

# 刑案隱性鏈結關聯模式

An Association Model for Implicit Crime Link Analysis

王朝煌

中央警察大學資管系教授兼教務長



# 內容大綱

- 緒論
- 文獻探討
- 刑案關聯模式
- 實驗結果與分析
- 結語



# 緒論 - 鏈結分析(Link Analysis)

i2 Analyst's Notebook 7 - [圖表 1]

檔案(F) 編輯(E) 檢視(V) 插入(I) 格式(O) 工具(T) 分析(A) 資料(D) 視窗(W) 說明(H)

竊盜 7, 竊盜 8, 竊盜 12, 竊盜 14, 竊盜 15, 竊盜 16, 竊盜 17, 竊盜 18, 竊盜 19, 竊盜 20, 竊盜 21, 竊盜 22, 竊盜 23, 竊盜 24, 竊盜 25, 竊盜 26, 竊盜 27, 竊盜 28, 竊盜 29, 竊盜 30, 竊盜 31, 竊盜 32, 竊盜 33

犯罪

- 會議
- 綁架
- 腳印
- 誇大
- 蒐證
- 輪胎胎紋
- 縱火
- 類別
- 竊盜

- 血型: A
- 血型: B
- 血型: 未知
- 可卡因
- 海洛因
- 縱火
- ▲ 槍擊
- 血型: AB
- 血型: O
- ← 大麻
- ← 狂喜
- ← 搶劫
- 竊擊

標準

生日

格線 NUM



## 緒論 - 隱性鏈結

- 顯性鏈結(explicit social link)
  - 個體(entity)間之實體鏈結(physical link)，如親屬關係，師生關係，電話通聯關係，金融轉帳關係等等。
- 隱性鏈結(implicit link)
  - 個體間運用探勘資料技術推導建立的鏈結，如相似度，共現關係(co-occurrence)等等。



# 緒論 - 研究背景與目的

- 少部份習慣犯罪者犯了相當大比例的犯罪案件。
  - 犯罪者基於安全考量、容易得手及實質利益等因素，往往會反覆運用自己最得意的手段與方法從事犯罪行為(理性選擇理論)。
  - 犯罪間的關係與案發地點間距成反比關係。
  - 犯罪間的關係與案發時間間隔亦相關(反比關係)。
- 運用資料探勘技術，進一步提供隱性鏈結資訊，輔助研判未破案件是否為同一人/集團之犯罪。



# 緒論 - 研究方法

- 蒐集實務機關刑案紀錄。
- 統計分析刑案紀錄內容，建構犯罪關聯模式。
- 實驗測試犯罪關聯模式。
- 透過問卷請實務偵查人員評估模式之效用。



# 緒論 - 刑案紀錄內容與結構

- 基本資料
  - 案件類別
  - 案發地點
  - 案發時間
  - 嫌犯及被害人資料
- 犯罪模式
  - 犯罪成因
  - 犯罪習癖
  - 準備措施
  - 犯罪方法
  - 犯罪工具



# 文獻探討

- 犯罪模式
- 犯罪地緣剖繪
- 犯罪時間分析





# 犯罪模式

- 作案手法
  - 原出於拉丁文的Modus Operandi，意指慣行犯於作案時所慣用的犯罪手段、方法或型式。
- 將犯罪者之獨特性格與行為特質，歸納分析其犯罪模式建立作案手法模式，據以比對可能之涉嫌者或疑似犯罪集團之連續犯行。
- (參考資料: B. E. Turvey, *Criminal Profiling*, Third Edition: *An Introduction to Behavioral Evidence Analysis*, Academic Press; 3 edition April 30, 2008)



# 犯罪地緣剖繪

- 在連續性暴力犯罪中，藉由分析犯案地點以確認犯罪者最可能居住地點的犯罪偵查與管理策略。
- 嫌犯的犯罪地點通常與渠住居所有「距離遞減」效應，越遠離犯罪居住地，犯案次數越少。
- 未破案件，因為缺乏犯罪者的住居地，依距離遞減效應，當刑案之發生地點越相近時，可推估為同一犯罪者所犯案件之機率較高。
- (參考資料:陳仁智，*地緣剖繪技術應用於連續街頭強盜搶奪犯罪偵查之研究*，中央警察大學刑事警察研究所碩士論文，2004)



# 犯罪時間分析

- 以5:00AM作為一天的起始，因為夜生活的人在清晨5點大多已經返家，而早起的人也大多在5點開始活動。
- 利用中位數及第1、3四分位數作為犯罪分析的指標，利用這些指標分析犯罪的集中與離散程度。
- (參考資料: Marcus Felson, Erika Poulsen, 2003, "Simple indicators of crime by time of day", *International Journal of Forecasting*, Volume 19, Issue 4)



# 犯罪時間分析(續)

| Crime<br>Time of day     | 紐約搶奪案 | 台北市搶奪案 | 台北市住宅竊盜 |
|--------------------------|-------|--------|---------|
| Start                    | 05:00 | 05:00  | 05:00   |
| 1 <sup>st</sup> quartile | 15:00 | 12:35  | 11:00   |
| Median                   | 20:30 | 16:50  | 15:10   |
| 3 <sup>rd</sup> quartile | 00:35 | 20:50  | 21:00   |



# 刑案關聯模式

- 犯罪模式關聯
- 犯罪地點關聯
- 犯罪時間關聯
- 結合犯罪模式、犯罪地點、及犯罪時間之關聯模式



# 犯罪模式變數

- 刑案紀錄犯罪模式之組成(變數)
  - 犯罪成因(crime cause, CC)
  - 犯罪習癖(crime habit, CH)
  - 準備措施(preparation action, PA)
  - 犯罪方法(crime method, CM)
  - 犯罪工具(crime tool, CT)



