

A Vector Threshold Model for the Simultaneous Spread of Correlated Influence

Yong Zhuang and Osman Yağın

Department of Electrical and Computer Engineering and CyLab,
Carnegie Mellon University, Pittsburgh, PA, 15213 USA
{yongzhua, oyagan}@andrew.cmu.edu

Abstract—Spread of influence is one of the most widely studied propagation processes in the literature on complex networks. Examples include the rise of collective action to join a riot and diffusion of beliefs, norms, and cultural fads, to name a few. Most existing works on modeling influence propagation consider a single *content* (e.g., an opinion, decision, product, political view, etc.) spreading over a network independent from everything else. However, most real-life examples involve multiple correlated contents spreading simultaneously and exhibiting positive (e.g., opinions on same-sex marriage and gun control) or negative (e.g., opinions on universal health care and tax-relief for the “rich”) correlation. To accommodate these cases, this paper proposes the *vector threshold model*, as an extension of the widely used Watts threshold model for complex contagions. Here, the *state* of a node is represented by a binary vector representing their opinion on a number of content items. Nodes switch their states based on the *influence* they receive from their neighbors in the network. The *influence* is represented by a vector containing the proportion of neighbors who support each content; both positively and negatively correlated contents can be captured in this formulation by using different rules for switching node states. Our main result is concerned with the expected size of *global cascades*, i.e., cases where a randomly chosen node can initiate a propagation that eventually reaches a positive fraction of the whole population. We also derive conditions on network structure for global cascades to be possible. Analytic results are supported by a numerical study.

Index Terms—Influence Propagation; Correlated Opinion Spread; Social Networks.

I. INTRODUCTION

In recent decades, mathematical modeling of propagation processes over networks have been studied in a wide range of contexts including cascading failures [1], [2], epidemics and social contagions [3]–[7], systemic risk in banking networks [8], to name a few. Contagion processes are typically modeled and studied under two different categories referred to as *simple* and *complex* contagions, respectively [9]. Simple contagion models are used for cases where a single source of exposure is enough for an individual to get *infected* and to start spreading the content to their contacts; e.g., news articles, disease spreading, etc. On the other hand, complex contagion models are used in cases where social reinforcement plays a key role in the spreading process. In other words, complex contagion models are used when multiple sources of exposure to a content (e.g., an opinion, a product, a political view, etc.) are needed for individuals to change their action or state. For example, an individual may not adopt a new behavior (or,

change their opinion) after seeing only one friend doing so, but the situation might change if the ratio of their contacts doing it exceeds a certain level.

This paper focuses on complex contagions for which several models have been proposed in the literature. Perhaps the most widely known among them is the linear threshold model proposed by Watts in [10]. In this model, each node belongs to one of two states, *inactive* or *active*, and has a threshold τ in $(0, 1]$ which is drawn from a distribution $P(\tau)$. This threshold indicates the required fraction of active neighbors for an inactive node to turn active. Starting from a state where all nodes are initially inactive, a small number of nodes are chosen uniformly at random and made active. Then, an inactive node with degree d of which m are active will get activated with probability

$$F[m, d] \triangleq P\left[\frac{m}{d} > \tau\right], \quad (1)$$

where $F[m, d]$ is referred to as the *response function*. This model allows studying the propagation of *binary* influence (where individuals are either active or inactive) over *monoplex* networks (where there exists only one link type). Under these assumptions, this model overly simplifies the dynamics for link relationships and possible states of each node. With this motivation several works have proposed and studied models that involve a richer set of node states and link types. For instance, Yağın and Gligor [11] extended this vanilla threshold model to *multiplex* networks in order to incorporate the fact that there may exist more than one type of edges in the network; e.g., friendship, colleagueship, family, etc. With this observation, they proposed a content-dependent multiplex threshold model. In this model, the network consists of r different link types with each link type having certain influence *weight* that varies by the content that is being spread. For instance, their models make it possible to capture the fact that video games might be more likely to spread among high-school friends rather than parents, while the opposite might be true for political ideas. From a different perspective, Melnik et al. [12] extended the vanilla linear threshold model to *multi-stage* (i.e., non-binary) influence propagation. In their model, nodes can belong to a richer a set of states, i.e., inactive, active, hyper-active, etc.

