CS 579: Online Social Network Analysis

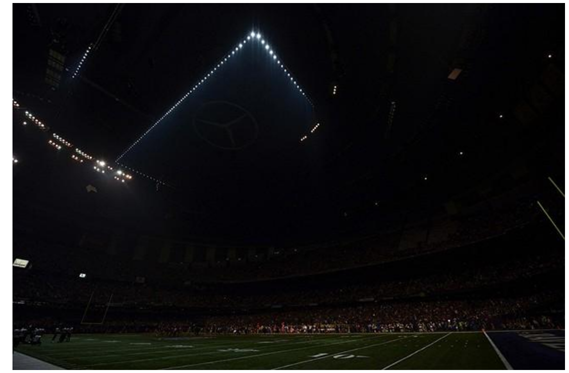
Information Diffusion in Social Media

Kai Shu

Spring 2022

What is Information Diffusion?

In February 2013, during the third quarter of Super Bowl, a power outage stopped the game for 34 minutes.



- Oreo, a sandwich cookie company, tweeted during the outage:
  - “Power out? No Problem. You can still dunk it in the dark”.
- The tweet caught on almost immediately, reaching
  - 15,000 retweets and 20,000 likes on Facebook in less than 2 days.
- Cheap advertisement reaching a large population of individuals.
  - Companies spent as much as 4 million dollars to run a 30 second ad during the super bowl.

Example of **Information Diffusion**

Information Diffusion

- Information diffusion: a process by which a piece of information (knowledge) is spread and reaches individuals through interactions

- Studied in a myriad of sciences

- We discuss methods from
  - Sociology, epidemiology, and ethnography
  - All are useful for social media mining

