Chapter 2

Background and Related Work

In this chapter, the study summarizes online social networks, describing their characteristics, the reasons behind the growth in their popularity, and the range of user interactions they allow. Then, further describes the applications of online social networks, motivating why understanding their structure and properties, is the necessary step to building future applications. The study provides the background on metrics for the analysis of complex graphs.

2.1 About Online Social Media Networks (OSMN)

Our study begins by defining online social networks, providing a brief history of their growth in popularity, and detail the mechanisms that today’s online social networks provide for users to connect and share content.

2.1.1 Definition and purpose

By definition, an online social network to be a system where (a) users are first class entities with a semi-public profile, (b) users can create explicit links with other users or content items, and (c) users can navigate the social network by
browsing the links and profiles of other users. This definition is consistent with that used in previous studies [22].

*Online social media networks* (OSMNs) help us in several ways, but the study found three most common functions provided by all the social networking sites (SNA), *first*, OSMNs are accustomed to preserving and restore the present social links, or to create new social ties. These SNS permit end-users to be “expressive and build noticeable peer networks”, through “connecting with other online people or groups whom they want to be part of their protracted (or lengthen) social network” [22], *second*, OSMNs also function as a *content holder*, on which every connected member uploads his or her matter or content. It’s noticed that the uploaded content differs from one social media site to another and also the online user’s personal profile differs and *last*, OSMNs are used to discover innovative, striking content by refining or filtering, prescribing or recommending, arranging and planning the uploaded content by the online users.

### 2.1.2 A brief history

The site Classmates.com [23] is regarded as the first website that allowed users to connect to other users. It began in 1995 as a site for users to reconnect with previous classmates and currently it has over 40 million registered users. However, *Classmates.com* did not allow users to create links to other users; rather, it allowed users to link to each other only via schools and colleges they had attended. In 1997, *SixDegrees.com* [24] was created, which was the first social networking site that allowed users to create links directly to other users. As such,
SixDegrees.com is the first site that meets the definition of an online social network from above.

Online social networks began to grow in popularity as more users became connected to the Internet. In the early 2000s, a number of general-purpose sites for finding friends were established, the most notable of which is Friendster [25]. The Friendster was focused on allowing friends-of-friends to meet, beginning as a rival to the online dating site Match.com [26]. Other, similar sites created in the same time frame include Facebook, Orkut [27] and LinkedIn [28].

In 2003, MySpace [29] was created as an alternative to Friendster and the others. MySpace allowed users to heavily customize the appearance of their profile, which proved very popular with users, causing MySpace to quickly become the largest online social network. As of this writing, MySpace has 247 million user accounts, over twice as many as the second most popular network, Facebook. For a complete history and analysis of the evolution of online social networks, the study refers the reader to the numerous papers by Boyd [30, 31 and 32].

With the rise in popularity of online social networks, many other types of sites began to include social networking features. Examples include multimedia content sharing sites (Facebook, Flickr, YouTube, and Zoomr [33]), blogging sites (Live-Journal and BlogSpot), professional networking sites (LinkedIn [34] and Ryze [35]), and news aggregation sites (Digg [36], Reddit [37], and del.icio.us [38]). All of these sites have different goals, but employ the common strategy of exploiting the social network to improve their sites. The list above is not meant to be exhaustive, as new sites are being created regularly. For a more complete and up-to-date list of the notable online social networking sites, the study refers the reader to Wikipedia [39].
The sociological aspects of the rapid growth and adoption of social networking sites are also the subject of much scholarship. One of the primary reasons that have been noted for the popularity of social networking sites is their user-centric nature. The content that is shared on social networking sites is often information about the users themselves, such as their status, photos, and so forth. For more details, the study refers the reader to the work by Boyd [40].

2.1.3 Mechanisms and policies

In a brief overview of the mechanisms and policies that most online social networks provide.

Users

To have the maximum presence, support and sharing in OSMNs requires online users to make themselves register as a fake or pretended identification with the connected online network, though few social networking sites (SNS) allow us to look through or access the public data without registering themselves. It is the end-user wish to share or upload his or her personal content (for example their birthday date, residing place, hobbies, etc.,) all this information combines to form an end user’s profile.

