

Bridging the short-term and long-term dynamics of economic structural change

Corresponding Author: Dr James McNerney

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

General comment:

In this work, the authors introduce a network model of an economy's structural transformation by which showing the connection between the two fields of study of the economic complexity, i.e., the study of the economic activities and the creation of proper metrics for the identification of the productive knowledge. The work is mathematically sound and the valuable contribution to the RCA approach is of particular interest when focusing on niche productive sectors. I endorse the publication in some form, as I argue that this work can actually be useful for all the academics engaged in the study of economic complexity and, even more, for decision makers aiming to use the resulting information. However, the arrangement of the arguments obstacles the transfer of knowledge to any of these stakeholders. The main text is not readable as a stand-alone work without previous knowledge on network science or economic complexity and requires a back-and-forth movement with the SI to understand the framework better. In this form, it exits the aim and scope of a multi-disciplinary journal such as Nature Communications. Thus, the significant revision here presented relates to the need for the authors to rearrange the argumentation and discussion to foster the engagement of the reader, rearranging the logical structure of how results and comments in the text.

Major revisions for text:

1. Applicable to all text: cross-references to the methods and SI lack and details are thus lost in the reading). There's also a missing piece of information that comes with the figure, and often the text refers to detail which are just given in the caption of the figure (e.g., Figure 2, the description of the network is only given in the caption, but clear references are in the text without further information on how the network is constructed).

2. Introduction:

a. As structured, the introduction is unbalanced, pushing toward argumentative text in support of the findings of the work (see pages 3 and 4 – that can be suitable for the discussion section instead) rather than presenting a proper context on where to locate these findings. An example is given by the GENEPI metrics, which is not presented in the introduction, while it is cited in the caption of figure 1 (the index is not introduced until the results section despite being a recurrent reference). The same applies to the concept of complexity, latest paragraph: here, the readers are introduced to something which is still to be adequately defined, and that remains suspended until the end of the Results. Actually, the concept of complexity, and how this is intended with respect to the field of study, is something that can be of interest to be briefly detailed here in the introduction section.

b. When introducing ECI and Fitness, the authors describe the bipartite using the sentence "which locations produce which products". The wording is imprecise compared to the methods. It is worth amplifying the discussion regarding production and export to unfold the difference between the two economic activities and why the RCA method is recursive in the metrics definitions (see description at the end of page 4, Results section).

c. Fig. 1 can be deleted considering adding a reference to its equal in the SI. It isn't easy to understand its meaning here at the beginning of the manuscript without knowing any detail regarding the variables on display.

3. Results:

a. Paragraph of Eq. (2): can you detail which network you refer to with the sentence "The PoR treats diversification as a process of spreading on a network."? I guess this one is the one from which you obtain a Laplacian network for figure 2 (and so it is the bipartite network of countries and exports). Still, details might be helpful to ease the reading of the text without

recurring to the SI.

b. Some more math references in the main text can help the non-expert reader to follow along. The comment applies to all the sections. The main text gives so many details for granted, and it should be supported instead to allow the general public access to the results. (eg. The ordering of the eigenvalues).

Other comments and food for thoughts

- As shown in Fig.3, tax havens and resource-rich countries are evident free riders in the economic development panorama. How truthful and valid is it to use your model on the export data knowing that these two kinds of countries can bias the results? This is particularly important when comparing the metrics' results with the GDP. This comment also relates to the note the authors made on oil: as described in the manuscript, oil extraction does require a broad set of knowledge, yet, the presence of the resource first defines the activity.

- To the previous related: besides the export ones, with which other kinds of data can the PoR model be used to extract information on countries' growth?

- Fig. 3:

- the border between R2 and R3 is at value (circa) $A=0.4$ value. Does it have a mathematical meaning and interpretation?

- Despite the full disclosure regarding the dynamics of the ensemble of countries, there's reason to believe that the wind speed at the interface among regions can be motivated by the trajectories of some specific countries rather than others. Have you explored these results further and detailed the contribution of each country's trajectory to the ensemble?

- All text and Section S1.8: how do the present arguments relate to the micro-model of structural change from Hausmann, Hwang & Rodrik? I argue that this analysis can further add value to your work.

- Can you show the time series of the correlation among the metrics?

- In detailing the comparison with the GENEPI, the authors comment on the difference between the two metrics, considering the nature of the proximity matrices upon which the metrics are constructed. Referring to the PoR, isn't there a shrinking of information by deriving countries' coordinates through the projection onto the axes?

- Since the GENEPI also identified stages of growth, how are these related to the Regions of growth you identify with the evolution of the terms A and b? In this sense, following the trajectories of the countries between the two planes could be of interest.

Reviewer #2

(Remarks to the Author)

The paper develops an integrative framework for measuring economic complexity (using national export data) to describe the growth and transformation of economies. There have been many measures of economic complexities (diversity, fitness, genepi, ECI, and entropy), and the paper provides how these measures correlate (FIG 1). The paper also provides how these compositional (diversity) measures imply the temporal evolution of the economy (as in Eq 1-4) and show the coordinate along the evolution (Fig 2 and 3). And finally, the paper shows how this evolution is decomposed into two parts indicating two coordinates (as in Eq 6), and thus corresponding to economic quantities: GDP and diversity (FIG 3).

My overall impression of the paper is very good. The authors demonstrated well their knowledge of the subject matter, and their derivation is technically sound, but it still needs to be articulated and goes deeper to understand fundamental mechanisms beyond mathematical interpretation. My major concern is 1) how much value is in the paper's contribution, and 2) if it is aligned with the intellectual interest of the readership of Nature Communication. In particular, it is unclear whether the manuscript is supposed to be a review paper or a new contribution as much of the content is recapturing the previous papers, including Sciarra et al. 2020, and Mealy et al. 2019. Therefore, as it currently stands, the manuscript contains several shortcomings (listed below) that render it inappropriate for a high-impact journal like Nature Communication.

The manuscript summarizes many different measures, presenting their correlations (two groups) nicely in Fig 1. This result was, however, already discussed in Mealy et al. 2019 and Sciarra et al. 2020, and the orthogonality between diversity and ECI has also been clearly discussed in the aforementioned papers. The temporal evolution is also discussed in the previous papers in a clear way (sections "spectral clustering" and "diffusion map distance" in Mealy et al. 2019). The paper's contribution is to make this discussion more explicit in time. But there was no further discussion of underlying mechanisms or independent observations beyond the simple mathematical recasting. The inferred coordinates and dynamics are again already discussed in Mealy et al. and Sciarra et al. I like the current manuscript because it rewrites the previous findings in their own language/framework, but the contribution does not address mechanisms beyond the equations. Therefore, it is hard to see if the current version is an enough substantial contribution to be published.

Minor issues:

1. Please rephrase ambiguous sentences and phrases. For example, the authors use the term 'our theory' without a concrete definition of their theory. Examples include sentences 'here we present a simple theory that suggests these branches describe the same dynamic process of develop... Our theory begins by making explicit the dynamic models that are implicit...' Sentences are on and on with no explicit definition of the authors' theory. What is the theory here by the way? Even to the end of the manuscript, I couldn't figure out what the authors mean by a theory. I can only guess them, but still I wasn't sure if the observation or explanation are meant to be a theory. Derivation is not a theory, and explanation is not a theory. So, I am really left unsatisfied. There are many cases. Please be specific and clear about what is being discussed.

2. Secondly, please be cautious about using emphatic adverbs/adjectives. In particular, instead of telling what readers are supposed to feel, please state what exactly they are and be informative. For example, in the abstract, 'these metrics take on radically new meanings, ... profound mathematical connection between...' There are many other cases, including: "...there is a surprisingly close correspondence...", "interestingly, these two groups of quantities resonate surprisingly well...", and

“somewhat alarmingly...” Of course, I do not think every use of those words is bad, but the paper uses them more than enough. Please use them with caution.

3. How does the decomposition of short-term and long-term time scales of dynamics differ from detrending in statistics?

4. The manuscript is somewhat narrowly constructed. It is written only those who are interested in ECI and product space because the manuscript is intended to integrate the literature of the field of ECI and product space. I put this as one of the two major concerns not because it is a deal breaker but because I hope the authors recognize this issue. It is not a deal breaker because I am sure that the authors can easily fix it by providing the context for the broader readership.

Reviewer #3

(Remarks to the Author)
Referee report for Nature Comm

I have answered to the “template” questions suggested by the Nature Communication below.

I think this is a great paper. It is literally a cornerstone of the literature on economic complexity and the principle of relatedness, and within that literature it will quickly become a classic. The paper is also technically very solid.

I have two main comments:

1) To what extent is it okay to select only one eigenvector? (I mean, one more than the uniform one). In other words, what is the distribution of (the first few) eigenvalues? The authors never discuss this. Ideally, I would have like to get a sense of the “effective rank” of the proximity matrix (see “low-rank hypothesis of complex systems” <https://arxiv.org/pdf/2208.04848.pdf>) Table 2). This is asking too much, but I hope the authors can at least plot the spectrum or describe the first few eigenvalues. Also, just to be clear, in S1.3.1, you would select the complex conjugate of the second eigenvector, and create a composite? Adding a sentence on mode selection in S1.3.1 might be helpful, although I recognize that this is out-of-scope given the choice of a symmetric matrix in the main text empirical application.

2) Do PoR studies use a constant Φ matrix, or is it updated over time? Is the Φ matrix more or less constant empirically? The assumption of constant Φ is necessary/convenient/clever (to solve Eq.(1) and move forward with this approach). But I would have appreciated a little bit more justification. I was not very convinced by the last paragraph in S1.8 – maybe it just needs to be expanded/clarified. This is a somewhat sensitive issue, because the authors never really expand on the details of all the time scales involved (see point 1), and one could think that Φ changes “quite quickly” compared to “how quickly” countries export baskets change conditional on a fixed Φ .

I have a number of smaller comments – these are either typos, or suggestions that the authors should feel free to ignore:

- It would be helpful to define “diversity” the first time it is used.
- In Eq. (1), use $=$ or $\equiv$, or please define “ $\sim$ ”. I don’t think you do mean “have the same scaling as $N \rightarrow \infty$ ”? I guess you need to add a noise term or write the LHS as a conditional expectation?
- After Eq. (3), make it clear that the last term can be negative (e.g.: replace “ability in p rises” by “ability in p rises, otherwise it falls”). I suspect the declines in ability are much rarer, so much more informative.
- In fig. 4, there are some dots on the left that are vertically aligned: They all take the same value for the ECI, but different values for the ECI*. Is there something interesting going on here, whereby ECI* is better able to capture detailed patterns, while the ECI fails because it uses uniform weights?
- P.7: “analyzed it in terms” -> “analyzed in terms”
- Fig 3: Consider giving more details for how Panel e) is produced. I suspect it might be too complicated to report – are you posting the code? Citing a software package?
- P11 : “rise in number” -> “rise in the number”
- P12: “higher wealth” -> I think you mean “higher income”(GDP is a flow, not a stock)
- P. 17: The paragraph starting with “Fourth” is interesting but a bit hand-wavy (or maybe it is clear to the authors and I just don’t see the point as clearly): can this be sharpened?
- P.19: “Differences between countries shrink first among quickly-decaying dimensions...”: There is probably a connection to the literature on “convergence clubs”.
- Section III tends to repeat ideas that were exposed previously.
- S1.3.3: “coorainte” -> coordinate
- Fig. S2 could have had a panel with $R_0 = 0.115$
- S18, p.13, paragraph starting with “First”. Wouldn’t the elements of v_2 depend on the choice of R_0 ? This would be a helpful reminder/caveat.

What are the noteworthy results?

The authors show how the principle of relatedness relates to the complexity metrics. They proceed as follows:

- Start from a dynamic equation that is successful in empirical research

- Solve it (i.e., express it as a function of time and initial conditions)
- Diagonalize the solution, and express the approximate solution in terms of the first 2 eigenmodes
- Study these eigenmodes – it turns out they are very related or identical to classic “complexity” metrics!

Will the work be of significance to the field and related fields? How does it compare to the established literature? If the work is not original, please provide relevant references.

This is an absolutely major study in the field of economic complexity metrics and relatedness. Within this field, I think it is the most important one in the last few years. It is at least as significant, in my view, as the recent studies in similar journals (Sciarra et al [27], Mealy et al [50]).

Does the work support the conclusions and claims, or is additional evidence needed?

Yes, this is a very solid paper. I have made some suggestions – I'd like to see a bit more of the spectrum of L_{Φ} . But the core conclusion is already supported by substantial and technically solid analysis.

Are there any flaws in the data analysis, interpretation and conclusions? - Do these prohibit publication or require revision?

As far as I can tell, there are no flaws that would prohibit publication. I have made some suggestions for revisions although I think the study could almost be published “as is”.

Is the methodology sound? Does the work meet the expected standards in your field?
Yes, and yes.

Is there enough detail provided in the methods for the work to be reproduced?
Yes, it appears so (I haven't tried), except for the details of how to create Fig. 3e.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Before moving through the details of this second review round, I'd like to formally apologize with the Authors for the delay of this report. The due date that this report needed was conflicting with cumulated sum of workload and unexpected life events. Again, apologies.

Reviewer #1

With respect to the first version, I appreciated the effort the Authors put in implementing the requested changes making the article more readable to unspecialized audience. Yet, despite the improvements, I'm still questioning whether the contribution the Authors claim to make truly is groundbreaking as the journal would require, or re-enters in the debate characterizing economic complexity while increasing the noise around the amount of metrics available.

My discussion stems from what the Authors claim and what literature offers.

What the Authors argue to do in tying together two areas of work within the economic complexity field appears to be a little bit highfalutin. I do agree that the way the framework introduced in this work cleans up the many contributions the economic complexity literature has (Table 1). Moreover, it provides a clearer interpretation about the long- and short-term dynamics of structural change. However, I would reduce the flattering arguments. In fact:

The principle of relatedness lies beyond all the metrics of economic complexity, and the main difference lies in the way this principle is embedded in the proximity/similarity matrices the short-term metrics are built upon. The EC metrics from Hausmann and Hidalgo are the original examples for this statement. The productive space in the Hidalgo et al. Science, 2007, provides the ground for the ECI-PCI metrics to have mathematical reasoning. I may agree there is always room for improvement, but if one has to identify the original example of the principle of relatedness, and the way it relates to the metrics of complexity, that is where I'd search for. The same Authors refer to these pieces of literature, and there is clear mathematical resemblance in among the metrics the Authors present in the Supplementary Material – S7, and the Science paper from Hidalgo et al. Thus, I am left with doubts about the importance of this contribution.

At page three, the Authors claim surprising the fact that currently in use metrics of economic complexity either relate to the diversity or composition of the economy. Is it though, surprising? This result has been repeatedly shown over and over, with examples in Mealy et al., Science Advances, 2019; Mariani et al., The Eur. Phys. Jour.B, 2015; Sciarra et al., Nat. Comms., 2020, and many others – as many referring works this manuscript has, especially considering the correlation with the diversity. Moreover, the GENEPIY metrics with its two component has introduced the idea that the two classical metrics of economic complexity either measure diversity or composition of the export baskets, and the composition of the two leads to the multi-dimensional metric of complexity.

Yet, the addition of the metrics the Authors introduce is new and marks my following point.

The authors make clear that contrarily to other research works their aim is to introduce a dynamical modelling framework from which to extract the long-term implications of short-term dynamics keeping distance from the definition of new complexity metrics. Still, in introducing the A and b components, they target the comparison with other economic complexity metrics, as the ECI, Fitness and GENEPIY, even defining the variables with resounding names such as ECI*, for example. Beyond the different mathematical frameworks, the convergency of the model's components follows that of the iterative

Method of Reflection in Hidalgo-Hausmann PNAS, 2014, with the first eigenvector resulting to be uniform and the second component taken for more informative results. Again, what can we learn anew with respect to previous contributions? I am afraid of the risk of falling within the circular reasoning of “729 new measures of economic complexity” - <https://arxiv.org/ftp/arxiv/papers/1708/1708.04107.pdf>.

I have a point about discussing structural change with respect to the presented framework. This point is not a critique moved for the Authors their-selves, because I can see it paves the way for a larger and more philosophical debate exiting the aim of this work. Yet I'd like to raise it here as the Authors can discuss through since “it is not an intrinsic requirement of our model” – page 5.

The economic complexity field has been marked by the presence – in its methods – of the Revealed Comparative Advantage. The role of the RCA is two-fold: it allows us to remove intrinsic noise related to the presence of outlying trade operations and to reveal the significance of the exports. There's plenty of literature, e.g., as Authored by Giovanni Dosi to cite one, discussing about how this procedure biases the results. My comment follows the same argument. Can we actually speak about long-term structural changes when the RCA erases important information? Note that the following reasoning works even considering the log transformation the Authors introduce, because the critical issue here stands in the RCA threshold ($RCA=1$), the authors still use it to map the values onto the log scale. Here's an example to follow. Let us analyze the following trade matrix among the countries c_1, c_2, c_3, c_4 of products p_1, p_2, p_3 . The trade flows range among three different orders of magnitude.

case (a)

```
trade p1 p2 p3
c1 100 10 0
c2 100 10 0
c3 10 0 1
c4 0 0 10
```

with RCA values

```
RCA p1 p2 p3
c1 1.043 1.095 0.000
c2 1.043 1.095 0.000
c3 1.043 0.000 1.992
c4 0.000 0.000 21.909
```

Let us just play with the component $c_2 p_3$ as in the example below.

case (b)

```
trade p1 p2 p3
c1 100 10 0
c2 100 10 10
c3 10 0 1
c4 0 0 10
```

with RCA values

```
RCA p1 p2 p3
c1 1.087 1.141 0.000
c2 0.996 1.046 0.996
c3 1.087 0.000 1.087
c4 0.000 0.000 11.952
```

As shown, when changing the element from $c_2 p_3=0$ – case (a) – to $c_2 p_3=10$, case (b):

In case b, the RCA value of $c_2 p_1$ changes from being greater than one, to be smaller than one, being one the threshold of competitive advantage. Thus, even though the country c_2 has not changed its performance in that particular product, it may result in some kind of structural change when analyzing the binary trade matrix.

The value $c_2 p_3$ increases by one order of magnitude – possibly implying that there has been a sort of structural change in the economy – which we are though not able to capture due to the RCA threshold impeding us from catching that 0.996 value.

So, my question to the Authors is: if the RCA threshold is not an intrinsic requirement of the model, and considering that the framework you present is claimed to be novel, can you show the results of your framework on long-term and short-term dynamics without considering the RCA?

Reviewer #1 on behalf of Reviewer #3

All points raised by the Reviewer have been properly addressed and answered.

Reviewer #2

(Remarks to the Author)

I'm thrilled to see substantial improvements in the paper, and I'm pleased to note that my two primary concerns regarding the paper's contribution and the readership of Nature Communication have been largely addressed. The new introduction is way better than the previous one. However, there are still some concerns that remain, listed below in order of importance. Besides the comments below, the paper's logics and methodologies seem sound and warrant publication.

1. Modeling Framework in Post-hoc Patterns: The paper asserts that it uncovers long-term and short-term dynamics by decomposing the bipartite network R_{cp} . Therefore, the derived long-term and short-term dynamics are sort of baked in the construction of R_{cp} , with the frontier of the R_{cp} network structure (associated with a large 'A' axis in Fig. 3) relying mainly on the economic activities of developed countries like the USA and Germany. The rest are catching up on the 'A' axis with diffusions on the 'b' axis in Fig. 3. This is what I refer to as post-hoc patterns. It doesn't imply that developed countries haven't slowed down, recessing, or wiggling in the frontiers, as evidenced by small, curled trajectories in region 2. The observation here is that underdeveloped/developing countries and un-usual countries (like tax-haven) have been catching up with developed countries in their product compositions over time (Region 1 and 3). Does this mean there is only one long-term progression of economic development, mostly set by developed countries? Is this the only path to development enough to be called a fundamental mechanism? Is there room for "Leapfrogging dynamics" in the framework? I would caution against labeling this as 'a modeling framework,' as it might mislead others to accept the post-hoc patterns without question.

2. Beyond Post-hoc Analysis: To extend the analysis, one could examine the long-term and short-term dynamics from R_{cp} in the past (1962-1999) and compare them to the dynamics between 2000 and 2018. How much do the long-term and short-term dynamics hold for extrapolation analyses? In this way, we can say the findings are beyond post-hoc patterns, and there may be something more fundamental to the backward projection. A simple comparison between R_{cp} (<2000) with R_{cp} (≥ 2000) could provide valuable insights.

3. Related to the retrospect projection, the categorization problem should be addressed: The unit of analysis is projected from the SITC classification to analyze almost 60 years, during which many classifications have shifted, changed, and restructured due to technological innovation. Have any data attrition issues been overlooked in dealing with this extended longitudinal evolution? How does projecting current dimensions onto previous activities affect the findings? Could this explain why we see some non-trivial trajectory curls in Region 2 where the developed countries (contributing the frontiers of the network structure) are located? Will these findings (curls in Region 2 in Fig. 3) still hold 30 years later with improved categorization systems? What are the chances the curls in Region 2 are methodological artifacts (using not-updated categories)? In other words, would these findings (curls in Region 2) still hold in 30 years later when we have updated categorization systems that better capture economic activities/products than now?

4. Nuance in Analysis: The nuance above is subtle, and this nuance by no means implies that the current analysis is incorrect. On the contrary, I am encouraging the deep discussion above because the current analyses and findings can shed light on a long-overdue shortcoming in the field, which has been overly focused on the complexity index from trade datasets. I was excited to see the paper thoughtfully state the exact same concerns that I have had in the field. However, to prevent misdirection in future research, I encourage the paper carefully articulate its scope and contributions to economic development theory.

5. Lastly, the paper could benefit from further tightening and trimming. For example, section D (on page 14) could be simplified by putting detailed analyses and expanding results, explanations and interpretation on pages 16 and 17. Another example would be that I find page 20 is the most exciting part of the paper, and I would like it to come slightly earlier than now. Also, for minor comments (related to the nuance point 4 above), I would suggest using formalizing/formalism rather than theoretical/theory throughout the main text whenever appropriate in order to avoid misguidance. For example, Table 1, Derived from a model $\rightarrow$ derived from a factorized matrix formalism. However, this use of the term is the authors' choice and please ignore my suggestion if the authors believe theoretical/theory is more appropriate.

I look forward to seeing these concerns addressed to improve the paper.

Reviewer #4

(Remarks to the Author)

I was recently asked to provide additional review on this paper, after the first round of reviews was already completed. In important ways, I found the paper interesting and a contribution to our understanding of economic complexity and relatedness. In other ways, in particular in its representation of the existing literature and the interpretation of the results, I found the paper falls short in important ways. My assessment is that this paper would be of interest of those working in economic complexity, but that as is, the paper is not ready for publication, and would benefit from a thorough revision.

