

Discovering social networks from free text

Jonathan Chang¹, Jordan Boyd-Graber², David M. Blei²

¹:Department of Electrical Engineering, Princeton University

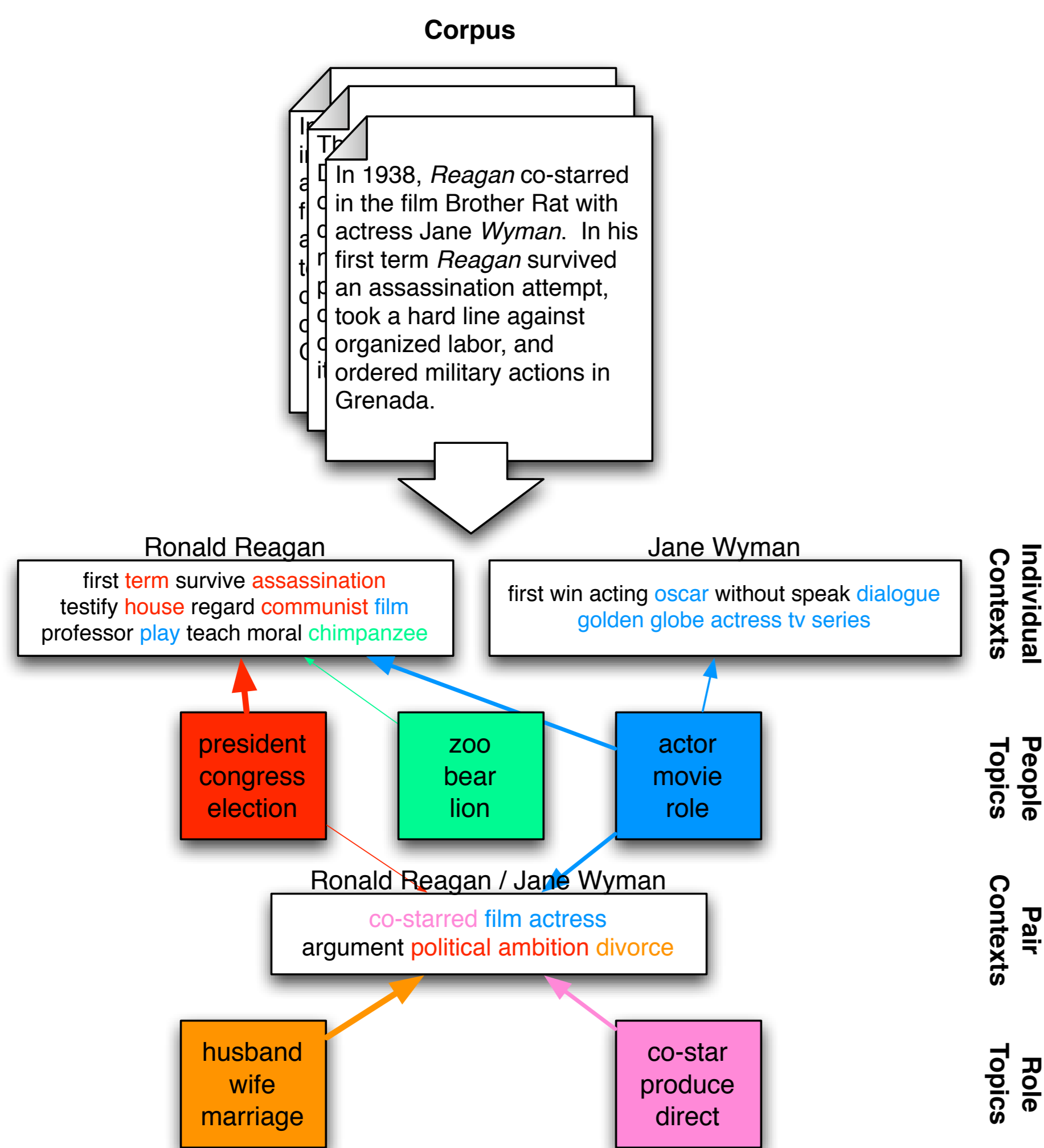
²:Department of Computer Science, Princeton University

Abstract

A lot of data are relational; they express relationships between pairs of entities (people, places, genes, businesses). Despite the inherently relational nature of data of interest, they are often not expressed relationally but rather as free text. To uncover this latent network structure from text, we posit a topic model that discovers common ways that relationships between entities are expressed. Using these automatically identified relationships, we are able to reconstruct social networks from free text.

