Similarity and generalization as discounted integration over higher-order paths in semantic networks

Valerio Rubino (valerio.rubino@unitn.it)[1] **& Manuela Piazza**[1]
[1]Center for Mind/Brain Science, University of Trento, 39068 Italy

Abstract

Similarity and generalization are often assumed to reflect distance in an underlying representational space. In semantic networks, however, distance can be defined by both shortest paths or by multiple indirect pathways. Previous research has shown that non-shortest, higher-order paths influence similarity judgements. Here, we extend these findings by re-analyzing a publicly available dataset to compare shortest-path models with models integrating over discounted higher-order paths. The latter better accounted for similarity judgements; error analyses indicate that this advantage arose from integrating multiple indirect paths rather than relying on shortest connections. These random-walk models implement the same core computation as the successor representation (SR), which has been proposed to support human generalization. Consistently, an SR-based model outperformed alternatives in accounting for inductive generalization judgements from the same dataset. These findings suggest that similarity and generalization are both shaped by higher-order connectivity between concepts.

Keywords: similarity; generalization; concepts; semantic networks; Successor Representation

Introduction

Evolution has shaped the human mind to perceive similarities around us, to survive in novel and uncertain environments. For instance, when facing a *zebra*, it can be crucial to determine whether it is more similar to a *horse* or to a *tiger*. Classic work by Shepard (1987) indicated that perceived similarity and, in turn, the probability of generalization between stimuli decay exponentially with their distance in an underlying psychological space. Subsequently, Nosofsky (1986) showed that this psychological space reflects a warped representation of stimulus features shaped by attentional weighting of feature dimensions. However, these studies used physical objects or artificial concepts defined in highly controlled experimental settings, and did not address how real-world semantic knowledge encoded in word meanings shapes perceived similarity and generalization across concepts.

According to a long-standing tradition, semantic knowledge is conceived as a network (or graph), in which nodes represent concepts and edges encode associations between them. Within this tradition, the original model by Collins and Quillian (1969) proposed that concept nodes are linked by class–inclusion relations, yielding a largely taxonomic network structure. More recent proposals, based on data from large-scale free association paradigms in which participants produce words in response to a cue model semantic knowledge as a network of graded associations of multiple kinds (Steyvers & Tenenbaum, 2005). This raises the question of how distance should be defined on such semantic graphs, and how well different notions of distance account for similarity judgements and generalization between real-world concepts.

Measures of global connectivity on semantic graphs

Semantic networks afford multiple graph-theoretic conceptualizations of semantic similarity between their concept nodes. In an influential paper, Collins and Loftus (1975) distinguished two notions of semantic closeness: one based on the shortest path connecting two concepts, and one based on their aggregate connectivity across all possible paths. Under the former view, similarity is determined by the most direct associative route between concepts; under the latter, it reflects the number and strength of longer, indirect (higher-order) pathways linking them. For instance, the semantic similarity between the words *zebra* and *tiger* might be determined solely by their direct associative strength. Alternatively, it may reflect the integration of multiple longer, indirect paths, such as *zebra-stripes-tiger* or *zebra-safari-tiger*. Building on the former, Kenett et al. (2017) defined similarity between concept nodes as the correlation between their shortest-path distances to all other nodes in the semantic network, and showed that this measure predicts both semantic relatedness judgements and priming effects. Subsequently, Kumar et al. (2020) demonstrated that shortest-path distance itself provides a better account of priming than correlations over shortest-path distances. In contrast, De Deyne et al. (2016, 2018) introduced semantic similarity measures that aggregate over all possible paths in the network, inspired by Katz centrality (Aprahamian et al., 2015). This random-walk–based measure sums over all paths connecting two nodes, weighting each path's contribution by its length (i.e., number of steps), and admits a closed-form expression (assuming a random policy):

$$G_{RW} = \sum_{k=0}^{\infty} \alpha^k T^k = (I - \alpha T)^{-1}, \tag{1}$$

where T is the transition matrix of a random walk on the semantic graph, such that T_{ij} denotes the probability of transitioning from concept node i to node j in a single step, and $\alpha \in (0, 1)$ is a discount parameter that downweights longer, indirect paths. Each row of G_{RW} thus represents the discounted probability distribution over nodes reachable from

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

a given starting concept. De Deyne et al. (2016) defined semantic similarity as the cosine similarity between rows of G_{RW} and showed that this measure accounts for *triadic preference judgements*, in which participants are presented with three concepts and select the most similar pair. This random-walk-based similarity measure also predicts pairwise similarity judgements better than alternative models based on co-occurrence in text corpora (De Deyne et al., 2017, 2018).

