



SpreadPy: A Python tool for modelling spreading activation and superdiffusion in cognitive multiplex networks

Salvatore Citraro ^a^{*,1}, Edith Haim ^{b,1}, Alessandra Carini ^b, Cynthia S.Q. Siew ^c, Giulio Rossetti ^a, Massimo Stella ^b

^a Institute of Information Science and Technologies “A. Faedo” (ISTI), National Research Council (CNR), Pisa, 56127, Italy

^b CogNosco Lab, Department of Psychology and Cognitive Science, University of Trento, Rovereto, 38068, Italy

^c Department of Psychology, National University of Singapore, Singapore, 117570, Singapore

ARTICLE INFO

Keywords:

Spreading activation
Multiplex networks
Computational cognitive modelling
Python simulation tool
Open-source

ABSTRACT

We introduce SpreadPy as a Python library for simulating spreading activation in cognitive single-layer and multiplex networks. Our tool enables systematic testing of structure–function relationships in cognition by comparing simulated activation dynamics outcomes with established theories in knowledge modelling. We demonstrate the library's utility through three case studies: (1) Spreading activation on associative knowledge networks distinguishes students with high versus low math anxiety, revealing anxiety-related structural differences in conceptual organization; (2) Activation trajectories in a creativity task vary with task difficulty, revealing how cognitive load modulates lexical access; (3) In individuals with aphasia, simulated activation patterns on lexical networks correlate with empirical error types (semantic vs. phonological) during picture-naming tasks, linking network structure to clinical impairments. SpreadPy's flexible framework allows researchers to model these processes using empirically derived or theoretical networks, providing mechanistic insights into individual differences and cognitive impairments. The library is openly available, supporting reproducible research in psychology, neuroscience, and education research.

1. Introduction

Spreading activation is a network-based model describing how activation propagates through interconnected entries during information search. The original model, introduced in cognitive science (Anderson, 1983; Collins & Loftus, 1975), considers the search process as the spread of an activation signal propagating on a cognitive or information network. In cognitive networks, nodes represent concepts and/or words while links indicate semantic, phonological, syntactic and other types of conceptual relationships (Siew, 2019). In information networks, nodes could represent websites, posts or documents and they could be linked according to their similarities (Crestani, 1997). In spreading activation, the activation signal originates from one or multiple initial nodes and spreads outward, diminishing with each step according to predefined parameters such as decay or thresholds. The final pattern of activation identifies structurally relevant nodes, related to the origin (Collins & Loftus, 1975).

In cognitive science, spreading activation models human cognitive processes (Anderson, 1983; Collins & Ellis, 1992; Collins & Quillian, 1969), including lexical access and memory retrieval. For instance,

when individuals read or hear a word, activation may be spreading along a networked cognitive repository of knowledge, a system known as mental lexicon and investigated across psychology, linguistics and artificial intelligence (Aitchison, 2012). Spreading activation can explain phenomena such as semantic priming, where exposure to a word (e.g., “doctor”) speeds up the recognition of a related word (e.g., “nurse”) (Aitchison, 2012; McRae et al., 2005). Activation spreading also accounts for lexical retrieval errors, such as tip-of-the-tongue states, where partial activation of a target word leads to the retrieval of semantically or phonologically similar words, thereby hindering the full activation and retrieval of the target word (Burke et al., 1991). The modelling framework of activation spread has profoundly influenced our understanding of semantic memory, lexical access, and associative reasoning (Aitchison, 2012; McClelland & Rumelhart, 1981; Vitevitch, 2008). However, previous implementations of spreading activation have been limited to computational frameworks in single-layer networks and via R only (Siew, 2019).

In computer science, spreading activation algorithms have been mainly applied to model semantic search and information retrieval

* Corresponding author.

E-mail address: salvatore.citraro@isti.cnr.it (S. Citraro).

¹ These authors contributed equally.

tasks. They operate by activating nodes representing search queries and propagating this activation through networks of interconnected documents, concepts, or data entities. Highly activated nodes correspond to documents or entities semantically related to the original query (Crestani, 1997). For example, Crestani (1997) demonstrated how spreading activation enhances search efficiency by retrieving documents contextually similar to user queries based on their semantic relationships. Similar works proposed a spreading activation-based navigation and information-foraging method to support users in exploring large document collections (Pirulli et al., 1996) or the World Wide Web (Crestani & Lee, 2000; Rocha et al., 2004). Spreading activation was even featured in digital library systems to improve retrieval accuracy and contextual relevance among library entries, notes and digital books (Ševcech et al., 2014). Wherever multiple types of nodes or links can exist, such as in the digital library example, spreading activation can be extended beyond single-layer structures. Multiplex networks, comprising multiple interconnected layers of information, provide a powerful framework for modelling spreading activation in these complex relationships (Stella et al., 2024).

To bridge cognitive insights with novel computational applications across multiplex networks, we here present SpreadPy, a Python library designed to implement spreading activation models on both single-layer and multiplex network structures. SpreadPy facilitates the modelling of complex cognitive processes and enhances algorithmic performance in semantic search tasks. This makes it a unified framework that is valuable for both cognitive science research and artificial intelligence applications.

1.1. Spreading activation in knowledge modelling

Despite being introduced at the fringe of cognitive science and cognitive psychology, spreading activation has been extensively adopted and further developed in computer science. The works by Crestani and colleagues (Crestani, 1997; Crestani & Lee, 2000) investigated spreading activation techniques in relation to information retrieval. Crestani (1997) introduced spreading activation as a computational mechanism to enhance semantic relevance and retrieval precision in document databases, highlighting its ability to naturally model associative thinking. Extending this concept, Crestani and Lee (2000) developed a constrained spreading activation algorithm tailored for web searching. They significantly improved query specificity by limiting activation propagation through selective semantic paths, thereby balancing retrieval comprehensiveness and efficiency.

Further developments expanded spreading activation techniques to more structured semantic domains. Rocha et al. (2004) presented a hybrid search approach integrating traditional keyword matching and spreading activation within semantic web frameworks. Their method improved retrieval performance by activating relevant semantic pathways derived from structured metadata, demonstrating the advantage of combining symbolic reasoning with spreading activation for handling complex information structures on the semantic web. Similarly, Ševcech et al. (2014) leveraged spreading activation in combination with user-generated annotations to enhance the contextual relevance of document searches within digital libraries. Consequently, they reinforced the adaptability of spreading activation to leverage both formal semantics and user-driven metadata.

In contrast to structured semantic approaches, Pirulli et al. (1996) applied spreading activation within user-centric interfaces for navigating large text collections. Their “Scatter/Gather” browsing approach used spreading activation to dynamically cluster (gather) and present specific document groups (scatter) based on topical relevance, supporting users’ natural cognitive strategies of incremental exploration and information foraging. This approach differed fundamentally from structured semantic methods by emphasizing user interactions and cognitive relevance directly rather than explicit semantic structures.

Collectively, the above studies illustrate the versatility and efficacy of spreading activation, underscoring its applicability across diverse information retrieval contexts, ranging from structured semantic databases to user-driven interactive environments. While our introduced tool SpreadPy can account for a wide variety of knowledge modelling approaches, we here focus on testing it within the field of cognitive network science, a novel research area adopting network science to investigate cognitive phenomena (Castro & Siew, 2020; Siew et al., 2019; Stella, 2022).

1.2. Spreading activation in cognitive network science

The spreading activation theory implemented in SpreadPy is based on the seminal work of Collins and Loftus (1975), whose spreading activation theory of semantic processing is built directly on Quillian’s earlier model of memory search (Quillian, 1967). Quillian’s theory introduced the idea that concepts are represented as nodes in a network and that activation spreads along associative links between them. Once a concept is stimulated through reading, hearing or mental activation (i.e. thinking about it), its activation spreads to related concepts along these links. Activation weakens with distance and time, and a threshold of cumulated activation must be reached for a concept to become accessible.

In contrast to Quillian’s model (Quillian, 1967), Collins and Loftus’ spreading activation model (Collins & Loftus, 1975) was not just intended as a theoretical explanation but as a computational framework to simulate cognitive search processes (Collins & Loftus, 1975). It shared key goals with early artificial intelligence since it aimed to understand knowledge representation and memory search in humans as well as simulate human-like behaviour in computers. The model has a lasting influence on cognitive psychology, where it provided explanations for semantic priming (Groot, 1983; McRae et al., 2005), lexical decision-making (Balota & Lorch, 1986), and memory retrieval phenomena (Saunders & MacLeod, 2006), among others. It offered a more flexible alternative to previous models of memory and inspired empirical research into associative strength, inhibition mechanisms, reaction times, and the structure of semantic networks (Saunders & MacLeod, 2006; Siew, 2019).

One key advance in this area was the recent work by Siew (2019), which introduced the R package *spreadr* to transform theoretical considerations into computational, psychological simulations. Siew’s work was the first implementation of spreading activation into an R environment, able to explain and reproduce psychological results from real-world data (Siew, 2019). While Siew’s *spreadr* package constitutes a significant advancement in network research, this tool operated only on single-layer networks, leaving a gap in the field for multiplex representations. Such advanced network models are relevant since cognitive processes, especially those involved in language and memory, have been shown to be influenced by multiple types of associations interacting with each other, such as semantic and phonological (Castro & Siew, 2020). For instance, during the lexical decision task, which heavily relies on semantic information, studies have found phonological priming effects that aid in lexical retrieval (Collins & Ellis, 1992; Coltheart et al., 1979).

To capture this complexity, and fully concretize Collins and Loftus’ original idea of featuring more networks linked together (Collins & Loftus, 1975), we use cognitive multiplex networks (Stella et al., 2017, 2018, 2024). These network models can incorporate multiple aspects of the mental lexicon (e.g., semantic associations and phonological similarities) at the same time. Despite their theoretical promise, computational tools for simulating spreading activation within network-based cognitive models remain underdeveloped. SpreadPy extends this approach to capture the multiplex nature of cognitive systems in which concepts are linked through diverse relations (e.g., semantic, phonological, or contextual). Fig. 1A depicts this multiplex structure. Each

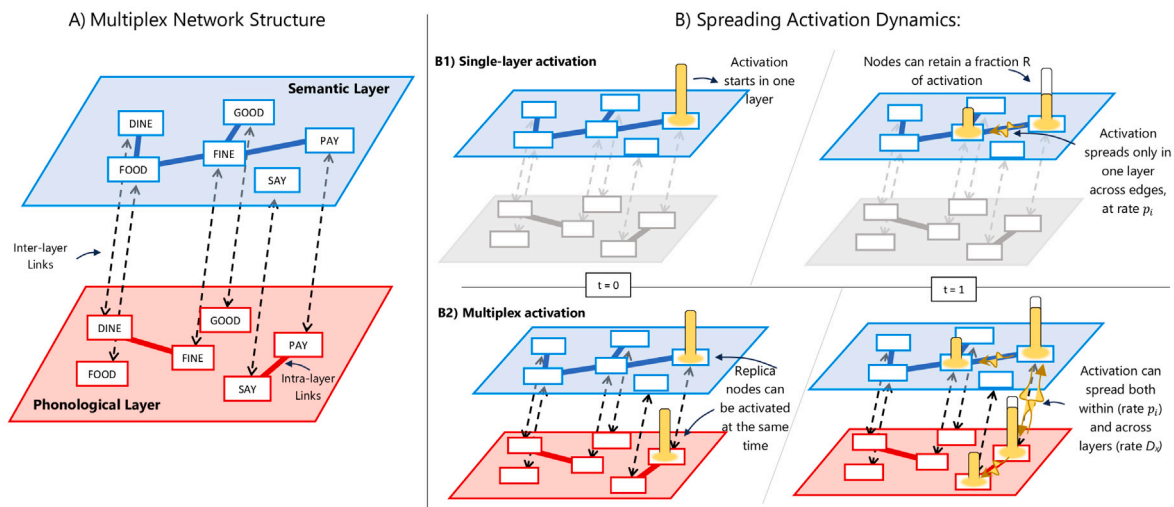


Fig. 1. Infographics for the functioning of SpreadPy over a cognitive network. (A) A cognitive network can be a multiplex network with a semantic and a phonological layer, where nodes are replicated (replica nodes) across layers. In a given layer, replica nodes are connected by intra-layer connections, e.g. phonological similarities or free associations (“good”-“fine” on semantic layer). Replicas of the same node/concept are connected by inter-layer connections (“good” on semantic layer -“good” on phonological layer). (B) As input, SpreadPy uses individual layers or multiplex networks to perform two types of activation spreading: (i) single-layer or (ii) multiplex.

node represents a word, existing as replicas simultaneously in both semantic and phonological layers, connected by inter-layer links. Within each layer, nodes are connected based on layer-specific similarities: meaning-related associations in the semantic layer, and phonological similarities in the phonological layer (see Fig. 1A). This multiplex design allows activation to spread not only within a single type of relationship (within only one layer), but also across different association types (across multiple layers), reflecting the interactive, complex nature of cognitive processes (Castro & Siew, 2020). Importantly, access to simulations based on cognitive multiplex networks also provides access to exploring the phenomenon of superdiffusion (Gomez et al., 2013) within cognitive science. Superdiffusion can be explained in two ways, one appealing to network scientists and another more appealing to psychologists and cognitive scientists. From a network science perspective, superdiffusion occurs when activation spreads faster across a multiplex network than it would on even the fastest individual layer alone. In this way, superdiffusion is a phenomenon accelerating the spread of activation even beyond the fastest limit posed by the network layer fastest at spreading the signal (Gomez et al., 2013). In cognitive science terms, superdiffusion in multiplex networks can be likened to how the mental lexicon draws on multiple, interconnected representational layers, e.g. phonological, and semantic (Baker et al., 2023), when retrieving a word. If each layer were accessed independently, retrieval speed would be constrained by the “slowest” pathway (e.g., phonological bottlenecks in tip-of-the-tongue states, Castro & Siew, 2020; Vitevitch & Aljasser, 2021). But because these layers can be coupled in the mental lexicon, activation can spread in parallel and converge, leading to faster-than-expected retrieval, much like multiplex superdiffusion accelerates diffusion relative to the slowest single network. This resonates with interplay models of lexical access where word retrieval arises from interactive activation across semantic and phonological levels (Dell, 1986; Dell & O’Seaghdha, 1991) and with models like the spreading-activation framework (Collins & Loftus, 1975), which show that coupling between layers enables more efficient navigation of the lexicon than isolated access would allow. Think of it like having multiple ways to recognize a word: through its meaning, its sound, or its visual form; sometimes using these recognition systems together allows an individual to access a word faster than using just the fastest single pathway alone.

1.3. Spreading activation with spreadpy

Our SpreadPy approach can model both single layer activation (Fig. 1 B1) and multiplex activation (Fig. 1 B2). In single-layer activation (Fig. 1 B1), a node spreads the activation it has received only within one layer. For example, receiving a payment notification on the smartphone may activate the concept “pay” on the semantic layer. The activated node “pay” retains a fraction of its original activation (e.g. 50 percent of activation are retained) and spreads the remaining activation within the semantic layer to the neighbouring concept “fine”. At timestep $t = 1$, both nodes now hold 50 percent of the initial activation. From there, activation will continue to spread within the semantic layer to adjacent nodes. In contrast, with a multiplex model (Fig. 1 B2), we can model all replicas of a node on different layers to receive activation at the same time. For instance, while receiving a payment notification on the phone, one may read the word “pay” in the pop-up, activating the replica nodes of “pay” both on the semantic and phonological layers simultaneously. From both replicas of “pay”, activation starts spreading along intra-layer links to the adjacent nodes “fine” on the semantic layer and “say” on the phonological layer. Additionally, activation can now also spread across layers, such that “fine” on the semantic layer also activates its replica on the phonological layer along an inter-layer link.