Intuition

- ▶ We learn topics associated with individuals.
- ▶ Those topics can filter out person-specific words in pair contexts. The remaining words express something about the relationship between the entities.
- ▶ We train a topic model on these words to learn relationships.

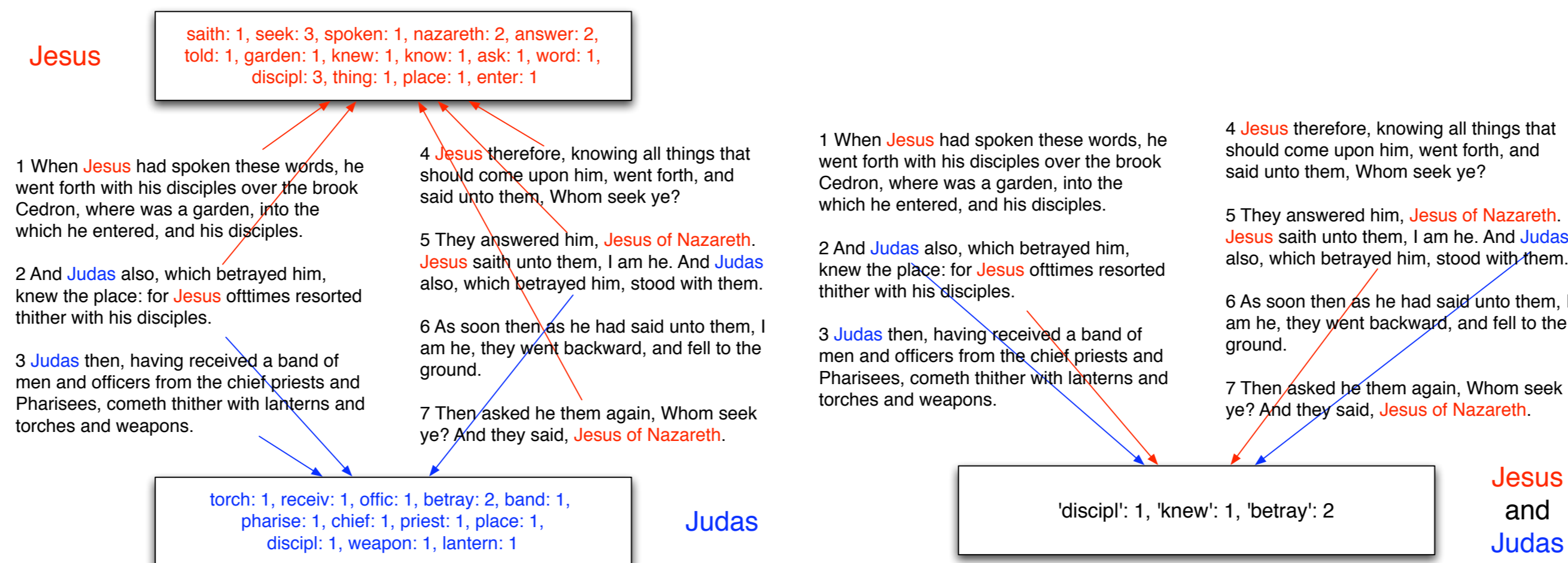


We evaluate our model based on how well it:

- ▶ predicts which entities participate in a relationship (given contexts, which entities should appear).
- ▶ describes the relationship between the entities (given entities, what context should appear).

