Overcoming Language Variation in Sentiment Analysis with Social Attention

Yi Yang

Bloomberg LP

Work performed at Georgia Tech with Jacob Eisenstein.

Language variation in sentiment analysis

"I would like to believe he's **sick** rather than just mean and evil."

Language variation in sentiment analysis

"I would like to believe he's **sick** rather than just mean and evil."

"You could've been getting down to this **sick** beat."

Language variation in sentiment analysis

Language variation in sentiment analysis

I am sick and weak

THIS IS SO SICK THANK U

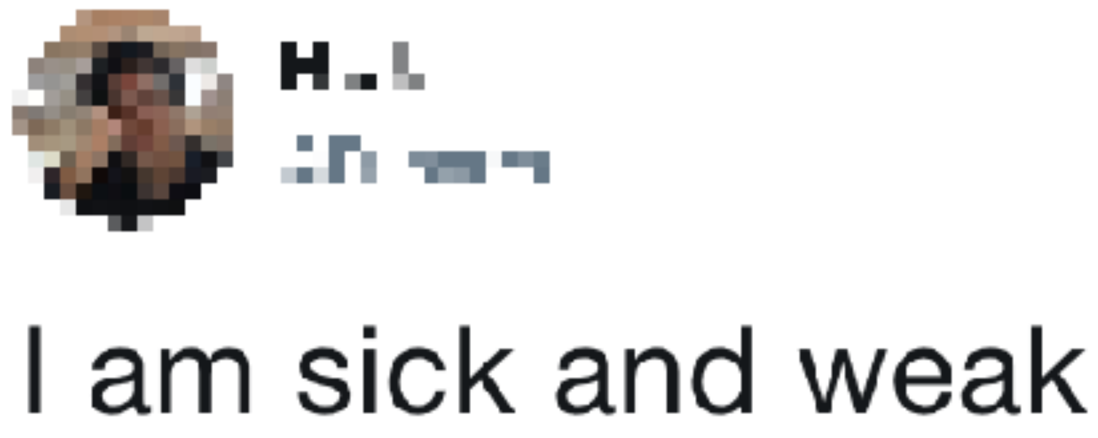
Language variation in sentiment analysis

