Introduction:

Objectives:

Methods:

Workflow:

Features:

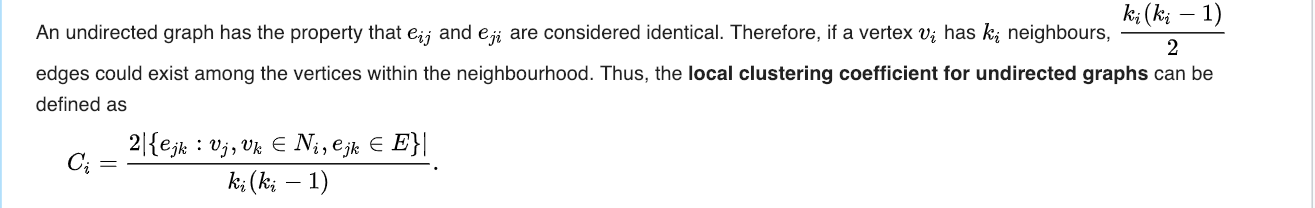
**Topological Features:**

1. Average Shortest Path:

Step 1: Find the geodesic distance from a non-disease node to each disease node

Step 2: Take the average of all distances as the Average Shortest Path of a non-disease node to all disease nodes

1. Local Clustering Coefficient:

Implemented according to the formula

The cluster E is the disease cluster, and N is the non-disease nodes cluster

1. Degree Centrality

Implemented by NetworkX library

Rationale: Higher, more likely to be involved in a more important functional module

1. Closeness Centrality

Implemented by NetworkX library

Rationale: Higher, more functionally important as needs to be communicated quickly.

1. Betweeness Centrality

Implemented by NetworkX library

Rationale: An important node will lie on a higher proportion of the paths.

1. Eigenvector Centrality

Implemented by NetworkX library

Rationale: the influence of a node in a network

1. Percolation centrality

Implemented by NetworkX library. This method has problems; thus the library method couldn’t successfully be run.

Rationale: importance of a node in purely topological terms, despite the network dynamic

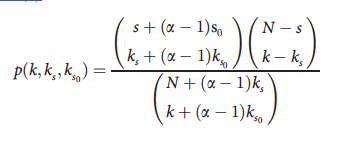
1. Pagerank

Implemented by NetworkX library.

Rationale: the notion of how central a node is in a network relative to a particular node

1. Connectivity Significance

Implemented using the formula given in the DIAMOnD paper



1. Articulation Point

Implemented using the algorithm given in this website (<https://www.geeksforgeeks.org/articulation-points-or-cut-vertices-in-a-graph/> ), the code has been slightly modified.

1. Modularity

Step 1: NetworkX library computes the modularity matrix

Step 2: Summing over all modularity from a node (row, in Gc.nodes()) to a disease node (column), exclude the node itself.

Step 3: Take the average for each node ( in Gc.nodes()).

**Sequence based features:**

**Functional Features:**

**Disease Module Identification**

**Dataset (After Data Cleaning, Data Balancing and Stratified Split)**

|  |  |  |
| --- | --- | --- |
| Disease Node | Non Disease Node (Essential and Non-essential) | Total |
| 1441 | 4329 | 5770 |

**Dataset 1**

|  |  |  |
| --- | --- | --- |
| Train set | | |
| Disease Node | Non Disease Node (Essential and Non-essential) | Total |
| 1009 | 1008 | 2017 |
| Test set | | |
| Disease Node | Non Disease Node (Essential and Non-essential) | Total |
| 432 | 433 | 865 |

**Dataset 2**

|  |  |  |
| --- | --- | --- |
| Train set | | |
| Disease Node | Non Disease Node (Essential and Non-essential) | Total |
| 1009 | 1008 | 2017 |
| Test set | | |
| Disease Node | Non Disease Node (Essential and Non-essential) | Total |
| 432 | 433 | 865 |

**Dataset 3**

|  |  |  |
| --- | --- | --- |
| Train set | | |
| Disease Node | Non Disease Node (Essential and Non-essential) | Total |
| 1008 | 994 | 2002 |
| Test set | | |
| Disease Node | Non Disease Node (Essential and Non-essential) | Total |
| 433 | 426 | 859 |

**Experiment1: K means clustering with Principle Component Analysis (on the whole dataset)**

* Dataset 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Condition | Accuracy | Precision | Recall | F1 |
| Cluster 0 is considered as disease | 0.45107564 | 0.41198502 | 0.22900763 | 0.29438002 |
| Cluster 0 is considered as non-disease | 0.54892436 | 0.53387794 | 0.77099237 | 0.63089154 |
| Cluster 1 is considered as disease | 0.52602359 | 0.52856055 | 0.48160999 | 0.50399419 |
| Cluster 1 is considered as non-disease | 0.47397641 | 0.47609943 | 0.51839001 | 0.49634551 |
| Cluster 2 is considered as disease | 0.52290076 | 0.54296875 | 0.28938237 | 0.3775464 |
| Cluster 2 is considered as non-disease | 0.47709924 | 0.48438978 | 0.71061763 | 0.57609001 |

