

# A new tool in the toolbox to measure polarisation in networks? A recommendation for "Measuring changes in polarisation using Singular Value Decomposition of network graphs"

# *Mario Angst* based on peer reviews by *Yasaman Asgari* and 2 anonymous reviewers

Sage Anatasi, Giulio Dalla Riva (2025) Measuring changes in polarisation using Singular Value Decomposition of network graphs. arXiv, ver. 4, peer-reviewed and recommended by Peer Community in Network Science. https://doi.org/10.48550/arXiv.2403.18191

Submitted: 24 April 2024, Recommended: 14 May 2025

### Cite this recommendation as:

Angst, M. (2025) A new tool in the toolbox to measure polarisation in networks? A recommendation for "Measuring changes in polarisation using Singular Value Decomposition of network graphs". *Peer Community in Network Science*, 100217. https://doi.org/10.24072/pci.networksci.100217

Published: 14 May 2025 Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

Societal polarisation, which S. Anastasi and G. D. Riva are concerned about in "Measuring changes in polarisation using Singular Value Decomposition of network graphs" [1], has been a talking point and sometimes diffuse object of concern for a while. In parallel, the issue has received sci- entific attention from a variety of perspectives, ranging from political to complexity science and from normative to empirical approaches [2]. From an empirically oriented, applied social science perspective, a key area of discussion in the literature has been the measurement of polarization. Polarisation is a concept, which can be critically approached, not least because the there can be a danger of oversimplified normative orientations toward the value of social cohesion, which is not necessarily some sort of ideal state for a functional democracy, and also because sometimes the presence of conflict in a society may highlight unjust differences in material conditions [3]. Thus, measurement of polarization has to be put into context of how polarization is understood. It is here, where S. Anastasi and G. D. Riva [1] add a crucial element to the discussion, by pro- viding a suggestion for a measurement of polarization based on network data, and by shining a light on how a geographical focus of the literature on polarization may have led to suboptimal measurements. Seemingly, concern about polarization is an international phenomenon. However, the actual literature, as S. Anastasi and G. D. Riva [1] convincingly argue, is heavily influenced by research in the context of the United States of America, which has a legacy of seeing polarisation solely through the prism of differences between two political parties. S. Anastasi and G. D. Riva [1] are clear in their conceptual approach to polarisation. They focus on a bi-modal (two-group) understanding of polarization, but argue for a diagnosis of increasing polarization if two polarized groups diverge across different axes. In this way, their understanding of polarization also connects to research that aims to distinguish differing or overlapping political cleavages in societies [4] and is grounded by the researchers' own understanding of the context of New Zealand, where their research takes place. In this context, the proposal both of utilizing network data and Singular Value Decomposition (SVD) seems well justified and innovative, as is the suggestion to define the process of polarization as "the loss of dimensionality of a graph observerd over time" [1, p.9]. Differently put, network data can be unique in embedding cleavages in relational structure. And modeling social networks through a graph embedding (Random Dot Product Graphs in the case of the paper), enables the computation of both an optimal embedding dimension and SVD entropy (as a measure of network complexity), which link well to the specific understanding of polarization. S. Anastasi and G. D. Riva [1] demonstrate their approach with examples using both empirical and simulated data. For their empirical case study, the rely on data from the social media platform pre-viously knows as Twitter, showing an interesting finding of potentially increasing polarisation in New Zealand climate debates between 2017-2020 and 2020-2023. Given that the social media plat- form in question is not particularly useful for researchers as a data source on social interactions anymore, this makes one think about potential for future studies using the method introduced in the paper. S. Anastasi and G. D. Riva [1] provide the means to replicate their approach by means of Julia code, which should give a good starting point for future studies (although reproducibility would have been even higher through the provision of a way to reproduce the computational environment, such as Docker or Nix). As the proposed approach is both elegant and widely gen- eralizable to other networks, it would be fascinating to see it applied in different contexts. Bibliography [1] S. Anastasi and G. D. Riva, "The process of polarisation as a loss of dimensionality: measuring changes in polarisation using Singular Value Decomposition of network graphs." [Online]. Available: https://arxiv.org/abs/2403.18191 [2] S. A. Levin, H. V. Milner, and C. Perrings, "The dynamics of political polarization," Proceedings of the National Academy of Sciences of the United States of America, vol. 118, no. 50, Dec. 2021, https://doi.org/10.1073/pnas.2116950118. [3] D. Kreiss and S. C. McGregor, "A review and provocation: On polarization and platforms," New media & society, vol. 26, no. 1, pp. 556–579, Jan. 2024, https://doi.org/10.1177/14614448231161880. [4] E. Borbáth, S. Hutter, and A. Leininger, "Cleavage politics, polarisation and participation in Western Europe," West European politics, vol. 46, no. 4, pp. 631-651, Jun. 2023, https://doi.org/10.1080/01402382.2022.2161786.

