Dynamics on Hypergraphs

University of Oxford

Jessica Toudic

A dissertation submitted for the

Honor School of Mathematics Part C

Trinity 2021

Abstract

In this dissertation, we develop a model based on random walks on hypergraphs to represent the thinking process behind recovery from mental illness using self-directed neuroplasticity (the ability to harness the power of thought to change the structure of the brain). We suggest through numerical simulations that our model emulates two key characteristics of successful engagement in self-directed neuroplasticity, namely that recovery is more successful the better able you are to (1) reappraise, and (2) refocus attention away from, maladaptive thoughts.

Our work makes novel contributions to the development of hypergraph theory, and may find broader applications in hypergraph science. Firstly, we introduce a new dynamical process, an edge-centric random walk on an hypergraph, and find its stationary distribution. We also define a certain projection of an hypergraph onto its hyperedges, which we call the contracted network, and show that an edge-centric random walk on an hypergraph is equivalent to a random walk on the corresponding contracted network. Furthermore, we introduce a marking process for hypergraphs, and propose a novel mechanism for generating random hypergraphs based on preferential attachment to hyperedges rather than nodes.

Contents

1 Introduction **3**

2 Introducing Hypergraphs **4**
- 2.1 Basic Definitions . . . 4
- 2.2 Introducing the Thought Hypergraph . . . 6

3 A Train of Thought **8**
- 3.1 Node-centric Random Walks . . . 8
- 3.2 Edge-centric Random Walks . . . 10
- 3.3 Projecting the Random Walks . . . 13
- 3.4 Reflection . . . 15

4 Refocusing **16**
- 4.1 Defining Refocusing . . . 16
- 4.2 Stationary Distribution with Refocusing . . . 18
- 4.3 Reflection . . . 20

5 Recovery **21**
- 5.1 Reappraisal . . . 21
- 5.2 Recovery Process . . . 22
- 5.3 Recovery Time . . . 23
- 5.4 Efficiency . . . 25
- 5.5 Reflection . . . 25

6 Numerical Simulations **26**
- 6.1 Generating Hypergraphs . . . 26
- 6.2 Simulating the Recovery Process . . . 27
  - 6.2.1 Reappraisal . . . 28
  - 6.2.2 Refocusing . . . 29
- 6.3 Reflection . . . 30

7 Discussion **31**

References **32**

A Appendix **35**
- A.1 Knowledge Toolkit . . . 35
- A.2 Recovery Process Additional Figure . . . 35
- A.3 Terminology . . . 36
- A.4 Sketch-Proof of Theorem 5.1 . . . 37
- A.5 Mechanisms for Generating Hypergraphs . . . 41
- A.6 Additional Figures . . . 44

1 Introduction

Cognitive behavioural therapy (CBT) is the among the most effective therapeutic treatments for anxiety disorders and depression [1]. Its strength lies in the ability of our mind to harness the power of thought to change the structure of our brain, known as self-directed neuroplasticity (SDN) [2, 3]. Successful engagement in SDN requires two processes: (1) using mindful awareness to recognise maladaptive thoughts and positively reappraise them, and (2) refocusing attention away from pathological thoughts to 'disengage from the topic of distress' [1, 3]. However, this means the efficacy of CBT falls short for individuals who lack adequate metacognitive awareness (the ability to 'think about thinking') [4].

Research has shown that visual explanations improve understanding of abstract concepts [5], and enhance conceptual understanding more effectively than verbal explanations alone [6]. This suggests that a visual model conceptualising the thinking process required for SDN (and hence CBT) may help those with poor metacognitive awareness recover from anxiety disorders or depression. However, such a model would only be useful if the process it abstracted emulated the two principles of successful recovery stated above, namely that recovery is more successful the better an individual is at reappraising their thoughts, and the better they are at disengaging from maladaptive thoughts [2, 3].

In this dissertation, we develop such a model. We base our model on hypergraphs, given their ease of visual representation, and define dynamics on hypergraphs to represent the process of recovery. Our model differs from any model in the literature because attempts to understand mental illness using mathematics, from biophysical models of the brain [7] to networks of pathological symptoms [8], have all been motivated by a goal to elucidate key mechanisms of development and maintenance of disorders, rather than emulate a recovery process.

The dissertation is structured as follows: in Section 2, we introduce basic definitions concerning hypergraphs, before introducing our model. In Section 3, we define dynamics representing thought. This includes the introduction of an edge-centric random walk process, a new projection onto the hyperedges of an hypergraph, and a proof of the equivalence between an edge-centric random walk on an hypergraph and a random walk on its corresponding projected network. In Section 4, we incorporate refocusing into our dynamical process. Section 5 sees us define (positive) reappraisal via a marking process, which leads us to the overall dynamical process of recovery. We also explain two measures that quantify the success of the recovery process. In Section 6, we explore the dynamics through numerical simulations. This involves presenting a novel process for generating random hypergraphs. Finally, we conclude with a discussion of our results, evaluation of the limitations of our model, and a suggestion of possible future directions.

2 Introducing Hypergraphs

2.1 Basic Definitions

We begin by introducing the basic definitions and notation required to define our model. The definitions are adapted from [9, 10, 11], and notation adopted mainly from [9, 11].

Definition 1 (**Hypergraph**). *An hypergraph $H = (V, E)$ is defined by a set of n nodes $V = \{v_1, v_2, \dots, v_n\}$ and a set of m hyperedges $E = \{E_1, E_2, \dots, E_m\}$. Each hyperedge $E_\alpha \in E$ is a non-empty subset of V. The size $|E_\alpha|$ of an hyperedge $E_\alpha \in E$ is equal to the number of nodes it contains. An hypergraph is simple if no hyperedge is a subset of another, and connected if each hyperedge shares at least one node with another hyperedge.*

As remarked in [9], if $|E_\alpha| = 2$ for all $\alpha \in \{1, \dots, m\}$, then the hypergraph reduces to a standard network. This leads us to think of an hypergraph as a generalisation of a network, in which each edge has an arbitrary positive number of nodes. Note also that a simple hypergraph restricts hyperedge sizes to be at least 2.

We can assign weights to the nodes and hyperedges of an hypergraph. In [11], Chitra and Raphael introduce edge-dependent node weights, whereby each node is assigned a distinct weight for each hyperedge it belongs to. However, we only assign nodes a single weight.

Definition 2 (**Weighted Hypergraph**). *A weighted hypergraph $H = (V, E, w, \gamma)$ is an hypergraph with associated node weights $\gamma = \{\gamma(1), \gamma(2), \dots, \gamma(n)\}$ and associated hyperedge weights $w = \{w(1), w(2), \dots, w(m)\}$, where $\gamma(i)$ is the weight of node v_i and $w(\alpha)$ is the weight of hyperedge E_α. Each weight is a positive real number.*

Given Definition 2, we note that the hypergraph in Definition 1 can be interpreted as a weighted hypergraph with all node and hyperedge weights set equal to 1.

In addition to assigning weights to nodes, we can also assign them states.

Definition 3 (**States**). *The state of node $v_i \in V$ is $\lambda_i \in \mathbb{R}$. This defines an hypergraph $H = (V, E, w, \gamma, \lambda)$ with associated node states $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$.*

States will become central in Section 5 when we consider reappraisal.

Rather than specifying an hypergraph by listing its nodes and hyperedges, we can encode it using a matrix.

Definition 4 (**Incidence Matrix**). *An $n \times m$ incidence matrix $\mathbf{e}$ encodes an hypergraph $H = (V, E, w, \gamma, \lambda)$ with n nodes and m hyperedges. It has entries*

$$e_{i\alpha} = \begin{cases} 1 & \textit{if } v_i \in E_\alpha, \\ 0 & \textit{otherwise.} \end{cases} \tag{1}$$

This allows us to define the adjacency matrix.

Definition 5 (**Adjacency Matrix**). *The $n \times n$ adjacency matrix of an hypergraph $H = (V, E, w, \gamma, \lambda)$ is $\mathbf{A} = \mathbf{e}\mathbf{W}\mathbf{e}^\top$ where $\mathbf{W}$ is an $m \times m$ diagonal matrix with α-th main-diagonal entry $W_{\alpha\alpha} = w(\alpha)$.*

The entry A_{ij} of the adjacency matrix equals the total weight of hyperedges that contain both nodes v_i and v_j.

We can define the degree of each node in the hypergraph analogously to the definition of node degree in a network, namely the sum of weights of edges incident to it.

Definition 6 (**Node degree**). *The degree $d(i)$ of a node $v_i \in V$ is given by*

$$d(i) = \sum_{\alpha} w(\alpha) e_{i\alpha}. \tag{2}$$

We can also define the degree of each hyperedge.

Definition 7 (**Hyperedge degree**). *The degree $\delta(\alpha)$ of an hyperedge $E_\alpha \in E$ is given by*

$$\delta(\alpha) = \sum_{i} \gamma(i) e_{i\alpha}. \tag{3}$$

That is, $\delta(\alpha)$ is the total weight of nodes contained in hyperedge E_α.

Throughout this dissertation, whenever we say 'hypergraph', we mean a simple, connected, weighted hypergraph. An example is shown in Figure 1. We will also assume that n and m are finite. Furthermore, following notation in [9], node indices will be denoted by Latin letters, and hyperedge indices by Greek letters (so that unless otherwise stated, Latin indices run from 1 to n and Greek indices run from 1 to m). With these definitions now in our toolkit, we can define the hypergraph model we will be exploring during this dissertation.

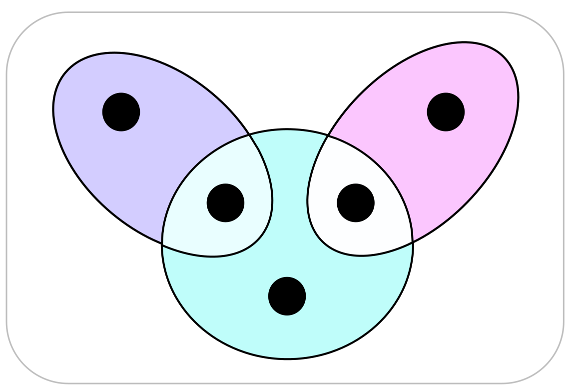


Figure 1: An hypergraph $H = (V, E, w, \gamma, \lambda)$ with 5 nodes and 3 hyperedges. Nodes are black dots and hyperedges are coloured. For consistency, all diagrams will use this hypergraph as an example.

2.2 Introducing the Thought Hypergraph

The model we define for our thoughts is based on the notion that all thought is combinatorial and associative [12]. Combinatorial means that thoughts consist of combinations of existing concepts, and associative refers to thoughts arising by direct association with concepts making up currently-active thoughts [12]. So, a concept acts as a stimulus of a thought it is associated to. In this way, a concept can be interpreted either as an external stimulus, including a physical or behavioural response, or as an internal stimulus, such as an emotion, feeling or belief [13]. Rather than specifying the nature of each concept, we will refer to all simply as 'concepts' throughout.

We now state the definition of our model, which is depicted visually in Figure 2.

Definition 8 (**Thought Hypergraph (TH)**)**.** *A thought hypergraph is a (simple, connected, weighted) hypergraph $H = (V, E, w, \gamma, \lambda)$, where each node represents a concept and each hyperedge represents a thought. The thought hypergraph consists of a finite number of concepts. Each hyperedge $E_\alpha \in E$ has weight $w(\alpha) = |E_\alpha|$, and each node $v_i \in V$ has state $\lambda_i \in \{0, 1\}$. We call the state of $v_i \in V$ maladaptive if $\lambda_i = 0$, and adaptive if $\lambda_i = 1$. The state of an hyperedge $E_\alpha \in E$ is*

$$\Lambda_\alpha = \frac{\sum_i \lambda_i \gamma(i) e_{i\alpha}}{\delta(\alpha)}. \tag{4}$$

We call the state of an hyperedge $E_\alpha \in E$ adaptive if $\Lambda_\alpha = 1$, maladaptive if $\Lambda_\alpha = 0$ and conflicted if $\Lambda_\alpha \in (0, 1)$. Finally, the thought hypergraph is adaptive if all nodes have state 1, maladaptive if all nodes have state 0, and otherwise conflicted.

By defining thoughts as hyperedges (i.e. as combinations of nodes), this guarantees our thoughts are combinatorial. The assumption that the number of concepts is finite means that the thought hypergraph we consider is a sub-hypergraph of a far larger hypergraph (the entire mind). For the purposes of our model, we assume that the hypergraph is static (conceptual combinations are pre-existing and ingrained in our minds, and span all possible relevant thoughts associated to the concepts), since the dissertation concerns dynamics *on* hypergraphs, and not dynamics *of* hypergraphs. Finally, the choice of hyperedge weights will become clear in the next section, when we consider the dynamics of thoughts.

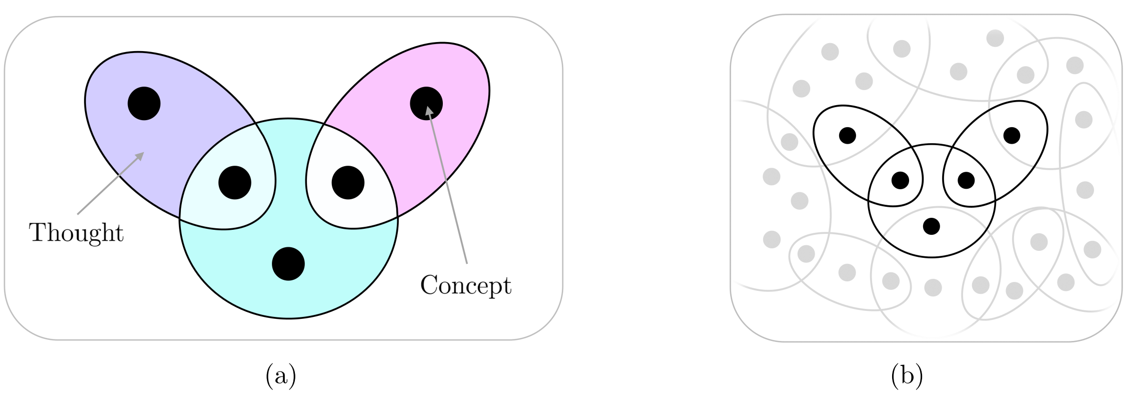


Figure 2: An example of a thought hypergraph. (a) A concept and thought labelled. (b) A thought hypergraph (black) is sub-hypergraph of the rest of the mind (grey).

The definition of maladaptive and adaptive states allows us to define recovery in terms of our model.

Definition 9 (Recovery). *Recovery is the process of a maladaptive thought hypergraph becoming adaptive. All concepts $v_i \in V$ initially have state $\lambda_i = 0$, and recovery has occurred when all concepts $v_i \in V$ have state $\lambda_i = 1$.*

The dynamical process we use to represent recovery is our focus for the sections that follow. Our first step, explored in the next section, is to define a process representing thinking.

Remark 1. To ensure connections to both the TH and hypergraph theory are made, we state definitions using mathematical terms, and follow them with a connection to our model. To guarantee clarity in our discussion, we have included a table comparing mathematical and model terminology in the Appendix.

3 A Train of Thought

Having introduced the TH, we can now begin exploring hypergraphs as a dynamic model for thought. Based on associativity of thoughts [12], an active concept acts as a stimulus for the next thought (combination of concepts). The activation of this thought then triggers another concept to act as a stimulus for the next thought, and so on, creating a 'train of thought' [12]. This process can be described mathematically by defining a random walk on the TH.

3.1 Node-centric Random Walks

A random walk (RW) on the nodes of a standard network involves a single choice: the random walker chooses an edge incident to the node it currently occupies, and moves to the other node contained in that edge. However, when considering RWs on the nodes of an hypergraph, the random walker faces an additional choice: since each hyperedge contains at least two nodes, the random walker must choose which node in the selected hyperedge to move to.

