Excursion Representation of Stochastic Block Model

Vitalii Konarovskyi

Bielefeld University

GPSD 2023 — Essen

joint work with David Clancy and Vlada Limic

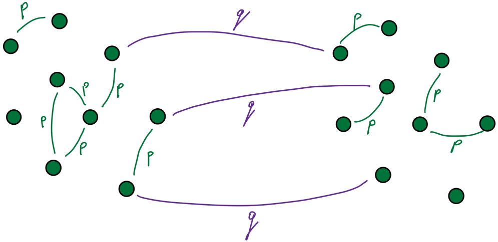
Table of Contents

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 $q = q_n$.

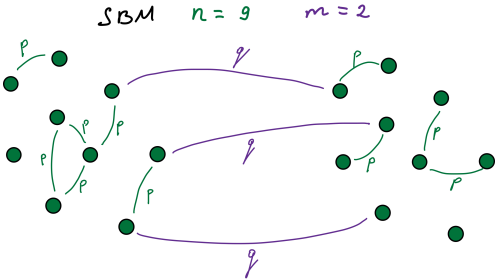
SBM $n = 9$ $m = 2$



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:

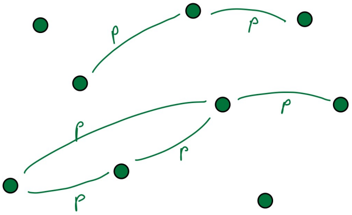
- If $p_n = q_n = \frac{a}{mn}$, then SBM is an Erdős-Rényi graph for which:
  - for $a > 1$, $C_1(n) \sim \Theta(n)$;
  - for $a < 1$, $C_1(n) \sim \Theta(\ln n)$; (Erdős, Rényi '60, '61)
  - for $a = 1$, $C_1(n) \sim \Theta(n^{2/3})$.
- If $p_n = \frac{a}{mn}$, $q_n = \frac{b}{mn}$, then
  - $a + (m-1)b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

$G(n,p)$ – an 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

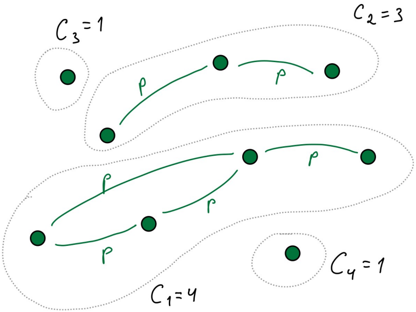
$G(n, p)$ – an Erdős-Rényi random graph with n vertices and edges appearing with prob.

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

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)**

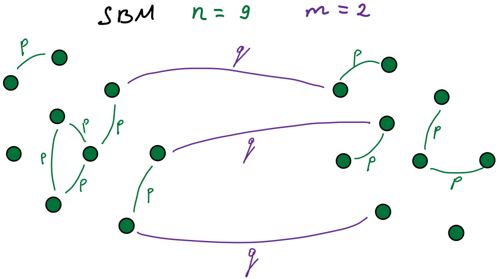
For every $t \in \mathbb{R}$ the sequence $X^{(n)}(t)$ converges in l^2 in distribution to a **standard Multiplicative Coalescent** (MC) $X^*(t)$, where

$X^*(t)$ is the ordered excursion lengths of the Brownian motion with parabolic drift

$$B^t(r) := B(r) - \frac{1}{2}r^2 + tr, \quad r \geq 0,$$

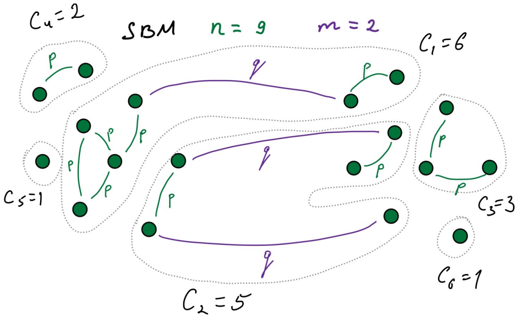
above past minima.

