Graph distinction through GENEOs and Permutants

Giovanni Bocchi[1] Massimo Ferri [2] Patrizio Frosini[3]

[1]Dept. of Environmental Science and Policy, University of Milano

[2]Dept. of Mathematics, University of Bologna

[3]Dept. of Computer Science, University of Pisa

31/01/2025

Preliminaries

To formally introduce GENEOs, let's consider two functional spaces whose functions are defined on some topological spaces X and Y.

- $\Phi = \{\varphi\colon X \to \mathbb{R}\} \quad \mathrm{dom}(\Phi) = X$
- $\Psi = \{\psi\colon Y \to \mathbb{R}\} \quad \mathrm{dom}(\Psi) = Y$

Then let's consider a subgroup G (resp. K) of the group of all homeomorphisms of X (resp. Y) that are Φ-preserving (resp. Ψ-preserving), i.e. those $g \in \mathrm{Homeo}(X)$ such that $\varphi \circ g \in \Phi$ and $\varphi \circ g^{-1} \in \Phi$ for all $\varphi \in \Phi$. Finally fix a homomorphism T between G and K.

GENEOs definition

Definition (GENEO)

A Group Equivariant Non-Expansive Operator F is a map between Φ and Ψ that, for a fixed homomorphism of groups T, has these two properties:

- **Equivariance:** $F(\varphi \circ g) = F(\varphi) \circ T(g)$ for all $\varphi \in \Phi$ and for all $g \in G$.
- **Non-Expansivity:** $||F(\varphi_1) - F(\varphi_2)||_\infty \leq ||\varphi_1 - \varphi_2||_\infty$ for all $\varphi_1, \varphi_2 \in \Phi$.

Usually if $\mathrm{dom}(\Phi) = \mathrm{dom}(\Psi)$ the natural choice is $T = \mathrm{id}_G$.

GENEOs definition (cont.)

1. **Equivariance** encodes the fact that a GENEO must commute with a specific group of transformations of the data domain. In some sense we can say that GENEOs are able to filter out those transformations.
2. **Non-expansivity** implies that GENEOs tend to simplify the metric structure of data, so in some sense they provide a simpler representation of data. Moreover it is important to derive some topological properties of the space of GENEOs.

Topological properties of GENEOs space

$\mathscr{F}$ is the space of all GENEOs between (Φ, G) and (Ψ, K) w.r.t. T.

$$D_{\mathrm{GENEO}}(F_1, F_2) = \sup_{\varphi \in \Phi} ||F_1(\varphi) - F_2(\varphi)||_\infty$$

Theorem (Compactness)

If both Φ and Ψ are compact in the topology induced by the sup norm distance then also $\mathscr{F}$ is compact in the topology induced by the metric D_{GENEO}.

Theorem (Convexity)

If Ψ is convex than the space $\mathscr{F}$ is also convex.

[Generalized] permutants

Definition (Generalized permutant)

A finite subset $H \subseteq X^Y$ is a generalized permutant for $T \colon G \to K$ if $gHT(g^{-1}) \subseteq H$ for every $g \in G$ or if H is the empty set.

Proposition (A generalized permutant defines a GENEO)

If H is a generalized permutant for T then the operator $F_H \colon \Phi \to \Psi$

$$F_H(\varphi) = \begin{cases} \frac{1}{|H|} \sum_{h \in H} \varphi h & \text{if } |H| > 0 \\ 0 & \text{otherwise} \end{cases}$$

is a linear GENEO provided that $F_H(\Phi) \subseteq \Psi$.

If $\mathrm{dom}(\Phi) = \mathrm{dom}(\Psi) = X$, $H \subseteq \mathrm{Homeo}(X)$ and $T = \mathrm{id}_G$ we denote them simply permutants.

Problem (Graph isomorphism)

Given a pair of undirected and unweighted graphs $\Gamma_1 = (V_1, E_1)$ *and* $\Gamma_2 = (V_2, E_2)$ *decide whether there exists a bijection* $f : V_1 \to V_2$ *such that* $\{u, v\} \in E_1$ *if and only if* $\{f(u), f(v)\} \in E_2$*.*

L. Babai proved that the GI problem is solvable in quasi-polynomial time[1]. Anyway, **no polynomial time** algorithm is currently known and it is also unclear if it is a NP-complete problem (most experts believe it to be an NP-intermediate problem).

