PROFEAT 2016

List of Newly Added Descriptors

Protein Binding Descriptors

❖ Feature Group 4 [G4]: Composition, Transition, Distribution (CTD) Descriptors

1. [G4.x.9] Protein-protein Interface hotspot propensity-Bogan
2. [G4.x.10] Protein-protein Interface (PPI) propensity-Ma
3. [G4.x.11] Protein-DNA Interface propensity-Schneider
4. [G4.x.12] Protein-DNA Interface propensity-Ahmad
5. [G4.x.13] Protein-RNA Interface propensity-Kim
6. [G4.x.14] Protein-RNA Interface propensity-Kim
7. [G4.x.15] Protein-RNA Interface propensity-Phipps
8. [G4.x.16] Protein-ligand binding site propensity-Khazanov
9. [G4.x.17] Protein-ligand valid binding site propensity -Khazanov
10. [G4.x.18] Propensity for Protein-ligand polar and atom-Imai
11. [G4.x.19] Molecular weight
12. [G4.x.20] CLogP
13. [G4.x.21] No. of hydrogen bond donor in side chain
14. [G4.x.22] No. of hydrogen bond acceptor in side chain
15. [G4.x.23] Solubility in water
16. [G4.x.24] Amino acid flexibility index

Where, x = 1: Composition, 2: Transition, 3: Distribution

Network-Based Descriptors

According to PROFEAT feature group indexing, each network descriptor is indexed as [X, Y, Z], where node-level descriptors are indexed by X=G10 and network-level descriptors are indexed by X=G11. Secondarily, each descriptor is labelled as un-weighted, edge-weighted, node-weighted, or directed by Y=1, 2, 3, 4 respectively. The feature calculated by the normalized weight is labelled by an extra "*N*" in Y. Thirdly, Z represents the descriptor ID# in the table below.

For example, [G10, 1, 7] is the node-level un-weighted neighbourhood connectivity; [G11, 2N, 145] is the network-level normalized edge-weighted transitivity; [G10, 4, 43] is the node-level directed local clustering coefficient.

In the following table, all the descriptors are grouped into different feature categories according to their algorithm definitions, and each column lists the computed descriptors for each input network type. Some descriptors can be defined by either un-weighted connection information or weighted information. Therefore, some notations are given: "○" represents the features calculated based on un-weighted network structure, "★" represents the features calculated based on edge weight, "●" represents the features calculated based on node weight, and "↗" represents the features calculated based on directed information.

ID	Feature Category	Network Descriptor Name	Network Type			
			Un-Directed		Directed	
			Un-Weighted	Edge Weighted	Node Weighted	Un-Weighted
Node-Level Descriptors [Local]						
1	Connectivity Profiles to the Immediate Neighbors	Degree	○	○	○	
2		Scaled Connectivity	○	○	○	
3		Number of Selfloops	○	○	○	↗
4		Number of Triangles	○	○	○	↗
5		Z Score	○	○	○	
6	Connectivity Profiles to the Next Immediate Neighbors	Clustering Coefficient	○	○	○	
7		Neighborhood Connectivity	○	○	○	
8		Topological Coefficient	○	○	○	
9		Interconnectivity	○	○	○	
10		Bridging Coefficient	○	○	○	
11	Distance Relationships to All Other Nodes	Average Shortest Path Length	○	○ ★	○	
12		Distance Sum	○	○ ★	○	
13		Eccentricity	○	○ ★	○	
14		Eccentric	○	○ ★	○	
15		Deviation	○	○ ★	○	
16		Distance Deviation	○	○ ★	○	
17		Radiality	○	○ ★	○	
18	Centrality based on Degree or Distance to all Other Nodes	Degree Centrality	○	○	○	
19		Closeness Centrality (avg)	○	○ ★	○	
20		Closeness Centrality (sum)	○	○ ★	○	
21		Eccentricity Centrality	○	○ ★	○	
22		Harmonic Closeness Centrality	○	○ ★	○	
23		Residual Closeness Centrality	○	○ ★	○	
24	Centrality based on Shortest Paths Passing thru the Studied Node	Stress Centrality	○	○ ★	○	
25		Betweenness Centrality	○	○ ★	○	
26		Normalized Betweenness	○	○ ★	○	
27		Bridging Centrality	○	○ ★	○	
28	Centrality based on Connectivity and Neighbors' Centrality	Page Rank Centrality	○	○	○	
29		Eigenvector Centrality	○	○	○	
30	Edge-Weighted Descriptor	Strength		★		
31		Assortativity		★		
32		Disparity		★		
33		Geometric Mean of Triangles		★		
34		Barrat's Local Clustering Coefficient		★		
35		Onnela's Local Clustering Coefficient		★		
36		Zhang's Local Clustering Coefficient		★		
37		Holme's Local Clustering Coefficient		★		
38	Node-Weighted Descriptor	Node Weight			●	
39		Node Weighted Cross Degree			●	
40		Node Weighted Local Clustering Coefficient			●	
41	Directed Descriptor	In-Degree				↗
42		Out-Degree				↗
43		Directed Local Clustering Coefficient				↗