# Reviews

# **Evaluation round #3**

DOI or URL of the preprint: https://doi.org/10.48550/arXiv.2403.18191 Version of the preprint: 3

## Authors' reply, 17 April 2025

Good afternoon,

Thank you for giving us the opportunity to fix these minor issues. The title has been updated, the code has been released with a ReadMe, and the correct DOI for the code repository has been added to the PCI submission. We have also added a couple more references.

Best regards, Sage Anastasi

## Decision by Mario Angst <sup>(i)</sup>, posted 31 March 2025, validated 01 April 2025

Dear authors,

Thank you again for taking the time to smoothing out the remaining rougher edges of this manuscript by addressing all the suggestions from the reviewers.

In my opinion, after reading through the point to point responses, remaining unaddressed concerns by the reviewers are mostly valid differences of style, in which case the authors should have the final say in my opinion.

For this reason, I have decided not to send the article out for review again and want to thank the two reviewers again at this point for their careful contributions to this manuscript. I am in the process of writing a recommendation for this manuscript, but still want to return it for one final revision to you, because of two remaining issues unrelated to content, which would however still make it impossible for me to recommend the preprint in its current form.

- The title of the referenced version for this on arXiv (https://arxiv.org/abs/2403.18191v3) has not been updated to the newer, shorter version, even though the track changes show it. Please make sure the version rendered for upload to arXiv aligns with the version referenced in PCI Netsci. Generally, v3 on arXiv does not show many of the changes to the paper outlined in track changes (for example the nice new entry paragraph). Make sure your next revision (4) matches the version referenced in PCI Netsci, which then will allow me to recommend it.
- Even though the move to zenodo has been a step in the right direction, reproducibility of the analysis currently does not adhere to minimum standards I would set for recommending a paper. Please, at the minimum:
  - separate issues of computational reproducibility from the manuscript text (see section "Code", where instructions for installing Julia packages are given). A replicator should be able to find all information on reproducing the analysis in the repository.
  - To this effect, add a README.md (or similar document with instructions for reproducing the analysis) to the zenodo repository, which describes, at minimum:
    - ★ contents of the repository
    - **\*** description of computational requirements and software needed
    - \* steps that a replicator needs to undertake to replicate the analysis
    - \* what elements of the results reported in the manuscript are reproduced
  - These are minimum requirements I would also encourage the use of Docker or Nix to actually ensure a larger degree of reproducibility of the computation environment.
  - To implement these changes, you may simply create a newer version of the repository on zenodo and point to it.

Best regards,

Dr. Mario Angst

## **Evaluation round #2**

DOI or URL of the preprint: https://doi.org/10.48550/arXiv.2403.18191 Version of the preprint: 2

## Authors' reply, 07 March 2025

Download author's reply Download tracked changes file

## Decision by Mario Angst <sup>(b)</sup>, posted 06 January 2025, validated 07 January 2025

Dear authors,

The two reviewers of your manuscript have again agreed to take the time to take a careful look at it. Again, in my opinion, their reviews meet the standard PCI network science sets for being kind, respectful and constructive. And also, they again have paid much attention to detail, which I was very happy about.

In my opinion, the reviews suggest that the manuscript is quite close in form and content to be recommended. Comments are now mostly not on a general level anymore, but relate to (at times crucial however) detail. This indicates to me that the manuscript likely requires only one last, relatively minor final revision.

I would suggest to you that you address the reviewers' comments in detail, should you choose to revise your manuscript. Some of the reviewer's comments may refer to writing style or normative or empirical tangents perceived as non-necessary (eg. the reference to the banned Venn diagrams, which I personally also found a bit off), which you should not feel obligated to incorporate. However, I would appreciate a quick discussion, if you choose not to do so for a point. I would appreciate if you thoroughly addressed point 1.2. by reviewer 1 ("more complex than polarized"), which I personally also found hit a crucial issue.

