

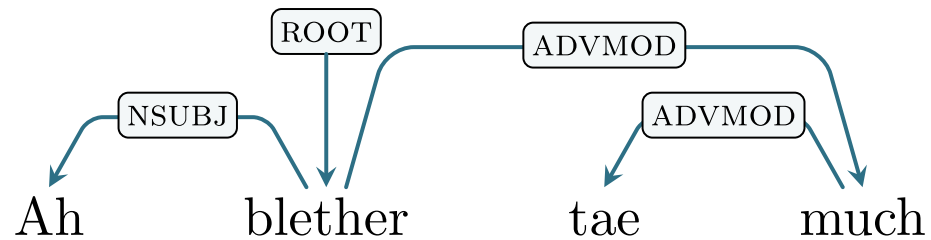
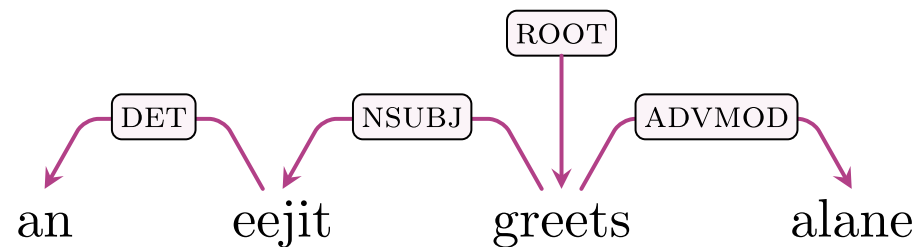
# REPLICATING AND EXTENDING "*BECAUSE THEIR TREEBANKS LEAK*": GRAPH ISOMORPHISM, COVARIANTS, AND PARSER PERFORMANCE

---

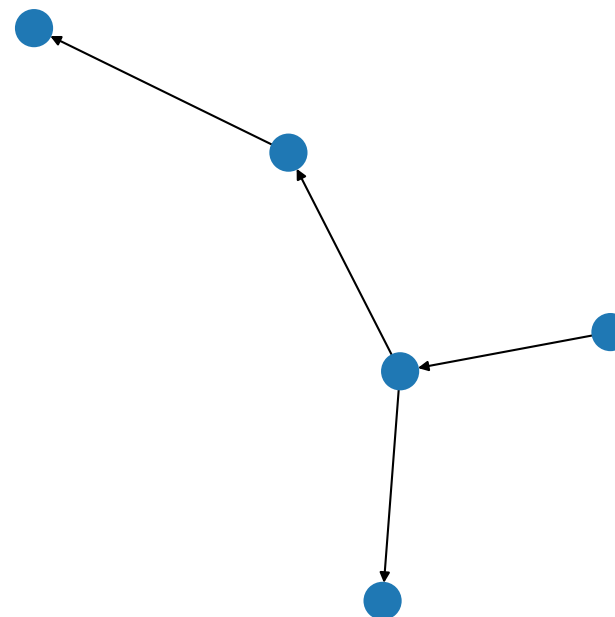
Mark Anderson, Anders Søgaard, Carlos Gómez Rodríguez

# DIRECTED UNLABELLED GRAPH ISOMORPHISM (DUG)

## Dependency trees



## Directed unlabelled graph



# DATA AND SYSTEMS

---

## All use Universal Dependency treebanks

- CoNLL18 - best performing systems, 33 treebanks
- UDPipe 1.2 - transition-based NN, 94 treebanks (UD v2.5)
- UDPipe 2.0 - graph-based NN, competitive, 90 treebank (UD v2.7)

# EXPLAINED VARIANCE FROM LR FIT

	Original			10 Seeds		
	CoNLL18	UDPipe 1.2	UDPipe 2.0	CoNLL18	UDPipe 1.2	UDPipe 2.0
Training size	0.014	0.100	0.060	-0.019	-0.346	-0.005
+ DUG	0.228	0.061	0.097	-0.004	-0.553	0.091
+ $\langle L_{\text{test}} \rangle$	0.195	0.169	0.146	-0.007	-0.370	0.140
All	-0.078	0.157	0.086	-0.413	-0.138	0.106

