GROUP D:

PAULO SOUSA

SOFIA MALPIQUE

YANNIK BAUER

ASSIGNMENT 6

TASK 1.1

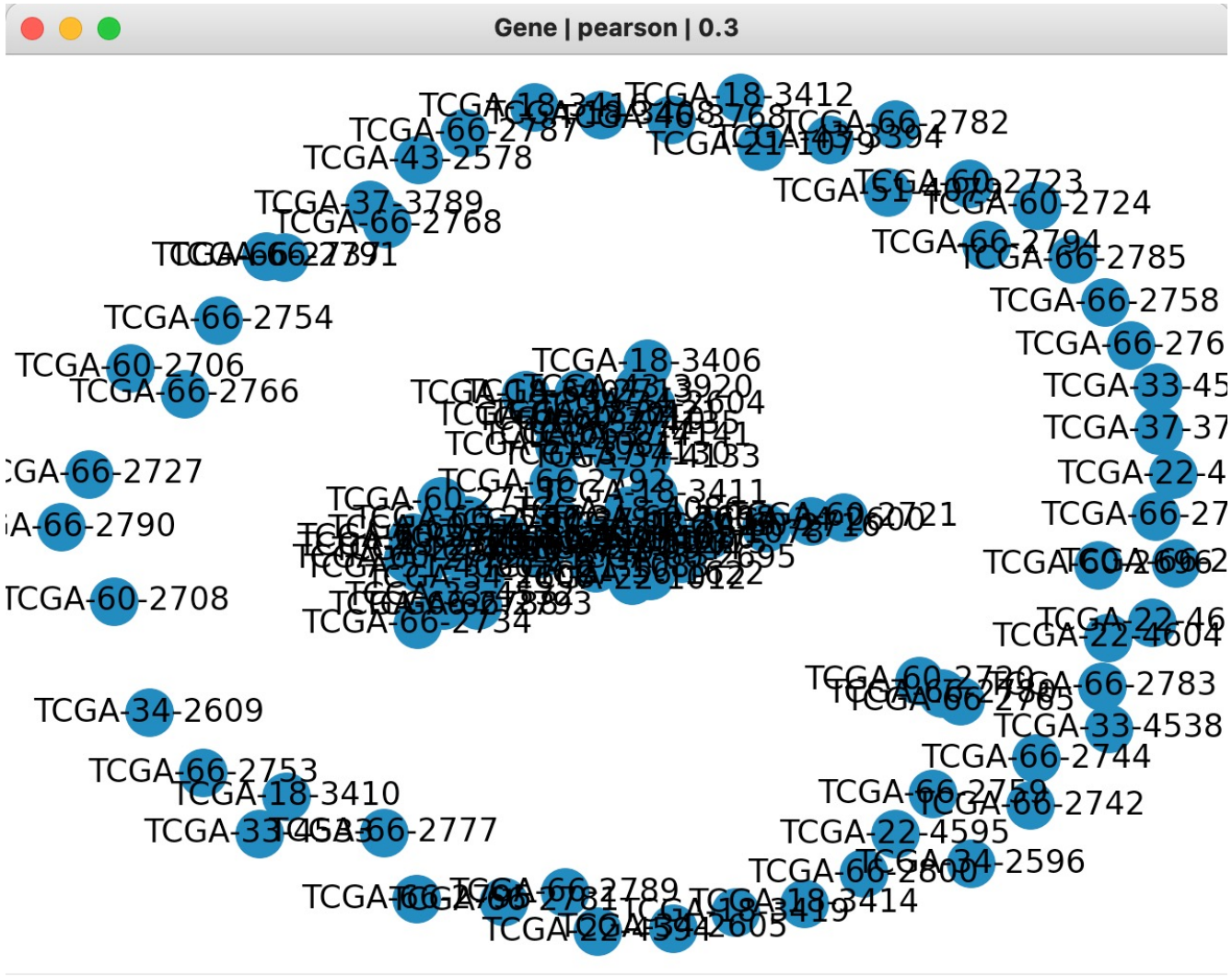
ONE CORRELATION MATRIZ:

Correlation Matriz | datatype: Gene | correlation= pearson | :

	TCGA-18-3406	TCGA-18-3407	...	TCGA-66-2795	TCGA-66-2800
TCGA-18-3406	1.000000	0.104163	...	0.147598	-0.050254
TCGA-18-3407	0.104163	1.000000	...	-0.001384	0.002215
TCGA-18-3408	0.090739	0.071621	...	0.143832	0.187454
TCGA-18-3410	0.043099	-0.081285	...	-0.076058	-0.067875
TCGA-18-3411	0.088210	-0.079021	...	0.182132	0.109491
...	...	...	...	...	...
TCGA-66-2792	-0.196908	0.078084	...	0.033762	0.160412
TCGA-66-2793	-0.122325	-0.124143	...	-0.103974	-0.064065
TCGA-66-2794	-0.190241	-0.122525	...	0.038960	0.187828
TCGA-66-2795	0.147598	-0.001384	...	1.000000	0.117730
TCGA-66-2800	-0.050254	0.002215	...	0.117730	1.000000

