Appendix on all confusion matrices supporting the analysis

Name Similarity Cluster of Size 5

	391 GROUND TRUTHS			
*** BASE LINE ***	GT. Pos. 272	GT. neg. 119		
POSITIVE 391	True Pos. 272	False Pos. 119	Precision 0.696	False Discovery Rate 0.304
NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
	Recall 1.0	Fall-out 1.0	Positive L. Ratio 0.696	F1 score — Accuracy 0.821 — 0.696

Table 9: Confusion matrix for link-networks of size 5 (MCC).

	391 GROUND TRUTHS			
	GT. Pos. 272	GT. neg. 119		
POSITIVE 289	True Pos. 231	False Pos. 58	Precision 0.799	False Discovery Rate 0.201
NEGATIVE 102	False Neg. 41	True Neg. 61	F. Omission Rate 0.402	Neg. Predictive Value 0.598
	Recall 0.849	Fall-out 0.487	Positive L. Ratio 1.641	F1 score — Accuracy 0.824 — 0.747

Table 10: Confusion matrix for link-networks of size 5 (Evaluation).

Name Similarity Cluster of Size 6

	224 GROUND TRUTHS			
*** BASE LINE ***	GT. Pos. 139	GT. neg. 85		
POSITIVE 224	True Pos. 139	False Pos. 85	Precision 0.621	False Discovery Rate 0.379
NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
	Recall 1.0	Fall-out 1.0	Positive L. Ratio 0.621	F1 score — Accuracy 0.766 — 0.621

Table 11: Confusion matrix for link-networks of size 6 (MCC).

	224 GROUND TRUTHS			
	GT. Pos. 139	GT. neg. 85		
POSITIVE 145	True Pos. 116	False Pos. 29	Precision 0.8	False Discovery Rate 0.2
NEGATIVE 79	False Neg. 23	True Neg. 56	F. Omission Rate 0.291	Neg. Predictive Value 0.709
	Recall 0.835	Fall-out 0.341	Positive L. Ratio 2.346	F1 score — Accuracy 0.817 — 0.768

Table 12: Confusion matrix for link-networks of size 6 (Evaluation).

Name Similarity Cluster of Size 7

		96 GROUND TRUTHS			
	*** BASE LINE ***	GT. Pos. 50	GT. neg. 46		
PREDICT	POSITIVE 96	True Pos. 50	False Pos. 46	Precision 0.521	False Discovery Rate 0.479
	NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
		Recall 1.0	Fall-out 1.0	Positive L. Ratio 0.521	F1 score — Accuracy 0.685 — 0.521

Table 13: Confusion matrix for link-networks of size 7 (MCC).

		96 GROUND TRUTHS			
		GT. Pos. 50	GT. neg. 46		
PREDICT	POSITIVE 54	True Pos. 42	False Pos. 12	Precision 0.778	False Discovery Rate 0.222
	NEGATIVE 42	False Neg. 8	True Neg. 34	F. Omission Rate 0.19	Neg. Predictive Value 0.81
		Recall 0.84	Fall-out 0.261	Positive L. Ratio 2.98	F1 score — Accuracy 0.808 — 0.792

Table 14: Confusion matrix for link-networks of size 7 (Evaluation).

Name Similarity Cluster of Size 8

		66 GROUND TRUTHS			
	*** BASE LINE ***	GT. Pos. 35	GT. neg. 31		
PREDICT	POSITIVE 66	True Pos. 35	False Pos. 31	Precision 0.53	False Discovery Rate 0.47
	NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
		Recall 1.0	Fall-out 1.0	Positive L. Ratio 0.53	F1 score — Accuracy 0.693 — 0.53

Table 15: Confusion matrix for link-networks of size 8 (MCC).

		66 GROUND TRUTHS			
		GT. Pos. 35	GT. neg. 31		
PREDICT	POSITIVE 32	True Pos. 27	False Pos. 5	Precision 0.844	False Discovery Rate 0.156
	NEGATIVE 34	False Neg. 8	True Neg. 26	F. Omission Rate 0.235	Neg. Predictive Value 0.765
		Recall 0.771	Fall-out 0.161	Positive L. Ratio 5.241	F1 score — Accuracy 0.806 — 0.803

Table 16: Confusion matrix for link-networks of size 8 (Evaluation).