In short, the paper uses a standard linear model of the form $(d/dt(\text{vector})) \sim \text{vector} + \text{matrix} * \text{vector}$ to draw some connections between measures of economic complexity and relatedness. The authors call this an economic model, but that is not the case, since it is a standard first order approximation in linear algebra, and not a model with rational agents, first and second order conditions, prices, etc. It is also not a generative model like other models used in this literature (original Lego model of the 2009 PNAS paper (solved analytically in 2011) or the 2017 Fink et al. model in Nature Communications). The authors then use this linear approximation to rederive well known concepts in the economic complexity literature: a measure of

diversity, a measure of economic composition (complexity), rediscovering the well-known fact that these measures of complexity are eigenvectors of proximity matrices. After that, they try to position their results in the existing literature in pages that largely misinterpret previous work. If the authors were to build on previous work, which they could because their results are very consistent with previous efforts, the paper would be a contribution. Doing so would require rewriting completely most non-technical parts of the paper (e.g. intro, interpretation, discussion). I believe the authors are capable of doing these changes, for which I recommend revising and resubmitting the paper.

In addition to this general concern, I also have specific concerns that the authors could address in a revised version.

1. The authors provide a linear approximation model predicting that in the very long run, all countries will produce all products. This is nonsensical from an economic perspective (and comes from the model lacking any strategic competition or comparative advantage generating mechanism). Can the authors justify their approximation better? (e.g., as a first-order matrix Taylor expansion of another model), Or tone down their claims of using an economic model to finally put the literature on a mathematical foundation?

2. Regarding the math, the work validates important previous work, adding to the robustness and validity of previous findings, but the authors do not describe it in such a way. Instead of identifying connections to previous work when interpreting their results, the authors provide mischaracterizations of previous approaches to differentiate their work more strongly. This is something that I would suggest the authors revise in a future version of their manuscript. A few examples are clear in Table 1

a. When the authors try to compare the “complexity inference view” with their “dynamics systems, view,” the first entry says that the “complexity view” is restricted to solving an inference problem and that their dynamical systems view brings in the “novelty” of characterizing activity baskets. That is incorrect, as an important motivation in the early work on economic complexity (going back more than a decade) is the characterization of baskets of economic activities.

b. Another example from the same table, which reads rather unkindly, is that in the complexity view the principle of relatedness play “no role in particular.” Yet, in a recent review of the field published in Nature Review Physics, this connection is clearly described as one of several known facts (Page 9: “metrics of complexity are connected to the idea of relatedness because they are derived from the spectrum of proximity matrices.”).

c. The same table also describes the derivation of complexity metrics as “motivated by a data challenge.” Again, reading the introduction of the original papers and books shows that this is clearly not the case. The original 2009 complexity paper starts with Adam Smith and the division of Labor. The 2014 book with MIT Press motivates complexity and relatedness by discussing the division of knowledge. Etc.

Overall, the authors seem far from understanding the motivation and goals of the original work. They would benefit from either not saying them, or learning to articulate their contribution in a constructive manner with the literature (since none of what they are saying is novel in term of motivation and goals is novel for an author experienced in this line of work).

3. Throughout the paper, the authors also repeatedly mention that the quantities they derive have little to do with complexity (e.g. page 17, page 4, page 21), but do not justify that claim. What is the definition of complexity they have in mind? It must be different than that of previous authors working in economic complexity, as in previous work, complexity measures are estimates of economic structure and composition that are relevant to understand the process of economic development. The latter is the interpretation the authors give to their own metric (b), and that coincides with the interpretation of ECI. I don't understand the need to invent a difference where there is a coincidence.

4. There are some interesting results that the authors do not seem to emphasize, but to me, seem more important than some of what they have now in their interpretation. One of them is showing that measures of economic composition (ECI, ECI*) are robust to changes in definitions of the proximity matrices. This seems to be a valuable contribution that is present in the results section but does not make it to the interpretation and discussion. This result is valuable as it adds to the literature and tells us that these metrics, as general concepts, are robust to variations in their implementation.

5. Another result along those lines is the authors showing that it is the composition variable (b, ECI*, ECI, second eigenvector, etc.), and not diversity (A, Fitness), the one that is more strongly related to GDP per capita (the authors do not study growth). This is important in two ways. First, this distinction is foundational to this literature. The original ECI paper compares ECI (composition) with diversity (HH, Entropy), to show that these measures of complexity capture information of composition that goes beyond diversity alone. Second, the original validation of these metrics was their ability to predict income and growth. The authors seem to concede that it is the compositional metrics (b, ECI*, ECI) the ones that explain this relationship (although this is an empirical claim derived from a figure, and not a mathematical demonstration).

6. On that note. If diversity metrics can be obtained by simply counting the number of activities an economy is specialized in. Why do we need a metric like Fitness? Seems like a complicated way of counting something as simple as the number of activities. The authors do not interpret this result.

7. On page 18 the authors claim the use of a system of equations to calculate complexity could be unnecessary. But that system of equations can be solved on six lines of code, while the linear approximation model proposed by the authors requires many more complicated steps to reach a very similar result (ECI*). Why would a simpler way to reach the same result be unnecessary?

8. The authors claim on page 21 that one of the advantages of their approach is that they derive these metrics from an economic model. But the authors do not present an economic model (no optimizing agents, prices, first and second order conditions, etc.). They present a non-mechanistic linear algebra model in which a vector changes according to a mixing matrix (the proximity matrix).

9. On that note, the model seems to lack any validation, other than helping generate variables that are correlated with metrics of diversity (e.g. Fitness) and complexity (e.g. ECI). The model would also predict that—over the very long run—all countries would produce all products (which would be a bad prediction). Therefore, it lacks any dynamics of comparative advantage or specialization. So, why do we need the model? Could we have just derived b or ECI^* from the eigenvectors of the mixing matrix (ϕ_{pp}). What is the model giving in addition to an alternative way of deriving indicators that we knew? Is it a principled model? Or just the first order approximation of a linear system ($Vector = Vector + Matrix * Vector$)?

10. The argument that the “fundamental” matrix is that of products seems incidental (page 19). Technological relationships between products also change over time (e.g. textiles were related to water power in the eighteenth century but not today). The “model” is based on this mixing matrix because the authors chose to build a model for a column of the specialization matrix R_{cp} (R_p), instead of a row (R_c). If they would have made the linear model about changes in the spatial distribution of products, instead of the specialization of countries, the mixing matrix of the model would be mixing economies (e.g. Φ_{cc}). So the claim of having discovered which matrix is “fundamental” using this linear model seems an overreach.

11. In the passage about oil (page 19) the authors seem to confuse complexity of products and processes. Crude petroleum extraction is a complex and difficult process that involves many high complexity inputs (drilling equipment, offshore rigs, etc.). But the product itself (crude petroleum) is not complex (it is whatever was underground). Moreover, since all of the things that are needed to extract petroleum can be brought into an offshore rig in the middle of an ocean, this shows that it is an activity that does not require any locally pre-existing complexity. Similarly, works-of-art do not require sophisticated drilling equipment, but they do take place in cities rich in knowledge flows and sophisticated markets, flourishing primarily in complex economies (e.g. New York, Paris, Tokyo), so their production should be correlated with the complexity of the economies exporting them.

12. In page 10 the authors refer repeatedly to $A(t)$ as ability, but since this has the same units as $R(t)$ ($R(t) = A(t)1 + \dots$), it is an indicator of specialization (RCA). The change of language is confusing and jumps into an interpretation of the variable. It may be simpler to stay with what it is descriptively.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Remarks to the Authors:

I thank the Authors for the scientific discussion we established throughout these review rounds, a discussion I am truly satisfied of, and I might say, it was definitely useful. The paper has benefited from this debate, with significant additional information included in the manuscript and supplementary information. If compared with the original version, it now better addresses the dilemma in the economic complexity fields, detailing the controversy between narrative and mathematical framework, and provides a better understanding of the reasoning behind the research question and how to address it. I feel the Authors' efforts in making their point in their Reply letters, a point that can finally be now read in the manuscript, even though I still have a few, minor, concerns. Regarding the text, I found the Reply letters to the Reviewers more straightforward and convincing than the tone of the article itself; especially in the background literature and discussion. I would suggest incorporating some of their defending statements into the article, if possible, especially regarding the above-mentioned dichotomy between narrative and mathematical aspects.

A few notes:

- I have a further point on the example the Authors report regarding product complexity. As detailed on page 19, the Authors claim that “pottery and works-of-art are clearly low-complexity goods”, countering the results of the analysis on product complexity. This is a bold statement. To date, there is little study-knowledge-agreement on which are the intrinsic characteristics of products that make them “complex”, and complexity from the metrics translate into capabilities of production, which I can suppose from the text that the Authors are imputing to supply chain, translocation of work, skills and knowledge. While I agree with the interpretation of the product complexity results concerning structural change that the Authors provide, and with the statement per se, I don't feel this statement is properly put in the current literature framework, and I'd reduce the emphasis on it.

- Something that I think is missing though is a ranking of countries according to the principle of relatedness, it would be interesting to see some results. I don't know if this is intentional, yet there is an interest in the rankings fueling the hype around complexity metrics, and that might be of interest to practitioners and stakeholders, as the Authors have pointed out. - Countries' specific comments on the trajectories shown are also lacking, beyond aggregation through income classification. This is important, especially considering the outreach of the Journal.

- I do agree that the GENEPI components and the A and b vectors here presented stem from a different perspective, mathematical reasoning, and narrative. Yet, I would like to see the correlation values between the X_1 and X_2 values without the transformation to project the values onto Fitness and ECI. How do these values alone figure in your grouping?

The actual representation of the stages was made onto this two rather than their projection onto the axis. If the correlation is high, or low, what further information would thus this measure explain? Following that, how would you explain the lack of correlation of the $X_2/\sqrt{d}$?

- In Supplementary Note S1 regarding the RCA, I'd rather not include the comment on the RCA effect as a note (which is not even admissible by the NC guidelines), better placing it in the main text.
- I would recommend checking editorial requirements

Again, I wish to thank the Authors for the high-level discussion, which touches a few streams of research from the past and ongoing, and I'd be glad to go further with the discussion in other rooms.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)

Bridging the short-term and long-term dynamics of economic structural change is an interesting paper that adds to the ongoing literature on economic complexity. The revised manuscript presents an improvement from the previous version, but it still fails to incorporate some of the revisions suggested by the reviewers. At these point, these are changes that mostly require rewording some sections. I hope the authors take these changes into consideration and use them to improve their next version of this manuscript.

1. The authors seem to misunderstand the use of the word complexity as they see their derived measure of complexity $b(t)$ as a measure of composition. Maybe, this is because the authors are unfamiliar with the etymology of the word. According to the online etymology dictionary complexity refers to the "composite nature, quality or state of being composed of interconnected parts." Wikipedia's entry for complexity also uses "a system or model whose components interact in multiple ways." Oxford's dictionary entry for complex defines it as "consisting of many different and connected parts." Complexity is therefore a noun that correctly describes a composite system (e.g. made of multiple industries, products, etc.), which is what metrics of complexity have long tried to communicate (a description of economic composition). Their measure of complexity $b(t)$ is a measure of the long term composition of the economy. The authors should positively leverage this connection with the literature and the etymology of the word instead of antagonizing it.

2. While the work does connect to several efforts in the literature, there is recent work that has advanced our mathematical understanding of economic complexity metrics that would be interesting to see discussed in this context. In particular, the works of Bottai et al. and Servedio et al. providing interpretations of economic complexity. These contributions give a closed form solution to the objective function optimized by complexity (the sum of $ECI-PCI^2$) and build on the bipartite clustering interpretation of Mealy et al. (2019)

Bottai, Carlo, Jacopo Di Iorio, and Martina Iori. "Reinterpreting Economic Complexity: A co-clustering approach." arXiv preprint arXiv:2406.16199 (2024).

Servedio, Vito DP, Alessandro Bellina, Emanuele Calò, and Giordano De Marzo. "Economic Complexity in Mono-Partite Networks." arXiv preprint arXiv:2405.04158 (2024).

3. Figure 3 is an important part of the paper that would benefit from a properly labeled panel (a scatter plot with ISO codes for each country). I understand this would require a larger panel, but it would be very for economists with an intuition of each economy. For instance, where are low diversity high complexity economies such as Finland or Singapore located? Are they considered tax heavens? Pushing this to the appendix obscures valuable information for the reader. This should be an easy comment to incorporate.

4. From the early days, the validation of metrics of economic complexity has involved econometric models where complexity metrics are used to explain macroeconomic outcomes (first growth, but then inequality, emissions, etc.). The paper would benefit from including such a model in the appendix and mentioning its results briefly in the main text. Does ECI^* behave equally to ECI when it comes to predicting growth? Worse? Better? How about other outcomes? The reader is left wondering about the impact of this work (does it change the ability of ECI to explain an outcome that economists care about (e.g. growth)?). Adding at least a growth regression should help in this regard.

5. On page 9 the authors try to build a distinction between a system made of two coupled equations (F and G, where F can clearly go into G resulting in a single equation) and a system made of a single equation. Basic college math tells us that coupled systems are a way to write a system instead of a fundamentally different idea (e.g. a second order differential equation can be expressed as two coupled first order differential equations, a Hamiltonian uses two equations and a Lagrangian one, but you can go from one to the other). The authors should again try to leverage the connection with the literature, or remove this section.

6. On page 9 the authors also seem to confuse the complexity of processes and products. Crude oil is not a sophisticated product, but its extraction is a complex process involving complex products (Eg. Excavation machinery (HS8431) is a high complexity product (about one standard deviation above the world's average). Trade data is fine grained enough to differentiate between the two (e.g the countries that export oil extraction service activities and machinery are more sophisticated than those that export the oil). This seems to be a false "conundrum." I would suggest removing this section.

Smaller points.

Page 2. "our results suggest they re not in competition, but offer complementary information." This point was already made by Sciarra et al. in the GENEPEY paper. Should be acknowlegde as such, instead of as a new point.

Page 9. What is the point of calling previous research pointless? Isn't the previous research motivating your current work?

Page 1. "We do not endorse the view that the PoR offers a complete description of long-term economic transformation". Is there such as view? Could the authors provide a reference for it? If not, they should remove this sentence.

(Remarks on code availability)

Version 3:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Dear Authors,

Thank you very much for your work. As I mentioned previously, the work has improved in many ways. I fully endorse publishing this work, from which our community will definitely benefit.

Carla Sciarra

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)

No further comments.

(Remarks on code availability)

No further comments.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source. The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Reviewer #1 (Remarks to the Author):

General comment:

In this work, the authors introduce a network model of an economy's structural transformation by which showing the connection between the two fields of study of the economic complexity, i.e., the study of the economic activities and the creation of proper metrics for the identification of the productive knowledge. The work is mathematically sound and the valuable contribution to the RCA approach is of particular interest when focusing on niche productive sectors. I endorse the publication in some form, as I argue that this work can actually be useful for all the academics engaged in the study of economic complexity and, even more, for decision makers aiming to use the resulting information. However, the arrangement of the arguments obstacles the transfer of knowledge to any of these stakeholders. The main text is not readable as a stand-alone work without previous knowledge on network science or economic complexity and requires a back-and-forth movement with the SI to understand the framework better. In this form, it exits the aim and scope of a multi-disciplinary journal such as Nature Communications. Thus, the significant revision here presented relates to the need for the authors to rearrange the argumentation and discussion to foster the engagement of the reader, rearranging the logical structure of how results and comments in the text.

We thank the reviewer for their extensive feedback and their many suggestions to improve the paper below. We made changes to try to make our arguments easier to follow, and make the text more readable as a standalone work for readers outside network science and economic complexity. In many cases we did this by pursuing the suggestions of the reviewer.

Major revisions for text:

1. Applicable to all text: cross-references to the methods and SI lack and details are thus lost in the reading). There's also a missing piece of information that comes with the figure, and often the text refers to detail which are just given in the caption of the figure (e.g., Figure 2, the description of the network is only given in the caption, but clear references are in the text without further information on how the network is constructed).

We thank the reviewer for underscoring this problem with the text. We significantly revised and reorganized our material to make it understandable without frequent recourse to the SI:

- A significant amount of material from the SI was moved to the Methods section of the main text to make it easier to find.
- References to the Methods now explicitly give the relevant section or equation.
- References to the SI now explicitly give the relevant section.
- We eliminated the need for one of the SI references by including additional algebraic steps to reach Eq. 5.

We reviewed references to figures to ensure that critical information is given in the text. In the case of figure 2 we verified that the text doesn't reference quantities or facts that only appear in the caption. The quantities discussed in the caption (revealed comparative advantage, the conditional probabilities $Q_{pp'}$, the proximity matrix Φ^M , the normalized proximity matrix $\tilde{\Phi}$, and the eigenvectors of $\tilde{\Phi}$) are defined in the text itself.

2. Introduction:

a. As structured, the introduction is unbalanced, pushing toward argumentative text in support of the findings of the work (see pages 3 and 4 – that can be suitable for the discussion section instead) rather than presenting a proper context on where to locate these findings. An example is given by the GENEPI metrics, which is not presented in the introduction, while it is cited in the caption of figure 1 (the index is not introduced until the results section despite being a recurrent reference). The same applies to the concept of complexity, latest paragraph: here, the readers are introduced to something which is still to be adequately defined, and that remains suspended until the end of the Results. Actually, the concept of complexity, and how this is intended with respect to the field of study, is something that can be of interest to be briefly detailed here in the introduction section.

We thank the reviewer for highlighting the imbalance and the need for more context. We significantly revised the abstract and introduction to try to address these issues. We now open the abstract with the topic of structural transformation generally (rather than the specific approach to this topic represented by the emerging field of economic complexity). We expanded on the concept of complexity and its significance for understanding the development process. We refer to the GENEPI metric by name. We condensed the material on pages 3-4 to give it less prominence. (We retained it in the introduction because we think it will help readers understand what findings we are going to develop later in the article.)

b. When introducing ECI and Fitness, the authors describe the bipartite using the sentence "which locations produce which products". The wording is imprecise compared to the methods. It is worth amplifying the discussion regarding production and export to unfold the difference between the two economic activities and why the RCA method is recursive in the metrics definitions (see description at the end of page 4, Results section).

We greatly appreciate these suggestions. We made the description of the bipartite network in the introduction more precise. In section 2A (page 5), we added a sentence noting that exports are often analyzed because of the availability of harmonized data across countries of very different levels of development, and because the ability to export competitively is often considered an indicator of production ability. In section 2D (page 18), we added sentences explaining why several complexity metrics use recursive definitions.

c. Fig. 1 can be deleted considering adding a reference to its equal in the SI. It isn't easy to understand its meaning here at the beginning of the manuscript without knowing any detail regarding the variables on display.

We thank the reviewer for highlighting this; we agree the inclusion of this figure could be confusing. We added language to better state its point, which is that nearly all complexity metrics empirically fall into just two kinds of empirical behavior (this is not known), and that this is surprising because of the variety of theoretical arguments that have been put forward to arrive at different metrics, and because these theoretical arguments are what have mainly received attention in justifying one metric over another. We note this figure raises the interest of an important part of the audience; debates about the differences between complexity indices have gotten quite heated at times, and the figure indicates that, in great part, this is not because one index outperforms others or is more theoretically sound, but because different metrics shed light on two complementary aspects of economic development.

3. Results:

a. Paragraph of Eq. (2): can you detail which network you refer to with the sentence "The PoR treats diversification as a process of spreading on a network."? I guess this one is the one from which you obtain a Laplacian network for figure 2 (and so it is the bipartite network of countries and exports). Still, details might be helpful to ease the reading of the text without recurring to the SI.

Thank you for pointing out the possible confusion here. We now say above equation 2 that the spreading process, in general, occurs on a network of economic *activities*. We refer to the network of exported *products* in figure 2 as one possible example.

b. Some more math references in the main text can help the non-expert reader to follow along. The comment applies to all the sections. The main text gives so many details for granted, and it should be supported instead to allow the general public access to the results. (eg. The ordering of the eigenvalues).

We thank the reviewer for making this suggestion. We added citations to relevant math references. Above equation 2, we added citations to works that discuss relevant background in network analysis and the method of eigenmode decomposition in dynamical systems. (For the latter case we cite a work that deals with eigenmode decomposition specifically in the context of dynamical network models, making it particularly suitable as a background reference.)

Other comments and food for thoughts

- *As shown in Fig.3, tax havens and resource-rich countries are evident free riders in the economic development panorama. How truthful and valid is it to use your model on the export data knowing that these two kinds of countries can bias the results? This is particularly important when comparing the metrics' results with the GDP. This comment also relates to the note the authors made on oil: as described in the manuscript, oil extraction does require a broad set of knowledge, yet, the presence of the resource first defines the activity.*

In fact, we think an advantage of our framework is that we can avoid strong statements about what drives a country's evolution (whether a tax haven or resource-rich country, or a more "ordinary" case), and focus on describing what countries do empirically. We show that the Principle of Relatedness, which has been repeatedly supported in the literature, theoretically implies that changes of two kinds – tracked by the coordinates A and b in our paper – describe particularly difficult-to-achieve changes in industry baskets. But however one arrives at these coordinates, once derived, we can ask what countries are doing in this particular coordinate system, and ultimately these quantities characterize simple ideas (A relates to overall diversity of output, b to progress along a particular axis of compositional change, largely an agriculture-to-manufacturing shift). As it turned out, tax havens and resource-rich countries display a distinct behavior from other countries under these measures (i.e. a useful difference to expose).

Oil extraction exemplifies our point that our approach can rationalize the mysterious "complexity" value of oil. If oil extraction is complex, then one would like complexity metrics to detect this whether or not one also needs (additionally) oil to be in the ground. As it stands, these metrics score oil extraction as being low in complexity. This is at odds with the usual interpretation of these metrics but not with the interpretation we explored here, where oil's low value in these metrics instead reflects the fact that as a country develops this product should become less prominent in the export basket over time in relative terms. (I.e. the country will diversify and develop additional, manufacturing-centric export categories.)

- *To the previous related: besides the export ones, with which other kinds of data can the PoR model be used to extract information on countries' growth?*

We mention some of these directions in the original version of the sentence below. We now expanded it to refer to two other types of data that the PoR has been applied to (technology patenting and research publications):

"We explored our results in the setting of country exports, but similar analyses could be carried out in networks of occupations \cite{Muneepeerakul2013}, industry output \cite{Neffke2011, Hausmann2021}, technology patenting \cite{Kogler2013}, research publications \cite{guevara2016research}, or in networks of related locations \cite{Bahar2014}, where the roles of locations and activities could be reversed by projecting the bipartite network of locations and activities onto location nodes instead of activities."

