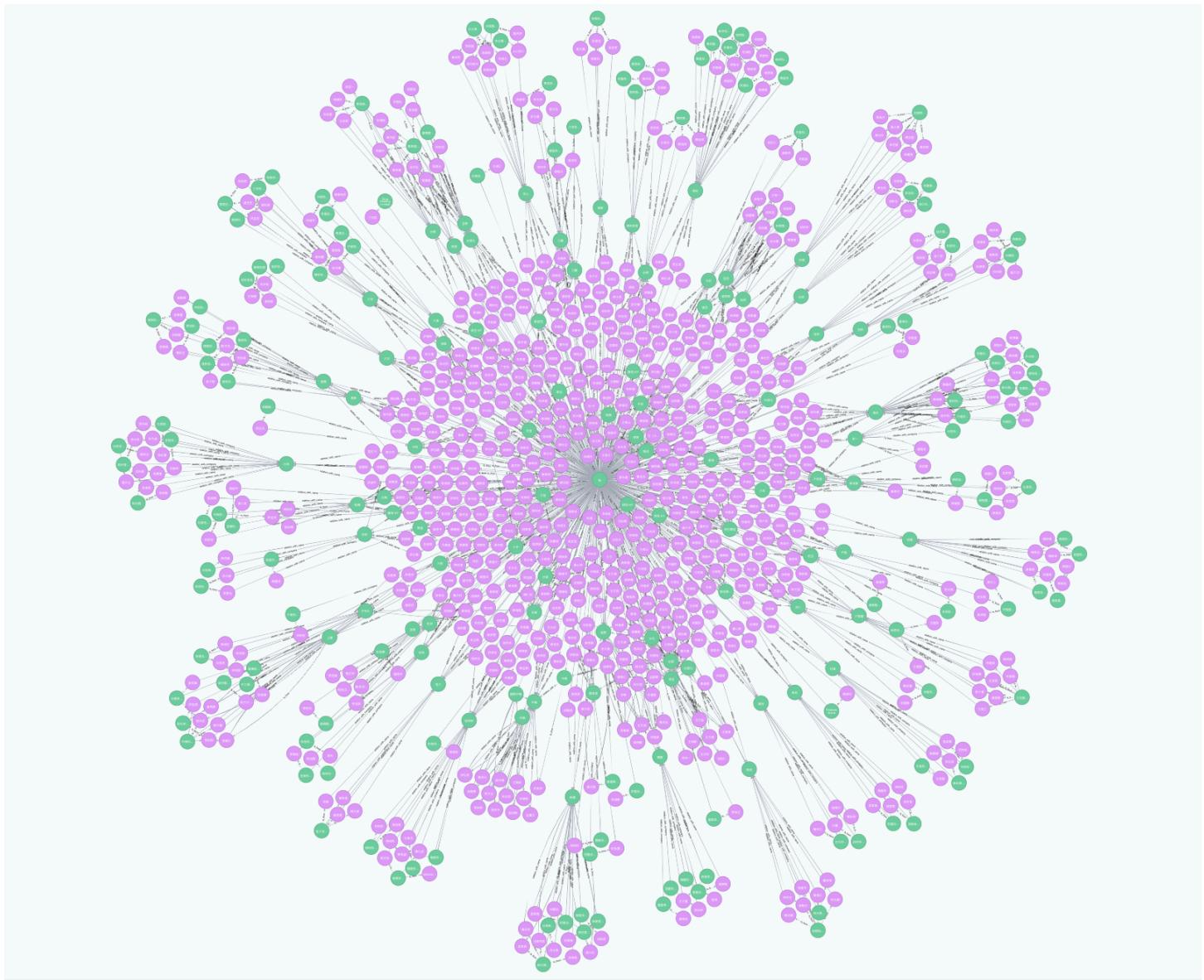


SNA 期末報告

組員：胡弘林、吳衡祐、黃柏嶧

指導老師：胡筱薇



目錄

一、 主題

二、 小組報告簡報連結

三、 資料來源與說明

四、 資料匯入程序

五、 簡易資料探索分析

六、 資料探索中的洞察報告

七、 結論 & 心得

一、主題

台灣上市上櫃公司產業群聚分析

二、小組報告簡報連結

~Youtube連結

<https://www.youtube.com/watch?v=RgWhDTrDknM>

三、資料來源與說明

以2018年台灣的上市上櫃公司董事監察名單為目標對象

資料來源：台灣證券交易所公開資訊觀測站

<http://mops.twse.com.tw/mops/web/t100sb07>

上市上櫃公司編碼來源：台灣證券交易所 - 證券編碼公告

<http://www.twse.com.tw/zh/page/products/stock-code2.html>

我們使用Python的Selenium套件去爬取網站內所有上市上櫃公司的資料，下圖為其中一家上市公司的示例，我們爬取的欄位有公司代號、公司名稱、職稱、姓名(或代表人姓名)、所代表法人姓名

