UNIVERSITY OF BIRMINGHAM
University of Birmingham Research at Birmingham

Graph characterisation using graphlet-based entropies

Aziz, Furqan; Akbar, Mian Saeed ; Jawad, Muhammad ; Malik, Abdul Haseeb ; Uddin, M. Irfan ; Gkoutos, Georgios

DOI:
10.1016/j.patrec.2021.03.031

License:
Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Document Version
Peer reviewed version

Citation for published version (Harvard):
Aziz, F, Akbar, MS, Jawad, M, Malik, AH, Uddin, MI & Gkoutos, G 2021, 'Graph characterisation using graphlet-based entropies', *Pattern Recognition Letters*, vol. 147, pp. 100-107.
https://doi.org/10.1016/j.patrec.2021.03.031

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 19. Apr. 2024

Pattern Recognition Letters
journal homepage: www.elsevier.com

Graph Characterisation using Graphlet-based Entropies

Furqan Aziz[a,b,c,d,**], Mian Saeed Akbar[e], Muhammad Jawad[f], Abdul Haseeb Malik[f], Irfan Uddin[g], Georgios V. Gkoutos[a,b,c,d,h,i,j]

[a]*Centre for Computational Biology, University of Birmingham, B15 2TT, Birmingham, United Kingdom.*
[b]*College of Medical and Dental Sciences, Institute of Cancer and Genomic Sciences, University of Birmingham, B15 2TT, Birmingham, United Kingdom*
[c]*Institute of Translational Medicine, University Hospitals Birmingham NHS Foundation Trust, B15 2TT, Birmingham, United Kingdom*
[d]*MRC Health Data Research UK (HDR), Midlands, United Kingdom*
[e]*Institute of Management Sciences, Peshawar, Pakistan.*
[f]*Department of Computer Science, University of Peshawar, Peshawar, Pakistan.*
[g]*Institute of Computing, Kohat University of Science and Technology, Kohat, Pakistan.*
[h]*NIHR Experimental Cancer Medicine Centre, B15 2TT, Birmingham, United Kingdom*
[i]*NIHR Surgical Reconstruction and Microbiology Research Centre, B15 2TT, Birmingham, United Kingdom*
[j]*NIHR Biomedical Research Centre, B15 2TT, Birmingham, United Kingdom*

ABSTRACT

In this paper, we present a general framework to estimate the network entropy that is represented by means of an undirected graph and subsequently employ this framework for graph classification tasks. The proposed framework is based on local information functionals which are defined using induced connected subgraphs of different sizes. These induced subgraphs are termed graphlets. Specifically, we extract the set of all graphlets of a specific sizes and compute the graph entropy using our proposed framework. To classify the network into different categories, we construct a feature vector whose components are obtained by computing entropies of different graphlet sizes. We apply the proposed framework to two different tasks, namely view-based object recognition and biomedical datasets with binary outcomes classification. Finally, we report and compare the classification accuracies of the proposed method and compare against some of the state-of-the-art methods.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

The field of graph theory and complex networks has recently emerged as an important area of science helping researchers of different disciplines to understand the characteristics of complex systems. One of the reasons of the increasing interest in complex network science is due to the fact that many complex systems can be represented in the form of graphs where vertices represent elements and edges represent interactions between elements. The increasing availability of large amount of data generated at rapid speeds across a numerous domains has shifted research efforts toward the generation of very large graphs that form a representation of such complex systems. Such graphs are typically termed *complex networks*. Despite obtaining such networks from diverse applications, several structural properties (such as power law degree distribution) are being shared accross them. The study of complex network allows scientists to investigate different properties of complex systems. Complex network has many applications in diverse disciplines, including; text analysis (1), computer vision (2), chemoinformatics (3), biological network analysis (4), and social network analysis (5; 6). Once a system has been represented as a complex network, a number of existing graph-based methods can be used to understand the complex structure of the underlying system and cater decisions or improvements.

The complexity of a network is defined as the summary of the underlying graph structure (7). It can be measured in different ways and helps scientists to quantify properties of networks for example by comparing networks for the identification of structural similarities (8). The simplest measure for comparing graphs is via using the number of nodes, the number of edges and the degree distributions. For example, Barabási et al. (9) have shown that many networks follow a power-law degree distribution, and hence the degree distribution of a network can be used to distinguish between different networks. Degree

**Corresponding author:
e-mail: f.aziz@bham.ac.uk (Furqan Aziz)

distribution is easy to compute but it has been proved that it is not a reliable indicator for network classification. For example, Estrada (10) has shown that the degree distribution of a network fails in uniquely quantifying the differences between real-world networks and has defined the heterogeneity index of a network to gain further insight in 52 different real-world networks. Escolano et al. (11) have proposed another complexity measure, called the thermodynamic depth complexity of a network, that is based on the heat flow process defined on a network. The resultant measure was successfully tested to analyse the structure of a protein-protein interaction (PPI) networks in terms of depth.

Among various complexity measures, entropy based methods have achieved promising accuracy in characterising the structure of a network. For example, Estrada et al. (12) have defined the Shannon entropy of a network that is computed from the eigenvalues of the adjacency matrix of a network and have analyzed the structure and dynamics of complex networks. In (13), the Shannon entropy of a graph is defined based on walks on the graph and the oriented line graph. The connection between quantum statistical mechanics and complex networks have motivated Passerini et al. (14) to define the von Neumann entropy. von Neumann entropy is computed as $S = -\sum_{i=1}^{n} \lambda_i \log \lambda_i$, where λ_i is the i^{th} eigenvalue of the normalised Laplacian of the graph. Passerini et al. have concluded that the von Neumann entropy of a network increases with the regularity properties of the network. Lin et al. (15) have shown the applications of von Neumann entropy to compare graphs that are obtained from images and biomedical datasets. They have also provided an efficient approximation to von Neumann entropy that uses degree distribution of a graph. Later, Ye et al. (16) have extended the use of the von Neumann entropy to directed graphs.

