Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 May 2018 doi:10.20944/preprints201805.0302.v1

Peer-reviewed version available at *Entropy* **2018**, *20*, 481; doi:10.3390/e20070481

Article

Relating Vertex and Global Graph Entropy in Randomly Generated Graphs

Philip Tee [1,2†,‡], George Parisis [3‡], Luc Berthouze [3‡], and Ian Wakeman [3‡]*

1 Moogsoft Inc; phil@moogsoft.com
2 University of Sussex; p.tee@sussex.ac.uk
3 University of Sussex; {g.parisis,l.berthouze,ianw}@sussex.ac.uk
* Correspondence: p.tee@sussex.ac.uk;
† Current address: Moogsoft Inc, 1265 Battery Street, San Francisco CA 94111
‡ These authors contributed equally to this work.

Abstract: Combinatoric measures of entropy capture the complexity of a graph, but rely upon the calculation of its independent sets, or collections of non-adjacent vertices. This decomposition of the vertex set is a known NP-Complete problem and for most real world graphs is an inaccessible calculation. Recent work by Dehmer *et al* and Tee *et al*, identified a number of alternative vertex level measures of entropy that do not suffer from this pathological computational complexity. It can be demonstrated that they are still effective at quantifying graph complexity. It is intriguing to consider whether there is a fundamental link between local and global entropy measures. In this paper, we investigate the existence of correlation between vertex level and global measures of entropy, for a narrow subset of random graphs. We use the greedy algorithm approximation for calculating the chromatic information and therefore Körner entropy. We are able to demonstrate close correlation for this subset of graphs and outline how this may arise theoretically.

Keywords: Graph Entropy, Chromatic Classes, Random Graphs

---

1. Introduction and Background

1.1. Overview

Global measures of graph entropy, defined combinatorially, capture the complexity of a graph. In this context complexity is a measure of how distinct or different each vertex is in terms of its interconnection into the rest of the graph. In many practical applications of network science, which can range from fault localization in computer networks to cancer genomics, this difference in connectivity can indicate that certain vertices in a graph are in some way more important to the correct functioning of the network the graph represents. In this paper we explore the potential relationships between vertex level entropy measures and global graph entropy. This is an important problem for a very simple reason. Global graph entropy, as defined originally by Körner [1] is expensive to compute, as it relies upon the calculation of independent sets of the graph. This is a known $NP-Complete$ problem, and in most real world graphs the computation of graph entropy is prohibitive. As stated before though, the value of this graph metric though is high, as it fundamentally captures a measure of the complexity of the graph that has a range of practical applications.
If it were possible to approximate the value of graph entropy with a much more easily computable metric, it would make it possible to use entropy in these and potentially many other applications. The fundamental barrier to computability is the fact that the entropy calculation depends upon combinatoric constructs across the whole graph. If instead, a metric were available that was intrinsically local, that is computable for each vertex in the graph with reference to only the local topology of the graph, it would then be possible to efficiently calculate a value for each vertex and then simply sum these values across the graph to obtain an upper bound for the entropy of the graph. The fact

Peer-reviewed version available at *Entropy* **2018**, *20*, 481; doi:10.3390/e20070481

that it is an upper bound is possible to assert by simply appealing to the sub additivity property of graph entropy. That is, for any two graphs G_1 and G_2, the entropy of the union of the graphs obeys $H(G_1 \cup G_2) \leq H(G_1) + H(G_2)$.

Precisely such a metric has been advanced in work by Dehmer *et al* [2–6], and developed by many other authors including in recent work we published [7] exploring the utility of vertex entropies in the localization of faults on a computer network. The formalisms used by Dehmer*et al* and in our previous work differ in the construction of the entropies, and how graphs are partitioned into local sets. A primary motivation for this difference was motivated by the practical application of these measures described in [7]. The relation between the vertex entropy formalisms introduced by Dehmer, and global graph entropies have been analyzed [8] and the central focus of this paper is to explore how closely our definitions of vertex entropy approximate global entropies for two classes of Random Graphs, the traditional Gilbert graph [9], and the Scale-Free graphs first advanced by Barabási [10]. We restrict our experimental investigation to simple connected graphs, which in the case of Scale Free graphs arise naturally, for the Gilbert graphs we chose to focus on the Giant Component (GC) of the generated graph. For statistical significance this further restricts the choice of connection probability to a range above the critical threshold at which a GC emerges.

In this section we present an overview of both global and vertex graph entropy, before discussing the experimental analysis in Section 2. The data analysis produces a perhaps surprisingly close correlation between vertex and global entropy, which we seek to explain in Section 3, by comparing the expected values of Chromatic Information as a proxy for Graph Entropy and the expected values for vertex entropy when considering an ensemble of random graphs. The main result of this paper is that for random graphs the correlation is strong and explained by the co-dependence of both metrics on edge density. We conclude in Section 4 and point to further directions in this research. In particular, if the correlation we describe in this paper holds as a general result this opens up the use of vertex level measures to frame entropic arguments for many dynamical processes on graphs, including network evolution. Such models of network evolution have been advanced by a number of authors including Peterson *et al* [11] and ourselves [12].

Before presenting the experimental and theoretical analysis of the possible correlation between local node and global measures of entropy, in the rest of this section we will briefly survey the necessary concepts.

1.2. Global Graph Entropy

The concept of the entropy of a graph has been widely studied ever since it was first proposed by Janos Körner in his 1973 paper on Fredman-Komlós bound [1]. The original definition rested upon a graph reconstruction based upon an alphabet of symbols, not all of which are distinguishable. The construction begins by identifying with each member of an alphabet of n possible signals $X = \{x_1, x_2, \ldots x_n\}$, with a probability of emission $P_i, i \in 1, 2, \ldots, n$ in a given fixed time period. Using this basic construction the regular Shannon entropy [13] is defined in the familiar way:

$$H(X) = -\sum_{i=1}^{n} P_i \log_2 P_i \tag{1}$$

In [1], and beautifully explained in [14], János Körner introduced the concept of the entropy of a graph in terms of a modified version of Shannon's original argument. Considering the alphabet X, as defined above, imagine that not all of the signals are distinguishable. In the analysis going forward we adopt the normal notation of a graph $G(V, E)$ as the combination of a set of vertices V, and the set of edges E that exist between the vertices. Further all of our analysis is restricted to simple graphs containing no self-edges (that is edges that connect a vertex to itself). A graph can be constructed by mapping to the vertex set V each of the signals in the alphabet, so that $v_i \in V$ equates to x_i and naturally associated with each vertex is a probability of emission of a signal $P(v_i) = P_i$, which is a fixed property of each vertex. Now, each of the vertices are connected with an edge $e_{i,j} \in E$, if and only if the two signals x_i, x_j

