Randomized Deep Structured Prediction
for Discourse-level Processing

Manuel Widmoser*[1], Maria Leonor Pacheco*[2], Jean Honorio[2], Dan Goldwasser[2]

[1] University of Salzburg, Austria

[2] Purdue University, USA

*equal contribution

Problems

- Many discourse-level tasks require accounting for longer texts
  [instead of just sentences]
- Learning highly expressive models is challenging
  [deep structured prediction]
- Constrained inference comes at high computational cost
  [the more dependencies, the higher the cost]

Contributions

- Combine expressive representations and inference
- Substitute exact inference by randomized inference
- Explore two challenging structured prediction problems:
  1. Argument Mining
  2. Stance Prediction

Argument Mining

essay

Living and studying overseas

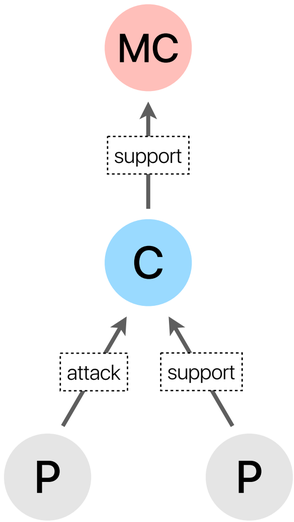
It is every student's desire to study at a good university and experience a new environment. While some students study and live overseas to achieve this, some prefer to study home because of the difficulties of living and studying overseas. In my opinion, one who studies overseas will gain many skills throughout this experience for several reasons.

First, living and studying overseas is an irreplaceable experience when it comes to learn standing on your own feet. One who is living overseas will of course struggle with loneliness, living away from family and friends but those difficulties will turn into valuable experiences in the following steps of life. Moreover, the one will learn living without depending on anyone else. Also, employers are mostly looking for people who have international and language skills. Becoming successful in this study will give the student an edge in job market. Therefore, one who has studied and lived overseas will become more eligible for the job than his/her peers.

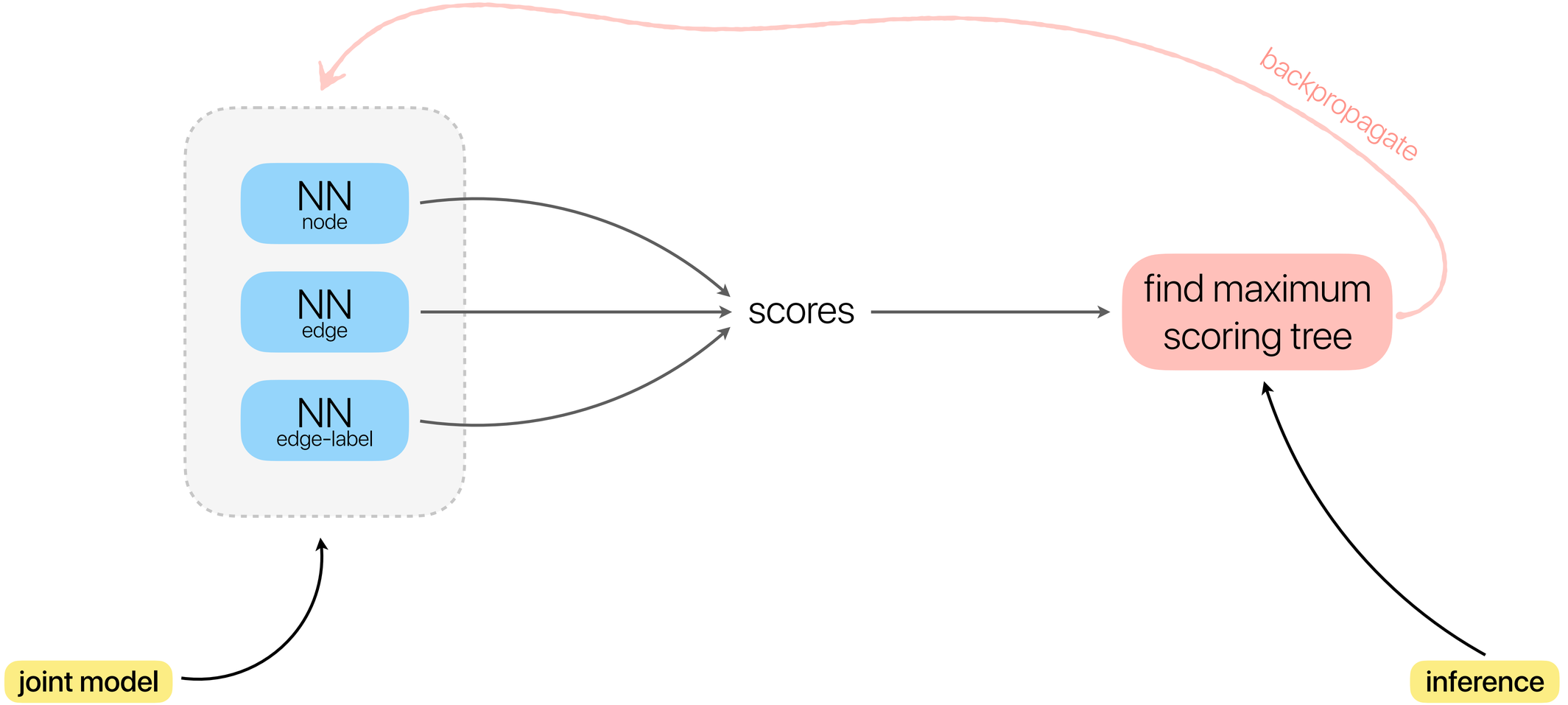
...

