Phase Transitions in Artificial Intelligence

Alan M. Luu

May 8, 2017

Abstract

Artificial intelligence often involves search and computation over large networks. This essay discusses a family of closely related papers on the application of ideas from the physics of phase transitions to artificial intelligence algorithms, specifically, heuristic search and activation spreading nets. The authors of these papers build theoretical models of these network algorithms with parameters associated with general properties of the network. Additionally, they run computational simulations to show that these algorithms display divergent behavior characteristic of phase transitions for certain values of the parameters.

1 Introduction

The turn of the millenium has seen one of the greatest and most dramatic shifts in human history: the explosion of computational power and available data. Computation has already shaped daily life for billions of people, created entire industries, and revolutionized nearly every field of study; however, still looming over the horizon is perhaps one of the most ambitious goals of computation: artificial intelligence.

Many artificial intelligence algorithms often involve search and computation over networks. In recent times, modern networks can reach macroscopic sizes, with some researchers having estimated the amount of all communication over the internet to be of order 2×10^{21} bytes back in 2011 [1]. This number has increased many-fold since then, and will only grow larger, faster, with the advent of new technologies on the horizon such as the Internet-of-Things and human genome sequencing. The field of physics is no stranger to massive numbers of interacting elements; as such, the tools of physicists, designed to understand

general properties of systems with 10^{23} particles, may perhaps be applied to these networks.

Statistical mechanics teaches us that in certain situations, general macroscopic properties of large, complex systems may be understood even if one knows little about the exact microscopic state. Additionally, these systems may exhibit sudden changes in behavior in response to small changes in parameter that cannot be anticipated from observing similar but small-scale systems. In a series of papers, Tad Hogg et al. attempt to find out whether these principles also hold true for algorithms over graphs consisting of macroscopic numbers of nodes and edges [2, 3, 4, 5].

2 Interacting Processes

We will first consider a model, created by Huberman and Hog, of a large number of processes that interact locally with each other. This model can be represented by a set of trees, where the root represents an initial process and the branches represent other processes affected by the initial process. These trees have the form of nodes connected by edges. The quantity of interest is the size of the tree, and the parameter of interest is z, the average branching ratio of the tree. We will show that there are two qualitatively different phases of this system depending on the value of z.

Let N(u) be the average number of nodes a distance u from the root. This distance measures the number of branching processes that occurred to create the node in question. Therefore, the distance of the root is 0, the distance of the root's children is 1, and so on. Since a tree has a unique path to every node, and each node has on average z children,

$$N(u) = z^u$$

In order to find the average number of nodes A in a tree, we take

$$A = \sum_{u=0}^{\infty} N(u) = \frac{1}{1-z}$$

for $z < 1$. For $z \geq 1$, A diverges to infinity. The authors generated a graph of A(z) for an infinite system, which is shown in figure 1. However, for a finite number of nodes, just as in finite physical systems, the divergence is

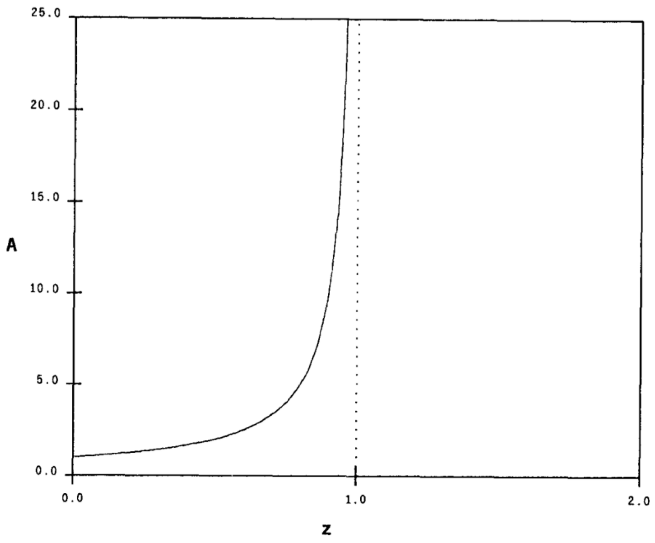

Figure 1: The average tree size A as a function of z

smoothed out. In real physical systems, this occures because the correlation length cannot exceed the size of the physical finite system. Likewise, in the model of interacting processes, the number of nodes is finite depending on the maximum depth of the tree.

$$A = \frac{(1-z)^d}{1-z}$$

A plot of A(z, d) generated by the authors is given in figure 2.

