Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

What can quantitative measures of semantic distance tell us about creativity?

Yoed N Kenett

Semantic distance plays an important role in the creative process: The farther one 'moves away' from a conventional idea, the more creative the new idea will likely be. Although intuitive, the role of semantic distance in creativity has been only indirectly examined due to the challenge of its measurement. Recent studies have started applying quantitative measures of semantic distance in creativity research. Such studies complement standard subjective measures of creativity; provide objective measures of the creative output; and also allow to more directly examine the role of semantic memory, and distance, in creativity. An overview of the main approaches that are being used will be described and the advantages of using such quantitative measures in creativity will be discussed.

Address
Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA

Corresponding author: Kenett, Yoed N (yoedk@sas.upenn.edu)

Current Opinion in Behavioral Sciences 2018, **27**:11–16

This review comes from a themed issue on **Creativity**

Edited by **Hikaru Takeuchi** and **Rex Jung**

https://doi.org/10.1016/j.cobeha.2018.08.010

Introduction

The role of semantic distance in creativity is intuitively embedded in a theory of creativity, through the notion that the farther one moves from a concept in a semantic memory space, the more novel or creative the new concept will be [1,2,3•]. For example, a key feature of divergent thinking — the widely applied method to measure creative potential — is moving away from conventional to more distant, weakly related responses [4].

While the role of connecting more distant concepts in memory in creativity is very intuitive, it is difficult to examine empirically, due to the difficulty of measuring semantic memory structure and semantic distance [5]. However, an increasing number of studies have applied computational methods to derive quantitative methods of semantic distance in relation to creativity [6,7•,8••,9••]. The aim of the current paper is to review such recent studies and to discuss how the application of quantitative measures of semantic distance in behavioral research can contribute to the ongoing investigation of the creative process.

