

NEW ALGORITHMS FOR THE GRAPH ALIGNMENT PROBLEM

Marc Lelarge

INRIA, DI/ENS, PSL Research University

joint work with Luca Ganassali, Laurent Massoulié and Waïss Azizian

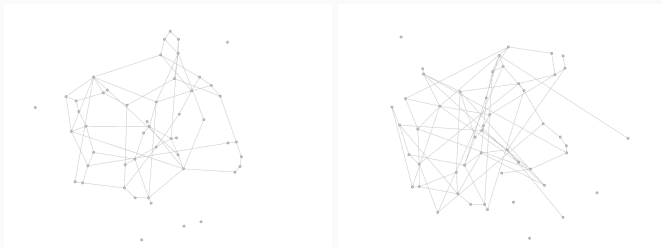
Simons Institute - Oct 2021

Alignment of Graphs

Problem : alignment of graphs

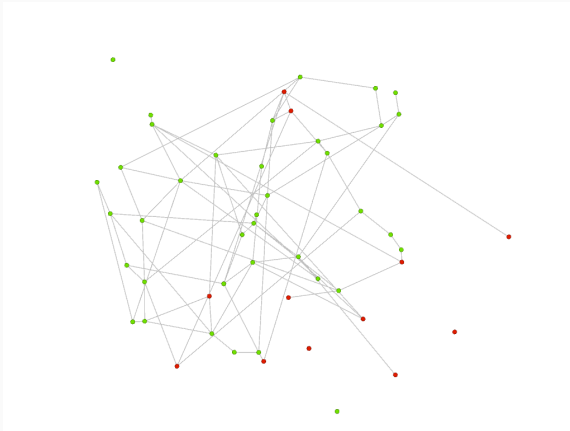
From graph 1 (on the left), put indices on its vertices, perturb the graph by adding and removing a few edges and remove indices to obtain graph 2 (on the right).

Task : recover the indices on vertices of graph 2.



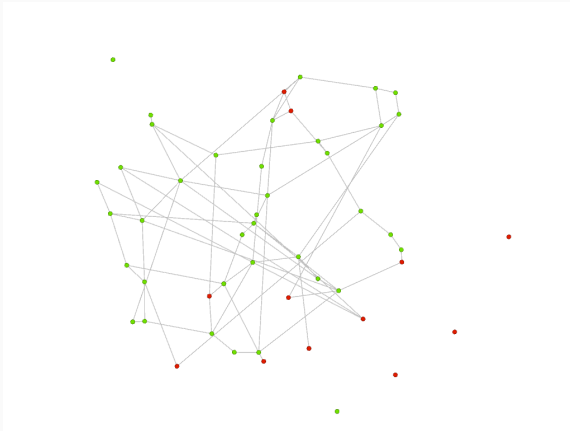
Result with FGNN

Green vertices are good predictions. Red vertices are errors (graph 2).



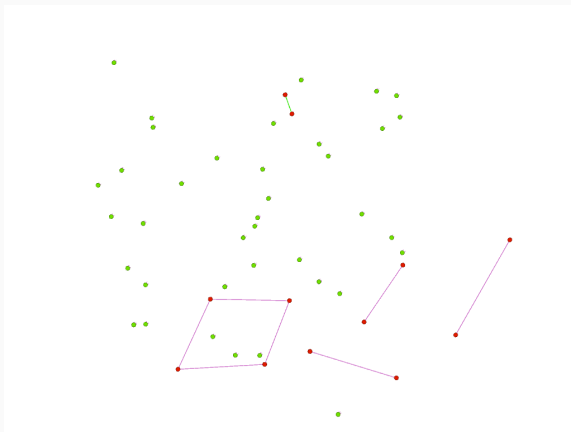
Result with FGNN

Green vertices are good predictions. Red vertices are errors (graph 1).



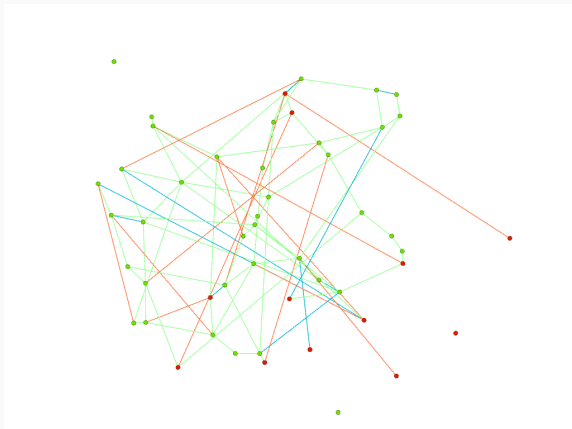
Result with FGNN

Here are the 'wrong' matchings or cycles.



