Stochastic modeling of cascade dynamics: A unified approach for simple and complex contagions across homogeneous and heterogeneous threshold distributions on networks

Fabian Tschofenig, Daniel Reisinger, Georg Jäger, Marie Lisa Kogler, Raven Adam, and Manfred Füllsack
Department of Environmental Systems Sciences, University of Graz, Graz, Styria, Austria

(Received 1 August 2023; accepted 20 March 2024; published 10 April 2024)

The COVID-19 pandemic has underscored the importance of understanding, forecasting, and avoiding infectious processes, as well as the necessity for understanding the diffusion and acceptance of preventative measures. Simple contagions, like virus transmission, can spread with a single encounter, while complex contagions, such as preventive social measures (e.g., wearing masks, social distancing), may require multiple interactions to propagate. This disparity in transmission mechanisms results in differing contagion rates and contagion patterns between viruses and preventive measures. Furthermore, the dynamics of complex contagions are significantly less understood than those of simple contagions. Stochastic models, integrating inherent variability and randomness, offer a way to elucidate complex contagion dynamics. This paper introduces a stochastic model for both simple and complex contagions and assesses its efficacy against ensemble simulations for homogeneous and heterogeneous threshold configurations. The model provides a unified framework for analyzing both types of contagions, demonstrating promising outcomes across various threshold setups on Erds-Rényi graphs.

DOI: 10.1103/PhysRevE.109.044307

I. INTRODUCTION

Worldwide pandemics have historically had profound consequences on societies, economics, and public health. Understanding the interactions between disease dynamics and human behavior is pivotal for effectively managing infectious diseases. With the emergence of COVID-19, there has been an upsurge in scientific work on the underlying spreading mechanisms [1,2]. Central to this understanding are the fundamental mechanisms driving transmission. In scientific literature, disease transmission is often characterized as simple contagion [3,4], wherein infection can occur through direct contact with just one individual. However, this likelihood of direct biological transmission can be mitigated to some extent through social measures, including mask-wearing, adherence to health guidelines, social interventions, distancing practices, and vaccination. While the implementation of these measures significantly impacts epidemic control, the willingness of the population to engage in these actions is not solely determined by biological factors but also by social processes. The dissemination of beliefs, as highlighted in literature, aligns with the concept of complex contagions, requiring interaction with multiple adopters for transmission to occur. The seminal work of Centola *et al.* [4,5] coined the difference between simple and complex contagions. Simple contagions can spread through a population after a single exposure or contact, resembling the rapid spread of viral diseases or information; the term "simple" denotes the minimal exposure needed for transmission, yet it does not imply that the contagion lacks complexity in its behavior. Conversely, complex contagions typically require multiple exposures or reinforcements from various sources to propagate, commonly observed in the adoption of behaviors or cultural norms. Although the term "complex" signifies the requirement for multiple steps or validations for transmission, it does not suggest that this is the sole factor contributing to the intricacy of spreading phenomena. In the context of complex contagions, reinforcing connections between individuals, who have already adopted the behavior or norm, can increase the likelihood of adoption by those who are still undecided or resistant.

Contagion processes extend beyond diseases or information; they encompass a broad spectrum of phenomena including malware, memes, emotions, and diverse behaviors [6–16]. Moreover, certain biological infections may also adhere to complex contagion patterns, particularly due to coinfections [17]. There is also an intertwining of biological and social contagion mechanisms, such as in the case of the COVID-19 pandemic, which was accompanied by an infodemic [18], deluging society with misinformation, also involving complex contagion mechanisms [19].

Complex contagion dynamics show chaotic behavior due to the existence of feedback loops, network structure, emergent behavior, and individual variability such as adoption thresholds. Considering networks that capture real-world social interactions, they typically feature clustering, in addition to other network characteristics, that make modeling of complex contagions challenging [4,20]. Clustered networks often exhibit high levels of triadic closure (the tendency of nodes to form triangles). Unlike treelike networks where mean-field theoretic approaches are more applicable, the presence of triangles in clustered networks creates complex dependencies among nodes that cannot be adequately captured by traditional mean-field approximations [20]. Indeed, clustering introduces a crucial component for complex contagion processes while at the same time resulting in nonlinear dynamics, thereby complicating the application of simple analytical solutions.

To address the issue of a lack of analytical approaches in complex contagion processes, we present a stochastic model for simple and complex contagions, depending on the following parameters: average degree $\langle k \rangle$ of the graph, initial seed

size $d(0)$, and distinct contagion thresholds T. The evaluation of its validity is benchmarked against various simulations on Erds-Rényi graphs.

II. THRESHOLD MODELS

In 1978, Granovetter [13] proposed the notion of adoption thresholds, arguing that within social contagion processes, these thresholds represent the minimum level of social reinforcement required for an individual to adopt the behavior, while taking into account individual nuances. Based on this concept, in 2002 Watts [21] proposed a model to investigate the structural and threshold conditions that make contagion processes susceptible to global adoption cascades on random networks. Here, agents observe the current binary state (active or inactive) of linked neighbors. If a threshold proportion ϕ of these k neighbors are active, agents adopt the active state; otherwise, they remain inactive. Several analytical approaches have been proposed for such threshold models [5,22], some of which are of stochastic nature [23,24]. Improving upon these models of threshold-based contagion, Centola and Macy [4] proposed the concept of complex contagions, where thresholds T are independent of the node degree k. This modification involves intriguing differences in system dynamics. For example, the Watts model requires for an agent with high node degree a high number of active neighbors to surpass its activation threshold. On the other hand, in the complex contagion model, the activation threshold remains constant, regardless of the node's degree. This results in a lower resistance to activation for high degree nodes. For nodes with a low degree, the scenario is revered. This contrasting behavior in the degree dependence of these two models results in significant differences in spreading paths and contagion dynamics. Consequently, this underscores the need for distinct research endeavors into these threshold concepts.