In (17), Dehmer has proposed a novel approach for estimating the structural information content (graph entropy) of a network that is represented by an undirected and connected graph. The method is based on local vertex information functionals that are derived from topological properties of the graph. These vertex functionals are then used to assign a probability value to each vertex of the graph. The resulting probability distribution is then used to obtain the Shannon entropy of the network. In (18), information functional is used to develop a general framework to compute graph entropy. The author has defined three types of information functionals f_v, f_p and f_c. f_v which can be obtained by the definition of local information graph $\mathscr{L}_G(v_i, j)$. f_p is based on path centrality, while f_c is based on vertex centrality measures. Cao et al. (19) have defined graph entropies based on independent sets and matching of graphs. Recently Aziz et al. (20) have presented an information functional that is defined using closed random walks and cycle functionals. They have applied their method to time-series networks and have shown that the obtained measures is more accurate in comparing graphs when compared to alternate information functionals defined in (18). More literature about network entropy measures can be found in (21).

In this paper, the objective is to construct a novel feature vector for a graph that can be used for comparing graphs for structural similarities. The components of the feature vectors are obtained by representing a graph into a transformed space using its connected induced subgraphs of fixed size, called graphlets. We next define a novel information functional that is based on the degree statistics of the graphlet and can be used to obtain a probability distribution over different types of fixed-size graphlets. The resulting probability distribution is then used to estimate the graph entropy in the transformed space. Our main contribution lies with the definition of a general framework that can be used to estimate network entropy, offering several advantages over other approaches. Firstly, our definition of information functional is not only restricted to the vertices of a graph but can be extended to higher order graphlets. This allows us to capture the structure of a graph in a transformed space, where a graph is represented using higher-order graphlets. Secondly, unlike some other approaches, such as (17; 18), that only consider subgraphs obtained by considering a node and all its reachable nodes at a fixed distance, our approach takes into account all possible types of small size graphlets. This helps in providing a richer representation of the structure of a graph. Finally, while most graphlet-based methods for analysing graph (22) utilises only small-order graphlets, our approach also incorporates higher order graphlets. This is possible due to'' the novel definition of the information functional that does not require us to compare graphlets for isomorphism. We have performed numerous experiments on real-world datasets and have empirically demonstrated that the proposed framework is more effective in characterising the structure of a network when compared to some other state-of-the-art methods.

The rest of the paper is organised as follows. In section 2, basic definitions and explanation of the proposed framework are given. Dehmer's framework for computing graph entropy and the definition of information functionals are also provided. Section 3 explains experimental evaluations, where the proposed framework is applied to spatial graphs and have reported classification accuracies. These spatial graphs are extracted from 2D images and chemical compounds. Finally in Section 4, the paper is concluded.

2. Method

The goal of this section is to explain the general framework proposed in the paper for computing the graph entropy and use it to construct a feature vector. The basic definitions of the graph entropy based on the local information functional that is computed from fixed size graphlets of a graph is given and will be useful through the paper.

Definition 1. *A* graph $G = (V, E)$ *is a finite non-empty set of vertices V and a finite set of edges E. Edges are ordered pair of vertices, i.e.* $E \subseteq V \times V$. *A* directed *graph is a graph in which the edges have no direction. An undirected graph is* connected *when there is a path between every pair of vertices. The* degree *of a vertex,* $v \in V$, *in a graph* $G(V, E)$ *is the number of edges incident with v.* □

Definition 2. *A* subgraph $S = (V_S, E_S)$ *of a graph G is a graph whose vertex set and edge set is a subset of the vertex set and*

edge set of G, i.e., $V_S \subseteq V$ and $E_S \subseteq E$. If the subgraph S has the property that whenever two vertices of S are joined by an edge in G they are also joined by an edge in S, then S is called an induced subgraph *of G.* □

Definition 3 (Shannon (23)). *The Shannon entropy of a variable X is defined as:*

$$I(X) = -\sum_{x} P(x)log(P(x)) \tag{1}$$

where P(x) is the probability that X is in the state x. □

2.1. Graph Entropy and Information Functionals

Based on Shannon's definition of entropy, Dehmer et al. (17) have proposed a general framework for defining the graph entropy. Their framework is based on local information functional that captures the local neighbourhood structure of vertices of a graph. They have defined the concept of j-sphere as.

Definition 4 (Dehmer (17)). *Let G be a graph. The set*

$$S_j(v_i, G) := \{v \in V | d(v_i, v) = j, j \geq 1\}, \tag{2}$$

is called the j-sphere of v_i regarding G. □

Based on the definition of j-sphere, the local information functional is defined as follows:

Definition 5 (Dehmer (17)). *Let G be an undirected simple graph. For a vertex $v_i \in V$, its local information functional is defined as*

$$f(v_i) = \alpha^{c_1|S_1(v_i,G)|+c_2|S_2(v_i,G)|+\cdots+c_\rho|S_\rho 1(v_i,G)|}, c_k > 0, 1 \leq k \leq \rho, \alpha > 0. \tag{3}$$

The c_k are arbitrary, real positive coefficients. □

The definition of local information functional allows us to associate a probability distribution to the vertices of a graph.

Definition 6 (Dehmer (17)). *Let G be a simple unlabelled graph. For a vertex $v_i \in V$, its probability is defined as*

$$p(v_i) = \frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)}, \tag{4}$$

where f represents an information functional. Since $p(v_1) + p(v_2) + p(v_3) + \cdots + p(v_n) = 1$, p represents a valid probability distribution. □

Once a valid probability of vertex set is obtained, Dehmer et al. (17) have used Shannon entropy to characterise the structure of a graph.

