

How to draw the Sankey diagram and forest plot

A. The Sankey diagram

1. Step 1: getting data from the link

Download 100-top cited articles for the two journals in this study at

<http://www.healthup.org.tw/html100/spine2journals.htm>

	A	B	C	D	E	F	G	H	I
1	Citation		Year	Country	Category	Journal	Type	PMID	
2	374		2011	U.S	F5.Physiology	Spine J	Journal Article	21729796	374
3	213		2012	U.S	F1.Pain&Prognosis	Spine (Phila Pa 1976)	Journal Article	21311399	213
4	172		2012	U.S	F2.Statist &Data	Spine (Phila Pa 1976)	Comparative Study	22045006	172
5	163		2012	U.S	F1.Pain&Prognosis	Spine (Phila Pa 1976)	Journal Article	22146287	163
6	138		2013	U.S	F5.Physiology	Spine (Phila Pa 1976)	Evaluation Study	23970107	138
7	135		2013	U.S	F2.Statist &Data	Spine (Phila Pa 1976)	Journal Article	23722572	135
8	120		2011	U.S	F1.Pain&Prognosis	Spine (Phila Pa 1976)	Journal Article	21952190	120

2. Step 2: SNA approach to get the element counts and relations between elements

	A	B	C	D	E	F	G	H
1					1			
2	C1	2011	49	21729796	1		2011	U.S
3	C1	2013	42	21311399	1		U.S	F5.Physiol
4	C1	2012	40	22045006	1		F5.Physiol	Spine J
5	C1	2014	34	22146287	1		Spine J	Journal Art
6	C1	2015	24	23970107	1		Journal Art	21729796
7	C1	2016	8	23722572		citation	2012	U.S
8	C1	2017	3	21952190			U.S	F1.Pain&P
9	C2	U.S	107	21192221			F1.Pain&P	Spine (Phil
10	C2	Japan	11	25839387		Count only	Spine (Phil	Journal Art
11	C2	South Korea	11	23537454			Journal Art	21311399
12	C2	Netherland	9	24113358			U.S	F2.Statist&
13	C2	China	8	26208232	1		F2.Statist&	Spine (Phil
14	C2	Canada	7	24412416		CSTR(number)	Spine (Phil	Comparativ
15	C2	U.K	7	23369494			Comparativ	22045006
16	C2	Australia	7	24239490	1		Journal Art	22146287
17	C2	Germany	4	23830297			2013	U.S
18	C2	Sweden	3	23369495		x-index Top 5	F5.Physiol	Spine (Phil
19	C3	F2.Statist&	74	24412032			Spine (Phil	Evaluation
20	C3	F5.Physiol	46	21192288			Evaluation	23970107