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 May 2018 doi:10.20944/preprints201805.0302.v1

Peer-reviewed version available at Entropy 2018, 20, 481; doi:10.3390/e20070481

are distinguishable. The automorphism groups of this graph are naturally related to the information lost (and hence entropy gained), by certain signals not being distinguishable. To avoid the definition involving complex constructions using these automorphism groups, Mowshowitz *et al* [15] recast the definition in terms of the mutual information between the independent sets of the graph, where an independent set is equivalent to a chromatic class, that is a collection of vertices that are not adjacent. To establish this definition, let us imagine a process whereby we randomly select a vertex from the graph, according to a probability distribution $P(V)$ for each vertex, which as the process of selection is uniform will be identically $\frac{1}{n}$ for each vertex in a graph of size n. Each vertex will in turn be a member of an independent set $s_i \in S$ (S is chosen to represent the independent sets to avoid confusion with I the mutual information). The conditional probability $P(V|S)$ is the probability of selecting a vertex when the independent set that it belongs to is known. These probabilities capture important information concerning the structure of the graph. Associated with $P(V|S)$ is a measure of entropy $H(V|S)$, or the uncertainty in the first occurrence of selecting a vertex when the independent set is known. Using these quantities we define Structural Entropy as follows:

Definition 1. *The Structural Entropy of a Graph $G(V,E)$, over a probability distribution $P(V)$, $H(G,P)$, is defined as:*

$$H(G,P) = H(P) - H(V|S) , \tag{2}$$

where S is the set of independent sets of G, or equivalently the set of Chromatic Classes.

Closely related to this definition of entropy is Chromatic Information. This is defined in terms of the colorings of the graphs, that divide the graph into subsets of V where each vertex in V has the same color label. Each graph has an optimal minimum set of colorings which can be achieved, the number of such sets being referred to as the Chromatic Number of the graph χ. These subsets are called *Chromatic Classes* C_i, with the constraint that $\bigcup_i C_i = V$. Chromatic information is then naturally defined as:

Definition 2. *The Chromatic Information of a graph of n vertices is defined as:*

$$I_c(G) = \min_{\{C_i\}} \left[-\sum_i \frac{|C_i|}{n} \log_2 \left(\frac{|C_i|}{n} \right) \right] , \tag{3}$$

where the minimization is over all possible collections of chromatic classes, or colorings, of the graph C_i.

Crucially, the chromatic information is closely related to the second term in Equation (2), $H(V|S)$, and if we assume that the probability distribution P is uniform, we can relate the two quantities through the following identity.

$$H(G,P) = \log_2 n - I_c(G) \tag{4}$$

We will make use of this identity in Section 2 as chromatic information is much more readily calculable than entropy in standard network analysis packages. It is also common in this measure of information to drop the P in $H(G,P)$, as we are assuming the probability is uniform.

1.3. Local Entropy Measures

The computational challenges in calculating global entropy measures stem from the calculation of the independent sets of a graph, which is a known NP-Hard problem. Recent work by Dehmer *et al* [2,3], provided a framework for the definition of a form of graph entropy defined at the purely local level of a node. In essence, Dehmer introduces the concept of a j-Sphere, S_i^j, centered at the i^{th} node. Dehmer's original definition relied upon subsets of vertices of a fixed distance from a given vertex v_i.

where distance $d(v_i, v_j)$ is the shortest distance between distinct vertices v_i and v_j (i.e. $i \neq j$). For a node $v_i \in V$, we define, for $j \geq 1$, the 'j-Sphere' centered on v_i as:

$$S_i^j = \{v_k \in V | d(v_i, v_k) = j\} \tag{5}$$

On these j-Spheres, Dehmer defined certain probability-like measures, using metrics calculable on the nodes such as degree as a fraction of total degree of all nodes in the j-Sphere, from which entropies can be defined. This locality avoids the computationally challenging issues present in the global forms of entropy.

In recent work [7], the authors extended this definition to introduce some specific local measures for Vertex Entropy, that is the graph entropy of an individual node in the graph. The analysis that was followed was based upon the concept of locality introduced in Dehmer *et al*, using the concept of a j-Sphere. In this work we will expand upon that analysis, and, instead consider the vertices of a graph as part of an ensemble of vertices, and the graph itself in turn as part of an ensemble of Graphs.

Returning to the fundamental definition of entropy, it is a measure of how incompletely constrained a system is microscopically, when certain macroscopic properties of the system are known. For example, if we have an ensemble of all possible simple, connected graphs of order N, $\mathcal{G}(N)$, (that is $|V| = N$), we potentially have a very large collection of graphs. Further, we could go on to prescribe a further property such as average node degree $\langle k \rangle$ for the whole ensemble, and ask what is the probability of randomly selecting a member of the ensemble $G_i(N) \in \mathcal{G}(N)$, that shares a given value of this property $\langle k \rangle$, and denote that as $P(G_i)$. Following the analysis in Newman *et al* [16] we can then define the Gibbs entropy of the ensemble as:

$$S_{\mathcal{G}(N)} = - \sum_{G_i \in \mathcal{G}(N)} P(G_i) \log_2 P(G_i), \tag{6}$$

which is maximized subject to the constraint, $\sum_{G_i \in \mathcal{G}(N)} P(G_i) \langle k_i \rangle = \langle k \rangle$. This analysis allows us to go from the observed value $\langle k \rangle$ to the form of degree distribution and then on to other properties for the whole ensemble. In essence in our work, we take a different approach in two regards. Firstly, we restrict ourselves to the vertex level, where we consider that the vertices of an individual graph are themselves a randomized entity, which can be assembled in many ways to form the end graph. For example, if we were to decompose a given graph into the degree sequence of the nodes, there will be many nodes of the same degree, which does not completely prescribe *which* node in the graph we are considering. In that way a measure of uncertainty and therefore entropy naturally arises.

