Noname manuscript No.
(will be inserted by the editor)

Theoretical and Computational Characterizations of Interaction Mechanisms on Facebook Dynamics Using a Common Knowledge Model

Chris J. Kuhlman · Gizem Korkmaz · S. S. Ravi · Fernando Vega-Redondo

Received: date / Accepted: date

Abstract Web-based interactions enable agents to coordinate and generate collective action. Coordination can facilitate the spread of contagion to large groups within networked populations. In game theoretic contexts, coordination requires that agents share common knowledge about each other. Common knowledge emerges within a group when each member knows the states and the thresholds (preferences) of the other members, and critically, each member knows that everyone else has this information. Hence, these models of common knowledge and coordination on communication networks are fundamentally different from influence-based unilateral contagion models, such as those devised by Granovetter and Centola. Moreover, these models utilize different mechanisms for driving contagion. We evaluate three mechanisms of a common knowledge model that can represent web-based communication among groups of people on Facebook, using nine social (media) networks. We provide theoretical results indicating the intractability in identifying all node-maximal bicliques in a network, which is the characterizing network structure that produces common knowledge. Bicliques are required for model execution. We also show that one of the mechanisms (named PD2) dominates another mechanism (named ND2). Using simulations, we compute the spread of contagion on these networks in the Facebook model and demonstrate that different mechanisms can produce widely varying behaviors in terms of the extent of the spread and the speed of contagion transmission. We also quantify, through the fraction of nodes acquiring contagion, differences in the effects of the ND2 and PD2 mechanisms, which depend on network structure and other simulation inputs.

Chris J. Kuhlman
Biocomplexity Institute & Initiative
University of Virginia
Charlottesville, VA
E-mail: hugo3751@gmail.com

Gizem Korkmaz
Biocomplexity Institute & Initiative
University of Virginia
Arlington, VA
E-mail: gkorkmaz@virginia.edu

S. S. Ravi
Biocomplexity Institute & Initiative
University of Virginia
Charlottesville, VA
E-mail: ssravi0@gmail.com

Fernando Vega-Redondo
Department of Decision Sciences
Bocconi University
Milan, Italy
E-mail: fernando.vega@unibocconi.it

1 Introduction

1.1 Background and Motivation

Contemporary waves of uprisings (e.g., Black Lives Matter, Women's March, Occupy Wall Street) are commonly characterized by significant use of social media to share information prior to, as well as during, protests to reach a critical number of participants. The purpose of this information is often to convince or influence others to support a cause (see Related Work, Section 2). The goal of understanding how local online interactions through social networks can facilitate information sharing in a way that generates common knowledge and coordination within large groups has motivated the construction of models of mobilization. While the exemplar in this work is protests, other applications of mobilization are family decisions to evacuate in the face of hurri-

canes and forest fires, and to participate in demonstrations for equality. Our work applies to these examples as well.

There are many social influence-based models that have been proposed and evaluated, e.g., Schelling (1960, 1971, 1978); Granovetter (1978); Oliver (1993); Watts (2002); Kempe et al. (2003); Dodds and Watts (2005); Centola et al. (2006); Centola and Macy (2007); Siegel (2009, 2010); Gonzalez-Bailon et al. (2011); Romero et al. (2011); Aral et al. (2013); He and Liu (2017). These references cover a range of topics; specifically, the spread of segregation, fads, revolts, protests, information on Twitter, and product marketing. Watts (2002) argues for the use of social influence models in a wide range of scenarios.

In many models, including those mentioned above, an agent is assigned a threshold that quantifies the amount of influence needed for it to mobilize or activate. In a networked population, an agent or network node i transitions from an inactive state (state 0) to an active state (state 1) if at least a **threshold** θ_i number of its neighbors (connections) are already in state 1. Once in state 1, a node remains in that state; this is called a *progressive* model (Kempe et al., 2003). In the above models, agents make individual decisions to change state based on their own thresholds and the states of their neighbors, irrespective of the decisions of their neighbors at the current time, and hence are referred to as **unilateral** models.

In contrast, in game-theoretic models of collective action, agents' decisions to transition to state 1 depend on their expectations of what others will do (and each agent's utility generated from her decision is based not only on what she does, but also on what decisions her neighbors make). That is, agents need to know each others' willingness to participate (defined by the threshold θ) and this information needs to be *common knowledge* among a group of agents (Korkmaz et al., 2014). A threshold is also called a *preference*, i.e., as a person's threshold decreases, her preference to participate increases. **Common knowledge** (CK) emerges within a group, if the members have communication channels to other members such that they form a particular network structure (e.g., cliques or bicliques, depending on model), and when each member knows the states and attributes (e.g., preference) of the other members, and critically, each member knows that everyone else knows her attributes. Common knowledge enables a group of agents to *coordinate* their actions, thus enabling them to transition state *simultaneously* and *cooperatively*, if it is mutually beneficial to do so.

In the context of collective action (e.g., protests), two CK models (Chwe (2000) and Korkmaz et al. (2014)) combine social structure and individual incentives together in a coordination game of incomplete information. Each provides a rigorous formalization of common knowledge. The authors study which network structures are conducive to coordination, and the local spread of knowledge and collective action.

CK models are fundamentally different from unilateral models as (*i*) contagion can *initiate* in CK models—that is, contagion can be generated when no contagion previously existed—whereas it does not in unilateral models (unless an agent's threshold is zero); (*ii*) CK models may utilize multiple mechanisms at graph geodesic distances of 1 and 2, whereas unilateral models often use influence from distance-1 neighbors, and (*iii*) the characterizing (social) network substructure for threshold-based models is a star subgraph centered at the ego node making a decision, while those for CK models include distance-2 based stars and other substructures such as cliques (Chwe, 1999) and bicliques[1] (i.e., complete bipartite graphs) (Korkmaz et al., 2014). Nodes in the biclique jointly change states if their thresholds are all less than the size of the CK set.

1.2 Overview of Common Knowledge on Facebook

Here, we provide the key concepts of Facebook communications on a social network, in the context of common knowledge. This accomplishes several things. First, it illuminates the fundamental differences between information sharing in the Common Knowledge on Facebook (CKF) model and the unilateral contagion models mentioned above. Two examples of the dynamics of the CKF model are presented: one to describe informally (to build intuition) the different mechanisms of the CKF model; one to compare CKF dynamics with those of the Granovetter (1978) type. Second, it provides the ideas that are the foundation of the operationalized CKF model formally presented in Section 3. Third, it enables us to summarize our contributions in Section 1.4 and to set up related work in Sectioni 2.

1.2.1 Conceptual Model of Facebook With Common Knowledge

In this work, we evaluate the Common Knowledge on Facebook (CKF) model (Korkmaz et al., 2014); see Figure 1. CKF models communication on Facebook (through

[1] A *biclique* contains two disjoint sets of nodes, where each node in one set has an edge to every node in the other set, while there are no edges between nodes in the same set. Examples include a cycle with 4 nodes (where each of the two disjoint sets has two nodes) and a star graph of size n where the center node (hub) is in one set, and the remaining $n-1$ nodes (spokes) are in the other set.

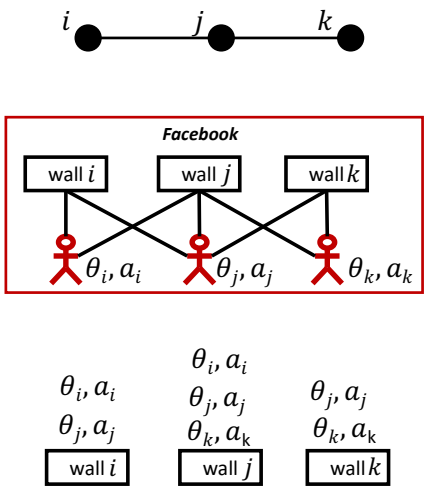


Fig. 1: Concepts for a social network representing Facebook. *Top:* three nodes, i, j, and k of a larger social network, connected by two edges. The network edges are for communication, so the social network is a communication network, e.g., among friends. *Middle:* The Facebook structure represented by a social network. Each node represents a person and that person's Facebook wall or timeline. As in Facebook, a person can post information to, and read information from, their own wall and their friends' walls. Here, the information is each node ℓ's threshold θ_ℓ and state (or action) a_ℓ. Therefore, i and k can each write on, and read from, their own walls and that of j. Node j can write to and read from all three walls. *Bottom:* the information that can be posted to, and read from, each wall. By interacting through the wall of j, i and k—at distance-2 in the social network—can communicate.

a "wall" or "timeline") as a means to generate CK and facilitate coordination. Geodesic distance-2 communication is achieved as follows: two individuals i and k do not directly communicate, but each communicates with person j. This means that if, for example, i writes information about herself on j's wall, then k knows i's information by reading j's wall, without directly communicating with i. The information thus travels two hops in a social network, that is, from i to j to k. These ideas lead to different mechanisms for driving contagion through a social network.

As indicated in Figure 1, the information shared among nodes are thresholds (preferences) θ and states or actions a. These assumptions are also in Chwe's face-to-face model of CK (Chwe, 1999, 2000). However, these conditions are not sufficient to form CK among nodes. An additional necessary condition in the CKF model is that the members have communication channels such that they form a biclique in the communication network (Korkmaz et al., 2014). If these conditions are met, CK—common knowledge of thresholds and states of group members—can form within a group. Furthermore, initial experiments on human subjects (Korkmaz et al., 2018a) indicate that individuals do behave in accordance with our model when they know their neighbors' thresholds and states, and are situated within biclique substructures.

In the CKF model, stochasticity is introduced using a parameter called **online probability** p which is the probability that an agent is online (available) during a time step to communicate with other agents. If the node is offline, then the agent cannot participate in the collective action; however, the graph structure is preserved. In other words, a person's Facebook wall exists and can be used (by others) whether or not the person is online. For example, in Figure 1, even if agent j is offline at time t, agents i and k can write on j's wall and learn about each other's thresholds and states, and can coordinate.

1.2.2 Common Knowledge Communication Mechanisms

Multiple mechanisms are operative in the CKF model, including CK itself, network dynamics, and local and global interactions. Hence, it is of interest to understand the effects of mechanisms on the spread of contagions. We aim to develop computational models of the CKF mechanisms to study these mechanisms individually and in combination, to quantify their effects on the spread of collective action. Table 1 describes these mechanisms, which are formalized in Section 3. These mechanisms are explained using examples below.

1.2.3 Illustration of Common Knowledge Contagion Dynamics

Figure 2 provides an example illustrating all three mechanisms summarized in Table 1. In this network, there are 7 people with different thresholds. Based on the CKF model (Korkmaz et al., 2014) summarized in Section 3, for agents to participate (i.e., transition to state 1), they need to share common knowledge with a group of people (they need to form a complete bipartite graph), and their thresholds should be less than the size of the common knowledge set (i.e., the group they share common knowledge with). In this example, agents 1, 2, 3, and 4 each have a threshold of 3, indicating that each needs to have at least 3 other people to participate (i.e., transition to state 1) for them to participate. These four people form a complete bipartite graph (a 4-cycle)

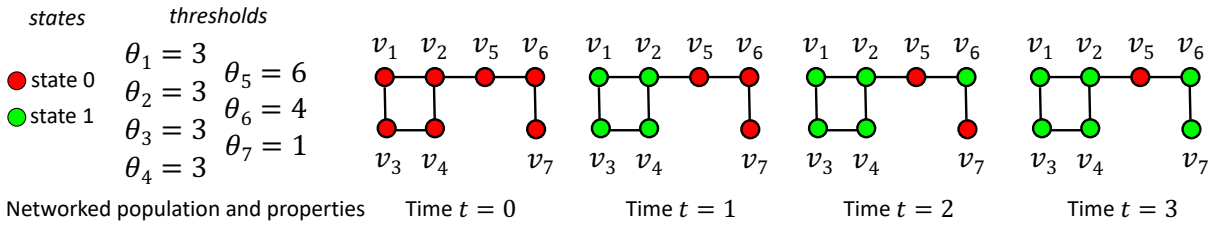


Fig. 2: Spread of contagion on a 7-node graph illustrating the mechanisms of Table 1. Each operative mechanism is evaluated independently, at each time step t. At $t = 3$, the *spread size* is 6 (6 nodes in green), and the *spread fraction* is 6/7. The dynamics resulting from the different mechanisms are discussed in the text.

