Psych204 Report: Bayesian Network Properties With Complex Network Topologies

Howon Lee (howonlee@stanford.edu)

Abstract

Bayesian networks are a formalism for a subset of all probabilistic programs. We investigate some large bayesian networks already created as cognitive models, mostly MUNIN, a probabilistic model for the use of electromyography in diagnosis, in order to find whether the recurrent patterns of other complex networks, such as a skewed degree distribution and certain graph properties, appear. We have mixed results, and discuss cognitive implications.

Keywords: complex networks; bayesian networks

Introduction

The initial impetus for this study is an observation: the observed degree distribution of a variety of hand-crafted Bayesian networks is incompatible with a thin-tailed distribution of the degrees. That is, typical thin-tailed distributions like a Gaussian or a Poisson do not suffice to represent the degree distribution of the network. Of course, these networks are used for statistical modelling, not cognitive modelling, but it seems to be an intriguing property of the formalism, and it might extend to cognitive modelling because the networks themselves were hand-crafted to be expert systems and therefore try to represent cognitive aspects of domain knowledge.

This fat-tailled property is very related to a host of other properties of complex networks. Two properties of cognitive consequence are the robustness to random insult but fragility to systematic attack on the highest degree nodes, and the propensity towards percolation matching.

When we say that the Bayes network is robust or fragile, we mean one of two meanings: the weak connectedness of the underlying graph and of the constancy of the probability distribution which the Bayes network encodes. We test both meanings and discuss cognitive implications.

We also test the propensity of the nodes towards percolation graph matching, which is an approximate method of subgraph isomorphism. Although subgraph isomorphism is NP-complete, real world complex networks, notably social networks, are vulnerable to an approximation which depends upon a starting seed determined by nonlinear optimization and then to percolation throughout the graph via inductive means. We find that this network is not amenable to such a matching to a generated null complex network, but are not surprised.

Data

Although we found many Bayesian networks available as cognitive models, we found only a few which were big enough to be of interest. The one which we will concentrate on in this project is called MUNIN, and is for statistical inferential use, in the diagnosis of electromyographical data (Andreassen, Woldbye, Falck, & Andersen, 1987). This is because it is the largest network found, of 1041 nodes. Of course to *make* or to *determine* such a network from evidence is not within the scope of the class, which is why a pre-existing network was used.

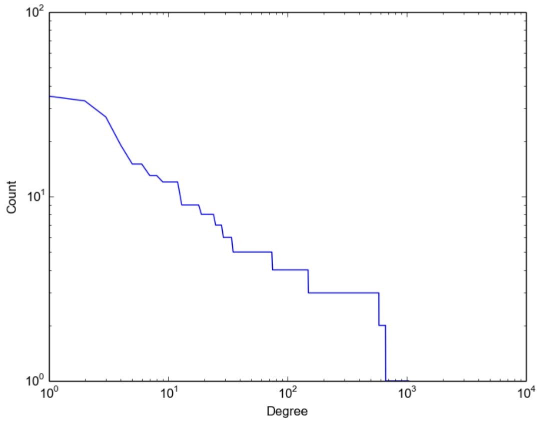

Figure 1: Outdegree of MUNIN, construed as a directed graph

Table 1: Average path length for different Bayes nets

Munin	Pathfinder	Diabetes	Pigs
8.176	2.000	7.518	6.551

Initial investigation of the topological structure of the network uncovered a fat tailed degree distribution in the outdegree, although there is not such a fat tail in the indegree. By ”fat tailed”, we mean that there is probability mass where there wouldn’t be if the degree distribution was a one-tailed Gaussian distribution. Although the outdegree structure is not mentioned in the original MUNIN paper, the indegree structure is, because they wished to reduce the parameter space to avoid having to fill in large amounts of parameters from the sparse data that they had.

It is notable that, although the smaller networks we investigated had systematic patterns of topology, they had a different outdegree distribution often, although they still had fat tails on that degree distribution.

