
 行政法人 國家災害防救科技中心  
National Science and Technology Center  
for Disaster Reduction

# 防災科技應用

## 社群大數據於災情動態分析之應用

國家災害防救科技中心  
災防資訊組

# 大綱



- 社群網路發展與簡介
  - 社群網路的應用概況
  - 社群網路災情蒐整所面臨之問題
- 社群資料分析方法
  - 文本分析
  - 關鍵字詞分析
- 社群網路應用於災害防救
  - 社群數據災情蒐整應變流程
  - 社群網路應用案例

2

## 社群網路發展及簡介 (1/2)



### 社群網路服務(亦稱社群網站):

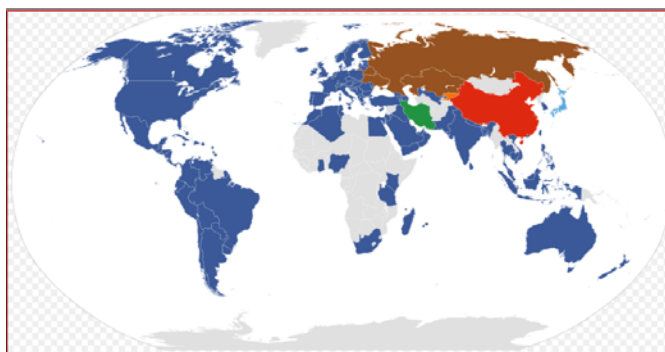
- ◆ 主要作用是為一群擁有相同興趣與活動的人建立線上社群，並提供多種使用者互動的方式
  - ◆ 如: 聊天、寄信、影音及檔案分享、部落格、新聞群組等
- ◆ 台灣地區社群網站熱門使用網站
  - ◆ 社群平台:
    - ◆ 臉書 (Facebook)、Google+、痞客邦 (Pixnet)
  - ◆ 討論區:
    - ◆ Mobile01、批踢踢 (PTT)、Yahoo奇摩知識+



## 社群網路發展及簡介 (2/2)

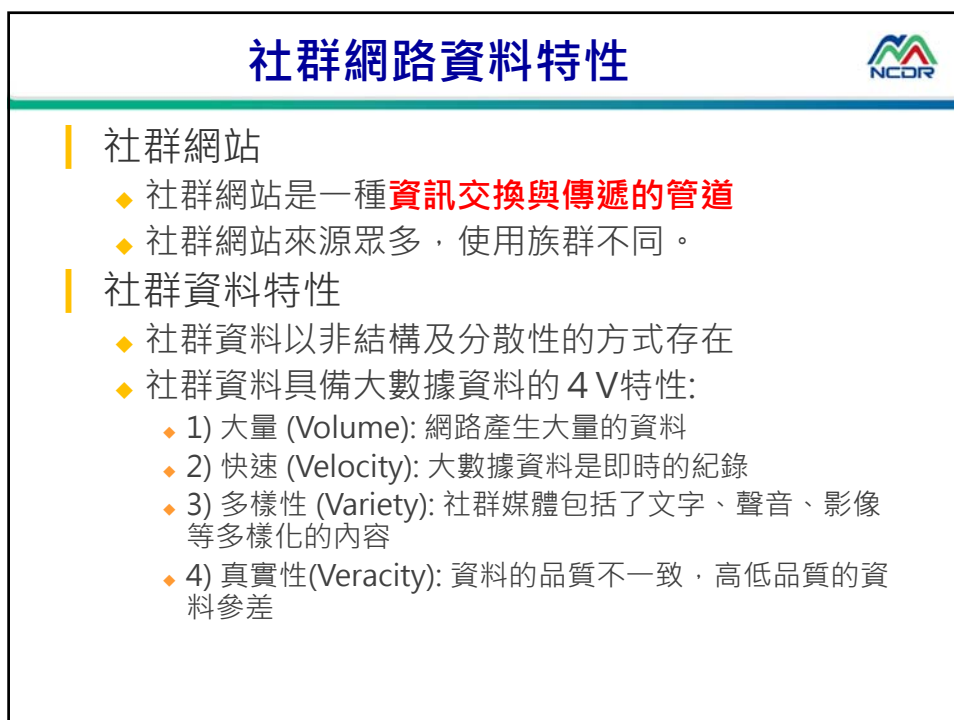
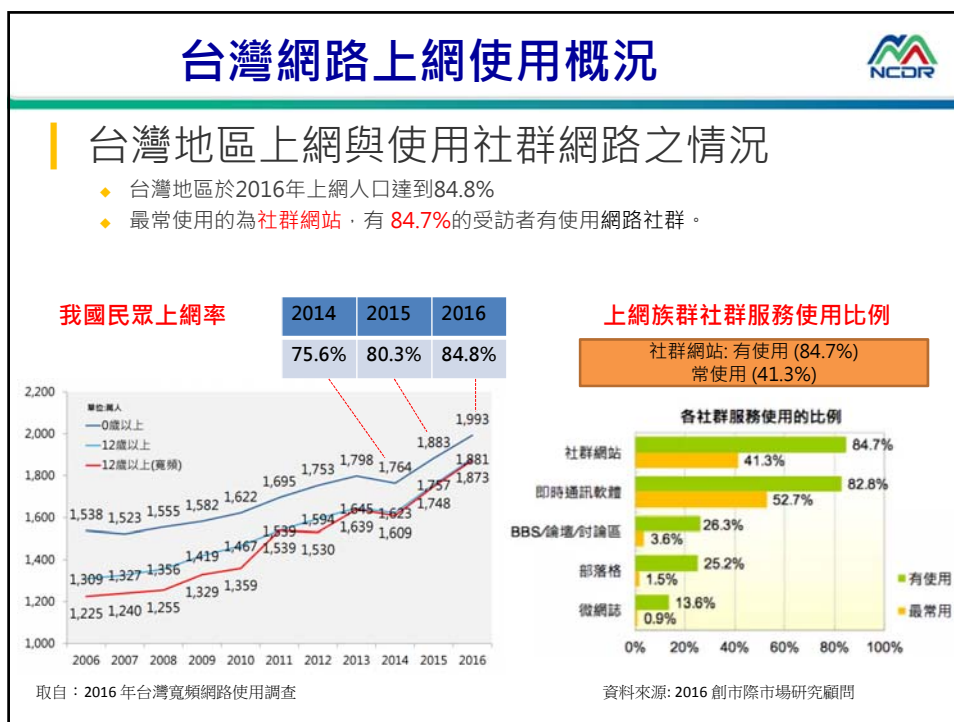


