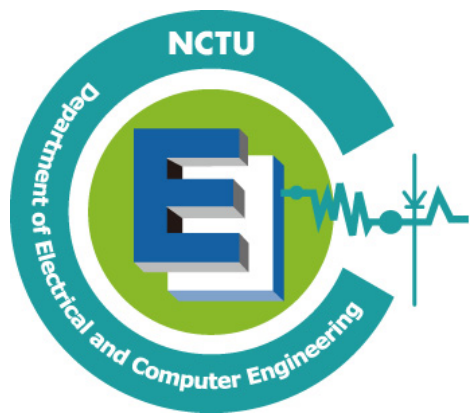


Social Network influence topology construction based on Transaction Data



指導教授：帥宏翰

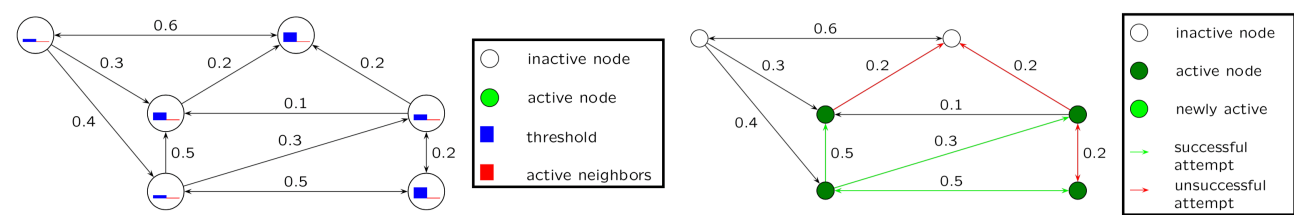
學生：電機07 王斯右

相關理論：

社群網路的影響力模型通常分為兩種，Linear Threshold Model 和 Independent Cascade Model [1]，Linear Threshold Model 一個節點是否動作是由周遭的節點不同比例的影響，若周遭總和的影響力超過一定的threshold 就算成功。而Independent Cascade Model 則是透過結點間的影響機率推移最終分配影響力到整張社群網路。

而結點間的機率也有不同的計算方式 [2]，可藉由單純分析兩節點的互動建立機率，也可以考慮整體周圍的節點綜合而成整體的機率，其中時間也會影響機率，大多以指數曲線衰退，但是在不同狀況下可能會有不同的衰退狀況[3]。

- Linear Threshold
- Independent Cascade



專題設計：

此次專題必須從無到有，透過交易資料建立出社群網路拓撲，而交易資料間本身並不具有明確的社群關係，因此透過相似但時間有先後的交易，來反推出兩個用戶間的影響機率。最後得出的結點間機率則視為Independent Cascade Model 機率推移的最終結果，從而建立整個社群網路拓撲。

未來規劃：

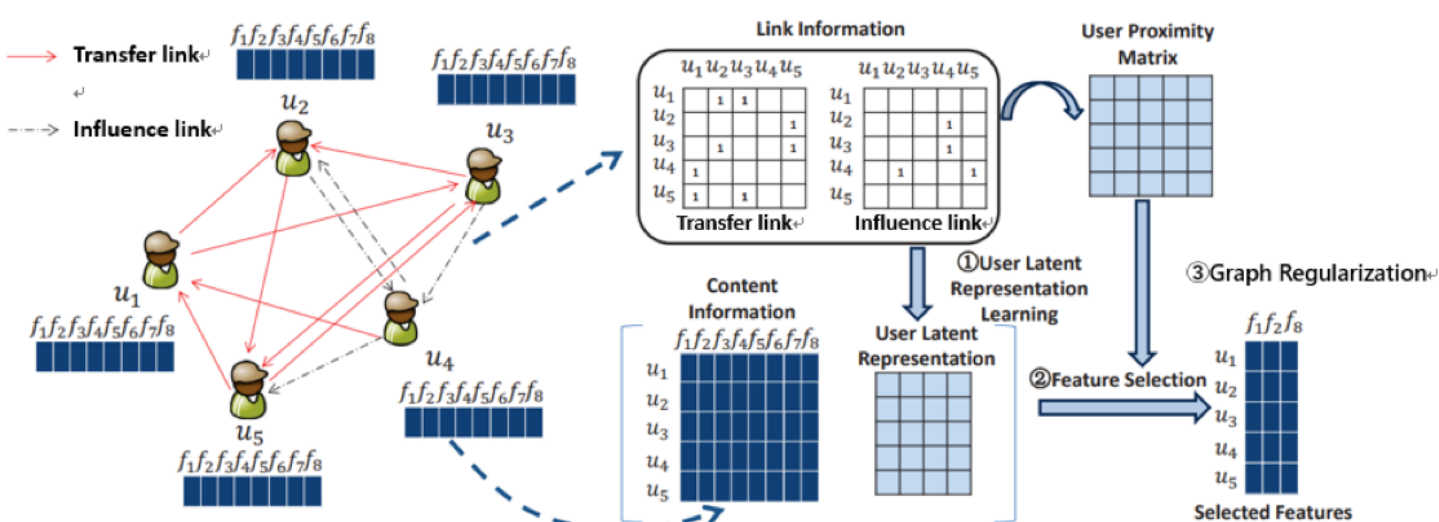
優化拓撲：詳細區分不同種類交易，判斷是否有影響力，以及其影響力經過時間的衰退，計算出更精準的影響力權重。