Separately, studies based on shortest-path distance (Kenett et al., 2017; Kumar et al., 2020) and Katz-style random-walk measures (De Deyne et al., 2016, 2017, 2018) showed that semantic similarity depends on global network structure. However, these approaches have not been directly compared on the same behavioural dataset. De Deyne et al. (2016), in particular, showed that similarity ratings are influenced by non-shortest indirect paths. This finding predicts that models integrating over multiple indirect paths should outperform models that rely only on the shortest route between concepts, but this prediction has not been directly tested against explicit shortest-path alternatives.

Generalizing with Successor Representation

The random-walk–based semantic similarity measure introduced by De Deyne et al. (2016, 2017, 2018) is equivalent to the Successor Representation (SR; Dayan, 1993; Gershman, 2018), as both quantify the expected discounted visitation of nodes under a random walk on a graph. Models based on the SR have been shown to capture how human participants learn and use transition structure to predict and control future outcomes (Momennejad et al., 2017), navigate hierarchically organized virtual environments (Wientjes & Holroyd, 2024), and acquire the structure of social networks (Son et al., 2024).

According to a recent proposal, these empirical successes reflect the fact that human generalization relies on SR or related diffusion-based models (Wu et al., 2025). These models capture higher-order graph structure by integrating information over multiple indirect paths between states. In controlled graph-structured tasks, such as subway prediction and graph-armed bandit problems, these models successfully capture how participants generalize across stimuli (Wu et al., 2019, 2020). However, it remains unclear whether higher-order connectivity similarly predicts generalization in more naturalistic settings, where generalization depends on semantic similarity between concepts expressed by symbols (i.e. words).

Goal and scope

The Katz-style random-walk measures used to define similarity in semantic networks are closely related to the SR, which has been proposed to support generalization over structured state spaces. Both approaches integrate over discounted higher-order paths through a transition structure. This formal connection suggests that similarity and generalization may rely on a common mechanism: the discounted integration of indirect paths.

In the present study, we use a model comparison approach

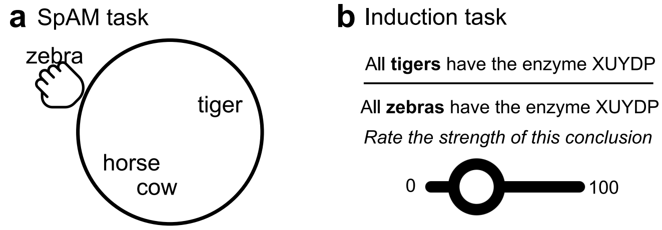

Figure 1: **a** Illustration of the SpAM task, where participants had to drag and drop words in a two-dimensional arena according to their semantic similarity. **b** Illustration of the inductive task, where participants rated the strength of inductive arguments.

to test whether semantic similarity and generalization are better characterized by distance defined in terms of shortest paths on a semantic network or by discounted integration over higher-order paths. To do so, we re-analyze a publicly available dataset comprising similarity and inductive generalization judgements from the same participants (Douven et al., 2023) and compare models that instantiate these competing notions of distance. Consistent with this prediction, our results show that a model based on discounted integration over multi-step paths provides a better account of both similarity and generalization judgements than shortest-path–based alternatives. More broadly, they point toward a potential integration between network-based accounts of semantic knowledge and SR-based models of predictive representations.

Methods

Materials

Dataset We re-analyzed the dataset from Douven et al. (2023), which includes semantic similarity judgements and inductive strength ratings from the same participants. Inductive ratings were missing for one participant, who was therefore excluded from the analyses, leaving a final sample of $N = 58$ participants. Similarity judgements were collected with the Spatial Arrangement Method (SpAM; Fig. 1a; Goldstone, 1994; Hout et al., 2013). In the SpAM task, participants arranged 20 animal words in a 2D arena such that spatial proximity reflected semantic similarity. For each participant, we first normalized item coordinates by screen width and height, then computed pairwise Euclidean distances between items. Distances D were converted to similarity using an inverse-distance transform:

$$S_{\mathrm{behav}}(i, j) = \frac{1}{D(i, j) + \epsilon}, \tag{2}$$

where $\epsilon = 10^{-12}$ is a small numerical constant to prevent similarity values from becoming arbitrarily large when distances are very small.

In the induction task (Fig. 1b), participants evaluated the strength of similarity-based inductive arguments adapted from Douven et al. (2023). Each argument consisted of a premise stating that all animals from a species share a certain property (e.g., *All tigers have the enzyme XUYDP*), followed by a conclusion attributing the same property to a different animal species (e.g., *All zebras have the enzyme XUYDP*). On a scale from 0 to 100, participants rated how strongly the conclusion followed from the premise for each argument. In this task, inductive strength ratings are known to depend primarily on the perceived similarity between the premise and conclusion animal species, with stronger generalization of the property to more similar animals (Marjieh et al., 2024; Osherson et al., 1990; Rips, 1975).

