Tampere University

Ana María Soto Blázquez

DETECTING TIE STRENGTH FROM SOCIAL MEDIA DATA IN A CONFERENCE SETTING

Knowledge Management
Master´s Thesis
June 2019

Abstract

SOTO BLÁZQUEZ, ANA MARÍA: Detecting tie strength from social media data in a conference setting
Master of Science Thesis
Tampere University
Master´s Degree Programme in Industrial Engineering
Professor Hannu Kärkkäinen
June 2019

The concept of tie strength was introduced by Granovetter as "a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie". Since the publication of this seminal study, several studies have been conducted incorporating the concept of tie strength in numerous fields.

The growing rise of social media in recent years has shaped a new way of establishing and maintaining ties between people. As a result, studies have been conducted that, based on social media data, are focused on the evaluation of tie strength between users. Social media has also positioned itself as a key tool in the development of events such as conferences, as it is consolidated as the communication platform through which to disseminate information and knowledge and networking.

Therefore, in the present study, it is sought to evaluate tie strength using publicly available Twitter data in the context of a conference. Specifically, the aim is to analyse the potential of implicit networks (particularly, mentions networks) generated in social media sites (particularly, Twitter) when evaluating tie strength and social ties, with special emphasis on weak ties and latent ties. Ultimately, the aim is to obtain conclusions that result in the demonstration of the utility and the advantages of implementing this analysis in the recommendation systems in conferences.

To address the main statement problem, this study starts with a review of the existing literature related to the topic. Subsequently, as regards the empirical part of the study, a case study approach is conducted. Specifically, a longitudinal single-case analysis is analysed, since the mentions networks generated from the publicly available Twitter data of the conference HICSS along nine editions (from 2010 to 2018) are studied. Different measures of social network analysis have been used to obtain results and conclusions.

Based on the analysis, different potentially useful measures for the evaluation of mentions networks and social ties are identified. These measures have served to analyse the social structures formed in a conference setting (highlighting star structures that reflect the information disseminating role of certain nodes), to identify the most relevant and influential participants (which generally correspond to important roles of the conference, as organizers or speakers), or to observe tendencies and groupings in communities according to common interests, among others.

Keywords: Tie strength, Social ties, Weak ties, Latent ties, Social media, Implicit networks, Twitter, Mentions networks, Conference

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Preface

This thesis explores the concept of tie strength in the context of a conference by analysing implicit networks generated from publicly available Twitter data.

I would like to thank in the first place Prof. Hannu Kärkkäinen for providing me the opportunity to work in this interesting topic and for giving me the support, the confidence, the guidance and the dedication during the entire development process of the Master's thesis. Secondly, I would also like to thank Jayesh Prakash Gupta for his support and help whenever I needed throughout this process. And finally, I would like to thank my family and friends for their continuous support and for having allowed me to have and enjoy this international experience during this year.

Tampere, 20 June 2019

Ana María Soto Blázquez

Contents

List of Illustrations

List of Figures

List of Tables

List of Symbols and Abbreviations

API	Application Programming Interface
HICSS	Hawaii International Conference on System Sciences
SH	Structural Holes Theory
SWT	Strength of Weak Ties

1 Introduction

1.1 Research background

The concept around which the present work revolves is tie strength. This term is introduced by Granvoetter in 1973 in his seminal study "The Strength of Weak Ties". Along with it, Granovetter focuses on a main distinction according to the type of tie strength. That is, Granovetter introduces the concepts of weak ties and strong ties.

From this seminal study, several studies have been subsequently conducted addressing this concept, being many the application fields of it. Education, recruitment, academic research, economics, business, information science are just some of the fields in which this concept has been used, being the main motivations for its association to these fields the networking, and transfer of information and knowledge (Levin and Cross, 2004; Zhang *et al.*, 2017).

On the other hand, in recent years, social media has acquired an indisputable importance both in the transfer of information and in the establishment, maintenance and development of new connections among people. That is, social media has been established as a fundamental tool for managing the machinery that encloses the concept of social ties and tie strength (Haythornthwaite, 2005; Boyd and Ellison, 2007; Kaplan and Haenlein, 2010; Wollan, Smith and Zhou, 2010; Nisar, Prabhakar and Strakova, 2019).

For the moment, there are several studies that have been conducted taking into account personal or available social media data for the evaluation of tie strength. However, the same number of studies is not found if the focus of the analysis is on the implicit networks built on social media sites.

Implicit networks offer a new perspective in the analysis and evaluation of social ties and tie strength, since they offer a scenario in which users present non-evident connections by sharing some characteristic, interest or common behaviour. That is to say, with the incorporation of the study of implicit networks to the evaluation of tie strength, a new research door is opened, being its focus on the identification of weak and latent ties.

Conferences are events which participants use mainly as a means of networking and obtaining educational benefits, among others (Severt *et al.*, 2007). That is, among the main motivating factors to attend a conference, are those in which its engine is again the transfer of novel information and knowledge, and networking. This fact, together with the growing rise of social media sites, has led in recent years to an increasing use of social media as a communication platform in conferences. With this, there are numerous advantages provided both to the organizers (e.g. creation of tailored conference content, recognition of relevant participants, improvement of the organizing and planning of conferences, enhancement of tailored networking opportunities or identification of interesting topics) and the rest of the conference participants (e.g. obtainment of novel information and relevant sources, development of professional career or establishment and maintenance of new potentially useful contacts) (Ebner, Rohs and Schön, 2010; Ross *et al.*, 2011; Jussila *et al.*, 2013; Aramo-Immonen, Jussila and Huhtamäki, 2014, 2015; Aramo-Immonen *et al.*, 2016).

From the compendium of the ideas mentioned in the previous paragraphs, the usefulness of the analysis and evaluation of social ties and tie strength from social media data in the context of a conference can be understood. In particular, this practice may involve the improvement and development of automated recommendation systems in a conference setting, which allow the organizers to improve their tasks and increase the advantages obtained by the rest of the participants.

In this way, as already mentioned, in the present work it is intended to provide a method for the analysis of implicit networks based on social media data (specifically, Twitter data), that allows the analysis and evaluation of tie strength in the context of conferences, so that useful applications and conclusions of these results can be obtained. Therefore, based on these ideas, the current study tries to address the research gap that is shown in the following illustration (Illustration 1).

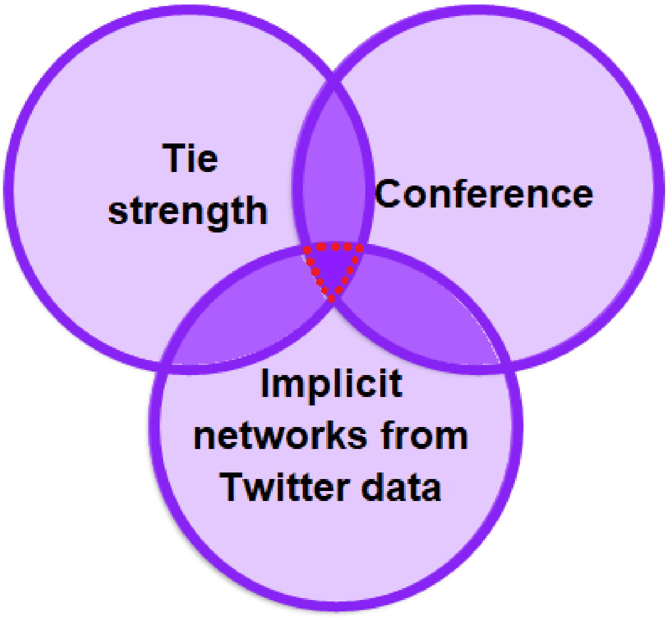

Illustration 1. Venn diagram indicating the research gap

1.2 Research questions

According to what is mentioned in the research background section (section 1.1), the current study seeks to evaluate and analyse tie strength and social ties in the context of a conference, through the analysis of implicit networks obtained from publicly available Twitter data. The description of this research gap has led to the formulation of the main problem statement that guides the research process of this work. This problem statement is reflected in the primary ontological question of this study, which is shown below:

How can implicit networks be used to recognize social ties in the context of conferences?

To be able to address the above question, five supporting research questions have also been formulated, which help to direct and focus the research process throughout the progress of this work. These five supporting research questions are the ones that are shown below.

RQ1 – What kind of information can be obtained from the Twitter data in the context of a conference?

RQ2 – What ways are there currently to identify social ties and evaluate tie strength from social media?

RQ3 – Which Twitter data items are related to social ties and tie strength?

RQ4 – How can that information be used to obtain interesting conclusions related to tie strength and social ties and useful social ties?

RQ5 – How can the analysis of implicit networks (mentions networks) in individual and sequential conferences be used in the recognition of social ties and useful social ties?

These questions mainly focus on the available information and data, the existing methods, as well as the potential of the implicit networks for the accomplishment of the current research. Specifically, the first three research questions focus on the analysis and evaluation of the existing literature, so that it can be concluded and understood what kind of information, what methods and what data are available. The fourth research question covers a more practical context, since it seeks to understand and explain the relevance of all that available information, trying to show in turn the possible conclusions that derive from it. Finally, the last supporting research question focuses on the implicit networks in the context of the empirical part of the present work, that is, in the context of the longitudinal analysis of different annual editions of the same conference, seeking in turn the obtaining of relevant conclusions about the recognition of social ties and useful social ties.

Therefore, it can be concluded that the formulation of these five supporting research questions helps to understand the current state of the available information, as well as contribute to obtain conclusions that derive from the empirical part conducted throughout the present work and research process.

1.3 Structure of the thesis

This section aims to present the overall structure of the thesis. At a general level, the structure of this thesis can be divided into four different parts: introduction, literature review, empirical study and conclusions (see Illustration 2).

The first chapter of the present document provides the introduction of the thesis by presenting an overall overview of the topic and indicating the research gap. Also, in this chapter, the identification of the main problem statement is presented, as well as the formulation of the supporting research questions based on the identified research gap.

The second part of the thesis is composed of three chapters in which the literature review is addressed. In particular, chapter 2 refers to the main concepts and theories related to the main topic of the present study, tie strength. Chapter 3, on the other hand, offers information related to social media, as well as to the important role that this tool plays in the identification and analysis of implicit networks and tie strength. And, finally, chapter

4 alludes to the context in which the analysis of the present work is framed, that is, the conferences. It provides the different reasons why people attend conferences, as well as giving a justification of the relevance of tie strength analysis in that context.

The third part of the thesis, in which the empirical part of the study is addressed, includes two chapters. In chapter 5, the analysis methods used are exposed, as well as the data collection process is explained. Chapter 6 presents the results obtained from the analysis of the case study.

Finally, the last part of the thesis is addressed in chapter 7. In this last chapter, the discussion of the research questions and the conclusions obtained from the study are presented. This chapter also includes the limitations found, as well as the topics for future research.

The figure shown below (Illustration 2) presents a schematic diagram in which the structure of the thesis can be visualized in a simplified way.

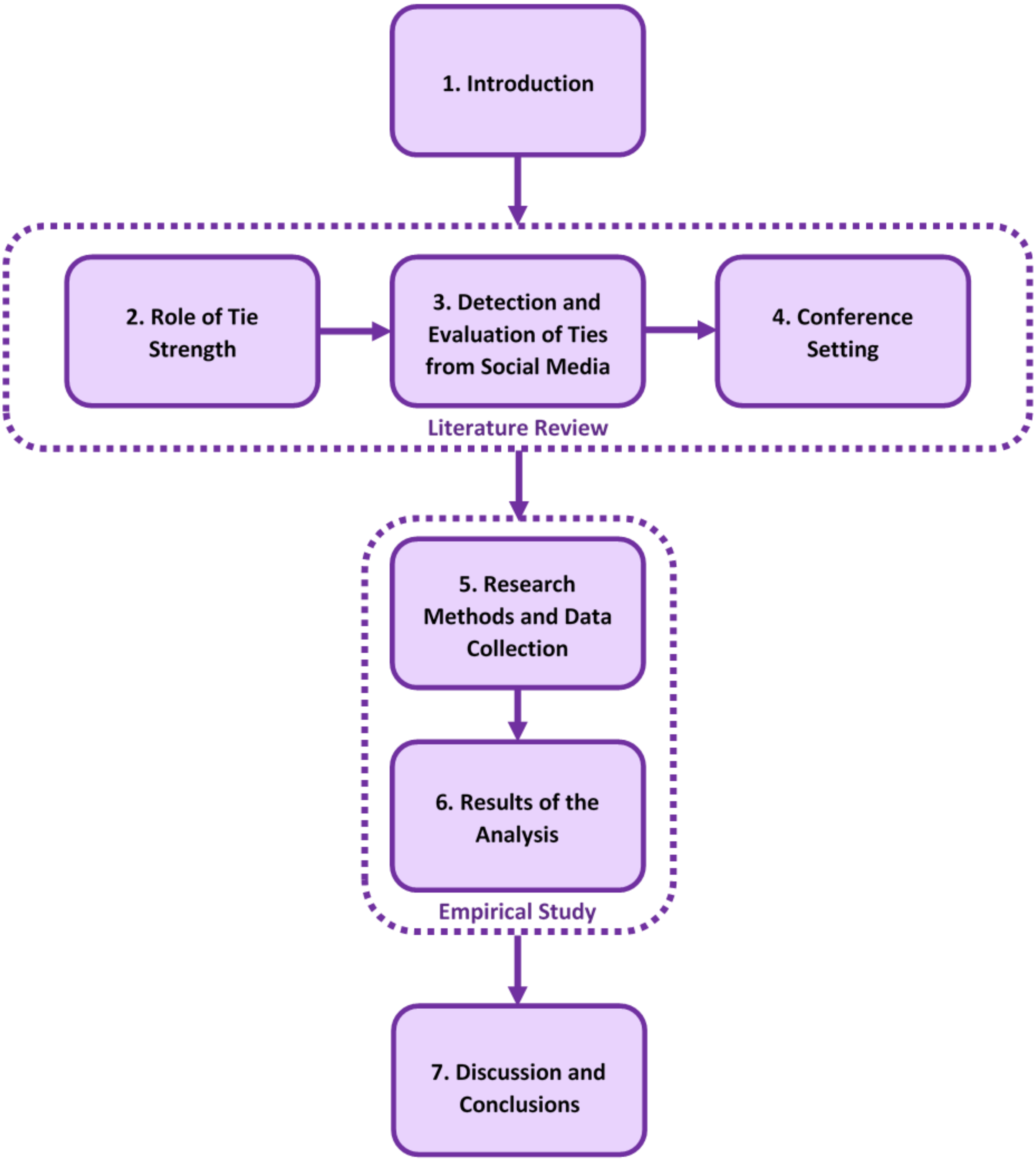

Illustration 2. Schematic diagram with the structure of the thesis

2 Role of Tie Strength

2.1 Definition of tie strength

This section seeks to define the concept of tie strength, using various studies in which the term has been utilized.

Firstly, it is essential to mention the seminal article, which was published by Granovetter in 1973, *The Strength of Weak Ties*, since this was the precursor of the topic that concerns here. In this work, Granovetter introduces the concept of tie strength, providing a set of factors and characteristics that allow the understanding of this concept. For Granovetter, tie strength is “a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie” (Granovetter, 1973).

Nevertheless, as can be seen, Granovetter did not provide a completely concrete and precise definition of the term tie strength, since he introduced the concept by defining it in terms of its indicators (Marsden and Campbell, 1984). That is why other authors have continued in this direction and have tried to describe the meaning of this concept in a more accurate way.

In this way, Krackhardt published in 1992 his work *The Strength of Strong Ties: The Importance of Philos in Organizations* in which he presents an alternative definition of tie strength. Krackhardt argues that tie strength depends on the degree to which the following three conditions are met: interaction (individuals must interact with each other to build a strong tie), affection (individuals must feel affection for each other to build a strong tie) and time (to build a strong tie it is required to have a history of interactions that have lasted over an extended period of time, it cannot be something instantaneous) (Krackhardt, 1992).

But, however, the majority of authors has continued with the definition provided by Granovetter. Thus, for example, Gilbert defends that tie strength refers to the feeling of closeness with another person (Gilbert, 2012), and Marsden supports the basis of his study on the definition provided by Granovetter.

2.2 Different kind of ties

In this section, a general view about the different types of ties that have been defined in the existing literature is given.

Already in the first work found about the topic, Granovetter introduced the different types of ties that can be found between two individuals. In his study, Granovetter presents a classification that is linked to the level of strength that these ties have. That is, the author differentiates between strong, weak or absent ties (Granovetter, 1973).

Starting from this idea introduced by Granovetter, other authors have continued the study and have deepened and completed the definition of such classification. In particular, subsequent studies, such as Haythornthwaite´s, offer a new perspective, finding a new potential in some of the formerly called "absent ties", so that now they would be called latent ties. These latent ties are, therefore, ones that can technically exist but have not been activated yet (Haythornthwaite, 2002). In this way, the types of ties can be classified and defined as shown below:

- **Strong ties:** These refer to people you really trust, people with whom you largely share your social circles (Gilbert and Karahalios, 2009).
- **Weak ties:** This type of ties, however, refers to mere acquaintances, who generally have different social circles than your own closest ones (Gilbert and Karahalios, 2009).
- **Absent ties:** Within this classification are included both the lack of any kind of relationship and the ties without substantial significance (i.e. a “nodding” relationship between people living on the same street or the “tie” to the vendor from whom one customarily buys a morning newspaper) (Granovetter, 1973).
- **Latent ties:** These are defined as the ties for which a connection is available technically but that has not yet been activated by social interaction (Haythornthwaite, 2002). So, for the moment they are “absent ties” but they have the possibility of becoming “weak ties”.

In the existing literature, other types of classifications have been made, which include other categories such as “dormant ties” (Levin, Walter and Murnighan, 2011) or “intermediate ties” (Retzer, Yoong and Hooper, 2010). However, the present work is focused on the study of the above classification.

Within the classification realized, it can be highlighted that the study interest falls on the categories of strong tie, weak tie and latent tie. On the other hand, as has been mentioned, latent tie plays a role of predecessor to weak tie. Therefore, there is a greater emphasis on the definitions of strong tie and weak tie, since their different characteristics will be those that lead to determine the interest and potential of each of these categories, depending on the context in which they are framed.

Thus, a deeper explanation about these categories is presented below, as well as the importance and strength of each of them.

The category of **strong ties** includes examples such as friends, co-workers or team-mates. They experience frequent and multiple types of interactions (emotional and instrumental); they show a high level of intimacy and self-disclosure, and present reciprocity in their exchanges. These people tend to be similar and they travel in the same social circles, so the experience, information, attitudes, resources and contacts come from the same source. Thus, this feature becomes a negative aspect, since there is no diversity. However, the positive factor presented by this type of ties is the great willingness and motivation to share what information and resources they have, that is, the great willingness to help each other. (Haythornthwaite, 2005)

On the other hand, **weak ties** refer to acquaintances, casual contacts or others in an organization. In this case the interactions are infrequent and primarily instrumental. Unlike strong ties, these people tend to be different from each other and they travel in distinct social circles. And this becomes the positive aspect and the strength of this type of ties, since information, experience, resources, attitudes and contacts come from different social spheres. Nevertheless, the weak point of this type of ties is the fact that they have less motivation and willingness to share all this information. (Haythornthwaite, 2005)

The table presented below summarizes the most important characteristics of each of these types of tie.

	Strong ties	**Weak ties**
Examples	Friends, co-workers, team-mates	Acquaintances, casual contacts
Tie characteristics	- Tend to be similar to each other - Same social circles	- Tend to be different from each other - Distinct social circles
Type of interaction	- Frequent - Emotional and instrumental - High level of intimacy and self-disclosure - Reciprocity	- Infrequent - Primarily instrumental
Weakness	- Experience, information, resources come from the same source	- Low motivation to share what information and resources they have

Strength	- Great willingness and motivation to share what information and resources they have and to help each other	- Experience, information, resources come from different social spheres (diversity)

Table 1. Strong and Weak Ties

After reviewing the existing literature, since the publication of Granovetter's work, a proof of the importance and relevance of weak ties was left. And, taking into account that a person generally presents a greater percentage of weak ties than of strong ties (Granovetter, 1973), and considering the greater potential utility of the information transmitted in weak ties due to its diversity, it leads again to emphasize the importance of weak ties. This has been also stated by other authors in later studies, affirming that weak ties often provide access to novel information, information not circulating in the closely knit network of strong ties (Gilbert and Karahalios, 2009).

2.3 Significance of measuring tie strength

Already in Granovetter´s seminal work, it is presented the importance of the evaluation of tie strength. In particular, this paper presents an analysis of the type of tie strength at an interpersonal level that has the best repercussion among job seekers to find the source of information for a new job.

After this seed initiated by Granovetter, several studies have been conducted over the years, expanding both the level of such analysis (group level, intra-organizational level and inter-organizational level), as well as the field or area of application of this analysis (e.g. education, recruitment, journalism, academic research). (Zhang *et al.*, 2017)

To help demonstrate the significance of the measurement and analysis of tie strength, a table is shown below to help draw a more complete idea about examples of the fields and levels at which studies have been made where the concept of tie strength has been used (Gupta, 2016; Zhang et al., 2017).

Level of analysis	Context of analysis
Individuals	- Job search - Collaboration - Innovation - Information and knowledge sharing - Information and knowledge access - Sources sharing - Motivation - Social media -Knowledge quality

	- Social influence - Social similarity - Productivity
Group and Intra-organizational	- Information flow and knowledge transfer - Social support - Social network analysis - Collaboration - Productivity and performance - Interpersonal interactions
Inter-organizational	- Inter-firm networking - Intercultural communication - Interpersonal interactions - General management - Information and knowledge transfer - Collaboration and trust - New business development - Multipartner alliances - Marketing - Innovation - Social media

Table 2. Examples of contexts and fields of analysis in studies related to measuring tie strength

In general, after reviewing the existing literature, it can be concluded that there are three main different ways in which weak ties generate a competitive advantage. These three main roles of the weak ties are: "searching for contacts with new information and knowledge", "efficient transferring of information and knowledge" and "spreading information to large group of actors". (Zhang *et al.*, 2017)

Therefore, this section serves to record the relevance and usefulness of the application of the tie strength concept when analysing certain social phenomena. Specifically, in the present work, the study will focus on the context of the conference setting, so that it will seek to obtain relevant conclusions in this field and at an individual level.

2.4 Major theories and models related to tie detection and tie strength

This section aims to present the main theories related to tie strength, and tie detection and identification between individuals within a network.

