POLITECNICO DI MILANO
DEPARTMENT OF ELECTRONICS, INFORMATION AND BIO-ENGINEERING
DOCTORAL PROGRAMME IN INFORMATION TECHNOLOGY

A VISUAL FRAMEWORK FOR THE EMPIRICAL ANALYSIS OF SOCIAL INFLUENCERS AND INFLUENCE

Doctoral Dissertation of:
Ajaz Hussain

Advisor:
Prof. Chiara Francalanci

Tutor:
Prof. Barbara Pernici

The Chair of the Doctoral Program:
Prof. Carlo Fiorini

2015 – XVIII

Acknowledgments

First and foremost, I am immensely thankful to Almighty God for letting me pursue and fulfill my dreams. Nothing could have been possible without His blessings.

I would like to express my special appreciation and thanks to my advisor - Prof. CHIARA FRANCALANCI, you have been a tremendous mentor for me by encouraging my research and for allowing me to grow as a research scientist. Your advice on both research and as well as on my professional career have been priceless. I really appreciate the time and effort of thesis external reviewer - Prof. RODOLFO BAGGIO for your valuable feedback and helpful suggestions. I am also grateful to my tutor - Prof. BARBARA PERNICI, for your kind guidance during course work and valuable assistance during Ph.D. phase.

I am highly grateful to Erasmus Mundus STRoNGTiES Committee for providing me an excellent mobility opportunity that had made a positive impact on my life. I am also grateful to my colleagues and seniors in *Information System Group* (ISGroup), who helped me during my research phase through consistent guidance. I am highly grateful to the NavigTweet Pilot committee members from DEIB for your valuable comments and suggestion.

I am highly thankful to my family, parents, and in-laws for your kind support and prayers with all of the sacrifices that you've made on my behalf. Last, but not least, I truly appreciate my beloved wife RABIA AJAZ, for your kind love, patience, understanding, consistent encouragement and motivation throughout my life. I would like to dedicate my thesis to my family, my wife and my son - MUHAMMAD ARSAL HASHMI.

Ajaz Hussain

Abstract

SOCIAL media have become pervasive and ubiquitous and represent a source of valuable information. The literature on social media makes a distinction between influencers and influence. The former are social media users with a broad audience. For example, influencers can have a high number of followers on Twitter, or a multitude of friends on Facebook, or a broad array of connections on LinkedIn. The term influence is instead used to refer to the social impact of the content shared by social media users.

The majority of these studies has focused on the role of influencers. Our claim is that while the information shared by influencers has a broader reach, the content of messages plays a critical role and can be a determinant of the social influence of the message irrespective of the centrality of the message's author. This thesis starts from the observation that social networks of influence follow a power-law distribution function, with a few hub nodes and a long tail of peripheral nodes, consistent with the so-called small-world phenomenon. In social media, hub nodes represent social influencers , but influential content can be generated by peripheral nodes and spread along possibly multi-hop paths originated in peripheral network layers.

This thesis provides a conceptual framework and related software tool to assess influence and identification of influencers. The assessment of influence and influencers is performed in two steps. First, an empirical analysis is conducted in order to verify the assumption that content can have an impact on influence. We propose a visual approach to the graphical representation and exploration of peripheral layers and clusters by exploiting the theory of k-shell decomposition analysis and power-law based modified force-directed method to clearly display local multi-layered neighborhood clusters around hub nodes. We put forward few hypotheses that tie specificity, frequency of tweets and frequency of retweets and are tested on data samples of roughly one million tweets. Overall, results highlight the effectiveness of our approach, providing interesting visual insights on how unveiling the structure of the periphery of the network can visually show the potential of peripheral nodes in determining influence and content relationship. Secondly, this thesis aims to provide a novel visual framework to analyze, explore and interact with Twitter 'Who Follows Who' relationships, by visually browsing the friends' network to identify the key influencers based upon the actual influence of the content they share. As part of this research, we have developed NavigTweet, a novel visualization tool for the influence-based exploration of Twitter network. The core concept of the proposed approach is to identify influencers by browsing through a user's friends' network. Then, a power-law based modified force-directed method is applied to clearly display the network graph in a multi-layered and multi-clustered way. To gather some insight into the user experience with the pilot release of NavigTweet, we have conducted a qualitative pilot user study. We report on the study and its results, with initial pilot release.

Riassunto

I SOCIAL MEDIA sono diventati uno strumento pervasivo e ampiamente diffuso, rappresentando così una fonte di informazioni preziose. La letteratura sui social media evidenzia una distinzione tra influencer e influence. I primi sono utenti dei social media con un vasto pubblico. Ad esempio, gli influencer possono avere un alto numero di follower su Twitter, un elevato numero di amici su Facebook, o una vasta gamma di connessioni su LinkedIn. Il termine influence viene invece usato per indicare l'impatto sociale dei contenuti condivisi dagli utenti dei social media.

La maggior parte di questi studi si è concentrata sul ruolo degli influenzatori, le cui informazioni condivise hanno una portata molto ampia. La nostra tesi invece, si concentra sul contenuto dei messaggi, che gioca un ruolo critico e può essere un fattore determinante dell'influenza sociale del messaggio indipendentemente dalla centralità dell'autore. Questa tesi parte dalla constatazione che le reti sociali di influenza seguono una funzione di distribuzione power-law, con pochi nodi hub e con una lunga coda di nodi periferici, coerenti con il cosiddetto fenomeno small-world. Nel contesto dei social media, i nodi hub rappresentano influenzatori sociali, tuttavia il contenuto influente può essere generato da nodi periferici e diffondersi coì lungo possibili percorsi multi-hop nati in livelli della rete periferica.

Questa tesi fornisce un quadro concettuale e un relativo strumento software al fine di valutare l'influenza e di identificare gli influenzatori. La valutazione di influenza e influenzatori viene eseguita in due fasi. In primo luogo, viene condotta un'analisi empirica per verificare l'ipotesi che il contenuto possa avere un impatto sull'influence. Proponiamo dunque un approccio visivo per la rappresen-

tazione grafica e l'esplorazione di layer periferici e cluster, sfruttando la teoria dell'analisi k-shell decomposition, mentre per quanto riguarda la visualizzazione di local multi-layered neighbourhood cluster intorno ai hub-nodes viene applicato il metodo force-directed modificato e basato sulla distribuzione power-law. Vengono inoltre presentate alcune ipotesi che legano la specificità, la frequenza di tweet e la frequenza di retweet, testate su campioni di dati di circa un milione di tweet. Nel complesso, i risultati evidenziano l'efficacia del nostro approccio, fornendo interessanti spunti visivi su come comprendere la struttura della periferia della rete, mostrando il potenziale dei nodi periferici nel determinare l'inflence e il contenuto relazionale.

In secondo luogo, questa tesi si propone di fornire un innovativo quadro visivo con lo scopo di analizzare, esplorare ed interagire con relazioni di Twitter di tipo 'Who Follows Who', navingando visivamente la rete di amici, per identificare gli influencer chiave basati sulla influence effettiva del contenuto che condividono. Come parte di questa ricerca, abbiamo sviluppato NavigTweet, uno nuovo strumento di visualizzazione per l'esplorazione dell'influence-based dei network di Twitter. Il concetto di base del metodo proposto è quello di identificare gli influencer navigando attraverso la rete di amici di un utente. Successivamete, viene applicato un metodo di force-directed modificato e basato sulla distribuzione power-law, con lo scopo di visualizzare in modo chiaro il grafico di rete tramite un approccio multi-layer e multi-cluster. A fine di ottenere conoscenza dall'esperienza degli utenti con il rilascio pilota NavigTweet, abbiamo condotto uno studio pilota qualitativo dell'utente. Diamo un report sullo studio e sui suoi risultati insieme al rilascio caso pilota iniziale.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

Social media have become pervasive and ubiquitous and represent a source of valuable information. There is a growing need for information visualization, which has recently become a popular subject of research [56, 92, 119]. In general, information visualization aims at showing information in an easy, user-friendly and graphical way. However visualizing information properly is not trivial and becomes a challenge when the focus is large social networks, such as Twitter. Twitter has been defined as the key role player of the change on how information dissemination is accomplished [4, 12, 44]. Its influence on information dissemination has led to research exploring on how this is achieved. According to [96] the unicity of direction in twitter connection provides the key driver of information dissemination via word of mouth (WoM) in retweets.

The literature on social media makes a distinction between influencers and influence. The former are social media users with a broad audience. For example, influencers can have a high number of followers on Twitter, or a multitude of friends on *Facebook*, or a broad array of connections on *LinkedIn*. The term influence is instead used to refer to the social impact of the content shared by social media users. The breadth of the audience was considered the first and foremost indicator of influence for traditional media, such as television or radio. However, traditional media are based on broadcasting rather than communication, while social media are truly interactive. It is very common that influencers say something totally uninteresting and, as a consequence, they obtain little or no attention. On the contrary, if social media users are interested in something, they typically show it by participating in the conversation with a variety of mechanisms and, most commonly, by sharing the content that they have liked. [30, 119] has noted that a content that has had an impact on a user's mind is shared. Influencers are prominent social media users, but we cannot expect that the content that they share is bound to have high influence, as discussed by [21, 114].

Previous research [17, 33] has shown how the content of messages can play a critical role and can be a determinant of the social influence of a message irrespective of the centrality of the message's author. Results suggest that peripheral nodes can be influential. This thesis starts from the observation made by [43] that social networks of influence follow a power-law distribution function [13], with a few hub nodes and a long tail of peripheral nodes, consistent with the so-called small-world phenomenon as noted by [159]. In social media, hub nodes represent social influencers [132], but influential content can be generated by peripheral nodes and spread along possibly multi-hop paths originated in peripheral network layers.

In this thesis we investigate the relation between 'content' and dynamics of social 'influence', by dealing with issues connected with influence and influencers on social media. The thesis considers information shared by "influencers" and seeks to show that the role played by the content of the message is higher than the centrality of authors for the spread and the reach of the message. The specific research questions which we concern about are stated as under:

- Is content shared by influencers is bound to have high influence?
- Is content a driver of social media influence?
- What are the characteristics of content that help to increase social media influence?

The research aims to understand how influential content spreads across the network. For this purpose, identifying and positioning hub nodes is not sufficient, while we need an approach that supports the exploration of peripheral nodes and of their mutual connections. Our claim is that Influential content can be generated by peripheral nodes and spreads along possibly multi-hop paths originated in peripheral network layers. We provide a conceptual visual framework and related software tool to the assessment of influence and identification of influencers. Considering the dynamics of influence and sparse complex nature of social networks, we need a visual approach in order to understand the connections and spread of the social networks through a visual exploration framework. Such disseminated information over widely-spread social networks, specifically Twitter, led us to adopt the visual platform, in order to better understand the network flow and spread of the influence. The framework of networks can be very useful when thinking about social dynamics, in this way, the people are represented as the nodes and their interactions are the connections (edges) between them. Using this model we demonstrate the influence spread and how someone can become influential in a social context by investigating characteristics of shared content that are an outcome of behavioral decisions made by social media users.

In this thesis, we exploit a modified power-law based force-directed algorithm [84] to highlight the local multi-layered neighborhood clusters around hub nodes. The algorithm is based on the idea that hub nodes should be prioritized in laying out the overall network topology, but their placement should depend on the topology of peripheral nodes around them. In our approach, the topology of periphery is defined by grouping peripheral nodes based on the strength of their link to hub nodes, as well as the strength of their mutual interconnections, which is metaphor of k-shell decomposition analysis [39, 91].

A growing research stream has recognized that destinations are complex dynamic systems whose characteristics need to be analyzed and understood in order to better address planning and management actions [14,102,138]. The destination is the fundamental unit of analysis in any modeling of a tourism system [11, 94] used for understanding the tourism phenomenon. In spite of the importance of destination as unit for study and management tourism, tourism research has given little attention to the concept of destination from tourist perspective [102, 126]. The tourist choice defined the role of a destination as "central" or "peripheral" within a network [11]. Tourists build their own networks around nodal destinations, even if they are geographically distant. Thus, tourist mobility affects the shape, the dimension, and the structure of the networks, where tourists are different for characteristics, trip-related behaviors, and type of holiday chosen. The network science approach has uncovered important outcomes concerning destinations' structures, the functioning of collaborative and cooperative groups, the diffusion of information or knowledge across the system or the relationships between the physical and the virtual components of a destination. The network perspective can offer a number of useful outcomes for tourism studies, but has also shown some limitations mainly due to the difficulty of collecting the data needed to perform a full analysis [102].

A part of our research focuses on the connections among Italian tourist destinations in order to understand complex dynamics of tourism and destination branding. We propose a visual approach to highlight the local multi-layered neighborhood clusters around hub nodes in aesthetic graph layout. The approach is tested on a large sample of tweets expressing opinions on a selection of Italian locations relevant to the tourism domain. Tweets have been semantically processed and tagged with information on a) the location (called brand) to which they refer, b) the number of retweets, c) the identifier of the retweeting author, and d) the topics (called category). With this information, we draw corresponding multi-mode networks highlighting the connections among authors (retweeting) and their interests (brand, and category) by aesthetically pleasant layouts. By visually exploring and understanding the multi-layered periphery of nodes in clusters, we also propose a few content-related hypotheses in order to understand network behavior and relationship among frequency, specificity, and retweets in tweets. In particular, we focus on three behavioral variables: *content specificity*, *frequency of sharing*, and *frequency of retweets*. The first variable represents the level of detail with which a user comments on a given subject of interest, while the second one represents the amount of contents shared by user in tweets and the third one is the frequency of retweets upon shared content. Insights on the relationship among *content specificity*, *frequency of sharing*, and *frequency of retweets* would help social media users to make their behavioral decisions. Fundamental goal of any social media user is to post content that is shared frequently, by many other users and over extended periods of time before fading [12, 56]. However, the literature does not provide systematic and visual evidence on how behavioral decisions regarding *content specificity*, *frequency of sharing*, and *frequency of retweets* exert an impact on influence. This thesis provides preliminary evidence from Twitter. We put forward three hypotheses that tie *specificity*, *frequency* and *frequency of retweets* and are tested on data samples of roughly one million tweets. Empirical and visual results show a significant relationship between influence and behavioral decisions on content. The relationship is found to be consistently significant across both data samples.

Based upon our proposed approach, visualizing the complex tourism destinations networks, and understanding the dynamics of tourism, reveals many aspects in order to better understand tourism phenoman. The empirical and visual evidence raises theoretical challenges and encourages further research to understand the relationship between content and influence on social media. The main innovative aspect of our approach is that we use statistics (hypotheses) and visualization together. One can visually verify the proposed hypotheses on graphs. Overall, results highlight the effectiveness of our approach, providing interesting visual insights on how unveiling the structure of the periphery of the network can visually show the potential of peripheral nodes in determining influence and content relationship. One can visually identify the actual and potential influencers by visualizing tourism network, and further practical implications to the study can be helpful in *co-branding*, *brand fidelity*, *brand promotion* strategies and provides a visual platform to explore the complex tourism network where people talk about different brands or tourism categories.

The ultimate goal of our research is to provide a novel visual framework to analyze, explore and interact with Twitter '*Who Follows Who*' relationships, by browsing the friends' network to identify the key influencers based on the actual influence of the content they share. As part of this research, we have developed *NavigTweet* [60, 61], a visual tool for the influence-based exploration of Twitter friends' network. It helps to identify the key players, and follow them directly through the NavigTweet. The user can explore its own Friend-of-a-Friend (FOAF) network in order to find interesting people to be followed. The top-influencers are identified by both user-level (e.g. number of followers, number of tweets, etc.) and content-based (number of hashtags, number of URLs, etc.) parameters, thoroughly described in Chapter 4. Based upon these parameters, the tool adopts the Analytical Hierarchy Process (AHP) technique, to rank Twitter users, as our NavigTweet user explores his/her FOAF network. The NavigTweet users can find influencers within their friends' network through a visual interface and iteratively explore FOAF network to find more influencers. To gather preliminary feedback on the NavigTweet user experience with a pilot release of NavigTweet, we have conducted a survey targeting a reference group of academic experts in the social media domain who have been asked to use the application in a real time environment. In order to visualize the twitter network in an aesthetically pleasant, multi-layered and multi-clustered graph layout, we exploit our modified power-law based force-directed graph drawing layout technique, as discussed in [58, 61].

The structure of the thesis is described in the following paragraphs.

Chapter 2 presents a critical review of the state of the art related to the main topics. The concept of social media and social networks is explained by discussing the main variables used to describe social influence and to identify social influencers. Previous research works addressing the relationship among the variables object of this study are discussed, in particular the relationship between influencers and influence. The dynamics of tourism industry are being discussed by relating with brand fidelity and perceptual mapping. A review of the existing literature on network visualization techniques and aesthetics by covering force-directed, clustering, and power-law based techniques is also discussed following the literature upon both semantic and social networks visualization tools. From the literature, a gap clearly emerges on the interconnection and the impact that content can have on the diffusion of the information on social media. The research presented in this thesis focuses on this gap.

Chapter 3 describes the empirical analysis of social influencers and influence by addressing the relationship among these three specific content-based variables: *content specificty*, *frequency of tweets*, and *frequency of retweets*. This chapter also takes a behavioral perspective by proposing a few content-based hypotheses and by investigating characteristics of shared content that are an outcome of behavioral decisions made by social media users. We put forward three hypotheses that tie *specificity*, *frequency*, and *retweets* which are tested on data samples of roughly one million tweets. Further, empirical results upon proposed research model along with statistical results and discussions is also provided.

Chapter 4 addresses the novel visual framework to analyze, explore and interact with Twitter '*Who Follows Who*' relationships, by browsing the friends' network to identify the key influencers based on the actual influence of the content they share. In this chapter, we describe the influence parameters provided by Twitter, which we considered to identify influencers within social network. Further, we discuss the user ranking methodology based on these parameters. We also discuss the *Analytical Hierarchy Processing* technique, which we used to implement our ranking mechanism. Finally, we discuss the overall ranking.

As a part of this research, we developed NavigTweet [60, 82] - a novel visualization tool for the influence-based exploration of Twitter network, discussed in Chapter 5. We discuss the application overview, objectives, architecture, and main building blocks. Finally, we discuss the implementation results and also provide a qualitative comparison of NavigTweet with existing applications.

Chapter 6 illustrates a detailed pilot test execution and results of NavigTweet. To assess the user experience with NavigTweet, we have conducted a qualitative pilot study. The comprehensive evaluation and results of a pilot test and subsequent large-scale test of NavigTweet is discussed in this chapter. At the end, we report a web-analytics to analyze the traffic behavior of NavigTweet Website [82].

Finally, Chapter 7 presents some concluding remarks and analysis of the main results of this thesis. This work presents some limitations, which leave the research field open for future work, as discussed contextually.

CHAPTER 2

State of the Art

THis chapter introduces the aspects and reviews upon the state of the art related to social media and social networks, influence and influencers analysis, and existing network visualization techniques and tools. Section 2.1 introduces the concepts of social media and social networks. Section 2.3 discusses the concepts of social influence and influencer analysis by discussing the main variables used to describe social influence. Section 2.4 presents a review of the existing literature on network visualization techniques and aesthetics by covering force-directed, clustering, and power-law based techniques. Section 2.5 discusses the existing literature upon both semantic and social networks visualization tools. Finally, Section 2.6 summarizes the literature gap that is the focus of this thesis.

2.1 Social Media and Social Networks

Social media have a strong impact on the way users interact and share information. The process through which users create and share opinions on brands, products, and services, i.e., the electronic word-of-mouth (eWOM) is gaining increasing attention [47, 74, 105, 111]. In the online context, the eWOM takes place in a private context (i.e. one-to-one) as well as more complex interactions (i.e.

one-to-many). This reveals the power of eWOM. The reach of information shared through eWOM can be both broad and fast [125]. Organizations are now aware of the fact that the dynamics of information sharing are difficult to predict and there is a need for improving control. For that reason, there is a growing interest in understanding the complex structure of social networks, as this can affect the dynamics of user interactions and information sharing [26, 97, 105, 154].

In this context, the process through which users create and share opinions on brands, products, and services, i.e. the electronic word-of-mouth (eWOM) is gaining increasing attention [72, 79, 104]. In the online context, the eWOM can give use of one-to-many complex interactions. This represents the most powerful aspect of the eWOM. The reach of information sharing through eWOM can be both broad and fast [3, 17, 33]. Companies know that controlling the dynamics of information sharing is very difficult. This need for improving control is one of the reasons why there is a growing interest in understanding how the structure of a social network can affect the dynamics of user interaction and information sharing [97].

In early 1900s, Simmel proposed the first theory about social networks, introducing the concept of social phenomena [140]. Then, in 1934, a formal representation of social networks in terms of graphs containing nodes and edges was proposed by [118]. Harary and Cartwright [40, 77] introduced the concept of direct graph. With the introduction of directed arcs between nodes they were able to explain complex social patterns called *sociograms*. Further, Milgram introduced the concept of the six degrees of separation [116], in which he proposed the idea of what he called "small world phenomenon", particularly interesting in understanding the power of eWOM (electronic Word-Of-Mouth).

2.2 Network Analysis and Tourism Dynamics

A growing research stream has recognized that destinations are complex dynamic systems whose characteristics need to be analyzed and understood in order to better address planning and management actions [14, 102, 138]. The destination is the fundamental unit of analysis in any modeling of a tourism system [11, 94] used for understanding the tourism phenomenon. Destinations are closely associated with the concept of network, which can be used to describe the various relationships and transactions that characterize a tourism system. In spite of the importance of destination as unit for study and management tourism, tourism research has given little attention to the concept of destination from tourist perspective [102, 126]. The tourist choice defined the role of a destination as "central" or "peripheral" within a network [11]. Tourists build their own networks around nodal destinations, even if they are geographically distant. Thus, tourist mobility affects the shape, the dimension, and the structure of the networks, where tourists are different for characteristics, trip-related behaviors, and type of holiday chosen. The network science approach has uncovered important outcomes concerning destinations' structures, the functioning of collaborative and cooperative groups, the diffusion of information or knowledge across the system or the relationships between the physical and the virtual components of a destination. Additionally, the network approach has been extended to design simulation models with which different scenarios can be obtained in order to explore the possible effects of different managerial or governance actions. This provides all those interested in the life of a tourism destination with powerful tools to inform their policy or management strategies. The network perspective can offer a number of useful outcomes for tourism studies, but has also shown some limitations mainly due to the difficulty of collecting the data needed to perform a full analysis [102].

Based on this discussion, and according to the literature reviewed, this study focuses on the connections among Italian tourist destinations in order to understand complex dynamics of tourism and destination branding. We propose a visual approach to highlight the local multi-layered neighborhood clusters around hub nodes in aesthetic graph layout. We use Anholt's National Brand Index Model [9] to form the tourism network. The approach is tested on a large sample of tweets expressing opinions on a selection of Italian locations relevant to the tourism domain. Tweets have been semantically processed and tagged with information on a) the location (called brand) to which they refer, b) the number of retweets, c) the identifier of the retweeting author, and d) the topics (called category).

Understanding the complex dynamics of tourism network and destinations analysis, reveals further intrinsic qualities and complex role-relation hierarchy of connected destination networks. Visualizing such tightly coupled complex tourism network and understanding the destinations interactions can help us in devising a tourism promotional strategy. Tourism practitioners can devise a brand promotion or tourism marketing strategy through visual approach. Such visual approach may help them in analysis of brand fidelity, co-branding which is defined as the pairing of two or more constituent brands [1], brand promotion and tourism destination branding which is defined as "selecting a consistent element mix to identify and distinguish it through positive image building" [36].

2.3 Influence and Influencers Analysis

The literature on social media makes a distinction between influencers and influence. The former are social media users with a broad audience. For example, influencers can have a high number of followers on Twitter, or a multitude of friends on *Facebook*, or a broad array of connections on *LinkedIn*. The term influence is instead used to refer to the social impact of the content shared by social media users [37]. The breadth of the audience was considered the first and foremost indicator of influence for traditional media, such as television or radio. However, traditional media are based on broadcasting rather than communication, while social media are truly inter-active. It is very common that influencers say something totally uninteresting and, as a consequence, they obtain little or no attention. On the contrary, if social media users are interested in something, they typically show it by participating in the conversation with a variety of mechanisms and, most commonly, by sharing the content that they have liked. [30, 119] has noted that a content that has had an impact on a user's mind is shared. Influencers are prominent social media users, but we cannot expect that the content that they share is bound to have high influence, as discussed by [21, 59, 114].

In previous research, Bruni et al. [33] have shown how the content of messages can play a critical role and can be a determinant of the social influence of a message irrespective of the centrality of the message's author. Results suggest that peripheral nodes can be influential: this thesis starts from the observation made by [43] that social networks of influence follow a power-law distribution function [13], with a few hub nodes and a long tail of peripheral nodes, consistent with the so-called small-world phenomenon, as noted by [159]. In social media, hub nodes represent social influencers, but influential content can be generated by peripheral nodes and spread along possibly multi-hop paths originated in peripheral network layers.

Most network visualization methodologies and tools focus on identifying network hubs. Hubs represent central nodes connecting sets of more peripheral nodes that are rather sparse and separate from each other, as discussed by [159]. The literature has focused on measuring centrality and provides a broad array of centrality metrics, each of them highlighting a different aspect of a hub's prominent role. As discussed by [63], *degree centrality* measures the absolute number of connections of a node, *closeness centrality* measures how far a node is from all other nodes in the network along the overall shortest paths, while *betweenness centrality* assesses the role of a node as a hub of information by analyzing the extent to which the node connects separate subnetworks. These metrics represent the underlying principles of many network visualization tools. The assumption that most tools make to visualize large networks is that hubs represent the main driver of the structure of networks and, if they exist, they should be clearly highlighted to cope with complexity and obtain a nice and intuitive representation of the network.

While the literature provides consolidated approaches supporting the identification and characterization of hub nodes i.e. influencers in a social network, research on information spread, which is related to the multi-layered distribution of peripheral nodes, is limited. The literature mainly focuses on the concept of influencers, while there is a need for effective visualization techniques in social networks, which enable users to visually explore large-scale complex social networks to identify the users who are responsible for influence. The ultimate goal of our research is to understand how influential content spreads across the network. For this purpose, identifying and positioning hub nodes is not sufficient, while we need an approach that supports the exploration of peripheral nodes and of their mutual connections.

2.3.1 Structural Variables

Traditionally, the literature characterizes a social media user as an influencer on the basis of structural properties. *Centrality metrics* are the most widely considered parameters for the structural evaluation of a user's social network. The centrality of a concept has been defined as the significance of an individual within a network [56]. Centrality has attracted a considerable attention as it clearly recalls concepts like social power, influence, and reputation. A node that is directly connected to a high number of other nodes is obviously central to the network and likely to play an important role [17]. [63] introduced the first centrality metrics, named as degree centrality, which is defined as the number of links incident upon a node. A distinction is made between in-degree and out-degree centrality, measuring the number of incoming and outgoing connections respectively. This distinction has also been considered important in social networks. For example, Twitter makes a distinction between friends and followers. Normally, on Twitter, users with a high in-degree centrality (i.e. with a high number of followers) are considered influencers.

In addition to degree centrality, the literature also shows other structural metrics for the identification of influencers in social networks. [101] presented an approach, where users were identified as influencers based on their total number of retweets [42]. Results highlighted how the number of retweets are positively correlated with the level of users' activity (number of tweets) and their in-degree centrality (number of followers). The PageRank score [32] has also been frequently adopted to evaluate influencers. It has been empirically found that a tweet has a larger reach if its author has a higher PageRank score [12, 96, 144]. The authors' ranking provided by the PageRank algorithm has been proved to be similar to that obtained with the number of followers. However, it has been found to be different from the ranking provided by the volumes of retweets [96, 103].

2.3.2 Content-based Variables

Besides structural metrics, the more recent literature has associated the complexity of the concept of influence with the variety of content. Several research works have addressed the need for considering content-based metrics of influence [25, 120, 148]. Content metrics such as the number of mentions, URLs, or hashtags have been proved to increase the probability of retweeting [16]. Considering the domain of recommendation, [45] proposed an approach by highlighting three properties: *recency of content*, *explicit interaction* among users and *user-generated content*.

[128] also discusses content-based features such as the number of mentions, URLs, or trending words. There metrics have been proved to increase the ability of predicting retweeting probability. In studies on information propagation, inclusion of URLs or hashtags is extensively used to define models for predicting mentions [65], retweeting probability [16,148] and topic adoption [106,151]. The dynamics of retweeting process is also discussed in a few studies [2,52,98,121,145].