- $N = \{(N_1, \text{MC}), (N_2, \text{C}), (N_3, \text{P}), (N_4, \text{P}), \ldots\}$
- $E = \{(E_1, (N_2, N_1)), (E_2, (N_3, N_2)), (E_3, (N_4, N_2)), \ldots\}$
- $L = \{(E_1, \text{support}), (E_2, \text{attack}), (E_3, \text{support}), \ldots\}$

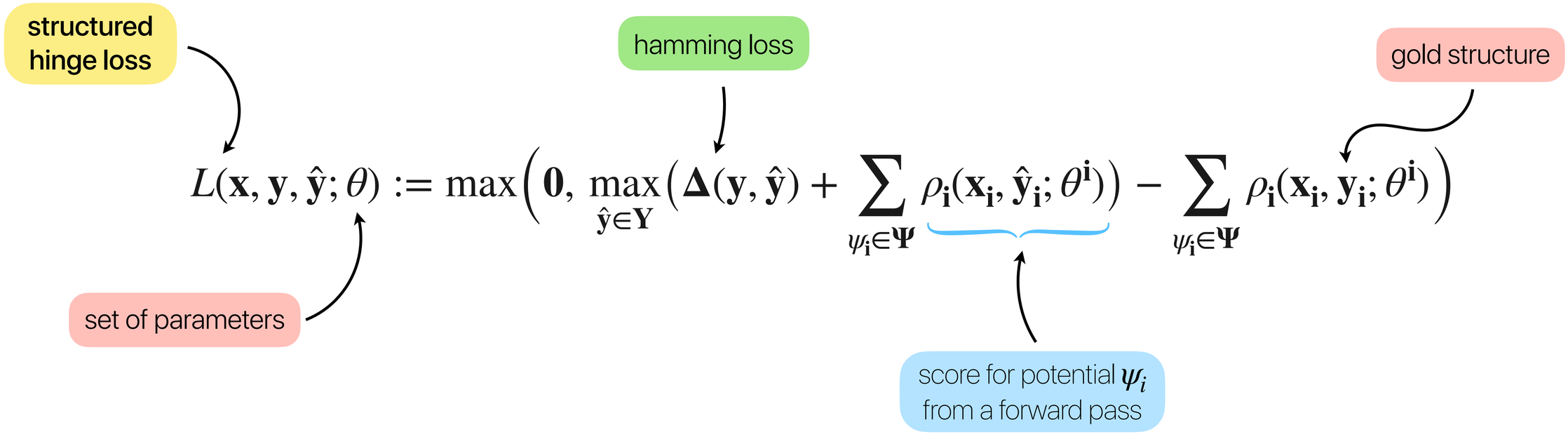
predict those



Learning


backpropagate
NN
node
NN
edge
NN
edge-label
scores
find maximum
scoring tree
joint model
inference

