[ONLINE APPENDIX]

UNDERSTANDING THE CONTRIBUTION OF RECOMMENDATION ALGORITHMS ON MISINFORMATION RECOMMENDATION AND MISINFORMATION DISSEMINATION ON SOCIAL NETWORKS

Royal Pathak[1], Francesca Spezzano[1], and Maria Soledad Pera[2]

[1] ***Boise State University,*** [2] ***Technische Universiteit Delft, The Netherlands***

Table 1. Number of superspreaders and non-superspreaders in each dataset when thresholds for the definition of superspreaders are changed.

	Politifact FakeNewsNet dataset		HealthStory FakeHealth dataset	
Threshold	Number of Superspreaders	Number of non-superspreaders	Number of Superspreaders	Number of non-superspreaders
50%	639	389	281	5125
60%	636	392	92	5314
70%	603	425	51	5355
80%	547	481	9	5397
90%	424	604	0	5406
100%	305	723	0	5406

Table 2. Average in-degree for superspreaders and non-superspreaders users and for various thresholds used to define superspreaders, and statistical significance test results to check whether the average in-degree is the same between superspreaders and non-superspreaders (Politifact FakeNewsNet dataset).

	Superspreaders	Non-Superspreaders	Hypothesis Test	
Threshold	In-degree	In-degree	p-value	Test outcomes
50%	6.362	3.955	0.002	Significant difference
60%	6.356	3.987	0.002	Significant difference
70%	6.396	4.153	0.003	Significant difference
80%	6.479	4.356	0.004	Significant difference
90%	6.325	4.918	0.069	Not significant difference
100%	5.690	3.955	0.036	Significant difference

Table 3. Average in-degree for superspreaders and non-superspreaders users and for various thresholds used to define superspreaders, and statistical significance test results to check whether the average in-degree is the same between superspreaders and non-superspreaders (HealthStory FakeHealth dataset).

	Superspreaders	Non-Superspreaders	Hypothesis Test	
Threshold	In-degree	In-degree	p-value	Test outcomes
50%	1.274	1.654	0.435	Not significant difference
60%	1.681	1.636	0.955	Not significant difference
70%	2.560	1.628	0.382	Not significant difference
80%	5.001	1.633	0.272	Not significant difference

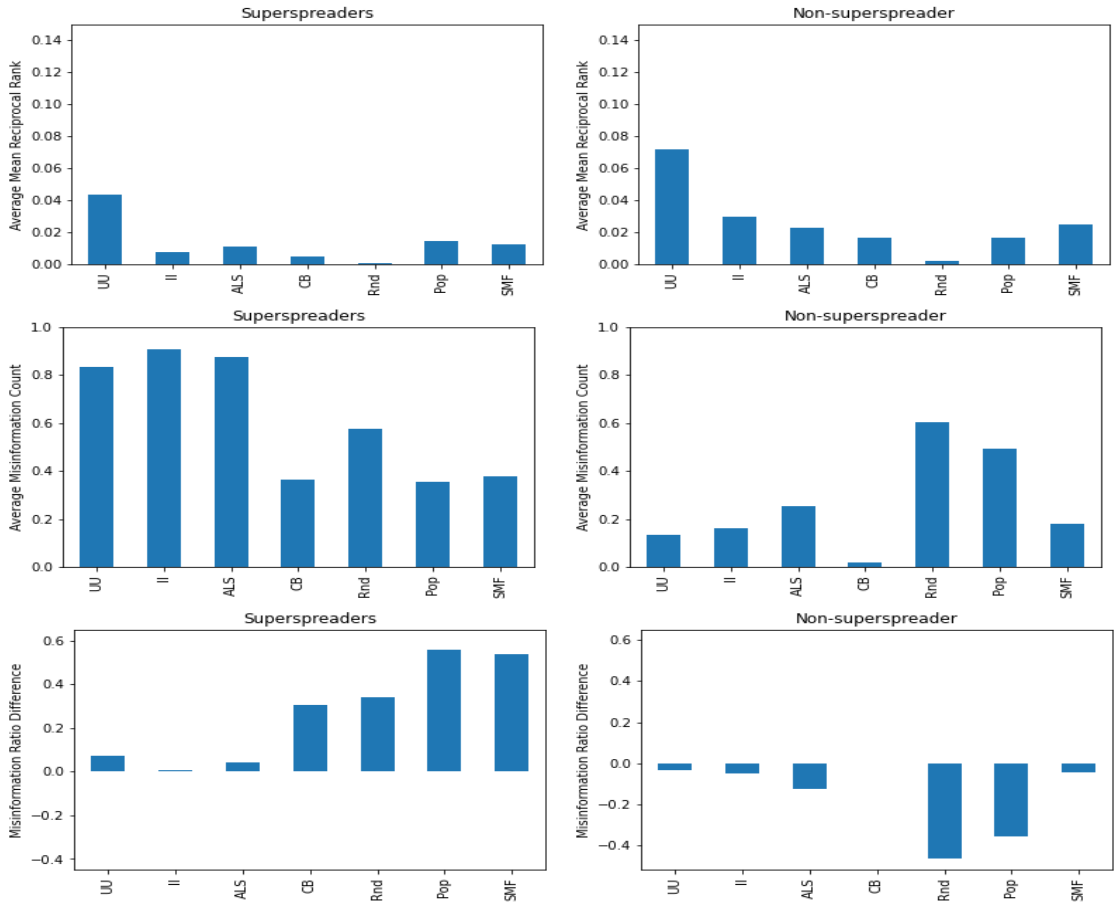

Fig. 9. Average MRR, Average MC, and MRD for each of RA, for superspreaders (defined by $\theta = 50\%$), and non-superspreaders in the Politifact FakeNewsNet dataset.

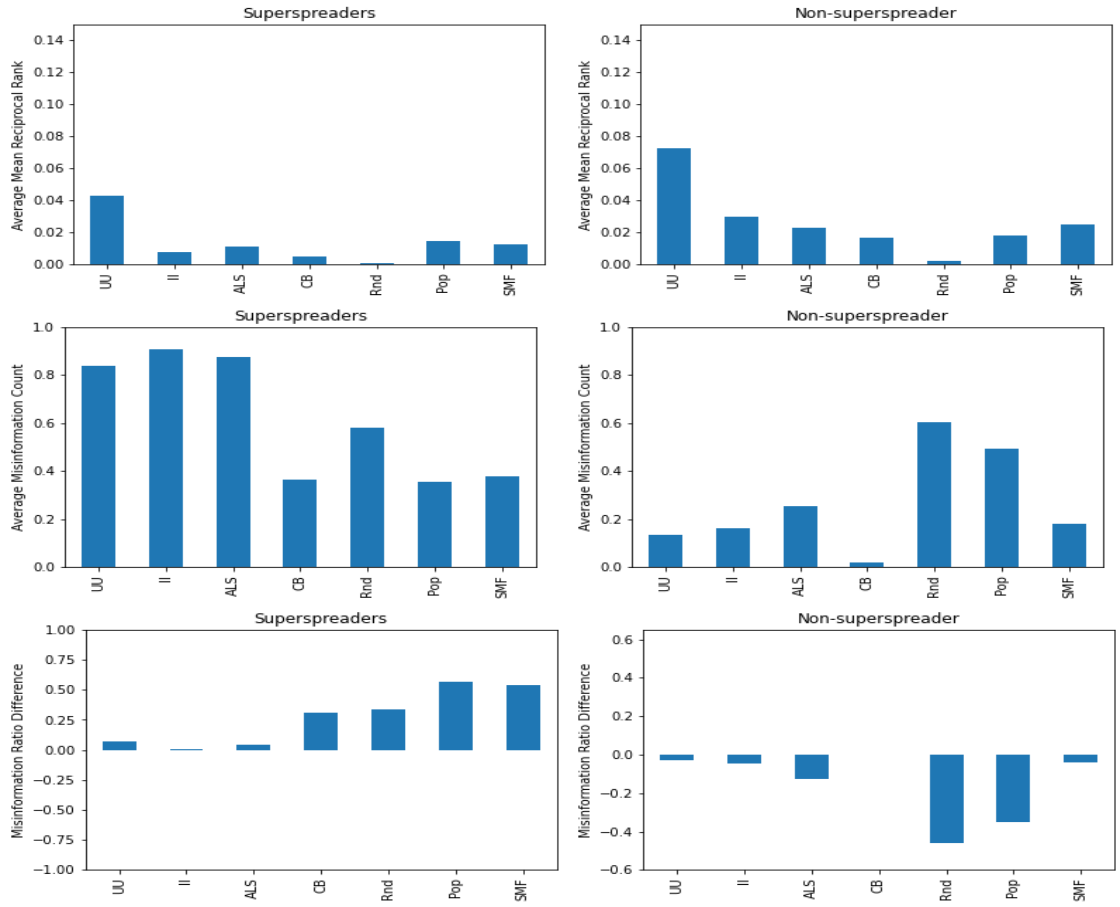

Fig. 10. Average MRR, Average MC, and MRD for each of RA, for superspreaders (defined by $\theta = 60\%$), and non-superspreaders in the Politifact FakeNewsNet dataset.

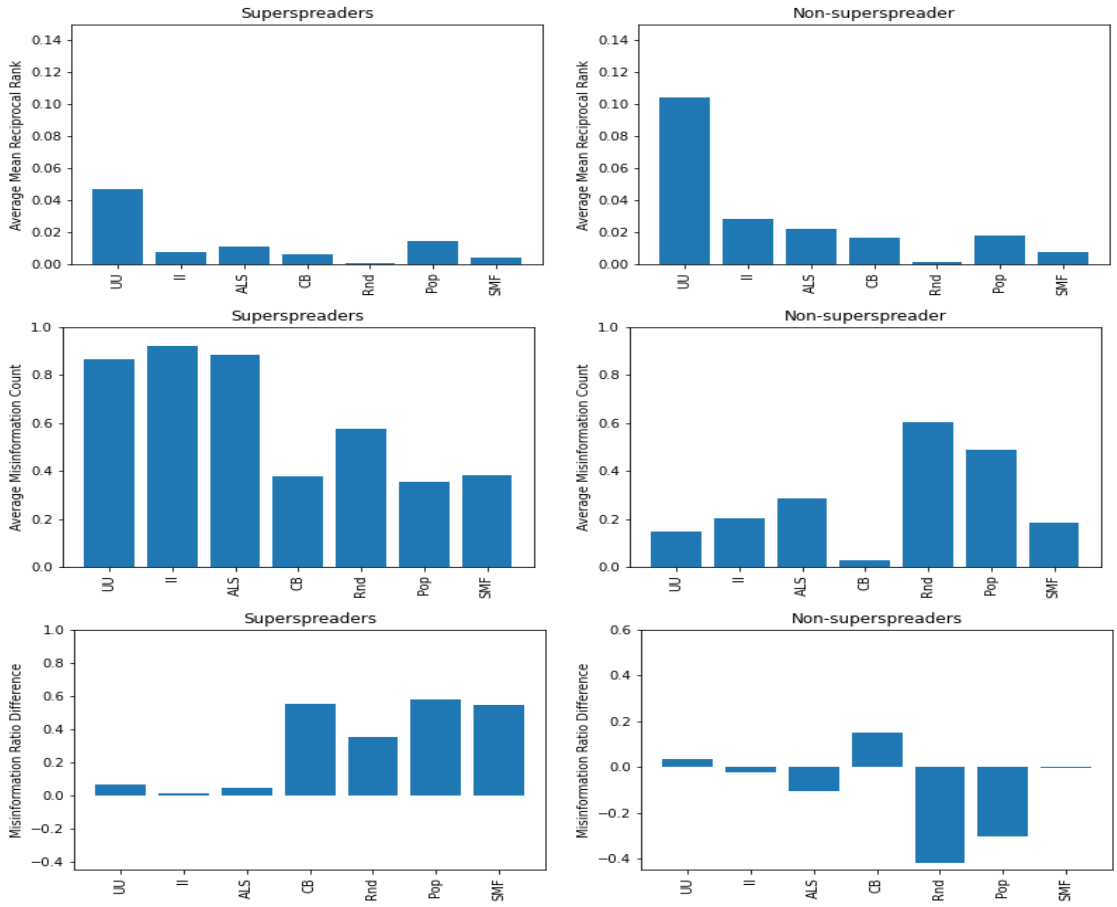

Fig. 11. Average MRR, Average MC, and MRD for each of RA, for superspreaders (defined by $\theta = 70\%$), and non-superspreaders in the Politifact FakeNewsNet dataset.

