Social media crawler and the generation of point of interest (POI)/ region of interest (ROI)

(社群網路資料攀爬與興趣點/範圍之生成)

Chiao-Ling Kuo (郭巧玲)

Assistant Research Fellow

Center for GIS, Research Center for Humanities and Social Sciences (RCHSS), Academia Sinica

April 27th , 2018

地理資訊科學研究專題中心
Center for GIS, RCHSS, Academia Sinica

About me

- **Ph.D. degree, Dept. of Geomatics, NCKU.**
- **Visiting PhD student, Institute for Geoinformatics (ifgi), Universität Münster, Germany (Sandwich program).**
- **Postdoc, Center for GIS, RCHSS, Academia Sinica.**
- **Postdoc, Institute of Geography, Universität Heidelberg, Germany (PRAP program).**
- **Research topics**
  - Semantics, Ontology, Ontology integration
  - Volunteered geographic information (VGI), Crowdsourcing, Social media
  - WebGIS, Mobile GIS, OpenGIS, Data standard, Metadata
  - Big Geospatial Data, algorithm

MOST-DAAD Sandwich program

- **The cooperation between MOST and DAAD**
  - Ministry of Science and Technology (MOST)
  - The German Academic Exchange Service (DAAD)
- **Applicants: Ph.D. Candidates**
- **Duration: 6-18 months**
- **Number of places: 30 people/Year**
- **German Language course: 2 months**
  - Goethe Institute in Bonn
- **Where to go: Germany**

MOST-DAAD Taiwan-German Summer Institute

- **Ms. and Ph.D. students, 2 months, 10 people /Year**

MOST PRAP Program (千里馬博後)

- Postdoctoral Research Abroad Program.
- Applicants: Ph.D.
- Duration: 12 months or 24 months
- Number of places: ca. 80~100 people/Year
- Where to go: All of the world

Graduate Students Study Abroad Program (千里馬博士生)

- Ph.D. students or Ph.D. Candidates

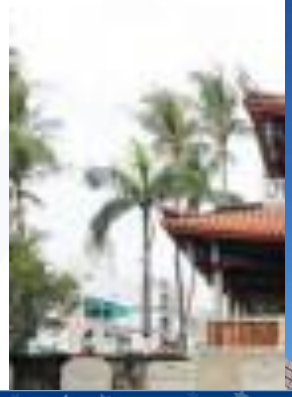
Social media

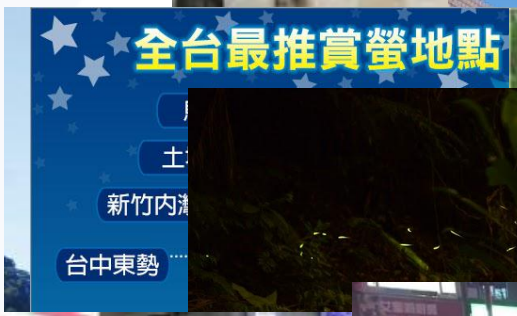
- Flickr

…

Background

- **Types of users' interest**
  - Tourist attraction
  - Landmark, Scenic point or area
  - Seasonal spots / events
  - Fad/bandwagon effect
    - e.g. specific object,
    -



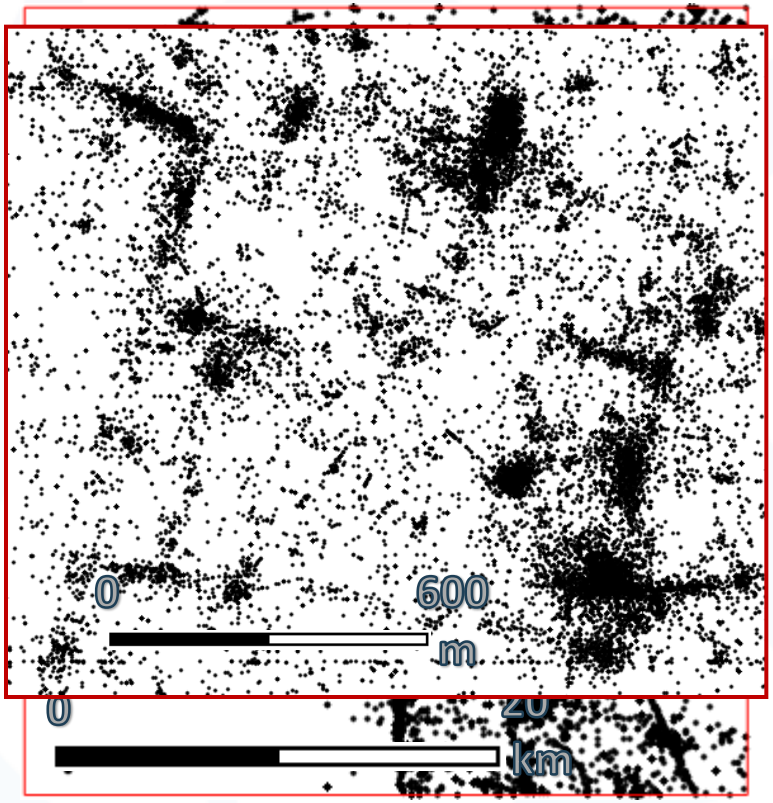


(tilted mailboxes in 2015)

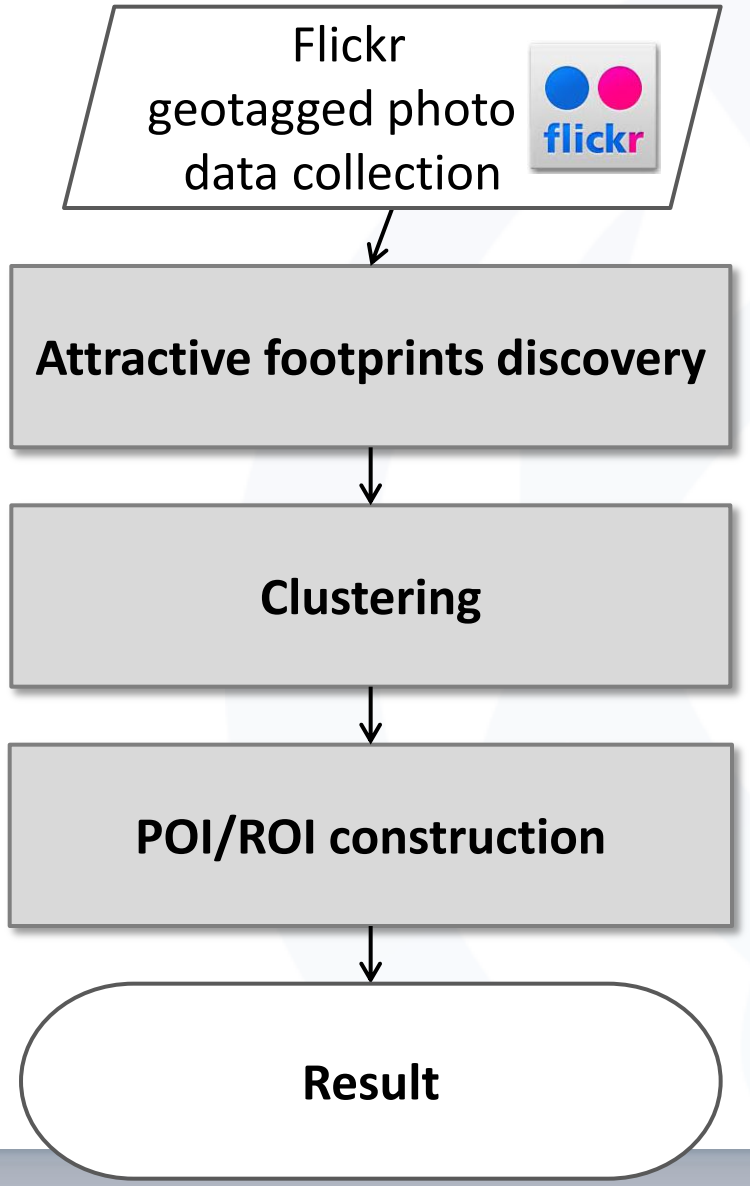
https://tung.hakka.gov.tw

Motivation & Goals

- **How to retrieve users' interest from geotagged photos as GI which helps in trend prediction and decision making?**
  - **Where** is the users' interest?
    - **Location**
  - **What** is it?
    - **Name**
  - **What** is the **range** of it?
    - **POI and ROI**
  - **When** does it appear?
    - **Life cycle**



