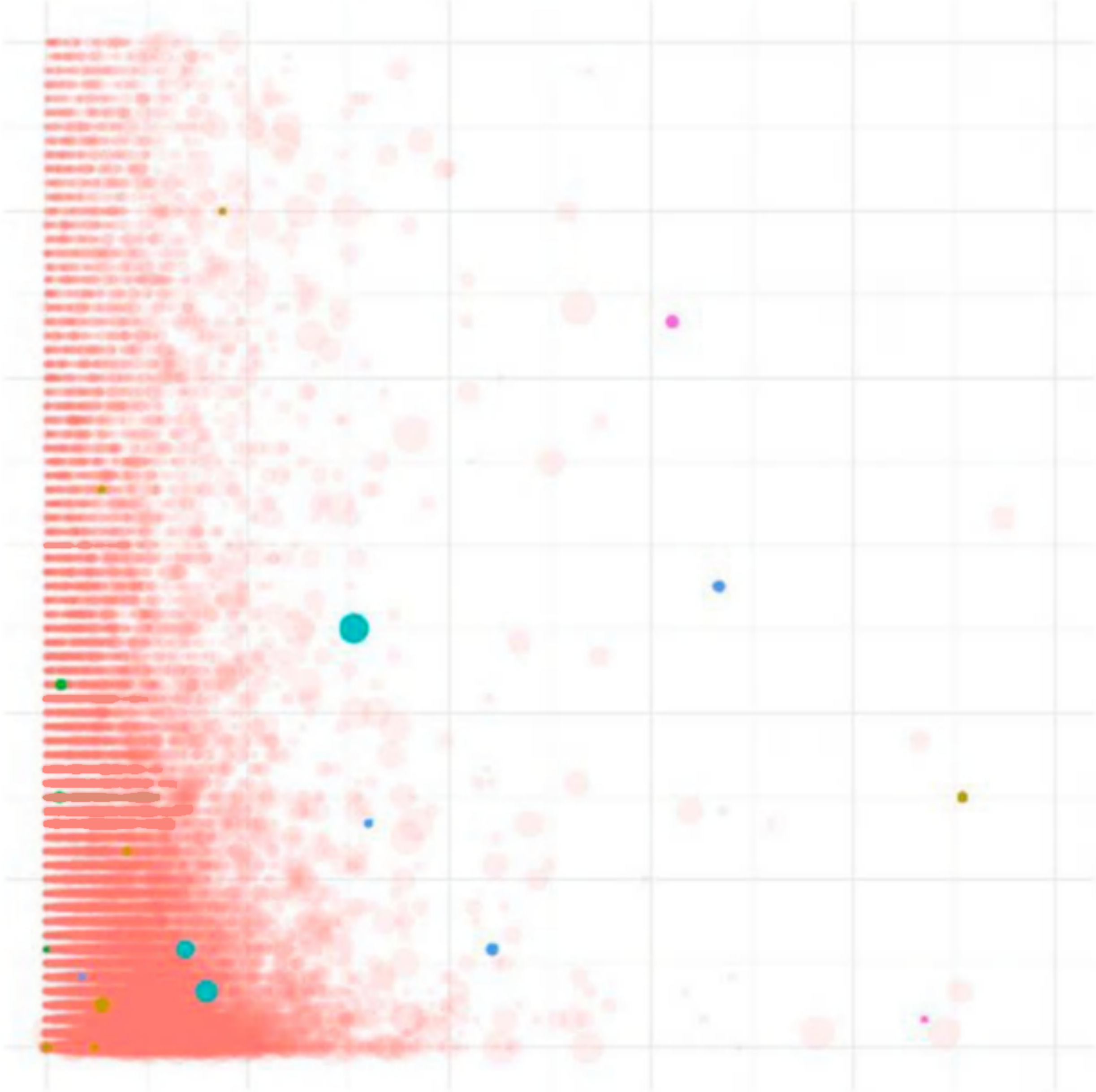
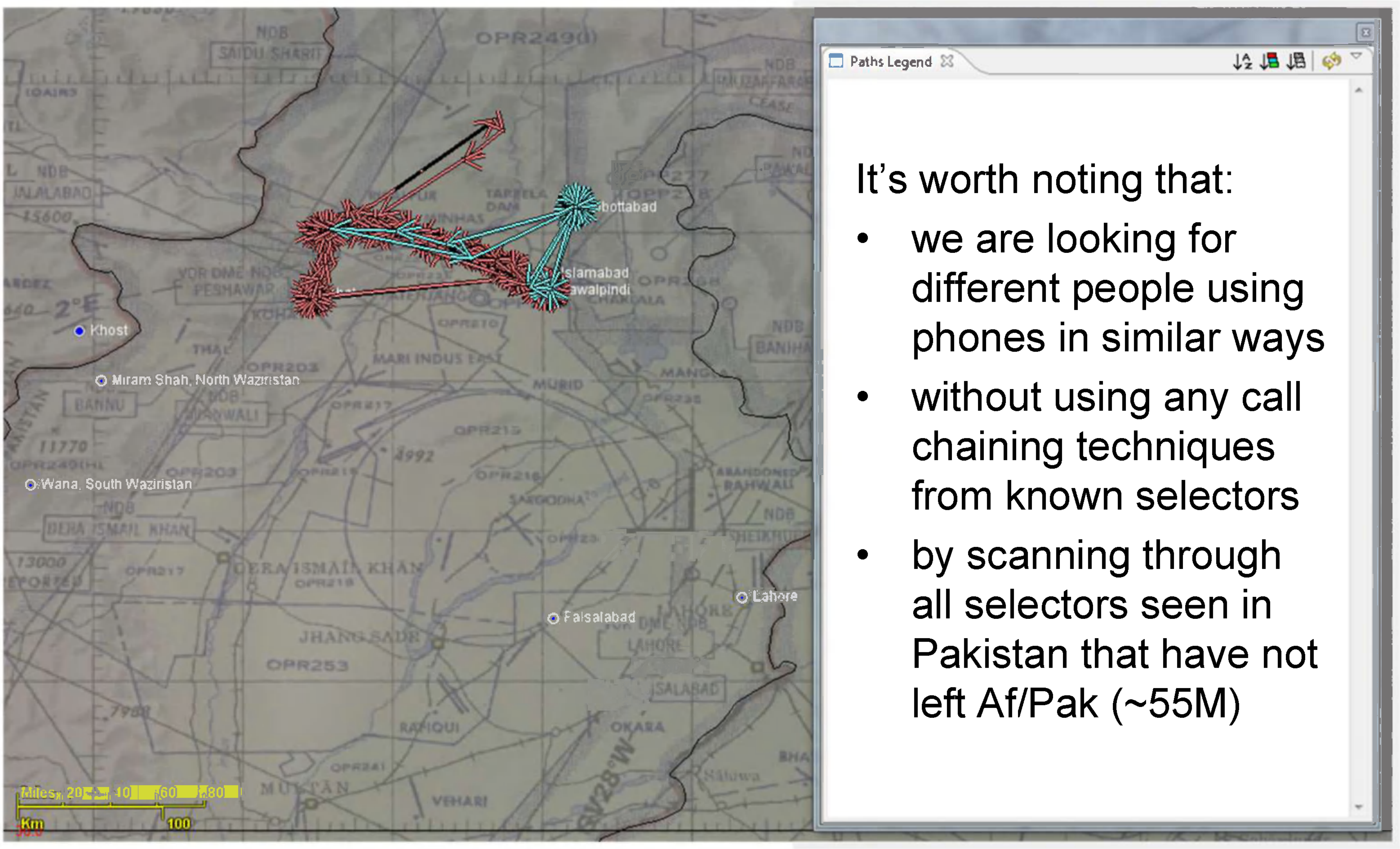
SKYNET:
Courier Detection via Machine Learning



[redacted], R66F/JHU
[redacted], R66F
[redacted], R66F
[redacted], T1211
[redacted], T1211
[redacted], S2I51
[redacted], S2I5/TD

