Universitat
de les Illes Balears

MASTER'S THESIS

CREATIVITY AS A RANDOM WALK SEARCH ON A SEMANTIC NETWORK

Nicoleta Kyosovska

Master's Degree in Physics of Complex Systems

Centre for Postgraduate Studies

Academic Year 2021-22

CREATIVITY AS A RANDOM WALK SEARCH ON A SEMANTIC NETWORK

Nicoleta Kyosovska

Master's Thesis

Centre for Postgraduate Studies

University of the Balearic Islands

Academic Year 2021-22

Key words:

Remote Associates Test, word association networks, creative problem-solving

Thesis Supervisors: Raúl Vicente, Víctor M. Eguíluz

Tutor: Claudio Mirasso

Summary

This thesis studies creativity in the context of network science. It investigates whether a random walk is a good model for the problem-solving process in humans during the Remote Associates Test, a multiply-constrained semantic creativity task. The work introduces two search models, which consist of a number of random walkers starting from each task cue and traversing a semantic network for a given number of time steps. We keep track of coincidences, defined as the nodes where the walkers' trajectories intersect asynchronously, and we look for the task solutions amongst the earliest or most frequent coincident words. Our simulations show that the model with multiple walkers starting from each cue matches average human performance and correlates with individual example accuracy for a given set of its parameters. The results are validated by running the simulations on randomised versions of the original network. We have also shown that the model is consistent with findings about the structural differences between association networks of high- and low- creative individuals.

These results agree with analysis of human data which indicates that local undirected search processes might lead to solving convergent creative tasks. We argue that the model is grounded in established theories of creativity in cognitive science and is plausible from a neuroscience perspective. The findings from this study can be the basis for further investigation into the dynamics behind the observed variance in creative ability in humans.

Contents

1 Introduction

Creativity is a fascinating quality of our minds. It is an essential ingredient for doing innovation, creating beauty or solving everyday problems. "Little-C" creativity is particularly indispensable in everyday cognitive functioning, yet research still hasn't come up with a full account of the psychological and neurobiological mechanisms underpinning it. Since it is multi-faceted and is, by definition, a very high-level process, it is a challenge to create a model which encompasses all stages of the creative process and unifies all its low-level mechanisms.

In this work we focus on studying creative thinking under the associative theory of creativity. We do so by using a simple dynamical model for undirected search. We evaluate it on a widely-used test for creativity. We want to see whether a simple local search process on a network can lead to finding the solution of a multiply-constrained problem. By using an open-ended process on a closed-solution task we attempt to achieve a unification of the distinct aspects of creative ability.

In the Introduction section, we start with an outline of the most established theories of creativity and the creative process. We proceed with the introduction of multiply-constrained problems. We then describe one of the most widely used tests for creativity: the Remote Associates Test, which is used for the evaluation of the model in the present report. Next, relevant work done in network science in the context of creativity is presented, in order to motivate the chosen approach in this work, together with a brief review of other computational approaches. We end the Introduction by specifying the research question and hypotheses, and by outlining the structure of the report.

1.1 Theories of Creativity

1.1.1 What is creative cognition?

Creativity in people is ability and quality: the ability to create and the quality of being creative. Some authors make a distinction between the so called "big-C" creativity, which belongs to the truly brilliant minds in the human history, and "little-C" creativity, the cognitive capacities that allow "ordinary" brains to solve everyday problems [Mekern et al., 2019]. According to Margaret Boden [Boden, 2004], a creative idea is one which is new, surprising and valuable. It can be new and valuable to one individual or to an entire society, and it can have different degrees of surprise. She argues that "big-C" creativity involves "H-novelty" - novelty with respect to the whole of human history, and as such, it cannot have a psychological or systematic explanation. Thus, in science, we focus on "big-C" creativity and on "P-novelty" - novelty for the individual or for society.

Based on how it comes about, creativity can be combinatorial and non-combinatorial [Boden, 2004]. Combinatorial creativity is making novel combinations of familiar ideas, for example what we see in poetic imagery or when making analogies. Non-combinatorial creativity can be exploratory - the generation of novel ideas by the exploration of structured conceptual spaces, or transformational - the same as exploratory but including a transformation of some dimension of the space so that new structures can be generated.

Another very common line of division of creative processes is the distinction between convergent and divergent thinking [Guilford, 1967]. When involved in divergent thinking, a person is thinking in an open-ended way, while convergent thinking implies having a specific goal in mind. While in research, tests and models often target only one of the two types, in reality creativity is usually a mix of both convergent and divergent thinking, as well as other cognitive processes [Mekern et al., 2019]. Testing for one type at a time is easier to operationalise, but doing so prevents us from obtaining a holistic explanation of creativity.

1.1.2 What does the creative process imply?

Helie et al distinguish between four stages in the creative problem solving: preparation, incubation, illumination (insight), and verification [Hélie and Sun, 2010]. The first stage consists in using logical reasoning to do search in various directions. If the problem is simple, a solution is found at this stage and there isn't a need to go through the rest of them. If the problem is more difficult, this stage usually does not lead to a good solution. The incubation period then begins - the person stops actively thinking about the task. This period can last for a very short time, like a few minutes, or a very long time - years. It has been demonstrated in experiments that incubation increases the probability of finding the solution. The next station, illumination, insight, or also called the "aha" moment, is the sudden realisation that a solution to the problem has been found. The last stage, verification, refers to the process of checking whether the insight is indeed the correct solution. Verification involves logical reasoning like the preparation stage. If it leads to the conclusion that the found solution is incorrect, the person repeats the process.

It is difficult to attribute these stages to either convergent or divergent thinking. Since it is a problem solving process, the assumption is that it involves convergent thinking: the problem solver is thinking about the problem, having in mind that there is an existing solution. We could label the preparation and the verification stages as convergent since they use logical thinking, but divergent thinking could also involve reasoning. Instead, we make a different distinction: of explicit and implicit cognitive processing. Explicit knowledge is easily accessible and verbalised and is thought to be often symbolic and clear, while implicit knowledge is less accessible and relatively hard to verbalise, and is often subsymbolic and vague [Hélie and Sun, 2010]. According to the Explicit-Implicit Interaction Theory [Hélie and Sun, 2010], explicit processes involve some form of rule-based reasoning, while implicit processes are associative and often satisfy soft constraints.

Another widely-known theory for the creative process is the associative theory of creativity by Mednick. According to him, creative thinking is the process of associating elements into new combinations, which are either useful in a certain way or meet specified requirements [Mednick, 1962]. Mednick argues that "the more mutually remote the elements of the new combination, the more creative the process or solution." He attributes different levels of creativity amongst people to different associative hierarchies. People with a steeper associative hierarchy can't reach more remote associations because stereotypical solutions come to mind with much higher probability than unstereotypical ones. On the other hand, people with a flatter associative hierarchy can think of solutions which are nearby or much further away with similar probabilities. Therefore, they reach more creative solutions. Mednick defines three ways of achieving a creative solution: serendipity (combining elements by accident), similarity (grouping elements by common features) and mediation (identifying a link that different elements share). He operationalises these ideas by introducing a test for creativity where a participant is asked to associate distinct elements by finding a common connection between them and thus to provide new meaningful combinations. This is called the Remote Associates Test and it will be presented in more detail in the next section.

1.2 Operationalising Creativity

1.2.1 Multiply-constrained creative problems

Tasks calling for a creative solution often involve multiple constraints to be satisfied. Problems with multiple constraints exhibit two important characteristics: each constraint defines a qualitatively different and mutually uninformative goal, and there isn't an explicit way of weighing the goals [Smith et al., 2013]. People tend to find solutions to such problems easily and this can misleadingly lead to the conclusion that the problems aren't difficult. Nevertheless, the computational load can be very high. The combinatorics of these tasks are usually huge and an exhaustive search of the solution space is inefficient. How do people direct their search? There are multiple theories on how this is done. The following high-level description of the process is often proposed: people first find a candidate solution, then they test the solution against all the criteria. If it is considered as acceptable, they stop the search process, otherwise they repeat it.

This is akin to the process formalised by Helie et al: the last stage here is the verification stage. Since finding a candidate solution can involve both implicit and explicit thinking and is not confined in a specific time frame, the first stage here can encompass all the previous stages from Helie et al's work: the preparation, the incubation and the insight stages.

1.2.2 The Remote Associates Test

Mednick defines a creative idea as an idea generated by making a new combination of elements via association. The result needs to be useful and/or to meet certain criteria, in order to be regarded as creative. Moreover, the more remote the association between elements, the more creative it is. In order to test for creative ability, he introduces the Remote Associates Test: participants are given three words as cues (for example blue, rat and cottage) and they need to find one word which is linked to all three (in this case it is cheese). The associative relationship between the cues and the solution in the RAT can vary: it can be a compound word such that each cue and the solution form a new word (e.g., firefly); it can be semantically related (e.g., water and ice); or it can form an expression (e.g., mind game).

RAT is the second most used test amongst 45 published imaging studies of creative cognition [Arden et al., 2010]. It has various advantages as a test for creativity - it is simple, and it is therefore easy to control, all constraints are of the same type (word-word associations), it lasts for a short time, so many examples can be given to many participants, and it has a single unique solution, making it easy to score.

It is a test for convergent thinking. It covers both the process of exploration, or finding a candidate solution, and verification, or testing the solution against all criteria. It can involve both explicit and implicit thinking, making it an insight problem. One of the central features that insight problems possess is the "aha" feeling experienced upon finding the right word [Bowden and Jung-Beeman, 2003].

It exhibits key features of multiply-constrained problems: each cue indicates a different aspect of the target word, and a way to prioritise connections to certain cues does not exist. [Smith et al., 2013]. This makes the test suitable not only for evaluation of creativity, but of evaluating a more general cognitive process, such as searching for candidate solutions in order to solve a multiply-constrained problem.

Although it is a very widely accepted creative problem solving test, RAT has received various criticisms.

An objection to the test is the proposition that it might be better suited for verbal fluency [Ansburg, 2000]. In this respect, versions in other languages have been developed. It has been shown that a less analytical approach to finding the right solution helped only people who were doing the test in their native language [Aiello et al., 2012]. Nevertheless, when building a model to solve the test, as long as it isn't a developmental model, its semantic knowledge is obtained a priori. This makes linguistic fluency a feature of the semantic knowledge, rather than the model. While it is important to acknowledge these criticisms, the operational environment for studying creativity that RAT offers is very good, and it is suitable for the research question of the present work.

1.3 Computational Models of Creativity

In this report we choose to represent semantic knowledge with a network, since this is most consistent with early theoretic work on semantic memory [Kajić et al., 2017]. Random walks are the most simple dynamical process on networks, and they have been used to model a variety of cognitive processes. In the context of creativity, it has been shown in a free recall task that a random, undirected search process can be a good model for semantic search behaviour in humans [Abbott et al., 2015]. Analysis of the search behaviour of humans in a multiply-constrained task suggests that local undirected search might also be central to solving a convergent thinking task [Smith et al., 2013]. Consequently, it seems reasonable to approach the question of associative creative thinking from the point of view of network science. We present a few works in this context.

1.3.1 Creativity in network science

Since a network is an abstract structure, the set of mathematical and computational tools for analysing it can be applied to any field. This is powerful because regardless of the specifics of the components of the system, one can find developed and understood analysis tools.

Classic cognitive theory in language and memory usually talks about networks of knowledge, however, network science hasn't been used to study cognitive phenomena as much as it has been used in neuroscience [Kenett and Faust, 2019]. Structuring semantic memory as a network allows us to do so. Network theory can be directly applied to quantitavely examine cognitive theory and to help improve our understanding of the mechanisms behind cognitive processes such as creative problem-solving.

The following discovery is made by Abbott et al in [Abbott et al., 2015]: they show how a one-stage search process (a random walk), which operates on a semantic network, is similar to a two-stage search process (optimal foraging), which operates in a semantic space (another representation). The two distinct processes in question are "clustering", during which the agent stays within a cluster of semantically related words, and "switching", when the agent occasionally changes the cluster. The authors show that the reason for the good fit of the random walk to the human data is that the semantic network captures well the clustering of its elements. This result is interesting, when put in the context of creativity and the tension in unifying convergent and divergent mechanisms, since it shows that an undirected search process can be a minimal model for directed, strategic behaviour.

The first time that an experiment with interim answers to the RAT was carried out was done a few years ago by Smith et al [Smith et al., 2013]. The sequences of answers given by participants allowed for the analysis of the search process. Two strategies were found: people evoke solutions mainly linked to one of the three cues at a time, and they adopt a local strategy - new guesses are affected by previous guesses. The results expand on the work by Abbott et al described above, showing evidence for a local search process in a multiply-constrained task. They differ from it in the fact that while local search is present, clustering behaviour is not exhibited. Moreover, the authors argue that it is not yet clear why local search strategy is adopted by people.

Regarding the first result in [Smith et al., 2013], that answers tend to be semantically similar only to one of the word constraints at a time, it is discussed that this might be an effect of the process of overtly reporting guesses. In addition, there might be additional implicit processing happening in the brain. Otherwise, if only taking the number of reported words in account, it seems that with a significant portion of the solution space left unexplored, it would be unlikely to find the correct solution. Thus, when modelling the problem solving process in a multiply-constraint task with a random walk, it seems unfavorable to constrain the search to one cue at a time.

Another interesting take on creativity research in network science is the work by Kenett et al [Kenett et al., 2018]. Here the authors compare the semantic networks of high and low creative individuals (as evaluated through a battery of creativity tasks), created from free association data. They use network percolation analysis (removal of links in a network whose strength is below an increasing threshold) to show that the network of high creative individuals is more robust. Following the assumption that a robust network means higher flexibility, they manage to quantify flexibility of thought, which is theorized to be an important requirement for finding creative solutions. The theory is further supported by the finding that what makes the networks of high creative individuals more robust is the higher strength of the links between different components. This might be important for the ability to find more remote associations, which also supports Mednick's definition for a creative association. Further analysis of the difference between the two ends of creative ability is the simulation of random walks on the semantic networks [Kenett and Faust, 2019]. This showed that a walk on the network of high creative people visits more unique nodes and moves further through the network for a set number of steps. Again, this is consistent with Mednick's theory.

1.3.2 Other computational models

We limit the review of other computational models to such which were evaluated on the RAT. Although in a different architectural setting, they shed a light on important aspects of the problem solving process.