• Fig. 3:

• *the border between R2 and R3 is at value (circa) $A=0.4$ value. Does it have a mathematical meaning and interpretation?*

There is not a particular meaning to $A=0.4$, though this value was chosen to (roughly) separate OECD countries horizontally from non-OECD countries, as shown in Fig. 3c. Most OECD countries lie to the right of $A=0.4$. Fig 3b and 3c were provided in part to help provide intuition about the values of A and b . Fig. 3b connects values of A to values of a standard diversity metric used in the literature. (Where *diversity* is the count of products in which a place has a revealed comparative advantage exceeding 1.) Fig. 3b shows that countries with a value $A=0.4$ typically export 20-30 such products, whereas the most diverse countries export a little over 100 (out of a maximum of 235 product categories in our data).

• *Despite the full disclosure regarding the dynamics of the ensemble of countries, there's reason to believe that the wind speed at the interface among regions can be motivated by the trajectories of some specific countries rather than others. Have you explored these results further and detailed the contribution of each country's trajectory to the ensemble?*

We agree, and we have put in the SI a plot that shows countries' individual trajectories (in Supplementary Note S2). We have added a further commentary in this section to note again that our quiver and streamline plots in Fig. 3e of the main text depict the average movement of countries in different parts of the (A,b) plane. These averages are dominated by just a few countries (or just one) in parts of the plane where we only observe a few places. We mention an instance where this small-numbers effect occurs.

• *All text and Section S1.8: how do the present arguments relate to the micro-model of structural change from Hausmann, Hwang & Rodrik? I argue that this analysis can further add value to your work.*

We thank the reviewer for this suggestion. Our approach is rather different. Hausman, Hwang & Rodrik build a micro-style model of local cost discovery. To provide empirical evidence for their theoretical arguments they introduce metrics (EXPY, PRODY) that proxy for variables in the model. These metrics mathematically ended up being precursors to later complexity metrics, and similar to complexity metrics, they produce a ranking of countries and products along a particular dimension (in their case, a productivity level associated to different countries and products).

Our primitive instead is the product space. We take as a given that there is a product space that captures the structure of relatedness between products, and on this space a place can move to new products. This space has a non-trivial network structure. Because of this, some patterns of movement should be expected to take longer than others, under the dynamics

reported by PoR studies. The network structure indicates which directions of change should be the slowest, and yields the coordinates we describe for where a country lies along these directions. The coordinates we derive therefore hinge on what this network looks like, and in general, there are many possible directions, corresponding to different patterns of change. Thus our framework doesn't have a micro model as such nor does it emphasize any ranking or 1-dimensional characterization of activities. However, we do cite a paper, Schetter 2019, that examines conditions (within a general-equilibrium trade model) under which a country's export basket can be effectively summarized with a single measure (which in their analysis turns out to closely resemble the ECI).

- *Can you show the time series of the correlation among the metrics?*

We thank the reviewer for this suggestion. We now show the time evolution of correlations among the metrics (Supplementary Note S4, Figure S4). These show that the correlations within the two groups of metrics were high across time (with one exception we note below). In addition, they show an extra effect that the correlations between these two groups were higher in early years than in recent ones. We added a figure showing why, displaying the scatter of regions and showing that in these years regions of our data overwhelmingly lie in a diagonal band where a region's A coordinate is strongly related to its b coordinate (i.e. there are few instances of places with low values of A and high values of b). An exception to the high within-group correlations is the quantity $X_2 / \sqrt{d}$ (where X_2 is a component of the GENEPI metric), which in early years is weakly correlated with other compositional quantities. GENEPI was originally studied on data in the years 1995 - 2017, and thus we are examining this quantity in a period far outside the one in which it was tested (we go back to 1966).

- *In detailing the comparison with the GENEPI, the authors comment on the difference between the two metrics, considering the nature of the proximity matrices upon which the metrics are constructed. Referring to the PoR, isn't there a shrinking of information by deriving countries' coordinates through the projection onto the axes?*

We thank the reviewer for this question. Since we don't assume that our coordinates (or any other metrics) necessarily capture complexity, we are not concerned that we are losing information due to us using a two-step procedure for obtaining country coordinates, where we first get the directional axes, and then project countries onto them. (Whereas GENEPI uses the elements of an eigenvector directly, which comes from a country-to-country similarity matrix.) Our method is an analytical recipe that comes from dynamical systems theory. It tells one how to derive slow directions of change in any linear dynamical system, and for such a goal, it would have been wrong for us to follow any other procedure. This is why – given the similarity of the coordinates we derive using this procedure with complexity metrics – we've questioned complexity metrics, and theorized that if there is a loss of information it goes the other way: Complexity metrics find success because they characterize slow changes that economies commonly undergo during development.

- *Since the GENEPI also identified stages of growth, how are these related to the Regions of growth you identify with the evolution of the terms A and b? In this sense, following the trajectories of the countries between the two planes could be of interest.*

Roughly, undeveloped countries in the bottom-left of the GENEPI plane correspond to our Region 1, and developed countries in the top-right of the GENEPI plane correspond to our Region 2. It is not clear what part of the GENEPI plane corresponds to our Region 3. This has to do with a non-linear transformation that approximately translates between the quantities (X1 and X2) that define the GENEPI plane and our (A,b) coordinates. In the (A,b) coordinate system, Region 3 countries deviate from others more so than they do in the (X1, X2) coordinates.

Beyond this technical issue of addressing the non-linear transformation between the two planes, we think a direct comparison of stages of growth could generate confusion because of a key conceptual difference in how we approached our (A, b) plane from how the GENEPI work approaches its (X1, X2) plane. In the GENEPI work, X1 and X2 are treated as different, imperfect measures of the same fundamental quantity (i.e. complexity). In contrast, our axes are different quantities from one another, and neither is assumed to have a relationship to complexity. We noted the formal mathematical connections between our coordinates and GENEPI and its components, along with similar connections to ECI and country Fitness, to underscore that it is possible to arrive at quantities resembling complexity metrics from the very different starting point of a dynamical model of country diversification. Detailing the mapping of our country evolution figures to the GENEPI stages of growth in particular (which is a framing of growth trajectories unique to this particular paper) takes us outside the aims of our work, and we also worry that it suggests that we are aiming to improve on the GENEPI plane. (Whereas we are not trying to improve the methods of any complexity metrics.) Where we go into detail about how our coordinates relate to existing complexity metrics, what we mainly want readers to understand is that there are such mathematical connections, and that existing complexity metrics may capture important information not because they are inferring the underlying complexity of places and products, because they are describing well-known, long-term changes that countries undergo during the process of development.

Reviewer #2 (Remarks to the Author):

The paper develops an integrative framework for measuring economic complexity (using national export data) to describe the growth and transformation of economies. There have been many measures of economic complexities (diversity, fitness, genepy, ECI, and entropy), and the paper provides how these measures correlate (FIG 1). The paper also provides how these compositional (diversity) measures imply the temporal evolution of the economy (as in Eq 1-4) and show the coordinate along the evolution (Fig 2 and 3). And finally, the paper shows how this evolution is decomposed into two parts indicating two coordinates (as in Eq 6), and thus corresponding to economic quantities: GDP and diversity (FIG 3).

My overall impression of the paper is very good. The authors demonstrated well their knowledge of the subject matter, and their derivation is technically sound, but it still needs to be articulated and goes deeper to understand fundamental mechanisms beyond mathematical interpretation. My major concern is 1) how much value is in the paper's contribution, and 2) if it is aligned with the intellectual interest of the readership of Nature Communication. In particular, it is unclear whether the manuscript is supposed to be a review paper or a new contribution as much of the content is recapturing the previous papers, including Sciarra et al. 2020, and Mealy et al. 2019. Therefore, as it currently stands, the manuscript contains several shortcomings (listed below) that render it inappropriate for a high-impact journal like Nature Communication.

We greatly appreciate the reviewer's feedback. We have made changes to try to explain the paper's value and novelty more clearly than before. As a start, we worked to be more explicit about the central contribution of this work. Our core contribution is to develop a modeling framework for structural change. We do not aim here to create a framework to integrate the wide array of complexity measures, *per se* – though our modeling framework does carry implications for interpreting many of these measures. We revised our introduction and language throughout to more clearly label our proposal of a modeling framework for structural change. This included, following the reviewer's suggestion below, improving our language where we used phrases like 'our theory', 'our derivation', and so on.

Our central contribution is a dynamical modeling framework that links three areas of literature:

- 1) The literature on the Principle of Relatedness (PoR)
- 2) Classic observations of long-run structural change
- 3) Complexity metrics

We can summarize the linkages to these areas one-by-one. First, we develop a dynamical model (Eq. 2) that is based on regressions used in the literature on the PoR. Surprisingly, although these regressions describe a dynamical process, no one in the last 15 years formally rendered the regression formulas as dynamical models, in order to exploit well-established methods in dynamical systems theory and further understand implications of the PoR. Our starting point was to do so, developing a model theoretically, and taking it to exports data.

In doing this, we framed the PoR literature in a somewhat atypical way as a study of structural change (i) on short (e.g. year-to-year) time scales (ii) using granular sectoral data. A major concern of ours was whether a dynamical model derived from this literature would be consistent with what is already known about structural change through decades of research; the second area above. What we show is that, in fact, our model is consistent with this research, implying long-run diversity and compositional changes in an economy that are recognized as integral to the structural transformation process. We find this is remarkable; apparently, one can derive correct information about the long-term, permanent development of economies from a simple model of year-to-year dynamics. This did not need to be the case, since the model could have worked on short time scales, while failing badly in its long-run predictions. By bridging short- and long-run structural evolution with one equation, we think we have arrived at a modeling framework that many researchers should be aware of and consider building upon. In general, the details of our specific dynamical model can change (and be improved hopefully), but one should always be able to follow the analytical recipe we use, which includes analyzing the long-term behavior of the model using dynamical systems techniques, and developing coordinates that describe an economy's progress in executing various patterns of change.

Finally, for researchers who are most familiar with economic complexity research, the functional forms of the diversity and compositional coordinates that we derive clearly resemble complexity metrics that have become popular in the last 13 years; the 3rd area above. These metrics are used in many applications spanning industrial policy, growth forecasting, and sustainability, but have also generated controversy over their interpretation and method of calculation. We connect our modeling framework to these metrics, showing how most existing metrics theoretically and empirically resemble one or the other of the two coordinates that emerged, supporting the empirical relevance of these metrics as measures that quantify important information about economic transformation.

The mutual reconciliation we reached among these areas gives us confidence that our modeling framework is sound and interesting. Besides proposing this framework itself, we point out some immediate implications that it has for complexity metrics. While supporting the empirical relevance of these metrics, our framework tells us to question whether their success truly stems from the reasons that are usually given. (I.e. that they infer the complexity of an economy.) We instead suggest that these metrics have predictive value for growth because they are useful in summarizing overt changes in the output mix of countries as they become richer. This would also mean that the competition between these metrics could be irrelevant, since the main competing choices really give complementary information, and none necessarily capture the originally-intended concept of complexity. Other insights emerge from our modeling framework as well. For instance, we exploit the intuition that comes from thinking about the dynamics of the model to offer an explanation why some sectors, such as oil extraction and works-of-art, have “complexity” values that seem surprising when one adopts the usual interpretation of these metrics.

While our main contribution is our dynamical modeling framework, even if we focus on our contributions to interpreting complexity metrics, we go beyond what others do in at least two ways. First, we don't re-interpret just one metric (as in Mealy et al.) but many. Second, we directly connect complexity metrics for the first time to a model of an evolutionary process, rather than describing a formal mathematical unity (as in Sciarra et al.) Our ability to generate insights about complexity metrics is closely tied to the fact that we began from a model. While works such as Mealy et al. and Sciarra et al. looked deeply at the methodology of complexity metrics, here we could also exploit intuition that came from thinking about the evolutionary process that our coordinates were derived from.

We agree that these points were not adequately developed and that the novelty of our contribution was not fully clear. To better convey these points, we made a number of revisions, starting with the abstract and introduction. We now begin the abstract with the issue of understanding structural change generally. (Rather than the specific approach to this topic represented by the emerging economic complexity field.) We more explicitly and more consistently refer to the contribution of our modeling framework. We increased the prominence of the classic stylized facts as a prong of literature that we are connecting with, and condensed introductory material on complexity metrics. Below we discuss other detailed issues raised by the reviewer.

The manuscript summarizes many different measures, presenting their correlations (two groups) nicely in Fig 1. This result was, however, already discussed in Mealy et al. 2019 and Sciarra et al. 2020, and the orthogonality between diversity and ECI has also been clearly discussed in the aforementioned papers.

Thank you for this comment. We agree that there is some similarity to Mealy et al. and Sciarra et al., who do look at a few of these correlations, though they do this in order to make very different points:

- Mealy et al. show the correlation of diversity and ECI (Figure 4 in their work). They do this to point out that diversity and ECI can have any correlation (positive, negative, or near-zero) despite the well-known orthogonality result by Kemp-Benedict that the reviewer refers to.
- Sciarra et al. look at the correlations of their $X_{c,1}$ quantity with Fitness and $X_{c,2}$ with the ECI. They do this to show that it is theoretically possible to approximate both country Fitness and ECI, simultaneously, with eigenvectors from a single country-to-country similarity matrix.

Neither of these texts make the point that most complexity metrics fall into two categories of behavior. We revised the reference to this figure in the introduction to better explain its purpose. We note that it is surprising that most complexity metrics fall into two kinds of behavior, given the variety of theoretical arguments that have been put forward to arrive at different metrics, and the emphasis on these arguments to justify one metric over another. We thank the reviewer for the opportunity to clarify this.

Regarding the orthogonality result, it is important to note that Mealy et al.'s orthogonality is different from the orthogonality (or really linear independence) that arises in our work. When we discuss diversity and ECI as capturing two axes, we're referring to the linear independence of the eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ of our framework, which represent two different kinds of transformation patterns that countries can evolve along. An economy's basket of activities can evolve in direction $\mathbf{v}_1$ independent of movement along $\mathbf{v}_2$ and vice versa. (I.e. a linear independence between two abstract data directions, described by vectors with dimensionality equal to the number of products.) Movement along one represents change in a place's diversity and movement along the other represents change in output composition. Mealy et al. note instead a built-in constraint that the ECI metric has within any given data set, an issue first reported by Kemp-Benedict 2014. This constraint is that the vector of measured values of ECIs for a given set of places will always be orthogonal to the vector of measured diversities of these places, due to how the ECI is formulated. (I.e. an orthogonality between vectors of observed values, whose dimensionality is the number of countries, as a constraint on obtainable values for a particular complexity metric. The constraint is weak, though, for reasons noted in Mealy et al.)

The temporal evolution is also discussed in the previous papers in a clear way (sections "spectral clustering" and "diffusion map distance" in Mealy et al. 2019). The paper's contribution is to make this discussion more explicit in time. But there was no further discussion of underlying mechanisms or independent observations beyond the simple mathematical recasting. The inferred coordinates and dynamics are again already discussed in Mealy et al. and Sciarra et al.

Mealy et al. were not claiming a diffusion map describes the evolution of countries. They just mention this technique as another dimensionality reduction method that corresponds to the construction of the conventional ECI. "Time" is a fiction in the method of diffusion maps that acts as a parameter to tune how strongly one wants to collapse data, and Mealy et al. do not indicate they are promoting "time" to something real. Even if they had, we are not just re-using the math of diffusion maps, and simply saying time is real. Eigenmode decomposition leads to distinct coordinates, ones that are justified by dynamical systems theory.

Sciarra et al. also are not making our points about temporal evolution. They show how the components of the GENEPI metric (X_1 and X_2) evolve (viewing these components as complexity measures), and argue that countries all roughly follow the same pathway in the space of these measures. This differs from the temporal dynamics we are reporting:

- The X_1 and X_2 measures scatter countries differently than our A and b coordinates do. The relationship between X_1 and X_2 and our A and b coordinates is approximate and non-linear, with important consequences for how countries get distributed. For example, differences between region 3 countries and others are more apparent in our A and b coordinates than in the (X_1 , X_2) coordinates. In general, in our coordinates, countries look more diverse in their behaviors.

- The GENEPY work is viewing X1 and X2 as measures of complexity. The GENEPY plane is plotting two methods of estimating complexity against one another. (Note that the GENEPY work does not, for example, ask what it means to vary independently in each dimension, e.g. what it means when a country is low in X1, but high in X2, etc.; such an analysis makes less sense to do if both axes are regarded as measuring the same fundamental quantity.) Sciarra et al. say that all countries roughly move along one pathway (i.e. the country data is 1-dimensional to some approximation), and that the X1 and X2 components capture complexity, i.e. they “track the path toward prosperity of countries as driven by economic complexity” (Sciarra et al. page 5). In contrast, in our work our axes are simply different quantities (and neither is assumed to represent complexity). They have intuitive meanings; one axis accounts for the diversity of the economic basket and the other its composition. Countries clearly can in principle vary independently in each direction, and we see that they do. We also observe that many, but certainly not all, follow a pathway that corresponds to well-known stylized facts of structural change, and also that a variety of countries simply do not fit this mold.

I like the current manuscript because it rewrites the previous findings in their own language/framework, but the contribution does not address mechanisms beyond the equations. Therefore, it is hard to see if the current version is an enough substantial contribution to be published.

We appreciate this feedback. We have tried to explain better why this paper is not simply rewriting earlier findings. As we note above our central contribution is a dynamical modeling framework for structural change that mutually reconciles: (1) The Principle of Relatedness; (2) classic studies of structural change; and (3) complexity metrics. In area 2, our framework affirms classic understanding, while in area 3, our framework revises existing understanding.

We strongly agree that there is more to be said about mechanisms underlying structural change and diversification. We do not delve into this, simply because there is enough to sort through to analyze existing findings in new ways (e.g. with the tools of dynamical systems). However, a model like the one here can help the goal of uncovering deeper mechanisms, by providing a tested theoretical setup one can elaborate on further, with the assurance that it is capturing important aspect of both short-term and long-run change. (Incidentally, some of the coauthors are also currently working on new efforts to model deeper mechanisms, in light of the dynamical model here.)

Minor issues:

1. Please rephrase ambiguous sentences and phrases. For example, the authors use the term ‘our theory’ without a concrete definition of their theory. Examples include sentences ‘here we present a simple theory that suggests these branches describe the same dynamic process of develop... Our theory begins by making explicit the dynamic models that are implicit...’

Sentences are on and on with no explicit definition of the authors' theory. What is the theory here by the way? Even to the end of the manuscript, I couldn't figure out what the authors mean by a theory. I can only guess them, but still I wasn't sure if the observation or explanation are meant to be a theory. Derivation is not a theory, and explanation is not a theory. So, I am really left unsatisfied. There are many cases. Please be specific and clear about what is being discussed.

We thank the reviewer for underscoring this problem. We made changes to clarify what we are referring to, revising phrases like “our theory”, etc. to be more explicit and more consistent. We revisited each of these phrases and in the majority of cases revised them to refer to our dynamical modeling framework for structural change, which consists of a dynamical model that is based on the empirical work of the PoR literature, and analyzed with standard techniques from dynamical system theory to extract the long-term implications of short-term dynamics. We see this as a modeling framework, and not just as one model, because others can (and hopefully will) improve the dynamical model at the core of the framework to make it describe country evolution ever more accurately, while being able to carry out all the other parts of our analysis (eigenmode decomposition, the analysis of different modes of economic transformation, the derivation of summary coordinates of long-term change).

2. Secondly, please be cautious about using emphatic adverbs/adjectives. In particular, instead of telling what readers are supposed to feel, please state what exactly they are and be informative. For example, in the abstract, ‘these metrics take on radically new meanings, ... profound mathematical connection between...’ There are many other cases, including: “...there is a surprisingly close correspondence...”, “interestingly, these two groups of quantities resonate surprisingly well...”, and “somewhat alarmingly...” Of course, I do not think every use of those words is bad, but the paper uses them more than enough. Please use them with caution.

We thank the reviewer for drawing attention to this. We’ve reduced the number of emphatic words.

3. How does the decomposition of short-term and long-term time scales of dynamics differ from detrending in statistics?

In detrending one decomposes change into long-term and short-time dynamics that occur together, while here, short-term and long-term do not refer to a decomposition but to the behavior of one model on different time scales. We derive the long-term behavior resulting from the short-term dynamics of the model (Eq. 2).

4. The manuscript is somewhat narrowly constructed. It is written only those who are interested in ECI and product space because the manuscript is intended to integrate the literature of the field of ECI and product space. I put this as one of the two major concerns not because it is a

deal breaker but because I hope the authors recognize this issue. It is not a deal breaker because I am sure that the authors can easily fix it by providing the context for the broader readership.

We thank the reviewer for pointing out this issue. This may stem from a tension between communicating with two types of readers we hope to reach, a broad audience of researchers interested in growth and development, who need more background on the methods and concepts used by the economic complexity approach to these topics, and the subset of these researchers who come from the field of economic complexity. The latter are deeply familiar with the methods and debates we address and may want us to “cut to the chase”.

We made several kinds of changes to try to make the broader interest of our work more apparent:

- In the abstract, we now open with the issue of structural transformation of economies. This comes first, before we turn to the specific approach to this problem represented by the emerging field of economic complexity.
- In the introduction, we elaborated on the concept of complexity to say more about what this term means and why it has become of interest for understanding the development of economies.
- We revised the introduction to more prominently highlight the connection to classic findings about structural change (which of course go back decades and lie outside the recent approaches to this topic in economic complexity), and condensed the discussion of complexity metrics.
- We more explicitly highlighted our central contribution being a modeling framework for structural change. (I.e. not a framework to re-interpret complexity metrics per se.)

Reviewer #3 (Remarks to the Author):

Referee report for Nature Comm

I have answered to the “template” questions suggested by the Nature Communication below.

I think this is a great paper. It is literally a cornerstone of the literature on economic complexity and the principle of relatedness, and within that literature it will quickly become a classic. The paper is also technically very solid.

We thank the referee for their support and for their helpful feedback below.

I have two main comments:

1) To what extent is it okay to select only one eigenvector? (I mean, one more than the uniform one). In other words, what is the distribution of (the first few) eigenvalues? The authors never discuss this. Ideally, I would have like to get a sense of the “effective rank” of the proximity matrix (see “low-rank hypothesis of complex systems” <https://arxiv.org/pdf/2208.04848.pdf> Table 2). This is asking too much, but I hope the authors can at least plot the spectrum or describe the first few eigenvalues.

We thank the referee for this great suggestion. We expanded our discussion of this in the SI in a new section (Supplementary Note S3) and plot the eigenvalues. In brief, there is nothing in our framework that says one must stop at the 2nd eigenvector. A main issue is that there is uncertainty in the proximity matrix, which impacts both the eigenvectors and eigenvalues. The proximity matrices in the literature have strong agreement only in the structure of the first two eigenvectors, and do not straightforwardly resolve how many eigenvectors matter. To resolve these questions, one has to determine the right proximity matrix. (And from there our framework provides tools to analyze this matrix.) We discuss these points in the new SI section, and offer some ways future work can address them. Given that our theory now provides a clear motivation to study (and a framework to interpret) the eigenvectors beyond the 2nd, we think one of the next questions is precisely the one the reviewer is asking, to determine the effective rank given a proximity matrix (which of course will depend on the method of estimation, the granularity of analysis, the amount of noise in the data, etc.)

