Is Holism A Problem For Inductive Inference? A Computational Analysis

Maxwell A. Bertolero (bertolero@berkeley.edu)
Thomas L. Griffiths (tom_griffiths@berkeley.edu)
Department of Psychology & Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA 94720 USA

Abstract

We investigate whether holism presents a problem for inductive inference by examining the relationship between the size of a Bayesian network that represents human conceptual knowledge and the computational complexity of probabilistic inference in that network. We find that, despite prior claims, holism may not be a problem for inductive inference, as computational cost does not increase exponentially as the network grows. While the network we analyze is holistic, it has a modular organization and grows in a way that potentially makes efficient inductive inference possible.

Keywords: holism; inductive inference; Fodor; Bayesian networks; graph theory; complex networks; ConceptNet

Introduction

Fodor (1983) claims that, while processes like vision can function adaptively almost purely based on sensory input from the eyes, central cognitive processes like decision-making only function adaptively if they integrate every bit of relevant information (i.e., if *every* variable is connected). Fodor argues that this holism—that everything we experience is potentially relevant to every conclusion we might want to draw—makes inductive inference intractable. This argument revisits Hume's (1739) observation that there are no rules of induction—one cannot know *a priori* what variables to evaluate for evidence of a belief, as every variable is potentially relevant to the belief in question. This seems to make inductive inference intractable, as one must do an exhaustive analysis of every connected variable during inductive inference.

If the intractability of inductive inference was not sufficiently concerning by itself, Fodor (2001) raises the stakes by arguing that holism may also make cognitive science a futile effort. If central cognitive processes involve inductive inference, then they are potentially holistic. Science carves nature at its joints, dividing and conquering. However, holistic systems have no joints, so Fodor argues that scientists should give up studying central cognitive processes, like decision making, where holism is an issue. He recommends focusing on non-holistic processes like vision.

Concerns about the impact of holism are based on the idea that all human concepts are connected (more precisely, that any one proposition involving particular concepts could bear on the confirmation of any other proposition, even ones involving different concepts). However, research in statistical machine learning has shown that variables being connected is not intrinsically a problem for inductive inference. Work on Bayesian networks shows that the complexity of probabilistic inference increases as a (potentially exponential) function of the size of the largest "clique" (i.e., fully connected set of nodes) in the graph of dependencies between variables (Koller & Friedman, 2009). Thus, the connectivity structure of our concepts might not render inductive inference intractable if the size of the largest clique only increases slowly as the network grows.

Here, we examine whether the connectivity structure of human concepts supports efficient probabilistic inference, providing a potential response to concerns about holism. We analyze a large network of concepts intended to capture human common-sense knowledge, using tools from graph theory and results concerning the computational complexity of inference in Bayesian networks. By examining the rate at which the largest clique grows—and the factors that contribute to the growth—we determine whether probabilistic inference could be conducted efficiently over human concepts.

Bayesian Networks and Holism

A graph is a mathematical model that represents a network. This model represents the network's entities as nodes, and connections between entities as edges that connect the nodes (see Figure 1). A Bayesian network is probabilistic graphical model that represents a set of variables as nodes and the dependencies between those variables as edges in a graph (Pearl, 1988). The graph structure can be used to support probabilistic inference, calculating the probability distribution over the variables given the observed values of a subset of those variables.

Concerns about holism are based on the fact that all human concepts seem potentially related. If we imagine these concepts being arranged as the nodes in a graph, with edges indicating relationships, then we might be concerned about holism if the graph is connected—if we can find a sequence of edges that links any pair of nodes. But the fundamental concern is about the tractability of inductive inference—whether it is possible to make inductive inferences under these circumstances. In computer science, problems are considered intractable if the amount of time taken to solve the problem increases exponentially in the size of the problem. This establishes a basic criterion for determining whether holism is a challenge for inductive inference: determining whether the time taken to perform probabilistic inference in a Bayesian network that captures the dependency structure of human concepts increases exponentially as the network grows.

