The weighted dependent random connection model

Community detection in a spatial preferential attachment model

[GeoSto26] — [Rouen], [24/09/2026]

The model: vertices, ages, connections

Vertices. In dimension $d \geq 2$: let η be a Poisson point process with intensity $\lambda \, \mathrm{d}x \, \mathrm{d}u$ on $\mathbb{R}^d \times (0,1]$. A point $X = (x,u) \in \eta$ carries

- a location $x \in \mathbb{R}^d$;
- an age (or weight) $u \in (0,1]$: *u small = old = attractive*.

Preferential attachment kernel.

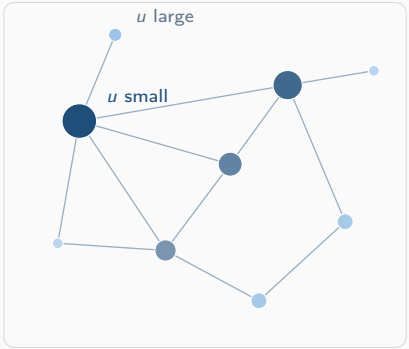
$$g(u,v) = (u \wedge v)^{\gamma} \, (u \vee v)^{1-\gamma}, \qquad \gamma \in (0,1).$$

Effective distance between $X = (x,u)$ and $Y = (y,v)$:

$$\kappa_{XY} := g(u,v) \, |x-y|^d.$$

Edges. Given η, drawn *independently* through a non-increasing profile $\varphi : [0,\infty) \to [0,1]$:

$$\mathbb{P}\big(X \sim Y \mid \eta\big) = \varphi(\kappa_{XY}).$$



Size $\propto$ seniority: old vertices attract long edges.

Analogy: a Boolean model with pair-dependent radii

Take the hard profile $\varphi = 1_{[0,1]}$ and attach to each $X = (x, u)$ the radius $R_X := (1/u)^{1/d}$ (large for old vertices). Then

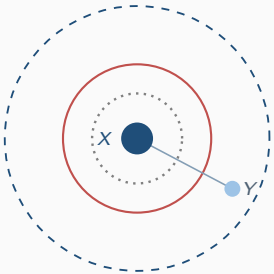
$$\kappa_{XY} \le 1 \iff |x - y| \le (R_X \vee R_Y)^{\gamma} (R_X \wedge R_Y)^{1-\gamma}.$$

Reading

A Boolean-type model whose connection range is the weighted geometric mean of the two radii.

- $\gamma = 1$ range $= R_X \vee R_Y$: *Boolean model*
- $\gamma = \frac{1}{2}$ range $= \sqrt{R_X R_Y}$: product kernel
- $\gamma \to 0$ range $= R_X \wedge R_Y$: most restrictive

Two differences with the Boolean model: the range is not carried by a single vertex.



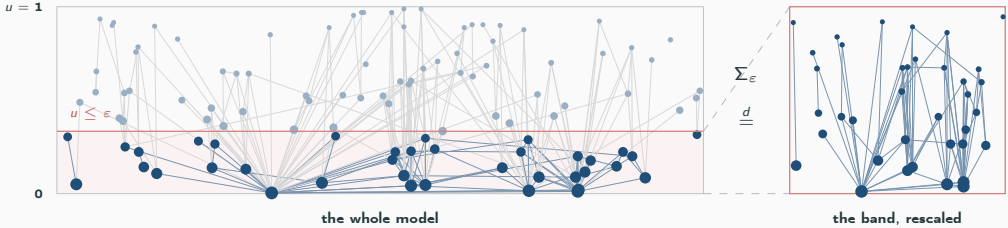
X old, Y young: linked if $\gamma = 1$, not if $\gamma = \frac{1}{2}$.

Self-similarity: the model reproduces itself

Lemma (scale invariance in age)

For every $\varepsilon \in (0, 1]$, the rescaling $\Sigma_\varepsilon : (x, u) \mapsto (\varepsilon^{1/d} x,\ u/\varepsilon)$ maps the induced subgraph on the ages $u \le \varepsilon$ onto an exact copy of the whole model: same law, same parameters $(\lambda, \gamma, \varphi)$.

The intensity is unchanged, $\lambda \cdot \varepsilon \cdot \varepsilon^{-1} = \lambda$, and κ_{XY} is invariant: $g(u/\varepsilon, v/\varepsilon) = \varepsilon^{-1} g(u, v)$ while $|\varepsilon^{1/d}(x - y)|^d = \varepsilon |x - y|^d$.