Also, just to be clear, in S1.3.1, you would select the complex conjugate of the second eigenvector, and create a composite? Adding a sentence on mode selection in S1.3.1 might be helpful, although I recognize that this is out-of-scope given the choice of a symmetric matrix in the main text empirical application.

Correct, if one had a complex eigenvector we would advocate selecting this eigenvector together with its complex conjugate and treat them as one mode. This is a standard trick that

leads to a mode with a (real) decay factor and a (real) sinusoidal variation factor, and shows how the final data vector is guaranteed to be real even though it's decomposed in terms of complex eigenvectors. We added a sentence (in what is now Supplementary Note S1) to clarify this.

2) Do PoR studies use a constant Φ matrix, or is it updated over time? Is the Φ matrix more or less constant empirically? The assumption of constant Φ is necessary/convenient/clever (to solve Eq.(1) and move forward with this approach). But I would have appreciated a little bit more justification. I was not very convinced by the last paragraph in S1.8 – maybe it just needs to be expanded/clarified. This is a somewhat sensitive issue, because the authors never really expand on the details of all the time scales involved (see point 1), and one could think that Φ changes “quite quickly” compared to “how quickly” countries export baskets change conditional on a fixed Φ .

To our knowledge PoR studies almost always assume a constant Φ matrix to make predictions. We're aware of one investigation into how good an approximation this is (Hidalgo 2009, which we now cite). Using a fixed Φ and showing that it successfully predicts transitions does indirectly probe Φ 's stability, but we agree this is not necessarily satisfying and more study / direct characterization would help. We revised the last paragraph of this section (now labeled as Supplementary Note S6) to emphasize this as a direction of future exploration.

I have a number of smaller comments – these are either typos, or suggestions that the authors should feel free to ignore:

- It would be helpful to define “diversity” the first time it is used.

Thank you for this suggestion. Fully defining diversity the first time (which happens in the introduction) was challenging because the definition also involves defining RCAs and discussing how they are thresholded. But we added an example to help explain what ‘diversity’ can mean in practice.

- In Eq. (1), use = or $\equiv$, or please define “ $\sim$ ”. I don't think you do mean “have the same scaling as $N \rightarrow \infty$ ”? I guess you need to add a noise term or write the LHS as a conditional expectation?

Thank you for pointing out the possible confusion here; we modified this expression to clarify this. We changed the “ $\sim$ ” sign to an equals sign, and added a noise term to emphasize that these PoR models are regression models.

- After Eq. (3), make it clear that the last term can be negative (e.g.: replace “ability in p rises” by “ability in p rises, otherwise it falls”). I suspect the declines in ability are much rarer, so much more informative.

Thank you for this suggestion, we added a statement that this term can be negative, corresponding to a decline in ability.

- In fig. 4, there are some dots on the left that are vertically aligned: They all take the same value for the ECI, but different values for the ECI*. Is there something interesting going on here, whereby ECI* is better able to capture detailed patterns, while the ECI fails because it uses uniform weights?

Yes, it is related to the fact that the ECI uses uniform weights from binarizing the RCAs. Some countries have the same binarized RCA structure in many years. For example, the left-most vertical dots are Iraq, which in several years has only the SITC industry 333 crude oil with $RCA > 1$. Each year results in the same ECI of -0.1253. (Also relevant is that the ECI we use in Fig 4 is not normalized with the mean and standard deviation across countries in each year, as is often done, which would ordinarily wash out this effect.)

- P.7: “analyzed it in terms” -> “analyzed in terms”

Thank you for pointing this out, we fixed this.

- Fig 3: Consider giving more details for how Panel e) is produced. I suspect it might be too complicated to report – are you posting the code? Citing a software package?

Yes, we are posting the code. We also put additional details in the caption. These state that the figure was produced in Matlab, with the ‘quiver’ and ‘streamslice’ functions, and the other key decisions to reproduce this: what data we use (we use all countries, products, and years from Comtrade) and how we binned it – we use 20-year movements over time, and divide the space into 10 equal-size bins horizontally and 12 equal-size bins vertically.

- P11 : “rise in number” -> “rise in the number”

Thank you, we fixed this.

- P12: “higher wealth” -> I think you mean “higher income”(GDP is a flow, not a stock)

Thank you, we fixed this.

- P. 17: *The paragraph starting with “Fourth” is interesting but a bit hand-wavy (or maybe it is clear to the authors and I just don’t see the point as clearly): can this be sharpened?*

Thank you, we looked at this carefully and did not find a simple way to sharpen it. We hope it does not confuse the reader.

- P.19: *“Differences between countries shrink first among quickly-decaying dimensions...”: There is probably a connection to the literature on “convergence clubs”.*

Thank you for pointing this out. We thought about pointing out a parallel with the phenomenon of convergence clubs because we agree that there is a resemblance. We decided against it because we thought it might be too speculative and merits a more extensive investigation.

- *Section III tends to repeat ideas that were exposed previously.*

Thank you for pointing this out. Responding as well to other comments, we condensed some statements in the introduction that might (perhaps) have been ones that felt redundant with section 3. We hope this reduces the feeling of repetition since some of these points will not have been mentioned previously.

- S1.3.3: *“coordainte” -> coordinate*

Thank you, we fixed this.

- *Fig. S2 could have had a panel with $R_0 = 0.115$*

Thank you for this suggestion, we now include a panel with $R_0 = 0.115$.

- S18, p.13, paragraph starting with “First”. *Wouldn’t the elements of v_2 depend on the choice of R_0 ? This would be a helpful reminder/caveat.*

Thank you for this suggestion. We now include a reminder to the reader here that there are still numerical and sign differences between complexity metrics and our v_2 (and ECI^*). There are differences in how these quantities are all constructed that impact the details of v_2 (and ECI^*). (Including choices about the transformation of RCAs and the R_0 parameter, as well as what proximity matrix is used, and the use RCAs at all as measures of ability.)

What are the noteworthy results?

The authors show how the principle of relatedness relates to the complexity metrics. They proceed as follows:

- Start from a dynamic equation that is successful in empirical research*
- Solve it (i.e., express it as a function of time and initial conditions)*
- Diagonalize the solution, and express the approximate solution in terms of the first 2 eigenmodes*
- Study these eigenmodes – it turns out they are very related or identical to classic “complexity” metrics!*

Will the work be of significance to the field and related fields? How does it compare to the established literature? If the work is not original, please provide relevant references.

This is an absolutely major study in the field of economic complexity metrics and relatedness. Within this field, I think it is the most important one in the last few years. It is at least as significant, in my view, as the recent studies in similar journals (Sciarra et al [27], Mealy et al [50]).

Does the work support the conclusions and claims, or is additional evidence needed?

Yes, this is a very solid paper. I have made some suggestions – I’d like to see a bit more of the spectrum of L_{Φ} . But the core conclusion is already supported by substantial and technically solid analysis.

Are there any flaws in the data analysis, interpretation and conclusions? - Do these prohibit publication or require revision?

As far as I can tell, there are no flaws that would prohibit publication. I have made some suggestions for revisions although I think the study could almost be published “as is”.

Is the methodology sound? Does the work meet the expected standards in your field?

Yes, and yes.

Is there enough detail provided in the methods for the work to be reproduced?

Yes, it appears so (I haven't tried), except for the details of how to create Fig. 3e.

We thank the reviewer for their feedback. As discussed above we now provide the details to recreate Fig. 3e, and also provide the eigenvalue spectrum.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Before moving through the details of this second review round, I'd like to formally apologize with the Authors for the delay of this report. The due date that this report needed was conflicting with cumulated sum of workload and unexpected life events. Again, apologies.

Reviewer #1

With respect to the first version, I appreciated the effort the Authors put in implementing the requested changes making the article more readable to unspecialized audience. Yet, despite the improvements, I'm still questioning whether the contribution the Authors claim to make truly is groundbreaking as the journal would require, or re-enters in the debate characterizing economic complexity while increasing the noise around the amount of metrics available.

My discussion stems from what the Authors claim and what literature offers.

What the Authors argue to do in tying together two areas of work within the economic complexity field appears to be a little bit highfalutin. I do agree that the way the framework introduced in this work cleans up the many contributions the economic complexity literature has (Table 1). Moreover, it provides a clearer interpretation about the long- and short-term dynamics of structural change. However, I would reduce the flattering arguments. In fact:

The principle of relatedness lies beyond all the metrics of economic complexity, and the main difference lie in the way this principle is embedded in the proximity/similarity matrices the short-term metrics are build upon. The EC metrics from Hausmann and Hidalgo are the original examples for this statement. The productive space in the Hidalgo et al. Science, 2007, provides the ground for the ECI-PCI metrics to have mathematical reasoning. I may agree there is always room for improvement, but if one has to identify the original example of the principle of relatedness, and the way it relates to the metrics of complexity, that is where I'd search for. The same Authors refer to these pieces of literature, and there is clear mathematical resemblance in among the metrics the Authros present in the Supplementary Material – S7, and the Science paper from Hidalgo et al. Thus, I am left with doubts about the importance of this contribution.

Thank you again for taking the time to review our work and providing a constructive review. We greatly appreciate the push to have us be clear about why our work is groundbreaking. We also desire to avoid adding noise around this topic. Many papers have come out making various points about complexity metrics, making the subject matter complicated.

Before we address the feedback we discuss a bit of context. As the reviewer notes, two papers in this area are seminal:

- **HKBH 2007:** Hidalgo, C. A., Klinger, B., Barabási, A. L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science*, 317(5837), 482-487.

- **HH 2009:** Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the national academy of sciences*, 106(26), 10570-10575.

Essentially, the HKBH 2007 paper showed that economies tend to move into new industries that are related to ones they have already. This was early evidence for a general phenomenon that has since been called “The Principle of Relatedness” (PoR).

The 2009 paper did two main things:

- (i) It proposed a *conceptual narrative* of country development. Countries grow by acquiring ‘capabilities’. Different combinations of capabilities enable production of different products. Capabilities are largely hidden, however, and not directly observed – their existence must be inferred through the output basket of a country.
- (ii) It proposed a data method to infer the number of capabilities a country has. This proposal was the first complexity metric, the Economic Complexity Index (ECI). The name reflects the idea that countries with more capabilities are (by definition) more complex.

In the short time since 2007 these two papers have been enormously successful. They have thousands of citations, and spurred a research agenda now spanning thousands of papers across several disciplines (including economics, geography, physics, and environmental science). Further, these papers are having an impact on policy in many countries and in organizations including the IMF, the UN, and the EU. Ideas in these papers guided the largest place-based regional development subsidy scheme in the world (EU’s smart specialization platform). By ranking industries for industrial policy, complexity metrics are factored into decisions that move hundreds of millions of dollars.

The problem is confusion and debate over how these metrics work. How is it that they accomplish their chief selling point – that they infer an economy’s (or product’s) hidden capability count?

This is a highly non-trivial inference problem. Inference is difficult in general, but beyond this, there has been a contentious methodological debate specific to these metrics. Many practitioners are unaware of this debate. Some of it has involved proposals for other metrics (the most notable example being country Fitness, proposed by [3]). Other work has investigated the interpretation of complexity metrics (e.g. Mealy et al. [4] for the ECI specifically) or simply tried to decide between the main competing options [4]. The ECI has received the most scrutiny, though the problem (in our view) is more general – claims about complexity metrics are loose across the board. “Derivations” of complexity metrics are driven by heuristic arguments and then interpreted heavily within the narrative of capabilities introduced in HH 2009. Notably, this narrative sets up certain expectations about how these metrics ought to behave in data. As members of one of the groups that champions the ECI, we’re often asked

about cases where this metric does not behave as expected. (We elaborate below.) Given how these metrics are increasingly used for policy, it is potentially dangerous to not have clarity about what these metrics are doing. In our paper, we are seeking to constructively give a plausible explanation for this.

With this context, we think this part of the feedback above is key:

“In fact: The principle of relatedness lies beyond all the metrics of economic complexity, and the main difference lie in the way this principle is embedded in the proximity/similarity matrices the short-term metrics are build upon. The EC metrics from Hausmann and Hidalgo are the original examples for this statement. The productive space in the Hidalgo et al. Science, 2007, provides the ground for the ECI-PCI metrics to have mathematical reasoning. I may agree there is always room for improvement, but if one has to identify the original example of the principle of relatedness, and the way it relates to the metrics of complexity, that is where I’d search for.”

This is not what HH 2009 did. When HH 2009 put forward the ECI, they were not deriving the implications of the Principle of Relatedness for a country’s evolution. They were looking for an inference measure to estimate the number of capabilities a country has. HH 2009 proposed that countries’ observed industries are the outcome of two unobserved variables: the capabilities each country possesses, and the capabilities each industry needs. (With rich countries likely having many capabilities and poor countries few.) This narrative was asserted to be tied to a set of equations that lead to the ECI. Heuristic arguments were (and still are) used to motivate the mathematical form of a quantity that could accomplish the desired inference of a country’s (or product’s) capability count. These arguments vary depending on which source one looks at. The original discussion in HH 2009 emphasized how a complexity metric should relate to a country’s diversity. Another starting point has been to say that a country’s complexity is the average complexity of its exported products, and a product’s complexity is the average complexity of the countries that export it, leading to a set of simultaneous equations,

$$ECI_c = \sum_p \frac{M_{cp}}{\sum_{p'} M_{cp'}} PCI_p$$

$$PCI_p = \sum_c \frac{M_{cp}}{\sum_{c'} M_{c'p}} ECI_c$$

where M_{cp} is 1 if country c exports product p and zero otherwise. In any case, as noted by Caldarelli et al. 2012 [6] and Mealy et al. 2019 [5], these motivating heuristic arguments lead to an eigenvalue problem where the ECIs and PCIs are elements of eigenvectors. (The arguments by themselves still do not pin down the ECIs and PCIs uniquely, and to do so it is additionally argued – confusingly, in our view – that we should focus on the 2nd eigenvector from the eigenvalue problem, because it is “more informative” than the first. Why one should rule out the 3rd, 4th, etc. eigenvectors is never made clear.) Regardless of the problems these arguments

have, the intent was not the same as ours – they were not trying to work out direct evolutionary implications of the PoR.

HH 2009 proposed the ECI as a way to extract information about capabilities implied by the structure of a bipartite network of countries and products. As constructed, the arguments would make the methodology applicable to bipartite networks describing any set of places, such as cities or regions, and any set of economic activities, such as occupations of employment or technological areas of patenting. The arguments do not hinge on there being a particular process of evolution, with places evolving in time in a particular way like the PoR. Similarly, follow-on works [3,6-8] have not used the PoR to obtain complexity metrics either. Country Fitness, proposed by Tacchella et al (2012) [3], uses a mirror image of the original argument, making a small change in one of the equations above. (This work argued that one of the arithmetic averages should instead be a harmonic average.) In either the ECI or country Fitness proposals, these authors never used the PoR as a generative model that gets you to a metric. The hope has been that one could obtain an inference measure that correlates with the hidden capabilities we don't directly observe, and the resulting quantities have been consistently framed this way.

Our starting point is fundamentally different from HH 2009, Refs [3,6-8], or any prior complexity metric proposals for that matter. In fact, we actually didn't set out to understand complexity metrics at all. Our explorations began by asking ourselves how one could quantify structural change. Could one summarize the "distance traversed" by a country as a number? Our approach started with the PoR, because the PoR contains ideas for mapping out which industries are close to each other and which are far away. We asked ourselves what the long-term behavior is of a country that obeys the PoR, using this expected behavior as the "ruler" by which to decide what kinds of structural change are "large". To our knowledge, no papers in economic complexity have asked what the long-term expectations of the PoR are. Notably, this question has a concrete answer regardless of whether "complexity" is an actual property of economies to be quantified. Our analytical approach to this question began with the regression models used to study the PoR, viewing them instead as dynamical models. This is also something we have not seen done elsewhere, at least not explicitly; people just do not think of the PoR in such terms. (Indeed if they had, other aspects of the analysis would be done differently, e.g. one would not binarize RCAs, or add regression control variables). This step was crucial because you can then rely on standard methodologies. Having written the PoR as a dynamical model, we apply the old technique of eigenmode analysis. This is a general way to understand the behavior of a dynamical model (with of course nothing to do with economic complexity). Our question was simple – what is the long-run behavior of a country that obeys the Principle of Relatedness?

Because the PoR is a linear dynamical model, one can answer this with techniques undergraduates learn as a homework problem. We show that under the PoR, the two longest time scales of country evolution are marked by changes in the *diversity* of industries a country has and changes in their *composition* – roughly, a general movement out of agriculture industries and into manufacturing. The results are remarkable for two reasons:

- It mathematically connects research in the new field of economic complexity on the PoR – which has described countries’ industrial changes on year-to-year time scales and at a highly granular level of hundreds of industries – with classic work in development economics, where (at a coarse-grained level) these diversity and composition changes have long been understood as canonical outcomes of development. We show that the short-term changes of the PoR actually *imply* the long-term canonical development path of increased diversity and a shift from agriculture to manufacturing. The fine-grained work in economic complexity seems to have never asked whether its findings were consistent with these coarse descriptions from classic development economics.
- The PoR naturally leads to two coordinate measures that describe these canonical changes, and *very closely resemble the two main competing complexity metrics* – even though we were not trying to create such measures, and made no use of the conceptual narratives about country capabilities that originally motivated these metrics.

The first point means that classic macro trends follow precisely from short-term dynamics. Many studies of development describe transition across broad sectors, i.e. agriculture to manufacturing to services. But the PoR describes industry change in terms of hundreds of sectors. One needs the micro-level product transition matrix used in the new field of economic complexity to describe these dynamics, something prior economic studies had no access to, adding nuance to this coarsened view of the world. There is nothing that would have guaranteed that applying eigenmode decomposition to the detailed descriptions of industry used with the PoR would result in something so consistent with prior work. It is great confirmation of a consistency between new science and old. We also think this makes the modeling framework here an appealing starting point for future investigations that e.g. enrich the simple dynamical model we study, which only attempts to capture the Principle of Relatedness and no other mechanisms of country evolution. (Obviously real country evolution is much more complex than this very useful but quite stylized principle.) It is also nice to extract specific and practical measures (our A and b coordinates) that summarize structural change quantitatively (our original and much more narrow goal).

The second point contains two surprises for users of complexity metrics. The main complexity metrics are the ECI and country Fitness, and these have generally been understood to be competing inference methods for gleaning information about a country’s capabilities. Yet our paper suggests:

- (i) They are not in competition. They provide complementary information about industrial baskets, each one being incomplete without the other.
- (ii) *Neither* necessarily infers information about country capabilities. Instead, each may just summarize surface-level changes in the diversity or composition of export baskets.

These quantities that closely resemble complexity metrics just emerge naturally from considering the long-run implications of the PoR. Moreover, it isn’t just one metric (such as the

ECI) that emerges – both major families of metrics do. They come out as numbers that describe how far a country has progressed in executing two slow, broad patterns of shifts across industries. They are *descriptions* of how far a country has progressed in these directions, akin to how the coordinates of a principal component analysis describe the main axes of data variation. (Here, the axes are about temporal changes rather than data variation per se.) Under our derivation, these metrics *are not inference methods that inform us about hidden productive capabilities* – “capabilities” are nowhere in our math. Ironically, our derivation of complexity metrics is nothing more than an implication of the PoR (i.e. a tenet of the field of economic complexity). Our derivation uses standard techniques, in contrast to the original arguments that have caused 14 years of debate.

Other considerations support the idea that complexity metrics just describe what happens to a country in the long run, rather than inferring complexity per se. Some products have always had “complexity” values that did not quite make sense. (Like the examples of oil production and works of art we discuss in the paper.)

These examples are more common than acknowledged, but occur in data that often goes unpublished. As members of a group that champions the ECI, we are often asked about its behavior in various datasets. The ECI/PCI methodology has long seemed to “work” in international trade data. That is, if you compute the ECIs of countries, it seems to rank them plausibly, and if you compute the PCIs of exported products, it also seems to (mostly) rank them plausibly – products that we intuitively tend to regard as complex (e.g. advanced manufactured goods) come out high, while intuitively non-complex products (e.g. cattle-raising) come out low.

However, when taking this methodology to other data, people often scratch their heads. Usually, the trouble is with the PCI portion of the methodology. While the ECIs still rank places plausibly, and have the expected correlation with GDP, the PCIs rank industries in ways that people find puzzling. They often assign high complexity values to seemingly low-complexity activities (e.g. home nannies, dog walkers). This outcome makes little sense if a PCI quantifies a product’s complexity. On the other hand, it makes quite a lot of sense if this vector instead quantifies what activities a place performs in the long run (whether complex or not):

- Home servants and dog-walkers are assigned positive values in our directional vector $\mathbf{v}_2$ (which would ordinarily be interpreted as the vector of “PCI”s) simply because developed economies acquire these activities later on as they develop, and become wealthy enough to harbor these industries. The PCI is not picking out the complex industries; it’s picking out the ones that come later, whether complex or not.
- The ECI still behaves as expected, correlating with GDP, because it is only wealthy societies that can afford to have home servants and dog-walking as industries.

Our different way of obtaining the ECI and PCI provides an explanation for these odd cases.

This summarizes what we do and how it is unlike prior work (and why it matters). As the Reviewer notes there is a close mathematical resemblance of our metrics to the ECI and PCI. We described this resemblance in detail in the Supplementary Material and the main text to make the point that a fundamentally different mathematical foundation (i.e. without invoking the concept of “capabilities”) can generate quantities that closely resemble “complexity metrics”. As a broad observation, some of the confusion around complexity metrics might derive from the fact that the first complexity metric, the ECI, is a solution to a rather general-looking eigenvector problem. A consequence is that there may be a variety of ways to conceptually motivate this eigenvector problem or similar ones, and arrive at quantities resembling complexity metrics, some not requiring the concept of capabilities.

The Reviewer found that the paper, in tying together two branches of work in economic complexity, felt a little pretentious. We can offer a guess as to why. “Unifying” two branches of research could seem to satisfy intellectual curiosity, but does not necessarily represent an important achievement in itself. It is possible we over-emphasized the value of unifying things unto itself.

To address these issues and alleviate confusion that our previous exposition brought up about why our results are significant, we have tried to bring in more of the context and logic outlined above. We substantially rewrote the Introduction and revised the Discussion, to better emphasize a number of points: (1) that the capabilities narrative is a “narrative”, and that complexity metrics are intended as data inference methods under the concepts of that narrative; (2) that the mathematical linkage between the metrics and the narrative is not tight, as it currently stands; (3) that while the PoR and complexity metrics represent new approaches to understanding structural change, it’s not clear how they relate to the classic understanding we already have; (4) that our work sets up such a relationship, in the process adding granularity and nuance to classic observations; (5) that our results point toward a different interpretation of complexity metrics in which they simply describe overt changes in activities that many economies undergo, rather than infer something deep about capabilities; (6) and that such changes in how we view complexity metrics could have a bearing on how they are used in policy settings.