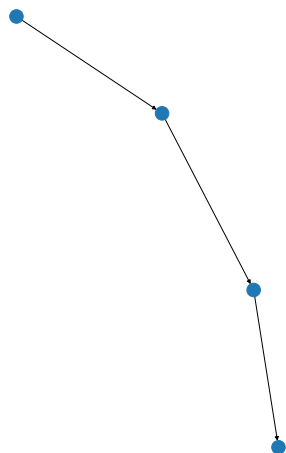
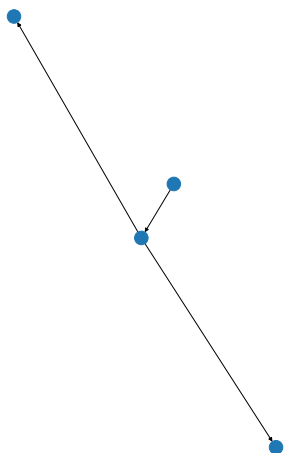
# BETTER WITH LOG

	CoNLL18	UDPipe 1.2	UDPipe 2.0
<b>log-size</b>	0.055	0.319	0.126
<b>+ DUG</b>	0.132	0.410	0.277
<b>+ <math>\langle L_{\text{test}} \rangle</math></b>	0.106	0.452	0.294
<b>All</b>	-0.184	0.412	0.229

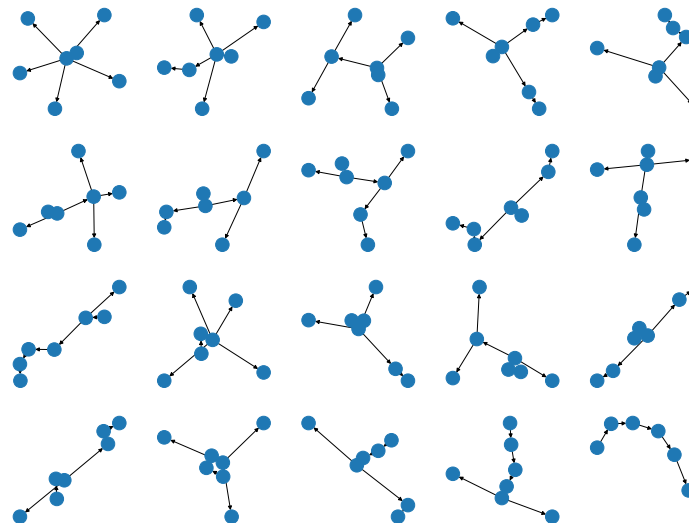
10 seeds CV (3 splits)