Workflow


Flickr
geotagged photo
data collection
flickr
Attractive footprints discovery
Clustering
POI/ROI construction
Result

Method- data collection

- **Flickr API** (http://www.flickr.com/services/rest/?method=...)
  - flickr.photos.search
    - PhotoID, ownerID, title, lat, lng, accu
  - flickr.photos.getExif
    - DateTimeOriginal, focus distance, et

id	owner	secret	server	farm	title	ispublic	isfriend	is
18826423800	12646296@N04	2e3e4eec42	405	1	IMG_5770LR	1	0	0
17919791200	94118792@N06	e3737993cc	7680	8	迷你點	1	0	0
15022638700	30661345@N05	bb1187f535	3906	4	DSCN4020	1	0	0

exif

{"photo":
{"id":"18826423800","secret":"2e3e4eec42"...
{"photo":
{"id":"17919791200","secret":"e3737993cc"...
{"photo":
{"id":"15022638700","secret":"bb1187f535"...
{"photo":
{"id":"14534901300","secret":"3bdc019e25"...
{"stat":"fail","code":2,"message":"Permission der

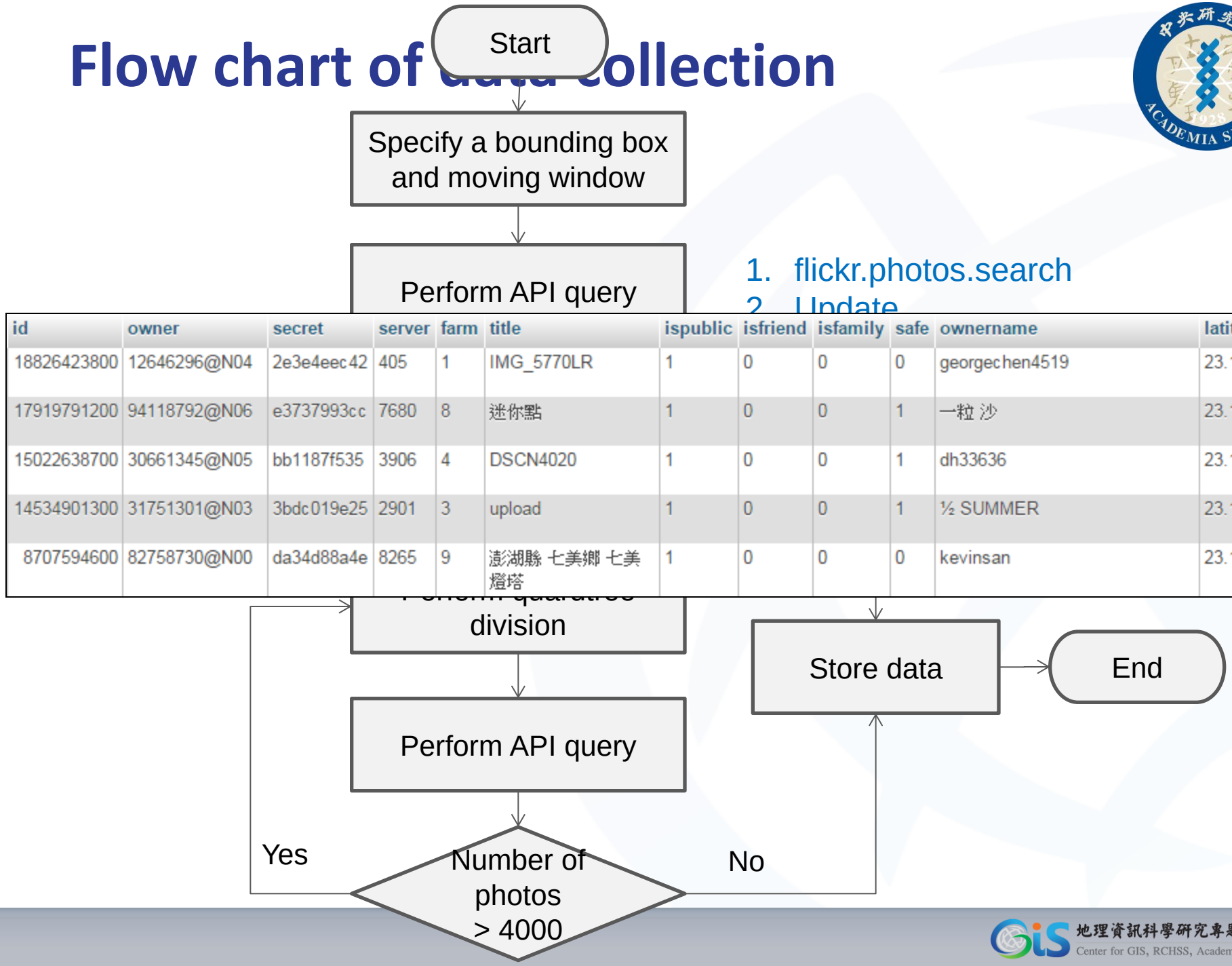

```
{"photo":
 { "id":"18826423800","secret":"2e3e4eec42","server":"405","farm":1,"camera":"Canon EOS 6D",
   "exif":[
   {"tagspace":"IFD0","tagspaceid":0,"tag":"Make","label":"Make","raw":{"_content":"Canon"}},
   {"tagspace":"IFD0","tagspaceid":0,"tag":"Model","label":"Model","raw":{"_content":"Canon EO
   {"tagspace":"IFD0","tagspaceid":0,"tag":"XResolution","label":"X-Resolution","raw":{"_conte
   {"tagspace":"IFD0","tagspaceid":0,"tag":"YResolution","label":"Y-Resolution","raw":{"_conte
   {"tagspace":"IFD0","tagspaceid":0,"tag":"ResolutionUnit","label":"Resolution Unit","raw":{"
   {"tagspace":"IFD0","tagspaceid":0,"tag":"Software","label":"Software","raw":{"_content":"Ad
   {"tagspace":"IFD0","tagspaceid":0,"tag":"ModifyDate","label":"Date and Time (Modified)","ra
   {"tagspace":"ExifIFD","tagspaceid":0,"tag":"ExposureTime","label":"Exposure","raw":{"_conte
   {"tagspace":"ExifIFD","tagspaceid":0,"tag":"FNumber","label":"Aperture","raw":{"_content":"
   {"tagspace":"ExifIFD","tagspaceid":0,"tag":"ExposureProgram","label":"Exposure Program","ra
   {"tagspace":"ExifIFD","tagspaceid":0,"tag":"ISO","label":"ISO Speed","raw":{"_content":"100
   {"tagspace":"ExifIFD","tagspaceid":0,"tag":"SensitivityType","label":"Sensitivity Type","ra
```

Method- data collection

- **Flickr API** (http://www.flickr.com/services/rest/?method=...)

