Supplementary Information for: Stirling Numbers of Uniform Trees and Related Computational Experiments

Amir Barghi and Daryl DeFord

1 Autocorrelation of Weighted Tree Sampling

One point of concern regarding the sampling-based algorithms for generating trees is the autocorrelation between members of a set of consecutive trees generated by this algorithm. As we see in Tables 1 and 2, using `tsaplots.plot_acf` function from `statsmodels` *Python* library, we generated the autocorrelation plots for sampling from the tree space prioritizing betweenness and closeness, respectively, with 100000 iterations, lags up to 1000000, $p = 0.01$, $q \in \{0.2, 0.4, 0.6, 0.8\}$, and $r \in \{0.2, 0.4, 0.6, 0.8\}$, where each point is displayed in increments of 50. That is, we modified the cycle basis walk on spanning trees to preferentially accept trees with better values under the corresponding metric. The blue ribbons in these plots are made of the 95% confidence intervals where the standard deviation is computed according to Bartlett's formula as provided in the following link.

We experimented with with 4 other starting trees for each of these measures and we did not see any discernible patterns with the ones included in Tables 1 and 2. We have included the GIF files for all of these sets of plots based on initial trees and global measure of centrality on the GitHub repository for comparison. As we see in Tables 1 and 2, as r increases the width of the blue ribbons in these plots for the 95% confidence intervals decreases; however, we do not see a discernible pattern as q increases.

Table 1: Autocorrelation plots for uniform sampling from the tree space based on betweenness with 100000 iterations, lags up to 1000000, $p = 0.01$, $q \in \{0.2, 0.4, 0.6, 0.8\}$, and $r \in \{0.2, 0.4, 0.6, 0.8\}$, where each point is displayed in increments of 50.

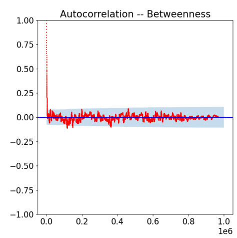
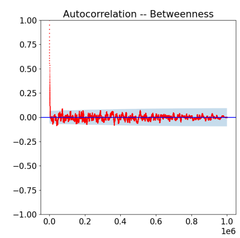
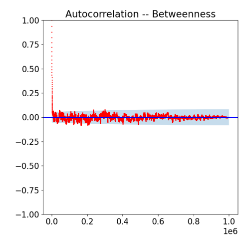
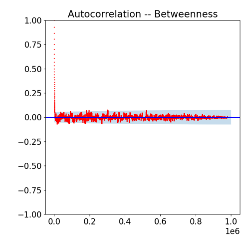
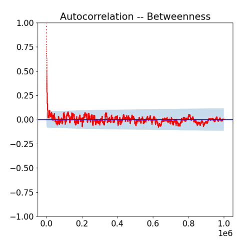
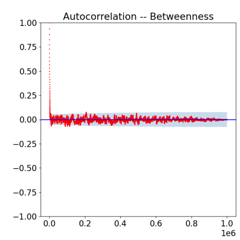
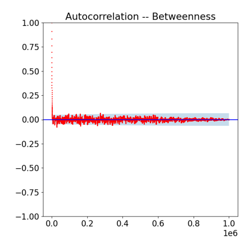
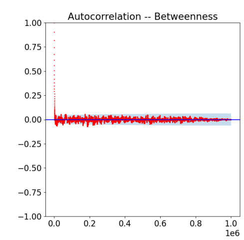
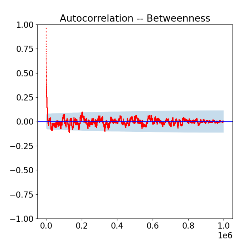
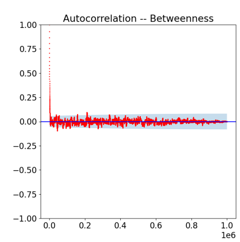
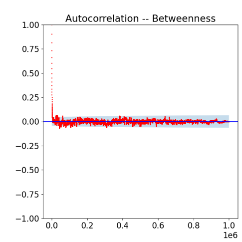
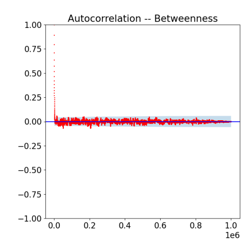
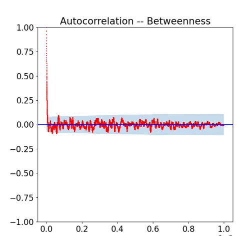
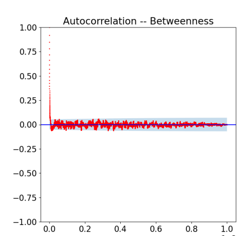
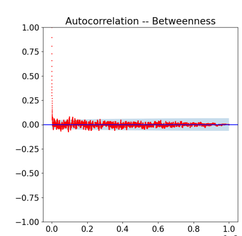
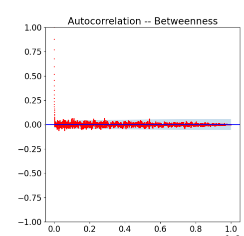
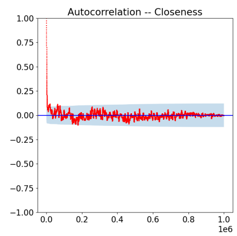
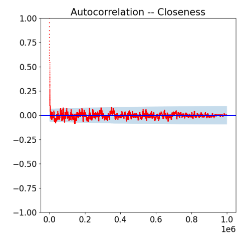
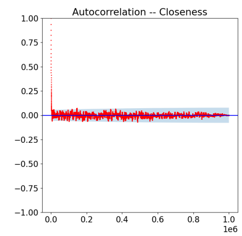
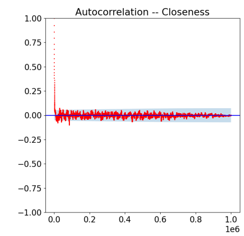
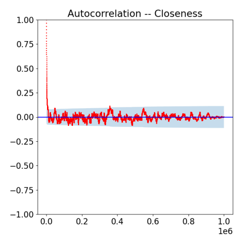
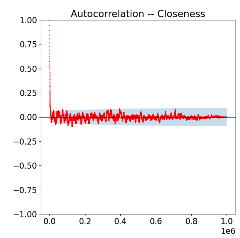
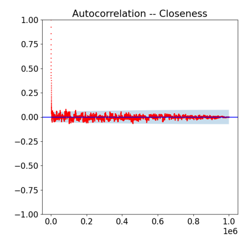
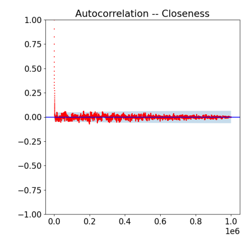
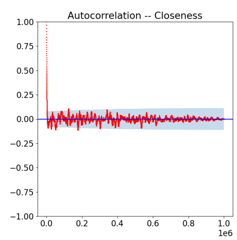
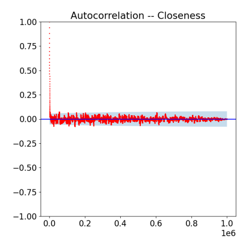
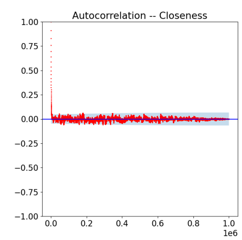
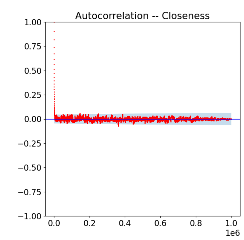
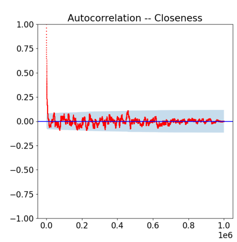
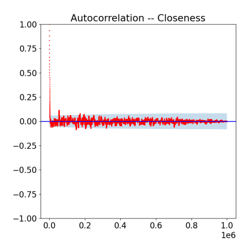
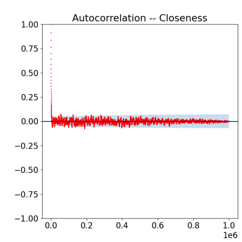
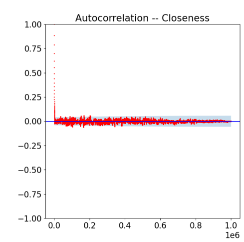
$q \setminus r$	0.2	0.4	0.6	0.8
0.2				
0.4				
0.6				
0.8				

