

cascade Behaviours and Network Effects

LEARNING OBJECTIVES

After completing the chapter, the readers are expected to

- Learn how to model an information diffusion pattern for predictive analysis.
- Understand the evolution of dynamic models for information diffusion.
- Compare the pros and cons of existing diffusion and cascade models.
- Design a new diffusion model for a spreading phenomenon.

Information spread on the social network through various user responses such as retweets/shares, likes/dislikes, upvotes/downvotes, and gradually form *information cascade*. Studying information cascade thus deals with modeling the flow of dynamic information in networks. It explains how information such as current affairs, opinions, rumors, fake news, and misinformation spread across nodes in a social network. Throughout the history, there have been several instances that have motivated us to study how content diffuses on a social network, or how individuals start exhibiting a *herd behaviour* (getting influenced by their immediate neighbours). This includes empirical studies in the field of sociology, termed as *diffusion of innovations*. Ryan and Gross (1943) presented a classical study of information diffusion addressing the adoption of hybrid seed corn by farmers in Iowa. They were influenced by their neighbours in the community network. Coleman et al. (1966) represented the process of adoption of a new drug by doctors in Illinois. It turned out that, along with the clinical evaluations, assurance from the social peer connections was also responsible for the doctors to adopt the drug, establishing that the success/failure of an innovation is highly guided by the structure of the network formed by the initial adopters.

There are numerous real-world scenarios that come under the category of information diffusion in networks. Many times, a social media post goes viral and forms a propagation tree. For instance, the chain of posts that were shared during the #MeToo movement (a social movement that was not initiated by a centralised authority) formed the different cascade patterns. Usually people get influenced by their peers in following certain trends, e.g., course recommendations while selecting elective courses during graduate studies, viral marketing of a product among people belonging to a common age group, or the psychological effect of the people's opinions.

Even in the medical domain, the propagation of a disease or an epidemic (e.g., the recent COVID-19 pandemic) comes under cascading behaviour of networks. There is a class of information cascade models wherein the decision to participate in the diffusion of information is driven by the attainment of direct benefits/payoffs such as the acceptance of mobile phones (better means of communication) worldwide.

At times, it becomes difficult for information to propagate within a community of individuals with very similar characteristics. This is primarily since information on innovations usually originates at an external source that might be opposed collectively by the community members, in turn causing the cascade to fail. On the other hand, most of the cascades with high depth are originated and regulated by groups of individuals, intending to spread certain propaganda. These groups of users form 'echo chambers' in a social network.

In this chapter, we begin by describing some basic terminologies that form the basis for modelling information cascades. We then dig deeper into the class of decision-based cascade models covering both single and multiple-choice decision models. Furthermore, we discuss a real-world case study to motivate the interest of readers in this topic. Next, we look into the theoretical aspect of probabilistic models followed by epidemic models and a brief discussion about the influence of external exposure on cascade networks. In the end, we touch upon cascade prediction discussing some state-of-the-art approaches based on the recent neural networks.

7.1 PRELIMINARIES AND IMPORTANT TERMINOLOGIES

Some basic terminologies that form the basis for modelling information cascades are listed as follows:

- Contagion:** An entity that spreads across a network. Figure 7.1(a) provides a representation of the same.

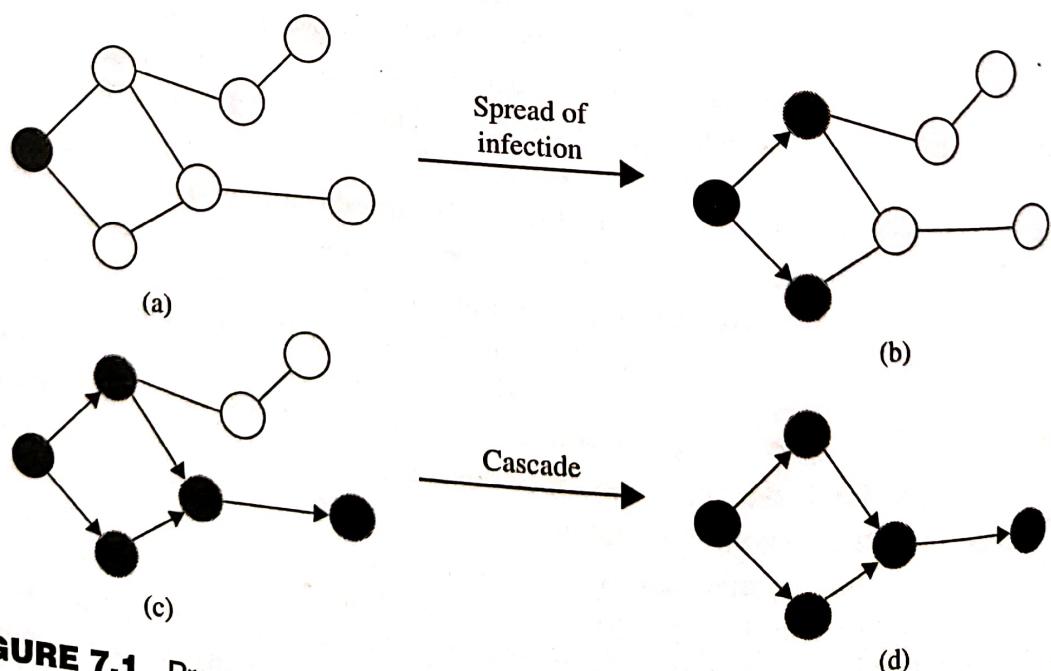


FIGURE 7.1. Propagation of cascade through a network.

2. **Adoption:** This refers to the event of infection or diffusion. It is also known as 'activation'. Figure 7.1(b) provides a representation of the same.
3. **Adopters:** These represent the final set of infected nodes. Figure 7.1(c) provides a representation of the same.
4. **Cascade:** The final propagation tree obtained by the spread of the infection is known as a cascade. Figure 7.1(d) provides a representation of the same.

7.2 CASCADE MODELS

In this section, we explore two models that describe cascade behaviour on networks – (a) decision-based models, and (b) probabilistic models.

In *decision-based models*, given a network, each node has the freedom to decide whether to adopt a contagion or not. The decision at each node is influenced by the behaviour of nodes in its neighbourhood. For instance, an individual user may end up buying a product based on the recommendation received from her friends. On the contrary, in the case of *probabilistic models*, given a network, each node would have to adopt the contagion with some probability. For instance, in the present scenario, if a patient infected by coronavirus is present in a gathering, probably people who are in contact with them would also be affected.

In the following section, we shall go much deeper into understanding how the decision-based cascade models work, followed by the same for probabilistic models.

7.2.1 Decision-based Models

Decision-based cascade models are originated from the idea of *local interaction models*, described by Morris (2000). This model is based on the fact that nodes in a network decide to adopt a new behaviour or contagion driven by a *direct benefit* or *payoff*. For a node, the payoff obtained by adopting a contagion is directly proportional to the number of its neighbours that have adopted the same contagion.

We can explain this using a two-player *coordination game*.¹ For an undirected network $G(V, E)$, such that V represents the set of nodes and E represents the set of edges, let $u, v \in V$ and $\langle u, v \rangle \in E$. A and B represent the two possible behaviours/strategies that each node in G could adopt. For the coordination game in which u and v are the players, the payoff obtained for each adoption strategy is summarised in Table 7.1, showing a positive incentive for u, v in case they agree on their adoption decision.

Table 7.1 provides the payoff dynamics for one edge $\langle u, v \rangle$ in the network G . It should be noted that each node u will play its own independent game with all its neighbours, and its final payoff would be determined by the sum of payoffs for all the games (one game with each neighbour) played by it. Therefore, the *choice of the strategy adopted by node u would be influenced by the choices made by all of its neighbours*.

¹Given a number of strategies, the end goal of the players is to coordinate on the same strategy to maximise their payoffs.

TABLE 7.1. Payoff distribution for different adoption strategies in a two-player coordination game.

S. No.	u 's decision	v 's decision	Payoff
1.	A	A	a^*
2.	B	B	b^*
3.	A	B	0
4.	B	A	0

* a and b are positive (> 0) constants.

Let us further calculate the required threshold at which a node u would decide to go with strategy A . Figure 7.2 illustrates the same. Here, we assume that u has d neighbours such that p fraction of nodes adopt strategy A (i.e., dp number of nodes adopt A) and the remaining $(1 - p)$ fraction of nodes adopt strategy B (i.e., $d(1 - p)$ number of nodes adopt B). As a result of the earlier assumptions and the payoff distribution given in Table 7.1, the following is true,

1. adp is the total payoff for node u if it goes with strategy A .
2. $bd(1 - p)$ is the total payoff for node u if it goes with strategy B .

For u to adopt A , $adp \geq bd(1 - p)$, which further reduces to,

$$p \geq \frac{b}{a + b} \quad (7.1)$$

This shows that, if at least $\frac{b}{a + b}$ fraction of u 's neighbours adopt A , then node u would go forward with adopting behaviour A as well.

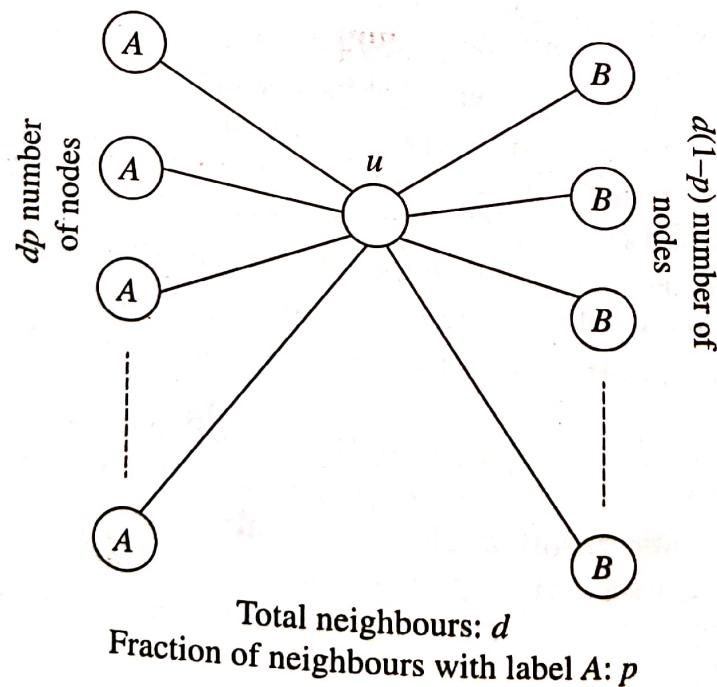


FIGURE 7.2. Node u would decide between strategies A and B based on its neighbours' decisions.

In order to understand how an information cascade is formed in a decision-based model, we present a toy example. Let us say, in India, most of the people use Amazon Prime Video as a service for streaming movies and TV episodes. A few of them visit the United States and switch their preference to Netflix, a decision taken regardless of what is reflected through the payoff/direct-benefit calculations. This situation is captured in the network represented by Figure 7.3, showing that the initial set of adopters (those who went to the United States) for Netflix is $U = \{A, G, I\}$. Also assume that the threshold (fraction of neighbours following the adoption strategy) for a switch from Amazon Prime Video to Netflix at a node is 0.50.

Initially, as shown in Figure 7.4(a), nodes B and H switch to Netflix since the fraction of their neighbours following Netflix is at least 0.50 (0.50 and 1.00, respectively). This further influences nodes C and J to make the change (fraction of neighbours following Netflix is 0.50 and 0.67, respectively).

Example 7.1

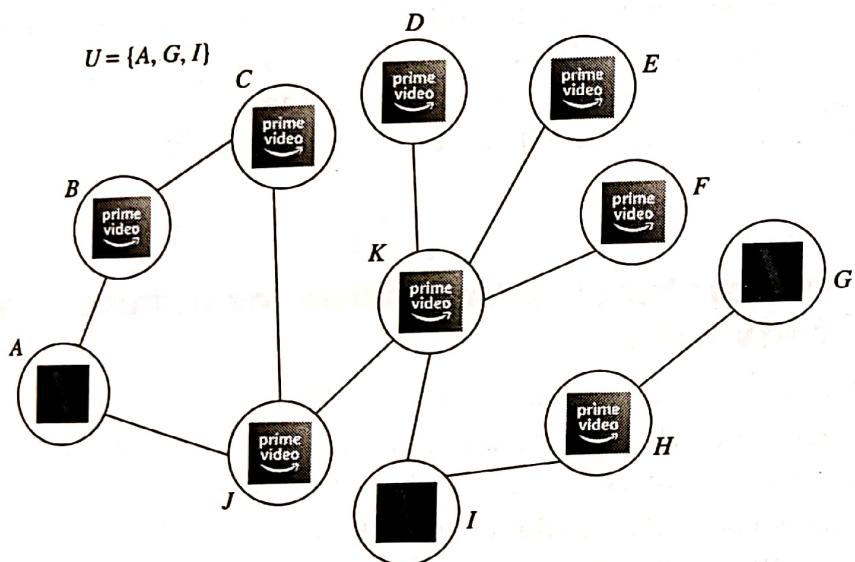


FIGURE 7.3. A toy example to show the propagation of a contagion obtained using the decision-based cascade model.

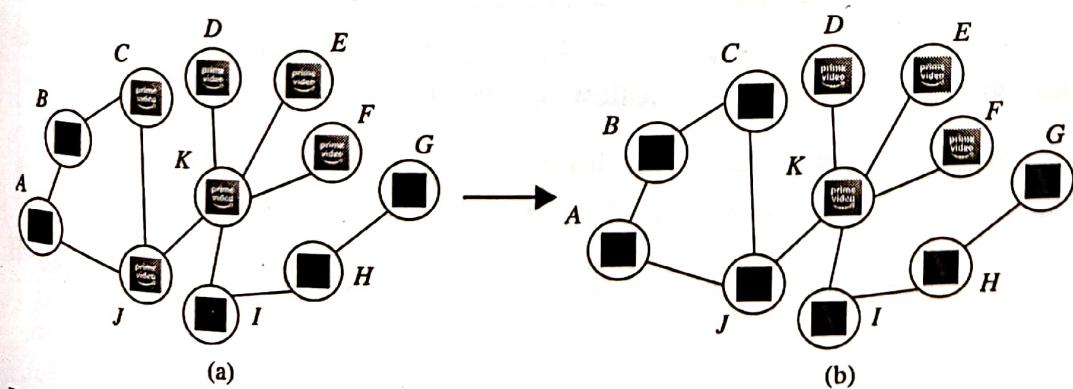


FIGURE 7.4. Propagation of the new "Netflix" contagion throughout the network.

TABLE 7.2. Payoff for a multiple-choice decision model.

S. No.	u 's decision	v 's decision	Payoff
1.	AB	A	a^1
2.	AB	B	b^1
3.	AB	AB	$\max(a, b)$

¹ a and b are positive (> 0) constants.

7.2.2 Multiple-choice Decision-based Model

Till now, we have only discussed a *single choice decision-based model*, the case wherein a user/node could only select/adopt one of the possible behaviours/strategies presented to her. In this section, we introduce the *multiple-choice decision-based model* that allows a node to adopt more than one strategy/behaviour. This means that, instead of selecting one among the two competing strategies, A and B , as described in Section 7.2.1, a node is allowed to adopt both behaviours. The payoff distribution for node u in network G becomes equivalent to Table 7.2.