At page three, the Authors claim surprising the fact that currently in use metrics of economic complexity either relate to the diversity or composition of the economy. Is it though, surprising? This result has been repeatedly shown over and over, with examples in Mealy et al., Science Advances, 2019; Mariani et al., The Eur. Phys. Jour.B, 2015; Sciarra et al., Nat. Comms., 2020, and many others – as many referring works this manuscript has, especially considering the correlation with the diversity. Moreover, the GENEPI metrics with its two component has introduced the idea that the two classical metrics of economic complexity either measure diversity or composition of the export baskets, and the composition of the two leads to the multi-dimensional metric of complexity.²

The reviewer is correct that the coincidence between Fitness and diversity is no surprising new finding. However, the point we wanted to make here is a different one: What is surprising is the correspondence of complexity metrics with how economists have long described structural change and economic development. Again we need to back up and summarize what the field has been saying:

- In the usual narrative, everyone wants their favored complexity metric to be positively correlated with countries' *diversities*. The rationale originates from HH 2009: More diverse countries should generally be more complex ones. So, as a very basic smell test, it has become common to check that a given complexity metric correlates with diversity. HH 2009 argued that the correlation of the ECI with diversity meant it passes a basic check. (Probably giving this weak test too much credence, in our view.) The creators of country Fitness made the same point about their metric, and strengthened the point – they touted their *higher* correlation with diversity as evidence that country Fitness was a *better* metric. As there is nowhere to go to find the objectively “correct” complexity ranking of countries, the field has used tests like this to decide when a metric looks better. (Incidentally, both groups appear to agree that diversity itself should not be used as a measure of country complexity. This implies that one doesn't want a complexity metric to have *perfect* [i.e. $\rho=1$] correlation with diversity. However, it has never been made clear how high the correlation “ought” to be and what would be too high.)
- The question of how complexity metrics relate to the *composition* of an export basket was not a topic of discussion in the way we're bringing it up. For example, no one made a particular point to say that the signage pattern of the PCIs matches the classic agriculture-manufacturing trajectory. No one *would* make such a point – the interesting thing about the ECI (which equals the average PCI of a country) is supposed to be that it conveys something deep and hard-to-observe (complexity), not something surface-level, such as how much an export basket emphasizes agriculture versus manufacturing. To be sure, everyone who glances at the values of PCIs does notice that agricultural products tend to have low values and manufacturing products high values, but they interpret this within the usual narrative as a statement that, evidently, agriculture products are often low in complexity and manufacturing products high. (This mindset is also expressed in the way the PCIs are re-centered after their “raw” values are computed. Since the PCIs are regarded as just having a complexity meaning relative to one another, they can be conveniently re-centered to have zero mean without harm. This changes the sign of a few products but doesn't change the relative complexity estimates of countries. If instead you wanted to use the PCIs to weight composition it would be odd to re-center them, changing their signs and magnitudes. In our theory, these signs are not just toss-away information but are meaningful – products with minus signs are ones that become relatively less important in the output basket over time and ones with plus signs become more important.)

The papers mentioned above are not discussing the point about composition or a connection of the ECI with the classic agriculture-to-manufacturing transition, but rather are expanding on the patterns of thought above that are driven by the usual complexity narrative. Two of the three papers (Mealy et al. and Mariani et al.) include discussions of the relationship of complexity metrics with diversity under the usual goal of asking whether metrics pass this basic test thought to be a requirement for a “good” complexity metric. Mariani et al. 2015 [5] evaluates the two main competing methodologies (the ECI/PCI and country Fitness / product Quality) based on how well they rank countries and products. Since the true ranking is unknown, they use smell tests to decide which methodology’s rankings appear more plausible. One such test is how well each country metric correlates with diversity. *Both* metrics are correlated with diversity, with Country Fitness’s correlation being higher. The main result of Mealy et al. 2019 is that the Method of Reflections is equivalent to spectral data clustering. The Method of Reflections arrays countries on a 1D axis such that countries with similar export baskets have similar ECIs. As part of their exploration into the interpretation of the ECI, they study the relationship of the ECI to diversity. They find the ECI is positively correlated with diversity in country export data, but fails to exhibit this property in two regional data sets (their Figure 4). (An observation that works just fine for our results – we don’t describe the ECI foremost as a quantity that tracks diversity, but as one that tracks structural transformation.)

Sciarra et al. is in a different category since its goal is to propose a new metric rather than evaluate existing ones. They proposed the GENEPI metric, which is a composite of two components ($X_{c,1}$ and $X_{c,2}$). However, numerically these components are less related to diversity and composition change than our A and b coordinates, and conceptually Sciarra et al. were not attempting to associate $X_{c,1}$ and $X_{c,2}$ with diversity and composition change. The connection they are interested in is to country Fitness and the ECI. Each of these is taken at face value as a complexity metric. The paper establishes mathematical formulas relating (1) $X_{c,1}$ to country Fitness, and (2) $X_{c,2}$ to the ECI. Their proposal is that *both* the ECI and country Fitness have value – as inference measures of complexity. They do not question that these metrics are inference methods that measure the underlying complexity of places. They assume both metrics have some ability to infer complexity but that each does so imperfectly. They propose that a particular mixture of these metrics (GENEPI) will give better complexity measurements than either metric alone. Conceptually, the 2D plane discussed in this work, using $X_{c,1}$ and $X_{c,2}$ as coordinates, is not explained or implied as describing diversity and compositional variation. Numerically, $X_{c,1}$ is less correlated with diversity than country Fitness is, and $X_{c,2}$ is less correlated with the ECI’s agriculture-manufacturing compositional split than the ECI is. One way to see this is that the mathematical transformations (shown in their paper or our SI section S4) taking $X_{c,1}$ to country Fitness and $X_{c,2}$ to the ECI distort these variables, making $X_{c,1}$ and $X_{c,2}$ themselves less quantitatively related to diversity and composition.

In contrast, the “surprise” we point out at page 3 takes the conversation outside these discussions driven by the capabilities narrative. We ask, what if we just think of these metrics as offering a surface-level characterization (as opposed to a shift in deeper, unobserved capabilities) of the export basket? It is surprising that complexity metrics, touted as quantities

that accomplish a deep inference task about capabilities, look just like they track two classic patterns of development that we have known about for many years.

Yet, the addition of the metrics the Authors introduce is new and marks my following point. The authors make clear that contrarily to other research works their aim is to introduce a dynamical modelling framework from which to extract the long-term implications of short-term dynamics keeping distance from the definition of new complexity metrics. Still, in introducing the A and b components, they target the comparison with other economic complexity metrics, as the ECI, Fitness and GENEPI, even defining the variables with resounding names such as ECI, for example. Beyond the different mathematical frameworks, the convergency of the model's components follows that of the iterative Method of Reflection in Hidalgo-Hausmann PNAS, 2014, with the first eigenvector resulting to be uniform and the second component taken for more informative results. Again, what can we learn anew with respect to previous contributions? I am afraid of the risk of falling within the circular reasoning of "729 new measures of economic complexity" - <https://arxiv.org/ftp/arxiv/papers/1708/1708.04107.pdf>.*

We agree with the reviewer that it would not be particularly groundbreaking to add another, marginally different complexity metric to the existing set and the labeling of our variable may indeed risk conveying this impression. Here, it is important to note that one important audience is researchers from economic complexity. This field has pursued a valuable approach to structural change that studies the process at a granular level involving hundreds of industries, on short year-to-year time scales, and an important output of this work has been the PoR, the stylized observation that became the starting point for our investigations. However, this is the same field that also produces and uses complexity metrics. The set of researchers who study the PoR is nearly identical to that familiar with complexity metrics. Having derived implications of the PoR and obtained the resulting A and b coordinates, it would be apparent to these researchers that the structural change coordinates we derive look like complexity metrics in form and numerical behavior. This alignment demands interpretation – what should one think of this? Did the authors just reinvent the wheel? Did they find a new route to the same quantities? Is this new route just a refinement of some kind, or is it importantly different? (As we emphasized it's the latter: We derived our coordinates with no use at all of the motivating concept of capabilities.) These debates are important and technical, and so we gave our variables names that would help researchers compare our math with prior math.

This comparison is meant to facilitate the points we are making about complexity metrics. We learn from these results that the field's perspective on these metrics may be fundamentally misguided. The leading complexity metrics are complementary quantities; each is incomplete without the other. Moreover, neither may be measuring complexity at all. Rather, they may just be tracking diversity and compositional changes that (i) are direct implications of the PoR (a major tenet of economic complexity itself) and (ii) are classically understood as aspects of a canonical development path. (Which, in turn, would help explain desirable properties,

especially their correlation with GDP, as well as the anomalous complexity values of some industries.)

The PoR is a simple and rich model of structural change, and so in studying its implications we could have directed our attention in a number of directions. But we focused on comparing the structural change coordinates (A and b) that emerge from the PoR with complexity metrics for two reasons. First, complexity metrics have become an influential tool of policy. In a short period, these metrics and the conceptual narrative they're paired with have had real impact on policy in many countries and international organizations. But there is confusion and debate over how these metrics work. Our results give a different explanation for this, while also

- making sense of their anomalous values,
- explaining why we have two families of metrics,
- explaining why both would be correlated with GDP,
- showing how these metrics relate to and align with classic facts about structural change.

Our results accomplish all this with much more conventional arguments than the original ones that have created years of debate.

We were not certain what is meant by “the convergency of the model’s components follows that of the iterative Method of Reflection in Hidalgo-Hausmann PNAS, 2014, with the first eigenvector resulting to be uniform and the second component taken for more informative results.” To be sure we do not use math that is the same or equivalent to the Method of Reflections. We also do not take only the second eigenvector. (And we don’t say we do so simply for more informative results. Such an argument, though, is made to reach the *conventional* ECI; e.g. Hausmann, Ricardo, et al. *The Atlas of Economic Complexity: Mapping paths to prosperity*. MIT Press, 2014. We look at both the first and second eigenvectors and do so because eigenmode analysis assigns meaning to both.)

We were also not sure what is meant by the circular reasoning of the referenced paper. There is a circularity (by design) in the way the two main competing complexity metrics are derived, in that these metrics use a system of coupled averages as a framework for developing complexity metrics. We actually don’t take issue with such circularity per se, but to better explain the differences between our mathematics and prior math, we did note that our approach doesn’t have this circularity; deriving our coordinates comes from a once-through calculation that (1) begins from the definition of the network spreading process, (2) derives directions of change, and (3) derives coordinates of change for these directions (Fig 5).

I have a point about discussing structural change with respect to the presented framework. This point is not a critique moved for the Authors their-selves, because I can see it paves the way for a larger and more philosophical debate exiting the aim of this work. Yet I’d like to raise it here as the Authors can discuss through since “it is not an intrinsic requirement of our model” – page 5. The economic complexity field has been marked by the presence – in its methods – of the Revealed Comparative Advantage. The role of the RCA is two-fold: it allows us to remove

intrinsic noise related to the presence of outlying trade operations and to reveal the significance of the exports. There's plenty of literature, e.g., as Authored by Giovanni Dosi to cite one, discussing about how this procedure biases the results. My comment follows the same argument. Can we actually speak about long-term structural changes when the RCA erases important information? Note that the following reasoning works even considering the log transformation the Authors introduce, because the critical issue here stands in the RCA threshold ($RCA=1$), the authors still use it to map the values onto the log scale. Here's an example to follow. Let us analyze the following trade matrix among the countries c_1, c_2, c_3, c_4 of products p_1, p_2, p_3 . The trade flows range among three different orders of magnitude.

case (a)

trade p_1 p_2 p_3

c_1 100 10 0

c_2 100 10 0

c_3 10 0 1

c_4 0 0 10

with RCA values

RCA p_1 p_2 p_3

c_1 1.043 1.095 0.000

c_2 1.043 1.095 0.000

c_3 1.043 0.000 1.992

c_4 0.000 0.000 21.909

Let us just play with the component c_2 p_3 as in the example below.

case (b)

trade p_1 p_2 p_3

c_1 100 10 0

c_2 100 10 10

c_3 10 0 1

c_4 0 0 10

with RCA values

RCA p_1 p_2 p_3

c_1 1.087 1.141 0.000

c_2 0.996 1.046 0.996

c_3 1.087 0.000 1.087

c_4 0.000 0.000 11.952

As shown, when changing the element from $c_{2p_3}=0$ – case (a) – to $c_{2p_3}=10$, case (b): In case b, the RCA value of c_{2p_1} changes from being greater than one, to be smaller than one, being one the threshold of competitive advantage. Thus, even though the country c_2 has not changed its performance in that particular product, it may result in some kind of structural change when analyzing the binary trade matrix.

The value c_{2p_3} increases by one order of magnitude – possibly implying that there has been a sort of structural change in the economy – which we are though not able to capture due to the RCA threshold impeding us from catching that 0.996 value.

So, my question to the Authors is: if the RCA threshold is not an intrinsic requirement of the model, and considering that the framework you present is claimed to be novel, can you show the results of your framework on long-term and short-term dynamics without considering the RCA?

We thank the Reviewer for raising this topic. We fully agree the choice of RCA as a measure of “ability”, while standard, comes with side effects. We added a new discussion about this in the SI (Supplementary Note S1) and summarize its points here.

Some of these side effects are results of using RCAs at all and some arise from the additional step of binarizing / thresholding. The latter in particular can produce exactly the kinds of counter-intuitive effects that the example above demonstrates, where a smooth change in RCAs produces non-smooth outcomes in the thresholded RCAs, and (most confusingly) in which a country’s performance in one product becomes regarded as drastically lower just because it improved its performance in another product.

First, focusing on our own results, in the places where thresholding was done at all it does not impact our findings. There are two places in our calculations where thresholding of RCAs could in principle come up. The dynamical Eq. 2 is

$$\dot{\vec{R}}(t) = \gamma(t) \vec{R}(t) - \frac{1}{\tau} L_{\Phi} \vec{R}(t),$$

where export data is used to construct both the proximity matrix Φ (and by extension the Laplacian matrix L_{Φ}) and the ability state vector $\vec{R}(t)$. In principle, either calculation could involve thresholding RCAs. For the first calculation, we tried three methods used in the literature (described in the text around Figure 7). Two of these (Φ^P and Φ^M) use thresholded RCAs (resulting in the “M” matrix) while the third (Φ^C) does not. As we show in the text, we don’t find large differences in our results whichever matrix is used. (For an extensive empirical investigation of different specifications of the proximity matrix, see Li & Neffke 2024 [9].) Thus, thresholding by itself need not have much impact through its effects on the proximity matrix. The second place where thresholding could in principle be done is constructing the state vector

$\vec{R}$. While this is often done in the literature, we never do it here. The model we study is a continuous time model, with elements of $\vec{R}$ that evolve continuously. We think it is a useful feature of our modeling approach that it expresses evolution in ability over a continuum of values.

As the Reviewer notes, to construct the elements of $\vec{R}$ we take raw RCAs and re-map them with the linear-to-log transformation $g(R) \sim \log(1 + R'/R_0)$ described in the text. Unlike binarizing, this is an invertible transformation with no information loss; one could fully recover the raw RCA values from the transformed ones. In this transformation, the R_0 parameter provides an RCA benchmark akin to the threshold RCA when binarizing (without the loss-of-information that occurs when binarizing). R_0 is approximately the RCA scale at which the transformation shifts from linear to log behavior. It has the value noted in our Methods of $R_0 = 0.115$, which was obtained through the procedure described in the SI that maximizes the descriptive power of the 2nd eigenvector. (Thus, the most rapid gains in transformed RCAs occur when the raw RCA is between 0 and 0.115, with a tapering off/diminishing returns above this value.) We note it is an open question at what level an RCA should be regarded as an indicator of “significant” ability in a product.

Beyond the questions that binarization raises, there are more general questions about the use of RCAs at all. The Reviewer’s example shows there are still counter-intuitive effects (albeit smaller in magnitude) where an improvement in ability in one product results in a measurement of dis-improvement in ability in another product. In our new SI section we make a few points here. For exports (as well as other measures of country activity) there are two sources of variation in output that we would typically like to remove: variation because some economies are bigger than others (in GDP or population say), and variation because some products are naturally produced in larger volumes than others. Removing these sources of variation feels like a natural step toward a measure of a place’s ability in a product. The RCAs are one way to do this and follow a simple idea of two normalizations – compute activity shares (i.e. normalize by a country’s total output across all products) and divide by the average share for each product (i.e. normalize by the average such share across all countries). This approach is sufficiently intuitive that it has been re-invented or newly-motivated in several places. (The same quantity is called the Location Quotient within geography, and has been recently motivated on information-theoretic grounds as an instance of pointwise-mutual information.) However, other approaches have also been used to normalize exports data. Revealed per Capita Advantage (RpCA) replaces the country’s total output with its population to factor out country size variation. One could also use a full regression model to form the expected value of a place’s output in a product, as done in Brummitt et al. [10]. To avoid sprawling the scope of the paper, we don’t extend our robustness beyond what we describe, but we now call out these issues and summarize the choices in the SI section.

Since our goals are to develop implications of the choices that the field is currently making, we deliberately took many standard choices “off the shelf” to carry out our calculations (such as using conventional RCAs). However, we hope the new discussion helps elevate to the surface some choices that researchers have about how ability in products gets quantified.

Reviewer #1 on behalf of Reviewer #3

All points raised by the Reviewer have been properly addressed and answered.

We thank Reviewer 1 for checking the manuscript with respect to Reviewer 3’s comments as well.

[1] Hidalgo, C. A., Klinger, B., Barabási, A. L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science*, 317(5837), 482-487.

[2] Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106(26), 10570-10575.

[3] Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., & Pietronero, L. (2012). A new metrics for countries' fitness and products' complexity. *Scientific Reports*, 2(1), 723.

[4] Mealy, P., Farmer, J. D., & Teytelboym, A. (2019). Interpreting economic complexity. *Science Advances*, 5(1), eaau1705.

[5] Mariani, M. S., Vidmer, A., Medo, M., & Zhang, Y. C. (2015). Measuring economic complexity of countries and products: which metric to use?. *The European Physical Journal B*, 88, 1-9.

[6] Caldarelli G, Cristelli M, Gabrielli A, Pietronero L, Scala A, Tacchella A (2012). A Network Analysis of Countries' Export Flows: Firm Grounds for the Building Blocks of the Economy. *PLoS ONE* 7(10): e47278.

[7] Schetter, U. (2019). A structural ranking of economic complexity. *CID Research Fellow & Graduate Student Working Paper*, (119).

[8] Sciarra, C., Chiarotti, G., Ridolfi, L., & Laio, F. (2020). Reconciling contrasting views on economic complexity. *Nature communications*, 11(1), 3352.

[9] Li, Y. & Neffke, F. M. (2024). Evaluating the principle of relatedness: Estimation, drivers and implications for policy. *Research Policy* 104952.

[10] Brummitt, C. D., Gómez-Liévano, A., Hausmann, R., & Bonds, M. H. (2020). Machine-learned patterns suggest that diversification drives economic development. *Journal of the Royal Society Interface*, 17(162), 20190283.

Reviewer #2 (Remarks to the Author):

I'm thrilled to see substantial improvements in the paper, and I'm pleased to note that my two primary concerns regarding the paper's contribution and the readership of Nature Communication have been largely addressed. The new introduction is way better than the previous one.

We thank the Reviewer for their comments and greatly appreciate their careful review. We are very happy the introduction is reading better and that the primary concerns have been largely addressed.

However, there are still some concerns that remain, listed below in order of importance. Besides the comments below, the paper's logics and methodologies seem sound and warrant publication.

1. Modeling Framework in Post-hoc Patterns: The paper asserts that it uncovers long-term and short-term dynamics by decomposing the bipartite network R_{cp} . Therefore, the derived long-term and short-term dynamics are sort of baked in the construction of R_{cp} , with the frontier of the R_{cp} network structure (associated with a large 'A' axis in Fig. 3) relying mainly on the economic activities of developed countries like the USA and Germany. The rest are catching up on the 'A' axis with diffusions on the 'b' axis in Fig. 3. This is what I refer to as post-hoc patterns. It doesn't imply that developed countries haven't slowed down, recessing, or wiggling in the frontiers, as evidenced by small, curled trajectories in region 2. The observation here is that underdeveloped/developing countries and un-usual countries (like tax-haven) have been catching up with developed countries in their product compositions over time (Region 1 and 3). Does this mean there is only one long-term progression of economic development, mostly set by developed countries? Is this the only path to development enough to be called a fundamental mechanism? Is there room for "Leapfrogging dynamics" in the framework? I would caution against labeling this as 'a modeling framework,' as it might mislead others to accept the post-hoc patterns without question.

We thank the reviewer for raising these points. We used them to elaborate on the relationship between short- and long-term in the paper, the sense in which our approach is a “modeling framework”, and the implications for “canonical development pathways” and “leapfrogging”.

First, it's important to note how we get the short- and long-term dynamics. The short-term dynamics are essentially taken from the literature. A body of literature reports that countries, at some level of fidelity, obey a stylized fact that has come to be known as the Principle of Relatedness (PoR). One of our initial points is that the PoR represents a description of what countries do year-to-year, i.e. it's a short-term description of country dynamics. Obviously, the PoR is an extremely simplified description of how countries evolve. But it is a widely

corroborated principle, and usefully, it is a straightforward exercise of dynamical systems theory to work out its long-run implications (i.e. the long-run behavior of a country that obeys the PoR). While straightforward, this appears never to have been done before.

It is also important to note that, when we numerically compute the mathematical quantities that emerge from the theory (i.e. to describe a country's long-term behavior), we are using information from just one year of data alone. Below, the Reviewer suggested splitting the data into an earlier period and a later period, using the data in the earlier period to predict directions of change in the later period. The Reviewer suggests splitting across the year 2000, but we did something far more aggressive – we use data in just the single year 1962 (i.e. the first available year in our dataset). The matrix $R_{cp}(t)$ matrix varies by year t and captures the full history of each country's intensity in each product. However, we use only the data in $R_{cp}(1962)$ to construct the matrix between products $\tilde{\Phi}_{pp}$, whose eigenvectors give the two directional axes of long-term change on which we plot the evolution of countries. Thus, in later years t , we ask where countries lie on axes that are defined by 1962 data alone. We constructed the proximity matrix between products from data in just one year because this is typical of empirical studies of the PoR, and it is of course interesting to see that data from the past can say something about the future. As an aside, it is plausible for directions of long-term change to be extracted from just one year of data because the cross-section of countries spans ones of very different levels of development (in a sense spanning a large amount of development time), and the network constructed from their exported baskets in this year gives an axis of change that also appears to describe many countries longitudinally.