Conclusion: Disease set: cluster 1; Non-disease set: cluster 0 and cluster 2

* Dataset 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Condition | Accuracy | Precision | Recall | F1 |
| Cluster 0 is considered as disease | 0.45697432 | 0.42091837 | 0.22900763 | 0.29662921 |
| Cluster 0 is considered as non-disease | 0.54302568 | 0.52955195 | 0.77099237 | 0.62786098 |
| Cluster 1 is considered as disease | 0.51804303 | 0.53412073 | 0.28244275 | 0.36949614 |
| Cluster 1 is considered as non-disease | 0.48195697 | 0.48773585 | 0.71755725 | 0.58073575 |
| Cluster 2 is considered as disease | 0.52498265 | 0.52694611 | 0.48854962 | 0.50702197 |
| Cluster 2 is considered as non-disease | 0.47501735 | 0.4767141 | 0.51145038 | 0.49347171 |

Conclusion: Disease set: cluster 2; Non-disease set: cluster 0 and cluster 1

* Dataset 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Condition | Accuracy | Precision | Recall | F1 |
| Cluster 0 is considered as disease | 0.52079692 | 0.54475703 | 0.29562804 | 0.38326586 |
| Cluster 0 is considered as non-disease | 0.47920308 | 0.48821549 | 0.70437196 | 0.57670455 |
| Cluster 1 is considered as disease | 0.44669696 | 0.40482574 | 0.20957668 | 0.27617741 |
| Cluster 1 is considered as non-disease | 0.55330304 | 0.53853428 | 0.79042332 | 0.64060742 |
| Cluster 2 is considered as disease | 0.52883607 | 0.53488372 | 0.49479528 | 0.51405912 |
| Cluster 2 is considered as non-disease | 0.47116393 | 0.47643979 | 0.50520472 | 0.49040081 |

Conclusion: Disease set: cluster 2; Non-disease set: cluster 0 and cluster 1

General Conclusion for Kmeans:

Cluster 2 is the disease set, and cluster 0 and cluster 1 are non-disease cluster

**Experiment2: Support Vector Machine with RBF Kernel**

* Dataset 1

Cross Validation Score

|  |  |
| --- | --- |
| CV1 | 0.62376238 |
| CV2 | 0.60891089 |
| CV3 | 0.61881188 |
| CV4 | 0.54455446 |
| CV5 | 0.58910891 |
| CV6 | 0.59405941 |
| CV7 | 0.58415842 |
| CV8 | 0.58415842 |
| CV9 | 0.51741294 |
| CV10 | 0.48 |

Test Set Score

|  |  |  |  |
| --- | --- | --- | --- |
| Accuracy | Precision | Recall | F1 Score |
| 0.55491329 | 0.55119826 | 0.58564815 | 0.56790123 |

* Dataset 2

Cross Validation Score

|  |  |
| --- | --- |
| CV1 | 0.55445545 |
| CV2 | 0.53960396 |
| CV3 | 0.57920792 |
| CV4 | 0.56435644 |
| CV5 | 0.56930693 |
| CV6 | 0.55940594 |
| CV7 | 0.64356436 |
| CV8 | 0.55940594 |
| CV9 | 0.60696517 |
| CV10 | 0.55 |

Test Set Score

|  |  |  |  |
| --- | --- | --- | --- |
| Accuracy | Precision | Recall | F1 Score |
| 0.59306358 | 0.5952381 | 0.5787037 | 0.58685446 |

* Dataset 3

Cross Validation Score

|  |  |
| --- | --- |
| CV1 | 0.51741294 |
| CV2 | 0.53233831 |
| CV3 | 0.59701493 |
| CV4 | 0.62686567 |
| CV5 | 0.66 |
| CV6 | 0.66 |
| CV7 | 0.61 |
| CV8 | 0.55 |
| CV9 | 0.6281407 |
| CV10 | 0.5678392 |

Test Set Score

|  |  |  |  |
| --- | --- | --- | --- |
| Accuracy | Precision | Recall | F1 Score |
| 0.56810244 | 0.562 | 0.64896074 | 0.60235798 |

**Experiment3: Random Forest**

* Dataset 1

Cross Validation Score

|  |  |
| --- | --- |
| CV1 | 0.71287129 |
| CV2 | 0.72277228 |
| CV3 | 0.70792079 |
| CV4 | 0.73267327 |
| CV5 | 0.68316832 |
| CV6 | 0.66831683 |
| CV7 | 0.72277228 |
| CV8 | 0.71287129 |
| CV9 | 0.68159204 |
| CV10 | 0.64 |

Test Set Score

|  |  |  |  |
| --- | --- | --- | --- |
| Accuracy | Precision | Recall | F1 Score |
| 0.68786127 | 0.68786127 | 0.68287037 | 0.68604651 |