董事及監察人出(列)席董事會情形

公司代號：4737 公司名稱：華廣

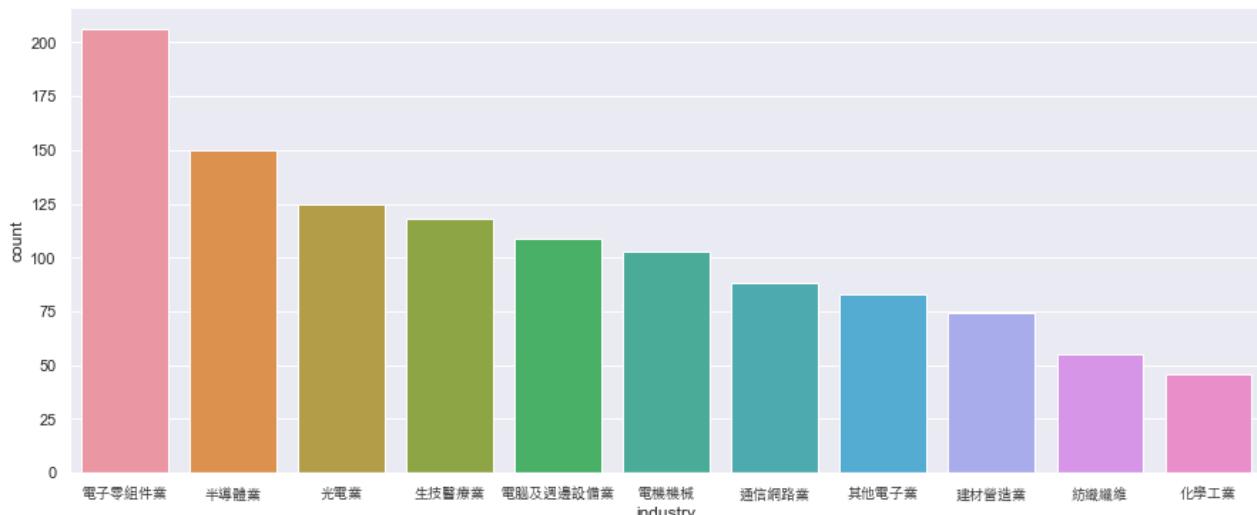
本屆董事會任期：105年02月05日至108年11月07日，迄今已開會18次，董事監察人出列席情形如下：

職稱	姓名(或代表人姓名)	所代表法人姓名	實際出(列)席次數(B)	委託出席次數	應出(列)席次數(A)	實際出(列)席%(B/A)	備註
董事長本人	黃椿木		18	0	18	100	
董事本人	江台雄		18	0	18	100	
董事之法人代表人	林明祥	華榮電線電纜股份有限公司	13	5	18	72	
董事之法人代表人	李聰	通化東寶藥業股份有限公司	17	1	18	94	
監察人本人	戴宗凱		17	0	18	94	
監察人本人	蘇錦彬		12	0	18	67	
監察人本人	林建興		16	0	18	89	
獨立董事本人	蔡淦仁		17	1	18	94	
獨立董事本人	郭莉真		18	0	18	100	
獨立董事本人	呂學裕		15	3	18	83	