Stochastic Block Model



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

Stochastic Block Model



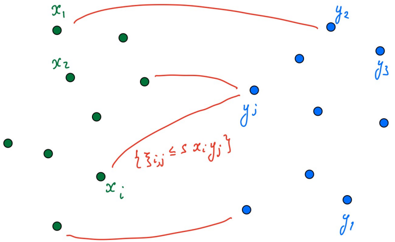
$$p = p_n(t) = \frac{1}{n} + \frac{t}{n^{4/3}} \quad q = q_n(s) = \frac{s}{n^{4/3}}, \quad t \in \mathbb{R}, \quad s \geq 0.$$

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

For $s \geq 0$ and a fixed family of indep. r.v. $\xi_{i,j} \sim \mathrm{Exp}(\mathit{rate}\ 1)$, $i, j \geq 1$, 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 $i, j \geq 1$ draw an edge between x_i and y_j iff $\xi_{i,j} \leq s x_i y_j$;
- define $\mathrm{RMM}_s(x, y)$ as the vector of the ordered masses of connected components.

Scaling Limit of SBM

Recall

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

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

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.

Scaling Limit of SBM

Recall

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

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

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.

We will call $\mathrm{RMM}_s(X^*(t), Y^*(t))$ an **Interacting Multiplicative Coalescent**

Table of Contents

Main Problem

Question: Does the scaling limit of the SBM admit an excursion description?

Main Problem

Question: Does the scaling limit of the SBM admit an excursion description?

Naive Guess: Probably, interacting MC can be described via “excursions” of a random field or a family of Brownian motions with interactions.

Standard MC as Jumps of Hitting Times

$X^*(t)$ is the ordered excursion lengths of the Brownian motion with parabolic drift

$$B^t(r) := B(r) - \frac{1}{2}r^2 + tr, \quad r \geq 0,$$

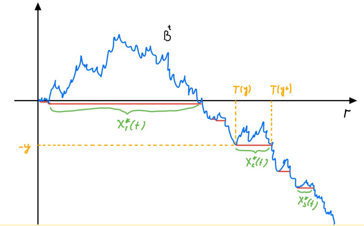
above past minima.

Standard MC as Jumps of Hitting Times

$X^*(t)$ is the ordered excursion lengths of the Brownian motion with parabolic drift

$$B^t(r) := B(r) - \frac{1}{2}r^2 + tr, \quad r \geq 0,$$

above past minima.



Define for $y \geq 0$

$$T(y) := \min\{r : \ B^t(r) = -y\} = \min\{r : \ \underline{B}^t(r) = -y\}$$

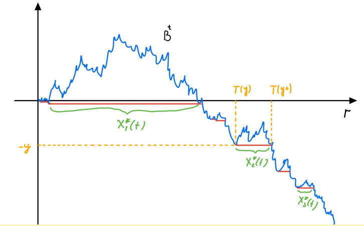
where $\underline{B}^t(r) = \min_{[0,r]} B^t$.

Standard MC as Jumps of Hitting Times

$X^*(t)$ is the ordered excursion lengths of the Brownian motion with parabolic drift

$$B^t(r) := B(r) - \frac{1}{2}r^2 + tr, \quad r \geq 0,$$

above past minima.



Define for $y \geq 0$

$$T(y) := \min\{r : \ B^t(r) = -y\} = \min\{r : \ \underline{B}^t(r) = -y\}$$

where $\underline{B}^t(r) = \min_{[0,r]} B^t$.

Observation

$X^*(t)$ is the collection of decreasingly ordered jumps $T(y+) - T(y)$, $y \geq 0$.

Hitting times for fields

Consider $\vec{X} : [0, \infty)^m \to \mathbb{R}^m$ defined by

$$X_i(r_1, \ldots, r_m) = X_{i,i}(r_i) + \sum_{j \neq i} X_{i,j}(r_j),$$

such that

1. $X_{i,i}$ are continuous
2. $X_{i,j}$, $i \neq j$, are non-decreasing and continuous

Hitting times for fields

Consider $\vec{X} : [0, \infty)^m \to \mathbb{R}^m$ defined by

$$X_i(r_1, \dots, r_m) = X_{i,i}(r_i) + \sum_{j \neq i} X_{i,j}(r_j),$$

such that

1. $X_{i,i}$ are continuous
2. $X_{i,j}$, $i \neq j$, are non-decreasing and continuous

Lemma (Chaumont, Marolleau '20)