2.4 Network Visualization Techniques and Aesthetics

Social media have become pervasive and ubiquitous. There is a growing need for information visualization, which has recently become a popular subject of research [92, 96, 119]. In general, information visualization aims at showing information in an easy, user-friendly and graphical way. However visualizing information properly is not trivial and becomes a challenge when the focus is social networks, such as Twitter. The graphical visualization of both social and semantic networks is considered complex due to their structure and role-relations defined in large-scale networks. A node can be linked to many other nodes, so the user can hardly understand the structure of the semantic net. In order to create a clear mental map, users need to apply a layout to the visualization. The nodes and edges should be placed in some order or semantic manner, that the user can understand clearly. Thus, a graph drawing layout is needed for any visualization in order to understand its structure in a meaningful way. We provide a conceptual visual framework and related software tool to the assessment of influence and identification of influencers. Considering the dynamics of influence and sparse complex nature of social networks, we need a visual approach in order to understand the connections and spread of the social networks through a visual exploration framework. Such disseminated information over widely-spread social networks, specifically Twitter, led us to adopt the visual platform, in order to better understand the network flow and spread of the influence. The framework of networks can be very useful when thinking about social dynamics, in this way, the people are represented as the nodes and their interactions are the connections (edges) between them. Using this model we demonstrate the influence spread and how someone can become influential in a social context by investigating characteristics of shared content that are an outcome of behavioral decisions made by social media users.

The existing literature provide details about various implementations, which have been proposed over the last few decades [35, 66, 69, 109, 155]. Due to computation constraints, scalability issues, and lack of aesthetic measures, these techniques or visualization tools [24, 99] have common shortfalls. In this section, we review the prominent graph drawing layout algorithms along with corresponding criteria to evaluate their effectiveness.

2.4.1 Network Visualization Techniques

The first spring-embedded model for network visualization has been proposed by [50, 51, 64], who have simplified the formula used to compute spring forces, and made significant improvements by using a cooling schedule to limit nodes maximum displacement. However, the repulsive force was still computed between all node pairs, yielding to an overall computational complexity of $O(N^2)$ for a network with N nodes. Subsequent studies that took a similar approach are the Online Force Directed Animated Visualization (OFDAV) technique by [81], and the edge-edge repulsion approach by [108]. More recently, [156] has proposed the over relaxation algorithm for force directed drawing. Despite these efforts, these force-directed algorithms are still considered non-scalable and unsuitable for large networks, as noted by [73].
Several research efforts in network visualization have targeted power-law algorithms and their combination with the traditional force-directed techniques, as for example in [6, 29, 43, 70, 84, 88, 107] and [7]. Among these approaches, the most notable is the Out-Degree Layout (ODL) for the visualization of large-scale network topologies, presented by [43, 127]. The core concept of the algorithm is the segmentation of the network nodes into multiple layers based on their out-degree, i.e. the number of outgoing edges of each node. The positioning of network nodes starts from those with the highest out-degree, under the assumption that nodes with a lower out-degree have a lower impact on visual effectiveness.

The topology of the network organization plays an important role such that there are plausible circumstances under which the highly connected nodes or the highest-betweenness nodes have little effect on the range of a given spreading process [5, 117]. For example, if a hub exists at the end of a branch at the periphery of a network, it will have a minimal impact in the spreading process through the core of the network, whereas a less connected person who is strategically placed in the core of the network will have a significant effect that leads to dissemination through a large fraction of the population. To identify the core and the multi-layered periphery of the clustered network we use a technique which is a metaphor of the *k-shell* (also called *k-core*) decomposition of the network, as discussed in [39, 91]. Examining this quantity in a number of social networks enables us to identify the best individual spreaders in the multi-layered periphery of a clustered network when the spreading originates in a single hub node.

2.4.2 Network Visualization Aesthetics

There are certain graph drawing constraints that a visualization must follow and these constraints are applied to graph in-order to obtain better understandability and structure of graph. For an effective and clear visualization of semantic nets, we need to discuss these graph drawing aesthetics.

[130, 131] discuss widely accepted aesthetic criteria and general graph drawing principles. These metrics are used to determine the quality of a graph or to define cost function of algorithms used in visualization. The literature provides some *generic* aesthetic guidelines that a network visualization should follow to improve understandability and enable visual scalability. The following represent the most common and widely accepted *generic* guidelines, as thoroughly discussed by [49, 130, 131, 150] and [88]:

- *Edge Bends* are defined as internal points of an edge whose coordinates do not lie on the straight line between the two end nodes of the edge. Edge-bends should be minimized to obtain a clear visualization.

- *Edge Crossings* represent points on a plane where two edges intersect. In a visualization, the number of edge crossings should be minimum.

- *Edge Orthogonality* is defined as the extent to which edges and edge segments follow the lines of an imaginary Cartesian grid. This should be maximum in an effective visualization.

- *Node Orthogonality* is defined as the extent to which nodes and bend points make maximum use of the grid points in an imaginary Cartesian grid. In a visualization, node orthogonality should be maximized.

- *Symmetry* - a graph should be symmetric along a single axis, for both nodes and edges. Symmetry should be maximized in order to improve aesthetics.

- *Consistent Flow* - In a directed graph, the flow of visualization should be consistently bound to a direction. Consistency should be maximized.

In addition to the afore-mentioned *generic* criteria, the literature provides *specific* criteria that should be followed to ensure visual effectiveness in identifying peripheral layers of nodes.

- *Local Minima* - Traditional force-directed techniques produce a graph whose total energy is a local minimum. Forces are too weak to spread peripheral layers across the graph. The local minima problem should be addressed to clearly identify peripheral layers of nodes.
- *Angular Resolution* - The angle between two incident edges leaving a node should be maximum to improve clarity, especially in multi-layered graphs.
- *Node Cluttering* - Nodes should not overlap on the graph canvas, to avoid view cluttering. Node cluttering should be minimized for the effective representation of multi-layered peripheral nodes around hub nodes.
- *Node Diameter Scaling* - To provide an effective visual representation and to distinguish the hub nodes from the multi-layered peripheral nodes on the canvas, the diameter of each node should be scaled upon the node's degree (i.e., the nodes with higher degree will have a larger diameter).

The guidelines surveyed in this section will be used in Chapter 3 to assess the visual effectiveness of our approach in the identification and exploration of layers of peripheral nodes.

2.5 Network Visualization Tools

This section summarizes the discussion upon state-of-the-art network visualization techniques both for semantic networks and social networks, respectively.

2.5.1 Semantic Network Visualization Tools

The most common and successful visualization tools are surveyed in [28, 67, 89, 115, 141] and [149]. *Cytoscape*, *OntoGraf*, *OntSphere*, *Giny*, *graphViz*, *HyperGraph*, *rdfGravity*, *IsaViz*, *Jambalaya*, *Owl2Prefuse*, *Flow-inspector*, *Gephi* and *SocNetV*. There is no one-to-one mapping between techniques and tools. This section discusses usage results from the literature or from experimental evidence that we made with the tools.

Existing tools are not highly scalable and with large-scale graphs, they are time inefficient or produce ambiguous layouts. Many visualization tools support graphs up to a few hundred nodes, such as *rdfGravity* [73], *Jambalaya* [147], *GraphViz* [53], and *Flow-inspector* [31]. With large-scale graphs, they are time inefficient or produce ambiguous layouts, as observed by [73] with rdfGravity.

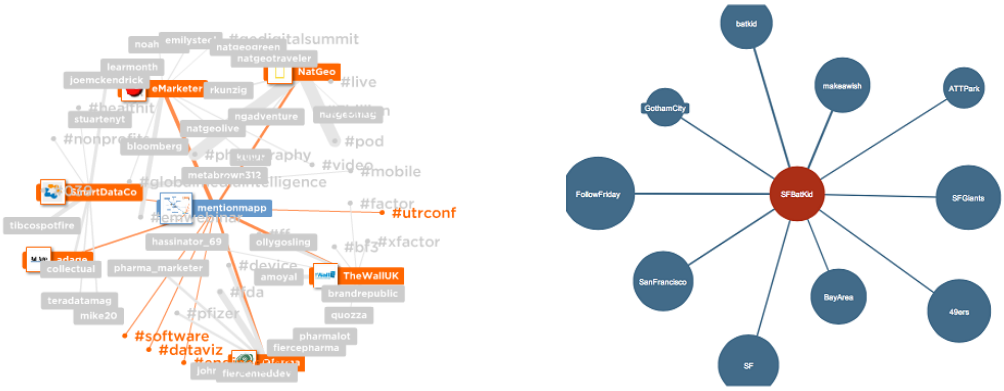
Node cluttering issues and edge overlap issues are common, as in *Prefuse* [78], *Gephi* [18], *GraphViz*, and *OntoGraf* [55]. Most tools do not support the description of role-relation hierarchies, as in *OntoGraf*, a Protégé plugin. Force-directed and spring layouts are implemented in several visualization tools, but local minima problems are common, as observed in *SocNetV* [87], *Gephi*, and in *Flow-inspector*.

The most practical shortcomings in existing tools, as noted by [88, 108, 131], are scalability, computational complexity, poor aesthetics, local-minima problem, complex topology layout, and convergence.

2.5.2 Social Network Visualization Tools

Social networks, more specifically, Twitter analytics tools generally aim at finding, analyzing and then optimizing a person's social growth. For example, *Twitonomy* [152] is an independent website, unaffiliated with Twitter that allows users to search for the Twitter history of accounts by entering a Twitter handle into a search box. Similarly, *Followerwonk* [57] is a web application which helps a user explore and grow his social graph. As discussed in [93], *Klout* is a system-generated tool for measuring influence; in other words it is a potential rating system that can be used as a measure of credibility. A user's Klout score is measured based on three components: true reach (how many people a user influences), amplification (how much the user influences them), and network impact (the influence of the user's network) (about [93]). Klout scores have a range of $1 - 100$, with a higher score indicating a higher level of influence. [90] discusses additional analytics tools including The *Archivist*, *SocialBro*, *Twenty Feet*, *TweetStats*, *Twitter Counter*, *Tweetstats*, and *TweepsMaps*. In [76, 143] authors also discusses about *NodeXL* tool which is NodeXL is a free, open-source template for Microsoft Excel 2007, 2010 and 2013 that makes it easy to explore network graphs. With *NodeXL*, a user can enter a network edge list in a worksheet, click a button and see its graph, all in the familiar environment of the Excel window.

The literature on social network visualization tools indicate that there exist only a few visualization tools. [90, 95] review existing tools, including *TouchGraph*, *MentionMap*, and *Hashtagify*. *TouchGraph* is a real-time web application which provides a cluster visualization of a user's Facebook network. It provides information for each friend and group of friends. The groups are clustered in different colors, but the representation is not friendly and a user cannot navigate or browse the network of other friends. Similarly, *MentionMap* provides a neat and interactive visualization, although sometimes it is hard to navigate due to ambiguous and cluttered graph layout, as shown in Figure 2.1a. It tends to discover the people who are more active in Twitter and the topics that they are talking about. The maximum depth of the graph is 2-level, as when a user browses another user's network, his/her own network disappears from visualization. Finally, *Hashtagify* [22] allows a user to visualize a network based on a Twitter Hashtag. Although the layout is not cluttered, as shown in Figure 2.1b as compared to MentionMap in Figure 2.1a, the tool does not allow the visualization of a user's friends or followers.



(a) *MentionMap* **(b)** *Hashtagify*

Figure 2.1: *Twitter visualization tools.*

2.6 Literature Gap

In this chapter we reviewed the literature in different fields related to the domain of application of this thesis. The literature on social media makes a distinction between influencers and influence. As discussed in the previous sections, the literature mainly focuses on the concept of influencers, while the relationship between content and influence is rather unexplored. This thesis takes a behavioural perspective by investigating characteristics of content, both at structural and content level, that are an outcome of behavioural decisions made by social media users. What we found missing from previous research is that content-based metrics of influence [25] do not measure influence based on quantitative properties of a user's activity within a social networks [12, 23, 38]. We think that these numerical properties help to a great extent in the discovery of influential people. These "numbers" provide us a lot of information, which, if it is correctly processed, will help us complement network topology based metrics [4, 8].

A further goal of this work is to design a scalable and robust power-law graph drawing technique in order to visualize complex social networks. We have discussed existing network visualization techniques and general graph drawing aesthetic criteria. This thesis will contribute to the quality of social network analysis by providing a visual framework to iteratively explore and interact with the network.

Finally, the literature has highlighted a few social network visualization tools. This survey will provide the basis to compare our tool with previous visualization approaches.

CHAPTER 3

Empirical Analysis of Social Influencers and Influence

In previous research, [17, 33, 92, 114] have shown how the content of messages can play a critical role and can be a determinant of the social influence of a message irrespective of the centrality of the message's author. Results suggest that peripheral nodes can be influential. This chapter starts from the observation made by [43] that social networks of influence follow a power-law distribution function [13], with a few hub nodes and a long tail of peripheral nodes, consistent with the so-called small-world phenomenon as noted by [159]. In social media, hub nodes represent social influencers [132], but influential content can be generated by peripheral nodes and spread along possibly multi-hop paths originated in peripheral network layers. The ultimate goal of our research is to understand how influential content spreads across the network. For this purpose, identifying and positioning hub nodes is not sufficient, while we need an approach that supports the exploration of peripheral nodes and of their mutual connections.

In this chapter, we exploit a modified power-law based force-directed algorithm [58, 62] to highlight the local multi-layered neighborhood clusters around hub nodes. The algorithm is based on the idea that hub nodes should be prioritized in laying out the overall network topology, but their placement should depend on the topology of peripheral nodes around them. In our approach, the topology of periphery is defined by grouping peripheral nodes based on the strength of their link to hub nodes, as well as the strength of their mutual interconnections, which is a metaphor of the k-shell decomposition analysis [39, 91].

The approach is tested on a large sample of tweets expressing opinions on a selection of Italian locations relevant to the tourism domain. Tweets have been semantically processed and tagged with information on *a)* the location to which they refer, *b)* the number of retweets, and *c)* the identifier of the retweeting author. With this information, we draw corresponding multi-mode networks highlighting the connections among authors (retweeting) and their interests (brand, category, and sentiment) by means of aesthetically pleasant layouts. By visually exploring and understanding the multi-layered periphery of nodes in clusters, we also propose a few content-related hypotheses in order to understand the relationship among *frequency*, *specificity*, and *retweets* (these variables will be defined below). Insights on the relationship among frequency, specificity, and influence would help social media users make their behavioral decisions. Social media users are aware that a post is influential if it raises attention from other users [8, 106]. Results highlight the effectiveness of our approach, providing interesting visual insights on how unveiling the structure of the periphery of the network can visually show the potential of peripheral nodes in determining influence.

This chapter also takes a behavioral perspective by proposing a few content-based hypotheses and by investigating characteristics of shared content that are an outcome of behavioral decisions made by social media users. In particular, we focus on three behavioral variables: *content specificity*, *frequency of sharing*, and *frequency of retweets*. The first variable represents the level of detail with which a user comments on a given subject of interest, while the second one represents the amount of contents shared by user in tweets and third one presents frequency of retweets of tweets shared by a given user. Insights on the relationship among *content specificity*, *frequency of sharing*, and *frequency of retweets* would help social media users to make their behavioral decisions. Fundamental goal of any social media user is to post content that is shared frequently, by many other users and over extended periods of time before fading [12, 56]. However, the literature does not provide systematic and visual evidence on how behavioral decisions regarding content specificity, frequency of sharing and expressed sentiment exert an impact on influence. This chapter provides preliminary evidence from Twitter. We put forward three hypotheses that tie *specificity*, *frequency*, and *retweets* which are tested on data samples of roughly one million tweets.

Empirical and visual results show a significant relationship between influence and behavioral decisions on content. This empirical and visual evidence raises theoretical challenges and encourages further research to understand the relationship between content and influence on social media. The main innovative aspect of our approach is that we use statistics (hypotheses) and visualization together. One can visually verify the proposed hypotheses on graphs produced by our proposed visualization approach.

The remainder of this chapter is structured as follows. Section 3.1 presents the proposed research hypotheses. Section 3.2 briefs about the implementation aspects of proposed visualization technique. Section 3.3 presents the data sample, network models and visualization results. Section 3.4 presents the empirical results. Discussions about results are in Section 3.5. Conclusions are drawn in Section 3.6.

3.1 Research Hypotheses

The literature indicates that social media are associated with a long-tail effect, with a variety of smaller communities [113]. While general content has a broad audience, there exists a variety of smaller communities who are interested in specific content. Such long-tail effect suggests that these communities are numerous and their specific interests are virtually limitless [56]. Social media users also consider specificity as an important metric for making behavioral decisions [33]. The specificity of social media users can be described as the level of detail with which a user comments on a given subject of interest. [92] has shown how the content of messages can play a critical role and can be a determinant of the social influence of a message irrespective of the centrality of the message's author. Twitter users with a high volume of tweets can be referred to as '*information sources*' or '*generators*' [85]. The literature also shows that social media users intend to post content that is shared frequently by many other users [12]. Social media users wish to be influential [44]. Intuitively, since users want to be interesting to many, if a user talks a lot, he/she will probably address the needs of multiple specific communities, i.e. multiple topics. Consequently, our first hypothesis posits a positive association between *frequency of tweets* and *content specificity* in multiple topics.

- *H1*: *Authors tweeting with a high frequency of tweets is positively associated with multiple topics (brands or categories) (i.e. visually, potential influencers are peripheral authors).*

If a speaker builds an audience around specific shared interests, content specificity may have a positive, as opposed to negative impact on audience attention. The literature suggests that social media user intend to post content that shared frequently by many other users [12]. The literature explains that retweeting is associated with information sharing, commenting or agreeing on other peoples' messages and entertaining followers [30]. [96] also shows that the most trending topics have an active period of one week, while half of retweets of a given tweet occurs within one hour and 75% within one day. The frequency of retweets is a major factor for estimating the quality of posts. It can be an important criterion since users tend to retweet valuable posts [44]. In the communities of people who are interested in specific content, users share specific content that followers are more likely to retweet. Intuitively, if a user tweets about multiple topics, interesting to many specific and active communities, he/she is most likely to get more retweets. Consequently, in the following hypothesis we posit a positive association between the *number of topics* and the *frequency of retweets*.

- *H2*: *Tweeting about multiple topics (brands or categories) is positively associated with the frequency of retweets (i.e. visually, peripheral authors, connected to multiple topics, are actual influencers).*

The breadth of the audience was considered the first and foremost indicator of influence for traditional media [46, 63, 80, 144], such as television or radio. However, traditional media are based on broadcasting rather than communication, while social media are truly interactive [21]. In traditional media, influencers intend to target a large audience by broadcasting frequently. Similarly, in social media, e.g. in Twitter, influencers intend to be more interactive by showing their presence and frequently tweeting [33]. If social media users are interested in something, they typically show it by participating in the conversation with a variety of mechanisms and, most commonly, by frequently sharing the content that they have liked [132]. A content that has had an impact on a user's mind is shared and gathers the attention of others [160]. The volumes of retweets are positively correlated with the level of users' activity (number of tweets) and their in-degree centrality (number of followers), as noted by [101, 157]. In social media, users are referred to as '*generalists*' or '*information sources*' if they talk about multiple topics [85]. On the contrary, there exist users, who are very specific in sharing content related to a specific topic or brand. These specific authors seems to be potential influence spreaders [56]. We posit that, these authors have to be active participants in each community by talking a lot. Our third hypothesis posits that such authors have a greater probability of being retweeted due to frequent tweets, and can be both potential and actual influencers.

- *H3*: *Tweeting more frequently (with a high frequency) about a single topic (brand or category) is positively associated with the frequency of retweets (i.e. visually, authors, drawn closer to single topic, are both actual and potential influencers).*

We posit the aforementioned three hypotheses that tie content specificity, frequency of tweets and frequency of retweets. Visually, hypothesis *H1* can be verified by observing the peripheral authors positioned in the outer-most layers of each cluster (lowest l-shell value, $l_s = 1$), which are only connected to one cluster hub (brand or category). These authors seem to be talking about a single brand or category. Such outlier authors can be *potential* influencers, if they further connect to other authors via content sharing and tweeting about multiple topics (brands or categories). Similarly, hypothesis *H2* can be visually verified by observing authors who are placed in between multiple clusters, connected to multiple clusters' hubs (brands or categories), and seem to be talking about multiple topics. These authors are *actual* influencers as they receive a high number of retweets by tweeting about multiple topics. Moreover, hypothesis *H3* can be visually verified by observing those authors who are positioned in the inner-most periphery of a single cluster (highest ls value) and seem to be placed close to the cluster hub (brand or category). Such authors are both *actual* and *potential* influencers as they are most specific about content sharing. These authors tweet frequently about a single topic (brand or category) and receive a high number of retweets.

3.2 Power-Law Algorithm

This section provides a high-level description of the graph layout algorithm used in this paper. An early version of the algorithm has also been presented by [58,61,83]. This paper improves the initial algorithm by identifying multiple layers of peripheral nodes around hub nodes according to the k-shell decomposition approach. The power-law layout algorithm belongs to the class of force-directed algorithms, such as the one by [43, 64, 88]. In this algorithm, we adopt a preprocessing method aimed at distinguishing hub nodes from peripheral nodes. This step is performed by pre- identifying hub nodes as N_h, which represents one of the following two sets:

1. A set of predefined tourism destinations, called brands, i.e. *Amalfi, Amalfi Coast, Lecce, Lucca, Naples, Palermo* and *Rome* (7 in total).
2. A set of predefined brand drivers of a destination's brand, called *categories*. Examples of categories are *Art & Culture*, *Food & Drinks*, *Events & Sport*, *Services & Transports*, etc., as explained in Section 3.3.

Figure 3.1 provides a general work-flow of the whole algorithm by showing its main building blocks. The proposed approach is aimed at the exploitation of the power-law degree distribution of author nodes (N_p). Provided that the distribution of the degree of the nodes follows a power law, we partition the network into a bipartite graph by distinguishing the set of predefined hub nodes N_h , which represents topics (brands or categories), and the set of peripheral nodes N_p, which represents authors, such that $\mathrm{N} = \mathrm{N_h} \cup \mathrm{N_p}$ with $\mathrm{N_h} \cap \mathrm{N_p} = \phi$.

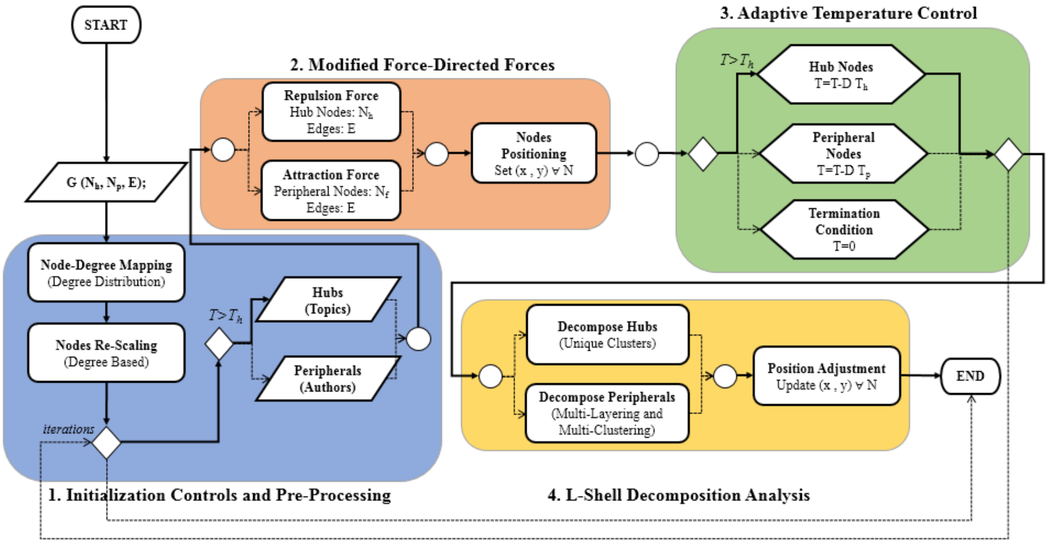


Figure 3.1: *Power-Law algorithm work-flow.*

The *Initial Controls and Pre- Processing* step is responsible for rescaling the size of each node in the graph, based upon the degree [139]. The higher the degree of a node, the greater the size and vice versa. This step is also responsible for partitioning the network into two predefined disjoint sets of vertices, (i.e. hub nodes - *topics*, and peripheral nodes - *authors*). The *Modified Force-Directed Forces* step calculates attraction and repulsion forces, based upon the value of T_h, which is a threshold value that can be tuned to optimize the layout, by providing maximum forces exerted upon Hub nodes N_h (*Adaptive Temperature Control*).

We introduce a customized dynamic temperature cool down scheme, which adapts the iterative step based on the current value of temperature. The temperature is supposed to be initialized at a value T_{start}, and then to be reduced by a variable Δt based on the current value of the temperature itself. This approach provides a convenient way to adapt the speed of iteration of the algorithm to the number of nodes to be processed. While processing hub nodes (a few), the temperature decreases slowly; while processing peripheral nodes (many), the temperature decreases more rapidly to avoid expensive computations for nodes that are not *central* to the overall graph layout. The formulae of attraction and repulsion forces are similar to those used in traditional force-directed approaches, such as [43]. In this paper, the forces formulae have been taken from the power-law based modified force-directed algorithm presented in [58, 59].

The *L-Shell Decomposition Analysis* step is responsible for the calculation of the l-shell value of author nodes in N_p, in order to create a multi-layered hierarchy of author nodes around the topics' nodes. This step also performs the final placement of nodes on graph canvas based on the computation of forces among nodes and l-shell mechanism. We tuned this technique by means of the metaphor of k-shell decomposition analysis [39, 91], in order to define the concept of level of each node in the multi-layered periphery of our graphs. This process assigns an integer as level index (l_s) to each node, representing its location according to successive layers (l shells) in the network. In this way, the author nodes who tweeted once about a specific topic, will have ($l_s = 1$) forming the out-most layer around that topic, and those who tweeted twice will have ($l_s = 2$) forming the inward successive layer, and so on. By this metaphor, small values of (l_s) define the periphery of the network (outliers), while the innermost network levels correspond to greater values of l_s, containing those authors who tweeted most frequently, as shown in Figure 3.2.

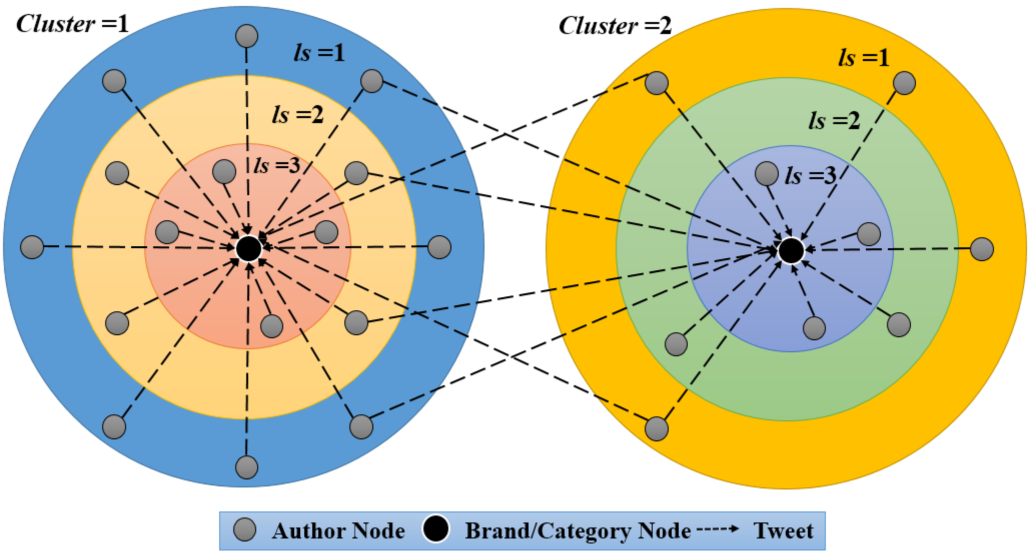


Figure 3.2: *Metaphor of k-shell decomposition analysis.*

3.3 Methodology

In this section, we will present the dataset that we have used in our experiment and the network models that we have built from the dataset. Empirical evaluations and related visualization results are also presented in this section

3.3.1 Variable Definition and Operationalization

Each graph $G(A,T)$ has a node set A representing authors and an edge set T representing tweets. We define as $N_T(a)$ the total number of tweets posted by author a. We define as $N_R(a)$ the total number of times author a has been retweeted. Tweets can refer to a brand b or to a category c. We define as $N_B(a)$ the total number of brands mentioned by each author a, in all his/her tweets, i.e. *brand specificity*. Similarly, $N_C(a)$ represents the total number of categories mentioned by each author a, in all his/her tweets, i.e. *category specificity*.