## 犯罪模式關聯- 變數權重

- 分別計算犯罪成因、犯罪習癖、準備措施、犯罪方法及犯罪工具等變數之資訊熵值 (information entropy)，代表變數之權重。
- 隨機變數 $X$ 的值域為 $\{x_1, x_2, x_3, \dots, x_n\}$ ，則其資訊熵值 $E(X)$ 定義為：

$$E(X) \equiv - \sum_{i=1}^n p(x_i) \log_2 p(x_i)$$

$p(x_i)$  為值出現的機率。 (參考:C. E. Shannon, and W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, Urbana, IL, 1949)





## 犯罪模式變數資訊熵值 - 以犯罪習癖為例

| 犯罪習癖         | 出現頻率 | $p(x_i)$ | $-\log_2 p(x_i)$ | 資訊熵值 |
|--------------|------|----------|------------------|------|
| 趁人不備或熟睡或入浴   | 638  | 0.424202 | 1.24             | 0.52 |
| 未填代碼         | 578  | 0.384309 | 1.38             | 0.53 |
| 無犯罪習癖        | 146  | 0.097074 | 3.36             | 0.33 |
| 有把風或共犯       | 66   | 0.043883 | 4.51             | 0.20 |
| 其他習癖         | 44   | 0.029255 | 5.10             | 0.15 |
| 闖空門(趁人不在、無人) | 6    | 0.003989 | 7.97             | 0.03 |
| 自行攜帶工具作案     | 6    | 0.003989 | 7.97             | 0.03 |
| 獨行盜          | 4    | 0.00266  | 8.55             | 0.02 |
| 自備交通工具搬運贓物   | 4    | 0.00266  | 8.55             | 0.02 |
| 專偷商店         | 4    | 0.00266  | 8.55             | 0.02 |
| 帶面具手套作案      | 3    | 0.001995 | 8.97             | 0.02 |
| 恐嚇不得報警       | 2    | 0.00133  | 9.55             | 0.01 |
| 遺留攜帶物        | 2    | 0.00133  | 9.55             | 0.01 |
| 專偷特定物品       | 1    | 0.000665 | 10.55            | 0.01 |
| 合計           | 1504 |          | 犯罪習癖的資訊熵值        | 1.91 |





# 犯罪模式關聯-模式變數值的權重

- 權重：以每一變數值的機率分布作為其權重之計算基礎。

$$\log \frac{p_{V_{PF}}}{p_{VF}}$$

- $p_{V_{PF}}$  為值  $V$  在習慣犯罪者  $P$  的犯罪模式變數  $F$  出現的機率，而  $p_{V_F}$  為值  $V$  在所有案件的犯罪模式變數  $F$  出現的機率 (參考:資訊檢索領域”詞頻/反文件詞頻”，G. Salton, Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer, Addison Wesley, 1989. )。



## 犯罪模式關聯(續)

- 犯罪模式關聯係數：結合變數權重與變數值的權重，計算犯罪者P與案件Q之關聯：(假設變數間兩兩獨立)

$$\sum_{F \in \{CC, CH, PA, CM, CT\}} E(F) \times match ( P_F , Q_F )$$

$$match ( P_F , Q_F ) = \log \frac{p_{V_{PF}}}{p_{V_F}}$$

$$\text{if } P_F = Q_F \quad = \quad \log \frac{p_{V_{PF}}}{p_{V_F}} \quad , 0 \text{ otherwise}$$



## 犯罪地點關聯

- 統計慣犯(犯案10次以上)前後兩案犯罪地點之距離，以其距離分布作為發展數學關聯模式之基礎。
- 求得每個案件與連續搶奪犯模型(重心)的歐幾理得距離，並以數學關聯模式計算犯罪地點關聯。

