Cliques analysis

Dario Garcia-Gasulla, in collaboration with David Torrents, Luisa Delgado, Juan Blanco Heredia, Armand Viltalta and Ferran Parés.

February 26, 2018

1

Assuming a window size W, we process each sample creating vertices. Two breaks belong to the same vertex if the distance between them is smaller than W. Thus, the range of a vertex is [min_pos_break - W, max_pos_break + W].

Given these vertices, we compute the graph and subgraphs generated for each sample, and look for cliques. Cliques of 3 vertices are triangles. Cliques of 4 vertices are squares with two diagonals. For each sample, we count how many of those we get, and then we compute the mean and standard deviation per cancer type, and in general. The following tables show that information for several windows sizes.

Some data files are empty. Those are listed in the tables as empty instances.

Table 1: Window size = 100

Cancer type	Valid instances	Empty instances	Mean of 3cliques	StdDev of 3cliques	Mean of 4cliques	StdDev of 4cliques
DLBC	7	0	0.0	0.0	0.0	0.0
PAEN	70	16	0.03	0.17	0.0	0.0
GBM	40	0	1.55	3.19	0.02	0.16
BOCA	60	6	1.07	3.78	0.08	0.64
BLCA	23	0	0.7	2.25	0.09	0.41
UCEC	47	2	0.43	0.87	0.0	0.0
THCA	30	18	0.03	0.18	0.0	0.0
LINC	28	0	0.04	0.19	0.0	0.0
CESC	20	0	0.2	0.68	0.0	0.0
MALY	99	1	0.23	1.14	0.0	0.0
LIHC	53	1	0.21	0.53	0.0	0.0
READ	16	0	0.19	0.53	0.0	0.0
LIRI	258	0	0.17	0.98	0.0	0.0
HNSC	44	0	0.2	0.55	0.0	0.0
STAD	38	0	0.79	2.82	0.0	0.0
SKCM	37	0	4.43	10.14	0.73	2.01
GACA	34	4	0.24	0.55	0.0	0.0
COAD	44	0	0.14	0.4	0.0	0.0
PACA	237	2	0.3	0.94	0.0	0.06
RECA	70	4	0.04	0.2	0.0	0.0
LUSC	48	0	0.48	0.91	0.0	0.0
CLLE	91	5	0.0	0.0	0.0	0.0
LICA	6	0	0.17	0.37	0.0	0.0
CMDI	21	47	0.0	0.0	0.0	0.0
OV	112	0	0.5	0.89	0.0	0.0
BTCA	12	0	0.25	0.6	0.0	0.0
ORCA	12	1	0.0	0.0	0.0	0.0
SARC	34	0	11.0	39.5	2.59	11.84
KIRP	27	6	0.15	0.76	0.0	0.0
LGG	17	2	0.0	0.0	0.0	0.0
LAML	8	18	0.0	0.0	0.0	0.0
PRAD	214	1	0.24	0.71	0.0	0.0
LUAD	36	1	0.14	0.42	0.0	0.0
EOPC	64	5	0.05	0.28	0.0	0.0
BRCA	209	2	0.61	1.26	0.0	0.0
KIRC	37	0	0.19	0.69	0.0	0.0
PBCA	204	32	0.04	0.22	0.0	0.0
MELA	69	0	2.04	3.86	0.06	0.29
KICH	42	3	0.14	0.77	0.0	0.0
ESAD	87	2	0.82	1.31	0.01	0.11
ALL	2605	179	0.55	5.03	0.05	1.41