[1] Time complexity in the class $2^{\mathcal{O}((\log n)^C)}$

Motivation (cont.)

Definition (Isomorphism test)

An isomorphism test is a function that, given a pair of graphs (Γ_1, Γ_2) returns 0 if they are isomorphic or 1 otherwise.

1. Exact tests: return the **correct** result but they are usually **slow** (no polynomial algorithm available).
2. Inexact tests: usually provide only a sufficient condition for non isomorphism (can be **inconclusive** for some instances) but they are usually **fast** (also linear complexity).

Setting

From now on we will consider **undirected** and **unweighed** graphs with (at most) N nodes $\mathcal{G}_N$ seen as subgraphs of the complete graph K_N.

1. As domain $X = \{\{v_i, v_j\}\colon i \neq j; i, j \in \mathbb{N}_N\}$.
2. As functional space $\Phi = \{\varphi\colon X \to \{0, 1\}\}$.
3. As group $G = \{g\colon X \to X \mid g(\{v_i, v_j\}) = \{v_{\sigma(i)}, v_{\sigma(j)}\}\}$ where σ is a permutation of $\{1, \ldots, N\}$.
4. A graph $\Gamma \in \mathcal{G}_N$ is represented by $\varphi \in \Phi$ denoted by $\Gamma \sim \varphi$.
5. Γ_1, Γ_2 isomorphic $\iff$ if it exists $g \in G$ s.t. $\varphi_1 = \varphi_2 \circ g$.

Subgraphs

We will also consider auxiliary graphs Λ seen as **subgraphs** of a smaller complete graph K_k with $k \ll N$.

1. As domain $Y = \{\{v_i, v_j\} : i \neq j; i, j \in \mathbb{N}_k\}$.
2. Two functional spaces: $\Psi_0 = \{\varphi : Y \to \{0, 1\}\}$ and $\Psi = \{\varphi : Y \to [0, 1]\}$.
3. A subgraph Λ is represented by $\psi_0 \in \Psi_0$ denoted by $\Lambda \sim \psi_0$.

Subgraph permutants

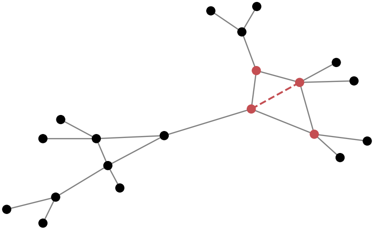
Using subgraphs, we introduced the following set that proves to be a **generalized permutant**. Given $\Lambda \sim \psi_0$ as a subgraph of $\Gamma \sim \varphi$ is the set of functions from Y to X of the form:

$$H_\Lambda^\varphi = \{h\colon Y \to X \mid (\varphi \circ h)(\{w_i, w_j\}) = 1 \iff \psi_0(\{wi, wj\}) = 1 \\ \{w_i, w_j\} \in Y, h \text{ is injective}\}$$

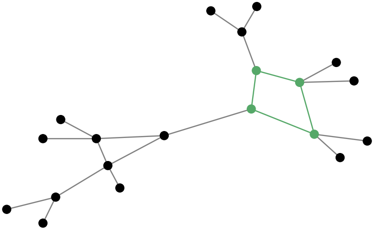
Proposition (H_Λ^φ is a generalized permutant)

Any set H_Λ^φ is a generalized permutant for $T\colon G_\Gamma \to \{id_Y\}$.

Subgraph permutants (cont.)



(a) $\Gamma_1 \quad H_\Lambda^{\varphi_1} = \emptyset$



(b) $\Gamma_2 \quad |H_\Lambda^{\varphi_2}| = 8$

Figure: Considering Λ as a square, the set $H_\Lambda^{\varphi_1}$ is empty since there is no strict embedding of Λ to Γ_1. On the other hand, there are 8 strict embeddings of Λ into Γ_2 (i.e. as many as the self-isomorphisms of Λ).

Subgraph GENEOs

We proved this new proposition in order to admit the case in which the permutant is φ dependent:

Proposition

The operator F_Λ defined as

$$F_\Lambda(\varphi) = \frac{1}{n_\Lambda |G_\Lambda|} \sum_{h \in H_\Lambda^\varphi} \varphi \circ h$$

is a linear GENEO between the pairs (Φ, G), $(\Psi, \{id_Y\})$ with respect to the trivial homomorphism $T : G \to \{id_Y\}$, where n_Λ is the number of occurrences of Λ in a complete graph with N nodes.

