Structural Discrimination of Networks by Using Distance, Degree and Eigenvalue-Based Measures

Matthias Dehmer[1]*, Martin Grabner[1], Boris Furtula[2]

1 UMIT, Institute for Bioinformatics and Translational Research, Hall in Tyrol, Austria, **2** University of Kragujevac, Faculty of Science, Kragujevac, Serbia

Abstract

In chemistry and computational biology, structural graph descriptors have been proven essential for characterizing the structure of chemical and biological networks. It has also been demonstrated that they are useful to derive empirical models for structure-oriented drug design. However, from a more general (complex network-oriented) point of view, investigating mathematical properties of structural descriptors, such as their uniqueness and structural interpretation, is also important for an in-depth understanding of the underlying methods. In this paper, we emphasize the evaluation of the uniqueness of distance, degree and eigenvalue-based measures. Among these are measures that have been recently investigated extensively. We report numerical results using chemical and exhaustively generated graphs and also investigate correlations between the measures.

Introduction

Structural analysis of graphs has been an outstanding problem in graph theory for several decades [1–4]. A challenging problem in this theory is to investigate structural features of the graphs and their characterization. Another important task is to quantify the structural features of graphs, as well as their complexity [2,3,5,6]. The former relates to developing measures such as the clustering coefficient or the average distance of a graph [7]. The latter relates to deriving complexity indices for graphs, which are often called structural descriptors/measures or topological indices [8–11].

In this paper, we deal with evaluating the uniqueness, discrimination power or degeneracy of special graph measures for investigating graphs holistically (in contrast to local graph measures) [12]. A descriptor is called degenerate if it possesses the same value for more than one graph. In view of the large body of literature on structural graph measures [2,3,5,13], the degeneracy problem has been somewhat overlooked in graph theory. In fact, the uniqueness of structural descriptors has been investigated in mathematical chemistry and related disciplines for discriminating the structure of isomeric structures and other chemical networks [14–16]. A detailed survey on the uniqueness of topological indices by using isomers and hexagonal graphs has been given by Konstantinova [16]. For more related work, see also [17].

To date, no complete graph invariant, i.e., a measure that is fully unique on general graphs, has been found. Indeed, some measures turned out to be complete by using special sets of graphs [15,17,18]. In a more general context, i.e., by using graphs without structural constraints, any topological graph measure has a certain kind of degeneracy, which also depends on the mathematical

method to define the measure, see [19,20]. A highly discriminating graph measure is desirable for analyzing graphs; hence, measuring the degree of its degeneracy is important for understanding its properties, limits and quality.

The main contribution of this paper is to investigate to what extent known degree, distance and eigenvalue-based measures are degenerate. Among the measures we examine (see Table 1) are the recently developed geometric-arithmetic indices [21,22], the atom-bond connectivity index [23] and the Estrada index [24], which is based on the eigenvalues of a special graph-theoretical matrix [25], here the adjacency and Laplacian matrix. It turns out that some of the measures based on distances and eigenvalues are highly unique in exhaustively generated graphs (e.g., see Table 2). Using these graphs is a greater challenge than only using isomeric structures, as exhaustively generated graphs do not possess any structural constraints. However, it is clear that other distance or eigenvalue-based measures exist that possess only low discrimination power [26], implying that the uniqueness of a measure crucially depends on its mathematical composition and the graph class under consideration.

Methods and Results

Uniqueness of Topological Descriptors

In this section, we present numerical results when evaluating the uniqueness of certain topological descriptors. Note that a summary of the topological indices used in this paper can be found in Table 1. As mentioned, the discrimination power of these measures has not yet been evaluated extensively on a large scale. Therefore, the results might be useful for gaining deeper insights

Table 1. The topological indices used for determining the value distributions and correlation plots.

Index Name	Symbol
Atom-bond connectivity index [23]	ABC
Augmented Zagreb index [40]	AZI
Variable Zagreb index [41]	VZI
Modified Zagreb index [42]	MZI
Narumi-Katayama index [43]	NK
Distance degree centric index [8,44]	I_C
Offdiagonal complexity [45]	OdC
Medium articulation [46]	MA_R
Degree-degree association index [29]	$I_{f_{exp}^{\lambda}}^{\lambda}$
First geometric-arithmetic index [21]	GA_1
Second geometric–arithmetic index [22]	GA_2
Third geometric–arithmetic index [31]	GA_3
Efficiency complexity [26]	C_e
Graph energy [47]	E
Laplacian energy [48]	LE
Estrada index [24]	EE
Laplacian Estrada index [49]	LEE
Spectral radius [10]	λ_{max}
Graph index complexity [26]	C_r
Balaban index [19]	J
Degree information index [8]	I_δ
Topological information content [6]	I_a
Vertex complexity [50]	I_V

doi:10.1371/journal.pone.0038564.t001

into these measures and for enabling implications when designing novel topological descriptors. As usual, we use the measure

$$S_I = (|\mathcal{G}| - \mathrm{ndv})/|\mathcal{G}|, \qquad (1)$$

which was called the sensitivity by Konstantinova [15], for evaluating the uniqueness of an index I. Clearly, S_I depends on a graph class $\mathcal{G}$; ndv are the values that cannot be distinguished by I, and $|\mathcal{G}|$ is the size of the graph set. Now, we start interpreting the results by considering Table 2 and observe that we have arranged the used descriptors into four groups. We also emphasize that the values in Table 2 have been calculated by using the graph classes N_i, $i=8,9,10$. These are the classes of exhaustively generated non-isomorphic, unweighted and connected graphs with i vertices each. The cardinalities $|N_i|$ are also depicted in Table 2.

For the degree-based indices, it is not surprising that these measures have only little discrimination power, as many graphs can be realized by identical degree sequences. This effect is even stronger if the cardinality of the underlying graph set increases, see Table 2. The highest discrimination power among the indices of this class has the ABC index. This is in accordance with the well-known fact that the degeneracy of topological descriptors decreases in the following order: firstgeneration(e.g.,NK) $\geq$ secondgeneration(e.g.,ABC) $\geq$ third-generation, see [27]. Recall that first-generation indices are integer measures derived from integer local vertex invariants such as vertex degrees or distances sums [28]. Second-generation indices are real numbers derived from integer local vertex invariants [28]. Third-

Table 2. Exhaustively generated sets of non-isomorphic and generated graphs.
$|N_8| = 11117$, $|N_9| = 261080$ and $|N_{10}| = 11716571$.

Index	N_8 ndv	S	N_9 ndv	S	N_{10} ndv	S
Degree-based Measures						
ABC	8520	0,233606	241793	0,073874	11539714	0,015095
AZI	8520	0,233606	241777	0,073935	11539377	0,015123
GA_1	8522	0,233426	242009	0,073047	11542066	0,014894
VZI	10500	0,055501	258286	0,010702	11704386	0,001040
MZI	10496	0,055860	258293	0,010675	11704428	0,001036
NK	10974	0,012863	260925	0,000594	11716377	0,000017
Information-theoretic Measures						
I_C	11116	0,000090	261079	0,000004	11716570	0,000000
OdC	10731	0,034722	259967	0,004263	11713337	0,000276
MA_R	10879	0,021409	260576	0,001930	11715462	0,000095
$I_{f_{exp}^{\lambda}}^{\lambda}$	385	0,965368	6016	0,976957	609204	0,948005
Distance-based Measures						
GA_2	1044	0,906090	40014	0,846737	3693236	0,684785
GA_3	663	0,940362	15228	0,941673	673972	0,942477
C_e	11076	0,003688	261020	0,000230	11716455	0,000010
Eigenvalue-based Measures						
E	1628	0,853558	47577	0,817769	2413055	0,794048
EE	751	0,932446	26457	0,898663	1460054	0,875386
LE	5098	0,541423	59542	0,771940	2338347	0,800424
LEE	1013	0,908878	23393	0,910399	718156	0,938706
λ_{max}	2003	0,819825	48120	0,815689	2137087	0,817601
Non-information-theoretic Measures						
C	10950	0,015022	260861	0,000839	11716146	0,000036
Cr	1779	0,839975	44652	0,828972	2098604	0,820886

doi:10.1371/journal.pone.0038564.t002

generation indices are real numbers derived from real local vertex invariants [28].

Most of the information-theoretic measures (e.g., I_C, OdC) we have evaluated in this study are based on grouping elements (e.g., vertices, degrees, etc.) in equivalence classes [6,8] to determine probability values. We observe that the uniqueness of these measures is also low. In contrast, the degree-degree association index $I_{f_{exp}^{\lambda}}^{\lambda}$ [29] is highly discriminating for all three graph classes [30]. Surely, a reason for this is the fact that this measure is non-partition-based, as probability values have been assigned to each vertex in the graph by using the special information functional f^{Δ}, see [29]. Note that N_{10} contains almost 12 million graphs. Calculating the discrimination power of the distance-based measures, such as the second or third geometric-arithmetic indices [22,31], leads to a somewhat surprising result: the uniqueness for N_8, N_9 and N_{10} is very high, but recall that they belong to the class of so-called second-generation indices [27]. Again, we see that the composition of the graph invariant (here, distances) to define the measure is crucial.

