LONGEST AND HEAVIEST PATHS IN DIRECTED RANDOM GRAPHS

Sergey Foss

based on joint work with Takis Konstantopoulos
and several other collaborators:
James Martin, Denis Denisov, Sanjay Ramassamy, Bastien Mallein,
Artem Pyatkin, and others

12 February 2021

What this is about

Erdős-Rényi graph: the most known random graph model
Barak-Erdős graph: a directed version of the above
Additional stuff like
– structure on the vertex set (e.g. a natural deterministic metric or a deterministic partial ordering)
– weights on the edges and/or the vertices
– make probabilities depend on the structure of the vertex set

We like to consider longest paths from a vertex to a "faraway" vertex and examine how the length grows. If we count weights instead of lengths, we say "heaviest" rather than "longest" path

Questions similar to last passage percolation models

"Dual model " – Infinite bin model for Barak-Erdős graph

"Perfect" simulation algorithm for IBM and its generalisations

Why?

Interesting mathematics and probability. In particular, interesting limits.

Models in statistical physics.

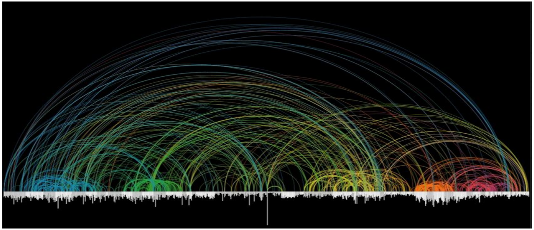
Appears in connection to parallell processing systems.

Mathematical biology: ecology (food chains) models.

Computer science (but questions/regimes may be different).

The Barak-Erdős graph

Vertex set $= \mathbb{Z}$. For $i < j$, declare (i, j) is an edge with probability p, independently (think of iid RVs α_{ij}: here $\alpha_{ij} = 1$ if the edge exists or $\alpha_{ij} = -\infty$, otherwise)



Let L_n be the maximum length of all paths from 0 to n.
Example. Let $n = 5$. Assume there are edges $(0, 1), (0, 2), (0, 4), (0, 5), (1, 2), (1, 3), (1, 5), (2, 3), (2, 4), (3, 4)$.
Then the longest path is $(0, 1, 2, 3, 4)$ and $L_5 = 4$.

Basic results

Proposition (based on the Kingman's subadditive ergodic theorem). For any $p \in [0,1]$ and as $n \to \infty$,

$$\frac{L_n}{n} \to C(p) \text{ a.s. and in } \mathcal{L}^1, \text{ i.e.}$$

$$\mathbb{E}\left|\frac{L_n}{n} - C(p)\right| \to 0.$$

Further, $p \mapsto C(p)$ is a deterministic continuous increasing function of p and $C(p) > p$ for $0 < p < 1$.
[This result holds even in a stationary-ergodic framework.]

Idea:

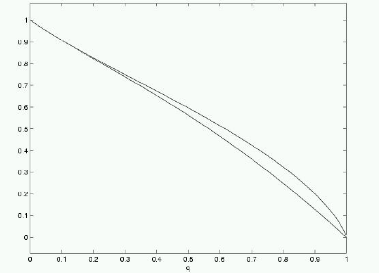
$L_n \equiv L_{0,n}$. Introduce further $L_{n,n+m}$ as the max path length within n and $n+m$. Then

$$L_{0,n+m} \leq L_{0,n} + 1 + L_{n,n+m},$$
$$L_{0,n+m} + 1 \leq (L_{0,n} + 1) + (L_{n,n+m} + 1).$$

This is the sub-additivity.

On the function $C(p) \equiv C(1-q)$

$C(1-q)$ as a function of $q = 1-p$: analytically obtained upper and lower bounds

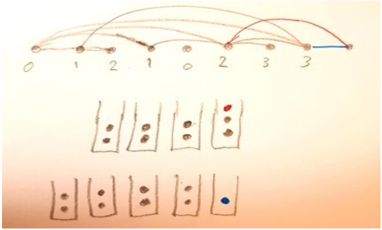


Methods used: extended renovation theory, construction of infinite bin model.