Definition 7 (Dehmer (17)). *Given a graph $G = (V, E)$, its entropy is defined as*

$$I_f(G) := -\sum_{i=1}^{|V|} \frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)} \log\left(\frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)}\right), \tag{5}$$

where $f(v_i)$ is an arbitrary local vertex information functional. □

Inspired by Dehmer's work, we have developed a novel framework to compute the graph entropy. Instead of defining information functional for a single vertex, here we introduce the idea of defining local information functional for a subset of vertices of a graph. We commence by defining the probability of a fixed size subset of vertices $V^k \subseteq V$.

Definition 8. *Let G be a simple unlabelled graph. Let V^k be a subset of V with size k, i.e., $V^k \subseteq V$ and $\left|V^k\right| = k$. For a fixed value of k, we define:*

$$p(V^k) = \frac{f(V^k)}{\sum_{U^k \subseteq V} f(U^k)}, \tag{6}$$

where f represents a graphlet information functional defined for a subset of vertices of V (See Definition 10). Here U^k also represents a subset of V with size k and the sum is taken over all possible subsets of size k. □

Fixing the value of k allows us to define the graph entropy for different types of graphlets. It is important to mention here that Equation 5 also allows us to define the graph entropy based on subgraphs by using local vertex functional. However, it only considers those subgraphs that are obtained by taking a vertex v and considering all vertices that are at a fixed distance from v. So, for example, if a vertex v has four neighbours, the local vertex functional will only consider subgraphs with five nodes and will ignore all the other subgraphs of smaller size. Our definition, on the other hand, allows us to define graph entropy based on all possible types of subgraphs. Once the value of k is fixed and an information functional is defined, we obtained the following definition of the entropy of a graph:

Definition 9. *Let $G = (V, E)$ be a graph and f be an information functional. We define the entropy of the graph G by*

$$I_f^k(G) = -\sum_{V^k \subseteq V} \frac{f(V^k)}{\sum_{U^k \subseteq V} f(U^k)} \log\left(\frac{f(V^k)}{\sum_{U^k \subseteq V} f(U^k)}\right), \tag{7}$$

where V^k is a subset of vertices V of G with size k. □

Note that when $k = 1$, Equation 7 reduces to Equation 5. In other words, our definition of entropy can be considered as a generalised definition. It is also important to note that when we define k=2, and use the connectivity constraint, we obtain the edge entropy of the graph. Here $f(v_i, v_j)$ can be considered as local edge information functional. Therefore, in order to define the entropy of a graph for a specific value of k, we need to define the information functional that accepts k vertices of a graph as an input argument.

In this paper our objective is to define graph entropy for different values of k, and use the resulting values as components of a feature vector that can be used to characterise graph in a low-dimensional feature space. Here we define information functional using induced subgraphs of a graph, which are also called graphlets. Using graphlets for characterising graph is a well studied problem in the field of network analysis (24). Graphlet-based methods are generally based on counting the

frequencies of isomorphic graphlets and then using the resultant frequencies to embed the graph in a low-dimensional feature space. Figure 1 shows all possible graphlets of size 3, 4 and 5.

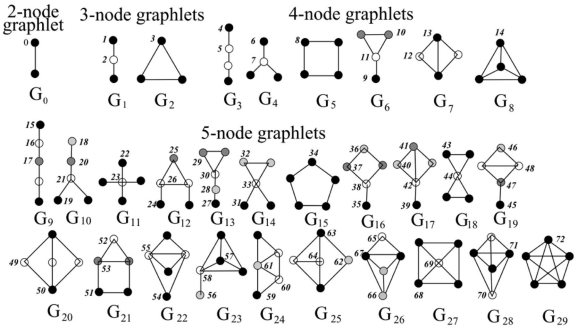


Fig. 1: Set of all possible graphlet of size 3, 4 and 5 (25).

The frequency of small order graphlets can generally be efficiently computed, however, computing the frequency of higher-order graphlets is a hard problem. Therefore, most of these methods use the frequencies of first few graphlets. There are 2 graphlets of size 3, 6 of size 4 and 21 of size 5. Note that graphlet of size 1 is simply a node while a graphlet of size 2 is an edge. We now define information functional for a graphlet.

Definition 10. *Let g_i^k represent the i^{th} graphlet of size k. For this graphlet, we define the information functional as:*

$$f(g_i^k) = e^{-\left(\frac{1}{d_1^i+1}+\frac{1}{d_2^i+1}+\frac{1}{d_3^i+1}\cdots\frac{1}{d_k^i+1}\right)}, \tag{8}$$

where $d_1^i, d_2^i, ..., d_k^i$ is the degree sequence of the graphlet g_i^k. □

The general idea behind computing the information functional, $f(V^k)$, for a given graphlet g_i^k, is that we consider the set all possible subset of a graph G, with size k, and discard those that are not fully connected. In other words, we find the set of all possible connected subgraphs G, which is the set of all possible graphlets of size k. For each graphlet, we compute its information functional using its degree statistics, according to Definition 10. Finally, using Equation 7, the graph entropy for a fixed value of k is computed. Unfortunately, finding the set of graphlets for all possible values of k is computationally infeasible. For this reason, we only consider those values of k for which we can efficiently find the set of all graphlets. In experimental evaluation, we set the value of $k = 1, 2, 3, 4, 5, |V|-1, |V|$. Using Equation 7 and Equation 8, we compute the value of $I_f^k(G)$ for a fixed value of k.

