

Network Denoising in Social Media

Huiji Gao

Computer Science and Engineering, Arizona State University, USA

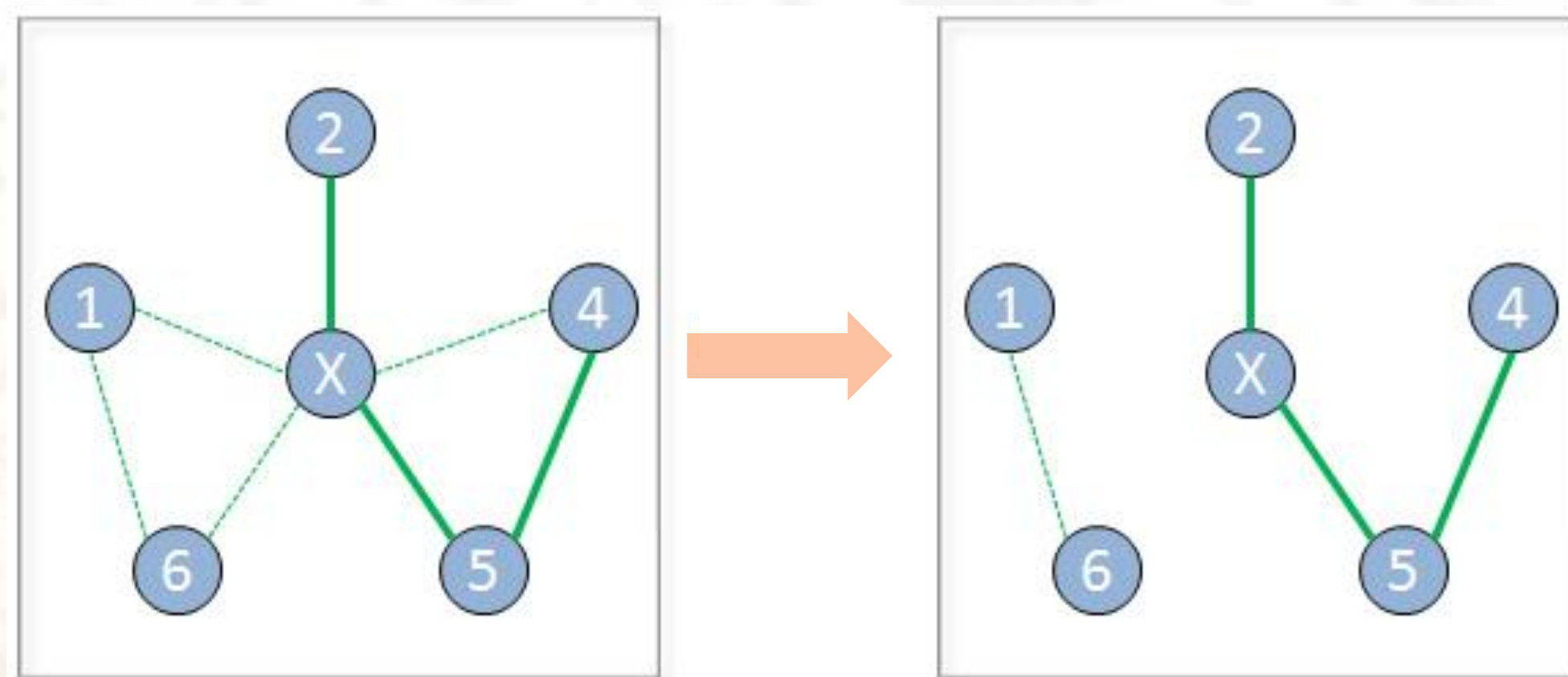
Huiji.Gao@asu.edu



Network Denoising

Social media expands the ways people communicate with each other. On a popular social media website, a user typically has hundreds of contacts (or friends) on average. As a person's social network grows, friend management is increasingly important for effective communications. Often one can only afford to maintain close friendship in a small scale due to limited time and other resources. In other words, the majority of one's connections are so-so friends and do not hold strong influence on the user. We propose to denoise the individuals' social network by removing noisy links.

