Near Minimal Spanning Tree and Scaling Exponent

Marc Lelarge
ENS-INRIA

joint work with **David Aldous** and **Charles Bordenave**

Journées ALEA
23 Mars 2007

"Regular" Optimization Problem

– Cost function $F : \mathbb{R}^d \longrightarrow \mathbb{R}$ reaching its minimum at x^*.

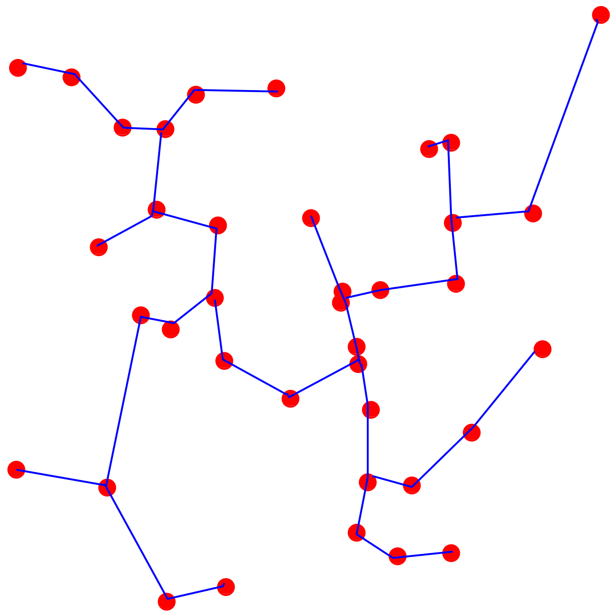
– Relation between the distance $\delta = |x - x^*|$ and the difference $\epsilon = F(x) - F(x^*)$:
if F is "smooth",

$$\epsilon(\delta) = \inf\{F(x) - F(x^*),\ |x - x^*| \geq \delta\} \sim C\delta^2.$$

In this case, we say that the scaling exponent corresponding to this problem is 2.

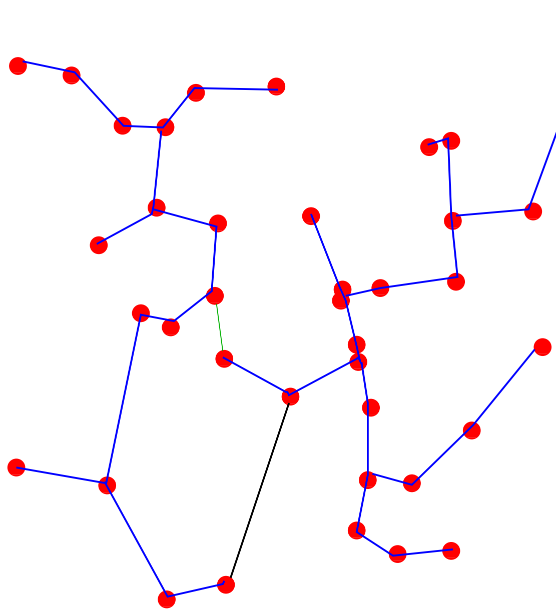
Combinatorial Optimization: the Minimal Spanning Tree

Cost function is the length of the spanning tree, $len(T_n)$
minimum reached by the Minimal Spanning Tree: MST_n



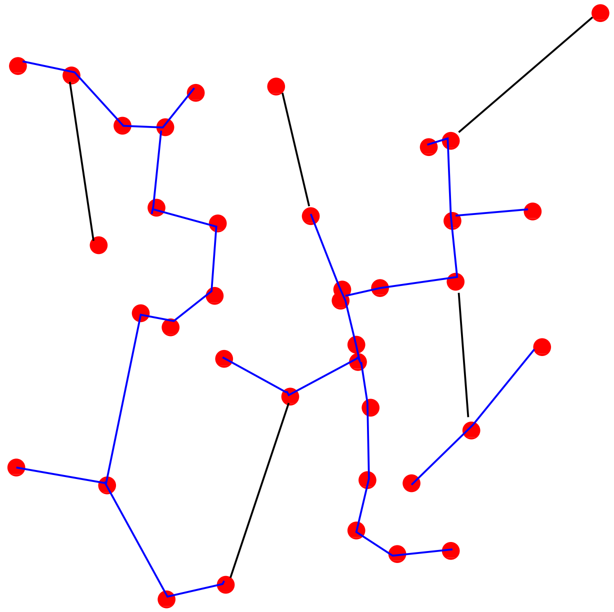
Perturbation of the minimal spanning tree (1)

Perturbation of the optimal solution: adding an edge creates a cycle
$\longrightarrow$ removing an edge of the cycle gives a spanning tree.