* Dataset 2

Cross Validation Score

|  |  |
| --- | --- |
| CV1 | 0.62871287 |
| CV2 | 0.62376238 |
| CV3 | 0.67821782 |
| CV4 | 0.65346535 |
| CV5 | 0.65841584 |
| CV6 | 0.62871287 |
| CV7 | 0.65841584 |
| CV8 | 0.65841584 |
| CV9 | 0.70646766 |
| CV10 | 0.66 |

Test Set Score

|  |  |  |  |
| --- | --- | --- | --- |
| Accuracy | Precision | Recall | F1 Score |
| 0.69710983 | 0.69767442 | 0.69444444 | 0.69605568 |

* Dataset 3

Cross Validation Score

|  |  |
| --- | --- |
| CV1 | 0.6119403 |
| CV2 | 0.71641791 |
| CV3 | 0.69154229 |
| CV4 | 0.68159204 |
| CV5 | 0.715 |
| CV6 | 0.75 |
| CV7 | 0.7 |
| CV8 | 0.625 |
| CV9 | 0.69346734 |
| CV10 | 0.66834171 |

Test Set Score

|  |  |  |  |
| --- | --- | --- | --- |
| Accuracy | Precision | Recall | F1 Score |
| 0.67520373 | 0.6673913 | 0.70900693 | 0.68756999 |

**Experiment4: Deep Learning**

* Dataset 1

Cross Validation Score

|  |  |
| --- | --- |
| CV1 | 0.58910891 |
| CV2 | 0.58910891 |
| CV3 | 0.53465347 |
| CV4 | 0.52475248 |
| CV5 | 0.57920792 |
| CV6 | 0.56930693 |
| CV7 | 0.51485149 |
| CV8 | 0.59405941 |
| CV9 | 0.55223881 |
| CV10 | 0.52 |

Test Set Score

|  |  |  |  |
| --- | --- | --- | --- |
| Accuracy | Precision | Recall | F1 Score |
| 0.54219653 | 0.53501946 | 0.63657407 | 0.58139535 |

* Dataset 2

Cross Validation Score

|  |  |
| --- | --- |
| CV1 | 0.56435644 |
| CV2 | 0.49009901 |
| CV3 | 0.54950495 |
| CV4 | 0.56435644 |
| CV5 | 0.61881188 |
| CV6 | 0.51485149 |
| CV7 | 0.53960396 |
| CV8 | 0.4950495 |
| CV9 | 0.60199005 |
| CV10 | 0.59 |

Test Set Score

|  |  |  |  |
| --- | --- | --- | --- |
| Accuracy | Precision | Recall | F1 Score |
| 0.56878613 | 0.57040573 | 0.55324074 | 0.56169213 |

* **Dataset 3**

Cross Validation Score

|  |  |
| --- | --- |
| CV1 | 0.55223881 |
| CV2 | 0.48258706 |
| CV3 | 0.63681592 |
| CV4 | 0.56716418 |
| CV5 | 0.67 |
| CV6 | 0.605 |
| CV7 | 0.56 |
| CV8 | 0.555 |
| CV9 | 0.5678392 |
| CV10 | 0.5678392 |

Test Set Score

|  |  |  |  |
| --- | --- | --- | --- |
| Accuracy | Precision | Recall | F1 Score |
| 0.58207218 | 0.58447489 | 0.59122402 | 0.58783008 |

