Word Activation Forces Map Word Networks

Jun Guo[1], Hanliang Guo[2], Zhanyi Wang[1]

[1]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications

No.10 Xitucheng Road, Haidian District, Beijing, China

[2] Aerospace and Mechanical Department, Viterbi School of Engineering, University of Southern California

University Park Campus, USC, Los Angeles, CA 90089 USA

Words associate with each other in a manner of intricate clusters[1-3]. Yet the brain capably encodes the complex relations into workable networks[4-7] such that the onset of a word in the brain automatically and selectively activates its associates facilitating the language understanding and generation[8-10]. One believes that the activation strength from one word to another forges and accounts for the latent structures of the word networks. This implies that mapping the word networks from brains to computers[11,12], which is necessary for various purposes[1,2,13-15], may be achieved through modeling the activation strengths. However, although a lot of investigations on word activation effects have been carried out[8-10,16-20], modeling the activation strengths remains open. Consequently, huge labor is required to do the mappings[11,12]. Here we show that our found *word activation force*s, statistically defined by a formula in the same form of the universal gravitation, capture essential information on the word networks, leading to a superior approach to the mappings. The approach compatibly encodes syntactical and semantic information into sparse coding directed networks, comprehensively highlights the features of individual words. We find that based on the directed networks sensible word clusters and hierarchies can be efficiently discovered. Our striking results strongly suggest that the word activation forces might reveal the

encoding of word networks in the brain.

Mapping word associations, which constitute complex networks, from brains to computers has been highly labor-consuming. A practically useful word network may cost thousands of people's labor to perform free association[11] or hundreds of man-months of labeling the complex word clusters[12]. Can such kind of tasks be automatically accomplished? To answer this question here we show an approach based on modeling the strengths of word activation in the brain.

Previous research suggests that the effects of word activation in the brain are trained by the language experience of the target human[21,22]. For a given word pair, their respective occurrence probabilities and co-occurrence probability are the key factors in the training[10,22,23]. By using well-designed large text corpora including the British National Corpus (BNC)[24] and the American National Corpus (ANC)[25] to approximate human language experiences, we studied the relations between the word activation effects and the statistics of word occurrence and co-occurrence, found a sort of statistics that reliably predict the word activation effects (see Supplementary Information, Section 3, for the details of the prediction).

Specifically, given the frequencies f_i and f_j and co-occurrence frequency f_{ij} of a pair of words i and j in the corpus used to simulate the language experience of the target human, we predict the strength of the activation that word i exerts on word j through the statistic $(f_{ij} / f_i)\ (f_{ij} / f_j\ )/d_{ij}^{2}$, where d_{ij} is the average distance by which word i precedes word j in their co-occurrences. Seeing the ratios of f_{ij} to f_i and f_{ij} to f_j as masses, we identify that the statistic is defined in the same form of the universal gravitation. Therefore we name it as *word activation force* from i to j, shortly waf_{ij}. According to the definition, the magnitude of *waf*s is unitarily quantified among [0,1]. Taking waf_{ij} for example, *zero* means that word i

is never followed by word j closer than L words in the language experience, while *one* means that words i and j are always immediately adjacent like a compound ($f_{ij} = f_i = f_j$, $d_{ij} = 1$). Readily, given a vocabulary, the *waf*s of every pair of the words constitute a squared but asymmetrical matrix **WAF** = $\{waf_{ij}\}$, i.e. a directed word network, where nonzero elements in the ith row give the out-links of the ith node (from word i to others), while nonzero elements in the ith column the in-links of it (from others to word i).

Now we demonstrate that a **WAF** mapped word network can be highly sensible and valuable through an example in which the **WAF** is created with the statistics of a vocabulary of 10,000 frequent English words in the BNC (see Supplementary Information Section 2.1 for details). In the **WAF**, the distribution of either in- or out-link strengths of the node (word) is heavy tailed, i.e. the words high-selectively distribute their link strength. For a particular word, the major fraction of the link strengths is only related to a few words, which are usually its partners in the relations of compound, phrase, head-modifier, subject-verb, verb-object, synonym, antonym etc. It is shown that the *waf*s highlight the key features of individual words while the **WAF** captures the overall associations between words. Meanwhile the heavy tailed distributions allow a sparse coding on the **WAF**, i.e. cutting off the meaninglessly weak links at a threshold T. In our experiment, with a T = 1.0E-6, we cut off 96.36% of the links (from 21,244,909 to 773,468) remaining 96.67% of the total strength of all links (from 39.94 to 38.61). Fig. 1 shows three ordinary nodes in the **WAF** after the sparse coding (see Supplementary Data1 for detailed data). The complete network of the sparse **WAF** is acquired by a program in Supplementary Programs, and its statistical features, which are completely consistent with the well-known natures of word networks, are provided in Supplementary Information Section 2.1.