The common aspect of [11] of [12] is the assumption that there is only one content being spread in network. Alternatively, Borodin et al. [13] proposed a threshold model for

competitive influence. If there are two contents, A and B , spreading over networks, then nodes satisfying the condition of supporting A (resp. B) will turn to the state of supporting A (resp. B). If nodes satisfy both of the conditions for supporting A and B , then the nodes will randomly choose a content to support. Although there are multiple contents in the propagation process, there is limited correlation between the contents arising from the specific constraint that nodes can not support A and B simultaneously. Aside from this, the spreading of one content has little impact on the spreading of the other. Additionally, even though some researchers studied the spread of correlated contents (e.g., competition, cooperation, etc.), they focused on simple contagions rather than complex contagions [14].

The discussion given above indicates that most works on complex contagions studied the spreading of either a single content (e.g., opinion, rumor, product, political view, etc.), or multiple contents spreading simultaneously but with *limited* correlation between each other. However, real-life influence propagation processes can have multiple contents spreading simultaneously with positive or negative correlation, and individuals might *support* more than one content at the same time. For example, the spread of the purchase behaviors of different products from the same company might exhibit positive correlation; e.g., an individual who already purchased an iPhone might be more easily influenced by their friends to buy the Apple HomePod. In contrast, one's opinions on universal health care and proposed tax relief for "wealthy" individuals would be expected to have *negative* correlation.

With this motivation, this work aims to initiate a study on complex contagions where multiple contents spread simultaneously in a correlated manner. To this end, we propose an extension of the Watts threshold model [10] allowing correlated contents to spread at the same time, referred to as the *the vector threshold model*; see Section II-B for description of the model. Under this model, we derive analytic results for the expected size of *global cascades*, i.e., cases where a randomly chosen node can initiate a propagation that eventually reaches a positive fraction of the whole population. We also derive conditions on the contagion parameters under which global cascades take place with positive probability. Our analytic results are supported by a numerical study through which we demonstrate how the correlations between spreading contents affect the expected size of global cascades. In particular, we find that when the mean degree is *low*, correlations have limited impact on the size of global cascades. In contrast, we show that different levels and types of correlations could result in global cascades with significantly different size when the mean degree is *high*.

The rest of the paper is organized as follows. In Section II, we introduce the network model, the proposed vector threshold model, and the problem of interests. In Section III, we give the derivations of the expected size of global cascades and the condition for the existence of global cascades. Then, in Section IV, we use numerical experiments to support our analysis and discuss the impact of correlations among contents on the

expected size of global cascades. In Section V, we conclude our work and provide a brief discussion on future directions.

II. MODEL AND PROBLEM DEFINITION

A. The network model: The configuration model

The degree sequence or the degree distribution is one of the most important parameters of networks. To be able to incorporate arbitrary degree distributions, we use the configuration model to construct networks. The configuration model is a widely used reference model for real-world networks, and is also analytically tractable [15]–[17]. Let $\mathcal{N} = \{1, 2, \dots, n\}$ denote the vertex set where n is the number of vertices. For each node in $\mathcal{N}$, its degree is assigned by a prescribed degree distribution $P(d)$ where d is the random variable for the degree. After the assignment, each node has the same number of *half-edges* as its degree. A half-edge accounts for an edge one end of which is connected to the previously assigned vertex while the other end of which is free to connect any other half-edges. Then, we randomly choose two half-edges to form an edge connecting two vertices until no half-edges are left. This model requires that the sum of the degrees is even. With the above process, the configuration model could generate networks by selecting a graph uniformly at random among all possible graphs that have the same degree distribution; see [15], [16] for more details. In addition, self-loops or multi-edges could be omitted in the limit of large network size [17], which simplifies the analysis in Section III-A.

B. The vector threshold model: A threshold model with multiple correlated contents

As mentioned in Section I, there are many works that model the influence propagation in networks. However, few of them consider the simultaneous spread of multiple correlated contents. Therefore, we introduce a general threshold model, which is called the *vector threshold model*. Without loss of generality, we assume that there are two correlated contents or opinions which are denoted by Content-1 and Content-2, respectively, and that nodes could have two choices for each of the contents, support or not support. Then, with two contents spreading in a network, there exist four possible states for each node: 0) not supporting both of the contents; 1) only supporting Content-1; 2) only supporting Content-2; and 3) supporting both of the contents. For notational convenience, we use state- i to indicate the i -th state where $i = 0, 1, 2, 3$, and name state-0 as *inactive* while the other states as *active*. In addition, once a node gets activated, it will not change its state. Although we assume there are two correlated contents, the arguments could be extended to an arbitrary number of correlated contents.

In an influence propagation process, each inactive node keeps receiving influence from its contacts. Since its contacts could belong to different states, we can classify the influence by the state of its neighbors into different categories, and represent the received influence in a vector. In our model, the influence from contacts is measured by the "perceived" proportion [18], i.e., the fraction of contacts in specific states.

