Facultad de Ingeniería en Electricidad y Computación

Detecting Damaged Regions after Natural Disasters using Mobile Phone Data: The Case of Ecuador

Galo Castillo-López
gadacast@fiec.espol.edu.ec

Fabricio Layedra
flayedra@espol.edu.ec

María-Belén Guaranda
mguarand@espol.edu.ec

Carmen Vaca
cvaca@fiec.espol.edu.ec

INTRODUCTION

Large scale natural disasters involve budgetary problems for governments even when local and foreign cash donations are available. Prioritizing investment requires near real time information about the impact of the hazard in different locations. Such information is not available through sensors or other devices specially in developing countries that do not have such infrastructure.

A rich source of information is the data resulting from mobile phone activity that citizens in affected areas start using as soon as they become available after the disaster, hours in some cases. We exploit such source of information in this work to conduct different analyses and ML techniques in order to identify the affected zones after the earthquake that took place in the Ecuadorian province of Manabí on April 16th, 2016.

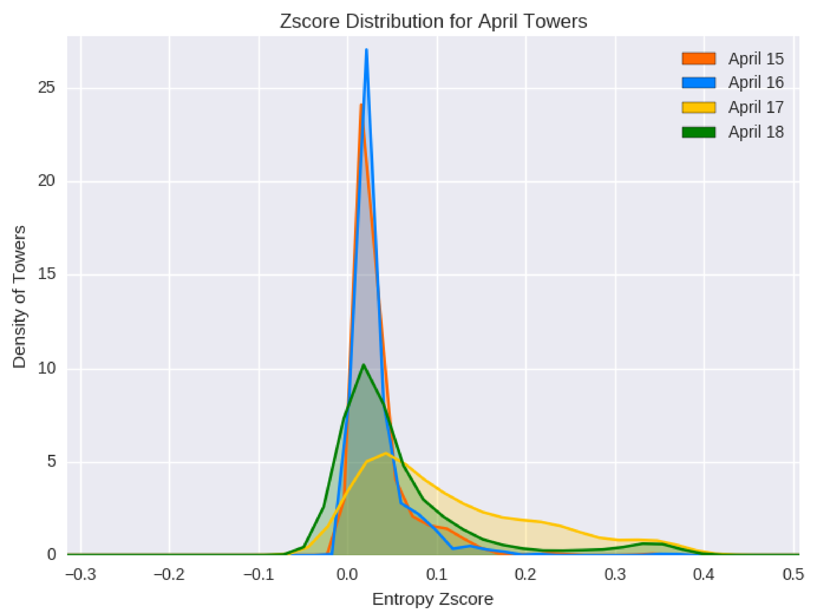
RQ: Can we identify the most affected zones in a shorter period than governments, using mobile phone data?

METHODS

We calculated three metrics to characterize the damage in each canton to be later used as features for three ML models.

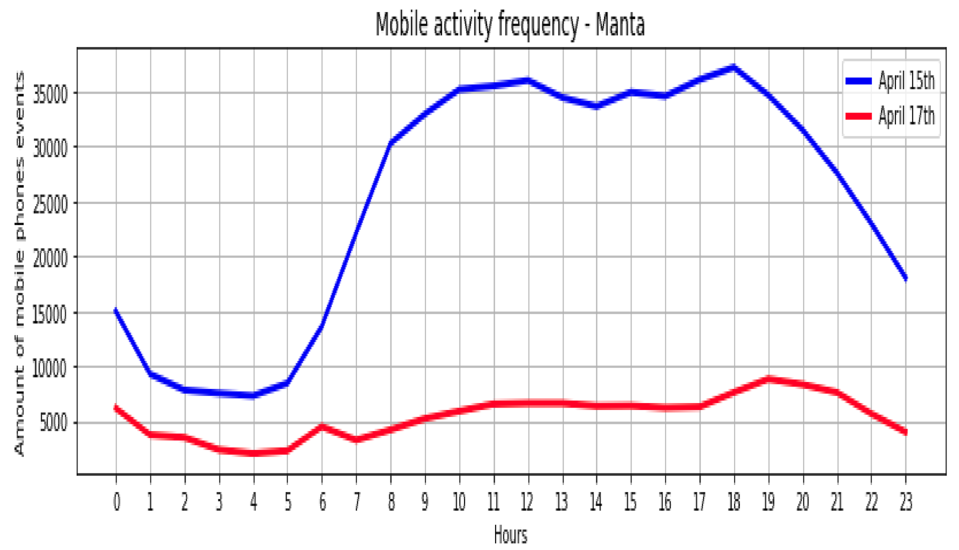
Visitors Diversity Index - Spatial feature