Same density, same fans, same hubs, the space is contracted by $\varepsilon^{1/d}$, so the copy lives in a smaller window. Grey edges leave the band and are not part of it.

Sparse regime and power-law degrees

Connection probabilities do not depend on the network size n. Keeping the mean degree finite as $n \to \infty$ forces the integrability condition

$$\int_{\mathbb{R}^d} \int_0^1 \int_0^1 \varphi\big(g(u,v)\,|x|^d\big)\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}x < \infty.$$

Mean degree ($\varphi = \rho$, $\rho(r) = 1 \wedge r^{-\delta}$, $\delta > 1$)

$$D(u) = \frac{\lambda \omega_d \delta}{\delta - 1}\left[\frac{1}{1-\gamma} + \frac{u^{-\gamma} - 1}{\gamma}\right],$$

$$\mathbb{E}[D] = \frac{2\lambda\omega_d\delta}{(\delta-1)(1-\gamma)}.$$

Finite *iff* $\delta > 1$ and $\gamma < 1$; proportional to $\lambda/(1-\gamma)$.

Degree tail

$D(u) \sim C\,u^{-\gamma}$ as $u \to 0$, hence

$$\mathbb{P}(D > k) \asymp k^{-1/\gamma}$$

Infinite variance iff $\gamma > \frac{1}{2}$, with a finite mean degree.

Where the preferential attachment hides

Read the age as a birth time: at time t the graph is $\{u \le t\}$, and $t = 1$ is the present.

Degree at time t of a vertex born at s

$$D_t(s) = \frac{C}{\gamma}\left[\left(\frac{t}{s}\right)^{\gamma} - 1\right] \asymp \left(\frac{t}{s}\right)^{\gamma}, \qquad C = \frac{\lambda \omega_d \delta}{\delta - 1}.$$

The equivalent form of the profile therefore reads

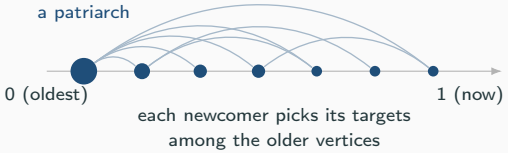
$$\mathbb{P}\big((x,t) \sim (y,s)\big) \;\approx\; \varphi\left(\frac{t\,\|x-y\|^d}{D_t(s)}\right) :$$

a newcomer attaches with a probability increasing in the current degree of its target, damped by distance so the rich get richer, locally.

Out-degree vs in-degree

$$D(u) = \underbrace{\frac{C}{1-\gamma}}_{\text{edges made at birth}} + \underbrace{\frac{C\,(u^{-\gamma}-1)}{\gamma}}_{\text{edges received later}}$$

The first term is $O(1)$: every vertex is born with a bounded number of edges. All the heavy tail sits in the second one, $\asymp u^{-\gamma}$: it is accumulated, not chosen.

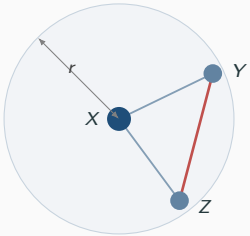


Many triangles: never locally tree-like

Geometry closes the triangles. With the hard profile and three vertices of comparable age u, the range is $r = (1/u)^{1/d}$ and

$$X \sim Y,\ X \sim Z \implies |y - z| \leq 2r,$$

while $Y \sim Z$ as soon as $|y - z| \leq r$.
Hence the local weak limit is not a tree it is the opposite of Erdős–Rényi, of the configuration model and of the usual SBM.



Two neighbours of X are close to each other, hence likely neighbours.

What it breaks. Belief propagation, non-backtracking spectra, cavity computations, all assume local tree-likeness, none of them transfers. Short paths share edges, so path counts are not sums of independent terms.

What it gives. Those same triangles carry the homophily when local signal tree-like models lack it.

Why this model?

Real social networks are all of these at once.

	ER / SBM	configuration	RGG	**WDRCM**
sparse	✓	✓	✓	✓
heavy-tailed degrees	×	✓ **imposed**	×	✓ **emergent**
geometry: homophily by distance	×	×	✓	✓
many triangles, clustering	×	×	✓	✓
growth by preferential attachment	×	×	×	✓

Why it matters here

The standard community-detection toolbox was built on the first column. The WDRCM is where every one of those assumptions fails at once, and where we can ask what survives.