N.B. max length from 0 *to* n *and max length within* $[0, n]$ *are asymptotically identical (hence every two far apart vertices are a.s. connected)*

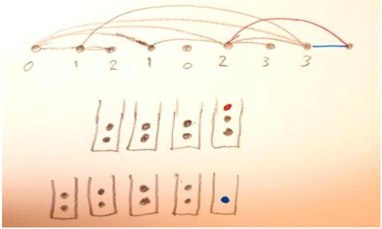
The Infinite Bin Model

Idea: Grow the graph little by little. Start with one point. Then go from graph on $\{0, \ldots, n\}$ to $\{0, \ldots, n+1\}$.
If *we only care about maximal paths* then we only need to know the "mark" of each of $i = 0, \ldots, n$ (this is the maximum of lengths of all paths ending at i).
$X_n(0)$, number of vertices having mark L_n
$X_n(-k)$, number of vertices having mark $L_n - k$, for $k = 0, \ldots, L_n$.



Here are vertices $0, \ldots, 7$ and we add vertex 8.

Dynamics of IBM



Each time n we have $n+1$ balls placed in L_n+1 bins numbered $0,-1,\dots,-L_n$ (and all these bins are non-empty).
Each time n each of $n+1$ balls becomes "active" with probability p, independently of anything else.
Find the right-most "active" bin with an active ball.
Then place a new ball into the bin to the right of the active bin.
If there is no active balls, place the new one in the left-most bin.

Stability

As n increases to infinity, the number of non-empty bins increases to infinity too. We can represent the state of the system as an infinite-dimensional vector with non-negative integer-valued coordinates where only finitely many of them are positive. In the limit (if it exists), all coordinates of vectors are positive.

Theorem $X_n = [X_n(0), X_n(-1), \ldots, X_n(-L_n)]$ *is a Markov process in* $\mathbb{N}^* \cup \mathbb{N}^\infty$ *that has a unique stationary version (process with values in* $\mathbb{N}^\infty$*) (and we have convergence in TV of projections).*

Functional Limits

FCLT1:

$$\left\{ n^{-1/2} \left(\sum_{k=0}^{[nC(p)t]} X_n(-k) - nt \right), t \geq 0 \right\} \to (W_t, t \geq 0),$$

weakly, where W is a Brownian motion.
FCLT0:

$$\left\{ n^{-1/2} \left(\sum_{k=0}^{[L_n t]} X_n(-k) - nt \right), 0 \leq t \leq 1 \right\} \to (W_t^0, 0 \leq t \leq 1),$$

weakly, where W^0 is a Brownian bridge.
Corollary: CLT FOR L_n

In-depth study of the IBM and $C(p)$ by Mallein + Ramassamy

MR1 generalized IBM to IBM(μ) where the next ball is selected by means of a random variable with distribution μ.
$\mu =$ geometric $\rightarrow$ IBM(μ).
$\mu =$ uniform the IBM(μ) $\rightarrow$ Aldous+Pitman 1983.
$v_\mu =$ speed of the front of IBM(μ). $v_{\text{geom}(p)} = C(p)$:

$$C(p) = ep - \frac{e\pi^2}{2} p(\log p)^{-2} + o(p(\log p)^{-2}).$$

Obtained by coupling the infinite-bin model with uniform distribution with a continuous-time branching random walk with selection. A series representation of $C(p)$ is in MR2.
MR1: Barak-Erdős graphs and the infinite-bin model
MR2: Two-sided infinite-bin models and analyticity for Barak-Erdős graphs (2019)

When $L_{n+1} = L_n + 1$?