Probabilistic inference in Bayesian networks is *#P*-hard, and the (worst-case) computational complexity of known exact (e.g. Koller & Friedman, 2009) and approximate algorithms (e.g., Koller, Lerner, & Angelov, 1999; Murphy & Weiss, 2001) for solving this problem is exponential with the size of the largest clique in the moralized and triangulated version of the graph. In other words, the size of the largest

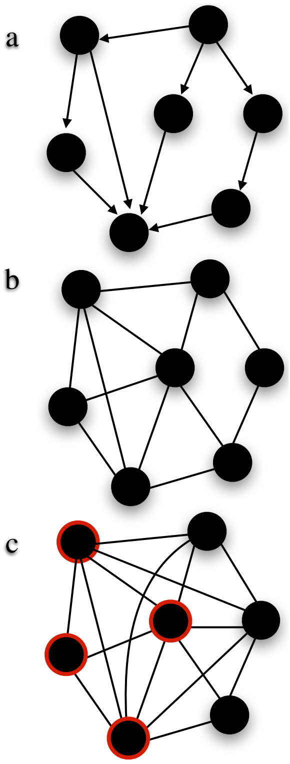

Figure 1: Graphs. (a) A directed graph, with nodes indicating concepts and edges indicating relationships. (b) The moralized graph. (c) The triangulated graph, largest clique nodes outlined in red.

clique determines the computational complexity of inference across the entire network. Maximum clique size is simply the size of the largest set of fully connected nodes, where every node has an edge to every other node in the set. When we moralize the graph, we connect any two nodes that have edges that point to a common node, and then we drop the directionality of edges. When we triangulate the graph, we make it "chordal," in that there is always an edge between nodes in loops greater than three. Thus, there are no chordless loops (loops without edges across nodes) greater than three in the graph. Figure 1 shows the progression from a directed graph to a moralized graph to a triangulated graph.

Using these results, we can formally define the criterion for holism to present an obstacle to inductive inference. Measuring clique size as the graph grows makes it possible to extrapolate about the computational complexity of probabilistic inference for graphs approaching the scale of human knowledge: if the largest clique grows near linearly as nodes are added to the graph, then holism is a real concern; if it grows sub-linearly, then connections between facts might not pose a real challenge for inductive inference, as the time required to calculate relevant probabilities would increase subexponentially. In the remainder of the paper we explore which of these regimes characterizes the properties of a graph based on human concepts.

Methods

Constructing a Graph of Human Concepts

To execute our analyses, we needed a network that represents human conceptual knowledge. We chose the ConceptNet database, which contains 280,000 facts, such as "eating is motivated by survival," contributed by a large online community (Speer & Havasi, 2012). This database was made to support common sense reasoning in artificial intelligence, but it can also potentially provide clues about aspects of cognition and the connectivity structure of our concepts.

We constructed a graph from the ConceptNet database. We took concepts to be nodes, and relationships between concepts to be edges. While a node is clearly a different notion of a concept than is assumed by Fodor, it is a working assumption that allows us to begin a quantitative exploration of these issues. There are 27 different types of relationships in ConceptNet; we treated all relationships as equal. For example, "eating" and "survival" are nodes, and "is motivated by" is an edge from "eating" to "survival". ConceptNet relations are judged by an online community, which allowed us to only use the highly rated relationships between concepts. This results in a directed and unweighted graph with 25,561 nodes and 33,768 edges that represents the network of human concepts.

We reduced the graph in this way for two reasons: accuracy and computational limits. ConceptNet contains many relationships that simply are not valid, so we wanted to make use of the ratings that the online community gave the relationships between concepts. Moreover, some measurements, like triangulation and clique finding, become highly computationally expensive on networks much larger than 10,000 nodes.