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
4	F5.Physiol	Spine J	12	#000000	C3					C3	0.77	C1	2012	40	#ffcc00	C1
5	Spine J	Journal Art	76	31	#000000	C4						C1	2014	34	#80ff00	C1
6	Journal Art	21729796	1	#ff0000	C5							C1	2015	24	#993300	C1
7	2012	U.S	7	#000000	C1							C1	2016	8	#A2B2E2	C1
8	U.S	F1.Pain&P	2	#000000	C2							C1	2017	3	#5e8437	C1
9	F1.Pain&P	Spine (Phil	6	#000000	C3						AAC	C2	U.S	107	#8A2BE2	C2
10	Spine (Phil	Journal Art	73	#ff0000	C4	To generate Unknowns edge						C2	Japan	11	#DEB887	C2
11	Journal Art	21311399	1	#000000	C5							C2	South Korea	11	#00ff00	C2
12	U.S	F2.Statist&	6	#ff0000	C1							C2	Netherland	9	#FBF5EF	C2
13	F2.Statist&	Spine (Phil	29	#ff0000	C1	1 Unknown	1 Known flow					C2	China	8	#0000ff	C2
14	Spine (Phil	Comparativ	5	#000000	C4							C2	Canada	7	#A2B2E2	C2
15	Comparativ	22045006	1	#000000	C7							C2	U.K	7	#A52A2A	C2
16	Journal Art	22146287	1	#000000	C2	2 Check	Edge color					C2	Australia	7	#DEB887	C2
17	2013	U.S	11	#000000	C1							C2	Germany	4	#5F9EAD	C2
18	F5.Physiol	Spine (Phil	16	#000000	C3							C2	Sweden	3	#7FFF00	C2
19	Spine (Phil	Evaluation	1	#000000	C1	3 Goto DD2	Chisquare c					C3	F2.Statist&	74	#D2691E	C3
20	Evaluation	23970107	1	#000000	C1							C3	F5.Physiol	46	#FF7F50	C3
21	Journal Art	23722572	1	#000000	C5			1.00				C3	F3.Spine&P	33	#6495ED	C3
22	Journal Art	21952190	1	#000000	C5			1.00	23537454			C3	F1.Pain&P	29	#FFB6C1	C3
23	Journal Art	21192221	1	#000000	C5			1.00				C3	F4.Physio	18	#DC143C	C3
24	2015	Canada	1	#000000	C1							C4	Spine J	100	#00FFFF	C4
25	Canada	F5.Physiol	1.16	#000000	C2							C4	Spine (Phil	100	#00008B	C4
26	Journal Art	25839387	1	#000000	C5							C5	Journal Art	160	#008B8B	C5
27	F1.Pain&P	Spine J	18	#000000	C3							C5	Comparativ	22	#B8860B	C5
28	Journal Art	23537454	1	#000000	C5							C5	Clinical Tr	7	#A9A9A9	C5
29	2015	U.S	11	#000000	C1							C5	Evaluation	4	#A9A9A9	C5

Excel module to draw the Sankey

To arrange data as the form below:

- Step 3: to draw the Sankey diagram

	A	B	C
1			
2	. [0.0001] C0		
3	C0 [0.0001] C1		
4	. [49] 2011		
5	. [42] 2013		
6	. [40] 2012		
7	. [34] 2014		
8	. [24] 2015		
9	. [8] 2016		
10	. [3] 2017		
11	2013 [23] F2.Statist&Date		
12	2012 [17] F2.Statist&Date		
13	2014 [9] F2.Statist&Date		
14	2015 [12] F2.Statist&Date		
15	2016 [4] F2.Statist&Date		
16	2017 [2] F2.Statist&Date		

Codes generated to fit the requirement of Sankey format at

<https://sankeymatic.com/>

- Step 4: Copy and Paste them onto the Sankey maker

<https://sankeymatic.com/build/>

Inputs:

- U.S. #5e8437
- Australia #8A2BE2
- Germany #0EB887
- Sweden #00F80
- F2.Statist&Data #009933
- F5.Physiology #fcc00
- F3.Spine&Surgery #f0086
- F1.Pain&Prognosis #80ff00
- F4.Phiopathology #993300
- Spine J #009933
- Spine (Phila Pa 1978) #fcc00
- Journal Article #009933
- Comparative Study #fcc00
- Clinical Trial #f0086
- Evaluation Study #80ff00
- Case Reports #993300
- Guideline #8A2BE2
- Congress #5e8437
- Comment #8A2BE2

Visit the [Manual](#) for a detailed guide

Nodes:

Width: 39 Spacing: 0 Max

Border: 0 Opacity: 0 1

Default Node Colors:

Use one color: [] or a Theme:

- Categories
- Tableau10
- Dark
- Varied

Flows:

Opacity: 0 1 Curviness: 1 ()

Default Flow Colors:

Use one color: [] or colors from:

- each flow's Source
- each flow's Target
- the outermost nodes (flowing in)

Diagram Size & Background:

Width: 2600 Height: 1800 Background Color: [] Transparent

Margins: Left 12 Right 12 Top 18 Bot 20

Move Nodes by dragging. Double-click a Node to reset, or: [Reset all moved Nodes](#)

"U.S [0] F5.Physiology #000000": Amounts must be greater than 0.

"2011 [0] Sweden #000000": Amounts must be greater than 0.

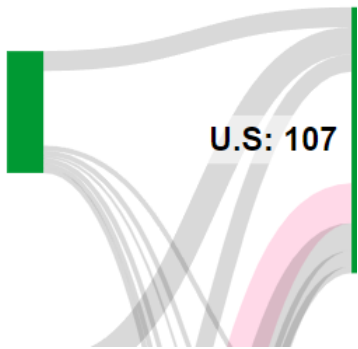
"2016 [0] China #000000": Amounts must be greater than 0.