If we compare the sensitivity values (using Equation 1) of some second-generation indices, e.g., the geometric-arithmetic indices with some of the third-generation indices (information-theoretic and eigenvalue-based measures), we observe that the uniqueness of

Table 3. Chemical isomers with $|V| = 11, 12$. $|C_{11}^{iso}| = 160294$, $|C_{12}^{iso}| = 738928$.

Index	C_{11}^{iso}		C_{12}^{iso}	
	ndv	S	ndv	S
Degree-based Measures				
GA_1	160063	0,001441	738685	0,000329
ABC	160089	0,001279	738714	0,000290
AZI	160093	0,001254	738721	0,000280
VZI	160290	0,000025	738924	0,000005
MZI	160290	0,000025	738924	0,000005
NK	160293	0,000006	738927	0,000001
Information-theoretic Measures				
I_C	160292	0,000012	738925	0,000004
OdC	160281	0,000081	738916	0,000016
MA_R	160291	0,000019	738926	0,000003
$I_{f_{exp}^A}^{\lambda}$	1479	0,990773	18852	0,974487
Distance-based Measures				
GA_2	23548	0,853095	118000	0,840309
GA_3	11046	0,931089	60597	0,917993
C_e	160036	0,001610	738454	0,000641
Eigenvalue-based Measures				
E	24417	0,847674	110075	0,851034
EE	19590	0,877787	88842	0,879769
LE	22982	0,856626	104151	0,859051
LEE	10062	0,937228	39634	0,946363
λ_{max}	28195	0,824104	117781	0,840606
Non-information-theoretic Measures				
C	160293	0,000006	738927	0,000001
Cr	21432	0,866296	91321	0,876414

Table 4. Chemical trees with $|V| = 20, 21, 22$. $|C_{20}| = 366319$, $|C_{21}| = 910726$, $|C_{22}| = 2278658$.

Index	C_{20}		C_{21}		C_{22}	
	ndv	S	ndv	S	ndv	S
Degree-based Measures						
GA_1	366257	0,000169	910662	0,000070	2278593	0,000029
ABC	366303	0,000044	910710	0,000018	2278640	0,000008
AZI	366303	0,000044	910710	0,000018	2278640	0,000008
VZI	366318	0,000003	910722	0,000004	2278657	0,000000
MZI	366318	0,000003	910722	0,000004	2278657	0,000000
NK	366318	0,000003	910725	0,000001	2278657	0,000000
Information-theoretic Measures						
I_C	366283	0,000098	910688	0,000042	2278608	0,000022
OdC	366311	0,000022	910718	0,000009	2278652	0,000003
MA_R	366317	0,000005	910725	0,000001	2278657	0,000000
$I_{f_{exp}^A}^{\lambda}$	196124	0,464609	544432	0,402200	39396	0,982711
Distance-based Measures						
GA_2	362628	0,010076	904971	0,006319	2266566	0,005307
GA_3	362171	0,011323	904971	0,006319	2270582	0,003544
C_e	319073	0,128975	813531	0,106723	2081010	0,086739
Eigenvalue-based Measures						
E	93204	0,745566	228831	0,748738	479746	0,789461
EE	87656	0,760711	224579	0,753407	525472	0,769394
LE	544	0,998515	880	0,999034	1275	0,999440
LEE	292	0,999203	509	0,999441	842	0,999630
λ_{max}	130783	0,642981	318330	0,650466	675147	0,703708
Non-information-theoretic Measures						
C	366318	0,000003	910725	0,000001	2278657	0,000000
Cr	69592	0,810024	160051	0,824260	316572	0,861071

e.g., GA_2, GA_3 is unexpectedly high. In particular, the high uniqueness of GA_3 for graphs $\in N_i$, $i = 8, 9, 10$, is probably caused by the fact that its calculation is based on distances between edges. As the number of edges lies in the interval $[n-1, n(n-1)/2]$, the range of the third geometric-arithmetic index is 0 to $n(n-1)/2$ [32], and the probability that two graphs have different index values is certainly larger than in the case when the number of edges would be fixed. This hypothesis can be supported by comparing the values of the sensitivity index (using Equation 1) of the GA_3 index shown in Tables 2 and 4. Thus, the sensitivity index resulting from GA_3 shown in Table 2 is greater than 0.94 ($\geq 94\%$), while, if the number of edges is fixed, see Table 4, the corresponding sensitivity index is less than 0.02 ($\leq 2\%$). Using this idea again, it can be understood why the sensitivity index of GA_3 (see Table 2) does not decrease with the number of vertices.

Let us turn to the uniqueness of some eigenvalue-based measures such as the graph energy E, the Estrada index EEo and the Laplacian Estrada index LEE. As expected, it is high because these measures belong to the class of third-generation indices (e.g., information-theoretic measures). We point out that the sensitivity index of the graph energy E and Laplacian energy LE could be affected by rounding errors. The reason for this is based on the fact that the difference between the values of E and LE for some graphs is less than 10^{-8} [33]. However, since the number of such graphs is very small, see [33], this does not strongly affect the computation of the uniqueness of E and LE measured by S and ndv. In particular, the Estrada and Laplacian Estrada indices possess high uniqueness for all three graph classes N_i. To give some arguments for this, recall their definitions, namely

$$EE = \sum_{i=1}^{n} e^{\lambda_i}, \qquad (2)$$

$$LEE = \sum_{i=1}^{n} e^{\mu_i}, \qquad (3)$$

where λ_i and μ_i are the eigenvalues of the adjacency and Laplacian matrices, respectively. Knowing that e is irrational and transcendental, it can be presumed that any power and the sum thereof is also irrational and transcendental. Hence, the graphs with the same Estrada (Laplacian Estrada) index are isospectral.

In addition, the uniqueness of these measures is quite stable, and the same holds for $I_{f_{exp}^A}^{\lambda}$. This means that there is only very little dependency between their uniqueness and the cardinality of the underlying graph set. Clearly, this result demonstrates that certain
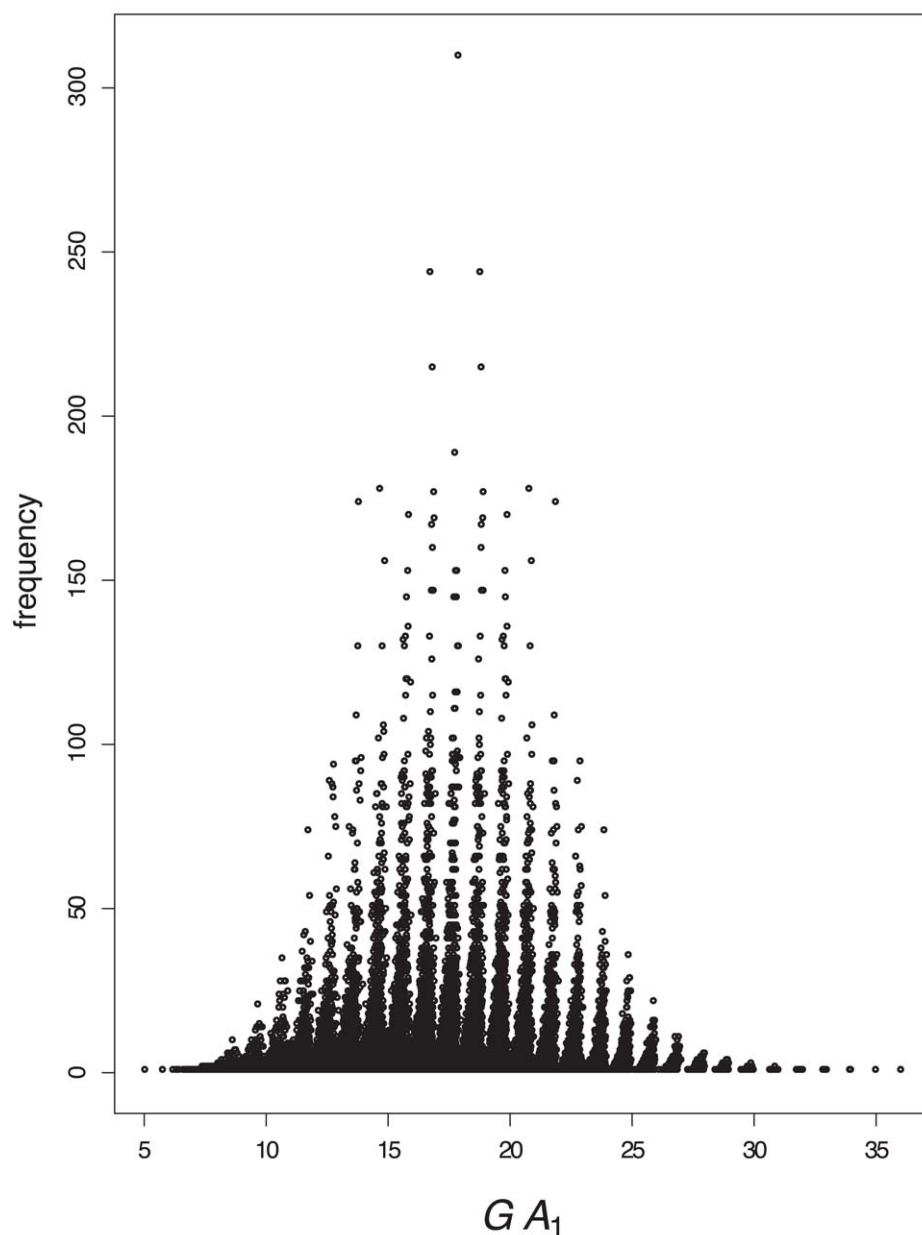
Figure 1. Value distribution for GA_1.
doi:10.1371/journal.pone.0038564.g001

measures/functions based on the eigenvalues of graphs possess a high discrimination power. This contradicts the widely assumed hypothesis that graph spectra are not feasible to discriminate graphs properly because of the existence of isospectral graphs, see [34,35]. Another positive example can be found in [36] where Dehmer et al. presented spectrum-based measures based on a probability distribution of structural values with low degeneracy.

