Cross-Multiplicative Coalescence and Minimal Spanning Trees of Irregular Graphs

Yevgeniy Kovchegov
Oregon State University

joint work with
Peter T. Otto (Willamette U.) and
Anatoly Yambartsev (IMS, U. of São Paulo)

A quote.

Aldous (1998): *It turns out that there is a large scientific literature relevant to the Marcus-Lushnikov process, mostly focusing on its deterministic approximation. Curiously, this literature has been largely ignored by random graph theorists*.

Erdős-Rényi random graph.

Erdős-Rényi random graph on K_n:

• Edge e is an associated a uniform random variable U_e over $[0, 1]$. The random variables $\{U_e\}_e$ are assumed to be independent.

• For the "time" parameter $p \in [0, 1]$, an edge e is considered "open" if $U_e \leq p$. Obtaining the Erdős-Rényi random graph $G(n, p)$.

Minimal spanning tree in K_n:

• A spanning tree in K_n with minimal $\sum_e U_e$ is called the **minimal spanning tree**. Let $L_n = \sum_e U_e$ denote its length.

The length of the minimal spanning tree in K_n.

A. Frieze (1985) using the results of **P. Erdős and A. Rényi (1960)** showed for Erdős-Rényi random graph on K_n:

The limiting mean length of the minimal spanning tree in K_n is

$$\lim_{n\to\infty} E[L_n] = \sum_{k=1}^{\infty} \int_0^{\infty} \frac{k^{k-2}t^{k-1}}{k!} e^{-kt} dt = \zeta(3),$$

where

$$\zeta(3) = \sum_{k=1}^{\infty} \frac{1}{k^3} = 1.202\ldots$$

is the value of the Riemann zeta function at 3.

The minimal spanning tree in a regular graph.

Frieze and McDiarmid (1989) derive the formula for the limiting mean length of the minimal spanning tree of the Erdős-Rényi random graph process on **regular graphs**, as multiples of $\zeta(3)$.

In particular, for the Erdős-Rényi random graph process on $K_{n,n}$,

$$\lim_{n \to \infty} E[L_n] = 2\zeta(3)$$

Main results

Novel approach to finding minimal spanning tree asymptotics via the hydrodynamic limits and the analysis of the corresponding **'reduced'** Smoluchowski coagulation equations.

Theorem [YK, Otto, and Yambartsev, 2017]. Let $\alpha, \beta > 0$, $\gamma = \alpha/\beta$, and $L_n = L_n(\alpha, \beta)$ be the length of a minimal spanning tree on a complete bipartite graph $K_{\alpha[n],\beta[n]}$ with partitions of size

$$\alpha[n] = \alpha n + o(\sqrt{n}) \quad \text{and} \quad \beta[n] = \beta n + o(\sqrt{n})$$

and independent uniform edge weights over $[0,1]$. Then

$$\lim_{n\to\infty} E[L_n] = \gamma + \frac{1}{\gamma} + \sum_{i_1\geq 1;\ i_2\geq 1} \frac{(i_1+i_2-1)!}{i_1!i_2!} \frac{\gamma^{i_1} i_1^{i_2-1} i_2^{i_1-1}}{(i_1+\gamma i_2)^{i_1+i_2}}.$$

A formula for $E[L_n]$.

S. Janson (1995) for K_n

Beveridge, Frieze, and McDiarmid (1998) in general:

For all connected graphs with i.i.d. uniform $[0,1]$ edge lengths,

$$E[L_n] = \int_0^1 E[\kappa(G(n,p))]dp - 1,$$

where $\kappa(G(n,p))$ is the number of components in the Erdős-Rényi random graph $G(n,p)$.

Coalescent processes.

• The process begins with n singletons (clusters of mass one).

• The cluster formation is governed by a symmetric collision rate kernel $K(i,j) = K(j,i) > 0$.

• A pair of clusters with masses (weights) i and j coalesces at the rate $K(i,j)/n$, independently of the other pairs, to form a new cluster of mass $i+j$.

• The process continues until there is a single cluster of mass n.

Famous kernels: Kingman's $K(i,j) \equiv 1$, Additive $K(i,j) = i+j$, and Multiplicative $K(i,j) = ij$.

Marcus-Lushnikov processes.

The Marcus-Lushnikov process

$$\mathbf{ML}_n(t) = \Big(\zeta_{1,n}(t), \zeta_{2,n}(t), \ldots, \zeta_{n,n}(t), 0, 0, \ldots\Big)$$

is an auxiliary process to the corresponding coalescent process that keeps track of the numbers of clusters in each weight category.

