Reviewer #2

– Generic NN formulation and relation to RNN: Our (14) is a highly structured neural network (NN) due to the Mori-Zwanzig moment closure (please note the form of $\boldsymbol{f}$ in Eq.(7)). This enables us to learn the diffusion parameter and estimate influence (both for the diffusion network) using the mean-field dynamical system (14). This is different from learning a generic RNN which is not interpretable and does not serve the purpose of learning diffusion parameters.

– Relevance to NeurIPS: Learning diffusion parameters and predicting influence are one of central problems in the ML community. Please see [14-17],[21-23].

– Input is only the initial condition: This is because we need to estimate the influence of source set during testing phase, where no additional temporal data is available. Hence we can only use the source set as input of our network during training and testing. This is different from the standard RNN applications.

– Map to standard RNN/LSTM: Yes, we can map a dynamical system to other types of RNNs. However, our purpose is to derive a structured dynamical system for influence prediction, and need to integrate the diffusion parameters in the dynamical system for training in this work.

– A more precise definition of $\boldsymbol{\varepsilon}$: in the present work, we set it to a generic NN to approximate the MZ memory kernel. We will state this with more details if a revision is allowed.

Reviewer #4

– Splitting the linear and nonlinear part in $\boldsymbol{\varepsilon}$: This is because the linear part is a consequence of the kernel approximation of $\boldsymbol{h}$ in Eq.(10) and $\boldsymbol{K}$ in Theorem 3. In this case, we can interpret the matrices $\boldsymbol{B}$ and $\boldsymbol{C}$ (and impose proper regularization or prior information if possible). The remainder part of $\boldsymbol{\varepsilon}$ is set to generic NN, but not interpretable anymore.

– Connection to PMP: Unlike the standard optimal control and Ref.[37], our formulation does not allow time-varying control variables. This requires a modification of the PMP to interpret our network training. By introducing the total Hamiltonian, we showed that the standard back-propagation is directly equivalent to maximizing the total Hamiltonian. This is different from Ref.[37] which relies on successive approximation (different from back-propagation) and is computationally more expensive,

– Error of NMF decrease as time increase: this is because as time goes, all nodes will eventually be infected (if the diffusion network is connected). Therefore, all methods (not only NMF) will have lower prediction error as time goes to infinity. But NMF appears to be more accurate in predicting the error in early to middle stage of the propagation (rather than asymptotical error), which is of most interests in practice (e.g., one would like to predict the spread of news in three days rather than a month).

– $e(t)$ is not defined: $e(t)$ is defined in Eq.(5) above Theorem 1.

Reviewer #5

– Evaluation of approximation error: please note that it is *infeasible* in practice to evaluate the original MZ memory term, because exact evaluation involves solving an ODE system of size 2^n (n is the size of the diffusion network). Instead, we use MC simulations to approximate ground truth as reference, and compared our results against it, as shown in our experiment part. This has been the standard approach for comparison in the literature.

– Using InfluLearner in influence maximization: Our comparison with InfluLearner on influence prediction demonstrated the significant improvement in accuracy. Therefore its performance cannot exceed NMF in influence maximization which heavily relies on the accuracy of influence prediction. Moreover, the computation complexity of InfluLearner is much higher than NMF, and not suitable for large scale problems and dense networks in influence maximization.

– Pseudo-code: the NMF network structure is given in Eq.(14), which can be solved by standard network training. We will provide more details if a revision is allowed.

– MAE only on 100 source sets: the 100 source sets refers to the testing data, whereas 1,000 refers to the training data.

Reviewer #6

– Examples: the network platform (like Twitter), campaign company, or advertiser are often interested in the influence of individuals (users) on networks—if the individual posts a news/advertisement, then it will be seen and retweeted by his/her followers, and then the followers of them, and so on. Accurate influence estimation proposed in our work helps the campaign company or advertiser to decide which individuals to select, at a certain monetary cost, to spread the news as fast as possible (to maximal amount of people on the network). It also has applications in disease intervention where