# Aligning sparse random graphs

Guilhem Semerjian

LPENS

11.08.2023 / Cargese

Giovanni Piccioli, GS, Gabriele Sicuro, Lenka Zdeborová, 2112.13079

Luca Ganassali, Laurent Massoulié, GS, 2209.13723

Andrea Muratori, GS, in preparation

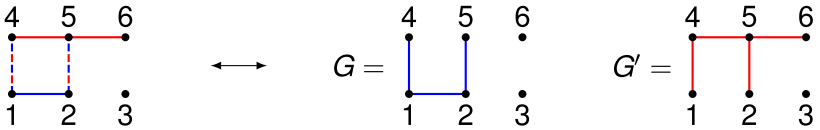
# Outline

# Outline

# Correlated Erdős-Rényi random graphs

$G = (V, E)$, $G' = (V, E')$ $\qquad$ $(G, G') \sim \mathrm{ER}(n, p, s)$ :

- $V = [n] = \{1, \dots, n\}$
- for each pair $i < j$, independently:
  - $\{i,j\} \in E$, $\{i,j\} \in E'$ with probability $ps$
  - $\{i,j\} \in E$, $\{i,j\} \notin E'$ with probability $p(1-s)$
  - $\{i,j\} \notin E$, $\{i,j\} \in E'$ with probability $p(1-s)$
  - $\{i,j\} \notin E$, $\{i,j\} \notin E'$ with probability $1 - p(2-s)$




$G \sim \mathrm{ER}(n, p)$, $G' \sim \mathrm{ER}(n, p)$

$s$ : correlation parameter (identical for $s = 1$, independent for $s = p$)

# An inference problem

- $(G, G') \sim \mathrm{ER}(n, p, s)$
- choose $\pi^*$, an uniformly random permutation of $V$
- set $H = (G')^{\pi^*}$, re-label the vertices through $\pi^*$
- given the observation of $(G, H)$, can one infer $\pi^*$ ?

  i.e. can one “align” the graphs ?

  Bayesian setting, the model is known to the observer

Motivations :

- De-anonymization of social networks

  [Narayanan, Shmatikov 08]
- Analysis of graph-structured data (e.g. biological networks)

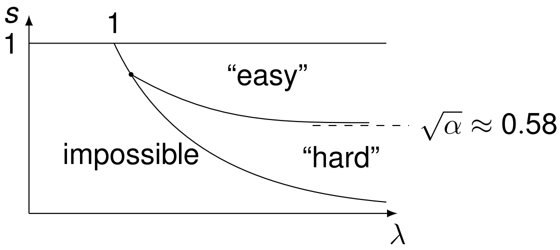
  [Singh, Xu, Berger 08]
- benchmark problem for graph neural networks

  [Nowak, Villar, Bandeira, Bruna 18]

# Phase diagram in the sparse regime

$p = \lambda/n$, $n \to \infty$, with $\lambda$ fixed

task : infer correctly a positive fraction of the elements of $\pi^*$




$\alpha$ : Otter's constant [Otter 48]

- $\lambda s < 1 \Rightarrow$ impossible [Ganassali, Massoulié, Lelarge 21a]
- $\lambda s > 1 \Rightarrow$ information-theoretically possible [Ding, Du 22]
- "easy" : polynomial-time message passing algorithm

  [Ganassali, Massoulié, Lelarge 21b]
  [Piccioli, GS, Sicuro, Zdeborová 21]
  [Ganassali, Massoulié, GS 22]

other results for $p = \Theta\left(\frac{\log n}{n}\right)$ and $p = \Theta\left(\frac{n^\alpha}{n}\right)$ with $\alpha \in (0,1]$

(almost) exact recovery of $\pi^*$ becomes possible

[Mao, Wu, Xu, Yu 22]
[Ding, Li 22]

# Outline

# Algorithm

goal : build $\widehat{\pi} = \widehat{\pi}(G, H)$, "as close as possible" from $\pi^*$

if $\widehat{\pi}$ is a function from $[n]$ to $[n]$ (not necessarily bijective), and the loss is $d(\widehat{\pi}, \pi^*) = \sum_i \mathbb{1}(\widehat{\pi}(i) \neq \pi^*(i))$

then optimal estimator : $\widehat{\pi}(i) = \text{argmax}_{i'} \mathbb{P}(\pi(i) = i' | G, H)$

posterior untractable, use instead a truncation :

$\widehat{\pi}(i) = \text{argmax}_{i'} \mathbb{P}(\pi(i) = i' | G_i^{(d)}, H_{i'}^{(d)})$ with

- $G_i^{(d)}$, depth $d$ neighborhood of $i$ in $G$
- $H_{i'}^{(d)}$, depth $d$ neighborhood of $i'$ in $H$