In case a node prefers to go with both strategies A and B , it would incur an additional cost c . For instance, in Example 7.1, an individual would incur an extra charge in maintaining both the services, namely, Amazon Prime Video and Netflix, in the form of increased subscription charges, internet usage, mobile app storage, etc.

7.2.3 Understanding Cascades for Infinite Chain Networks

Let there be a network as shown in Figure 7.5 having an infinite chain of nodes following behaviour B , with the exception that a few of them are hard-wired to adopt behaviour A . Each node is allowed to adopt only one strategy among A and B . We would now establish a relation between the payoff constants a, b for behaviour A to spread through the entire network. Let us investigate this with the help of Example 7.2.

Example 7.2

Consider the case where $a = 3$ and $b = 2$. For the given network, node u , as marked in Figure 7.5, has the following two choices:

1. Stick with behaviour B , resulting in a total payoff of $0 + 2 = 2$.
2. Switch to behaviour A , resulting in a total payoff of $3 + 0 = 3$.

As a result, node u would adopt behaviour A and drop B to maximise its payoff (since $3 > 2$).

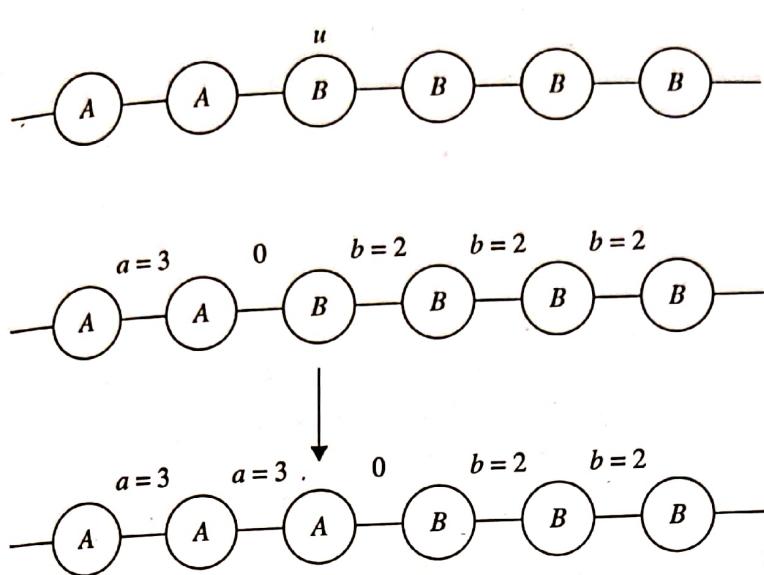


FIGURE 7.5. Infinite chain network – establishing a relation between a and b for behaviour A to cascade through the entire network with the strategy set $S = \{A, B\}$.

Through Example 7.2, we can generalise that, for a setting/network similar to the one shown in Figure 7.5, we would observe behaviour A to propagate through the entire network if $a \geq b$, or the threshold obtained previously in Section 7.2.1 is ≤ 0.50 .

Now, let us consider a case wherein a node could also choose to keep both behaviours A and B , i.e., multiple-choice decision-based model. Thus, the possible set of strategies now becomes $S = \{A, B, AB\}$. Again, we would establish a relation between the payoff constants a, b , and cost variable c for behaviour A to spread through the entire network. Let us investigate this with the help of Examples 7.3 and 7.4.

Consider the case where $a = 3, b = 2, c = 1$ as shown in Figure 7.6. For the given network, node u has the following three choices:

Example 7.3

1. Stick with behaviour B , resulting in a total payoff of $0 + 2 = 2$.
2. Switch to behaviour A , resulting in a total payoff of $3 + 0 = 3$.
3. Switch to behaviour AB , resulting in a total payoff of $3 + 2 - 1 = 4$.

As a result, node u would adopt behaviour AB to maximise its payoff ($4 > 2$). It should further be noted that the cascade would stabilise with no change in node behaviour after Figure 7.6(b). This is due to the fact that, for node v , one step further in the network chain, behaviour B would still bring the maximum possible

payoff of $2+2=4$ (behaviour A ($3+0=3$) and AB ($3+2-1=4$) do not result in a better/higher payoff), as represented by Figure 7.6(c) Alt. Choice (I) and (II).

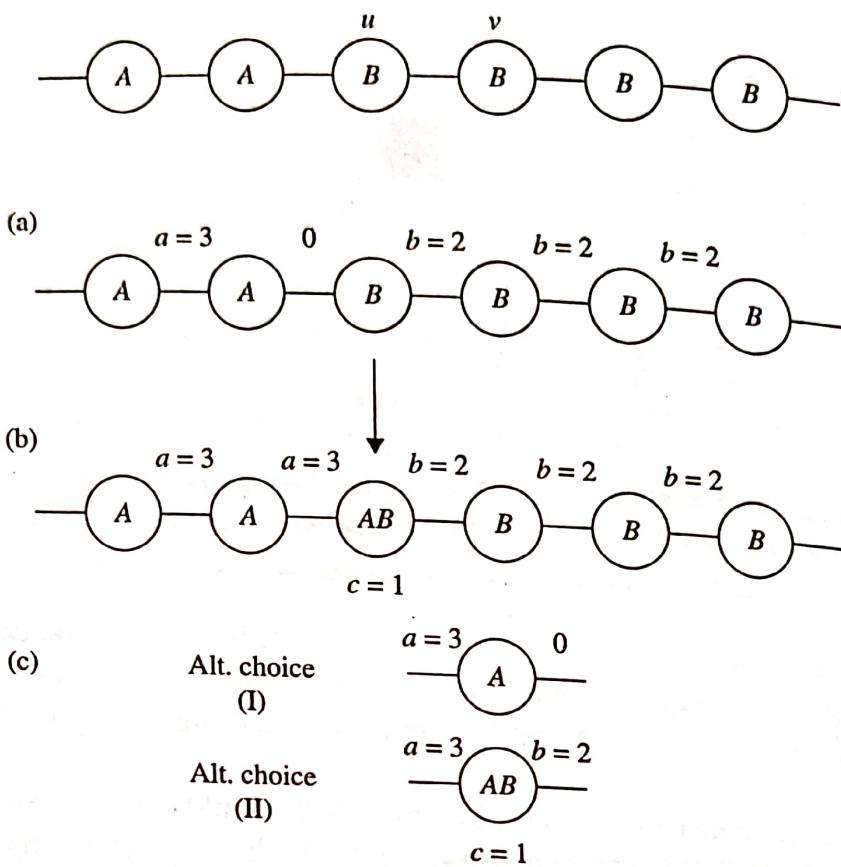


FIGURE 7.6. Infinite chain network – establishing a relation between a and b for behaviour A to cascade through the entire network with the strategy set as $S = \{A, B, AB\}$.

Example 7.4

Consider another case wherein $a = 5$, $b = 3$, $c = 1$, described in Figure 7.7. As shown in Figure 7.7(b), node u would switch from B to AB for a better payoff of 7 ($5+3-1$), instead of 3 ($0+3$) by sticking to B , or 5 ($5+0$) by switching to A . Similarly, in the next step presented in Figure 7.7(c), node v would also switch to behaviour AB (payoff from B : $3+3=6$; payoff from A : $5+0=5$; payoff from AB : $5+3-1=7$). Now, in the next iteration, Figure 7.7(d), node u would change its earlier adopted behaviour from AB (payoff = $5+5-1$) to A (payoff = $5+5$) in order to drop B for not incurring the extra cost c ($= 1$). As a result, behaviour A would propagate through the entire chain, switching each node's strategy in the following manner: $B \rightarrow AB \rightarrow A$.

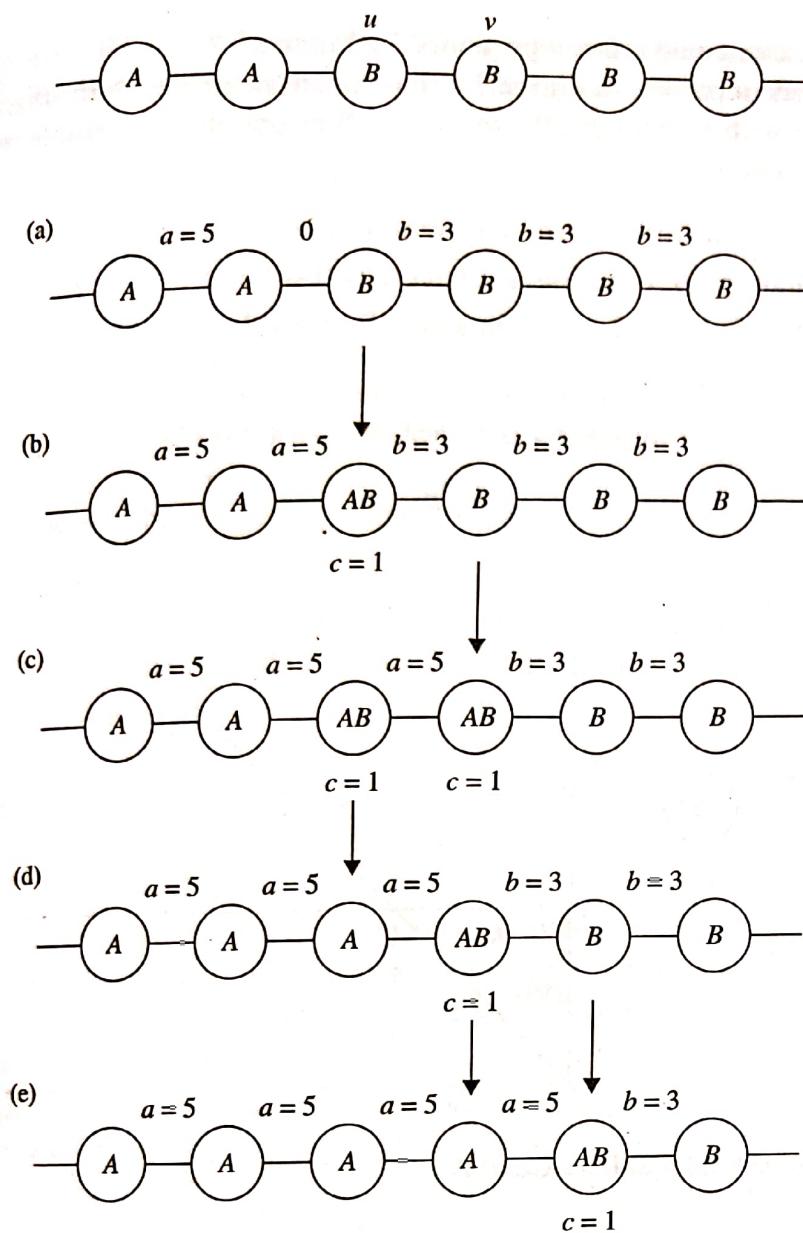


FIGURE 7.7. Infinite chain network – establishing a relation between a and b for behaviour A to cascade through the entire network with the strategy set $S = \{A, B, AB\}$.

Generic Model for Infinite Chain Networks

In this section, we would generalise the cascading behaviour for the network with infinite nodes as defined in the previous section with the strategy set $S = \{A, B, AB\}$. For simplicity, we would only consider the snapshot of the network to the right-hand side of nodes that are hard-wired to adopt strategy A (the same behaviour would be duplicated for nodes on the left-hand side). Given that $b = 1$, we would look for values of a and c for which strategy A would propagate through the entire network.

Now two cases could arise, represented by Figures 7.8 and 7.9.

Case A: For node u present in Figure 7.8 (having left neighbour with strategy A and right neighbour with strategy B), the following three options are possible in terms of adopting a contagion:

1. Adopt behaviour A , resulting in a payoff of $a + 0 = a$.
2. Adopt behaviour B , resulting in a payoff of $0 + 1 = 1$.
3. Adopt behaviour AB , resulting in a payoff of $a + 1 - c$.

Depending on payoffs obtained from options (1) and (2), node u could decide between A or B ; based on payoffs obtained from options (2) and (3), node u could

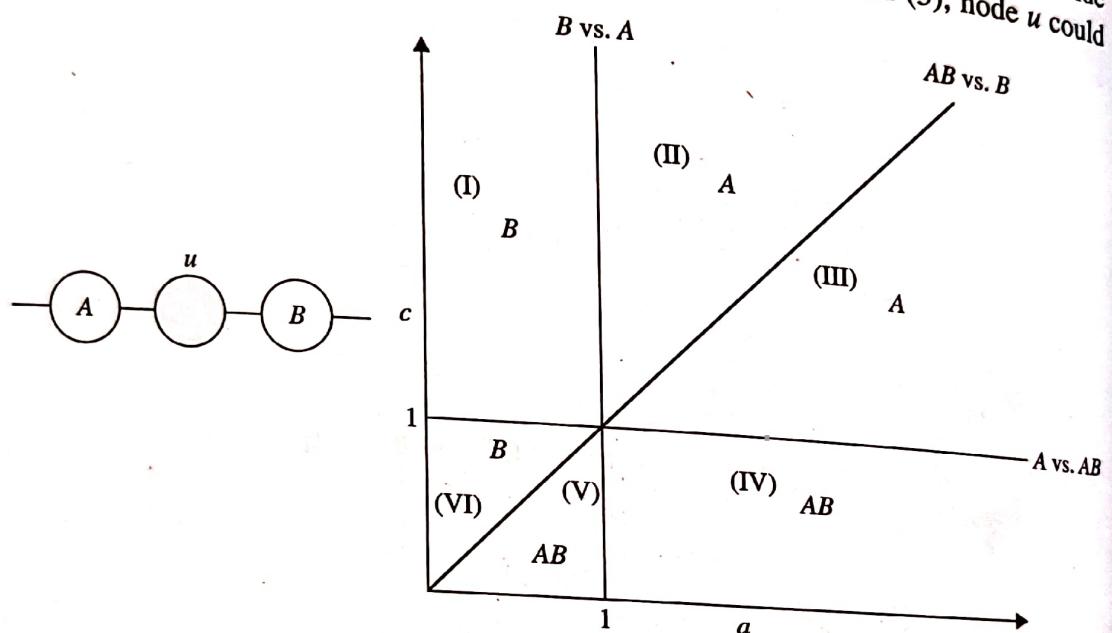


FIGURE 7.8. Generalisation: Case A.