"Spine J [0] C5 #ffffff": Amounts must be greater than 0.

122.04

2011: 49

U.S: 107



Download it

Prefix = [] Suffix = []

Number Format: 1,000,000.00

Layout Options:

Note: these options only apply to diagrams with 3 or more columns of nodes:

Place all flow origins at the left edge Place all flow endpoints at the right edge

Reverse the graph (flow right-to-left)

Diagram Scale = 0.6981 per pixel (200.0001/286.5px)

For fair comparisons between diagrams:

1) Use the same units for each, and 2) Make their Diagram Scales match as closely as possible.

Export Diagram:

Once you are satisfied with your diagram, you can export it as an image or as vector code (SVG).

PNG Image:

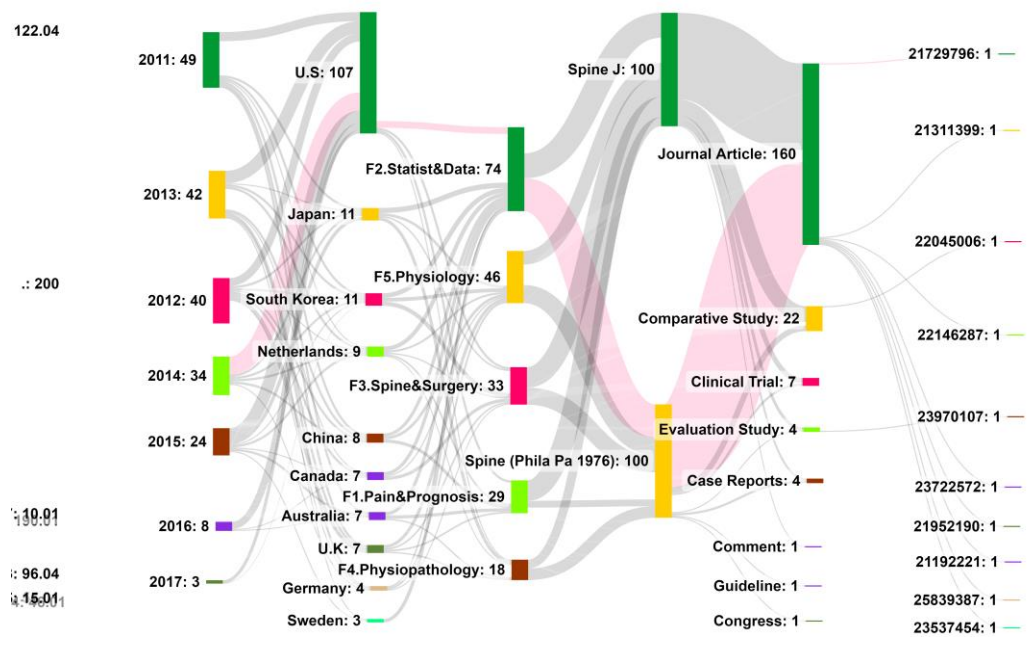
Size: 2x - Medium

[Export 3200 x 1200 PNG](#)

