Chapter 1: Introduction

1.1 Social networks

Although social media networking is a concept started during the decade 1960s, it gained popularity in recent years when people frequently used it for communication purposes. The explosion of new devices and internet-based education is the main cause of increased usage of online social media. A growing number of individuals using OSNs as a contemporary communication podium for sharing their opinions and news around the world. OSNs associate individuals across the world under one platform. After the evolution of the internet and online media, the tremendous change in how people communicate with each other. Online media allow people to propagate information faster than any other media. Figure 1.1 illustrate the various online platform used for data dissemination.

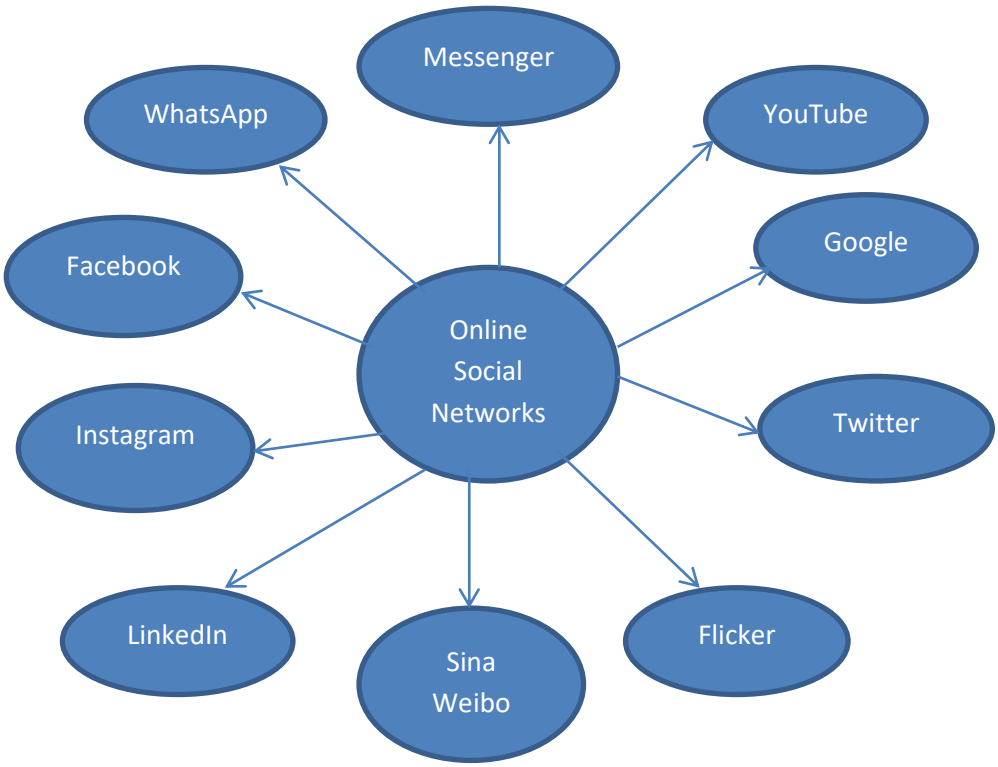

Fig. 1.1 Widely used online Social media networking sites (https://smartblogger.com/social-media-sites/)

The information is the base to connect people and share their information. If an individual disseminates any information and their neighbor shows interest in it, then the same information will be broadcast among them. Such people with similar interests connect and a community group is formed. Nowadays, Internet technology has become an essential part of our daily lives. Traditional information media like newspapers, televisions have become less prominent with the times. Undoubtedly the development of community sites has played an essential role in such revolution. Social networking sites such as Twitter, Facebook, Sina Weibo, Google, YouTube, Instagram, LinkedIn, WhatsApp and many more are popular and widely used. According to researchers 62% of individuals get news from social means of which 18% do it frequently (David et al., 2019). An analysis shows 4,406,156,312 about 4.4 billion users use the Internet globally along with 1,732,159,590 about 1.7 billion websites (Internetlive stats). Each second witnesses 2.86 million emails floating through the global web, 79K individuals searching on Google engine, 4.61 million blogs getting posted daily, 8.7K tweets being shared, 81.6K YouTube videos being viewed, 955 photos being uploaded on Instagram and 4.2K Skype video calls taking place. These numbers enhance as the social media platforms develop. A case study discussed shows an increase in social media usage and its significance to news ingesting (Newman et. al, 2013). Figure 1.2 describes the monthly active users in millions using various online platforms.

Facebook has the maximum visitors i.e. 2375 million and followed by YouTube, Instagram. Sina Weibo, a famous Chinese microblog site has 462 million active users. Twitter is mainly used for academic research and has 330 million users worldwide. WeChat, a famous Chinese messaging app has 1,112 million active users. Social platforms are the leading broadcasting medium for reporters and citizens to get the latest breaking and political news stories (Zubiaga et al., 2018).