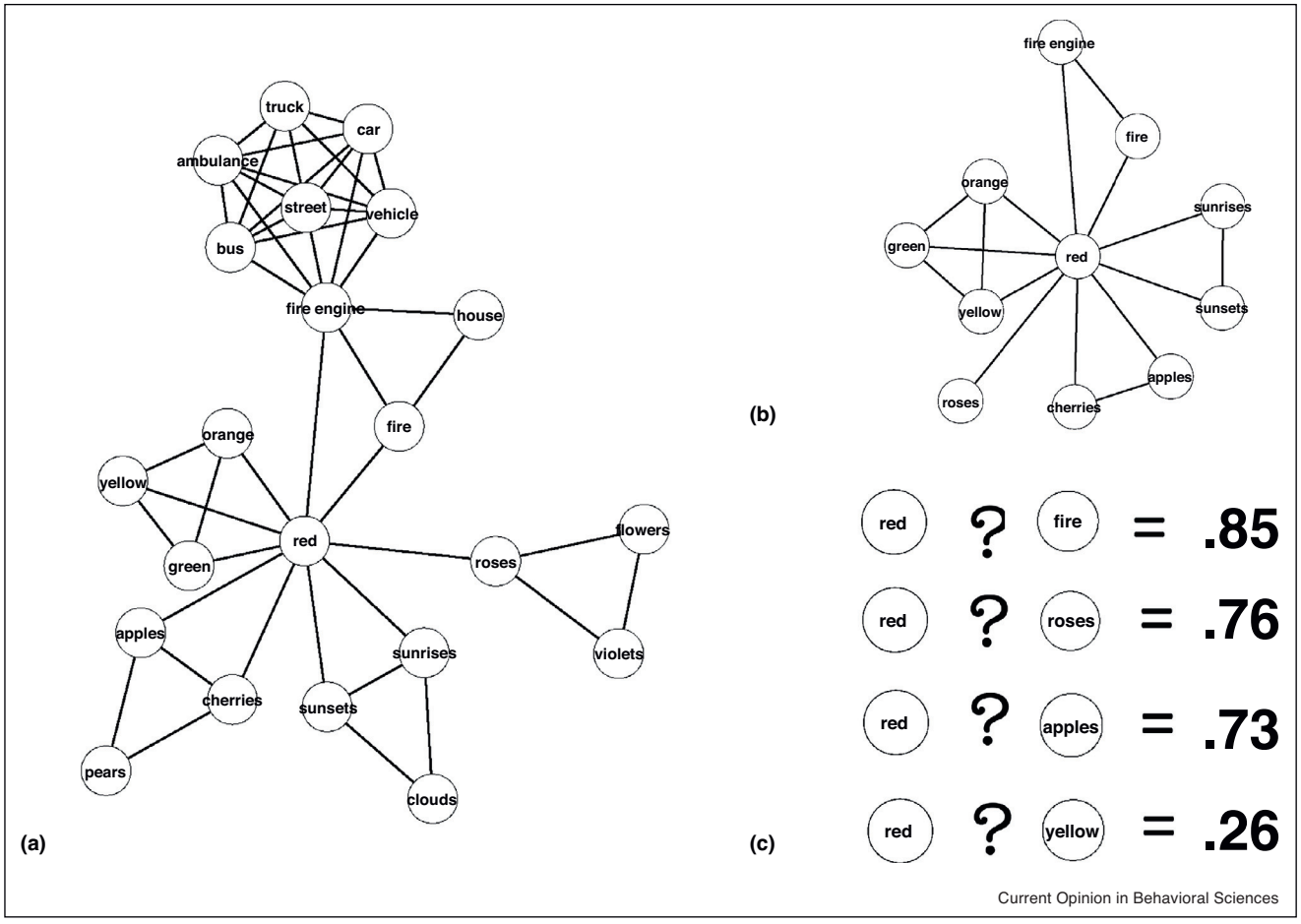
Theory

Theoretically, semantic distance plays a role in various models of the creative process. The main theory that related creative thinking to semantic memory structure is the associative theory of creativity [1]. According to this theory, creativity involves the connection of weakly related, or remote concepts into novel and applicable concepts. The farther apart the concepts are, the more creative the new combination will be. More recently, the top-down, executive functions account, has been proposed [10,11]. This account emphasizes the importance of cognitive control and executive functions, such as fluid intelligence and retrieval abilities, in the creative process [11–15]. According to this theory, cognitive control supports creativity via more effective memory retrieval and strategy implementation. A third line of relevant theories includes the dual-process theories of creativity [16–21]. These theories attempt to formalize the creative process based on its two main components — novelty and applicability: A generation stage that divergently generates multiple ideas and an evaluation stage that prunes these ideas.

In all these theories of creativity, semantic memory plays a role. Its role is either direct — varying degree of organization of semantic memory facilitating uncommon, weakly related, combinations [1,19,21] — or indirect — the basis upon which executive functions such as cognitive control operate on during the creative process [11,22]. Thus, semantic memory structure plays an important role in the creative process [23]. It is important to note that the role of semantic memory in creativity may be only relevant in verbal creative output, and not in other forms such as visual creative output [4]. However, the general theories on creativity described above do not make any such distinctions. Yet, currently the nature and structure of semantic memory remains an open issue [5].

A classic cognitive model of semantic memory structure was proposed by Collins and Loftus [24]. According to their model, concepts in memory are structured as a network, according to a semantic similarity principal (Figure 1a). The more semantic properties two concepts

Figure 1



An illustration of semantic memory structure, adapted from Collins and Loftus [24]. **(a)** A visualization of a theoretical example of the concept red and its surround concepts, organized based on overlap of semantic features. **(b)** The concept *red* and only its directly connected neighbors in the theoretical example provided by Collins and Loftus [24]. **(c)** LSA-based semantic distance scores between the concept *red* and some of its directly connected neighbors in the theoretical example provided by Collins and Loftus [24].

share, they argue, the more links there are between them [24]. Accordingly, the authors define semantic distance as the 'shortest path [direct or indirect] between two nodes' [24; p. 412, note 3]. Furthermore, the authors argue for a spreading activation model, where once a concept in the semantic network is activated, activation spreads from it to all of its directly connected neighbors, activation which quickly decays over time and space [24]. Therefore, this model converges strongly with the main theories of creativity, which largely attribute the creative process to an active process operating upon semantic memory structure [3•].