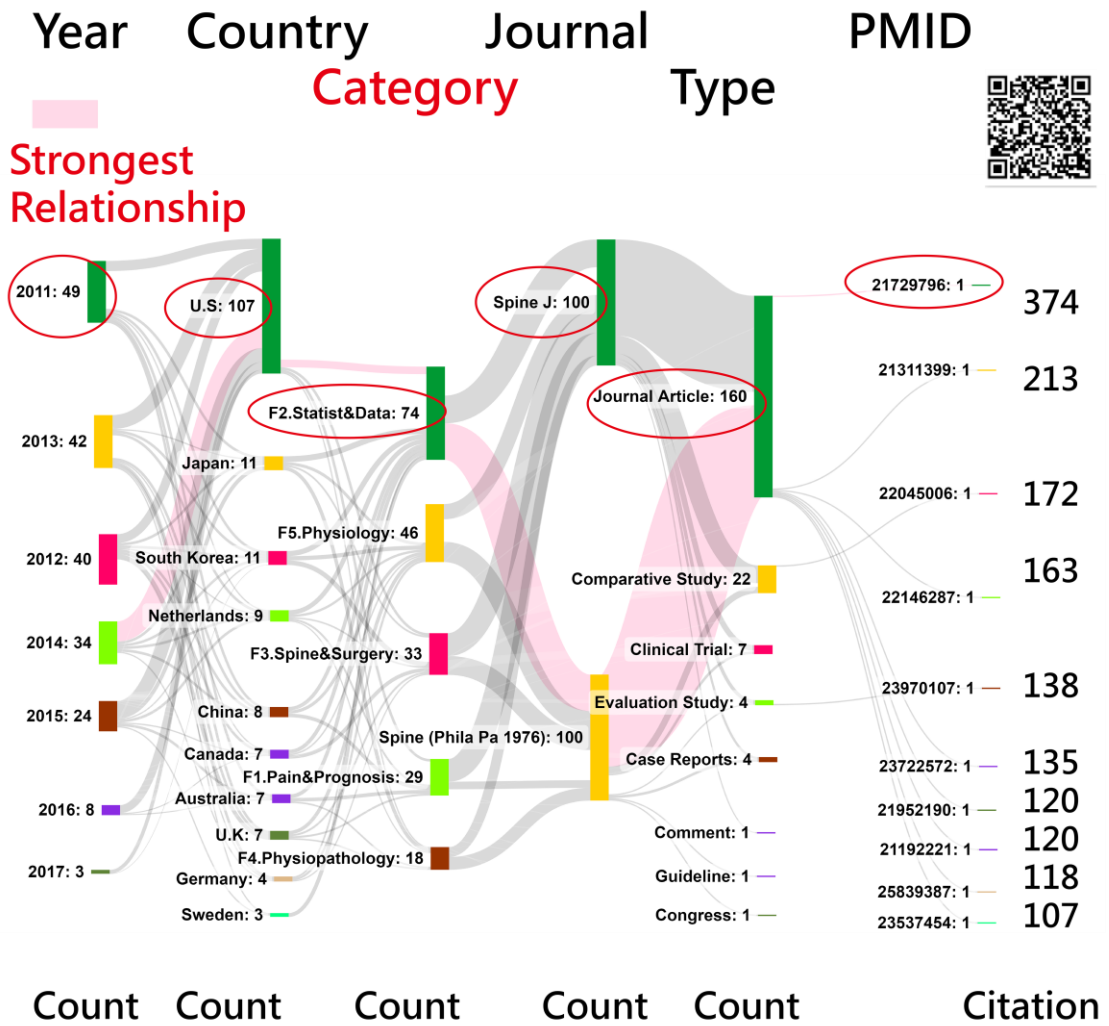
Dimensions for IMG tag:

```
<img width="1600" height="600" ... />
```

SVG Code...



Top 10 dominant elements in entities



If data were pasted on the website at <http://www.healthup.org.tw/kpiall/forestplot.asp>, the Sankey can be drawn in this way below:

responses(rows for person and columns for items)

