TINB: a topical interaction network builder from WWW

Atul Srivastava, Anuradha Pillai, Deepika Punj, Arun Solanki & Anand Nayyar

Wireless Networks
The Journal of Mobile Communication, Computation and Information

ISSN 1022-0038

Wireless Netw
DOI 10.1007/s11276-020-02469-y

Springer

Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media, LLC, part of Springer Nature. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Author's personal copy

Wireless Networks
https://doi.org/10.1007/s11276-020-02469-y

TINB: a topical interaction network builder from WWW

Atul Srivastava[1] · Anuradha Pillai[2] · Deepika Punj[2] · Arun Solanki[3] · Anand Nayyar[4]

Accepted: 21 September 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract
Social network is a collection of people generally called 'actors' who are connected to each other based on some association criteria like a friend, follow, co-authorship, co-workers, etc. Interaction networks are the generalization of social networks. In recent developments of data sciences, analytics has applications in every significant area such as economy, general elections, epidemics, terrorism detection, clustering, marketing, etc. All of these areas require interaction data of various entities. Though the social network is a significant reservoir for such data, it covers only one segment of the information. A right amount of information is available on the web, but it is not useful for analytics in its raw form. This paper presents a framework that collects information from www using a parameterized crawler and prepares the social network-like structure of web pages, called interaction network. The interaction network prepared is similar to any traditional social network in every aspect. The web pages are selected based on contexts of the URLs found in the nearby vicinity of URLs, decided by predefined parameters. The proposed crawler is tested over several topics covering thousands of pages. More than 50 percent harvest rate is achieved by the proposed crawler. Properties of the interaction network such as degree distribution, clustering coefficient, modularity, distribution of communities, diameter and page rank have been investigated to establish the fact that it behaves like any traditional social network. The idea of preparing interaction network is extendible to the field of newage technologies like IoT, big data, deepweb, prediction models etc.

Keywords Focused crawler · Search engine · Social network · Social graph · IoT

1 Introduction

For better readability of the paper, Table 1 must be referred for abbreviations.

The social network is the interconnection of socially active individuals who are generally called actors. Social networks provide a platform where people can connect and share ideas, thoughts, opinions, likes, and dislikes. People are offered to maintain societal relationships among themselves on social networks. The interaction network is a network of entities that can be connected by the fact that they have some information in common. For instance, consider a social network, in which two users are considered having a connection between them if they are friends to each other or one of them is following the other. Users on social networks make connections if they have some common attributes such as common workplace, similar political view, same study place, common interests in sports or music, etc. Social graphs obtained from prominent social networks are used for many social studies such

✉ Arun Solanki
ymca.arun@gmail.com

Atul Srivastava
atul.nd2@gmail.com

Anuradha Pillai
anuangra@yahoo.com

Deepika Punj
deepikapunj@gmail.com

Anand Nayyar
anandnayyar@duytan.edu.vn

[1] Department of Computer Science and Engineering, Pranveer Singh Institute of Technology, Kanpur, India

[2] Department of Computer Engineering, JC BOSE UST, YMCA, Faridabad, India

[3] Department of Computer Science and Engineering, Gautam Buddha University, Greater Noida, India

[4] Duy Tan University, Da Nang, Vietnam

Published online: 06 October 2020

Author's personal copy

Table 1 List of abbreviations

SNA	Social network analysis
OSN	Online social network
WWW	World wide web
URL	Uniform resource locator
IP	Internet protocol
TINB	Topical interaction network builder
ASTRO-PH	Astro-physicist
COND-MAT	Condensed matter
HEP-TH	High energy physics theory
GF-QC	General relativity and quantum cosmology
ODP	Open directory project
RDF	Resource description framework
DMOZ	directory.mozila.org
OT	Online tutorial

as terrorism detection, monitoring epidemics, exit polls, etc. [1].

Social networks provide an open platform to analyze and understand the behavior of real-world entities, their interaction patterns and propagation of information. The explosive growth of Online Social Networks (OSNs) has assured the possibility of prominent outcomes of social network analysis. Social networks can be referred to as trivial interaction networks because they provide a platform for people to connect with each other for social activities like sharing information, photo, video, and feelings. The purpose of creating a social network was not to market analysis or any other analytical processing. Later the information and knowledge available on social networks made it extremely useful for applications related to various fields of analytical research. Interaction networks can be constructed with a similar phenomenon for non-trivial domains also. For example, there are many websites that can be put in one category based on some criteria which are similar such as the content they are showing, the technology on which the websites are developed, etc. Web pages are considered as vertices, and if two webpages have common information (showing similar content/using the same technology), an undirected edge can be put between them or if one page contains hyperlinks to other pages then directed edges can be put among them [2, 3]. The interconnected network of web pages is called a web graph. The web graph created in this manner is an interaction network of web pages which is similar to the social graph obtained from any social network. This interaction network of web pages can be used for many purposes such as clustering the web pages and identifying a particular category of websites such as pornographic websites or websites spreading fanatic ideas or the recent technologies being used by the sites.

The idea can be elaborated with the following example. For instance, consider video games for computers and mobile phones. Information can be created for games based on genre. Games based on similar genres or having the same features can be connected. Christopher John Kneifer [4] discussed the psychological effect of violent games on the brains of kids. The interaction network of games can be used to make conclusions based on the centrality of games, degree distribution, and communities formed in the interaction network of games. For instance, suppose a particular game G1 has a feature which is the root cause of a particular psychological disorder, then other games that fall in the same community in which G1 falls must also be scrutinized as a precautionary major. This can be done by applying community detection techniques, which are part of social network analysis, on the interaction network of games.

Crawling the social web has always been one of the significant options to gather data for social studies. The data on social websites are directly usable for social research because of its structural properties. As the actors are connected, they have capabilities to influence others or get influenced by others. This makes social networks to be the centre of attraction for several application areas. In recent developments of data sciences, analytics has applications in every significant area such as economy, general elections, epidemics, terrorism detection, clustering, marketing, etc. Normal web crawler crawls through URLs and gathers contents of the web pages in the repository. This content is generally used for indexing the web pages for retrieval against a search item, but it does not have the structure required for social studies. One step further to this process can be to form an interaction network of the web pages and the content they contain.

The objective of the paper can be summed up in the following points:

(1) To present basic discrimination in the process of crawling of normal web and crawling of the social web.
(2) To develop a focused web crawler, which produces an interaction network of web pages as an outcome.
(3) To establish the fact that the interaction network prepared contains structural characteristics similar to any traditional social network.
(4) To establish the fact that the interaction network prepared can be treated as any traditional social network. All social network analysis tools and techniques are directly applicable to the interaction network prepared in this paper.

Author's personal copy

(5) To present the idea of preparing an interaction network developed in this paper as extendable for other non-social domains (profiling of entities).

Section 2 briefly explains the interaction networks prepared from various non-social domains in the past and the significant topical crawling paradigms. Section 3 compares crawling paradigms, i.e. crawling the social web and crawling the normal web. Section 4 discusses the proposed methodology. It includes the idea of a parameterized crawler, the realization of regular webpages as actors of social network and an abstract view of the architecture of the proposed crawler, TINB (Topical Interaction Network Builder). Section 5 thoroughly discusses the implementation specifications and experimental results obtained. Section 6 concludes the paper with future scope.

2 Related work

Social networks are extensively popular for analytical research. Along with conventional social networks, researchers have tried to prepare social networks from unconventional domains. For instance, a citation network has been prepared in [5], which covers papers from 1993 to 2003. If a paper *i* cited a paper *j,* then there is a directed edge from *i* to *j*. A similar network of authors collaborating in the field of astrophysics ASTRO-PH, condensed matter COND-MAT, general relativity GR-QC, high energy physics theory HEP-TH have been prepared in [6]. A collaboration network of authors in the DBLP database is maintained in terms of ground truth communities [7]. Communities include authors collaborating on the same topic. A community having less than 3 nodes is removed. Another unconventional social network of individuals is LiveJournal [8], where individual members can maintain journals, blogs and can socialize with other authors or members. An interaction network is prepared that includes a network of e-mail conversations in [8]. A similar network has been prepared from comments of Wikipedia in [9] where an edge from node i to node j represents that user i have at least edited the comment page of j. Another network from Wikipedia is a network of rfa (request for admin) [9], where the edge (i, j) represents that i has voted for j. For market research purposes, a network [10] of items that are bought together on Amazon has been prepared. A temporal network of interactions on idea exchange website 'Stack Overflow' has been prepared in [11]. Based on the format of questions asked and comments three kinds of interactions have been considered in this network. A web graph of web pages from berkely.edu and Stanford.eduis prepared in [7]. The network is directed, and edges represent hyperlinks. A who-trust-whom network of individuals involved in BitCoin transactions is prepared in [12]. The users are allowed to rate others on the scale of -10 to $+10$. The edges are weighted with the accumulated scores. This is the first weighted interaction network. A hyperlink network is prepared from Reddit in [13], where the edges represent a directed connection between two communities (subreddits). CollegeMsg [14] is a temporal interaction network of users sending private messages to each other. The edge (u, v, t) means that user u had sent a message to user v at time t. An interaction network of images is prepared in [15] which is based on the metadata of images available on Flickr.com. Edges between images represent the same location, posted in the same group, gallery of the set, or images are having similar tags, etc. A review network of wines [16] from the cellar tracker is prepared which includes user reviews of the wine. A face to face interaction network is prepared in [17] of the people playing the game Resistance. If an individual speaks while looking at individual v then the edge (u, v) is weighted with the probability of participation of u while looking at v. In [18] a network of online news documents of blog pages sharing memes or a quotation frequently is prepared. If two documents share a common meme they are connected by an edge. It enables the monitoring of news documents of blog pages that share a particular meme.

Since the web has gained popularity, researchers started exploring different ways to gather information from web pages which are later used for indexing the web pages and aiding the search engines. Focused crawling is one of the most popular approaches to do this. Researchers started grilling on link contexts in web pages. Link contexts are used for topical crawling, classification of pages and search engines. Anchor text is used by McBryan [19] for indexing the URLs at www worm.com.

Similarly, in Google Search Engine [20], anchor text is used to index pages [21]. Craswel et al. [22] proved the effectiveness of anchor text for making the web pages for a search engine solely based on pages indexed by link contexts. The topical locality of the web is tested by Davison [23]. Web pages contain hyperlinks to the pages with a similar context.