Then, there are two types of influence, one from supporting Content-1 while the other from supporting Content-2. Because state-0 neighbors do not exert any influence and state-3 neighbors support both of the contents, the perceived proportion from state- i is defined as

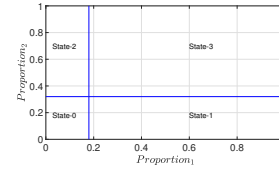
$$\text{Proportion}_i = \frac{\# \text{ of neighbors in state-}i \text{ and state-3}}{\# \text{ of neighbors}}, \quad i = 1 \text{ or } 2. \quad (2)$$

Therefore, each inactive node in the vector threshold model could receive a vector of perceived proportions, $\mathbf{v} = [\text{Proportion}_1, \text{Proportion}_2]$. Based on the received vector of an inactive node, we could determine which state the inactive node will turn to by a response function (3), i.e., the probability of changing from the inactive state to the other active states. Under the vector threshold model, given an inactive node with degree d of which m_1 , m_2 , and m_3 neighbors for state-1, 2, and 3, respectively, the probability it turns to state- i is given by

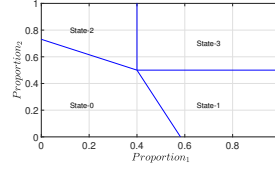
$$F_i[\mathbf{m}, d] \triangleq P[(\text{Proportion}_1, \text{Proportion}_2) \in \text{Space}_i], \quad (3)$$

where $\mathbf{m} = (m_1, m_2, m_3)$, Proportion_i accounts for the perceived proportion of state- i neighbors (2), and Space_i means the parameter space of being in state- i . The parameter space is formed by the perceived proportions, i.e., $[\text{Proportion}_1, \text{Proportion}_2]$, and is illustrated in Figure 1. By the flexibility of the design of the parameter space, this model offers a general solution for different correlations among contents.

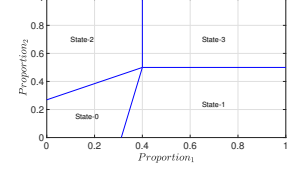
In general, there are three widely recognized correlations among contents: independent, positive, and negative correlation. We could split the parameter space accordingly to achieve these correlations. First, the independent correlation means the spread of contents is independent from each other. For example, the spread of newly released video games is usually independent from the spread of new norms. This correlation could be achieved by dividing the parameter space as Figure 1(a). In Figure 1(a), the threshold required by supporting Content-2 (resp. Content-1) is not affected by the proportion of state-1 contacts, i.e., Proportion_1 (resp. Proportion_2). Second, the positive correlation could be interpreted as a spreading content would help the spread of other contents. For instance, the spread of opinions on same-sex marriage and gun control could be a positive correlation. If we support same-sex marriage, we may more easily accept gun control. A positive correlation between the spread of contents is illustrated in Figure 1(b). In this figure, the more perceived proportion of contacts with one content an individual receives, the lower threshold it requires to turn to the other states. Third, the negative correlation means the spread of one content would impede the spread of other contents. For example, the spread of opinions on universal healthcare and tax-relief for the “rich” could be a negative correlation. Because the more we support universal healthcare, the less we will support tax-relief for the “rich”. This correlation is illustrated in Figure 1(c). As we can observe from the figure, the more perceived proportion



(a) Independent Opinions



(b) Positively Correlated Opinions



(c) Negatively Correlated Opinions

Fig. 1: An illustration of different splits on the parameter space for different correlations.

of state-1 neighbors an individual has, the higher threshold it requires to turn to state-2.

Moreover, we can have an even more complex split on the parameter space to model a non-trivial correlation between the spread of contents, like non-linear boundaries for splitting the parameter space. In this work, we only focus on a linear relationship between the perceived proportions and the boundaries of the states.

C. Problem Definition

In this work, we consider an influence propagation process under the proposed vector threshold model. In particular, assume that all nodes are initially inactive, i.e., state-0, a node is chosen uniformly at random and is set as state-3. Then, other nodes start changing their states according to (3) synchronously at times $t = 0, 1, \dots$, i.e., the influence starts propagating over networks. Since the contagion process is monotone (i.e., an active node can never switch back to inactive or other states), it will eventually stop, i.e., a steady-state will be reached.

For analyzing a spread of influence, there is a major metric of interest. The metric is the expected size of global cascades; see Definition II.2 for more details. In other words, for given contents (e.g., opinions, rumors, products, etc.), we aim at calculating the fraction of people eventually adopting these contents, in the cases where a global spreading event is possible. We can regard the expected size of global cascades as a measure of the extent of the propagation process. With this measure, we could answer or predict how widely an influence would reach, and then determine if we need to take further actions to control an influence. In addition, this metric could help us derive the condition of the existence of global cascades, which could help us control the propagation of an influence.

