**RESEARCH**　　　　　　　　　　　　　　　　　　　　　**Open Access**

# Thinking spatially in computational social science

## Commentary on Yong-Yeol Ahn (2021): "Representation learning for computational imagination"

Aliakbar Akbaritabar[1*] (iD)

*Correspondence:
akbaritabar@demogr.mpg.de
[1]Max Planck Institute for
Demographic Research (MPIDR),
Konrad-Zuse Strasse 1, 18057
Rostock, Germany

**Abstract**

Deductive and theory-driven research starts by asking questions. Finding tentative answers to these questions in the literature is next. It is followed by gathering, preparing and modelling relevant data to empirically test these tentative answers. Inductive research, on the other hand, starts with data representation and finding general patterns in data. Ahn suggested, in his keynote speech at the seventh International Conference on Computational Social Science ($IC^2S^2$) 2021, that the way this data is represented could shape our understanding and the type of answers we find for the questions. He discussed that specific representation learning approaches enable a meaningful embedding space and could allow spatial thinking and broaden computational imagination. In this commentary, I summarize Ahn's keynote and related publications, provide an overview of the use of spatial metaphor in sociology, discuss how such representation learning can help both inductive and deductive research, propose future avenues of research that could benefit from spatial thinking, and pose some still open questions.

**Keywords:** Computational social science; Data representation; Word embedding vectors; Spatial metaphor; Mobility

## 1 Introduction

Computational Social Science (CSS) [1, 2] is a scientific discipline where computational tools and techniques are used to answer research questions that could be *social* in nature. This area has attracted two groups of scientists [3]: a) social scientists with computational skills, and b) computational scientists with an interest in questions related to social phenomena. The advent of social media and online social networks has led to increased digitization of human interactions and exponential growth of the digital traces left behind in these interactions [3–5]. Both of these groups of scientists know the value of these digital trace data and have the skills to analyze them [6, 7].

Generally, research projects, including in CSS, could follow two main approaches: data-driven (inductive) or theory-driven (deductive) [4, pp 61-62]. While the data-driven approach might start from an *inductive* or bottom-up exploration of the data to find general patterns, the theory-driven approach -predominantly used in social sciences- usually starts with a question and follows a few steps. The order of these steps might be modified based on research circumstances and should not be considered linear as one might have a few backs and forts between specific steps. *Question and theorize*: first, one needs to ask relevant questions and identify the data that could help in answering them. Theorizing tentative answers for these questions based on previous findings presented in the literature guides the empirical investigation and hypothesis testing. These theories and tentative answers could come from different disciplines such as sociology, political sciences, and psychology, to name a few. *Gather and pre-process data*: to answer research questions, one needs data. Gathering data in CSS could mean web scraping, sending requests to large-scale databases using Structured Query Language (SQL), building Application Programming Interface (API) solutions or connecting to them through programming languages, or building an Agent-Based Model (ABM) [8, 9] or using other types of simulation methods to create artificial data. Online surveys [10, 11] are designed and run e.g., by recruiting respondents through social media advertisement [12]. Online experiments [13, 14] could be another way of gathering data. Organizing, reshaping, cleaning, and making the data ready to be used is next. It can be lengthy and cumbersome if the previous step has not been done reasonably since the digital trace data is not always created for research purposes [4, 15]. *Model*: one then finds the best way to model the data to prepare answers to the research questions and compare these answers with the tentative ones from the literature. Modelling the data could be done with an ever-growing set of approaches such as statistical, relational (e.g., network analysis), structural approaches or text analysis and topic modelling depending on the nature of the data and questions. *Report and publish*: At the end, one prepares scientific manuscripts describing the above steps, the main findings, and whether they support or refute the previous findings. Of course, this process could be continued by future extensions, replication or refutation of research results.

In sum, in each of these steps of a deductive research project, a researcher makes decisions and chooses tools and techniques to achieve their goals. The type of data that is gathered, the data representation method that is used, and the way the data is modelled [16] could affect the way we think about the phenomenon and bind our understanding and imagination. On the other hand, inductive research follows a different set of steps and workflow. For instance, there are representation learning approaches used in inductive research that use the data, learn its features, and enable specific interpretation of this data by creating an embedding space. The main focus of Yong-Yeol Ahn's keynote [17] is a call to spend more time on the implications of such embedding space, its meaning and interpretability, and going beyond the boundaries of our computational imagination in the CSS research process.

## 2 Summary of Ahn's keynote: "Representation learning for computational imagination"

Yong-Yeol Ahn delivered a keynote speech at the seventh International Conference on Computational Social Science (IC$^2$S$^2$), held in Eidgenössische Technische Hochschule Zürich (ETH Zürich) in 2021 [video recording available in 17].

Ahn started his keynote by indicating that the three usual steps in the workflow of a CSS [1, 2] research project with the inductive approach are data, features (representation) and models. While in the computational side of CSS and for a long time different methods of representation learning are developed and used [e.g., see a review in 18], still much emphasis has been on the third step and models. This is because the modelling step answers our questions. He calls for an outside-the-box thinking in the *representation* step before we move onto the modelling step and encourages using representation learning methods by providing illustrative examples from his and others' works. The gap in CSS research that Ahn's keynote fills is in the data representation step and his main suggestion is to focus on the embedding space that is obtained from the representation learning. He emphasizes that this embedding space could be interpreted as a meaningful space with orientation, and could enable adopting a spatial metaphor as described below.