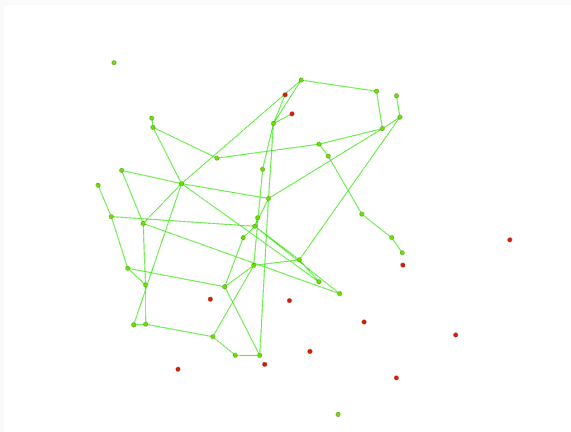
Result with FGNN

Superposing the 2 graphs : green edges in both, orange and blue edges in graph 1 and 2 resp.

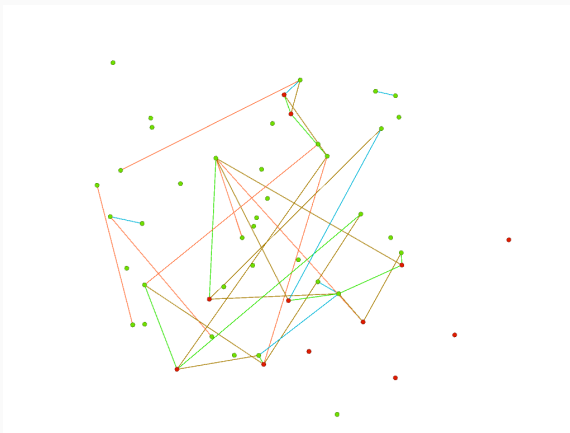


Result with FGNN

Matched edges.

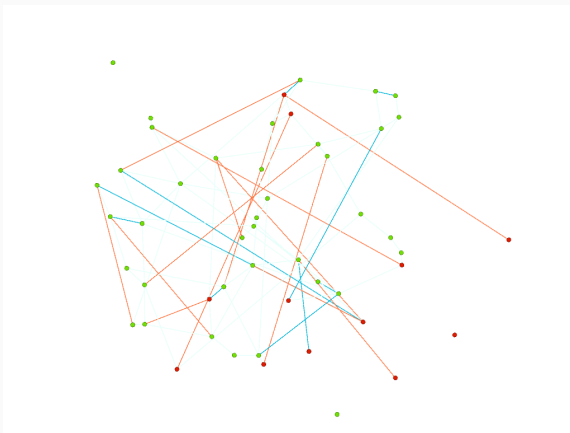


Mismatched edges.



Result with FGNN

Green vertices are well paired vertices. Red vertices are errors.



Theoretical results

Correlated Erdős-Rényi (1)

Two graphs \mathcal{G} and \mathcal{G}' with the same set of nodes $[n]$ and with respectively blue and red edges. The blue and red edges are obtained by sampling uniformly at random :

- with probability $\lambda s/n$ to get two-colored edges ;
- with probability $\lambda(1-s)/n$ to get a blue (monochromatic) edge ;
- with probability $\lambda(1-s)/n$ to get a red (monochromatic) edge ;
- with probability $1 - \lambda(2-s)/n$ to get a non-edge,

where $\lambda > 0$ and $s \in [0, 1]$ are fixed parameters when n will tend to infinity.

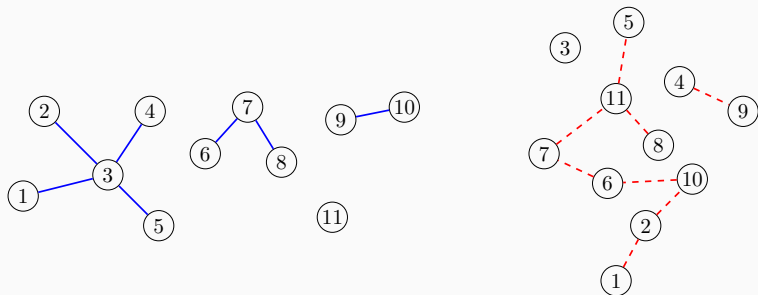


Correlated Erdős-Rényi (2)

We then relabel the vertices of the red graph \mathcal{G}' with an uniform independent permutation $\pi^* \in \mathcal{S}_n$, and we observe \mathcal{G} and $\mathcal{H} := \mathcal{G}'\pi^*$.