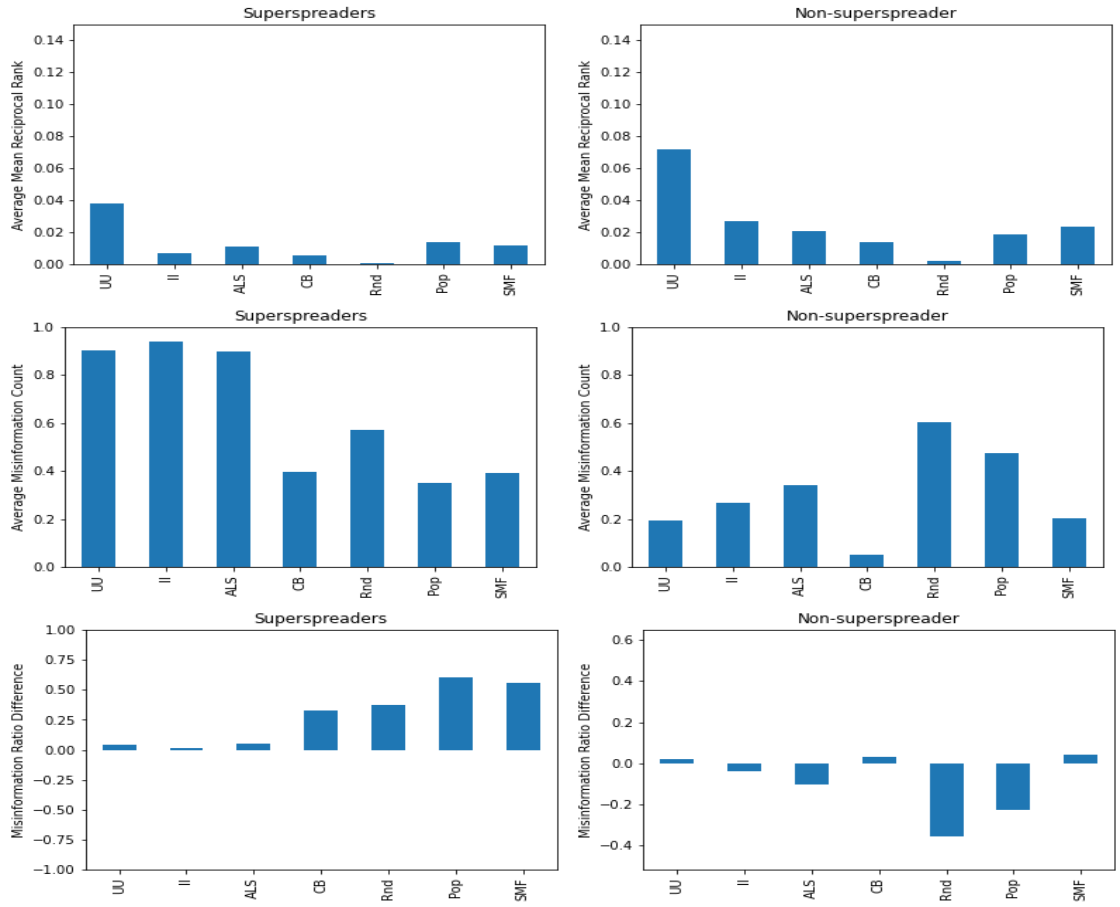

Fig. 12. Average MRR, Average MC, and MRD for each of RA, for superspreaders (defined by $\theta = 80\%$), and non-superspreaders in the Politifact FakeNewsNet dataset.

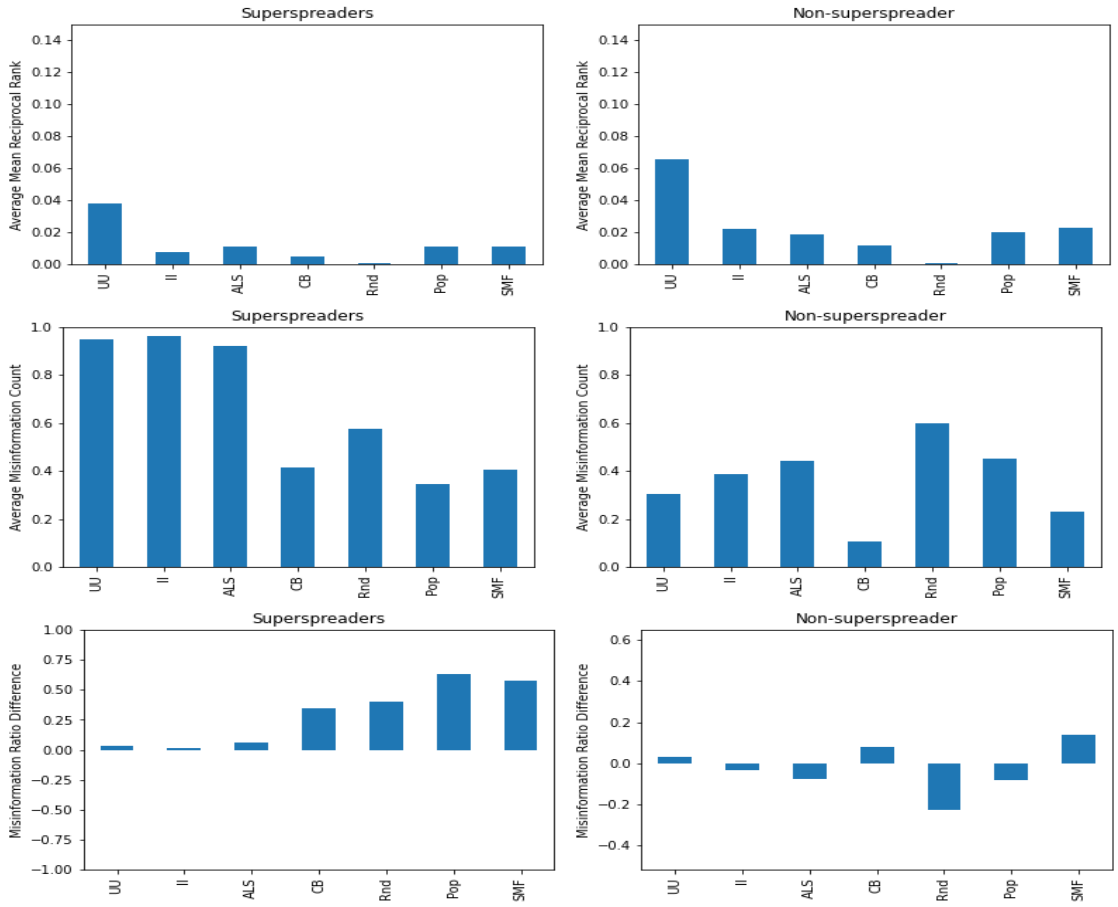

Fig. 13. Average MRR, Average MC, and MRD for each of RA, for superspreaders (defined by $\theta = 90\%$), and non-superspreaders in the Politifact FakeNewsNet dataset.

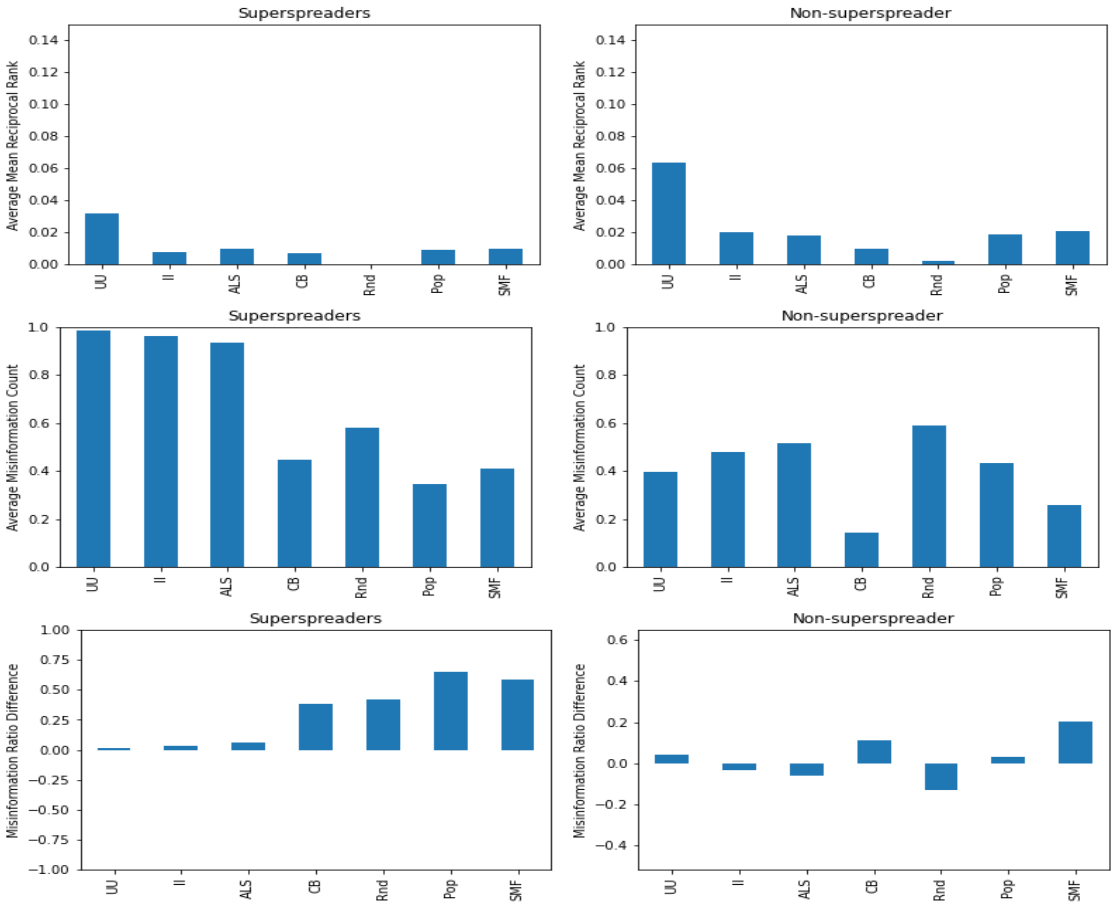

Fig. 14. Average MRR, Average MC, and MRD for each of RA, for superspreaders (defined by $\theta = 100\%$), and non-superspreaders in the PolitiFact FakeNewsNet dataset.