Contrary to simple contagions, the topological structure of the network has a high impact on the diffusion dynamics of threshold models, including complex contagions, and has been researched intensively [24–31]. Threshold models illuminate the importance of network topology and individual node thresholds, emphasizing the critical role of key nodes and structures like wide bridges [4,32]. Such phenomena have been observed in multiple occasions where even within a connected component it is impossible for a complex contagion to spread from one node to the other [33–35].

The aim of our proposed model is to build on these findings on complex contagion cascade dynamics and provide an analytic approach in form of an iterative stochastic model. The proposed model aims to (i) provide a stochastic approach to contagion cascades, (ii) unify its applicability for both simple and complex contagions,—two theories that are typically studied independently—and (iii) extend the model to heterogeneous threshold systems.

III. SIMPLICIAL CONTAGIONS, HIGHER ORDER SYSTEMS AND HYPERGRAPHS

Simplicial contagions [36], a specialized subset of complex contagions, delve into contagion dynamics through higher-order interactions within groups, moving beyond conventional pairwise node connections. These models, alongside threshold-based contagions, have contributed unique perspectives to our understanding of complex contagion dynamics [37–39].

While threshold models emphasize the critical role of key nodes and structures like wide bridges, simplicial contagions utilize higher-order systems and hypergraphs to intricately model group interactions and collective behaviors, showcasing their significant impact on contagion processes [37]. The adoption of hypergraphs, in particular, facilitates a more nuanced representation of these complex interactions, enhancing our comprehension of contagion dynamics [38]. Notably, higher-order systems and hypergraphs reveal unique types of transitions in contagion spreads that differ from those continuous transitions observed in traditional threshold models, presenting theoretical and practical implications for understanding and managing contagion phenomena in complex networks [39].

Together, these fundamentally different models provide a robust framework for examining complex contagion propagation, encompassing the micro-level details of individual nodes and structures and the macro-level complexities of group dynamics and higher-order interconnections. Furthermore, mean-field approaches for the simplicial contagion model have been achieved to express the temporal evolution of the density of infected nodes for the simplicial contagion model [36]. However, for the common threshold-based complex contagion model examined in this paper, this has not been done. Our proposed stochastic model closes this gap and allows us to approximate $d(t)$ using a stochastic formulation of this threshold-based model.

IV. METHODOLOGY

The proposed model is tested on random graphs with Poisson degree distribution, namely, the Erds-Rényi model [40]. Despite the fact that such randomly generated graphs are not considered to be particularly realistic [41], they can serve as a basis for advancing towards more realistic networks [42,43]. Due to this fact and the stochastic character of the Erds-Rényi model, it makes sense to test our stochastic model on a graph with a similar fundamental concept.

To configure the graph to match real-world network structures, the average degree is $\langle k \rangle \approx 19$, which coincides with the average degree taken from 59 real-world social networks obtained from a large network data repository [44]. The contagion thresholds of the nodes for the homogeneous simulation are in the set of $T_i \in \{1, 2, 3, 4, 5\}$, matching empirical research on various complex contagions [10,45]. However, cascades become progressively difficult to trigger for bigger threshold values [24]. Regarding seeding strategies in terms of partial activation on the network, there are typically two ways this is done for complex contagion simulations. First, by initialization of a selected part of the network, for example, in the neighborhood of one or more preselected nodes [17,32,46,47]. Second, by activation of a randomly selected portion $d(0)$ of the network before simulation start, which is often done for investigation of cascades [23,24,48]. The latter principle is employed in this paper.

First, a stochastic complex contagion model for the case of homogeneous threshold distributions is derived (see Appendix A 3). The model is then expanded to encompass complex contagion systems with heterogeneous threshold distributions (see Appendix A 4). Simulations are carried out to determine the model's effectiveness in both homogeneous and heterogeneous threshold systems.

V. MODEL

We assume networks of size N and average node degree $\langle k\rangle$ and that contagion dynamics evolve over time steps t. Nodes have a binary state (either 0 or 1) which corresponds to an activation density $d(t)$ on the graph. The initial activation density $d(t=0)$ is reflected by a probability that is given by the seeding procedure and can vary from complete activation to complete inactive state of all nodes, $d(0) \in [0, 1]$. The heterogeneous activation thresholds T_i for nodes i are independent of the node link degree, as is typical for complex contagions. Considering the macroscopic parameters and, at the same time, scaling these attributes down to individual nodes, the macroscopic density $d(t)$ can be interpreted as the likelihood that each neighboring node of node i is active by a probability $d(t)$. This means that every node is active with a probability $p \approx d(t)$. According to this assumption for each node's neighborhood, we extent this stochastic approach by defining the probability of finding x active neighbors for a randomly selected node in the graph. Testing nodes for their binary state gives a sequence of outcomes whose respective probabilities are described by Bernoulli trials [49] as binomial distributions. Thus, the problem of determining the probability of exactly x active nodes in a neighborhood of size $\langle k\rangle$ yields

$$\mathrm{Pr}_{\mathrm{binom}}(\langle k\rangle, x, p) = \binom{\langle k\rangle}{x} p^x (1-p)^{\langle k\rangle - x}. \tag{1}$$

According to the concept of complex contagions, every node i in this graph has a threshold T_i which needs to be overcome by the number of active neighbors in its neighborhood. We can extend our binomial model to the question of the probability to find T_i or less active neighbors in the neighborhood of every node given the spreading density $d(t)$ at time step t.

$$\mathrm{Pr}_{\mathrm{binomcdf}}(\langle k\rangle, x, p) = \sum_{x=0}^{T_i} \binom{\langle k\rangle}{x} p^x (1-p)^{\langle k\rangle - x}. \tag{2}$$

From this, we can determine if there are at least T_i active nodes in the neighborhood of every node i, which is equivalent to the cumulative binomial distribution's complementary probability:

$$\mathrm{Pr}_a(x \geqslant T_i | \langle k\rangle, p, T_i) = 1 - \mathrm{Pr}_{\mathrm{binomcdf}}(\langle k\rangle, T_i - 1, p). \tag{3}$$

Notable is the $T_i - 1$ in the sum, which derives from the fact that the cumulative binomial distribution is for $x \leqslant T_i$, necessitating $T_i - 1$ to include T_i in the complementary probability.