### 世界各國社群網路使用情況



各國最流行的社群網站 ■ Facebook ■ Twitter ■ VKontakte ■ QQ空間 ■ Odnoklassniki ■ Facenama ■ 沒有資料

<https://zh.wikipedia.org/wiki>



## 社群網路資料特性



### 資料的應用:

- ◆ 資料透過分析及重整等方式，創造新的應用價值，同時協助解決與優化政府施政效率與民生各類議題。

### 資料加值與階層性

- ◆ 資料
- ◆ 資訊
- ◆ 知識
- ◆ 智慧



## 海地地震網路志工社區協作

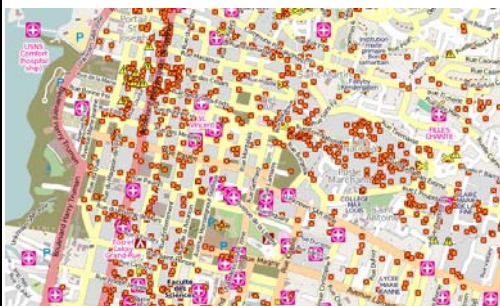


### 2010年海地大地震

- ◆ 當地地圖已經20年未進行更新，許多地點的資訊，付之闕如
- ◆ OSM 的人道救援小組 (HOT) · 花了七天時間進行田野勘查與地圖繪製。
- ◆ 以一週完成的地圖，成為救災的重要依據。

Text your needs to

**4636**  
in Haiti



## 海燕颱風 OpenStreetMap 網路協作




---

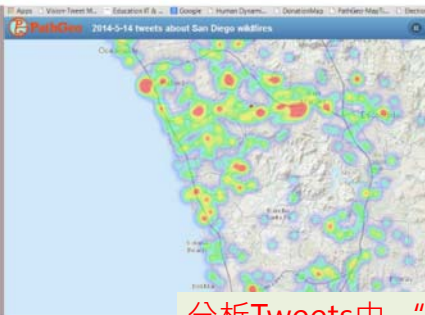
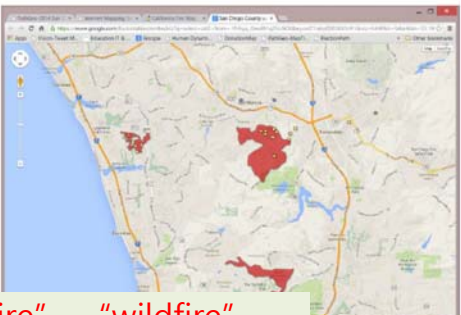
### 菲律賓海燕颱風

- ◆ 國際紅十字會人道救援組織第一時間就趕往災區協助救災及復原工作，但遭遇到缺法當地地圖的困境。
- ◆ 在Humanitarian OpenStreetMap Team發起，同時由410名國際OSM志工的協助下，迅速將災區基本地圖建立起來，提供現場救援組織人員自由運用。