Table 2: Autocorrelation plots for uniform sampling from the tree space based on closeness with 100000 iterations, lags up to 1000000, $p = 0.01$, $q \in \{0.2, 0.4, 0.6, 0.8\}$, and $r \in \{0.2, 0.4, 0.6, 0.8\}$, where each point is displayed in increments of 50.

<table>
<tr><th>$q \setminus r$</th><th>0.2</th><th>0.4</th><th>0.6</th><th>0.8</th></tr>
<tr><td>0.2</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
<tr><td>0.4</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
<tr><td>0.6</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
<tr><td>0.8</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
</table>

2 Validating the Probabilistic Approach Using Wilson's Sampling Algorithm

Table 3: Statistics from 100 samples from the probabilistic approach with $m = 10,000$ for the k-th Stirling number on uniformly sampled trees of order $n \in \{7, 8, \dots, 19\}$ and $n-2 \geq k \geq \lceil \frac{n}{2} \rceil$

n	k	Mean	SD	Skewness
7	4	0.000380	0.0590	-0.0661
7	5	0.00412	0.0666	0.0438
8	4	-0.00198	0.0390	0.386
8	5	0.0219	0.141	0.0250
8	6	0.0213	0.0930	-0.0851
9	5	-0.00873	0.118	-0.0909
9	6	-0.00200	0.242	-0.330
9	7	0.00822	0.125	0.469
10	5	0.00363	0.0618	1.56
10	6	0.0325	0.326	0.299
10	7	0.0415	0.421	0.152
10	8	-0.0348	0.158	-0.222
11	6	0.00352	0.230	-0.681
11	7	0.103	0.529	-0.160
11	8	-0.0348	0.652	-0.169
11	9	0.000100	0.176	0.270
12	6	0.0125	0.118	0.462
12	7	-0.108	0.561	-0.454
12	8	0.192	0.971	0.287
12	9	0.109	0.791	0.0619
12	10	-0.0221	0.230	-0.251
13	7	-0.0381	0.439	-1.16
13	8	-0.100	1.32	0.399
13	9	-0.166	1.75	-0.300
13	10	0.147	1.04	-0.149
13	11	-0.0161	0.267	0.281
14	7	0.0205	0.146	0.537
14	8	0.242	1.07	0.191
14	9	-0.0753	2.36	-0.157
14	10	-0.155	2.93	-0.101
14	11	0.0128	1.46	0.197
14	12	0.0270	0.335	0.873
15	8	0.00390	1.33	0.0955
15	9	0.170	3.35	0.142
15	10	0.0157	4.49	0.358
15	11	0.256	3.71	0.367
15	12	-0.159	1.37	-0.298
15	13	0.0169	0.317	0.0832
16	8	0.0263	0.336	-2.00
16	9	-0.189	2.66	-0.0130
16	10	0.812	5.22	0.348
16	11	-0.434	6.26	0.208
16	12	-0.401	6.21	-0.0340
16	13	-0.133	2.20	-0.132
16	14	-0.0288	0.419	0.00154
17	9	-0.174	1.26	-0.800
17	10	0.569	5.74	-0.146
17	11	-0.273	11.5	-0.281
17	12	0.384	11.0	0.0288
17	13	-0.965	7.17	0.436
17	14	0.0135	2.54	-0.284
17	15	0.00844	0.461	0.249
18	9	-0.00293	0.471	-5.38
18	10	0.417	4.28	0.155
18	11	2.09	15.0	-0.114
18	12	-0.485	22.5	-0.337
18	13	3.14	21.9	0.225
18	14	-0.194	11.1	-0.185
18	15	0.0178	3.54	-0.0625
18	16	0.0469	0.499	0.307
19	10	0.0204	2.54	-1.50
19	11	-0.0856	11.6	-1.30
19	12	-0.346	23.8	0.170
19	13	1.80	39.1	0.0508
19	14	2.74	26.6	-0.0568
19	15	0.743	13.1	0.167
19	16	0.360	3.85	0.0763
19	17	-0.00189	0.605	-0.397

3 Tree-based Classification

Table 4: Accuracy scores and Matthews correlation for different classification methods with closeness, betweenness, and $\begin{bmatrix} T \\ k \end{bmatrix}$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as predictors for all non-isomorphic trees of order $n = 12$