A. Homogeneous thresholds

For the homogeneous case $T_i = T\ \forall i \in G$, to calculate the spreading density $d(t+1)$ for the next time step $t+1$, we use Eq. (3) to formulate a change rate for spreading density, such that

$$d(t+1) = d(t) + (1 - d(t))\mathrm{Pr}_a(x \geqslant T | \langle k\rangle, d(t), T). \tag{4}$$

Without the implementation of a recovery from active to inactive states in the model, the first part $d(t)$ represents all nodes that are activated, which will further accumulate for $t+1$. In the latter part of the equation, the share $(1 - d(t))$ of inactive nodes are considered. This means that every inactive node in the neighborhood of an inactive node has at least one inactive neighbor. Instead of having at least T active neighbors in $\langle k\rangle$, we have T active neighbors in $\langle k\rangle - 1$.

This has the effect of decreasing the probability of finding a sufficient number of active neighbors in the neighborhood, resulting in

$$d(t+1) = d(t) + (1 - d(t))\mathrm{Pr}_a(x \geqslant T) \tag{5}$$

by usage of the short form

$$\mathrm{Pr}_a(x \geqslant T) = \mathrm{Pr}_a(x \geqslant T | \langle k\rangle - 1, d(t), T). \tag{6}$$

Finally, if the condition of a sufficient number of active neighbors is met, the probability for activation is instantiated to adhere to typical contagion process models. This instantiation is intended to match the formulation of discrete susceptible-infected (SI) models, particularly in the case of a simple contagion where $T = 1$ [50]. For this, the transmission probability $\beta \in (0, 1]$ is introduced, resulting in the final form of the underlying equation for the homogeneous threshold distribution case:

$$d(t+1) = d(t) + (1 - d(t))\mathrm{Pr}_a(x \geqslant T)\beta. \tag{7}$$

Note that the interval on the lower bound is closed because $\beta = 0$ would prevent contagion processes.

B. Heterogoneous thresholds

Additionally, heterogeneous complex contagion systems are also supported by the proposed model. To clarify what is meant by heterogeneous threshold configurations, let A be the set of all nodes in G. Let B_T be the subsets of A, such that every subset B_T contains all nodes i in A with threshold $T_i = T$. Consequently each node can only be an element in a single subset B_T, therefore the number of nodes $N = \sum_{B_T \forall T \in \{1,2,3,4,5\}} |B_T|$ is conserved.

This model can be described as follows. First, $d(t)$ must be distinct for each subset B_T, denoted as $d^T(t)$. Consequently, $d^T(t)$ is the spreading density in each subset B_T. This term replaces $d(t)$ in the initial parts of the equation; however, the macroscopic density is still dependent on all activity states in the system and not just within each subset of nodes with the same T. The mean spreading density in the entire graph is therefore calculated by averaging the densities $d^T(T)$ weighted $|B_T|$, which represents the amount of nodes in each subset. This holds

$$d^T(t+1) = d^T(t) + (1 - d^T(t))Pr_a(x \geqslant T | \langle k\rangle - 1, d(t), T)\beta \tag{8}$$

$$d(t) = \sum_{T \in G} d^T(t) \frac{|B_T|}{N}. \tag{9}$$

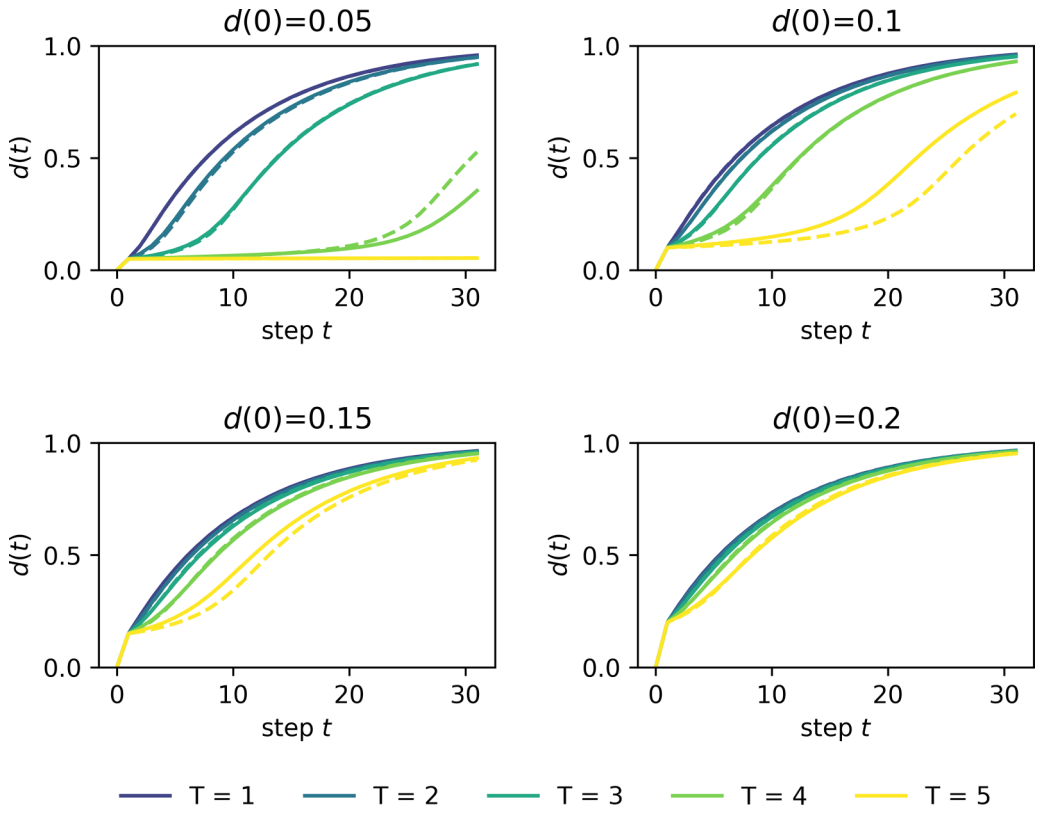

FIG. 1. Results for the time evolution of the spreading density $d(t)$ for different seed sizes $d(0)$ for homogeneous threshold configurations $T \in \{1, 2, 3, 4, 5\}$ with the solid line representing the averages over the ensemble simulations and the dashed line representing the result of the stochastic model.

