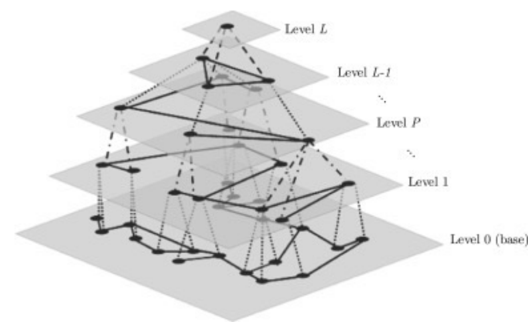
"Hierarchical graph embedding in vector space by graph pyramid," S. Mousav, et al.



NICE 2022 Lightning talk: Graph Embedding Using Cortical Like Sparse Distributed Representations

DAN HAMMERSTROM (PORTLAND STATE) dwh@pdx.edu

DMITRI NIKONOV (INTEL)

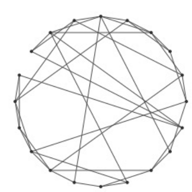
MOHAMED ABIDALREKAB (PORTLAND STATE)

This work was supported in part by the Center for Brain Inspired Computing (C-BRIC),one of six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA.

Center for Brain Inspired Computing

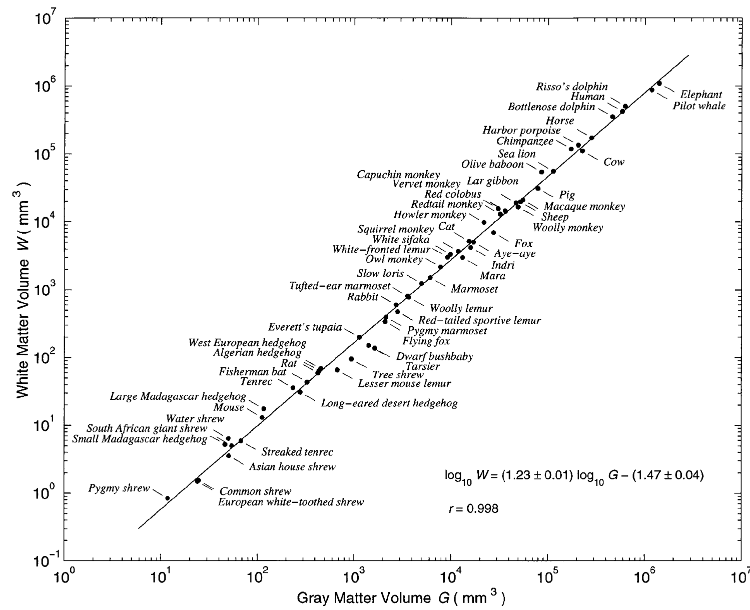
Research Goal

- Cognition involves accessing and performing inference over complex, structured knowledge
  - Capture knowledge, represent it, perform inference over it
  - Graphs are a natural representation of such sparse information
- Research goal: explore the mapping of graphs to basic cortical like arrays
  - Start with a simplified cortical model - incrementally add cortical characteristics
  - Sparse, distributed data representations
  - Competing objective functions:
    - Increased mapping quality amb performance on standard graph queries (here label inference)
    - Connectivity amb number and length of connections
- Most real-world graphs (semantic, knowledge, social networks, …) have small world/scale free characteristics
  - A scale-free network is a network whose degree distribution follows a power law
  - In a small-world network the typical distance between two randomly chosen nodes grows according to log(N)
  - Has been observed in neural circuits