June 5, 2012

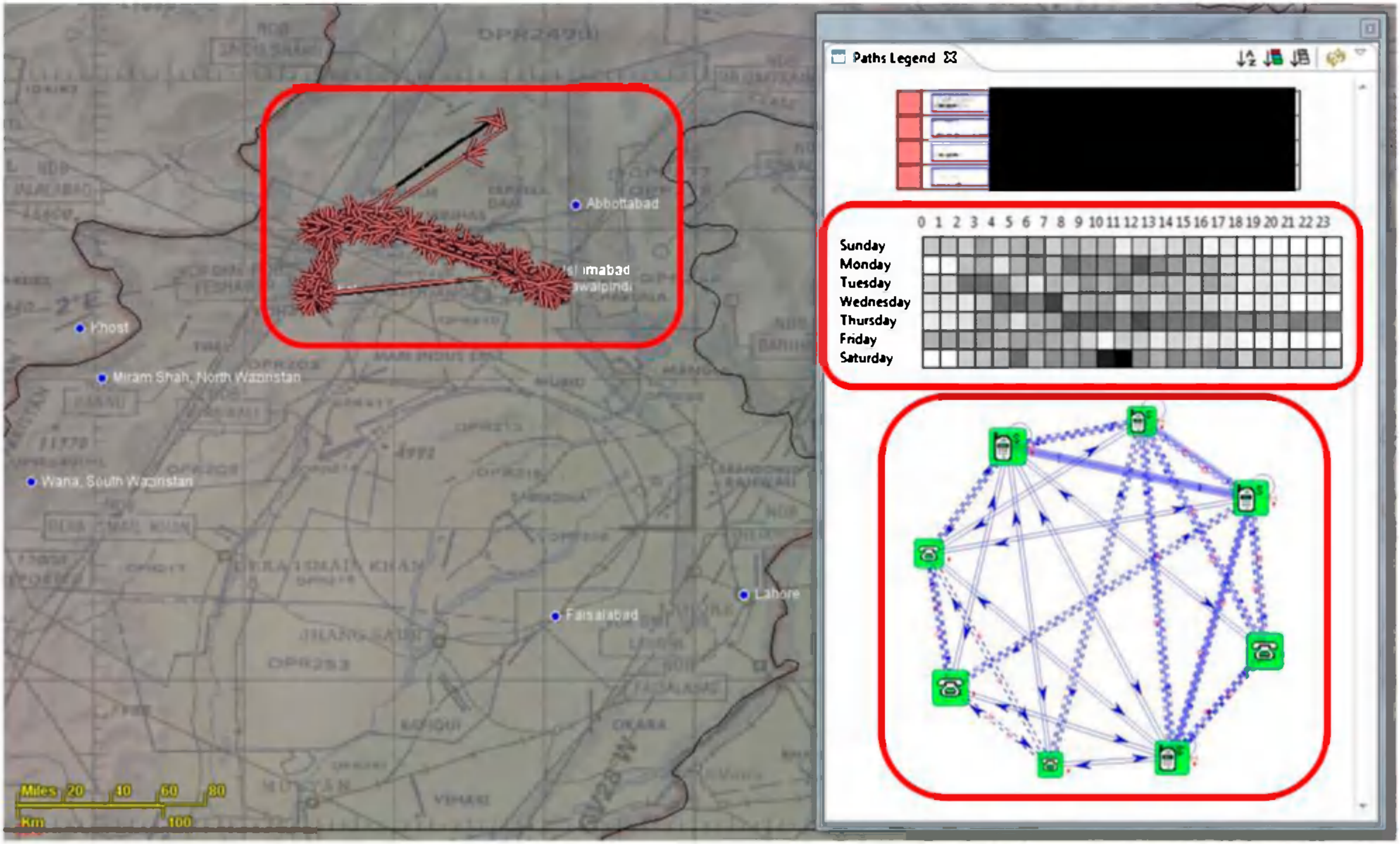
Given a handful of courier selectors, can we find others that “behave similarly” by analyzing GSM metadata?



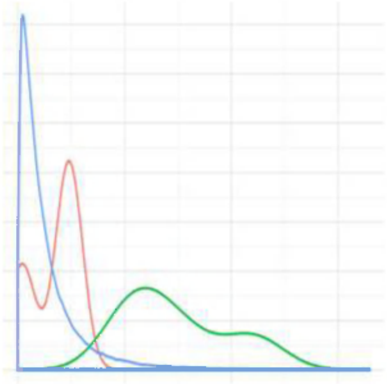
It’s worth noting that:

- we are looking for different people using phones in similar ways
- without using any call chaining techniques from known selectors
- by scanning through all selectors seen in Pakistan that have not left Af/Pak (~55M)

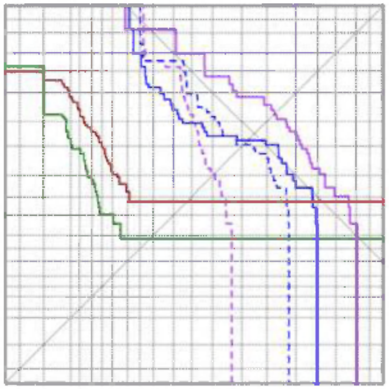
From GSM metadata, we can measure aspects of each selector's pattern-of-life, social network, and travel behavior



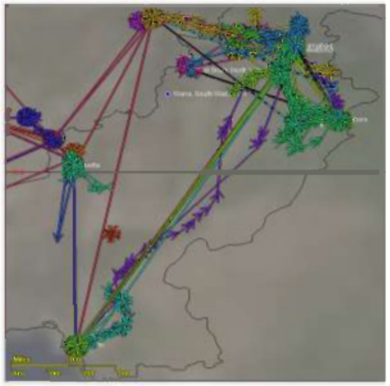
This presentation describes our search for AQSL couriers using behavioral profiling