The average density of active nodes $d(t)$ is computed as a weighted sum over all possible contagion thresholds T in the set G. For each threshold T, the density of active nodes $d^T(t)$ is multiplied by the proportion of nodes $\frac{|B_T|}{N}$, which have the respective threshold. Summing these products for all T yields $d(t)$. Thus, $d(t)$ represents the ratio of active nodes to the total nodes in the graph, whereas $d^T(t)$ is similar to $d(t)$ but focuses solely on the subset of nodes with the corresponding threshold T. In the case that the threshold distribution is homogeneous, Eqs. (8) and (9) become essentially equal to Eq. (7). This means that the heterogeneous model is applicable to configurations of homogeneous distributions as well, as the homogeneous case is essentially a heterogeneous case with one and only one distinct subset $B_T \subseteq A$ and $A \subseteq B_T$.

VI. RESULTS AND DISCUSSION

We show ensemble simulation setups on Erds-Rényi graphs with $N = 5000$ and $p = \frac{k}{N-1}$ resulting in an average degree $\langle k \rangle \approx k$ and compare the ensemble means to the iterative results of our proposed stochastic model. The ensemble size is 50; for each simulation, an Erds-Rényi graph is generated and every node is assigned a random threshold T_i. The threshold distributions are given by the simulation scenario. After graph generation, the activity states are initialized by a seeding procedure which selects $d(0)N$ nodes randomly and sets them to an active state. Other nodes occupy the inactive state. The simulation runs until time step $t_{\max} = 30$ and allows for diffusion of active states according to the complex contagion mechanism of Sec. II. The transmission probability β is set to 0.2, which is correlated to reasonable transmission velocity and transient times until saturation. Moreover, a too small transmission probability may result in higher numeric errors due to the iterative nature of our proposed analytic model, as described in Sec. VI C. The higher this parameter, the faster the contagion spreads throughout the network. The lower the transmission probability is, the slower it spreads, resulting in more steps, therefore increasing the propagated error in the analytic model in comparison to the simulation runs.

A. Homogeneous threshold systems

Figure 1 presents the results for the simulations (solid lines) as well as for the analytic solutions (dashed lines) for different seed sizes $d(0)$. For each of these different seed sizes, homogeneous systems with $T \in \{1, 2, 3, 4, 5\}$ are initialized and simulated 50 times, i.e,. all nodes $i \in G$ have the same threshold $T_i = T$ as stated in Appendix A 3. Therefore, each line in the graph visualizes the average of the 50 simulation runs for each T, where every node in the graph has the same threshold T. Figure 1 illustrates that our model intends to provide a solution for both simple, i.e., $T = 1$, and complex contagions, i.e., $T \geqslant 2$. For simple contagions and complex contagions with low thresholds, the stochastic model results correspond almost perfectly with the simulation results. For complex contagions with higher thresholds, the stochastic and simulation results overall show a good match and only deviate

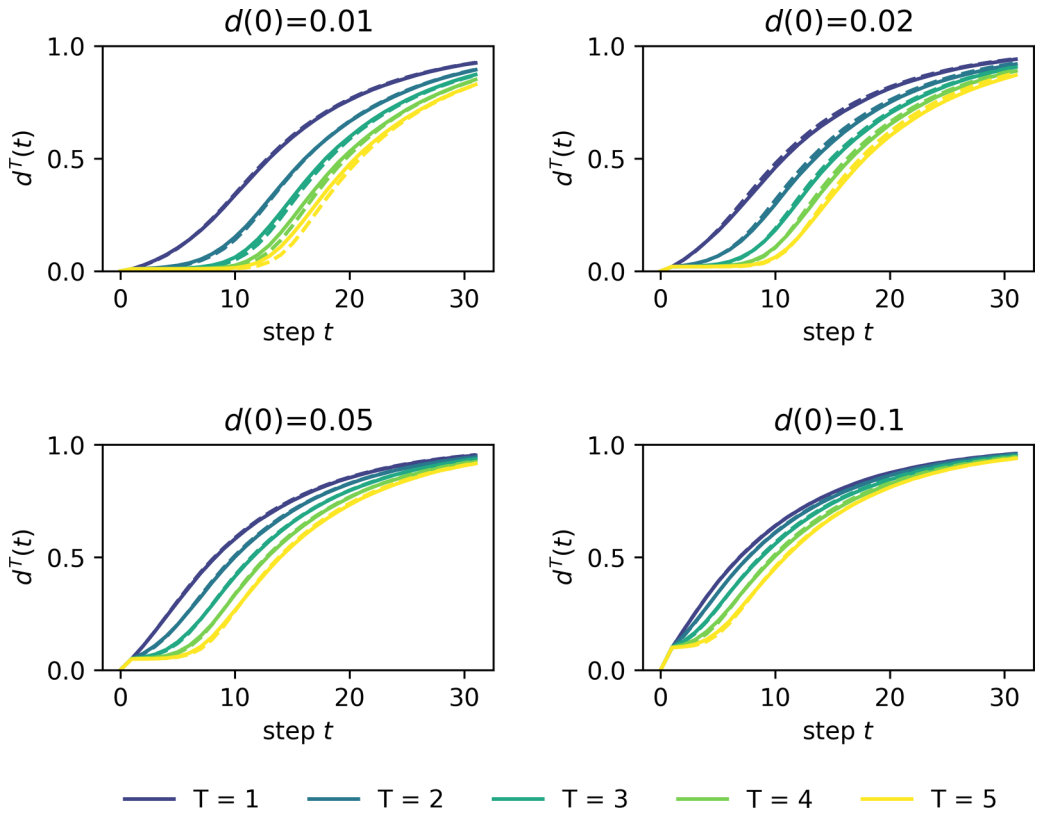

FIG. 2. Results for the time evolution of the spreading densities $d^T(t)$ of each equally sized subset B_T with thresholds $T \in \{1, 2, 3, 4, 5\}$ (heterogeneous case) starting from initial seed sizes $d(0) \in \{0.01, 0.02, 0.05, 0.1\}$ with the solid line being the averages over the multiple simulation runs and the dashed line the result of the stochastic model. Disappearing dashed lines correspond to a perfect match between the simulation runs and the stochastic model.

