Stochastic block model in a new critical regime

Vitalii Konarovskyi

Bielefeld University & Leipzig University

Summer School: Mathematics of Large Networks

joint work with Vlada Limic

UNIVERSITÄT BIELEFELD

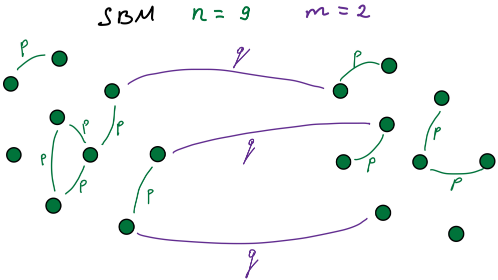
UNIVERSITÄT LEIPZIG

Table of Contents

Stochastic Block Model

Stochastic Block Model $G(n, p, q)$ **is a random graph such that:**

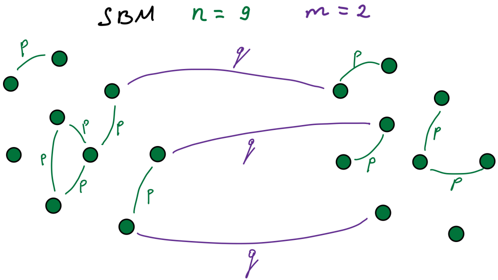
- consists of nm vertices divided into m subsets $(m = 2)$;
- edges are drown independently;
- intra class edges appear with probability $p = p_n$;
- inter class edges appear with probability $q = q_n$.



Stochastic Block Model

Stochastic Block Model $G(n, p, q)$ **is a random graph such that:**

- consists of nm vertices divided into m subsets ($m = 2$);
- edges are drown independently;
- intra class edges appear with probability $p = p_n$;
- inter class edges appear with probability $q = q_n$.



We are interested in the scaling limit as $n \to \infty$ and $p_n, q_n \to 0$.

Largest connected component of SBM

$C_1(n)$ is the size of the largest connected component of the SBM

Largest connected component of SBM

$C_1(n)$ is the size of the largest connected component of the SBM

It is well-known:

- If $p_n = q_n = \frac{a}{mn}$, then SBM is an Erdős-Rényi graph for which:
  - for $a > 1$, $C_1(n) \sim \Theta(n)$;
  - for $a < 1$, $C_1(n) \sim \Theta(\ln n)$; (Erdős, Rényi '60, '61)
  - for $a = 1$, $C_1(n) \sim \Theta(n^{2/3})$.

Largest connected component of SBM

$C_1(n)$ is the size of the largest connected component of the SBM

It is well-known:

- If $p_n = q_n = \frac{a}{mn}$, then SBM is an Erdős-Rényi graph for which:
  - for $a > 1$, $C_1(n) \sim \Theta(n)$;
  - for $a < 1$, $C_1(n) \sim \Theta(\ln n)$; (Erdős, Rényi '60, '61)
  - for $a = 1$, $C_1(n) \sim \Theta(n^{2/3})$.
- If $p_n = \frac{a}{mn}$, $q_n = \frac{b}{mn}$, then
  - $a + (m-1)b > m$, $C_1(n) \sim \Theta(n)$;
  - $a + (m-1)b \leq m$, $C_1(n) \sim o(n)$. (Bollobás, Janson, Riordan '07)

Largest connected component of SBM

$C_1(n)$ is the size of the largest connected component of the SBM

It is well-known:

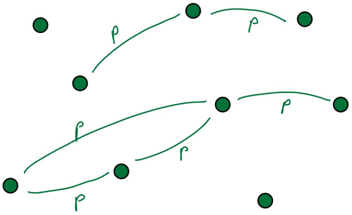
- If $p_n = q_n = \frac{a}{mn}$, then SBM is an Erdős-Rényi graph for which:
  - for $a > 1$, $C_1(n) \sim \Theta(n)$;
  - for $a < 1$, $C_1(n) \sim \Theta(\ln n)$; (Erdős, Rényi '60, '61)
  - for $a = 1$, $C_1(n) \sim \Theta(n^{2/3})$.
- If $p_n = \frac{a}{mn}$, $q_n = \frac{b}{mn}$, then
  - $a + (m-1)b > m$, $C_1(n) \sim \Theta(n)$;
  - $a + (m-1)b \leq m$, $C_1(n) \sim o(n)$. (Bollobás, Janson, Riordan '07)

We are interested in the new critical regime: $q_n \ll p_n \sim \frac{1}{n}$.

