

Results

November 6, 2020

1 Main Results

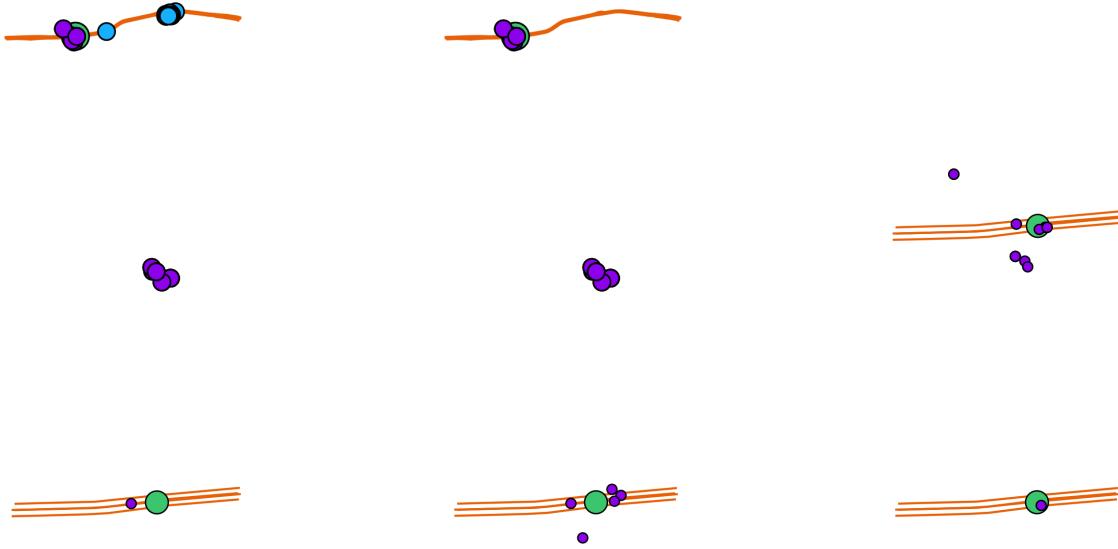


Figure 1: In Paper: Figure 1

Table 1: In Paper: Table 1

	Any Location Captured by Algorithm Close to True Crash Location		Crash Location Determined by Algorithm Close to True Crash Location	
	Recall	Precision	Recall	Precision
LNE _x	0.674	0.686	0.129	0.132
Alg., Raw Gaz	0.695	0.757	0.57	0.761
Alg., Aug Gaz	0.798	0.857	0.656	0.813
Alg., Aug Gaz [Cluster]			0.666	0.777

‘N Crashes’ refers to the number of correctly identified crashes. ‘Raw Gaz’ refers to the raw gazetteer (ie, dictionary of landmarks with original names) and ‘Aug Gaz’ refers to the augmented gazetteer. We use our raw gazetteer as an input into LNE_x, which implements its own augmentation process. For LNE_x, the crash location is determined by taking the centroid of all locations captured by the algorithm. Locations are considered close if they are within 500 meters of each other.