For every $\vec{y} \in [0, \infty)^m$ there exists a (component-wise) minimal solution $\vec{T} = \vec{T}(y) \in [0, \infty]^m$ to the equation

$$X_i(\vec{T}) = -y_i, \quad \forall i \text{ such that } T_i < \infty,$$

denoted by

$$\vec{T}(y) := \min\left\{\vec{r} : \ \vec{X}(\vec{r}) = -\vec{y}\right\}.$$

Scaling limit of SBM as Jumps of Hitting Times

For fixed $t \in \mathbb{R}$ and $s \geq 0$ define

$$X_{i,i}(r) := B_i^t(r) = B_i(r) - \frac{1}{2}r^2 + tr, \quad r \geq 0$$

$$X_{i,j}(r) = sr, \quad i \neq j, \quad r \geq 0.$$

Scaling limit of SBM as Jumps of Hitting Times

For fixed $t \in \mathbb{R}$ and $s \geq 0$ define

$$X_{i,i}(r) := B_i^t(r) = B_i(r) - \frac{1}{2}r^2 + tr, \quad r \geq 0$$

$$X_{i,j}(r) = sr, \quad i \neq j, \quad r \geq 0.$$

Let $m = 2$ and

$$X_1(r_1, r_2) = X_{1,1}(r_1) + X_{1,2}(r_2) = B_1^t(r_1) + sr_2,$$

$$X_2(r_1, r_2) = X_{2,1}(r_1) + X_{2,2}(r_2) = sr_1 + B_2^t(r_2),$$

Scaling limit of SBM as Jumps of Hitting Times

For fixed $t \in \mathbb{R}$ and $s \geq 0$ define

$$X_{i,i}(r) := B_i^t(r) = B_i(r) - \frac{1}{2}r^2 + tr, \quad r \geq 0$$

$$X_{i,j}(r) = sr, \quad i \neq j, \quad r \geq 0.$$

Let $m = 2$ and

$$X_1(r_1, r_2) = X_{1,1}(r_1) + X_{1,2}(r_2) = B_1^t(r_1) + sr_2,$$

$$X_2(r_1, r_2) = X_{2,1}(r_1) + X_{2,2}(r_2) = sr_1 + B_2^t(r_2),$$

Set for $y \geq 0$

$$\vec{T}(y) := \min\{(r_1, r_2) : \ X_1(r_1, r_2) = -y, \ \ X_2(r_1, r_2) = -y\}$$

Scaling limit of SBM as Jumps of Hitting Times

For fixed $t \in \mathbb{R}$ and $s \geq 0$ define

$$X_{i,i}(r) := B_i^t(r) = B_i(r) - \frac{1}{2}r^2 + tr, \quad r \geq 0$$

$$X_{i,j}(r) = sr, \quad i \neq j, \quad r \geq 0.$$

Let $m = 2$ and

$$X_1(r_1, r_2) = X_{1,1}(r_1) + X_{1,2}(r_2) = B_1^t(r_1) + sr_2,$$

$$X_2(r_1, r_2) = X_{2,1}(r_1) + X_{2,2}(r_2) = sr_1 + B_2^t(r_2),$$

Set for $y \geq 0$

$$\vec{T}(y) := \min\{(r_1, r_2) : \ X_1(r_1, r_2) = -y, \quad X_2(r_1, r_2) = -y\}$$

Theorem. (Clancy, K., Limic '23)

For every $t \in \mathbb{R}$ and $s \geq 0$, the distribution of $\mathrm{RMM}_s(X^*(t), Y^*(t))$ coincides with the law of decreasingly ordered sequence of norms of jumps $\|\vec{T}(y+) - \vec{T}(y)\|_1$, where $\|\vec{r}\|_1 = r_1 + r_2$, $r_i \geq 0$.

Construction of a continuous Curve

Note that

$$\vec{T}(y) = \min\{(r_1,r_2):\ X_1(r_1,r_2) = -y,\ \ X_2(r_1,r_2) = -y\}$$
$$= \min\{(r_1,r_2):\ \underline{X}_1(r_1,r_2) = -y,\ \ \underline{X}_2(r_1,r_2) = -y\},$$

where $\underline{X}_1(r_1,r_2) = \underline{B}_1^t(r_1) + sr_2$, $\underline{X}_2(r_1,r_2) = sr_1 + \underline{B}_2^t(r_2)$, $\underline{B}_i^t(r) = \min_{[0,r]} B_i^t$.