We know that either $L_{n+1} = L_n$ or $L_{n+1} = L_n + 1$.
Further, the conditional probability

$$\mathbb{P}(L_{n+1} = L_n + 1 \mid X_n(0) = k) = 1 - (1-p)^k.$$

Lemma let $X = (X(0), X(-1), \ldots, X(-k), \ldots)$ be the stationary (limiting) random vector for Markov chain $\{X_n\}$. Then

$$C(p) = \mathbb{E}(1 - (1-p)^{X(0)}).$$

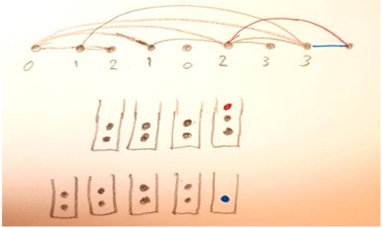
Corollary. Let $X^{(1)}, X^{(2)}, \ldots$ be i.i.d. random variables having the same distribution with $X(0)$. Let $Y_i = 1 - (1-p)^{X^{(i)}}$, for $i = 1, 2, \ldots$ and $\overline{Y}_n = (Y_1 + \ldots + Y_n)/n$ their averages. Then

$$\mathbb{E}\overline{Y}_n = C(p) \text{ and } \overline{Y}_n \to C(p)$$

exponentially fast, both a.s. and in $\mathcal{L}^1$.
Same for stationary, but slower convergence rate.

Regenerative points for the IBM

Each time n take a new numbering of all balls in all bins, starting from the bottom right to the top left. For example, in



we have, for $n = 7$, two balls in box 0 get numbers 1 and 2; then two balls in box (-1) numbers 3 and 4, etc. Then, for $n = 8$, we have different numberings, depending on placing the new ball. Let β_{n+1} be the number of the first "active" ball at time n. Then we may assume that β_{n+1} has a geometric distribution with parameter p.

Consider the following events:

$$A_{n,n+m} = \{\beta_{n+1} \le 1, \beta_{n+2} \le 2, \ldots, \beta_{n+m} \le m\},$$
$$A_n = \{\beta_{n+i} \le n+i, \text{ for all } i \ge 1\}.$$

(here $\beta_{n+1} \le 1$ means $\beta_{n+1} = 1$).
Observation: Given $A_{n,n+m}$, the new balls (after the n'th) are placed independently of locations of the earlier balls.
Next: with $q = 1 - p$,

$$\mathbb{P}(A_{n,n+m}) = \prod_{i=1}^{m}(1 - q^i) \text{ and}$$
$$\mathbb{P}(A_n) = \prod_{i=1}^{\infty}(1 - q^i) > 0 \text{ if } q \in [0, 1).$$

Thus, random times n when events A_n occur form regenerative epochs.

Perfect simulation for $C(p)$

This is a simulation of i.i.d. samples having mean $C(p)$.
Consider events $A_{-k,0}$.
Run k backwards:

$$\nu = \min\{k : \mathbf{I}(A_{-k,0}) = 1\}.$$

Starting from "time" $-\nu$, build the IBM forwards: we consider vertices numbered
$-\nu, -\nu + 1, \ldots, 0$.
Then at "time" 0 look at the number Y of balls in the right-most bin.

Proposition Y has the same distribution with $X(0)$, the right-most coordinate of the limiting vector.

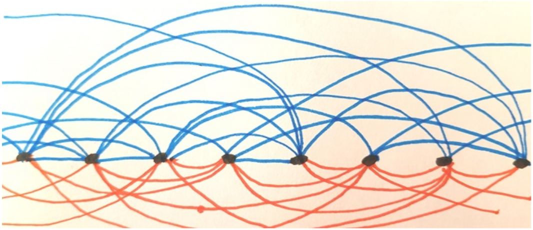
(?) How to produce many samples? I.i.d. or stationary?

Second model: Binary decoration

Model suggested by M+R. Vertex set $\mathbb{Z}$. Let x vary in $[-\infty, \infty)$.

