 0$ otherwise

Interference Queueing Dynamics



$\{x_i(t)\}_{i \in \mathbb{Z}^d} \in \mathbb{N}^{\mathbb{Z}^d}$ Queue lengths at time $t \geq 0$

Independent rate λ Poisson arrivals

Rate of departure from queue $i \in \mathbb{Z}$ at time t $\frac{x_i(t)}{\sum_{j \in \mathbb{Z}^d} a_j x_{i-j}(t)}$

If 'neighboring' queues are large, instantaneous departure rate is small.

Questions -

1) For what λ and $\{a_i\}_{i \in \mathbb{Z}^d}$, is the process $\{x_i(t)\}_{i \in \mathbb{Z}^d}$ 'stable' ?

2) Characterize the steady state ??

Main Results

1. Stability

If $\lambda \sum_{j \in \mathbb{Z}^d} a_j < 1$, then system is stable

Main Results

1. Stability

If $\lambda \sum_{j \in \mathbb{Z}^d} a_j < 1$, then system is stable

2. Moments

Let $\{y_i\}_{i \in \mathbb{Z}^d}$ be the minimal stationary solution to the dynamics

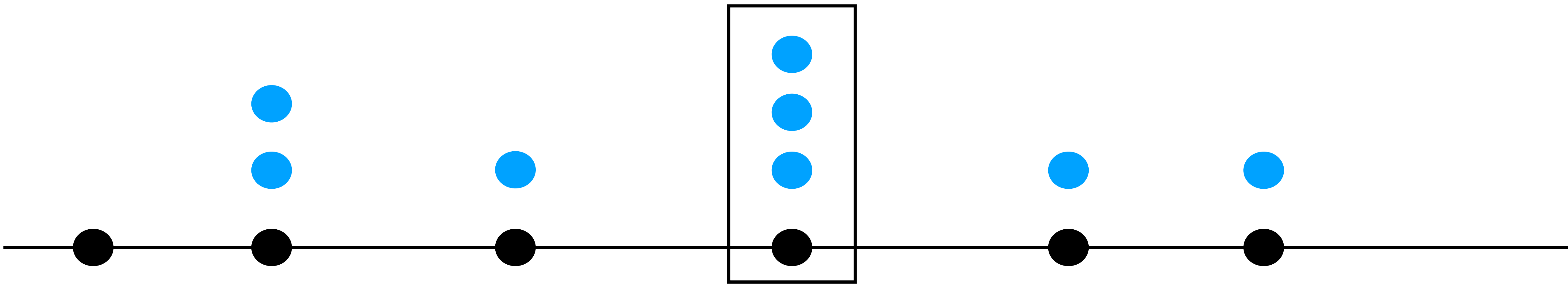
If $\lambda \sum_{j \in \mathbb{Z}^d} a_j < 1$, then $\mathbb{E}[y_0] = \dfrac{\lambda}{1 - \lambda \sum_{j \in \mathbb{Z}^d} a_j}$

If $\lambda \sum_{j \in \mathbb{Z}^d} a_j < \dfrac{2}{3}$ then $\mathbb{E}[y_0^2] < \infty$

[Shneer and Stolyar'18] established this for the entire stability range

In upcoming work, we establish exponential moments exist in the entire range

Intuition



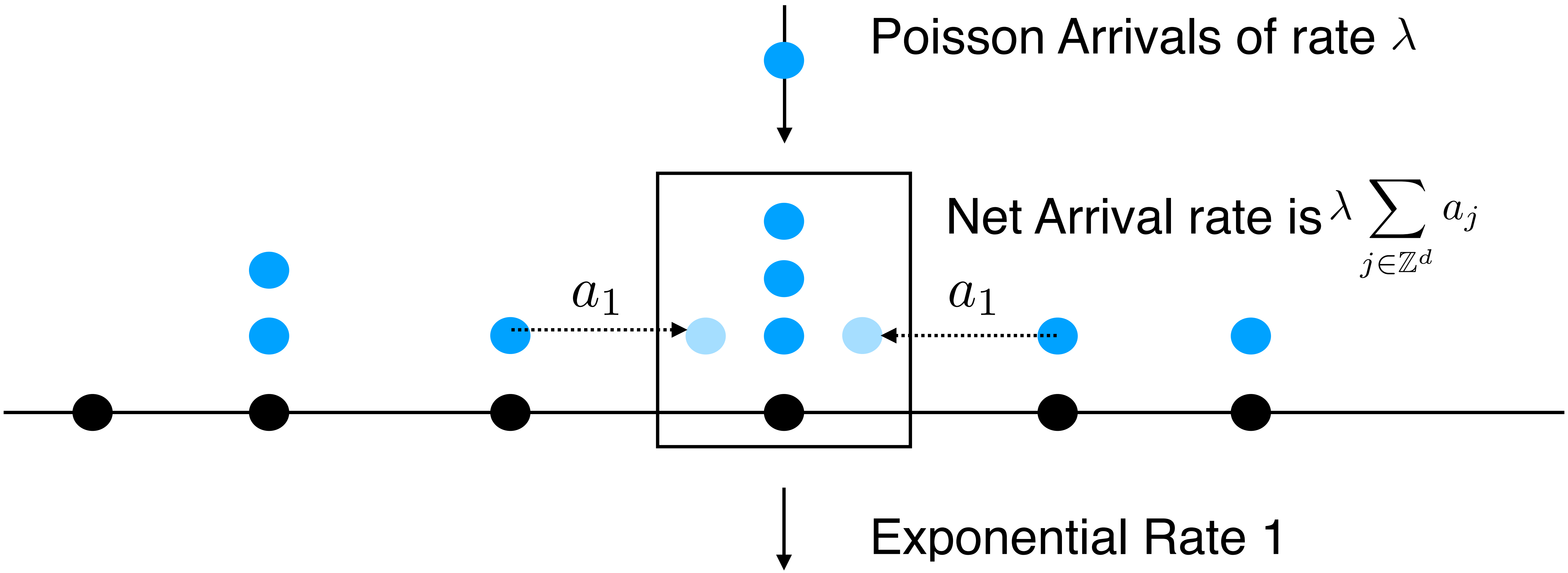
Consider any local maximum queue i, i.e. $x_i(t) = \max\{x_{i-j}(t) : a_j > 0\}$

Its instantaneous departure rate is $\dfrac{x_i(t)}{\sum_{j\in\mathbb{Z}^d} a_j x_{i-j}(t)} \geq \dfrac{1}{\sum_{j\in\mathbb{Z}^d} a_j}$