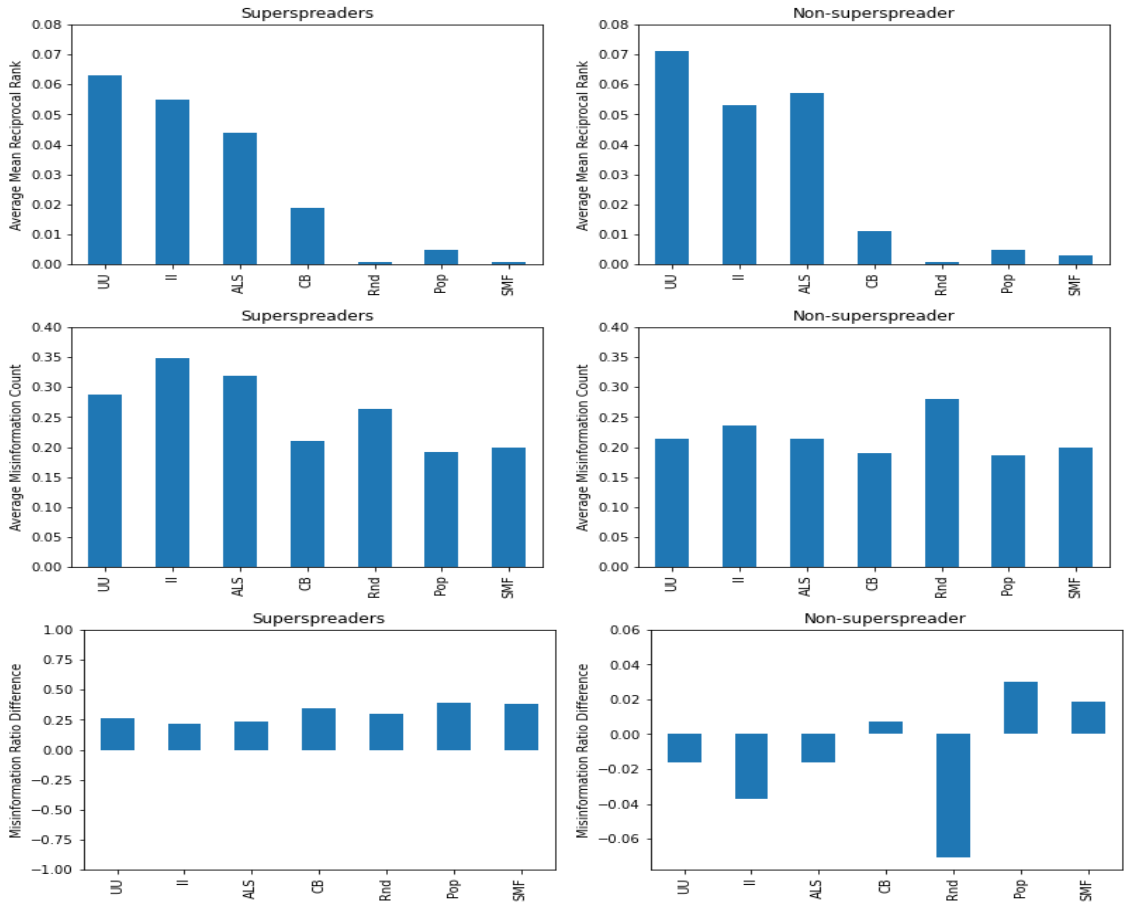

Fig. 15. Average MRR, Average MC, and MRD for each of RA, for superspreaders (defined by $\theta = 50\%$), and non-superspreaders in the HealthStory FakeHealth dataset.

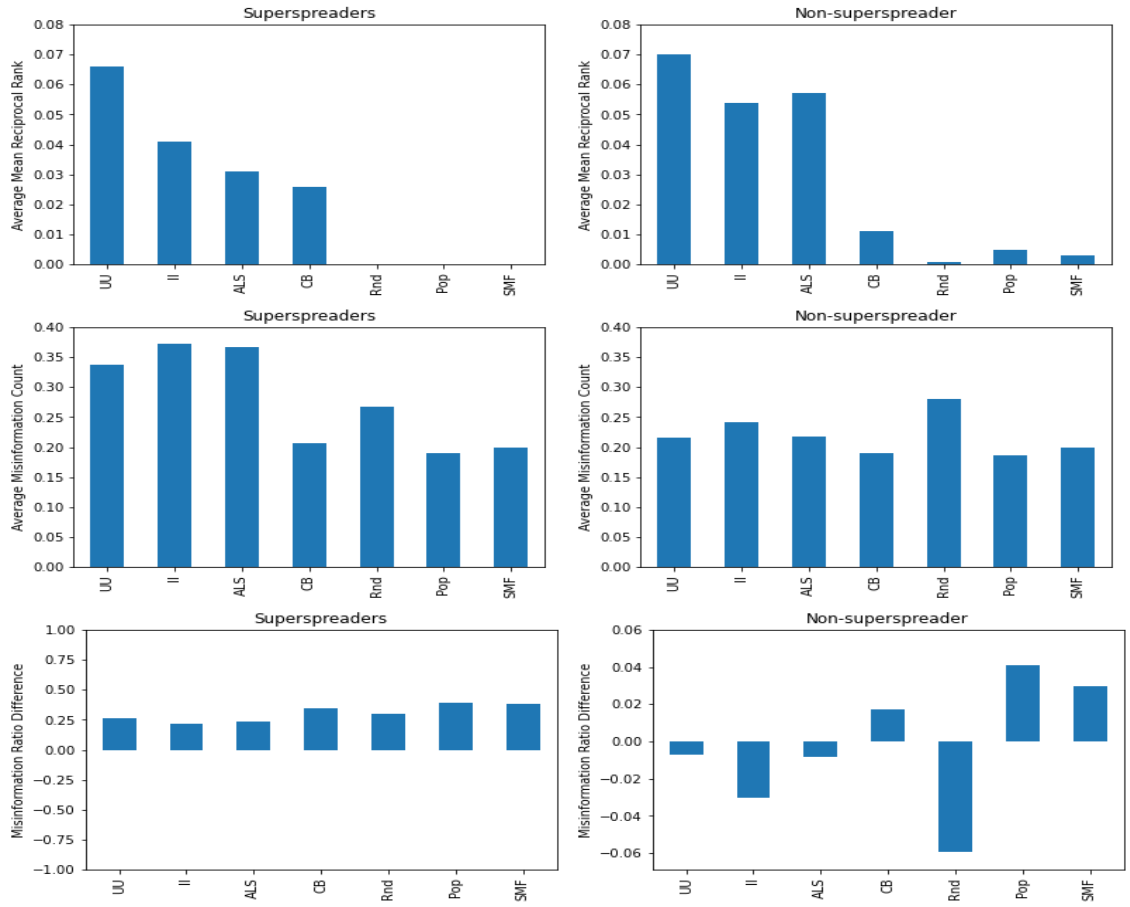

Fig. 16. Average MRR, Average MC, and MRD for each of RA, for superspreaders (defined by $\theta = 60\%$), and non-superspreaders in the HealthStory FakeHealth dataset..

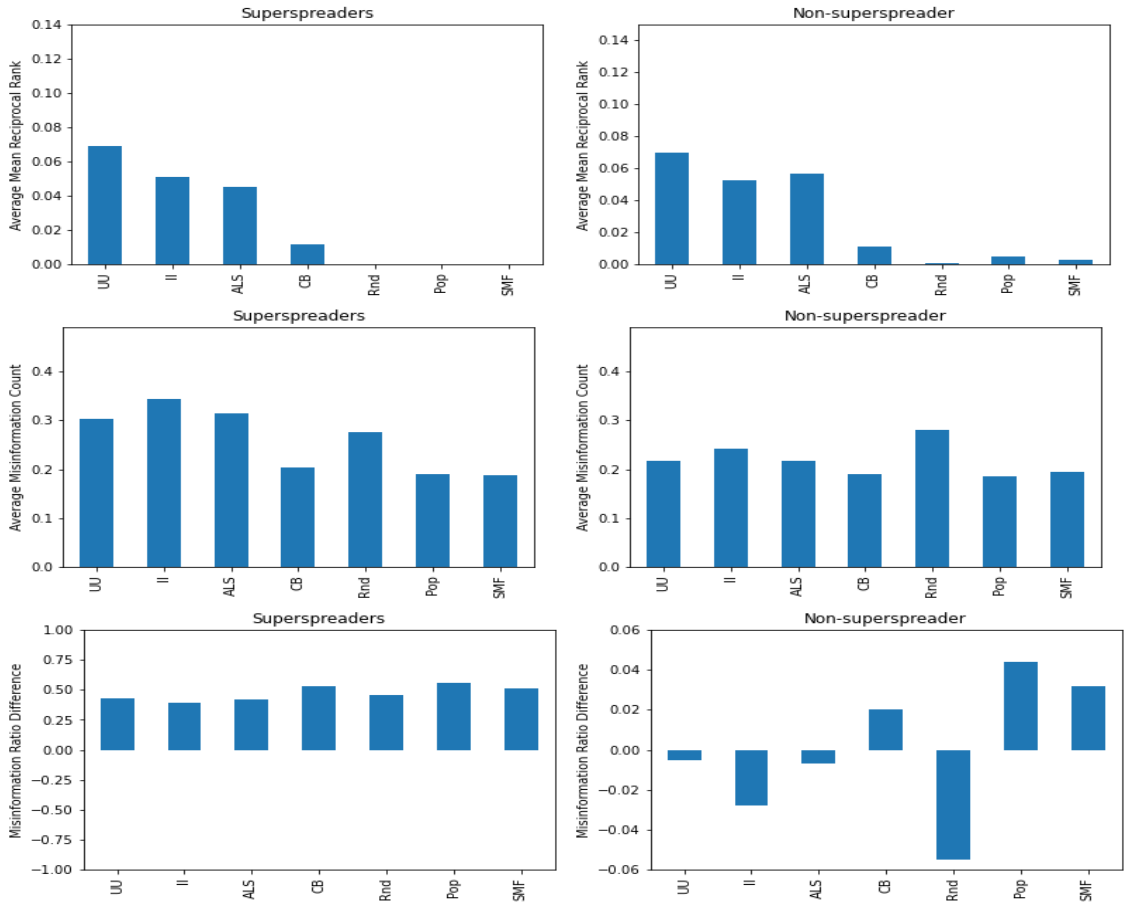

Fig. 17. Average MRR, Average MC, and MRD for each of RA, for superspreaders (defined by $\theta = 70\%$), and non-superspreaders in the HealthStory FakeHealth dataset.