Methods and Results

The two different possible meanings of insult and attack in the Bayes net were investigated by attacking and insulting the network in two different ways.

The first way was to investigate the weak connectedness

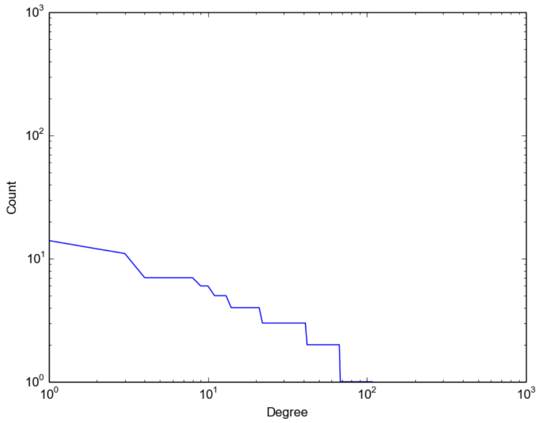

Figure 2: Outdegree of Pathfinder, construed as a directed graph

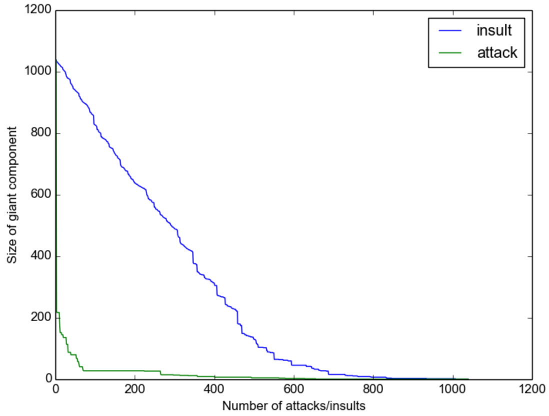

Figure 3: Resilience testing of nodes under random insult and focused attack

property of the network and did not deal with the conditional probability tables at all, just the removal of nodes from the graph via random insult or focused attack. The connectedness property was measured by measuring the size of the "giant component", which is a weak percolation throughout the graph which occurs in many random graphs and occurs in this graph, according to the method of (Albert, Jeong, & Barabási, 2000). As can be seen in 3, there is extremely different behavior towards random insult and focused attack, where focused attack fragments the network and makes its largest component very small almost immediately, whereas random insult does not fragment the graph nearly as quickly.

The second way, to investigate the behavior of the probability distribution which the Bayes net encodes, deconditionalized the probability distributions which the attacked or insulted node encoded, to a default, uniform prior, while leaving the node itself in the network. Then, the Bayes

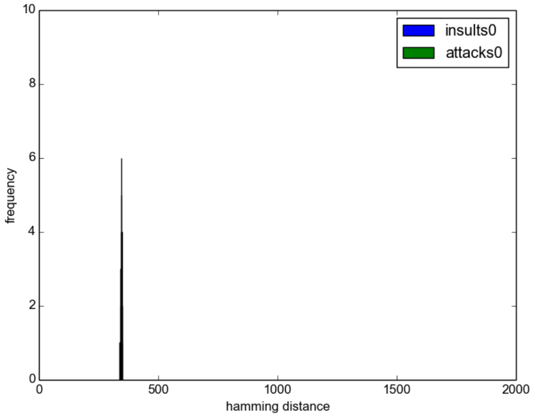

Figure 4: Resilience testing of produced state under random insult and focused attack, 0 nodes removed

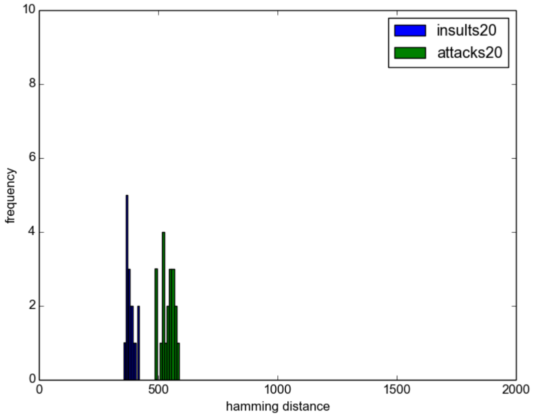

Figure 5: Resilience testing of produced state under random insult and focused attack, 20 nodes removed

net was sampled without conditioned variables by normal sampling (because there were no conditioned variables, no rejection sampling was needed, and MCMC was not used because samples are correlated from MCMC).