Table 1: Communication mechanisms of the CKF model evaluated in this work. These mechanisms may be operative in contagion initiation, propagation, or both. Mechanism abbreviations are shown within square brackets (i.e., [·]). They are studied individually and in combination. Each of these is different from classic diffusion mechanisms, e.g., Granovetter (1978); Watts (2002); Centola and Macy (2007); Gonzalez-Bailon et al. (2011).

Mechanism	**Description**
Common knowledge [CK]	This is a common knowledge mechanism characterized by *bicliques* in social networks. This mechanism can *initiate* contagion, and can drive contagion *propagation*. No seeded nodes with contagion are required. It is a coordination mechanism where agents make joint decisions.
Neighborhood dynamics [ND2]	This is influence (communication) produced by neighbors within distance-2 of an ego node. This mechanism, also referred to as distance-2 mechanism, *propagates* contagion. This is a unilateral mechanism.
Population dynamics [PD2]	Since agents (nodes) know both state and thresholds of agents within distance-2, an agent can infer information about the numbers of agents currently in state 1, even when these other agents are at geodesic distances of 4 or more. This mechanism, also referred to as threshold inference mechanism, *propagates* contagion. This is a unilateral mechanism.

that allows them to generate common knowledge about their willingness to participate. They know each others' thresholds and know that they are sufficiently low for them to jointly participate and achieve mutual benefits. Hence, they transition to state 1 at $t = 1$. This is referred to as the *common knowledge* [CK] mechanism. On the other hand, agent 5, who shares common knowledge of thresholds with agents 1, 2, and 4 (through the 4-node star network centered at agent 2), has threshold of 6 which is not low enough for him to participate with the other 3 players that he shares CK with. Agent 5 also is part of CK node sets $\{2, 5, 6\}$ (a 3-node star centered at agent 5) and $\{5, 6, 7\}$ (a 3-node star centered at agent 6), but cannot transition to state 1 for the same reason. Similarly, persons 6 and 7 do not transition to state 1 at $t = 1$.

Since agent 2 is within distance-2 of agent 6 (friend-of-friend), agent 6 knows agent 2's threshold and state (action) through the Facebook wall or timeline of agent 5. At $t = 2$, agent 2's state is 1 and her fixed threshold is 3. Thus, agent 6 knows that at least four agents are in state 1. Agent 6's threshold is satisfied (i.e., $\theta_6 \leq 4$) and she transitions to state 1. This is the *population dynamics* [PD2] mechanism.

Finally, at $t = 3$, person 7 will transition to state 1 as a result of the *neighborhood dynamics* [ND2] mechanism: it has one activated neighbor (agent 6) within distance-2 to meet its threshold of 1. (Agent 7 can also transition due to the [PD2] mechanism, which illustrates that a node may transition to state 1 at one time due to multiple mechanisms.) All of the state transitions in this example are made formal in Section 3.

We note that applying a Granovetter (1978) type threshold model, zero nodes would transition to state 1, given the setup in Figure 2.

1.2.4 Comparison Between Granovetter Threshold Contagion Dynamics and Those from Common Knowledge

Figure 3 shows two cases of contagion dynamics produced on the same network, where thresholds of nodes are the same, but where the contagion model is different. At the top, the dynamics are for the CKF model. At the bottom, Granovetter (1978) threshold model dy-

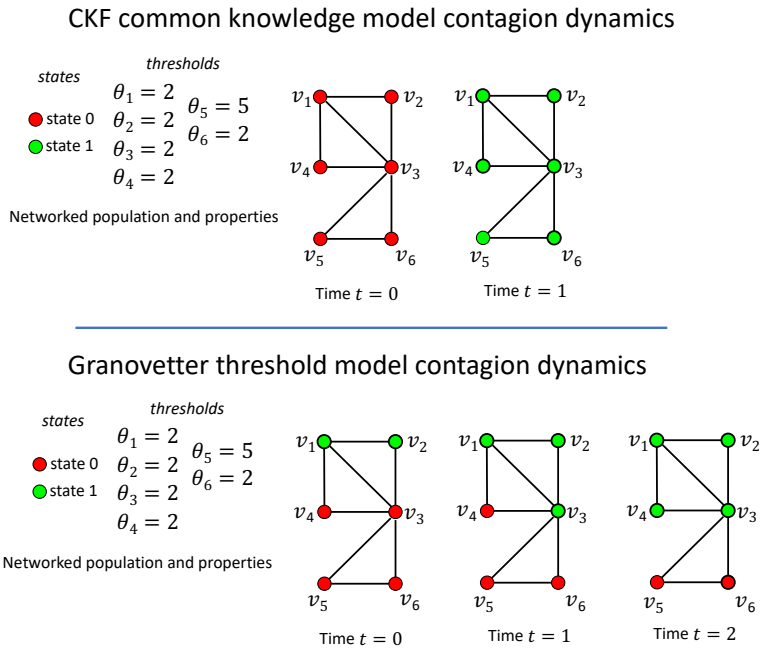


Fig. 3: Comparison of contagion spreading for the CKF model (Korkmaz et al., 2014) (top) and the Granovetter threshold model (Granovetter, 1978) (bottom). The networks and threshold assignments are the same in both cases. The seeding of nodes in the active state are different: there are no seed nodes for the CKF model, while there are two seed nodes (namely, v_1 and v_2) for the Granovetter threshold model. The CKF model shows greater diffusion, even without seeding. The final spread fractions are 1.0 (CKF) and 0.67 (Granovetter threshold).

namics are provided. The initial conditions are also different: all nodes are in the inactive state (red color) for the CKF model, whereas two nodes (v_1 and v_2) are in the active state 1 (green) at time $t = 0$ for the Granovetter model. At the end of $t = 1$, all six nodes are active in the CKF model, while only three nodes are for the Granovetter model, i.e., the fraction of nodes activated are 1.0 versus 0.50, respectively, for the two models. At $t = 1$, the CKF model produces a *fixed point*, meaning that no more node state transitions are possible. At $t = 2$, v_4 transitions in the Granovetter model, and this system now reaches a fixed point. Consequently, the final spread fractions are 1.0 and 0.67 for the CKF and Granovetter threshold models, respectively.

We explore the mechanisms for these state changes. In the top graphics in Figure 3 for the CKF model, all nodes transition at $t = 1$ because a star subgraph exists with the hub node v_3. Since this is a biclique and all nodes have $\theta_i \leq 5$, $1 \leq i \leq 6$, which is less than the number of nodes in the biclique (6), all nodes transition state, as described in the previous example. In contrast, the Granovetter model uses distance-1 influence to change the states of two nodes (v_3 and v_4) over two time steps. These differences give insights into the two models. Moreover, the same dynamics in the CKF model would occur if all thresholds were increased to five; but if this was the case, then no nodes would transition state in the Granovetter threshold model. Formal models are presented in Section 3.

1.3 Goals of Our Work

We demonstrated through two examples in the previous subsections that (i) there are multiple mechanisms by which vertices (or agents, nodes) can transition state in the CKF model, and (ii) contagion can spread more rapidly and more extensively in the CKF model compared to that for unilateral threshold-based models such as the Granovetter (1978) model. The goals of this work are to understand and quantify, both theoretically and through computational experiments, the contributions of these mechanisms to CKF contagion dynamics.

1.4 Contributions of This Work

We divide our contributions into two parts: theoretical results and computational results. The computational results helped facilitate the formulation of questions for our theoretical study.

A. Theoretical characterizations of the CKF model dynamics. Our results are as follows; see Section 4 for details.

1. All bicliques in a network must be enumerated because they are inputs to the CKF model. The problem of listing all of the bicliques in a graph takes exponential time in the worst-case.
2. Given a social network G under the CKF model where all nodes have the same threshold $\theta \geq 0$ and the model is deterministic (i.e., the online probability $p = 1$ for all nodes), the population dynamics (threshold inference) mechanism [PD2] dominates the neighborhood dynamics (distance-2) mechanism [ND2]. However, the converse may not hold.
3. For a social network G under the deterministic CKF model where all nodes have the same threshold θ with θ being either 0 or 1, the threshold inference mechanism [PD2] is equivalent to the distance-2 mechanism [ND2].
4. The previous two results may not hold when all nodes do not have the same threshold. This emphasizes the importance of the uniform threshold assumption.
5. For the stochastic version of the CKF model (i.e., for at least one node, the online probability $p < 1$) where all nodes have the same threshold $\theta \geq 1$, the [PD2] mechanism dominates the [ND2] mechanism.

B. Computational results for the CKF model. We quantify contagion dynamics on nine web-based social networks (presented in Section 5) that range over three orders of magnitude in numbers of nodes and of edges, 5× in average degree (i.e., over a factor of 5), three orders of magnitude in maximum degree, 80× in average clustering coefficient, and 3× in graph diameter. Based on simulations on these networks (in Section 6), our results are presented below. In this discussion, we use d_{ave} and d_{max} to denote respectively the average and maximum node degrees of a network. We provide a qualitative summary of several computationally-driven observations; a more detailed listing is given in Section 6.2.

1. The [PD2] mechanism is a more dominant mechanism in driving contagion in networks than the [ND2] mechanism. Over all the networks, once the two mechanisms [CK]+[PD2] are active, there is no benefit (i.e., there is no further increase in contagion spreading) by adding the [ND2] mechanism. In some cases, the contribution to contagion of the [ND2] mechanism rivals that of the [PD2] mechanism.
2. The [PD2] mechanism always increases the contagion spread over that of the [CK] only mechanism. The [ND2] mechanism sometimes increases the contagion spread over that of the [CK] only mechanism. The amounts of these contributions vary widely across networks, for different conditions.
3. As p increases from 0.03 to 0.4 for a single network, the differences between the [CK]-generated spread fractions and the [CK]+[PD2]-generated spread fractions (over the population) decrease. Lesser online probabilities favor a mechanism whereby a node can transition to state 1 via influence of a single neighbor, as is the case for [PD2]. As p increases, more nodes are online and conditions are more conducive for the CK mechanism, resulting in [CK]-driven contagion spread catching up.
4. The ranking of networks, from least able to spread contagion to most able (for threshold $\theta = \lceil d_{ave} \rceil$), is characterized by the increasing order of the ratio d_{max}/d_{ave}. See Section 6.3.4 for details.

1.5 Extensions from Preliminary Version

A preliminary version of this paper appeared as Kuhlman et al. (2020). This version contains a number of extensions to the previous version, including the following: (*i*) expansion of Introduction and Related Work; (*ii*) all theoretical results are new (Section 4); (*iii*) four new networks are analyzed in Section 5, in addition to the original five; and (*iv*) new simulation results in Section 6 (the number of results plots in the paper has increased from eight to 43).

1.6 Paper Organization

The paper is organized as follows. Section 2 contains related work. Formal definitions concerning the model are presented in Section 3. Theoretical results are provided in Section 4. Section 5 contains the social networks that we study using computational experiments. The agent-based simulation (ABS) process and results are presented in Section 6. Section 7 presents some concluding remarks.

2 Related Work

Many *unilateral* models and applications discussed in Section 1 are not repeated here. Here, we focus on models and data from social media and game-theoretic common knowledge models. See Oliver (1993) for a review of coordination and collective action.

There are several studies that model web-based social media interactions. The spread of hashtags on Twitter is studied using a threshold model in Romero et al.

(2011). Diffusion on Facebook is modeled in Sun et al. (2009), and a similar type of mechanism on Facebook is used to study the resharing of photographs in Cheng et al. (2014). None of these works uses the "wall" or "timeline" mechanism of Facebook that is considered here under the CKF model.