3.3.2 Data Sample

The tourism domain has been used as a running example as it is one of the most common industries in the field of social networks [112]. We collected a sample of tweets over a two-month period (December 2012 - January 2013). For the collection of tweets, we queried the public Twitter APIs by means of an automated collection tool developed ad-hoc. Twitter APIs have been queried with the following crawling keywords, representing tourism destinations (i.e. brands): *Amalfi*, *Amalfi Coast*, *Lecce*, *Lucca*, *Naples*, *Palermo* and *Rome*. Two languages have been considered, *English* and *Italian*. Collected tweets have been first analyzed with a proprietary semantic engine in order to tag each tweet with information about *a)* the location to which it refers, *b)* the location's brand driver (or category) on which authors express an opinion, *c)* the number of retweets (if any), and *d)* the identifier of the retweeting author.

Our data sample is referred to the tourism domain. We have adopted a modified version of the Anholt Nation Brand index model to define a set of categories of content referring to specific brand drivers of a destination's brand [9]. Examples of brand drivers are *Art & Culture*, *Food & Drinks*, *Events & Sport*, *Services & Transports*, etc. A tweet is considered Generic if it does not refer to any Specific brand driver, while it is considered *Specific* if it refers to at least one of Anholt's brand drivers. Tweets have been categorized by using an automatic semantic text processing engine that has been developed as part of this research [17,33]. The semantic engine can analyze a tweet and assign it to one or more semantic categories. The engine has been instructed to categorize according to the brand drivers of Anholt's model, by associating each brand driver with a specific content category described by means of a network of keywords. Each tweet can be assigned to multiple categories. We denote with N_C the number of categories each tweet *w* is assigned to; the specificity $S(w)$ of a given tweet *w* is defined in Equation 3.1 as follows:

$$S\left(w\right)=\left\{\begin{array}{l}0,N_c=0\\1,N_c>0\end{array}\right\} \tag{3.1}$$

The data collection step has been followed by a preliminary data analysis aimed to the statistical exploration of the characteristics of data distribution. Results highlighted that all variables follow a power-law distribution [122]. Since the Structural Equation Modeling (SEM) tool [15, 34, 146] adopted for model verification provides only linear regression to model variable relationships, each variable has been represented on a logarithmic scale and standardized. For the sake of clarity, the values reported in Table 3.1 refer to the descriptive statistics of the original non-linear variables.

Table 3.1: *Basic descriptive statistics of our data set.*

Variable	Value	S.D.
Number of tweets	957,632	–
Number of retweeted tweets	79,691	–
Number of tweeting authors	52,175	–
Number of retweets	235,790	–
Number of retweeting authors	66,227	–
Average number of tweets per author	10.07	± 86.83
Average number of retweeted tweets per author	1.525	± 4.67
Average number of retweets per author	1.40	± 4.52
Average frequency of retweets per author	0.58	± 0.38
Average content Specificity per author	0.35	± 0.46

3.3.3 Network Models

In order to verify the effectiveness of the proposed algorithm with respect to the goal of our research, we have defined different network models based on the data set described in the previous section. Figure 3.3 provides an overview of the adopted network models.

- *Author → Brand (N_1)*: This model considers the relationship among authors and domain brands, i.e., touristic destinations in our data set. The network is modeled as an undirected affiliation two-mode network, where an author node n_a is connected to a brand node n_b whenever author a has mentioned brand b in at least one of his/her tweets.

- *Author → Category (N_2)*: This model considers the relationship among authors and domain brand drivers (categories), i.e., city brand drivers in our data set (namely, *Arts & Culture*, *Events & Sports*, *Fares & Tickets*, *Fashion & Shopping*, *Food & Drink*, *Life & Entertainment*, *Night & Music*, *Services & Transport*, and *Weather & Environmental*). The network is modelled as an undirected affiliation two-mode network, where an author node n_a is connected to a category node n_c whenever author a has mentioned a subject belonging to category c in at least one of his/her tweets.

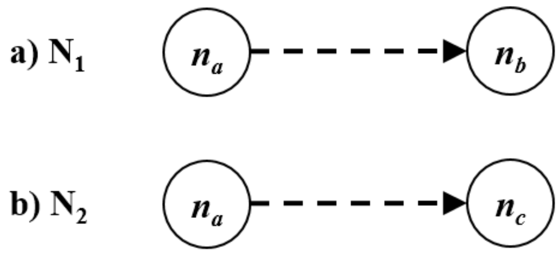


Figure 3.3: *Network Models a) N_1: Author → Brand. b) N_2: Author → Category.*

3.3.4 Network Visualization Results

The empirical results and discussions on network visualization will adopt network N_1 network (i.e. Author → Brand) as reference example. Figure 3.4 provides an enlarged view of network N_1 visualized by means of the proposed power-law layout algorithm. A summary description for N_1 and N_2 networks is presented in Table 3.2, where $N_R(a)$ represents the total number of retweets, $N_B(a)$ shows the total number of tweets in which author a talked about brand B (N_1 network), $N_c(a)$ shows the total number of tweets in which author a talked about category C (N_2 network), and $N_T(a)$ represents the total frequency of author a (i.e. the total number of tweets of author a).

The network visualization depicted in Figure 3.4 adopts multicolored nodes to represent authors, and highlighted encircled blue (dark) nodes to represent the tourism destinations (i.e. brands) on which authors have expressed opinions in their tweets. The layout of the network produced by the power-law layout algorithm clearly highlights that author nodes aggregate in several groups and subgroups based on their connections with brand nodes, which in this case are the hub nodes.

The groups of author nodes cluster together all those authors that are connected to the same hubs (i.e. brands) referred to as *cluster*. Our approach provides a visual clustering for those authors who have tweeted about the same brand.

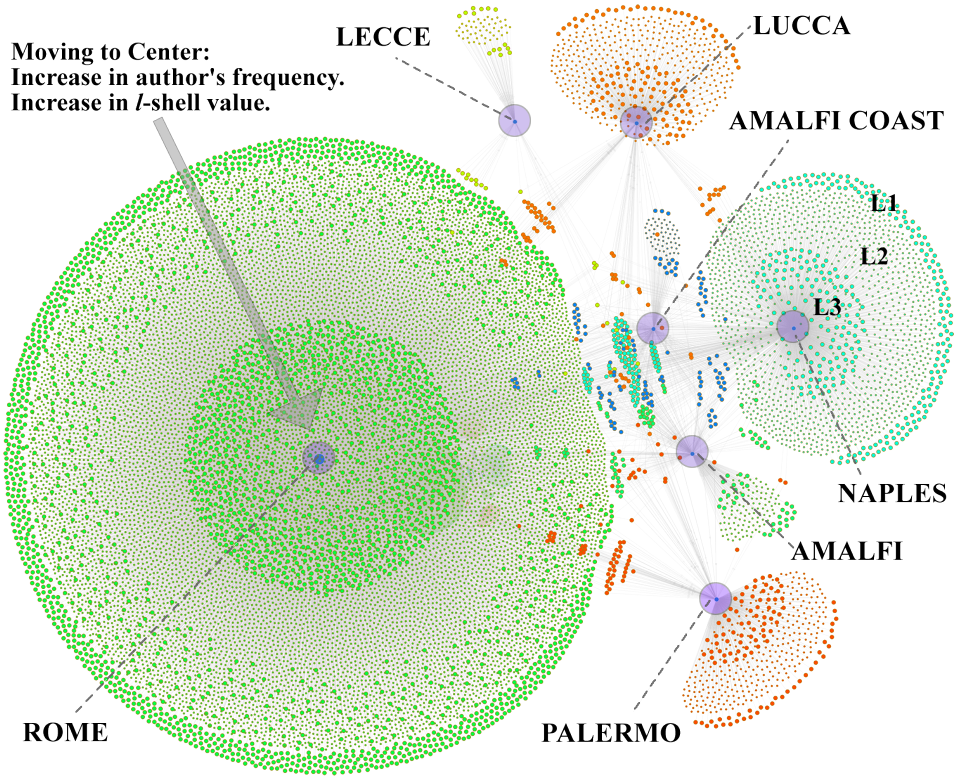


Figure 3.4: *N_1 Network: Author → Brand (Enlarged View).*

Table 3.2: *Descriptive statistics on the dimensions of N_1 and N_2 networks.*

Authors	$N_R(a)$	N_1	N_2	$N_T(a)$
		$N_B(a)$	$N_C(a)$	
398	92	856	1,913	2,769
1,662	364	2,905	5,959	8,864
10,710	2,907	12,559	18,498	31,057
18,711	5,329	21,140	29,842	50,982
30,310	8,690	33,684	46,120	79,804
37,626	10,529	41,620	56,960	98,580
47,295	12,833	52,208	71,667	1,23,875

3.4 Empirical Results

This section reports on the empirical testing and evaluation of the proposed hypotheses. First, we discuss our research model and then we present empirical results.

3.4.1 Research Model

AMOS 20 [10] has been used to analyze the research model that we adopted for estimation analysis (see Figure 3.5. In Figure 3.5, we report each variable relationship only in its standardized regression coefficient's sign (note that signs are consistent between the two data sets N_1 and N_2). In this model, $N_T(a)$ represents a dependent variable as it is measured with multiple independent variables, which are $N_R(a)$, $N_B(a)$, and $N_C(a)$.

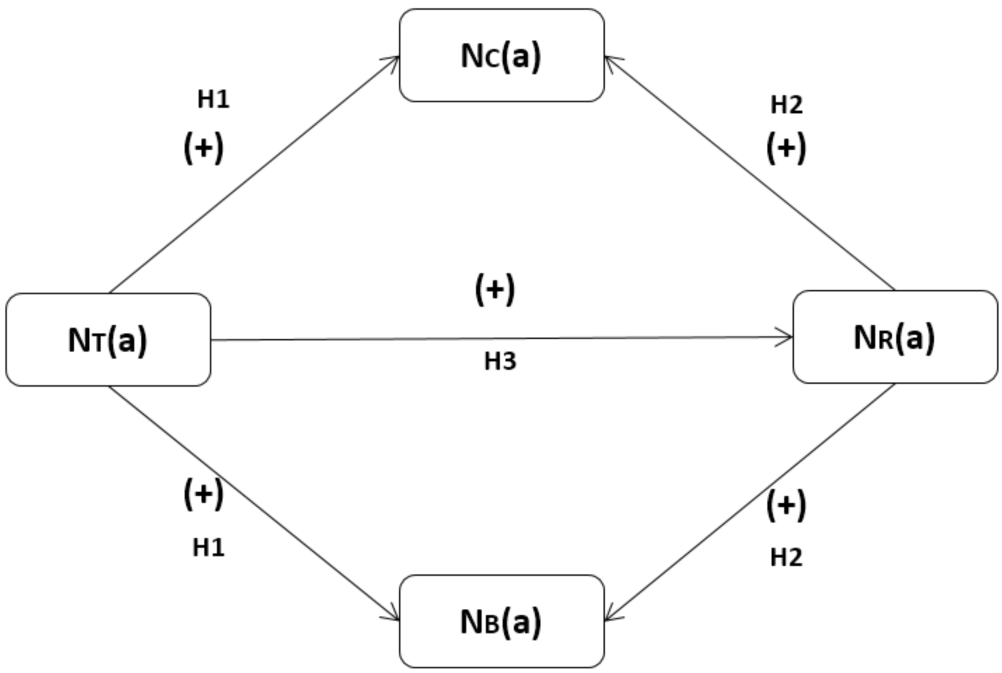


Figure 3.5: *Research Model.*

Statistical Results

All statistical analyses have been performed with SPSS 20 [123]. Correlation and Regression analyses have been performed on our data set. Tables 3.3 reports the descriptive statistics of each variable from our dataset that we used for statistical analysis and to validate our proposed research hypotheses, as discussed in Section 3.1.

Table 3.3: *Descriptive statistics of each variable from dataset.*

	$N_R(a)$	$N_B(a)$	$N_C(a)$	$N_T(a)$
Mean	1.37	1.04	1.53	2.78
S.E of Mean	0.009	0.000	0.001	0.002
S.D	10.817	0.283	1.109	1.837
Variance	117.007	0.080	1.230	3.375

Figure 3.6a presents the histogram of $N_T(a)$, indicating that most authors have a low number of tweets, and a few have a very large number of tweets, according to the small world phenomenon and power-law distribution [122] in a network, i.e. there are a few hub nodes. Similarly, Figure 3.6b represents the histogram of $N_R(a)$, indicating that most authors have a low number of retweets. Those who have a large number of retweets, are actually influencers or hub-nodes with a large number of tweets. Similarly, Figure 3.6c represents the histogram of *brand specificity*. We can observe whether a brand's specificity is high or low. Likewise, in Figure 3.6d, we can observe *category specificity*.

Table 3.4 presents the correlation matrix among the persistence variables used in our analysis. From Table 3.4, it follows that correlation is significant at 0.01 level (2-tailed). All persistence variables are positively correlated with each other, and thus, have a significant impact upon each other.

Table 3.4: *Correlation matrix of persistence variables (Pearson Index).*

	$N_T(a)$	$N_R(a)$	$N_B(a)$	$N_C(a)$
$N_T(a)$	1	0.326	0.590	0.898
$N_R(a)$	0.326	1	0.254	0.235
$N_B(a)$	0.590	0.254	1	0.392
$N_C(a)$	0.898	0.235	0.392	1

The regression estimation results of the research model are shown in Table 3.5. All relationships between persistence metrics (i.e. $N_R(a)$, $N_B(a)$, and $N_C(a)$) and the persistence latent variable (i.e. $N_T(a)$) are significant, with $p < 0.001$. This confirms that factorization was performed correctly over fitted research model.

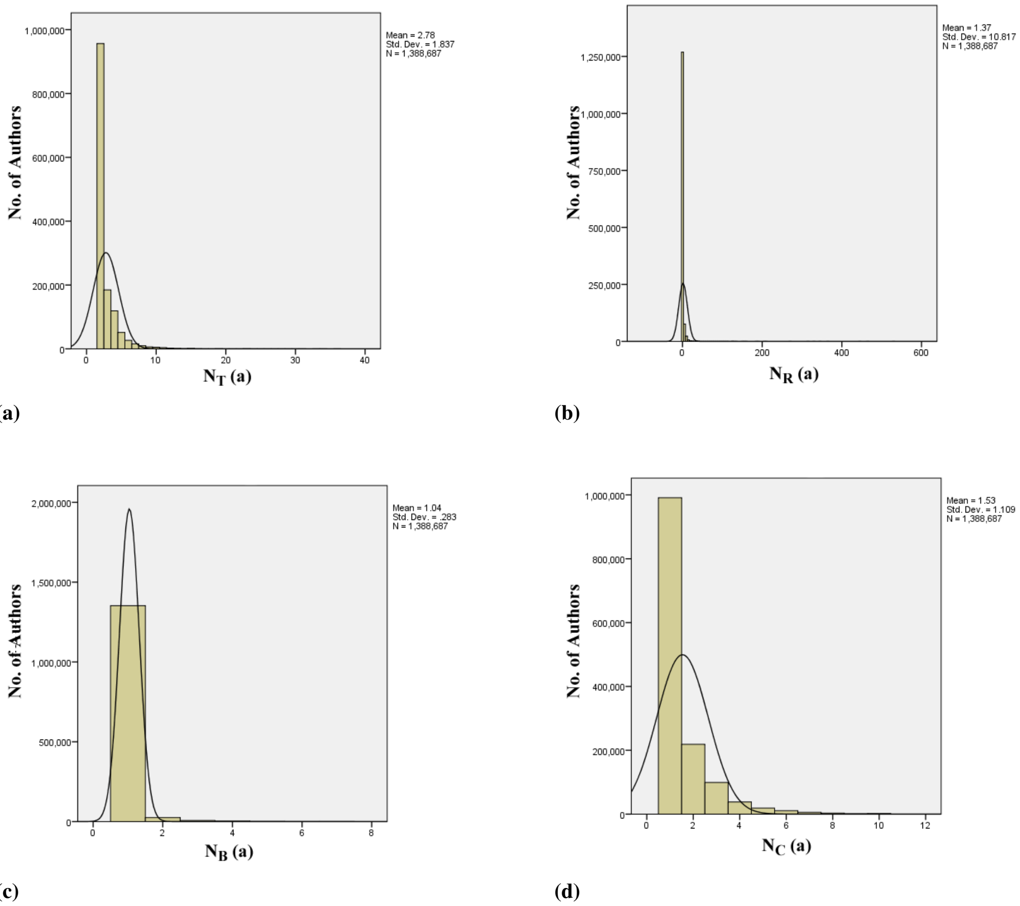


Figure 3.6: *Frequency distribution of a) the number of tweets and b) the number of retweets c) brand specificity, d) category specificity.*

Table 3.5: *Estimates of regression weights for the research model.*

$V_{Dependent}$	$V_{Independent}$	R_W	S.E	$p - value$
$N_R(a)$	$N_T(a)$	0.082	0.000	< 0.001
$N_B(a)$	$N_T(a)$	0.000	0.000	< 0.001
$N_C(a)$	$N_T(a)$	0.000	0.000	< 0.001

- *HYPOTHESIS H1*
  Hypothesis *H1* "*Tweeting with a high frequency of tweets is positively associated with number of topics (brands or categories) (i.e. visually potential influencers are the peripheral authors)*" has been tested through correlation. As shown in Table 3.4, both $N_C(a)$ and $N_B(a)$ have positive correlation of 0.898 and 0.590, respectively, with $N_T(a)$ at 0.01 level of significance. Hence, both correlation values support hypothesis *H1*. It means that *generalist* authors, who tweet about multiple topics (brands or categories), are more likely to be *content specifiers*. Such authors, by having a greater probability of sharing contents, can be *potential influencers* in their network.

  Similarly, through visualization results we can observe the big sized author nodes, who tweet a lot about multiple brands (Figure 3.4) or about multiple categories (Appendix 1).

- *HYPOTHESIS H2*
  Similarly, hypothesis *H2*, "*Tweeting about multiple topics (brands or categories) is positively associated with the frequency of retweets (i.e. visually, peripheral authors, connected to multiple topics, are actual influencers)*", has been tested through correlation. By Table 3.4, both $N_C(a)$ and $N_B(a)$ have a positive correlation of 0.254 and 0.235, respectively, with $N_R(a)$ at 0.01 level of significance. Hence, both correlation values support hypothesis *H2*. This means that, authors, who have a large number of retweets, are also *content specifiers* or can also be '*information sources*' or '*generators*'. Such authors can be actual influencers in spreading information within a network, as they receive a large number of retweets by tweeting about multiple topics.

  From a visualization standpoint, if we explore the graph shown Figure 3.4), we can note how authors who seem to be big sized nodes (visually drawn in-between multiple cluster peripheries) talking about multiple topics (brands or categories), also have a high number of retweets.

- *HYPOTHESIS H3*
  Similarly, hypothesis *H3*, "*Tweeting more frequently about a single topic (brand or category) is positively associated with the frequency of retweets (i.e. visually, authors drawn closer to single topic, are both actual and potential influencers)*", has been tested through correlation. By observing values from Table 3.4, $N_T(a)$ and $N_R(a)$ have a positive correlation of 0.326 at 0.01 level of significance. Although the correlation coefficient is not high, the p-value in Table 3.5 on significance and seems to support a positive (though weak) correlation between $N_T(a)$ and $N_R(a)$. As per descriptive statistics of networks, presented in Table 3.2, we can observe that as the

number of tweets increases, the number of retweets also increases for each size or network topology.

From a visual standpoint, as shown in Figure 3.4, we know that the nodes (which are drawn closer to a single brand in the innermost periphery of distinct clusters) are those authors who tweet most frequently about a specific brand in their cluster. Such authors are connected closer to cluster hubs (brands or categories), by having a *high l-shell value* and a *high number of tweets* (as discussed earlier in Section 3.2).

3.5 Discussion

The network layout shows that clusters are placed at a different distance from the visualization center based on the number of hubs to which they are connected. In other words, the most peripheral clusters are those in which nodes are connected to only one hub, while the central cluster is the one in which nodes are connected to the highest number of hub nodes. Within a single cluster, multiple layers seem to be formed. By implementing the *l-shell* decomposition methodology, the outside layer consists of author nodes who posted a tweet only once, as we move inward towards the brand node (hub), the frequency of tweeting increases. Hence, the closest nodes to a hub represent the authors who tweeted most about that brand and are both *potential* and *actual influencers*. The power-law layout algorithm has provided a net-work layout that is very effective in highlighting a specific property of authors which was not a measured variable in our dataset, i.e. their specificity (or generality) with respect to a topic (i.e. a brand as in Figure 3.4 or category in Appendix 1). Authors belonging to different clusters are in fact those who are more *generalist* in their content sharing, since they tweet about multiple different brands. On the contrary, authors belonging to the innermost clusters are those who are very *specific* in sharing content related to one brand.

Since the *specificity* (generality), *frequency of tweets* and *retweets* of authors was not an explicitly measured variable in our dataset, it is possible to posit that the pro-posed network layout algorithm can be considered as a powerful visual data analysis tool, since it is effective in providing visual representations of networks that help unveiling specific (implicit) properties of the represented networks.

We also noticed that, as the graph size increases, more peripheral layers seems to be formed surrounding hub nodes, which increase the influence spread across newly formed peripheral layers in multi-layered form. Authors seem to evolve by tweeting about multiple topics among multiple peripheries. We can visually identify the increase in influence spread, as shown in Figures 6 and 7, which are larger graphs of the N_1 type network, as compared to Figure 3.4, where the addition of more multi-layered peripheral nodes around hub-nodes (i.e. brands) increases the influence spread across those peripheral layers. The outlier authors along the periphery can be potential influence spreaders, if they connect with other clusters through retweeting and, thus, play a critical role in determining influence. As presented in Figure 4, network N_1 is related to the relationship between authors and brands, i.e., touristic destinations. In this case, the clustering of nodes provides a distinct clustering of those authors who have tweeted about the same destination. The layering of nodes around brands is instead related to the intensity of tweeting about a given destination; i.e., authors closer to a brand node tweet a higher number of times about that destination with respect to farther authors. The emerging semantics of the network visualization in this case is related to the *brand fidelity* of authors. The visualized network layout supports the visual analysis of those authors who have a higher fidelity to a given brand, or those authors who never tweet about that brand.

This paper's findings have some practical implications on how to design a strategy to promote tourism destinations. For example, findings suggest that to promote a specific brand, WoM may become more efficient by linking that specific brand with other brands, as this seems to increase reach and influence. For example, they can share posts comparing their brand with other competing and non-competing brands [19,54,100]. Similarly, they can identify the most popular and least popular brands, as the multi-layered peripheral network of author nodes reveals potential and actual influencers. They can target authors in the periphery who can be *information spreaders* and, thus, connect to other communities in order to increase reach. Tourism practitioners can also identify the most widely discussed topics (categories) and focus on them in their advertising campaigns. For example, while addressing a specific brand (e.g. Rome), they can relate it with a specific category (e.g. Arts & Culture), in order to increase the specificity of their posts. From a visualization standpoint, tourism practitioners can also identify the key players in the network and classify them as *information spreaders*, *sources*, or *seekers*. *Information spreaders* can either be *generalist* authors who are connected to multiple communities and discuss about multiple topics, as they have a broad reach and a significant influence. Becoming an engaging member of relevant communities will give social media users a chance to promote content to a targeted audience and in-crease their actual influence.

3.6 Conclusion

We have discussed a novel visual approach to the analysis and exploration of social networks in order to identify and visually highlight influencers (i.e., hub nodes), and influence (i.e., spread of multi-layer peripheral nodes), represented by the opinions expressed by social media users on a given set of topics. Results show that our approach produces aesthetically pleasant graph layouts, by highlighting multi-layered clusters of nodes surrounding hub nodes (the main topics). These peripheral node clusters represent a visual aid to understand influence. Empirical testing and evaluation results show that the proposed three hypothesis that tie *content specificity*, *frequency of tweets* and *retweets* are supported. Moreover, *specificity*, *frequency*, and *retweets* are also mutually correlated, and have a significant impact on an author's influence and encourage us to further explore a social network's intrinsic characteristics.

Such outcomes can be leveraged by tourism practitioners, marketing departments or social media community. For example, one can analyze the most competitive locations, events or initiatives in the market. Social media marketing managers can also visually identify major key players in the network, like *information spreaders* and *information sources*. In social media communities, users like *information seekers*, would be able to visually identify the actual and potential influencers and can further follow them.

APPENDIX 1

Figures 3.7, 3.8,3.9, and 3.10 provide few more visualization of networks N_1 and N_2 of our dataset. An enlarged and zoomable version of the network layouts can be accessed online at the following URL: `http://goo.gl/97v8zu`.