The second difference to the approach taken by Newman *et al*, is that rather than work from a measurable constraint and maximized form of entropy back to the vertex probability, we ask what probabilities we can prescribe on a vertex. In the interests of computational efficiency we construct a purely local theory of the graph structure constrained to those vertex properties that are measurable in the immediate (that is $j = 1$) neighborhood of the vertex. Out of the possible choices we have selected node degree, node degree as a fraction of the total edges in the network, and local clustering coefficient. In our work, we also redefined the local clustering coefficient C_1^i of a node i as the fraction of existing edges amongst nodes in the $j = 1$ neighborhood to possible edges in that neighborhood. This is different to the traditional measure in that it includes the edges between the node v_i and its neighbors, a choice that was made to avoid a zero clustering coefficient for nodes that are the 'center' of a star-like network. Following the analysis in our prior work [7], we summarize the considered probabilities below.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 May 2018 doi:10.20944/preprints201805.0302.v1

Peer-reviewed version available at Entropy 2018, 20, 481; doi:10.3390/e20070481

- **Inverse Degree** In this case we denote the vertex probability as:

$$P(v_i) = \frac{Z}{k_i^{\alpha}}\text{ , with} \tag{7}$$

$$Z^{-1} = \sum_j k_j^{-\alpha} \text{ to ensure normalization.} \tag{8}$$

  This type of vertex probability mirrors the attachment probability of the scale free model and leads to a power law of node degree. In the standard scale free model $\alpha = 3$, but for the purposes of our experimentation, and for simplicity, we set $\alpha = 1$. In essence very large hubs are less probable, which intuitively captures the notion that they carry more of the global structure, and therefore information, of the graph. Graphs comprised of nodes with similar degrees will maximize entropy using this measure, reflecting the fact less information is carried by knowledge of the node degree.
- **Fractional Degree** We use in this case the following for vertex probability:

$$P(v_i) = \frac{k_i}{2|E|}. \tag{9}$$

  This probability measure captures the likelihood that a given edge in the network terminates or originates at the vertex v_i. Nodes with a high value of this probability will be more highly connected in the graph, and graphs which have nodes with identical values of the probability will have a higher entropy. This reflects the fact that the more similar nodes are the less information is known about the configuration of a given node by simply knowing its fractional degree.
- **Clustering Coefficient** The clustering coefficient measures the probability of an edge existing between the neighbors of a particular vertex. However, its use in the context of a vertex entropy needs to be adjusted by a normalization constant $Z = \sum_i C_1^i$ to be a well behaved probability measure and sum to unity. For simplicity we omit this constant and we assert:

$$P(v_i) = C_1^i \tag{10}$$

  The local clustering coefficient captures the probability that any two neighbors of the node v_i are connected. The larger this probability the more the local one hop subgraph centered at v_i is to the perfect graph. Again graphs comprised of nodes with similar clustering coefficient will maximize this entropy, reflecting the fact that the graph is less constrained by knowledge of the nodes clustering coefficients.

In essence for a given graph $G(N) \in \mathcal{G}(N)$, we specify a measured quantity for vertex v_i , as $x(v_i) = x_i$, and ask what is the probability of a random vertex v_i having this value. We denote this probability as $P(v_i)_{x(v_i)=x_i}$. This allows us to define entropy at the vertex level , and for the whole graph as:

$$S(v_i) = -P(v_i)\log_2 P(v_i) \tag{11}$$

$$S(G) = -\sum_{v_i \in G} P(v_i)\log_2 P(v_i) \tag{12}$$

In our analysis we compute the values of each of these variants of vertex entropy, summed across the whole graph as described in Equation 12. Because of the local nature of the probability measures the vertex entropy values are far quicker to compute than any of the global variants, and do not involve any known $NP - Complete$ calculations. In Section 2 we describe the approach taken to perform this analysis. The results indicate that the vertex entropy approximation yields a value that is closely correlated with the global values of chromatic information (and therefore entropy).
In Section 3 we will attempt to investigate how such a correlation could arise.

Peer-reviewed version available at Entropy 2018, 20, 481; doi:10.3390/e20070481

2. Experimental Analysis

2.1. Method and Objectives

We seek to establish whether there is any significant relationship between the global graph entropy of a graph, and the entropy obtained by summing the local node values of entropy. Because of the computational limits involved in calculating global values of entropy, we are restricted to graphs of moderate size, which rules out analyzing repositories of real world graphs such as the Stanford Large Network Dataset [17], or the Index of Complex Networks (ICON) [18]. A more tractable source of graphs are randomly generated graphs, where we can control the scale.
Random graphs are well understood to replicate many of the features of real networks, including the 'small world' property, clustering and degree distributions. We consider in our analysis two classes of randomly generated graphs, the Gilbert Random Graph $G(n, p)$ [9], and the scale-free graphs generated with a preferential attachment model [10]. For both types of graphs we simulated a large number of graphs with varying parameters to generate many possible examples of graphs that share a fixed number of vertices, with the choice of Gilbert graphs permitting varying edge counts and densities. For the purposes of our analysis we fixed the vertex count at $n = |V| = 300$ and in each case we included only fully connected, simple graphs. For the Gilbert graphs, this entailed analyzing only the Giant Component (GC) to ensure a fully connected graph.
We will discuss the results for each type of graph in more detail below, but the data revealed a close correlation between the chromatic information of the graph (and therefore structural entropy), and the value obtained by summing the local vertex entropies. We considered three variants of the vertex entropy and also included the edge density of the graph, defined as:

$$C(G) = \frac{2|E|}{n(n-1)}. \tag{13}$$

This measures the probability of an edge existing between any two randomly selected vertices.
To quantify the nature of the relationship between each of the aforementioned local entropies (including edge density) and chromatic information of the graphs, we adopted a model selection approach by performing polynomial regression using polynomials of increasing order up to 5, referred to as $H_{1,\ldots,5}$ henceforth, using the technique of least squares to numerically calculate the polynomial coefficients. The best model (within the family of models considered) was assessed based on the Bayesian Information Criteria (BIC), and Akaike Information Criteria (AIC) [19].
To calculate the measures we make the assumption that the distribution of the errors are identically and independently distributed, and reduced the likelihood function to the simpler expressions:

$$BIC = n \log_e(\hat{\sigma}_r^2) + k \log_e n \tag{14}$$

$$AIC = n \log_e(\hat{\sigma}_r^2) + 2k \tag{15}$$

$$\hat{\sigma}_r^2 = \frac{1}{n}\sum_{i=1}^{i=n}(\hat{y}_i - y_i)^2, \tag{16}$$

$$\tag{17}$$

where $\hat{y}_i$ is the prediction by model H, $\hat{\sigma}_r^2$ is the residual sum of squares, k the number of parameters in the model (in this case for H_j, $k = j + 1$), and n the number of data points.
The graph generation, analysis and model selection was all performed with a mixture of Java and MATLAB code, which is available upon request from the authors.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 May 2018 doi:10.20944/preprints201805.0302.v1