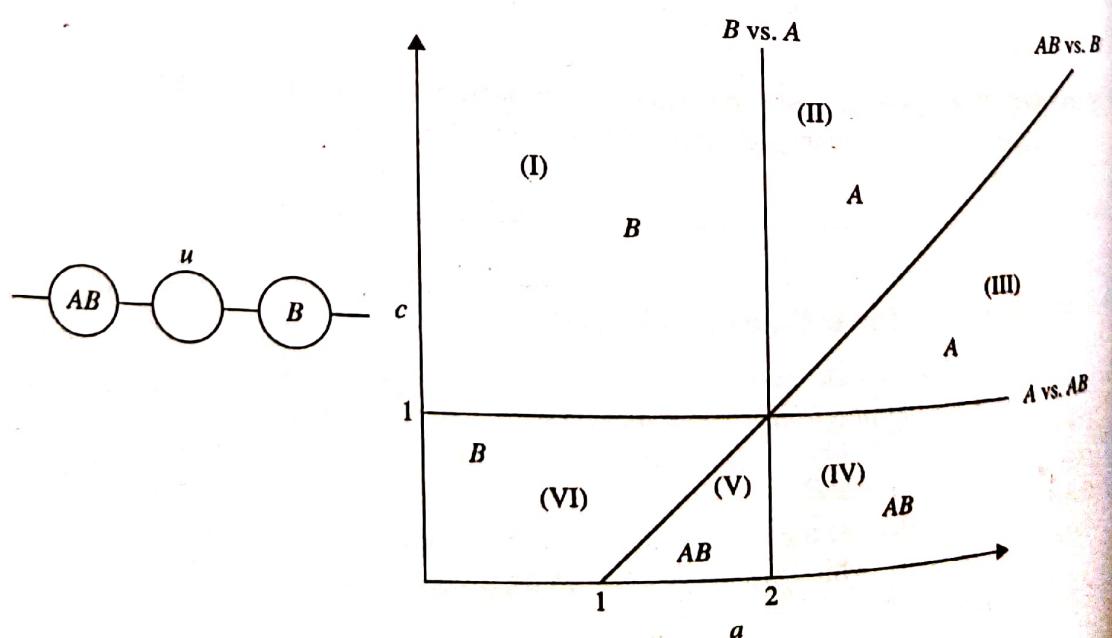


FIGURE 7.9. Generalisation: Case B.

decide between B or AB ; based on payoffs obtained from options (3) and (1), node u could decide between AB or A , giving rise to the following three break-point equations:

$$B \text{ vs. } A: a = 1; a < 1: \text{prefer } B, a > 1: \text{prefer } A \quad (7.2)$$

$$AB \text{ vs. } B: a = c; a < c: \text{prefer } B, a > c: \text{prefer } AB \quad (7.3)$$

$$A \text{ vs. } AB: c = 1; c < 1: \text{prefer } AB, c > 1: \text{prefer } A \quad (7.4)$$

With the help of Equations 7.2, 7.3 and 7.4, we can divide the (a, c) region into six parts, as shown in Figure 7.8.

Let us compute the behaviour adopted within the six regions obtained earlier:

1. **Region (I):** Since $a < 1, c > 1$ through Equations 7.2 and 7.4, we see that a choice has to be made between strategies B and A . In this case, B would be preferred. B would provide a higher payoff (i.e., 1) as compared to the payoff from A ($a < 1$).
2. **Region (II):** Since $a > 1, a < c$ through Equations 7.2 and 7.3, we see that a choice has to be made between strategies A and B . In this case, A would be preferred. A would provide a higher payoff (i.e., a) as compared to the payoff from B (i.e., 1) because $a > 1$.
3. **Region (III):** Since $c > 1, a > c$ through Equations 7.4 and 7.3, we see that a choice has to be made between strategies A and AB . In this case, A would be preferred. The payoff through AB would be $a + 1 - c$ that would be $< a$ (payoff obtained from strategy A) because $c > 1$ ($1 - c < 0$).
4. **Region (IV):** Since $a > 1, c < 1$ through Equations 7.2 and 7.4, we see that a choice has to be made between strategies A and AB . In this case, AB would be preferred. The payoff through AB would be $a + 1 - c$ that would be $> a$ (payoff obtained from strategy A) because $c < 1$ ($1 - c > 0$).
5. **Region (V):** Since $a < 1, a > c$ through Equations 7.2 and 7.3, we see that a choice has to be made between strategies B and AB . In this case, AB would be preferred. The payoff through AB would be $a + 1 - c$ that would be > 1 (payoff obtained from strategy B) because $a > c$ ($a - c > 0$).
6. **Region (VI):** Since $a < c, c < 1$ through Equations 7.3 and 7.4, we see that a choice has to be made between strategies B and AB . In this case, B would be preferred. The payoff through AB would be $a + 1 - c$ that would be < 1 (payoff obtained from strategy B) because $a < c$ ($a - c < 0$).

Case B: For node u present in Figure 7.9 (having left neighbour with strategy AB and right neighbour with strategy B), the following three options are possible in terms of adopting a contagion:

1. Adopt behaviour A , resulting in a payoff of $a + 0 = a$.
2. Adopt behaviour B , resulting in a payoff of $1 + 1 = 2$.
3. Adopt behaviour AB , resulting in a payoff of $a + 1 - c$, if $\max(a, 1) = a$.

Based on payoffs obtained from options (1) and (2), node u could decide between A or B ; based on payoffs obtained from options (2) and (3), node u could decide between B or AB ; based on payoffs obtained from options (3) and (1), node u could decide between AB or A , giving rise to the following three break point equations:

$$B \text{ vs. } A: a = 2; a < 2: \text{prefer } B, a > 2: \text{prefer } A \quad (7.5)$$

$$AB \text{ vs. } B: a - c = 1; a - c < 1: \text{prefer } B, a - c > 1: \text{prefer } AB \quad (7.6)$$

$$A \text{ vs. } AB: c = 1; c < 1: \text{prefer } AB, c > 1: \text{prefer } A \quad (7.7)$$

With the help of Equations 7.5, 7.6 and 7.7, we can divide the (a, c) region into six parts, as shown in Figure 7.9.

Let us compute the behaviour adopted within the six regions obtained earlier:

1. **Region (I):** Since $a < 2, c > 1$ through Equations 7.2 and 7.4, we see that a choice has to be made between strategies B and A . In this case, B would be preferred. B would provide a higher payoff (i.e., 2) as compared to the payoff from A ($a < 2$).
2. **Region (II):** Since $a > 2, a - c < 1$ through Equations 7.2 and 7.3, we see that a choice has to be made between strategies A and B . In this case, A would be preferred. A would provide a higher payoff (i.e., a) as compared to the payoff from B (i.e., 2) because $a > 2$.
3. **Region (III):** Since $c > 1, a - c > 1$ through Equations 7.4 and 7.3, we see that a choice has to be made between strategies A and AB . In this case, A would be preferred. The payoff through AB would be $a + 1 - c$ ($\max(a, 1) = a$; $a > 2; b = 1$) that would be $< a$ (payoff obtained from strategy A) because $c > 1$ ($1 - c < 0$).
4. **Region (IV):** Since $a > 2, c < 1$ through Equations 7.2 and 7.4, we see that a choice has to be made between strategies A and AB . In this case, AB would be preferred. The payoff through AB would be $a + 1 - c$ ($\max(a, 1) = a$; $a > 2; b = 1$) that would be $> a$ (payoff obtained from strategy A) because $c < 1$ ($1 - c > 0$).
5. **Region (V):** Since $a < 2, a - c > 1$ through Equations 7.2 and 7.3, we see that a choice has to be made between strategies B and AB . In this case, AB would be preferred. The payoff through AB would be $a + 1 - c$ ($\max(a, 1) = a$; $a < 2; b = 1$) that would be > 2 (payoff obtained from strategy B) because $a - c > 1$.
6. **Region (VI):** Since $a - c < 1, c < 1$ through Equations 7.3 and 7.4, we see that a choice has to be made between strategies B and AB . In this case, B would be preferred. The payoff through AB would be $a + 1 - c$ that would be < 2 (payoff obtained from strategy B) because $a - c < 1$.

In order to find the suitable values of (a, c) for which strategy A would propagate through the entire network, we combine the results captured by Cases A and B as

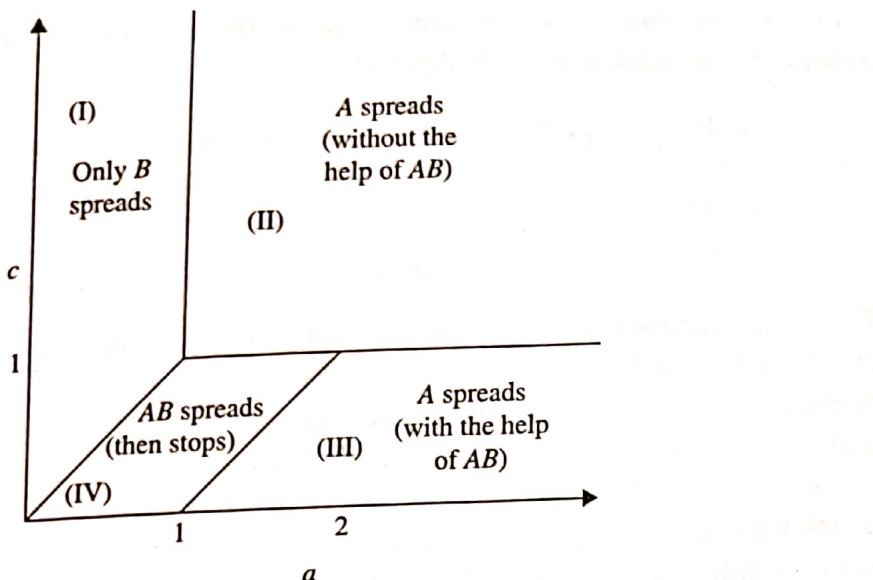


FIGURE 7.10. Combining Cases A and B.

shown in Figure 7.10. We can therefore conclude that the $a - c$ plane divides into four final regions:

1. **Region (I):** For very small values of a , the nodes stick to strategy B .
2. **Region (II):** If both a, c are high then it would not be favourable for the nodes to adopt AB (not feasible to incur the high-cost c) or B , resulting in the direct achievement of behaviour A throughout the network.
3. **Region (III):** In this region, A spreads with the help of infiltration via AB . Therefore, the nodes initially adopt AB and then drop B to receive a better payoff by eliminating cost c . Example shown in Figure 7.7 is an instance of this case.
4. **Region (IV):** In this region, AB spreads by 1 step in the first iteration but all the nodes afterward stick with B . Example showed in Figure 7.6 is an instance of this case.

7.3 CASE STUDY – THE “INDIGNADOS” MOVEMENT

In 2011, Spain witnessed one of the highest unemployment rates across Europe, responsible for a staggering share of 21.3%, bringing the total amount to 4,910,299 at the end of March 2011. This number had risen from about 214,000 from the previous quarter.² A lot of measures were introduced by the parties of Spain to help overcome this tough state of affairs. Unfortunately, all the fixes reportedly contained glaring loopholes that resulted from giving access to cheap labour to business owners or increasing the retirement age of the working population. Neither of these policies was acceptable to the citizens of Spain. These spurred a series of events, which was named as the Anti-Austerity Movement or the *Indignados Movement*. This event is of great interest to us because the protesters/participants coordinated movements using the popular social network – *Twitter*. The dynamics of the participation of the citizens

²https://en.wikipedia.org/wiki/Anti-austerity_movement_in_Spain

can be viewed through the lens of a cascade problem by formulating the problem, taking into consideration the following:

1. When does one generally start to tweet about the protests?
2. Does the neighbour have any influence on the answer to the aforementioned question?
3. If yes, how much is the influence?

The three questions have been answered by González-Bailón et al. (2011) using the concept of *cascade networks*. To study the aforementioned problem, a collection of relevant tweets are needed. In order to do so, the authors first identified 70 hashtags that were used by the protesters. Tweets containing these hashtags were collected for one month. A total of 581,750 tweets were collected, and 87,569 users were identified as relevant. The condition for relevancy is – any user who tweeted with the specified set of hashtags and has followers and followees who did the same.

7.3.1 Network Formulation

Two kinds of networks were created, consisting of uni-directed followers.

1. **Full network:** This network consists of edges, each of that represents a *follow* i.e., if a user follows another user, an edge will be formed between them connecting the two users in the network.
2. **Symmetric network:** This network consists of edges that represent *mutual follows* – if both the users follow each other, a single edge will be formed between the users in the network. This network represents the *strong connections* between the users.

For one month (across which the tweets were collected), the network is assumed to be static.

7.3.2 Common Terminologies

First, let us get the common terminologies out of the way. We shall use these terms frequently in the subsequent sections.

User Activation Time

This is the time when the user gets activated, i.e., starts tweeting about the protests.

K_{in}

This denotes the number of neighbours of a user when she becomes active.

K_a

This denotes the number of *active* neighbours (tweeting about the protests) of a user when she becomes active.

Activation Threshold

This denotes the fraction of active neighbours (tweeting about the protests) of a user when she becomes active. It is given by $\frac{K_a}{K_{in}}$.

7.3.3 Recruitment Threshold

Recruitment

It is defined as the event wherein a user tweets about an ongoing protest in Spain. The growth of fraction of active users is given in Figure 7.11

We want to answer – how much role does “social pressure” or “peer pressure” play in the recruitment of an individual (to tweet about the protests)? This can be quantified using the amount of social pressure by defining it as the fraction of neighbours who are already recruited, i.e., actively tweeting about the protests.

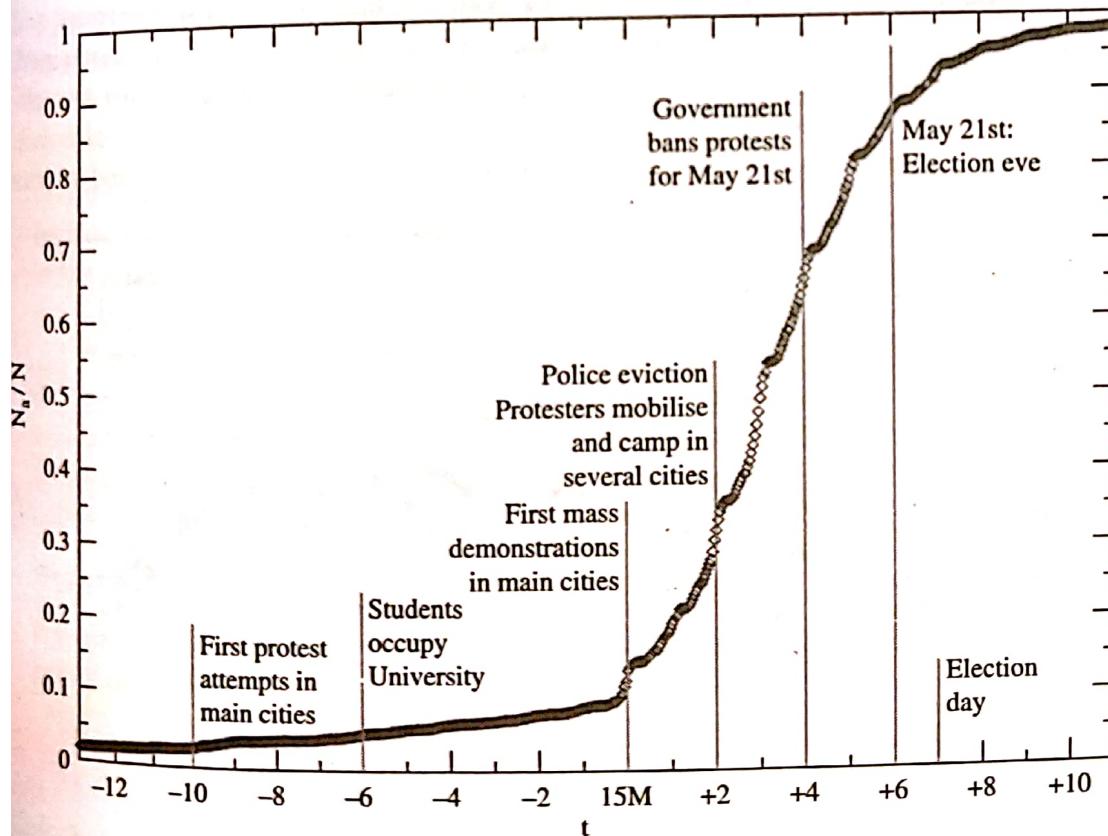


FIGURE 7.11. Fraction of recruited users over time. The x-axis denotes the time over one month. The y-axis denotes the number of active users normalised by the total users. We can see that curve reaches 0.98 by the end of the month. This shows that over the course of the month, 98% of the users tweeted about the protest. The vertical labels denote some of the important events that took place over the course of the month (González-Bailón et al. 2011). Reprinted with permission from González-Bailón et al. (2011).