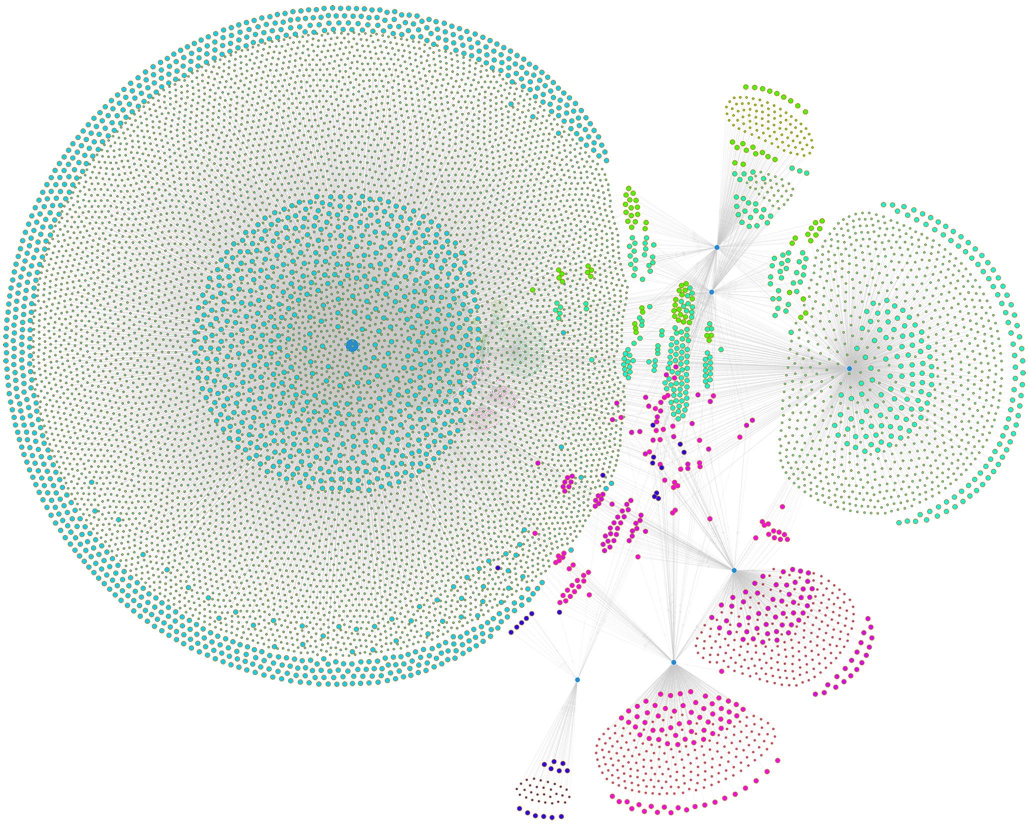


Figure 3.7: *Network visualizations of N_1 (Author $\rightarrow$ Brand).*

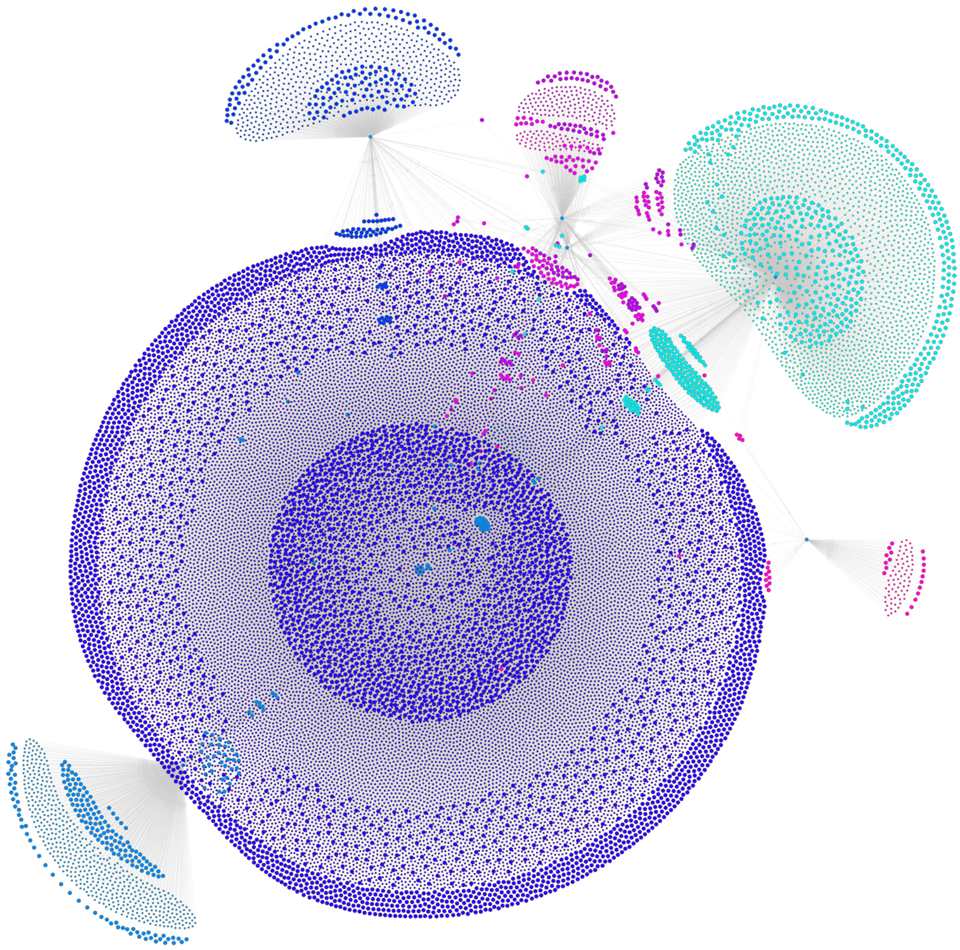


Figure 3.8: *Network visualizations of N_1 (Author $\rightarrow$ Brand).*

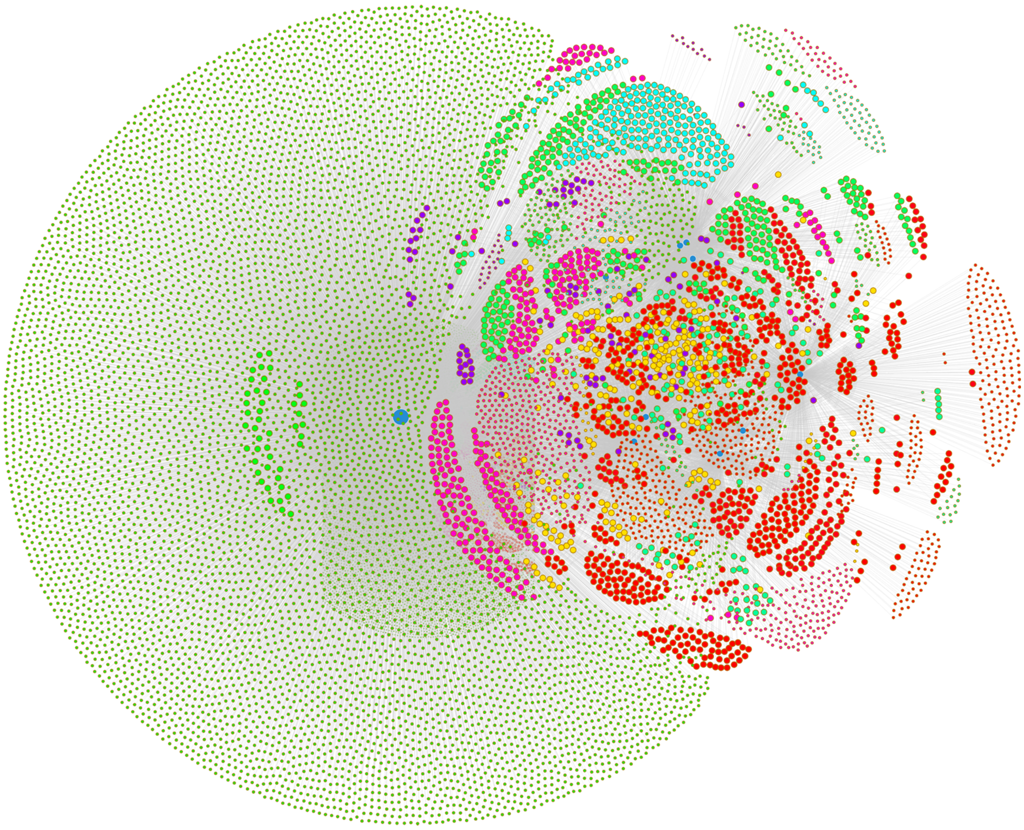


Figure 3.9: *Network visualizations of network N_2 (Author $\rightarrow$ Category).*

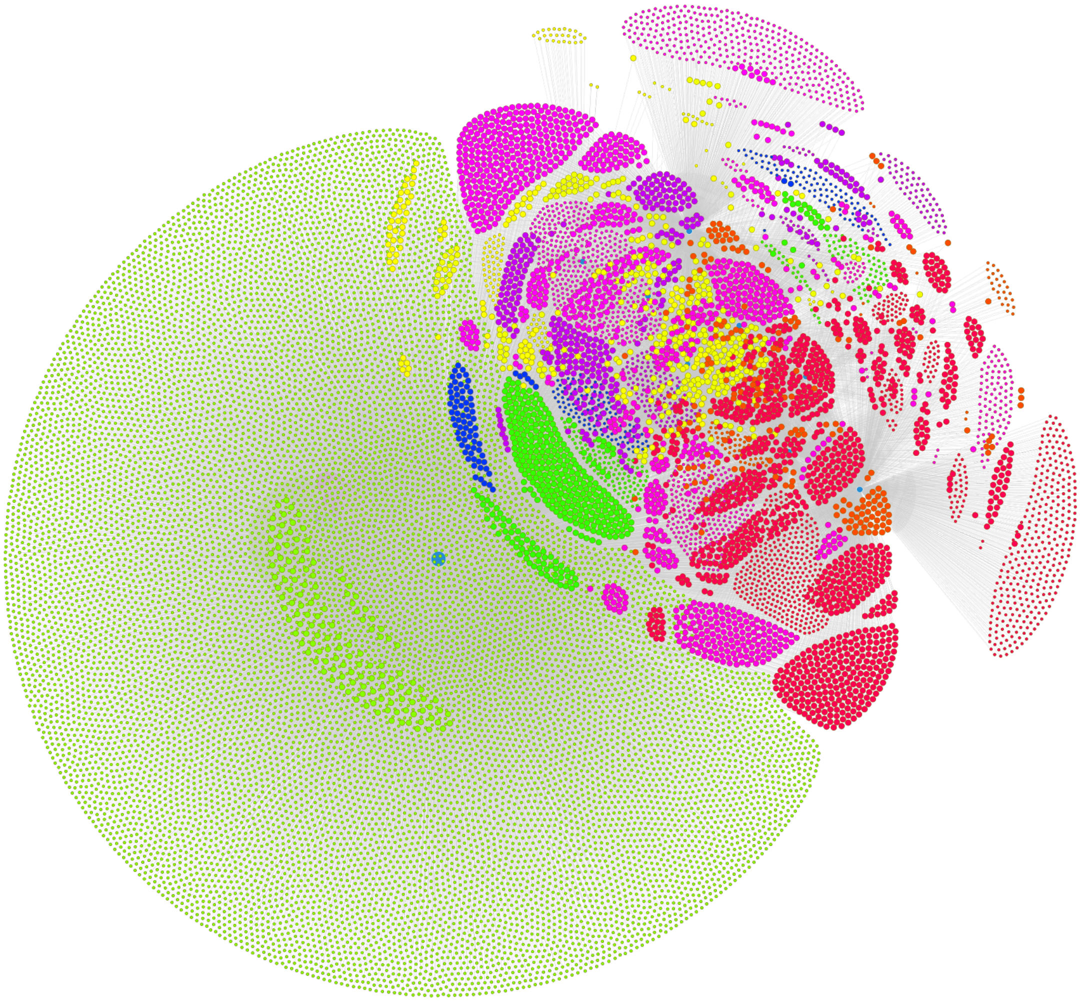


Figure 3.10: *Network visualizations of network N_2 (Author $\rightarrow$ Category).*

CHAPTER 4

Influence-based Exploration of Social Networks

This chapter describes the influence parameters provide by Twitter network, which we considered to identify influencers within social network. Further, we discuss the user ranking methodology based upon these parameters. We also discuss the *Analytical Hierarchy Processing* technique, which we used to implement our ranking mechanism. Finally we discuss the overall ranking.

The existing literature indicates that many researchers endeavor to focus on the shared content provided by users in social networks [41, 65, 68], in order to provide a ranking based on the actual influence of the content that they share (e.g.ranking by number of followers). The social media literature makes a distinction between influencers and influence. Influencers are prominent social media users with a broad audience. For example, social users with a high number of followers and retweets on Twitter [146], or a multitude of friends on Facebook [142], or a broad network of connections on LinkedIn [134]. The term influence refers to the social impact of the content shared by social media users. If social media users seem to be interested in something, they normally show it by participating in the conversation with a variety of mechanisms, mostly by sharing the content that they have liked. [9, 119] has noted that a content that has an impact on a user's mind is usually shared. Influencers are prominent social media users, but we cannot be certain that the their shared content has influence, as discussed by [21].

Social media have become pervasive and ubiquitous. There is a growing need for information visualization, which has recently become a popular subject of research [56, 92, 119]. In general, information visualization aims at showing information in an easy, user-friendly and graphical way. However visualizing information properly is not trivial and becomes a challenge for large social networks, such as Twitter. Twitter has been defined by many researches as the key role player of the change on how information dissemination is accomplished. Its influence on information dissemination has led to research exploring how this is achieved. According to [96] the unicity of direction in Twitter connections provides the key driver of information dissemination via word of mouth (WoM) in retweets [75, 133]. The ultimate goal of our research is to provide a novel visual framework to analyze, explore and interact with Twitter's 'Who Follows Who' relationships by browsing the friends' network to identify the key influencers upon the actual influence of the content they share. The ultimate goal of our research is to provide a novel visual framework to analyze, explore and interact with Twitter's 'Who Follows Who' relationships, by browsing friends' networks to identify the key influencers based upon the actual influence of the content that they share.

4.1 Influence-based Parameters

Several research works have addressed the need for considering content-based metrics of influence [25]. Content metrics such as the number of mentions [128], URLs [65, 148], or hashtags [98, 151] have been proved to increase the probability of retweeting [16, 121]. Twitter has been the most common dataset for researches on user influence. For example, [44, 96] measure the influence of Twitter users based on the sheer number of retweets spawned from the users' tweets. Recently, [158] have studied the elite users who control a significant portion of the production, flow, and consumption of information in the Twitter network. In [158] a top-down approach is used by identifying top users based on how frequently these appear in user-generated lists.

What we found missing from previous research is that generally they do not base the influence measurement on numerical properties of the activity of a tweeter [12, 23, 38]. We think that these numerical properties help to a great extent in the discovery of possibly influential people. These "numbers" provide us a lot of information, which if it is correctly processed will help us in finding possibly influential users [4, 8]. We provide a list of the most important types of Twitter properties that can be used to assess influence. We defined 10 parameters for which we wanted to retrieve data in order to have a deeper insight and to further specify our ranking algorithm. We decided to categorize our parameters in two main categories, as follows:

- **User-Level Resources**: This category includes all the parameters that describe a Twitter user, such as number of favorite tweets, number of followers, number of people that the use is following, etc.

- **Tweet-Level Resources**: This category includes all the parameters that can be obtained by analyzing the social dynamics of the most recent tweets of each user, such as total number of recent tweets, th number of times the tweets have been marked as favorite by other users (number of favorited), the number of times each tweet has been retweeted, etc.

4.2 User Ranking Methodology

Our goal is to provide a ranking of users of Twitter, based on their influence parameters, as discussed in Table 4.1. We believe that each parameter plays an important role in identifying influencers within the network. The problem is to define how important each of these parameters is. For this purpose, we have used the method proposed by Saaty [135,136], called *Analytic Hierarchy Process* (AHP). This method has been used for decades and is widely accepted by the scientific community. We will explain how the method works in the next section. Our ranking methodology is shown in Figure 4.1. The outcome of AHP is a vector of weights for different. NavigTweet provides an aggregated score of each user as a weighed sum of different parameters using the weights obtained from AHP. The higher the score the higher the rank, and vice versa.

4.2.1 Analytical Hierarchy Process (AHP)

The Analytical Hierarchy Process (AHP) method is useful for decision makers when there is one objective, many criteria to make the decision and many alternatives. The interesting feature of AHP is that it compares criteria in a pairwise fashion. The decision maker who provides estimates of the relative importance of criteria by comparing them pairwise. Figure 4.2 presents an overview of the AHP method. The outcome of AHP is the weight vector, containing a weight for each parameter.

Table 4.1: *Influence-based parameters of Twitter.*

USER-LEVEL RESOURCES	
Number of Following	"Following" someone means you will see their tweets (Twitter updates) in your personal timeline. Twitter lets you see who you follow and also who is following you.
Number of Favorites	Favoriting a tweet can let the original poster know that you liked their tweet. A user marks a tweet as favorite, in order to save it and have the possibility to check it later. We intend to measure volumes of those tweets that have been marked as 'favorite'.
Number of Tweets	The total number of posts that the user has made since the time of Twitter sign-up.
Number of Lists	Lists are a shorter way of having information regarding a topic of interest, the user is interested in. A user subscribes to such lists or creates them in Twitter. It also depicts user engagement in reading what other people post.
Number of Followers	The number of users engaged in posts from the particular user, or subscribed to receives updates from that particular user.
TWEET-LEVEL RESOURCES (200 Recent Tweets)	
Number of URLs	Tweets that include web-links are more likely to be retweeted. This makes us believe that a user with influential characteristics should have a high volume of such tweets that include URLs.
Number of Hashtags	A hashtag is used to mark keywords regarding specific topics of interest. Using hashtags shows that the user is quite friendly to twitter topics and likely to have a certain amount of influence.
Number of Retweets	A retweet is used to share a post that someone else posted before. It is considered as some information the user likes and considers worth retweeting. Measuring the volume of the retweets provides some useful information on whether the posts of the user are followed by the other users too.
Number of Favourited	A user makes a tweet as favorite, in order to save it and have the possibility to check it any time later. Calculating the number of times a tweet has been marked as favorite by other users shows that the user is influent to some other people.
Number of Mentions	A user is mentioned in a tweet when the tweet is thought to be of his interest or just to be included in the message sharing. Mentioning users requires knowing what information other users like and would like to be involved in, and also considered as a factor that increases user influence.

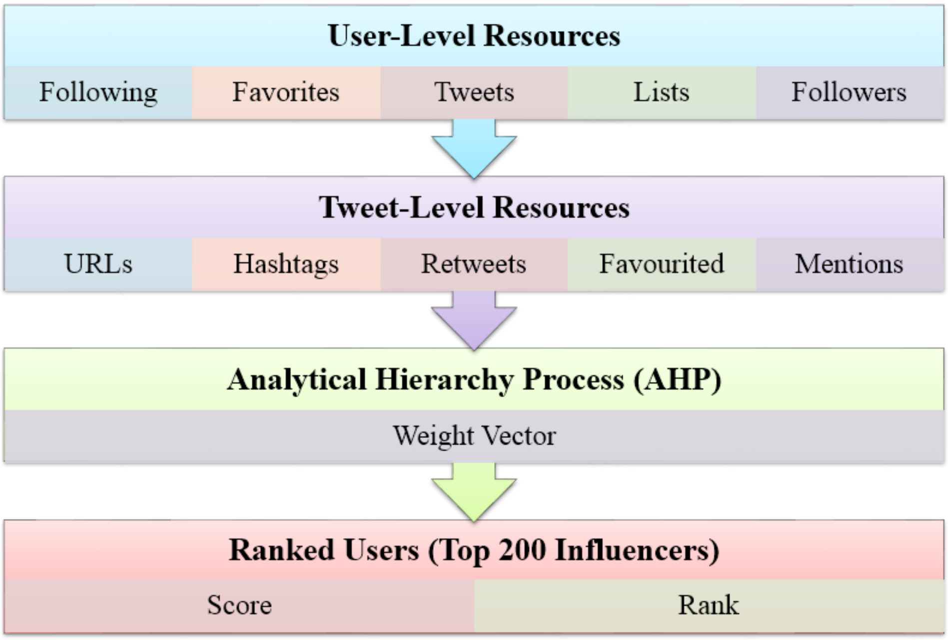


Figure 4.1: *Influence-based Ranking.*

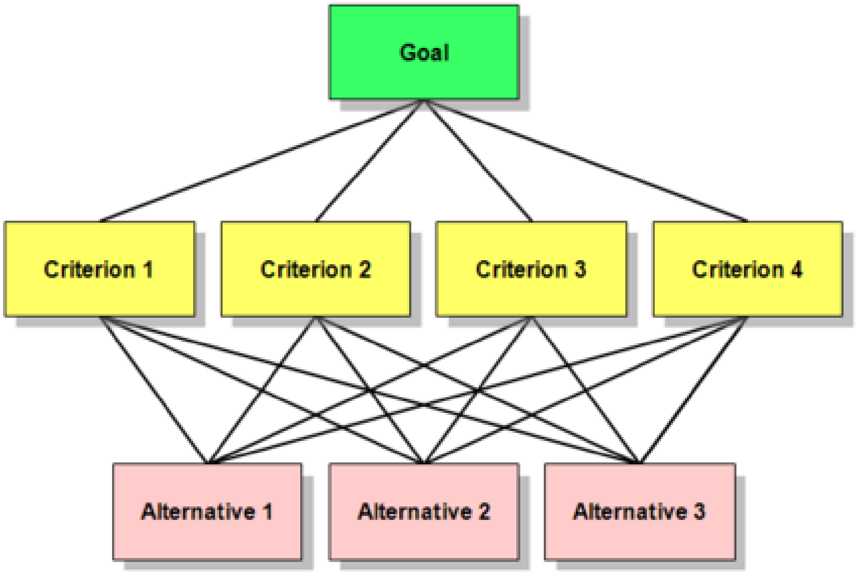


Figure 4.2: *Analytic Hierarchy Process Overview.*

4.3 User Ranking Algorithm

Algorithm 1 outlines the user ranking algorithm, which we have implemented in NavigTweet, as discussed in next chapter. The algorithm calculates a ranking on the basis of both tweet-level and user-level influence parameters. As an input, the algorithm takes a Twitter user node N, and object U provided by Twitter API, and, as an output, it provides a final ranking value of U.

The *UserBasedScore(u)* method provides a score value of user-level parameters and returns a user-level score value. Similarly, *TweetBasedScore(u)* method provides a score value of tweet-level parameters (last 200 fetched-tweets) and returns a tweet-level score value. After scoring each user, the algorithm provides a ranking value for each user by sorting all users based upon the score value.

Algorithm 1: User ranking algorithm of NavigTweet.

Data:

N = Twitter user node;
U = Twitter user object, retrieved from Twitter API;
(AHP based Weight Vectors)

CONSTANT $W_{parameter}$ as DOUBLE;
Input : (N, U)
Output: Final Ranking value (Score) of each node $n \in N$.

begin
  function $UserBasedScore(u)$:= **do begin**
  (User-level influence parameters ranking)
  (Product sum of weight and values)
  $DOUBLE\ d \leftarrow \sum(W * U.Value)$;
  $N.userRank \leftarrow d$;
  return d;
  end do
  function $TweetBasedScore(u)$:= **do begin**
  (Tweet-level influential parameters ranking)
  for $i \Leftarrow 1 \rightarrow 200$ **do**
    (Summing up values for each parameter)
    $DOUBLE\ f \leftarrow \sum(W * U.Value)$;
    $N.tweetsRank \leftarrow f$;
    return f;
  end
  foreach $u \in U$ **do**
    $u.AHPScore = UserBasedScore(u) + TweetsBasedScore(u)$;
  end
  (Descending sort of nodes by their AHP Score)
  (assign i^{th} indexed value as node's AHP Rank)
end

CHAPTER 5

NavigTweet: Visual Exploration of Twitter Network

This chapter describes the architecture and implementation aspects of the tool - NavigTweet [60, 82], which has been developed as a part of our research. We discuss the application overview, objectives, architecture, building blocks, and, finally, the implementation results.

5.1 Application Overview

NavigTweet is a novel influence-based visualization tool for exploring Twitter network. Twitter users can visually explore and interact with their own network and as well as their friends' network, i.e. Friend of a Friend (FOAF) network [27, 71, 124]. NavigTweet provides a way to visually identify the Social Influencers (prominent users) within their friends' networks by means of an influence-based ranking mechanism.

The intended audience is people who wish to find interesting information regarding their Social Network and enlarge their social network by identifying interesting people to follow. The intended audience may find influencers within their social networks through NavigTweet by exploring friend-of-a-friend

(FOAF) relationships. The main functionalities that the tool offers are:

- A visual interface to explore Twitter network.
- Influence-based detection of Influencers or prominent users among Twitter users.
- A platform to access the interesting information regarding influencers in a graphical form.

The salient features of the application are summarized in Figure 5.1.

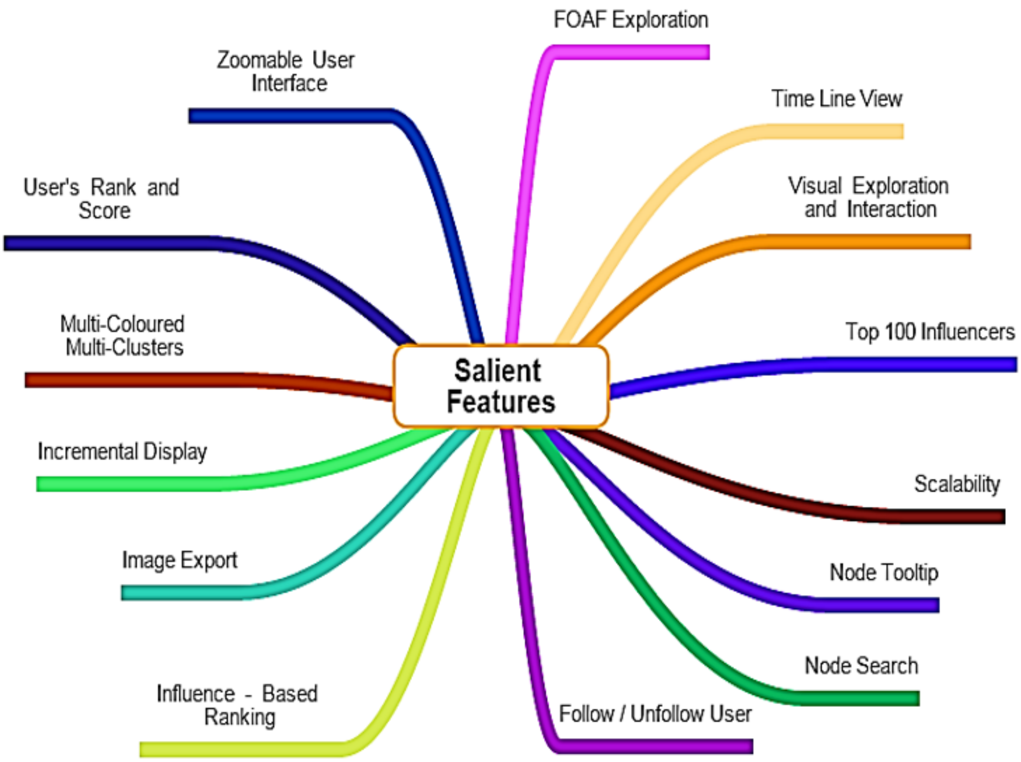


Figure 5.1: *NavigTweet's Salient Features*

5.2 Application Objectives

NavigTweet aims to provide a visual interface to interact and explore the Twitter network. It helps to identify the key players or prominent Twitter users, and follow them directly through application interface.

The main objectives of NavigTweet, are the following:

- To provide a visual interface to explore Twitter network and, particularly the FOAF networks.
- To provide influence-based rank for Twitter users and to help users to identify prominent users in Twitter network.
- To provide aesthetically pleasant layout for generated network graph.
- To directly add *follow* relationships on Twitter through application interface.

5.3 Application Architecture

The work-flow of NavigTweet is provided in Figure 5.2. The basic work-flow steps of the application are the following:

1. **Twitter Authentication**: NavigTweet uses the OAuth protocol for Twitter user authentication, using the Pin-based mechanism provided by Twitter APIs. This module is responsible for handling user authentication for successful login.
2. **User Node**: After successful login, the application creates a user node on the graph canvas, corresponding to the user who has logged in.
3. **Twitter Data Streaming**: This module is responsible for fetching the data of a user's friends. Due to the rate-limit of Twitter APIs, we fetch a maximum number of 500 friend IDs and 100 User objects in one API call.
4. **Graph Model Processing**: This module creates nodes and edges for parsed friends on the graph canvas. As a result, a local neighborhood cluster of friends' nodes around a user's node is created on graph canvas.
5. **AHP-Based Ranking**: This module provides each node's AHP-based score and rank, by using both user-level and tweet-level influence parameters provided by Twitter API, as shown in Figure 5.2. Due to the rate-limit of Twitter APIs, NavigTweet fetches the last 200 tweets of users in order to calculate their tweet-level rank.
6. **Graph Controller**: Finally, this module handles event related functionalities (e.g. mouse double-click event) and applies the power-law based graph layout. Whenever the user double-clicks on any node, the application repeats from step 3 and fetches the friends of the node on which the user has double clicked.

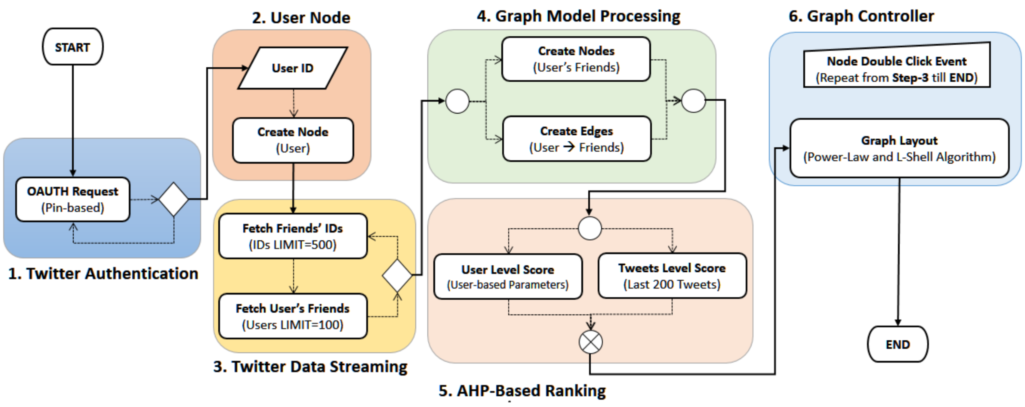


Figure 5.2: *NavigTweet Work-flow.*

5.4 Basic Building Blocks of NavigTweet

The basic building blocks of NavigTweet are the graph layout algorithm, the content-based user ranking methodology and the ranking algorithm. These modules are briefly described in the following.

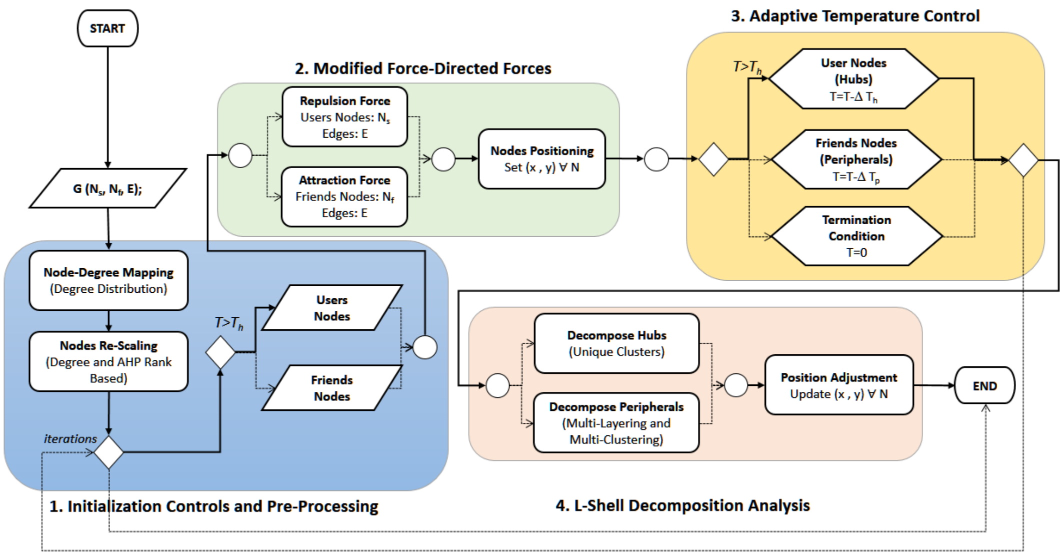


Figure 5.3: *Power-Law algorithm Work-flow.*

5.4.1 Power-Law Algorithm (Graph Layout Technique)

In order to draw the Twitter network in an aesthetically pleasant way, NavigTweet uses a modified force-directed graph layout, as presented earlier in [58, 61]. The general work-flow of the power-law algorithm is presented in Figure 5.3.