Method	Criterion	Pruning	Train Score	Test Score	Train Corr.	Test Corr.
DecisionTreeClassifier	Gini	0	1.0	0.93841	1.0	0.87676
DecisionTreeClassifier	Gini	1	0.93455	0.94203	0.87445	0.88653
DecisionTreeClassifier	Entropy	0	1.0	0.92754	1.0	0.85505
DecisionTreeClassifier	Entropy	1	0.92364	0.90942	0.8531	0.82917
ExtraTreeClassifier	Gini	0	1.0	0.93478	1.0	0.86939
ExtraTreeClassifier	Gini	1	0.93455	0.94203	0.87445	0.88653
ExtraTreeClassifier	Entropy	0	1.0	0.92754	1.0	0.85509
ExtraTreeClassifier	Entropy	1	0.86545	0.88768	0.7307	0.77507
Bagging	NA	NA	0.98909	0.9529	0.97818	0.90667
RandomForestClassifier	Gini	0	1.0	0.96377	1.0	0.9275
RandomForestClassifier	Gini	1	0.94545	0.95652	0.89093	0.91393
RandomForestClassifier	Entropy	0	1.0	0.94928	1.0	0.89857
RandomForestClassifier	Entropy	1	0.94182	0.95652	0.88419	0.91393
ExtraTreesClassifier	Gini	0	1.0	0.96377	1.0	0.92744
ExtraTreesClassifier	Gini	1	0.92	0.92754	0.83984	0.85539
ExtraTreesClassifier	Entropy	0	1.0	0.9529	1.0	0.90581
ExtraTreesClassifier	Entropy	1	0.92364	0.92391	0.8472	0.84838

Table 5: Accuracy scores and Matthews correlation for different classification methods with closeness, betweenness, and $\begin{bmatrix} T \\ k \end{bmatrix}$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as predictors for a sample of size 500 of non-isomorphic trees of order $n = 18$

Method	Criterion	Pruning	Train Score	Test Score	Train Corr.	Test Corr.
DecisionTreeClassifier	Gini	0	1.0	0.884	1.0	0.76776
DecisionTreeClassifier	Gini	1	0.892	0.876	0.78384	0.75163
DecisionTreeClassifier	Entropy	0	1.0	0.888	1.0	0.7753
DecisionTreeClassifier	Entropy	1	0.912	0.884	0.83149	0.76835
ExtraTreeClassifier	Gini	0	1.0	0.856	1.0	0.71109
ExtraTreeClassifier	Gini	1	0.892	0.876	0.78384	0.75163
ExtraTreeClassifier	Entropy	0	1.0	0.872	1.0	0.74337
ExtraTreeClassifier	Entropy	1	0.868	0.852	0.73517	0.70862
Bagging	NA	NA	0.996	0.896	0.992	0.79138
RandomForestClassifier	Gini	0	1.0	0.896	1.0	0.79135
RandomForestClassifier	Gini	1	0.888	0.872	0.77681	0.74801
RandomForestClassifier	Entropy	0	1.0	0.888	1.0	0.7753
RandomForestClassifier	Entropy	1	0.928	0.888	0.85553	0.7761
ExtraTreesClassifier	Gini	0	1.0	0.892	1.0	0.78384
ExtraTreesClassifier	Gini	1	0.868	0.868	0.73676	0.73933
ExtraTreesClassifier	Entropy	0	1.0	0.896	1.0	0.79135
ExtraTreesClassifier	Entropy	1	0.912	0.876	0.82637	0.75318

Table 6: Accuracy scores and Matthews correlation for different classification methods with closeness, betweenness, and $\left[{T \atop k}\right]$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as predictors for a sample of size 500 of non-isomorphic trees of order $n = 18$ generated using uniform sampling

Method	Criterion	Pruning	Train Score	Test Score	Train Corr.	Test Corr.
DecisionTreeClassifier	Gini	0	1.0	0.86	1.0	0.71902
DecisionTreeClassifier	Gini	1	0.852	0.844	0.70437	0.69409
DecisionTreeClassifier	Entropy	0	1.0	0.856	1.0	0.71594
DecisionTreeClassifier	Entropy	1	0.964	0.856	0.92814	0.71269
ExtraTreeClassifier	Gini	0	1.0	0.84	1.0	0.68529
ExtraTreeClassifier	Gini	1	0.824	0.852	0.67385	0.71456
ExtraTreeClassifier	Entropy	0	1.0	0.816	1.0	0.63058
ExtraTreeClassifier	Entropy	1	0.832	0.776	0.67639	0.57864
Bagging	NA	NA	1.0	0.864	1.0	0.72703
RandomForestClassifier	Gini	0	0.996	0.896	0.992	0.79213
RandomForestClassifier	Gini	1	0.88	0.848	0.76152	0.70296
RandomForestClassifier	Entropy	0	1.0	0.9	1.0	0.80226
RandomForestClassifier	Entropy	1	0.964	0.904	0.92902	0.81216
ExtraTreesClassifier	Gini	0	1.0	0.9	1.0	0.80126
ExtraTreesClassifier	Gini	1	0.84	0.884	0.67925	0.77618
ExtraTreesClassifier	Entropy	0	1.0	0.876	1.0	0.75238
ExtraTreesClassifier	Entropy	1	0.904	0.876	0.80782	0.76011

4 Support Vector Classification

Table 7: Accuracy scores and Matthews correlation for support vector classification methods with closeness, betweenness, and $\left[{T \atop k}\right]$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as predictors for all non-isomorphic trees of order $n = 12$

Method	Criterion	Pruning	Train Score	Test Score	Train Corr.	Test Corr.
Linear SVC	NA	NA	0.89818	0.90942	0.79668	0.8189
Quadratic SVC	NA	NA	0.95636	0.94928	0.91269	0.89842

Table 8: Accuracy scores and Matthews correlation for support vector classification methods with closeness, betweenness, and $\left[{T \atop k}\right]$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as predictors for a sample of size 500 of non-isomorphic trees of order $n = 18$

Method	Criterion	Pruning	Train Score	Test Score	Train Corr.	Test Corr.
Linear SVC	NA	NA	0.852	0.868	0.70287	0.73531
Quadratic SVC	NA	NA	0.9	0.86	0.79933	0.71938

Table 9: Accuracy scores and Matthews correlation for support vector classification methods with closeness, betweenness, and $\left[{T \atop k}\right]$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as predictors for a sample of size 500 of non-isomorphic trees of order $n = 18$ generated using uniform sampling

Method	Criterion	Pruning	Train Score	Test Score	Train Corr.	Test Corr.
Linear SVC	NA	NA	0.896	0.868	0.7917	0.73716
Quadratic SVC	NA	NA	0.9	0.876	0.79987	0.75418

5 Regression

Table 10: Different scores for various regression methods with $\log_{10}(P(T;2,1)$, closeness, betweenness, and class as predictors and $\left[{T \atop k} \right]$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as response for all non-isomorphic trees of order $n = 12$

