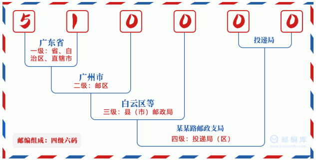
Structural Information and Dynamical Complexity of Networks

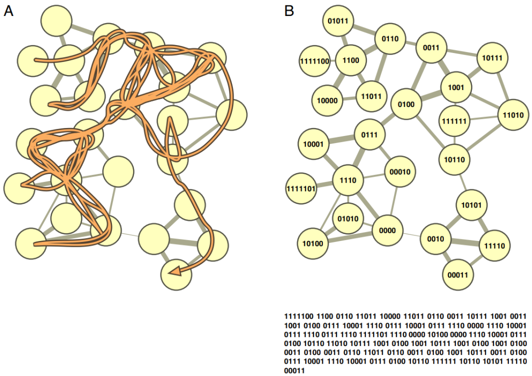
Fanhui Meng

COIN,
School of Computer Science and Engineering,
Sun Yat-sen University

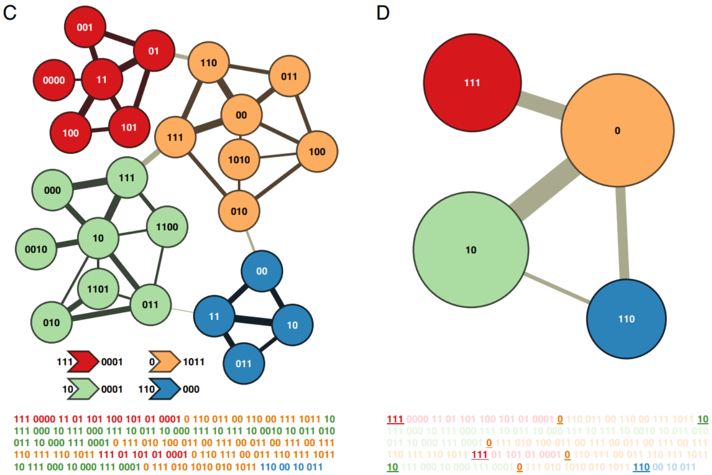
April 12, 2023

Postal Code and Random Walk





Random Walk and Coding Theory



Global Measures

For a connected graph G with n nodes,

$$\mathcal{I}(G) = -\sum_{i=1}^{k} \frac{n_i}{n} \log_2 \frac{n_i}{n}. \tag{1}$$

where n_i is the number of topologically equivalent vertices in the i-th vertex orbits of G, and k is the number of different orbits.

Local Entropy Measures

For each i, define the entropy of i in G by

$$\mathcal{I}^G(i) = -\sum_{j=1}^{n} \frac{d(i,j)}{d(i)} \log_2 \frac{d(i,j)}{d(i)}. \tag{2}$$

where $d(i,j)$ is the distance between i and j in G, and $d(i)$ is the sum of $d(i,j)$ for all j.

Parametric Graph Entropy

Given a network $G = (V, E)$ and a function f from V to positive real numbers, define

$$p(i) = \frac{f(i)}{\sum_{j=1}^{n} f(i)}. \tag{3}$$

The parametric measure is the Shannon entropy of the distribution $(p(1), p(2), \cdots, p(n))$.

Gibbs Entropy

Given a microcanonical ensemble, the Gibbs entropy is defined as

$$\Sigma = \frac{1}{n} \log N. \tag{4}$$

where N is the number of the networks in the ensemble.

The Gibbs entropy of a network ensemble is the number of bits needed to determine the code of the network generated by the ensemble.

- ◇ ensemble
- ◇ microcanonical ensemble

Shannon Entropy

For a network of n nodes, in which each pair of nodes (i, j) of weight α is created with probability $\pi_{ij}(\alpha)$. Then the probability of the canonical undirected network ensemble, which is defined by its adjacency matrix $a_{i,j}$, is defined by

$$\mathbf{\Pi} = \prod_{i<j} \pi_{i,j}(a_{i,j}),$$

for which the log-likelihood function is given by

$$\mathcal{L} = -\sum_{i<j} \log \pi_{i,j}(a_{i,j}).$$

Shannon Entropy

The entropy of a canonical ensemble is defined as

$$S = \langle \mathcal{L} \rangle_{\mathbf{\Pi}} = -\sum_{i<j} \sum_{\alpha} \pi_{i,j}(\alpha) \log \pi_{i,j}(\alpha). \tag{5}$$

For undirected network where $\alpha = 0, 1$,

$$S = -\sum_{i<j} p_{i,j} \log p_{i,j} - \sum_{i<j} (1 - p_{i,j}) \log(1 - p_{i,j}),$$

where $p_{i,j} = \pi_{i,j}(1)$ is the probability of the existence of the edge (i, j).

The Shannon entropy of a network ensemble is the number of bits needed to determine the expression of the network generated by the ensemble.

Von Neumann Entropy

Define a dense matrix as

$$\rho = L / \sum_{i,j} a_{i,j},$$

where L is the Laplacian matrix of the network with $L_{i,j} = \sum_r a_{ir}\delta_{i,j} - a_{a,j}$.

Define the von Neumman entropy of an ensemble as

$$S_{VN} = -\langle \mathrm{Tr}\rho \log \rho \rangle_{\mathbf{\Pi}}. \tag{6}$$

Structural Entropy of Models of Networks

Given a random graph model $\mathcal{M}$, let $\mathcal{S}$ be the set of all graphs of the same type (isomorphic) generated by model $\mathcal{M}$. The structural entropy $H_{\mathcal{S}}$ of $\mathcal{S}$ is the Shannon entropy of the distribution $p(G)$ for all $G \in \mathcal{S}$,

$$H_{\mathcal{S}} = -\sum_{G \in \mathcal{S}} p(G) \log p(G), \tag{7}$$

where $p(G)$ is the probability of generating a graph G by the model $\mathcal{M}$.