SWOW We constructed a semantic graph using free association norms from the Small World of Words (SWOW) project, which comprises more than 12,000 cue words and the corresponding elicited responses from over 90,000 participants (De Deyne et al., 2018). The semantic graph consists of concept nodes connected by weighted edges: two concepts are linked if one was produced as an associate of the other, and edge weights correspond to association strength (i.e., the normalized response frequency). Following Abramski et al. (2025), we lowercased all words, corrected spelling errors, and pruned rare associations and concepts not present in WordNet. We made the graph undirected by symmetrizing edge weights between concept pairs and then defined its transition matrix T by row-normalizing the resulting adjacency matrix, so that each row sums to one.

To model the current task using this transition structure, we extracted the nodes corresponding to the 20 mammals presented to participants by Douven et al., 2023. The resulting subgraph was not fully connected, and was therefore augmented using a k-nearest-neighbor expansion procedure. For each target concept, we added its top-k associated concepts (based on symmetrized association strength). The value of k was increased incrementally until all target concepts belonged to a single connected component. This procedure yielded a fully connected semantic graph with $n = 111$ nodes.[1] We then merged obvious morphological variants (*striped* and *stripe*), spelling variants (*shepard* and *shepherd*), and gendered versions of the same animal (*lion* and *lioness*), resulting in a graph with $n = 108$ concept nodes.

Computational modeling

Semantic similarity models We defined three models of semantic similarity, differing in how they incorporate the higher-order structure of the reduced SWOW graph. The first is the shortest-path distance (SP) model, which quantifies distance as the hop distance, defined as the unweighted shortest-path length between two concept nodes, that is, the minimum number of edges separating them, irrespective of edge weights, following prior work on semantic network distance (Kumar et al., 2020). Unlike earlier formulations that use hop distance directly, we introduce an explicit decay parameter α_{SP}, governing how rapidly similarity decreases with increasing hop distance. The model defines similarity between two concepts i and j as:

$$\hat{S}_{SP}(i, j) = \exp(-\alpha_{SP} d_A(i, j)) \tag{3}$$

where $d_A(i, j)$ is the minimum number of graph edges between the concepts i and j, and is computed by finding their shortest-path distance on a binarized version of the semantic graph adjacency matrix. The SP model is a direct analogue of exponential models of similarity in space, usually taking the form of $\exp(-\alpha d(i, j))$, where $d(i, j)$ is either Euclidean or Manhattan distance[2]. The second is the shortest weighted path (SWP) model, which defines similarity as:

$$\hat{S}_{SWP}(i, j) = \exp(-\alpha_{SWP} D_T) \tag{4}$$

where D_T is the cost of the shortest path between the two nodes, computed as the minimum summed cost along any path, with each edge cost defined as the negative logarithm of its transition probability. The decay of similarity with shortest weighted path distance is controlled by the free parameter α_{SWP}.

Lastly, we used an SR model, defining similarity from the discounted expected future occupancy of states under the transition matrix T (Dayan, 1993; Gershman, 2018). The SR is defined as

$$M = (I - \gamma T)^{-1}, \tag{5}$$

where $\gamma \in (0, 1)$ controls the contribution of multi-step transitions. Because SpAM-derived similarity judgements are symmetric, we derived symmetric similarity estimates by first row-normalizing the SR matrix and then symmetrizing it (by averaging the SR matrix with its transpose). SR values were subsequently log-transformed.

The three models differ in how they incorporate indirect paths in the semantic network (Fig. 2a), but they all predict a decay of similarity with increasing graph distance (Fig. 2b), governed by their respective decay parameters α_{SP}, α_{SWP}, and γ. Intuitively, the SR captures how likely a random walk starting from one concept is to visit another over time, with more distant, indirect transitions contributing less due to discounting.

Model fitting and evaluation Similarity predictions for each pair of concepts were calibrated to each participant's SpAM-derived behavioural similarity through a linear transformation, $S^\star = \beta_0 + \beta_1 X$, where β_0 and β_1 capture the offset and scaling of the predicted similarities, respectively. To fit the three models to behavioural similarity judgements, we estimated model parameters using a two-step procedure separately for each participant. Because all pairwise measures were symmetrized, analyses were conducted on the upper triangle of the similarity matrix (i.e., one observation per unordered concept pair). Each of the SP, SWP, and SR models

[1] To rule out the possibility that results were driven by this particular graph size, we replicated all analyses with augmented graphs of size $n = 250$, $n = 500$ and $n = 1000$, obtaining qualitatively identical results.