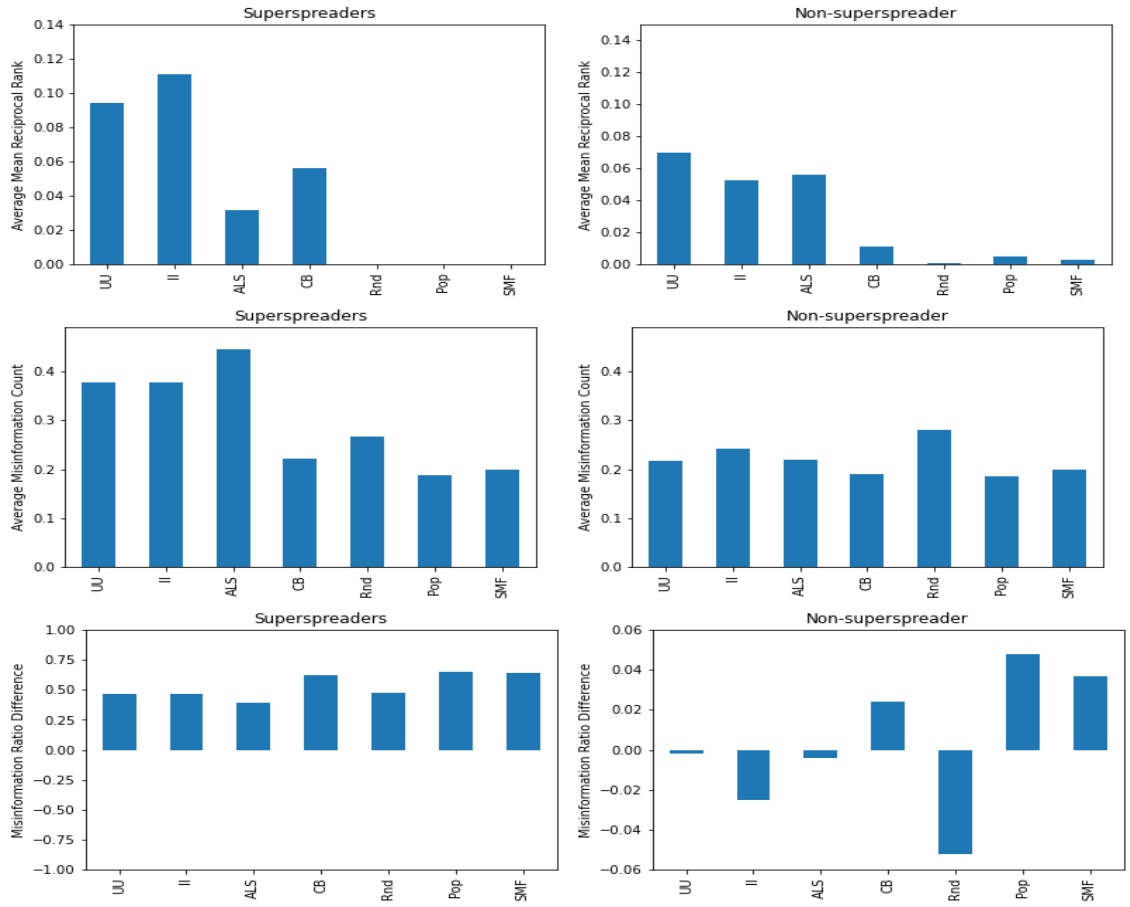

Fig. 18. Average MRR, Average MC, and MRD for each of RA, for superspreaders (defined by $\theta = 80\%$), and non-superspreaders in the HealthStory FakeHealth dataset.

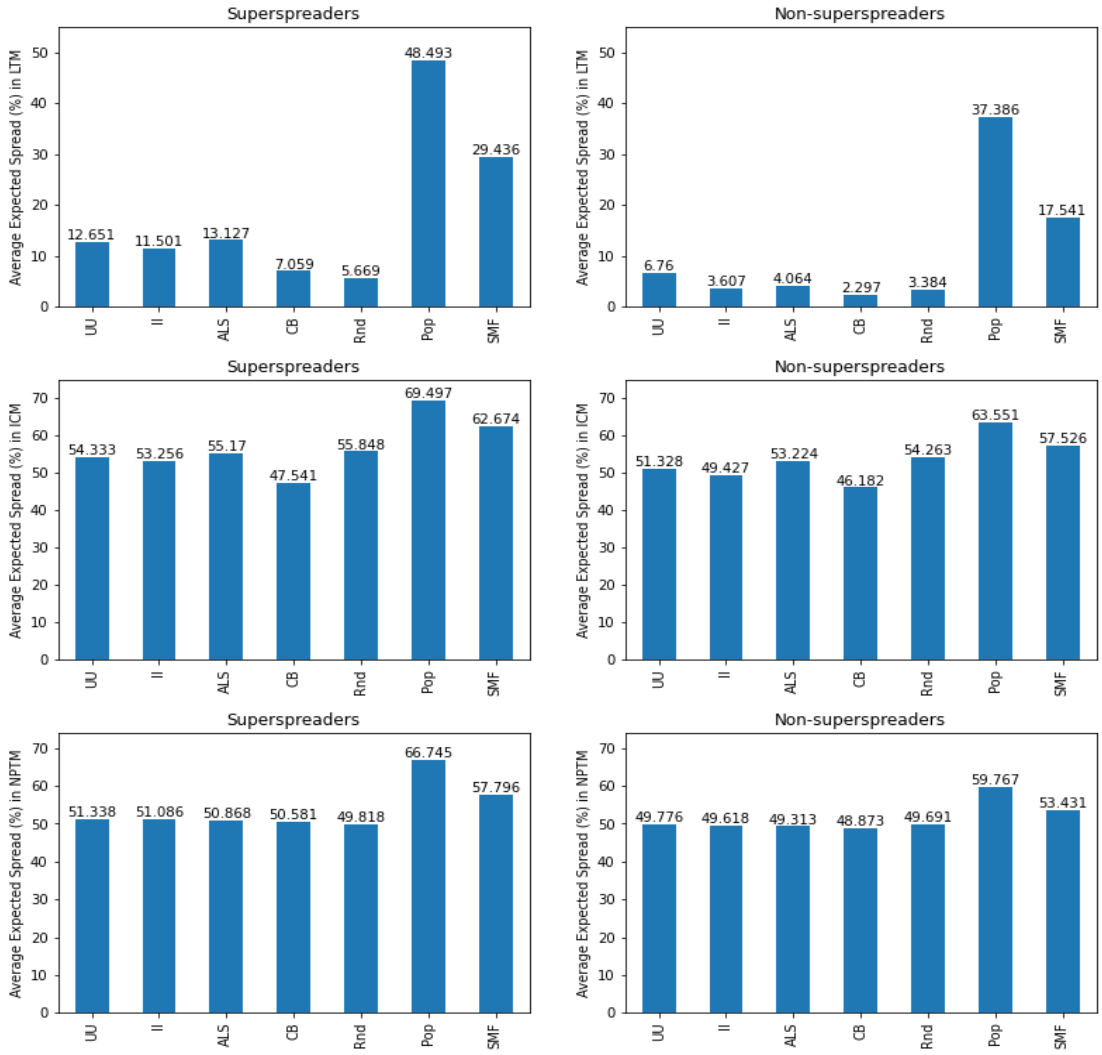

Fig. 19. Average expected spread for each of RA, for all users, superspreaders (defined by $\theta = 50\%$), and non-superspreaders in the Politifact FakeNewsNet dataset according to the Linear Threshold model (LTM – top row), Independent Cascade model (ICM – middle row), and the Node Profile Threshold Model (NPTM – bottom row).

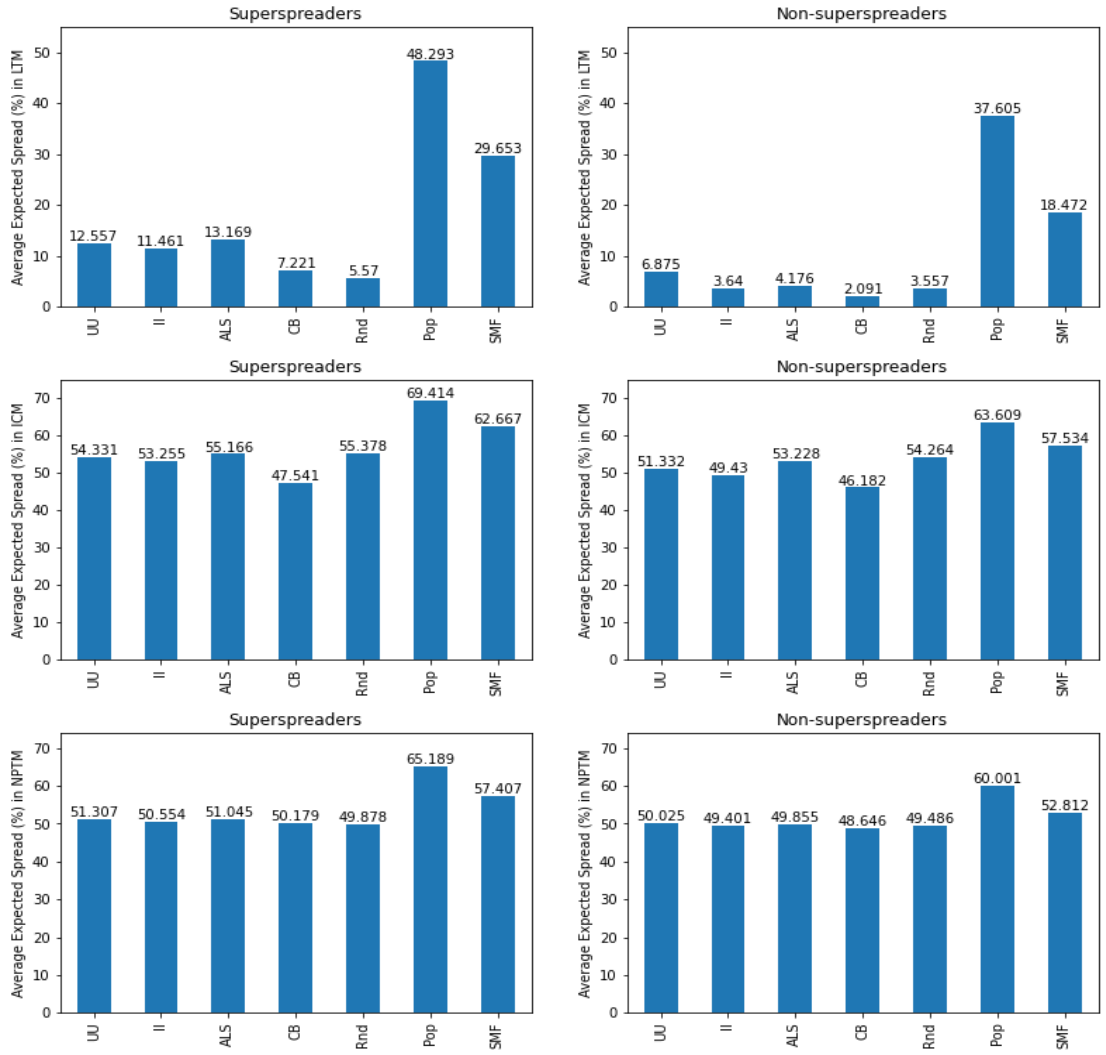

Fig. 20. Average expected spread for each of RA, for all users, superspreaders (defined by $\theta = 60\%$), and non-superspreaders in the Politifact FakeNewsNet dataset according to the Linear Threshold model (LTM – top row), Independent Cascade model (ICM – middle row), and the Node Profile Threshold Model (NPTM – bottom row).