in the case of high T and low $d(0)$. This deviation arises from the fact that initiating cascades becomes increasingly difficult at higher thresholds due to the decreased likelihood of sufficient active neighbors during the early stages of the spreading dynamics. This is consistent with research on relative threshold systems [24]. Nonetheless, we argue that this effect diminishes with increasing network size, as the significance of spread-inhibiting node configurations of initially activated nodes becomes less and less relevant.

B. Heterogeneous threshold systems

For the heterogeneous threshold simulation configuration, the system is initialized with equally distributed thresholds $T \in \{1, 2, 3, 4, 5\}$, where 20% of the total nodes possess a threshold of $T = 1$; another 20% have a threshold of $T = 2$. This pattern continues for all nodes and their respective thresholds result in equally sized subsets B_T (see Appendix A 4). Again, 50 simulations are executed and plotted against the stochastic model's solution. The results of the comparison are shown in Fig. 2. Here, unlike the visualizations from Fig. 1, the lines represent the time evolution of the subset specific spreading density $d^T(t)$ simulation runs, as opposed to each subgroup having its own simulation runs. Results for the heterogeneous case also show a near-perfect fit. The simulations demonstrate an excellent correlation with the stochastic results, with the correlation increasing as the initial seed size $d(0)$ increases, as can be seen in Fig. 2. For the case in which $d(0) = 0.02$, it can be observed that the simulations are, on average, slightly below stochastic results. This underperformance can be attributed to the local network structure, which may mitigate spreading, as the stochastic model assumes that every active node in the graph possesses complete spreading potential. However, active nodes tend to cluster together because a node in the neighborhood of an active node is more likely to change state under otherwise identical conditions than one in the neighborhood of an inactive node. This leads to a local clustering pattern and a tendency for an excess of active nodes in a given neighborhood compared to the minimum required for the activation of a particular node i with threshold T_i. As a consequence of this excessive activation, clustered active nodes fail to trigger subsequent cascades to their maximum capacity. Furthermore, the clustering hypothesis is challenged by the simulation's tendency to perform better as $d(0)$ increases, thereby reducing the significance of clustering effects. Notably, the analytic solution not only provides highly accurate approximations despite being an iterative process, but also closely mirrors the simulation curves displaying the spreading densities $d^T(t)$ in the subsets B_T with similar T for all investigated seed sizes.

C. Error classification

In the real world, heterogeneous network attributes are of high relevance, therefore the following investigation of the estimation of the accuracy of the model is carried out on

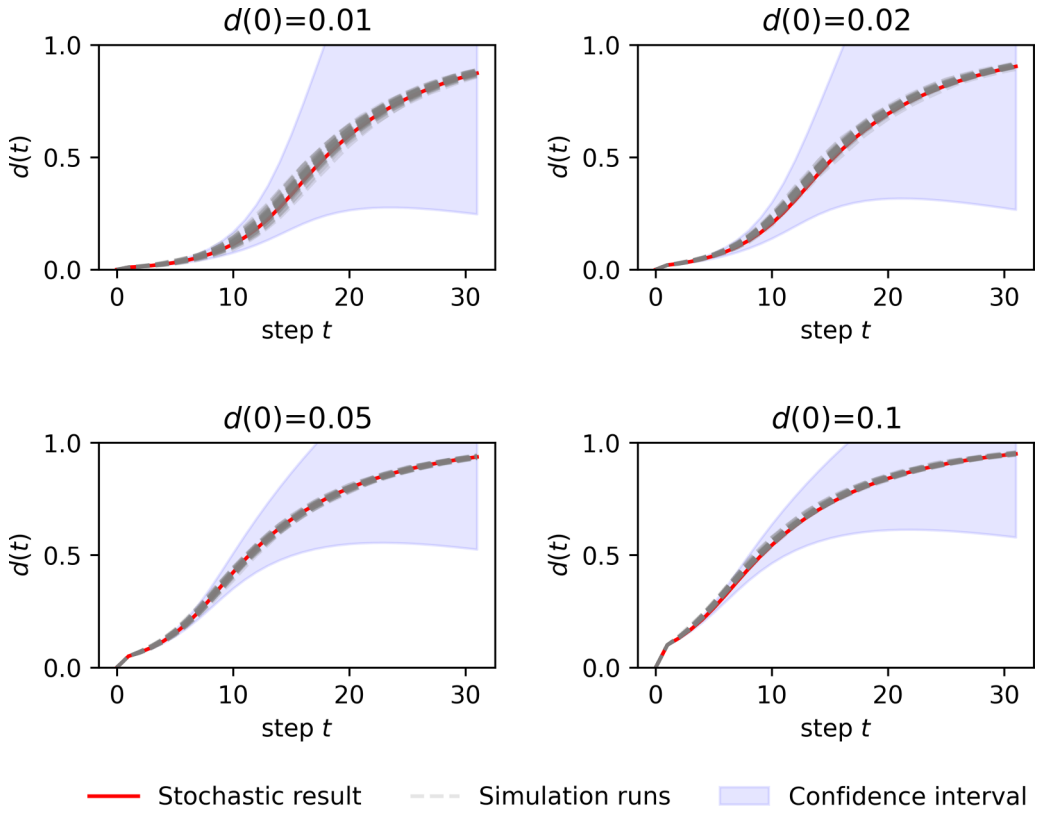

FIG. 3. Results for the time evolution of the spreading density $d(t)$ for different seed sizes $d(0)$ for the heterogeneous threshold configuration with equal amount of $T \in \{1, 2, 3, 4, 5\}$ with the dashed gray lines being the average spreading density over multiple simulation runs and the red line the result of the stochastic model. The blue area marks the confidence intervals around the stochastic solution determined via Eqs. (10) and (11).