The problem of community detection

Take a "social network". You observe the vertices and the connections, nothing else. *Are there communities? How many? Who is in which?*

Here: two communities. Given η, attach i.i.d. uniform labels $\sigma_X \in \{\pm 1\}$, and draw the edges independently with

$$\mathbb{P}\big(X \sim Y \mid \eta, \sigma\big) = \begin{cases} \varphi_{\mathrm{in}}\big(\kappa_{XY}\big) & \sigma_X = \sigma_Y, \\ \varphi_{\mathrm{out}}\big(\kappa_{XY}\big) & \sigma_X \neq \sigma_Y, \end{cases}$$

both non-increasing, with $\varphi_{\mathrm{in}} \geq \varphi_{\mathrm{out}}$ (homophily); the signal is $\psi := \varphi_{\mathrm{in}} - \varphi_{\mathrm{out}}$ and $\bar{p} := \frac{1}{2}(\varphi_{\mathrm{in}} + \varphi_{\mathrm{out}})$.

Neighbours are a thinning of η: the same- and cross-community counts are independent Poisson,

$$D^{\pm}(u) \sim \mathrm{Poi}\Big(\tfrac{\lambda}{2} \iint \varphi_{\mathrm{in/out}}\Big), \qquad \mathbb{E}D^{+} > \mathbb{E}D^{-}.$$

That gap is all we have.

what you observe

what you want to recover

Both communities are everywhere: space says nothing, only the edges do.

Overlap and recovery

Observe the model in a window W_n of volume n: $\eta_n := \eta \cap (W_n \times (0,1])$, $N_n := |\eta_n|$, E_n the edges inside. An estimator $\hat{\sigma}^{(n)}$ is a deterministic function of (η_n, E_n) and the labels are *not* observed.

Empirical overlap

$$\mathrm{ov}_n(\sigma, \hat{\sigma}) = \frac{1}{N_n} \Big| \sum_{x \in \eta_n} \sigma_x \hat{\sigma}_x \Big| \ \in [0,1].$$

The absolute value absorbs the global sign flip. Close to 1: almost perfect reconstruction. Tending to 0: no better than a coin toss.

Weak recovery

Solvable if there are estimators $\hat{\sigma}^{(n)}$ and a constant $\alpha > 0$ with $\lim_{n\to\infty} \mathbb{P}[\mathrm{ov}_n \geq \alpha] = 1$. It is like beating the random guesser.

Strong recovery

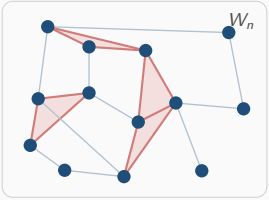
Solvable if for *every* $\varepsilon > 0$ there are estimators with $\lim_{n\to\infty} \mathbb{P}[\mathrm{ov}_n \geq 1 - \varepsilon] = 1$, i.e. $\mathrm{ov}_n \to 1$ in probability. For a family indexed by λ: reached in the high-intensity limit if for every $\varepsilon > 0$ there is $\lambda(\varepsilon) < \infty$ such that this holds for all $\lambda \geq \lambda(\varepsilon)$.

Distinguishability: always, and with no threshold

Distinguishable if, for *every* $h : \mathbb{R}^+ \to [0,1]$, one can tell with probability $\geq \frac{1}{2} + \alpha - o_n(1)$, some $\alpha > 0$, whether (η_n, E_n) comes from the planted partition or from the community-free model $H_{\lambda,h,d}$ (uniform prior on the two).

Theorem (mutual singularity)

For every $\lambda > 0$, every d, every $h : \mathbb{R}_+ \to [0,1]$ and all profiles with $1 \geq \varphi_{\text{in}} \geq \varphi_{\text{out}} \geq 0$ and $|\{\varphi_{\text{in}} \neq \varphi_{\text{out}}\}|_{\text{Leb}} > 0$, the laws of the one-community model G^1 and of the two-community model G^2 on $\mathbb{M}_{\mathcal{G}}(\mathbb{R}^d)$ are mutually singular.



triangle density in $W_n \xrightarrow[n\to\infty]{\text{a.s.}}$

a deterministic constant (ergodicity)

Edges can be faked: $h = \bar{p}$ matches every pair statistic.
Triangles statistic differ, at effective distances r_{12}, r_{13}, r_{23},

$$T^{(2)} - T^{(1)} = \tfrac{1}{8}\,\psi(r_{12})\,\psi(r_{13})\,\psi(r_{23}) \;>\; 0.$$