- How to model information diffusion

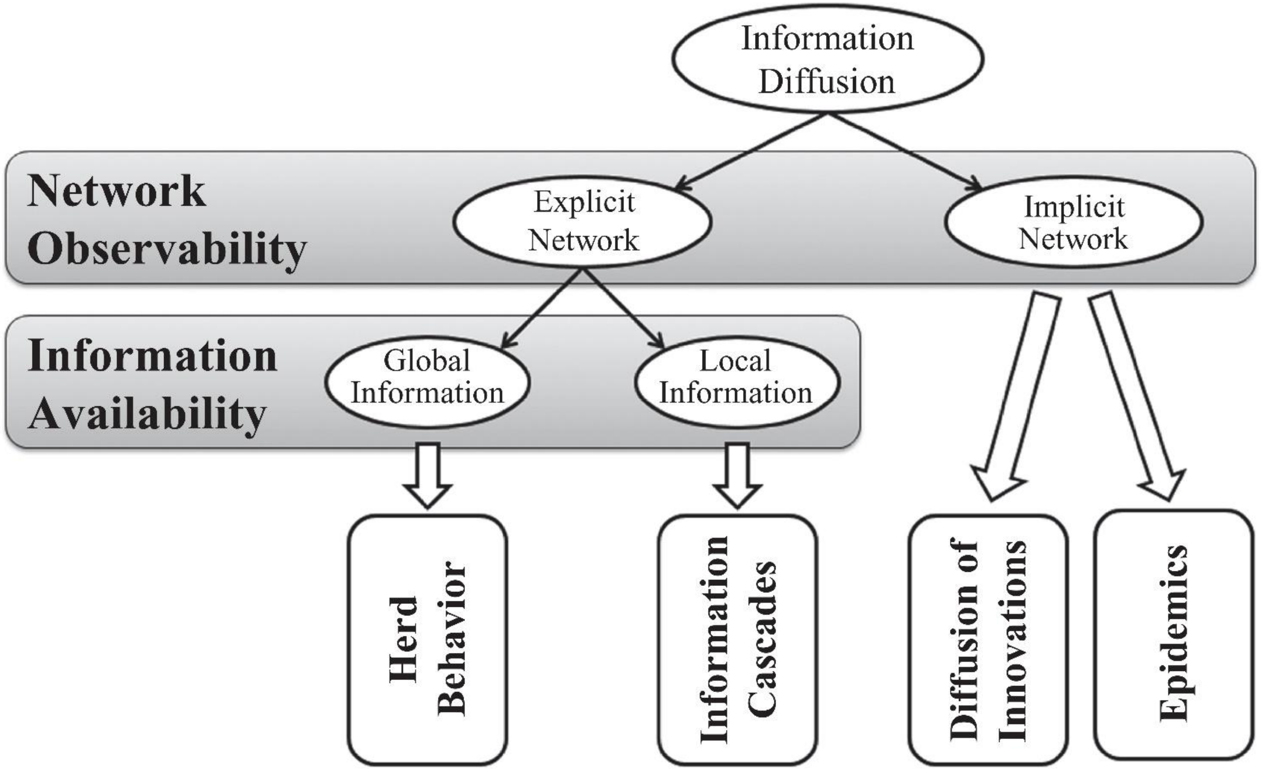
Information Diffusion

- **Sender(s).** A sender or a small set of senders that initiate the information diffusion process;

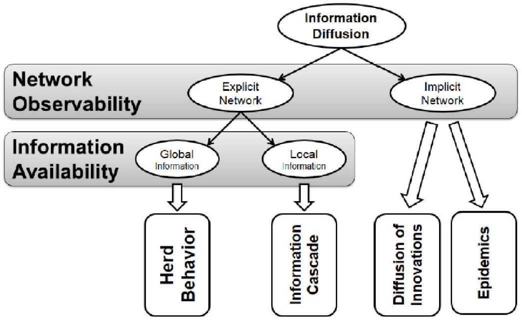
- **Receiver(s).** A receiver or a set of receivers that receive diffused information. Commonly, the set of receivers is much larger than the set of senders and can overlap with the set of senders; and

- **Medium.** This is the medium through which the diffusion takes place. For example, when a rumor is spreading, the medium can be the personal communication between individuals.

Types of Information Diffusion



Intervention is the process of interfering with information diffusion by expediting, delaying, or even stopping diffusion



Herd Behavior

- **Network is observable**
- **Only public information is available**

Herd Behavior: Milgram's Experiment

Stanley Milgram asked one person to stand still on a busy street corner in New York City and stare straight up at the sky

- About 4% of all passersby stopped to look up.

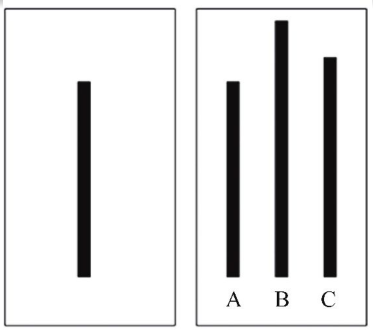
When 5 people stand on the sidewalk and look straight up at the sky, 20% of all passersby stopped to look up.

Finally, when a group of 18 people look up simultaneously, almost 50% of all passersby stopped to look up.

Herd Behavior: Popular Restaurant Example

- Assume you are on a trip in an area that you are not familiar with.
- Planning for dinner, you find restaurant **A** with excellent reviews online and decide to go there.
- When arriving at **A**, you see **A** is almost empty and restaurant **B**, which is next door and serves the same cuisine, almost full.
- Deciding to go to **B**, based on the belief that other diners have also had the chance of going to **A**, is an example of herd behavior

Herd Behavior: Solomon Asch's Experiment



- Groups of students participated in a vision test
- They were shown two cards, one with a single line segment and one with 3 lines
- The participants were required to match line segments with the same length
- Each participant was put into a group where all other group members were collaborators with Asch.
- These collaborators were introduced as participants to the subject.
- Asch found that in **control groups** with no pressure to conform, only 3% of the subjects provided an incorrect answer.
- However, when participants were surrounded by individuals providing an incorrect answer, up to 32% of the responses were incorrect.

Herding: Asch Elevator Experiment

https://www.youtube.com/watch?v=3vAKfdan0ao

Herd Behavior

Herd behavior describes when a group of individuals performs actions that are highly correlated without any plan

Main Components of Herd Behavior

- A method to transfer behavior among individuals or to observe their behavior
- A connection between individuals

Examples of Herd Behavior

- Flocks, herds of animals, and humans during sporting events, demonstrations, and gatherings

Network Observability in Herd Behavior

In herd behavior, individuals make decisions by observing all other individuals' decisions

- In general, herd behavior's network is close to a complete graph where nodes can observe most other nodes and they can observe public information
  - For example, they can see the crowd

Designing a Herd Behavior Experiment

1. There needs to be a decision made
   - For example, *going to a restaurant*

2. Decisions need to be in a sequential order

3. Decisions are not mindless and people have private information that helps them decide; and

4. No message passing is possible. Individuals don't know the private information of others, but can infer what others know from what is observed from their behavior.