Perturbation of the minimal spanning tree (2)

Perturbation of δn edges:
What is the order of magnitude of the variation of the length of the tree?



Motivation

– n random points on $\mathbb{R}^2$ such that the nearest-neighbor distances are order one.

– We define the random variable:

$$\epsilon_n(\delta) = \min \left\{ \frac{len(T_n) - len(MST_n)}{n},\ |T_n \backslash MST_n| \geq \delta n \right\},$$

where the minimum is taken over spanning trees T_n.

– What is the value of α?

$$\lim_n \mathbb{E}[\epsilon_n(\delta)] \equiv C\delta^{\alpha}$$

Plan

1. Probability models for n random points
2. Upper Bound
3. Some properties of the Minimal Spanning Tree
4. Lower Bound
5. Result

Model 1: the disordered lattice

Start with the discrete 2-dimensional cube $\mathbb{C}_m^2 = [1, 2, \ldots, m]^2$, so there are $n = m^2$ vertices and there are 4 edges at each non-boundary vertex. Then take the edge-lengths to be i.i.d. random variables ξ_e, whose common distribution ξ has finite mean and some bounded continuous density function $f_\xi(\cdot)$.

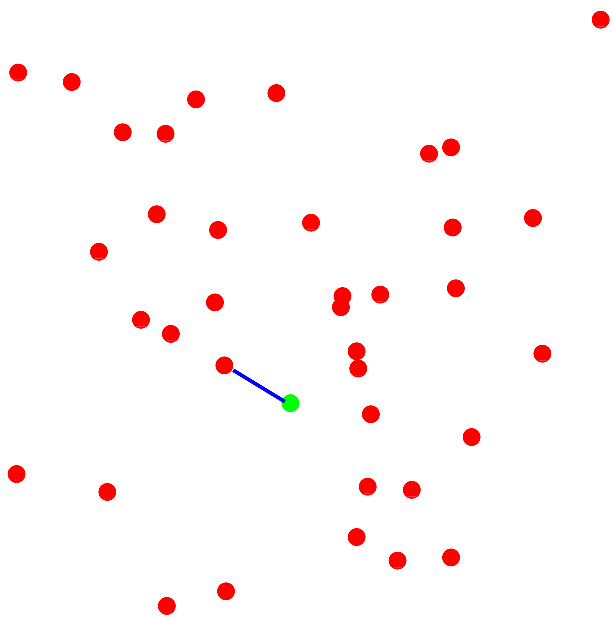
Model 2: Uniform Euclidean model

Take the continuum 2-dimensional square $[0, \sqrt{n}]^2$ of volume n. Put down n independent uniformly distributed random points in this square. Take the complete graph on these n vertices, with Euclidean distance as edge-lengths.

