Asymptotic Laws for Maximum Coloring of Sparse Random Geometric Graphs

Milan Bradonjić

Bell Labs, USA
joint work with Sem Borst

SGA 2016 in honor of prof. Dragoš Cvetković

(i) random geometric graphs

(ii) sparse regime

(iii) coloring

(iv) constant number of colors

(v) asymptotic laws

Random Geometric Graphs

Definition ("uniform model")

Let $\mathcal{I}_n = \{x_1, x_2, \ldots, x_n\}$ be n points uniformly and independently distributed in $[0,1]^d$. The random geometric graph has the node set $\mathcal{I}_n$, and the edge set where every two nodes are adjacent if within distance $||x_i - x_j|| \leq r(n)$.

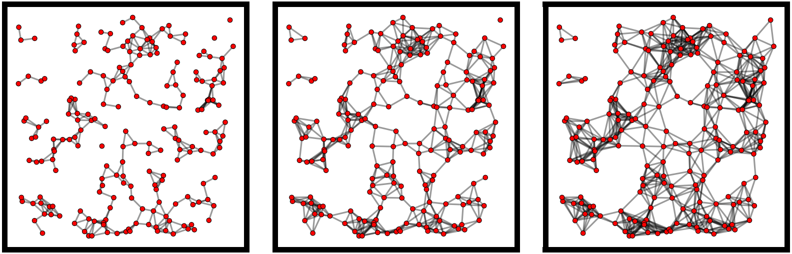


Figure : Number of nodes $n = 200$ and $r' = 0.075$, $r'' = 0.1$, $r''' = 0.125$. As r increases, the graph evolves (in the number of edges).

Random Geometric Graphs

Definition ("Poisson model")

Let $\mathcal{X}_\lambda = \{x_1, x_2, x_3, \dots\}$ be a Poisson point process in $\mathbb{R}^d$ with intensity $\lambda > 0$. Let $n \in \mathbb{N}$. The random geometric graph has the node set $\mathcal{X}_\lambda \cap [0, n^{1/d}]^d$, and the edge set where every two nodes are adjacent if within distance $||x_i - x_j|| \le r(n)$.

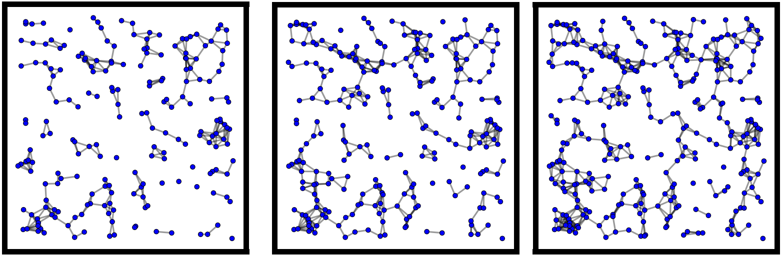
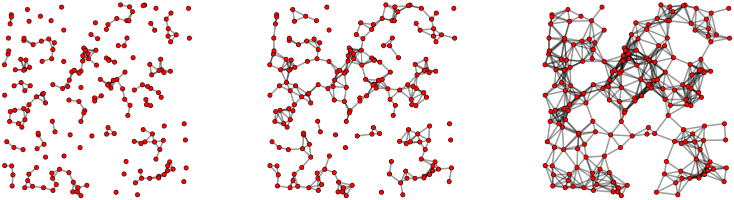


Figure : volume 200 and $r = 1$, densities: $\lambda' = 1$, $\lambda'' = 1.25$, $\lambda''' = 1.5$. As λ increases, the graph evolves (both the number of nodes and edges).

The structure of RGGs

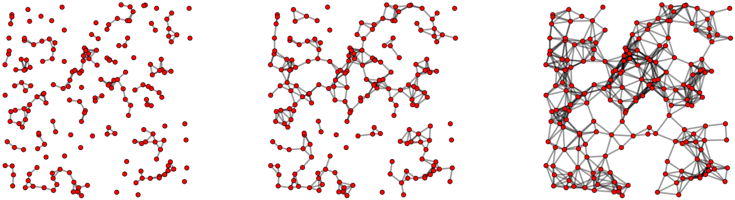
As r increases, two sharp thresholding phenomena appear w.h.p.