Apart from random walks on semantic networks, more complicated cognitive and neurobiological models have been recently developed. One is a cognitive architecture which comprises a knowledge base of word pairs. They are modelled in the second conceptual level of a three-level architecture, whose first level is a subsymbolic level of feature spaces, represented in a distributed way [Mekern et al., 2019]. Modelling an associative search of the knowledge base involves activating all the word pairs that involve the three cue words and letting the architecture converge on an answer found in three different word pairs.

Interestingly, it has been shown in a similar architecture that performance in the RAT was higher when increasing the activation of a single cue. Free recall thus led to better results than a cued-retrieval method. This means that it is possible for a divergent activity to converge on a correct solution via outputting the most activated element(s) [Mekern et al., 2019].

A similar result was obtained in a spiking neuron model for solving the RAT [Kajić et al., 2017]. The model involves a selection network which randomly activates one of the three RAT cues at a time, regularly switching between them. The representation of words, over which activation spreads, is distributed. It starts spreading from all the neurons which take part of the representation of the cue word. A winner-takes-all mechanism, subject to an auto-inhibitory signal, takes the activation patterns as input and outputs the most activated word. The performance of the model had a good match with the human data, which confirms the idea that a link between divergent and convergent processes could be a winner-takes-all mechanism.

1.4 Problem Specification

In the present report we choose to work with the RAT since it is a widely accepted test for creativity in humans. We represent semantic knowledge with a network, where the nodes in the network are words and the links between them exist if two words can be associated. This representation implies the presence of clusters of words, which is conducive for the modelling of optimal search behaviour [Abbott et al., 2015]. We use a random walk since this is the most simple dynamical process on networks, and yet, it has been shown to be a good model for various cognitive processes, including creative associative thinking [Abbott et al., 2015]. There exists a tension between incorporating divergent and convergent processing in a model for creativity. For a convergent task such as the RAT, it might seem that focused, strategic search is required. However, there are indications following analysis of human behaviour, that local search is central for finding the correct solutions [Smith et al., 2013]. Results from cognitive and neurobiological architectures hone in on this idea, showing evidence that it is possible for a divergent processing model to converge on a correct solution [Mekern et al., 2019]. We will therefore check whether a random walk on a semantic network (an open-ended dynamical process) will match human performance on the multiply-constrained Remote Associates Test (a convergent task).

Specifically, we try to answer the question: is a random walk on a semantic network, derived from human associations, a good model for creative problem solving? We evaluate the model on the Remote Associates Test to check whether the model's performance is comparable to human performance. We look at the percentage of solved examples and the accuracy per example for a range of the model's parameters.

1.5 Report Structure

The report follows this outline: we first formally specify the random walk process. We proceed to describe the semantic networks used as a knowledge base. We then categorise their structure. What follows are the algorithms behind the simulations, and the evaluation procedures used in them. Next, we present the results. We conclude with a discussion, limitations and future work.

2 Methods

2.1 Network Theory

A network is a group of points linked by lines. A point is referred to as a node and a link is referred to as an edge. Many systems in the physical, biological and social sciences can be understood as networks. They can be broadly classified as technological, biological, social, information, etc. [Newman, 2018].

In the work here we use semantic networks, which are a type of information networks. The most famous example of an information network is the World Wide Web where we don't have physical objects as nodes and edges, but web pages and hyperlinks. The World Wide Web can give an insight into relationships between content and topics, as well as the structure of human knowledge.

Semantic networks can be considered as information networks because they represent knowledge. Semantic networks have been used in psychology, philosophy and linguistics, as well as in automated reasoning, machine translation and to model cognition [Sowa, 1987]. Some types are very informal, while others contain formally defined logic.

Semantic networks of associations between words have recently gathered attention in the study of semantic memory structure, and specifically, its relation to creativity. They are a very simple kind of semantic network: nodes are words which are linked, if the two words are associated.

2.1.1 Characterising networks

A network can be directed or undirected, depending on whether the direction of an edge is important or not. It can be weighted, where the weight signifies the strength of a link, or unweighted, where a link can be either 0 (not present) or 1 (present). A network that does not contain more than one edge between nodes, and that does not contain self-edges, is called a simple network.

Another categorisation for networks is whether they are connected or disconnected. If a network is disconnected, it is divided by two or more subsets, called components, such that there is no path between two nodes in different components. If all nodes belong to the same component, the graph is connected.

The adjacency matrix is a mathematical representation of a network. If every node gets an integer label $1...n$, the matrix is defined to be an $n \times n$ matrix with elements A_{ij} such that $A_{ij} = 1$ if there is an edge between the nodes i and j, and $A_{ij} = 0$ otherwise.

The graph Laplacian is another very widely used representation of a graph. It is relevant for the theory of random walks on networks, diffusion, dynamical systems, etc.

The graph Laplacian for a simple, undirected, unweighted network is a $n \times n$ symmetric matrix L with elements

$$L_{ij} = \begin{cases} k_i & \text{if } i = j, \\ -1 & \text{if } i \neq j \text{ and there is an edge between nodes } i \text{ and } j, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where k_i is the degree of node i [Newman, 2018].

We can also define it as an equation that includes the Kronecker delta δ_{ij}, which is 1 if $i = j$ and 0 otherwise:

$$L_{ij} = k_i \delta_{ij} - A_{ij}, \tag{2}$$

where A_{ij} is an element of the adjacency matrix.

In matrix form, L becomes

$$\mathrm{L} = \mathrm{D} - \mathrm{A}, \tag{3}$$

where D is the diagonal matrix with the node degrees:

$$\mathrm{D} = \begin{pmatrix} k_1 & 0 & 0 & \cdots \\ 0 & k_2 & 0 & \cdots \\ 0 & 0 & k_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{4}$$

The graph Laplacian can also be defined for weighted networks: one simply replaces the adjacency matrix with the weighted adjacency matrix and the degree k_i of a node by the sum $\sum A_{ij}$ of the respective matrix elements.

$\boldsymbol{L}$ has some useful properties:

1. The sum of each row (column) is zero $\sum_j L_{ij} = \sum_j (k_i\delta_{ij} - A_{ij}) = k_i - k_i = 0$
2. All the eigenvalues of L are non-negative
3. At least one eigenvalue is $\lambda = 0$
4. Vector $\mathbf{1} = (1, 1, \ldots, 1)$ is the eigenvector relative to $\lambda = 0$
5. There are as many $\lambda = 0$ as many connected components in the network

2.1.2 Structural parameters

There are a number of essential measures that characterise the structure of a network (graph).

Number of nodes and edges The simplest measures for a network are the number of its nodes and its edges. They are usually denoted as n and m and they give an idea of the size of the network.

Degree The degree of a node in an undirected graph is the number of edges connected to it.

The degree of node i is usually denoted as k_i. It can be written in terms of the adjacency matrix as

$$k_i = \sum_{j=1}^{n} A_{ij}. \tag{5}$$

If we count all the degrees in the network, we count all the edges twice: if a network consists of two nodes i and j and they are connected, the number of edges is $m = 1$ while the sum of their degrees would be 2. Thus we have

$$2m = \sum_{i=1}^{n} k_i = \sum_{ij} A_{ij}. \tag{6}$$

Density The density ρ of the network is the fraction of the total number of possible edges in a network that are present:

$$\rho = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)} = \frac{c}{n-1} \tag{7}$$

where c is the mean degree.

Since the density lies in the range $0 \leq \rho \leq 1$, it can be understood as the probability that a randomly picked pair of nodes is connected by an edge.

A network is considered sparse if, when taking the limit $n \to \infty$, the density approaches 0. One can't do such an approximation with real-world networks, but informally, a network is considered sparse if ρ is very small [Newman, 2018].

Global Clustering The two most widely used clustering coefficients are the Watts-Strogatz clustering coefficient, that defines global clustering as local clustering on average, and the Newman transitivity index that defines global clustering in terms of the fraction of paths of length two that are closed.

The global clustering coefficient is defined as

$$C_N = \frac{\text{number of closed paths of length two}}{\text{number of paths of length two}}$$

If $t = |C_3|$ is the total number of triangles and $|P_2|$ is the number of paths of length two, then

$$C_N = \frac{3|C_3|}{P_2} \tag{8}$$

Local Clustering The second definition for a network's clustering, defined by Watts and Strogatz, is the mean of the local clustering coefficients. Local clustering represents the average probability that a pair of one's connections are also connected between themselves:

$$C_i = \frac{\text{number of pairs of neighbours of i that are connected}}{\text{number of pairs of neighbours of i}}$$

If t_i is the number of triangles attached to node i of degree k_i then we have

$$C_i = \frac{2t_i}{k_i(k_i - 1)} \tag{9}$$

Thus the Watts-Strogatz clustering becomes:

$$C_{WS} = \frac{1}{n} \sum_i C_i \tag{10}$$

Average Path Length A walk in a network is a sequence of nodes where each node is connected to the previous one by an edge. A path is a walk that does not repeat nodes. The length of a path between two nodes is the number of edges that need to be traversed along it.

The average shortest path is the average number of edges it takes to get from one node to another. The shortest path length is the shortest walk between two nodes.

Diameter The diameter is the longest shortest path. A small diameter is a property of "small-world" networks.

2.2 Random Walks

A random walk is a trajectory across the nodes and edges of a network by repeatedly taking random steps. We start at an initial node j at time $t = 0$ and choose randomly one of its edges k_j, proportionally to their weights. For an unweighted graph, an edge would be chosen uniformly. Then we walk across the chosen edge e_{ji} and find ourselves at node i at time $t = 1$. We then repeat the same for t_{max} steps. A random walk can traverse nodes and edges multiple times and can consequently move along the same edge in the opposite direction more than once (if the network is undirected or if there is an edge in both directions in a directed network) [Newman, 2018].

In an unweighted network, the probability $p_i(t)$ that a walker is at node i at time t is the probability that it is at a neighbour node j at time $t - 1$ times the probability to step on one of its edges, which is $1/k_j$:

$$p_i(t) = \sum_j \frac{A_{ij}}{k_j} p_j(t-1) \tag{11}$$

In matrix form this can be written as $\mathrm{p}(t) = \mathrm{AD}^{-1}\mathrm{p}(t-1)$ where p is the vector with elements p_i and D is the diagonal matrix with the degree of the nodes. This expression is for an undirected, unweighted network. For a weighted network, we use the sum of all the weights of node i's edges, instead of its degree in equation 11.

The distribution after a long time over nodes is given by equation 4 with t equal to infinity: $p_i(\infty) = \sum_j A_{ij} p_j(\infty)/k_j$, or in matrix form:

$$\mathrm{p} = \mathrm{AD}^{-1}\mathrm{p} \tag{12}$$

Here, for the weighted version we have the sum of all the weights of node i's edges for each diagonal element i in D.

If we rearrange equation 12, it becomes

$$\left(\mathrm{I} - \mathrm{AD}^{-1}\right)\mathrm{p} = (\mathrm{D} - \mathrm{A})\mathrm{D}^{-1}\mathrm{p} = \mathrm{LD}^{-1}\mathrm{p} = 0 \tag{13}$$

From this it follows that $\mathrm{D}^{-1}\mathrm{p}$ is an eigenvector of the Laplacian with eigenvalue 0 [Newman, 2018].

Using properties 4. and 5. of the Laplacian, we have that $\mathrm{D}^{-1}\mathrm{p} = a\mathbf{1}$, where a is a constant, or equivalently $\mathrm{p} = a\mathrm{D}\mathbf{1}$, so that $p_i = ak_i$. If a is such that the probabilities $\sum p_i = 1$, we obtain

$$p_i = \frac{k_i}{\sum_j k_j} = \frac{k_i}{2m} \tag{14}$$

by using equation 6.

As one would guess, a walker is more likely to traverse a node if its degree is high. In the context of diffusion, where we have multiple walkers on a network, we can understand from this equation the topological effect on the diffusive process: it stabilises on the most connected nodes [Barrat et al., 2008].

2.3 Model Description and Evaluation

We model associative search during the RAT by simulating random walks on a semantic network and tracking coincidences between them. We describe the two algorithms to do so below.

2.3.1 Number of coinciding walkers

For each example, a number of walkers W starts from each cue word. Then the walk evolves according to the dynamics described in subsection 2.2. We simulate walks with W between 10 and 1000 and for time steps t_{max} 5 and 100.

After a simulation has finished, we find asynchronous coincidences between walkers. We define an asynchronous coincidence as a node that was traversed at any time step during the simulation (apart from the initial time step $t = 0$), by at least one walker from each cue word. The time of a coincidence is the average of the times of the walkers involved in it. We then measure two quantities:

Earliest Coincidences These are the top coincident nodes according to earliest time.

Most Frequent Coincidences These are the top coincident nodes according to the total number of walkers, who traversed it, until a time step t.

2.3.2 Independent coincidences between triplets of walkers

We define a second way to simulate asynchronous coincidences between walkers on a network. A number of walkers $W = 3$ is let to walk on the network at a time, and each walker starts from each cue word as before. The walk ends when the number of coincidences is $c = 10$. Here a coincidence is defined as a node traversed by exactly $W = 3$ walkers and its time is the average of the times of the three walkers. We repeat this for i iterations, or number of walks of $W = 3$, in order to find the earliest and most frequent coincidences.

The key difference between the two simulations is that in the first case, we find the top coincident nodes according to one iteration, amongst many walkers. For any node, the number of coincidences is fixed, it is a single one, as per the definition in *2.3.1 Number of coinciding walkers*. What we count is the number of walkers who traversed the node (i.e. who were part of the coincidence). In the second case, we find the top coincident nodes amongst a number of iterations, each of them involving a fixed number of walkers that coincide, $W = 3$. What is not fixed is the total number of coincidences a node can have (the maximum possible being the number of iterations i). Thus, we count how many times $W = 3$ walkers traversed the node independently.

The statistics that we gather are the following:

Earliest Cardinal Coincidences These are the coincidences with the earliest time amongst all i iterations.

This is different from the earliest coincidences statistic in *2.3.1 Number of coinciding walkers*, since it considers the time of coincidences, collected from a number of independent iterations, instead of a single one.

Most Frequent Coincidences These are the nodes which were coincident in the highest number of iterations.