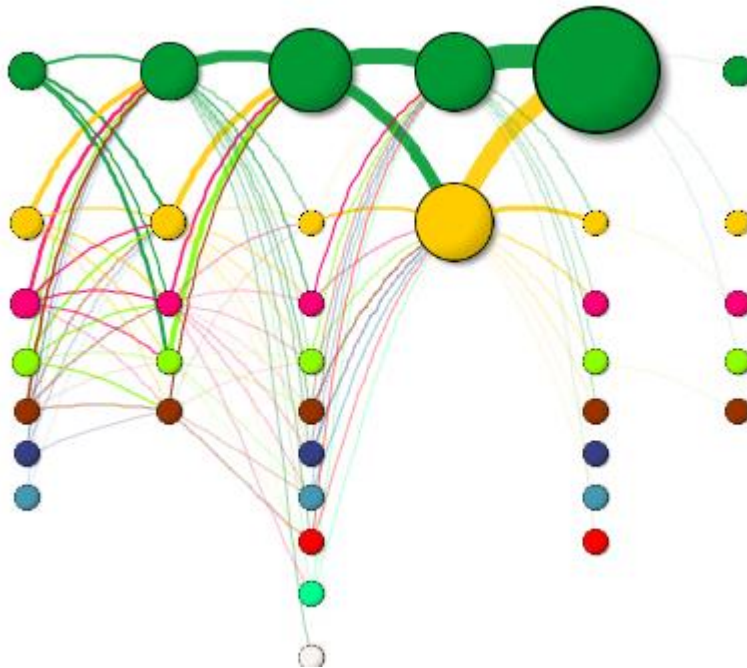
```
11 2011 49
11 2013 42
11 2012 40
11 2014 34
11 2015 24
11 2016 8
11 2017 3
13 F2.Statist&Date 74
13 F5.Physiology 46
13 F3.Spine&Surgery 33
13 F1.Pain&Prognosis 29
13 F4.Phiopathology 18
15 U.S 107
15 Japan 11
15 South Korea 11
15 Netherlands 9
```

For the format from differcne sources, see below examples...

```
2011 F5.Physiology 26.00000
F5.Physiology U.S 50.00000
U.S Spine J 122.00000
Spine J Journal Article 180.00000
Journal Article 21729796 2.00000
2012 F1.Pain&Prognosis 12.00000
F1.Pain&Prognosis U.S 44.00000
U.S Spine (Phila Pa 1976) 92.00000
Spine (Phila Pa 1976) Journal Article 140.00000
Journal Article 21311399 2.00000
2012 F2.Statist&Date 30.00000
F2.Statist&Date U.S 82.00000
Spine (Phila Pa 1976) Comparative Study 38.00000
Comparative Study 22045006 2.00000
Journal Article 22146287 2.00000
2013 F5.Physiology 16.00000
```

Copy from MS Excel and Paste them onto the boxes: one for data, another for sample size, respectively. The first plot immediately appears on Google Maps.

Forest Source



B. The Forest plot <http://www.healthup.org.tw/kpiall/forestplot.asp>

← → ↻ 🏠 ⚠ 不安全 | healthup.org.tw/kpiall/forestplot.asp

Item01	94	22	92	20
item02	98	21	92	22
Item03	98	28	88	26
Item04	94	19	82	17
Item05	98	21	88	22
Item06	96	21	92	22

For the format from differcne sources, see below examples...

60 60
65 65
40 40
200 200
50 45
85 85

Copy from MS Excel and Paste them onto the boxes: one for data, another for sample size, respectively
Forest plot immediately appears on Google Maps.

Forest Source

Clear

Data1:

Item01	94	22	92	20
item02	98	21	92	22
Item03	98	28	88	26
Item04	94	19	82	17
Item05	98	21	88	22
Item06	96	21	92	22

Data2 for sample:

60 60
65 65
40 40
200 200
50 45
85 85

Data were pasted onto the website.

the first plot immediately appears on Google Maps.

Forest Source

Examples for use in Forest plot

SE adjustment(9:

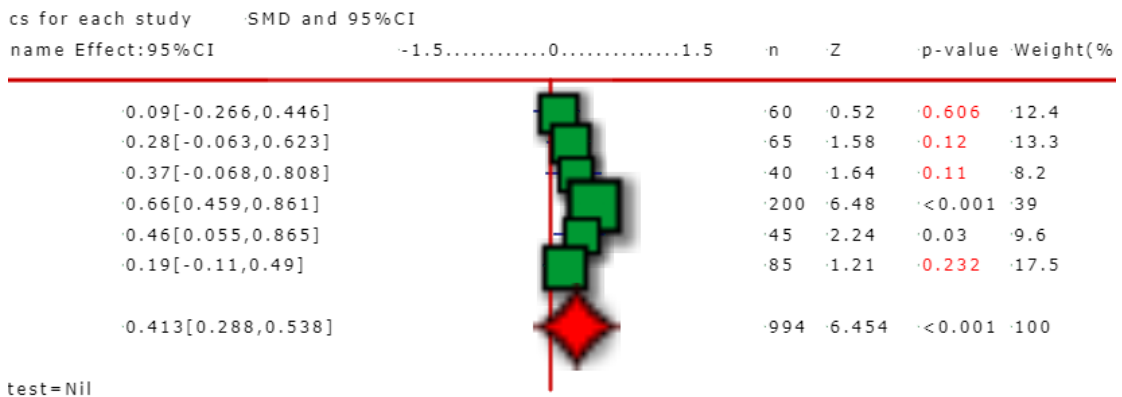
Item01,0.09,0.18,-0.266,0.446,60,0.52,0.606,12.43
 item02,0.28,0.18,-0.063,0.623,65,1.58,0.12,13.34
 Item03,0.37,0.22,-0.068,0.808,40,1.64,0.11,8.21
 Item04,0.66,0.1,0.459,0.861,200,6.48,0.001,38.95
 Item05,0.46,0.21,0.055,0.865,45,2.24,0.03,9.6
 Item06,0.19,0.15,-0.11,0.49,85,1.21,0.232,17.48
 Overall,0.413,0.06,0.288,0.538,994,6.454,0.001,100

