L-Università ta' Malta
Faculty of Information & Communication Technology
Department of Artificial Intelligence

Predicting links in a social network based on recognised personalities

Andrew Aquilina

Supervisor: Dr. Charlie Abela

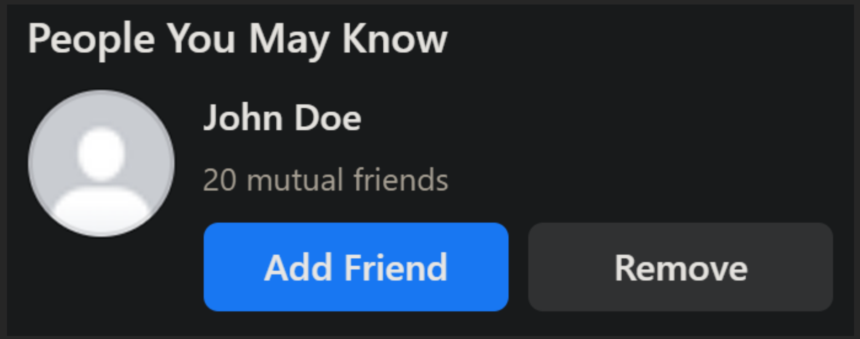
Outline

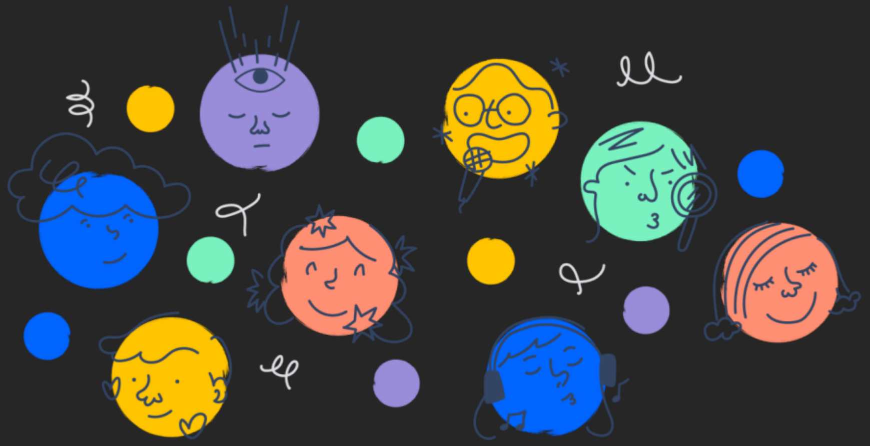
- Motivation & Aim
- Personality Recognition from Text
- Personality-Aware Link Prediction
- Results & Findings

Motivation & Aim

Introduction

- The revolution of Online Social Networks
- Friend recommendations ensure the growth of their network
- The impact of personality towards friend selection





Motivation & Aim

Research Question

“Can link prediction precision improve when the users’ followee personality preferences are taken into consideration?”