Behavioral Feature Extraction

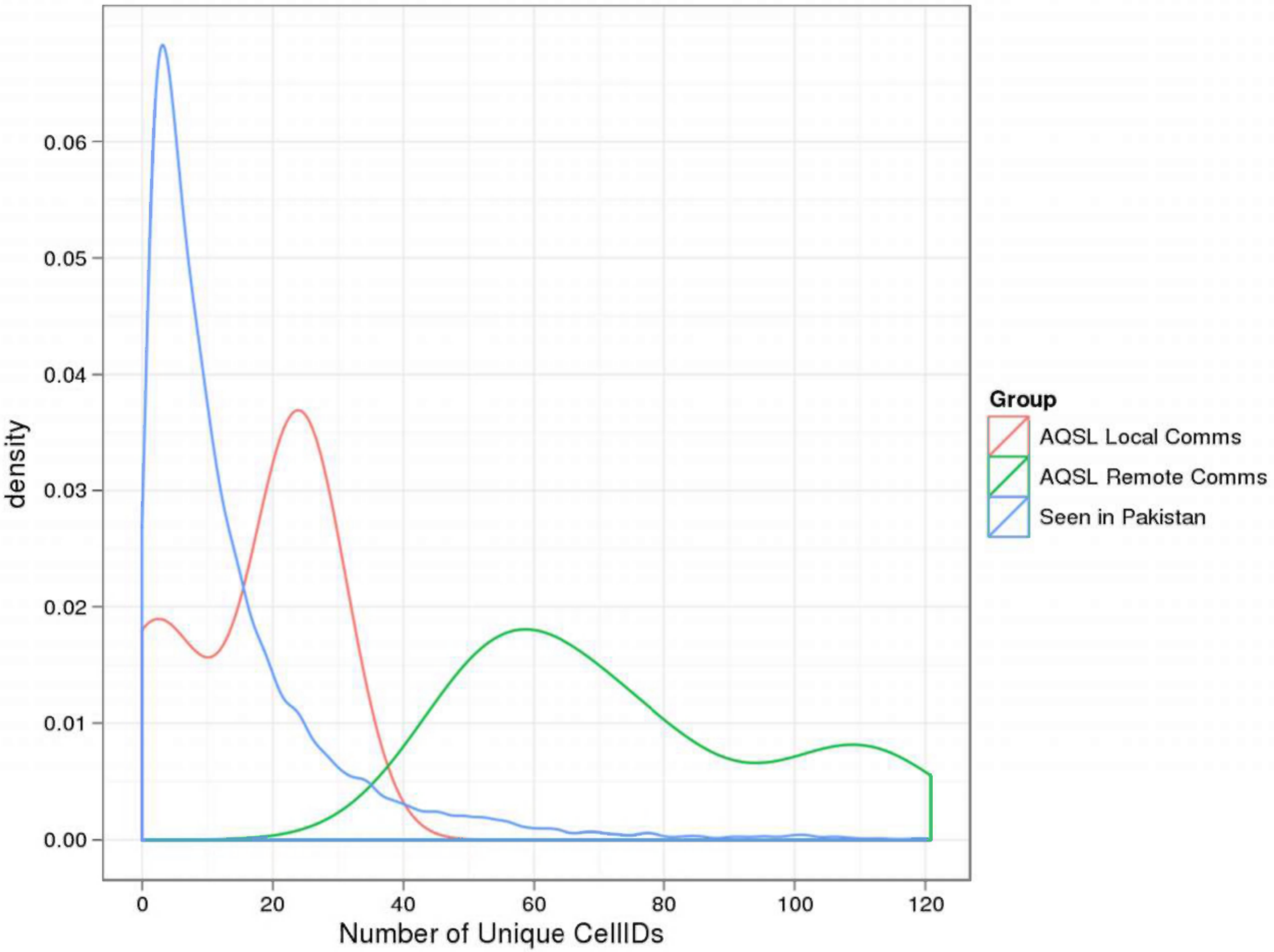


Cross Validation Experiment on AQSL Couriers

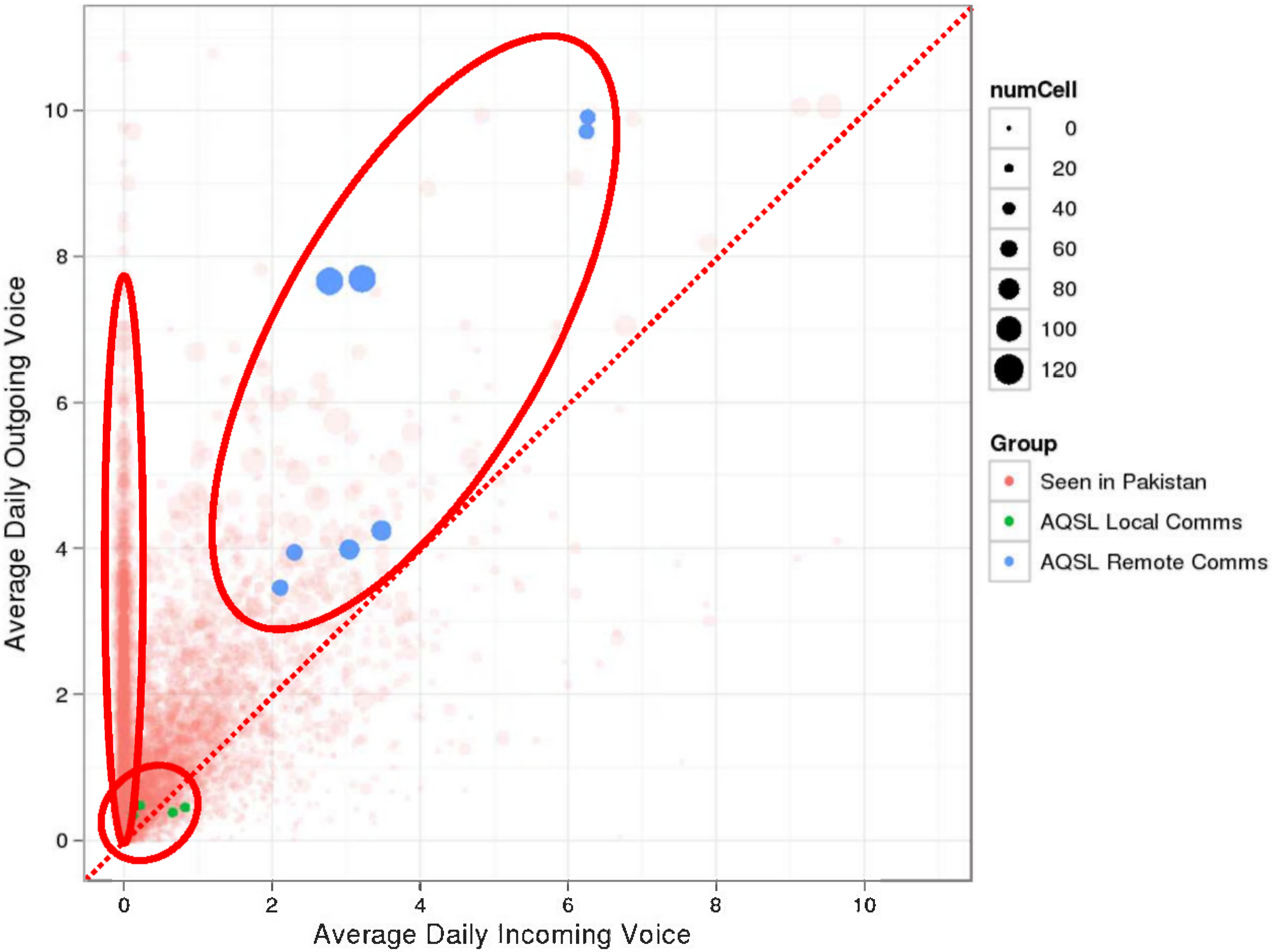


Preliminary SIGINT Findings

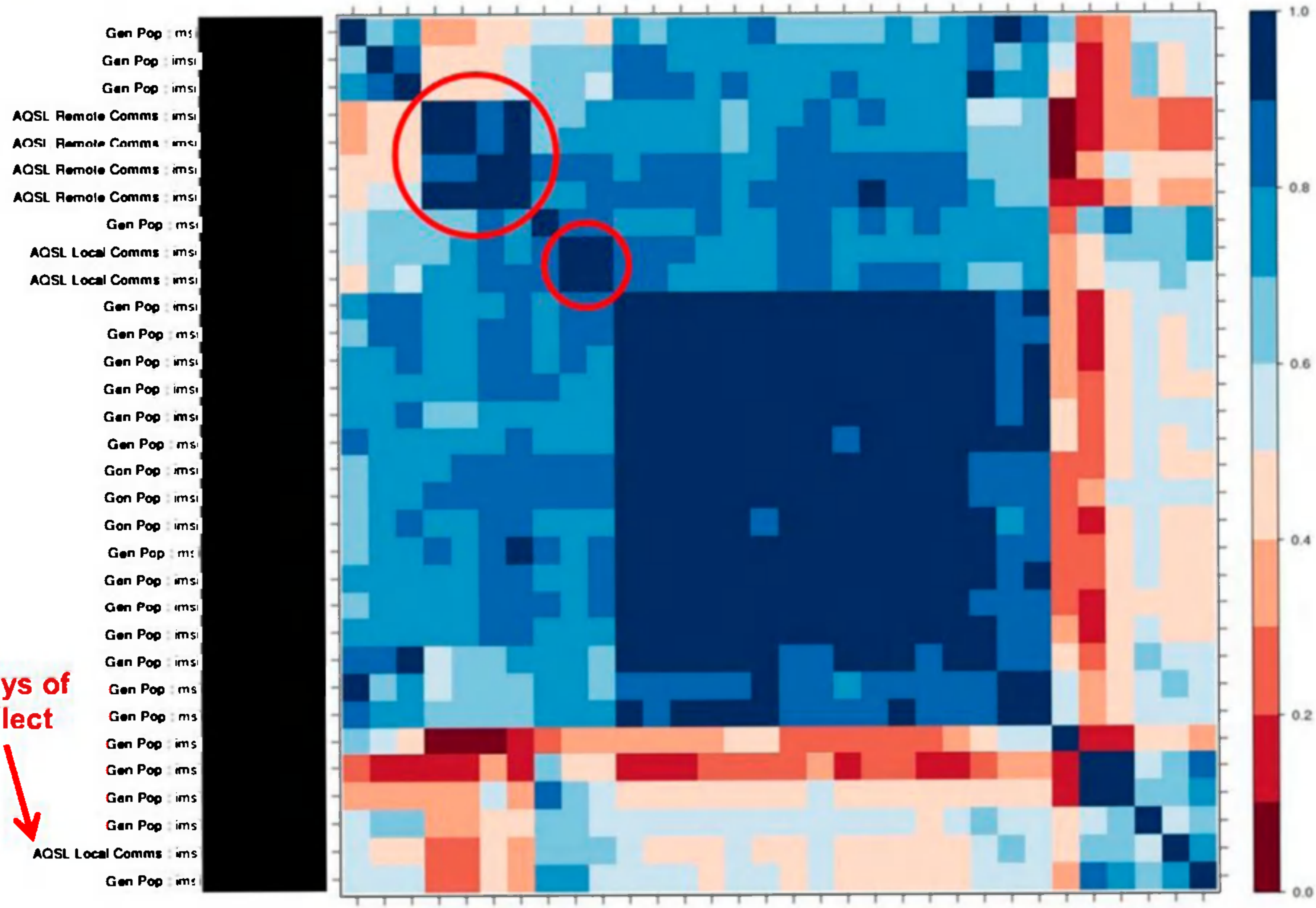
Counting unique UCELLIDs shows that couriers travel more often than typical Pakistani selectors



By examining multiple features at once, we can see some indicative behaviors of our courier selectors