Accordingly, the following formulations were introduced.

1. If $\frac{K_a}{K_{in}} \approx 0$, it is said that the user has joined the current protests under little or no social pressure. This can be translated as "very few of the user's neighbours were tweeting about the protests when she decided to join".
2. If $\frac{K_a}{K_{in}} \approx 1$, it is said that the user has joined the current protests under a lot of social pressure. This can be translated as "a lot of the user's neighbours were tweeting about the protests when she decided to join".

Now, the next logical step would be to peek beneath the hood and see how the activation threshold $\frac{K_a}{K_{in}}$ is distributed across the population of active users. To get an in-depth understanding, two plots are presented; Figure 7.12 has the distribution of the activation threshold separately plotted for the full network and the symmetrical network. Figure 7.13 has two plots for different timelines. The month-long timeline is broken down into *pre-15M* and *post-15M* timelines. *15M* (15 May) denotes the first day of the mass demonstrations.

We observe from Figure 7.12 that for both the networks, (a) full network and (b) symmetric network, a uniform distribution of activation threshold is maintained, with the exception of two local peaks: (a) at 0, representing the leaders or the self-active users of the protests, and (b) at 0.5 (many users join the protests after half of their neighbours do). From Figure 7.13, the researchers concluded (and we can

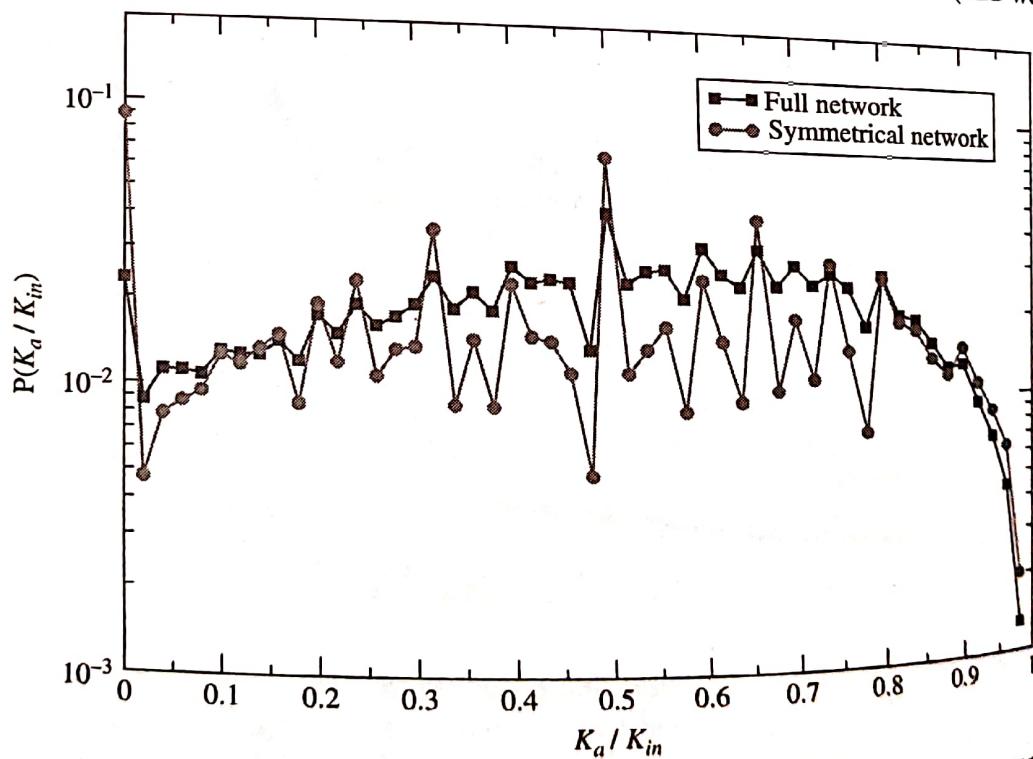


FIGURE 7.12. x-axis denotes the activation value. y-axis denotes the corresponding fraction of total users having that as the activation threshold. These two plots are for full network and symmetric network defined in Section 7.3.1 (González-Bailón et al. 2011). Reprinted with permission from González-Bailón et al. 2011.

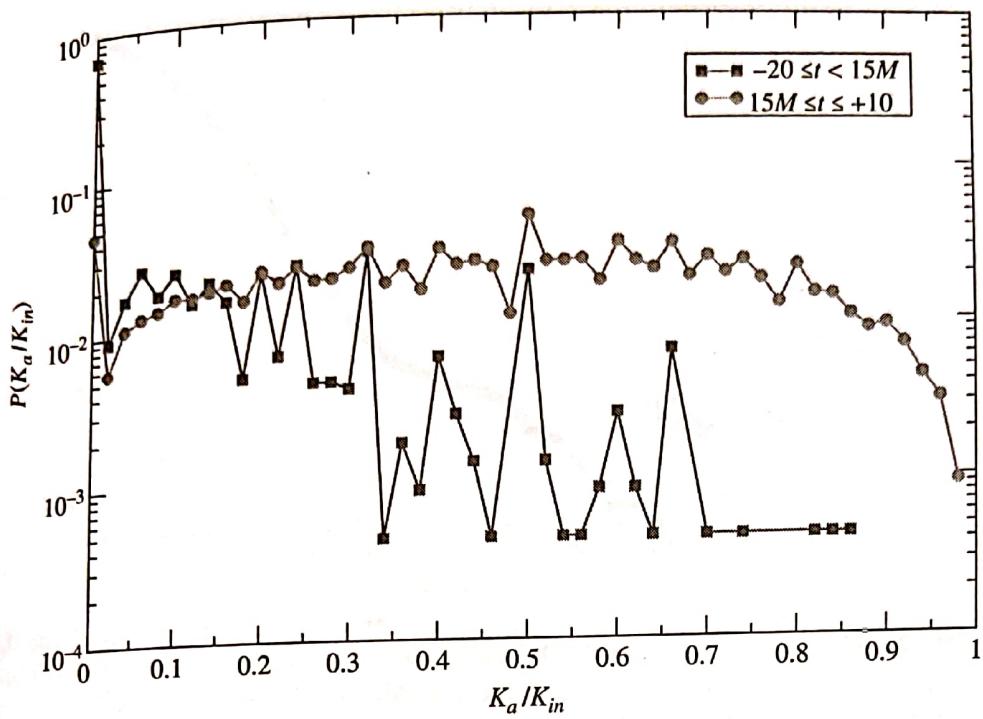


FIGURE 7.13. The x -axis denotes the activation value. y -axis denotes the corresponding fraction of total users having that as the activation threshold. These two plots are for *pre-15M* and *post-15M* timelines (González-Bailón et al. 2011). Reprinted with permission from González-Bailón et al. (2011).

observe) that the activation threshold distribution did not change much amongst the early activated, low-threshold users after *15-M*. At the same time, the media coverage of the protests increased *post-15M*. Therefore, one can conclude that the media did not significantly influence the recruitment of early activated low-threshold users.

Now, the next question to be asked is whether there is any discernible effect of the neighbour's activation time on the user. More formally, whether the hypothesis “*if several neighbours of a user suddenly become active, then the user is more likely to become active*” is true or false. To study this, the researchers first defined a quantity called *Burstiness*.

Burstiness

Burstiness is defined as the relative increase in the active neighbours. This is mathematically defined as,

$$\frac{\Delta K_a}{K_a} = \frac{(K_a^{t+1} - K_a^t)}{K_a^t} \quad (7.8)$$

where t denotes a day.

The cumulative distribution of the burstiness of the users looks like Figure 7.14. We can observe that the slope is very steep for high activation threshold users as compared to the low activation threshold users. This indicates that users with high activation threshold are more likely to join the protests if they see a sudden increase in participation amongst their neighbours. The same is not true for users with lower activation thresholds.

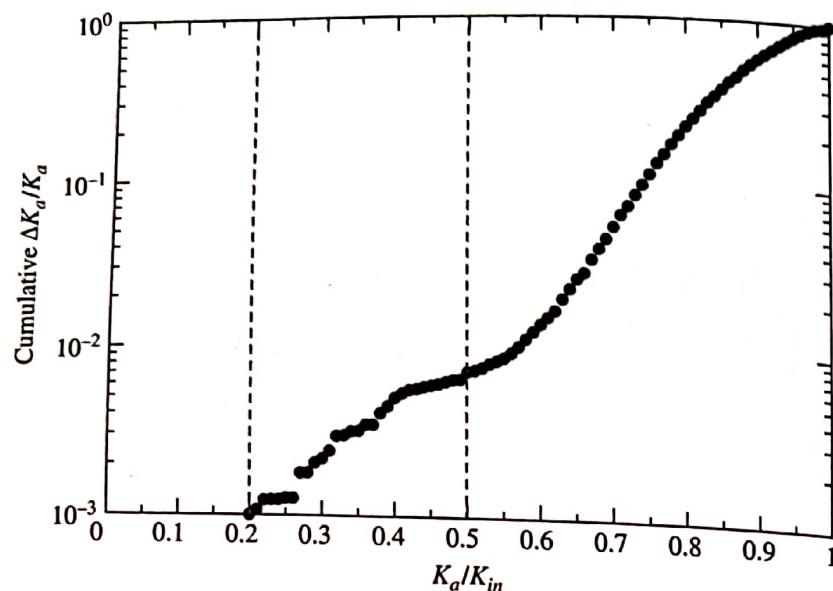


FIGURE 7.14. x-axis denotes the activation value, y-axis denotes the corresponding cumulative burstiness (González-Bailón et al. 2011). Reprinted with permission from González-Bailón et al. (2011).

7.3.4 Information Cascades

Now, we need to define the process of information cascading for our case study. There is no native definition of cascade structure or cascading in this graph similar to the one observed in some other settings such as the spreading of diseases. Here, information cascade is defined as: “*if the messages of a user and her followers lie within the time difference Δt , then the user and the followers are said to form a cascade*”. We notice that the size of the cascade increases, and at the same time, the number of such cascades decreases as shown in Figure 7.15.

Now, we have got the answers to the following questions:

1. When does a user get influenced by its neighbours?
2. Does the influence resemble a cascading structure?

The answer to both the questions stated here is a resounding *yes*. Furthermore, we turn our sight to the third most pressing question – *who starts successful cascades?* This question is very broad. We shall try to answer a particular case of this question. *Does the centrality of the node starting the cascade have a relation with the success of the cascade?* Here, “success” is defined as the ability to start a cascade that consistently coalesces to increase in size and does not remain stagnant for a long time. This question can be answered using *K*-core decomposition (refer Chapter 2).

K-core Decomposition

It refers to the extraction of the maximal subgraph of the parent graph such that all the vertices in the subgraph have a degree of at least K . Figure 7.16 gives us a better picture of what a K -core is.

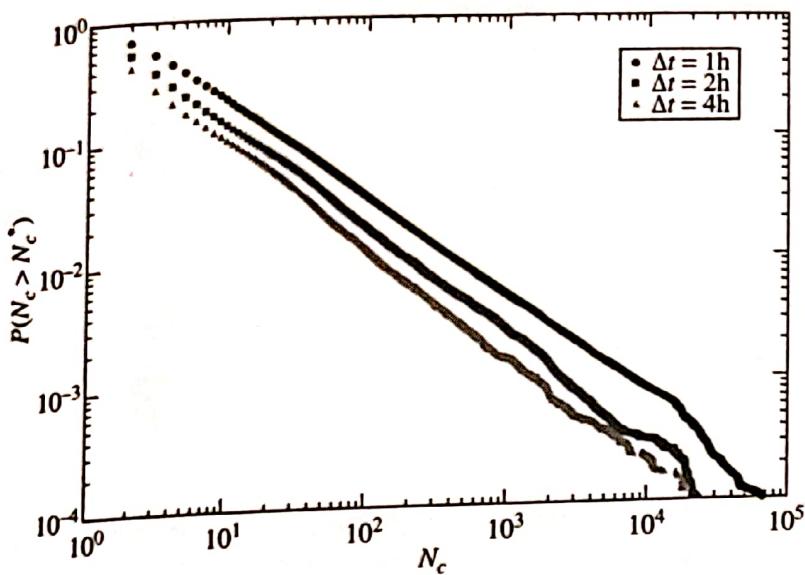


FIGURE 7.15. x-axis denotes the size of cascades. y-axis denotes the fraction of cascades having the corresponding size for different Δt s (González-Bailón et al. 2011). Reprinted with permission from González-Bailón et al. (2011).

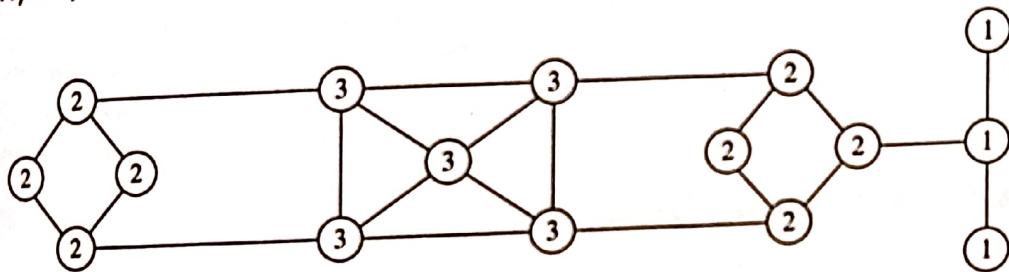


FIGURE 7.16. Nodes with label 3 denote a 3-core subgraph. Nodes with label 2 denote a 2-core subgraph. Nodes with label 1 denote a 1-core subgraph.

Researchers use the method of K -core decomposition on the asymmetric network to find if the value of K has any correlation with starting the cascade. The method of K -core decomposition is implemented by iteratively removing nodes with a degree less than K . K -core can also be thought of as a connected chunk of a network having higher centrality with respect to the rest of the network. We notice that cascades belonging to higher K -core subgraphs (i.e., the central users) are indeed responsible for starting successful cascades as shown in Figure 7.17.