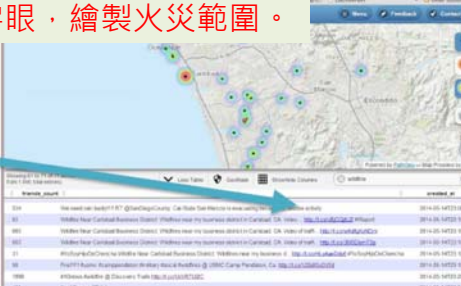



## 加州森林火災 Twitter



分析Tweets中 “fire” , “wildfire” , “evacuation等字眼，繪製火災範圍。

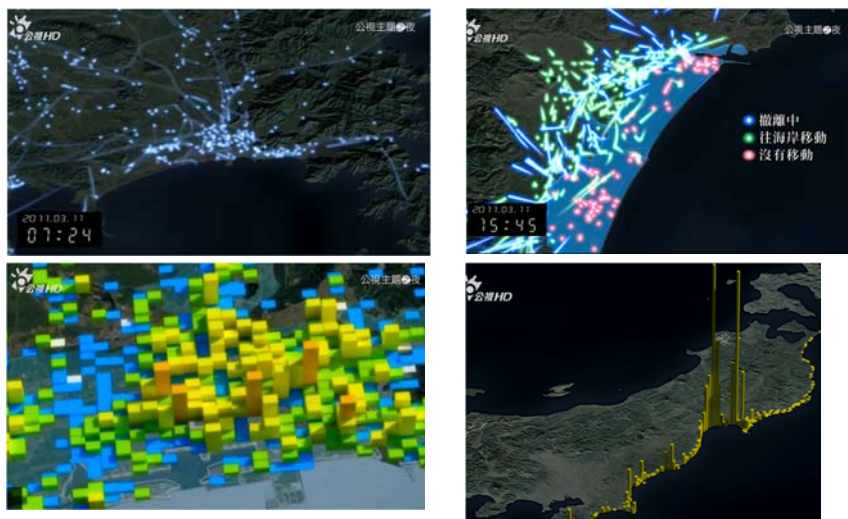





## 311日本地震汽車GPS軌跡



分析汽車及行動基地台之GPS資料,了解災害當下民眾的移動軌跡。



## 社群網路與災害情資



災害發生期間  
如何快速掌握第一手災害現場狀況?



哪裡有災害發生?  
發生何種災害?



民眾於災害期間災情回報利用119或1999專線方式通報

12



## 社群網路災情蒐整所面臨之問題



**網站來源蒐集之困難**

- ◆ 社群網站來源眾多
  - ◆ 資料彙整不易
  - ◆ 網站資料量不同
  - ◆ 使用者群族不同
  - ◆ 各網站討論主題不同

**社群網路熱點**

- ◆ 隨時間動態產生新來源
- ◆ 不容易發現
- ◆ 具有群聚效應
- ◆ 災情資訊量多
- ◆ 災害主題性明確

社群網站



臉書上關於蘇迪勒颱風災情之活動



圖片來源：Facebook、Plymouth Devon14

## 社群網路災情蒐整所面臨之問題



### 社群資料彙整之議題

- ◆ 災害訊息量多，資料處理耗時過多
- ◆ 災害類型繁多，描述品質不一
- ◆ 緊急與重要性不一
- ◆ 地理位置資訊處理困難



路樹、招牌倒塌，影響道路



河道水位快接近橋梁



道路、街道淹水

圖片來源：Facebook 15



行政法人 國家災害防救科技中心  
National Science and Technology Center  
for Disaster Reduction

## 社群資料分析方法



## 文本分析 (Text Mining)



- | Text Mining 是應用 Data Mining 來進行文件內文字資料的分析，並期望取得文字之間的關聯性。
  - ◆ Text Mining 針對文字資料進行分析，其資料結構多屬半結構化或非結構資料。
  - ◆ 一般處理流程
    - ◆ 進行資料前處理 ( Pre-Processing )
      - ◆ 資料收集(Collection)、清理(cleaning)、標記(tagging)
    - ◆ 運用統計方法或演算法
      - ◆ 例如：Term Frequency - Inverse Document Frequency · 簡稱 TF-IDF

17

## 文本分析 (Text Mining)



- | 文本資料收集來源
  - ◆ 社群網路資料
    - ◆ 各類型社群網站、討論區
  - ◆ 特定網站資料
    - ◆ 氣象局、新聞稿
  - ◆ 紙本資料
    - ◆ 訪談調查、滿意度調查
- | 社群網路資訊收集技術
  - ◆ 網路爬蟲 (Web Clawer)
  - ◆ 網路爬蟲服務
    - ◆ Kimono
      - ◆ Kimono是個雲端爬蟲服務，並能以API的形式提供抓取的資料



