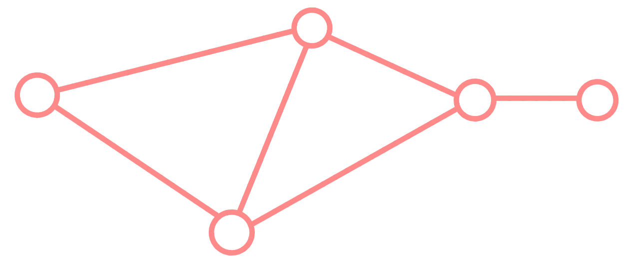
Probabilistic aspects of minimum spanning trees

Louigi Addario-Berry

November 3, 2016

Duke University

I. The MST problem



$G=(V,E)$ finite connected graph.

A spanning forest is a subgraph $F=(V,E')$, $E'\subseteq E$, with no cycles. It is a spanning tree if it is also connected.

For fixed positive weights $(U_e)_{e\in E}$, the weight of a spanning tree T of G is $\omega(T)=\sum_{e\in E(T)} U_e$

Say T is the MST of G if

$$\omega(T)=\min\{\omega(T') : T' \text{ a spanning tree of } G\}$$

If weights U_e are all distinct then MST is unique.

II. MST Algorithms.

Greedy: Kruskal's algorithm

Order $E = \{e_1, e_2, \dots, e_m\}$ s.t. $U_{e_1} < U_{e_2} < \cdots < U_{e_m}$

Set $F_0 = (V, \phi)$; empty forest.

Aim: Maintain that $F^i = (V, E^i)$ is the minimum spanning forest of $G^i = (V, \{e_1, \dots, e_i\})$

componentwise MST

Let $E^{i+1} \begin{cases} E^i \cup \{e_{i+1}\} & \text{if this creates no cycle} \\ E^i & \text{otherwise.} \end{cases}$

Output: $F^m = (V, E^m)$ which is the minimum spanning tree.

[Aside: Kruskal's algorithm finds a min-weight basis for any matroid, and this characterizes matroids.]

II. MST Algorithms.

Local: Prim's algorithm

Fix a starting vertex (a root) v.

Set $T_1 = (\{v\}, \phi)$; a tree with 1 vertex.

Aim: Maintain that $T_i = (V_i, E_i)$ is the minimum spanning tree of $G|_{V_i} = (V_i, \{uv \in E : u \in V_i, v \in V_i\})$

For $1 \leq i < |V|$, let $e_{i+1} = uv$ be the smallest weight edge from V_i to $V \setminus V_i$.

Let $V_{i+1} = V_i \cup \{u, v\}$, $E_{i+1} = E_i \cup \{e_{i+1}\}$.

Then $T_{|V|}$ is the MST.

Interesting hard question: when do local algorithms yield (near)-optimal outputs?

III. Randomization.

Average-case analysis: understand behaviour of an algorithm when averaged over all possible inputs.

[Smoothed analysis: understand behaviour on random perturbation of input data.]

Algorithmic complexity

$\neq$

Average-case complexity

$\neq$

Probabilistic difficulty

MST is algorithmically easy but probabilistically challenging.

III. Randomization.

Mean-field

Graph = K_n = complete graph, n vertices

Edge weights: independent, identically distributed, Uniform $[0,1]$

[Kruskal's algorithm only uses edge ordering; so any IID continuous weights give same law for MST viewed as an unweighted graph.]

Theorem (Frieze, 1985):

Let $T_n = \mathrm{MST}(K_n)$.

Then $w(T_n) \xrightarrow{P} \zeta(3)$.

Euclidean

$\mathbb{Z}^d$, nearest-neighbour edges.

