Physica A 404 (2014) 341–358

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Correlation of automorphism group size and topological properties with program-size complexity evaluations of graphs and complex networks

Hector Zenil [a,c,*], Fernando Soler-Toscano [b,c], Kamaludin Dingle [d,e], Ard A. Louis [d]

[a] Unit of Computational Medicine, Karolinska Institute, Sweden
[b] Grupo de Lógica, Lenguaje e Información, Universidad de Sevilla, Spain
[c] Algorithmic Nature Group, LABoRES, Paris, France
[d] Rudolf Peierls Centre for Theoretical Physics, University of Oxford, UK
[e] The Systems Biology DTC, University of Oxford, UK

HIGHLIGHTS

- It is numerically shown that graphs have same complexity as dual graphs.
- Is shown a correlation between automorphism size and complexity.
- It is shown that real-world graphs can be classified by algorithmic complexity.
- We advance a method by algorithmic probability alternative to compression.
- Is shown that Kolmogorov complexity is sensitive to topological differences.

ARTICLE INFO

Article history:
Received 9 April 2013
Received in revised form 19 February 2014
Available online 6 March 2014

Keywords:
Kolmogorov complexity
Complex networks
Graph automorphisms
Algorithmic probability
Compressibility
Network biology

ABSTRACT

We show that numerical approximations of Kolmogorov complexity (K) of graphs and networks capture some group-theoretic and topological properties of empirical networks, ranging from metabolic to social networks, and of small synthetic networks that we have produced. That K and the size of the group of automorphisms of a graph are correlated opens up interesting connections to problems in computational geometry, and thus connects several measures and concepts from complexity science. We derive these results via two different Kolmogorov complexity approximation methods applied to the adjacency matrices of the graphs and networks. The methods used are the traditional lossless compression approach to Kolmogorov complexity, and a normalised version of a Block Decomposition Method (BDM) based on algorithmic probability theory.

1. Introduction

Graphs are an important tool for mathematically analysing many systems, from interactions of chemical agents, to ecological networks, to representing data objects in computer science [1,2]. An interesting perspective regarding such graphs is to investigate the complexity [3,4] or information content of a graph [5]. While Shannon information theory [6,5,7] and

* Corresponding author at: Unit of Computational Medicine, Karolinska Institute, Sweden.
E-mail addresses: hzenilc@gmail.com, hectorz@labores.eu (H. Zenil).

http://dx.doi.org/10.1016/j.physa.2014.02.060
0378-4371/

Fig. 1. An example of a graph which has no symmetries (i.e. automorphism group size of 1, cf. Definition 2), despite being intuitively simple, i.e. just a string of nodes with a single side node.

counting symmetries [8,9] have been applied to measure the information content/complexity of graphs, little has been done, by contrast, to demonstrate the utility of Kolmogorov complexity as a numerical tool for graph and real-world network investigations. Some theoretical connections between graphs and algorithmic randomness have been explored (e.g. Ref. [10]), but these are mostly related to formal properties of random graphs.

Here we computationally study some of these numerical and real-world directions and show how Kolmogorov complexity can capture group-theoretic and topological properties of abstract and empirical graphs and networks. We do this by introducing a measure of graph complexity based on approximating the Kolmogorov complexity of the adjacency matrix representation of a graph, which we achieve by applying our recently developed *Block Decomposition Method* (below) [11]. A theoretical advantage of using Kolmogorov complexity K is that the measure K is designed to capture *all* structure (i.e. non-randomness) in an object, such as a graph. In contrast, only looking at symmetries, for example, can miss structure and potential simplicity in a graph. For example, Fig. 1 shows a graph with no symmetries, despite being far from random, and indeed is intuitively simple.

This paper is structured as follows: the next two sections give background definitions and theorems; then in Section 4 we consider the question of whether approximations of Kolmogorov complexity can characterise group-theoretic and topological properties of graphs and networks, more specifically, we study the number of automorphisms of a graph and introduce a normalised measure of graph complexity; following this in Section 5 we apply our measure to real-world networks; and finally we study in Section 6 algorithmic information-theoretic similarities of networks with similar topologies created using different mechanisms (e.g. random versus preferential attachment).

2. Preliminaries

2.1. Graph notation

A graph $g = (V, E)$ consists of a set of vertices V (also called nodes) and a set of edges E. Two vertices, i and j, form an edge of the graph if $(i, j) \in E$. Let the binary adjacency matrix of g be denoted by $Adj(A)$. A graph can be represented by its adjacency matrix. Assuming that the vertices are indices from 1 to n, that is, that $V = \{1, 2, \ldots, n\}$, then the adjacency matrix of g is an $n \times n$ matrix, with entries $a_{i,j} = 1$ if $(i, j) \in E$ and 0 otherwise. The distance $D(g)$ of a graph g is the maximum distance between any 2 nodes of g. The size $V(g)$ of a graph g is the vertex count of g; similarly $E(g)$ denotes the edge count of g.

Definition 1. Two graphs g and h are *isomorphic* if and only if there exists a permutation λ such that $\lambda(g) = h$. (That is, g and h are topologically equivalent.)

The general problem of graph isomorphism appears, for example, in chemistry [12,13] and biology [14,15].

Definition 2. An *automorphism* of a graph g is a permutation λ of the vertex set V, such that the pair of vertices (i, j) forms an edge if and only if the pair $(\lambda(i), \lambda(j))$ also forms an edge.

The set of all automorphisms of an object forms a group, called the *automorphism group*. Intuitively, the size of the automorphism group $A(g)$ provides a direct measure of the abundance of symmetries in a graph or network. Every graph has a trivial symmetry (the identity) that maps each vertex to itself.

A clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together (for example, friends in social networks [16]).

Definition 3.

$$C(v_i) = \frac{2\,|E(N_i)|}{n_i(n_i - 1)}$$

where $E(N_i)$ denotes the set of edges with both nodes in N_i, and $C(v_i)$ is the local clustering coefficient of node v_i.

3. Approximating Kolmogorov complexity

3.1. Background results

Before describing our graph complexity measure, we provide some pertinent definitions and results from the theory of algorithmic randomness.