$d = O(\log n)$ in the following

# Algorithm

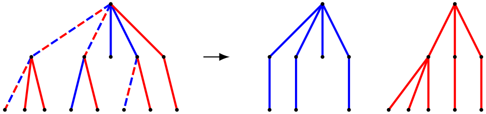
$$\widehat{\pi}(i) = \text{argmax}_{i'} \mathbb{P}(\pi(i) = i' | G_i^{(d)}, H_{i'}^{(d)})$$

$$\begin{aligned}
\mathbb{P}(\pi(i) = i' | G_i^{(d)}, H_{i'}^{(d)}) &= \frac{\mathbb{P}(\pi(i) = i', G_i^{(d)}, H_{i'}^{(d)})}{\mathbb{P}(G_i^{(d)}, H_{i'}^{(d)})} \\
&= \frac{\mathbb{P}(G_i^{(d)}, H_{i'}^{(d)} | \pi(i) = i')}{\mathbb{P}(G_i^{(d)}, H_{i'}^{(d)})} \mathbb{P}(\pi(i) = i') \\
&= \frac{\mathbb{P}(G_i^{(d)}, H_{i'}^{(d)} | \pi(i) = i')}{\mathbb{P}(G_i^{(d)}, H_{i'}^{(d)})} \frac{1}{n}
\end{aligned}$$

in the large $n$ limit, ratio of the probabilities of two neighborhoods with aligned roots vs random roots

# Algorithm

- if $i' = \pi^*(i)$, $G_i^{(d)}$ and $H_{i'}^{(d)}$ are correlated Galton-Watson trees



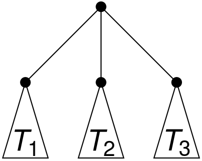
  three types of offsprings, Poisson laws of parameters $\lambda s$, $\lambda(1-s)$, $\lambda(1-s)$
  joint law of the neighborhoods : $P_1^{(d)}(T, T')$
- otherwise they are (essentially) independent Galton-Watson trees, offsprings Poisson of mean $\lambda$
  neighborhoods have law $P_0^{(d)}(T) P_0^{(d)}(T')$

if $P_1^{(d)}$ is sufficiently distinct from $P_0^{(d)} \otimes P_0^{(d)}$ when $d$ grows,
one can pick the right $i'$

# Algorithm

$L^{(d)}(T,T') = \frac{P_1^{(d)}(T,T')}{P_0^{(d)}(T)P_0^{(d)}(T')}$ likelihood ratio, recursive computation :

$T =$




$\ell$ : degree of the root of $T$, $T = (T_1, \dots, T_\ell)$

idem for $T' = (T'_1, \dots, T'_{\ell'})$

$$L^{(d)}(T,T') = f(\{L^{(d-1)}(T_i,T'_{i'})\}_{i\in[\ell]}^{i'\in[\ell']})$$

$$f(\{L_{i,i'}\}) = \sum_{k=0}^{\min(\ell,\ell')} e^{\lambda s}(1-s)^{\ell+\ell'} \left(\frac{s}{\lambda(1-s)^2}\right)^k \sum_{I,I',\sigma} \prod_{i\in I} L_{i,\sigma(i)} ,$$

with $|I| = |I'| = k$ and $\sigma : I \to I'$ bijective
generalized permanent of the $\ell \times \ell'$ matrix,
computational cost grows factorially with the degrees

# Algorithm

can be turned into a message passing algorithm :

- compute the "scores" $L_{ii'}^{(d)} = L^{(d)}(G_i^{(d)}, H_{i'}^{(d)})$ for all pairs of vertices
- from messages $L_{ii' \to jj'}^{(t)}$ with $t = 1, \dots d$ (likelihood ratio between the neighborhood of $i$ deprived from its neighbor $j$ in $G$ and the neighborhood of $i'$ deprived from its neighbor $j'$ in $H$)
- return $\widehat{\pi}(i) = \mathrm{argmax}_{i'} L_{ii'}^{(d)}$

quadratic number of messages, with update cost factorial in the degrees, $\ell_{\max} = \Theta\left(\frac{\log n}{\log \log n}\right)$ hence still $\mathrm{poly}(n)$

if $P_1^{(d)}$ is sufficiently distinct from $P_0^{(d)} \otimes P_0^{(d)}$ when $d$ grows, $L_{i\pi^*(i)}^{(d)} \gg L_{ii'}^{(d)}$ for $i' \neq \pi^*(i)$ (with positive probability), hence $\pi^*(i)$ can be recovered