$$VDI(i) = \frac{-\Sigma_{a=1}^{A} \rho_{ia} log(\rho_{ia})}{log(A)}$$

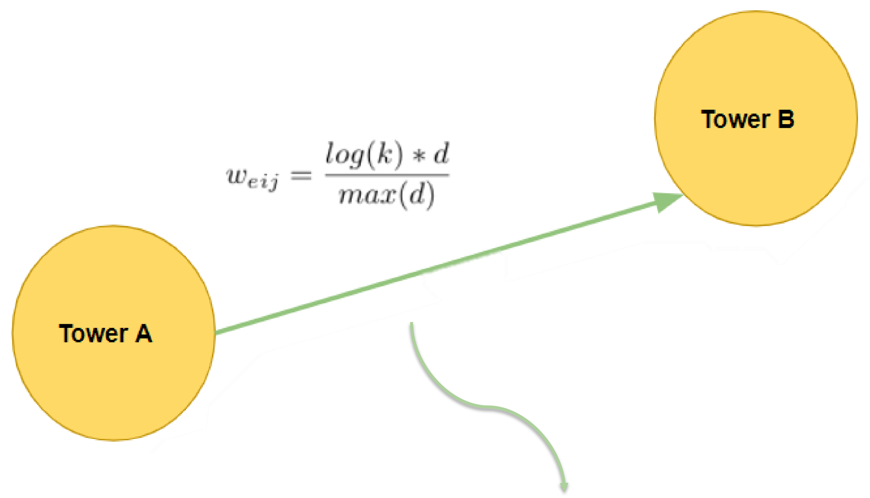


Time Series Dissimilarity - Temporal feature

$$dsm = \sqrt{\sum_{i=0}^{23} (a(t_i) - b(t_i))^2}$$



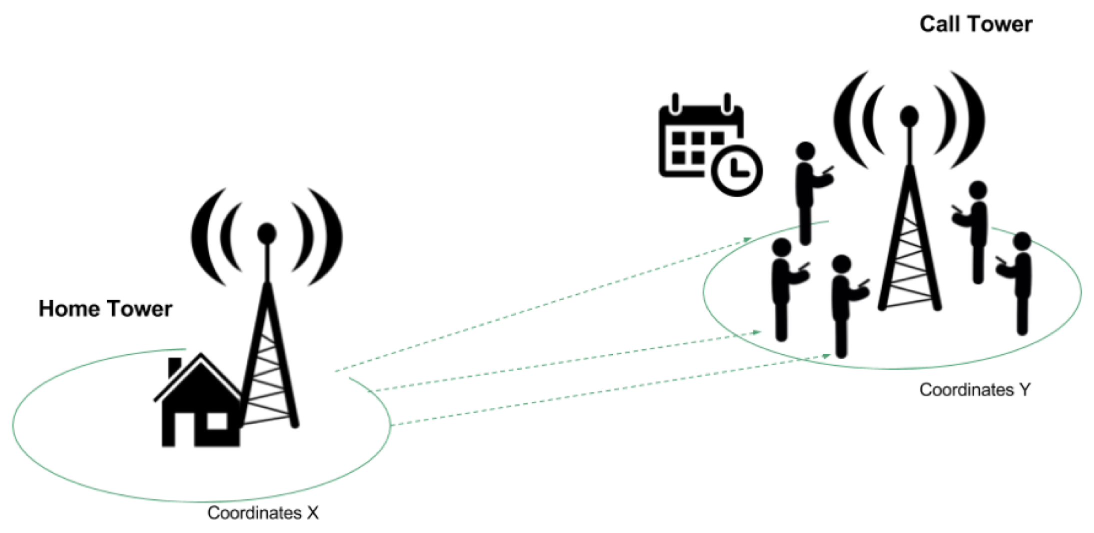
Eigenvector centrality (Networks feature)