Latent semantic analysis

A popular computational method to represent semantic distance in creativity research is through Latent Semantic Analysis [LSA; 25,26]. LSA quantifies the semantic similarity between words in a given high-dimensional semantic space by determining the probability of a given word co-occurring in a specific context (e.g. a paragraph of a document). The semantic distance between a pair of words is determined by subtracting the LSA similarity score from 1 [27]. For example, the LSA-based semantic distance between pairs of words in the theoretical example provided by Collins and Loftus [24] can be computed (Figure 1c). However, despite the concepts *red* and *fire* being directly connected in [24], the LSA-based semantic distance between these concepts is quite high (.85). In comparison, and as expected, the LSA-based semantic distance between *red* and *yellow* is quite short (.26).

In the past few years, several studies have used LSA-based measures of semantic distance to study creativity [28–30]. For example, Prabhakaran *et al.* [27] examined LSA semantic distances of responses in a verb generation task, where participants were required to produce verbs in response to a series of nouns, either a verb in a cued 'be creative' condition or a verb in an un-cued 'typical'

condition. The authors found that the LSA-based semantic distance of the verbs to the nouns were higher in the cued 'be creative' condition compared to the non-cued 'typical' condition. Furthermore, the semantic distances between the verbs and nouns generated in the cued 'be creative' condition were also related to measures of fluid intelligence, divergent thinking and creative achievement. Heinen and Johnson [8••] recently showed that LSA-based measures of semantic distance relate to measures of novelty and appropriateness, common subjective measures of creative output [4]. Furthermore, the authors show that such LSA scores were sensitive to instruction manipulation and changed when participants were required to generate creative responses. Such creative responses had an average intermediate LSA-based semantic distance score, compared to a low average score for common responses or high average score for random responses [8••].

Other studies have used LSA-based measures of semantic distance to examine different aspects of the cognitive processes involved in creativity. Beaty *et al.* [31] used LSA to assess participants associative abilities, by averaging semantic distance values of their responses generated during verbal fluency tasks to specific cue words. This measure of associative ability, along with several other measures of cognitive ability, was used to examine the contribution of both semantic memory structure and executive function processes to creative ability. Using structural equation modelling, the authors found joint effects of both associative and executive abilities on the fluency and subjectively rated creativity scores of divergent thinking responses. Finally, Hass [7•,32] has recently applied LSA-based measures of semantic distance to 'track' the dynamics of generating divergent thinking responses in a continuous divergent thinking task. Thus, these studies demonstrate the significance of computing LSA-based measures of semantic distance to examine creative output.

However, objections have been raised regarding the validity of this approach as a measure of semantic distance and in predicting semantic priming effects [33–36]. Research has also indicated that performance of LSA models strongly depends on the choice and scope of the text corpus used, which can become the determining factor in how well the model captures human performance [35]. Furthermore, while more advanced computational models of semantic spaces are being developed [37•], the validity of estimating semantic distance based on analysis of textual corpora has yet to be determined [38]. Finally, Forthmann, Oyebade, Ojo, Günther, and Holling [39] have shown how elaboration in DT responses (number of words of a DT response) can bias their LSA-based measures of semantic distance. The authors caution with interpretation of LSA-based measures of semantic distance of DT responses and offer ways to correct for such potential biases [39]. Thus, while useful, application of LSA-based measures of semantic distance in creativity research can only provide information on the creative output. Furthermore, researchers should be aware of potential methodological pitfalls.

Semantic networks

A different computational approach to represent semantic distances might be realized through path lengths in semantic networks. Network science is based on mathematical graph theory, providing quantitative methods to investigate complex systems as networks [40–42]. A network is comprised of nodes, which represent the basic units of a system (e.g. semantic memory) and links, or edges, that signify the relations between them (e.g. semantic similarity). By structuring language and memory as a network, network science can directly and quantitatively examine classic cognitive theory and the operations of cognitive processes that take place in memory retrieval and associative thought. For example, the theoretical example provided by Collins and Loftus [24] can be represented as a semantic network (Figure 1a) or just a subset of that network. For example, Figure 1b plots the directly connected neighbors of the concept *red* from the theoretical example provided by Collins and Loftus [24].