Topical crawlers generally rely on the context of hyperlinks. Topical crawling works in [[24], [25]] are influential link context works. Iwazume et al. [26] used ontology along with anchor text. To guide topical crawlers, various heuristics are also used. Topical crawling works [27–29] have used standard information retrieval techniques such as cosine similarity to queries, profiles or on topic examples.

Bedi et al. [30] proposed a multithreaded semantic focused crawler (SFC) which utilizes domain ontology to specify the topic and decide seed URLs. Dong and Hussain [31] also made use of semantic focused crawling and ontological learning by designing an unsupervised framework for vocabulary based ontology learning and designed

Author's personal copy

a hybrid algorithm for matching semantically relevant topics. Du et al. [32] designed a seed URL selection approach based on user interests ontology is used to expand topic queries. Yang [33] used an ontology supported website model which provides a solution to an information agent, i.e. ontology supported information shell.

The idea of interaction networks directly fits into the newage technological field, Internet of Things (IoT). The reason is the structure of network in IoT applications. Some of the significant applications developed using IoT are as follows: In [34], Turjman has proposed a mobile IoT based model for secure data delivery. The framework uses highly dynamic topologies to save the energy consumption. A Canonical Particle Swarm (CPS) optimised data delivery system has been proposed in [35]. This framework targets multimedia data in the form of images and videos. It is an industry oriented IoT framework. An approach based on integration of Markovian process and IEEE 802.16 to improve connectivity of vehicles to roadside units has been proposed in [36]. The framework succesfully minimizes communication delay and error rates and improves throughput and QoS. A thourough study of how IoT enables smart and efficient parking systems in highly populated cities has been presented in [37]. The different techniques proposed in literature have been compared on parameters like sensors, communication protocols, software systems, security and privcacy, interoperability etc. In [38], Ullah et al. have proposed an android clone detection system for IoT devices. In all of the above applications of IoT, an interaction network of IoT devices can be created and SNA techniques can be directly applied on them.

There are several instances of preparing interaction networks from/for non-social domains to achieve various research or analytical objectives as discussed earlier in this section. Almost every interaction network is prepared from specific domain activities. In the case of the normal web, the domain is not limited to a specific topic or activity. The limitation of work done so far is that the social representation of non-social domains is done in a very restricted domain which is already focused. In this work, the target space is the World Wide Web, which covers every domain possible. The preparation of the interaction network is made focused on a specific topic by considering only topic-specific web pages while ignoring the irrelevant web pages.

3 Crawling social web vs crawling normal web

The web crawler is a program that autonomously visits web pages through hyperlinks available on the web pages and collects information available on the web pages in terms of text, images, videos and contextual information. In a way, the basic crawlers which work on normal web can be called conventional crawlers as the social web came into existence in the recent past. Furthermore, for the social web, the conventional crawler requires an ample amount of changes to handle structural differences of the social web. The conventional crawler is a program that starts from a seed and proceeds further with the help of links present at the seed. A social web crawler is similar to a conventional crawler that starts with a particular user which is a seed node for the crawler. Then it explores its friend list and finds more nodes to be crawled further. The crawler for the normal web starts with a seed URL. It collects all the information present at this seed URL page and extracts URLs from this content. These new URLs found at the current crawled page are added to the to-be-crawled-next list (URL frontier) of the crawler as shown in Fig. 1.

Both the above crawling strategies are the same in terms of basic concept but different at implementation level because a regular web page and a page of the social network have different structures and different content to be crawled. In the case of the normal web, all the content present at the page is downloaded, and the content is used for indexing the web pages. In social web page, the content lies inside different categories, e.g. the friend list of the user containing new users for the crawler to be crawled further, general information of the user like hometown, study place, workplace, etc., the content posted by the user, the pages/groups/communities liked or joined by the user, etc. The crawler may be interested in one or many of such categories.

Another difference between normal web crawlers and social web crawlers is the protection/security checks the crawler has to deal with. Normal web crawlers do not face much resistance because there are not many options available to block a crawler on a web site. One way to block a crawler on a website is protecting the website by a password. Deep webs [39, 40], where the content is hidden behind a form, are also hard to crawl. Another way is blocking IP addresses of every crawler provided that IP addresses of every crawler are known (which may or may not be a big deal). On the other hand, social web crawlers face bigger challenges such as almost every social web is password protected. Crawler requires a valid user name and password to enter into the social web. Moreover, the users on social sites get security features like hiding their friend lists, making their posts private; hiding the pages/groups/communities they have joined, hiding the personal information, protecting their profile pictures, etc. A well-protected user is a dead-end for the social web crawler. Architectures of normal web crawling and social web crawling are shown in Figs. 1 and 2. In [41], Srivastava et al. proposed an unbiased crawler for social networks.

Author's personal copy

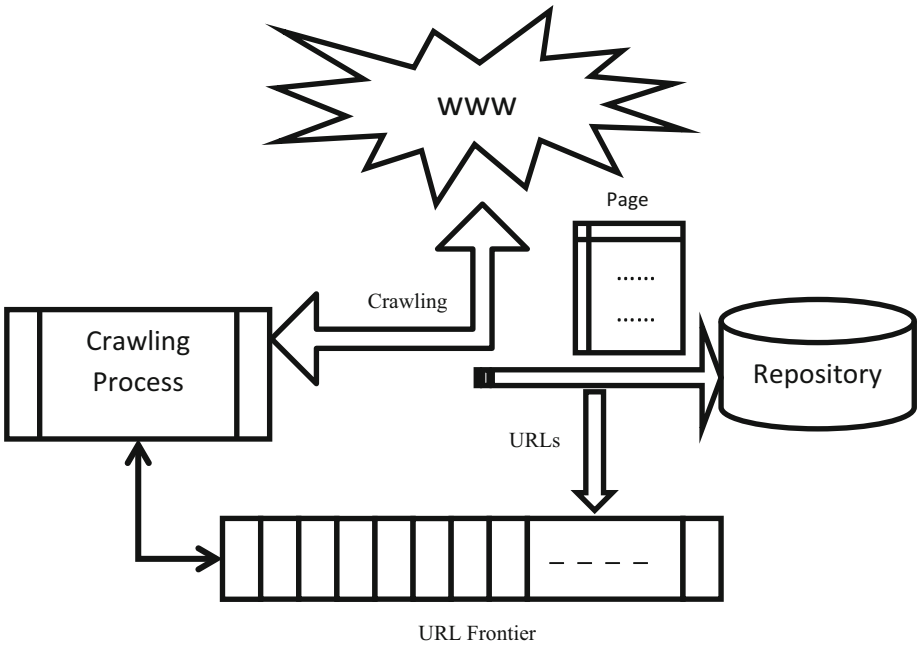


Fig. 1 Normal web crawling architecture

4 Proposed work: parameterised crawler for normal web

A normal web crawler starts with a seed URL. It visits the web page of the seed URL and extracts all the URLs present at this page for further crawling. Instead of adding every URL found at the parent page to the frontier, the crawler selects only those URLs which satisfy a certain set of parameters.

Predefined parameters are used to create a boolean expression. Boolean expression is the definition of the crawler's focus. The keywords related to the focus of the crawler with their positive or negative orientations are used as truth values in the expression. These truth values are ANDed and ORed to create an appropriate definition of the focus of the crawler. Every URL is tested with this expression, and only those URLs whose contexts satisfy the expression are visited by the crawler.

The context of a link in the text present around the links on the web page [24, 25]. One of the early attempts confines the context to the anchor text only [26], i.e. the URL and the text used within the anchor tag are considered as context for the URL. As shown in Fig. 3, the text was written between anchor tag (*Click for best anchor tag*

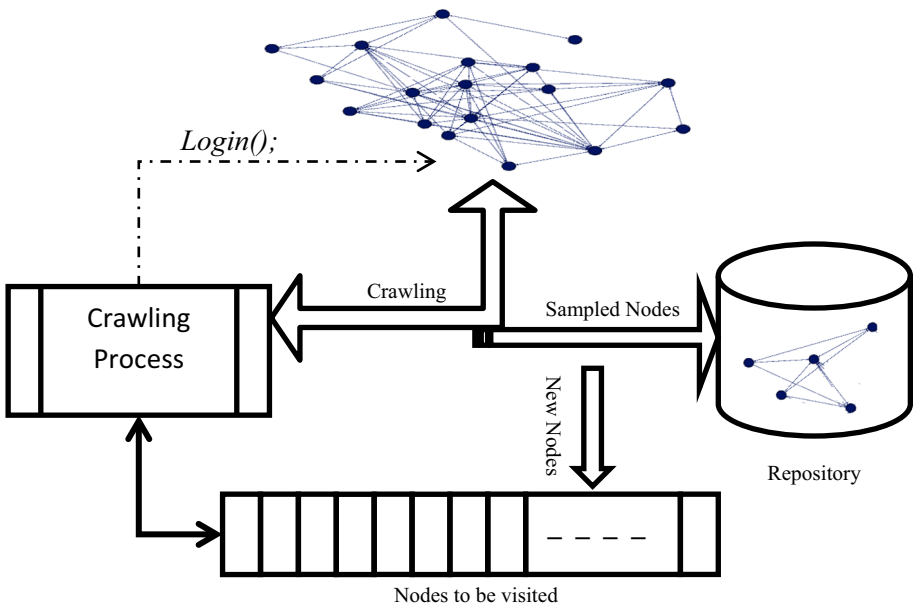


Fig. 2 Social web crawling architecture

Author's personal copy

Fig. 3 Anchor tag example

```
<a href="https://www.demoanchortag.com">Click for best anchor tag example.</a>
```

example.) and the link (https://www.demoanchortag.com) is considered as context for URL. Only this information is tokenized to generate keywords, which are passed into the Boolean expression for determining the focus.

Another developed approach used anchor text and some the text in its vicinity (or text window) [27] as a context for the URL. A text window is selected and captured within which the target URL lies. As shown in Fig. 4, the text coming in the vicinity of the window is considered as context for the URL.