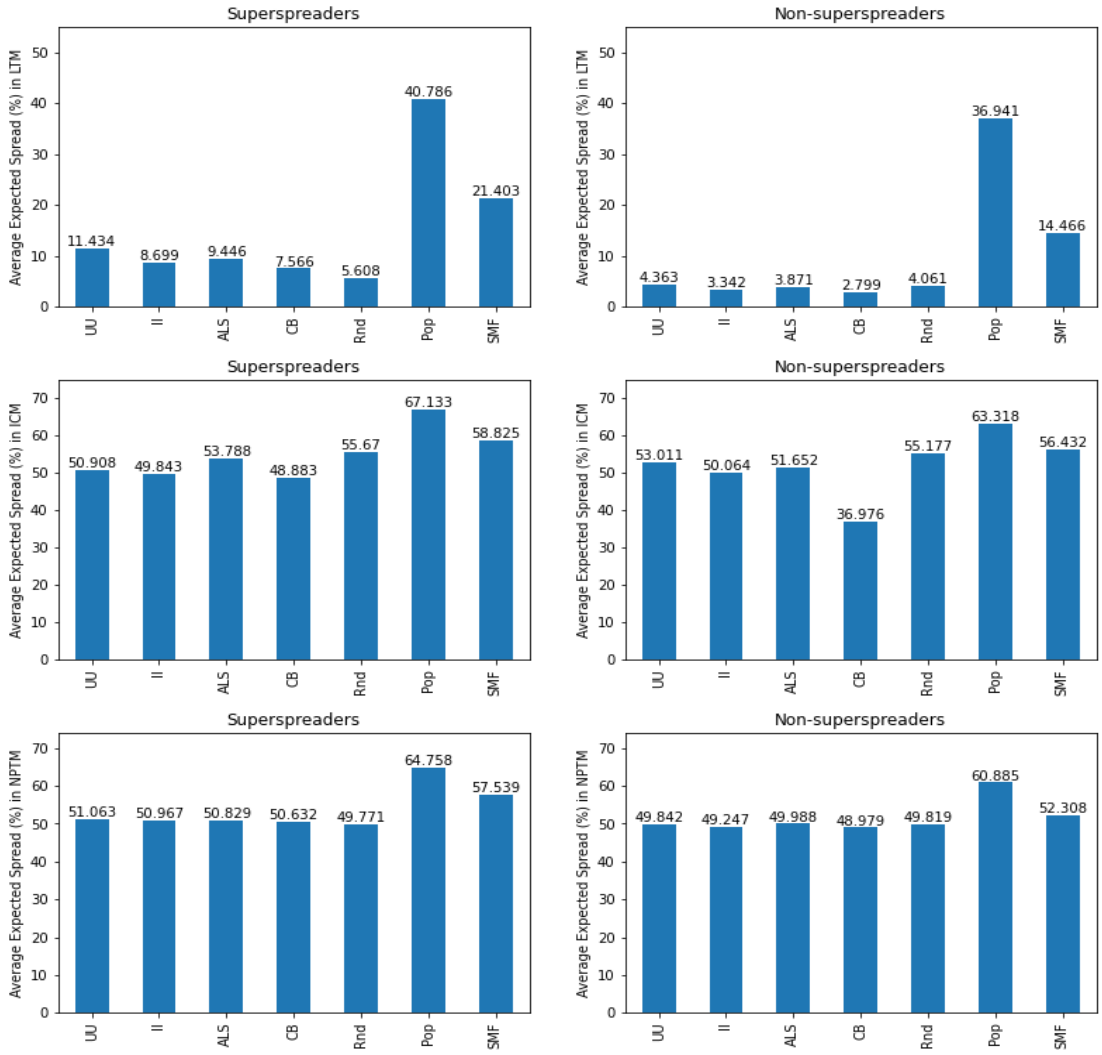

Fig. 21. Average expected spread for each of RA, for all users, superspreaders (defined by $\theta = 70\%$), and non-superspreaders in the PolitiFact FakeNewsNet dataset according to the Linear Threshold model (LTM – top row), Independent Cascade model (ICM – middle row), and the Node Profile Threshold Model (NPTM – bottom row).

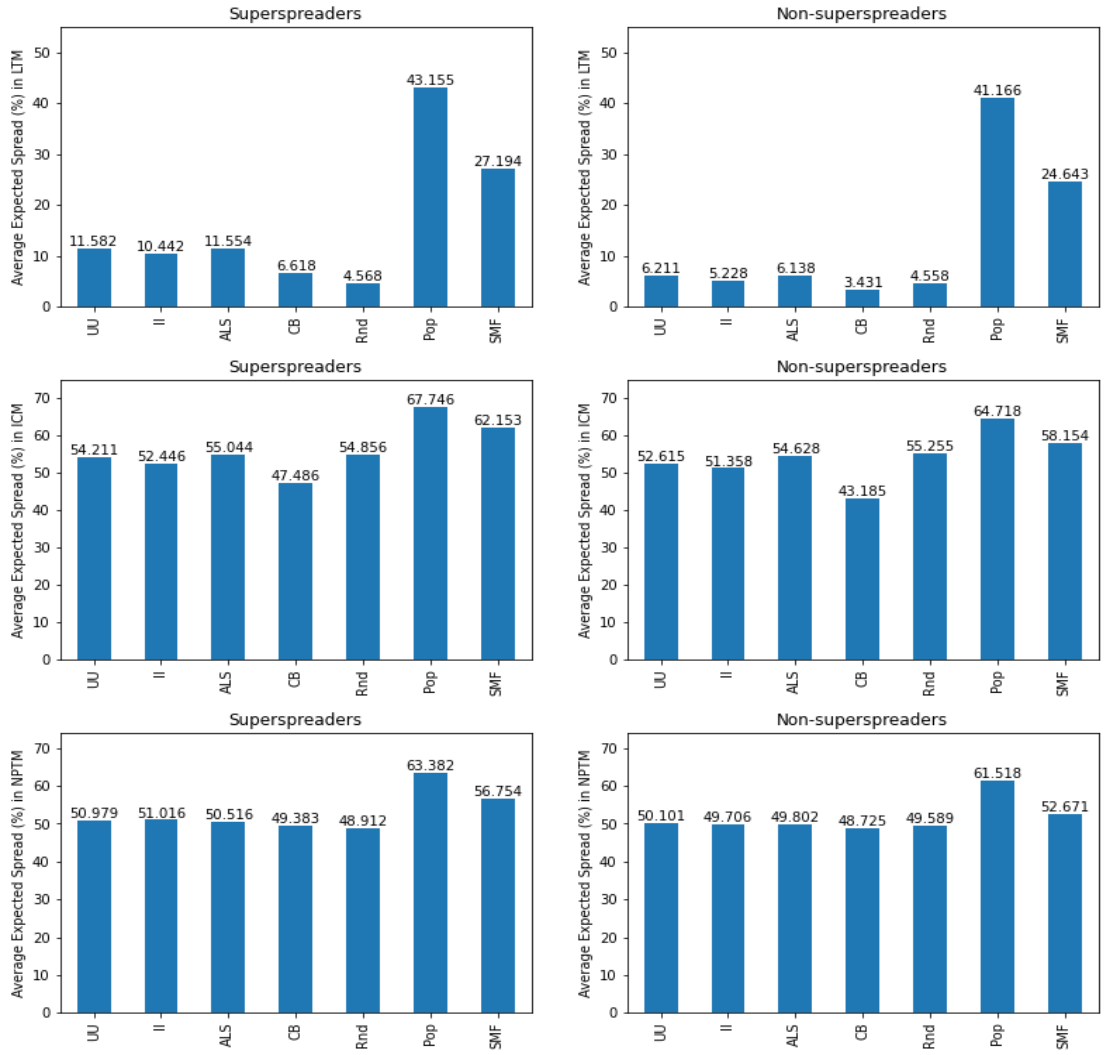

Fig. 22. Average expected spread for each of RA, for all users, superspreaders (defined by $\theta = 80\%$), and non-superspreaders in the Politifact FakeNewsNet dataset according to the Linear Threshold model (LTM – top row), Independent Cascade model (ICM – middle row), and the Node Profile Threshold Model (NPTM – bottom row).