I am sick and weak

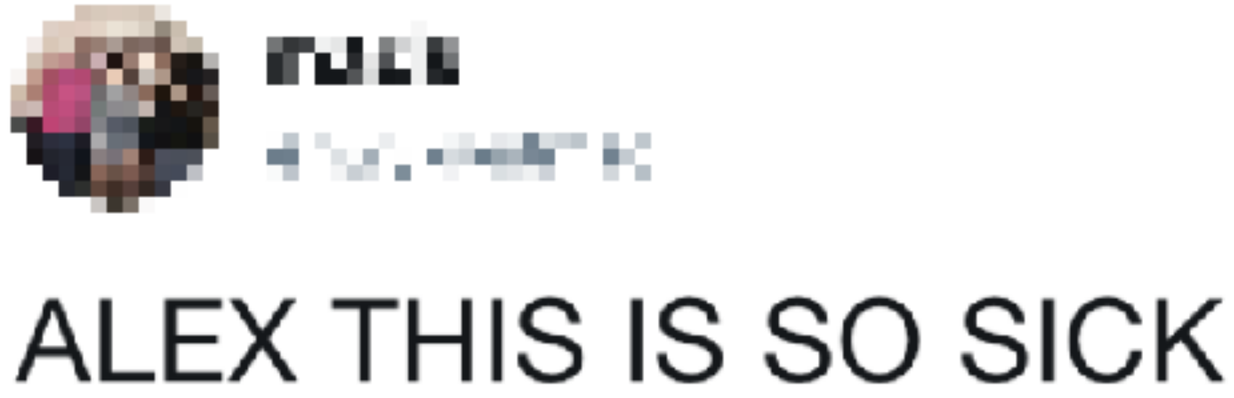
THIS IS SO SICK THANK U

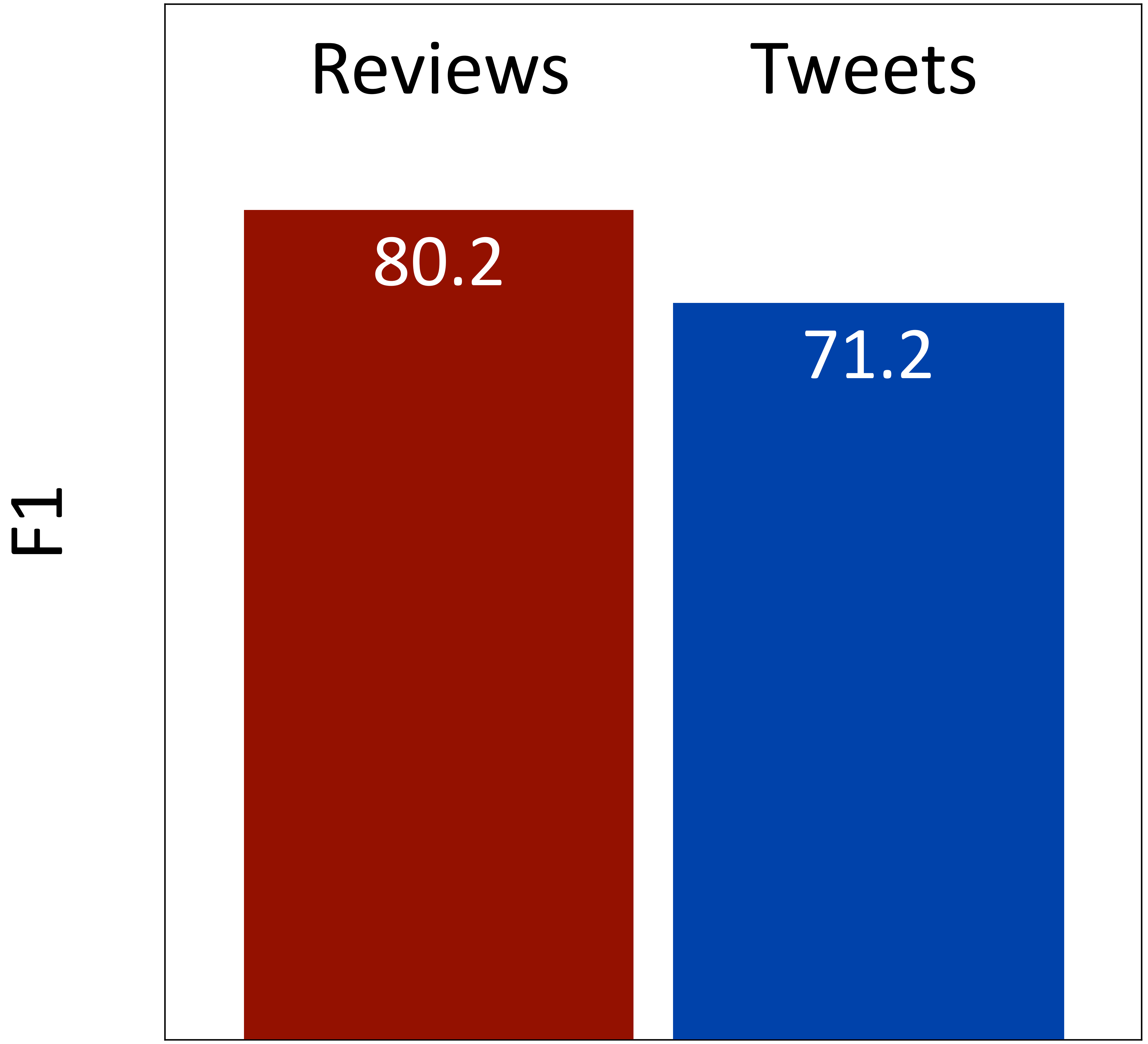
ALEX THIS IS SO SICK

Language variation in sentiment analysis





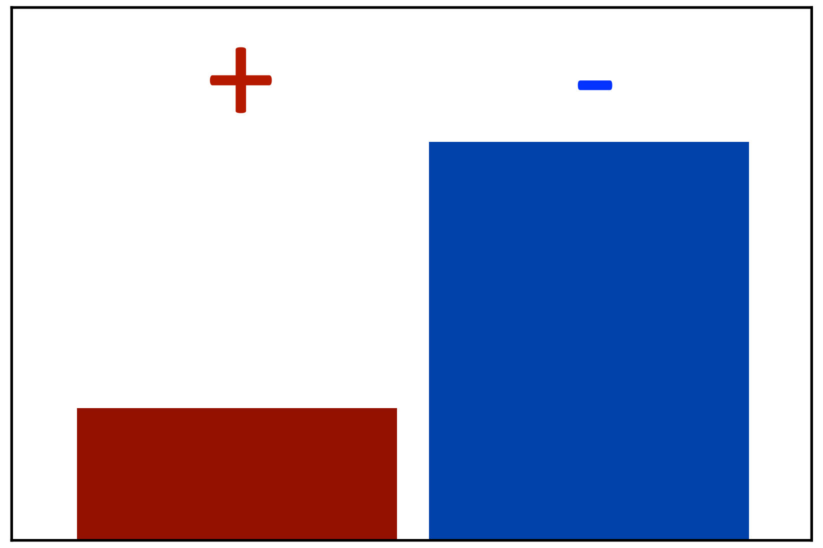


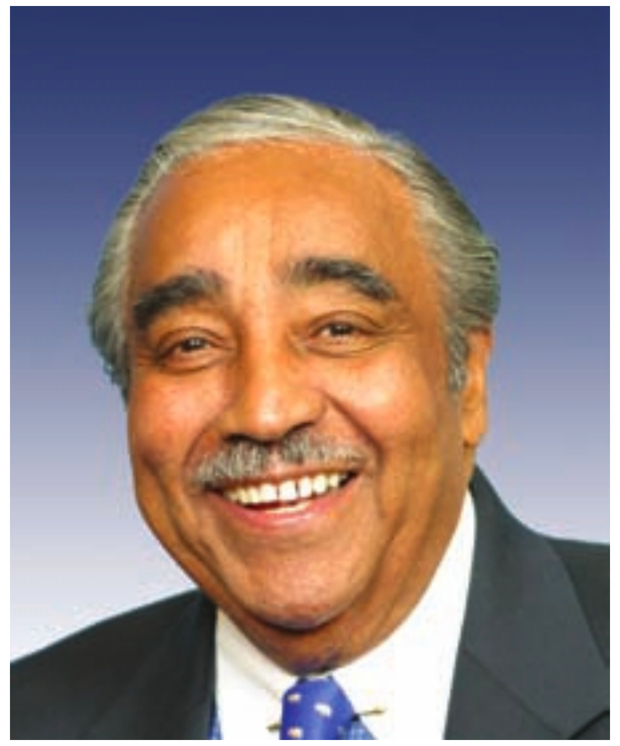
Personalized sentiment analysis

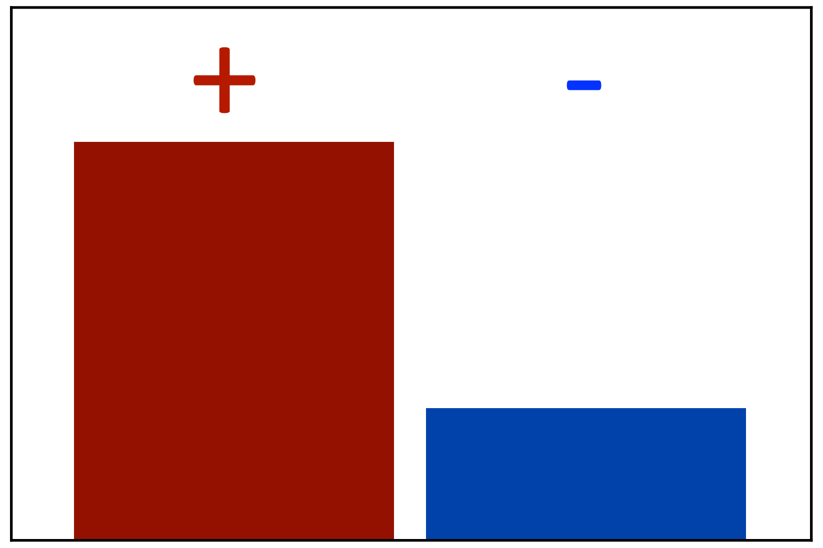
- **Goal**: personalized conditional likelihood, $p(y|\mathbf{x}, a)$.
  - $\mathbf{x}$ is the text, and a is the author.

Personalized sentiment analysis

- **Goal**: personalized conditional likelihood, $p(y|\mathbf{x}, a)$.
  - $\mathbf{x}$ is the text, and a is the author.