資料樣態：

	公司名稱	公司代碼	市場類別	產業類別	職稱	姓名	所代表法人姓名
0	台泥	1101	上市	水泥工業	董事長之法人代表人	張安平	嘉利實業股份有限公司
1	台泥	1101	上市	水泥工業	董事之法人代表人	余俊彥	中成開發投資股份有限公司
2	台泥	1101	上市	水泥工業	董事之法人代表人	辜公怡	泰和興業股份有限公司
3	台泥	1101	上市	水泥工業	董事之法人代表人	李鐘培	財團法人辜公亮文教基金會
4	台泥	1101	上市	水泥工業	董事之法人代表人	王伯元	富品投資股份有限公司

最後一共爬取16,600筆資料，其中我們發現一共有928上市公司以及766家上櫃公司，而產業類別的分佈圖如下，以電子零組件業、半導體業等科技業為大宗。



四、資料匯入程序

我們將16,600筆資料匯入Neo4j資料庫，匯入語法如下：

```
1 USING PERIODIC COMMIT 1000
2 LOAD CSV WITH HEADERS FROM "file:///董事監名單_clean2.csv" AS row
3
4 MERGE (c:Company {Name:row.公司名稱, Type: row.產業類別})
5 MERGE (a:Actor {Name: row.姓名, Character: row.職稱})
6 MERGE (r:Company {Name: row.所代表法人姓名})
7
8 MERGE (c)<-[:relation_with_name]-(a)
9 MERGE (c)<-[:relation_with_company]-(r)
10 MERGE (a)<-[ :is_from ]-(r);
```