Table 2: Window size = 1,000

Cancer type	Valid instances	Empty instances	Mean of 3cliques	StdDev of 3cliques	Mean of 4cliques	StdDev of 4cliques
DLBC	7	0	0.29	0.45	0.0	0.0
PAEN	70	16	0.27	0.92	0.0	0.0
GBM	40	0	5.2	7.61	0.5	1.43
BOCA	60	6	5.58	7.91	0.22	0.58
BLCA	23	0	2.17	5.97	0.35	1.63
UCEC	47	2	3.64	5.39	0.21	0.77
THCA	30	18	0.17	0.37	0.0	0.0
LINC	28	0	1.0	1.22	0.0	0.0
CESC	20	0	1.95	2.85	0.35	1.15
MALY	99	1	1.63	8.66	0.67	4.66
LIHC	53	1	5.15	8.56	0.38	1.62
READ	16	0	1.75	3.7	0.0	0.0
LIRI	258	0	1.22	4.24	0.11	1.0
HNSC	44	0	2.41	3.89	0.09	0.36
STAD	38	0	3.97	8.27	0.16	0.71
SKCM	37	0	10.38	15.83	1.76	3.91
GACA	34	4	3.15	5.24	0.5	2.54
COAD	44	0	1.64	3.84	0.09	0.36
PACA	237	2	2.63	5.67	0.31	2.67
RECA	70	4	0.13	0.48	0.0	0.0
LUSC	48	0	5.31	9.43	0.83	4.44
CLLE	91	5	0.03	0.23	0.0	0.0
LICA	6	0	1.5	2.93	0.33	0.75
CMDI	21	47	0.14	0.47	0.0	0.0
OV	112	0	4.71	6.77	0.09	0.39
BTCA	12	0	0.42	0.95	0.0	0.0
ORCA	12	1	0.08	0.28	0.0	0.0
SARC	34	0	35.0	83.32	11.15	45.94
KIRP	27	6	0.59	1.91	0.04	0.19
LGG	17	2	0.18	0.38	0.0	0.0
LAML	8	18	0.0	0.0	0.0	0.0
PRAD	214	1	1.67	3.75	0.02	0.14
LUAD	36	1	1.25	2.02	0.03	0.16
EOPC	64	5	0.91	3.53	0.11	0.87
BRCA	209	2	3.67	6.29	0.07	0.3
KIRC	37	0	0.59	1.62	0.0	0.0
PBCA	204	32	0.12	0.46	0.0	0.0
MELA	69	0	7.19	13.47	0.78	2.8
KICH	42	3	0.86	3.98	0.05	0.3
ESAD	87	2	4.41	6.31	0.23	0.97
ALL	2605	179	2.8	11.88	0.34	5.64

Table 3: Window size = 10,000

Cancer type	Valid instances	Empty instances	Mean of 3cliques	StdDev of 3cliques	Mean of 4cliques	StdDev of 4cliques
DLBC	7	0	0.43	0.73	0.0	0.0
PAEN	70	16	0.39	1.47	0.0	0.0
GBM	40	0	9.62	17.68	1.2	3.74
BOCA	60	6	6.37	9.93	0.28	0.73
BLCA	23	0	2.17	5.97	0.35	1.63
UCEC	47	2	3.6	5.1	0.26	0.96
THCA	30	18	0.27	0.51	0.0	0.0
LINC	28	0	0.96	1.18	0.0	0.0
CESC	20	0	1.2	2.2	0.25	1.09
MALY	99	1	1.17	5.42	0.56	4.13
LIHC	53	1	5.25	8.86	0.57	2.81
READ	16	0	1.62	3.24	0.0	0.0
LIRI	258	0	1.31	5.26	0.17	1.79
HNSC	44	0	1.84	3.91	0.16	0.9
STAD	38	0	3.47	7.67	0.16	0.81
SKCM	37	0	11.78	20.24	1.38	2.95
GACA	34	4	2.62	3.63	0.09	0.37
COAD	44	0	2.0	5.69	0.32	1.41
PACA	237	2	2.27	5.02	0.32	3.0
RECA	70	4	0.4	1.56	0.0	0.0
LUSC	48	0	5.9	10.48	1.17	5.12
CLLE	91	5	0.08	0.63	0.0	0.0
LICA	6	0	0.5	1.12	0.0	0.0
CMDI	21	47	0.14	0.47	0.0	0.0
OV	112	0	3.89	5.77	0.12	0.48
BTCA	12	0	1.0	2.2	0.08	0.28
ORCA	12	1	0.08	0.28	0.0	0.0
SARC	34	0	38.18	82.39	12.79	50.57
KIRP	27	6	1.0	2.99	0.15	0.59
LGG	17	2	0.41	0.97	0.0	0.0
LAML	8	18	0.0	0.0	0.0	0.0
PRAD	214	1	2.96	4.62	0.03	0.19
LUAD	36	1	1.53	2.73	0.08	0.36
EOPC	64	5	1.53	2.98	0.06	0.5
BRCA	209	2	4.25	6.72	0.16	0.74
KIRC	37	0	0.84	2.42	0.05	0.32
PBCA	204	32	0.16	0.84	0.0	0.0
MELA	69	0	9.52	19.9	1.75	7.42
KICH	42	3	0.69	3.37	0.05	0.3
ESAD	87	2	4.64	7.67	0.52	3.19
ALL	2605	179	3.12	12.51	0.42	6.35