To identify word clusters based on the distinctive directed word network **WAF**, we introduce a word *affinity measure* A^{waf} from a unique perspective that deviates from the currently popular ones of semantic space models[26,27] (perspectives of vector space). A^{waf} is defined as the geometric average of the mean overlap rates of the in-links and out-links of the inquired two words (see Methods Summary for the formula). In advance, we validated that A^{waf} produced affinities and the human judged synonymies on a set of benchmark word pairs[28] are significantly correlated through an auxiliary experiment [0.52(p = 1.13E-5), see Supplementary Information Section 3.2 for details].

In our main experiment, we applied A^{waf} to every pair of words in the vocabulary to generate a symmetric affinity matrix, i.e. an undirected word network. By sorting the words according to their affinities to each word, we found that the word network is incredibly consistent with human knowledge: Almost every node (word) keeps strong links to its relatives but no link or weak ones to the irrelatives. Especially, across parts of speech, granularity of the concepts and popularity of the words, a large amount of the words possess the strongest links to their best partners, such as *a~the*, *abbey~monastery*, *aberdeen~dundee, ability~capacity, above~below, abroad~elsewhere, abruptly~swiftly, absence~presence, abundance~diversity, abuse~violence, academic~scientific, academy~institution, accept~recognise, acceptable~reasonable, accommodate~adapt,* etc. Reasonably, nouns and verbs usually keep strong links to their siblings in changed forms, e.g. *arm~arms*, *arrive~arrives*, and *arriving~arrive*. In addition, the absolute values of the affinities make sense on the closeness between words. It means that the strengths of links (affinities) of different nodes are mutually comparable, and that a uniform threshold can be adopted to remove the weak links when necessary.

To present the significant structure of the word network in a visualised way, we group every word and its top 5 neighbors forming diverse 6-word-clusters based sub-networks, and provide the complete results in Supplementary Data2. Here through a few examples we show that the 6-word-clusters and the sub-networks are striking. Fig. 2 shows the 6-word-clusters with different parts of speech. In a wider scope, Fig. 3 shows a complex sub-network including various hierarchies in the domain of *science and art*. Such a sub-network presents an inherent complex structure of word networks that includes intricate semantic relations. Additional instances of the complex sub-networks are provided in Supplementary Information Section 2.2. Note that, for showing the essences of the word network, we adopted an intuitive clustering method. Obviously, by combining the top *N* clustering with a reasonable threshold of affinity, or by using the specialised clustering algorisms for complex networks[1,2,29,30] instead, better clusters and hierarchies can be found.

Comparing the neighboring words in our networks and the associated words in manual free association makes more sense. To this end, targeting 3,269 words in our vocabulary which are overlapped by the words used in the free association of ref. 11, we compare their top 3 neighbors in our network and top 3 associates in the free association. The comparison shows that our results are mostly comparable to the ones of the free association. Table 1 presents a small part of the comparison for the target words of common used nouns and verbs in daily life (see Supplementary Data3 for the complete comparison). From Table 1, a big proportion of the target words our network and the free association give common answers, such as for *beer*, *wine*, *walk*, *talk* etc. In total this proportion is some 1/4 (798:3,269). For the rest, although both the neighbors and the associates are sensible, they present respective characters. In contrast to the freedom of the associates in categories, the neighbors are exerted more syntactic constraints, leading to that they are much more consistent in parts of speech while they are occasionally loose in semantics (e.g. for *eat*). Notably, our results are merely based on the BNC, a

100 million word corpus, thereby they are naturally characterised by its specific contexts for each words which are unnecessarily consistent with common sense. For example, in our network, *apple*'s top three neighbors are *microsoft*, *novell* and *ibm,* it suggests that the word is mainly in the contexts of computer industry rather than daily life in the corpus.