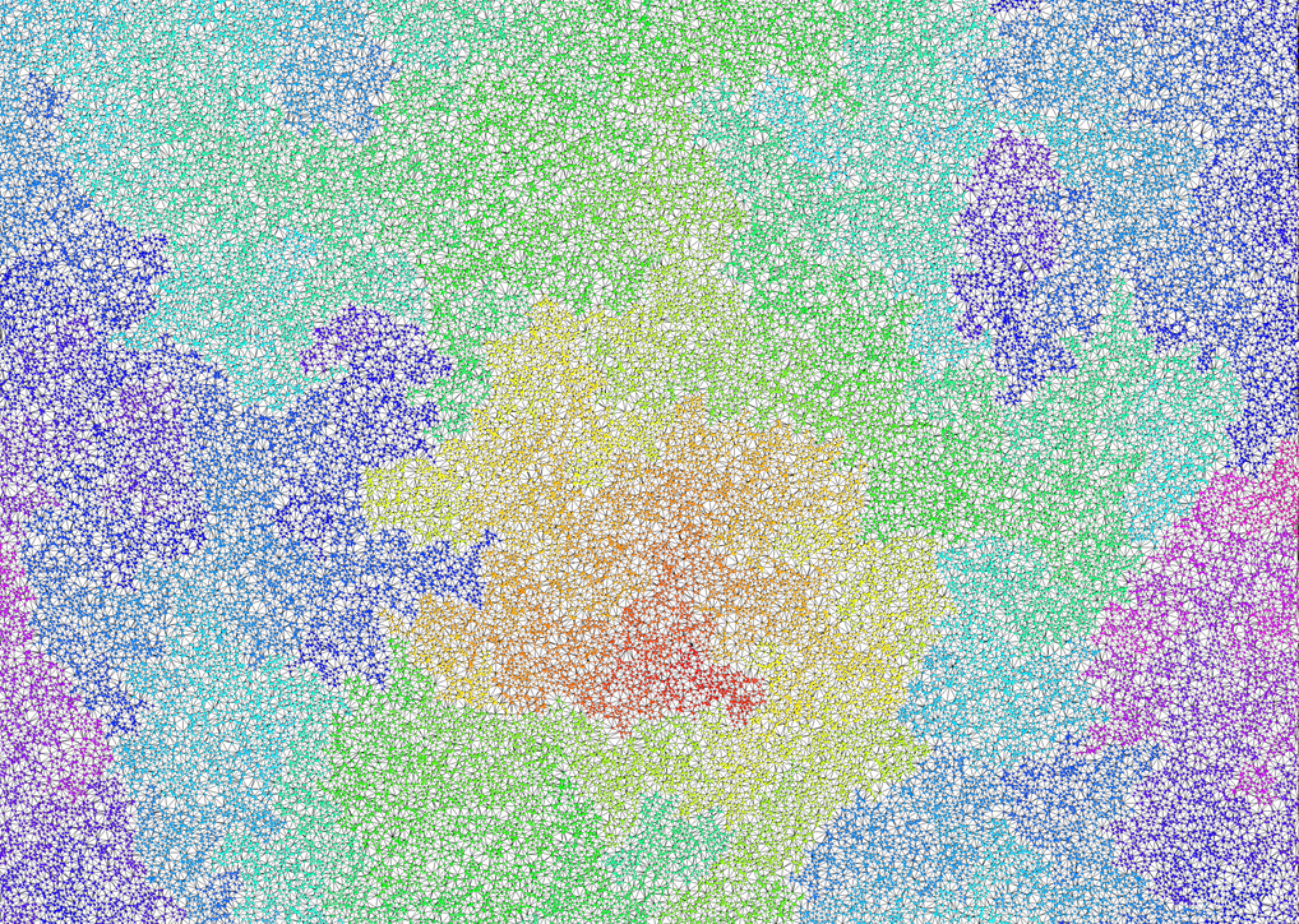
Edge weights: independent, identically distributed, Uniform $[0,1]$

[On ∞ graphs some care is needed: Kruskal not obviously well-defined; Prim well-defined but T_∞ may not reach all vertices]

"Invasion Percolation"

Theorem (Newman, 1995):

T_∞ has 0 density $\Longleftrightarrow$ No percolation at criticality.

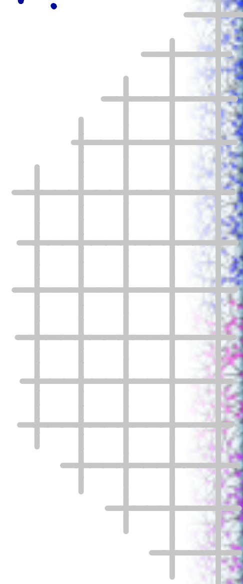


IV. Percolation

Fix connected graph $G=(V,E)$, IID Uniform$[0,1]$ edge weights $(U_e, e\in E)$.

Set $E_p = \{e\in E : U_e \le p\}$, $G_p = (V, E_p)$.

Assume G transitive: $\forall\, u,v\in V\ \exists$ an automorphism φ of G with $\varphi(u)=v$.