$$\begin{aligned} I_f^k(G) &= -\sum_{V_G^k \subseteq V_G} \frac{f(V_G^k)}{\sum_{U_G^k \subseteq V_G} f(U_G^k)} \log\left(\frac{f(V_G^k)}{\sum_{U_G^k \subseteq V_G} f(U_G^k)}\right), \\ &= -\sum_{g_i^k \in g^k} \frac{f(g_i^k)}{\sum_{g_j^k \in g^k} f(g_j^k)} \log\left(\frac{f(g_i^k)}{\sum_{g_j^k \in g^k} f(g_j^k)}\right), \end{aligned} \tag{9}$$

where g^k represents the set of all connected graphlets of size k that are present in G. Note that the above equation can be further simplified, if the frequency of each graphlet is known. Let $\left|g_i^k\right|$ be the frequency of the graphlet g_i^k in G. Then the above equation can be expressed as:

$$I_f^k(G) = -\sum_{i=1}^{T} \frac{\left|g_i^k\right| f(g_i^k)}{\sum_{j=1}^{T} \left|g_j^k\right| f(g_j^k)} \log\left(\frac{f(g_i^k)}{\sum_{j=1}^{T} \left|g_j^k\right| f(g_j^k)}\right), \tag{10}$$

where T is the frequency of different types of graphlets of size k. We now compute the value of $I_f^k(G)$, for $k = 1, 2, 3$. For $k = 1$, there is only one graphlet of size 1, i.e., a single node of a graph. For this graphlet $f(g_1^1) = e^{-1}$. Therefore

$$\begin{aligned} I_f^1(G) &= -\sum_{v_i \in V_G} \frac{f(v_i)}{\sum_{v_j \in V_G} f(v_j)} \log\left(\frac{f(v_i)}{\sum_{v_j \in V_G} f(v_j)}\right), \\ &= -\sum_{v_i \in V_G} \frac{e^{-1}}{\sum_{v_j \in V_G} e^{-1}} \log \frac{e^{-1}}{\sum_{v_j \in V_G} e^{-1}}, \\ &= \log |V_G|. \end{aligned} \tag{11}$$

Similarly, for $k = 2$, there is also only one graphlet of size 2, which is a single edge of a graph in this case. For this graphlet $f(g_1^2) = e^{-1}$. Therefore,

$$\begin{aligned} I_f^2(G) &= -\sum_{e_i \in E_G} \frac{f(e_i)}{\sum_{e_j \in E_G} f(e_j)} \log\left(\frac{f(e_i)}{\sum_{e_j \in E_G} f(e_j)}\right), \\ &= -\sum_{e_i \in E_G} \frac{e^{-1}}{\sum_{e_j \in E_G} e^{-1}} \log \frac{e^{-1}}{\sum_{e_j \in E_G} e^{-1}}, \\ &= \log |E_G|. \end{aligned} \tag{12}$$

For $k = 3$, there are two types of connected graphlets, i.e., a line and a triangle as shown in Figure 1. Let g_1^3 and g_2^3 be the line and triangle graphlets respectively. Since the degree distributions of g_1^3 is $\langle 1, 1, 2\rangle$ and that of g_2^3 is $\langle 2, 2, 2\rangle$, therefore $f(g_1^3) = e^{-\frac{4}{3}}$ and $f(g_2^3) = e^{-1}$. Thus the entropy $I_f^3(G)$ can be computed as follows:

$$\begin{aligned} I_f^3(G) &= -\frac{\left|g_1^3\right| e^{-\frac{4}{3}}}{\left|g_1^3\right| e^{-\frac{4}{3}} + \left|g_2^3\right| e^{-1}} \log\left(\frac{e^{-\frac{4}{3}}}{\left|g_1^3\right| e^{-\frac{4}{3}} + \left|g_2^3\right| e^{-1}}\right), \\ &\quad - \frac{\left|g_2^3\right| e^{-1}}{\left|g_1^3\right| e^{-\frac{4}{3}} + \left|g_2^3\right| e^{-1}} \log\left(\frac{e^{-1}}{\left|g_1^3\right| e^{-\frac{4}{3}} + \left|g_2^3\right| e^{-1}}\right), \\ &= -\frac{\left|g_1^3\right|}{\left|g_1^3\right| + \left|g_2^3\right| e^{\frac{1}{3}}} \log\left(\frac{1}{\left|g_1^3\right| + \left|g_2^3\right| e^{\frac{1}{3}}}\right) \\ &\quad - \frac{\left|g_2^3\right| e^{\frac{1}{3}}}{\left|g_1^3\right| + \left|g_2^3\right| e^{\frac{1}{3}}} \log\left(\frac{e^{\frac{1}{3}}}{\left|g_1^3\right| + \left|g_2^3\right| e^{\frac{1}{3}}}\right), \\ &= \left(\left|g_1^3\right| + \left|g_2^3\right| e^{\frac{1}{3}}\right)^{-1} \\ &\quad \left[\left(\left|g_1^3\right| + \left|g_2^3\right| e^{\frac{1}{3}}\right) \log\left(\left|g_1^3\right| + \left|g_2^3\right| e^{\frac{1}{3}}\right) - \frac{1}{3}\left|g_2^3\right| e\right]. \end{aligned} \tag{13}$$

2.2. Feature Vector Construction

In order to embed the graph into a low-dimensional feature space, a feature vector is constructed whose components represent the the graph entropies. The main objective is to obtain a powerful and invariant graph representation that can also be computed efficiently. Such representation can be used to classify graphs with higher accuracy. A feature vector of length n is proposed as $\vec{v}_G = [I_f^1(G), I_f^2(G), I_f^3(G), ..., I_f^n(G)]$. In other words, the k^{th} component of the feature vector $\vec{v}_G$ is obtained by computing the entropy $I_f^k(G)$. However, since the computation of all $I_f^k(G)$ for all values of k is generally not practical, only a subset of values that could be efficiently computed for a given graph are selected. Finally, a normalisation step is performed, where the feature vector is normalised by dividing the components of the feature vector by the sum of all the components of the feature vector. This step is usually performed in literature in order to make the feature vector invariant of the size of the graph (26).