Remarkably, our approach is highly efficient. Thanks to the sharply skew distributions of the *waf*s, the encoding of the **WAF** is very sparse. It makes the computation of 10^8 affinities between every pair of words tractable under general conditions. Our computation was accomplished within 16 hours on a platform of PC (4*2.66GHz Quad CPU) + Matlab. Besides the practical benefit, the high efficiency also suggests that the principles underlying the **WAF** are plausible.

From a broader angle of view, our approach might reveal the learning mechanism of linguistic neural networks. The hypotheses that statistical information underlies linguistic neural networks have long been proposed[21,23]. However, what statistics are crucial for the development of the neural networks remains unclear. Our found statistics *word activation forces* efficiently capture the substantial associations between words, automatically leading to human knowledge consistent word networks. This suggests that it is likely to turn to a promising direction of understanding the learning mechanism with the hint of the newly found statistics. Practically, with the high effectiveness and efficiency, our approach has the immediate future of application in various tasks such as word clustering, thesaurus establishment, word sense discrimination, and query extension in information retrieval and so on.

METHODS SUMMARY

Counting occurrences and co-occurrences of words. Word frequencies are counted under the condition without stemming verbs or changing nouns between plural and singular forms but with changing all upper cases into lower cases. For example, *change*, *changed*, *changing*, *term*, *terms* were treated as 5 different words, but *CAT*, *Cat* and *cat* the same word *cat*. To count the co-occurrence frequency f_{ij}, the limit of the farthest position (indicated by word number) where word j appears behind word i, referred to as L, should be predetermined. Referring to previous work, L is tested around 5 in this study. We found that values of *waf*s are not sensitive to L ranging from 4 to 5 and 6. Therefore we only give the results of L = 5 in the main text. To ensure the ratios of f_{ij} to f_i and f_{ij} to f_j are less than or equal to 1, we only count the co-occurrences of word i and word j where neither word i nor word j appears in the intervening words.

The affinity measure between the words in the WAF. For a pair of words i and j in the directed word network **WAF**, we define their affinity as:

$$\mathrm{A}_{ij}^{waf}=[\frac{1}{|\mathbf{K}_{ij}|}\sum_{k\in\mathbf{K}_{ij}}OR(waf_{ki},waf_{kj})\cdot\frac{1}{|\mathbf{L}_{ij}|}\sum_{l\in\mathbf{L}_{ij}}OR(waf_{il},waf_{jl})]^{1/2}$$

where $\mathbf{K}_{ij}=\{k\,|\,waf_{ki}>0\ or\ waf_{kj}>0\}$, $\mathbf{L}_{ij}=\{l\,|\,waf_{il}>0\ or\ waf_{jl}>0\}$, and $OR(x,y)=min(x,y)/max(x,y)$. Readily, $\mathbf{K}_{ij}$ and $\mathbf{L}_{ij}$ are the sets of the labels of the words connected by the in-links and the out-links of word i or word j, respectively. And $OR(x,y)$ is an overlap rate function of x and y. That is, we define A_{ij}^{waf} as the geometric average of the mean overlap rates of the in-links and out-links of the words i and j in the **WAF**. Obviously, $\mathrm{A}_{ij}^{waf}=\mathrm{A}_{ji}^{waf}$. Therefore, using this measure we can acquire an undirected word network whose links represent word affinities from the directed one **WAF**. Notably, since **WAF** is sparse, the computation of A_{ij}^{waf} is efficient.

References

1. Palla, G., Derenyi, I., Farkas, I. and Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, 814–818 (2005).

2. Ahn, Y. Y., Bagrow, J. P. & Lehmann, S. Link communities reveal multiscale complexity in networks. *Nature* **466**, 761–764 (2010).

3. Roweis, S. T. & Saul, L. K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**, 2323-2326 (2000).

4. Crutch, S. J. & Warrington, E. K. Abstract words are represented in an associative network whilst concrete words are represented in a categorical network. *Brain & Lang.* **91** 13-14 (2004).

