Convergence and limits of finite trees

Gábor Elek and Gábor Tardos

“Warwick”, October 23, 2020

limit of (dense) graphs

Lovász, Szegedy at al.

- random k-sample of vertices of a fixed graph G

↓ induces

- random graph R on k (labeled) vertices

↓

- subgraph densities: d(K,G) = Pr(R=K)

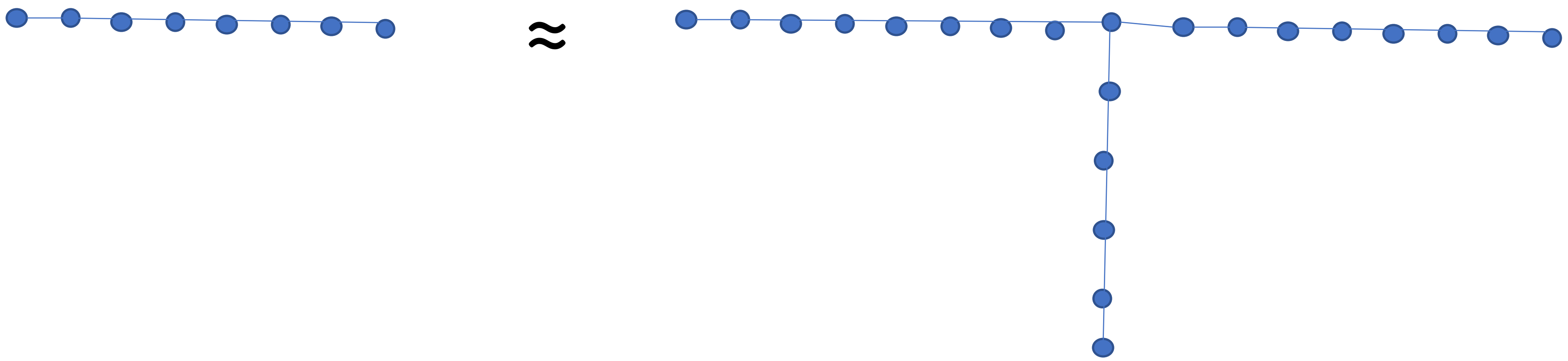
- (G_i) is a convergent sequence if $d(K,G_i)$ is convergent for every K

What is the limit? Need $G_i \to G^*$ with $d(K,G_i) \to d(K,G^*)$

Graphons: $G^*: [0,1]^2 \to [0,1]$ symmetric, measurable (not unique)

limit of trees

- trees are sparse → random samples induce the empty graph
- sparse graph limits (Benjamini-Schramm):
  small neighborhood of a random point
  too local



- we want global limits instead

limit of trees

- trees are sparse → random samples induce the empty graph
- consider them as a dense structure:

graph distance

need: normalization
here: with diameter

- random k-sample of vertices of a fixed tree T

↓ distances

- random k by k real matrix M_k – distributed as $\tau_k(T)$

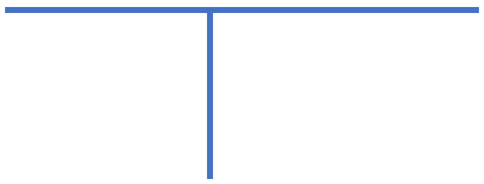
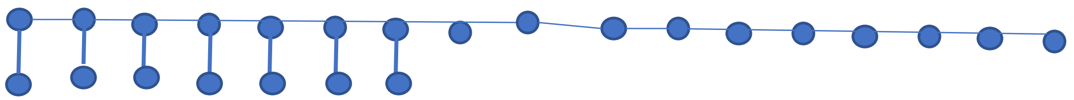
↓

- convergence in distribution

(T_i) is a convergent sequence if $\tau_k(T_i)$ weakly converges for all k

limit of trees

intuitive examples

- paths   (P_n) — clearly convergent
  limit: unit interval:
  uniform
  for sampling we need distance and distribution on it
- 3 equal paths from common root — also convergent
  
  limit: 3 intervals:
  distance: each has length ½
  distribution: uniform
- combs
  
  — convergent
  limit: unit interval:
  distribution: not uniform
- stars   ($K_{1,n}$) — clearly convergent
  almost every distance = diameter

measured real trees

many equivalent definitions

- Real tree = complete metric space (X,d)

  +

  ∀x,y ∃ isometric embedding of length d(x,y) interval connecting them

  +

  ∀ intermediate point of embedding separates x from y

- Measured real tree = real tree + probability Borel measure

  (a metric measure space)

- sampling makes sense,

  so does the distribution $\tau_k(T)$ for a measured real tree T

- $(T_i) \to T$ iff $\forall\, k\colon \tau_k(T_i) \to \tau_k(T)$

Ideally we had

1. Each convergent sequence of finite trees has a limit measured real tree.
2. Every measured real tree (of diameter $\leqq 1$) is a limit of finite trees.
3. The limit is unique up to isomorphism.