3 Heuristic Search

A concrete application of the interacting process model is the heuristic search, where one tries to find a specific node in a tree starting from the root using a guiding heuristic that is able to eliminate some implausible branches of the tree. Suppose we have a tree with depth d and branching ratio b. Next, assume that we have a heuristic applied at each node that removes each incorrect branch with probability $1 - p$. We can then construct a new tree with effective branching ratio $z = bp$. A pictoral representation of this process is given in figure 3. To find the total number of nodes explored, we note that, except for the heuristic, the search proceeds randomly. Therefore, on average, our search algorithm will

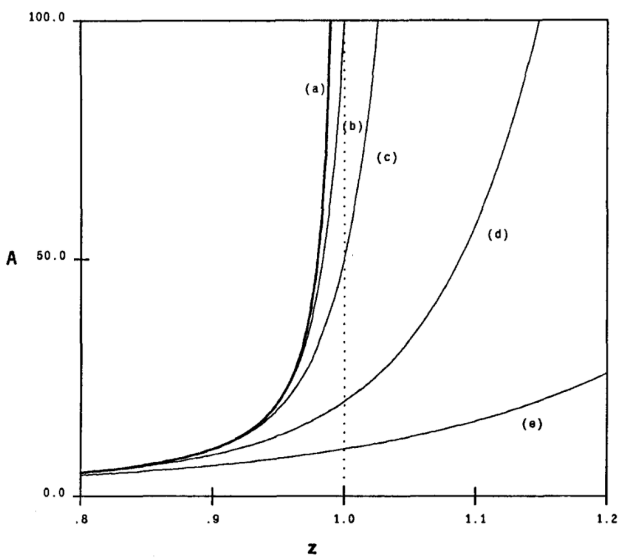

Figure 2: The average size of trees with depth d. $A = \frac{(1-z)^d}{1-z}$. The different curves show A(z, d) with varying values of d. (a) represents the infinite tree, (b) d = 100, (c) d = 50, (d) d = 20, and (e) d = 10.

try half of the incorrect branches before proceeding down the correct branch. Let $A(d)$ be the average size of a tree with depth d. From the root, the average size of each of the implausible branches is $A(d-1)$, and there are $(b-1)p$ incorrect branches. After incorrectly trying on average $\frac{1}{2}(b-1)p$ branches each of size $A(d-1)$, the algorithm then proceeds one step down the correct branch, giving a total of $1 + \frac{1}{2}(b-1)pA(d-1)$ nodes up to that point. Afterwards, the algorithm is faced with the same exact problem as we had initially, but this time with a tree of depth $d-2$, and so on. The total number of nodes searched is

$$N = d + \frac{1}{2}(b-1)p\sum_{s=1}^{d-1} A(s)$$

for $z < 1$, and with $n = \frac{N}{d}$, this gives

$$n = \frac{2-z-p}{2(1-z)}.$$

When $z \geq 1$, n becomes infinite, which signals the transition between linear

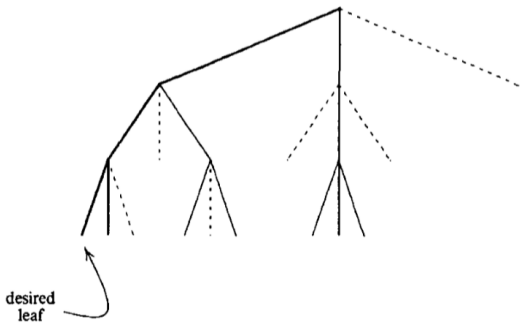

Figure 3:

time search and exponential time search. Therefore, near the critical point $z = 1$, even a small change in the effectiveness of the heuristic can create a large shift in the overall character of the search algorithm.

4 Activation Spreading Networks

Activation spreading networks are algorithms in which signals, usually numeric quantities, are passed between connected nodes. The nodes themselves represent states or items; for example, in a neural network for image recognition, a node can represent the intensity of a pixel or a transformation on pixel intensities. The characteristic features of the nodes are the number of connections to other nodes as well as the weight of each connection. The weights between nodes usually signify the strength or importance of the interaction. At each time step, a number of nodes are "activated" by an external source. That is, all nodes have an associated number that is initialized to zero, and some nodes have their associated number incremented a finite amount at every time step. This number then decays downward a set amount each time step, and each of the nearby neighbors increments its own current value according to the weight of the edge connected to the activated node. As the system evolves over time, a steady equilibrium state may emerge, and qualitative features may be understood with respect to the general parameters of the system.

To formalize this system, suppose we have a graph with undirected, weighted edges. Let n be the number of nodes, μ be the average number of edges per node, γ be the relaxation rate at which the activity of a node dies down, and α be the amount of activation flowing into the node from all neighbers per unit time.

