Referee #2: Sibel Eker

This manuscript presents a literature review on social norm diffusion/contagion on social networks, and whether a critical mass similar to a Pareto effect is observed. The manuscript makes a relevant and important contribution to the expanding literature on social tipping by aggregating the available empirical evidence on tipping dynamics. I also appreciate that authors clarify and describe tipping-related concepts and terms in Section 2, yet I believe that manuscript would benefit from a more condensed version of it.

My main concern is how "social tipping" is implicitly defined in the manuscript, and that it is not clearly aligned with the real-life examples of what social tipping in the climate change and sustainability context entails. For instance, the manuscript focuses on social networks and complex contagion, but does it mean that all social tipping process, such as rapid reduction of EV battery costs, coverage of climate change in the school curricula, as identified in the authors' own paper from 2020, have to include contagion dynamics through a social network? Even if we can rightfully represent many of those in networks, do we have to? Is the presence of a network structure a necessary condition for tipping? How does this relate to the existing social tipping and "positive" tipping literature that does not necessarily focus on networks? Answering such questions would improve the manuscript.

Firstly a big thanks for the considerate and in-depth comments and feedback. They will certainly help us improve the manuscript.

Regarding the focus on social tipping, the questions you pose relate to the underlying difficulty in disentangling complex systems. Quickly summed, as mentioned in lines starting from 170, we seek to use the tools of network theory as a basis to discuss and investigate social tipping processes. Inasmuch as social interaction is required for social tipping, where the former implies some network structure, we don't have to represent it as such. We do this because of the large body of evidence supporting it as a useful tool for doing so (Berner et al., 2023; Guilbeault et al., 2018; Sayama et al., 2013; Smaldino, 2023, Winkelmann et al., 2022). We hope to relate to existing literature by seeing how previous findings, and social tipping in general can also be explained through network theory, which may improve or provide alternate understandings vs existing methods. Also to clarify; this paper focuses purely on the social element of tipping, i.e. regarding social change. We do not consider socio-economic, technological, or other kinds of tipping.

I hope that helps answer your questions.

We will also include these answers in the text at your suggestion, as we agree these are not explained enough.

Relatively minor comments:

Lines 75-88 : This paragraph does a fair job in listing the objectives, yet they sound a bit too ambitious. For instance, "we identify critical factors influencing tipping processes in social systems" is eventually only limited to a handful of studies included in your review on social networks, therefore I suggest reformulating these objectives more consistently with the actual methods and results.

We agree and will amend these statements to be restricted to what our methods and results show.

Figure 1a: I don't find this figure very clear, especially considering the accompanying text. If the “shaded area represents the domain where the system exhibits a nonlinear response” would any nonlinear curve above the y=x line mean "tipping"? How about the zone below the line? It might also include very nonlinear changes.

We agree this is not really communicated properly. Firstly any curve above or below the y=x line would be a “spillover” or some form of non-linear response but it’s ambiguous as to whether this means tipping per our definitions. We will address this. Secondly, the zone below the line also represents the opposite of a “spillover”, also a non linear change. We will address this by defining the line above y=x as a positive non-linear change, or a “positive spillover”, and below the line a negative non-linear change, or “negative spillover” we will address whether this constitutes tipping in the sense of our own definitions and refer more specifically to how this graph relates. We will also address the text in line 115 to 125 which does not make this clear.

Line 121-122: “where roughly a 20% change in a system control parameter results in an 80% change in the system state at equilibrium”. I am not sure if this is the right formulation of your idea. What happens as lambda moves from 0.1 to 0.12, which is a 20% change in the control parameter?

This is correct and the statement is indeed misleading and wrong. The section you point to has a few issues we already plan to address, such as the unconventional use of the current adopter fraction as a system parameter and not a variable, as well as the potentially confusing use of a cdf. We will also amend the lines to indicate the change is referring not to the current system control parameter but to the maximum potential fraction of adopters.

Line 126-127: “the shown Pareto CDF depicts a scenario in which a minority of actors have convinced a large majority to switch to another social norm.” I think it doesn't show that one, because the y axis is F when t goes to infinity, the final state. When lambda=0.8, we end up at >0.8 and no tipping occurs. The pareto cdf line covers many different scenarios not necessarily those where a minority convinces the others.