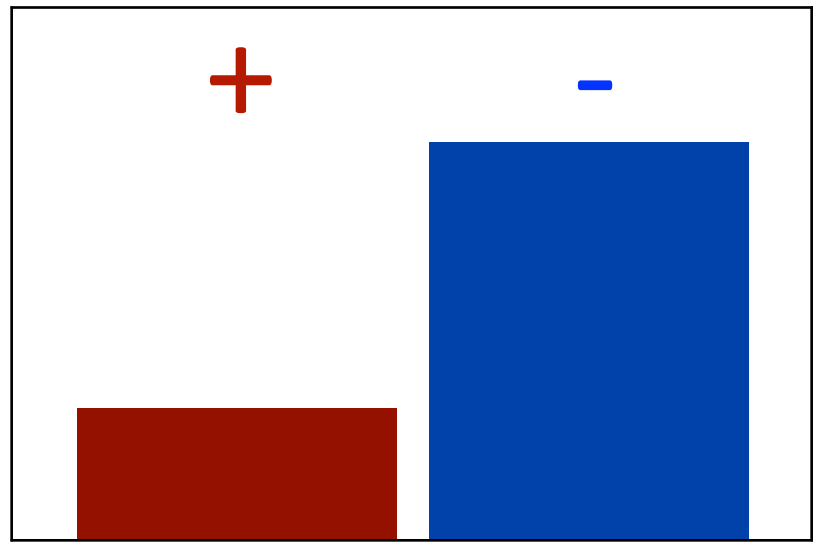


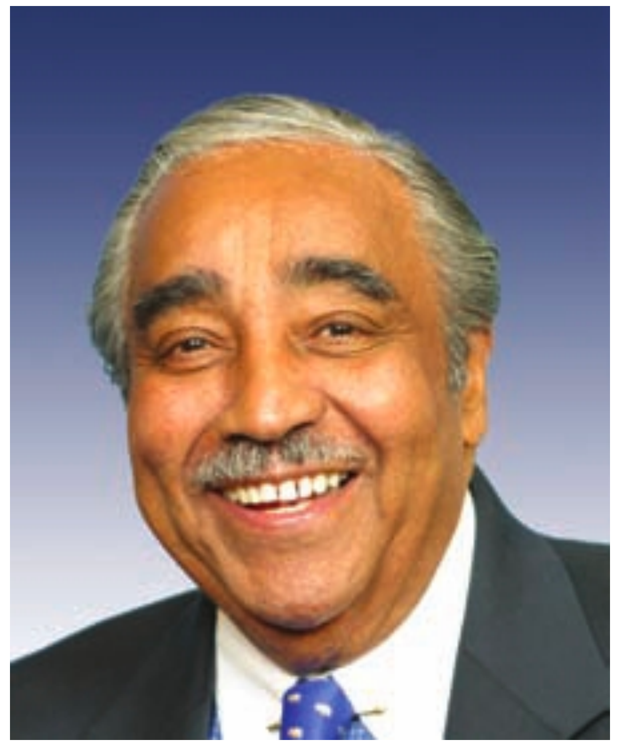


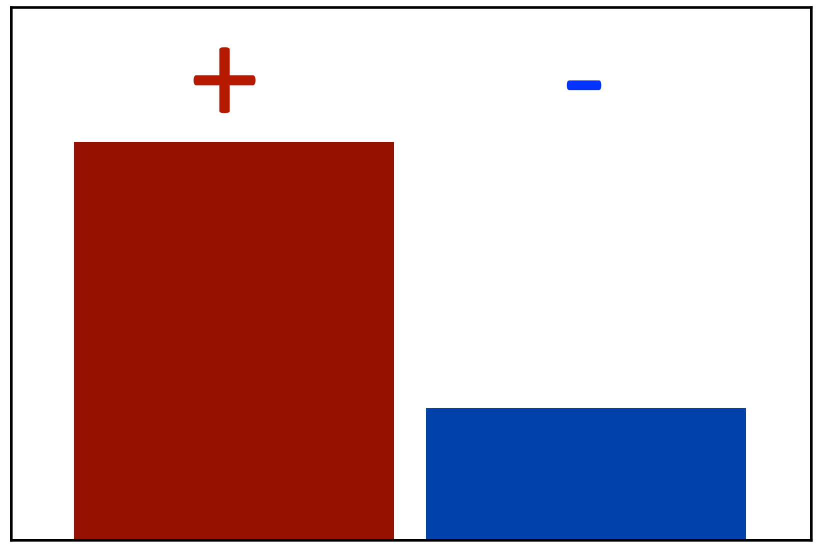


Personalized sentiment analysis

- **Goal**: personalized conditional likelihood, $p(y|\mathbf{x}, a)$.
  - $\mathbf{x}$ is the text, and a is the author.







- **Problem**: we have labeled examples for only a few authors.

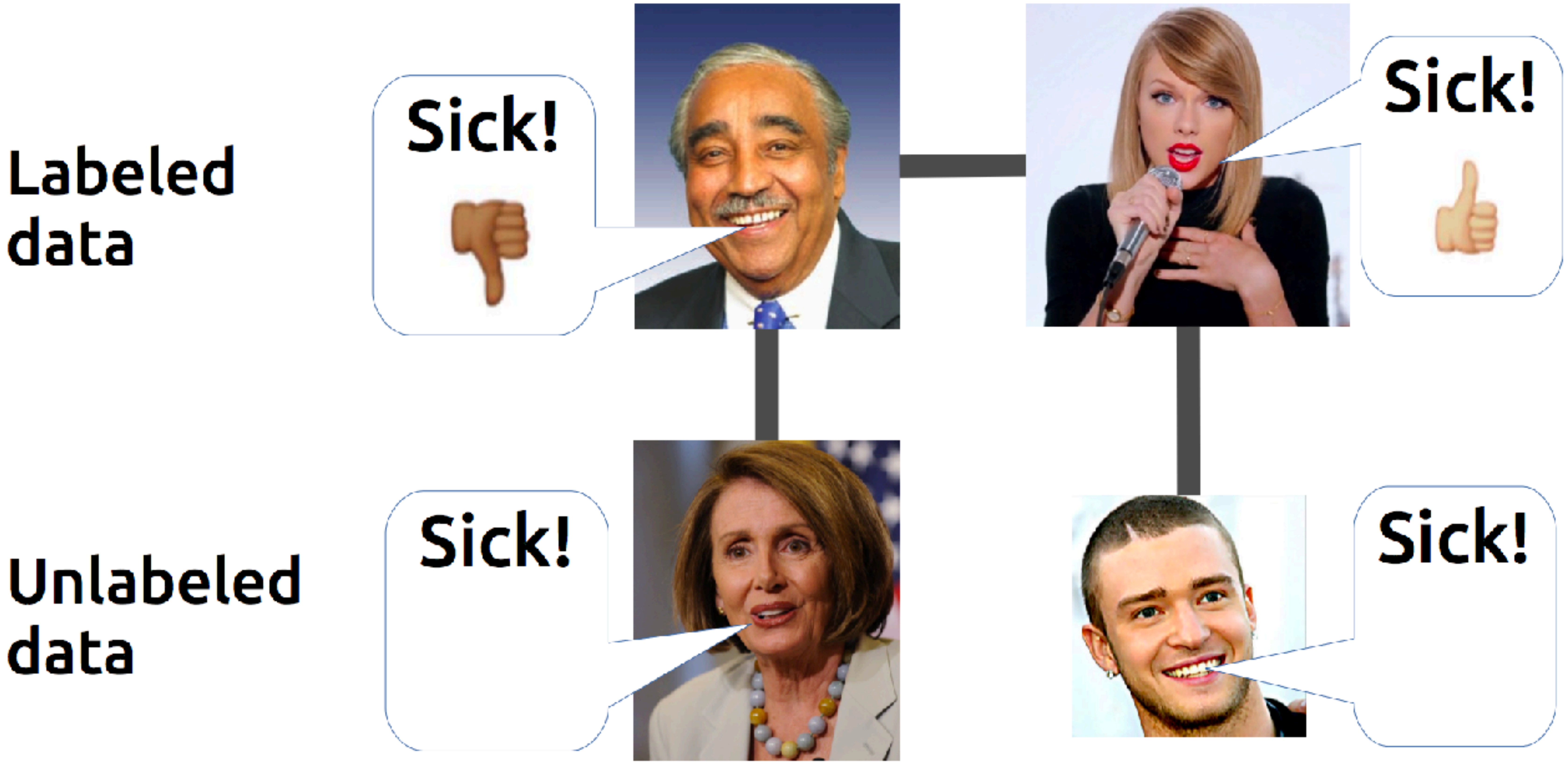
Homophily to the rescue?

Homophily: neighbors have similar properties.

Thelwall (2009); Al Zamal et al. (2012)

Homophily to the rescue?

Homophily: neighbors have similar properties.



Thelwall (2009); Al Zamal et al. (2012)

Evidence for linguistic homophily

Pilot study: is classifier accuracy assortative on the Twitter social network?