A growing number of studies have applied network science methodologies to study creativity, focusing on the role of semantic memory structure in the creative process [2]. These studies have shown how differences in semantic memory structure relate to individual differences in creativity, both at the group level [43] and at the individual level [44•]; simulated search processes over the semantic networks of low and high creative individuals [45]; examined the relations of semantic memory structure to creative achievement and fluid intelligence [46]; and related flexibility of thought to the robustness of their semantic networks to attack [47]. These studies have quantitatively shown how the semantic memory structure of high creative individuals is more connected and more flexible, allowing for more efficient spreading of activation throughout such a semantic network structure [2].

A method to measure semantic distance with network science tools was recently developed [48••]. The authors demonstrate how path length over a semantic network can serve as a measure of semantic distance. The authors developed a semantic distance task, in which participants judged whether two words were related to each other. The word pairs constructed for this study varied in the path length between them, based on a large-scale network analysis of the Hebrew lexicon [48••]. This study found a differential effect of path length on participant's performance: When up to four steps separated between word-pairs, participants exhibited an increase in RT and decrease in the percentage of word-pairs judged as related. From four steps onwards, participants exhibited

a significant decrease in RT and the word-pairs were dominantly judged as unrelated. Path length was also related to performance in memory retrieval — as the path length between the word-pairs grew, it was harder for participants to retrieve them in a subsequent memory recall test. The authors also found that path length was significantly correlated to subjective ratings of the word-pairs, judged by an independent group of judges: As path length between the word-pairs grew, they were judged to be more weakly related to each other. Finally, the authors show how path length outperforms LSA-based measures of semantic distance in predicting participant's performance on the semantic distance task [48••]. As such, this method can be applied in creativity research, to examine theories on individual differences in creativity related to semantic memory structure [1].

An effort to examine divergent thinking from a semantic network perspective has been developed by Runco *et al.* [49,50]. Acar and Runco [49] examined the semantic distance of divergent thinking responses based on data collected from three different sources of semantic networks. Each of the divergent thinking responses was classified as close, remote or highly remote, depending on the number of concepts in the responses that were found in a specific network. The authors show that this approach is more objective and reliable in assessing divergent thinking responses. However, this approach is not based on a network science methodology per se, but rather matching of divergent thinking responses to close or remote concepts in textual corpora. In this sense, the method is more similar to LSA-based measures and its caveats as described above.

Conclusions

What can quantitative measures of semantic distance applied in behavioral research tell us about creativity? The application of such methods can potentially provide a more objective measure of the output of creative thinking [6,8••] and it can quantitatively examine the role of semantic memory in creative thinking which has been mostly under-examined [2]. Finally, it can shed novel light on the interaction between bottom-up and top-down cognitive processes that realize creative thinking [3•,9••,31,46]. Importantly, for behavioral research, these methods offer powerful quantitative and objective measures that can supplement and strengthen standard subjective scoring methods of the creative process and its output [4].

However, it is important to be aware that LSA-based and network-based measures of semantic distance do not overlap. As can be seen from Figure 1, based on the theoretical example provided by Collins and Loftus [24], several of the directly related neighbors of the concept *red* (Figure 1b) generate quite large LSA-based semantic distance scores (Figure 1c). More methodological work is needed to identify how both of these types of measures relate to the theoretical construct of semantic distance and to each other. Furthermore, given that semantic memory varies across individuals and is dynamic [51], and that creative thinking leads to dynamic shifts in its structure [52], individual-based quantitative approaches need to be developed (e.g. [44•]).

In conclusion, a growing number of studies are applying quantitative measures of semantic distance in creativity research. Such methods are providing objective measures of the creative output and a more direct investigation of the role of semantic memory structure, and semantic distance, in the creative process. While these measures are still far from replacing subjective scoring methods of creativity [8••], they provide a powerful way to quantitatively study creativity [3•,9••].