However, it is first necessary to explain what is meant by a network. A network consists of a set of actors or nodes along with a set of ties of a specified type (such as friendship) that link them. These ties are interconnected through shared endpoints to form paths that indirectly link nodes that are not directly linked. (Borgatti and Halgin, 2011) A simplified example of a network is shown in the following illustration.

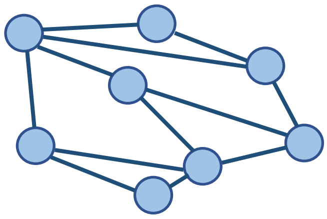

Illustration 3. Example of network

It is important to note that a network is not the same thing as a group. These two terms are distinguished mainly for two reasons. The first one is that, unlike groups, networks do not have natural boundaries. And the second one is that networks do not have to be connected, that is, it is possible that a network is disconnected when some of its nodes cannot reach certain others by any path, so that the network would be fragmented into different components. (Borgatti and Halgin, 2011)

Once the concept of network is understood, it is now necessary to refer to the concept of network theory. A network theory alludes to the mechanisms and processes that interact with network structures to yield certain outcomes for individuals and groups. That is, it refers to the consequences of network variables. (Borgatti and Halgin, 2011)

At this point, the two main theories existing in the literature related to tie detection and tie strength can already be introduced. These are the Strength of Weak Ties theory (SWT) and the Structural Holes theory (SH). Both theories will be addressed and explained in the following parts of this section.

2.4.1 Triadic Closure and the Strength of Weak Ties

The principle that governs this type of models is: “If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future” (Easley and Kleinberg, 2010). This idea is known under the nomenclature of “triadic closure” and it is presented graphically in the following illustration:

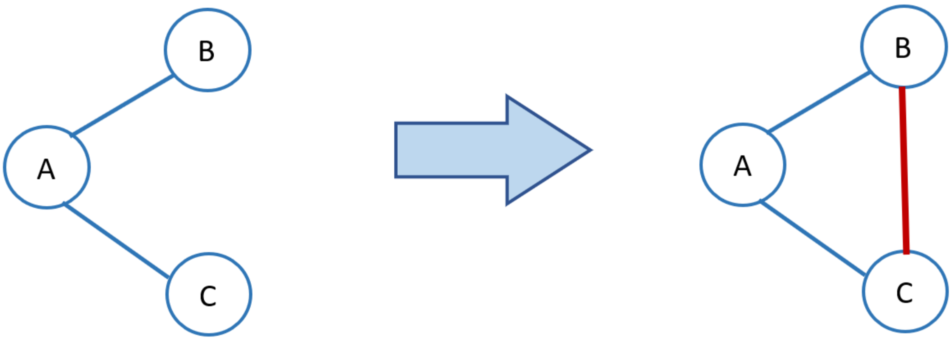

Illustration 4. Triadic Closure

Granovetter´s work shows the idea that people get, or at least hear about, new jobs mostly through acquaintances, that is, through weak ties, instead of strong ties. Therefore, Granovetter argues that strong ties are unlikely to be a source of novel information (Granovetter, 1973; Borgatti and Halgin, 2011). This statement is reached after the assumption of various premises, which led Granovetter to formulate his theory: Strength of Weak Ties (SWT).

These premises are mainly two. The first one is that “the stronger the tie between two people is, the more likely their social worlds will overlap”. This assertion leads to conclude that if A and B have a strong tie, and B and C have a strong tie, then A and C have an increased chance of having at least a weak tie (Borgatti and Halgin, 2011). This links to the concept of triadic closure presented in the beginning of this section. That is to say, Granovetter already introduced this concept within the premises that underlie the formation of his SWT theory.

The second premise on which Granovetter's SWT is based is that “bridging ties are a potential source of novel ideas”. (Borgatti and Halgin, 2011) To understand this, first reference must be made to the concept of bridge. A bridge is understood here as an edge joining two nodes that cannot be deleted without causing its separation in two different components, that is, this edge is the only route between its endpoints, the two nodes (Easley and Kleinberg, 2010). In this way, “a bridging tie is a tie that links a person to someone who is not connected to his or her other friends. The idea is that, through a bridging tie, a person can hear things that are not already circulating among his close friends” (Borgatti and Halgin, 2011).

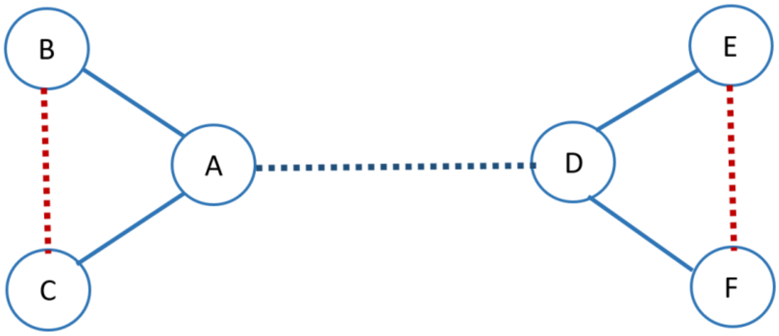

Illustration 5. Strength of Weak Ties. Triadic Closure and Bridging Ties

In the example illustrated above (Figure 2) it can be observed in a very simplified way the application of the two premises explained in the previous paragraphs, which serve as the basis for the construction of the SWT.

In this figure, the solid lines represent strong ties, while the dashed lines represent weak ties. First, it can be observed that the triadic closure principle is met, since, for example, when presenting A and B, and A and C strong ties respectively, B and C present at least one weak tie. On the other hand, it is observed that there is a weak tie between A and D, being this the only way that allows connecting the two endpoints. Therefore, it can be collected under the name of bridging tie, through which it will be possible the exchange of information and resources between two groups of individuals that are not part of the same social sphere, fact by which it can be foresee a bidirectional enrichment of informative and instrumental nature.

2.4.2 Structural Holes Theory

The Structural Holes (SH) theory of social capital is contributed thanks to Burt's work.

Firstly, social capital can be defined as the resources and advantages that result from the social structure, that is in other words, the social structure is a kind of capital that can create a competitive advantage for individuals when it comes to achieving their objectives. (Burt, 1992)

Once the definition of social capital is exposed, the Structural Holes theory provided by Burt puts again the emphasis on weak ties. Burt states that the weaker connections between groups are holes in a social structure and that these structural holes create a competitive advantage for those individuals whose relationships span the holes. (Burt, 1992)

Structural holes between two groups does not mean that individuals in those groups are unaware of one another, but it only means that they are focused on their own activities. Individuals on either side of a structural hole move in different flows of information, therefore, structural holes constitute an opportunity to achieve novel information and resources and gain diversity through bringing together people from opposite sides of the hole. (Burt, 1992)

According to the definition previously given, it can be observed that the description provided for the relationships existing between groups separated by structural holes agrees with the definition of weak ties. Therefore, it can be understood that this theory is based on the clouds of nodes surrounding a given node, as well as the types of strength ties that exist between them. (Borgatti and Halgin, 2011) For the best understanding of this theory, a figure that illustrates this idea is shown below.

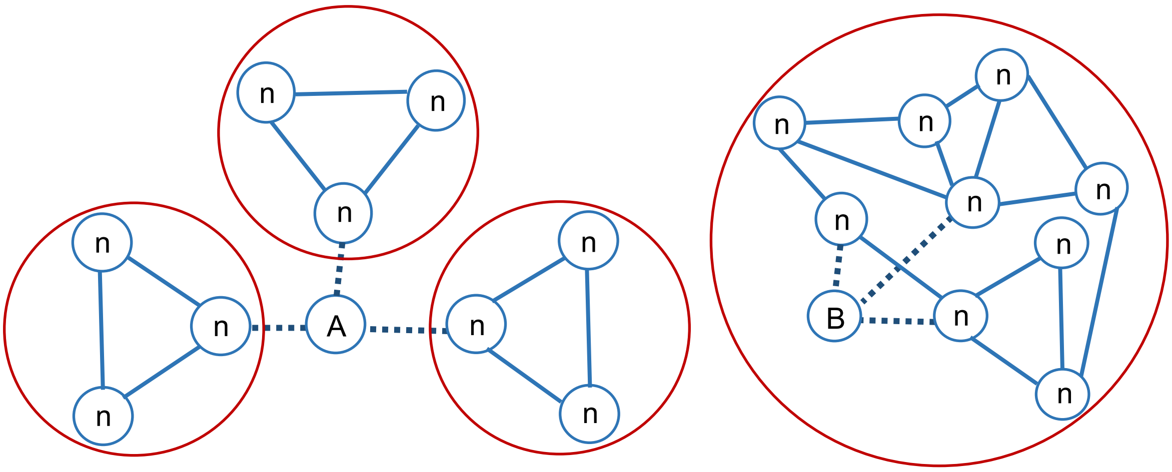

Illustration 6. Structural Holes Theory. Two different situations

In this illustration, two different situations are observed. On the one hand, node A presents connections with three nodes and it can be seen that these connections lead to clouds of independent nodes between them. On the other hand, node B presents all its connections linked to nodes that belong to the same cloud of nodes. In view of this situation, it can be concluded that node A will have greater opportunities to obtain novel and nonredundant information and resources, and therefore, it will be able to benefit from a greater competitive advantage than in the case of the situation presented by node B.

2.4.3 Comparison between Strength of Weak Ties and Structural Holes

It can be concluded that both the theory presented by Granovetter and the one presented by Burt are certainly related, since both focus on the contributions and positive implications of the existence of weak ties within a social network.

That is, it can be considered that between both theories there is a change in the language, but that, nevertheless, the consequences obtained are the same. In this way, in the example presented in Illustration 6, it can be seen that, under the denomination established by Granovetter, A has more bridging ties than B; while using Burt's terminology, A has more structural holes than B, that is, more nonredundant ties. However, in both cases, reference is made to the fact that A has a greater number of ties that provide diversity and more sources from which to obtain new information and resources.

2.5 Dimensions, indicators and predictors of tie strength

Once mentioned the significance and the potential applications in numerous fields (see section 2.3) that supposes the knowledge of the different tie strengths, there arises in this point the need to answer the question: *How can tie strength be measured and based on what factors?* That is why the different dimensions, indicators and predictors to take into account when measuring tie strength are presented in this section.

In this field, many authors have provided with different contributions. Below, some of the ideas presented by some of the most outstanding authors and papers will be presented. And, after having analysed the existing literature, a summary of the dimensions, indicators and predictors that are considered more relevant will be made.

It is essential to start this analysis with the seminal article *The Strength of Weak Ties*. Granovetter already shows, through its definition of tie strength, the main dimensions that comprise this concept. In particular, tie strength is initially defined through the combination of four dimensions: the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie (Granovetter, 1973).

However, after the work of Granovetter, new contributions have been made over the years that have allowed to go in depth in the analysis, so that the list of dimensions and relevant factors of tie strength have increased considerably.

Among subsequent research, authors like Lin et al., who introduce the concept of an individual´s social resources within a social network, stand out. He defines them as “the wealth, status, power as well as social ties of those persons who are directly or indirectly linked to the individual”. Therefore, Lin et al. defend that the access a person has to social resources is also an important factor to take into account (Lin, Ensel and Vaughn, 1981). Lin et al. state that according to the level of social resources available to an individual, he or she will be found in one level or another of the pyramid, being at the upper levels those individuals with greater power in terms of social resources, that is, with a greater " prestige". In general, it can be concluded that Lin et al. affirm that this social distance between individuals is influenced by factors such as socioeconomic status, race, gender, political affiliation or education level that each of them possesses. And it is precisely this social distance, in turn, that mainly influences tie strength (Lin, Ensel and Vaughn, 1981; Gilbert and Karahalios, 2009; Liberatore and Quijano-Sanchez, 2017).

For their part, Marsden et al., in their work "Measuring tie strength", continue with the idea proposed by Granovetter, considering as main dimensions the ones suggested by him. But, in particular, Marsden et al. consider that the measurement of "closeness" or "intesity" is the best indicator of tie strength, which is mainly defined by two factors: the time and depth of the relationship (Marsden and Campbell, 1984).

Marsden et al. difference between two types of variables: indicators and predictors. He considers that indicators refer to real components of tie strength (duration, closeness, frequency, mutual confiding, breadth of topics) (Petroczi, Nepusz and Bazsó, 2007), while predictors refer to aspects of relations, that are related to, but are not components of, tie strength (Marsden and Campbell, 1984). That is, predictors refers to contextual contingencies such as neighbourhood, workplace (Marsden and Campbell, 1984), similar socio-economic status, affiliation (Gilbert and Karahalios, 2009), occupation prestige (Petroczi, Nepusz and Bazsó, 2007), social distance (Lin, Ensel and Vaughn, 1981), recency of communication (Lin, Dayton and Greenwald, 1978), interaction frequency (Granovetter, 1973), possessing at least one mutual friend (Shi, Adamic and Strauss, 2007) or communication reciprocity (Friedkin, 1980).

Wellman and Wortley claim that providing emotional support acts as a signal that indicates stronger ties (e.g., offering advice on family problems) (Wellman and Wortley, 1990; Gilbert and Karahalios, 2009). And Burt defends that structural factors also influence tie strength (e.g. informal social circles or network topology) (Burt, 1995; Gilbert and Karahalios, 2009).

Subsequently, Gilbert et al. conduct a work of analysis and compilation of the information collected and studied so far, and they do so by summarizing the dimensions identified to date, resulting in a list of seven main dimensions. These are: intensity, intimacy, duration, reciprocal services (Granovetter, 1973), structural variables, emotional support and social distance (Gilbert and Karahalios, 2009).

Xiang et al. proposed a model to infer relationship strength based on profile similarity and interaction activity (Xiang, Neville and Rogati, 2010; Liberatore and Quijano-Sanchez, 2017). Up to that moment, to determine tie strength, methods that required the intervention of individuals had been used, based on answers to different questions. However, the suggested model in this work opens a new door, proposing an unsupervised model, which allows to infer a continuous-valued relationship strength for links (Xiang, Neville and Rogati, 2010).

Subsequently, Gilbert et al., in their work *Predicting Tie Strength in a New Medium*, reinforce their theory presented in *Predicting Tie Strength with Social Media* (Gilbert and Karahalios, 2009), by presenting evidence that their Facebook tie strength model can be extrapolated and generalized to new social mediums (e.g. Twitter) (Gilbert, 2012).

Hossmann et al. affirm that social, mobility and communication ties are related (Hossmann *et al.*, 2012; Liberatore and Quijano-Sanchez, 2017), leaving a new contribution when analysing and identifying new indicators or predictors of tie strength and the relationship among them. In particular, Hossmann et al. conclude that the three dimensions of tie strength mentioned (social, meeting and communication) depend on each other (Hossmann *et al.*, 2012).

There are several works that have continued to make contributions to this field, as is the case of Rodríguez et al. (2014), whose study proposes the analysis and evaluation of the context and the strength of the individual's ties by using signs of interaction available form social sites APIs (e.g. retweets or private messages in Twitter) (Servia-Rodríguez *et al.*, 2014).

Therefore, after analysing and reviewing the existing literature, the following sections propose the dimensions, indicators and predictors that have been considered most relevant as a result of the studies and articles published on the topic.

2.5.1 Different dimensions of tie strength

This section presents the main dimensions, which include the main factors to be taken into account when measuring tie strength, as well as a brief explanation of each of them for its correct understanding.

1) Amount of time:

This dimension is mainly measured through two indicators, duration and frequency of contact (Granovetter, 1973; Lin, Ensel and Vaughn, 1981; Marsden and Campbell, 1984; Gilbert and Karahalios, 2009; Liberatore and Quijano-Sanchez, 2017). Referring to the seminal work of Granovetter, it is presented that the higher frequency and duration of interactions between individuals, the greater their feelings of friendship will be between them (Granovetter, 1973).

2) Emotional intensity:

This dimension is defined as the degree, force or amount of strength that something has (Liberatore and Quijano-Sanchez, 2017). Therefore, individuals with a higher degree of emotional intensity will present a higher tie strength (Granovetter, 1973; Lin, Ensel and Vaughn, 1981; Mathews *et al.*, 1998; Gilbert and Karahalios, 2009).

3) Intimacy (mutual confiding):

Petróczi et al. defend this dimension as the most important factor of tie strength (Petroczi, Nepusz and Bazsó, 2007). It refers to the state of having a private or a very personal relationship (Liberatore and Quijano-Sanchez, 2017). This dimension can also be understood as the existence of mutual trust and confiding between individuals, so it is related to indicators such as the willingness to offer support to another person or the breadth of topics discussed (Granovetter, 1973; Marsden and Campbell, 1984; Gilbert and Karahalios, 2009).

4) Reciprocal services:

This dimension refers to actions carried out in common between individuals (Liberatore and Quijano-Sanchez, 2017). This reciprocity dimension assumes that individuals with higher tie strength have greater willingness to share what knowledge, information and resources they have (Granovetter, 1973; Haythornthwaite, 2005; Gilbert and Karahalios, 2009).

5) Structural variables:

This dimension refers to variables such as the overlapping social circles, shared organization affiliation, social homogeneity, network topology or informal social circles

(Alba and Kadushin, 1976; Lin, Ensel and Vaughn, 1981; Burt, 1995; Petroczi, Nepusz and Bazsó, 2007; Gilbert and Karahalios, 2009). Therefore, a higher tie strength tends to connect similar people and that are in the same social structures (Haythornthwaite, 2005; Gilbert and Karahalios, 2009).

6) Emotional support:

This dimension is associated with variables such as offering advice or help in family concerns, or showing empathy or caring about another person (Marsden and Campbell, 1984; Wellman and Wortley, 1990; Petroczi, Nepusz and Bazsó, 2007; Gilbert and Karahalios, 2009). Thus, individuals who possess a high degree of emotional support, present a higher probability of having a stronger tie between them.

7) Social distance:

This dimension refers to variables such as the degree of similarity in educational level, political orientation, gender, race or socioeconomic status (Lin, Ensel and Vaughn, 1981; Marsden and Campbell, 1984; Petroczi, Nepusz and Bazsó, 2007; Gilbert and Karahalios, 2009). Thus, individuals with less social distance between them, have a greater disposition or probability of presenting a higher tie strength.

2.5.2 Indicators and predictors of tie strength

In this section, the main indicators that have been considered the most relevant after the evaluation of the existing literature are presented.

Indicators	References
Closeness	(Marsden and Campbell, 1984, 2012; Perlman and Fehr, 1987; Blumstein and Kollock, 1988; Mathews *et al.*, 1998; Petroczi, Nepusz and Bazsó, 2007)
Duration	(Granovetter, 1973; Marsden and Campbell, 1984, 2012; Perlman and Fehr, 1987; Blumstein and Kollock, 1988; Petroczi, Nepusz and Bazsó, 2007)
Frequency of contact	(Granovetter, 1973; Lin, Ensel and Vaughn, 1981; Marsden and Campbell, 1984, 2012; Perlman and Fehr, 1987;

	Blumstein and Kollock, 1988; Mathews *et al.*, 1998; Benassi, Greve and Harkola, 1999; Petroczi, Nepusz and Bazsó, 2007)
Breadth of discussion topics	(Granovetter, 1973; Marsden and Campbell, 1984, 2012; Perlman and Fehr, 1987; Blumstein and Kollock, 1988; Petroczi, Nepusz and Bazsó, 2007)
Mutual confiding (trust)	(Granovetter, 1973; Marsden and Campbell, 1984, 2012; Mathews *et al.*, 1998; Petroczi, Nepusz and Bazsó, 2007)
Depth of relation	(Marsden and Campbell, 1984)
Mutual acknowledgement of contact/ Reciprocity	(Granovetter, 1973; Friedkin, 1980; Marsden and Campbell, 1984; Perlman and Fehr, 1987; Blumstein and Kollock, 1988; Mathews *et al.*, 1998; Petroczi, Nepusz and Bazsó, 2007)
Multiplexity	(Granovetter, 1973; Marsden and Campbell, 1984; Petroczi, Nepusz and Bazsó, 2007)
Provision of emotional support and aid offered and received within the relationship	(Granovetter, 1973; Wellman, 1982; Marsden and Campbell, 1984; Lin, Woelfel and Light, 1985; Perlman and Fehr, 1987; Blumstein and Kollock, 1988; Wellman and Wortley, 1990; Petroczi, Nepusz and Bazsó, 2007; Gilbert and Karahalios, 2009)
Social homogeneity of those joined by a tie	(Lin, Ensel and Vaughn, 1981; Marsden and Campbell, 1984)
Overlap of social circles	(Granovetter, 1973; Alba and Kadushin, 1976; Marsden and Campbell, 1984)

Voluntary investment in the tie	(Perlman and Fehr, 1987; Blumstein and Kollock, 1988; Petroczi, Nepusz and Bazsó, 2007)
Desire for companionship	(Perlman and Fehr, 1987; Blumstein and Kollock, 1988; Petroczi, Nepusz and Bazsó, 2007)
Sociability/conviviality	(Petroczi, Nepusz and Bazsó, 2007)

Table 3. Indicators of tie strength

On the other hand, also in this section, the main generic predictors considered in the literature are shown. These predictors are those presented in the following table.

Predictors	References
Kinship status	(Feld, 1982; Marsden and Campbell, 1984, 2012)
Neighbour status	(Feld, 1982; Marsden and Campbell, 1984, 2012)
Co-worker status	(Marsden and Campbell, 1984)
Overlapping organizational memberships	(Marsden and Campbell, 1984)
Socioeconomic status	(Lin, Ensel and Vaughn, 1981; Gilbert and Karahalios, 2009; Liberatore and Quijano-Sanchez, 2017)
Political affiliation	(Alba and Kadushin, 1976; Lin, Ensel and Vaughn, 1981; Beggs and Hurlbert, 1997; Petroczi, Nepusz and Bazsó, 2007; Gilbert and Karahalios, 2009; Liberatore and Quijano-Sanchez, 2017)
Education level	(Lin, Ensel and Vaughn, 1981; Gilbert and Karahalios, 2009; Liberatore and Quijano-Sanchez, 2017)

Occupation prestige	(Petroczi, Nepusz and Bazsó, 2007)
Recency of communication	(Lin, Dayton and Greenwald, 1978; Gilbert and Karahalios, 2009)
Interaction frequency	(Granovetter, 1973; Gilbert and Karahalios, 2009)
Possessing at least one mutual friend	(Shi, Adamic and Strauss, 2007; Gilbert and Karahalios, 2009)
Communication reciprocity	(Friedkin, 1980; Gilbert and Karahalios, 2009)
Shared social circles	(Alba and Kadushin, 1976; Beggs and Hurlbert, 1997; Petroczi, Nepusz and Bazsó, 2007)

Table 4. Predictors of tie strength

3 Detection and Evaluation of Ties from Social Media

3.1 What is social media?

When talking about Social Media, reference should also be made to two highly related concepts: Web 2.0 and User Generated Content. On the one hand, Web 2.0 consists of the ideological and technological foundation (Kaplan and Haenlein, 2010), that is, it refers to the tools and technologies that allow users to communicate, create content and share it easily online (Jussila, Kärkkäinen and Aramo-Immonen, 2014). And, on the other hand, User Generated Content represents the sum of all the ways in which people make use of Social Media (Kaplan and Haenlein, 2010).