In experimental evaluation a feature vector

$$\vec{v}_G = [I_f^1(G), I_f^2(G), I_f^3(G), I_f^4(G), I_f^5(G), I_f^{|V|-2}(G), I_f^{|V|-1}(G)]$$

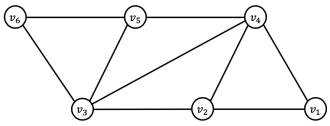
is constructed [1]. The reason for considering the first five components is the low computational cost. As mentioned earlier, the graphlets of sizes up to 5 can be computed very efficiently. These small graphlets will capture the local structure of a graph. Note that the alternate graphlet based method also consider small order graphlets due to computation cost.

To compute $I_f^{|V|-1}(G)$, a single vertex of G is removed. If the resulting graph is connected, then the value of the information functional using its degree distribution is computed. This process is repeated for all the vertices of the graph. Finally these values are substituted in Equation 9 to obtain graph entropy.

To demonstrate this process, consider the graph shown in Figure 2. Let g_U^j represents a graphlet of G with size j, that includes all the nodes of G except nodes in the set U, where $U \subseteq V$. To compute the value of $f(g_{v_1}^5)$, the node v_1 is removed from the graph. The resultant connected graph with 4 vertices and 5 edges has degree sequence $2, 2, 3, 3, 4$. Therefore $f(g_{v_1}^5) = e^{-\frac{41}{30}}$. It is worth mentioning here that the entropy for graphlets of size $|V|-1$ can be computed very efficiently. This is because the total number of graphlets of size $|V|-1$ are bounded by $O(|V|)$. Computing information functional for each graphlet requires $O(|V|^2)$ time, as this operation requires computing the degree of each vertex of the graphlet. Therefore the worst case running time for computing entropy in this case is bounded by $O(|V|^3)$. The value of $I_f^{|V|-2}(G)$ can be computed in a similar way. In this case a maximum of $\binom{|V|}{2} = \frac{|V|(|V|-1)}{2} = O(|V|^2)$ connected graphlets of size $|V| - 2$ are computed. Therefore the running time of computing entropy in this case is bounded by $O(|V|^4)$. Note that, graphs of size greater than five are considered. For a graph of size five, set of all of its graphlets can be efficiently computed.

The reason for including the last two components is to obtain a richer representation of the graph that is less prone to the problem of failing to distinguish pair of non-isomorphic graphs that have same frequency of lower-order graphlets. This is illustrated with the help of a simple example. Consider the pair of graphs, G_1 and G_2, shown in Figure 3.

[1] Code to compute $\vec{v}_G$ is available at https://github.com/azizfurqan/GE

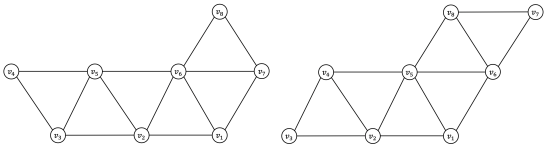


(a)

$$f(g_{v_1}^5) = e^{-\frac{41}{30}}, f(g_{v_6}^5) = e^{-\frac{41}{30}}, f(g_{v_2}^5) = e^{-\frac{19}{12}}, f(g_{v_5}^5) = e^{-\frac{19}{12}}$$

$$f(g_{v_3}^5) = e^{-\frac{7}{4}}, f(g_{v_4}^5) = e^{-\frac{7}{4}}$$

Fig. 2: Information functional for a simple graph G with 6 vertices and 9 edges. Here g_U^j represents a connected subgraph of G of size j, that includes all the nodes of G except nodes in U, where $U \subseteq V$.



(a) G_1 (b) G_2

Fig. 3: Pair of non-isomorphic graphs

Clearly, the graphs G_1 and G_2 are non-isomorphic, since the unique vertex of degree 5 in G_1 is adjacent to a vertex of degree 2 which is not the case for G_2. Both the graphs G_1 and G_2 share the same number of vertices, edges, unique graphlets of size 3, and unique graphlets of size 4. However, the frequency of unique graphlets of size greater than 4 is different. Therefore, graphlet kernel, when used with graphlets of size up to 4 will fail to distinguish the two non-isomorphic graphs shown in Figure 3. On the other hand, the proposed method will produce the feature vectors $\vec{v}_{G_1} = [3.0000, 3.7004, 4.3745, 4.8794, 4.5480, 4.4924, 2.9834]$ and $\vec{v}_{G_2} = [3.0000, 3.7004, 4.3745, 4.8794, 4.5278, 4.4931, 2.9835]$ for graph G_1 and G_2 respectively. Note that the higher-order components of $\vec{v}_{G_1}$ and $\vec{v}_{G_2}$ are different which means that inclusion of higher-order graphlets can increase the discriminative power of graph representation.

It is worth noting that most of the graphlet based methods are based on counting the frequency of isomorphic graphlets. This requires the testing of which pair of graphlets are isomorphic. This problem is one of few standard problems in computational complexity theory that belongs to NP, but not known to belong to either P or NP-Complete. While testing of isomorphism property for small graphlets is computationally feasible, but becomes computationally expensive for large graphlets. For this reason, graphlet based methods generally utilise graphlets of small sizes. One of the advantages of the proposed feature vector is that its components can be computed using Equation 9, which does not require to compute the frequencies of each graphlet. So, although the set of all graphlets of size k are to be searched, there is no need to find the fre-

quencies of each graphlet which requires to test graphlets for isomorphism (a problem known to be NP-complete). This allows the computation of entropy for high order graphlets. In the experiment section a feature vector is constructed that utilises both low-order and high-order graphlets to obtain a richer graph representation.