Additionally, let $\mathbf{A}$(N) be a vector whose ith component denotes the activity of node i at the Nth time step, let $\mathbf{C}$(A) be a vector whose ith component denotes the activation of the ith node due to an external source, and let $\mathbf{R}_{ij}$ be a matrix that denotes the weight of the edge between the ith and jth nodes. The time evolution of the net is given by

$$\mathbf{A}(N) = C(N) + (1-\gamma)\mathbf{A}(N-1) + \alpha\mathbf{R}\ \mathbf{A}(N-1)$$

For small values of $\frac{\alpha}{\lambda}$, the activity near externally driven nodes will propagate outwards, but relax down to an asymptotic value, localizing the influence of the external source in time and space. This phase corresponds to section I in figure 4. However, as $\frac{\alpha}{\gamma}$ is increased, the relaxation rate decreases. If $\frac{\alpha}{\gamma}$ passes through the critical point in which amplification from neighbors wins over relaxation, the cluster of connected nodes around the external source will activate each other indefinitely, driving the total activity higher and higher. However, although the relaxation time diverges, for small μ, the activity is still localized in space around the external source. The authors call the range of influence of the external source the horizon, and make the distinction between horizons in space and horizons in time. The phase in which the time horizon diverges while the space horizon is finite is denoted as phase II in figure 4. As μ is increased, the spacial horizon increases until it diverges to infinity at $\mu \geq 1$. This phase, denoted phase III in figure 4, displays divergent behavior both in time and in space.

The authors ran 4 simulations to experimentally verify the phase transition behavior of activation spreading networks. Each experiment consisted of 25 spreading activation trials over a graph containing 100 nodes. A single node was chosen to be the external driving source, and the activity of the entire network, defined as the sum of the activities of all nodes, was recorded throughout the simulation. Two parameters were introduced by the authors in order to determine the point at which the network settled into its fixed point. When the total activity changed by less than the settling threshold e, set to 0.001, between two consecutive iterations, the network was considered to have settled. Additionally, when the network failed to settle, an upper bound B on the total activity was set, and the simulation was stopped when this bound was reached.

Experiments 1 and 2 correspond to varying the parameter μ from 0.4 to 2.0 with $\alpha = 0.4$ and $\gamma = 0.6$. This corresponds to taking the path on the phase diagram from phase II to phase III. In figure 5, the fraction of nodes activated as well as the standard deviation of activated nodes is plotted with respect to

the number of links. We see that divergent behavior, somewhat rounded due to the finite size of the system, starts to set in past the predicted critical point for the infinite system (denoted by the arrow on the graph). Additionally, near the critical point, the fluctuations become larger, which is also expected of systems near criticality.

Experiments 3 and 4 correspond to varying $\frac{\alpha}{\gamma}$ according to the paths shown on figure 4. The experimental results are shown in figure 6. The solid curve corresponds to $\mu = 0.4$ and the dashed curve corresponds to $\mu = 1.4$. In graph A, the total activity was measured, and in graph B, the number of iterations required to settle (for $\frac{\alpha}{\gamma} < 1$) or reach the upper bound $\frac{\alpha}{\gamma} > 1$ was measured. Again, we see the divergent behavior as predicted by theory.

5 Conclusion

The authors showed theoretically and experimentally how phase transitions can occur in two important artificial intelligence algorithms. However, this is not too surprising, as most of these results seem to stem from the fact that these models are similar to those of percolation models [6], Ising models [7], and random graphs [8], which are already known to display phase transitions. Perhaps the most novel direction from this work is the interaplay between temporal and spacial horizons. I would have liked to see more experiments run with the parameters in the vicinity of the "triple point" where phases I, II and III meet. The results of such a study may be useful for systems with dynamical as well as topological considerations such as epidemics and biological networks.

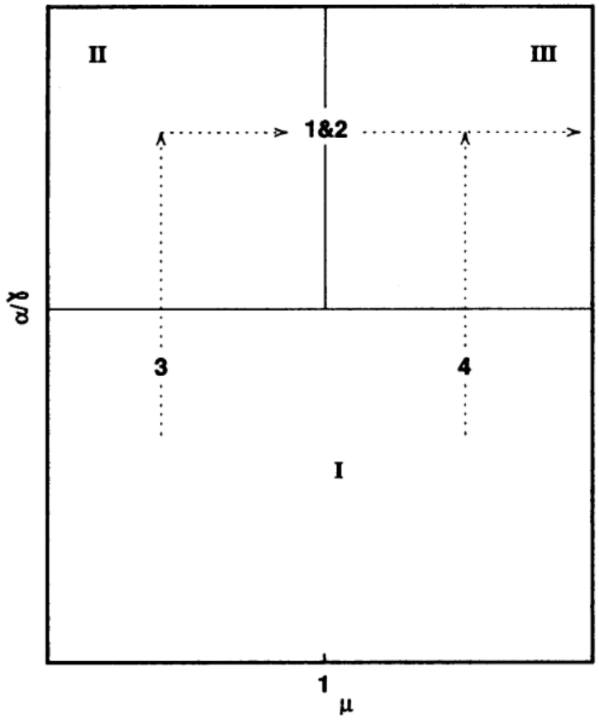

Figure 4: The phase diagram for the spreading activation net of an ideal infinite system. The horizontal line corresponds to the trajectory through parameter space of experiments 1 and 2, while the vertical lines correspond to experiments 3 and 4.