Conflict of interest statement

Nothing declared.

Acknowledgements

I thank Boris Forthmann for enlightening discussions that contributed to this manuscript. I also thank Roger Beaty for constructive comments on previous versions of this manuscript.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as

- • of special interest
- •• of outstanding interest

1. Mednick AS: **The associative basis of the creative process**. *Psychol Rev* 1962, **69**:220-232.

2. Kenett YN: **Investigating creativity from a semantic network perspective**. In *Exploring Transdisciplinarity in Art and Science.* Edited by Kapoula Z, Renoult J, Volle E, Andreatta M. Springer; 2018:49-75.

3. • Volle E: **Associative and controlled cognition in divergent thinking: theoretical, experimental, neuroimaging evidence, and new directions**. In *The Cambridge Handbook of the Neuroscience of Creativity.* Edited by Jung RE, Vartanian O. Cambridge University Press; 2018:333-362.

This chapter provides an in-depth neurocognitive review of divergent thinking, reviewing both cognitive and neural findings. Specifically, the chapter examines the role of semantic memory structure and executive processes in divergent thinking. The chapter proposes a new theory — the Distance Dependent Representation Activation Model, highlighting the interaction of both semantic memory and executive functions in the creative process.

4. Runco MA, Acar S: **Divergent thinking as an indicator of creative potential**. *Creat Res J* 2012, **24**:66-75.

5. Jones MN, Willits J, Dennis S: **Models of semantic memory**. In *Oxford Handbook of Mathematical and Computational Psychology.* Edited by Busemeyer J, Townsend J. Oxford University Press; 2015:232-254.

6. Green AE: **Creativity, within reason: semantic distance and dynamic state creativity in relational thinking and reasoning**. *Curr Dir Psychol Sci* 2016, **25**:28-35.

7. • Hass RW: **Tracking the dynamics of divergent thinking via semantic distance: analytic methods and theoretical implications**. *Mem Cognit* 2016, **45**:233-244.

This study is the first to examine LSA-based semantic distance measures of responses generated in a continuous divergent thinking task. This was done to examine how the LSA-based semantic distance changes

between sequential divergent thinking responses and compared to the cue. The method developed in this study offers a way to examine semantic distances between multiple word sentences, thus advancing the application of such quantitative methods in studying more naturalistic creative output.

8. •• Heinen DJP, Johnson DR: **Semantic distance: an automated measure of creativity that Is novel and appropriate**. *Psychol Aesthet Creat Arts* 2017 http://dx.doi.org/10.1037/aca0000125.

This study demonstrates, for the first time, how subjective measures of creativity such as novelty and appropriateness can be related to LSA-based measures of semantic distance. Furthermore, the authors demonstrate how instruction manipulation can affect LSA-based semantic distances. The authors propose that LSA-based measures of semantic distance can serve as objective, automatic, measures of creativity.

9. •• Kenett YN: **Going the extra creative mile: the role of semantic distance in creativity – theory, research, and measurement**. In *The Cambridge Handbook of the Neuroscience of Creativity.* Edited by Jung RE, Vartanian O. Cambridge University Press; 2018:233-248.

A more in-depth review on the role of semantic distance in creativity and how it is measured. This chapter provides further details on the LSA-based and network-based approaches, including a more traditional, frequency-based approach to measuring semantic distance.

10. Beaty RE, Silvia PJ: **Why do ideas get more creative over time? An executive interpretation of the serial order effect in divergent thinking tasks**. *Psychol Aesthet Creat Arts* 2012, **6**:309-319.

11. Benedek M, Jauk E, Sommer M, Arendasy M, Neubauer AC: **Intelligence, creativity, and cognitive control: the common and differential involvement of executive functions in intelligence and creativity**. *Intelligence* 2014, **46**:73-83.

12. Jauk E, Benedek M, Neubauer AC: **The road to creative achievement: a latent variable model of ability and personality predictors**. *Eur J Pers* 2014, **28**:95-105.

