Nick Petty
Z23296080

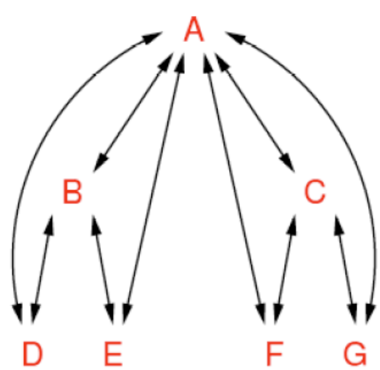
Homework 2 Node Similarity and Community Detection

Question 1 [1.5 pts] Please use your own language to briefly explain the following concepts:

PageRank score: a way of determining the importance of a web page by counting the number and quality of links to that page.
Rooted PageRank: a modification to PageRank that keeps the random walker focused on a particular area of the graph. This is done by randomly resetting the walker back to the root node being investigated.
Network community: a group of nodes in a network which are densely connected to each other compared to their connections with the rest of the network.
Clique: a subset of nodes in a network that forms a complete subgraph. Each node in the group is connected to all other nodes in the group.
k-Clique: a subgraph where the shortest path between any two nodes is <= k, and which cannot maintain this feature if more nodes are added.
Low-rank approximation: by computing the rank-k matrix of a graph, choosing a low value of k, and using another community detection method, the noise in a graph can be reduced. This is also used in conjunction with Singular Value Decomposition to find significant data amid noise.

Question 2 [2 pts]: Given seven web pages with the following link structure,

1. Please use "Power Iteration" (a.k.a simple iteration) to calculate the PageRank scores for each website. (You only need to show the first and the second iterations results, with the initial PageRank scores for each node being set as 1/n=0.15) [1 pt].
2. Please also use Eigenvector based approach to calculate PageRank scores for each web page [1 pt] (please show your solutions.)



1. The adjacency matrix is normalized and the PageRank Power Iteration is applied.

	A	B	C	D	E	F	G
A	0	1	1	1	1	1	1
B	1	0	0	1	1	0	0
C	1	0	0	0	0	1	1
D	1	1	0	0	0	0	0
E	1	1	0	0	0	0	0
F	1	0	1	0	0	0	0

	A	B	C	D	E	F	G
A	0.0000	0.3333	0.3333	0.5000	0.5000	0.5000	0.5000
B	0.1667	0.0000	0.0000	0.5000	0.5000	0.0000	0.0000
C	0.1667	0.0000	0.0000	0.0000	0.0000	0.5000	0.5000
D	0.1667	0.3333	0.0000	0.0000	0.0000	0.0000	0.0000
E	0.1667	0.3333	0.0000	0.0000	0.0000	0.0000	0.0000
F	0.1667	0.0000	0.3333	0.0000	0.0000	0.0000	0.0000

G	1	0	1	0	0	0	0

G	0.1667	0.0000	0.3333	0.0000	0.0000	0.0000	0.0000

	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5	Iteration 6	Iteration 7
A	0.1500	0.4000	0.2667	0.3444	0.2963	0.3272	0.3070
B	0.1500	0.1750	0.1417	0.1694	0.1491	0.1633	0.1536
C	0.1500	0.1750	0.1417	0.1694	0.1491	0.1633	0.1536
D	0.1500	0.0750	0.1250	0.0917	0.1139	0.0991	0.1090
E	0.1500	0.0750	0.1250	0.0917	0.1139	0.0991	0.1090
F	0.1500	0.0750	0.1250	0.0917	0.1139	0.0991	0.1090
G	0.1500	0.0750	0.1250	0.0917	0.1139	0.0991	0.1090

2. The eigenvalues and eigenvector were calculated with bluebit.gr, with the highest value highlighted to indicate the corresponding eigenvector. The first table is from the adjacency matrix; the second table is from the normalized adjacency matrix.

Eigenvalues	Eigenvector
3.2361	0.6015
2.0000	0.3717
1.2361	0.3717
1.4142	0.3008
1.4142	0.3008
0.0000	0.3008
0.0000	0.3008

Eigenvalues	Eigenvector
1.0000	0.7171
0.3333	0.3586
0.6667	0.3586
0.5773	0.2391
0.5773	0.2391
0.0000	0.2391
0.0000	0.2391

In both methods, the nodes are ranked as follows:

Rank 1 – A
Rank 2 – B, C
Rank 3 – D, E, F, G

Question 3 [1 pt]: In Quesiton 2, please use rooted PageRank to calculate similarity between each pair of nodes. Each time, the random walker has a probability 1-α (where α=*0.2*) to return back to an original node. (Please show your solutions).

The formula for rooted PageRank is (1 - α)(I - α * (D^{-1}*A)$)^{-1}$ and is calculated below.

A	A	B	C	D	E	F	G
A	0	1	1	1	1	1	1
B	1	0	0	1	1	0	0
C	1	0	0	0	0	1	1
D	1	1	0	0	0	0	0
E	1	1	0	0	0	0	0

I	A	B	C	D	E	F	G
A	1	0	0	0	0	0	0
B	0	1	0	0	0	0	0
C	0	0	1	0	0	0	0
D	0	0	0	1	0	0	0
E	0	0	0	0	1	0	0

D	A	B	C	D	E	F	G
A	6	0	0	0	0	0	0
B	0	3	0	0	0	0	0
C	0	0	3	0	0	0	0
D	0	0	0	2	0	0	0
E	0	0	0	0	2	0	0

F	1	0	1	0	0	0	0
G	1	0	1	0	0	0	0

F	0	0	0	0	0	1	0
G	0	0	0	0	0	0	1

F	0	0	0	0	0	2	0
G	0	0	0	0	0	0	2

D^{-1}	A	B	C	D	E	F	G
A	0.167	0.000	0.000	0.000	0.000	0.000	0.000
B	0.000	0.333	0.000	0.000	0.000	0.000	0.000
C	0.000	0.000	0.333	0.000	0.000	0.000	0.000
D	0.000	0.000	0.000	0.500	0.000	0.000	0.000
E	0.000	0.000	0.000	0.000	0.500	0.000	0.000
F	0.000	0.000	0.000	0.000	0.000	0.500	0.000
G	0.000	0.000	0.000	0.000	0.000	0.000	0.500