- $n \cdot r_c(n)^d = \lambda_c$ (Fig. 2) the largest component giant of order n (sparse regime; constant degree)
- $n \cdot r_t(n)^d = \gamma_d^{-1} \log n$ (Fig. 3) connectedness (dense regime)
- λ_c not known!
- dim=2: experimentally $\lambda_c \approx 1.44$ Quintanilla, Torquato '97; exact bounds $\lambda_c \in [0.696, 3.372]$ Meester, Roy '96; improvement $\lambda_c > 4/(3\sqrt{3}) \approx 0.7698$ Kong, Yeh '06.

The structure of RGGs

As r increases, two sharp thresholding phenomena appear w.h.p.



- $n \cdot r_c(n)^d = \lambda_c$ (Fig. 2) the largest component giant of order n (sparse regime; constant degree)
- $n \cdot r_t(n)^d = \gamma_d^{-1} \log n$ (Fig. 3) connectedness (dense regime)
- λ_c not known!
- dim=2: experimentally $\lambda_c \approx 1.44$ Quintanilla, Torquato '97; exact bounds $\lambda_c \in [0.696, 3.372]$ Meester, Roy '96; improvement $\lambda_c > 4/(3\sqrt{3}) \approx 0.7698$ Kong, Yeh '06.

Why RGGs?

Good model for:

1. Wireless networks
   two radios can communicate only if within range of each other
2. Relational data-sets
   higher dimensional data-set $\{x_1, x_2, x_3, \ldots\} \subset \mathbb{R}^d$,
   where coordinates of x_i represent attributes,
   distance $\|x_i - x_j\|$ measures the similarity among elements.
3. Cluster analysis
   dividing a large collection of individuals into groups
4. Statistical physics
   finite range interaction model

Gilbert '61, Penrose '03.

Why RGGs?

Good model for:

1. Wireless networks
   two radios can communicate only if within range of each other
2. Relational data-sets
   higher dimensional data-set $\{x_1, x_2, x_3, \ldots\} \subset \mathbb{R}^d$,
   where coordinates of x_i represent attributes,
   distance $\|x_i - x_j\|$ measures the similarity among elements.
3. Cluster analysis
   dividing a large collection of individuals into groups
4. Statistical physics
   finite range interaction model

Gilbert '61, Penrose '03.

Chromatic number

- Minimal number of colors $\chi(G)$ needed to color **all** nodes of a given graph G, so adjacent nodes receive different colors.
- **Applications**: assigning radio frequencies, job scheduling, etc.

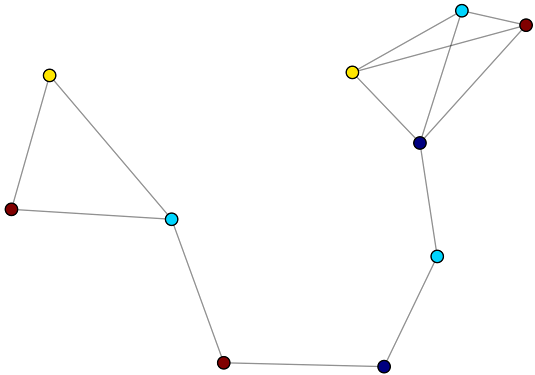


Figure : A small and sparse RGG, with $n = 10$, exp. deg. 2.6, and $\chi = 4$.

Chromatic number of RGGs

Result by McDiarmid, Müller '05, '09:

$$\chi(G_{n,r}) = \Theta\left(\frac{\log n}{\log\left(\frac{\log n}{nr^d}\right)}\right).$$

In the thermodynamic limit, when $nr^d = const$,

$$\chi(G_{n,r}) = \Theta\left(\frac{\log n}{\log\log n}\right) \longrightarrow \infty .$$

Additional inspiration to use only a constant k number of colors!