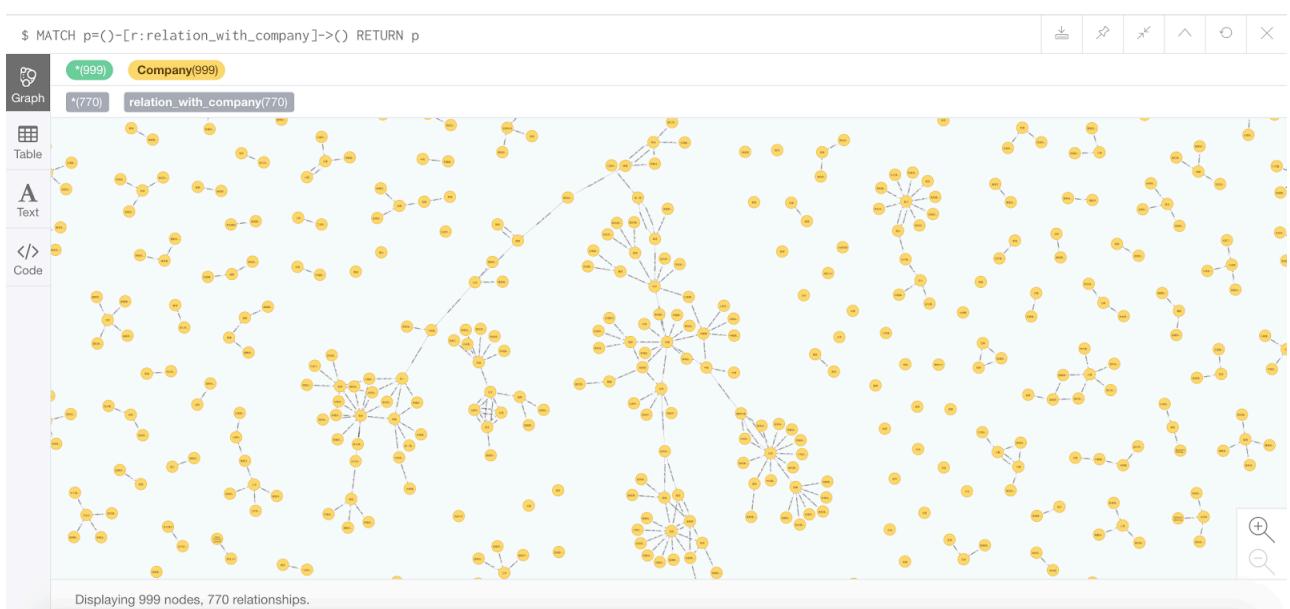
五、簡易資料探索分析

前提假設：

→ 不論是董事法人之代表人或是監察人，我們假設他所代表的公司就是與該家公司有所關聯

1. 查看資料中的關聯性

我們透過Cypher語法將資料匯入Neo4j資料庫後，首先查看全部資料中所有公司與其代表公司間的關聯，實際跑出的圖如下



我們發現在所有公司間的關聯中，出現許多孤立點亦或是只有兩三間公司相連的關聯圖，不過當中卻有一群關係錯綜複雜的網絡圖出現在這之中，因此我們想進一步分析這群網絡。

2. 篩選出最大 Group 的企業進行進一步的分析。



鋼鐵工業	水泥產業	化學工業	油電燃氣業	紡織纖維業	金融保險業
中鋼	亞泥	東聯	大台北	新纖	王道銀行
中鋼構	東泥	中碳	欣天然	新紡	遠東銀
中鴻	台泥	國炭科	新海	遠東新	華票
春雨	嘉泥	信昌化			
春源					
第一銅					
唐榮					
其他業	航運業	電器電纜業	電機機械	通信網路業	生技醫療業
中聯資源	裕民	聲寶	瑞智	遠傳	華廣
中鼎	台船	華榮	中宇	亞太電	展旺
崑鼎					
光電業	半導體業	橡膠工業	貿易百貨業	資訊服務業	
友輝	合晶	國際中橡	遠百	新鼎	

我們發現這群網絡以水泥、鋼鐵、化學產業為主，一共有533筆資料，42家公司，因此我們將此份關聯圖從所有資料中獨立出來（去除其他孤立點和subgroup），後續的指標以及演算法我們將會使用Python的NetworkX套件進行分析。

3. 進行簡易指標計算

(a) 整張網絡圖的平均最短距離

```
nx.average_shortest_path_length(G)
```

```
5.2383498854087085
```

(b) 直徑與半徑

```
nx.diameter(G)
```

```
10
```

```
nx.radius(G)
```

```
5
```

(c) Degree數量前10家公司

```
sorted(G.degree, key=lambda x: x[1], reverse=True)[0:10]
```

```
[('台泥', 16),  
 ('亞泥', 14),  
 ('中鋼', 14),  
 ('大台北', 14),  
 ('新纖', 9),  
 ('王道銀行', 8),  
 ('遠東新', 8),  
 ('裕民', 8),  
 ('中宇', 8),  
 ('華榮', 8)]
```

(d) In degree數量前10家公司

```
sorted(G.in_degree, key=lambda x: x[1], reverse=True)[0:10]
```

```
[('中鋼', 8),  
 ('亞泥', 6),  
 ('華榮', 6),  
 ('台泥', 4),  
 ('遠東新', 4),  
 ('裕鼎實業股份有限公司', 4),  
 ('新纖', 3),  
 ('財團法人新光吳火獅文教基金會', 3),  
 ('中鼎', 3),  
 ('經濟部', 3)]
```

(e) Out degree數量前10家公司

```
sorted(G.out_degree, key=lambda x: x[1], reverse=True)[0:10]
```

[('台泥', 12),
 ('大台北', 12),
 ('亞泥', 8),
 ('王道銀行', 7),
 ('中聯資源', 7),
 ('欣天然', 7),
 ('中宇', 7),
 ('新纖', 6),
 ('中鋼構', 6),
 ('中鋼', 6)]

(f) 中心點

```
nx.center(G)
```

['中聯資源']

→ 中聯資源為中鋼集團的成員之一，主要專注於中鋼製程副產物爐石用途創新及加值化工作，領域跨足水泥、鋼鐵、化學、環保等產業，因此確實有可能位於其網絡的中心點位置

(g) 邊緣點，最大的Eccentricity

```
nx.periphery(G)[:10]
```

['欣田崎生化科技股份有限公司',
 '國軍退除役官兵輔導委員會',
 '欣欣客運股份有限公司',
 '通產企業股份有限公司',
 '新海瓦斯股份有限公司',
 '朋萊股份有限公司',
 '工興實業股份有限公司',
 '光偉投資股份有限公司',
 '欣淼投資股份有限公司',
 '濟真股份有限公司']