This statistic differs from the one in *2.3.1 Number of coinciding walkers* since frequency isn't defined by the number of walkers but by the number of independent runs.

Earliest Ordinal Coincidences Having stored the order at which the coincidences happened for each iteration, we take an ordinal approach: the first earliest coincidence is the node, which was the first coincidence for the highest number of iterations, the second earliest coincidence is the node, which was the second coincidence for the highest number of iterations, and so on. We gather the first c coincidences for a number of iterations (i.e. walks) i. We then count how many times a node was a coincidence and return the top rt nodes for each ordinal value.

A difference here is that we look whether the solution is present within $c{*}rt$ coincident nodes, and not c, thus we expect the performance to be higher.

2.3.3 Evaluation

It is important to note that our models do not have an explicitly defined "evaluation stage", which is usually part of a creative process, according to cognitive science. Although we have two algorithms, that we call our models, we have more than one kind of output, from which to quantify a model's success. This leads to an implicit division of the models into sub-models. The object of evaluation of the two models that we discuss next (evaluation here referring to the research process) is therefore different not only depending on the model, but also depending on the statistic (broadly, earliest or most frequent coincidences).

Once we have the different kinds of coincidences according to the two algorithms, as well as the different kinds of outputs that we look at, we calculate the percentage of RAT examples (25 in total), where the solution word is present amongst the top ones. We also calculate how many times each of the examples was solved. We compare these results with the human results, taken from [Smith et al., 2013]. We calculate the Pearson correlation coefficient for the accuracy by example, as done in [Kajić et al., 2017]. We find the parameter ranges, where the results are the closest: we look for the parameters for which the percentage solved examples is within an error of 5% around the human result, and for which the Pearson correlation coefficient is $r \geq 0.5$.

We validate the result by performing the simulations on randomised versions of the FAN, where the randomisation consists in performing a number of double-edge swaps. A swap removes two randomly chosen edges u-v and x-y and creates new edges u-x and v-y. This removes correlations in the network while keeping the degree distribution fixed. If the structure of the association network is important for the creative process, we expect to see poor results with these randomised networks.

The FAN network is an aggregate of multiple people's associations across many years of free association experiments. In RAT experiments with individuals, some people perform better than others. We account for this variance in creative ability by checking for a hypothesis on the network structure level. It has been shown that high semantic creative individuals have association networks with stronger connections between the network's components than low semantic individuals, while the connections within components are not different [Kenett et al., 2018]. We check whether this hypothesis holds for the model in our work by running the simulations with either the inter- or intralinks of the network lowered by 50%. We use two algorithms for community detection in networks in order to categorise the links: Louvain [Blondel et al., 2008] and Infomap [Rosvall et al., 2009]. We expect to see worse performance with lower interlink weights.

We proceed to describe the semantic network we simulated the walks on. In section *3.1 Structural Analysis of the FAN* we characterise its structure by measuring the properties introduced in section *2.1.2 Structural parameters*.

2.4 The Free Association Norms Network

The network we used in this work is derived from the Free Association Norms dataset [Nelson et al., 2004]. The dataset was created via a large number of experiments for free associations with over 6000 people during a few decades. Participants were given a cue word and had to report the first word that came to mind. A distribution of associate words was thus created for each cue, where the density of the associate word was the number of times it was given as an answer, normalised by the number of participants. The number of words in the network is 5018. The original graph is weighted and directed: more people can give a cue word as an associate to another than the other way around. We consider the unweighted version by setting a weight to 1 if it is more than 0. We make the networks undirected (or symmetric) by summing the matrix with its transpose.

The authors Kajic et al have shown that the FAN is a good network to model the RAT test [Kajić et al., 2017]. We used their code to load the association matrices and create the unweighted and symmetric versions of the network.

We have focused on the 25 RAT problems as in Smith et al [Smith et al., 2013] where they were selected from multiple sources with the constraint that all of the cues and answers were unique, therefore representing multiply-constrained tasks. These are listed in table 3 in the Appendix.

2.5 Implementation

The structural analysis of the networks, as well as their randomisation, was made with the Python library networkX [Hagberg et al., 2008]. Louvain community detection was done with the framework CDlib [Rossetti et al., 2019], which builds functions for community detection and evaluation methods on top of networkX. Infomap community detection was done with the Python library Infomap [Edler et al.,]. The loading of the association matrix FAN was done with the code developed in [Kajić et al., 2017]. The implementation of the random walk and the rest of the results was done by the author in Python.

3 Results

3.1 Structural Analysis of the FAN

Table 1: Structure of the FAN.

Parameter	Unweighted FAN	Weighted FAN	ER
Nodes	5018	5018	5018
Edges	55 232	55 232	55 272(250)
Mean degree	22	-	22.0(1)
Mean link weight	-	0.035	-
Density	0.0044	-	0.0044
Transitivity	0.08	-	0.0044(1)
(Local) Clustering	0.008	0.19	0.00441(9)
Average Path Length	3.04	0.03	3.031(4)
Diameter	5	0.595	4.2(4)

An Erdos-Renyi network with the same number of nodes and a probability for the creation of a link, proportional to the number of edges of the FAN, was created in order to compare its structure to the FAN. The results are shown in table 1 and discussed below.

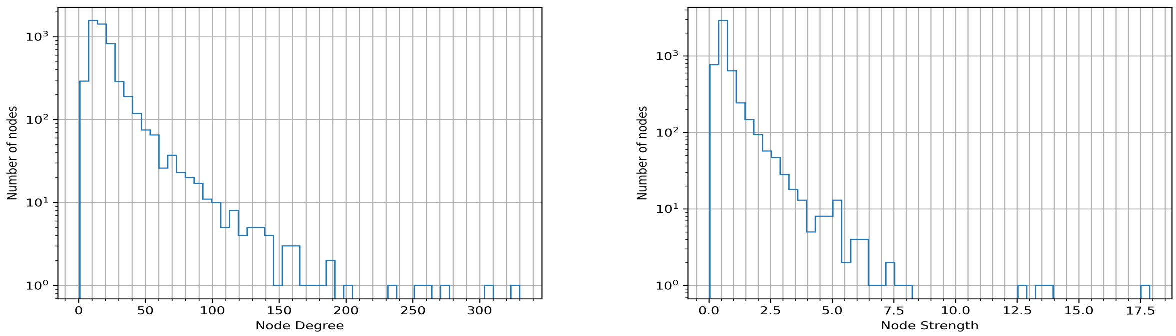
Number of Nodes and Edges The number of nodes of the FAN is $n = 5018$ and the number of edges is $m = 55,232$.

Degree and Total Weight The unweighted network has the most nodes with degree between 1 and 30, and very few nodes with degree higher than 150 (figure 1a).

For the weighted network, instead of the degree distribution, we calculate the distribution of the node strengths. The strength of a node is the sum of the weights of its edges and represents its total weight.

The distribution of node strengths has a similar shape to the degree distribution. There are many nodes with total weight between 0 and 2.5 and very few with total weight above 7.5 (figure 1b).

Density The density of the unweighted FAN is 0.0044, which is quite low.



(a) Unweighted FAN

(b) Weighted FAN

Figure 1: The degree distribution of the unweighted FAN and the node strength distribution of the weighted FAN

Global Clustering The Newmann transitivity of the unweighted FAN is 0.083, which is higher than the transitivity of the corresponding ER network - 0.00440(11). This shows that the fraction of paths of length two that are closed is much higher in the FAN, thus it is more connected.

Local Clustering Local clustering can be calculated for both the weighted graph: 0.19, and the unweighted graph: 0.008.

The graph has a considerably lower probability of a pair of node's connections to be connected between themselves, when the average weight between a pair of nodes is considered. The result for the weighted local clustering can vary depending on the generalisation method [Saramäki et al., 2007], but this is out of the scope of this work.

The local clustering of the equivalent Erdos-Renyi network is 0.00441(9) which means that there are much more triangles between nodes in the FAN than at random, whereas this might not be the case for the weighted network.

Average Path Length and Diameter The average path length for the unweighted network is 3.04 and its diameter is 5. For the weighted network these quantities are 0.03 and 0.595 respectively.

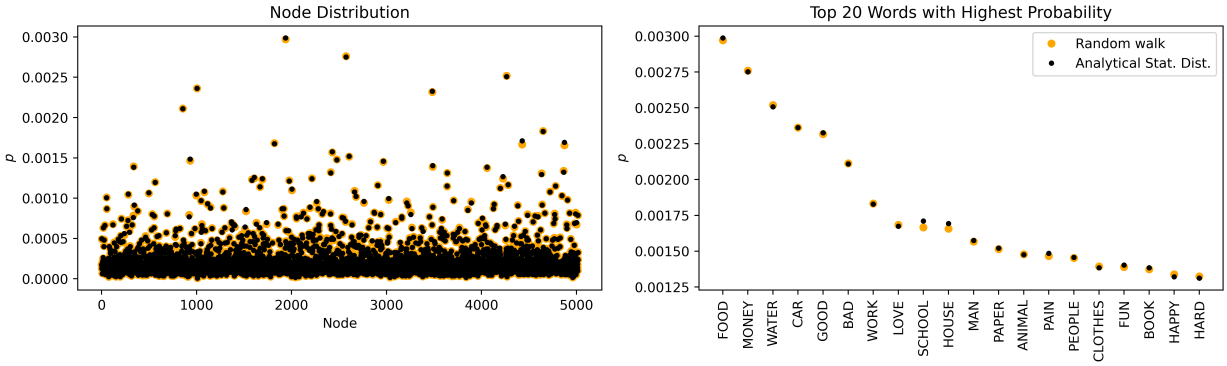
The average path length of the equivalent ER graph is 3.0305(41) and its diameter is 4.2(4). Thus, the unweighted FAN is identical to a random graph in this respect.

Comparing the ratios of the quantities between the two networks, the ratio between the average distance between nodes and the maximum distance in the case of the weighted network is much lower. This might indicate that there are a few nodes which are much further away in the weighted network than in the unweighted.

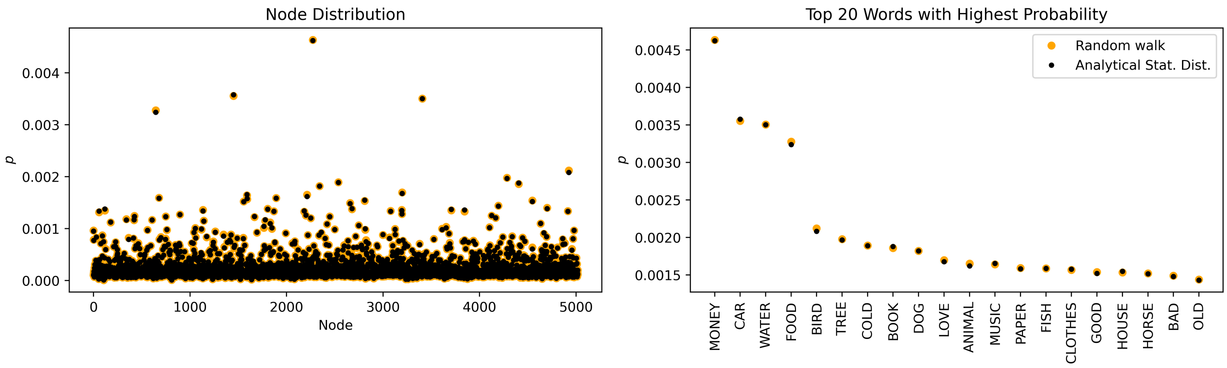
3.1.1 Stationary distribution

We simulate random walks on the networks for a long time to confirm that it matches with the analytical stationary distribution. The analytical distribution for the unweighted network is equation 14 from section *2.2 Random Walks* and the generalisation $p_i = \frac{S_i}{\sum_{l=1}^{N} S_l}$ is calculated for the weighted network, where S_i is the node strength.

The density of a node from the random walk was calculated by taking the total number of times it was



(a) Unweighted FAN



(b) Weighted FAN

Figure 2: The stationary distributions for the unweighted and weighted FAN. The simulation was made with $W = 100$ walkers starting from random nodes on the network for number of steps $t_{max} = 10^5$.

traversed amongst the walkers, divided by the total number of time steps for all walkers:

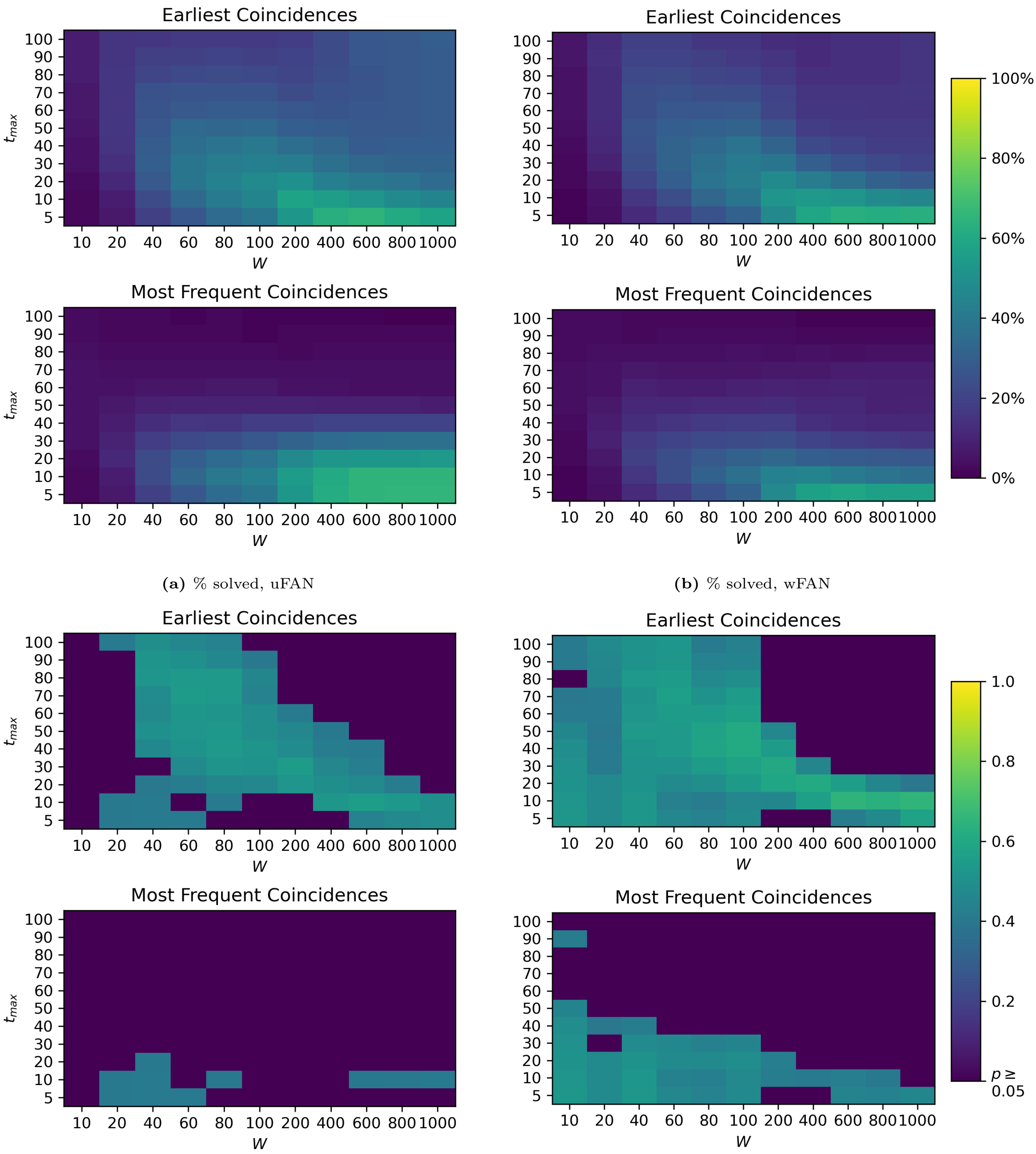
$$p_i = \frac{\sum_{w=1}^{W} T_i^w}{t_{max} W}, \tag{15}$$

where T_i^w is the number of times walker w traversed node i and t_{max} is the length of the simulation.