Name Similarity Cluster of Size 9

		45 GROUND TRUTHS			
	*** BASE LINE ***	GT. Pos. 21	GT. neg. 24		
PREDICT	POSITIVE 0	True Pos. 0	False Pos. 0	Precision -	False Discovery Rate -
	NEGATIVE 45	False Neg. 21	True Neg. 24	F. Omission Rate 0.467	Neg. Predictive Value 0.533
		Recall 0.0	Fall-out 0.0	Positive L. Ratio -	F1 score — Accuracy - — 0.533

Table 17: Confusion matrix for link-networks of size 9 (MCC).

		45 GROUND TRUTHS			
		GT. Pos. 21	GT. neg. 24		
PREDICT	POSITIVE 26	True Pos. 21	False Pos. 5	Precision 0.808	False Discovery Rate 0.192
	NEGATIVE 19	False Neg. 0	True Neg. 19	F. Omission Rate 0.0	Neg. Predictive Value 1.0
		Recall 1.0	Fall-out 0.208	Positive L. Ratio 3.883	F1 score — Accuracy 0.894 — 0.889

Table 18: Confusion matrix for link-networks of size 9 (Evaluation).

Name Similarity Cluster of Size 10

		24 GROUND TRUTHS			
	*** BASE LINE ***	GT. Pos. 8	GT. neg. 16		
PREDICT	POSITIVE 0	True Pos. 0	False Pos. 0	Precision -	False Discovery Rate -
	NEGATIVE 24	False Neg. 8	True Neg. 16	F. Omission Rate 0.333	Neg. Predictive Value 0.667
		Recall 0.0	Fall-out 0.0	Positive L. Ratio -	F1 score — Accuracy - — 0.667

Table 19: Confusion matrix for link-networks of size 10 (MCC).

		24 GROUND TRUTHS			
		GT. Pos. 8	GT. neg. 16		
PREDICT	POSITIVE 7	True Pos. 7	False Pos. 0	Precision 1.0	False Discovery Rate 0.0
	NEGATIVE 17	False Neg. 1	True Neg. 16	F. Omission Rate 0.059	Neg. Predictive Value 0.941
		Recall 0.875	Fall-out 0.0	Positive L. Ratio -	F1 score — Accuracy 0.933 — 0.958

Table 20: Confusion matrix for link-networks of size 10 (Evaluation).

Name Similarity Cluster of Size 3: MCC vs. Expert

*** BASE LINE ***	64 GROUND TRUTHS			
	GT. Pos. 56	GT. neg. 8		
POSITIVE 64	True Pos. 56	False Pos. 8	Precision 0.875	False Discovery Rate 0.125
NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
	Recall 1.0	Fall-out 1.0	Positive L. Ratio 0.875	F1 score — Accuracy 0.933 — 0.875

Table 21: Confusion matrix for link-networks of size 3 (MCC).

	64 GROUND TRUTHS			
	GT. Pos. 56	GT. neg. 8		
POSITIVE 60	True Pos. 54	False Pos. 6	Precision 0.9	False Discovery Rate 0.1
NEGATIVE 4	False Neg. 2	True Neg. 2	F. Omission Rate 0.5	Neg. Predictive Value 0.5
	Recall 0.964	Fall-out 0.75	Positive L. Ratio 1.2	F1 score — Accuracy 0.931 — 0.875

Table 22: Confusion matrix for link-networks of size 3 (Expert).

Name Similarity Cluster of Size 4: MCC vs. Expert

*** BASE LINE ***	24 GROUND TRUTHS			
	GT. Pos. 19	GT. neg. 5		
POSITIVE 24	True Pos. 19	False Pos. 5	Precision 0.792	False Discovery Rate 0.208
NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
	Recall 1.0	Fall-out 1.0	Positive L. Ratio 0.792	F1 score — Accuracy 0.884 — 0.792

Table 23: Confusion matrix for link-networks of size 4 (MCC).

	24 GROUND TRUTHS			
	GT. Pos. 19	GT. neg. 5		
POSITIVE 22	True Pos. 18	False Pos. 4	Precision 0.818	False Discovery Rate 0.182
NEGATIVE 2	False Neg. 1	True Neg. 1	F. Omission Rate 0.5	Neg. Predictive Value 0.5
	Recall 0.947	Fall-out 0.8	Positive L. Ratio 1.023	F1 score — Accuracy 0.878 — 0.792

Table 24: Confusion matrix for link-networks of size 4 (Expert).

Name Similarity Cluster of Size 5: MCC vs. Expert

		15 GROUND TRUTHS			
	*** BASE LINE ***	GT. Pos. 14	GT. neg. 1		
PREDICT	POSITIVE 15	True Pos. 14	False Pos. 1	Precision 0.933	False Discovery Rate 0.067
	NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
		Recall 1.0	Fall-out 1.0	Positive L. Ratio 0.933	F1 score — Accuracy 0.966 — 0.933

Table 25: Confusion matrix for link-networks of size 5 (MCC).