**Feature Selection**

**Feature Importance using the model that gives the best result**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Dataset 1 | Dataset 2 | Dataset 3 | Average | Met  Average | Average Score Ranking |
| Average Shortest Path to all Disease genes | 0.04346083 | 0.03424906 | 0.031861 | 0.0365236 | 1 | 2 |
| BetweennessCentrality | 0.01667202 | 0.01703063 | 0.01930463 | 0.0176691 | 0 | 35 |
| ClosenessCentrality | 0.02228344 | 0.02340866 | 0.02744768 | 0.0243799 | 0 | 18 |
| DegreeCentrality | 0.01317933 | 0.01338562 | 0.01184739 | 0.0128041 | 0 | 36 |
| EigenvectorCentrality | 0.02404728 | 0.02516761 | 0.02079586 | 0.0233369 | 0 | 30 |
| HarmonicCentrality | 0.02559715 | 0.02451262 | 0.02254047 | 0.0242167 | 0 | 20 |
| Modularity | 0.12664342 | 0.11366804 | 0.13332279 | 0.1245448 | 1 | 1 |
| PageRank | 0.0245738 | 0.02959554 | 0.02306586 | 0.0257451 | 0 | 12 |
| BP | 0.02930925 | 0.04047054 | 0.03008127 | 0.033287 | 1 | 4 |
| CC | 0.02815931 | 0.02608422 | 0.02585718 | 0.0267002 | 0 | 7 |
| MF | 0.02505777 | 0.02997702 | 0.02251853 | 0.0258511 | 0 | 11 |
| FrequencyA | 0.02282336 | 0.02592521 | 0.02355146 | 0.0241 | 0 | 23 |
| FrequencyC | 0.02604764 | 0.0257333 | 0.02863298 | 0.0268046 | 0 | 6 |
| FrequencyD | 0.0230632 | 0.02453177 | 0.02503815 | 0.024211 | 0 | 21 |
| FrequencyE | 0.03329916 | 0.03024063 | 0.03877915 | 0.0341063 | 1 | 3 |
| FrequencyF | 0.02312303 | 0.02213389 | 0.0225686 | 0.0226085 | 0 | 33 |
| FrequencyG | 0.02781013 | 0.02360264 | 0.02279414 | 0.0247356 | 0 | 16 |
| FrequencyH | 0.02493357 | 0.02560892 | 0.02747443 | 0.0260056 | 0 | 10 |
| FrequencyI | 0.02552015 | 0.02288315 | 0.02295234 | 0.0237852 | 0 | 26 |
| FrequencyK | 0.02260602 | 0.0255323 | 0.02641962 | 0.0248526 | 0 | 15 |
| FrequencyL | 0.02359026 | 0.02509347 | 0.02257635 | 0.0237534 | 0 | 27 |
| FrequencyM | 0.02351522 | 0.02450154 | 0.02178546 | 0.0232674 | 0 | 31 |
| FrequencyN | 0.02481608 | 0.02290447 | 0.025167 | 0.0242959 | 0 | 19 |
| FrequencyP | 0.02556854 | 0.02590297 | 0.02413133 | 0.0252009 | 0 | 14 |
| FrequencyQ | 0.02476733 | 0.02491347 | 0.02291031 | 0.024197 | 0 | 22 |
| FrequencyR | 0.02260081 | 0.02390277 | 0.02531955 | 0.023941 | 0 | 25 |
| FrequencyS | 0.02403117 | 0.02507819 | 0.02660608 | 0.0252385 | 0 | 13 |
| FrequencyT | 0.02762817 | 0.02418846 | 0.02779702 | 0.0265379 | 0 | 8 |
| FrequencyV | 0.02291558 | 0.02496835 | 0.02426742 | 0.0240505 | 0 | 24 |
| FrequencyW | 0.02485183 | 0.02533806 | 0.02842926 | 0.0262064 | 0 | 9 |
| FrequencyX | 0.02477384 | 0.02698961 | 0.02235954 | 0.0247077 | 0 | 17 |
| Aromaticity | 0.0240424 | 0.02275794 | 0.02095999 | 0.0225868 | 0 | 34 |
| Isoelectric | 0.02889065 | 0.03005274 | 0.02974957 | 0.0295643 | 1 | 5 |
| SSfractionHelix | 0.02501649 | 0.02274011 | 0.02255569 | 0.0234374 | 0 | 29 |
| SSfractionTurn | 0.02226102 | 0.02332593 | 0.02377288 | 0.0231199 | 0 | 32 |
| SSfractionSheet | 0.02252077 | 0.02360054 | 0.02475901 | 0.0236268 | 0 | 28 |

1/36 = 0.02777

**Independent T Test**

Dataset 1

Ttest\_indResult(statistic=-0.18277642374064065, pvalue=0.8549863129780713)

Ttest\_indResult(statistic=-0.3149051960389861, pvalue=0.752856449004072)

Ttest\_indResult(statistic=0.07862849334802469, pvalue=0.9373335856726093)

Ttest\_indResult(statistic=-0.2965791204554409, pvalue=0.7668092390408194)

Ttest\_indResult(statistic=-0.6484670342445745, pvalue=0.5167346246141626)

Ttest\_indResult(statistic=-0.47899885148173726, pvalue=0.6319758107692448)

Ttest\_indResult(statistic=-6.932983289790123, pvalue=5.065949059727641e-12)

Ttest\_indResult(statistic=0.05781528000985333, pvalue=0.9538997776847222)

Ttest\_indResult(statistic=4.289755487242454, pvalue=1.8475936362579164e-05)

Ttest\_indResult(statistic=2.283746436521454, pvalue=0.022458928996439035)

Ttest\_indResult(statistic=2.823171484732739, pvalue=0.004787773688851957)

Ttest\_indResult(statistic=-0.03193402099565115, pvalue=0.974526880443391)

Ttest\_indResult(statistic=-2.231430091842398, pvalue=0.025729329540794647)

Ttest\_indResult(statistic=2.0413731266092325, pvalue=0.041304714610277966)

Ttest\_indResult(statistic=5.403450188970156, pvalue=7.070806848722212e-08)

Ttest\_indResult(statistic=-0.9822605142850782, pvalue=0.32605402908888004)

Ttest\_indResult(statistic=-3.506759425006363, pvalue=0.00046053207857954536)

Ttest\_indResult(statistic=-2.889155066792971, pvalue=0.0038916376035779444)

Ttest\_indResult(statistic=1.1954937721292205, pvalue=0.23199259421721594)

Ttest\_indResult(statistic=3.9916668159817164, pvalue=6.724845698795733e-05)

Ttest\_indResult(statistic=-0.00800287277873035, pvalue=0.9936152538058017)

Ttest\_indResult(statistic=0.49689026652013957, pvalue=0.619304362282586)