Objective Function



Inference

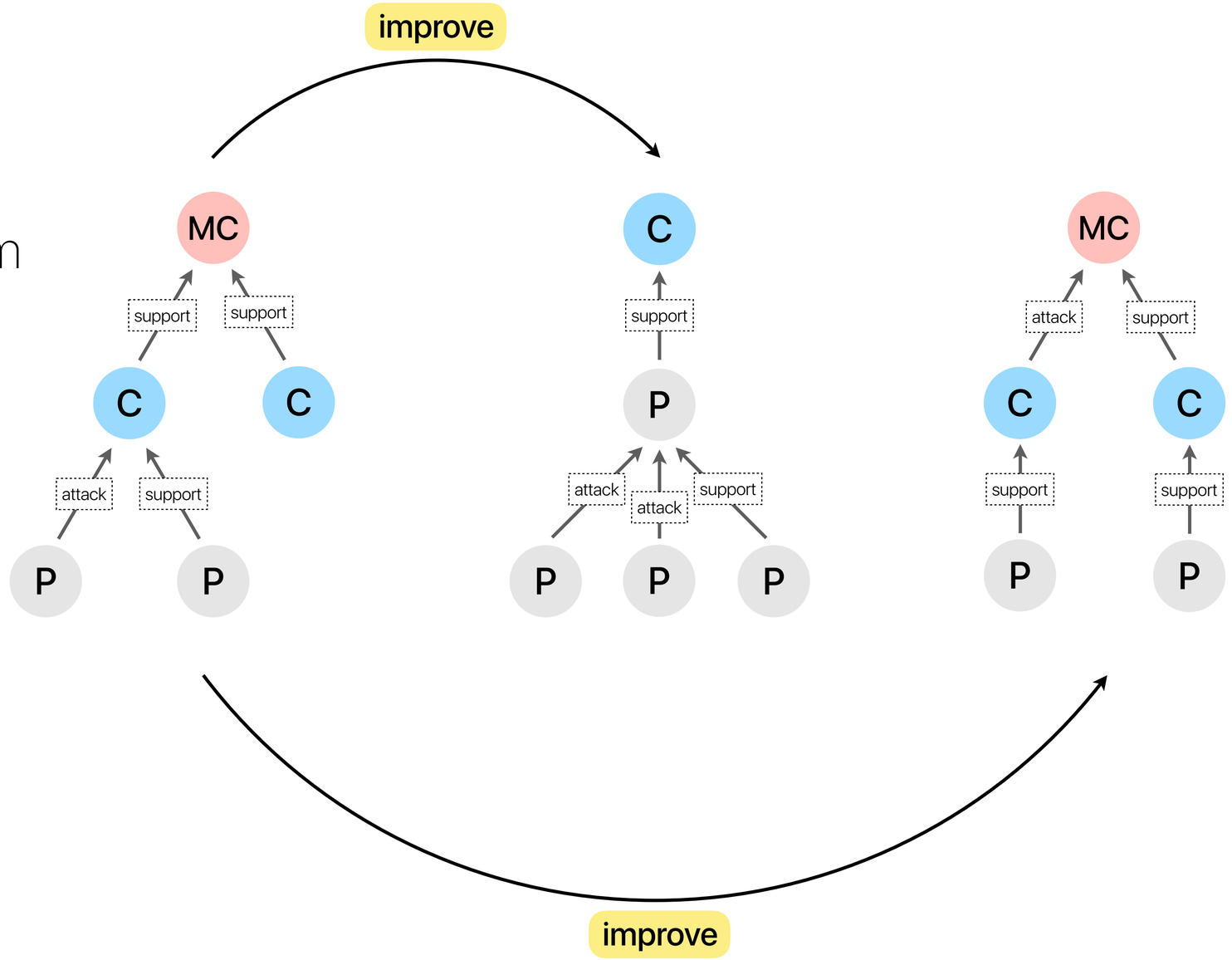
- ILP define inference problem as integer linear program
  [exact inference]
- AD^3 translate ILP into an AD^3 instance
  [approximate inference]
- Rand randomized inference

Can we get a competitive model?