In our simulations, activation is seeded at cue words and spreads through the network over discrete time steps. We measure two key metrics: the maximum activation level a node accumulates and the time it takes to reach this peak. Fig. 2A shows an example of how we trace the activation level of the target word “hope” over time, peaking at time t_m with its maximum activation α_m .

Fig. 2B further illustrates diffusive processes and the differences between normal and superdiffusion regimes in our simulations. In complex networks, diffusive processes are relative to the spread of information, signals, social norms, diseases and other attributes across nodes in a network (Haim & Haim, 2023; Newman, 2018). When activation spreads along linked nodes, network topology drives the diffusion dynamics, i.e. how quickly a signal spreads across nodes in the network. Thus, diffusion dynamics can be modelled in terms of adjacency/links between nodes, encapsulated within a binary adjacency matrix A , and in terms of network degree of individual nodes, encapsulated within a diagonal matrix D with degrees on its main diagonal (Haim & Haim, 2023; Newman, 2018). Node degrees can indicate how much signal in

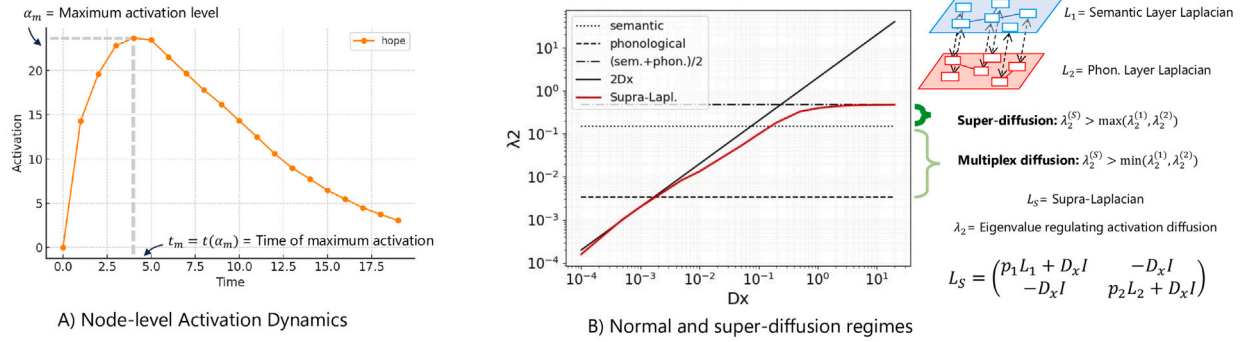


Fig. 2. (A) Example of activation for “hope”. At the node level, activation arrives as a function of time, peaking at time t_m (maximum activation time) and reaching at most activation level of α_m . (B) A multiplex structure can accelerate diffusion – activation on the multiplex network spreads faster than on its slowest layer. It can also cause a superdiffusion process – activation on the multiplex network spreads faster than on its fastest layer.

total can accumulate over nodes, while individual links can indicate potential incoming or escaping routes for diffusive signals. One can encapsulate both these elements in a matrix called Laplacian, $L = D - A$, which can be normalized in several ways. The eigenspectrum of the Laplacian L contains insights about the diffusion dynamics. In particular, the second smallest eigenvalue λ_2 encodes information about the characteristic diffusion time τ , i.e. the time it takes for diffusion to spread along a specific region of a complex network. Through series expansion, one can check that in simple diffusion processes on complex networks, $\lambda_2 \propto 1/\tau$, i.e. the larger the λ_2 eigenvalue of the Laplacian, the shorter the time required for the diffusive signal to reach a whole region of the network. In other words, higher λ_2 means faster diffusion (Haim & Haim, 2023; Koponen, 2021). In multiplex networks, one can build a supra-Laplacian matrix, see also Fig. 2B, and compare the eigenvalue $\lambda_2^{(s)}$ of the whole multiplex network against the eigenvalues $\lambda_2^{(1)}$ and $\lambda_2^{(2)}$ of the individual layers.

1.4. Superdiffusion and its relevance for knowledge modelling

Multiplex diffusion occurs when the diffusion on the whole multiplex network can efficiently exploit inter-layer jumps to spread at a rate faster than on the slowest layer. Inter-layer jumps can occur with a parameter D_x , whose magnitude regulates the rate of diffusion among replicas of the same physical node compared to the rates, p_1 and p_2 , of diffusion between different replica nodes in the same layer, either layer 1 or layer 2. In other words, multiplex diffusion occurs when $\lambda_2^{(s)}$ is greater than the smallest respective eigenvalue for the Laplacians of the individual layers. In multiplex diffusion, multiplexity is beneficial because it provides shortcuts facilitating the spread of signals even among nodes being remote in individual layers. The combination of two layers together thus speeds up diffusion compared to what would happen in the slowest layer (Gomez et al., 2013; Koponen, 2021). However, in this regime, diffusion over the whole multiplex remains slower than in the fastest layer. In formulas, $\lambda_2^{(s)}$ (highlighted in red in Fig. 2B) would be higher than the $\min(\lambda_2^{(1)}, \lambda_2^{(2)})$ but lower than the $\max(\lambda_2^{(1)}, \lambda_2^{(2)})$.

Superdiffusion occurs when the two layers are so strongly entwined that diffusion on the multiplex network is faster even than diffusion in the faster layer, if considered in isolation. In formulas, $\lambda_2^{(s)}$ would be higher than $\max(\lambda_2^{(1)}, \lambda_2^{(2)})$ but still lower than the upper bound given by the superposition of Laplacians (see Fig. 2B). This superposition posits an upper theoretical bound to the speed of diffusion, which is the fastest when all edges in the multiplex networks are fully available at the same time, i.e. when the coupling of layer $D_x \rightarrow +\infty$ (Koponen, 2021).

The phenomenon of superdiffusion offers a novel lens for understanding cognitive processes. In cognitive systems, representations are often distributed across multiple modalities (e.g. semantic, phonological, and contextual layers), each contributing uniquely to information

processing. Superdiffusion arises when interactions between these layers create synergistic effects, enabling rapid and coordinated activation across the network. This contrasts with traditional single-layer models, where activation is confined to a single modality, potentially underestimating the speed and flexibility of cognitive operations (Haim & Haim, 2023; Koponen, 2021).

When modelling mental search, superdiffusion acts in terms of mixing the topological information (e.g., cycles, shortest paths, neighbourhoods) of the considered network layers. Superdiffusion would thus be relative to a highly interactive set of cognitive processes taking place on all considered network layers, e.g., phonological associations and semantic associations both contributing to producing mixed errors during picture naming by people affected by aphasia (Dell et al., 1997). By modelling single-layer diffusion, multiplex and superdiffusion, SpreadPy enables researchers to quantify the impact of both network structure and spreading dynamics over a wide variety of phenomena. We further explore and test this in our case studies.

1.5. Manuscript aims and outline

This work introduces SpreadPy, a Python library that simulates spreading activation in single-layer and multiplex cognitive networks, offering researchers a flexible platform to test hypotheses about structure–function relationships in the mental lexicon. By integrating principles from cognitive psychology and network science, SpreadPy enables the exploration of how cognitive processes are modulated by network topology, such as clustering, centrality, and layer coupling. We choose the Python implementation to position ourselves among the broad landscape of novel network modelling toolkits in cognitive science, such as SNAFU (Zemla et al., 2020) and forager (Kumar et al., 2024), which support memory search and semantic network construction from fluency tasks, and EmoAtlas (Semeraro et al., 2025), for constructing emotional forma mentis networks.

We test and demonstrate the utility of SpreadPy through three case studies: (1) examining the impact of math anxiety on associative knowledge networks, (2) simulating activation dynamics to differentiate item difficulty in a creativity task, and (3) modelling lexical retrieval errors in aphasia. Particularly, we demonstrate that SpreadPy on multiplex networks is essential for creativity and aphasia case studies. In creativity tasks, phonological cues can hop into the semantic layer under superdiffusion, improving item difficulty discrimination. In aphasia, formal errors’ activations showed weak correlations with empirical error frequencies on phonological layers alone but stronger correlations in the multiplex network, revealing that phonological mistakes actually emerge from semantic–phonological interference rather than purely phonological processes. Both phenomena are invisible to single-layer models. These applications highlight how SpreadPy can uncover mechanistic insights into cognitive phenomena and clinical conditions, advancing both theoretical and applied research in cognitive data science.

Table 1

Summary of network data used in the three case studies, based on behavioural forma mentis networks (FMN) reconstructed from human and GPT participants (Math Anxiety), and on the Largest Viable Cluster (LVC) of a semantic–phonological multiplex representation of the mental lexicon (Creativity and Aphasia).

Case Study	Network Type	Dataset	Description
Math Anxiety – ITA	Semantic (FMN)	$N=72 + N_{GPT}=300$	{math., stat.} → network
Creativity – ENG	Multiplex (LVC)	SWoW + IPA	$ V = 4118, E = 26229$
Aphasia – ENG	Multiplex (LVC)	SWoW + IPA	$ V = 4118, E = 26229$

2. Materials and methods

This Section outlines the implementation of diffusion and superdiffusion regimes in SpreadPy, and the common network data used for simulations in this manuscript across the three case studies.

Single-layer Diffusion. The base model operates on a simple, undirected graph, $G = (V, E)$, where V is the set of nodes, and $E \subseteq V \times V$ denotes the set of edges as unordered pairs of nodes. Each node holds an amount of activation energy that can be transferred to its neighbours over time. At each time step t , an active node $u \in V$ with non-zero energy $e_{u,t}$ retains a fraction R of its energy and distributes the remaining portion uniformly among its neighbours. The energy transfer from node u to a neighbour $v \in \Gamma_u$, being Γ_u the set of u 's neighbours, is given by the following equation:

$$\varphi(u, v) = \frac{e_{u,t} \cdot (1 - R)}{|\Gamma_u|}, \quad (1)$$

where $|\Gamma_u|$ is the number of u 's neighbours, and R is the retention factor controlling the energy kept within u .

Multiplex Diffusion. The network is extended to a multiplex graph, $\mathcal{M} = (V, E, L)$, where L is the set of layers. Each node $u \in V$ can exist as a replica $u^{[\ell]}$ in one or more layers, indexed by $\ell \in L$. Edges can be within the same layer (intra-) or between replica nodes across layers (inter-). A coupling parameter $D_x > 0$ balances intra- and inter-layer diffusion, with probabilities $p_{\parallel} = \frac{1}{1+D_x}$ and $p_{\perp} = \frac{D_x}{1+D_x}$, respectively.

Intra-layer spreading mirrors the single-layer case, scaled by $p_{\parallel}$. The energy transfer from node $u^{[\ell]}$ to a neighbour $v^{[\ell]} \in \Gamma_u^{[\ell]}$, being $\Gamma_u^{[\ell]}$ the set of u 's neighbours on layer ℓ , is given by the equation:

$$\varphi(u^{[\ell]}, v^{[\ell]}) = \frac{e_{u^{[\ell]},t} \cdot (1 - R)}{|\Gamma_u^{[\ell]}|} \cdot p_{\parallel}, \quad (2)$$

where $|\Gamma_u^{[\ell]}|$ is the number of u 's neighbours on layer ℓ .

Inter-layer spreading allows a node to transfer activation energy to its replica in another layer, after which the energy continues to spread among the neighbours in the new layer. This process can be described by dividing it into two parts: replica energy transfer and neighbours inter-layer spreading. Energy transfer from node $u^{[\ell]}$ to its replica $u^{[\ell']}$, with $\ell' \neq \ell$, is modelled as follows:

$$\varphi(u^{[\ell]}, u^{[\ell']}) = e_{u^{[\ell]},t} \cdot (1 - R) \cdot p_{\perp}. \quad (3)$$

Neighbours inter-layer spreading describes activation passing from a node on one layer to its replica on another layer, and then continues spreading to that replica's neighbour within the second layer. Such neighbours inter-layer spreading occurs in ℓ' , where the replica energy is transferred among u 's neighbours in $\Gamma_u^{[\ell']}$ as follows:

$$\varphi(u^{[\ell]}, v^{[\ell']}) = \frac{\varphi(u^{[\ell]}, u^{[\ell']})}{|\Gamma_u^{[\ell']}|}, \quad (4)$$

where $|\Gamma_u^{[\ell']}|$ is the number of neighbours of u on layer ℓ' .

This formulation ensures that the total transferred energy (after retention) is divided between intra- and inter-layer interactions according to the parameter D_x , represented as a probability, which governs the strength of coupling across layers. The parameter D_x acts as a superdiffusion parameter: higher values enhance cross-layer coupling and accelerate diffusion. As shown in Gomez et al. (2013), by analysing

the second smallest eigenvalue λ_2 of the supra-Laplacian matrix, one can assess whether the coupling strength D_x facilitates faster diffusion across the multiplex network compared to individual layers. This enables analytical tuning of D_x to describe superdiffusive dynamics.

Network Representations. The data used to represent the mental lexicon as a cognitive multiplex network are shared between Case Study 2 and Case Study 3. Instead, in Case Study 1, we use a single-layer, semantic network, following the approach described in Ciringione et al. (2024). A summary of network data is highlighted in Table 1.

Network Data for Case Study 1. We use behavioural forma mentis networks (Stella et al., 2019) to reconstruct the network of memory recalls produced by students with high and low math anxiety, as captured by a psychometric approach (Math Anxiety Scale) (Ciringione et al., 2024). Behavioural forma mentis networks are obtained from valence norms and free association data collected from 72 Italian undergraduate psychology students and from 300 GPT-simulated students. For these simulations, OpenAI's GPT 3.5 was used, as in the past work by Ciringione and colleagues (Ciringione et al., 2024). To build the free associations in forma mentis networks, the authors (Ciringione et al., 2024) designed a task consisting of two parts: a free association game, and a valence evaluation. The task is roughly structured as the following instruction: *Write a list of 3 words that come to your mind when you think of CUE_WORD and assign a valence score to each word as either "positive", "negative", or "neutral".* Cue words are STEM-related, e.g., they include concepts such as *mathematics*. A network G is then constructed using the word lists provided by each participant, connecting with an edge the cue word to each of the provided responses. A statistical analysis of word valence is performed to assign a unique label to each word — either "positive", "negative", or "neutral" (Ciringione et al., 2024; Stella et al., 2019). These labels are represented as colours in the mindset stream plots, cf. Brian and Stella (2023) and Fig. 4.