Herding: Urn Experiment

- There is an urn with three marbles in it

Majority Blue
P[B,B,R]=50%

Majority Red
P[R,R,B]=50%

- During the experiment in a large class, each student comes to the urn, picks one marble, and checks its color in private
- The student predicts **majority blue** or **majority red**, writes her prediction on the blackboard, and puts the marble back in the urn
- Students cannot see the color of the marble taken out and can only see the predictions made by different students regarding the majority color on the board

Urn Experiment: First and Second Student

- First Student:
  - *Board:*
    - Observed: **B** ⇨ Guess: **B**

    -or-

    - Observed: **R** ⇨ Guess: **R**
- Second Student:
  - *Board:* ***B***
    - Observed: **B** ⇨ Guess: **B**

    -or-

    - Observed: **R** ⇨ Guess: **R**/**B** (flip a coin)

Urn Experiment: Third Student

- *If board: **B**, **R***
  - Observed: **B** ⇨ Guess: **B**, or
  - Observed: **R** ⇨ Guess: **R**

- *If board: **B**, **B***
  - Observed: **B** ⇨ Guess: **B**, or
  - **Observed: R ⇨ Guess: B (Herding Behavior)**

The forth student and onward

- Board: **B**,**B**,**B**
- **Observed: B/R ⇨ Guess: B**

Bayes's Rule in the Herding Experiment

Each student tries to estimate the conditional probability that the urn is **majority-blue** or **majority-red**, given what she has seen or heard

- She would guess **majority-blue** if:

  P[majority-blue|what she has seen or heard] > 1/2

- From the setup of the experiment we know:

  P[majority-blue] = P[majority-red]=1/2

  P[blue|majority-blue] = P[red|majority-red]=2/3

Bayes's Rule in the Herding Experiment

P[majority-blue|blue] = P[blue|majority-blue] x P[majority-blue] / P[blue]

P[blue] = P[blue|majority-blue] × P[majority-blue]
+ P[blue|majority-red] × P[majority-red]
= 2/3 × 1/2 + 1/3 × 1/2 = 1/2

P[majority-blue|blue] = (2/3 × ½)/(1/2)

- So the first student should guess **blue** when she sees **blue**
- The same calculation holds for the second student

Third Student

P[majority-blue | blue, blue, red] =
P[blue, blue, red | majority-blue] × P[majority-blue] / P[blue, blue, red]