7.4 PROBABILISTIC CASCADES

Earlier in Section 7.2.1, we saw that the cascade is structured as a decision-based model, where the infection is dependent on the payoff model. The cascade will continue to grow only when its growth is associated with the highest reward amongst each of the nodes. But in many real-world scenarios, such hard decision-making criteria or payoff functions are not present. For example, let us consider the infection

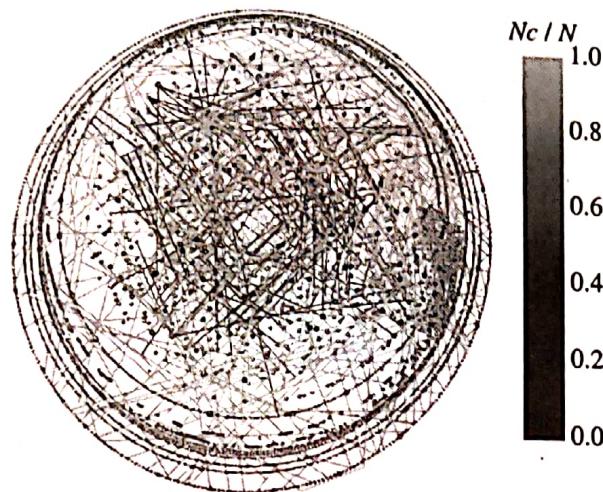


FIGURE 7.17. Size of the node denotes the degree centrality, i.e., k -core decomposition, and the shade of the node denotes the maximum size of the cascade generated by that node (user) (González-Bailón et al. 2011). Reprinted with permission from González-Bailón et al. (2011).

spreading mechanism of the novel COVID-19 virus. If we consider each individual of the population as a node, we can see the spread of the virus as a cascade. This cascade continues to grow over time. In other words, this cascade growth is not in the hands of the node. It is very complex and unobservable. However, you need not fret, for the probabilistic model of information cascade would come to our rescue now!

7.4.1 Modelling Cascade

We shall model an epidemic spread using the probabilistic assumptions that we enumerate in the subsequent sections. We model the spread on ‘random trees’. Random trees look similar to those shown in Figure 7.18.

We employ three assumptions to make the analysis easier.

1. Person at the top (root) of the random tree is always infected.
2. Each person in the random tree meets d new people. So, the random tree is a d -nary tree.
3. Each person, on meeting an infected person, has the probability of getting infected as q ($q > 0$).

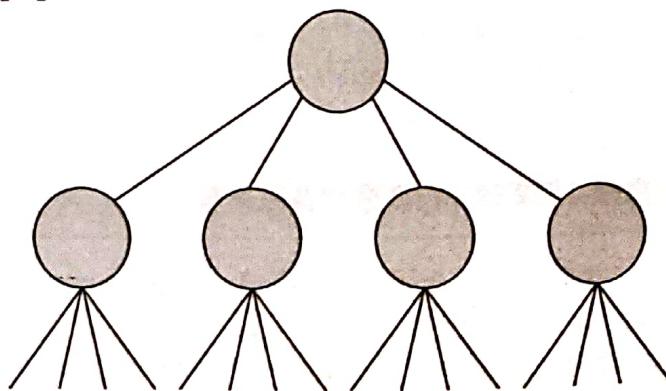


FIGURE 7.18. A random tree with $d = 4$.

Now, we will dig deeper into how the two quantities (d, q) enumerated in the assumptions influence the spread of the epidemic. We can safely conclude that,

- For the virus to stay active and keep on spreading, the probability of a node being infected at a depth h should be a positive real number. Also, this condition should hold true for all the levels of the random tree. If we know, at any level h , there exists an infected node, we can be sure that there is at least one infected node in the previous level, i.e., in level $h - 1$. So, for the virus to stay active, there should be a non-zero probability of an infected node present at the bottom of the tree. Mathematically, we can write this as,

$$\lim_{h \rightarrow \infty} P[\text{a node at depth } h \text{ is infected}] > 0 \quad (7.9)$$

- Drawing along the same lines, the condition for the cascade to die out will be that, at any particular level h , no node is infected. Mathematically,

$$\lim_{h \rightarrow \infty} P[\text{a node at depth } h \text{ is infected}] = 0 \quad (7.10)$$

Now, all we need is to model is the left-hand side of Equations 7.9 and 7.10, i.e., $\lim_{h \rightarrow \infty} p_h$ in terms of q and d . This quantity can be thought of as a recurrence relation. Once the probability at $h - 1$ is known, we can calculate the probability at h . Let the probability of a node being infected at level $h - 1$ be p_{h-1} . Then the probability of being infected at level h is given by,

$$p_h = 1 - (1 - q \times p_{h-1})^d \quad (7.11)$$

where $q \times p_{h-1}$ denotes the probability of one particular node being infected at level h . Subtracting that quantity from 1 gives us the probability of that node not being infected. There are d such nodes in that particular level; so we raise this quantity to the d^{th} power for all the nodes. Now, we subtract this quantity from 1 to get the probability of at least one node being infected at level h . This is what p_h denotes.

Now, if the root of the random tree is not infected, then the epidemic dies out at the first node itself. In order to watch our analysis in action, we need to consider that the root node is infected with a probability equal to 1. Therefore, we can write p_2 as,

$$p_2 = 1 - (1 - q \times 1)^d \quad (7.12)$$

Similarly, p_3 can be written as,

$$p_3 = 1 - (1 - q \times p_2)^d \quad (7.13)$$

Using certain approximations, we can morph this into a recurrence relation that is a function of x such that:

$$f(x) = 1 - (1 - q \times x)^d \quad (7.14)$$

where x represents the level/height in the random tree at which the probability is being calculated. Now, we notice three properties of Equation 7.14. These are

1. $f(x)$ is a **monotonic** function. This is because $\forall t : f'(t) > 0$.
2. $f'(x)$ is **non-increasing**, i.e., $f'(x) = q \times d \times (1 - q \times x)^{d-1}$.
3. $f'(x)$ is **monotone non-increasing** on $[0, 1]$.

The reason we enumerated these properties is that we can then quantify the necessary key conditions.

We can observe that $f(0) = 0$ and $f'(0) = q \times d$. Using the third property enumerated here we can conclude that $f'(x) \leq q \times d; \forall x$. For the epidemic to die out, we want the probability of infection at a lower depth to decrease, i.e., $f(x) < x \Rightarrow f'(x) < 1 \Rightarrow q \times d < 1$. This is obtained by the application of the mean value theorem used in conjunction with the third point enumerated here. Now, we can focus our attention on the quantity $q \times d$. In the existing literature, this quantity is called *Reproductive Number* (denoted as R_0). Now, the two conditions arising out of different values of R_0 are

1. If $R_0 \geq 1$, the epidemic grows in an exponential manner.
2. If $R_0 < 1$, the epidemic spreads at a rate that reduces constantly and eventually dies out.

We can use two methods to contain the spread of the epidemic.

1. We can restrict the virus from spreading to nodes by keeping the already infected nodes in isolation. This would reduce the value of d .
2. We can restrict the virus by reducing the transmission rate by promoting better hygiene practices. This would reduce q .

Now, it would be nice to confirm if the empirical value of R_0 correlates with the formula estimate derived earlier. A study was conducted to estimate R_0 in real-world scenarios by Cha et al. (2008). The dataset used was crawled from the Flickr social network where each user is connected to other users through a link denoting that they are friends. Users on Flickr interact by “liking” each other’s images. Data amounting to 100 days of activity were collected across ~ 2 million users interacting with a total of ~ 11 million images and sharing around ~ 34 million ‘like’ interactions. A cascade in such a network is defined as: *if a user in her friend list ‘liked’ a picture before you, then the user’s like is a part of the social cascade of the photo*.

To empirically calculate the value of R_0 , we calculate the fraction of infected nodes at the root for each cascade. To find out the true estimate, the formula that is employed is:

$$R_0 = q \times d \times \frac{\text{avg}(d_i^2)}{\text{avg}(d_i)^2} \quad (7.15)$$

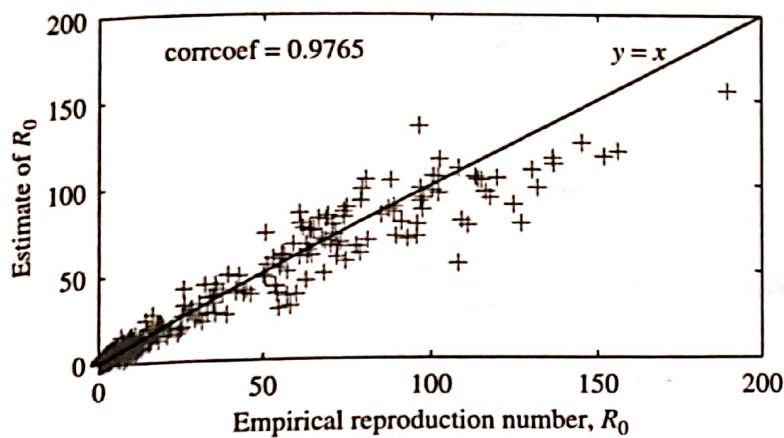


FIGURE 7.19. Correlation of empirical R_0 with the calculated one (Cha et al. 2008). Reprinted with permission from Cha et al. (2008).

where the term $\frac{\text{avg}(d_i^2)}{\text{avg}(d_i)^2}$ is introduced in the original estimate to account for the skewed degree distribution. To estimate q , the mean of the fraction of nodes infected across all infected nodes was taken. We observe that there is a strong correlation ($= 0.9765$) between the calculated and the empirical value. This is shown in Figure 7.19.

7.5 EPIDEMIC MODELS

Real-life contagion/virus is associated with epidemics that are a tad bit different from the ones discussed earlier. This is because once a node gets infected if it heals before transmitting the virus, the effective transmission rate is 0. To account for this, we introduce a new epidemic model that uses two parameters to simulate the propagation. These are:

1. Birth Rate (β): The probability with which a neighbour node attacks another node with the virus.
2. Death Rate (δ): The probability with which an infected node heals.

7.5.1 SEIR Model

$S+E+I+R$ (or SEIR) model gives us a generalised framework to model the spread of epidemics. SEIR is an acronym of Susceptible (those who may become infected), Exposed (those who are infected, but not yet capable of spreading the infection/idea), Infected (those who are capable of further propagating the infection/idea), and Recovered (those who have recovered from or become immune to the infection/idea). Another state is Skeptics (Z , susceptibles who no longer follow the infection/idea).

This is shown in Figure 7.20. There are many variations of the aforementioned framework. Two of the variations that we shall discuss are SIR and SIS models.

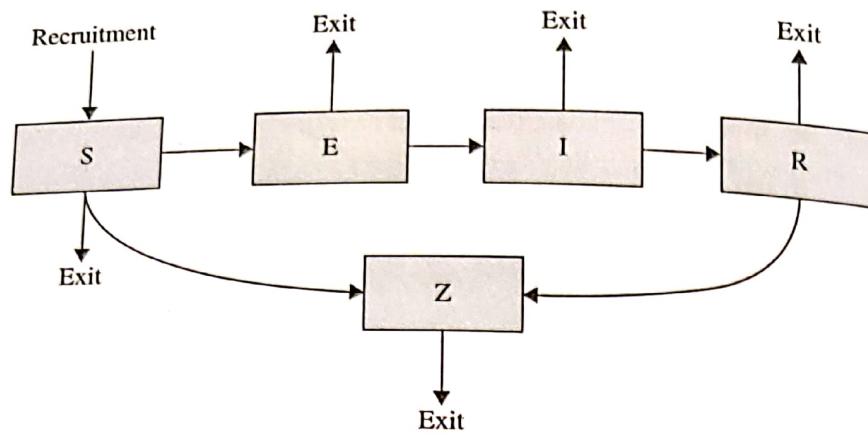


FIGURE 7.20. Concept diagram of SEIR model.

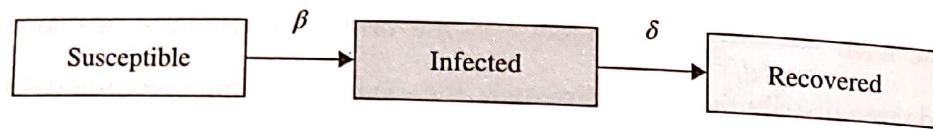


FIGURE 7.21. Concept diagram of SIR model highlighting the transition rates.

7.5.2 SIR Model

Certain rules which apply to this framework are as follows:

1. A node can go through only three stages: (a) Susceptible, (b) Infected, and (c) Recovered, i.e., there is a direct transition from the population of susceptible to those who can transmit the infection/idea.
2. The transition between stages are governed by transition rates elaborated earlier in this section. This is further clarified through Figure 7.21.

Employing fundamental concepts from the differential equations, we can formulate the following equations.

1. Rate of change of ‘susceptible population’ is given by,

$$\frac{dS}{dt} = -\beta \times S \times I \quad (7.16)$$

2. Rate of change of ‘recovered nodes’ is given by,

$$\frac{dR}{dt} = \delta \times I \quad (7.17)$$

3. Rate of change of ‘infected population’ is given by,

$$\frac{dI}{dt} = \beta SI - \delta I \quad (7.18)$$

7.5.3 SIS Model

In the SIS model, as the abbreviation suggests, a node can go through the phases of 'susceptible' to 'infected' to 'susceptible' again. This primarily means that nodes that have recovered will be immediately made susceptible again. For example, a disease such as the common cold which can recur with a high probability can be modelled by SIS. This is shown in Figure 7.22.

Similar to the SIR model, we shall use differential equations to model the growth of the susceptible and the infected population. Each virus has a strength of β/δ . Rate of change of 'susceptible population' is given by,

$$\frac{dS}{dt} = -\beta SI + \delta I \quad (7.19)$$

Rate of change of 'infected population' is given by,

$$\frac{dI}{dt} = \beta SI - \delta I \quad (7.20)$$

One might wonder, if a recurring epidemic ever dies out at all, especially, since we added the condition of a recovered node being susceptible again. The answer to this very crucial question is a resounding yes with some conditions. To answer this question, we introduce new terminology, the *epidemic threshold*, denoted by τ . If the virus strength is less than τ , then the epidemic eventually dies out. τ is nothing but the reciprocal of the largest eigenvalue of the adjacency matrix representing the underlying network. Mathematically, for the epidemic to die out, the required condition is

$$\frac{\beta}{\delta} < \tau = \frac{1}{\lambda_{1,A}} \quad (7.21)$$

where A denotes the adjacency matrix and $\lambda_{1,A}$ denotes the largest eigenvalue of A . An experimental result from Wang et al. (2018) confirms the aforementioned result as shown in Figure 7.23.

The next pressing question is – would having a large number of infected nodes in the initial stages stop the epidemic from dying down? Wang et al. (2018) also answered this question via Figure 7.24. It shows that increasing or decreasing the number of initial carriers will make no difference in extending or reducing the duration of the epidemic.

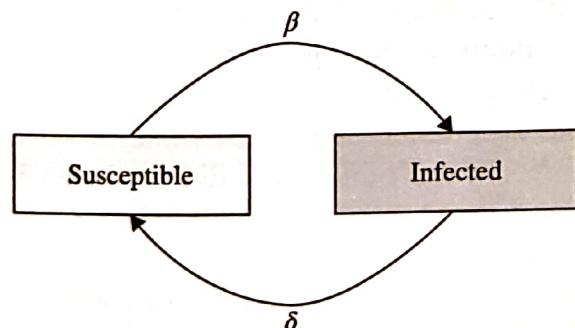


FIGURE 7.22. Concept diagram of SIS model highlighting the transition rates.

