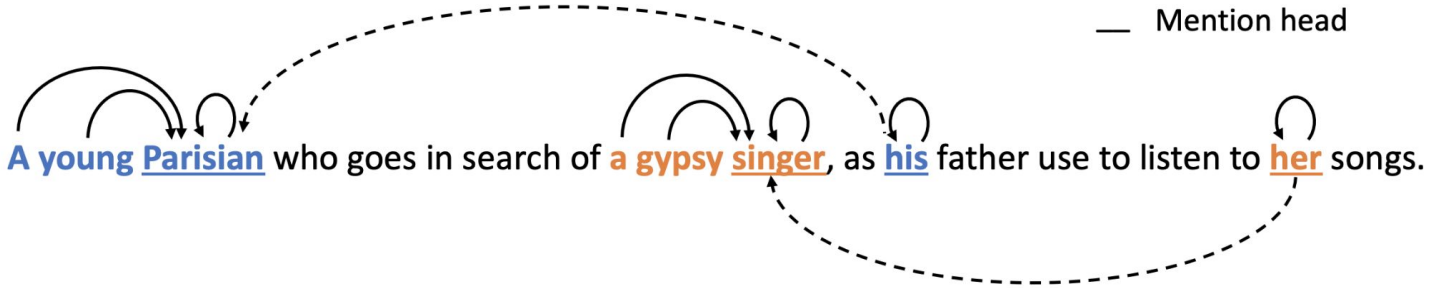
Transformers as Graph-to-Graph Models

Andrei Catalin Coman

Graphs in Transformers



- Transformers are essentially latent graph processing models, with sequences just being a special case.
- Attention weights are functionally equivalent to graph edges, since they determine how information propagates.
- Attention weights are computed from a set of vectors.
- Graphs are encoded in Transformers' set-of-vector embeddings.

James Henderson, Alireza Mohammadshahi, Andrei Coman, and Lesly Miculicich. 2023. Transformers as Graph-to-Graph Models. In Proceedings of the Big Picture Workshop, pages 93–107, Singapore, Singapore. Association for Computational Linguistics.

idiap RESEARCH INSTITUTE

"Transformers are latent graph models"

Graph2Graph Transformers (G2GT)

- Our G2GT makes this graph processing ability explicit.
- Observed graphs are input to the attention weight computations as relation embeddings.
- Predicted graphs are output with attention-like functions.
- Latent graphs are computed by pretrained Transformer weights.

	A	young	Parisian	...	a	gypsy	singer	,	as	his	use	her	songs
A	0	0	0	0	0	0	0	0	0	0	0	0	0
young	0	0	0	0	0	0	0	0	0	0	0	0	0
Parisian	1	0	0	0	0	0	0	0	0	0	0	0	0
...	0	0	0	0	0	0	0	0	0	0	0	0	0
a	0	0	0	0	0	0	0	0	0	0	0	0	0
gypsy	0	0	0	0	0	0	0	0	0	0	0	0	0
singer	0	0	0	0	1	0	0	0	0	0	0	0	0
,	0	0	0	0	0	0	0	0	0	0	0	0	0
as	0	0	0	0	0	0	0	0	0	0	0	0	0
his	0	0	2	0	0	0	0	0	0	1	0	0	0
use	0	0	0	0	0	0	0	0	0	0	0	0	0
to	0	0	0	0	0	0	0	0	0	0	0	0	0
her	0	0	0	0	0	0	2	0	0	0	0	1	0
songs	0	0	0	0	0	0	0	0	0	0	0	0	0

query-key | query-relation | relation-key

$$e_{ij} = \frac{1}{\sqrt{d}}\left[x_i \boldsymbol{W^Q}(x_j \boldsymbol{W^K})^T + x_i \boldsymbol{W^Q}(r_{ij} \boldsymbol{W_1^R})^T + r_{ij} \boldsymbol{W_2^R}(x_j \boldsymbol{W^K})^T\right]$$