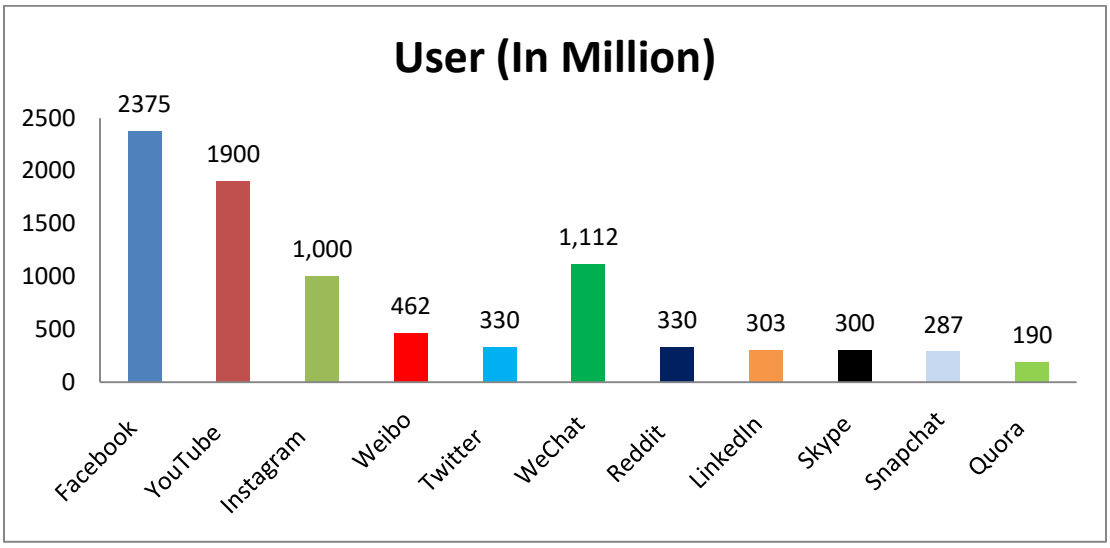

Fig. 1.2 Number of monthly active users among various online social media platform (https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/)

These sites play a vital role in politics, education, economics, marketing, government departments, news collection and much more. Compared with traditional media, these sites provide more control over users to allow more flexibility for collecting the latest news and other information. Social sites are playing a significant role in intensifying the financial world, stock market. Sentiments shown within tweets analyse market opinions, trends and examine the effect of post on product and brand (Bharucha et al., 2017). Researchers collect data from social sites and then try to remove bias from collected data. Social media is the best platform to collect the people's views regarding any current issues, like during elections to judge the leading party in exit polls.

1.2 OSNs Structure

The structure of online social networks is just like a graph G, which consists of edges and nodes. The Eq. (1.1) shows the structure of the graph:

G = {E,V} , E= Edge, V=Vertices (1.1)

The number of edges and vertices represent the structure of the graph. With the structure of the graph, it is possible to visualize social networks efficiently. Through the edges and vertices information is propagated from one node to another. The information dissemination in social networks studied to understand the following points:

1. What kind of data is disseminated more among the online media?
2. What are the reasons to spread such information?
3. How the information is disseminated?
4. Who plays the main role in information dissemination?
5. Why and how to control extensive information dissemination?

1.3 Main Contributors of OSNs in Information Dissemination

Online users spread the information from one place to another. Mainly there are three types of social network users:

1.3.1 Actors

The individual users of OSNs who can produce, accept and disseminate information over the network known as actors.

1.3.2 Neighbors

One or more individuals who can send and receive information from the actor are the neighbor of an actor. A social structure is created only when the actor finds and attaches to the neighbor node.

1.3.3 Community

Group of people that share information with each other through online sites, chat rooms, email and so on forms a community.

1.4 Information dispersion process among OSNs

The explosion of social media services such as online blogging facility and abundant revolution on new generation devices such as IoT devices and smartphones playing a prominent role in analyses information dispersion process, social actions and social performance. Through social sites immense amount of data flows from one source to another. The flow of information is very fast because of the development of different applications with different viewpoints.

In the social network when the receiver receives the message, he/she further spread it or ignores it. If the individual spread the message further then the receiver becomes the propagator otherwise the message is ignored. Every time the same process is followed. This is the general spreading process in online social media (Li et al., 2013).

1.5 Spreading Process

Susceptible-Infected-Recovered (SIR) model mostly used to analyze the spreading phenomenon. Recently, the SIR model is extensively used in the network associated investigations on learning information and rumour dispersion in social networks (Zheng et al., 2005). There are three types of property associated with every node in this model. Susceptible nodes are those nodes not infected by any type of disease yet. In future these types of nodes become infected through any other infected node and soon they recover themselves by removing the malware. Infected nodes (I) are nodes spreading misinformation among susceptible node and recovered nodes (R) are those types of nodes not spread any misinformation among other nodes.

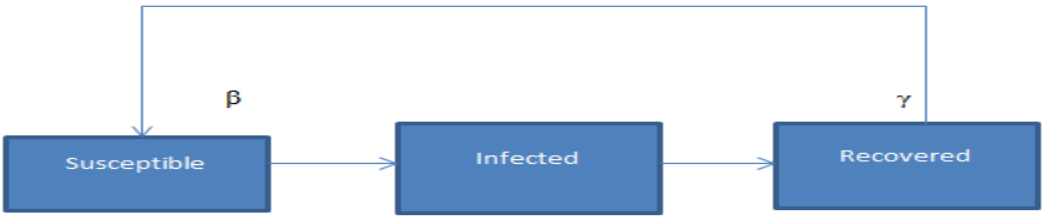

Fig. 1.3 SIR Model (Zheng et al., 2005)

Initially, all the nodes are susceptible except only one node that works as a spreader (I). With the passages of time spreader node infect other nodes with probability β. Further these infected nodes recovered with probability γ (Jin et al., 2013). The spreader of OSNs is categorized into the following:

1.5.1 Naive Spreader

Naive spreader spread the information only to their neighbor.