Peer-reviewed version available at Entropy 2018, 20, 481; doi:10.3390/e20070481

2.2. Scale-Free Graphs

The data for the scale-free graphs is displayed in Figure 1. We have segregated the plots by the calculated chromatic number for the graphs, and overlaid the optimal least squares fit model. It is suggestive that for graphs that share the same value of χ, there is a non-trivial relationship between the metrics. To gain insights into the nature of this relationship, the data was fitted using a least squares approach to polynomials up to the 5^{th} degree. In Tables 1, 2, 3, 4 we applied both the Bayesian Information Criteria and Akaike Information Criteria to identify the best model. We have highlighted the row for the model with the strongest performance in the BIC test, and the $\Delta_{BIC}/\Delta_{AIC}$ is measured from the H_1 model which performs worse in both analyses. In each case there is strong support for the existence of a relationship between the vertex entropy measures and chromatic information. Both AIC and BIC have a marginal preference for higher order polynomial models, with rejection of higher order indicating a preference for H_2 for Inverse Degree and Edge Density. The other measures require higher order fitting, over H_4, being necessary for Fractional Degree and H_3 for Cluster Entropy.

Model	Bayes Information Criteria	Δ_{BIC}	Akaike Information Criteria	Δ_{AIC}
H_1	-734.33	0.00	-740.33	0.00
H_2	**-830.12**	**-95.80**	**-839.14**	**-98.80**
H_3	-825.16	-90.84	-837.18	-96.85
H_4	-821.95	-87.62	-836.97	-96.63
H_5	-818.63	-84.31	-836.66	-96.32

Table 1. Model Selection Analysis for Inverse Degree Entropy for Scale-Free Graphs of constant $|V|$

Model	Bayes Information Criteria	Δ_{BIC}	Akaike Information Criteria	Δ_{AIC}
H_1	-708.84	0.00	-714.84	0.00
H_2	-715.24	-6.41	-724.25	-9.41
H_3	-719.49	-10.66	-731.51	-16.66
H_4	**-719.62**	**-10.79**	**-734.64**	**-19.80**
H_5	-715.31	-6.47	-733.33	-18.49

Table 2. Model Selection Analysis for Fractional Degree Entropy for Scale-Free Graphs of constant $|V|$

Model	Bayes Information Criteria	Δ_{BIC}	Akaike Information Criteria	Δ_{AIC}
H_1	-728.34	0.00	-734.35	0.00
H_2	-796.47	-68.12	-805.48	-71.13
H_3	**-798.84**	**-70.50**	**-810.86**	**-76.51**
H_4	-794.89	-66.54	-809.90	-75.55
H_5	-793.77	-65.43	-811.80	-77.44

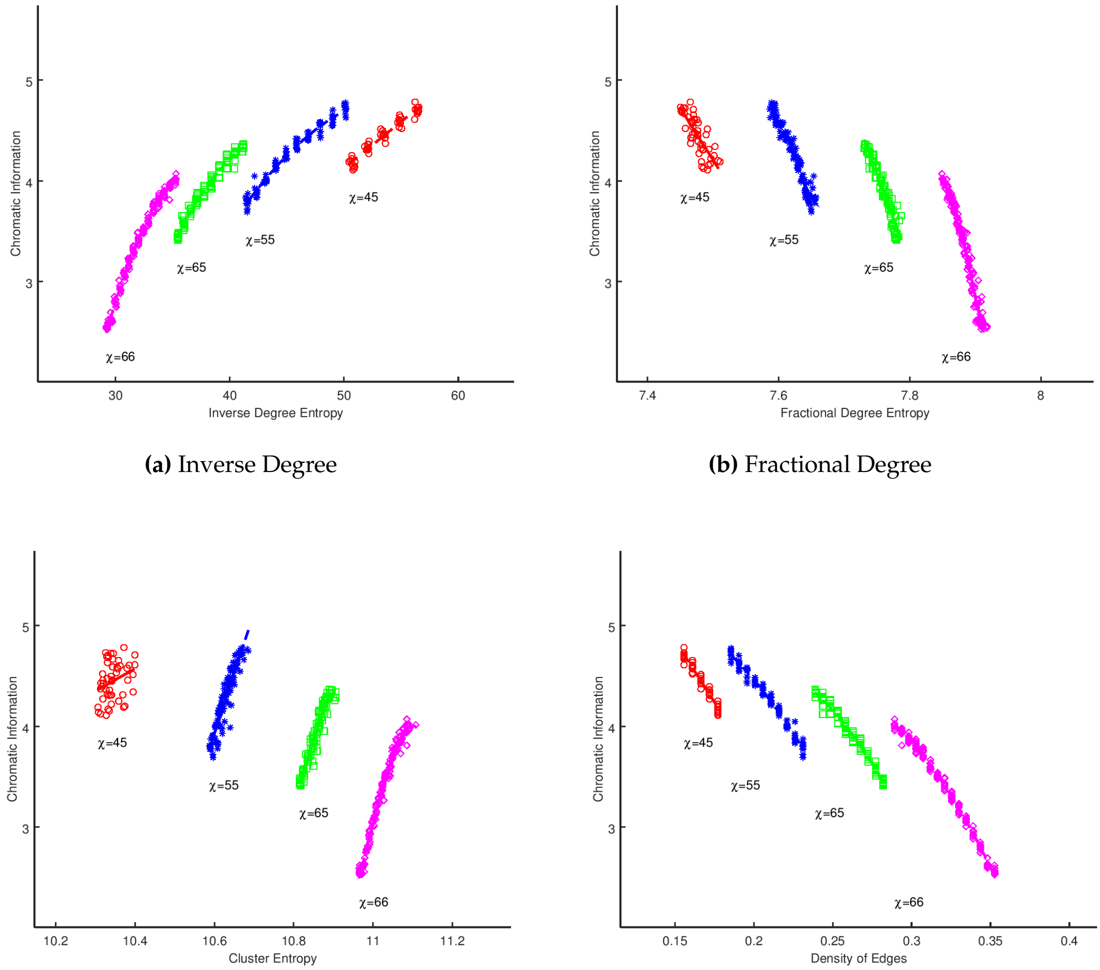
Table 3. Model Selection Analysis for Cluster Entropy for Scale-Free Graphs of constant $|V|$

Model	Bayes Information Criteria	Δ_{BIC}	Akaike Information Criteria	Δ_{AIC}
H_1	-777.77	0.00	-783.78	0.00
H_2	**-844.94**	**-67.17**	**-853.96**	**-70.18**
H_3	-842.39	-64.62	-854.40	-70.63
H_4	-839.21	-61.43	-854.23	-70.45
H_5	-836.87	-59.10	-854.89	-71.11

Table 4. Model Selection Analysis for Density of Edges for Scale-Free Graphs of constant $|V|$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 May 2018 doi:10.20944/preprints201805.0302.v1

Peer-reviewed version available at *Entropy* **2018**, *20*, 481; doi:10.3390/e20070481



(a) Inverse Degree

(b) Fractional Degree

(c) Clustering Coefficient Entropy

(d) Edge Density

Figure 1. Sum of vertex entropies for whole graph vs. chromatic information for Barabási-Albert scale-free graphs of constant $|V|$.