We agree that Figure 1.a is currently a bit ambiguous and potentially misleading. As stated above, we will address these concerns as part of the whole section, which includes a slight redesign of the plot for a few reasons including that stated above. The axis points (0,0) are also mismatched and will be corrected. We are currently considering two potential directions for the plot. Both use more traditional approaches similar to recursion or cobweb plots from system dynamics. We refer to two examples from the literature: the former, shown in Figure 2, and the latter, shown in Figure 1, of the respective papers(Efferson et al., 2020; Wiedermann et al., 2020). Hopefully using existing and familiar representations will aid understanding, reduce confusion, and clarify our main message with the plot.

Figure 1b: Appreciating this plot of tipping dynamics over time, yet it is the typical s-curve of transitions/innovation adoption. Could you expand on what makes it “tipping”? Furthermore, it is not clear on the figure what the tipping threshold is, as suggested in Box 1.

The point of this figure is to try and reconcile the concepts that are introduced or will be introduced in the paper with concepts existing in the literature on social tipping. Similar to the transition/innovation and adoption, social tipping events resolve as the typical s-curve due to inherently similar spreading dynamics which occur in these systems. Not to mention that the former usually occur through social interaction. What this figure represents under the “tipping phase” can be seen as tipping under the notion of the original authors from where we draw the language for each phase. You are correct if you are implying it does not qualify for our definition of “Tipping event” as per Box 1, where the figure 1b shows the transient dynamics, and our definition is based on the steady state. This is an oversight and will be addressed so as to reduce ambiguity. One approach would be to state that social tipping used in the sense of the notions from the original authors has a slightly different meaning to our definition. Another would be to provide a different definition of the “social tipping” stage, for example more closely drawing on the definitions(section 4.2 and SI) from work by Winkelamnn, et al. (2022).

Line 182: Using cascade and social tipping interchangeably. “Cascade” has been used (in the GTPR) recently to refer to the tipping dynamics in different systems affecting each other, both for climate tipping points and positive tipping points. This terminology choice might be confusing for the readers

We agree and note that we plan to improve the clarity and strictness in the use of terms in general to reduce confusion. Regarding this specific point, we will either remove the concurrent use of cascade or change the term to something less confusing such as “contagion”.

Figure 3: Not clear what the x- and y- axes represent.

The y axis represents the probability of the population having a certain threshold fraction which is represented by the x axis. These labels are missing and will be added.

Line 368: 21 papers were discarded because they did not include complex contagion. It would be useful to mention what they included instead.

This is an oversight and will be explicitly stated, but we imply here that they include only simple contagion dynamics.

Table 4: Very useful table summarizing the reviewed studies. To strengthen the conclusions of the manuscript, we need to know more about these studies, though. Could you specify sample size/network size, geographic location, what the nodes and edges are, socioeconomic and demographic background etc.?

We agree, however a lot of these things are left out due to the economy of space on what is already a very full manuscript. We will address this by adding an extended version of the table with important characteristics in a supplementary file. Additionally we will possibly add more columns to Table 4. according to their perceived importance in the literature.

Figure 6: Very informative figure especially because it distinguishes between model-based and empirical studies! The sample size is confusing, though, since the caption says that it includes 87 papers and the text earlier mentioned the eventual # of studies in the review were ~30. Please clarify. It raises the question of what made the empirical studies with a similar lambda around 20% but not tipped (bottom left). This figure could have been improved in a bubble chat format, where the dot size refers to the network size in each study. “Population size” came up as one of the most important drivers in Fig.4, and we do observe its relevance. For instance, the global percentage of the vegetarian population is 20%, but no tipping is observed.

Firstly thank you for the useful suggestion regarding the dot size representing the network size, we will implement this in the figures. The number 87 in Fig 6. Refers to the number of trajectories from the studies in the last stage of the review which were reduced to 12 articles. As you indicate, this is not distinguished properly and will be made clear through specific notation i.e N_p for papers and N_t for trajectories. The small group of empirical studies that did not tip, and in fact ended up with less supporters than at the critical mass illustrates how tipping is highly dependent on the social system and context. We will include an entry in the results which addresses this cluster and explains the factors that may have prevented a successful positive tipping event.

Line 518- Evidence of critical mass: For the empirical studies covered in this paper, it would be very informative for the strength of evidence whether they are in lab settings or in a real-life experimental setting or contain large-scale data?

