

שימוש בטוויטר להתראה הקרובה לזמן אמת ולניתוח נזק של אסונות טבע בישראל וסביבתה

מוטי זוהר, אוניברסיטת חיפה
אביגדור גל, הטכניון
עוז שורר, אוניברסיטת חיפה
בר גנוסר, הטכניון
אפרת מורין, האוניברסיטה העברית
לאה ויטנברג, אוניברסיטת חיפה
רן נוף, המכון הגיאולוגי

מרכז הידע והמחקר הלאומי בתחום ההיערכות למצבי חירום

סיכום שנה א'

מבוא

- למעלה מ-500 מיליון משתמשים (subscribers) ברחבי העולם¹
- בישראל למעלה ממיליון עם גידול של כ-100000 לשנה בשנים האחרונות²
- החל מ-2015: ציוץ יכול להכיל טקסט בן (עד) 280 תווים
- לכל ציוץ נוספים נתונים נלווים (metadata) הכוללים בין היתר מאפיינים של הציוץ, פרופיל המשתמש, נתוני עוקבים, רטוויטים (retweets), האש-טאג'ס (#) ולעיתים גם נתוני מיקום (מדויק או משוער)
- רשת טוויטר מאפשרת גישה לחלק מהציוצים (עד שבועיים מרגע שליחתם) באמצעות שירותי רשת ו-API

¹ <https://www.statista.com/statistics/282087/number-of-monthlyactivetwitter-users/>

² <https://www.statista.com/statistics/558324/number-oftwitter-users-in-israel>

מטרות המחקר

- פיתוח שיטות וכלים לחילוץ המיקום (geolocation) של ציורים המתארים אסונות טבע ובחינת פריסתם הגיאוגרפית
- פיתוח מדד להערכת חומרה בהסתמך על מיקום וסנטימנט
- ניצול הרשת החברתית בניהול אירועים מתגלגלים במהלך ולאחר אסון טבע

שאלות למענה במסגרת שנת המחקר הראשונה

- האם טוויטר יכולה לשמש כפלטפורמה להתראה הקרובה לזמן אמת להתרחשות אסונות טבע בישראל?
- מהם דפוסי הפיזור וההבדלים בין ציוצים בעלי מיקום (geolocated) לאלו בעלי מיקום מחולץ (georeferenced)?
- מהם האזורים המושפעים מן האסון (לפי טוויטר) ועד כמה הם תואמים מידע המגיע מסנסורים/דיווחי אמת מהשטח?

מקרי בוחן

- בישראל וסביבתה:

- השריפה בחיפה נובמבר 2016

- שטפון ואסון נחל צפית אפריל 2018

- רצף רעידות האדמה יולי 2018

- להשוואה בארה"ב:

- Valleyfire ו-Buttefire קליפורניה ספטמבר 2015

- שטפנות במיסיסיפי דצמבר 2015

- רעידות האדמה בינואר (קליפורניה) ויולי (אלסקה) 2015

שיטת המחקר

- רכש ציוצים מ-Twitter בפרקי הזמן הרלוונטים ולפי קריטריונים לסינון (בשפות עברית, ערבית ואנגלית)
- איתור נתונים כמותיים מסנסורים ואיכותניים ממקורות נוספים
- עיבוד הנתונים ואחסונם בבסיס נתונים רלציוני ע"ג פלטפורמת ענן (Azure)
- שימוש בכלי למידת מכונה ועיבוד שפה טבעית לסינון ציוצים לא רלוונטים
- הערכת אמינות, חילוץ המיקום גיאוגרפי, סטטיסטיקה וניתוח סיכומי
- השוואה לנתוני סנסורים/מידע נוסף וניתוח מבוסס זמן ומרחב