Definition II.1 (Global Cascades). Global cascades mean a randomly chosen node can initiate a propagation that eventually reaches a positive fraction of active nodes among the

whole population. In particular, we define a random variable S for the fraction of active nodes as

$$S \triangleq \frac{\# \text{ of active nodes at steady-state}}{n},$$

where n is the number of nodes in the network. In this case, a *global* cascade means $S > 0$ when n approaches infinity. In our discussion, active nodes indicate nodes in state-1, state-2, and state-3.

Definition II.2 (The Expected Size of Global Cascades). Given global cascades are possible, the expected global cascades size is defined as

$$\lim_{n \rightarrow \infty} \mathbb{E}[S \mid S > 0], \quad (4)$$

where the random variable S is defined in Definition II.1.

III. MAIN RESULTS

A. Analysis of the expected size of global cascades

In this section, we aim to derive the expected size of global cascades, i.e., the final fraction of active nodes in a network; see Definition II.2 for more details. If a global cascade exists, we could obtain the final fraction of active nodes by calculating the probability of that a randomly chosen node is active (i.e., in state-1, state-2, or state-3), which is expressed as (5).

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}[S \mid S > 0] &= P[\text{a randomly chosen node is active}] \\ &= \sum_d P_d \sum_{m_1}^{d-1} \sum_{m_2}^{d-1-m_1} \sum_{m_3}^{d-1-m_1-m_2} \binom{d}{m_1} \binom{d-m_1}{m_2} \times \\ &\quad \binom{d-m_1-m_2}{m_3} q_1^{m_1} q_2^{m_2} q_3^{m_3} (1-q_1-q_2-q_3)^{d-m_1-m_2-m_3} \\ &\quad \times \{F_1[\mathbf{m}, d] + F_2[\mathbf{m}, d] + F_3[\mathbf{m}, d]\}, \end{aligned} \quad (5)$$

where q_1 (resp. q_2 and q_3) indicates the probability of that the neighbors of a chosen node is in state-1 (resp. state-2 and state-3), P_d is the probability of having degree d for a randomly chosen node, and m_1 (resp. m_2 and m_3) in $\mathbf{m} = [m_1, m_2, m_3]$ is the number of neighbors in state-1 (resp. m_2 and m_3). We explain the validity of (5) as follows. First, for a randomly chosen node, the probability of having d contacts is P_d . Given the node with d contacts, by a combinatorial argument, the probability of having $\mathbf{m} = (m_1, m_2, m_3)$ neighbors is

$$\binom{d}{m_1} \binom{d-m_1}{m_2} \binom{d-m_1-m_2}{m_3} \times q_1^{m_1} q_2^{m_2} q_3^{m_3} (1-q_1-q_2-q_3)^{d-m_1-m_2-m_3}. \quad (6)$$

If a node has $\mathbf{m} = (m_1, m_2, m_3)$ active neighbors among d contacts, the probability of turning active is

$$F_1[\mathbf{m}, d] + F_2[\mathbf{m}, d] + F_3[\mathbf{m}, d], \quad (7)$$

where $F_i[\mathbf{m}, d]$ is the response function (3) of turning to state- i from state-0, $i = 1, 2, 3$. Therefore, (5) is obtained by first iterating all possible $\mathbf{m}$ and then taking the expectation of the multiplication of (6) by (7) with respect to d .

From the expression of (5), it is clear that we need the help of q_1 , q_2 , and q_3 to calculate the expected size of

global cascades. We could obtain q_1 , q_2 , and q_3 by the tree-approximation approach [19]. The tree-approximation approach was developed to get a mean-field solution to the Ising model [20]. In this approach, we assume that networks have a tree-structure. That is, each node has only one parent node and several children nodes. Then, we label each layer of the tree from the bottom to the top, $0, 1, \dots$. For each node at layer ℓ , there is only one parent node at layer $\ell+1$ and several children at layer $\ell-1$. The number of children follows the excess degree distribution mentioned; see [16] for more details. In addition, the states of nodes at layer ℓ will not update their states until all nodes at layer i , $i = 0, 1, \dots, \ell-1$ finish updating. In this case, parent nodes at $\ell+1$ are always inactive before the updates of the states of nodes at layer ℓ . With this tree-assumption, we could derive the probability of being in each state for the contacts of the initially activated nodes in the network. First, we define $q_{i,\ell}$ as the probability that a randomly chosen node at layer ℓ turns to state- i , $i = 0, 1, 2, 3$. Obviously, the probability that the chosen node at layer ℓ is in state-0 is $1 - q_{1,\ell} - q_{2,\ell} - q_{3,\ell}$. With these probabilities, we can recursively express the probabilities of being in any states for nodes at layer $\ell+1$.