Scaling limit of Erdős-Rényi Graphs

$G(n, p)$ – a Erdős-Rényi random graph with n vertices and edges appearing with prob.

$$p = p_n(t) = \frac{1}{n} + \frac{t}{n^{4/3}}, \quad t \in \mathbb{R}$$



Scaling limit of Erdős-Rényi Graphs

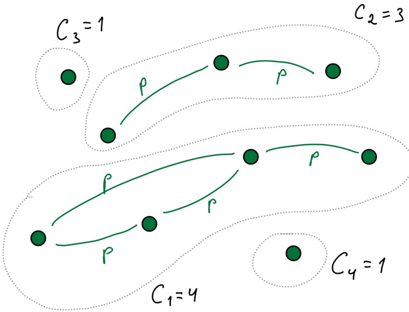
$G(n, p)$ – a Erdős-Rényi random graph with n vertices and edges appearing with prob.

$$p = p_n(t) = \frac{1}{n} + \frac{t}{n^{4/3}}, \quad t \in \mathbb{R}$$

Define

$$X^{(n)}(t) := \frac{1}{n^{2/3}}(C_1, C_2, \dots, C_k, 0, 0, \dots \dots),$$

where $C_k = C_k(n, t)$ is the size of the k-th largest connected component.



Scaling limit of Erdős-Rényi Graphs

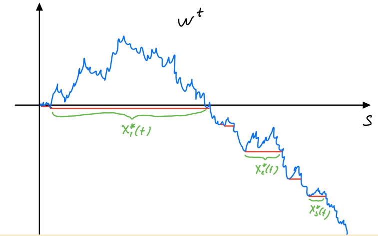
Theorem. (Aldous '97, Anmerdariz '01, Limic '98,'19)

For every $t \in \mathbb{R}$ the sequence $X^{(n)}(t)$ converges in l^2 to $X^*(t)$ in distribution,

where $X^*(t) = (X_i^*(t))_{i \geq 1}$ are the ordered excursion lengths of the Brownian motion with parabolic drift

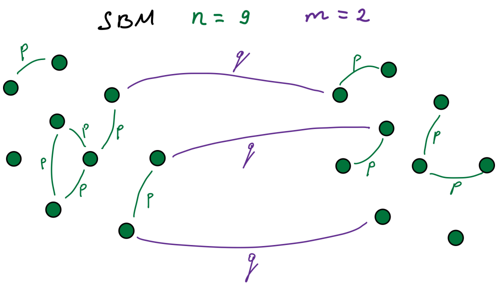
$$W^t(s) := W(s) - \frac{1}{2}s^2 + ts, \quad s \geq 0,$$

above past minima.



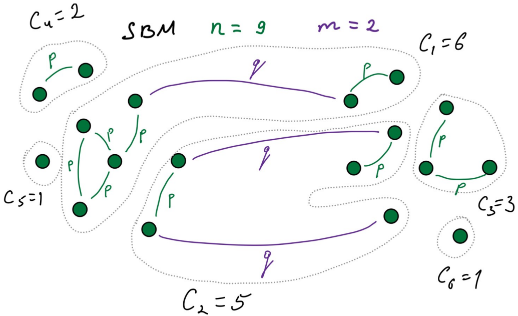
$X^*(t)$, $t \in \mathbb{R}$, is called the **standard Multiplicative coalescent**, and is a Markov process in l^2.