Figure 2 shows the node distribution for both networks, according to the simulation and the analytical result. It also shows the top 20 words with highest probability to be traversed, which are also the 20 most central nodes according to the degree centrality of the network (total weight centrality for the weighted one). The networks share 12 out of their top 20 most central nodes.

3.2 Coincidences between Random Walkers

We proceed to discuss the results from the two proposed algorithms which attempt to model the search for solutions during the RAT. Due to the fact that we have more than one kind of output, from which to quantify a model's success, we have a division of the model into two categories. Although we abstain from



(a) % solved, uFAN

(b) % solved, wFAN

(c) r, uFAN

(d) r, wFAN

Figure 3: Results for the "Multiple Walkers" model, collected amongst the top 10 earliest coincidences and the top 10 most frequent nodes for different values of the number of walkers per cue W and time steps t_{max}. They are averaged over 100 runs. **(a. & b.)**: Percentage solved RAT examples for the unweighted (a) and weighted (b.) FAN. **(c. & d.)**: Pearson correlation coefficient r between the human accuracy per example and the model accuracy per example for the unweighted (c) and weighted (c.) FAN.

explicitly defining them as sub-models, the language we'll use will allude to this idea. We might refer to a statistic as "performing" or "having a performance", which seems appropriate in this context.

3.2.1 Multiple walkers

We compare the percentage solved examples and the accuracy by example between the models and the human data (taken from [Smith et al., 2013]) in order to evaluate the proposed random walks models. We start with the model *2.3.1 Multiple Walkers* on the unweighted and weighted Free Association Norms network. The results for a different number of walkers W and time steps t_{max} are averaged over 100 runs. We look for the correct solution within the top 10 earliest coincident nodes (EC) and the top 10 most frequent coincidences (FC), and we calculate the percentage solved examples and the accuracy per example.

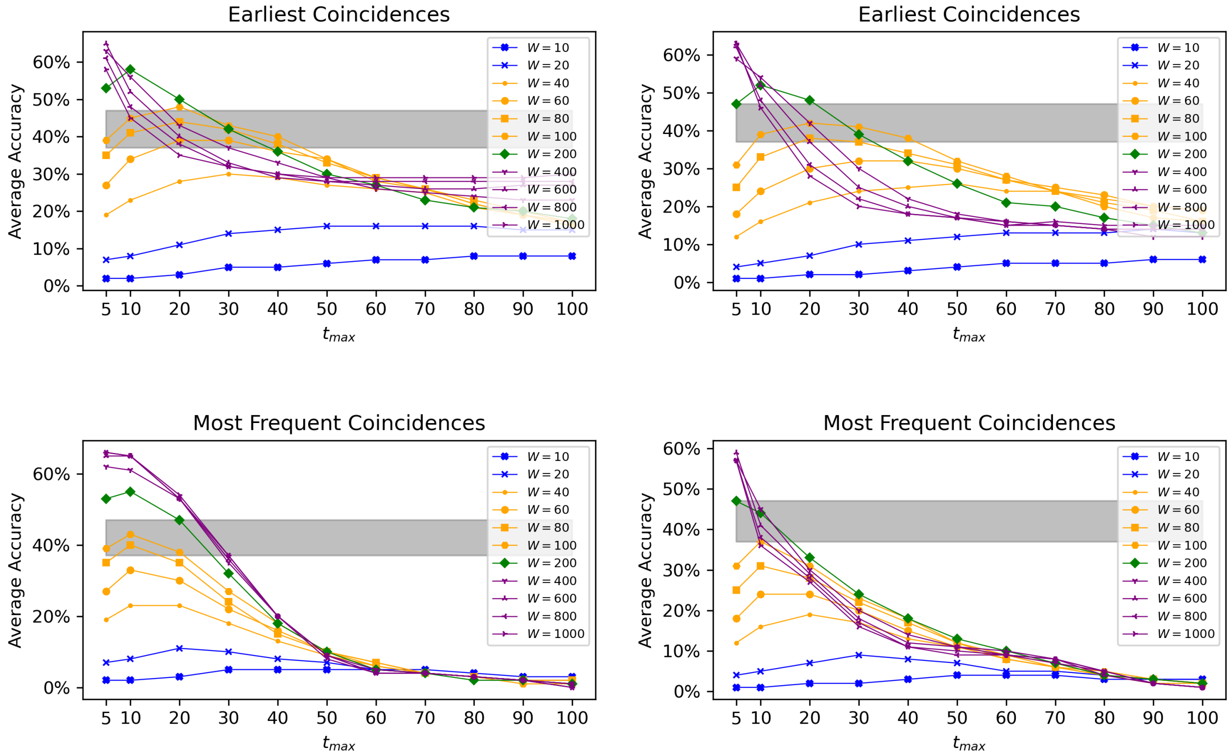
We first look at the model performance across all RAT examples regardless of the average human performance. The results of the simulations are shown as heatmaps in figure 3a for the unweighted FAN and in figure 3b for the weighted FAN. For both statistics, the highest accuracy is achieved with a higher number of walkers. The earliest coincidences statistic shows moderate accuracy for a high number of time steps, if the number of walkers is low, while in the case of most frequent coincidences, the performance is poor with a number of time steps higher than 50, for any number of walkers. FC for the unweighted network achieves the highest accuracies overall. EC shows a pattern of a negative relationship between number of time steps and number of walkers.

The performance as a function of time steps for different walkers W can be seen in figure 4. Let's focus on EC (the earliest coincidences) for the unweighted network, the top plot in figure 4a. The blue colour denotes W for which the performance increases with higher t_{max}. These are the lowest W. For $W \in [40, 200]$ the function increases until reaching a peak, $t_{max} = 20$ in yellow or $t_{max} = 10$ in green, and then decreases. When the walkers are $W > 200$, the performance decreases with t_{max}, shown in purple. The case is similar for the weighted FAN, the top plot in figure 4b.

FC (the most frequent coincidences statistic) show a different pattern. Performance is the highest for $t_{max} \in [5, 10]$ across all W, with values similar to these for EC. The negative slope of the drop of performance is steeper, and reaches a low value of around 5% at $t_{max} = 50$. Consequently, FC shows poorer performance across all W for $t_{max} > 50$. This difference between the statistics for high t_{max} can be explained with the fact that when the walkers are let to run for a longer time, they start reaching the stationary distribution of the network, meaning that the majority of the nodes they visit the most are the most central nodes of the network and are rarely the solution to the RAT tasks (see fig. 10 for an example of the outputs of single runs, the examination of which led to this conclusion). For low values of t_{max}, it seems that an increase in time affects the performance relatively in the same way, while this is not the case for EC, where the shape differs depending on W.

Let's now investigate whether the model performance is comparable to human performance. We have chosen as the margin of interest the average human performance, $42 \pm 5\%$. This is denoted with the grey area in figure 4. We can see from here that certain parameter sets fall in this area. These might be the parameters that make the model better for matching the human data, and not the ones which achieve the optimal average performance.

The second way we compare the human data to the model is by looking at the Pearson correlation coefficient between the example accuracies, which we can see in figures 3c (unweighted) and 3d (weighted). The colour indigo in c) and d) denotes results which are not considered due to their p-value being $p \geq 0.05$. We see that EC has a wider range of parameters with a good correlation of $r \geq 0.5$. EC shows a similar trend to the average accuracy: a negative relationship between the time and the number of walkers. EC for the weighted network seems to have the highest $r \in [0.6, 0.7]$ for certain values of W and t_{max}.



(a) Unweighted FAN

(b) Weighted FAN

Figure 4: "Multiple Walkers" model. Average accuracy as a function of the maximum number of time steps t_{max} for different values of the number of walkers W. The area in grey is the human performance, 42%, within an error of 5%.

It can be argued that, after inspecting the coincidences, this relationship between W and t_{max} is due to the simulations reaching the stationary distribution, similarly to FC. The resulting pattern is different because the coincidence time is taken as an average between walkers. Few walkers need more time to coincide, and the probability to coincide on a solution grows until reaching a peak. With many walkers, the earliest intersections have high probability to land on the solution. With more time, the average of the traversal times changes, and the earliest landings become the words with the highest network frequency.

It is interesting to note how FC has the highest average performance, yet the correlation by example is low. This might be an indication that having more walkers means reaching the solutions to certain examples with almost perfect accuracy, making the average higher, but moving away from the human accuracy for these examples.

We now attempt to find the parameters for which both average performance and performance by example lead to the best match with human data. Figure 6 shows the discrete versions of the heatmaps in figure 3. While we are interested in the highest accuracy by example correlation, coloured in dark and light green in c) and d), we don't necessarily want to have the highest average percentage solved, but rather the one that falls within a range of the human result. This is the range $42 \pm 5\%$, coloured in green in a) and b).

Figure 6 shows the intersection of the parameters, for which these two constraints are fulfilled: $42 \pm 5\%$ for average performance and $r \geq 0.5$. The optimal parameter sets for the unweighted FAN fall between

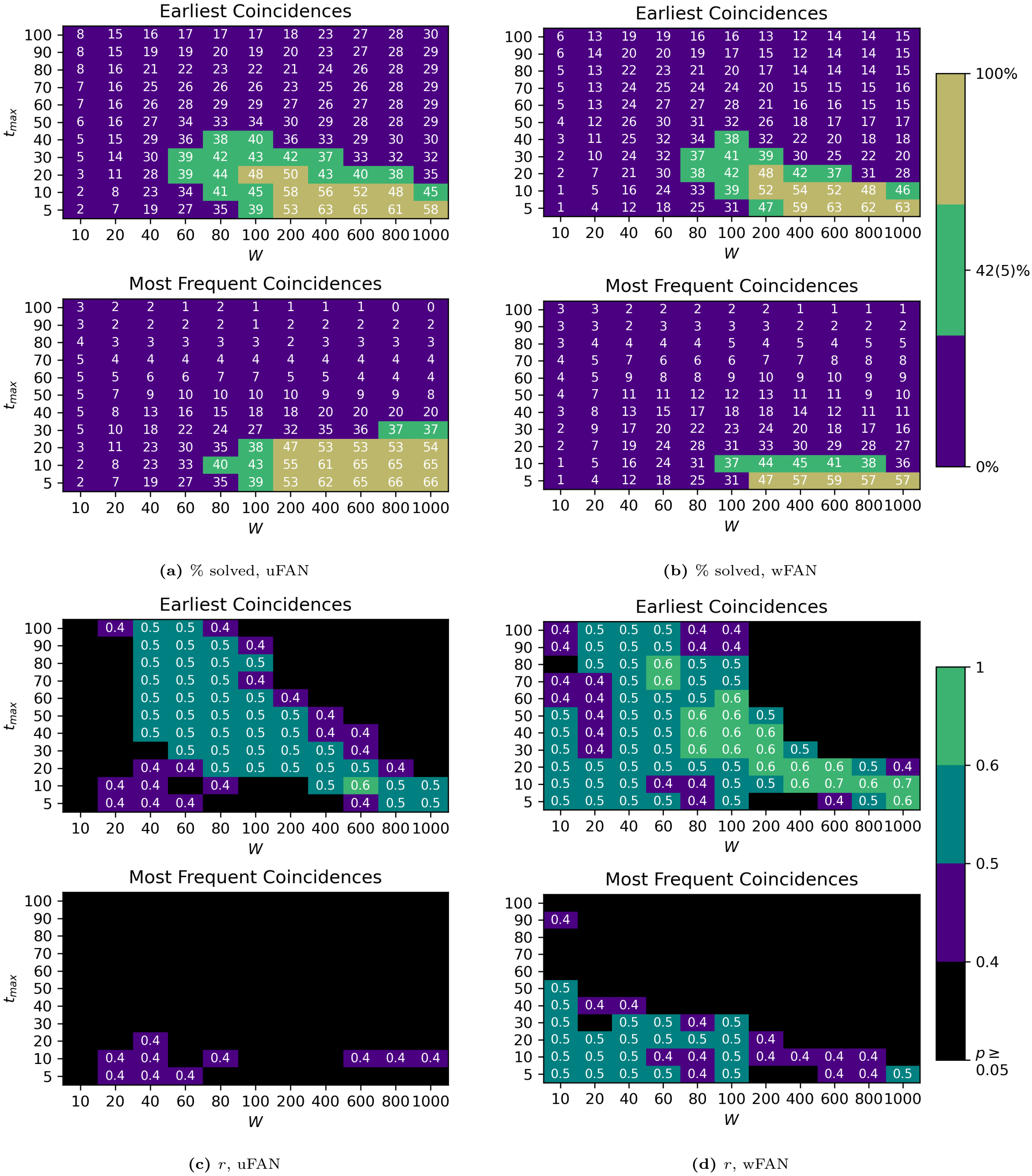
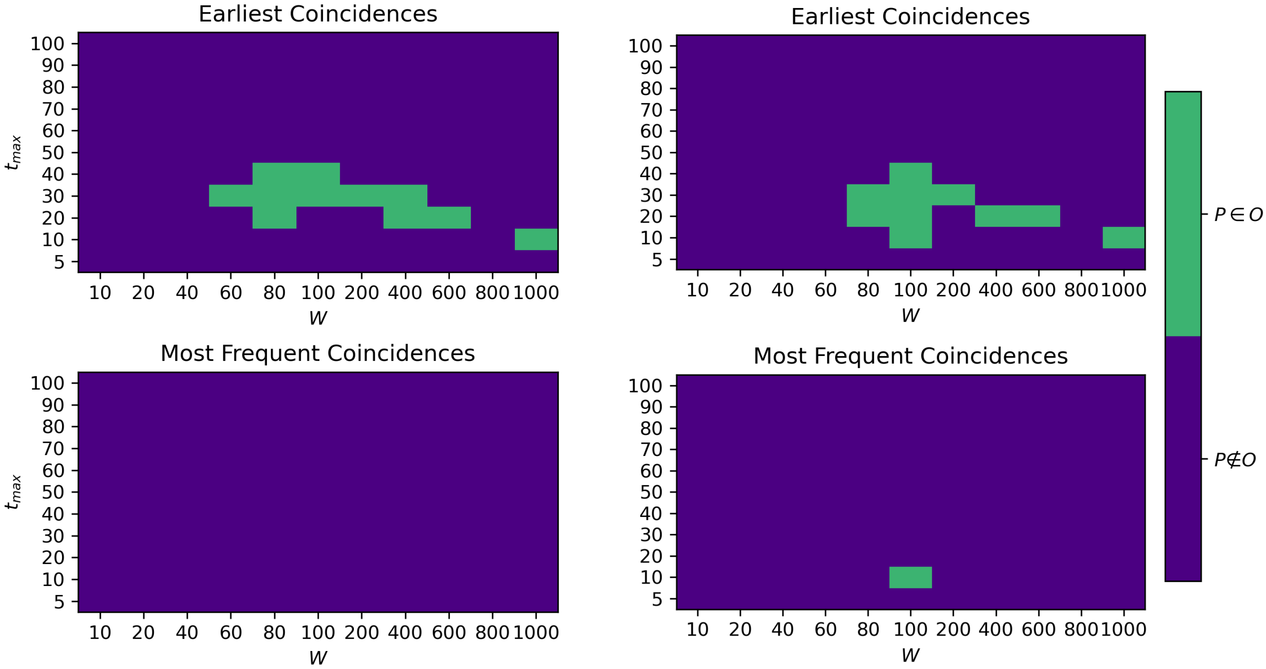