Best regards and thanks again for submitting to the PCI community - I look forward to seeing this manuscript enter its final stages before recommendation.

Dr. Mario Angst

## Reviewed by Yasaman Asgari, 05 January 2025

Dear Authors,

I want to thank the authors for their efforts in revising the paper following the first round of feedback. Your revisions show significant progress, and I appreciate the thought and care you have put into addressing the initial comments. I now invite you to refine the manuscript further based on the detailed feedback provided below. In particular, I encourage you to focus on improving the introduction and abstract, as these sections are critical for capturing the reader's attention and clearly articulating the value of your work. I am confident that with these changes, especially to the introduction and abstract, this paper will become a truly strong contribution to the field.

"The process of polarisation as a loss of dimensionality: measuring changes in polarisation using Singular Value Decomposition of Random Dot Product Graphs"

The title is so long and not catchy enough. Making it shorter is recommended.

Abstract is way too long. Try to be concise and only show the relevant information that a researcher needs to read to choose to further explore your paper.

"as well as a wide range of countries [2]. [3] [4] [5] [6]."

Make the citation like [2-6].

"Social and political polarisation is an issue of increasing concern in New Zealand [1], as well as a wide range of countries [2]. [3] [4] [5] [6]. While it was initially thought that New Zealand had avoided the populist takeover

seen in countries such as the USA [7], radical changes in the government makeup in both 2020 and 2023 have since called this into question. This paper will explore existing methods of measuring polarisation and propose a new measurement based on Singular Value Decomposition of social Random Dot Product Graphs." Start with a general paragraph. Remove the sentences about your method.

"0 [14] [15] [16]"

Citations like [14-16].

"Polarisation scholarship in New Zealand is particularly influenced from research in the USA." Why?

"As such, finding a method of measuring polarisation that can better handle this type of situation would be beneficial to researchers in New Zealand and similar countries."

NZ is not the only country with multi-polar polarisation. Please address the literature on measures that find polarisation in multi-party systems.

"This was first proposed by Esteban & Ray [30]."

The sentence is quite misplaced. Rewrite the sentence with the sentence before it.

"This concerns us because of issues such as potential polarisation against groups such as refugees;" The sentence has some emotional content. Rewrite for clarity.

"especially in countries that are seeing influxes of refugees due to war or climate change."

Relate climate change to refugees. It is not evident how they are related. Climate change polarisation is usually about mitigation strategies or the existence of climate change, not about its consequences like refugee flows. "Computational data science..."

What is computational data science? Rewrite the sentence for clarity.

"their speed when analysing"

Speed is not a great term. You can say they are fast or mention that they are not computationally expensive or discuss the algorithm's storage/time complexity.

"There are fast and simple algorithms for detecting communities that can then be investigated for polarisation, such as the Louvain method [31]."

Repetitive. Merge with previous sentences or elaborate further.

"In some cases, clustering algorithms are used to measure polarisation; there are a number of ways of measuring distances between the clusters..."

Try to rewrite for clarity.

"but we are unconvinced by this; as the number of polarised groups increases, so does the number of dimensions required for their clusters to all be equidistant from each other, which increases the complexity of the space so much that it becomes more complex than what we would consider to be 'polarised'."

This is not a scientific argument. Rewrite for clarity and justify your claims. This is vague and cannot serve as a justification for your concerns.

"computational analyses"

Avoid this vague term. Replace with something more specific.

"dimension reduction"

Use the term "dimensionality reduction."

"As Principal Component Analysis or Canonical Correspondence Analysis, in order to create a single dimension that can be evaluated for bimodality."

Sometimes these methods reduce to two dimensions for visual inspection of concentrations.

"Measuring latent positions empirically means projecting them in lower-dimensional spaces, and then assessing the resulting first dimension for bimodality [33]."

Repetitive.

"These analyses do not treat the correlation structure delivered by the PCA or CCA as informative in and of itself; they are simply used to discover the largest principal component or axis (i.e. the largest dimension)." Unclear. Explain what is meant by "correlation structure." "An advantage of this approach is that the issue driving polarisation is generated from the data, rather than researchers presuming what it is and risking choosing incorrectly."