TASK 1.2

- ONE GRAPH:



TASK 1.3

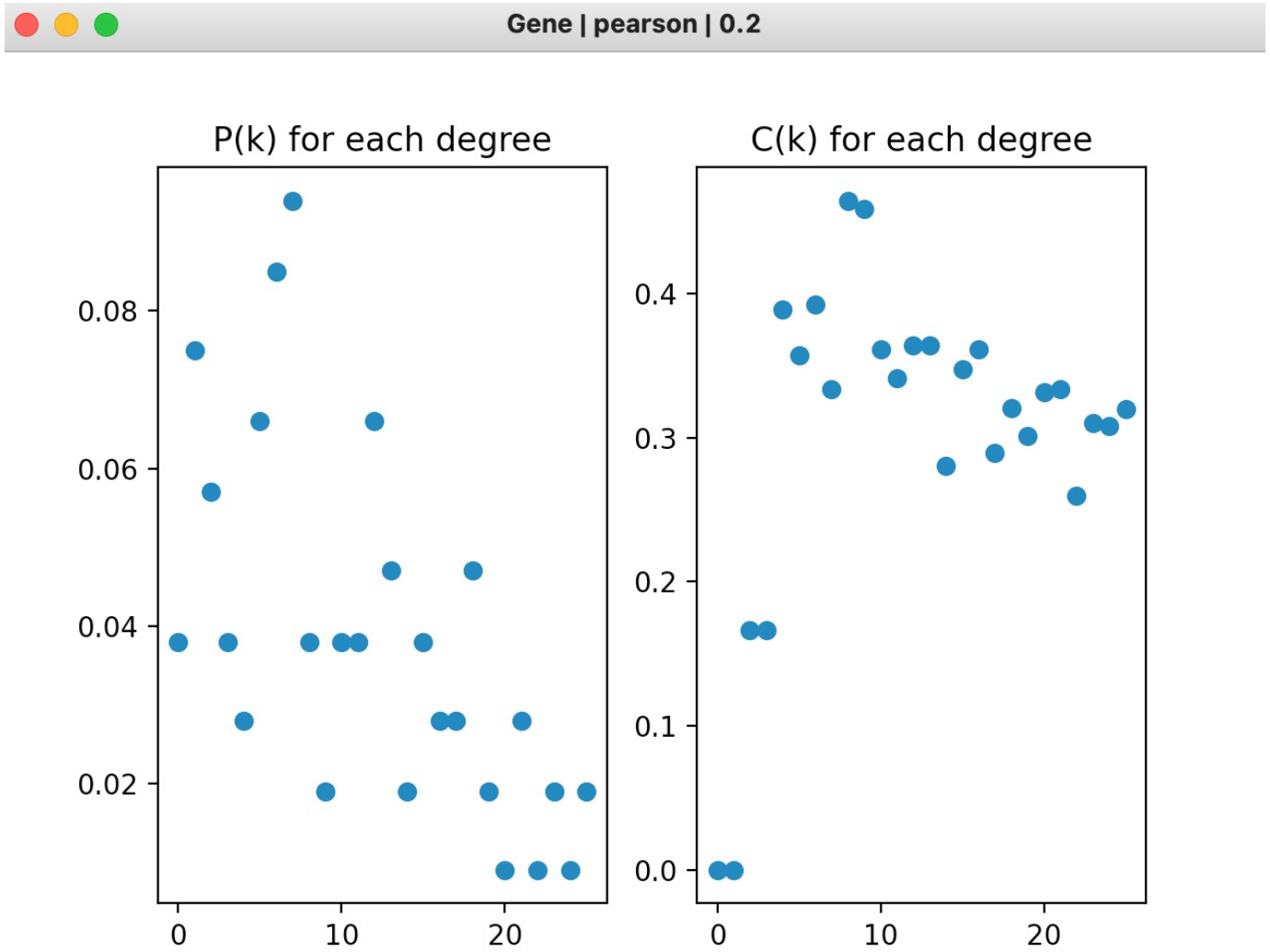
- Está feito no codigo python, mas não houve tempo para fazer o slide

TASK 1.4

0.1	pearson	106	2177	41.08	0.48
0.2	pearson	106	513	9.68	0.30
0.3	pearson	106	72	1.36	0.08
0.4	pearson	106	5	0.09	0.00
0.5	pearson	106	0	0.00	0.00
0.1	spearman	106	2341	44.17	0.50
0.2	spearman	106	654	12.34	0.34
0.3	spearman	106	99	1.87	0.11
0.4	spearman	106	8	0.15	0.02
0.5	spearman	106	0	0.00	0.00