Chromatic number of RGGs

Result by McDiarmid, Müller '05, '09:

$$\chi(G_{n,r}) = \Theta\left(\frac{\log n}{\log\left(\frac{\log n}{nr^d}\right)}\right) .$$

In the thermodynamic limit, when $nr^d = const$,

$$\chi(G_{n,r}) = \Theta\left(\frac{\log n}{\log \log n}\right) \longrightarrow \infty .$$

Additional inspiration to use only a constant k number of colors!

Chromatic number of RGGs

Result by McDiarmid, Müller '05, '09:

$$\chi(G_{n,r}) = \Theta\left(\frac{\log n}{\log\left(\frac{\log n}{nr^d}\right)}\right).$$

In the thermodynamic limit, when $nr^d = const$,

$$\chi(G_{n,r}) = \Theta\left(\frac{\log n}{\log\log n}\right) \longrightarrow \infty.$$

Additional inspiration to use only a constant k number of colors!

Chromatic number of RGGs

Result by McDiarmid, Müller '05, '09:

$$\chi(G_{n,r}) = \Theta\left(\frac{\log n}{\log\left(\frac{\log n}{nr^d}\right)}\right).$$

In the thermodynamic limit, when $nr^d = const$,

$$\chi(G_{n,r}) = \Theta\left(\frac{\log n}{\log\log n}\right) \longrightarrow \infty .$$

Additional inspiration to use only a constant k number of colors!

Chromatic number of RGGs

Result by McDiarmid, Müller '05, '09:

$$\chi(G_{n,r}) = \Theta\left(\frac{\log n}{\log\left(\frac{\log n}{nr^d}\right)}\right) .$$

In the thermodynamic limit, when $nr^d = const$,

$$\chi(G_{n,r}) = \Theta\left(\frac{\log n}{\log\log n}\right) \longrightarrow \infty .$$

Additional inspiration to use only a constant k number of colors!

Why sparse regime?

- Our question much harder in the sparse regime.
  For a connected graph (dense regime) the answer tends to 0.
- Wireless networks:
  capacity $= f$ (number of users n, number of channels χ)
  increasing in n, decreasing in χ
- Many (real) networks (data sets) are 'very sparse';
  experiments on networks 5K-14M nodes and 6K-100M edgees
  Leskovec et all '09.

Why sparse regime?

- Our question much harder in the sparse regime.
  For a connected graph (dense regime) the answer tends to 0.

- Wireless networks:
  capacity $= f$ (number of users n, number of channels χ)
  increasing in n, decreasing in χ

- Many (real) networks (data sets) are 'very sparse';
  experiments on networks 5K-14M nodes and 6K-100M edgees
  Leskovec et all '09.

Main Question

What is the **maximum number** of nodes in a sparse **Random Geometric Graph** that can be properly colored with a **constant** number k of colors?

Maximum $M_{k,r}(V)$, given any $k, d \in \mathbb{N}$, set of nodes V, and $r > 0$.

Optimization problem: $M_{k,r}(V)$ is the maximum and integer. There are $|V|^{k+1}$ configurations. Interested when $|V| \to \infty$.

Main Question

What is the **maximum number** of nodes in a sparse **Random Geometric Graph** that can be properly colored with a **constant** number k of colors?

Maximum $M_{k,r}(V)$, given any $k, d \in \mathbb{N}$, set of nodes V, and $r > 0$.

Optimization problem: $M_{k,r}(V)$ is the maximum and integer. There are $|V|^{k+1}$ configurations. Interested when $|V| \to \infty$.

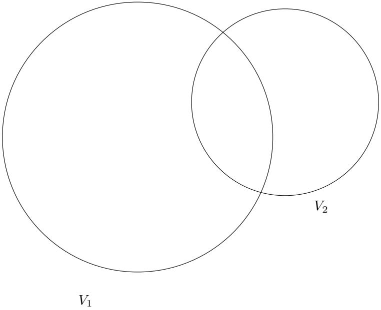
Main Question

What is the **maximum number** of nodes in a sparse **Random Geometric Graph** that can be properly colored with a **constant** number k of colors?

Maximum $M_{k,r}(V)$, given any $k, d \in \mathbb{N}$, set of nodes V, and $r > 0$.

Optimization problem: $M_{k,r}(V)$ is the maximum and integer. There are $|V|^{k+1}$ configurations. Interested when $|V| \to \infty$.