The proposed approach is aimed at the exploitation of the power-law degree distribution of user nodes (N_s). Provided that the distribution of the degree of the nodes follows a power law, we can partition the network into two disjoint sets of vertices N, i.e. the set of Twitter users' nodes N_s, and the set of friends' nodes N_f, such that $\mathrm{N} = \mathrm{N_s} \cup \mathrm{N_f}$ with $\mathrm{N_s} \cap \mathrm{N_f} = \phi$. Algorithm 2 provides a high-level overview of the whole algorithm by showing its main building blocks.

Algorithm 2: High-level power-law algorithm of NavigTweet.

Data:
N_s = Twitter user node: $\{v|v \in V, v \notin N_f\}$;
N_f = Twitter user's friend node: $\{w|w \in W, w \notin N_s\}$;
$|V|$ = Number of vertices: $|N_s| + |N_f|$;
A = $W * L$ (W and L are the width and length of the graph canvas.);
k = Optimal distance between vertices: $\sqrt[c]{A/\,|V|}$;
c = experimentally found constant;
x = distance between two vertices;
f_a = Attractive force: $x^2 k$;
f_r = Repulsive force: $-k^2/x$;
E = Edges connecting Twitter users and friends;
$G(N_s, N_f, E)$ = Bipartite graph containing two disjoint vertices and edges;
d = Degree of node representing the number of edges connected to it;
l_s = l-shell index value representing node's location according to successive layers (l-shells) on the graph canvas;
C_n = Cluster index value representing cluster where node n belongs. Nodes having same ls value belongs to same cluster with distinct value C_n;
T = Energy / Temperature variable;
T_h = Temperature threshold, to control simulation;
CONSTANT MAX_DIM as INTEGER;
CONSTANT MIN_SIZE as INTEGER;
Input : $G(N_s, N_f, E)$ with initial random placements of nodes on graph canvas
Output: $G(N_s, N_f, E)$ with a multi-clustered hierarchical layout with multi-layered peripheries

begin
  (To avoid local-minima and convergence issues, repeat till iterations)
  for $i \leftarrow 1$ **to** $iterations$ **do**
    if $T > T_h$ **then**
      (calculate f_a and f_r for N_s with maximum T value) **call function** $repulsiveForces(N_s)$;
      call function $attractiveForces(E)$;
    end
    else
      (Calculate f_a and f_r for N_f with relatively low T value.)
      call function $repulsiveForces(N_f)$;
      call function $attractiveForces(E)$;
    end
    (limit the maximum displacement to the temperature t)
    (and then prevent from being displaced outside frame A.)
    call function $nodeDegreeMapping(V)$;
    call function $LShellDecompose(N_f)$;
    call function $NodesPlacement(V)$;
    call function $coolDown(T)$;
  end
  call function $resetNodesSizes(V)$;
end

Repulsion Force

It pulls the nodes away from each other. The underlying physical model is that of the Coulomb's law, which assumes that two nodes are electrically charged with electric loads of the same sign and thus repel each other with a force $f_r(z) = k^2/x$, with k a real positive number representing the characteristic constant of the electric field. We have set $k = 100$ for our experimental analyses. The detailed steps of this method are presented earlier in [58, 61].

Attraction Force

It attracts nodes if they are connected by an edge E. The underlying physical model is that of the Hooke's law [[110]], which assumes that the edge between the two nodes behaves like a spring that is attracting the two nodes with a force x^2/k, where k is a real positive number representing the spring's characteristic constant. We have set $k = 40$ for our experimental analyses. The detailed steps of this method are presented in [58, 61].

Node-Degree Mapping

This method simply calculates all nodes' degrees and sorts them by creating a degree distribution. It provides a sorted map of nodes with their degrees.

L-Shell Decomposition

This method creates multi-layered peripheries of N_f nodes around nodes N_s. It assigns each node a unique l_s and C_s value, based upon degree. Algorithm 3 presents overall structure of this method.

Algorithm 3: L-Shell Decomposition Method.

```
begin
    C_s = 1;
    for v ∈ N_s do
        v.l_s = C_s; C_s = C_s + 1;
    end
    for w ∈ N_f do
        ∀e ∈ w.E do begin:=
        if ∃e.N_s.C_s ∈ e.N_f.ClustersMap < v, C_s > then
            increment e.N_f.l_s;
        end
        else
            w.l_s = 1;
            add w.l_s ∈ e.N_f.LevelsMap < v, l_s >;
            add e.N_s.C_s ∈ e.N_f.ClustersMap < v, C_s;
        end
    end
end
```

Node Placement

This step is responsible for placing nodes on canvas according to the force model. The detailed description of this method can be found in [58, 61].

Temperature Cool Down

This step is responsible for cooling down the system temperature, in order to make the algorithm converge. We introduce a customized dynamic temperature cool down scheme, which adapts the cool down step based on the current value of the temperature. As shown in Figure 5.4, the temperature is supposed to be initialized at a value T_{start}, and then to be reduced by a variable cool down step Δt based on the current value of the temperature itself. This approach provides a convenient way to adapt the speed of iteration of the algorithm to the number of nodes to be processed. While processing user nodes N_s (a few), the temperature decreases slowly; while processing peripheral friends' nodes N_f (many), the temperature decreases more rapidly to avoid expensive computations for nodes that are not "central" to the overall graph layout. The reference temperature value T_c is used as convergence threshold, i.e. when the temperature reaches that point the iteration is stopped.

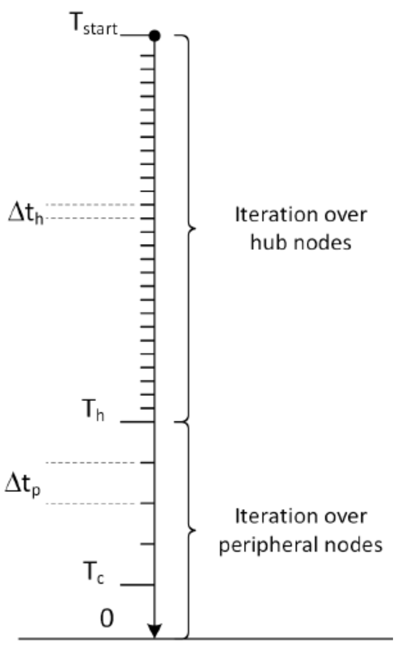


Figure 5.4: *Adaptive temperature cool down mechanism.*

Algorithm 4 presents the general overview of the temperature cool down scheme. Variables Δt_h and Δt_p may be parameterized to adapt the algorithm behavior to properly fit the requirements given by the context of analysis. The values that we used for our experimental analyses are $\Delta t_h = 0.0005$ and $\Delta t_p = 0.05$.

Algorithm 4: Temperature Cool down

begin
  if $Temperature > T_h$ **then**
    $Temperature = Temperature - \Delta t_h$;
  else
    $Temperature = Temperature - \Delta t_p$;
  end
  if $Temperature \leq T_c$ **then**
    $Temperature = 0$;
  end
end

Reset Nodes Sizes

This method is responsible for rescaling the size of each node in the graph, based on their degree and AHP rank. The higher the degree and rank of a node, the greater the size and vice versa. Algorithm 5 outlines the detailed steps of this method.

Algorithm 5: Nodes Reset Method

begin
  for $v \in V$ **do**
    $part = (v.TotalFriends - v.Rank) + 1$;
    $size = (part * (MAX\_DIM/v.TotalFriends)) + MIN\_SIZE$;
    $v.size = size$;
  end
end

5.4.2 User Ranking Methodology

The ranking methodology that we have adopted in NavigTweet is summarized in Figure 4.1. NavigTweet initially collects influence parameters, at both user-level and tweet-level. To weigh different parameters, based upon their relative importance, we have adopted the Analytical Hierarchy Process (AHP) method proposed by [135, 136]. As explained in section 4.2.1 The outcome of AHP is a vector of weights of parameters. NavigTweet provides an aggregated score for each user, as a weighed sum of different parameters using the weights obtained from AHP. The higher the score the higher the rank, and vice versa. Figure 4.1, presented earlier in Section 4.2, summarizes the ranking methodology implemented by NavigTweet.

5.4.3 Influence-based User Ranking Algorithm

Algorithm 1 shows the outlines the ranking algorithm adopted by NavigTweet. The algorithm calculates the ranking on the basis of both tweet-level and user-level influence parameters, as mentioned earlier in Section 4.1.

5.5 Visual Elements

The main visual elements of NavigTweet are the color-scheme, the graph layout and the node tool-tip. These are described in next sections.

5.5.1 Color Scheme

NavigTweet uses a node color-scheme to distinguish different types of nodes (see excerpt in Table 5.1). There are two types of nodes, currently selected user nodes and the friend nodes. The nodes with a higher influence according to AHP ranking are red with green bold stroke. Similarly, selected users' nodes are represented in blue color with white thin stroke and friend nodes are represented by any random color other than red and blue (with think stroke). For more clarity, we applied distinct node-stroke colors with variant width (wide stroke for influencer nodes, and thin stroke for other nodes).

Table 5.1: *Excerpt of NavigTweet node color scheme.*

Type	Color	Stroke
Selected user	Blue	White and Thin
Influencer	Red	Green and Thick
User's friends	Random	Brown and Thin

5.5.2 Graph Visualization

In order to create an aesthetically pleasant layout in the multi-clustered and multi-layered peripheral network, we apply the power-law based modified force-directed algorithm, discussed earlier in Section 5.4.1. This algorithm tends to arrange nodes in such a way that highly connected user nodes are placed in a more central position while the less-connected friend nodes are placed in the periphery around their user node. In this way, each node has its own cluster of multi-layered friend nodes. Graph layouts generated with this technique are usually perceived as aesthetically pleasant, since all edges have roughly the same length and tend to avoid edge crossings.

Figure 5.5 shows a small graph containing two different clusters of nodes around nodes (1)) and (2). NavigTweet identifies common friends, if any, who are connected to more than one user (e.g. node (4) in Figure5.5). In the example, the top user with the highest rank is node (3).

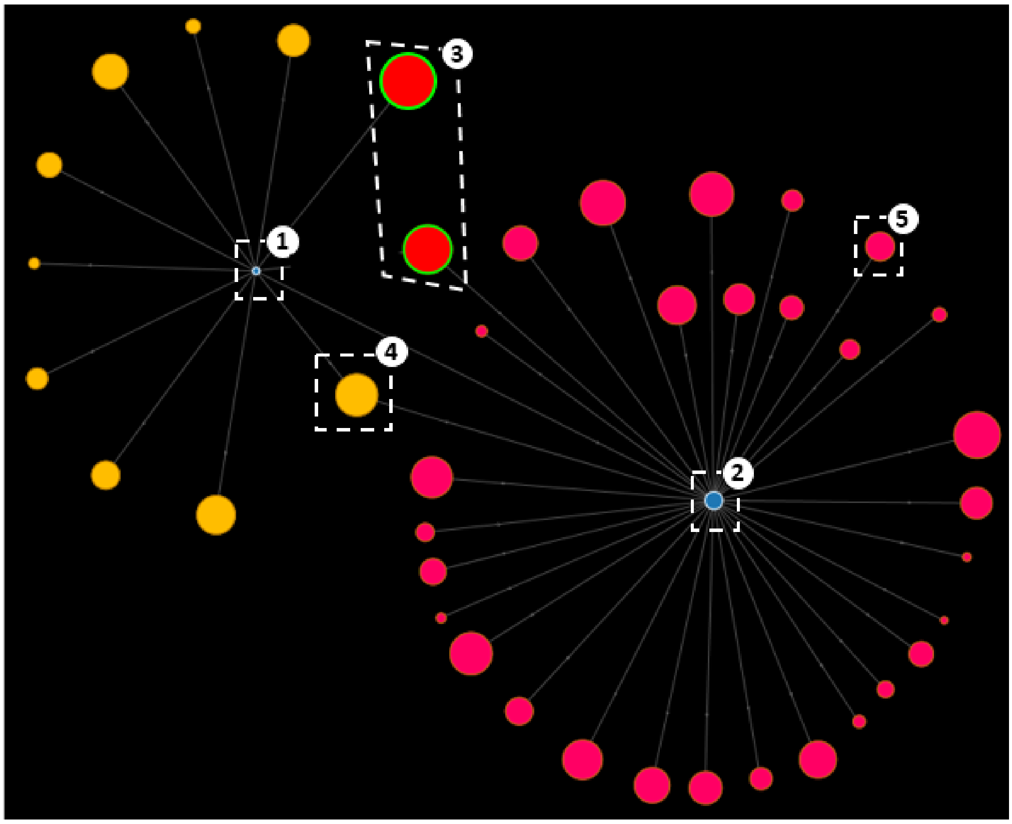


Figure 5.5: *Sample network graph.*

In the graph, the size of nodes varies, as shown in Figure 5 and Figure 5.6, based upon its relative rank within the local neighborhood cluster of friends. The higher the node's rank, the larger the node size, and vice versa.

5.5.3 Node Tool-tip

NavigTweet provides information for all nodes in a tool-tip. When the user brings the mouse over a particular node, the tool-tip shows the node information along with its rank, see node (1) in Figure 5.6. Additional user profile information such as photo, screen name, location etc. is also displayed in the tool-tip.

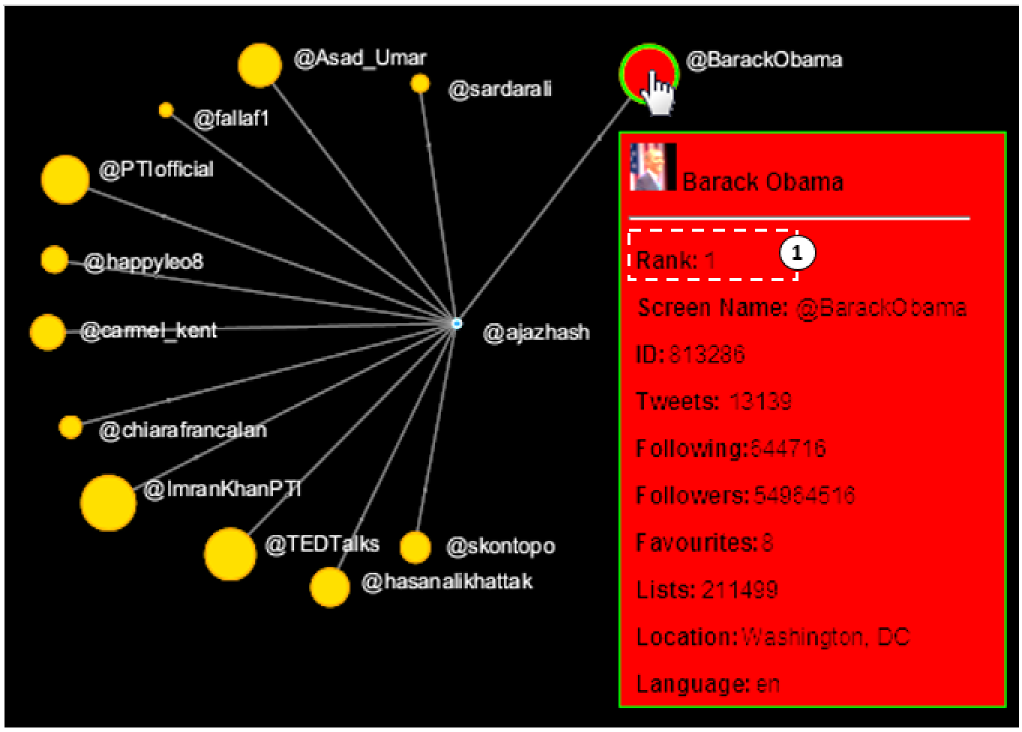


Figure 5.6: *Varying nodes' sizes and node tool-tip.*

5.6 Implementation and Results

We have implemented NavigTweet as a desktop application. The application is written in JAVA using Twitter4j [153] – a JAVA-based library, and Piccolo 2D [20] – a JAVA based 2D Graphic library. Figure 5.7 represents the main components of NavigTweet. The application has a GUI compatible with multiple operating systems (Windows, MAC OS, and Linux/Unix) and contains a runnable JRE file. The only pre-requisite of NavigTweet is the JAVA Runtime Environment. During installation, the setup will automatically install the JRE Bundle package, if missing. NavigTweet uses OAuth-based protocol for user authentication provided by Twitter API. The OAuth protocol allows Twitter users to approve the application and allow it to act on their behalf without sharing their password. Then, NavigTweet can require an Access Token from Twitter. This initial configuration is a one-time process. Further details can be found on NavigTweet website [82].

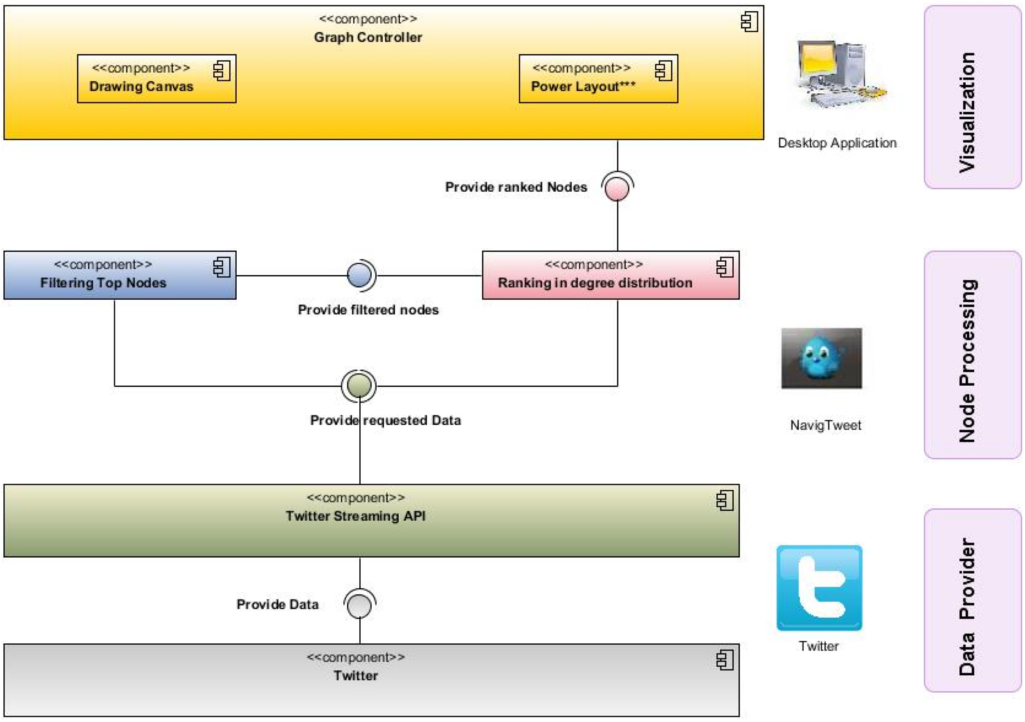


Figure 5.7: *NavigTweet Components.*

5.6.1 Application Interface

Figure 5.8 shows the main screen of NavigTweet. The user interface consists of three panels: left, center and bottom. The left panel shows the influencers, as well as Twitter and control options. The center panel shows the graph canvas, where the user can explore and interact with the graph. The bottom panel provides the timeline and console panes for the currently selected node.

Left Panel

The *Left Panel* provides three further sub-panels: *Influencers Panel*, *Twitter Panel* and *Control Panel*. The *Influencers Panel* is dedicated to show both graph-level and user-level top-20 influencer list, as shown in Figure 5.8. A user can directly follow or un-follow Twitter users from the top-20 influencer list. The *Twitter Panel* displays the user's timeline and provides a bird-eye view of the whole graph. Moreover a user can post tweets directly to his or her timeline and can also send direct message to his or her followees. The *Control Panel* provides

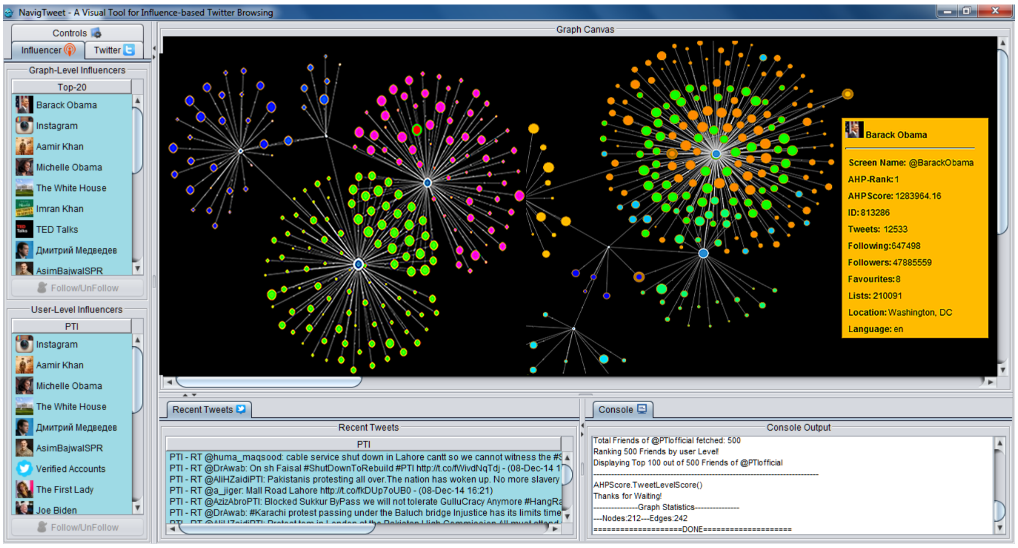


Figure 5.8: *Main Screen of NavigTweet Interface.*

various button controls for print, data export, and show / hide node labels. It also allows searching any node in the network graph.

Center Panel

The *Center Panel* of NavigTweet shows the graph canvas, where a user can explore and interact with the graph. Users can smoothly zoom in to explore the whole network topology. They can pan the background and move elements around via drag and drop, to further optimize the graph visualization and adapt it to their needs. Whenever a node is dragged and released, the rest of the nodes are repositioned with animated transitions according to the power-law based force-directed algorithm. When a user double-clicks on any node, the application fetches the node's friends in real-time and shows them on the graph.

Bottom Panel

The Bottom Panel of NavigTweet displays the most recent 20 tweets of the currently selected user. When a user selects a different node, the recent tweets of new selected user will be dynamically displayed.

5.6.2 Application Functionalities - Summary

The main functionalities of NavigTweet are classified into four different categories based upon the type of functionality. These functionalities are outlined in

Table 5.2.

Table 5.2: *Main functionalities of NavigTweet.*

Categories	Features
User Profile Management	• User authentication. • OAuth protocol • Access token generation for login. • Profile information access
Interaction with Twitter	• Follow / Unfollow user. • Display friend network graph. • View timeline. • Post a tweet. • Explore social network at any depth. • User search. • Top-20 user-level influencers (i.e. influencers that are selected among any node on canvas.). • Top-20 graph-level influencers (i.e. influencers that are selected among users connected with a followee relation with currently selected users). • View user analytics. • Send direct messages

Influence-based Social Network	• Perform AHP-based ranking of each user. • Show mutual-follower(s). • Browse FOAF network. • Identify top-100 influencers among each user's browsed network
Interface and Controls	• Zoomable user interface. • Node tooltip. • Show/Hide node labels. • Bird's Eye View of Graph Canvas. • Print Graph. • Apply Power-Layout. • Console Output/Log. • Multi-Color Clusters. • Export Data (CSV). • Mouse Events (drag, scroll, over, click, etc.).

5.6.3 Application Configuration

Twitter uses OAuth Authentication Protocol to provide authorized access to its APIs. Hence, NavigTweet uses Twitter's OAuth authorization protocol that allows Twitter users to approve application to act on their behalf without sharing their password. Using the OAuth protocol, NavigTweet requires the Access Token from Twitter and saves it locally for future access. Figure 5.9 explains this protocol.

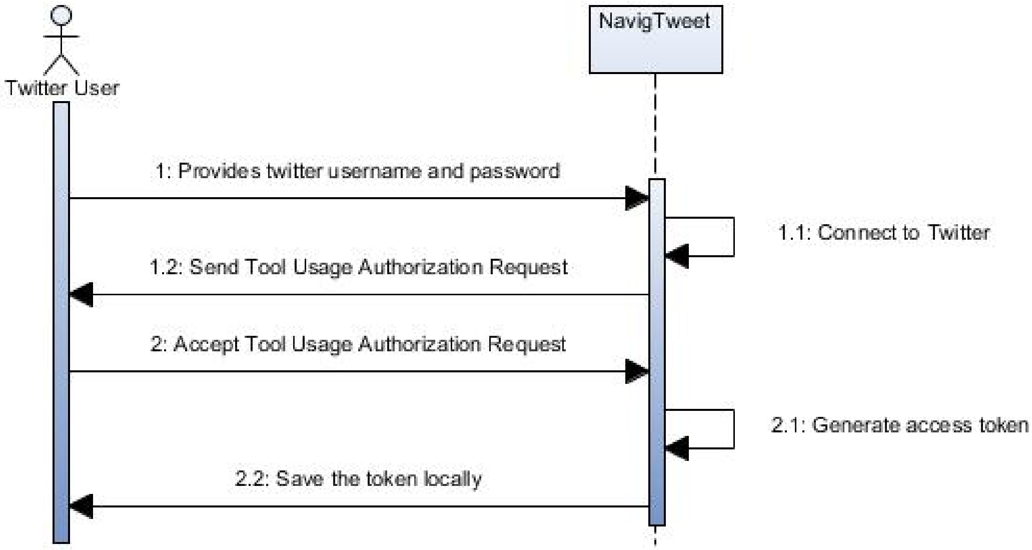


Figure 5.9: *NavigTweet OAuth Work-flow.*

5.6.4 Comparison with Existing Applications

As noted in Section 2.5, there exist a few visualization tools supporting the explanation of Twitter's network. Twitter changes its APIs periodically, which enforces developers to continuously update their tools. This represents the main reason why the number of tools is practically limited. Table 5.3 shows the highlights of the comparison between NavigTweet and the tools that we have been able to test. We have considered various features in the comparison covering main requirement types, related to performance, service, support, interface, and scalability.

Table 5.4 provides a more qualitative analysis of the usability of literature benchmark tools, including NavigTweet. We have considered several factors related to aesthetics, such as color-scheme, distance between nodes, amount f information, zoom-ability of graph canvas, node shapes, mouse controls, etc.

Table 5.3: *Quantitative Comparison between NavigTweet and similar Twitter explanation tools.*

Criteria x Tool	TouchGraph	MentionMap	InMaps	NavigTweet
Real-time	No	Yes	No	Yes
Graph depth	1	2	1	n-level
Response time	<5s	<5s	>5s	>5s
Initial load time	>5s	<5s	<5s	>5s
Open source	No	No	No	TODO
Pre / Freemium	Both	Freemium	Freemium	Currently Freemium
Social Network	Facebook	Twitter	LinkedIn	Twitter
Platform	Web	Web	Web	Currently Desktop
Help and support	Feedback	Feedback	FAQ/Feedback	Feedback/Tutorial

Table 5.4: *Qualitative Comaprison between NavigTweet and similar Twitter explanation tools.*

Criteria x Tool	TouchGraph	MentionMap	InMaps	NavigTweet
Network browsing	Self & Others	Self & Others	Self	Self & Others
Friendly colors	Somehow	Yes	Yes	Somehow
Clusters clarity	Yes	Yes	Somehow	Yes
Multi-color cluster	No	No	Yes	Yes
Zoom-able Interface	Yes	Yes	Yes	Yes
Pan & drag	Yes	Yes	Yes	Yes
Information quantity	A lot	Normal	Normal	Normal
Information placement	ToolTip	None	ToolTip	Tooltip
Default information	Name + Photo	Name + Photo	Name	Screen Name
Node shape	Circular	Rectangular	Circular	Circular

CHAPTER 6

NavigTweet: Pilot Test Execution and Results

This chapter describes the results of a pilot test and subsequent large-scale test of NavigTweet. A pilot is a test of a proposed product or service with the intended user base. A pilot can be used to:

- Determine the feasibility of a proposed product prior to the official launch.
- Identify faults in the developed product to help determine if it is ready for general release.