The text in the vicinity of the text window is now considered as context for the URL. It is processed as follows:

1. *This text window is parsed and tokenized to find significant keywords. The text may contain several insignificant words such as 'is,' 'the,' 'a,' 'an,' 'on,' 'to'* etc. *It may also contain adjectives and adverbs, which are insignificant to determine the focus area. Such tokens are ignored during the identification of keywords.*
2. *Every other URL in this text window is ignored. Only the target URL and other text are considered. The other URLs are treated independently when they are about to be visited by TINB.*
3. *The keywords obtained in the text window are categorized in different categories corresponding to the Boolean expression; it has to satisfy. The Boolean expression defines the focus area of TINB. The boolean expression contains contextual information of the focus area defined in terms of keywords. If similar keywords are discovered around the target URL then this URL is related to the focus area. The keywords obtained from parsing are then tested against the Boolean expression. Attributes in the Boolean expression are defined next.*
4. *The expression has attributes like domain_of_URL, key_in_URL, Key_in_left_of_URL, Key_in_right_of_URL and Key_in_anchor.*
5. *Attribute domain_of_URL maps target domains of the URLs. Mostly the domain of the URL is informative to the topic to which the web site is related. For instance, domain.ac.in is generally used for university web sites. Therefore the domain of the URL is considered as one of the significant attributes.*
6. *Attribute key_in_URL attribute identifies the URL against the focused area. For instance, if our focus area is 'Travel' then keywords like outing, journey, hotel, bus, train,* etc. *in URL imply a higher probability of such URLs being related to our focused area.*
7. *The text window is captured around the URL. Text appearing to the left of the URL is called the left context of the URL and the text appearing to the right of the URL is called the right context of the URL.*
8. *In various situations, the relevance of the URL may be more affected by the left context or may be more affected by the right context of the URL. It depends on the focus area and the web page. Therefore, if it is not sure that which of the left or right context has more effect on the relevance of the URL, both should be given the same importance.*
9. *Attribute key_in_anchor contains only those keywords which are within the anchor tag. This is sometimes as relevant as the key_in_URL keywords, or sometimes it may contain some more trivial text like 'click here'* etc.

The focus area for the crawler is defined in terms of the values of these attributes. The values of attributes can be predefined as per the focus area. An expression is generated with these attributes, their predefined parameters and logical operators (AND, OR, NOT). The expression is generated in a fully disjunctive normal form (OR of ANDs), which means that each attribute appears in every conjunction of the expression. A prototype of the expression is shown in Eq. 1.

$$\begin{aligned}&((domain\_of\_URL =' .ac.in' \; AND \; key\_in\_URL =' university' \; AND \; Key\_in\_left\_of\_URL \\ &=' academic' \; AND \; Key\_in\_right\_of\_URL =' academic' \; AND \; Key\_in\_anchor =' academic') \\ &OR(domain\_of\_URL =' .edu.in' \; AND \; key\_in\_URL =' university' \; AND \; Key\_in\_left\_of\_URL \\ &=' academic' \; AND \; Key\_in\_right\_of\_URL =' academic' \; AND \; Key\_in\_anchor =' academic')OR\ldots)\end{aligned} \tag{1}$$

Author's personal copy

Fig. 4 Text window example

$$P(u) = \frac{Number\ of\ conjunctions\ satisfied\ by\ u}{Total\ number\ of\ conjunctions} \quad (2)$$

Every URL extracted from web pages has its own set of keywords for each attribute of the expression. These keywords for the attributes are tested against every conjunction of the expression. The URL is assigned a score based on the number of conjunctions satisfied by the URL. The score portrays the probability of the target URL is relevant to the focus area. Relevance probability of a URL u is defined as shown in Eq. 2.

A threshold value can be set for the acceptance of the probability. Any URL having relevance probability equal to or greater than the threshold value is considered relevant and it is added to the URL frontier to be crawled further. The threshold for relevance probability depends on the focus area and availability of web pages in that focus area.

The strategy for focused crawling has been devised in this section which will be used by TINB during crawling and building interaction networks. The purpose of TINB is not just to crawl www and collect data in datasets, but TINB is dedicated to convert this data into a social network of web pages. This social network of web pages is called the interaction network of web pages which is in all respect similar to any traditional social network. This implies that every web page in the interaction network is the same as a user in a traditional social network. The interaction network must be exactly similar to any traditional social network so that all SNA techniques apply to the interaction network without any modification. To achieve this, every web page is treated as a user, and it is given a profile as a user of the social network. This is called the profiling of web pages, which is explained in the next section.

4.1 Realising normal web as social web (profiling the URLs)

In any trivial social network, the users or actors have a general profile. The profile contains the basic information of that user as attributes of the user such as the home town, study place, workplace, interests, political orientation, etc. The users have ties between them which are realized as directed or undirected edges depending upon the kind of associations.

For instance, if the user i is a friend of user j then edge (i, j) is undirected because friendship is a bidirectional relationship. On the other hand, if the user i follows j, it is not necessary that j also follows i and hence edge (i, j) is undirected. Associations among users are highly affected by the attributes of the users. It has been already discussed how values of these attributes motivate interactive association among users. All SNA techniques make use of this profile oriented structure of the social network. SNA techniques use associations among users and their profile information for attribute values to derive conclusions.

The interaction network prepared from the non-social domain must also have the same structure if SNA techniques are to be applied to that. The interaction network prepared from WWW shall be eligible for SNA techniques to be applied to it without any modification in the

Author's personal copy

technique only if the interaction network has the same structure as does by any traditional social network. Therefore, in the realization of the normal web as a social network, URLs are portrayed as users. Each URL has its profile. URL profile contains general information of the URL in the form of its attributes such as *domain_of_URL, key_in_URL, Key_in_left_of_URL, Key_in_right_of_URL* and *Key_in_anchor*. Figure 5 shows a prototype of the interaction network of the normal web.

The interaction network of the normal web contains the exact same structure of associations as well as profile oriented arrangement of URLs. Every SNA technique can be applied to this interaction network without any modification in the technique. This step is called the profiling of URLs. In the next section complete architecture of the TINB crawler has been explained in detail.

4.2 Architecture of topical interaction network builder (TINB)

As shown in Fig. 6, TINB is a multithreaded four-layered crawler. TINB is designed to be multithreaded to achieve distributed computing. Each thread has a sequence of four layers namely, the Fetch layer, Processing Layer, Filter Layer, Processing Layer, and Representation Layer.

Every layer has a specific task to complete. The functioning of layers is as follows:

(A) *Fetch Layer*

The lowest layer is the Fetch layer. It is responsible for establishing the connection with the internet. URL frontier contains the URLs to be fetched further with their relevance probability (Eq. 3). Redundancy has been handled, and the URL frontier contains non-redundant URLs only. Initially, seed URLs are stored in the URL frontier. URL frontier is maintained by the Filter layer. URL frontier is a shared resource. It is shared among threads. Fetch layer selects a URL from the frontier which has the highest relevance probability. The relevance is the measurement of the relevance of the URL to the topic of focus. The fetched page is stored in a temporary repository for further processing.

(B) *Processing Layer*

Processing layer is the most responsible layer, and it performs the following tasks:

1. *It parses the complete web page and identifies URLs on the page. These URLs are extracted from the page.*
2. *The extracted URLs are checked for redundancy. Already visited URLs are discarded.*
3. *The text window is captured around the new URLs found on the page.*
4. *The text window of every URL is parsed and tokenized.*
5. *The tokens are processed to eliminate unnecessary tokens. Stemming is performed on tokens to get the tokens in their root words, e.g.that 'likes' and 'liked' are converted into the root word 'like'. Significant*