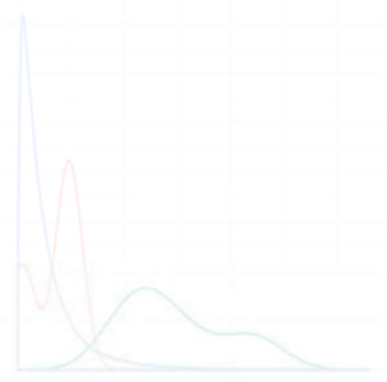
Looking at a hierarchical clustering derived from all 80 features, the AQSL groups mostly stay together



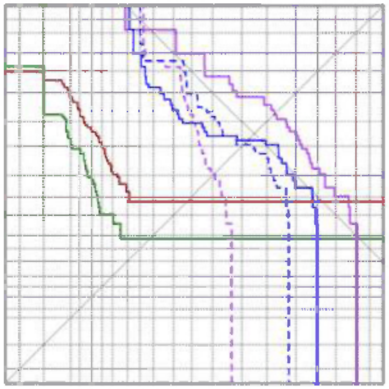
[redacted]

4 days of collect

Now, we’ll describe a cross validation experiment on the AQSL selectors that we were provided



Behavioral Feature Extraction



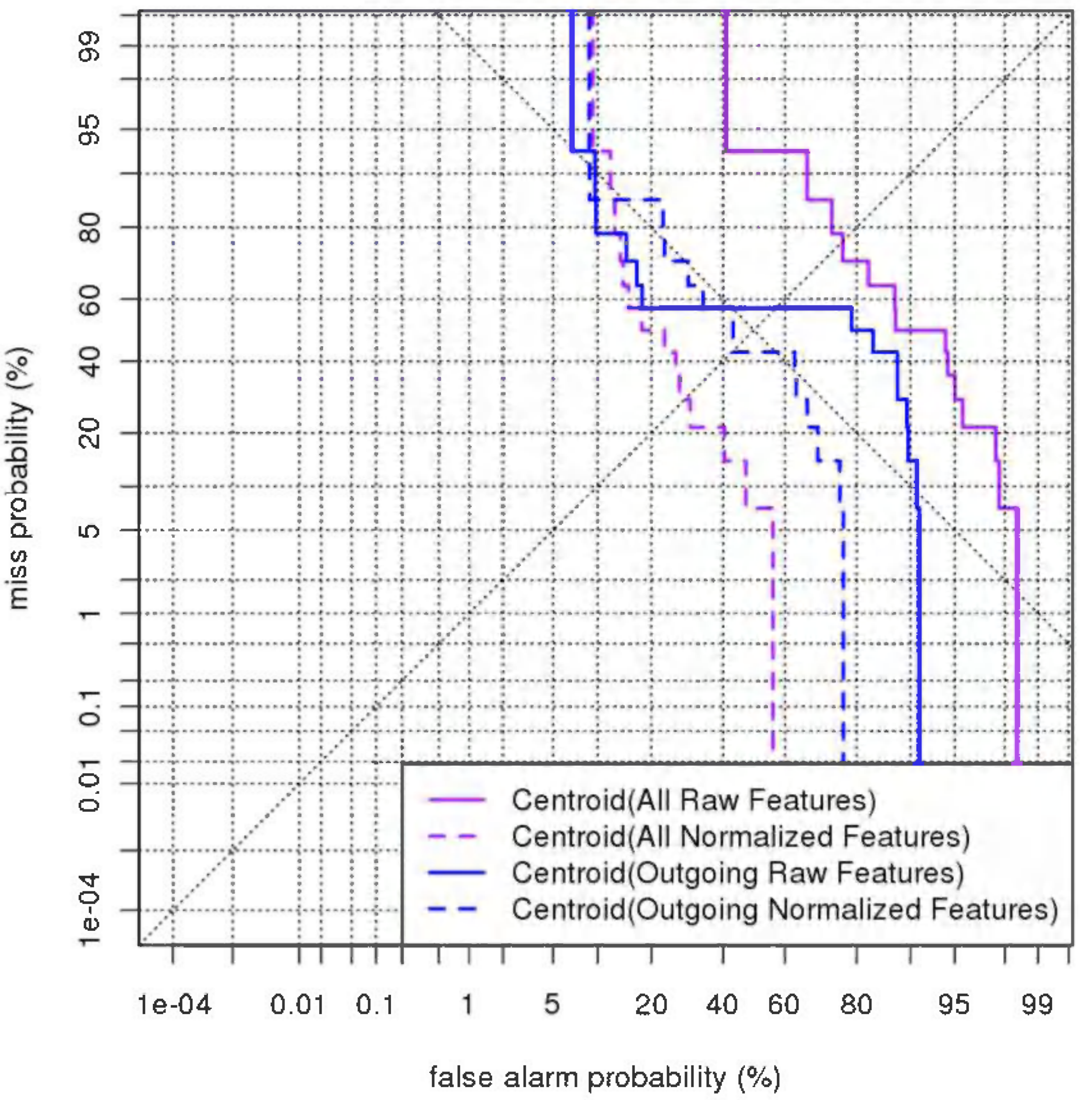
Cross Validation Experiment on AQSL Couriers

Preliminary SIGINT Findings

Our initial detector uses the centroid of the AQSL couriers to “find other selectors like these”

AQSL Cross-Validation Experiment

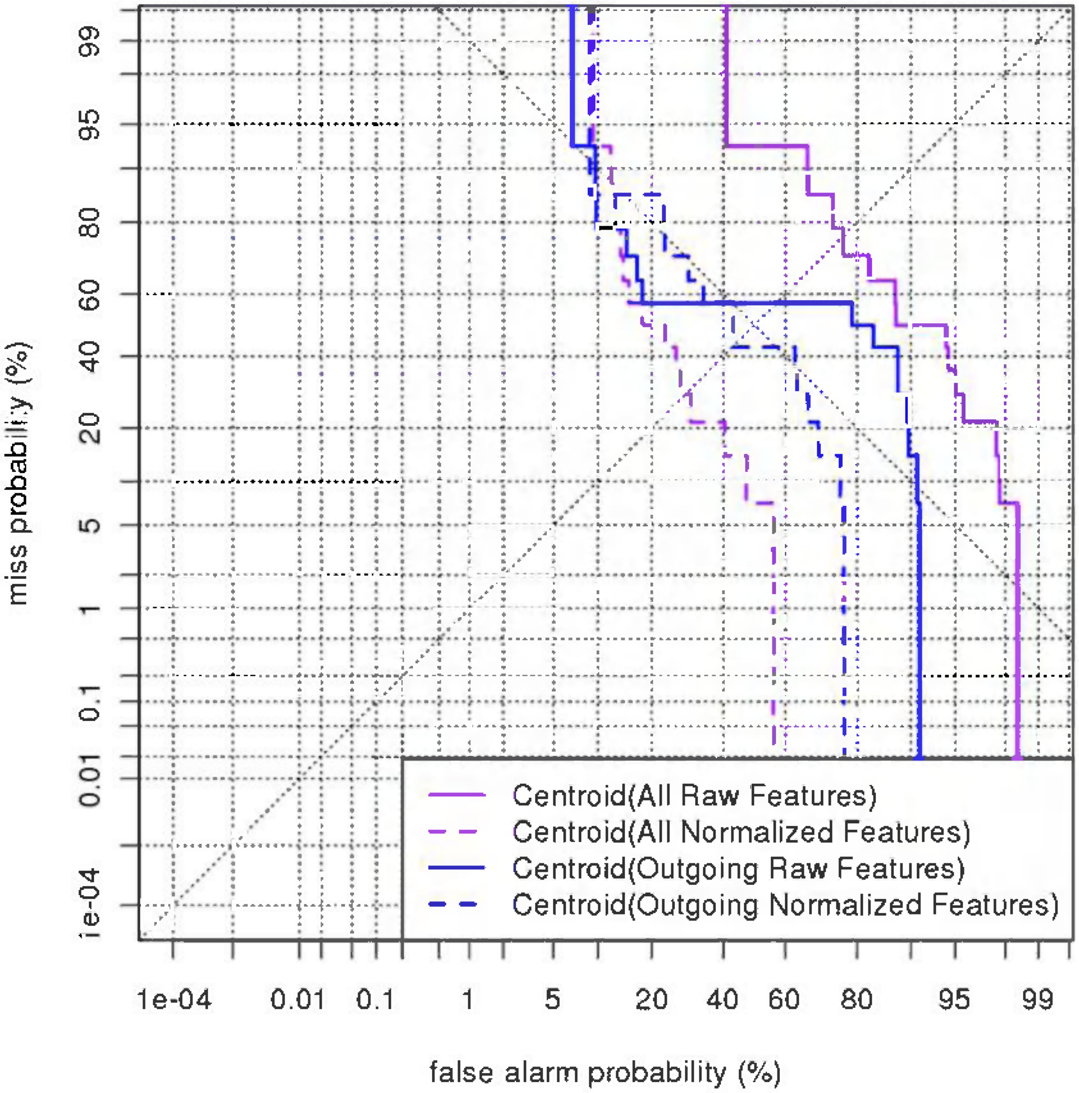
- 7 MSISDN/IMSI pairs
- Hold each pair out and score them when training the centroid on the rest
- Assume that random draws of Pakistani selectors are nontargets
- How well do we do?