Cortical Interconnect Architecture

- What role does cortical interconnect architecture play in cortical functionality?
- Connections are expensive in biology
- Cortex has, most likely found the “minimum” connectivity - number and length of connections - required by cortex to do its job
- “A universal scaling law between gray matter and white matter of cerebral cortex,” K. Zhang and T. Sejnowski, PNAS, May 9, 2000, Vol. 97 No. 10 5621–5626
  - Neocortex has a similar layered architecture in species over a wide range of brain sizes
  - Larger brains => longer fibers to communicate between distant cortical areas
  - White matter volume increases disproportionally (4/3 power) faster than gray matter volume
  - Power law accounts for empirical data spanning several orders of magnitude in brain sizes for various mammalian species



$$W = \frac{C}{T} G^{4/3}$$

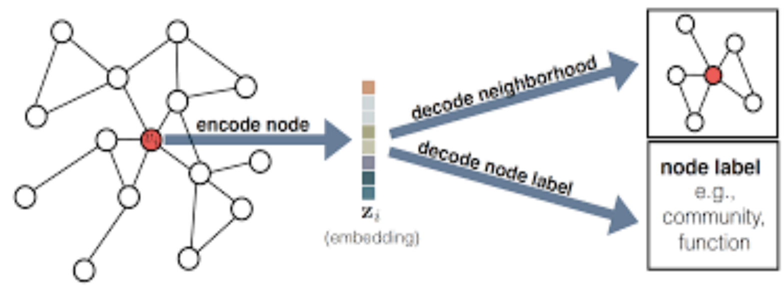
Portland State
UNIVERSITY

C-BRIC

Current Approaches (Deep Neural Networks)

- Deep Network architectures for “graph embedding” Graph Neural networks (GNNs):Node2Vec, Graph2Vec, TransE, RecGNNs, Graph Convolutional Neural Networks

"Representation Learning on Graphs: Methods and Applications",
W. Hamilton, et al., Department of Computer Science Stanford University.



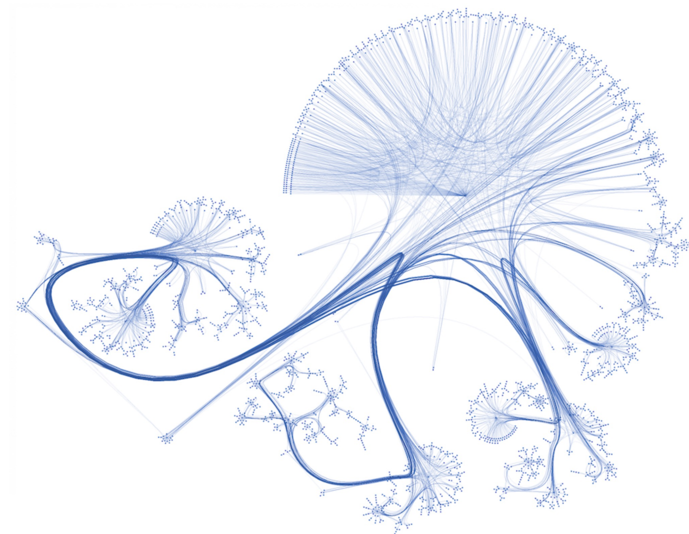
- Gradient descent results in broad data distributions
  - Representing sparse graphical relationships can be less efficient
- Other disadvantages: long training times, scalability, the need to retrain when there is new information
- Biological inspired algorithms address some of these disadvantages
  - Numenta’s “complementary sparseness”: sparse interconnect and sparse activation

Portland State UNIVERSITY

C-BRIC

Cognitive Graph (COG)

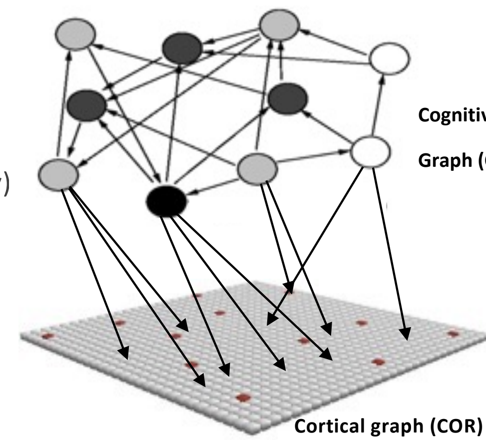
- A source graph is read into the system and an internal data structure, COG, is created
- Graph data sets used (labeled directed graphs)
  - adjnoun (A network of word adjacencies of common adjectives and nouns in the novel "David Copperfield" by Charles Dickens)
  - webkb-wisc (A dataset that includes web pages from computer science departments of various universities)
  - CORA (A dataset of scientific publication citations classified into one of seven classes)
  - Human Protein Interaction (A protein-protein interaction network for Homo Sapiens, where nodes represent proteins, and edges indicate the biological interaction between a pair of proteins)