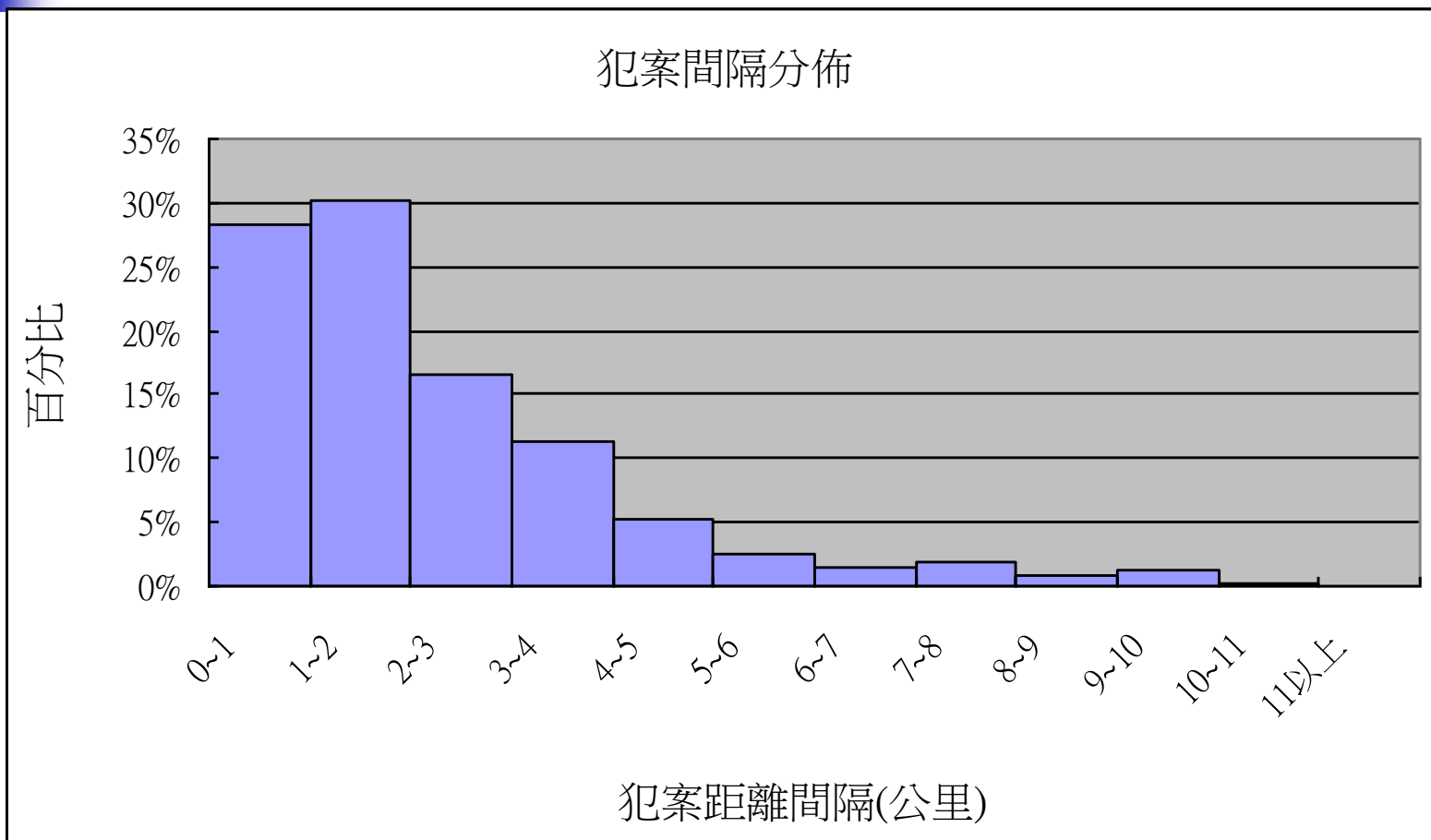


## 犯罪地點之距離分布

| 距離(公里) | 個數 | 百分比    |
|--------|----|--------|
| 0~1    | 92 | 28.31% |
| 1~2    | 98 | 30.15% |
| 2~3    | 54 | 16.62% |
| 3~4    | 37 | 11.38% |
| 4~5    | 17 | 5.23%  |
| 5~6    | 8  | 2.46%  |
| 6~7    | 5  | 1.54%  |
| 7~8    | 6  | 1.85%  |
| 8~9    | 3  | 0.92%  |
| 9~10   | 4  | 1.23%  |
| 10~11  | 1  | 0.31%  |
| 11以上   | 0  | 0.00%  |