$D^{-1}*A$	A	B	C	D	E	F	G
A	0.000	0.167	0.167	0.167	0.167	0.167	0.167
B	0.333	0.000	0.000	0.333	0.333	0.000	0.000
C	0.333	0.000	0.000	0.000	0.000	0.333	0.333
D	0.500	0.500	0.000	0.000	0.000	0.000	0.000
E	0.500	0.500	0.000	0.000	0.000	0.000	0.000
F	0.500	0.000	0.500	0.000	0.000	0.000	0.000
G	0.500	0.000	0.500	0.000	0.000	0.000	0.000

$\alpha * (D^{-1}*A)$	A	B	C	D	E	F	G
A	0.000	0.033	0.033	0.033	0.033	0.033	0.033
B	0.067	0.000	0.000	0.067	0.067	0.000	0.000
C	0.067	0.000	0.000	0.000	0.000	0.067	0.067
D	0.100	0.100	0.000	0.000	0.000	0.000	0.000
E	0.100	0.100	0.000	0.000	0.000	0.000	0.000
F	0.100	0.000	0.100	0.000	0.000	0.000	0.000
G	0.100	0.000	0.100	0.000	0.000	0.000	0.000

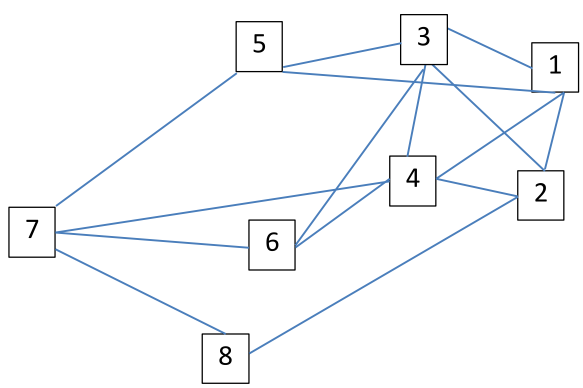
$I - \alpha * (D^{-1}*A)$	A	B	C	D	E	F	G
A	1.000	-0.033	-0.033	-0.033	-0.033	-0.033	-0.033
B	-0.067	1.000	0.000	-0.067	-0.067	0.000	0.000
C	-0.067	0.000	1.000	0.000	0.000	-0.067	-0.067
D	-0.100	-0.100	0.000	1.000	0.000	0.000	0.000
E	-0.100	-0.100	0.000	0.000	1.000	0.000	0.000
F	-0.100	0.000	-0.100	0.000	0.000	1.000	0.000
G	-0.100	0.000	-0.100	0.000	0.000	0.000	1.000

$(I - \alpha * (D^{-1}*A))^{-1}$	A	B	C	D	E	F	G
A	1.020	0.041	0.041	0.037	0.037	0.037	0.037
B	0.083	1.017	0.003	0.071	0.071	0.003	0.003
C	0.083	0.003	1.017	0.003	0.003	0.071	0.071
D	0.110	0.106	0.004	1.011	0.011	0.004	0.004
E	0.110	0.106	0.004	0.011	1.011	0.004	0.004
F	0.110	0.004	0.106	0.004	0.004	1.011	0.011
G	0.110	0.004	0.106	0.004	0.004	0.011	1.011

$(1 - \alpha)(I - \alpha * (D^{-1}*A))^{-1}$	A	B	C	D	E	F	G
A	0.816	0.033	0.033	0.029	0.029	0.029	0.029
B	0.066	0.813	0.003	0.056	0.056	0.002	0.002
C	0.066	0.003	0.813	0.002	0.002	0.056	0.056
D	0.088	0.085	0.004	0.809	0.009	0.003	0.003
E	0.088	0.085	0.004	0.009	0.809	0.003	0.003
F	0.088	0.004	0.085	0.003	0.003	0.809	0.009
G	0.088	0.004	0.085	0.003	0.003	0.009	0.809

Question 4 [1.5 pts]: The following networks show connections between 8 individuals in a small community. For node pairs (1, 7) and (1, 6), please use following measures to calculate their similarity (or distance) value and conclude which pair is more likely to form a link.

a. Jacarrd's Coefficient (0.25 pt)
b. Adamic/Adar (0.25 pt)
c. Preferential attachment (0.25 pt)
d. Katz (with β=0.05) (0.25 pt)
e. SimRank score with C=1 (please show the SimRank score after the 1st iteration). (0.5 pt)



Solutions are in the following image:

a. Jaccard's coefficient $J(1,7)$ vs $J(1,6)$

$\Gamma(1)=\{2,3,4,5\}$ $\Gamma(6)=\{3,4,7\}$ $\Gamma(7)=\{4,5,6,8\}$

$$J(1,7)=\frac{|\Gamma(1)\cap\Gamma(7)|}{|\Gamma(1)\cup\Gamma(7)|}=\frac{|\{4,5\}|}{|\{2,3,4,5,6,8\}|}=\frac{2}{6}=\frac{1}{3}$$

$$J(1,6)=\frac{|\Gamma(1)\cap\Gamma(6)|}{|\Gamma(1)\cup\Gamma(6)|}=\frac{|\{3,4\}|}{|\{2,3,4,5,7\}|}=\frac{2}{5}$$