Method	k	Train Score	Test Score	Train EVS	Test EVS
LinearRegression	6	0.13958	0.12107	0.13958	0.12714
LinearRegression	7	0.55036	0.56972	0.55036	0.5704
LinearRegression	8	0.82921	0.85098	0.82921	0.85131
LinearRegression	9	0.90362	0.9143	0.90362	0.916
LinearRegression	10	0.85581	0.80199	0.85581	0.8056
RidgeRegression	6	0.13285	0.11073	0.13285	0.11533
RidgeRegression	7	0.51905	0.50996	0.51905	0.51001
RidgeRegression	8	0.78598	0.79391	0.78598	0.79393
RidgeRegression	9	0.86939	0.89259	0.86939	0.89316
RidgeRegression	10	0.84817	0.81435	0.84817	0.81739
LassoRegression	6	0.0	-0.00165	0.0	0.0
LassoRegression	7	0.45861	0.43879	0.45861	0.43884
LassoRegression	8	0.73281	0.73048	0.73281	0.73073
LassoRegression	9	0.85104	0.87699	0.85104	0.87734
LassoRegression	10	0.81973	0.77018	0.81973	0.77349
ElasticNet	6	0.0	-0.00165	0.0	0.0
ElasticNet	7	0.45926	0.43836	0.45926	0.4384
ElasticNet	8	0.7321	0.73214	0.7321	0.73241
ElasticNet	9	0.85022	0.87607	0.85022	0.87639
ElasticNet	10	0.82043	0.77198	0.82043	0.77533
Quadratic	6	0.20854	0.18904	-2211.23588	-3396.58377
Quadratic	7	0.66974	0.71566	-4.53711	-5.92245
Quadratic	8	0.86704	0.89947	0.73299	0.72764
Quadratic	9	0.9416	0.95824	0.85022	0.87639
Quadratic	10	0.94721	0.96705	-5.98409	-3.54293
SGDRegressor	6	-32750.58111	-38668.59559	-2211.23588	-3396.58377
SGDRegressor	7	-103.36196	-94.4036	-4.53711	-5.92245
SGDRegressor	8	-1.8125	-1.43	0.73299	0.72764
SGDRegressor	9	0.85022	0.87607	0.85022	0.87639
SGDRegressor	10	-16.63892	-9.96178	-5.98409	-3.54293

Table 11: Different scores for various regression methods with $\log_{10}(P(T;2,1)$, closeness, betweenness, and class as predictors and $\begin{bmatrix} T \\ k \end{bmatrix}$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as response for a sample of size 500 of non-isomorphic trees of order $n = 18$

Method	k	Train Score	Test Score	Train EVS	Test EVS
LinearRegression	9	0.04004	0.06002	0.04004	0.06295
LinearRegression	10	0.2578	0.27202	0.2578	0.27893
LinearRegression	11	0.51556	0.52358	0.51556	0.52617
LinearRegression	12	0.71913	0.73921	0.71913	0.74057
LinearRegression	13	0.84149	0.86708	0.84149	0.86746
LinearRegression	14	0.88017	0.90363	0.88017	0.9037
LinearRegression	15	0.86546	0.89044	0.86546	0.89045
LinearRegression	16	0.81859	0.85547	0.81859	0.8555
RidgeRegression	9	0.03289	0.04404	0.03289	0.04725
RidgeRegression	10	0.24556	0.247	0.24556	0.25451
RidgeRegression	11	0.48492	0.48516	0.48492	0.4884
RidgeRegression	12	0.67736	0.69483	0.67736	0.69679
RidgeRegression	13	0.79677	0.81848	0.79677	0.81923
RidgeRegression	14	0.84042	0.85336	0.84042	0.85362
RidgeRegression	15	0.83458	0.84412	0.83458	0.84424
RidgeRegression	16	0.79921	0.81885	0.79921	0.81897
LassoRegression	9	0.0	-0.00529	0.0	0.0
LassoRegression	10	0.21456	0.22894	0.21456	0.23741
LassoRegression	11	0.42754	0.44882	0.42754	0.45246
LassoRegression	12	0.68765	0.70526	0.68765	0.70718
LassoRegression	13	0.80995	0.8315	0.80995	0.8322
LassoRegression	14	0.84007	0.85336	0.84007	0.85363
LassoRegression	15	0.82351	0.8325	0.82351	0.83265
LassoRegression	16	0.79176	0.81329	0.79176	0.81347

Table 12: Different scores for various regression methods with $\log_{10}(P(T;2,1)$, closeness, betweenness, and class as predictors and $\begin{bmatrix} T \\ k \end{bmatrix}$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as response for a sample of size 500 of non-isomorphic trees of order $n = 18$

Method	k	Train Score	Test Score	Train EVS	Test EVS
ElasticNet	9	0.0	-0.00529	0.0	0.0
ElasticNet	10	0.21497	0.23138	0.21497	0.23967
ElasticNet	11	0.42397	0.4472	0.42397	0.45096
ElasticNet	12	0.61372	0.64995	0.61372	0.65234
ElasticNet	13	0.74975	0.78217	0.74975	0.78315
ElasticNet	14	0.81401	0.83017	0.81401	0.83053
ElasticNet	15	0.82345	0.83267	0.82345	0.83283
ElasticNet	16	0.79201	0.81313	0.79201	0.81329
Quadratic	9	0.08376	0.13838	-117320.55672	-76041.64184
Quadratic	10	0.3801	0.46307	-72.05863	-47.40627
Quadratic	11	0.66417	0.69875	0.32732	0.42638
Quadratic	12	0.80725	0.82679	0.35951	0.36154
Quadratic	13	0.87327	0.8889	0.29051	0.29365
Quadratic	14	0.89519	0.91035	0.41809	0.4248
Quadratic	15	0.89819	0.91062	0.82345	0.83283
Quadratic	16	0.88915	0.89481	-57.04841	-64.3344
SGDRegressor	9	-7395672.57298	-4419046.29853	-117320.55672	-76041.64184
SGDRegressor	10	-4989.73181	-3126.1376	-72.05863	-47.40627
SGDRegressor	11	-44.8419	-31.93281	0.32732	0.42638
SGDRegressor	12	-0.55429	-0.21132	0.35951	0.36154
SGDRegressor	13	0.00935	-0.06662	0.29051	0.29365
SGDRegressor	14	0.78867	0.8154	0.78991	0.81581
SGDRegressor	15	0.82345	0.83267	0.82345	0.83283
SGDRegressor	16	-2112.74264	-2194.25821	-57.04841	-64.3344

Table 13: Different scores for various regression methods with $\log_{10}(P(T;2,1)$, closeness, betweenness, and class as predictors and $\left[{T \atop k} \right]$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as response for a sample of size 500 of non-isomorphic trees of order $n = 18$ generated using uniform sampling