Therefore, Social Media can be understood as “a group of Internet-based applications that build on the ideological and technological foundation of Web 2.0 and that allow the creation and the exchange of User Generated Content” (Kaplan and Haenlein, 2010; Jussila, Kärkkäinen and Aramo-Immonen, 2014). This digital Social Media is also characterized by being highly scalable, accessible and by operating in real time, that is, it can be considered a tool accessible to everyone, everywhere and at every time (Wollan, Smith and Zhou, 2010). Thus, social media helps users to overcome difficulties that derive from time and distance barriers (Petroczi, Nepusz and Bazsó, 2007).

There are different types of Social Media depending on the functionality offered, the rules of use or how to use it by users (Jussila, Kärkkäinen and Aramo-Immonen, 2014). Among these types, some can be highlighted such as blogs, social network sites (e.g. Twitter, Facebook, LinkedIn), virtual social worlds, content communities (e.g. Youtube), collaborative projects (e.g. Wikipedia) or virtual game worlds (Kaplan and Haenlein, 2010; Jussila, Kärkkäinen and Aramo-Immonen, 2014). However, in this work the focus of study and interest falls on the social network sites, specifically on the Twitter platform.

Social network sites refers to applications that allow users to connect with each other, through the creation of a personal profile through which they can share information, multimedia content; they can exchange messages, ideas, comments, opinions, recommendations; they can create a list of other users with whom they share a connection, etc (Boyd and Ellison, 2007; Kaplan and Haenlein, 2010; Wollan, Smith and Zhou, 2010; Coşkun and Ozturan, 2018; Nisar, Prabhakar and Strakova, 2019). Apart from the differences from a technological perspective (Liberatore and Quijano-Sanchez, 2017), there are numerous social network sites, since they obey very varied objectives, features, interests or practices. Many of them focus on maintaining pre-existing relationships, but others, however, help strangers get in touch based on the fact that they share interests, hobbies, thoughts or tendencies (Boyd and Ellison, 2007).

Another of the most important characteristics about social network sites, and one that is of great interest for the present work, is the fact that through them users are allowed to articulate and make visible their social networks. This allows to create connections between individuals that otherwise would not be connected, so that in some way they allow the formation of "latent ties", a concept that has been explained in section 2.2 (Haythornthwaite, 2005; Boyd and Ellison, 2007).

Particularizing this section of definitions to the platform of interest of study in this work, Twitter can be described succinctly as a web-based microblogging service which allows users to share textual messages of up to 280 characters. These messages are called “tweets” (Servia-Rodríguez *et al.*, 2014).

3.2 Motivation of social media use especially Twitter

In this section it is presented in a general way some of the most important reasons why the analysis, study and evaluation of social media is interesting.

Firstly, reference must be made to the great growth that this new way of communication is experiencing in the last years and to its constantly growing. The following figure (Figure 1) shows a graph that indicates the number of social media users worldwide in recent years, as well as a forecast for the coming years.

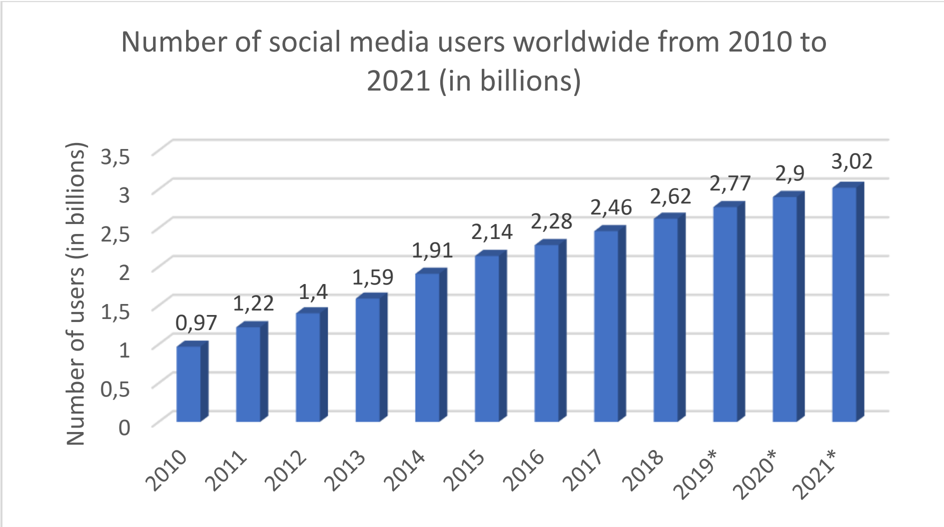

Figure 1. Number of social media users worldwide from 2010 to 2021

In view of these results, it is evident the need that exists today to adapt to new media, being able to participate and get the best benefit from the utilization of social media.

Specifically, in the current work, the analysis of data from the social network Twitter will be performed. Regarding the evolution of this platform, we find the results that are shown in the following graph (Figure 2).

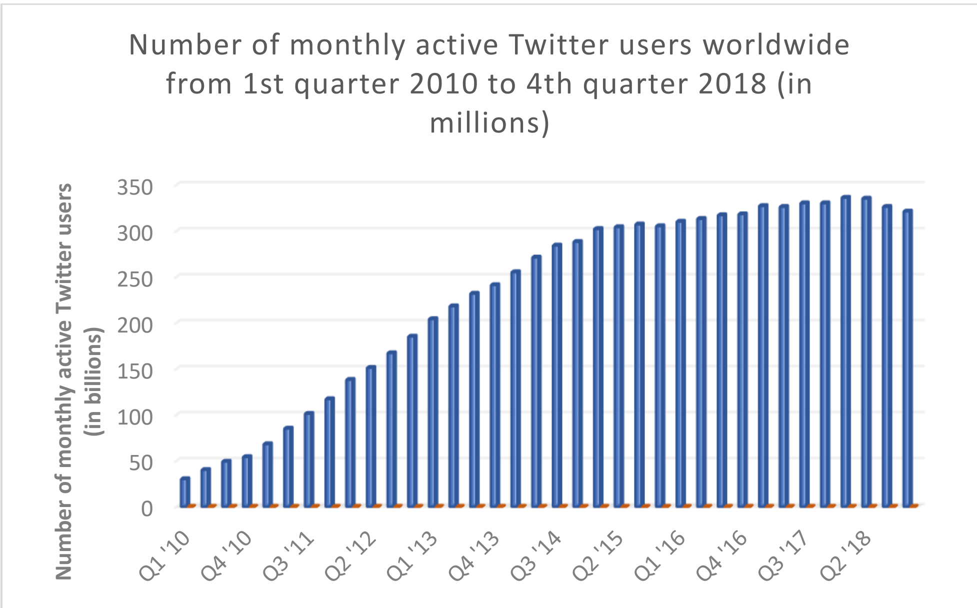

Figure 2. Number of monthly active Twitter users worldwide from first quarter of 2010 to fourth quarter of 2018

This graph shows how the number of Twitter users has multiplied by 100 (from 30 to 321 millions) from 2010 to 2018 (*Number of monthly active Twitter users worldwide from 1st quarter 2010 to 4th quarter 2018 (in millions)* | *Statista*), so it can be concluded the great significance of this social network worldwide.

Due to this widespread growth of social networks worldwide, there are many objectives and benefits that can be achieved through them, both at the personal level and at the organizational level. Social media opens a new door to new opportunities for communication, collaboration, learning and interaction (Jussila, Kärkkäinen and Aramo-Immonen, 2014; Liberatore and Quijano-Sanchez, 2017).

Concentrating on the focus of interest of the current work, social network sites allow managing the ties between individuals. In turn, consequently, it is possible to analyse such ties and perform sociological studies, so that social behaviours can be shown. Such analysis can be beneficial, as is the case, for example, of organizations that take advantage of this information in order to improve their performance to increase their benefits. That is, social network sites provide a source of data on user behaviour (Boyd and Ellison, 2007), which means a source of information that can be very useful in improving the productivity and profitability of some task (Nisar, Prabhakar and Strakova, 2019) or that serve as a potential of benefit to conduct studies based on that data (Coşkun and Ozturan, 2018).

Example and proof of the comments in the previous paragraph is the use of social measures in recommender systems or decision-making processes. These social measures are performed through the analysis of the users´ profiles, as well as of their contact lists and tie strength estimations (Golbeck, 2006; Quijano-Sánchez, Díaz-Agudo and Recio-García, 2014; Liberatore and Quijano-Sanchez, 2017).

Therefore, it can be concluded that, for the current study, the main reason for the analysis of social network sites data (specifically, Twitter data) is to analyse the behaviour of users who seek to maintain existing relationships and network through these platforms. In particular, the large amount of personal information that Twitter users post can be analysed to deduce the tie strength between users (Arnaboldi, Guazzini and Passarella, 2013; Liberatore and Quijano-Sanchez, 2017).

3.3 Measures of tie strength in Twitter

After reviewing the existing literature, it is important to note that one of the earliest studies in which tie strength is analysed from Twitter data is the one done by Gilbert in 2012 (Gilbert, 2012). Following this study, and reviewing subsequent research, the most relevant measures of tie strength that use the Twitter platform as data source are shown below (Table 5).

Dimesnion	Measures using Twitter	References
Amount of time	- Days since last communication - Days since first communication	(Gilbert, 2012)
Emotional intensity	- Initiated @-replies - Direct message headers - @-reply words exchanged - Private messages exchanged	(Gilbert, 2012; Servia-Rodríguez *et al.*, 2014)
Intimacy (mutual confiding)	- Private messages exchanged - Days since last communication - @-reply intimacy words	(Gilbert, 2012; Servia-Rodríguez *et al.*, 2014)
Reciprocal services	- @-reply words exchanged - Mutual followers - Retweets friend´s tweets - Marking as favorite friend´s tweets	(Gilbert, 2012; Servia-Rodríguez *et al.*, 2014)
Structural variables	- Mean tie strength of mutual friends - Sharing the same Hashtag - Taking part of the same list - Retweets the same tweets - Common Followers (network overlap) - Common Followees (network overlap)	(Gilbert, 2012; Servia-Rodríguez *et al.*, 2014)
Emotional support	- @-reply intimacy words - Private messages exchanged	(Gilbert, 2012; Servia-Rodríguez *et al.*, 2014)

Social distance	- Following count - Follower difference - Retweets the same tweets - Sharing the same Hashtag - Taking part of the same list - Marking as favorite the same tweets	(Gilbert, 2012; Servia-Rodríguez *et al.*, 2014)

Table 5. Measures of tie strength using Twitter

The measures indicated in the table above are based on the existing literature; however, for the subsequent analysis of data in the present work, data that can be extracted from Twitter will be indicated, as well as the measures that will be carried out. This is due to some of the measures presented in this table may no longer be available or may have changed or evolved at the time of the current work.

3.4 Implicit networks from Twitter data

As already mentioned, the popularity of social media has become a global trend. And due to this popularity, social media becomes a tool or a means with great potential to conduct the study of online social networks and the communities that emerge from them (Sousa, Sarmento and Rodrigues, 2010).

Social networks allow the creation of explicit social networks through the acceptance of "connection requests". Nevertheless, it is of special interest in the present study to take into account the implicit connections that derive from the actions carried out by users in social networks, that is, activities such as commenting on a photo or a profile, tagging a photo or sending a message are some examples of actions that lead to the emergence of implicit networks. These networks are also known as *activity networks*, since they are networks that derive from the current interactions between users, rather than from the mere declaration of friendship (Sousa, Sarmento and Rodrigues, 2010). Specifically, in the case of Twitter, it can be considered activities or actions such as retweets, mentions to other users, answering a comment or photo, reacting to a publication, etc.

In the present work, the focus is on the construction of the implicit networks that emerge from the mentions made between users on Twitter, specifically in the context of conferences. That is to say, the implicit networks of study are the networks of mentions that are obtained taking into account the mentions made by some users to others, being these indicated with the "@" symbol within the published tweets. In this way, directional networks are obtained, in which a user of origin mentions a destination user.

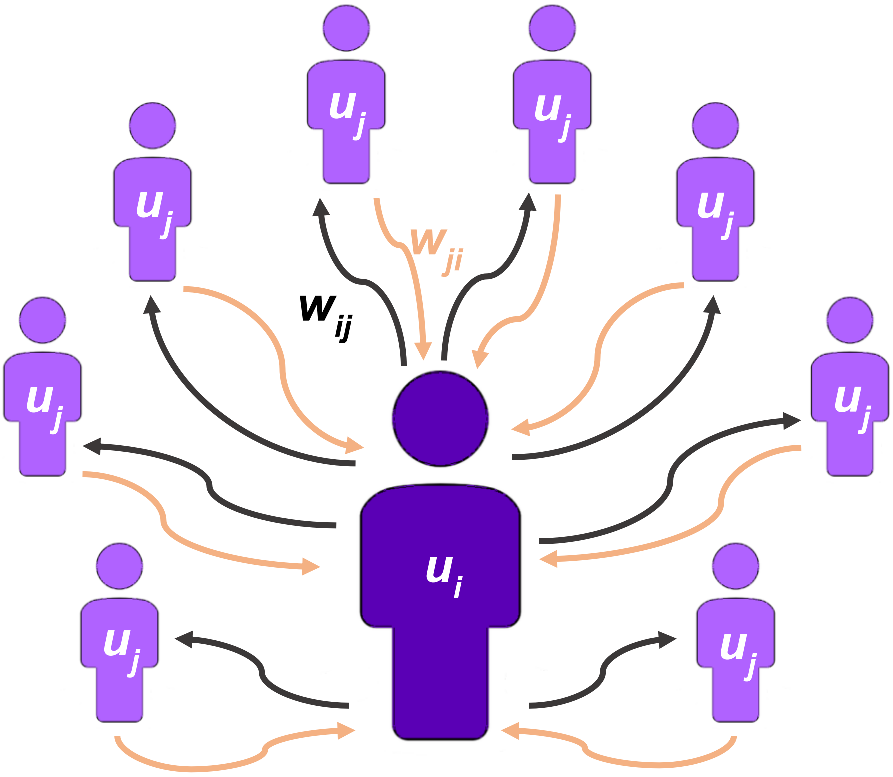

Illustration 7. Ego-centric implicit network of user u_i, representing mentions interactions with users u_j

To simplify the visualization of the concept, in the previous illustration (Illustration 7) an example of an ego-centric implicit network constructed according to the mentions interactions is graphically indicated. As can be observed, the links between users are shown with the indication of the directions of such connections and, in turn, each edge in the illustration has an associated weight *w*, which is directly proportional to the number of mentions made.

Therefore, as indicated, in the present work the focus is on the analysis of implicit networks, specifically, in the analysis and evaluation of mentions networks that emerge from the interactions between users through the Twitter platform in the context of a conference.

4 Significance of Conference Setting for Tie Strength

4.1 Motivations for attending conferences

In this section the main reasons why people attend conferences are presented. In this sense, this section helps to explain the relevance of tie strength evaluation in an event, such as a conference.

To address this section, reference is made to Severt's work (Severt *et al.*, 2007), in which it is indicated that there are five main dimensions that encompass the motivating factors for attending a conference. These five dimensions are: networking, activities and opportunities, convenience of the conference, education benefits and products and deals (Severt *et al.*, 2007).

Specifically, within these five dimensions, the most relevant for the current study are networking and education benefits, since tie strength evaluation is particularly important in them.

Networking dimension refers to the possibility and opportunity to meet and exchange thoughts and ideas with other people who may share similar interests or may contribute with significant information or knowledge. In this way, participants can establish contacts that may be useful in the future; a utility that can materialize, for example, in new job opportunities (Severt *et al.*, 2007; Zhang, Bhardwaj and Karger, 2016). Some authors stress the significance of networking as one of the most important motives for competing conferences (Oppermann and Chon, 1997; Rittichainuwat, Beck and Lalopa, 2001; Malek, Mohamed and Ekiz, 2011). This statement helps to understand the importance of the analysis and evaluation of tie strength in a conference setting.

On the other hand, educational benefits dimension is related to the educational opportunities that a participant has when attending a conference. That is, for example, the possibility of discover valuable information and knowledge or access to career enhancement opportunities, among others (Severt *et al.*, 2007; Zhang, Bhardwaj and Karger, 2016). In this sense, this dimension also becomes important in tie strength analysis among participants between whom these information and knowledge flows occur, being able to evaluate, for example, what degree of tie strength provides more valuable information or resources in a conference setting.

4.2 Current ways of identifying potentially useful contacts

This section presents the current ways in which participants find potential useful contacts in a conference.

After reviewing the existing literature, it can be concluded that, in general, there are three ways of identifying potentially useful contacts in a conference setting. The first one focuses on the work of the organizers, the second is based on the chance that a fortuitous encounter occurs with a potentially relevant person, and the third is based on the use of technology as a means to facilitate relevant connections.

Regarding the first of the ways described in the previous paragraph, the conference organizers adopt the role of planning and designing the conference structure, as well as the framework and aspects surrounding that event. In this way, the conference organizers contribute to the generation of potentially useful connections through the use of practices that facilitate such meetings. That is, activities such as the collection and analysis of data from previous conferences and participant feedback, as well as the use and application of such information in the conference planning, contribute to meetings between people who have a high probability of useful connections being formed between them. A practical example that allows to visualize this organizers´ task would be the disposition of participants in different coffee tables according to the previously gathered information about their interests, so that they could initiate fruitful discussions (Aramo-Immonen, Jussila and Huhtamäki, 2014, 2015; Aramo-Immonen *et al.*, 2016).

The second of the proposed ways refers to the fortuitous encounter between two people potentially generating a relevant connection. That is, sometimes encounters are produced by chance, however, this type of encounters do not have a high probability of occurrence, since no system or procedure is followed whereby people with similar interests coincide in the same space and time of a conference (Ross *et al.*, 2011).

Finally, the last of the proposed ways is based on the use of technology as a facilitator of useful connections. This type of technology based mechanisms focuses on recommendation systems that, through the collection and evaluation of information about the participants, propose suggestions to meet other participants that are outlined as possible useful contacts (Hornick and Tamayo, 2012; Zhong, Yang and Nugroho, 2015; Zhang, Bhardwaj and Karger, 2016).

4.3 Current ways of Twitter data use in conferences

The aim of this section is to summarize the main uses of the Twitter platform and Twitter data in a conference setting.

Firstly, it is important to highlight here the important role played by social media, specially Twitter, in the performance and development of a conference. Twitter is presented as a tool for information dissemination (Aramo-Immonen, Jussila and Huhtamäki, 2014, 2015; Aramo-Immonen *et al.*, 2016) that allows the participation and interaction of all the participants through non-verbal communication, in real time, and without interrupting the main channel communication during the conference, that is, without disrupting the speaker (Ross *et al.*, 2011; Jussila *et al.*, 2013).

Furthermore, it should be noted that there are different types of Twitter users in a conference setting. In particular, the following roles can be distinguished in a conference setting: organizer, speaker, attendee and online attendee (Reinhardt *et al.*, 2009; Ross *et al.*, 2011), and they can all become Twitter users within the context of the conference.

Taking into account the two previous paragraphs, we conclude the idea of transition from "traditional conferences", in which most of the participants embodied a passive role of mere listening to the speaker, towards a new way of understanding the conferences, in which each of the participants becomes an active piece of the puzzle, being able to participate by asking questions, providing information or commenting (Ross *et al.*, 2011; Jussila *et al.*, 2013).

In addition, the above advantages are not limited to the time of the conference itself, but extend to the previous and subsequent moments of it. This provides greater possibilities by allowing the configuration of a more personalized conference content for attendees, as well as greater opportunities for networking and subsequent maintenance of the relationships established in a conference setting (Reinhardt *et al.*, 2009; Ross *et al.*, 2011; Jussila *et al.*, 2013, 2014).

After reviewing the existing literature, it can be concluded that the main uses of Twitter in a conference setting are framed within one of the following three broad categories: communication between conference participants, communication between the audience and speakers or organizers, and for reporting to non-attendees about the conference (Ebner, Rohs and Schön, 2010; Jussila *et al.*, 2013). And within such uses, some of them can be highlighted, including commenting on presentations, establishing discussions, formulating questions, efficient sharing of information, sharing resources or jotting notes (Reinhardt *et al.*, 2009; Ross *et al.*, 2011; Jussila *et al.*, 2013).

All this data that is published through the Twitter platform allows it to be used in a way that supposes some beneficial contribution for the user. That is, both organizers and the rest of the conference participants can use Twitter data in the context of conference setting for different purposes. As for the organizers, it is highlighted the use of such data for performing tasks such as creating tailored conference content, recognizing relevant participants, identifying interesting topics, enhance tailored networking opportunities or improve the organizing and planning of conferences (Ebner, Rohs and Schön, 2010; Jussila *et al.*, 2013). On the other hand, the rest of the participants can use Twitter data in activities such as obtaining information and relevant sources, developing professional career or establishing and maintaining new relevant contacts (Ebner, Rohs and Schön, 2010; Ross *et al.*, 2011; Jussila *et al.*, 2013).

4.4 Possibility for a tie strength-based approach

This section seeks to explain the relevance of the incorporation of social media analysis techniques in a tie strength evaluation approach in the context of conference setting.

Firstly, referring to the aforementioned, one of the main reasons, and the one of greatest interest for the current work, for which people attend conferences is to be able to access and meet new people, being able to form new connections that will be potentially useful for both parties. That is to say, the focus is on networking.

Furthermore, as already indicated, networking is facilitated thanks to the use of recommendation systems. And, in turn, these recommendation systems are nourished by the information and data that is collected through social media, specifically Twitter.

Nevertheless, at this point, it is questionable whether it would be possible to improve such recommendation systems, since up to now, according to the literature, the data analysis methods used mainly focus on data from the registration forms of the participants (Hornick and Tamayo, 2012; Zhong, Yang and Nugroho, 2015). For this reason, the idea of incorporating a tie strength evaluation based approach on recommendation systems arises at this point, in a way that contributes to a better and more efficient networking task and can increase the chances of contacting potentially useful people in a conference setting, that is, achieve better recommendations.