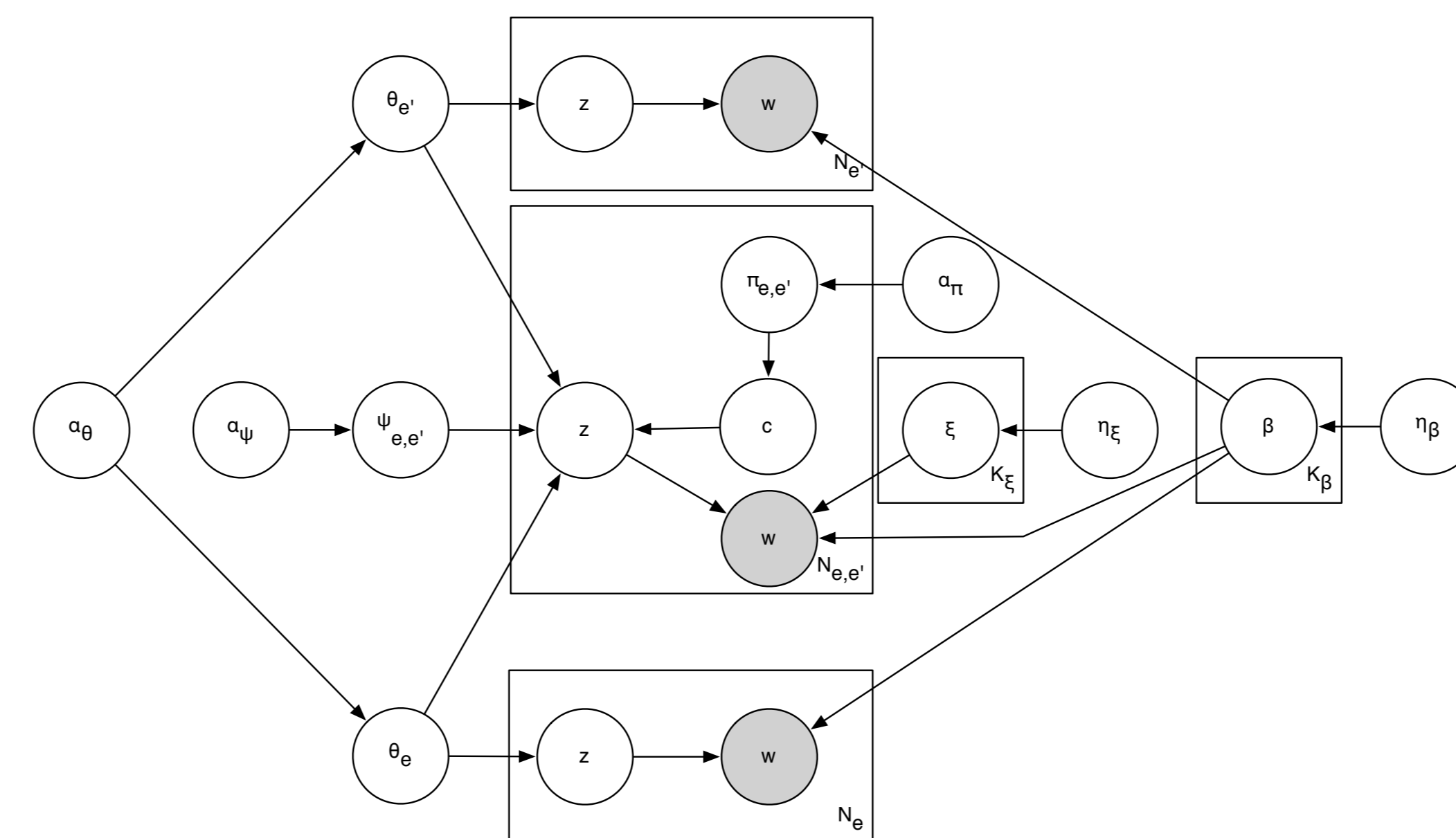
Creating the model input from free text

- ▶ We assume we have raw text where entities have been identified (either automatically or manually).
- ▶ We extract per-person snippets and per-pair snippets from this raw text.



- ▶ The model uses these snippets to generate networks of entities and descriptions of the relationships between them.