What do you mean? Provide justifications or references to cases where researchers chose incorrectly.

"In their tests of Pearson correlations of pairs of issues in the American National Election Study, they did not find that there was increasing correlation in the ideological field of US-American voters pre-2004." Details not needed here. It is a paragraph about methodology.

"They argue that, 'In a colonized country, the presence of the dominant power is every day made evident through a variety of contents: differences of dress, of language, of skin colour, of customs [...] the colonizer is discursively constructed as the anti-colonized." (p. 128)."

Too much detail. Not relevant to your main argument.

"The starkest demonstration of this principle is in the actions of the 1970s Argentinian military dictatorship, which banned Venn diagrams from being taught in primary schools because they 'were feared to encourage subversive models of collectivity' [38]."

Not relevant to your argument.

"Laclau & Mouffe and Baldassarri & Gelman"

Citations should follow the same concise format.

"correlation structure"

Unclear. Needs explanation.

"We represent the conversation happening on a social platform (Twitter, Facebook, Instagram, etc.) as a network. Each user that took part in the conversation is mapped to a node."

Make this a subsection on network representation. It is not needed as a standalone paragraph. Your first results hold for any network. Start with the mathematical formulation instead.

"Methods"

Add a paragraph outlining the methods section.

"Introduction"

Include an outline of the rest of the paper, advertising what readers will learn.

"Having established these networks, we model them as Random Dot Product Graphs (RDPGs) [39]. RDPGs are used instead of other graph embeddings because their optimal embedding dimension (see below) is established a priori to the analysis, so it is independent of the network's size."

Rewrite for clarity and connect it to previous information. Explain why you chose this model.

"metric spaces"

Specify whether values in the metric spaces are real numbers, natural numbers, or complex numbers. Clearly define how L and R are formulated.

"Li and Ri are in general not directly observable."

Unclear. Do you mean they are not given or that they must be inferred?

"Li · Rj = P(i □ j)."

Explain that L\_i and R\_j are vectors and clarify their dimensions. Are L and R matrices? What are their properties? "We do not parametrise the network for this analysis."

Unclear. Rewrite for clarity.

"Given a network, we can assess its graph complexity by computing its SVD entropy. We chose this measure instead of others because it is based on  $\Sigma$ ."

Provide a stronger justification and connect this to the previous paragraph.

"This is based on the same principles as the view that the process of polarisation is one of increasing correlation."

Explain this better.

"We complement this definition by also observing the complexity of the graph, as determined by its SVD entropy, and notice whether it corresponds to an increase or decrease of polarisation."

This statement lacks clarity and detail. Rewrite to make it more precise. It is a crucial part of your work.

"We divided the data into two-time windows that corresponded to equally sized, large networks: between 2017 and 2020, and 2020 to 2023. For each time frame, we built a network by considering each user (identified by their unique IDs) as a node, and any mention, reply, or quote tweet between two users as an edge. Retweets were excluded in order to maintain focus on the New Zealand network and prevent the network from expanding to international discussions of climate change. There were 6767 tweets in the 2017-2020 network and 6172 in the 2020-2023 network; as such, the network sizes given are based on the number of nodes unless otherwise stated. We analysed the two networks independently."

Convert this into a table showing:

Time period

Number of tweets

Number of nodes

Number of edges

Size of the giant component

"Figure 1: Plot comparing ^d of NZ climate change tweets in 2017-2020 to ^d in 2020-2023." Add to the caption what the plot reveals or how it contributes to the argument.

Show all entropy values with a maximum of three digits of precision.

"4.1. Acknowledgments"

Remove if empty or add relevant acknowledgments.

COP Discussion on Twitter:

Put all results (e.g., entropy values, dimensions with 100 or 1,000 SVD values, number of tweets, nodes, edges, clustering coefficients, and average degree) into a single table for better clarity and understanding.

Finally, I kindly suggest using apps like Grammarly to discover writing issues. I know that the author's native language is English, but this helps to bring some clarity as well.