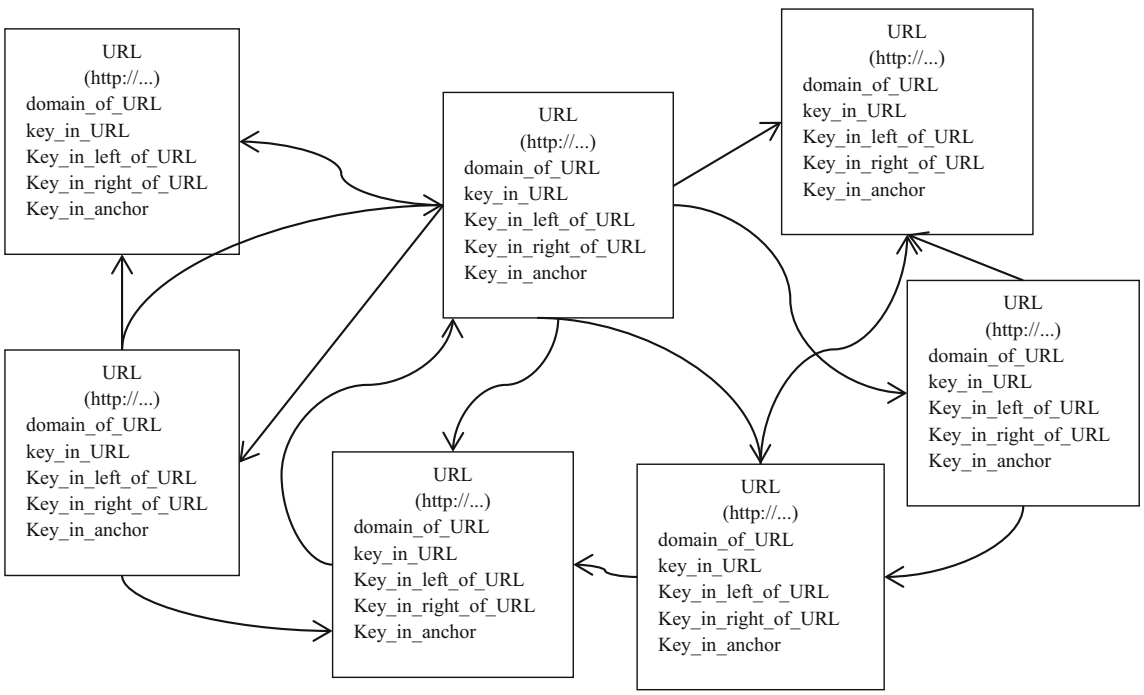


Fig. 5 Prototype of interaction network of normal web with URL profiling

Author's personal copy

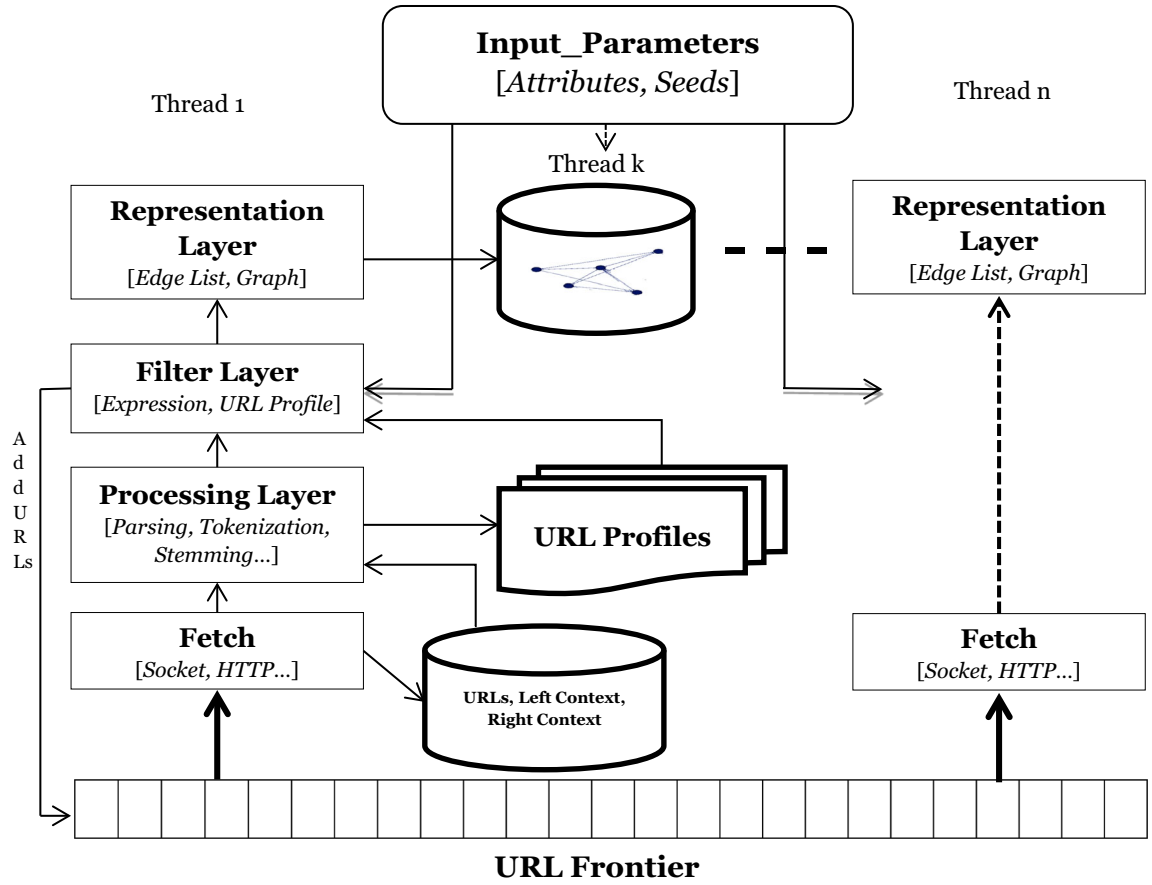


Fig. 6 Multithreaded high-level model of crawling architecture

tokens are designated to corresponding contexts of the URL.

6. *The profile of each URL is prepared. The URL profile contains attributes domain_of_URL, key_in_URL, Key_in_left_of_URL, Key_in_right_of_URL and Key_in_anchor. The tokens found in the text window are filled as values to these attributes. The association information is also maintained in the form of an adjacency list. This adjacency list is considered temporary because the URL profiles prepared in this layer will be evaluated for their relevance in the next layer, i.e. Filter layer. The URLs which are discarded in the Filter layer will not be part of the interaction network, and therefore their associations with other URLs will also be discarded. The prepared URL profiles are stored in a profile repository, URL Profiles.*

(C) *Filter Layer*

URL profiles prepared by the Processing layer are input to the filter layer. The filter layer is responsible for the quantification of the relevance of URLs to the focus topic. It performs the following tasks:

1. *This layer takes predefined parameters for the attributes and the expression defined over attributes. The Boolean expression is the definition of the focus topic. It is defined in terms of predefined values of attributes of URLs. The attributes and their predefined values are used to construct a Boolean expression. The expression has several conjunctions of attributes that are to be satisfied by the URL.*
2. *Each URL profile has values corresponding to its attributes in the form of contexts of URL. The values in attributes of URL and definition of focus area (Boolean expression) are matched.*
3. *Each URL is assigned a score called relevance probability-based on the number of conjunctions satisfied by the URL (Eq. 3).*
4. *The filter layer discards every URL having relevance probability below a threshold value.*
5. *A threshold can be set for filtering out the URLs which seem to be irrelevant to the focus area before promoting them to the next layer.*
6. *The URLs which survive the filtering process are added to the URL frontier.*

Author's personal copy

(D) *Representation Layer*

The representation layer is responsible for realizing the network of the normal web as a social web. This network is similar to any other interaction network. It performs the following tasks:

1. *It takes filtered URL profiles and from the filter layer and the temporary adjacency list prepared by the Processing layer.*
2. *The nodes which have been rejected by the Filter layer are removed from the adjacency list.*
3. *Remaining associations are appended into a central data set of the interaction network.*
4. *In the central dataset, the associations are maintained as an edge list.*
5. *URLs with their profiles are users of the interaction network.*

The above architecture of TINB successfully crawls WWW and prepares a focussed interaction network. The interaction network is similar to any other traditional social network, and hence any SNA technique is directly applicable to it.

5 Experiments and results

TINB is tested over several topics fetching thousands of pages. The exact statistics of the results have been given case wise in further text along with other outcomes. The topics are picked from the Open Directory Project (ODP).[1] ODP is a directory where URLs are maintained category wise. The directory is maintained manually, hence unbiased. The data fetched from ODP is in the RDF format. The current work advocates the use of the data and URLs in this format to initialize parameters for the focused crawl. Few topics are arbitrarily decided by defining the parameters and few known seeds. The crawler produces a collection of pages fetched from WWW. Every page crawled by the crawler is part of the interaction network. The performance of the crawler is determined by two objectives of the crawler.

- *Does the crawler maintain the focussed direction of the crawl?*

  The pages need to be evaluated in terms of their relevance to the focus area.
- *Does the crawler produce a real-world network of web pages?*

  The objective of the TINB is to prepare an interaction network of web pages which is similar to any trivial social network and keep that network as focused as possible.

To test the performance TINB, following performance metrics are identified from the literature [25]:

(A) *Harvest Rate*

Harvest rate measures the fraction of the crawled pages which are relevant to the topic of focus. Any manual judgment of relevance is not feasibly possible to conduct on millions of pages. Suppose the crawler collects t pages during the crawl, harvest rate is defined as Eq. 3:

$$h = \frac{1}{t}\sum_{i=1}^{t} P(u_i) \tag{3}$$

where $P(u_i)$ denotes relevance probability of ith page represented by URL u_i. Relevance probability ranges from 0 to 1. 0 means the page is completely irrelevant to the focus area, and one means the page satisfies every parameter of the focus area.

Figure 7 shows the harvest rate of crawls done on topics picked from the Open Directory Project (ODP) and topics configured manually. Almost 50% harvest rate is maintained by the crawler in crawls for topics selected from ODP, and a 60% harvest rate is maintained in case of topics set manually. As the web pages have several common hyperlinks on each page and much-repeated text around hyperlinks 50–60% harvest rate is significantly a good outcome.

(B) *Degree Distribution*

The degree of the vertex in a network is the number of edges incident on it. Let us define P_k be the fraction of nodes having degree k. In other words, it can be said as, P_k is the probability of any vertex chosen randomly which has degree k. A plot of P_k for any network gives the distribution of degree. For any random network, the degree distribution follows binomial or poison distributions [42–44]. But the real-world networks are unlikely to show their degree distribution as Random networks. The degrees of vertices in real-world networks are highly skewed towards the right. There are two ways to plot degree distribution. One is to construct a histogram with exponentially growing bin sizes (e.g. 1, 2–3, 4–7, 8–15…). In this scheme, the histogram is plotted on the logarithmic scale and the width of each bin remains constant. Another way is to plot the cumulative distribution of degrees represented by following Eq. 4.

$$P_k = \sum_{k' \geq k} P_{k'} \tag{4}$$

where P_k is the probability of a node of degree greater than or equal to k. This method represents all the data whereas

[1] https://dmoz-odp.org.

