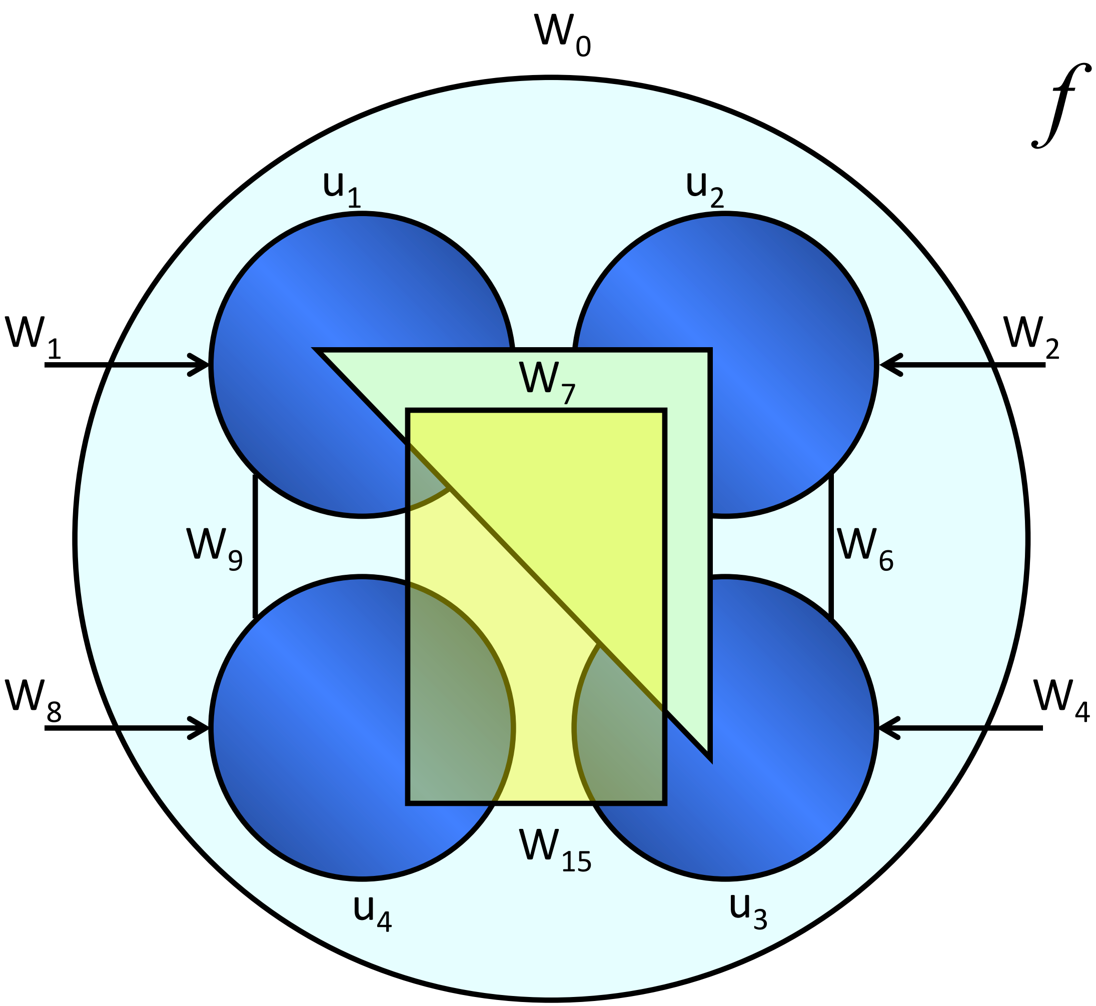
Structure Discovery in MOHNs

Kevin Swingler