$J(1,7)<J(1,6) \Rightarrow (1,6)$ is more likely to form a link

b. Adamic/Adar $A(1,7)$ vs $A(1,6)$

$\Gamma(1)\cap\Gamma(7)=\{4,5\}$ $\Gamma(1)\cap\Gamma(6)=\{3,4\}$

$|\Gamma(3)|=|\{1,2,4,5,6\}|=5$ $|\Gamma(4)|=|\{1,2,3,6,7\}|=5$ $|\Gamma(5)|=|\{1,3,7\}|=3$

$$A(1,7)=\sum_{z\in\{4,5\}}\frac{1}{\log|\Gamma(z)|}=\frac{1}{\log 5}+\frac{1}{\log 3}=1.431+2.096=3.527$$

$$A(1,6)=\sum_{z\in\{3,4\}}\frac{1}{\log|\Gamma(z)|}=\frac{1}{\log 5}+\frac{1}{\log 5}=\frac{2}{\log 5}=2.861$$

$A(1,7)>A(1,6)\Rightarrow (1,7)$ is more likely to form a link

c. Preferential attachment $P(1,7)$ vs $P(1,6)$

$P(1,7)=|\Gamma(1)|\cdot|\Gamma(7)|=4\cdot 4=16$

$P(1,6)=|\Gamma(1)|\cdot|\Gamma(6)|=4\cdot 3=12$

$P(1,7)>P(1,6)\Rightarrow (1,7)$ is more likely to form a link

d. Katz $K(1,7)$ vs $K(1,6)$ $\beta=0.05$

$\sum_{l=1}^{\infty}\beta^{l}\cdot|paths_{x,y}^{\langle l\rangle}|=\sum_{l=1}^{3}0.05^{l}\cdot|paths_{x\to y}^{\langle l\rangle}|$ using $l\to 3$

$K(1,7)=0.05^1\cdot 0+0.05^2\cdot 2+0.05^3\cdot 5=0.005625$

$K(1,6)=0.05^1\cdot 0+0.05^2\cdot 2+0.05^3\cdot 7=0.005875$

$K(1,7)<K(1,6)\Rightarrow (1,6)$ is more likely to form a link

---

e. SimRank with $C=1$ $S(1,7)$ vs $S(1,6)$

$S_0(a,b)=\begin{cases}0 & \text{if } a\neq b\\ 1 & \text{if } a=b\end{cases}$ $S_{k+1}(a,b)=\frac{C}{|\Gamma(a)|\cdot|\Gamma(b)|}\cdot\sum_{i\in\Gamma(a)}\sum_{j\in\Gamma(b)}S(i,j)$

$|\Gamma(1)|=|\{2,3,4,5\}|=4$ $|\Gamma(6)|=|\{3,4,7\}|=3$ $|\Gamma(7)|=|\{4,5,6,8\}|=4$

$$S_0(1,7)=0\rightarrow S_1(1,7)=\frac{1}{4\cdot 4}\cdot\sum_{i\in\Gamma(1)}\sum_{j\in\Gamma(7)}S(i,j)=\frac{1}{4\cdot 4}(1+1)=\frac{2}{16}=\frac{1}{8}$$

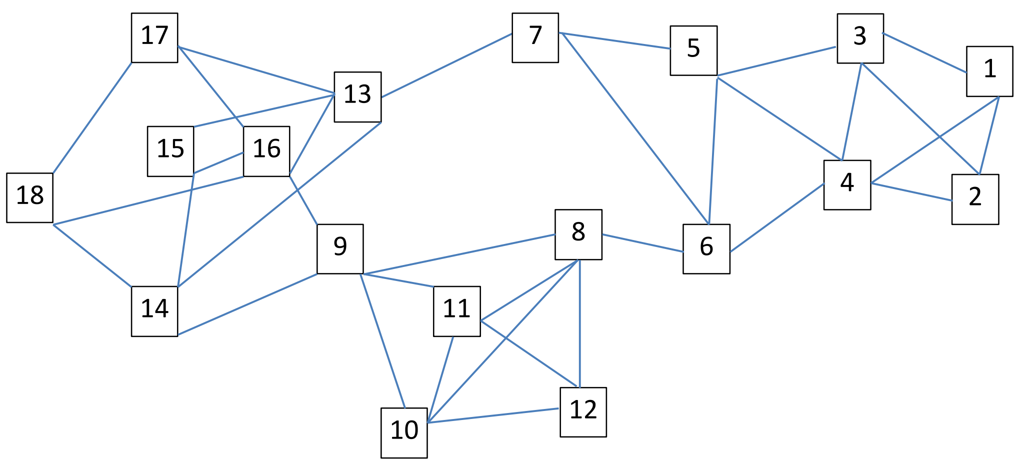
$$S_0(1,6)=0\rightarrow S_1(1,6)=\frac{1}{4\cdot 3}\cdot\sum_{i\in\Gamma(1)}\sum_{j\in\Gamma(6)}S(i,j)=\frac{1}{12}\cdot(1+1)=\frac{2}{12}=\frac{1}{6}$$

$S(1,7)<S(1,6)\Rightarrow (1,6)$ is more likely to form a link.