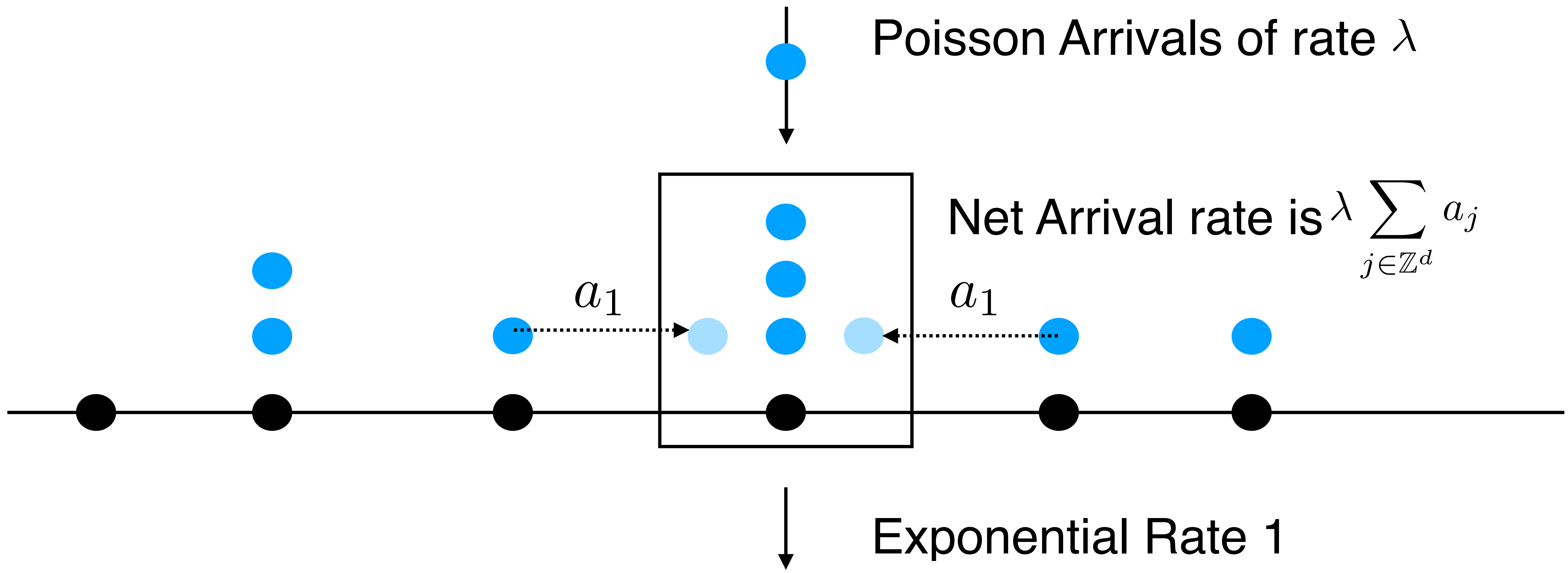
The arrival rate at every queue is λ

if $\lambda \sum_{j\in\mathbb{Z}^d} a_j < 1$, then this local maximum queue has negative drift

Intuition


Poisson Arrivals of rate λ
Net Arrival rate is $\lambda \sum_{j \in \mathbb{Z}^d} a_j$
a_1
a_1
Exponential Rate 1

Intuition



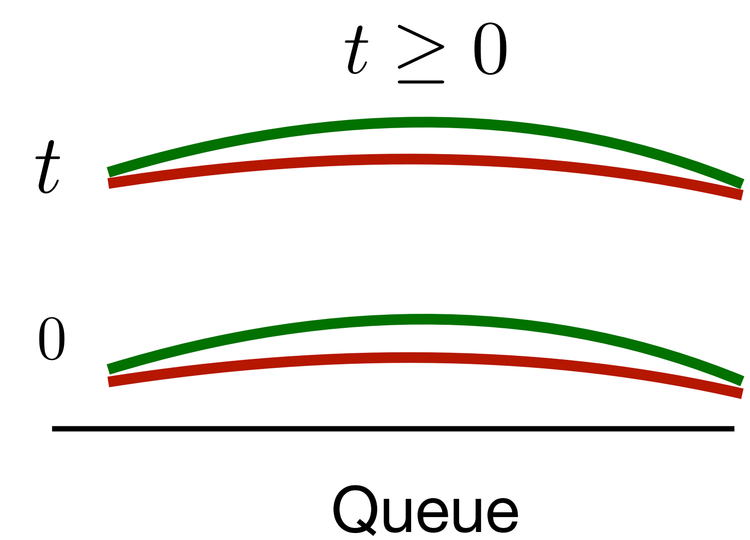
Stability - $$\lambda \sum_{j \in \mathbb{Z}^d} a_j < 1$$

Mean Queue Length - $$\frac{\lambda \sum_{j \in \mathbb{Z}^d} a_j}{1 - \lambda \sum_{j \in \mathbb{Z}^d} a_j} \cdot \frac{1}{\sum_{j \in \mathbb{Z}^d} a_j}$$

M/M/1 *Fraction of solid balls*

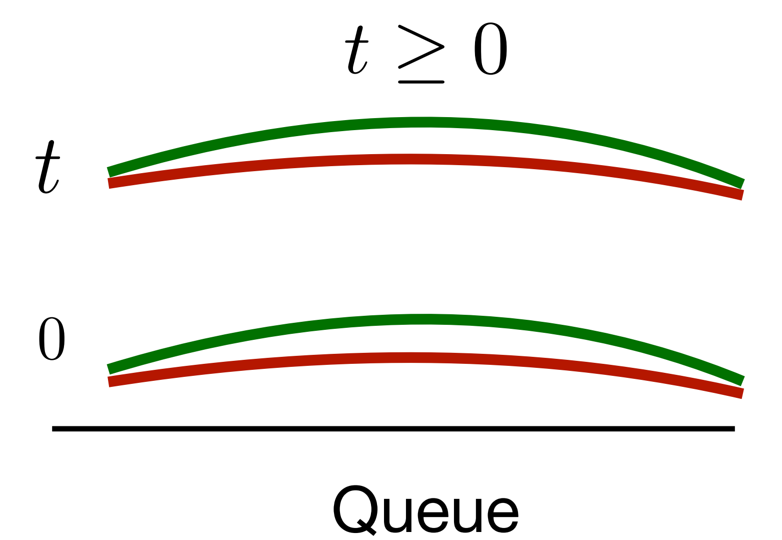
Monotonicity

If two initial conditions $\{x_i(0)\}_{i\in\mathbb{Z}^d}$ and $\{y_i(0)\}_{i\in\mathbb{Z}^d}$ s.t. for all $i \in \mathbb{Z}^d$ $x_i(0) \leq y_i(0)$, then there exists a coupling such that almost-surely $\forall t \geq 0$, $\forall i \in \mathbb{Z}^d$ $x_i(t) \leq y_i(t)$.



Monotonicity

If two initial conditions $\{x_i(0)\}_{i\in\mathbb{Z}^d}$ and $\{y_i(0)\}_{i\in\mathbb{Z}^d}$ s.t. for all $i \in \mathbb{Z}^d$ $x_i(0) \leq y_i(0)$, then there exists a coupling such that almost-surely $\forall t \geq 0$, $\forall i \in \mathbb{Z}^d$ $x_i(t) \leq y_i(t)$.