Sub-additivity:



$M_{k,r}(V_1 \cup V_2) \le M_{k,r}(V_1) + M_{k,r}(V_2)$, for any V_1, V_2.

There is **no scale-invariance** (i.e. homogeneity),

$$M_{k,r}(\alpha V) \neq \alpha M_{k,r}(V)\,.$$

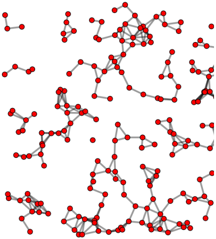


Figure : V

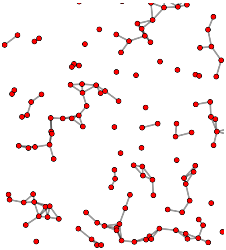


Figure : $V \to 1.25V$

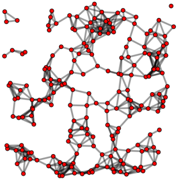


Figure : $V \to 0.75V$

Why are scale-invariance and sub-additivity important?

Consider any function L that maps a finite subset of points $x_1, x_2, x_3, \ldots$ from $\mathbb{R}^d$ to $\mathbb{R}_+$, and is *monotone*, *scale-invariant*, *translation-invariant*, and *sub-additive*.

Theorem (Steele, PTCO Theorem 3.1.1)

If x_i are independent random variables with the uniform distribution on $[0,1]^d$ then with probability one

$$\lim_{n\to\infty} n^{-(1-1/d)} L(x_1, x_2, \ldots, x_n) = \beta_L(d),$$

where $\beta_L(d)$ is a positive constant, which depends both on the dimension d and the functional L.

Why are scale-invariance and sub-additivity important?

Consider any function L that maps a finite subset of points $x_1, x_2, x_3, \ldots$ from $\mathbb{R}^d$ to $\mathbb{R}_+$, and is *monotone*, *scale-invariant*, *translation-invariant*, and *sub-additive*.

Theorem (Steele, PTCO Theorem 3.1.1)

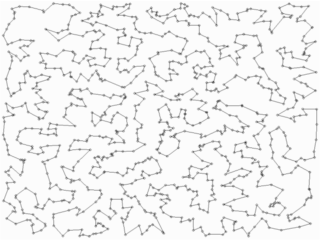
If x_i are independent random variables with the uniform distribution on $[0,1]^d$ then with probability one

$$\lim_{n\to\infty} n^{-(1-1/d)} L\left(x_1, x_2, \ldots, x_n\right) = \beta_L(d)\,,$$

where $\beta_L(d)$ is a positive constant, which depends both on the dimension d and the functional L.

Representative example

Euclidean traveling salesman problem (TSP)



$$\lim_{n\to\infty} n^{-(1-1/d)} L_{TSP}(x_1, x_2, \ldots, x_n) = \beta(d) \int_{x\in\mathbb{R}^d} f(x)dx$$

Beardwood, Halton, Hammersley '59.

Our object $M_{k,r}(V)$ is neither scale-invariant nor smooth!

Existing frameworks

to obtain weak and strong law of large numbers, central limit theorem, etc. on some 'well behaved' Euclidean functionals

- Steele "Probability Theory and Combinatorial Optimization"
- Rhee-Talagrand isoperimetric inequalities for smooth functionals (e.g. TSP, MST)
- boundary zero cost methods: Frieze, Yukich, "Probabilistic analysis of the TSP", '02
- (i) Laws of large numbers for smooth, superadditive Euclidean functionals (Thm 8.1)
  (ii) General 'umbrella theorem' for smooth, subadditive Euclidean functionals (Thm 8.3)
  Yukich "Limit theorems in discrete stochastic geometry", '09
- stabilization methods: Penrose, Yukich, Baryshnikov
- . . .