The OSMNs consist of ties between connected end-users. Few SNS permit end-users to connect any other random user without taking permission from the tie
recipient, while some sites obey a two-phase communication rule that creates a connection only when both end-users agree upon. Few social sites for example Facebook, have the networks which have direct connections (means a connection from A to B does not interpret the presence of a vice-verse connection), whereas others sites like Orkut, having the networks of undirected connections.

End-users connect to other online users for numerous reasons. The target of a link may be a real-world acquaintance, a business contact, a virtual acquaintance, someone who shares the same interests, someone who uploads interesting content, and so on. In fact, some users even consider the acquisition of many links to be a goal in itself [31]. When compared to links on the Web, links in online social networks combine the functionality of both hyperlinks and bookmarks.

A user’s links, along with her profile, are usually visible to those who visit the user’s account. Thus, users are able to navigate the online social media network (OSMN) by following end-user-to-end-user relationships, surfing the end user’s personal profile and any shareable or uploaded content of visited end-users as they go. Few sites such as LinkedIn, grant surfing of end-users profiles within the user’s own community, while some sites, like Flickr, permit end-users to open and view any other user’s profile in the network.

Groups

Many social networking sites, like Usenet newsgroups [41], which allow end-users to build exclusive interest communities. All the online users not only can broadcast their messages to their linked groups (visible to all group members) but also can upload shared content to the social network group. Many online groups
have administrator (a single person) for controlling and maintaining the group, while few groups are accessible for any end-user or member to join. At present, all the social networking sites (SNS) need specific group acknowledgements by the end-users; end-users must manually constitute groups, assign administrators (if required), and acknowledge which groups they belong to. Certain sites (such as Facebook) create a few pre-populated groups based on the domain of users’ email addresses, but the majority of groups do not fall into this category.

The primary use of groups in today’s networks is to either express access control policies or to provide a forum for shared content. Examples of the former include sites like Facebook, which, by default, allows only users located in the same geographic location or organization to view each other’s profiles. Examples of the latter are more common, including Flickr’s shared photo groups and Orkut’s communities feature.

**Content**

Once an identity is created, the end-users can upload and share their content onto the social networking sites (SNS) in their respective accounts. Many of the social networking sites (SNS) allow the end-users, either make their content public (accessible to everyone) or make it personal (only visible by their peer group or immediate “friends”). Most of the SNS, for example, YouTube and Facebook, permit end-users to share online an immense or indefinite quantity of content, while few SNS like Flickr, where users have to minimal subscription fee to make his content shareable with the limit constraints. Almost all the uploaded content by the end user is catalogued or indexed at the end user’s personal account, or the profile for the other online users to access and explore fresh content through the
social networking site. By using textual search all the referenced or the indexed content can be made publicly accessible to other online users. For example, the Facebook picture or image search which permits online users to find pictures or photos by seeking through the comments and tag.

Once on the site, users can submit their uploaded content into groups that they are a member of. The privacy settings often allow for the content to be accessible only by group members. Moreover, the sites generally allow users to browse the content uploaded to the groups they are members of.

Users are also often allowed to create favourites lists, which link to a user’s favourite content uploaded by other users. These favourite lists are also generally publicly accessible from the user’s profile page. Similarly, most sites allow users to comment on pieces of content, much like an Usenet posting, and the comments appear alongside the piece of content itself.

Finally, many sites contain most popular content lists, which contain the most popular content items (in terms of the number of views, comments, or ratings) that have been recently uploaded. Users can browse these lists to find new content to view. A notable example is Facebook’s top-100 lists, where popularity is based on the number of views, comments, or favourite-markings a video has recently received.
2.2 Significance of study Online Social Media Networks (OSMN)

Online social networking is still very much in its infancy, yet it has already formed the basis for some enormously popular applications. As this paradigm matures, the study expects more sophisticated applications to naturally emerge. It is not inconceivable that social networking systems will eventually become de-facto portals for both personal and commercial online interactions. Below, the study outlines a few of the many potential applications that could benefit from understanding the structure of and information flow in these networks. Additionally, the study speculates on how the data collected in this thesis could be relevant to researchers in other disciplines.