A couple of data mining studies have used Facebook walls, including an experimental study (Devineni et al., 2015). Features of cascades on Facebook are studied using user wall posts (Huang et al., 2013), but again, these are cascades of the conventional social influence type; there is no assessment of CK-based coordination.

Most experimental studies of social influence involve unilateral interactions, where one user sends messages to influence one or more others. Web-based experimental studies of unilateral influence phenomena include those on Twitter (Gonzalez-Bailon et al., 2011), Facebook (Bakshy et al., 2012; Dow et al., 2013; Kramer et al., 2014; Cheng et al., 2014; Adamic et al., 2016), LinkedIn (Chen et al., 2017), Digg (Hodas and Lerman, 2014), Doodle (Romero et al., 2017), Stack Overflow (Upadhyay et al., 2017), and Wikipedia (Backstrom et al., 2013), among others (e.g., Centola (2010, 2011)). We argue that our work on CK is valuable in complementing the large number of studies, such as those cited here, on unilateral models and mechanisms.

A study of the spread of videos on YouTube (Susarla et al., 2012) finds four mechanisms of social influence operative: awareness, attention, homophily, and information. They find significant influence (contagion) effects and network structure effects in determining whether videos are successful and to what degree. Our study also seeks to understand individual mechanisms.

The Chwe common knowledge model was introduced in Chwe (1999, 2000); it was conceived for face-to-face or in-person interactions that produce common knowledge. Several situations in which common knowledge may be produced are given in Chwe (1998). Korkmaz et al. (2018a) conducts an online experimental study testing the Chwe common knowledge model.

There have been a few works on the CKF model introduced in Korkmaz et al. (2014). Details of the game-theoretic formulation are provided there. For example, simulations of the CKF model cannot be carried out efficiently; we point out in Section 4 that listing all bicliques in a network takes exponential time in the worst-case. This makes studying CKF on very large networks (e.g., with 1 million or more nodes) extremely difficult. See Section 5 for computational details. An approximate and computationally efficient CKF model is specified in Korkmaz et al. (2016a). Studies involving both Chwe and CKF models include basic simulation results (Korkmaz et al., 2014). CK dynamics on networks without key players is studied in Korkmaz et al. (2016b, 2018b). *None* of these works investigates the individual and combinations of mechanisms shown in Table 1.

3 Model

3.1 Preliminaries

This section provides a formal description of the Common Knowledge on Facebook (CKF) model (Korkmaz et al., 2014) studied in this paper. The population is represented by a communication network $G(V, E)$. There is a set $V = \{1, 2, \ldots, n\}$ of n nodes (people) and an edge set E where an undirected edge $\{i, j\} \in E$ means that nodes $i, j \in V$ can communicate with each other. Each person $i \in V$ at time t is in a state $a_{it} \in \{0, 1\}$: if $a_{it} = 1$, person i is in the active state (e.g., joining a protest), and $a_{it} = 0$ otherwise (e.g., staying at home). We use *progressive* dynamics (Kempe et al., 2003), such that once in state 1, nodes do not transition back to 0. Each node i has a threshold θ_i that indicates its inclination/resistance to activate.

3.2 Common Knowledge Formalism

Given person i's threshold θ_i and the system state at t, denoted by the n-vector $a_t = (a_{1t}, a_{2t}, \ldots, a_{nt})$, her utility is given by

$$U_{it} = \begin{cases} 0 & \text{if } a_{it} = 0 \\ 1 & \text{if } a_{it} = 1 \ \wedge \ |\{j \in V : a_{jt} = 1\}| \geq \theta_i \\ -z & \text{if } a_{it} = 1 \ \wedge \ |\{j \in V : a_{jt} = 1\}| < \theta_i \end{cases} \quad (1)$$

where $-z < 0$ is the penalty a person gets if she activates and not enough people join her. Thus, a person will activate as long as she is sure that there is a sufficient number of people (in the population) in state 1 at t. A person always gets utility 0 by staying in state 0 regardless of what others do since we do not consider free-riding problems. When she transitions to the active state, she gets utility 1 if the total number of other people activating is at least θ_i. (Note that these "others" need not be neighbors of i, as in unilateral models.)

The CKF model describes Facebook-type (friend-of-friend) communication in which friends write to and read from each others' Facebook walls. This information is also available to their friends of friends. An overview of the Facebook setup was provided in Section 1.2 and Figure 1. The mechanisms and their implications are described below. The communication network indicates that if undirected edge $\{i, j\} \in E$, then node i (resp., j) communicates (θ_i, a_{it}) (resp., (θ_j, a_{jt}))

to node j (resp., i) over edge $\{i,j\}$ at time t. This information is available to j's (resp., i's) neighbors. The communication network helps agents to coordinate by creating common knowledge at each t.

3.3 Facebook Common Knowledge Model Mechanisms

Here we describe the three mechanisms in this model (cf. Table 1), and their implications. (Figure 2 of Section 1 illustrates these mechanisms.)

The CKF model describes a Facebook type communication which allows for distance-2 communication: two nodes, i and j, with $\{i,j\} \notin E$ can communicate by posting to and reading from the wall of a common neighbor k, provided $\{i,k\},\{j,k\} \in E$. Thus, all $i \in V$ can communicate with all nodes $j \in V$ if their geodesic distance $\delta(i,j)$ is at most 2. All three mechanisms make use of this Facebook communication structure.

For the **common knowledge [CK]** mechanism of Table 1, the biclique subgraph is the structure necessary for creation of CK among a group of people (Korkmaz et al., 2014), and allows them to jointly activate. We first compute all node-maximal bicliques in G. Let $M^{biclique}$ denote the set of nodes of G that forms a biclique. Then, V in Equation (1) is replaced with $M^{biclique}$. At each t, Equation (1) is computed for each $i \in V$ in each CK set $M^{biclique}$ for which $i \in M^{biclique}$.

The **neighborhood dynamics [ND2]** mechanism (Table 1) is similar to the Granovetter (1978) unilateral contagion model, but with interaction at both distance-1 and distance-2. Let the neighbors j of i within distance-2 be defined by $N_i^2 = \{j : \delta(i,j) \leq 2\}$. The [ND2] mechanism is given by

$$a_{it} = \begin{cases} 1 & \text{if } a_{i,t-1} = 1 \text{ or } |\{j \in N_i^2 : a_{j,t-1} = 1\}| \geq \theta_i \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Finally, the **population dynamics [PD2]** mechanism indicates that a node i that is in state 0 can infer a minimum number of nodes already in state 1 if a node j in N_i^2 is already in state 1, by knowing θ_j. Formally,

$$a_{it} = \begin{cases} 1 & \text{if } a_{i,t-1} = 1 \text{ or} \\ & \quad (\max \theta_j : j \in N_i^2, a_{j,t-1} = 1) + 1 \geq \theta_i \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Assume $a_{i,t-1} = 0$. If $j \in N_i^2$ and $a_{j,t-1} = 1$, with θ_j, then i can infer that at least $\theta_j + 1$ nodes are in state 1. Now, if $\theta_i \leq \theta_j + 1$, then i will transition to state 1; i.e., $a_{it} = 1$.

At each time $t-1$, all operative mechanisms are evaluated, independently, for each $i \in V$ in which $a_{i,t-1} = 0$. If any of the three mechanisms causes i to transition, then $a_{it} = 1$.

3.4 Online Probability

Each agent i's presence or absence on the social network (i.e., online or offline) is captured by the **online probability** $0 \leq p_i \leq 1$. This node probability is the probability that a node is online (i.e., is available for communication on Facebook) or the contagion dynamics at each t. A node i that is online at t, and in state 0, can possibly transition state $0 \to 1$ and, if $a_{it} = 1$, then i can influence other nodes to transition state $0 \to 1$. A node that is offline at any t cannot transition state nor influence others to transition due to absence of communication. Note that the graph structure in Figure 1 is preserved irrespective of whether a node is online.

In computing contagion dynamics at each time t, each node i first conducts a Bernoulli trial with probability p_i. After all nodes have completed this step, the nodes that are online at this time step are known. Only these nodes that are considered in the mechanisms of Section 3.3.

4 Theoretical Results

4.1 Overview

In this section, we present theoretical results for the CKF model. First, we observe that a core problem in the simulation of the CKF model, namely the generation of all the bicliques in a graph, is intractable in general. We then present several results that point out dominance relationships between the threshold inference ([PD2]) and distance-2 ([ND2]) mechanisms for propagating a contagion. We also examine the effect of the online probability on the dominance relationships.

4.2 Intractability of Listing all the Bicliques

Proposition 1 *The problem of listing all the bicliques in a graph takes exponential time in the worst-case.*

Proof As proven by Prissner (2000), there are graphs on n nodes where the number of bicliques is *exponential* in n. In particular, Prissner presents a class $\mathcal{C}$ of graphs where each graph has n nodes and the number of bicliques in the graph is $3^{n/3}$. Thus, any algorithm

that lists each of the bicliques in a graph must use exponential time when the input to the algorithm is from the class $\mathcal{C}$ of graphs identified in Prissner (2000). ■

4.3 Relationships between Threshold-Inference and Distance-2 Mechanisms

Proposition 2 *Given a social network G under the CKF model where all nodes have the same threshold $\theta \geq 0$ and the model is deterministic (i.e., the online probability $p = 1$ for all nodes), the threshold inference mechanism [PD2] dominates the distance-2 mechanism [ND2]; that is, if [ND2] causes nodes to transition state, then [PD2] will also cause these same nodes to transition state. However, the converse may not hold.*

Proof We first show that whenever [ND2] causes a node to transition state, so does [PD2]. If the threshold value $\theta = 0$ for all the nodes, then the result is trivial since all the nodes will transition to 1 in the very first time step. So, assume that $\theta \geq 1$. We first note that both mechanisms require the existence of at least one node in state 1 for causing a node state transition. Suppose the [ND2] mechanism is operative at a node j and it causes the node j to transition from 0 to 1. Then, by definition of the [ND2] mechanism, there are at least θ nodes within a distance of 2 from j which are in state 1 during the time instant before j transitions to 1. Since $\theta \geq 1$, we conclude that there is at least one node i which is in state 1 before j's transition and which is within a distance of 2 from j. Since the state of i is 1, and $\theta_j = \theta_i$, the conditions are met for j to transition to state 1 by [PD2]. Thus, whenever [ND2] causes a node state transition, so does [PD2].

To prove that the converse does not hold, consider the following example. Suppose the graph is a tree with five nodes as shown in Figure 4 and all the nodes have the same threshold $\theta = 2$. Suppose at time 0, nodes c, d and e are in state 1 while nodes a and b are in state 0. Under [ND2] mechanism, node a has only one node (namely, node c) which is within a distance of 2 from a and which is in state 1. Since the threshold of node a is 2, a cannot transition from 0 to 1 using the [ND2] mechanism. However, since node a has one node c which is in state 1 and c is within a distance of 2 from a, the [PD2] mechanism is operational at a, and consequently, a can transition to 1. Hence, even though the [ND2] mechanism does not cause node a to transition, the [PD2] mechanism does cause node a to transition to state 1. ■

We now consider a special case where all the nodes have the same threshold which is either 0 or 1.

Proposition 3 *Given a social network G under the deterministic CKF model where all nodes have the same threshold θ and θ is either 0 or 1. Then the threshold inference mechanism [PD2] is equivalent to the distance-2 mechanism [ND2]. In other words, if one of these mechanisms causes a node to transition to state 1, then so does the other.*

Proof When all the nodes have the same threshold $\theta = 0$, the proposition is trivially true since each node changes to 1 in the first time step. So, assume that for all the nodes, the threshold $\theta = 1$. In this case, from Proposition 2, it follows that whenever [ND2] causes a node state transition, so does [PD2]. Thus, we only need to prove the converse. Assume that some node i changes from 0 to 1 through the [PD2] mechanism. Thus, there is at least one node j within distance 2 of i whose state is 1. When $\theta = 1$, this is exactly the condition needed under [ND2] for node i to change from 0 to 1. Thus, [ND2] is also causes transition at i, and this completes our proof of Proposition 3. ■

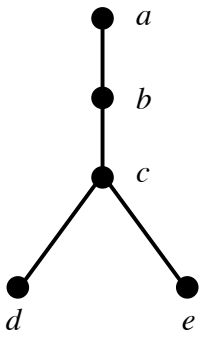


Fig. 4: Example tree used in proving Proposition 2.