Ttest\_indResult(statistic=-1.5427300591295616, pvalue=0.12300614961335822)

Ttest\_indResult(statistic=-1.9839111244168541, pvalue=0.04736054057398115)

Ttest\_indResult(statistic=1.722942494508443, pvalue=0.08500626880017957)

Ttest\_indResult(statistic=4.1987413266013505, pvalue=2.7653847852305085e-05)

Ttest\_indResult(statistic=-2.7194710137784615, pvalue=0.006577879341981658)

Ttest\_indResult(statistic=-2.6985622805118488, pvalue=0.007004616469558606)

Ttest\_indResult(statistic=0.4939698805936146, pvalue=0.6213651215414472)

Ttest\_indResult(statistic=-2.0321280884788084, pvalue=0.04223228760579554)

Ttest\_indResult(statistic=-3.141112547618865, pvalue=0.001700150033144769)

Ttest\_indResult(statistic=-2.7749790755896266, pvalue=0.005556155181927847)

Ttest\_indResult(statistic=0.4383524254116494, pvalue=0.6611637051115553)

Ttest\_indResult(statistic=-0.7291674097316232, pvalue=0.4659585567018689)

Ttest\_indResult(statistic=-4.618599131256598, pvalue=4.032870106702551e-06)

Ttest\_indResult(statistic=-4.618599131256598, pvalue=4.032870106702551e-06)

Dataset 2:

Ttest\_indResult(statistic=1.6850277511760048, pvalue=0.09209155786926618)

Ttest\_indResult(statistic=-1.2793656067963917, pvalue=0.2008714002708827)

Ttest\_indResult(statistic=-0.7454353084385115, pvalue=0.45606943624813967)

Ttest\_indResult(statistic=-0.022420376494383207, pvalue=0.9821141795206039)

Ttest\_indResult(statistic=-0.7025355511781524, pvalue=0.4824020161524859)

Ttest\_indResult(statistic=0.1891117512535865, pvalue=0.8500185569361836)

Ttest\_indResult(statistic=-7.539981895988298, pvalue=6.260511821699951e-14)

Ttest\_indResult(statistic=3.4264649793757305, pvalue=0.0006200621790114283)

Ttest\_indResult(statistic=6.160676401028748, pvalue=8.252639105877047e-10)

Ttest\_indResult(statistic=2.1828444244521896, pvalue=0.029127766166405396)

Ttest\_indResult(statistic=4.456397792920994, pvalue=8.652923452794624e-06)

Ttest\_indResult(statistic=-0.4471906021957826, pvalue=0.6547710746979074)

Ttest\_indResult(statistic=-2.8020926839341045, pvalue=0.005111168448899066)

Ttest\_indResult(statistic=1.675094051309323, pvalue=0.09402434169473275)

Ttest\_indResult(statistic=5.0651897856955, pvalue=4.3372982743642405e-07)

Ttest\_indResult(statistic=-0.659166035923022, pvalue=0.5098418536951119)

Ttest\_indResult(statistic=-3.4847226454278393, pvalue=0.0005000013320963884)

Ttest\_indResult(statistic=-2.6547657942191236, pvalue=0.007980030800323579)

Ttest\_indResult(statistic=0.21093563094406423, pvalue=0.8329524233334842)

Ttest\_indResult(statistic=3.4834289543651846, pvalue=0.0005024137685241842)

Ttest\_indResult(statistic=-0.6792815662650938, pvalue=0.4970140609866417)

Ttest\_indResult(statistic=0.600471326652482, pvalue=0.5482394120371392)

Ttest\_indResult(statistic=-0.56076777552052, pvalue=0.5749994794654787)

Ttest\_indResult(statistic=-2.18658953740151, pvalue=0.02885270751086501)

Ttest\_indResult(statistic=1.4178326792294318, pvalue=0.15634780393195236)

Ttest\_indResult(statistic=3.482192720570059, pvalue=0.0005047292014293054)

Ttest\_indResult(statistic=-1.129860441230848, pvalue=0.2586291263906128)

Ttest\_indResult(statistic=-1.8022972333741063, pvalue=0.07160317059632147)

Ttest\_indResult(statistic=0.341863682850796, pvalue=0.7324784243733373)

Ttest\_indResult(statistic=-1.9375976591579935, pvalue=0.052769974269556465)

Ttest\_indResult(statistic=-0.659267059218383, pvalue=0.5097770002066442)

Ttest\_indResult(statistic=-1.3282554921051428, pvalue=0.1841989635966047)

Ttest\_indResult(statistic=-0.16863538816799437, pvalue=0.866095285306817)

Ttest\_indResult(statistic=-0.7762613111804965, pvalue=0.43765845519766233)

Ttest\_indResult(statistic=-3.648051488674246, pvalue=0.00026892273860660453)

Dataset 3:

Ttest\_indResult(statistic=1.6384569774309812, pvalue=0.10143643612963418)

Ttest\_indResult(statistic=-2.306597038574143, pvalue=0.021148464909536042)

Ttest\_indResult(statistic=1.6706475968451309, pvalue=0.09490075872095237)