$\alpha_{i,i}$, $i < j$, iid Bernoulli RVs, $\mathbb{P}(\alpha_{i,j} = 1) = p$.

$w^x_{i,j} = \alpha_{i,j} + x(1 - \alpha_{i,j})$, weight of (i, j).



blue edge (probability p) has weight 1

red edge (probability $1 - p$) has weight x, which could be negative

Binary decoration II

The corresponding weighted graph is called $G_p(x)$. We are interested in heavy paths:

$$W_{i,j}^x = \text{ maximum weight of all paths from } i \text{ to } j.$$

Then $G_p(-\infty)$ is the BE graph insofar as the quantity $W_{i,j}^{-\infty}$ (= maximum length of all paths from i to i) is of interest.
Theorem *For $x \geq -\infty$,*

$$\lim_{n\to\infty} W_{0,n}^x/n = \lim_{n\to\infty} (W_{0,n}^x)^+/n = C_p(x),$$

where $C_p(x)$ is deterministic with $\lim_{x\to-\infty} C_p(x) = C_p$, the Mallein-Ramassamy analytic function of p.

Binary decoration III

We are interested in the behavior of $C_p(x)$ as a function of x for fixed p. Let
$\Gamma_1 = \min\{n \geq 1 : \mathbf{I}(A_n) = 1\}$,
$\Gamma_2 = \min\{n > \Gamma_1 : \mathbf{I}(A_n) = 1\}$,
$\gamma = (\mathbb{E}(\Gamma_2 - \Gamma_1))^{-1}$,
$W^x_{\Gamma_1,\Gamma_2}$, the maximal path weight between Γ_1 and Γ_2.

Theorem *For $x < 2$,*

$$C_p(x) = \gamma \mathbb{E}[W^x_{\Gamma_1,\Gamma_2}],$$

– *$C_p(x)$ is a convex function of x,*
– $\lim_{x\to\infty} C_p(x)/x = C_{1-p}(0)$,
– *Right/Left derivatives $D^{\pm} C_p(x)$ exist at every x,*
– *If x is irrational then $D^+ C_p(x) = D^- C_p(x)$*

Binary decoration IV: main result

$C_p(x)$ is differentiable at every irrational x.
But the converse is not true.

Theorem *The set of x for which $C_p(x)$ fails to be differentiable is equal to the union of*
1) nonpositive rationals;
2) positive integers except 1;
3) the reciprocals of positive integers except 1.

Binary decoration: the intuition behind

Let $a, b, c, d > 0$ and $b/a \neq d/c$.
Then $\max(a + bx, c + dx)$ is differentiable everywhere except x_0: $a + bx_0 = c + dx_0$.
Same for any finitely many...
Now: for any x, $W^{x}_{\Gamma_1,\Gamma_2}$ is the maximum of a finitely many linear functions almost surely.
And if two or more functions with different slopes provide max at some x – no differentiability...

Perfect simulation for $C_p(x)$ in the case $x < 1$.

We follow again the backward-forward procedure.

Step 1. Simulate ν backwards for the "$-\infty$–or–1" Barak-Erdos graph.

Step 2. We have already simulated, which edges within $[-\nu, 0]$ have weight 1 and which edges have weight $-\infty$.
Now replace weight $-\infty$ by weight x and run the algorithm forward.
Then at time 0 consider all paths that start from time $-\nu$, and look at the set A of all edges $i \in \{-\nu, \dots, 0\}$ whose weight, W_i, differs from the maximal weight, W_{max} by 1 or less.
Then let

$$\Delta_p(x) := \max_{i \in A}(W_i + w_{i,1} - W_{max})^+.$$

Theorem We have $\mathbb{E}\Delta_p(x) = C_p(x)$.

Further models

More general decorations on BE graphs

We can add weights with an arbitrary distribution on the edges of a BE graph.