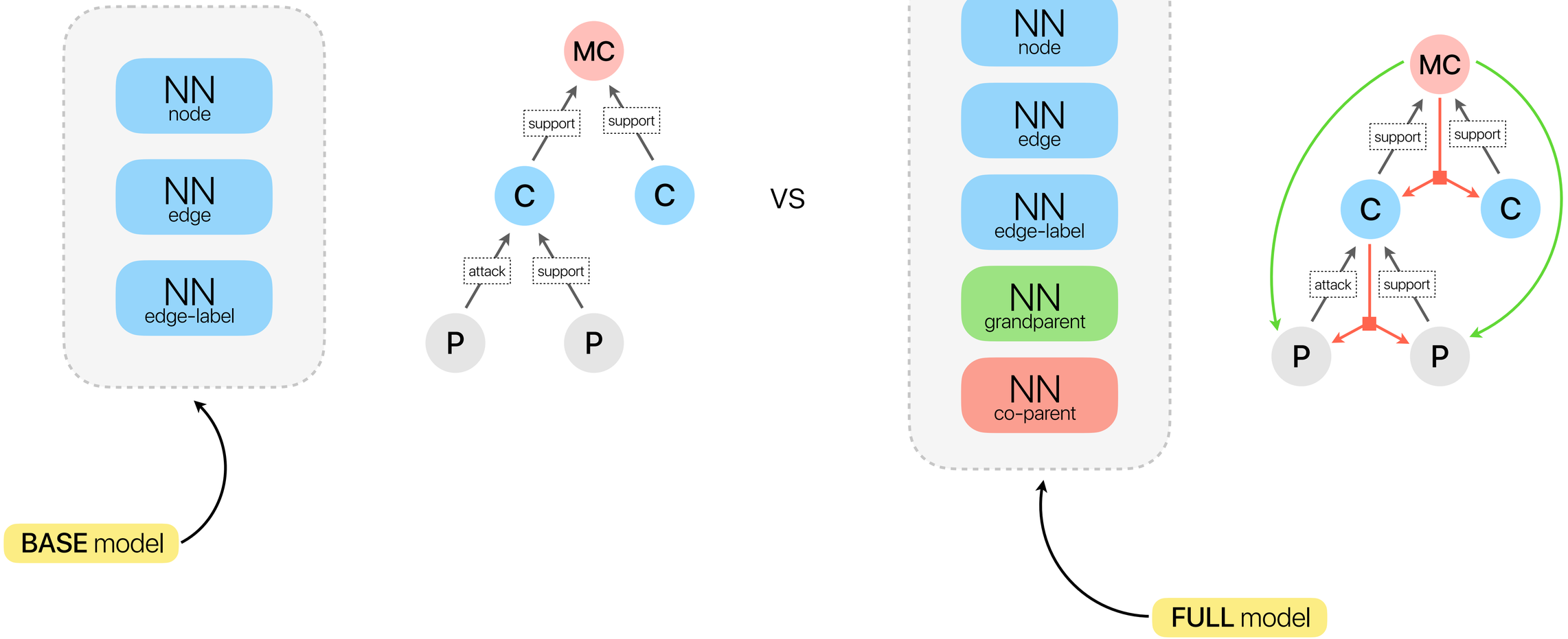
AD^3: Martins et al., 2015: http://dl.acm.org/citation.cfm?id=2789288

Randomized Inference

- Draw tree structure at random
- Label and score the tree
- Improve greedy locally
- Repeat