2.2.1 Trust

Nearby nodes in a social graph tend to rely upon each other more than random pairs of users in the network. A number of research systems have already been proposed to exploit this trust. Trust relationships are being used in the PGP web of trust [42] to eliminate the need for a trusted certificate authority. SybilGuard [43] and SybilLimit [44] uses the social network to mitigate Sybil [45] attacks on distributed systems, exploiting the fact that real people tend to have a diverse set of social relations. RE [46] determines the social network distance between the sender and the receiver of an email to aid SPAM detection. The study admits that a profound understanding of the basic or fundamental topology is a crucial beginner’s level step, not only in the designing but also in exploring trust and other prominence metrics for these systems.
2.2.2 Mutual Interest

Neighbouring end-users with any social media network share common interests. The others tried to explore adjacent regions of their social network because there is a high probability of finding the content that is of their interest. Online media systems such as PeerSpective [47], Yahoo My Web [48] and Google Coop [49] use social networks to rank online search outcomes relative to the interests of the end-user. By using both one, viewed uploaded matter and second, clicked on online search options by the connected members of the social media network, these online media systems will rank accordingly the outcomes of the end-users perspective explorations more efficiently and accurately.

To understand the framework or structure of OSMNs as well as the correct definition that will define them properly, the study needs to understand various efficient algorithms to fetch the user’s reliability in terms of relationships with other users and also find the degree of mutual interest between two online connected users. It is also crucial to understand the strong physical make-up of the social media networks to sustain the deliberated attempts of modification. These are not the concern areas of this thesis, but the bottom line is to understand the basic fundamental of the OSMNs which is the foremost requirement of this thesis.

2.2.3 Content Exchange

The phenomenal popularity of social networking sites like Facebook, YouTube, Flickr, and MySpace represents a shift in how content is published, located, and distributed on the Internet. Understanding how content diffuses through these networks and becomes popular over time is not only of academic interest, but is
increasingly important in commercial advertising, in political campaigning, and ultimately to society. In fact, a number of research efforts [50, 51, 52, 53, 54 and 55] have proposed viral marketing campaigns to leverage the word-of-mouth effect. In 2007 alone, $1.2 billion was spent on advertising in online social networks [56] worldwide, and this is expected to triple by 2011.

Understanding how information flows among users of online communities is an important step toward the design and analysis of future information dissemination systems. Understanding how information flows in online social networks can also aid designers of current social networking systems. If, for example, one can predict the relative popularity of newly introduced objects, caching and pre-fetching schemes can be created to reduce the latency and bandwidth required by the site. Since many of the currently popular sites rely primarily on advertising for revenue, reducing distribution costs for multimedia content is clearly a pressing issue.

Understanding how content flows through social networks also has the potential to improve search algorithms. By examining the content that users view or mark as a favourite, sites may be able to suggest other content that may be of interest to the user. Many have noted [57] that the age of the Internet has enabled much greater diversity in preferences and tastes; using online social networks appears to be a natural approach to further discover and refine tastes.

Finally, understanding how content is exchanged in online social networks can help guide the designers of future systems. Social networks have already proven to be useful in a number of different contexts, and we are seeing new sites popping using social networks to predict music preferences, find potential job applications, and share content. By understanding the user structure and the
properties of information flow, designers of future systems have an empirical basis for designing and provisioning their systems.

2.2.4 Other Disciplines

As mentioned before, our work has relevance beyond computer systems. To sociologists, online social networks offer an unprecedented amount of data. These systems represent the complete evolution of a large, contained online social network, with the accompanying timeline of every event that occurred within them. Sociologists can examine this data not only to validate existing theories of communication but also search for fresh and innovative forms of communication.

To political scientists and marketing specialists, studying how information flows through social networks may help improve techniques such as targeted advertising and viral marketing. Political candidates have already realized the importance of blogs in recent elections [58]. Similarly, marketing specialists are already experimenting with paid viral marketing to better promote products and companies. Clearly, a better understanding of how content is currently being exchanged in these systems holds the potential to improve these approaches.

