1 Intro

Our current method makes two simplifying assumptions:

1. **Perfect emission**: nodes that have not yet been emitted are emitted with probability 1 on first hit
2. **Perfect censoring**: nodes that have been previously emitted are censored with probability 1 on all future hits

The following method generalizes to allow imperfect censoring or emissions with a fixed probability.

2 Method

Let p_{emit} be the probability of emitting a node when it is encountered on a random walk and has not been emitted before. Let p_{censor} be the probability of censoring a node when it is encountered on a random walk and has been emitted before. Both parameters should typically be close to 1.

Previously, to calculate $P(X_{t+1}|X_1...X_t)$ we treat the set of previously emitted nodes, $A = \{X_1...X_t\}$, as non-absorbing nodes and the set of previously un-emitted nodes, $B = \{X_{t+1}...X_k\}$, as absorbing nodes (k is the number of elements in X).

Let P be the transition matrix of graph G. We re-arrange the states of P into:

$$P' = \begin{bmatrix} Q & R \\ S & T \end{bmatrix}$$

where Q denotes the transitions from A to A, R denotes transitions from A to B, S denotes transitions from B to A, and T denotes transitions from B to B. To account for imperfect censoring, we can treat A as non-absorbing with probability p_{censor} and as absorbing with probability $1 - p_{censor}$. Similarly, to account for imperfect emission, we can treat B as absorbing with probability p_{emit} and as non-absorbing with probability $1 - p_{emit}$. We compose new matrices:

$$Q' = \begin{bmatrix} p_{censor} \cdot Q & (1 - p_{emit}) \cdot R \\ p_{censor} \cdot S & (1 - p_{emit}) \cdot T \end{bmatrix}$$

$$R' = \begin{bmatrix} (1 - p_{censor}) \cdot Q & p_{emit} \cdot R \\ (1 - p_{censor}) \cdot S & p_{emit} \cdot T \end{bmatrix}$$

$$P'' = \begin{bmatrix} Q' & R' \\ 0 & I \end{bmatrix}$$

Q' denotes transitions to all non-absorbing nodes that will not be emitted on the random walk. R' denotes transitions to all absorbing nodes that will be emitted on the random walk. Note that if P is n x n large, P'' is $2n$ x $2n$ large to allow each node to be both absorbing and non-absorbing.

Given values for p_{emit} and p_{censor}, we can now solve as we have been doing. Or we might want to define a prior distribution for each parameter to marginalize over.

3 Discussion

From a psychological perspective, imperfect censoring is necessary to account for perseverations. p_{censor} reflects some monitoring component, where lower values correspond to poorer monitoring. It could also be used to model disruptions to working memory. Imperfect emission may reflect a weak connection between semantic and lexical (or motor) nodes. It could be used to model certain types of anomia. Ideally, we would want to perform inference to recover these parameters, since they have some psychological meaning.

From an ML perspective, perfect censoring may throw out informative data when constructing a graph. Otherwise, if the raw data contain perseverations, we must throw them out as if they didn't happen, distorting the data. On the other hand, imperfect emission does not throw out any data, but may provide better fits if the underlying process does contain imperfect emissions.