

First, distribution of components

	CM1	JM1	KC1	MC1	PC1	PC3	PC4
true	48	2102	325	470	76	160	178
false	457	8776	1782	23056	1031	1403	1280
total	505	10878	2107	23526	1107	1563	1458

dense components

	CM1	JM1	KC1	MC1	PC1	PC3	PC4
true	48	1839	302	206	69	137	178
false	208	6888	1308	15360	1007	534	235
total	256	8727	1610	15566	1076	671	413

sparse components

	CM1	JM1	KC1	MC1	PC1	PC3	PC4
true	0	263	23	264	7	23	0
false	249	1888	472	7696	24	869	1045
total	249	2151	495	7960	31	892	1045

test set sizes (approximately)

	CM1	JM1	KC1	MC1	PC1	PC3	PC4
all	50	1087	210	2352	110	156	145
dense	25	872	161	1556	107	67	41

From confusion matrix...

- a is the number of **correct** predictions that an instance is **negative**,
- b is the number of **incorrect** predictions that an instance is **positive**,
- c is the number of **incorrect** of predictions that an instance **negative**, and
- d is the number of **correct** predictions that an instance is **positive**.

		Predicted	
		Negative	Positive
Actual	Negative	a	b
	Positive	c	d

The problem is that we must say correct or incorrect (correct that it's positive, or incorrect that it's positive, etc.)

For instance...

SF = number of "false" modules in sparse components

ST = number of "true" modules in sparse components

TD = testing on dense components

As an example, use MC1 (2352 approx. test set size training & testing on all components)

if $a = SF$ and $b = ST$ then TD correctly makes 7696 predictions for "false" and incorrectly makes 264 predictions about "false" in addition

if $a = SF$ and $c = ST$ then TD correctly makes 7696 predictions for "false" and incorrectly makes 264 predictions about "true" in addition

if $a = SF$ and $d = ST$ then TD correctly makes 7696 predictions for "false" and correctly makes 264 predictions about "true" in addition

if $b = SF$ and $d = ST$ then TD incorrectly makes 7696 predictions for "false" and correctly makes 264 predictions about "true" in addition...

In essence, doing this is not at all accurate.

```

----- PD -----
#key, ties, win, loss, win-loss
TrDTeD, 2, 1, 0, 1, 0.0, 34.0, 73.2, 92.7,100.0,[----- |      +--- ]
TrATeA, 2, 1, 0, 1, 0.0, 31.5, 71.4, 93.9,100.0,[----- |      +--- ]
TrATeD, 2, 1, 0, 1, 0.0, 27.0, 66.7, 95.7,100.0,[----- |      +--- ]
TrDTeA, 0, 0, 3, -3, 0.0, 19.4, 64.7, 90.5,100.0,[----- |      +--- ]

----- PF -----
#key, ties, win, loss, win-loss
TrDTeD, 2, 1, 0, 1, 0.0, 7.3, 26.7, 65.9,100.0,[--- |      +----- ]
TrATeA, 2, 1, 0, 1, 0.0, 6.1, 26.7, 68.4,100.0,[--- |      +----- ]
TrATeD, 2, 1, 0, 1, 0.0, 4.6, 33.3, 74.8,100.0,[-- |      +----- ]
TrDTeA, 0, 0, 3, -3, 0.0, 0.0, 14.0, 52.6,100.0,[ |      +----- ]

```

Normalized Data Set Sizes - All Vs. Dense Components

