

# Link Prediction from Heterogeneous Opinion Mining Networks with Multi-Domain Applications



Bernardo M. Costa, Aime Nobrega, Adailton Araujo, Ricardo Marcacini

Institute of Mathematics and Computer Sciences (ICMC), University of São Paulo (USP)



## Motivation and Research Question

Opinion Mining and Analysis: core task in NLP

State-of-the-art results on mining opinion from text obtained with feature extraction and sentiment analysis methods

The **laptop** has an **excellent** **screen** but the **battery** is **short-lived**

$e_i$                        $s_{ijkl}$                        $a_{ij}$                        $a_{i(j+1)}$                        $s_{i(j+1)kl}$

"laptop" is the entity ( $e_i$ )  
 "screen" is the aspect ( $a_{ij}$ ) with positive sentiment  
 "battery" is another aspect ( $a_{i(j+1)}$ ) with negative sentiment

Requires large labeled datasets for the training or model fine-tuning step, where fine-tuned models for a particular domain often fail in new domains

**How to mine opinions from pre-trained and general-purpose models in different domain applications?**

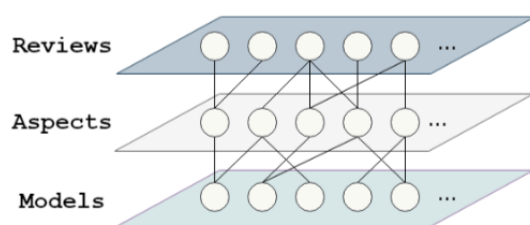
## Our Approach - Heterogeneous Network

Heterogeneous opinion mining network combines opinion data from different domains, integrating models trained for different applications

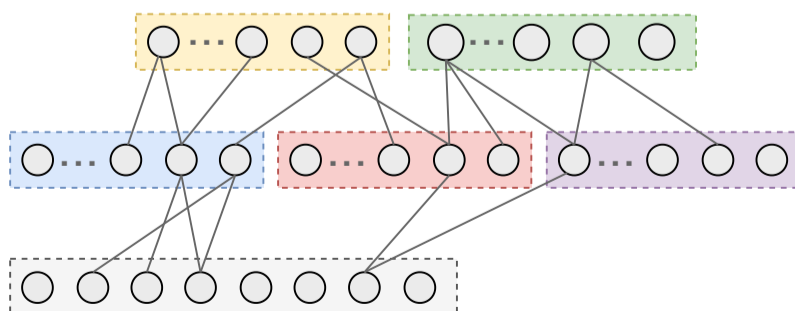
Overview of the datasets used in the experimental evaluation

Dataset	sentences	features	distinct features
eBay	294	206	167
Evernote	367	295	259
Facebook	327	242	204
Netflix	341	262	201
Photo editor	154	96	80
Spotify	227	180	145
Twitter	183	122	99
WhatsApp	169	118	100

General network structure



Network layers composition



- Train set sentences
- Positive samples - obtained from train set
- Negative samples - obtained from train set
- Test set sentences
- Extracted requirements from test sentences
- RE-BERT trained models, each one on an app train dataset

**Link prediction explores the network topology**

## Results

Network topology carries information across domains

GCN Embeddings can be used as valuable information

### Evaluation Results (Accuracy)

Algorithm	GCN-HOMN-LM		GCN-HOMN	
	Mean	Std	Mean	Std
<b>Sample Size (<math>p</math>)</b>				
<b>0.05 (498)</b>	0.824	$\pm 0.023$	0.808	$\pm 0.017$
<b>0.10 (945)</b>	0.810	$\pm 0.025$	0.818	$\pm 0.010$
<b>0.15 (1338)</b>	0.772	$\pm 0.063$	0.818	$\pm 0.009$
<b>0.20 (1680)</b>	0.790	$\pm 0.026$	0.808	$\pm 0.012$
<b>0.25 (1968)</b>	0.793	$\pm 0.011$	0.801	$\pm 0.004$
<b>0.30 (2205)</b>	0.775	$\pm 0.023$	0.782	$\pm 0.007$

Link prediction with and without Language Modeling enrichment results

## Conclusions

Heterogeneous Networks are a natural way of representing and combining different domains applications on tasks in Opinion Mining

Network embeddings are potentially useful for mining opinions

## Future work

Investigate network embeddings and apply to a wide range of different tasks in opinion mining