Method- data collection

Flow chart of data collection



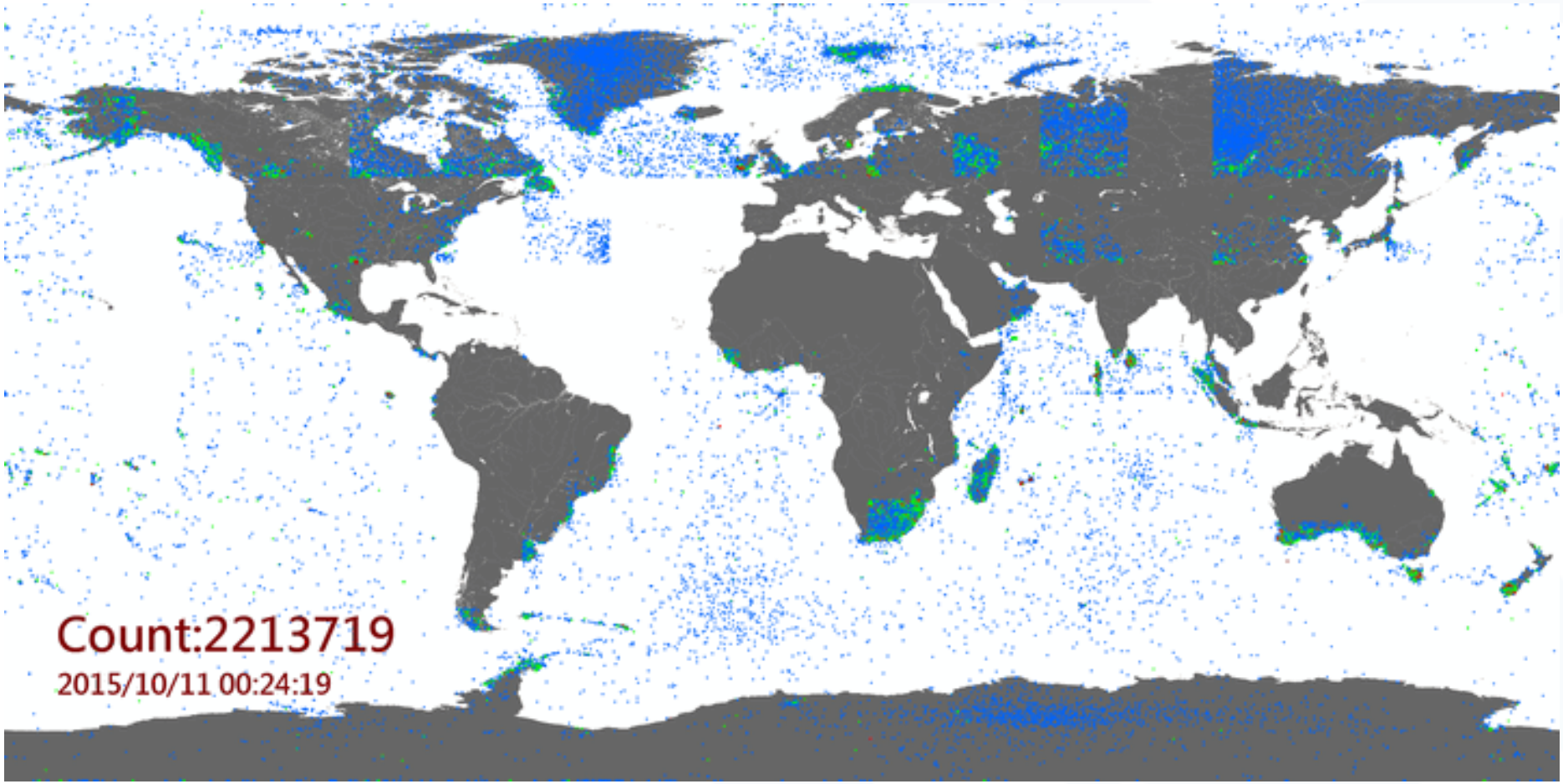
id	owner	secret	server	farm	title	ispublic	isfriend	isfamily	safe	ownername	latitude
18826423800	12646296@N04	2e3e4eec42	405	1	IMG_5770LR	1	0	0	0	georgechen4519	23.189638
17919791200	94118792@N06	e3737993cc	7680	8	迷你點	1	0	0	1	一粒 沙	23.195861
15022638700	30661345@N05	bb1187f535	3906	4	DSCN4020	1	0	0	1	dh33636	23.197568
14534901300	31751301@N03	3bdc019e25	2901	3	upload	1	0	0	1	½ SUMMER	23.195758
8707594600	82758730@N00	da34d88a4e	8265	9	澎湖縣 七美鄉 七美燈塔	1	0	0	0	kevinsan	23.190231

```
$api_key=getKey();
ec
```

```
API key: cfa0db16daa3889cfa458dc443d4703c
Array
(
    [id] => 22016384222
    [owner] => 97099263@N03
    [secret] => ed63c67327
    [server] => 5770
    [farm] => 6
    [title] => ha_2015-10-08-1050
    [ispublic] => 1
    [isfriend] => 0
    [isfamily] => 0
    [safe] => 1
    [ownername] => ncsmsun
    [latitude] => 35.165083
    [longitude] => 136.899694
    [accuracy] => 16
    [context] => 0
    [place_id] => MGYBCxxTWr.m19JQPg
    [woeid] => 28546749
    [geo_is_family] => 0
    [geo_is_friend] => 0
    [geo_is_contact] => 0
    [geo_is_public] => 1
    [media] => photo
    [media_status] => ready
    [url] => http://farm6.static.flickr.com/5770/22016384222_ed63c67327.jpg
)
```

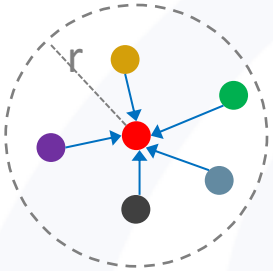
```
."&method=flickr.photos.search
."&extras=owner_name,geo,media"
//."&min_taken_date=1970-01-01%2000%3A00%3A00"
."&safe_search=1"
."&min_upload_date=1970-01-01 00:00:00"
//."&max_upload_date:2015-06-30 00:00:00"
."&nojsoncallback=1"
."&page={$page}"
."&per_page={$per_page}"  //最多每頁500筆
."&sort=date-posted-desc" //sort:date-posted-desc  interestingness-desc
."&src=js"
."&ticket_number=2");
```