The marginals \mathcal{G}, \mathcal{H} are Erdős-Rényi random graphs with average degree λ .

The goal is to estimate the latent vertex correspondence π^* .



Dependence of joint distribution in $e(\mathbf{G} \wedge \mathbf{G}')$:

$$\mathbb{P}(\mathcal{G} = \mathbf{G}, \mathcal{G}' = \mathbf{G}') \propto \left[\frac{s(n - \lambda(2 - s))}{\lambda(1 - s)^2} \right]^{e(\mathbf{G} \wedge \mathbf{G}')} .$$

MAP estimator of π^* is the solution of the QAP problem :

$$\arg \max_{\Pi} \langle \mathbf{A}, \Pi \mathbf{B} \Pi^T \rangle$$

Dependence of joint distribution in $\mathbf{e}(\mathbf{G} \wedge \mathbf{G}')$:

$$\mathbb{P}(\mathcal{G} = \mathbf{G}, \mathcal{G}' = \mathbf{G}') \propto \left[\frac{s(n - \lambda(2 - s))}{\lambda(1 - s)^2} \right]^{e(\mathbf{G} \wedge \mathbf{G}')} .$$

MAP estimator of π^* is the solution of the QAP problem :

$$\arg \max_{\Pi} \langle \mathbf{A}, \Pi \mathbf{B} \Pi^T \rangle$$

Only feasible when $\lambda = \Omega(\log n)$ Cullina and Kiyavash (2017), in polynomial time when noise $1 - s$ is small Ding et al. (2018) Mao et al. (2021).

Dependence of joint distribution in $\mathbf{e}(\mathbf{G} \wedge \mathbf{G}')$:

$$\mathbb{P}(\mathcal{G} = \mathbf{G}, \mathcal{G}' = \mathbf{G}') \propto \left[\frac{s(n - \lambda(2 - s))}{\lambda(1 - s)^2} \right]^{e(\mathbf{G} \wedge \mathbf{G}')} .$$

MAP estimator of π^* is the solution of the QAP problem :

$$\arg \max_{\Pi} \langle \mathbf{A}, \Pi \mathbf{B} \Pi^T \rangle$$

Only feasible when $\lambda = \Omega(\log n)$ Cullina and Kiyavash (2017), in polynomial time when noise $1 - s$ is small Ding et al. (2018) Mao et al. (2021).

In the rest of this talk, we consider the sparse regime where the average degree λ is not scaling with n .

Planted graph alignment

For any subset $\mathcal{C} \subset [n]$, the performance of any one-to-one estimator $\hat{\sigma} : \mathcal{C} \rightarrow [n]$

$$\text{ov}(\pi^*, \hat{\sigma}) := \frac{1}{n} \sum_{i \in \mathcal{C}} \mathbf{1}_{\hat{\sigma}(i) = \pi^*(i)}.$$

Note that the estimator $\hat{\sigma}$ only consists in a partial matching. The *error fraction* of $\hat{\sigma}$ with the unknown permutation π^* is defined as

$$\text{err}(\pi^*, \hat{\sigma}) := \frac{1}{n} \sum_{i \in \mathcal{C}} \mathbf{1}_{\hat{\sigma}(i) \neq \pi^*(i)} = \frac{|\mathcal{C}|}{n} - \text{ov}(\pi^*, \hat{\sigma}).$$

Planted graph alignment

For any subset $\mathcal{C} \subset [n]$, the performance of any one-to-one estimator $\hat{\sigma} : \mathcal{C} \rightarrow [n]$

$$\text{ov}(\pi^*, \hat{\sigma}) := \frac{1}{n} \sum_{i \in \mathcal{C}} \mathbf{1}_{\hat{\sigma}(i) = \pi^*(i)}.$$

Note that the estimator $\hat{\sigma}$ only consists in a partial matching. The *error fraction* of $\hat{\sigma}$ with the unknown permutation π^* is defined as

$$\text{err}(\pi^*, \hat{\sigma}) := \frac{1}{n} \sum_{i \in \mathcal{C}} \mathbf{1}_{\hat{\sigma}(i) \neq \pi^*(i)} = \frac{|\mathcal{C}|}{n} - \text{ov}(\pi^*, \hat{\sigma}).$$

A sequence of injective estimators $\{\hat{\sigma}_n\}_n$ is said to achieve

- *Partial recovery* if there exists some $\epsilon > 0$ such that $\mathbb{P}(\text{ov}(\pi^*, \hat{\sigma}) > \epsilon) \xrightarrow{n \rightarrow \infty} 1$,
- *One-sided partial recovery* if it achieves partial recovery and $\mathbb{P}(\text{err}(\pi^*, \hat{\sigma}) = o(1)) \xrightarrow{n \rightarrow \infty} 1$.