Network Data for Case Studies 2 and 3. We represent the mental lexicon as a multiplex network $\mathcal{M}$ (Stella et al., 2024), where concepts are nodes connected via two layers, one for representing semantic associations and one for phonological similarities. The semantic layer is proxied by free associations — i.e., memory recall patterns of pairs of words — based on over 1 million responses from the Small World of Words (SWoW) dataset (Deyne et al., 2019), specifically, the SWoW-EN18 release.² The phonological layer connects words based on phonological similarities, which are encoded here, as in past works (Stella et al., 2017, 2018; Vitevitch, 2008), by connecting any two words whose IPA (International Phonetic Alphabet) transcriptions differ by the addition, substitution or deletion of one phoneme only. By considering them in word forms (rather than IPA transcriptions), one can have the same set of nodes replicated across layers (see Fig. 1A). This alignment makes it possible to consider a multiplex network representation of the mental lexicon that was already investigated in past approaches (Stella, 2020). Other approaches might rather represent word forms on a semantic layer and phonological IPA transcriptions on another layer, however this would lead to multilayer networks different from the one considered here. For the difference between multilayer and multiplex networks we refer the interested reader to Stella et al. (2024). To constrain simulations along nodes being connected on both the phonological and the semantic layers, we considered only the Largest Viable

² <https://smallworldofwords.org/it/project/research>

Cluster (LVC) of this multiplex representation, including $|V| = 4118$ nodes and $|E| = 26229$ edges. The LVC is the subgraph composed of all nodes for which exists a path in the multiplex network that does not exist in the layers separately, following (Stella, 2020; Stella et al., 2018). Algorithmically, the LVC can be computed through a recursive procedure that first identifies the largest connected component within each layer, then finds their intersection to obtain common nodes. The algorithm constructs a subgraph from these common nodes, extracts its largest connected component, and iteratively refines this process until the set of common nodes converges, ensuring the final LVC represents the maximal subset of nodes that are simultaneously connected across all specified layers. On this 2-layer representation we explore the inter-layer diffusion parameter D_x , which regulates superdiffusion, and it is shown in Fig. 2B.

3. Results

We test SpreadPy in 3 case studies, modelling (1) math anxiety patterns in human- and LLM-based behavioural data; (2) item-level difficulty in a creativity task; and (3) picture naming data from individuals with aphasia in psycholinguistic experiments. We treat each case study as independent from the others, with its own aims and scope, results and discussions.

3.1. Case study 1: SpreadPy and psychometric data from math anxiety

Main goal. This section investigates activation patterns related to the concepts of “mathematics” and “statistics”, and their connections with psychological variables such as anxiety, stress, and depression. Our model was used to simulate lexical activation across single-layer networks derived from different participant groups. This allowed us to identify and study group-based differences in the organization of concepts in these networks. Specifically, we compared students with high math anxiety (HMA) and low math anxiety (LMA) to examine how varying levels of math anxiety influence both the structure and the activation dynamics of conceptual networks. Particular attention was paid to how mathematical and statistical concepts connect with affective states such as anxiety, stress, and depression.

Cognitive networks and math anxiety. Math anxiety is a multifaceted phenomenon that impairs cognitive functioning and emotional well-being, often manifesting as worry, tension, and avoidance in response to math-related situations (Ashcraft, 2002; Hembree, 1990). Cognitive network science offers powerful tools to investigate how such anxiety is structured and sustained within the mental lexicon (Stella, 2022). Networks built from free association and valence attribution tasks can model this associative knowledge and capture how individuals mentally represent concepts related to science or mathematics and their affective tone (Stella, 2022; Stella et al., 2019). In individuals with high math anxiety, such networks tend to show more fragmented associative structures and stronger links between math-related concepts negative emotions such as stress or fear (Stella, 2022). These associative patterns not only reflect memory recall but may also reinforce cognitive biases like avoidance behaviour, which contributes to the persistence of anxiety (Gangemi et al., 2012; Hayes et al., 1996). By modelling these associations as networks, researchers can analyse how affective content shapes knowledge structure and how anxiety constrains the cognitive complexity around the field of mathematics (Stella, 2022).

Data and modelling setup. We simulated how the activation of the trigger node “mathematics” propagates through cognitive networks of interconnected concepts, affecting nodes associated with emotional states, such as stress, anxiety, and depression. To examine whether math anxiety alters associative pathways in cognitive networks, we simulated spreading activation from the node “mathematics” across group-based single-layer semantic networks. These networks were built from Italian free association data (see Section 2) collected from 72 Italian undergraduate psychology students, who were classified as having

high (HMA) or low math anxiety (LMA) based on the Mathematics Anxiety Scale. Participants responded to 40 cue words related to STEM, mental health, academic evaluation, and personal attitudes. Additional networks were generated by GPT-3.5 simulating students with high (GPT-HMA) or low math anxiety (GPT-LMA). This choice is justified by past works (Abramski et al., 2023) showing that GPT can capture the associative structures underlying math anxiety. Here, we want to observe that these models can be used as computational proxies to compare with human data and test whether affective-cognitive patterns generalize beyond experiential samples. For each of these four datasets (human HMA, human LMA, GPT-HMA, GPT-LMA) a separate single-layer network was built.

SpreadPy was used to simulate spreading activation in these single-layer networks across different retention rates ($R = 0.2, 0.4, 0.6, 0.8$) to analyse possible differences in the propagation patterns; decay and suppress parameters were set to 0. We then observed the spreading of activation through each network independently from one another and up to 400 time steps. We tracked the maximum activation level α_m and the time step of maximum activation t_m , representing the peak in activation following each triggered node. Furthermore, metrics such as node degree and shortest path lengths were collected as key elements of the structural configuration of the network. The networks were then visualized through mindset streams, as described by Brian and Stella (2023), which allow to highlight connections between two words in conceptual maps. These subgraphs are created from all the shortest paths between the concepts and allow associating linked nodes in a network to study their affective patterns. The valence of the connections within the stream is determined by the colour of the links: blue for positive valence, red for negative valence, and grey for neutral concepts (Brian & Stella, 2023).

Stronger and earlier activation of “anxiety” in high-math-anxiety students. Fig. 3 illustrates the different activation levels for the nodes “anxiety” and “stress” across networks at various retention rates (0.2, 0.4, 0.6 and 0.8). When activation is injected into the node “mathematics” and allowed to diffuse, the HMA network drives the target word “anxiety” to its peak within the first ten time steps. The amplitude is $\approx 15\text{--}25\%$ higher than in the low-math-anxiety (LMA) network across all four retention regimes ($R = 0.2 \rightarrow 0.8$). The LMA curve rises more slowly and plateaus at a markedly lower level, indicating a weaker semantic-affective coupling between mathematics and anxiety in non-anxious students. The network itself displays an anxiety bias in HMA learners through the prevalence of negative associations given to “mathematics”. Lower retention ($R = 0.2$) produces the steepest overshoot — activation in the high-math anxiety network peaks above 35 units (see Fig. 3). As retention increases to $R = 0.8$, the curves flatten and converge on a common asymptote, yet the vertical gap between HMA and LMA remains around 5–6 units. This robustness implies that the difference is not a fragile artifact but a structural property of the students’ associative knowledge structure. Replacing the probe with “stress” (Fig. 3 bottom four panels) yields peak amplitudes two orders of magnitude smaller and a much slower approach to equilibrium. Crucially, the HMA-LMA separation persists but contracts to a few tenths of an energy unit. The lexicon therefore encodes “mathematics” as explicitly *anxiogenic* rather than generically negative: only the “anxiety” concept is strongly and quickly co-activated, while “stress” shows weak, delayed coupling.

GPT-simulated students replicate the human activation hierarchy but with dampened intensity. The GPT-simulated forma mentis networks successfully reproduce the overall rank ordering observed in human data: when activation spreads from the cue word “mathematics”, the target concepts “anxiety” and “stress”, respectively, reach their highest activation levels in the high math anxiety (HMA) group, followed by the low math anxiety (LMA) group ($HMA > LMA$). This same pattern holds for GPT-simulated participants: GPT-HMA consistently shows higher activation than GPT-LMA. However, the overall intensity of activation is markedly reduced in GPT-generated

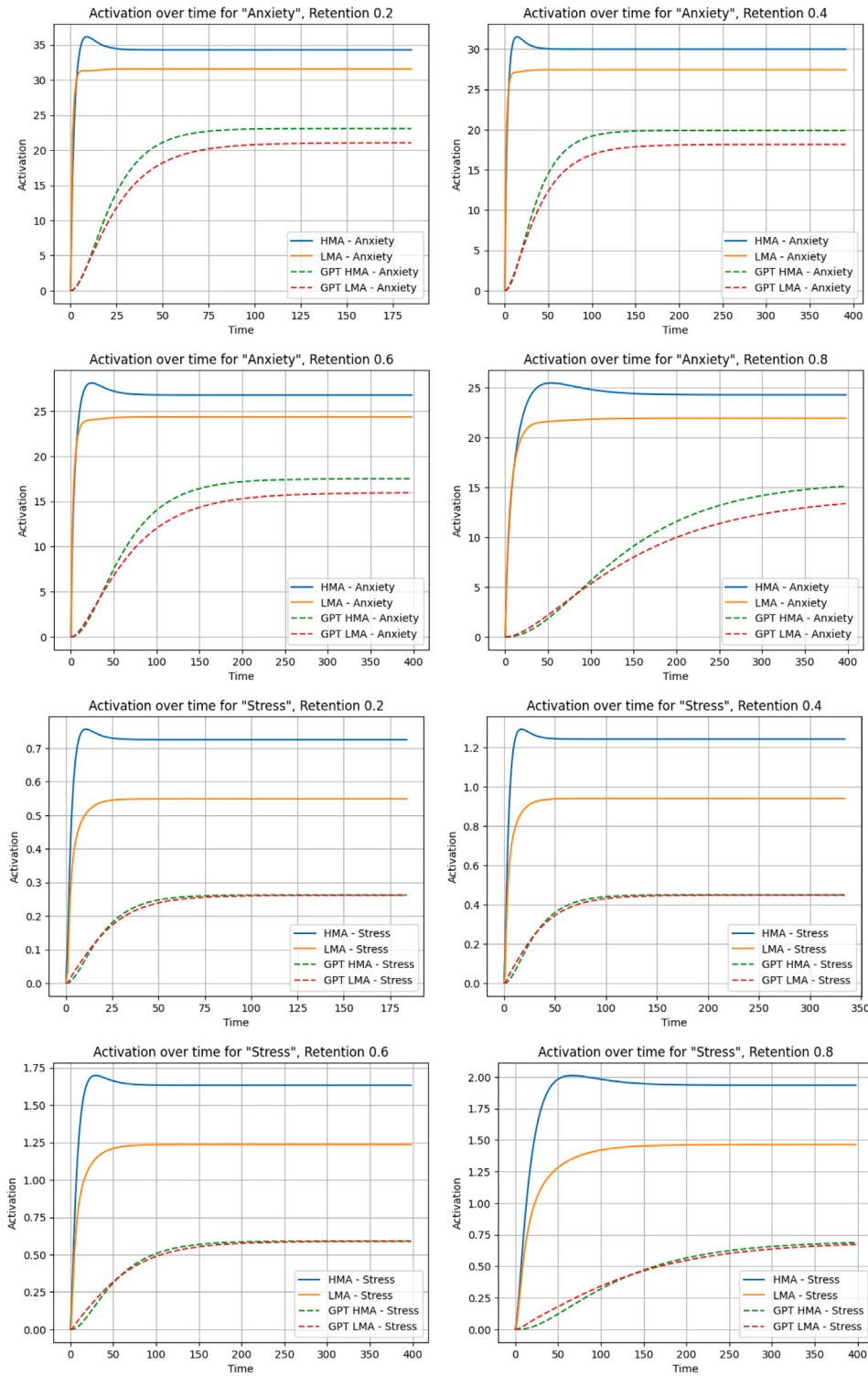


Fig. 3. Overall activation in HMA and LMA students and GPT-HMA and GPT-LMA students for the words “anxiety” (top 4 panels) and “stress” (bottom 4 panels) after the initial activation of the cue “mathematics” at different retention rates.

networks compared to human ones, and the diffusion unfolds more slowly over time. As shown in Fig. 3, the GPT curves sit below the human counterparts and exhibit delayed and flatter growth, taking longer to reach peak activation. In some conditions, the difference between GPT-HMA and GPT-LMA becomes almost negligible (see Fig. 3, bottom panels). This reduced contrast indicates that the models capture the directionality of affective links but underestimate their

strength, suggesting that purely text-based learning misses experiential reinforcement present in real students.

Mindset streams reveal different pathways from “mathematics” to “anxiety/stress/depression” in HMA and LMA networks. Mindset streams are a collection of all shortest paths that join two concepts inside a forma mentis network (Brian & Stella, 2023). They provide a graphical thought-traces, exposing both the semantic bridges and the emotional tone (valence) that link those ideas, revealing how

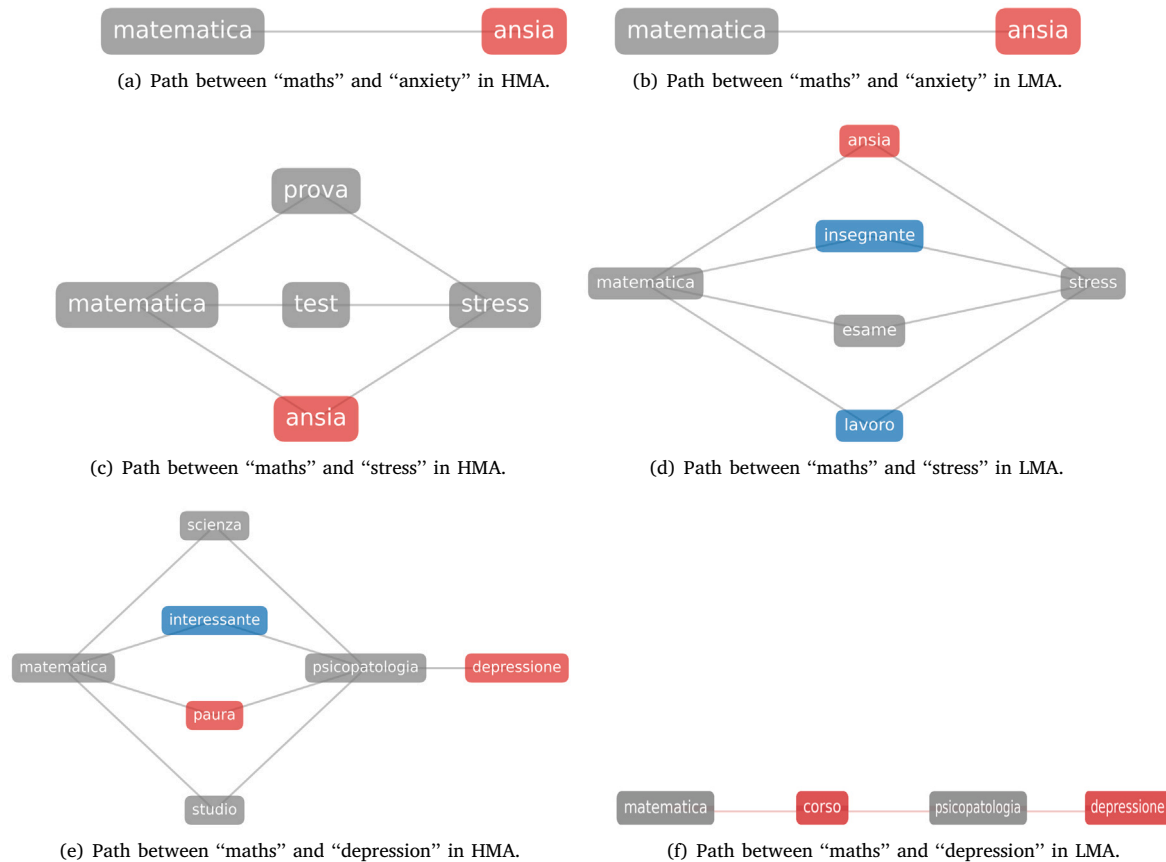


Fig. 4. Mindset streams between “mathematics” (matematica) and “anxiety” (ansia), “stress”, and “depression” (depressione) in HMA and LMA networks. The concepts are represented by Italian words because the networks were built from Italian data (see main text).

directly concepts are associated and which intermediary ideas bridge them (Brian & Stella, 2023).