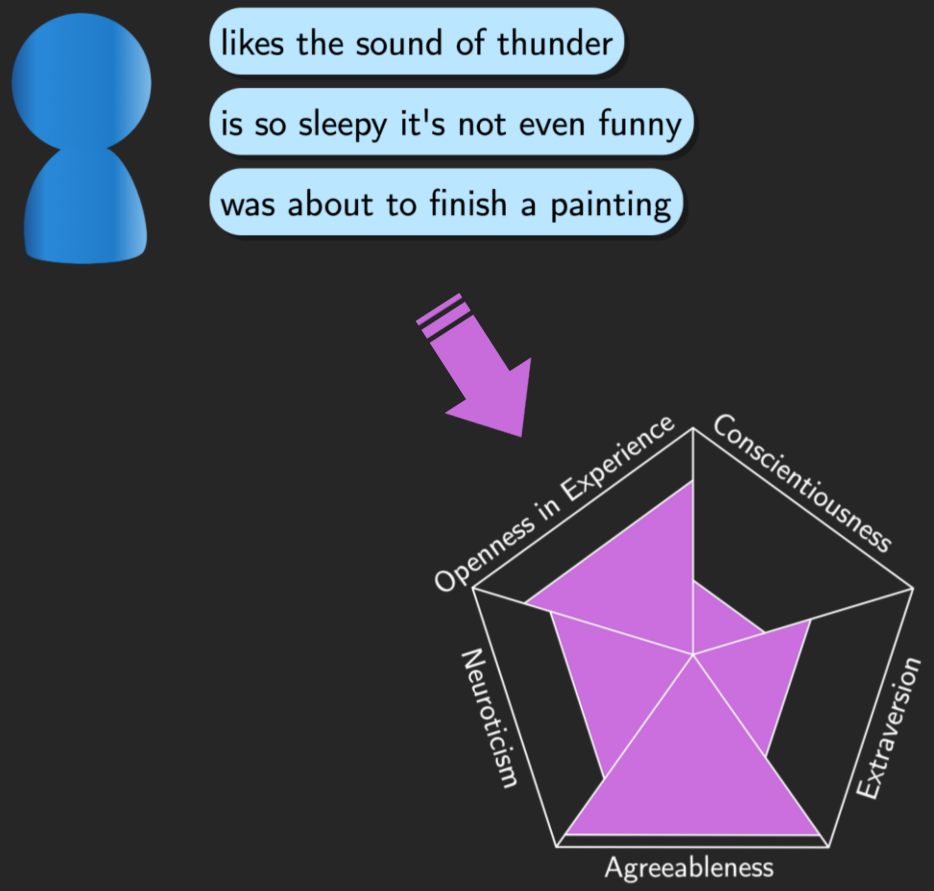
Objective 1: Employ the relationship between language and personality

Objective 2: Determine whether followee personality preferences relate to the created links

Motivation & Aim

Personality Recognition from Text

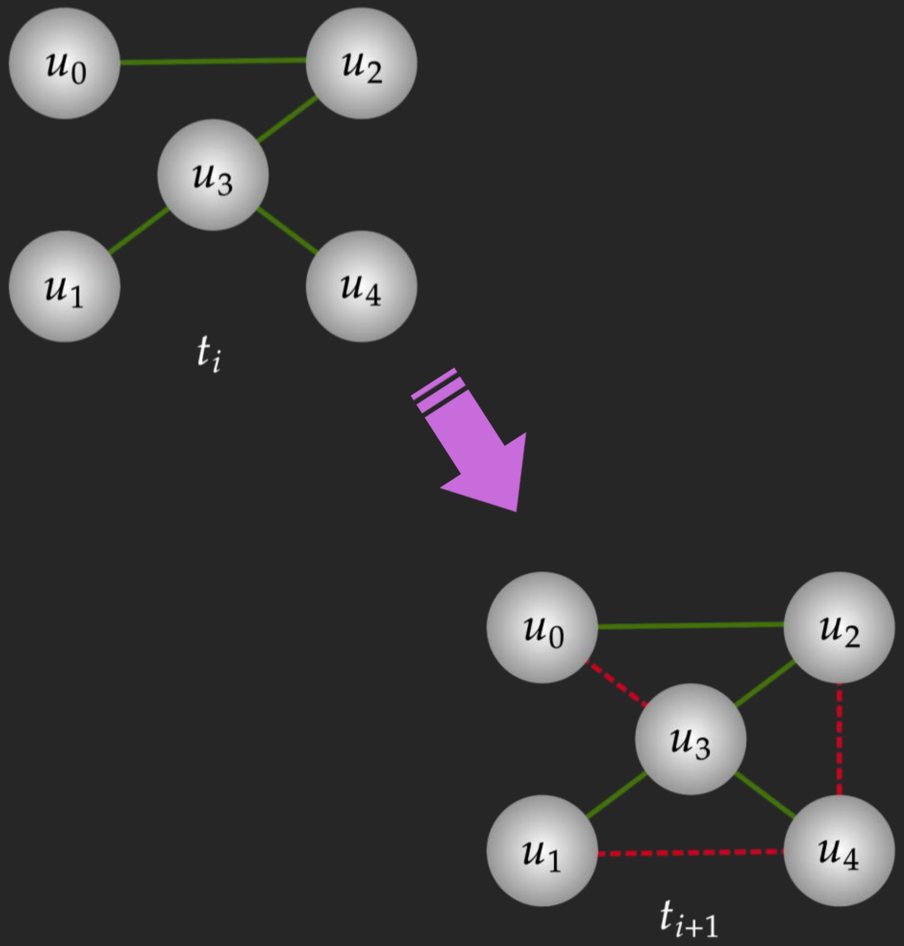
- The Big Five personality model
- Every Big Five dimension is encoded in language
- Questionnaires are time-consuming and impractical