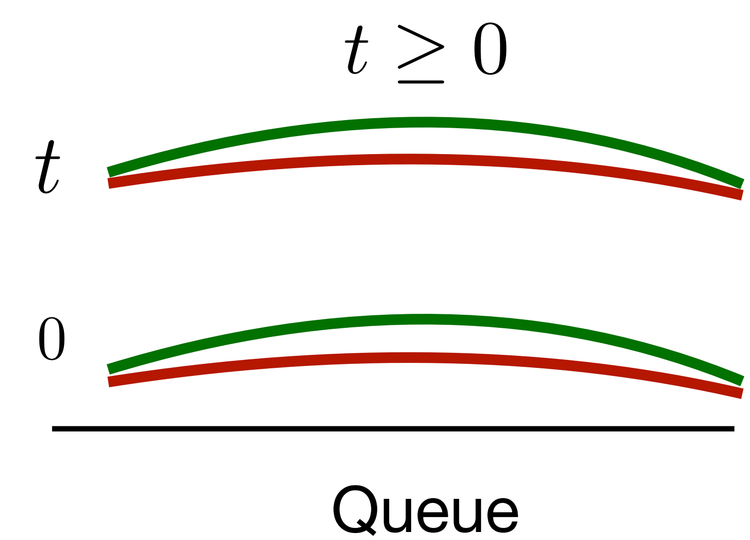


<u>Proof</u> Induction

Arrivals retain the ordering

Monotonicity

If two initial conditions $\{x_i(0)\}_{i\in\mathbb{Z}^d}$ and $\{y_i(0)\}_{i\in\mathbb{Z}^d}$ s.t. for all $i \in \mathbb{Z}^d$ $x_i(0) \leq y_i(0)$, then there exists a coupling such that almost-surely $\forall t \geq 0$, $\forall i \in \mathbb{Z}^d$ $x_i(t) \leq y_i(t)$.



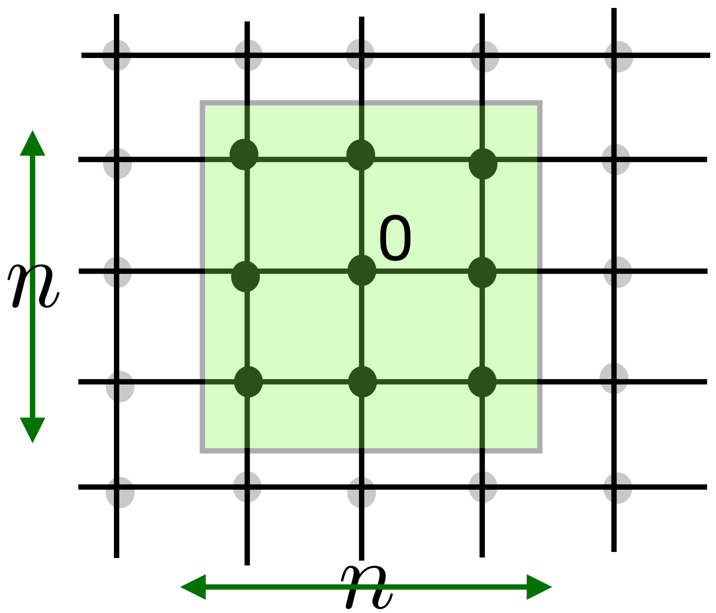
<u>Proof</u> Induction

Arrivals retain the ordering

Two queues are equal - higher interference system has smaller departure

Unequal queues - Retains ordering as at-most one customer departs

Proof Steps



1. Consider a spatial truncation - finite dimensional

2. If $\lambda \sum_{j \in \mathbb{Z}^d} a_j < 1$ => Stability

   Max queue length - Lyapunov function

3. Rate Conservation Principle

$\lambda \sum_{j \in \mathbb{Z}^d} a_j < 1$ => $\mathbb{E}[y_0^{(n)}] = \frac{\lambda}{1 - \lambda \sum_{j \in \mathbb{Z}^d} a_j} - o_n(1)$ *Tightness of* $\{y_0^{(n)}\}_{n \in \mathbb{N}}$

4. Switch of limits in time and space *Coupling from the past*

5. Monotone Convergence to yield the moment formula

Large Initial Conditions

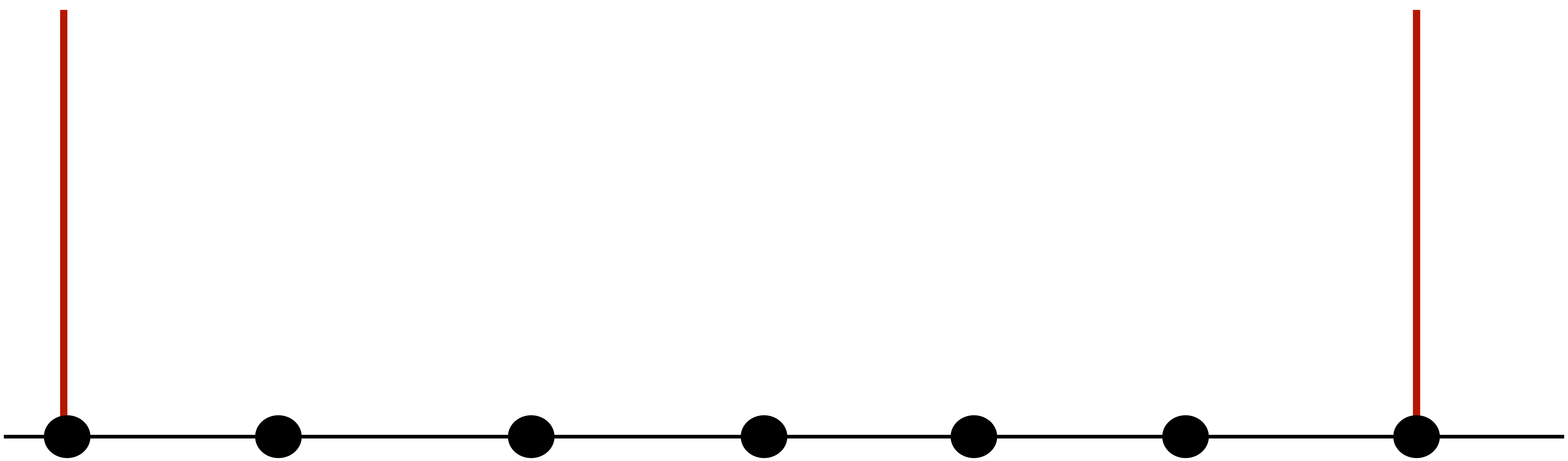
Theorem

For every λ, there exists a probability distribution on $\mathbb{N}$ such that if the initial condition is $\{x_i(0)\}_{i\in\mathbb{Z}^d}$ i.i.d. from this distribution, then $\forall i \in \mathbb{Z}^d$, $\lim_{t\to\infty} x_i(t) = \infty$ almost-surely.