We now turn to the questions of whether our results imply only one long-term progression of development and whether leapfrogging is possible in our modeling framework. Whether the specific model Eq. 2 allows leapfrogging depends on how leapfrogging is defined. If leapfrogging means jumps in GDP per capita, then this model does allow for this. The relationship in Figure 3a (between the ECI* and GDP per capita) is an empirical one, i.e. it is not something that emerges from the model, which makes no attempt to internalize the complex relationship between GDP and industry portfolio. It is entirely possible that, in the future, this relationship would change (i.e. growth in wealth would no longer be associated with this direction of movement but others). On the other hand, if leapfrogging is defined by jumps into industries that are unrelated to ones a country currently does, then this is explicitly not allowed by Eq. 2 – this equation only captures the dynamics of the PoR, which (by definition) is a principle of spreading locally within the network of industries.

Saying that countries only change according to the PoR is a highly simplifying assumption. (Which is why it was so surprising to us how much this assumption gets right about long-term evolution.) But this is a restriction of the PoR, not of the modeling framework, where the remedy is the usual one – add terms to Eq. 2 to account for other processes of change (in particular in this case, ones that would enable long-distance jumps). Then one would follow the same analytical recipe as before: Identify the dynamical modes / expected long-run behavior of the resulting differential equation. This is why we call this a 'framework', i.e. one can modify the

model while carrying out the same steps of finding the most prominent dynamical modes and using them to summarize long-run evolution in low-order terms. Changing the model will modify the long-term predictions, along with the preferred coordinate system to summarize country changes.

Nevertheless, this potential for elaborating on Eq. 2 does not change the fact that when we make the very simple assumption that countries obey only the PoR, the implied long-run evolution is the “canonical” development pathway. It is interesting that this pathway emerges out of the PoR alone. We now comment on the points above in a paragraph in Section 2C. We also briefly revisit them in the Discussion where we reiterate that this very simple assumption yields the canonical pathway, but is clearly incomplete, and that more complete models could be studied in the same way.

2. Beyond Post-hoc Analysis: To extend the analysis, one could examine the long-term and short-term dynamics from R_{cp} in the past (1962-1999) and compare them to the dynamics between 2000 and 2018. How much do the long-term and short-term dynamics hold for extrapolation analyses? In this way, we can say the findings are beyond post-hoc patterns, and there may be something more fundamental to the backward projection. A simple comparison between R_{cp} (<2000) with R_{cp} (≥ 2000) could provide valuable insights.

We thank the Reviewer for this suggestion. As we noted above we actually did an even more drastic data split because we only used the first year of the data to identify directions of long-term change. Like other studies, the proximity matrix we use to describe the network of products comes from data from just one year. (In our case 1962, the first available year in our dataset.) Thus, our analysis sets up an expectation about country evolution decades into the future on the basis of information contained in just the first year of data alone. The $R_{cp}(t)$ matrix is computable for each year of data t , and building the proximity matrix Φ_{pp} , roughly means multiplying the matrix $R(1962)$ with its transpose. (Several functional forms have been used as we explored as a robustness check.) Whichever particular formula is used, the proximity matrix is then row-normalized to give $\tilde{\Phi}_{pp}$ (the matrix in Eq. 2) whose first two eigenvectors give the two longest-term directions of country change. That these directions are informative of long-term changes is a remarkable demonstration that the PoR not only is useful for describing short-term change, but also contains information about long-term country evolution.

3. Related to the retrospect projection, the categorization problem should be addressed: The unit of analysis is projected from the SITC classification to analyze almost 60 years, during which many classifications have shifted, changed, and restructured due to technological innovation. Have any data attrition issues been overlooked in dealing with this extended longitudinal evolution? How does projecting current dimensions onto previous activities affect the findings? Could this explain why we see some non-trivial trajectory curls in Region 2 where the developed

countries (contributing the frontiers of the network structure) are located? Will these findings (curls in Region 2 in Fig. 3) still hold 30 years later with improved categorization systems? What are the chances the curls in Region 2 are methodological artifacts (using not-updated categories)? In other words, would these findings (curls in Region 2) still hold in 30 years later when we have updated categorization systems that better capture economic activities/products than now?

We thank the Reviewer for raising these points. We strongly agree 60 years is a long time to maintain a single classification system in the data. We would not expect the details of these curls to survive changes in the classification system but in general we are not sure how strongly one should interpret small changes in the locations of countries on these plots, beyond the broad patterns that we point out. To get a feel for this issue, we carried out a new robustness check to see how country variation would look in these figures if we were using a different product classification scheme. This is shown in what is now Supplementary Note S3. While the new scheme is also fixed in time (so we are not directly addressing changes in classification through time), we think that replacing the full classification scheme helps to get at the issue of how our results are impacted by the classification system. In the new SI note we show the same figure panels using exports under the HS classification system (as compared to SITC used in the original Fig 3), for which we are able to look at the shorter period 1995 - 2018. (In the SITC data we are able to look at 1962 - 2018.) HS classifications give very similar results. Income increases mainly as a function of the b coordinate (holding A fixed) while diversity increases mainly as a function of the A coordinate (holding b fixed). We see a similar arrangement of different categories of economies and a similar concentration of export data in different parts of the plane. The movement of countries is also similar but noisier, with differences among countries in some parts of Region 1, where in the 1995 - 2018 time period a number of countries with A values between about 0.3 and 0.6 progressed toward lower values of A during these years. This is due to the difference in time period (SITC data covers 1962 - 2018 while HS covers 1995 - 2018) rather than the classification system. Movement patterns under the two classifications are very similar when confining the SITC data to 1995 - 2018. Note that while the A and b coordinates summarize classic patterns of structural change, in which countries following a canonical development path move toward higher income and product diversity, not every country followed this path, or did so in every time period. Evidently, movement consistent with the canonical path was realized by more countries in the full 1962 - 2018 period than in just 1995 - 2018.

4. Nuance in Analysis: The nuance above is subtle, and this nuance by no means implies that the current analysis is incorrect. On the contrary, I am encouraging the deep discussion above because the current analyses and findings can shed light on a long-overdue shortcoming in the field, which has been overly focused on the complexity index from trade datasets. I was excited to see the paper thoughtfully state the exact same concerns that I have had in the field. However, to prevent misdirection in future research, I encourage the paper carefully articulate

its scope and contributions to economic development theory.

We thank the Reviewer for the encouragement. Indeed these are subtle issues that economic complexity analyses often do not delve into. The field has not always been particularly careful at distinguishing what counts as an exercise in describing historical patterns in data versus predicting what will happen. These issues impinge on the question of when we are talking about a description of development (contingent on a particular time period and data type) and when we are talking about something that approaches a more general theory. Here we straddle both places. The empirical plots in Fig. 1 are a description, yet the long-term term behavior of the PoR (i.e. the coordinate axes that define these plots) are a theory outcome – they are mathematical implications of the PoR. (While the fundamental drivers of the PoR itself and its limits remain issues that are being explored by the field.)

5. Lastly, the paper could benefit from further tightening and trimming. For example, section D (on page 14) could be simplified by putting detailed analyses and expanding results, explanations and interpretation on pages 16 and 17. Another example would be that I find page 20 is the most exciting part of the paper, and I would like it to come slightly earlier than now. Also, for minor comments (related to the nuance point 4 above), I would suggest using formalizing/formalism rather than theoretical/theory throughout the main text whenever appropriate in order to avoid misguidance. For example, Table 1, Derived from a model -> derived from a factorized matrix formalism. However, this use of the term is the authors' choice and please ignore my suggestion if the authors believe theoretical/theory is more appropriate.

I look forward to seeing these concerns addressed to improve the paper.

We thank the Reviewer for the encouragement to tighten the paper. We searched for places throughout the paper to reduce verbiage. We also reviewed our use of “theory” and “model”. “Theory” was already nearly banished from the main text in the previous round. (Not counting generic phrases like “theory results”, “theoretical findings, etc.) One instance remained in the Methods, which we have now fixed. Many of these instances were replaced in the last round with “dynamical modeling framework”. In Table 1 we made the phrase “Derived from a model” a bit more specific. The full cell now reads “Derived from a dynamical model: Coordinates emerge from a dynamical system describing economic diversification.”

We furthermore thank the Reviewer for the ideas to simplify section D (“Comparing our structural change coordinates with complexity metrics”), which goes into detail about how the structural change coordinates numerically and theoretical resemble complexity metrics, the intuition underlying this, and the relationship to dimensionality reduction perspectives in recent literature. Please note that we don’t think of this as a complexity metrics paper per se, but a paper for a broader audience that is interested in ways to model and quantify economic structural change. However, it is interesting that our results potentially resolve heated discussions in the economic complexity field’s approach to these problems. Moreover, we

believe that our results are highly relevant to practitioners in this area. We therefore wanted to be sure that these readers are served properly. In particular, given how widely used complexity metrics are to advocate for how to prioritize industrial policy, we expect policy makers to be interested in the detailed discussion of the coordinates and the extent to which these coordinates resemble complexity metrics derived with the narrative of “capabilities” numerically and theoretically. To serve these different audiences, we believed the cleanest sequence was to (i) fully establish the long-run implications of the dynamical model; (ii) establish that the structural change coordinates that emerge from this model look like complexity metrics; (iii) interpret this resemblance.

Reviewer #4 (Remarks to the Author):

I was recently asked to provide additional review on this paper, after the first round of reviews was already completed. In important ways, I found the paper interesting and a contribution to our understanding of economic complexity and relatedness. In other ways, in particular in its representation of the existing literature and the interpretation of the results, I found the paper falls short in important ways. My assessment is that this paper would be of interest of those working in economic complexity, but that as is, the paper is not ready for publication, and would benefit from a thorough revision.

We greatly thank the Reviewer for providing an additional review and appreciate the time they have taken to give a thorough look at our paper.

In short, the paper uses a standard linear model of the form $(d/dt(\text{vector}) \sim \text{vector} + \text{matrix} * \text{vector})$ to draw some connections between measures of economic complexity and relatedness. The authors call this an economic model, but that is not the case, since it is a standard first order approximation in linear algebra, and not a model with rational agents, first and second order conditions, prices, etc.

We thank the Reviewer for this comment. We do not pose the model as a model of optimizing agent behavior. Some phrases we used are

- “a network model of an economy’s structural transformation”
- “models for economic evolution”
- “network models that operationalize this principle [the Principle of Relatedness]”
- “a dynamical model of economic diversification”

Most of all, we repeatedly called it a dynamical model (a standard way to describe a model of the form $d/dt(\text{vector}) = \dots$). We also emphasized that this dynamical model was based on the regression models widely used to capture the PoR (models many readers would be aware of and understand not to be economic models in the rational-agent sense). We found just one place where we use the phrase “economic model” (on page 21), which is also referred to by the Reviewer further below. To limit possible confusion, we changed that phrase.

It is also not a generative model like other models used in this literature (original Lego model of the 2009 PNAS paper (solved analytically in 2011) or the 2017 Fink et al. model in Nature Communications).

We’re not sure in what way our model is not a generative model. It certainly satisfies the usual notion (one could certainly create synthetic country data using this model).

Given the two models mentioned by the Reviewer, we wonder if the concern is that our model doesn't feature country "capabilities", like these other models do. But this is fundamental to the points we make. Using a model without country capabilities, we obtain quantities that closely resemble complexity metrics, but have a different interpretation.

The authors then use this linear approximation to rederive well known concepts in the economic complexity literature: a measure of diversity, a measure of economic composition (complexity), rediscovering the well-known fact that these measures of complexity are eigenvectors of proximity matrices.

To be sure, this differs from our own framing of the results. What we claim is roughly the opposite – we say our results tend to give a different understanding of concepts in the economic complexity literature. We did not set out to find complexity metrics; it's just that what we found looks like complexity metrics. We derived quantities that resemble complexity metrics with a derivation very different from and more conventional than existing ones. We argued this calls into question the usual interpretation of complexity metrics.

After that, they try to position their results in the existing literature in pages that largely misinterpret previous work. If the authors were to build on previous work, which they could because their results are very consistent with previous efforts, the paper would be a contribution. Doing so would require rewriting completely most non-technical parts of the paper (e.g. intro, interpretation, discussion). I believe the authors are capable of doing these changes, for which I recommend revising and resubmitting the paper.

We think of our paper as reinterpreting and not misinterpreting prior work. In particular, as far as complexity metrics are concerned, we want to make the point that these metrics need not infer information about the hidden productive capabilities of countries (even though the literature has this interpretation), but rather describe long-term changes that countries commonly undergo as they develop, which may or may not be driven by capabilities.

In addition to this general concern, I also have specific concerns that the authors could address in a revised version.

1. The authors provide a linear approximation model predicting that in the very long run, all countries will produce all products. This is nonsensical from an economic perspective (and comes from the model lacking any strategic competition or comparative advantage generating mechanism). Can the authors justify their approximation better? (e.g., as a first-order matrix Taylor expansion of another model), Or tone down their claims of using an economic model to finally put the literature on a mathematical foundation?

We agree that our model does not capture strategic interactions between economic agents. To be sure, we did not intend to pose the model as a model of optimizing agent behavior. We say we use a simple dynamical model based on the regression models used to capture the Principle of Relatedness. None of these models have the aforementioned features (e.g. strategic competition, or many other real-world processes). We point out that this dynamical model (indeed, despite its simplicity) connects

- (1) observations of short-term movement across industries (the PoR),
- (2) observations of long-term movement across industries (classic facts of development), and
- (3) the mathematical form (though not necessarily conceptual meaning) of complexity metrics.

To help readers understand the assumptions involved, we added statements to the effect that the PoR is assuredly not a complete description of short-term dynamics, which by extension means that the 1st and 2nd eigenvectors are just approximate descriptions of long-term country behavior. With more information on short-run dynamics, one could adjust the dynamical model, with correspondingly better long-run predictions. In future research, this would be an interesting direction to go in, and the dynamical modeling framework here is then helpful to think through such an endeavor. However, this doesn't change our finding that simply modeling short-run dynamics with the PoR alone leads to long-run behavior in agreement with stylized facts and coordinates of change that closely resemble complexity metrics.

2. Regarding the math, the work validates important previous work, adding to the robustness and validity of previous findings, but the authors do not describe it in such a way. Instead of identifying connections to previous work when interpreting their results, the authors provide mischaracterizations of previous approaches to differentiate their work more strongly. This is something that I would suggest the authors revise in a future version of their manuscript. A few examples are clear in Table 1

We regret our work is perceived this way, but also would raise objections. For reasons we describe after each example below, we don't think we are mischaracterizing previous work.

a. When the authors try to compare the "complexity inference view" with their "dynamics systems, view," the first entry says that the "complexity view" is restricted to solving an inference problem and that their dynamical systems view brings in the "novelty" of characterizing activity baskets. That is incorrect, as an important motivation in the early work on economic complexity (going back more than a decade) is the characterization of baskets of economic activities.

Although we cannot assess the original authors' implicit motivations, providing a direct characterization of a country's basket of activities itself surely isn't what made complexity metrics a major interest to researchers. The authors of country Fitness, for instance, do not claim that this metric's key feature is that it measures a country's industrial diversity. The

excitement around complexity metrics is that they let us infer hidden properties of economies that we don't directly see. This is described as information about the count or sophistication of a place's productive capabilities, understood within the conceptual framework of a tripartite network of countries, products, and capabilities. Alternatively, if complexity metrics were just a characterization of export baskets themselves, why would this methodology be preferable over, say, PCA? This is not explained anywhere.

b. Another example from the same table, which reads rather unkindly, is that in the complexity view the principle of relatedness play “no role in particular.” Yet, in a recent review of the field published in Nature Review Physics, this connection is clearly described as one of several known facts (Page 9: “metrics of complexity are connected to the idea of relatedness because they are derived from the spectrum of proximity matrices.”).

We regret that our writing caused offense, but are somewhat puzzled why this would be the case. In this cell of the table, our aim is to emphasize that the evolutionary Principle of Relatedness wasn't used to derive the mathematical form of complexity metrics. The quote from page 9 of the mentioned review paper underscores this point, noting that several complexity metrics can be derived from proximity matrices. But proximity between economic activities is defined in the literature in a way that requires no assumptions about how economies change with time (just being measured by observing product collocations). Similarly, the arguments that motivate the ECI, for example, depend just on statements about what relationship one should expect between a place or product's complexity, on the one hand, and the static structure of this proximity network of products on the other. The PoR is not being used to directly derive complexity metrics. Finally, whereas country complexity measures are derived from country spaces, product complexity measures are derived from product spaces. In contrast, in our derivations, we never construct a country space, and instead use the A and b coordinates to measure how far along the major axes of the *product space* a country has progressed. This also explains why the b coordinate (ECI*) is not identical to the ECI.

c. The same table also describes the derivation of complexity metrics as “motivated by a data challenge.” Again, reading the introduction of the original papers and books shows that this is clearly not the case. The original 2009 complexity paper starts with Adam Smith and the division of Labor. The 2014 book with MIT Press motivates complexity and relatedness by discussing the division of knowledge. Etc.

We are unsure what the Reviewer's concern is in this case. The full text of this cell of the table says that under the usual complexity inference view, the derivation of complexity metrics is “Motivated by a data challenge: Infer complexity from data about what goods are produced where.” This strikes us as an uncontroversial statement about the derivation of complexity metrics.

For sure, the context for wanting complexity metrics at all was a desire to understand mechanisms underlying economic development (such as how knowledge is divided across occupations and brought together, etc.) But these discussions do not constitute a derivation of the metrics themselves. To try to be a little clearer, we re-labeled the row “Mathematical derivation” instead of just “Derivation”.

Overall, the authors seem far from understanding the motivation and goals of the original work. They would benefit from either not saying them, or learning to articulate their contribution in a constructive manner with the literature (since none of what they are saying is novel in term of motivation and goals is novel for an author experienced in this line of work).

For reasons mentioned above and below we think we understand the motivations of the original work.

3. Throughout the paper, the authors also repeatedly mention that the quantities they derive have little to do with complexity (e.g. page 17, page 4, page 21), but do not justify that claim. What is the definition of complexity they have in mind? It must be different than that of previous authors working in economic complexity, as in previous work, complexity measures are estimates of economic structure and composition that are relevant to understand the process of economic development. The latter is the interpretation the authors give to their own metric (b), and that coincides with the interpretation of ECI. I don't understand the need to invent a difference where there is a coincidence.

We respectfully disagree with the reviewer on this point. Complexity metrics have been put forward as quantities that infer hidden information about the productive capabilities of countries (their economic complexity), but the coordinates we derive need not have such a meaning. Our coordinates can just quantify long-term industry shifts – period – without inferring anything about country “capabilities”. (Indeed our coordinates have a concrete interpretation whether or not “capabilities” and “complexity” are actually meaningful properties of countries to be quantified.)

Complexity metrics are not just any estimates that characterize the structure of economies relevant to development. Their stated purpose and main selling point has been that they can infer hidden information about the productive capabilities of an economy from observable data.

The quotes below exemplify this understanding of the goal of complexity metrics (bold emphasis ours):

- Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the national academy of sciences*, 106(26), 10570-10575.

*“...the question of economic complexity is equivalent to asking **whether we can infer properties such as the diversity and exclusivity of the Lego pieces inside a child’s bucket** by looking only at the models that a group of children, each with a different bucket of Legos, can make.”*

*“...we develop a method to characterize the structure of bipartite networks, which we call the Method of Reflections, and apply it to trade data to illustrate how it can be used to **extract relevant information about the availability of capabilities** in a country.”*

- Caldarelli G, Cristelli M, Gabrielli A, Pietronero L, Scala A, et al. (2012) A Network Analysis of Countries’ Export Flows: Firm Grounds for the Building Blocks of the Economy. *PLoS ONE* 7(10): e47278. doi:10.1371/journal.pone.0047278

*“Therefore the origin of the differences in the wealth of countries **can be inferred by the record of trading activities analyzed as the expressions of the capabilities of the countries.**”*

*“Hidalgo and Hausman (HH) have introduced in [13,14] the fundamental idea that the complex set of **capabilities of countries** (in general hardly comparable between different countries) **can be inferred from the structure of matrix M (that we can observe)**. In this spirit, ubiquitous products require few capabilities and can be produced by most countries, while diversified countries possess many capabilities allowing to produce most products.”*

- Cristelli M, Gabrielli A, Tacchella A, Caldarelli G, Pietronero L (2013) Measuring the Intangibles: A Metrics for the Economic Complexity of Countries and Products. *PLoS ONE* 8(8): e70726. doi:10.1371/journal.pone.0070726

*“...recent data-driven works [1–3] **have proposed a complexity approach to measure the intangible elements which drive the competitiveness of countries** starting from the dataset of international trade..”*

*“Due to the difficulty in categorizing, quantitatively analyzing and comparing capabilities, **exported products by each country become in such a scenario the main proxy to infer the level of complexity of a productive system, that is the endowment of capabilities**. In some sense the basket of exported products of a country contains encrypted information about its fundamental capabilities, i.e., the peculiar social and economic substrate on which the complexity of the national economic system is built. In some sense **the basket of exported products of a country contains encrypted information about its fundamental capabilities**, i.e., the peculiar social and economic substrate on which the complexity of the national economic system is built.”*

- Mealy, P., Farmer, J. D., & Teytelboym, A. (2019). Interpreting economic complexity. *Science Advances*, 5(1), eaau1705.

*“A novel pair of measures known as the economic complexity index (ECI) and the product complexity index (PCI) were recently introduced **to infer information about countries’ productive capabilities from their export baskets** (2, 3).”*

- Sciarra, C., Chiarotti, G., Ridolfi, L., & Laio, F. (2020). Reconciling contrasting views on economic complexity. *Nature Communications*, 11(1), 3352.

*“**Within this new class of metrics, the productive knowledge owned by each country – which embeds capabilities, finance, technology, human capital and resources, and determines the country’s potential for economic growth – can be extracted from easy-to-find data.**”*

*“The grounding **hypotheses about the hidden capabilities of countries and on how these can be deducted by looking at the export baskets of countries upon which the EC algorithms are built** - are preserved in our framework.”*

*“This result is confirmed by the reordering of the matrix N according to the GENEPI values (Supplementary Fig. 6, Supplementary Table 1), showing that **countries with higher complexity share similar sets of capabilities, as their export baskets are similar.**”*

- Hidalgo, C. A. (2021). Economic complexity theory and applications. *Nature Reviews Physics*, 3(2), 92-113.

*“**Economic complexity metrics can be used to measure the presence of multiple economic factors in a way that is agnostic about what these factors might be.**” (p. 8)*

- Balland, P. A., Boschma, R., Crespo, J., & Rigby, D. L. (2018). Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional Studies*.

*“**Hidalgo and Hausmann (2009)**, in their analysis of product exports by country, **offer an alternative measure of complexity** that reflects the difficulty of mastering the capabilities required to export a particular commodity (indexed by the rarity of exports of a given type), the diversity of capabilities held by different countries and the relatedness between them. **An analogy with the board game Scrabble is often used to explicate understanding of this measure.** Thus, those **countries (players) that have rare capabilities (letters) tend to***

produce more complex outputs (words), especially when those countries possess broader sets of related capabilities (a greater variety of letters that may be combined to produce many different words)."