2.3. *Gilbert Random Graphs* $G(n, p)$

In addition to scale-free graphs we analyzed random graphs. The results are displayed in Figure 2, overlaid with the least squares optimized best fit. On visual inspection it appears that there is a systematic and non-trivial relationship between the metrics. In order to gain insights into this relationship, the data was again fitted using a least squares approach to polynomials up to the 5^{th} degree. In Tables 5, 6, 7, 8 we applied both the Bayesian Information Criteria and Akaike Information Criteria to select the best model (among the models considered). As in the case of the scale-free graphs, we have highlighted the row in bold corresponding to the best model from a BIC perspective, and $\Delta_{BIC}/\Delta_{AIC}$, is measured against the worst performing model H_1. For Gilbert graphs, AIC and BIC both support the hypothesis of a relationship between the metrics. In the case of all but Inverse Degree, it would appear that H_2 is an optimal choice of model. Inverse Degree, however would appear to be best fitted by H_3, in contrast to the behavior of the scale-free graphs.

Peer-reviewed version available at Entropy 2018, 20, 481; doi:10.3390/e20070481

Model	Bayes Information Criteria	Δ_{BIC}	Akaike Information Criteria	Δ_{AIC}
H_1	-2004.92	0.00	-2008.75	0.00
H_2	-2181.68	-176.76	-2189.33	-180.59
$\mathbf{H_3}$	**-2182.62**	**-177.70**	**-2194.10**	**-185.35**
H_4	-2176.82	-171.90	-2192.12	-183.37
H_5	-2171.14	-166.22	-2190.27	-181.52

Table 5. Model Selection Analysis for Inverse Degree Entropy for Random Graphs of constant $|V|$

Model	Bayes Information Criteria	Δ_{BIC}	Akaike Information Criteria	Δ_{AIC}
H_1	-1806.14	0.00	-1809.96	0.00
$\mathbf{H_2}$	**-1874.29**	**-68.15**	**-1881.94**	**-71.98**
H_3	-1868.70	-62.56	-1880.17	-70.21
H_4	-1859.34	-53.20	-1874.64	-64.68
H_5	-1856.25	-50.11	-1875.38	-65.42

Table 6. Model Selection Analysis for Fractional Degree Entropy for Random Graphs of constant $|V|$

Model	Bayes Information Criteria	Δ_{BIC}	Akaike Information Criteria	Δ_{AIC}
H_1	-2109.61	0.00	-2113.44	0.00
$\mathbf{H_2}$	**-2146.19**	**-36.58**	**-2153.84**	**-40.40**
H_3	-2140.43	-30.82	-2151.91	-38.47
H_4	-2134.61	-25.00	-2149.92	-36.48
H_5	-2128.86	-19.25	-2147.99	-34.56

Table 7. Model Selection Analysis for Cluster Entropy for Random Graphs of constant $|V|$

Model	Bayes Information Criteria	Δ_{BIC}	Akaike Information Criteria	Δ_{AIC}
H_1	-951.67	0.00	-955.49	0.00
$\mathbf{H_2}$	**-985.93**	**-34.26**	**-993.58**	**-38.08**
H_3	-980.15	-28.48	-991.63	-36.13
H_4	-974.37	-22.71	-989.68	-34.18
H_5	-969.80	-18.13	-988.93	-33.43

Table 8. Model Selection Analysis for Edge Density for Random Graphs of constant $|V|$

3. Theoretical Discussion of the Results

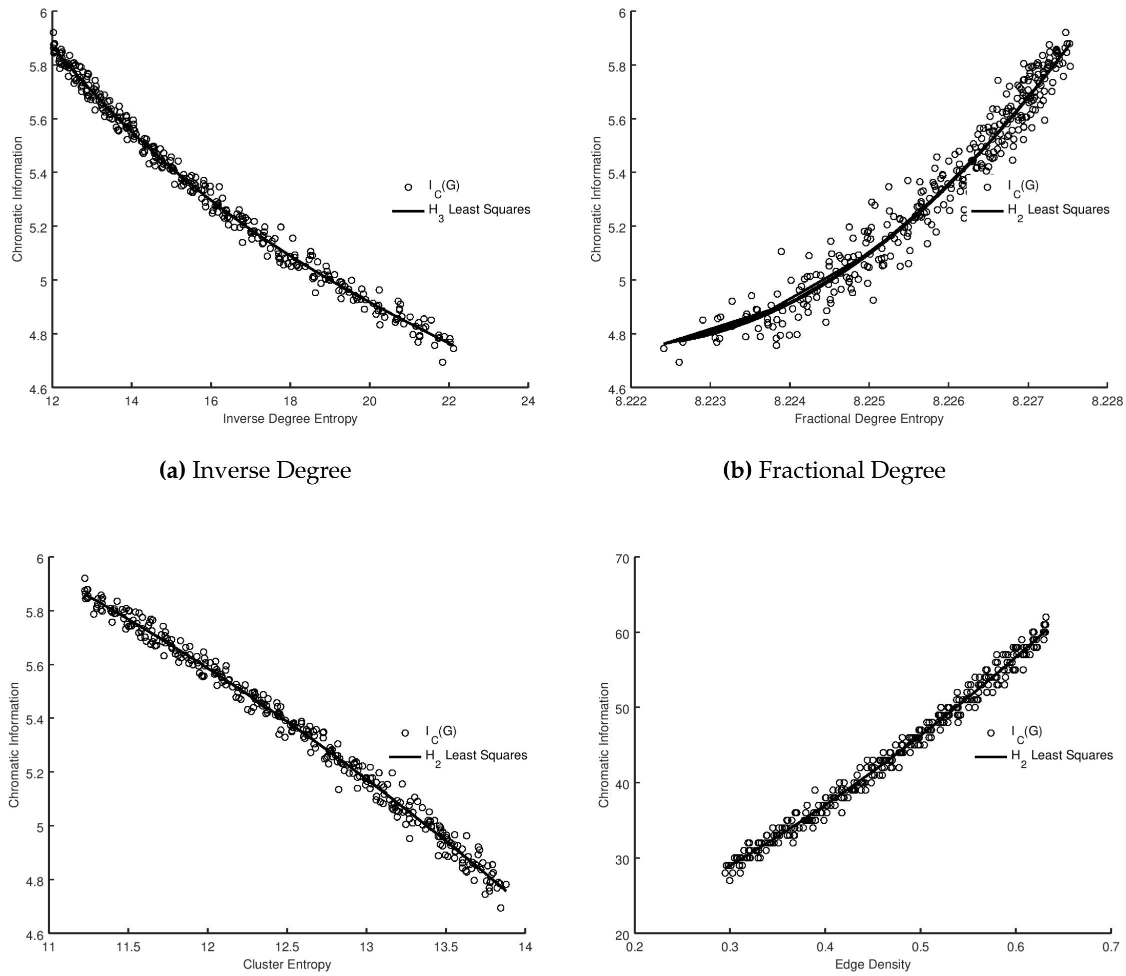
The strong correlation between chromatic information and the various forms of vertex entropy derived graph entropies may at first seem paradoxical. The first quantity is combinatorial in nature, and depends upon the precise arrangement of edges and vertices to produce the optimal coloring which dictates its value. The summed vertex entropies, at least to first order, depend solely upon the individual node degrees and take little or no account of the global arrangement of the graph.