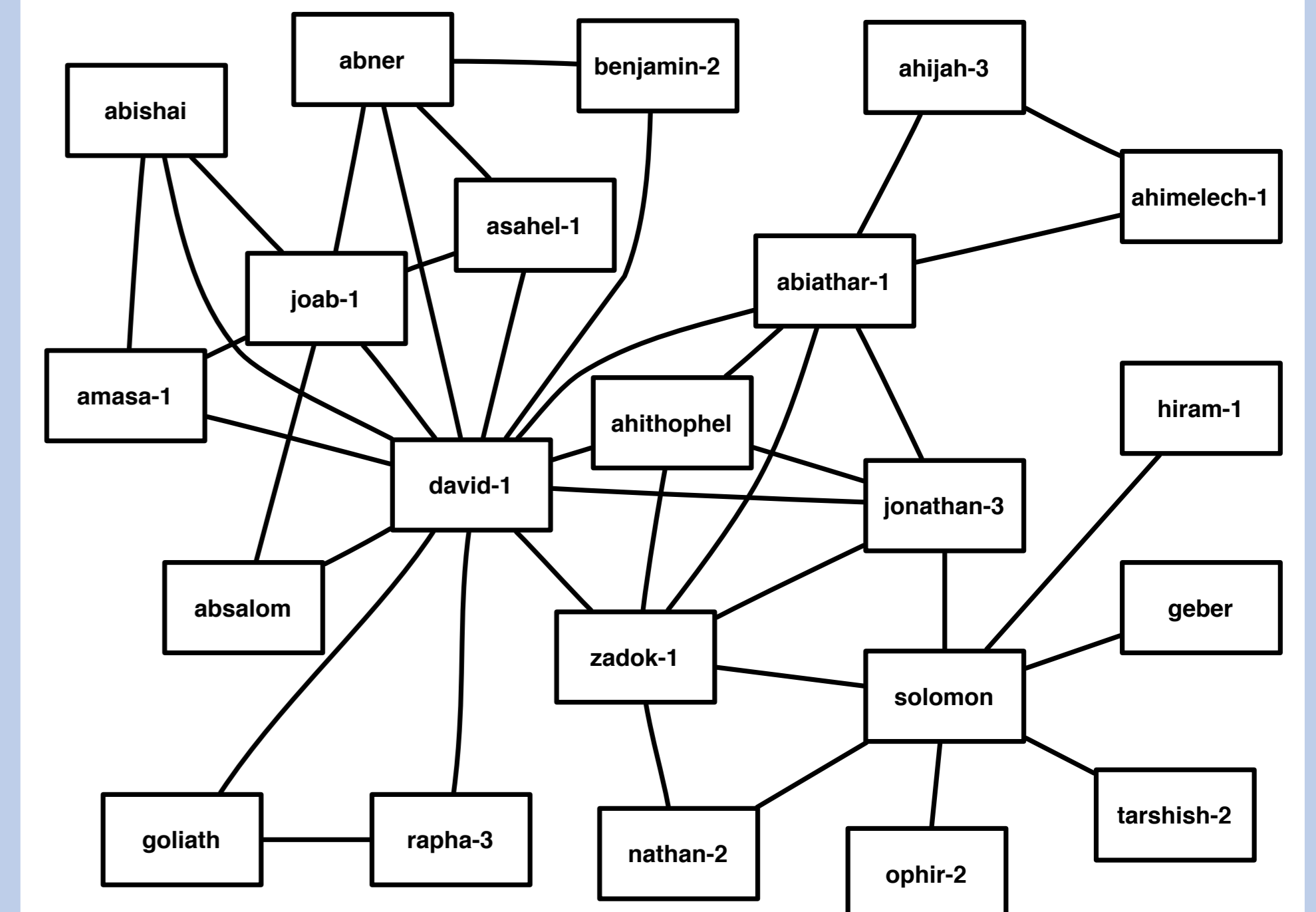
Model



- For each entity e ,
 - Draw $\theta_e \sim \text{Dir}(\alpha_\theta)$;
 - For each word associated with this entity,
 - Draw $z_{e,n} \sim \text{Mult}(\theta_e)$;
 - Draw $w_{e,n} \sim \text{Mult}(\beta_{z_{e,n}})$.
- For each pair of entities e, e' ,
 - Draw $\psi_{e,e'} \sim \text{Dir}(\alpha_\psi)$, $\pi_{e,e'} \sim \text{Dir}(\alpha_\pi)$;
 - For each word associated with this entity pair,
 - Draw $c_{e,e',n} \sim \text{Mult}(\pi_{e,e'})$;