A (local) algorithm

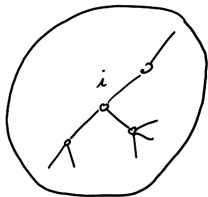
i
o

$\tau^*(i)?$

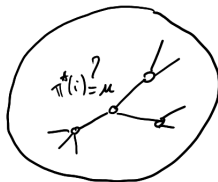
eg

\mathcal{H}

A (local) algorithm

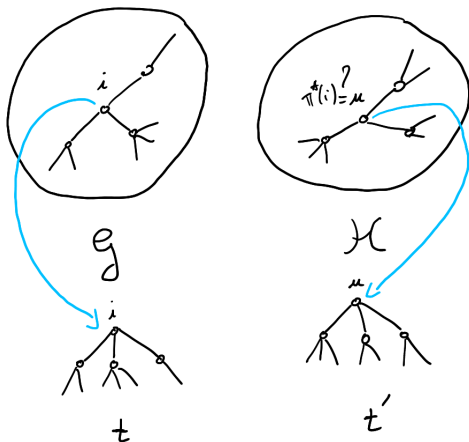


g

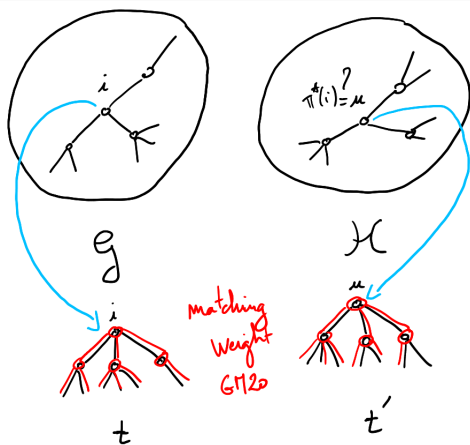


H

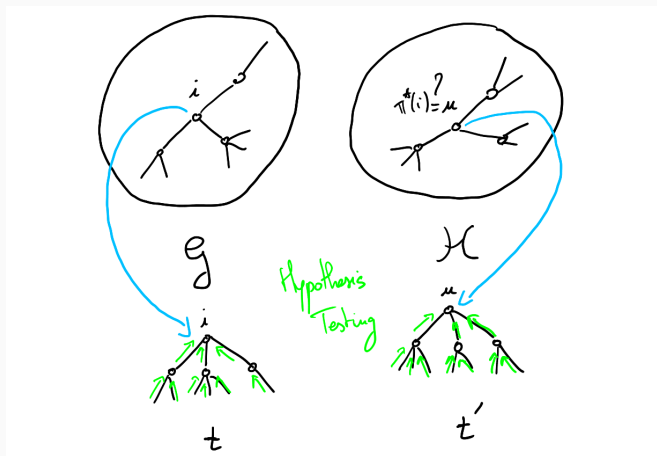
A (local) algorithm



A (local) algorithm



A (local) algorithm

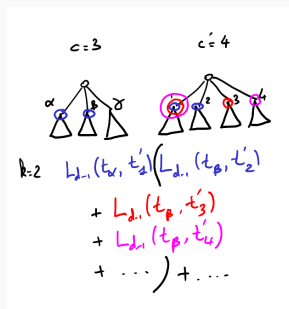


For $i \in V(\mathcal{G})$, $u \in V(\mathcal{H})$, look at the neighborhoods \mathcal{N}_i and \mathcal{N}_u at depth d :

- if $u = \pi^*(i)$, $(\mathcal{N}_i, \mathcal{N}_u) \simeq$ GW trees of offspring $\text{Poi}(\lambda)$, with intersection of offspring $\text{Poi}(\lambda s)$ (model $\mathbb{P}_{1,d}$);
- if $u \neq \pi^*(i)$, $(\mathcal{N}_i, \mathcal{N}_u) \simeq$ independent GW trees of offspring $\text{Poi}(\lambda)$ (model $\mathbb{P}_{0,d}$).