Here $\zeta_{k,n}(t)$ denotes the number of clusters of weight k at time $t \geq 0$.

Since the coalescent process begins with n singletons,

$$\mathbf{ML}_n(0) = (n, 0, 0, \ldots).$$

Marcus-Lushnikov processes.

For the multiplicative kernel $K(i,j) = ij$, the process $\mathbf{ML}_n(t)$ describes cluster size dynamics of the Erdős-Rényi random graph process $G(n,p)$ on K_n with $p = 1 - e^{-t/n}$.

$$\begin{aligned}
\lim_{n\to\infty} E[L_n] &= \lim_{n\to\infty} \int_0^1 E[\kappa(G(n,p))]\,dp - 1 \\
&= \lim_{n\to\infty} \int_0^\infty \frac{1}{n} E[\kappa(G(n, 1-e^{-t/n}))] e^{-t/n} dt - 1 \\
&= \lim_{n\to\infty} \int_0^\infty \sum_{k=1}^{\infty} \frac{E[\zeta_{k,n}(t)]}{n} e^{-t/n} dt - 1.
\end{aligned}$$

Smoluchowski coagulation equations.

Smoluchowski coagulation equations for the multiplicative kernel $K(i,j) = ij$ are

$$\frac{d}{dt}\zeta_k = -k\zeta_k \sum_{j=1}^{\infty} j\zeta_j + \frac{1}{2}\sum_{j=1}^{k-1} j(k-j)\zeta_j\zeta_{k-j}, \qquad \zeta_k(0) = \delta_{1,k}.$$

They are suppose to describe the deterministic dynamics of the limiting fractions in the Marcus-Lushnikov processes,

$$\zeta_k(t) = \lim_{n\to\infty} \frac{\zeta_{k,n}(t)}{n},$$

BUT...

Problem with conservation of mass.

McLeod (1962) showed that the Smoluchowski coagulation equations

$$\frac{d}{dt}\zeta_k = -k\zeta_k \sum_{j=1}^{\infty} j\zeta_j + \frac{1}{2}\sum_{j=1}^{k-1} j(k-j)\zeta_j\zeta_{k-j}, \qquad \zeta_k(0) = \delta_{1,k}$$

have no solution past $T_{gel} = 1$.

Issue: Conservation of mass $\sum_{j=1}^{\infty} j\zeta_j(t) = 1$.

The problem is solved by introducing the reduced Smoluchowski system also known as the Flory's coagulation system:

$$\frac{d}{dt}\zeta_k = -k\zeta_k + \frac{1}{2}\sum_{j=1}^{k-1} j(k-j)\zeta_j\zeta_{k-j}, \qquad \zeta_k(0) = \delta_{1,k}.$$

Flory's coagulation system.

The solutions of the two systems

$$\frac{d}{dt}\zeta_k = -k\zeta_k \sum_{j=1}^{\infty} j\zeta_j + \frac{1}{2}\sum_{j=1}^{k-1} j(k-j)\zeta_j\zeta_{k-j}, \qquad \zeta_k(0) = \delta_{1,k}$$

and

$$\frac{d}{dt}\zeta_k = -k\zeta_k + \frac{1}{2}\sum_{j=1}^{k-1} j(k-j)\zeta_j\zeta_{k-j}, \qquad \zeta_k(0) = \delta_{1,k}$$

coincide up until the gelation time $T_{gel} = 1$.

They are

$$\zeta_k(t) = \frac{k^{k-2}t^{k-1}}{k!}e^{-kt}.$$

Flory's coagulation system.

The solutions of Flory's coagulation (reduced Smoluchowski) system

$$\zeta_k(t) = \frac{k^{k-2} t^{k-1}}{k!} e^{-kt}$$

satisfy the conservation of mass up until $T_{gel} = 1$:

$$\begin{cases} \sum_{k=1}^{\infty} k\zeta_k(t) = 1 & \text{if } t \leq T_{gel} \\ \sum_{k=1}^{\infty} k\zeta_k(t) < 1 & \text{if } t > T_{gel}. \end{cases}$$

This phenomenon is known as gelation. It reflects the emergence of a unique giant component in the Erdős-Rényi random graph process.

The hydrodynamic limit.