1.5.2 Gateway Spreaders

This spreader work as a bridge to spread information among different communities. Gateway spreader spread information in huge amount and playing a vital role in information dissemination.

1.5.3 Herding Spreaders

These spreaders are the individual actors who reach the maximum number of people and make them infected and their neighbor also.

1.5.4 Non-Spreaders

These individuals never spread any type of information among others.

1.6 Parameters to study structure of online media

Usually OSNs are scale-free systems with power-law degree distribution. OSNs have numerous different community structures having complex clustering coefficient values (Jin, 2008). New models are necessary to study this type of structure. Numerous models and approaches were developed in the previous years to study the configuration of networks. Various centrality/coreness techniques are proposed to study the significance of a node in the system.

1.6.1 Degree centrality

Total occurrence of edge on a particular node known as degree centrality. A higher degree depicts that node having higher influence.

1.6.2 Betweenness centrality

Betweenness centrality is the centrality based on shortest paths (Linton, 1978). The betweenness centrality is calculated by the Eq. (1.2).

$$g(v)= \sum_{s\#v\#t}^{n} \boldsymbol{\sigma}st(v)/\boldsymbol{\sigma}st \qquad (1.2)$$

$\boldsymbol{\sigma}st = \text{Shortest path from node s to t}$

$\boldsymbol{\sigma}st(v) = \text{Total path pass through v}$

1.6.3 Closeness centrality

Closeness centrality specifies how near a node is to another nodes in the network. In another word, information reachability from one node to another is known as closeness centrality (Sabidussi, 1966).

1.6.4 K-shell centrality

K-shell centrality estimates the spreading capability of a node by using its location and position (Miorandi et al., 2010).

Apart from above approaches there are numerous other techniques proposed to identify the structural significance of a node. Certain of them are Eigenvector centrality (Canright and Engø-Monsen, 2006), Mixed Degree decomposition (Zangh, 2013), k-hop centrality (Niu et al.,2014),Extended neighborhood coreness, Weighted neighborhood coreness (Wang et al., 2017) and many more.

1.7 Rumours among Online Networks

Though online sites play a vital role to transmit the information from one source to another. The advancement of online sites is also the main cause of rumour and

disinformation dissemination among the internet. Recent research illustrates how the rumour spreading process is faster among online sites as compared to other information (Doerr et al., 2012). Anybody having access to the internet can share any type of information. In the lack of systematic control of information dissemination in such networks rumours are simply created and disseminated across the system. Rumours create trouble and are destructive for society. Finding its core or identifying the reliability of such rumours needs a lot of determination in terms of data transmission. Various rumors related to breaking news and events damage people's beliefs in a short while. These news are unverified when they are transmitted among the media, later on these news are proven false.

1.7.1 Rumour Definitions

Different researchers provide different terminology to rumour. According to Allport and Postman “Trust of people on a specific topic mostly on current topics without knowing its truth value. Such news is transmitted by word of mouth from one person to another” (Allport and Postman, 1947).

According to DiFonzo “rumour is defined as unverified and instrumentally appropriate information that spreads uncertainty, hazard or potential danger” (DiFonzo and Bordia, 2007). Rumour is a part of information for which its trustworthiness is not checked when it is disseminated among the online media.

1.7.2 Rumours Characterization

General characterization of rumour is divided among three parts.

1.7.2.1 Information

Rumour is a message which has inconsistency related to someone, something.

1.7.2.2 Propagation

Rumours are spread by various people among the network for various reasons, like to create panic, issue of current interest and many more.

1.7.2.3 Veracity

Veracity of rumour not checked at the time of its dissemination and later on, it also not proved to be false. Once rumour spread among the media more people participate without knowing its veracity. Hence most of the time rumour spreads anxiety and chaos in the society. For example, false rumour about "Obama injured in two blasts at White House" caused 10 billion USD losses in the market (peter, 2013).

1.8 Rumours during different web waves

During the origin of the World Wide Web, numerous online sites developed for information transmission hence also increasing the rumours among online media. The history of the World Wide Web started in 1989, huge development had been made about the internet and various technologies related to it. The Figure 1.4 describes the development of various web technologies starting from web 1.0 till web 5.0 future technology. The technology process started with web 1.0 development, various tools and techniques used during various web development periods for social communication and for information sharing are described. The following section illustrates the description about web 1.0 to 5.0.

1.8.1 Web 1.0 (The Web of Contents)

The first generation of web developed in 1989 in Switzerland known as read only web. Very little interaction of users was there and interaction with the websites was not possible. Static web pages were used, it was not compatible with the machine and the source of contact was the email, fax, phone number and addresses. All the links were assigned manually by the webmaster and the role was very inactive (Newman et al., 2013).

1.8.2 Web 2.0 (THE SOCIAL WEB)

Web 2.0 recognised as a read-write web started with a conference brainstorming session, users upload and share content during this period, many websites like Myspace, Facebook, Twitter, Orkut were developed during this generation. Development of web 2.0 provides convenience to disseminate information through various communication channels, meanwhile rapid spread of rumour, fake news, hoaxes spread among the internet (Aghaei et al., 2012).

During web 2.0 generation several rumours about Barack Obama were spread by his opponents through the emails. Many of these emails related to donations include "special offer". If the people will donate $50 then 8 items related to Obama provided to them like fleece jackets (Cogburn et al., 2011).