(h) 關節點

```
nx.minimum_node_cut(G)
```

{'中鋼'}

(i) 關聯邊 (Bridge)

```
nx.minimum_edge_cut(G)
{('中鋼', '高瑞投資股份有限公司')}
```

→ 中鋼這家公司在這個網絡中扮演聯繫不同群之間的角色，因為若將他去除此網絡將無法成型，但從另一方面來看在這群網絡中有許多的公司為中鋼集團旗下的子公司，因此我們也可以猜測中鋼集團在這個產業鍊當中同時也扮演著龍頭的角色。

4. Centrality

Centrality這個指標可以用來衡量它在整個網絡中的結構重要性

(a) Degree Centrality

```
degCent = nx.degree_centrality(G)
sorted(degCent.items(), key=lambda x: x[1], reverse=True)[:5]
[('台泥', 0.09803921568627451),
 ('中鋼', 0.09150326797385622),
 ('大台北', 0.09150326797385622),
 ('亞泥', 0.08496732026143791),
 ('新纖', 0.058823529411764705)]
```

(b) Close Centrality

```
closeCent = nx.closeness_centrality(G, wf_improved=False)
sorted(closeCent.items(), key=lambda x: x[1], reverse=True)[:5]
[('中聯資源', 0.3227848101265823),
 ('中鋼', 0.30297029702970296),
 ('台泥', 0.2953667953667954),
 ('中碳', 0.2781818181818182),
 ('中鋼構', 0.2756756756756757)]
```

(c) Betweenness Centrality

```
btwnCent = nx.betweenness_centrality(G)
sorted(btwnCent.items(), key=lambda x: x[1], reverse=True)[:5]

[('中聯資源', 0.6111397775484461),
 ('台泥', 0.5233058135534916),
 ('中鋼', 0.5145625501662652),
 ('中宇', 0.35178018575851394),
 ('新纖', 0.3394392844857241)]
```

```
1 btwnCent_edge = nx.edge_betweenness_centrality(G)
2 sorted(btwnCent_edge.items(), key=lambda x: x[1], reverse=True)[:5]
```

```
[(['台泥', '中聯資源'], 0.4334521687462864),
 (['中聯資源', '中鋼'], 0.37387884446707986),
 (['台泥', '新纖'], 0.3313810372633902),
 (['中鋼', '中宇'], 0.3242084712672948),
 (['中聯資源', '亞泥'], 0.2398353280706223)]
```

(d) Edge Betweenness Centrality

小結：

→ 我們使用了課堂中老師所教的不同種方式來計算每個點的centrality，可以發現主要都是中鋼、中聯資源、台泥等大公司，如上一段所述中心點、關節點相同，這些點在整個網絡中佔有重要的地位。

六、資料探索中的洞察報告

Link prediction

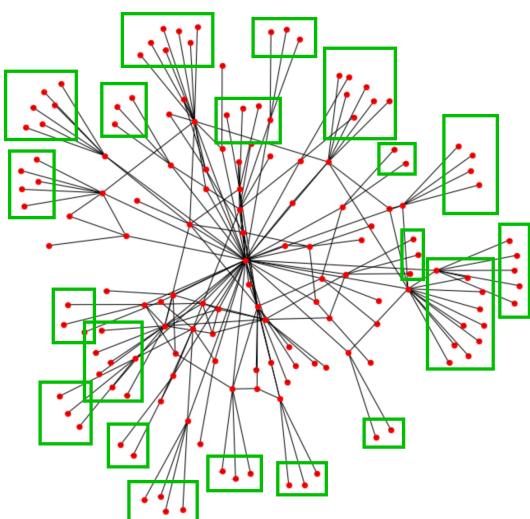
1. Common neighbors

```
import operator
common_neigh = [(e[0], e[1], len(list(nx.common_neighbors(G, e[0], e[1]))))
sorted(common_neigh, key=operator.itemgetter(2), reverse=True)[:5]
[('遠百', '東聯', 3),
 ('富品投資股份有限公司', '中信投資股份有限公司', 2),
 ('遠東新', '裕利投資股份有限公司', 2),
 ('台船', '唐榮', 2),
 ('中碳', '台泥', 2)]
```