Table 4: Window size = 100,000

Cancer type	Valid instances	Empty instances	Mean of 3cliques	StdDev of 3cliques	Mean of 4cliques	StdDev of 4cliques
DLBC	7	0	0.86	2.1	0.0	0.0
PAEN	70	16	0.56	1.75	0.0	0.0
GBM	40	0	18.68	33.07	6.98	24.76
BOCA	60	6	17.58	50.31	3.7	19.07
BLCA	23	0	2.91	5.9	0.35	1.63
UCEC	47	2	2.85	3.8	0.19	0.49
THCA	30	18	0.4	0.84	0.0	0.0
LINC	28	0	1.32	1.63	0.0	0.0
CESC	20	0	1.0	2.24	0.25	1.09
MALY	99	1	1.09	4.53	0.07	0.38
LIHC	53	1	5.17	8.67	0.58	2.02
READ	16	0	1.69	4.77	0.0	0.0
LIRI	258	0	2.0	8.69	0.28	2.4
HNSC	44	0	2.02	4.19	0.23	1.06
STAD	38	0	9.71	38.0	1.5	7.55
SKCM	37	0	16.84	35.49	4.32	14.91
GACA	34	4	4.59	8.91	0.74	3.37
COAD	44	0	2.11	6.14	0.32	1.28
PACA	237	2	2.8	6.47	0.33	2.06
RECA	70	4	0.79	3.72	0.01	0.12
LUSC	48	0	7.1	15.76	1.35	4.6
CLLE	91	5	0.31	2.61	0.0	0.0
LICA	6	0	0.5	1.12	0.0	0.0
CMDI	21	47	0.14	0.47	0.0	0.0
OV	112	0	3.48	4.54	0.2	0.75
BTCA	12	0	2.58	4.37	0.5	1.66
ORCA	12	1	0.08	0.28	0.0	0.0
SARC	34	0	193.38	290.12	142.35	409.7
KIRP	27	6	1.07	3.16	0.22	0.83
LGG	17	2	0.41	0.97	0.0	0.0
LAML	8	18	0.0	0.0	0.0	0.0
PRAD	214	1	6.71	11.01	0.79	3.12
LUAD	36	1	3.61	6.03	0.19	0.52
EOPC	64	5	2.45	4.06	0.19	1.03
BRCA	209	2	10.37	19.64	1.84	7.09
KIRC	37	0	22.41	129.95	4.68	27.56
PBCA	204	32	0.3	1.58	0.05	0.64
MELA	69	0	26.25	71.56	11.71	52.1
KICH	42	3	0.86	4.3	0.05	0.3
ESAD	87	2	6.14	10.48	0.82	3.02
ALL	2605	179	7.55	46.36	2.9	50.64