In Table 3 and Table 4, we have also evaluated the discrimination power of the measures using isomers and chemical trees. In particular, we use the isomeric classes C_{11}^{iso} and C_{12}^{iso} containing all isomers with 11 and 12 vertices, see Table 3. The numerical results are quite similar to Table 2. However, when evaluating the indices by using the classes of chemical trees C_{20}, C_{21} and C_{22}, we see that the discrimination power of $I_{f_{exp}^{\Delta}}^{\lambda}$ deteriorates significantly. To better understand this, note that the

information functional $f_{exp}^{\Delta}(v_i)$ relies on determining the shortest paths for all $v_i \in V$ and, then, degree-degree associations thereof resulting in $f_{exp}^{\Delta}(v_i)$, see [29]. Finally, when applying this measure to trees, the reason for the deterioration of its uniqueness could be understood by the occurrence of a large number of paths possessing similar length and, hence, resulting in very similar probability values and entropies. Interestingly, the eigenvalue-based measures LE and LEE possess high uniqueness, and whose values are almost independent of the cardinality of the graph sets. Thus, these measures turned out to be quite feasible to discriminate chemical trees uniquely.

Value Distributions

In order to tackle the question of what kind of degeneracy the measures possess, we plot their characteristic value distributions.
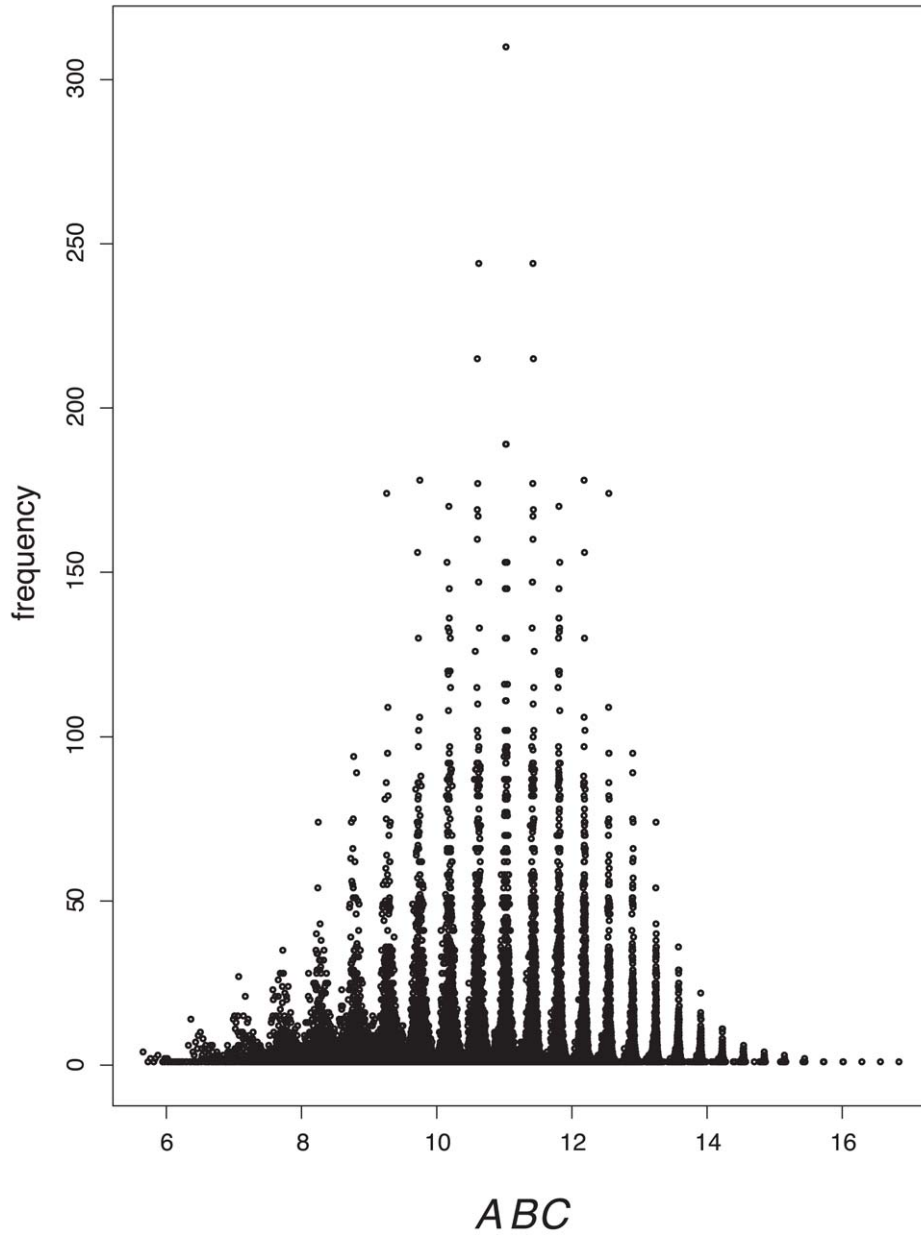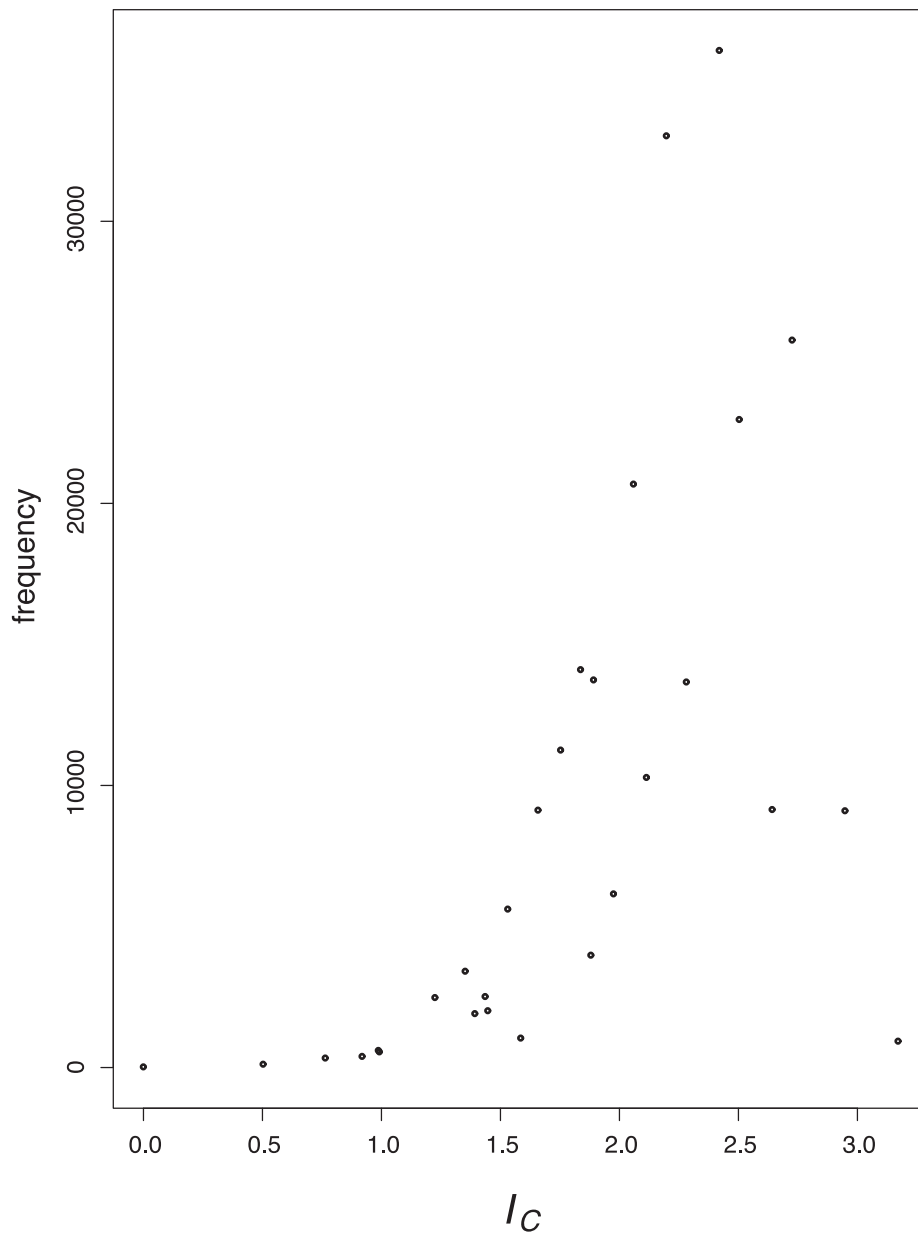
Figure 2. Value distribution for *ABC.*

Figure 3. Value distribution for I_C.