3. Experiments

In this section the proposed method is applied to spatial graphs that are extracted from two different types of datasets. The first type of graphs are extracted from images while the second type of graphs are extracted from biomedical datasets. To evaluate the performance of the proposed method, the following state-of-the-art methods are used for comparison, i.e., Ihara Coefficients (27), Random Walk Kernel (28), Graphlet Kernel (GK) (22), Shortest Path Kernel (SPK) (29), ShapeDNA (30) and Von Neumann Entropy (14).

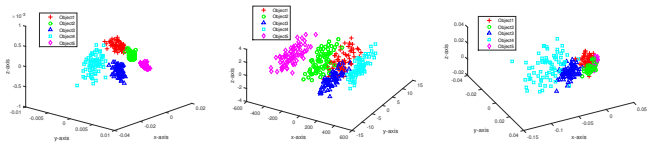
To compare the classification accuracies of the proposed and the alternate methods, r-fold cross-validation is used. In r-fold cross validation the original data is partitioned into r subsamples of equal size. In these r subsamples a single subsample is retained as the validation set for testing the model, while the remaining $r-1$ subsamples are used as the training set. This process of cross-validation is then repeated r times. For our experiments, $r = 10$ repeats, widely applied in similar experiments, were applied. The classification accuracies are estimated using a Support Vector Machine (SVM) with a standard Gaussian kernel. SVM is a discriminative classifier that, given labeled training data, outputs an optimal hyperplane which is used to categorise new examples.

3.1. View based object recognition

In this section experiments are performed on simple unlabelled graphs that are extracted from images available within the Columbia Object Image Library (COIL) (31). COIL consists of 20 different object each with 72 different views. These views are taken under controlled environment and lightning conditions. To extract an unlabelled graph from each view, the feature points of the image are located using the Harris corner detector (32). These feature points are then treated as vertices of the graph and a Delaunay Triangulation (33) is obtained to get the edges. A Delaunay triangulation (DT) for a set P of points in a Euclidean space is a triangulation, DT(P), such that no point in P is inside the circumcircle of any triangle in DT(P). DT are $md2$ graphs ($md2$ graphs are those graphs whose nodes have degree at least two). In the experiments only five different objects are used with all their views. These objects are shown in Figure 4.

Fig. 4: Objects from COIL used in the experiments.

For each graph a normalised feature vector is computed as, $\vec{v}_G = [I_f^1(G), I_f^2(G), I_f^3(G), I_f^4(G), I_f^5(G), I_f^{|V|-2}(G), I_f^{|V|-1}(G)]$. To visualise the multi-dimensional data to 3D, the dimensionality reduction technique, principal component analysis (PCA) is applied on the resulting feature vectors. *PCA* is mathematically defined (34) as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. Figure 5a shows the embedding results in a three-dimensional feature space.



(a) Proposed Feature Vector (b) Ihara Coefficients (c) Graphlet Kernel

Fig. 5: PCA embedding for Delaunay Triangulations for the proposed and alternate methods

To compare the performance of the proposed method, a similar procedure to Ihara coefficients, which are considered a powerful tool in characterising $md2$ graphs (35), is applied. A feature vector similar to one proposed by Ren et al. (27) is used. The visualisation results for Ihara coefficients are shown in Figure 5b. Finally for comparison purposes, the visualisation results of a graphlet kernel are also shown in Figure 5c. These visualisation results suggest that, the proposed method not only provides a better separation between objects of different classes, but it also provides low interclass variation as compared to alternate techniques. To evaluate the performance of the proposed method, its accuracy is estimated using support vector machine with a standard Gaussian kernel and have validated the result of Support Vector Machine (SVM using 10-fold cross validation. A detailed comparison of the resultant average accuracies is provided in Table 1

Table 1: Classification accuracies for Delaunay Triangulations

Methods	Accuracy
Proposed	**97.78 ± 0.02**
Ihara Coefficients	94.17 ± 1.03
Graphlet kernel	86.94 ± 0.04

These results demonstrate the effectiveness of the proposed method in distinguishing between graphs. Note that the accuracy of the proposed method is considerably higher than that of a Graphlet kernel which suggests that by considering higher order graphlets the classification accuracy can be significantly improved. The classification accuracy of the proposed method is also higher than that of Ihara coefficients which is generally considered a very powerful tool in classifying Delaunay triangulations (27; 35).

Table 2: Classification accuracies of different methods on bioinformatics datasets

Datasets	Proposed	Lower order	Graphlet	Random Walk	Shortest Path	ShapeDNA	VNE
MUTAG	88.33 ± 2.18	85.15 ± 2.03	85.06 ± 2.76	85.64 ± 1.93	86.05 ± 2.14	86.50 ± 2.22	85.01 ± 2.69
AIDS	99.15 ± 0.20	97.00 ± 0.35	96.45 ± 0.46	90.55 ± 2.01	99.50 ± 0.01	92.65 ± 0.55	85.64 ± 2.22
NCI1	70.02 ± 0.69	65.84 ± 0.7242	68.74 ± 0.63	63.92 ± 0.65	69.68 ± 0.55	68.35 ± 0.60	60.26 ± 0.36
NCI109	70.90 ± 0.95	68.57 ± 0.8254	69.74 ± 0.53	62.22 ± 0.73	68.74 ± 0.58	67.92 ± 0.58	59.58 ± 0.69
PTC-MR	63.08 ± 2.00	58.75 ± 1.85	59.77 ± 1.68	58.38 ± 2.38	58.53 ± 2.75	57.59 ± 1.95	58.20 ± 2.36
PTC-MM	65.51 ± 2.13	63.43 ± 1.42	65.48 ± 1.62	62.86 ± 3.04	63.76 ± 2.51	65.48 ± 1.67	62.82 ± 1.90
PTC-FR	68.97 ± 1.77	65.83 ± 1.22	68.57 ± 1.04	66.95 ± 1.96	66.39 ± 1.73	67.25 ± 1.11	62.98 ± 2.22
PTC-FM	63.34 ± 1.30	57.29 ± 1.57	59.17 ± 2.60	59.89 ± 1.63	60.73 ± 1.39	60.72 ± 1.96	58.75 ± 2.12
COX2	78.59 ± 0.83	79.03 ± 1.27	78.16 ± 1.03	77.10 ± 0.69	78.58 ± 0.71	78.17 ± 0.66	78.16 ± 0.03