Large Initial Conditions

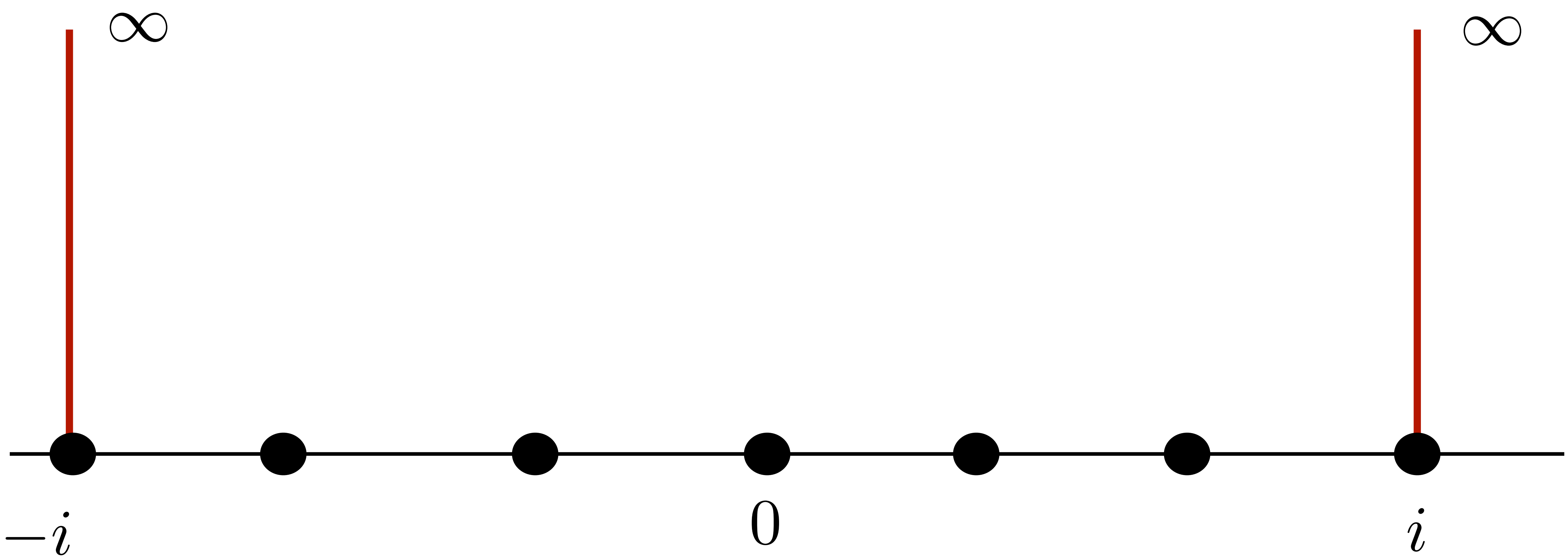
Theorem

For every λ, there exists a probability distribution on $\mathbb{N}$ such that if the initial condition is $\{x_i(0)\}_{i\in\mathbb{Z}^d}$ i.i.d. from this distribution, then $\forall i \in \mathbb{Z}^d$, $\lim_{t\to\infty} x_i(t) = \infty$ almost-surely



If “large” frozen boundary is present, then stationary queue length at 0 is also “large” with “high probability”

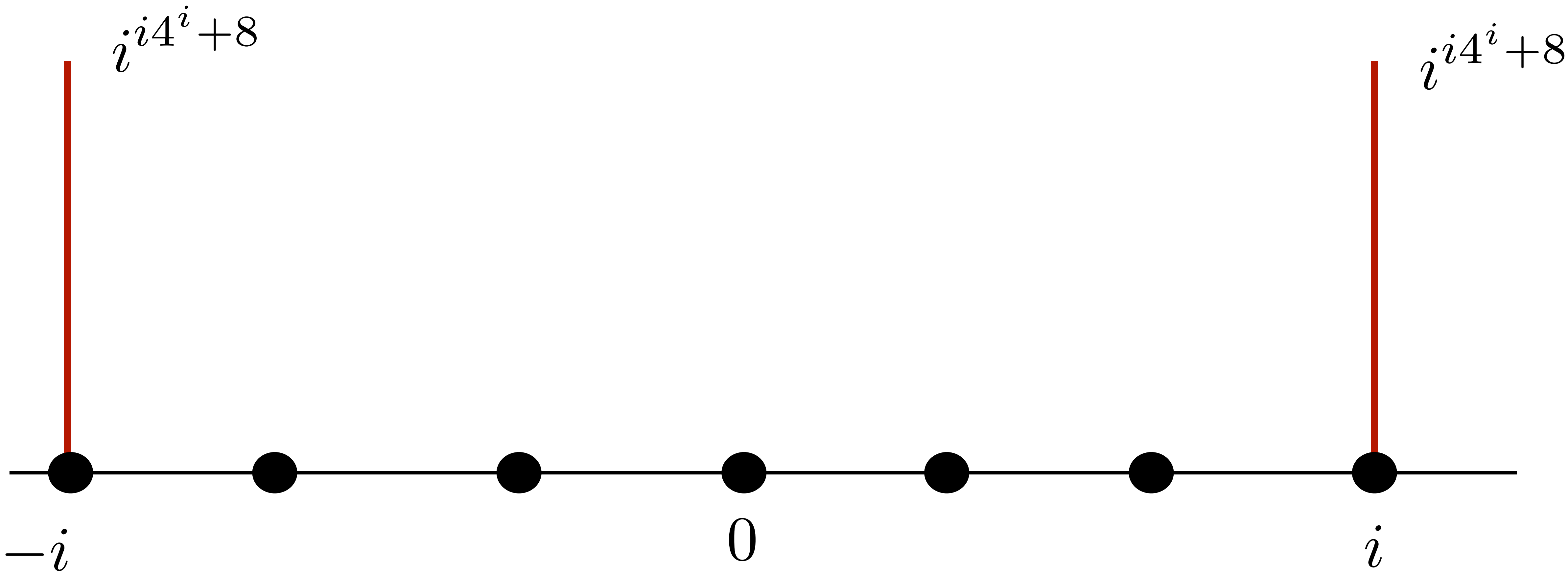
Convergence to Stationary Solutions



$$\exists (t_i)_{i\in\mathbb{N}} \text{ s.t. } t_i \to \infty \text{ s.t. } \mathbb{P}[x_0(t_i) < i] \le i^{-4}$$

Because of the infinite barrier, all queues diverge to infinity at a linear rate

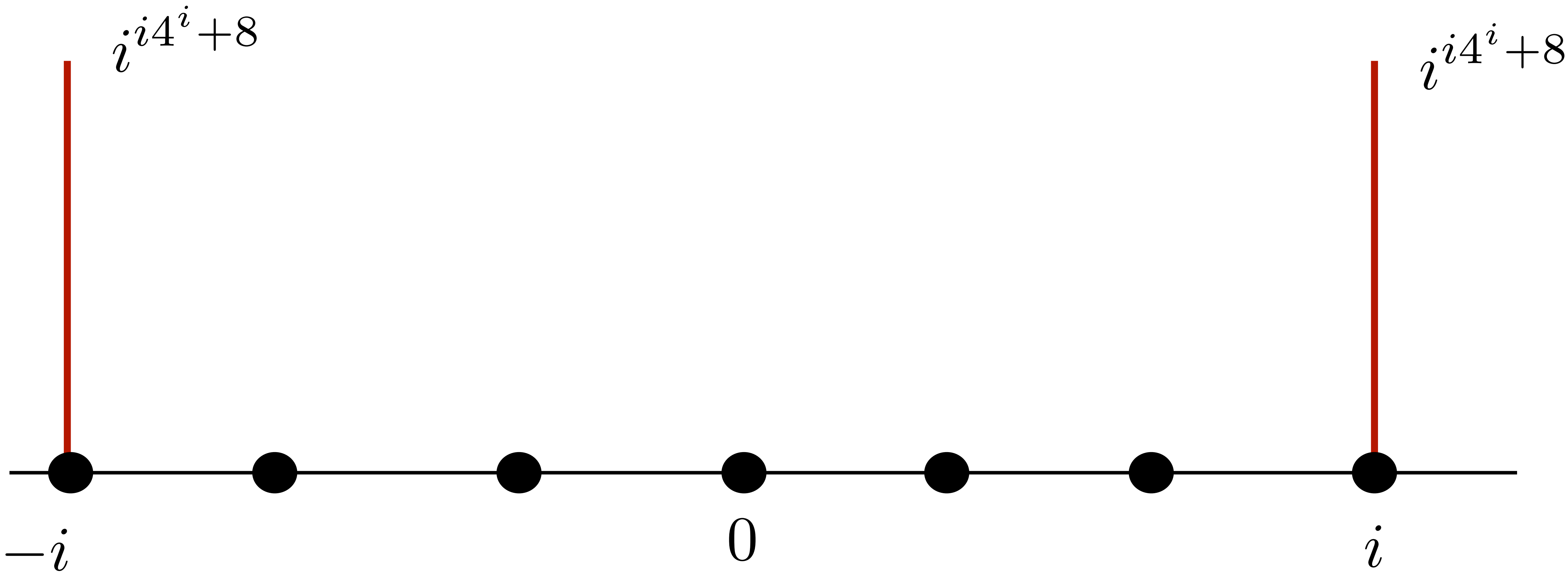
Convergence to Stationary Solutions



$$\exists (t_i)_{i\in\mathbb{N}} \text{ s.t. } t_i \to \infty \text{ s.t. } \mathbb{P}[x_0(t_i) < i] \le i^{-3}$$