GENEO based isomorphism test

Theorem (GENEO-based isomorphism test)

Given two graphs $\Gamma_1 \sim \varphi_1$ and $\Gamma_2 \sim \varphi_2$, for every choice of $p \geq 1$ and $\Lambda_1, \ldots, \Lambda_p$ the map:

$$F_p(\varphi_1, \varphi_2) = \max_{i=1}^{p} \left\| \frac{1}{2} |F_{\Lambda_i}(\varphi_1) - F_{\Lambda_i}(\varphi_2)| \right\|_\infty$$

is a GENEO from $(\Phi \times \Phi, G \times G)$ to $(\mathbb{R}, \{id_S\})$ with respect to the trivial homomorphism $T \colon G \times G \to \{id_S\}$ ($\mathbb{R}$ is identified with the set of real-valued functions on a singleton S). Moreover, the followings hold true:

1. *If Γ and Γ_2 are isomorphic then $F_p(\varphi_1, \varphi_2) = 0$.*
2. *$F_p(\varphi_1, \varphi_2) \neq 0$ then Γ and Γ_2 are not isomorphic.*

GENEO based isomorphism test (cont.)

Thus the function $\mathbb{1}(F_p(\varphi_1, \varphi_2))$ is an inexact isomorphism test,
if $\mathbb{1}(F_p(\varphi_1, \varphi_2)) = 1$ then the two graphs are surely non isomorphic.

Thus, choosing wisely the subgraphs $\Lambda_1, \ldots, \Lambda_p$, we can hope to obtain a **simple** and **efficient** isomorphism test.

But how do we choose $\Lambda_1, \ldots, \Lambda_p$?

r-regular graphs: a case studio

We choose to test our method using r-regular graphs, which are usually hard to distinguish for methods like the Weisfeiler-Leman test (1-WL) and (Messagge Passing) Graph Neural Networks (GNNs).

Definition (r-regular graph)

A graph is called r-regular for $r \geq 2$ if every node has degree exactly equal to r.

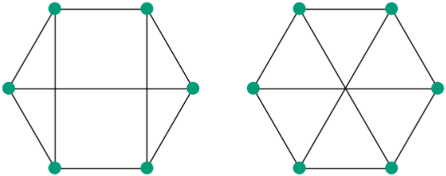


Figure: 3-regular non isomorphic graphs indistinguishable for 1-WL.

Initial pool of Λ_j

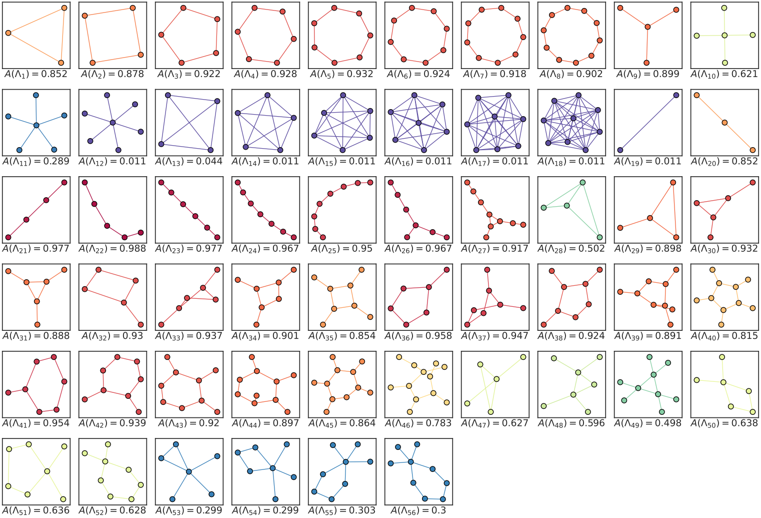


Figure: Initial pool of Λ_j with accuracies.