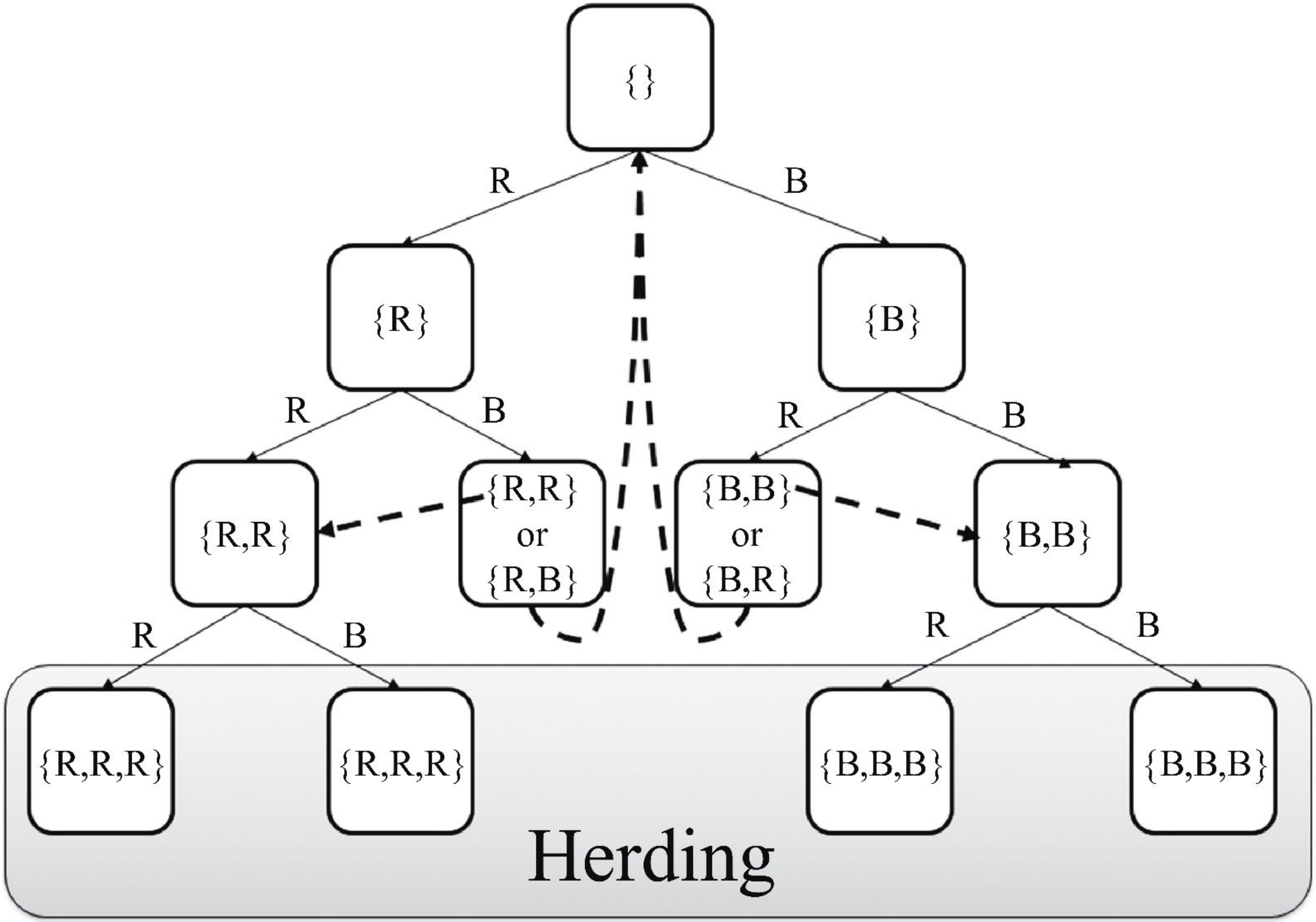
P[blue, blue, red | majority-blue] = $2/3 \times 2/3 \times 1/3 = 4/27$

P[blue, blue, red] = P[blue, blue, red | majority-blue] × P[majority-blue]
+ P[blue, blue, red | majority-red] × P[majority-red]
= $(2/3 \times 2/3 \times 1/3) \times 1/2 + (1/3 \times 1/3 \times 2/3) \times 1/2 = 1/9$

P[majority-blue | blue, blue, red] = $(4/27 \times 1/2) / (1/9) = 2/3$

- So the third student should guess **blue** even when she sees **red**
- All future students will have the same information as the third student

Urn Experiment

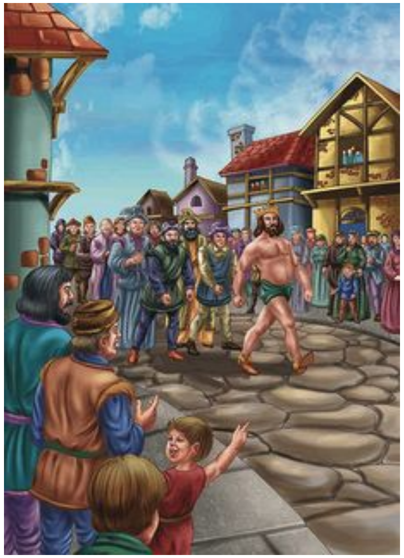


Herding Intervention

In herding, we only have access to public information.

- Herding may be intervened by releasing private information which was not accessible before

The boy in **"The Emperor's New Clothes"** story ***intervenes*** the herd by shouting **"The king is naked"**



Herding Intervention

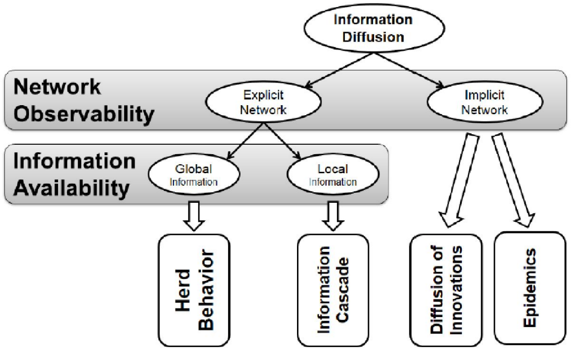
To intervene the herding effect, we need one person to *tell the herd that there is nothing in the sky* or *simply ask what they're looking at*, and the first person started looking up was probably to stop his nose bleeding ☺️

How Does Intervention Work?