Since interested only in finite time t_i, can bring down the barrier to a finite value at a small penalty in probability

Convergence to Stationary Solutions



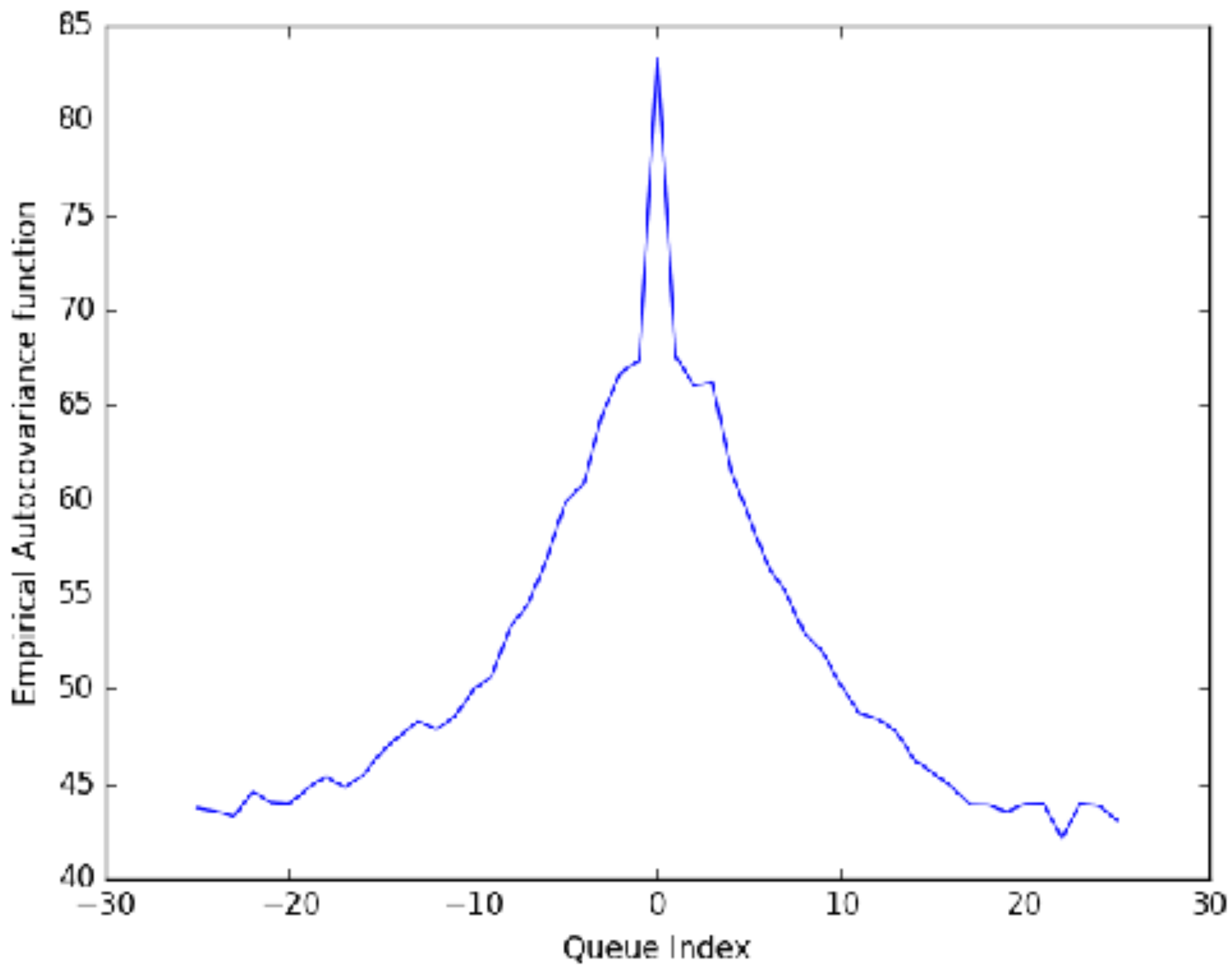
$$\exists (t_i)_{i\in\mathbb{N}} \text{ s.t. } t_i \to \infty \text{ s.t. } \mathbb{P}[x_0(t_i) < i] \leq i^{-3}$$

Since interested only in finite time t_i, can bring down the barrier to a finite value at a small penalty in probability

Borel-Cantelli to conclude the proof

Open Questions

If $\lambda \sum_{j \in \mathbb{Z}^d} a_j < 1$, then what moments of $x_{0,\infty}(0)$ exists ? How do correlations decay ? i.e., how does $k \to \mathbb{E}[x_0 x_k] - (\mathbb{E}[x_0])^2$ decay ?



d=1, n=51

$\lambda = 0.1419, \lambda_c = 1/7$

No propagation of chaos even in an infinite system !

Open Questions

Uniqueness of Stationary Solution

Existence/construction of other non-degenerate stationary solutions ?

Convergence to Stationary Solution

Do other initial conditions apart from all empty converge to a stationary limit ?

Prediction of bad outage events propagating from 'far out' in space

Summary of the Talk

Two problems in networking

- Introduced new mathematical models and questions
- Demonstrate the effectiveness of the model

  Contextual Data in Graph Clustering
  Scalability of wireless protocols

- New tools and techniques in the analysis of the proposed models

Papers

1. *Social Learning in Multi Agent Multi Armed Bandits with* S. Shakkottai and A. Ganesh, Preprint, 2019

2. *Interference Queueing Networks on Grids* with F. Baccelli and S. Foss, In Annals of Applied Probability, to appear

3. *Community Detection on Euclidean Random Graphs*, with F. Baccelli and E. Abbe
   Journal version under submission
   In ACM-SIAM Symposium on Discrete Algorithms (SODA), 2018
   Extended Abstract in Allerton, 2017

4. *Spatial Birth-Death Wireless Networks,* with F. Baccelli
   In IEEE Transactions on Information Theory, 2017.
   Extended abstract in Allerton, 2016

5. *Performance-oriented association in large cellular networks with technology diversity*, with F. Baccelli and J. Woo Cho,
   In International Teletrafic Congress (ITC 28), 2016

6. *CSMA k-SIC : A Class of Distributed MAC protocols and their performance evaluation*, with F. Baccelli,
   In INFOCOM, 2015

Acknowledgements

Thank You

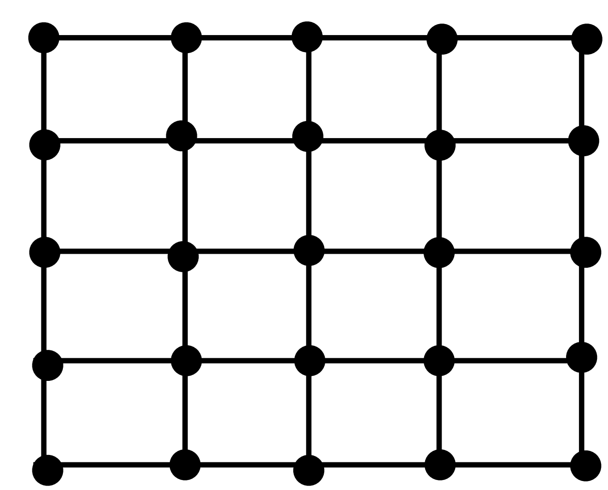
Main Proof Idea - Stability

Two systems on $B \subset \mathbb{Z}^d$ with the same dynamics.
All queues in $B^{\complement}$ are frozen without activity.