Back to the stochastic block model



$$p = p_n(t) = \frac{1}{n} + \frac{t}{n^{4/3}} \quad q = q_n(s) = \frac{s}{n^{4/3}}, \quad t \in \mathbb{R}, \quad s \geq 0.$$

Back to the stochastic block model



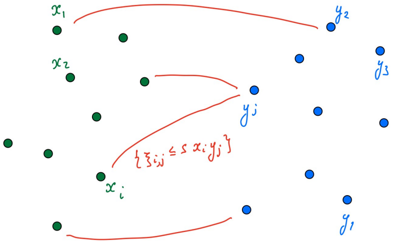
$$p = p_n(t) = \frac{1}{n} + \frac{t}{n^{4/3}} \quad q = q_n(s) = \frac{s}{n^{4/3}}, \quad t \in \mathbb{R}, \quad s \geq 0.$$

Define

$$Z^{(n)}(t,s) := \frac{1}{n^{2/3}}(C_1, C_2, \dots, C_k, 0, 0), \quad t \in \mathbb{R}, \quad s \geq 0,$$

where $C_k = C_k(n,t,s)$ is the size of the k-th largest connected component of the SBM $G(n, p_n, q_n)$

Restricted multiplicative merging



Let $l^2_\downarrow = \{x = (x_i)_{i\ge 1} \in l^2 : \; x_1 \ge x_2 \ge \cdots \ge 0\}$.

For $s \ge 0$ and a fixed family of indep. r.v. $\xi_{i,j} \sim \mathrm{Exp}(\textit{rate } 1)$, $i, j \ge 1$, define a random map $\mathrm{RMM}_s : l^2_\downarrow \times l^2_\downarrow \to l^2_\downarrow$:

- consider coord. of $x, y \in l^2_\downarrow$ as a masses of corresponding vertices of a graph;
- for every $i, j \ge 1$ draw an edge between x_i and y_j iff $\xi_{i,j} \le s x_i y_j$;
- define $\mathrm{RMM}_s(x, y)$ as the vector of the ordered masses of connected components.

The main result

Remind

$$p = p_n(t) = \frac{1}{n} + \frac{t}{n^{4/3}} \quad q = q_n(s) = \frac{s}{n^{4/3}}, \quad t \in \mathbb{R}, \quad s \geq 0.$$

$$Z^{(n)}(t,s) := \frac{1}{n^{2/3}}(C_1, C_2, \dots, C_k, 0, 0), \quad t \in \mathbb{R}, \quad s \geq 0,$$

where $C_k = C_k(n,t,s)$ is the size of the k-th largest connected component of the SBM $G(n, p_n, q_n)$

The main result

Remind

$$p = p_n(t) = \frac{1}{n} + \frac{t}{n^{4/3}} \quad q = q_n(s) = \frac{s}{n^{4/3}}, \quad t \in \mathbb{R}, \quad s \geq 0.$$

$$Z^{(n)}(t,s) := \frac{1}{n^{2/3}}(C_1, C_2, \dots, C_k, 0, 0), \quad t \in \mathbb{R}, \quad s \geq 0,$$

where $C_k = C_k(n,t,s)$ is the size of the k-th largest connected component of the SBM $G(n, p_n, q_n)$

Theorem. (K., Limic '21)

For every $t \in \mathbb{R}$ and $s \geq 0$ the process $Z^{(n)}(t,s)$ converges in l^2 in distribution to $\mathrm{RMM}_s(X^*(t), Y^*(t))$, where X^*, Y^* are independent standard multiplicative coalescents that are independent of ξ