Growth Models

We used two different methods to simulate the growth of knowledge, allowing us to examine the properties of the resulting graph as it increased in size. One method randomly added nodes, the other method randomly added edges. We used these two different growth methods as they reflect different assumptions about learning. Both methods keep the graph connected during growth.

Node Sampling In the *node sampling* growth method, we randomly select a node to initialize the graph. Next, we randomly select one edge that connects to this first initial node, and add the node at the other end of the edge to the graph. The graph is now initialized. After this, we randomly select a node that already exists in the graph, and then randomly select one of that node's edges to find a new node that is not in the graph yet. We add the new node that was found at the end of the randomly selected node's edge, along with any edges the new node has to nodes that exist in the graph. Thus, at each step, one node is added, and at least one edge is added. This growth method reflects a learning model in which we learn a concept, and we acquire all its relations to concepts we already know.

Edge Sampling In the *edge sampling* growth method, we randomly select an edge to initialize the graph. Both nodes get added to the graph. The graph is now initialized. After this, we randomly sample from all the edges that connect to nodes that exist in the graph. If an edge is picked that connects two nodes that already exist in the graph, only the edge is added. If an edge is picked that is between a node that exists in the graph and a node that has not been added to the graph yet, both the new edge and the node that does not exist in the graph yet are added. Thus, at each step, one edge is added, and either zero or one node is added. This growth method reflects a learning model in which we sequentially learn relationships between concepts, acquiring concepts when we learn a relation from a known concept to an unknown concept.

Graph-Theoretic Analyses

Clique Size We measured the largest clique size in a moralized and triangulated version of the graph. As described above, moralizing the graph first creates edges between nodes with edges that direct to the same target node, and then drops the directionality of edges, resulting in an undirected graph. Triangulating the graph ensures that there are no loops greater than three. Triangulation can be performed by selecting an ordering of the nodes, then proceeding through these nodes in turn, adding connections between all nodes to which they are connected. The ordering influences the size of the resulting cliques, and finding the ordering that minimizes the size of the largest clique is *NP*-hard. Consequently, we used the heuristic of starting with the nodes with the fewest edges, which is standard in the Bayesian network literature (Koller & Friedman, 2009).

Degree The degree of a node is the number of edges that connect to it. We were primarily interested in degree measurements of the graph as it grows, but we also measured the degree distribution of the entire graph. Moreover, we wanted to know when higher degree nodes are added to the graph. Thus, when a node is added, we recorded its degree in the full ConceptNet graph.

Modularity The graph-theoretic notion of modularity measures how decomposable a network is. Many different networks, from metabolic networks (Jeong, Tombor, Albert, Oltvai, & Barabasi, 2000; Wagner & Fell, 2001) to the world wide web (Faloutsos, Faloutsos, & Faloutsos, 1999; Albert, Jeong, & Barabasi, 1999; Broder et al., 2000), exhibit a modular structure, which means that the network's nodes can be grouped into smaller communities such that within-community edges are denser than between-community edges. Modularity quantifies the ability of the network to decompose into separable communities, and is an essential property found in many complex networks that allows the network to easily evolve, develop, and engage in flexible and dynamic behaviors (Meunier, Lambiotte, & Bullmore, 2010).

We measured modularity using Newman's Q (Newman & Girvan, 2004), which is the fraction of the edges in the network that connect nodes from within the same (automatically-identified, as described below) communities minus the expected value of the same quantity in a network with the same community divisions but random connections between the nodes. If the fraction of within-community edges to between-community edges is no better than random, we will get a Q value of 0. The Q value increases as that fraction becomes greater than random. Q values approaching 1, which is the maximum, indicate a very modular structure. Modular networks identified in the previous research mentioned above exhibit Q values from 0.3 to 0.7.