Author's personal copy

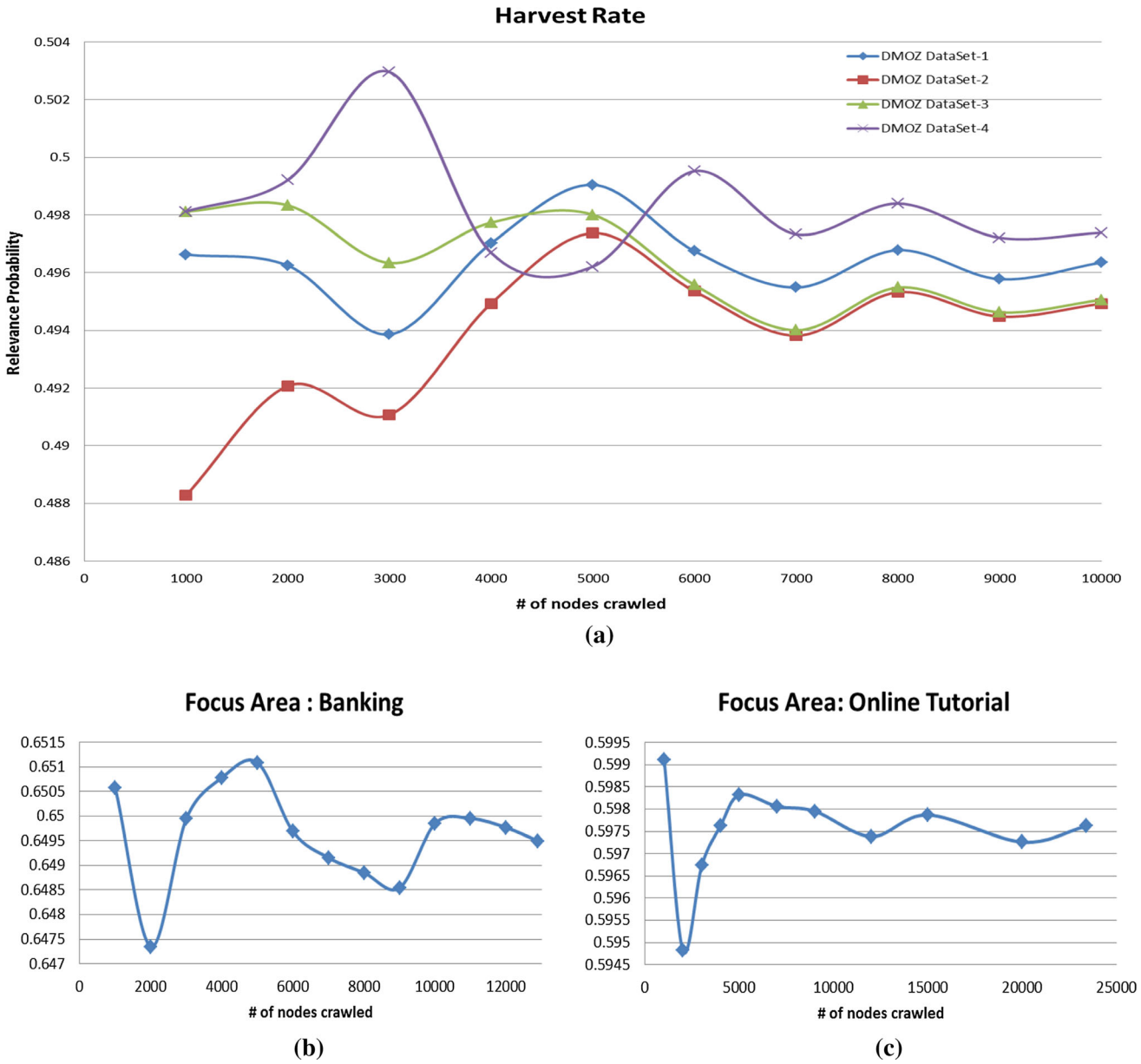


Fig. 7 Average harvest rate. The vertical axis shows the average relevance probability. **a** Topics picked from open directory project, **b** focus area is banking, **c** focus area is an online tutorial

in binning, any difference between the values of data points is lost if they fall into the same bin. Social networks are also real-world networks, and they also showcase the same degree distribution. Degree distribution of every real-world network follows the power law of degree distribution defined as Eq. 5.

$$P_k \propto k^{-\alpha} \tag{5}$$

P_k is the probability of a node having degree k and α is some constant(value of alpha depends on density and size of social network). Networks with power-law degree distribution are referred to as scale-free networks [45]. It specifies that the nodes with a higher degree in the network will be in less and nodes with a low degree will be more in number. The value of alpha depends on the density and size of the social network. Figure 8 shows the degree distribution of various interaction networks prepared by the crawler. All distributions are right-skewed and follow a power law of degree distribution.

Here, results of four crawls for topics selected from ODP and results of two crawls for topics configured manually are visualized using Gephi. Figures eight and Figs. 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20 show visualization of these interaction networks and structural analytical outcomes. The visualization represents the

Author's personal copy

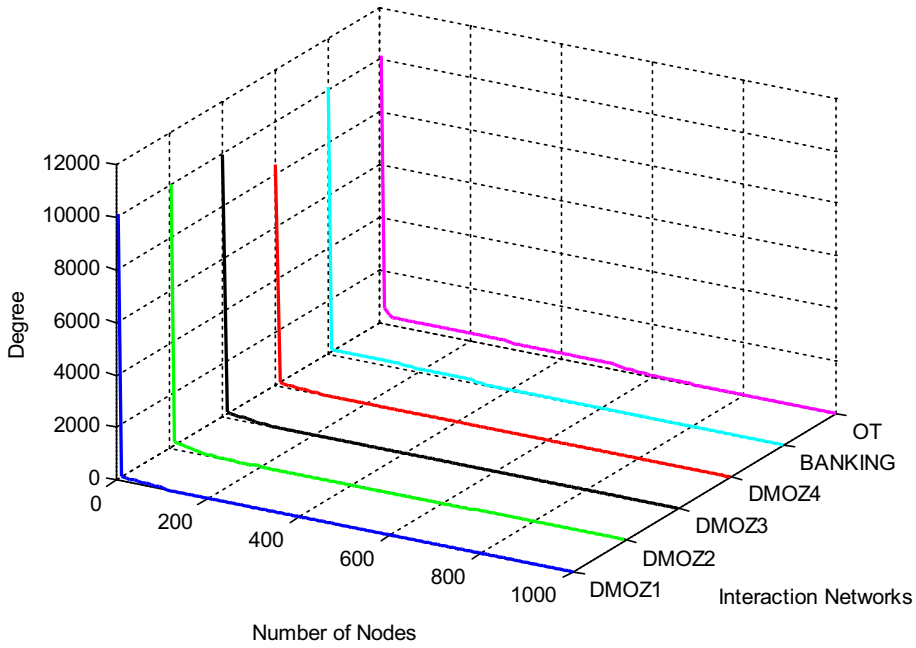


Fig. 8 Degree distribution for six different networks. The y-axis is vertex degree k and x-axis is a number of vertices having a degree greater than or equal to k. Z-axis represents six different interaction networks as follows: first four networks of topics picked from Open Directory Project, Network of topic: Banking and Network of topic: Online Tutorial

Fig. 9 DMOZ dataset-1 visualization

interconnection of the nodes graphically. The nodes represent a certain web page (URL profile), and edges represent the connection between these web pages. In all visualizations, one of several metrics of social graphs is highlighted with the help of color coding or the size of the nodes.

Figure 9 represents the interaction network prepared for the topic selected from the DMOZ directory. The topic of focus for this network is 'Travel.' Seed URLs are picked from the DMOZ directory. The network contains 20303 nodes. The average path length is 2.75 and the diameter of the network is 6. This network contains 70 communities with modularity 0.906 [46]. Figure 10 shows distribution of communities in the network. The average Clustering Coefficient is 0.017. Figure 8 shows degree distribution in the network, which follows the power law of degree distribution.

Figure 11 represents the interaction network prepared for the topic selected from the DMOZ directory. The topic of focus for this network is 'Computer Science

Author's personal copy

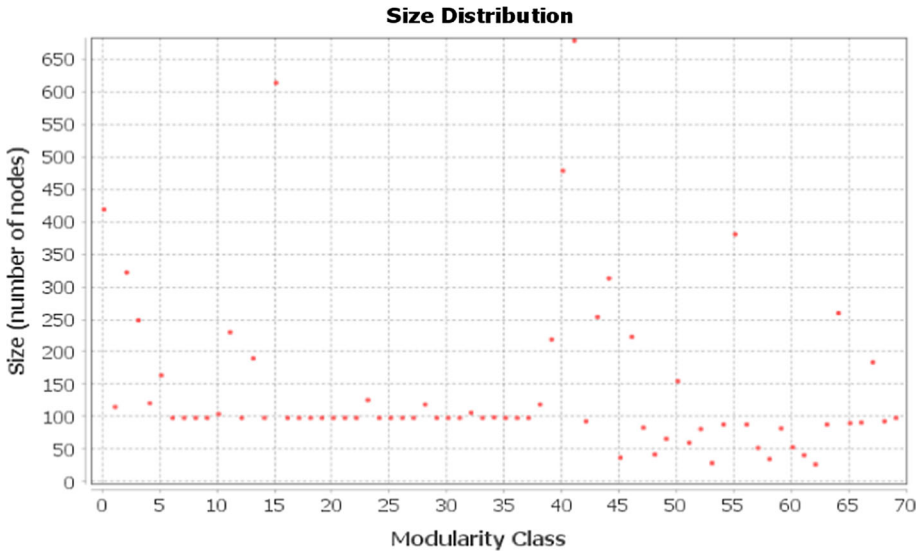


Fig. 10 Distributions of communities in DMOZ dataset-1

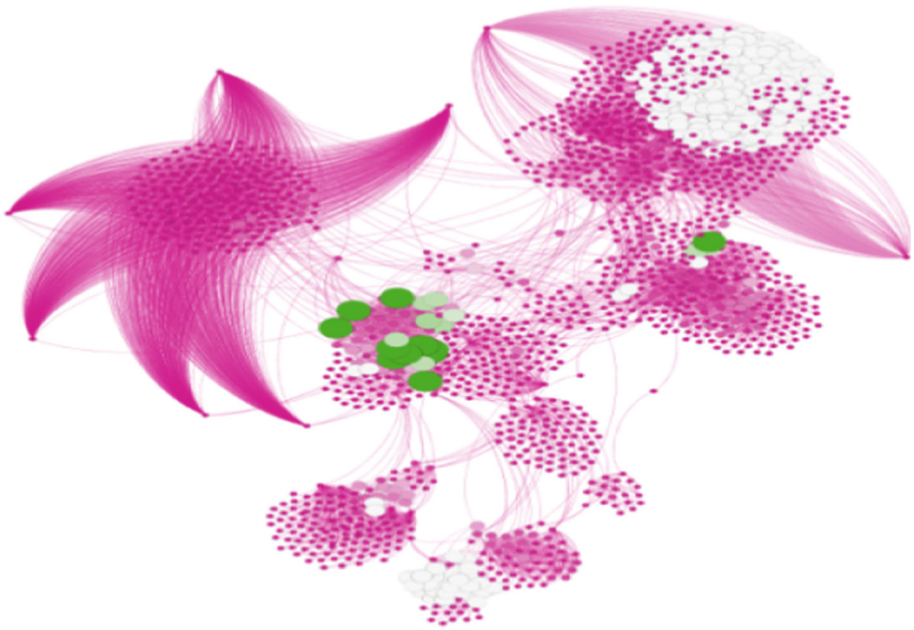


Fig. 11 DMOZ dataset-2 visualization

Engineering.' Seed URLs are picked from the DMOZ directory. The network contains 10039 nodes. The average path length is 2.17 and the diameter of the network is 5. This network contains 60 communities with modularity 0.870. Figure 12 shows distribution of communities in the network. The average Clustering Coefficient is 0.019. Figure 8 shows degree distribution in the network, which follows the power law of degree distribution.

Figure 13 represents the interaction network prepared for the topic selected from the DMOZ directory. The topic of focus for this network is 'Photography, Food, Health.' Seed URLs are picked from the DMOZ directory. The network contains 10014 nodes. The average path length is 2.14 and the diameter of the network is 5. This network contains 42 communities with modularity 0.899. Figure 14 shows distribution of communities in the network. The average Clustering Coefficient is 0.022. Figure 8 shows degree distribution in the network which follows power law of degree distribution.