It is certainly beyond the scope of this work, and indeed to the opinion of the authors, intractable to calculate a precise relationship between the two quantities. It is possible, however, to construct an argument as to *why* the two quantities might be in such close correlation.

To approach an explanation, it must first be noted that the experimental data is generated by sampling a number of randomly generated graphs of varying size. The only relationship between the graphs is the manner of their construction, and, crucially the resultant degree distributions of the graphs. Let us

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 May 2018 doi:10.20944/preprints201805.0302.v1

Peer-reviewed version available at *Entropy* **2018**, *20*, 481; doi:10.3390/e20070481



(a) Inverse Degree

(b) Fractional Degree

(c) Clustering Coefficient Entropy

(d) Edge Density

Figure 2. Sum of vertex entropies for whole graph vs. chromatic information for Gilbert graphs $G(n, p)$ for $p \in [0.31, 0.7]$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 May 2018 doi:10.20944/preprints201805.0302.v1
Peer-reviewed version available at Entropy 2018, 20, 481; doi:10.3390/e20070481

consider an ensemble of graphs $\mathscr{G}(G(V,E))$, with degree distributions $P(k)$, and fixed order $n = |V|$. For this ensemble, a random member G of order n, of chromatic number $\chi(G)$ will have an average chromatic information as follows:

$$\langle I_C(G)\rangle \approx -\chi(G) \times \frac{\langle|C_\alpha|\rangle}{n}\log_2\frac{\langle|C_\alpha|\rangle}{n} \tag{18}$$

where $\langle|C_\alpha|\rangle$, is the average size of the chromatic class.
Using the definitions of vertex entropy described in Section 1, we can also similarly compute an average value of each vertex entropy for a member of the ensemble $\mathscr{G}$, where we have taken the continuum approximation in the integral on the r.h.s:

$$\langle S_{vertex}\rangle = n \times \langle S(v_i)\rangle = n\int_1^\infty P(k)S(k)dk \tag{19}$$

In the case of the scale free graphs, this yields analytically soluble integrals for inverse degree, fractional degree vertex entropies, and edge densities but for the clustering coefficient entropies, and for all $G(n,p)$ random graphs the integrals are not solvable directly. To simplify the analysis we use the approximate continuum result for the scale free degree distributions at $t \to \infty$, $P(k) = \frac{2m^2}{k^3}$, where m is the number of nodes a new nodes connects to during attachment. For the clustering coefficient we can make a very rough approximation in the case of scale free networks of $4m/n$ by arguing that for an average node degree of $\langle k\rangle = 2m$, each of the neighbors shares the average degree and has a probability if $2m/n$ of connecting to another neighbor of the vertex. The quantity is not exactly calculable, but in [10] a closer approximation gives $C_i^1 \propto n^{-3/4}$, though for simplicity we will use our rough approximation. Where an exact solution is not available we can roughly approximate the value of $\langle S(v_i)\rangle$, by replacing the exact degree of the node by the average degree and then asserting:

$$\langle S_{vertex}\rangle = n \times S(\langle k\rangle) \tag{20}$$

We summarize these expressions in Table 9

Vertex Entropy Measure	Scale Free Graphs	Random Graphs $G(n,p)$
Inverse Degree	$2nm^2/9\ln 2$	$p^{-1}\log_2(pn)$
Fractional Degree	$m\log_2(2mn)$	$\log_2 n$
Clustering Coefficient	$4m\log_2(n/4m)$	$-np\log_2 p$

Table 9. Average Entropies across Random Graphs

In section 2 we presented the analysis of samples of randomly generated graphs created using three schemes. Each of these showed a surprisingly strong correlation between the vertex entropy measures, summed across the whole graph, and the chromatic information obtained using the greedy algorithm. The greedy algorithm is well known to obtain a coloring of an arbitrary graph which is close, but not optimal. Indeed the chromatic number of the graph obtained from the greedy algorithm $\chi_g(G_)$ is an upper bound of the true chromatic number $\chi(G)$. For a full description see [9,20].