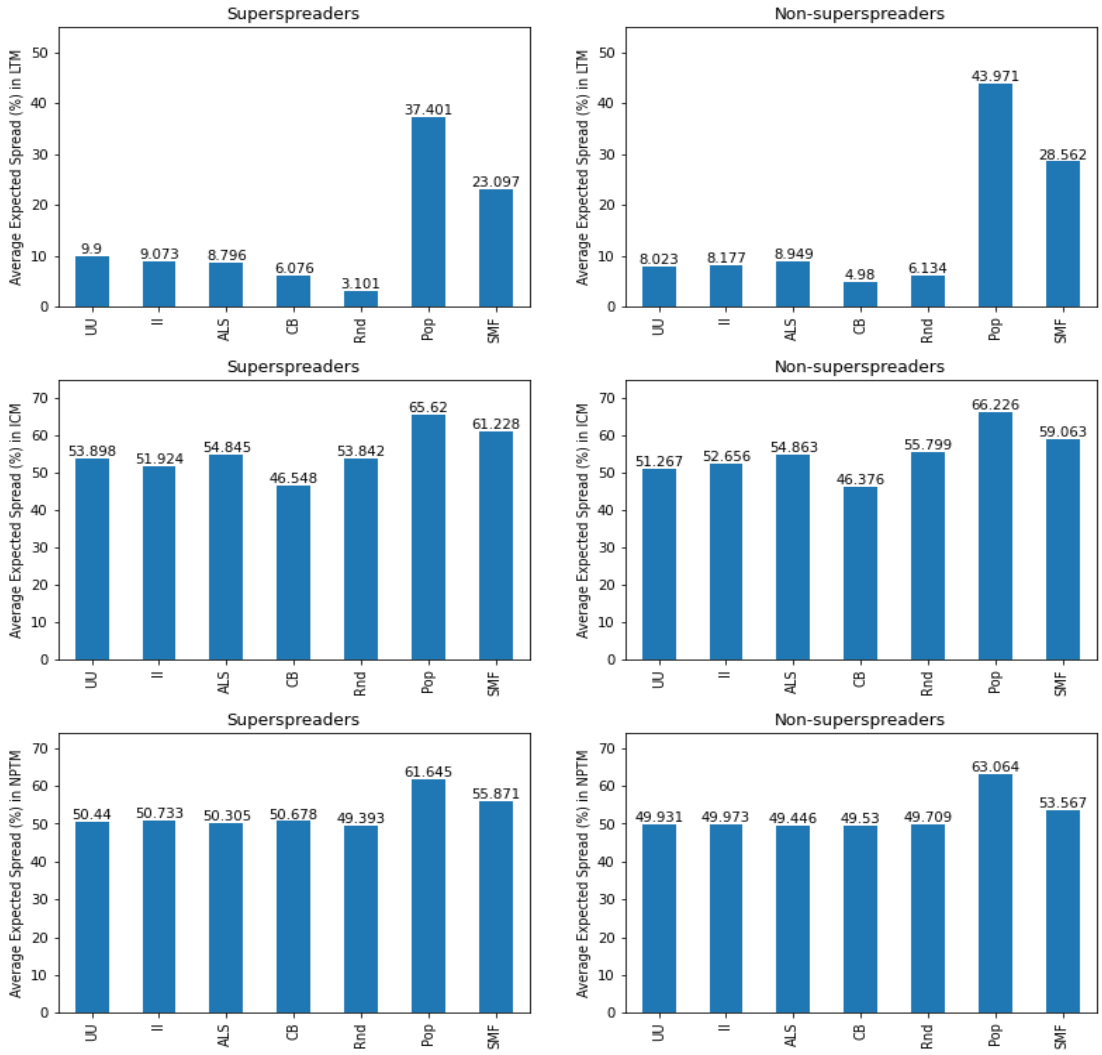

Fig. 23. Average expected spread for each of RA, for all users, superspreaders (defined by $\theta = 90\%$), and non-superspreaders in the Politifact FakeNewsNet dataset according to the Linear Threshold model (LTM – top row), Independent Cascade model (ICM – middle row), and the Node Profile Threshold Model (NPTM – bottom row).

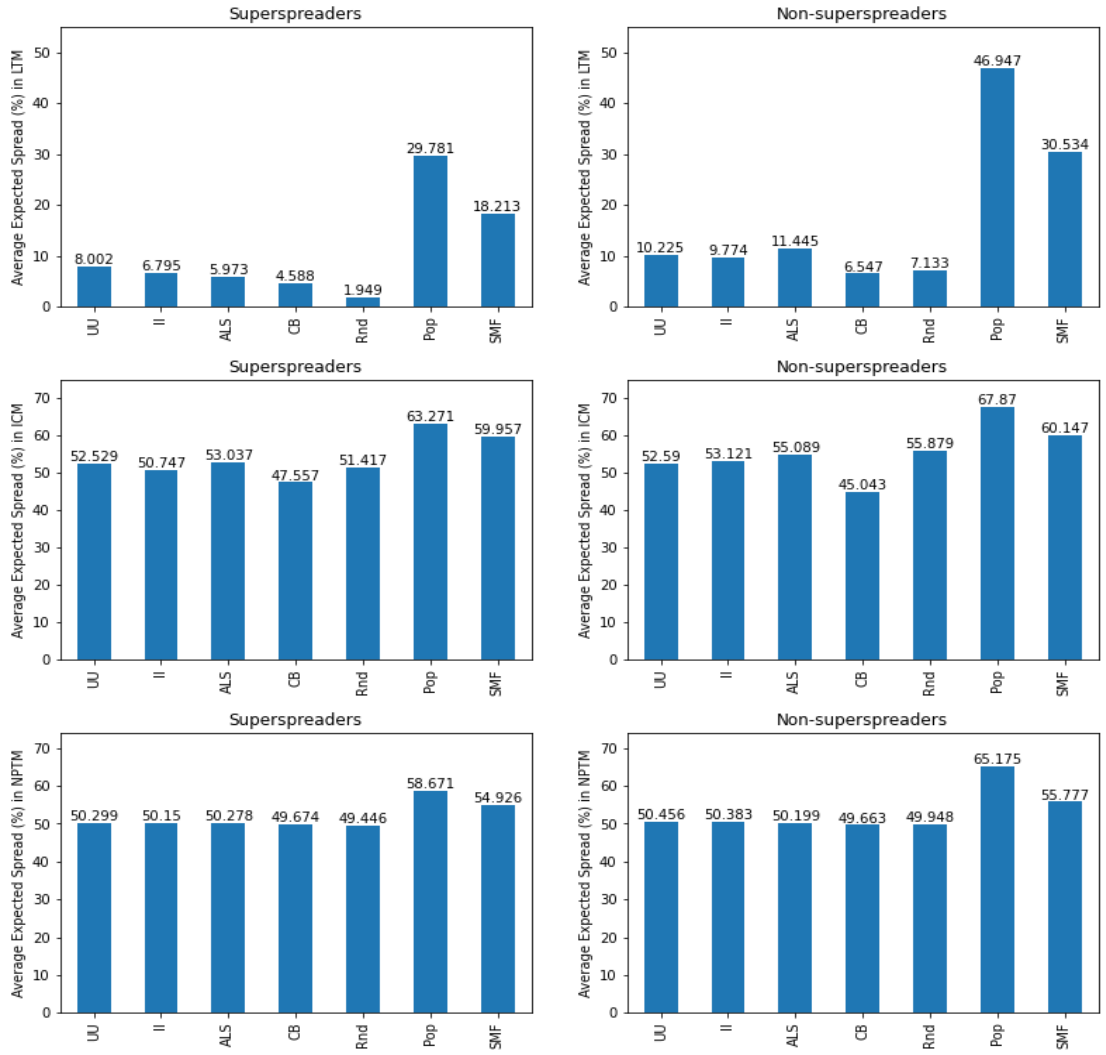

Fig. 24. Average expected spread for each of RA, for all users, superspreaders (defined by $\theta = 100\%$), and non-superspreaders in the Politifact FakeNewsNet dataset according to the Linear Threshold model (LTM – top row), Independent Cascade model (ICM – middle row), and the Node Profile Threshold Model (NPTM – bottom row).

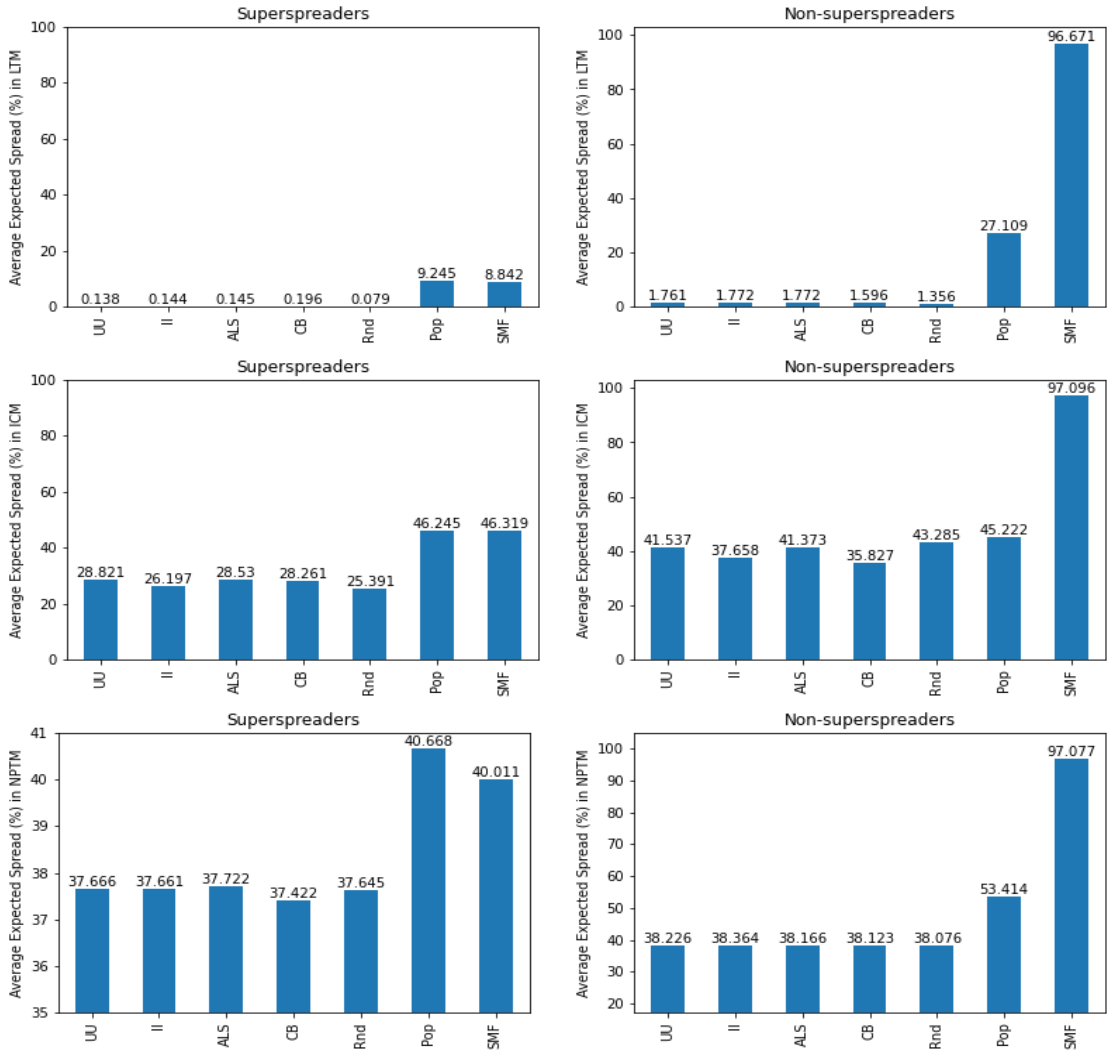

Fig. 25. Average expected spread for each of RA, for all users, superspreaders (defined by $\theta = 50\%$), and non-superspreaders in the HealthStory FakeHealth dataset according to the Linear Threshold model (LTM – top row), Independent Cascade model (ICM – middle row), and the Node Profile Threshold Model (NPTM – bottom row).