## Reviewed by anonymous reviewer 1, 05 December 2024

Thank you for revising the article. I believe that it has much improved compared to the previously submitted version. I'm satisfied with how most of the concerns which I raised in the last review were addressed. Nevetherless, I'd like the authors to address the following remarks before I can recommend the article for acceptance:

- 1. The new introduction is much clearer than the earlier version. However, there are some statements which I found rather obscure:
  - (a) "The most important is that use of hypothesis tests can be used to detect when a distribution is polarised, but after the first significant result it is difficult to robustly show that any further increase in polarisation is also significant. That is to say, while subsequent results may also be statistically significant, it is difficult to show that the increase from the first result to the second is itself significant."

What I understand here, is that such hypothesis tests often do not report significant increases in polarisation. How is this an issue with the test? If the increase is not significant, then it's existence is not supported by the empirical data (at least as long as the statistical assumptions hold).

(b) "It is possible to argue that three or more groups can all be polarised against each other, but we are unconvinced by this; as the number of polarised groups increases, so does the number of dimensions required for their clusters to all be equidistant from each other, which increases the complexity of the space so much that it becomes more complex than what we would consider to be "polarised"

Perhaps I don't understand this point but couldn't the same be said about your approach? In your work, polarisation is the dimension of a latent space. In the quote above, complexity seems to stand for dimension. So how does higher complexity imply that the data is "more complex than polarised"? And what does that even mean?

Compare also "Our method gives a value that corresponds to the network's capacity for complexity, and is inversely related to its level of polarisation" on page 7 where complexity is reciprocal to polarisation. I feel this contradicts the quote above.

(c) "The starkest demonstration of this principle is in the actions of the 1970s Argentinian military dictatorship, which banned Venn diagrams from being taught in primary schools because they "were feared to encourage subversive models of collectivity" [38].

I understand that the main motivation for the paper stems from political science or sociology and I have sympathies for the author's concern for the degredation of democratic principles in NZ. Nevertheless, I'd personally find the introduction more convincing if it was written in a more sober style. This is just meant as feedback for the authors not as something that, in my opinion, must necessarily be changed.

- 2. Section 2.1:
  - (a) L and R are introduced as metric space and later being referred to as matrices.
  - (b) Please make the assumption explicit that L and R take values in the interval [0,1].
  - (c) As I remarked before, Σ must also be truncated in the definitions of ^L and ^R. It is okay to reference other works for details but the things you make explicit should be correct.
- 3. Answer to point 2 in my previous review: I think you did a good job expanding your discussion of previous work. I think that criticising scientific approaches by others is a cornerstone of scientific practise. As long as you don't make any unfounded or polemic claims (which you didn't do), everyone will be fine with reading your criticism.
- 4. Answer to point 4 in my previous review: The networks which are compared in Section 3.1 are of the same size (1000 nodes sampled). This makes it in my eyes at least plausible that you can compare the dimensionality value across networks. Conversely, I feel it's plausible that a larger network should in expectation have larger dimension, without necessarily being more polarised.

Are the networks considered in Section 3.2 of the same size? Please report the network sizes (I couldn't readily find them in [33]). Also, if they are different, I'd suggest to same sample-size subnetworks from them as you did in Section 3.1 in order to make the computed dimensionality values comparable.

Typos & minor remarks

- 1. incomplete sentence in second paragraph on page 4
- 2. second paragraph, page 6: "beyond just they ideological field"
- 3. page 9, displaymath environment: In should be \ln
- 4. page 14: misplaced [h]
- 5. The paper [33] has a appeared in a journal. Please replace arXiv references by peer-reviewed references, whenever possible. Also it would be great to include DOIs, whenever possible.

## **Evaluation round #1**

DOI or URL of the preprint: https://doi.org/10.48550/arXiv.2403.18191 Version of the preprint: 1

## Authors' reply, 24 October 2024

#### Download author's reply

#### Decision by Mario Angst <sup>(i)</sup>, posted 24 June 2024, validated 26 June 2024

## An interesting article in need of some polishing

#### Dear authors,

I am super happy that I was able to find two researchers who have given your preprint a thorough review. In my opinion, their reviews have met the standard PCI network science sets for being kind, respectful and constructive. Beyond this the reviews shine in the attention to detail.

In my opinion, the reviews suggest that there is no fundamental problem with this preprint the should prevent it from eventually being recommended, but it still has some way to go to get there. I hope you share this assessment after reading the reviews. I would very much like to see a revised version of this article and eventually recommend it.