Fig. 4(a) shows a short mindset stream linking “mathematics” and “anxiety” through a single direct connection, indicating a strong and immediate association between the two concepts. In contrast, Fig. 4(e) illustrates a longer stream – three links connecting “mathematics” to “depression” – suggesting a more distant relationship that relies on intermediate concepts.

An interesting pattern emerges in GPT-LMA and GPT-HMA networks, where mindset streams between “mathematics” and “anxiety” are longer and involve more intricate intermediary links in the high-anxiety (GPT-HMA) condition. This complexity makes it difficult to predict spreading activation from mindset streams alone. The phenomenon likely reflects GPT’s lexical-semantic knowledge, which privilege semantic associations over contextual ones. Mindsets streams for GPT networks are provided in Supplementary Material.

Applied to the human networks of highly anxious and lowly anxious students, the streams show that “mathematics → anxiety” is a one-step direct connection in both groups. This indicates that the subject of mathematics alone already evokes apprehension in students. In contrast, the paths from “mathematics” to “stress” diverge: In HMA students the path runs through “test/attempt → anxiety”, pinpointing test anxiety as the main channel. In LMA students it detours via “job → exam → teacher”, mixing neutral and positive cues and signalling a milder valence. Paths to “depression” are longest: in HMA they wind through academically flavoured but largely neutral terms (“science → study/interesting/fear/study → psychopathology”), and in LMA they are even more indirect, making depression cognitively distant from mathematics. Together, these condensed insights indicate a tight, direct coupling between mathematics and anxiety, whereas links to “stress”

and “depression” are weaker and context-dependent, with test-related notions amplifying distress in the high-anxiety group.

Discussion of results. The simulations reveal a strong and immediate semantic link between “mathematics” and “anxiety” in the networks of high-math-anxiety (HMA) students. This suggests that interventions should go beyond general reduction of stress-related vocabulary and instead focus on weakening this direct associative pathway. For example, this could be achieved by embedding more positive or neutral concepts between “mathematics” and “anxiety” in the students’ mental lexicon. The spreading activation model not only captures these affective trajectories in detail, but also offers a measurable framework to evaluate whether psychological or pedagogical interventions, such as mindset training or curricular changes, have effectively altered the structure and emotional tone of students’ conceptual networks related to “mathematics” and “anxiety”.

Summary and implications. The math anxiety simulations showed that high-anxiety networks feature direct links between “mathematics” and “anxiety”, whereas lower-anxiety networks spread activation more diffusely. This suggests that network structure shapes how negative affect is sustained. Finally, it opens possibilities for testing interventions with our SpreadPy package, such as embedding neutral or positive associations, to weaken maladaptive pathways.

3.2. Case study 2: Modelling creativity data with SpreadPy

Main goal. This case study assesses whether the SpreadPy library can predict item-level difficulty in a classic creativity task, the Remote Associates Test (RAT) (Mednick, 1962), by simulating activation dynamics on a multiplex network. We investigated whether different patterns of activation spreading across semantic and phonological

layers could explain why some RAT items are easier to solve than others. Specifically, we asked whether the two activation-based outcomes maximum activation (α_m) and time of maximum activation (t_m) could help differentiate between easy, medium and hard difficult RAT items. We further tested how this relationship is modulated by key parameters of the SpreadPy model, including retention rate, diffusion regime (standard vs. superdiffusion) and initial activation layer (semantic, phonological or both).

Creativity and multiplex networks. Creativity research has increasingly adopted tools from network science to model how people generate novel ideas by connecting remote concepts (Beatty et al., 2023; Kenett & Faust, 2019). The Remote Associates Test (RAT) was developed by Mednick (1962) and is widely used to assess convergent creativity. Solving the RAT evaluates the ability to identify a word that links three seemingly unrelated cues (e.g. cottage - Swiss - cake $\rightarrow$ cheese). During the test, participants must often inhibit misleading initial associations and rely on more distant, non-obvious connections. This makes the RAT particularly sensitive to the structure of the mental lexicon (Bowden & Jung-Beeman, 2003; Bowers et al., 1990). Especially difficult RAT items typically require longer and more indirect associative chains than easy items, retrieving a common target through less obvious associative pathways. This has made the RAT a valuable tool for studying insight problem solving, associative thinking, and creative cognition (Bowden & Jung-Beeman, 2003; Marko et al., 2019; Wu et al., 2020).

While spreading activation has been proposed as a relevant cognitive mechanism in creative performance, few models have implemented this framework over multiplex networks. Prior studies have used single-layer networks to simulate activation spreading through the semantic network (Siew, 2019). In contrast, SpreadPy allows us to simulate time-sensitive activation diffusion over a multiplex network. To our knowledge, this is the first simulation study to investigate how activation dynamics on a semantic-phonological multiplex network relate to difficulty levels of RAT items.

Data and modelling setup. We selected a subset of established RAT items for which difficulty ratings were available (Bowden & Jung-Beeman, 2003; Bowers et al., 1990; Lee et al., 2014; Wu et al., 2020). To ensure compatibility with our multiplex network, we retained only items for which all four relevant words (three cues and one target solution) were present in both network layers. This constraint was necessary because the absence of any node would disrupt the spreading process across the multiplex network. In each simulation, activation was seeded at the three cue nodes and allowed to spread across both semantic and phonological layers. We monitored only the solution node, recording two outcome variables per item: maximum activation (α_m) and time of maximum activation (t_m).

Research hypothesis. Our central hypothesis is that easier RAT items should allow the solution node to accumulate activation more quickly (lower t_m) and result in higher maximum activation (higher α_m). Conversely, difficult items were expected to show slower (higher t_m) and weaker (smaller α_m) activation build-up. This is hypothesized to occur due to greater conceptual distance between cues and target, as well as more interfering and misleading concepts in between. To further understand how item difficulty influences the activation patterns, we varied key simulation parameters to analyse their effect on α_m and t_m . We examined the influence of these three parameters on the activation patterns (see Fig. 5): (i) Retention rate (set to 0.2, 0.5, or 0.8); (ii) Diffusion mode (standard diffusion vs. superdiffusion); (iii) Initial activation layer (phonological only, semantic only, or both layers simultaneously).

Single-layer versus multiplex diffusion. SpreadPy enables diffusion simulations either on individual layers, see Table 2, or on the whole multiplex structure, see Table A.1. When separating difficulty levels in terms of α_m or t_m , the strongest effect sizes are found in terms of multiplex diffusion ($d \approx 0.8$ for easy vs. hard, $p < 0.05$), either when starting from the semantic layer or when activating both replicas

or words on both layers. When diffusion is restricted only on the semantic layer, differences between RAT difficulty levels decrease while remaining statistically significant ($d \approx 0.6$ for easy vs. hard, $p < 0.05$). Interestingly, diffusion on the phonological layer only (see Table A.1) never gives rise to statistically significant differences. These numerical evidences indicate that: (i) the RAT difficulty levels, reflecting varying creativity, encode features reflected by both semantic and phonological similarities between words; (ii) semantic features are predominant, compared to phonological features, when determining the difficulty of searching for solutions in a RAT task. The first point motivates us further to explore multiplex diffusion with the RAT creativity data (see Table 3).

Peak magnitude outperforms activation timing in separating difficulty levels. Across all simulation conditions, maximum activation (α_m) provided the clearest separation of RAT item difficulty. This was especially evident in comparisons of easy vs. hard items (averaged Cohen's d across diffusion regimes and retention rates of 0.73), but also visible in easy versus medium items (averaged Cohen's $d = 0.44$). By contrast, the time to reach maximum activation (t_m) showed fewer significant contrasts ($p < 0.05$), with only easy versus hard comparisons consistently reaching moderate effect sizes (averaged Cohen's d of 0.55). Comparisons of medium versus hard items were generally not significant ($p > 0.05$) for both α_m and t_m . These results suggest that peak activation strength, rather than retrieval speed, may better reflect the associative accessibility of a correct solution in a convergent creative task such as the RAT.

Easier items reach higher peak activation. As visible in Fig. 5, maximum activation levels were significantly higher for easy RAT items compared to medium and hard items, across all retention and diffusion conditions. When starting the activation on the phonological layer (Fig. 5, top panel) there are only small effects under multiplex diffusion ($D_x = 0.1$; mean $d = 0.27$ for easy vs medium; mean $d = 0.36$ for easy vs hard). Under superdiffusion ($D_x = 1, 10$) and when starting from either the semantic layer or both layers simultaneously, effect sizes increase to moderate or strong with mean $d = 0.46$ (easy vs medium) and mean $d = 0.77$ (easy vs strong). Noticeably, when cue words were activated on both layers simultaneously, differences between diffusion modes decreased and effect sizes became more stable (see Table 2). No reliable differences were found between medium and hard items.

Easier items reach peak activation faster. Throughout the simulations, easy RAT items consistently reached their peak activation earlier (lower t_m) than medium or hard items. This pattern stayed consistent across different starting layers, retention rates and diffusion regimes. This indicates that retrieval latency, i.e. the time it takes to retrieve an item from semantic memory which is here captured by t_m (time until maximum activation), is a valuable indicator of semantic accessibility (Collins & Quillian, 1969). The comparison of easy vs hard items consistently showed significant differences with moderate effect sizes (mean $d = 0.55$). Differences between easy and medium items were inconsistent, with significant findings only in some conditions (phonological seed layer, $D_x = 0.1$; both seed layers, $D_x = 1$, $R = 0.5$; both seed layers, $D_x = 10$, $R = 0.2$) with weak to moderate effect sizes (mean $d = 0.44$). Taken together, these findings indicate that t_m robustly distinguishes easy from hard RAT items, regardless of seeding condition, retention rate and diffusion strength.

Retention rates modulate activation timing, not activation magnitude. When changing retention rates (0.2; 0.5; 0.8) in the simulations, maximum activation levels remained largely unaffected. This was somewhat unexpected, considering a higher retention rate of 0.8 means a node retains 80 percent of ingoing activation for itself, spreading only 20 percent to its neighbouring nodes. In that case, less activation should keep spreading throughout the network, resulting in potentially less activation accumulating at the solution node. Thus, one might expect to observe higher maximum activation levels for lower retention rates (more activation is spread on) than for higher retention rates (less activation is spread on). However, even though we found

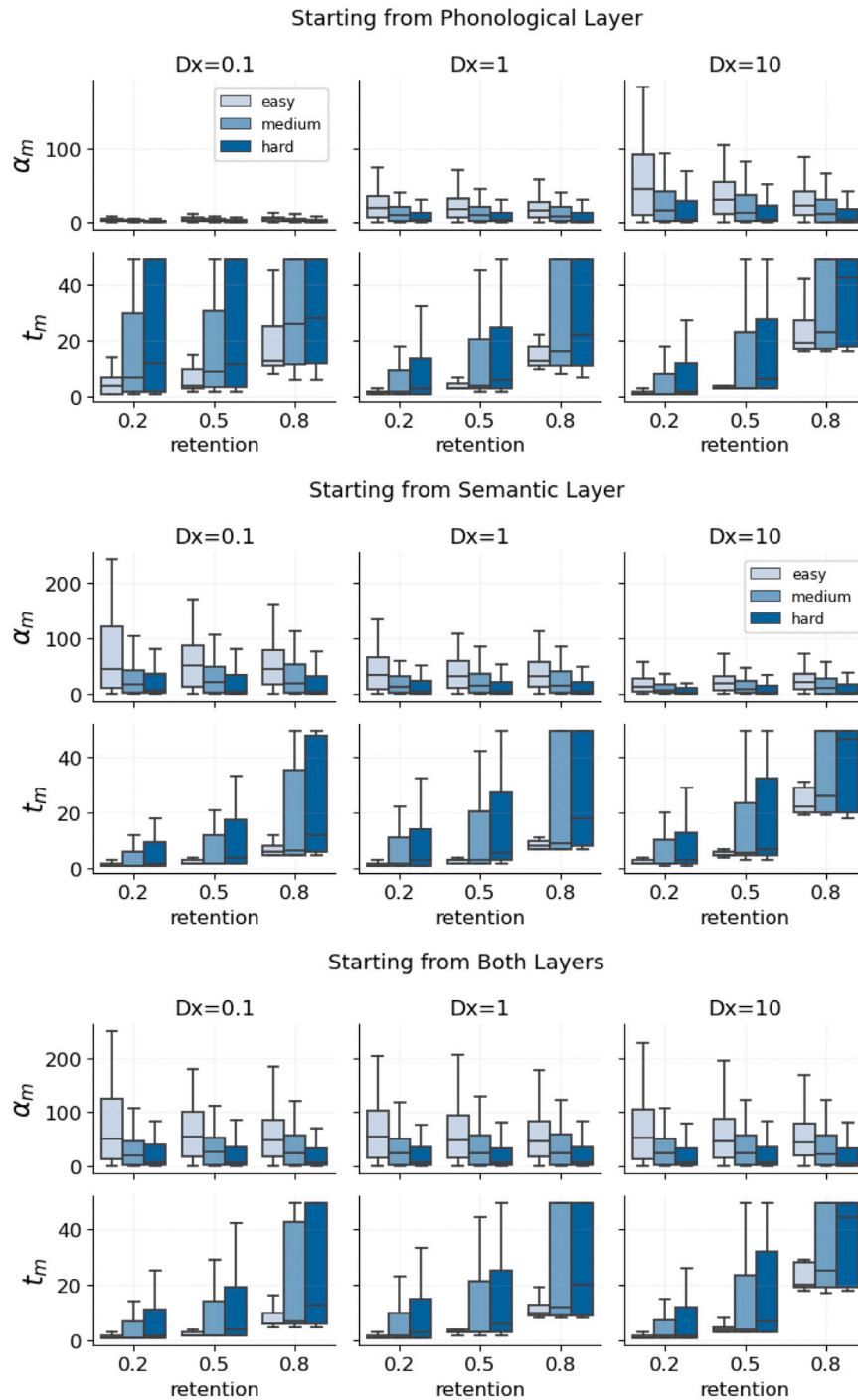


Fig. 5. Creativity: Boxplots displaying maximum activation (α_m) and time to reach it (t_m) at different diffusion ($D_x = 0.1, 1, 10$) and retention ($R = 0.2, 0.5, 0.8$) rates, starting from different layers when diffusion is seeded (phonological or semantic replica, or both).

some indications for such a pattern, our results indicate no strong variations in maximum activation levels according to modulations in retention rate.

In contrast, we found considerable correlations between retention values and the time it took to reach peak activation. Across all diffusion regimes and starting layers, higher retention consistently increased the latency of maximum activation. This effect was especially pronounced in the superdiffusion regime ($D_x = 10$), where increasing retention from 0.2 to 0.8 prolonged average and minimum t_m values for all three item difficulty types. For example, when seeding the activation at the phonological layer under superdiffusion ($D_x = 10$), the contrast in t_m

between easy and hard items increased from $d = 0.37$ (at $R = 0.2$) to $d = 0.61$ at ($R = 0.8$), while corresponding α_m differences remained stable (mean $d = 0.78$). This decoupling suggests that retention does not alter whether a solution becomes accessible (i.e. if it crosses a certain activation threshold), but when it becomes accessible (i.e. how fast activation can accumulate at the solution node). In behavioural terms, this suggests a failure to access a solution node might arise not due to insufficient activation levels, but due to a temporal delay, especially when participants are under time pressure (Collins & Quillian, 1969; Maril et al., 2001).

Table 2

Creativity: Cohen's d values for different starting layers. Bold values indicate statistically significant Kruskal–Wallis results ($p < 0.05$).