PhD. Talk 2015

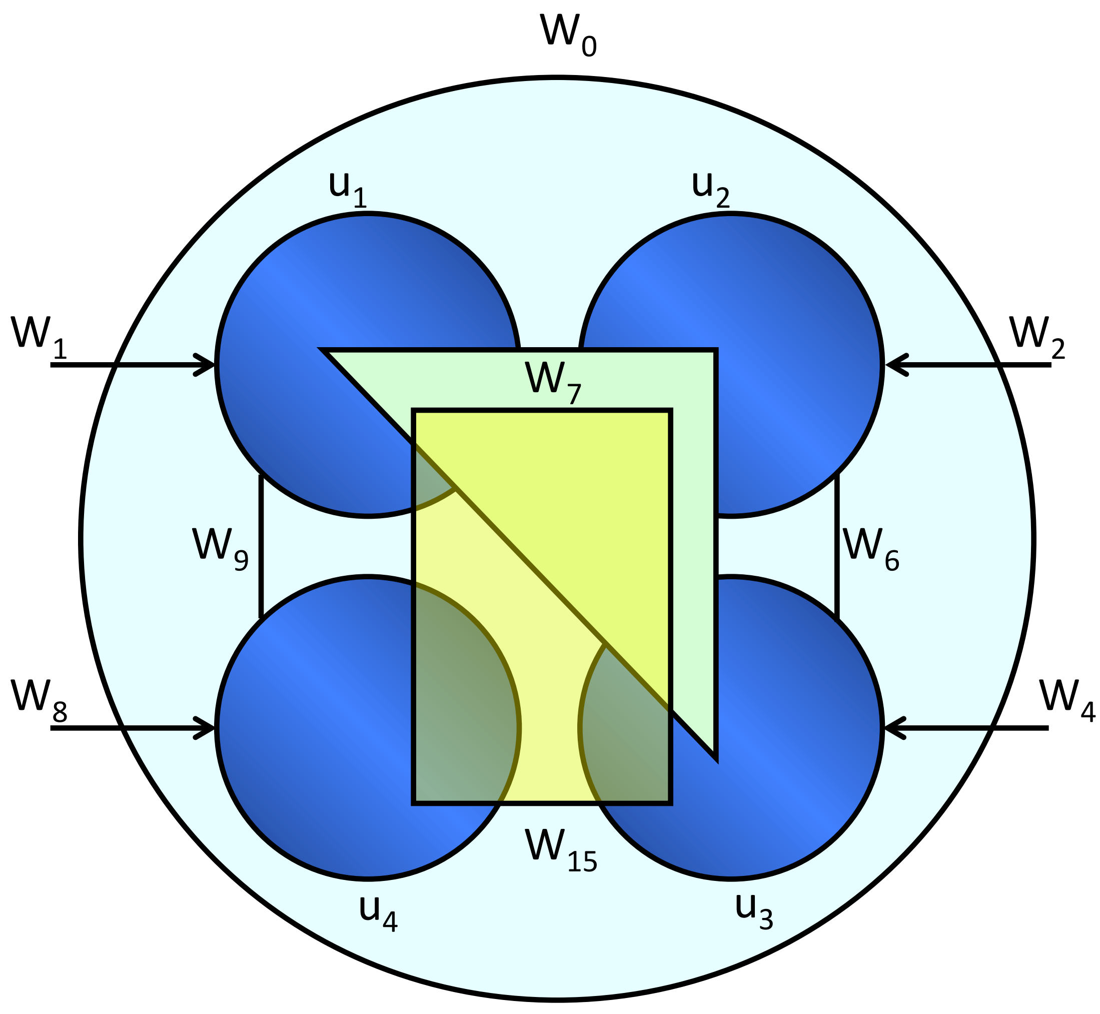
Mixed Order Hyper Network