5. Crutch, S. J. & Warrington, E. K. The semantic organisation of proper nouns: the case of people and brand names. *Neuropsychol*. **42**, 584–596 (2004).

6. Carreiras, M., *et al*. An anatomical signature for literacy. *Nature* **461**, 983-988(2008).

7. Szathmary, E. & Szamado, S. Being human: language: a social history of words. *Nature* **456**, 40-41(2008).

8. Turkeltaub, P. E., Gareau, L., Flowers, D. L., Zeffiro, T. A.&Eden, G. F. Development of neural mechanisms for reading. *Nature Neurosci*. **6**, 767-773 (2003).

9. Balota, D. A. & Lorch, R. F. Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *J. Exp. Psychol. Learn. Mem. Cogn.***12**, 336-345 (1986).

10. McKoon, G. & Ratcliff, R. Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *J. Exp. Psychol. Learn. Mem. Cogn.* **18**, 1155-1172 (1992).

11. Nelson, D. L., McEvoy, C. L. & Schreiber, T. A. The University of South Florida word association, rhyme, and word fragment norms. http://w3.usf.edu/FreeAssociation/ (1998).

12. Fellbaum, C. *WordNet: An Electronical Lexical Database* (The MIT Press, Cambridge, MA 1998).

13. Cano, P., *et al.* Nearest-Neighbor automatic sound annotation with a WordNet taxonomy. J. Intelligent Information Systems **24**, 99-111(2005).

14. Degemmis, M., Lops, P. & Semeraro, G. A content-collaborative recommender that exploits WordNet-based user profiles for neighborhood formation. *User Model User-Adap Inter*. **17**, 217-255 (2007).

15. Kolte, S. G. & Bhirud, S. G. WordNet: a knowledge source for word sense disambiguation. *I. J. Recent Trends in Eng.* **2**, 213-217 (2009).

16. Kutas, M. & Hillyard, S. A. Brain potentials during reading reflect word expectancy and semantic association. *Nature* **307**, 161-163 (1984).

17. Henson, R., Shallice, T. & Dolan, R. Neuroimaging Evidence for Dissociable Forms of Repetition Priming. *Science* **287**, 1269-1272 (2000).

18. Crinion, J., *et al*. Language Control in the Bilingual Brain. *Science* **312**, 1537-1540 (2006).

19. Gennari, S. P., MacDonald, M. C., Postle, B. R. & Seidenberg, M. S. Context-dependent interpretation of words: Evidence for interactive neural processes. *NeuroImage* **35**, 1278-1286 (2007).

20. Hoeks, J. C. J. *et al.* Seeing words in context: the interaction of lexical and sentence level information during reading. *Cogn. Brain Res.* **19**, 59-73 (2004).

21. Seidenberg, M. S. Language Acquisition and Use: Learning and Applying Probabilistic Constraints. *Science* **275**, 1599-1603 (1997).

22. Church, K. & Hanks, P. Word association norms, mutual information, and lexicography. *Comput. Linguist.* **16**, 22-29 (1990).

23. Mitchell, T. M. *et al*. Predicting Human Brain Activity Associated with the Meanings of Nouns. *Science* **320**, 1191-1195 (2008);

24. British National Corpus. http://www.natcorp.ox.ac.uk/ (2009).

25. American National Corpus. http://www.americannationalcorpus.org/ (2002).

26. Will, L. Towards a theory of semantic space. Proc. of the 23rd Annual Conf. of the Cognitive Science Society, 576–581 (Edinburgh, Scotland, 2001).

27. Padó, S. & Lapata, M. Dependency-based construction of semantic space models. *Comput. Linguist.* **33**, 161-199 (2007).

28. Rubenstein, H. & Goodenough, J. B. Contextual correlates of synonymy. *Comm. of the ACM* **8**, 627–633 (1965).

29. Clauset, A., Moore, C. & Newman, M. E. J. Hierarchical structure and the prediction of missing links in networks. *Nature* **453**, 98–101 (2008).

30. Guimera, R. & Newman, M. E. J. &Moore, C. Finding community structure in very large networks. *Phys. Rev. E***70**, 066111 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank Guang Chen for discussions inspiring this study; Yang Luo and Junliang Bai for their assistance in initial data analysis. Research supported by Chinese 111 program of 'Advanced Intelligence and Network Service' under grant no. B08004.

