Graph Entropy and Conditioning

Arthur Ramer, Marian Grendar, and Colette Padet

Abstract

How to perform conditioning when certain letters/outcomes are not distinguishable? Distinguishability being specified by a graph, we follow [4], [5], applying Körner's graph entropy and related information divergence on graphs to address this question.

1. Setup, graph entropy

Let a memoryless source emit letters from a finite n-element alphabet $\mathcal{X}$, with probability distribution P. We mark the pairs of letters that can be distinguished by edges of graph G on the vertex set $V(G) = \mathcal{X}$. In this setting, the relevant information theoretic entropy is graph entropy $H(G, P)$. Intorduced by Körner [2], it can be conveniently defined as [1]:

$$H(G, P) \triangleq \min_{a \in \mathsf{VP}(G), a>0} \sum_{i=1}^{n} p_i \log \frac{1}{a_i},$$

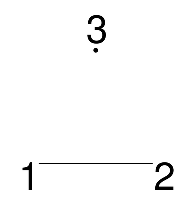
where the strictly positive vector a goes over $\mathsf{VP}(G)$, the vertex packing polytope of the graph G. Recall that the vertex packing polytope of a graph G is the convex hull of the characteristic vectors of stable sets of G.

2. Graph entropy for simple graphs

Graph numbers refer to technical report [6].

2.1 Three vertex graphs

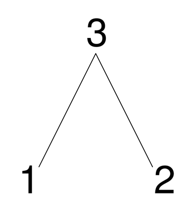
2.1.1 Graph 1



$$H(G, P) = H(p_1, p_2, p_3) - H(p_1 + p_2, p_3).$$

There, $H(\cdot)$ denotes the Shannon's entropy.

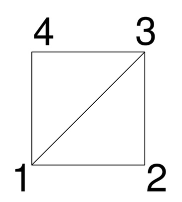
2.1.2 Graph 2



$$H(G, P) = H(p_1 + p_2, p_3).$$

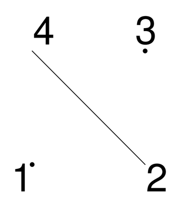
2.2 Some of four vertex graphs

2.2.1 Graph 3



$$H(G, P) = H(p_1, p_2 + p_4, p_3).$$

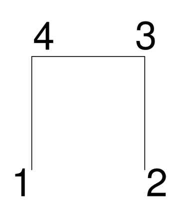
2.2.2 Graph 4



$$H(G, P) = H(p_1, p_2, p_3, p_4) - H(p_1, p_2 + p_4, p_3).$$

2.2.3 Graph 13

The smallest graph exhibiting phase transition is the following one:



If $p_1 p_2 > p_3 p_4$,

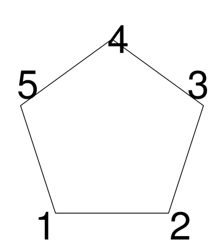
$$H(G, P) = H(p_1, p_2, p_3, p_4) - H(p_1 + p_4, p_2 + p_3).$$

If $p_1 p_2 \leq p_3 p_4$,

$$H(G, P) = H(p_1 + p_3, p_2 + p_4).$$

2.3 A five vertex graph

Graph 21.



If

$$\begin{aligned} p_1 + p_2 + p_3 &> p_4 + p_5, \\ p_1 + p_4 + p_5 &> p_2 + p_3, \\ p_1 + p_2 + p_4 &> p_3 + p_5, \\ p_1 + p_2 + p_5 &> p_3 + p_4, \\ p_3 + p_4 + p_5 &> p_1 + p_2, \end{aligned}$$

then

$$H(G, P) = H(p_1, p_2, p_3, p_4, p_5) - \log 2.$$

There are also several boundary cases. For instance, if $p_1 + p_2 + p_3 \leq p_4 + p_5$, then

$$H(G, P) = H(p_1, p_2, p_3, p_4, p_5) - H(p_1 + p_2 + p_3, p_4 + p_5).$$

3. Information divergence of graph and conditioning

Following Sanov Theorem for graphs it is natural to define *graph information divergence* $I(p \,||\, q; G) \triangleq -H(p; G) + L(q \,||\, p)$, where $L(q||p) \triangleq -\sum_{i=1}^{n} p_i \log q_i$ is L-divergence.