2.3 Related Work

In this section, the study describes prior work related to the topics presented in this thesis. As this thesis covers a number of different topics, the related work has been grouped into sections detailing (a) work that examines the static structural
properties of complex networks, (b) work that examines how complex networks evolve, (c) work that identifies and uses communities in online social networks, (d) work that tries to explore of the communication between online groups and communities, and (e) work that tries to search communities in a personalize web.

2.3.1 Complex Network Structure

Our study begins by examining the work that characterizes the structure of static snapshots of large scale networks. In following chapters, the study examines the static snapshots of multiple online social networks. In order to ground our analysis, the study compares our results to those from other large-scale complex networks such as the Web and the Internet. Thus, we describe related work that studies these networks after describing work that studies social networks.

Social Networks

Sociologists have studied many of the properties of off-line social networks, and also our study only briefly describes a few of the relevant findings. The book as written by Wasserman [59] which shows the overall view of the structure of the off-line social networks and various analysis techniques associated with them. Milgram [60] demonstrated that the average distance between two internet users was six hops or nodes, demonstrating that social networks can be classified as small-world. Pool and Kochen [61] provided an analysis of how the small-world property of social networks affects contacts and influence. The influential paper by Granovetter [62] puts a valid point that a social media network can be partitioned into ‘strong’ (which are strongly coupled or highly clustered) and
‘weak’ links (which represent long-distance relationships). The studies were able to verify that online social networks have similar properties, with short path lengths and strong clusters connected by long distance links.

As online social networks gained popularity, researchers have begun to investigate their properties. At Stanford University, Adamic et al. [63] studied and explored an early OSMN, and found that the network has small-world characteristics as well as a significant clustering coefficient. Liben-Nowell et al. [64] found a strong correlation between friendship and the geographic location of users by working on the data provided by LiveJournal. Kumar et al. [65] examined two online social networks from Yahoo! And found that both possessed a dominant SCC. Girvan and Newman observed that the online users in the OSMN have the tendency to create highly coupled groups [66], evidenced by a high clustering coefficient. Our study will be exploring all of these properties from the Facebook higher education groups during our research study.

Finally, researchers have also examined how the activity network, or the pattern of interactions between users, compares with the social network. In particular, Wilson et al. [67] studied the activity network of samples of the Facebook network and found that in contrast to the social network; the activity network is much sparser and has a significantly lower maximal degree. Chun et al. [68] found similar properties in the for the Korean social interaction network CyWorld [69]. In our work, the study focuses only on the social network, but our approach and methods could be naturally applied to the real-world social network as well.

Other Information Networks
Many research experiments were done to explore the framework or structure of information networks like the graph of Web pages and the Internet’s routing topology. A prominent study of Web structure [70] showed that the Web has a “bow-tie” shape, consisting of a dominant Strong Connected Component (SCC), and clusters of vertices or nodes that can either reach to SCC or can be fetched from the SCC. Our study has shown the SCC in our research on online social media networks. Faloutsos et al. [71] found that the degree distribution (DD) of the Internet’s routing topology obeys a power-law rule whereas, Siganos et al. [72] showed the Internet’s high-level structure resembles like a “jellyfish”.

Kleinberg [73] showed that high-degree nodes can be observed as either hub (pages containing valuable citations on a subject) or powerful nodes in the control (pages having appropriate information on a subject). He also presented an algorithm [74], which, when given a graph on Web pages, can infer pages function as hubs and as authorities. In [75], the most popular PageRank algorithm was discussed which uses the internet web structure to find pages which are considered reputable. Those online social networks, which have a high degree of linkage symmetry, may prevent such algorithms from working since the hubs are automatically also the authorities.

2.3.2 Complex Network Growth

In addition to the study of the static structure of various information networks, researchers have also examined the evolution of networks, looking at the processes by which links are formed and removed. Consistent with previous work, our study refers to these processes as growth models. In our work, the study collects detailed data on the growth of online social networks. Thus, in this
section, the study describes related work on various growth models and detail the extent to which they have been validated on real data.

**Growth Models**

Growth models for complex networks can be partitioned into structural models (i.e., models that only take into account the structure of the network to predict link formation or removal) and explanatory models (i.e., models that consider external factors, such as human factors in online social networks, to predict links). The study describes each of these types of models below.