# Algorithm

more formally : [Ganassali, Massoulié, Lelarge 21b]

if for some value of $(\lambda, s)$, as $d \to \infty$

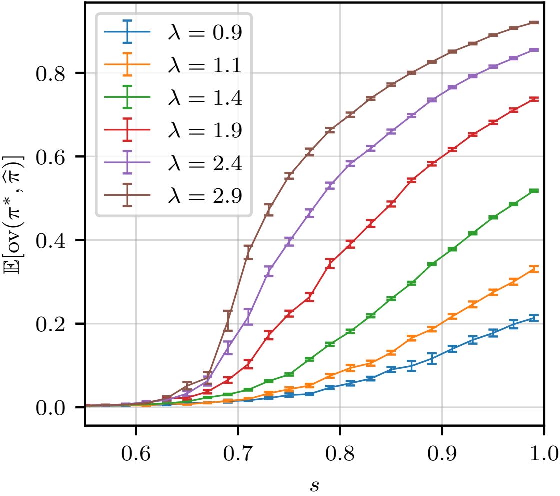
$$\mathrm{KL}(P_1^{(d)} || P_0^{(d)} \otimes P_0^{(d)}) = \mathbb{E}_1[\log L^{(d)}(T, T')] \to \infty$$

then the partial recovery of $\pi^*$ is feasible in polynomial time
(with a slightly different algorithm)

corresponds to the one-sided feasibility of the hypothesis testing problem on trees

# Algorithm

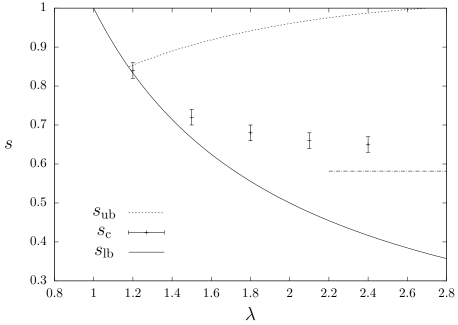
graphs of size $n = 2048$




crossover around $s \approx 0.6$ for most of these $\lambda$

# Algorithm

from the study of the tree problem, divergence of $\mathrm{KL}(P_1^{(d)}||P_0^{(d)} \otimes P_0^{(d)})$, phase diagram :

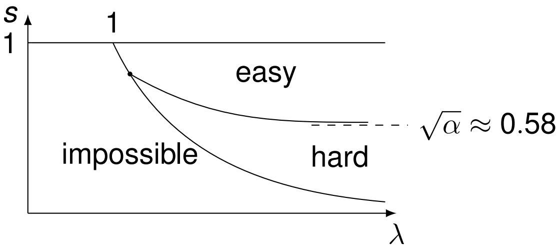



dot-dashed line at $\sqrt{\alpha} \approx 0.58$, limit of the transition line for $\lambda \to \infty$ ?
this Otter threshold appeared before in the detection problem

[Mao, Wu, Xu, Yu 21]

# Outline

# Otter's threshold in the large $\lambda$ limit




definition of Otter's constant $\alpha$ : $\alpha^{-1} = \lim_{n\to\infty} \frac{1}{n} \log(A_n)$

with $A_n$ the number of rooted, unlabelled trees, on $n$ vertices

- for all $s < \sqrt{\alpha}$, all $\lambda$, $\limsup \mathrm{KL}(P_1^{(d)} || P_0^{(d)} \otimes P_0^{(d)}) < \infty$ as $d \to \infty$
- for all $s > \sqrt{\alpha}$, all $\lambda > \lambda_c(s)$, $\mathrm{KL}(P_1^{(d)} || P_0^{(d)} \otimes P_0^{(d)}) \to \infty$ as $d \to \infty$

[Ganassali, Massoulié, GS 22]
[Mao, Wu, Xu, Yu 22]

# Otter's threshold in the large $\lambda$ limit

Some ideas of the proof :

- $\lambda \to \infty$ (after $n \to \infty$) should bring some Gaussianity
- for $d = 1$, $T \equiv \ell$ (degree of the root)

$$\frac{\mathrm{Po}(\lambda)-\lambda}{\sqrt{\lambda}} \xrightarrow[\lambda\to\infty]{\mathrm{d}} \mathcal{N}(0,1)$$

$$\mathrm{KL}(P_1^{(1)} || P_0^{(1)} \otimes P_0^{(1)}) \underset{\lambda\to\infty}{\longrightarrow} \mathrm{KL}\left(\mathcal{N}\left(0, \begin{pmatrix} 1 & s \\ s & 1 \end{pmatrix}\right) \middle|\middle| \mathcal{N}\left(0, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)\right) = -\frac{1}{2}\log(1-s^2)$$

# Otter's threshold in the large $\lambda$ limit

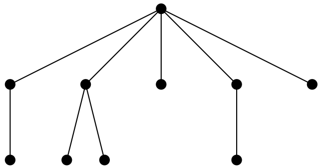
- for larger $d$ :

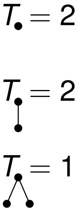
  $\chi_d = \{$rooted unlabelled trees of depth at most $d\}$