Hypothesis testing : Can we test $\mathbb{P}_{1,d}$ versus $\mathbb{P}_{0,d}$?

Computing the likelihood ratio



For two trees of depth d , the likelihood ratio $L_d(t, t') := \frac{\mathbb{P}_{1,d}(t, t')}{\mathbb{P}_{0,d}(t, t')}$ verifies

$$L_d(t, t') = \sum_{k=0}^{c \wedge c'} \psi(k, c, c') \sum_{\substack{\sigma \in \mathcal{S}(k, c) \\ \sigma' \in \mathcal{S}(k, c')}} \prod_{i=1}^k L_{d-1}(t_{\sigma(i)}, t'_{\sigma'(i)}),$$

where c and c' are the number of children of the roots,
 $\psi(k, c, c') = e^{\lambda s} \times \frac{s^{k_5^{c+c'} - 2k}}{\lambda^k k!}$, and $\mathcal{S}(k, \ell)$ denotes the set of injective mappings from $[k]$ to $[\ell]$.

One-sided tests : tests $\mathcal{T}_d : \mathcal{X}_d \times \mathcal{X}_d \rightarrow \{0, 1\}$ such that $\mathbb{P}_{0,d}(\mathcal{T}_d = 0) = 1 - o(1)$ and $\liminf_d \mathbb{P}_{1,d}(\mathcal{T}_d = 1) > 0$ (i.e. vanishing type I error and non vanishing power).

Theorem

Let

$$KL_d := KL(\mathbb{P}_{1,d} \parallel \mathbb{P}_{0,d}) = \mathbb{E}_{1,d} [\log(L_d)].$$

Then the following propositions are equivalent :

- (i) There exists a one-sided test for deciding $\mathbb{P}_{0,d}$ versus $\mathbb{P}_{1,d}$,
- (ii) $\lim_{d \rightarrow \infty} KL_d = +\infty$ and $\lambda_S > 1$,
- (iii) with probability $1 - p_{\text{ext}}(\lambda_S) > 0$, L_d diverges to $+\infty$ with rate $\Omega\left(\exp\left(\Omega(1) \times (\lambda_S)^d\right)\right)$.

One-sided partial graph alignment

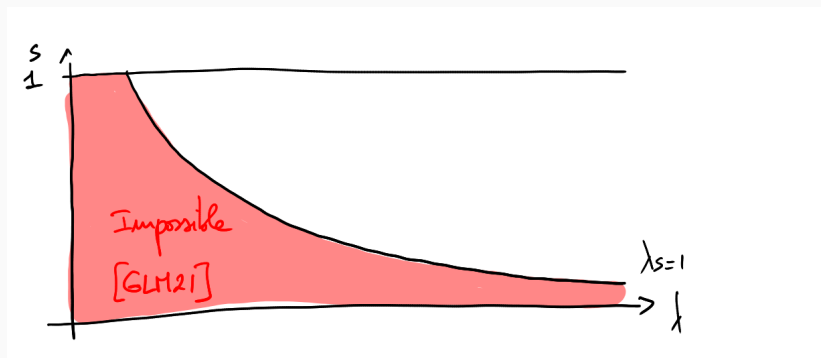
Recall : estimator $\hat{\sigma} : \mathcal{C} \rightarrow [n]$ is said to achieve

- *Partial recovery* if there exists some $\epsilon > 0$ such that
$$\mathbb{P}(\text{ov}(\pi^*, \hat{\sigma}) > \epsilon) \xrightarrow{n \rightarrow \infty} 1,$$
- *One-sided partial recovery* if it achieves partial recovery and
$$\mathbb{P}(\text{err}(\pi^*, \hat{\sigma}) = o(1)) \xrightarrow{n \rightarrow \infty} 1.$$

Theorem

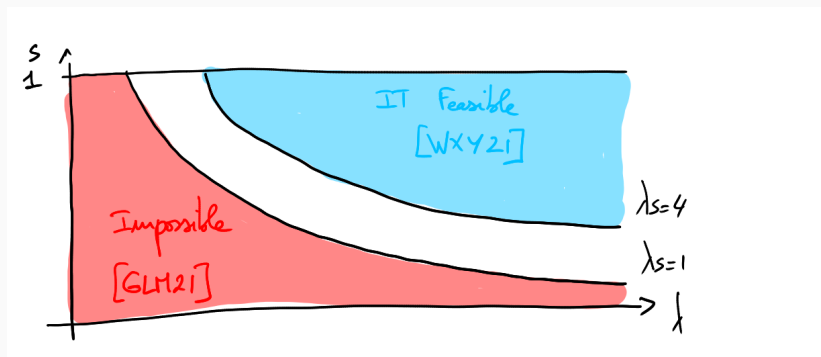
For given (λ, \mathbf{s}) , if one-sided correlation detection is feasible, then one-sided partial alignment in the correlated Erdős-Rényi model $\mathcal{G}(n, \lambda/n, \mathbf{s})$ is achieved in polynomial time by our algorithm.