18

## 文本分析 - 商業上之應用案例



### 威名百貨WAL-MART分析案例

- ◆ 銷售商品之間的關係
  - ◆ 購買尿布同時購買最多商品的是啤酒。
  - ◆ 市場調查: 美國的太太常叮嚀丈夫在下班前幫嬰兒買尿布，而40%的先生在買完尿布後，又會隨手拎兩罐啤酒
  - ◆ 商場策略: 各個賣店把啤酒和尿布擺在一起，結果尿布和啤酒的銷售量增加三成
- ◆ 溫度與商品銷售的關係:
  - ◆ 氣溫在24度到27度時，鰻魚、冰品和防曬乳賣得好。
  - ◆ 溫度在22度到25度時，涼飲、冰咖啡和殺蟲劑就不可少。
  - ◆ 溫度在17度到20度時，布丁、沙拉和優格則很受歡迎。

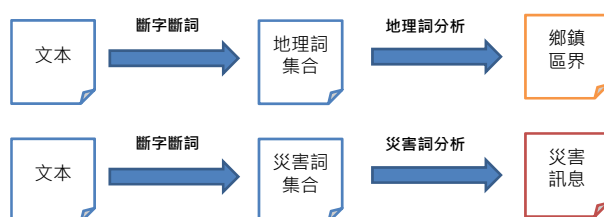
19

## 文本分析-關鍵字詞分析



### 中文斷字斷詞關鍵字詞分析技術

- ◆ 使用中研院「中文斷字系統」及結巴 (jieba) 中文斷字斷詞模組



綠色隧道淹水了！請小心，先不要走這裡



原文: 綠色隧道淹水了，請小心，先不要走這裡

原文: [綠色隧道][淹水][了]·[請][小心]·[先][不要][走][這裡]

20

## 社群網路關鍵字詞觀測設定



- **社群網路關鍵字詞觀測**
  - 預先設定相關之觀測關鍵字
  - 依據不同觀測關鍵字搜尋相關文章
  - 同時也能觀測社群網路即時的聲量

**查詢條件**

期間：  
 最近15天  
 最近30天  
 2015/08  
 08/08 00:00:00  
 08/09 00:00:00

來源：  
 討論區  
 社群網站  
 facebook  
 facebook-event  
 bbs  
 新聞

**關鍵字設定主題**

主題：  
 空業品質  
 與泰華有名詞  
 武漢  
 數基災害  
 八卦  
 深坑山體滑坡  
 test

關鍵字篩選：

**相關文章列表**

文章列表 查詢條件：2015/08/08 - 2015/08/09 / 論壇文章 / 討論區 - 社群網站 - facebook - facebook - 共192篇結果

排序：  
 日期  
 相關度  
 熱門度  
 好評  
 負評  
 點閱率  
 結果：

- 大仁科大竟又淹水了～過不去了～  
 她～暑假就好(在忙)  
 共 3 則回覆 · 本新聞 2015-08-08 23:39:44 · 共 1 次點閱 · 0.0 · 0.0  
[facebook事件](#) · [新增對談簿](#) · [一起看歷史事件](#) · [新增](#) · [庫存內容](#) · [加入追蹤](#)
- 台南有那淹水或封路的嗎？等等要從大港轉車到海佃上班  
 那裡實際上全是招牌倒下的牌子  
 共 3 則回覆 · &Hu Xing 2015-08-08 23:35:59 · 共 1 次點閱 · 0.0 · 33.2  
[facebook事件](#) · [新增對談簿](#) · [一起看歷史事件](#) · [新增](#) · [庫存內容](#) · [加入追蹤](#)
- 台南有那淹水或封路的嗎？等等要從大港轉車到海佃上班  
 永和路二段  
 共 3 則回覆 · &Hu Xing 2015-08-08 23:34:59 · 共 1 次點閱 · 0.0 · 0.0  
[facebook事件](#) · [新增對談簿](#) · [一起看歷史事件](#) · [新增](#) · [庫存內容](#) · [加入追蹤](#)



行政法人 **國家災害防救科技中心**  
 National Science and Technology Center  
 for Disaster Reduction

## 社群網路應用於災害防救

4-11

11



## 社群數據攀爬即時資料蒐整倉儲



### 社群資料量統計-蒐整社群網站來源頻道 (213個)

- ◆ 單月資料量約500-700萬筆
- ◆ 平均每月資料儲存量約300G
- ◆ 資料即時性約20分內。



批踢踢(PTT)、臉書(FB)、Youtube  
社群網站



巴哈姆特、LineQ、Mobile01、FashionGuide、  
卡提諾王國、伊莉討論區、PCDVD、Plurk搜尋、  
校園聊天-Dcard  
討論區