3.1. Gilbert Random Graphs

Let us first consider the case of the Gilbert random graphs. We follow the same treatment and notation as in [9] and [21]. We construct the graph starting with n vertices, and each of the $\frac{1}{2}n(n-1)$ possible links are connected with a probability p. The two parameters n and p completely describe the parameters of the generated graph graph, and we denote this family of graphs as $G(n,p)$. It is well known that equation (18) is maximized when each of the chromatic classes of the graph C_α are uniform. That is, if the cardinality fo a chromatic class is denoted by $|C_\alpha|$, and χ is the chromatic number of the graph, we have:

$$|C_\alpha| = \frac{n}{\chi}, \forall\, C_\alpha \tag{21}$$

This chromatic decomposition is only obtained from the perfect graph on n vertices, K_n, and proof of this upper bound is outlined in [7]. For a given random graph $G(n,p)$, we denote the coloring obtained in this way, the homogenized coloring $\bar{C}_\alpha$ of $G(n,p)$, and we assert that $I_C(G) \leq \bar{I}_C(G)$. It is straight forward to verify that this yields as an expression for the chromatic information the following:

$$I_C(G) \leq \bar{I}_C(G) = \log \chi \tag{22}$$

To build upon the analysis, we consider a randomly selected chromatic class C_α, which has $c = \frac{n}{\chi}$ nodes. We denote the probability that c randomly selected nodes do not posses a link between them as $\bar{P}(C_\alpha, \chi)$, and the probability that *at least* one link exists between these nodes as $P(C_\alpha, \chi)$. We consider a large ensemble of random graphs, generated with the same Gilbert graph parameters n, p, which we write as $\mathscr{G}(n,p)$. For a randomly chosen member of this ensemble, the criteria for the graph $G(n,p) \in \mathscr{G}(n,p)$ to possess a chromatic number χ, is simply that it is more likely for c randomly selected nodes in $G(n,p)$ to be disconnected. That is:

$$\bar{P}(C_\alpha, \chi) \geq P(C_\alpha, \chi) \tag{23}$$

To estimate the first term, we note that for c randomly chosen nodes to have no connecting edges is given by:

$$\bar{P}(C_\alpha, \chi) = (1-p)^{\frac{1}{2}c(\chi)(c(\chi)-1)} \tag{24}$$

For brevity of notation we write $c(\chi)$ as c, and ask the reader to remember that c is a function of the chromatic number of the graph. We can also estimate the second term, by factoring the probability of a link from any of the nodes $v_i \in C_\alpha$, connecting to another node in C_α. The total probability is accordingly the product of:

$$P(C_\alpha, \chi) = P(\text{of any link}) \times P(\text{ link connects two nodes } v_i, v_j \in C_\alpha) \times c$$

Feeding in the standard parameters from the Gilbert graphs, we obtain:

$$\begin{aligned} P(C_\alpha, \chi) &= p \times \frac{c(c-1)}{n(n-1)} \times c \\ &= \frac{pc^2(c-1)}{n(n-1)} \\ \text{for } n \gg c, &\approx \frac{pn}{\chi^3} \end{aligned}$$

If we substitute back in the dependence of c upon the chromatic number, we obtain the following inequality, as the criteria for a given chromatic number χ to support an effective random coloring of the graph, we obtain the following inequality that can be used to estimate χ:

$$(1-p)^{\frac{1}{2}c(c-1)} \geq \frac{pc^2(c-1)}{n(n-1)} \tag{25}$$

It is easy to verify that equality is only ever reached when $p \to 1$, which yields the perfect graph in which $c = 1$, as every node is adjacent to every other node. The left hand side of Equation(25) is an increasing function of χ, whereas the right hand side is a decreasing function, so as χ increases we arrive at a minimum value such that the inequality is satisfied. We take this to be our estimate of the value of χ. The obtained value is not a strict upper bound, as we see in Section 2, but it is a reasonable

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 May 2018 doi:10.20944/preprints201805.0302.v1

Peer-reviewed version available at Entropy 2018, 20, 481; doi:10.3390/e20070481

estimate of the value. Although Equation (25) as a transcendental equation is not directly soluble, we can numerically solve to determine the value of χ for a fixed link probability p at which equality is reached. We present the results of the analysis in Figure 3, together with the optimal least squares fit of the relationship. We can also evaluate the best model for the relationship between our estimate and the measured values using BIC and AIC in the same manner as with the metrics. We present the results in Table 10. It is evident that a cubic relationship offers the best choice of model, and that fit is overlaid on the experimental data in Figure 3.

Model	Bayes Information Criteria	Δ_{BIC}	Akaike Information Criteria	Δ_{AIC}
H_1	-90.00	0.00	-92.25	0.00
H_2	-138.79	-48.79	-143.28	-51.04
$\mathbf{H_3}$	**-146.53**	**-56.53**	**-153.28**	**-61.03**
H_4	-142.72	-52.72	-151.71	-59.47
H_5	-138.60	-48.60	-149.84	-57.59

Table 10. Model selection analysis for computed χ versus measured for $|V| = 300$.

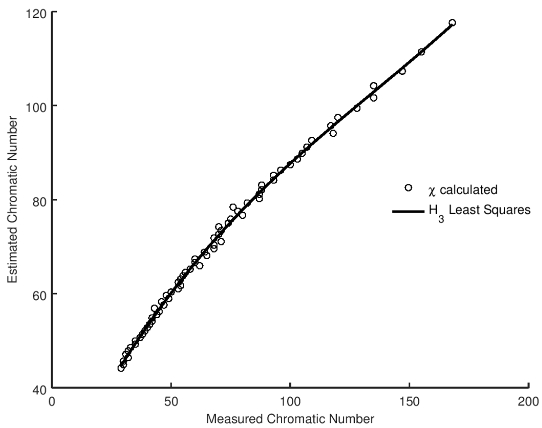


Figure 3. Calculated χ versus measured χ, for Gilbert graphs $G(n, p)$ with $n = 300$ and $p \in [0.3, 1.0,]$. Overlaid is the Least Squares Fit for H_3.

Having established that we can use Equation (25) to generate a good estimate of the chromatic number of a Gilbert graph, we can now attempt to explain how the chromatic information obtained from the greedy algorithm correlates with the vertex entropy measures. We begin by simplifying Equation (25) to determine the minimum value of χ at which equality is reached, by assuming the limit of $c \ll n$, and $c, n \gg 1$ to obtain:

$$(1-p)^{\frac{n^2}{2\chi^2}} = \frac{pn}{\chi^3}$$

Taking the logarithm of both sides of this equation and manipulating we arrive at the following expression:

$$3 \log \chi = \log pn - \frac{n^2}{2\chi^2} \log(1-p) \tag{26}$$

$$\bar{I}_C G = \frac{1}{3} \log pn - \frac{n^2}{6\chi^2} \log(1-p) \tag{27}$$

Equation 27 represents an approximation for the chromatic information of a random graph. Numerical experimentation indicates that the first term dominates for small values of p, and as $p \to 1.0$ the second terms becomes numerically larger. Inspection of Table 9, shows that Equation 27 contains terms that reflect a number of the expressions for the vertex entropy quantities we have considered. Indeed as the experiments were conducted at a fixed value of $n = 300$, for small p, by elementary manipulation one can see that $\bar{I}_C G \propto pS_{VE}$, or alternatively $\bar{I}_C G \propto S_{CE}/p$. Although this is a far from rigorous analytical derivation of the dependence of the chromatic information on the vertex entropy terms, the analysis does perhaps go some way to making the close correlation experimentally make sense in the context of this theoretical analysis. Deriving an exact relationship between the two quantities is beyond the scope of this work.