		15 GROUND TRUTHS			
		GT. Pos. 14	GT. neg. 1		
PREDICT	POSITIVE 14	True Pos. 13	False Pos. 1	Precision 0.929	False Discovery Rate 0.071
	NEGATIVE 1	False Neg. 1	True Neg. 0	F. Omission Rate 1.0	Neg. Predictive Value 0.0
		Recall 0.929	Fall-out 1.0	Positive L. Ratio 0.929	F1 score — Accuracy 0.929 — 0.867

Table 26: Confusion matrix for link-networks of size 5 (Expert).

Name Similarity Cluster of Size 6: MCC vs. Expert

		19 GROUND TRUTHS			
	*** BASE LINE ***	GT. Pos. 14	GT. neg. 5		
PREDICT	POSITIVE 19	True Pos. 14	False Pos. 5	Precision 0.737	False Discovery Rate 0.263
	NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
		Recall 1.0	Fall-out 1.0	Positive L. Ratio 0.737	F1 score — Accuracy 0.848 — 0.737

Table 27: Confusion matrix for link-networks of size 6 (MCC).

		19 GROUND TRUTHS			
		GT. Pos. 14	GT. neg. 5		
PREDICT	POSITIVE 19	True Pos. 14	False Pos. 5	Precision 0.737	False Discovery Rate 0.263
	NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
		Recall 1.0	Fall-out 1.0	Positive L. Ratio 0.737	F1 score — Accuracy 0.848 — 0.737

Table 28: Confusion matrix for link-networks of size 6 (Expert).

Name Similarity Cluster of Size 7: MCC vs. Expert

*** BASE LINE ***	12 GROUND TRUTHS			
	GT. Pos. 10	GT. neg. 2		
POSITIVE 12	True Pos. 10	False Pos. 2	Precision 0.833	False Discovery Rate 0.167
NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
	Recall 1.0	Fall-out 1.0	Positive L. Ratio 0.833	F1 score — Accuracy 0.909 — 0.833

Table 29: Confusion matrix for link-networks of size 7 (MCC).

	12 GROUND TRUTHS			
	GT. Pos. 10	GT. neg. 2		
POSITIVE 10	True Pos. 10	False Pos. 0	Precision 1.0	False Discovery Rate 0.0
NEGATIVE 2	False Neg. 0	True Neg. 2	F. Omission Rate 0.0	Neg. Predictive Value 1.0
	Recall 1.0	Fall-out 0.0	Positive L. Ratio -	F1 score — Accuracy 1.0 — 1.0

Table 30: Confusion matrix for link-networks of size 7 (Expert).

Name Similarity Cluster of Size 8: MCC vs. Expert

*** BASE LINE ***	4 GROUND TRUTHS			
	GT. Pos. 4	GT. neg. 0		
POSITIVE 4	True Pos. 4	False Pos. 0	Precision 1.0	False Discovery Rate 0.0
NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
	Recall 1.0	Fall-out -	Positive L. Ratio -	F1 score — Accuracy 1.0 — 1.0

Table 31: Confusion matrix for link-networks of size 8 (MCC).

	4 GROUND TRUTHS			
	GT. Pos. 4	GT. neg. 0		
POSITIVE 4	True Pos. 4	False Pos. 0	Precision 1.0	False Discovery Rate 0.0
NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
	Recall 1.0	Fall-out -	Positive L. Ratio -	F1 score — Accuracy 1.0 — 1.0

Table 32: Confusion matrix for link-networks of size 8 (Expert).

Name Similarity Cluster of Size 9: MCC vs. Expert

		9 GROUND TRUTHS			
	*** BASE LINE ***	GT. Pos. 8	GT. neg. 1		
PREDICT	POSITIVE 9	True Pos. 8	False Pos. 1	Precision 0.889	False Discovery Rate 0.111
	NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
		Recall 1.0	Fall-out 1.0	Positive L. Ratio 0.889	F1 score — Accuracy 0.941 — 0.889

Table 33: Confusion matrix for link-networks of size 9 (MCC).

		9 GROUND TRUTHS			
		GT. Pos. 8	GT. neg. 1		
PREDICT	POSITIVE 8	True Pos. 8	False Pos. 0	Precision 1.0	False Discovery Rate 0.0
	NEGATIVE 1	False Neg. 0	True Neg. 1	F. Omission Rate 0.0	Neg. Predictive Value 1.0
		Recall 1.0	Fall-out 0.0	Positive L. Ratio -	F1 score — Accuracy 1.0 — 1.0

Table 34: Confusion matrix for link-networks of size 9 (Expert).