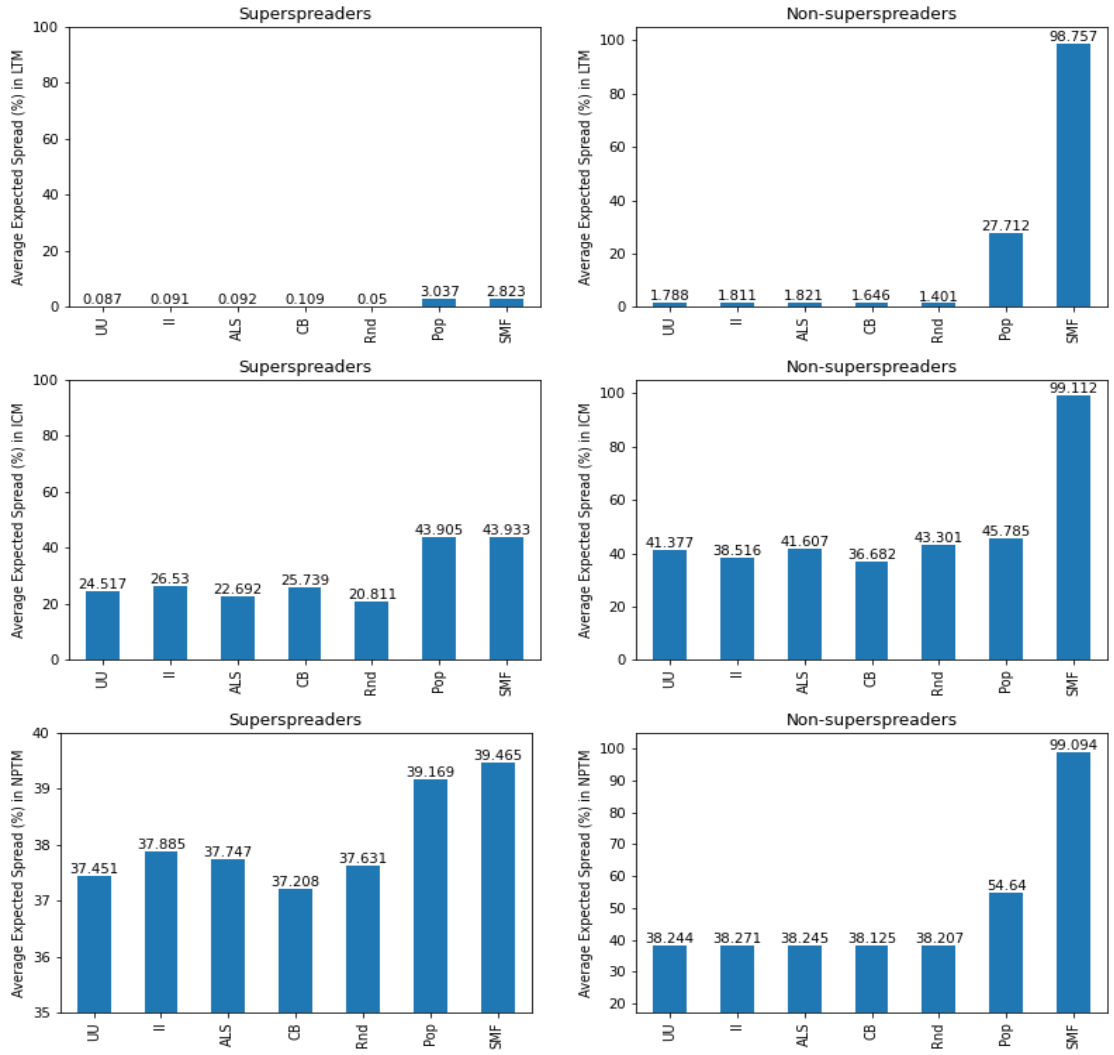

Fig. 26. Average expected spread for each of RA, for all users, superspreaders (defined by $\theta = 60\%$), and non-superspreaders in the HealthStory FakeHealth dataset according to the Linear Threshold model (LTM – top row), Independent Cascade model (ICM – middle row), and the Node Profile Threshold Model (NPTM – bottom row).

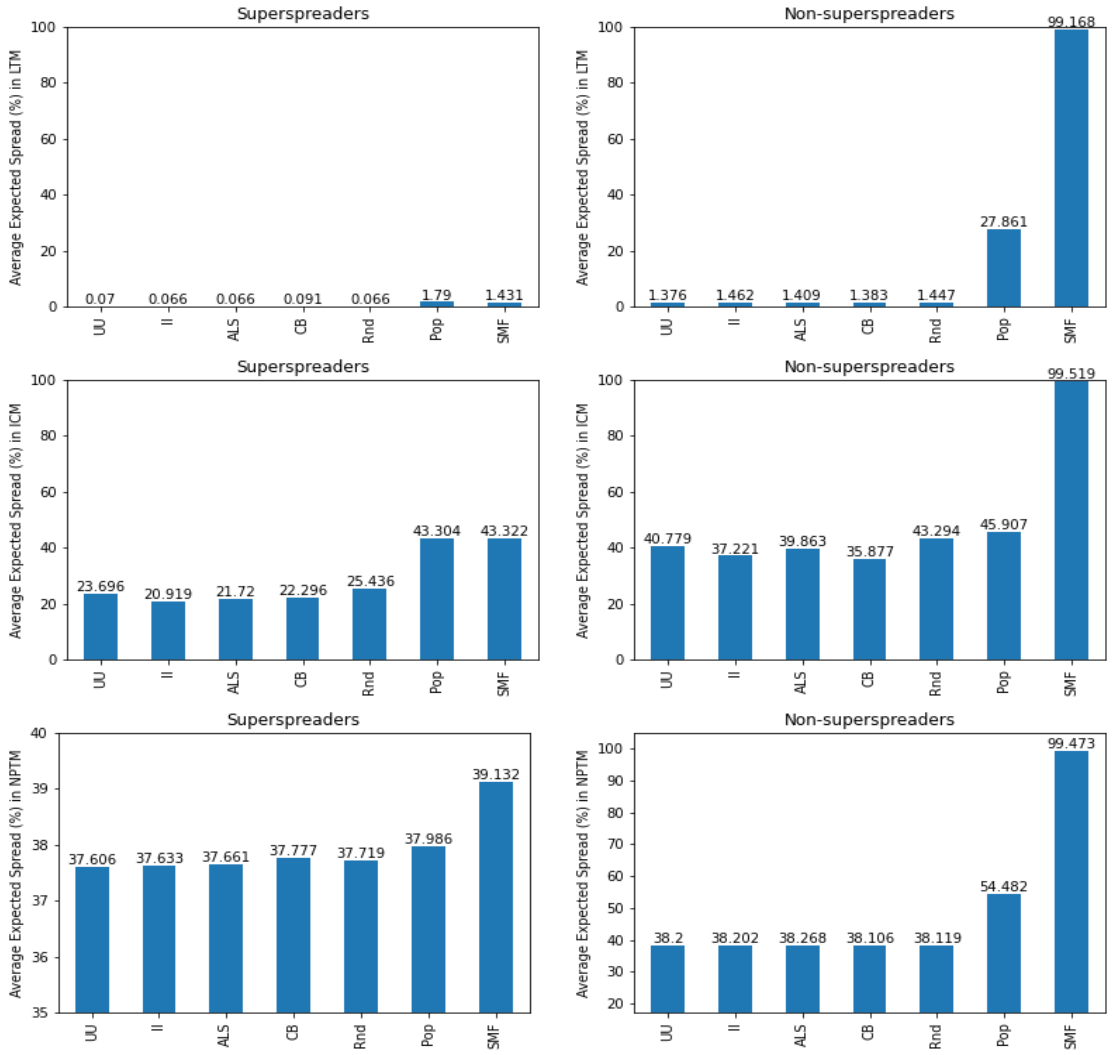

Fig. 27. Average expected spread for each of RA, for all users, superspreaders (defined by $\theta = 70\%$), and non-superspreaders in the HealthStory FakeHealth dataset according to the Linear Threshold model (LTM – top row), Independent Cascade model (ICM – middle row), and the Node Profile Threshold Model (NPTM – bottom row).

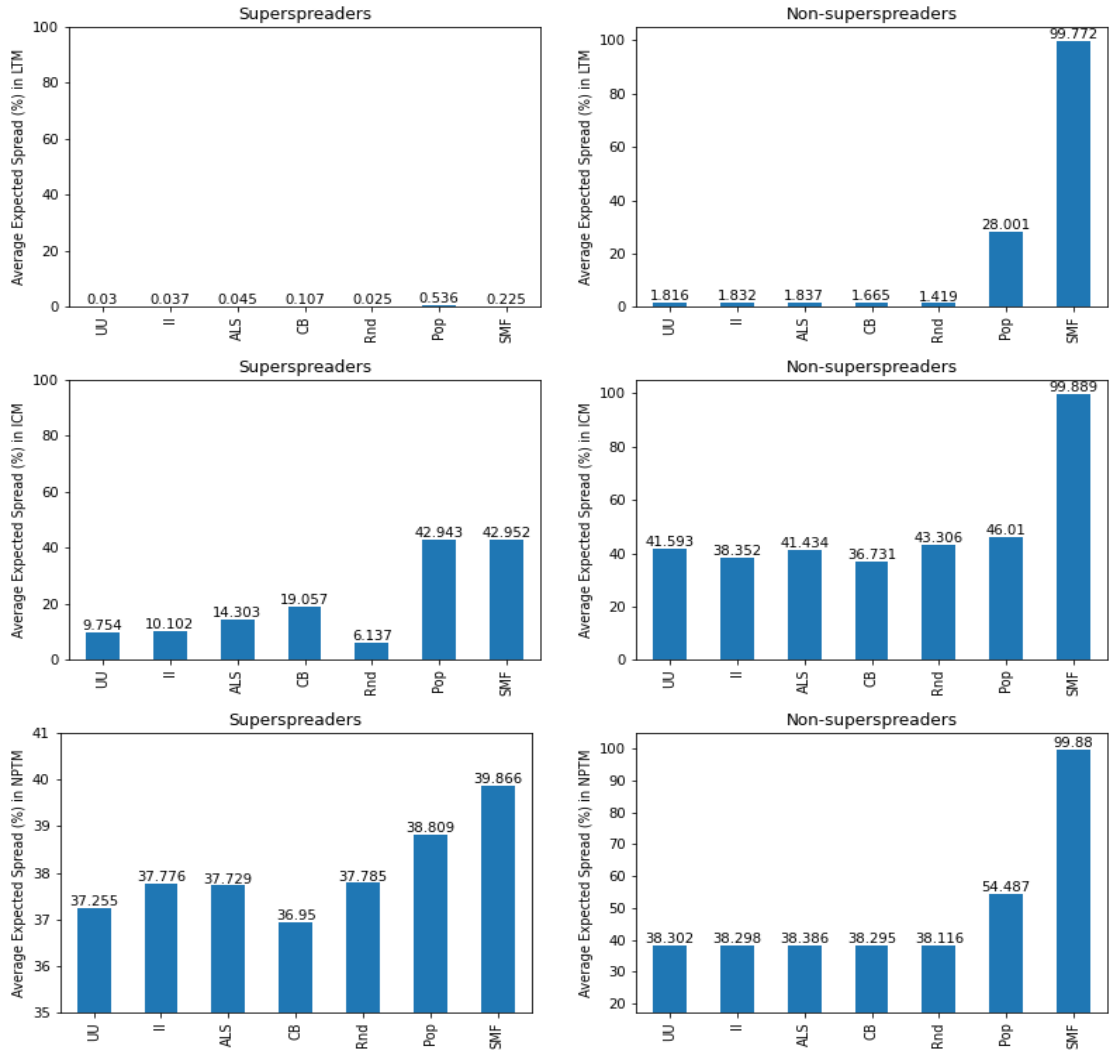

Fig. 28. Average expected spread for each of RA, for all users, superspreaders (defined by $\theta = 80\%$), and non-superspreaders in the HealthStory FakeHealth dataset according to the Linear Threshold model (LTM – top row), Independent Cascade model (ICM – middle row), and the Node Profile Threshold Model (NPTM – bottom row).