1.8.3 Web 3.0 (THE SEMANTIC WEB)

The web 3.0 was considered a read, write and execute web. This wed provide various content and context aware features, high searching capabilities and the main motive were the linked dataset (Nath et al., 2014). Numerous rumours related to brands spread due to evolution of web 3.0, negative information about the brands produce fear among the customers (Chiou et al., 2014).

1.8.4 Web 4.0(Internet of Things)

Web 4.0 comes under the category of open, linked and smart web and its comparison is done with the human brain. It is known as a symbiotic web whose purpose is to interact between human and machine in symbiosis (Kambil, 2008).

1.8.5 Web 5.0 (The Web of Thoughts)

It is a sensory and sensitive web planned to develop machines that interact with humans. Highly intelligent web 5.0 comes after 2030 and intelligent things such as brain implants will be extremely popular.

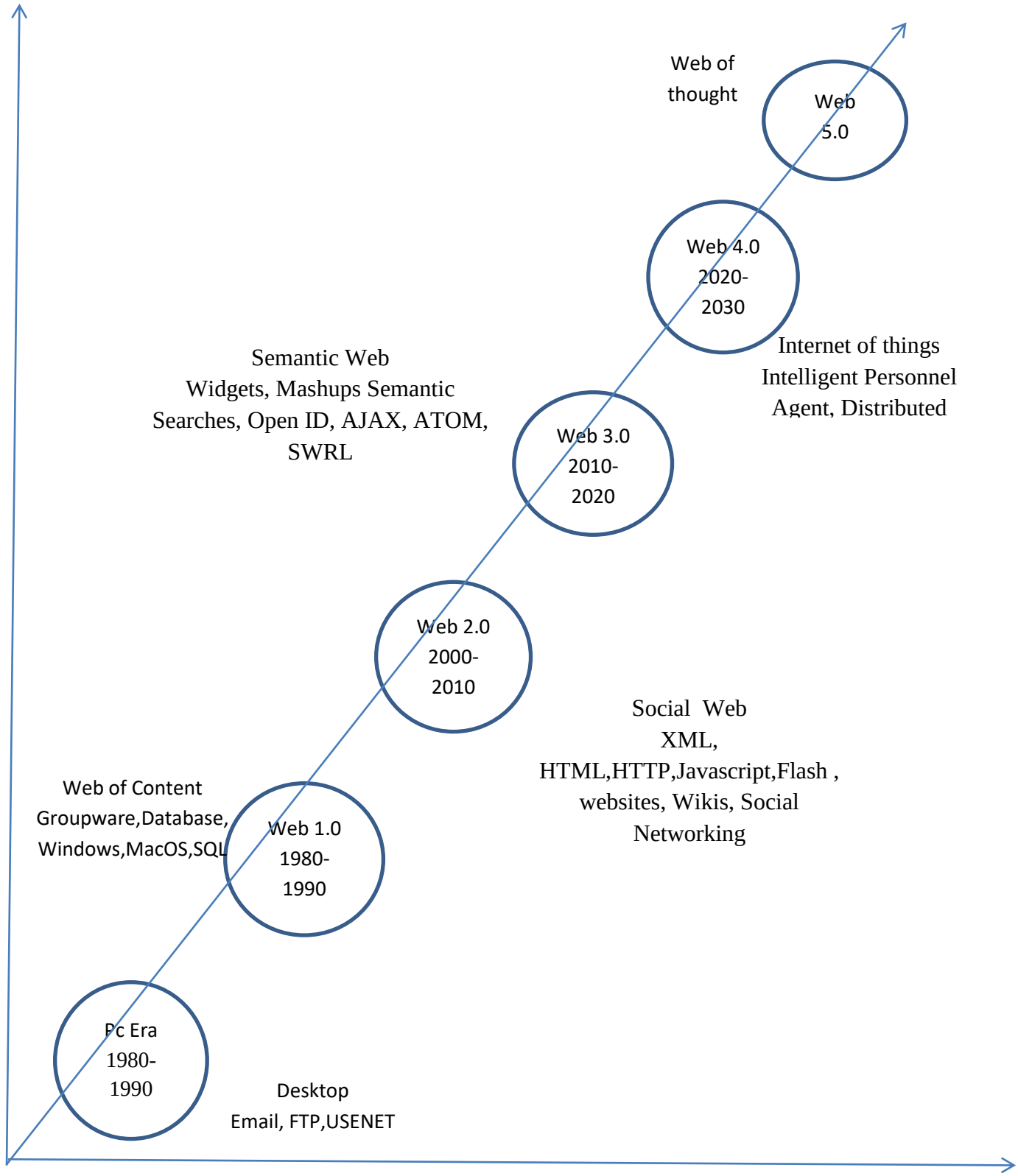

Fig. 1.4 Development of Web Waves (Patel, 2013)

Table 1.1 summarizes the different web technologies developed during various web generation periods and their business impact.