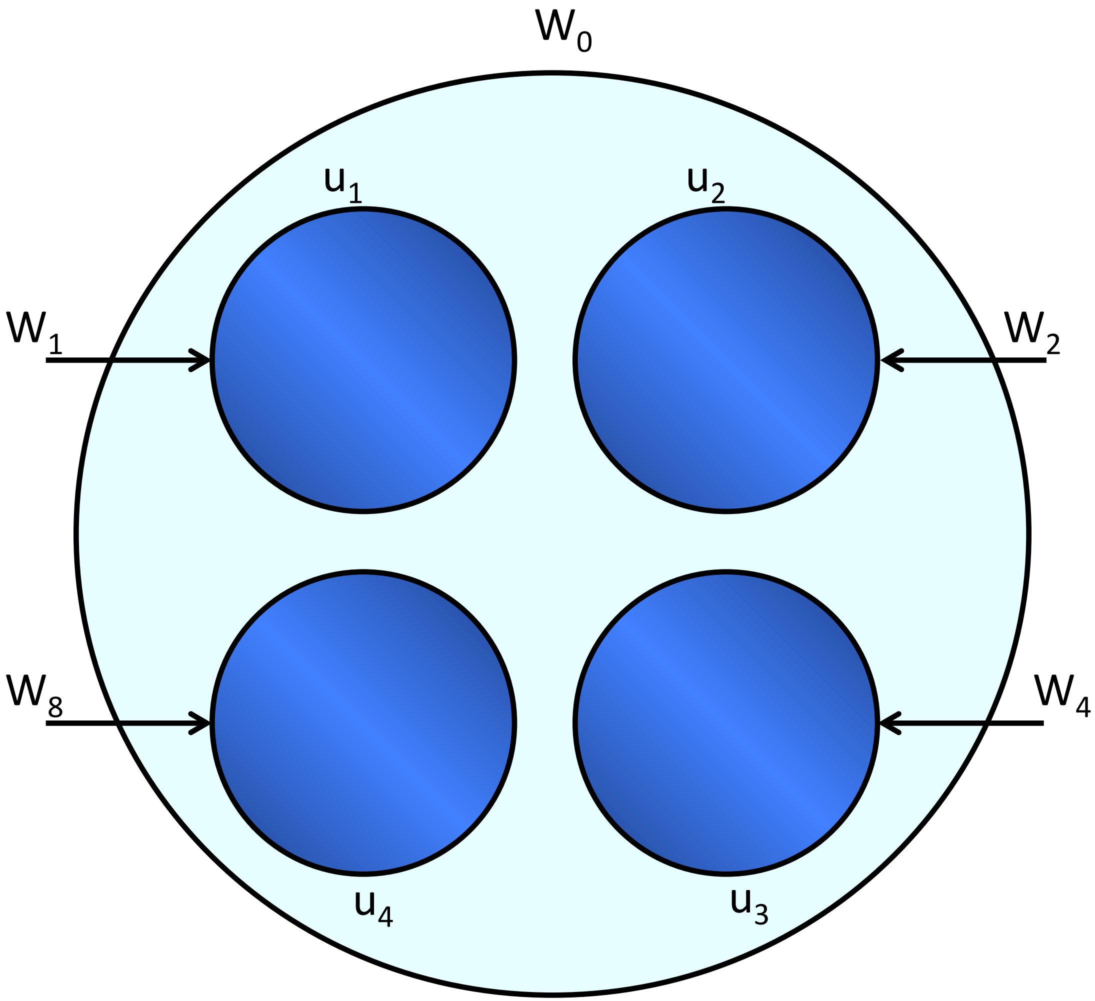
$$f(x) = \sum_{i \in N} W_i \prod_{u \in Q_i} u$$