BLOG 聯合新聞網、中時電子報\_即時新聞、蘋果日報、自  
由時報、蕃新聞、MSN新聞、Ettoday、壹電視新  
聞台、工商時報  
部落格

25

## 社群即時災情數據分析-關鍵字詞分析



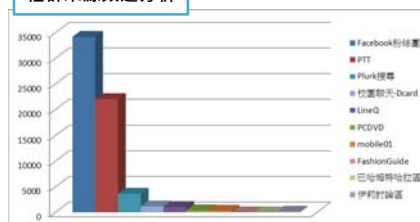
### 資訊擷取與過濾技術 (以關鍵字詞組為基礎)

- ◆ 使用邏輯運算符號
  - ◆ AND (交集)、OR(聯集)、!(排除)
- ◆ 使用萬用字元符號
  - ◆ ? (0 或 1個字元)、\* (0~5個字元)
- ◆ 淹水災害: (淹?水|積水)
- ◆ 風力災害: (路樹|招牌)&(掉落|倒塌)

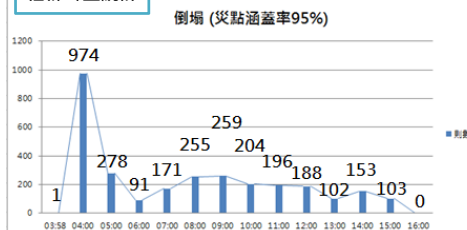
### 災情趨勢分析

- ◆ 依據災害類型建置災害之關鍵字詞組
- ◆ 以社群聲量作為計算基礎
  - ◆ 統計某一時間區間內的文章數量

社群來源頻道分析



社群聲量統計



26

## 社群數據災情蒐整應變流程





社群災中應變流程

災情擷取

災情彙整


發佈災情至  
決策輔助系統

### 社群技術導入與整合

- ◆ 社群應變作業平台
  - ◆ 即時觀測**社群資料**功能
  - ◆ 擷取**重點災情**
- ◆ 多人協作地圖技術(Collector)
  - ◆ 即時上傳災情
- ◆ 發佈災情至**決策輔助系統**

27

## 社群數據災情蒐整統計 (105年度)



### 颶風事件

颶風事件	社群災情攀爬	重點災情圖	發佈至決策輔助系統
尼伯特	45,649	94	20
莫蘭蒂及馬勒卡	98,868	182	120
梅姬	128,161	144	84

### 地震事件

地震事件	社群災情攀爬	重點災情圖	發佈至Portal
高雄美濃地震	62,914	25	14

### 豪雨事件

豪雨事件	社群災情攀爬	重點災情圖	發佈至Portal
0516	2,344	35	18

28

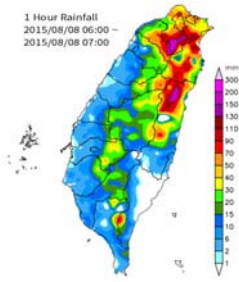


行政法人 國家災害防救科技中心  
National Science and Technology Center  
for Disaster Reduction

## 社群網路災情綜整於災害防救應用案例

### 蘇迪勒颱風

發生時間:  
陸上警報於8月6日20:30~ 8月9日08:30時解除。  
並於8月7日大約17:00時暴風半徑逐漸接處陸地  
，實際侵台日期為8月8日。



## 社群網路資料來源範圍

定期資料攀爬 - OpView 平台 (收集)

- 數據統計時間: 8/6 ~ 8/10

Facebook 「蘇迪勒吹殘 一起為受災樹木發聲」

- 數據統計時間: 8/7 ~ 8/17

「颱風災情資料表」

- 數據統計時間: 8/7 ~ 8/10

定期攀爬 - OpView 平台  
(資料來源包含社群網站、討論區、部落格)  
- 統計8/8颱風資料聲量最多的前十名網站



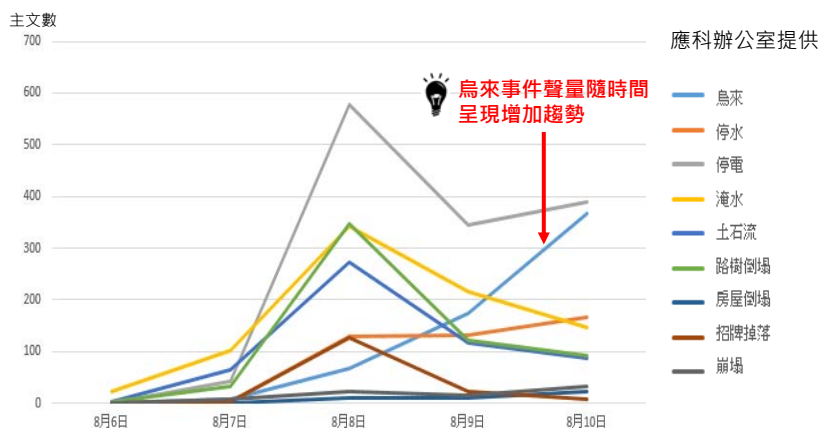
時間	內容	用戶	圖片
08-08 05:03	蘇迪勒風山崩電燈 斷了	李小姐	
08-08 05:02	台中市之斷電區域 第10區	Zyng	
08-08 05:01	彰化市中正路二段 117號	JJ	
08-08 05:00	蘇迪勒	Mighty	