Method	k	Train Score	Test Score	Train EVS	Test EVS
LinearRegression	9	0.08897	0.05384	0.08897	0.05715
LinearRegression	10	0.3601	0.32037	0.3601	0.32673
LinearRegression	11	0.56487	0.55916	0.56487	0.56091
LinearRegression	12	0.71863	0.71352	0.71863	0.71412
LinearRegression	13	0.81612	0.81248	0.81612	0.81287
LinearRegression	14	0.86837	0.86104	0.86837	0.86142
LinearRegression	15	0.88418	0.869	0.88418	0.86938
LinearRegression	16	0.87212	0.85174	0.87212	0.85204
RidgeRegression	9	0.06423	0.0549	0.06423	0.05785
RidgeRegression	10	0.31307	0.32593	0.31307	0.33154
RidgeRegression	11	0.51467	0.53814	0.51467	0.53949
RidgeRegression	12	0.6705	0.67828	0.6705	0.67866
RidgeRegression	13	0.77511	0.77597	0.77511	0.77619
RidgeRegression	14	0.83713	0.83307	0.83713	0.83329
RidgeRegression	15	0.86283	0.85345	0.86283	0.85368
RidgeRegression	16	0.85894	0.84634	0.85894	0.84651
LassoRegression	9	0.0	-0.00087	0.0	0.0
LassoRegression	10	0.29234	0.30624	0.29234	0.31206
LassoRegression	11	0.49424	0.51482	0.49424	0.51646
LassoRegression	12	0.65951	0.66425	0.65951	0.66472
LassoRegression	13	0.76675	0.76532	0.76675	0.76559
LassoRegression	14	0.83088	0.8262	0.83088	0.82647
LassoRegression	15	0.85799	0.84945	0.85799	0.84966
LassoRegression	16	0.84367	0.83285	0.84367	0.83287

Table 14: Different scores for various regression methods with $\log_{10}(P(T;2,1)$, closeness, betweenness, and class as predictors and $\left[{T \atop k}\right]$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as response for a sample of size 500 of non-isomorphic trees of order $n = 18$ generated using uniform sampling

Method	k	Train Score	Test Score	Train EVS	Test EVS
ElasticNet	9	0.0	-0.00087	0.0	0.0
ElasticNet	10	0.29247	0.30621	0.29247	0.31209
ElasticNet	11	0.49438	0.51515	0.49438	0.51669
ElasticNet	12	0.65292	0.6595	0.65292	0.65998
ElasticNet	13	0.76225	0.76256	0.76225	0.76283
ElasticNet	14	0.82954	0.82557	0.82954	0.8258
ElasticNet	15	0.85773	0.84918	0.85773	0.84936
ElasticNet	16	0.84619	0.83589	0.84619	0.83592
Method	k	Train Score	Test Score	Train EVS	Test EVS
Quadratic	9	0.16998	0.01776	-6731.47459	-7931.10983
Quadratic	10	0.46845	0.32863	-5.42954	-6.2534
Quadratic	11	0.64128	0.59245	0.41915	0.44271
Quadratic	12	0.76229	0.73372	0.23965	0.24575
Quadratic	13	0.83971	0.82047	0.22388	0.2272
Quadratic	14	0.88477	0.8686	0.38344	0.38486
Quadratic	15	0.90273	0.88763	0.85773	0.84936
Quadratic	16	0.90023	0.88975	-67.52073	-64.27628
SGDRegressor	9	-1818115.52563	-2011671.98016	-6731.47459	-7931.10983
SGDRegressor	10	-2219.9779	-2362.99147	-5.42954	-6.2534
SGDRegressor	11	-28.85835	-29.07875	0.41915	0.44271
SGDRegressor	12	0.22408	0.24056	0.23965	0.24575
SGDRegressor	13	-3.39823	-3.48302	0.22388	0.2272
SGDRegressor	14	-10.57673	-10.29922	0.38344	0.38486
SGDRegressor	15	0.85773	0.84918	0.85773	0.84936
SGDRegressor	16	-11243.71688	-10143.426	-67.52073	-64.27628

6 Tree-based Regression

Table 15: Different scores for various tree-based methods with $\log_{10}(P(T;2,1)$, closeness, betweenness, and class as predictors and $\begin{bmatrix} T \\ k \end{bmatrix}$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as response for all non-isomorphic trees of order $n = 12$

Method	Pruning	k	Train Score	Test Score	Train EVS	Test EVS
DecisionTreeRegressor	0	6	0	-0.00165	1	-0.5116
DecisionTreeRegressor	0	7	0	-0.00154	1	0.39931
DecisionTreeRegressor	0	8	0	-0.00348	1	0.74175
DecisionTreeRegressor	0	9	0	-0.00061	1	0.878
DecisionTreeRegressor	0	10	0	-0.00056	1	0.88265
DecisionTreeRegressor	1	6	0.64682	-0.04145	0.64682	-0.0403
DecisionTreeRegressor	1	7	0.65092	0.64405	0.65092	0.64767
DecisionTreeRegressor	1	8	0.80918	0.82022	0.80918	0.8208
DecisionTreeRegressor	1	9	0.7977	0.7942	0.7977	0.79628
DecisionTreeRegressor	1	10	0.83642	0.8018	0.83642	0.80206
ExtraTreeRegressor	0	6	0.14775	0.14974	1	-0.6099
ExtraTreeRegressor	0	7	0.47595	0.49433	1	0.30985
ExtraTreeRegressor	0	8	0.56732	0.56735	1	0.69405
ExtraTreeRegressor	0	9	0.59366	0.57804	1	0.8685
ExtraTreeRegressor	0	10	0.5576	0.49368	1	0.86261
ExtraTreeRegressor	1	6	0.4447	0.01845	0.4447	0.02018
ExtraTreeRegressor	1	7	0.48565	0.55644	0.48565	0.56094
ExtraTreeRegressor	1	8	0.55472	0.56813	0.55472	0.56816
ExtraTreeRegressor	1	9	0.74425	0.79456	0.74425	0.79456
ExtraTreeRegressor	1	10	0.72	0.68131	0.72	0.68197
Bagging	NA	6	0.84028	-0.09897	0.84031	-0.0945
Bagging	NA	7	0.91596	0.63513	0.91629	0.6363
Bagging	NA	8	0.96455	0.84023	0.96461	0.84064
Bagging	NA	9	0.9788	0.90819	0.9788	0.90906
Bagging	NA	10	0.9734	0.8944	0.97366	0.89709