Dx	Groups	α_m			t_m		
		R=0.2	R=0.5	R=0.8	R=0.2	R=0.5	R=0.8
0.1	easy vs hard	0.27	0.38	0.43	0.69	0.72	0.57
0.1	easy vs medium	0.19	0.29	0.33	0.46	0.55	0.52
0.1	medium vs hard	0.06	0.10	0.14	0.26	0.23	0.08
1	easy vs hard	0.78	0.76	0.72	0.53	0.55	0.62
1	easy vs medium	0.48	0.46	0.44	0.40	0.42	0.47
1	medium vs hard	0.27	0.29	0.29	0.18	0.19	0.16
10	easy vs hard	0.82	0.79	0.74	0.37	0.48	0.61
10	easy vs medium	0.51	0.45	0.42	0.22	0.35	0.38
10	medium vs hard	0.26	0.30	0.30	0.17	0.15	0.23

(a) Starting From Phonological Layer

Dx	Groups	α_m			t_m		
		R=0.2	R=0.5	R=0.8	R=0.2	R=0.5	R=0.8
0.1	easy vs hard	0.81	0.82	0.78	0.43	0.46	0.64
0.1	easy vs medium	0.50	0.49	0.45	0.30	0.25	0.45
0.1	medium vs hard	0.23	0.28	0.30	0.17	0.26	0.22
1	easy vs hard	0.82	0.80	0.77	0.49	0.49	0.65
1	easy vs medium	0.50	0.47	0.43	0.33	0.33	0.44
1	medium vs hard	0.26	0.29	0.31	0.21	0.20	0.21
10	easy vs hard	0.69	0.74	0.73	0.44	0.57	0.59
10	easy vs medium	0.37	0.43	0.42	0.34	0.45	0.36
10	medium vs hard	0.28	0.29	0.30	0.13	0.16	0.23

(b) Starting From Semantic Layer

Dx	Groups	α_m			t_m		
		R=0.2	R=0.5	R=0.8	R=0.2	R=0.5	R=0.8
0.1	easy vs hard	0.81	0.82	0.78	0.51	0.52	0.64
0.1	easy vs medium	0.51	0.50	0.46	0.40	0.42	0.50
0.1	medium vs hard	0.23	0.28	0.29	0.16	0.14	0.17
1	easy vs hard	0.81	0.80	0.76	0.52	0.53	0.64
1	easy vs medium	0.50	0.47	0.44	0.38	0.41	0.46
1	medium vs hard	0.27	0.29	0.30	0.18	0.17	0.18
10	easy vs hard	0.80	0.77	0.74	0.44	0.54	0.60
10	easy vs medium	0.49	0.45	0.42	0.29	0.41	0.37
10	medium vs hard	0.27	0.30	0.30	0.19	0.17	0.23

(c) Starting From Both Layers

Table 3

Parameter estimates for the multinomial logistic regression using the *easy* class as baseline.

Variable	Category	Estimate	Std. Error	z-value	p-value	Significance
Intercept	hard	0.6523	0.0566	11.516	<0.0001	***
Intercept	medium	0.0938	0.0647	1.450	0.1470	
r	hard	-0.7475	0.1588	-4.706	<0.0001	***
r	medium	-0.5731	0.1793	-3.196	0.0014	**
α_{max}	hard	-0.0139	0.0011	-12.813	<0.0001	***
α_{max}	medium	-0.0069	0.0011	-6.362	<0.0001	***
t_{max}	hard	0.0255	0.0028	9.079	<0.0001	***
t_{max}	medium	0.0201	0.0031	6.402	<0.0001	***
$s = phon$	hard	0.2582	0.0505	5.114	<0.0001	***
$s = phon$	medium	-0.0172	0.0578	-0.297	0.7662	
$s = sem$	hard	0.4329	0.0538	8.042	<0.0001	***
$s = sem$	medium	0.1201	0.0615	1.953	0.0508	.
S	hard	0.0706	0.0080	8.769	<0.0001	***
S	medium	0.0279	0.0091	3.060	0.0022	**

Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$.

Reversed patterns across phonological and semantic layers. We find reversed patterns for maximum activation α_m when modulating the diffusion parameter and when starting activation either from the phonological or the semantic layer. When seeding the activation at the phonological layer, nearly no difficulty-related differences emerge at multiplex diffusion ($D_x = 0.1$) with small effect sizes (mean Cohen's $d = 0.26$ for maximum activation). This indicates that spreading activation confined to phonological similarity alone fails to capture the semantic associations necessary for solving the Remote Associates Task. However, under superdiffusion ($D_x = 10$), phonological cues gain access to the semantic layer via enhanced inter-layer coupling, allowing them to quickly *hop* into the semantic layer, resulting in larger effect

sizes (mean Cohen's $d = 0.62$ for maximum activation). This suggests that phonological structure alone does not encode RAT difficulty, but rather becomes informative when it allows shortcuts into the semantic network.

Conversely, when seeding the activation at the semantic layer, we find the strongest separation of item difficulty in the result variable of maximum activation at multiplex diffusion ($D_x = 0.1$) with moderate Cohen's d values (mean $d = 0.64$ for maximum activation). Allowing activation to diffuse excessively into the phonological layer under superdiffusion ($D_x = 10$) reduces these effect sizes (mean $d = 0.56$). Keeping activation within the semantic layer seems to preserve the network's capacity to discriminate easy versus hard items,

whereas adding excessive diffusion into the phonological layer blurs that discrimination. This pattern aligns with the reliance of the RAT on semantic, not phonological, associations (Bowden & Jung-Beeman, 2003; Mednick, 1962). While phonological activation may facilitate access to the solution under certain conditions, it does not appear to encode RAT item difficulty as effectively or directly as the semantic layer.

Multinomial logistic regression complements size analysis. We ran a multinomial logistic regression to identify how different predictors influence the likelihood of the categories *hard* and *medium*, relative to a baseline category (*easy*). Significant negative coefficients for r (retention rate) and α_{max} (maximum activation level) indicate that higher values on these predictors decrease the odds of responses being classified as *hard* or *medium*, with particularly strong effects for *hard*. Conversely, significant positive coefficients for t_{max} (time of maximum activation), $s = sem$ (start from the semantic layer), $s = phon$ (start from the phonological layer), and S (superdiffusion parameter) indicate that increases in these predictors enhance the likelihood of responses being categorized as *hard* or, to a lesser extent, *medium*. The lack of significant effect for $s = phon$ in predicting the *medium* category indicates this predictor primarily differentiates between *hard* and *easy* categories. Overall, predictors show consistently stronger and highly significant effects in distinguishing *hard* responses relative to the baseline category, thus agreeing with the results on size analysis.

Discussion of results. Our simulations demonstrate that easier RAT items consistently reached higher peak activation levels (higher α_m) and did so more quickly (lower t_m) than medium and hard items, indicating that easy items were structurally more accessible. This is supported by the spreading activation theory, which posits that nodes closer to the initial activated node accumulate activation more rapidly than remote ones (Collins & Loftus, 1975). The distinction between easy and harder items was robust across parameters, suggesting that temporal and magnitude-based metrics like t_m and α_m are valid indicators of structural accessibility. However, the lack of distinction between medium and hard items implies that standard RAT difficulty levels may reflect a binary division of “readily accessible” vs. “remote” rather than a fine-grained continuum. Retention analyses revealed that higher retention delays peak activation without affecting its strength. This finding highlights the importance of temporal constraints in retrieval of a solution node (Collins & Quillian, 1969; Maril et al., 2001). In practical terms, participants may fail to retrieve a correct solution not because activation never reaches a threshold, but because it does so too late under time-limited conditions. Phonological seeding only aided discrimination under superdiffusion, highlighting that RAT performance primarily relies on semantic pathways. This aligns with previous research, identifying semantic network structure as a core driver of creative problem solving during verbal creativity tasks like the RAT (Bowden & Jung-Beeman, 2003; Lee et al., 2014; Mednick, 1962).

Future studies should examine whether individual differences in retrieval speed and brain activation patterns can predict item difficulty and creativity measures, linking network dynamics to real-world creative performance (Beatty et al., 2023; Kenett & Faust, 2019).

Summary and implications. The creativity case study shows that SpreadPy can differentiate item-level difficulty in the Remote Associates Test by capturing activation dynamics across semantic and phonological layers. Differences in structural accessibility within the network caused easier RAT items to spread activation more quickly and strongly. Our SpreadPy package can use both the speed and strength of activation to simulate why certain creative problems may be more readily solved than others. This illustrates the tool’s utility for predicting task difficulty and connecting cognitive models of creativity with network-based simulations.

3.3. Case study 3: Modelling aphasic data with SpreadPy

Main goal. This case study assessed whether SpreadPy can reproduce and quantitatively dissociate the three canonical picture-naming errors observed in anomic aphasia: semantic, formal (phonological) and mixed paraphasias. Specifically, we asked whether diffusion dynamics on a multiplex lexical network predict empirical error frequencies and their characteristic types (formal, semantic or mixed), and which combinations of retention (R) and inter-layer coupling (D_x) maximize diagnostic separation.

Multiplex networks and aphasia. Aphasia syndromes can come with various deficits, such as anomia, which describes problems with accessing a desired word. In anomic aphasia, patients usually only suffer from word retrieval problems while other linguistic domains remain largely unimpaired. For example, a patient with anomic aphasia could struggle during a simple picture naming task, show slower reaction times and increased chances for mistakes, such as saying “waterfall” for a picture reporting a “bottle” (Harnish, 2018; Kohn & Goodglass, 1985). Psycholinguists can search for patterns in mistaken productions within picture naming tasks, labelling mistakes as predominantly semantic, formal/phonological, or mixed, among other types of mistakes (e.g. no production at all, non-word) (Kittredge et al., 2008; Kohn & Goodglass, 1985; Wilshire, 2002). Semantic mistakes take place when individuals can understand some semantic features of the target word – the word represented in the picture – but cannot wholly activate, retrieve or produce it entirely. Consequently, the produced mistake shares some semantic similarity with the desired target. An example of a semantic mistake would be naming a picture with a “table” as “supporter” (Kittredge et al., 2008). Analogously to semantic mistakes being influenced by semantic similarities, formal or phonemic mistakes occur when individuals produce similar sounding words to the desired target. If picture naming is performed verbally, formal mistakes are mostly driven by phonological sound similarities. An example could be naming a picture showcasing a “bunny” as “funny” (Wilshire, 2002). Mixed mistakes are relative to experimenters identifying both patterns of semantic and/or phonological similarities, e.g. saying “rat” for “cat”. Identifying the types and rates of these mistakes can help psycholinguists to better understand the extent and severity of anomic aphasia within a given individual, as well as better identify strategies for language re-training (Kittredge et al., 2008; Kohn & Goodglass, 1985; Wilshire, 2002). Early interactive activation models (see Dell et al., 1997; Schwartz et al., 2006) captured aphasic paraphasias with single-layer architectures, but struggled to explain cross-modal error blends and latency patterns (Dell et al., 1997). Subsequent two-step and production-monitor approaches improved descriptive coverage yet still lacked an explicit representation of the dual semantic–phonological topology that psycholinguistic and neuroimaging work now supports (Foygel & Dell, 2000; Schwartz et al., 2006). Multiplex network science, with its formal treatment of inter-layer diffusion and superdiffusion, offers a principled way to model how weakened semantic–phonological coupling prolongs lexical retrieval and skews the error mix. SpreadPy operationalises this framework, allowing us to map classical aphasia phenomena onto analytically tractable network parameters.

Data and modelling setup. We combined the free association network from the Small World of Words project (semantic layer) with a phonological similarity network constructed by single-phoneme edit distance. We focused on nodes that are connected on both layers, i.e. we focused on the Largest Viable Cluster of this semantic/phonological multiplex network. This yielded a two-layer network of $N = 4118$ shared word-nodes. A trial-level error corpus from the Philadelphia Naming Task project provided picture naming data from chronic anomic speakers ($N \approx 1200$ trials; equalized across error classes), used here as behavioural ground truth. For each stimulus we seeded 1.0 energy on the target word either in the semantic or phonological layer and ran deterministic diffusion for 50 time-steps under three

Table 4

Aphasia: Cohen's d , Starting from Phonological and Semantic Layers; Bold values if Kruskal–Wallis values are statistically significant, $p < 0.05$; Column L stands for “Target Layer”; α_m stands for Max Activation, and t_m stands for Time Max Activation; Legend in Errors: S: Semantic; F: formal; M: Mixed; Legend in L: Ph: Phonological, S: Semantic.

Dx	Groups	L	Starting: Phonological Layer						Starting: Semantic Layer					
			α_m			t_m			α_m			t_m		
			R=0.2	0.5	0.8	R=0.2	0.5	0.8	R=0.2	0.5	0.8	R=0.2	0.5	0.8
0.1	S vs F	Ph	0.01	0.00	0.03	0.18	0.11	0.26	0.29	0.26	0.11	0.65	0.58	0.38
0.1	S vs M	Ph	0.03	0.02	0.02	0.40	0.32	0.38	0.23	0.23	0.13	0.48	0.51	0.41
0.1	F vs M	Ph	0.02	0.01	0.01	0.25	0.23	0.14	0.08	0.03	0.02	0.15	0.07	0.03
0.1	S vs F	S	0.31	0.30	0.16	0.76	0.72	0.55	0.11	0.01	0.07	0.90	1.13	1.47
0.1	S vs M	S	0.26	0.28	0.17	0.61	0.64	0.55	0.10	0.00	0.08	0.93	1.17	1.46
0.1	F vs M	S	0.06	0.03	0.01	0.12	0.07	0.01	0.01	0.01	0.01	0.00	0.01	0.03
1	S vs F	Ph	0.08	0.09	0.00	0.37	0.41	0.09	0.20	0.20	0.08	0.42	0.49	0.37
1	S vs M	Ph	0.09	0.10	0.01	0.31	0.35	0.05	0.23	0.22	0.10	0.40	0.48	0.37
1	F vs M	Ph	0.01	0.01	0.01	0.09	0.09	0.05	0.02	0.02	0.02	0.03	0.02	0.00
1	S vs F	S	0.18	0.23	0.12	0.65	0.69	0.57	0.03	0.07	0.07	0.72	0.79	0.93
1	S vs M	S	0.23	0.26	0.15	0.70	0.70	0.57	0.05	0.09	0.09	0.77	0.82	0.95
1	F vs M	S	0.04	0.02	0.02	0.03	0.00	0.01	0.01	0.01	0.02	0.04	0.02	0.05
10	S vs F	Ph	0.01	0.06	0.03	0.39	0.52	0.14	0.12	0.14	0.04	0.31	0.40	0.05
10	S vs M	Ph	0.03	0.07	0.01	0.42	0.49	0.19	0.18	0.16	0.01	0.27	0.35	0.09
10	F vs M	Ph	0.02	0.01	0.02	0.02	0.03	0.05	0.05	0.01	0.03	0.05	0.06	0.04
10	S vs F	S	0.00	0.18	0.02	0.57	0.69	0.43	0.02	0.07	0.02	0.47	0.57	0.31
10	S vs M	S	0.04	0.20	0.06	0.64	0.69	0.49	0.05	0.09	0.01	0.50	0.56	0.38
10	F vs M	S	0.03	0.01	0.03	0.06	0.00	0.08	0.02	0.01	0.02	0.01	0.01	0.07

Table 5

Aphasia, multiplex: Kendall Correlation between Time Max Activation (t_m) and Real Frequency Error (all values statistically significant).