CORA
https://graphsandnetworks.com/the-cora-dataset/

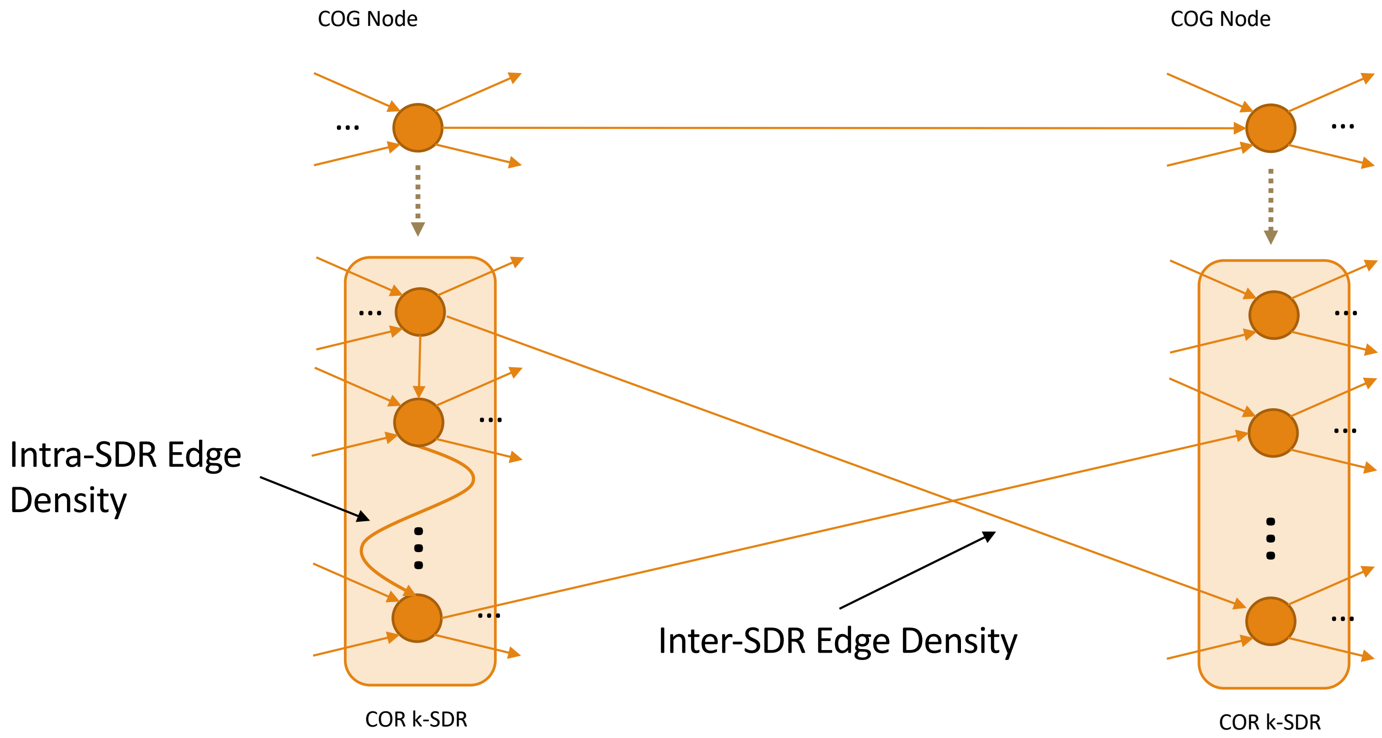
Cortical Graph (COR)