Figure 5: Results for the "Multiple Walkers" model on the unweighted and weighted FAN, collected amongst the top 10 earliest coincidences and the top 10 most frequent nodes for different values of the number of walkers per cue W and time steps t_{max}. They are averaged over 100 runs. **(a. & b.)**: Percentage solved RAT examples; the parameter range, for which the average performance is the closest to the human result, is 42(5)% (in cyan). **(c. & d.)**: Pearson correlation coefficient r between the human accuracy per example and the model accuracy per example.



(a) O, uFAN

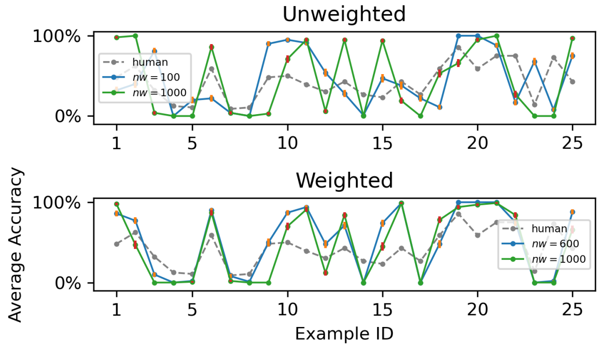
(b) O, wFAN

Figure 6: Results for the "Multiple Walkers" model on the unweighted and weighted FAN. The parameter sets P, where the percentage solved examples is within the human range $42 \pm 5\%$, and where the Pearson correlation for accuracy by example is $p \geq 0.5$

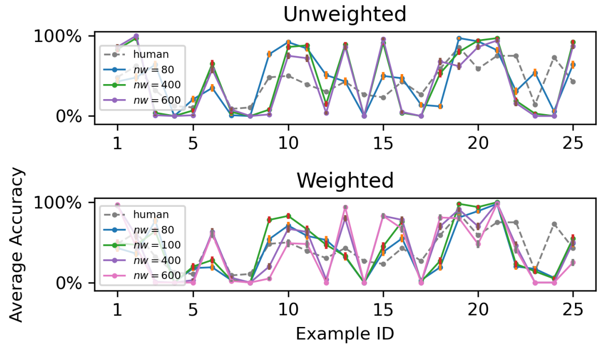
$W \in [60, 600]$ and $t_{max} \in [20, 30, 40]$ for EC. There is an additional parameter set for a bigger value of walkers and a smaller time, $W = 1000$ and $t_{max} = 10$. The optimal parameters for the weighted FAN have similar ranges for the time. The minimum for number of walkers is slightly higher, $W = 80$. Additionally, the weighted FAN has a value within the constraints for FC, $W = 100$ and $t_{max} = 10$. Overall, we see a variety of parameters for which the model performance is comparable to the human performance according to these metrics. A discussion about the differences between the results for the two networks can be found in section *4.2 Network Comparison*.

Figure 7 shows the accuracy by example against the human result for the optimal parameter sets according to the conditions above. It looks like there are examples for which the model generally performs better than the humans and examples for which it almost never finds the solutions. The results are categorised by the number of time steps since there is a qualitative pattern in the fits. It seems that there are a few examples which are very easy for the model for a few time steps, but they become harder for more time steps, and they get closer to the human values. On the other hand, the ones which are difficult for the model do not vary as much as a function of the time steps. This is addressed further in section *4 Discussion*.

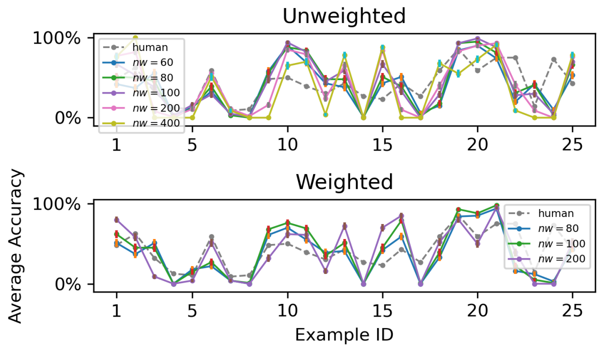
We perform the same simulations with randomised versions of the FAN to validate the results. The networks are randomised by swapping the edges between a number of node pairs. This is done for a fraction of the number of edges m in the network $f_m \in [0.001, 0.01, 0.1]$. The set of parameters shown are for $W \in [10, 20, 40, 60, 80, 100]$ and number of time steps $t_{max} \in [5, 25, 50, 75, 100]$, as well as $W \in$



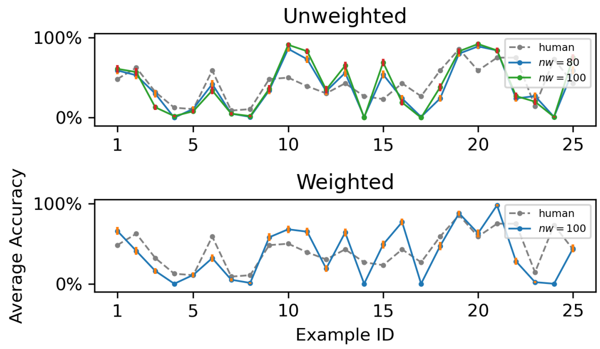
(a) $t_{max} = 10$



(b) $t_{max} = 20$



(c) $t_{max} = 30$



(d) $t_{max} = 40$

Figure 7: The average accuracy by example for the best parameter sets for the earliest coincidences of the "Multiple Walkers" model on the FAN. The human results are shown in grey. Each subplot shows the sets for a specific value of the number of maximum steps t_{max} for the unweighted network (top) and the weighted one (bottom).

$[200, 400, 600, 800, 1000]$ and $t_{max} = 100$ [1]. We see the results in figure 8. Here the purple points are runs with $f_m = 0.001$, the blue points are with $f_m = 0.01$ and the yellow ones are with $f_m = 0.1$. Overall, randomisation leads to worse performance, proportionally to the number of swapped edges. The weighted network (left and right triangle markers) appears to be more sensitive to randomisation, since it varies less with f_m. The unweighted network, on the other hand, shows similar outputs to the original one with a small f_m. Overall, these results show that the structure of the association networks is important for the RAT performance.

We show specific RAT examples in figures 10 and 11 to demonstrate the two statistics we gather in the simulations and to show how randomisation affects the results. The statistics, earliest coincidences and most frequent coincidences, are gathered by sorting all the coincidences amongst the walkers in the simulation by the average time of the walkers traversing the node, on one hand, and by the number of walkers having traversed the node, on the other hand. The x-axis shows the coincident word, the left y-axis (the bars) shows the percentage trajectories (or walkers), where the node was present (that traversed the node), and the right y-axis (the yellow stars) shows the time of the coincidence (the mean of the times the walkers traversed the

[1] The first set of simulations are averaged over 25 random networks per fraction and 50 independent runs per network and the second set - over 20 network realisations and 25 runs per realisation. This was done in order to shorten the runtime.

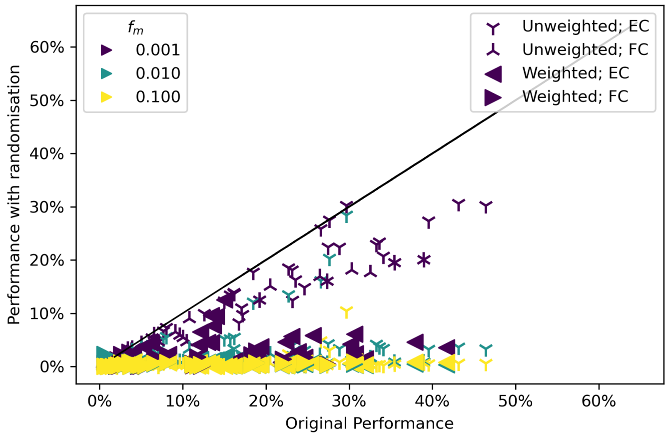


Figure 8: Performance versus performance with a randomised network. Each data point is a run for a $(W,\, t_{max})$ pair. The colours denote the level of randomisation: with either 0.001, 0.01 or 0.1 of the edges swapped. The markers denote the kind of network, unweighted or weighted, and the kind of output, earliest coincidences (EC) or most frequent coincidences (FC). The black line is the equality between the original and randomised networks' performance.

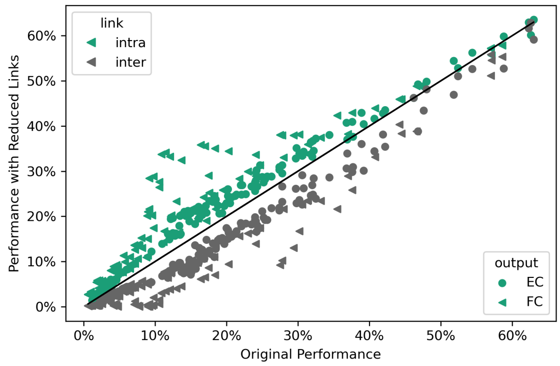
node).

Figure 10 shows example 22 from the RAT, with the three cues being "tooth", "potato" and "heart", and the solution "sweet". Subfigure 10a shows the result of the original FAN, where the solution was found (annotated with a red circle). For a small number of time steps, below 40, a lot of the coincident words in both statistics overlap, since the number of coincident words is smaller for shorter simulation time. Figure 10b shows the result after randomising a fraction of the edges $f_m = 0.001$. As in figure 8, the result is the average for 25 random networks and 50 simulations per network. Generally, the randomised networks produce different trajectories, due to the different links and weights between the nodes. With a small fraction of the edges affected, the solution word is often times still found, even if the trajectories are overall different. In example 22, only 2 out of the 10 earliest and 2 out of the 10 most frequent coincidences overlap with these of the original FAN (for this specific simulation), yet one of them is the solution word "sweet". With more swappings of edges, this becomes the case more rarely: the example given for $f_m = 0.01$ shows no overlapping with the outputs of the original FAN. Another result of the randomisation is less overlap on the same coincident node between walkers. Note that the highest percentage of trajectories for a coincidence is 8% in the case of the original FAN, while it is 3.5% in the case with $f_m = 0.01$.

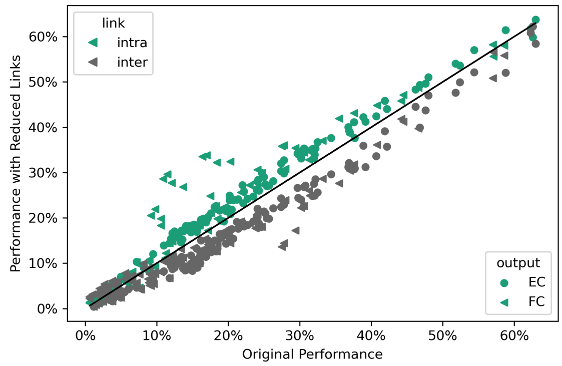
We see no results found after randomisation in the second example given in figure 11, for the task with cues "due", "life", "tense" and solution "past".

Finally, we have accounted for variance in creative ability by running the random walks on a slightly different structure. The original results were compared to the results of the simulations on versions of the FAN network with 50% lower weights of the links between components, as well as with versions of the FAN with 50% lower weights of the links within the components. We used Louvain and Infomap community detection to characterise the types of links. The modularity of the resulting split according to Louvain was 0.64, while for Infomap it was lower: 0.55 for depth 1 and 0.56 for depth 2.