Dx	Err.	Layer	Starting Phon			Starting Sem			Starting Both		
			R=0.2	0.5	0.8	R=0.2	0.5	0.8	R=0.2	0.5	0.8
0.1	F	Ph	-0.18	-0.18	-0.20	-0.17	-0.20	-0.20	-0.18	-0.19	-0.20
0.1	M	Ph	-0.08	-0.15	-0.18	-0.18	-0.19	-0.19	-0.15	-0.18	-0.17
0.1	S	Ph	-0.04	-0.13	-0.21	-0.22	-0.25	-0.26	-0.15	-0.22	-0.25
0.1	F	S	-0.14	-0.16	-0.18	-0.04	-0.09	-0.15	-0.13	-0.15	-0.19
0.1	M	S	-0.19	-0.20	-0.20	-0.19	-0.19	-0.19	-0.18	-0.19	-0.19
0.1	S	S	-0.26	-0.25	-0.24	-0.25	-0.25	-0.24	-0.24	-0.24	-0.24
1	F	Ph	-0.20	-0.20	-0.20	-0.20	-0.20	-0.20	-0.20	-0.20	-0.20
1	M	Ph	-0.17	-0.18	-0.18	-0.18	-0.19	-0.19	-0.18	-0.19	-0.19
1	S	Ph	-0.23	-0.24	-0.26	-0.23	-0.24	-0.25	-0.23	-0.24	-0.26
1	F	S	-0.18	-0.19	-0.19	-0.17	-0.18	-0.19	-0.17	-0.18	-0.19
1	M	S	-0.20	-0.21	-0.20	-0.20	-0.20	-0.21	-0.21	-0.20	-0.21
1	S	S	-0.25	-0.25	-0.26	-0.25	-0.25	-0.25	-0.25	-0.25	-0.26
10	F	Ph	-0.19	-0.19	-0.17	-0.20	-0.20	-0.17	-0.20	-0.20	-0.17
10	M	Ph	-0.19	-0.19	-0.18	-0.18	-0.18	-0.19	-0.19	-0.19	-0.18
10	S	Ph	-0.23	-0.24	-0.23	-0.22	-0.24	-0.22	-0.22	-0.24	-0.22
10	F	S	-0.17	-0.19	-0.18	-0.18	-0.19	-0.18	-0.17	-0.19	-0.18
10	M	S	-0.21	-0.20	-0.18	-0.19	-0.20	-0.19	-0.21	-0.21	-0.18
10	S	S	-0.24	-0.25	-0.26	-0.24	-0.26	-0.26	-0.24	-0.26	-0.26

coupling regimes ($D_x = 0.1, 1, 10$) and three retention levels ($R = 0.2, 0.5, 0.8$). Outcome measures were maximum activation (α_m) and time-to-maximum activation (t_m) in the target layer; group contrasts were quantified with Cohen's d and evaluated using Kruskal–Wallis tests.

Correlations with behavioural outcomes. The correlations in [Table 5](#) show that our simulation measures can be meaningfully related to behavioural outcomes. In particular, time-to-maximum activation (t_m) can be interpreted as reaction times in picture naming tasks. The consistent negative Kendall correlations in [Table 5](#) indicate that items which reach peak activation earlier in the network are also those that more often lead to naming errors in patients. This is consistent with the idea that premature activation of competitors may disrupt target retrieval and underlie error production. Likewise, by correlating simulation dynamics with empirical error rates, we capture a parallel to accuracy measures typically reported in behavioural studies. These results demonstrate that SpreadPy can generate predictions that are comparable to established behavioural metrics.

Single-layer versus multiplex diffusion when distinguishing errors. When differentiating semantic, formal and mixed mistakes, activation spreading on the multiplex structure (see [Table 4](#)) and on single layers (see [Table A.3](#)) provide different results. The strongest

differences in t_m are found when activation spreads on the semantic layer only ($d \approx 1.94, p < 0.05$), followed by diffusion involving both multiplex layers but starting from the semantic one ($d \approx 1.47, p < 0.05$). These differences persist across all tested retention rates and are relative to distinguishing semantic and formal (phonological) mistakes. Considerably smaller differences are found on the phonological layer alone or on the multiplex network when starting from the phonological layer ($d \approx 0.4, p < 0.05$). These robust findings indicate that semantic and phonological mistakes might be differentiated by temporal dynamics more strongly related to the semantic layer only than to a mix of phonological similarities and free associations.

Single-layer versus multiplex diffusion for error frequency. In terms of predicting the frequency of error types based on their diffusion dynamics, the multiplex network provides only marginally higher correlations ($\tau \approx -0.26$, see [Table 5](#)) compared to the semantic layer only ($\tau \approx -0.25$, see [Table A.4](#)) for semantic mistakes. This finding confirms that the semantic layer determines most of the diffusion dynamics determining activation patterns related to semantic mistakes. Interestingly, free associations encode also activation patterns relative to mixed mistakes, since t_m on both the semantic layer only and on the multiplex network exhibit similar correlations ($\tau \approx -0.20$) with the frequency of mixed mistakes. This similarity might be due to free

associations being a multidimensional set of semantic but also mixed associations, as reported in prior works (Baker et al., 2023). Where multiplex diffusion provides significantly stronger results than single-layer diffusion is for the correlations with the frequency of formal mistakes. These correlations are at most $\tau \approx -0.12$ in the phonological layer only, while they increase to $\tau \approx -0.20$ in the multiplex case. These patterns interestingly indicate that formal mistakes might be qualitatively different from semantic mistakes. Whereas semantic mistakes tend to occur with frequencies correlating mostly to time delays contained mostly in the semantic layer, formal mistakes occur with frequencies correlating more strongly with time delays exhibited by a multiplex phonological/semantic multiplex structure. Hence, formal mistakes might be due to an interference between semantic and phonological similarities, as posited also by past relevant works in aphasia research (Baker et al., 2023; Dell, 1986; Dell et al., 1997).

Single-layer versus multiplex diffusion. The above findings indicate that single-layer diffusion never outmatches multiplex diffusion in terms of distinctions between error classes or correlations with error frequencies. Instead, multiplex diffusion can provide significant information about formal mistakes, information that cannot be retrieved in single-layer diffusion. For these reasons, in the remained of this Section, we focus on multiplex diffusion.

Temporal dynamics – not peak magnitude – best discriminate aphasic error classes. With reference to Table 4, across all diffusion-coupling conditions ($D_x = 0.1, 1, 10$) Cohen's d for time-to-maximum activation is consistently an order of larger magnitude than for maximum activation. When activation is seeded in the phonological layer and measured on the same layer, t_m reaches medium effect sizes (up to $d \approx 0.40$) whereas α_m seldom exceeds $d = 0.03$. The pattern is even clearer in the semantic layer: with semantic seeding, between semantic vs. formal and semantic vs. mixed errors, t_m climbs from $d \approx 0.90$ at weak coupling ($D_x = 0.1$) to $d \approx 0.57$ at strong coupling ($D_x = 10$), while the corresponding α_m contrasts remain ≤ 0.11 . Hence, the latency with which a concept reaches peak activation carries the diagnostic signal, not the height of that peak.

Semantic errors diverge earliest and most strongly. For $D_x = 0.1$, differences in t_m between semantic and formal errors in the semantic layer reach $d = 1.47$ when retention is high ($R = 0.8$), representing a very strong effect. Even at moderate retention ($R = 0.5$) the effect remains large ($d \approx 1.13$). No other error contrast approaches this magnitude, indicating that semantic-level competition is the primary driver of delayed naming in anomic aphasia.

Superdiffusion modulates, but does not erase, temporal separation. Increasing the coupling from $D_x = 0.1 \rightarrow 1$ halves the largest t_m effect sizes (e.g. d from $1.47 \rightarrow 0.93$ for semantic vs. formal), and a further ten-fold increase ($D_x = 10$) reduces them to the upper-medium range ($d \approx 0.57$). The persistence of sizeable t_m gaps even under strong coupling, i.e. in the superdiffusion regime, suggests that lexical-phonological shortcuts cannot fully compensate for degraded semantic access.

Robust inverse link between activation time and error prevalence. With reference to Table 5, across every simulated configuration the Kendall τ between t_m and the empirical frequency of naming errors is negative ($\tau \approx -0.08, \dots, -0.26$). Thus, items whose activation peaks earlier within the network are more likely to elicit a paraphasia in patients. This finding is consistent with interactive models in which quickly rising but mis-targeted activation out-competes the correct lemma before speech monitoring can intervene. Within the semantic layer τ reaches the largest magnitudes for semantic errors, ($-0.24 \leq \tau \leq -0.26$), followed by formal ($-0.14 \leq \tau \leq -0.20$) and mixed errors ($-0.18 \leq \tau \leq -0.20$). The gradient mirrors clinical incidence rates (semantic > mixed > formal) and indicates that SpreadPy's diffusion measures on the current layers are most diagnostic for the frequency of semantic selection errors. Increasing R incrementally strengthens the inverse correlation. By contrast, being in multiplex or superdiffusion regime has negligible impact on τ . Whether activation is injected

into the phonological, semantic or both layers simultaneously, τ estimates differ by ≤ 0.03 within each condition. This suggests a lack of dependency on the initial multiplex or single-layer seeding.

Discussion of results. Within every D_x condition, larger retention values systematically inflate t_m effect sizes, particularly in the semantic layer. High retention therefore prolongs intra-node accumulation, stretching the time axis and enhancing sensitivity to processing delays characteristic of different error types. Importantly, our numerical experiments indicate that, when modelling picture naming in people with aphasia, researchers should prioritize t_m -based metrics and explore low-to-moderate inter-layer coupling with higher retention. These settings maximize the model's ability to reproduce the empirically observed separation between semantic, formal and mixed errors. A lack of variation in the correlations with error frequencies indicates that superdiffusion, and hence cross-layer shortcuts, matter less than temporal dispersion within layers. The stable, negative t_m -frequency association validates t_m as a mechanistic proxy for error susceptibility. It further supports the hypothesis that aphasic naming failures are driven by prematurely peaking competitors rather than by excessively strong ones. Because the effect is parameter-robust, t_m can serve as a reliable optimization target when adapting SpreadPy to individual patients or exploring therapeutic interventions that aim to prolong lexical competition in damaged semantic networks.

Summary and implications. In the aphasia case study, SpreadPy successfully modelled semantic, formal, and mixed naming errors by using single-layer and multiplex network structures. While semantic errors were best captured within the semantic layer, formal errors emerged more clearly when interactions between semantic and phonological layers were taken into account in a multiplex simulation. These results highlight SpreadPy's clinical relevance: it can link error types to network structure, providing a computational testbed for hypotheses about aphasic impairments.

4. Discussion

The present work puts the spreading activation model proposed by Collins and Loftus (1975) into computational form encompassing Python and multiplex cognitive networks: SpreadPy. Collins and Loftus' spreading activation model describes how concepts in memory are activated by energy spreading across links between conceptual representations. SpreadPy's central advantage is its ability to simulate spreading activation within both single-layer and multiplex cognitive networks (Stella et al., 2024). To our knowledge, this represents a first pioneering attempt to implement spreading activation across multiple interconnected layers. This is relevant because multiplex cognitive network allow for the representation of different types of information such as semantic and phonological relationships, that are represented on distinct but interacting layers that can influence the activation dynamics of each other (Stella et al., 2017, 2018, 2024). This enables researchers to simulate how different types of information, such as meaning and sound, interact during cognitive processes like word associations, creativity, or error production. SpreadPy builds upon the functionality of the existing *spreadr* tool, implemented in R (Siew, 2019). Our Python-based framework replicates all features of *spreadr* for single-layer networks, but significantly extends the tool by supporting multiplex architectures. To our knowledge, there is currently no equivalent open-source Python tool that enables spreading activation simulations on single-layer and multiplex networks.

SpreadPy supports modelling networks in multiple languages, as demonstrated through simulations using English or Italian lexical networks. This allows researchers to work directly with data in the original language, avoiding the need for translation, which is a common limitation in many modelling tools. Modelling cognitive data in the original language is crucial, as linguistic structures such as word associations or semantic distances vary significantly across languages. Thus, translations may fail to preserve key cognitive or structural patterns specific to a certain language (Vigliocco & Kita, 2006).

4.1. Technical simulation scenarios

As a computational framework, the SpreadPy library is flexible enough to adapt to several theories and several different parameters. For instance, spreading activation can be used in networks of document similarities (Crestani, 1997) or in cognitive networks (Collins & Loftus, 1975; Dóczy, 2019) as extensively tested in the case studies of this paper.

SpreadPy enables researchers to simulate a wide range of cognitive phenomena by flexibly adjusting key parameters in line with theoretical assumptions. For instance, varying the retention parameter R allows to model how sustained activation might influence lexical retrieval. For example, in modelling the tip-of-the tongue phenomenon, low retention may simulate a temporary failure to access a given word (Burke et al., 1991). In aphasia, different retention values may reflect impairments in maintaining activation on target nodes (Dell et al., 1997).

Adjusting the diffusion parameter D_x (Gomez et al., 2013) allows control over the strength of interlayer coupling. By doing so, researchers can define whether activation spreads more readily within or across layers. This could be used for investigating mixed errors in aphasia (Dell et al., 1997), where interaction between layers is essential. High coupling simulates strong interactions between layers, while low coupling reflects more isolated layer functioning.

Choosing the starting point of activation allows modelling task-dependent differences. For instance, tasks involving category verification (Mulligan & Peterson, 2008) may start from the semantic layer as they focus on meaning, while phonological similarity judgements (Siew & Castro, 2023) may begin from the phonological layer due to their reliance on phonological information. Experiments involving both semantic and phonological information at the same time, such as auditory stimuli during a lexical decision task, can be modelled with activation starting from both layers simultaneously.

Another key modelling choice lies in selecting which layer/s the network is built from. Researchers focusing on phonology can simulate activation in a phonological single-layer network. This approach is well suited for experiments investigating phonological similarity judgements (Siew & Castro, 2023), spoken-word recognition (Vitevitch & Aljasser, 2021) or processing spoken nonwords (Vitevitch et al., 1997). By isolating the phonological layer, SpreadPy allows users to examine effects of phonological similarity or neighbourhood density without interference from semantic associations (Siew, 2013; Siew & Castro, 2023). To model phonological impairments, researchers can lower the retention parameter to simulate rapid decay of activation on the phonological layer, similar to degraded phonological working memory (Grivol & de Vasconcellos Hage, 2011).

Likewise, researchers interested in meaning-based processing can model activation spreading in semantic single-layer networks. This is relevant for investigating semantic priming effects (Groot, 1983), category verification (Mulligan & Peterson, 2008), or the role of semantic distance in judgements of semantic relatedness (Kenett et al., 2017). Moreover, SpreadPy allows researchers to add further information as weight to nodes or links. For instance, nodes can be weighted according to their frequency which can help in explaining strong or weak priming effects (Yap et al., 2009).

SpreadPy's support for multiplex cognitive networks (Stella et al., 2024) means that experimenters can simulate scenarios in which phonological and semantic layers interact dynamically. This is particularly relevant for studies of lexical decision (Coltheart et al., 1979), speech-in-noise recognition (Bradlow & Alexander, 2007) or anomic aphasia (Baker et al., 2023), where both types of information have been found to influence performance (Baker et al., 2023; Castro & Siew, 2020; Collins & Ellis, 1992; Coltheart et al., 1979).

Beyond cognitive science, SpreadPy is also suitable for applications in information retrieval (Crestani, 1997). For instance, computer scientists can simulate semantic search processes in digital libraries, where nodes represent documents and edges reflect shared features such as

authorship, genre, or keyword overlap (Ševcech et al., 2014). In such contexts, spreading activation models can help evaluate how relevant documents are retrieved based on initial query activation, and how these signals evolve over time.