Community Detection For many networks, including our conceptual network, the partition of the network into communities is not known. Thus, we extract the communities of the graph by using Louvain heuristics (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) until modularity (Q) is maximized. First, we assign a different community to each node of the network, so the initial partition contains as many communities as nodes. We then iterate through two steps. Step 1: for each node x, we consider the neighbors y of x, and we evaluate the modularity score gain of removing x from its community and by placing it in the community of y. The node x is then placed in the community that maximizes the gain in modularity, but only if this gain is positive. If no positive gain is possible, x stays in its original community. This process is applied until modularity is maximized, such that no nodes can be moved to increase modularity. Step 2: each community is treated as a node, and Step 1 is repeated. The algorithm stops once invoking Step 2 does not increase modularity. The output is a modularity value of the network and a hierarchical partition of the network into communities. This process has been validated by applying it to networks that have a known community structure, with the percentage of correctly grouped nodes at 0.98 (Blondel et al., 2008).

Results

Our primary focus was on the growth of the largest clique, but we used the additional graph theory analyses as a source of insight into the patterns of clique growth. We will thus present these analyses in turn. Results were very similar across the two growth models, so they are presented together.

Clique Size Growth

In both growth models, we found that the largest clique initially grew roughly linearly, but the rate of growth decreased significantly as the number of nodes in the graph increased (see Figure 2). The increase in clique size is thus sub-linear, and the increase in the computational cost of probabilistic inference is sub-exponential. In the remainder of this section we consider the factors that contributed to this growth pattern.

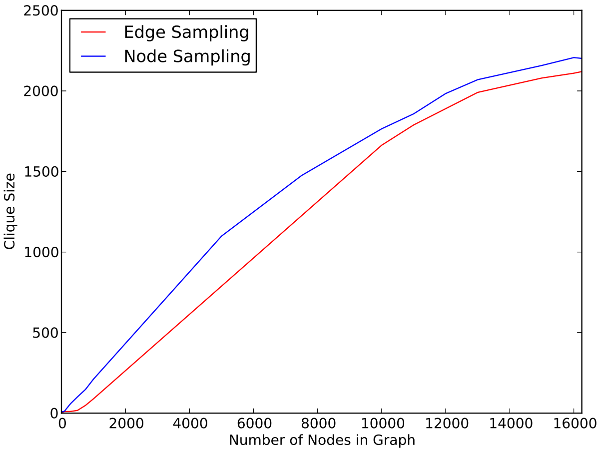

Figure 2: Size of the largest clique as a function of the number of nodes in the graph, for both growth methods.

Degree

Our analysis of the degree of nodes included evaluating the overall degree distribution in ConceptNet, and looking at the role that degree played in the growth of the graphs.

Overall Degree Distribution Taking the full ConceptNet graph, we found that the overall degree distribution is heavy-tailed (consistent with a power-law or log-normal distribution), as shown in Figure 3. This is consistent with the degree distribution observed in other graphs based on human semantic knowledge (Steyvers & Tenenbaum, 2005), and is a common observation in other complex networks (Sporns, 2009). It indicates that while many nodes have low degree, a few nodes have very large degree, forming connections with a large number of other nodes.

Degree of Nodes added to Graph We observed that, despite the random sampling process behind both of our growth methods, high degree nodes were added to the graph early (see Figure 4). If nodes are sampled randomly, the nodes with the most edges are most likely to be found and added to the graph early. If edges are sampled randomly, the nodes with the highest degree are more likely to be added to the graph, as they have the most edges. This result is a priori predictable, assuming a heavy tailed degree distribution, which we observed. As high-degree nodes are added early, the average degree of the network is higher while the graph is small. As the graph grows, and lower degree nodes are added, the average degree begins to shrink. As soon as the average degree begins to shrink, clique size growth slows.

The observation that high-degree nodes are added early provides some validation for the assumptions behind our growth models. Previous models (e.g., Steyvers & Tenenbaum, 2005) have made predictions about the time course of semantic acquisition. These models show that the order in which concepts are acquired is crucial in determining their connectivity—concepts that enter the network early are expected to show higher degree. This has been verified against age of acquisition norms (Gilhooly & Logie, 1980; Ellis & Ralph, 2000), which show that high-degree concepts are learned early. In this growing network model, the degree of a node decreases as a function of the time since it was first connected to the network. We observed this same phenomenon—as the graph grew, the degree of nodes added to the graph decreased.