3.2. Scale free graphs

In the case of scale free graphs we can follow a similar analysis to the pure random case. To derive the probability of a link, for a randomly selected node the average probability of attachment $p = \langle k \rangle / 2mt$ by appeal to the original preferential attachment model. The dependence upon the connection valence m drops out and the average probability of a link existing between two nodes becomes $p = 1/n$. Following an identical argument to the random graph case we arrive at the following relationship:

$$\bar{I}_C G = \frac{n^2}{4\chi^2} \log_2 \left(\frac{n}{n-1} \right) \tag{28}$$

Again, it is important to stress that this is not a rigorous derivation of a relationship between chromatic information and vertex entropy, but it is possible to explain some of the structure in the results presented in Figure 1. For a given fixed value of χ the experiments represent graphs produced with increasing edge densities. As edges are added to the graph that *do not* increase the chromatic number, the size of the chromatic classes will evidently equalize (a discussion of this point can be found in [7]). This will have the effect if increasing the chromatic information, until a point is reached where $\chi \to \chi + 1$. At this point the denominator of Equation 28 will increase causing a drop in chromatic information. So, as each vertex entropy measure is fundamentally dependent upon the number of edges in a fixed sized graph, we would expect to see a series of correlations for each value of chromatic information, which is indeed what is demonstrated in Figure 1.

4. Conclusion

In this work we have principally been interested in investigating what, if any, correlation exists between purely local measures of graph entropy and global ones. It is not possible to make a general statement that for any graph this correlation exists, but for the two classes of random graphs considered it is persuasive that a relationship exists.

This is an interesting and important result. Interest in the informational content of graphs has a wide range of application, both in terms of network dynamics as a model of network evolution, but also outside of the field of network science it is under investigation in many different fields. For example, quantum gravity fundamentally relies upon spacetime becoming graph like at the so called Planck Length [22], and with entropy becoming posited as a potential origin of Gravity [23] the entropy of the spacetime graph is of interest. It would be convenient if large scale information content was largely driven by the local graph structure of a typical node if a tractable entropic theory of quantum gravity is to become possible. If the entropy of the spacetime mesh were not definable locally, any theory obtained would suffer from pathological lack of locality, a key feature of most modern field theories. Although far from settled, this paper does at least illustrate that for certain types of graph, the local environment of a typical node may indeed dominate the information content of the graph. In future work we intend to investigate other implications of vertex entropy on the dynamical processes possible on a graph.

1. Körner, J. Fredman–Komlós bounds and information theory, 1986.
2. Dehmer, M. Information processing in complex networks: Graph entropy and information functionals. *Applied Mathematics and Computation* **2008**, *201*, 82–94.
3. Dehmer, M.; Mowshowitz, A. A history of graph entropy measures. *Information Sciences* **2011**, *181*, 57–78.
4. Mowshowitz, A.; Dehmer, M. Entropy and the complexity of graphs revisited. *Entropy* **2012**, *14*, 559–570.
5. Cao, S.; Dehmer, M.; Shi, Y. Extremality of degree-based graph entropies. *Information Sciences* **2014**, *278*, 22–33.
6. Emmert-Streib, F.; Dehmer, M.; Shi, Y. Fifty years of graph matching, network alignment and network comparison. *Information Sciences* **2016**, *346-347*, 180–197, [1410.7585].
7. Tee, P.; Parisis, G.; Wakeman, I. Vertex Entropy As a Critical Node Measure in Network Monitoring. *IEEE Transactions on Network and Service Management* **2017**, *14*, 646–660.
8. Dehmer, M.; Mowshowitz, A.; Emmert-Streib, F. Connections between classical and parametric network entropies. *PLoS ONE* **2011**, *6*, 1–8.
9. Bollobás, B. *Random Graphs*, 2nd ed.; Cambridge University Press, 2001.
10. Albert, R.; Barabási, A.L. Statistical mechanics of complex networks. *Review of Modern Physics* **2002**, *74*.
11. Peterson, J.; Dixit, P.D.; Dill, K.A. A maximum entropy framework for nonexponential distributions. *Proceedings of the National Academy of Sciences* **2013**, *110*, 20380–20385, [arXiv:1501.01049v1].
12. Tee, P.; Wakeman, I.; Parisis, G.; Dawes, J.; Kiss, I.Z. Constraints and Entropy in a Model of Network Evolution **2016**. [1612.03115].
13. Shannon, C.E. A Mathematical Theory of Communication. *The Bell System Technical Journal* **1948**, *27*, 379–423.
14. Simonyi, G. Graph entropy: a survey. *Combinatorial Optimization* **1995**, *20*, 399–441.
15. Mowshowitz, A.; Mitsou, V. Entropy, Orbits, and Spectra of Graphs. *Analysis of Complex Networks: From Biology to Linguistics* **2009**, pp. 1–22.
16. Park, J.; Newman, M.E.J. Statistical mechanics of networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* **2004**, *70*, 1–13, [arXiv:cond-mat/0405566].
17. Leskovec, J.; Krevl, A. SNAP Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data, 2014.
18. Clauset, A.; Tucker, E.; Sainz, M. The Colorado Index of Complex Networks. http://icon.colorado.edu/, 2016.
19. Burnham, K.P.; Anderson, D.R. Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research* **2004**, *33*, 261–304, [arXiv:1011.1669v3].
20. Bollobàs, B. *Modern Graph Theory*; Springer-Verlag New York, 1998; pp. XIV, 394.
21. Barabási, A.L. *Network Science*; Cambridge University Press; 1 edition (August 5, 2016), 2016; p. 474.
22. Trugenberger, C.A. Quantum gravity as an information network self-organization of a 4D universe. *Physical Review D - Particles, Fields, Gravitation and Cosmology* **2015**, *92*, 1–12, [1501.0140].
23. Verlinde, E.P. Emergent Gravity and the Dark Universe. *arXiv* **2016**, pp. 0–50, [1611.02269].