In the above proposition, the proof of dominance of the [PD2] mechanism over the [ND2] mechanism relies on the assumption that all nodes have the same threshold θ. Our next result points out that the assumption of uniform thresholds is essential.

Proposition 4 *There exists a social network G under the deterministic CKF model where not all the nodes have the same thresholds such that the [PD2] mechanism does not dominate the [ND2] mechanism; that is, for some node u of G, the [ND2] mechanism causes u to transition to state 1, but the [PD2] mechanism does not cause transition.*

Proof Let G be the star graph with $n \geq 4$ nodes where the center node u has degree $n-1$ while each of the other nodes is adjacent only to u; thus, the nodes other than u are leaves with degree 1. Suppose u is in state 0 and its threshold $\theta_u = n-1$. Let W denote the set consisting of all the $n-1$ leaf nodes. Let each node $w \in W$ be in state 1 and have a threshold of 1; note that $|W| = \theta_u = n-1 \geq 3$. In this example, the threshold inference mechanism [PD2] will not be activated for node u because all nodes in W have a threshold of 1 which is less than θ_u. However, the distance-2 mechanism [ND2] will be activated because of the following: (i) all nodes $w \in W$ are within distance-2 of u (in fact, they are at distance-1 from u); and (ii) $|W| = \theta_u$. Thus, node u's threshold is met by the [ND2] mechanism, and it will transition to state 1. ■

For uniform thresholds, the dominance of the [PD2] mechanism over the [ND2] mechanism indicated in Proposition 2 also holds under the stochastic version of the CKF model (i.e., for at least one node $p < 1$), as shown below.

Corollary 1 *Given a social network G, and the CKF model where all nodes have the same threshold $\theta \geq 1$ and the model is stochastic (i.e., for at least one node, the online probability $p < 1$). Then, the [PD2] mechanism dominates the [ND2] mechanism.*

Proof The proof is similar to the one presented for Proposition 2. Suppose all the nodes have threshold $\theta \geq 1$ and a node i changes from 0 to 1 using the [ND2] mechanism. We will show that the [PD2] mechanism is also operational at i.

Since the [ND2] mechanism is operational at i, there are at least $\theta \geq 1$ nodes which satisfy all of the following properties: (i) these nodes are all in state 1, (ii) they are all within distance 2 of i and (iii) these nodes are all online during the time step when i changes from 0 to 1. These conditions readily imply that there is at least one node j satisfying all the following properties: (i) j is in state 1, (ii) $\theta_j = \theta_i$, (iii) j is within distance 2 of i, and (iv) j is online during the time step when i changes from 0 to 1. These are exactly the conditions needed for the [PD2] mechanism to cause node i to transition. The corollary follows. ■

4.4 Sensitivity of [ND2] and [PD2] Mechanisms to Online Probability

Here, we consider systems under the stochastic CKF where all nodes have the same threshold $\theta \geq 1$ and each node decides to be online with probability p independently of other nodes. We will explain through an example that the chances of a node transitioning to 1 through the [PD2] mechanism is at least as large as the transition through the [ND2] mechanism.

The example is again a star graph similar to the one considered in the proof of Proposition 4. Let u be the center node of a star graph and let θ (which is the threshold of each node) be the number of leaves in the star graph. Assume that at a certain time t, u is in state 0 and is online. For each leaf node w, let the state and online probability be 1 and p respectively. In this situation, both [ND2] and [PD2] mechanisms may apply depending on how many nodes are online.

For the [ND2] mechanism to apply, since the threshold of u is θ, all the θ leaf nodes must be online at time t; since each node decides to be online independently, the probability of that event is p^θ. For the [PD2] mechanism to apply, it is enough to have at least one of the leaf nodes online. To compute this probability, note that the probability that none of the θ leaf nodes is online is $(1-p)^\theta$; therefore, the probability that at least one node is online at time t is $1-(1-p)^\theta$. To see the relationship between the probability expressions p^θ and $1-(1-p)^\theta$, first note that since $p > 0$, $1-p < 1$. Thus, for any $\theta \geq 1$, we have

$$(1-p) \;\geq\; (1-p)^\theta.$$

Therefore,

$$1-(1-p)^\theta \;\geq\; p \;\geq\; p^\theta,$$

where the last inequality follows from the fact that $p \leq 1$. In the final inequality above, the left and right sides represent respectively the transition probabilities for u due to the [PD2] and [ND2] mechanisms. We thus conclude that in the above example, transition due to [PD2] is at least as likely as that due to [ND2].

Finally, we relate these theoretical results to the forthcoming simulation results in Section 6. In simulations, we always include the [CK] mechanism because this is the only mechanism that does not require seeding of nodes in state 1 in order for contagion to appear (i.e., to initiate) and to spread. Hence, including the [CK] mechanism obviates the need for seeding. The results of this section apply to the simulations in Section 6 because our results here do not depend on how nodes reach state 1; only that at least one node is in state 1 so that [PD2] and [ND2] mechanisms can operate.

Table 2: Characteristics of web-based social networks analyzed. If there are multiple connected components in a graph, we use only the giant component. Here, n and m are numbers of nodes and edges, respectively; d_{ave} and d_{max} are average and maximum degrees; c_{ave} is average clustering coefficient; Δ is graph diameter; n_{bbc} is the number of non-star bicliques; s_{bbc} is the average size of the non-star bicliques; and s_{bc}^{max} is the maximum size of *any* biclique. Several properties are computed with the codes in the *net.science* cyberinfrastructure (Ahmed et al., 2020), using methods in SNAP (Leskovec and Sosič, 2016) and NetworkX (Hagberg et al., 2008).

Network	**Type**	n	m	d_{ave}	d_{max}	d_{max}/d_{ave}	c_{ave}	Δ	n_{bbc}	s_{bbc}	s_{bc}^{max}
Jazz	Musicians	198	2,742	27.7	100	3.61	0.617	6	22,228	16.7	101
SF1	Stylized	4,956	45,031	18.2	270	14.84	0.0780	8	833,918	10.8	271
Wiki	Online Voting	7,115	100,762	28.3	1,065	37.63	0.141	7	355,012,343	17.5	1,066
Ca-Hepth	Co-authorship	8,638	24,806	5.74	65	11.32	0.482	18	207	8.8	65
FHS	Human Health	10,430	37,103	7.11	78	10.97	0.530	18	1,688	13.0	79
P2P	Peer Comms.	10,876	39,940	7.34	103	14.03	0.00622	10	5,663	5.9	104
Astroph	Co-authorship	17,903	196,972	22.0	504	22.91	0.633	14	82,795	27.0	505
Enron	Email	33,696	180,811	10.7	1,383	129.3	0.509	17	22,264,629	18.1	1,384
FB	Facebook	43,953	182,384	8.30	223	26.87	0.115	18	258,668	7.1	224

5 Social Networks Used and Their Structural Properties

In this study, we use nine web-based social networks summarized in Table 2. For example, FB is a Facebook user network (Viswanath et al., 2009), P2PG is a peer-to-peer network, Wiki is a Wikipedia network of voting for administrators, and Enron is an Enron email network (Leskovec and Krevl, 2014). All but the SF1 network are real (i.e., mined) networks. SF1 is a scale free (SF) network generated by a standard preferential attachment method (Barabasi and Albert, 1999) to fill in gaps of the real networks. Most networks come from Leskovec and Krevl (2014).

Structural properties were generated with SNAP (Leskovec and Sosič, 2016) and NetworkX (Hagberg et al., 2008) with their implementations in the *net.science* cyberinfrastructure for network science (Ahmed et al., 2020). Structural properties are presented in columns labeled n (number of nodes in network) through Δ (graph diameter). The last three columns provide data on bicliques found in networks; see the Table 2 caption for property definitions.

For networks possessing multiple components, we use the giant component of that network for all of our work. These networks have wide-ranging structural properties: over three orders of magnitude in numbers of nodes and of edges, 5× in average degree (i.e., over a factor of 5), three orders of magnitude in maximum degree, 80× in average clustering coefficient, and 3× in graph diameter. Hence, they represent a broad sampling of web-based mined network features.

Figure 5 shows the average degrees for selected networks in the original graphs G, corresponding to geodesic distance of 1, and in the square of the graphs G^2. These increases are an order of magnitude or more, and are important for the [ND2] and [PD2] mechanisms that operate over distance-2.

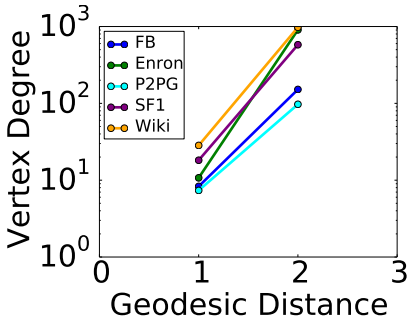


Fig. 5: Average vertex degree for geodesic distances 1 and 2 (i.e., for G^1 and G^2), illustrating the large increase in node degree that occurs in going from distance-1 neighborhoods to distance-2 neighborhoods.

Per Section 3, the biclique is the network substructure that is required to form CK in the CKF model. Therefore, the simulations in Section 6 require as input the bicliques for each network. As discussed in Section 4, the number of bicliques in a graph can be exponential in the graph size. This means that it is computationally intensive to identify all bicliques in a network. Here, we use the method and code from Liu et al. (2006). We ran the serial code on Dell PowerEdge C6420 2.666 GHz hardware nodes, with 384 GB RAM and 40 cores per node. We ran the code on each network for up to seven days of wall-clock time. (For a shared network servicing over 100 users, this is a large amount of time.) Results are printed out as computations proceed, and

computations for a few networks (Astroph and FHS) did not finish in seven days, and so we use the results generated up to that point.

To give an idea of the size of these computations, the Enron and Wiki networks produced biclique files of sizes 18 GB and 129 GB (gigabyte), respectively. The Astroph and FHS networks each produced output files of bicliques between 1 and 2.5 TB (terabytes) in size. Even the smallest network, Jazz, produced a biclique output file of size 115 GB. These results are not surprising in light of the theoretical result in Section 4.2. So, in this sense, the graphs in Table 2 are large. The number of unique bicliques determined with the cited code is in column n_{bbc} of the table. Specifically, the Liu et al. (2006) code iterates through a graph such that the same biclique can be identified multiple times, which is largely responsible for the large file sizes. We iterate through the resulting output file of bicliques, keeping only unique bicliques.

6 Simulations and Results

In this section, we describe the process, parameters and results of our agent-based simulations.

6.1 Agent-Based Simulation Process and Parameters

We conduct discrete time agent-based simulations using the CKF model described in Section 3, the web-based networks given in Table 2, and the bicliques determined as part of structural characterization of the networks. Table 3 summarizes the parameters and their values used in simulations. A **simulation** consists of a set of 30 runs, where a **run** consists of the spread of contagion from an initial configuration (or state) with all nodes in state 0 at time $t = 0$, to a specified maximum time t_{max}. Differences among runs are due to the stochasticity in models. Note that the process simulated on these networks is conceptually the same as the contagion dynamics in Figure 2 and in the upper graphic of Figure 3.

In the results below, we report the average of all runs of a simulation. In some plots, we provide error bars to demonstrate that variability across runs of a simulation is quite small. In many of the plots with multiple curves, error bars are not shown since they make the results more cluttered with no additional insights because curves can be very close together.