This procedure ensures that new application is tested thoroughly before being released to the public users. The document analyses the opportunity to conduct a pilot project, defines the main characteristics of the pilot and identifies the activities, schedule, and feedback from interviewees. Then, the chapter discusses the large-scale test.

6.1 Scope of Pilot

In this section, we provide an overview of the functionalities of NavigTweet. Further detailed application features are also discussed in the requirement analysis section. Finally, we discuss the expected outcomes and risks associated with the pilot activity.

6.1.1 Application Features

The pilot activity targets the salient features of NavigTweet, as provided in Table 6.1. The pilot activity is intended to verify the usability of NavigTweet and the degree of user satisfaction.

Table 6.1: *Salient Features of NavigTweet.*

Feature Code	Feature	Description	Functionality Type
$[F_1]$	Visual Exploration and Interaction	NavigTweet will provide a visual interface to the user to explore and interact with his/her Twitter network.	Core
$[F_2]$	Aesthetically Pleasant Graph Layout	NavigTweet will layout the generated graph in an aesthetically pleasant way for user's better understanding.	Ancillary
$[F_3]$	Users Profiling	NavigTweet will access and manage user's profiles through Twitter API via secure authentication protocol.	Core
$[F_4]$	Browse Friends and FOAF network.	User will be able to explore not only his own friends' network but FOAF network to find prominent users.	Core
$[F_5]$	Identification of Influencers within network.	NavigTweet will identify the key influencers among the network, which user will browse through his friends' network.	Core
$[F_6]$	Direct Interface with Twitter.	NavigTweet will also provide a direct interface to the user to access his/her timeline, to post tweets, to send private messages to his/her followers, and to directly follow/unfollow any user in his/her network.	Ancillary

$[F_7]$	User Level Access	NavigTweet will be able to access any user's profile data, which is available through Twitter API, in order to get user-level parameters, like no. of followers, followees.	Core
$[F_8]$	Tweets Level Access	NavigTweet will be able to access any user's tweets, in order to fetch parameters like no. of urls, hashtags, mentions etc.	Core
$[F_9]$	Influence-based Users' Ranking	NavigTweet will provide rank of each user, which will be calculated on the basis of influencing parameters like no. of retweets, no. of followers, no. of hashtags, etc.	Core
$[F_{10}]$	Multi-Control User Interface	Navigtweet will provide various options, like zoomable user interface, node-search, export data, print graph, and mouse events.	Ancillary

6.1.2 Requirements Analysis

The pilot activity has been carried out in order to make sure that each functional and non-functional requirement of the product is being tested. During our analysis, we categorized the detailed set of both functional and non-functional requirements into main entities, as shown in Table 6.2. For each category, we list functional requirements, mentioned as $[FR_i]$ and non-functional requirements, mentioned as $[NFR_i]$.

Table 6.2: *Requirement Analysis of NavigTweet.*

Code	Requirement	Description	Functionality Type
User Profile Management			
$[FR_1]$	User Authentication	Secure authorized access via OAuth protocol provided by Twitter API.	Core

$[FR_2]$	Access Token Generation for Login.	For Login, Pin code generation through OAuth protocol provided by Twitter API.	Core
$[FR_3]$	Profile information access	Access of user-level and tweet-level information, permissible information access through Twitter API.	Core
Direct Interface with Twitter			
$[FR_1]$	Follow / Unfollow User	Directly follow / unfollow any user in network.	Ancillary
$[FR_2]$	Display Friends' network graph.	To display Friends' network cluster on graph canvas.	Core
$[FR_3]$	View Timeline	View recent tweets	Ancillary
$[FR_4]$	Post a Tweet	To post a tweet directly through application interface.	Ancillary
$[FR_5]$	Explore Social network at any depth.	To explore any FOAF network.	Core
$[FR_6]$	User Search.	To search any user on graph canvas.	Ancillary
$[FR_7]$	Display top-20 User-Level Influencers	To show panel containing top-20 influencers of selected user.	Core
$[FR_8]$	View User analytics	To show all available information of the user.	Core
$[NF_1]$	Send Direct Messages	To send private message to the followers.	Ancillary
Influence–based Social Network			

$[FR_1]$	Perform Ranking of each user	To calculate each node's rank and score by influence-based parameters.	Core
$[FR_2]$	Show Mutual-Follower(s)	To show mutual Follower(s) on graph canvas.	Core
$[FR_3]$	Browse FOAF network.	To explore FOAF network of any user.	Core
Interface and Controls			
$[FR_1]$	Zoom-able User Interface	The graph canvas should be zoom-able	Core
$[FR_2]$	Node Tool-tip	Node tool-tip to show user-related information and user's rank.	Core
$[FR_3]$	Show/Hide Node Labels	Node's screen name.	Ancillary
$[FR_4]$	Bird's Eye View of Graph Canvas.	To show abstract view of whole graph canvas.	Ancillary
$[FR_5]$	Print Graph.	To print/save graph as PNG image.	Core
$[FR_6]$	Apply Power-Layout	To apply power-layout on generated graph to adjust it further for ease.	Ancillary
$[FR_7]$	Console Output/Log	To show Console output/log panel, in order to monitor activity.	Ancillary
$[FR_8]$	Multi-Colour Clusters	Distinct colour of nodes for each cluster.	Core
$[FR_9]$	Real-Time social networks.	To show real-time networks of each selected user.	Core

$[NF_1]$	Load time of any user's network $< 5s$	To take minimal time to load each selected user's friend network.	Core
$[NF_2]$	Initial Load time $< 5s$	After user-login, canvas should appear within no time.	Ancillary
$[NF_3]$	Exception handling on Twitter Rate-Limit Issue.	Exception throw when Twitter's rate-limit exceeds.	Core
$[NF_4]$	Double-click Event	Double-click the node to load it's friends' network.	Core
$[NF_5]$	Mouse Drag and Over events.	Nodes should be dragged by mouse, and by Mouse-over tool-tip should appear.	Ancillary
$[NF_6]$	Application Installer	The installer application should be easy and provide guided interface.	Ancillary
$[NF_7]$	Multi-OS Support	Application should run on Windows and MAC.	Core

6.1.3 Risks

The scope of the pilot activity clearly states which functions, requirements, and features will be included. All the functions, requirements, and features listed are included in pilot. The expected outcome is the fulfillment of all functional requirements of the application in a normal scenario. The following are a few risks and contingency plans associated with the application, which may arise during the pilot.

- **JRE Dependency:** The application has been built by using JAVA technologies, so in order to run the application on a client machine, it is necessary to have JRE installed. The application installer has the capability to detect the pre-requisite and will auto install JRE, if missing.

- **Twitter Rate-Limit Exceed:** Twitter imposes constraints to fetch user data. The application provides strong exception handling mechanisms, if any rate-limit is exceeded. The application allows users to wait, until Twitter resets. The application will auto-resume after Twitter's rate-limit resets.

- **Loss of Internet Connection:** When users run the application, the application itself checks the availability of an Internet connection, but it doesn't continuously monitor the Internet connection status. So, if a user's Internet connection is interrupted during the application execution, then the user will not be able to load more nodes on the graph.

6.2 Pilot Objectives

The objectives help us to identify criteria for measuring the success of pilot activity. Following are the objectives of the pilot.

- Ensure that the application meets the general expectations of target users.
- Ensure that the application has advantages over existing applications, if any.
- Ensure that the application adds value to the user.
- Test the deployment process.
- Ensure that all functional and non-functional requirements meet user expectations.
- Gather further functional/non-functional requirements from users.
- Test the application installation.
- Gather user reviews of existing functionalities.
- Check application reliability and supportability constraints.
- Check real-time estimated results and associated risks, if any.
- Verify the user attention and interest towards the application.
- Ensure that the application performs as per expected outcome in a real-time environment.
- Understand the user interest in the application, e.g. to know whether the application is interesting, informative, innovative, usable or not.
- Ensure the effectiveness of the application in exploring Twitter network.

- Ensure the clarity.
- Measure the level of user satisfaction.
- Ensure that the end-user can identify social influencers among his/her Twitter network, through browsing hop-to-hop networks.
- Gather further improvement suggestions and feature updates, if any.

6.3 Pilot Team

Initially, we targeted a reference group of 8 people from academia, who are expert specific research areas. All participants were familiar with the idea of graph visualization and/or had some knowledge in related topics, such as social network analysis, Social influencers or Twitter. One participant never used Twitter, but was expert in the domain of graph theory and graph visualization. Another participant was also not familiar with social network analysis or influencers, but had expertise in algorithms, software architecture and web technologies. Table 6.3 provides details about pilot participants. The participants could interact with NavigTweet implementation themselves, while they were gradually told about interactive features. We intended to demonstrate the application in a real-time environment, to gather their feedback about the application.

Table 6.3: *Pilot Participants*

Resource	Role	Research Line
Participant 1	Full Professor	Information System
Participant 2	Full Professor	Dynamics of Complex Systems
Participant 3	Full Professor	Graph Theory, Information Visualization
Participant 4	Associate Professor	Data, Web and Society
Participant 5	Associate Professor	Information Systems
Participant 6	Associate Professor	Data, Web and Society
Participant 7	Associate Professor	Information Systems
Participant 8	Associate Professor	Advanced Software Architecture

6.3.1 Responsibilities

The responsibilities of pilot resources are listed below:

- Testing application installer.

- Application pre-requisites review.
- Application installation testing.
- Configuration testing.
- Functional unit testing.
- User acceptance testing.
- Feedback review survey.
- Application website review.
- User guide review.
- OS compatibility test (MAC, Windows, Linux).

6.4 Pilot Tasks

In this section, we describe the pilot summary from each participant, by highlighting the key-findings and their remarks against each requirement category as listed below. Each participant was involved in testing application features and provided his/her remarks, which we will discuss in this section. The general categories for application interfaces or set of requirements, which are also discussed in Section 6.1.2, are as follows:

- **User Profile Management**: The set of requirements which fall under this category are related to application interfaces like *user-authentication*, *OAuth based Token generation* and *user profile information*.
- **Direct Interface with Twitter**: The interfaces like *posting a tweet*, *show user-time line*, *send direction messages* to the user, and *top-20 influencers* fall under this category.
- **Influence-based Social Network**: The set of requirements which fall under this category are related to *ranking of users*, *identification of mutual follower(s)*, and *browsing of FOAF networks*.
- **Application Control Interfaces**: The control options or features like *node tool-tip*, *ZUI*, *Bird's eye view* [137], *Print Graph* and *Export data* fall under this category.

We questioned each participant about the application features and we recorded their responses. They tested each feature of the application in a real-time environment and recorded their run-time feedback. Table 6.4 provides the pilot summary response from all participants. The general categories of application requirements were taken from Table 6.2.

Table 6.4: *Summary of Pilot Responses.*

<table>
<tr><td>FIXED TESTS</td><td>Application functionality is acceptable.
Application performs with acceptable speed and response.
Application interface is user-friendly.
Browse FOAF networks is useful.
Notify any Twitter Rate-Limit Exceed Exception is useful.
Follow / Unfollow user in real time is effective.</td></tr>
<tr><td>SUMMARY</td><td>Each participant randomly tested every feature of the NavigTweet. The test response was satisfactory. Summary about each set of requirements is as follows:
User Profile Management:
Successful user authentication via OAuth protocol, along with request token generation. After successful login, user profile was accessible.
Direct Interface with Twitter:
Successful working of each feature, Top-20 Influencers, Time-Line view. Tweet posting was speedy and responsive.
Influence-based Social Network:
User ranking and score found to be accurate. Mutual friends also been found by exploring FOAF networks.
Interfaces and Controls:
All application interfaces worked perfectly, no Twitter API related exception thrown by application. Tested graph print feature as well.</td></tr>
<tr><td>END REMARKS</td><td>Most Features liked:
Power-law based graph drawing technique.
Graph animation.
Users relative ranking.
Top-20 Influencers panel.
Interface with Twitter.
Nodes' color-scheme.
Recent Tweets Panel.
Node's tool-tip.
Improvements / suggestions:
Graph Nodes' Legend panel.
Display Top-20 Graph-level influencers panel, to highlight influencers.
Export graph data into CSV.
Graph Nodes' Legend.
Follow / unfollow any user at run-time.
Web-Interface of NavigTweet.</td></tr>
</table>

6.5 Pilot Phases

The methods through which we assessed the quality of NavigTweet are described below:

6.5.1 Phase 1: Face-to-Face Interviews

During the pilot, we have performed one-to-one, face-to-face interviews. We had the opportunity to brief the interviewees about the application scenario, installation, and application flow. With each participant, we obtained real-time feedback by running the application. The discussion sessions with each participant took around 1–1.5 hours. During each session, we tested the application thoroughly. To provide some rough guidance through the features of NavigTweet and ensure touching upon a wide range of visualization aspects, a number of questions were asked. Participants were asked these questions to give them an incentive to look at NavigTweet features and aspects of the visualization and interface, and to induce suggestions on missing or desired features. The results showed that the participants had no problems in understanding the visualization, or interface.

6.5.2 Phase 2: Feedback Survey

The pilot activity also involved a structured feedback survey, provided in Table 6.5, which we have administered after the face-to-face meeting. The questionnaire was designed to cover the qualitative aspects of NavigTweet user interface. Each participant was encouraged to provide us his/her opinion and remarks by answering these questions. They provided us high quality feedback and helped us to identify issues and suggest new features.

6.6 Pilot Evaluation and Results

In this section, we describe the pilot results. Firstly, we provide an overview of the answers in Section 6.6.1. Then, we discuss the issues identified during the pilot and the list of new requirements which have been raised during the pilot.

6.6.1 User Interviews and Rating

Figure 6.1 shows the summary evaluation of different functional areas of NavigTweet. Each pilot participant evaluated existing features of the application and proposed new requirements, both functional and non-functional.

Comments were generally favorable towards NavigTweet (*"Really useful, and aesthetically pleasant graphs with nice color-scheme"*, *"Innovative and Informative tool"*, *"User Ranking and Influencers Identification over graphs is quite*

Table 6.5: *Feedback Survey.*

QUESTIONS	ANSWER CRITERIA
QUALITATIVE ANALYSIS	
Do you find NavigTweet interesting? ***(User Interest)***	• Funny • Boring • Helpful • Informative • Innovative • Useful • Usable
How would you rate the effectiveness of NavigTweet, as an interactive tool to explore your Twitter social networks? ***(User Interaction)***	Low/High 5 point scale.
How would you rate the clarity for NavigTweet? ***(Clarity Perception)***	Low/High 5 point scale.
Do you find NavigTweet helpful in exploring and identifying the influencers (prominent twitter users)? ***(Influencers Identification)***	Yes/No/Somehow
Would you browse other users' friends' networks via NavigTweet? ***(Network Browsing Level)***	Yes/No/Somehow
How would you rate NavigTweet overall? ***(User Satisfaction)***	Low/High 5 point scale.
USER INTERFACE	
Do you like the User Interface of NavigTweet? ***(Graphical User Interface)***	• Graph representation • Friendly color-scheme • Cluster clarity • Informative node tool-tip
Which color scheme in clusters you prefer? ***(Clusters color-scheme)***	Same/Different
How much information is displayed per user node? ***(User Information Quality)***	Too little/Normal/Too much

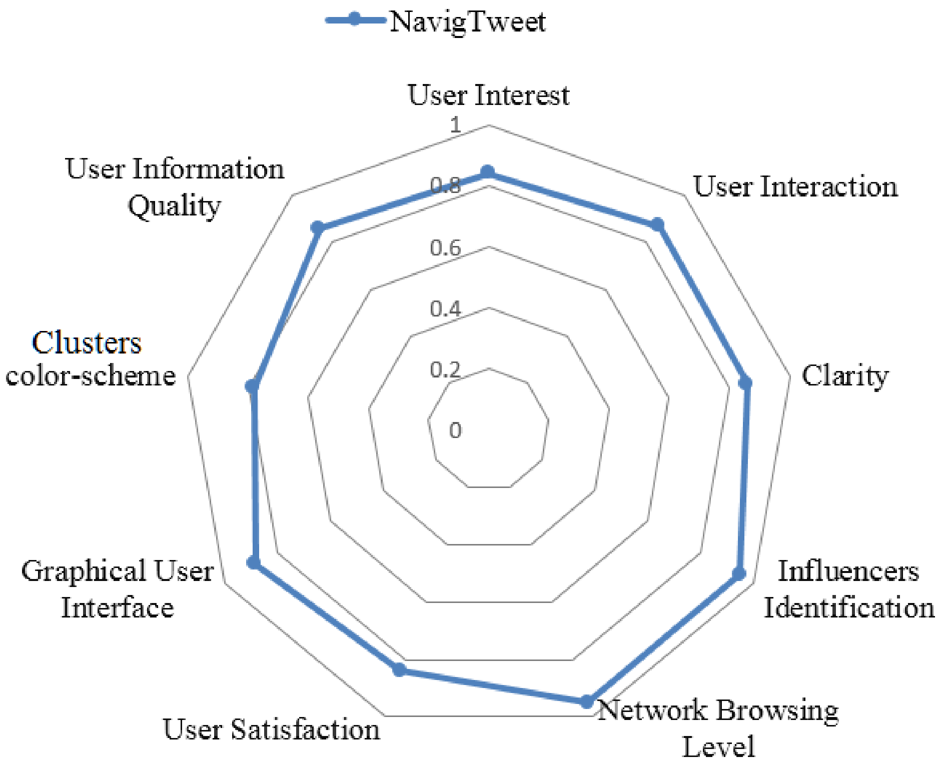


Figure 6.1: *Rating of NavigTweet based on qualitative criteria.*

wonderful!"), which was especially praised for *User Interface*, *Graph Animated Layout*, *Multi-colored Clustering Scheme*, *Dynamic Top-20*, *User- and Graph-Level panel*, *Browsing Friends' List*, *Mutual-Friends Identification*. Several participants pointed out that the tool identifies actual influencers that are visualized in a novel and easy-to-understand way.

6.6.2 Identified Issues

Overall we received positive feedbacks from all pilot participants, but following are some issues which were identified during the pilot.

- **Application Installation Problem on MAC OS**:
  During the pilot, the application could not be installed on MAC OS, although the application installer package aimed for multi-OS. The issue was due to a problem in the installer package, which was built for MAC OS. The issue has been resolved, the application installer for MAC OS has been fixed, and tested successfully.

- **Twitter Rate-Limit Issue**:
  During pilot activity, we also faced Twitter rate-limit issue, but NavigTweet is able to handle such exception by a dialog box alert and requests the user to wait for a few seconds. Some participants were not happy with rate-limit, but it is beyond our control as per Twitter API policy, although Participants appreciated the exception handling feature.

A pilot participant advised to reduce tool-tip contents and to eliminate some information panels, as the tool itself is self-explanatory and provides an understandable work flow. Three pilot participants insisted about web-based interface, which we are considering for the next release of NavigTweet. Another pilot participant advised to introduce a new panel with graph-level influencers among all selected users and their connections. We also received advice on introducing a data export feature prior to public release.

6.6.3 Feature Updates

Each participant evaluated existing requirements and features of the application and also proposed new requirements, including both functional and non-functional requirements. Table 6.6 explains the new requirements that have been highlighted by users. Prior to the public release, we implemented most of the features shown in Table 6.6.

Table 6.6: *Feature Updates Summary*

Requirement / Feature	Update Status	Comments / Remarks
Display Top-20 Graph-Level Influencers	YES	Added in left Influencers Panel.
Data export feature	YES	Button provided in left Control Panel to export the CSV file.
Graph node legend	YES	The graph legend panel is added in left Control Panel.
Node tool-tip content update	YES	Revised contents for clear user understanding.
Application stand-alone JRE file	YES	The Runnable JRE can be downloaded from NavigTweet website.
User guide in installer package	YES	PDF file is added in installer package.
Webpages of user-guide	YES	User Guide webpage created.
Web-based interface	NO	Will be considered in next version.
Influencers Graphs/Pie charts	NO	Will be considered in next version.
Twitter Analytics	NO	Will be considered in next version.

6.7 Extensive Survey: Summary of Responses

We created an online survey[1], as shown in Table 6.5, which is available on NavigTweet website. In order to get real-time feedback from end users, we circulated information within communities by engaging in online conversation.

6.7.1 Rollout Strategy

We circulated the survey via different communication channels and targeted various end-user segments. The communication channels which we adopted for NavigTweet survey rollout are listed below:

- **Emails**: Direct release notification emails sent to the user base, along with survey.
- **Social Media**: NavigTweet release news are posted on social media like FB, Twitter, LinkedIn, Google+. Additionally, various technology blogs, Twitter/FB/Linkedin are also targeted.
- **Direct Communication**: Release information circulated directly to social circle including friends, co-workers, etc. which they further circulated within their social circle.

Target Market Segmentation

The market segmentation which is targeted for NavigTweet survey rollout is described as follows:

- **Students**: University students of bachelor, Masters and PhD researchers were covered. Emails were sent to mailing lists of students in universities at local and international level.
- **Academic Faculty**: We targeted faculty resources of various universities, and sent release notification to their mailing lists.
- **Industry segment**: The industry related people were also notified through email. The emails and other information have been collected through various sources.

6.7.2 Target Audience

We targeted social networks like Facebook, LinkedIn and posted on various blogs, and communities and groups. We also targeted a variety of forums. The detail of targeted communities, blogs, and universities is listed as follows:

[1] http://goo.gl/azdMZ5

LinkedIn

We circulated NavigTweet information and survey to the following groups on LinkedIn:

- Twitter Strategies
- Semantic Web Research
- Sentiment Analysis
- Semantic Technologies
- Graph Theory and Algorithms
- Semantic Web for Life
- Web 3.0 Applications
- Social Network Analysis in Practice

Facebook

Following pages were targeted:

- Twitter Tab
- Twitter Inc.
- Scholarship Networks.
- Personal Wall.
- Friends' Walls.

Google Forum(s)

The following forums were targeted:

- Twitter4j Forum.
- Piccolo2D Forum.

Universities

We have targeted various universities, where we circulated the survey among different mailing lists of Students, Researchers, Faculty Members etc. Following is the list of those universities:

- POLIMI Italy.
- NUST, Pakistan.
- SZABIST, Pakistan.
- COMSATS, Pakistan.
- KHU, South Korea.
- UQU, Saudi Arabia.

6.7.3 Questionnaire Results

So far, we have collected 102 questionnaires from end-users. Figure 6.2 shows the daily response rate to the survey during a 3-month period, where peak responses are during the period when we spreaded information about NavigTweet on social channels. The questionnaire survey was catagorized into three parts: 1)*Qualitative Analysis* 2) *User Interface* and 3) *Comments and Suggestions*. To better understand user profiles, we also obtained demographics of users with dedicated questions.

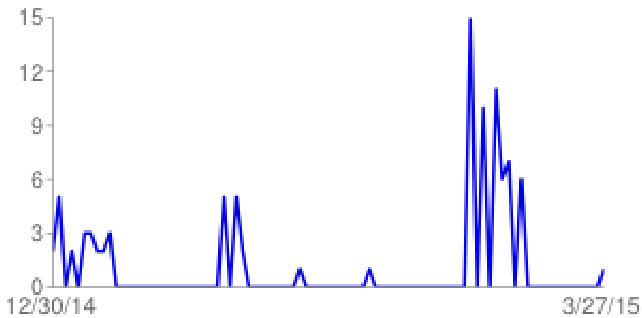


Figure 6.2: *Daily response rate over 3-months period.*

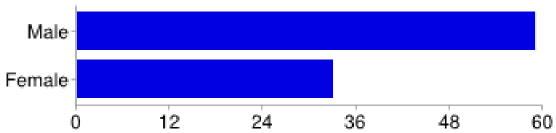
Demographics

In order to understand user demographics (e.g. age, gender, education, job profile, etc.), we asked a few questions in the questionnaire, which helped us to know the end-users. Among all users, 64 % are male under age of 31-35 years (27.2%) with Masters' degree qualification (41%). Most users normally use Twitter on a daily basis (41.3 %). Table 6.7 shows the users' demographics.

Table 6.7: *Summary of Responses: Demographics.*

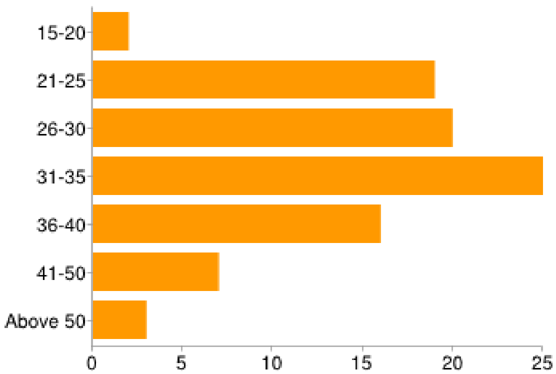
DEMOGRAPHICS

Gender



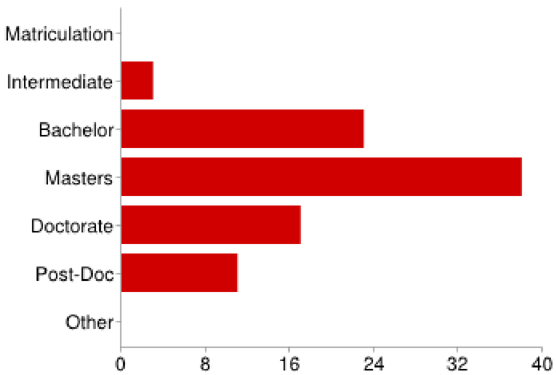
Male	64.1 %
Female	35.9 %

Age Group



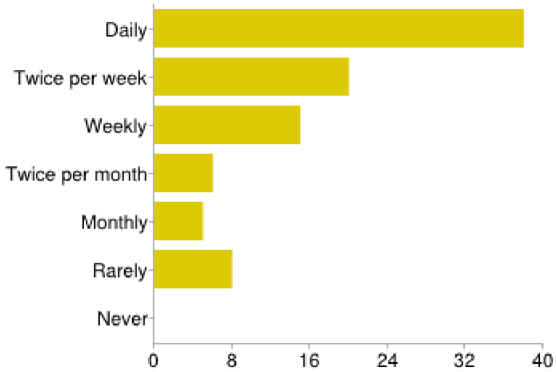
15-20	2.2 %
21-25	20.7 %
26-30	21.7 %
31-35	27.2 %
36-40	17.4 %
41-50	7.6 %
Above 50	3.3 %

Qualification Level



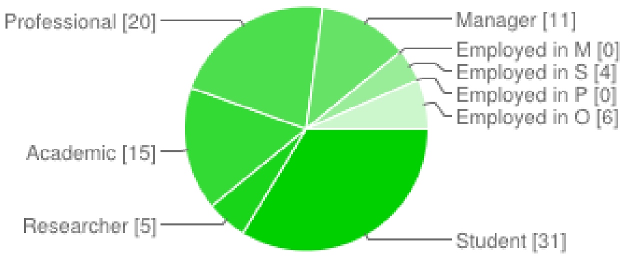
Intermediate	3.3 %
Bachelor	25 %
Masters	41.3 %
Doctorate	18.5 %
Post-Doc	12 %
Other	0 %

Twitter Usage



Daily	41.3 %
Twice per week	21.7 %
Weekly	16.3 %
Twice per month	6.5 %
Monthly	5.4 %
Rarely	8.7 %
Never	0 %

Role



Student	33.7 %
Researcher	5.4 %
Academic	16.3 %
Professional / Consultant	21.7 %
Manager	12 %
Manufacturing Industry	0 %
Services	4.3 %
Public Administration	0 %
Others	6.5 %

Nationality

French, Pakistani, Irish, Bangladeshi, Jordanian, American, Lebanese, British, Turkish, Saudi, Italian, Spanish, German, Indian, Korean, Omani, Greek, Egyptian, Emirati, Swiss, Canadian, Mexican, Iranian, Swedish

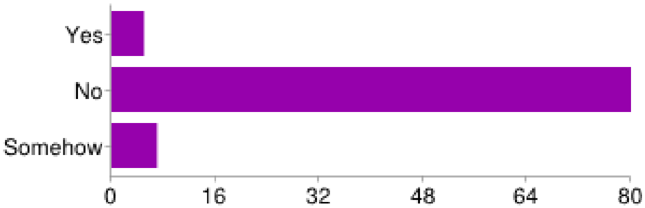
Qualitative Analysis

More than 80 % of the respondents found NavigTweet as an interesting tool (helpful, informative, innovative, easy to use, etc.). 82.6 % of users rated the effectiveness of NavigTweet by scoring 4 or 5 on a 1-5 scale. Overall, 86.9 % users seems to be satisfied with NavigTweet. The detailed summary of responses against qualitative analysis from the survey is discussed in Table 6.8.