The natural way to see if a distribution Q faithfully represents another distribution P or not is Kullback-Leibler divergence(Kullback & Leibler, 1951), defined as:

$$D_{KL}(P||Q) = \sum_i P(i) \ln \frac{P(i)}{Q(i)}$$

where the divergence is greater if Q is a worse simulation of P.

However, KL-divergence was not used, because a graph was misread. Although we saw that the indegree was constrained to be very small for all nodes in the network, which meant that the marginalizing computation for computing KL-divergence was in actuality easy, we misread a graph, think-

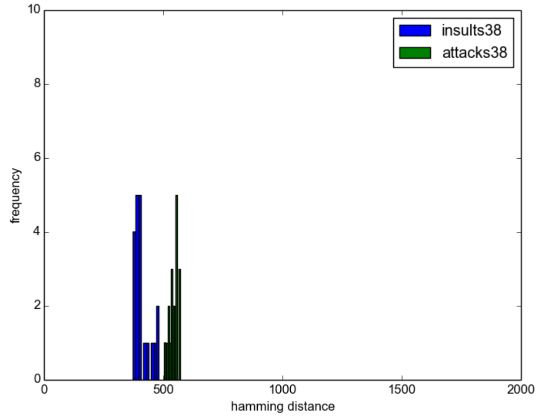

Figure 6: Resilience testing of produced state under random insult and focused attack, 38 nodes removed

ing it said that the typical indegree of nodes that had indegrees was on the order of 10^2, which would make marginalization impossible.

Therefore, an ad hoc method was used instead. The MAP estimate was taken for each random variable in the Bayes net from the samples, where 5000 samples were taken from the null Bayes net (MUNIN without any attack or insult) and 5000 samples were taken from the attacked or insulted Bayes net. A divergence of the two Bayes nets was taken to be the Hamming distance of the two MAP estimates, repeated 20 times, and a histogram of the Hamming distances plotted in 5.

Although the results are not null (5), and it is clearly seen that there is more difference between the null Bayes net and the changed Bayes net in the attacked Bayes net, as opposed to the randomly insulted one, these results are heavily clouded by the fact that this MAP estimate method is not found in the literature in actuality and vulnerable to all the criticisms of the MAP estimate in the literature(Sorenson, 1980). What can be said is that the conjecture that there is a differential effect of focused attack versus random insult in the probability distribution encoded by the Bayes net has not been falsified.

The amenability of the network to percolation graph matching was tested. A Random Typing Graph (Akoglu & Faloutsos, 2009) of 1041 nodes was generated and a seed of 30 nodes fitted by inspection, as per (Narayanan & Shmatikov, 2009), and then percolation graph matching was run to match the generated graph with MUNIN. With percolation graph matching, there is usually a phase transition between matching nearly no nodes and matching nearly all the nodes (Yartseva & Grossglauser, 2013), and in 20 tries no more than 40 nodes were ever matched, whereas in social network de-anonymization it is routine that 70, 80% of the nodes are matched (Narayanan & Shmatikov, 2009).

Discussion

Herbert Simon tells a parable about the essential nature of hierarchy in any evolutionary process which is to make complex systems. (Simon, 1996) Two watchmakers are to make watches compoesed of 1000 parts, but are interrupted by a stochastic process which completely destroys any partial work. One watchmaker makes those 1000 parts by constructing sub-parts of 100 parts each which are durable against interruption, which have sub-assemblies of 10 parts each, which are themselves durable against interruption: the other must construct the whole watch, 1000 pieces, without interruption. With fairly conservative estimates of the chance of interruption, this gives a radical advantage to the watchmaker who uses hierarchy: Simon estimates a 4000-fold advantage. Moreover, Simon later applies this to abstract problem solving in the form of constraint satisfaction problems, which are highly related to Bayesian networks and probabilistic programming in general.