Methodology



$$\min_{w_i \geq 0} \sum_{i=1}^n (\|N_i w_i - f_i\|_2^2 + \lambda \|w_i\|_1)$$

$$\min_{w_i \geq 0} \sum_{i=1}^n (\alpha_1^2 \|N_i^1 w_i - f_i^1\|_2^2 + \alpha_2^2 \|N_i^2 w_i - f_i^2\|_2^2 + \lambda e^T w_i)$$

$$\min_{w_i \geq 0} \sum_{i=1}^n (\|\theta_i w_i - F_i\|_2^2 + \lambda e^T w_i)$$

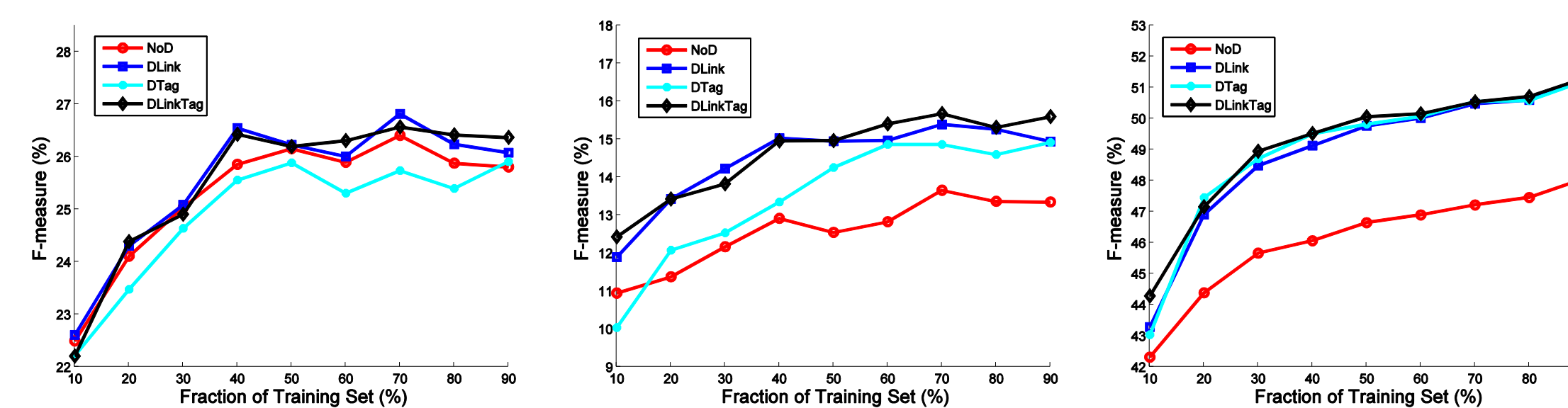
$$\theta_i = (\alpha_1 N_i^1, \alpha_2 N_i^2, \dots, \alpha_m N_i^m)^T \quad F_i = (\alpha_1 f_i^1, \alpha_2 f_i^2, \dots, \alpha_m f_i^m)^T$$

Behavioral Inference

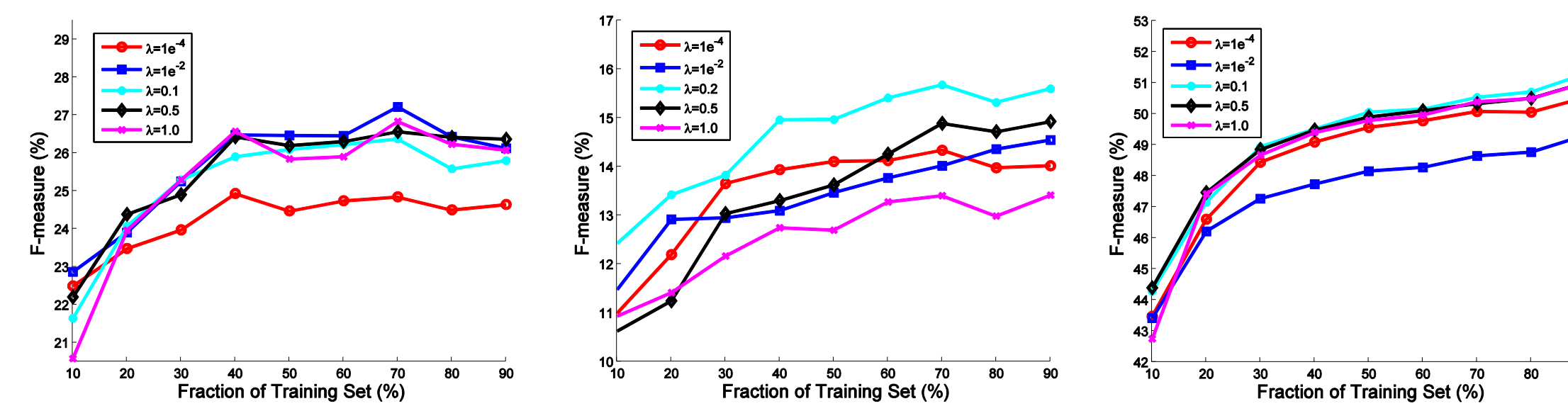
Social Media Datasets

Dataset	BlogCatalog	Flickr	BlogMI
Users	8,797	8,465	6,069
Edges	290,059	195,847	523,642
Unique Tags	7,418	7,303	5,161
Classes	59	169	70
Density	7.5×10^{-3}	5.5×10^{-3}	2.8×10^{-2}
Avg. Degree	66	46	173