Ttest\_indResult(statistic=0.03564759834234488, pvalue=0.9715658430176732)

Ttest\_indResult(statistic=-0.8603675178189164, pvalue=0.38965862656593864)

Ttest\_indResult(statistic=-0.38302550925781403, pvalue=0.7017293253829867)

Ttest\_indResult(statistic=-7.381120274216576, pvalue=2.047251952535262e-13)

Ttest\_indResult(statistic=0.6778119670906309, pvalue=0.49794577106923)

Ttest\_indResult(statistic=5.4027142588480315, pvalue=7.103659696471656e-08)

Ttest\_indResult(statistic=1.1439239573935958, pvalue=0.25275094096814893)

Ttest\_indResult(statistic=2.8790019977582832, pvalue=0.004018993255706891)

Ttest\_indResult(statistic=-0.31830128813971603, pvalue=0.7502796558060774)

Ttest\_indResult(statistic=-2.6461824293846354, pvalue=0.008185205658209465)

Ttest\_indResult(statistic=1.9739024816592206, pvalue=0.04848896119302228)

Ttest\_indResult(statistic=6.344545617065992, pvalue=2.5848339253721784e-10)

Ttest\_indResult(statistic=-1.4319862872041669, pvalue=0.1522570593919467)

Ttest\_indResult(statistic=-4.049668578550852, pvalue=5.2653383106572256e-05)

Ttest\_indResult(statistic=-3.025902692678568, pvalue=0.0025010040574126266)

Ttest\_indResult(statistic=0.32295304519565327, pvalue=0.7467544195373239)

Ttest\_indResult(statistic=6.209827349136399, pvalue=6.073979866754787e-10)

Ttest\_indResult(statistic=0.22073880480078165, pvalue=0.8253115765939747)

Ttest\_indResult(statistic=0.287572320549006, pvalue=0.7736949773773915)

Ttest\_indResult(statistic=-0.8692370241562105, pvalue=0.3847904659991874)

Ttest\_indResult(statistic=-2.1127393893327238, pvalue=0.03470954507176236)

Ttest\_indResult(statistic=1.6613616922591758, pvalue=0.09675046617132385)

Ttest\_indResult(statistic=4.235925700043015, pvalue=2.348028547944756e-05)

Ttest\_indResult(statistic=-4.131437627880463, pvalue=3.7084658363001634e-05)

Ttest\_indResult(statistic=-4.110283836804894, pvalue=4.0629166588291366e-05)

Ttest\_indResult(statistic=-0.10289055203062494, pvalue=0.9180570522968077)

Ttest\_indResult(statistic=-2.969464314452029, pvalue=0.0030080004910241417)

Ttest\_indResult(statistic=-1.0713268116852452, pvalue=0.28411293600905646)

Ttest\_indResult(statistic=-2.3082788821572233, pvalue=0.02105466367098374)

Ttest\_indResult(statistic=0.9655223889228307, pvalue=0.33436488931293007)

Ttest\_indResult(statistic=-0.8934561166508481, pvalue=0.3716881623573115)

Ttest\_indResult(statistic=-5.448175235366589, pvalue=5.521638990703887e-08)

**Mann Whitney U Test**

Dataset 1:

MannwhitneyuResult(statistic=1032184.5, pvalue=0.39315089464094777)

MannwhitneyuResult(statistic=1004985.5, pvalue=0.0644137250549062)

MannwhitneyuResult(statistic=1034583.0, pvalue=0.43497131976325554)

MannwhitneyuResult(statistic=1029021.0, pvalue=0.3378996402975891)

MannwhitneyuResult(statistic=1029223.5, pvalue=0.34322049999878007)

MannwhitneyuResult(statistic=1016832.0, pvalue=0.168910620121143)

MannwhitneyuResult(statistic=911328.0, pvalue=6.4889790936366685e-09)

MannwhitneyuResult(statistic=1035242.0, pvalue=0.4466115988951819)

MannwhitneyuResult(statistic=945230.5, pvalue=1.5622267536446604e-05)

MannwhitneyuResult(statistic=973178.5, pvalue=0.001790314286837703)

MannwhitneyuResult(statistic=968914.5, pvalue=0.000955175861641887)

MannwhitneyuResult(statistic=1029890.0, pvalue=0.3542600297132102)

MannwhitneyuResult(statistic=995380.5, pvalue=0.027497987737792183)

MannwhitneyuResult(statistic=988511.0, pvalue=0.01299180551408405)

MannwhitneyuResult(statistic=916647.5, pvalue=2.6061224758438396e-08)

MannwhitneyuResult(statistic=1018128.5, pvalue=0.18394628886017156)

MannwhitneyuResult(statistic=970426.5, pvalue=0.0011981739068224109)

MannwhitneyuResult(statistic=986840.5, pvalue=0.010688786288739354)

MannwhitneyuResult(statistic=1019966.0, pvalue=0.20663298836319244)

MannwhitneyuResult(statistic=965313.5, pvalue=0.0005472239588938336)