44		Neighbourhood Connectivity (only in)				↗
45		Neighbourhood Connectivity (only out)				↗
46		Neighbourhood Connectivity (in & out)				↗
47		Average Directed Neighbour Degree				↗
Node-Level Descriptors [Global]						
1	Basic Global Connectivity Profiles	Number of Nodes	○	○		○
2		Number of Edges	○	○		○
3		Number of Selfloops	○	○		↗
4		Maximum Connectivity	○	○		
5		Minimum Connectivity	○	○		
6		Average Number of Neighbours	○	○		
7		Total Adjacency	○	○		
8		Network Density	○	○		↗
9		Average Clustering Coefficient	○	○		
10		Transitivity	○	○		
11		Heterogeneity	○	○		
12		Degree Centralization	○	○		
13		Central Point Dominance	○	○		
14	Network Measure Based on all Shortest Paths	Total Distance	○	○ ★		
15		Network Diameter	○	○ ★		
16		Network Radius	○	○ ★		
17		Shape Coefficient	○	○ ★		
18		Characterisitc Path Length	○	○ ★		
19		Network Eccentricity	○	○ ★		
20		Average Eccentricity	○	○ ★		
21		Network Eccentric	○	○ ★		
22		Eccentric Connectivity	○	○ ★		
23		Unipolarity	○	○ ★		
24		Integration	○	○ ★		
25		Variation	○	○ ★		
26		Average Distance	○	○ ★		
27		Mean Distance Deviation	○	○ ★		
28		Centralization	○	○ ★		
29		Global Efficiency	○	○ ★		
30	Topological Index Based on Connectivity	Edge Complexity Index	○	○		
31		Randic Connectivity Index	○	○		
32		Atom-Bond Connectivity Index	○	○		
33		Zagreb Index 1	○	○		
34		Zagreb Index 2	○	○		
35		Zagreb Index Modified	○	○		
36		Zagreb Index Augmented	○	○		
37		Zagreb Index Variable	○	○		
38		Narumi-Katayama Index	○	○		
39		Narumi-Katayama Index (log)	○	○		
40		Narumi Geometric Index	○	○		
41		Narumi Harmonic Index	○	○		
42		Alpha Index	○	○		
43		Beta Index	○	○		

44		Pi Index	○	○		
45		Eta Index	○	○		
46		Hierarchy	○	○		
47		Robustness	○	○		
48		Medium Articulation	○	○		
49	Topological Index Based on Shortest Paths	Complexity Index A	○	○ ★		
50		Complexity Index B	○	○ ★		
51		Wiener Index	○	○ ★		
52		Hyper-Wiener	○	○ ★		
53		Harary Index 1	○	○ ★		
54		Harary Index 2	○	○ ★		
55		Compactness Index	○	○ ★		
56		Superpendentic Index	○	○ ★		
57		Hyper-Distance-Path Index	○	○		
58		BalabanJ Index	○	○ ★		
59		BalabanJ-like 1 Index	○	○ ★		
60		BalabanJ-like 2 Index	○	○ ★		
61		BalabanJ-like 3 Index	○	○ ★		
62		Geometric Arithmetic Index 1	○	○		
63		Geometric Arithmetic Index 2	○	○ ★		
64		Geometric Arithmetic Index 3	○	○ ★		
65		Szeged Index	○	○ ★		
66		Product Of Row Sums	○	○ ★		
67		Product Of Row Sums (log)	○	○ ★		
68		Schultz Topological Index	○	○ ★		
69		Gutman Topological Index	○	○ ★		
70		Efficiency Complexity	○	○ ★		
71	Entropy-Based Complexity Descriptors	Information Content (Degree Equality)	○	○		
72		Information Content (Edge Equality)	○	○		
73		Information Content (Edge Magnitude)	○	○		
74		Information Content (Distance Degree)	○	○		
75		Information Content (Distance Degree Equality)	○	○		
76		Radial Centric Information Index	○	○		
77		Distance Degree Compactness	○	○		
78		Distance Degree Centric Index	○	○		
79		Graph Distance Complexity	○	○		
80		Information Layer Index	○	○		
81		Bonchev Information Index 1	○	○		
82		Bonchev Information Index 2	○	○		
83		Bonchev Information Index 3	○	○		
84		Balaban-like Information Index 1	○	○		
85		Balaban-like Information Index 2	○	○		
86	Eigenvalue-Based Connectivity Descriptors	Graph Energy	○	○		
87		Laplacian Energy	○	○		
88		Spectral Radius	○	○		
89		Estrada Index	○	○		
90		Laplacian Estrada Index	○	○		
91		Quasi-Weiner Index	○	○		