5 Research Methods and Data Collection

5.1 Introducing the research methods

5.1.1 Case study approach

In the present work, the research methodology known as "case study approach" is followed. The term "case study approach" can be defined as a tool that "allows in-depth, multi-faceted explorations of complex issues in their real-life settings" (Sarah Crowe, Kathrin Cresswell, Ann Robertson, Guro Huby, 1993). In this document, the approach is specifically oriented to a single-case design, in particular, a longitudinal case, since the same case is studied at various points in time.

Specifically, in the present work, data sources (Twitter data) are used, which, through different methods related to social network analysis, allow to draw conclusions and help to answer the research questions posed. In section 5.2 of this chapter, the case selection is deepened, as well as the obtaining of the data and the nature of them.

5.1.2 Social Network Analysis

Social network analysis provides a powerful model for social structure by utilizing network and graph theories (Otte and Rousseau, 2002). As already mentioned in section 2.4, the structure of a network is composed of nodes and edges that connect them. And, therefore, social network analysis is used to visualize and study these social structures.

In the particular case of the present work, the analysis focuses on the study of the relations among conference participants (nodes) and the interactions that are established among them (edges), so the use of social network analysis is appropriate for the present study.

To conduct this analysis, the tool used to obtain the visualization of the network and to analyse different parameters of it is Gephi. Gephi is an open source software for exploring and manipulating networks, which allows its visualization and analysis in order to obtain conclusions from it (Bastian, Heymann and Jacomy, 2009). In the results section (chapter 6) of this document, the networks obtained with this tool are shown, based on the user mentions in tweets. Also, in that chapter, the different parameters analysed thanks to Gephi are explained, showing in turn the results obtained.

However, in order to use Gephi, it is firstly necessary the construction of the network of interest from the available data (see section 5.2.3 of this chapter). To carry out this process, the programming language PYTHON has been used, with which, through the integrated development environment Spyder (*Scientific Python Development Environment*) that is included with Anaconda (a free and open-source distribution of the Python and R programming languages) (*Spyder — Anaconda 2.0 documentation*; *Anaconda Distribution — Anaconda 2.0 documentation*), the mentions networks have been built. The programming code used for this task is presented in Appendix 1.

In addition, the data analysis and visualization platform called Tableau has also been used. This tool helps to organize and interpret raw data. In particular, this tool has been used to conduct a part of the descriptive data analysis, in which the Twitter activity levels of conference participants (nodes) are analysed (see section 6.1.6).

5.2 Conducting the research

5.2.1 Case selection

The case study selection in the present work is motivated by the main objective of the study, that is, the evaluation and analysis of tie strength using publicly available social media data (specifically Twitter) in a conference setting. The case selected to achieve this goal is the case HICSS, *The Hawaii International Conference on System Sciences*.

HICSS is one of the longest-standing scientific conferences (*ScholarSpace at University of Hawaii at Manoa: Hawaii International Conference on System Sciences - HICSS*). This fact of prolongation in time through different annual editions makes possible the fulfillment of one of the requirements for the achievement of the proposed objective. That is, as it is a conference held over several consecutive years, it is possible to perform the longitudinal analysis sought through the data of each of the editions.

Moreover, given the topic of this conference, framed in an environment of technology, information, computer and system sciences (*ScholarSpace at University of Hawaii at Manoa: Hawaii International Conference on System Sciences - HICSS*), it can be can presumed the willingness of participants to use social networks as communication platform during the conference. This predisposition translates into a high level of activity in social networks, a fact that increases the data with which to work and allows the obtaining of networks with more complete information and that reflects more closely the reality of the conference context.

In addition, being a conference embedded in a learning and interactive work environment (*ScholarSpace at University of Hawaii at Manoa: Hawaii International Conference on System Sciences - HICSS*), it is presumed in turn the predisposition of the participants to networking, looking for contacts, resources and information with potential utility for them.

Last but not least, another of the main reasons for choosing this conference is the possibility of accessing the data. That is, the social media data related to this conference is publicly available, which is essential to carry out the desired analysis.

Therefore, the compendium of these four main ingredients (long-standing conference, participants with profiles predisposed to the use of social networks and networking, and publicly available data) makes the case HICSS a choice that meets the requirements of the analysis sought in the present work.

5.2.2 Data collection

Obtaining the necessary Twitter data for the analysis is carried out through the use of the APIs (*Application Programming Interface*) of Twitter. In order to access these APIs, the creation of a Twitter developer account is firstly required.

The Twitter APIs allow access to different types of data. For the present study, the focus is on the datasets that can be obtained from the standard version of the APIs, that is, the free and public version. With it, it is possible to access information related to accounts and user´s profiles, which are translated into metadata that include information such as user´s names, their description, or their place of origin, among others. Also, it contains information related to tweets, being able to filter them through the realization of searches for specific keywords or requesting a sample of Tweets from specific accounts.

Originally, the Twitter APIs were mainly classified into three large groups, which are described below (Kumar, Morstatter and Liu, 2014; Weller *et al.*, 2014; Pfeffer, Mayer and Morstatter, 2018):

The Streaming API

- It is a push-based system: it provides a subset of tweets in real time.
- There are 3 different bandwidths:
    - “Spritzer”: 1% of the tweets.
    - “Gardenhose”: 10% of the tweets (not generally available).

    - "Firehose": 100% of the tweets (not generally available).
- There are 2 different methods, as points of access to data:
    - Sample: up to 1% or 10% of all tweets, selected at random.
    - Filter: the track, follow, and locations parameters can be used to select specific results from the stream.
        - Track: only returns tweets that include those words.
        - Follow: only tweets from a set of users represented by their collective comma-delimited user IDs.
        - Locations: for researchers interested in geographically bounded research.

The REST API

- It is a pull-based system: it allows to access the core Twitter data.

The Search API

- It is a pull-based system: it allows to access Twitter search.
- It is possible to filter the search using, for example, language or localization.

However, over the years, these APIs have been restructured under different denominations, as well as new features have been added that, in general, are under new payment requirements. That is, the new categories of "enterprise" and "premium" are added to the catalogue, in addition to the existing "standard" category, which is free and public and provide basic query functionalities and foundational access to Twitter data.

However, as already mentioned, this work focuses on the use of the most accessible category of Twitter APIs, so that the APIs of interest are currently framed under the new name indicated below:

- **Filter realtime tweets:** new way to call the Streaming API.
- **Search tweets:** new denomination for the Search API.
- **API reference index:** that contains what was formerly called the REST API.

Furthermore, in addition to this restructuring of denominations, it is important to take into account some important changes that have taken place over the years in terms of the queries that can be made by the Twitter data requestors to the platform through the APIs. Specifically, the most prominent changes refer to the purpose of increasing the protection of user data, thereby restricting access to them. Some of the most important changes are shown in the table presented below:

<table>
<tr>
<td rowspan="3">2006 - 2010

(API v1)</td>
<td>- The Streaming API:

It is a continuous stream that provides tweets in real time. The speed of reception of tweets has fluctuations that depend on the bandwidth of the two ends of the connection and the overload of the Twitter servers.

In its standard free mode, this API allows access to 1% of the total of tweets, which would be sufficient for the context of conferences as is the case of this work.</td>
</tr>
<tr>
<td>- The REST API:

The most important characteristic to keep in mind within this type of API is that it presents a significant restriction, and this is that it is a rate-limited resource.

During this time period, the existing limitation consists of 150 requests per hour and user (350 when logged in to Twitter via OAuth).</td>
</tr>
<tr>
<td>- The Search API:

While, in theory, some historical collection of data is possible through the Search API, in practice its utility is severely limited due to the data availability time limitation of seven days. So, the most important characteristics are the following:

o Data availability time limitation: within seven days of being posted
o Does not require authentication</td>
</tr>
<tr>
<td>2010 - 2012

(API v1)</td>
<td>From this moment, all third-party applications that request user data must authenticate to the Twitter API using the OAuth protocol. The main motivation for performing this measure is for security and protection reasons of user data.</td>
</tr>
<tr>
<td>2012 - 2013

(API v1.1)</td>
<td>OAuth authentication became mandatory for all endpoints and a new API version v1.1 is released.

In general, the main changes in Twitter API version v1.1 include:

- Required authentication on every API endpoint:</td>
</tr>
</table>

	In version 1.0, it was possible to access certain API endpoints without authentication, which allowed access to public information from the Twitter API without being identified. To avert malicious uses of the data and improve their protection and security, in version 1.1, authentication is required on every API endpoint. - A new per-endpoint rate-limiting methodology: In version 1.0, the number of authenticated requests was limited to 350 per hour and user, regardless of the type of information requested. Nevertheless, in version 1.1, there is a limitation of the rate depending on the endpoint of the API in question. Thus, most individual API endpoints are rate limited at 60 requests per hour and endpoint. However, there is a set of high-volume endpoints related to Tweet display, profile display and user lookup where it is possible to make up to 720 requests per hour and endpoint.	
2013 - Present **(API v1.1)**	New limitations in requests rates	- **Filter realtime tweets** (The Streaming API): o Limitation: being a continuous flow, the restriction is applied to the received flow rate, which will never be greater than 50 tweets per second
		- **API reference index** (The REST API): o Variable limitation depending on the requested method. This restriction is measured in requests per 15 minutes, and the values range between 15 and 900 (the most restrictive methods are those that provide the lists of followers and followed by 15 requests, and the least restricted are the queries for user tweets by 900 requests) o The REST API still has the same features, but it is less precise and exposes the data in an aggregated or limited manner

		- **Search tweets** (The Search API): ○ Limitation: 180 requests per 15 minutes

Table 6. APIs evolution

Specifically, for the present work, the API used is the REST API, and the collected data obey the tweets collected under the hashtag "#hicss" followed by the corresponding year in each case. This data is obtained in *.json* format, which is processed thanks to the PYTHON programming language, as already indicated in section 5.1.2.

5.2.3 Dataset description

In order to conduct the evaluation and analysis of the data, it is necessary first to clean them. For this, the first essential step is to know the structure presented by these data and to understand the meaning of each of its variables.

The data present a structure made up of nested lists and dictionaries. The main list shows an enumeration of tweets, which in turn contains diverse information about each tweet. That is, the main structure is as shown in the Illustration 8 presented below.

```
[
{information related to tweet 1}
{information related to tweet 2}
                :
{information related to tweet n}
]
```

Illustration 8. Main structure of available data

Inside "information related to tweet X" the information related to each tweet appears, information that includes the publication time, data about the user's profile, information related to retweets and mentions, or the content of the tweet, among others. Therefore, once the data structure is understood, it is necessary to select the data that are of interest for the purpose of this study.

As previously mentioned, this work focuses on the construction of mentions networks, so the information that is useful is the one related to this purpose. That is, it is proceeded to build a code that allows the construction of networks in which an origin node (user) mentions (presents a directional connection to) another destination node (user). The code implemented for such mentions networks construction task is presented in Appendix 1.

6 Results of the Analysis

As already mentioned in the previous chapter, in order to visualize the connections and determine the type of relationship that exists between the users that participate in the conference, a tool called Gephi has been used. This tool allows to build graphs that represent networks made up of nodes and links between them. Specifically, in the present work, the analysis has been conducted through the construction of mentions networks in the context of a conference setting thanks to the Twitter platform. In other words, to analyse the relationships between users, it has been chosen to explore the potential of implicit networks, specifically, the mentions made by each of the users in their tweets have been used. In this way, with the help of PYTHON programming language, networks have been built which, thanks to the Gephi tool, have been converted into graphs that facilitate their visualization.

The tasks mentioned in the previous paragraph have been conducted for different consecutive years of the same conference in order to obtain a more general and solid vision about the conclusions that can be obtained after making the graphs. In particular, as already mentioned, the object of study has been the Hawaii International Conference on System Sciences (HICSS) (an annual conference for Information Systems and Information Technology academics and professionals sponsored by the University of Hawaii at Manoa) since 2010, until 2018.

Therefore, this section aims to show the results obtained after the analysis of the mentions networks obtained. For this, three subsections are presented below with which it is intended to give a complete view of the results. In particular, these three subsections obey a structure in levels, in such a way that it starts from a global vision, to end up in a detailed vision of the networks.

In the first part, the objective is to give a general view of the evolution of the networks throughout the different editions of the conference of study, showing possible trends and behaviours. The second part of the analysis focuses on the study of specific cases to understand the interpretation of the results obtained. That is, in this second part, the study is focused on three years (2014, 2015 and 2016), in order to deepen the analysis and show conclusions after its interpretation. And, finally, the third level of analysis focuses on an example of a specific network, seeking to give a more precise and qualitative interpretation of the results.

6.1 Longitudinal descriptive analysis

Firstly, as indicated, a descriptive and global analysis will be conducted to obtain a general vision of the evolution of the networks throughout the nine years of study. This analysis includes aspects such as the number of users and the number of connections, as well as certain defining parameters of networks that have been considered relevant for the study of these networks.

6.1.1 Number of nodes and edges

In the first place, the evolution over the nine years of study has been analysed for both the number of nodes and the number of connections or edges. The results obtained from this analysis are shown in Figure 3 and 4.

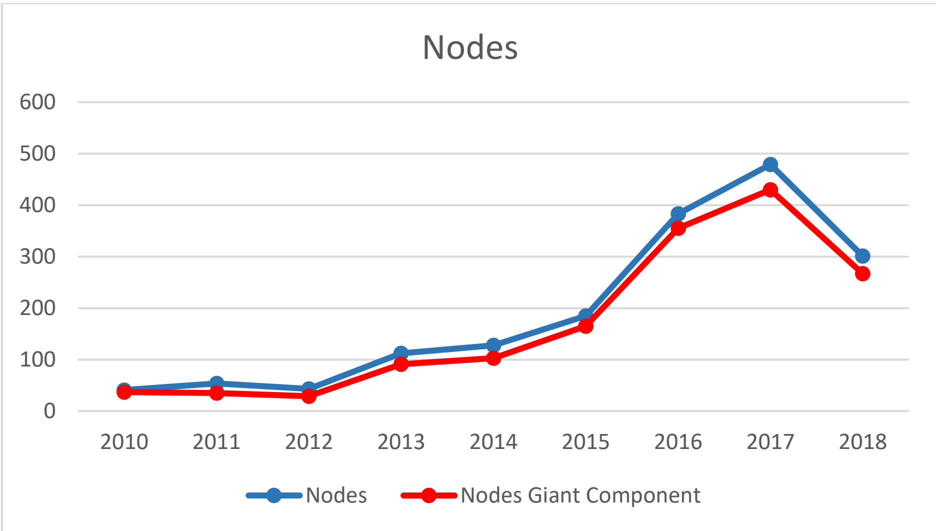

Figure 3. Evolution of number of nodes

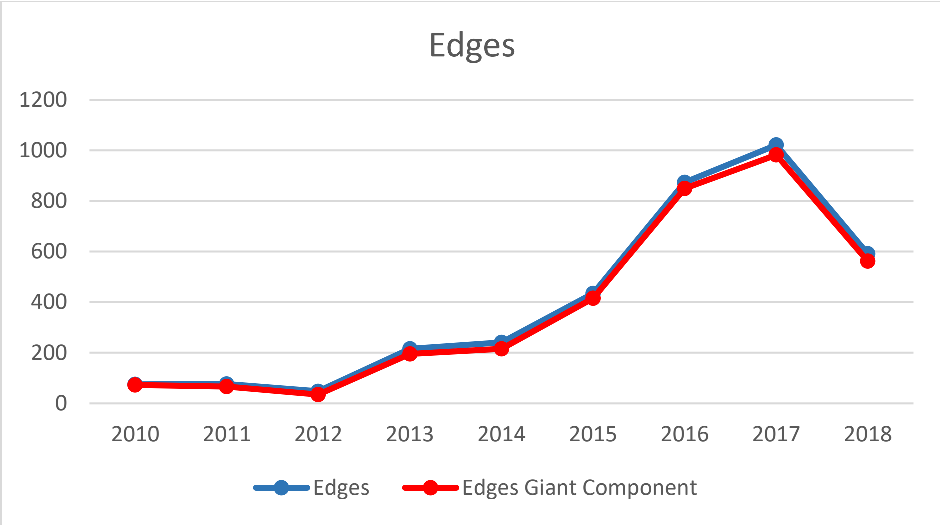

Figure 4. Evolution of number of edges

As can be seen, two series appear in these graphs. This is due to the fact that the same analysis has been conducted for both the entire network and the network formed only by what is known as "Giant Component". This "Giant Component" is that part of the network that constitutes the central cloud or mass, eliminating nodes that orbit around it. In this way, the values obtained for the original network are indicated with blue colour, while the values obtained once the Giant Component filter has been applied are red. This has been done in order to compare both networks and see if the minority of orbiting nodes distort the results obtained or not.

In both cases, a similar trend is observed, so that similar conclusions can be drawn from both results. In particular, a general tendency to the appearance of a greater number of users over the years can be concluded, as well as a greater number of connections between them. However, it should be noted a fall in these values in the last year, 2018, falling back to values lower than those obtained in 2016.

The most immediate interpretation for this isolated case that can be considered is the loss of interest on the part of the users in the conference of study. However, going back to the section explaining the motivation of using Twitter for the present study (section 3.2), a comparison can be made between Figure 2, which shows the active Twitter users worldwide from 2010 to 2018, and Figure 3 with the number of nodes in the analysed network. In this comparison, the coincidence in temporary terms of the decrease in the number of users is highlighted. This observation leads to raise a new motivation for such a decline of nodes in the network built in this section, that is, this decline in 2018 can also be derived from a general decline in the use of Twitter.

Once the magnitude of the networks and their evolution have been visualized, the next step is to analyse some of the most outstanding parameters, so that the characteristics that define the networks in each of the years of study can be understood and possible trends can be observed. Again, the visualization of the obtained results has been conducted through graphs in which the original network is compared with the filtered network constituted by the central cloud of nodes (Giant Component).

6.1.2 Average Degree

The first parameter to be analysed is the Average Degree. The degree of a node can be defined as the number of edges that are adjacent to that node, that is, it is the sum of edges of a node. In this way, and thanks to the Gephi tool, the average of the degree of each node has been obtained for each of the nine years of study. The values obtained are reflected in Figure 5 shown below.

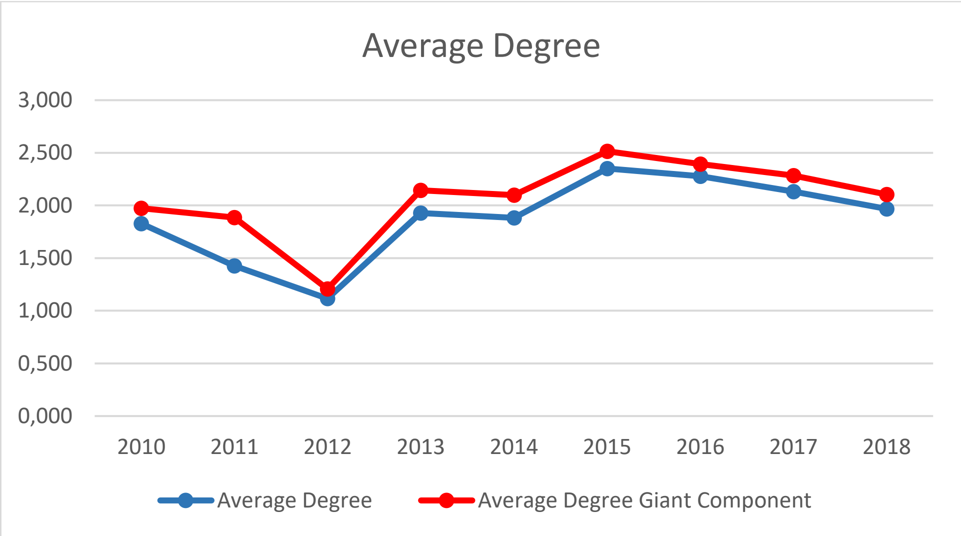

Figure 5. Evolution of Average Degree

In both cases (original network and Giant Component) a similar trend is observed, so that similar conclusions can be drawn from both results. In particular, it can be seen that the range of values is between 1 and 2,5, so, although the graph has certain peaks given the scale of the axes, these values remain fairly similar over the years of study. From this it can be concluded that, in general, nodes in the networks obtained have a similar structure in terms of number of connections, being able to see slight increases that coincide with periods of increase in the number of nodes.

6.1.3 Average Weighted Degree

Following the previous idea, the second parameter analysed is the Average Weighted Degree. Since the previous parameter did not offer much information, an attempt to go a step further in that direction has been made by now analysing the weighted degree of the nodes. This parameter is also based on the number of edges for a node, however, in this case such value is ponderated by the weight of each edge. That is, the sum of weights or intensity of each connection of a node is performed. The average of those sums for each node is the result shown in the following graph (Figure 6).

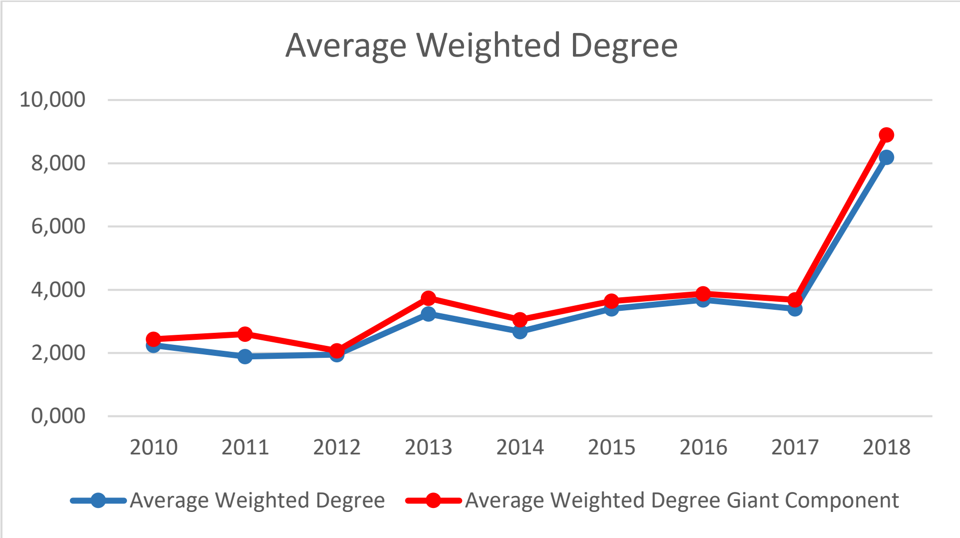

Figure 6. Evolution of Average Weighted Degree

Despite not obtaining an excessively wide range of values, in this case a certain positive trend can be observed. In particular, it highlights the value obtained in the last year, 2018, in which values greater than 8 are reached. It is interesting to compare this result with the one obtained for the number of nodes, since it is observed that, in spite of the decrease in the number of users that appears in 2018, there is an increase in the weights of the connections that year. This can be interpreted as the nodes that have ceased to be part of the network in 2018 in the context of the conference are nodes that formed low weight connections with the rest. In other words, in view of the results, it can be interpreted that there has been a "filtering" of users in 2018, so that some of the weakest connected to the network or to the rest of nodes have stopped participating in it in the last conference edition.