Our Probabilistic Objectives:

As $t \to \infty$ and $n \to \infty$, examine mean, variance, laws of large numbers, and limiting distribution of the following objects:

$$\begin{aligned} \text{Poisson case:} \quad & F_{k,\lambda}(t) & := & \; M_{k,1}([0,t]^d \cap \mathcal{X}_\lambda)\,, \\ \text{Uniform case:} \quad & H_{k,\nu}(n) & := & \; M_{k,\sqrt[d]{\nu/n}}(\mathcal{I}_n)\,. \end{aligned}$$

Hurdle:

$F_{k,\lambda}(t)$ and $H_{k,\nu}(n)$ are maximum colorings: global, non-stabilizing functionals. Dependency propagates through local interactions.

Our Probabilistic Objectives:

As $t \to \infty$ and $n \to \infty$, examine mean, variance, laws of large numbers, and limiting distribution of the following objects:

$$\begin{array}{rrcl} \text{Poisson case:} & F_{k,\lambda}(t) & := & M_{k,1}([0,t]^d \cap \mathcal{X}_\lambda)\,, \\ \text{Uniform case:} & H_{k,\nu}(n) & := & M_{k,\sqrt[d]{\nu/n}}(\mathcal{I}_n)\,. \end{array}$$

Hurdle:

$F_{k,\lambda}(t)$ and $H_{k,\nu}(n)$ are maximum colorings: global, non-stabilizing functionals. Dependency propagates through local interactions.

Proposition

For any $k \geq 1$, $\lambda > 0$, the limit of $\mathbb{E}\left\{F_{k,\lambda}(t)/\lambda t^d\right\}$ *exists and equals*

$$a_{k,\lambda} := \lim_{t\to\infty} \mathbb{E}\left\{F_{k,\lambda}(t)/\lambda t^d\right\} = \inf_{t>0} \mathbb{E}\left\{F_{k,\lambda}(t)/\lambda t^d\right\}.$$

Proposition

For any $k \geq 1$, $\nu > 0$, the limit of $\mathbb{E}\left\{H_{k,\sqrt[d]{\nu/n}}(n)/n\right\}$ *exists*

$$\lim_{n\to\infty} \mathbb{E}\left\{H_{k,\sqrt[d]{\nu/n}}(n)/n\right\} = a_{k,\nu}.$$

Note. The expectations of the coloring ratios tend to the same limiting object (depends on k and density).

Proposition

For any $k \geq 1$, $\lambda > 0$, *the limit of* $\mathbb{E}\left\{F_{k,\lambda}(t)/\lambda t^d\right\}$ *exists and equals*

$$\begin{aligned} a_{k,\lambda} &:= \lim_{t\to\infty} \mathbb{E}\left\{F_{k,\lambda}(t)/\lambda t^d\right\} \\ &= \inf_{t>0} \mathbb{E}\left\{F_{k,\lambda}(t)/\lambda t^d\right\}. \end{aligned}$$

Proposition

For any $k \geq 1$, $\nu > 0$, *the limit of* $\mathbb{E}\left\{H_{k,\sqrt[d]{\nu/n}}(n)/n\right\}$ *exists*

$$\lim_{n\to\infty} \mathbb{E}\left\{H_{k,\sqrt[d]{\nu/n}}(n)/n\right\} = a_{k,\nu}.$$

Note. The expectations of the coloring ratios tend to the same limiting object (depends on k and density).

Convergence in probability

Proposition

For any $k \geq 1$, $\lambda > 0$, the random variable $F_{k,\lambda}(t)/\lambda t^d$ converges to $a_{k,\lambda}$ in probability as $t \to \infty$:

$$(\forall \delta > 0) \qquad \mathbb{P}\left\{ \left| \frac{F_{k,\lambda}(t)}{\lambda t^d} - a_{k,\lambda} \right| \geq \delta \right\} \xrightarrow{t\to\infty} 0\,.$$

Proposition

For any $k \geq 1$, $\nu > 0$, the random variable $H_{k,\sqrt[d]{\nu/n}}(n)/n$ converges to $a_{k,\nu}$ in probability as $n \to \infty$.

