Information content and maximum entropy of compartmental systems in equilibrium

The authors introduce new indices linked to information theory, with application to continuous-time Markov chains in the field of compartmental systems. The examples and applications appear as a mere illustration of the theoretical results.

Compartmental models are not new in Markov chains theory, visited compartments are just visited states, as some of the references point out.
Still, in information theory, as explained in the annex, non absorbing chains are considered, with a likelihood with respect to a classical Borel measure, and the entropy rate is $\lim_T H(X_t, 0<t<T)/T$. This quantity is well-known to inherit the properties of Shannon entropy for random variables.
Here, the Markov chains are absorbing ones, so the usual notion of entropy rate par unit time does not apply. Instead, the authors consider a reference measure induced by all the possible trajectories of the chain before its absorption. They determine the likelihood of the model with respect to this measure, and then compute the Shannon entropy, called H(X).
The new indices are:
H(X)/E T, where T is the time to absorption; this is not a rate per unit time (which Shannon entropy rate is).
H(X)/E N, where N is the number of jumps before absorption.

The properties of Shannon entropy for random variables (convexity, symmetry, etc.) are the basis of the MaxEnt method. These properties are compulsory, allowing for a unique maximum, uniformity, and identifiability. Any new index that is to be used for MaxEnt has to be proven to satisfy these properties. Unfortunately, this would be very difficult to obtain when dividing entropy by the expectation of a quantity depending on the process itself.

The examples could bring some arguments for using these new indices. Unfortunately, nothing helps the reader to understand whether their values confirm some known facts on the behavior of the systems, or if the comments are just comments of the figures in the tables and graphs. Section 3.1 is a clear illustration of the complexity of the maximum: a theoretical higher maximum for a higher number of states tells nothing on the real level of entropy of a given system. Furthermore, line 579 *Usually, entropy is maximized when the system is highly symmetric,* this is not usual by chance but as a direct consequence of the structural definition of entropy. The same is true on the comments in line 592 actually linked to the structural properties of the function x log x.

No mathematical argument is given in the paper to ensure that the following claim (line 336) holds, not even some empirical explanation is given that could justify a conjecture.
While the path entropy measures the uncertainty of the entire path, entropy rates measure the average uncertainty of the instantaneous future of a particle while it is in the system: for the entropy rate per unit time the uncertainty entailed by the infinitesimal future, and for the entropy rate per jump the uncertainty entailed by the next jump.

The discussion and conclusion following the examples show that indeed these conditions are not satisfied. Commenting on the poor results for Example 3.4 that
This example is only supposed to give a first impression of how the maximum entropy principle can be used in combination with entropy rates or path entropy in similar situations. Practical examples usually have a high level of complexity such that existence and uniqueness of a maximum entropy model have to be studied on a case-by-case basis.
is just the opposite of the powerful concept of entropy that is known to fit any situation when handled pertinently, and actually the opposite of any scientific tool.

The level of mathematics and probability necessary for a reader to understand the paper is rather high, especially on Markov theory and information theory. Therefore, many lengthy sections on the very basics of entropy is of no use, and indeed most of the material is used nowhere in the paper. Since the authors are no experts in information theory, a reference to Cover and Thomas, and some definitions closely linked to the paper would be more pertinent than Sections 2.1 and A1, beginning of Section 2.6, and so on. Of course, Kolmogorov, topological or graphical entropies (lines 59-71) have nothing to do here, Markov processes are linked to Shannon entropy from his original paper giving birth to information theory. By the way, the determination of the reference measure is much less classical and could be detailed.

Some local comments:
--The end of Section 1 is repeated at the beginning of Section 2. Section 2 is far too long, mixing definitions, main theoretical results and examples.
--Strangely, right in the middle of the paper, Section 2.7 is devoted to the estimation of the number of compartments, with no link to entropy, even though its title is *Structural model identification via MaxEnt.*
-- Why does MaxCal only appear in the conclusion? It would have been welcome in the introduction, as a method linked to both entropy and Markov chains.

All in all, my advice is to reject the paper, not finalized enough for publication.