**Structural Growth Models**

Many researchers wanted to analyse the interesting high-level structural properties across different online social media networks by hypothesizing that the networks are the result of a few common structural growth processes at work. Many models of these processes have been proposed and analysed to explain the structure of complex networks.

The well-known Barabási-Albert (BA) model [76], based on preferential attachment, has shown power-law degree distributions in the OSMN. In the BA model, by using a probability distribution calculated by node degree, it is easier to add or attach new links to the respective nodes, resulting in linear preferential attachment. Many new innovations have been proposed to the BA model (i.e., to add a tunable level of clustering [77]). The study is able to verify that the growth
of online social networks follows linear preferential attachment, but not in the way that the BA model proposes.

Another set of models that generate power-law networks are based on local rules, such as firstly, Random Walk model [78 & 79], in which nodes picks new neighbours by taking random walks; secondly, Common Neighbours model, in which nodes picks new neighbours by short-listing those nodes with whom they share many friends in common; and lastly, finite memory model [80], in which nodes eventually become inactive and stop receiving any new links. All of these models not only exhibit preferential attachment which means the degree of selecting high-degree nodes are mostly, but also shows a higher level of local clustering than the BA model [79]. The study demonstrates that, while these models are more accurate at predicting the destination of new links in our data than the BA model, the overall accuracy of these models remains very low. For a more detailed treatment all of these models and others, the study refers the reader to a paper by Mitzenmacher [81].

**Explanatory Growth Models**

Some recent studies, particularly on online social networks, have proposed explanatory models of the network growth. Unlike structural growth models, which try to model growth solely as a function of the network structure, explanatory models seek to account for the underlying sociological factors that cause the links to be established. For example, an explanatory growth model for Flickr, a photo-sharing social network, could be based on an understanding of how users behave when sharing pictures.
Examples of work on explanatory growth models include Kumar [65], who divided users into ones who are active and passive, and presented a model describing their behaviour in an online social network. Jin et al. [82] presented a model of social networks based on known human interactions. Backstrom et al. [83] looked at snapshots of group membership in LiveJournal and presented a framework for the growth of end-user in groups over time based on understandings of peer pressure. Finally, Chang et al. [84] proposed a model for the growth in connectivity of the Internet topology, modelling the decision processes of the administrators of autonomous systems.

Compared to structural growth models, explanatory models are more detailed, but they also tend to be specific to the network being investigated. For example, the reasons why autonomous systems connect to each other in the Internet topology are very different from the reasons why users in Facebook connect to each other. By being agnostic to these factors, structural growth models are inherently less accurate. But, they are far more general, and can be compared across different types of networks. In this thesis, the study focuses only on structural growth models.

**Validation of Growth Models**

It is important to note here that both structural and explanatory growth models are, by and large, intuitive models that can explain the observed structural properties of the networks. But, they have not been significantly validated using empirical data. Mitzenmacher [85] poses this as one of the biggest challenges facing the future of power-law research. One of the contributions of this thesis lies in collecting data that can be used to determine how well these processes, predict what actually occurs in different real-world networks at scale.
Observations of Network Growth

With the growth in popularity of online social networks, a few studies have examined the properties of the networks over time. We briefly describe these studies below.

A few studies have looked at how links are formed in social networks. Kossinets and Watts [86] demonstrated that fresh or new ties in the social media network were established between the nodes which were lying nearer to each other. In physics, Nowell et al. [87] experimented networks having co-authorship, to predict future collaborations by analysing different graph proximity metrics. Newman [88] and Jeong et al. [89] examined the properties of scientific collaboration networks and found evidence of preferential attachment. Peltom¨aki and Alava [90] worked on a movie actor network and scientific collaboration network to find the strongest evidence of sub-linear preferential attachment.