Name Similarity Cluster of Size 10: MCC vs. Expert

		1 GROUND TRUTHS			
	*** BASE LINE ***	GT. Pos. 1	GT. neg. 0		
PREDICT	POSITIVE 1	True Pos. 1	False Pos. 0	Precision 1.0	False Discovery Rate 0.0
	NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
		Recall 1.0	Fall-out -	Positive L. Ratio -	F1 score — Accuracy 1.0 — 1.0

Table 35: Confusion matrix for link-networks of size 10 (MCC).

		1 GROUND TRUTHS			
		GT. Pos. 1	GT. neg. 0		
PREDICT	POSITIVE 1	True Pos. 1	False Pos. 0	Precision 1.0	False Discovery Rate 0.0
	NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
		Recall 1.0	Fall-out -	Positive L. Ratio -	F1 score — Accuracy 1.0 — 1.0

Table 36: Confusion matrix for link-networks of size 10 (Expert).

50 meters proximity

		92 GROUND TRUTHS			
	*** BASE LINE ***	GT. Pos. 30	GT. neg. 62		
PREDICT	POSITIVE 0	True Pos. 0	False Pos. 0	Precision -	False Discovery Rate -
	NEGATIVE 92	False Neg. 30	True Neg. 62	F. Omission Rate 0.326	Neg. Predictive Value 0.674
		Recall 0.0	Fall-out 0.0	Positive L. Ratio -	F1 score — Accuracy - — 0.674

Table 37: Confusion matrix for link-networks of size 3, 50m before (MCC).

		92 GROUND TRUTHS			
		GT. Pos. 30	GT. neg. 62		
PREDICT	POSITIVE 45	True Pos. 26	False Pos. 19	Precision 0.578	False Discovery Rate 0.422
	NEGATIVE 47	False Neg. 4	True Neg. 43	F. Omission Rate 0.085	Neg. Predictive Value 0.915
		Recall 0.867	Fall-out 0.306	Positive L. Ratio 1.888	F1 score — Accuracy 0.693 — 0.75

Table 38: Confusion matrix for link-networks of size 3, 50m before (Evaluation).

50 meters proximity + name

		31 GROUND TRUTHS			
	*** BASE LINE ***	GT. Pos. 27	GT. neg. 4		
PREDICT	POSITIVE 31	True Pos. 27	False Pos. 4	Precision 0.871	False Discovery Rate 0.129
	NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
		Recall 1.0	Fall-out 1.0	Positive L. Ratio 0.871	F1 score — Accuracy 0.931 — 0.871

Table 39: Confusion matrix for link-networks of size 3, 50m after (MCC).

		31 GROUND TRUTHS			
		GT. Pos. 27	GT. neg. 4		
PREDICT	POSITIVE 19	True Pos. 19	False Pos. 0	Precision 1.0	False Discovery Rate 0.0
	NEGATIVE 12	False Neg. 8	True Neg. 4	F. Omission Rate 0.667	Neg. Predictive Value 0.333
		Recall 0.704	Fall-out 0.0	Positive L. Ratio -	F1 score — Accuracy 0.826 — 0.742

Table 40: Confusion matrix for link-networks of size 3, 50m after (Evaluation).

50 meters proximity: all remain the same except this

		92 GROUND TRUTHS			
	*** BASE LINE ***	GT. Pos. 31	GT. neg. 61		
PREDICT	POSITIVE 0	True Pos. 0	False Pos. 0	Precision -	False Discovery Rate -
	NEGATIVE 92	False Neg. 31	True Neg. 61	F. Omission Rate 0.337	Neg. Predictive Value 0.663
		Recall 0.0	Fall-out 0.0	Positive L. Ratio -	F1 score — Accuracy - — 0.663

Table 41: Confusion matrix for link-networks of size 3, 50m before (MCC).

		92 GROUND TRUTHS			
		GT. Pos. 31	GT. neg. 61		
PREDICT	POSITIVE 45	True Pos. 27	False Pos. 18	Precision 0.6	False Discovery Rate 0.4
	NEGATIVE 47	False Neg. 4	True Neg. 43	F. Omission Rate 0.085	Neg. Predictive Value 0.915
		Recall 0.871	Fall-out 0.295	Positive L. Ratio 2.034	F1 score — Accuracy 0.711 — 0.761

Table 42: Confusion matrix for link-networks of size 3, 50m before (Expert).

