Link Assessment in Social Networks Using Multiple Associated Interaction Networks

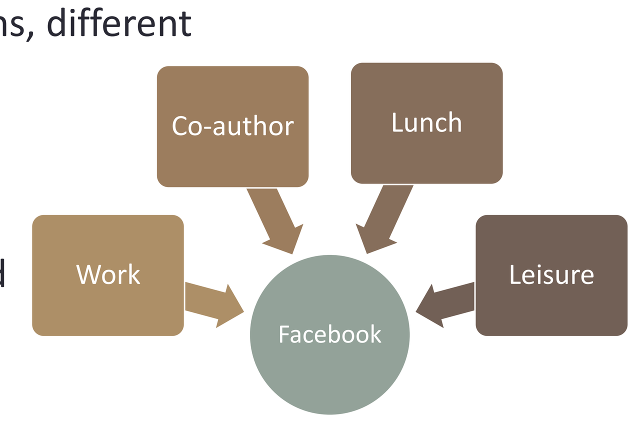
Mohammed Abufouda and Katharina A. Zweig
Graph Theory and Complex network Analysis group, Computer Science Department, University of Kaiserslautern
{abufouda, zweig}@cs.uni-kl.de

ABSTRACT

Many complex network systems suffer from noise that disguises the structure of the network and hinders an accurate analysis of these systems. Link assessment is the process of identifying and eliminating the noise from network systems in order to better understand these systems. In this paper, we address the link assessment problem in social networks that may suffer from noisy relationships. We employed a machine learning classifier for assessing the links in the social network of interest using the data from the associated interaction networks around it. The method was tested with two different data sets: each contains the social network of interest, with ground truth, along with the associated interaction networks. The results showed that it is possible to effectively assess the links of a social network using only the structure of a single network of the associated interaction networks and also using the structure of the whole set of the associated interaction networks. The experiment also revealed that the assessment performance using only the structure of the social network of interest is relatively less accurate than using the associated interaction networks. This indicates that link formation in the social network of interest is not only driven by the internal structure of the social network, but also influenced by the external factors provided in the associated interaction networks.

BACKGROUND AND RESEARCH QUESTIONS

- **Link Assessment Problem**: Given a network, assess the links in this network using its structure and/or using the structure of any extrinsic information (Associated Interaction Networks).
- **Causes:** low cost of making links, automatic invitations, different motivations
- **Homophily: "**Birds of feather flock together"
- **Tie formation process:** is guided by both internal and external homophily [1].



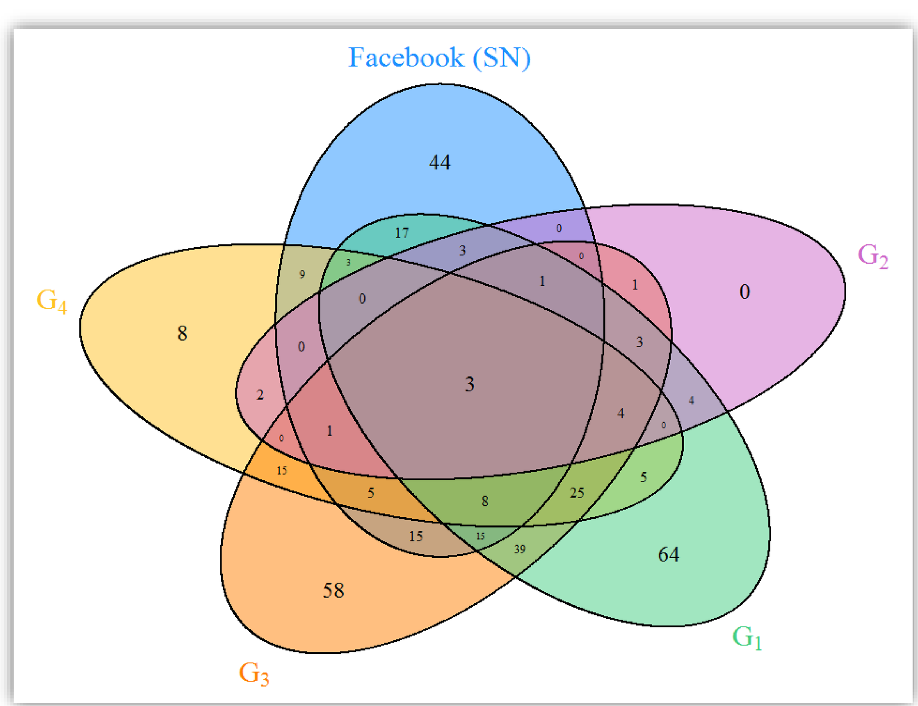


Figure 1: The Venn diagram for edge overlapping between the Facebook social network SN and the other associated interaction networks G for the research group data set.