92		Mohar Index 1	○	○		
93		Mohar Index 2	○	○		
94		Graph Index Complexity	○	○		
95		Adjacency Matrix HM	○	○		
96		Adjacency Matrix SM	○	○		
97		Adjacency Matrix ISM	○	○		
98		Adjacency Matrix PM	○	○		
99		Adjacency Matrix IPM	○	○		
100		Laplacian Matrix HM	○	○		
101		Laplacian Matrix SM	○	○		
102		Laplacian Matrix ISM	○	○		
103		Laplacian Matrix PM	○	○		
104		Laplacian Matrix IPM	○	○		
105		Distance Matrix HM	○	○		
106		Distance Matrix SM	○	○		
107		Distance Matrix ISM	○	○		
108		Distance Matrix PM	○	○		
109		Distance Matrix IPM	○	○		
110		Distance Path Matrix HM	○	○		
111		Distance Path Matrix SM	○	○		
112		Distance Path Matrix ISM	○	○		
113		Distance Path Matrix PM	○	○		
114		Distance Path Matrix IPM	○	○		
115		Augmented Vertex Degree Matrix HM	○	○		
116		Augmented Vertex Degree Matrix SM	○	○		
117		Augmented Vertex Degree Matrix ISM	○	○		
118		Augmented Vertex Degree Matrix PM	○	○		
119		Augmented Vertex Degree Matrix IPM	○	○		
120		Extended Adjacency Matrix HM	○	○		
121		Extended Adjacency Matrix SM	○	○		
122		Extended Adjacency Matrix ISM	○	○		
123		Extended Adjacency Matrix PM	○	○		
124		Extended Adjacency Matrix IPM	○	○		
125		Vertex Connectivity Matrix HM	○	○		
126		Vertex Connectivity Matrix SM	○	○		
127		Vertex Connectivity Matrix ISM	○	○		
128		Vertex Connectivity Matrix PM	○	○		
129		Vertex Connectivity Matrix IPM	○	○		
130		Random Walk Markov Matrix HM	○	○		
131		Random Walk Markov Matrix SM	○	○		
132		Random Walk Markov Matrix ISM	○	○		
133		Random Walk Markov Matrix PM	○	○		
134		Random Walk Markov Matrix IPM	○	○		
135		Weighted Struct. Func. Matrix IM1 HM	○	○		
136		Weighted Struct. Func. Matrix IM1 SM	○	○		
137		Weighted Struct. Func. Matrix IM1 ISM	○	○		
138		Weighted Struct. Func. Matrix IM1 PM	○	○		
139		Weighted Struct. Func. Matrix IM1 IPM	○	○		

140		Weighted Struct. Func. Matrix IM2 HM	○	○		
141		Weighted Struct. Func. Matrix IM2 SM	○	○		
142		Weighted Struct. Func. Matrix IM2 ISM	○	○		
143		Weighted Struct. Func. Matrix IM2 PM	○	○		
144		Weighted Struct. Func. Matrix IM2 IPM	○	○		
145	Edge-Weighted Descriptors	Weighted Transitivity		★		
146		Barrat's Global Clustering Coefficient		★		
147		Onnela's Global Clustering Coefficient		★		
148		Zhang's Global Clustering Coefficient		★		
149		Holme's Global Clustering Coefficient		★		
150	Node-Weighted Descriptor	Total Node Weight			●	
151		Node Weighted Global Clustering Coefficient			●	
152	Directed Descriptor	Average In-Degree				↗
153		Maximum In-Degree				↗
154		Minimum In-Degree				↗
155		Average Out-Degree				↗
156		Maximum Out-Degree				↗
157		Minimum Out-Degree				↗
158		Directed Global Clustering Coefficient				↗