- $\{y_i(t)\}_{i \in B}$: the set B is a torus.

- $\{z_i(t)\}_{i \in B}$: the set B has boundary effects.

Interference is lower at the boundaries.



B

Main Proof Idea - Stability

Two systems on $B \subset \mathbb{Z}^d$ with the same dynamics.
All queues in $B^{\complement}$ are frozen without activity.

- $\{y_i(t)\}_{i \in B}$: the set B is a torus.

- $\{z_i(t)\}_{i \in B}$: the set B has boundary effects.

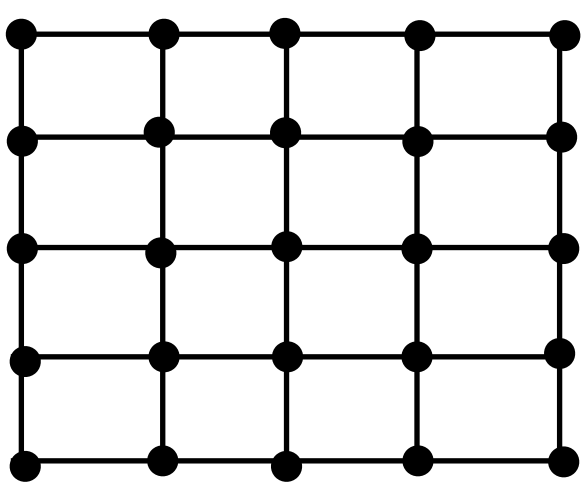
Interference is lower at the boundaries.

$\forall t \ \forall i \in B$

1) $x_i(t) \geq z_i(t)$

2) $y_i(t) \geq z_i(t)$

Monotonicity



B

Finite Torus System

$\{y_i(t)\}_{i \in B}$ process on a torus.

<u>Theorem</u> - If $\lambda \sum_{j \in \mathbb{Z}^d} a_j < 1$, then $\{y_i(t)\}_{i \in B}$ is Positive Recurrent and the stationary distribution possess exponential moments. Furthermore, the mean queue length satisfies $\mathbb{E}[y_0(t)] = \frac{\lambda}{1 - \lambda \sum_{j \in \mathbb{Z}^d} a_j}$

Finite Torus System

$\{y_i(t)\}_{i\in B}$ process on a torus.

<u>Theorem</u> - If $\lambda \sum_{j\in\mathbb{Z}^d} a_j < 1$, then $\{y_i(t)\}_{i\in B}$ is Positive Recurrent and the stationary distribution possess exponential moments. Furthermore, the mean queue length satisfies $\mathbb{E}[y_0(t)] = \dfrac{\lambda}{1-\lambda\sum_{j\in\mathbb{Z}^d} a_j}$

<u>Proof Idea of Stability</u>

$$\frac{d}{dt} y_i = \lambda - \frac{y_i}{\sum_{j\in\mathbb{Z}^d} a_j y_{(i-j)/B}(t)}$$

Fluid scale equation

Consider the maximal queue $i^*(t) := \arg\max_{i\in B} y_i(t)$

$$\frac{d}{dt} y_{i^*(t)} = \lambda - \frac{y_{i^*(t)}}{\sum_{j\in\mathbb{Z}^d} a_j y_{i^*(t)-j}(t)}$$

This has negative drift

$$\leq \lambda - \frac{1}{\sum_{j\in\mathbb{Z}^d} a_j} < -\epsilon$$

Can upper bound by a stable Single server queue.

Finite Torus System

Rate Conservation - "On Average what comes in is what goes out".

For Ex. $$\lambda = \mathbb{E}\left[\frac{y_0(t)}{\sum_{j \in \mathbb{Z}^d} a_j y_{j/B}(t)} \mathbf{1}_{y_0(t)>0}\right]$$

Avg arrival rate equals avg departure rate.

<u>Key Idea:</u>

Consider $I(t) := y_0(t) \sum_{j \in \mathbb{Z}^d} a_j y_j(t)$ in stationarity and solve $\frac{d}{dt}\mathbb{E}[I(t)] = 0$

Average increase due to arrivals - $\lambda + \lambda(\sum_{j \in \mathbb{Z}^d} a_j)\mathbb{E}[y_0(t)]$

Average decrease due to departures - $\mathbb{E}[y_0(t)]$

Equating the two yields $$\mathbb{E}[y_0(t)] \in \left\{\frac{\lambda}{1 - \lambda \sum_{j \in \mathbb{Z}^d} a_j}, \infty\right\}$$

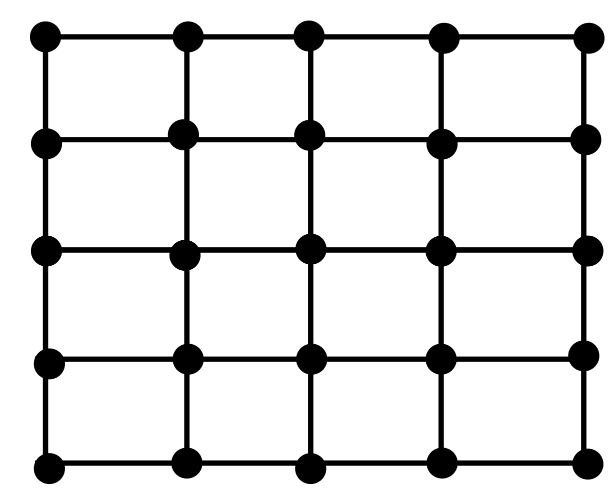
Coupling From the Past

$\{z_i(t)\}_{i \in B}$, process where the set B has boundary effects.