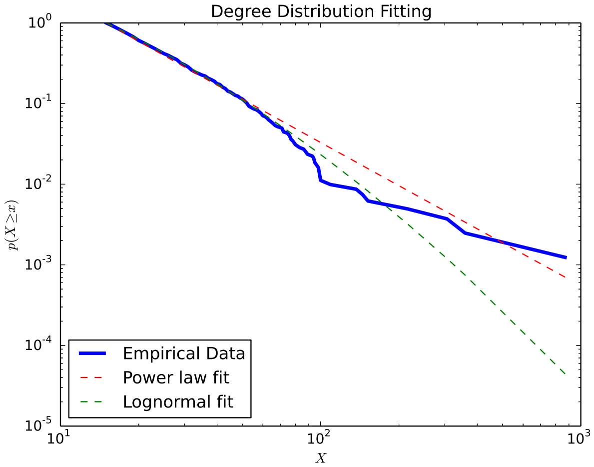

Figure 3: Overall degree distribution for ConceptNet, plotted on log-log axes. The near-linear relationship is consistent with a heavy-tailed distribution.

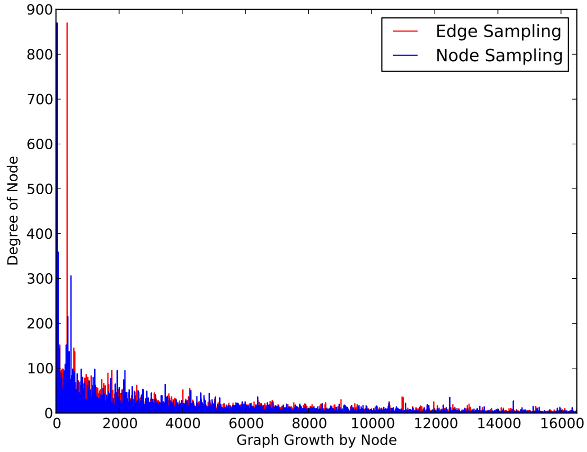

Figure 4: Degree of nodes during growth, shown in order of addition to the graph under both growth models. High-degree nodes tend to be added early.

Modularity

The modularity of the full ConceptNet graph is very high ($Q = 0.79$). During the initial stages of growth, we observed that

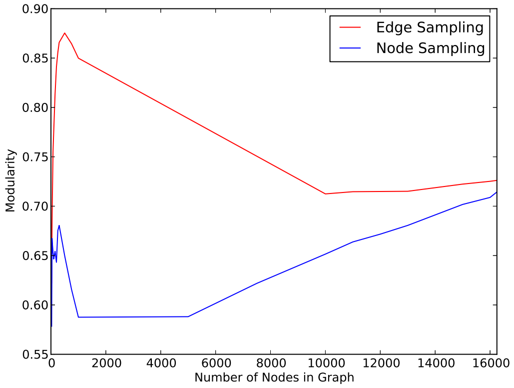

Figure 5: Modularity as a function of the number of nodes in the graph, for both growth methods.

modularity decreased as the clique size grew. After this initial growth, modularity stabilized, and clique size growth slowed (see Figures 2 & 5). This suggests a relationship between modularity and clique size growth. A modular network will have smaller cliques than a non-modular network, as edges between modules are limited in a modular network, which results in many small cliques, assuming the sizes of the communities stay stable during growth. In a non-modular network, there are no decomposable entities, resulting in one large clique. We speculate that, as long as network growth does not lead to decreases in modularity, clique size will continue to increase slowly as nodes are added beyond the range that we considered in our analysis.