  $\chi_{d+1} = \mathbb{N}^{\chi_d}$ : number of copies of subtrees under the root

  $T \in \chi_{d+1} = \{T_t\}_{t \in \chi_d}$

  Example :

$T =$ 



# Otter's threshold in the large $\lambda$ limit

shift, rescale and rotate : $$y_\beta = \sum_{t \in \chi_d} f_\beta^{(d)}(t) \frac{T_t - \lambda \mathbb{P}_0^{(d)}(t)}{\sqrt{\lambda}} \qquad \beta \in \chi_d$$

$(y, y')$ becomes (infinite-dimensional) Gaussian vector as $\lambda \to \infty$ with covariance :

- diagonal under $P_0^{(d+1)} \otimes P_0^{(d+1)}$
- $2 \times 2$ block-diagonal under $P_1^{(d+1)}$

$$\mathrm{KL}(P_1^{(d+1)} || P_0^{(d+1)} \otimes P_0^{(d+1)}) \underset{\lambda \to \infty}{\longrightarrow} -\frac{1}{2} \sum_{\beta \in \chi_d} \log(1 - s^{2|\beta|})$$

with $|\beta|$ the number of vertices in the ("dual") tree $\beta$

when $d \to \infty$, the sum diverges if $s^2 \alpha > 1$

the $f_\beta^{(d)}(t)$ are orthogonal polynomials (with respect to the Galton-Watson measure $P_0$), generalizing the Charlier polynomials, defined by recursion on $d$

# Outline

# Simplified algorithm

[Muratori, GS in preparation]

the orthogonal polynomials $f_\beta^{(d)}(t)$ “diagonalize” the likelihood ratio :

$$L^{(d)}(t,t') = \sum_{\beta \in \chi_d} s^{|\beta|-1} f_\beta^{(d)}(t) f_\beta^{(d)}(t')$$

recall that $\chi_d = \{$rooted unlabelled trees of depth at most $d\}$

inspired by the “low degree polynomial method”, introduce for $m \geq 2$

$$L_m^{(d)}(t,t') = \sum_{\beta \in \chi_{d,m}} s^{|\beta|-1} f_\beta^{(d)}(t) f_\beta^{(d)}(t')$$

in $\chi_{d,m}$, restrict the number of offsprings (of the dual trees $\beta$) to be $\leq m$

discards some information, not $\geq 0$ anymore

# Simplified algorithm

recursive nature of $\chi_{d,m}$ translates into recursive computation of $L_m^{(d)}$ :

$$L_m^{(d)}(T,T') = f_m(\{L_m^{(d-1)}(T_i,T'_{i'})\}_{i\in[\ell]}^{i'\in[\ell']})$$

$$f(\{L_{i,i'}\}) = \sum_{k=0}^{\min(\ell,\ell')} e^{\lambda s}(1-s)^{\ell+\ell'}\left(\frac{s}{\lambda(1-s)^2}\right)^k \sum_{I,I',\sigma}\prod_{i\in I} L_{i,\sigma(i)} \ ,$$

$f_m = f$ truncated to order $s^m$ : much faster to compute
( in $O(ll')$ operations for $m = 2, 3$)

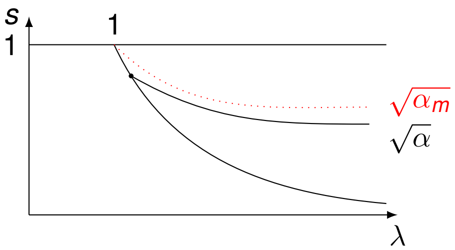
Otter's modified constant on the growth of the number of trees with offspring $\leq m$ :

$\sqrt{\alpha_2} \approx 0.63$, $\sqrt{\alpha_3} \approx 0.60$ [Otter 48]

not so far from $\sqrt{\alpha} \approx 0.58$

# Simplified algorithm

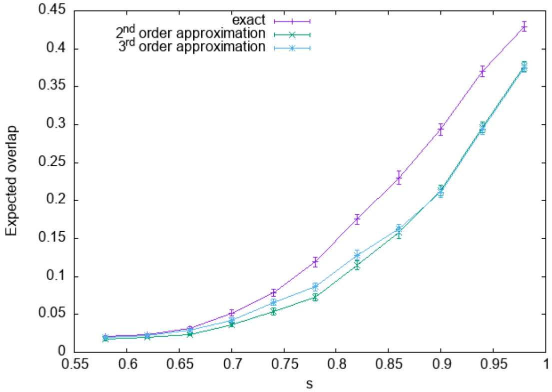
expected phase transition for the simplified algorithm at level $m$:



# Simplified algorithm

some preliminary numerical results :

- $n = 512$, $\lambda = 1.2$ :



some preliminary numerical results :

- $n = 1024$, threshold 0.05 on the overlap :