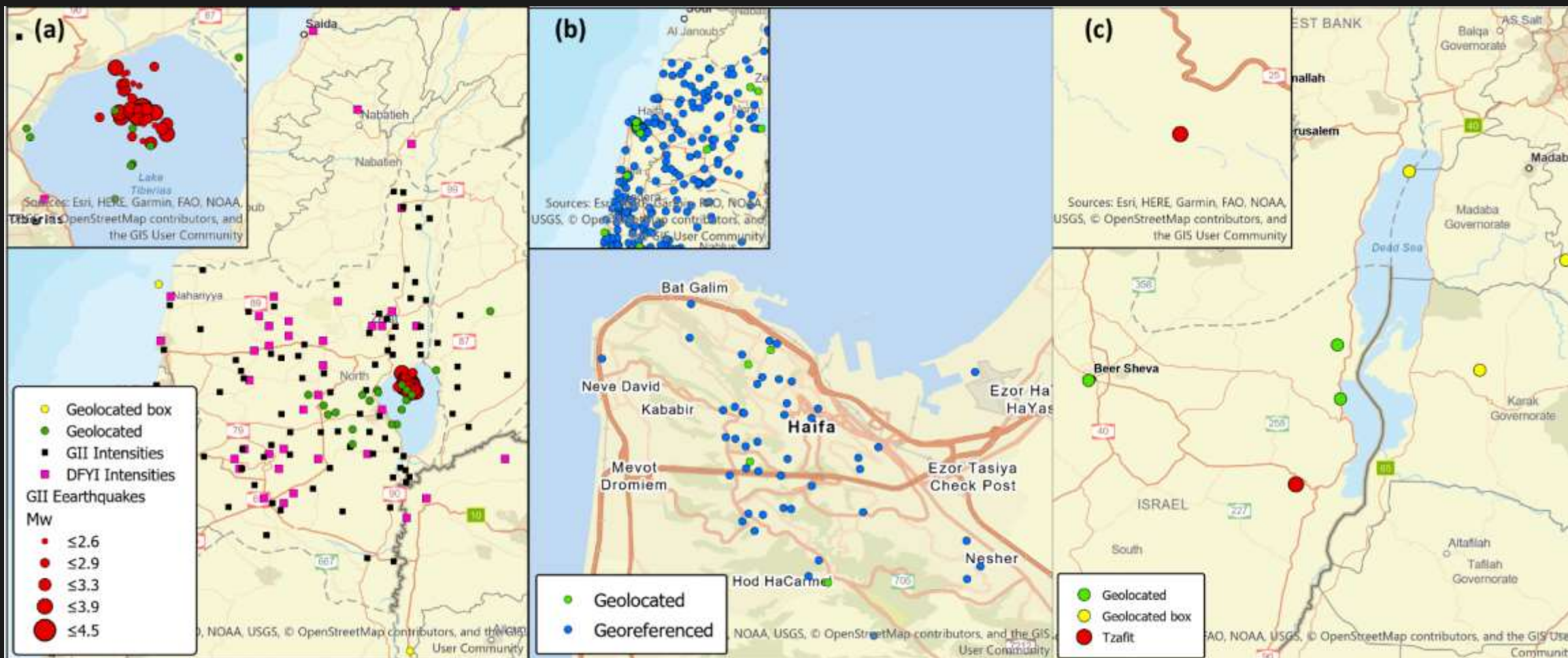
נתונים גולמיים - ציוצים

Event		Period	Total tweets	Geolocated Tweets
IL	Sequence of earthquakes, July 2018	2018-07-03 00:00:00 – 2018-07-29 23:58:57	3255699	46307
	The fire in Haifa, November 2016	2016-11-24 12:40:00 – 2016-11-27 23:58:59	2388724	6249
	The flood at Zafit river, April 2018	2018-04-23 00:00:00 – 2018-07-08 15:59:59	884053	864
USA	The earthquake in January 2015	2015-01-28 20:00:00 – 2015-01-29 21:59:59	1650436	1186483
	The earthquake in July 2015	2015-07-29 01:00:00 – 2015-07-30 03:09:59	2208124	188417
	The fires in September 2015	2015-09-12 01:00:00 – 2015-09-18 23:59:59	12591787	1140314
	Floods in December 2015	2015-12-23 00:00:00 – 2015-12-29 23:59:59	12064353	814914