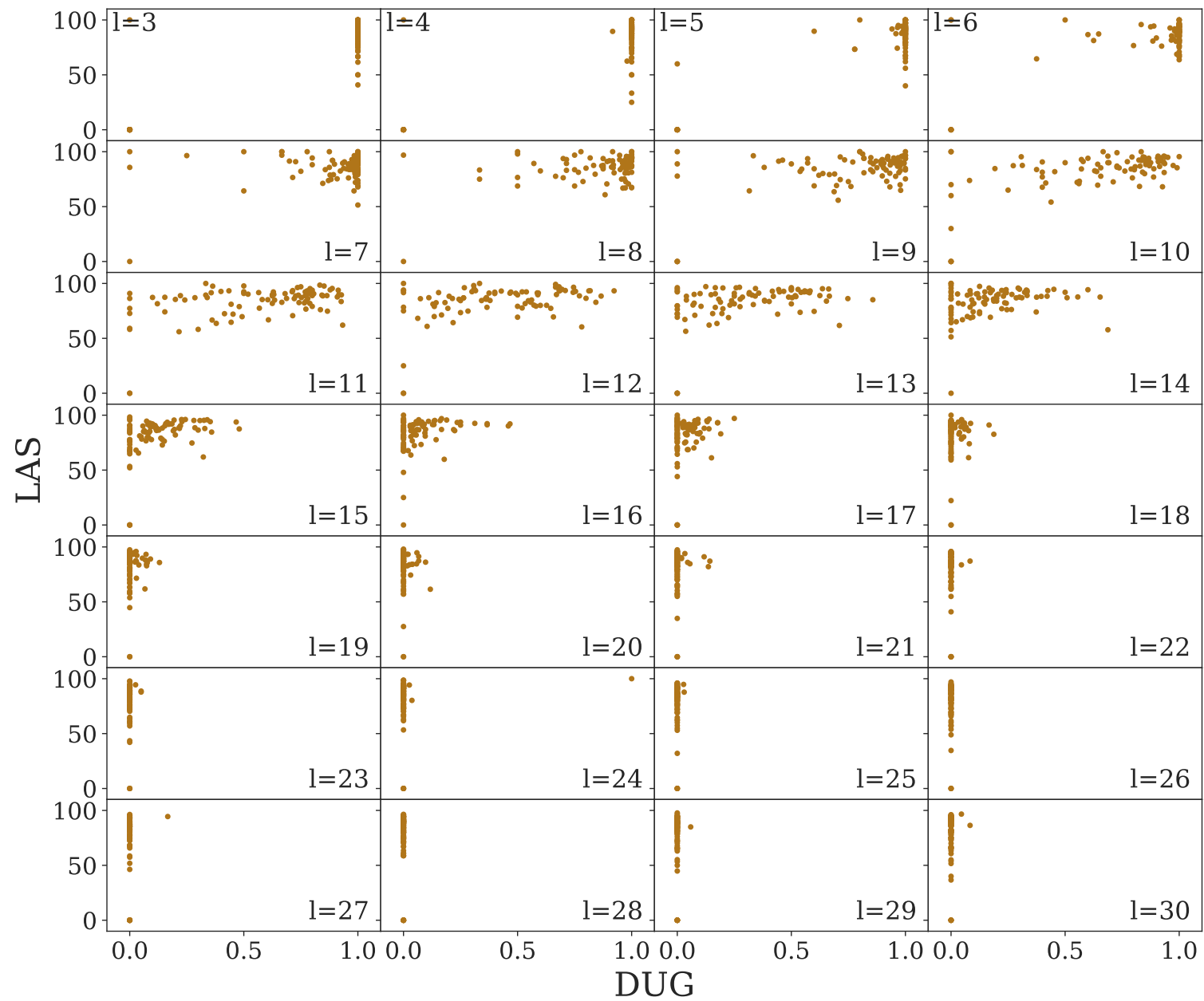
# ISOMORPHISMS

Sentences with 3 tokens



Sentences with 6 tokens





# DUG COEFFICIENTS

	CoNLL18	UDPipe 1.2	UDPipe 2.0
<b>LAS</b>	-0.13 (p=0.458)	-0.13 (p=0.213)	-0.18 (p=0.083)
<b>Training size</b>	0.44 (p=0.011)	0.42 (p<0.001)	0.46 (p<0.001)
<b><math>\langle L_{\text{test}} \rangle</math></b>	-0.96 (p<0.001)	-0.91 (p<0.001)	-0.92 (p<0.001)