Selection of Λ_j

Algorithm 1: Forward Selection

Data: $\Lambda_1, \ldots, \Lambda_p$, $((\varphi_1, \varphi_2)_k)_{k=1}^m$

Result: $S \subseteq \{\Lambda_1, \ldots, \Lambda_p\}$

$l \leftarrow 1$; $i_l \leftarrow \underset{t \in \{1,\ldots,p\}}{\arg\max}\, A\Big(F_{\{j\}}\Big)$;

while $\max\limits_{t \in \{1,\ldots,p\} \setminus \{i_1,\ldots,i_l\}} A\Big(F_{\{i_1,\ldots,i_l\} \cup \{t\}}\Big) > A\Big(F_{\{i_1,\ldots,i_l\}}\Big)$ **do**

$\quad i_{l+1} \leftarrow \underset{t \in \{1,\ldots,p\} \setminus \{i_1,\ldots,i_l\}}{\arg\max} A\Big(F_{\{i_1,\ldots,i_l\} \cup \{t\}}\Big)$;

end

$S = \{\Lambda_{i_1}, \ldots, \Lambda_{i_l}\}$;

Selection of Λ_j (cont.)

After running the selection algorithm we found that maximum accuracy could be achieved with only three operators F_{22}, F_{21} and F_4:

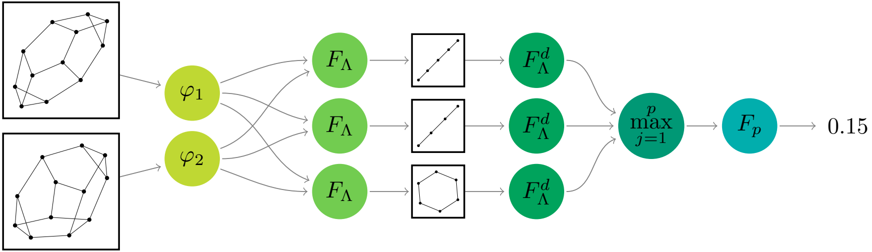


Figure: Architecture of the subgraph permutant GENEO-based model.

Comparison

Finally we tested the network against other SOTA methods, both exact and inexact, considering both **expressivity** and **time efficiency**.

Table: Average results for $r = 3$ with timeout of 10s.

Method	N					
	100		1000		10000	
	Time	Acc	Time	Acc	Time	Acc
GENEO-1	0.014	0.990	0.138	0.986	1.338	0.984
GENEO-2	0.019	0.996	0.193	0.994	1.866	0.994
GENEO-3	0.051	**1.000**	0.539	**1.000**	5.202	**1.000**
NTX-FASTER	0.000	0.000	0.000	0.000	0.003	0.000
NTX-FAST	0.001	0.719	0.011	0.758	0.128	0.754
NTX-COULD	0.003	0.719	0.114	0.758	9.747	0.754
NTX-IS	4.719	0.691	10.000	0.000	10.000	0.000
1-WL	0.001	0.000	0.008	0.000	0.090	0.000
2-WL	7.134	0.000	10.000	0.000	10.000	0.000
3-WL	10.040	0.000	10.000	0.000	10.000	0.000

Comparison (cont.)

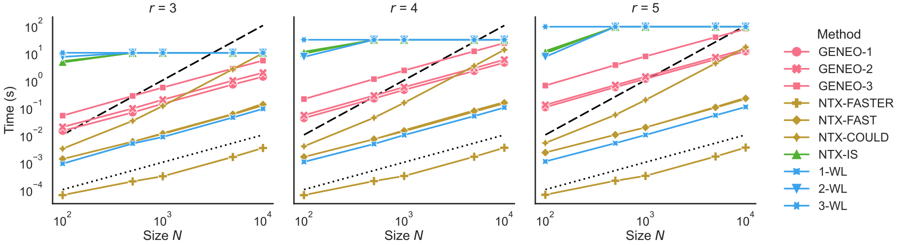


Figure: Time complexity of the models compared. Results are in double logarithmic scale.

Take home message

[Generalized] permutants and GENEOs are versatile tools to be used in ML and AI applications when some prior knowledge is available. In this talk they were employed to learn an isomorphism test based on the search of subgraphs.

Thank you for your attention!

References I

- G. Bocchi, S. Botteghi, M. Brasini, *et al.*, "On the finite representation of linear group equivariant operators via permutant measures," *Annals of Mathematics and Artificial Intelligence*, vol. 91, no. 4, pp. 465–487, 2023, ISSN: 1012-2443. DOI: 10.1007/s10472-022-09830-1.
- F. Ahmad, M. Ferri, and P. Frosini, "Generalized Permutants and Graph GENEOs," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1905–1920, 2023. DOI: 10.3390/make5040092.
- G. Bocchi, M. Ferri, and P. Frosini, "A novel approach to graph distinction through GENEOs and permutants," preprint at arXiv, 2024. arXiv: 2406.08045.