מימוש תהליכי למידת מכונה

- סינון ראשוני של ציוצים בהתאם לאוסף חוקים הקשורים במועד הציוץ, המיקום ממנו נשלח והופעת מילות מפתח.
- 'ניקוי' ציוצים (אותיות גדולות קטנות, מילות יחס, תווים מיוחדים ואימוג'ים, מספרים, רווחים וכיו"ב)
- עיבוד הציוצים הנותרים בכלים מעולם עיבוד השפה הטבעית.
- למידת ה-features המשמעותיים אשר מגדירים רלוונטיות, למשל מילים דומיננטיות אשר מהוות אינדיקציה חזקה לכך שהציוץ בעל ערך: **'רעידת אדמה' - < רונאלדו**
- ביצוע prediction על ציוצים חדשים וסיווגם בהתאם לרלוונטיות.

מקרי בוחן - ישראל

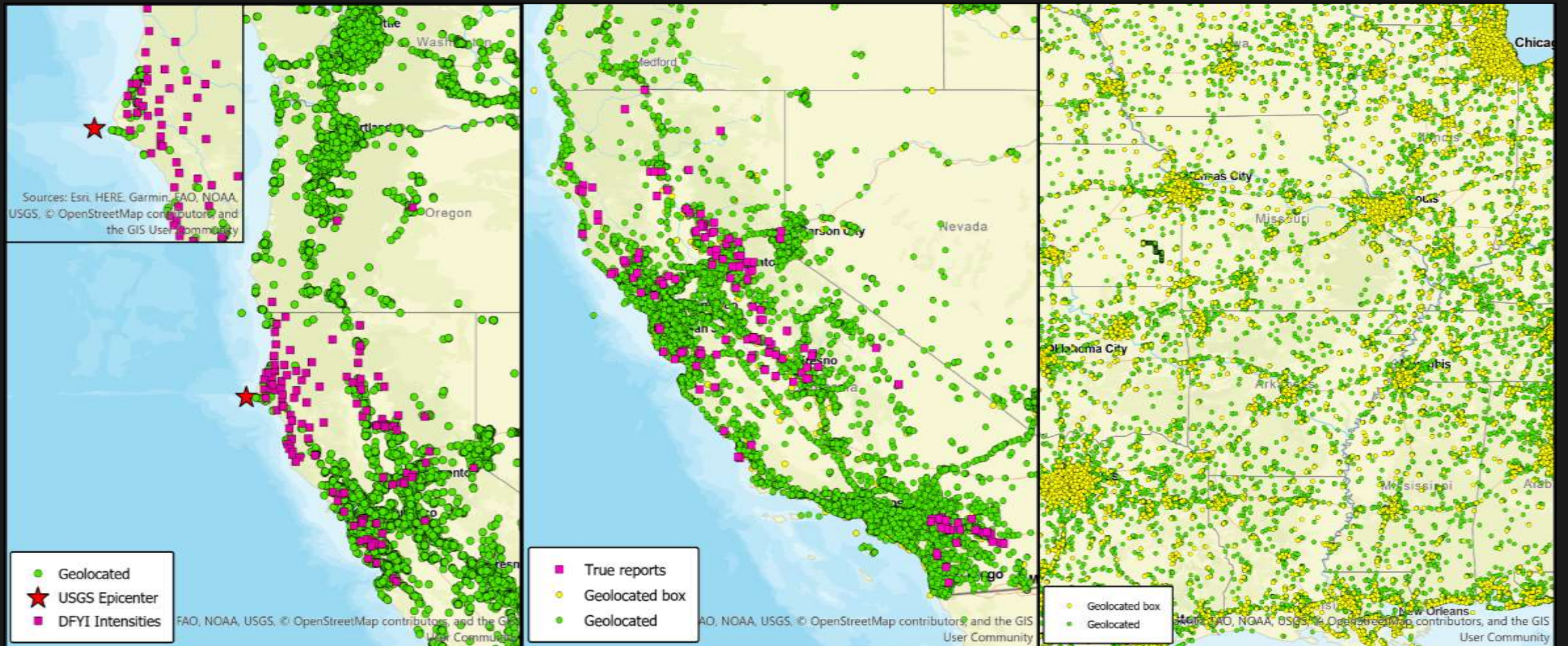