Research Questions:
- **RQ1:** Does every interaction network encompass sufficient information to assess the social network SN?
- **RQ2**: Does the whole set of the interaction networks provide a better link assessment than using a single interaction network?
- **RQ3**: Which is better for assessing the links of a social network **SN**, using only the structure of a social network **SN** or combining it with the structures of the associated interaction networks?

METHOD

Our aim is to assess the links in a social network using the associated interaction networks of the same actors. We employ machine learning classification techniques by building a feature data model (FDM) of the associated interaction networks and the social network of interest. The idea of the proposed framework is to convert the link assessment problem into a machine learning classification problem where the classifier should indicate whether a particular link is noise or a real link via learning from the associated interaction networks. Figure 2 shows the FDM framework components.

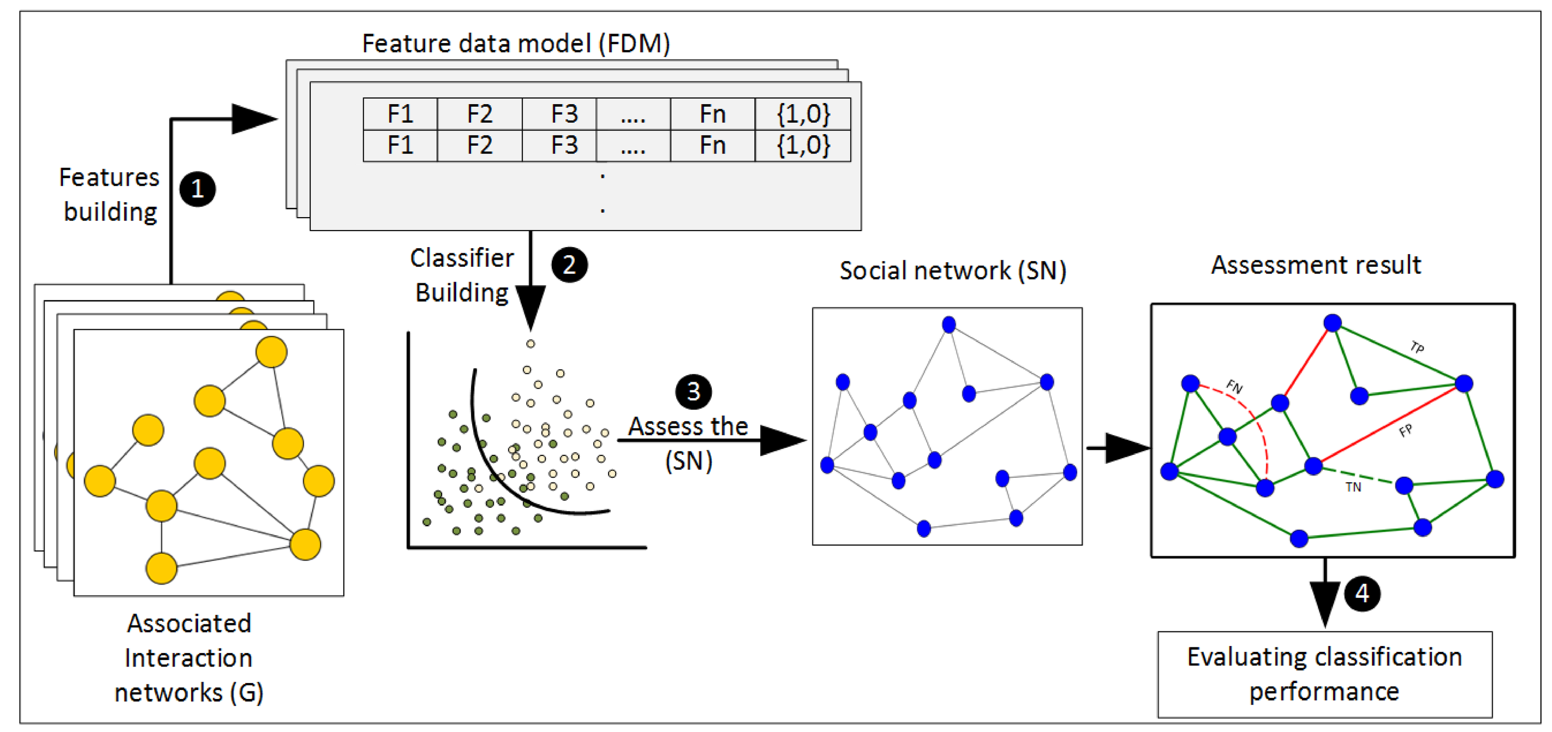


Figure 2: The proposed framework for link assessment using associated interaction networks