Figure 4. Value distribution for *OdC.*
doi:10.1371/journal.pone.0038564.g004
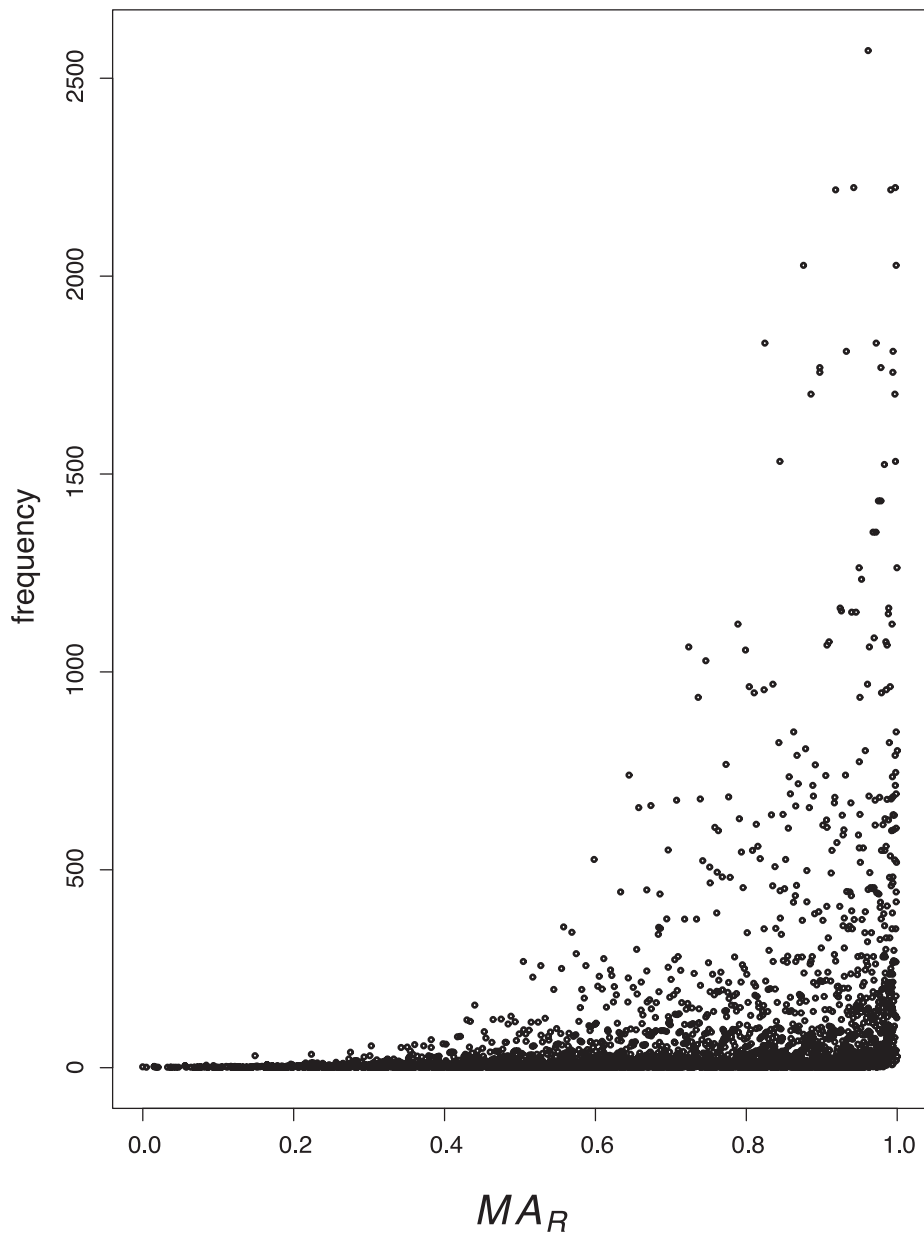
Figure 5. Value distribution for *MA.*
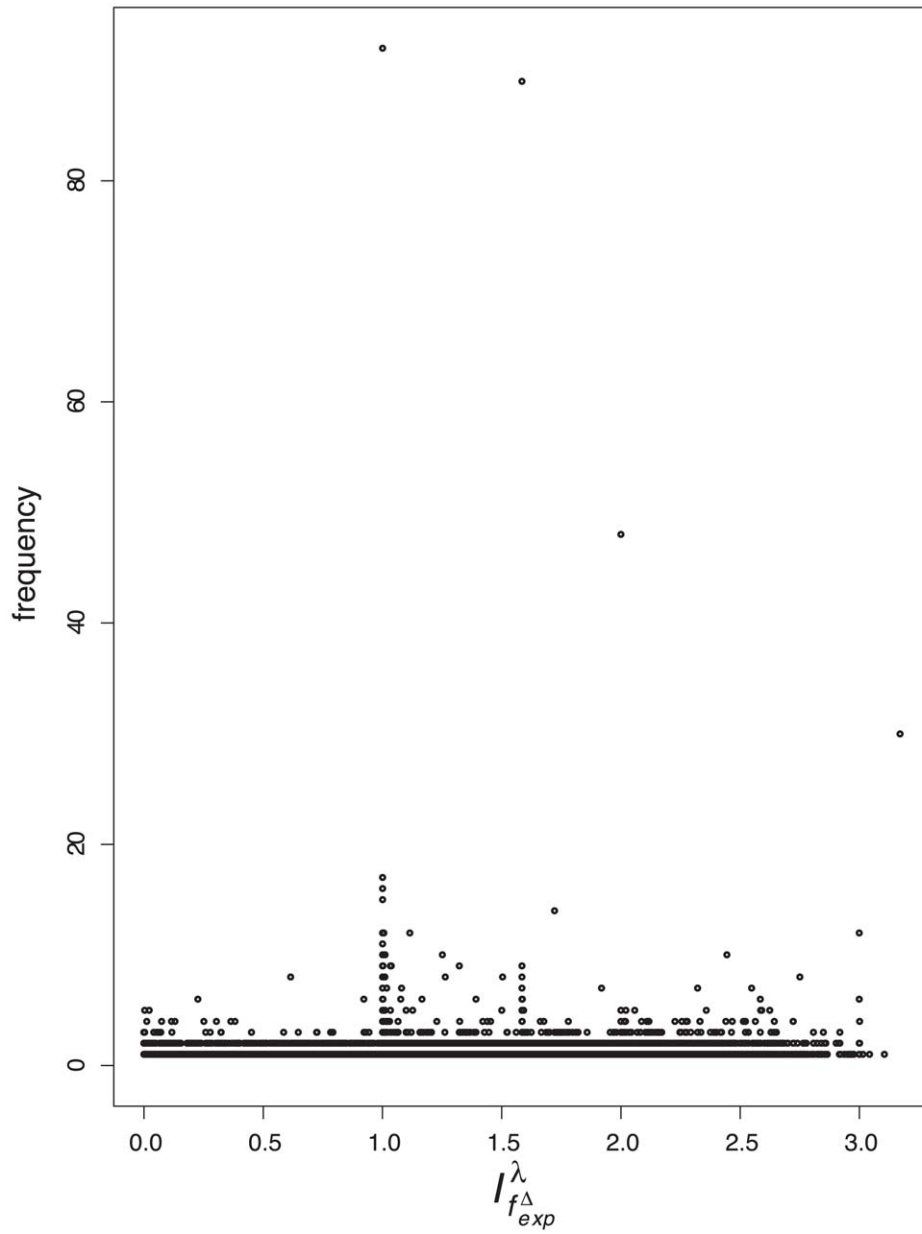
Figure 6. Value distribution for $I^{\lambda}_{f^{\Delta}_{exp}}$.
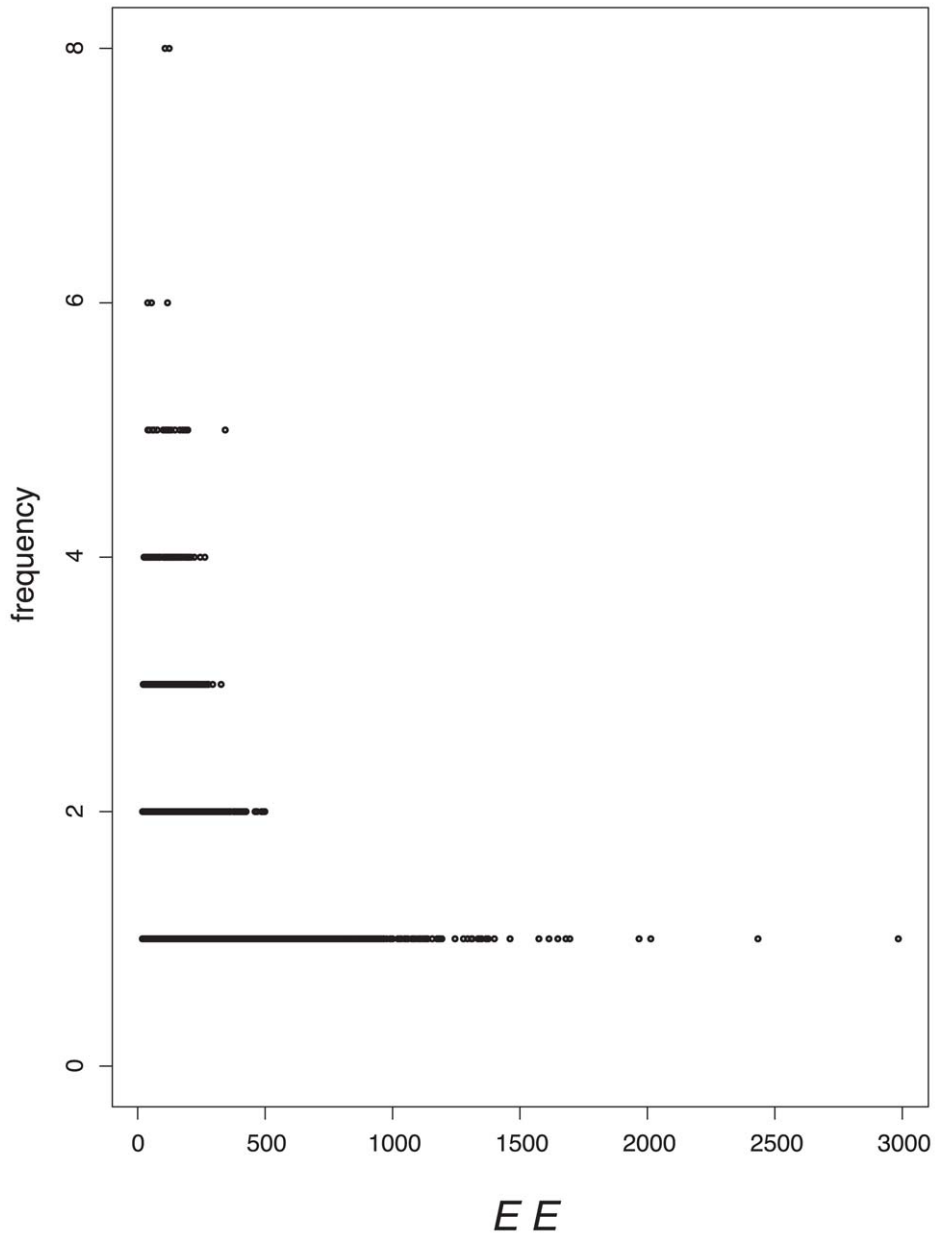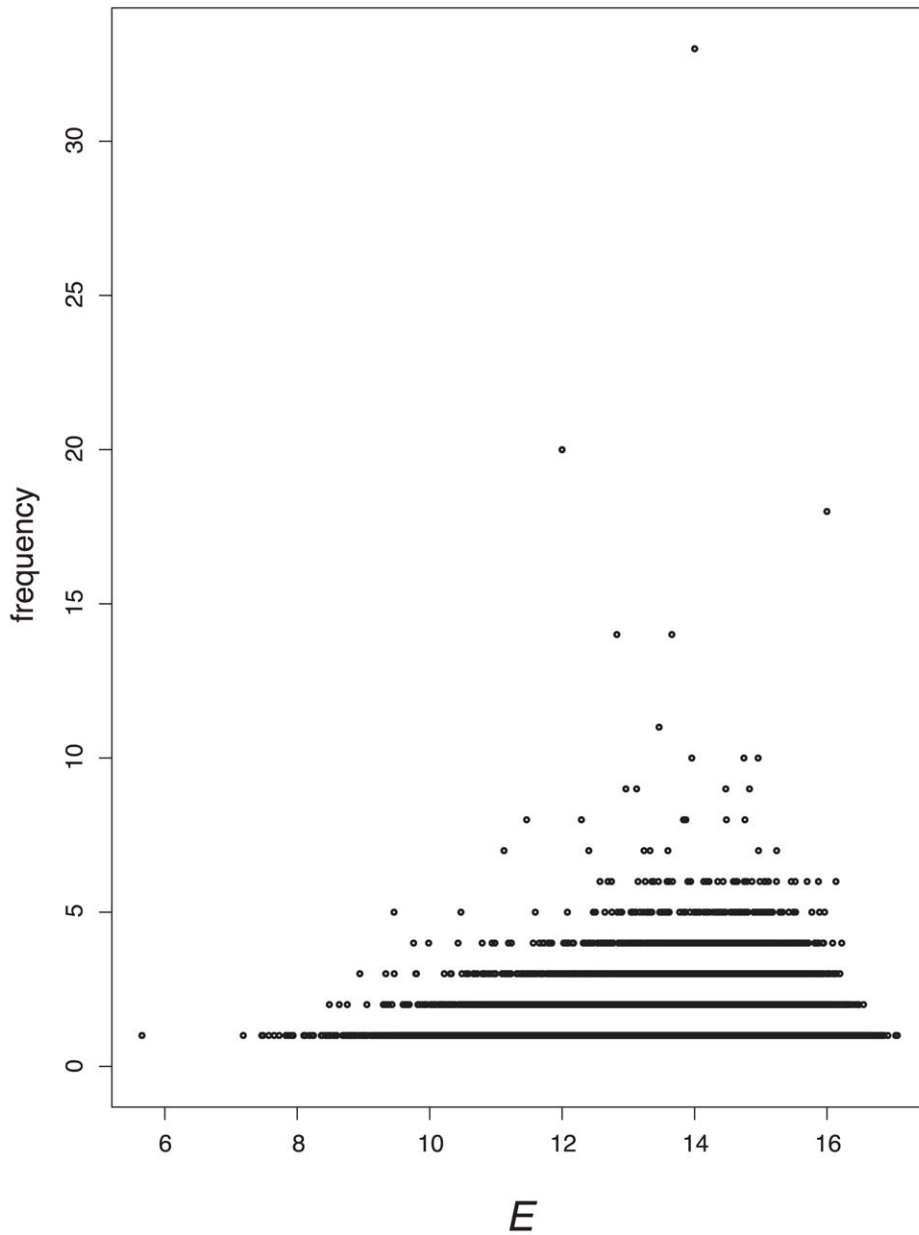
Figure 7. Value distribution for *E*.

