NEURAL INFORMATION PROCESSING SYSTEMS

Aalto University School of Science

compositional PAC-Bayes:

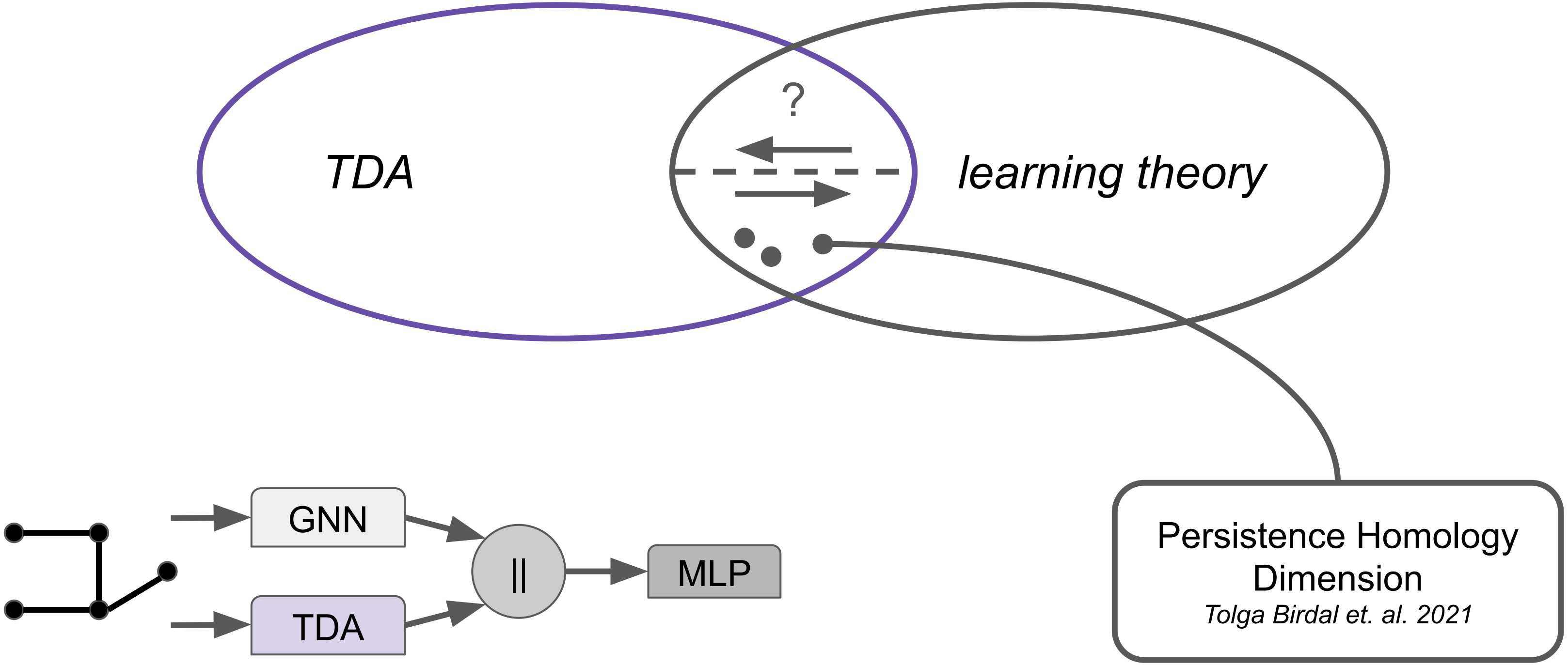
generalization of GNNs with persistence and beyond