Ahn's keynote and its title were motivated by the *sociological imagination* of C. Wright Mills [19]. Mills emphasized the importance of the relationship between personal experience and the wider society which required designed research frameworks to gather data and study it. This is similar to the deductive approaches in the research described in the previous section. With the advent of the Internet, online contexts appeared and hosted people's lives. Information and communications technologies (ICTs) allow for gathering the digital trace of people's online presence [20] leading to new data sources. New tools were developed to leverage these data sources and led to an increased reputation for computational sociology [3, 5, 14] with which we can expand the sociological imagination to the online world. This new *observational* data sources are not always structured and cleaned as they are not built by scientists for research [4, 15]. Representation learning methods that are mainly developed in computational sciences could learn features of this unstructured data and encode them into an embedding space. A different set of models could take these embedding space and decode them and generate output data which has features similar to the used and encoded input. Computational imagination in Ahn's keynote means the ability to be open-minded and delving deeper into the data representation space. This is a lesson from deep learning that machines can learn the representation of the data. He asks if we can interpret this representation space as a literal and physical space. This enables adopting a spatial metaphor. While the data representation lives in a vector space, he asks if we can interpret this vector space literally as a space?

As one example of such a meaningful embedding space, he refers to the use of word embedding vectors, or word2vec [21, 22], to construct sentences by sequences of events that are similar to words in a sentence [23]. These vectors could represent different phenomena and the closeness or distance between these vectors identifies similar or dissimilar things. As an example, two sets of related data i) country names and ii) country capital cities could be represented as two vectors. Word2vec models find similarities between these vectors and place a country name closer to their capital city in the vector space. He adds that the idea of word2vec is not new, but now we have enough computational power to use representation learning of data in the form of vectors in a high-dimensional space and model them.

He continues to highlight the benefits of focusing on the interpretability of the embedding space in terms of having a meaningful "orientation" i.e., a sense of direction in space. Using this orientation, one can speak about directions and the relative placement of objects in a landscape. He describes that this spatial interpretation of the embedding space

is intuitive to us as "we think and imagine spatially". He gives examples of how in our language, we use this direction-based view to describe things similar to metaphors we live by [24]. For instance, *up* is good while *down* is bad. Or *right* is good while *left* could be bad. This spatial interpretation could be applied to a space of words, materials, faces, and knowledge to name a few. To adopt such a spatial metaphor, we need meaningful orientations and some representation learning methods enable having such an orientation. As an example, the word2vec type of vectors allows constructing semantic axes to facilitate this orientation and movement in this space.

Ahn is encouraging CSS researchers to pay more attention to the interpretability of the data representation and embedding space. By doing so, representation learning could enable the adoption of a spatial metaphor. He emphasizes the importance of choosing the right observation context and representation learning for the phenomenon, as it will shape what we see and understand.

Ahn and his collaborators have previously used network-based and relational representation of data [e.g., 25–28] and in this keynote, he emphasizes that while a network-based representation is useful, a spatial interpretation of the embedding space could expand it further. In comparison to network models, using this spatial metaphor to interpret the representation space allows an easier understanding and modelling of movement. Using networks, it is harder to represent and model a very high-dimensional movement similar to animals foraging a space. The nodes in a network and the edges between them, as sequences of nodes, could similarly be represented as vectors in the semantic space by following a random walk i.e., a node2vec. Then, the same spatial logic could be used to analyze this embedding space further and combine them with other vectors and attributes that might not be in a network format.

Representation learning, for instance in deep learning type of models, uses vectors embedded in a high-dimensional vector space and Ahn asks whether we can interpret this embedding space "literally" as a space. As an example, using sentences in word2vec allows us to consider sequences of words, and when these sentences are meaningful, we can model the probability of each word by knowing the previous or surrounding words. This could be considered similar to Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformer (GPT) language models. However, large language models could be more sophisticated than the simplified example discussed here. Contextualizing words by investigating other words used alongside them in the same context allows one "to know a word by the company it keeps" [17]. Word2vec is a combination of sequences and tokens and this simple idea can be extended to many different use cases and contexts.

Ahn and his collaborators have used this type of spatial interpretation of representation space in various domains. In their "science genome" project, they look at the space of knowledge and where are disciplines located in the space [29]. As an example, imagine where sociology would be placed on a landscape relative to math and physics. As another example about scientific journals and conferences, what could be the result of adding sociology to a specific computer science conference, e.g., The International Conference on Web and Social Media (ICWSM) by Association for the Advancement of Artificial Intelligence (AAAI). What journal or event would be closest to the outcome of this addition? In his keynote, he provides examples of such hypothetical addition or subtractions. In a separate set of works, their group investigates mobility and how scientists move in the

geographical, physical, linguistic and organizational space represented as different sets of vectors in a high-dimensional vector space [23].