Table of Contents

Different construction of the SBM

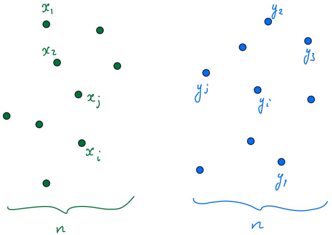
$$p = \frac{1}{n} + \frac{t}{n^{4/3}}, \quad q = \frac{s}{n^{4/3}}$$

$$Z^{(n)}(t,s) := \frac{1}{n^{2/3}}(C_1, C_2, \dots, C_k, 0, 0),$$

Different construction of the SBM

$$p = \frac{1}{n} + \frac{t}{n^{4/3}}, \quad q = \frac{s}{n^{4/3}}$$

$$Z^{(n)}(t,s) := \frac{1}{n^{2/3}}(C_1, C_2, \ldots, C_k, 0, 0),$$

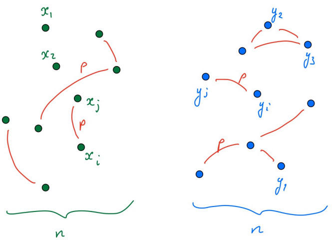


1. Prescribe the mass $x_i = y_i = n^{-\frac{2}{3}}$ to every vertex;

Different construction of the SBM

$$p = \frac{1}{n} + \frac{t}{n^{4/3}}, \quad q = \frac{s}{n^{4/3}}$$

$$Z^{(n)}(t,s) := \frac{1}{n^{2/3}}(C_1, C_2, \ldots, C_k, 0, 0),$$

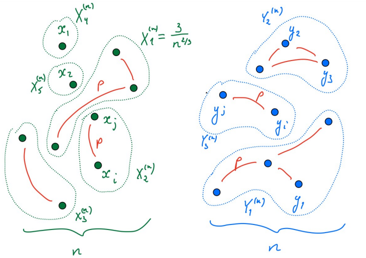


1. Prescribe the mass $x_i = y_i = n^{-\frac{2}{3}}$ to every vertex;
2. Independently add **only intra** class edges between x_i, x_j and between y_i, y_j with probability p

Different construction of the SBM

$$p = \frac{1}{n} + \frac{t}{n^{4/3}}, \quad q = \frac{s}{n^{4/3}}$$

$$Z^{(n)}(t,s) := \frac{1}{n^{2/3}}(C_1, C_2, \ldots, C_k, 0, 0),$$

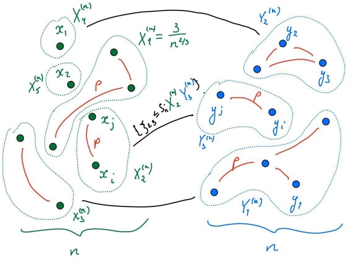


1. Prescribe the mass $x_i = y_i = n^{-\frac{2}{3}}$ to every vertex;
2. Independently add **only intra** class edges between x_i, x_j and between y_i, y_j with probability p
3. Let $X^{(n)}$ and $Y^{(n)}$ be the ordered connected component masses of green and blue graphs, resp.
   Remark: $X^{(n)} \to X^*(t)$ and $Y^{(n)} \to Y^*(t)$, where X^*, Y^* are independent standard multiplicative coalescents.

Different construction of the SBM

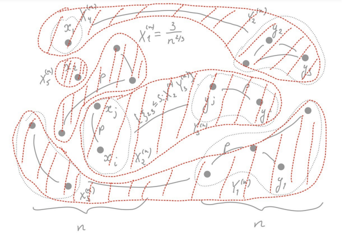
$$p = \frac{1}{n} + \frac{t}{n^{4/3}}, \quad q = \frac{s}{n^{4/3}}$$