Motivation & Aim

Personality-Aware Link Prediction

- The Link Prediction problem
- Users tend to have their own followee personality preferences
- Incorporate such preferences towards a number of link predictors

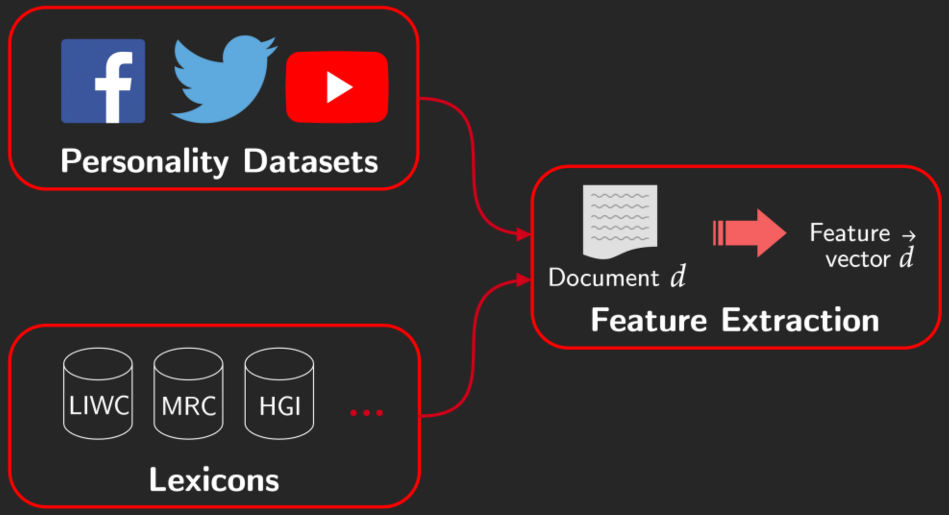


Personality Recognition from Text

Closed vs. Open-vocabulary approaches

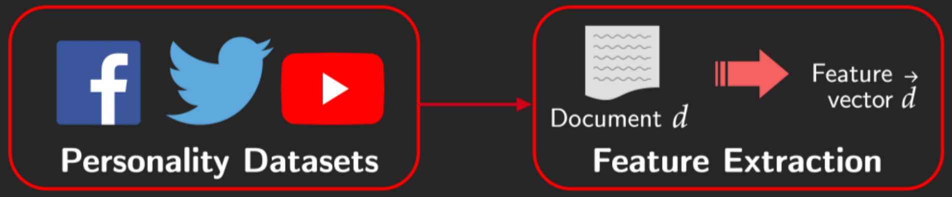
Closed-vocabulary

Using lexicons to derive intermediary features from text



Open-vocabulary

Derives data-driven features (not limited to pre-defined word lists)



Personality Recognition from Text

Methodology

Closed-vocabulary

- A feature extraction component was built from various lexicons
- Regression models were trained using personality-annotated datasets
- Optimisation techniques