Ahn goes beyond that and asks "how far can we push this spatial metaphor and spatial interpretation?" He presents examples of "frames" and "framing" that could essentially be semantic axes similar to a yardstick allowing us to measure bias towards specific frames in space. For instance, the legality frame used in studies by him and collaborators [30, 31], which shows how the meaning of a word could change by the community that uses it. The new meaning could over time become prevalent and substitute previous ones and he gives "awful", "gay" and "broadcast" as examples of words whose meaning has changed over time which was driven by the community who has been using them with specific intended meanings.

Ahn shows examples from their study of different types of mobility, including scientists' moves between academic organizations [23], that as a result of considering the embedding space as a literal space with meaningful axes, word2vec could be considered an equivalent to the gravity model [32]. This is because of the mathematical properties of word2vec that align well with the gravity law of mobility. It places more similar objects closer to each other in the vector space. Similar to gravity law that the size of the population in origin and destination pairs of locations could explain the probability of movements [32].

On the limitations of such a spatial interpretation of the representation space, Ahn adds that the embedding space constructed is highly dependent on the quality of the data. If one inputs garbage, what comes out of the representation learning is a non-meaningful semantic landscape and leads to a garbage output with low-quality results. Furthermore, we should be mindful of algorithmic bias, e.g., in the process of data generation or storage of the digital traces [4, 33], that could affect the data and representation learning of it. Ahn quotes from the literature the example of face recognition algorithms that fail at detecting black faces.

To sum up, Ahn's keynote, the step-by-step discussion of the pros and cons of the use of representation learning methods which enables a spatial approach, word2vec models as one example of such representation learning of data in CSS are a good example of speculating about different steps of research and how they can be improved beyond the usual methods and traditions. This in-depth look at the data representation step is necessary as it shows how different representation learning methods [18] could enable or bind the way we see the phenomenon, the methods we use to model it [16], our results, and their interpretation.

## 3  Thinking spatially and use of spatial metaphor in social sciences

My background is in [computational] sociology and sociology of science. In this section, I provide a summary of the uses of spatial metaphor in sociology and methods or theories of social research [34, 35].

It is important to note that research projects in sociology could predominantly follow a deductive approach, as described in the introduction, with exceptions such as the grounded theory that could be considered a data-first approach [4, pp 61-62]. This type of research traditionally uses *designed data* that are specifically created and carefully gathered for research. An example is designing a survey and gathering responses to empirically test a theory and hypothesis. Thinking spatially in the deductive approach entails that depending on the research question and adopted theory, some concepts might be formulated in spatial form. This means the researcher has a spatial view of the phenomenon

at the start which guides the operationalization of the concepts, measurement and data-gathering step. For instance, the researcher might ask geometrically relevant questions about distance among objects under study and this spatial view of the phenomenon guides the modelling step. This is similar to a deductive view to network analysis that depending on the research question, the researcher formulates and gathers a different type of network data, and defines nodes and ties in a form that might be different from other studies.

The use of the spatial metaphor in sociology as described by Silber [36] is a distinguishable feature of contemporary social theory, which could be attributed to the increased availability of spatial data [37]. Nevertheless, I need to emphasize that, based on Healy and Moody's [38] description, sociology seems to lag in terms of the best practices in visualization to represent data and communicate insights. In addition, my description of uses of spatial metaphor in sociology is not an attempt at establishing precedence [39].

Social theorists use a spatial metaphor to show how concepts could be positioned and interpreted in relative placements to and from each other. For instance, Bourdieu [chapter three of 40, 41, p 452] used a spatial metaphor to show how different forms of capital could be represented in the same social space allowing to cluster different groups of people possessing this combination of capitals (e.g., a combination of high cultural capital with low economic capital as an illustrative example).

Schelling's [42, 43, pp 135-166] segregation model could be considered as one of the most famous uses of a spatial metaphor in social research to explain the macro patterns of ethnic and racial segregation by investigating micro individual behaviors and the process of emergence leading to this macro phenomenon [9, 44]. In his formulation, the relation between individuals and groups is defined, observed and modelled in a spatial setup that enables a macro picture of the segregation.

The landscape view of the scientific fields and space-based visualization has been used often in the sociology of science, scientometrics and bibliometrics. For instance, the Visualization of Similarities (VOS) algorithm [45, 46] uses co-occurrence matrices and places topics used often together closer on a spatial landscape and farther from less-frequently-used topics. This is complemented by a network of terms co-occurrence that is a network representation of the simultaneous usage of these terms in a document by adding ties between them. Contour maps are used also to represent scientific fields, concepts, and closer/farther distances between them [47] or to overlay it over the networks of terms or collaborations to enable a spatial interpretation of closeness or distance [48]. This use of contour maps is similar to the works by geographers of science [e.g., 49–51] which show how scientific collaboration networks could be overlaid over their spatial context and represent the mutual effect of the network ties and geographical distance [52–54] and similar to the works reviewed by Small and Adler [34] which investigate the role of spatial context in affecting the formation of network ties.

The notion of *organizational foci* with a local or global focus [55] has been used to investigate whether identification with the global foci could enable a boundary-crossing behavior in advice-seeking of employees in an organizational context that crosses the formal hierarchies and boundaries of the organizational affiliation. This highlights how location and positioning could be defined with *local* and simultaneously with *global* criteria and framing practices.