Figure 8. Value distribution for *EE*.
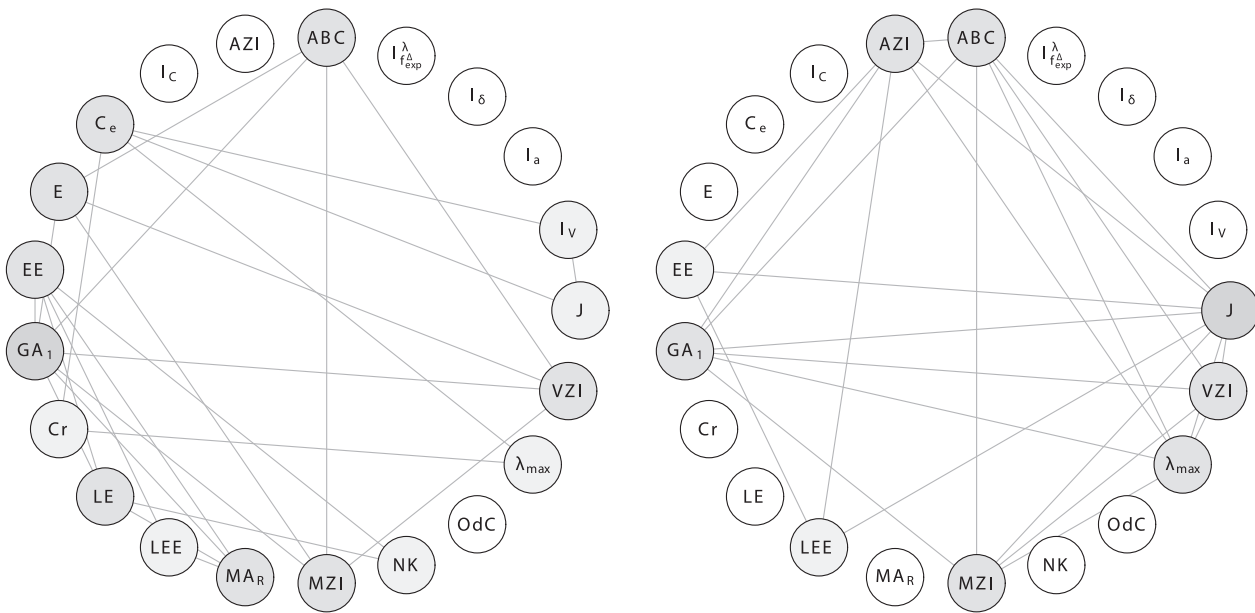doi:10.1371/journal.pone.0038564.g008