Question 5 [4 pts]: In the following network,

1. Please find the complete set of communities by using 3-clquie [0.25pt], 3-club [0.25pt], and 3-core [0.25pt], respectively (If there are multiple sets, please just report the top three sets with the maximum number of nodes).
2. Please calculate the Geodesic distance between each pair of nodes, and use Multidimensional Scaling (MDS) to convert the network into a two dimensional space. Please report the values of all nodes in the two dimensional space and draw all nodes in the two dimensional space [1.25 pt].
3. Implement a k-means clustering algorithm (selecting k=2 and using node 18 and node 1 as the initial centers), and report the community structures after 10 iterations (You may use any other third party tools for k-means clustering. Or you can follow the k-means Excel implementation in the following URL to calculate the results) [2 pts]

k-means: http://www.csse.monash.edu.au/courseware/cse5230/2004/assets/clustering.pdf



1. 3-cliques:

{17, 16, 15, 14, 13, 11, 10, 9, 8, 7, 6, 5} = 12 nodes
{18, 17, 16, 15, 14, 13, 11, 10, 9, 8, 7} = 11 nodes
{1, 2, 3, 4, 5, 6, 7, 8} = 8 nodes

3-clubs:
{17, 16, 15, 14, 13, 11, 10, 9, 8, 7, 6, 5} = 12 nodes, substructure diameter <= 3
{18, 17, 16, 15, 14, 13, 11, 10, 9, 8} = 10 nodes, substructure diameter <= 3
{1, 2, 3, 4, 5, 6, 7, 8} = 8 nodes, substructure diameter <= 3

3-cores:
{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18} = 18 nodes
All nodes have degree >= 3, so they all connect to at least 3 other nodes in the same substructure, and are therefore in the same 3-core.

2. Several matrices are needed for this calculation, the first is the geodesic distance and square of geodesic distance, D:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0	1	1	1	2	2	3	3	4	4	4	4	4	5	5	5	5	6
2	1	0	1	1	2	2	3	3	4	4	4	4	4	5	5	5	5	6

3	1	1	0	1	1	2	2	3	4	4	4	4	3	4	4	4	4	5
4	1	1	1	0	1	1	2	2	3	3	3	3	3	4	4	4	4	5
5	2	2	1	1	0	1	1	2	3	3	3	3	2	3	3	3	3	4
6	2	2	2	1	1	0	1	1	2	2	2	2	2	3	3	3	3	4
7	3	3	2	2	1	1	0	2	3	3	3	3	1	2	2	2	2	3
8	3	3	3	2	2	1	2	0	1	1	1	1	3	2	3	2	3	3
9	4	4	4	3	3	2	3	1	0	1	1	2	2	1	2	1	2	2
10	4	4	4	3	3	2	3	1	1	0	1	1	3	2	3	2	3	3
11	4	4	4	3	3	2	3	1	1	1	0	1	3	2	3	2	3	3
12	4	4	4	3	3	2	3	1	2	1	1	0	4	3	4	3	4	4
13	4	4	3	3	2	2	1	3	2	3	3	4	0	1	1	1	1	2
14	5	5	4	4	3	3	2	2	1	2	2	3	1	0	1	2	2	1
15	5	5	4	4	3	3	2	3	2	3	3	4	1	1	0	1	2	2
16	5	5	4	4	3	3	2	2	1	2	2	3	1	2	1	0	1	1
17	5	5	4	4	3	3	2	3	2	3	3	4	1	2	2	1	0	1
18	6	6	5	5	4	4	3	3	2	3	3	4	2	1	2	1	1	0

D	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0	1	1	1	4	4	9	9	16	16	16	16	16	25	25	25	25	36
2	1	0	1	1	4	4	9	9	16	16	16	16	16	25	25	25	25	36
3	1	1	0	1	1	4	4	9	16	16	16	16	9	16	16	16	16	25
4	1	1	1	0	1	1	4	4	9	9	9	9	9	16	16	16	16	25
5	4	4	1	1	0	1	1	4	9	9	9	9	4	9	9	9	9	16
6	4	4	4	1	1	0	1	1	4	4	4	4	4	9	9	9	9	16
7	9	9	4	4	1	1	0	4	9	9	9	9	1	4	4	4	4	9
8	9	9	9	4	4	1	4	0	1	1	1	1	9	4	9	4	9	9
9	16	16	16	9	9	4	9	1	0	1	1	4	4	1	4	1	4	4
10	16	16	16	9	9	4	9	1	1	0	1	1	9	4	9	4	9	9
11	16	16	16	9	9	4	9	1	1	1	0	1	9	4	9	4	9	9
12	16	16	16	9	9	4	9	1	4	1	1	0	16	9	16	9	16	16
13	16	16	9	9	4	4	1	9	4	9	9	16	0	1	1	1	1	4
14	25	25	16	16	9	9	4	4	1	4	4	9	1	0	1	4	4	1
15	25	25	16	16	9	9	4	9	4	9	9	16	1	1	0	1	4	4
16	25	25	16	16	9	9	4	4	1	4	4	9	1	4	1	0	1	1
17	25	25	16	16	9	9	4	9	4	9	9	16	1	4	4	1	0	1
18	36	36	25	25	16	16	9	9	4	9	9	16	4	1	4	1	1	0

Next is $I - (1/n)ee^T$, which is also denoted as J:

J	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0.944	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056
2	-0.056	0.944	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056
3	-0.056	-0.056	0.944	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056
4	-0.056	-0.056	-0.056	0.944	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056
5	-0.056	-0.056	-0.056	-0.056	0.944	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056
6	-0.056	-0.056	-0.056	-0.056	-0.056	0.944	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056
7	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	0.944	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056
8	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	0.944	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056
9	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	0.944	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056
10	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	0.944	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056
11	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	0.944	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056
12	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	0.944	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056
13	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	0.944	-0.056	-0.056	-0.056	-0.056	-0.056
14	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	0.944	-0.056	-0.056	-0.056	-0.056
15	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	0.944	-0.056	-0.056	-0.056
16	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	0.944	-0.056	-0.056
17	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	0.944	-0.056
18	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	-0.056	0.944

From this, we calculate (-1/2)J*D*J = P:

P	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	9.515	9.015	7.293	6.293	3.710	3.154	0.821	0.654	-2.401	-1.762	-1.762	-0.623	-2.123	-5.985	-5.290	-6.068	-5.290	-9.151
2	9.015	9.515	7.293	6.293	3.710	3.154	0.821	0.654	-2.401	-1.762	-1.762	-0.623	-2.123	-5.985	-5.290	-6.068	-5.290	-9.151
3	7.293	7.293	6.071	4.571	3.488	1.432	1.599	-1.068	-4.123	-3.485	-3.485	-2.346	-0.346	-3.207	-2.512	-3.290	-2.512	-5.373
4	6.293	6.293	4.571	4.071	2.488	1.932	0.599	0.432	-1.623	-0.985	-0.985	0.154	-1.346	-4.207	-3.512	-4.290	-3.512	-6.373
5	3.710	3.710	3.488	2.488	1.904	0.849	1.015	-0.651	-2.707	-2.068	-2.068	-0.929	0.071	-1.790	-1.096	-1.873	-1.096	-2.957
6	3.154	3.154	1.432	1.932	0.849	0.793	0.460	0.293	-0.762	-0.123	-0.123	1.015	-0.485	-2.346	-1.651	-2.429	-1.651	-3.512
7	0.821	0.821	1.599	0.599	1.015	0.460	1.127	-1.040	-3.096	-2.457	-2.457	-1.318	1.182	0.321	1.015	0.238	1.015	0.154
8	0.654	0.654	-1.068	0.432	-0.651	0.293	-1.040	0.793	0.738	1.377	1.377	2.515	-2.985	0.154	-1.651	0.071	-1.651	-0.012
9	-2.401	-2.401	-4.123	-1.623	-2.707	-0.762	-3.096	0.738	1.682	1.821	1.821	1.460	-0.040	2.099	1.293	2.015	1.293	2.932
10	-1.762	-1.762	-3.485	-0.985	-2.068	-0.123	-2.457	1.377	1.821	2.960	2.460	3.599	-1.901	1.238	-0.568	1.154	-0.568	1.071
11	-1.762	-1.762	-3.485	-0.985	-2.068	-0.123	-2.457	1.377	1.821	2.460	2.960	3.599	-1.901	1.238	-0.568	1.154	-0.568	1.071
12	-0.623	-0.623	-2.346	0.154	-0.929	1.015	-1.318	2.515	1.460	3.599	3.599	5.238	-4.262	-0.123	-2.929	-0.207	-2.929	-1.290
13	-2.123	-2.123	-0.346	-1.346	0.071	-0.485	1.182	-2.985	-0.040	-1.901	-1.901	-4.262	2.238	2.377	3.071	2.293	3.071	3.210
14	-5.985	-5.985	-3.207	-4.207	-1.790	-2.346	0.321	0.154	2.099	1.238	1.238	-0.123	2.377	3.515	3.710	1.432	2.210	5.349
15	-5.290	-5.290	-2.512	-3.512	-1.096	-1.651	1.015	-1.651	1.293	-0.568	-0.568	-2.929	3.071	3.710	4.904	3.627	2.904	4.543
16	-6.068	-6.068	-3.290	-4.290	-1.873	-2.429	0.238	0.071	2.015	1.154	1.154	-0.207	2.293	1.432	3.627	3.349	3.627	5.265

17	-5.290	-5.290	-2.512	-3.512	-1.096	-1.651	1.015	-1.651	1.293	-0.568	-0.568	-2.929	3.071	2.210	2.904	3.627	4.904	6.043
18	-9.151	-9.151	-5.373	-6.373	-2.957	-3.512	0.154	-0.012	2.932	1.071	1.071	-1.290	3.210	5.349	4.543	5.265	6.043	8.182

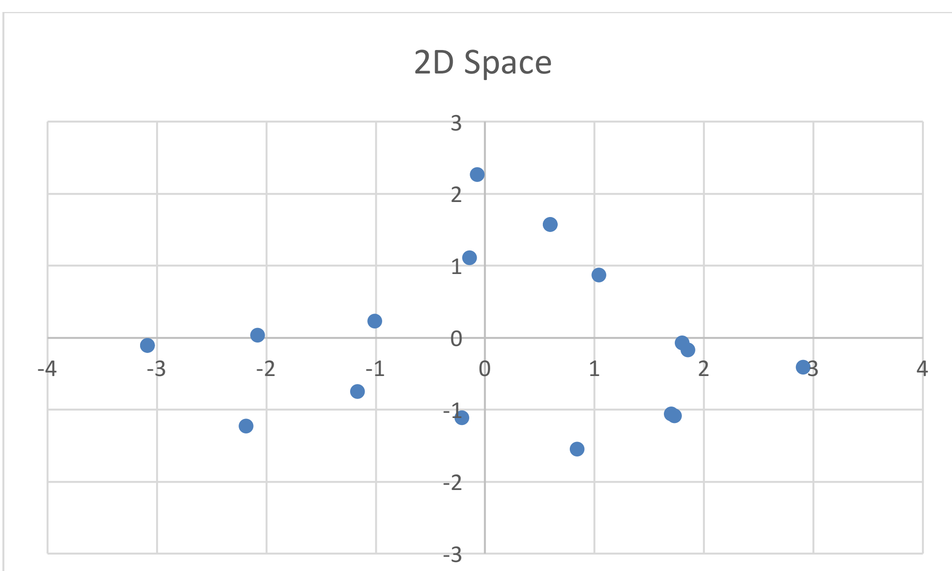
Finally, $S(x, y) = V*\lambda^{1/2}$, where S is the 2D coordinates of each node, λ is the diagonal matrix of the 2 top eigenvalues of P, and V is the corresponding 2 eigenvectors of P:

V	X	Y
1	-0.420	-0.024
2	-0.420	-0.024
3	-0.297	-0.273
4	-0.283	0.007
5	-0.159	-0.165
6	-0.137	0.050
7	-0.029	-0.247
8	-0.019	0.246
9	0.141	0.193
10	0.081	0.349
11	0.081	0.349
12	-0.010	0.502
13	0.114	-0.343
14	0.245	-0.017
15	0.231	-0.234
16	0.252	-0.037
17	0.235	-0.241
18	0.395	-0.091