Seed nodes are nodes that are assigned state 1 at time $t = 0$. We have no seed nodes in simulation instances (i.e., runs), for all simulations in this work. Thus, at $t = 0$, all agents are in state 0, and the only mechanism by which nodes can transition $0 \rightarrow 1$ is [CK]. In fact, [CK] is the only mechanism that can result in a state transition when all nodes are in state 0, regardless of t. Once at least one node is in state 1, all three mechanisms may be operative.

Table 3: Summary of contagion study parameters.

Parameter	**Description**
Networks	See networks in Table 2.
Agent thresholds θ	Uniform threshold values for each simulation: all nodes in a network have the same value. Values range from $\theta = 6$ through $\theta = 29$ across simulations and networks.
Online probabilities p	Uniform value for all nodes in a simulation. Values are in the range of 0.03 to 0.4.
Model mechanisms	[CK], [ND2], and [PD2] mechanisms described in Table 1. [CK] is always operative to initiate contagion.
Seed vertices	No specified seed vertices; all vertices initially in state 0. CK model initiates contagion without seeds.
Simulation duration t_{max}	30 and 90 time steps.

6.2 Summary of Findings

Our findings are summarized as follows.

1. The [PD2] mechanism is a more dominant mechanism in driving contagion in networks than the [ND2] mechanism. Over all the networks, once the two mechanisms [CK]+[PD2] are active, there is no benefit (i.e., there is no further increase in contagion spreading) by adding the [ND2] mechanism.
2. The [PD2] mechanism always increases the contagion spread over that of the [CK] only mechanism; amounts vary from a roughly 25% increase (for the Enron network) to a 900% increase (for the P2P network).
3. The [ND2] mechanism sometimes increases the contagion spread over that of the [CK] only mechanism; amounts vary from a roughly no increase (for Enron, Wiki) to a 100% increase (for P2P).
4. Recall that online probability p is the probability that an agent is online during a time step to communicate with other agents. Over all the networks, online probability values as low as 0.03 can produce some level of contagion spread in seven of nine networks. Contagion spread for all mechanisms asymptotically reach their upper bounds for $p = 0.4$.
5. As p increases from 0.03 to 0.4 for a single network, the differences between the [CK]-generated spread

fractions and the [CK]+[PD2]-generated spread fractions (over the population) decrease. Lesser online probabilities favor a mechanism whereby a node can transition to state 1 via influence of a single neighbor, as is the case for [PD2]. As p increases, more nodes are online and conditions are more conducive for the CK mechanism, resulting in [CK]-driven contagion spread catching up.

6. The ranking of networks, from least able to spread contagion to most able (for threshold $\theta = \lceil d_{ave} \rceil$), is characterized by the increasing order of the ratio d_{max}/d_{ave}. See Section 6.3.4 for details.
7. Explaining CKF model contagion dynamics is complex. The previous point explains the behaviors of all networks, except the FB network. Also, the Jazz network shows virtually no contagion spread, for $\theta = \lceil d_{ave} \rceil$, up through $p \leq 0.1$, whereas FB shows significant spreading for $p = 0.1$. Yet, with further increase in online probability to $p = 0.4$, the ranking switches: the spread of contagion is far more on Jazz than on FB network.
8. Average degree is a key parameter in understanding contagion dynamics. Four networks have $d_{ave} \leq 8.3$. Each of these networks generates no contagion spreading for the [CK]-only mechanism for $\theta = 8$. The remaining five networks with greater d_{ave} generate significant contagion. Adding [PD2] to the [CK] mechanism generates contagion in three of the four networks that generate no contagion with only the [CK] mechanism.

6.3 Results

This section presents results from our simulation experiments. These results are broken down by topics as indicated in Table 4. We also discuss the links between simulation results and the theoretical results in Section 4.

6.3.1 Effects of CKF model mechanisms on contagion dynamics for the Wiki network

Figures 6g through 6i—the last row of plots in Figure 6—show cumulative fractions of agents in state 1 in the Wiki network as a function of simulation time. The threshold of all nodes is $\theta = \lceil d_{ave} \rceil = 29$. There are four curves in each of the three plots in each row of Figure 6. The [CK] mechanism is the blue curve; the [CK]+[ND2] mechanisms is the magenta curve; the [CK]+[PD2] mechanisms is the green curve; and the [ALL] mechanisms is the orange curve. The three plots correspond to different online probabilities p, increasing from 0.03 to 0.05 and 0.1. In each plot, the [CK] and [CK]+[ND2] curves are close, and the [CK]+[PD2] and [ALL] curves are close. These curves suggest that, as a first broad conclusion for Wiki, the [PD2] mechanism makes a significant difference when added to the [CK] mechanism (i.e., the green and orange curves are higher in the plots than are the blue and magenta curves). The second broad conclusion for Wiki is that the [ND2] has a lesser effect when added to the [CK] mechanism (i.e., the magenta curve is only slightly higher in the plots than the blue curve).

Table 4: Results are grouped into the following subsections.

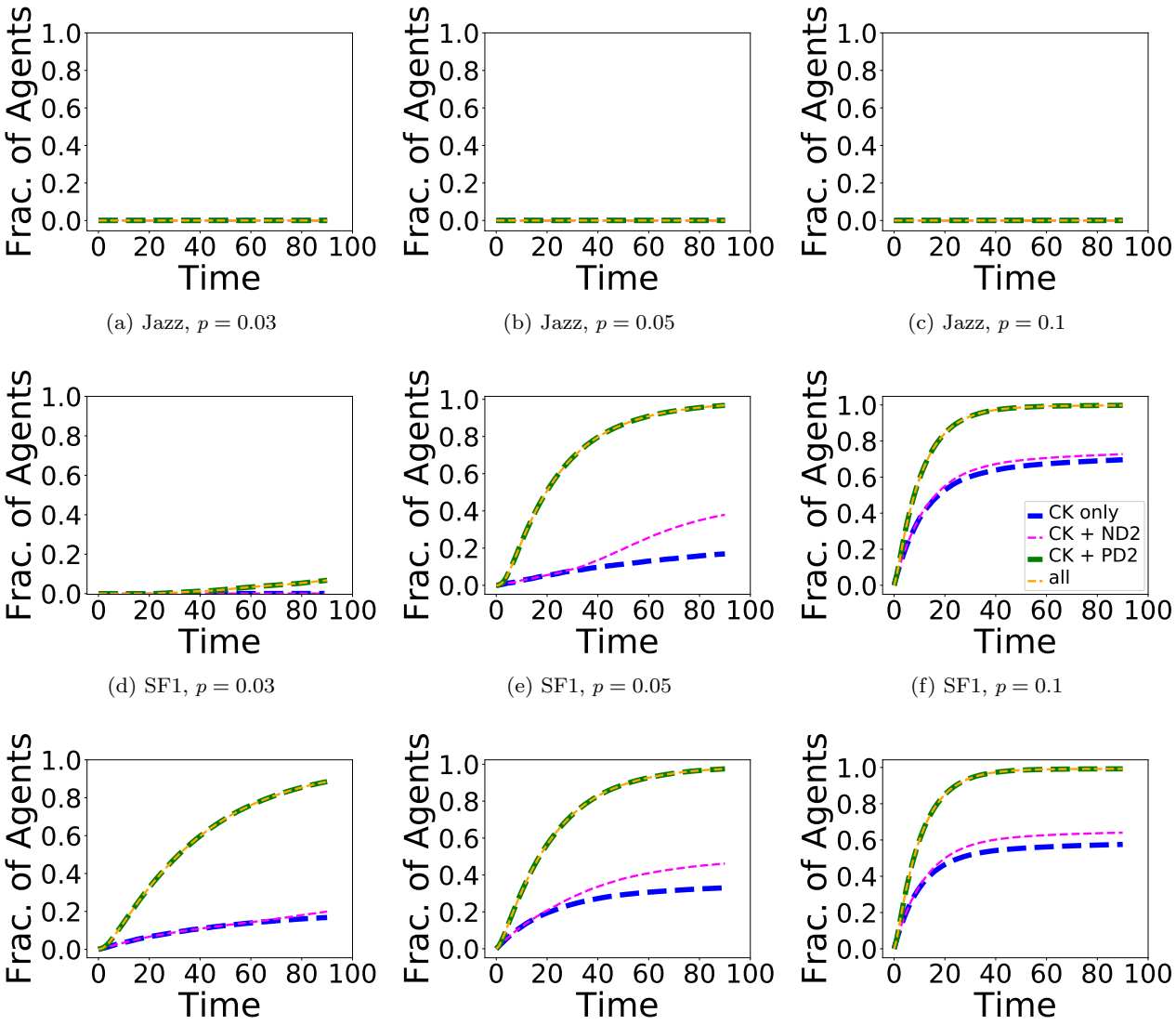
Results Section	Results Description
6.3.1	Effects of CKF model mechanisms on contagion dynamics for Wiki network.
6.3.2	Effects of online probability p on contagion dynamics for Wiki network.
6.3.3	Effects of CKF model mechanisms on contagion dynamics across all networks.
6.3.4	Explaining CKF contagion spreading across networks with network structure.
6.3.5	Saturated behavior governed by online probability p.
6.3.6	Strength of CKF mechanisms with changes in online probabilities p.
6.3.7	The effect of CK-only mechanism on contagion dynamics compared to the full model across networks.
6.3.8	Comparisons of final contagion spread at time $t = 30$ across networks and mechanisms.

6.3.2 Effects of online probability p on contagion dynamics for Wiki network

Staying with Figures 6g through 6i, as we move left to right in the plots where online probability p increases from 0.03 to 0.1, the curves rise, as expected. The intuition is straight-forward: as p increases, the expectation is that more agents (nodes) are online on Facebook and hence each engaged agent has more neighbors with whom to generate CK. Thus, each agent has a better chance of transitioning state to 1, and this results in greater fractions of agents in state 1.

6.3.3 Effects of CKF model mechanisms on contagion dynamics across all networks

Figures 6 through 8 contain time histories for the cumulative fractions of nodes in state 1 over time. Each row in each figure corresponds to results for one network, and the three columns in each figure correspond to $p = 0.03$, 0.05, and 0.1. The threshold is constant for