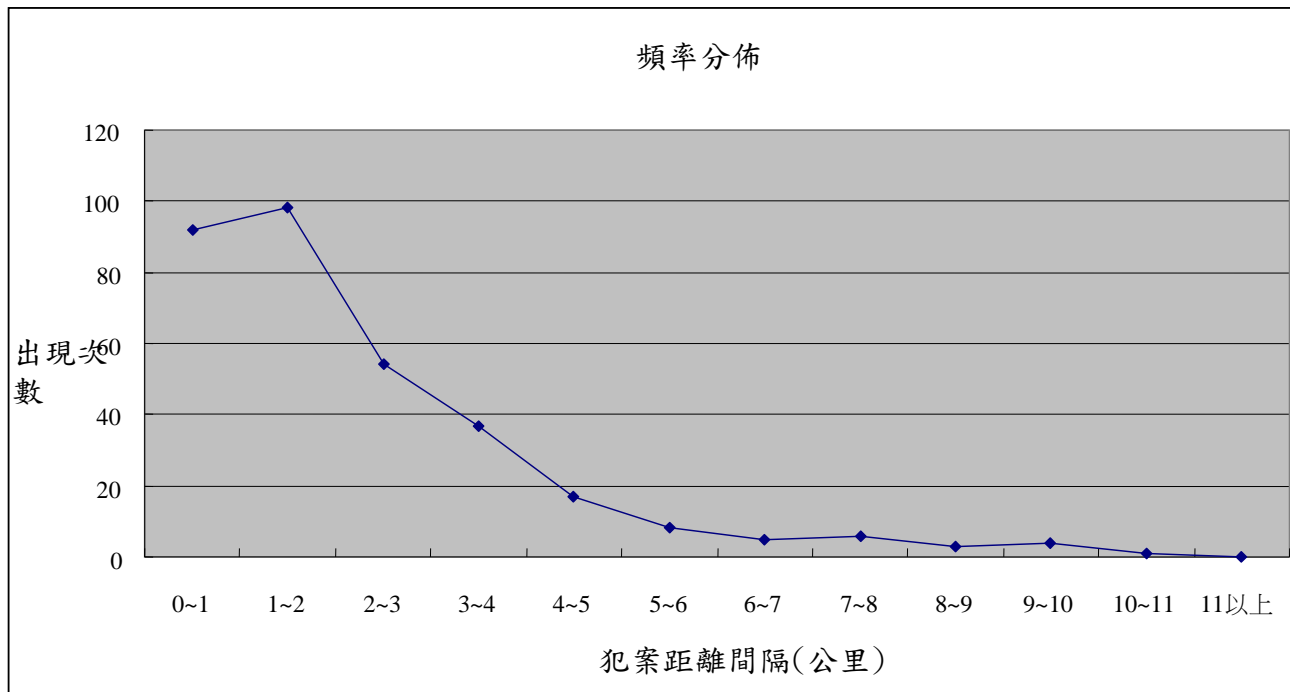


# 犯罪地點之距離分布(續)





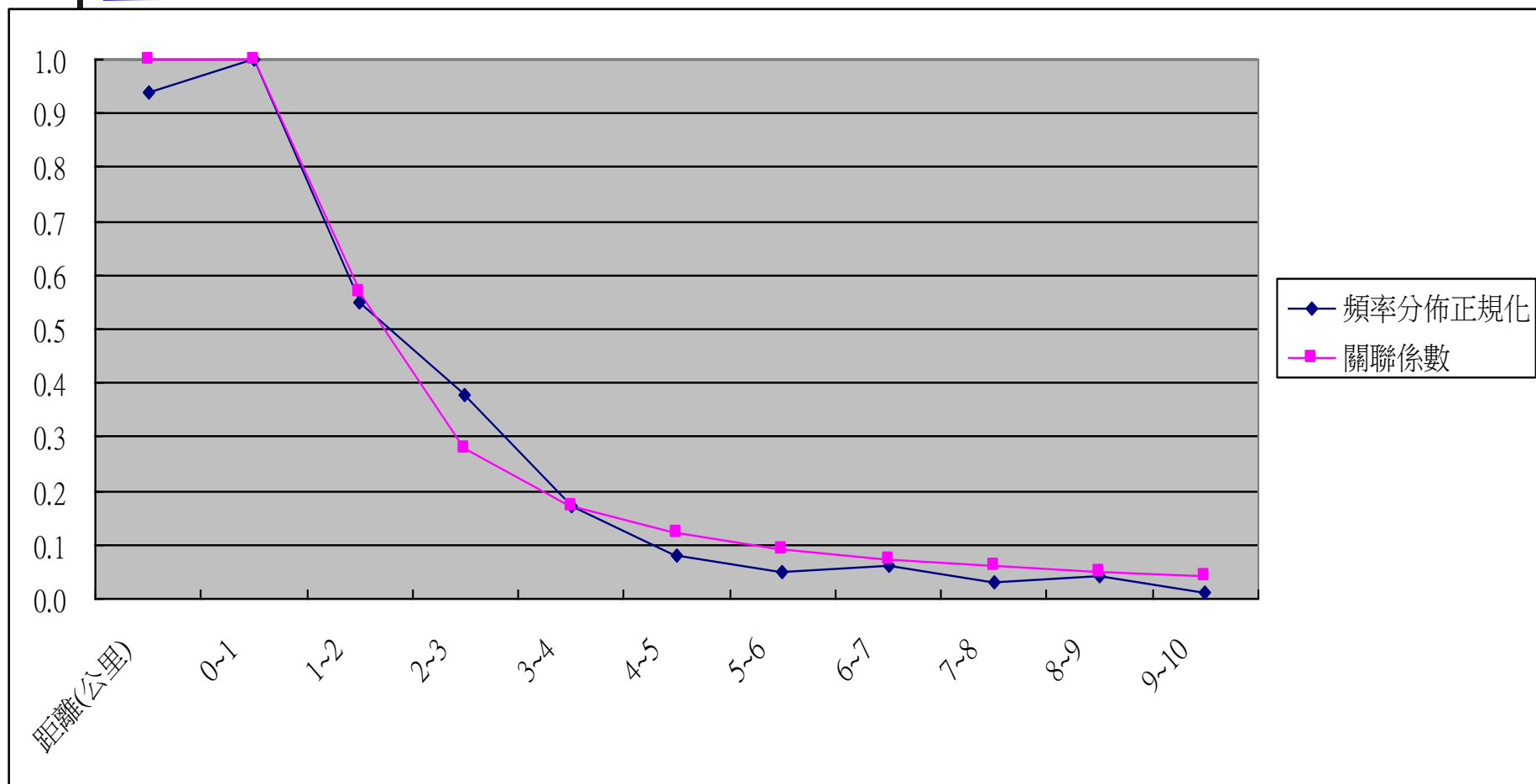
# 犯罪地點關聯模式



|          |   |  |
|----------|---|--|
| ◻        | } | 1, 犯案地點間隔在 2 公里以內◻                       |
| 關聯係數 = ◻ |   | $\frac{1}{(d-1)^{1.4}}$ , 犯案地點間隔大於 2 公里◻ |



# 犯罪地點關聯模式 (續)

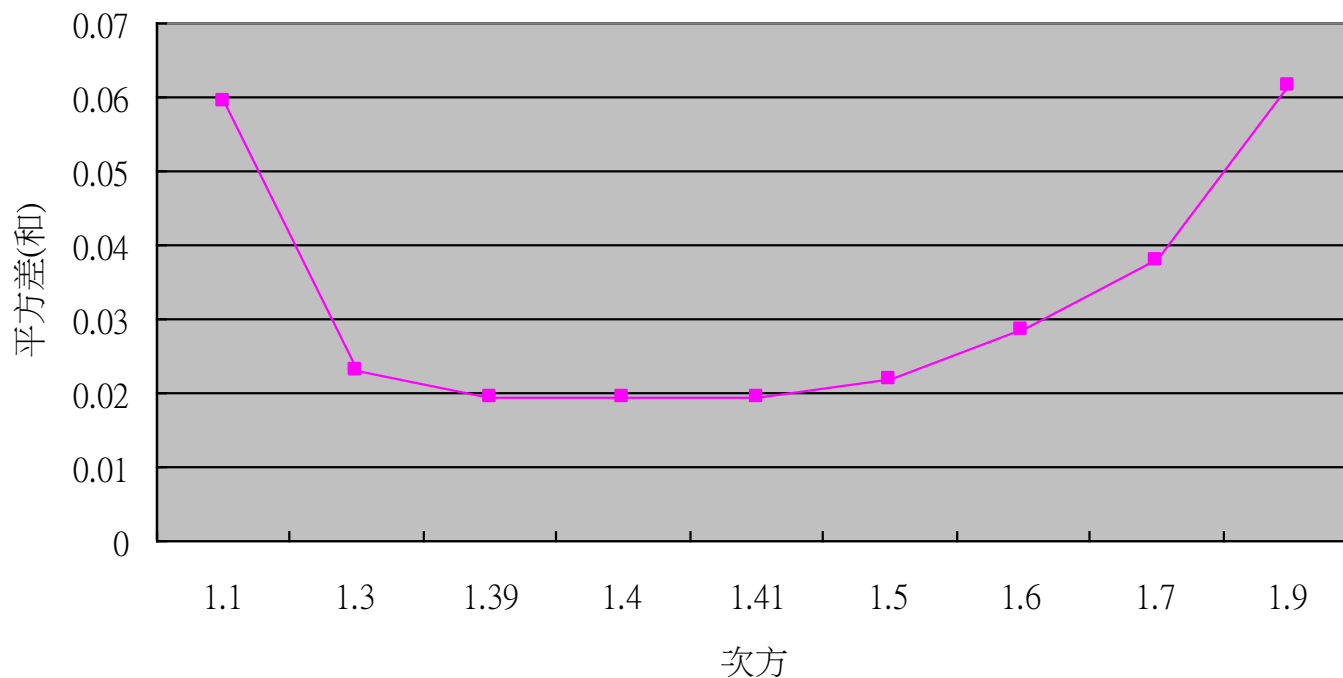




# 犯罪地點關聯模式 (續)

| 次方   | 1.1      | 1.3      | 1.39     | 1.4             | 1.41     | 1.5      | 1.6      | 1.7      | 1.9      |
|------|----------|----------|----------|-----------------|----------|----------|----------|----------|----------|
| 平方差和 | 0.059356 | 0.022940 | 0.019452 | <b>0.019402</b> | 0.019412 | 0.021831 | 0.028382 | 0.037740 | 0.061386 |