[2] For a lattice graph, the shortest-path distance $d_A(i, j)$ corresponds to the Manhattan distance between nodes i and j.

included a single free parameter (respectively, α_{SP}, α_{SWP}, or γ) controlling the decay of similarity as a function of path distance or higher-order semantic structure. For each model, we first fit this decay parameter by maximizing Pearson's r between model predictions and behavioural similarity. With the decay parameter fixed, we then estimated calibration parameters β_0 and β_1 via ordinary least squares (OLS). Model parameters were fit by maximizing correlation with observed similarity judgements, as this measure is invariant to scale and captures rank-order structure. Model performance was then evaluated using cross-validated root mean squared error (cv-RMSE) to assess absolute predictive accuracy. For cross-validation, the twenty concepts were divided into five folds (four concepts each): models were fit on pairs comprising training concepts only and evaluated on pairs with one training and one held-out test concept.

We then used the same similarity models to predict inductive strength judgements. Strength ratings were linearly rescaled to the open interval (0, 1) to be modelled with beta regression. The decay parameters were fixed at the values estimated from the similarity task, and a beta regression was fit to link model-predicted similarity to induction ratings. Predictive accuracy was assessed using 5-fold cross-validation, with the 30 induction judgements divided into five folds (six judgements each). For each iteration, models were fit on four training folds and evaluated on the held-out judgements to compute cv-RMSE.

Results

SR best predicts semantic similarity

Model comparison We compared Pearson's r correlations between each model's predicted similarities and SpAM-derived semantic similarity ratings (Fig. 2a). Across participants, the SR model achieved the highest correlation (mean $r = 0.58 \pm 0.02$), significantly outperforming both the SP model ($r = 0.46 \pm 0.02$; paired two-tailed t-test on Fisher z-transformed coefficients: $t(57) = 8.5$, $p < .001$, $d = 1.12$) and the SWP model ($r = 0.50 \pm 0.02$; $t(57) = 11.5$, $p < .001$, $d = 1.5$). The SR model also achieved significantly lower cv-RMSE than both SP ($t(57) = -5.8$, $p < .001$, $d = -0.76$) and SWP ($t(57) = -8.1$, $p < .001$, $d = -1.06$). The difference between SP and SWP was not significant ($t(57) = -1.5$, $p = .139$, $d = 0.02$).

Residuals analyses identify concept pairs best predicted by SR We compared residuals of SR and SWP, which achieved the second-best correlation with behavioural similarity. We quantified the advantage of SR as the difference in residuals between SR and SWP, and examined how this difference varied with hop distance on the unweighted graph (Fig. 3a). To do so, we used mixed-effects linear regression; to account for non-independence in the data, we included random intercepts for participants and for each concept node (entered separately for both elements of each pair), thereby controlling for repeated observations within participants and dependencies arising from shared concepts across pairs. We found

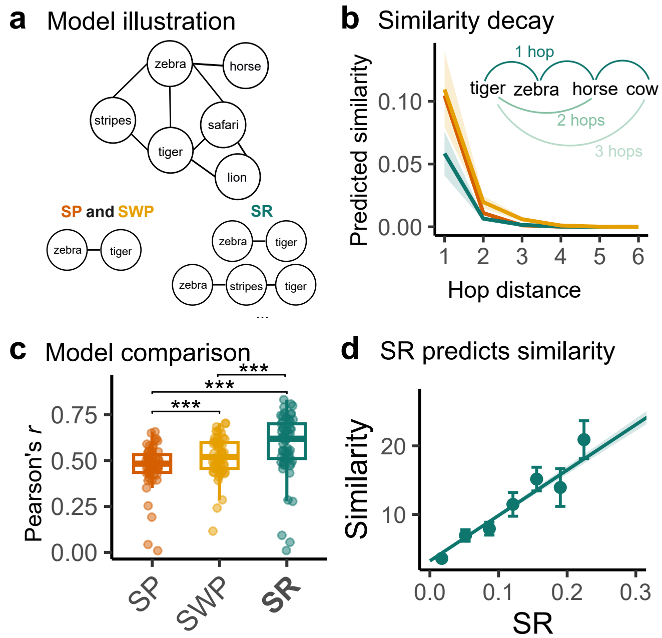

Figure 2: **a** Illustration of how the three models derive similarity on a small subgraph of the semantic network: SP and SWP models depend only on the shortest route between nodes, whereas SR integrates over multiple indirect paths. **b** Predicted similarity vs. hop distance (number of edges between nodes). **c** Correlation between model predictions and SpAM similarity (paired two-tailed t-tests on Fisher z; $*p < .05$, $**p < .01$, $***p < .001$). **d** Predictions from the best-fitting model (SR).