This extra choice has led to a number of varying definitions of RWs on the nodes of hypergraphs. In the seminal paper by Zhou, Huang and Schölkopf [14], hyperedge choice is made proportional to hyperedge weight, and node choice uniform across all nodes in a selected hyperedge. Chitra and Raphael [11] generalised this work by introducing node choices that are proportional to node degree. It is this generalisation that we focus on.

Definition 10 (**Node-centric Random Walk** [11]). *Let $H = (V, E, w, \gamma, \lambda)$ be an hypergraph. At time t, the random walker is at node $v_t = v_i$, and will:*

1. *Select an hyperedge $E_\alpha \in E$ containing node v_i, chosen with probability $w(\alpha)/d(i)$.*
2. *Select a node $v_j \in E_\alpha$, chosen with probability $\gamma(j)/\delta(\alpha)$.*
3. *Walk to node $v_{t+1} = v_j$ at time $t + 1$.*

This is a Markov chain on V with transition probabilities

$$T_{ij} = \sum_{\alpha} \frac{w(\alpha)}{d(i)} \frac{\gamma(j)}{\delta(\alpha)} e_{i\alpha} e_{j\alpha}, \tag{5}$$

which can be encoded in an $n \times n$ transition matrix $\mathbf{T} = (T_{ij})$.

In the context of the TH, a node-centric random walk is a concept-centric train of thought, since nodes represent concepts. Thoughts are activated in transitions between the concepts acting as stimuli. The following definition is equivalent to Definition 10, but stated in terms of the TH to aid understanding.

Definition 11 (**Concept-centric Train of Thought**). *Let $H = (V, E, w, \gamma, \lambda)$ be a TH. At time t:*

1. *The current stimulus $v_t = v_i$ activates one thought $E_\alpha \in E$ containing concept v_i. Thought E_α is chosen with probability $w(\alpha)/d(i)$.*

2. *Thought E_α triggers a concept $v_j \in E_\alpha$ to act as the next stimulus, chosen with probability $\gamma(j)/\delta(\alpha)$.*

3. *Concept $v_{t+1} = v_j$ acts as the stimulus at time $t+1$.*

A diagram of this process is shown in Figure 3. Note that it is possible that $v_{t+1} = v_t$, so the RW is *lazy* [11]. Chitra and Raphael [11] also consider non-lazy RWs, in which the subsequent node choice must be distinct from the current node, however we choose to use a lazy process as it better supports the notion of a train of thought: a thought does not have any memory about which concept activated it.

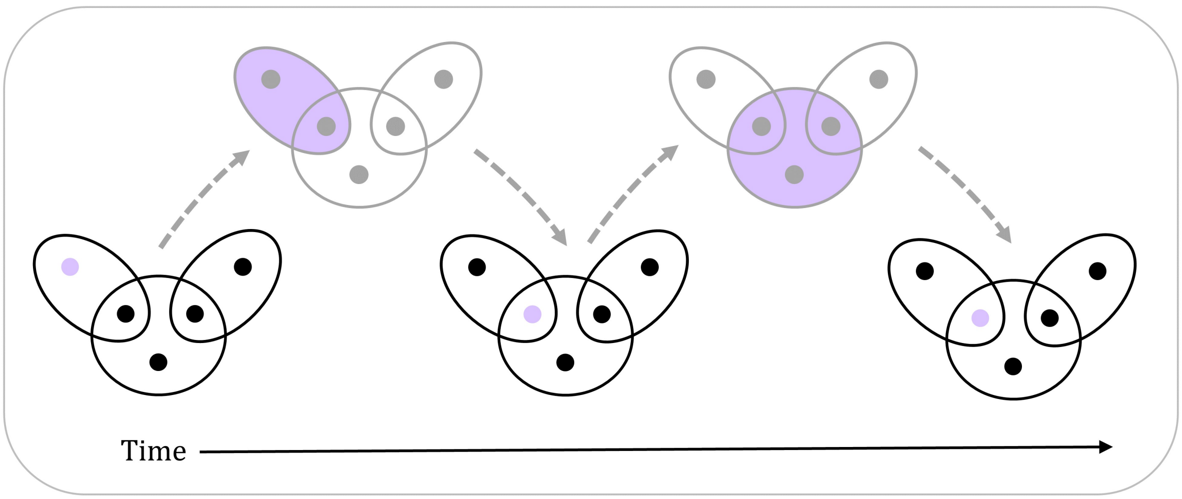


Figure 3: Node-centric random walk process. Bottom layer: purple node is the current position of the random walker. Top layer: purple hyperedge is the hyperedge activated during the transition between nodes. In the context of the TH, the purple node is the current stimulus, and the purple hyperedge is the thought activated by the stimulus during the transition.

Although we defined the RW in terms of arbitrary hyperedge weights, recall that in Definition 8 we chose $w(\alpha) = |E_\alpha|$ for each $E_\alpha \in E$. This choice was made because Carletti et al. [15] demonstrated that such a choice represents a bias of the random walker to remain in large hyperedges for longer. This directly reflects the notion that a thought occurs more often if more stimuli can activate it.

Remark 2. The Markov chain has finite state space since we are assuming n is finite, and is irreducible since H is connected. Hence, it is positive recurrent. Therefore, its stationary distribution exists and is unique.

Given Definition 11 and Remark 2, the stationary distribution of the RW is of interest to us, because it tells us the long-run proportion of time that a certain concept is used to stimulate a new thought.

Theorem 3.1 (**Node-Centric Random Walk Stationary Distribution** [11])**.** *Let* $\mathbf{T} = (T_{ij})$ *be the transition matrix of the node-centric random walk on an hypergraph* $H = (V, E, w, \gamma, \lambda)$*. The stationary probability of node* $v_j \in V$ *is*

$$\pi_j = \frac{\gamma(j)d(j)}{\sum_k \gamma(k)d(k)}. \tag{6}$$

Proof. By definition, the stationary probability of node $v_j \in V$ satisfies

$$\pi_j = \sum_i \pi_i T_{ij}. \tag{7}$$

This is satisfied by (6), since substituting (5) and (6) into the right-hand side of (7), and recalling the definitions of degree, gives

$$\begin{aligned}
\pi_j &= \sum_i \pi_i T_{ij} \\
&= \sum_i \frac{\gamma(i)d(i)}{\sum_k \gamma(k)d(k)} \sum_\alpha \frac{w(\alpha)}{d(i)} \frac{\gamma(j)}{\delta(\alpha)} e_{i\alpha} e_{j\alpha} \\
&= \frac{\gamma(j)}{\sum_k \gamma(k)d(k)} \sum_\alpha \frac{w(\alpha)}{\delta(\alpha)} e_{j\alpha} \sum_i \gamma(i) e_{i\alpha} \\
&= \frac{\gamma(j)}{\sum_k \gamma(k)d(k)} \sum_\alpha \frac{w(\alpha)}{\delta(\alpha)} \delta(\alpha) e_{j\alpha} \\
&= \frac{\gamma(j)}{\sum_k \gamma(k)d(k)} \sum_\alpha w(\alpha) e_{j\alpha} \\
&= \frac{\gamma(j)d(j)}{\sum_k \gamma(k)d(k)}.
\end{aligned}$$

By uniqueness of the stationary distribution, the stationary probability of $v_j \in V$ is $\pi_j = \gamma(j)d(j)/\sum_k \gamma(k)d(k)$, as required. □

Hence, the stationary probability of a concept is proportional the product of its weight and degree. We now introduce a new type of random walk on hypergraphs.

3.2 Edge-centric Random Walks

A consideration of edge-centric random walks on hypergraphs, in which the random walker moves between hyperedges (rather than nodes), has received a lack of attention in the literature. However, this alternative type of RW is of interest to us when considering the TH: an edge-centric RW on a TH focuses on the thoughts rather than the concepts, enabling us to gain insight into the long-term proportion of time spent thinking a certain thought. This is in contrast to the node-centric RW, in which the activated thoughts are hidden in transitions between stimuli.

We adapt Definition 11 to introduce an edge-centric random walk process on an hypergraph. Figure 4 demonstrates it visually.

Definition 12 (**Edge-centric Random Walk**). *Let $H = (V, E, w, \gamma, \lambda)$ be an hypergraph. At time t, the random walker is at hyperedge $E_t = E_\alpha$, and will:*

1. *Select a node $v_i \in E_\alpha$, chosen with probability $\gamma(i)/\delta(\alpha)$.*
2. *Select an hyperedge $E_\beta \in E$ containing node v_i, chosen with probability $w(\beta)/d(i)$.*
3. *Walk to hyperedge $E_{t+1} = E_\beta$ at time $t+1$.*

This process is a Markov chain on E with transition probabilities

$$T_{\alpha\beta} = \sum_i \frac{\gamma(i)}{\delta(\alpha)} \frac{w(\beta)}{d(i)} e_{i\alpha} e_{i\beta}, \tag{8}$$

which can be encoded in an $m \times m$ transition matrix $\mathbf{T} = (T_{\alpha\beta})$.

Analagously to the node-centric process, we can phrase the edge-centric process in terms of our TH using terms in Table A1 (in the Appendix), and call it a *thought-centric train of thought*. Note that this process is again lazy, so that the same thought can occur repeatedly.

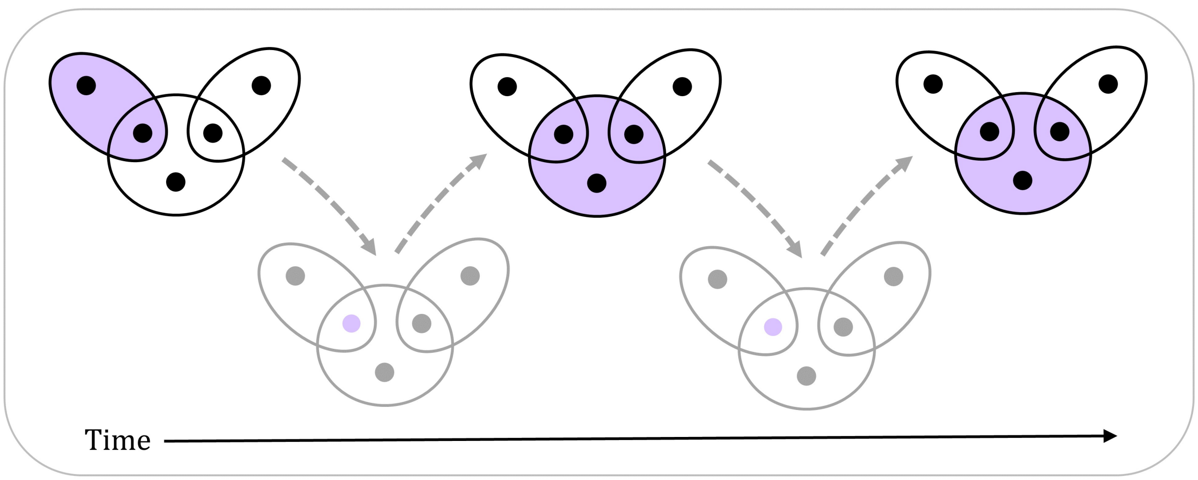


Figure 4: Edge-centric random walk process. Top layer: purple hyperedge is the current position of the random walker. Bottom layer: purple node is the node chosen during the transition. In the context of the TH, the purple hyperedge is the currently-active thought and the purple node is the stimulus between thoughts.

Furthermore, we can consider the stationary distribution of this process (which exists and is unique by Remark 2, as $|E| = m$ is finite).

Theorem 3.2 (**Edge-Centric Random Walk Stationary Distribution**). *Let $\mathbf{T} = (T_{\alpha\beta})$ be the transition matrix of the edge-centric random walk on an hypergraph $H = (V, E, w, \gamma, \lambda)$. The stationary probability of hyperedge $E_\beta \in E$ is*

$$\pi_\beta = \frac{w(\beta)\delta(\beta)}{\sum_\sigma w(\sigma)\delta(\sigma)}. \tag{9}$$

Proof. By definition, the stationary probability of hyperedge $E_\beta \in E$ satisfies

$$\pi_\beta = \sum_\alpha \pi_\alpha T_{\alpha\beta}. \tag{10}$$

Substituting (8) and (9) into the right-hand side of (10) gives

$$\begin{aligned}
\pi_\beta &= \sum_\alpha \pi_\alpha T_{\alpha\beta} \\
&= \sum_\alpha \frac{w(\alpha)\delta(\alpha)}{\sum_\sigma w(\sigma)\delta(\sigma)} \sum_i \frac{\gamma(i)}{\delta(\alpha)} \frac{w(\beta)}{d(i)} e_{i\alpha} e_{i\beta} \\
&= \frac{w(\beta)}{\sum_\sigma w(\sigma)\delta(\sigma)} \sum_i \frac{\gamma(i)}{d(i)} e_{i\beta} \sum_\alpha w(\alpha) e_{i\alpha} \\
&= \frac{w(\beta)}{\sum_\sigma w(\sigma)\delta(\sigma)} \sum_i \frac{\gamma(i)}{d(i)} d(i) e_{i\beta} \\
&= \frac{w(\beta)}{\sum_\sigma w(\sigma)\delta(\sigma)} \sum_i \gamma(i) e_{i\beta} \\
&= \frac{w(\beta)\delta(\beta)}{\sum_\sigma w(\sigma)\delta(\sigma)},
\end{aligned}$$

showing that (8) satisfies (10). By uniqueness of the stationary distribution, (9) holds, as required. □

As we now show, the edge-centric RW is consistent with the node-centric RW, in the sense that the stationary distributions provide the same information: the stationary probability of an hyperedge is equal to the sum of the stationary probabilities of the nodes that make it up. Since a node can belong to multiple hyperedges, we must weight the stationary distribution of a node by a factor proportional to the hyperedge weight. Performing this calculation for hyperedge $E_\alpha \in E$, we have

$$\begin{aligned}
\sum_i \frac{w(\alpha)}{d(i)} \pi_i e_{i\alpha} &= \sum_i \frac{w(\alpha)}{d(i)} \frac{\gamma(i) d(i)}{\sum_k \gamma(k) d(k)} e_{i\alpha} \\
&= \frac{w(\alpha)}{\sum_k \gamma(k) d(k)} \sum_i \gamma(i) e_{i\alpha} \\
&= \frac{w(\alpha)\delta(\alpha)}{\sum_k \gamma(k) d(k)},
\end{aligned} \tag{11}$$

and normalising this gives π_α, as required. In the context of the TH, this is good news: it tells us that both processes defined above provide the same information about the long-term proportion of time spent thinking a particular thought. More generally, such an interplay between the processes suggests that the edge-centric random walk has potential to be used in broader applications of hypergraphs.

3.3 Projecting the Random Walks

Considering the stationary distributions (6) and (9), we see that they depend only on the weights and degrees of, respectively, concepts and thoughts. This leads us to question whether it is possible to define these processes on a standard network in which concepts or thoughts are the nodes, and the transition probabilities between nodes are defined by the TH. Doing so would provide a representation of the TH that is amenable to results extended from network science, whilst still requiring the model to be defined as an hypergraph (i.e. in order to obtain correct transition probabilities).

Such a representation can be found by considering 'projections' of an hypergraph onto its nodes or hyperedges, and seeing whether there exists a RW on the projection which is *equivalent* to that on the hypergraph (has the same state space and same transition probabilities between states). This was explored by Chitra et al. [11], who were interested in understanding whether a RW on the nodes of an hypergraph possessed higher-order properties that prevented it from being equivalent to a RW on a projected network. Although not our focus here, their work serves as useful for answering our question above.

To distinguish the separate projections, we call the projection onto the nodes a *clique graph*, and the projection onto the hyperedges a *contracted network*. The former is defined in [11], and the latter is our own definition. However, we note that the general consideration of a projection of an hypergraph onto its hyperedges is not a new concept (for example, see [16]).