Crawler



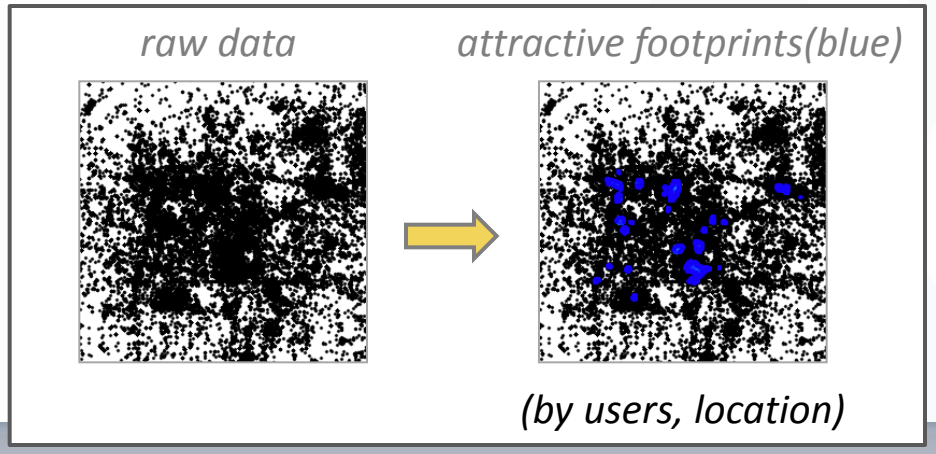
Method- Attractive footprints discovery

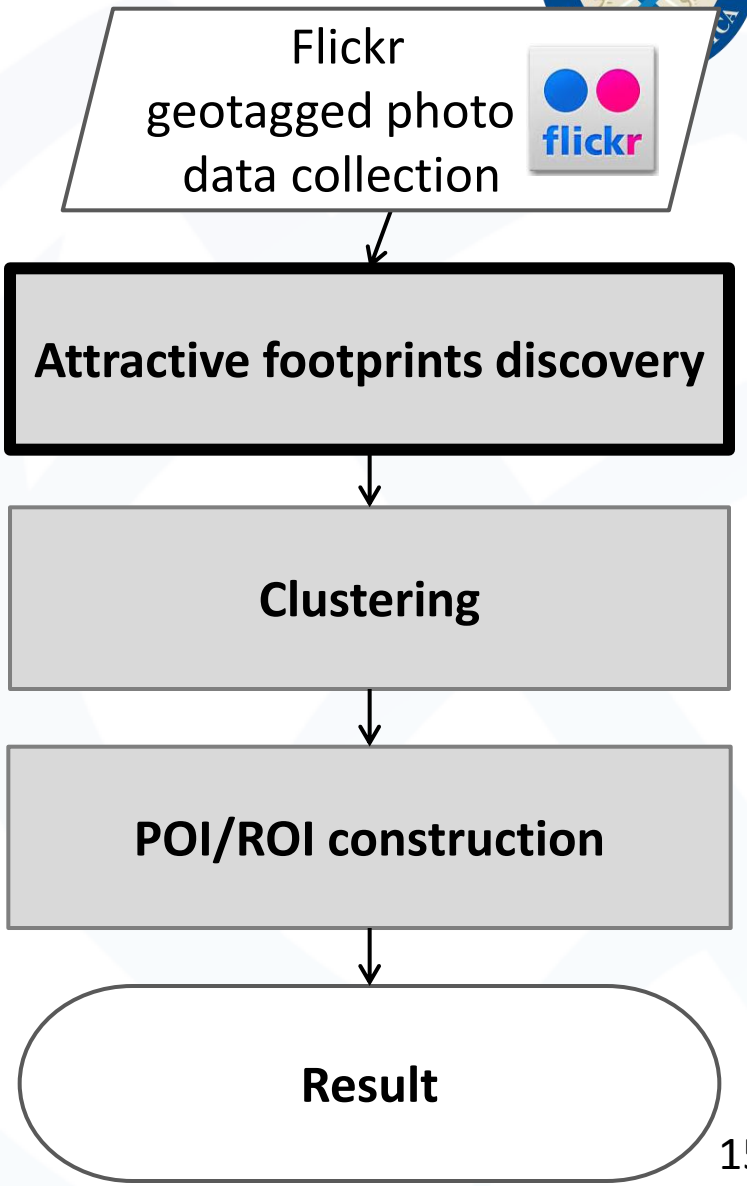
- **Voting value *υ* calculation**
  - Gaussian distance
    - σ =16 while r=50 m



$$\upsilon_{p_i} = \sum_{i=1}^{j} \omega_{ij}, \omega_{ij} = e^{-\frac{\|i-j\|^2}{2\sigma^2}}$$

  - Non duplicated users
  - Attractive footprints
    - $\upsilon \geq T_1$ *(e.g. 30)*





Method- Clustering

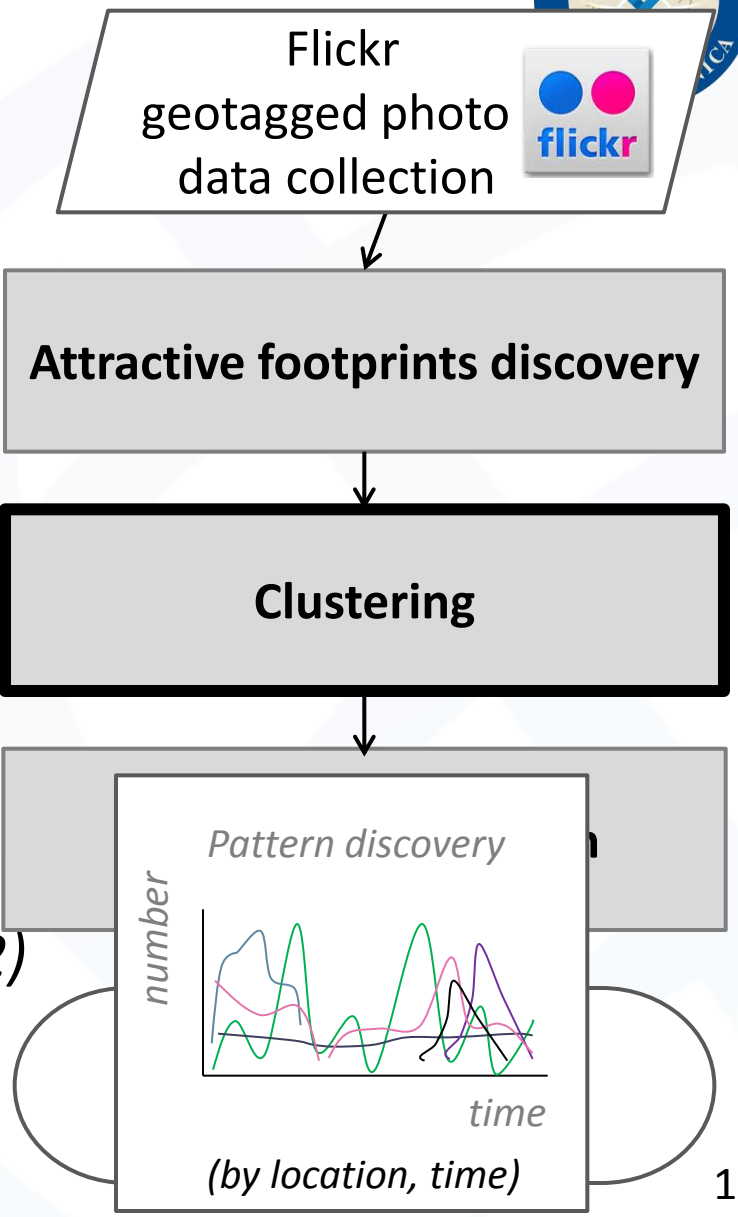
- **Pattern discovery**
  - Valid time, attributes
  - 12 months (r=50 m)

$$X_N = \frac{(X - X_{min})}{(X_{max} - X_{min})}, X_N \in [0,1]$$

$$P_{diff} = \sqrt{\sum_{i=1}^{j} (X_{N_i} - Y_{N_i})^2}$$

$$P_{diff_N} = \frac{(P_{diff} - P_{diff_{min}})}{(P_{diff_{max}} - P_{diff_{min}})}, P_{diff_N} \in [0,1]$$

  - Similar pattern: $P_{diff_N} \leqq T_2$ *(e.g. 0.2)*

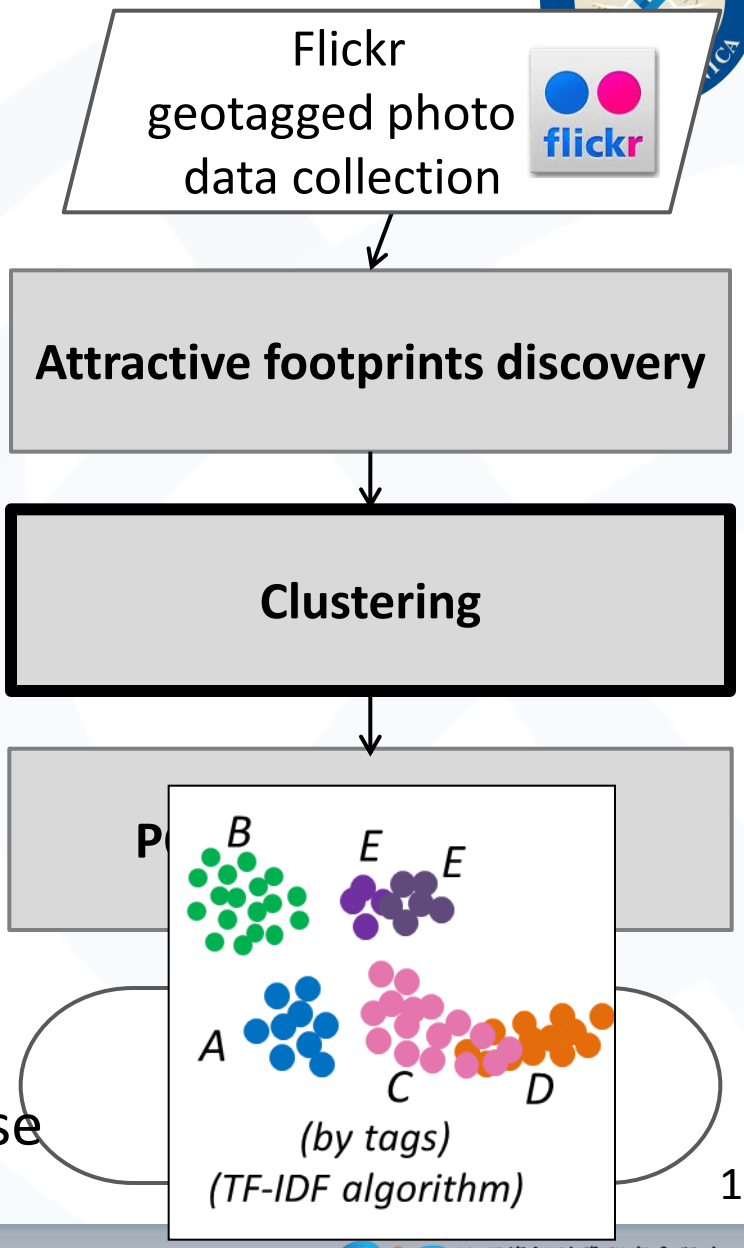


Method - Clustering

- **Clustering with spatial overlap algorithm (SO algorithm)**
  - T_3 is set as a buffer
  - Attractive footprints can be processed/grouped again while $T_2 - T_3 \leqq P_{diff_N} \leqq T_2$
- **Naming**
  - TF-IDF (Liu and Yang, 2012)