## 社群網路各類型災害聲量分佈



### 依各種災害統計社群網路聲量分佈

- 災害聲量多集中在8/8當日
- 烏來事件於8/8當日災害聲量不多，但隨時間呈現聲量增加的趨勢



烏來事件反映出社群網路於災情彙整時，對於災害造成的通訊、網路中斷，或是偏鄉地區的災情收集上還有一定的困難度

## 社群網路災情綜整分析 - 內文深度分析



### Facebook活動「蘇迪勒吹殘 一起為受災樹木發聲」

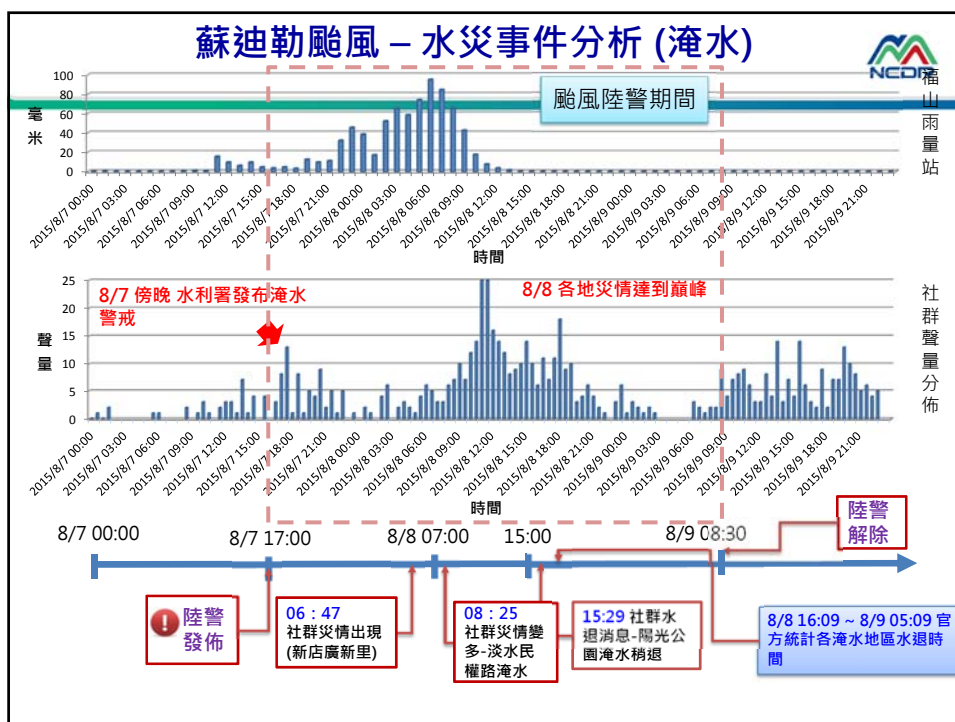
- ◆ 分析取樣 (710筆) · 經人工判釋分析 (資料相關度: 57.1%)
  - ◆ 風災相關資料(319)、淹水相關資料(61)、河/溪水暴漲相關資料(26)
  - ◆ 非災害相關資料(311)


#### ◆ 災情內文深度分析

- ◆ 地點精準度: 描述地點包含地標或路名 (精準度高)、描述地點包含縣市/鄉鎮區 (精準度低)
- ◆ 災害情資品質: 描述災害類型及災害程度 (品質高)、只描述災害類型 (品質低)

資料項目	風災資料品質分佈	水災資料品質分佈	河/溪水暴漲資料品質分佈
地點描述精準度高	256 (80.2%)	51 (83.6%)	25 (96%)
地點描述精準度低	63 (19.7%)	10 (16.3%)	1 (4%)
災害情資品質高	171 (53.6%)	38 (62.2%)	20 (76.9%)
災害情資品質低	148 (46.3%)	23 (37.7%)	6 (23%)

網民在描述地點約有8成以上會描述到地標或路名/地名/河名，災害情資描述約6成左右會描述災害類型 (如: 淹水) 及受災程度 (如: 半個輪胎高)

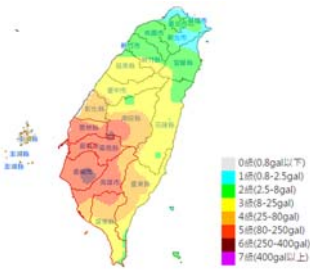



 行政法人 國家災害防救科技中心  
 National Science and Technology Center  
 for Disaster Reduction