Construction of a continuous Curve

Note that

$$\vec{T}(y) = \min\{(r_1, r_2): \ X_1(r_1, r_2) = -y, \ \ X_2(r_1, r_2) = -y\}$$
$$= \min\{(r_1, r_2): \ \underline{X}_1(r_1, r_2) = -y, \ \ \underline{X}_2(r_1, r_2) = -y\},$$

where $\underline{X}_1(r_1, r_2) = \underline{B}_1^t(r_1) + sr_2, \quad \underline{X}_2(r_1, r_2) = sr_1 + \underline{B}_2^t(r_2), \quad \underline{B}_i^t(r) = \min_{[0,r]} B_i^t.$

We define a curve $\gamma : [0, \infty) \to [0, \infty)^2$ by

$$\underline{X}_1(\gamma(u)) = \underline{X}_2(\gamma(u)), \quad \|\gamma(u)\|_1 = u$$

Construction of a continuous Curve

Note that

$$\vec{T}(y) = \min\{(r_1,r_2): \; X_1(r_1,r_2) = -y, \quad X_2(r_1,r_2) = -y\}$$
$$= \min\{(r_1,r_2): \; \underline{X}_1(r_1,r_2) = -y, \quad \underline{X}_2(r_1,r_2) = -y\},$$

where $\underline{X}_1(r_1,r_2) = \underline{B}_1^t(r_1) + sr_2, \quad \underline{X}_2(r_1,r_2) = sr_1 + \underline{B}_2^t(r_2), \quad \underline{B}_i^t(r) = \min_{[0,r]} B_i^t.$

We define a curve $\gamma : [0,\infty) \to [0,\infty)^2$ by

$$\underline{X}_1(\gamma(u)) = \underline{X}_2(\gamma(u)), \quad \|\gamma(u)\|_1 = u$$

Lemma

Take $g_i(r) := sr - \underline{B}_i^t(r)$, $\kappa := \left(g_1^{-1} + g_2^{-1}\right)^{-1}$. Then γ is uniquely determined by $\gamma_i(u) = g_i^{-1} \circ \kappa(u)$. Moreover, for every $y \geq 0$ the hitting time

$$S(y) := \inf\{u : \; X_i \circ \gamma(u) = -y\}, \quad i = 1,2$$

satisfies $\vec{T} = \gamma \circ S$, $\|\vec{T}\|_1 = \|\gamma \circ S\|_1 = S$

Construction of a continuous Curve

Note that

$$\vec{T}(y) = \min\{(r_1, r_2) : \ X_1(r_1, r_2) = -y, \ \ X_2(r_1, r_2) = -y\}$$
$$= \min\{(r_1, r_2) : \ \underline{X}_1(r_1, r_2) = -y, \ \ \underline{X}_2(r_1, r_2) = -y\},$$

where $\underline{X}_1(r_1, r_2) = \underline{B}_1^t(r_1) + sr_2, \quad \underline{X}_2(r_1, r_2) = sr_1 + \underline{B}_2^t(r_2), \quad \underline{B}_i^t(r) = \min_{[0,r]} B_i^t.$

We define a curve $\gamma : [0, \infty) \to [0, \infty)^2$ by

$$\underline{X}_1(\gamma(u)) = \underline{X}_2(\gamma(u)), \quad \|\gamma(u)\|_1 = u$$

Lemma

Take $g_i(r) := sr - \underline{B}_i^t(r)$, $\kappa := \left(g_1^{-1} + g_2^{-1}\right)^{-1}$. Then γ is uniquely determined by $\gamma_i(u) = g_i^{-1} \circ \kappa(u)$. Moreover, for every $y \geq 0$ the hitting time

$$S(y) := \inf\{u : \ X_i \circ \gamma(u) = -y\}, \quad i = 1, 2$$

satisfies $\vec{T} = \gamma \circ S$, $\|\vec{T}\|_1 = \|\gamma \circ S\|_1 = S$

$$\Longrightarrow \quad \|\vec{T}(y+) - \vec{T}(y)\|_1 = S(y+) - S(y).$$

Excursion Description of SBM along the Curve

Theorem. **(Clancy, K., Limic '23)**

Let

$$g_i(r) = sr - \underline{B}_i^t(r) = sr - \min_{[0,r]} B_i^t,$$
$$\kappa = \left(g_1^{-1} + g_2^{-1}\right)^{-1}$$
$$\gamma_i = g_i^{-1} \circ \kappa.$$

Then for every $t \in \mathbb{R}$ and $s \geq 0$, the distribution of the scaling limit of the stochastic block model $\mathrm{RMM}_s(X^*(t), Y^*(t))$ coincides with the law of ordered excursion lengths of $X_i \circ \vec{\gamma}$ above past minima, where

$$X_1(\vec{r}) = B_1^t(r_1) + sr_2, \quad r_i \geq 0$$

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.

D. Clancy, V. Konarovskyi, V. Limic
Excursion representation of Stochastic Block Model.
In preparation (2023)

Thank you!