The concept of core and periphery in the world system theory represented as concentric circles [56] and positioning of different actors (e.g., countries) in a landscape as peripheries surrounding the core and semi-periphery is similarly using a spatial metaphor.

To sum up, these were only some examples highlighting that in deductive research in social sciences, a spatial view is often used in formulating concepts at the start of the research. The use of spatial metaphor is not new and different theorists or scholars have used a metaphor of landscape to represent their ideas. But the metaphors used by them are rather different from the representation learning discussed by Ahn [17]. As explained earlier, in representation learning, a spatially meaningful representation is adopted to the data to learn its features and the outcome is an embedding space, which as Ahn encourages CSS researchers to do by providing word2vec as an illustrative example, could be interpreted as a literal space with meaningful orientation.
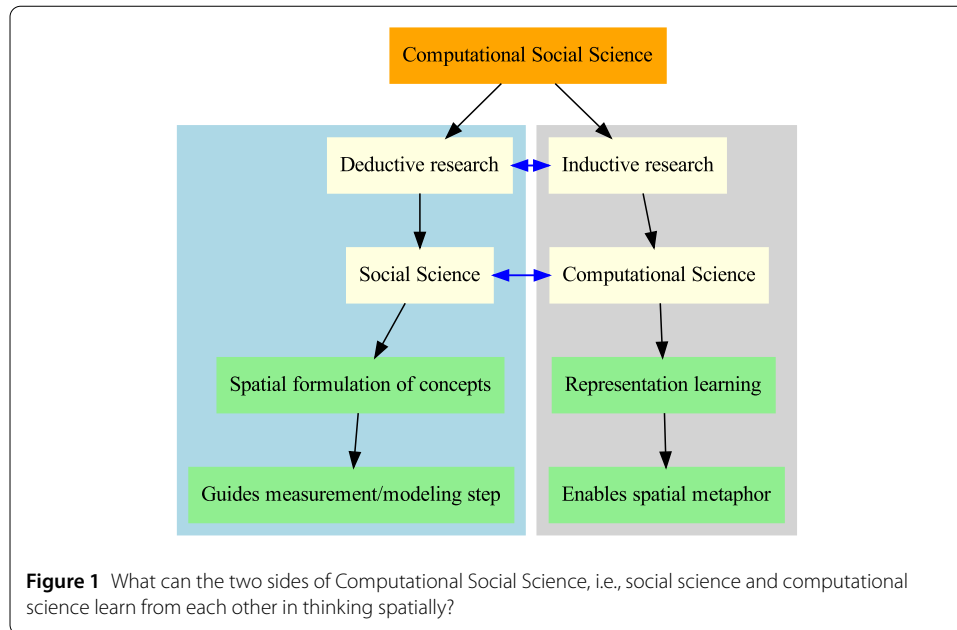
See [57] as an illustrative example that combines the inductive approach by using word embedding vectors trained on large corpora of text to measure different dimensions of the concept of "social class" in culture and its change, and the deductive approach in using a survey to empirically evaluate the results of the word2vec and the distances between vectors obtained using a questionnaire that respondents filled out on Amazon MTurk.

Briefly, in social research, principally, data is designed and gathered for research purposes and the spatial metaphor is a way of thinking about the concepts, gathering data, and presenting the relationship among concepts and results. But representation learning that Ahn discusses adapts itself to the data and learns from the data features and enables a spatial metaphor to be adopted. These go in two different directions since in the former, data is gathered in the specific way that the researcher designs to answer the research question and to test a hypothesis, i.e., if the conceptual formulation is spatial, the data gathering, operationalization and measurement considers this and one would ask about distance and geometrically relevant questions. In the latter, data could be observational, already gathered, and sometimes it is not built for research [4], and the researcher uses a spatially aware representation learning for it to find general patterns in this data in an inductive way. I think that the computational and social science side of CSS could learn from these two seemingly different but in my opinion complementary approaches. I elaborate on how this learning can happen in future research in the next section.

## 4 Future direction of research

It is necessary to reiterate, as Ahn does in his keynote, that the use of representation learning and specifically word2vec models as an example is not new. But the computational power and tools that are now widely spread and accessible allow better adoption of Ahn's suggestion in interpreting the embedding space as a meaningful space enabling answering spatial questions. In addition, Ahn's focus on how word2vec and similar representation learning methods [18] work and probing into their mathematical properties i.e., that word2vec aligns well with gravity law of human mobility [23], is a good example of the need to investigate how deep learning and similar black-box models function.

I think the two approaches to research in the CSS community, i.e., inductive and deductive could lend concepts, tools and techniques to each other as depicted in Fig. 1 to help in understanding how these representation learning methods function. For instance, I think if a deductive research question with a spatial formulation of concepts is adopted to be tested by inductive approach and using observational data and representation learning of

**Figure 1** What can the two sides of Computational Social Science, i.e., social science and computational science learn from each other in thinking spatially?

the data could allow testing whether the spatial qualities and hypotheses formulated deductively are observed and proven in the obtained embedding space? This could be similar to the research framework that was employed by [57] to cross-validate the results of word embedding with a survey.