$$a_{ij} = \log(tf_{ij} + 1.0) * \log(\frac{N + 1.0}{n_j})$$

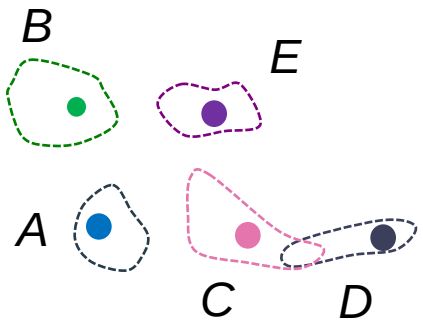
- **Merge**
  - the same name and spatially close



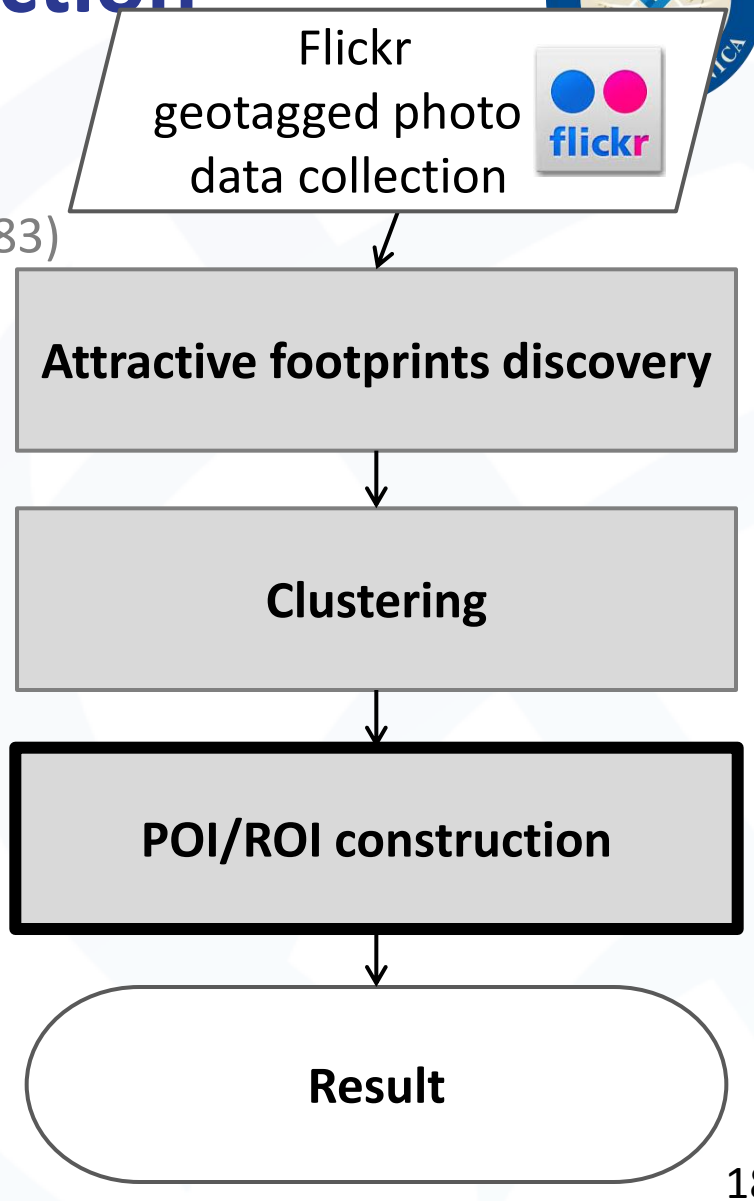
Method- POI/ROI construction

- **Region of interest (ROI)**
  - Alpha-shape (Edelsbrunner et al., 1983)
- **Point of interest (POI)**
  - Peak value of cluster

Result: POI/ROI construction



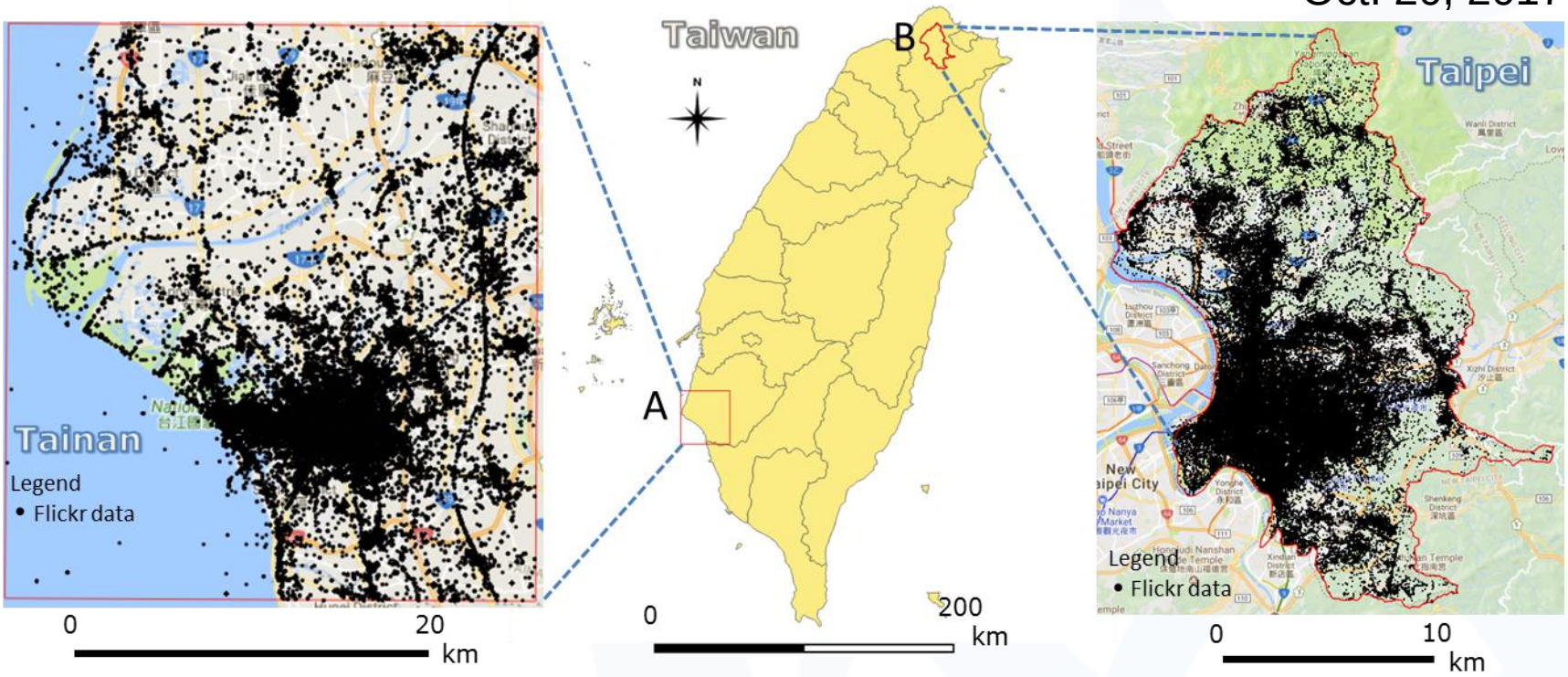
(ROI: Alpha-shape algorithm
POI: peak value of cluster)