Figure 15 represents the interaction network prepared for the topic selected from the DMOZ directory. The topic of focus for this network is 'Breaking News, Current News.' Seed URLs are picked from the DMOZ directory. The network contains 8427 nodes. The average path length is 2.81 and the diameter of the network is 5. This network contains 38 communities with modularity 0.880. Figure 16 shows the distribution of communities in the network. The average Clustering Coefficient is 0.004. Figure 8 shows

Author's personal copy

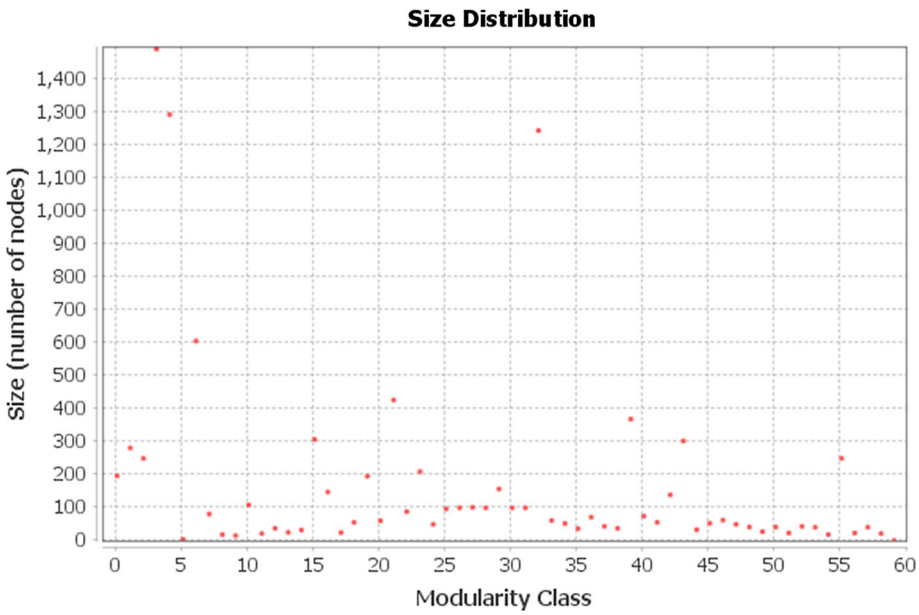


Fig. 12 Distributions of communities in DMOZ dataset-2

Fig. 13 DMOZ dataset-3 visualization

degree distribution in the network which follows the power law of degree distribution.

Figure 17 represents the interaction network prepared for the topic 'Banking.' Seed URLs are fed manually. The network contains 28945 nodes. The average path length is 2.04 and the diameter of the distribution of communities in the network. The average Clustering Coefficient is 0.044. Figure 8 shows degree distribution in the network which follows power law of degree distribution in-network is 4. This network contains 16066 communities with modularity 0.513. Figure 18 shows distribution of communities in the network. The average Clustering Coefficient is 0.044. Figure 8 shows degree distribution in the network which follows the power law of degree distribution.

Figure 19 represents the interaction network prepared for the topic 'Online Tutorial.' Seed URLs are fed

Author's personal copy

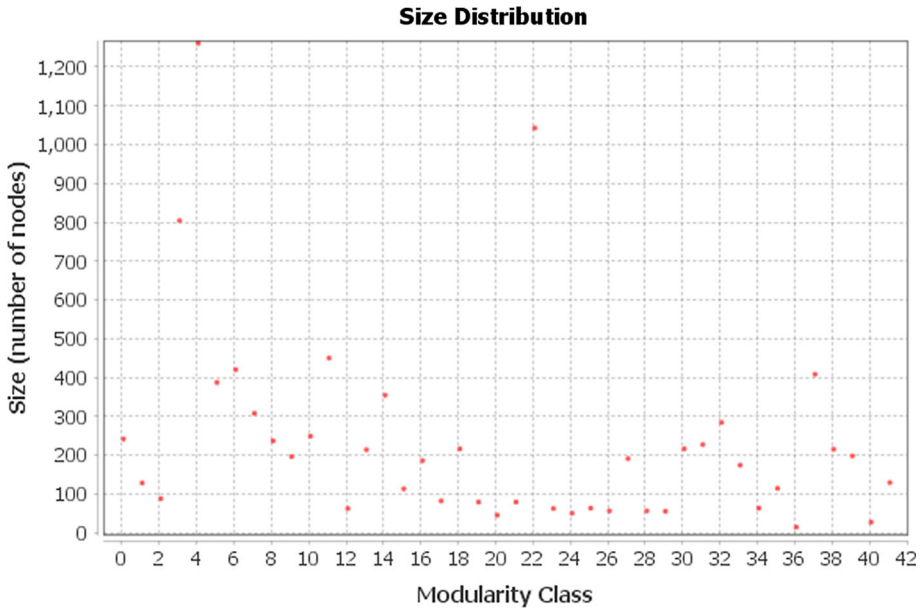


Fig. 14 Distributions of communities in DMOZ dataset-3

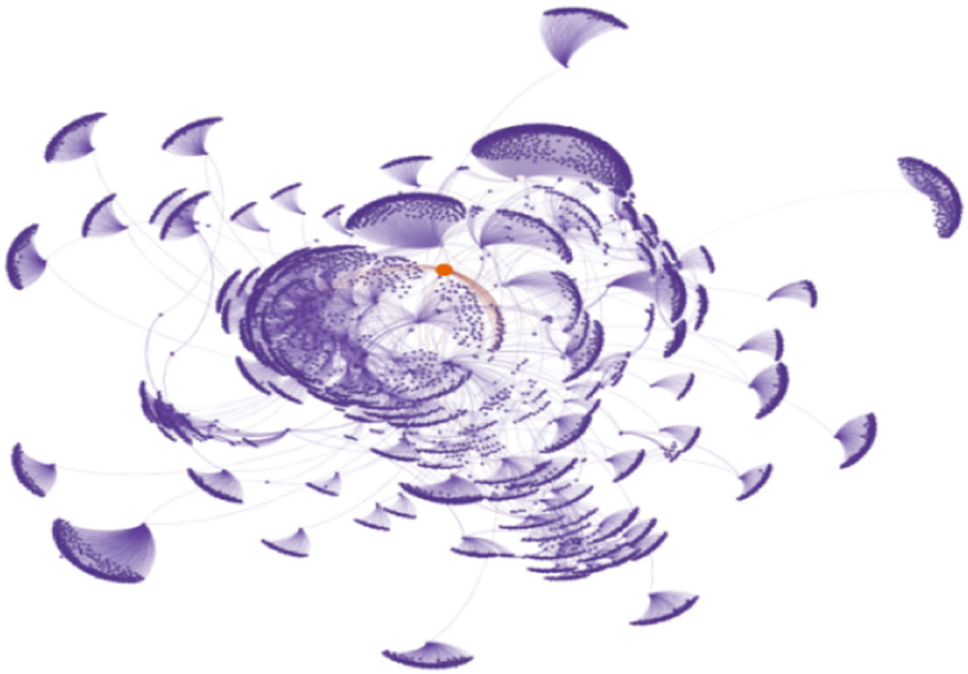


Fig. 15 DMOZ dataset-4 visualization

manually. The network contains 24278 nodes.The average Clustering Coefficient is 0.005 [20]. Figure 20 shows distribution of page rank in the network [21]. Figure 8 shows degree distribution in the network which follows the power law of degree distribution.

The above discussion establishes the fact that the interaction networks prepared by TINB for various topics of focus have structural properties exactly similar to any traditional social network. Various structural analytical procedures have been applied, and promising results are achieved from the interaction networks. Therefore, it can be concluded that the interaction networks prepared by TINB are suitable for any social network analysis procedure to be applied for social research.

6 Conclusion

Crawling a normal web and crawling a social web are two different paradigms yet having several similarities. Techniques for contextually focused crawling, available for normal web, have been efficiently used with little domain-