Author Contributions J. G. presented the formulae of the word activation force and the word affinity measure, constructed the word networks and wrote the manuscript. H. G. prepared materials, compared our results with those of others, and revised the manuscript. Z. W. counted the frequencies of occurrence and co-occurrence of the words in the BNC, and accomplished the illustration of the word networks.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.G. (guojun@bupt.edu.cn).

Table 1. Top 3 neighbors and associates with target words from our network and free association

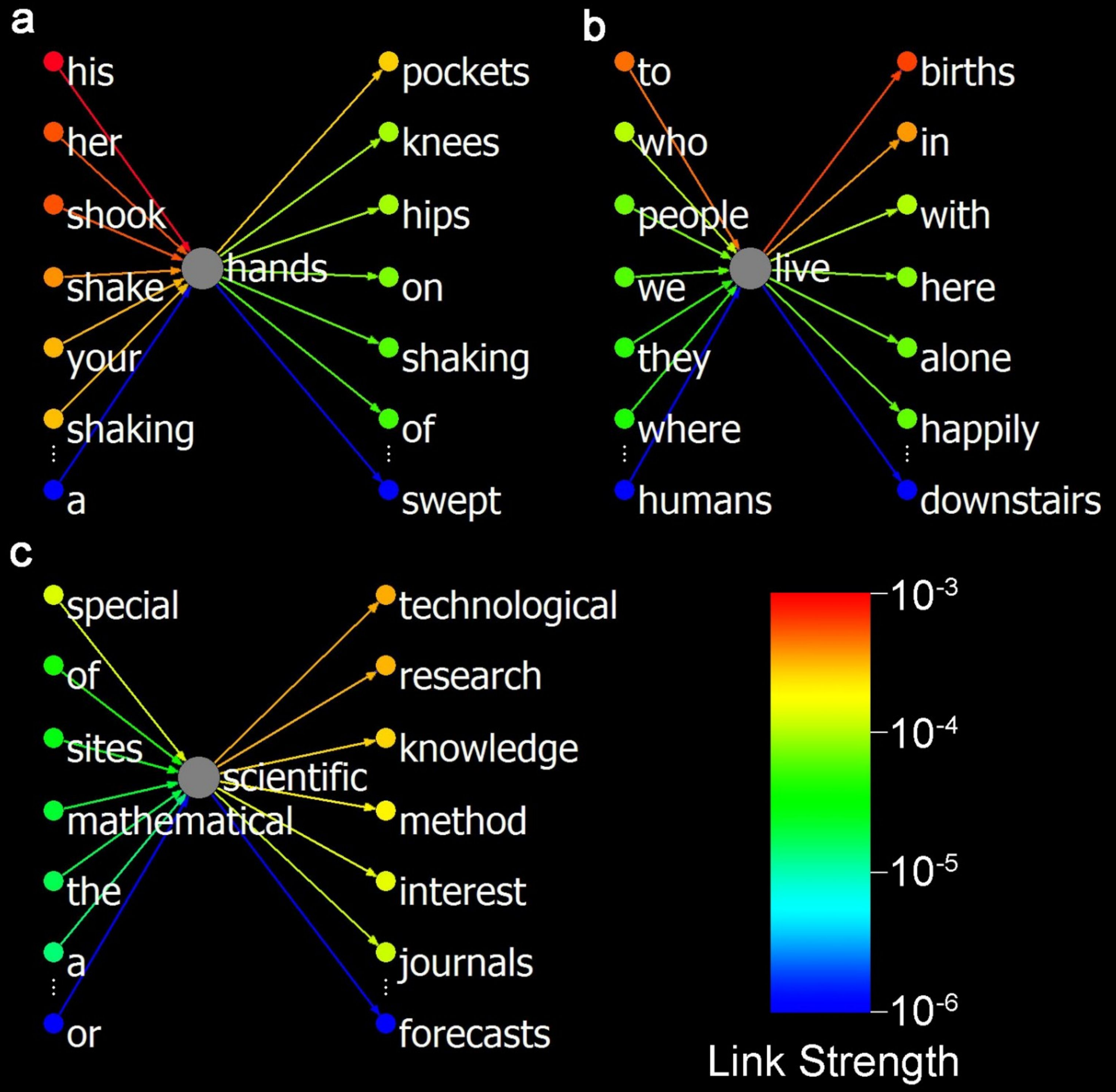
Targets	Neighbors of our network	Associates of free association
bread	meat cheese toast	butter dough loaf
butter	cream cheese flour	bread margarine milk
milk	meat cream wine	cow drink honey
drink	drinking coffee sleep	water beer thirst
beer	wine whisky champagne	drink wine drunk
wine	coffee beer champagne	beer drink dine
drunk	asleep alone guilty	alcohol beer drive
drive	driving walk push	car fast way
walk	walking move run	run talk stroll
run	running play move	walk jog fast
sleep	talk drink bed	dream rest awake
talk	speak talking leave	speak listen chatter
leave	stay talk stop	come go arrive
live	lived stay play	die life dead
play	playing played move	fun ball game
move	turn moved talk	leave away stay
ball	shot match straight	bat round throw
throw	pull pick push	ball catch toss
catch	pick throw pull	fish throw ball
fish	animals birds species	water swim sea
water	food light air	drink cool wet
food	material water land	eat drink hunger
eat	talk pick lose	food drink fat
fat	sugar butter diet	skinny thin cat

