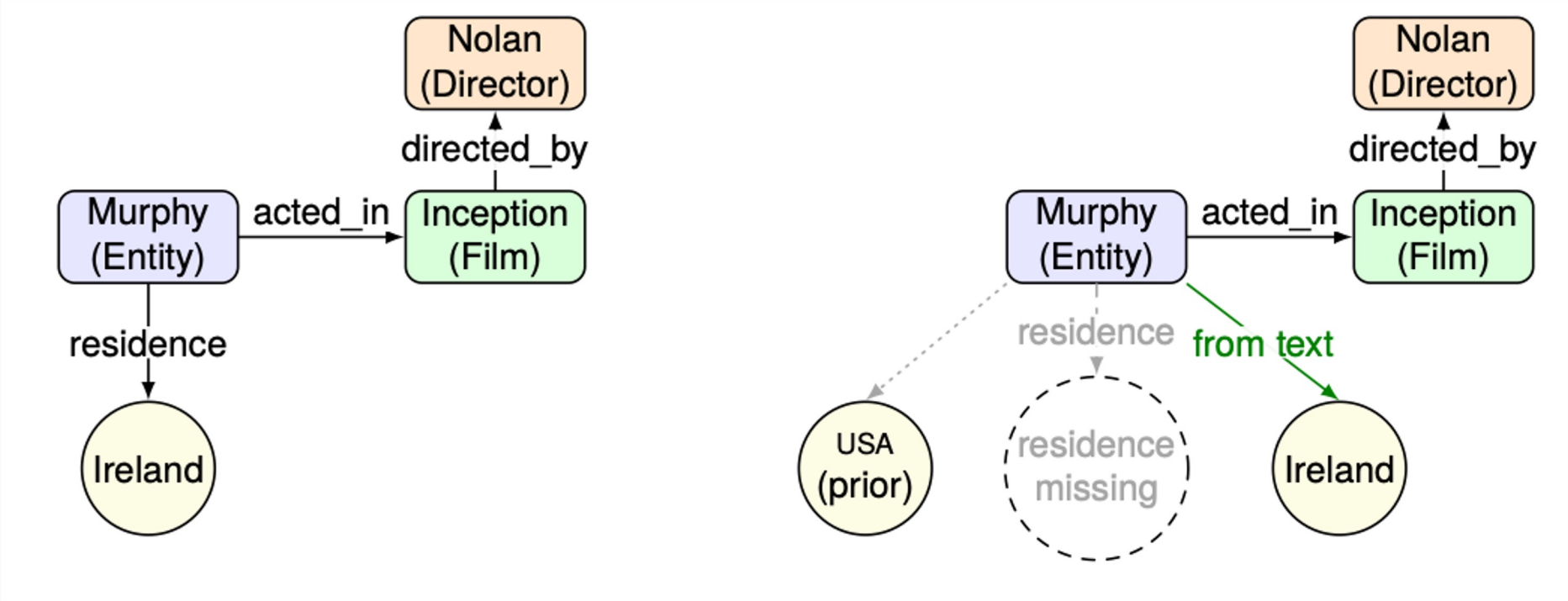
TOR VERGATA
UNIVERSITÀ DEGLI STUDI DI ROMA

LREC 20 Palma 26

Linguistic Initialization for Inductive Reasoning in Heterogeneous Knowledge Graphs

D. Pasquini, D. Croce, R. Basili

Structural equivalence



The "Initialization Gap" in GNNs

Most GNNs start with a 'Tabula Rasa'

- Nodes are often initialized with unmeaningful IDs
- Semantics must be reconstructed entirely from topology
- **The risk**: In sparse or noisy graphs, the model learns co-occurrence, not meaning

Limitations of Structure-Only Models

- **Hub Bias**: Models default to high-degree 'Popularity' nodes
- **Cold Start**: New entities are invisible to the message-passing mechanism
- **Missing Bridges**: Identifiers encode identity, not semantics

Our core philosophy

Language defines the starting state, structure constrains its evolution.

- Inject **semantics** at initialization
- Let the Language Model provide the **'prior'**
- let the Graph provide the **'evidence'** to refine that prior

Intrinsic vs. Contextual Meaning

- **Intrinsic View** t_v^{node} : The node's name or definition
- **Contextual View** t_v^{ctx} : Verbalizing the local neighborhood

$$\mathbf{x}_v = f_{\text{text}}\left(t_v^{\text{node}} || t_v^{\text{ctx}}\right)$$

Example: "*Murphy is an Irish Actor who acted in the movie Inception directed by Christopher Nolan [...]*"

We combine the node label with a verbalized summary of who it's connected to. This ensures even sparse nodes have a 'semantic footprint' from the start.