that SR's advantage decreased with unweighted hop distance ($\beta = -0.36$, 95% CI $[-0.39, -0.34]$, $p < .001$; Fig. 3b). Crucially, this reduction in SR's advantage was not driven by a loss of accuracy for more distant concept pairs: on the contrary, SR absolute prediction error ($|S_{\text{SR}} - S_{\text{behav}}|$) decreased with hop distance (linear mixed-effects model: $\beta = -1.17$, $[-1.27, -1.06]$, $p < .001$). Instead, SR's diminished advantage at larger hop distances reflected increasing convergence between SR and shortest weighted path predictions: the correlation between the two models rose monotonically with hop distance (Fig. 3c).

In addition to this linear effect, SR's advantage was larger for directly connected concept pairs (hop=1; $\beta = 0.20$, $[0.13, 0.28]$, $p < .001$), and including a binary hop=1 predictor significantly improved model fit ($\chi^2(1) = 28.31$, $p < .001$). Restricting the analysis to directly connected pairs, SR's advantage increased with (mean-centered) average two-step connectivity between nodes (mean $(T_{ij} + T_{ji})/2$; $\beta = 1.19$, $[0.81, 1.57]$, $p < .001$), while the effect of direct connectivity strength was not significant ($\beta = -0.24$, $[-0.53, 0.06]$, $p = .120$)[3]. Thus, SR captures the influence of indirect two-step paths that shortest-path models overlook.

[3]Although the two regressors are correlated, collinearity was low (variance inflation factor = 1.89).

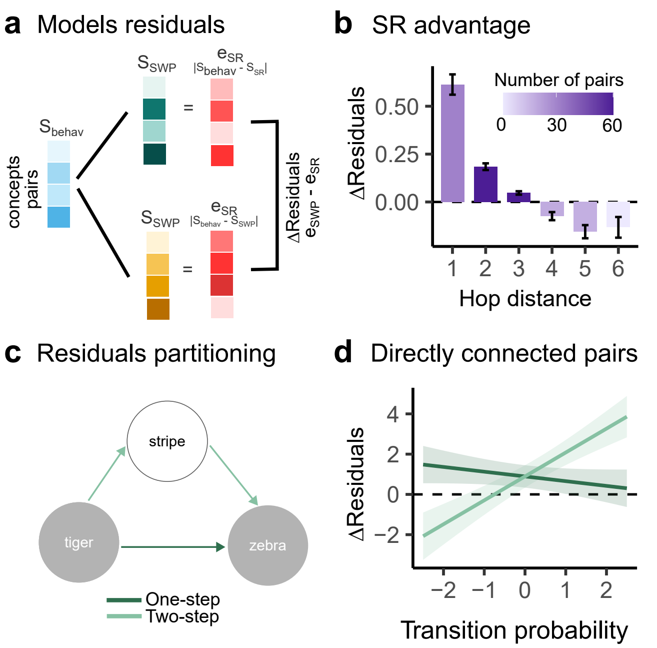

Figure 3: **a** Schematic illustration of the residual analysis procedure, comparing prediction errors of the SR and SWP models for individual concept pairs. **b** Advantage of the SR model over SWP decreases with hop distance. **c** Illustration of direct one-step and indirect two-step (mediated by another node) connections between concept nodes. **d** Residual advantage of SR for directly connected concept pairs (hop=1) decreases with one-step transition probability and increases with two-step transition probability.

Accordingly, when controlling for the two-step connectivity across all pairs, the positive hop=1 effect reversed in direction ($\beta = -0.20$, $[-0.26, -0.14]$, $p < .001$). Thus, the apparent hop advantage is fully explained by the contributions of indirect paths.

Similarity-fitted SR transfers to inductive generalization

In the dataset from Douven et al., 2023, participants also rated the strength of 30 inductive arguments. We tested whether the three models fitted to behavioural similarity generalize to these induction ratings. Specifically, we fixed each model's decay parameter to the value fit on the similarity task and evaluated predictive accuracy with 5-fold cross-validated beta regression (Fig. 4a).

Across participants, the SR model achieved the lowest test error (cv-RMSE = 3.97), and outperformed both SP ($t(57) = -2.8$, $p = .006$, $d = 0.37$) and SWP ($t(57) = -3.4$, $p = .001$, $d = 0.44$). Instead, we found no significant difference between SP and SWP predictive accuracy ($t(57) = 1.2$, $p = .231$, $d = 0.16$).