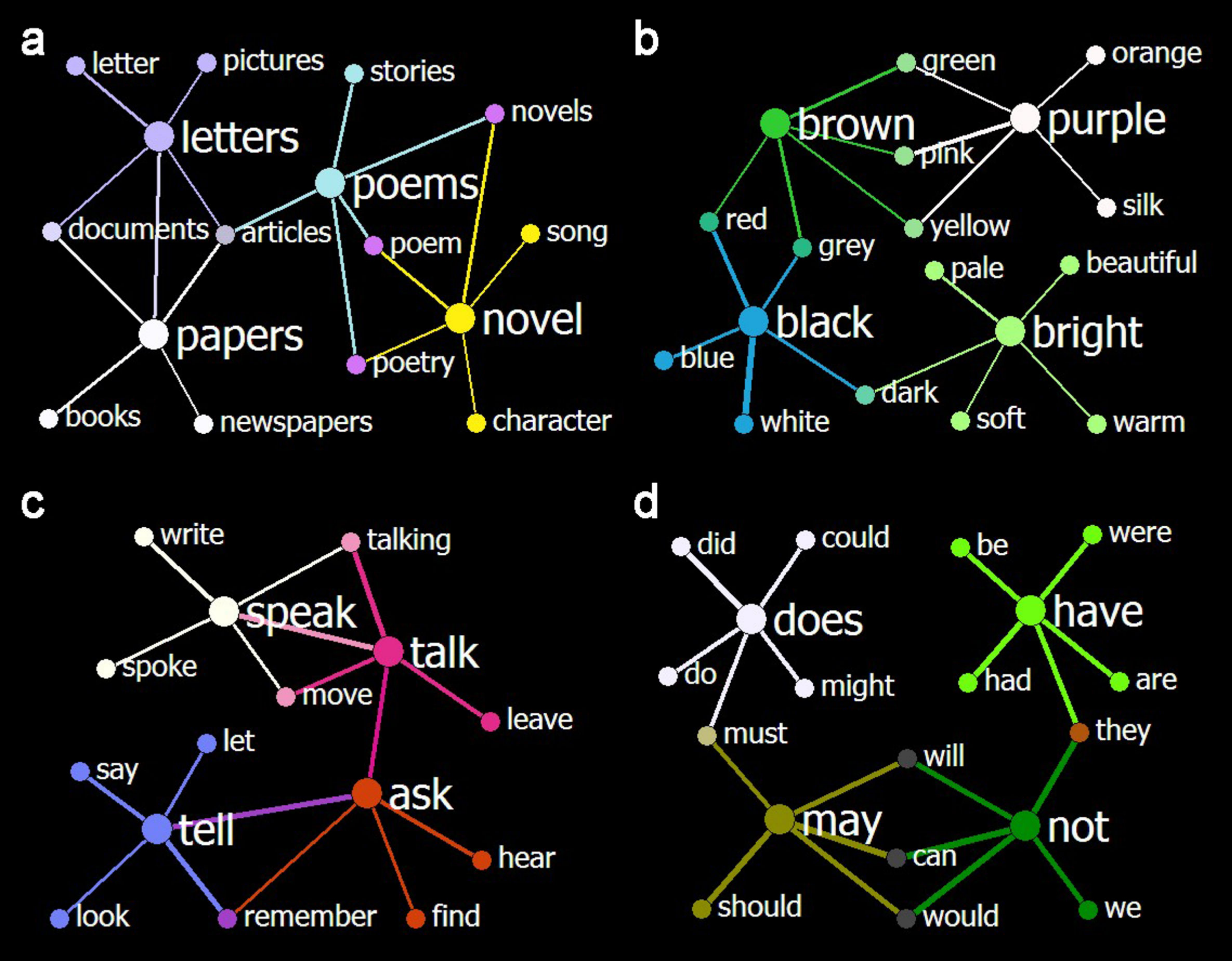
Except the first one, the targets are iteratively chosen from the previous associates or neighbors.