While the use of representation learning and as an example word2vec models offer possibilities and avenues to nurture our creativity and broaden CSS researchers' computational imagination, there are still open questions that could be investigated in future research. Broadly, what other existing representation learning methods [18] create an embedding space that could enable a meaningful spatial interpretation? For the specific case of word2vec models, we could ask how robust, replicable or stochastic are the results of these models. Are there evaluation methods for the results of these models? How different are the results of word2vec models in comparison to text analysis methods such as topic modelling or less elaborated analysis such as noun-phrase-clause identification methods? Are the similarities found by topic models, that group similar concepts under the same topic, comparable to a shorter distance between vectors in word2vec? How can we evaluate the *meaningfulness* of the semantic landscape obtained from the word2vec model? Some variables such as gender or occupation are nominal or categorical and discrete variables, by using a spatially meaningful representation learning method, e.g., word embedding vectors, these variables could be represented as a continuum (vector) and allow learning of features of multiple variables simultaneously. How is the interaction between these variables different from more traditional modelling approaches and does this interaction give a higher explanation power to our models? Most of the previous uses of spatial metaphors (e.g., the ones described above) have a two-dimensional definition of the landscape and space. What other types of space could be used, for instance, three or multi-dimensional spaces, and do they increase our analytical power or unnecessarily increase the complexity?

One question that follows after adopting Ahn's [17] suggestion in interpreting the embedding space in spatial terms is how to measure the distance between different points

**Figure 2** Different measures of distance [58, image is used by the original author's permission]
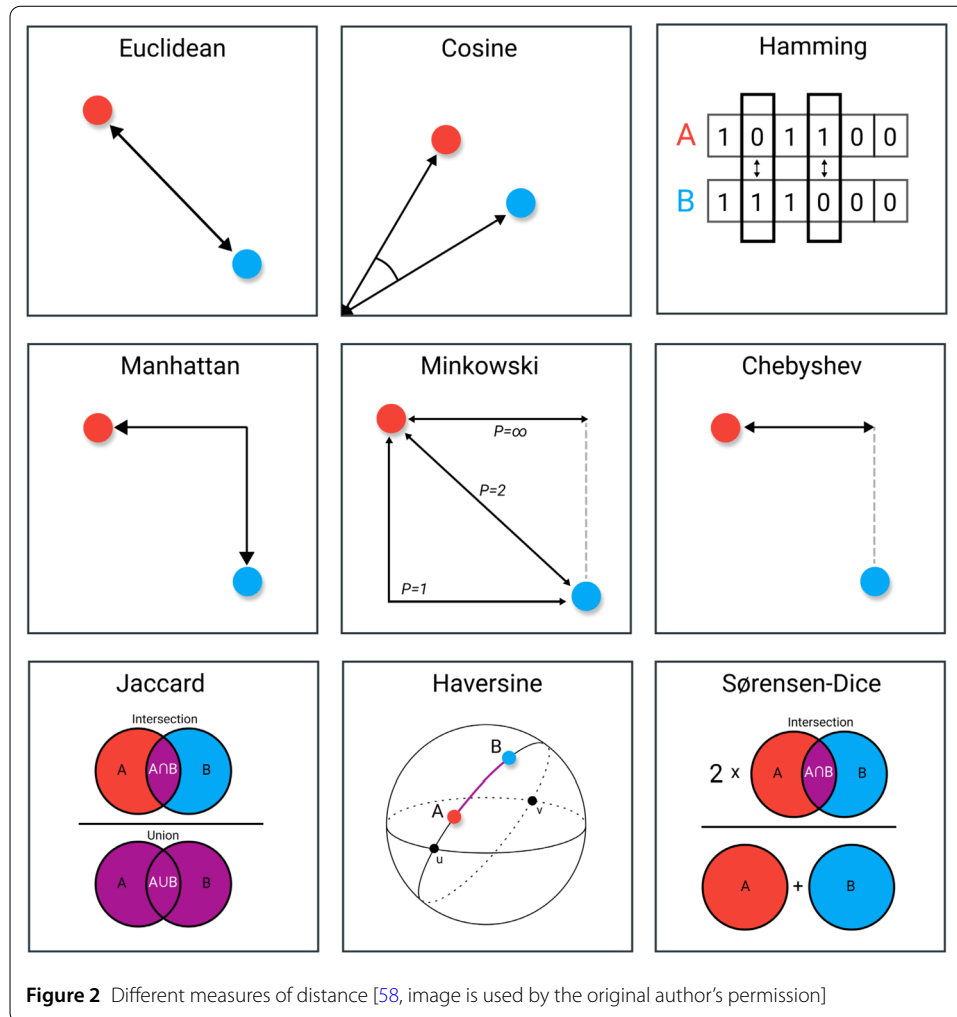
in the landscape [37]. There are different measures of distance proposed in the literature e.g., see a visualization in Fig. 2 [image is used by the original author's permission from 58]. Which definitions could we use and how are they going to affect the representation learning and understanding of the spatial landscape? Some of the distance measures assume an equal distance between any two points, i.e., a symmetrical definition, but the spatial metaphor allows considering asymmetric definitions of distance. Logan [37] advocates that the uses of different measures of distance should be informed by theory. How are these uses going to affect the representation learning and the embedding space and results generated by the model?