Our research work shares similar goals and methodology as the above studies and the fetched dataset allowed us to analyse social media higher education group network properties and its growth. The study analyses daily snapshots of Facebook ego-networks, and weekly snapshots of the Facebook Higher Education Groups. From Wikipedia, we have sufficient data to create a snapshot of the network at the precise second a new link is established. Since the growth models rely solely on the current network structure to predict new link formation, having frequent snapshots of the network is crucial for validating the models with high accuracy.
Researchers have also studied the high-level properties of graph evolution, looking for evolution trends at the global level. For instance, Leskovec et al. [91] examined the evolution of a number of real-world graphs, including collaboration networks and recommendation networks. They found that the graphs tend to denser and that the average path length tends to shrink (instead of growing in proportion to the number of nodes). Additionally, Kumar et al. [92] observed the early evolution of the blogosphere, and found that it is rapidly increasing in both scale and connectedness. This line of work is largely complementary to our work, as the study focuses on the local link formation phenomena which might lead to these global observations.

2.3.3 Detecting Communities

The study now turns our attention to the detection of communities in online social networks. A community is a group of the online users in a social media network that is more tightly interconnected (highly coupled) than the overall network [93]. Thus, all the work described in this section tries to detect densely connected components of graphs. At a high level, the approaches can be divided into global or local approaches, the first one which speculates knowledge of the entire graph and later, which only speculate detailed knowledge of a specific region of the network, respectively. After briefly describing how communities were detected classically in sociology, the study describes the global and local approaches. Then, we describe empirical studies of social networks that have looked for the presence of Communities.
Classical Community Detection

Classical community detection in sociology took the approach of partitioning the vertices in a social media network into various tightly coupled communities by reducing the number of links between communities. Within this approach, there are two main algorithms: Label Propagation Algorithm (LPA) [94] and the Fast Network Community Algorithm (FCNA) [95]. Both the algorithms divide the graph into the best two communities possible, and then further subdivide those two until reaching the user-specified number of communities. However, both algorithms require the user to specify the sizes of the two communities initially and the final number of communities desired. There are other algorithms that work on the same notion is spectral bisection [96] and the Kernighan-Lin algorithm [97].

Global Community Detection

One of the first community detection algorithms that did not assume pre-existing knowledge of the community structure was proposed by Girvan and Newman [98]. In short, their algorithm works by calculating the “most important” link in the network and then removing it. The algorithm then repeats this step until the each and every partition in a social network graph behaves as a community. Continuing to run the algorithm over the various partitions will produce even finer communities until all links are removed from the network.

From the above description, it is clear that the selection of the most important link is integral to the functioning of the algorithm. A good metric of importance can quickly partition the graph into its various communities while a bad metric can
simply disconnect nodes one-by-one and produce degenerate partitions. Girvan and Newman suggested using the metric of betweenness centrality. The intuition behind Girvan and Newman’s algorithm is simple: if the study assumes that the social network is divided into densely connected communities the betweenness centrality metric looks for links that bridge communities. Since communities are, by definition, denser than the graph as a whole, these bridging links will naturally have a higher betweenness centrality. Once they are removed from the graph, the underlying community structure emerges.

Newman later proposed a faster, alternate approach, based on the greedy optimization of modularity. The algorithm initiates with each node in a separate community, and finally fuses all pairs of communities, selecting at each phase the pair that would generate the highest increase or smallest decrease in modularity. Clauset et al. [99] proposed an innovative and a different, faster variant of this algorithm by further improving and optimizing the different operations with the use of more efficient data structures. These improvements in speed are important, as the running time of the original algorithm prohibited it from being used on graphs with more than a few thousand links.

Tyler et al. [100] presented a variant of the algorithm by Girvan and Newman, which improved the speed of the algorithm at the cost of accuracy. Instead of calculating the total betweenness centrality score by considering all paths starting at every vertex in the graph, Tyler et al. Suggest that the betweenness centrality is calculated by summing over only a subset of the vertices, thereby obtaining a partial betweenness centrality score for all other edges. The same algorithm can execute multiple times, producing many community partitions and are then combined into one wholesome community partition by using the technique which was proposed by Wilkinson et al. [101].
Another algorithm was proposed by Radicchi et al. [102] which was based on the similar approach of Girvan and Newman. This approach uses a concept of local approximation to identify the edges to be shunned out, which can be estimated quickly and, therefore, performs better and execute quickly. For every respective edge, it approximates the betweenness centrality by the number of loops of length three (i.e., triangles) that include the edge. Inter-community edges are unlikely to belong to many triangles because they require another edge between the communities to complete the loop.