Definition 13 (**Clique Graph** [11])**.** *An hypergraph $H = (V, E, w, \gamma, \lambda)$ defines a clique graph $G^H = (V, \mathcal{E}, \omega)$, where V is the node set, $\mathcal{E} = \{\{v_i, v_j\} : v_i, v_j \in V$ and $v_i, v_j \in E_\alpha$ for some $E_\alpha \in E\}$ is the edge set (including self-loops), and $\omega = \{\omega(e) : e \in \mathcal{E}\}$ is a set of prescribed weights on the edges $e \in \mathcal{E}$.*

Definition 14 (**Contracted Network**)**.** *Given an hypergraph $H = (V, E, w, \gamma, \lambda)$, the corresponding contracted network is $G_c^H = (E, \hat{\mathcal{E}}, \Omega)$, where the node set of G_c^H is the set of hyperedges E of H, the edge set (including self-loops) is $\hat{\mathcal{E}} = \big\{\{E_\alpha, E_\beta\} : E_\alpha, E_\beta \in E, |E_\alpha \cap E_\beta| \neq \emptyset\big\}$ and the edge weights are $\Omega = \{\Omega(e) : e \in \hat{\mathcal{E}}\}$.*

In these definitions, the existence of self-loops reflects the 'laziness' of the RWs. Figure 5 shows a visual representation of the projections. In the context of our TH, the clique graph is a *concept network*, and the contracted network a *thought network*.

These definitions lead us to two theorems which answer our question above: there *do* exist random walks on the projected networks equivalent to those on the hypergraph. The first theorem corresponds to Theorem 4 in [11], and the second is our own. The proofs follow the same structure, so we only prove our theorem here, and refer the reader to Appendix B of [11] for proof of the first.

Theorem 3.3 (**Node Projection** [11])**.** *Let $H = (V, E, w, \gamma, \lambda)$ be an hypergraph. Then there exist edge weights ω on the corresponding clique graph $G^H = (V, \mathcal{E}, \omega)$ such that a random walk on G^H is equivalent to the node-centric random walk on H.*

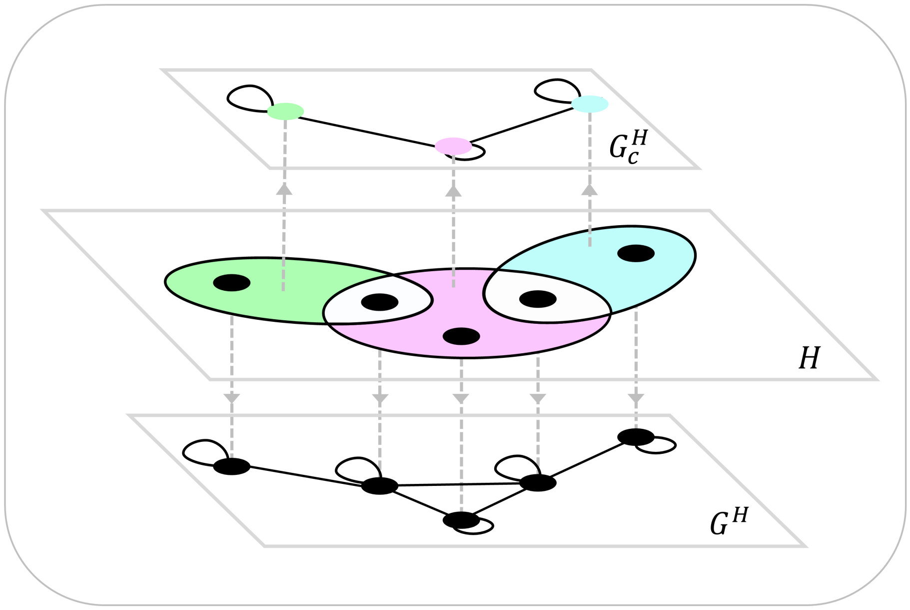


Figure 5: Projections of an hypergraph H down onto its clique graph G^H and up onto its contracted network G_c^H.

Theorem 3.4 (**Hyperedge Projection**). *Let $H = (V, E, w, \gamma, \lambda)$ be an hypergraph. Then there exist edge weights Ω on the corresponding contracted network $G_c^H = (E, \hat{\mathcal{E}}, \Omega)$ such that a random walk on G_c^H is equivalent to the edge-centric random walk on H. Moreover, the weight of edge $\{E_\alpha, E_\beta\} \in \hat{\mathcal{E}}$ is $\pi_\alpha T_{\alpha\beta}$, where π and $\mathbf{T} = (T_{\alpha\beta})$ are respectively the stationary distribution and transition matrix of the edge-centric random walk on H.*

The proof of Theorem 3.4 requires the following Lemma from [11], for which we recall the notion of a reversible Markov chain from SB3.1 Applied Probability.[1]

Lemma 3.5. *Let M be an irreducible Markov chain with finite state space E and transition probabilities $T_{\alpha\beta}$ for $E_\alpha, E_\beta \in E$. M is reversible if and only if there exists a weighted, undirected network G with node set E such that a random walk on G is equivalent to M.*

Proof. Let M be a Markov chain satisfying the conditions of the Lemma, and suppose M is reversible. By irreducibility, M has a unique stationary distribution π, and $\pi_\alpha \neq 0$ for any $E_\alpha \in E$. Let G be a network with nodes E and weight $w_{\alpha\beta} = \pi_\alpha T_{\alpha\beta}$ on each edge $\{E_\alpha, E_\beta\}$. Since M is reversible, $\pi_\alpha T_{\alpha\beta} = \pi_\beta T_{\beta\alpha}$ for all $E_\alpha, E_\beta \in E$. Therefore, $w_{\alpha\beta} = w_{\beta\alpha}$ for all $E_\alpha, E_\beta \in E$, so the edge weights are well-defined. In a random walk on G, the probability of going from node E_α to node E_β in a timestep is

$$\frac{w_{\alpha\beta}}{\sum_\sigma w_{\alpha\sigma}} = \frac{\pi_\alpha T_{\alpha\beta}}{\sum_\sigma \pi_\alpha T_{\alpha\sigma}} = \frac{T_{\alpha\beta}}{\sum_\sigma T_{\alpha\sigma}} = T_{\alpha\beta}, \tag{12}$$

since $\mathbf{T} = (T_{\alpha\beta})$ is right-stochastic. Hence, the transition probabilities and state space of the random walk on G are equal to that of M, so equivalence holds. The converse holds since a random walk on an undirected network is always reversible [17]. $\square$

[1] $\pi_\alpha T_{\alpha\beta} = \pi_\beta T_{\beta\alpha}\ \forall E_\alpha, E_\beta \in E$.

Proof of Theorem 3.4. We first show that an edge-centric random walk on $H = (V, E, w, \gamma, \lambda)$ satisfies the conditions of the Lemma, and then show that it is reversible.

Since H is connected, any random walk on E is irreducible. The state space of the random walk is finite because the number of edges $|E|$ is finite. By Kolmogorov's Theorem [18], reversibility is equivalent to

$$T_{\alpha_1\alpha_2}T_{\alpha_2\alpha_3}\cdots T_{\alpha_k\alpha_1} = T_{\alpha_1\alpha_k}T_{\alpha_k\alpha_{k-1}}\cdots T_{\alpha_2\alpha_1} \tag{13}$$

for any set of k edges $\{E_{\alpha_1}, \ldots, E_{\alpha_k}\}$. Recalling that the transition probabilities are

$$T_{\alpha\beta} = \sum_i \frac{\gamma(i)}{\delta(\alpha)}\frac{w(\beta)}{d(i)}e_{i\alpha}e_{i\beta} = \frac{w(\beta)}{\delta(\alpha)}\sum_i \frac{\gamma(i)}{d(i)}e_{i\alpha}e_{i\beta}, \tag{14}$$

we have (defining $\alpha_{k+1} = \alpha_1$)

$$\begin{aligned}
T_{\alpha_1\alpha_2}T_{\alpha_2\alpha_3}\cdots T_{\alpha_k\alpha_1} &= \left(\frac{w(\alpha_2)}{\delta(\alpha_1)}\sum_i \frac{\gamma(i)}{d(i)}e_{i\alpha_1}e_{i\alpha_2}\right)\cdots\left(\frac{w(\alpha_1)}{\delta(\alpha_k)}\sum_i \frac{\gamma(i)}{d(i)}e_{i\alpha_k}e_{i\alpha_1}\right) \\
&= \prod_{j=1}^{k}\frac{w(\alpha_{j+1})}{\delta(\alpha_j)}\sum_i \frac{\gamma(i)}{d(i)}e_{i\alpha_j}e_{i\alpha_{j+1}} \\
&= \prod_{j=1}^{k}\frac{w(\alpha_j)}{\delta(\alpha_{j+1})}\sum_i \frac{\gamma(i)}{d(i)}e_{i\alpha_{j+1}}e_{i\alpha_j} \\
&= \left(\frac{w(\alpha_k)}{\delta(\alpha_1)}\sum_i \frac{\gamma(i)}{d(i)}e_{i\alpha_1}e_{i\alpha_k}\right)\cdots\left(\frac{w(\alpha_1)}{\delta(\alpha_2)}\sum_i \frac{\gamma(i)}{d(i)}e_{i\alpha_2}e_{i\alpha_1}\right) \\
&= T_{\alpha_1\alpha_k}T_{\alpha_k\alpha_{k-1}}\cdots T_{\alpha_2\alpha_1}.
\end{aligned}$$

So by Kolmogorov's Theorem, an edge-centric random walk on H is reversible.

Hence, by Lemma 3.5, there exists a weighted, undirected network G with with node set E and well-defined edge weights $w_{\alpha\beta} = \pi_\alpha T_{\alpha\beta}$ such that random walks on G and the hyperedges of H are equivalent. Finally, note that the equivalence implies that $w_{\alpha\beta} > 0$ if and only if $T_{\alpha\beta} > 0$, and so G is the contracted network G_c^H, as required. □

3.4 Reflection

We have now defined 'train of thought' processes on our TH. We showed that the stationary distributions of the edge-centric and node-centric processes are consistent, demonstrating that whichever process we use, the long-term proportion of time spent thinking a particular thought is the same. Furthermore, we showed that both processes are equivalent to random walks on their respective projected networks. Such an equivalence is valuable, because it not only makes the RW processes amenable to network-based methods, but it also enables one to focus on different 'levels' of an hypergraph model to suit a particular interest; this could have wider applications in the study of multi-layer networks. These multiple levels will become useful in Section 5 when we consider the recovery process. Overall, we believe the edge-centric random walk process holds exciting potential for further research.

4 Refocusing

In the previous section, we defined train of thought processes, which continued forever in the TH. However, once a concept is stimulated and a train of thought begins, at some point the mind *refocuses* to a thought outside of the TH. Although we don't need to model thoughts outside of the TH, we do need to model the transitions of leaving and re-entering the TH. In this section, we consider these transitions.

4.1 Defining Refocusing

To encorporate refocusing into our RW processes, we take inspiration from the method of teleportation used in the PageRank algorithm [19]: at each step of the RW, with some probability η the random walker follows the RW as usual, and with probability $1-\eta$ the random walker *teleports* independently of the underlying network structure to another node, according to a choice given by a probability distribution $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ called the *preference vector*. Teleportation reflects the concept of refocusing well: during 'teleportation', we assume the mind (random walker) has refocused and is wandering around thoughts outside of the TH (see Figure 6). The node it teleports to corresponds to the node it re-enters the TH on. Similarly to a thought being hidden between transitions in the concept-centric train of thought process, refocusing is hidden in a teleportation transition.

We can define teleportation in terms of both the node-centric (Definition 10) and edge-centric (Definition 12) random walks.

Definition 15 (**Node Teleportation Process**). *Let $1-\eta \in [0,1]$ be the probability of teleportation, and let $\mathbf{u} = (u_1, \ldots, u_n)$ be the $1 \times n$ preference vector of nodes of an hypergraph $H = (V, E, w, \gamma, \lambda)$. At time t, the random walker is at node $v_t = v_i$, and will:*

- *with probability η, perform steps 1–3 in the node-centric random walk process to walk to some node $v_{t+1} = v_j$ at time $t+1$;*
- *with probability $1-\eta$, walk to some node $v_{t+1} = v_j$, chosen with probability u_j.*

This is a Markov chain on V with transition probabilities

$$T_{ij}^{(\eta)} = \eta T_{ij} + (1-\eta) u_j, \tag{15}$$

where T_{ij} is given by (5). The transition probabilities can be encoded in an $n \times n$ node refocusing transition matrix $\mathbf{T}^{(\eta)} = \eta \mathbf{T} + (1-\eta)\mathbf{1}^\top \mathbf{u}$, where $\mathbf{1}$ is a $1 \times n$ vector of ones.

Definition 16 (**Edge Teleportation Process**). *Let $1-\eta \in [0,1]$ be the probability of teleportation, and let $\mathbf{u} = (u_1, \ldots, u_m)$ be the $1 \times m$ preference vector of hyperedges of an hypergraph $H = (V, E, w, \gamma, \lambda)$. At time t, the random walker is at hyperedge $E_t = E_\alpha$, and will:*

- *with probability η, perform steps 1–3 in the edge-centric random walk process to walk to some hyperedge $E_{t+1} = E_\beta$ at time $t+1$;*
- *with probability $1-\eta$, walk to some hyperedge $E_{t+1} = E_\beta$, chosen with probability u_β.*

This is a Markov chain on E with transition probabilities

$$T^{(\eta)}_{\alpha\beta} = \eta T_{\alpha\beta} + (1-\eta)u_{\beta}, \tag{16}$$

where $T_{\alpha\beta}$ is given by (8). The transition probabilities can be encoded in an $m \times m$ hyperedge refocusing transition matrix $\mathbf{T}^{(\eta)} = \eta\mathbf{T} + (1-\eta)\mathbf{1}^{\top}\mathbf{u}$, where $\mathbf{1}$ is a $1 \times m$ vector of ones.

These definitions can be rephrased in terms of the TH by substituting terms, as explained in Table A1.

Remark 3. The teleportation processes can equally be defined on their respective projected networks. This is because the preference vectors $\mathbf{u}$ are independent of the underlying structure of H and its projections, and so since the RW transition matrices are equal on H and the corresponding projection by Theorems 3.3 and 3.4, $\mathbf{T}^{(\eta)} = \eta\mathbf{T} + (1-\eta)\mathbf{1}^{\top}\mathbf{u}$ is equal on H and the corresponding projection too.

Note that $\eta = 0$ refers to no train of thought inside the TH, so that each time we enter the TH we refocus outside of it, while $\eta = 1$ corresponds to no refocusing. This parameter will become central to our later discussion of recovery.

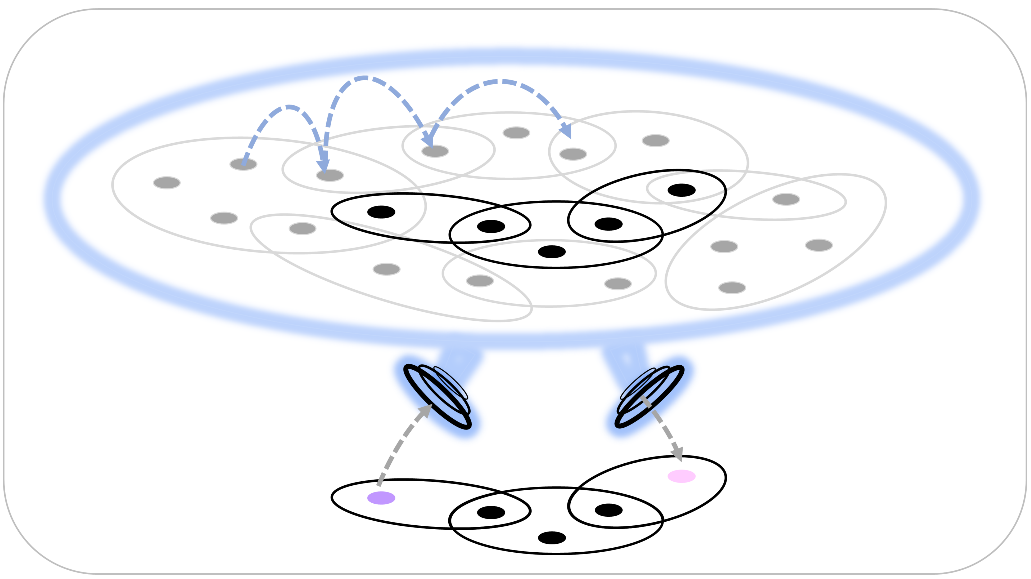


Figure 6: Example of a node teleportation process. The random walker leaves the (thought) hypergraph on the purple node and re-enters on pink node. During the transition, the random walker is assumed to be walking around nodes external to the (thought) hypergraph being considered, which is shown in the blue teleport ring.