$\lambda^{1/2}$	1	2
1	7.353	0.000
2	0.000	4.514

S	X	Y
1	-3.088	-0.108
2	-3.088	-0.108
3	-2.184	-1.232
4	-2.081	0.032
5	-1.169	-0.745
6	-1.007	0.226
7	-0.213	-1.115
8	-0.140	1.111
9	1.037	0.871
10	0.596	1.576
11	0.596	1.576
12	-0.074	2.266
13	0.838	-1.548
14	1.802	-0.077
15	1.699	-1.056
16	1.853	-0.167
17	1.728	-1.088
18	2.905	-0.411

From S, the 2D Space graph can be created:



3. 10 iterations of a k-means clustering algorithm are shown in the tables below. There are 2 clusters, labeled K1 and K2, with centroids at (K1_X, K1_Y) and (K2_X, K2_Y), respectively. After each iteration, the centroids are updated to (NK1_X, NK1_Y) and (NK2_X, NK2_Y). In this graph, the centroids are stable by the second iteration. The members of the clusters are:

K1 = {1, 2, 3, 4, 5, 6, 7, 8, 12}
K2 = {9, 10, 11, 13, 14, 15, 16, 17, 18}

I_1	X	Y	D_K1	D_K2	Clust.	K1		K2	
1	-3.088	-0.108	0.000	6.001	K1	-3.088	-0.108		
2	-3.088	-0.108	0.000	6.001	K1	-3.088	-0.108		
3	-2.184	-1.232	1.443	5.154	K1	-2.184	-1.232		
4	-2.081	0.032	1.017	5.005	K1	-2.081	0.032		
5	-1.169	-0.745	2.022	4.087	K1	-1.169	-0.745		
6	-1.007	0.226	2.108	3.963	K1	-1.007	0.226		
7	-0.213	-1.115	3.046	3.196	K1	-0.213	-1.115		
8	-0.140	1.111	3.191	3.403	K1	-0.140	1.111		
9	1.037	0.871	4.240	2.265	K2			1.037	0.871
10	0.596	1.576	4.051	3.046	K2			0.596	1.576
11	0.596	1.576	4.051	3.046	K2			0.596	1.576
12	-0.074	2.266	3.838	4.004	K1	-0.074	2.266		
13	0.838	-1.548	4.182	2.359	K2			0.838	-1.548
14	1.802	-0.077	4.890	1.152	K2			1.802	-0.077
15	1.699	-1.056	4.880	1.368	K2			1.699	-1.056
16	1.853	-0.167	4.942	1.079	K2			1.853	-0.167
17	1.728	-1.088	4.915	1.357	K2			1.728	-1.088
18	2.905	-0.411	6.001	0.000	K2			2.905	-0.411

K1_X	K1_Y	K2_X	K2_Y
-3.088	-0.108	2.905	-0.411

NK1_X	NK1_Y	NK2_X	NK2_Y
-1.449	0.036	1.450	-0.036

I_2	X	Y	D_K1	D_K2	Clust.	K1		K2	
1	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
2	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
3	-2.184	-1.232	1.466	3.826	K1	-2.184	-1.232		
4	-2.081	0.032	0.632	3.532	K1	-2.081	0.032		
5	-1.169	-0.745	0.830	2.714	K1	-1.169	-0.745		
6	-1.007	0.226	0.481	2.472	K1	-1.007	0.226		
7	-0.213	-1.115	1.689	1.983	K1	-0.213	-1.115		
8	-0.140	1.111	1.694	1.960	K1	-0.140	1.111		

9	1.037	0.871	2.623	0.997	K2			1.037	0.871
10	0.596	1.576	2.560	1.824	K2			0.596	1.576
11	0.596	1.576	2.560	1.824	K2			0.596	1.576
12	-0.074	2.266	2.620	2.761	K1	-0.074	2.266		
13	0.838	-1.548	2.783	1.631	K2			0.838	-1.548
14	1.802	-0.077	3.253	0.354	K2			1.802	-0.077
15	1.699	-1.056	3.332	1.050	K2			1.699	-1.056
16	1.853	-0.167	3.309	0.424	K2			1.853	-0.167
17	1.728	-1.088	3.370	1.088	K2			1.728	-1.088
18	2.905	-0.411	4.377	1.502	K2			2.905	-0.411

K1_X	K1_Y	K2_X	K2_Y
-1.449	0.036	1.450	-0.036

NK1_X	NK1_Y	NK2_X	NK2_Y
-1.449	0.036	1.450	-0.036

I_3	X	Y	D_K1	D_K2	Clust.	K1		K2	
1	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
2	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
3	-2.184	-1.232	1.466	3.826	K1	-2.184	-1.232		
4	-2.081	0.032	0.632	3.532	K1	-2.081	0.032		
5	-1.169	-0.745	0.830	2.714	K1	-1.169	-0.745		
6	-1.007	0.226	0.481	2.472	K1	-1.007	0.226		
7	-0.213	-1.115	1.689	1.983	K1	-0.213	-1.115		
8	-0.140	1.111	1.694	1.960	K1	-0.140	1.111		
9	1.037	0.871	2.623	0.997	K2			1.037	0.871
10	0.596	1.576	2.560	1.824	K2			0.596	1.576
11	0.596	1.576	2.560	1.824	K2			0.596	1.576
12	-0.074	2.266	2.620	2.761	K1	-0.074	2.266		
13	0.838	-1.548	2.783	1.631	K2			0.838	-1.548
14	1.802	-0.077	3.253	0.354	K2			1.802	-0.077
15	1.699	-1.056	3.332	1.050	K2			1.699	-1.056
16	1.853	-0.167	3.309	0.424	K2			1.853	-0.167
17	1.728	-1.088	3.370	1.088	K2			1.728	-1.088
18	2.905	-0.411	4.377	1.502	K2			2.905	-0.411