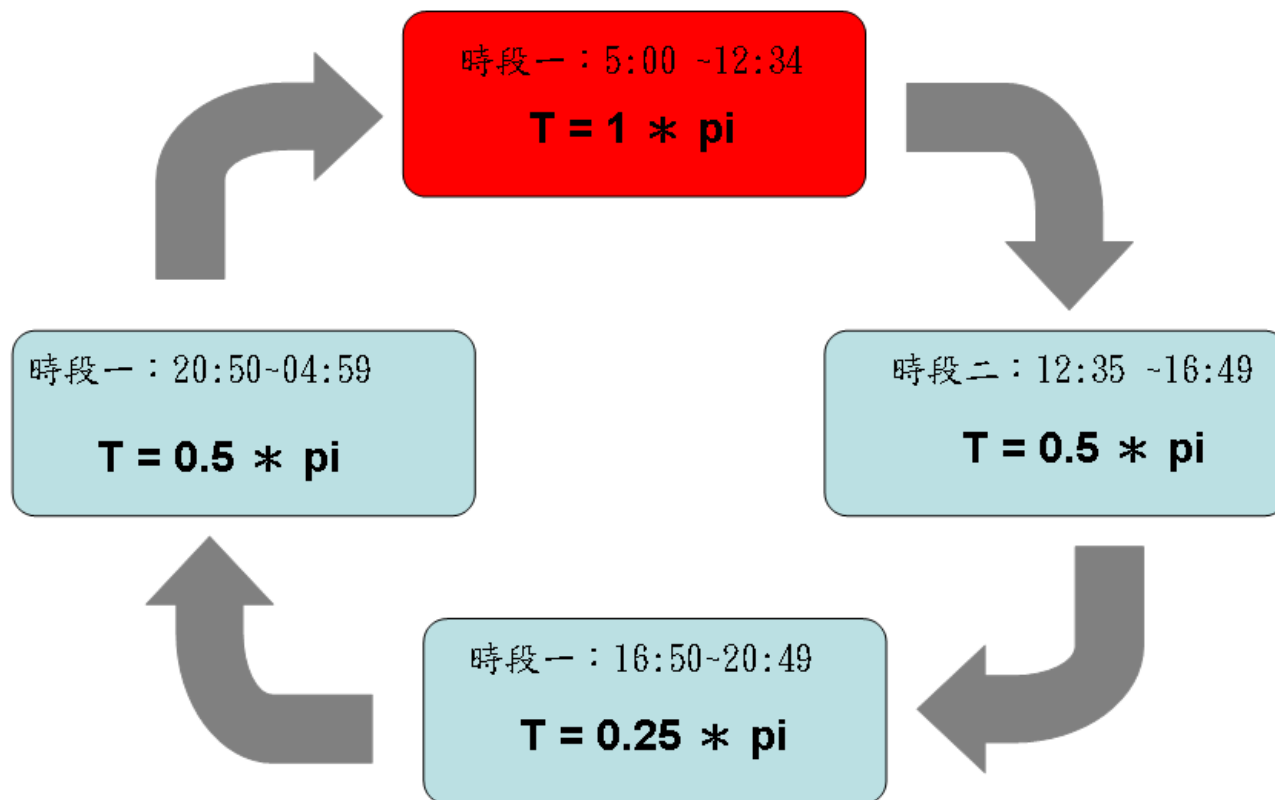
區域特性關聯係數驗證







# 犯罪時間關聯模式





# 刑案關聯模式

- 以實驗尋找
  - 犯罪模式關聯(Profile, P)
  - 犯罪距離關聯(Distance, D)
  - 犯罪時間關聯(Time, T)
- 之最佳組合，發展刑案關聯模式。



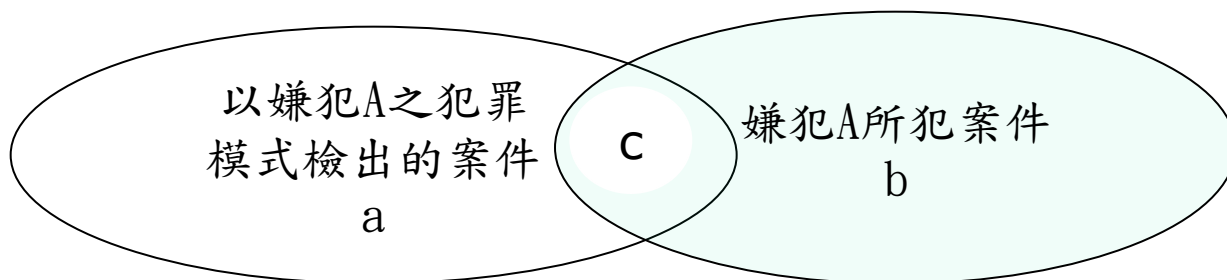
# 實驗設計

- 實驗分為單一因素，雙因素組合，及三因素組合，其中雙因素與三因素分以1:2:4比例組合計算刑案關聯：
  - 單一因素：分別以P, D, T計算刑案關聯。
  - 雙因素組合：以1:2:4比例兩兩組合方式計算(PD, PT, DT)，共計15組。
  - 三因素組合：將PDT以1:2:4等比例組合，共計19組(剔除重複部份)。