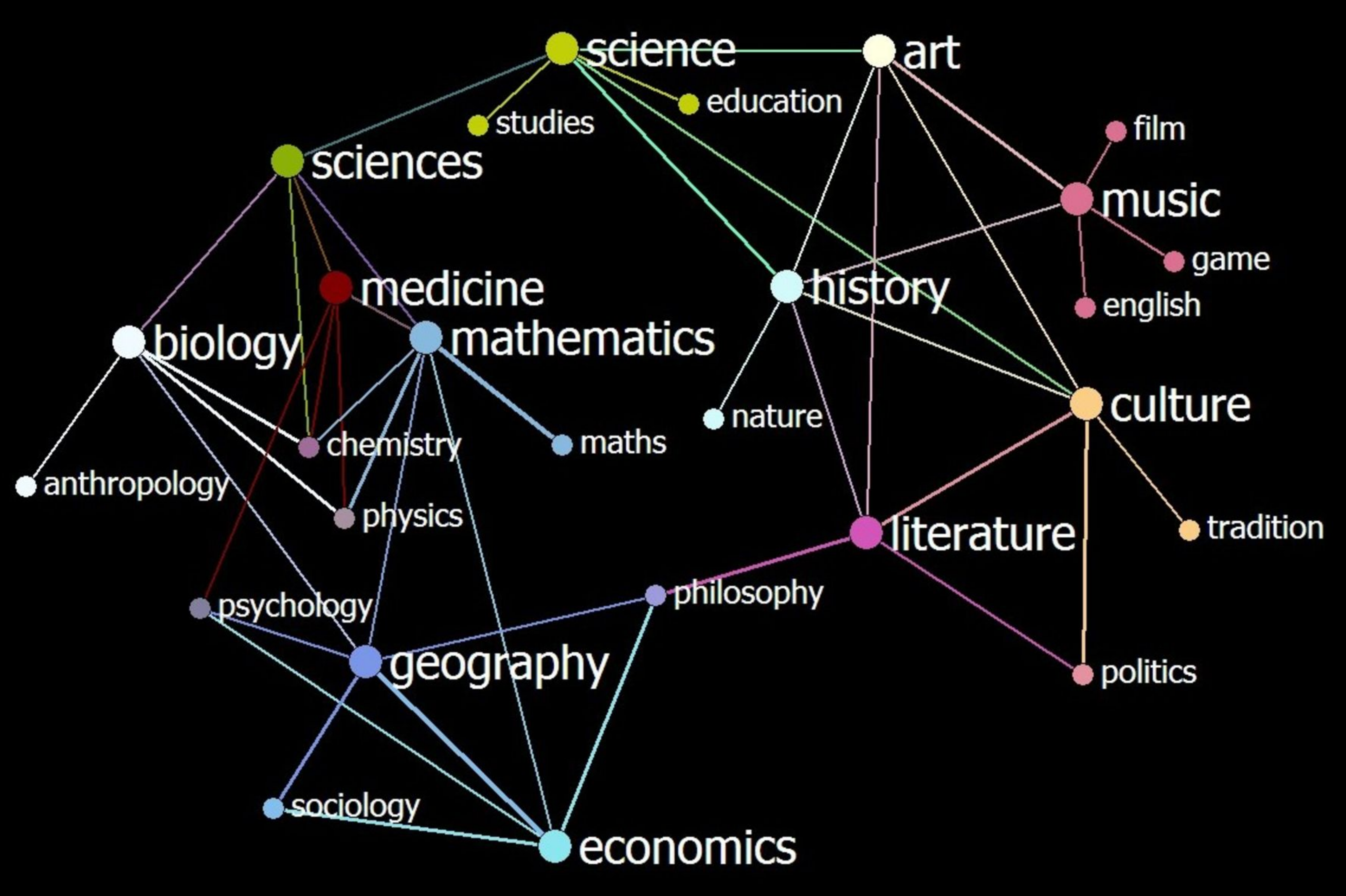
Figure 1 | Three ordinary nodes and their in- and out-links in the sparse WAF. For every node (word), the strongest 6 and the weakest 1 in- and out-links are presented, showing the sharply descending strengths and the most forceful restraints to the meanings of the nodes. **a,** ‘hands’ (noun, 164 in-links and 141 out-links in total) is characterised by the forceful links of modifiers (*his, her, your*), corresponding verbs (*shook, shake, shaking*), associates (*pockets, knees, hips*), etc. **b,** ‘live’ (verb, 129, 153) by the links of subjects (*who, people, we, they*), syntactic restraints (*to, in, with, here, alone, happily, where*) and associates (*births*). **c,** ‘scientific’ (adjective, 70, 185) by the links of the words composing phrases (*research, knowledge, method, interest, journals*), near-synonyms (*technological, mathematical*), and syntactic restraints (*of, the, a*). The unbalanced link strengths can be seen, for example, the strong in-links of ‘hands’ and the weak in-links of ‘scientific’ are in contrast. Note that the coloured strengths are at exponential scales.