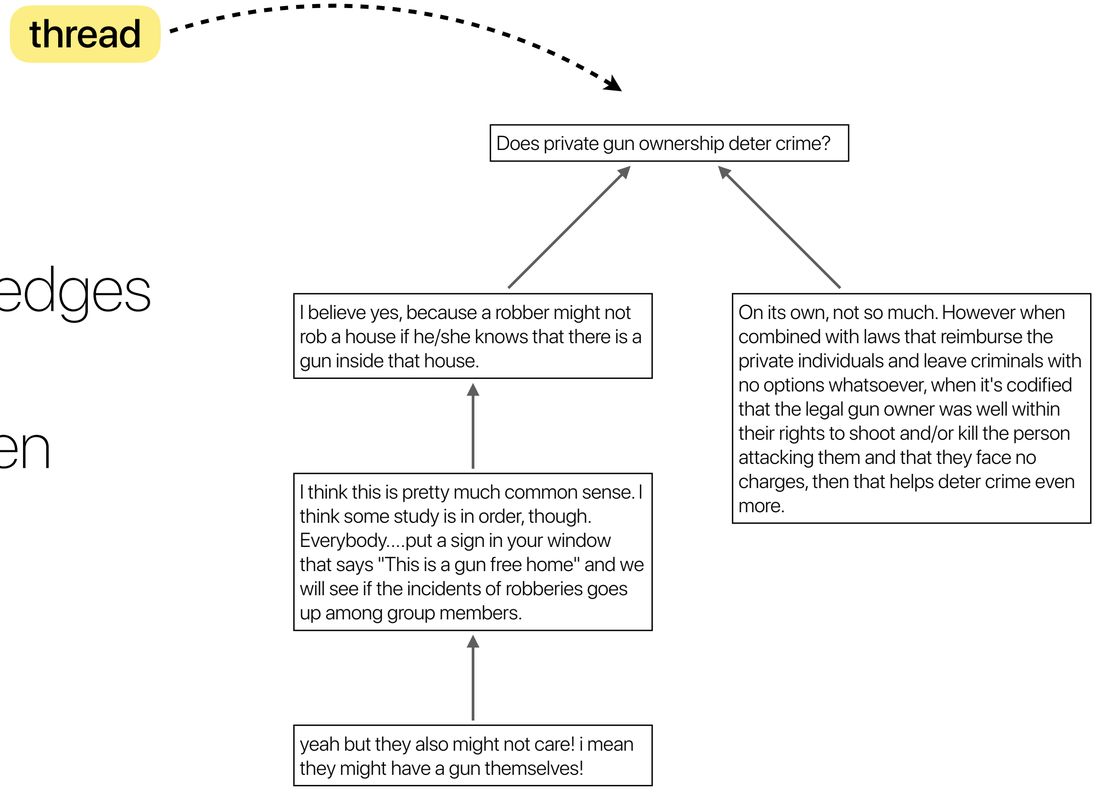


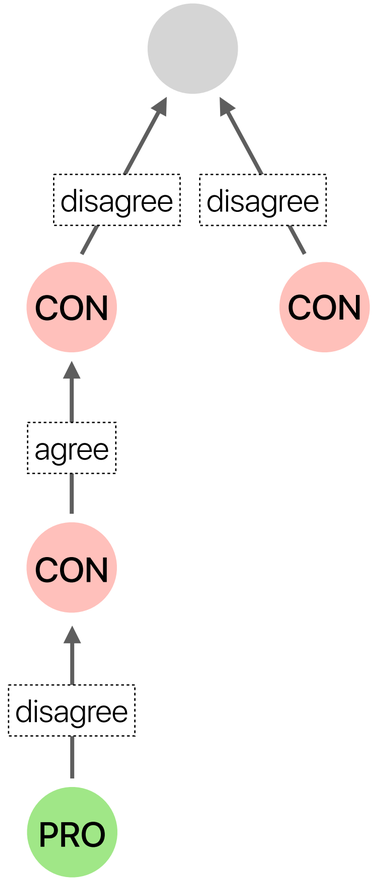
Base vs Full


NN
node
NN
edge
NN
edge-label
BASE model
MC
support
support
C
C
attack
support
P
P
vs
NN
node
NN
edge
NN
edge-label
NN
grandparent
NN
co-parent
FULL model
MC
support
support
C
C
attack
support
P
P

Stance Prediction

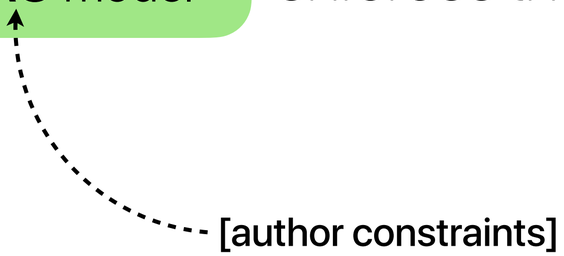
- Label nodes and edges
- Tree structure given
- Much larger trees





Base vs AC

- **BASE** model learns node and edge labels
- **AC** model enforces the same stance for all posts by the same author



Experiments: Argument Mining

		nodes	edges	edge-labels
BASE model	ILP	83.0	57.6	68.0
	AD^3	83.2	58.2	68.4
	Rand	82.8	58.4	68.4
FULL model	ILP	83.1	61.2	69.2
	AD^3	83.7	62.0	68.5
	Rand	83.8	62.6	68.4