## 5 Conclusions

In this commentary, I summarized Ahn's keynote speech [17] and his suggestion to use representation learning methods which enable a meaningful interpretation of embedding space and allow adopting a spatial metaphor in CSS inductive research projects. I provided an overview of the use of spatial metaphors in deductive research in social sciences and sociology which guides the formulation of concepts, data gathering and spatial interpretation of results. I introduced some future avenues of research and how the social science and computational science side of CSS could learn from each other.

Interpreting the embedding space generated by the representation learning in a spatial framework as encouraged by Ahn [17] could allow for resolving some of the still open paradoxes in science. As an example from my area of research, using the example of word2vec models might help in resolving the direction of causation between scientific mobility and collaboration. Some works in the literature have advocated for an effect of mobility on collaboration [59], while others have found a bidirectional [60, 61] or an inconclusive [62] effect between the two. Using word embedding vectors for the trajectory of mobility and collaboration (similar to Ahn's research e.g., [23]) might help to resolve this paradox by enabling the consideration of multiple vectors (e.g., for mobility, collaboration, and similar other factors such as geographic distance or linguistic and cultural similarities) in a unified framework to compare the strengths of their effects.

**Abbreviations**
AAAI, Association for the Advancement of Artificial Intelligence; ABM, Agent-Based Model; API, Application Programming Interface; BERT, Bidirectional Encoder Representations from Transformers; CSS, Computational Social Science; ETH Zürich, Eidgenössische Technische Hochschule Zürich; GPT, Generative Pre-trained Transformer; $IC^2S^2$, International Conference on Computational Social Science; ICTs, Information and communications technologies; ICWSM, The International Conference on Web and Social Media; SQL, Structured Query Language; VOS, Visualization of Similarities.

**Data availability**
Not applicable.

# Declarations

**Competing interests**
The author declares no competing interests.

**Author contributions**
AA: Conceptualization, Writing – original draft, Writing – review & editing. The author read and approved the final manuscript.

**Authors' information**
Aliakbar Akbaritabar is a computational social scientist with a background in sociology. He is currently a research scientist and the Deputy Head for Training at the Department of Digital and Computational Demography at the Max Planck Institute for Demographic Research (MPIDR). Previously, he was a postdoctoral researcher at the German Centre for Higher Education Research and Science Studies (DZHW) and German Competence Centre for Bibliometrics. He holds a PhD in economic sociology and labor studies from the University of Milan, and a PhD in sociology (social welfare) from the Allameh Tabataba'i University of Tehran. His work focuses on themes related (but not limited) to the science of science, scholarly migration, social networks, collaboration networks, and computational social science.

**References**
1. Lazer D, Pentland A, Adamic L, Aral S, Barabasi A-L, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, Jebara T, King G, Macy M, Roy D, Van Alstyne M (2009) SOCIAL SCIENCE: computational social science. Science 323(5915):721–723. https://doi.org/10.1126/science.1167742
2. Lazer DMJ, Pentland A, Watts DJ, Aral S, Athey S, Contractor N, Freelon D, Gonzalez-Bailon S, King G, Margetts H, Nelson A, Salganik MJ, Strohmaier M, Vespignani A, Wagner C (2020) Computational social science: obstacles and opportunities. Science 369(6507):1060–1062. https://doi.org/10.1126/science.aaz8170
3. Edelmann A, Wolff T, Montagne D, Bail CA (2020) Computational social science and sociology. Annu Rev Sociol. https://doi.org/10.1146/annurev-soc-121919-054621
4. Salganik MJ (2018) Bit by bit: social research in the digital age. Princeton University Press, Princeton. OCLC: on1012406622

5. Evans J, Foster JG (2019) Computation and the sociological imagination. Contexts 18(4):10–15. https://doi.org/10.1177/1536504219883850

6. Alburez-Gutierrez D, Zagheni E, Aref S, Gil-Clavel S, Grow A, Negraia DV (2019) Demography in the digital era: new data sources for population research. SocArXiv

7. Kashyap R, Rinderknecht RG, Akbaritabar A, Alburez-Gutierrez D, Gil-Clavel S, Grow A, Kim J, Leasure DR, Lohmann S, Negraia DV, Perrotta D, Rampazzo F, Tsai C-J, Verhagen MD, Zagheni E, Zhao X (2022) Digital and computational demography. Technical report, SocArXiv. https://doi.org/10.31235/osf.io/7bvpt. https://osf.io/preprints/socarxiv/7bvpt/. Accessed 2022-04-21

8. Gao D, Akbaritabar A (2021) Using agent-based modeling in routine dynamics research: a quantitative and content analysis of literature. Rev Manag Sci. https://doi.org/10.1007/s11846-021-00446-z

9. Macy MW, Willer R (2002) From factors to actors: computational sociology and agent-based modeling. Annu Rev Sociol 28(1):143–166. https://doi.org/10.1146/annurev.soc.28.110601.141117

10. Akbaritabar A (2022) Dichotomy of offline versus online social capital, Rochester, NY. https://papers.ssrn.com/abstract=4311832. Accessed 2022-12-26