Our initial detector uses the centroid of the AQSL couriers to “find other selectors like these”

AQSL Cross-Validation Experiment

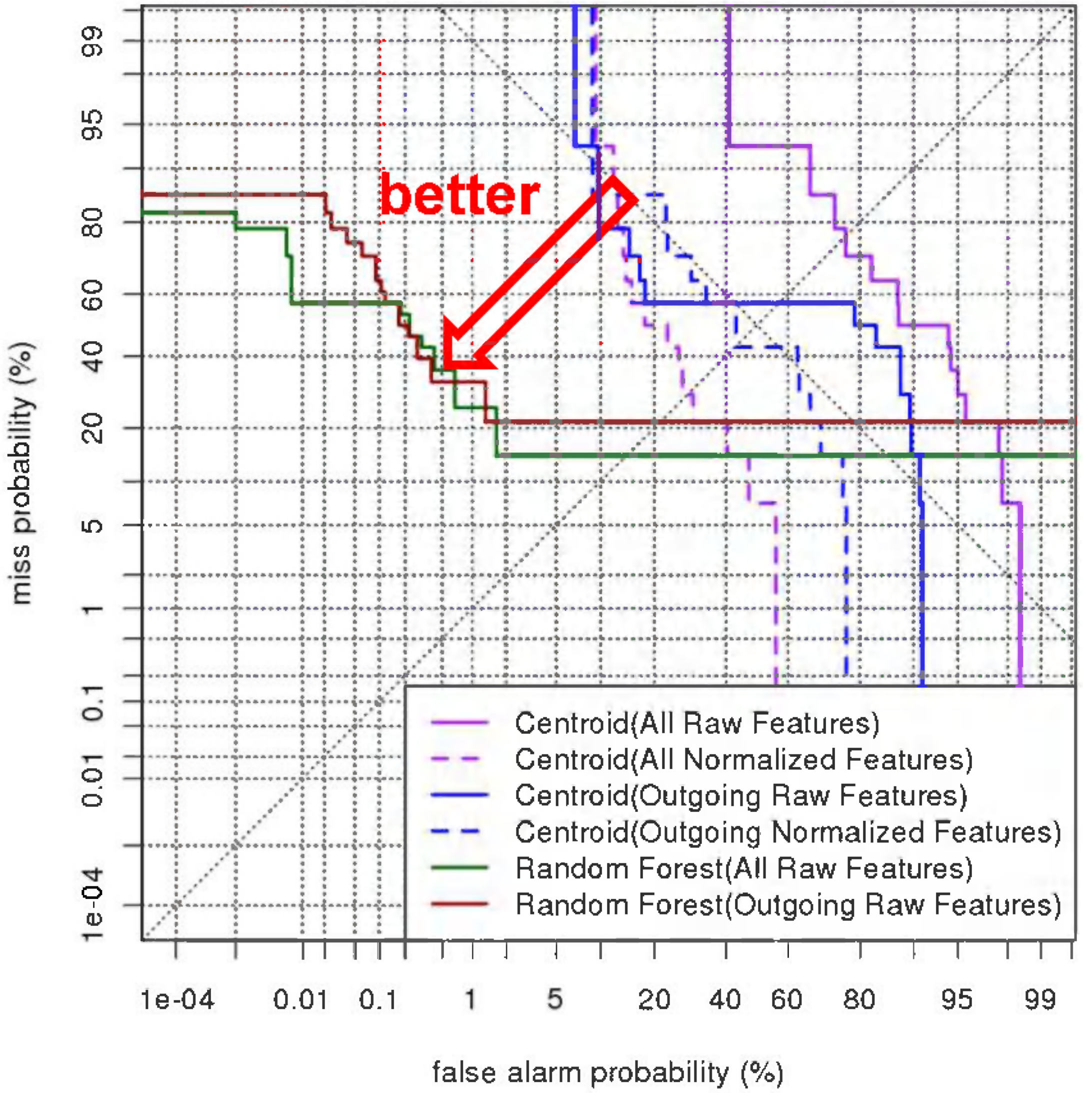
- Initial experiments showed EER in 10-20% range
- Here, performance is much worse against these nontargets:
    - Seen in Pakistan
    - Not seen outside of Af/Pak
    - Not FVEY selectors



Statistical algorithms are able to find the couriers at very low false alarm rates, if we’re allowed to miss half of them

Random Forest Classifier

- 7 MSISDN/IMSI pairs
- Hold each pair out and then try to find them after learning how to distinguish remaining couriers fro n other Pakistanis
  (using 100k random selectors here)
- Assume that random draws of Pakistani selectors are nontargets
- 0.18% False Alarm Rate at 50% Miss Rate



We've been experimenting with several error metrics on both small and large test sets

			100k Test Selectors		55M Test Selectors	
Training Data	Classifier	Features	False Alarm Rate at 50% Miss Rate	Mean Reciprocal Rank	Tasked Selectors in Top 500	Tasked Selectors in Top 100
None	Random	None	50%	1/23k (simulated)	0.64 (active/Pak)	0.13 (active/Pak)
Known Couriers	Centroid	All	20%	1/18k		
		Outgoing	43%	1/27k		
	Random Forest		0.18%	1/9.9	5	1
+ Anchory Selectors						

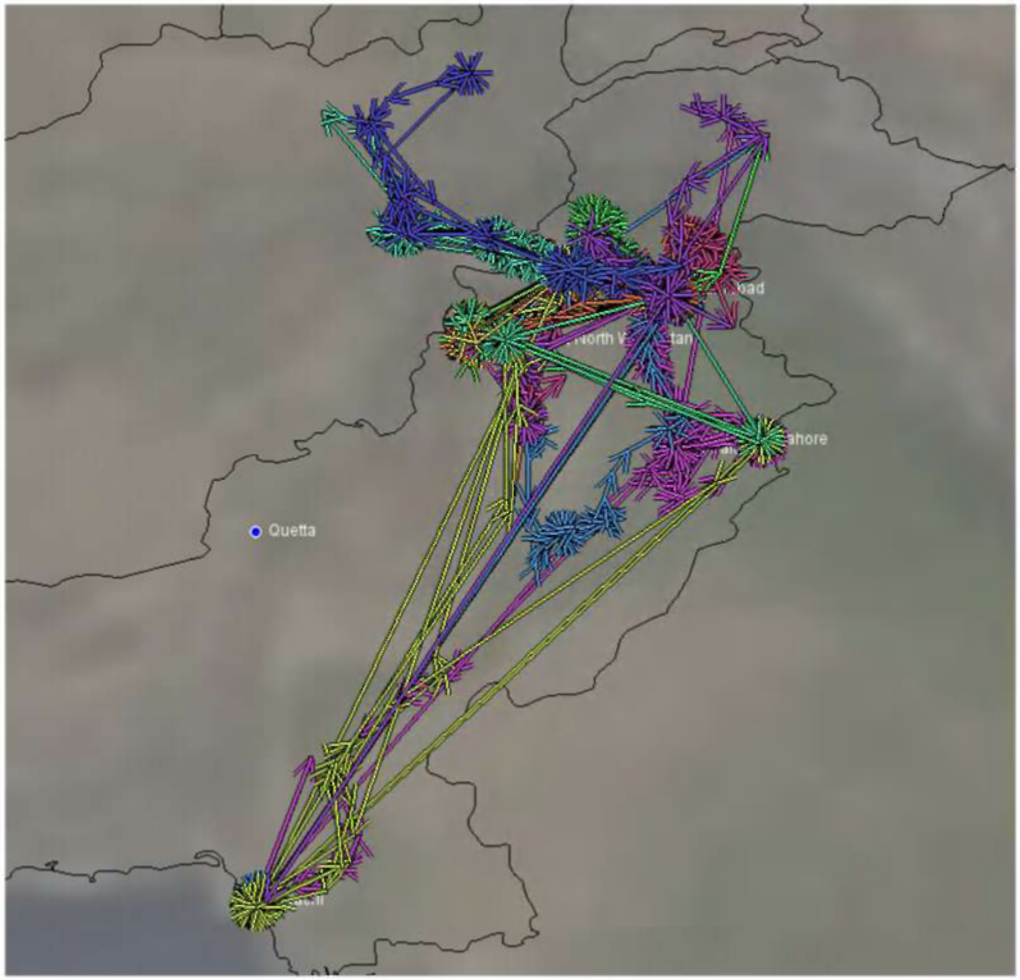
Random Forest:

- 0.18% false alarm rate at 50% miss rate
- 7x improvement over random performance when evaluating its tasked precision at 100

To get more training data we scraped selectors from S2I11 Anchory reports containing keyword “courier”

Anchory Selectors

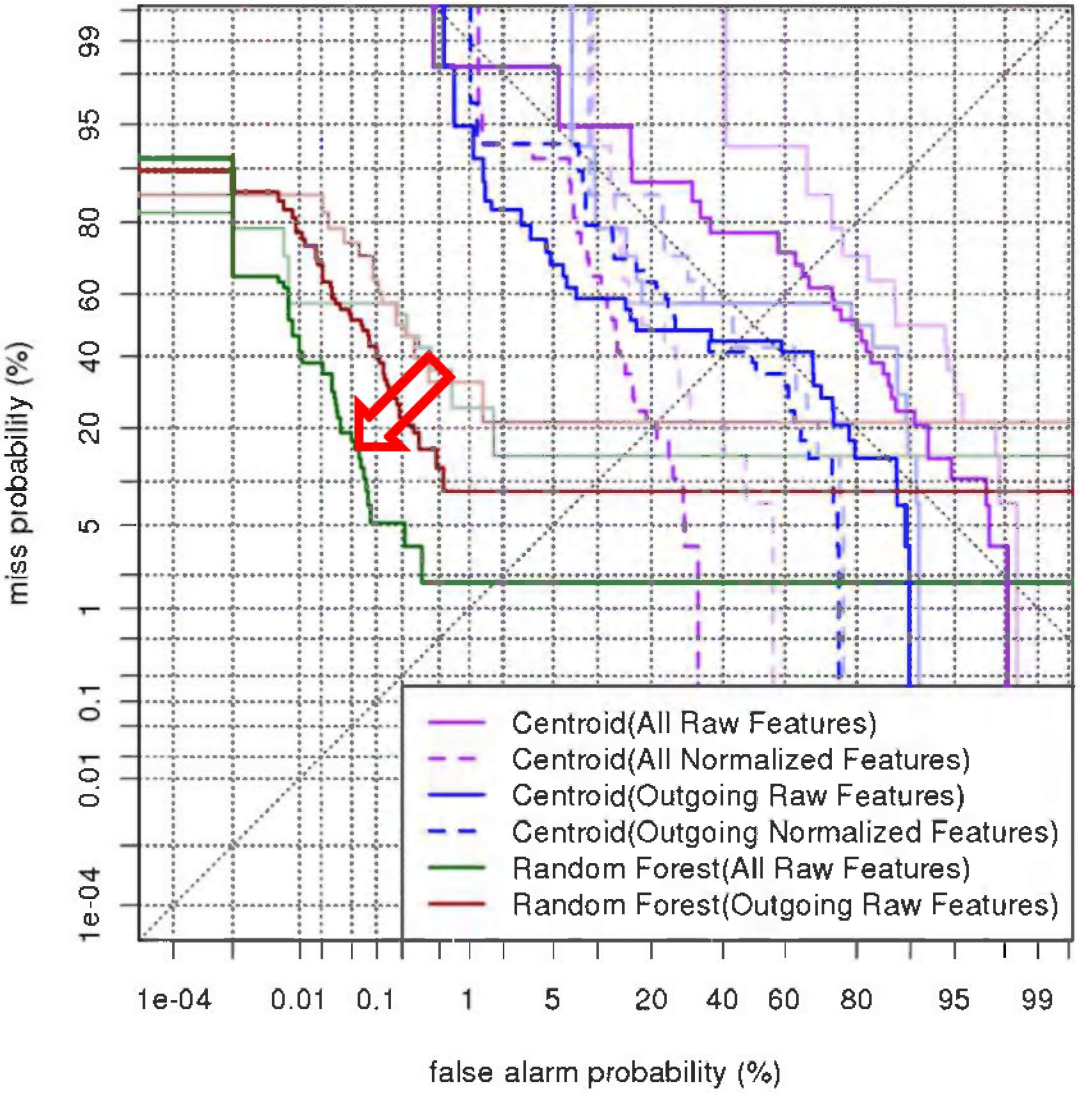
- Searched for reports containing “S2I11” AND “courier”
- Filtered out non-mobile numbers and kept selectors with “interesting” travel patterns seen in SmartTracker



Adding selectors from Anchory reports to the training data reduced the false alarm rates even further

Anchory Selectors

- Searched for reports containing “S2I11” AND “courier”
- Filtered out non-mobile numbers and kept selectors with “interesting” travel patterns seen in SmartTracker



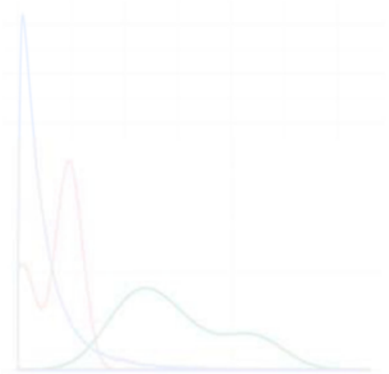
We’ve been experimenting with several error metrics on both small and large test sets

			100k Test Selectors		55M Test Selectors	
Training Data	Classifier	Features	False Alarm Rate at 50% Miss Rate	Mean Reciprocal Rank	Tasked Selectors in Top 500	Tasked Selectors in Top 100
None	Random	None	50%	1/23k (simulated)	0.64 (active/Pak)	0.13 (active/Pak)
Known Couriers	Centroid	All	20%	1/18k		
		Outgoing	43%	1/27k		
	Random Forest		0.18%	1/9.9	5	1
+ Anchory Selectors			0.008%	1/14	21	6

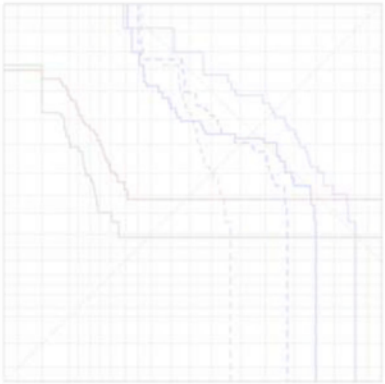
Random Forest trained on Known Couriers + Anchory Selectors:

- 0.008% false alarm rate at 50% miss rate
- 46x improvement over random performance when evaluating its tasked precision at 100

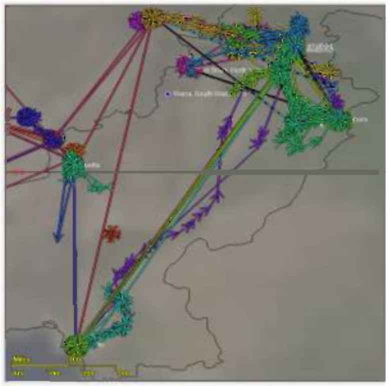
Now, we'll investigate some findings after running these classifiers on +55M Pakistani selectors via MapReduce