Table 16: Different scores for various tree-based methods with $\log_{10}(P(T;2,1)$, closeness, betweenness, and class as predictors and $\left[{T \atop k}\right]$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as response for all non-isomorphic trees of order $n = 12$

Method	Pruning	k	Train Score	Test Score	Train EVS	Test EVS
RandomForestRegressor	0	6	0.85136	0.02653	0.85143	0.0295
RandomForestRegressor	0	7	0.94131	0.62677	0.94133	0.62755
RandomForestRegressor	0	8	0.97187	0.84723	0.97189	0.84723
RandomForestRegressor	0	9	0.98071	0.91172	0.98075	0.91263
RandomForestRegressor	0	10	0.98446	0.88945	0.98452	0.89175
RandomForestRegressor	1	6	0.7218	0.11583	0.72191	0.11961
RandomForestRegressor	1	7	0.68679	0.681	0.6868	0.6839
RandomForestRegressor	1	8	0.80041	0.81145	0.80042	0.81245
RandomForestRegressor	1	9	0.82265	0.82049	0.82269	0.82197
RandomForestRegressor	1	10	0.88392	0.82502	0.88393	0.82631
ExtraTreesRegressor	0	6	1	-0.07947	1	-0.07486
ExtraTreesRegressor	0	7	1	0.60749	1	0.60815
ExtraTreesRegressor	0	8	1	0.83322	1	0.83336
ExtraTreesRegressor	0	9	0.99999	0.91283	0.99999	0.91303
ExtraTreesRegressor	0	10	1	0.9142	1	0.91511
ExtraTreesRegressor	1	6	0.34398	0.18782	0.34398	0.19831
ExtraTreesRegressor	1	7	0.56552	0.5799	0.56552	0.58223
ExtraTreesRegressor	1	8	0.75558	0.77591	0.75558	0.77643
ExtraTreesRegressor	1	9	0.79986	0.81911	0.79986	0.81921
ExtraTreesRegressor	1	10	0.79632	0.75925	0.79632	0.75959

Table 17: Different scores for various tree-based methods with $\log_{10}(P(T;2,1)$, closeness, betweenness, and class as predictors and $\begin{bmatrix} T \\ k \end{bmatrix}$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as response for a sample of size 500 of non-isomorphic trees of order $n = 18$

Method	Pruning	k	Train Score	Test Score	Train EVS	Test EVS
DecisionTreeRegressor	0	9	0	-0.00465	1	-0.54132
DecisionTreeRegressor	0	10	0	-0.01492	1	0.41453
DecisionTreeRegressor	0	11	0	-0.02512	1	0.52418
DecisionTreeRegressor	0	12	0	-0.02618	1	0.64173
DecisionTreeRegressor	0	13	0	-0.0232	1	0.74065
DecisionTreeRegressor	0	14	0	-0.01884	1	0.74874
DecisionTreeRegressor	0	15	0	-0.01356	1	0.76847
DecisionTreeRegressor	0	16	0	-0.00765	1	0.75666
DecisionTreeRegressor	1	9	1	-0.67872	1	-0.6782
DecisionTreeRegressor	1	10	0.65553	0.4225	0.65553	0.4233
DecisionTreeRegressor	1	11	0.64248	0.64779	0.64248	0.64809
DecisionTreeRegressor	1	12	0.77588	0.71853	0.77588	0.72152
DecisionTreeRegressor	1	13	0.81426	0.7873	0.81426	0.79089
DecisionTreeRegressor	1	14	0.75671	0.76274	0.75671	0.77079
DecisionTreeRegressor	1	15	0.75715	0.7296	0.75715	0.73502
DecisionTreeRegressor	1	16	0.7876	0.74801	0.7876	0.74915
ExtraTreeRegressor	0	9	0.15035	0.00191	1	-0.27273
ExtraTreeRegressor	0	10	0.43513	0.41503	1	0.34463
ExtraTreeRegressor	0	11	0.4534	0.51338	1	0.55004
ExtraTreeRegressor	0	12	0.52557	0.52463	1	0.69089
ExtraTreeRegressor	0	13	0.56806	0.59171	1	0.69355
ExtraTreeRegressor	0	14	0.5681	0.52539	1	0.77871
ExtraTreeRegressor	0	15	0.54067	0.52246	1	0.75528
ExtraTreeRegressor	0	16	0.50088	0.4653	0.99977	0.7531
ExtraTreeRegressor	1	9	1	-0.42045	1	-0.40754
ExtraTreeRegressor	1	10	0.60832	0.4858	0.60832	0.48656
ExtraTreeRegressor	1	11	0.60988	0.64387	0.60988	0.64559
ExtraTreeRegressor	1	12	0.6545	0.65126	0.6545	0.6523
ExtraTreeRegressor	1	13	0.74688	0.73766	0.74688	0.74431
ExtraTreeRegressor	1	14	0.69637	0.71055	0.69637	0.71324
ExtraTreeRegressor	1	15	0.69514	0.67783	0.69514	0.68764
ExtraTreeRegressor	1	16	0.65165	0.67041	0.65165	0.67191
Bagging	NA	9	0.75714	-0.09375	0.75735	-0.08879
Bagging	NA	10	0.89316	0.5495	0.89316	0.55217
Bagging	NA	11	0.92803	0.69051	0.92809	0.69757
Bagging	NA	12	0.95229	0.77297	0.95247	0.77681
Bagging	NA	13	0.96163	0.82512	0.96165	0.82675
Bagging	NA	14	0.96926	0.8613	0.96929	0.86139
Bagging	NA	15	0.96587	0.85212	0.96604	0.8524
Bagging	NA	16	0.95214	0.80449	0.95235	0.80616

Table 18: Different scores for various regression methods with $\log_{10}(P(T;2,1)$, closeness, betweenness, and class as predictors and $\begin{bmatrix} T \\ k \end{bmatrix}$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as response for a sample of size 500 of non-isomorphic trees of order $n = 18$