13. Lee CS, Therriault DJ: **The cognitive underpinnings of creative thought: a latent variable analysis exploring the roles of intelligence and working memory in three creative thinking processes**. *Intelligence* 2013, **41**:306-320.

14. Nusbaum EC, Silvia PJ: **Are intelligence and creativity really so different?: Fluid intelligence, executive processes, and strategy use in divergent thinking**. *Intelligence* 2011, **39**:36-45.

15. Silvia PJ, Beaty RE, Nusbaum EC: **Verbal fluency and creativity: General and specific contributions of broad retrieval ability (Gr) factors to divergent thinking**. *Intelligence* 2013, **41**:328-340.

16. Barr N, Pennycook G, Stolz JA, Fugelsang JA: **Reasoned connections: A dual-process perspective on creative thought**. *Think Reason* 2014, **21**:61-75.

17. Finke RA, Ward TB, Smith SM: *Creative Cognition: Theory, Research, and Applications*. MIT Press; 1992.

18. Gabora L: **Revenge of the "neurds": Characterizing creative thought in terms of the structure and dynamics of memory**. *Creat Res J* 2010, **22**:1-13.

19. Simonton DK: **Creative thought as blind variation and selective retention: Why creativity is inversely related to sightedness**. *J Theor Philos Psychol* 2013, **33**:253-266.

20. Simonton DK: **On praising convergent thinking: Creativity as blind variation and selective retention**. *Creat Res J* 2015, **27**:262-270.

21. Sowden PT, Pringle A, Gabora L: **The shifting sands of creative thinking: Connections to dual-process theory**. *Think Reason* 2014, **21**:40-60.

22. Mendelsohn GA: **Associative and attentional processes in creative performance**. *J Pers* 1976, **44**:341-369.

23. Abraham A, Bubic A: **Semantic memory as the root of imagination**. *Front Psychol* 2015, **6**:325.

24. Collins AM, Loftus EF: **A spreading-activation theory of semantic processing**. *Psychol Rev* 1975, **82**:407-428.

25. Landauer TK, Dumais ST: **A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge**. *Psychol Rev* 1997, **104**:211-240.

26. Landauer TK, Foltz PW, Laham D: **An introduction to latent semantic analysis**. *Discourse Process* 1998, **25**:259-284.

27. Prabhakaran R, Green AE, Gray JR: **Thin slices of creativity: Using single-word utterances to assess creative cognition**. *Behav Res Methods* 2014, **46**:641-659.

28. Bourgin DD, Abbott JT, Griffiths TL, Smith KA, Vul E: **Empirical evidence for Markov chain Monte Carlo in memory search**. *Proc 36th Annu Meet Cogn Sci Soc*. 2014.

29. Dumas D, Dunbar KN: **Understanding Fluency and Originality: A latent variable perspective**. *Think Ski Creat* 2014, **14**:56-67.

30. Forster EA, Dunbar KN: **Creativity evaluation through latent semantic analysis**. *Proc 31st Annu Meet Cog Sci Soc*. 2009:602-607.

31. Beaty RE, Silvia PJ, Nusbaum EC, Jauk E, Benedek M: **The roles of associative and executive processes in creative cognition**. *Mem Cognit* 2014, **42**:1-12.

32. Hass RW: **Semantic search during divergent thinking**. *Cognition* 2017, **166**:344-357.

33. Hutchison KA, Balota DA, Cortese MJ, Watson JM: **Predicting semantic priming at the item level**. *Q J Exp Psychol* 2008, **61**:1036-1066.

34. Mandera P, Keuleers E, Brysbaert M: **How useful are corpus-based methods for extrapolating psycholinguistic variables?** *Q J Exp Psychol* 2015, **8**:1628-1642.

35. Recchia G, Jones MN: **More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis**. *Behav Res Methods* 2009, **41**:647-665.

36. Simmons S, Estes Z: **Using latent semantic analysis to estimate similarity**. *Proc 28th Annu Meet Cog Sci Soc*. 2006.