Communities

The community-finding analysis identified clear partitions of the graph into subsets that had many internal connections but few external connections (see Figure 6). We also observed a linear growth trend in the number of communities in the graph, suggesting the dynamic ability of the network of concepts to form new, more specific communities as more concepts are added, which ensures that modularity stays stable or increases.

More specific communities allows for more efficient searches of what connected variables are critical for an inference, which is required as the graph grows. For example, an early, small version of the graph might contain the following nodes in one community: “horse”, “President Obama”, “Richard Feynman”, “redwood”, “dog”, “Carl Sagan”, “cactus”. All of these concepts can coherently exist in one community, as they are all living organisms. However, as more concepts are added, three new communities might be created, each only containing plants, animals, or humans, or even a further division of humans into politicians and scientists. Thus, the search space for important connected variables remains constant, even though the number of variables in the graph continues to grow.

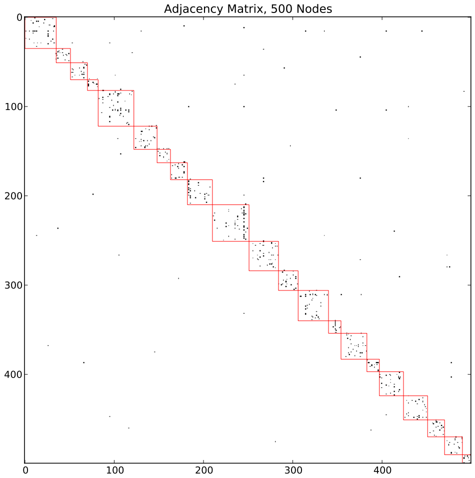

Figure 6: Adjacency matrix during growth under the node sampling growth model, with 500 Nodes. Black indicates where edges exist, white where edges do not exist. The red boxes represent communities. The graph is highly modular, with few connections outside communities.

Discussion

Hume’s solution to the problem of induction was that Nature builds in the proper connections between concepts, claiming that “Nature, by an absolute and uncontrollable necessity has determined us to judge as well as to breathe and feel”. Hume was on the right track. Holism does not present a problem for inductive inference, as the nature of the acquisition of our concepts and the modular nature of the connections between them, as well as the heavy tailed degree distribution, lead to sub-linear growth in clique size, meaning that efficient probabilistic inference is potentially possible.

We observed that, because our conceptual network has a heavy tailed degree distribution, high-degree nodes are added early. While these nodes are added, modularity decreases, and clique size growth is linear. However, when lower-degree nodes are added to the graph, modularity stabilizes and slowly increases (in part by increasing the number of communities), and clique size growth slows. This suggests that the timecourse of semantic acquisition and the degree distribution and modular structure of our conceptual network makes inductive inference tractable. In other words, the connectivity of a variable (i.e., a concept) constrains what variables one must observe to form a belief regarding that variable, as the connectivity structure of our conceptual network is modular, which limits the size of cliques. Thus, it is not the case that one must do an exhaustive search of every connected variable.

Fodor claimed that, because our central processing capacities, like reasoning and decision making, potentially require the truth of any proposition to provide confirmation to any other, inductive inference is intractable. Perhaps worse, the holistic nature of our conceptual network makes the study of central processing capacities a futile effort, as one can not tease apart holistic systems. However, we found evidence that the assumptions of Fodor's claims are false. We found that, while our conceptual network is holistic, in that it is connected, this does not present a theoretical problem for inductive inference or the study of central processes, as this network is modular in the mathematical sense. While it might not be informationally encapsulated—the criterion that Fodor believes makes analysis of certain psychological faculties possible—this does not mean that it is not decomposable. Finally, Fodor differentiates two aspects of holism: confirmation is isotropic, in that any proposition can in principle bear on any other, and also "Quinean", in that the confirmation of a proposition could depend on properties determined by the entire belief system, such as its simplicity. While our analysis has focused on this isotropic aspect of holism, Quinean concerns are dependent on the network being isotropic; thus, we suspect that the modular structure of human knowledge may also address the Quinean aspect of holism, as global properties like simplicity might not be as global as Fodor assumes.