FMS11: If $\mathbb{E}U^2 < \infty$ then skeleton points still exist and regenerative structure allows the proof of a LLN and a CLT (more conditions needed for CLT). Different limiting results can be obtained under the assumptions that $\mathbb{E}U^2 = \infty$, and $\mathbb{P}(U > x)$ is regularly-varying with index $0 < s < 2$ (FMS11).

We can add weights on the edges (U) and on the vertices (V) of a BE graph. FK18: A regenerative structure, and hence limit theorems, is possible under the assumptions $P(V \geq 0) = 1$, $\mathbb{E}V < \infty$, $\mathbb{E}U > 0$, $\mathbb{E}(U^+)^2 < \infty$. Under the same assumptions, under Poissonian scaling, we can prove that the decorated BE graph converges weakly to a decorated PWIF.

Elastic graph

$p \equiv p_{i,j} \equiv p(j-i)$, $-\infty < i < j < k$ such that $\sum_{k=1}^{\infty} k(1-p_1)\cdots(1-p_k) < \infty$ (p_k can drop to 0 but not too fast)

SKELETON POINTS: $\mathcal{S}$ be the set of vertices i s.t. for any $i' < i < i''$ there is a path $i' \to i''$ containing i.

Theorem *$\mathcal{S}$ is a stationary renewal process (I mean, really, that the random measure $\xi = \sum_{i \in \mathcal{S}} \delta_i$ is stationary and its Palm version is a renewal process). Moreover, the graph regenerates over $\mathcal{S}$. The rate of $\mathcal{S}$ is $\gamma = \prod_{j=1}^{\infty}(1-(1-p_1)\cdots(1-p_j))^2$.*

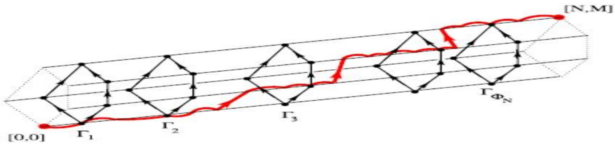
Corollary *Max length of all paths from 1 to n is a sum of i.i.d. random variables plust two negligible independent random variables.*

Ergo, maxima of random variables become sums of i.i.d. random variables and we're at home.

LLN and CLT: $L_n/n \to C$ and $(L_{[nt]} - Cnt)/\sqrt{\gamma\sigma^2 n}$ converges to a standard BM. Here, $\sigma^2 = \mathrm{var}(L(\Gamma_1, \Gamma_2) - C(\Gamma_2 - \Gamma_1))$.

Elastic slabgraph

$(I, \preceq)$ some finite partially ordered set with a least element 0 and a greatest element M and let $H(I, \preceq)$ be its Hasse diagram.



Limiting process of the FCLT: Let B^{ι}, $\iota \in I$ i.i.d. BMs.
$\iota = (0 = \iota_0 \prec \iota_1 \prec \cdots \prec \iota_r = M)$ path im Hasse diagram.

$$Z^{(\iota)}(t) :=$$
$$\sup \left\{ B^{(\iota_0)}(t_0) + [B^{(\iota_1)}(t_1) - B^{(\iota_1)}(t_0)] + \cdots + [B^{(\iota_r)}(t_r) - B^{(\iota_r)}(t_{r-1})] \right\},$$

where sup is over all $0 \leq t_0 \leq \cdots \leq t_r = t$,

$$Z(t) = \max_{\iota} Z^{(\iota)}(t).$$

Elastic slabgraph and GUE (Gaussian unitary ensemble)

Suppose that $I = \{0, 1, \ldots, M-1\}$, $\prec$ is the natural order. E.g, with $M = 2$,

$$Z(t) = \sup_{0 \le s \le t} \{B^{(0)}(s) + B^{(1)}(t) - B^{(1)}(s)\}.$$

Let λ_M be the largest eigenvalue of an $M \times M$ GUE random matrix. Then

$$Z(t) \overset{(\mathrm{d})}{=} \sqrt{t}\lambda_M.$$

Letting $M \to \infty$,

$$M^{1/6}(\lambda_M - 2\sqrt{M}) \to F_{TW},$$

in distribution, where F_{TW} is the Tracy-Widom distribution.