the heterogeneous threshold simulations from Sec. VI B. Due to the iterative nature of the proposed model, each step is dependent on all preceding steps; consequently, errors may accumulate. To estimate the upper and lower error boundaries, 1000 simulations for the heterogeneous simulations initial conditions are conducted. Starting with the seed step and moving to the first step, we assume that the simulation must be as stochastic as possible due to its random initialization. For all subsequent steps after initialization, node activations adhere to the rules outlined in Sec. I and Appendix A 2. Consequently, it makes sense to identify the error in the initial step and then extrapolate the upper and lower limits in all subsequent steps based on it. The average relative error in the first step between the stochastic result and the simulations is $\Delta d(0)$. It is determined by running 1000 simulations for all $d(0)$ used in the heterogeneous simulation and calculating its average absolute difference to the stochastic solution after a single step. From this, the upper boundary $d^{\text{upper}}(t)$ as well as the lower boundary $d^{\text{lower}}(t)$ in step t can be estimated via

$$d^{\text{upper}}(t) = d(t)(1 + \Delta d(0))^t, \tag{10}$$

$$d^{\text{lower}}(t) = d(t)(1 - \Delta d(0))^t. \tag{11}$$

In Fig. 3, the heterogeneous simulation from Fig. 2 is displayed by showing the mean spreading density $d(t)$ against the ensemble simulation runs. The area between the upper and lower boundaries is colored in blue. All simulations within this area are within the expected boundaries of error propagation originating from the single step error $\Delta d(0)$. As shown in Fig. 3, simulation runs increasingly deviate from one another, the smaller the seed size $d(0)$ in the beginning. However, the error in a single step between simulation and analytic solution and therefore the upper and lower boundaries are very similar and independent from the seed size. This indicates that the model performs really well besides a small numerical error from step to step.

D. Comparison to SI models

Both SI models and complex contagion models function as binary state contagion models, addressing the dynamics of how active nodes influence and activate their connected inactive neighbors.

In traditional SI models, the probability that a node i gets activated by its connection to an infectious node j is represented as $P_{i\to j} = \beta$. Contrarily, in complex contagion models, the activation of node i depends on whether it is connected to an adequate number of active neighbors. In the case of $T = 1$, a simple contagion is present and the contagion is comparable to discrete SI models. Hence, to replicate this similarity in our model, we implement the transmission probability β.

This incorporation essentially aligns our model closely with the classical discrete SI model, but with an added nuance: it accounts for the probability of having an adequate number of active neighbors and can therefore be considered as a form of SI model implementing the conditions of complex

contagions. This nuance elucidates why the presented stochastic results resemble SI models.

E. Comparison of simplicial contagions, hypergraphs, and higher order systems

The model described in this paper deviates significantly from simplicial contagion models, foremost due to the fact that it depends on node-level activation thresholds, which simplicial approaches do not include. Each node is given a threshold T, which specifies the minimal number of active neighbors needed for activation. This threshold mechanism introduces a deterministic contagion spread pathway that is highly dependent on the topology of the network and the initial dispersion of active nodes. Higher threshold contagions have been shown to require extensive connectivity or wide bridges to spread, highlighting the model's reliance on particular structural configurations and the unreachability of some nodes in the network [32–35]. On the contrary, simplicial contagion models do not incorporate the individual threshold criterion; rather, they emphasize contagion through group interactions within simplices [36]. These models prioritize the collective behavior of node groups over the discrete state of individual nodes, capturing dynamics where contagion transmission is governed by the collective state of multiple interacting nodes rather than meeting individual node thresholds. The absence of a node-level threshold in simplicial models permits a more fluid contagion, emphasizing the role of group interactions and the probabilistic nature of contagion transmission within complex network structures. Originating from these group interactions, hypergraphs offer a viable way of modeling such higher-order interactions [38]. Hypergraphs extend traditional graph theory by allowing edges to connect more than two nodes, thus encapsulating the higher-order interactions intrinsic to simplicial models. These interactions, represented as hyperedges in a hypergraph, embody the group dynamics critical to simplicial contagion models, where contagion spread is influenced by the collective state of nodes within these hyperedges. The absence of node-level thresholds facilitates a contagion spread that is less constrained by individual node states and more influenced by the overall configuration and connectivity of nodes within so-called simplices [36]. The distinction between the threshold-based model discussed here and the simplicial contagion models, particularly in the context of hypergraphs, highlights the diverse methodologies employed to understand contagion dynamics. While the threshold model emphasizes the role of individual node thresholds and network topologies like wide bridges, simplicial models, and their hypergraph counterparts focus on the collective behavior of groups of nodes, offering a broader perspective on complex contagion spread mechanisms within networks.

VII. CONCLUSIONS

In conclusion, the proposed stochastic approach yields promising results for both simple and complex contagion systems on homogeneous and heterogeneous threshold configurations applied on Erds-Rényi graphs. Therefore, this approach provides a unified model for both simple and complex contagion systems. The accuracy of the analytical results improves with larger graph sizes as well as bigger seed sizes.

However, due to the fact that the model is of discrete nature in terms of $\langle k \rangle$, it means that $\langle k \rangle$ has to be an integer number. As $\langle k \rangle$ is the average degree of the graph, it is most likely not an integer but rather a floating point number, so it needs to be either ceiled to the next higher integer or floored to the next lower one, which already induces inaccuracies in the starting conditions of the model as well as for every consecutive step. Further research into solving the problem via transitioning to a continuous model or by approximating the intermediate points between two integers via polynomial interpolation could show another improvement in the model performance.

A significant benefit of the stochastic model is its lesser susceptibility to the graph size in terms of its computational complexity. Furthermore, we expect that the larger the graph, the better the match between the simulation results and the analytic solutions due to the law of large numbers.