# 評估指標

- Precision(P): 以嫌犯A之犯罪模式於某一相似度檢出的案件中，真正為嫌犯A所犯案件的比率，即 $c/a$ 。
- Recall(R): 以嫌犯A之犯罪模式於某一相似度檢出的案件中，其中真正為嫌犯A所犯案件，佔嫌犯A所犯案件的比率，即 $c/b$ 。
- F-measure =  $2PR/(P+R)$





## 實驗資料

- 本研究以臺北市轄內已破的搶奪案件資料作為分析與實驗對象，嫌犯總數為690人，犯案總數為1504人次。
- 分析累計犯案10次以上的嫌疑人共計23人，犯案總數為347人次，占犯案總數23.07%。
- 取**連續犯案10次**以上的14組嫌犯建立犯罪模型。
- 以關聯係數0.9，0.8，0.7及0.6作為關聯門檻值，並分別計算14個搶奪模型之準確率(Precision, P)、檢出率(Recall, R)及綜效(F-Measure)做為評估指標。



## 研究結果與分析

- 單一因素以犯罪距離關聯模式(D)關聯門檻0.9時，綜效為0.45最佳。

| D           |             |      |      |      |
|-------------|-------------|------|------|------|
| 關聯門檻 \ 評估指標 | 0.9         | 0.8  | 0.7  | 0.6  |
| P           | 0.32        | 0.31 | 0.31 | 0.29 |
| R           | 0.75        | 0.75 | 0.78 | 0.79 |
| F           | <b>0.45</b> | 0.44 | 0.44 | 0.42 |



## 研究結果與分析（續）

- 雙因素組合以犯罪距離關聯模式與犯罪時間關聯模式1:2(D1T2)關聯門檻0.8時，綜效為0.55最佳。

| D1T2        |      |             |      |      |
|-------------|------|-------------|------|------|
| 關聯門檻 \ 評估指標 | 0.9  | 0.8         | 0.7  | 0.6  |
| P           | 0.50 | 0.42        | 0.25 | 0.17 |
| R           | 0.54 | 0.78        | 0.89 | 0.94 |
| F           | 0.52 | <b>0.55</b> | 0.40 | 0.29 |



## 研究結果與分析（續）

- 三因素組合有多種態樣，可以依據需求，選取準確性較高，亦或檢出率較高之刑案關聯模式：

| 變數比例         | P1D2T4      | P1D1T2      | P1D4T4      | P1D1T2      |
|--------------|-------------|-------------|-------------|-------------|
| 關聯門檻<br>評估指標 | 0.8         | 0.8         | 0.9         | 0.7         |
| P            | 0.49        | <b>0.57</b> | <b>0.60</b> | 0.40        |
| R            | 0.68        | <b>0.55</b> | 0.49        | <b>0.80</b> |
| F            | <b>0.57</b> | <b>0.56</b> | 0.54        | 0.54        |