Table 1.1 Categorization of various web waves

Web Wave	Name of the phase	Business Impact
Web 1.0	Basic internet Web	Web 1.0 was the basic internet web developed by Berners-Lee (Berners-Lee et al., 1992). It was associated with many corporations. Web 1.0 was not so interactive, its usage was bound for publishing corporate information, for maintaining sales and market strategies, and transactions among consumers.
Web 2.0	The social Web	Web 2.0 was totally different from web 1.0 as it provides the facility for information sharing among different platforms. Web based applications, networking services, video call, blogs and many more services are provided by web 2.0 (O'Reilly, 2005).
Web 3.0	Semantic Web	During the web 3.0 generation human beings and AI techniques joined together to offer more significant, useful information. The semantic, 3D, geospatial web, various artificial techniques developed during this period. To promote business policies this version of the web provides many facilities (Zeldman, 2006).
Web 4.0	The symbiotic web (IoT)	The fourth generation of the web is based on wireless communications. During this period everything is operated through the internet. This web will communicate with individuals in

		a similar way as human beings communicate among each other. Shortly fourth generation web will be replaced by remote Web 5.0.
Web 5.0	Sensory emotive Web	Web 5.0 developed to design such types of computers that communicate with human beings. It is an emotive web that will measure human feelings with various techniques. One website is www.wefeelfine.org which identifies sensitive expressions among the web and categories them. Web 5.0 will absolutely be more sociable than its predecessors.

Hence large numbers of social sites are created through the origin of the World Wide Web. People use these sites for information collection as well as to maintain good relationships among family and friends.

1.9 Propagation of Rumour

Dissemination of rumour negatively affects the society and creates chaos. Hence in recent years rumour spreading has become an important topic of research. Rumours are spread from one community to another, if the primary spreader is prominent then the chances of rumour spreading increase. Rumour dispersion possesses the same properties of infectious epidemic dispersion in the population. So mostly rumour spreading models created by the inspiration from epidemic models. Numerous models have been created to mitigate the problem of rumour dispersion. The following are some epidemic models:

1.9.1 SI

SI is a simple and classical model where nodes contain only two states Susceptible (S), Infectious (I). S nodes are susceptible means unaffected nodes and I nodes are

spreading disinformation among other nodes. Many models such as SIS and SIR are derived from the SI model (Anderson, 1992).

Fig. 1.5 SI Rumour Spreading Model (Anderson, 1992)

1.9.2 SIS

In SIS nodes are either in a susceptible or infectious stage. The susceptible node transmits after contacting the infectious node. In this model the node becomes susceptible after some period (Pastor-Satorras and Vespignani, 2002).

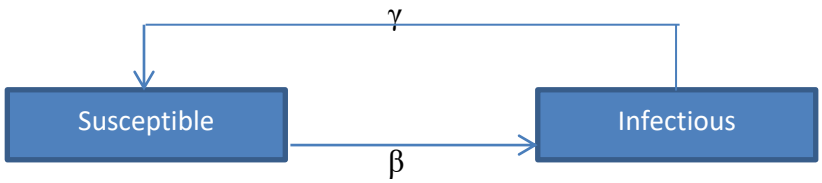

Fig. 1.6 SIS Rumour Spreading Model (Pastor-Satorras and Vespignani, 2002).

1.9.3 SIR

The SIR model contains susceptible (S), infected (I), and recovered® states. Susceptible nodes are free from any misinformation, infected nodes disperse information among susceptible nodes, and recovered nodes do not spread any type of misinformation among other nodes. Fig 1.3 represents the SIR model. In the initial state all nodes are unaffected, only one node behaves like a spreader. With the passage of time, infected nodes infect other neighbor nodes. Later on these infected nodes recover.

1.9.4 Independent Cascade

In this model nodes infect another node in an independent manner. Once a node infests their neighbor, then neighbor nodes with time duration spread misinformation among another node. Once infected the neighbor node starts misinformation dispersion among other nodes (Kempe et al., 2003).

Apart from the above discussed model, researchers invent various other models such as linear threshold (Richardson and Domingos, 2002) and voter model (Afassinou , 2014). Among all models SIR is widely used in rumour detection tasks. Numerous models developed by considering the SIR as a base model.

1.10 Actors of Rumour transmission

This section defines various kinds of actors playing a vital role to initiate or transfer rumour in online social media (Schneider-Mizell and Sander,2009).

1.10.1 Bots

Bots are an automated computer program that runs scripts from the internet. Bots are generated for web crawling purposes and are the part of Botnet that is managed by a 3rd party organization (Zannettou et al., 2019).

1.10.2 Criminals/Terrorist Groups

To create panic or spread the hazard criminal/terrorist usually spread the rumour among the ONS (Boshmaf et al., 2011).

1.10.3 Political parties

Political persons often spread the rumour to empower their image and to hurt the image of other political parties (Al-khateeb et al., 2015).

1.10.4 Trolls

The meaning of the Troll is to annoy or disrupt someone. Dispersion of trolling is faster as compared to other text or images. Troll individuals disseminate rumour to incite others or impose emotional pressure.

1.10.5 Journalists

To enhance the popularity of the journal journalist often twist the real story, mix misinformation with the actual story and then disperse it among the online media. Misinformation mix with the actual story (Mihaylov et al., 2015).

1.11 Reasons behind Rumour Dissemination

Rumour spreader disperses the rumour for various purposes. Table 1.2 illustrates the main reasons for rumour dissemination.