Other approaches have looked at finding multiple, overlapping community structures from a global perspective. This is in contrast to the previously discussed approaches, which were only concerned with finding the best way to partition the nodes into single, non-overlapping set of communities. The overlapping approaches include work by Palla et al. [103], which used k-cliques to find overlapping communities at different scales. Baumes et al. [104 & 105] proposed a similar approach for finding overlapping communities by first looking for dense collections of nodes in the graph. Du et al. [106] had proposed an algorithm to find communities in wide-scale social networks by taking into account the concept of overlapped communities. Finally, Li et al. [107] proposed a separate approach for overlapping community detection based on triangle formation and clustering based on text similarity.

**Local Community Detection**

One of the potential disadvantages of the global approaches used in community identification is that the whole structure of the network must be known a priori, that is why these approaches are expensive, as real-world networks are large and
complicated. For that reason, only many researchers looked upon local approaches to detect communities, which use local domain knowledge to construct a community around a group of nodes. Moreover, local approaches have more applicable and scalable to much complex and larger graphs, as well as graphs which are not completely visible due to privacy restrictions. Moreover, local approaches to community detection also hold the potential to detect multiple community structures – global approaches assign each node to exactly one community, even if multiple such structures exist. Finally, local approaches allow for natural decentralization, as the computation can be trivially divided up and distributed. Clauset [108] proposed one of the first local approaches to community detection, which was based on the greedy construction of a community around a source node. The algorithm creates a community by adding vertices one-by-one, choosing the vertex at each step that maximizes the ratio of intra-community links or connections to inter-community links or connections for the (vertices) nodes on the “fringe” of the community. Thus, this algorithm tries to create a strong community by greedily picking nodes that have many links inside the community.

Barrow et al. [109] proposed an alternative algorithm, which adds all the $k$-hop vertices at each step until the ratio of inter-community to intra-community links falls below a threshold. Both of these were shown to detect communities in synthetic graph and a real-world product recommendation network. Recently, Wakita et al. [110] proposed a modification to the Clauset algorithm, which has the potential to identify communities in social media networks with up too few million end-users. Although, his work did not provide any strong foundation for community structure implied from the social media network.

Additionally, two new local community detection algorithms have been proposed to improve the speed and performance of community detection. Luo et al. [111]
recommended an algorithm which was similar to Clauset’s, with the exception that it iteratively adds and removes nodes, continuing until adding or removing any single vertex would not result in a better community. Barrow [112] evaluated the performance of the various algorithms (and one additional newly proposed one), and found that the algorithm of Leo et al. Performed the best on synthetically generated graphs.

2.4 Scope of Our Research Work

As the title suggests that the work in this research focuses on the key aspect that is analysing social media networks for the betterment of higher education community (HE) or Higher Education Institutions (HEIs). The research focused on extracting and fetching real data sets by using software application tools for the social media mining. The greatest challenge is the software tools which are going to be used for Social Media (SM) as well as on the user generated content which is changing day by day, then how can we find, access and share information. At the same time, the easy availability of large, social data makes it possible for the computer and social scientists to pose new and innovative research questions for HE on the scale that were never done before. This, however, requires us to think the approaches to algorithmic approaches to analyse the SM data.

As far as the current literature survey is concerned, some research has been done on the social network analysis, but for the specific characteristics like academic performance, student behaviour or retention, etc. But nobody has done the research to establish effectiveness thresholds in the use of these techniques to
obtain more and better outcomes in the application of social network analysis (SNA) for the recommendations in academia using social media. The research will eventually represent an additional element to give more valuable recommendations or the predictions about the online social media networks (OSMN).

This research does not aim to replace the traditional method of lectures at the university. However, it is more towards promoting dynamic study culture of the campus. By this study, it will be shown how social network data analysis discovers the patterns of students’ participation in social networking and relating their participation patterns with their personal behaviour. The study will also try to implement the system where data matching can be manipulated to predict the decision regards the best knowledge practice.