Behavioral Feature Extraction

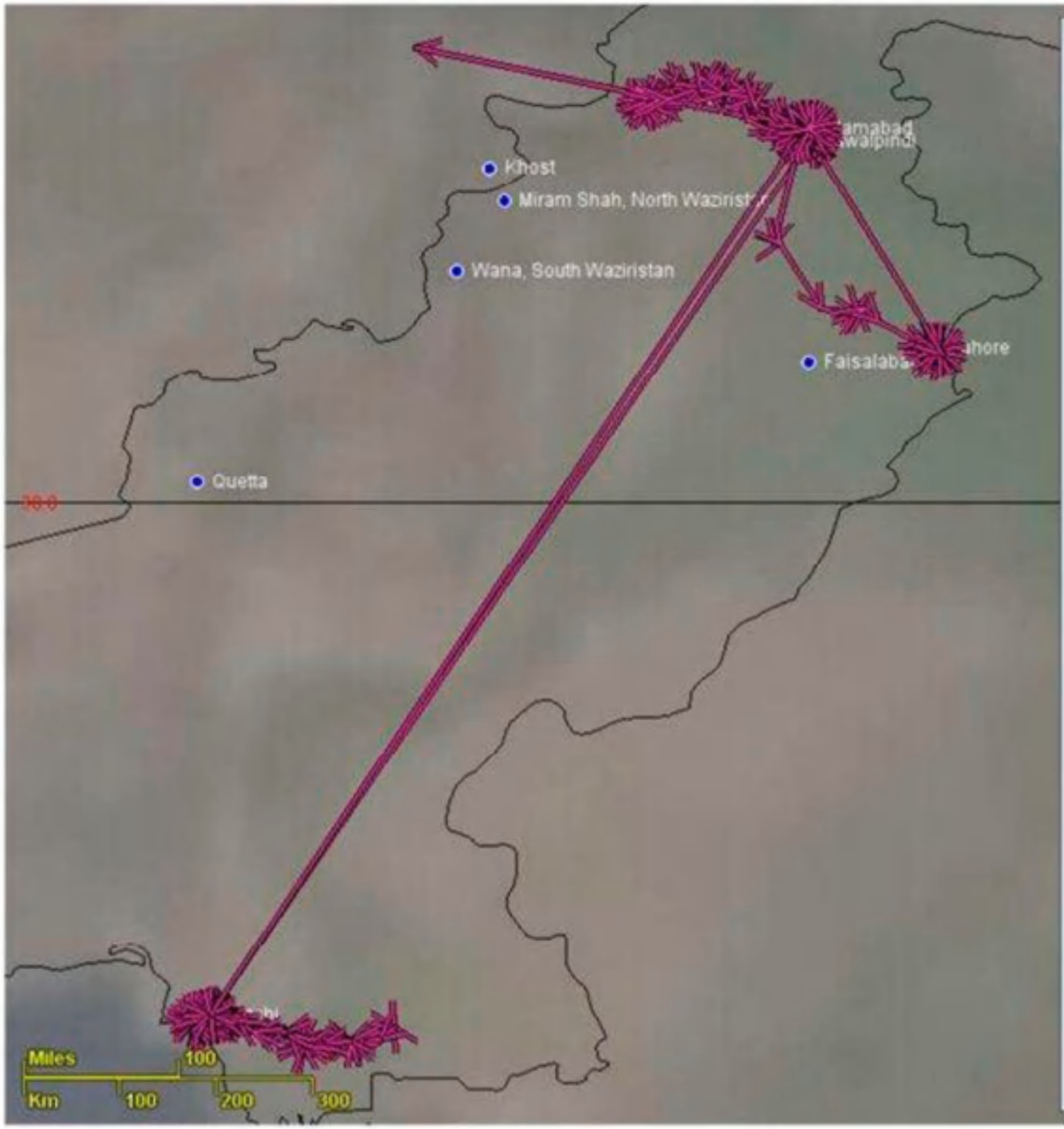


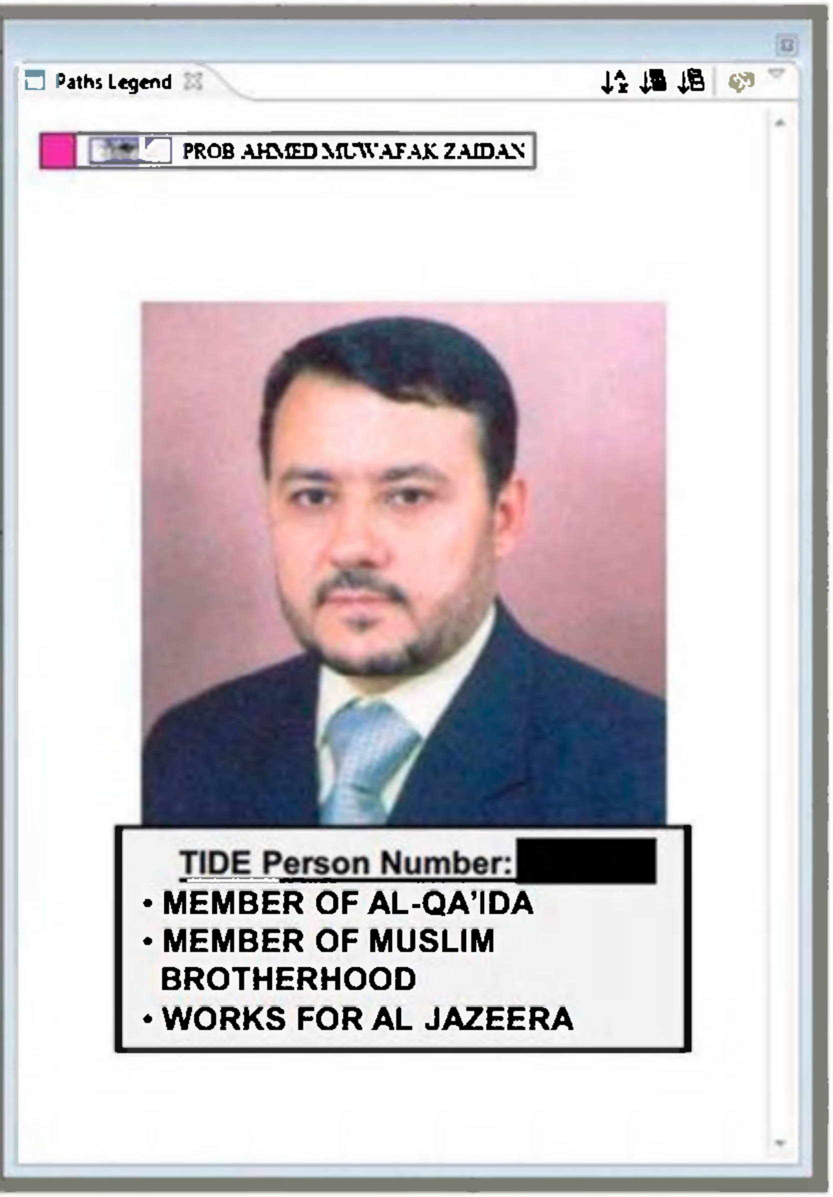
Cross Validation Experiment on AQSL Couriers



Preliminary SIGINT Findings

The highest scoring selector that traveled to Peshawar and Lahore is PROB AHMED ZAIDAN