3.2. Biomedical Datasets:

In our last set of experiment, we apply the proposed method on biomedical datasets. For this purpose we have chosen publicly available datasets, including MUTAG (36), AIDS (37), NCI1 and NCI109 (38), PTC (39) and COX2 (40). We note here that the graphs in these datasets are undirected labelled graphs, whose vertices are labelled with the atom names and whose edges are labelled with the bond type between the atoms. Since, our method is developed for unlabelled graphs, we have converted all the graphs into unlabelled graphs by ignoring the graph nodes and edge labels. The same unlabelled datasets were used when the performance of the other state-of-the-art methods was assessed.

In order to evaluate the performance of the proposed method, we apply it on graphs obtained from publicly available datasets. This results in embedding the graph in a 7-dimensional feature space. To estimate the classification accuracies, we have used support vector machine with standard Gaussian kernel. As with the COIL dataset, the results of support vector machines were validated using a 10-fold cross validation. To compare the performance of the proposed method, we have applied a similar procedure to estimate the average accuracies of other state-of-the-art methods discussed earlier. Note that we have not considered the Ihara coefficients in comparison, as Ihara coefficients may fail in characterising graphs that may have vertices with degree one (35). Table 2 compares the resulting accuracies of the proposed and alternate methods. The gray cells in the table indicate the best performance, while the light gray cells correspond to the second best performance. In order to demonstrate the importance of higher order graphlets, we have also computed the accuracies of the proposed feature vector that was constructed using graphlets of size up to five. These accuracies are reported in the second column of the table.

The classification accuracies reported in Table 2 suggest that the proposed method can be used to classify graphs with higher accuracies when compared to other state of the art methods. Our proposed method has outperformed all other assessed methods across all the datasets bar the AIDS and COX2 ones, where the proposed method resulted in the second best performance. It is important to note that, in all cases, the proposed method always results in superior performance when compared to a graphlet kernel. In most cases, the difference between classification accuracies is significant. Finally, the second column of the table shows that the classification accuracies are significantly reduced if the higher order components of the feature vectors are ignored. This suggest that the inclusion of higher-order graphlets in our framework has significantly increased the performance of the classifier.

4. Conclusion

The paper presented a novel technique for estimating the entropy of a complex network and has explored its applications in understanding the structure of complex networks. The framework is based on the assumption that the complex network is represented by an undirected and unlabelled graph. The proposed framework is based on the information functional, which can be defined for any size graphlets. The advantages of estimating entropies at different levels is that it enables the graph to be embedded in a lower dimension feature space, where traditional machine learning algorithms can be used to compare graphs for similarities (or dissimilarities). To demonstrate the applications of the proposed framework, the approach is applied to networks extracted from images and biomedical datasets. The experimental results show that the proposed method can classify graphs with higher accuracy compared to some current state-of-the-art techniques. Potential future research on this work could include the exploration of whether the definition of other information functionals can be used to classify networks with higher accuracy. It would also be interesting to extend the framework to labelled and/or weighted graphs as well as other general data structures, such as directed graphs and hypergraphs.

Acknowledgement

GVG and FA acknowledge support from the NIHR Birmingham ECMC, NIHR Birmingham SRMRC, Nanocommons H2020-EU (731032) and the NIHR Birmingham Biomedical Research Centre and the MRC Heath Data Research UK (HDRUK/CFC/01), an initiative funded by UK Research and Innovation, Department of Health and Social Care (England) and the devolved administrations, and leading medical research charities. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, the Medical Research Council or the Department of Health.

References

[1] H. Ferraz de Arruda, F. Nascimento Silva, V. Queiroz Marinho, D. Raphael Amancio, L. da Fontoura Costa, Representation of texts as complex networks: a mesoscopic approach, Journal of Complex Networks 6 (1) (2017) 125–144.

[2] J. J. de Mesquita Sá, A. R. Backes, P. C. Cortez, Texture analysis and classification using shortest paths in graphs, Pattern Recognition Letters 34 (11) (2013) 1314 – 1319.

[3] B. Gaüzère, L. Brun, D. Villemin, Two new graphs kernels in chemoinformatics, Pattern Recognition Letters 33 (15) (2012) 2038 – 2047.

[4] F. Aziz, A. Acharjee, J. A. Williams, D. Russ, L. Bravo-Merodio, G. V. Gkoutos, Biomarker prioritisation and power estimation using ensemble gene regulatory network inference, International Journal of Molecular Sciences 21 (21) (2020) 7886.

[5] K. R. Žalik, B. Žalik, Memetic algorithm using node entropy and partition entropy for community detection in networks, Information Sciences 445-446 (2018) 38 – 49.

[6] F. N. Silva, C. H. Comin, T. K. Peron, F. A. Rodrigues, C. Ye, R. C. Wilson, E. R. Hancock, L. da F. Costa, Concentric network symmetry, Information Sciences 333 (2016) 61 – 80.

[7] K. Anand, G. Bianconi, S. Severini, Shannon and von neumann entropy of random networks with heterogeneous expected degree, Phys. Rev. E 83 (2011) 036109.