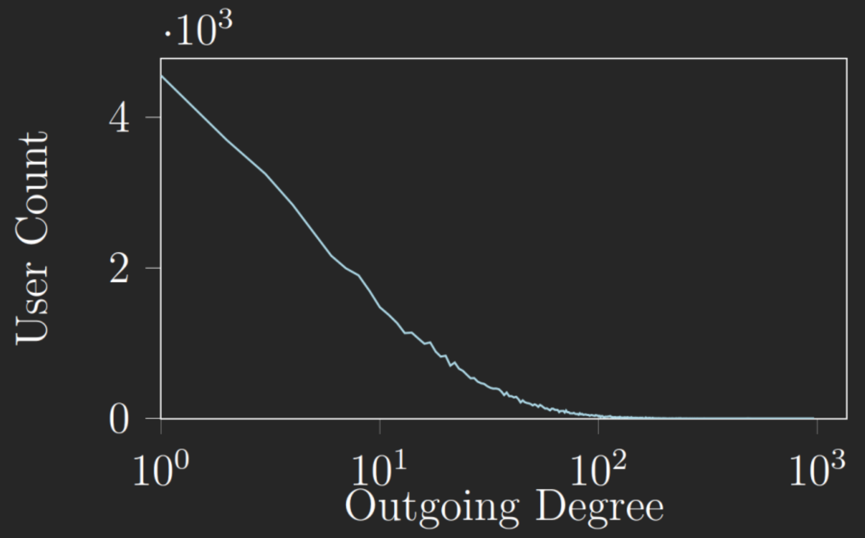
Open-vocabulary

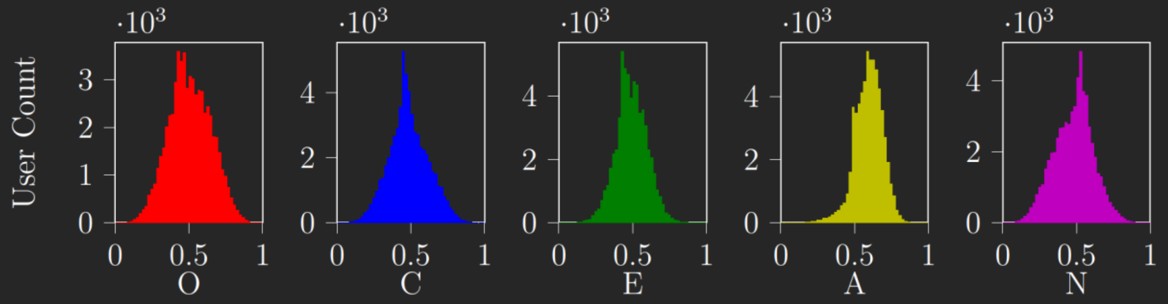
- Adopted the Differential Language Analysis model
- Training and testing were conducted using the same datasets

Personality-Aware Link Prediction

Data Collection

- ‘Twitter-ego’ dataset (containing ~2M edges and ~80K users)
- Utilisied the Twitter API to collect their tweets
- Using the best personality recogniser, the users’ personalities were recognised



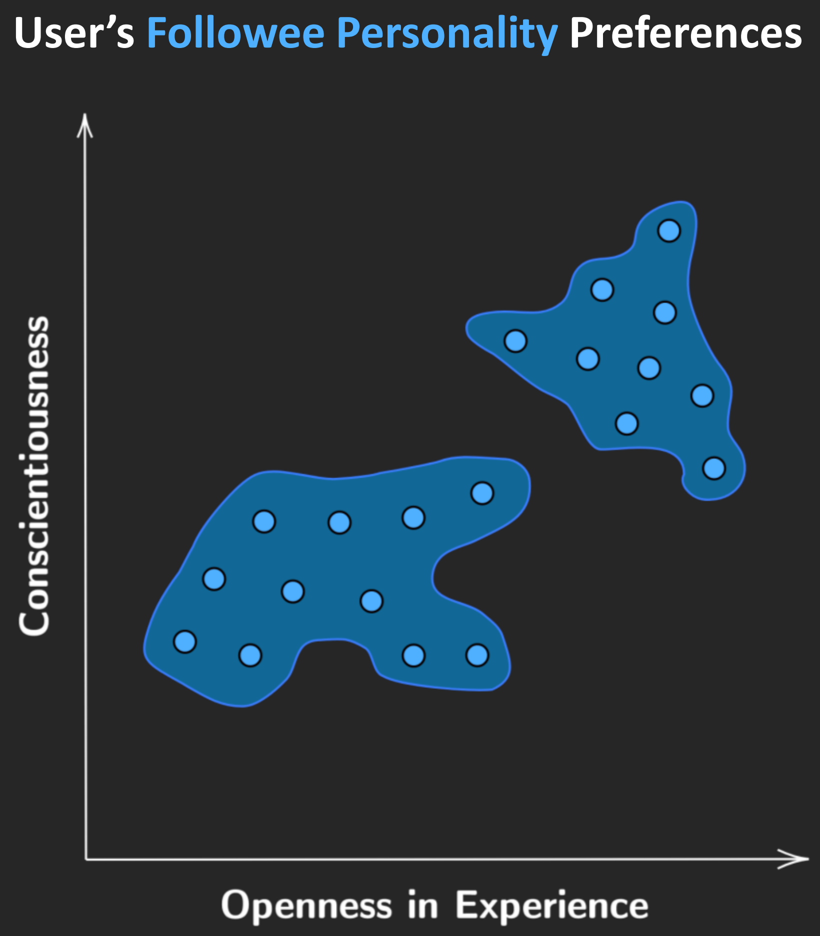


Personality-Aware Link Prediction

Methodology

silhouette scoring

- k-Means algorithm to cluster every user's followees' personalities
- Scores potential followees based on cluster proximity
- Topological and path-based predictors were aggregated with PALP scores