## 社群網路應用案例


### 0206震災

發生時間:  
 105年2月6日03:57於高雄市美濃區  
 (北緯:22.92度、東經120.54度)  
 發生芮氏規模6.6地震，深度14.6公里。



0級(0.8g以下)  
 1級(0.8-2.5g)  
 2級(2.5-5g)  
 3級(5-25g)  
 4級(25-80g)  
 5級(80-250g)  
 6級(250-400g)  
 7級(400g以上)

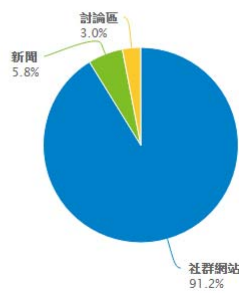
## 0206震災社群資料收集



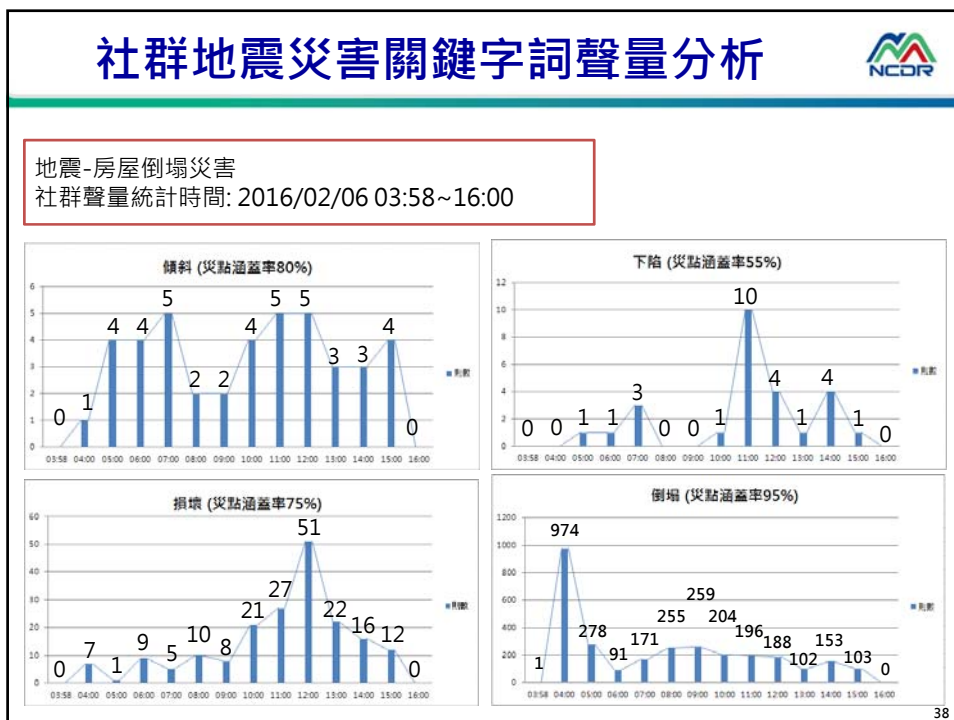
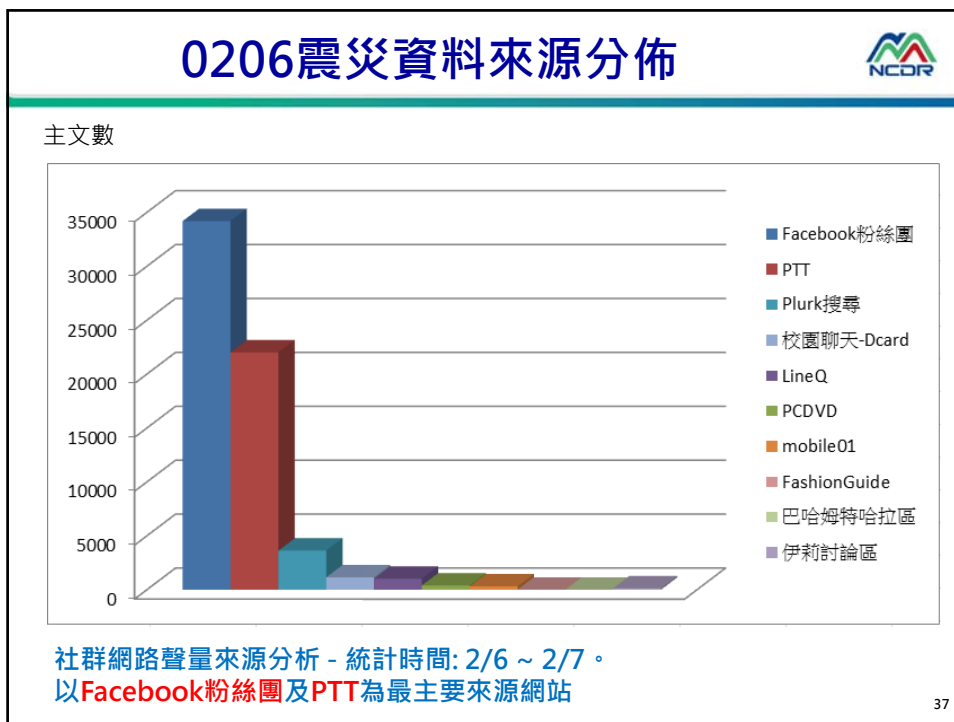
搜尋範圍: 社群網站、新聞、討論區  
 統計時間: 2016 2/6 ~ 2/7 · 總共62,914筆