Percolation threshold (for G infinite):

$$p_c(G) = \inf\{p\in[0,1] : \mathbb{P}(G_p \text{ has an infinite connected component}) > 0\}$$

$p_c(\mathbb{Z}) = 1$; $p_c(\mathbb{Z}^2) = \frac{1}{2}$; $p_c(\mathbb{Z}^d) = \frac{1}{2d} + O\left(\frac{1}{d^2}\right)$, d large.

For $v\in V$ write $v \xleftrightarrow{G_p} \infty$ if v in an infinite component of G_p.

Let $\Theta_p(G) = \mathbb{P}(o \xleftrightarrow{G_p} \infty)$. Doesn't depend on root/origin $o\in V$.

Definition: G percolates at criticality if $\Theta_{p_c}(G) > 0$.

Conjecture: No percolation at criticality in $\mathbb{Z}^d$, all $d\ge 2$. Known for $d=2, d>10$.

V. Percolation on finite graphs.

The percolation process $(G_p, 0 \le p \le 1)$ still makes sense but all components are finite for every p.

Many possible definitions of percolation threshold.

One possibility which "works" for many geometries:

Order components of G_p in decreasing order of size as $(C_p(i), i \ge 1)$.

Let $p_c = \inf\{p : \mathbb{E}[|C_p(2)|/|C_p(1)|] \le 1/2\}$.

(Intuition: Once giant component appears, all other components much smaller.)

<u>Mean-field</u>: $G = K_n$. Then percolation on G gives the Erdős-Rényi random graph process, typically denoted $(G_{n,p}, 0 \le p \le 1)$.

VI. Percolation and MSTs.

Assume $G=(V,E)$ finite.

IID Uniform$[0,1]$ edge weights $(U_e, e\in E)$ now and forever after.

Recall With $E=\{e_1,e_2,\dots,e_m\}$ ordered s.t. $U_{e_1}<U_{e_2}<\cdots<U_{e_m}$,

Kruskal's algorithm maintains that $F^i=(V,E^i)$ is the minimum spanning forest of $G^i=(V,\{e_1,\dots,e_i\})$; $F^m=(V,E^m)$ is MST.

With $p_i=U_{e_i}$, then $G^i=G_{p_i}$ for all i

Set $E_p^m=\{e\in E_m : U_e\le p\}$, $F_p=(V,E_p^m)$.

Then the vertex sets of connected components of F_p and G_p are identical.

VII Dimension and volume growth

Distances determined by major routes, so known by time p_c

Heuristic [Percolation threshold p_c
=
point at which long-range correlations / global structure appear.

At times $p \neq p_c$, all finite/nongiant components are small:

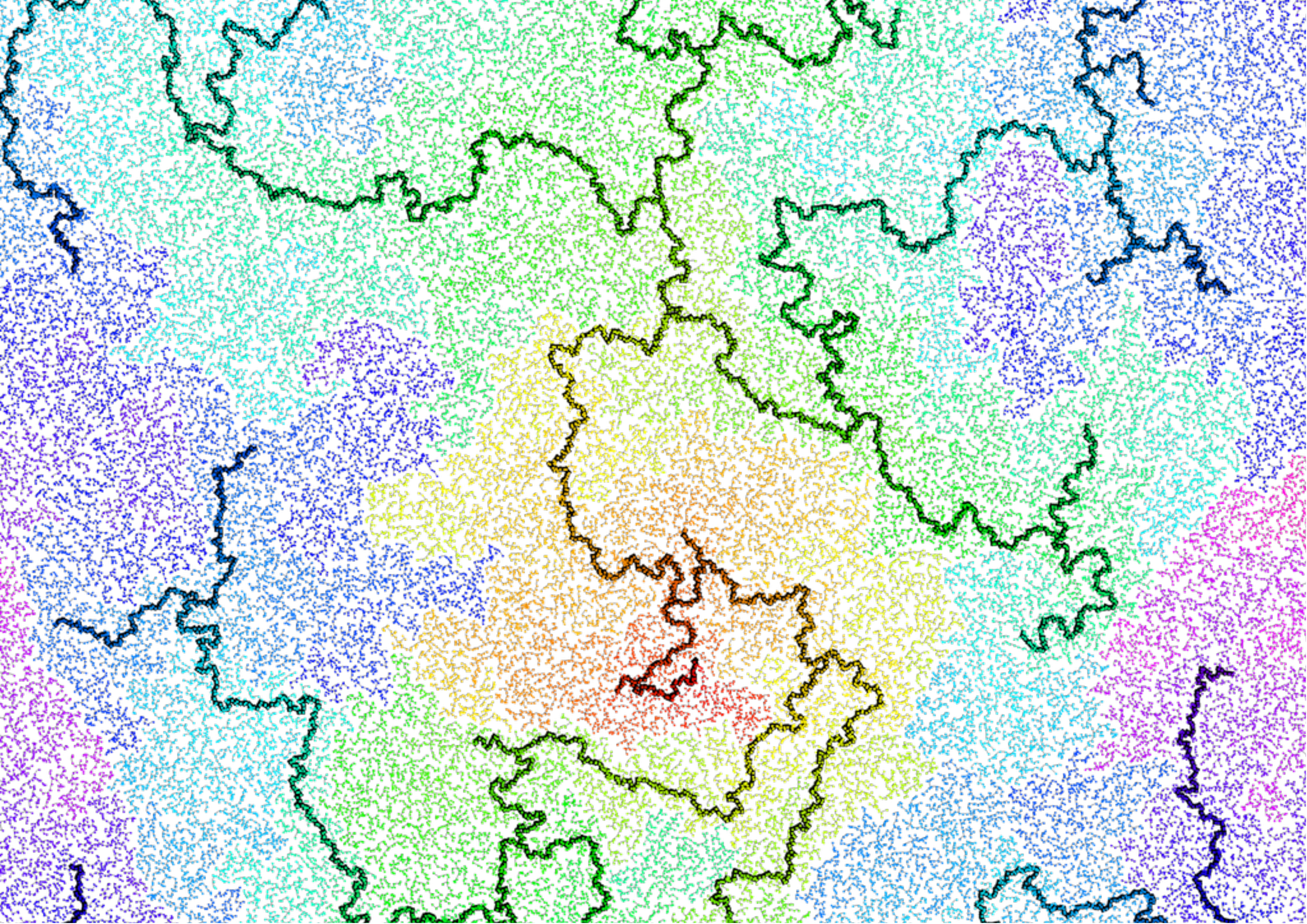
$$\mathbb{E}[\text{Size of component } C \text{ of } G_p \text{ containing } v \mid C \neq C_p(1)] = O(1)$$

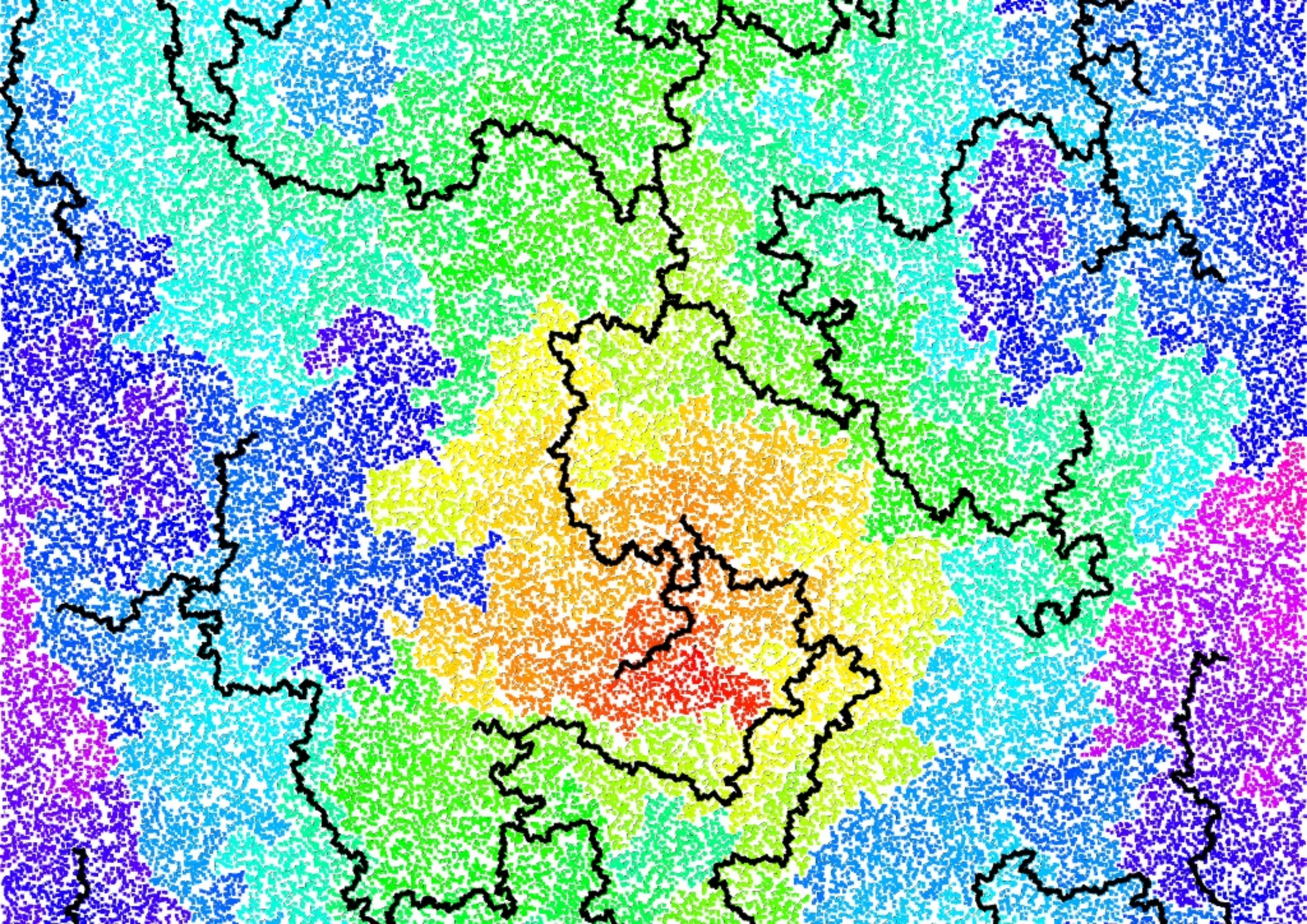
Percolation-MST coupling then suggests:

Construction of major routes of MST completed by time $\approx p_c$
Rural routes connect to nearby highways.

If $d_{MST}(v,w) :=$ Graph distance between v and w in MST
$d_{HW}(v,w) :=$ # highway edges on v-w path (edges of $C_{p_c}(1)$, say)

Then $\sup\{\mathbb{E}(d_{MST}(v,w)/d_{HW}(v,w)) : v,w \in V\} < \infty.$





VII Dimension and volume growth

On n-vertex graph $G=(V,E)$ with percolation threshold p_c,

$$\sup\{d_{MST}(v,w) : v,w \in C_p(1)\} \asymp n^{\beta}$$

Then expect MST to have

- diameter $O(n^{\beta})$
- volume growth $\frac{1}{\beta}$ $\quad \mathbb{E}\left[\#\{w\in V : d_{MST}(v,w)\leq k\}\right] \asymp k^{1/\beta}$
- scaling limit with "intrinsic dimension" $1/\beta$.

VIII Results.

$\xrightarrow{d}$: converges in distribution

Let T_n be MST of K_n. Then

- $\dfrac{\mathrm{diam}(T_n)}{n^{1/3}} \xrightarrow{d} D$, for some continuous r.v. D, and $\mathbb{E}[\mathrm{diam}(T_n)/n^{1/3}] \longrightarrow \mathbb{E}D$.

- There is an infinite random rooted tree (T,ρ) such that $(T_n,\rho_n) \xrightarrow{d} (T,\rho)$ in the Benjamini-Schramm sense (convergence of local graph statistics) and $\mathbb{E}|B(\rho,k)| \asymp k^3$

- There is a random compact measured metric space $\mathcal{T} = (\mathcal{T}, d, \mu)$ with Minkowski dimension 3, such that the following holds.