(a) Jazz, $p = 0.03$ (b) Jazz, $p = 0.05$ (c) Jazz, $p = 0.1$

(d) SF1, $p = 0.03$ (e) SF1, $p = 0.05$ (f) SF1, $p = 0.1$

(g) Wiki, $p = 0.03$ (h) Wiki, $p = 0.05$ (i) Wiki, $p = 0.1$

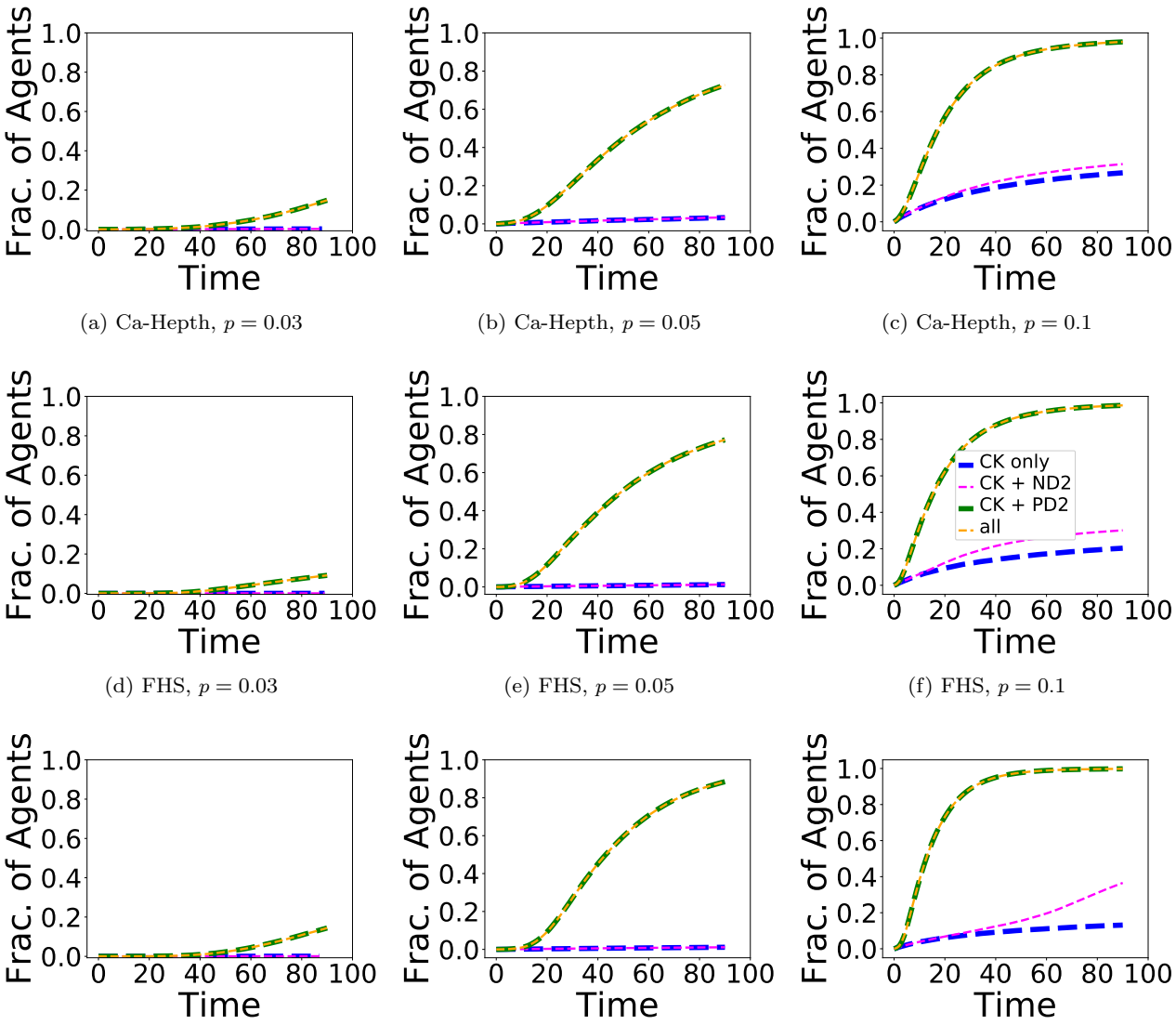
Fig. 6: Cumulative fraction of agents in state 1 plotted as a function of time in simulations, for combinations of different mechanisms (from Table 1). The threshold value is constant for each network, where $\theta = \lceil d_{ave} \rceil$. Each propagation mechanism is isolated for different simulations and is represented by a different curve; however, [CK] (labeled CK) is always operative. Online probability is $p = 0.03$, 0.05, and 0.1 in each of the three columns. The legend in plot (f) applies to all plots. The magenta and orange curves are purposely thinner than the blue and green curves, in order to compare curves when they overlay.

each network, but differs across networks; specifically, for each network, we let $\theta = \lceil d_{ave} \rceil$.

There is significant variation across networks. Some plots, like those for Jazz in Figures 6a through 6c, and those for FB in Figures 8g through 8i, show very little contagion spreading over 90 time steps. Other networks, like Wiki and Enron in Figures 6g through 6i and in Figures 8d through 8f, show large spread fractions.

Several networks show no spreading for $p = 0.03$ but show contagion spreading for $p \geq 0.05$. All three networks in Figure 7 demonstrate that $p = 0.05$ is sufficiently large to produce contagion spreading for [CK]+[PD2], but not for either [CK] or [CK]+[ND2].

In all of these plots, there is close association between the [CK] and [CK]+[ND2] curves, and between the [CK]+[PD2] and [ALL] curves. These support the conclusions in Section 6.3.1: the [PD2] mechanism con-



(a) Ca-Hepth, $p = 0.03$ (b) Ca-Hepth, $p = 0.05$ (c) Ca-Hepth, $p = 0.1$

(d) FHS, $p = 0.03$ (e) FHS, $p = 0.05$ (f) FHS, $p = 0.1$

(g) P2P, $p = 0.03$ (h) P2P, $p = 0.05$ (i) P2P, $p = 0.1$

Fig. 7: Cumulative fraction of agents in state 1 plotted as a function of time in simulations, for combinations of different mechanisms (from Table 1). The threshold value is constant for each network, where $\theta = \lceil d_{ave} \rceil$. Each propagation mechanism is isolated for different simulations and is represented by a different curve; however, [CK] (labeled CK) is always operative. Online probability $p = 0.03$, 0.05, and 0.1. The legend in plot (f) applies to all plots. The magenta and orange curves are purposely thinner than the blue and green curves, in order to compare curves when they overlay.

tributes more to the [CK] mechanism than does the [ND2] mechanism.

Results in this section and Section 6.3.1 informed the theoretical results in Section 4.3.

6.3.4 Explaining the behavior of CKF contagion spreading across networks

The issue is how to reason about ordering the networks from least to most with respect to the amount of contagion spread. A qualitative ordering, based on the amount of spreading and on spreading at lower p, is provided in Table 5. That is, these two criteria, over Figures 6 through 8, are used to rank and group networks

Table 5: Ranking of networks by amount of contagion spreading into four categories, in increasing order, and the corresponding value of d_{max}/d_{ave}.

Spreading Category	**Networks**	d_{max}/d_{ave}
Least spreading	Jazz, FB	3.61, 26.87
Intermediate spreading	FHS, Ca-Hepth, P2P	10.97, 11.32, 14.03
Greater spreading	SF1	14.84
Greatest spreading	Astroph, Wiki, Enron	22.91, 37.63, 129.3

into the four categories in the first column of Table 5. For example, the "least spreading" category contains the two networks with very little contagion spread for $p \leq 0.05$. The networks in the "greatest spreading" category are those with significant contagion spreading for $p = 0.03$. For each network, d_{max}/d_{ave} is provided.

The d_{max}/d_{ave} ratio does a good job of reproducing this ranking, meaning that contagion spreading correlates with this ratio. The greater the value of d_{max}/d_{ave}, the greater the amount of contagion spreading.

The motivation for this ratio is the following. The denominator is the threshold in Figures 6 through 8. This is because a CK set size must be at least $\theta + 1$, so that the threshold of each node in a CK set can be satisfied. The numerator is one less than the size of the largest CK set in a network. That is, in these networks, the largest biclique is the star subgraph where the hub node is the node with greatest degree (compare columns d_{max} and s_{bc}^{max} in Table 2). This is determined by data analysis. For a given p and network, the CK set that has the best opportunity to initiate contagion, in expectation, is the largest biclique. So the larger the value of d_{max}/d_{ave}, the more likely the contagion is to start spreading. The same basic results in Table 5 are obtained when using the difference $(d_{max} - d_{ave})$ rather than the ratio (d_{max}/d_{ave}).

The exception to this observation is FB. It has a large $d_{max}/d_{ave} = 26.87$, but it has the second smallest spreading among all networks in Figures 6 through 8. So this exception is a big one. We conjecture that the reason the spreading is so small in FB is that the average size of a biclique (that is not a star subgraph) is $s_{bbc} = 7.1 < d_{ave} = 8.3$, per Table 2. There are a lot of these bicliques (see column n_{bbc}) in FB. Thus, there are a lot of bicliques that are too small to initiate and propagate contagion, leading to smaller spread sizes.

An interesting outcome of this set of results is how for the CKF model, one may reason about simpler network structural quantities like d_{max}/d_{ave} in order to explain contagion dynamics.

6.3.5 Saturated behavior governed by online probability

Figures 6 through 8 provide the curves for cumulative fractions of nodes in state 1 for all networks up through $p = 0.1$. Not all networks exhibited saturated fractions of nodes in state 1, so the issue is to understand saturation behavior.

Figure 9 shows three plots that provide the three categories of behavior, all at $p = 0.4$. Thresholds are still $\theta = \lceil d_{ave} \rceil$. Most networks show the behavior displayed in Figure 9a. These networks show saturation for all curves, and as stated earlier, the curves with [PD2] attain greater saturation levels than curves with [ND2].

Figure 9b presents behavior that applies to Jazz, P2P, and Astroph. In these results, the [ND2] mechanism does provide additional contagion spread, beyond that produced by [CK] alone, which is not the case in Figure 9a. Admittedly, the amount of additional spread is not large, but it does contribute to contagion spread.

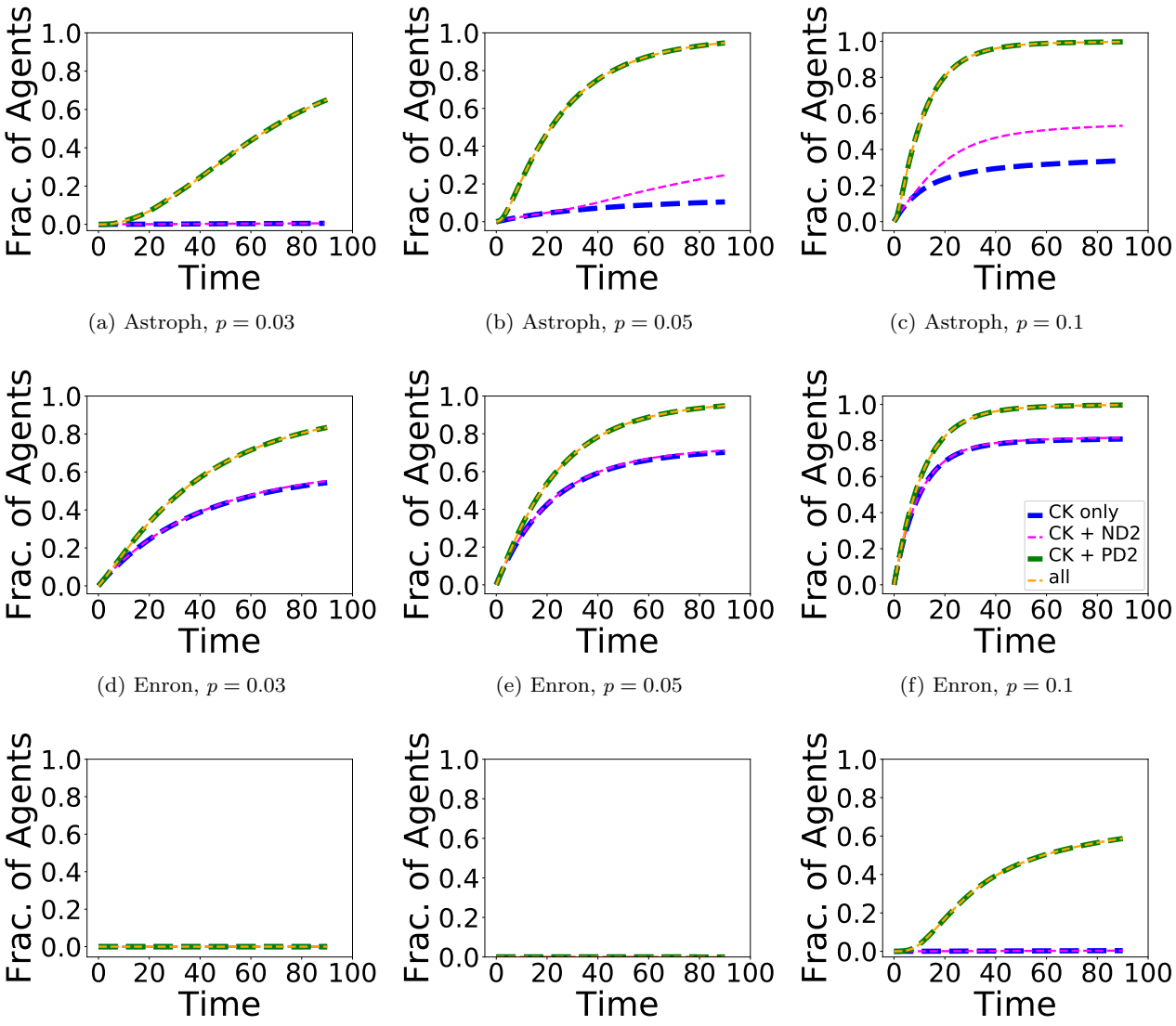
Figure 9c applies only to FB, which has the second-to-least spreading of any network (Jazz has the least spreading for $p \leq 0.1$). In fact, this behavior is somewhat surprising. Comparing Jazz and FB in Figures 6a through 6c and in Figures 8g through 8i, we see that Jazz has less spreading. So, Figure 9 demonstrates that the spreading in Jazz overtakes that of FB as p increases from 0.1 to 0.4.