Table 1.2 Motivation behind information contamination

Reason	Explanation
Profit	Different actors achieve profit by broadcasting the rumours. Bots and paid organizers are directly being salaried for such jobs. Few individuals like reporters, political people are indirectly getting financial benefits.
Passion	Some individuals are keen supporters of a person, group or entity. When data reaches such persons then they spread it further without knowing its truth value.
Political Influence	Political parties spread the rumour for their own benefits and to destruct Political parties spread the rumour to enhance their image and to destroy other parties. For example, during the 2016 presidential election mostly individuals influenced by the false news disperse among online media(Allcott and Gentzkow, 2017).
Malicious Intent	Sometimes rumours are spread to hurt or damage the image of another person, organization.
Fun	Some person spread the rumour for entertainment/fun purpose.

1.12 How Rumour Spread

Efficient usage of social sites has reduced the communication gap. However it has also increased rumour messages. Rumour spreading is a diffusion process and online media is the main channel for rumour spreading. A rumour is an intentionally created false message by someone spread on the online media. Social media users then forward it to

other sites by believing that it is a genuine message. Figure 1.7 depicts the communication of rumour via various social media sites.

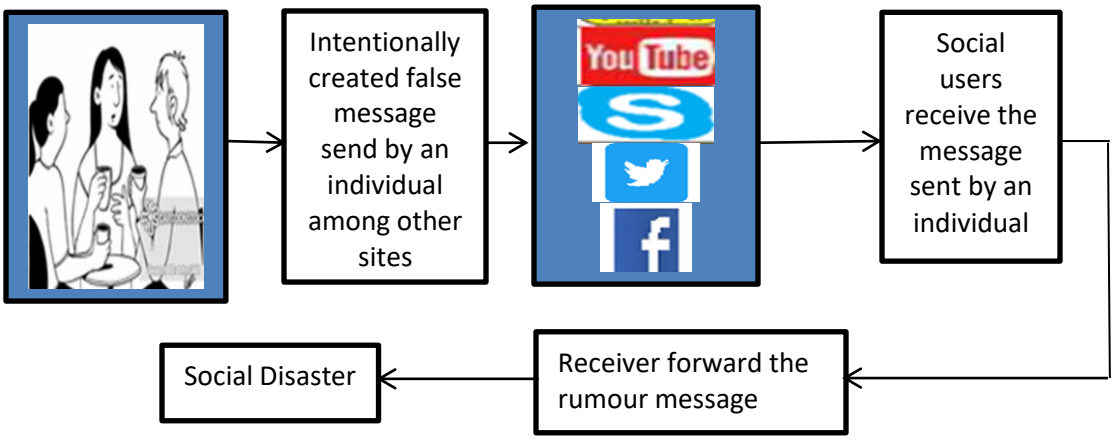

Fig. 1.7. Spreading of rumour on social media sites (Bordia and DiFonzo, 2005)

Figure1.7 describes the rumour spreading phenomena. Initially, someone creates a false message intentionally and disseminates it on various networking sites. Other social media users receive that false message and spread it further on other channels. In this way the false message spreads from one source to another, destroying information quality. People spread rumours for various purposes like to gain attention, for political purposes, enhance the stock market, for fun, and many more. Psychological views motivate people to disseminate rumour among online media (Bordia and DiFonzo, 2005). Three reasons that motivate people to spread rumors are fact-finding, maintaining relationships and self-improvement. People inspired by fact-finding understand the reasonable solution of the rumour by a problem-solving procedure; those motivated by relationship building want to interact with a person to spread false information through the media, and those inspired by self-enhancement just spread rumours either intentionally or unintentionally. Hence we can say rumour spreads for many purposes that adversely affect society. The following section discusses some case studies that identify the adverse impact of rumour among society.

In August 2012 an online rumour saying that the Assamese population dwelling in Bangalore suffered a violent attack on the last day of Ramadan created fear and panic in public. Consequently the Indian government banned text message services to handle the situation (Oh et al., 2018).

The rumour in March 1991 appeared in major Northern California newspapers about tropical fantasy soda manufactured by Ku Klux-Klan causing sterilization in black men spiraled down its sales by 70. It led to multiple attacks on the brand's delivery trucks. Similarly, another rumour related to soda was that pop rocks candy eaten with soda created a stomach problem (DiFonzo and Bordia, 2011).

Rumour changes public opinion about anything. The fake news about the Fukushima nuclear accident changed people's attitude related to nuclear energy (Kim, 2017).

A case study discussed a financial rumour related to the Bombay stock market. Such rumour affects the functionality of the market and the stockholders (Majumdar and Bose, 2018).

During the COVID-19 pandemic, the world witnessed rumours like rinse mouth with salty water, eating oregano or drinking bleach as preventive measures, and the virus being created by the US (Pulido et al., 2020).These rumors increased racist attitudes and behavior putting the government at significant risk and triggering to implement prevention measures.

Rumours and fake news spread like wildfire through WhatsApp (Farooq, 2017). The case study shows WhatsApp was the only reason for the Muzaffarnagar riots in 2013.

Brands are also vulnerable to fake news. Rumours create challenges for a brand, sometimes leading to the retirement of the brand (Borges-Tiago et al., 2020). The recent famous case was Nestle India’s Maggi noodles(Dhanesh and Sriramesh, 2018).

During the hurricane Harvey and Irma disaster event, the rumour in the media was that immigration status would be checked at the shelter. Another rumour creating a negative effect during the disaster was about plague in flooded water and shortage of fuel(Hunt et al., 2020).

1.13 Identification of rumour

The solid growth of the web waves and smart devices has started several online social applications. Hence several pieces of information are shared among the media. It is very difficult to separate false information from true information. So, identification of rumour with greater accuracy in OSNs is a recent investigation interest among industry and academic groups.