- When a new piece of private information is released,
  - The herd reevaluates their guesses and this may create completely new results

- The Emperor's New Clothes
  - When the boy gives his private observation, other people compare it with their observation and confirm it
  - This piece of information may change others guess and ends the herding effect

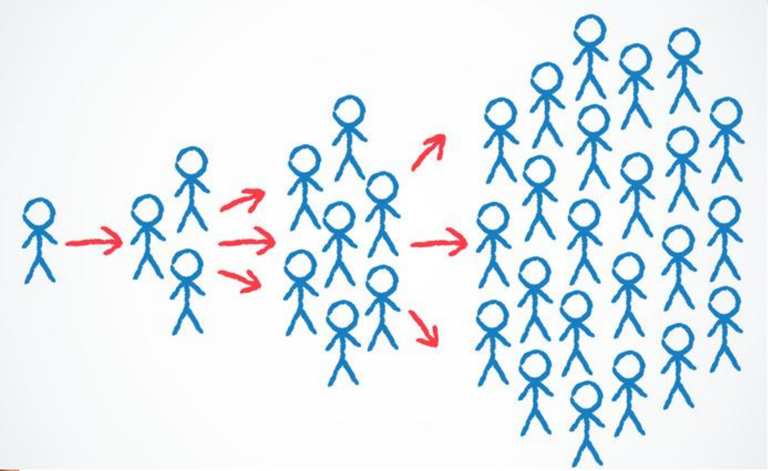
- In the urn experiment, intervention is possible by
  1. A private message to individuals informing them that the urn is majority blue (i.e., cheating ;) or
  2. Writing the observations next to predictions on the board.



Information Cascade

- In the presence of a network
- Only local information is available

Information Cascade



- Users often repost content posted by others in the network.
  - Content is often received via immediate neighbors (friends).

- Information propagates through friends

An information cascade: a piece of information/decision cascaded among some users, where

individuals are connected by a network and

individuals are only observing decisions of their immediate neighbors

Cascade users have less information available

- Herding users have almost all information about decisions

Notable example

- Between 1996/1997,
  - Hotmail was one of the first internet business's to become extremely successful utilizing viral marketing
  - By inserting the tagline "*Get your free e-mail at Hotmail*" at the bottom of every e-mail sent out by its users.

- Hotmail was able to sign up **12 million users** in 18 months.

- At the time, this was the fastest growth of any user- based company.
  - By the time Hotmail reached **66 million** users, the company was establishing **270,000** new accounts each day.

--
Get your free Email at Hotmail

Underlying Assumptions for Cascade Models

- The network is a *directed* graph.
  - Nodes are actors
  - Edges depict the communication channels between them

- A node can only influence nodes that it is connected to
- Decisions are binary; a node can be
  - **Active**: the node has adopted the behavior/innovation/decision
  - **Inactive**

- An activated node can activate its neighboring nodes; and
- Activation is a progressive process, where nodes change from inactive to active, but not vice versa

Independent Cascade Model (ICM)

- **Independent Cascade Model (ICM)**
  - A ***sender-centric*** model of cascade
  - Each node has **one chance** to activate its neighbors

- In ICM, nodes that are active are senders and nodes that are being activated as receivers
  - The *linear threshold model* concentrates on the receiver (to be discussed later in Chapter 8)

ICM Algorithm

- Node activated at time t, has one chance, at time step $t+1$, to activate its neighbors

- If v is activated at time t
  - For any neighbor w of v, there's a probability p_{vw} that node w gets activated at time $t+1$.

- Node v activated at time t has a single chance of activating its neighbors
  - The activation can only happen at $t+1$