Figure 9. Left: Correlation network $G^1_{\geq 0.9}$ inferred from C_{21}. Right: Correlation network $G^2_{\geq 0.9}$ inferred from N_9.
doi:10.1371/journal.pone.0038564.g009

The y-axis is the absolute frequency of the graphs, with a certain index value depicted on the x-axis. For a graph class, we use the class of exhaustively generated non-isomorphic, connected and unweighted graphs denoted by N_9. We start with Figures 1 and 2 and observe the vertical strips, indicating that a large number of graphs have quite similar index values discretely distributed on a certain interval. In addition, the hull of these value distributions looks like a Gaussian curve. This means that by using GA_1 and ABC, there exist many degenerate graphs possessing quite similar index values where the hull of the distributions forms a Gaussian curve.

As we can see from Figures 3, 4, 5, 6, the value distribution (and in fact the distribution of degenerate graphs) when considering the information-theoretic measures is significantly different. We start with I_C, and see that the value distribution is quite scattered, i.e., there are no regions in which the graphs are closely clustered. In contrast, the values of OdC are rather clustered. Similarly, this also holds for MA_R and observe that all three measures (I_C, OdC and MA_R) are highly degenerate on N_9. But, the degree-degree association index $I^\lambda_{f_{exp}}$ possesses a high discrimination power (see Figure 6). In particular, we see that there exist only a very few degenerate graphs whose index values exploit the entire domain.

The results of plotting the value distributions for the eigenvalue-based measures graph energy E and Estrada index EE are depicted in Figures 7 and 8. We see that they possess a high
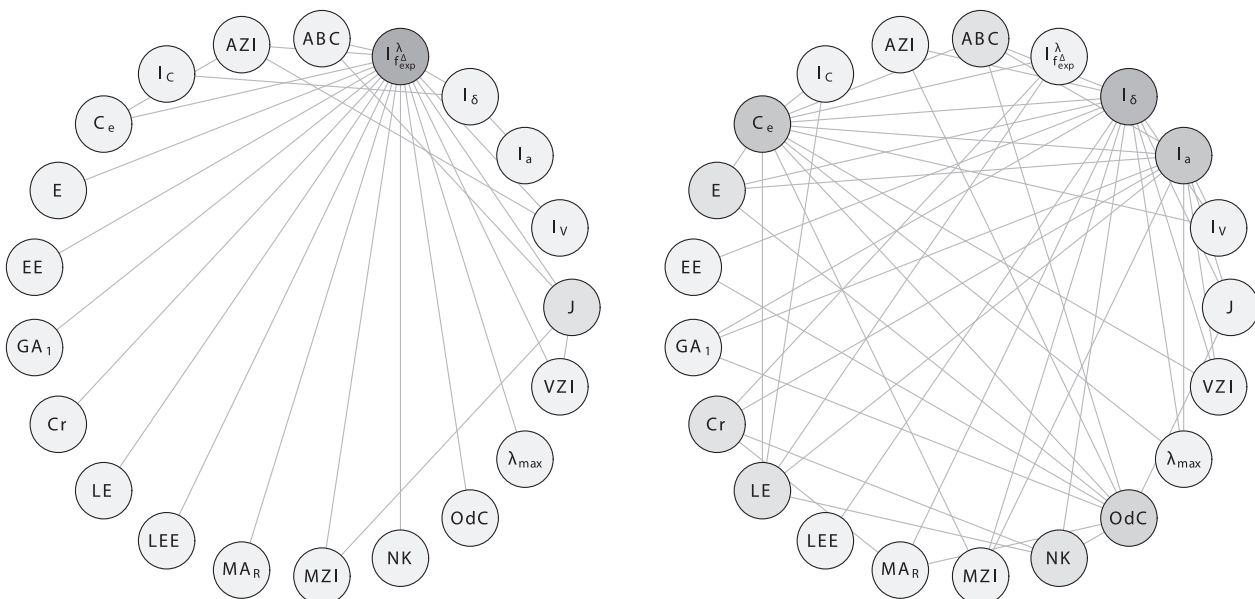

Figure 10. Left: Correlation network $G^1_{\leq 0.2}$ inferred from C_{21}. Right: Correlation network $G^2_{\leq 0.2}$ inferred from N_9.
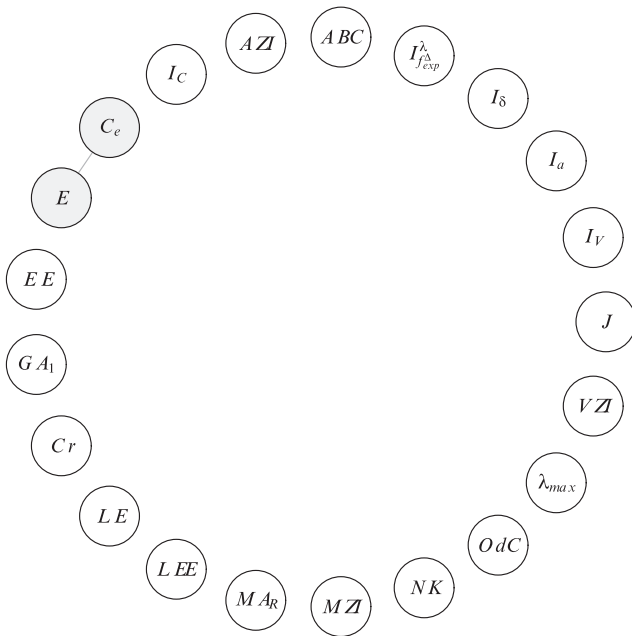doi:10.1371/journal.pone.0038564.g010

Figure 11. Correlation network $G_{\leq 0.01}$ inferred from N_9.
doi:10.1371/journal.pone.0038564.g011

Figure 12. Correlation network $G_{\leq 0.01}$ inferred from C_{21}.
doi:10.1371/journal.pone.0038564.g012

discrimination power and observe the horizontal strips. This means that a certain number of graphs (e.g., 2, 4, etc.) possess index values in a certain domain. When considering Figure 7, the horizontal strip for $y=1$ indicates the low degeneracy of this measure. This is similar for the EE shown in Figure 8.

Correlations Between Indices

In order to investigate the correlation ability of the topological indices, we calculate the linear correlation between them and depict the results as correlation networks. More precisely, the linear correlation between the descriptor values of two data vectors has been computed according to the method of Pearson [37]. In the depicted plots of the correlation networks, the calculated Pearson Product-Moments have then been used as edge weights for labeling the edges connecting the vertices representing the compared descriptor pairs. The correlation networks are shown in Figures 9, 10, 11, 12, 13, 14.

We use the graph classes C_{21} and N_9, and choose different thresholds for the correlation coefficient, resulting in different networks.

Definition 1. Let $\{I_1,\ldots,I_k\}$ be a set of topological indices defined on a graph class $\mathcal{G}$ and let $\theta \leq 1$. The vertex and edge set of the correlation network $G_{\geq \theta} := (V,E)$ inferred from $\mathcal{G}$ is defined by

$$V := \{I_1,\ldots,I_k\} \quad and \quad E := \big\{\{I_i,I_j\} : |\rho(I_i,I_j)| \geq \theta\big\}, \quad (4)$$

where $\rho \in [-1,1]$ is the correlation coefficient.