6.1.4 Graph Density

The next parameter analysed is Graph Density. This parameter measures how close the network is to be "completed". A complete graph has all possible edges and its density is equal to 1. The results obtained are shown in the following graph (Figure 7).

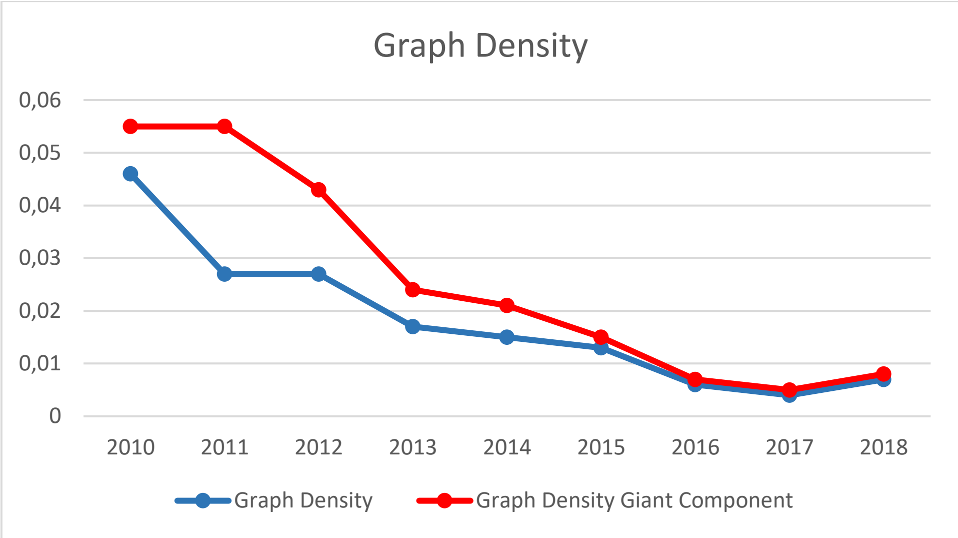

Figure 7. Evolution of Graph Density

When visualizing the obtained results, it can be observed that in networks with the largest number of nodes, the density of the graph decreases. That is, with the increase in the number of nodes, the number of edges does not increase enough to maintain the density of the network, but rather that the increase follows an approximately linear trend, as shown in Figure 8.

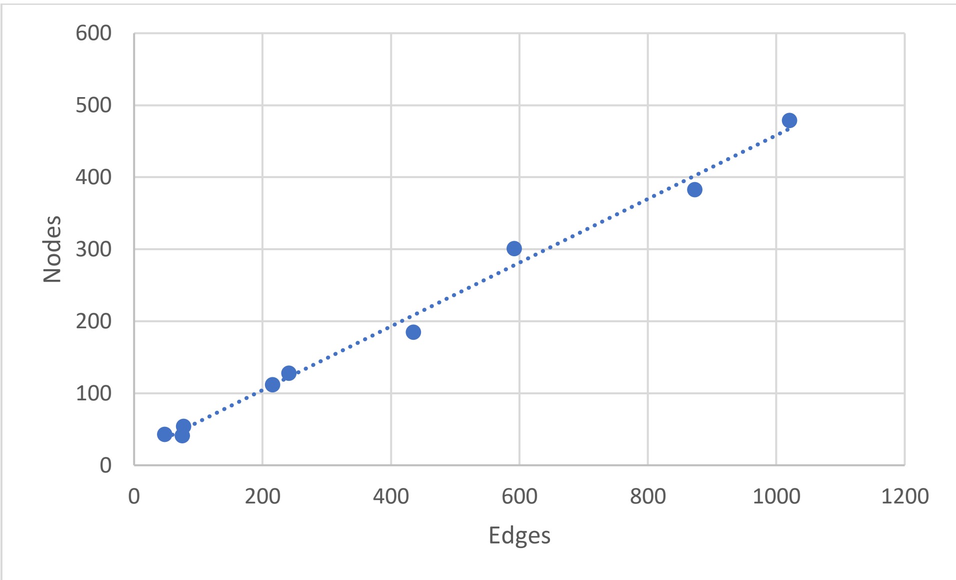

Figure 8. Tendency of growth of edges with respect to growth of nodes

Therefore, a negative trend is concluded regarding the density of the network. And with this, it is deduced the incorporation of nodes with similar connection characteristics to those already existing nodes. That is, new nodes have a similar number of connections to the existing ones and, therefore, not enough to equal the increase of connection possibilities between them.

6.1.5 Number of communities

Finally, it is also analysed the number of communities that results in each network after applying the modularity algorithm provided by the Gephi tool. This number of communities results from the division of the network into sub-networks according to the existing connections between the nodes, obtaining groups of nodes that have similar characteristics. That is, the formation of clusters takes place, taking into account that nodes belonging to a cluster must be as similar as possible, and nodes belonging to different clusters must be as different as possible in terms of the connections they have.

Therefore, the number of communities serves as an indicator of the number of "profiles" that can be found within the network and, therefore, serves as a measure of diversity within it. The results obtained are shown in Figure 9.

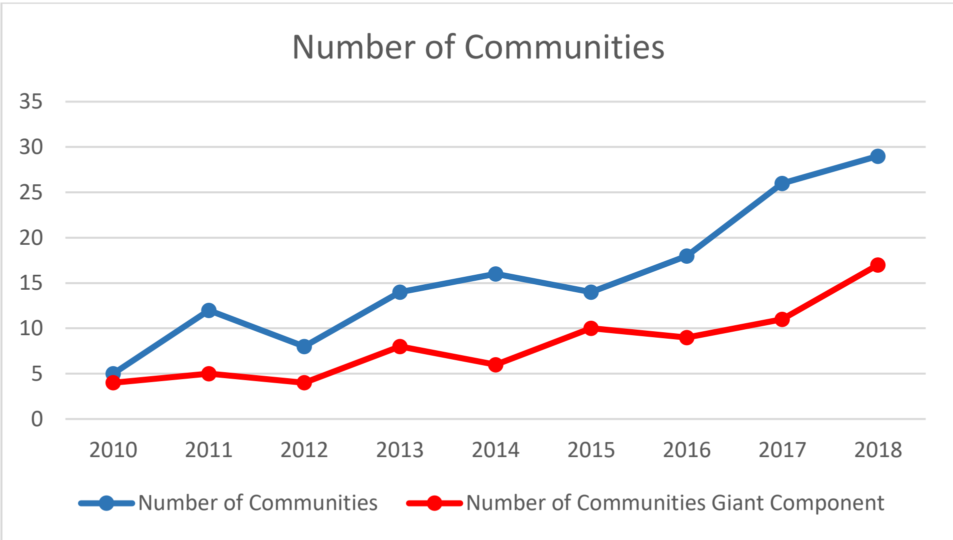

Figure 9. Evolution of number of communities

As can be seen, the trend follows a positive increase according to the increase in the number of nodes in the network. Therefore, taking into account this result and that obtained when analysing the previous network density parameter, it can be concluded that, although it is true that the nodes that are incorporated into the network over the years present similar characteristics in regarding the number of connections, these connections do not occur in the same environment or context. That is, the interpretation of the results shows that new nodes appear with different characteristics from those already existing, so that the formation of new communities that reflect and identify in a more approximate or faithful way the characteristics and features of each node emerge.

6.1.6 Activity Levels

To improve the contextualization of the situation in each of the editions of the conference of study, within the descriptive analysis, it is also interesting to analyse the activity levels before, during and after the conference. For this, it is proceeded to the construction of timelines of the number of tweets in each of the years of study. The tool used to perform this task is Tableau, with which the timelines obtained from the number of tweets per day are presented in Appendix 2. However, this subsection focuses on the analysis of the comparison of the different years, with the objective of helping to build an overview of the trend and situation in each of them.

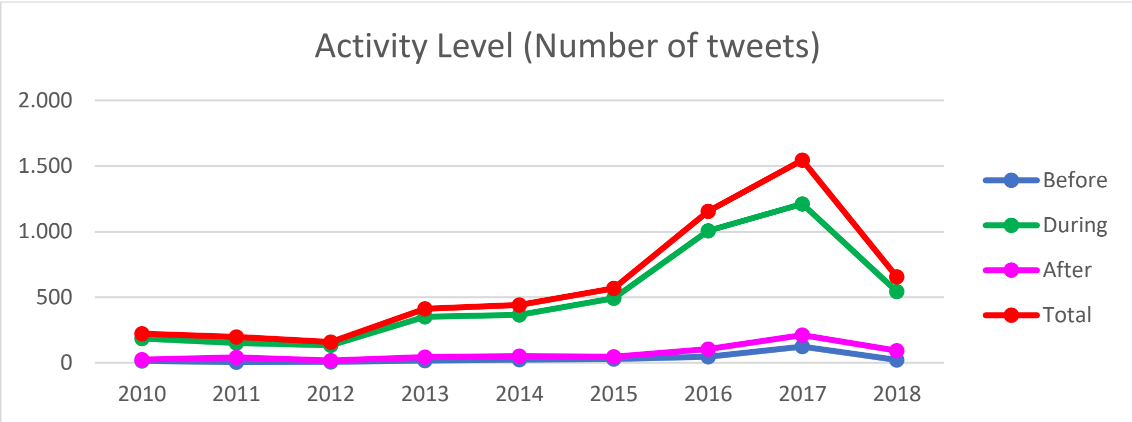

Figure 10. Activity Level (number of tweets) along the years

Figure 10 shows four series that indicate the activity in number of tweets before, during and after the conference in the corresponding edition, as well as the total of such three periods of time. As can be observed, again with the exception of what happened in the last year, a positive trend can be observed in activity level, a fact that agrees with the increase in the number of users (nodes) previously presented (Figure 3).

In addition, to complete and visualize this part of the analysis, the activity level is shown below (Figure 11) taking into account the percentage of such activity that occurs before, during and after the conference in each of the years.

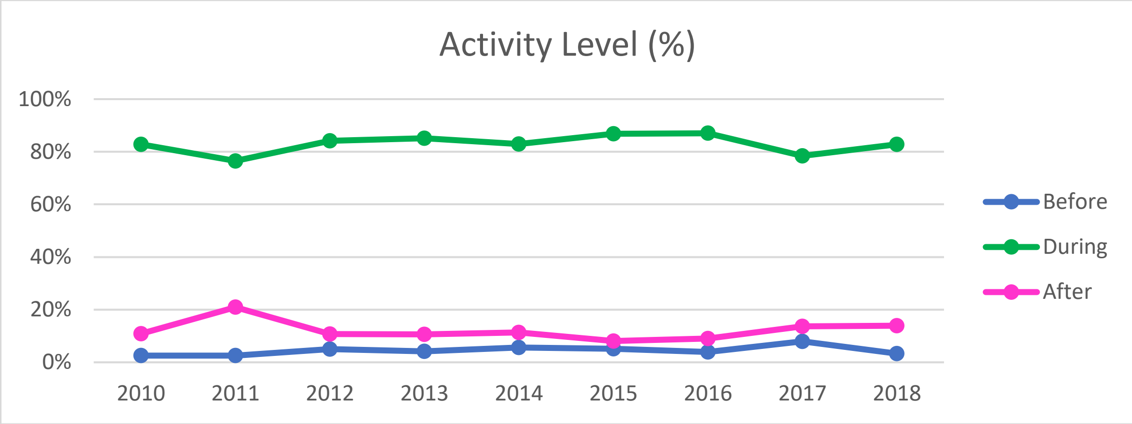

Figure 11. Activity Level (%) along the years

As can be observed, a relatively constant trend is maintained, which illustrates the great difference in the activity that exists between the period comprising the days of the conference itself and the rest of the days. It can also be noted the higher activity after the conference days than in the previous days of it.

6.2 Networks analysis

As already indicated in the introduction of this section, the networks have been built on the basis of mentions among users in the context of the HICSS conference through the social network Twitter. As previously mentioned, this task seeks to explore the potential of implicit networks through the analysis of mentions networks obtained in nine editions of the same conference. However, in this section the analysis focuses on three years (2014, 2015 and 2016), showing the graphs obtained for the remaining years in Appendix 3.

The choice of these three years within the available spectrum is motivated by the best proportion of number of nodes when analysing and drawing conclusions. That is, the first years show a lower number of users (nodes) and, therefore, the extraction of information may be limited; whereas in the case of recent years, its greater number of nodes adds complexity when it comes to visualizing the conclusions obtained.

In the graphs obtained, a unique identifier number has been assigned to each node. This has been done for two main reasons: firstly, the need to maintain the anonymity of Twitter users, and secondly, the greater readability of node labels in these graphs.

At the same time, also to facilitate the visualization of the graphs and focus the attention and the interest of the work, in the graphs of each year, labels have been indicated only for the nodes that constitute users that have already appeared in previous years; except for the first year, 2010, in which the identification numbers of all the nodes are indicated (Appendix 3).

The graphs obtained show different groups or communities, being these groups formed following the modularity principle based on the mentions made among users. That is, the modularity algorithm groups the nodes that have more common connections (more common mentions), thus classifying users according to different communities. In this way, in the graphs, nodes (users) are coloured by community and edges (connections) are coloured in the colour of the user who makes the mention. In turn, the size of the nodes is proportional to the number of connections that node has, so those with the greatest impact on the network highlight.

The layout presented by the graphs follows a forces algorithm, that is, the nodes with more common connections are attracted and those with fewer common connections repel each other. Therefore, this algorithm makes the distance between the nodes inversely proportional to the common connections, that is, at a greater distance, less interaction between the nodes.

Therefore, as already mentioned above, in order to focus the analysis and correctly understand how to interpret the results, it will be commented in greater depth on what has been observed in 2014, 2015 and 2016. This analysis is shown below.

6.2.1 Graphs

The graphs obtained for the years 2014, 2015 and 2016 are shown below (Illustration 9, 10 and 11).

Year 2014:

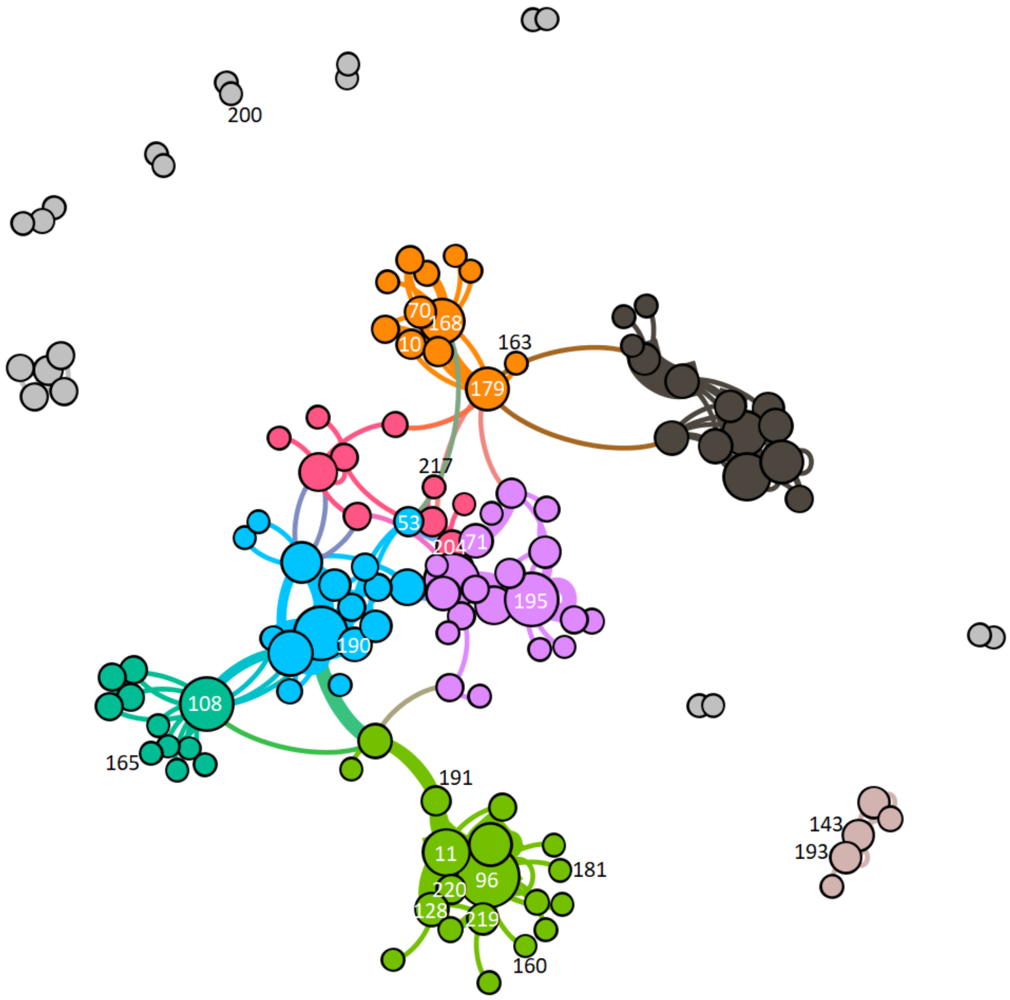

Illustration 9. Network 2014

Year 2015:

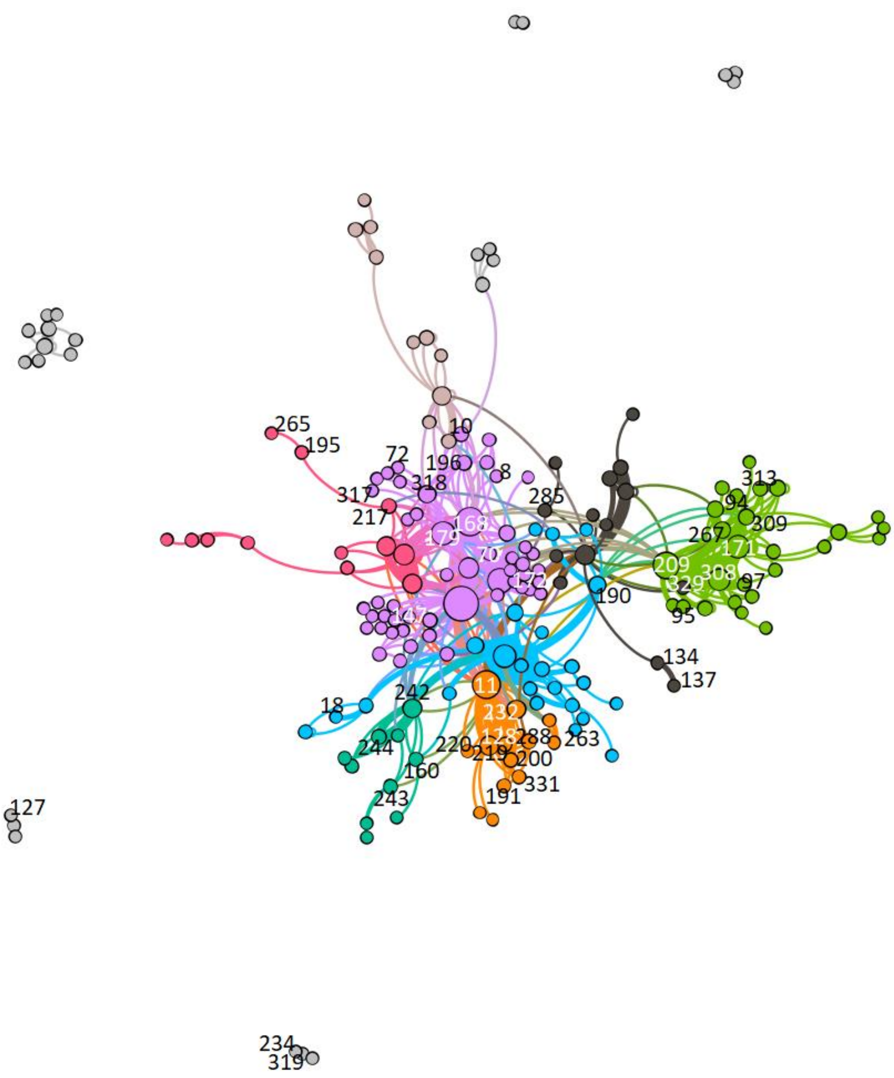

Illustration 10. Network 2015

Year 2016:

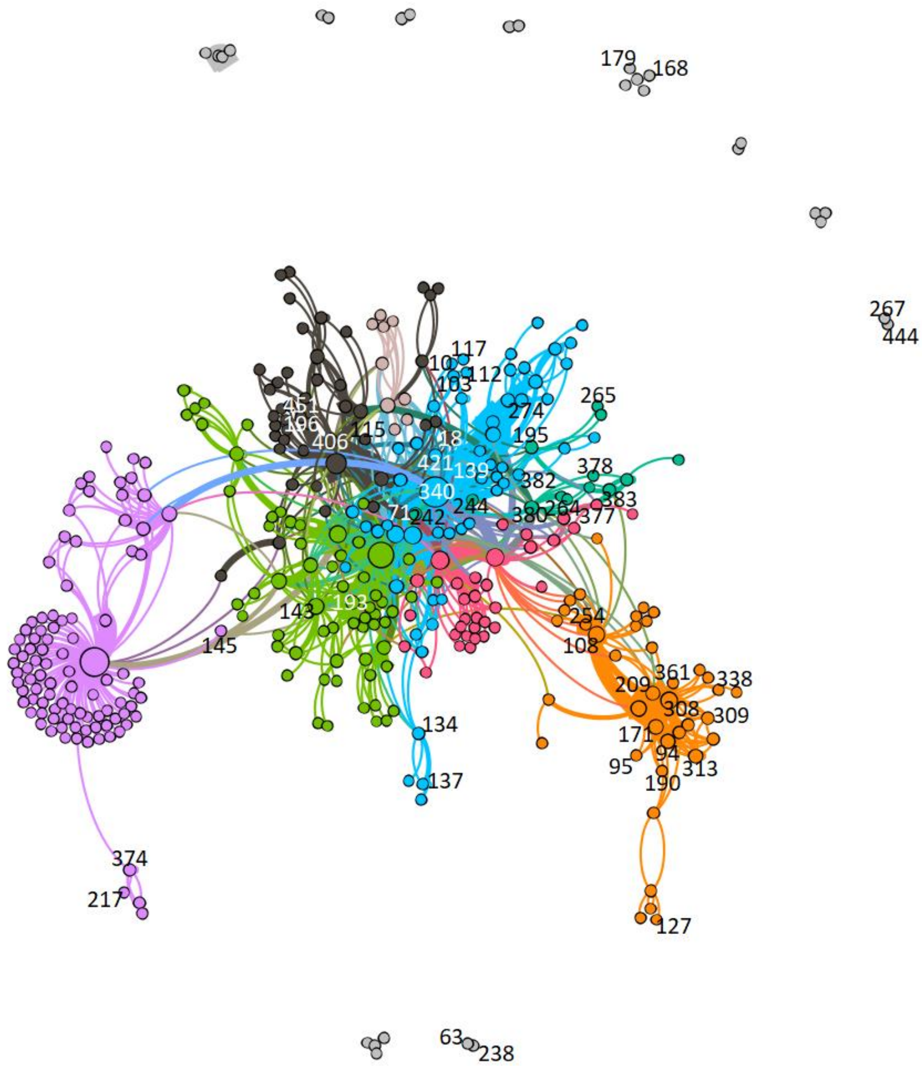

Illustration 11. Network 2016

Once the graphs have been obtained, a comparative analysis of them has been conducted. For this, a comparison of the years in pairs has been made, that is, 2014-2015 and 2015-2016. In each of these X-Y comparisons, in the graph of year Y, the labels of nodes (users) that have already appeared in any of the previous years are indicated. However, in the graph of year X, only the labels of nodes that appear also in the following year (Y) are shown. In this way it is tried to facilitate the interpretation of the results while trying to give an overview of the results obtained.