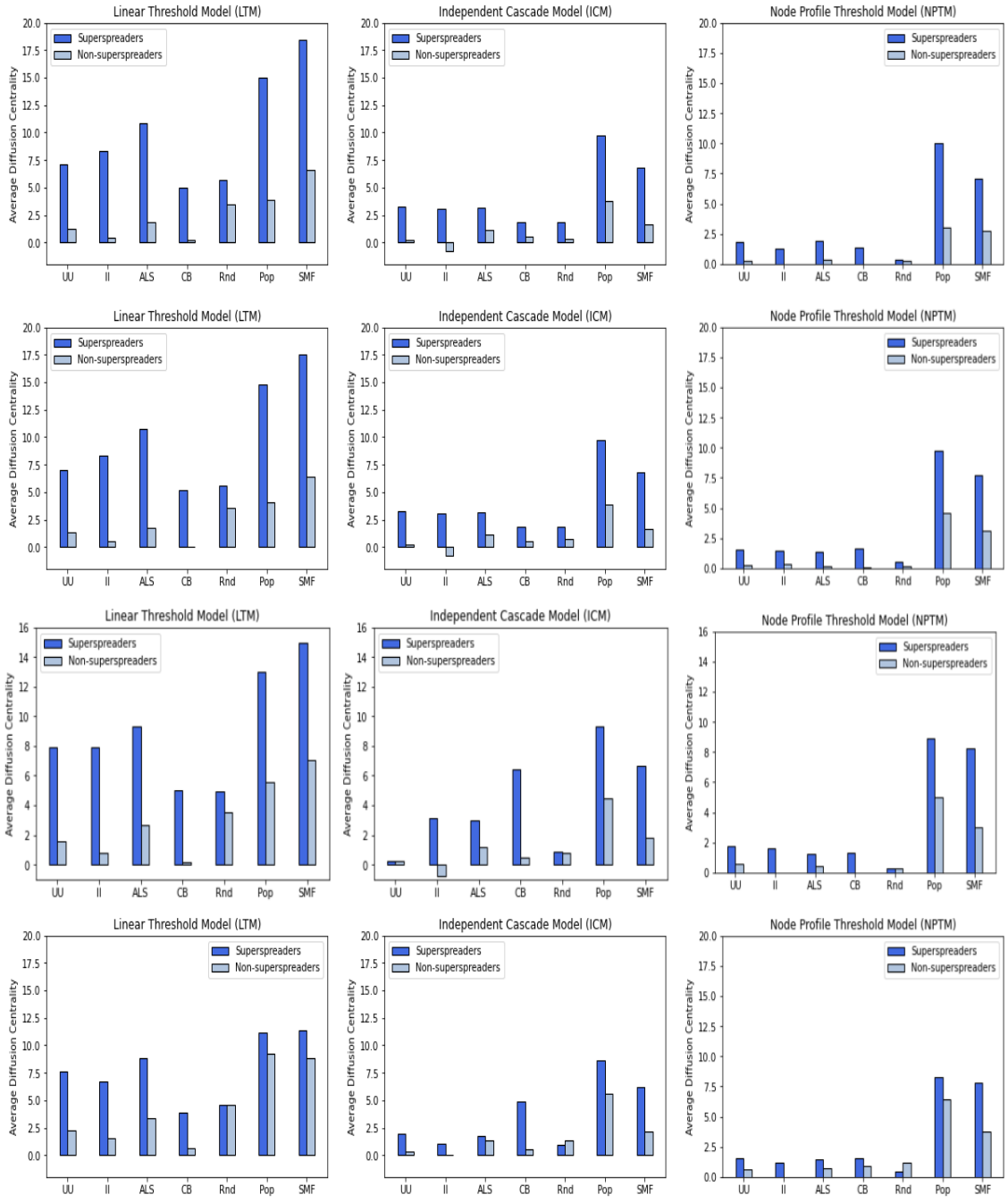

Fig. 29. Average Diffusion Centrality for different user types in the PolitiFact FakeNewsNet dataset according to different thresholds θ for defining superspreaders (first row θ=50%,second row θ=60%, third row θ=70%, and fourth row θ=80%).

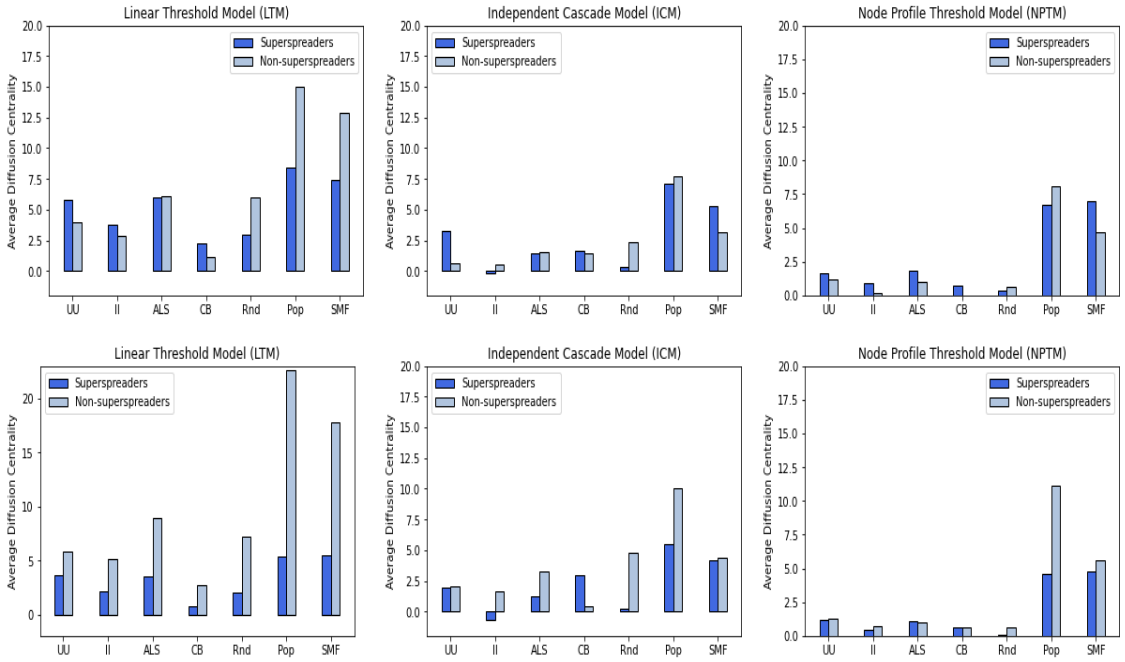

Fig. 30. Average Diffusion Centrality for different user types in the PolitiFact FakeNewsNet dataset according to different thresholds θ for defining superspreaders (first row θ=90%, and second row θ=100%).

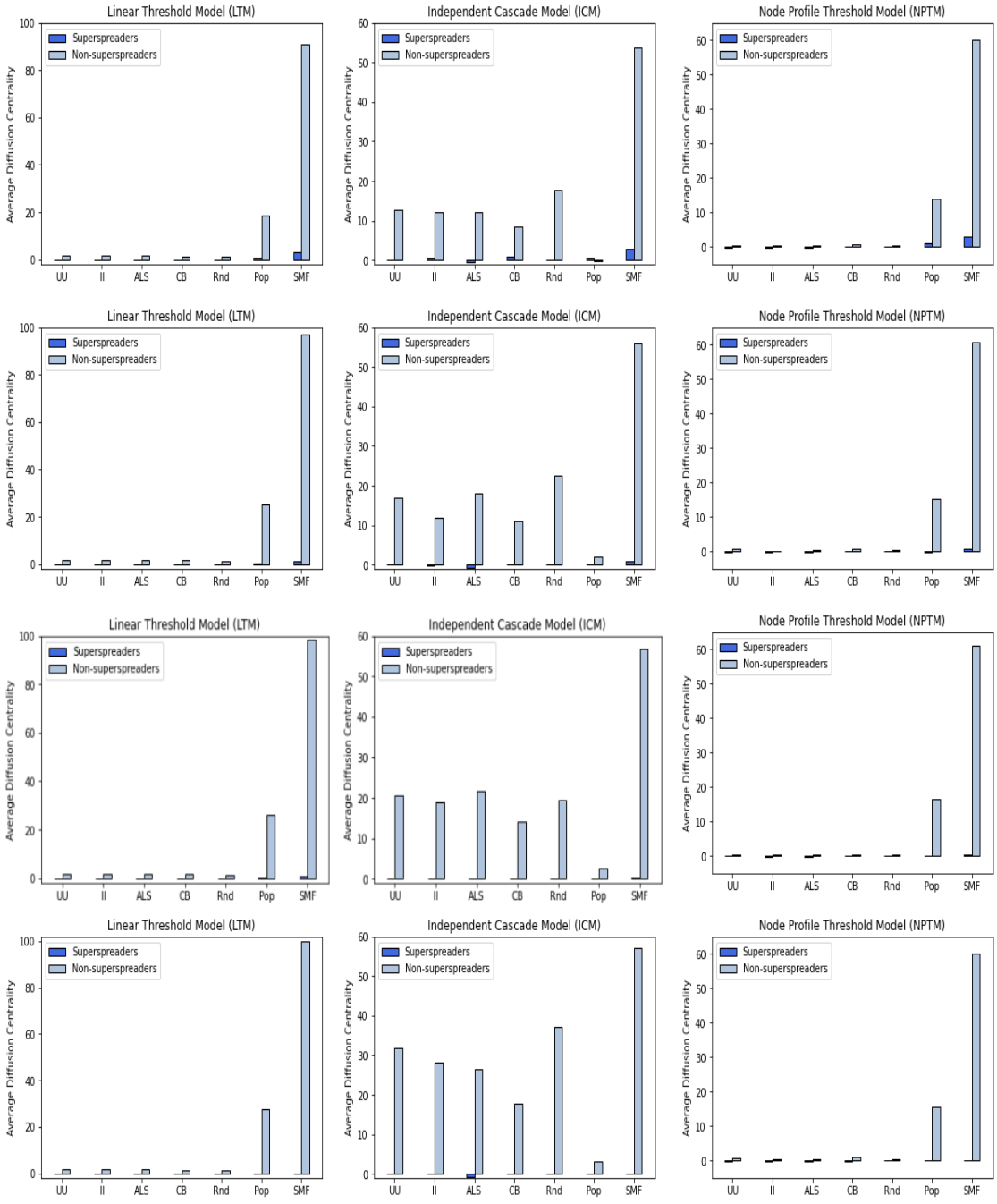

Fig. 31. Average Diffusion Centrality for different user types in the HealthStory FakeHealth dataset according to different thresholds θ for defining superspreaders (first row θ=50%,second row θ=60%,third row θ=70%, and fourth row θ=80%).