 - Draw $z_{e,e',n} \sim \text{Mult}(\delta_0(c_{e,e',n})\theta_e + \delta_1(c_{e,e',n})\theta_{e'} + \delta_2(c_{e,e',n})\psi_{e,e'})$;
 - Draw $w_{e,e',n} \sim \text{Mult}((\delta_0(c_{e,e',n}))\beta_{z_{e,n}} + \delta_1(c_{e,e',n})\beta_{z_{e',n}} + \delta_2(c_{e,e',n})\xi_{z_{e,e',n}})$.

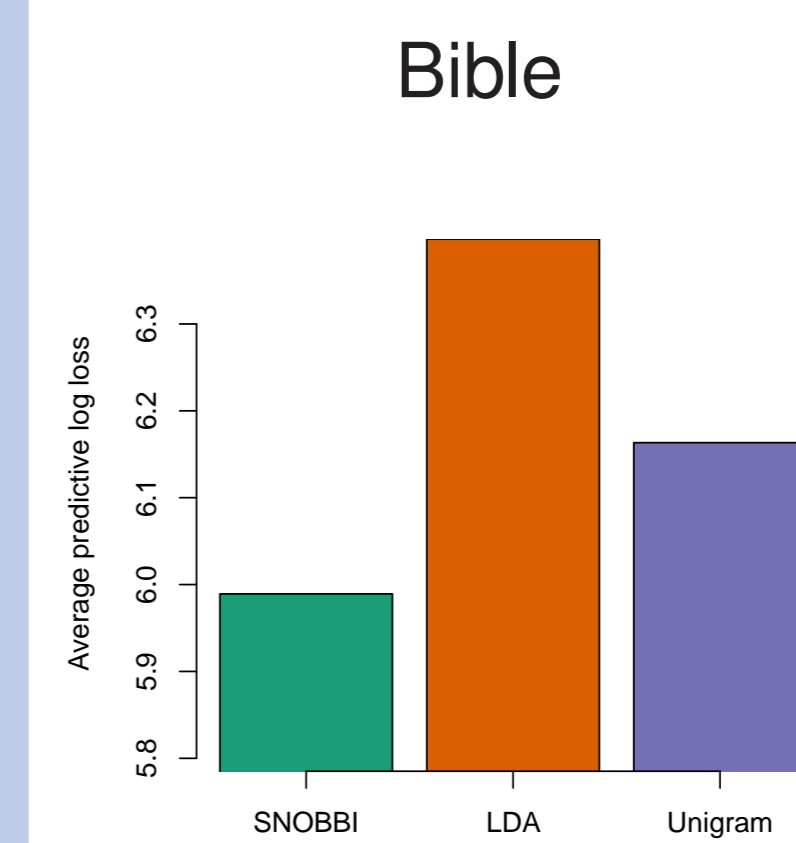
- ▶ θ_e characterizes entity e and $\theta_{e'}$ characterizes e' . Words in the pair snippet for (e, e') come from $\theta_e, \theta_{e'}$, or $\psi_{e,e'}$.
- ▶ $\pi_{e,e'}$ expresses the strength of the relationship between e and e' , and $\psi_{e,e'}$ reveals what that relationship is.