Evidence for linguistic homophily

Pilot study: is classifier accuracy assortative on the Twitter social network?

$$\text{assort}(G) = \frac{1}{\#|G|} \sum_{(i,j)\in G} \delta(y_i = \hat{y}_i)\delta(y_j = \hat{y}_j) + \delta(y_i \neq \hat{y}_i)\delta(y_j \neq \hat{y}_j)$$

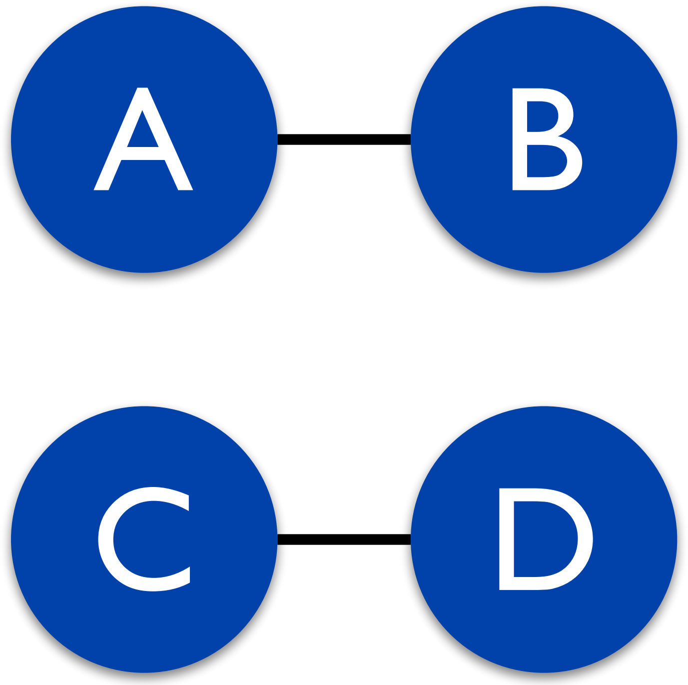
- Whether a sentiment classifier tends to make consistent predictions for social neighbors.

Evidence for linguistic homophily

Network rewiring: degree-preserving randomization

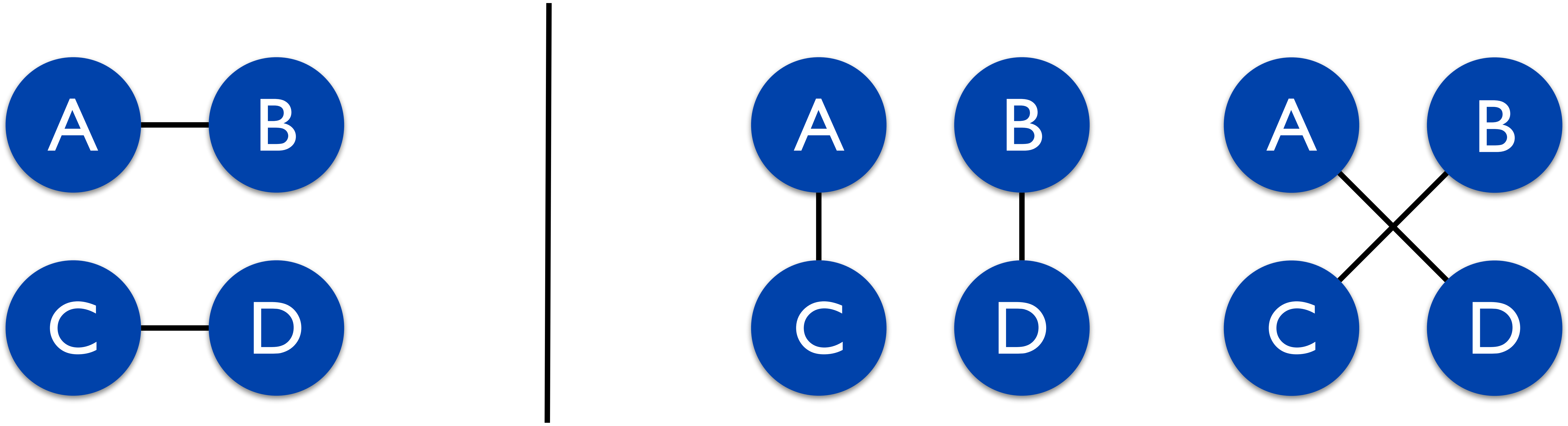
Evidence for linguistic homophily

Network rewiring: degree-preserving randomization



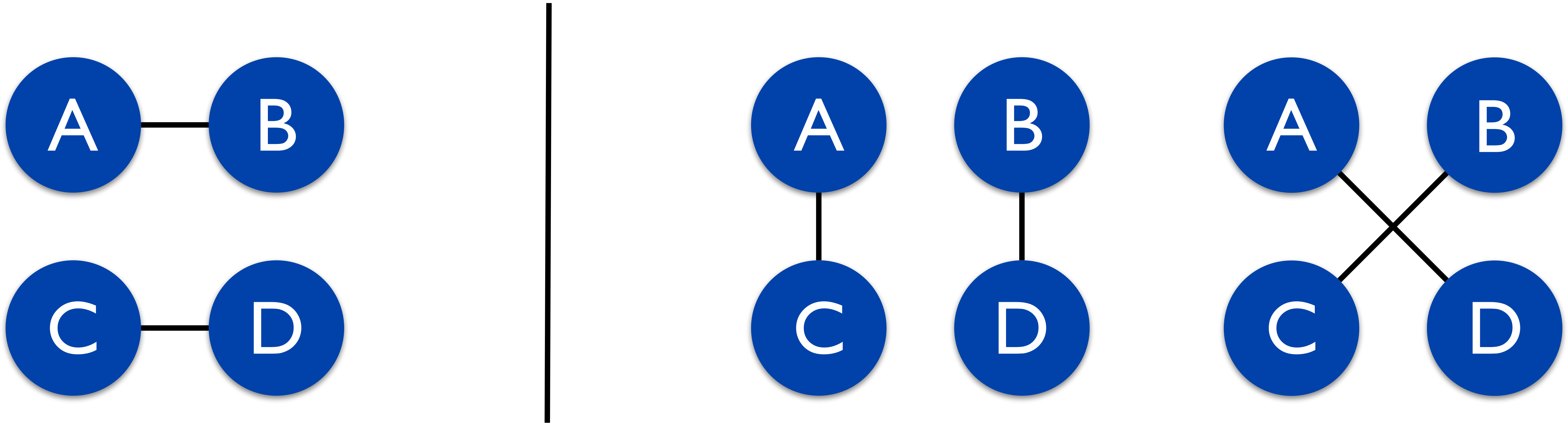
Evidence for linguistic homophily

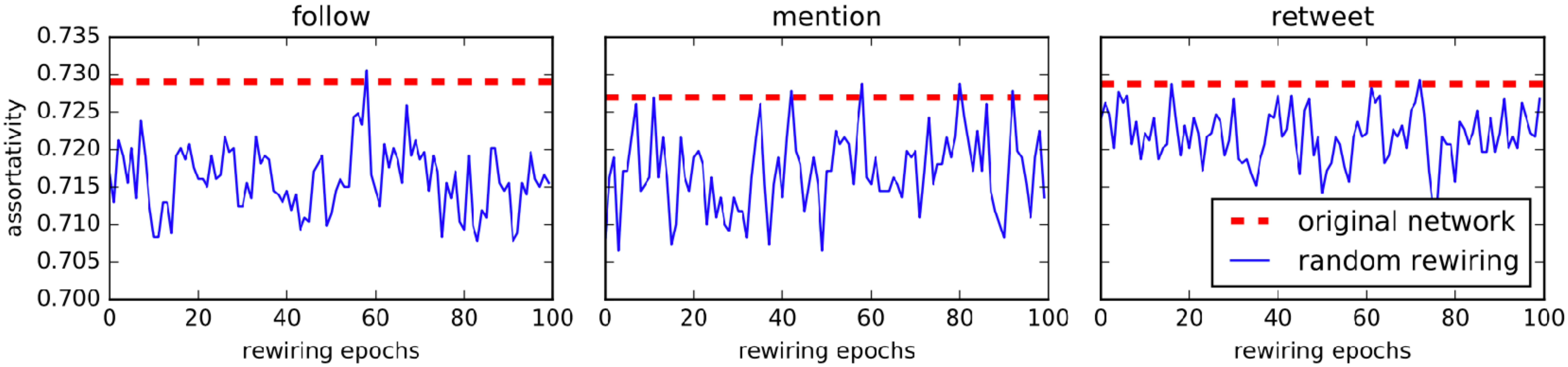
Network rewiring: degree-preserving randomization