Convergence in probability

Proposition

For any $k \geq 1$, $\lambda > 0$, the random variable $F_{k,\lambda}(t)/\lambda t^d$ converges to $a_{k,\lambda}$ in probability as $t \to \infty$:

$$(\forall \delta > 0) \qquad \mathbb{P}\left\{\left|\frac{F_{k,\lambda}(t)}{\lambda t^d} - a_{k,\lambda}\right| \geq \delta\right\} \xrightarrow{t\to\infty} 0\,.$$

Proposition

For any $k \geq 1$, $\nu > 0$, the random variable $H_{k,\sqrt[d]{\nu/n}}(n)/n$ converges to $a_{k,\nu}$ in probability as $n \to \infty$.

to conclude this part

In the limit as $t \to \infty$ and $n \to \infty$:

convergence of expectations:

$$\frac{\mathbb{E}\left\{F_{k,\lambda}(t)\right\}}{\lambda t^d} \to a_{k,\lambda} \qquad \text{and} \qquad \frac{\mathbb{E}\left\{H_{k,\sqrt[d]{\nu/n}}(n)\right\}}{n} \to a_{k,\nu}$$

On average, the constant fraction of nodes $a_{k,\lambda}$ (i.e. $a_{k,\nu}$) can be properly assigned one of k colors.

convergence in probability:

$$\frac{F_{k,\lambda}(t)}{\lambda t^d} \xrightarrow{\mathrm{p}} a_{k,\lambda} \qquad \text{and} \qquad \frac{H_{k,\sqrt[d]{\nu/n}}(n)}{n} \xrightarrow{\mathrm{p}} a_{k,\nu}$$

The maximum number of colored nodes is concentrated for both models.

to conclude this part

In the limit as $t \to \infty$ and $n \to \infty$:

convergence of expectations:

$$\frac{\mathbb{E}\left\{F_{k,\lambda}(t)\right\}}{\lambda t^d} \to a_{k,\lambda} \qquad \text{and} \qquad \frac{\mathbb{E}\left\{H_{k,\sqrt[d]{\nu/n}}(n)\right\}}{n} \to a_{k,\nu}$$

On average, the constant fraction of nodes $a_{k,\lambda}$ (i.e. $a_{k,\nu}$) can be properly assigned one of k colors.

convergence in probability:

$$\frac{F_{k,\lambda}(t)}{\lambda t^d} \xrightarrow{\mathrm{p}} a_{k,\lambda} \qquad \text{and} \qquad \frac{H_{k,\sqrt[d]{\nu/n}}(n)}{n} \xrightarrow{\mathrm{p}} a_{k,\nu}$$

The maximum number of colored nodes is concentrated for both models.

to conclude this part

In the limit as $t \to \infty$ and $n \to \infty$:

convergence of expectations:

$$\frac{\mathbb{E}\left\{F_{k,\lambda}(t)\right\}}{\lambda t^d} \to a_{k,\lambda} \qquad \text{and} \qquad \frac{\mathbb{E}\left\{H_{k,\sqrt[d]{\nu/n}}(n)\right\}}{n} \to a_{k,\nu}$$

On average, the constant fraction of nodes $a_{k,\lambda}$ (i.e. $a_{k,\nu}$) can be properly assigned one of k colors.

convergence in probability:

$$\frac{F_{k,\lambda}(t)}{\lambda t^d} \xrightarrow{\text{p}} a_{k,\lambda} \qquad \text{and} \qquad \frac{H_{k,\sqrt[d]{\nu/n}}(n)}{n} \xrightarrow{\text{p}} a_{k,\nu}$$

The maximum number of colored nodes is concentrated for both models.

- We know $F_{k,\lambda}(t)$ is concentrated 'around' $a_{k,\lambda}\lambda t^d$ (for any $t > 0$).
- Can we further describe $F_{k,\lambda}(t)$?
- What can we say about the variance $\sigma^2(t) := \mathbb{V}ar\left\{F_{k,\lambda}(t)\right\}$?