Results in the regime with constant mean degree



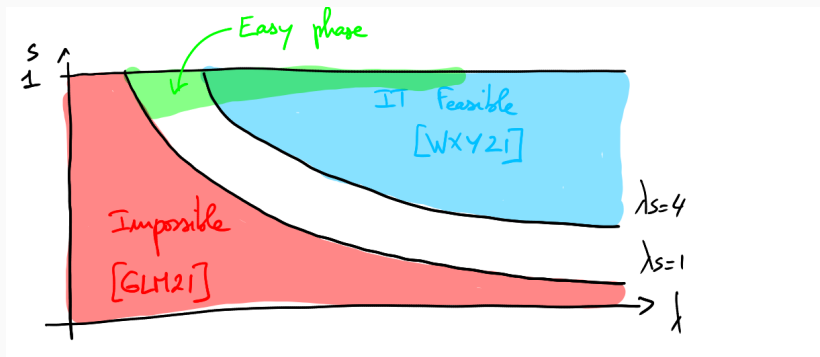
Ganassali et al. (2021b)

Results in the regime with constant mean degree



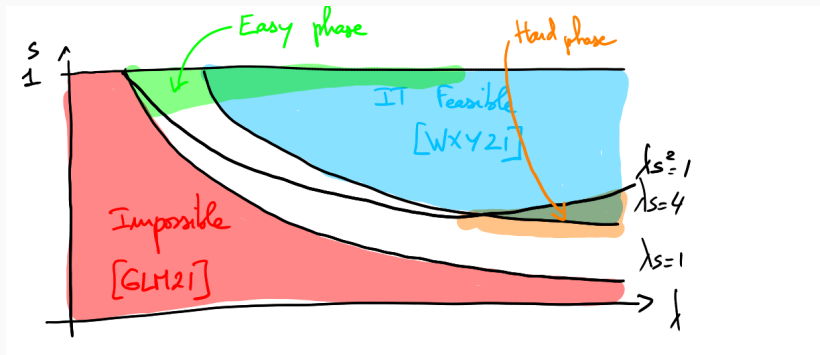
Wu et al. (2021)

Results in the regime with constant mean degree



Sufficient conditions for the existence of one-sided test based on the Kullback-Leibler divergence or the number of automorphisms of GW trees. Ganassali et al. (2021a)

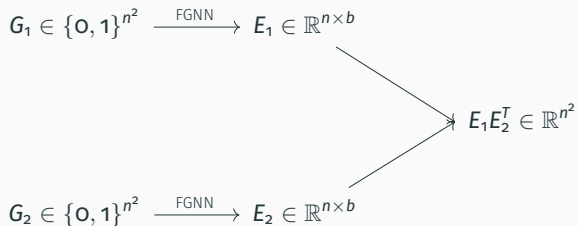
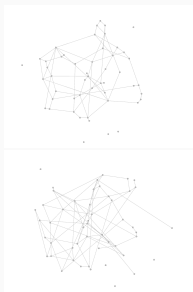
Results in the regime with constant mean degree



Conjectured hard phase based on the impossibility of one-sided test because KL_d is bounded. Ganassali et al. (2021a)

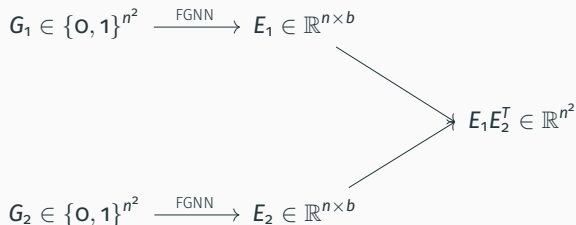
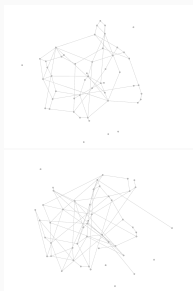
A learning algorithm

Learning the graph alignment problem with Siamese FGNNs



- The same FGNN is used for both graphs.

Learning the graph alignment problem with Siamese FGNNs



- The same FGNN is used for both graphs.
- We added a LAP solver to get a permutation from $\mathbf{E}_1 \mathbf{E}_2^T$.