Both algorithms show similar results (fig. 9): performance is slightly worse with the weakening of the inter-links, and slightly better with the weakening of the intralinks. Quantitavely comparing the two algorithms, the change in results is slightly bigger for Louvain. Weakening intralinks is effectively creating the opposite



(a) Louvain

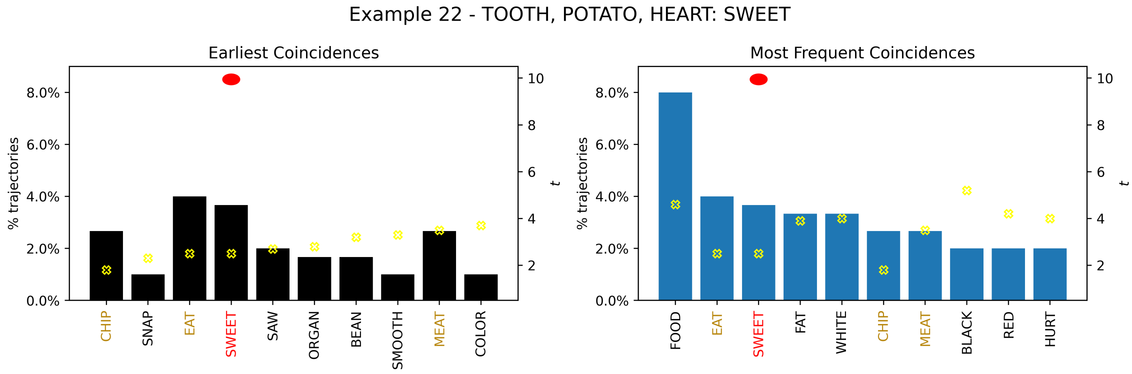


(b) Infomap

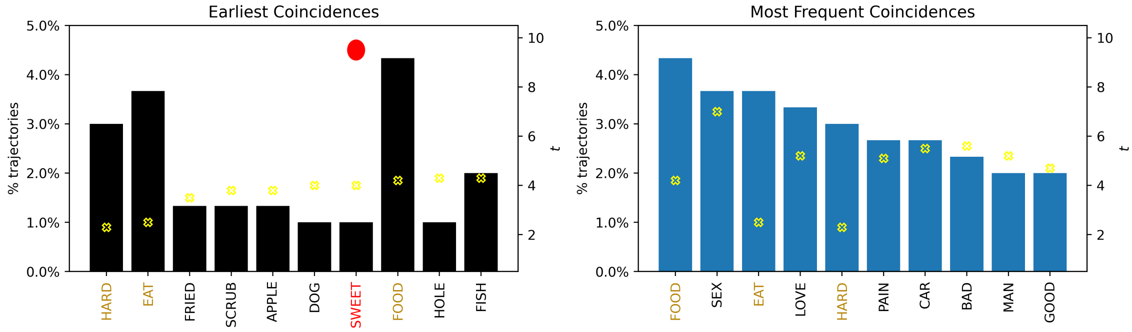
Figure 9: The ratio between the original result and the result with reduced intralinks (blue) and interlinks (grey). The categorisations of the links are done with Louvain community detection (a) and Infomap (b). Each data point is a run for a particular number of walkers W and t_{max}, for either earliest coincidences (circles) or most frequent coincidences (triangles). The values of W and t_{max} are these used for the main result in figure 3.

change in relative difference between inter- and intra-links, leading to the opposite effect. In terms of output statistic, the most frequent coincidences (the triangle marker) appears to be more sensitive to the change in links for some parameters.

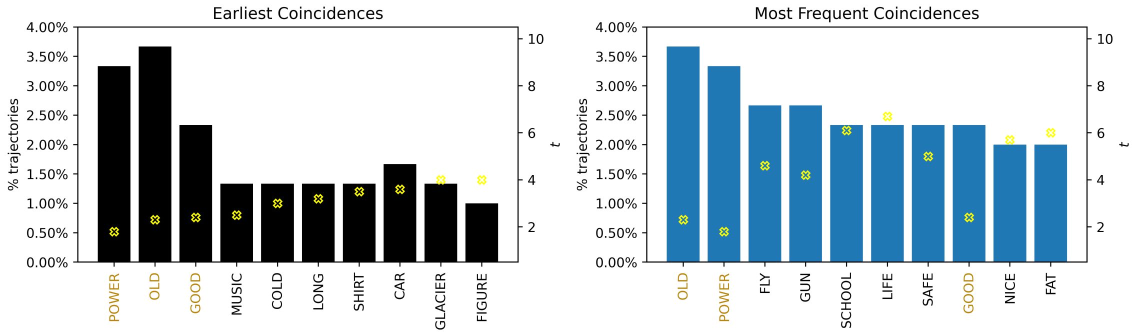
This result shows that the "Multiple Walkers" model is compatible with the findings that stronger interlinks in highly creative individuals lead to better performance [Kenett et al., 2018].



(a) Unweighted FAN

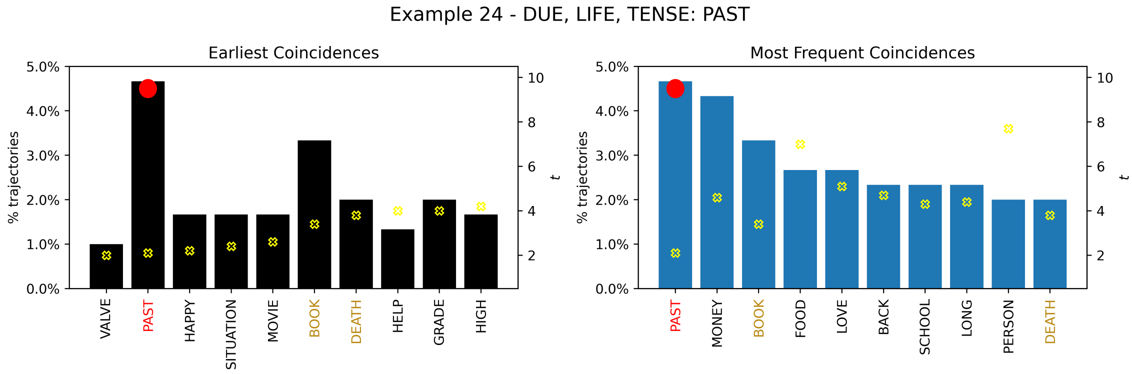


(b) Randomised with $f_m = 0.001$

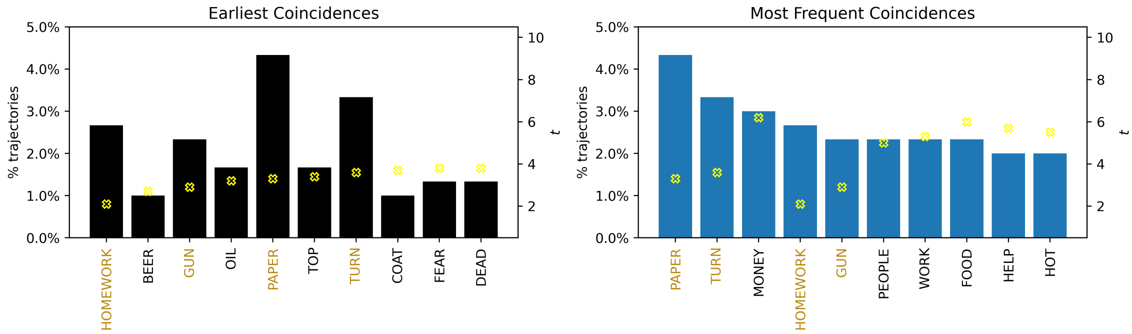


(c) Randomised with $f_m = 0.01$

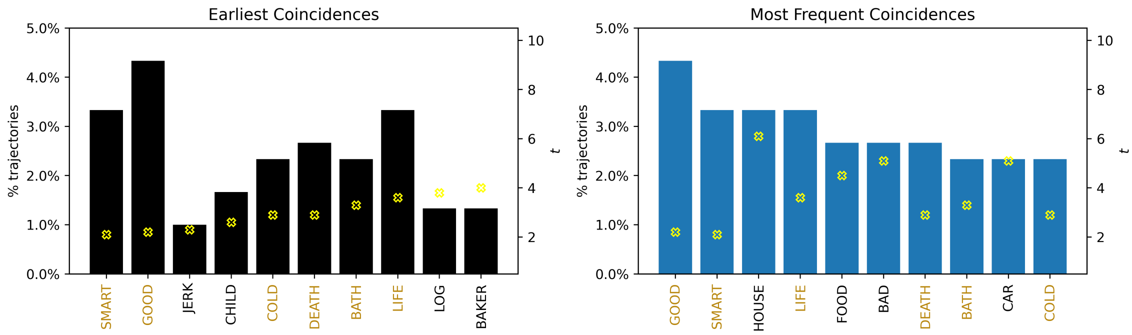
Figure 10: Top 10 earliest coincidences (left) and top 10 most frequent coincidences (right) for one example of the RAT for the unweighted FAN, as well as for a randomised FAN with fraction of edges randomised $f_m = 0.001$ and $f_m = 0.01$. Number of walkers per cue is $W = 100$. The x-axis shows the word, the left y-axis shows the percentage trajectories, where the node was present, and the right y-axis shows the time of the coincidence, which is the mean of the times the walkers traversed the node. The bars show the % trajectories and the stars show the coincidence time. The solution is annotated with a small red circle above the word's bar (if present) and its label is in red. The labels of the words which are present in both statistics (the left and the right plot) are in ochre.



(a) Unweighted FAN



(b) Randomised with $f_m = 0.001$



(c) Randomised with $f_m = 0.01$

Figure 11: Top 10 earliest coincidences (left) and top 10 most frequent coincidences (right) for one example of the RAT for the unweighted FAN, as well as for a randomised FAN with fraction of edges randomised $f_m = 0.001$ and $f_m = 0.01$. Number of walkers per cue is $W = 100$. The x-axis shows the word, the left y-axis shows the percentage trajectories, where the node was present, and the right y-axis shows the time of the coincidence, which is the mean of the times the walkers traversed the node. The bars show the % trajectories and the stars show the coincidence time.

3.2.2 Triplets of walkers

The model "Triplets of Walkers" performs considerably worse. Table 2 shows the percentage solved performance for the weighted and unweighted FAN, for number of iterations $i \in [10, 25, 50, 100, 200, 400]$, averaged over 100 simulations.

An example of the Earliest Ordinal Coincidences output for a single run is shown in fig. 12 in the Appendix.

We haven't performed the rest of the analysis which was done for the model "Multiple Walkers" in 3.2.1, since the percentage solved is considerably lower than the human result. More about the results of this algorithm and its comparison to the first one follow in the section *4.1 Model Discussion*.

Table 2: Percentage of solved RAT examples for earliest (cardinal) coincidences and the most coincident nodes, according to Algorithm 2, for the weighted and unweighted FAN, as well as the earliest ordinal coincidences. The nodes are aggregated from a number of walks of 3 walkers. The results are averaged over 100 runs.

		Percentage Solved Amongst		
weighted	walks	Earliest Coincident Cardinal	Most Coincident	Earliest Coincident Ordinal
0	10	1.36(23)%	0.20(9)%	3.84(33)%
0	25	2.44(29)%	0.32(11)%	9.20(51)%
0	50	3.08(31)%	0.20(9)%	10.24(49)%
0	100	3.60(37)%	0.04(4)%	6.2(4)%
0	200	2.36(32)%	0(0)%	3.88(34)%
0	400	2.08(28)%	0(0)%	2.56(27)%
1	10	1.08(19)%	0(0)%	3.96(39)%
1	25	2.48(30)%	0.16(8)%	4.80(42)%
1	50	3.08(36)%	0(0)%	5.64(48)%
1	100	3.20(34)%	0.04(4)%	4.64(37)%

4 Discussion

Having started this work with the goal of modelling creativity, we chose a semantic association problem and adopted a network science approach. More specifically, we modelled the cognitive process of solving a semantic creativity problem with random walks on a semantic network of word associations. The model consists of letting multiple walkers start from each given cue of the remote associates task and perform random walks on the semantic network for a given time. This is implemented in two algorithms on a weighted and an unweighted associations network. The walkers coincide on nodes which are often the solution, and their performance is comparable to human performance for certain parameter values on both networks for the algorithm which involves multiple rather than triplets of walkers at a time. What follows is a discussion of the results and the neural plausibility of the proposed search process, as well as a section on limitations and future work.

4.1 Model Comparison

We chose two different algorithms for the random walks: "Multiple Walkers", where the solution is found amongst coincidences between multiple walkers starting from each stimulus in a single run, and "Triplets of Walkers", where the solution is found amongst coincidences between 3 walkers aggregated from multiple runs. We saw that the "Multiple Walkers" performance was significantly higher, and was comparable to the average human performance, while "Triplets of Walkers" was not. This leads to the conclusion that having multiple walkers seems to be a better way to exploit the knowledge base of associations. A hypothesis for the good performance of the first algorithm might be that the response is more closely related to one of the stimuli and the majority of coinciding walkers started from this stimulus. In theory, this aspect can't be leveraged in the second algorithm. It would require further analysis to investigate this.

We proceed to discuss more broadly the results of "Triplets of Walkers", which were briefly commented in the previous section. The statistic "most coincident" nodes performs poorly because the responses belong to the stationary distribution, even for a small number of walks such as $i = 10$ (this has not been quantified but was a rather qualitative observation, when looking at single run outputs like in figures 10 and 11). Generally, it seems that the walkers steer away from the stimuli with the progress of time and the coincidences seem more random than these for the "Multiple Walkers". If we look at aggregating across several iterations as making multiple attempts, it seems that this kind of resetting does not help the process find the solution as much as having more walkers traverse the association network simultaneously.

It is worth noting that the "earliest ordinal" statistic performs slightly better than the "earliest cardinal", with results seeing a more pronounced improvement for i around 50. This might simply arise from the fact that it is more probable to find a solution amongst more coincidences, meaning that the logic to look at earliest time from an ordinal perspective does not lead to better results.