Table 1: Model Features (Nodes pair-wise dependent features)

Common Neighbors	$\mathcal{CN}(v,w) = \lvert\Gamma(v)\cap\Gamma(w)\rvert$
Resource Allocation	$\mathcal{RA}(v,w) = \sum_{\substack{z\in\{\Gamma(v)\cap\Gamma(w)\}\\ z\neq v\neq w}} \frac{1}{\lvert\Gamma(z)\rvert}$
Adamic-Adar coefficient	$\mathcal{AAC}(v,w) = \sum_{\substack{z\in\{\Gamma(v)\cap\Gamma(w)\}\\ z\neq v\neq w}} \frac{1}{log\lvert\Gamma(z)\rvert}$
Jaccard index	$\mathcal{JI}(v,w) = \frac{\lvert\Gamma(v)\cap\Gamma(w)\rvert}{\lvert\Gamma(v)\cup\Gamma(w)\rvert}$
Preferential Attachment	$\mathcal{PA}(v,w) = \lvert\Gamma(v)\rvert\cdot\lvert\Gamma(w)\rvert$

Table 2: Data sets

- ***RG*** is the research group data set
- ***LF*** is a law firm offline social network

dataset	Networks	n	m	$cc(G_i)$	$\eta(G_i)$
RG	SN **Facebook**	32	248	0.48	0.5
	G_1 Work	60	338	0.34	0.19
	G_2 Co-author	25	42	0.43	0.14
	G_3 Lunch	60	386	0.57	0.21
	G_4 Leisure	47	176	0.34	0.16
LF	SN **Friends**	69	339	0.43	0.07
	G_1 Co-work	71	726	0.41	0.15
	G_2 Advice	71	717	0.42	0.14

RESULTS

Table 3: **Link assessment of the SN using single network for both the real data and the corresponding random networks of the same degree sequence**

Data Set	Real Data			Random Networks		
	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$
RG	$\psi(\mathrm{FDM}_{G_1})$ =0.83	0.83	0.83	0.01	0.024	0.02
	$\psi(\mathrm{FDM}_{G_2})$ =0.78	0.44	0.43	≈ 0	≈ 0	≈ 0
	$\psi(\mathrm{FDM}_{G_3})$ =0.83	0.84	0.84	0.02	0.036	0.03
	$\psi(\mathrm{FDM}_{G_4})$ =0.83	0.81	0.82	0.035	0.027	0.03
LF	$\psi(\mathrm{FDM}_{G_1})$ =0.89	0.91	0.89	0.098	0.167	0.13
	$\psi(\mathrm{FDM}_{G_2})$ =0.84	0.91	0.87	0.08	0.14	0.1

RESULTS (CONT'D)

Table 4: **Link assessment of the SN using the interaction networks and the social network itself**

Data set	$\mathrm{FDM}_{\mathcal{G}}$			FDM_{SN}		
	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$
$\psi(\mathrm{FDM}_{RG})$	0.84	0.84	0.84	0.82	0.83	0.82
$\psi(\mathrm{FDM}_{LF})$	0.89	0.90	0.90	0.84	0.87	0.85

Table 5: **Classifiers benchmarking for the link assessment using Logistic Regression, Naïve Bayes, Random Forests, and Alternating Decision Trees**

Data set	Classifier	Prediction performance				
		$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	$\mathcal{ACC}$	AU-ROC
RG	$\mathcal{LR}$	0.834	0.86	0.834	83.47	0.85
	$\mathcal{NB}$	0.83	0.78	0.78	77.62	0.85
	$\mathcal{RF}$	0.79	0.79	0.79	79.032	0.79
	$\mathcal{ADT}$	0.8	0.798	0.8	79.84	0.77
LF	$\mathcal{LR}$	0.84	0.915	0.874	91.5	0.84
	$\mathcal{NB}$	0.89	0.90	0.82	90.54	0.85
	$\mathcal{RF}$	0.89	0.91	0.89	91.7	0.83
	$\mathcal{ADT}$	0.89	0.92	0.89	91.67	0.79