I would suggest to you that you address the reviewers' comments in detail, should you choose to revise your manuscript. Beyond this, three broad themes stand out I would suggest you consider on a more general level from my reading of the combined reviews, which I would consider necessary to move forward with this manuscript:

- The review of existing approaches to measure polarization would benefit from more depth and completeness, both theoretically and technically. The reviewers have made some good starting suggestions in this regard.

- The introduction and description of methods used needs increased mathematical rigour and could benefit from some more depth. Again, the reviewers have made some suggestions here.

- More rigour is also required in a more thorough description of both empirical and synthetic test procedures, which includes the materials put in place to ensure the reproducibility of results.

Best regards and thanks again for submitting to the PCI community - it has been a pleasure handling this preprint up until now and I hope our conversation on your work will continue.

Dr. Mario Angst

## Reviewed by anonymous reviewer 2, 28 May 2024

This paper introduces new dimensions of polarization, yet it requires several modifications to enhance clarity and accuracy:

1. Title: The phrase "network graphs" might be redundant as 'networks' and 'graphs' typically refer to similar concepts. A more precise term would improve clarity.

2. The abstract is considerably lengthy. Typically, abstracts should not exceed 300 words. Consider condensing it without compromising the content's integrity.

3. Literature Review: The discussion on polarization methods within networks is incomplete. Notable methods such as matrix factorization and network embeddings are not mentioned. For instance, network embeddings' utility in understanding polarization and their relation to community detection methods like von Neumann entropy are crucial topics that are overlooked. Additionally, the discussion on polarization related to climate change, particularly across various social media platforms in countries like Germany, China, and the

USA, is missing. A comprehensive summary of these works could provide better insight into the underlying causes of polarization and the current state of research.

4. Materials Section: The explanation of Von Neumann's entropy, its relation to Pielou's evenness, and the preference for this method over Shannon entropy need clearer justification. The frequent use of "communication network" may cause confusion; consider specifying its context. Moreover, the mathematical framework and language require refinement. Consistency in the spelling of "polarization/polarisation" throughout the text will enhance the professional tone.

5. Case Study on New Zealand Twitter: Clarification is needed on the use of 'age' as a keyword, the number of tweets and unique users collected per period, and whether these collections represent networks sized by nodes or edges. Consider visualizing these networks and explaining the absence of other potential keywords like "COP2x". The methodologies used for mentions, replies, quotes, and why retweets were excluded should be clearly justified( also what is a mention?). More details about the dataset's timeline, and representative analysis are also required.

6. COP Discussion on Twitter: It appears that yearly data was intended for use, yet no graphs show the polarization process over these years. Details on the number of users, nodes, tweets per category and year, and network visualizations for each "COP2x" discussion should be included to avoid potential misrepresentations due to small sample sizes.

7. Synthetic Data: You mentioned that we start with 1000 nodes and partition them with fractions 0.5, 0.2, 0.1, and finally 0.01. This results in a group of just 10 nodes, which allows for a maximum of 45 possible edges among them. Given a connection probability ranging from 0.3 to 0.45, this results in approximately 13 to 20 edges, leading to a sparsely connected component that might even become disconnected. Conversely, the remaining 990 nodes could potentially form 9900 edges, but with a 0.01 probability, only about 99 edges would actually form. This creates a significant imbalance between intra-group and inter-group connections. Therefore, it would be more effective to focus on adjusting edge density instead of merely connection probability to better understand polarization. Additionally, it appears you have not compared these results with other existing methodologies, which could provide further insights into the study's robustness.

8. Consider integrating all findings into a single figure with multiple subplots rather than separate figures for each result.

9. The conclusion discussing unexpected findings related to common axioms of polarization is unclear. Could you elaborate on these findings and their implications?

Finally, I would like to kindly ask you to revise the way that the manuscript is structured; this could make a more fluent story and better showcase the great work and the results.

## Reviewed by anonymous reviewer 1, 22 May 2024

The paper proposes a method for measuring polarisation in a social network based on singular value decompositions of the network's adjacency matrix. While previous approaches had focused on divisions into two opposing groups, the method described in this paper assigns to a social network the dimension of a latent space which serves as a measure for polarisation into multiple groups. Thereby, multiple reasons for fragmentation of a social network can be explained. After introducing their statistical framework, the authors consider multiple real-world and synthetical networks. Based on these instances, the authors argue that their notion of polarisation can indeed serve as a measure for polarisation in social networks.