Method	Pruning	k	Train Score	Test Score	Train EVS	Test EVS
RandomForestRegressor	0	9	0.81094	-0.10542	0.81094	-0.1024
RandomForestRegressor	0	10	0.91686	0.5702	0.91688	0.57253
RandomForestRegressor	0	11	0.9514	0.70896	0.95149	0.71406
RandomForestRegressor	0	12	0.96588	0.79557	0.96601	0.79953
RandomForestRegressor	0	13	0.97248	0.84121	0.97251	0.84216
RandomForestRegressor	0	14	0.97594	0.86128	0.97598	0.8614
RandomForestRegressor	0	15	0.97607	0.84373	0.97609	0.84381
RandomForestRegressor	0	16	0.96905	0.82081	0.96914	0.82138
RandomForestRegressor	1	9	0.80441	-0.02738	0.80441	-0.02367
RandomForestRegressor	1	10	0.696	0.54106	0.69638	0.54247
RandomForestRegressor	1	11	0.71935	0.68813	0.71947	0.69076
RandomForestRegressor	1	12	0.80678	0.75245	0.80678	0.75593
RandomForestRegressor	1	13	0.8239	0.81741	0.82391	0.81936
RandomForestRegressor	1	14	0.80592	0.81496	0.80593	0.81888
RandomForestRegressor	1	15	0.80496	0.79344	0.80497	0.79685
RandomForestRegressor	1	16	0.81934	0.75287	0.81937	0.75338
ExtraTreesRegressor	0	9	1	-0.22035	1	-0.21757
ExtraTreesRegressor	0	10	1	0.56388	1	0.56478
ExtraTreesRegressor	0	11	1	0.72141	1	0.7241
ExtraTreesRegressor	0	12	1	0.7967	1	0.79838
ExtraTreesRegressor	0	13	1	0.83847	1	0.83926
ExtraTreesRegressor	0	14	1	0.86075	1	0.8608
ExtraTreesRegressor	0	15	1	0.87028	1	0.87029
ExtraTreesRegressor	0	16	0.99999	0.8326	0.99999	0.83346
ExtraTreesRegressor	1	9	1	-0.15666	1	-0.15393
ExtraTreesRegressor	1	10	0.71928	0.58376	0.71928	0.58463
ExtraTreesRegressor	1	11	0.70942	0.68865	0.70942	0.6907
ExtraTreesRegressor	1	12	0.73674	0.74092	0.73674	0.74252
ExtraTreesRegressor	1	13	0.77283	0.78652	0.77283	0.78848
ExtraTreesRegressor	1	14	0.74342	0.76883	0.74342	0.77219
ExtraTreesRegressor	1	15	0.74557	0.78178	0.74557	0.78487
ExtraTreesRegressor	1	16	0.73848	0.72709	0.73848	0.72794

Table 19: Different scores for various tree-based methods with $\log_{10}(P(T;2,1)$, closeness, betweenness, and class as predictors and $\left[T \atop k \right]$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as response for a sample of size 500 of non-isomorphic trees of order $n = 18$ generated using uniform sampling

Method	Pruning	k	Train Score	Test Score	Train EVS	Test EVS
DecisionTreeRegressor	0	9	0.0	-0.00087	1.0	-1.06174
DecisionTreeRegressor	0	10	0.0	-0.00019	1.0	-0.21226
DecisionTreeRegressor	0	11	0.3893	0.29412	1.0	0.15898
DecisionTreeRegressor	0	12	0.0	-0.00486	1.0	0.38989
DecisionTreeRegressor	0	13	0.0	-0.00675	1.0	0.6024
DecisionTreeRegressor	0	14	0.0	-0.00751	1.0	0.70046
DecisionTreeRegressor	0	15	0.0	-0.00784	1.0	0.73328
DecisionTreeRegressor	0	16	0.0	-0.00819	1.0	0.77361
DecisionTreeRegressor	1	9	1.0	-0.90217	1.0	-0.90196
DecisionTreeRegressor	1	10	0.3353	0.2591	0.3353	0.26415
DecisionTreeRegressor	1	11	0.59277	0.53568	0.59277	0.53679
DecisionTreeRegressor	1	12	0.74032	0.67822	0.74032	0.67888
DecisionTreeRegressor	1	13	0.80733	0.76866	0.80733	0.76868
DecisionTreeRegressor	1	14	0.83276	0.74644	0.83276	0.74683
DecisionTreeRegressor	1	15	0.84077	0.79667	0.84077	0.7973
DecisionTreeRegressor	1	16	0.83353	0.78909	0.83353	0.78961
ExtraTreeRegressor	0	9	0.13994	-0.06657	1.0	-0.90196
ExtraTreeRegressor	0	10	0.25011	0.28822	1.0	-0.32458
ExtraTreeRegressor	0	11	0.3893	0.29412	1.0	0.26655
ExtraTreeRegressor	0	12	0.48973	0.41411	1.0	0.56446
ExtraTreeRegressor	0	13	0.55093	0.48672	1.0	0.62739
ExtraTreeRegressor	0	14	0.58265	0.47388	1.0	0.76592
ExtraTreeRegressor	0	15	0.59167	0.47672	1.0	0.76236
ExtraTreeRegressor	0	16	0.61117	0.50781	1.0	0.75028
ExtraTreeRegressor	1	9	1.0	-0.90217	1.0	-0.89152
ExtraTreeRegressor	1	10	0.24562	0.34714	0.24562	0.35677
ExtraTreeRegressor	1	11	0.46337	0.4037	0.46337	0.40379
ExtraTreeRegressor	1	12	0.7168	0.64912	0.7168	0.64949
ExtraTreeRegressor	1	13	0.74787	0.6936	0.74787	0.6937
ExtraTreeRegressor	1	14	0.76329	0.72076	0.76329	0.7219
ExtraTreeRegressor	1	15	0.70857	0.64602	0.70857	0.64602
ExtraTreeRegressor	1	16	0.74398	0.71352	0.74398	0.71394
Bagging	NA	9	0.81709	-0.18804	0.81733	-0.17914
Bagging	NA	10	0.86867	0.15476	0.86902	0.15835
Bagging	NA	11	0.91459	0.45655	0.91459	0.45701
Bagging	NA	12	0.94783	0.6492	0.94791	0.6494
Bagging	NA	13	0.95903	0.75299	0.95903	0.753
Bagging	NA	14	0.97159	0.81731	0.9716	0.81748
Bagging	NA	15	0.96829	0.83075	0.96832	0.83087
Bagging	NA	16	0.9724	0.8491	0.97248	0.84916

Table 20: Different scores for various regression methods with $\log_{10}(P(T;2,1)$, closeness, betweenness, and class as predictors and $\begin{bmatrix} T \\ k \end{bmatrix}$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as response for a sample of size 500 of non-isomorphic trees of order $n = 18$ generated using uniform sampling