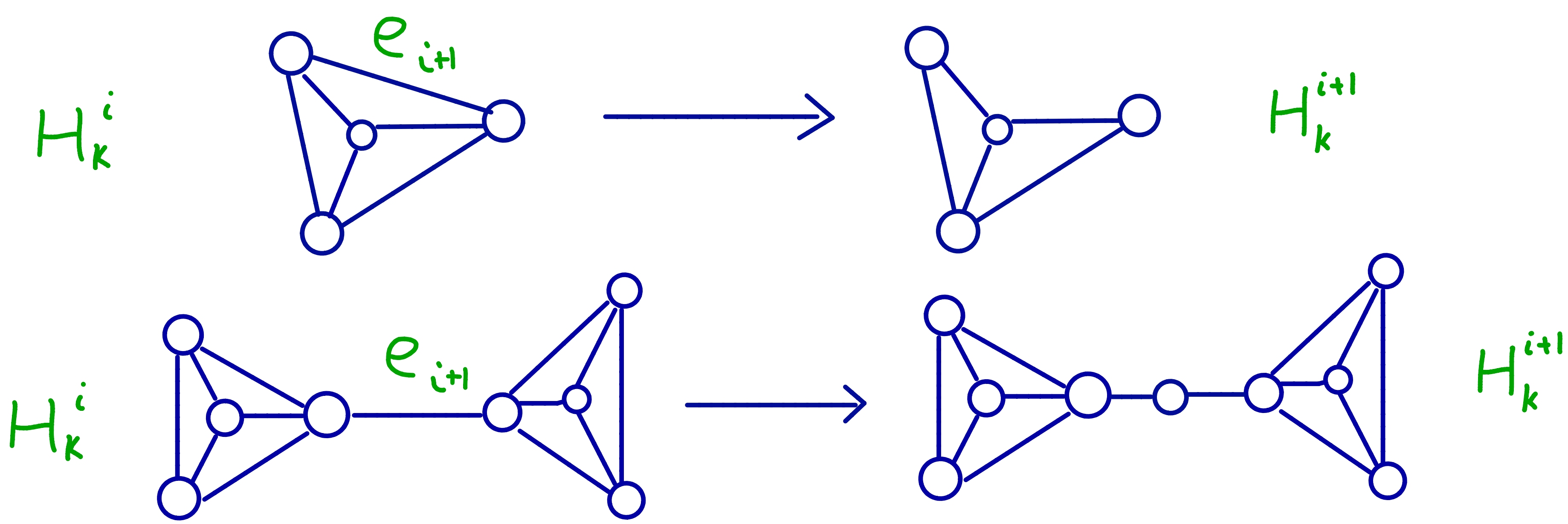
  Write V_n = vertices of T_n = vertices of K_n.

  Let $d_n = n^{-1/3} \cdot$ (graph distance in T_n), $\mu_n = \frac{1}{n} \sum_{v \in V_n} \delta_v$ = empirical measure on V_n

  Then $(V_n, d_n, \mu_n) \xrightarrow{d} \mathcal{T}$ in the Gromov-Hausdorff sense.

IX Alternate description of scaling limit $\mathcal{T}$

- Let H_k^0 be a random 3-regular graph with k vertices.
- Given H_k^i, choose an edge e_{i+1} of H_k^i uniformly at random.
  - Remove e_{i+1} to create H_k^{i+1} unless this would disconnect H_k^i



- Repeat until no more cycles; call result H_k.
- Scale distances by $c \cdot k^{-1/3}$, let $k \to \infty$. Result has same law as $\mathcal{T}$.

<u>Corollary</u>: $\mathcal{T}$ is also the scaling limit of the MST of random 3-regular graphs.

X Universality, extensions

The tree $\mathcal{T}$ is expected to be the MST scaling limit on any "high-dimensional" graph, e.g.

- random d-regular graphs
- the hypercube
- the lattice torus $(\mathbb{Z}/n\mathbb{Z})^d$ for $d > 8$

For low dimensions the picture is blurry.

$d=2$: The MST scaling limit (as an embedded object) exists and has dimension $\in (1+\varepsilon, 7/4)$. (Garban, Pete, Schramm 20??)

Numerics suggest dimension 1.22... no non-numerical predictions in physics literature.

$3 \leq d \leq 6$, Wide open. $6 < d \leq 8$ Opinions vary.

XI References

LAB, Nicolas Broutin, Bruce Reed, Critical random graphs and the structure of a minimum spanning tree, Random Structures and Algorithms, Vol 35, Issue 3 (2009), 323-347.

LAB, Nicolas Broutin, Christina Goldschmidt, Critical random graphs: limiting constructions and distributional properties, Electronic Journal of Probability, Vol 15 (2010), 741-775.

LAB, Nicolas Broutin, Christina Goldschmidt, The continuum limit of critical random graphs, Probability Theory and Related Fields, Vol 152 (2012), 367-406.

LAB, Ross Kang, Simon Griffiths, Invasion percolation on the Poisson-weighted infinite tree, Annals of Applied Probability, Vol 22 (2012), 931-970.

LAB, Partition functions of discrete coalescents: from Cayley's formula to Frieze's ζ(3) limit theorem, in XI Symposium on Probability and Stochastic Processes. Progress in Probability Volume 68 (2015). Birkhauser, Basel.

LAB, Nicolas Broutin, Christina Goldschmidt, Grégory Miermont, The scaling limit of the minimum spanning tree of the complete graph, Annals of Probability, to appear.


IMPACASTORINA
IMPA·NWL - NET
Thank you !