רעידות אדמה גליל 2018

שריפה חיפה 2016

שטפון צפית 2018

מקרי בוחן – ארה"ב



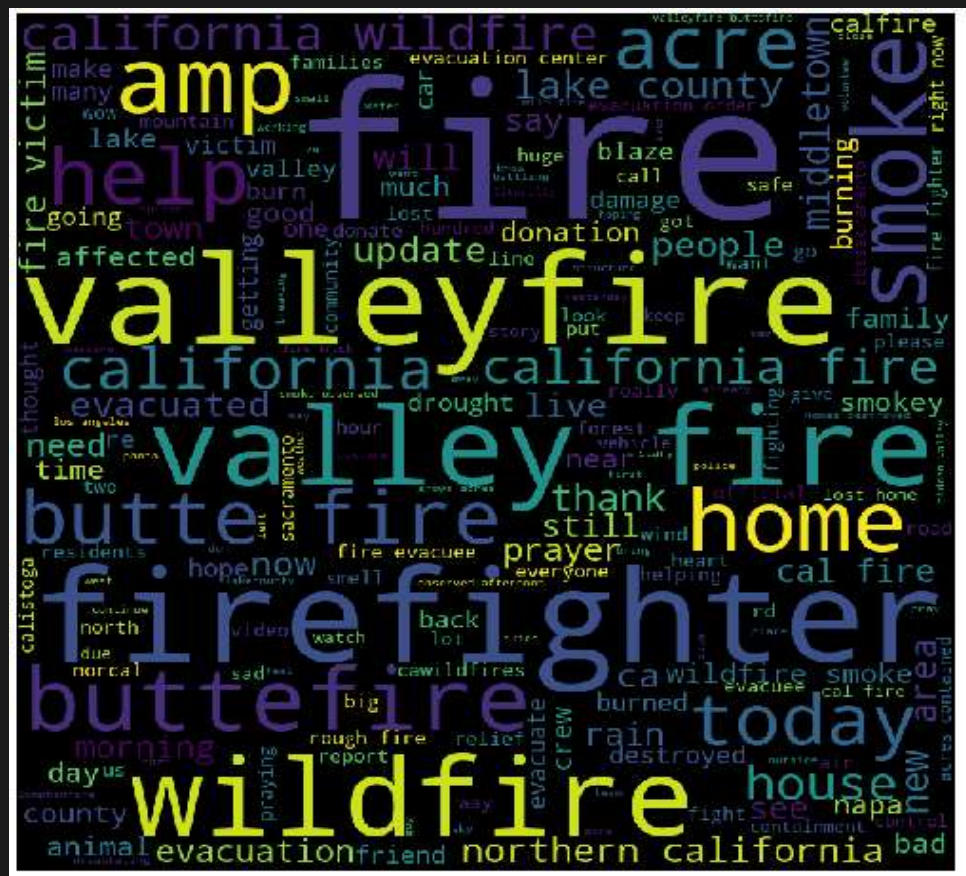
רעידת אדמה קליפורניה 2015

שריפות קליפורניה 2015

שטפונות מיסיסיפי 2015

ממצאים ראשוניים

ניתוח סנטימנט – השריפה בקליפורניה ספט' 2015



- אזכורים של מוקדי התרחשות מרכזיים
- יצירת ז'רגון של מושגים הנוגעים לאירוע
- אפיון שינויים בסנטימנט החברתי כפי שמשתקף מהציוצים על גבי ציר הזמן

ניתוח ז'רגון – השריפה בקליפורניה ספט' 2015

FR (Valley) (12.09.2015 13:00:00)

1 hour before

4 hours after

135

1166

Token

Freq.