Method	Pruning	k	Train Score	Test Score	Train EVS	Test EVS
RandomForestRegressor	0	9	0.84444	-0.11222	0.84456	-0.11184
RandomForestRegressor	0	10	0.90138	0.18277	0.90151	0.18513
RandomForestRegressor	0	11	0.93689	0.50666	0.93699	0.50669
RandomForestRegressor	0	12	0.95739	0.66172	0.9574	0.66181
RandomForestRegressor	0	13	0.96768	0.75372	0.96768	0.75373
RandomForestRegressor	0	14	0.9718	0.81334	0.97181	0.81337
RandomForestRegressor	0	15	0.97661	0.83958	0.97662	0.83966
RandomForestRegressor	0	16	0.97817	0.85302	0.97819	0.85304
RandomForestRegressor	1	9	0.83204	-0.10604	0.83249	-0.10589
RandomForestRegressor	1	10	0.41345	0.28867	0.41354	0.29043
RandomForestRegressor	1	11	0.63208	0.57193	0.63215	0.57254
RandomForestRegressor	1	12	0.73836	0.68498	0.73838	0.68537
RandomForestRegressor	1	13	0.81871	0.77385	0.81872	0.77393
RandomForestRegressor	1	14	0.84174	0.76737	0.84175	0.76758
RandomForestRegressor	1	15	0.85565	0.79267	0.85565	0.79273
RandomForestRegressor	1	16	0.85791	0.80573	0.85791	0.80574
ExtraTreesRegressor	0	9	1.0	-0.18572	1.0	-0.18155
ExtraTreesRegressor	0	10	1.0	0.17717	1.0	0.18458
ExtraTreesRegressor	0	11	1.0	0.49111	1.0	0.49174
ExtraTreesRegressor	0	12	1.0	0.65931	1.0	0.65996
ExtraTreesRegressor	0	13	1.0	0.75786	1.0	0.75802
ExtraTreesRegressor	0	14	1.0	0.81299	1.0	0.81305
ExtraTreesRegressor	0	15	1.0	0.84033	1.0	0.84035
ExtraTreesRegressor	0	16	0.99997	0.84559	0.99997	0.8456
ExtraTreesRegressor	1	9	1.0	-0.18452	1.0	-0.17711
ExtraTreesRegressor	1	10	0.23152	0.26341	0.23152	0.26768
ExtraTreesRegressor	1	11	0.56321	0.55777	0.56321	0.55953
ExtraTreesRegressor	1	12	0.70842	0.6747	0.70842	0.6748
ExtraTreesRegressor	1	13	0.78746	0.75229	0.78746	0.75231
ExtraTreesRegressor	1	14	0.82496	0.78926	0.82496	0.78927
ExtraTreesRegressor	1	15	0.81946	0.7729	0.81946	0.77348
ExtraTreesRegressor	1	16	0.83073	0.77582	0.83073	0.77615

7 Support Vector Regression

Table 21: Different scores for support vector regression methods with $\log_{10}(P(T;2,1)$, closeness, betweenness, and class as predictors and $\left[{T \atop k}\right]$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as response for all non-isomorphic trees of order $n = 12$

Method	k	Train Score	Test Score	Train EVS	Test EVS
Linear SVR	6	-0.00089	-0.00414	-0.00055	-0.00052
Linear SVR	7	0.4843	0.47968	0.48761	0.48295
Linear SVR	8	0.74575	0.73116	0.74716	0.73184
Linear SVR	9	0.85438	0.87814	0.85569	0.8807
Linear SVR	10	0.81524	0.74682	0.81983	0.75746
Quadratic SVR	6	0.0004	-0.00266	0.00071	0.00084
Quadratic SVR	7	0.42478	0.41774	0.4331	0.42585
Quadratic SVR	8	0.72592	0.72358	0.72718	0.72361
Quadratic SVR	9	0.87837	0.89873	0.87975	0.89938
Quadratic SVR	10	0.8948	0.85257	0.89549	0.85548

Table 22: Different scores for support vector regression methods with $\log_{10}(P(T;2,1)$, closeness, betweenness, and class as predictors and $\left[{T \atop k}\right]$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as response for a sample of size 500 of non-isomorphic trees of order $n = 18$

Method	k	Train Score	Test Score	Train EVS	Test EVS
Linear SVR	9	-0.64046	-0.04416	0.00138	0.0012
Linear SVR	10	-0.15898	-0.18713	0.01581	0.01195
Linear SVR	11	0.38722	0.33633	0.40674	0.38075
Linear SVR	12	0.60521	0.61186	0.60965	0.62518
Linear SVR	13	0.7331	0.75056	0.73397	0.75344
Linear SVR	14	0.19705	0.24421	0.23256	0.24772
Linear SVR	15	0.7576	0.79896	0.76372	0.80355
Linear SVR	16	0.68692	0.74476	0.69938	0.75607
Quadratic SVR	9	-0.61231	-0.03826	0.00366	0.00397
Quadratic SVR	10	-0.1599	-0.18839	0.01364	0.00986
Quadratic SVR	11	0.32036	0.23869	0.35007	0.29929
Quadratic SVR	12	0.55402	0.50941	0.5597	0.52845
Quadratic SVR	13	0.70675	0.69494	0.70744	0.70022
Quadratic SVR	14	0.79368	0.81197	0.79541	0.81201
Quadratic SVR	15	0.82382	0.85878	0.82853	0.86136
Quadratic SVR	16	0.80338	0.83925	0.8099	0.84503

Table 23: Different scores for support vector regression methods with $\log_{10}(P(T;2,1)$, closeness, betweenness, and class as predictors and $\left[{T \atop k}\right]$ for $n-1 \geq k \geq \lceil \frac{n}{2} \rceil + 1$ as response for a sample of size 500 of non-isomorphic trees of order $n = 18$ generated using uniform sampling

Method	k	Train Score	Test Score	Train EVS	Test EVS
Linear SVR	9	-0.0005	-0.00432	0.00113	0.00085
Linear SVR	10	0.27336	0.30712	0.2917	0.31186
Linear SVR	11	0.49668	0.51849	0.49708	0.51883
Linear SVR	12	0.65326	0.65902	0.65349	0.66038
Linear SVR	13	0.76125	0.76058	0.76266	0.76326
Linear SVR	14	0.82866	0.823	0.8311	0.82689
Linear SVR	15	0.85654	0.84786	0.8586	0.85075
Linear SVR	16	0.85017	0.84142	0.853	0.84479
Quadratic SVR	9	0.00322	-0.00059	0.00422	0.00341
Quadratic SVR	10	0.2495	0.25836	0.27606	0.2689
Quadratic SVR	11	0.43728	0.4548	0.43923	0.45538
Quadratic SVR	12	0.57836	0.59192	0.58025	0.59436
Quadratic SVR	13	0.69864	0.71269	0.69887	0.71298
Quadratic SVR	14	0.78881	0.80226	0.78967	0.80317
Quadratic SVR	15	0.85359	0.85629	0.85414	0.85693
Quadratic SVR	16	0.87402	0.86878	0.8768	0.87153