4.2 Stationary Distribution with Refocusing

As before, we are interested in the long-term proportion of time spent thinking a particular thought, that is, the hyperedge stationary distribution. In particular, it is interesting to understand how this changes as a function of the refocusing parameter η. The following Theorem tells us the stationary distribution, and we prove it by expanding on Section 2 in [20]. The proof relies on the definition of strict diagonal dominance and the Levy-Desplanques theorem, which we first state and prove for completeness.

Definition 17 (**Strictly Diagonally Dominant** [21])**.** *An $m \times m$ matrix $\mathbf{A} = (A_{ij})$ is strictly diagonally dominant if*

$$|A_{ii}| > \sum_{j \neq i} |A_{ij}| \;\; \forall i \in \{1, \ldots, m\}. \tag{17}$$

Lemma 4.1 (**Levy-Desplanques Theorem** [21])**.** *Let $\mathbf{A} = (A_{ij})$ be a strictly diagonally dominant $m \times m$ matrix. Then $\mathbf{A}$ is non-singular.*

Proof. Suppose that $\mathbf{A}$ is singular. Then there exists a non-zero vector $\mathbf{v}$ such that $\mathbf{A}\mathbf{v} = \mathbf{0}$, so $\sum_{j=1}^{m} A_{ij} v_j = 0$ for all $i \in \{1, \ldots, m\}$. Let $|v_M| = \max\{|v_1|, \ldots, |v_m|\}$. This implies $|A_{MM}||v_M| = |A_{MM} v_M| = |\sum_{j \neq M} A_{Mj} v_j| \leq \sum_{j \neq M} |A_{Mj}||v_j| \leq |v_M| \sum_{j \neq M} |A_{Mj}|$, where the first inequality uses the triangle inequality, and the second inequality is by definition of $|v_M|$. This provides $|A_{MM}| \leq \sum_{j \neq M} |A_{Mj}|$, contradicting strict diagonal dominance. $\square$

Theorem 4.2. *Let $\mathbf{u} = (u_1, \ldots, u_m)$ be the hyperedge preference vector, and let π_β and $\mathbf{T}$ be, respectively, the stationary probability of hyperedge $E_\beta \in E$ and the transition matrix of the edge-centric random walk on $H = (V, E, w, \gamma, \lambda)$. Furthermore, let $\mathbf{I}$ be the $m \times m$ identity matrix. The stationary probability $\pi_\beta^{(\eta)}$ of hyperedge $E_\beta \in E$ under the edge teleportation process is*

$$\pi_\beta^{(\eta)} = \begin{cases} u_\beta & \text{if} \quad \eta = 0, \\ \pi_\beta & \text{if} \quad \eta = 1, \\ (1-\eta) \sum_\alpha u_\alpha \big[(\mathbf{I} - \eta \mathbf{T})^{-1} \big]_{\alpha\beta} & \text{if} \quad \eta \in (0, 1). \end{cases} \tag{18}$$

Proof. Firstly, note that for each $\eta \in [0, 1]$, the stationary distribution $\pi^{(\eta)}$ is unique, since H is connected and hence the random walks are irreducible. The stationary distribution $\pi^{(\eta)}$ satisfies

$$\pi^{(\eta)} = \pi^{(\eta)} \mathbf{T}^{(\eta)}, \tag{19}$$

and by definition of $\mathbf{T}^{(\eta)}$ this becomes

$$\pi^{(\eta)} = \eta \pi^{(\eta)} \mathbf{T} + (1 - \eta) \mathbf{u}, \tag{20}$$

since $\pi^{(\eta)} \mathbf{1}^\top = 1$.

- For $\eta = 0$, (20) becomes $\pi^{(0)} = \mathbf{u}$. By uniqueness, $\pi^{(0)} = \mathbf{u}$.
- For $\eta = 1$, (20) becomes $\pi^{(1)} = \pi^{(1)} \mathbf{T}$, which is satisfied by the stationary distribution of the edge-centric random walk π. By uniqueness, $\pi^{(1)} = \pi$.

- For $\eta \in (0,1)$, rearranging (20) gives

$$\pi^{(\eta)}(\mathbf{I} - \eta\mathbf{T}) = (1-\eta)\mathbf{u}, \tag{21}$$

We now note that $(\mathbf{I} - \eta\mathbf{T})$ is strictly diagonally dominant: $|(\mathbf{I} - \eta\mathbf{T})_{ii}| = |1 - \eta T_{ii}| = 1 - \eta T_{ii}$ since $0 < \eta T_{ii} < 1$. Furthermore, $\sum_{j\neq i} |(\mathbf{I} - \eta\mathbf{T})_{ij}| = \sum_{j\neq i} |0 - \eta T_{ij}| = \sum_{j\neq i} \eta T_{ij} = \eta(1 - T_{ii})$, where the third equality holds since the row-sum of $\mathbf{T}$ is 1. Since $\eta < 1$, we have $\eta(1 - T_{ii}) \leq 1 - \eta T_{ii}$, and so $\mathbf{I} - \eta\mathbf{T}$ is strictly diagonally dominant. By Lemma 4.1, $(\mathbf{I} - \eta\mathbf{T})^{-1}$ exists, so we can right-multiply (21) by this inverse to get

$$\pi^{(\eta)} = (1-\eta)\mathbf{u}(\mathbf{I} - \eta\mathbf{T})^{-1}. \tag{22}$$

Therefore,

$$\pi^{(\eta)}_\beta = (1-\eta)\sum_\alpha u_\alpha \left[(\mathbf{I} - \eta\mathbf{T})^{-1}\right]_{\alpha\beta}. \tag{23}$$

□

To understand the long-term proportion of time spent thinking a certain thought, we need to choose a specific form of preference vector. We know a thought is activated by a stimulated concept, and a concept is chosen with probability proportional to its weight. Therefore, a thought is more likely to occur if the total weight of its concepts is high (i.e. if its degree is high). Hence, we choose the preference vector to be proportional to hyperedge degrees, giving $u_\alpha = \delta(\alpha)/\sum_\sigma \delta(\sigma)$ for each $E_\sigma \in E$. This means the stationary distribution (18) becomes

$$\pi^{(\eta)}_\beta = \begin{cases} \frac{\delta(\beta)}{\sum_\alpha \delta(\alpha)} & \text{if} \quad \eta = 0, \\ \frac{w(\beta)\delta(\beta)}{\sum_\alpha w(\alpha)\delta(\alpha)} & \text{if} \quad \eta = 1, \\ (1-\eta)\sum_\alpha \frac{\delta(\alpha)}{\sum_\sigma \delta(\sigma)}\left[(\mathbf{I} - \eta\mathbf{T})^{-1}\right]_{\alpha\beta} & \text{if} \quad \eta \in (0,1). \end{cases} \tag{24}$$

So, unless all hyperedges are the same size, the stationary distribution varies with η. In terms of thinking, this demonstrates that our ability to refocus affects which thoughts are more common in the long-run. In particular, when η is close to 1 we rarely refocus, and in the long-run we are more likely to be thinking of thoughts of large size (i.e. containing a large number of concepts). Therefore, if one does not refocus often, it becomes easier to fixate on large thoughts. This will be important in Section 5.

Remark 4. Theorem 4.2 can equivalently be stated for $\pi^{(\eta)}_i$ in terms of the node teleportation process, with $\mathbf{u} = (u_1, \dots, u_n)$ and indices i, j instead of α, β. Moreover, the preference vector is the vector of node degrees, for the same reason we chose the hyperedge preference vector.

4.3 Reflection

Overall, this section saw us introduce the process of 'refocusing' via teleportation on hypergraphs. This work may find broader applications in hypergraph science. For example, if an hyperedge were to represent a community of nodes, using the hyperedge stationary distributions we could rank whole communities, rather than just ranking nodes. In the context of our model, we found that our ability to refocus determines which thoughts become more common in the long-run, and that infrequent refocusing can lead to fixation. Next, we introduce reappraisal and the full recovery process.

5 Recovery

Now that we have seen how to traverse our TH and refocus, we consider reappraisal and the recovery process. This can be explained as follows [1]: the series of trains of thought are taking place on our TH, and we have no control over which thoughts emerge into our conscious awareness. At some point, we may engage in mindful awareness (pay attention to our thoughts non-judgmentally). By doing so, we may identify an unhelpful or maladaptive thought. We can then choose to use our executive attention to focus on that thought and reappraise it by assigning it a more helpful meaning. Having attempted to do so, we want to refocus our mind to outside of the TH.

We now proceed to define this process in terms of our model.

5.1 Reappraisal

We can introduce reappraisal by defining a process similar to the mathematical notion of a RW marking process on a graph, in which a random walker 'marks' each node v_i it visits with some probability p_i until all nodes are marked [22]. The concept of 'marking' can be defined in terms of binary node states: a node is marked if its state is set equal to 1, and unmarked if its state is 0. As our random walks are defined on hypergraphs, the marking probabilities can depend on hyperedges as well as nodes.

Definition 18 (**Reappraisal**). *Suppose that a node-centric random walk is talking place on an hypergraph $H = (V, E, w, \gamma, \lambda)$. Say that at time t, the random walker is at node $v_t = v_i$ and has selected hyperedge $E_\alpha \in E$. Then with probability $p_{i,\alpha}$, called the reappraisal probability, the random walker sets $\lambda_i = 1$. Otherwise, the random walker leaves λ_i unchanged.*

This is a marking process on an hypergraph. In terms of the TH, each concept v_i is reappraised with probability $p_{i,\alpha}$ each time it acts as a stimulus for thought E_α. A simple choice is to let the reappraisal probabilities $p_{i,\alpha} = p_i$ be *thought-independent* (i.e. not depend on the stimulated thought $E_\alpha \in E$). However, we can also let the reappraisal probabilities depend on the process.

Definition 19 (**Thought-Dependent Reappraisal Probability**). *The thought-dependent reappraisal probability of thought $E_\alpha \in E$ stimulated by concept $v_i \in V$ is*

$$p_{i,\alpha} = \Lambda_\alpha, \tag{25}$$

where we recall that $\Lambda_\alpha = \frac{\sum_i \lambda_i \gamma(i) e_{i\alpha}}{\delta(\alpha)}$.

That is, $p_{i,\alpha}$ is proportional to the total weight of adaptive concepts in E_α. This definition makes intuitive sense; the reappraisal probability is higher the more adaptive the thought.

5.2 Recovery Process

Given Definition 18, we are now in a position to define the overall recovery process.

Definition 20 (**Recovery Process**). *Let $H = (V, E, w, \gamma, \lambda)$ be a TH, and suppose it is maladaptive at time $t = 0$. Let $p_{i,\alpha}$ be the reappraisal probability of $v_i \in E_\alpha$, η the refocusing parameter and $\mathbf{u} = (u_1, \ldots, u_n)$ the concept preference vector. At time t:*

1. *Thought: the current stimulus $v_t = v_i$ activates one thought $E_\alpha \in E$ containing concept v_i. Thought E_α is chosen with probability $w(\alpha)/d(i)$.*
2. *Reappraisal: with probability $p_{i,\alpha}$ the state of $v_i \in E_\alpha$ is set to be adaptive ($\lambda_i = 1$).*
3. *Refocusing: with probability $1 - \eta$, one concept $v_k \in V$ is chosen with probability u_k to be the stimulus at time $t + 1$. Otherwise, the train of thought continues inside the TH: thought E_α triggers a concept $v_j \in E_\alpha$, chosen with probability $\gamma(j)/\delta(\alpha)$, to act as the stimulus at time $t + 1$.*

The recovery process is complete when H is adaptive.

Therefore, the recovery process is the node teleportation process (Definition 15) with reappraisal (Definition 18). Note that we have defined the recovery process in terms of the node-centric random walk, as it is the states of the concepts that are being reappraised. By doing so, we can observe the recovery process at both the concept and thought level (i.e. on both projected networks), since the state of a thought is a function of the states of its concepts. Therefore, despite the recovery process being defined on the concepts, it can be interpreted as changing our *thoughts* from maladaptive to adaptive. Figure 7 shows this visually.

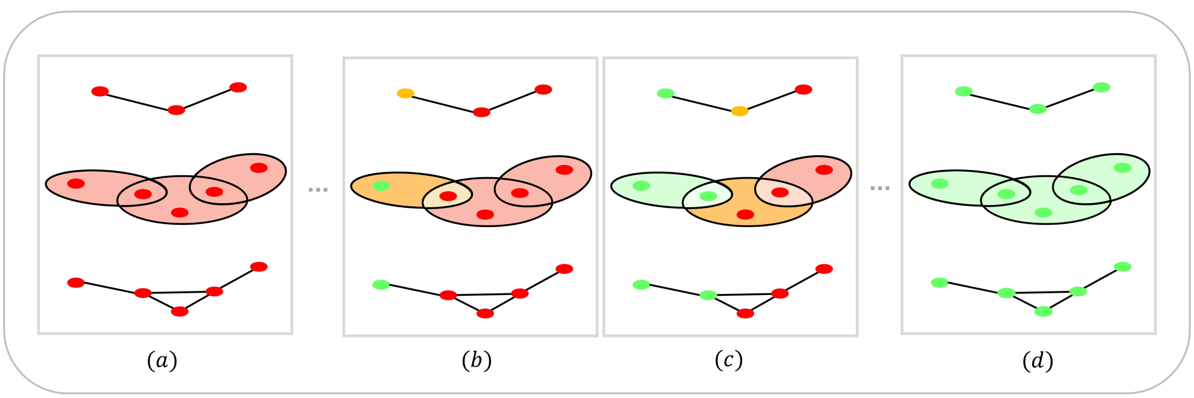


Figure 7: Visual explanation of the recovery process. Top is the contracted network, middle is the TH and bottom is the clique graph. (Self-loops have been removed for visual clarity.) Red is maladaptive, orange is conflicted and green is adaptive. The aim is to turn the TH from red to green. (a) The TH is initially maladaptive. (b) After some time, one concept has been reappraised. The thought it is contained in is now conflicted. (c) Both concepts in the left hyperedge have been reappraised and are now adaptive. The left hyperedge is now adaptive and the middle one conflicted. (d) The recovery process ends when the TH is adaptive. (See Appendix for a colour blind friendly version.)

Since the process is ergodic,[2] provided all reappraisal probabilities are positive, the probability of recovery is 1. Therefore, our model portrays the positive message that recovery *is* possible. However, the nature of the recovery process is affected by the reappraisal probabilities and the refocusing parameter. We want to see whether our model is consistent with the behaviour expected from clinical theory [2, 3]: a high reappraisal probability and frequent refocusing leads to more successful recovery. In the next few subsections, we define the two quantities which we use to assess this success.

5.3 Recovery Time

The most obvious quantity to assess the recovery process by is the *recovery time*: the number of thoughts activated inside the TH during the recovery process before the TH becomes adaptive. Intuitively, the quicker the recovery, the more successful it is. Since recovery occurs by reappraisal, which is a marking process, recovery time is synonymous with *marking time*. Moreover, since the recovery process is stochastic, we are interested in the expected recovery time (ERT).