6.3.6 Strength of CKF mechanisms with changes in online probabilities p.

In Figures 6 through 8, and in Figure 9, it is clearly observed that as p increases from 0.03 to 0.4, the differences between the [CK] curves and the [CK]+[PD2] curves decrease. This is explained formally Section 4.4. When p is smaller, e.g., $p \approx 0.05$, few nodes are online. Accordingly, the [PD2] mechanism has the advantage in that a node u needs only one node v that is online and in state 1 for u to transition to state 1. However, as p increases to 0.4, there are many more nodes that are online, and so conditions are conducive for these many nodes of CK sets to be online and thereby transition via the [CK] mechanism.

6.3.7 The effect of CK-only mechanism on contagion dynamics compared to the full model across networks

We analyze the fraction of activated nodes over time under the [CK]-only mechanism and under the [ALL] mechanisms of the CKF model combined. We use a smaller threshold, $\theta = 8$, for all networks. This fixed threshold is used in order to compare behaviors of different networks with different d_{ave} values. Five of the



(a) Astroph, $p = 0.03$ (b) Astroph, $p = 0.05$ (c) Astroph, $p = 0.1$

(d) Enron, $p = 0.03$ (e) Enron, $p = 0.05$ (f) Enron, $p = 0.1$

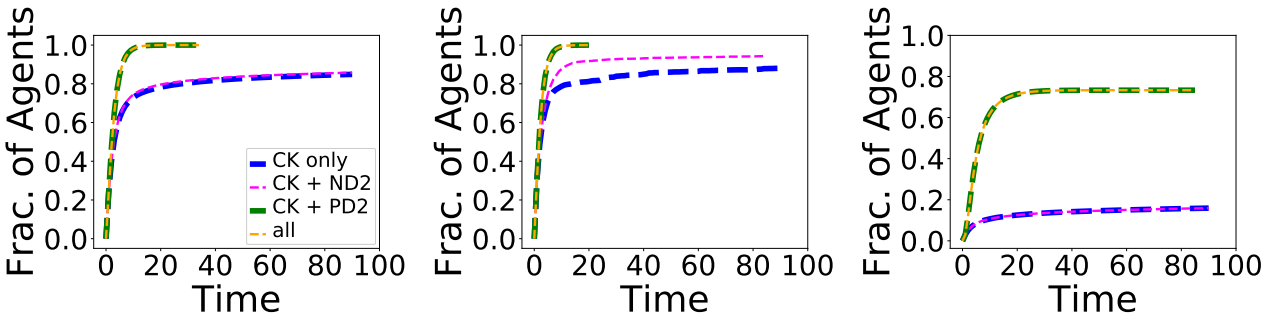
(g) FB, $p = 0.03$ (h) FB, $p = 0.05$ (i) FB, $p = 0.1$

Fig. 8: Cumulative fraction of agents in state 1 plotted as a function of time in simulations, for combinations of different mechanisms (from Table 1). The threshold value is constant for each network, where $\theta = \lceil d_{ave} \rceil$. Each propagation mechanism is isolated for different simulations and is represented by a different curve; however, [CK] (labeled CK) is always operative. The four largest networks from Table 2 are shown in rows and different p are in columns. The legend in plot (f) applies to all plots. The magenta and orange curves are purposely thinner than the blue and green curves, in order to compare curves when they overlap.

nine networks have d_{ave} values near or less than $\theta = 8$; the other four networks have $d_{ave} > \theta$. Figure 10 provides results for all networks.

The four networks with the smallest d_{ave} all have flat curves (i.e., little spreading) for the [CK] mechanism (see the plots in the first column). Enron, with $d_{ave} = 10.7$, has much greater spreading for [CK] than does Jazz with a much greater $d_{ave} = 27.7$; but Enron has a much greater d_{max}. This behavior is thus consistent with the explanation offered above.

The plots of the second column of Figure 10 illustrate that adding the [PD2] and [ND2] mechanisms to the [CK] mechanism generates contagion in three of the four networks that showed no contagion with [CK] alone: FHS, Ca-Hepth, and P2P. But the spreading in FB remains minimal.



(a) Ca-Hepth, $p = 0.4$ (b) Jazz, $p = 0.4$ (c) FB, $p = 0.4$

Fig. 9: Cumulative fraction of agents in state 1 plotted as a function of time in simulations, for combinations of different mechanisms (from Table 1). The threshold value is constant for each network, where $\theta = \lceil d_{ave} \rceil$. Results are for three representative behaviors for $p = 0.4$, which is greater than the p values in Figures 6, 7, and 8 (values of 0.03, 0.05, and 0.1). These curves demonstrate saturation in the fraction of activated agents. Each propagation mechanism is isolated for different simulations and is represented by a different curve; however, [CK] (labeled CK) is always operative. (a) all networks show this behavior, with the following exceptions. (b) Jazz, P2P, and Astroph show this behavior where the [CK+ND2] is slightly above the [CK] curve. (c) FB is alone in showing such a large discrepancy between curves with and without the [PD2] mechanism. The legend in plot (a) applies to all plots.

6.3.8 Comparisons of final contagion spread at time $t = 30$ across networks and mechanisms

Figure 11 provides spread fractions at $t = 30$ for all nine networks under different combinations of mechanisms (specified on x-axis): from left to right, [CK] only, [CK]+[ND2], [CK]+[PD2], and [ALL]. The uniform threshold θ for each network is given by $\theta = \lceil d_{ave} \rceil$. So, θ is different across networks. In each plot, curves are for $p = 0.05$, 0.1, 0.2 and 0.4.

Error bars are shown in each plot, and the Jazz network shows large variations in Figure 11a. This is because the combination of $\theta = 28$ and $p = 0.2$ for Jazz are such that contagion is just on the cusp of spreading. So, some runs produce contagion, and others do not, leading to significant standard deviation in results. The error bars for all conditions for all other networks show very little variance.

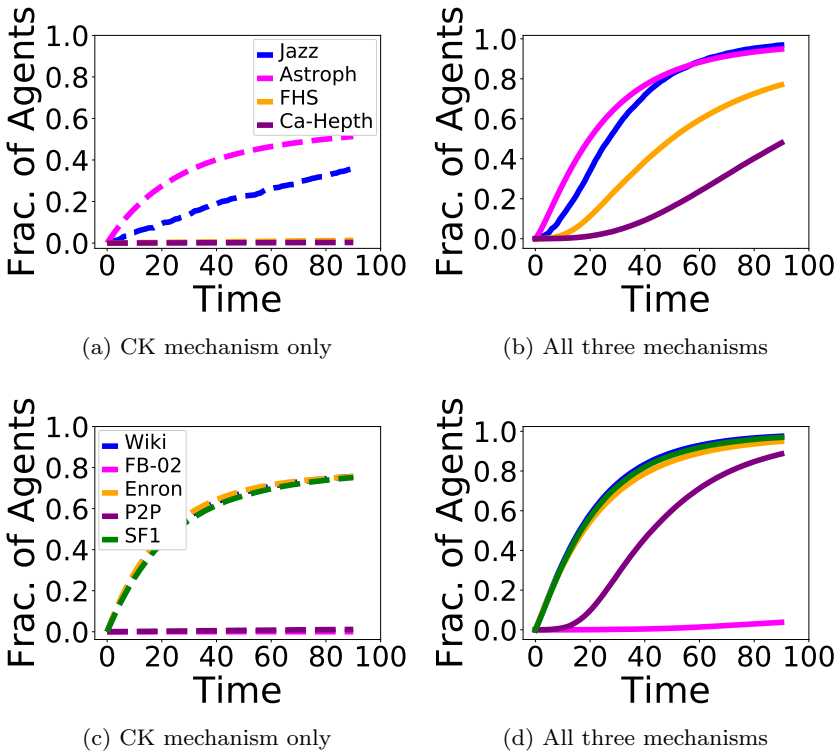
FB of Figure 11i has the smallest spread sizes. The [CK] mechanism in isolation can drive contagion through appreciable fractions of the other networks, depending on p. The largest effects of the [ND2] mechanism on spread fractions occur for Astroph, FHS, P2P, and Wiki, but only for particular p. In all the nine plots of Figure 11, the [PD2] mechanism contributes significantly to the driving force for contagion spread (the positive slopes of curves from "+ND2" to "+PD2" on the x-axis), except perhaps when [CK] or [CK]+[ND2] produce very large spread fractions. Finally, we observe that the curves are flat in going from "+PD2" to "All" on the x-axis, where the difference is the addition of the [ND2] mechanism.

Results in this section informed the theoretical results in Sections 4.3 and 4.4.

7 Conclusions

In this paper, we evaluate the Common Knowledge on Facebook (CKF) contagion model on a set of nine networks with wide ranging properties, for a range of thresholds and online probabilities. We model and investigate multiple mechanisms of contagion spread (initiation and propagation), as well as the full model. We find evidence that the [CK] and [PD2] mechanisms are the major driving forces for contagion initiation and spread, compared to the [ND2] mechanism. These types of results are being used to specify conditions for impending human subject experiments that will evaluate CK and its mechanisms (e.g., Korkmaz et al. (2018a)), and will be used to assess the predictive ability of the models.

The theoretical results in Section 4 were guided by the simulation results of Section 6. For example, the observations of the role of the [ND2] mechanism directed us to conjecture that this mechanism was weaker than the [PD2] mechanism, and in fact that it may be dominated by [PD2]. Section 4 shows that this is indeed the case. Thus, the theoretical results provide rigorous proofs of observed behaviors, and the simulation results of Section 6 provide examples, across a wide range of



(a) CK mechanism only (b) All three mechanisms

(c) CK mechanism only (d) All three mechanisms

Fig. 10: Cumulative fraction of agents in state 1 as a function of simulation time, for $p = 0.05$ and for $\theta = 8$. The nine networks are arbitrarily broken up such that each network occurs in plots on either the first row or the second row (9 curves on one plot is too cumbersome). In the first column, curves are for the CK mechanism only. In the second column, curves are for [ALL] mechanisms of the model. The results show the sensitivity of outbreak size on average degree d_{ave}. The average degrees for FHS, Ca-Hepth, FB and P2P have $d_{ave} \leq \theta$ (d_{ave} for FB is slightly greater than $\theta = 8$); these networks have small outbreaks due to CK only. The remaining networks all have $d_{ave} > \theta$, and these networks have larger spread fractions. The legend in plot (a) applies to plot (b). The legend in plot (c) applies to plot (d).

networks, of the particulars of spread fractions for different mechanisms and simulation conditions.

Ongoing work includes extending the model to directed networks and to other media platforms.

Acknowledgments: We thank the reviewer for providing helpful comments. This work is partially supported by University of Virginia Strategic Investment Fund award number SIF160, by NSF Grants CMMI-1916670 (CRISP 2.0), ACI-1443054 (DIBBS), IIS-1633028 (BIG DATA), CMMI-1745207 (EAGER), OAC-1916805 (CINES), CCF-1918656 (Expeditions), and IIS-1908530, and by the Air Force Office of Scientific Research under award number FA9550-17-1-0378. Any opinions, finding, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Air Force.