Kirill Brilliantov
ETH Zurich

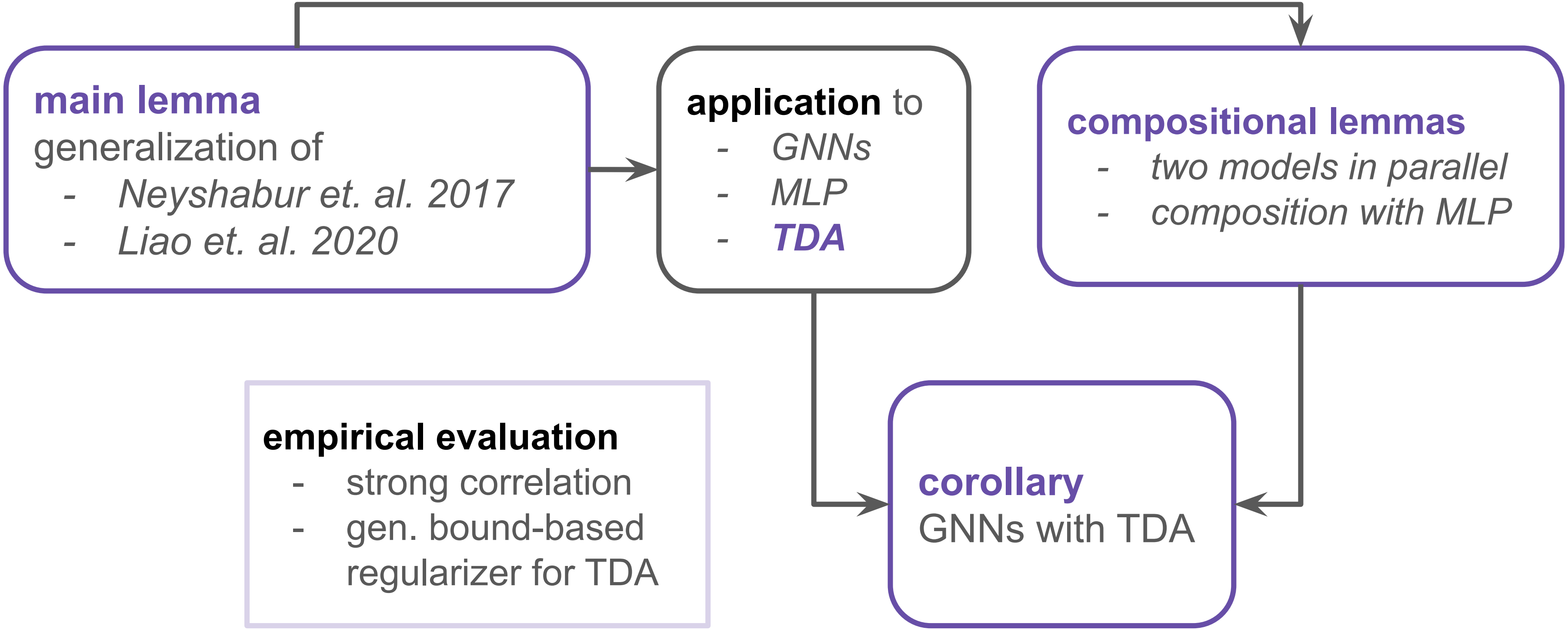
Amauri H. Souza
Federal Institute of Ceará

Vikas K. Garg,
YaiYai Ltd & Aalto University

motivation


TDA
?
learning theory
Persistence Homology
Dimension
Tolga Birdal et. al. 2021
GNN
||
MLP
TDA

overview of the contribution



main lemma: PAC Bayes

simplified PAC-Bayes *[Neyshabur et. al. 2017]*

let $f : \mathscr{X} \to \mathbb{R}^k$ be **any** model with parameters $\boldsymbol{w}$ and let $\mathcal{P}$ be a distribution over parameters, **independent** of the $\mathscr{D}$.

then for any γ, $\delta > 0$ with probability at least $1 - \delta$ over i.i.d. training set $\mathscr{S}$ of size m drawn according to $\mathscr{D}$, for any $\boldsymbol{w}$ and any random perturbation $\boldsymbol{u} \sim \mathcal{Q}(\boldsymbol{w})$
such that $\mathbb{P}(\sup_{\mathbf{x}} |f(\boldsymbol{x}; \boldsymbol{w}) - f(\boldsymbol{x}; \boldsymbol{w} + \boldsymbol{u})|_\infty < \frac{1}{4}\gamma) > \frac{1}{2}$ w.r.t $\mathcal{Q}$ we have

generalization gap^2 = $\mathcal{O}\{ D_{KL}(\mathcal{Q}(\boldsymbol{w}) \,||\, \mathcal{P});\ 1/m;\ \log(6m/\delta) \}$

main lemma

let $f : \mathscr{X} \square \mathbb{R}^k$ be **any** model with parameters $\boldsymbol{w}$ = *vec*$\{W_1, \ldots, W_n\}$; and
$T > 0$, $S_1, \ldots, S_n > 0$ which may depend on $\boldsymbol{w}$ and **parameter-independent**
$C_1 > 0$, $C_2 > 0$, $0 < \eta_1, \ldots, \eta_n < 1$ s.t.

- we can upper bound $\sup_{\mathbf{x}} | f(\boldsymbol{x}; \boldsymbol{w}) |_2$ with C_1, T
- for $\boldsymbol{u}$ = *vec*$\{U_1, \ldots, U_n\}$ with $|| U_i ||_2 < \eta_i S_i$ we can upper bound
  $\sup_{\mathbf{x}} | f(\boldsymbol{x}; \boldsymbol{w}) - f(\boldsymbol{x}; \boldsymbol{w} + \boldsymbol{u}) |_2$ with C_2, T, S_i, $|| U_i ||_2$
- T and S_i are **connected**: we can upper bound the sum of S_i with T
- η_i and C_1, C_2 are **connected**: we can upper bound the min of η_i with C_1, C_2

then for any γ, $\delta > 0$ with probability at least $1 - \delta$ over i.i.d. training set S of size m drawn according to $\mathscr{D}$ we provide generalization gap depending on γ, δ, n, m, C_1, C_2, T, S_i, $|\boldsymbol{w}|_2$, min of η_i.