Mostly, the rumour classification is a binary or higher classification problem (Castillo et al., 2011). These classification models deliberate one and many characteristics of data being spread for this classification. Early detection of rumour prevents network damage. Moreover, it is problematic to control once the rumour spreads among a larger number of persons in the system.

There are various examples in actual life. One example, an event happened in Ferguson during 2014, which spread the rumour fast among the media. During this event a student named Micheal Brown was shot and killed by an army officer. Very soon various rumours spread about the event on social media. At the result of this incident, public prosecutor stated that this rumour diffusion among media made inquiry more problematic within a period of 24-hour (Vosoughi et al., 2017). Therefore, accurate classification of rumours in the earlier stages is very important to reduce its negative impact.

1.14 Tools of rumour detection

Rumour discovery and resolution is a protuberant research interest amongst academicians and industry specialists. Numerous applications are developed for rumour identification tasks. The following are some practical examples to handle the rumour.

Table 1. 3 Tools of rumour detection

RumourLens	RumourLens was a Twitter based application developed during 2014 to help the journalist to identify the rumour post from Twitter (Qazvinian et al., 2011).

TweetCred	TweetCred applications access the credibility of Twitter posts (Hamidian and Dia, 2019).
Check	It is a platform that checks the breaking news online. It is an open source, its trial version is available for registered users.
PHEME	It is a NLP based research project developed for rumour detection tasks. This project is funded by the European commission and numerous research works are published through this research(Zhao et al., 2015).
Emergent	It is a web based automatic rumour detection application that detects URL related rumour. Human involvement is essential to identify the rumour (Craig, 2015).

There are many other applications to combat the rumour like ClaimBuster (Zubiaga et al., 2017) Hoaxy (Vijeev et al., 2018) , REVEAL (Sivasangari et al., 2018) and so on.

1.15 Feature Used For Rumour Detection

Rumour definition is a binary classification task. Most literature on this work uses the following paradigm for supervised classification: 1) Extracting prominent features from various platforms. 2) Training a suitable classifier with an appropriate dataset. 3) Validating the dataset. The first phase for such a procedure pertains to mining the prominent characteristics. Figure 1.8 describes all kinds of features used for rumour detection.

1.15.1 Content Feature

Content features are fetched from image and text. Content is derived from three different aspects: textual content, different video and images and hashtag (#), references of users (@), outer sources links and emoji icons. Content features are further divided into two parts: textual and visual features.

1.15.2 Textual Features

These are generated via the text's linguistics and generally used in natural language processing.

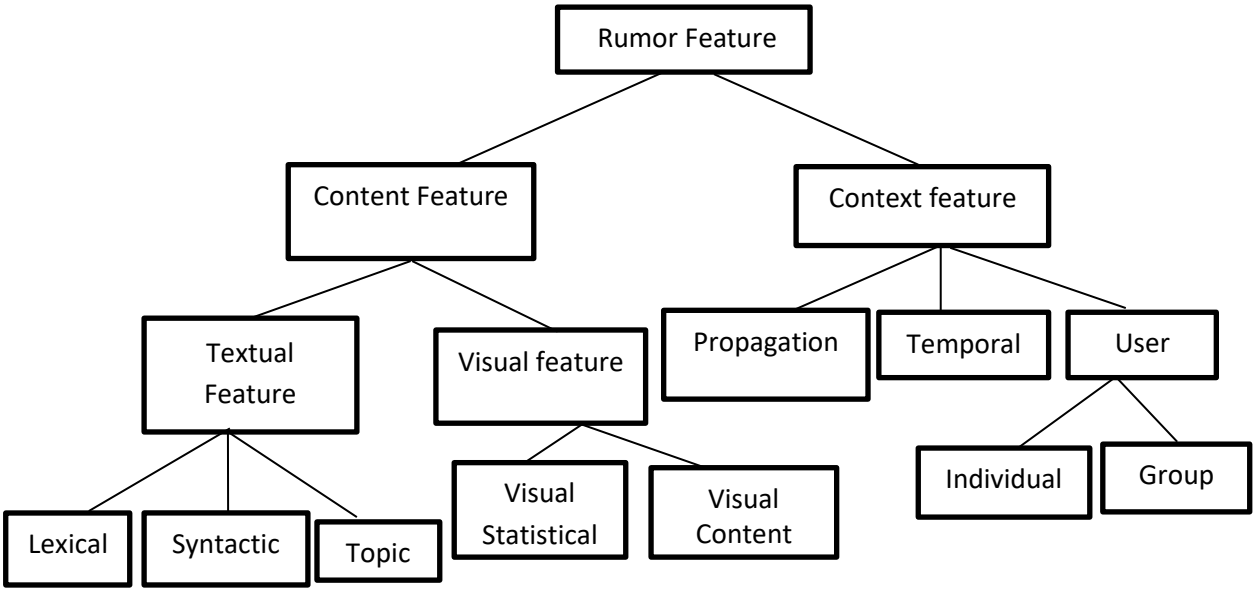

Fig. 1.8 Features used for rumour detection (Castillo et al., 2011)

1.15.3 Lexical features

These features extracted on word level. A researcher used the total amount of words, characters, length of a word for the rumour identification task (Castillo et al., 2011). Another researcher worked upon the tweet fraction, including the pronoun used by a person (Kwon et al., 2013).