Group if necessary from 1 to n at least 5 observed number for each group)

EffectSize: Scale(>0;eg 0.9 or 1) Toward the right(>0) Mul

Extended to Two sides(<>0) Extreme=

To confirm the parameters as shown below and click on the submit icon.



Appendices:

Word cloud tutorial in Excel

<https://help.xlstat.com/6698-word-cloud-tutorial-excel>

Latent Semantics Analysis (LSA) in Excel tutorial

<https://help.xlstat.com/6470-latent-semantic-analysis-lsa-excel-tutorial>

k-means clustering in Excel tutorial

Text Mining⁵

- [Latent Semantics Analysis \(LSA\) in Excel tutorial](#)
- [Cleaning text data in Excel tutorial](#)
- [Convert text data to lower or upper case in Excel](#)
- [Word cloud tutorial in Excel](#)
- [Feature extraction tutorial in Excel](#)
 - <https://help.xlstat.com/tutorial-guides/text-mining>

Feature extraction tutorial in Excel

- <https://help.xlstat.com/6751-feature-extraction-tutorial-excel>

	A	B	C	D	E	F	G	H
1	PMID	surgeri	method	effect	advers	patholog	imag	diagnost
2	26689395	1	0	0	0	0	0	0
3	26555839	2	2	0	0	0	0	0
4	26335676	2	0	0	0	0	0	0
5	26208232	0	0	0	0	0	0	0
6	26020847	3	0	0	0	0	0	0
7	25955086	4	2	0	0	0	0	0
8	25893353	0	0	0	0	2	1	0
9	25839387	0	0	0	0	0	0	0
10	25394317	3	2	0	0	0	0	0
11	24480958	4	0	0	0	0	0	0
12	24270931	0	0	0	0	0	4	4
13	24253796	0	0	0	0	0	0	0

BF	BG	BH	BI	BJ	BK	BL
view	class	F1	F2	F3	F4	F5
0	2	-0.32101	0.03681	-0.0707	2.33759	-0.09017
0	5	-0.39668	-0.28528	-0.39899	-0.22724	-0.33354
0	5	-0.4033	-0.19003	-0.2521	-0.27256	-0.29651
0	2	-0.15271	-0.09869	-0.57433	-0.12948	-0.32873
0	5	-0.42683	0.1435	1.70921	2.36492	-0.47225
0	5	-0.60801	-0.3051	-0.21353	-0.44124	-0.49399
0	2	-0.09026	-0.26332	-0.71941	-0.23679	0.06833
0	2	-0.21022	-0.46192	-0.23579	-0.84111	0.33493
0	5	-0.44818	-0.30172	-0.31189	-0.2248	-0.33361
0	5	-0.35832	-0.18718	-0.18751	-0.06884	-0.09843
0	4	-0.5094	-0.37439	-1.29245	0.43195	-0.36751
0	2	0.59565	-0.00099	-0.69276	-0.18936	-0.6921
0	5	-0.326	-0.21978	-0.38035	-0.15048	-0.21591
0	2	-0.28894	-0.22263	-0.77322	0.0557	-0.4598
0	2	-0.20421	-0.11514	-0.48723	-0.12704	-0.32881
0	4	-0.5094	-0.37439	-1.29245	0.43195	-0.36751