4.2. Complexity analysis and parameter sensibility

The update rules of SpreadPy are implemented as local, edge-based operations: at each iteration, only nodes with non-zero activation and their neighbours are updated. This ensures that complexity scales with the number of active edges, $\mathcal{O}(|E_a|)$, rather than with the full graph size. For sparse lexical and semantic networks, where $|E| \ll |V|^2$, this results in nearly linear-time behaviour. In the multiplex case, memory usage grows as $\mathcal{O}(|V| \cdot |L|)$, since each node is represented in every active layer, while computation scales linearly with intra- and inter-layer edges.

Moreover, previous discussion on technical simulation scenarios highlights how analysts must carefully consider parameter sensitivity. In this regard, SpreadPy allows systematic exploration of the effects of R , D_x , initial activation nodes, and layer selection. In the three selected case studies (cf. next paragraph), small changes in R affect activation decay and retention on target nodes, influencing peak activation and retrieval latency. Adjusting D_x modulates inter-layer coupling, affecting diffusion dynamics in multiplex versus single-layer networks. Similarly, the choice of starting nodes and activated layers can alter activation pathways and outcomes. Parameter sweeps or sensitivity analyses are recommended to ensure robust and interpretable results.

4.3. Case study insights

We demonstrated the utility of SpreadPy through three distinct case studies investigating spreading activation in relation to: (i) math anxiety, (ii) creativity, and (iii) aphasia.

In the affective cognition case study, SpreadPy was used to model the semantic associations of participants with varying levels of math anxiety. We used psychometric scores to classify participants into high- and low-math-anxiety groups. On group-level single-layer networks constructed from free association data, we activated the node “mathematics” and measured peak activation α_m and time of peak activation t_m on the nodes “anxiety”, “stress” and “depression”.

The simulations revealed that in high math-anxious individuals, “mathematics” triggered rapid activation of the concept “anxiety”, reflecting direct and strong links in the semantic-affective network. In contrast, individuals without math anxiety showed more diffuse activation paths, involving intermediate nodes that “softened” the link between math and negative affect. These findings highlight a mirroring between the structure of memory, a sustained activation of negative affect, and individual differences in math anxiety. Individuals that were classified as having high math anxiety through psychometric scores tend to maintain stronger activation of negative concepts, and their cognitive networks reflect this negative bias (Stella, 2022). SpreadPy thus allows researchers to operationalize how affective features are encoded in the structure of cognitive networks and to explore how these structures shape activation flow over time.

In the creativity case study, we investigated if SpreadPy could differentiate between items with varying difficulty in the Remote Associates Test (RAT). We constructed multiplex networks consisting of a semantic and a phonological layer. On either the semantic, the phonological or both layers simultaneously, we activated all three cue words (e.g. “cottage - Swiss - cake”) at the same time and measured peak activation and time of peak activation at the solution node (e.g. “cheese”). SpreadPy could effectively differentiate items by difficulty, predicting faster and higher activation peaks for easier items (see Fig. 5). Higher activation could reach solution nodes more quickly at shorter distances from the initially activated nodes. This interpretation can explain why easier RAT solutions exhibited lower t_m (time of max activation) and higher

α_m (peak activation). In contrast, more difficult items typically involve more remote associations, according to Mednick's Associative Theory of Creativity, which proposes that creative thinking involves bridging distant semantic concepts (Mednick, 1962). Our SpreadPy simulations provide a quantitative match to this theory: items rated as more difficult required more time and exhibited lower peak activation compared to easy items.

Beyond supporting established creativity theories, this case study also highlights the advantages of modelling cognitive search mechanisms relevant for creativity on a multiplex network. Our results showed that differences in RAT item difficulty were most pronounced when activation diffused through a multiplex structure combining semantic and phonological layers (see Table 2). Although the semantic layer alone produced significant distinctions, reflecting the task's inherently semantic nature, results from the multiplex network showed greater differentiation between the difficulty levels. In contrast, diffusion on the phonological layer alone produced no significant results, further underscoring that the RAT task relies mostly on semantic memory relationships rather than sound-based cues (Bowden & Jung-Beeman, 2003; Lee et al., 2014). However, the combined semantic–phonological multiplex network provided additional discriminative strength, suggesting that while semantic pathways drive solution access, phonological connections may further support retrieval or strengthen activation of the solution indirectly.

This finding supports broader psycholinguistic claims that semantic and phonological layers are tightly coupled (Collins & Ellis, 1992; Coltheart et al., 1979; Dóczy, 2019; Levelt, 2001) and that even semantically dominated tasks may benefit from models that can incorporate phonological information and allow for its interaction with semantic information.

The aphasia case study demonstrated how SpreadPy can be applied in clinical settings, e.g. to simulate picture-naming errors in individuals affected by disorders from the aphasic spectrum (Baker et al., 2023). We constructed multiplex networks with a semantic layer and a phonological layer. We started activation at the target word (name of the object in the picture, e.g. “bottle”) and measured how quickly (t_m) and how much maximum activation (α_m) arrived at the mistaken solution (e.g. “waterfall”).

We investigated 3 different aphasic error types: semantic, formal and mixed errors. Semantic errors involve the production of words that are conceptually related to the target (e.g. saying “supporter” for “table”) (Kittredge et al., 2008), while formal errors are driven by phonological similarity (e.g. saying “funny” for “bunny”) (Wilshire, 2002). Mixed errors combine both semantic and phonological overlap (e.g. saying “rat” for “cat”) (Baker et al., 2023). Our simulations indicate that the temporal feature of activation spreading, i.e. the time to reach peak activation, was better at discriminating between error types compared to maximum activation itself. Semantic errors were best captured using the semantic single-layer network, with clear distinctions from other error types based on t_m .

This suggests a crucial advancement for aphasia studies. Our SpreadPy simulations show quantitatively that semantic mistakes arise primarily from semantic representations, the multiplex network performed almost as well in identifying semantic errors as the semantic layer alone. In contrast, formal errors were poorly differentiated from other error types in the phonological single-layer network. However, in the multiplex network, formal mistakes were far more distinguishable, suggesting that semantic information interferes with phonological retrieval. Even though formal errors may ostensibly appear to be a by-product of processes that occur on the phonological layer alone, our simulations suggest a more nuanced mechanism by which formal errors emerge from the interaction between semantic and phonological processes, rather than simply being confined to the phonological layer (Castro & Stella, 2019; Dell et al., 1997). This makes the multiplex network especially valuable: it models how semantic errors arise within the semantic layer, and also captures the interference of semantic activation on the phonological layer, which helps explain formal errors.

4.4. Investigation of superdiffusion across experiments

In multiplex cognitive networks (Stella et al., 2024), activation spreads simultaneously across semantic and phonological layers, and their interaction produces a synergistic acceleration, i.e. superdiffusion, that goes beyond what either pathway alone can achieve. Psychologically, superdiffusion corresponds to the ability of speakers to retrieve words more rapidly by drawing on multiple representational codes in parallel, semantic associations (e.g., “dog” → “cat”) and phonological cues (e.g., “mat” → “cat”), which converge to amplify activation of the target word. Superdiffusion from multiplex networks is a process-level account of how interactive activation across modalities boosts efficiency in word finding. Hence superdiffusion consistently aligns with spreading activation theories (Collins & Loftus, 1975) and interactive lexical access models (Dell & O'Seaghdha, 1991).

Importantly, superdiffusion in cognitive multiplex networks, as modelled with SpreadPy, corresponds most directly to accelerated and more efficient lexical access rather than to a general reduction in cognitive load. The simulations show that when activation is allowed to spread jointly across semantic and phonological layers, target nodes (e.g., solution words in creativity tasks, lexical items in aphasia data) not only reach higher activation peaks but also do so faster than in single-layer or standard multiplex regimes. In Case Study 2, easier Remote Associates Test items reached higher and earlier activation under superdiffusion. In Case Study 3, the timing of activation peaks predicted error types in aphasia. Hence, in our experiments with cognitive/psychological data, superdiffusion manifests behaviourally as shorter retrieval latencies and increased likelihood of correct access under time constraints, both elements mappable in terms of faster access and not directly indicative of lower cognitive load. Rather than reducing the overall cognitive effort required, superdiffusion provides a mechanistic explanation for why certain lexical retrievals are both quicker and more accurate: synergistic cross-layer interactions allow information to accumulate at targets more rapidly than through either semantic or phonological pathways alone.

Superdiffusion also has implications for learning and memory. In educational settings, students with well-integrated knowledge networks (e.g., strong connections between mathematical concepts and real-world applications) may exhibit superdiffusive activation, facilitating problem-solving learning. Conversely, conditions like math anxiety, which fragment associative networks, could disrupt superdiffusion, leading to slower and less efficient cognitive processing (Artemenko et al., 2015; Stella, 2022). By modelling these dynamics, SpreadPy enables researchers to quantify the impact of network structure on cognitive performance, offering insights into individual differences and potential interventions.

Finally, superdiffusion provides a mechanistic framework for studying higher-order cognition, such as creativity and reasoning. In creative tasks, for example, the ability to combine distant concepts may rely on superdiffusive activation across semantic and contextual layers. SpreadPy's multiplex simulations allow researchers to test these hypotheses, exploring how network topology and layer coupling influence cognitive flexibility. By bridging network science and cognitive psychology, superdiffusion models can open new avenues for understanding the neural and computational principles underlying human thought.

A promising future direction is to design experiments that orthogonally manipulate network coupling and task load to disentangle whether superdiffusion reflects faster lexical access or reduced cognitive effort. For example, items with high versus low semantic–phonological coupling (predicting faster access in SpreadPy via earlier and stronger activation peaks) could be tested under varying secondary task demands (e.g., n-back or articulatory suppression). Measuring both retrieval speed/accuracy and independent load markers such as pupil dilation or EEG theta power would allow researchers to determine whether superdiffusion selectively enhances access speed (earlier

responses, higher drift rates in diffusion models) without reducing physiological indicators of effort, thus clarifying whether its primary psychological signature is accelerated retrieval rather than lowered cognitive load.

4.5. Limitations

Although SpreadPy provides a flexible and extensible framework, several limitations should be acknowledged. First, all case studies presented here relied on language-based networks, limiting the generalizability of our demonstrations to other domains of cognition. Secondly, the datasets used in the present work are restricted in scope, focusing on English and Italian lexical networks. While SpreadPy technically supports cross-linguistic simulations even in less-studied languages and other writing systems, its application in other languages needs to be tested. Furthermore, the quality of outcomes remains dependent on the availability and representativeness of the underlying network data. Finally, as with all computational models, the predictive value of SpreadPy hinges on the validity of the input parameters and network structure. The library does not impose a single theory of cognition, which increases flexibility but also places responsibility on researchers to construct and validate networks that are theoretically grounded and empirically supported.

While SpreadPy is designed as a computational tool for simulating cognitive activation dynamics, its output is only as robust as the network it is given. Because the tool depends on the structural quality of the input networks, whether constructed from human association data or Large Language Models, users must pay attention to the validity and representativeness of their input networks. SpreadPy does not prescribe a single architecture or theory, and this openness is a strength, but this places responsibility on the researcher to construct transparent and valid networks. Nevertheless, at the same time, network structure alone may not fully capture the richness of cognitive processes, which also depend on temporal, contextual, and neural dynamics. A possible future extension would be to integrate external sources of information, such as neural data, behavioural measures, or contextual embeddings, into the network representation, and to combine structural and attribute-based features within the model. Such developments would allow SpreadPy to move beyond purely structural simulations and toward more comprehensive accounts of cognitive phenomena.

4.6. Impact for future research

Our findings not only demonstrate SpreadPy's capacity for modelling cognitive dynamics in multiplex systems, but also point toward several promising directions for methodological development and future research.

From a practical standpoint, SpreadPy addresses several limitations of classical cognitive modelling frameworks. Models like Dell's interactive activation of word production (Dell, 1986; Dell & O'Seaghdha, 1991) are restricted by small-scale lexicons, while McClelland and Rumelhart's interactive activation model of context effects (McClelland & Rumelhart, 1981) rely on seemingly arbitrary parameter settings that hinder generalization and reproducibility. Meanwhile, Collins and Loftus' (1975) spreading activation theory remains primarily verbal and difficult to operationalize. SpreadPy offers a clear computational advantage by supporting simulations on realistically sized networks with transparent and interpretable parameters and flexible parameter tuning.

Notably, the case studies presented in this paper make use of only a subset of SpreadPy's functionality. Several built-in features remain available for future exploration. For example, SpreadPy supports the integration of node- or edge-level attributes such as valence or word frequency. Future work could modulate activation dynamics using such attributes, aligning simulations more closely with known patterns

of language use and word recognition. Moreover, although the current case studies employed a uniform diffusion across the network, the SpreadPy tool is able to simulate weighted spreading based on these attributes. Furthermore, inhibitory mechanisms are also implemented in the tool. We inherited the suppression parameter from Siew's *spreadr* (Siew, 2019), which globally reduces activation below a given threshold. Future refinements of this feature could simulate inhibition of specific nodes or pathways, as seen in lexical competition or error correction.

Furthermore, in contrast to many traditional cognitive models that provide static outcomes, SpreadPy allows for the tracking of activation over time. Thus, it may be combined with techniques such as EEG, eye-tracking, and mouse-tracking which emphasize the temporal unfolding of cognitive processes rather than static end states (Spivey et al., 2005). In addition to tracking peak activation and latency of peak activation, SpreadPy allows activation trajectories to be computed and visualized across time points and for each individual node. This enables a far more detailed comparison between empirical data and simulated dynamics than traditional models that only predict endpoint behaviour, such as reaction times or accuracy rates.

SpreadPy addresses distinct methodological needs across multiple research communities. Psychology and cognitive science researchers can leverage the tool to operationalize theories of semantic memory and spreading activation, quantifying how interventions alter associative structures while bridging psychometric measures with computational simulations. Computer science applications encompass algorithm validation, network analysis methodologies, and developing algorithmic models that simulate human semantic processing. Neuroscience benefits from SpreadPy's capacity to model network-level impairments, exploring how semantic degradation or compromised lexical access manifest in aphasia, dementia, or developmental disorders, with temporal dynamics serving as proxies for neural processing efficiency and connectivity patterns.

Finally, SpreadPy opens new opportunities for cross-linguistic simulations. The tool supports the simulation on networks in any language, making it well suited to test predictions, e.g. word recognition tasks across languages with different writing systems (e.g., alphabetic vs. logographic). For instance, researchers could compare how orthography–phonology–semantic interactions differ between English and Chinese, testing extensions of the Seidenberg and McClelland framework (Seidenberg & McClelland, 1989; Yang et al., 2006) within a multiplex network environment.

4.7. Conclusion

Finally, SpreadPy serves as an accessible and fully open-source tool for simulating how minds activate, associate and sometimes err through the structure of their mental networks. It facilitates theory testing, and supports empirical designs that go beyond static outcomes to capture the dynamics of cognition as it unfolds.

CRedit authorship contribution statement

Salvatore Citraro: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Edith Haim:** Writing – original draft, Visualization, Resources, Methodology, Investigation, Formal analysis, Data curation. **Alessandra Carini:** Writing – original draft, Visualization, Investigation, Formal analysis, Data curation. **Cynthia S.Q. Siew:** Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Giulio Rossetti:** Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Formal analysis. **Massimo Stella:** Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Code availability

The SpreadPy library and all code used for the simulations are openly available at the following link: <https://github.com/dsalvaz/SpreadPy>.

Consent to participate

Not applicable. No new participants were recruited as part of this study.