- Can, M., & Gozgor, G. (2017). The impact of economic complexity on carbon emissions: evidence from France. *Environmental Science and Pollution Research*, 24, 16364-16370.

*"The ECI is calculated with trade data from the United Nations (UN) COMTRADE; thus, the data are only based on the export basket (not all goods produced). **This indicator aims to demonstrate the production characteristics of the economic system by taking into account the capabilities of a country** (Hausmann et al. 2011; Minondo and Requena-Silvente 2013). Therefore, the economic complexity is an indicator of a knowledge-based, skill-based, and sophisticated production of a country. In other words, economic complexity can be defined as a knowledge-based and skill-based production structure of a country with an efficient output level. The high value of the ECI is a sign of the highly sophisticated manufacturing capabilities of the country's production structure (Sweet and Maggio 2015)."*

The field sees complexity metrics as accomplishing an inference task. (Notably, the quotes vary in how they describe precisely what is inferred. This is an aspect of a looseness that exists in the field in how these metrics are being described.) We are making the point that our metrics have a straightforward meaning as coordinates, and need not have any extra meaning for inferring what the literature calls "capabilities" and "complexity". We didn't need to use these concepts at all. Nevertheless, using conventional arguments from dynamical systems theory, we arrived at two coordinate measures that closely resemble the two main competing "complexity metrics".

4. There are some interesting results that the authors do not seem to emphasize, but to me, seem more important than some of what they have now in their interpretation. One of them is showing that measures of economic composition (ECI, ECI) are robust to changes in definitions of the proximity matrices. This seems to be a valuable contribution that is present in the results section but does not make it to the interpretation and discussion. This result is valuable as it adds to the literature and tells us that these metrics, as general concepts, are robust to variations in their implementation.*

We are glad the Reviewer sees value in showing that the b coordinate (a.k.a. the ECI*) resembles the ECI under different definitions of the proximity matrix. Strictly, we do not show that the conventional ECI is also robust to changes in the definition of the proximity matrix (though it would not surprise us if it were), as we always used the same proximity matrix to compute the conventional PCI. (I.e. we varied the proximity matrix that we used to compute PCI*, and then compare the resulting ECI* against a single fixed conventional ECI, which was always computed from the Φ^P matrix.) We added a comment to the Methods to clarify this.

5. Another result along those lines is the authors showing that it is the composition variable (b , ECI^* , ECI , second eigenvector, etc.), and not diversity (A , $Fitness$), the one that is more strongly related to GDP per capita (the authors do not study growth). This is important in two ways. First, this distinction is foundational to this literature. The original ECI paper compares ECI (composition) with diversity (HH , $Entropy$), to show that these measures of complexity capture information of composition that goes beyond diversity alone. Second, the original validation of these metrics was their ability to predict income and growth. The authors seem to concede that it is the compositional metrics (b , ECI^* , ECI) the ones that explain this relationship (although this is an empirical claim derived from a figure, and not a mathematical demonstration).

We agree it is intriguing that the ECI^* , controlling for A , is the variable that correlates with GDP. However, we want to avoid a discussion that leans toward saying this makes this metric more important than A , for several reasons. First, even though there two groups we don't want to gloss over all distinctions between quantities and equate the behavior of ECI^* with ECI and A with every diversity-like metric. There are differences between these quantities and they should be taken into account. Second, one of our most important points is that ECI^* and A were derived without a model that contains capabilities. In this sense, neither the ECI^* nor A is derived in a way that supports the original intent of these metrics and the narrative by which they were correlated with GDP (note that we do not exclude that capabilities play a role, we just explain that a simple alternative explanation exists). Third, our results also emphasize that these metrics are incomplete without one another. However, we do note in the text the distinction between ECI^* and A . We highlight that income increases with b , and the association with A is weak. We also make it explicit that "Note that, although countries often diversify as they rise in income\cite{Imbs2003}, the observations above do not directly associate higher income with higher diversity of exported products. Rather, they suggest that what matters more for a country's GDP per capita is the composition of exports\cite{Hausmann2007}."

6. On that note. If diversity metrics can be obtained by simply counting the number of activities an economy is specialized in. Why do we need a metric like $Fitness$? Seems like a complicated way of counting something as simple as the number of activities. The authors do not interpret this result.

If diversity-like complexity metrics are effectively just counting the number of activities, then indeed metrics like $Fitness$ would be overly-complicated ways of measuring diversity. But as we note, our results don't completely rule out the possibility that complexity metrics are also accomplishing an inference task, i.e. in addition to characterizing the diversity and composition of economic baskets. Rather our results present a route to complexity metrics that doesn't require them to be inference measures. So whether we need a metric like country $Fitness$ depends on whether there are reasons to think that $Fitness$ is also accomplishing an inference in addition to characterizing diversity, something our dynamical model can't definitively

address. In addition, we did not dwell on this point because the correlation between diversity and Fitness is known. On the other hand, our results do highlight that the ECI is incomplete without some kind of diversity metric. (Notably many countries can have high ECI without being developed, but if a country has both high ECI *and* high diversity then it is developed.)

7. On page 18 the authors claim the use of a system of equations to calculate complexity could be unnecessary. But that system of equations can be solved on six lines of code, while the linear approximation model proposed by the authors requires many more complicated steps to reach a very similar result (ECI). Why would a simpler way to reach the same result be unnecessary?*

We are not sure if the Reviewer refers to the complexity of computing these quantities (ECI or ECI*) from data or the complexity of the underlying math. For computing these quantities, the amount of code is similar (e.g. a handful of matrix manipulations, eigenvector calculations, dot products, etc.)

In this passage we contrast the structure of our theory with a common mathematical framework for complexity metrics. Our mathematical derivation might (?) involve more algebra, though we're not sure it matters how many steps a derivation has. What we think matters is how simple the starting point is and how sound the derivation is. We start from a well-known empirical regularity, and work out its implications with standard math. By contrast, while apparently simple, the system-of-equations approach has generated years of debate. In addition, a point that may get under-emphasized is that this approach needs further explanation or steps to be completed. For example, to reach the ECI, one needs to explain why we should focus on just the 2nd eigenvector that satisfies these equations, and also why we should reject higher order eigenvectors that satisfy these equations as well. To reach country Fitness, one needs to resolve issues of computational convergence of the simultaneous equations to a fixed point.

8. The authors claim on page 21 that one of the advantages of their approach is that they derive these metrics from an economic model. But the authors do not present an economic model (no optimizing agents, prices, first and second order conditions, etc.). They present a non-mechanistic linear algebra model in which a vector changes according to a mixing matrix (the proximity matrix).

We revised the phrase on page 21 to avoid possible misperception. The phrase was changed to an "evolutionary model". Together with other descriptions of the theory (noted earlier) we think this will avoid confusion about the type of model we study.

9. On that note, the model seems to lack any validation, other than helping generate variables that are correlated with metrics of diversity (e.g. Fitness) and complexity (e.g. ECI). The model

would also predict that—over the very long run—all countries would produce all products (which would be a bad prediction). Therefore, it lacks any dynamics of comparative advantage or specialization. So, why do we need the model? Could we have just derived b or ECI^* from the eigenvectors of the mixing matrix (ϕ_{pp}). What is the model giving in addition to an alternative way of deriving indicators that we knew? Is it a principled model? Or just the first order approximation of a linear system ($Vector = Vector + Matrix * Vector$)?

We thank the Reviewer for this comment. This asks about both the agreement of the model with data and about our added value, so we take these issues in turn (starting with the latter).

First, the model we study is not exactly new – it’s a dynamical version of regression models for the PoR used throughout the literature. We make relatively small modifications to set these up as a dynamical model in order to work out their long-run implications. But the majority of the modeling effort has been done, in work across the literature that develops and validates the PoR. We see our contribution in reposing these models in dynamical terms and understanding their long-run implications.

As far as these implications relate to complexity metrics, we obtain a new derivation of these metrics that matters for how we interpret them. The derivation here doesn’t need the concepts of capabilities and complexity and doesn’t carry any suggestion that they infer something hidden. We just ask “What is the long-run behavior of economies that obey the PoR?” This question has a concrete answer whether or not “complexity” is even a quantity of economies to be measured. Answering it leads to coordinates that closely resemble complexity metrics, but just track diversity and compositional changes that countries undergo over long times – i.e. these metrics need not have a deeper meaning for inferring a hidden property of economies.

Incidentally, to buy a “complexity” interpretation of complexity metrics, we think more structure on the problem is needed. Perhaps the ECI or ECI^* could measure complexity in addition to directly characterizing an export basket’s composition, but this requires more assumptions (e.g. about Mcp matrix). The main assumption we use here is that a country obeys the empirically-validated Principle of Relatedness for extended periods of time.

Regarding agreement with observation, given that models never agree 100% with the world, we were very surprised that such a simple model connects these three things:

- (1) observations of short-run country evolution (the PoR),
- (2) observations of long-run country evolution (classic economic studies of structural change),
- (3) the functional form (if not necessarily conceptual meaning) of complexity metrics.

Our framework is in agreement with the first set of observations by design (it was our starting point). However, it was not obvious that the long-run behavior of the PoR would align with classic stylized facts of long-run development, nor that it would yield quantities resembling complexity metrics. Nevertheless, like all models, this model will also deviate from what we see

in the world. To correct shortcomings of the particular model of Eq. 2, the remedy would be the usual one of constructing a more sophisticated model (for example, by adding terms that capture other processes of change). Afterwards, one would follow the steps we followed here of identifying the model's long-run behavior and deriving coordinates that summarize it.

We like that the model features long-run diversification to all products. A point that the field of economic complexity has been at pains to make is that, to first order, economic development entails broad diversification, not specialization, as traditional economic theories would suggest (to the detriment of policy). The tendency toward diversification has been demonstrated empirically (e.g. Imbs & Wacziarg), and at higher levels of product aggregation it is the case that developed economies produce nearly every good. This is not to disregard specialization as a force in shaping export baskets. To build this in, the remedy (as noted earlier) is to create a more accurate or complete short-run model. (For example, one could estimate non-linearities in the PoR, or look for discrepancies between short- and medium-run predictions of the PoR or, going deeper, build mechanisms for specialization according to economic principles.) Reaching a more accurate representation of short-term behavior should improve long-run predictions as well. However, this wouldn't change the fact that countries by and large diversify to all products (at broad levels of aggregation) as they develop.

10. The argument that the “fundamental” matrix is that of products seems incidental (page 19). Technological relationships between products also change over time (e.g. textiles were related to water power in the eighteenth century but not today). The “model” is based on this mixing matrix because the authors chose to build a model for a column of the specialization matrix R_{cp} (R_p), instead of a row (R_c). If they would have made the linear model about changes in the spatial distribution of products, instead of the specialization of countries, the mixing matrix of the model would be mixing economies (e.g. $\Phi_{cc'}$). So the claim of having discovered which matrix is “fundamental” using this linear model seems an overreach.

We were not claiming to have discovered that the product matrix is fundamental. What we say here is that making such an assumption represents a difference between our theoretical framework and the usual one:

“Under the mindset we adopt to derive our results, this matrix [the product matrix] is the fundamental one, capturing technological relationships between products. The country similarity matrix is incidental, capturing current similarities between countries depending on where they happen to be in their development.”

We edited the lead-in to this paragraph to make this clearer. We are describing a shift in thinking that the theory here would have us take. Whether this different mindset is better than the usual one is yet to be determined. Notably though, taking the product matrix to be fundamental is the implicit mindset of the Principle of Relatedness. There are versions of the PoR where a country space is added, but it is less clear why products would diffuse primarily

across countries that are similar in the Country Space (as opposed to between geographical neighbors, or countries that interact heavily through trade). If the reason were that these countries are similar in capability structure, then the Product Space would still be the one that drives the dynamics.

We agree technological relationships change with time, and indeed we think this is an interesting direction to think along (SI Section 1.8, last paragraph).

11. In the passage about oil (page 19) the authors seem to confuse complexity of products and processes. Crude petroleum extraction is a complex and difficult process that involves many high complexity inputs (drilling equipment, offshore rigs, etc.). But the product itself (crude petroleum) is not complex (it is whatever was underground). Moreover, since all of the things that are needed to extract petroleum can be brought into an offshore rig in the middle of an ocean, this shows that it is an activity that does not require any locally pre-existing complexity. Similarly, works-of-art do not require sophisticated drilling equipment, but they do take place in cities rich in knowledge flows and sophisticated markets, flourishing primarily in complex economies (e.g. New York, Paris, Tokyo), so their production should be correlated with the complexity of the economies exporting them.

Product complexity metrics are said to inform us about the capabilities an economy needs to make a product. Their aim is not to characterize physical features of the product itself. We agree that oil production requires complex production capabilities. For that reason, it is surprising that oil is rated as low in complexity by the PCI (if the PCI is understood with its usual interpretation).

That oil extraction can be done in countries with few productive capabilities likely reflects the fact that this activity is so profitable that companies can afford to transplant expensive equipment and entire skill ecosystems to the point of production. That doesn't mean oil extraction doesn't need complex knowledge inputs.

To put it differently, why would we say that oil rigs are not complex (since they don't strictly require local pre-existing capabilities), but not say the same about e.g. car plants and many other activities? There are many other economic activities that *do* have high PCI values that clearly need complex knowledge inputs, but just would strictly require moving complex equipment and knowledgeable workers into one place.

It is evident that paintings can be done in places without hugely complex infrastructure. We are suggesting that such activities nevertheless have a high PCI value because they are most readily supported (from the demand side) in economies that are rich and developed. We don't say this is the only explanation for the anomalies in the PCI ranking but it is a parsimonious one: The PCI may be measuring which industries grow in emphasis over the course of development, not complexity per se.

12. In page 10 the authors refer repeatedly to $A(t)$ as ability, but since this has the same units as $R(t)$ ($R(t)=A(t)1 + \dots$), it is an indicator of specialization (RCA). The change of language is confusing and jumps into an interpretation of the variable. It may be simpler to stay with what it is descriptively.

We actually used the same language for both quantities. We say the RCAs (i.e. the R_{cp} 's) are treated as measures of inferred ability in an exported product. In a new SI section (Supplementary Note S1), we discuss the RCA further, and among other things we note various ways in which the literature describes RCAs, as showing e.g. when a country is an effective exporter of a product, with the country able to make a product at a competitive enough level to realize significant exports [more than the country's "fair share"]. While the RCAs have been described with a variety of phrasings (comparative advantage, technological advantage, competitiveness, comparative performance, etc.) we do not have strong opinions about this. Since we did not come up with this metric, we do not wish to defend or interpret it strongly, and our main goal is to follow the literature in using RCAs to quantify some form of ability that is spreading across a network.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Remarks to the Authors:

I thank the Authors for the scientific discussion we established throughout these review rounds, a discussion I am truly satisfied of, and I might say, it was definitely useful. The paper has benefited from this debate, with significant additional information included in the manuscript and supplementary information. If compared with the original version, it now better addresses the dilemma in the economic complexity fields, detailing the controversy between narrative and mathematical framework, and provides a better understanding of the reasoning behind the research question and how to address it. I feel the Authors' efforts in making their point in their Reply letters, a point that can finally be now read in the manuscript, even though I still have a few, minor, concerns. Regarding the text, I found the Reply letters to the Reviewers more straightforward and convincing than the tone of the article itself; especially in the background literature and discussion. I would suggest incorporating some of their defending statements into the article, if possible, especially regarding the above-mentioned dichotomy between narrative and mathematical aspects.

We thank the Reviewer for their feedback, and for the suggestion to include more of the language from the last reply. This is tricky. In being more explicit, the language we used in our reply can also be read as more aggressive and confrontational than we intend to be. We worry how it affects an already contentious discussion, and the reception of the paper. There is also a severe length challenge. In our last reply, our first response (for instance) took six pages, summarizing what previous papers had and had not done. In trying to incorporate language from the reply, we kept finding it difficult to put in individual sentences without inserting statements that feel unsupported and tendentious.

We did tentatively add a new paragraph to start off the Discussion that draws on language from the previous reply letter. The paragraph reiterates points on the dichotomy between narrative and math in economic complexity, and the takeaways from our work. The points in this paragraph are made elsewhere in our paper and we think the Discussion can work with or without it. We're not sure if the Discussion / paper reads better with or without and have asked the editor for guidance on which route is preferred (i.e. including or not including the new paragraph).

A few notes:

- I have a further point on the example the Authors report regarding product complexity. As detailed on page 19, the Authors claim that "pottery and works-of-art are clearly low-complexity goods", countering the results of the analysis on product complexity. This is a bold statement. To date, there is little study-knowledge-agreement on which are the intrinsic

characteristics of products that make them “complex”, and complexity from the metrics translate into capabilities of production, which I can suppose from the text that the Authors are imputing to supply chain, translocation of work, skills and knowledge. While I agree with the interpretation of the product complexity results concerning structural change that the Authors provide, and with the statement per se, I don’t feel this statement is properly put in the current literature framework, and I’d reduce the emphasis on it.

We thank the Reviewer for pointing out the potential for issues here. We treated this comment in combination with one from Reviewer 4 on oil production, which was the other example we gave in this paragraph of a product whose complexity estimate (by the ECI and product Quality metrics) defies expectations.

We made two kinds of changes. First, in the main text, we reworded the text to soften the original statement, noting that “pottery and works-of-art” is a product class that is *plausibly* low in complexity.

Second, we noted that the intuitive expectation we set up for these products (i.e. that “pottery and works-of-art” should have low complexity, and oil extraction high complexity) is consistent with data on occupational diversity and wages. Occupations are frequently taken in the economic complexity literature as kinds of production capabilities. In a new section in the SI (Supplementary Note 8), we look at the occupational diversity of the two NAICS industries “Oil and gas extraction” and “Independent artists, writers, and performers” with data from the US Bureau of Labor Statistics. We find (not surprisingly) that oil extraction ranks near the top in occupation diversity (top 13% of industries), and that artistic performance ranks relatively low (bottom 39%). In addition, we examine the median wages of occupations engaged in each industry, because it has been argued that the wages of an industry should typically be higher in complex industries. We find similar results here too, with oil and gas extraction ranking near the top (4th place) and artistic performance ranking relatively low (173rd) out of 247 industries.

We agree there is no full agreement on how to measure the complexity of products and the discussion here underscores this. The PCI and product Quality are intended as inferred measures of complexity by gleaning a signal of complexity using e.g. export data. Occupational diversities are a direct measure that count some productive capabilities, but they miss any capabilities that aren’t occupations. Wages (under the assumptions of the literature) are thought to be a *correlate* of complexity. So no measure is without drawbacks, but if we cannot mention products whose estimated complexities using complexity metrics defy expectations that are set by other means (whether intuition using the field’s narrative about complexity, or other imperfect quantitative measures), then complexity metrics simply “define” complexity and there is no way to validate them. Moreover, we offer an alternative interpretation of these metrics that resolves the counter-intuitive results. That is, we are not nit-picking about individual observations in the dataset (any statistical method will produce outliers), but are instead pointing out that these observations can be reconciled in our framework.

- Something that I think is missing though is a ranking of countries according to the principle of relatedness, it would be interesting to see some results. I don't know if this is intentional, yet there is an interest in the rankings fueling the hype around complexity metrics, and that might be of interest to practitioners and stakeholders, as the Authors have pointed out.

Thank you, we added the ranking to the SI (Supplementary Note 6). This is the ranking of countries according to where they lie along the ECI* axis. It is similar to the ECI ranking (shown for comparison). We added this table with a note saying that our findings would suggest interpreting these rankings (and potentially those of other metrics) differently than usual (summarizing how much the economy is weighted towards some goods over others/has progressed along conventional axes of development, rather than offering an inference about the hidden variable of country complexity).

-Countries' specific comments on the trajectories shown are also lacking, beyond aggregation through income classification. This is important, especially considering the outreach of the Journal.

Thank you for making this suggestion. We added comments on several countries from different parts of the (A,b) plane to SI Supplementary Note 3 where we show individual country trajectories. Hopefully, this makes it easier for readers to think about country movement in this plane. We think the new text helps to pull out some interesting cases and patterns; for example, similar movement of China and Korea leaving later stages of Region 1 and entering Region 2; initially erratic movements of Region 1 countries such as Egypt, Albania, and Ethiopia before embarking on steadier development upward and to the right in the plane, following the canonical trajectory.

- I do agree that the GENEPI components and the A and b vectors here presented stem from a different perspective, mathematical reasoning, and narrative. Yet, I would like to see the correlation values between the X_1 and X_2 values without the transformation to project the values onto Fitness and ECI. How do these values alone figure in your grouping? The actual representation of the stages was made onto this two rather than their projection onto the axis. If the correlation is high, or low, what further information would thus this measure explain? Following that, how would you explain the lack of correlation of the $X_2/\sqrt{d}$?

Thank you for this suggestion, we added the correlations of bare X_1 and X_2 (without the transformations involving diversity d) to SI Fig. S4. We see that

- Bare X_2 is composition-like (somewhat less than the combination $X_2 / \sqrt{d}$).
- Bare X_1 is neither composition- nor diversity-like (like GENEPI as a whole).

We're not sure what further information one gets from the GENEPI metric or the GENEPI plane, whether we view this metric as a complexity inference as originally intended, or as a way

to characterize overt structural change. For the purpose of characterizing structural change, we have been able to interpret the information in the A and b coordinates in two ways: (i) by exploiting their theoretical meaning within the dynamical model of country change we studied (the PoR); and (ii) by looking at their empirical relationships to diversity and per capita GDP. On the theory front, the X_1 and X_2 components of GENEPLY don't emerge from a model of country behavior. (They emerge from a mathematical unification of the ECI and Fitness metrics using linear algebra.) On the empirical front, X_1 and X_2 are correlated to diversity and per capita GDP, like other country metrics here, though these correlations happen to be weaker. (The Spearman correlation of X_1 to diversity is 0.726, and that of X_2 to GDP per capita is 0.549. The corresponding correlations of A to diversity and b to per capita GDP are 0.961 and 0.653, respectively.) As a result, we don't see any particular informational advantages of X_1 / X_2 / GENEPLY over the (A,b) coordinates. It might be that X_1 and X_2 are also useful axes on which to describe country evolution, conveying the information of development in some other way that is interesting, but it is just not obvious how.

For the purpose of inferring complexity, we're also not sure whether the GENEPLY metric / GENEPLY plane gives further information than existing complexity metrics. If one (i) accepts the premise of Sciarra et al. that country Fitness and the ECI extract some hidden information about a country's capability count, and that (ii) the particular combination of these metrics represented by GENEPLY is a good combination, then, yes, GENEPLY would be a better inference metric than either country Fitness or the ECI. But country Fitness and the ECI are derived with heuristic arguments, neither one mathematically derived from a model with country capabilities, showing how these metrics ought to infer capability counts. The GENEPLY work does not resolve this fundamental issue. The mathematical connections it uncovers run strictly between the country Fitness, ECI, and GENEPLY metrics themselves, not between capabilities and any of these metrics.