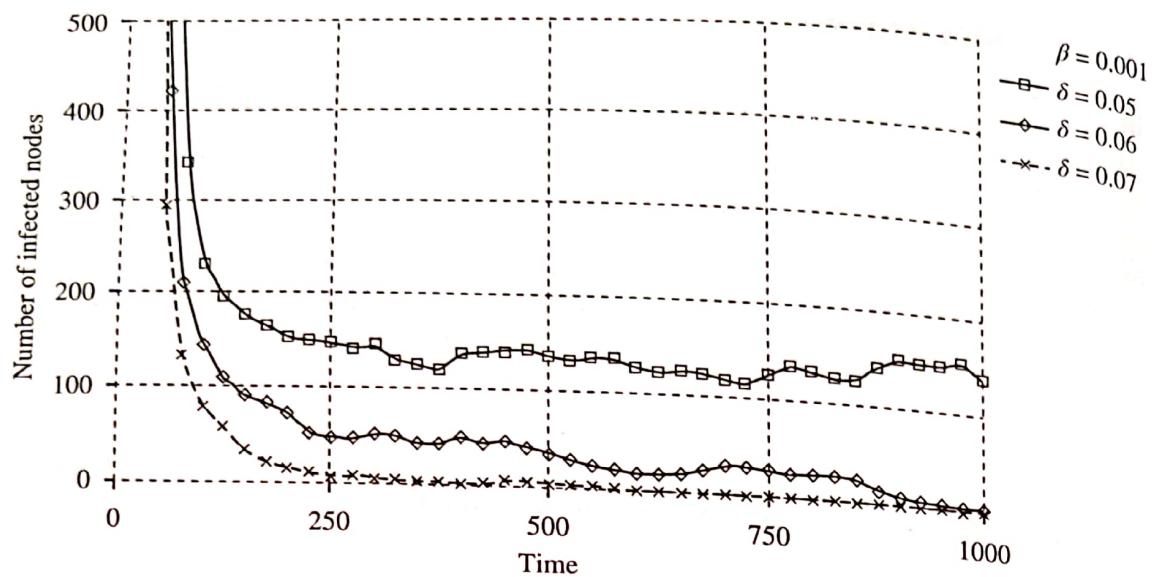


FIGURE 7.23. Infected nodes vs. time for different β values (Wang et al. 2018). Reprinted with permission from Wang et al. (2018).

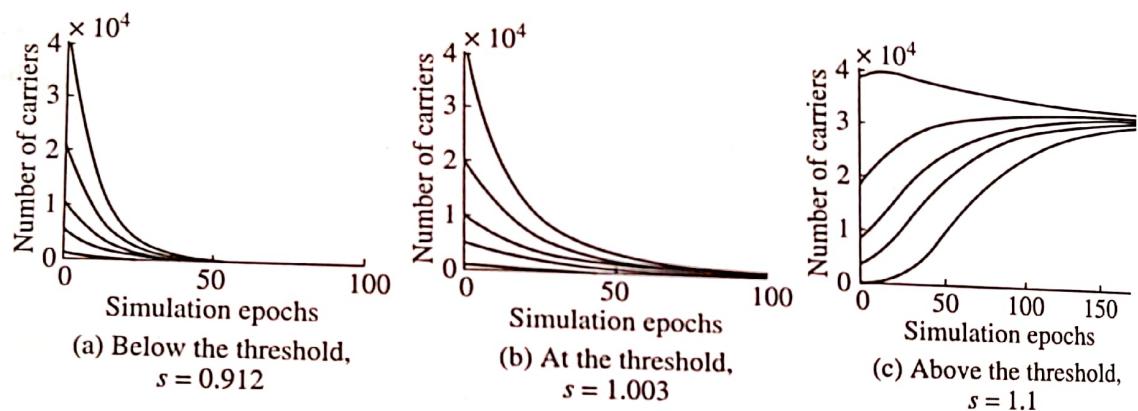


FIGURE 7.24. Infected nodes vs. time for different β values for the different number of initial nodes. Reprinted with permission from Wang et al. (2018).

7.5.4 Analysing Rumour Spread – SEIZ Model

Now that we are armed with the theoretical understanding of the SEIZ model, we can apply it to real-world scenarios. One such interesting application would be to analyse the spread of rumours in social networks such as Twitter. A rumour in many ways is like a disease. People who are active on social media can be considered to be ‘susceptible’. People who believe the rumour, can be considered to be ‘infected’. People who have seen/heard the rumour but did not believe it yet can be considered to be ‘exposed’. And people who did not believe can be considered ‘skeptics’. Jin et al. (2013) followed this paradigm.

Here, different parameters are as follows: $S \rightarrow I$ contact rate (β), $S \rightarrow Z$ contact rate (b), $E \rightarrow I$ contact rate (ρ), incubation rate (ϵ), $S \rightarrow Z$ probability is given contact with skeptics (I), and $S \rightarrow I$ probability is given contact with adopters (p).

Using the pre-defined notations that are denoted by S , I , E , and Z , the model was fitted to a dataset containing various real and fake news that trended on Twitter. The model was fitted such that the difference between the predicted number of tweets and the actual number of tweets for each timestamp was minimised. We can infer that SEIZ models the situation better than the SIS model. Jin et al. (2013) followed a modified SEIZ model parameterised as shown in Figure 7.25. Using this, they came up with a new metric, called R_{SI} , which is defined as,

$$R_{SI} = \frac{(1-p)\beta + (1-l)b}{\rho + \epsilon} \quad (7.22)$$

This metric can be considered as a kind of flux ratio between the ratio of effects entering a node that is being examined to those leaving that node. The flux ratio (as shown in Figure 7.26) was high for real-life events and low for events that were rumours.

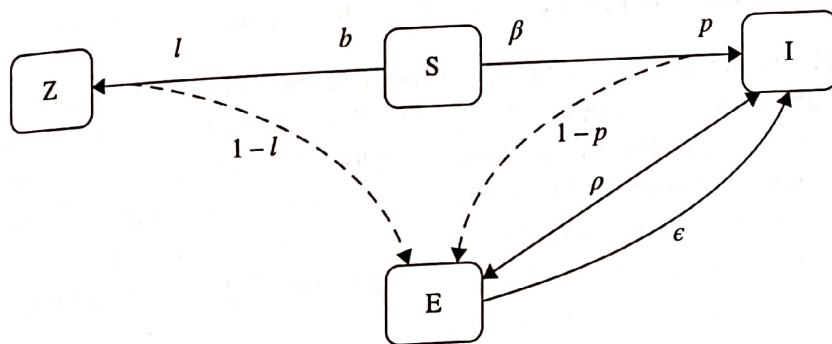


FIGURE 7.25. Parameterised SEIZ Model proposed by Jin et al. (2013).

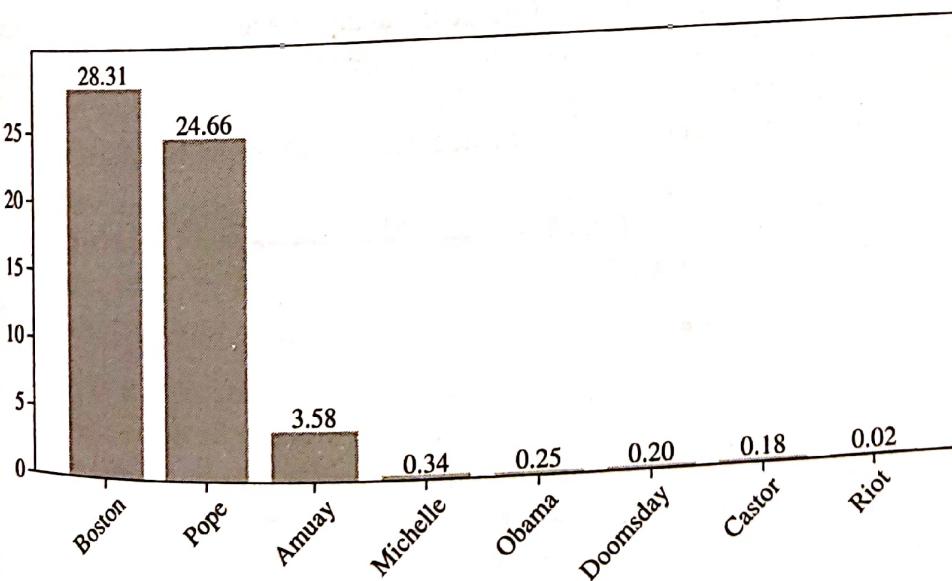


FIGURE 7.26. R_{SI} values for the 8 topics which capture the genuineness of the event. The plot is adopted from Jin et al. (2013).

7.6 INDEPENDENT CASCADE MODELS

Till now, we have looked at modelling the spread of infection with a uniform probability between any two nodes. But in real-life scenarios, the transmission of disease may be more probable between certain pairs of nodes than other pairs. To model this, we can assume an edge between u and v having a probability p_{uv} of transmission between them. A toy network for illustration is presented in Figure 7.27.

To fully parameterise the model, we need to estimate all the edge probabilities. Goyal et al. (2010) showed that it is quite intractable to do so. One very tempting way to reduce the model complexity by making estimation easier is to assume equal weights for all the edges. But this would produce a model equivalent to the SIR epidemic model.

Myers et al. (2012a) proposed the separation of the act of exposure and the event of acting upon that exposure. Those two acts are defined as follows:

1. **Exposure:** The event of a node being exposed to a contagious incident.
2. **Adoption:** The event of the node acting on the contagious incident.

It was hypothesized that the probability of adoption is influenced by the number of neighbours who have adopted. For probabilistic models such as those capturing the spread of viruses and information, the plot of Exposure vs. Adoption is very smooth as shown in Figure 7.28(b). But for events involving decision making that is binary in nature, there exists a kink in the curve, which represents the critical amount of exposure required for adoption as shown in Figure 7.28(a). The different types of exposure curves are shown in Figure 7.29.

The Exposure vs. Adoption dynamics also plays an important role in various marketing strategies. One way to analyse such strategies was researched by Leskovec et al. (2007), where they studied how referral programs assist in customer purchase habits. The dataset on which they performed the analysis was collected from a referral program of a major online retail store based in Seattle. It contained approximately 16 million recommendations across 4 million products and 0.5 million brands. The Exposure vs. Adoption curves for four categories are shown in Figure 7.30.

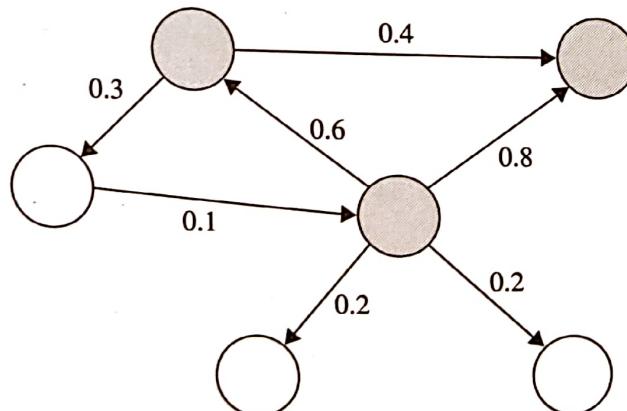


FIGURE 7.27. A toy example of an independent cascade model. Edge weights indicate the probability of transmissions.

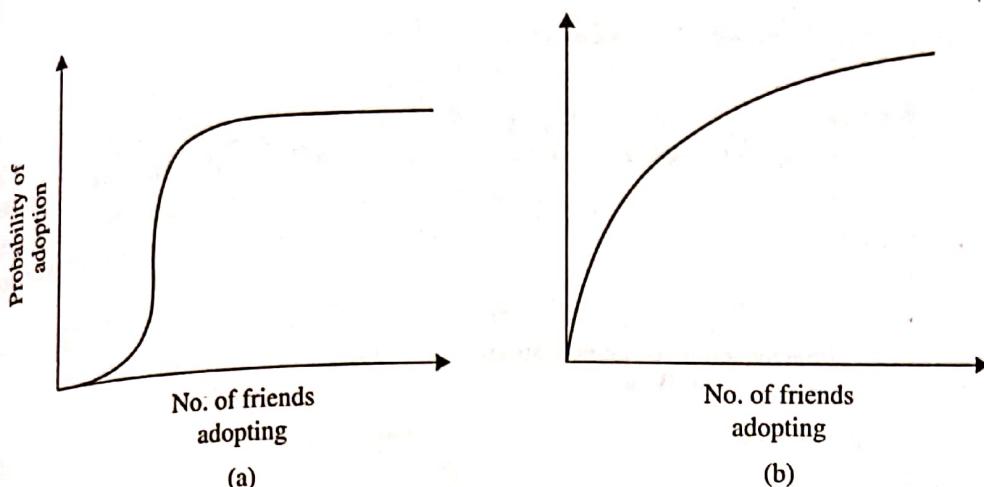


FIGURE 7.28. Exposure vs. Adoption: (a) Discrete model, and (b) Probabilistic model (Myers et al. 2012a).

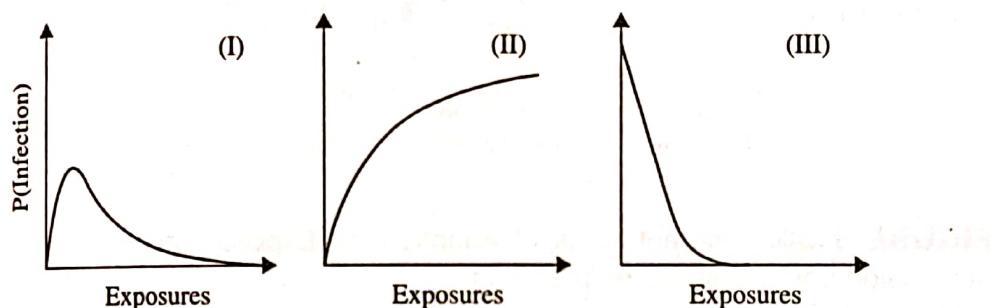


FIGURE 7.29. Different kinds of Adoption vs. Exposure curves. The plot is adopted from Myers et al. (2012a).

In another study by Romero et al. (2011), the exposure curves on Twitter were analysed. Here, ‘exposure’ meant seeing retweets of a particular tweet by different users, and ‘adoption’ meant retweeting that particular tweet after seeing other retweets. The dataset for this was collected by crawling Twitter for 6 months, containing 3 billion tweets from 60 million users. For ease of analysis, only the top 500 hashtags were used. An interesting observation was that the average exposure curve for the top 500 hashtags first peaks and then quickly decays, as shown in Figure 7.31. Before further analysis, we would introduce certain terms defined by Romero et al. (2011) to understand the study better.

1. **Width of the curve:** The length of extension of the curve along the x -axis is called the width of the curve. It is denoted by $\max(D(K))$.
2. **Stickiness of cascade:** It is the maximum height of the curve along the y -axis, i.e., the maximum probability of adoption it had reached the height of its virality. It is denoted by $\max(P_k)$.
3. **Persistence of cascade:** This is defined as the ratio of the area under the curve and the rectangle defined by the width and height of the curve. This reflects how long the cascade persisted.