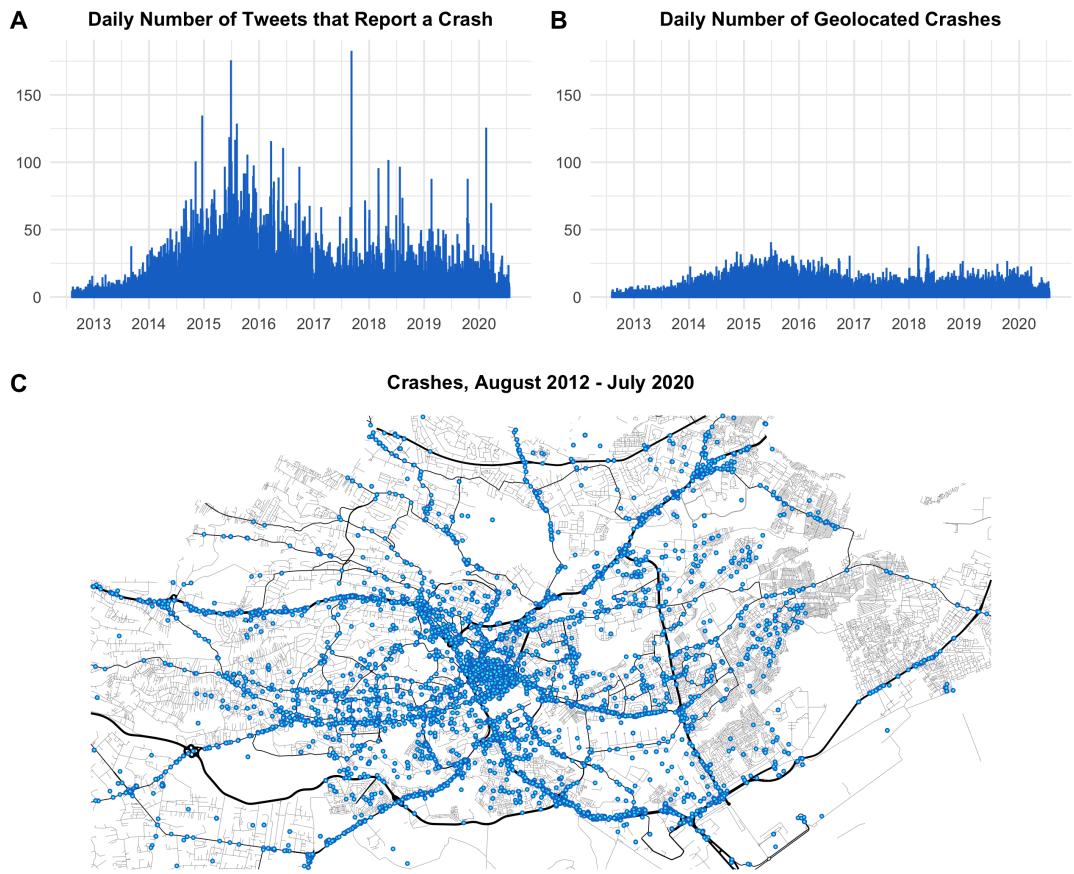


Figure 2: In Paper: Figure 2

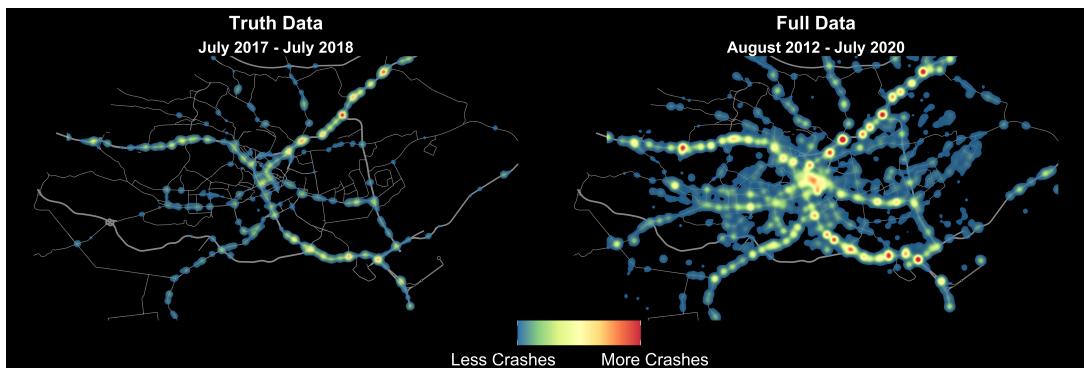


Figure 3: In Paper: Figure 3

2 Supplementary Information

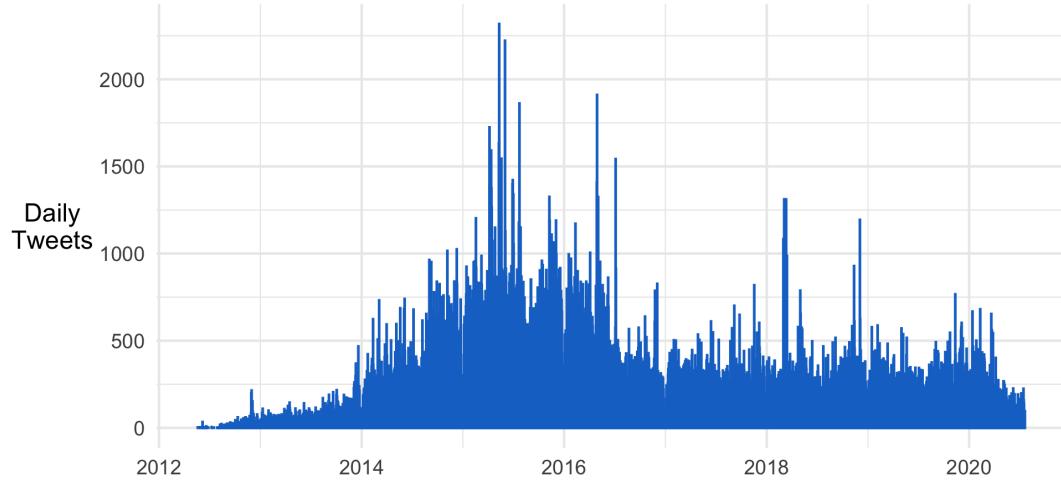


Figure 4: In Paper: Figure s1

Table 2: In Paper: Table s4

Precision	Recall	F1	Accuracy	N-Grams
Naive Bayes				
0.938	0.947	0.942	0.919	1
0.945	0.949	0.947	0.926	2
0.945	0.949	0.947	0.926	3
SVM				
0.935	0.963	0.948	0.927	1
0.94	0.966	0.953	0.934	2
0.939	0.967	0.953	0.934	3