Evidence for linguistic homophily

Network rewiring: degree-preserving randomization





Model

Personalization by ensemble

$$p(y|\mathbf{x}, a) = \sum_{k=1}^{K} \underbrace{Pr(Z_a = k|a, G)}_{\text{ensemble weights}} \times \underbrace{p(y|\mathbf{x}, Z_a = k)}_{\text{basis models}}$$

Personalization by ensemble

$$p(y|\mathbf{x}, a) = \sum_{k=1}^{K} \underbrace{Pr(Z_a = k|a, G)}_{\text{ensemble weights}} \times \underbrace{p(y|\mathbf{x}, Z_a = k)}_{\text{basis models}}$$

- Train each basis model with all the labeled data.
  - Employ ConvNets as basis models.

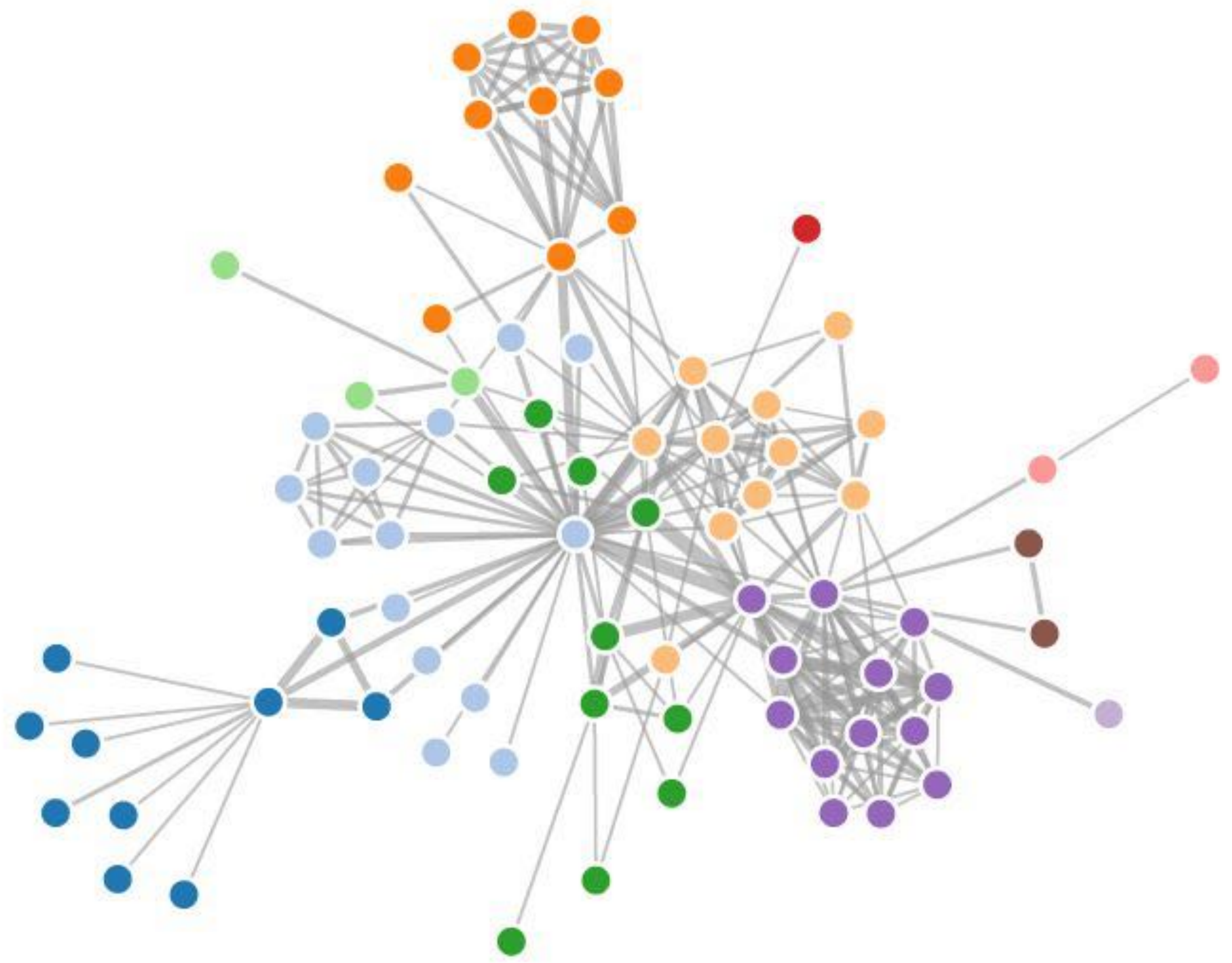
Personalization by ensemble

$$p(y|\mathbf{x}, a) = \sum_{k=1}^{K} \underbrace{Pr(Z_a = k|a, G)}_{\text{ensemble weights}} \times \underbrace{p(y|\mathbf{x}, Z_a = k)}_{\text{basis models}}$$

- Train each basis model with all the labeled data.
  - Employ ConvNets as basis models.
- Apply linguistic homophily:
  - Adopt similar ensemble weights for social neighbors.
  - De-correlate errors made by different basis models.