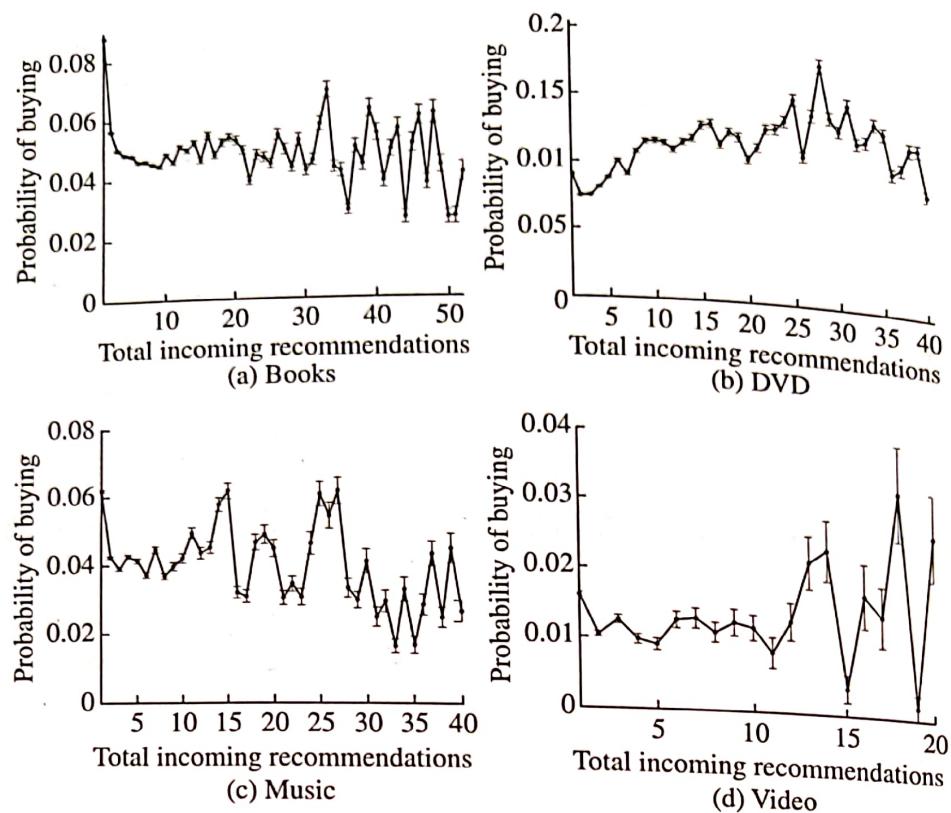


FIGURE 7.30. Different kinds of Adoption vs. Exposure curves. Reprinted with permission from Leskovec et al. (2007).

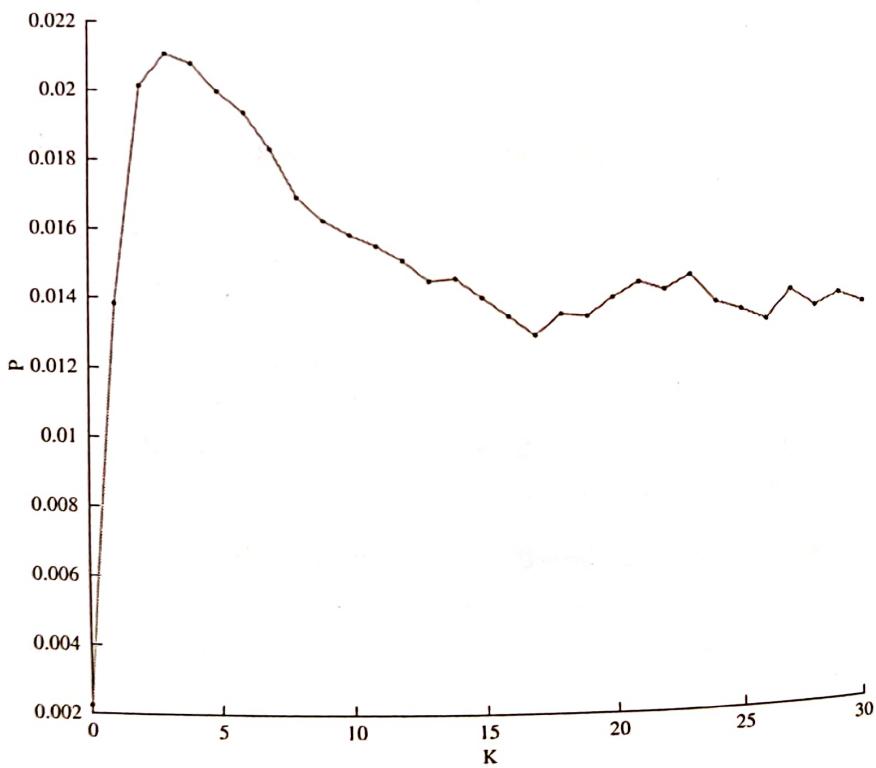


FIGURE 7.31. Average exposure curve for the top 500 hashtags. $P(K)$ denotes the fraction of individuals who adopt/use a particular hashtag after their K^{th} exposure to the same. Reprinted with permission from Romero et al. (2011).

For comparing the exposure curves corresponding to different Twitter hashtags, the authors identified eight topics, each consisting of more than or equal to 20 hashtags. They also defined a NULL set consisting of a random subset of the collected hashtags as a base set to compare with. The persistence of the different categories was compared against the NULL set as shown in Figure 7.32. Two very peculiar observations that came across are as follows:

1. Idioms and music persisted less than the topics in the NULL set, i.e., the positive effect of increased exposure to the related hashtags (causing the user to eventually use the hashtag) falls off more quickly than in the average case.
2. Politics and sports persisted more than the topics in the NULL set.

A similar analysis was performed for the stickiness metric.

Till now, one character trait of all the networks was that the origination of the exposure was within the network. But in a lot of real-world scenarios, an exposure can be injected externally also through media such as TV shows, newspapers, hearsay, etc. Myers et al. (2012b) studied this question by modelling external influences. To formalise the problem studied by the authors, we are given a graph G and a set of node adoption times which is timestamps of node adopting the URL. The goal of this study is to model the external event $\lambda_{ext}(t)$ which is the number of external exposures at time t and also find out the exposure curve. This problem statement is further clarified by Figure 7.33. The authors modelled this hypothesis on 3 billion tweets collected in the year 2011. They concluded that only 71% of exposure comes from within the Twitter network, and the rest is from external sources.

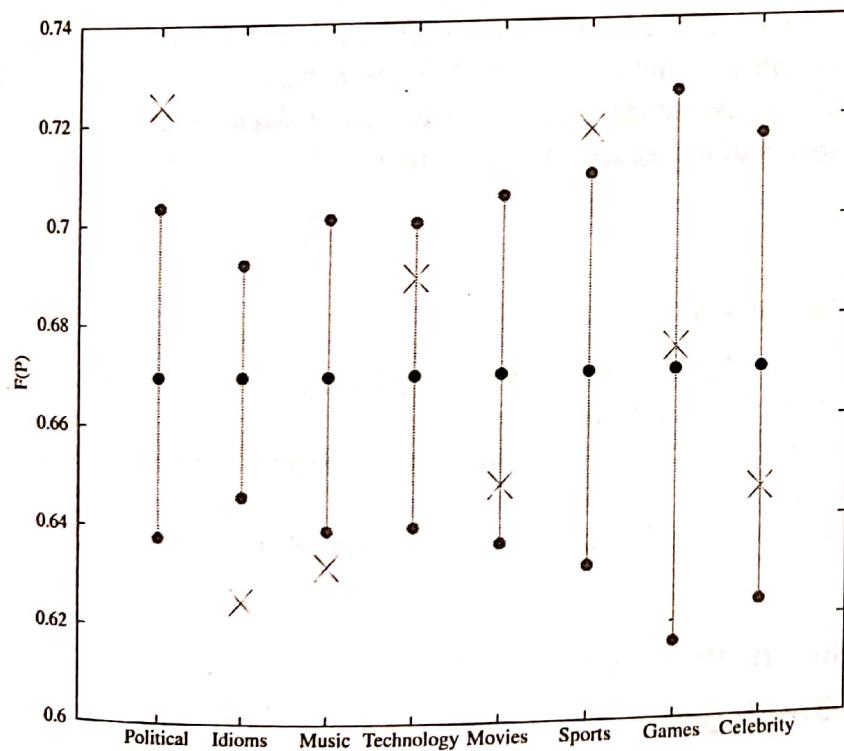


FIGURE 7.32. $F(P)$ denotes the persistence levels. The black dot (at the centres of the vertical lines) denotes the persistence of the NULL subset. The cross denotes the average persistence of the specific topic. The plot is taken from Romero et al. (2011).

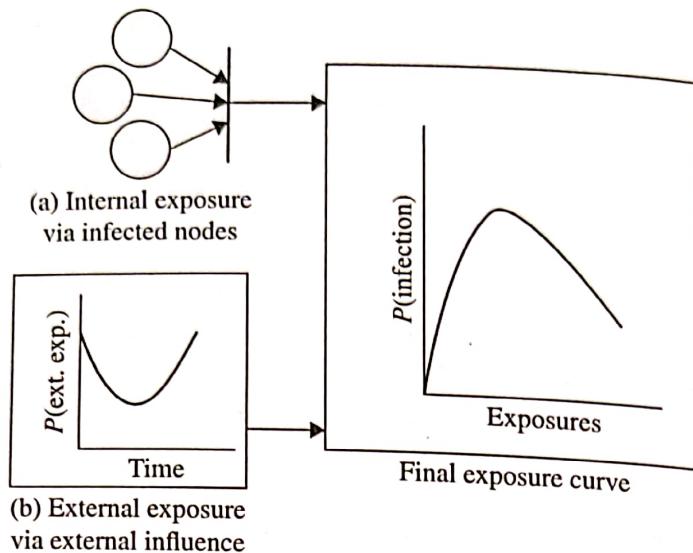


FIGURE 7.33. Modelling external exposures due to outside influences.

7.7 ♦ CASCADE PREDICTION

As discussed so far, information cascades play an essential role in various real-world scenarios, such as rumor spread, viral marketing, and trend prediction. Cascade prediction would benefit the authorities involved in the aforementioned applications. For example, by understanding the propagation of different marketing strategies of a product, marketing specialists would be able to fine-tune their tactics.

Studies in cascade prediction can be divided into two main categories: (a) *classification problem* in which we predict if the diffused information/content would become popular in the future, and (b) *regression problem* in which we learn different numerical aspects of a cascade in future such as the final size, growth, and shape of the cascade.

7.7.1 DeepCas

Here, we present a brief overview of **DeepCas**, proposed by Li et al. (2017). It is an end-to-end deep learning approach to predict the size of a cascade in the future. Instead of using prior knowledge limited to a particular problem statement or type of network to handcraft features, DeepCas uses a representation of the cascade network itself, becoming independent of feature representation.

An Overview of the Architecture

Cascade Graph (g_c^t)

Let C be a set of cascades that originate in a network $G(V, E)$ at time t_0 . $g_c^t = (V_c^t, E_c^t)$ denotes the cascade graph (as shown in Figure 7.34(a)) for a cascade $c \in C$ obtained after time t from t_0 , $V_c^t \subset V$; $E_c^t = E \cap (V_c^t \times V_c^t)$.

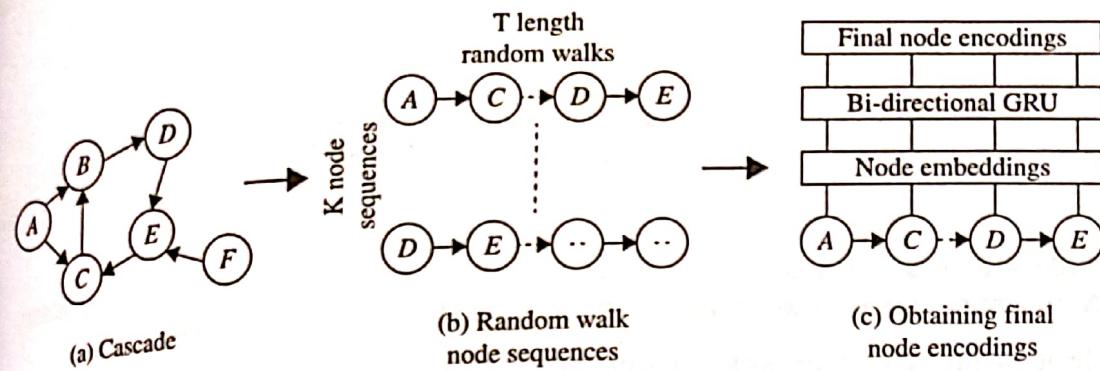


FIGURE 7.34. Flow diagram for DeepCas.

Size Increment of Cascade (Δs_c)

Δs_c represents an increase in the size of cascade c in a Δt time interval, i.e., $\Delta s_c = |V_c^{t+\Delta t}| - |V_c^t|$.

DeepCas presents a neural network framework such that given an input cascade graph g_c , it predicts the increment in the size of the cascade Δs_c . For a cascade graph g_c , DeepCas first represents it in the form of node sequences as shown in Figure 7.34(b). To predict the future size of a cascade, it is essential to consider the propagators/nodes in the pipeline. Instead of considering g_c as a bag of individual nodes, proper node sequences are considered to incorporate the local and global network structure characteristics to facilitate prediction. To sample node sequences, the authors performed random walks on g_c using the Markov chain model shown in Figure 7.35.

Transition Probability in Markov Chain (p)

When the random walker is at state N in the Markov chain and the current node is v , it uses a transition probability $p(u \in N_c(v) | v)$ to select a neighbour node $u \in N_c(v)$ to jump to ($N_c(v)$ represents the set of neighbours of v in the cascade graph g_c). Therefore,

$$p(u \in N_c(v) | v) = \frac{sc_t(u) + \alpha}{\sum_{s \in N_c(v)} (sc_t(s) + \alpha)} \quad (7.23)$$

where α represents the smoothing parameter; $sc_t(u)$ represents the scoring function for a node v to transfer to node u that can be defined as the degree of node u in G or in g_c .

Similarly, when the random walker is at the random node jump state J , the transition probability $p(u)$ of jumping to a random node is defined as,

$$p(u) = \frac{sc_t(u) + \alpha}{\sum_{s \in V_c} (sc_t(s) + \alpha)} \quad (7.24)$$

where V_c represents the set of nodes in g_c .

For determining T (length of node sequences) and K (number of node sequences), the following probabilities (mentioned in Figure 7.35) are learned

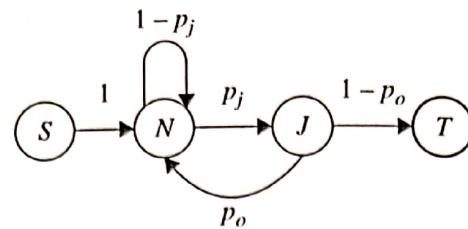


FIGURE 7.35. Markov chain model for a random walk on g_c , S : starting state; N : neighbour node state; J : random node jump state; T : terminal state.

through the neural network architecture itself, instead of tuning them as hyper parameters:

1. p_o : governs the expected number of sampled node sequences from g_c .
2. p_j : governs the expected length of the sampled node sequences from g_c .

After obtaining the node sequences for the cascade graph g_c , we use a node embedding to represent each node.

Node Embedding

Let N^{node} represent the number of nodes in a node sequence. We use one-hot encoding to denote each node, i.e., using a vector $q \in \mathbb{R}^{N^{node}}$.