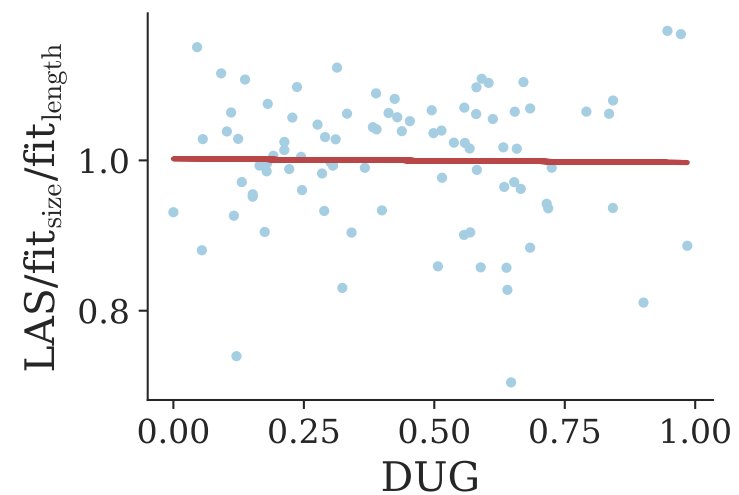
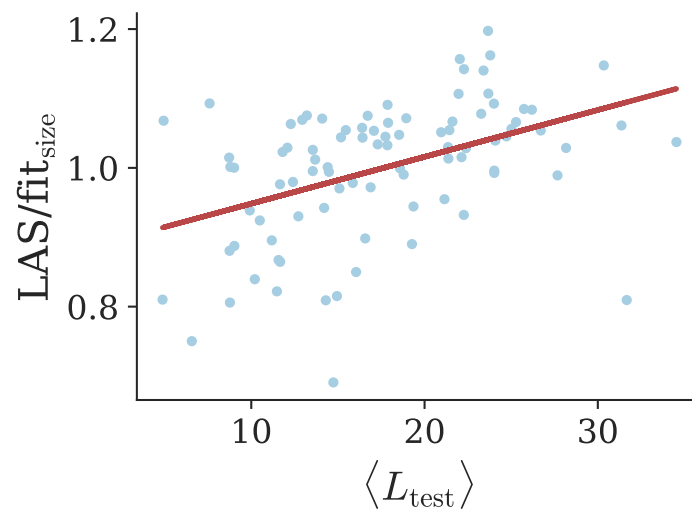
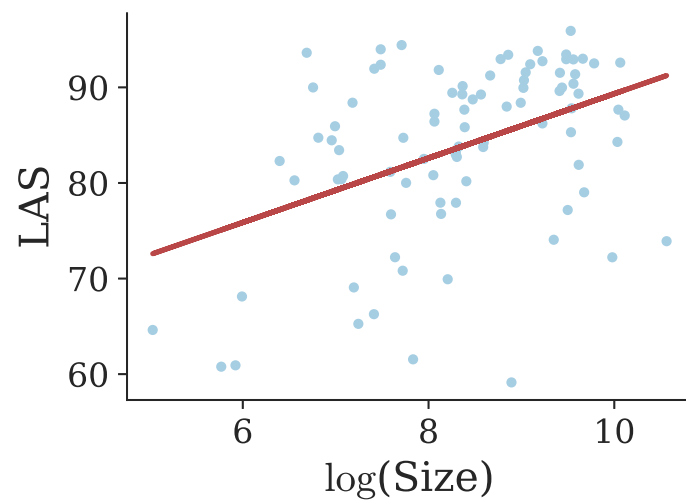
Spearman's  $\rho$



# PARTIAL COEFFICIENTS

	CoNLL18	UDPipe 1.2	UDPipe 2.0
<b>DUG</b>	-0.13 (p=0.458)	-0.13 (p=0.213)	-0.18 (p=0.083)
<b>Training size</b>	-0.44 (p=0.010)	-0.50 (p<0.001)	-0.46 (p<0.001)
<b><math>\langle L_{\text{test}} \rangle</math></b>	0.18 (p=0.329)	-0.13 (p=0.213)	0.21 (p=0.049)
<b>both</b>	-0.27 (p=0.126)	0.01 (p=0.915)	-0.12 (p=0.245)

# BACKGROUND REMOVAL



# CONTROLLED EXPERIMENT

- Find treebanks with sufficient number of sentences with 12 tokens (>1200).
- Create unique splits of 1000 training instances and 200 test.
- Train UDPipe 1.2 models for each.
- Measure LAS and DUG for each split.
- Spearman's  $\rho$ : 0.82 ( $p < 0.001$ )

$N_{\text{train\_trees}}=1000$ ,  $N_{\text{test\_trees}}=200$ , Sentence Length=12