To make this comparison between year X and year Y, a process of identification of nodes has been conducted, so that certain clusters or groups of nodes that show a tendency to remain close in consecutive years have been identified. These clusters are also represented in the illustrations shown below (Illustration 12 and 13).

Years 2014-2015:

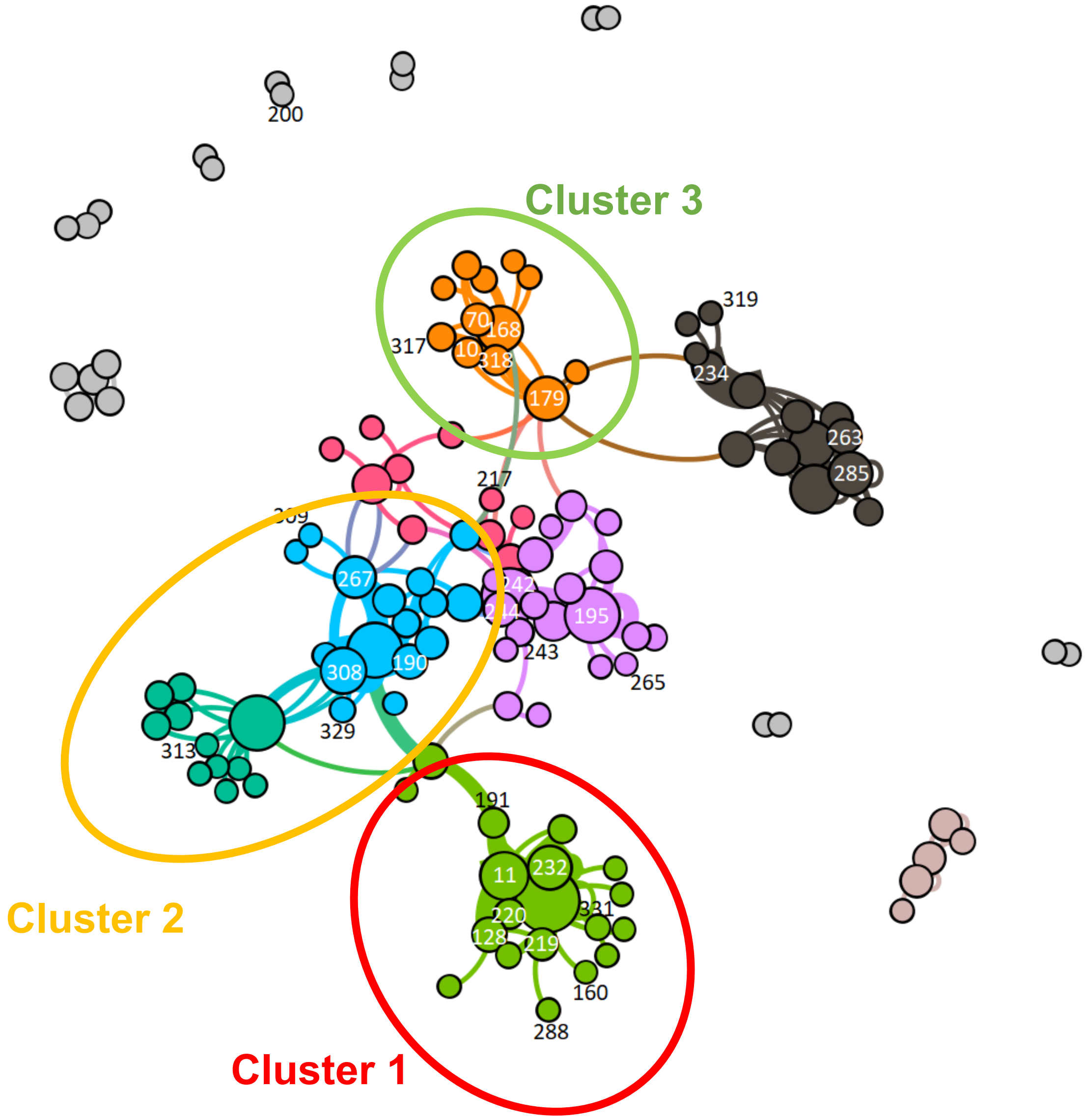

Illustration 12. Network 2014 with clusters (compared with network 2015)

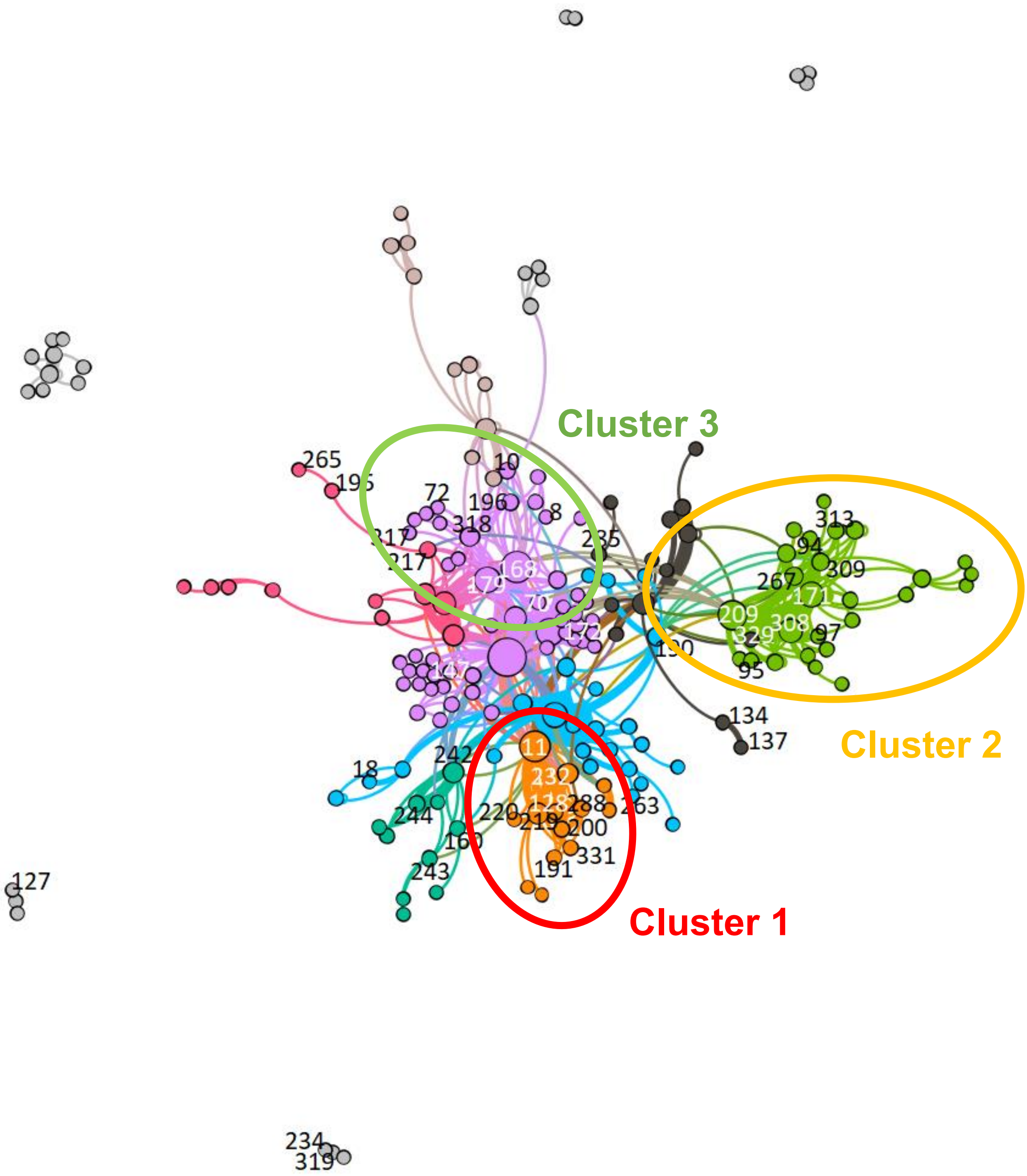

Illustration 13. Network 2015 with clusters

In the first place, it is observed that, as previously indicated in the descriptive analysis, the number of nodes (users) and interactions is lower in 2014 than in 2015. However, in spite of the fact that the number of nodes in 2015 is higher, it can be seen that there is a greater number of nodes with similarities between them, since the **number of communities** is smaller (therefore, being in general the number of nodes per community larger). Specifically, under the same resolution parameter in the Gephi tool, the number of communities (different colours) generated in 2014 is 16, while in 2015 it is 14.

However, in view of the graphs, it is observed that this result is due to small groupings that surround the central cloud. If these small communities that orbit around the large mass of nodes (Giant Component) are not taken into account, the number of communities in 2014 is 7, while in 2015 it is 9, being this result more expected as there are more nodes in 2015.

After analysing the number of communities in each of the years, the focus is on the analysis of the clusters identified with the nodes or users that participate in both editions of the conference of study. In the case of the network obtained for the year 2014, a more defined **disposition of the clusters** is observed and with greater differentiation among them than in the case of the year 2015. However, in both cases, this division of clusters obeys to a large extent the division into different communities with the modularity criterion of the Gephi tool.

Regarding the **structure** presented by each of these clusters, in 2014 a great tendency to the star arrangement can be observed, that is, there is a central node with a larger size (with a greater degree of mentions) to which the rest of the nodes within the same cluster are connected. However, this trend is reduced in the case of 2015, with a smaller difference in size between nodes (that is, degrees of mentions between nodes are balanced to some extent), as well as a greater interconnection between different nodes, establishing a configuration more of community than of star. However, in 2015, within the communities formed in each cluster, it is possible to differentiate some nodes that act as small central nodes, showing “small star configurations”, which are interconnected to give rise to the final community configuration.

This type of arrangement based on a star configuration leads to conclude the importance that certain nodes have within the network. This point is of great interest and usefulness, since these nodes have an influential profile and impact on other users. It is then clear the potential that the information of this result can suppose in recommendation systems, since the influence and intervention on these nodes of greater weight presents a greater probability of success and propagation, thanks to the influence of these nodes in others.

Years 2015-2016:

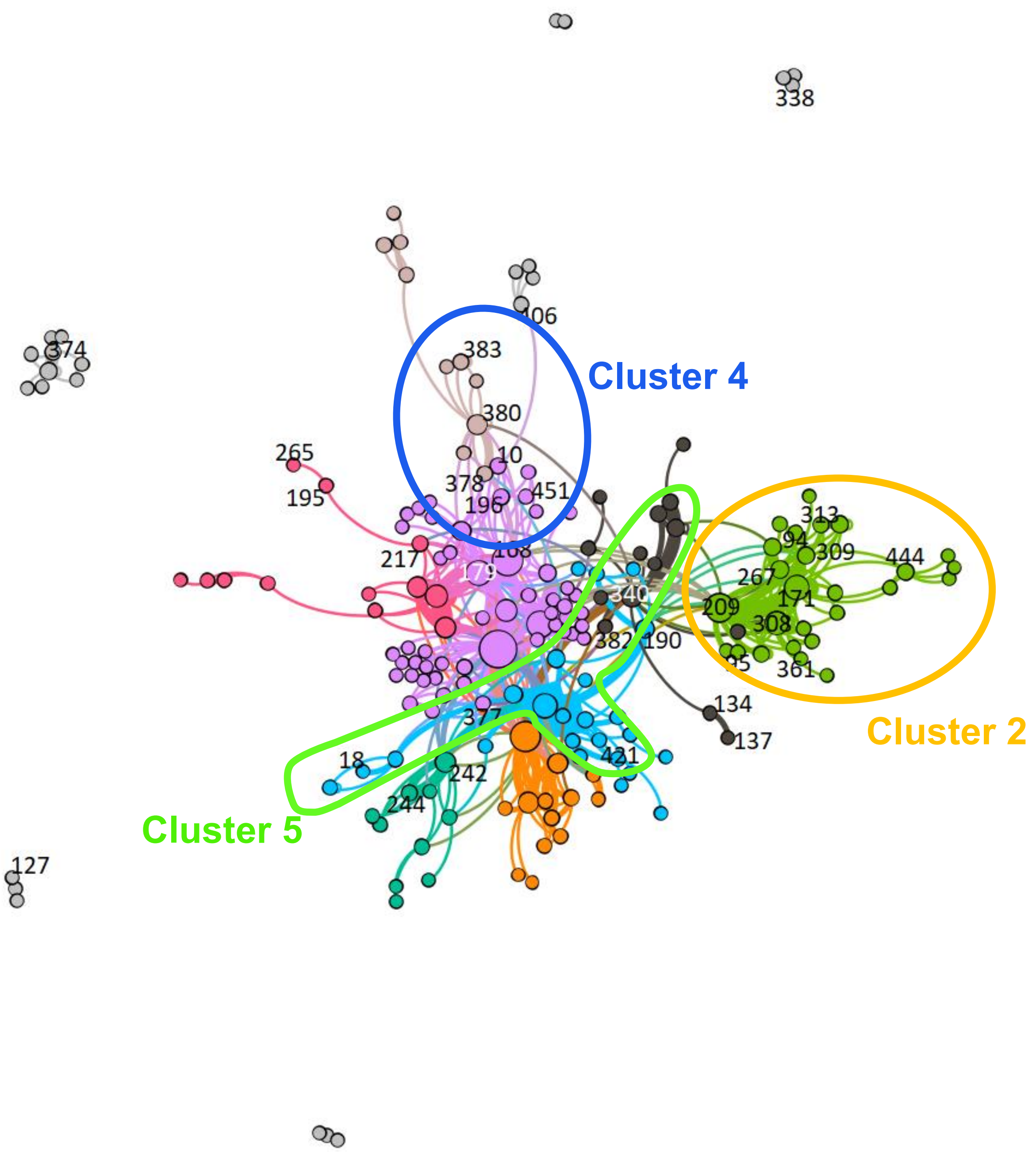

Illustration 14. Network 2015 with clusters (compared with network 2016)

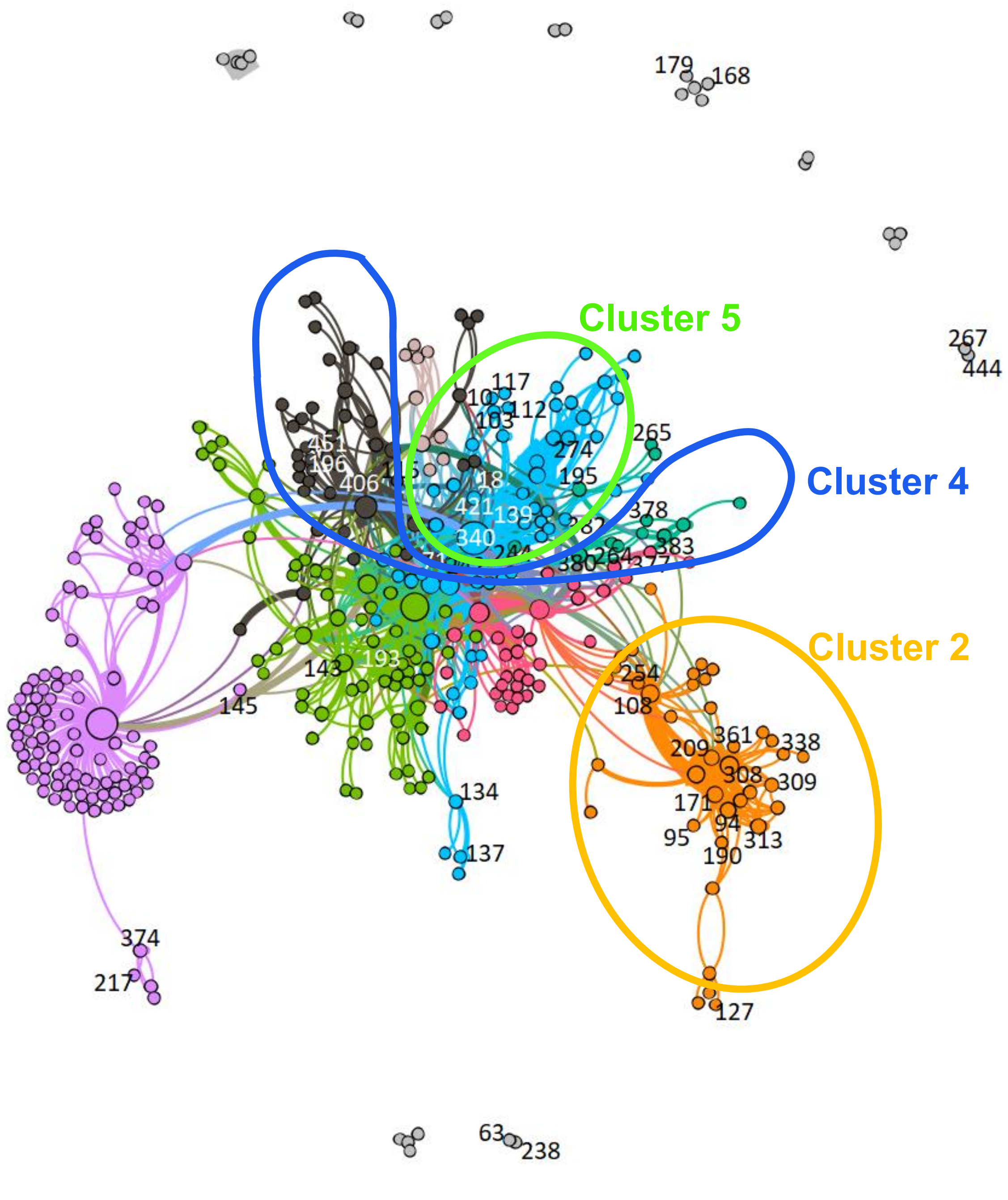

Illustration 15. Network 2016 with clusters

In this case, in terms of the **number of communities** generated by the Gephi tool, 14 communities are obtained in 2015, in contrast to the 18 obtained in 2016. However, in the same way as in the previous case, if the communities that orbit around the central cloud of the network are not taken into account, the number of communities becomes 9 in 2015 and 8 in 2016. The conclusions that derive from this result are similar to those of the previous case, that is, despite if there is a greater number of nodes in 2016 than in 2015, the proximity between them results in the final number of communities that make up the central cloud being similar in both cases.

Once again, once the number of communities in each of the years has been analysed, the focus is on the analysis of the clusters identified with the nodes or users that participate in both editions of the conference of study. In this case, the **disposition of the clusters** is similar in both cases. Specifically, Cluster 2 is more defined and distanced from the rest of nodes, but Cluster 4 and Cluster 5 are more interconnected with the rest of nodes of the central cloud. From this, it can be deduced that Cluster 2 presents greater independence that can be translated into aspects such as discussing topics that are more different than the rest of the cloud, or the need to influence some user of that cluster if it is wanted to reach that group of users. This last aspect is due to the fact that in case of Cluster 2, connections with the rest of the central mass are fewer than in the other two clusters, and, therefore, the bridges through which accessing such group are lower.

As for the **structure** of the clusters, as has already been mentioned in the previous case, in 2015 there is a community arrangement made up of different small star arrangements. In 2016, a similar structure can be distinguished, but in this case the number of connections increases, resulting in a more interconnected structure between nodes of the same cluster and between nodes of different clusters. In the same way as in the previous comparison, the importance of the potential of the information obtained with these graphs can be highlighted, being able to identify those most influential nodes, as well as the accessibility to different clusters through the bridges established by the existing connections.

6.2.2 Wordclouds

The next step that is taken in the analysis is the construction of wordclouds with the most mentioned words within the text content of the tweets of the users belonging to each of the identified clusters.

The objective that is sought with this task is to see if the different clusters address different topics. This is conducted to check if there is a profile of common interests to which the different users belonging to each group approach, or if, on the contrary, the proximity between nodes within a group is due solely to possible personal affinities. In addition, it is sought to observe if the tweets published by the users are mere tools to accomplish the organization of the event, or if, on the contrary, the topics related to the conference are discussed. For example, some of the most important topics discussed in the HICSS Conference are: *Digital Transformation*, *Blockchain in Business*, *Artificial Intelligence and its applications*, *Data Analytics and Business Intelligence*, *IT and*

Healthcare, *IT and Society*, *Innovation and Sustainability, Information Security and Privacy* or *Research Methodologies in Practice* (*ScholarSpace at University of Hawaii at Manoa: Hawaii International Conference on System Sciences 2019*); so the idea is to see if it exists a relation between the topics covered in the tweets and those discussed in the conference.