- The COR is constructed via a random walk through the COG (rapid learning)
  - Random walks (from Node2Vec*) with a weighted balance DFS (depth first search) and BFS (breadth first search)
- The cortical network consists of a 2D XY grid of "columns"
- Each column has some number of minicolumns (COR nodes)
  - During the random walk *k* minicolumns (*k*-SDR) recruited for each COG node (constrained random)
  - Axons are recruited randomly: a minicolumn can connect, on average, to P_a (axon connection density) other minicolumns (we use 1-5%)
  - Single layer columns, single level XY grid
- The COR representations of the COG nodes constitute a very sparse "graph embedding"
- Synaptic (edge) weights for the COR nodes are determined by simple Hebbian learning during the random walk
  - Are labels correlated or anti-correlated (can be + or -)
- The simulator allows for a range of COR allocation and interconnect techniques
  - SISO – Similar Input Similar Output (Representation)



*node2vec: Scalable Feature Learning for Networks. A. Grover, J. Leskovec, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2016.

COR Node Allocation



Graph Query: Label Inference

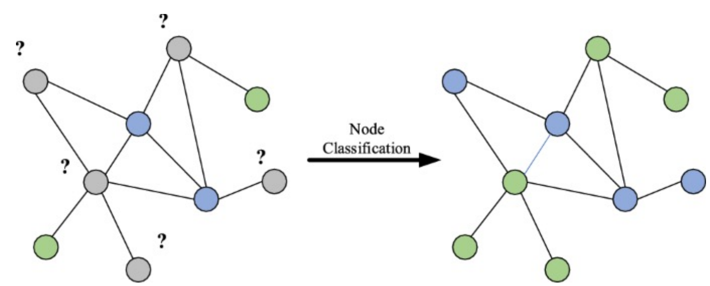
- COR nodes in each COG node's COR *k*-SDR share the same label

- For label inference, 50% of the nodes in each graph have their labels removed
  - These are then inferred from the remaining labels

- Label likelihoods ("beliefs") are propagated through the network
  - Likelihood propagation via message passing
  - Nodes accumulate <u>label log likelihoods</u> from neighbor nodes
  - COR: subsets of identically labeled nodes form *cell assemblies*, reinforcing each other "resonance"

- A standard label inference (node classification) algorithm
  - Node Classification in Social Networks, S. Bhagat et al.
  - No "high order information" (homophily) factored into process

Xiao, S., Wang, S., Dai, Y. *et al.* Graph neural networks in node classification: survey and evaluation. *Machine Vision and Applications* **33,** 4 (2022)



Preliminary Results

Data Set	adjnoun		webkb		CORA		Human Protein (PPI)
Cognitive graph (nodes/edge/labels)	112/425/2		265/822/5		3264/4536/7		3890/76584/50
Cortical graph (nodes/edge)	1024/1349		16384/1687		16384/3456		262K/49K
Cognitive graph, Macro-F	99%		77%		90%		10%
Node2Vec, Macro-F	83%		63%		75%		18%
Cortical graph, Macro-F	95%	88%	86%	83%	90%	82%	76%
Cortical graph. Connectivity	5%	1%	5%	1%	5%	1%	1%
Connection Effectiveness	0.49	0.84	0.31	0.45	0.33	0.38	1.43

- adjnoun k=4, all others k=8
- Connection effectiveness, which is only computed for the Cortical graph, is the network performance (Macro-F1 score) divided by the mean connections per node times the mean connection length

Example of cell assembly resonance

Next Steps

- COR architecture:
  - More “physical” diffusion model of label belief
  - Numenta-like dendritic segments that capture “higher order” information (homophily)
  - Multi-level hierarchies that capture more graph structure
  - Belief propagation by spikes
  - More complex multi-layered columns
  - Analysis to drive architecture
  - Investigate functional implications of Small World and Scale Free networks

- Expand the tasks, a wider variety of graph queries
  - Label inference (node classification, NC), given a partly labeled graph, infer or classify the unlabeled nodes
  - Edge inference (link prediction or graph completion), remove edges from a learned graph, infer missing edges
  - PageRank (PR) Used extensively in graph search