We us the weak law of large numbers of T. Kurtz to show the hydrodynamic limit

$$\lim_{n\to\infty} \frac{\zeta_{k,n}(t)}{n} = \zeta_k(t),$$

where for a fixed time $T > 0$ and a given integer $K > 0$, we have

$$\lim_{n\to\infty} \sup_{s\in[0,T]} \left| \sum_{k=1}^{K} n^{-1}\zeta_{k,n}(s) - \sum_{k=1}^{K} \zeta_k(s) \right| = 0 \qquad \text{a.s.}$$

using the limit theorems of T. Kurtz for density dependent population processes.
See **Kurtz (1981)** and **Ethier and Kurtz (1986)**.

The minimal spanning tree on K_n.

Informally,

$$\lim_{n\to\infty} E[L_n] = \lim_{n\to\infty} \int_0^\infty \sum_{k=1}^\infty \frac{E[\zeta_{k,n}(t)]}{n} e^{-t/n} dt - 1$$

$$= \sum_{k=1}^\infty \int_0^\infty \zeta_k(t)dt + \lim_{n\to\infty} \int_{T_{gel}}^\infty \frac{1}{n} e^{-t/n} dt - 1$$

$$= \sum_{k=1}^\infty \int_0^\infty \zeta_k(t)dt + \lim_{n\to\infty} e^{-T_{gel}/n} - 1 = \sum_{k=1}^\infty \int_0^\infty \zeta_k(t)dt.$$

Here $\int_{T_{gel}}^\infty \frac{1}{n} e^{-t/n} dt$ represents the emergence of one giant component at time $T_{gel} = 1$.

We formalize the above argument.

BIG picture.

We prove that

$$\lim_{n\to\infty} E[L_n] = \sum_{k=1}^{\infty} \int_0^{\infty} \zeta_k(t)dt = \sum_{k=1}^{\infty} \int_0^{\infty} \frac{k^{k-2}t^{k-1}}{k!} e^{-kt} dt = \zeta(3).$$

General graphs: Consider a Marcus-Lushnikov processes equivalent to the cluster size dynamics in a general graph, e.g. K_n, $K_{n,n}$, $K_{5n,7n}$, etc. The solutions $\zeta_{\mathbf{k}}(t)$ for the corresponding **reduced** Smoluchowski coagulation equations are considered with $\mathbf{k}$ in a certain index space. Then,

$$\lim_{n\to\infty} E[L_n] = \sum_{\mathbf{k}} \int_0^{\infty} \zeta_{\mathbf{k}}(t)d(t).$$

Erdős-Rényi process on $K_{\alpha n,\beta n}$.

For $\alpha, \beta > 0$, consider two integer valued functions, $\alpha[n] = \alpha n + o(\sqrt{n})$ and $\beta[n] = \beta n + o(\sqrt{n})$.

Consider an Erdős-Rényi random graph process on the bipartite graph $K_{\alpha[n],\beta[n]}$.

In the coalescent process corresponding to an Erdős-Rényi random graph process on $K_{\alpha[n],\beta[n]}$, each cluster is assigned a weight vector $\mathbf{i} = \begin{bmatrix} i_1 \\ i_2 \end{bmatrix}$.

The coalescence kernel for any pair of clusters with weight vectors $\mathbf{i} = \begin{bmatrix} i_1 \\ i_2 \end{bmatrix}$ and $\mathbf{j} = \begin{bmatrix} j_1 \\ j_2 \end{bmatrix}$ is

$$K(\mathbf{i}, \mathbf{j}) := i_1 j_2 + i_2 j_1.$$

Cross-multiplicative coalescent process.

• The process begins with $\alpha[n] + \beta[n]$ singletons, of which $\alpha[n]$ of weight $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\beta[n]$ of weight $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

• The cluster formation is governed by kernel

$$K(\mathbf{i},\mathbf{j}) := i_1 j_2 + i_2 j_1.$$

• A pair of clusters with weight vectors $\mathbf{i}$ and $\mathbf{j}$ would coalesce into a cluster of weight $\mathbf{i}+\mathbf{j}$ with rate $K(\mathbf{i},\mathbf{j})/n$.

• The process continues until there is a single cluster of weight $\begin{bmatrix} \alpha[n] \\ \beta[n] \end{bmatrix}$.

We will call this a cross-multiplicative coalescent process.