Challenges in designing a learning algorithm for graphs

For a permutation $\sigma \in \mathcal{S}_n$, for $\mathbf{G} \in \mathbb{F}^{n \times n}$ ($\mathbb{F} = \mathbb{R}^p$ feature space), we define :

$$(\sigma \star \mathbf{G})_{\sigma(i_1), \sigma(i_2)} = \mathbf{G}_{i_1, i_2}.$$

$\mathbf{G}_1, \mathbf{G}_2$ are isomorphic iff $\mathbf{G}_1 = \sigma \star \mathbf{G}_2$.

Challenges in designing a learning algorithm for graphs

For a permutation $\sigma \in \mathcal{S}_n$, for $\mathbf{G} \in \mathbb{F}^{n \times n}$ ($\mathbb{F} = \mathbb{R}^p$ feature space), we define :
 $(\sigma \star \mathbf{G})_{\sigma(i_1), \sigma(i_2)} = \mathbf{G}_{i_1, i_2}$.

$\mathbf{G}_1, \mathbf{G}_2$ are isomorphic iff $\mathbf{G}_1 = \sigma \star \mathbf{G}_2$.

Equivariance :

A function $f : \mathbb{F}^{n^2} \rightarrow \mathbb{F}^n$ is said to be **equivariant** if $f(\sigma \star \mathbf{G}) = \sigma \star f(\mathbf{G})$.

For the graph alignment problem, we used an equivariant GNN from $\{0, 1\}^{n \times n}$ to \mathbb{F}^n .

Challenges in designing a learning algorithm for graphs

For a permutation $\sigma \in \mathcal{S}_n$, for $G \in \mathbb{F}^{n \times n}$ ($\mathbb{F} = \mathbb{R}^p$ feature space), we define :
 $(\sigma \star G)_{\sigma(i_1), \sigma(i_2)} = G_{i_1, i_2}$.

G_1, G_2 are isomorphic iff $G_1 = \sigma \star G_2$.

Equivariance :

A function $f : \mathbb{F}^{n^2} \rightarrow \mathbb{F}^n$ is said to be **equivariant** if $f(\sigma \star G) = \sigma \star f(G)$.

For the graph alignment problem, we used an equivariant GNN from $\{0, 1\}^{n \times n}$ to \mathbb{F}^n .

Expressiveness : Azizian and Lelarge (2020)

FGNN has the **best power of approximation** among all architectures working with tensors of order 2 (MGNN or LGNN).

(Maron et al., 2019) to propose the **folklore graph layer (FGL)** :

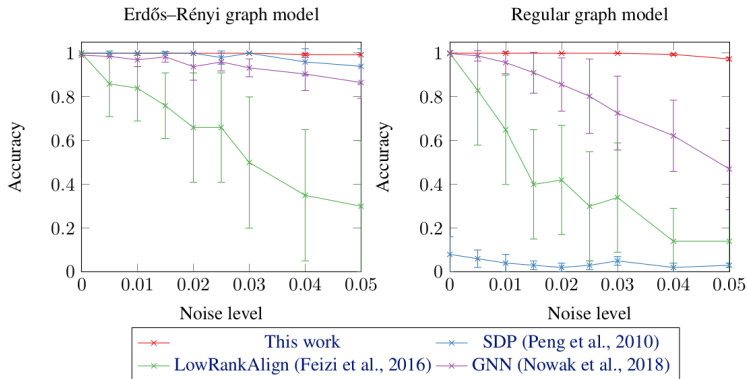
$$h_{i \rightarrow j}^{\ell+1} = f_0 \left(h_{i \rightarrow j}^{\ell}, \sum_{k \in V} f_1 \left(h_{i \rightarrow k}^{\ell} \right) f_2 \left(h_{k \rightarrow j}^{\ell} \right) \right),$$

where f_0, f_1 and f_2 are learnable functions.

For FGNNs, **messages are associated with pairs of vertices** as opposed to MGNN where messages are associated with vertices.

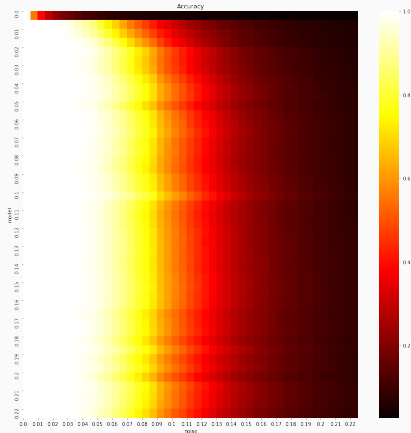
FGNN : a FGNN is the composition of FGLs and a final invariant/equivariant reduction layer from \mathbb{F}^{n^2} to \mathbb{F}/\mathbb{F}^n .