Aside: percolation in the community-free model

Long-range assumption. To isolate the effect of long edges, take the profile ρ regularly varying of index $-\delta$, $\delta > 1$:

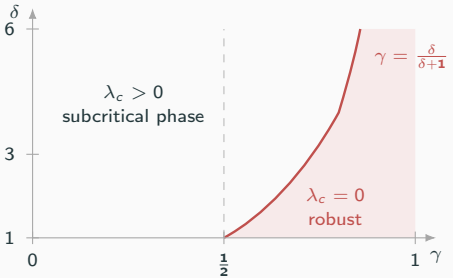
$$\lim_{r\to\infty} \frac{\rho(cr)}{\rho(r)} = c^{-\delta} \qquad \text{for all } c \geq 1.$$

Lighter tails and bounded support included are recovered from the limit $\delta \uparrow \infty$ by a comparison argument.

Theorem (Gracar–Lüchtrath–Mörters)

For the weight-dependent RCM with the preferential attachment kernel (also: sum kernel, min kernel), $\gamma \in (0,1)$, set $\gamma_c(\delta) := \frac{\delta}{\delta+1}$. Then

- $\gamma < \gamma_c$: $\lambda_c > 0$ a genuine subcritical phase, no infinite component for small λ;
- $\gamma > \gamma_c$: $\lambda_c = 0$ an infinite component for every $\lambda > 0$, however small.



Since $\delta > 1$, always $\gamma_c > \frac{1}{2}$: robustness requires infinite variance is not enough.

Impossibility: when the signal graph does not percolate

The signal graph

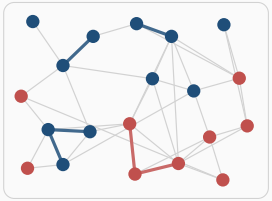
Decompose each pair independently, using $\varphi_{\text{in}} = \varphi_{\text{out}} + \psi$: with probability $\psi(\kappa_{XY})$ a signal edge, present *iff* $\sigma_X = \sigma_Y$; otherwise, with probability $\varphi_{\text{out}}(\kappa_{XY})$, a noise edge, present whatever the labels. G_{sig} is the community-free RCM with profile ψ: its components are monochromatic, and they are the only carriers of label information.

Proposition (L^1 impossibility)

Under suitable regularity, if

$$\gamma < \frac{\delta}{\delta + 1} \qquad \text{and} \qquad \mu_{\text{sig}} := \lambda \iiint \psi \;\leq\; 1,$$

then the optimal overlap satisfies $\alpha_n^* \to 0$: weak recovery is impossible, for any algorithm.



coloured: G_{sig} grey: noise edges

$\mu_{\text{sig}} \leq 1$: by a branching comparison G_{sig} is subcritical with finite monochromatic islands in label-free noise.

$\gamma < \frac{\delta}{\delta+1}$: the non-robust regime of the previous slide, which is what makes this phase non-empty.

Borrowing the Euclidean result

In the ERCM, the connection probability depends on the *distance only* ie $\varphi_{\text{in}}(\|x-y\|)$ or $\varphi_{\text{out}}(\|x-y\|)$, ages ignored.

Abbe, Baccelli, Sankararaman

Their GBG algorithm solves weak recovery on the ERCM, and its maximal overlap satisfies

$$\lim_{\lambda\to\infty} O_\lambda = 1.$$

On a thin slice, the WDRCM *is* an ERCM

Restrict to the ages $[t_0-\epsilon_0, t_0+\epsilon_0]$. If

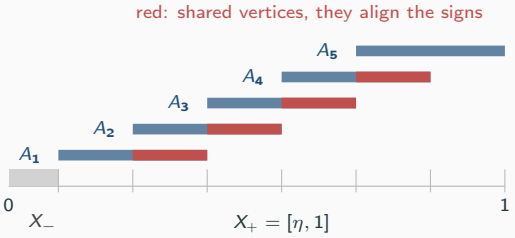
$$\epsilon_0 \;\le\; t_0\,\frac{1-x^a}{1+x^a}, \qquad a=\frac{1}{1+2\gamma}, \quad x=\frac{\tilde{r}}{R},$$

the age no longer distorts the geometry: the slice behaves like an ERCM of effective radius $\tilde{R}$, and $\lim_{\lambda\to\infty} O_{\lambda,t_0,\epsilon_0} = 1$.