Definition 21 (**Expected Marking Time (EMT)**). *Let $\mathcal{T}(\mathbf{p})$ denote the marking time of a RW marking process on n nodes $V = \{v_1, \dots, v_n\}$ with marking probabilities $\mathbf{p}$, and let π be the stationary distribution of the RW. The EMT is*

$$\mathbb{E}(\mathcal{T}(\mathbf{p})) = \sum_i \pi_i \mathbb{E}_i(\mathcal{T}(\mathbf{p})), \tag{26}$$

where $\mathbb{E}_i(\mathcal{T}(\mathbf{p}))$ is the EMT for marking processes starting from node $v_i \in V$.

In terms of the recovery process, this expression is the ERT. Therefore, we obtain the ERT by the total law of probability (because the recovery process must begin at some node), and we assume that the RW has been occurring for a sufficiently long time before the marking process (i.e. recovery process) begins, so that the probability of starting at node $v_i \in V$ is π_i. This reflects the existence of maladaptive thoughts long before recovery begins.

Marking times were explored by Banderier and Dobrow [22], as a generalisation of cover times of RWs graphs (the time it takes for a random walker to visit all nodes in a graph, maximised over all nodes [23]). Cover times have many practical applications, from biology to computer science, and have hence received a lot of attention in the mathematical literature [24]. Despite this, the majority of work has involved finding bounds on the expected cover time [17], with very few explicit results existing [24]. However, using probability generating functions allows some progress towards an explicit formula for the EMT (and hence the ERT) to be made [22, 25].

Since the theory in [22] is based on standard networks, the marking probabilities depend only on the nodes. Therefore, the formula we give below holds only for thought-independent reappraisal probabilities (i.e. $p_{i,\alpha} = p_i$ for all $v_i \in V$). Nevertheless, it serves useful to understand the simplest case of our recovery process.

We now introduce the formula for the EMT from [22]. The length of the proof prevents us from presenting it in its entirety, however in the Appendix we provide a sketch-proof to give intuition behind the formula and explain crucial steps omitted in [22]. Before stating the theorem, we introduce the operator $\mathcal{L}_S$.

[2] Irreducible by connectedness of H and aperiodic by RW laziness.

Definition 22 (**Operator $\mathcal{L}_S$** [22])**.** *Let $G = (V, E)$ be a network with n nodes $V = \{v_1, \ldots, v_n\}$. Let $\{x_1, \ldots, x_n\}$ be real variables and $S \subseteq \{1, \ldots, n\}$. Denote the marking probability of node $v_i \in V$ by p_i, and let $q_i = 1 - p_i$. For each $l \in \{1, \ldots, n\}$, $\mathcal{L}_S$ substitutes x_l for q_l if $l \in S$, and otherwise substitutes x_l for* 1*.*

Remark 5. In the following, we denote the size of a subset S by $|S|$, and $\mathbf{I}$ is the $n \times n$ identity matrix. The network $G = (V, E)$ has n nodes $V = \{v_1, \ldots, v_n\}$. The vector $\tilde{\mathbf{p}} = (p_1, \ldots, p_n)$ denotes the thought-independent reappraisal probabilities[3] (i.e. marking probabilities), and $\mathbf{T}$ is an $n \times n$ RW transition matrix. Furthermore, $\mathbf{X} = \mathrm{Diag}(x_1, \ldots, x_n)$ is an $n \times n$ diagonal matrix whose entries $\mathcal{L}_S$ acts on.

Theorem 5.1 (**EMT starting from node v_i** [22])**.** *The expected marking time of $G = (V, E)$ for a random walk starting at node $v_i \in V$ is*

$$\mathbb{E}_i\big(\mathcal{T}(\tilde{\mathbf{p}})\big) = \sum_{\substack{S \subseteq \{1, \ldots, n\}, \\ S \neq \emptyset}} (-1)^{|S|+1} \mathcal{L}_S \left(\sum_{j=1}^{n} x_i \big[\big(\mathbf{I} - \mathbf{T}\mathbf{X}\big)^{-1} \big]_{ij} \right) \tag{27}$$

Proof. See Appendix and [22]. □

This theorem allows us to introduce our own corollary.

Corollary 5.1.1 (**EMT**)**.** *The expected marking time of $G = (V, E)$ is*

$$\mathbb{E}\big(\mathcal{T}(\tilde{\mathbf{p}})\big) = \sum_{\substack{S \subseteq \{1, \ldots, n\}, \\ S \neq \emptyset}} (-1)^{|S|+1} \mathcal{L}_S \left(\sum_{i} \pi_i \sum_{j=1}^{n} x_i \big[\big(\mathbf{I} - \mathbf{T}\mathbf{X}\big)^{-1} \big]_{ij} \right) \tag{28}$$

Proof. Substituting (27) into our definition of the EMT (26) yields (28). □

Corollary 5.1.1 brings us to the ERT of a TH.

Theorem 5.2 (**ERT with thought-independent reappraisal probabilities**)**.** *Let $H = (V, E, w, \gamma, \lambda)$ be a maladaptive thought network, and let $\tilde{\mathbf{p}} = (p_1, \ldots, p_n)$ be thought-independent reappraisal probabilities. Denote the $n \times n$ node refocusing transition matrix by $\mathbf{T}^{(\eta)}$. Then the expected recovery time of H is*

$$\mathbb{E}\big(\mathcal{T}(\tilde{\mathbf{p}})\big) = \sum_{\substack{S \subseteq \{1, \ldots, n\}, \\ S \neq \emptyset}} (-1)^{|S|+1} \mathcal{L}_S \left(\sum_{i} \pi_i \sum_{j=1}^{n} x_i \big[\big(\mathbf{I} - \mathbf{T}^{(\eta)}\mathbf{X}\big)^{-1} \big]_{ij} \right) \tag{29}$$

Proof. Consider the concept network G^H from Definition 13. By Theorem 3.3 the random walk on G^H is equivalent to the node-centric random walk on H, so has the same transition matrix $\mathbf{T}$. By Remark 3, the random walk on G^H with refocusing has transition matrix $\mathbf{T}^{(\eta)} = \eta \mathbf{T} + (1 - \eta)\mathbf{1}^\top \mathbf{u}$. Applying Corollary 5.1.1 to G^H with $\mathbf{T}^{(\eta)}$ yields (29). □

[3]The tilde is to specify thought-independence.

5.4 Efficiency

Although the ERT provides a good quantitative measure of recovery, it fails to capture details about the process. Given that the aim of recovery is to reappraise maladaptive concepts, time spent stimulating adaptive concepts is in-effect wasting time and energy. Since recovery already requires a lot of energy [3], this is less than ideal. This suggests measuring a notion of efficiency.

Definition 23 (Efficiency). *The efficiency of the recovery process is*

$$\textit{Efficiency} = 1 - \frac{\textit{expected time spent stimulating adaptive concepts}}{\textit{expected recovery time}}, \tag{30}$$

where 'time' is the number of steps in the random walk.

Higher efficiency means proportionally more time is dedicated to trying to reappraise maladaptive thoughts, rather than fixating on already-reappraised ones; more expended energy contributes towards recovery.

5.5 Reflection

We have now introduced the full recovery process for our model. We were able to produce a formula for the ERT for the simplest case of the recovery process, which may find wider applications in the theoretical study of marking processes on hypergraphs. However, the more general case of thought-dependent reappraisal probabilities eludes such a formula, since the thoughts are latent to the concept network. Moreover, (29) provides little insight into the ERT, since it depends on the specific transition matrix $\mathbf{T}^{(\eta)}$, and an explicit formula requires much unwieldy algebra for even the smallest of hypergraphs. Therefore, we must resort to numerical simulations to gain insight into the qualitative behaviour of the ERT and efficiency of our recovery process. This is our focus for the next section.

6 Numerical Simulations

Having defined our recovery process and two measures to quantify its success, we perform numerical simulations. Before we doing so, we must generate hypergraphs on which to perform our simulations.

6.1 Generating Hypergraphs

Since thought is hard to observe [12] and "language is a system for expressing thought" [26], cognitive semantic networks, in which nodes represent words and edges represent relationships between them, have been used to explore thinking processes [27]. Analysis of these networks has suggested they exhibit a power-law degree-distribution (of the form $f(k) \propto k^{-\kappa}$) [28]; a distribution well-known to be generated in networks via growth and preferential attachment mechanisms, in which new nodes attach preferentially to nodes with high degree [29]. In particular, cognitive semantic networks have been suggested to have power-law exponent $\kappa = 3$ [27, 28]. Preferential attachment and power-law behaviour therefore motivate our mechanism for generating hypergraphs.

A variety of preferential attachment mechanisms have been explored for generating random hypergraphs [16, 30, 31], and their degree-distributions verified analytically as power-law. (See Table A3 for an overview.) All existing mechanisms that add hyperedges at each growth step preferentially choose nodes to include in each newly added hyperedge. However, we devised an algorithm that chooses an hyperedge preferentially based on hyperedge degree, before choosing nodes inside it uniformly at random (see Figure 8). This ensures the generated hypergraph is simple and connected. Furthermore, we state as an input the hyperedge degree-distribution (specifically, power-law with exponent $\kappa = 3$ to fit with the cognitive semantic network literature [28]). We explain our algorithm in more detail in the Appendix.

Algorithm 2: Preferential Hyperedge Growth Mechanism

Inputs: $D = [|E_1|, \ldots, |E_m|]$ (hyperedge size distribution), $m = \text{len}(D)$.
Output: Hypergraph H.
Initialise: Hypergraph H contains $|E_1|$ nodes in one hyperedge, E_1.
for $t = 1$ *to* $m - 1$ **do**
  Hypergraph Growth:
  Choose one hyperedge E_u in H with probability proportional to degree, $\delta(u)$;
  Choose l uniformly from $\{1, \ldots, \min\{|E_u| - 1, |E_{t+1}| - 1\}\}$;
  Choose l nodes uniformly in E_u;
  Hyperedge E_{t+1} contains these l nodes and $|E_{t+1}| - l$ new nodes;
  Add E_{t+1} to H.
end

Figure 8: Preferential hyperedge growth mechanism for generating hypergraphs. Note that it is called Algorithm 2 because Algorithm 1 calculates the power-law hyperedge size distribution $D = [|E_1|, \ldots, |E_m|]$, which is the input to this algorithm (see Figure A4).

Remark 6. Our algorithm generates hypergraphs in which $\gamma(i) = 1$ for all $v_i \in V$. So by construction, $\delta(\alpha) = |E_\alpha| = w(\alpha)$ for all $E_\alpha \in E$. Assigning node weights would interfere with the power-law hyperedge degree-distribution. To avoid this, we only consider $\gamma(i) = 1$ for all $v_i \in V$.

Remark 7. Figure 9 shows the average of 100 realisations of the node degree-distribution for hypergraphs with $10,000$ hyperedges on a log-log plot. The straight-line behaviour is characteristic of a power-law, however is only suggestive of such a node degree-distribution, and does not constitute a proof [32]. Since this distribution is not the focus of our dissertation, we suffice in concluding that it is heavy-tailed. We also remark that in the semantic network literature, verification of power-law distributions was based on plots [28], so may not be entirely reliable [32].

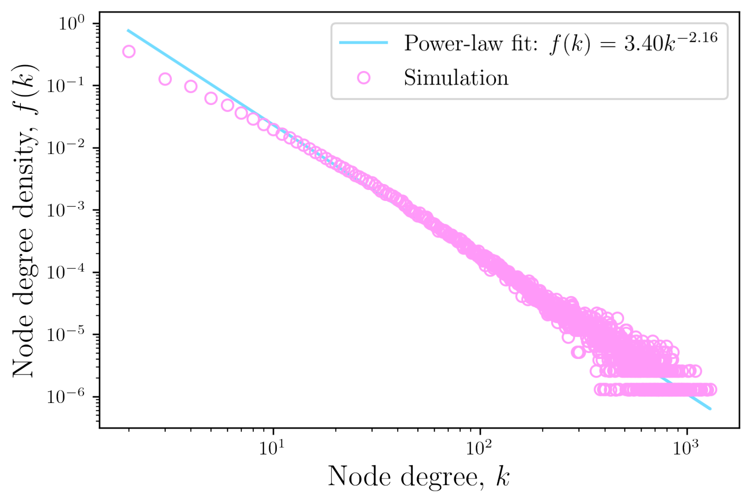


Figure 9: Log-log plot of node degree-distribution of hypergraphs with $10,000$ hyperedges and hyperedge power-law degree-distribution with exponent $\kappa = 3$. Generated using our preferential hyperedge attachment mechanism and averaged over 100 realisations. Plotted against a fitted power-law.

With hypergraphs generated, we can progress to our main focus for this section: simulating the recovery process.

6.2 Simulating the Recovery Process

We begin by verifying that (29) coincides with an average over realisations of the recovery process on random hypergraphs generated using our algorithm. Figure 10 demonstrates that this is the case.

We now want to understand whether our recovery process emulates the two characteristic features of successful SDN, which we recall from the introduction: the better you are able to (1) reappraise maladaptive thoughts and (2) refocus away from such thoughts, the more successful the recovery process. The former can be explored by varying the marking probabilities $p_{i,\alpha}$ of nodes, and the latter by varying the refocusing parameter η. Recall that (29) only holds for thought-independent marking probabilities $p_{i,\alpha} = p_i$, and so for thought-dependent marking probabilities we must resort to taking ensemble averages.

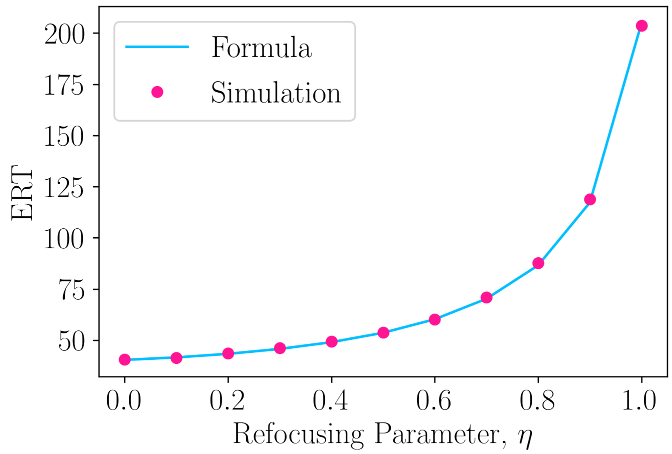
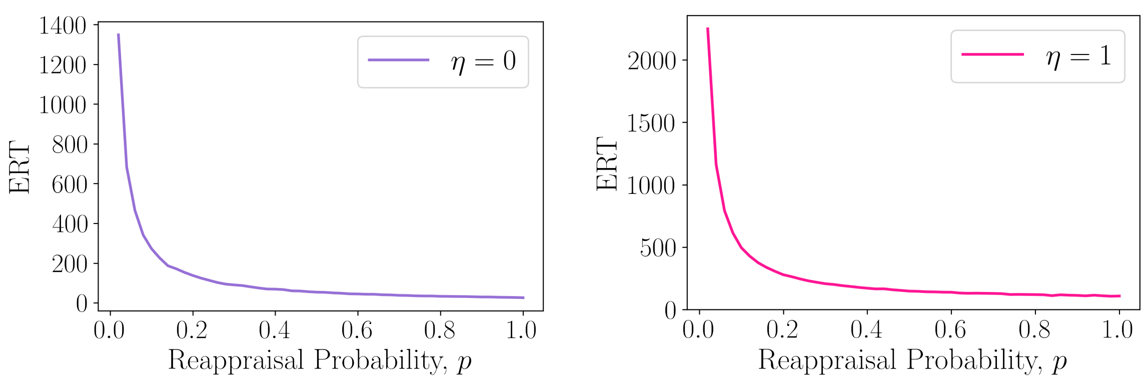


Figure 10: Simulation of recovery process on a TH with $m = 6$ and $\kappa = 3$ has ERT that matches the formula (29) in Theorem 5.2. Here, we chose $p_{i,\alpha} = 1$ for all $v_i \in V$. The simulation is averaged over 200 realisations of the recovery process starting from each node.