A major limitation of the proposed model is that the investigated Erds-Rényi graph with its Poisson degree distribution does not capture many network properties which are typically observed in real social networks. As many real-world graphs show a power-law degree distribution, we aim to test the proposed model on Barabási-Albert (BA) graphs in future research, which are also a form of stochastic graphs, however, they follow a power-law degree distribution and its generation is close to real world networks via its preferential attachment concept. We anticipate that our stochastic model will exhibit satisfactory accuracy when applied to graphs with a wider and more heterogeneous degree distribution. We assume that the incorporation of a correction factor to accommodate topological configurations of BA networks could be necessary, especially considering the significant deviation of the node degree from the network's average degree compared to Erds-Rényi graphs. Building on this investigation, benchmarking the model on real world networks would then also be a reasonable next step.

Furthermore, extending our proposed model to a SIR model holds considerable promise for further research, given its substantial relevance in various contagion processes [51].

ACKNOWLEDGMENTS

We would like to thank the University of Graz for their financial support.

APPENDIX: DEFINITIONS

1. Simple contagions

A simple contagion denotes a spreading mechanism where an encounter with an infected node is adequate for contagion transmission to a susceptible node. Formally, this can be described by

$$P_{i\to j} = \beta,$$

where

(1) $P_{i\to j}$ denotes the likelihood of node i transmitting the contagion to node j.
(2) β stands for the transmission likelihood.

2. Complex contagions

In contrast, a complex contagion necessitates to be in contact with multiple active or infected neighbors to also become active:

$$P_{i\ \text{activates}} = \begin{cases} \beta & \text{for } n_{\text{active neighbors}} \geqslant T_i \\ 0 & \text{otherwise.} \end{cases}$$

While in simple contagions, only a fraction β of interactions with an active node instigates an infection, this must also be applied to complex contagions for uniformity. A consistent approach is essential as a complex contagion with a threshold $T = 1$ is fundamentally analogous to a simple contagion. To circumvent the case distinction, the likelihood of satisfying the constraint $n_{\text{active neighbors}} \geqslant T_i$ can be introduced as $\Pr_a(i)$, leading to

$$P_{i\ \text{activates}} = \Pr_a(i)\beta,$$

where

(1) $P_{i\ \text{activates}}$ represents the likelihood of node i becoming active in the next step.
(2) $n_{\text{active neighbors}}$ indicates the count of active neighbors of node i.
(3) T is the activation threshold, marking the necessary count of active neighbors for node i activation.
(4) $\Pr_a(i)$ is the likelihood of node i having an adequate number of active neighbors $n_{\text{active neighbors}} \geqslant T_i$.
(5) β denotes the transmission likelihood.

3. Homogeneous thresholds

For a network with uniform thresholds:

$$T_i = T \quad \text{for all} \quad i \in G.$$

Here, T_i is node i's complex contagion threshold, T is a fixed value, and i denotes each network node. Essentially, all nodes i in the network G share an identical complex contagion threshold T.

4. Heterogeneous thresholds

For a network with diverse thresholds,

$$T_i \neq T_j \quad \text{for certain} \quad i, j \in G,$$

where

(1) T_i and T_j signify the complex contagion thresholds of nodes i and j correspondingly.
(2) Nodes i and j represent distinct nodes in the network.

This implies that the network contains at least two nodes, i and j, with differing complex contagion thresholds, T_i and T_j.

[1] C. M. Ferreira, M. J. Sá, J. G. Martins, and S. Serpa, The COVID-19 contagion–pandemic dyad: A view from social sciences, Societies **10**, 77 (2020).

[2] B. Murgante, G. Balletto, G. Borruso, G. Las Casas, P. Castiglia, and M. Dettori, Geographical analyses of COVID-19's spreading contagion in the challenge of global health risks, Mobil. Environ. **17**, 283 (2020).

[3] B. Min and M. San Miguel, Competing contagion processes: Complex contagion triggered by simple contagion, Sci. Rep. **8**, 10422 (2018).

[4] D. Centola and M. Macy, Complex contagions and the weakness of long ties, Am. J. Soci. **113**, 702 (2007).

[5] D. Centola, V. M. Eguíluz, and M. W. Macy, Cascade dynamics of complex propagation, Physica A **374**, 449 (2007).

[6] A.-L. Barabási, *Network Science* (Cambridge University Press, Cambridge, England, 2016).

[7] E. L. Corbett, C. J. Watt, N. Walker, D. Maher, B. G. Williams, M. C. Raviglione, and C. Dye, The growing burden of tuberculosis: Global trends and interactions with the HIV epidemic, Arch. Intern. Med. **163**, 1009 (2003).

[8] D. Balcan, H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J. J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van den Broeck, V. Colizza, and A. Vespignani, Seasonal transmission potential and activity peaks of the new influenza A(H1N1): A Monte Carlo likelihood analysis based on human mobility, BMC Med. **7**, 45 (2009).

[9] L. Hufnagel, D. Brockmann, and T. Geisel, Forecast and control of epidemics in a globalized world, Proc. Natl. Acad. Sci. USA **101**, 15124 (2004).

[10] B. State and L. Adamic, The diffusion of support in an online social movement: Evidence from the adoption of equal-sign profile pictures, in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Association for Computing Machinery New York, NY, US, 2015), pp. 1741–1750.

[11] B. Mønsted, P. Sapieżyński, E. Ferrara, and S. Lehmann, Evidence of complex contagion of information in social media: An experiment using Twitter bots, PloS one **12**, e0184148 (2017).

[12] M. S. Granovetter, The strength of weak ties, Am. J. Soc. **78**, 1360 (1973).

[13] M. Granovetter, Threshold models of collective behavior, Am. J. Soc. **83**, 1420 (1978).

[14] N. A. Christakis and J. H. Fowler, The spread of obesity in a large social network over 32 years, New Engl. J. Med. **357**, 370 (2007).