預測模型建立：結合轉帳、地理資料，建立自動調整特徵權重的模型，應用於使用者行為的預測 [5]。

$$\min_{U, V^T, V^\sigma} \|WX - U\| + \lambda_1 \|A^T - UV^T U^T\| + \lambda_2 \|A^\sigma - UV^\sigma U^T\| + \frac{\lambda_3}{2} \text{tr}(U^T L_c U)$$

其中 $X \in \mathbb{R}^{(N \times D)}$ ，為使用者特徵矩陣， V^T 為由轉帳資料建立的拓撲，

V^σ 則是本次專題的結果，由交易資料建立的影響力拓撲，加上 L_c 地點相似矩陣之Laplace矩陣，結合為共同分解模型，找出最佳的特徵權重。

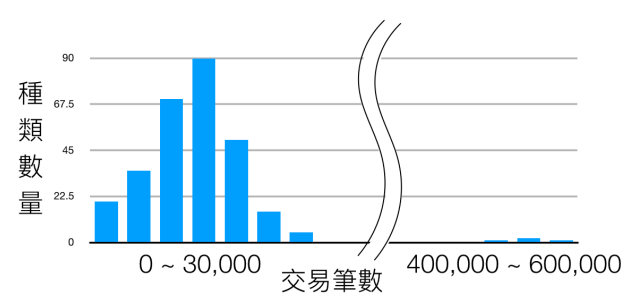


專題摘要：

透過社群網路的分析，可以在實際應用中達到許多目的，例如：行為預測、影響範圍推估。本次專題就透過交易資料建立人與人的連結，由交易行為的互相影響，反推出用戶可能存在相關性的機率，藉此建立社群網路拓撲，作為下一步行為預測的基礎。

專題結果：

篩選資料



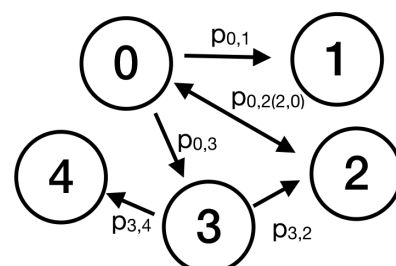
由交易筆數-種類數量的圖可得知，有某些種類有大量的交易，發現其中大多是百貨類、量販店、加油站...無法明確推估影響力，因此先暫時剔除。

交易比較



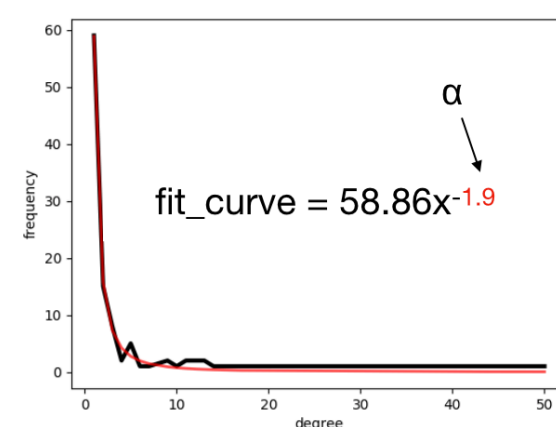
透過單一店家，每個用戶，一定時間區段內用戶先後消費的紀錄，計為成功影響，記錄用戶間影響的次數。

拓撲建立



統計所有用戶交易的次數 A_v ，並透過紀錄確認該用戶影響其他用戶的次數，得到 A_{v2u} ，計算兩者得到用戶間影響機率 $p_{v,u} = A_{v2u} / A_v$ ，推得拓撲連結的機率權重。

驗證拓撲



根據degree-frequency建立折線圖並回歸power law曲線，社群網路中的 α 值通常會在2~3之間 [4]。

困難與解法：

• 複雜度問題

總用戶數約為十萬名，平均每人交易為30筆，若每人每筆都要比對，複雜度為 $N^2 \times n^2$ (N:用戶 n:交易筆數) 無法有效完成。

解決方法是重新以時間作為考量，選擇小區段資料找出可能的影響行為，複雜度降低至 $t(N \times n)$ (t:時間區段)，大幅降低計算量。

• 影響力判定

交易資料本身無法得知交易明細，因此回推的過程需要篩選某些特定類別的交易，如：百貨、量販店、政府服務(稅金)...，排除這些無關交易被誤判為影響力行為，進而建出 power law distribution 的分佈，但是因為仍有許多大型商場(如：Uniqulo, Lativ)、定期繳費(如：kkbox, 電信費)被計入，因此結果的 α 值只有1.9，比一般 Scale-free network 更為密集，若要解決此問題可能需要個案處理特殊的店家。

參考資料：

- [1] D. Kempe, J. M. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In Proc. of the Ninth ACM Int. Conf. on Knowledge Discovery and Data Mining (KDD'03).
- [2] A Goyal, F Bonchi, LVS Lakshmanan. Learning Influence Probabilities In Social Networks. In Proc. of the third ACM int. conf. on Web search and data mining (WSDM '10).
- [3] SA Myers, C Zhu, J Leskovec. Information Diffusion and External Influence in Networks. In Proc. of the 18th ACM SIGKDD int. conf. on Knowledge discovery and data mining (KDD '12).
- [4] T Rastogi. A Power Law Approach to Estimating Fake Social Network Accounts. (2016) arXiv:1605.07984.
- [5] K Cheng, J Li, H Liu. Unsupervised Feature Selection in Signed Social Networks. In Proc. of the 23rd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD '17).