來源聲量分布 時間區間: 2016/02/06~2016/02/07 主題: 南部地震  
 資料來源: 討論區、社群網站、facebook、facebook-event、bbs、新聞

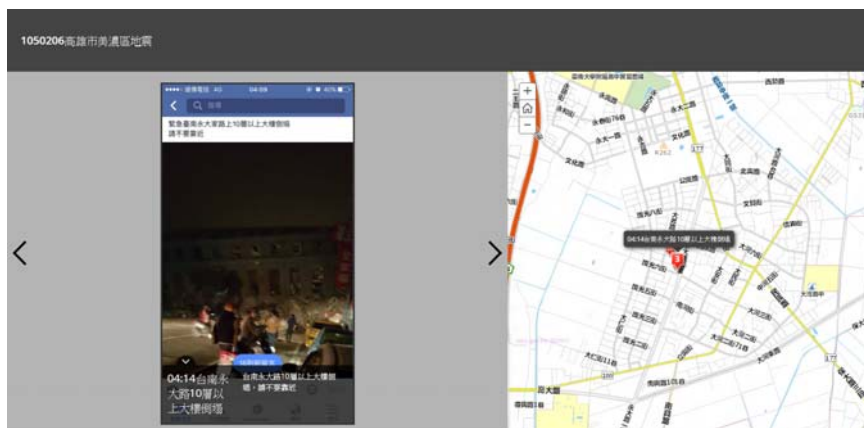
來源	聲量數	比例
社群網站	60,891	91.2%
新聞	3,859	5.8%
討論區	2,023	3.0%



36



## 0206震災災情發佈



2016/02/06 04:14 台南市永大路10層以上大樓倒塌  
(來源: PTT)

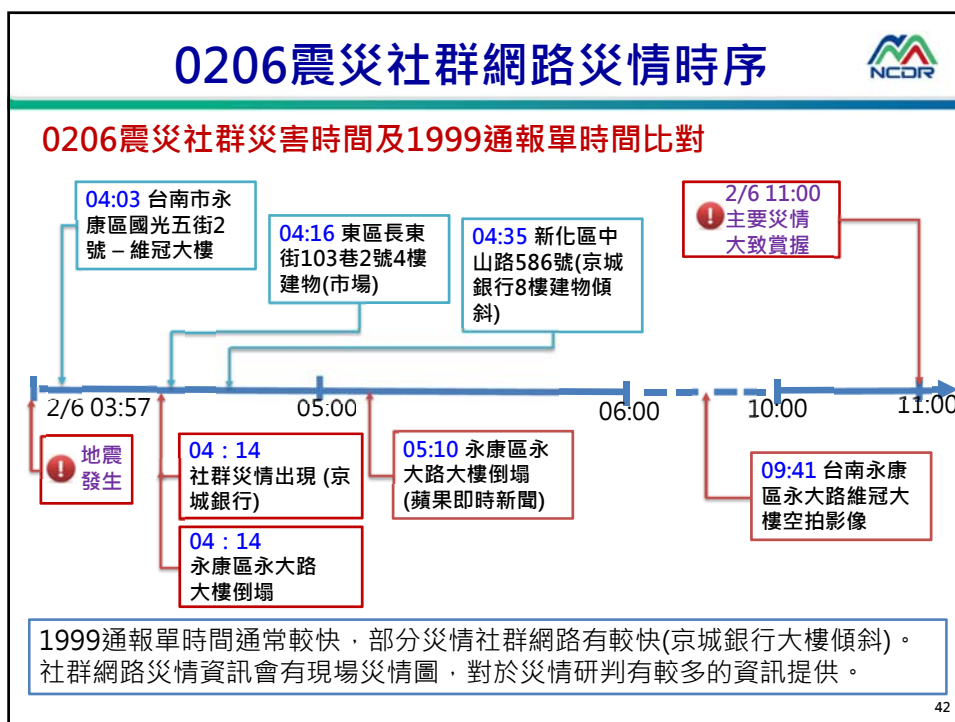
## 0206震災地理空間災情分佈



地理空間災情分佈 - 經人工處理繪製及綜整







感謝聆聽  
敬請指教