[Incidentally, in studying these correlations, we note that GENEPLY very strongly resembles its X_1 component (Spearman correlation 0.987), while the correlation with X_2 is a good deal smaller (Spearman correlation 0.493). This is not a surprise and can be seen immediately in Fig. 1 of Sciarra et al. Higher values of GENEPLY are associated with higher contours, and the ranking of countries by their position across these contours is visibly very similar to the ranking along the x-axis (X_1) by itself. The weaker correlation with X_2 can be expected from the knee-shape of the data noted by the authors. As one moves to higher contours, X_2 first decreases, then increases. At first glance, the fact that GENEPLY is so strongly correlates with X_1 could be argued to undercut the information benefits of using both X_1 and X_2, i.e. of plotting countries in the GENEPLY plane, rather than on simply a 1D scale using GENEPLY by itself or the X_1 component by itself. However, Sciarra et al. don't appear to claim that putting countries on the GENEPLY plane should reveal more information about them than just rank ordering them. (I.e. their goal in showing this plane might simply be to show the relationship of X_1 and X_2 with GENEPLY.) Evidently, the construction of GENEPLY is such that it more closely resembles X_1 than X_2. Under their assumptions, country Fitness (mathematically related to X_1) and ECI (mathematically related to X_2) are both succeeding in inferring some information about "capabilities", and the very strong resemblance to X_1 is not necessarily problematic if one

thinks a good combination of these two complexity metrics should derive more from country Fitness than from the ECI. Again, the key issue is that there is no math establishing the ability of country Fitness and ECI to infer information about capabilities in the first place.]

- In Supplementary Note S1 regarding the RCA, I'd rather not include the comment on the RCA effect as a note (which is not even admissible by the NC guidelines), better placing it in the main text.

We thank the Reviewer for this suggestion - we moved this material to the main text, placing it in the Methods right before we discuss how we compute the ability vectors R . We slightly condensed and reworked it, and drew on some of the Reviewer's language from the previous (i.e. 2nd) round of review (e.g. noting that the field of economic complexity has been "marked by the presence of Revealed Comparative Advantage in its methods" and that their use is meant to both filter out noise and reveal significant exports.)

- I would recommend checking editorial requirements

We checked the editorial requirements and indeed found a number of fixes to make. Thank you for this suggestion.

Again, I wish to thank the Authors for the high-level discussion, which touches a few streams of research from the past and ongoing, and I'd be glad to go further with the discussion in other rooms.

We thank the Review greatly for their feedback throughout the review rounds. We agree, it has significantly improved the paper.

Reviewer #4 (Remarks to the Author):

Bridging the short-term and long-term dynamics of economic structural change is an interesting paper that adds to the ongoing literature on economic complexity. The revised manuscript presents an improvement from the previous version, but it still fails to incorporate some of the revisions suggested by the reviewers. At these point, these are changes that mostly require rewording some sections. I hope the authors take these changes into consideration and use them to improve their next version of this manuscript.

1. The authors seem to misunderstand the use of the word complexity as they see their derived measured of complexity ($b(t)$) as a measure of composition. Maybe, this is because the authors are unfamiliar with the etymology of the word. According to the online etymology dictionary complexity refers to the "composite nature, quality or state of being composed of interconnected parts." Wikipedia's entry for complexity also uses "a system or model whose components interact in multiple ways." Oxford's dictionary entry for complex defines it as "consisting of many different and connected parts." Complexity is therefore a noun that correctly describes a composite system (e.g. made of multiple industries, products, etc.), which is what metrics of complexity have long tried to communicate (a description of economic composition). Their measure of complexity $b(t)$ is a measure of the long term composition of the economy. The authors should positively leverage this connection with the literature and the etymology of the word instead of antagonizing it.

We thank the Reviewer for their feedback here and below. We are aware of dictionary definitions of complexity. We are also familiar with a variety of ways this term is used in academic research. It probably goes without saying that this word is used in many ways (in and out of academic discourse). In our last round of responses, we showed a list of quotes (13 quotes from 8 sources) from the field of economic complexity, exemplifying the understood goal of complexity metrics. These quotes also showcase how complexity is understood in this particular literature. Complexity refers to the count and sophistication of country "capabilities", as understood in a tripartite network framework first laid out in Hausmann and Hidalgo 2009 [1]. Metrics of complexity, in turn, are meant to infer hidden information about these capabilities. However, the way we derive our two coordinates gives us no reason to ascribe such a meaning to the $b(t)$ coordinate (or the $A(t)$ coordinate). Indeed if we were to call either of these quantities 'complexity', we think it would be quite confusing.

2. While the work does connect to several efforts in the literature, there is recent work that has advanced our mathematical understanding of economic complexity metrics that would be interesting to see discussed in this context. In particular, the works of Bottai et al. and Servedio et al. providing interpretations of economic complexity. These contributions give a closed form solution to the objective function optimized by complexity (the sum of $ECI-PCI^2$) and build on the bipartite clustering interpretation of Mealy et al. (2019)

Bottai, Carlo, Jacopo Di Iorio, and Martina Iori. "Reinterpreting Economic Complexity: A co-clustering approach." arXiv preprint arXiv:2406.16199 (2024).

Servedio, Vito DP, Alessandro Bellina, Emanuele Calò, and Giordano De Marzo. "Economic Complexity in Mono-Partite Networks." arXiv preprint arXiv:2405.04158 (2024).

We thank the Reviewer for pointing us to these papers. We would like to point out that both papers came out as preprints only after the second round of this paper's review. Nevertheless, we now cite the first paper in Section 2D, where we write "While our results invite us to ask whether complexity metrics just summarize structural change rather than infer hidden information about capabilities, they do not rule out the possibility that these metrics play both roles simultaneously. Recent works explore this by making assumptions about the structure of capabilities and how they are acquired through the process of development." This citation helps reinforce the idea that these two possibilities are not mutually exclusive.

For the second paper we were not sure how it impacts our results. As the Reviewer notes this paper identifies an objective function that is maximized by the ECI algorithm. However, it does not show how optimizing this function leads to an inference of capability counts of countries and products. The paper's stated contribution is a novel extension of "economic complexity algorithms" to non-bipartite graphs. However, we're not sure how to think about this in terms of economies and how they change over short and long time scales – the two classes of nodes in the original network are "countries" and "products", distinctly different things that we wish to continue treating as different entities in order to understand structural change. (E.g. what does a country-country or product-product link convey in the more general case? The paper offers no interpretation.) The paper also contains results connecting its metrics to the vulnerability of networks to node removal. While interesting, we're not sure how to connect this to our findings.

3. Figure 3 is an important part of the paper that would benefit from a properly labeled panel (a scatter plot with ISO codes for each country). I understand this would require a larger panel, but it would be very for economists with an intuition of each economy. For instance, where are low diversity high complexity economies such as Finland or Singapore located? Are they considered tax heavens? Pushing this to the appendix obscures valuable information for the reader. This should be an easy comment to incorporate.

We thank the Reviewer for this idea. We agree, it would be interesting for many readers to see how individual countries were situated by our coordinates. We made this panel, and added it to Supplementary Note 3. We placed this figure in the SI just because the paper is quite long, and the identities of the data points in the figure don't directly bear on our findings. Mainly in Figure 3 we wish to (i) show how our axes relate empirically to diversity and composition, (ii) give further intuition about these axes by showing what kinds of countries (wealthy, poor, resource-based, etc.) occupy different regions of the plane, and (iii) show the trends of country movement. These help support one of our core findings, that these coordinates can be understood simply in terms of established facts of long-run development. The new figure uses ISO codes to identify countries. We kept the marker encoding from the main text Figure 3c to show which countries were classified as OECD, developing, tax havens, or resource-rich under

the criteria and data sources given in the caption to Figure 3. Regarding Finland and Singapore, both are situated in Region 2. Both have relatively lower values of the A coordinate compared with other countries in this Region (i.e. agreeing with the intuition that these are relatively lower diversity, high-complexity economies.)

4. From the early days, the validation of metrics of economic complexity has involved econometric models where complexity metrics are used to explain macroeconomic outcomes (first growth, but then inequality, emissions, etc.). The paper would benefit from including such a model in the appendix and mentioning its results briefly in the main text. Does ECI* behave equally to ECI when it comes to predicting growth? Worse? Better? How about other outcomes? The reader is left wondering about the impact of this work (does it change the ability of ECI to explain an outcome that economists care about (e.g. growth)?). Adding at least a growth regression should help in this regard.

We thank the Reviewer for this idea. To be sure, we never said our coordinates would be better predictors of growth. We emphasized that our coordinates are derived to be descriptors of development paths.

However, we understand the curiosity to explore this topic. Given the resemblance of our coordinates to complexity metrics, and given that complexity metrics have been used to predict growth since their introduction (as the Reviewer notes), it's natural to ask how our coordinates perform at this task.

We added results regressing growth against our coordinates in a new SI section (Supplementary Note 10). Our coordinates perform well in comparison to the most prominent metrics (ECI and country Fitness). Regressing 5-year, 10-year, and 25-year growth, as well as GDP per capita levels, the b coordinate (ECI*) exceeds the ECI in explained variance. We limit the sample of countries to those that are neither resource-dependent nor tax havens. However, using all countries does not change this conclusion. Similarly, we arrive at the same conclusion regardless of whether we fix PCI in time using the 1962 trade patterns or allow the PCI to change over time (which is what we report in the SI).

It is interesting that our metrics perform well at predicting growth (even beating other measures), though we honestly don't know what should be concluded from these results. They are orthogonal to what we are claiming. The usual interpretation of complexity metrics says these metrics infer hidden information about countries' underlying productive capabilities. The fact that these metrics are also able to forecast growth is then taken as a promising sign consistent with this narrative. However, as we emphasized, our coordinates can be derived without using the narrative of country capabilities. Our coordinates are derived just as descriptors of long-term changes economies undergo in development. The fact that they are able to forecast growth is certainly interesting, though perhaps not shocking. Evidently, growth rates are somehow connected to the basket of goods that an economy produces and how this basket evolves over time. But the question of course is how – how and why are some baskets of

goods more growth-enhancing than others. Our results do not resolve this. In addition to the new SI section we added text to the main paper referring to these results and noting the points above.

5. On page 9 the authors try to build a distinction between a system made of two coupled equations (F and G, where F can clearly go into G resulting in a single equation) and a system made of a single equation. Basic college math tells us that coupled systems are a way to write a system instead of a fundamentally different idea (e.g. a second order differential equation can be expressed as two coupled first order differential equations, a Hamiltonian uses two equations and a Lagrangian one, but you can go from one to the other). The authors should again try to leverage the connection with the literature, or remove this section.

We thank the Reviewer for this comment. We revised this paragraph on page 9 to clarify it, and also added discussion text to the SI (Supplementary Note 7) to be more explicit about the mathematical issues we are highlighting here.

In brief, the country- and product-level variables we derive in the paper cannot be written as a system of coupled equations like the pair of conceptual equations shown at the top of the paragraph on page 9. In the closest approach, our math contains something analogous to the first of these equations, but not the second. The first equation conceptually says that country complexity is a function of product complexities (e.g. in the case of the ECI, it is an average of these complexities). While we don't interpret our A and b coordinates in complexity terms, we made clear that these country-level variables are weighted averages of product-level variables, and thus the parallel in math is very close here. But we cannot in general write our product-level variables as a function of country-level variables, in analogy to the second equation on the top of page 9. The core reason is that our product variables (i.e. the elements of the v_2 eigenvector / PCI^* s) are computed from the product proximity matrix that is taken as a basic starting point for the dynamical model. This proximity matrix is a primitive of our model, and it can be estimated in numerous ways that cannot necessarily be duplicated by a calculation using country-level variables. (Indeed the proximity matrix between products might be constructed with different datasets entirely that have nothing to do with trade or export data. For instance, one can calculate the proximity between products from information on the occupations hired by the industries that produce them, while using export data to track a country's portfolio of products through time.) In the new SI section, we use the example of the ECI to make these points more explicit. We show mathematically where the parallels are in the calculations between the calculation of the ECI^* and the ECI, and where the two frameworks differ (i.e. the fact noted above, that the proximity matrix Φ can be estimated in many ways).

6. On page 9 the authors also seem to confuse the complexity of processes and products. Crude oil is not a sophisticated product, but its extraction is a complex process involving complex products (Eg. Excavation machinery (HS8431) is a high complexity product (about one standard deviation above the world's average). Trade data is fine grained enough to differentiate

between the two (e.g. the countries that export oil extraction service activities and machinery are more sophisticated than those that export the oil). This seems to be a false "conundrum." I would suggest removing this section.

We thank the Reviewer for this comment. A similar one came up in the last round. Our reply here echoes some of the points made previously. The key issue is that product complexity metrics are said to inform us about the capabilities an economy needs to make a product. They do not aim to characterize physical aspects of the product itself. We agree that crude oil itself is relatively simple in one sense (it's fairly uniform in physical properties), and that oil extraction is a complex process involving many inputs. It's for the latter reason we face a conundrum. Producing oil requires many production capabilities, yet oil is rated as low in complexity by the PCI.

For a similar reason, it does not matter that trade data distinguishes crude oil from excavation machinery. These are different products. The production process of oil extraction machinery is distinct from the production process for oil. Here we are discussing the latter. [As another example, like extraction machinery, a computer is a complex product used in the production of other products. But these other products can be simple (e.g. a holiday e-greeting card designed by one person) or complex (e.g. a 747). Whether one's data happens to distinguish computers from these other products does not matter if we want to measure the complexity of those products. To do that, we have to consider their production processes, not that of computers.]

To address both this comment and one by Reviewer 1, we reworded the text in this paragraph and supported our verbal statements with data. In the main text, we soften a statement about "pottery and works-of-art" a little and say that it is a product class that is *plausibly* low in complexity. We then note that this intuition is supported if we look at data that probes complexity by asking how many occupations are involved in production. These results are presented in a new SI section (Supplementary Note 8). Using data from the US Bureau of Labor statistics, we look at the NAICS industries "Oil and gas extraction" and "Independent artists, writers, and performers". We estimate the number of occupations engaged in these industries in several ways. Independent of the measure, oil extraction ranks near the top in occupation diversity and artistic performance ranks in the bottom 40%. We also examine median wages, since it has been argued that wages of an industry should usually be higher in complex industries. We find similar results here, with oil and gas extraction ranking near the top and artistic performance ranking relatively low. None of these measures are perfect, but they all align with the intuition we presented (originally just in words), and are in tension with the complexity values measured by the PCI.

Smaller points.

Page 2. "our results suggest they re not in competition, but offer complementary information." This point was already made by Sciarra et al. in the GENEPEY paper. Should be acknowledge as such, instead of as a new point.

Thank you for highlighting this sentence. We agree it could be read to think we are repeating an idea proposed in Sciarra et al. We revised it to say “Our results suggest they’re not in competition, but measure distinct and complementary properties of a country’s activity basket.”

Sciarra et al. assume the ECI and country Fitness measure the same thing (country complexity). They assume that each metric measures complexity imperfectly, and it is implied by their analysis that they assume each one’s measurement errors are independent. (Such that combining them into the GENEPI metric can result in a better estimate of complexity.) This is not the way in which we mean that the ECI and country Fitness are complementary. Here we’re saying that they simply measure different things. Each metric measures different aspects of a country’s export basket. (Roughly, one measures diversity, the other composition.)

Page 9. What is the point of calling previous research pointless? Isn't the previous research motivating your current work?

Thanks for this comment. To refer to the relevant sentences here we pasted them below:

“Second, our findings suggest debates between the main contending metrics could be pointless. Complexity metrics are typically viewed as being in competition, since they represent alternative methods to estimate the same quantity (country or product complexity). But clearly we would not view the coordinates $A(t)$ and $b(t)$ this way; these coordinates just summarize different, complementary information about activity baskets, focused either on their diversity ($A(t)$) or composition ($b(t)$).”

We do not call prior research pointless, but suggest the (heated) debates between the main contending metrics could be. Our target is the debate between the main contending metrics (ECI and country Fitness). We raise this “pointlessness” as a possibility; we see it as an open question that hinges on whether either of these metrics succeeds in inferring complexity.

Our paper asks whether these metrics just summarize different surface information (diversity vs. composition) about an economy’s activity basket. If this is the case, then the debate about which of the main metrics is a better complexity measure would indeed be pointless, since neither metric (in this case) would measure complexity.

Also, the idea that this is an open question, not fully resolved, is reinforced later in this section where we raise the possibility that existing metrics could perform two roles simultaneously. They might both characterize country export baskets at a surface level, and provide a deeper inference about the hidden variable of complexity.

Page 1. "We do not endorse the view that the PoR offers a complete description of long-term economic transformation". Is there such as view? Could the authors provide a reference for it? If not, they should remove this sentence.

We thank the Reviewer for pointing out the possible confusion. We dropped the phrase "endorse the view". The sentence now reads, "While the PoR is an incomplete description of long-term economic transformation, we explore what would happen if we take literally the long-run consequences that are implicit in how the PoR has been formulated."

Before moving through the details of this second review round, I'd like to formally apologize with the Authors for the delay of this report. The due cure that this report needed was conflicting with cumulated sum of workload and unexpected life events. Again, apologies.

Reviewer #1

With respect to the first version, I appreciated the effort the Authors put in implementing the requested changes making the article more readable to unspecialized audience. Yet, despite the improvements, I'm still questioning whether the contribution the Authors claim to make truly is groundbreaking as the journal would require, or re-enters in the debate characterizing economic complexity while increasing the noise around the amount of metrics available.

My discussion stems from what the Authors claim and what literature offers.

1. What the Authors argue to do in tying together two areas of work within the economic complexity field appears to be a little bit highfalutin. I do agree that the way the framework introduced in this work cleans up the many contributions the economic complexity literature has (Table 1). Moreover, it provides a clearer interpretation about the long- and short-term dynamics of structural change. However, I would reduce the flattering arguments. In fact:
 - The principle of relatedness lies beyond all the metrics of economic complexity, and the main difference lie in the way this principle is embedded in the proximity/similarity matrices the short-term metrics are build upon. The EC metrics from Hausmann and Hidalgo are the original examples for this statement. The productive space in the *Hidalgo et al. Science*, 2007, provides the ground for the ECI-PCI metrics to have mathematical reasoning. I may agree there is always room for improvement, but if one has to identify the original example of the principle of relatedness, and the way it relates to the metrics of complexity, that is where I'd search for. The same Authors refer to these pieces of literature, and there is clear mathematical resemblance in among the metrics the Authros present in the Supplementary Material – S7, and the *Science* paper from *Hidalgo et al.* Thus, I am left with doubts about the importance of this contribution.
 - At page three, the Authors claim *surprising* the fact that currently in use metrics of economic complexity either relate to the diversity or composition of the economy. Is it though, surprising? This result has been repeatedly shown over and over, with examples in *Mealy et al., Science Advances*, 2019; *Mariani et al., The Eur. Phys. Jour.B*, 2015; *Sciarra et al., Nat. Comms.*, 2020, and many others – as many referring works this manuscript has, especially considering the correlation with the diversity. Moreover, the GENEPIY metrics with its two component has introduced the idea that the two classical metrics of economic complexity either measure diversity or composition of the export baskets, and the composition of the two leads to the multi-dimensional metric of complexity. Yet, the addition of the metrics the Authors introduce is new and marks my following point.
 - The authors make clear that contrarily to other research works their aim is to introduce a dynamical modeling framework from which to extract the long-term implications of short-term dynamics *keeping distance from the definition of new complexity metrics*. Still, in introducing the *A* and *b* components, they target the comparison with other economic complexity metrics, as the ECI, Fitness and GENEPIY, even defining the variables with resounding names as ECI*, for example. Beyond the different mathematical framework, the convergency of the model's components follows that of the iterative *Method of Reflection* in *Hidalgo-Hausmann PNAS*, 2014, with the first eigenvector resulting to be uniform and the second component taken for more informative results. Again, what can we learn from anew with respect to previous contributions? I am afraid of the risk of falling within the circular reasoning of “729 new measures of economic complexity”
- <https://arxiv.org/ftp/arxiv/papers/1708/1708.04107.pdf>.

2. I do have a point about talking about structural change with respect to the presented framework. This point is not a critique moved for the Authors their-selves, because I can see paves the way for a larger and more philosophical debate exiting the aim of this work. Yet I'd like to raise it here as the Authors can discuss through since "it is not an intrinsic requirement of *our* model" – page 5.
- The economic complexity field has been marked by the presence – in its methods – of the Revealed Comparative Advantage. The role of the RCA is two-fold: it allows us to remove intrinsic noise related to the presence of outlying trade operations and to reveal the significance of the exports. There's plenty of literature, e.g., as Authored by *Giovanni Dosi* to cite one, discussing about how this procedure biases the results. My comment follows the same argument. Can we actually speak about long-term structural changes when the RCA erases important information? Note that the following reasoning works even considering the *log* transformation the Authors introduce, because the critical issue here stands in the RCA threshold ($RCA=1$), the authors still use to map the values onto the log scale. Here's an example to follow. Let us analyze the following trade matrix among the countries c_1, c_2, c_3, c_4 of products p_1, p_2, p_3 . The trade flows range among three different orders of magnitude. Let us just play with the component of the matrix c_3p_3 as in the example below. As shown, when changing the element from $c_3p_3 = 0$ – case (a) – to $c_3p_3 = 10$, case (b):
- In case b, the RCA value of c_2p_1 changes from being greater than one, to be smaller than one, being one the threshold of competitive advantage. Thus, even though the country c_2 has not changed its performance, it may result in some kind of structural change when analyzing the binarized matrix of trade.
 - The value c_3p_3 increases of one order of magnitude – possibly implying that there has been a sort of structural change in the economy – which we are though not able to capture due to the RCA threshold impeding us to catch that 0.996 value.

(a)

trade	p1	p2	p3
c1	100	10	0
c2	100	10	0
c3	10	0	1
c4	0	0	10

RCA	p1	p2	p3
c1	1.043	1.095	0.000
c2	1.043	1.095	0.000
c3	1.043	0.000	1.992
c4	0.000	0.000	21.909

(b)

trade	p1	p2	p3
c1	100	10	0
c2	100	10	10
c3	10	0	1
c4	0	0	10

RCA	p1	p2	p3
c1	1.087	1.141	0.000
c2	0.996	1.046	0.996
c3	1.087	0.000	1.087
c4	0.000	0.000	11.952

So, my question to the Authors is: if the RCA threshold is not an intrinsic requirement of the model, and considering that the framework you present is claimed to be novel, can you show the results of your framework on long-term and short-term dynamics without considering the RCA?

Reviewer #1 on behalf of Reviewer #3

All points raised by the Reviewer have been properly addressed and answered.