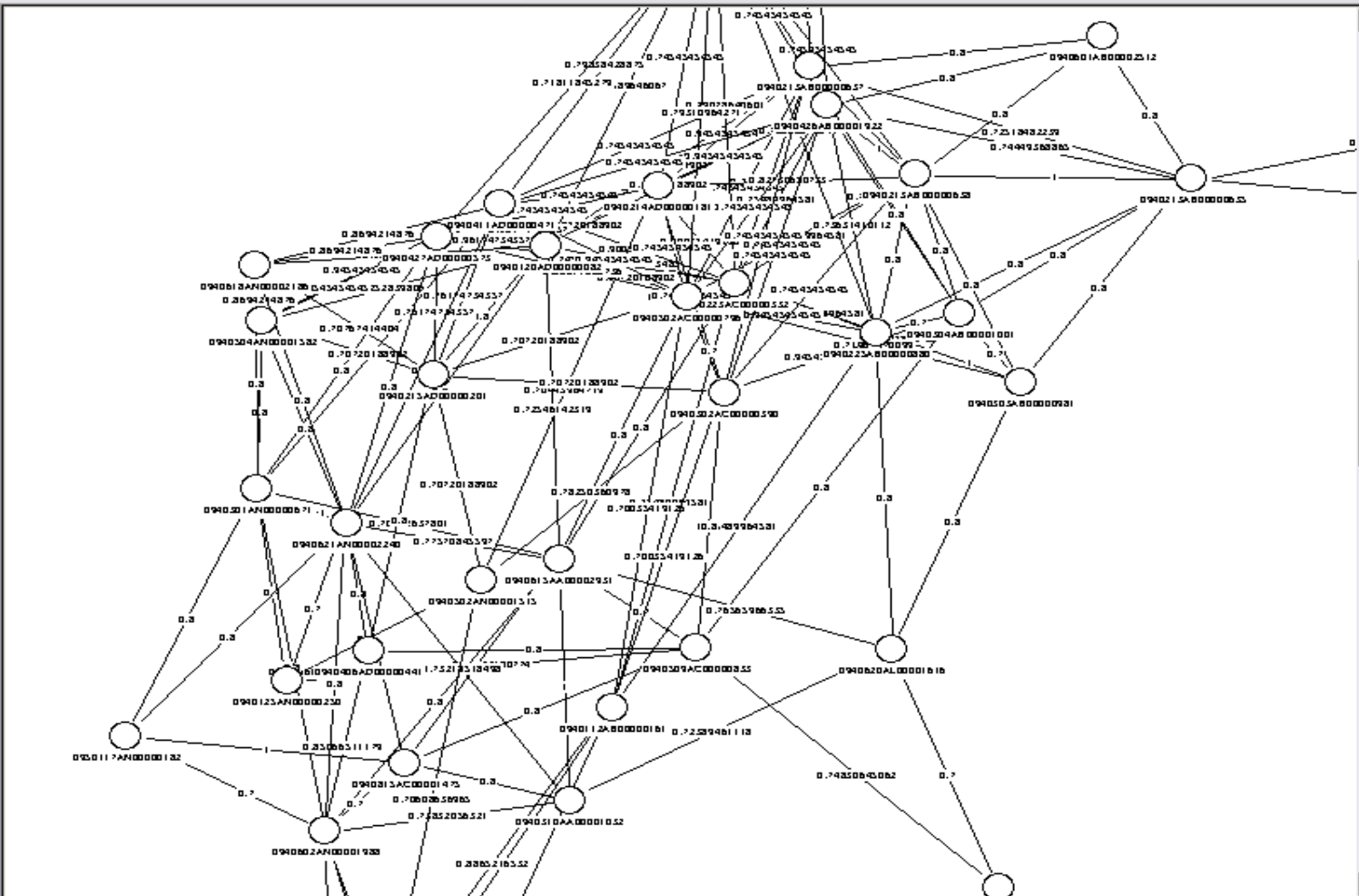
## 結語

- 藉由「刑案隱性鏈結關聯分析」，期協助偵查人員在看似毫無關聯的刑案，透過探勘隱性鏈結建立關聯，輔助顯性鏈結之不足。
- 藉由刑案隱性鏈結，判斷是否為相同嫌犯或前科慣犯再犯之可能性，擴大清查範圍。
- 可搭配鏈結視覺化軟體，輔助犯罪偵查工作。



i2 Analyst's Notebook 7 - [圖表 2]

檔案(F) 編輯(E) 檢視(V) 插入(I) 格式(O) 工具(T) 分析(A) 資料(D) 視窗(W) 說明(H)



常規

- 男性
- 女性
- 未具名
- 用戶
- 電話
- 行動電話
- 帳戶
- 信用卡
- 現金
- 文件

● 中東  
● 加勒比黑人  
● 亞洲人  
● 歐洲人  
 西班牙人  
 法國人  
 芬蘭人  
 英國人  
 挪威人  
 荷蘭人  
 愛爾蘭人