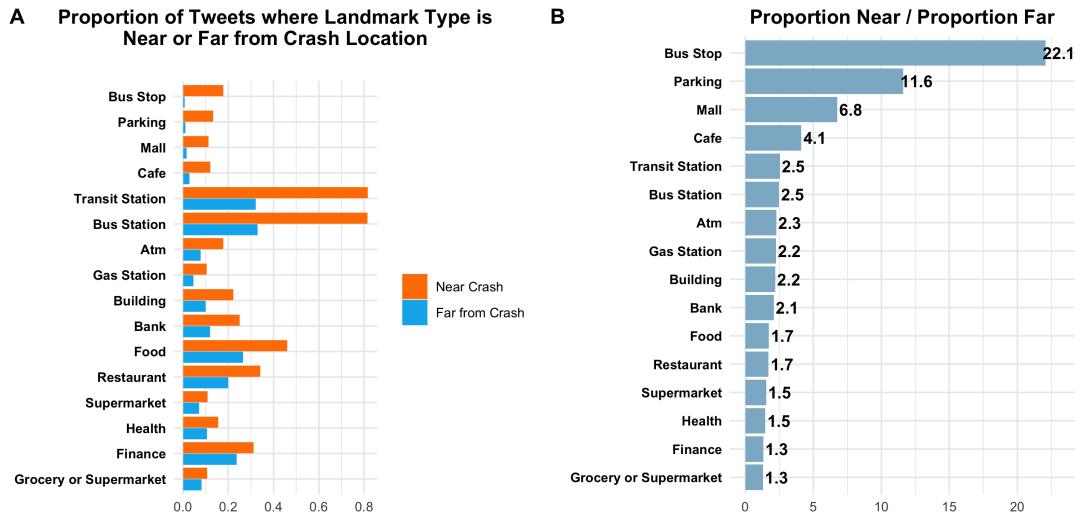


Figure 5: In Paper: Figure s2

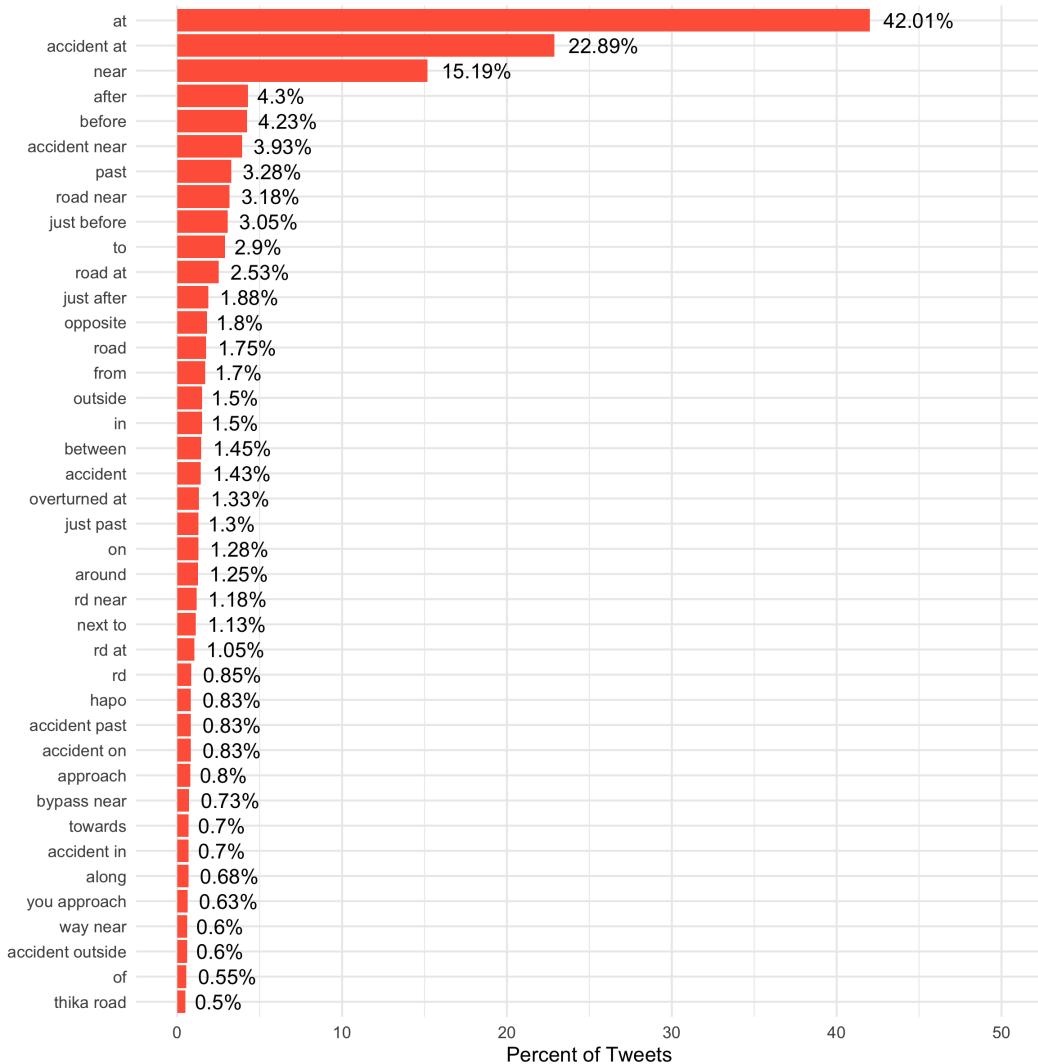


Figure 6: In Paper: Figure s3

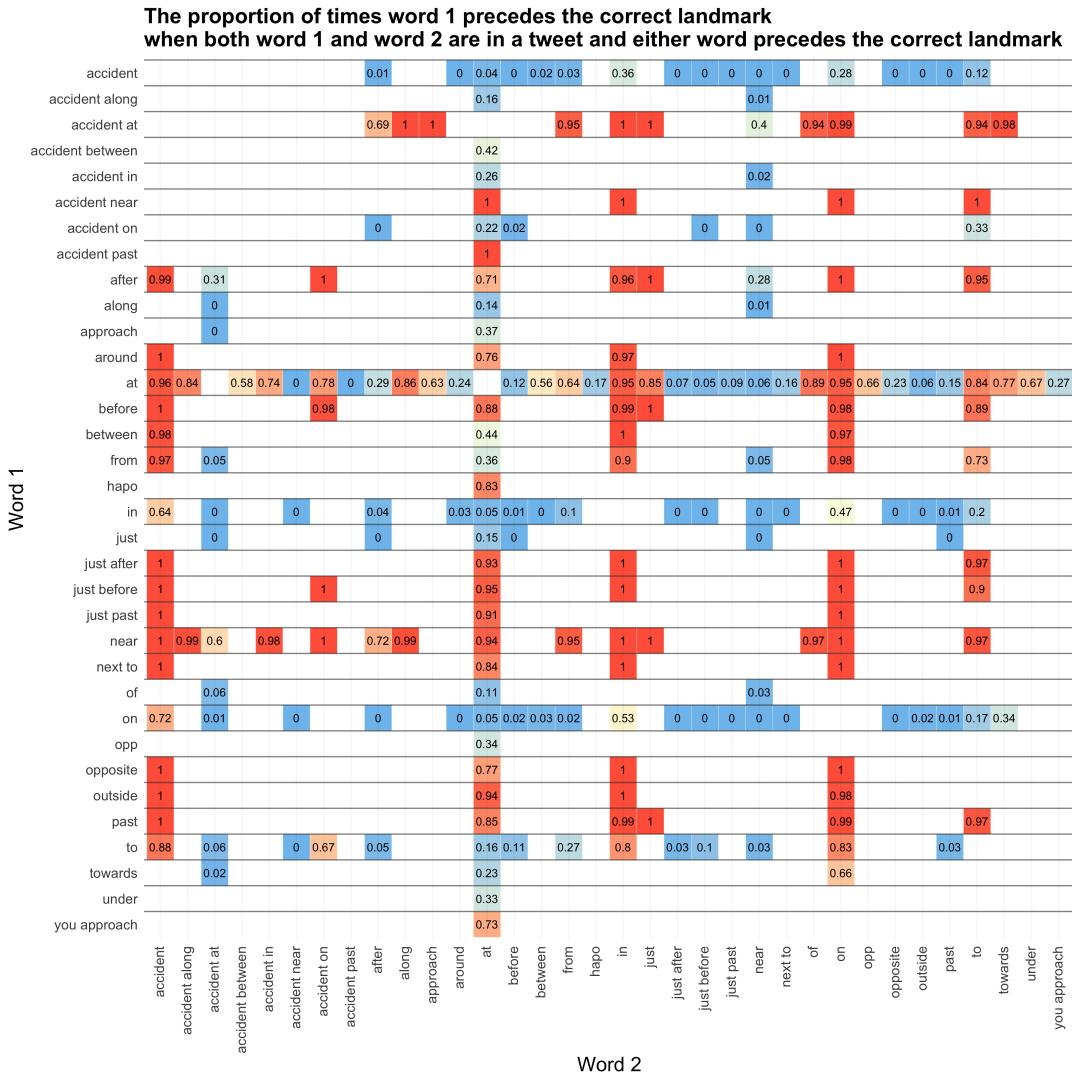


Figure 7: In Paper: Figure s4

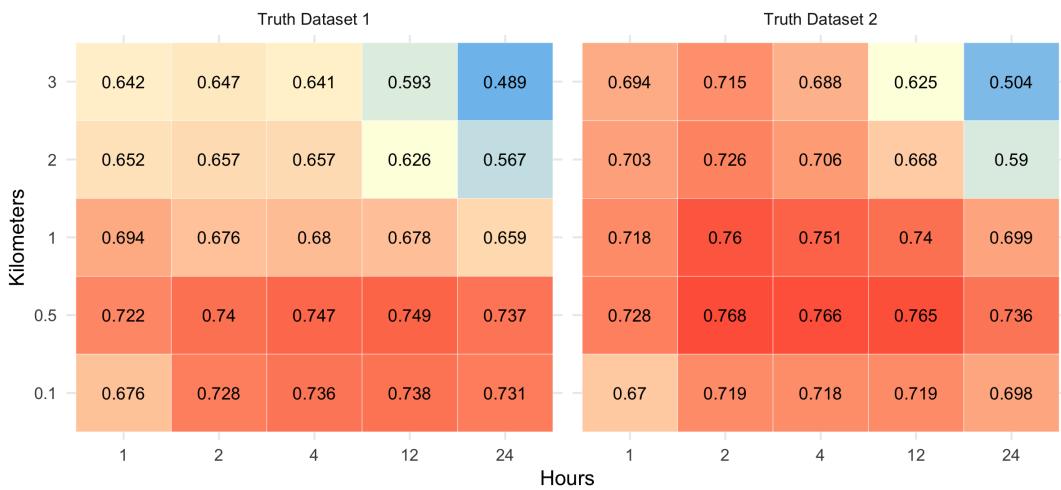
Table 3: In Paper: Table s5

Any Location Captured by Algorithm Close to True Crash Location			Crash Location Determined by Algorithm Close to True Crash Location		Algorithm Cluster Contains True Crash Loction	