MannwhitneyuResult(statistic=1028517.0, pvalue=0.3316664995989782)

MannwhitneyuResult(statistic=1027548.0, pvalue=0.3160759939173151)

MannwhitneyuResult(statistic=1001104.5, pvalue=0.04819396458250507)

MannwhitneyuResult(statistic=999836.5, pvalue=0.042771382891805904)

MannwhitneyuResult(statistic=1001375.0, pvalue=0.04941910269890824)

MannwhitneyuResult(statistic=957149.0, pvalue=0.00014138749920597878)

MannwhitneyuResult(statistic=976384.0, pvalue=0.0028079482749482575)

MannwhitneyuResult(statistic=965166.0, pvalue=0.000534601219092292)

MannwhitneyuResult(statistic=1028924.0, pvalue=0.3383048337445807)

MannwhitneyuResult(statistic=974836.5, pvalue=0.0022622619189020636)

MannwhitneyuResult(statistic=959641.5, pvalue=0.00021659101247698136)

MannwhitneyuResult(statistic=977010.5, pvalue=0.003059317078474078)

MannwhitneyuResult(statistic=1028734.5, pvalue=0.33520763954063165)

MannwhitneyuResult(statistic=1018661.5, pvalue=0.1903616671892166)

MannwhitneyuResult(statistic=946687.5, pvalue=2.0749769268702637e-05)

MannwhitneyuResult(statistic=946687.5, pvalue=2.0749769268702637e-05)

Dataset 2:

MannwhitneyuResult(statistic=1017173.5, pvalue=0.1727929217917945)

MannwhitneyuResult(statistic=1004776.5, pvalue=0.06329099773345542)

MannwhitneyuResult(statistic=1017942.0, pvalue=0.18173223776814984)

MannwhitneyuResult(statistic=1028775.0, pvalue=0.33359920693134926)

MannwhitneyuResult(statistic=997343.0, pvalue=0.03354617123947449)

MannwhitneyuResult(statistic=1020200.5, pvalue=0.2096421155392183)

MannwhitneyuResult(statistic=896437.0, pvalue=1.056611557110278e-10)

MannwhitneyuResult(statistic=987681.5, pvalue=0.011799046746643928)

MannwhitneyuResult(statistic=883552.5, pvalue=2.1701328660313372e-12)

MannwhitneyuResult(statistic=972573.0, pvalue=0.001640886602355619)

MannwhitneyuResult(statistic=947751.0, pvalue=2.5460885700292222e-05)

MannwhitneyuResult(statistic=1032473.5, pvalue=0.3981353479295282)

MannwhitneyuResult(statistic=965729.5, pvalue=0.0005842671987594864)

MannwhitneyuResult(statistic=1000643.5, pvalue=0.04616211107334913)

MannwhitneyuResult(statistic=924358.5, pvalue=1.7101612662581966e-07)

MannwhitneyuResult(statistic=1029171.0, pvalue=0.34235829866534057)

MannwhitneyuResult(statistic=967886.5, pvalue=0.0008167664220440808)

MannwhitneyuResult(statistic=979569.0, pvalue=0.004309507121700831)

MannwhitneyuResult(statistic=1037850.0, pvalue=0.49303442278068305)

MannwhitneyuResult(statistic=978155.5, pvalue=0.003571634869405955)

MannwhitneyuResult(statistic=1035476.0, pvalue=0.4507569173517414)

MannwhitneyuResult(statistic=1032226.5, pvalue=0.39387434149328837)

MannwhitneyuResult(statistic=1024775.0, pvalue=0.2733035195954282)

MannwhitneyuResult(statistic=989264.5, pvalue=0.014163753949628962)

MannwhitneyuResult(statistic=1014090.0, pvalue=0.1397954620849276)

MannwhitneyuResult(statistic=980475.5, pvalue=0.00485174918486991)

MannwhitneyuResult(statistic=1007877.0, pvalue=0.08700901740650102)

MannwhitneyuResult(statistic=986259.0, pvalue=0.009975107994241391)

MannwhitneyuResult(statistic=1028331.0, pvalue=0.32865011617379103)

MannwhitneyuResult(statistic=982397.5, pvalue=0.00620163499140749)

MannwhitneyuResult(statistic=1011522.0, pvalue=0.11580675591229284)

MannwhitneyuResult(statistic=1006080.5, pvalue=0.07495654971186332)

MannwhitneyuResult(statistic=1011101.0, pvalue=0.11217136251068038)

MannwhitneyuResult(statistic=1023494.5, pvalue=0.2545684332206299)

MannwhitneyuResult(statistic=953102.5, pvalue=6.899107222635008e-05)

MannwhitneyuResult(statistic=953102.5, pvalue=6.899107222635008e-05)

Dataset 3:

MannwhitneyuResult(statistic=1006478.5, pvalue=0.22577515455857444)

MannwhitneyuResult(statistic=1000988.5, pvalue=0.1532265780959839)

MannwhitneyuResult(statistic=996500.5, pvalue=0.11419658441039016)