An analogy suggests itself, to the sub-concepts embodied in a probabilistic program, or to the sub-concepts embodied in a Bayesian network model, and to the sub-components embedded in the watch by the intelligent watchmaker. But the structure of the *realized* Bayesian model in the MUNIN program is not arrayed in such an evenly hierarchical fashion, though it is hierarchical. Instead, there exist nodes of extremely high outdegree, which tend to affect many other nodes in the system, nearly of the order of the number of nodes: this seems to be somewhat analogous to the intelligent watchmaker creating a sub-part of 800 parts, with the caveat that this phenomenon only happens in the out-degree, so the analogy may not necessarily apply.

Clearly, this phenomenon deserves further study, with more rigorously created extremely large cognitive models. Part of the argument of this project is that one interesting path of further study on this phenomenon might be to investigate the differential effects of random failure of program components versus systematic attack.

The difference between the outdegree and the indegree of the large models is interesting, and may have cognitive implications: it is noted that there was a conscious attempt in the original building of the model to reduce the maximum indegree, to avoid the curse of dimensionality for each conditional probability table: it might be wondered to see if such a computational difference, and if such a difficulty in marginalization, holds in larger cognitive models in humans also. It must be noted that this is not a novel question, however(Goodman & Tenenbaum, 2015).

The failure of graph matching is interesting but not unexpected, because the many null models for complex networks (Kronecker graphs(Leskovec, Chakrabarti, Kleinberg, & Faloutsos, 2005), multifractal graphs(Palla, Lovász, & Vicsek, 2010), and the one we attempted, the random typing graph(Akoglu & Faloutsos, 2009)) also attempt to account for observed fat tails in the indegree distribution of many complex networks, such as social networks and WWW graphs, whereas the indegree distribution is not distributed in this way in these graphs. It remains to see if a more abstract domain-free model can be constructed and fitted in some way for this sort of Bayes network, obeying the observed regularities.

Acknowledgments

Thanks to Prof. Goodman and all the course staff for the course.

References

Akoglu, L., & Faloutsos, C. (2009). Rtg: a recursive realistic graph generator using random typing. *Data Mining and Knowledge Discovery*, *19*(2), 194–209.

Albert, R., Jeong, H., & Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, *406*(6794), 378–382.

Andreassen, S., Woldbye, M., Falck, B., & Andersen, S. K. (1987). Munin: A causal probabilistic network for interpretation of electromyographic findings. In *Proceedings of the 10th international joint conference on artificial intelligence-volume 1* (pp. 366–372).

Goodman, N., & Tenenbaum, J. (2015). *Probabilistic models of cognition*.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 79–86.

Leskovec, J., Chakrabarti, D., Kleinberg, J., & Faloutsos, C. (2005). Realistic, mathematically tractable graph generation and evolution, using kronecker multiplication. In *Knowledge discovery in databases: Pkdd 2005* (pp. 133–145). Springer.

Narayanan, A., & Shmatikov, V. (2009). De-anonymizing social networks. In *Security and privacy, 2009 30th ieee symposium on* (pp. 173–187).

Palla, G., Lovász, L., & Vicsek, T. (2010). Multifractal network generator. *Proceedings of the National Academy of Sciences*, *107*(17), 7640–7645.

Simon, H. A. (1996). *The sciences of the artificial* (Vol. 136). MIT press.

Sorenson, H. W. (1980). *Parameter estimation: principles and problems* (Vol. 9). M. Dekker.

Yartseva, L., & Grossglauser, M. (2013). On the performance of percolation graph matching. In *Proceedings of the first acm conference on online social networks* (pp. 119–130).