To test the specificity of the discount parameter γ, we reversed the value fitted to each participant in the similarity task and recomputed SR predictions using $\gamma_{\text{induction}} = 1 - \gamma_{\text{similarity}}$: if γ captures participant-specific discounting of indirect structure, reversing it should impair transfer. We then re-evaluated the reversed-γ SR model using the same cross-validation procedure. This manipulation increased test error (one-tailed t-test: $t(57) = 1.74$, $p = .044$; Fig. 4b). Although small, this effect indicates that the predictive benefit of SR arises from carrying over participant-specific weighting of multi-step semantic structure from similarity judgements to inductive generalization.

Discussion

Semantic knowledge guides our choices and structures our experience through a web of interconnected concepts. Links between concepts are commonly measured using free association paradigms, in which response frequencies provide an index of associative strength (Deese, 1962). These norms motivate semantic network models, which represent concepts as nodes and associations as weighted edges (Baronchelli et al., 2013; Borge-Holthoefer & Arenas, 2010b; Hills & Kenett, 2021; Steyvers & Tenenbaum, 2005). Such models account for a wide range of cognitive phenomena, including memory search (Abbott et al., 2015), semantic priming (Kenett et al., 2017; Kumar et al., 2020), word acquisition (Stella et al., 2017, 2018), and categorization (Borge-Holthoefer & Arenas, 2010a). Importantly, prior work has emphasized that semantic processing depends on the global structure of semantic networks: priming and similarity are shaped not only by direct associations, but also by indirect, multi-step paths mediated by intermediate concepts (De Deyne et al., 2016, 2017, 2018; Kenett et al., 2017; Kumar et al., 2020; Nelson et al., 1997).

The present study sought to identify which aspects of indirect connectivity in semantic networks shape judged similarity and inductive generalization. Across analyses, an SR-based model that integrates information over multiple paths outperformed shortest-path–based alternatives. Residual analyses showed that SR's advantage diminished with hop distance as model predictions converged for more distant concept pairs.

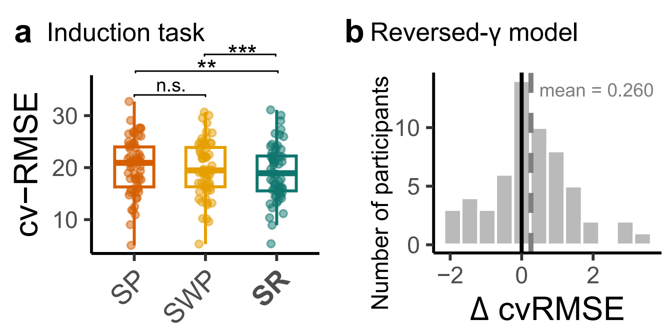

Figure 4: **a** Predictive performance (cv-RMSE) on generalization judgements of the three models using decay parameters fitted on similarity data. **b** Distribution of changes in predictive accuracy when reversing the discount parameter fitted on similarity data of each participant ($1 - \gamma$).

Furthermore, SR's advantage was disproportionately larger for directly connected concept nodes; this effect was driven by the presence of indirect two-step associations between directly connected nodes, which SWP and SP models overlook because such paths are not the shortest. Together, these findings suggest that models based on shortest-path distance (Kenett et al., 2017; Kumar et al., 2020) provide a less complete account of human similarity judgements than models that integrate information across multiple paths in semantic networks (De Deyne et al., 2016, 2017, 2018).

Previous work has modelled inductive strength using Bayesian inference on structured hypothesis spaces (Kemp & Tenenbaum, 2009). Our results suggest that such computational accounts can be complemented by SR, which offers a neurally plausible mechanism for learning and exploiting structured relationships (Wu et al., 2025). Indeed, closely related diffusion-based measures have been shown to capture how humans generalize values across nodes in controlled graph-structured tasks (Wu et al., 2019, 2020). The present findings extend this line of work by showing that related mechanisms also support generalization over naturalistic stimuli embedded in semantic knowledge.

The predictive accuracy of the SR-based model on inductive judgements depended, in part, on the transfer of the discount parameter γ fitted to similarity data. This relationship suggests that participants rely on comparable multi-step discounting of semantic structure when judging similarity and when generalizing properties across concepts. However, the dependence on the fitted γ was limited, likely due to a relatively flat correlation landscape during model fitting on similarity data itself. These results are consistent with the idea that discounted multi-step integration supports both similarity and generalization, although the weak identifiability of the parameter prevents drawing conclusions about whether the same discount parameter underlies both tasks within participants.