Token

Freq.

fire

65

fire

589

fired

14

fired

124

portland

12

amp

81

morning

9

like

63

amp

7

get

54

california

7

portland

48

rp

6

game

40

pdx

6

new

38

w

6

day

37

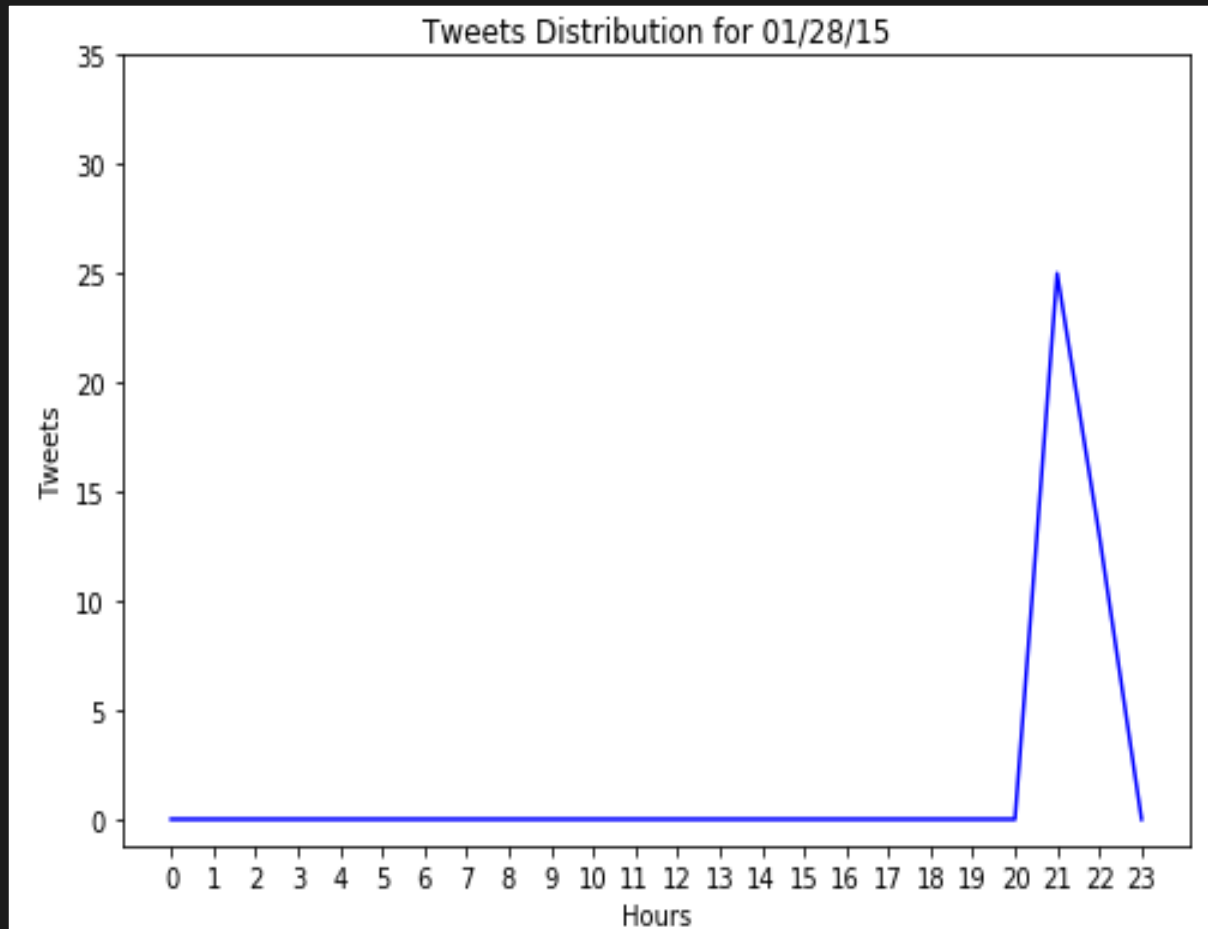
get

6

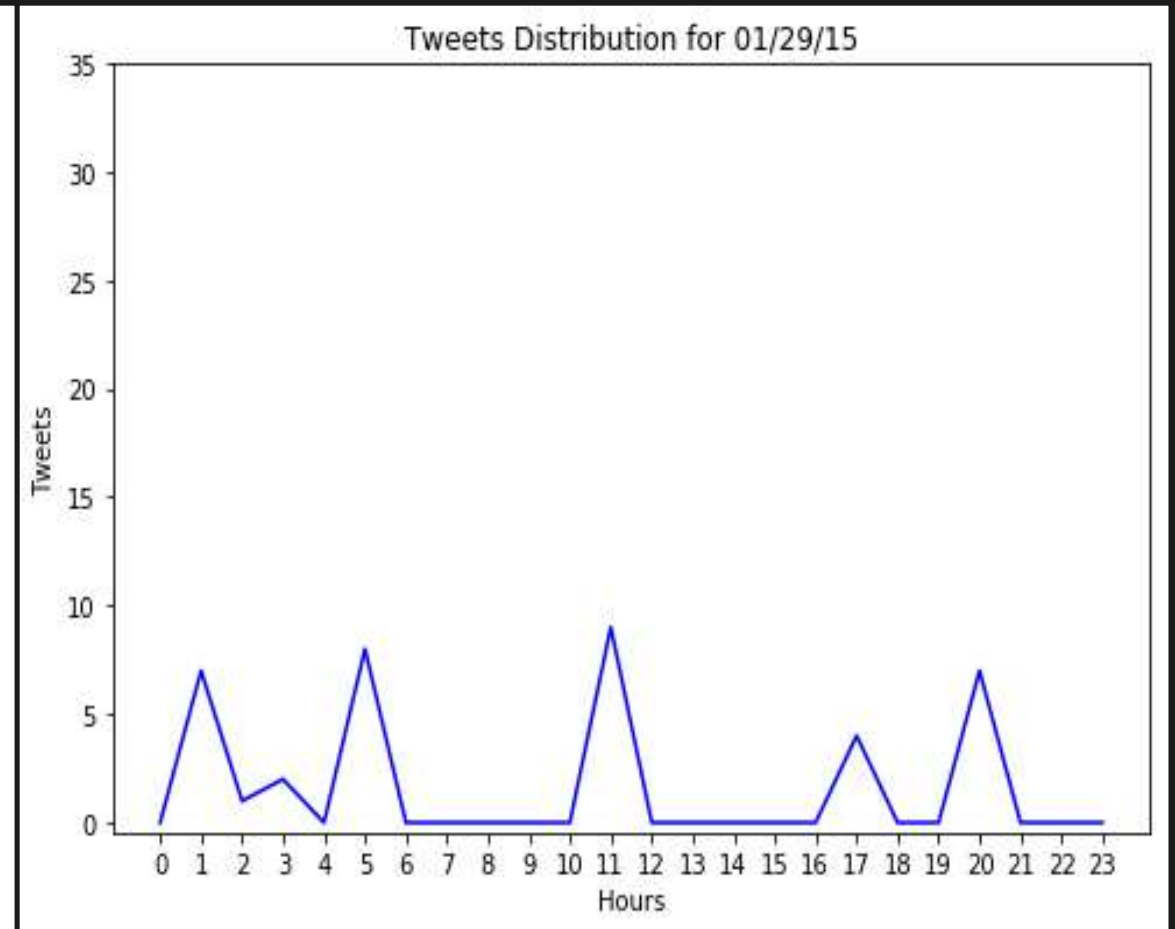
today

36

התפלגות עיתית - רעידת אדמה קליפורניה 2015