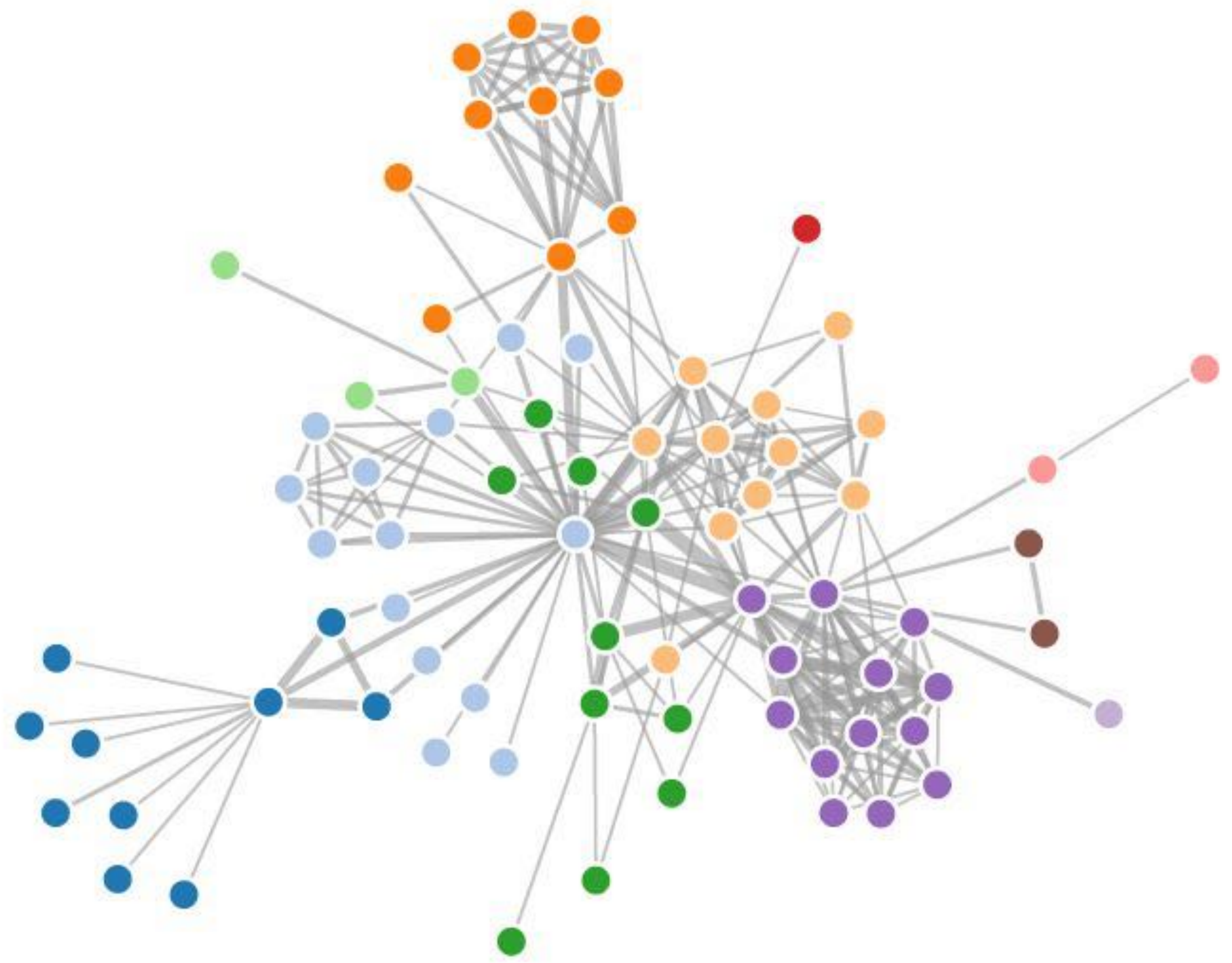
Network-driven personalization

- For each author, estimate a node embedding $\mathbf{v}_a$ (Tang et al., 2015).
- Nodes who share neighbors get similar embeddings.



Network-driven personalization

- For each author, estimate a node embedding $\mathbf{v}_a$ (Tang et al., 2015).
- Nodes who share neighbors get similar embeddings.



- **Social attention**:

$$Pr(Z_a = k | a, G) = \mathrm{SoftMax}(f(\mathbf{v}_a))$$

Learning

- Jointly train with cross-entropy loss:

$$\ell(\Theta) = -\sum_{t=1}^{T} \mathbf{1}[Y^* = t] \log \Pr(Y = t \mid \mathbf{x}, a)$$

Learning

- Jointly train with cross-entropy loss:

$$\ell(\Theta) = -\sum_{t=1}^{T} \mathbf{1}[Y^* = t] \log \Pr(Y = t \mid \mathbf{x}, a)$$

Problem: network information tends to be ignored.

Learning

- Jointly train with cross-entropy loss:

$$\ell(\Theta) = -\sum_{t=1}^{T} \mathbf{1}[Y^* = t] \log \Pr(Y = t \mid \mathbf{x}, a)$$

Problem: network information tends to be ignored.

- Pre-train basis models with instance-weighted losses:

$$\ell_k = -\alpha_{a,k} \sum_{t=1}^{T} \mathbf{1}[Y^* = t] \log \Pr(Y = t \mid \mathbf{x}, Z_a = k)$$

Experiments

Data

- SemEval Twitter sentiment analysis data.
  - 18,024 tweets
  - Follow, mention, retweet networks
  - Network expansion

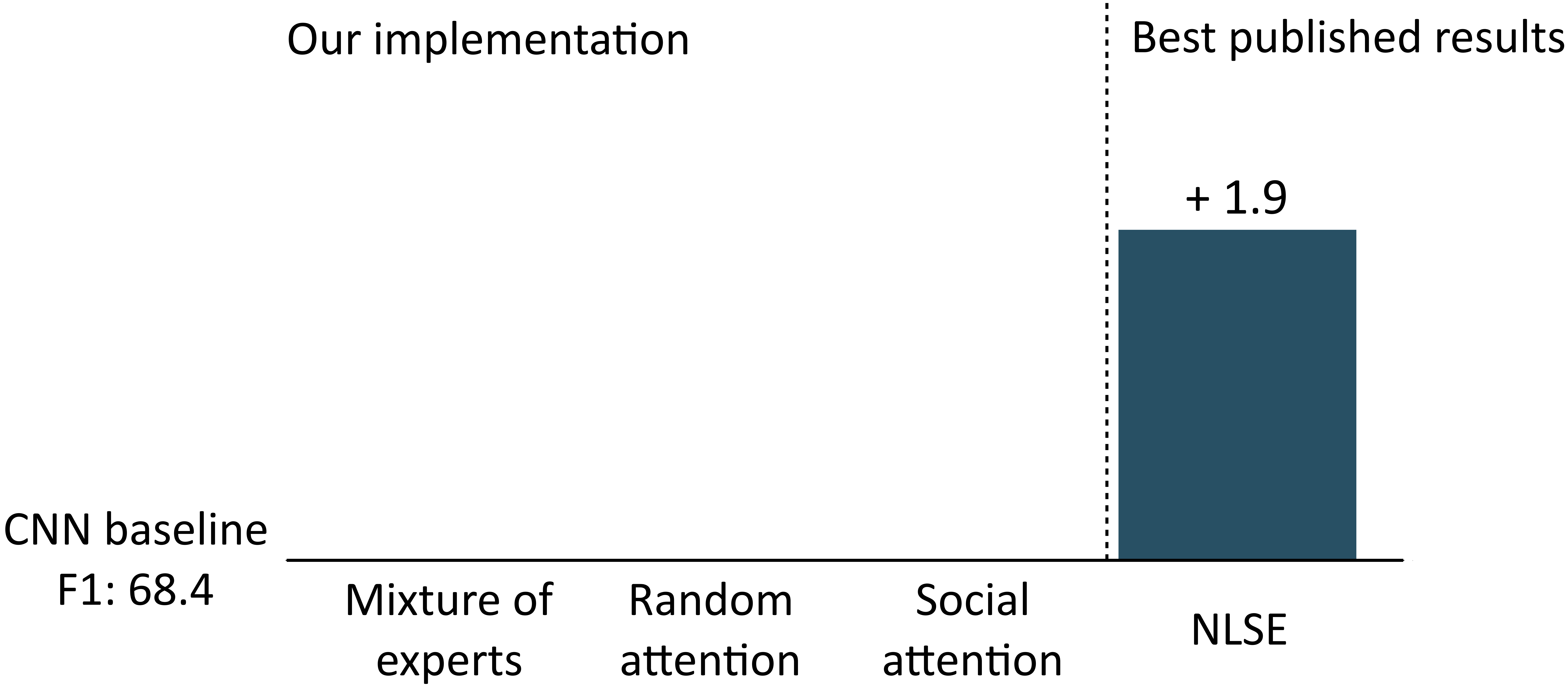
Nakov et al. (2013); Rosenthal et al. (2015); Tang et al. (2012)