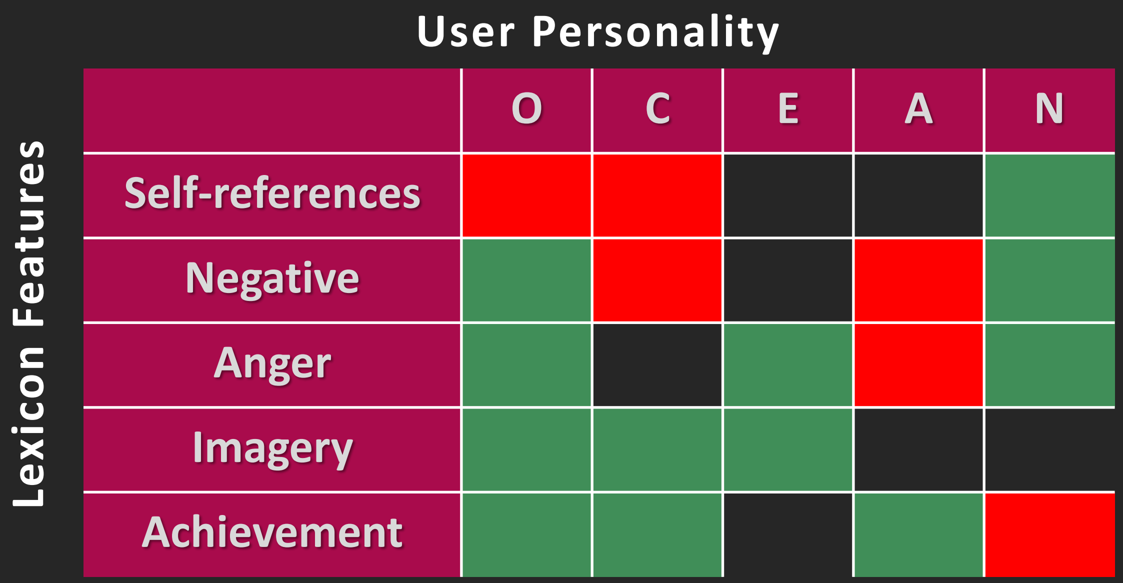
Results & Findings

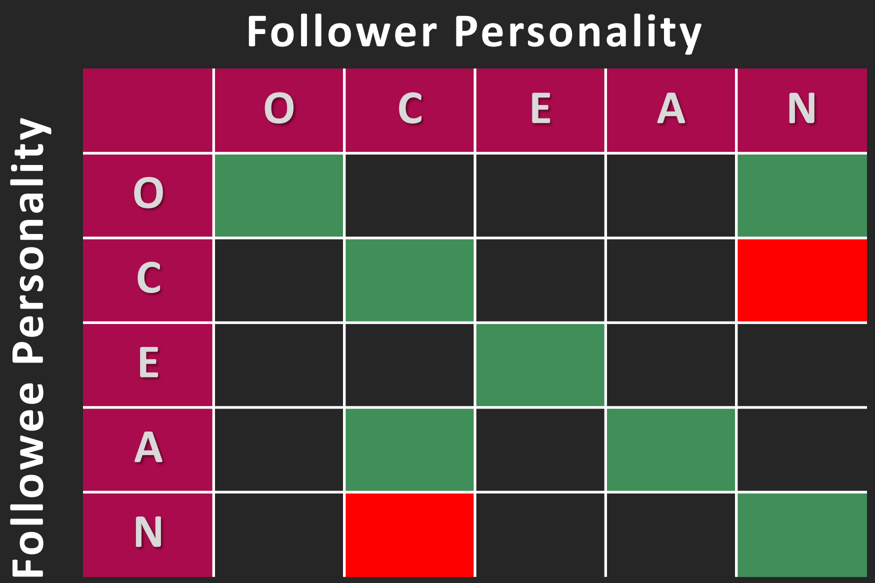
Correlation Analysis

Positive correlation

Negative correlation

Non-significant/small correlation





Results & Findings

Personality Recognition from Text

- The best model was found to be a SVM model with a Pearson VII function kernel