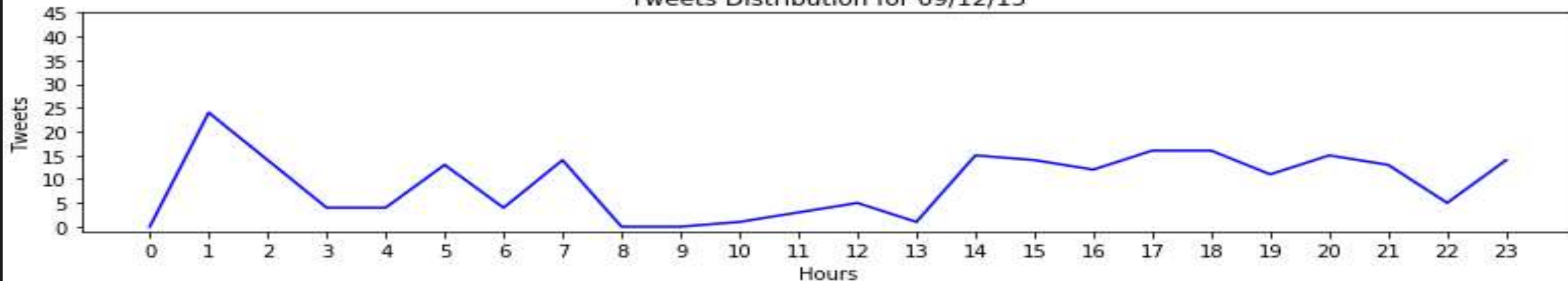
28/01/2015



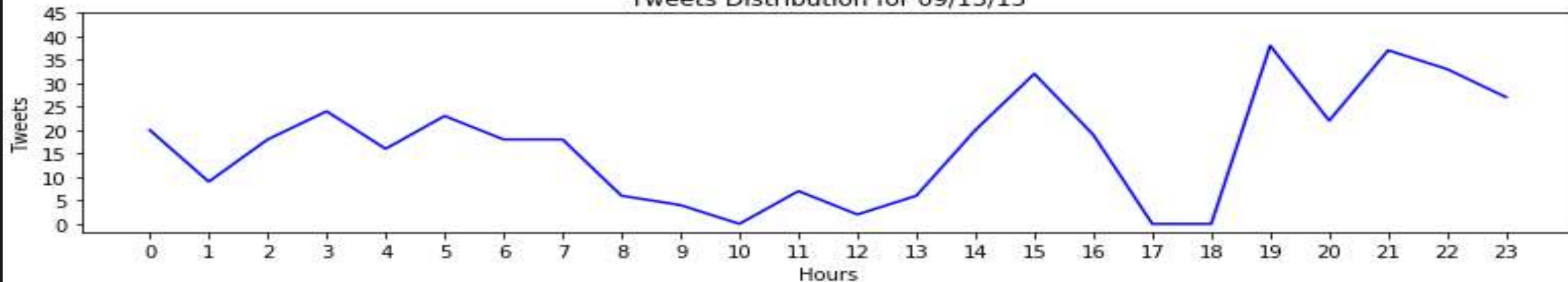
29/01/2015

התפלגות עיתית - שריפות קליפורניה 2015

Tweets Distribution for 09/12/15



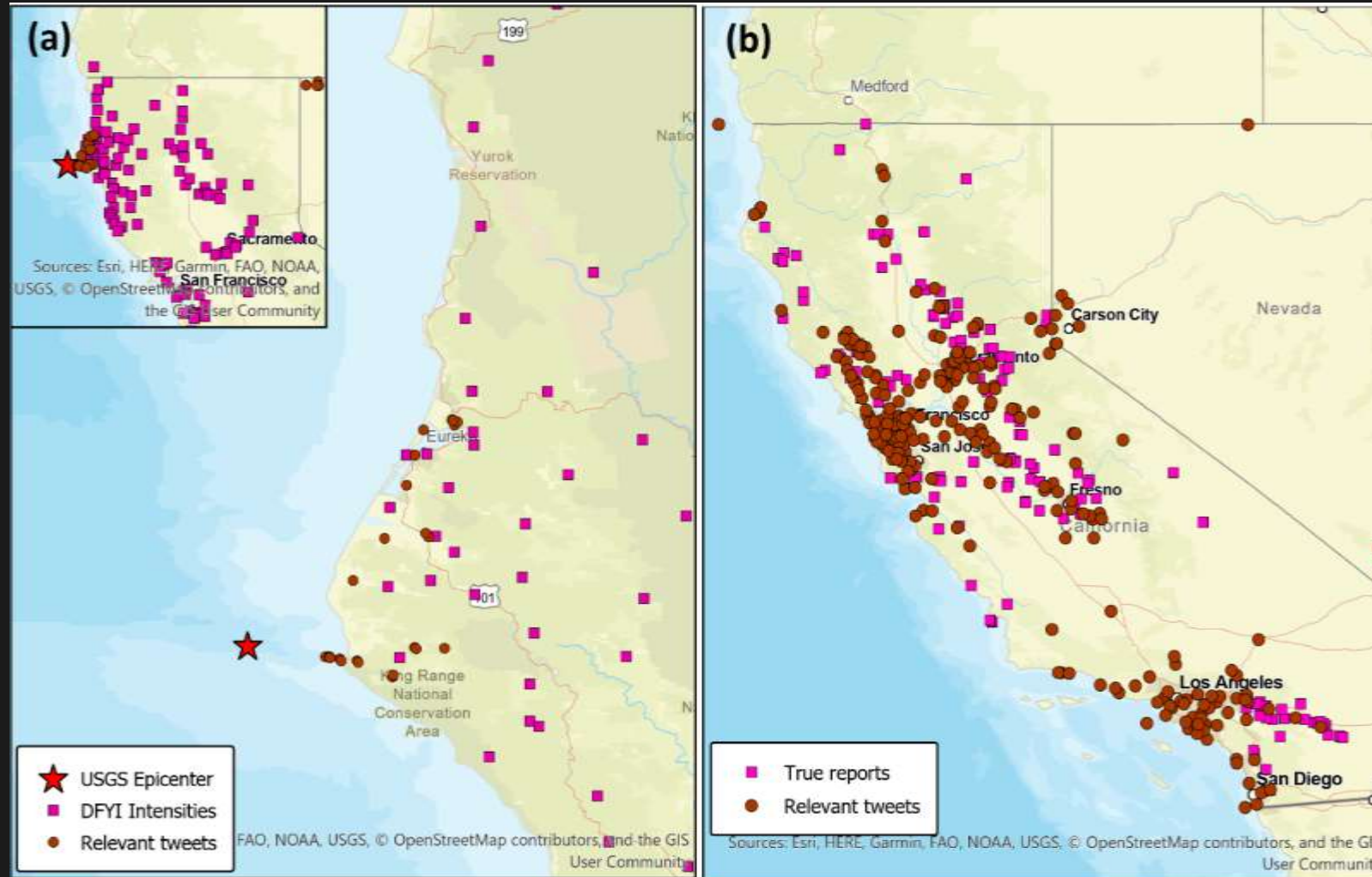
Tweets Distribution for 09/13/15



12-13/09/2015

פריסה מרחבית – ציורים לאחר עיבוד שפה*

רעידת
האדמה ינואר
2015



שריפות
קליפורניה
ספטמבר 2015
CAL FIRE | Office
of the
State Fire Marshal

* ללא חילוץ מיקום (Georeferencing)

סיכום וצעדים להמשך

- הרחבת השימוש בלמידת מכונה לכלים נוספים של עיבוד שפה טבעית
- חילוץ מיקום (Georeferencing) באמצעות מטא-דאטה ומערכי הגזטיר
OSM ו-Geonames
- השוואה לסנסורים ומקורות אנושיים (הידרוגרפים, יומני דיווח, קריאות אזרחים)
- איתור 'נקודות חמות' (hot spots), אזורים מושפעים וקשרים מרחביים בין הציוצים
- ניתוח סנטימנט להערכה איכותנית של חומרת האירוע
- מעקב אחר אירועים מתגלגלים בעת אסון ממושך

תודות

- מרכז הידע והמחקר הלאומי בתחום ההיערכות למצבי חירום
- משרד המדע והטכנולוגיה
- וורוניק עמירב, המכון הגיאופיסי לישראל
- *CAL FIRE* | Office of the State Fire Marshal
- יאיר רינת, השרות ההידרולוגי לישראל