PROBLEMS

- Measure 0 branch of a real tree has no effect on sampling
  Let's forbid measure 0 branches.
- What about the sequence of stars??
  do not converge to a separable real tree
  Let's add "distance" parameter from tree to sample
  the distribution is on $T\times[0,\infty)$
  $d((x,a),(y,b))=d(x,y)+a+b$
  for stars: $T = \{p\}$, the distribution concentrates on $(p,1/2)$

makes all such real trees separable

The limit object: dendron

a marked metric measure space
Depperschmidt, Greven, Pfaffelhuber

- Dendron: $D = (T,\mu)$, where
  $T=(T,d)$ is a (separable) real tree
  μ is a probability Borel measure on $T\times[0,\infty)$
  such that $\mu(B\times[0,\infty))>0$ for any branch B of T
- Sampling: k i.i.d. $(x_i,a_i) \sim \mu$
- Distance matrix $M_k = (d_{ij})$ with $d_{ij} = d(x_i,x_j)+a_i+a_j$, $d_{ii} = 0$
- $\tau_k(D)$ is the distribution of M_k
- We write $(T_i) \to D$ iff $\forall$ k: $\tau_k(T_i) \to \tau_k(D)$

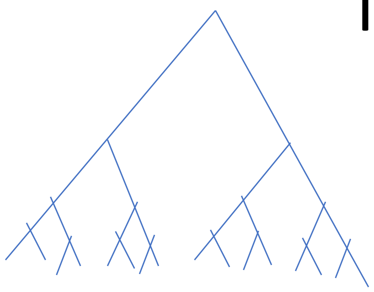
Results

1. Each convergent sequence of finite trees has a limit dendron.
2. Every dendron (of essential diameter $\leqq 1$) is a limit of finite trees.
3. The limit is unique up to isomorphism.
4. The space of dendrons of essential diameter $\leqq 1$ is compact.

More examples

Svante Janson

- Stars ($K_{1,i}$) → (T,μ), where T = {p}, μ concentrates on (p,1/2)
- complete binary tree of depth i: same limit as for stars
- comp. binary tree with edges on level j replaced by paths of length 2^{i-j}:
  limit = fractal tree with uniform distribution on Cantor set



- many paths from a single high degree vertex
  (every single path has o(n) vertices)
  limit = single point tree, but tricky distribution on [0,1/2]

Random tree limits

Svante Janson

- Conditioned critical Galton-Watson trees (finite variance)

↓ in distribution
with slightly different scaling

Aldous's Continuum Random Tree

↑ in distribution
with slightly different scaling

- Random tree on n labelled vertices

- Conditioned subcritical Galton-Watson trees (+mild conditions)

↓ in distribution
no scaling

D=(T,μ), with T={p}, μ = geometric distribution

Sampling dense graphs versus trees
An important difference

- Let G_1 and G_2 be two independent samples from the random graph G(n,1/2).

  G_1 and G_2 are very close in sampling: a random k-sample is close to G(k,1/2)
  (small cut distance too)
  G_1 and G_2 are very far in edit distance
  (need to add/delete cca $n^2/4$ edges to make G_1 isomorphic to G_2)

- If T_1 and T_2 are two n vertex trees that are close in sampling distance,

  then there exists a bijection between the vertex sets with small distortion
  in the graph distance

The proof technique
ultraproducts

See also:

proof of the Hypergraph Regularity Lemma by Elek and Szegedy

Advantages:

- Very “clean” finite-infinite transitions
- Conceptually straight forward
- Compact proofs of qualitative statements (like existence, unicity, etc.)

Disadvantages:

- Technical details
- Absolutely no quantitative bounds

The proof technique

ultraproducts

- Fix ultrafilter $\boldsymbol{\omega}$ on the positive integers: the family of LARGE sets
- ultralimit: $\lim_{\omega} x_i = x$ if every neighborhood of x contains a LARGE subsequence of x_i.
  ALWAYS exists and unique in a compact Hausdorff space
- ultraproduct: $\prod_{\omega} A_i$: combines a sequence of structures A_i, keeps properties held by a LARGE subsequence

A_i can be
– sets
– metric spaces, van den Dries, Wilkie
– probability measure spaces, Elek, Szegedy
– (marked) metric measure spaces, etc.

Proof sketch

Existence

- T_i are finite trees, therefore also measured real trees
- $T^* := \prod_{\omega} T_i$: also a real tree, but ugly (not separable)
  also a probability measure space but ugly (not a Borel measure)
- "Prune" T*: remove branches with measure 0.
  T := core of T* = what remains after pruning
- T is nice (separable)
- Some part of the probability distribution may be outside T
- Replace $x \in T^*$ with $(p,a) \in T \times [0,\infty)$, where p = projection of x to T
  a = distance of x from T
- Dist. on T* is transformed to μ on $T \times [0,\infty)$
- $D = (T,\mu)$ is the ultralimit dendron
- $\lim_{\omega} \tau_k(T_i) = \tau_k(D)$
- If (T_i) is convergent, then $\tau_k(T_i) \to \tau_k(D)$, $T_i \to D$, Q.E.D.

Proof sketch

Uniqeness

- T is a measured real tree (works for dendrons too)
- $(x_1,\ldots,x_k)$ is a k-sample from T,
  T_k is subtree spanned by these points.
- $d(x_i,x_j)$ is enough to reconstruct T_k upto isomorphism.
- The core of $\prod_{\omega} T_k$ is isomorphic to T almost surely.
- T is reconstructed from $\tau_k(T)$, therefore the limit is unique. Q.E.D.

Proof sketch

Finite trees are dense among dendrons (of essential diameter ≦ 1)

- T is a measured real tree (works for dendrons too)
- We saw $T_k \to T$, where T_k is a is a real tree spanned by k points
- So finitely spanned real trees are dense
- It is easy to approximate a finitely spanned real tree with finite trees.
  (if the essential diameter is ≦ 1)

Q.E.D.

Missing numeric results

- Trees close in sampling $\Rightarrow$ $\exists$ vertex-bijection with small distortion
- A k-sample is close in sampling to the full tree with high probability

THANK YOU