K1_X	K1_Y	K2_X	K2_Y
-1.449	0.036	1.450	-0.036

NK1_X	NK1_Y	NK2_X	NK2_Y
-1.449	0.036	1.450	-0.036

I_4	X	Y	D_K1	D_K2	Clust.	K1		K2	
1	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
2	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		

3	-2.184	-1.232	1.466	3.826	K1	-2.184	-1.232		
4	-2.081	0.032	0.632	3.532	K1	-2.081	0.032		
5	-1.169	-0.745	0.830	2.714	K1	-1.169	-0.745		
6	-1.007	0.226	0.481	2.472	K1	-1.007	0.226		
7	-0.213	-1.115	1.689	1.983	K1	-0.213	-1.115		
8	-0.140	1.111	1.694	1.960	K1	-0.140	1.111		
9	1.037	0.871	2.623	0.997	K2			1.037	0.871
10	0.596	1.576	2.560	1.824	K2			0.596	1.576
11	0.596	1.576	2.560	1.824	K2			0.596	1.576
12	-0.074	2.266	2.620	2.761	K1	-0.074	2.266		
13	0.838	-1.548	2.783	1.631	K2			0.838	-1.548
14	1.802	-0.077	3.253	0.354	K2			1.802	-0.077
15	1.699	-1.056	3.332	1.050	K2			1.699	-1.056
16	1.853	-0.167	3.309	0.424	K2			1.853	-0.167
17	1.728	-1.088	3.370	1.088	K2			1.728	-1.088
18	2.905	-0.411	4.377	1.502	K2			2.905	-0.411

K1_X	K1_Y	K2_X	K2_Y
-1.449	0.036	1.450	-0.036

NK1_X	NK1_Y	NK2_X	NK2_Y
-1.449	0.036	1.450	-0.036

I_5	X	Y	D_K1	D_K2	Clust.	K1		K2	
1	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
2	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
3	-2.184	-1.232	1.466	3.826	K1	-2.184	-1.232		
4	-2.081	0.032	0.632	3.532	K1	-2.081	0.032		
5	-1.169	-0.745	0.830	2.714	K1	-1.169	-0.745		
6	-1.007	0.226	0.481	2.472	K1	-1.007	0.226		
7	-0.213	-1.115	1.689	1.983	K1	-0.213	-1.115		
8	-0.140	1.111	1.694	1.960	K1	-0.140	1.111		
9	1.037	0.871	2.623	0.997	K2			1.037	0.871
10	0.596	1.576	2.560	1.824	K2			0.596	1.576
11	0.596	1.576	2.560	1.824	K2			0.596	1.576
12	-0.074	2.266	2.620	2.761	K1	-0.074	2.266		
13	0.838	-1.548	2.783	1.631	K2			0.838	-1.548
14	1.802	-0.077	3.253	0.354	K2			1.802	-0.077
15	1.699	-1.056	3.332	1.050	K2			1.699	-1.056
16	1.853	-0.167	3.309	0.424	K2			1.853	-0.167
17	1.728	-1.088	3.370	1.088	K2			1.728	-1.088
18	2.905	-0.411	4.377	1.502	K2			2.905	-0.411

K1_X	K1_Y	K2_X	K2_Y
-1.449	0.036	1.450	-0.036

NK1_X	NK1_Y	NK2_X	NK2_Y
-1.449	0.036	1.450	-0.036

I_6	X	Y	D_K1	D_K2	Clust.	K1		K2	
1	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
2	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
3	-2.184	-1.232	1.466	3.826	K1	-2.184	-1.232		
4	-2.081	0.032	0.632	3.532	K1	-2.081	0.032		
5	-1.169	-0.745	0.830	2.714	K1	-1.169	-0.745		
6	-1.007	0.226	0.481	2.472	K1	-1.007	0.226		
7	-0.213	-1.115	1.689	1.983	K1	-0.213	-1.115		
8	-0.140	1.111	1.694	1.960	K1	-0.140	1.111		
9	1.037	0.871	2.623	0.997	K2			1.037	0.871
10	0.596	1.576	2.560	1.824	K2			0.596	1.576
11	0.596	1.576	2.560	1.824	K2			0.596	1.576
12	-0.074	2.266	2.620	2.761	K1	-0.074	2.266		
13	0.838	-1.548	2.783	1.631	K2			0.838	-1.548
14	1.802	-0.077	3.253	0.354	K2			1.802	-0.077
15	1.699	-1.056	3.332	1.050	K2			1.699	-1.056
16	1.853	-0.167	3.309	0.424	K2			1.853	-0.167
17	1.728	-1.088	3.370	1.088	K2			1.728	-1.088
18	2.905	-0.411	4.377	1.502	K2			2.905	-0.411

K1_X	K1_Y	K2_X	K2_Y
-1.449	0.036	1.450	-0.036

NK1_X	NK1_Y	NK2_X	NK2_Y
-1.449	0.036	1.450	-0.036

I_7	X	Y	D_K1	D_K2	Clust.	K1		K2	
1	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
2	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
3	-2.184	-1.232	1.466	3.826	K1	-2.184	-1.232		
4	-2.081	0.032	0.632	3.532	K1	-2.081	0.032		
5	-1.169	-0.745	0.830	2.714	K1	-1.169	-0.745		
6	-1.007	0.226	0.481	2.472	K1	-1.007	0.226		
7	-0.213	-1.115	1.689	1.983	K1	-0.213	-1.115		
8	-0.140	1.111	1.694	1.960	K1	-0.140	1.111		
9	1.037	0.871	2.623	0.997	K2			1.037	0.871
10	0.596	1.576	2.560	1.824	K2			0.596	1.576
11	0.596	1.576	2.560	1.824	K2			0.596	1.576
12	-0.074	2.266	2.620	2.761	K1	-0.074	2.266		
13	0.838	-1.548	2.783	1.631	K2			0.838	-1.548