[8] C. Orsini, M. M. Dankulov, P. Colomer-de Simón, A. Jamakovic, P. Mahadevan, A. Vahdat, K. E. Bassler, Z. Toroczkai, M. Boguñá, G. Caldarelli, S. Fortunato, D. Krioukov, Quantifying randomness in real networks, Nature Communications 6 (2015) 8627.

[9] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.

[10] E. Estrada, Quantifying network heterogeneity, Phys. Rev. E 82 (2010) 066102.

[11] F. Escolano, E. R. Hancock, M. A. Lozano, Heat diffusion: Thermodynamic depth complexity of networks, Phys. Rev. E 85 (2012) 036206.

[12] E. Estrada, N. Hatano, Statistical-mechanical approach to subgraph centrality in complex networks, Chem. Phys. Letters 439 (1) (2007) 247 – 251.

[13] E. Estrada, A. José, N. Hatano, Walk entropies in graphs, Linear Algebra and its Applications 443 (2014) 235–244.

[14] F. Passerini, S. Severini, The von neumann entropy of networks, arXiv:0812.2597v2 (2008).

[15] L. Han, F. Escolano, E. R. Hancock, R. C. Wilson, Graph characterizations from von neumann entropy, Pattern Recognition Letters 33 (15) (2012) 1958 – 1967.

[16] C. Ye, R. C. Wilson, C. H. Comin, L. d. F. Costa, E. R. Hancock, Approximate von neumann entropy for directed graphs, Phys. Rev. E 89 (2014) 052804.

[17] M. Dehmer, A novel method for measuring the structural information content of networks, Cybernetics and Systems 39 (8) (2008) 825–842.

[18] M. Dehmer, Information processing in complex networks: Graph entropy and information functionals, Applied Mathematics and Computation 201 (1) (2008) 82 – 94.

[19] S. Cao, M. Dehmer, Z. Kang, Network entropies based on independent sets and matchings, Applied Mathematics and Computation 307 (2017) 265 – 270.

[20] F. Aziz, E. R. Hancock, R. C. Wilson, Graph entropy from closed walk and cycle functionals, in: S+SSPR, Springer, 2016, pp. 174–184.

[21] M. Dehmer, A. Mowshowitz, A history of graph entropy measures, Information Sciences 181 (1) (2011) 57 – 78.

[22] N. Shervashidze, S. Vishwanathan, T. Petri, K. Mehlhorn, K. Borgwardt, Efficient graphlet kernels for large graph comparison, in: Proceedings of Machine Learning Research, Vol. 5, PMLR, Florida USA, 2009, pp. 488–495.

[23] C. E. Shannon, W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, Urbana and Chicago, 1949.

[24] M. Bressan, F. Chierichetti, R. Kumar, S. Leucci, A. Panconesi, Counting graphlets: Space vs time, in: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, ACM, 2017, pp. 557–566.

[25] W. Hayes, K. Sun, N. Pržulj, Graphlet-based measures are suitable for biological network comparison, Bioinformatics 29 (4) (2013) 483–491.

[26] N. Pržulj, D. G. Corneil, I. Jurisica, Modeling interactome: scale-free or geometric?, Bioinformatics 20 (18) (2004) 3508–3515.

[27] P. Ren, R. C. Wilson, E. R. Hancock, Graph characterization via ihara coefficients, IEEE Trans. on Neural Networks 22 (2) (2011) 233–245.

[28] T. Gärtner, P. Flach, S. Wrobel, On graph kernels: Hardness results and efficient alternatives, in: Learning Theory and Kernel Machines, Springer, 2003, pp. 129–143.

[29] K. M. Borgwardt, H.-P. Kriegel, Shortest-path kernels on graphs, in: Data Mining, Fifth IEEE International Conference on, IEEE, 2005, pp. 8–pp.

[30] M. Reuter, F.-E. Wolter, N. Peinecke, Laplace–beltrami spectra as 'shape-dna' of surfaces and solids, CAD 38 (4) (2006) 342 – 366.

[31] S. Nayar, S. A. Nene, H. Murase, Columbia object image library (coil 100). department of comp, Science, Columbia University, Tech. Rep. CUCS-006-96 (1996).

[32] C. Harris, M. Stephens, A combined corner and edge detector., in: Alvey vision conference, Citeseer, 1988, pp. 10–5244.

[33] B. Delaunay, Sur la sphre vide, izvestia akademii nauk sssr, Otdelenie Matematicheskikh i Estestven-nykh Nauk (1934) 793–800.

[34] I. T. Jolliffe, Principal component analysis, Springer-Verlag, New York (1986).

[35] F. Aziz, R. C. Wilson, E. R. Hancock, Backtrackless walks on a graph, IEEE trans. on neural netw. and learning sys. 24 (6) (2013) 977–989.

[36] A. K. Debnath, R. L. Lopez de Compadre, G. Debnath, A. J. Shusterman, C. Hansch, Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. correlation with molecular orbital energies and hydrophobicity, J. of medicinal chemistry 34 (2) (1991) 786–797.

[37] K. Riesen, H. Bunke, Iam graph database repository for graph based pattern recognition and machine learning, in: S+SSPR, Springer, 2008, pp. 287–297.

[38] N. Wale, I. A. Watson, G. Karypis, Comparison of descriptor spaces for chemical compound retrieval and classification, Knowledge and Information Systems 14 (3) (2008) 347–375.

[39] C. Helma, R. D. King, S. Kramer, A. Srinivasan, The predictive toxicology challenge 2000–2001, Bioinformatics 17 (1) (2001) 107–108.

[40] J. J. Sutherland, L. A. O'brien, D. F. Weaver, Spline-fitting with a genetic algorithm: A method for developing classification structure- activity relationships, Chem. info. and computer sciences 43 (6) (2003) 1906–1915.