MannwhitneyuResult(statistic=1005214.0, pvalue=0.2059856089397778)

MannwhitneyuResult(statistic=1019533.5, pvalue=0.4357012039680279)

MannwhitneyuResult(statistic=1012481.5, pvalue=0.31522332589777197)

MannwhitneyuResult(statistic=892124.0, pvalue=1.482487215362354e-09)

MannwhitneyuResult(statistic=997623.5, pvalue=0.12431628231189307)

MannwhitneyuResult(statistic=905530.5, pvalue=5.11969601659638e-08)

MannwhitneyuResult(statistic=982561.5, pvalue=0.03321730143524073)

MannwhitneyuResult(statistic=959198.0, pvalue=0.0019074807821780678)

MannwhitneyuResult(statistic=1015392.5, pvalue=0.36342307669957186)

MannwhitneyuResult(statistic=955916.5, pvalue=0.0011764847218549753)

MannwhitneyuResult(statistic=967995.0, pvalue=0.006300395687026905)

MannwhitneyuResult(statistic=881034.0, pvalue=6.325715816126703e-11)

MannwhitneyuResult(statistic=1004927.5, pvalue=0.20524233023283023)

MannwhitneyuResult(statistic=923832.5, pvalue=3.495664819102729e-06)

MannwhitneyuResult(statistic=972343.0, pvalue=0.010779784888543904)

MannwhitneyuResult(statistic=1019619.0, pvalue=0.43722690843177875)

MannwhitneyuResult(statistic=904571.0, pvalue=4.028265542499379e-08)

MannwhitneyuResult(statistic=1000033.0, pvalue=0.14810412746680468)

MannwhitneyuResult(statistic=1022036.0, pvalue=0.4806214590436843)

MannwhitneyuResult(statistic=999329.5, pvalue=0.14086424788443141)

MannwhitneyuResult(statistic=958809.5, pvalue=0.0018033601598450525)

MannwhitneyuResult(statistic=999535.0, pvalue=0.14295379624485438)

MannwhitneyuResult(statistic=939260.5, pvalue=7.365058132364191e-05)

MannwhitneyuResult(statistic=924324.5, pvalue=3.8803640773743735e-06)

MannwhitneyuResult(statistic=932087.0, pvalue=1.8919123835537842e-05)

MannwhitneyuResult(statistic=1020613.0, pvalue=0.4550118874319015)

MannwhitneyuResult(statistic=937359.5, pvalue=5.175270355893612e-05)

MannwhitneyuResult(statistic=985782.0, pvalue=0.04554299614797883)

MannwhitneyuResult(statistic=978551.5, pvalue=0.021847849315715765)

MannwhitneyuResult(statistic=1016468.5, pvalue=0.381853387515923)

MannwhitneyuResult(statistic=1011470.5, pvalue=0.299145346399844)

MannwhitneyuResult(statistic=901663.5, pvalue=1.926384219378441e-08)

MannwhitneyuResult(statistic=901663.5, pvalue=1.926384219378441e-08)

**Previous result**

|  |  |
| --- | --- |
| **Feature** | **p-value** |
| betweennessCentrality | 0.0 |
| degreeCentrality | 0.0 |
| eigenvectorCentrality | 0.0 |
| pageRank | 0.0 |
| frequencyC | 5.0711617337839556e-272 |
| bp | 2.2551136554012031e-266 |
| frequencyP | 3.1809961982607242e-205 |
| modularity | 7.7864781074210468e-197 |
| frequencyG | 3.0935046358956275e-169 |
| frequencyA | 5.6546162404219853e-168 |
| frequencyR | 1.5852180244653675e-167 |
| frequencyQ | 6.6957218016767694e-161 |
| frequencyW | 1.5991249980870714e-158 |
| mf | 7.7018610203437641e-156 |
| isoelectric | 3.7564798742695547e-154 |
| ssfractionTurn | 1.9318871410386582e-143 |
| ssfractionSheet | 1.9318871410386582e-143 |
| frequencyE | 5.4394821456192792e-143 |
| frequencyK | 3.2295339759509463e-135 |
| frequencyS | 6.0544780905567766e-133 |
| frequencyM | 3.2746933783387814e-129 |
| frequencyH | 1.7654241315856116e-119 |
| cc | 1.7797019013339253e-116 |
| frequencyY | 1.5314214346519035e-105 |
| frequencyF | 1.0239487524761442e-102 |
| frequencyT | 3.0786315358753929e-97 |
| frequencyD | 4.9152092390936651e-96 |
| frequencyL | 3.203509005186527e-87 |
| frequencyV | 2.3207878305453458e-86 |
| avgSP | 3.2491338206333851e-82 |
| aromaticity | 1.1898493070107014e-77 |
| frequencyN | 2.3340655068775652e-70 |
| frequencyI | 8.0232673908148838e-69 |
| ssfractionHelix | 2.567340502716567e-42 |
| harmonicCentrality | 1.6702656930961926e-23 |
| closenessCentrality | 3.9056997044935054e-19 |