Denoising vs No Denoising



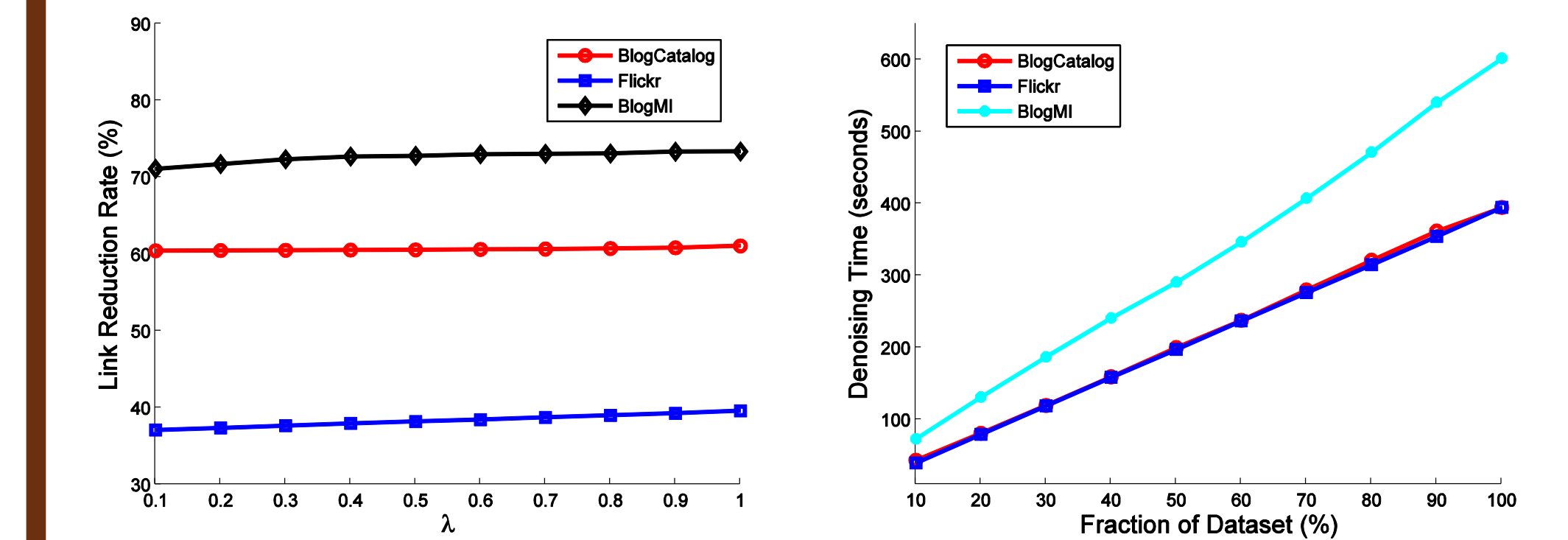
Denoising Performance vs λ



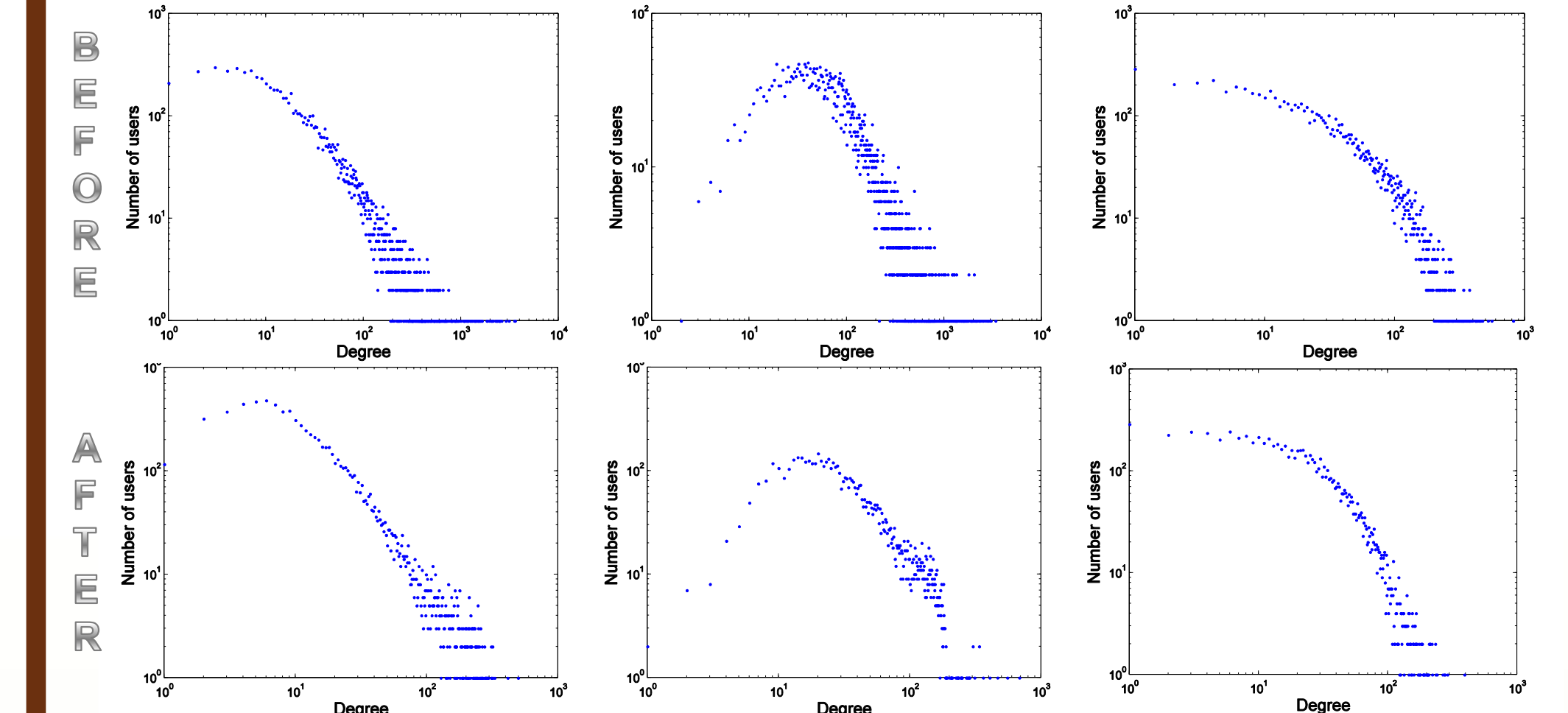
Denoising with Multiple Interactions

Datasets	Methods	F-Measure (%)	Link Reduction	Time (second)
BlogCatalog	No Denoising	25.81	290,059	101.92
	Linking	26.08 (+0.7%)	112,559 (-61.19%)	36.43 (-65.49)
	Tagging	25.91 (+0.4%)	39,775 (-86.29%)	47.44 (-54.48)
	Linking + Tagging	26.37 (+2.2%)	114,337 (-60.58%)	46.65 (-55.27)
Flickr	No Denoising	48.06	195,847	85.62
	Linking	51.28 (+7%)	96,442 (-50.76%)	64.67 (-20.95)
	Tagging	51.20 (+6.53%)	107,247 (-44.24%)	66.75 (-18.87)
	Linking + Tagging	51.29 (+6.72%)	123,130 (-37.13%)	67.07 (-18.55)
BlogMI	No Denoising	13.35	523,642	106.1
	Linking	14.94 (+11.91%)	137,261 (-73.79%)	32.34 (-73.76)
	Tagging	14.92 (+11.76%)	38,575 (-92.63%)	24.26 (-81.84)
	Commenting	15.36 (+15.06%)	49,139 (-90.62%)	54.13 (-51.97)
	Linking + Tagging	15.60 (+16.85%)	147,881 (-71.76%)	35.96 (-70.14)
	Tagging + Commenting	15.36 (+15.06%)	75,027 (-85.67%)	46.68 (-59.42)
Linking + Commenting	15.49 (+16.03%)	155,004 (-70.40%)	28.77 (-77.33)	
Linking + Tagging + Commenting	14.81 (+10.94%)	141,200 (-73.04%)	32.47 (-73.63)	

Link Reduction & Scalability



Statistics before and after Denoising



Measures	Ave. Degree		Clustering Coefficient		Ave. Sharing Tags	
	No Denoising	Denoising	No Denoising	Denoising	No Denoising	Denoising
BlogCatalog	66	26	0.46	0.20	0.27	0.40
Flickr	46	29	0.13	0.12	24.43	28.95
BlogMI	173	49	0.39	0.12	0.24	0.39

Conclusions and Future work

- Propose an efficient approach to denoise social networks for friend management intensifies.
- Good performance on behavioral inference, link reduction, and time efficiency.
- Integrating all actions vs. selectively integrating actions

This work is, in part, supported by ONR