TASK 1.5

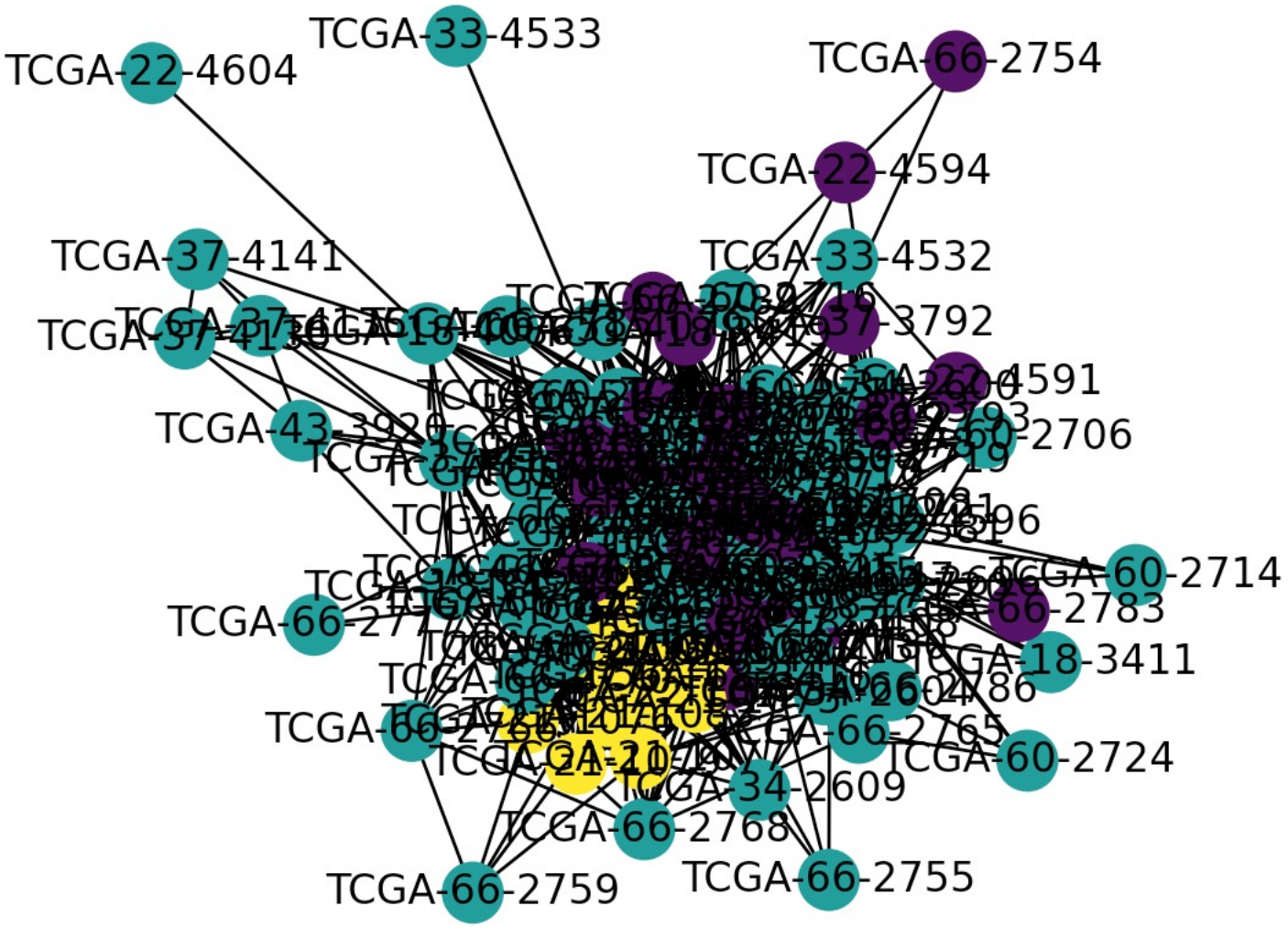
COM BASE NESTES GRAFICOS DECIDIMOS O NETWORK MAIS SEMELHANTE



TASK 2

Cluster obtido usando
- concate dos datasets

- kMeans com k=3



TASK 2

Coluna Survival

que está muito diferente

para os 3 clusters

	cluster1	cluster2	cluster3
count	30.000000	64.000000	12.000000
mean	586.300000	719.531250	1147.750000
std	588.585631	773.106863	1022.473038
min	12.000000	12.000000	53.000000
25%	128.750000	93.250000	462.000000
50%	443.500000	531.000000	922.000000
75%	748.250000	952.000000	1449.250000
max	2466.000000	3469.000000	3724.000000