[15] J. Zhang and D. Centola, Social networks and health: New developments in diffusion, online and offline, Ann. Rev. Soc. **45**, 91 (2019).

[16] D. Centola, *Change: How to Make Big Things Happen* (Little, Brown Spark, New York City, 2021).

[17] D. Guilbeault, J. Becker, and D. Centola, Complex contagions: A decade in review, *Complex Spreading Phenomena in Social Systems: Influence and Contagion in Real-World Social Networks*, edited by S. Lehmann and Y. Y. Ahn (Computational Social Sciences. Springer, Cham, 2018), pp. 3–25.

[18] D. J. Rothkopf, When the buzz bites back, The Washington Post, **11**, B1 (2003).

[19] P. Törnberg, Echo chambers and viral misinformation: Modeling fake news as complex contagion, PLoS one **13**, e0203958 (2018).

[20] D. J. P. O'Sullivan, G. J. O'Keeffe, P. G. Fennell, and J. P. Gleeson, Mathematical modeling of complex contagion on clustered networks, Front. Phys. **3**, 71 (2015).

[21] D. J. Watts, A simple model of global cascades on random networks, Proc. Natl. Acad. Sci. USA **99**, 5766 (2002).
[22] F. Karimi and P. Holme, Threshold model of cascades in empirical temporal networks, Physica A **392**, 3476 (2013).
[23] P. D. Karampourniotis, S. Sreenivasan, B. K. Szymanski, and G. Korniss, The impact of heterogeneous thresholds on social contagion with multiple initiators, PloS one **10**, e0143020 (2015).
[24] J. P. Gleeson and D. J. Cahalane, Seed size strongly affects cascades on random networks, Phys. Rev. E **75**, 056103 (2007).
[25] V. Barash, C. Cameron, and M. Macy, Critical phenomena in complex contagions, Soc. Networks **34**, 451 (2012).
[26] P. S. Dodds and J. L. Payne, Analysis of a threshold model of social contagion on degree-correlated networks, Phys. Rev. E **79**, 066115 (2009).
[27] J. P. Gleeson, Cascades on correlated and modular random networks, Phys. Rev. E **77**, 046117 (2008).
[28] P. Gai and S. Kapadia, Contagion in financial networks, Proc. R. Soc. A **466**, 2401 (2010).
[29] A. Galstyan and P. Cohen, Cascading dynamics in modular networks, Phys. Rev. E **75**, 036109 (2007).
[30] A. Hackett, S. Melnik, and J. P. Gleeson, Cascades on a class of clustered random networks, Phys. Rev. E **83**, 056107 (2011).
[31] Y. Ikeda, T. Hasegawa, and K. Nemoto, Cascade dynamics on clustered network, in *Journal of Physics: Conference Series* (IOP Publishing, Bristol, England, 2010), Vol. 221, p. 012005.
[32] D. Guilbeault and D. Centola, Topological measures for identifying and predicting the spread of complex contagions, Nat. Commun. **12**, 4430 (2021).
[33] D. Centola, The spread of behavior in an online social network experiment, Science **329**, 1194 (2010).
[34] D. Centola, An experimental study of homophily in the adoption of health behavior, Science **334**, 1269 (2011).
[35] D. Centola, *How Behavior Spreads: The Science of Complex Contagions* (Princeton University Press, Princeton, NJ, 2018), Vol. 3.
[36] I. Iacopini, G. Petri, A. Barrat, and V. Latora, Simplicial models of social contagion, Nat. Commun. **10**, 2485 (2019).
[37] Y. Chen, Y. R. Gel, M. V. Marathe, and H. V. Poor, A simplicial epidemic model for COVID-19 spread analysis, Proc. Natl. Acad. Sci. USA **121**, e2313171120 (2023).
[38] G. F. de Arruda, G. Petri, and Y. Moreno, Social contagion models on hypergraphs, Phys. Rev. Res. **2**, 023032 (2020).
[39] G. Ferraz de Arruda, G. Petri, P. M. Rodriguez, and Y. Moreno, Multistability, intermittency, and hybrid transitions in social contagion models on hypergraphs, Nat. Commun. **14**, 1375 (2023).
[40] P. Erdös and A. Rényi, On random graphs I, Publ. Math. Debrecen **6**, 18 (1959).
[41] S. H. Strogatz, Exploring complex networks, Nature (London) **410**, 268 (2001).
[42] R. Albert, H. Jeong, and A.-L. Barabási, Error and attack tolerance of complex networks, Nature (London) **406**, 378 (2000).
[43] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Network robustness and fragility: Percolation on random graphs, Phys. Rev. Lett. **85**, 5468 (2000).
[44] R. Rossi and N. Ahmed, The network data repository with interactive graph analytics and visualization, Proc. AAAI Conf. Artif. Intell. **29**, 4292 (2015).
[45] V. Barash and J. Kelly, Salience vs. commitment: Dynamics of political hashtags in Russian Twitter, Berkman Center Research Publication No. 2012-9, Available at SSRN https://ssrn.com/abstract=2034506.
[46] D. Kempe, J. Kleinberg, and É. Tardos, Maximizing the spread of influence through a social network, in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery New York, NY, US, 2003), pp. 137–146.
[47] Y. Youm, B. Lee, and J. Kim, A measure of centrality in cyclic diffusion processes: Walk-betweenness, PloS One **16**, e0245476 (2021).
[48] S. Melnik, J. A. Ward, J. P. Gleeson, and M. A. Porter, Multi-stage complex contagions, Chaos **23**, 013124 (2013).
[49] A. Papoulis, Bernoulli Trials: Probability, Random Variables, and Stochastic Processes (McGraw-Hill, New York, 1984), Vol. 57.
[50] A. L. Hill, D. G. Rand, M. A. Nowak, and N. A. Christakis, Infectious disease modeling of social contagion in networks, PLoS Comput. Biol. **6**, e1000968 (2010).
[51] C. Ji and D. Jiang, Threshold behaviour of a stochastic SIR model, Appl. Math. Model. **38**, 5067 (2014).