3.1 Conditioning via graph information divergence minimization

We propose to perform conditioning by means of minimization of graph information divergence (wrt p) under the constraints specifying which nodes are erased from the graph.

3.1.1 Conditioning 1: Graph 1

Let $p_3 = 0$. Then

$$\hat{p} = \arg \inf_{\substack{p_3=0, \\ p_1+p_2=1}} p_1 \log \frac{p_1}{q_1(p_1+p_2)} + p_2 \log \frac{p_2}{q_2(p_1+p_2)} + p_3 \log \frac{p_3}{q_3}.$$

is $\hat{p} = \left[\frac{q_1}{q_1+q_2}, \frac{q_2}{q_1+q_2}, 0\right]$.

Let $p_1 = 0$. The solution to

$$\hat{p} = \arg \inf_{\substack{p_1=0, \\ p_2+p_3=1}} p_1 \log \frac{p_1}{q_1(p_1+p_2)} + p_2 \log \frac{p_2}{q_2(p_1+p_2)} + p_3 \log \frac{p_3}{q_3}.$$

is indeterminate. If the additional restriction $p \geq q$ is imposed, then one of the coordinates of $\hat{p}$ is qual to $\min\{q_2, q_3\}$ and the other one is its complement.

3.1.2 Conditioning 3: Graph 21

Here the conditioning (with respect to $p_i = 0$, for any fixed i) is particularly simple. Let $p_1 = 0$. The minimization problem

$$\hat{p} = \arg \inf_{\substack{p_1=0, \\ p_2+p_3+p_4+p_5=0}} p_1 \log \frac{2p_1}{q_1} + \cdots + p_5 \log \frac{2p_5}{q_5}$$

has a unique solution $\hat{p} = \left[0, \frac{q_2}{q_2+q_3+q_4+q_5}, \ldots, \frac{q_5}{q_2+q_3+q_4+q_5}\right]$.

4. Summary

Few observations and conjectures:

1. For most graphs, the *functional form* of graph entropy depends on P, in the sense that it changes as P passes through a threshold conditions (c.f. Graph 13 for an illustration).
2. Graph entropy for odd $(2n+1)$-cycle with probabilities sufficiently close to uniform is of the form $H(G, P) = H(p_1, \ldots, p_{2n+1}) - \log n$.
3. It appears that graph entropy for any graph can be expressed as $+/-$ combination (linear with coefficients ± 1) of Shannon entropies over some partitions of the vertex set. The partition is probability-specific, per Item 1. We postulate that such a partition can be discovered by numerical analysis of the entropy values.
4. Conditioning via graph information divergence seems generally to behave in accordance with natural, commonsense expectations.

References

[1] Csiszár, I., Körner, J., Lovász, L, Marton, K., Simonyi, G. Entropy splitting for antiblocking corners and perfect graphs. *Combinatorica*, 10:27–40, 1990.

[2] Körner, J. Coding of an information source having ambiguous alphabet and the entropy of graphs. *Trans. 6th Prague Conference on Information Theory, 1971*, 411–425, Academia, Prague, 1973.

[3] Körner, J., Simonyi, G. Graph pairs and their entropies: modularity problems. *Combinatorica*, 20/2:227–240, 2000.

[4] Ramer, A. GraphMaxEnt. *Bayesian inference and maximum entropy methods in science and engineering*, A. Mohammad-Djafari (ed.), 179–187, AIP, Melville(NY), 2006.

[5] Ramer, A. Trust updating as belief revision. *IPMU'06 - 11th Conf. Information Processing and Management of Uncertainty*, Paris, July 2006.

[6] Ramer, A. and Grendar, M. Graph entropy for simple graphs and conditioning. UNSW Technical Report 2006.

[7] Simonyi, G. Graph Entropy: A Survey. DIMACS, 20:399–441, 1995.

5. Affiliation of the authors

Arthur Ramer is with School of Computer Science and Engineering, UNSW, Sydney 2032, Australia, e-mail: ramer@cse.unsw.edu.au. Marian Grendár, Dept. of Mathematics, FPV, UMB, Banská Bystrica, Slovakia, e-mail: marian.grendar@savba.sk. Colette Padet, Laboratoire de Thermomècanique, SUE - UTAP, Universitè de Reims Champagne-Ardenne, Facultè des Sciences - BP. Reims, France.