Results on synthetic data



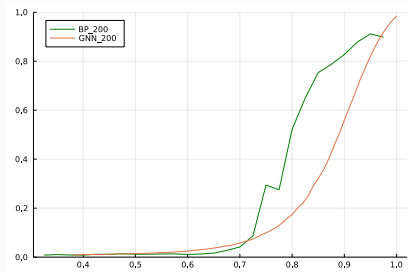
- Graphs : $n = 50$, density = 0.2
- Training set : 20000 samples
- Validation and Test sets : 1000 samples

Generalization for regular graphs



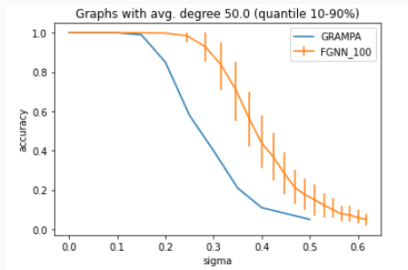
Each line corresponds to a model trained at a given noise level and shows its accuracy across all noise levels.

Comparison FGNN vs BPAIalign



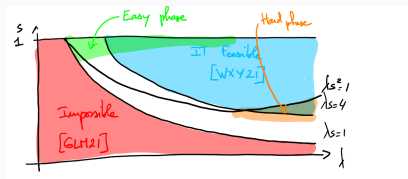
Overlap as a function of the correlation s for correlated E-R with average degree 3 (number of nodes : 200).

Comparison FGNN vs GRAMPA

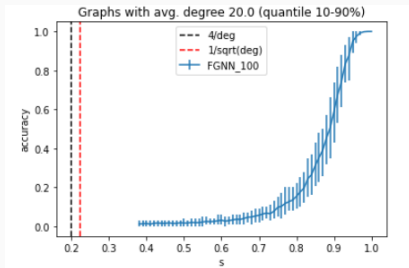
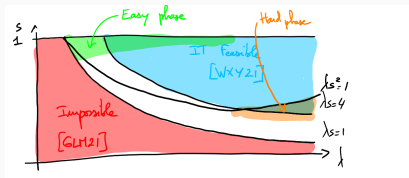


Overlap as a function of $\sigma = \sqrt{\frac{1-s}{1-\lambda/n}}$ for correlated E-R with average degree 50 (number of nodes : 100). Fan et al. (2019)

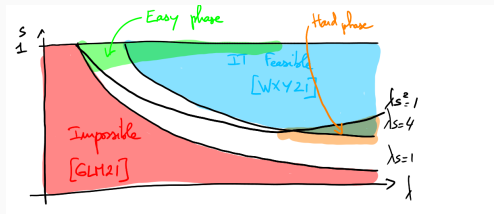
From theory to practice



From theory to practice



Conclusion



The Bitter Lesson by Rich Sutton

The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin.

Thank You!

Références

- W. Azizian and M. Lelarge. Expressive power of invariant and equivariant graph neural networks. *arXiv preprint arXiv :2006.15646*, 2020.
- D. Cullina and N. Kiyavash. Exact alignment recovery for correlated Erdős-Rényi graphs, 2017.
- J. Ding, Z. Ma, Y. Wu, and J. Xu. Efficient random graph matching via degree profiles. *arXiv e-prints*, art. arXiv :1811.07821, Nov 2018.
- Z. Fan, C. Mao, Y. Wu, and J. Xu. Spectral graph matching and regularized quadratic relaxations i : The gaussian model. *arXiv preprint arXiv :1907.08880*, 2019.
- L. Ganassali, L. Massoulié, and M. Lelarge. Correlation detection in trees for partial graph alignment. *arXiv preprint arXiv :2107.07623*, 2021a.
- L. Ganassali, L. Massoulié, and M. Lelarge. Impossibility of partial recovery in the graph alignment problem. *arXiv preprint arXiv :2102.02685*, 2021b.
- C. Mao, M. Rudelson, and K. Tikhomirov. Random graph matching with improved noise robustness. *arXiv preprint arXiv :2101.11783*, 2021.
- H. Maron, H. Ben-Hamu, H. Serviansky, and Y. Lipman. Provably powerful graph networks. In H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32 : Annual Conference on Neural*