Monotonicity => $x_i(t) \geq z_i(t)$ and $y_i(t) \geq z_i(t)$

Thus $\mathbb{E}[z_0(t)] \leq \dfrac{\lambda}{1 - \lambda \sum_{j \in \mathbb{Z}^d} a_j}$ ***Uniformly in the size of*** B

Consider $B_n \nearrow \mathbb{Z}^d$ and corresponding stationary $z_0^{(n)}(0)$

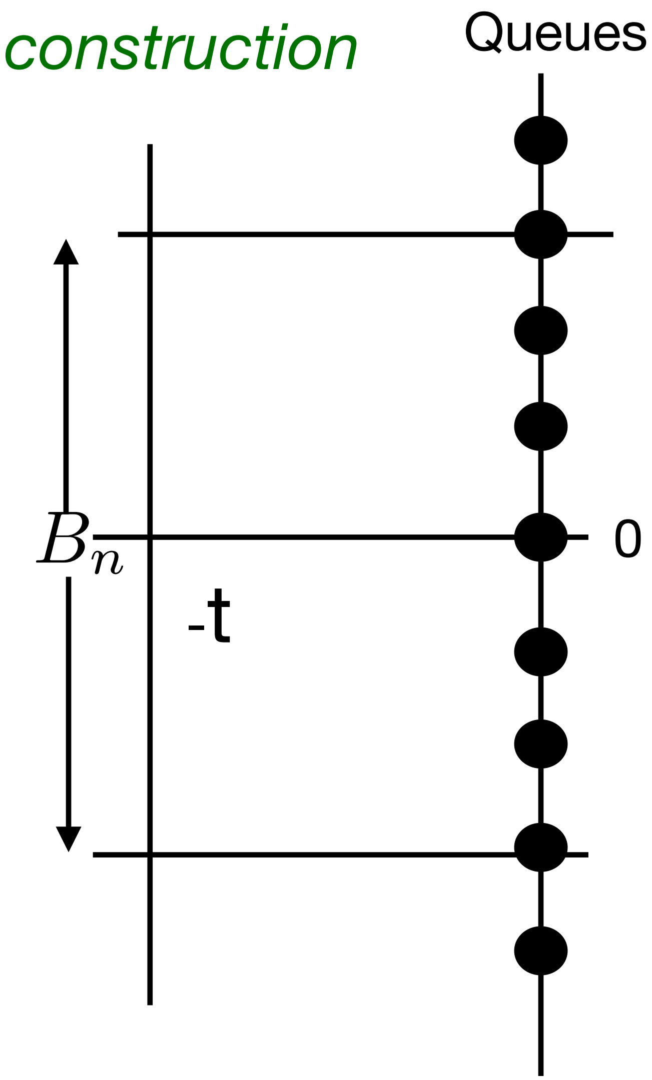


B

Coupling From the Past

Let $B_n \nearrow \mathbb{Z}^d$. $z_{0,t}^{(n)}(0)$ - the queue length of queue 0 at time 0, when the truncated B_n system is started empty at time -t.

Notice $\forall t \geq 0 \quad \lim_{n\to\infty} z_{0,t}^{(n)}(0) = x_{0,t}(0)$ *Corollary of the construction*



Monotonicity =>

$\lim_{t\to\infty} z_{0,t}^{(n)} := z_{0,\infty}^{(n)}$ and $\lim_{n\to\infty} z_{0,\infty}^{(n)} := z_{0,\infty}^{(\infty)}$ a.s.

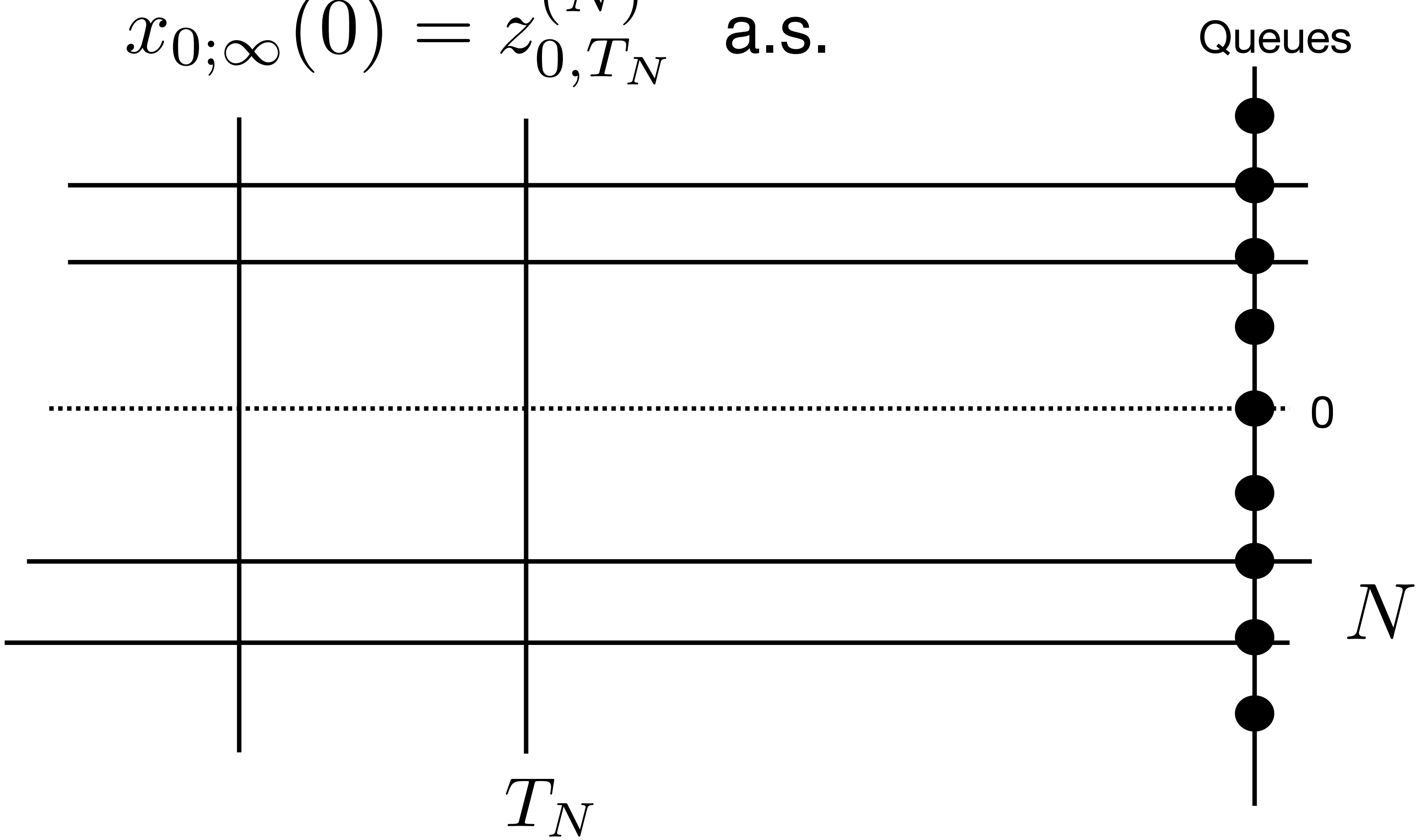
We know $\sup_{n\in\mathbb{N}} \mathbb{E}[z_{0,\infty}^{(n)}] \leq \dfrac{\lambda}{1-\lambda\sum_{j\in\mathbb{Z}^d} a_j}$

thus, $\mathbb{E}[z_{0,\infty}^{(\infty)}] < \infty$

Coupling From the Past

<u>Lemma</u> - If $\lambda \sum_{j \in \mathbb{Z}^d} a_j < 1$, then $\exists N \in \mathbb{N}$ and $\exists T_N < \infty$ random such that

$x_{0;\infty}(0) = z_{0,T_N}^{(N)}$ a.s.