Competitive results to the DLA open-vocabulary alternative

Optimisation results

Hyper-parameter tuning

Cross-media learning

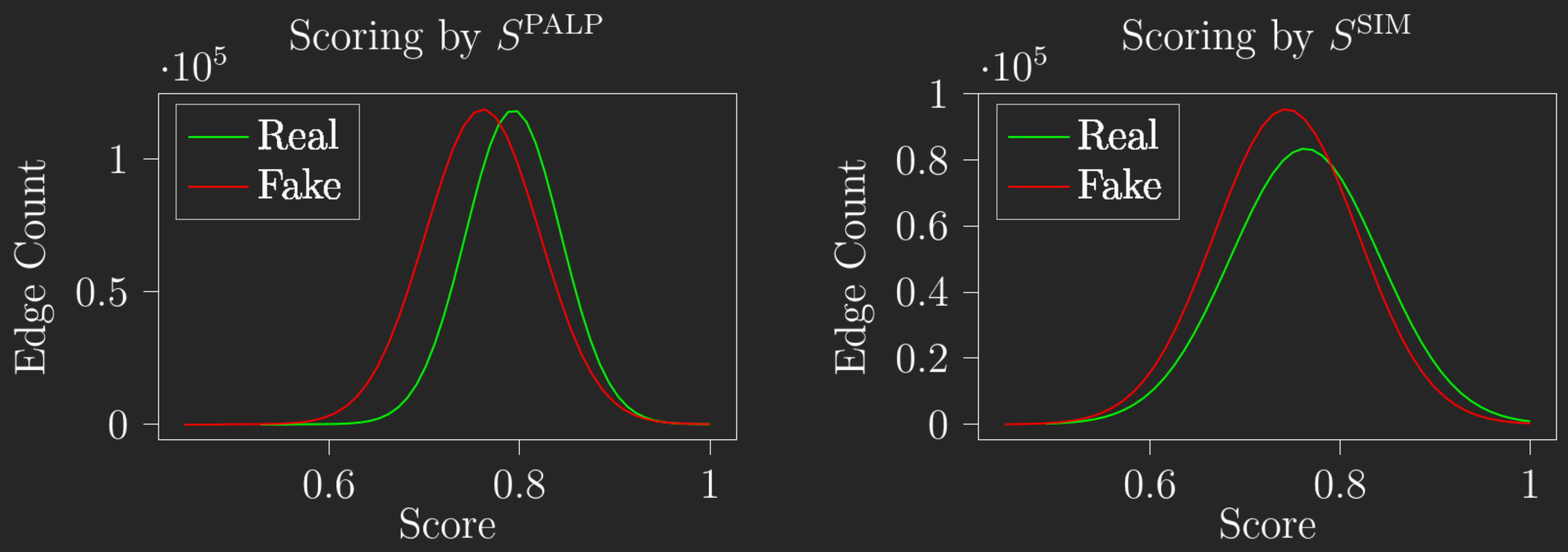
PCA feature construction

Results & Findings

Personality-Aware Link Prediction

Statistical testing determined that

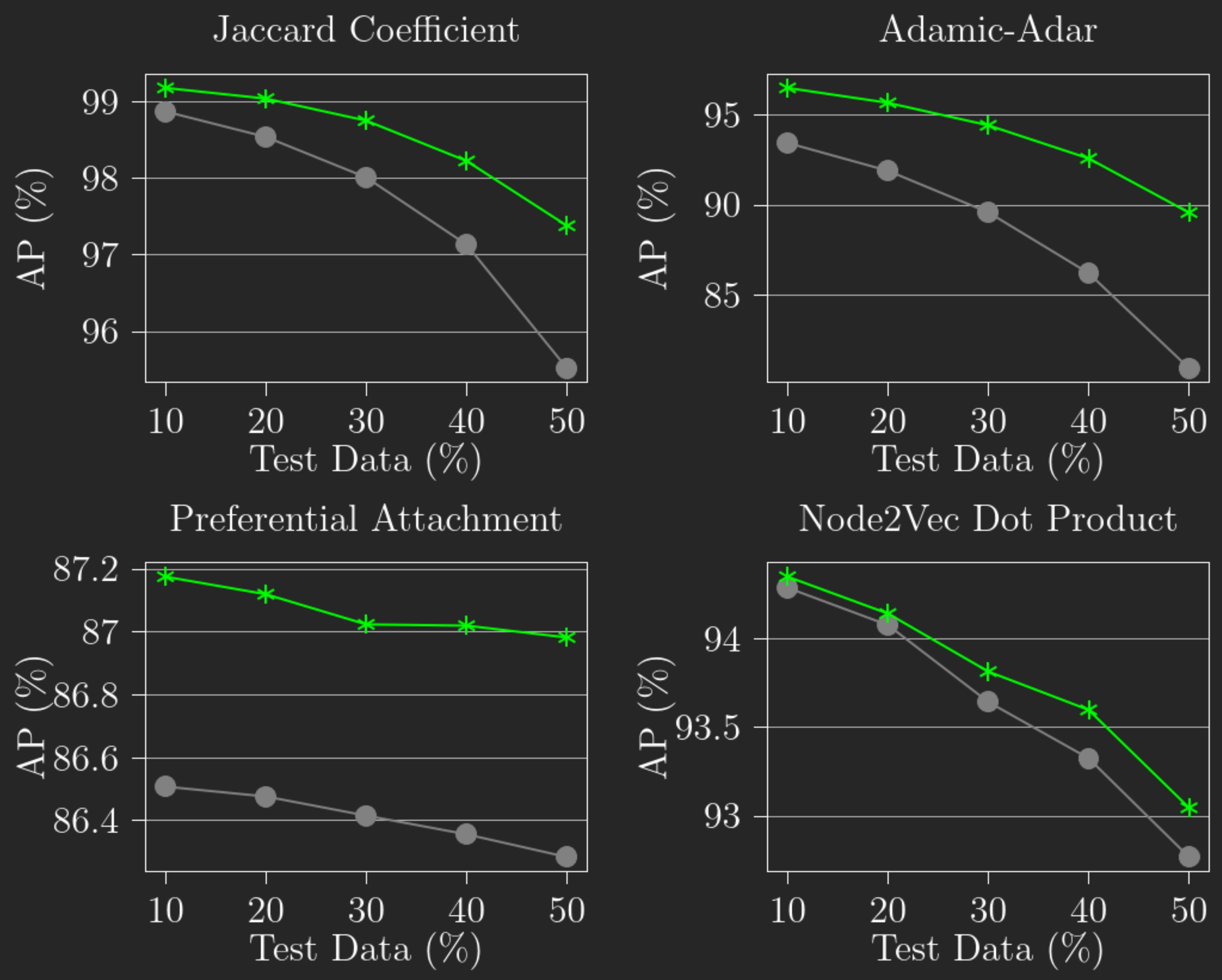
the proposed metric improves over personality similarity



Results & Findings

Personality-Aware Link Prediction

- Employed a variety of dataset splits (10% - 50% test data)
- Improved both Average Precision and Area Under the Curve
- The Adamic-Adar metric experienced an increase of ~10%



Conclusion

…and Future work

Link prediction precision has improved when followee personality preferences were incorporated