500 meters proximity

	249 GROUND TRUTHS			
*** BASE LINE ***	GT. Pos. 66	GT. neg. 183		
POSITIVE 0	True Pos. 0	False Pos. 0	Precision -	False Discovery Rate -
NEGATIVE 249	False Neg. 66	True Neg. 183	F. Omission Rate 0.265	Neg. Predictive Value 0.735
	Recall 0.0	Fall-out 0.0	Positive L. Ratio -	F1 score — Accuracy - — 0.735

Table 43: Confusion matrix for link-networks of size 3, 500m before (MCC).

	249 GROUND TRUTHS			
	GT. Pos. 66	GT. neg. 183		
POSITIVE 115	True Pos. 63	False Pos. 52	Precision 0.548	False Discovery Rate 0.452
NEGATIVE 134	False Neg. 3	True Neg. 131	F. Omission Rate 0.022	Neg. Predictive Value 0.978
	Recall 0.955	Fall-out 0.284	Positive L. Ratio 1.929	F1 score — Accuracy 0.696 — 0.779

Table 44: Confusion matrix for link-networks of size 3, 500m before (Evaluation).

500 meters proximity + name

	155 GROUND TRUTHS			
*** BASE LINE ***	GT. Pos. 148	GT. neg. 7		
POSITIVE 155	True Pos. 148	False Pos. 7	Precision 0.955	False Discovery Rate 0.045
NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
	Recall 1.0	Fall-out 1.0	Positive L. Ratio 0.955	F1 score — Accuracy 0.977 — 0.955

Table 45: Confusion matrix for link-networks of size 3, 500m after (MCC).

	155 GROUND TRUTHS			
	GT. Pos. 148	GT. neg. 7		
POSITIVE 127	True Pos. 125	False Pos. 2	Precision 0.984	False Discovery Rate 0.016
NEGATIVE 28	False Neg. 23	True Neg. 5	F. Omission Rate 0.821	Neg. Predictive Value 0.179
	Recall 0.845	Fall-out 0.286	Positive L. Ratio 3.441	F1 score — Accuracy 0.909 — 0.839

Table 46: Confusion matrix for link-networks of size 3, 500m after (Evaluation).

2000 meters proximity

*** BASE LINE ***	198 GROUND TRUTHS			
	GT. Pos. 61	GT. neg. 137		
POSITIVE 0	True Pos. 0	False Pos. 0	Precision -	False Discovery Rate -
NEGATIVE 198	False Neg. 61	True Neg. 137	F. Omission Rate 0.308	Neg. Predictive Value 0.692
	Recall 0.0	Fall-out 0.0	Positive L. Ratio -	F1 score — Accuracy - — 0.692

Table 47: Confusion matrix for link-networks of size 3, 2 km before (MCC).

	198 GROUND TRUTHS			
	GT. Pos. 61	GT. neg. 137		
POSITIVE 81	True Pos. 57	False Pos. 24	Precision 0.704	False Discovery Rate 0.296
NEGATIVE 117	False Neg. 4	True Neg. 113	F. Omission Rate 0.034	Neg. Predictive Value 0.966
	Recall 0.934	Fall-out 0.175	Positive L. Ratio 4.021	F1 score — Accuracy 0.803 — 0.859

Table 48: Confusion matrix for link-networks of size 3, 2 km before (Evaluation).

2000 meters proximity + name

*** BASE LINE ***	342 GROUND TRUTHS			
	GT. Pos. 322	GT. neg. 20		
POSITIVE 342	True Pos. 322	False Pos. 20	Precision 0.942	False Discovery Rate 0.058
NEGATIVE 0	False Neg. 0	True Neg. 0	F. Omission Rate -	Neg. Predictive Value -
	Recall 1.0	Fall-out 1.0	Positive L. Ratio 0.942	F1 score — Accuracy 0.97 — 0.942

Table 49: Confusion matrix for link-networks of size 3, 2 km after (MCC).

	342 GROUND TRUTHS			
	GT. Pos. 322	GT. neg. 20		
POSITIVE 279	True Pos. 274	False Pos. 5	Precision 0.982	False Discovery Rate 0.018
NEGATIVE 63	False Neg. 48	True Neg. 15	F. Omission Rate 0.762	Neg. Predictive Value 0.238
	Recall 0.851	Fall-out 0.25	Positive L. Ratio 3.928	F1 score — Accuracy 0.912 — 0.845

Table 50: Confusion matrix for link-networks of size 3, 2 km after (Evaluation).