標準

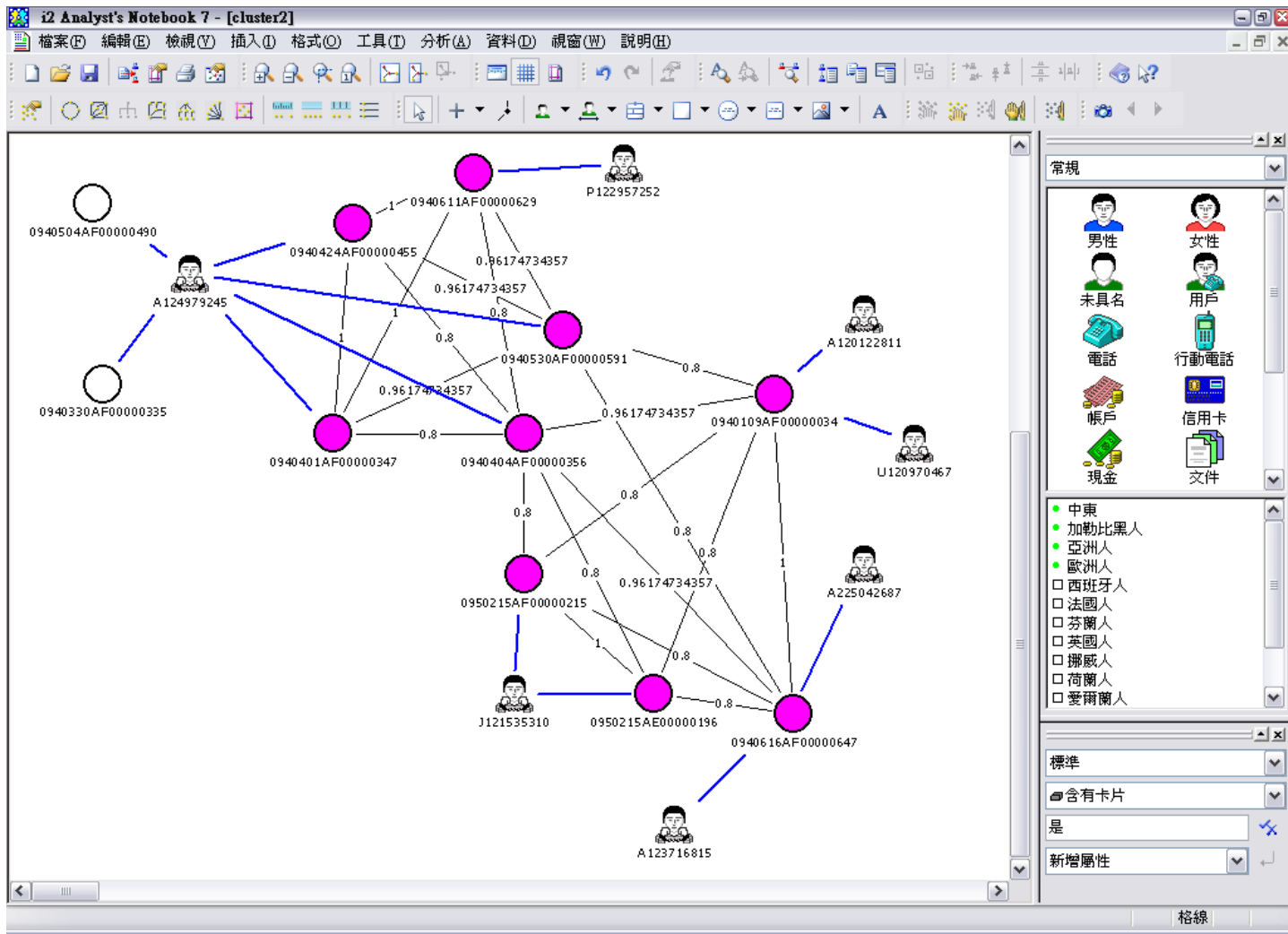
含有卡片

是

新增屬性

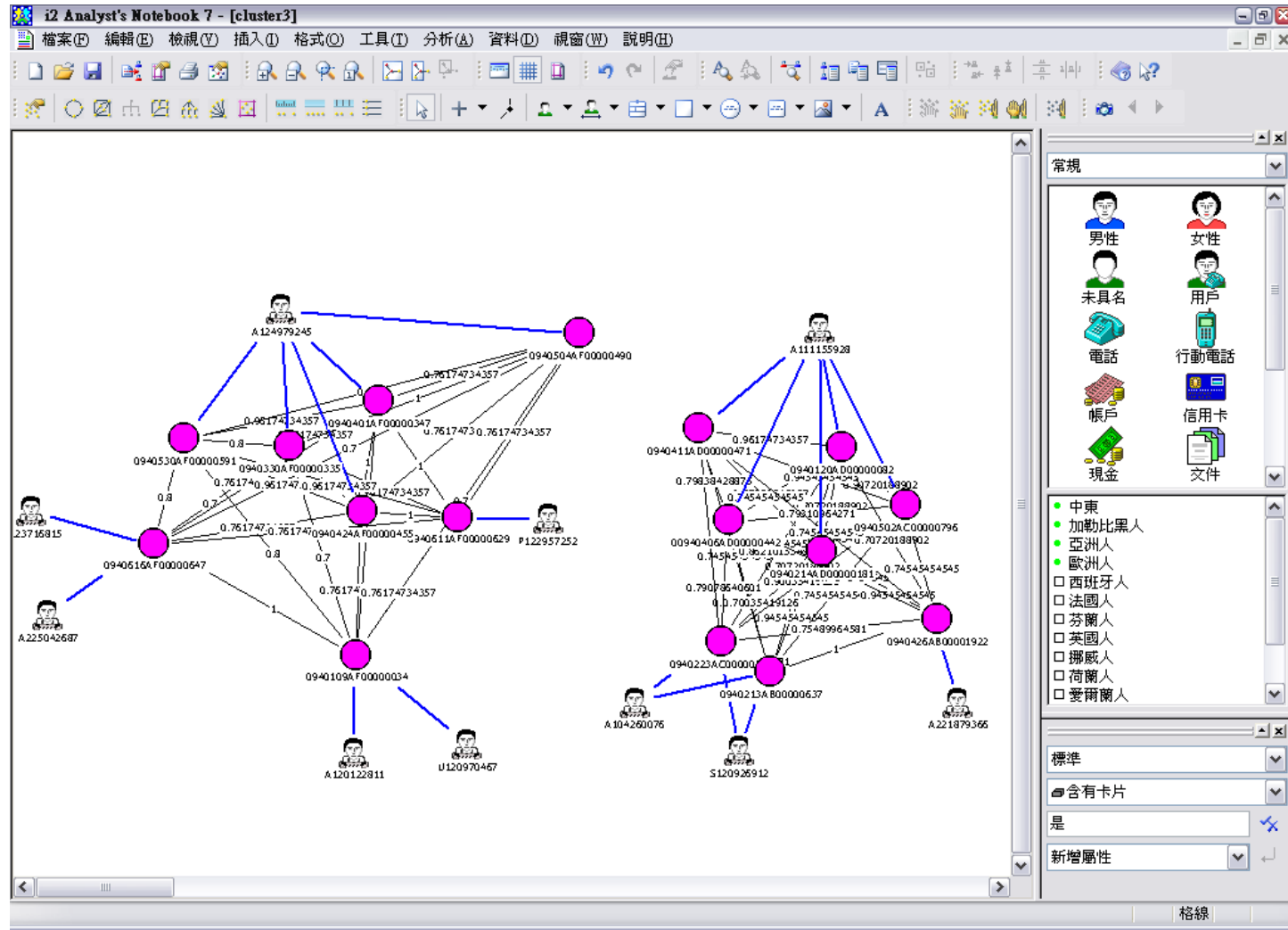


# 分群(Clustering) 與社會網絡分析





# 分群(Clustering) 與社會網絡分析(續)





# Related References

- M. J. Berry, and G. Linoff, *Data Mining Techniques: For Marketing, Sales, and Customer Support*, John Wiley & Sons, 1997.
- 陳仁智，地緣剖繪技術應用於連續街頭強盜搶奪犯罪偵查之研究，中央警察大學刑事警察研究所碩士論文，2004。
- A. Graycar, “Local Government and Crime Prevention” , *Proceeding of the character, Impact and Prevention of Crime in Regional Australia Conference*, Townsville, 2001.
- M. Felson, E. Poulsen, 2003, “Simple indicators of crime by time of day” , *International Journal of Forecasting*, Volume 19, Issue 4
- J. Han, and M. Kamber, *Data Mining: Concepts and Techniques*, Second Edition, Morgan Kaufmann, 2006.
- 黃富源、范國勇、張平吾，*犯罪學概論*，三民書局，2005。
- H. C. Lee, *Crime Scene Investigation*, Central Police University Press, Taoyuan, Taiwan, ROC, 1994.
- M. Palmiotto, “Crime Pattern Analysis: An Investigative Tool” , in *Critical Issues in Criminal Investigation*, 2nd ed., Pilgrimage, 1988.
- Y. T. Peng, and J. H. Wang, “Link Analysis Based on Webpage Co-occurrence Mining – a Case Study on a Notorious Gang Leader in Taiwan” , *Proceedings of the IEEE International Conference on Intelligence and Security Informatics*, 2008.
- C. E. Shannon, and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL, 1949.
- G. Salton, *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison Wesley, 1989.
- B. E. Turvey, *Criminal Profiling*, Third Edition: An Introduction to Behavioral Evidence Analysis, Academic Press; 3 edition (April 30, 2008)
- Jau-Hwang Wang, Bill T. Lin, Ching-Chin Lin, “Application of the Vector Space Model on Criminal Record Retrieval” , *1997國際科學警察學術研討會論文集*，中央警察大學，1997。
- S. Wasserman, and K. Faust, *Social Network Analysis and Applications*, New York and Cambridge, 1994.
- **J.-H. Wang, C.-L. Lin, „ An Association Model for Implicit Crime Link Analysis” , to appear in the Proceedings of PAISI 2010.**



---

# Q & A