Definition 2. Let $\{I_1,\ldots,I_k\}$ be a set of topological indices defined on a graph class $\mathcal{G}$ and let $\theta \leq 1$. The vertex and edge set of the correlation network $G_{\leq \theta} := (V,E)$ inferred from $\mathcal{G}$ is defined by

$$V := \{I_1,\ldots,I_k\} \quad and \quad E := \big\{\{I_i,I_j\} : |\rho(I_i,I_j)| \leq \theta\big\}, \quad (5)$$
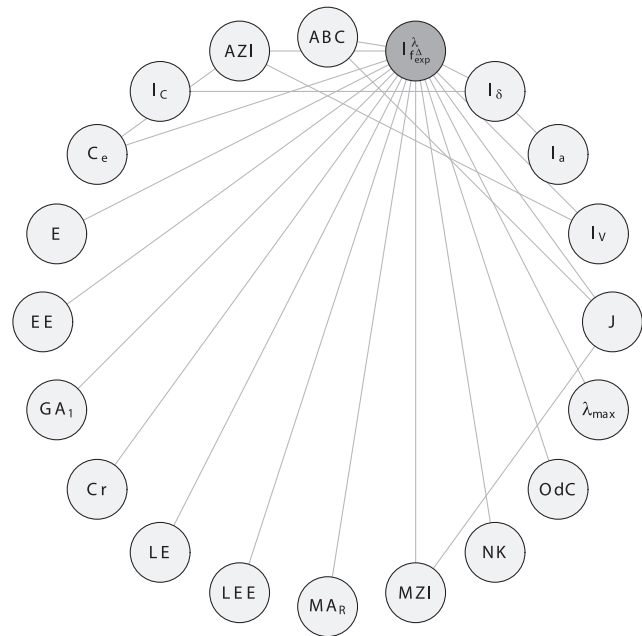
where $\rho \in [-1,1]$ is the correlation coefficient.

We start interpreting the results by considering the left-hand side of Figure 9. The vertices of the graph $G^1_{\geq 0.9}$ represent indices that are highly correlated (here, $|\rho| \geq 0.9$) by using the graph class C_{21}. In all correlation graphs, hub vertices, i.e., those with a high degree, are colored in gray. In particular, the grayer the color of a vertex is, the higher its degree.

In $G^1_{\geq 0.9}$, the first geometric-arithmetic index (GA_1) and other measures are highly correlated with other indices that belong to different groups, e.g., degree-based and eigenvalue-based, etc. In addition, graph energy (E) and Estrada index (EE) are highly correlated with other measures such as the Modified Zagreb index
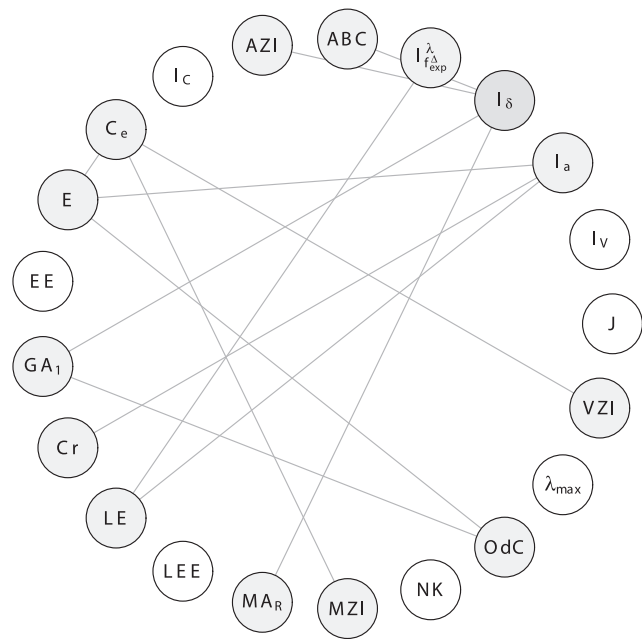

Figure 13. Correlation network $G_{\leq 0.05}$ inferred from N_9.
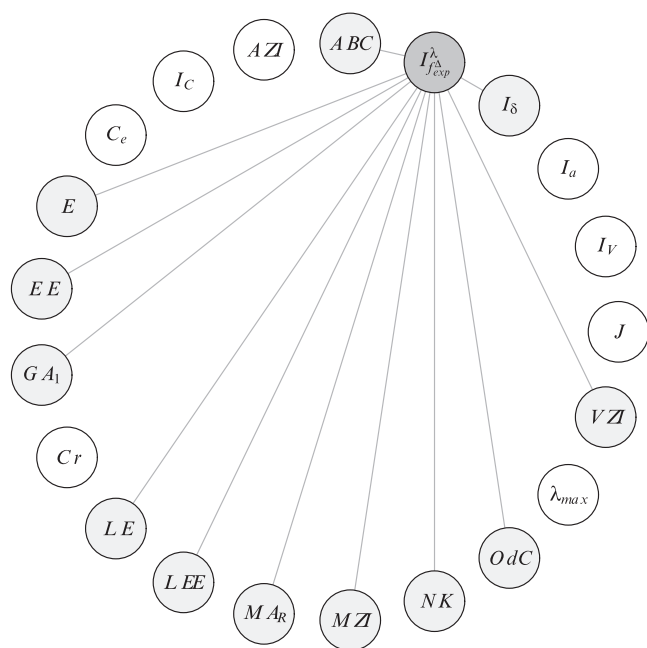doi:10.1371/journal.pone.0038564.g013

Figure 14. Correlation network $G_{\leq 0.05}$ inferred from C_{21}.
doi:10.1371/journal.pone.0038564.g014

(degree-based). By using the graph class N_9, we obtain the same type of correlation network denoted by $G^2_{\geq 0.9}$. Observe that the connectedness of this network is similarly high in $G^1_{\geq 0.9}$, however, there exist new hubs. For instance, the Balaban J and the augmented Zagreb index (AZI) index represent such vertices, i.e., they are highly correlated with other indices from different paradigms such as degree-based and eigenvalue-based measures. Interestingly, the uniqueness (measured by ndv and S) of, e.g., AZI and λ_{max} by using N_9 is higher than by taking C_{21} into account. Nevertheless, these indices (and others) possess larger neighborhoods compared to C_{21}. This means that they contain more highly correlated vertices adjacent to AZI and λ_{max} than by using C_{21}. One would have expected this in a reverse order as the isomers (C_{21}) are structurally more similar among each other than the graphs contained in N_9. It is likely that the reasons for this are different structural characteristics captured by the underlying graphs of N_9 and C_{21}.

For studying indices that are only slightly correlated, firstly consider $G^1_{\leq 0.2}$ in Figure 10. We see that the degree-degree association index ($I^\lambda_{f^\Delta_{exp}}$) is a hub vertex, i.e., there is only a small correlation. That means $I^\lambda_{f^\Delta_{exp}}$ (by using C_{21}) captures structural information significantly different compared to almost all other measures (representing vertices) in this network. If we consider N_9 as a graph set, we observe that $G^2_{\leq 0.2}$ has more hubs than $G^1_{\leq 0.2}$. For instance, I_Δ and I_a represent hubs and therefore possess only a small correlation with other measures from different paradigms. This also implies that the structural characteristics of the graphs $\in N_9$ are different to those $\in C_{21}$. Also, the hubs in $G^2_{\leq 0.2}$ could serve as potential candidates to be tested for solving QSAR/QSPR problems [38] as they capture structural characteristics differently (compared to classical indices) and some (e.g., efficiency complexity and offdiagonal complexity) have not yet been used in

mathematical chemistry and drug design. In addition, it would be interesting to examine their ability for classifying graphs optimally by using supervised learning techniques, e.g., see [39].

To finalize this section, we consider Figures 11, 12, 13, 14. We have also plotted the evolution of the correlation networks for $\theta = 0.01, 0.05$, and have obtained the networks $G_{\leq 0.01}$ and $G_{\leq 0.05}$ for both N_9 and C_{21}, respectively. From Figure 11, we see that by using N_9, the measures C_e and E are highly uncorrelated ($\theta = 0.01$). In addition, the degree-degree association index $I^\lambda_{f^\Delta_{exp}}$ and GA_1 are highly uncorrelated by using C_{21} ($\theta = 0.01$). If we now choose $\theta = 0.05$ for N_9 and C_{21}, the resulting networks (see Figures 13 and 14) also show highly uncorrelated indices. Starting with N_9 (see Figure 13), far more indices are highly uncorrelated ($\theta = 0.05$) compared with Figure 11. These indices belong to different paradigms (degree-based, information-theoretic, etc.). But when considering the graph class C_{21} (see Figure 14), only the degree-degree association index $I^\lambda_{f^\Delta_{exp}}$ is highly uncorrelated ($\theta = 0.05$) with many other indices. It is clear that the differences between these correlation networks are clearly induced by the structural differences (factors such as cyclicity and connectedness, which contribute to the complexity of the graphs) of the graph classes. Note that we obtained a similar result by comparing N_9 and N_{10} (instead of N_9 and C_{21}). Figure 14 expresses that by using trees, $I^\lambda_{f^\Delta_{exp}}$ captures structural information significantly different than many other non-information-theoretic indices such as E, EE, etc. We hypothesize that this result also holds for other tree classes as well. As mentioned above, the index $I^\lambda_{f^\Delta_{exp}}$ could be used to characterize graphs for problems in structural chemistry or QSAR, with the aim that it solves a particular problem (e.g., QSAR/QSPR) better than existing indices which have already been used.