$$u \in \{-1, 1\}$$

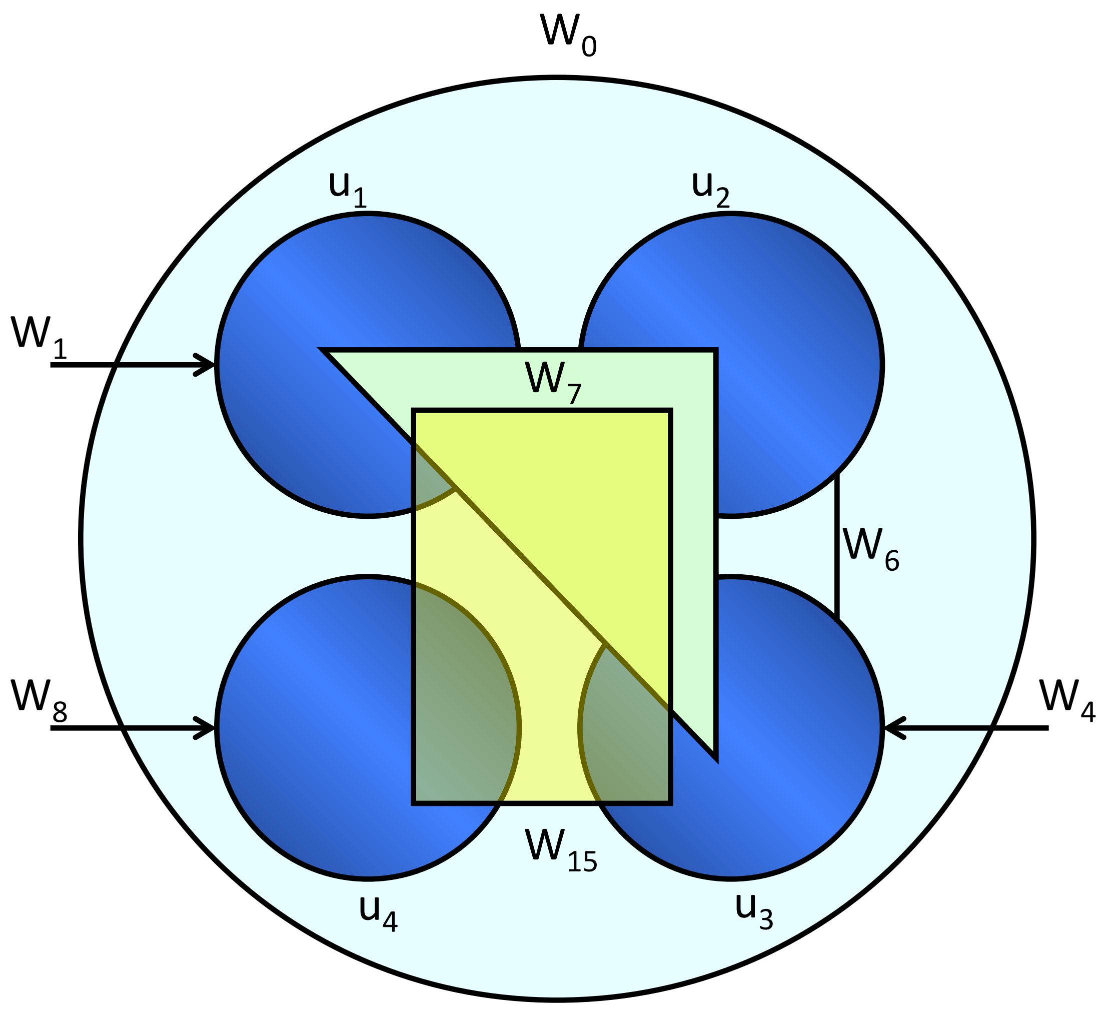
Mixed Order Hyper Network



Mixed Order Hyper Network



Mixed Order Hyper Network



What is the Right Structure?

2^{2^n} Possible configurations

Structure means which weights to include

Structure Discovery based on data

Some Possible Approaches

- Greedy
  - Pick the first order weights that are significant by some statistical test
  - Among those, look for second order weights that are significant
  - ...etc.
  - Problem - what if all significant weights are order 2, or higher?

A Structure Discovery Algorithm

- Maintain two distributions:

- Probability of interactions at each possible order being important: $P(m)$
- Probability of each input (u) being important: $P(u)$

The Algorithm

1. Pick an order, o, by sampling $P(m)$
2. Pick o neurons by sampling from the distribution $P(u)$
3. Add a weight, w, connecting the neurons from step 2.
4. Use a learning rule to set the value of w
5. Test the significance of the value of w and remove it if it is insignificant
6. Update the distributions $P(u)$ and $P(m)$
7. Repeat from 1 until convergence

Sampling from the Orders $P(m)$

- You can choose the shape of $P(m)$ to incorporate any prior knowledge, or to impose a sensible rule such as try low orders first:

$$P(m) = \lambda e^{-\lambda|c-m|}$$

- You might also have knowledge you can use to set $P(u)$

Evolving $P(m)$

- The mode of the exponential distribution for $P(m)$ is at c
- As lower orders are exhausted, and the number of samples available grows, c can increase to allow higher orders to be explored
- At the same time, $P(m)$ is reshaped according to how many weights at each order have proved significant