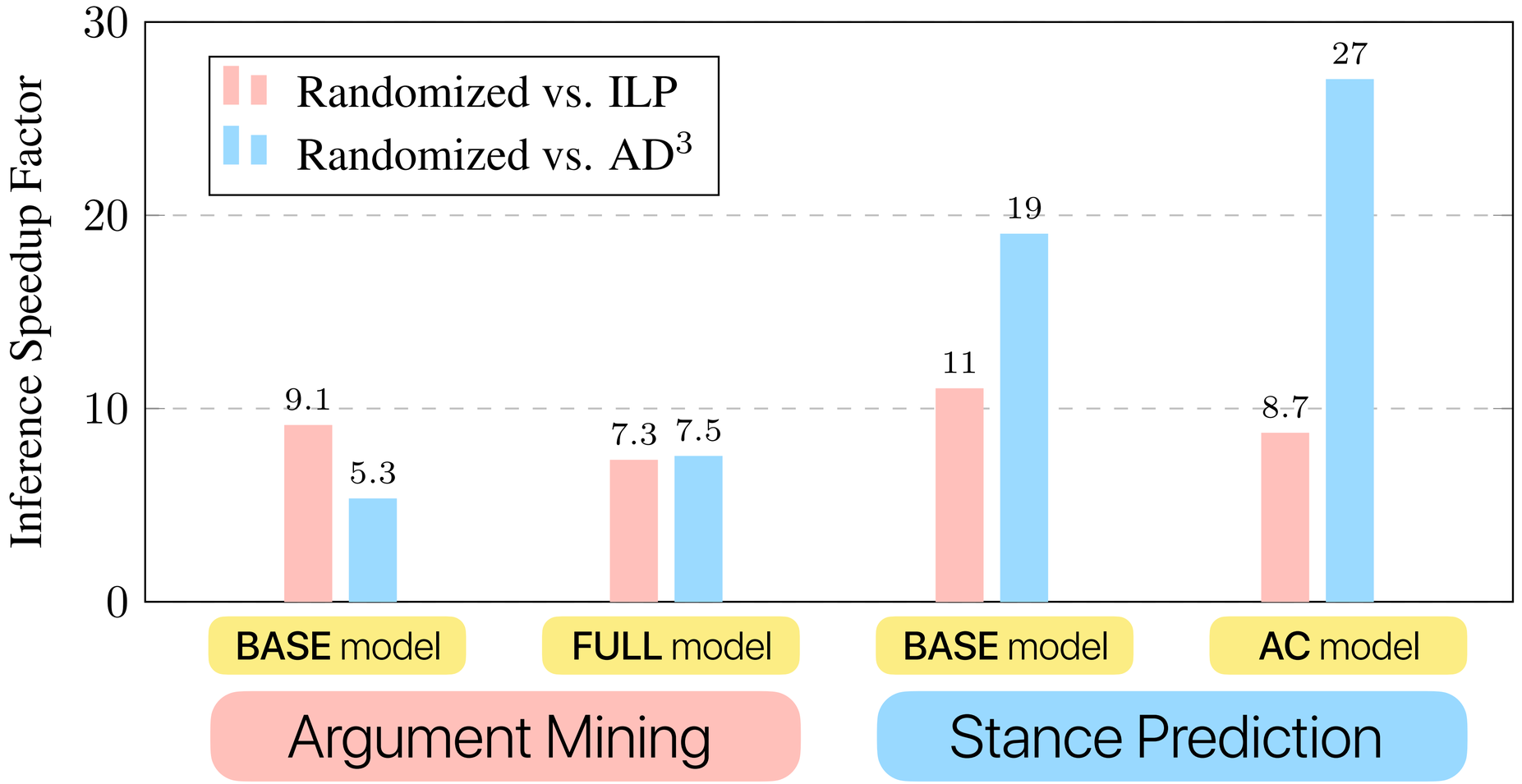
- UKP Dataset: Stab and Gurevych, 2017: https://doi.org/10.1162/COLI_a_00295
- reported scores: macro F1 for nodes and edge-labels, positive F1 for edges

Experiments: Stance Prediction

		nodes	edge-labels
BASE model	ILP	70.7	68.3
	AD^3	70.1	68.2
	Rand	68.9	69.0
AC model	ILP	82.3	81.5
	AD^3	81.8	79.5
	Rand	81.7	80.9

- 4Forums Dataset from the Internet Argument Corpus: Walker et al., 2012: https://www.aclweb.org/anthology/N12-1072/
- reported scores: averaged accuracies over 4 issues (abortion, evolution, gay marriage, gun control)

Inference Analysis



Conclusion

- We studied the effectiveness of randomized inference for deep structured prediction
- Positive results for two challenging discourse-level tasks
- Highly competitive results at a lower computational cost

[an efficient alternative to exact inference]