2. Jaccard_coefficient

```
L = list(nx.jaccard_coefficient(G))
L.sort(key=operator.itemgetter(2), reverse = True)
L[:10]
[('財團法人東南文化基金會', '東樹投資股份有限公司', 1.0),
 ('財團法人東南文化基金會', '東樹投資(股)公司', 1.0),
 ('財團法人東南文化基金會', '財團法人陳趙樹公益慈善基金會', 1.0),
 ('財團法人東南文化基金會', '長青(股)公司', 1.0),
 ('財團法人東南文化基金會', '立凱投資(股)公司', 1.0),
 ('台灣新光實業股份有限公司', '友輝', 1.0),
 ('台灣新光實業股份有限公司', '進賢投資股份有限公司', 1.0),
 ('台灣新光實業股份有限公司', '德岳實業股份有限公司', 1.0),
 ('台灣新光實業股份有限公司', '新光開發股份有限公司', 1.0),
 ('台灣新光實業股份有限公司', '綿豪實業股份有限公司', 1.0)]
```

→ 用此方法的時候會發現，有相當多的1，因為有許多的公司是只有跟一家公司有關聯的(下圖)，這樣的話用jaccard會讓他們很有可能連再一起。



3. Resource_allocation

```
L = list(nx.resource_allocation_index(G))
L.sort(key=operator.itemgetter(2), reverse=True)
L[:5]

[('遠百', '東聯', 0.5769230769230769),
 ('中宇', '中鋼構', 0.5714285714285714),
 ('中鼎', '柏惠投資股份有限公司', 0.5),
 ('三菱伸銅株式會社', '華榮', 0.5),
 ('嘉新國際股份有限公司', '台泥', 0.5)]
```

4. Adamic_adar_index

```
L = list(nx.adamic_adar_index(G))
L.sort(key=operator.itemgetter(2), reverse=True)
L[:5]

[('遠百', '東聯', 1.8582210984293646),
 ('中宇', '中鋼構', 1.8216182225789146),
 ('中鼎', '柏惠投資股份有限公司', 1.4426950408889634),
 ('三菱伸銅株式會社', '華榮', 1.4426950408889634),
 ('嘉新國際股份有限公司', '台泥', 1.4426950408889634)]
```

5. Preferential_attachment

```
L = list(nx.preferential_attachment(G))
L.sort(key=operator.itemgetter(2), reverse=True)
L[:5]

[('台泥', '大台北', 210),
 ('台泥', '中鋼', 210),
 ('中鋼', '大台北', 196),
 ('亞泥', '台泥', 195),
 ('亞泥', '中鋼', 182)]
```

小結：

→ 由Common neighbors中可以看到最有可能的連結為('遠百','東聯')，也就是共同有關聯公司最多的，不過他們共同有關連的公司也只有三間，可以看出在這個網絡之中不是很密集的，主要都是大公司們互相透過中聯資源所連接後，旗下的小公司連著大公司。

七、結論 & 心得

可以從以上的結果中看出，中聯資源在這整個產業鍊的重要性，像是台泥、中鋼也都派董事去督促這個公司，整個圖形就像是家族族譜。

對於link prediction結果感覺不適用這種公司的網絡，公司跟公司的關聯主要已經是有一套制度存在了，大公司連著中聯資源，小公司連著大公司，不像是人際的網絡，比較活的情況。

台泥、中鋼本身也沒有直接的相連，整個網絡也很少看到兩家公司互相派董事的情形，我認為這是個很好的現象，整個網絡都是有階層式的，像是我看到台塑的那區的網絡，就覺得怪怪的，感覺有在做不法勾當的味道飄出來。網絡分析真的方便，由圖像就可以輕鬆的看到很多資訊。