11. Akbaritabar A, Hezarjaribi J, Jullien N (2015) A tool for dynamic measurement of social capital embedded in online social networks. Grafo Working Papers 4(0):1–30. https://doi.org/10.5565/rev/grafowp.14. Accessed 2020-04-16

12. Grow A, Perrotta D, Del Fava E, Cimentada J, Rampazzo F, Gil-Clavel S, Zagheni E, Flores RD, Ventura I, Weber I (2022) Is Facebook's advertising data accurate enough for use in social science research? Insights from a cross-national online survey. J R Stat Soc, Ser A, Stat Soc. https://doi.org/10.1111/rssa.12948. https://onlinelibrary.wiley.com/doi/pdf/10.1111/rssa.12948

13. Boase J (2016) Augmenting survey and experimental designs with digital trace data. Commun Methods Meas 10(2–3):165–166. https://doi.org/10.1080/19312458.2016.1150975

14. Keuschnigg M, Lovsjö N, Hedström P (2018) Analytical sociology and computational social science. J Comput Soc Sci 1(1):3–14. https://doi.org/10.1007/s42001-017-0006-5

15. Lewis K (2015) Three fallacies of digital footprints. Big Data Soc 2(2):2053951715602496. https://doi.org/10.1177/2053951715602496

16. Torres AFC, Akbaritabar A (2022) Methodological monotheism across fields of science in contemporary quantitative research. https://doi.org/10.48550/arXiv.2208.05373. http://arxiv.org/abs/2208.05373. Accessed 2022-08-11

17. Ahn Y-Y (2021) Keynote: representation learning for computational imagination. In: 7th international conference on computational social science (IC2S2 2021), ETH Zurich. https://video.ethz.ch/events/2021/ic2s2/ea7a1b7c-fb20-4fa2-956b-8bc685db468e.html

18. Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35(8):1798–1828. https://doi.org/10.1109/TPAMI.2013.50

19. Mills CW (1959) The sociological imagination. Oxford University Press, London

20. McFarland DA, Lewis K, Goldberg A (2016) Sociology in the era of big data: the ascent of forensic social science. Am Sociol 47(1):12–35. https://doi.org/10.1007/s12108-015-9291-8

21. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. https://doi.org/10.48550/arXiv.1301.3781. http://arxiv.org/abs/1301.3781. Accessed 2022-12-28

22. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Proceedings of the 26th international conference on neural information processing systems – volume 2. NIPS'13. Curran Associates, Red Hook, pp 3111–3119

23. Murray D, Yoon J, Kojaku S, Costas R, Jung W-S, Milojević S, Ahn Y-Y (2021) Unsupervised embedding of trajectories captures the latent structure of mobility. arXiv:2012.02785 [physics]. Accessed 2021-07-10

24. Lakoff G, Johnson M (2008) Metaphors we live by. University of Chicago Press, Chicago

25. Ahn Y-Y, Bagrow JP, Lehmann S (2010) Link communities reveal multiscale complexity in networks. Nature 466(7307):761–764. https://doi.org/10.1038/nature09182

26. Park J, Wood IB, Jing E, Nematzadeh A, Ghosh S, Conover MD, Ahn Y-Y (2019) Global labor flow network reveals the hierarchical organization and dynamics of geo-industrial clusters. Nat Commun 10(1):3449. https://doi.org/10.1038/s41467-019-11380-w

27. Schich M, Song C, Ahn Y-Y, Mirsky A, Martino M, Barabási A-L, Helbing D (2014) A network framework of cultural history. Science 345(6196):558–562. https://doi.org/10.1126/science.1240064

28. Miao L, Murray D, Jung W-S, Larivière V, Sugimoto CR, Ahn Y-Y (2022) The latent structure of global scientific development. Nat Hum Behav 6:1206–1217. https://doi.org/10.1038/s41562-022-01367-x

29. Peng H, Ke Q, Budak C, Romero DM, Ahn Y-Y (2021) Neural embeddings of scholarly periodicals reveal complex disciplinary organizations. Sci Adv 7(17):9004. https://doi.org/10.1126/sciadv.abb9004

30. An J, Kwak H, Ahn Y-Y (2018) SemAxis: a lightweight framework to characterize domain-specific word semantics beyond sentiment. In: Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: long papers). Association for Computational Linguistics, Melbourne, pp 2450–2461. https://doi.org/10.18653/v1/P18-1228. https://aclanthology.org/P18-1228

31. Kwak H, An J, Jing E, Ahn Y-Y (2021) FrameAxis: characterizing microframe bias and intensity with word embedding. PeerJ Comput Sci 7:644. https://doi.org/10.7717/peerj-cs.644

32. Zipf GK (1946) The P1 P2/D hypothesis: on the intercity movement of persons. Am Sociol Rev 11(6):677–686. https://doi.org/10.2307/2087063

33. Wagner C, Strohmaier M, Olteanu A, Kıcıman E, Contractor N, Eliassi-Rad T (2021) Measuring algorithmically infused societies. Nature 595:197–204. https://doi.org/10.1038/s41586-021-03666-1

34. Small ML, Adler L (2019) The role of space in the formation of social ties. Annu Rev Sociol 45(1):111–132. https://doi.org/10.1146/annurev-soc-073018-022707