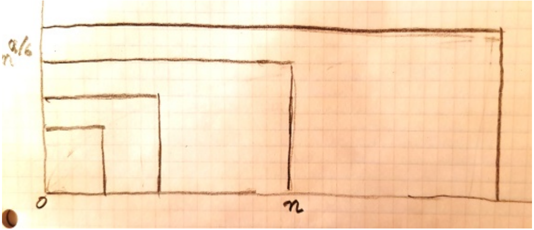
Obtaining F_{TW} via a single limit I

Vertex set = $\mathbb{Z}_+ \times \mathbb{Z}_+$, with standard partial order

Simplest case: Put a directed edge from a vertex to a bigger vertex with probability p.

Theorem *Let $L_{n,m}$ be the maximum length of all paths from $(0,0)$ to (n,m). Then, for some positive constants C, C_1, and any $0 < a < 3/14$, we have*

$$n^{a/6}\left(\frac{L_{n,n^a} - Cn}{C_1\sqrt{n}} - 2n^{a/2}\right) \to F_{TW}.$$



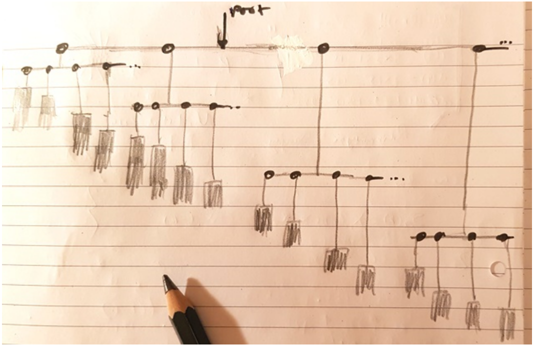
Obtaining F_{TW} via a single limit II

Method: Clipping the graph up to finite height allows us to use skeleton points. Then use the corresponding result for Brownian last passage percolation and Komlos-Major-Tusnady embedding.

The constant $3/14$ is not optimal. For the corresponding last passage percolation problem with iid exponential weights on vertices the constant is 1 (Johansson).

PWIT limit I

PWIT = Poisson-Weighted Infinite Tree (terminology due to Aldous and Steele, appears in combinatorial optimization problems)



The PWIT is an infinite infinitary tree.

PWIT limit II

Related to
CCM = continuum cascade model (Physics literature):
Let Φ^x, $x \geq 0$, be a collection of iid stationary Poisson processes.
Declare (x, y) to be an edge if y is a point of $x + \Phi^x$.

PWIT = the continuous connected component of CCM containing 0.

Hence CCM = PWIF = Poisson-Weighted Infinite Forest.
Theorem *Consider the BE graph G_n on $\frac{1}{n}\mathbb{Z}$ with $p = 1/n + o(1/n)$. Then $G_n \to$ PWIF, weakly, while $G_n^0 \to$ PWIF, weakly.*

Take-home message: In a Poissonian limiting regime, the BE graph looks like a random self-similar tree and hence recursions are possible.

Some open problems

1) Recursive difference-differential equations obtained from the decorated PWIT must be solved. How? What kind of hopefully meaningful approximations give?
2) The F_{TW} is obtained by the sliding flat window. Can we get rid of this restriction?
3) Understand the connection to random matrices directly and not via Brownian last-passage percolation.
4) 2D vertex sets give rise to GUE eigenvalues and the F_{TW}. What about 3D?
5) Let S_t be the point s at which the maximum $Z(t) = \max_{0 \le s \le t}\{B^{(0)}(s) + B^{(1)}(t) - B^{(1)}(s)\}$ is achieved. Find the law of S_t.
6) Applications are important. Think of meaningful questions.
7) What can we say about $C_p(x)$ jointly in (p, x)?