Overall, these findings suggest that semantic similarity and generalization depend on higher-order, multi-step connectivity in semantic networks. Although our study does not adjudicate between different models of how such connectivity is integrated, the SR provides a neurally plausible account of this process. This interpretation dovetails with proposals that SR-like mechanisms, although originally evolved for spatial navigation, are co-opted to support semantic cognition (Haga et al., 2025), and suggests a potential integration between network-based accounts of semantic knowledge and SR-based theories of predictive representation.

Limitations and future directions

Our results dovetail with prior work demonstrating that global, distant connectivity in semantic networks shapes similarity judgements (De Deyne et al., 2017, 2018; Kenett et al., 2017; Kumar et al., 2020) and also shows that integrating multiple higher-order paths is necessary to account for both similarity and generalization judgements. However, these findings are subject to several limitations, which in turn suggest promising directions for future research.

First, the similarity data re-analyzed from Douven et al. (2023) were collected using a single SpAM trial in which all 20 concepts were arranged in a two-dimensional space, which can distort nearest-neighbor structure (Tversky & Hutchinson, 1986). However, spatial embeddings impose assumptions that are known to be violated by perceived similarity (Tversky, 1977) and word association structure (Griffiths et al., 2007), for instance by implying symmetry, while human judgements are often asymmetric. In fact, network-based measures naturally capture asymmetry (Kumar et al., 2020), but addressing this issue requires collecting non-spatial similarity data, such as pairwise ratings (also obviating the need to transform Euclidean distances into similarity).

Second, we used inductive judgements to investigate generalization. In contrast to previous studies in which participants generalized learned reward values across graph-structured states (Wu et al., 2018, 2019, 2020), the present task did not involve rewards and instead directly probed generalization with argument strength ratings. Nevertheless, these ratings may encourage participants to treat the task as a similarity judgement; future work could employ paradigms that more clearly dissociate generalization from similarity, such as identification and classification (Shepard et al., 1961).

Third, the inductive generalization analysis was based on a relatively small number of argument pairs, which limits the stability of cross-validation estimates and calls for caution in interpreting predictive performance.

Fourth, parameter identifiability was limited, and inductive predictions showed weak dependence on the similarity-fitted discount parameter γ. Richer data and alternative model formulations may better capture the breadth of similarity and generalization.

Lastly, our analyses relied on a naturalistic semantic network derived from free association norms. While this choice enhances ecological validity, it may also introduce uncontrolled confounds related to word frequency, familiarity, or idiosyncratic differences across participants. Future work could assess the robustness of the present findings using alternative semantic resources or experimentally controlled learning paradigms in which network structure is manipulated directly.

Acknowledgments

AI tools were used to assist with writing analysis code. All modeling choices, analyses and interpretations were made by the authors.

References

Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2015). Random walks on semantic networks can resemble optimal foraging.

Abramski, K., Improta, R., Rossetti, G., & Stella, M. (2025). The "llm world of words" english free association norms generated by large language models. *Scientific Data*, *12*(1). https://doi.org/10.1038/s41597-025-05156-9

Aprahamian, M., Higham, D. J., & Higham, N. J. (2015). Matching exponential-based and resolvent-based centrality measures. *Journal of Complex Networks*, *4*(2), 157–176. https://doi.org/10.1093/comnet/cnv016

Baronchelli, A., Ferrer-i-Cancho, R., Pastor-Satorras, R., Chater, N., & Christiansen, M. H. (2013). Networks in cognitive science. *Trends in cognitive sciences*, *17*(7), 348–360.

Borge-Holthoefer, J., & Arenas, A. (2010a). Categorizing words through semantic memory navigation. *The European physical journal B*, *74*(2), 265–270.

Borge-Holthoefer, J., & Arenas, A. (2010b). Semantic networks: Structure and dynamics. *Entropy*, *12*(5), 1264–1302.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295x.82.6.407

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, *8*(2), 240–247.

Dayan, P. (1993). Improving generalization for temporal difference learning: The successor representation. *Neural Computation*, *5*(4), 613–624. https://doi.org/10.1162/neco.1993.5.4.613

De Deyne, S., Navarro, D. J., Perfors, A., & Storms, G. (2016). Structure at every scale: A semantic network account of the similarities between unrelated concepts. *Journal of Experimental Psychology: General*, *145*(9), 1228–1254. https://doi.org/10.1037/xge0000192

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2018). The "small world of words" english word association norms for over 12,000 cue words. *Behavior Research Methods*, *51*(3), 987–1006. https://doi.org/10.3758/s13428-018-1115-7

De Deyne, S., Perfors, A., & Navarro, D. J. (2017). Predicting human similarity judgments with distributional models: The value of word associations. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 4806–4810. https://doi.org/10.24963/ijcai.2017/671

Deese, J. (1962). On the structure of associative meaning. *Psychological review*, *69*(3), 161.