The measure for polarisation introduced by the paper seems to provide interesting insights into the example data sets. It is an interesting statistical tool, which addresses some of the conceptual shortcomings of previous works (bimodality).

However, the paper does not provide a clear hypothesis, which is justified in the introduction and addressed in the results section. Essential notions are introduced only in the results section without being mentioned in

the methods sections or introduction (von Neumann entropy). Moreover, the paper lacks mathematical rigour.

In my opinion, the paper should be revised substantially before being considered for publication, taking into account the suggestions below:

Title and abstract

Does the title clearly reflect the content of the article? [x] Yes, [] No (please explain), [] I don't know Does the abstract present the main findings of the study? [x] Yes, [] No (please explain), [] I don't know Introduction

Are the research questions/hypotheses/predictions clearly presented? [] Yes, [x] No (please explain), [] l don't know:

I could not identify a clear hypothesis from the introduction.

The goal seems to be to measure polarisation. However, it is not clear to me how the abstract properties of polarised communities that are derived from the cited political science literature are transformed into statistically testable properties. Correlation seems to be used as an antonym to polarisation without precise definition (correlation with respect to which probability distributions?).

Furthermore, in the results section, the von Neumann entropy of the networks plays a central role without being ever mentioned before.

In my eyes, the paper would benefit from a clear cut hypothesis or research question, which corresponds to the results section.

Does the introduction build on relevant research in the field? [] Yes, [x] No (please explain), [] I don't know: the authors mention some statistical techniques for detecting communities/polarisation in social networks such Principal Component Analysis, Canonical Correspondence Analysis, or clustering. However, a detailed discussion of the shortcomings of already existing techniques is missing.

Materials and methods

Are the methods and analyses sufficiently detailed to allow replication by other researchers? [] Yes, [x] No (please explain), [] | don't know:

The mathematical definitions of many concepts remain vague and imprecise (see below).

The Github repository lacks instructions for installing the package. Please consider depositing a docker image at a recognised research data management platform.

Are the methods and statistical analyses appropriate and well described? [] Yes, [x] No (please explain), [] l don't know: Several key concepts are not defined rigorously (see the detailed comments below).

In my eyes, the applicability of the techniques in the paper for longitudinal studies is not sufficiently justified. Polarisation values of multiple networks (same data set over time) are compared. The paper would benefit from a more detailed explanation of why these values can be compared.

For the results based on stochastic block model graphs, it would be great if some of the statistical properties considered here were computed analytically rather than experimentally. Computing expected values and error probabilities would significantly improve the results here.

Results

In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)? [x] Yes, [] No (please explain), [] I don't know not applicable

Are the results described and interpreted correctly? [x] Yes, [ ] No (please explain), [ ] I don't know Discussion

Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? [] Yes, [x] No (please explain), [] I don't know. The results are not compared to other techniques for community detection.

Are the conclusions adequately supported by the results (without overstating the implications of the findings)? [x] Yes, [ ] No (please explain), [ ] I don't know Issues with mathematical definitions:

Section 2.1

It should be specified what the vectors/matrices are indexed by. In particular, the crucial assumption that the dot product of the vectors L and R lies in the interval [0,1] should be justified. This is essential since these dot products are used to define probabilities.

Something is odd with the truncation: in the definition of \hat\L\ and \hat\R\, the dimension of Sigma does not match those of the truncated matrices. Better be more precise about what the matrices are indexed by, what their dimensions are.

In what way is \hat\d\ optimal? What does it mean for this value to robust with respect to network size? Why is \hat\d\ robust?

What is ||A||\*?

What happens with negative singular values in the first equation on page 7?

Section 2.2: \hat\d\ is an integer, not a vector space. As such, it does not have a "dimensionality".

Section 4: von Neumann entropy is first mentioned here. Please provide a formal definition or reference in the methods section.

All figures: axes, subplots are not sufficiently labelled. In particular, what do the error bars indicate? What was sampled in Figures 1 and 2? Aren't these just individual fixed deterministic networks?

Tables: the dimension d is always an integer. It should be written without .0.

Typos:

Page 5, last line: which should be whose Page 6: "this commonly" should be "this is commonly" Page 15, last line: reference missing