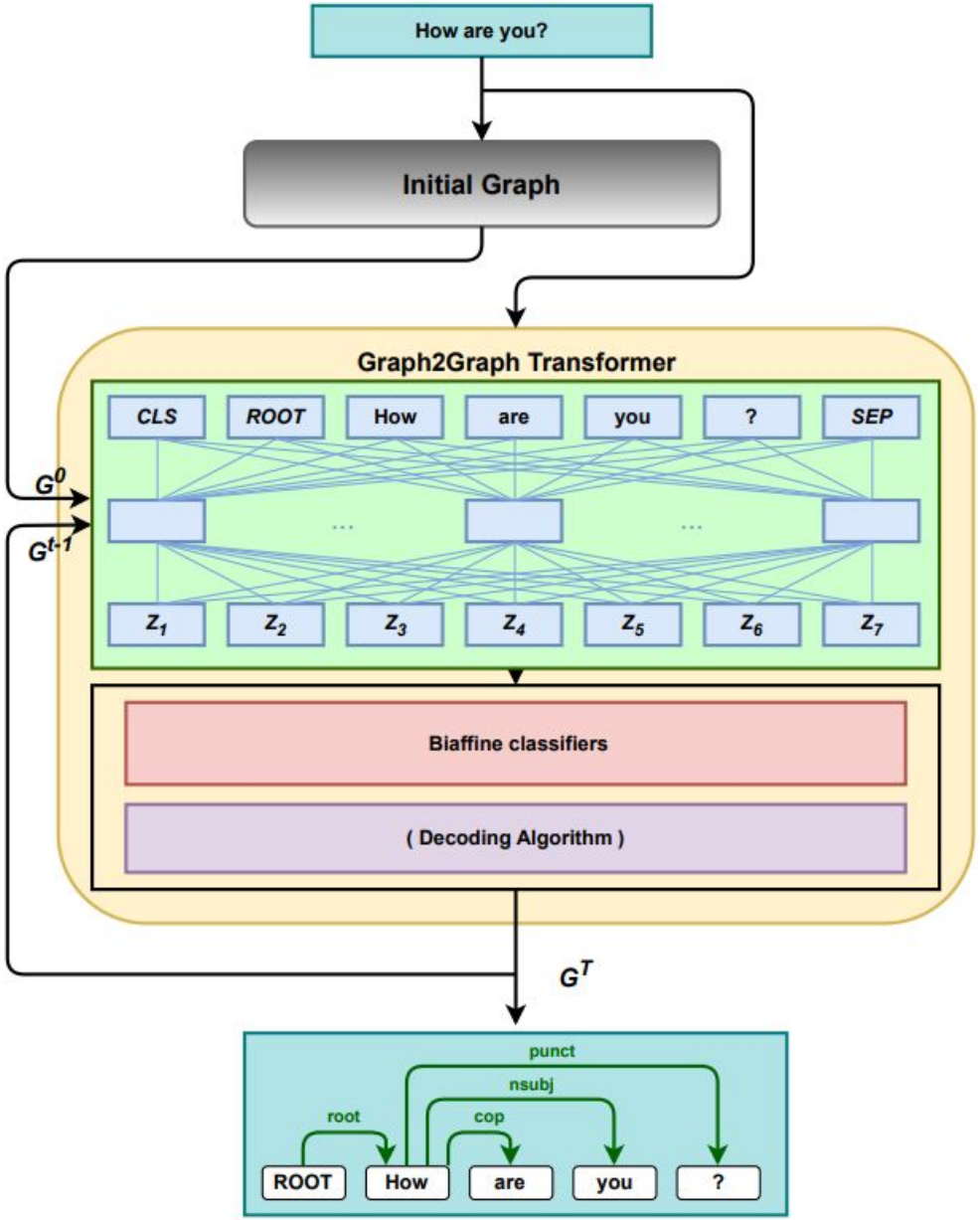
$$z_i = \sum_j \alpha_{ij}\underbrace{(x_j \boldsymbol{W^V} + r_{ij} \boldsymbol{W_3^R})}_{\text{value-relation}}$$

James Henderson, Alireza Mohammadshahi, Andrei Coman, and Lesly Miculicich. 2023. Transformers as Graph-to-Graph Models. In Proceedings of the Big Picture Workshop, pages 93–107, Singapore, Singapore. Association for Computational Linguistics.

"Transformers are latent graph models"

Efficient Global Graph Prediction

- G2GT encodes the observed graph, predicted graph, and latent graph in a single joint Transformer embedding.
- G2GT enables iterative refinement of the predicted graph to capture global patterns over the graph and text, without any bespoke pipeline or decoding strategy.



James Henderson, Alireza Mohammadshahi, Andrei Coman, and Lesly Miculicich. 2023. Transformers as Graph-to-Graph Models. In Proceedings of the Big Picture Workshop, pages 93–107, Singapore, Singapore. Association for Computational Linguistics.

idiap RESEARCH INSTITUTE