In the following, we give a detailed derivation of $q_{1,\ell}$ (9), since we can derive the other probabilities in the similar way. We can see the validity of (9) as follows. For nodes at layer ℓ , the probability of having d contacts is $\frac{dP_d}{\langle d \rangle}$, since we already know that there is one parent at layer $\ell+1$; see excess degree in [16]. Among these d contacts, there are $d-1$ children and 1 parent node. With these $d-1$ children, the probability of having $\mathbf{m} = (m_1, m_2, m_3)$ active neighbors is

$$\begin{aligned} &\binom{d-1}{m_1} \binom{d-1-m_1}{m_2} \binom{d-1-m_1-m_2}{m_3} q_{1,\ell-1}^{m_1} q_{2,\ell-1}^{m_2} \\ &\quad \times q_{3,\ell-1}^{m_3} (1 - q_{1,\ell-1} - q_{2,\ell-1} - q_{3,\ell-1})^{d-1-m_1-m_2-m_3}. \end{aligned} \quad (8)$$

Again, with $\mathbf{m} = (m_1, m_2, m_3)$ active neighbors in different states, the probability of turning to state-1 for the node is given by $F_1[\mathbf{m}, d]$. If we iterate all possible $\mathbf{m}$ and take an expectation on the product of (8) and $F_1[\mathbf{m}, d]$ with respect to the degree d , we will get (9). Then, using similar arguments, we could get the expressions for $q_{2,\ell}$ (10) and $q_{3,\ell}$ (11).

$$\begin{aligned} q_{1,\ell} &= \sum_d \frac{dP_d}{\langle d \rangle} \sum_{m_1}^{d-1} \sum_{m_2}^{d-1-m_1} \sum_{m_3}^{d-1-m_1-m_2} \\ &\quad F_1[(m_1, m_2, m_3), d-1, \ell-1] \end{aligned} \quad (9)$$

$$\begin{aligned} q_{2,\ell} &= \sum_d \frac{dP_d}{\langle d \rangle} \sum_{m_1}^{d-1} \sum_{m_2}^{d-1-m_1} \sum_{m_3}^{d-1-m_1-m_2} \\ &\quad F_2[(m_1, m_2, m_3), d-1, \ell-1] \end{aligned} \quad (10)$$

$$\begin{aligned} q_{3,\ell} &= \sum_d \frac{dP_d}{\langle d \rangle} \sum_{m_1}^{d-1} \sum_{m_2}^{d-1-m_1} \sum_{m_3}^{d-1-m_1-m_2} \\ &\quad F_3[(m_1, m_2, m_3), d-1, \ell-1] \end{aligned} \quad (11)$$

where

$$\begin{aligned}
\mathbf{F}_k[(m_1, m_2, m_3), d-1, \ell-1] = & \\
& \binom{d-1}{m_1} \binom{d-1-m_1}{m_2} \binom{d-1-m_1-m_2}{m_3} \\
& \times q_{1,\ell-1}^{m_1} q_{2,\ell-1}^{m_2} q_{3,\ell-1}^{m_3} \\
& \times (1 - q_{1,\ell-1} - q_{2,\ell-1} - q_{3,\ell-1})^{d-m_1-m_2-m_3} \\
& \times F_k[(m_1, m_2, m_3), d]. \tag{12}
\end{aligned}$$

With the above derivations, we obtain a non-linear system described by the equations (9) - (11). We could recursively solve the non-linear system, and then get $q_{1,\infty}$, $q_{2,\infty}$, and $q_{3,\infty}$. These quantities obtained from the tree-approximation technique correspond to q_1 , q_2 , and q_3 introduced at the beginning of this section. Next, we could replace these probabilities into (5) to get the expected size of global cascades. As discussed in previous works [21], the tree-approximation technique generates accurate results in the asymptotic limit $n \rightarrow \infty$. We will confirm the correctness of our analysis even when the number of nodes is finite in Section IV-A.

B. The condition of the existence of global cascades

In this section, we aim to find the condition of the existence of global cascades. As mentioned in Section III-A, the proposed vector threshold model could be described by a non-linear system given by (9) - (11). In this case, the behaviors of this non-linear system imply the state of the dynamics.