Quantification of Structural Information

To better understand the great challenge left by Shannon, we examine the following example.

Suppose that we are given a structured data $G = (V, E)$ which is a graph. We want to analyse G. According to Shannon's information, we can get the Shannon information of G only by the following approach:

(1) (De-structuring) To define a distribution $\mathbf{p}$ of G, such as the distribution of degrees, or distances of G.
(2) (Shannon's information) To compute the Shannon information of $\mathbf{p}$, i.e., $I = H(\mathbf{p})$.
(3) (Information of G) We obtain the information I of G.

Step (1) gives an unstructured vector $\mathbf{p}$ by removing the structure of G. This step may lose the most interesting properties of G. Step (3) gives us the Shannon information I of G, which is just a number. The question is: what properties of G can we find from the Shannon information I of G?

Therefore, Shannon information gives us only a number for every graph G. However, we can not analyse the properties of graph G from the Shannon number I.

By the same reason, none of the existing measures in Section II supports the analysis of structured data, because, all the measures are a specific form of the Shannon entropy.

The situation above becomes worse when we analyse the large-scaled networking data, and unstructured big data.

To solve the problem left by Shannon, we need a new metric of structural information that supports the analysis of graphs, networks, structured data and even unstructured big data. Here we provide such a metric.

Before introducing our metric, we describe the general problem we will solve as follows:

Given a graph G, suppose that

(i) G is a structured, but noisy data,
(ii) G is obtained from evolution,
(iii) there are rules controlling the evolution of G, and
(iv) there are random variations occurred in the evolution of G.

Our questions for analysis of G include:

(1) How to measure the amount of randomness in the evolution of G?
(2) How to extract exactly the part of G constructed by rules, excluding the random variations?
(3) Can we distinct the part of G generated by rules and the part of G perturbed by random variations?

Dynamical Complexity of Networks

1. Networks are complex. How to measure this complexity?
2. Naturally evolving networks are hard even to describe, to define or to store. All the existing notions of various entropy can be regarded as static measures of complexity.
3. The great challenge: measure the complexity of interactions, communications, operations and evolution of the real world networks.
4. Evolving in two ways: nature/society v.s. engineered.

Question: What is the dynamical complexity of networks?

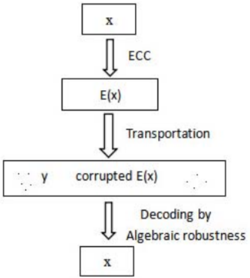
Intuition

Given a noisy or corrupted graph G, define the structural information H, the essential structure T, the true knowledge K.

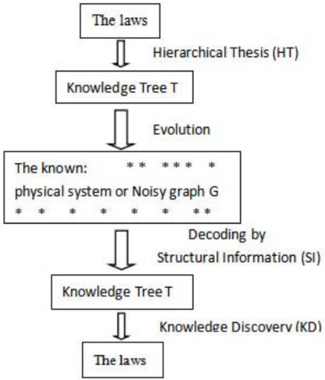
1. K is placed in T.
2. T is determined by H.
3. K consists of the rules, regulations and laws of G.
4. T is obtained from G by excluding the maximum amount of nondeterminism, uncertainty, and noise that occurred in G.

Inspiration

Decoding Error Correcting Code (ECC) : Given a string x



Decoding the Truth : For an object



Steps:

Given an object, suppose

1. Unknown laws of the object.
2. Hierarchical Thesis.
3. Physical or noisy graph.
4. Decoding by structural information.
5. Knowledge discovery.

One-Dimensional Structural Information

Definition 1 (One-Dimensional Structural Information of Connected and Undirected Graphs)

Let $G = (V, E)$ be an undirected and connected graph with n nodes and m edges. For each node $i \in (1, 2, \cdots, n)$, let d_i be the degree of node i in G, and let $p_i = \frac{d_i}{2m}$.
Then the stationary distribution of the random walk in G is described by probability verctor $\boldsymbol{p} = (p_1, p_2, \cdots, p_n)$. Define the one-demisional structural information of G or the positioning entropy of G as follows:

$$\begin{aligned} \mathcal{H}^1(G) = H(\boldsymbol{p}) &= H\left(\frac{d_1}{2m}, \cdots, \frac{d_2}{2m}\right) \\ &= -\sum_{i=1}^{n} \frac{d_i}{2m} \log_2 \frac{d_i}{2m} \end{aligned} \tag{8}$$

It measures the information required to determine the one-demisional code of the node that is accessible from random walk in G with stationary distribution.

One-Dimensional Structural Information

Definition 2 (Weighted Degree and Volume)

Given a network $G = (V, E)$, suppose that the weights assigned to edges is defined by a weight function $w : E \to \mathbb{R}^+$. Define the weighted degree of node u to be

$$d_u = \sum_{v \in N(u)} w((u, v)),$$

where $N(u)$ is the set of neighbors of u. A weighted graph has k-bounded weight if for every edge e, $w(e) \leq k$. For a subset $U \subseteq V$, define the volume of U to be

$$vol(U) = \sum_{v \in U} d_v.$$

Define the volume of G to be $vol(G) = \sum_{v \in V} d_v$.

One-Dimensional Structural Information

Definition 3 (One-Dimensional Structural Information of Weighted and Connected Networks)

Let

One-Dimensional Structural Information

Definition 4 (One-Dimensional Structural Information of Directed and Connected Graph)

Given

One-Dimensional Structural Information

Definition 5 (One-Dimensional Structural Information of Disconnected Graphs)

Given


thank you