Implementation

- **Test area: Tainan city and Taipei City**

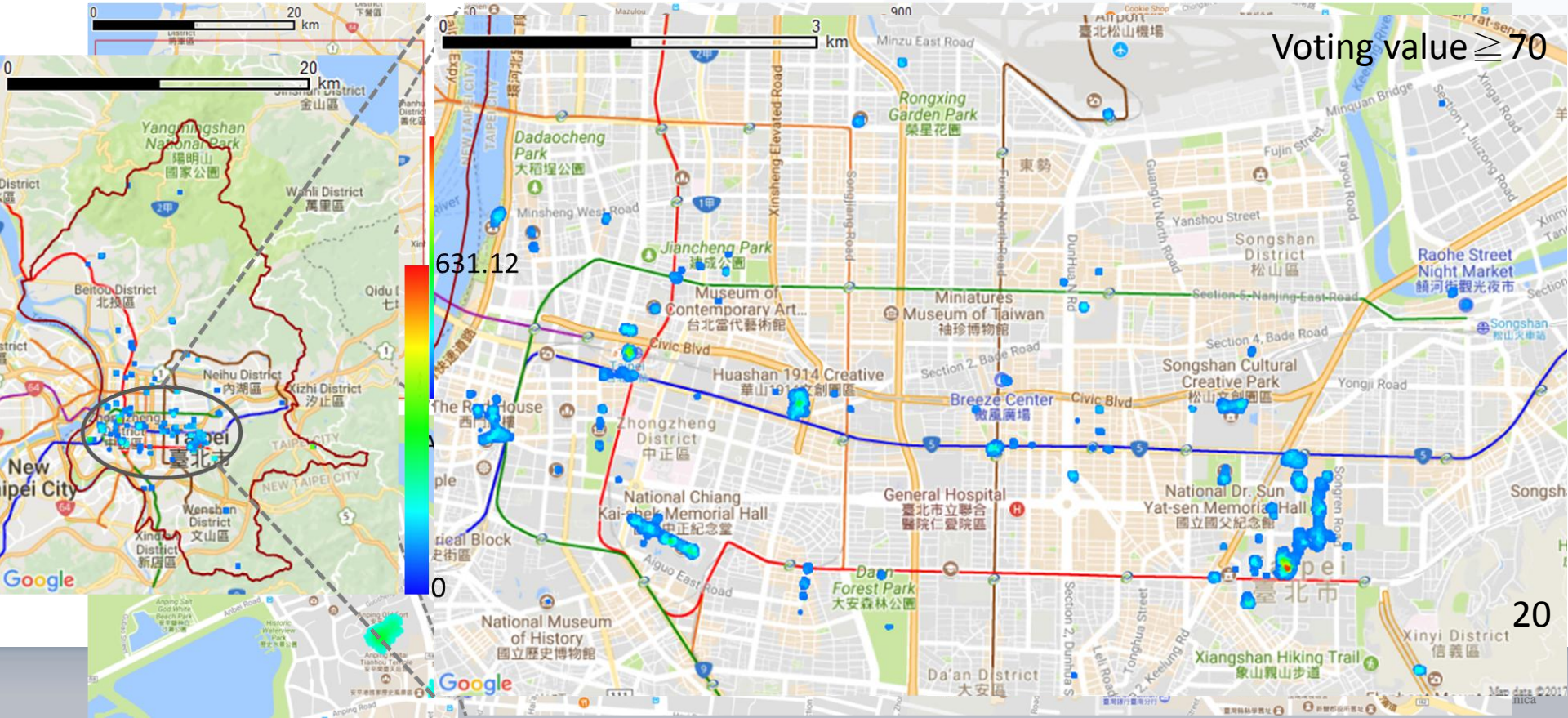
~ Oct. 20, 2017



Item	Study area A (ca. 991 km2)	Study area B(ca. 272 km2)
Total number of photos	276,018	1,956,980
Percentage of photos in Taiwan	3.44%	24.36%
Distinct contributed users	6,749	22,886
User tags (total/distinct)	925,761/34,140	2,918,749/97,803
Photos with user tags	144,249	406,461

Implementation

- **Test area (Accuracy 12~16, street level)**
  - Tainan City: 276,018→256,149 photos , 6,749→5,792 users
  - Taipei City: 1,956,980→ 1,895,042 photos, 22,886→20,566
  - Search radius: 50 m



Implementation

- **Top 10 POI/ROIs in Tainan**

林百貨(1)

正興咖啡館(2)

赤崁樓(3)

武廟(9)

安平樹屋(4)

德記洋行(6)

Photos from Flickr or Tourism Bureau of Tainan City Government (https://www.twtainan.ne

Implementation

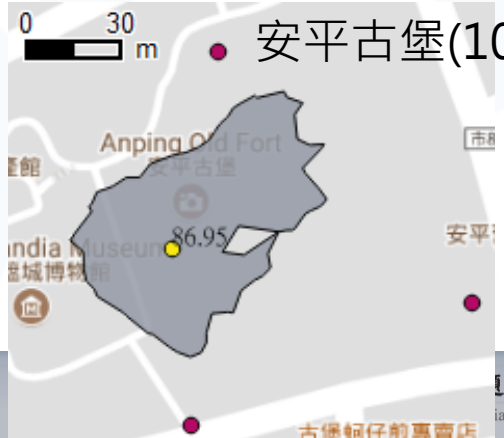
- **Top 10 POI/ROIs**

孔廟(5)

台灣文學館(7)

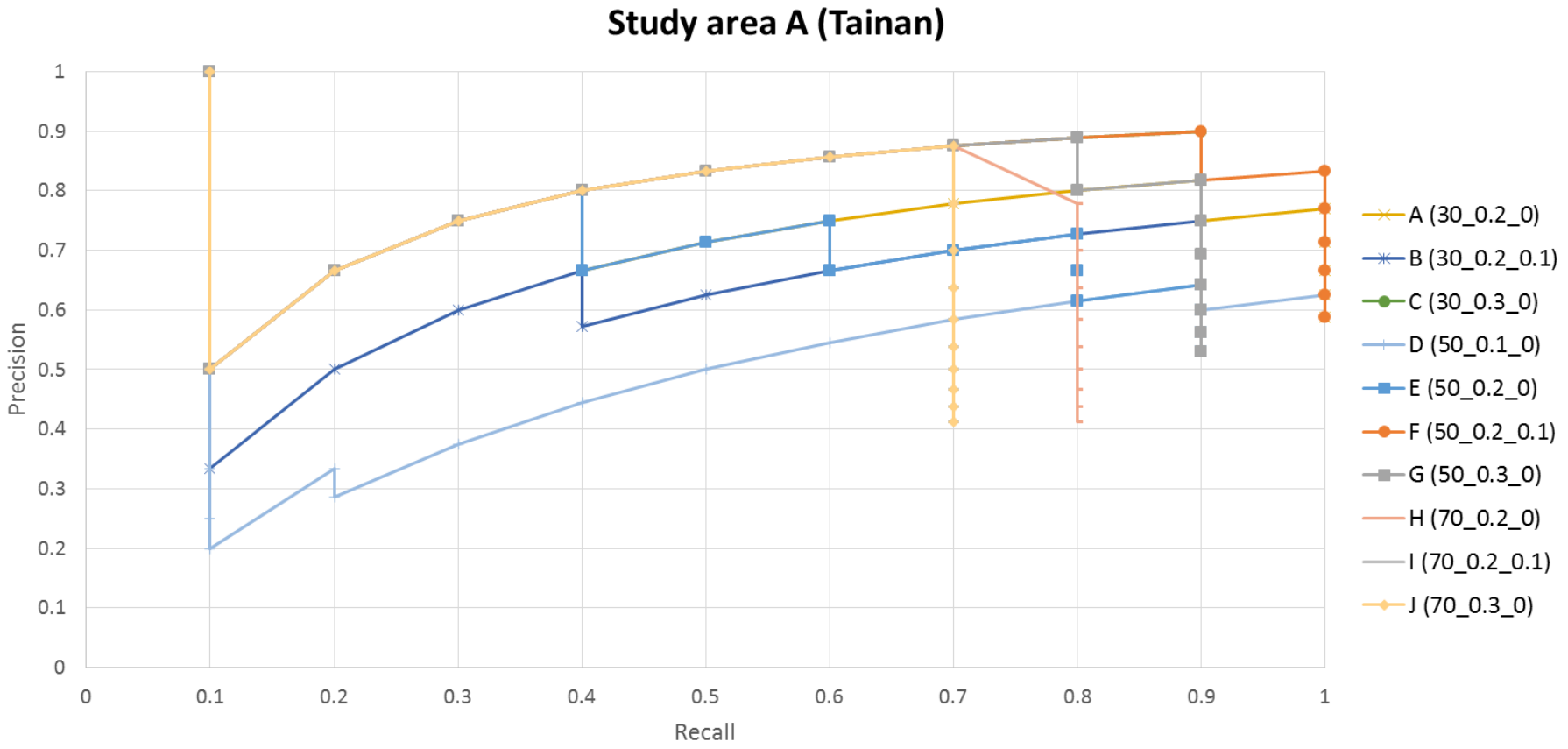
神農街(8)