Figure 2 | 6-word clusters and their local connections identified by our found *waf*s and affinity measure. Centre nodes in the clusters are in the bigger size for the eye. The nodes and links belonging to the same cluster are in the same colour except those that are shared by more than one cluster, whose colours are mixed. The thickness of a link represents the affinity between its nodes, ranging from 0.07 (*novel-poetry*) to 0.21 (*have-had*) in this figure. The length of a link means nothing. **a,** Local connections of noun clusters (related to literature). **b,** Adjective-noun clusters (colour). **c,** Verb clusters (talking). **d,** Function word clusters. Besides the plausibility of the clusters and their connections, the strong links between function words are notable.

Figure 3 | A complex sub-network composed by 6-word clusters in the domain of *science and art*. The sub-network shows that the affinities between the words are highly sensible, while the hierarchies in the word networks are intricate. In the sub-network, various types of hierarchies are included, for example the vertical hierarchies (*science, sciences, mathematics; art, culture, tradition*), the flat hierarchies (*history, culture, literature; mathematics, geography, economics*), and the hybrid hierarchies (*art, history, music; art, literature, politics*). The affinities range from 0.06 (*medicine-psychology*) to 0.13 (*mathematics-maths*).


a
his
her
shook
shake
your
shaking
a
hands
pockets
knees
hips
on
shaking
of
swept
b
to
who
people
we
they
where
humans
live
births
in
with
here
alone
happily
downstairs
c
special
of
sites
mathematical
the
a
or
scientific
technological
research
knowledge
method
interest
journals
forecasts
10^{-3}
10^{-4}
10^{-5}
10^{-6}
Link Strength


a
letter
pictures
stories
novels
letters
poems
documents
articles
poem
song
papers
novel
poetry
books
newspapers
character
b
green
orange
brown
purple
pink
red
silk
yellow
grey
pale
beautiful
black
bright
blue
dark
white
soft
warm
c
write
talking
speak
talk
spoke
move
leave
let
say
ask
tell
hear
look
remember
find
d
did
could
be
were
does
have
do
might
had
are
must
they
will
may
not
can
should
would
we


science
art
education
studies
sciences
film
music
game
medicine
history
english
biology
mathematics
nature
culture
chemistry
maths
anthropology
physics
literature
tradition
philosophy
psychology
geography
politics
sociology
economics