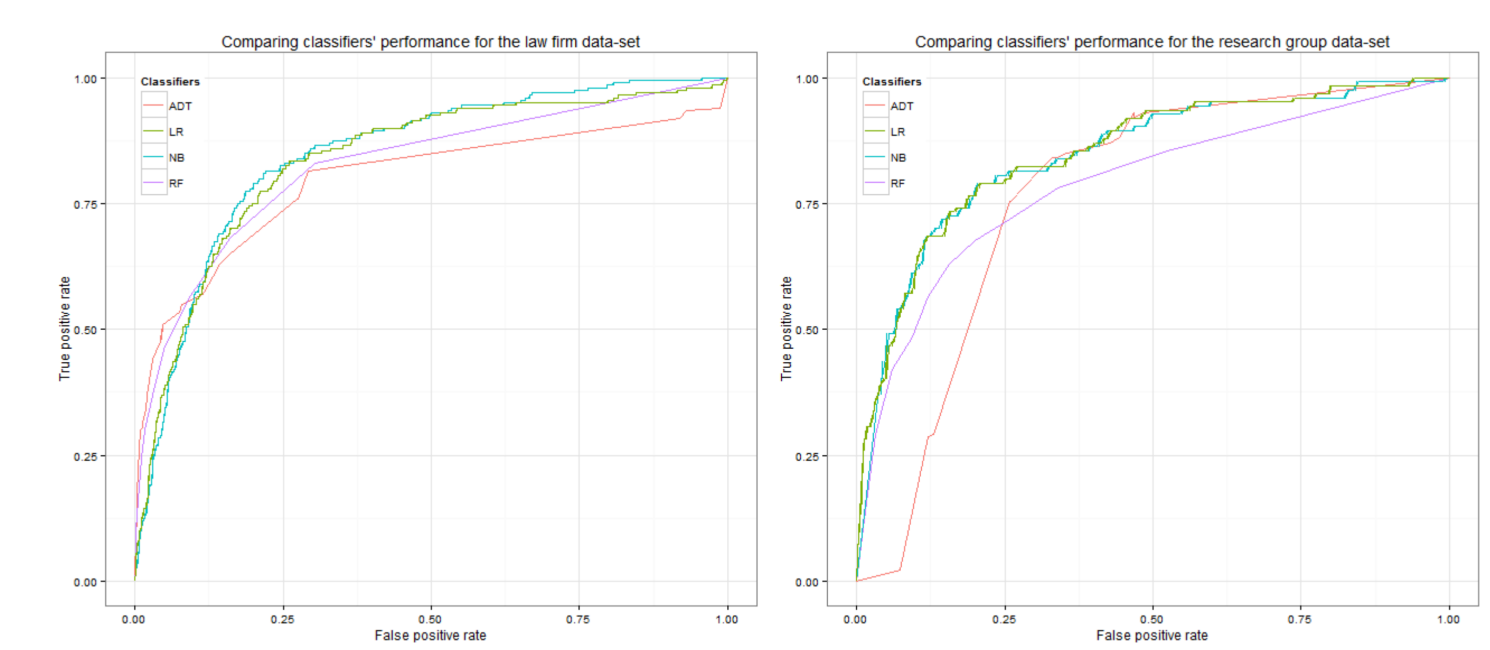


Figure 3: Classifiers' performance for the research group data set (left) and the law firm data set (right)

CONCLUSIONS

- The results suggest a correlation between the social network and the interaction networks.
- This correlation guides the tie formation process in the social network.
- More likes were correctly classified when using all of the associated interaction networks, which means that the external factors affects the structure of the social network.
- The used datasets have small number of nodes, but have a social network with ground truth.
- More information about the work in this poster is available in [2].

REFERENCES

[1] M. Abufouda and K. A. Zweig, "*Interactions around social networks matter: Predicting the social network from associated interaction networks*," *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, Beijing, 2014, pp. 142-145.
[2] M. Abufouda and K. A. Zweig, "*Are we Really Friends?: Link Assessment in Social Networks using Multiple Associated Interaction Networks*", Proceedings of the 24th International Conference on World Wide Web, 2015 ACM International Conference on, 2015, pp. 771--776.