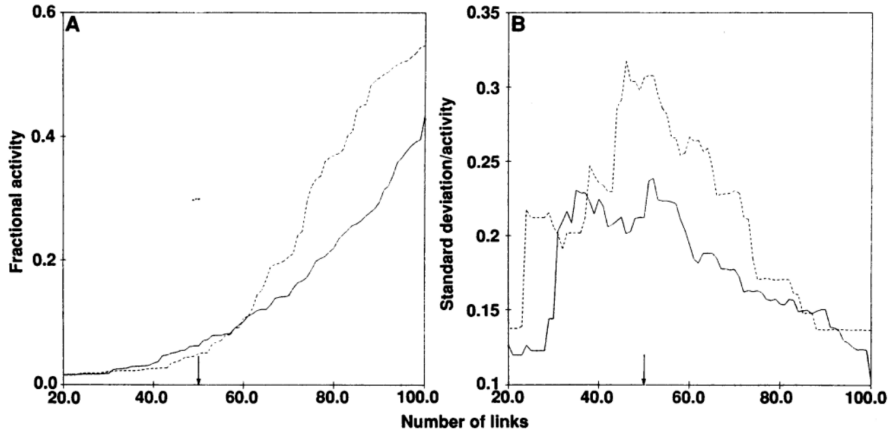

Figure 5: The network used in both experiments has 100 nodes, and the number of links was changed from 20 to 100, with $\alpha = 0.4$ and $\gamma = 0.6$. **A** shows the fraction of nodes active and **B** shows the standard deviation of the activity. The solid curve corresponds to B = 10 and the dashed curve corresponds to B = 1000

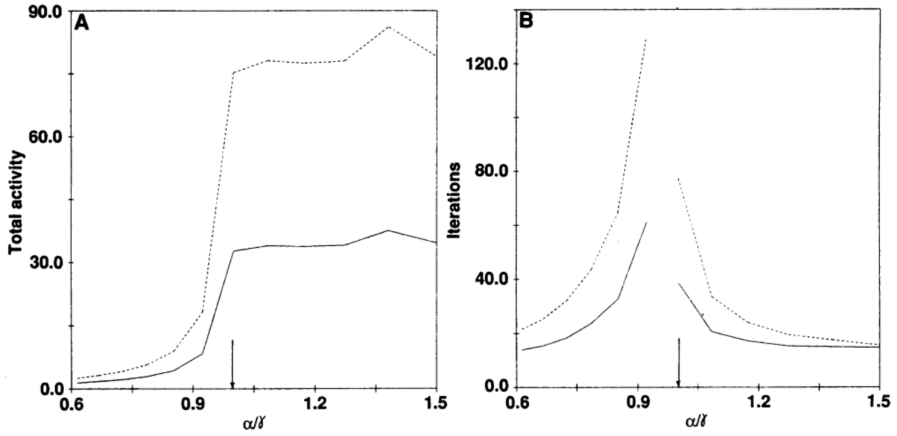

Figure 6: Graph A shows the total activity of the network. The solid line corresponds to $\mu = 0.4$ and the dashed line corresponds to $\mu = 1.4$. Graph B shows the completion time. To the left of the predicted phase transition, the graph shows settling time, and to the right of the phase transition, the time to reach the upper bound.

6 References

[1] Martin Hilbert1, Priscila Lopez, The Worlds Technological Capacity to Store, Communicate, and Compute Information. *Science* 01 Apr 2011: Vol. 332, Issue 6025, pp. 60-65

[2] Bernardo A. Huberman and Tad Hogg, Phase Transitions in Artificial Intelligence Systems. *Artificial Intelligence*, Volume 33, Issue 2, 155-171. (1987)

[3] Jeff Shrager *et al.,* Observation of Phase Transitions in Spreading Activation Networks. *Science* 236 (4805), 1092-1094. (1987)

[4] Tad Hogg *et al.,* Phase Transitions and the Search Problem. *Artificial Intelligence*, Volume 81 1-15. (1996)

[5] Tad Hogg, Refining the Phase Transition in Combinatorial Search. *Artificial Intelligence*, Volume 81 127-154. (1996)

[6] Broadbent, Simon; Hammersley, John (1957), Percolation processes I. Crystals and mazes. Proceedings of the Cambridge Philosophical Society, 53: 629641

[7] Onsager, Lars (1944), "Crystal statistics. I. A two-dimensional model with an order-disorder transition", Physical Review, Series II, 65 (34): 117149

[8] Erdos, P.; Rnyi, A. (1959). "On Random Graphs. I". Publicationes Mathematicae. 6: 290297.