	Recall	Precision	Recall	Precision	Recall	Precision
LNE_x						
LNE _x Aug Gaz	0.674	0.686	0.129	0.132	0.175	0.125
Algorithm - by Source						
Aug Gaz - Geonames	0.167	0.376	0.147	0.492	0.157	0.47
Aug Gaz - Google	0.79	0.853	0.65	0.812	0.661	0.78
Aug Gaz - OSM	0.518	0.693	0.431	0.73	0.446	0.691
Algorithm - All Sources						
Raw Gaz	0.695	0.757	0.57	0.761	0.582	0.723
Aug Gaz	0.798	0.857	0.656	0.813	0.666	0.777

Table 4: In Paper: Table s6

Variable	Min	Quartile 1	Median	Mean	Quartile 3	Max
Truth Dataset 1						
Hours Diff	0	0.133	0.55	1.68	1.693	23.776
KMs Diff	0	0	0.013	0.213	0.138	3.328
N Tweets	2	2	2	3.324	3	44
Truth Dataset 2						
Hours Diff	0	0.161	0.597	1.715	2.144	23.417
KMs Diff	0	0	0.024	0.276	0.199	3.13
N Tweets	2	2	2	3.275	3	19

Adjusted Rand Index



Jaccard Coefficient

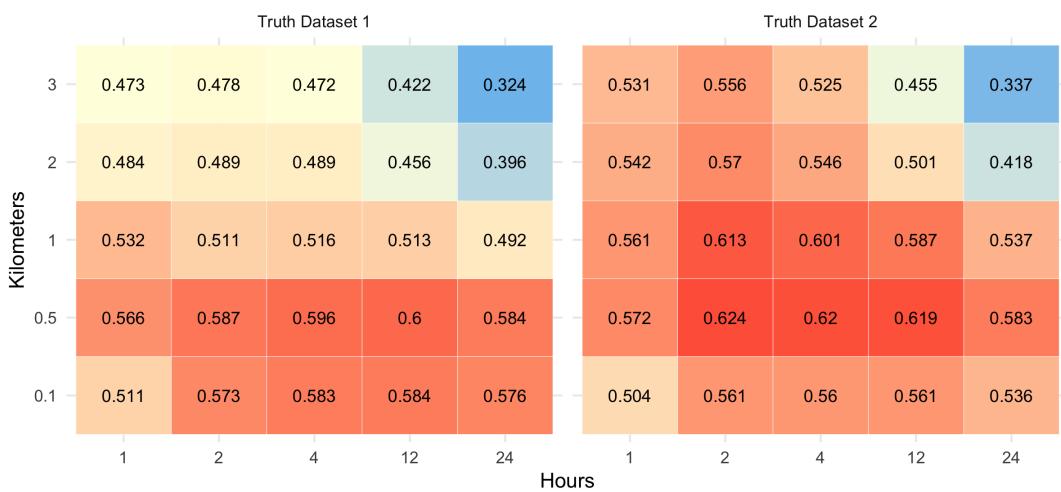


Figure 8: In Paper: Figure s5