Data

- SemEval Twitter sentiment analysis data.
  - 18,024 tweets
  - Follow, mention, retweet networks
  - Network expansion
- Ciao product review sentiment analysis data.
  - 100,000 reviews
  - User trust network

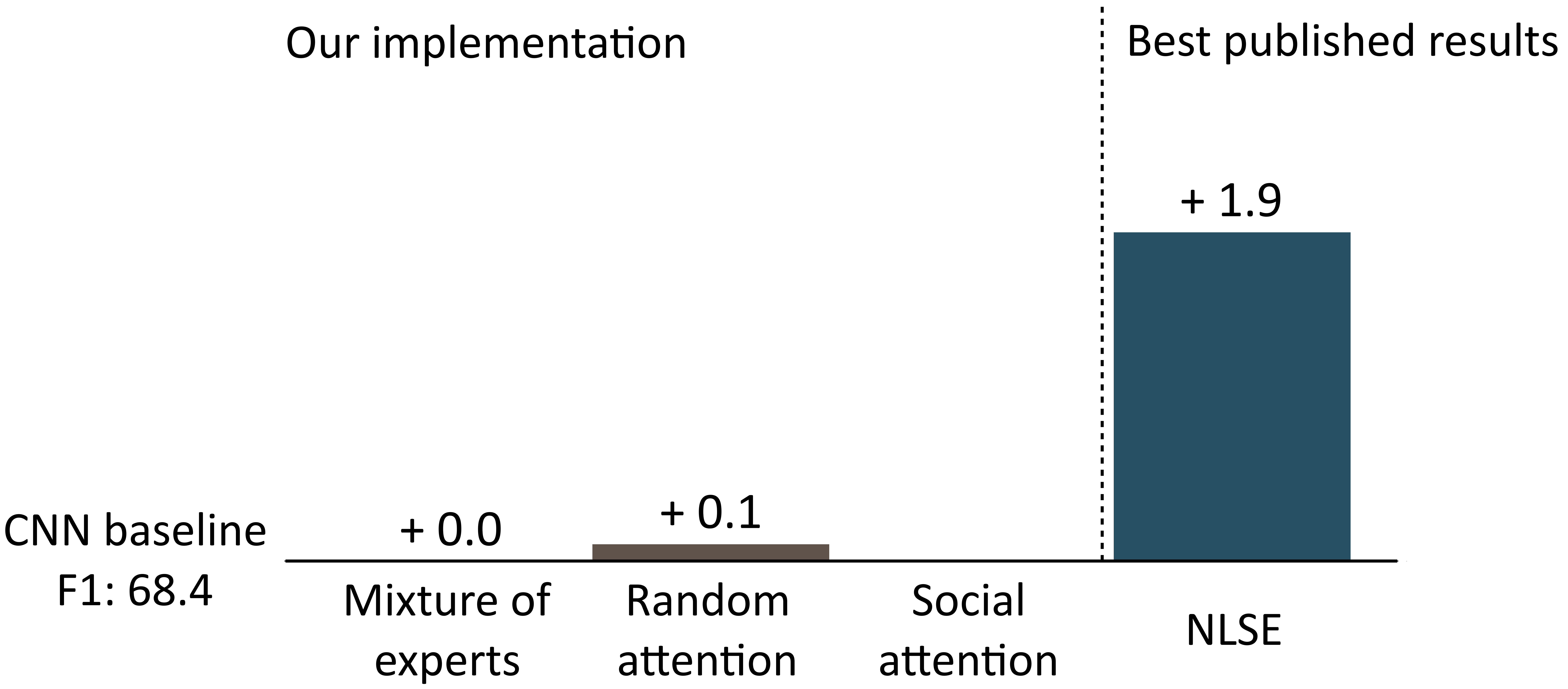
Nakov et al. (2013); Rosenthal et al. (2015); Tang et al. (2012)

Results: SemEval Twitter data



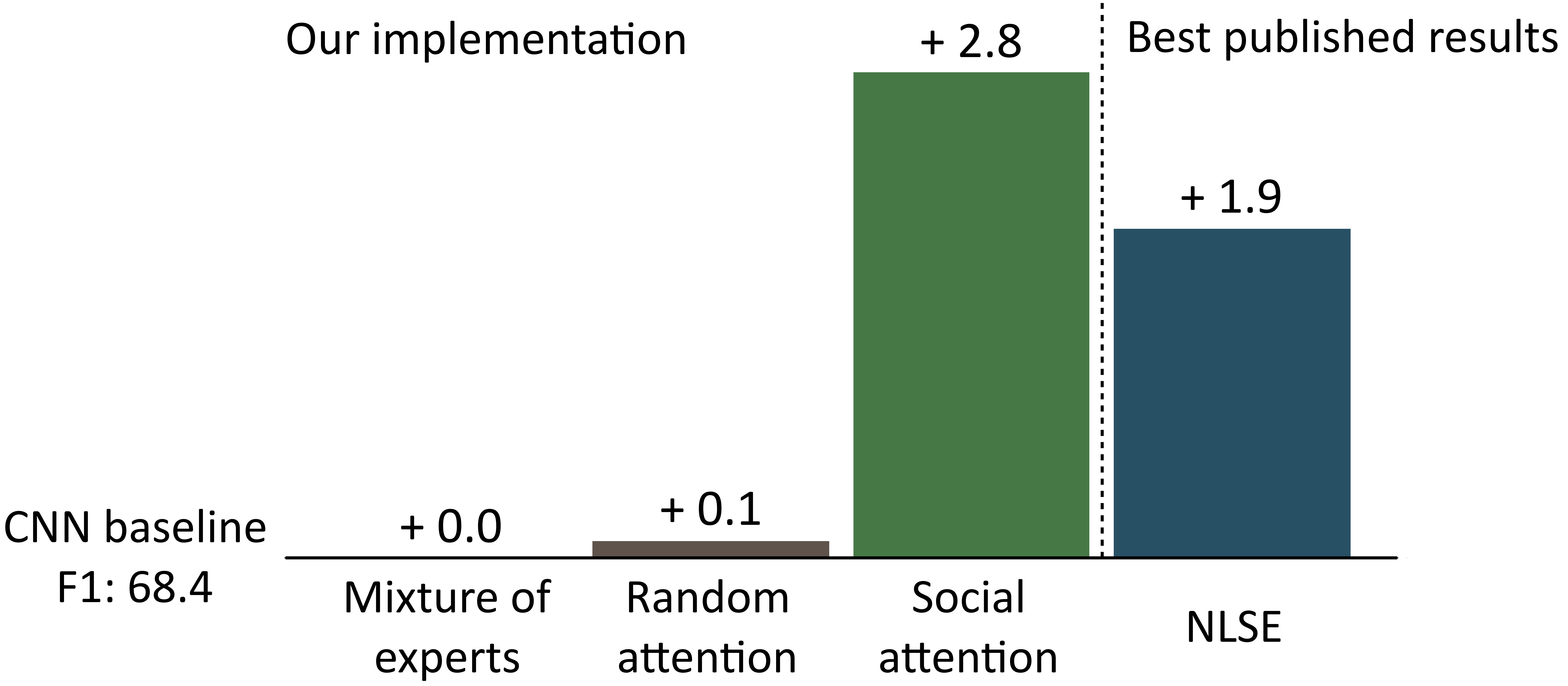
Astudillo et al. (2015)