We conclude with a comment on the different assumptions behind the two algorithms: one has many parallel network diffusive processes, while the other has a small limited number, with the option to reset. The "Multiple Walkers" performance for number of walkers $W > 100$ yields higher results than the human average for a short number of time steps. This begs the following question: how cognitively plausible is it to have a high number of parallel searches? We know that the human brain is incredibly powerful and information processing is done in a highly parallel fashion, however this result might indicate that for such a high-level model of cognition, there must be a limit to the level of implicit processing involved. Another indication that the human-like W might be $W < 100$ is that the time for reaching the best average performance and by example is around 40 time steps, which might be closer to human time. The analysis in this work would benefit from a direct comparison between model time and human time.

4.2 Network Comparison

In this work we have used both the unweighted and weighted undirected FAN networks. Their structural characterisation shows that they are not very different. They have similar degree/node strength distributions, with the weighted FAN having fewer nodes of considerably higher total weight strength. Both networks also share a lot of their most frequent words. By contrast, the weighted network has much higher clustering than the unweighted one and its random equivalent.

The simulations for the RAT yield similar results for both networks. The unweighted one has slightly higher average performance, more pronounced for the most frequent coincidences statistic. This one falls outside the average human range, however. The weighted network shows significantly higher Pearson correlation for accuracy by example across bigger ranges of the parameters on both metrics.

This second result does not lead to the weighted FAN having a larger optimal parameter range. They are mostly similar for the two networks, with the difference that the weighted network has an optimal value for FC and the unweighted does not. Qualitatively comparing the accuracies by example in figure 7, it might be the case that the weighted network yields slightly lower accuracies for the examples which the model finds easier, which are closer to the human values. It remains similar to the unweighted network on the examples that the model finds harder than humans.

These differences indicate that using the weights might be better for the purpose of modelling average human response, since the medium difficulty examples seems to be fit better, while ignoring the weights might lead to generally more optimal RAT performance. It would require further analysis of the correlations between the model and the human data according to level difficulty, in order to investigate this directly. This is addressed in subsection *4.5 Limitations*.

4.3 Neural Plausibility

In this section we argue that the model is plausible not only from a cognitive perspective, but from a neuroscientific perspective as well. We expand on the related literature on the cognition behind creativity, discussed in section *1 Introduction*, and we add to it by discussing a number of neuroscientific studies that the model is grounded in.

The random walk model in this work is a direct reflection of the theory of activation spreading introduced by Quillian and extended in the seminal paper by Collins and Loftus [Collins and Loftus, 1975]. The authors describe memory search as activation spreading from two or more nodes in a semantic network which stops when an intersection occurs. The search involves going through all the first neighbours of the starting nodes in parallel, then all the first neighbours of the first neighbours, etc., tagging each traversed node with the starting node and its predecessor. An intersection is found when a tag of another starting node is encountered, and one could trace the path leading to it by following the tags. The theory involves the notion of semantic relatedness: the more properties two concepts have in common, the more links there are between them, thus semantic relatedness is proportional to the number of interconnections between nodes. The neural activity behind these processes is still hard to measure, but there are multiple studies related to activation spreading and language circuits in the brain, its effect in recognition tasks, the effect of neuromodulators on activation processes, etc. which are shedding light on the neural underpinnings of the phenomenon ([Meade et al., 2007], [Foster et al., 2011], [Pulvermüller et al., 2009]).

Evaluating/aggregating the coincident words in a certain way to come up with solutions is akin to the second of the two main stages described in theories of creativity. One always involves the generation of ideas (it can be a few different stages which overall have this function), and another is responsible for the evaluation of these ideas. Neuroimaging studies suggest that the generation phase is controlled by the default-mode network (DMN) and the evaluation phase - by the executive control network (ECN). The DMN has been found to be involved in mind wandering, future thinking, memory retrieval, schematic memory, and semantic integration, as well as during divergent thinking (DT) tasks, where the individual applies internally-directed and defocused attention in order to create/find multiple solutions [Kleinmintz et al., 2019]. Executive control processes, such as initiation, inhibition, working memory, flexibility, planning and error detection are associated with neural activity in different prefrontal regions which form the ECN. Several studies have shown how the two networks are connected during creative tasks and how they change in time and according to the context and the task. It is suggested that a cyclic alternation between states of search and cognitive control is what underlies creative thinking [Kleinmintz et al., 2019].

Finally, the asynchronous aspect of the coincidences in the model can be motivated by the phenomenon of temporal integration, which happens in the brain. It has been showed that temporal integration and segregation are governed by the presence of a hierarchy of intrinsic neural timescales [Wolff et al., 2022] in both unimodal and transmodal regions of the brain.

4.4 Related Work

Apart from the findings described in the Introductory section 1.3.1 on creativity from a networks point of view, it is worth discussing a recent study on the RAT [Valba et al., 2021], which approaches the problem of modelling associative creativity in a similar way. Its findings could not be incorporated in this work, unfortunately, because the authors were not aware of them during the formulation of the thesis' research question and hypothesis.

Both works are grounded on Mednick's theory for creative association and build off the findings that local undirected search might be a good strategy for solving a convergent thinking task [Smith et al., 2013]. The authors examine how associative networks' structure is linked to performance via random walks. They show how a Markov chain of associations can predict RAT accuracy well.

While the used models are similar, Valba et al answer a different research question: they investigate how people optimise solving the RAT. They do so by evaluating variations of the random walk search algorithms, introducing different heuristics and considering problem difficulty. The algorithms include one walker at a time, which is another significant difference, and traversal of the same node more than once is restricted in some of them.

On the other hand, this thesis investigates whether the random walk is a good model for human performance in the RAT on average. It involves the classical random walk but with multiple trajectories, introducing the notion of asynchronous coincidences. Coincidences are key in the model and they rest on the phenomenon of activation spreading in cognitive science. Therefore, the work here attempts to provide a more generalised mechanism and it can be argued that it is more neuroscientifically motivated, in that asynchronicity and the integration of multiple parallel processes agree with the phenomenon of temporal integration in human neural activity.

It would be interesting to improve the model here, while keeping the neuroscientific grounding, in order to answer the question addressed by Valba et al: how do people optimise for the RAT. This could imply accounting for problem difficulty, and attempting to answer the more difficult research question of whether the structure or the dynamics produce variance in skill.

In terms of results, we can make a tentative comparison due to the fact that the battery of RAT examples used is bigger, the analysis differentiates between easy, medium and hard problems, as well as due to the other differences in the modelling approach. The random walk with resetting model and the random walk with attraction to stimuli have Pearson correlation coefficients of 0.718 and 0.742, which are higher than the highest coefficient in the "Multiple Walkers" model, $r = 0.656$.

An interesting plane of comparison is the definition of the evaluation stage. While Valba et al define the RAT as a first-passage problem, here we have a slightly different approach. We let the models run for a number of time steps or number of coincidences, depending on the chosen algorithm. We then have two statistics: one measures earliest coincidences, or aggregated earliest coincidences (for the "Triplets of Walkers" model), which does not assume that the task is necessarily a first-passage problem. Rather, the idea is that it is likely that the output word will be found during the earliest average passing time of the walkers through it. The second statistic takes aggregations of the coincident words during the entire simulation time, which could potentially be an indication for the presence of a more hollistic checking procedure in humans. The question whether this procedure is implicit or explicit remains open and its plausibility could be investigated further in future work.

4.5 Limitations

We start the section by discussing the time involved in finding a solution. It is not trivial to find a way to make model time comparable to human time, thus this was left out of the scope of the thesis. It would be interesting to incorporate it in further work.

An important limitation is that the work only uses 25 RAT problems. These were chosen from a study of semantic search during the RAT where participants were asked to report every word that came to mind during the problem solving process. We used these in order to be able to collect sequences of coincidences and to compare them to the human results. This extension to the dissertation's goals was not implemented due to lack of access to the human data. Apart from a fuller evaluation of the proposed algorithms, this would allow for an investigation into the ways in which locality in the search process arises, which was shown in other studies [Smith et al., 2013]. More on how this can be approached follows in section *4.6 Further Work*.

We have used a very simple model here with the goal to generalise the creative association process. It has led to results which match human performance on average, however, we can make more fine-grained models if we include problem difficulty in the evaluation. It would be informative to see whether the proposed model is better or worse for specific levels of difficulty. This might indicate that variation of the random walk model can be investigated in order to match performance on the other levels of difficulty, as was explored in [Valba et al., 2021]. We expand on potential directions for more complicated models in the section below.

Another limitations due to a coarser approach is related to the starting cue of coinciding walkers in the model. We have not investigated whether the coinciding walkers tend to start from the same stimulus, nor the structural relation between the stimulus and the response for the specific examples. If walkers from the same cue lead to the solution, this might confirm certain findings of the search in the RAT being focused on one cue at a time [Abbott et al., 2015]. Consequently, random walk models with some sort of cue switching might be considered.

Finally, this work only looks at one association network, which was shown to be fit for modelling the RAT [Kajić et al., 2017]. A comparison between the structures of different knowledge bases in relation to the dynamical process of choice running on them could provide a fuller explanation of the creative process in humans. It could accent on certain aspects of the problem solving dynamics which remain hidden by limiting the modelling work to one particular network. There are studies that have analysed the fitness of different knowledge bases for the RAT, which might help further investigation in this direction.

4.6 Further Work

Smith et al have conducted a RAT experiment with humans, where they ask them to overtly report candidate solutions before giving a final one [Smith et al., 2013]. The words are placed in a multidimensional space and word similarity is measured as the cosine angle between two words. The analysis has shown that responses tend to be more related to one of the three cues, that locality in the search is partially driven by cluster search, and that subsequent responses that are similar correspond with shorter response times. If access to the human responses is given, it would be interesting to compare their word and time frequencies to these of the "Multiple Walkers" model, taking the chain of coincidences between walkers as output sequences. In addition, a similar analysis of bunching around cues and cluster search could be made via cosine similarity, as well as diffusion distance. These metrics could also be useful for filtering candidate outputs in order to match their number to human data, since the random walk model might produce a higher number of candidate solutions, given a higher number of walkers. Finally, the candidate solutions sequences analysis can help find a more refined set of optimal parameters for the model, the number of walkers and the number of time steps.

We have chosen to work with a simple random model in order to investigate whether it is enough to explain associative creativity. We have seen indications that this is the case on average. By incorporating problem difficulty, we could try to expand on the analysis by looking into the ways people optimise their search. More complicated dynamical processes such as random walks on clusters or with jumps could be incorporated, or some sort of heuristics for the search, for example such that the walkers do not 'forget' the starting nodes [Valba et al., 2021].

Comparing how these algorithms of varying complexity deal with tasks of different difficulty might also shed a light on how people differ in their creative ability on a dynamical level. Currently this work attempts to account for this difference only on a structure level. The findings here show that the average performance suffers from weaker inter-component links, while it improves when these links are stronger. This is an indication that the "Multiple Walkers" model is compatible with variance in creative ability resulting from differences in knowledge. Further modelling work on the dynamics, as well as work with individual human data, is required to address the question of whether people can succeed or fail to exploit their knowledge bases.

A more systematic evaluation of the model might benefit from the comparison of a battery of networks. A recent study of modelling the RAT evaluates the theoretical adequacy of the five knowledge bases, as well as which ones lead to the best model performance [Schatz et al., 2022]. According to their findings, the best networks are the Human Brain Cloud (HBC) [Gabler,] and Small WOrld of Words in ENglish (SWOWEN) [Deyne et al., 2019] word association networks. Future work can use these to compare performance of the "Multiple Walkers" model.

An additional way to extend this work can be to incorporate the modelling of the insight or the "aha" moment during the creative search. Currently the mechanism behind the evaluation stage is somewhat implicitly defined. We look for the top 10 coincidences based on different metrics, and the stopping criterion is a number of time steps (in the "Multiple walkers" model). How does the aggregation of solutions happen in the brain? What is the stopping criterion in humans? Is associative thinking a first-passage problem [Valba et al., 2021], where the correct solution is instantly recognised, or does it involve a more sophisticated check for correctness? Investigating the phenomenon of insight is currently a challenge for neuroscience and it is not trivial from a networks perspective.

Finally, the random walk model can be evaluated on other associative creativity tests.

References

[Abbott et al., 2015] Abbott, J. T., Austerweil, J. L., and Griffiths, T. L. (2015). Random walks on semantic networks can resemble optimal foraging. In *Neural Information Processing Systems Conference; A preliminary version of this work was presented at the aforementined conference.*, volume 122, page 558. American Psychological Association.

[Aiello et al., 2012] Aiello, D. A., Jarosz, A. F., Cushen, P. J., and Wiley, J. (2012). Firing the executive: When an analytic approach to problem solving helps and hurts. *The Journal of Problem Solving*, 4(2):7.

[Ansburg, 2000] Ansburg, P. I. (2000). Individual differences in problem solving via insight. *Current Psychology*, 19(2):143–146.

[Arden et al., 2010] Arden, R., Chavez, R. S., Grazioplene, R., and Jung, R. E. (2010). Neuroimaging creativity: a psychometric view. *Behavioural brain research*, 214(2):143–156.

[Barrat et al., 2008] Barrat, A., Barthelemy, M., and Vespignani, A. (2008). *Dynamical processes on complex networks*. Cambridge university press.

[Blondel et al., 2008] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

[Boden, 2004] Boden, M. A. (2004). *The creative mind: Myths and mechanisms*. Routledge.

[Bowden and Jung-Beeman, 2003] Bowden, E. M. and Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior research methods, instruments, & computers*, 35(4):634–639.

[Collins and Loftus, 1975] Collins, A. M. and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, 82(6):407.

[Deyne et al., 2019] Deyne, S. D., Navarro, D. J., Perfors, A., Brysbaert, M., and Storms, G. (2019). Measuring the associative structure of english: The 'small world of words' norms for word association. *Behavior Resarch Methods*, 51(3):987–1006.

[Edler et al.,] Edler, D., Eriksson, A., and Rosvall, M. The mapequation software package.

[Foster et al., 2011] Foster, P. S., Yung, R. C., Branch, K. K., Stringer, K., Ferguson, B. J., Sullivan, W., and Drago, V. (2011). Increased spreading activation in depression. *Brain and Cognition*, 77(2):265–270.

[Gabler,] Gabler, K. Human brain cloud. a massively multiplayer word association game.

[Guilford, 1967] Guilford, J. P. (1967). Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior*, 1(1):3–14.