$x=\tilde{r}/R$, the geometric contrast: small x, thick slice; $x\to 1$, thin slice.

From one slice to the whole model

- $X_- = [0, \eta]$: set aside, default $+1$ (η arbitrarily small)
- X_+: overlapping slices $A_1, \ldots, A_m$
- shared vertices fix the global sign
- overlap $\geq 2\alpha - 1$, independant of m



Theorem

Let $d \geq 2$ and $|\{\varphi_{\text{in}} \neq \varphi_{\text{out}}\}|_{\text{Leb}} > 0$. There is a finite $\lambda_{\text{upper}}(\varphi_{\text{in}}, \varphi_{\text{out}}, d)$ such that for $\lambda \geq \lambda_{\text{upper}}$, GBG with propagation solves weak recovery on all of $[0, 1]$, and $\tilde{O}_\lambda^{[0,1]} \to 1$.

The robust regime $\gamma > \frac{\delta}{\delta+1}$: buy density with age

ρ regularly varying of index $-\delta$, as before. Here λ is fixed, possibly tiny and we cannot raise it. So we lower the age instead.

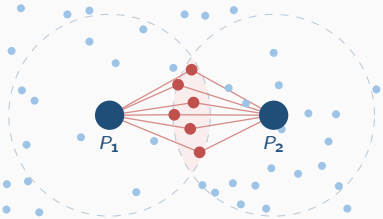
Elect Patriarchs of age $\asymp \epsilon$. Their reach grows like $\epsilon^{-\gamma/d}$, their spacing only like $(\lambda\epsilon)^{-1/d}$.

Test a pair by counting their common young relays:

$$N(\epsilon) \asymp (\lambda)^{1+\delta}\, \epsilon^{\delta-\gamma(1+\delta)}.$$

$\gamma > \frac{\delta}{\delta+1} \iff \delta - \gamma(1+\delta) < 0 \iff N(\epsilon) \to \infty$ as $\epsilon \to 0$.

Artificial ageing is a density amplifier: the local signal-to-noise ratio is made arbitrarily large *without* touching λ.



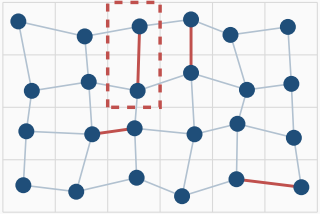
Two Patriarchs, their reach, and the young relays they share

Renormalise, then decode

- one Patriarch per box: a field on $\mathbb{Z}^d$
- each bond: a noisy reading of $\sigma_i\sigma_j$, error as small as we like, noise m_0-dependent
- decode by minimum weight, or locally from the observable syndrome

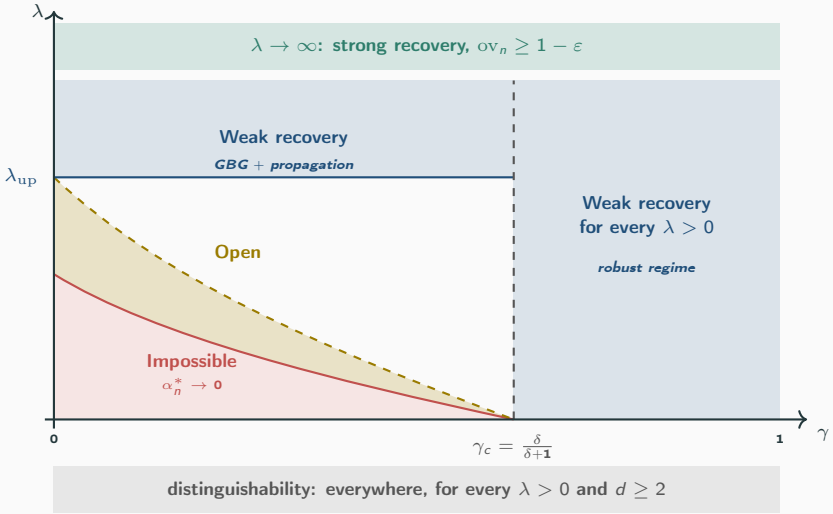
Theorem

If $\gamma > \frac{\delta}{\delta+1}$, weak recovery is solvable for every $\lambda > 0$.



One Patriarch per box; red bonds are misread, the dashed loop is the cluster they flag

Where we stand



Phase diagram in the plane (γ, λ), at fixed $\delta > 1$ and $d \geq 2$.