Author's personal copy

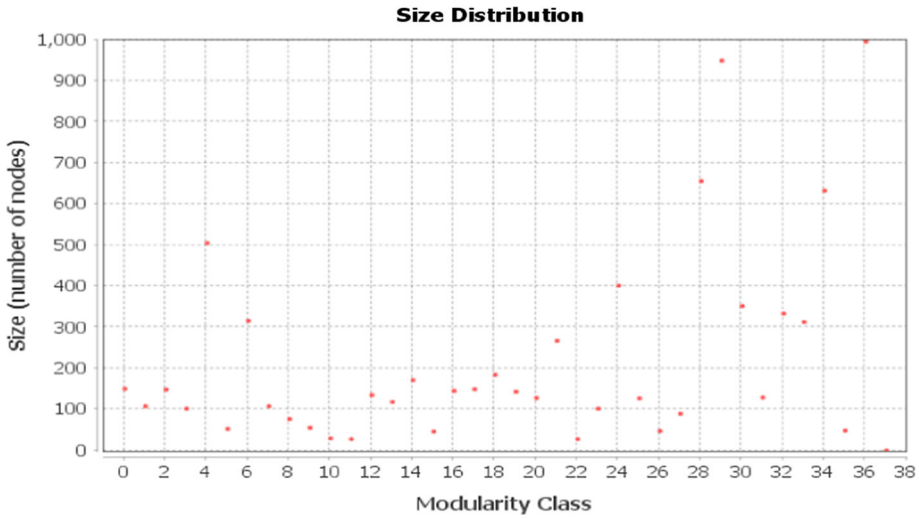


Fig. 16 Distributions of communities in DMOZ dataset-4

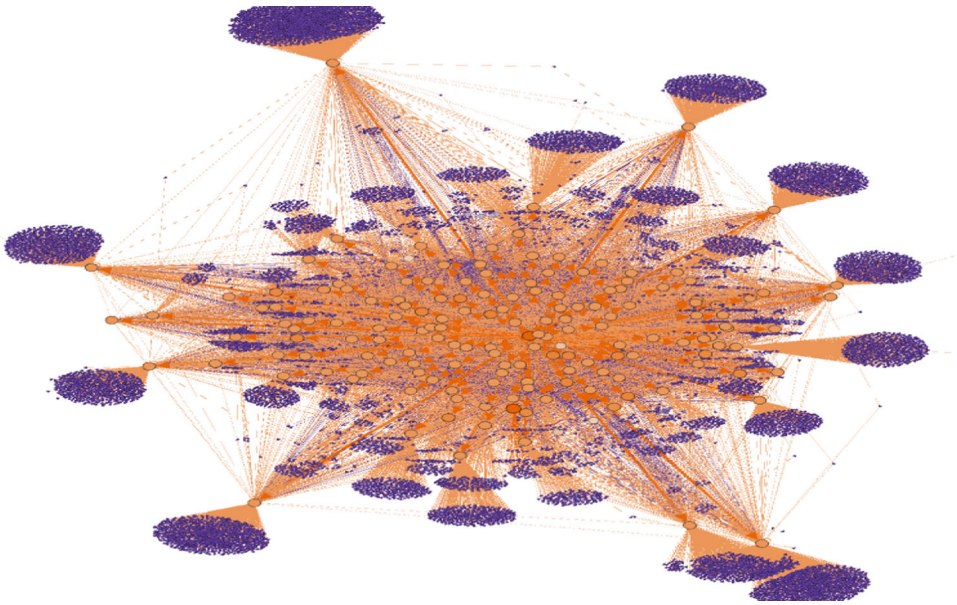


Fig. 17 Banking dataset visualization

Fig. 18 Distributions of communities in banking dataset

Author's personal copy

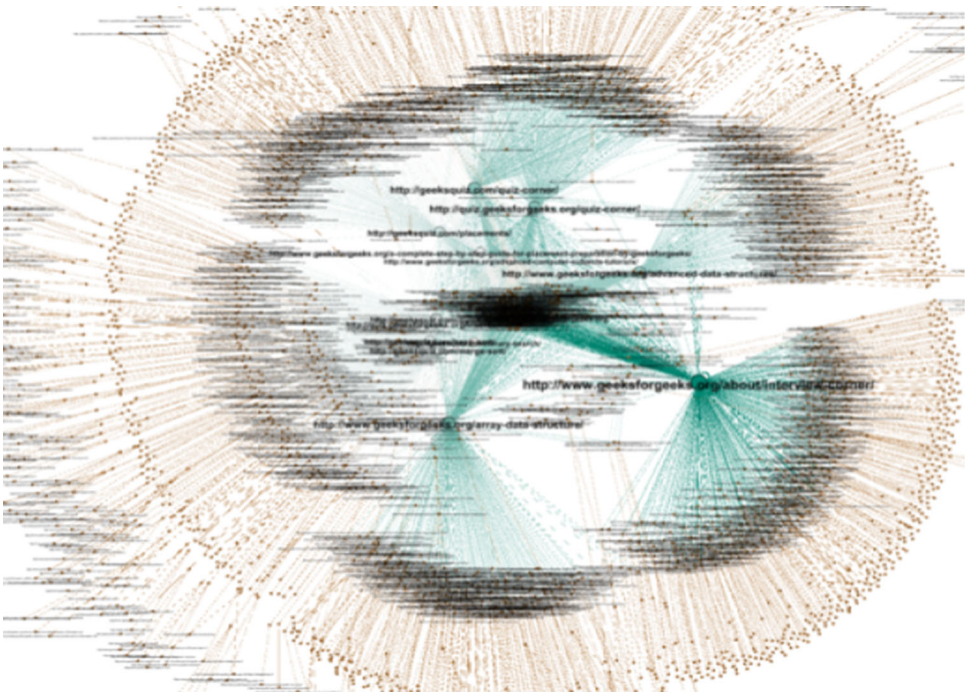


Fig. 19 Online tutorial dataset visualization

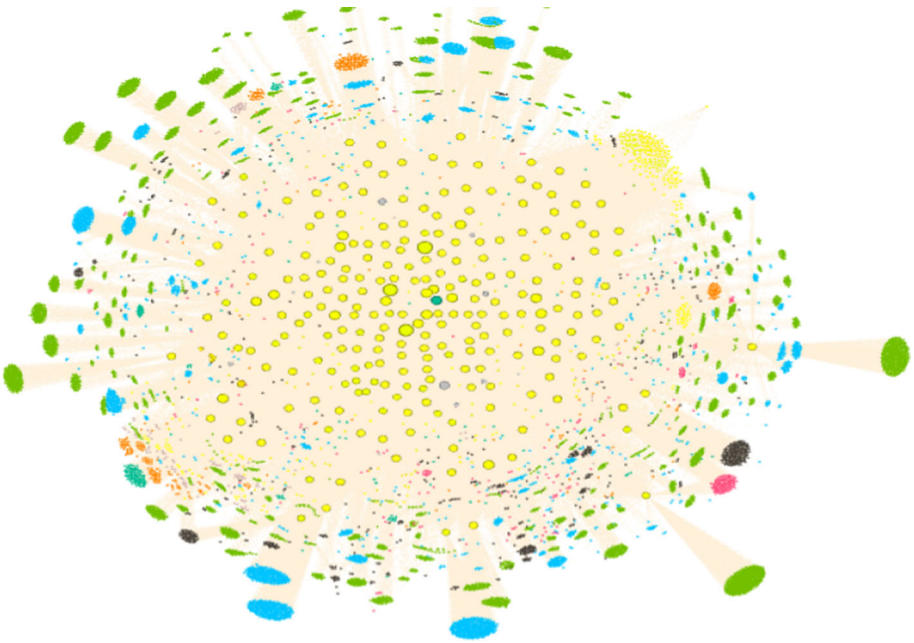


Fig. 20 Distributions of page rank online tutorial dataset

specific modification. The focused crawler TINB uses parameters based on the contextual information of the URLs to prepare an interaction network of web pages corresponding to a specific topic. This framework is tested for several topics by crawling thousands of pages for each topic. To give the structure of the social network to the network of the web pages, the URLs are treated as actors of the network with the proper profile. The attributes of the profile are parameters based on which further analytical procedures can be applied to the prepared interaction network like any other traditional social network. The performance metrics of the crawler are harvest rate, which testifies the efficiency of focused crawling and degree distribution. It also validates the scale-free structure of the prepared interaction network. The results of the experiments establish both objectives identified in the proposed work. Several other analytical experiments are performed on the prepared interaction networks to prove its applicability for social research and analytics. Every interaction network prepared by this framework contains more than

Author's personal copy

10,000 pages (some contain more than 25,000 pages). The prepared interaction networks are suitable for obtaining structural characteristics of the networks concerned with the analytics, such as clustering coefficient, modularity, communities, degree distribution, diameter, page rank, etc.

In the future, a generic interaction network of the web could be prepared in which the filtering of the pages can be done at the time of visualization or analysis. Moreover, the edges representing connections among web pages can be given weights to depict the similarity extend between two web pages. The proposed methodology seems aligned with next-generation technologies and has great potential of being applicable in fields like multimedia big data [35, 47, 48], software testing [49], content-based prediction and quality prediction [50, 51], etc.

References

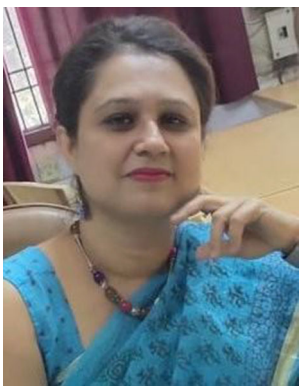
1. Srivastava, A., Pillai, A., & Gupta, D. J. (2014). Social network analysis: Hardly easy. In *2014 IEEE international conference on reliability, optimization and information technology (ICROIT)* (pp. 128–135). IEEE.
2. Choudhary, R., & Solanki, A. (2015). Improved vision based algorithm for deep web data extraction. *Journal of Web Engineering and Technology, 2*(2), 23–32.
3. Sharma, A., & Solanki, A. (2015). A hybrid page rank algorithm for web Pages. *International Journal for Scientific Research & Development, 3*(3), 3702–3708.
4. Kneifer, C. J. (2014). A comparison study on violent video games: Explained by the gamers themselves (Doctoral dissertation, University of South Florida).
5. Leskovec, J., Kleinberg, J., & Faloutsos, C. (2005). Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the 11th ACM SIGKDD international conference on knowledge discovery in data mining* (pp. 177–187).
6. Leskovec, J., Kleinberg, J., & Faloutsos, C. (2007). Graph evolution: Densification and shrinking diameters. *ACM transactions on Knowledge Discovery from Data (ACM TKDD), 1*, Article 2.
7. Yang, J., & Leskovec, J. (2015). Defining and evaluating network communities based on ground-truth. *ICDM, 42,* 181–213. https://doi.org/10.1007/s10115-013-0693-z.
8. Leskovec, J., Lang, K., Dasgupta, A., & Mahoney, M. (2009). Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics, 6*(1), 29–123.
9. Leskovec, J., Huttenlocher, D., Kleinberg, J. (2010). Predicting positive and negative links in online social networks. In *WWW*.
10. Leskovec, J., Adamic, L., & Adamic, B. (2007). The dynamics of viral marketing. *ACM Transactions on the Web (ACM TWEB) 1*(1), Article 1.
11. Paranjape, A., Benson, A. R., & Leskovec, J. (2017). Motifs in temporal networks. In *Proceedings of the tenth ACM international conference on web search and data mining* (pp. 601–610).
12. Kumar, S., Hooi, B., Makhija, D., Kumar, M., Subrahmanian, V. S., & Faloutsos, C. (2018). REV2: Fraudulent user prediction in rating platforms. In *11th ACM international conference on web search and data mining (WSDM)*.
13. Kumar, S.,Hamilton, W.L., Leskovec, J., & Jurafsky, D. (2018). Community interaction and conflict on the web. In *World wide web conference*.
14. Panzarasa, P., Opsahl, T., & Carley, K. M. (2009). Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. *Journal of the American Society for Information Science and Technology 60*, 911–932, Article 5.
15. McAuley, J., & Leskovec, J. (2012). Image labeling on a network: Using social-network metadata for image classification. In *ECCV*.
16. McAuley, J., & Leskovec, J. (2013). From amateurs to connoisseurs: modelling the evolution of user expertise through online reviews. In *WWW*.
17. Bai, C., Kumar, S., Leskovec, J., Metzger, M., Nunamaker, J. F., & Subrahmanian, V. S. (2019). Predicting visual focus of attention in multi-person discussion videos. In *International joint conference on artificial intelligence (IJCAI)*.
18. Leskovec, J., Backstrom, L., & Kleinberg, J. (2009). Meme-tracking and the dynamics of the news cycle. In *International conference on knowledge discovery and data mining ACM SIGKDD*.
19. McBryan, O. A. (1994). Genvl and WWWW: Tools for taming the web. *Computer Networks and ISDN Systems., 27*(2), 308.
20. Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology 25*, 163–177, Article 2.
21. Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems., 30,* 107–117.
22. Craswell, N., Hawking, D., & Robertson, S. E. (2001). Effective site finding using link anchor information. In *Proceedings of the 24th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 250–257).
23. Davison, B. D. (2000). Topical locality in the web. In *Proceedings of the 23rd annual international ACM SIGIR conference on research and development in information retrieval* (pp. 272–279).
24. Bra, P. M. E. D., & Post, R. D. J. (1994). Information retrieval in the world wide web: making client-based searching feasible. *Computer Networks and ISDN Systems, 27*(2), 183–192.
25. Chakrabarti, S., Berg, M. V. D., & Dom, B. (1999). Focused crawling: A new approach to topic-specific web resource discovery. *Computer Networks, 31*(11–16), 1623–1640.
26. Iwazume, M., Shirakami, K., Hatadani, K., Takeda, H., & Nishida, T. (1996). IICA: An ontology-based internet navigation system. In *Proceedings AAAI-96 workshop internet-based information systems*.
27. Hersovici, M., Jacovi, M., Maarek, Y. S., Pelleg, D., Shtalhaim, M., & Ur, S. (1998). The shark-search algorithm: An application: Tailored web site mapping. *Computer Networks and ISDN Systems, 30*(1–7), 317–326.
28. Menczer, F., Pant, G., Ruiz, M., & Srinivasan, P. (2001). Evaluating topic-driven web crawlers. In *Proceedings of the 24th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 241–249).
29. Subramanyam, M., Phanindra, G. V. R., Tiwari, M. & Jain, M. (2001). Focused crawling using TFIDF centroid. In *Hypertext retrieval and mining (CS610) class project*.
30. Bedi, P., Thukral, A., & Banati, H. (2012). A multi-threaded semantic focused crawler. *Journal of Computer Science and Technology, 27*(6), 1233–1242.
31. Dong, H., & Hussain, F. K. (2014). Self-adaptive semantic focused crawler for mining services information discovery. *IEEE Transactions on Industrial Informatics, 10*(2), 1616–1626.
32. Du, Y. J., Hai, Y. F., Xie, C. Z., & Wang, X. M. (2014). An approach for selecting seed URLs of focused crawler based on user-interest ontology. *Applied Soft Computing, 14*(Part C), 663–676.