Variance

Lemma

For all λ, d, k,

$$\inf_{t>3} \frac{\sigma^2(t)}{t^d} > 0.$$

Lemma

For any λ, d, k, t,

$$\sigma^2(t) \leq \lambda t^d.$$

Volume is the right order

Proposition

For all $\lambda > 0$ *and* $d, k \in \mathbb{N}$, *asymptotically as* t *tends to* ∞, *we have*

$$\sigma^2(t) = \Theta(t^d).$$

Variance

Lemma

For all λ, d, k,

$$\inf_{t>3} \frac{\sigma^2(t)}{t^d} > 0.$$

Lemma

For any λ, d, k, t,

$$\sigma^2(t) \leq \lambda t^d.$$

Volume is the right order

Proposition

For all $\lambda > 0$ *and* $d, k \in \mathbb{N}$*, asymptotically as* t *tends to* ∞*, we have*

$$\sigma^2(t) = \Theta(t^d).$$

Constant $a_{k,\lambda}$

- In the limit, the constant $a_{k,\lambda}$ fraction can be properly colored.
- For dim=1

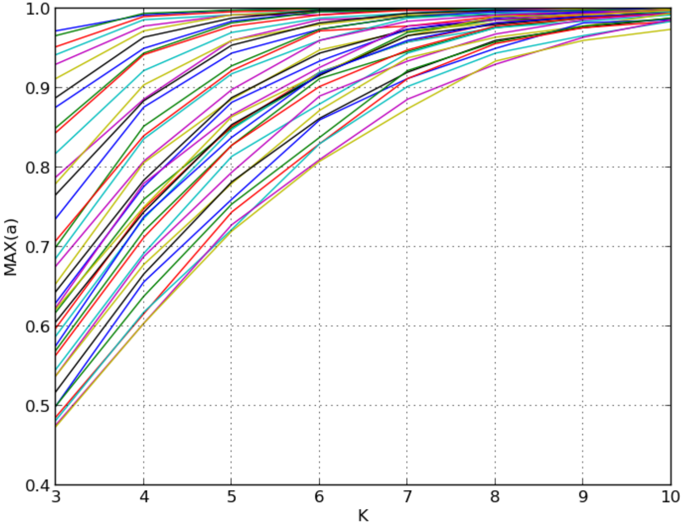
$$a_{k,\lambda} = 1 - \underbrace{\frac{\mathbb{P}\left\{Poisson(\lambda) = k\right\}}{\mathbb{P}\left\{Poisson(\lambda) \leq k\right\}}}$$

  the Erlang loss probability
- For dim=1, if both k and λ grow large, but finite,

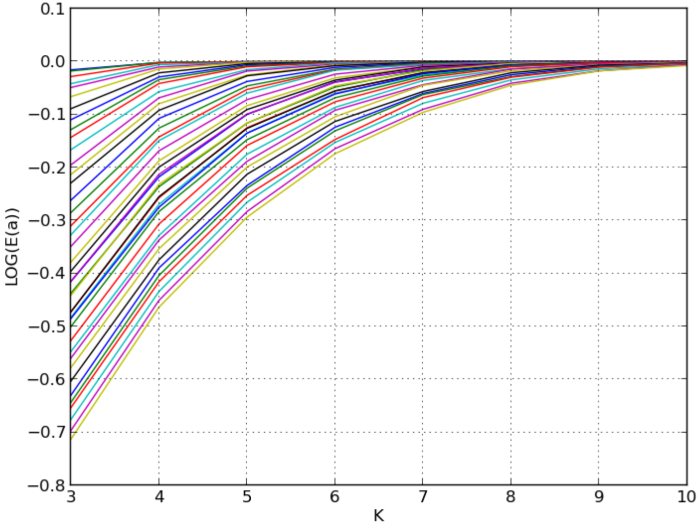
$$a_{k,\lambda} \quad \approx \quad \max\{k/\lambda, 1\}\,.$$

- For dim ≥ 2, we have non-tight bounds.
- Not knowing $a_{k,\lambda}$ is similar to the β_d-paradigm in TSP.

Simulations



Simulations



Simulations

- No guaranties that the global maximum can be found (algorithmically)
- Lower bounds on the coloring ratio for **finite** volume (number of nodes)
- Relate these results to the real asymptotic values of $a_{k,\lambda}$?

Questions

- Extend work to other functionals
- Values of the constant $a_{k,\lambda}$
- Models with long range interaction

Happy Birthday prof. Cvetković!

Thank You

ect.bell-labs.com/who/milan
milan@research.bell-labs.com