We agree and this will be added in some form, either in another column of Table 4, an additional table, or a Table entry in the appendix.

Line 531-533 “For modellers…” : Could you expand on “validation across modelling approaches” for computational efficiency? There are examples of system dynamics models which do not use a complex contagion and threshold approach, but show that the inflection point of the norm adoption function (the counterpart of the cdf of the probability of an agent adopting a norm for the fraction of agents who already adopted it) is the most important driver of large scale change in the diet context. How do we for instance cross-validate those? https://www.nature.com/articles/s41893-019-0331-1

As touched on in previous responses, a complex contagion approach is a modeling decision to represent a type of phenomena in this case social contagion or tipping processes. System dynamics models as you mentioned are also a modelling decision to do the same. In this case where the underlying processes are the same, i.e norm adoption, implying social interaction, it is often possible to directly compare representations of dynamics. In the case you mentioned, the norm adoption function is equivalent to the CDF as you already identified. It is in theory possible to construct this CDF even from a distribution of microscopic thresholds (threshold fractions) to obtain one function which governs the adoption rate in a social group undergoing a contagion event (Wiedermann et al., 2020). This would allow one to compare these two things across models. I believe this is beyond the scope of our paper but we will amend this section with more detail on how cross validation can be performed across model types under certain assumptions.

Line 542: Undermining what you did.

This is indeed perhaps underselling the results and will be reformulated to be less severe or removed.

Minor comments:

Line 45 “mechanisms. In which…” please watch the grammar.

This will be amended.

Line 174 Please clarify what “systems-dynamics” is, since it is not a commonly used term.

This will be changed to system dynamics.

Line 338: Which “Ref”?

This reference is missing and will be added in.

I suggest to put table captions above the table, since it is a more commonly used convention.

We will adopt this convention.

Line 497: table x

This will be changed to Table 5.

The first paragraph of Section 5: I suggest to divide this paragraph and several others, since they are too lengthy and contain more than 1 argument, main point etc.

We agree, section 5 and in general throughout the paper will be condensed and more structured, i.e broken up into multiple paragraphs.

References

Berner, R., Gross, T., Kuehn, C., Kurths, J., and Yanchuk, S.: Adaptive Dynamical Networks, https://doi.org/10.48550/arXiv.2304.05652, 12 April 2023.

Efferson, C., Vogt, S., and Fehr, E.: The promise and the peril of using social influence to reverse harmful traditions, Nat. Hum. Behav., 4, 55–68, https://doi.org/10.1038/s41562-019-0768-2, 2020.

Guilbeault, D., Becker, J., and Centola, D.: Complex Contagions: A Decade in Review, in: Complex Spreading Phenomena in Social Systems: Influence and Contagion in Real-World Social Networks, edited by: Lehmann, S. and Ahn, Y.-Y., Springer International Publishing, Cham, 3–25, https://doi.org/10.1007/978-3-319-77332-2_1, 2018.

Sayama, H., Pestov, I., Schmidt, J., Bush, B. J., Wong, C., Yamanoi, J., and Gross, T.: Modeling complex systems with adaptive networks, Comput. Math. Appl., 65, 1645–1664, https://doi.org/10.1016/j.camwa.2012.12.005, 2013.

Smaldino, P.: Modeling Social Behavior: Mathematical and Agent-Based Models of Social Dynamics and Cultural Evolution, Princeton University Press, 360 pp., 2023.

Wiedermann, M., Smith, E. K., Heitzig, J., and Donges, J. F.: A network-based microfoundation of Granovetter's threshold model for social tipping, Sci. Rep., 10, 11202, https://doi.org/10.1038/s41598-020-67102-6, 2020.

Winkelmann, R., Donges, J. F., Smith, E. K., Milkoreit, M., Eder, C., Heitzig, J., Katsanidou, A., Wiedermann, M., Wunderling, N., and Lenton, T. M.: Social tipping processes towards climate action: A conceptual framework, Ecol. Econ., 192, 107242, https://doi.org/10.1016/j.ecolecon.2021.107242, 2022.

(Berner et al., 2023; Guilbeault et al., 2018; Sayama et al., 2013; Smaldino, 2023)

(Wiedermann et al., 2020)