Now, given an embedding matrix $A \in \mathbb{R}^{H \times N^{node}}$, for a dimension H , the vector embedding representation $x \in \mathbb{R}^H$ for each node is obtained as,

$$x = Aq$$

The authors used a bi-directional *gated recurrent unit (GRU)*³ (Hochreiter and Schmidhuber 1997) to obtain an encoding of the node sequences by capturing the diffusion flow of information through the nodes (shown in Figure 7.34(c)). $h_i^k \in \mathbb{R}^{2H}$ represents an encoding of the i^{th} node in the k^{th} node sequence obtaining by concatenating the hidden vectors obtained from the forward and backward GRUs.

The authors learned the representation of g_c by assembling the encoded node sequences using an *attention mechanism* to learn K and T , as shown in Figure 7.36 (read Chapter 9 for some of the neural network concepts often used on networks).

To learn K , the sequences are divided into mini batches, and geometric distribution of *attention* is modelled. Therefore, learning the parameter p_{geo}^c would provide a good estimate for K . To incorporate the observation that the number of node sequences required has a correlation with the cascade graph size, the authors used $p_{geo}^c = \lfloor \log_2(\text{sz}(g_c) + 1) \rfloor$.

To learn T , a multinomial distribution of attention $(\lambda_1, \dots, \lambda_T)$ is used over the T length of nodes.

³GRUs are a gating mechanism in recurrent neural networks (Hochreiter and Schmidhuber (1997)). It is out of the scope of the book. Read deep learning books for more details.

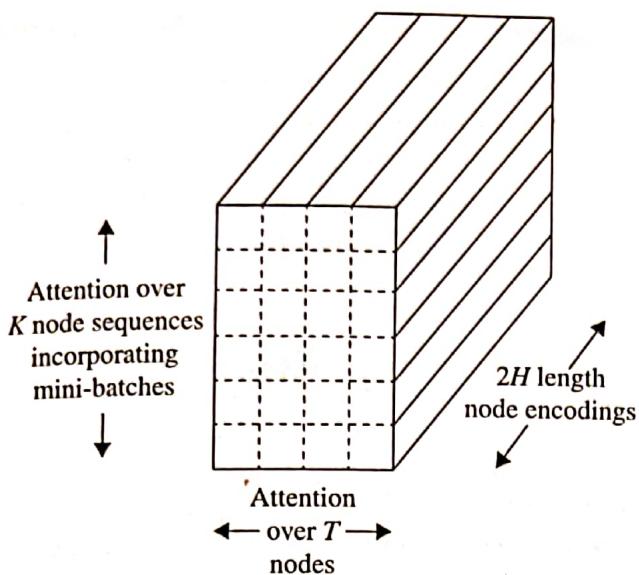


FIGURE 7.36. Attention mechanism to group the node sequences into a representation of g_c .

The final representation of the cascade graph g_c is given as,

$$h(g_c) = \sum_{k=1}^K \sum_{i=1}^T \left((1 - a_c)^{\lfloor k/B \rfloor} a_c \right) \lambda_i h_i^k \quad (7.25)$$

where B is the size of a mini-batch and $a_c = p_{geo}^{\lfloor \log_2(\text{sz}(g_c)+1) \rfloor}$.

The objective function to be minimised is:

$$\mathcal{O} = \frac{1}{|C|} \sum_c (f(g_c) - \Delta s_c)^2 \quad (7.26)$$

where $f(g_c) = \text{MLP}(h(g_c))$; MLP represents a multi-layer perceptron.

7.7.2 DeepHawkes

Let us now briefly discuss DeepHawkes, an end-to-end deep learning approach to predict the popularity of a piece of information diffused in an online social network via retweet prediction. Cao et al. (2017) proposed this model to combine the predictive power of models based on deep learning and the easy interpretation of information cascades provided by the Hawkes process.⁴

An Overview of the Architecture

To begin, let us first describe how the authors defined the popularity of a message or a piece of information.

Popularity of a Message (R_j^i). Given a set of messages M , a cascade $C^i = \{(u_j^i, v_j^i, t_j^i)\}$ is used to describe each message m_i . Each tuple (u_j^i, v_j^i, t_j^i) denotes the j^{th}

⁴Refer HAWKES (1971) to learn more about the Hawkes process.

retweet such that v_j^i retweets the message m_i of u_j^i wherein the time gap between the original post and the j^{th} retweet is t_j^i . The popularity of m_i till time t , denoted by R_t^i , is defined as the total number of retweets within that time frame.

Now, let us briefly discuss the self-exciting Hawkes process on which the DeepHawkes model is based. Hawkes process can be used to structure how new events would arrive in the future, e.g., how a tweet would gain popularity over time. The arrival rate can be modelled as,

$$\rho_t^i = \sum_{j: t_j^i \leq t} \mu_j^i \phi(t - t_j^i) \quad (7.27)$$

Here, ρ_t^i captures the rate at which new retweets arrive for a message tweet m^i at time t ; t_j^i represents the time gap between the original post, and the j^{th} retweet; μ_j^i denotes the users who will be influenced by the j^{th} retweet, e.g., in the case of Twitter it may comprise the followers (immediate neighbours) of the j^{th} retweeter; $\phi(t)$ represents the time decay function.

DeepHawkes is inspired by the following three factors of the Hawkes process captured in Equation 7.27, assembled into a neural network-based model:

- 1. Influence of users:** Messages retweeted by influential users tend to get further more retweets over time. Their model learns the representation of users to capture user influence in the same way as DeepCas learns node embeddings as described earlier in this section.
- 2. Self-exciting mechanism:** Every retweet of a message is responsible for attracting more retweets thereby increasing its popularity over time. As a result, the authors captured the influence of the entire retweet cascade/path to predict future popularity. These retweet paths are encoded using a GRU.
- 3. Time decay effect:** The influential exposure of retweets degrades as time passes. DeepHawkes learns the time decay effect through a non-parametric process not requiring any prior domain knowledge.

7.8 CHAPTER SUMMARY

In this chapter, we have gone beyond the static structural characteristics of networks and discussed dynamic information flow through network cascades. We have looked at different terminologies associated with the cascading phenomena. Through an introduction to cascade modelling, we have been able to discover cascades in two ways: decision-based cascade models and probabilistic models. We have discussed in depth different mechanisms and circumstances in which the spread of the cascade can be formalised with the help of various network scenarios. The first type was the pay-off model or the decision-based cascade model, wherein each node plays its own copy of the game with each of its neighbours to adopt a contagion for maximising its direct benefit. We have explored this setting in both single-choice and multi-choice flavours. Generalised relations have been established for supporting the situation in

which a single strategy/behaviour could propagate through the entire network of an infinite chain of nodes. Then, we have applied the theory to people's behaviour in the historical Spanish movement. We have explored who is responsible for starting cascades. We have also explored the probabilistic setting of the cascade model. We have discussed two methods of cascade spread in this setting, namely, the SIS and SIR models. We have discussed the applications of these models in predicting rumors spread on popular social network sites. We have also looked into what initial conditions are influential. We have then seen an application of these findings for popular marketing strategies. We have further discussed a study that increased the complexity of the model by inducing edge-specific probabilities and modelled social cascades. We have looked into cascade models wherein external influences are incorporated that change different cascade properties. In the end, two deep learning-based approaches have been discussed to provide a motivation for predicting cascades.

ADDITIONAL RESOURCES

- Important papers/blogs/reports
- A Survey of Information Cascade Analysis: Models, Predictions and Recent Advances Zhou et al. (2020).
- Contagion Morris (2000).
- Chapter 19: Cascading Behaviour in Networks (<http://www.cs.cornell.edu/home/kleinber/networks-book>).
- Cascade prediction (without deep learning): Cheng et al. (2014), Cheng et al. (2016), Kupavskii et al. (2012), Bakshy et al. (2011), Zaman et al. (2014).
- Cascade prediction (using generative models): Zhao et al. (2015), Kobayashi and Lambiotte (2016), Bao et al. (2015), Rizouli et al. (2017).
- Cascade prediction (with deep learning): Dutta et al. (2020), Masud et al. (2020).
- Extension of Wang et al. (2018): Prakash et al. (2012).
- Source codes
 - Implementation of DeepCas: <https://github.com/chengli-um/DeepCas>.
 - Implementation of DeepHawkes: <https://github.com/CaoQi92/DeepHawkes>.
 - Implementation of CasCN: <https://github.com/ChenNed/CasCN>.
 - Implementation of TiDeH (generative model for cascade prediction): <https://github.com/NII-Kobayashi/TiDeH>.
 - Twitter cascade dataset curated by SNAP Research: <http://snap.stanford.edu/seismic/>

EXERCISES

Objective Type Questions

1. In decision-based cascade models, a node's decision tries to optimise the reward for its local community. (True/False)
2. The rate of change of infected population in the SIS model depends solely on the birth rate of the virus β (assume perfect mixing of population). (True/False)
3. In case of decision-based cascades, the centrality of a node in the network has no effect on its ability to start a successful cascade. (True/False)

4. For a contagion whose spread is modelled using the SIS model, even if its strength is above the epidemic threshold, we can bring the number of infected nodes to zero, given we restrict the number of initially infected nodes to a minimum. (True/False)
5. Independent cascade model is infeasible for modelling probabilistic contagions because the network might not be in a single connected component. (True/False)

Short Answer Questions

- A new company is trying to compete with an old competitor. The company wants to strategize its plan of increasing the market share of its product/technology with the help of their customer network. What should they do?
 - Strategically choose a set of people and make them an early adopter of this technology.
 - Increase the payoff of adopting the technology.
 - Both (a) and (b).
 - None of these
- An epidemic is spreading in your town. With the help of the information obtained from the official authorities, you know that an infected person infects a new person with a probability of 0.05 and on an average a person gets in

Long Answer Questions

- You are a new researcher at Indian Median Association. You are given a job to model coronavirus. You decided that you would use an SIS model. Your assignment is to calculate dI/dS and explain it. I = number of people infected. S = number of people susceptible to the virus.
- What is the epidemic threshold and why do we need it for an SIS model? When does the epidemic die out?
- Delhi got its first coronavirus patient and you have got the job to model it. We assume the population of Delhi to be around 20 million. We assume if you recover once, you are not susceptible to corona. $S(t)$, $I(t)$, $R(t)$ are the number of susceptible, infected and recovered

contact with 22 other people. Is this epidemic going to die out?

- You need to model the spread of novel COVID-19 using one of the epidemic cascade models. What is the best model to work with (assuming you are still susceptible if you recover from the virus)?
 - SIS
 - SIR
 - SEIR
 - None of these
- Assume that there are two strategies A and B that can be adopted by individuals. B gives a payoff of 5; using both costs 5. What should be the minimum payoff for A, so that everyone goes with strategy A?

people at time t , respectively. $s(t) = S(t)/N$, $i(t) = I(t)/N$, $r(t) = R(t)/N$. Use the SIR model to answer the following:

- Can $S(t)$ increase with time? Why or why not?
- How the networks of $I(t)$, $S(t)$ and $R(t)$ look like?
- Calculate the maximum number of infected people anytime.
- Everyone in a town is using an iPhone. A new company ZETO launches its new phone, called ZPhone. Let us denote ZPhone by A and iPhone by B. Assume an infinite path model with all users using product B. Given the payoff for B as 1, for what pair of (a, c) ,

where a denotes the payoff for using product A; c represents the extra cost for using both A and B, does the usage of product A

spread throughout the town? Describe with the help of a network, given one user uses only product A.

BIBLIOGRAPHY

- Bakshy, E., Hofman, J. M., Mason, W. A., and Watts, D. J. (2011). Everyone's an influencer: Quantifying influence on twitter. In *WSDM* (pp. 65–74). New York, NY, USA: Association for Computing Machinery.
- Bao, P., Shen, H.-W., Jin, X., and Cheng, X.-Q. (2015). Modeling and predicting popularity dynamics of microblogs using self-excited Hawkes processes. In *WWW* (pp. 9–10).
- Cao, Q., Shen, H., Cen, K., Ouyang, W., and Cheng, X. (2017). DeepHawkes: Bridging the gap between prediction and understanding of information cascades. In *CIKM* (pp. 1149–1158). New York, NY, USA.
- Cha, M., Mislove, A., Adams, B., and Gummadi, K. P. (2008). Characterizing social cascades in Flickr. In *Proceedings of the First Workshop on Online Social Networks* (pp. 13–18). New York, NY, USA.
- Cheng, J., Adamic, L., Dow, P. A., Kleinberg, J. M., and Leskovec, J. (2014). Can cascades be predicted? In *WWW* (pp. 925–936).
- Cheng, J., Adamic, L. A., Kleinberg, J. M., and Leskovec, J. (2016). Do cascades recur? In *WWW* (pp. 671–681). Republic and Canton of Geneva, CHE.
- Coleman, J. S., Katz, E., and Menzel, H. (1966). *Medical innovation: A diffusion study*. New York: Bobbs-Merrill.
- Dutta, S., Masud, S., Chakrabarti, S., and Chakraborty, T. (2020). Deep exogenous and endogenous influence combination for social chatter intensity prediction. In *ACM SIGKDD* (pp. 1999–2008). New York, NY, USA: Association for Computing Machinery.
- González-Bailón, S., Borge-Holthoefer, J., Rivero, A., and Moreno, Y. (2011). The dynamics of protest recruitment through an online network. *Scientific reports*, 1, 197.
- Goyal, A., Bonchi, F., and Lakshmanan, L. V. (2010). Learning influence probabilities in social networks. In *WSDM* (pp. 241–250).
- HAWKES, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1), 83–90.
- Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.*, 9(8), 1735–1780.
- Jin, F., Dougherty, E., Saraf, P., Cao, Y., and Ramakrishnan, N. (2013). Epidemiological modeling of news and rumors on Twitter. In *Proceedings of the 7th Workshop on Social Network Mining and Analysis*.
- Kobayashi, R., and Lambiotte, R. (2016). Tideh: Time-dependent Hawkes process for predicting retweet dynamics. In *ICWSM* (pp. 191–200).
- Kupavskii, A., Ostroumova, L., Umnov, A., Usachev, S., Serdyukov, P., Gusev, G., and Kustarev, A. (2012). Prediction of retweet cascade size over time. In *CIKM* (pp. 2335–2338). New York, NY, USA: Association for Computing Machinery.
- Leskovec, J., Adamic, L. A., and Huberman, B. A. (2007). The dynamics of viral marketing. *ACM Trans. Web*, 1(1), 5–es.
- Li, C., Ma, J., Guo, X., and Mei, Q. (2017). DeepCas: An end-to-end predictor of information cascades. In *WWW* (pp. 577–586). Republic and Canton of Geneva, CHE.
- Masud, S., Dutta, S., Makkar, S., Jain, C., Goyal, V., Das, A., and Chakraborty, T. (2020). Hate is the new infodemic: A topic-aware modeling of hate speech diffusion on Twitter. *arXiv preprint arXiv:2010.04377*.
- Morris, S. (2000). Contagion. *Review of Economic Studies*, 67(1), 57–78.
- Myers, S. A., Zhu, C., and Leskovec, J. (2012). Information diffusion and external influence in networks. In *ACM SIGKDD* (pp. 33–41).