35. Swedberg R (2020) Using metaphors in sociology: pitfalls and potentials. Am Sociol 51(2):240–257. https://doi.org/10.1007/s12108-020-09443-3

36. Silber IF (1995) Space, fields, boundaries: the rise of spatial metaphors in contemporary sociological theory. Soc Res 62(2):323–355

37. Logan JR (2012) Making a place for space: spatial thinking in social science. Annu Rev Sociol 38:507–524. https://doi.org/10.1146/annurev-soc-071811-145531

38. Healy K, Moody J (2014) Data visualization in sociology. Annu Rev Sociol 40(1):105–128. https://doi.org/10.1146/annurev-soc-071312-145551
39. Holme P (2022) Good idea! But it's already been explored by geographers. https://petterhol.me/2022/06/04/good-idea-but-its-already-been-explored-by-geographers/. Accessed 2022-06-04
40. Bourdieu P, Wacquant LJD (1992) An invitation to reflexive sociology. Polity Press, Cambridge
41. Bourdieu P (1984) Distinction: a social critique of the judgement of taste. Harvard University Press, Cambridge
42. Schelling TC (1971) Dynamic models of segregation. J Math Sociol 1(2):143–186. https://doi.org/10.1080/0022250X.1971.9989794
43. Schelling TC (2006) Micromotives and macrobehavior. Norton, New York
44. Stadtfeld C (2018) The micro-macro link in social networks. SSRN Scholarly Paper ID 3211795, Social Science Research Network, Rochester, NY. https://papers.ssrn.com/abstract=3211795. Accessed 2020-04-20
45. van Eck NJ, Waltman L (2007) VOS: a new method for visualizing similarities between objects. In: Decker R, Lenz H-J (eds) Advances in data analysis. Studies in classification, data analysis, and knowledge organization. Springer, Berlin, pp 299–306. https://doi.org/10.1007/978-3-540-70981-7_34
46. van Eck NJ, Waltman L (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 84(2):523–538. https://doi.org/10.1007/s11192-009-0146-3
47. Moody J, Light R (2006) A view from above: the evolving sociological landscape. Am Sociol 37(2):67–86
48. Moody J (2004) The structure of a social science collaboration network: disciplinary cohesion from 1963 to 1999. Am Sociol Rev 69(2):213–238. https://doi.org/10.1177/000312240406900204
49. Maisonobe M, Grossetti M, Milard B, Eckert D, Jégou L, Hamilton P (2016) The global evolution of scientific collaboration networks between cities (1999–2014): multiple scales. Rev Fr Sociol 57(3):266–287
50. Maisonobe M, Eckert D, Grossetti M, Jégou L, Milard B (2016) The world network of scientific collaborations between cities: domestic or international dynamics? J Informetr 10(4):1025–1036. https://doi.org/10.1016/j.joi.2016.06.002
51. Bornmann L, de Moya-Anegón F (2019) Spatial bibliometrics on the city level. J Inf Sci 45(3):416–425. https://doi.org/10.1177/0165551518806119
52. Akbaritabar A (2021) A quantitative view of the structure of institutional scientific collaborations using the example of Berlin. Quant Sci Stud 2(2):753–777. https://doi.org/10.1162/qss_a_00131
53. Bidart C, Maisonobe M, Viry G (2022) Analysing personal networks in geographical space beyond the question of distance. Soc Incl 10(3):233–247. https://doi.org/10.17645/si.v10i3.5381
54. Akbaritabar A, Barbato G (2021) An internationalised Europe and regionally focused Americas: a network analysis of higher education studies. Eur J Educ 56(2):219–234. https://doi.org/10.1111/ejed.12446. https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejed.12446
55. Lomi A, Lusher D, Pattison PE, Robins G (2014) The focused organization of advice relations: a study in boundary crossing. Organ Sci 25(2):438–457. https://doi.org/10.1287/orsc.2013.0850
56. Wallerstein IM, Wallerstein SRI (2004) World-systems analysis: an introduction. Duke University Press, Durham
57. Kozlowski AC, Taddy M, Evans JA (2019) The geometry of culture: analyzing the meanings of class through word embeddings. Am Sociol Rev 84(5):905–949
58. Grootendorst M (2021) 9 distance measures in data science. https://www.maartengrootendorst.com/blog/distances/. Accessed 2022-10-06
59. Kato M, Ando A (2017) National ties of international scientific collaboration and researcher mobility found in Nature and Science. Scientometrics 110(2):673–694. https://doi.org/10.1007/s11192-016-2183-z
60. Boekhout HD, Traag VA, Takes FW (2021) Investigating scientific mobility in co-authorship networks using multilayer temporal motifs. Netw Sci 9(3):354–386. https://doi.org/10.1017/nws.2021.12
61. Chinchilla-Rodríguez Z, Miao L, Murray D, Robinson-García N, Costas R, Sugimoto CR (2018) A global comparison of scientific mobility and collaboration according to national scientific capacities. Front Res Metr Anal 3:17. https://doi.org/10.3389/frma.2018.00017
62. Netz N, Hampel S, Aman V (2020) What effects does international mobility have on scientists' careers? A systematic review. Res Eval 29(3):327–351. https://doi.org/10.1093/reseval/rvaa007

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.