As described in Section III-A, we have the relationship $q_1 := q_{1,\infty}$, $q_2 := q_{2,\infty}$, and $q_3 := q_{3,\infty}$. Then, the recursive equations (9) - (11) have the form

$$q_i = f_i(q_1, q_2, q_3), \quad i = 1, 2, 3. \tag{13}$$

From the equations (9) - (11), we can easily find that there exists a trivial fixed point $q_1 = q_2 = q_3 = 0$. If we replace this fixed point into (5), then we could obtain

$$P[\text{a randomly chosen node is active}] = 0, \tag{14}$$

which means there is no global cascade from Definition II.1. However, if there exists a non-trivial fixed point which leads to a positive value for (5), then it indicates the existence of global cascades. To check the existence of non-trivial fixed points, we could use the Jacobian matrix $\mathbb{J}$ (15) obtained by linearizing equations at $q_1 = q_2 = q_3 = 0$.

$$\mathbb{J} = \left. \frac{\partial f_i(q_1, q_2, q_3)}{\partial q_j} \right|_{q_1=q_2=q_3=0}. \tag{15}$$

If the special radius, i.e., the largest eigenvalue in absolute value, of $\mathbb{J}$ is larger than one, then it means the non-linear system is not stable at $q_1 = q_2 = q_3 = 0$. In this case, there exist non-trivial fixed points, which indicates there would exist global cascades. Otherwise, we could conclude that there do not exist global cascades.

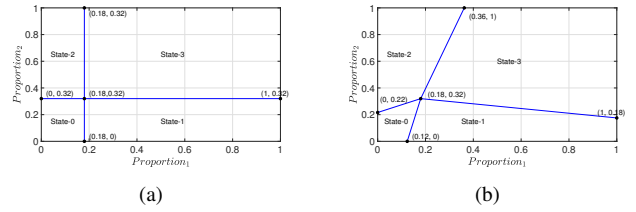


Fig. 2: The different region split strategies in Section IV-A.

IV. NUMERICAL EXPERIMENTS

A. The agreement between our analysis and experimental results

In this section, we aim to demonstrate the correctness of our analysis when the number of nodes is finite. We use Poisson distribution to assign the degree for each node. Namely, the probability that a node has degree d is given by:

$$P(d) = e^{-\lambda} \frac{\lambda^d}{d!}, \quad d = 0, 1, \dots, \tag{16}$$

where λ denotes the mean number of edges assigned for each node. We use $n = 2 \times 10^6$ as the number of nodes, and two different split strategies on the parameter space which are shown in Figure 2(a) and 2(b). We can interpret these two figures as a spread of uncorrelated contents and correlated contents, respectively, which has been discussed in Section II-B. Then, for each degree parameter, we conduct 1,000 experiments, take the average of the cascade size for experiments with *global* cascades,¹ and depict them in Figure 3 and 4. From both of Figure 3 and 4, we can observe that there is a good agreement between our analysis and experimental results. These observations confirm the correctness of our analysis of the expected size of global cascades.

In addition, we observe from Figure 3 and 4 that there exist two *phase transitions*, i.e., the place where the expected size of global cascades change from 0 to a positive fraction or vice versa. These two phase transitions are reported in many works [7], [10], [22] and provide insights on the impact of network connectivity on the expected size of global cascades. In short, the first transition indicates that global cascades exist only when the connectivity of a network achieve a certain value. In contrast, the second transition around *high* λ value appears because too much connectivity will increase the stability of nodes, i.e., when they have a large number of friends, individuals tend to be difficult to get influenced by a few active neighbors. In this case, the second phase transition appears.

B. The impact of correlations between spreading contents

As we have introduced in Section II-B, the proposed threshold model offers a general solution for different correlations among contents. In this section, we aim at exploring how

¹In a single experiment, if the fraction of active nodes is larger than 0.5%, then we regard the experiment as one with a *global* cascade.

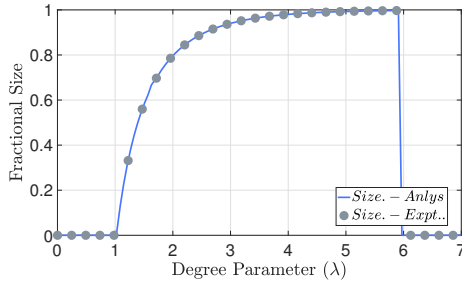


Fig. 3: There is an agreement between our analysis and numerical results. The solid curve indicates the results obtained from our analysis, while the symbols are the results obtained from numerical results. The parameter space is split as shown in Figure 2(a).