14	1.802	-0.077	3.253	0.354	K2			1.802	-0.077
15	1.699	-1.056	3.332	1.050	K2			1.699	-1.056
16	1.853	-0.167	3.309	0.424	K2			1.853	-0.167
17	1.728	-1.088	3.370	1.088	K2			1.728	-1.088
18	2.905	-0.411	4.377	1.502	K2			2.905	-0.411

K1_X	K1_Y	K2_X	K2_Y
-1.449	0.036	1.450	-0.036

NK1_X	NK1_Y	NK2_X	NK2_Y
-1.449	0.036	1.450	-0.036

I_8	X	Y	D_K1	D_K2	Clust.	K1		K2	
1	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
2	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
3	-2.184	-1.232	1.466	3.826	K1	-2.184	-1.232		
4	-2.081	0.032	0.632	3.532	K1	-2.081	0.032		
5	-1.169	-0.745	0.830	2.714	K1	-1.169	-0.745		
6	-1.007	0.226	0.481	2.472	K1	-1.007	0.226		
7	-0.213	-1.115	1.689	1.983	K1	-0.213	-1.115		
8	-0.140	1.111	1.694	1.960	K1	-0.140	1.111		
9	1.037	0.871	2.623	0.997	K2			1.037	0.871
10	0.596	1.576	2.560	1.824	K2			0.596	1.576
11	0.596	1.576	2.560	1.824	K2			0.596	1.576
12	-0.074	2.266	2.620	2.761	K1	-0.074	2.266		
13	0.838	-1.548	2.783	1.631	K2			0.838	-1.548
14	1.802	-0.077	3.253	0.354	K2			1.802	-0.077
15	1.699	-1.056	3.332	1.050	K2			1.699	-1.056
16	1.853	-0.167	3.309	0.424	K2			1.853	-0.167
17	1.728	-1.088	3.370	1.088	K2			1.728	-1.088
18	2.905	-0.411	4.377	1.502	K2			2.905	-0.411

K1_X	K1_Y	K2_X	K2_Y
-1.449	0.036	1.450	-0.036

NK1_X	NK1_Y	NK2_X	NK2_Y
-1.449	0.036	1.450	-0.036

I_9	X	Y	D_K1	D_K2	Clust.	K1		K2	
1	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
2	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
3	-2.184	-1.232	1.466	3.826	K1	-2.184	-1.232		
4	-2.081	0.032	0.632	3.532	K1	-2.081	0.032		
5	-1.169	-0.745	0.830	2.714	K1	-1.169	-0.745		
6	-1.007	0.226	0.481	2.472	K1	-1.007	0.226		
7	-0.213	-1.115	1.689	1.983	K1	-0.213	-1.115		

8	-0.140	1.111	1.694	1.960	K1	-0.140	1.111		
9	1.037	0.871	2.623	0.997	K2			1.037	0.871
10	0.596	1.576	2.560	1.824	K2			0.596	1.576
11	0.596	1.576	2.560	1.824	K2			0.596	1.576
12	-0.074	2.266	2.620	2.761	K1	-0.074	2.266		
13	0.838	-1.548	2.783	1.631	K2			0.838	-1.548
14	1.802	-0.077	3.253	0.354	K2			1.802	-0.077
15	1.699	-1.056	3.332	1.050	K2			1.699	-1.056
16	1.853	-0.167	3.309	0.424	K2			1.853	-0.167
17	1.728	-1.088	3.370	1.088	K2			1.728	-1.088
18	2.905	-0.411	4.377	1.502	K2			2.905	-0.411

K1_X	K1_Y	K2_X	K2_Y
-1.449	0.036	1.450	-0.036

NK1_X	NK1_Y	NK2_X	NK2_Y
-1.449	0.036	1.450	-0.036

I_{10}	X	Y	D_K1	D_K2	Clust.	K1		K2	
1	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
2	-3.088	-0.108	1.645	4.539	K1	-3.088	-0.108		
3	-2.184	-1.232	1.466	3.826	K1	-2.184	-1.232		
4	-2.081	0.032	0.632	3.532	K1	-2.081	0.032		
5	-1.169	-0.745	0.830	2.714	K1	-1.169	-0.745		
6	-1.007	0.226	0.481	2.472	K1	-1.007	0.226		
7	-0.213	-1.115	1.689	1.983	K1	-0.213	-1.115		
8	-0.140	1.111	1.694	1.960	K1	-0.140	1.111		
9	1.037	0.871	2.623	0.997	K2			1.037	0.871
10	0.596	1.576	2.560	1.824	K2			0.596	1.576
11	0.596	1.576	2.560	1.824	K2			0.596	1.576
12	-0.074	2.266	2.620	2.761	K1	-0.074	2.266		
13	0.838	-1.548	2.783	1.631	K2			0.838	-1.548
14	1.802	-0.077	3.253	0.354	K2			1.802	-0.077
15	1.699	-1.056	3.332	1.050	K2			1.699	-1.056
16	1.853	-0.167	3.309	0.424	K2			1.853	-0.167
17	1.728	-1.088	3.370	1.088	K2			1.728	-1.088
18	2.905	-0.411	4.377	1.502	K2			2.905	-0.411

K1_X	K1_Y	K2_X	K2_Y
-1.449	0.036	1.450	-0.036

NK1_X	NK1_Y	NK2_X	NK2_Y
-1.449	0.036	1.450	-0.036