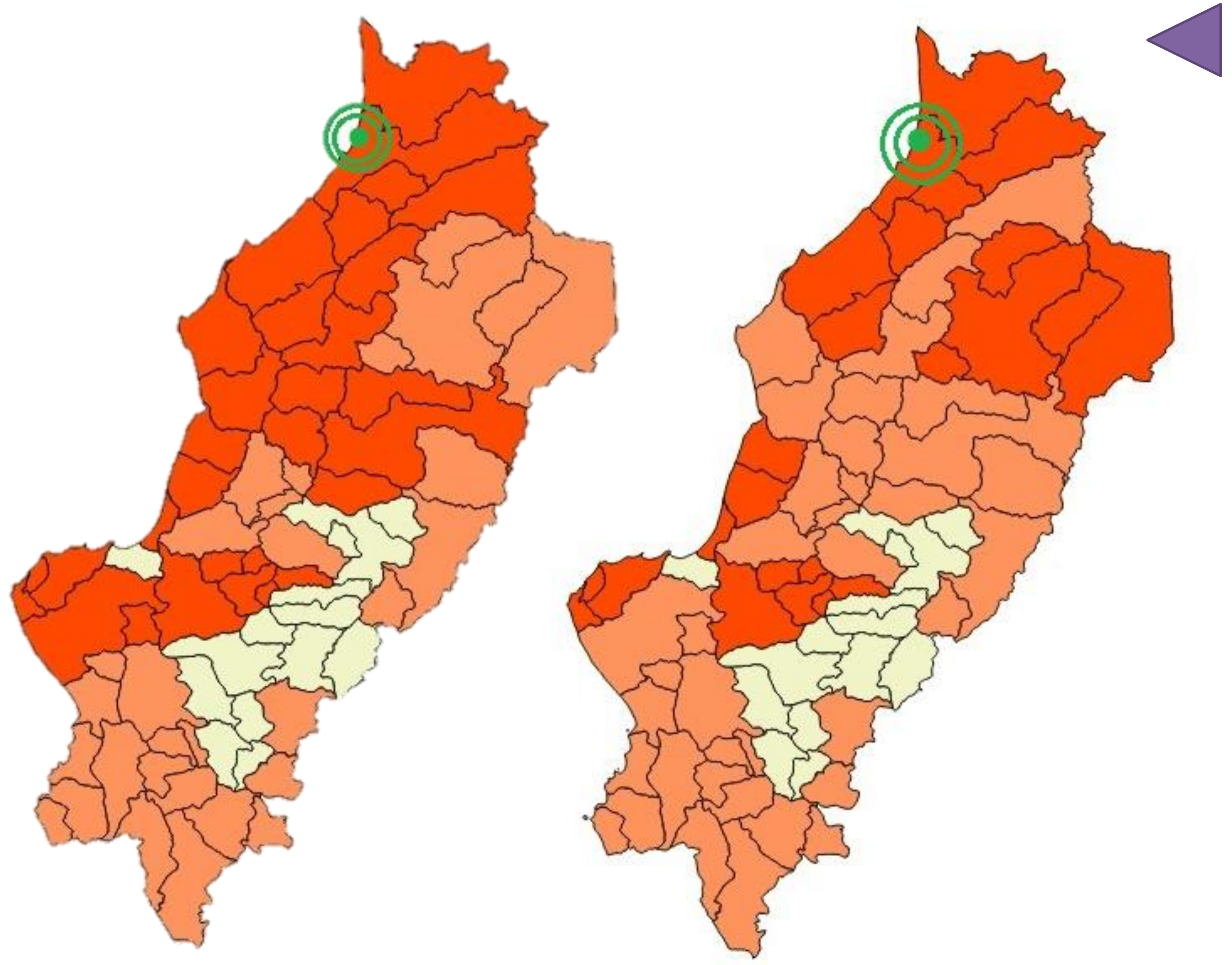
weij: Amount of users who have presented an event on B, having A as Home Tower.
k: amount of users.
d: distance between towers (in km)

DATASET

We used 11M aggregated records provided by *Telefónica*, a telecommunications service provider. Records were produced from April 15th to 18th .



RESULTS



Levels of damage on Manabí according to government (left) and our KNN model (right). A darker red represents a higher level of damage.

	Accuracy	Recall	F1 Score
Linear SVC	0.70	0.56	0.63
K-Nearest Neighbors	**0.75**	**0.67**	**0.71**
Logistic Regression	0.70	0.33	0.43

Classification models based on spatial, temporal and networks features. KNN presented the best evaluation metrics.

CONCLUSIONS

- Our method can be used as a complement to identify highly damaged zones within the first hours after a natural disaster.
- Our contribution includes the creation of the *Visitors Diversity Index* metric, which allows us to difference between highly and partially damaged zones.
- Mobile activity shows to be a rich source of metadata about people's activities, specially after disruptive events such as natural disasters.