Consent for publication

Not applicable. All data used were from publicly available sources with prior publication.

Ethics approval

Ethics approval was not required as no new data were collected. All case studies are based on previously published, anonymized datasets.

Declaration of the use of AI

During the preparation of this work the authors did not use AI for any part of the work related to the manuscript submitted.

Funding

MS acknowledges support from: (1) Dipartimento di Eccellenza of the Department of Psychology and Cognitive Science, University of Trento, (2) COGNASCO/COGNOSCO project funded by University of Trento's Call for Research 2024/2027.

This work is supported by SoBigData.it which receives funding from the European Union NextGenerationEU National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR) Project: SoBigData.it Strengthening the Italian RI for Social Mining and Big Data Analytics Prot. IR0000013 Avviso n. 3264 del 28/12/2021.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge Giuseppe Fonte for useful preliminary feedback on the creativity case study dataset.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.chbr.2026.100940>.

Availability of data and materials

All datasets used in the case studies are publicly available through the original publications. For case study 1 on math anxiety, we constructed the networks based on data from Ciringione et al. (2024). For the multiplex networks used in case study 2 and 3, we constructed the semantic layer from the Small World of Words (SWoW) dataset collected by Deyne et al. (2019). The phonological layer was constructed from phonological similarities collected by Vitevitch (2008). For case study 2 on creativity, we used RAT items reported in Bowden and Jung-Beeman (2003), Bowers et al. (1990), Lee et al. (2014), Wu et al. (2020). Finally, the database for case study 3 on aphasic errors was used from Castro and Stella (2019).

References

- Abramski, K., Citraro, S., Lombardi, L., Rossetti, G., & Stella, M. (2023). Cognitive network science reveals bias in GPT-3, GPT-3.5 turbo, and GPT-4 mirroring math anxiety in high-school students. *Big Data and Cognitive Computing*, 7(3), 124.
- Aitchison, Jean (2012). *Words in the mind: an introduction to the mental lexicon*. John Wiley & Sons.
- Anderson, John R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22(3), 261–295.
- Artemenko, Christina, Daroczy, Gabriella, & Nuerk, Hans-Christoph (2015). Neural correlates of math anxiety—an overview and implications. *Frontiers in Psychology*, 6, 1333.
- Ashcraft, Mark H. (2002). Math anxiety: Personal and educational, and cognitive consequences. *Current Directions in Psychological Science*, 11(5), 181–185.
- Baker, Oliver, Montefinese, Maria, Castro, Nichol, & Stella, Massimo (2023). Multiplex lexical networks and artificial intelligence unravel cognitive patterns of picture naming in people with anomia aphasia. *Cognitive Systems Research*, 79, 43–54.
- Balota, David A., & Lorch, Robert F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 12(3), 336.
- Beaty, Roger E., Kenett, Yoed N., Hass, Richard W., & Schacter, Daniel L. (2023). Semantic memory and creativity: The costs and benefits of semantic memory structure in generating original ideas. *Thinking & Reasoning*, 29(2), 305–339.
- Bowden, Edward M., & Jung-Beeman, Mark (2003). Normative data for 144 compound remote associate problems. *Behavior Research Methods, Instruments, & Computers*, 35, 634–639.
- Bowers, Kenneth S., Regehr, Glenn, Balthazard, Claude, & Parker, Kevin (1990). Intuition in the context of discovery. *Cognitive Psychology*, 22(1), 72–110.
- Bradlow, Ann R, & Alexander, Jennifer A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, 121(4), 2339–2349.
- Brian, Kieran, & Stella, Massimo (2023). Introducing mindset streams to investigate stances towards stem in high school students and experts. *Physica A. Statistical Mechanics and its Applications*, 626, Article 129074.
- Burke, Deborah M., MacKay, Donald G., Worthley, Joanna S., & Wade, Elizabeth (1991). On the tip of the tongue: what causes word finding failures in young and older adults? *Journal of Memory and Language*, 30(5), 542–579.
- Castro, Nichol, & Siew, Cynthia S. Q. (2020). Contributions of modern network science to the cognitive sciences: Revisiting research spirals of representation and process. *Proceedings: Mathematical, Physical and Engineering Sciences*, 476(2238), 1–25.
- Castro, Nichol, & Stella, Massimo (2019). The multiplex structure of the mental lexicon influences picture naming in people with aphasia. *Journal of Complex Networks*, 7(6), 913–931.
- Ciringione, Luciana, Reigl, Simone, D'Onofrio, Isaia, Poquet, Oleksandra, Gabriel, Florence, & Stella, Massimo (2024). Math anxiety and associative knowledge structure are entwined in psychology undergraduate students but not in large language models.
- Collins, Alan F., & Ellis, Andrew W. (1992). Phonological priming of lexical retrieval in speech production. *British Journal of Psychology*, 83(3), 375–388.
- Collins, Allan M., & Loftus, Elizabeth F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407.
- Collins, Allan M., & Quillian, M. Ross (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8(2), 240–247.
- Coltheart, Max, Besner, Derek, Jonasson, Jon Torfi, & Davelaar, Eileen (1979). Phonological encoding in the lexical decision task. *The Quarterly Journal of Experimental Psychology*, 31(3), 489–507.
- Crestani, Fabio (1997). Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review*, 11, 453–482.
- Crestani, Fabio, & Lee, Puay Leng (2000). Searching the web by constrained spreading activation. *Information Processing & Management*, 36(4), 585–605.
- Dell, Gary S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283.
- Dell, Gary S., & O'Seaghdha, Padraig G. (1991). Mediated and convergent lexical priming in language production: A comment on levelt et al. *Psychological Review*, 98(4), 604–614.
- Dell, Gary S., Schwartz, Myrna F., Martin, Nadine, Saffran, Eleanor M., & Gagnon, Deborah A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104(4), 801.
- Deyne, Simon De, Navarro, Danielle J., Perfors, Amy, Brysbaert, Marc, & Storms, Gert (2019). The "small world of words" english word association norms for over 12,000 cue words. *Behavior research methods*, 51, 987–1006.
- Dóczy, Brigitta (2019). An overview of conceptual models and theories of lexical representation in the mental lexicon. In *The Routledge handbook of vocabulary studies* (pp. 46–65).
- Foygel, Dan, & Dell, Gary S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language*, 43(2), 182–216.
- Gangemi, Amelia, Mancini, Francesco, & van den Hout, Marcel (2012). Behavior as information: "If i avoid, then there must be a danger". *Journal of Behavior Therapy and Experimental Psychiatry*, 43(4), 1032–1038.

- Gomez, Sergio, Diaz-Guilera, Albert, Gomez-Gardenes, Jesus, Perez-Vicente, Conrad J., Moreno, Yami, & Arenas, Alex (2013). Diffusion dynamics on multiplex networks. *Physical Review Letters*, 110(2), Article 028701.
- Grivol, Marcia Aparecida, & de Vasconcellos Hage, Simone Rocha (2011). Phonological working memory: a comparative study between different age groups. *Jornal Da Sociedade Brasileira de Fonoaudiologia*, 23, 245–251.
- Groot, Annette M. B. De (1983). The range of automatic spreading activation in word priming. *Journal of Verbal Learning and Verbal Behavior*, 22(4), 417–436.
- Haim, Edith, & Haim, Edith (2023). Cognitive networks for knowledge modelling: A gentle tutorial for data and cognitive scientists. *PsyArXiv Preprints*.
- Harnish, Stacy M. (2018). Anomia and anomia aphasia: implications for lexical processing. In *The Oxford handbook of aphasia and language disorders* (pp. 121–144).
- Hayes, Steven C., Wilson, Kelly G., Gifford, Elizabeth V., Follette, Victoria M., & Strosahl, Kirk (1996). Experiential avoidance and behavioral disorders: A functional dimensional approach to diagnosis and treatment. *Journal of Consulting and Clinical Psychology*, 64(6), 1152.
- Hembree, Ray (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, 21(1), 33–46.
- Kenett, Yoed N., & Faust, Miriam (2019). A semantic network cartography of the creative mind. *Trends in Cognitive Sciences*, 23(4), 271–274.
- Kenett, Yoed N., Levi, Efi, Anaki, David, & Faust, Miriam (2017). The semantic distance task: Quantifying semantic distance with semantic network path length. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(9), 1470.
- Kittredge, Audrey K., Dell, Gary S., Verkuilen, Jay, & Schwartz, Myrna F. (2008). Where is the effect of frequency in word production? insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, 25(4), 463–492.
- Kohn, Susan E., & Goodglass, Harold (1985). Picture-naming in aphasia. *Brain and Language*, 24(2), 266–283.
- Koponen, Ismo T. (2021). Systemic states of spreading activation in describing associative knowledge networks ii: generalisations with fractional graph laplacians and q-adjacency kernels. *Systems*, 9(2), 22.
- Kumar, A. A., Apsel, M., Zhang, L., Xing, N., & Jones, M. N. (2024). Forager: A python package and web interface for modeling mental search. *Behavior Research Methods*, 56(6), 6332–6348.
- Lee, Christine S., Huggins, Anne Corinne, & Theriault, David J. (2014). A measure of creativity or intelligence? examining internal and external structure validity evidence of the remote associates test. *Psychology of Aesthetics, Creativity, and the Arts*, 8(4), 446.
- Levelt, Willem J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, 98(23), 13464–13471.
- Mari, Anat, Wagner, Anthony D., & Schacter, Daniel L. (2001). On the tip of the tongue: An event-related fmri study of semantic retrieval failure and cognitive conflict. *Neuron*, 31(4), 653–660.
- Marko, Martin, Michalko, Drahomír, & Riečanský, Igor (2019). Remote associates test: An empirical proof of concept. *Behavior Research Methods*, 51(6), 2700–2711.
- McClelland, James L., & Rumelhart, David E. (1981). An interactive activation model of context effects in letter perception: I. an account of basic findings. *Psychological Review*, 88(5), 375.
- McRae, Ken, Cree, George S., Seidenberg, Mark S., & McNorgan, Chris (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37(4), 547–559.
- Mednick, Sarnoff (1962). The associative basis of the creative process. *Psychological Review*, 69(3), 220.
- Mulligan, Neil W., & Peterson, Daniel (2008). Attention and implicit memory in the category-verification and lexical decision tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(3), 662.
- Newman, Mark (2018). *Networks*. Oxford University Press.
- Pirolli, Peter, Schank, Patricia, Hearst, Marti, & Diehl, Christine (1996). Scatter/gather browsing communicates the topic structure of a very large text collection. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 213–220).
- Quillian, M. Ross (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science*, 12(5), 410–430.
- Rocha, Cristiano, Schwabe, Daniel, & Aragao, Marcus Poggi (2004). A hybrid approach for searching in the semantic web. In *Proceedings of the 13th international conference on world wide web* (pp. 374–383).
- Saunders, J. O., & MacLeod, Malcolm D. (2006). Can inhibition resolve retrieval competition through the control of spreading activation? *Memory & Cognition*, 34(2), 307–322.
- Schwartz, Myrna F., Dell, Gary S., Martin, Nadine, Gahl, Susanne, & Sobel, Paula (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language*, 54(2), 228–264.
- Seidenberg, Mark S., & McClelland, James L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523.
- Semeraro, A., Vilella, S., Improta, R., Duro, E. S. De, Mohammad, S. M., Ruffo, G., & Stella, M. (2025). EmoAtlas: An emotional network analyzer of texts that merges psychological lexicons, artificial intelligence, and network science. *Behavior Research Methods*, 57(2), 77.
- Siew, Cynthia S. Q. (2013). Community structure in the phonological network. *Frontiers in Psychology*, 4, 553.
- Siew, Cynthia S. Q. (2019). Spreadr: An R package to simulate spreading activation in a network. *Behavior Research Methods*, 51(2), 910–929.
- Siew, Cynthia S. Q., & Castro, Nichol (2023). Phonological similarity judgments of word pairs reflect sensitivity to large-scale structure of the phonological lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 49(12), 1989.
- Siew, Cynthia S. Q., Wulff, Dirk U., Beckage, Nicole M., & Kenett, Yoed N. (2019). Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity*, 2019(1), Article 2108423.
- Spivey, Michael J., Grosjean, Marc, & Knoblich, Günther (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, 102(29), 10393–10398.
- Stella, Massimo (2020). Multiplex networks quantify robustness of the mental lexicon to catastrophic concept failures, aphasic degradation and ageing. *Physica A: Statistical Mechanics and its Applications*, 554, Article 124382.
- Stella, Massimo (2022). Network psychometrics and cognitive network science open new ways for understanding math anxiety as a complex system. *Journal of Complex Networks*, 10(3), cnac012.
- Stella, Massimo, Beckage, Nicole M., & Brede, Markus (2017). Multiplex lexical networks reveal patterns in early word acquisition in children. *Scientific Reports*, 7(1), 46730.
- Stella, Massimo, Beckage, Nicole M., Brede, Markus, & Domenico, Manlio De (2018). Multiplex model of mental lexicon reveals explosive learning in humans. *Scientific Reports*, 8(1), 2259.
- Stella, Massimo, Citraro, Salvatore, Rossetti, Giulio, Marinazzo, Daniele, Kenett, Yoed N., & Vitevitch, Michael S. (2024). Cognitive modelling of concepts in the mental lexicon with multilayer networks: Insights, advancements, and future challenges. *Psychonomic Bulletin & Review*, 1–24.
- Stella, Massimo, Sarah De Nigris, Aleksandra Aloric, & Siew, Cynthia S. Q. (2019). Forma mentis networks quantify crucial differences in stem perception between students. *PloS One*, 14(10), Article e0222870.
- Vigliocco, Gabriella, & Kita, Sotaro (2006). Language-specific properties of the lexicon: Implications for learning and processing. *Language and Cognitive Processes*, 21(7–8), 790–816.
- Vitevitch, Michael S. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language, and Hearing Research*, 51(2), 408–422.
- Vitevitch, Michael S., & Aljasser, Faisal M. (2021). Phonotactics in spoken-word recognition. In *The handbook of speech perception* (pp. 286–308).
- Vitevitch, Michael S., Luce, Paul A., Charles-Luce, Jan, & Kemmerer, David (1997). Phonotactics and syllable stress: implications for the processing of spoken nonsense words. *Language and Speech*, 40(1), 47–62.
- Ševcech, Jakub, Mořo, R., Holub, Michal, & Bieliková, Mária (2014). User annotations as a context for related document search on the web and digital libraries. *Informatica*, 38(1).
- Wilshire, Carolyn E. (2002). Where do aphasic phonological errors come from? evidence from phoneme movement errors in picture naming. *Aphasiology*, 16(1–2), 169–197.
- Wu, Ching-Lin, Huang, Shih-Yuan, Chen, Pei-Zhen, & Chen, Hsueh-Chih (2020). A systematic review of creativity-related studies applying the remote associates test from 2000 to 2019. *Frontiers in Psychology*, 11, Article 573432.
- Yang, Jianfeng, Zevin, Jason D., Shu, Hua, McCandless, Bruce D., & Li, Ping (2006). A "triangle model" of Chinese reading. In *28th annual conference of the cognitive science society*. Mahwah, NJ.
- Yap, Melvin J., Tse, Chi-Shing, & Balota, David A. (2009). Individual differences in the joint effects of semantic priming and word frequency revealed by RT distributional analyses: The role of lexical integrity. *Journal of Memory and Language*, 61(3), 303–325.
- Zemla, J. C., Cao, K., Mueller, K. D., & Austerweil, J. L. (2020). SNAFU: The semantic network and fluency utility. *Behavior Research Methods*, 52(4), 1681–1699.