Wordclouds are constructed thanks to the use of the PYTHON programming language (Appendix 1), and the results obtained are shown in the following illustrations. Continuing with the comparative structure by pairs of years, the 2014-2015 clusters are shown first and then the 2015-2016 clusters.

Years 2014-2015:

- **Cluster 1:**

Illustration 16. Wordcloud Cluster 1 (2014)

Illustration 17. Wordcloud Cluster 1 (2015)

In the wordcloud of Cluster 1 in the case of 2014, words such as “learning”, “workshop”, “join”, “students”, “challenge”, “participate” or “training” highlight. While in the case of 2015, words such as “virtual teams”, “paper”, “inventions”, “participating” or “workshop” appear.

In this case, there is a relation between the topics addressed in the same cluster in consecutive years, which leads to the conclusion that there is a similarity in the interests of those users, and therefore, in their profile. In particular, the words found evoke a learning and research environment, as well as its implementation through teams and workshops. With this, it can be understood that they do address topics related to the conference, and not only to its organization, thus it is possible to associate this cluster with a profile interested in the subtopic of *Research Methodologies in Practice* within the conference.

On the other hand, it can be observed that wordclouds are also a useful tool when obtaining feedback from users, since in them it can be found some words that determine emotions that can be positive or negative, in order to help determine if the conference has been successful or not. In particular, in this case words such as “great” or “mind blowing” can be found, which indicate a positive perception on the part of the users.

- **Cluster 2:**

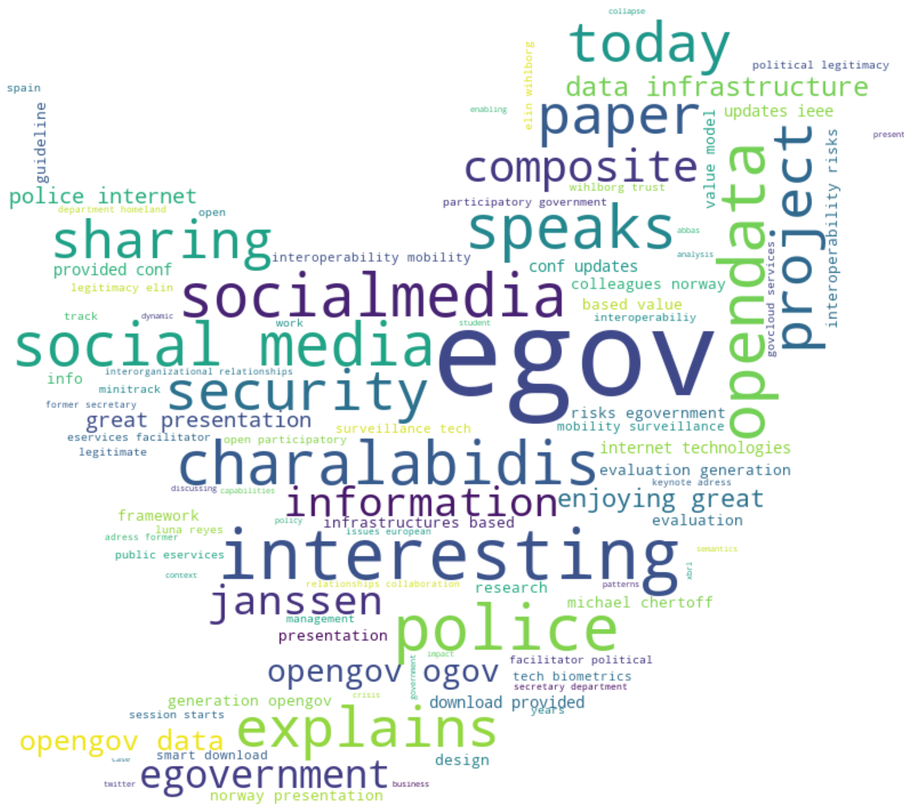

Illustration 18. Wordcloud Cluster 2 (2014)

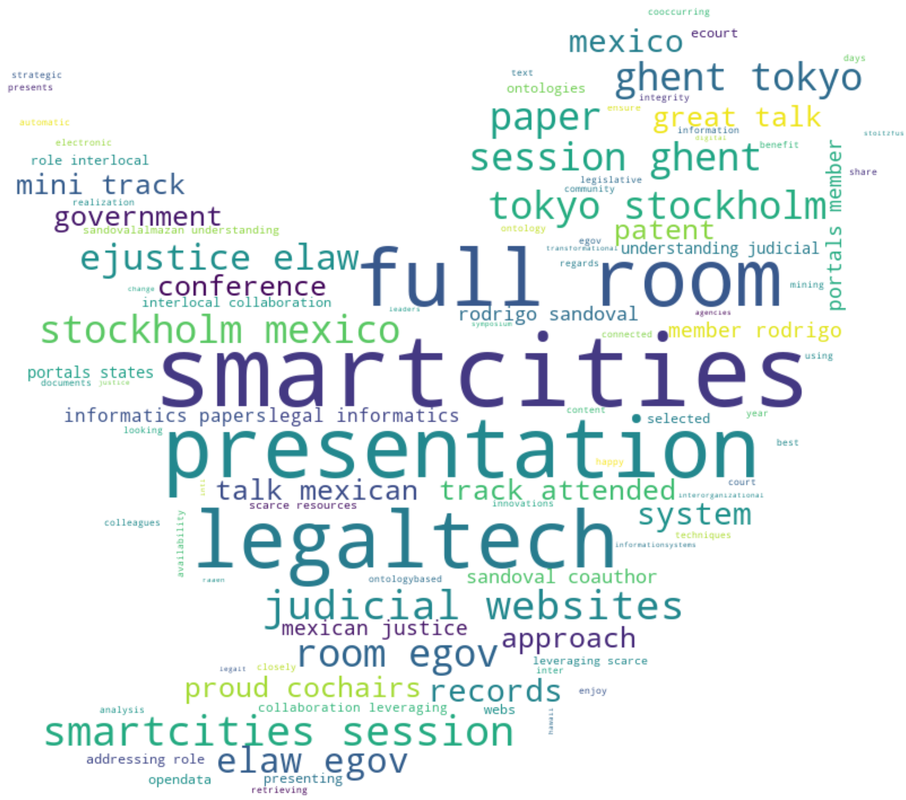

Illustration 19. Wordcloud Cluster 2 (2015)

In the case of Cluster 2, important differences are observed regarding the tweets content of the previous cluster. Specifically, in 2014 it highlights words such as “egoverment”, “opengov”, “egov”, “security”, “opendata”, “police” or “social media”. While in 2015 it highlights words such as “ejustice”, “elaw”, “legaltech”, “judicial”, “egov”, “patent”, “smartcities” or “government”, among others.

In this case, in both years, the words of this cluster evoke an environment of justice, legality, privacy and ethics. Therefore, it could be associated to a profile more interested in the subtopic of *Information Security and Privacy* within the conference topic.

On the other hand, it can be also found words like “great talk”, “enjoying” or “interesting”, that tip the scale of the success of the conference in positive values.

- **Cluster 3:**

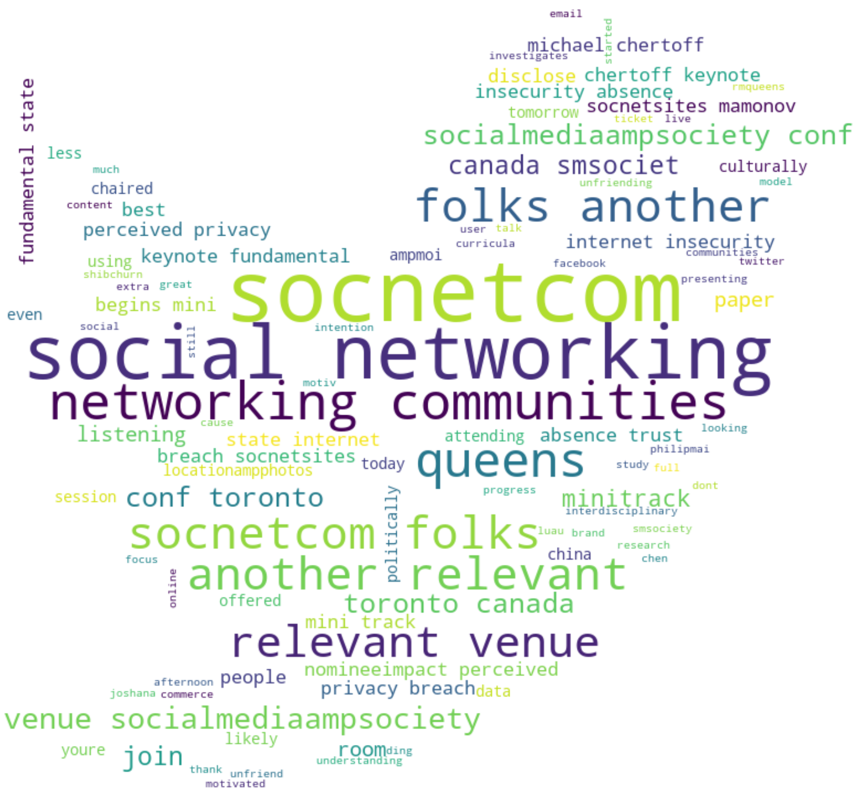

Illustration 20. Wordcloud Cluster 3 (2014)

Illustration 21. Wordcloud Cluster 3 (2015)

Again, in Cluster 3, there are differences with respect to the previous two clusters. In particular, in 2014 it highlights words such as "social networking”, “socnetcom”, “networking communities” or “join”. While in 2015 it highlights words such as “social media”, “media archiving”, “social journalism”, “news national”, “citizen journalism” or “survey report”.

In this case, a greater difference can be observed between the words obtained in 2014 and those obtained in 2015. Those of 2014 evoke a networking environment and the creation of communities through social media; while those of 2015 evoke an environment more related to the world of journalism, news and social surveys. However, in spite of these differences, in both years the topics addressed can be encompassed under a common denominator, being possible to correlate them with the *IT and Society* topic among the most important topics of the HICSS conference.

The fact that there is a more pronounced difference between years 2014 and 2015 in Cluster 3 than in Cluster 1 and 2 makes sense if the graph of 2015 is observed. That is, when analysing the disposition of clusters in the network, it is observed that Clusters 1 and 2 adopt a more independent disposition from the rest of communities, while Cluster 3 is more embedded in the core of the network, being closer and with greater communications with other communities, so it can exist certain influence or sharing of topics with neighbouring communities.

Years 2015-2016:

- **Cluster 2:**

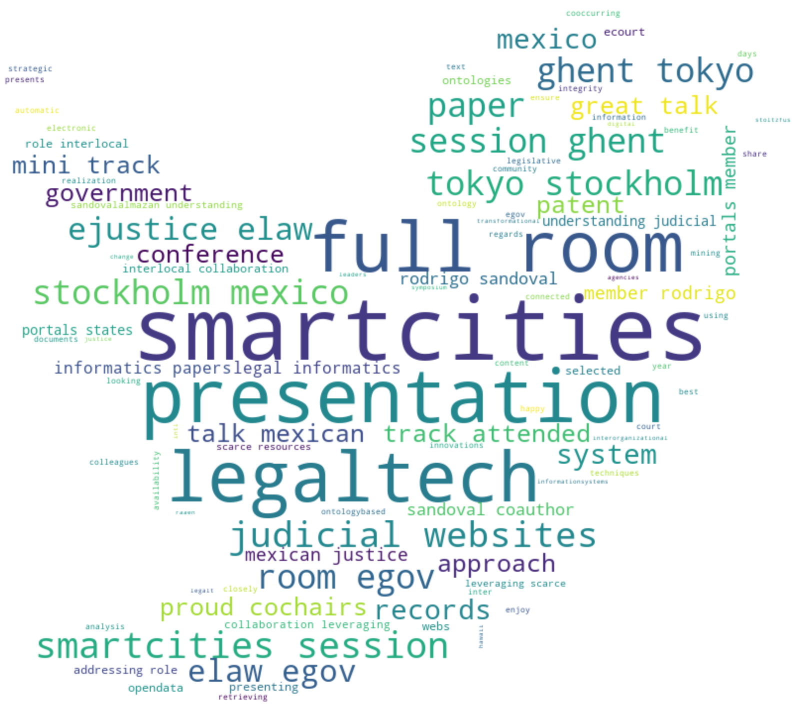

Illustration 22. Wordcloud Cluster 2 (2015)

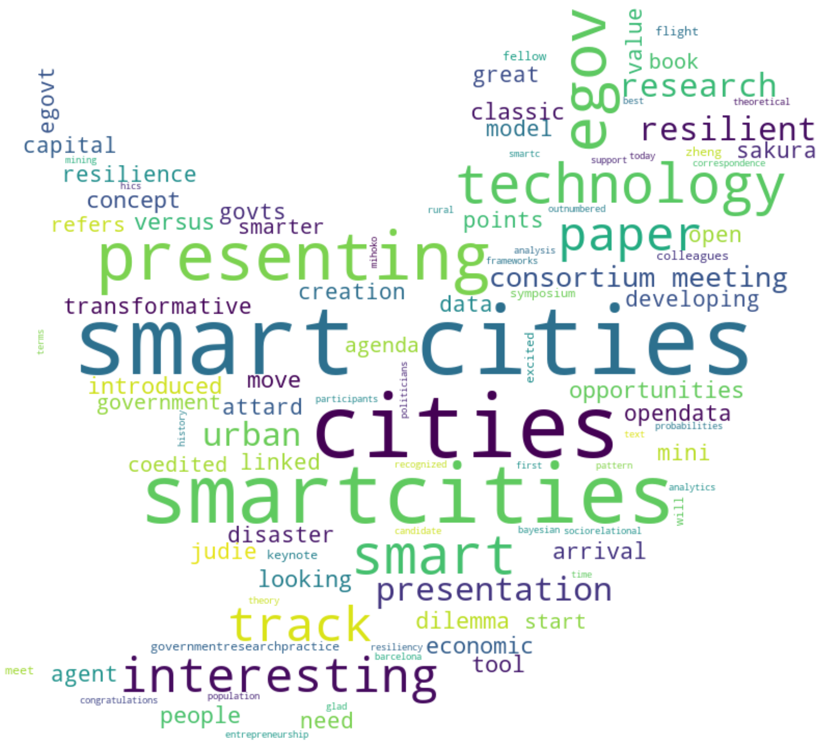

Illustration 23. Wordcloud Cluster 2 (2016)

In this case, as already mentioned in the previous comparison, Cluster 2 in 2015 highlights words such as “ejustice”, “elaw”, “legaltech”, “judicial”, “egov”, “patent”, “smartcities” or “government”. While in 2016 it highlights words such as “egov”, “smart cities”, “egov” or “technology”.

It can be observed that Cluster 2 continues with the same theme in the three years analysed. As can be seen in the graphs, in the three years Cluster 2 presents a quite independent disposition with respect to the rest of the communities of the graph, reason why it can be concluded that it is a delimited cluster whose interests and discussion topics are framed in a defined context.

- **Cluster 4:**

Illustration 24. Wordcloud Cluster 4 (2015)

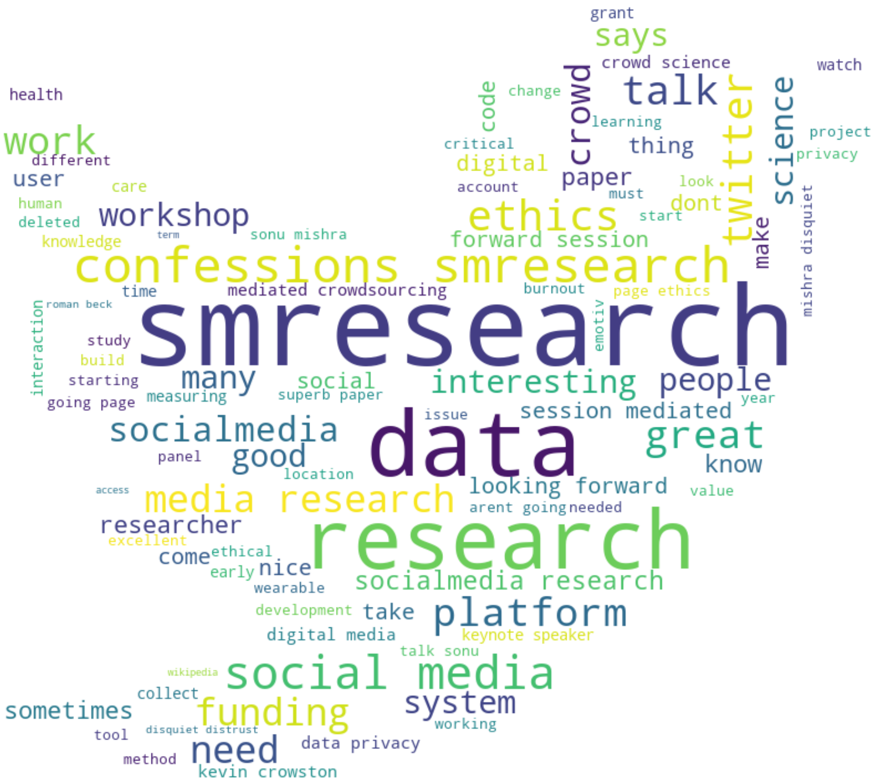

Illustration 25. Wordcloud Cluster 4 (2016)

In this case, a new cluster (Cluster 4) has been identified in 2015. This cluster is closer to the cluster identified in the subsequent year (2016) in terms of the nodes it contains. However, it is worth noting the similarity of this cluster with Cluster 3 in 2015, so it is expected a similar theme in the discussions of users.

Specifically, Cluster 4 in 2015 highlights words such as “social media”, “network”, “participation”, “student”, “work” or “analysis”. While in 2016 highlights words such as “social media”, “research”, “data”, “science”, “workshop”, “smresearch” or “work”.

Therefore, this cluster is part of an environment of social media research and data analysis common in both years, being at an intermediate point or combination between the subtopics of *IT and Society*, *Research Methodologies in Practice* and *Data Analytics*. The fact of approaching different topics can again be explained by the disposition of this cluster in the graph, especially in the year 2016. That is, in such graph it is observed that, unlike Cluster 2, Cluster 4 crosses the core of the network, so it can be influenced or share different topics with other neighbouring communities.

- **Cluster 5:**

Illustration 26. Wordcloud Cluster 5 (2015)

Illustration 27. Wordcloud Cluster 5 (2016)

Finally, in the case of Cluster 5, in 2015, words such as “paper”, “research”, “development”, “social media”, “data”, “change” or “future” are highlighted. While in 2016, the most relevant words are those such as “data analytics”, “science”, “paper”, “research”, “data” or “ecosystemanalytics”.

Therefore, a common topic is also defined in both years for the case of Cluster 5. In particular, the discussions can be framed in a combination of the main subtopics of *Digital Transformation*, since it talks about future and change, and *Data Analytics*.

In conclusion, it can be seen that, despite all the clusters are under a common denomination marked by the general theme of the conference, different subtopics can be distinguished within the network. It has been observed that this thematic differentiation is more delimited in the case of clusters that present a disposition with greater independence from the rest of the communities within the network. Therefore, this analysis of the wordclouds of each cluster has served to strengthen and reinforce the understanding of the disposition of each of the communities in the network graph.

In turn, it can be concluded the usefulness and importance of identifying the profile of each cluster at the time of influencing and impacting the different users, who constitute potential conference attendees in future editions. In other words, the use of wordclouds allows to segment the users to build a better recommendation system and plan a more consistent, fruitful and successful organization of the conference.

6.3 Individual analysis

Finally, to go a step further and deepen in the analysis, here it is intended to go one level below in the analysis by considering a single network in order to observe certain centrality features and obtain conclusions in a more qualitative and focused manner. In this case, the network obtained for 2015 is taken as the object of study, since it has an adequate proportion of nodes and connections to facilitate its interpretation and understanding.

Therefore, firstly, some node characteristic measures will be evaluated, visualizing the results through illustrations that facilitate their interpretation. In particular, these measures will be analysed to study the centrality in the network, so that the most important nodes can be identified in it.

Subsequently, it will be tried to conduct a more qualitative analysis, lowering the analysis to the level of evaluation of certain tweets made by the most important users (nodes) within the network. The objective is to try to identify the role of these nodes within the network and try to give a real interpretation to the reasons that lead these nodes to highlight from the rest.

6.3.1 Centrality Measures

In-degree and out-degree

In the presented networks of each one of the years of study (section 6.2.1 and Appendix 3), the degree of nodes was already represented through the size of them. However, in this section this measure will be disaggregated in two: the in-degree and the out-degree measures. This differentiation exists because the mentions network built is a directed network, that is, connections present directions that go from an origin node to a destination node.

The in-degree measure refers to the number of incoming edges in a node (that is, how many nodes point to that node), while the out-degree measure refers to the number of outcoming edges of a node (that is, to how many nodes point that node). It is therefore sought to observe if there are important nodes that only highlight in one of the two measures, so that an interpretation can be given in such case. That is, it could be the case that a node has a high degree, but such value is produced only by outcoming connections, which would indicate that such user mentions many others, but, however, is not mentioned by others.

The results obtained are shown in the following illustrations (Illustration 28 and 29). In them, the measures analysed are identified through the use of different colour intensities. Those nodes with higher values are represented with a higher intensity, while those with lower values have a lower colour intensity.