Remark 8. In this section, we perform simulations on hypergraphs with six hyperedges. To show the results are not specific to hypergraphs of this size, we include figures for hypergraphs of various sizes in the Appendix.

6.2.1 Reappraisal

In Figure 11, we plot the ERT for constant reappraisal probabilities (i.e. $p_i = p \in (0, 1]$ for all $v_i \in V$). As anticipated, the ERT is a decreasing function of p: as the probability of reappraisal increases, and so the individual is better at reappraising their thoughts, recovery is quicker. This fits with our mathematical intuition about marking processes, since if the probability of marking nodes is higher, we expect all nodes to be marked in a shorter amount of time. Overall, this suggests our model emulates characteristic 1.



(a) (b)

Figure 11: ERT for constant reappraisal probabilities $p \in (0, 1]$ averaged over 200 realisations of the recovery process starting from each node, on an hypergraph with $m = 6$ and $\kappa = 3$. (a) Refocusing parameter $\eta = 0$. (b) Refocusing parameter $\eta = 1$. (Different refocusing parameters shown to demonstrate that qualitative behaviour is independent of η.)

6.2.2 Refocusing

We first use the ERT to evaluate the dependence of recovery on the refocusing parameter η. As Figure 12 shows, for thought-independent reappraisal probabilities the ERT is a monotonic increasing function of η. This suggests that the higher the probability of refocusing, the quicker the recovery process. However, the thought-dependent case behaves slightly differently: the ERT is again high for large η but no-longer monotonic for small η, implying that a balance is required between paying attention to maladaptive thoughts for long enough to enable recovery, but not too long so as to fixate. Overall, the dramatic increase for large η but relatively low values for small η implies that more refocusing is far better than none.

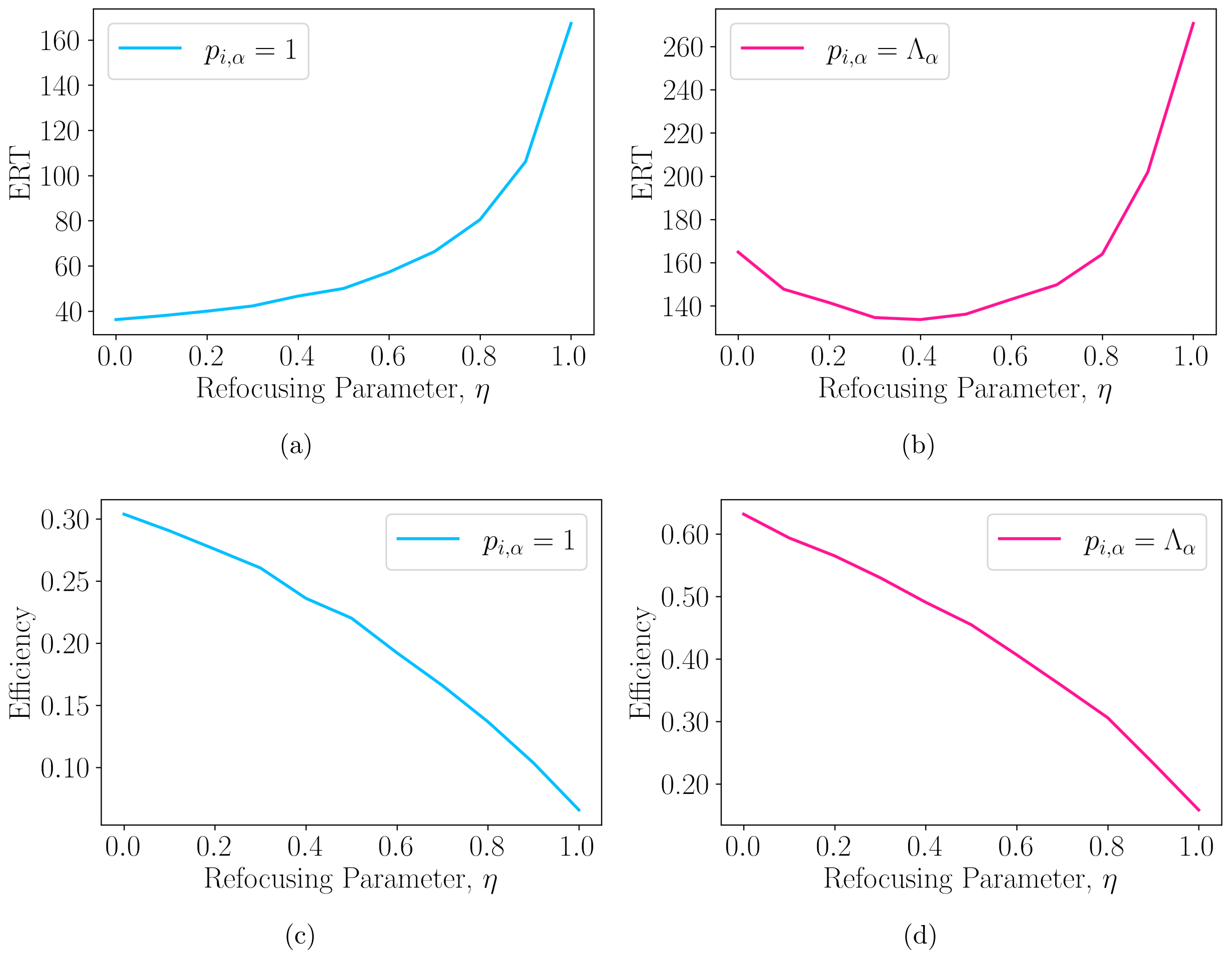


Figure 12: Recovery process for an hypergraph with $m = 6$ and $\kappa = 3$, averaged over 200 realisations per starting node. Blue represents thought-independent reappraisal probabilities and pink represents thought-dependent reappraisal probabilities. (a) ERT increases with refocusing parameter η. (b) ERT is very high for high η, and low for small η. (c) Efficiency is a decreasing function of η, so the recovery process is more efficient the more frequently refocusing occurs. (d) Efficiency is again a decreasing function of η. Overall, the thought-independent case suggests lots of refocusing is optimum for both the ERT and efficiency, whereas the thought-dependent case suggests a slight trade-off is required between minimising and the ERT and maximising efficiency.

Considering instead the efficiency, both the thought-independent and thought-dependent cases suggest that more frequent refocusing increases efficiency (see Figure 12). Hence, if we evaluate the recovery process based on efficiency, it suggests that our model emulates the refocusing characteristic of SDN.

6.3 Reflection

Overall, our simulations demonstrate that the speed of recovery increases with reappraisal ability, and that efficiency increases with refocusing frequency. Using these as the criteria for more successful recovery, our simulations exhibit the two characteristics of SDN we hoped for. However, they don't provide proof that our model emulates such characteristics. We discuss this limitation in the next section, where we conclude our dissertation by reviewing our work, discussing the limitations of our model, and suggesting possible future directions.

7 Discussion

In this dissertation, we developed a conceptual model representing the thinking process behind recovery from mental illness using SDN. Our motivation stemmed from the need for individuals to have sufficient metacognitive awareness in order to engage successfully in such a process [2, 4]. Since visual explanations improve conceptual understanding [5, 6], we wanted to see whether a simple visual dynamical model could emulate key characteristics of successful SDN: the better able we are to reappraise thoughts and refocus, the less we get stuck unnecessarily fixating on thoughts during recovery and the quicker the recovery process overall. By defining a recovery process based on random walks on hypergraphs, and performing numerical simulations, we provided evidence suggesting our model emulates these characteristics.

Although our numerical simulations showed what we were hoping for, they in no way constitute a proof that our model definitively exhibits the characteristics we wanted it to display. The lack of an analytical proof of the qualitative behaviour of recovery is a major limitation our model, and renders us unable to verify any general features of the dynamical process we defined. At best, we have shown that the model emulates the desired characteristics for the specific hypergraph configurations we used. However, as explicit results concerning marking times and cover times for networks are something that have so far eluded the mathematical community [24, 25], numerical simulations are the best we can do in this situation.

Furthermore, we have focused on dynamics *on* hypergraphs, but it would be interesting for any future models to explore dynamics *of* hypergraphs. This could, for example, incorporate Hebbian mechanisms for the generation of new thoughts and degradation of others, potentially using our random hypergraph generation algorithm. This would make the model better reflect true thinking processes; something our model somewhat over-simplifies.

Whilst developing our model, we also contributed to the development of hypergraph theory. We first introduced a novel dynamical process (edge-centric random walk on an hypergraph) and found its stationary distribution. Following this, we defined a certain projection of an hypergraph onto its hyperedges, the contracted network, and proved that random walks on edges of an hypergraph and its corresponding contracted network are equivalent. We also introduced a marking process for hypergraphs (although found no analytical results for it). Finally, we introduced a novel mechanism for generating random hypergraphs, based on preferential attachment to hyperedges rather than nodes. As it was not the focus of our dissertation, it would be interesting for future work to analytically derive the node degree distribution generated from this mechanism, and see whether such a mechanism has any applicability to the generation of empirical hypergraphs.

Despite its limitations, we hope our work provides some motivation for the creation of a visual model to help individuals develop an adequate metacognitive awareness in the context of recovery from mental illness. Having said this, the elusiveness of any model representing recovery from mental illness in the literature to date may be a reflection of the perceived utility of such a model in clinical practice. Therefore, we make no claim that our model is useful, but instead appreciate it as a tool for allowing us the opportunity to explore dynamics on hypergraphs.

References

[1] Aidan J Flynn. *Using Semantic Network Structure to Understand Repetitive Negative Thinking*. PhD thesis, Villanova University, 2020.

[2] Tim Klein, Beth Kendall, and Theresa Tougas. Changing brains, changing lives: Researching the lived experience of individuals practicing self-directed neuroplasticity. 2019.

[3] Jeffrey M Schwartz and Sharon Begley. *The mind and the brain*. Springer Science & Business Media, 2009.

[4] David Kronemyer and Alexander Bystritsky. A non-linear dynamical approach to belief revision in cognitive behavioral therapy. *Frontiers in Computational Neuroscience*, 8:55, 2014.

[5] Kyung J Min, John K Jackman, and Jason CK Chan. Visual models for abstract concepts towards better learning outcomes and self-efficacy. 2014.

[6] Eliza Bobek and Barbara Tversky. Creating visual explanations improves learning. *Cognitive research: principles and implications*, 1(1):1–14, 2016.

[7] John D Murray, Murat Demirtaş, and Alan Anticevic. Biophysical modeling of large-scale brain dynamics and applications for computational psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(9):777–787, 2018.

[8] Denny Borsboom. A network theory of mental disorders. *World psychiatry*, 16(1):5–13, 2017.

[9] Timoteo Carletti, Duccio Fanelli, and Renaud Lambiotte. Random walks and community detection in hypergraphs. *Journal of Physics: Complexity*, 2021.

[10] Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, and Giovanni Petri. Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, 874:1–92, 2020.

[11] Uthsav Chitra and Benjamin Raphael. Random walks on hypergraphs with edge-dependent vertex weights. In *International Conference on Machine Learning*, pages 1172–1181. PMLR, 2019.

[12] Nagendra Marupaka, Laxmi R Iyer, and Ali A Minai. Connectivity and thought: The influence of semantic network structure in a neurodynamical model of thinking. *Neural Networks*, 32:147–158, 2012.

[13] Judith N Mildner and Diana I Tamir. Spontaneous thought as an unconstrained memory process. *Trends in neurosciences*, 42(11):763–777, 2019.

[14] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. *Advances in neural information processing systems*, 19:1601–1608, 2006.

[15] Timoteo Carletti, Federico Battiston, Giulia Cencetti, and Duccio Fanelli. Random walks on hypergraphs. *Physical Review E*, 101(2):022308, 2020.

[16] Jian-Wei Wang, Li-Li Rong, Qiu-Hong Deng, and Ji-Yong Zhang. Evolving hypernetwork model. *The European Physical Journal B*, 77(4):493–498, 2010.

[17] David Aldous and Jim Fill. Reversible markov chains and random walks on graphs, 2002.

[18] Frank P Kelly. *Reversibility and stochastic networks*. Cambridge University Press, 2011.

[19] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[20] Taher Haveliwala, Sepandar Kamvar, and Glen Jeh. An analytical comparison of approaches to personalizing pagerank. Technical report, Stanford, 2003.

[21] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[22] Cyril Banderier and Robert Dobrow. A generalized cover time for random walks on graphs. *Proceedings of FPSAC'00, Springer, 113-124 (2000)*, 11 2001.

[23] Benjamin F. Maier and Dirk Brockmann. Cover time for random walks on arbitrary complex networks. *Physical Review E*, 96(4), Oct 2017.

[24] Marie Chupeau, Olivier Bénichou, and Raphaël Voituriez. Cover times of random searches. *Nature Physics*, 11(10):844–847, 2015.

[25] Kiyoshi Inoue, Sigeo Aki, and Balakrishnan Narayanaswamy. Generating functions of waiting times and numbers of visits for random walks on graphs. *Methodology and Computing in Applied Probability - METHODOL COMPUT APPL PROBAB*, 15, 06 2013.

[26] Steven Pinker and Ray Jackendoff. The faculty of language: what's special about it? *Cognition*, 95(2):201–236, 2005.

[27] Cynthia Siew, Dirk Wulff, Nicole Beckage, and Yoed Kenett. Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity*, 2019:1–24, 06 2019.

[28] Mark Steyvers and Joshua Tenenbaum. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, 29:41–78, 01 2005.

[29] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[30] Dajie Liu, Norbert Blenn, and Piet Van Mieghem. A social network model exhibiting tunable overlapping community structure. *Procedia Computer Science*, 9:1400–1409, 2012.

[31] F. Hu, J. Guo, F. Li, and H. Zhao. Hypernetwork models based on random hypergraphs. *International Journal of Modern Physics C*, 30:1950052, 2019.

[32] Anna D Broido and Aaron Clauset. Scale-free networks are rare. *Nature communications*, 10(1):1–10, 2019.

A Appendix

A.1 Knowledge Toolkit

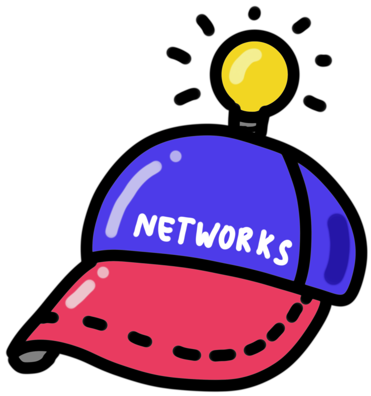


Figure A1: A networks 'thinking cap': you'll require knowledge of C5.4 Networks and its prerequisites (including SB3.1 Applied Probability). SC2 Probability and Statistics for Network Analysis would also be useful.