Fig. Results by EFA

EFA in Excel

<https://www.real-statistics.com/multivariate-statistics/factor-analysis/>

	A	B	C	D	E
4	low	0.935			
5	back	0.935			
6	pain	0.876			
7	prognosi	0.785			
8	psycholog	0.599			
9	factor	0.507			
10	review	0.301			
11	model	0.289			
12	patholog	0.136			
13	data		0.899		
14	statist		0.899		
15	numer		0.899		
16	econom		0.546		
17	cost		0.526		
18	health		0.368		
19	trend		0.154		
20	spine			0.531	
21	spinal			0.479	
22	fusion			0.478	

	BG	BH	BI	BJ	BK	BL
class	F1	F2	F3	F4	F5	
	2	-0.32101	0.03681	-0.0707	2.33759	-0.09017
	5	-0.39668	-0.28528	-0.39899	-0.22724	-0.33354
	5	-0.4033	-0.19003	-0.2521	-0.27256	-0.29651
	2	-0.15271	-0.09869	-0.57433	-0.12948	-0.32873
	5	-0.42683	0.1435	1.70921	2.36492	-0.47225
	5	-0.60801	-0.3051	-0.21353	-0.44124	-0.49399
	2	-0.09026	-0.26332	-0.71941	-0.23679	0.06833
	2	-0.21022	-0.46192	-0.23579	-0.84111	0.33493
	5	-0.44818	-0.30172	-0.31189	-0.2248	-0.33361
	5	-0.35832	-0.18718	-0.18751	-0.06884	-0.09843
	4	-0.5094	-0.37439	-1.29245	0.43195	-0.36751
	2	0.59565	-0.00099	-0.69276	-0.18936	-0.6921
	5	-0.326	-0.21978	-0.38035	-0.15048	-0.21591

Using the factor scores to perform the k-means clustering and select the number of clusters to five equal to the factor number.

The document related to the cluster can be obtained.

To classify articles into subject categories:

1. dataset

{PMID, MeSh terms, Keywords} using SNA and determine cluster number=5 that will be clustered as below:

That is, each article can be classified into each category using the SNA.

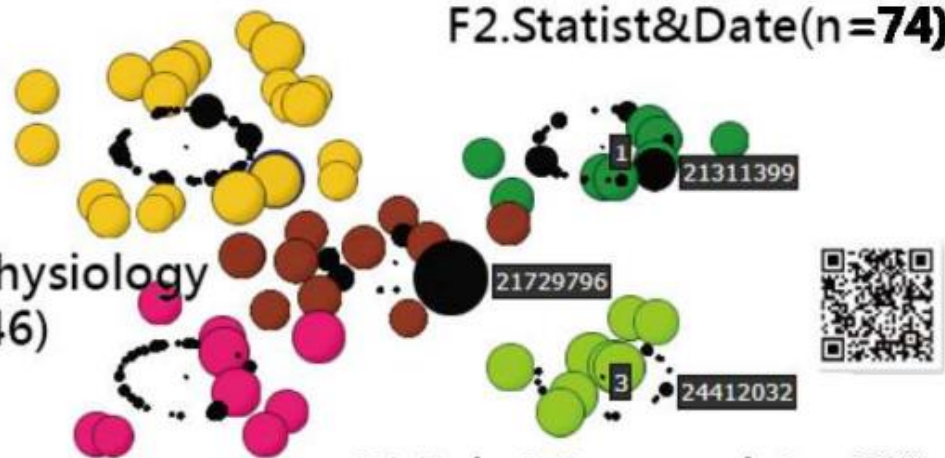
F3.Spine&Surgery(n=33)

F2.Statist&Date(n=74)

F5.Physiology (n=46)

F4.Physiopathology(n=18)

F1.Pain&Prognosis(n=29)



2. factor scores in two journals:

Because each articles with each factor score have been known, the SMD can be compared using the forest plot.

Similarly, the counts in each category for each journal can be obtained to perform the ChiSQ test for examining the count distributions equally across categories.