Summary and Conclusion

In this paper, we have explored to what extent degree and eigenvalue-based measures are degenerate. To tackle this problem, we used exhaustively generated undirected, connected and non-isomorphic graphs and chemical graphs. Interestingly, we found that some recently developed distance-based measures, e.g., $GA_{2,3}$, have a much better uniqueness than measures that are known to be highly unique for chemical graphs, e.g., the Balaban J index. Note that the results for the Balaban J index by using the classes N_i, $i = 8, 9, 10$, have been reported in an earlier paper [30]. Equally, some of the eigenvalue-based measures such as E, LE and LEE possess high discrimination power for all graph classes that we examined in this paper. This shows that such measures for discriminating graphs structurally can be feasible, despite the existence of isospectral graphs. A strong point of all measures (except the topological information content for large graphs, as it relies on determining their automorphism groups) used in this study is their polynomial time complexity. Hence, they could also be applied to large complex networks. First studies of examining the uniqueness of structural measures by using gene networks inferred from high-throughput data are under development. We will also examine the relationship between the uniqueness of a measure and the ability to classify graphs meaningfully.

Author Contributions

Analyzed the data: MG. Wrote the paper: MD BF MG.

References

1. Cvetković DM, Doob M, Sachs H (1980) Spectra of Graphs. Theory and Application. Deutscher Verlag der Wissenschaften. Berlin, Germany.
2. Dehmer M, editor (2011) Structural Analysis of Complex Networks. Birkhäuser Publishing.
3. Harary F (1969) Graph Theory. Addison Wesley Publishing Company. Reading, MA, USA.
4. Kuratowski K (1930) Sur le problème des courbes gauches en topologie. Fund Math Vol 15: 271–283.
5. Dehmer M, Emmert-Streib F, editors (2009) Analysis of Complex Networks: From Biology to Linguistics. Wiley VCH Publishing.
6. Mowshowitz A (1968) Entropy and the complexity of the graphs I: An index of the relative complexity of a graph. Bull Math Biophys 30: 175–204.
7. Newman MEJ, Barabási AL, Watts DJ (2006) The Structure and Dynamics of Networks. Princeton Studies in Complexity. Princeton University Press.
8. Bonchev D (1983) Information Theoretic Indices for Characterization of Chemical Structures. Research Studies Press, Chichester.
9. Dehmer M, Mowshowitz A (2011) A history of graph entropy measures. Inform Sciences 1: 57–78.
10. Todeschini R, Consonni V, Mannhold R (2002) Handbook of Molecular Descriptors. Wiley-VCH. Weinheim, Germany.
11. Diudea MV, Gutman I, Jäntschi L (2001) Molecular Topology. Nova Publishing. New York, NY, USA.
12. Emmert-Streib F, Dehmer M (2011) Networks for systems biology: Conceptual connection of data and function. IET Syst Biol 5: 185–207.
13. da F Costa L, Rodrigues F, Travieso G (2007) Characterization of complex networks: A survey of measurements. Adv Phys 56: 167–242.
14. Bonchev D, Mekenyan O, Trinajstić N (1981) Isomer discrimination by topological information approach. J Comp Chem 2: 127–148.
15. Konstantinova EV (1996) The discrimination ability of some topological and information distance indices for graphs of unbranched hexagonal systems. J Chem Inf Comput Sci 36: 54–57.
16. Konstantinova EV (2006) On some applications of information indices in chemical graph theory. In: Ahlswede R, Bäumer L, Cai N, Aydinian H, Blinovsky V, et al., editors, General Theory of Information Transfer and Combinatorics, Springer, Lecture Notes of Computer Science. 831–852.
17. Diudea MV, Ilić A, Varmuza K, Dehmer M (2011) Network analysis using a novel highly discriminating topological index. Complexity 16: 32–39.
18. Xu CYHL (1996) On highly discriminating molecular topological index. J Chem Inf Comput Sci 36: 82–90.
19. Balaban AT (1982) Highly discriminating distance-based topological index. Chem Phys Lett 89: 399–404.
20. Balaban AT (1987) Numerical modelling of chemical structures: Local graph invariants and topological indices. In: King RB, Rouvray DH, editors, Graph Theory and Topology, Elsevier. 159–176. Amsterdam, The Netherlands.
21. Vukičević D, Furtula B (2009) Topological index based on the ratios of geometrical and arithmetical means of end–vertex degrees of edges. J Math Chem 46: 1369–1376.
22. Fath-Tabar G, Furtula B, Gutman I (2010) A new geometric–arithmetic index. J Math Chem 47: 477–486.
23. Estrada E, Torres L, Rodríguez L, Gutman I (1998) An atom–bond connectivity index: Modelling the enthalpy of formation of alkanes. Indian J Chem 37A: 849–855.
24. Estrada E (2002) Characterization of the folding degree of proteins. Bioinformatics 18: 697–704.
25. Janežić D, Mileževič A, Nikolić S, Trinajstić N (2007) Graph-Theoretical Matrices in Chemistry. Mathematical Chemistry Monographs. University of Kragujevac and Faculty of Science Kragujevac.
26. Kim J, Wilhelm T (2008) What is a complex graph? Physica A 387: 2637–2652.
27. Balaban AT (2005) Can topological indices transmit information on properties but not on structures? J Comput Auid Mol Des 19: 651–660.
28. Balaban AT, Mills D, Kodali V, Basak SC (2006) Complexity of chemical graphs in terms of size, branching and cyclicity. SAR QSAR Environ Res 17: 429–450.
29. Dehmer M, Emmert-Streib F, Tsoy Y, Varmuza K (2011) Quantifying structural complexity of graphs: Information measures in mathematical chemistry. In: Putz M, editor, Quantum Frontiers of Atoms and Molecules, Nova Publishing. 479–498.
30. Dehmer M, Grabner M, Varmuza K (2012) Information indices with high discrimination power for arbitrary graphs. PLoS ONE 7: e31214.
31. Zhou B, Gutman I, Furtula B, Du Z (2009) On two types of geometric-arithmetic index. Chem Phys Lett 482: 153–155.
32. Das KC, Gutman I, Furtula B (2011) Survey on geometric–arithmetic indices of graphs. MATCH Commun Math Comput Chem 65: 595–644.
33. Miljković O, Furtula B, Radenković S, Gutman I (2009) Equienergetic and almost–equienergetic trees. MATCH Commun Math Comput Chem 61: 451–461.
34. Ivanciuc O, Ivanciuc T, Diudea MV (1999) Polynomials and spectra of molecular graphs. Roum Chem Quarterly Rev 7: 41–67.
35. Randić M, Vracko M, Novic M (2001) Eigenvalues as molecular descriptors. In: Diudea MV, editor, QSPR/QSAR Studies by Molecular Descriptors, Nova Publishing. 93–120. Huntington, NY, USA.
36. Dehmer M, Müller L, Graber A (2010) New polynomial-based molecular descriptors with low degeneracy. PLoS ONE 5: e11393.
37. Backhaus K, Erichson B, Plinke W, Weiber R (2003) Multivariate Analysemethoden. Springer. Berlin, Germany.
38. Devillers J, Balaban AT (1999) Topological Indices and Related Descriptors in QSAR and QSPR. Gordon and Breach Science Publishers. Amsterdam, The Netherlands.
39. Müller LAJ, Kugler KG, Graber A, Dehmer M (2011) A network-based approach to classify the three domains of life. Biology Direct 6: 140–141.
40. Furtula B, Graovac A, Vukičević D (2010) Augmented Zagreb index. J Math Chem 48: 370–380.
41. Miličević A, Nikolić S (2004) On variable Zagreb indices. Croat Chem Acta 77: 97–101.
42. Nikolić S, Kovačević G, Miličević A, Trinajstić N (2003) The Zagreb indices 30 years after. Croat Chem Acta 76: 113–124.
43. Narumi H, Katayama M (1984) Simple topological index. a newly devised index characterizing the topological nature of structural isomers of saturated hydrocarbons. Mem Fac Engin Hokkaido Univ 16: 209–214.
44. Bonchev D (1989) The concept for the center of a chemical structure and its applications. J Mol Struct: Theochem 185: 155–168.
45. Claussen JC (2007) Offdiagonal complexity: A computationally quick network complexity measure - Application to protein networks and cell division. In: Deutsch A, de la Parra RB, de Boer R, Diekmann O, Jagers P, et al., editors, Mathematical Modeling of Biological Systems, Volume II, Boston: Birkhäuser. 303–311.
46. Wilhelm T, Hollunder J (2007) Information theoretic description of networks. Physica A 388: 385–396.
47. Gutman I, Li X, Zhang J (2009) Graph energy. In: Dehmer M, Emmert-Streib F, editors, Analysis of Complex Networks: From Biology to Linguistics, Wiley-VCH. 145–174.
48. Gutman I, Zhou B (2006) Laplacian energy of a graph. Linear Algebra Appl 414: 29–37.
49. Fath-Tabar GH, Ashrafi AR, Gutman I (2009) Note on Estrada and L-Estrada indices of graphs, volume CXXXIX. 1–16 pp.
50. Raychaudhury C, Ray SK, Ghosh JJ, Roy AB, Basak SC (1984) Discrimination of isomeric structures using information theoretic topological indices. J Comput Chem 5: 581–588.