GNN Architecture

Moving from Text to Heterogeneous Feature Space

$$\mathbf{x}_v = f_\theta(\mathcal{T}(v)) \qquad \mathbf{h}_v^{(0)} = \mathrm{Drop}\Big(\mathrm{ReLU}\big(\mathbf{W}_{\phi(v)}\,\mathbf{x}_v + \mathbf{b}_{\phi(v)}\big)\Big)$$

Multi-Head Attention over Typed Edges:

$$\tilde{\mathbf{h}}_i^{r,k} = \mathbf{W}_{r,k}\mathbf{h}_i^{(l)} \qquad \alpha_{ij}^{r,k} = \mathrm{softmax}_{j\in\mathcal{N}_i^r}\Big(\mathrm{LeakyReLU}\big(\mathbf{a}_{r,k}^\top[\tilde{\mathbf{h}}_i^{r,k}\|\tilde{\mathbf{h}}_j^{r,k}]\big)\Big)$$

Preserving Semantics in Sparse Regions:

$$\mathbf{h}_i^{(l+1)} = \begin{cases} \mathrm{Drop}\big(\mathrm{ReLU}(\mathbf{m}_i^{\mathrm{tot}})\big) + \mathbf{h}_i^{(l)} & \text{if } \mathcal{N}_i \neq \emptyset, \\ \mathbf{h}_i^{(l)}, & \text{otherwise} \end{cases}$$

Decoding Typed Triples (u,r,v) $\qquad s_{uv}^r = \sigma(\mathbf{W}_{r,2}\ \mathrm{ReLU}\Big(\mathbf{W}_{r,1}[\mathbf{h}_u^{(L)}\|\mathbf{h}_v^{(L)}]\Big))$

GNN Architecture

Challenges of Heterogeneous Supervision

- Skewed relation distributions
- Hub nodes (e.g., "USA") causing memory overflow
- Parameter explosion with |R| relations

Bounded supervision

- **Positive Edge Capping**

$$\hat{M}_r = \min(M_r, K_{\text{cap}})$$

where K_{cap} is the maximum number of supervised positive edges retained for any single relation

- **Stochastic Relation Activation**

$$|\mathcal{R}_{\text{active}}| \leq R_{\max}$$

where R_{cap} is the maximum number of distinct relations considered per step

Semantic bucketing

- **Partition** the **relation space**:
  - Frequent head relations: $\mathbf{R_{head}}$
  - Rare tail relations: $\mathbf{R_{tail}}$
- R_{tail} are grouped into semantic buckets by their domain: $\mathbf{B_{tail}}$

1. Prevent the parameter explosion
2. Model shares strength across rare predicates semantically similar
3. Prevent the overfitting typical of sparse relations

Experimental setup

Evaluation on **500** Subgraphs

	Train	Validation	Test	Full
Counts				
# Graphs	350	75	75	500
# Nodes	479126	119374	105422	703922
# Edges	2551151	624008	551791	3726950

	In	Out	Total	Conn.%
Global	5.29	5.29	10.59	100
Entity	4.10	5.26	9.36	100
Type	**11.85**	5.49	17.34	100

Robustness to Imbalance

Setup/Init	Random Initialization				Text Initialization			
	Acc	**P**	**R**	**F1**	**Acc**	**P**	**R**	**F1**
Balanced	.95	.94	.95	.95	.97	.97	.98	**.97**
Weak	.97	.84	.85	.84	.98	.90	.92	**.91**
Realistic	.99	.55	.67	.60	.99	.70	.75	**.72**

The "Hard Set" Challenge

Ratio	Init	S+	S-	Acc	P	R	F1
1:1	Random	6,340	6,340	0.81	0.98	0.64	0.77
	Text-grounded	6,265	6,265	0.92	0.99	0.85	0.92
1:9	Random	6,340	57,060	0.96	0.88	0.64	0.74
	Text-grounded	6,265	56,385	0.98	0.92	0.85	0.89
1:100	Random	6,340	634,000	0.99	0.4	0.64	0.49
	Text-grounded	6,265	626,500	0.99	0.53	0.85	0.65

when the graph has nothing to tell you, language does

Predicate discrimination

Initialization	Category	P	R	F1	N
Random	Head	0.99	0.68	0.81	3,412
Random	Bucket	0.99	0.97	0.98	251
Text	Head	0.99	0.88	0.93	3,278
Text	Bucket	0.99	0.92	0.96	244

- For specific 'Head' relations, the Random model gets confused
- Linguistic grounding provides the fine-grained discrimination needed to pick the right relation

Qualitative insight

Feature	Cold-Start Node (Sparse)	Hub Node (Dense)
Connectivity	2 training neighbors	28 training neighbors
Key Relation	medicine.disease.risk_factors	authored_by, topic_type
Random Init	Fail: Not in Top 10 (P≈0.24)	Success: Rank 1 (P>0.99)
Text-Grounded	Success: Rank 1 (P=1.00)	Success: Rank 1 (P≈1.00)

Conclusion

- Resolved Structural equivalence
- Eliminated Hub Bias
- Scalable, Lightweight Integration
- Semantics should not be an afterthought for GNNs.
- By grounding initialization in language, we move from popularity-based guessing to semantically-informed reasoning

Thank you

Complexity

$$\mathcal{O}\big(L\,(Nd^2 + Md)\big) + \mathcal{O}\big(|\mathcal{R}_{\text{active}}|K_{\text{cap}}(1+\rho)d^2\big)$$

- The encoder cost (left) depends on the graph size (N,M) to preserve structural fidelity.
- The supervision cost (right) is strictly bounded by our hyperparameters $\mathbf{K_{cap}}$ and $\mathbf{R_{max}}$
- As the graph grows, the supervision overhead stays constant

Thank you