A.2 Recovery Process Additional Figure

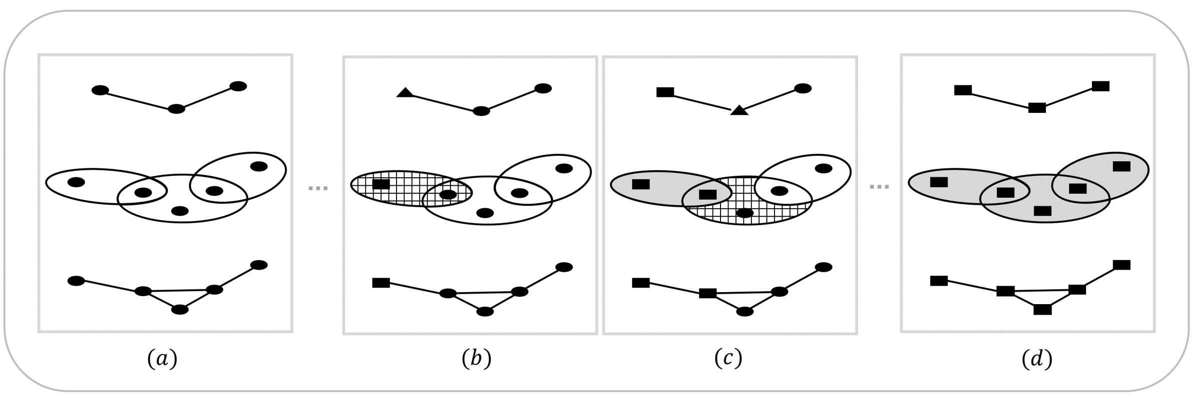


Figure A2: Colour blind friendly version of Figure 7 showing the recovery process. Circles represent maladaptive concepts, triangles represent conflicted concepts, and rectangles represent adaptive concepts. Hyperedges with no fill represent maladaptive thoughts, hashed hyperedges are conflicted thoughts, and grey hyperedges are adaptive thoughts. The aim is to get from the maladaptive TH (a) to the adaptive TH (d).

A.3 Terminology

Model Definition	**Mathematical Definition**
Concept	Node
Thought	Hyperedge
Maladaptive state (concept v_i)	$\lambda_i = 0$
Adaptive state (concept v_i)	$\lambda_i = 1$
Maladaptive state (thought E_α)	$\Lambda_\alpha = 0$
Conflicted state (thought E_α)	$\Lambda_\alpha \in (0, 1)$
Adaptive state (thought E_α)	$\Lambda_\alpha = 1$
Maladaptive TH	$\lambda_i = 0$ for all $v_i \in V$.
Adaptive TH	$\lambda_i = 1$ for all $v_i \in V$.
Conflicted TH	Mixture of some $\lambda_i = 0$ and some $\lambda_i = 1$.
Train of thought	Random walk on an hypergraph
Concept-centric train of thought	Node-centric random walk
Stimulus (concept-centric train of thought)	Node the random walker is on when choosing an hyperedge in the node-centric random walk.
Thought-centric train of thought	Edge-centric train of thought
Currently-active thought	Hyperedge random walker is currently on in edge-centric random walk.
Stimulus (thought-centric train of thought)	Node chosen in transition between hyperedges in edge-centric random walk.
Contracted network	Projection of an hypergraph onto its hyperedges, given by Definition 14.
Refocusing process	Teleportation as in PageRank algorithm.
Concept refocusing process	Node teleportation process
Thought refocusing process	Edge teleportation process
Concept preference vector	Preference vector of nodes in node teleportation process.
Thought preference vector	Preference vector of hyperedges in edge teleportation process.
Reappraisal	Marking process on an hypergraph using a node-centric random walk. A visit to a node may correspond to 'marking' the node by setting its state to 1.
Reappraisal probability	Probability a node is marked when the random walker visits it, denoted by $p_{i,\alpha}$.
Thought-independent reappraisal probability	Probability a node is marked when the random walker visits it is independent of hyperedges ($p_{i,\alpha} = p_i$ for all $v_i \in V$).
Recovery process	Marking process using node-centric random walk with teleportation.

Table A1: Table comparing model terminology to mathematical terminology.

A.4 Sketch-Proof of Theorem 5.1

We sketch a proof of Theorem 5.1, illuminating aspects omitted in [22]. The proof involves four steps, described in Figure A3. Table A2 explains relevant notation.

Notation	**Meaning**
G	Network $G = (V, E)$ with n nodes that the marking process occurs on.
p_l	Probability of marking node $v_l \in V$ when the random walker visits it.
$q_l = 1 - p_l$	Probability the random walker does not mark node v_l on a visit.
$1 - q_l^{r_l}$	Probability the random walker marks node v_l on a random walk in which node v_l is visited r_l times.
(k, v_i)-walk	Random walk of length k starting at node $v_i \in V$.
$\mathbf{X}$	An $n \times n$ diagonal matrix with i-th main diagonal entry equal to the real variable x_i.
$\mathcal{L}_S$	Operator defined in Definition 22 that acts on the variables x_i for $i \in \{1, \ldots, n\}$.
$\mathcal{T}(\tilde{\mathbf{p}})$	Marking time (i.e. the first time that all nodes have been marked) for a random walk with (thought-independent) reappraisal probabilities $\tilde{\mathbf{p}} = (p_1, \ldots, p_n)$.
$\mathbf{T}$	The $n \times n$ node transition matrix.
$\mathbf{I}$	The $n \times n$ identity matrix.
$F_i^{(k)}(x_1, \ldots, x_n)$	Multivariate probability generating function of walks of length k starting at node $v_i \in V$, where x_i corresponds to visits to node v_i.
L	Random variable specifying the length of a random walk on G. That is, $\{L = k\}$ is the event that a random walk is of length k.
$\mathbf{U} = (U_1, \ldots, U_n)$	Vector of random variables, where U_l is the number of times the random walker visits node v_l in a walk.
$\mathbf{r} = (r_1, \ldots, r_n)$	Vector specifying visit frequency for each node, so that the event $\{U_l = r_l\}$ corresponds to r_l visits to node v_l. Note that r_i does not include the visit to v_i at the beginning of the (k, v_i)-walk; it only nodes it visits by walking to them.
$\mathbb{P}(L = k, \mathbf{U} = \mathbf{r})$	Probability node v_l is visited r_l times on a walk of length k, for $l \in \{1, \ldots, n\}$. Note that $\mathbb{P}(L = k, \mathbf{U} = \mathbf{r}) = 0$ unless $\sum_{l=1}^{n} r_l = k$.
B	Event that a random walk results in all nodes being marked.
Y	Random variable with state space $\mathbb{N}$ such that the event $\{Y = k\} = \{B, L = k\}$ is the probability that a random walk of length k has marked all the nodes.
$F_+(z)$	Probability generating function of all random walks on the network starting at node v_i that mark all nodes.
$F_{\mathcal{T}}(z)$	Probability generating function of the marking time. (Usual notation $G(z)$ for probability generating function not used to avoid confusion with the network G).

Table A2: Notation required for the proof of Theorem 5.1

.

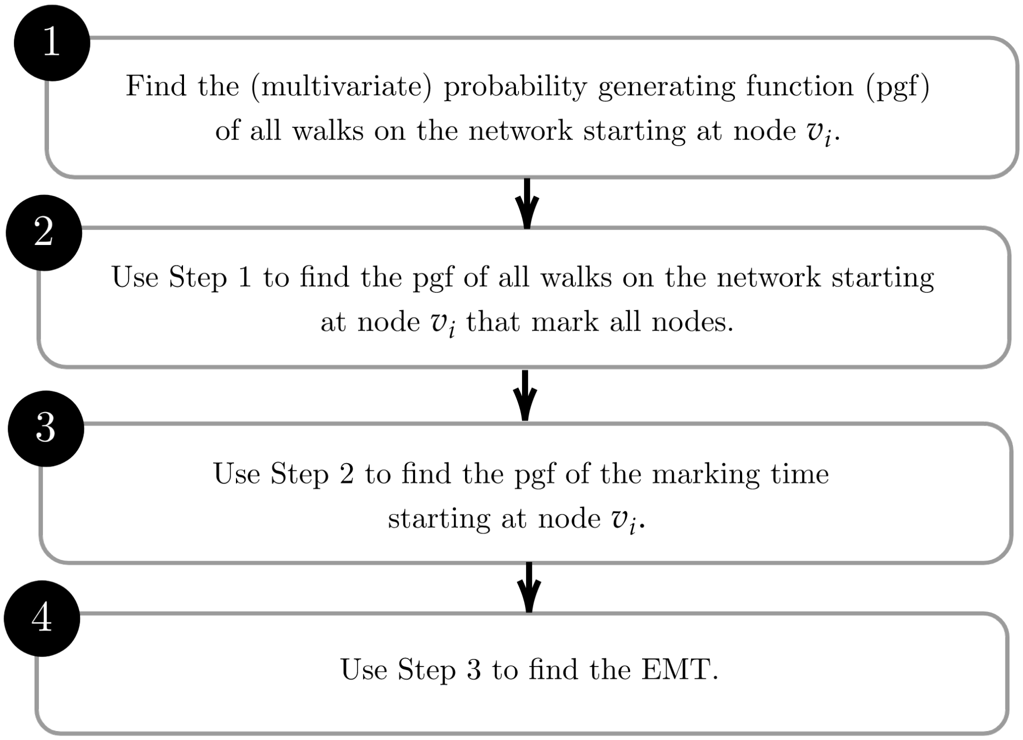


Figure A3: Steps making up the proof of Theorem 5.1.

Step 4 follows from Step 3 since $\mathbb{E}(\mathcal{T}(\tilde{\mathbf{p}})) = F'_{\mathcal{T}}(1)$. We now sketch the other steps.

A.4.1 Step 1

A (k, v_i)-walk can be specified by the frequency of visits to each node. Hence, the (multivariate) pgf of (k, v_i)-walks is

$$F_i^{(k)}(x_1, \ldots, x_n) = \sum_{\mathbf{r}} \mathbb{P}(L = k, \mathbf{U} = \mathbf{r}) x_i x_1^{r_1} \cdots x_n^{r_n} = \sum_{\mathbf{r}} \mathbb{P}(L = k, \mathbf{U} = \mathbf{r}) x_i \prod_{l=1}^{n} x_l^{r_l}, \quad (31)$$

where the sum is over all possible walks. This can be written in terms of the transition matrix: $[(\mathbf{TX})^k]_{ij}$ is a sum of monomials $x_1^{r_1} \cdots x_n^{r_n}$, each with coefficient equalling the probability of walking from v_i to v_j in k steps, such that node v_l is visited r_l times, excluding the start node. Therefore,

$$F_i^{(k)}(x_1, \ldots, x_n) = \sum_{\mathbf{r}} \mathbb{P}(L = k, \mathbf{U} = \mathbf{r}) x_i \prod_{l=1}^{n} x_l^{r_l} = \sum_{j=1}^{n} x_i [(\mathbf{TX})^k]_{ij}. \quad (32)$$

A.4.2 Step 2

Using the total law of probability,

$$\begin{aligned} \mathbb{P}(B, L = k) &= \sum_{\mathbf{r}} \mathbb{P}(B, L = k, \mathbf{U} = \mathbf{r}) && (33) \\ &= \sum_{\mathbf{r}} \mathbb{P}(B | L = k, \mathbf{U} = \mathbf{r}) \mathbb{P}(L = k, \mathbf{U} = \mathbf{r}). && (34) \end{aligned}$$

By independence of node marking, the probability of marking all nodes in a (k, v_i)-walk such that each node v_l is visited r_l times is

$$(1 - q_i^{r_i+1}) \prod_{l \neq i} (1 - q_l^{r_l}). \tag{35}$$

So,

$$\mathbb{P}(B | L = k, \mathbf{U} = \mathbf{r}) = (1 - q_i^{r_i+1}) \prod_{l \neq i} (1 - q_l^{r_l}). \tag{36}$$

Note that

$$\begin{aligned} \prod_{l=1}^{n} (1 - q_l^{r_l}) &= \sum_{S \subseteq \{1,\ldots,n\}} (-1)^{|S|} \prod_{l \in S} q_l^{r_l} && (37) \\ &= \sum_{S \subseteq \{1,\ldots,n\}} (-1)^{|S|} \mathcal{L}_S \bigg(\prod_{l=1}^{n} x_l^{r_l} \bigg), && (38) \end{aligned}$$

which allows us to re-write (36) as a sum:

$$\mathbb{P}(B | L = k, \mathbf{U} = \mathbf{r}) = \sum_{S \subseteq \{1,\ldots,n\}} (-1)^{|S|} \mathcal{L}_S \bigg(x_i \prod_{l=1}^{n} x_l^{r_l} \bigg). \tag{39}$$

Hence,

$$\begin{aligned} \mathbb{P}(B, L = k) &= \sum_{\mathbf{r}} \mathbb{P}(L = k, \mathbf{U} = \mathbf{r}) \sum_{S \subseteq \{1,\ldots,n\}} (-1)^{|S|} \mathcal{L}_S \bigg(x_i \prod_{l=1}^{n} x_l^{r_l} \bigg) && (40) \\ &= \sum_{S \subseteq \{1,\ldots,n\}} (-1)^{|S|} \mathcal{L}_S \bigg(\sum_{\mathbf{r}} \mathbb{P}(L = k, \mathbf{U} = \mathbf{r}) x_i \prod_{l=1}^{n} x_l^{r_l} \bigg) && (41) \\ &= \sum_{S \subseteq \{1,\ldots,n\}} (-1)^{|S|} \mathcal{L}_S \bigg(\sum_{j=1}^{n} x_i [(\mathbf{TX})^k]_{ij} \bigg), && (42) \end{aligned}$$

where (42) follows from (41) by Step 1.

Finally,

$$\begin{aligned} F_+(z) &= \sum_{k=0}^{\infty} z^k \mathbb{P}(B, L = k) && (43) \\ &= \sum_{k=0}^{\infty} z^k \sum_{S \subseteq \{1,\ldots,n\}} (-1)^{|S|} \mathcal{L}_S \bigg(\sum_{j=1}^{n} x_i [(\mathbf{TX})^k]_{ij} \bigg) && (44) \\ &= \sum_{S \subseteq \{1,\ldots,n\}} (-1)^{|S|} \mathcal{L}_S \bigg(\sum_{j=1}^{n} x_i \sum_{k=0}^{\infty} z^k [(\mathbf{TX})^k]_{ij} \bigg). && (45) \end{aligned}$$

Provided $(\mathbf{I} - z\mathbf{TX})^{-1}$ exists, this gives

$$F_+(z) = \sum_{S\subseteq\{1,\ldots,n\}} (-1)^{|S|} \mathcal{L}_S\left(\sum_{j=1}^{n} x_i \big[(\mathbf{I} - z\mathbf{TX})^{-1}\big]_{ij} \right), \tag{46}$$

which bears close resemblance to the formula being proven.

A.4.3 Step 3

Since the marking time is the *first* time all nodes have been marked,

$$\mathbb{P}(\mathcal{T}(\tilde{\mathbf{p}}) = k) = \mathbb{P}(Y = k) - \mathbb{P}(Y = k-1). \tag{47}$$

Then the pgf of $\mathcal{T}(\tilde{\mathbf{p}})$ is[4]

$$\begin{aligned} F_{\mathcal{T}}(z) &= \sum_{k=0}^{\infty} z^k \mathbb{P}(\mathcal{T}(\tilde{\mathbf{p}}) = k) && (48)\\ &= \sum_{k=0}^{\infty} z^k \big[\mathbb{P}(Y = k) - \mathbb{P}(Y = k-1)\big] && (49)\\ &= (1-z) \sum_{k=0}^{\infty} z^k \mathbb{P}(Y = k) && (50)\\ &= (1-z) F_+(z), && (51) \end{aligned}$$

where $F_+(z) = \sum_{k=0}^{\infty} z^k \mathbb{P}(Y = k)$ because $\{Y = k\} = \{B, L = k\}$.

The remainder of the proof is given in [22].

[4]Using the convention $\mathbb{P}(Y = -1) \equiv 0$.