[Hagberg et al., 2008] Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring network structure, dynamics, and function using networkx. In Varoquaux, G., Vaught, T., and Millman, J., editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA.

[Hélie and Sun, 2010] Hélie, S. and Sun, R. (2010). Incubation, insight, and creative problem solving: a unified theory and a connectionist model. *Psychological review*, 117(3):994.

[Kajić et al., 2017] Kajić, I., Gosmann, J., Stewart, T. C., Wennekers, T., and Eliasmith, C. (2017). A spiking neuron model of word associations for the remote associates test. *Frontiers in psychology*, 8:99.

[Kenett and Faust, 2019] Kenett, Y. N. and Faust, M. (2019). A semantic network cartography of the creative mind. *Trends in cognitive sciences*, 23(4):271–274.

[Kenett et al., 2018] Kenett, Y. N., Levy, O., Kenett, D. Y., Stanley, H. E., Faust, M., and Havlin, S. (2018). Flexibility of thought in high creative individuals represented by percolation analysis. *Proceedings of the National Academy of Sciences*, 115(5):867–872.

[Kleinmintz et al., 2019] Kleinmintz, O. M., Ivancovsky, T., and Shamay-Tsoory, S. G. (2019). The two-fold model of creativity: the neural underpinnings of the generation and evaluation of creative ideas. *Current Opinion in Behavioral Sciences*, 27:131–138.

[Meade et al., 2007] Meade, M. L., Watson, J. M., Balota, D. A., and Roediger III, H. L. (2007). The roles of spreading activation and retrieval mode in producing false recognition in the drm paradigm. *Journal of Memory and Language*, 56(3):305–320.

[Mednick, 1962] Mednick, S. (1962). The associative basis of the creative process. *Psychological review*, 69(3):220.

[Mekern et al., 2019] Mekern, V., Hommel, B., and Sjoerds, Z. (2019). Computational models of creativity: a review of single-process and multi-process recent approaches to demystify creative cognition. *Current Opinion in Behavioral Sciences*, 27:47–54.

[Nelson et al., 2004] Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (2004). The university of south florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36(3):402–407.

[Newman, 2018] Newman, M. (2018). *Networks*. Oxford university press.

[Pulvermüller et al., 2009] Pulvermüller, F., Shtyrov, Y., and Hauk, O. (2009). Understanding in an instant: neurophysiological evidence for mechanistic language circuits in the brain. *Brain and language*, 110(2):81–94.

[Rossetti et al., 2019] Rossetti, G., Milli, L., and Cazabet, R. (2019). Cdlib: a python library to extract, compare and evaluate communities from complex networks. *Applied Network Science*, 4(1):1–26.

[Rosvall et al., 2009] Rosvall, M., Axelsson, D., and Bergstrom, C. T. (2009). The map equation. *The European Physical Journal Special Topics*, 178(1):13–23.

[Saramäki et al., 2007] Saramäki, J., Kivelä, M., Onnela, J.-P., Kaski, K., and Kertesz, J. (2007). Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E*, 75(2):027105.

[Schatz et al., 2022] Schatz, J., Jones, S. J., and Laird, J. E. (2022). Modeling the remote associates test as retrievals from semantic memory. *Cognitive Science*, 46(6):e13145.

[Smith et al., 2013] Smith, K. A., Huber, D. E., and Vul, E. (2013). Multiply-constrained semantic search in the remote associates test. *Cognition*, 128(1):64–75.

[Sowa, 1987] Sowa, J. F. (1987). Semantic networks.

[Valba et al., 2021] Valba, O., Gorsky, A., Nechaev, S., and Tamm, M. (2021). Analysis of english free association network reveals mechanisms of efficient solution of remote association tests. *PloS one*, 16(4):e0248986.

[Wolff et al., 2022] Wolff, A., Berberian, N., Golesorkhi, M., Gomez-Pilar, J., Zilio, F., and Northoff, G. (2022). Intrinsic neural timescales: temporal integration and segregation. *Trends in cognitive sciences*.

5 Appendix

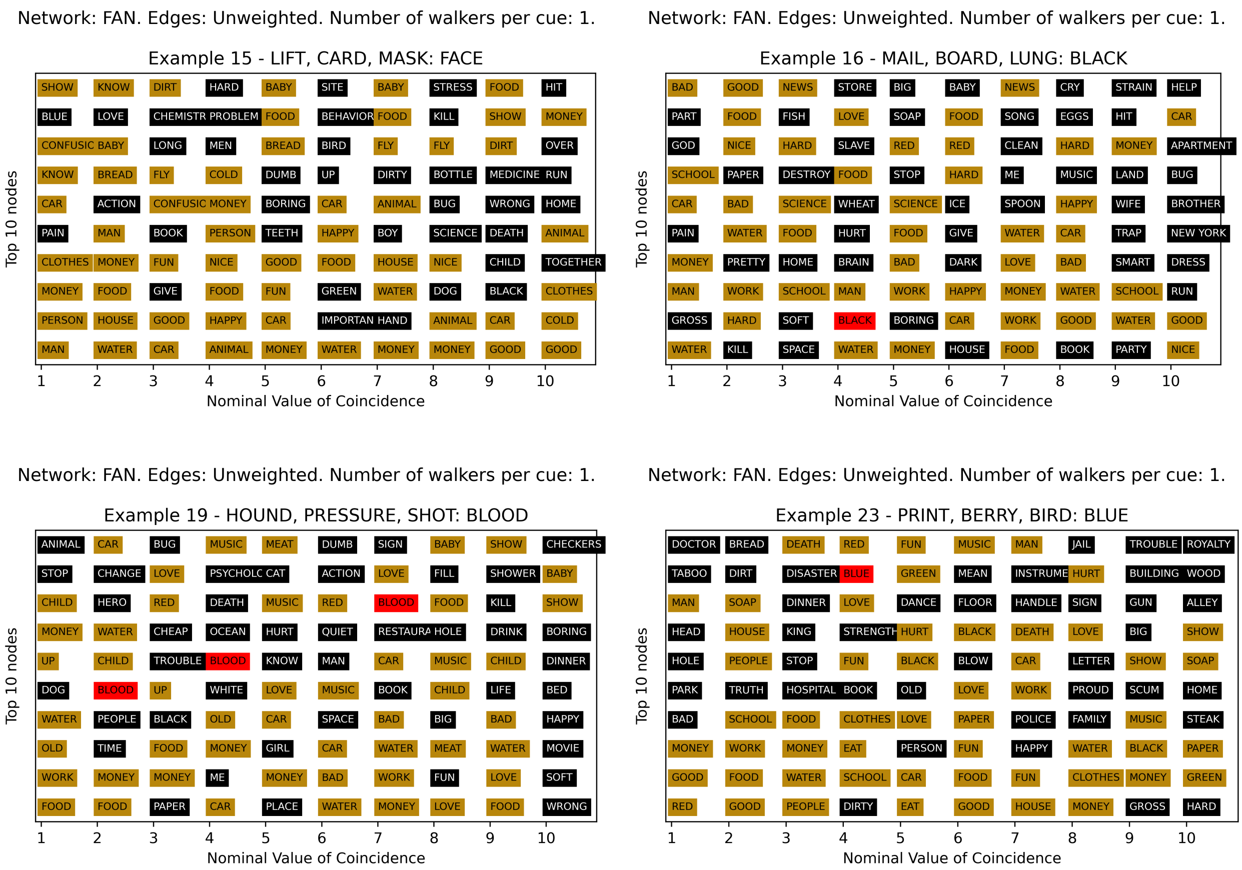


Figure 12: Earliest coincidences according to the ordinal approach for four RAT examples. The words in red are the found solutions. The words, which are repeated in different nominal positions, are in ochra.

Table 3: RAT examples.

	Cue 1	Cue 2	Cue 3	Solution
1	DEMON	LIMIT	WAY	SPEED
2	WIDOW	BITE	MONKEY	SPIDER
3	TIME	HAIR	STRETCH	LONG
4	CHEST	CAR	STORE	TOY
5	LAND	HAND	HOUSE	FARM
6	DATE	ALLEY	FOLD	BLIND
7	BUMP	EGG	STEP	GOOSE
8	HIGH	BOOK	SOUR	NOTE
9	BROKEN	CLEAR	EYE	GLASS
10	COIN	QUICK	SPOON	SILVER
11	STRIKE	SAME	TENNIS	MATCH
12	PINE	CRAB	SAUCE	APPLE
13	PIKE	COAT	SIGNAL	TURN
14	SPEAK	MONEY	STREET	EASY
15	WATER	TOBACCO	STOVE	PIPE
16	LIFT	CARD	MASK	FACE
17	MAIL	BOARD	LUNG	BLACK
18	WAGON	BREAK	RADIO	STATION
19	CHOCOLATE	FORTUNE	TIN	COOKIE
20	HOUND	PRESSURE	SHOT	BLOOD
21	SAFETY	CUSHION	POINT	PIN
22	DUST	CEREAL	FISH	BOWL
23	TOOTH	POTATO	HEART	SWEET
24	PRINT	BERRY	BIRD	BLUE
25	DUE	LIFE	TENSE	PAST

Table 4: Shortest path length and number of shortest paths. Unweighted graph

	cue1	shpl	nshp	cue2	shpl	nshp	cue3	shpl	nshp	solution	totShpl
1	DEMON	1	1	LIMIT	1	1	WAY	2	1	SPEED	4
2	WIDOW	1	1	BITE	2	4	MONKEY	1	1	SPIDER	4
3	TIME	1	1	HAIR	1	1	STRETCH	1	1	LONG	3
4	CHEST	2	1	CAR	1	1	STORE	3	38	TOY	6
5	LAND	1	1	HAND	2	1	HOUSE	1	1	FARM	4
6	DATE	1	1	ALLEY	2	2	FOLD	1	1	BLIND	4
7	BUMP	3	8	EGG	1	1	STEP	3	2	GOOSE	7
8	HIGH	3	20	BOOK	1	1	SOUR	3	1	NOTE	7
9	BROKEN	1	1	CLEAR	1	1	EYE	1	1	GLASS	3
10	COIN	1	1	QUICK	1	1	SPOON	1	1	SILVER	3
11	STRIKE	1	1	SAME	1	1	TENNIS	1	1	MATCH	3
12	PINE	2	1	CRAB	1	1	SAUCE	1	1	APPLE	4
13	PIKE	1	1	COAT	2	1	SIGNAL	1	1	TURN	4
14	SPEAK	3	13	MONEY	1	1	STREET	3	16	EASY	7
15	WATER	1	1	TOBACCO	1	1	STOVE	1	1	PIPE	3
16	LIFT	1	1	CARD	2	1	MASK	1	1	FACE	4
17	MAIL	2	1	BOARD	1	1	LUNG	3	14	BLACK	6
18	WAGON	1	1	BREAK	3	8	RADIO	1	1	STATION	5
19	CHOCOLATE	1	1	FORTUNE	1	1	TIN	1	1	COOKIE	3
20	HOUND	1	1	PRESSURE	1	1	SHOT	1	1	BLOOD	3
21	SAFETY	1	1	CUSHION	1	1	POINT	1	1	PIN	3
22	DUST	3	13	CEREAL	1	1	FISH	1	1	BOWL	5
23	TOOTH	1	1	POTATO	1	1	HEART	2	5	SWEET	4
24	PRINT	2	1	BERRY	2	1	BIRD	1	1	BLUE	5
25	DUE	1	1	LIFE	1	1	TENSE	1	1	PAST	3

Table 5: Shortest path length and number of shortest paths. Weighted graph

	cue1	shpl	nshp	cue2	shpl	nshp	cue3	shpl	nshp	solution	totShpl
1	DEMON	0.01	1	LIMIT	0.02	1	WAY	0.02	2	SPEED	0.054
2	WIDOW	0.01	1	BITE	0.02	1	MONKEY	0.01	1	SPIDER	0.043
3	TIME	0.01	1	HAIR	0.01	1	STRETCH	0.01	1	LONG	0.0395
4	CHEST	0.02	1	CAR	0.01	1	STORE	0.02	1	TOY	0.047
5	LAND	0.01	1	HAND	0.02	1	HOUSE	0.02	3	FARM	0.0485
6	DATE	0.02	2	ALLEY	0.01	1	FOLD	0.01	1	BLIND	0.0375
7	BUMP	0.03	1	EGG	0.03	1	STEP	0.02	1	GOOSE	0.075
8	HIGH	0.02	1	BOOK	0.01	1	SOUR	0.03	1	NOTE	0.0615
9	BROKEN	0.02	1	CLEAR	0.02	1	EYE	0.01	1	GLASS	0.043
10	COIN	0.01	1	QUICK	0.01	1	SPOON	0.02	1	SILVER	0.045
11	STRIKE	0.01	1	SAME	0.02	1	TENNIS	0.01	1	MATCH	0.0385
12	PINE	0.03	1	CRAB	0.02	1	SAUCE	0.01	1	APPLE	0.0535
13	PIKE	0.02	2	COAT	0.03	2	SIGNAL	0.02	1	TURN	0.0635
14	SPEAK	0.02	1	MONEY	0.01	1	STREET	0.02	1	EASY	0.049
15	WATER	0.02	1	TOBACCO	0.02	1	STOVE	0.01	1	PIPE	0.0545
16	LIFT	0.02	1	CARD	0.02	1	MASK	0.02	1	FACE	0.66
17	MAIL	0.02	1	BOARD	0.01	1	LUNG	0.03	1	BLACK	0.055
18	WAGON	0.03	3	BREAK	0.02	1	RADIO	0.02	1	STATION	0.0715
19	CHOCOLATE	0.01	1	FORTUNE	0.02	1	TIN	0.01	1	COOKIE	0.045
20	HOUND	0.01	1	PRESSURE	0.02	1	SHOT	0.01	1	BLOOD	0.045
21	SAFETY	0.03	3	CUSHION	0.03	1	POINT	0.02	1	PIN	0.073
22	DUST	0.02	1	CEREAL	0.03	2	FISH	0.01	1	BOWL	0.063
23	TOOTH	0.01	1	POTATO	0.01	1	HEART	0.02	1	SWEET	0.041
24	PRINT	0.02	1	BERRY	0.03	2	BIRD	0.01	1	BLUE	0.06
25	DUE	0.01	1	LIFE	0.01	1	TENSE	0.02	1	PAST	0.0365