安平古堡(10)



Discussion and evaluation

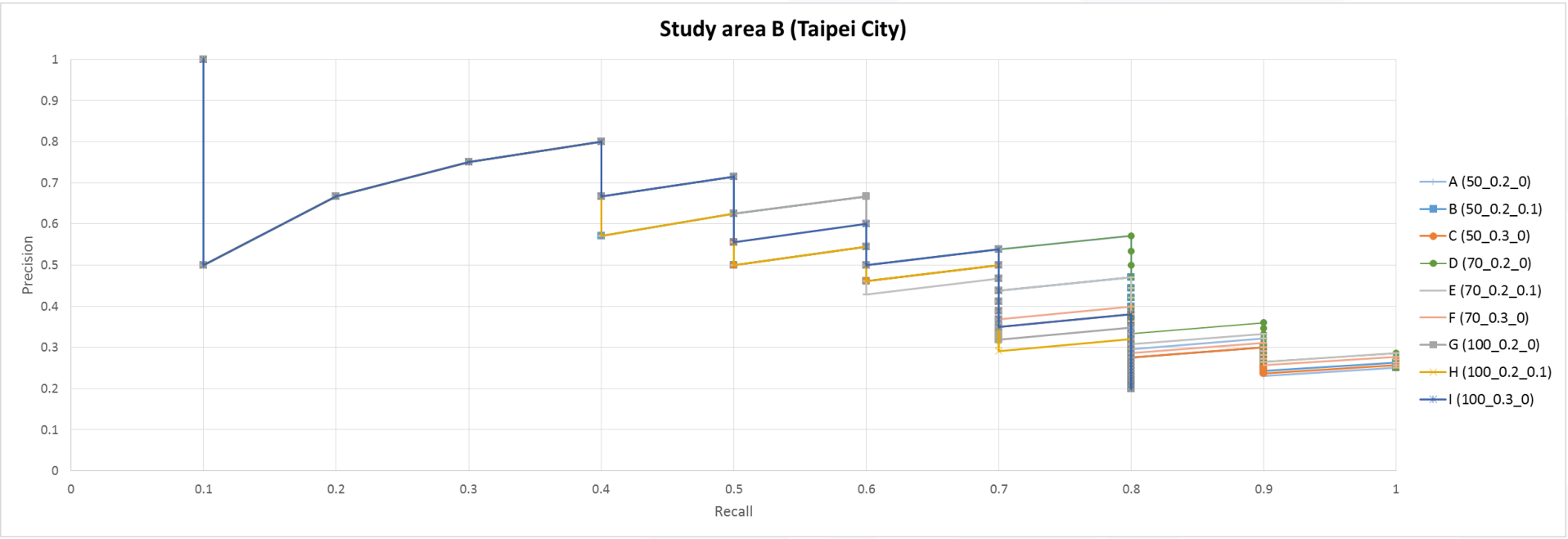
- T_1, T_2, and T_3 set as 50, 0.2, and 0.1, respectively has a much better performance in Tainan.



台南市觀光局:林百貨,赤崁樓,安平樹屋,安平古堡,祀典武廟,孔子廟,奇美博物館,神農街,德記洋行,國立台灣文學館

Discussion and evaluation

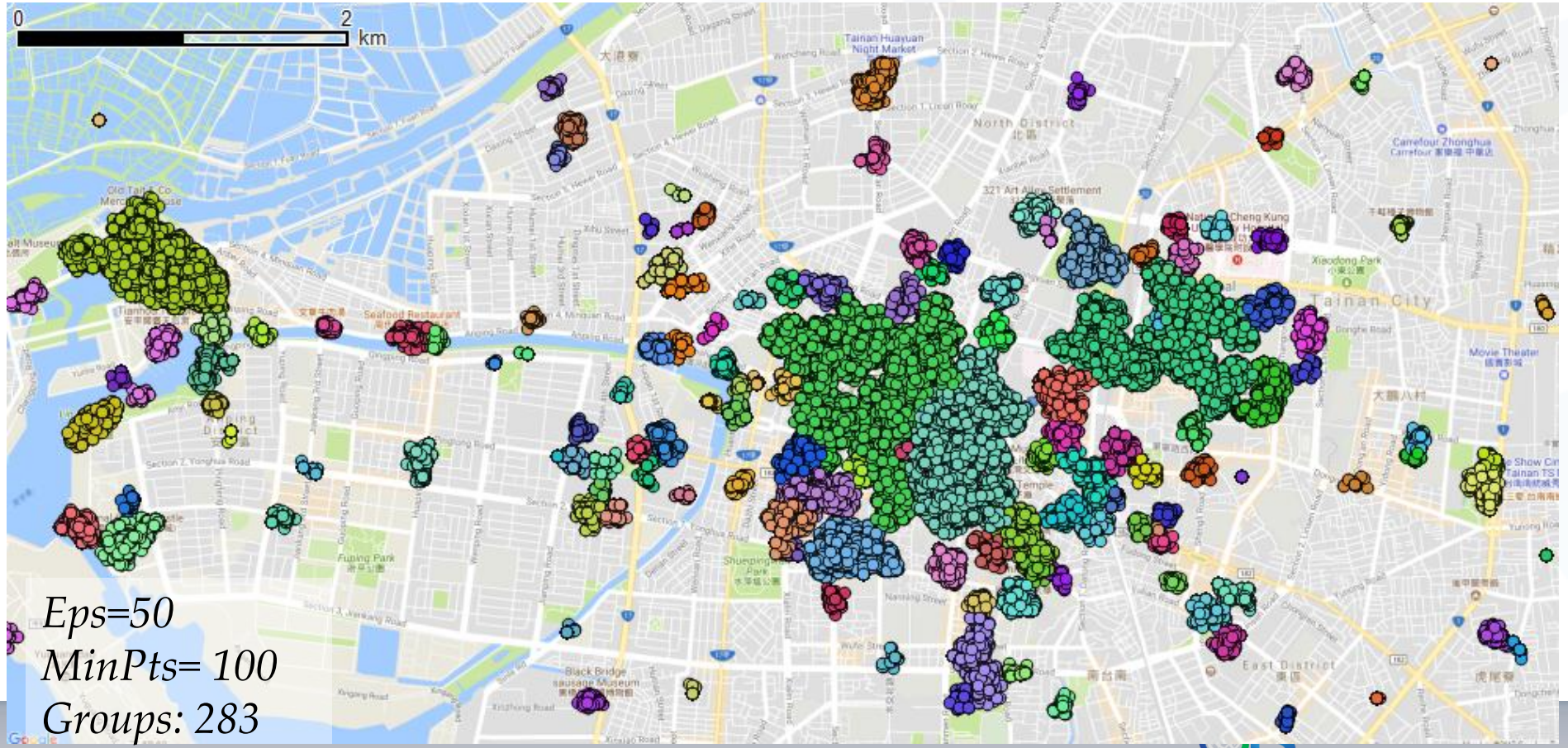
- **T_1, T_2, and T_3 set as 70, 0.2, and 0.1, respectively has a much better performance in Taipei.**



臺北市政府觀光傳播局 :國立中正紀念堂,國立國父紀念館,國立故宮博物院,松山文創園區, 華山1914文化創意產業園區,市立動物園, 臺北101觀景台, 西門紅樓, 艋舺龍山寺, 象山自然步道