Results: SemEval Twitter data



Astudillo et al. (2015)

Results: SemEval Twitter data



Astudillo et al. (2015)

Variable sentiment words

	More positive	More negative
1	banging loss fever broken fucking	dear like god yeah wow
2	chilling cold ill sick suck	satisfy trust wealth strong lmao
3	ass damn piss bitch shit	talent honestly voting win clever
4	insane bawling fever weird cry	lmao super lol haha hahaha
5	ruin silly bad boring dreadful	*lovatics* wish *beliebers ariana-tors kendall*

Results: Ciao review data

CNN baseline
F1: 74.4

+ 0.0

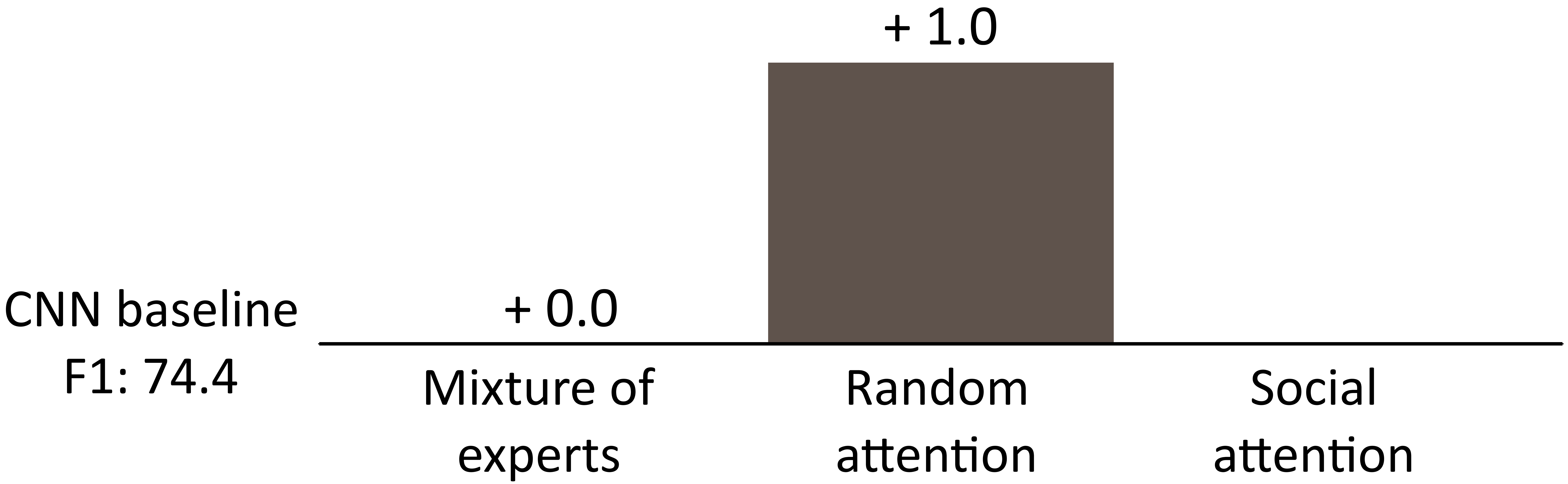
Mixture of experts

Random attention

Social attention

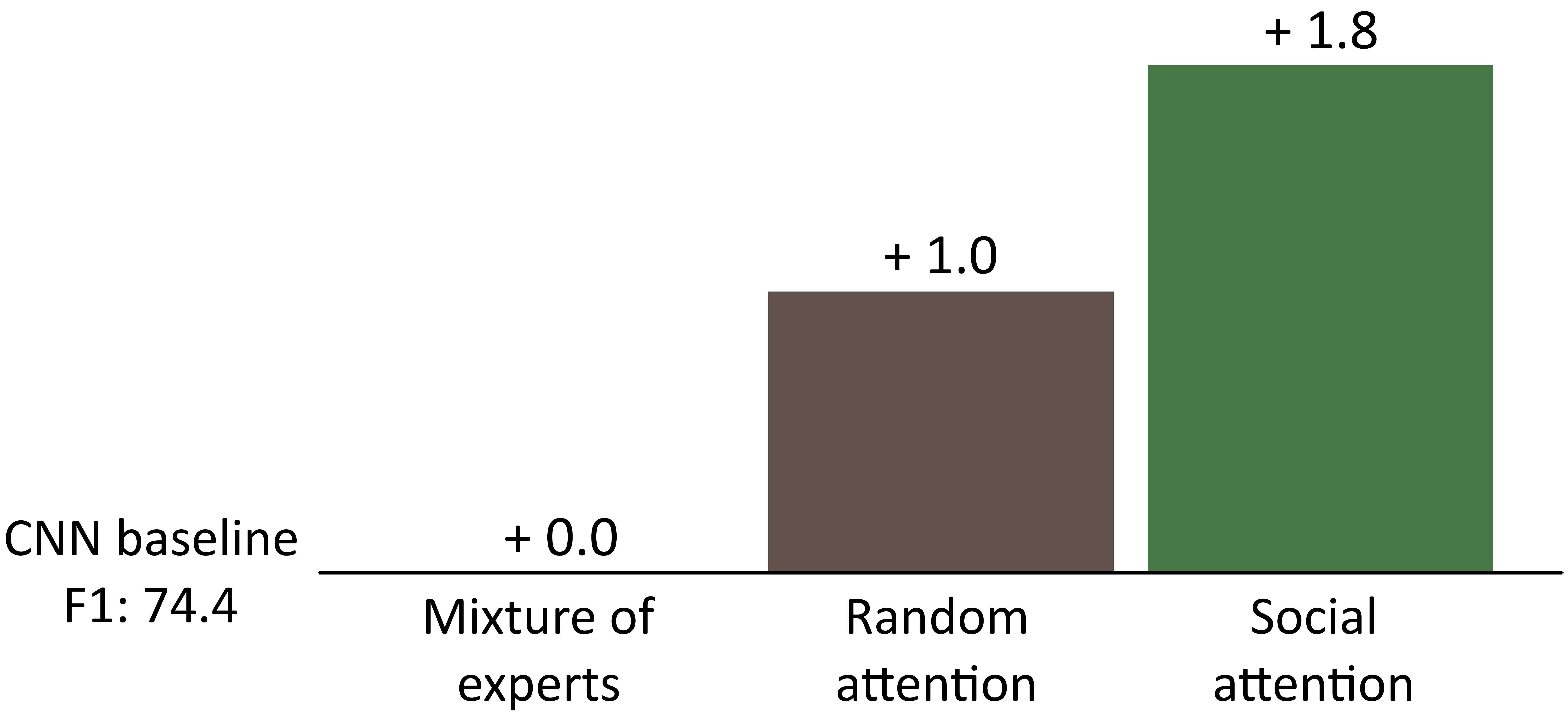
Tang et al. (2012)

Results: Ciao review data



Tang et al. (2012)

Results: Ciao review data



Tang et al. (2012)

Conclusions and future work

- Language variation poses challenges in sentiment analysis.
- Linguistic homopily alleviates the data sparsity issue for estimating personalized models.
- Social attention mechanism significantly improves accuracy.
- The socially-infused ensemble architecture can be applied to other tasks such as tagging, parsing, etc.