Holism is a significant challenge, and the present work just represents a first step towards evaluating its seriousness. One concern for proponents of inductive inference is that, while the size of the largest clique grows sub-linearly, the largest clique remains large enough to pose a significant computational challenge. Understanding how this continues to scale will require analyzing larger graphs. Future analyses should also include null growth models, as well as models that vary modularity, community growth, and degree distribution. This will allow us to make stronger claims as to what drives sub-linear growth in clique sizes. For example, a heavy tailed degree distribution and modularity might actually be independent but necessary attributes of sub-linear growth, or the heavy tailed degree distribution might determine the modular structure, which in turn limits clique size growth. Moreover, future analyses should use different knowledge networks and inference algorithms, as the theoretical significance of the current analysis depends on how well Bayesian networks capture inference, and on how well ConceptNet serves as an approximation of knowledge.

More broadly, most graph theory analyses focus on describing properties of graphs, rather than considering algorithmic processes that occur in graphs. The approach taken in this paper demonstrates the utility of going beyond investigating the "descriptive statistics" of complex networks, and asking questions about network structure motivated by algorithmic properties. We asked, giving the structure of this network, what algorithms are computationally possible? While we used Bayesian inference, there might be other, similar questions about algorithms and representations that scientists can ask that go beyond the traditional "complex network" approach. In particular, a similar algorithmically-motivated analysis could be applied to network analysis of the human brain.

Acknowledgments. This material is based upon work supported by a National Science Foundation Graduate Research Fellowship DGE-1106400 and National Science Foundation Grant SMA-1228541. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Albert, R., Jeong, H., & Barabasi, A. (1999). Diameter of the world-wide web. *Nature*, *401*, 130-131.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008 (12pp).

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., … Wiener, J. (2000). Graph structure in the web. *Comput. Netw.*, *33*, 309–320.

Ellis, A., & Ralph, M. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology*, *26*, 1103–1123.

Faloutsos, M., Faloutsos, P., & Faloutsos, C. (1999, August). On power-law relationships of the internet topology. *SIGCOMM Comput. Commun. Rev.*, *29*(4), 251–262. doi: 10.1145/316194.316229

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT/Bradford Press.

Fodor, J. (2001). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: MIT Press.

Gilhooly, K., & Logie, R. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods and Instrumentation*, *12*, 395–427.

Hume, D. (1739). *A treatise of human nature*. Oxford: Oxford University Press.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabasi, A. (2000). The large-scale organization of metabolic networks. *Nature*, *407*, 651-654.

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. MIT Press.

Koller, D., Lerner, U., & Angelov, D. (1999). A general algorithm for approximate inference and its application to hybrid bayes nets. *Proceedings 15th Conference on Uncertainty in AI*, 324-333.

Meunier, D., Lambiotte, R., & Bullmore, E. (2010). Modular and hierarchically modular organization of brain networks. *Frontiers in Neuroscience*, *4*, 200.

Murphy, K., & Weiss, Y. (2001). The factored frontier algorithm for approximate inference in dbns. *Proceedings of the 17th Conference on Uncertainty in AI*, 378-385.

Newman, M. E. J., & Girvan, M. (2004, Feb). Finding and evaluating community structure in networks. *Phys. Rev. E*, *69*, 026113.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.

Speer, R., & Havasi, C. (2012, may). Representing general relational knowledge in conceptnet 5. In N. C. C. Chair) et al. (Eds.), *Proceedings of the eight international conference on language resources and evaluation (lrec'12).* Istanbul, Turkey: European Language Resources Association (ELRA).

Sporns, E. B. . (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*, 186–198.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*, 41–78.

Wagner, A., & Fell, D. (2001). The small world inside large metabolic networks. *Proc. R. Soc. Lond. B*, *268*, 1803–1810.