37. • Mandera P, Keuleers E, Brysbaert M: **Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: a review and empirical validation**. *J Mem Lang* 2017, **92**:57-78.

In this study, the authors describe state-of-the-art methods to assess semantic distance based on textual corpora (similar to LSA). The authors then demonstrate how these new models predict behavioral performance in psycholinguistic tasks and address previous concerns raised against LSA. Importantly, the authors also provide semantic vectors in English and Dutch along side a graphical interface to work with.

38. De Deyne S, Verheyen S, Storms G: **Structure and organization of the mental lexicon: A network approach derived from syntactic dependency relations and word associations**. In *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks.* Edited by Mehler A, Blanchard P, Job B, Banish S. Springer; 2016:47-79.

39. Forthmann B, Oyebade O, Ojo A, Günther F, Holling H: **Application of latent semantic analysis to divergent thinking is biased by elaboration**. *J Creat Behav* 2018 http://dx.doi.org/10.1002/jocb.240.

40. Baronchelli A, Ferrer-i-Cancho R, Pastor-Satorras R, Chater N, Christiansen MH: **Networks in cognitive science**. *Trends Cogn Sci* 2013, **17**:348-360.

41. Borge-Holthoefer J, Arenas A: **Semantic networks: Structure and dynamics**. *Entropy* 2010, **12**:1264-1302.

42. Karuza EA, Thompson-Schill SL, Bassett DS: **Local patterns to global architectures: Influences of network topology on human learning**. *Trends Cogn Sci* 2016, **20**:629-640.

43. Kenett YN, Anaki D, Faust M: **Investigating the structure of semantic networks in low and high creative persons**. *Front Hum Neurosci* 2014, **8**:407.

44. • Benedek M, Kenett YN, Umdasch K, Anaki D, Faust M, Neubauer AC: **How semantic memory structure and intelligence contribute to creative thought: A network science approach**. *Think Reason* 2017, **23**:158-183.

This study developed a methodology to represent an individual's semantic network, based on semantic relatedness judgments. The authors then demonstrate how the network properties of individual semantic networks relate to individual differences in creative ability. Such an approach is important in advancing the investigation of the role of semantic memory structure, as well as semantic distance, in creativity.

45. Kenett YN, Austerweil JL: **Examining search processes in low and high creative individuals with random walks**. *Proc 38th Annu Meet Cogn Sci Soc*. 2016:313-318.

46. Kenett YN, Beaty RE, Silvia PJ, Anaki D, Faust M: **Structure and flexibility: Investigating the relation between the structure of the mental lexicon, fluid intelligence, and creative achievement**. *Psychol Aesthet Creat Arts* 2016, **10**:377-388.

47. Kenett YN, Levy O, Kenett DY, Stanley HE, Faust M, Havlin S: **Flexibility of thought in high creative individuals represented by percolation analysis**. *Proc Natl Acad Sci USA* 2018, **115**:867-873.

48. •• Kenett YN, Levi E, Anaki D, Faust M: **The semantic distance task: Quantifying semantic distance with semantic network path length**. *J Exp Psychol Learn Mem Cogn* 2017, **43**:1470-1489.

This study is the first to propose that path length over a semantic network can serve as an alternative quantitative measure to semantic distance. The authors demonstrate how path length between word-pairs, computed from a large-scale semantic network, predicts performance in a semantic relatedness judgment task and is related to success in recalling the judged word-pairs in a surprise memory recall task. Finally, the authors show how this measure outperforms LSA-based measures in predicting performance in their task.

49. Acar S, Runco MA: **Assessing associative distance among ideas elicited by tests of divergent thinking**. *Creat Res J* 2014, **26**:229-238.

50. Beketayev K, Runco MA: **Scoring divergent thinking tests by computer with a semantics-based algorithm**. *Eur J Psychol* 2016, **12**:210-220.

51. Yee E, Thompson-Schill SL: **Putting concepts into context**. *Psychon Bull Rev* 2016, **23**:1015-1027.

52. Schilling MA: **A "small-world" network model of cognitive insight**. *Creat Res J* 2005, **17**:131-154.