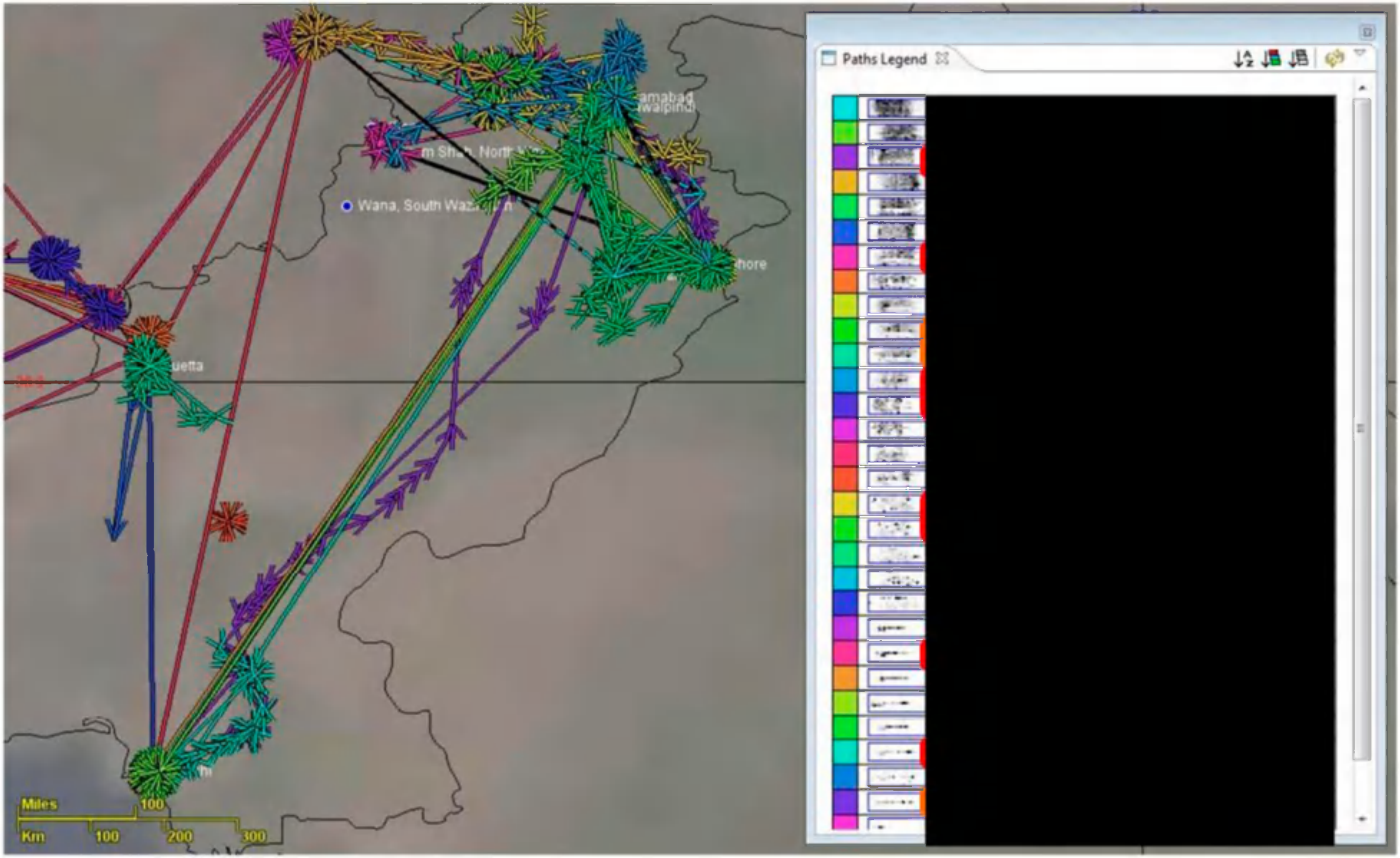




🟪 PROB AHMED MUWAFAK ZAIDAN

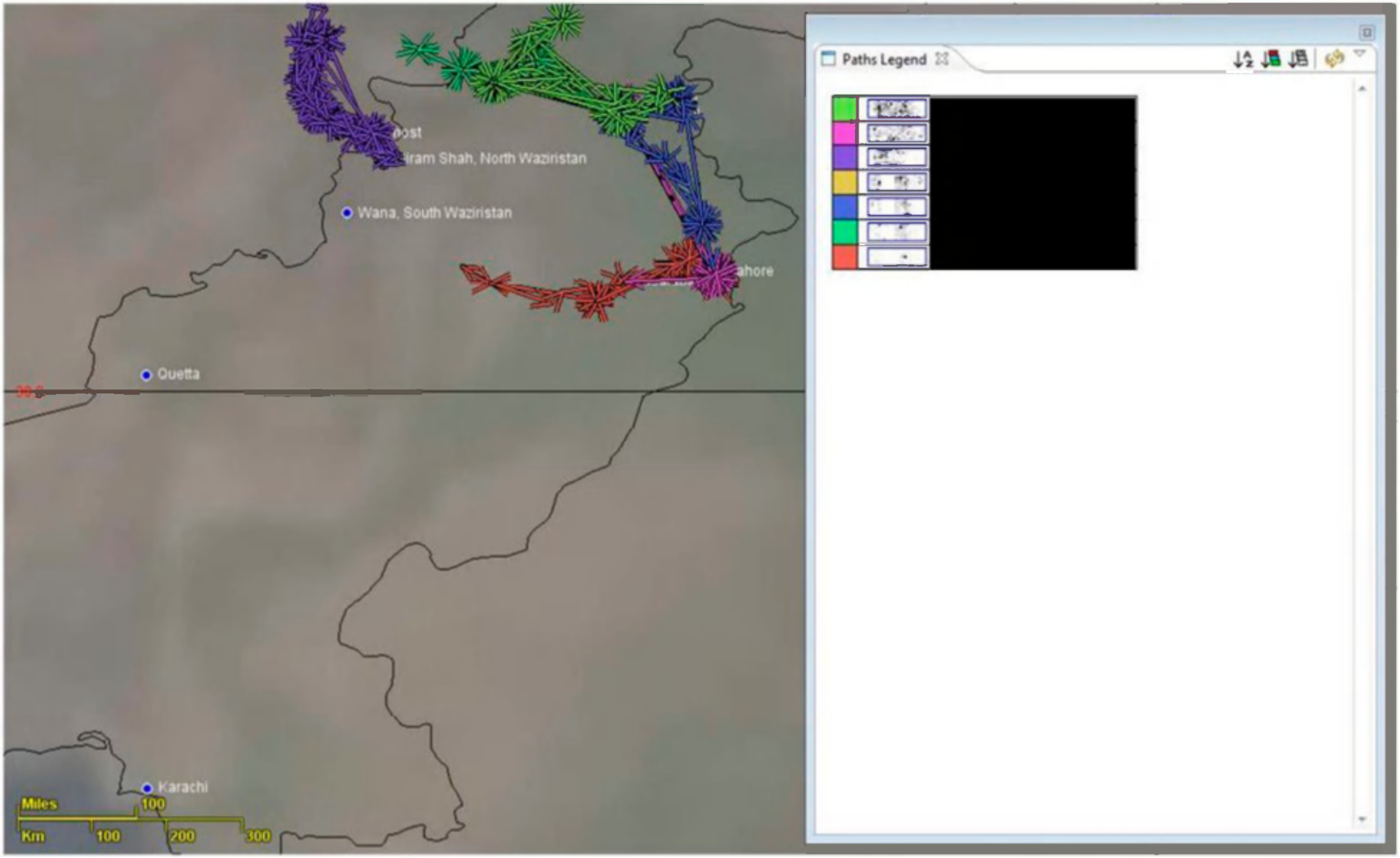
TIDE Person Number: [redacted]
- MEMBER OF AL-QA'IDA
- MEMBER OF MUSLIM BROTHERHOOD
- WORKS FOR AL JAZEERA

In the top 500 scoring selectors, 21 are tasked leading us to believe that we’re on the right track

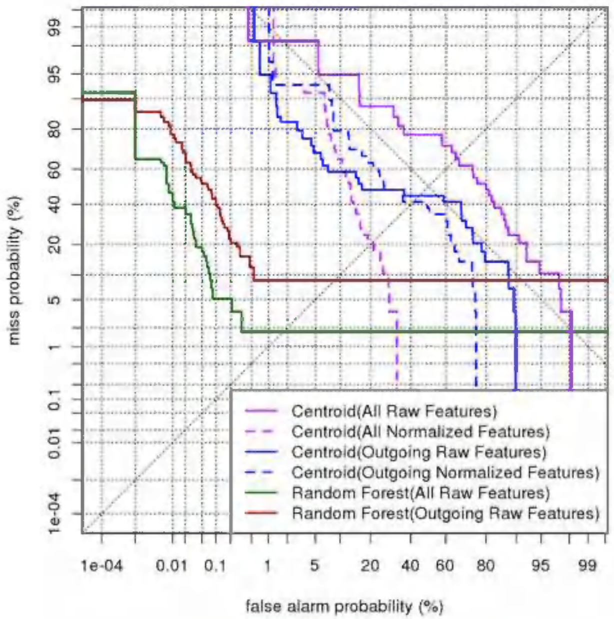


[redacted]

We have also discovered many untasked selectors with interesting travel patterns

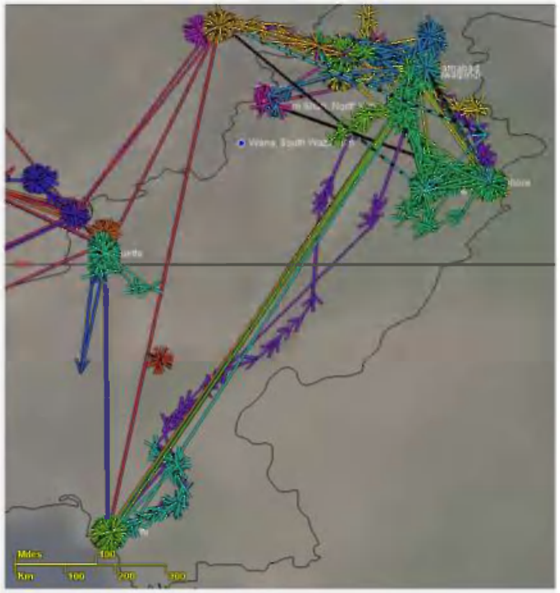


Preliminary results indicate that we're on the right track, but much remains to be done



Cross Validation Experiment:

- Random Forest classifier operating at 0.18% false alarm rate at 50% miss
- Enhancing training data with Anchory selectors reduced that to 0.008%
- Mean Reciprocal Rank is ~1/10



Preliminary SIGINT Findings:

- Behavioral features helped discover similar selectors with "courier-like" travel patterns
- High number of tasked selectors at the top is hopefully indicative of the detector performing well "in the wild"