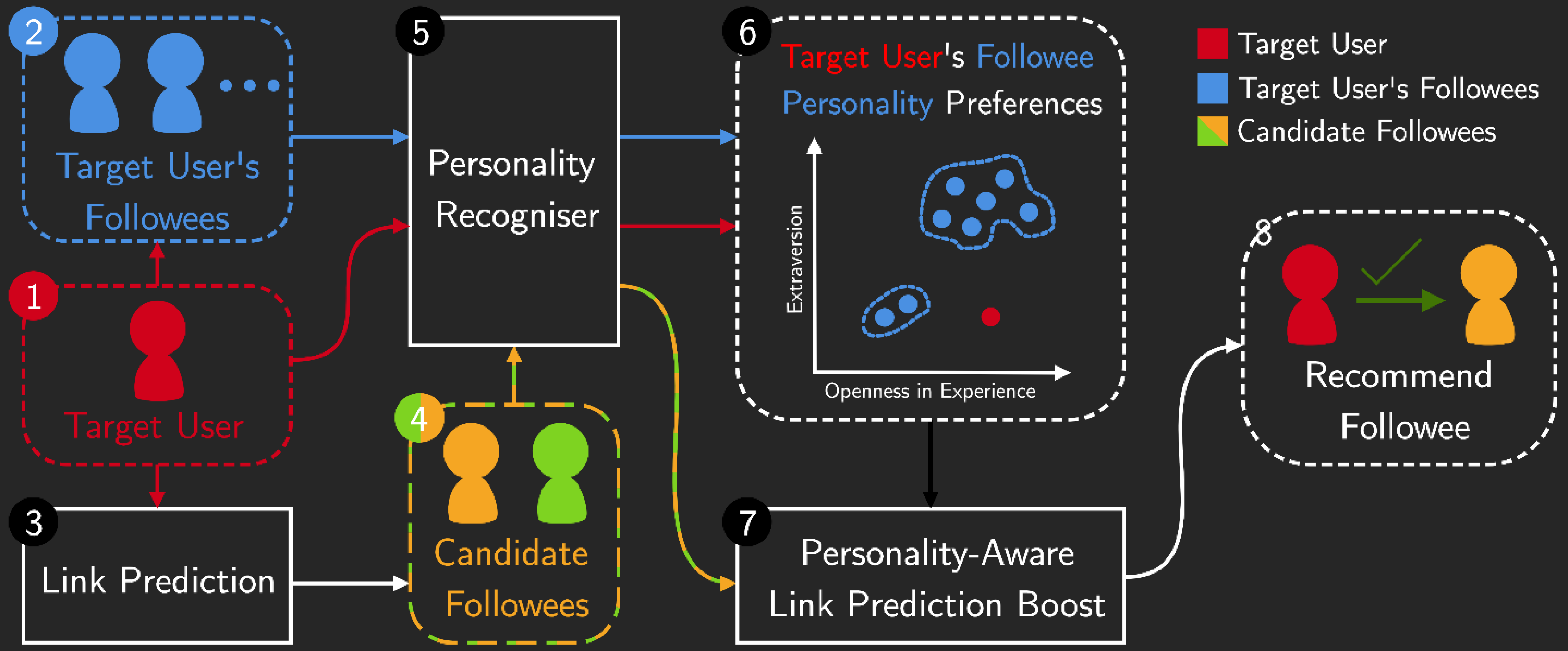
- Employing additional personality-annotated samples
- Substantiate findings on a wide-variety of social networks (Facebook, LinkedIn, etc.)

Thank you for listening!
Any questions?

---

https://github.com/wendru18/**big5-app**
https://github.com/wendru18/**palp-boost**

Supplementary Material



Supplementary Material

$$S^{\mathrm{PALP}}(u_i, u_j) = \sum_{c \in C_i} \frac{\mathrm{weight}(c)}{\mathrm{dist}(c, p_j) + \epsilon}$$

$$S^{\mathrm{SIM}}(u_i, u_j) = \frac{1}{\mathrm{dist}(p_i, p_j) + \epsilon}$$

Supplementary Material

Model	MAE O	C	E	A	N
LR	.131	.137	.157	.139	.168
GP	.121	.124	.141	.124	.150
M5Rules	.120	.123	.144	.125	.147
RF	.112	.113	.135	.113	.135
SVM-POL	.111	.112	.146	.114	.138
SVM-RBF	.111	.112	.131	.113	.134
SVM-PUK	**.105****	**.109***	.129	**.109***	**.130**
DLA [14]	.109	.110	**.121****	.114	.140

Supplementary Material

	Model	Test Data (%) 10	20	30	40	50
AP (%)	JC	98.87	98.54	98.01	97.13	95.52
	•JC	**99.18**	**99.04**	**98.75**	**98.23**	**97.38**
	AA	93.45	91.91	89.61	86.24	80.98
	•AA	96.49	95.66	94.42	92.59	89.59
	PA	86.50	86.47	86.41	86.35	86.28
	•PA	87.17	87.12	87.02	87.02	86.98
	N2V	94.28	94.07	93.64	93.32	92.77
	•N2V	94.34	94.14	93.81	93.59	93.04