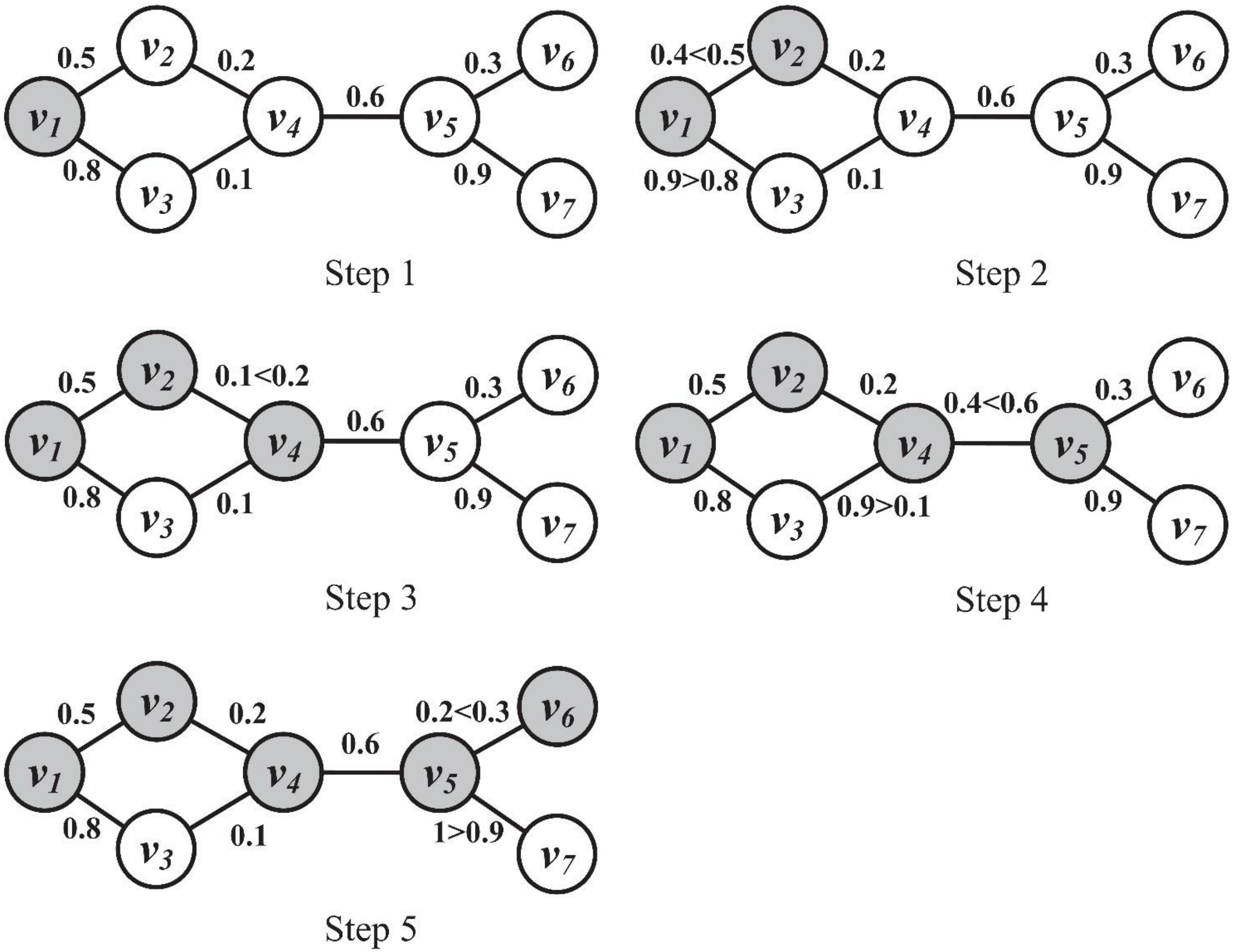
ICM Algorithm

Algorithm 1 Independent Cascade Model (ICM)

Require: Diffusion graph $G(V, E)$, set of initial activated nodes A_0, activation probabilities $p_{v,w}$

return Final set of activated nodes A_∞
i = 0;
while $A_i \neq \{\}$ **do**
  $i = i + 1;$
  $A_i = \{\};$
  for all $v \in A_{i-1}$ **do**
    for all w neighbor of $v, w \notin \cup_{j=0}^{i} A_j$ **do**
      rand = generate a random number in [0,1];
      if rand $< p_{v,w}$ **then**
        activate w;
        $A_i = A_i \cup \{w\};$
      end if
    end for
  end for
end while
$A_\infty = \cup_{j=0}^{i} A_j;$
Return A_∞;

Independent Cascade Model: An Example



- $r_{1,2} = 0.4$
- $r_{1,3} = 0.9$
- $r_{2,4} = 0.1$
- $r_{3,4} = 0.9$
- $r_{4,5} = 0.4$
- $r_{5,6} = 0.2$
- $r_{5,7} = 1.0$