Author's personal copy

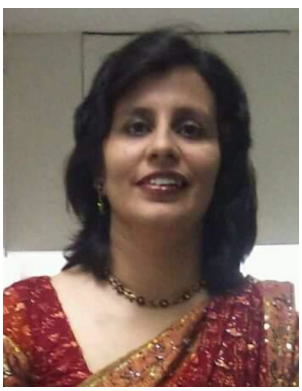
33. Yang, S. Y. (2010). A focused crawler with ontology-supported website models for information agents. In P. Bellavista, R. S. Chang, H. C. Chao, S. F. Lin, & P. M. A. Sloot (Eds.), *Advances in grid and pervasive computing. GPC* (Vol. 6104)., Lecture notes in computer science Berlin: Springer.
34. Al-Turjman, F. (2017). Energy-aware data delivery framework for safety-oriented mobile IoT. *IEEE Sensors Journal, 18*(1), 470–478.
35. Al-Turjman, F., & Alturjman, S. (2018). 5G/IoT-enabled UAVs for multimedia delivery in industry-oriented applications. *Multimedia Tools and Applications*, *79*, 1–22.
36. Al-Turjman, F. (2019). Smart-city medium access for smart mobility applications in Internet of Things. *Transactions on Emerging Telecommunications Technologies*. https://doi.org/10.1002/ett.3723.
37. Al-Turjman, F., & Malekloo, A. (2019). Smart parking in IoT-enabled cities: A survey. *Sustainable Cities and Society*, *49*, 101608.
38. Ullah, F., Naeem, H., Naeem, M. R., Jabbar, S., Khalid, S., Al-Turjman, F., & Abuarqoub, A. (2019). Detection of clone scammers in Android markets using IoT-based edge computing. *Transactions on Emerging Telecommunications Technologies*. https://doi.org/10.1002/ett.3791.
39. Singh, J., & Solanki, A. (2016). A deep web search engine for deep page. In *International conference on communication and computing systems (ICCCS-2016)*, Taylor and Francis, at Dronacharya College of Engineering, Gurgaon, 9–11 September (pp. 919–925).
40. Solanki, A. & Kumar, E. (2010). Online query submission for deep web in specific domains. In *Proceedings of 2nd International Conference on Computer Engineering and Technology, Chengdu, China, indexed in IEEE Digital Library* (vol. 3, pp. 32–34).
41. Srivastava, A., Pillai, A., & Gupta, D. J. (2018). Crawling social web with cluster coverage sampling. In M. Hoda, N. Chauhan, S. Quadri, & P. Srivastava (Eds.), *Software engineering Advances in intelligent systems and computing* (Vol. 731, pp. 103–114). Berlin: Springer.
42. Erdos, P., & Renyi, A. (1959). On random graphs. *Publ. Math. Debrecen., 6,* 290–297.
43. Erdos, P., & Renyi, A. (1960). On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutato Int. Kozl., 5,* 17–61.
44. Erdos, P., & Renyi, A. (1961). On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hungar., 12,* 261–267.
45. Barabasi, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286,* 509–512.
46. Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory Experiment, 10,* P10008.
47. Kumar, A., Sangwan, S. R., & Nayyar, A. (2020). Multimedia social big data: Mining. In S. Tanwar, S. Tyagi, & N. Kumar (Eds.), *Multimedia big data computing for IoT applications*. Intelligent Systems Reference Library (Vol. 163). Singapore: Springer. https://doi.org/10.1007/978-981-13-8759-3_11
48. Patel, D., Narmawala, Z., Tanwar, S., & Singh, P. K. (2018). A systematic review on scheduling public transport using IoT as tool. In B. Panigrahi, M. Trivedi, K. Mishra, S. Tiwari, & P. Singh (Eds.), *Smart innovations in communication and computational sciences. Advances in intelligent systems and computing* (Vol. 670, pp. 39–48). Singapore: Springer.
49. Nayyar, A. (2019). *Instant approach to software testing: Principles, applications, techniques, and practices*. Delhi: BPB Publications.
50. Diwaker, C., Tomar, P., Solanki, A., Nayyar, A., Jhanjhi, N. Z., Abdullah, A., et al. (2019). A new model for predicting component-based software reliability using soft computing. *IEEE Access, 7,* 147191–147203.
51. Gheisari, M., Panwar, D., Tomar, P., Harsh, H., Zhang, X., Solanki, A., et al. (2019). An optimization model for software quality prediction with case study analysis using MATLAB. *IEEE Access, 7,* 85123–85138.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Atul Srivastava is an Assistant Professor in the Department of Computer Science and Engineering, Pranveer Singh Institute of Technology, Kanpur, UP, India. He received Ph.D. in Computer Engineering from JC Bose University of Science and Technology, YMCA, Faridabad, Haryana, India. He Published more than 25 papers in reputed international journals and conferences. His subjects of interest include Data Mining, Information Retrieval, Hidden web, Web Mining and Social Networks.



Anuradha Pillai is an Assistant Professor in the Department of Computer Engineering, JC Bose University of Science and Technology, YMCA, Faridabad, Haryana, India. She received Ph.D. in Computer Engineering from Maharishi Dayanand University, Rohtak. She Published more than 60 papers in reputed international journals and successfully guided 4 Ph.D. students. Her subjects of interest include Data Mining, Information Retrieval, Hidden web, Web Mining and Social Networks.



Deepika Punj is working as Assistant Professor in Department of Computer Engineering at JC BOSE University of Science and Technology YMCA, Faridabad, India. She has done Ph.D in Computer Engineering. She is having 14 years of experience in teaching. She has published around 25 papers in National and International Journals. Her research interests include Data Mining, Deep Learning, Machine Learning and Internet Technologies.

Author's personal copy

Arun Solanki is working as Assistant Professor in the Department of Computer Science and Engineering, Gautam Buddha University, Greater Noida, India where he has been working since 2009. He has worked as Time Table Coordinator, member Examination, Admission, Sports Council, Digital Information Cell, and other university teams from time to time. He has received M.Tech. Degree in Computer Engineering from YMCA University, Faridabad, Haryana, India. He has received his Ph.D. in Computer Science and Engineering from Gautam Buddha University in 2014. He has supervised more than 60 M.Tech. dissertations under his guidance. His research interests span Expert System, Machine Learning, and Search Engines. He has published many research articles in SCI/Scopus indexed International journals/conferences like IEEE, Elsevier, Springer, etc. He has participated in many international conferences. He has been a technical and advisory committee member of many conferences. He has organized several FDP, Conferences, Workshops, and Seminars. He has chaired many sessions at International Conferences. Arun Solanki is working as Associate Editor in International Journal of Web-Based Learning and Teaching Technologies (IJWLTT)" IGI publisher. He has been working as Guest Editor for special issues in Recent Patents on Computer Science, Bentham Science Publishers. Arun Solanki is the editor of many Books with a reputed publisher like IGI Global, CRC and AAP. He is working as the reviewer in Springer, IGI Global, Elsevier, and other reputed publisher journals.

Anand Nayyar received Ph.D. (Computer Science) from Desh Bhagat University in 2017 in the area of Wireless Sensor Networks. He is currently working in Graduate School, Duy Tan University, Da Nang, Vietnam. A Certified Professional with 75+ Professional certificates from CISCO, Microsoft, Oracle, Google, Beingcert, EXIN, GAQM, Cyberoam and many more. Published more than 350 Research Papers in various National and International Conferences, International Journals (Scopus/SCI/SCIE/SSCI Indexed). Member of more than 50+ Associations as Senior and Life Member and also acting as ACM Distinguished Speaker. He has authored/co-authored cum Edited 25+ Books of Computer Science. Associated with more than 400 International Conferences as Programme Committee/Advisory Board/Review Board member. He is currently working in the area of Wireless Sensor Networks, MANETS, Swarm Intelligence, Cloud Computing, Internet of Things, Blockchain, Machine Learning, Deep Learning, Cyber Security, Network Simulation, Wireless Communications. Awarded 27+ Awards for Teaching and Research—Young Scientist, Best Scientist, Young Researcher Award, Outstanding Researcher Award, Excellence in Teaching and many more. He is acting as Editor-in-Chief of IGI-Global, USA Journal titled "International Journal of Smart Vehicles and Smart Transportation (IJSVST)".