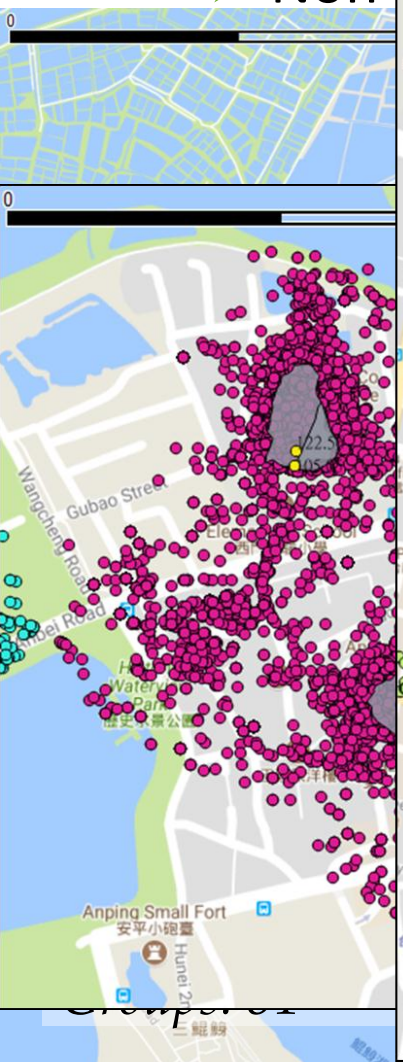
Discussion and evaluation

- **DBSCAN (Ester et al., 1996)**
  - Density-based spatial clustering of applications with noise: *Eps*, *MinPts*.

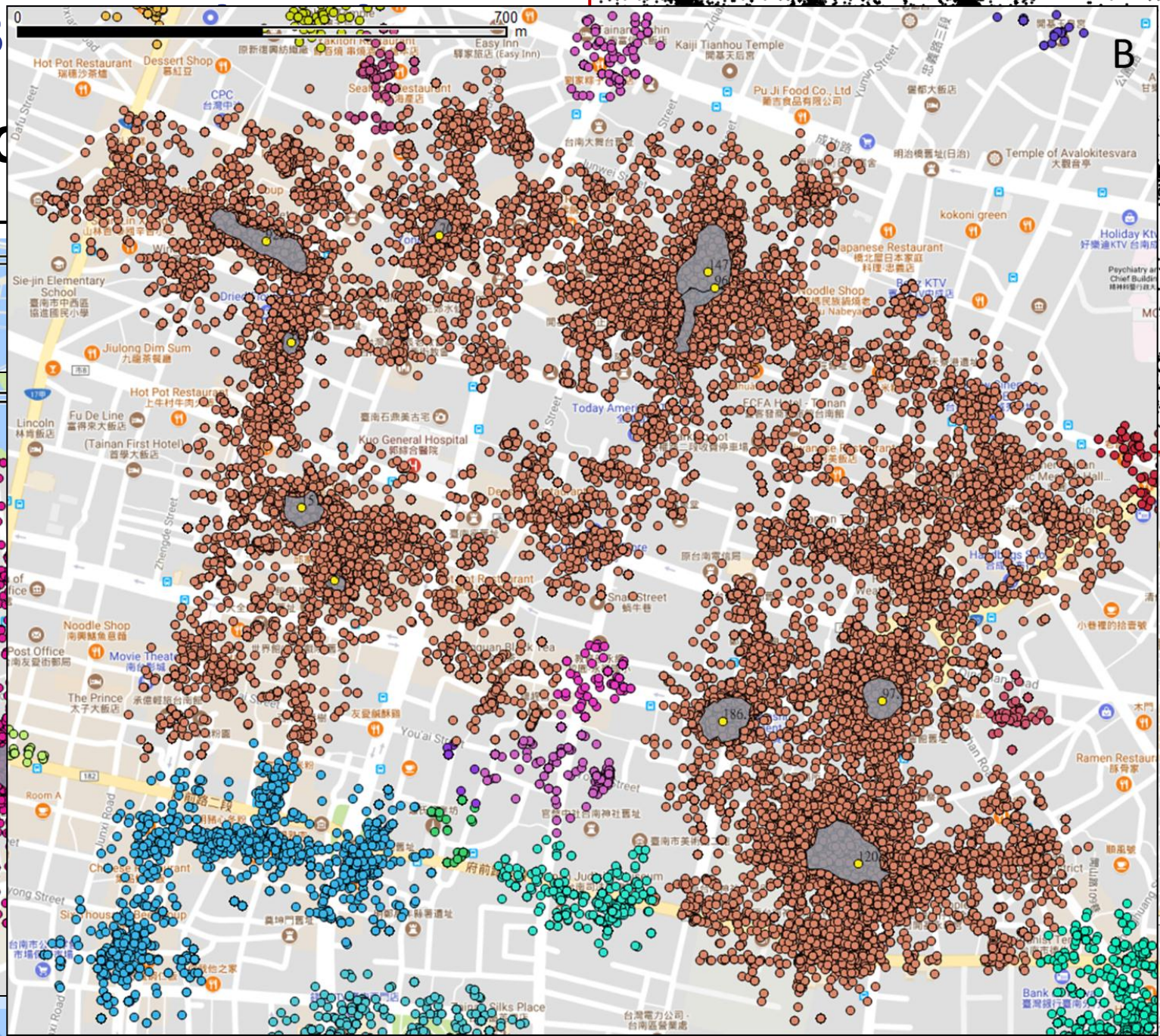


Discuss

- **P-DBSC**
  - Non-



A



Conclusions

- **A efficient m**
  temporal pr
  photos to di
  - It is especia
  - This approa

- Including mo
  - ☒ Location, l
  - ☐ Image, ori

- **Semantics re**
  - Semantics c
  - Semantics c

tag	nikename	count	weight
台灣	台灣	286796	0
square	square	105605	0
iphoneography			0
squareformat			0
instagramapp			0
uploaded:by=ins			0
friendlyflickr			0
台北			0
canon			0
flickrandroidapp:			0
nikon			0
uploaded:by=flic			0
kaohsiung			0
flickriosapp:filter			0
sony			0
台南			0
travel	travel	41022	0

```
[戶外] => 6.9859611604402
[屋頂] => 5.5641981474518
[建築] => 5.2702747148209
[赤崁樓] => 4.0653579755107
[建築物] => 3.324955380586
[taiwan] => 3.324955380586
[舊台南市] => 2.0453046478196
[花園] => 2.0453046478196
[flickriosapp:filter=nofilter] => 2.0453
[canoneos1v] => 2.0453046478196
[ef17404l] => 2.0453046478196
[rvp100] => 2.0453046478196
[tainancity] => 1.8697500041566
[tainan] => 1.6853587130773
[chikanlou] => 1.5363703610563
[fortprovintia] => 1.5363703610563
[westcentraldistrict] => 1.500402669842
[台南] => 1.3148784043832
[文字] => 1.3148784043832
[zhongxidistrict] => 1.0258481332283
[taiwan2015] => 1.0258481332283
[橫樑] => 1.0258481332283
[asia] => 1.0258481332283
[臺南市] => 1.0258481332283
```

International Journal of Geo-Information

Volume 7, Issue 3

ISPRS Int. J. Geo-Inf. **2018**, *7*(3), 121; doi:10.3390/ijgi7030121

Open Access Article

Efficient Method for POI/ROI Discovery Using Flickr Geotagged Photos

Chiao-Ling Kuo [1,*], **Ta-Chien Chan** [1], **I-Chun Fan** [1,2] and **Alexander Zipf** [3]

1 Research Center for Humanities and Social Sciences, Academia Sinica, Taipei 115, Taiwan
2 Institute of History and Philology, Academia Sinica, Taipei 115, Taiwan
3 Institute of Geography, Heidelberg University, 69120 Heidelberg, Germany

* Author to whom correspondence should be addressed.

Received: 17 January 2018 / Revised: 14 February 2018 / Accepted: 12 March 2018 / Published: 16 March 2018

View Full-Text | Download PDF [8164 KB, uploaded 16 March 2018] | Browse Figures

http://www.mdpi.com/2220-9964/7/3/121

Thanks for your attention!
Welcome any comments~

kuo@chiaoling.com

地理資訊科學研究專題中心
Center for GIS, RCHSS, Academia Sinica