$P(u)$ Exploring / Exploiting

- Two possible assumptions:
  1. All variables have some use, so those that have not yet been used should be explored

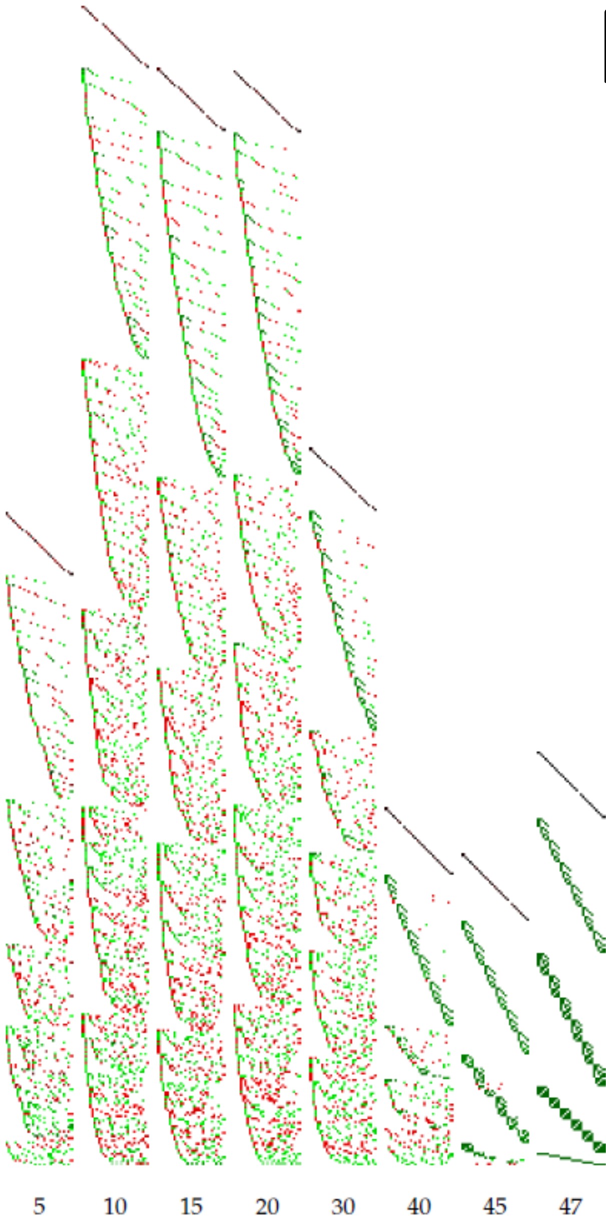
  2. Variables that have proved useful at lower orders are likely to be important at higher orders too

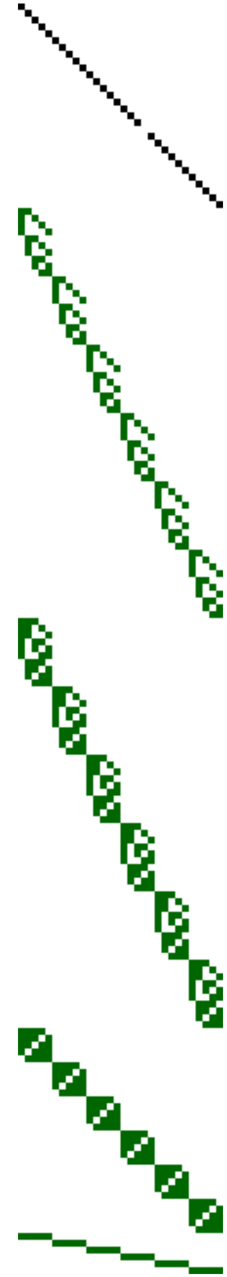
- Probability $P(u)$ can be manipulated to reflect which assumption is in force

Example Higher Order Function

- 5 bit-trap function:
- Variables interact in contiguous groups of four
- Interactions at orders 1,2,3,4,5
- Trap part means function seems to be simple first order, but is high order in some parts of input space

Results





Clustering Digits 1-10