Learned social network from the Bible

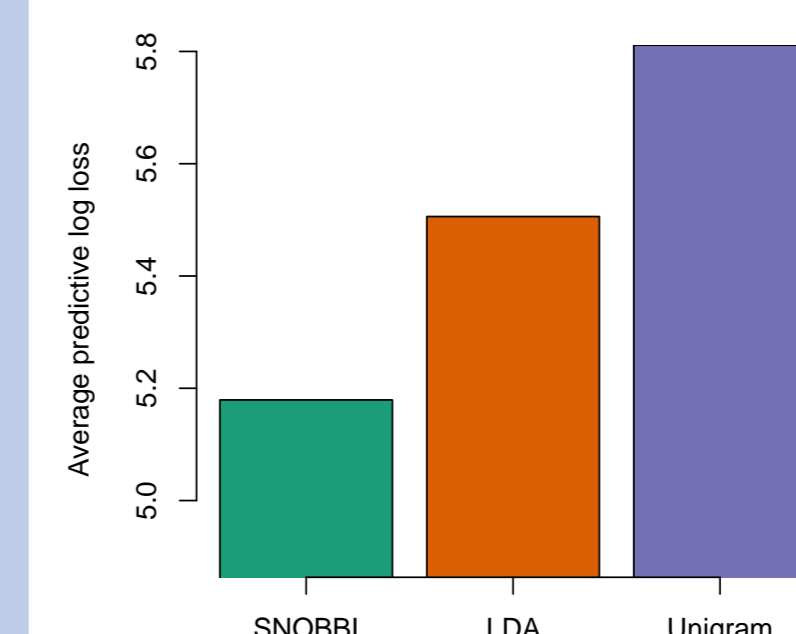


"kinship"	"military"
wife	king
call	fled
mother	lord
brother	citi
abram	smote
daughter	syrian
Miriam, Moses	Chedorlaomer, Abraham
Jesus, Mary	Gideon, Zebah

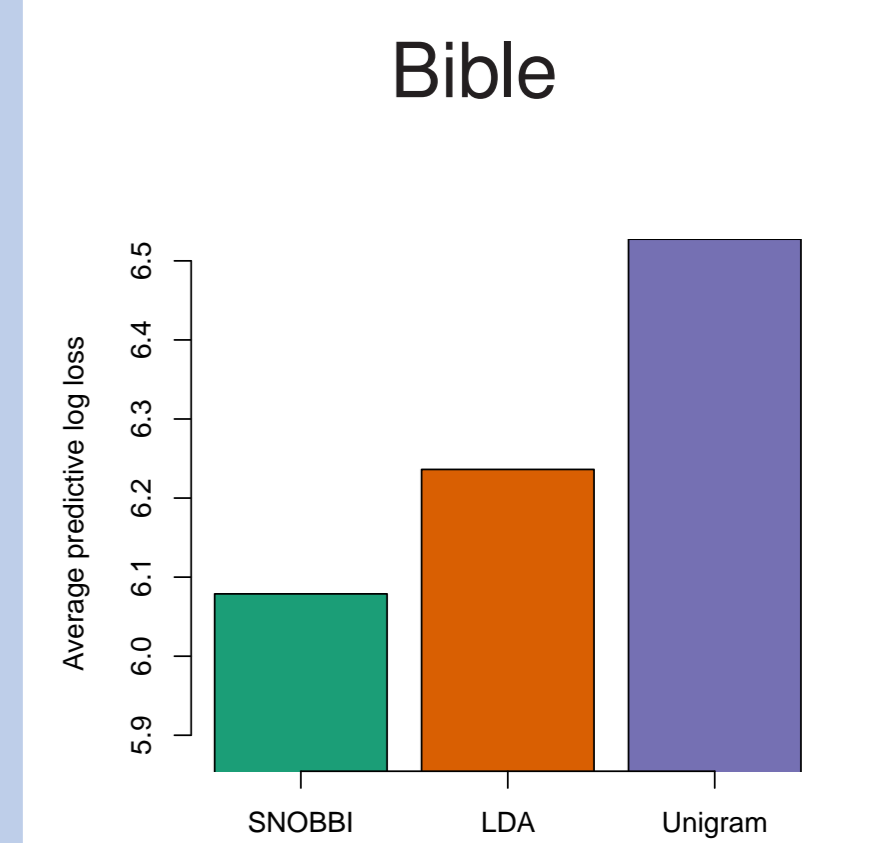
Entity prediction



Wikipedia



Relation prediction



Wikipedia