Ambiguity-free labeled graph $G = (V, E, len)$

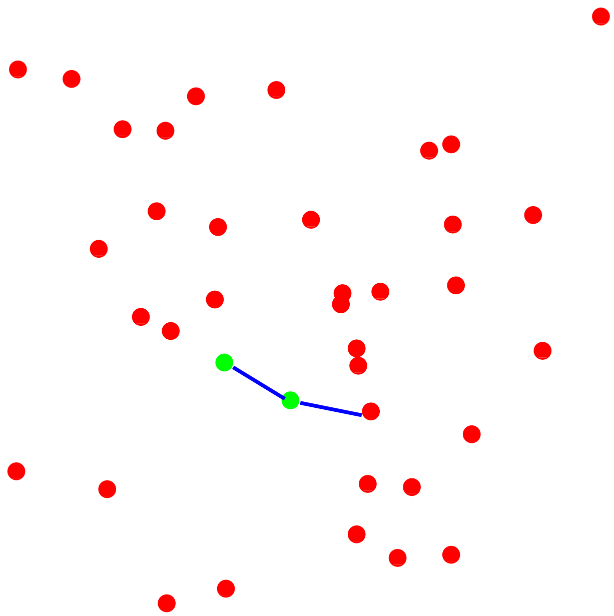
Prim's algorithm

The minimal spanning tree (in finite graphs) can be constructed by an inductive invasion procedure.



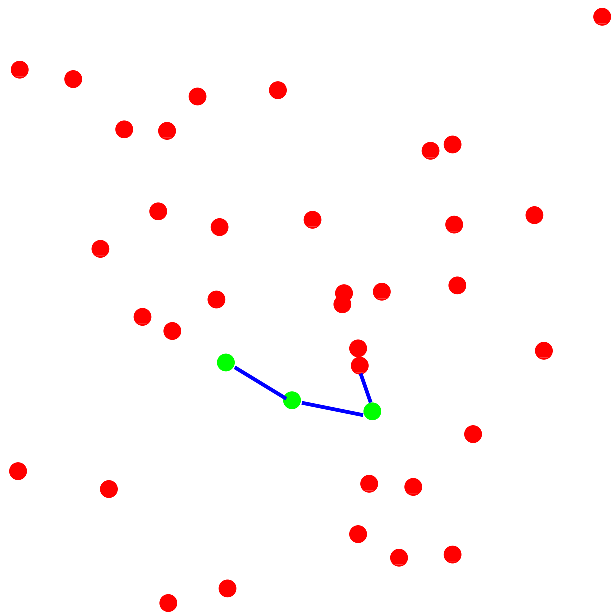
Prim's algorithm

Start with a vertex $I_0(v) = \{v\}$ and given $I_k(v)$, let $I_{k+1}(v)$ be $I_k(v)$ together with the shortest edge in the frontier of $I_k(v)$.



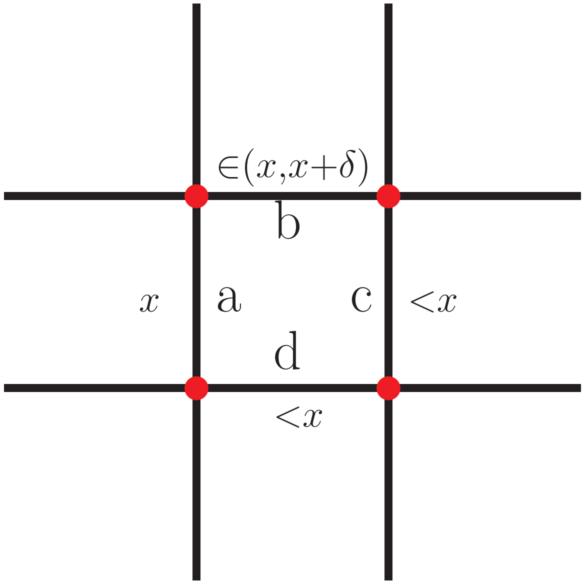
Prim's algorithm

After $n-1$ steps, we obtain the minimal spanning tree. In particular the set of invaded edges does not depend on the starting site (in finite graphs).



Upper Bound (for the lattice -1)

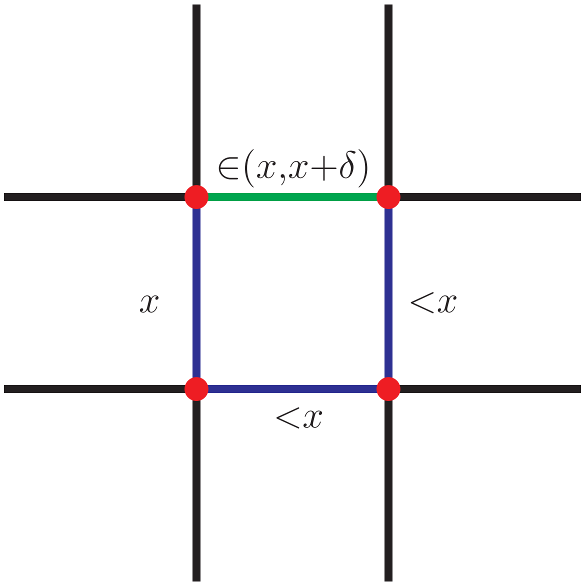
The upper bound rests upon a simple construction of near-minimal spanning trees:



The probability of seeing this configuration is $q(\delta) \sim c\delta$ with $0 < c < \infty$.

Upper Bound (for the lattice -2)

To obtain a perturbation of the MST: remove the edge of length x from the MST and add the green edge.



The increase in edge-length of spanning tree caused by the modification equals $r(\delta) \sim \frac{1}{2}\delta q(\delta) \sim C\delta^2$.

Upper Bound

Recall:

$$\epsilon_n(\delta) = \min\left\{\frac{len(T_n) - len(MST_n)}{n},\ |T_n \backslash MST_n| \geq \delta n\right\}.$$

In either model, there exists $C < \infty$ such that for δ sufficiently small,

$$\limsup_{n\to\infty} \mathbb{E}[\epsilon_n(\delta)] \leq C\delta^2.$$

The creek-crossing criterion (Alexander (95))

Consider an edge $e = (u, v)$ of a graph $G = (V, E, len)$ (finite or infinite). The following are equivalent (for our probabilistic models):

– 1) there exists no path from u to v with all edges shorter than e.

– 2) e is invaded either from u or from v.

Let G_t be the subgraph consisting of those edges e of G with $len(e) < t$ and define

$$perc(u, v) = \inf\{t,\ u \text{ and } v \text{ in same component of } G_t\},\ \text{then:}$$
$$e \text{ satisfies 1)} \Leftrightarrow len(e) = perc(e).$$

If G is finite, these assumptions are equivalent to

– e is an edge of the MST of G (Prim's algorithm).

Heuristic

An edge is in the MST iff

$$len(e) = perc(e)$$

where $perc(e)$ depends on all the other edge-lengths.
So an edge e NOT in the MST has an excess length: $len(e) - perc(e) > 0$.
Define the measure μ_n on $(0, \infty)$ in terms of the expectation (n random points)

$$\mu_n(0, x) = \frac{1}{n}\mathbb{E}\left[|e : \ 0 < len(e) - perc(e) < x|\right].$$

For any reasonable model with suitable scaling of edge-lengths we expect an $n \to \infty$ limit measure $\mu(\cdot)$, with a density $\nu(x) = d\mu/dx$ having a finite limit $\nu(0^+)$ as $x \downarrow 0$.
The condition $\nu(0+) < \infty$ is essentially sufficient for the scaling exponent to equal 2.

Local Weak Convergence (Aldous-Steele)

In each probabilistic model, we can let $n \to \infty$ and get an infinite point model:

– for the disordered lattice, the limit is just disordered $\mathbb{Z}^2$.

– for uniform Euclidean, the limit is Poisson-Euclidean.

MST convergence Theorem
Consider n-vertex connected graphs $G_n[X_n]$ that converges in distribution to some G_∞. Then jointly with this convergence, we have

$$MST(G_n) \to^d MST(G_\infty).$$

The *perc* functional is also 'continuous':

$$perc(e, G_n) \to^d perc(e, G_\infty).$$

Upper bound (lattice case)

Define for $e = (u, v)$,

$$\begin{aligned} perc^{-}(u,v) &= \inf\{t,\ u \text{ and } v \text{ in same component of } G_t \backslash \{e\}\}, \text{ so that,} \\ perc(e) &= \min(len(e), perc^{-}(e)). \end{aligned}$$

$$\begin{aligned} \mu(0,x) &= \mathbb{E}\sum_i \mathbb{1}(0 < len(O,\eta_i) - perc(O,\eta_i) \le x) \\ &= 4\int \mathbb{E}\mathbb{1}(s < len(e) \le s + x) f_{perc^{-}}(s)ds \\ &\le Cx, \end{aligned}$$

thanks to the independence of $len(e)$ and $perc^{-}(e)$.