applying **main lemma** to MLPs, GNNs and PersLay

defining:

- $T, S_1, \ldots, S_n$ as spectral norms of weight matrices
- $C_1, C_2; \eta_1, \ldots, \eta_n$ according to perturbation analysis [*Neyshabur et al. 2017; Liao et. al. 2020*]

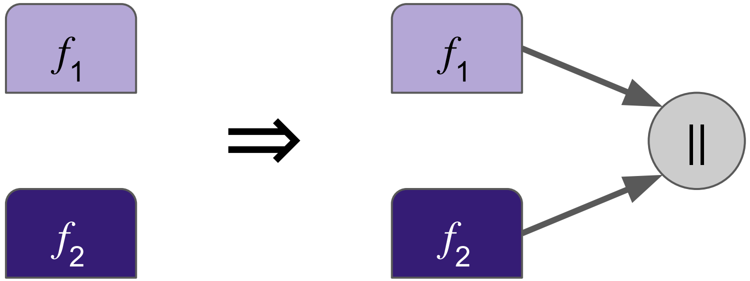
we can apply **main lemma** to MLPs and GNNs

for PersLay [*Carrière et. al. 2020*] we provide:

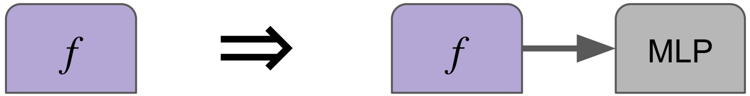
- perturbation analysis
- **main lemma** variables definition

compositional lemmas

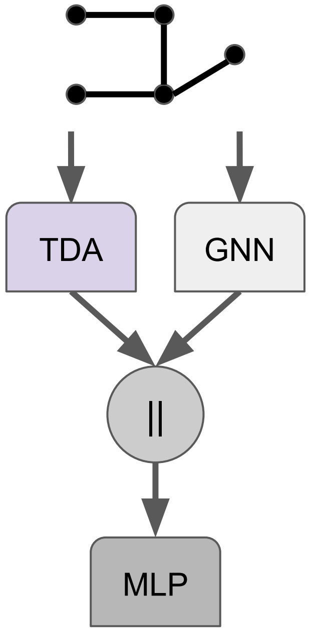
lemma (*two models in parallel*)



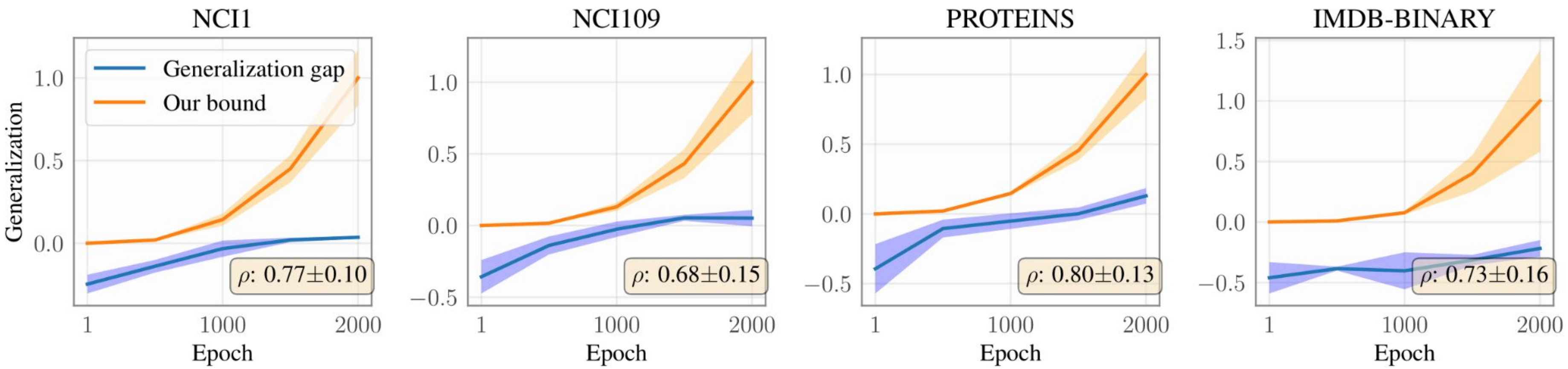
lemma (*composition with MLP*)



corollary (*GNNs with persistence*)



experiments



conclusion

- **main lemma**: general PAC-Bayes lemma
  - can be applied to well-studied models
  - can be applied to persistence homology
  - can be applied to **compositions**
- empirical evaluation shows **strong** correlation

Thank you for attention!