1.15.4 Syntactic features

These features address the rumour on sentence level. The number of keywords, sentiment score of the sentence, part of speech tagging are used by researchers (Hassan et al., 2010). The bag of words (BOW) model is an appropriate option to present a document depending upon its words. Researchers used a bag of words technique for rumour detection tasks (Ma et al., 2017).

1.15.5 Topic features

Topic features are obtained from the message level. The main aim of these features is to understand the message and its relationship within the corpus. The researchers used topic features to detect rumours from the weibo site, Latent Dirichlet Allocation (LDA) classifiers trained for the proposed work with 18 topics. Features were fetched at the message and topic level to detect rumour (Wu et al., 2015).

1.15.6 Visual Features

Visual features describe the visual content including videos and images among different aspects and are further categorised into visual statistical features and visual content.

1.15.7 Visual statistical features

These are image statistics attached with rumour messages. A research group discussed image and health-related issues (Zhang et al., 2015). Another group pointed out rumours mainly containing out-dated images by considering time series characteristics (Sun et al., 2013).

1.15.8 Visual content

Visual content features mine the image from graphical perception like pictorial clearness, consistency and diversity.

1.15.9 Social Context features

Social content features came into existence through social communication amongst users and content including tagging, posting, reposting, commenting. Numerous features are derived from the community association for the rumour discovery task. Social features are divided into three categories: user, propagation and temporal features.

1.15.10 User features

User features identified from social networks of users to describe the features of single and group of users. Individual features are fetched from a single user characteristic, including user profile, age, gender, occupation, online criteria, number of followers, and user message. A researcher used client and location features to describe the user posting behaviour (Yang et al., 2012). In another research, the authors assessed the user's consistency via checking the user's integrity, diversity, and source location (Liu et al., 2015).

1.15.11 Group feature

The group feature is described as those who play a similar role in the rumour diffusion process.

1.15.12 Propagation features

Propagation features are identified through the fusion network. Message statistics, propagation trees, depth and size of trees were included to identify features. Some researchers worked upon the network features which were created through the comment provider (Yang et al., 2015). Another group of researchers studied certain propagation features to find the pattern of rumour diffusion (Wang and Terano, 2015).

1.16 Dataset Used

Collecting appropriate data for examination is also the main issue posing difficulties to investigators. Social media businesses have strict rules for their provided dataset. For every task different policies for data collection are necessary. For such reasons benchmark dataset and repositories are publicly available today. News agencies, search engines and social media websites are sources from where researchers can collect data. The following section describes some representative datasets collected from real-world social platforms.

1.16.1 KWON Dataset

This dataset was published in 2013. It contains 47 events related to rumour and 57 non-rumour events. All the events are collected from twitter, sixty tweets are contained by every event. To confirm the validity of the event, event labeling is done via 4 participants (Kwon et al., 2013).

1.16.2 RUMDECT Dataset

This dataset, discharged during 2016, consists of Twitter and Sina Weibo dataset (Ma et al., 2016). The Weibo dataset is collected through Sina community organization, and another dataset is collected from Twitter posts (Castillo et al., 2011).

1.16.3 BuzzFeed News

This dataset contains complete information of news published on the Facebook platform related to nine news agencies. It's five Buzz feed journalists manually check each piece of information published by BuzzFeed (Potthast et al., 2016).

1.16.4 LIAR

This dataset is gathered from PolitiFact with the help of API. It contains 12,836 human-labelled statements collected from numerous contexts like news channels, television, radio interviews and dialect (Wang, 2017).

1.16.5 BS Detector

This dataset is composed of checking news veracity from a browser extension called BS detector. It refers to the entire link on websites for unreliable sources. Instead of a human editor BS indicator does the data labelling (Cai, 2017).

1.16.6 CREDBANK

It is the largest dataset containing 60 million tweets related to 1000 news events. Thirty annotators check the credibility of each tweet from Mechanical Turk (Mitra, 2015).

1.16.7 MULTI dataset

This dataset released in 2017 was about the rumour and non-rumour posts collected from Sina Weibo. The dataset focused on multimodal content includes textual and visual information (Jin, 2017).

1.16.8 FakeNews Net dataset

The FakeNews Net dataset contains 24,000 contents about social contexts including user reactions and comments (Shu et al., 2020).

1.16.9 PHEME Dataset

This dataset contains rumour and non-rumour related to nine events (Decrzynski, 2014). Professional journalists labelled these dataset true, false, unverified as tabulated in Table 1. 4.

Table 1. 4 Various dataset for rumour detection

Dataset	Platform	Rumour	Non-Rumour
KWON	Twitter	47	55
MediaEval	Twitter	9,000	6,000
MULTI	Sina Weibo	4,749	4,779
RUMDECT	Sina Weibo, Twitter	498	494
RUMOURREVAL	Twitter	145	74
Charlie Hebdo	Twitter	458	1621
Ottawa Shooting	Twitter	470	420
Sydney Siege	Twitter	522	699

1.17 Chapter Summary

This Chapter illustrates the online social networking sites and the information dissemination process among online media. It also specifies the structure of online media, main kinds of information propagators and various methods to define the information dissemination in OSNs. Further this chapter elaborate the rumour and their characteristics. The information related to rumour transmission, various transmitted models, actors and reasons for rumour dissemination are illustrated. Further chapter reveals identification and various tools used for rumour identification. Finally the chapter defines the various types of features and publicly available dataset used for rumour detection.