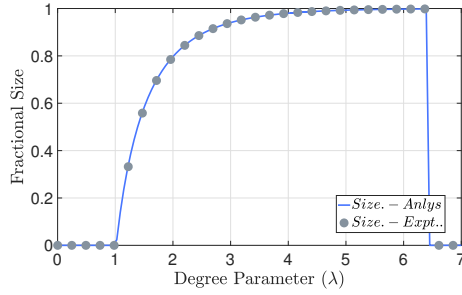


Fig. 4: There is an agreement between our analysis and numerical results. The solid curve indicates the results obtained from our analysis, while the symbols are the results obtained from numerical results. The parameter space is split as shown in Figure 2(b).

different correlations between contents affect the expected size of global cascades.

We make a comparison between three different correlations: independent, positive, and negative correlation, which are illustrated by the region split strategies in Figure 5. To make a fair comparison, we keep the thresholds right on the x - and y -axis the same. With this setting, when a node only receives Proportion_1 (resp. Proportion_2), i.e., $\text{Proportion}_2 = 0$ (resp. $\text{Proportion}_1 = 0$), the probability of being in state-1 (resp. state-2) are equal in these correlations.

What we can observe from Figure 6 are 1) there still exist two phase transitions; and 2) when the mean degree of nodes is at a low level, these correlations produce global cascades with similar size, while we see significant differences between these correlations when the degree of nodes is at a high level. This can be explained as follows. For the first observation, it happens with the same reason mentioned in Section IV-A. Limited connectivity results in the first phase transition, while high stability given by high degree leads to the second phase transition. Since different splits on the parameter space do not change the connectivity of networks, which is the reason behind the appearance of two phase transitions, two phase transitions still exist. We can explain the reason behind the second observation as follows. According to (3) and Figure 5, we find that the positive correlation makes nodes more easily be activated while it does not change the connectivity. When the mean degree is at low level, limited connectivity is the

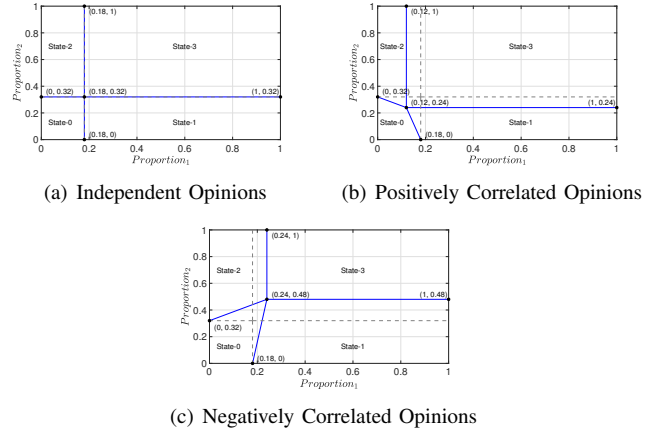


Fig. 5: The different region splits for different correlations between contents for Section IV-B. The gray dashed lines indicate the region split strategy of the independent correlation.

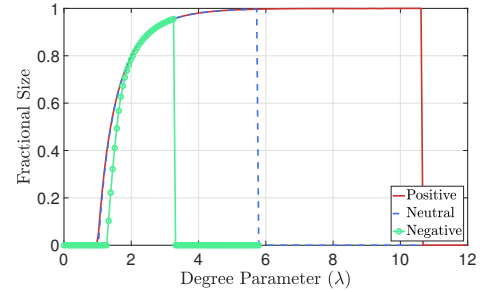


Fig. 6: The comparison between different correlations on the expected size of global cascades. These results are obtained by analytic results.

reason for determining the position of the first phase transition and the expected size of global cascades. Even a correlation makes nodes more easily be activated, the position and the size would not change much. In contrast, when the mean is at high level, there exists enough connectivity. However, high connectivity will increase the local stability of nodes; i.e., according to (3) and Figure 5, if a node has a large amount of friends, then the node is hard to be activated. The increased local stability would prevent cascades from happening. As observed in Figure 5(b), a positive correlation makes inactive nodes easier been activated. In this case, a positive correlation helps nodes to overcome the local stability caused by high degree, which makes global cascades possible again.

From these experiments, we can conclude that depending on the mean degree of nodes, the impact of different correlations between spreading contents on the expected size of global cascades will be significantly different. That is, when the mean degree is at a high level, the expected size of global cascades is more sensitive to the correlations between spreading contents.

V. CONCLUSION AND FUTURE WORK

In this work, we propose a vector threshold model that analyzes a simultaneous spread of correlated influence. We derive recursive relations characterizing the dynamics of in-

fluence propagation to compute the expected size of *global* cascades, i.e., cases where a single individual can initiate a propagation that eventually influences a positive fraction of the population. Also, we find what the conditions are in order for the existence of global cascades. Then, we use a numerical study to confirm and support our analytic results. In addition, we report an interesting observation that different correlations could lead to a significant difference on the expected size of global cascades only when the mean degree is at a high level.