A.5 Mechanisms for Generating Hypergraphs

	Authors	Growth Type	Initialisation	Growth Step	Preferential Choice
1	Wang et al. [16]	Hyperedge	m_0 nodes in 1 hyperedge.	m new nodes and 1 existing node (m fixed) form new hyperedge.	Existing node chosen with probability proportional to degree.
2	Hu et al. [31]	Hyperedge	m_0 nodes in 1 hyperedge.	1 new node and m existing nodes form new hyperedge, with m chosen according to some probability distribution (e.g. Uniform or Poisson).	Existing nodes chosen with probabilities proportional to degrees.
3	Hu et al. [31]	Hyperedge	m_0 nodes in 1 hyperedge.	m new nodes and r existing nodes form new hyperedge. m and r chosen according to some (possibly different) probability distributions.	Existing nodes chosen with probabilities proportional to degrees.
4	Liu et al. [30]	Node	L_0 nodes, N_0 hyperedges.	1 new node v_t added to $m_t - 1$ existing hyperedges, where distribution $M = [m_1, \ldots, m_L]$ pre-specified. L nodes added in total.	New node joins existing hyperedges chosen with probabilities proportional to hyperedge size.
5	Our mechanism	Hyperedge	$\|E_1\|$ nodes in one hyperedge, E_1.	$\|E_{t+1}\| - r$ new nodes, and r existing nodes in hyperedge E_u form new hyperedge, E_{t+1}. E_u chosen with probability proportional to its degree $\delta(u)$, and r uniform on $\{1, 2, \ldots, \min\{\|E_{t+1}\| - 1, \|E_u\| - 1\}\}$. Hyperedge size distribution pre-specified as $D = [\|E_1\|, \ldots, \|E_m\|]$, and $t \in \{1, \ldots, m-1\}$.	Existing hyperedge chosen with probability proportional to degree.

Table A3: Table describing preferential growth mechanisms for hypergraphs. All existing processes that add hyperedges preferentially attach to nodes of high degree. Our mechanism instead preferentially attaches to hyperedges of high degree. Mechanism 4 differs from the others by adding nodes instead of hyperedges at each growth step.

Algorithm 1: Hyperedge Size Distribution

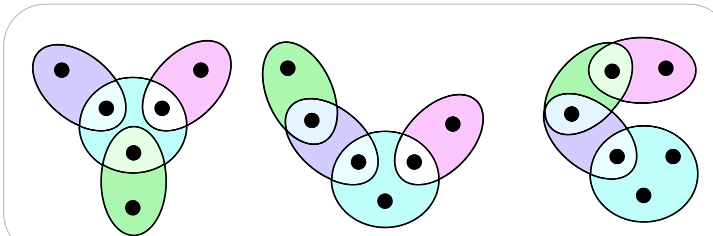
Inputs: m = number of hyperedges, κ = power-law exponent.
Output: Hyperedge size distribution $D = [|E_1|, \ldots, |E_m|]$.
Initialise: $r = 2$, $B = [m]$, $L = B$, $C = 2^{\kappa}$, $D = []$.

```
/* Increase size of largest hyperedge (r) until no longer a valid
   distribution                                                    */
while L[-1] ≠ 0 and sum(L) = m do
    r = r + 1;
    C = (∑_{h=2}^{r} h^{-κ})^{-1} ;              // Normalising constant
    B = L ;                                       // Set to previous list L
    /* Make new list of hyperedge sizes                                */
    L = [] ;                                      // Empty list
    for h = 2 to r do
        x = round(C m h^{-κ}) ;                   // Number of hyperedges of size h
        L.append(x) ;                             // Add to list
    end
end
/* Use B to make hyperedge size distribution                          */
for i = 0 to len(B) - 1 do
    for j = 1 to B[i] do
        D.append(i + 2);
    end
    D.sort(reverse = True) ;                      // Sort from largest to smallest
end
```

(a)

Step	r	B	C	L
0	2	$[4]$	8	$[4]$
1	3	$[4]$	$216/35$	$[3, 1]$
2	4	$[3, 1]$	$1728/307$	$[3, 1, 0]$

(b)



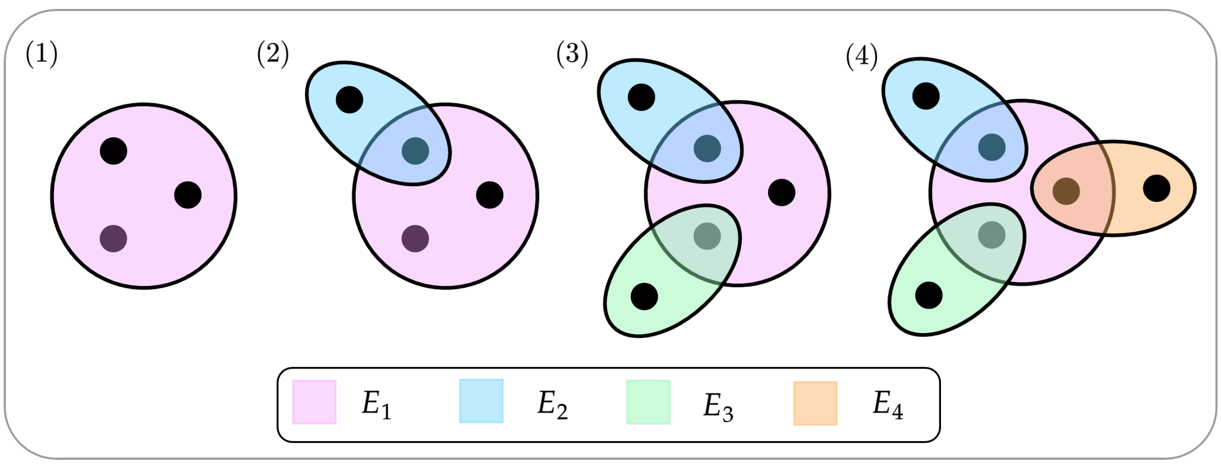
(c)

Figure A4: Algorithm for generating power-law hyperedge degree distribution $D = [|E_1|, \ldots, |E_m|]$. (a) Table showing algorithm. (Python code in A.7.) (b) Example of Algorithm 1 for $m = 4$, $\kappa = 3$ (while-loop). Output using for-loop is $D = [3, 2, 2, 2]$. (c) Hypergraphs with $D = [3, 2, 2, 2]$. (Note that 'degree distribution' is synonymous with 'size distribution' as we are assigning all nodes weight 1, and all hyperedges $E_\alpha \in E$ weight $|E_\alpha|$.)

Algorithm 2: Preferential Hyperedge Growth Mechanism

Inputs: $D = [|E_1|, \dots, |E_m|]$ (hyperedge size distribution), $m = \mathrm{len}(D)$.
Output: Hypergraph H.
Initialise: Hypergraph H contains $|E_1|$ nodes in one hyperedge, E_1.
for $t = 1$ *to* $m - 1$ **do**
  Hypergraph Growth:
  Choose one hyperedge E_u in H with probability proportional to degree, $\delta(u)$;
  Choose l uniformly from $\{1, \dots, \min\{|E_u| - 1, |E_{t+1}| - 1\}\}$;
  Choose l nodes uniformly in E_u;
  Hyperedge E_{t+1} contains these l nodes and $|E_{t+1}| - l$ new nodes;
  Add E_{t+1} to H.
end

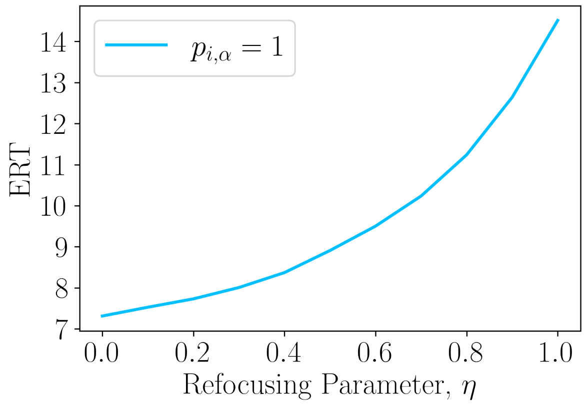
(a)



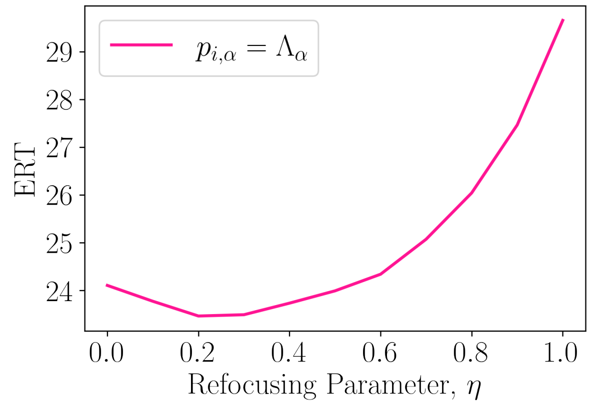
(b)

Figure A5: Algorithm for generating hypergraph from hyperedge size distribution $D = [|E_1|, \dots, |E_m|]$. (a) Table showing algorithm from Figure 8. (Python code in A.7.) (b) Example of Algorithm 2 for $D = [3, 2, 2, 2]$. (1) Initially H contains $|E_1| = 3$ nodes in one hyperedge, E_1. (2) Choose E_1 with probability 1. Choose $l = 1$ (since $\min\{|E_1| - 1, |E_{t+1}| - 1\} = \min\{2, 1\} = 1$). Then E_2 contains 1 node in E_1 and 1 new node. (3) Choose E_1 with probability 3/5 and E_2 with probability 2/5. Say E_1 is chosen. Then $l = 1$ (since, again, $\min\{2, 1\} = 1$). Then E_3 contains 1 node (chosen uniformly) from E_1 and 1 new node. (4) Choose E_1 with probability 3/7, E_2 with probability 2/7 and E_3 with probability 2/7. Say E_1 is chosen. Then E_4 contains 1 node in E_1 and 1 new node.

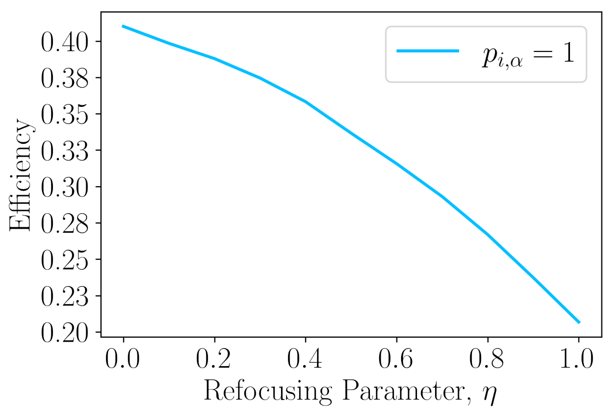
A.6 Additional Figures



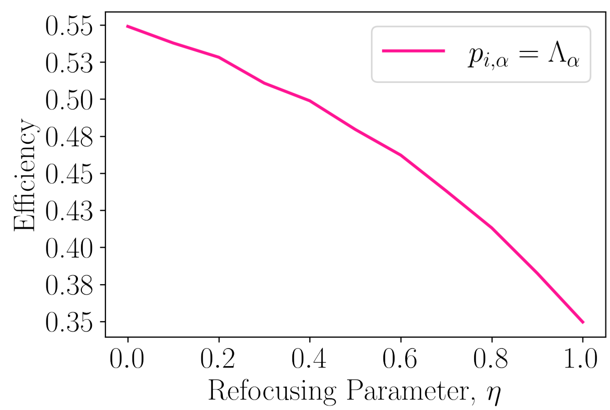
(a)



(b)



(c)



(d)

Figure A6: Recovery process for an hypergraph with $m = 2$ and $\kappa = 3$, averaged over 500 realisations per starting node. The behaviour matches that of an hypergraph with $m = 6$ and $\kappa = 3$. (a) ERT increases with refocusing parameter η for thought-independent reappraisal probabilities. (b) ERT for thought-dependent reappraisal probabilities is very high for high η and low for small η, but not monotonic. (c) Efficiency is a decreasing function of η for thought-independent reappraisal probabilities. (d) Efficiency is a decreasing function of η for thought-dependent reappraisal probabilities.

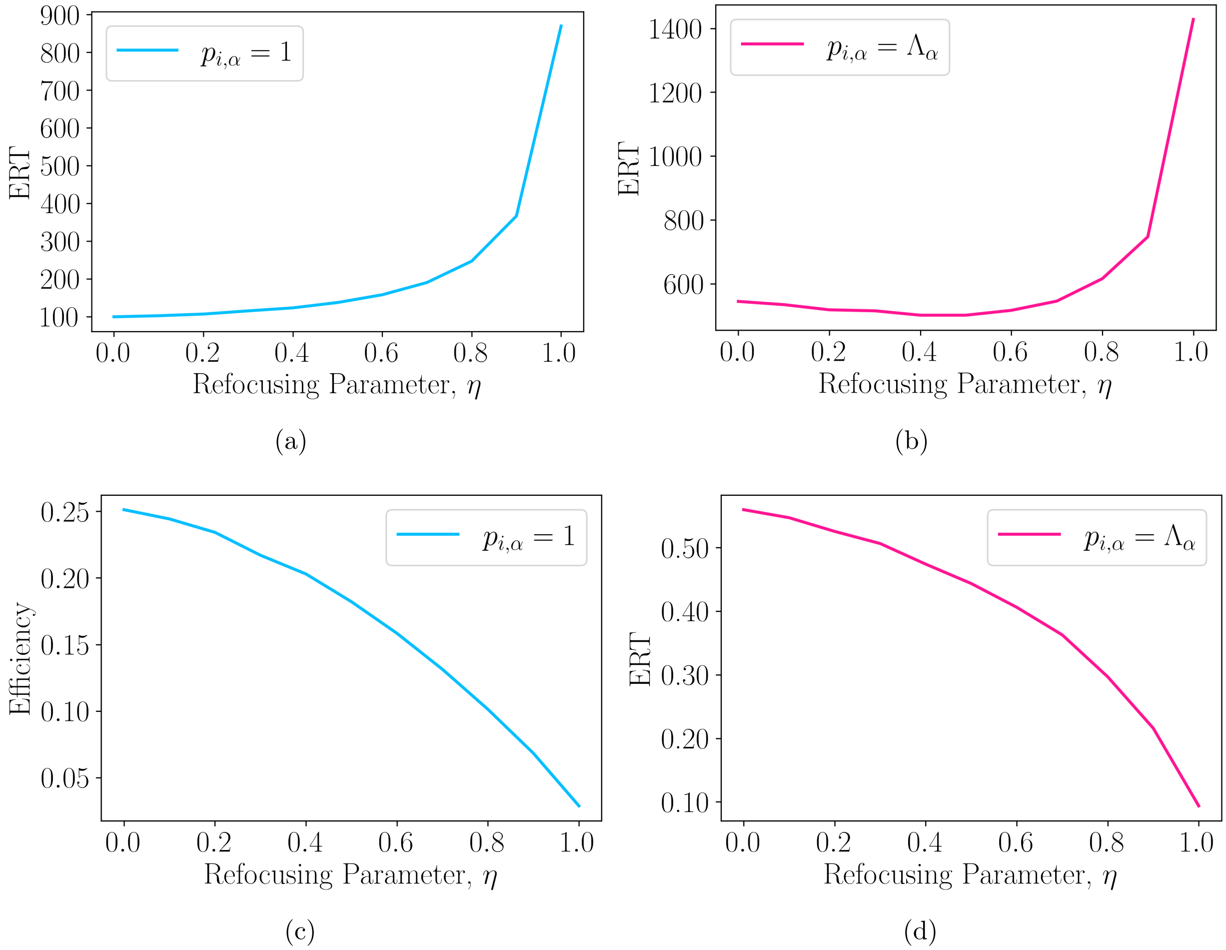


Figure A7: Recovery process for an hypergraph with $m = 12$ and $\kappa = 3$, averaged over 100 realisations per starting node. The behaviour matches that of hypergraphs with $m = 6$ and $m = 2$. (a) ERT increases with refocusing parameter η for thought-independent reappraisal probabilities. (b) ERT for thought-dependent reappraisal probabilities is low for small η and dramatically increases for large η. (c) Efficiency is a decreasing function of η for thought-independent reappraisal probabilities. (d) Efficiency is a decreasing function of η for thought-dependent reappraisal probabilities.