Smoluchowski coagulation equations.

$$\begin{aligned}\frac{d}{dt}\zeta_{i_1,i_2}(t) = &- \zeta_{i_1,i_2}(t)\sum_{j_1,j_2}(i_1 j_2 + i_2 j_1)\zeta_{j_1,j_2}(t) \\ &+ \frac{1}{2}\sum_{\substack{\ell_1,k_1:\ \ell_1+k_1=i_1,\\ \ell_2,k_2:\ \ell_2+k_2=i_2}}(\ell_1 k_2 + \ell_2 k_1)\zeta_{\ell_1,\ell_2}(t)\zeta_{k_1,k_2}(t)\end{aligned}$$

with the initial conditions $\zeta_{1,0}(0) = \alpha$ and $\zeta_{0,1}(0) = \beta$.

Gelation: T_{gel} solves

$$1 - (\alpha \wedge \beta)t + \ln((\alpha \vee \beta)t) = 0.$$

Reduced Smoluchowski system.

$$\frac{d}{dt}\zeta_{i_1,i_2}(t) = -(\beta i_1 + \alpha i_2)\zeta_{i_1,i_2}(t) \\ + \frac{1}{2} \sum_{\substack{\ell_1,k_1:\ \ell_1+k_1=i_1, \\ \ell_2,k_2:\ \ell_2+k_2=i_2}} (\ell_1 k_2 + \ell_2 k_1)\zeta_{\ell_1,\ell_2}(t)\zeta_{k_1,k_2}(t)$$

with the initial conditions $\zeta_{1,0}(0) = \alpha$ and $\zeta_{0,1}(0) = \beta$.

Solution:

$$\zeta_{i_1,i_2}(t) = \frac{i_1^{i_2-1} i_2^{i_1-1} \alpha^{i_1} \beta^{i_2}}{i_1! i_2!} e^{-(\beta i_1 + \alpha i_2)t} t^{i_1+i_2-1}$$

Used **F. Huang and B. Liu (2010)** generalization of Abel's binomial theorem.

Theorem [YK, Otto, and Yambartsev, 2017]. Let $\alpha, \beta > 0$ and $L_n = L_n(\alpha, \beta)$ be the length of a minimal spanning tree on a complete bipartite graph $K_{\alpha[n],\beta[n]}$ with partitions of size

$$\alpha[n] = \alpha n + o(\sqrt{n}) \quad \text{and} \quad \beta[n] = \beta n + o(\sqrt{n})$$

and independent uniform edge weights over $[0, 1]$. Then

$$\lim_{n\to\infty} E[L_n] = \sum_{i_1,i_2}^{\infty} \int_0^{\infty} \zeta_{i_1,i_2}(t) d(t),$$

where $\zeta_{i_1,i_2}(t)$ indexed by $\mathbb{Z}_+^2 \setminus \{(0,0)\}$ is the solution of

$$\frac{d}{dt}\zeta_{i_1,i_2}(t) = -(\beta i_1 + \alpha i_2)\zeta_{i_1,i_2}(t) + \frac{1}{2} \sum_{\substack{\ell_1,k_1:\ \ell_1+k_1=i_1, \\ \ell_2,k_2:\ \ell_2+k_2=i_2}} (\ell_1 k_2 + \ell_2 k_1)\zeta_{\ell_1,\ell_2}(t)\zeta_{k_1,k_2}(t)$$

with $\zeta_{i_1,i_2}(0) = \alpha\delta_{1,i_1}\delta_{0,i_2} + \beta\delta_{0,i_1}\delta_{1,i_2}$.

The length of the minimal spanning tree on $K_{\alpha n, \beta n}$.

Theorem [YK, Otto, and Yambartsev, 2017]. Let $\alpha, \beta > 0$, $\gamma = \alpha/\beta$, and $L_n = L_n(\alpha,\beta)$ be the length of a minimal spanning tree on a complete bipartite graph $K_{\alpha[n],\beta[n]}$ with partitions of size

$$\alpha[n] = \alpha n + o(\sqrt{n}) \quad \text{and} \quad \beta[n] = \beta n + o(\sqrt{n})$$

and independent uniform edge weights over $[0,1]$. Then

$$\lim_{n\to\infty} E[L_n] = \gamma + \frac{1}{\gamma} + \sum_{i_1\geq 1;\ i_2 \geq 1} \frac{(i_1+i_2-1)!}{i_1! i_2!} \frac{\gamma^{i_1} i_1^{i_2-1} i_2^{i_1-1}}{(i_1+\gamma i_2)^{i_1+i_2}}.$$

The above theorem recovers the result of **Frieze and McDiarmid (1989)** for $K_{n,n}$:

Corollary. If $\gamma = 1$, then $\lim_{n\to\infty} E[L_n] = 2\zeta(3)$.