We know $\sup_{n \in \mathbb{N}} \mathbb{E}[z_{0,\infty}^{(n)}] \leq \frac{\lambda}{1 - \lambda \sum_{j \in \mathbb{Z}^d} a_j}$. Thus $\mathbb{E}[x_{0,\infty}(0)] \leq \frac{\lambda}{1 - \lambda \sum_{j \in \mathbb{Z}^d} a_j}$

$\blacksquare$

Monotonicity

If two initial conditions $\{x_i(0)\}_{i\in\mathbb{Z}^d}$ and $\{y_i(0)\}_{i\in\mathbb{Z}^d}$ s.t. for all $i \in \mathbb{Z}^d$

$x_i(0) \leq y_i(0)$ are coupled with the same driving sequence $(\mathcal{A}_i, \mathcal{D}_i)_{i\in\mathbb{Z}^d}$

then $\forall t \geq 0$, $\forall i \in \mathbb{Z}^d$ $x_i(t) \leq y_i(t)$

<u>Proof</u> Induction on events.

Arrivals retain the ordering.

Consider a potential departure event at queue i.

If $x_i(t) \leq y_i(t) + 1$, then ordering retained as exactly 1 departure.

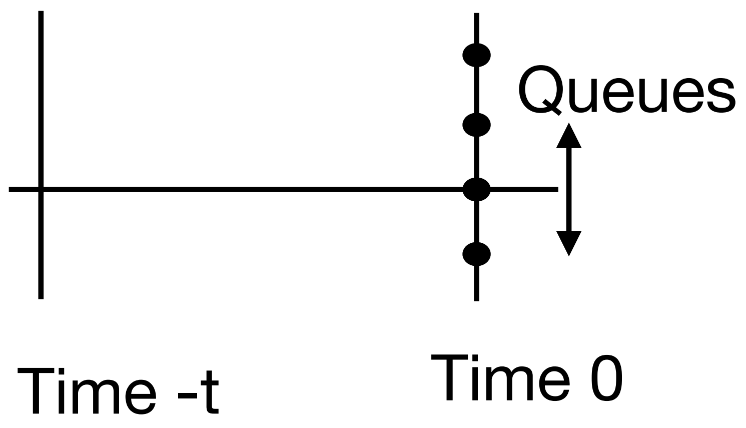
If $x_i(t) = y_i(t)$, then the departure probabilities are ordered by induction hypothesis

$$\frac{x_i(t)}{\sum_{j\in\mathbb{Z}^d} a_j x_{i-j}(t)} \geq \frac{y_i(t)}{\sum_{j\in\mathbb{Z}^d} a_j y_{i-j}(t)}$$

$\blacksquare$

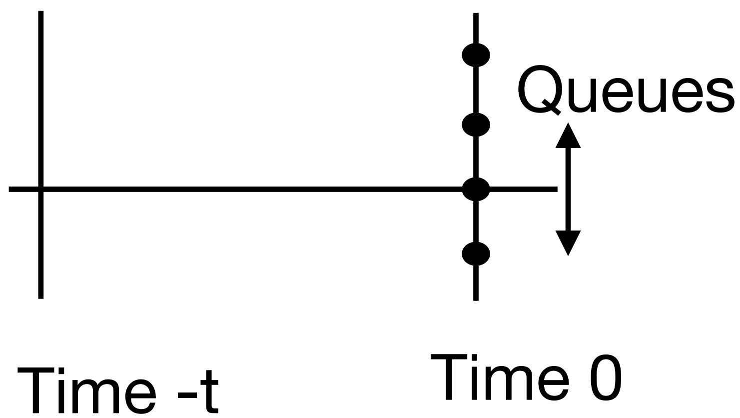
Stability Definition - Backward Construction

$x_{i;t}(0)$ Queue length of i at time 0 ***given*** the entire system was started empty at time -t



Stability Definition - Backward Construction

$x_{i;t}(0)$ Queue length of i at time 0 ***given*** the entire system was started empty at time -t

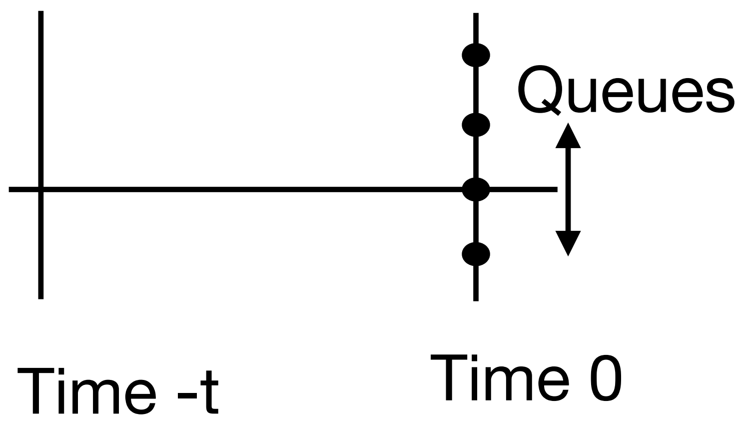


Monotonicity => $t \to x_{i;t}(0)$ is non-decreasing

$$x_{i,\infty}(0) := \lim_{t\to\infty} x_{i;t}(0) \quad \text{a.s.}$$

Stability Definition - Backward Construction

$x_{i;t}(0)$ Queue length of i at time 0 ***given*** the entire system was started empty at time -t



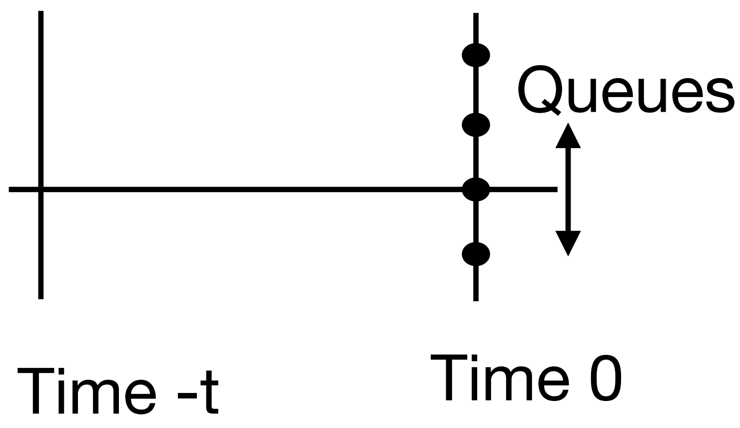
Monotonicity => $t \to x_{i;t}(0)$ is non-decreasing

$$x_{i,\infty}(0) := \lim_{t\to\infty} x_{i;t}(0) \quad \text{a.s.}$$

$\{x_{i,\infty}(0)\}_{i\in\mathbb{Z}^d}$ is a *minimal stationary solution* to the dynamics.

Stability Definition - Backward Construction

$x_{i;t}(0)$ Queue length of i at time 0 ***given*** the entire system was started empty at time -t



Monotonicity => $t \to x_{i;t}(0)$ is non-decreasing

$$x_{i,\infty}(0) := \lim_{t \to \infty} x_{i;t}(0) \quad \text{a.s.}$$

<u>0-1 Law</u> $\mathbb{P}[\cap_{i \in \mathbb{Z}^d} x_{i,\infty}(0) < \infty] \in \{0, 1\}$

$\boxed{\text{If } x_{i,\infty}(0) < \infty \text{ a.s.} \Rightarrow \underline{\text{System is stable}}}$

$\{x_{i,\infty}(0)\}_{i \in \mathbb{Z}^d}$ is a *minimal stationary solution* to the dynamics.