Table 6.8: *Summary of Responses: Qualitative Analysis.*

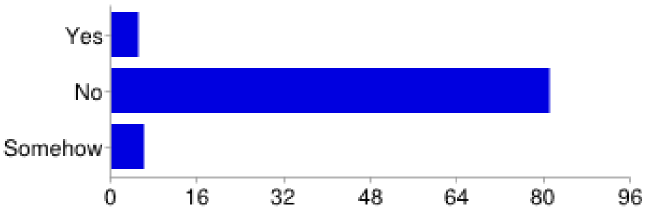
QUALITATIVE ANALYSIS

Funny



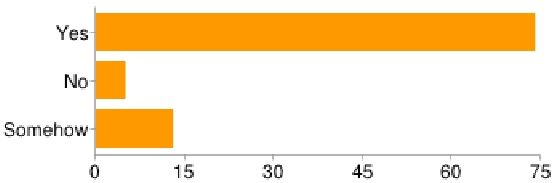
Yes	5.4 %
No	87 %
Somehow	7.6 %

Boring



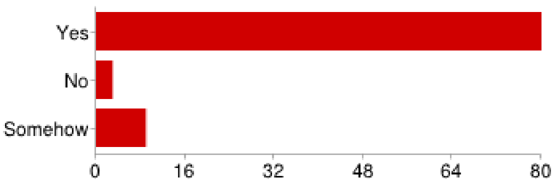
Yes	5.4 %
No	88 %
Somehow	6.5 %

Helpful



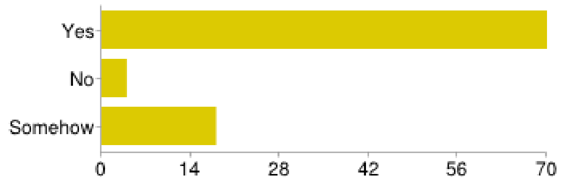
Yes	80.4 %
No	5.4 %
Somehow	14.1 %

Informative



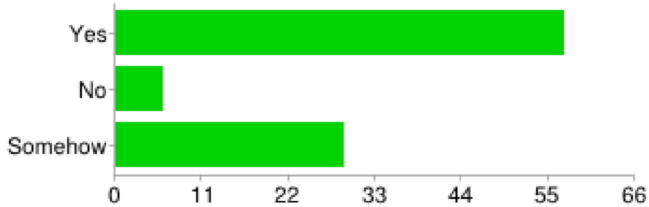
Yes	87 %
No	3.3 %
Somehow	9.8 %

Innovative



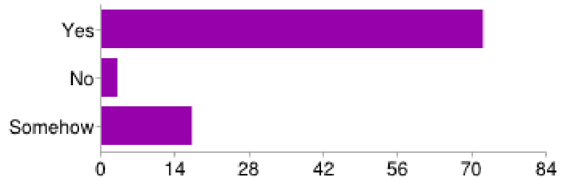
Yes	76.1 %
No	4.4 %
Somehow	19.6 %

Easy to Use



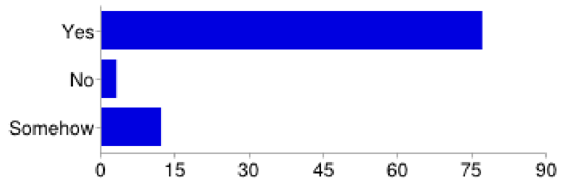
Yes	62 %
No	6.5 %
Somehow	31.5 %

Useful



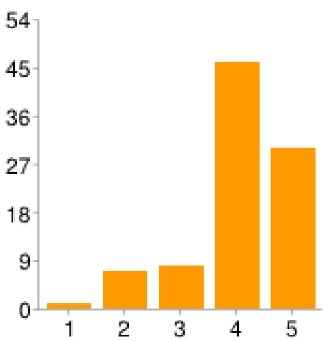
Yes	78.3 %
No	3.3 %
Somehow	18.5 %

Usable



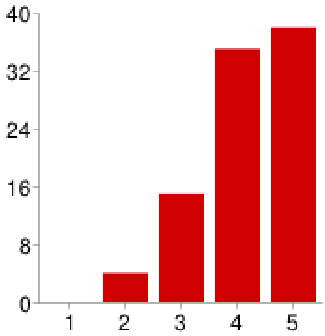
Yes	83.7 %
No	3.3 %
Somehow	13 %

Effectiveness



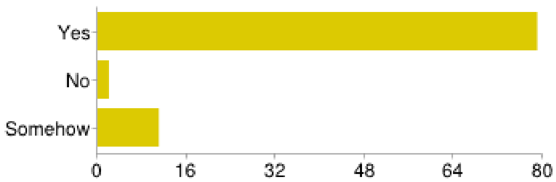
1	1.1 %
2	7.6 %
3	8.7 %
4	50 %
5	32.6 %

Clarity



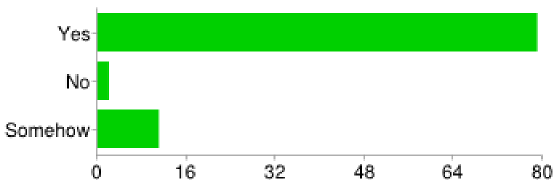
Yes	85.9 %
1	0 %
2	4.3 %
3	16.3 %
4	38 %
5	41.3 %

Influencers Identification



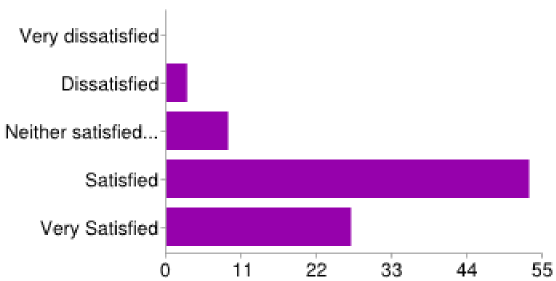
Yes	85.9 %
No	2.2 %
Somehow	12 %

Network Browsing



Yes	85.9 %
No	2.2 %
Somehow	12 %

Overall Rating



Very dissatisfied	0 %
Dissatisfied	3.3 %
Neither satisfied nor dissatisfied	9.8 %
Satisfied	57.6 %
Very Satisfied	29.3 %

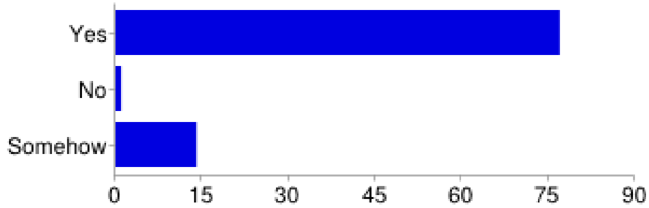
User Interface

Questions which cover aspects of NavigTweet user interface are asked in the user interface category of the feedback survey (e.g. color-scheme, tool-tip, graph representation, etc.). More than 90 % of users like the user interface of NavigTweet including graph representation, clarity, tool-tip, and color-scheme. 81.5 % of users prefer different color scheme in graph clusters. 93.5 % of users found information displayed per user node as normal. The detailed summary of responses against user interface from the survey is discussed in Table 6.9.

Table 6.9: *Summary of Responses: User Interface.*

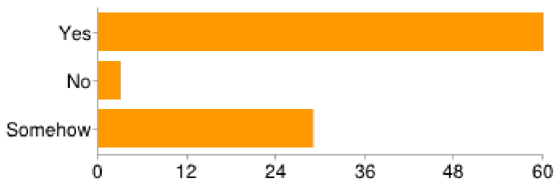
USER INTERFACE

Graph representation



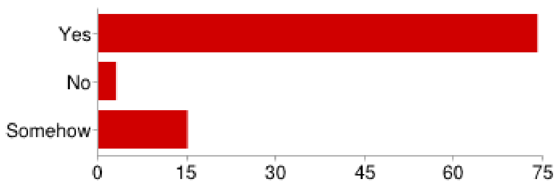
Yes	83.7 %
No	1.1 %
Somehow	15.2 %

Friendly color-scheme



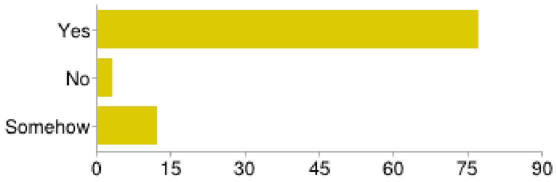
Yes	65.2 %
No	3.3 %
Somehow	31.5 %

Graph Clusters clarity



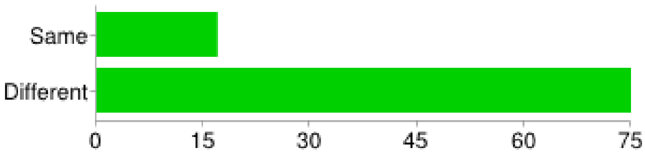
Yes	80.4 %
No	3.3 %
Somehow	16.3 %

Informative node tool-tip



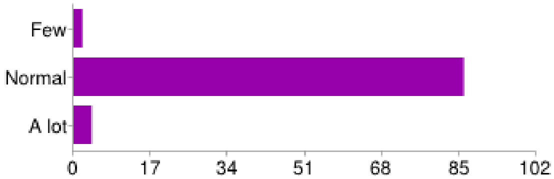
Yes	83.7 %
No	3.3 %
Somehow	13 %

Cluster's preferred color-scheme



Same	18.5 %
Different	81.5 %

Information display per node



Few	2.2 %
Normal	93.5 %
A Lot	4.3 %

6.8 Web Analytics (NavigTweet Website)

How deep do visitors navigate into the website? Are search engine visits more effective than referring site entries? We address these questions by time series analysis of *Google Analytics* [48] data. We aim to analyze the effectiveness of entries (visit behavior and length of sessions) depending on website traffic source: direct visit, in-link entries (for instance, en.wikipedia.org), and search engine visits (for example, Google). For this purpose, we used Google Analytics [86] to analyze the traffic behavior of NavigTweet Website [82]. Google Analytics is a free web analytics solution that provides webmasters with insightful information about how visitors interact with their websites.

Why use Google Analytics? Firstly, and most importantly, it is used because it provides time series data. Moreover, it is also employed because it is a free service offered by Google that generates detailed statistics about the visits to a website and a user friendly application. This tracking application, external to the website, records traffic by inserting a small piece of HTML code into every page of the website.

Google Analytics tells the web owner how visitors found the site and how they interacted with it. Users will be able to compare the behavior of visitors who were referred from search engines, from referring sites and emails, and direct visits, and thus gain insight into how to improve the site's content and design [129].

Figure 6.3 shows a snapshot of the Google analytics dashboard which provides an overview on NavigTweet website.

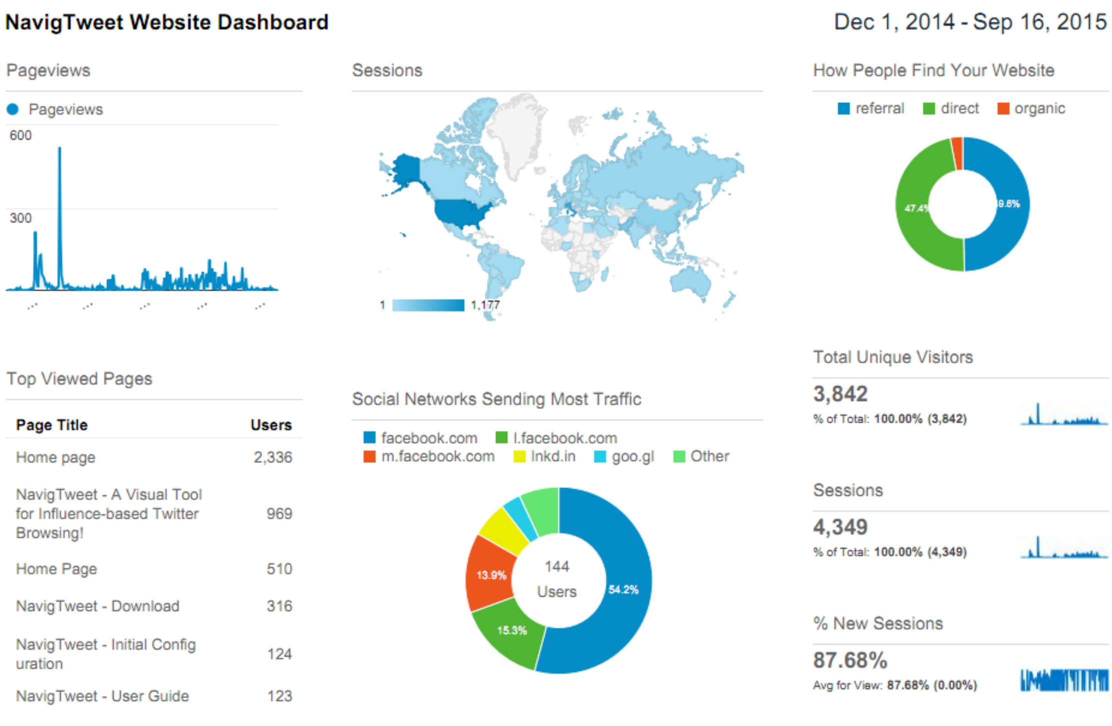


Figure 6.3: *NavigTweet Website Overview Dashboard.*

6.8.1 Website Profile

In November 2014, we launched the NavigTweet website [82]. This website encompasses information about NavigTweet's features, introduction, link to package download, user guide and online survey form. The number of pages per visit tells whether the visitors are attracted by the content or not (visit length), interested to download the tool, and willing to provide the survey or not.

6.8.2 Audience Overview

Google Analytics allows users to export report data in MS Excel format, which when transformed can be analyzed statistically. We analyzed the website performance from November 2014 to April 2015, enough for obtaining general audience overview. During this period, 1,584 sessions are found and, among those session visits, 72.6 % sessions (1,150) are created by new users, and 27.4 % sessions (434) are created by returning users. Figure 6.4 reflects the sessions overview. The percentage of first time visits is around 72.54 %.

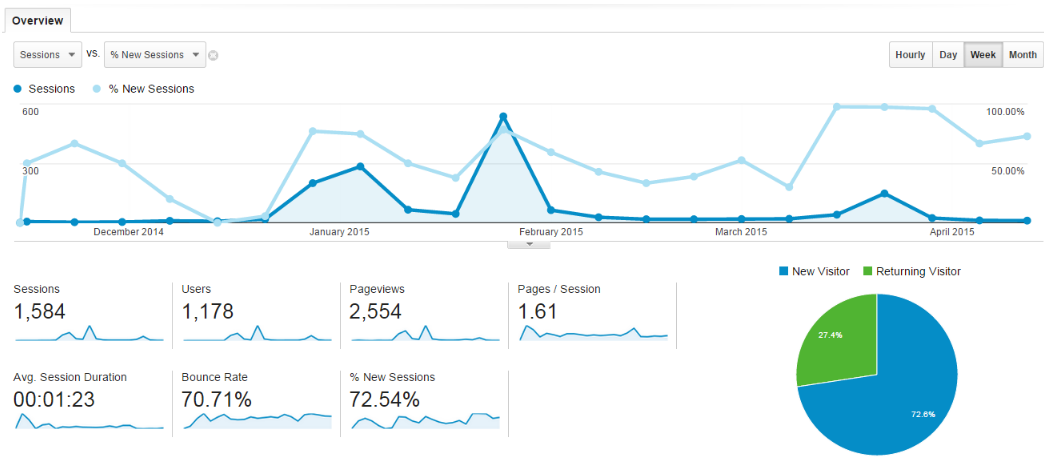


Figure 6.4: *Audience Overview.*

Geography

Figure 6.5 shows a geographical map of the users who have accessed NavigTweet's website. Users from different countries visited the website. The main continents are Europe (55.3 % sessions), Asia (26.6 % sessions), Americas (10.61 %). In total, NavigTweet's website has been accessed from 54 different countries. The highlighted country in the map is Italy, which shows 633 user sessions.

6.8.3 Web Traffic

Google Analytics allow users to analyze traffic sources and also provides customized reports on channels, referrals, and device technology reports. Figure 6.6 provides an overview of NavigTweet's traffic analysis.

Users found NavigTweet website from either direct, referral or organic search sources. Among 1,186 users, 62.5 % (741 users) are direct, 30 % (356 users) are referrals, and the remaining 7.5 % (89 users) found our website through organic search sources (e.g. google, bing etc.). The high percentage of direct users is due to our campaigns on NavigTweet release, via email and social channels. We can also observe people referring our tool to other users. Among social channels, the most active social channels are found to be Facebook and LinkedIn. Moreover,

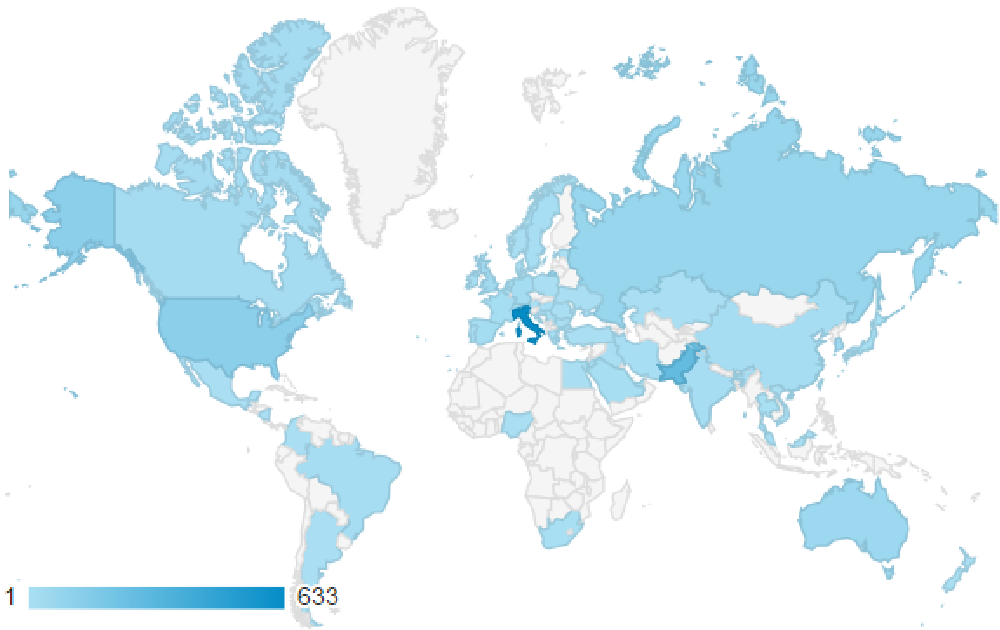


Figure 6.5: *Geographical Map Overlay.*

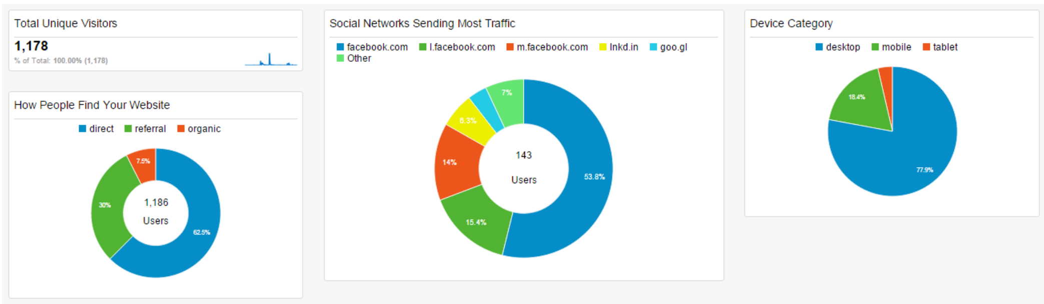


Figure 6.6: *Web Traffic Sources.*

users accessed our website from different devices, including desktop (77.9 %), mobile (18.4 %) and tablets (3.7 %).

6.8.4 Browsing Behavior

Google Analytics also allows users to analyze site content behavior by providing page views statistics. Users can obtain reports upon pageviews, bounce rate, landing pages, and behavioral flow. Figure 6.7 provides this report on NavigTweet website.

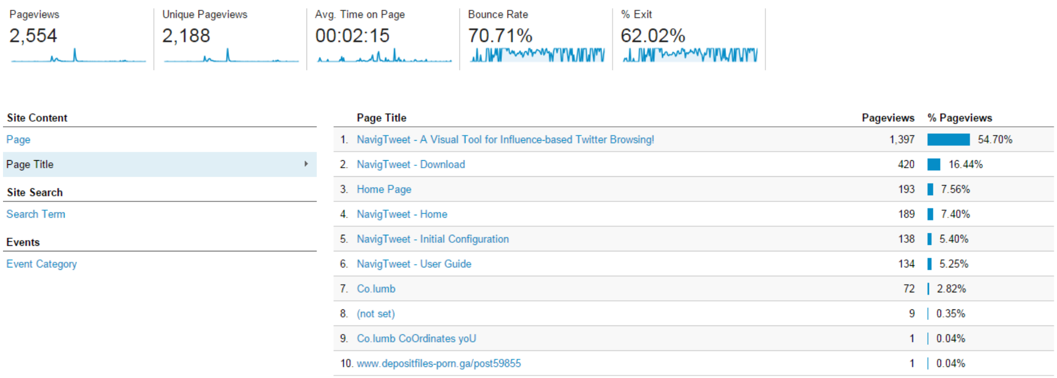


Figure 6.7: *Browsing Behavior Analytics.*

Among 2,554 total pageviews, the unique pageviews are 85.6 % (i.e.2,188). Such high percentage of unique pageviews shows that users have visited all pages of the website, rather than exiting from landing page. It shows a high interest of the user to browse the website by viewing all pages, considering average time spent on one single page is also 2 minutes and 15 seconds. Moreover, we can observe that 54.7 % pageviews are of main index page of the website (landing page), and second largest percentage is 16.44 % of the downloads page.

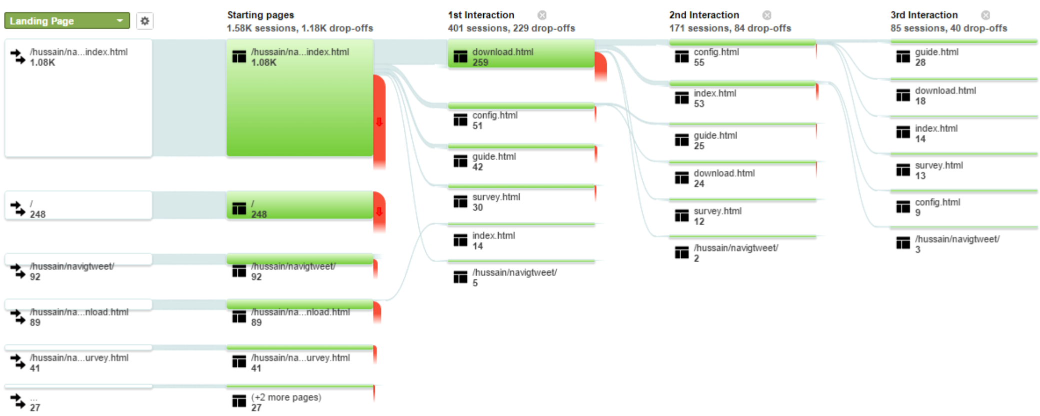


Figure 6.8: *Page Flow Behavior.*

CHAPTER 7

Discussion and Conclusion

The objective of this thesis was to investigate the relationship between content and influence on social media. From previous literature [17, 33, 92, 114], the content of messages can play a critical role and can be a determinant of the social influence of a message irrespective of the centrality of the message's author. In this thesis, we have put forwarded the hypothesis that peripheral nodes in the network play a critical role in spreading of content and have proposed an approach that supports the exploration of peripheral nodes and of their mutual connections.

We have exploited a modified power-law based force-directed algorithm [58] to highlight the local multi-layered neighborhood clusters around hub nodes. The algorithm is based on the idea that hub nodes should be prioritized in laying out the overall network topology, but their placement should depend on the topology of peripheral nodes around them.

7.1 Practical Implications

This thesis targets the research in two correlated perspectives, one perspective of the thesis focuses on the complex dynamics of tourism network and destination (brand) analysis, and the other focuses on behavioral perspective of content-based influence browsing and exploration through a software tool - NavigTweet. We provide a conceptual visual framework and related software tool to the:

- Influence Assessment (using network analysis of Tourism dynamics).
- Influence Maximization (by focusing on spread and reach).
- Influencers Identification

In the following we provide practical implications to both correlated aspects of our thesis.

7.1.1 Devising a Tourism Promotional Strategy

The work presented in regard of network analysis and tourism dynamics have some practical implications. In Chapter 3, we have discussed the approach with which the topology of the periphery is defined by grouping peripheral nodes based on the strength of their link to hub nodes, as well as the strength of their mutual interconnections, according to k-shell decomposition analysis [39, 91]. Results show that our approach produces aesthetically pleasant graph layouts, by highlighting multi-layered clusters of nodes surrounding hub nodes. These multi-layered peripheral node clusters represent a visual aid to understand influence. In Chapter 4, we have proposed three hypotheses that tie *content specificity*, *frequency of sharing* and *retweets*. The hypotheses have been tested on a sample of 1,000,000 tweets from the tourism domain. Results show that *specificity*, *frequency*, and *retweets* are mutually correlated, and have a significant impact on an author's influence and encourage us to further explore social network's intrinsic characteristics. This can be interpreted visually as our approach produces aesthetically pleasant graph layouts, by highlighting multi-layered clusters of nodes surrounding hub nodes. These peripheral nodes clusters represent a visual aid to understand influence.

The findings presented in this thesis are relevant not only theoretically but also 'practically'. Understanding what variables impact on the dynamics of information on social media platforms, like Twitter, is a prerequisite for marketers who seem interested in devising efficient social media strategies and optimizing the ways they engage with consumers on these platforms. In fact, the main innovative contribution of this work is on the different perspective on influence using a visualization approach, which is not focused only on the centrality of the author, but especially on the actual content shared by social media user. For example, one can analyses the most competitive locations, events or initiatives in the market with respect to particular market segment. Social media marketing managers can also visually identify major key players in the network, like information spreaders and information sources. In social media communities, users like information seekers, would be able to visually identify the actual and potential influencers and can further follow them. This work presents some limitations, which will be addressed as part of future work. A first comment that is worth considering is about the variables involved in this work i.e. *content specificity*, *frequency of sharing* and *retweets*. Chapter 3 addresses the relationship among these three content-based variables, but there may exist other content-based variables that can be tied with additional hypotheses, which have not yet been considered in the research model discussed in Chapter 3. For example, *number of following*, *number of followers*, *number of mentions*, etc.

As this thesis takes a behavioral perspective to understand visually the complex dynamics of tourism network and destination analysis, tourism practitioners can devise a tourism promotional strategy by using proposed approach. Table 7.1 presents a glance view of some practical implications or directions which can be adopted by tourism practitioners through the proposed approach.

Table 7.1: *Practical Guidelines - Devising a Tourism Promotional Strategy*

Approach	**Description**
Co-Branding	Link *specific* brand with *other* brand, in order to increase brand *reach*.
Brand Promotion	Comparing *competing* vs. *non-competing* brands.
Brand Fidelity	Identify most-popular vs. least-popular brands.
Cross-Promotion	Address specific *brands* with specific *categories*, in order to increase *spread*.
Target Key Players	e.g. identify the *Spreaders*, *Specifiers* for the Information Diffusion.

The proposed approach have also some challenges or limitations, for instance, the tourism businesses or industries on destination analysis are not widely spread and are small with less-structured or having in-sufficient resources. For instance, there might be a problem in data collection, or practically they might be adopting non-standard practices. These might be current challenges to the tourism industry.

Another limitation of this work is related to the Twitter case study and tourism domain data set. Clearly, tourism is one of the most complex, multi-faceted and mutable domains. Based on this, the generalizability of the conclusions that can be drawn from the observation of the real world should be carefully analyzed, especially when considering empirical studies such as those presented in this thesis. As already discussed in Sections 3.3, the datasets include tweets and retweets posted by users in a well defined time range, which cannot be considered as representative of the whole phenomenon. In addition to that, the research model proposed in this thesis and related insights should be validated across different social networks other than tourism. Although our experiment can be repeated with data from domains different from tourism, additional empirical work is needed to extend testing to multiple datasets and domains.