There are many possible directions for future work. First, this work only derives the expected size of global cascades. Giving analytic results for the probability of having global cascades may be a possible direction. Second, this work can be extended to more general network models than the configuration model used here. For instance, it would be interesting to consider networks that have high clustering coefficient. Finally, it would also be interesting to study vector threshold models using non-linear threshold models.

ACKNOWLEDGMENTS

We thank Gopaljee Atulya for discussions which inspired this work. Research was sponsored by the Army Research Office and was accomplished under Grant Number W911NF-17-1-0587. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government.

REFERENCES

- [1] O. Yağan, D. Qian, J. Zhang, and D. Cochran, "Optimal Allocation of Interconnecting Links in Cyber-Physical Systems: Interdependence, Cascading Failures and Robustness," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1708–1720, 2012.
- [2] Y. Zhang and O. Yağan, "Optimizing the robustness of electrical power systems against cascading failures," *Scientific reports*, vol. 6, p. 27625, 2016.
- [3] D. Qian, O. Yağan, L. Yang, and J. Zhang, "Diffusion of real-time information in social-physical networks," in *2012 IEEE Global Communications Conference (GLOBECOM)*, Dec 2012, pp. 2072–2077.
- [12] S. Melnik, J. A. Ward, J. P. Gleeson, and M. A. Porter, "Multi-stage complex contagions," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 23, no. 1, p. 013124, 2013.
- [4] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical review letters*, vol. 86, no. 14, p. 3200, 2001.
- [5] O. Yağan, D. Qian, J. Zhang, and D. Cochran, "Conjoining speeds up information diffusion in overlaying social-physical networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 6, pp. 1038–1048, 2013.
- [6] Y. Zhuang and O. Yağan, "Information propagation in clustered multi-layer networks," *IEEE Trans. Network Science and Engineering*, vol. 3, no. 4, pp. 211–224, 2016.
- [7] Y. Zhuang, A. Arenas, and O. Yağan, "Clustering determines the dynamics of complex contagions in multiplex networks," *Physical Review E*, vol. 95, no. 1, p. 012312, 2017.
- [8] T. R. Hurd and J. P. Gleeson, "A framework for analyzing contagion in banking networks," *arXiv preprint arXiv:1110.4312*, 2011.
- [9] D. Centola and M. Macy, "Complex contagions and the weakness of long ties," *American journal of Sociology*, vol. 113, no. 3, pp. 702–734, 2007.
- [10] D. J. Watts, "A simple model of global cascades on random networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 9, pp. 5766–5771, 2002.
- [11] O. Yağan and V. Gligor, "Analysis of complex contagions in random multiplex networks," *Physical Review E*, vol. 86, no. 3, p. 036103, 2012.
- [13] A. Borodin, Y. Filmus, and J. Oren, "Threshold models for competitive influence in social networks," in *International Workshop on Internet and Network Economics*. Springer, 2010, pp. 539–550.
- [14] W. Cai, L. Chen, F. Ghanbarnejad, and P. Grassberger, "Avalanche outbreaks emerging in cooperative contagions," *Nature physics*, vol. 11, no. 11, p. 936, 2015.
- [15] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random structures & algorithms*, vol. 6, no. 2-3, pp. 161–180, 1995.
- [16] M. E. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Physical review E*, vol. 64, no. 2, p. 026118, 2001.
- [17] M. Newman, *Networks: An Introduction*. Oxford university press, 2018.
- [18] M. Granovetter, "Threshold models of collective behavior," *American journal of sociology*, vol. 83, no. 6, pp. 1420–1443, 1978.
- [19] S. N. Dorogovtsev, A. V. Goltsev, and J. F. Mendes, "Critical phenomena in complex networks," *Reviews of Modern Physics*, vol. 80, no. 4, p. 1275, 2008.
- [20] J. P. Sethna, K. Dahmen, S. Kartha, J. A. Krumhansl, B. W. Roberts, and J. D. Shore, "Hysteresis and hierarchies: Dynamics of disorder-driven first-order phase transformations," *Physical Review Letters*, vol. 70, no. 21, p. 3347, 1993.
- [21] J. P. Gleeson and D. J. Cahalane, "Seed size strongly affects cascades on random networks," *Physical Review E*, vol. 75, no. 5, 2007.
- [22] Y. Zhuang and O. Yağan, "Multi-stage complex contagions in random multiplex networks," *arXiv preprint arXiv:1807.00454*, 2018.