Douven, I., Verheyen, S., Elqayam, S., Gärdenfors, P., & Osta-Vélez, M. (2023). Similarity-based reasoning in conceptual spaces. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1234483

Gershman, S. J. (2018). The successor representation: Its computational logic and neural substrates. *Journal of Neuroscience*, *38*(33), 7193–7200.

Goldstone, R. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers*, *26*(4), 381–386.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological review*, *114*(2), 211.

Haga, T., Oseki, Y., & Fukai, T. (2025). A unified neural representation model for spatial and conceptual computations. *Proceedings of the National Academy of Sciences*, *122*(11), e2413449122.

Hills, T. T., & Kenett, Y. N. (2021). Is the mind a network? maps, vehicles, and skyhooks in cognitive network science. *Topics in Cognitive Science*, *14*(1), 189–208. https://doi.org/10.1111/tops.12570

Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of spam: A fast, efficient, spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General*, *142*(1), 256–281.

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological review*, *116*(1), 20.

Kenett, Y. N., Levi, E., Anaki, D., & Faust, M. (2017). The semantic distance task: Quantifying semantic distance with semantic network path length. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(9), 1470–1489. https://doi.org/10.1037/xlm0000391

Kumar, A. A., Balota, D. A., & Steyvers, M. (2020). Distant connectivity and multiple-step priming in large-scale semantic networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(12), 2261–2276. https://doi.org/10.1037/xlm0000793

Marjieh, R., Jacoby, N., Peterson, J. C., & Griffiths, T. L. (2024). The universal law of generalization holds for naturalistic stimuli. *Journal of Experimental Psychology: General*, *153*(3), 573–589. https://doi.org/10.1037/xge0001533

Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., & Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature Human Behaviour*, *1*(9), 680–692. https://doi.org/10.1038/s41562-017-0180-8

Nelson, D. L., Bennett, D. J., & Leibert, T. W. (1997). One step is not enough: Making better use of association norms to predict cued recall. *Memory & Cognition*, *25*(6), 785–796.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*(2), 185–200. https://doi.org/10.1037/0033-295x.97.2.185

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 665–681. https://doi.org/10.1016/s0022-5371(75)80055-7

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychologi-*

cal Monographs: General and Applied, *75*(13), 1–42. https://doi.org/10.1037/h0093825

Son, J.-Y., Vives, M.-L., Bhandari, A., & FeldmanHall, O. (2024). Replay shapes abstract cognitive maps for efficient social navigation. *Nature Human Behaviour*, *8*(11), 2156–2167. https://doi.org/10.1038/s41562-024-01990-w

Stella, M., Beckage, N. M., & Brede, M. (2017). Multiplex lexical networks reveal patterns in early word acquisition in children. *Scientific reports*, *7*(1), 46730.

Stella, M., Beckage, N. M., Brede, M., & De Domenico, M. (2018). Multiplex model of mental lexicon reveals explosive learning in humans. *Scientific reports*, *8*(1), 2259.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, *29*(1), 41–78.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352. https://doi.org/10.1037/0033-295x.84.4.327

Tversky, A., & Hutchinson, J. W. (1986). Nearest neighbor analysis of psychological spaces. *Psychological Review*, *93*(1), 3–22. https://doi.org/10.1037//0033-295x.93.1.3

Wientjes, S., & Holroyd, C. B. (2024). The successor representation subserves hierarchical abstraction for goal-directed behavior (L. J. Graham, Ed.). *PLOS Computational Biology*, *20*(2), e1011312. https://doi.org/10.1371/journal.pcbi.1011312

Wu, C. M., Meder, B., & Schulz, E. (2025). Unifying principles of generalization: Past, present, and future. *Annual Review of Psychology*, *76*(1), 275–302.

Wu, C. M., Schulz, E., & Gershman, S. J. (2019). Generalization as diffusion: Human function learning on graphs. https://doi.org/10.1101/538934

Wu, C. M., Schulz, E., & Gershman, S. J. (2020). Inference and search on graph-structured spaces. *Computational Brain&; Behavior*, *4*(2), 125–147. https://doi.org/10.1007/s42113-020-00091-x

Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D., & Meder, B. (2018). Generalization guides human exploration in vast decision spaces. *Nature Human Behaviour*, *2*(12), 915–924. https://doi.org/10.1038/s41562-018-0467-4