Upper bound (Euclidean case)

We consider a Poisson point process $\Phi = \sum_i \delta_{\eta_i}$ of rate 1 on $\mathbb{R}^2$. Define the measure μ on $(0, +\infty)$ by

$$\begin{aligned} \mu(0,x) &= \mathbb{E}\sum_i \mathbb{1}(0 < len(O,\eta_i) - perc(O,\eta_i) \leq x) \\ &= 2\pi \int_0^\infty \mathbb{P}^{O,\underline{t}}(perc(O,\underline{t}) \in [t-x,t))tdt, \end{aligned}$$

where $\underline{t} = (0,t)$.
We need to prove that

$$\limsup_{x\to 0} \frac{\mu(0,x)}{x} < \infty.$$

Far from r_c

Let r_c be the critical radius for the Poisson continuum percolation model of density 1 and deterministic radius.
For any $\epsilon > 0$, we have

$$\begin{aligned} \text{for } 0 < t < r_c - \epsilon, \quad & \mathbb{P}^{O,\underline{t}}(perc(O,\underline{t}) \in [t-x,t)) \leq C_1 x, \\ \text{for } t > r_c + \epsilon, \quad & \mathbb{P}^{O,\underline{t}}(perc(O,\underline{t}) \in [t-x,t)) \leq C_1 x e^{-C_2 t}. \end{aligned}$$

On $(r_c - \epsilon, r_c + \epsilon)$

Denote by G^O the open cluster containing the origin. We have

$$\begin{aligned} \mathbb{P}^{O,\underline{t}}(perc(O,\underline{t}) \in [t-x,t)) &= \mathbb{P}^{O,\underline{t}}_{1,t}(A) - \mathbb{P}^{O,\underline{t}}_{1,t-x}(A), \text{ with } A = \{\underline{t} \in G^O\}, \\ &= \mathbb{P}^{O,\underline{t}}_{1,t}(A) - \mathbb{P}^{O,\underline{t-x}}_{1,t-x}(A) + \\ &\quad \mathbb{P}^{O,\underline{t-x}}_{1,t-x}(A) - \mathbb{P}^{O,\underline{t}}_{1,t-x}(A). \end{aligned}$$

The second term is $O(x)$ and for the first term

$$\frac{1}{x}\int_{r_c-\epsilon}^{r_c+\epsilon} (\mathbb{P}^{O,\underline{t}}_{1,t}(A) - \mathbb{P}^{O,\underline{t-x}}_{1,t-x}(A)) t dt \leq C,$$

by a simple change of variable.

Near Minimal Spanning Trees

We define the random variable:

$$\epsilon_n(\delta) = \min \left\{ \frac{len(T_n) - len(MST_n)}{n}, \ |T_n \backslash MST_n| \geq \delta n \right\},$$

where the minimum is taken on spanning trees T_n.

Theorem 1. *There exists positive constants* C_1, C_2, δ_0, n_0 *such that for* $0 < \delta < \delta_0$, $n \geq n_0$,

$$C_1 \delta^2 < \mathbb{E}\epsilon_n(\delta) < C_2 \delta^2.$$

– Scaling exponent is 2 for the minimal spanning tree!

– minimum matching and TSP are expected to have scaling exponent 3

– For a combinatorial optimization problem, a larger exponent means that there are more near-optimal solutions, suggesting that the algorithmic problem of finding the optimal solution is intrinsically harder.

Merci!

References

[1] D.J. Aldous and J.M. Steele. The objective method. In volume 110 of *Encyclopaedia of Mathematical Sciences*, pages 1–72. Springer-Verlag, 2003.

[2] K. S. Alexander. Percolation and minimal spanning forests in infinite graphs. *Ann. Probab.*, 23:87–104, 1995.

[3] R. Meester and R. Roy. *Continuum percolation* Cambridge University Press, Cambridge, 1996.