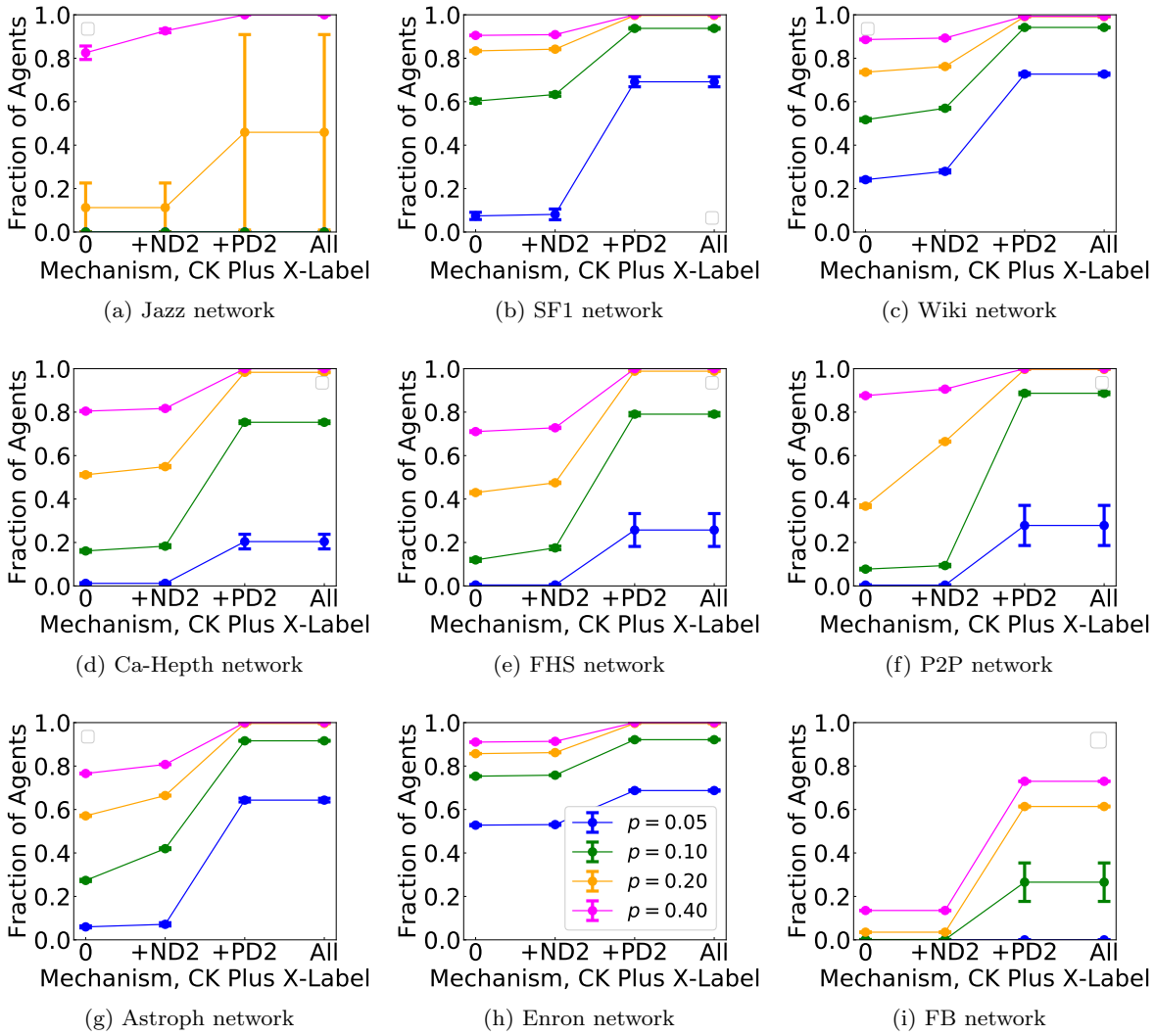
References

Adamic LA, Lento TM, Adar E, Ng PC (2016) Information evolution in social networks. In: WSDM, pp 473–482

Ahmed NK, Alo RA, Amelink CT, Baek YY, Chaudhary A, Collins K, Esterline AC, Fox EA, Fox GC, Hagberg A, Kenyon R, Kuhlman CJ, Leskovec J, Machi D, Marathe MV, Meghanathan N, Miyazaki Y, Qiu J, Ramakrishnan N, Ravi SS, Rossi RA, Sosic R, von Laszewski G (2020) net.science: A cyberinfrastructure for sustained innovation in network science and engineering. In: Gateway Conference

Aral S, Muchnik L, Sundararajan A (2013) Engineering social contagions: Optimal network seeding in the presence of homophily. Network Science 1:125–153

Backstrom L, Kleinberg J, Lee L, et al. (2013) Characterizing and curating conversation threads: Expansion, focus, volume, and re-entry. In: WSDM



(a) Jazz network (b) SF1 network (c) Wiki network

(d) Ca-Hepth network (e) FHS network (f) P2P network

(g) Astroph network (h) Enron network (i) FB network

Fig. 11: CKF model results. Cumulative fraction of agents in state 1 at time $t = 30$ as a function of mechanisms and p (same legend for all plots) for $\theta = \lceil d_{ave} \rceil$: (a) Jazz, $\theta = 28$. (b) SF1, $\theta = 19$; (c) Wiki, $\theta = 29$. (d) Ca-Hepth, $\theta = 6$; (e) FHS, $\theta = 8$; (f) P2P, $\theta = 8$; (g) Astroph, $\theta = 22$; (h) Enron, $\theta = 11$; and (i) FB, $\theta = 9$; The mechanisms on the x-axis always includes [CK] over all 30 time steps, where "0" corresponds to only the [CK] mechanism; "+ND2" means [CK]+[ND2]; "+PD2" means [CK]+[PD2]; and "All" means the [ALL] (full) model. The legend in plot (h) applies to all plots. The error bars for y-axis values represent one stdev. The data illustrate that [PD2] provides a much greater driving force for contagion spread than does [ND2]. Although [PD2] often generates a greater contribution to driving force than does [ND2], the latter can generate significant additional contagion. See for example the curve for P2P and $p = 0.2$.

Bakshy E, Rosenn I, Marlow C, Adamic L (2012) The role of social networks in information diffusion. In: WWW, pp 519–528

Barabasi A, Albert R (1999) Emergence of scaling in random networks. Nature 286:509–512

Centola D (2010) The spread of behavior in an online social network experiment. Science 329:1194–1197

Centola D (2011) An experimental study of homophily in the adoption of health behavior. Science 1269:1269–1272

Centola D, Macy M (2007) Complex contagions and the weakness of long ties. Am J Soc 113(3):702–734

Centola D, Eguiluz V, Macy M (2006) Cascade Dynamics of Complex Propagation. Physica A 374:449–456

Chen G, Chen BC, Agarwa D (2017) Social incentive optimization in online social networks. In: WSDM, pp 547–556

Cheng J, Adamic LA, Dow PA, Kleinberg J, Leskovec J (2014) Can cascades be predicted? In: WWW

Chwe MSY (1998) Culture, circles, and commercials publicity, common knowledge, and social coordination. Rationality and Society 10(1):47–75

Chwe MSY (1999) Structure and strategy in collective action. American Journal of Sociology 105:128–156

Chwe MSY (2000) Communication and coordination in social networks. Rev of Econ Stud 67:1–16

Devineni P, Koutra D, Faloutsos M, Faloutsos C (2015) If walls could talk: Patterns and anomalies in facebook wallposts. In: ASONAM, pp 367–374

Dodds PS, Watts DJ (2005) A generalized model of social and biological contagion. J of Theo Bio 232(4):587–604

Dow PA, Adamic LA, Friggeri A (2013) The anatomy of large facebook cascades. In: ICWSM, pp 145–154

Gonzalez-Bailon S, Borge-Holthoefer J, Rivero A, Moreno Y (2011) The dynamics of protest recruitment through an online network. Scientific Reports pp 1–7

Granovetter M (1978) Threshold Models of Collective Behavior. Am J of Soc 83(6):1420–1443

Hagberg AA, Schult DA, Swart PJ (2008) Exploring network structure, dynamics, and function using NetworkX. In: Proceedings of the 7th Python in Science Conference (SciPy2008), pp 11–15

He X, Liu Y (2017) Not enough data? Joint inferring multiple diffusion networks via network generation priors. In: Proceedings of the 10th ACM Symposium on Web Search and Data Mining (WSDM), pp 465–474

Hodas NO, Lerman K (2014) The simple rules of social contagion. Scientific Reports 4

Huang TK, Rahman MS, Madhyastha HV, Faloutsos M, et al. (2013) An analysis of Socware cascades in online social networks. In: WWW, pp 619–630

Kempe D, Kleinberg J, Tardos E (2003) Maximizing the spread of influence through a social network. In: KDD, pp 137–146

Korkmaz G, Kuhlman CJ, Marathe A, et al. (2014) Collective action through common knowledge using a Facebook model. In: AAMAS

Korkmaz G, Kuhlman CJ, Ravi SS, Vega-Redondo F (2016a) Approximate contagion model of common knowledge on Facebook. In: Hypertext, pp 231–236

Korkmaz G, Kuhlman CJ, Vega-Redondo F (2016b) Can social contagion spread without key players? In: BESC

Korkmaz G, Capra M, Kraig A, Lakkaraju K, Kuhlman CJ, Vega-Redondo F (2018a) Coordination and common knowledge on communication networks. In: Proceedings of the AAMAS Conference, July 10-15, Stockholm, Sweden, 1062–1070

Korkmaz G, Kuhlman CJ, Ravi SS, Vega-Redondo F (2018b) Spreading of social contagions without key players. World Wide Web Journal 21:1187–1221

Kramer ADI, Guillory JE, et al. (2014) Experimental evidence of massive-scale emotional contagion through social networks. PNAS 111(24):8788–8790

Kuhlman CJ, Korkmaz G, Ravi SS, Vega-Redondo F (2020) Effect of interaction mechanisms on facebook dynamics using a common knowledge model. In: International Conference on Complex Networks and Their Applications (COMPLEX NETWORKS), pp 395–407

Leskovec J, Krevl A (2014) SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data

Leskovec J, Sosič R (2016) SNAP: A general-purpose network analysis and graph-mining library. ACM Transactions on Intelligent Systems and Technology (TIST) 8(1):1

Liu G, Sim K, Li J (2006) Efficient mining of large maximal bicliques. In: LNCS 4081, Conf. DaWak 2006, pp 437–448

Oliver P (1993) Formal models of collective actions. Annual Review of Soc 19:271–300

Prissner E (2000) Bicliques in graphs I: Bounds on their number. Combinatorica 20(1):109–117

Romero D, Meeder B, Kleinberg J (2011) Differences in the mechanics of information diffusion. In: WWW

Romero D, Reinecke K, Robert L (2017) The influence of early respondents: Information cascade effects in online event scheduling. In: WSDM

Schelling T (1960) The Strategy of Conflict. Harvard U. Press

Schelling T (1971) Dynamic models of segregation. Journal of Mathematical Sociology 1:143–186

Schelling T (1978) Micromotives and Macrobehavior. W. W. Norton and Company

Siegel D (2009) Social networks and collective action. American Journal of Political Science 53:122–138

Siegel D (2010) When does repression work? collective action in social networks. The Journal of Politics 73:993–1010

Sun E, Rosenn I, Marlow CA, Lento TM (2009) Gesundheit! modeling contagion through facebook news feed. In: ICWSM

Susarla A, Oh JH, Tan Y (2012) Social networks and the diffusion of user-generated content: Evidence from youtube. Info Sys Research 23:23–41

Upadhyay U, Valera I, Gomez-Rodriguez M (2017) Uncovering the dynamics of crowdlearning and the value of knowledge. In: WSDM, pp 61–70

Viswanath B, Mislove A, Cha M, Gummadi KP (2009) On the evolution of user interaction in Facebook. In: WOSN

Watts D (2002) A simple model of global cascades on random networks. Proceedings of the National Academy of Sciences (PNAS) 99(9):5766–5771