7.1.2 Influence-based visual Exploration - NavigTweet

As a part of research, we have developed NavigTweet [82] - a novel visualization tool for the influence-based exploration of Twitter network, discussed in Chapter 5. NavigTweet embeds the concepts of content-based influence explored in Chapter 3. NavigTweet helps to identify the key players, and follow them directly through the NavigTweet. The user can explore its own Friend-of-a-Friend (FOAF) network in order to find interesting people to be followed. The top-influencers are identified by both user-level (e.g. number of followers, number of tweets, etc.) and content-based (number of hashtags, number of URLs, etc.) parameters, thoroughly described in Chapter 4. Based upon these parameters, the tool adopts the Analytical Hierarchy Process (AHP) technique, to rank Twitter users, as our NavigTweet user explores his/her FOAF network. The application overview, objectives, architecture, building blocks and finally the implementation results are discussed in chapter 5. To assess the user experience with NavigTweet, we have conducted a qualitative pilot study, which is also reported in Chapter 5.

To gather a preliminary feedback on the NavigTweet user experience with a pilot release of NavigTweet, we have conducted a survey targeting a reference group of academic experts in the social media domain who have been asked to use the application in real time environment. Chapter 6 describes the results of a pilot test and subsequent large-scale test of NavigTweet, where we present the results of feedback questionnaire collected through the survey. We found that pilot participants were positive about the functionalities and features of the tools along with novelty of the idea itself, and received favorable comments concerning NavigTweet. We have addressed the pilot comments by modifying and updating the tool accordingly. We are currently conducting an extensive survey, and so far, we have collected 102 questionnaires from end-users, summary of responses is discussed in extensive survey. The preliminary feedback that we have obtained suggests that NavigTweet identified top-influencers more accurately by understanding influential content provided by Twitter users. On the basis of content-based influence, both user-level and tweet-level parameters play a critical role in order to identify the top-influencers among the network.

NavigTweet can help general users in order to understand the influence dynamics by providing a visual exploration platform, by which users can browse through their own and FOAF networks at unlimited depth-level friends' network. Users can maximize their influence, through content-based ranking and scoring by comparing with other friends' score. As the provided ranking is relative, and it also provides the general guidance upon perimeters, to maximize the influence spread and reach. For example, John found that his score is low due to less number of retweets, so he can work on sharing more tweets by being more active, in that way, it's most likely to receive high number of retweets upon sharing more tweets in the network, this will help John to maximize the relative rank and thus maximizing the influence. NavigTweet helps users in providing practical guidelines to provide the content-based suggestions for the maximization of influence spread and reach. Users can make some behavioral decisions by analyzing user-level and tweets-level influential parameters in order to maximize their influence.

Considering the limited desktop implementation of NavigTweet, future work will also consider implementation of web-based interface of NavigTweet, in which we intend to incorporate additional navigation and analysis features. Any suggestions or reviews received from end-users, as part of the ongoing extensive survey, will also be considered in this second release. The results presented in this thesis should be interpreted as a basis on which even more general conclusions may be drawn with further research work aimed to a comprehensive visualization framework, describing the complex dynamics of influence on social media and identification of social influencers by analyzing, exploring and interacting with social networks.

Bibliography

[1] Jarle Aarstad, Håvard Ness, and Sven A Haugland. Network position and tourism firms' co-branding practice. *Journal of Business Research*, 2015.

[2] Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. Semantic enrichment of twitter posts for user profile construction on the social web. In *The Semanic Web: Research and Applications*, pages 375–389. Springer, 2011.

[3] Alessandro Acquisti and Ralph Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. In *Privacy enhancing technologies*, pages 36–58. Springer, 2006.

[4] Eugene Agichtein, Carlos Castillo, Debora Donato, Aristides Gionis, and Gilad Mishne. Finding high-quality content in social media. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pages 183–194. ACM, 2008.

[5] H. Akama, M. Miyake, and J. Jung. A new evaluation method for the graph clustering of semantic networks built from lexical co-occurrence information. In *The 18th International Congress of Linguists, Abstracts*, volume 1, pages 206–207, 2008.

[6] R. Andersen, F. Chung, and L. Lu. Drawing power law graphs using local/global decomposition. *Twelfth Annual Symposium on Graph Drawing*, 2004.

[7] R. Andersen, F. Chung, and L. Lu. Drawing power law graphs using a local global decomposition. *Algorithmica*, 47(4):397, 2007.

[8] Isabel Anger and Christian Kittl. Measuring influence on twitter. In *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*, page 31. ACM, 2011.

[9] Simon Anholt. *Competitive identity: The new brand management for nations, cities and regions*. Palgrave Macmillan, 2006.

[10] James L Arbuckle. Ibm spss amos 20 userŠs guide. *Amos Development Corporation, SPSS Inc*, 2011.

[11] Vincenzo Asero, Simona Gozzo, and Venera Tomaselli. Building tourism networks through tourist mobility. *Journal of Travel Research*, page 0047287515569777, 2015.

[12] Sitaram Asur, Bernardo A Huberman, Gabor Szabo, and Chunyan Wang. Trends in social media: Persistence and decay. In *ICWSM*, 2011.

[13] Rodolfo Baggio. Complex systems, information technologies, and tourism: A network point of view. *Information Technology & Tourism*, 8(1):15–29, 2005.

[14] Rodolfo Baggio, Noel Scott, and Chris Cooper. Network science: A review focused on tourism. *Annals of Tourism Research*, 37(3):802–827, 2010.

[15] Richard P Bagozzi and Claes Fornell. Theoretical concepts, measurements, and meaning. *A second generation of multivariate analysis*, 2(2):5–23, 1982.

[16] Eytan Bakshy, Jake M Hofman, Winter A Mason, and Duncan J Watts. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 65–74. ACM, 2011.

[17] D. Barbagallo, L. Bruni, C. Francalanci, and P. Giacomazzi. An empirical study on the relationship between twitter sentiment and influence in the tourism domain. In *Information and Communication Technologies in Tourism 2012*, pages 506–516. Springer, 2012.

[18] M. Bastian, S. Heymann, and M. Jacomy. Gephi: an open source software for exploring and manipulating networks. In *ICWSM*, 2009.

[19] Tom Baum. Themes and issues in comparative destination research: The use of lesson-drawing in comparative tourism research in the north atlantic. *Tourism Management*, 20(5):627–633, 1999.

[20] B.B. Bederson, J. Grosjean, and J. Meyer. Toolkit design for interactive structured graphics. *Software Engineering, IEEE Transactions on*, 30(8):535–546, 2004.

[21] F. Benevenuto, M. Cha, K.P. Gummadi, and H. Haddadi. Measuring user influence in twitter: The million follower fallacy. In *International AAAI Conference on Weblogs and Social (ICWSM10)*, pages pp. 10–17, 2010.

[22] Shea Bennett. Visually explore twitter hashtags and their relationships with hashtagify, 2012.

[23] Jonah Berger and Katy Milkman. Social transmission, emotion, and the virality of online content. *Wharton Research Paper*, 2010.

[24] Mike Bergman. Large-scale rdf graph visualization tools. http://www.mkbergman.com/414/large-scale-rdf-graph-visualization-tools.

[25] Carolina Bigonha, Thiago NC Cardoso, Mirella M Moro, Marcos A Gonçalves, and Virgílio AF Almeida. Sentiment-based influence detection on twitter. *Journal of the Brazilian Computer Society*, 18(3):169–183, 2012.

[26] Uldis Bojars, John G Breslin, Vassilios Peristeras, Giovanni Tummarello, and Stefan Decker. Interlinking the social web with semantics. *Intelligent Systems, IEEE*, 23(3):29–40, 2008.

[27] Stefano Bortoli, Heiko Stoermer, Paolo Bouquet, and Holger Wache. Foaf-o-matic-solving the identity problem in the foaf network. In *SWAP*, 2007.

[28] A. Bosca, D. Bonino, and P. Pellegrino. Ontosphere: more than a 3d ontology visualization tool. Swap, 2005.

[29] F. Boutin, J. Thievre, and M. Hascoet. Focus-based filtering + clustering technique for power-law networks with small world phenomenon. *SPIE-IS & T Electronic Imaging*, 6060, 2006.

[30] D. Boyd, S. Golde, and G. Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. *IEEE*, pages pp. 1–10, 2010.

[31] L. Braun, M. Volke, J. Schlamp, A. Von Bodisco, and G. Carle. Flow-inspector: a framework for visualizing network flow data using current web technologies. *Computing*, pages 1–12, 2012.

[32] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1):107–117, 1998.

[33] L. Bruni, C. Francalanci, P. Giacomazzi, F. Merlo, and A. Poli. The relationship among volumes, specificity, and influence of social media information. In *Proceedings of International Conference on Information Systems*, 2013.

[34] Heather E Bullock, Lisa L Harlow, and Stanley A Mulaik. Causation issues in structural equation modeling research. *Structural Equation Modeling: A Multidisciplinary Journal*, 1(3):253–267, 1994.

[35] Chen C. *Information visualization: Beyond the horizon*. Springer-Verlag New York Inc, 2004.

[36] Liping A Cai. Cooperative branding for rural destinations. *Annals of tourism research*, 29(3):720–742, 2002.

[37] Amparo E Cano, Suvodeep Mazumdar, and Fabio Ciravegna. Social influence analysis in microblogging platforms–a topic-sensitive based approach. *Semantic Web*, 2011.

[38] Francesca Carmagnola, Francesco Osborne, and Ilaria Torre. User data distributed on the social web: how to identify users on different social systems and collecting data about them. In *Proceedings of the 1st international workshop on information heterogeneity and fusion in recommender systems*, pages 9–15. ACM, 2010.

[39] Shai Carmi, Shlomo Havlin, Scott Kirkpatrick, Yuval Shavitt, and Eran Shir. A model of internet topology using k-shell decomposition. *Proceedings of the National Academy of Sciences*, 104(27):11150–11154, 2007.

[40] Dorwin Cartwright, Frank Harary, and R Norman. Structural models: An introduction to the theory of directed graphs. *New York*, 1965.

[41] Paula Carvalho, Luís Sarmento, Mário J Silva, and Eugénio De Oliveira. Clues for detecting irony in user-generated contents: oh...!! it's so easy;-). In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 53–56. ACM, 2009.

[42] Twitter Inc. Help center: Faqs about retweets (rt). Twitter help. https://support.twitter.com/articles/77606-faqs-about-retweets-rt, 2014.

[43] D.S.M. Chan, K.S. Chua, C. Leckie, and A. Parhar. Visualisation of power-law network topologies. In *Networks, 2003. ICON2003. The 11th IEEE International Conference on*, pages 69–74. IEEE, 2004.

[44] Jae-Young Chang. An evaluation of twitter ranking using the retweet information. *Journal of Society for e-Business Studies*, 17(2), 2014.

[45] Jilin Chen, Rowan Nairn, Les Nelson, Michael Bernstein, and Ed Chi. Short and tweet: experiments on recommending content from information streams. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1185–1194. ACM, 2010.

[46] Beomjin Choi, TS Raghu, and Ajay Vinze. An empirical study of standards development for e-businesses: A social network perspective. In *System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on*, volume 6, pages 139a–139a. IEEE, 2006.

[47] Shu-Chuan Chu and Yoojung Kim. Determinants of consumer engagement in electronic word-of-mouth (ewom) in social networking sites. *International journal of advertising*, 30(1):47–75, 2011.

[48] Brian Clifton. *Advanced web metrics with Google Analytics*. John Wiley & Sons, 2012.

[49] P. Eades. A heuristic for graph drawing. *Congress Numerantium*, 42:149–160, 1984.

[50] P. Eades. Navigating clustered graphs using force-directed methods. *Journal of Graph Algorithms and Applications*, 4(3):157–181, 2000.

[51] P. Eades and X. Lin. Spring algorithms and symmetry. *Computing and Combinatorics*, pages 202–211, 1997.

[52] Miles Efron. Hashtag retrieval in a microblogging environment. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 787–788. ACM, 2010.

[53] J. Ellson, E. Gansner, L. Koutsofios, S. North, and G. Woodhull. Graphviz Ů open source graph drawing tools. In *Graph Drawing*, pages 594–597. Springer, 2002.

[54] Michael J Enright and James Newton. Tourism destination competitiveness: a quantitative approach. *Tourism management*, 25(6):777–788, 2004.

[55] S. Falconer. Ontograf, 2010.

[56] Weiguo Fan and Michael D. Gordon. The power of social media analytics. *Communications of the ACM*, 57(6):74–81, 2014.

[57] Followerwonk.com. Followerwonk. https://followerwonk.com/, 2014.

[58] C Francalanci and A Hussain. A visual approach to the empirical analysis of social influence. In *DATA 2014?Proceedings of 3rd International Conference on Data Management Technologies and Applications. doi*, volume 10, page 0004992803190330, 2014.

[59] Chiara Francalanci and Ajaz Hussain. Discovering social influencers with network visualization: evidence from the tourism domain. *Information Technology & Tourism*, pages 1–23, 2015.

[60] Chiara Francalanci and Ajaz Hussain. *NavigTweet: A Visual Tool for Influence-Based Twitter Browsing*, volume 9073 of *Lecture Notes in Computer Science*, book section 12, pages 183–198. Springer International Publishing, 2015.

[61] Chiara Francalanci and Ajaz Hussain. *A Visual Analysis of Social Influencers and Influence in the Tourism Domain*, book section 2, pages 19–32. Springer International Publishing, 2015.

[62] Chiara Francalanci and Ajaz Hussain. A visual analysis of social influencers and influence in the tourism domain. In Iis Tussyadiah and Alessandro Inversini, editors, *Information and Communication Technologies in Tourism 2015*, pages 19–32. Springer International Publishing, 2015.

[63] Linton C Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1979.

[64] T.M.J. Fruchterman and E.M. Reingold. Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129–1164, 1991.

[65] Wojciech Galuba, Karl Aberer, Dipanjan Chakraborty, Zoran Despotovic, and Wolfgang Kellerer. Outtweeting the twitterers-predicting information cascades in microblogs. In *Proceedings of the 3rd conference on Online social networks*, pages 3–3, 2010.

[66] V. Geroimenko and C. Chen. *Visualizing the semantic web: XML-based internet and information visualization*, chapter Spring-Embedded Graphs for Semantic Visualization, pages 172–182. Springer-Verlag New York Inc, 2006.

[67] V. Geroimenko and C. Chen. *Visualizing the semantic web: XML-based internet and information visualization*. Springer-Verlag New York Inc, 2006.

[68] Rumi Ghosh and Kristina Lerman. Predicting influential users in online social networks. *arXiv preprint arXiv:1005.4882*, 2010.

[69] J. Golbeck and P. Mutton. *Force-Directed Drawing Algorithms*, chapter Force-Directed Drawing Algorithms, pages 8493–8957. CRC Press LLC, 2004.

[70] J. Golbeck and P. Mutton. *Force-Directed Drawing Algorithms*. CRC Press LLC, 2004.

[71] Jennifer Golbeck and Matthew Rothstein. Linking social networks on the web with foaf: A semantic web case study. In *AAAI*, volume 8, pages 1138–1143, 2008.

[72] Ronald E Goldsmith and David Horowitz. Measuring motivations for online opinion seeking. *Journal of interactive advertising*, 6(2):2–14, 2006.

[73] S. Goyal and R. Westenthaler. Rdf gravity (rdf graph visualization tool). *Salzburg Research, Austria*, 2004.

[74] Thomas W Gruen, Talai Osmonbekov, and Andrew J Czaplewski. ewom: The impact of customer-to-customer online know-how exchange on customer value and loyalty. *Journal of Business research*, 59(4):449–456, 2006.

[75] Pranjal Gupta and Judy Harris. How e-wom recommendations influence product consideration and quality of choice: A motivation to process information perspective. *Journal of Business Research*, 63(9):1041–1049, 2010.

[76] Derek Hansen, Ben Shneiderman, and Marc A Smith. *Analyzing social media networks with NodeXL: Insights from a connected world*. Morgan Kaufmann, 2010.

[77] Frank Harary. Structural models: An introduction to the theory of directed graphs. 2005.

[78] J. Heer, S.K. Card, and J.A. Landay. Prefuse: a toolkit for interactive information visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 421–430. ACM, 2005.

[79] Thorsten Hennig-Thurau, Kevin P Gwinner, Gianfranco Walsh, and Dwayne D Gremler. Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *Journal of interactive marketing*, 18(1):38–52, 2004.

[80] Liaquat Hossain, Andrè Wu, and Kenneth KS Chung. Actor centrality correlates to project based coordination. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 363–372. ACM, 2006.

[81] M.L. Huang, P. Eades, and J. Wang. On-line animated visualization of huge graphs using a modified spring algorithm. *Journal of Visual Languages & Computing*, 9(6):623–645, 1998.

[82] A. Hussain. Navigtweet. `http://home.deib.polimi.it/hussain/navigtweet/`, 2015.

[83] A. Hussain, K. Latif, and A.T. Rextin. Navigowl-plugin for protege 4.1 beta, 2011. `http://protegewiki.stanford.edu/wiki/NavigOWL`.

[84] Ajaz Hussain, Khalid Latif, A Rextin, Amir Hayat, and Masoon Alam. Scalable visualization of semantic nets using power-law graphs. *Applied Mathematics & Information Sciences*, 8(1):355–367, 2014.

[85] C. J. Hutto, Sarita Yardi, and Eric Gilbert. A longitudinal study of follow predictors on twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 821–830, 2013.

[86] Google Inc. Google analytics. `www.google.com/analytics/`, 2015.

[87] D Kalamaras. Social network visualizer (socnetv). http://socnetv.sourceforge.net, 2010.

[88] A. Katifori, C. Halatsis, G. Lepouras, C. Vassilakis, and E. Giannopoulou. Ontology visualization methodsŮa survey. *ACM Computing Surveys (CSUR)*, 39(4):10, 2007.

[89] A. Katifori, E. Torou, C. Halatsis, G. Lepouras, and C. Vassilakis. A comparative study of four ontology visualization techniques in protégé: Experiment setup and preliminary results. In *Information Visualization, IV 2006.*, pages 417–423. IEEE, 2006.

[90] G. Kilpatrick. Twitter analytics. `http://twittertoolsbook.com/10-awesome-twitter-analytics-visualization-tools/`, 2015.

[91] Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Hernán A. Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6(11):888–893, 2010.

[92] Christine Klotz, Annie Ross, Elizabeth Clark, and Craig Martell. *Tweet!–And I Can Tell How Many Followers You Have*, pages 245–253. Springer, 2014.

[93] Klout.com. Klout. `https://klout.com/home`, 2014.

[94] Matthew Krawczyk and Zheng Xiang. Perceptual mapping of hotel brands using online reviews: a text analytics approach. *Information Technology & Tourism*, pages 1–21.

[95] M. Kujawski. Twitter visualization tools. http://www.mikekujawski.ca/2014/08/13/two-free-twitter-network-visualization-tools/, 2014.

[96] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM, 2010.

[97] Fabio Landini, Franco Malerba, and Roberto Mavilia. The structure and dynamics of networks of scientific collaborations in northern africa. *Scientometrics*, pages 1–21, 2015.

[98] David Laniado and Peter Mika. Making sense of twitter. In *The Semantic Web–ISWC 2010*, pages 470–485. Springer, 2010.

[99] M. Lanzenberger, J. Sampson, and M. Rester. Visualization in ontology tools. In *International Conference on Complex, Intelligent and Software Intensive Systems*, pages 705–711. IEEE, 2009.

[100] Anna Leask. Destination competitiveness: A comparative study of hong kong, macau and singapore. *Tourism Analysis*, 2015.

[101] Alex Leavitt, Evan Burchard, David Fisher, and Sam Gilbert. The influentials: New approaches for analyzing influence on twitter. *Web Ecology Project*, 4(2):1–18, 2009.

[102] Sonja Sibila Lebe, Dr Prof. Matjaž Mulej, and Rodolfo Baggio. Complex tourism systems: a visibility graph approach. *Kybernetes*, 43(3/4):445–461, 2014.

[103] Changhyun Lee, Haewoon Kwak, Hosung Park, and Sue Moon. Finding influentials based on the temporal order of information adoption in twitter. In *Proceedings of the 19th international conference on World wide web*, pages 1137–1138. ACM, 2010.

[104] Mira Lee and Seounmi Youn. Electronic word of mouth (ewom) how ewom platforms influence consumer product judgement. *International Journal of Advertising*, 28(3):473–499, 2009.

[105] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. `http://snap.stanford.edu/data`, June 2014.

[106] Jingxuan Li, Wei Peng, Tao Li, Tong Sun, Qianmu Li, and Jian Xu. Social network user influence sense-making and dynamics prediction. *Expert Systems with Applications*, 41(11):5115–5124, 2014.

[107] C.C. Lin and H.C. Yen. A new force-directed graph drawing method based on edge-edge repulsion. *IEEE Computer Society*, 2005.

[108] Chun-Cheng Lin and Hsu-Chun Yen. A new force-directed graph drawing method based on edge-edge repulsion. *Journal of Visual Languages & Computing*, 23(1):29–42, 2012.

[109] J. Lin and C. Dyer. *Data-Intensive Text Processing with MapReduce*, chapter Graph Algorithms. University of Maryland, College Park, 2010.

[110] A. Love and Edward H. *A treatise on the mathematical theory of elasticity*. at the University Press, 1920.

[111] Anshu Malhotra, Luam Totti, Wagner Meira Jr, Ponnurangam Kumaraguru, and Virgilio Almeida. Studying user footprints in different online social networks. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, pages 1065–1070. IEEE Computer Society, 2012.

[112] Andreina Mandelli, Elena Marchiori, and Lorenzo Cantoni. Tourists and destination management organizations facing social media and eword-of-mouth. a research in italy. *The UCLA Anderson Business and Information Technologies (BIT) Project: A Global Study of Business Practice (2012)*, pages 225–225, 2012.

[113] Sharon Meraz. Is there an elite hold? traditional media to social media agenda setting influence in blog networks. *Journal of Computer-Mediated Communication*, 14(3):682–707, 2009.

[114] Johnnatan Messias, Lucas Schmidt, Ricardo Oliveira, and Fabrício Benevenuto. You followed my bot! transforming robots into influential users in twitter. *First Monday*, 18(7), 2013.

[115] B. Mike. Large-scale rdf graph visualization tools, 2008. `http://www.mkbergman.com/414`.

[116] Stanley Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.

[117] K. Misue, P. Eades, W. Lai, and K. Sugiyama. Layout adjustment and the mental map. *Journal of Visual Languages and Computing*, 6(2):183–210, 1995.

[118] Jacob Levy Moreno. Who shall survive?: A new approach to the problem of human interrelations. 1934.

[119] Seth A. Myers and Jure Leskovec. The bursty dynamics of the twitter information network. In *Proceedings of the 23rd international conference on World wide web*, pages 913–924, 2014.

[120] Mor Naaman, Jeffrey Boase, and Chih-Hui Lai. Is it really about me?: message content in social awareness streams. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 189–192. ACM, 2010.

[121] Meenakshi Nagarajan, Hemant Purohit, and Amit P Sheth. A qualitative examination of topical tweet and retweet practices. *ICWSM*, 2(010), 2010.

[122] Mark EJ Newman. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46(5):323–351, 2005.

[123] Julie Pallant. *SPSS survival manual: A step by step guide to data analysis using SPSS*. McGraw-Hill International, 2010.

[124] John C Paolillo and Elijah Wright. Social network analysis on the semantic web: Techniques and challenges for visualizing foaf. In *Visualizing the semantic web*, pages 229–241. Springer, 2006.

[125] Do-Hyung Park and Jumin Lee. ewom overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications*, 7(4):386–398, 2009.

[126] Douglas G Pearce. Toward an integrative conceptual framework of destinations. *Journal of Travel Research*, 53(2):141–153, 2014.

[127] Richard Perline. Strong, weak and false inverse power laws. *Statistical Science*, pages 68–88, 2005.

[128] Sasa Petrovic, Miles Osborne, and Victor Lavrenko. Rt to win! predicting message propagation in twitter. In *ICWSM*, 2011.

[129] Beatriz Plaza. Monitoring web traffic source effectiveness with google analytics: An experiment with time series. In *Aslib Proceedings*, volume 61, pages 474–482. Emerald Group Publishing Limited, 2009.

[130] H.C. Purchase. Metrics for graph drawing aesthetics. *Journal of Visual Languages & Computing*, 13(5):501–516, 2002.

[131] Helen C. Purchase, D. Carrington, and J. Allder. Empirical evaluation of aesthetics based graph layout. *Empirical Software Engineering*, 7(3):233–255, 2002.

[132] Zhuo-Ming Ren, An Zeng, Duan-Bing Chen, Hao Liao, and Jian-Guo Liu. Iterative resource allocation for ranking spreaders in complex networks. *EPL (Europhysics Letters)*, 106(4):48005–48005, 2014.

[133] Marsha L Richins. Negative word-of-mouth by dissatisfied consumers: A pilot study. *The journal of marketing*, pages 68–78, 1983.

[134] Matthew A Russell. *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. " O'Reilly Media, Inc.", 2013.

[135] Thomas L Saaty. How to make a decision: the analytic hierarchy process. *European journal of operational research*, 48(1):9–26, 1990.

[136] Thomas L Saaty and Luis G Vargas. *Models, methods, concepts & applications of the analytic hierarchy process*, volume 1. Springer, 2001.

[137] M. Sarkar and M.H. Brown. Graphical fisheye views of graphs. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 83–91. ACM, 1992.

[138] Noel Scott, Rodolfo Baggio, and Chris Cooper. *Network analysis and tourism: From theory to practice*, volume 35. Channel View Publications, 2008.

[139] Stephen B Seidman. Network structure and minimum degree. *Social networks*, 5(3):269–287, 1983.

[140] Georg Simmel. *Soziologie: untersuchungen über die formen der vergesellschaftung*. Duncker & Humblot, 1908.

[141] M. Sintek. Visualizing protege ontologies, 2003. `http://protegewiki.stanford.edu/wiki/OntoViz`.

[142] Justin Smith. 10 key seo strategies every facebook page owner should know, inside facebook. http://www.adweek.com/socialtimes/seo-facebook-pages-10-key-strategies/224907, July 2009.

[143] Marc A Smith, Ben Shneiderman, Natasa Milic-Frayling, Eduarda Mendes Rodrigues, Vladimir Barash, Cody Dunne, Tony Capone, Adam Perer, and Eric Gleave. Analyzing (social media) networks with nodexl. In *Proceedings of the fourth international conference on Communities and technologies*, pages 255–264. ACM, 2009.

[144] Raymond T Sparrowe, Robert C Liden, Sandy J Wayne, and Maria L Kraimer. Social networks and the performance of individuals and groups. *Academy of management journal*, 44(2):316–325, 2001.

[145] Kate Starbird and Leysia Palen. *Pass it on?: Retweeting in mass emergency*. International Community on Information Systems for Crisis Response and Management, 2010.

[146] James H Steiger. Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, 42(5):893–898, 2007.

[147] M. Storey, M. Musen, J. Silva, C. Best, R. Ernst, N.and Fergerson, and N. Noy. Jambalaya: Interactive visualization to enhance ontology authoring and knowledge acquisition in protégé. In *Workshop on Interactive Tools for Knowledge Capture (K-CAP-2001)*. Citeseer, 2001.

[148] Bongwon Suh, Lichan Hong, Peter Pirolli, and Ed H Chi. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *Social computing (socialcom), 2010 ieee second international conference on*, pages 177–184. IEEE, 2010.

[149] Y. Sure, J. Angele, and S. Staab. Ontoedit: Guiding ontology development by methodology and inferencing. *On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE*, pages 1205–1222, 2010.

[150] R. Tamassia. Constraints in graph drawing algorithms. *Constraints*, 3(1):87–120, 1998.

[151] Oren Tsur and Ari Rappoport. What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 643–652. ACM, 2012.

[152] Twitonomy.com. Twitonomy. http://www.twitonomy.com/, 2014.

[153] Twitter4j.org. Twitter4j api. http://twitter4j.org/en/index.html, 2010.

[154] Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P Gummadi. On the evolution of user interaction in facebook. In *Proceedings of the 2nd ACM workshop on Online social networks*, pages 37–42. ACM, 2009.

[155] C. Walshaw. A multilevel algorithm for force-directed graph drawing. *Journal of Graph Algorithms and Applications*, 7(3):253–285, 2003.

[156] Y. Wang and Z. Wang. A fast successive over-relaxation algorithm for force-directed network graph drawing. *Science China Information Sciences*, 55(3):677–688, 2012.

[157] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, 1994.

[158] Shaomei Wu, Jake M Hofman, Winter A Mason, and Duncan J Watts. Who says what to whom on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 705–714. ACM, 2011.

[159] X. Xu, N. Yuruk, Z. Feng, and T.A.J. Schweiger. Scan: a structural clustering algorithm for networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 824–833. ACM, 2007.

[160] Dejin Zhao and Mary Beth Rosson. How and why people twitter: the role that micro-blogging plays in informal communication at work. In *Proceedings of the ACM 2009 international conference on Supporting group work*, pages 243–252. ACM, 2009.