$$Z^{(n)}(t,s) := \frac{1}{n^{2/3}}(C_1, C_2, \ldots, C_k, 0, 0),$$



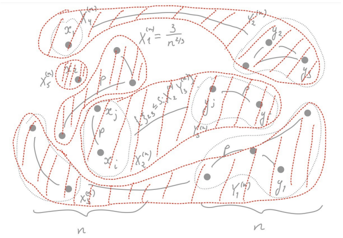
1. Prescribe the mass $x_i = y_i = n^{-\frac{2}{3}}$ to every vertex;
2. Independently add **only intra** class edges between x_i, x_j and between y_i, y_j with probability p
3. Let $X^{(n)}$ and $Y^{(n)}$ be the ordered connected component masses of green and blue graphs, resp.
   Remark: $X^{(n)} \to X^*(t)$ and $Y^{(n)} \to Y^*(t)$, where X^*, Y^* are independent standard multiplicative coalescents.
4. Add **inter** class edges between $X_i^{(n)}, Y_j^{(n)}$ if $\{\xi_{i,j} \leq X_i^{(n)} Y_j^{(n)} s_n\}$, where $q = \mathbb{P}\{\xi_{i,j} \leq x_i y_j s_n\}$ $(s_n \to s,\ n \to \infty)$

Different construction of the SBM

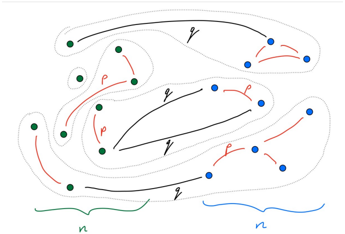


1. By the construction, the ordered connected component masses are $\mathrm{RMM}_{s_n}(X^{(n)}, Y^{(n)})$

Different construction of the SBM



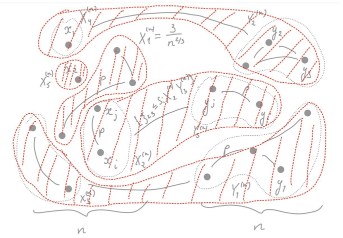
$\stackrel{d}{=}$



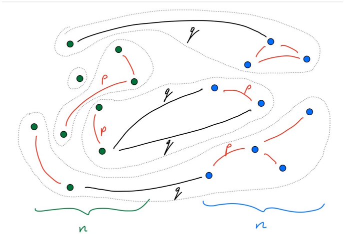
1. By the construction, the ordered connected component masses are $\mathrm{RMM}_{s_n}(X^{(n)}, Y^{(n)})$
2. Using the properties of exponential distributed random variables

$$\mathrm{RMM}_{s_n}(X^{(n)}, Y^{(n)}) \stackrel{d}{=} Z^{(n)}(t, s).$$

Different construction of the SBM



$\overset{d}{=}$



1. By the construction, the ordered connected component masses are $\mathrm{RMM}_{s_n}(X^{(n)}, Y^{(n)})$
2. Using the properties of exponential distributed random variables

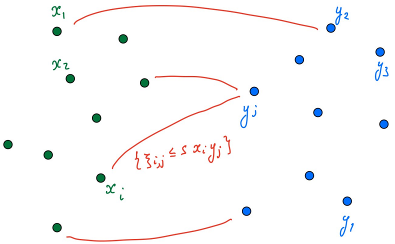
$$\mathrm{RMM}_{s_n}(X^{(n)}, Y^{(n)}) \overset{d}{=} Z^{(n)}(t, s).$$

3. Then having the continuity of RMM, we get

$$\mathrm{RMM}_{s_n}(X^{(n)}, Y^{(n)}) \to \mathrm{RMM}_s(X^*(t), Y^*(t)),$$

because $X^{(n)} \to X^*(t)$, $Y^{(n)} \to Y^*(t)$ and $s_n \to s$.

Continuity of RMM



$\mathrm{RMM}_s(x, y)$ is the vector of the ordered masses of connected components.

Proposition (K., Limic '21)

Let $x^n \to x$, $y^n \to y$ in $l^2_\downarrow$ and $s_n \to s$ in $[0, \infty)$. Then

$$\mathrm{RMM}_{s_n}(x_n, y_n) \to \mathrm{RMM}_s(x, y) \quad \text{in } l^2_\downarrow \text{ in probability}$$

as $n \to \infty$.

References

V. Konarovskyi, V. Limic
Stochastic Block Model in a new critical regime and the Interacting Multiplicative Coalescent.
Electronic Journal of Probability. Vol. 26 (2021), 23 pp.

Thank you!