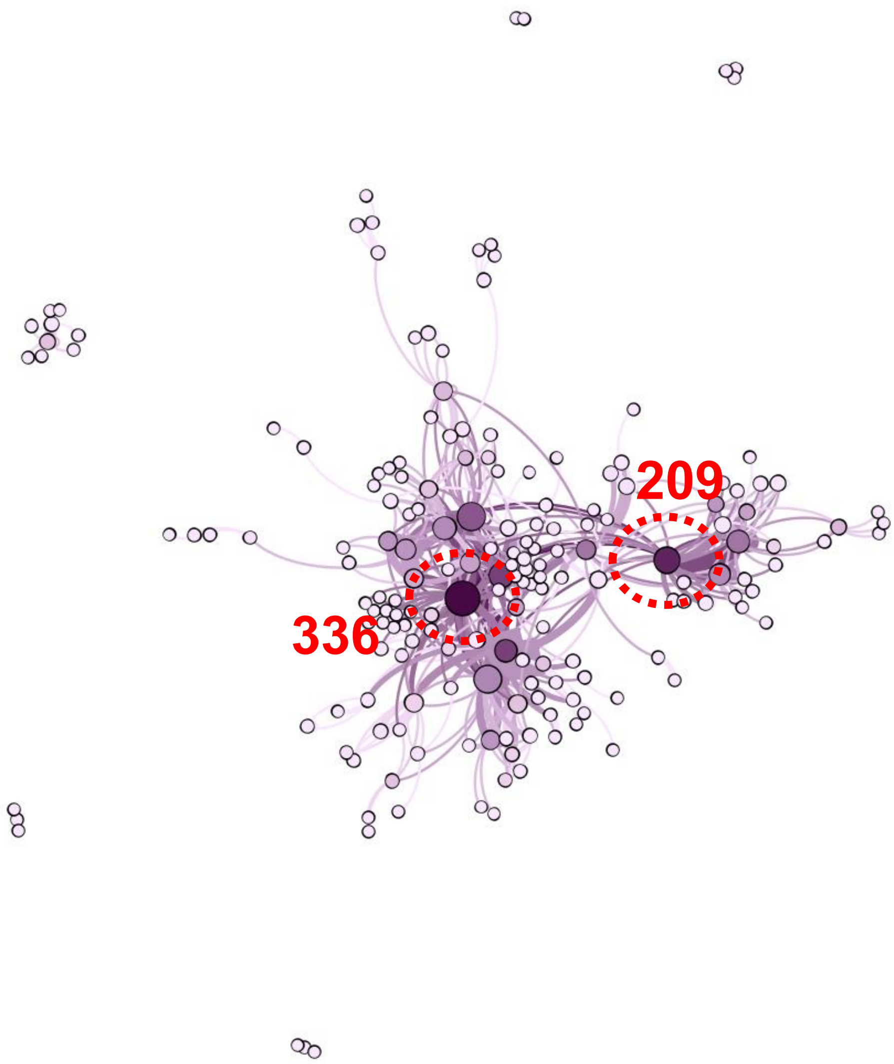

Illustration 28. In-Degree Measure

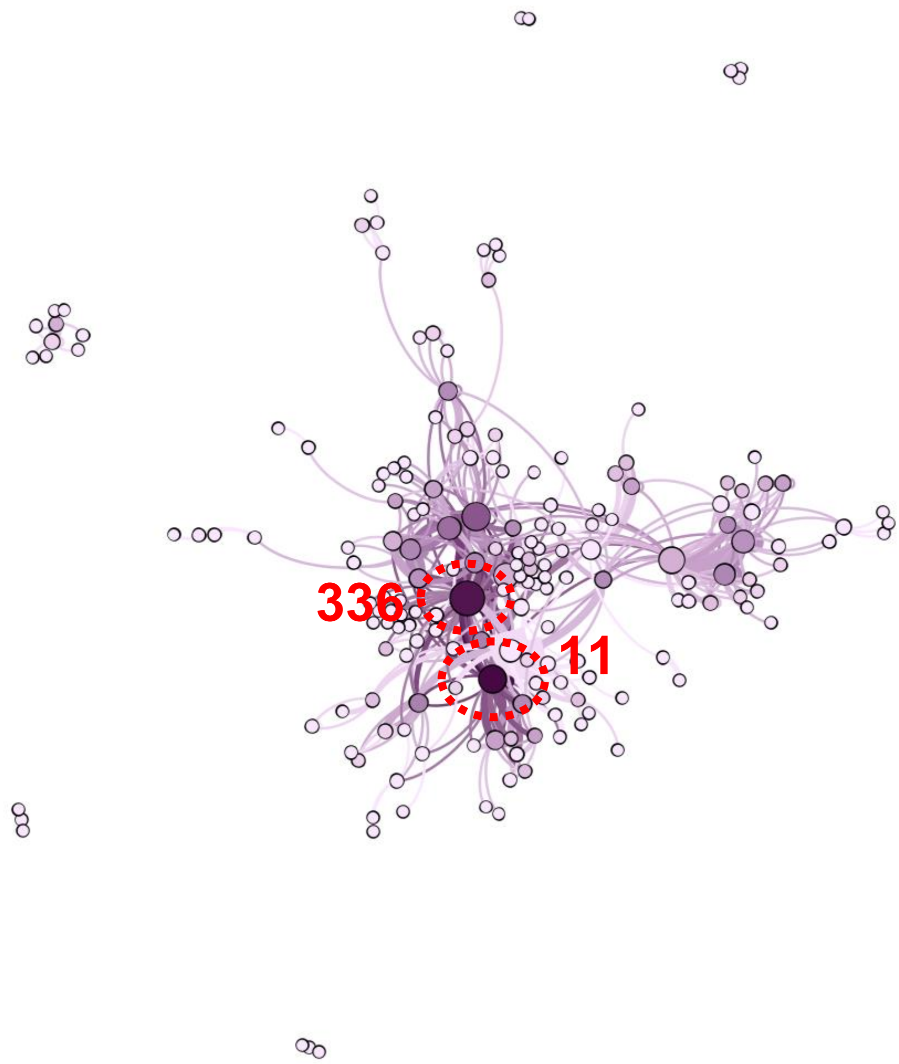

Illustration 29. Out-degree Measure

Certain differences can be distinguished between both representations of the network, since the nodes do not show in all the cases the same intensity of colour, which leads to conclude that some nodes are more mentioned and others make more mentions. Specifically, in Illustration 28 and 29, they have been indicated with a red dashed line those two nodes that highlight in each of the cases. As can be seen, one of the nodes is the same, while another is different. The identification of these nodes contributes to the selection of the nodes with greater centrality, which will be selected to conduct the subsequent qualitative analysis. Therefore, the identification of the most important nodes resulting from this analysis is discussed here, being the nodes with higher values of in-degree the nodes 336 and 209, and the nodes with higher values of out-degree the nodes 336 and 11.

Closeness Centrality

The next centrality measure to be analysed is the Closeness Centrality. This measure indicates the average distance from a given starting node to all other nodes in the network. However, it should be again taken into account the fact that the network of study is directed and not all nodes are able to access all the rest of nodes.

Closeness Centrality can be described with the following formula in its normalized form (Figure 12), where $C_C(i)$ is the closeness centrality measure of node i, $d(i,j)$ represents the distance measure (the shortest path) between the nodes i and j, and N is the number of nodes in the network graph (*Closeness Centrality (Centrality Measure) - GeeksforGeeks*).

$$C_C(i) = \frac{N-1}{\sum_{j=1}^{N} d(i,j)};\ \forall i,j \in \mathbb{N}$$

Figure 12. Closeness Centrality normalized formula

Anew, the representation of the graph obtained is shown indicating with different degrees of intensity of colour the different values of the measure. The result is shown in the following illustration (Illustration 30).

Illustration 30. Closeness Centrality

As can be seen, in this case there is not a small number of nodes that highlight from the rest of the network, but there are several nodes that show high values and these are distributed throughout the network. After observing the results, it can be concluded that this measure does not picture a good representation when it comes to identifying the central and most important nodes in the network. The reason for this is that, for being a directed network, not all nodes can reach any other, so the results obtained are not representative of the entire network. In other words, to understand what is happening, an example that refers to the nodes indicated with a green dashed line box in the network of Illustration 30 is shown below.

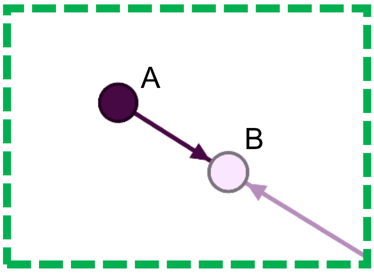

Illustration 31. Example of nodes in a directed network

In this illustration it can be seen the detail of what happens in that example. It can be observed how node A can only access node B within the network, since node B does not present an edge directed towards another node. Therefore, node A appears with the maximum colour intensity, since it can reach any other node that is reachable by it (only one node, node B) in the shortest possible distance, that is, in a single step. However, node B, since it cannot reach any other node within the network, its colour intensity, and therefore its value, is the lowest.

Therefore, it is concluded in this part of the analysis that, for the context in which the networks of study are framed, the closeness centrality measure is not the most adequate to identify the central nodes, so it will not be taken into account when selecting the candidate users for the subsequent qualitative analysis.

Betweenness Centrality

Another of the most important measures of centrality is the Betweenness Centrality. This refers to the measure of how often a node appears on shortest paths between nodes in the network. In some way, this measure serves to quantify how important are the nodes in their role of connectors or bridges for the rest of nodes to connect with each other.

Betweenness Centrality can be described with the following formula (Figure 13), where $C_B(k)$ is the betweenness centrality measure of node k, n_{ij} represents the total number of shortest paths from node i to node j, and $n_{ij}(k)$ indicates the number of those shortest paths that pass through node k. It should be noted that Betweenness Centrality can be calculated even if the nodes are not connected (*Betweenness Centrality (Centrality Measure) - GeeksforGeeks*).

$$C_B(k) = \sum_{i \neq j \neq k} \frac{n_{ij}(k)}{n_{ij}}; \ \forall i \neq j \neq k \ \in \mathbb{N}$$

Figure 13. Betweenness Centrality formula

Betweenness Centrality can also be normalized to the interval [0,1] by dividing by the number of pairs of nodes not including k. That is, the following expression (Figure 14) represents the formula in its normalized form in the case of directed graphs, being N the total number of nodes (*Betweenness Centrality (Centrality Measure) - GeeksforGeeks*).

$$C_B(k) = \frac{\sum_{i \neq j \neq k} \frac{n_{ij}(k)}{n_{ij}}}{(N-1)(N-2)}; \ \forall i \neq j \neq k \ \in \mathbb{N}$$

Figure 14. Betweenness Centrality normalized formula

Again, the results obtained for the network of study in this section are shown according to the intensity graduation of colour. The representation obtained is shown in the following illustration (Illustration 32).

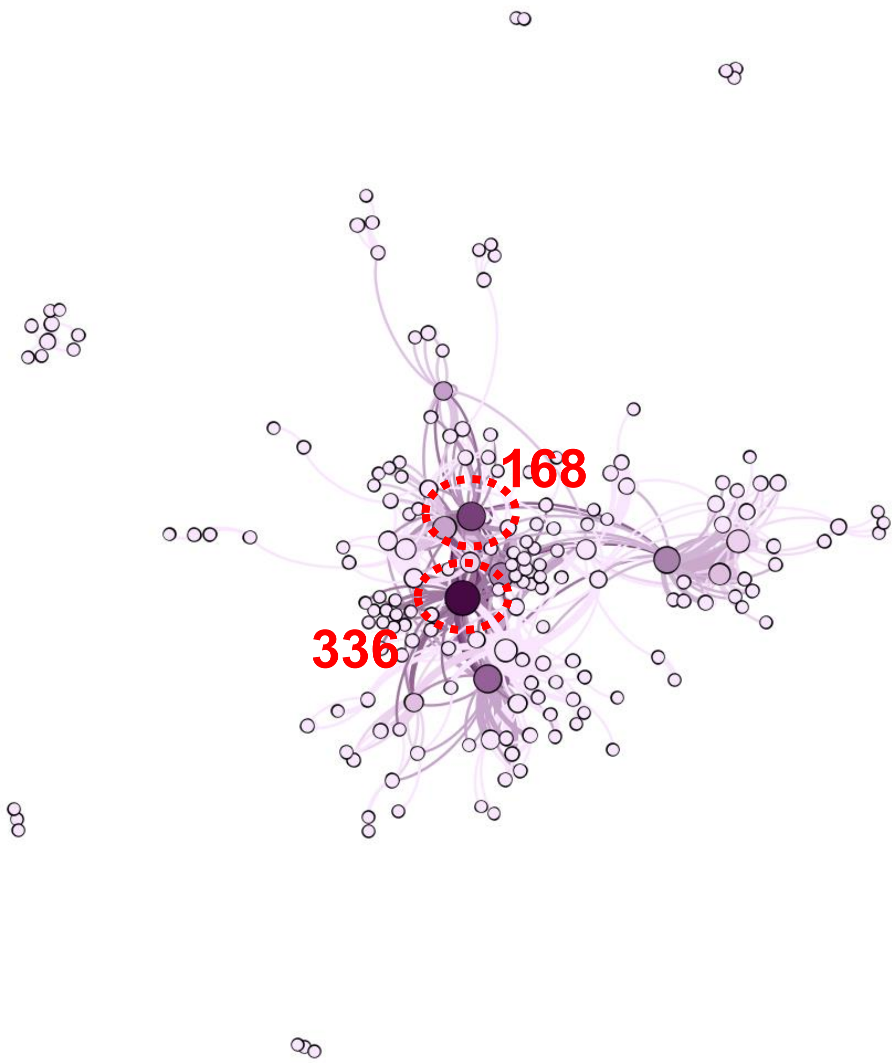

Illustration 32. Betweenness Centrality

In this case, a reduced number of nodes is highlighted from the rest. In particular, the two most noteworthy are indicated in Illustration 32. The identification numbers of those nodes are 168 and 336. It should be noted the coincidence of this last node with those identified in the representations of in-degree and out-degree measures (Illustration 28 and 29).

Therefore, this measure is similar to closeness centrality, but it is more useful in this case of a directed network, since betweenness centrality is also able to capture structural differences. That is, in some way, it is able to identify those nodes that are in a "key" position within the network.

PageRank

Finally, the PageRank measure is used. In a simplified way, this measure refers to the fact that the central nodes are those that, if a "random walk" is taken on this network, there is a high probability to pass through them. In some way, it measures the importance of the nodes within the network structure.

PageRank is an algorithm also used by Google Search to rank websites in their search engine. Specifically, this measure works by counting the number and quality of the links to a website to estimate how important the page is. The PageRank algorithm reflects a probability distribution that represents the probability that a person randomly clicking on links will reach a given page. The PageRank algorithm requires a process of several iterations to adjust the PageRank values (*Page Rank Algorithm and Implementation - GeeksforGeeks*). These websites represent the nodes and the links refer to the edges of the network.

Therefore, the importance of a node is determined by the sum of the PageRank scores of the nodes that point to it, these scores being indicators of the level of "prestige" or "authority" (*Page Rank Algorithm and Implementation - GeeksforGeeks*). In this way, the PageRank score can be expressed through the following formula (Figure 15), in which $P(i)$ indicates the PageRank score of node i, $P(j)$ represents the PageRank score of node j (which connects with node i), and $O(j)$ refers to the number of out-links of node j (whether or not it is directed towards i).

$$P(i) = \sum_{i \neq j} \frac{P(j)}{O(j)}; \ \forall i \neq j \ \in \mathbb{N}$$

Figure 15. PageRank formula

Once more, the colour intensity gradation is applied in the graph obtained to identify the values presented by each node in relation to this measure. The result obtained is shown in Illustration 33.

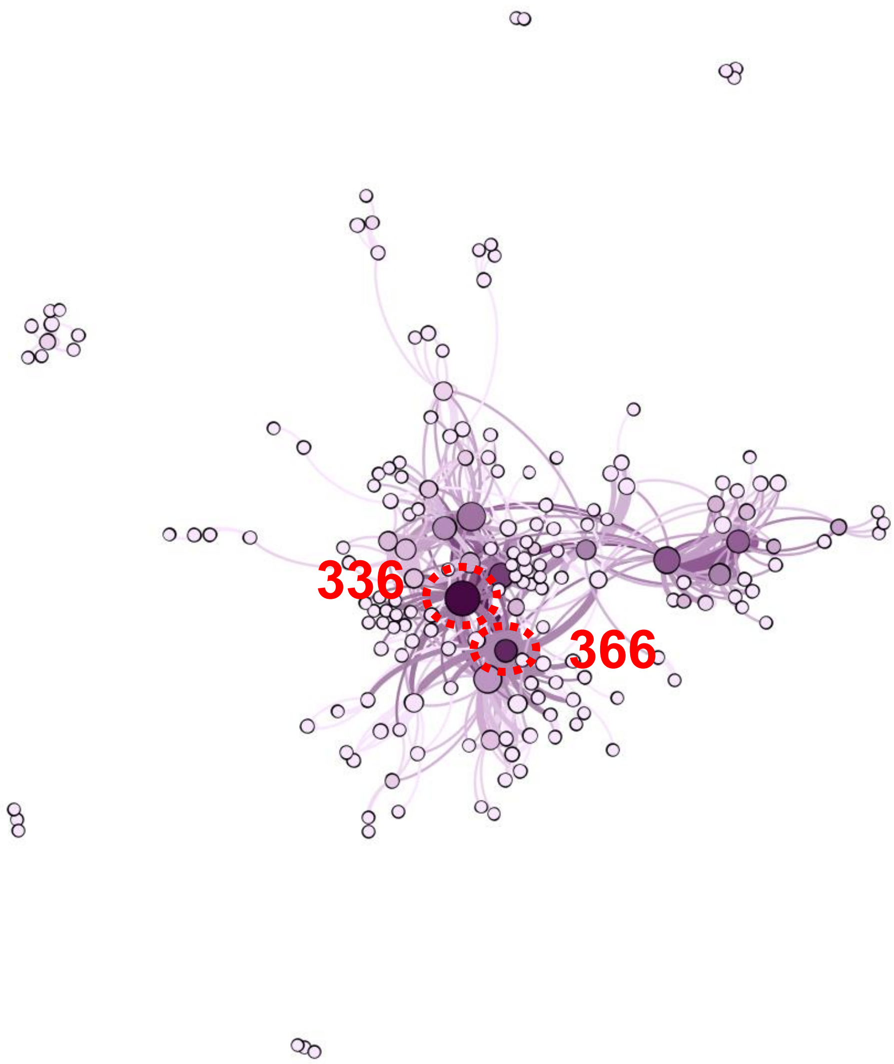

Illustration 33. PageRank

In view of the results, this measure indicates that the two most relevant nodes are those indicated by a dashed red line in illustration 25. These nodes correspond to the identifiers 336 and 366. It should be noted that node 336 appears again as the most important node in the network.

Once the centrality analysis of the network of study has been conducted and taking into account each of the relevant results achieved, it can be concluded that the main central nodes of the network are the nodes with identification numbers 336, 209, 11, 168 and 366.

6.3.2 Qualitative Analysis

Once the nodes with greater relevance within the network have been identified, as indicated in the introduction of this subsection, a qualitative analysis is conducted to obtain conclusions regarding the identification of the profiles of such key nodes.

As indicated, this task is conducted for the nodes with greater centrality and, therefore, with greater importance, identified within the network. For reasons of maintaining the anonymity of the users of the built networks, no identifying information such as names, descriptions or similar data is shown; but reference to certain terms or resources that denote the possible role of such user within the network is made.

From the qualitative analysis of the tweets it is observed that nodes 11, 168 and 336 present a high activity, being present in their tweets numerous references to links of web pages ("http://..."), as well as expressions or terms such as "today is about ...", "location", "not be missed", "starting soon" are identified, among others. They also show references to hours and places where different talks take place. With this information it can be deduced that the role played by these nodes within the network can be framed within the concept of organizer of the event (see the main roles that can be identified in a conference setting in section 4.3).

In addition, if the illustrations obtained for each of the centrality measures are observed (Illustration 25, 26, 29 and 30), it can be seen how, in particular, node 336 represents the node with the greatest centrality in all cases. Also, node 11 presents high centrality in the out-degree measure, while it presents a lower value in the in-degree measure, which leads to deduce the predominance of diffusion of information of such node.

As for node 366, it is a user who does not publish tweets, but appears on the network because of the numerous mentions it receives. Specifically, the nodes identified as organizers in previous paragraphs allude to node 366 to announce future talks or to thank and praise talks given. That is, with the information available in the content of tweets, it can be deduced that the role under which this node is framed is that of speaker, who makes certain presentations throughout the days in which the conference is held.

Finally, as regards node 209, its lower activity in the publication of tweets makes it difficult to identify the role it plays within the network. However, after the analysis of its tweets, it is observed that this node focuses on commenting on different talks of the conference once they have taken place. On the other hand, in view of Illustration 28, it is observed that node 209 receives many mentions, being also many of them made by the nodes commented in previous paragraphs (the nodes identified as organizers). With all this information, the role of this node within the network is not completely clear. However, it can be noted the high relation it has with the organizers, as well as its interest in providing information about different talks through the provision of links to web pages ("http://...") and comments about the place and time of them. That is, it can be interpreted that this node plays an important role in the creation of the conference, but that, nevertheless, it does not seem the main person in charge of its organization and dissemination in social networks.

Therefore, it is concluded that, with this more qualitative analysis, it has been identified that nodes with greater centrality within the network are those that have a greater link with the conference, that is, organizers, creators of the event or speakers. Its importance and its influence in different clusters or communities within the network, is information that may be relevant when implementing recommendation systems in a more efficient manner. That is, if it is wanted to influence a community of nodes or a particular node, it can be done through any of these "influential nodes", who play a strong role of "bridges", nodes connectors within the network.

7 Discussion and Conclusions

7.1 Discussion

In the first chapter of the present work an introduction on the addressed topic is made, as well as the research gap in which the present study is framed is presented. The identification of such research gap has shaped the formulation of the main research questions, which are also indicated in such first chapter.

Subsequently, chapters 2, 3 and 4 provide the theoretical basis of this study, while in chapters 5 and 6 the focus is on presenting the empirical part of it.

Once the previous chapters have been conducted, the purpose of this section is to combine the information gathered, that is, to discuss the results obtained in the previous chapters, so that conclusions can be drawn and, therefore, the research questions can be answered.

7.1.1 Answering to the research questions

***RQ1** – What kind of information can be obtained from the Twitter data in the context of a conference?*

The first research question posed focuses on the basis of the analysis, that is, on the available information. It is essential to know the information available and the usefulness that can be given to it to be able to focus and carry out the subsequent analysis. Specifically, the information available contains data related to user profiles (metadata) and published tweets, as well as information about the interactions that occur among them in terms of mentions or retweets made.

With this information that can be obtained from Twitter, both quantitative aspects (number of tweets, mentions, retweets, users, etc.) as well as more qualitative aspects (participant profiles, feelings and preferences of the participants regarding the topics covered in the conference, etc.) can be analysed.

With the compendium of both types of information, conclusions that are framed in the context of network analysis can be reached, since trends, behaviours, connections, identification of important nodes in the network, identification of communities of nodes with similar profiles or identification of main topics can be observed, among others.

RQ2 – What ways are there currently to identify social ties and evaluate tie strength from social media?

After the analysis of the theoretical foundation that serves as the basis for the present study, it can be concluded that there are two main theories on which the different methods for tie strength evaluation from social media data are based. These two theories are: the Strength of Weak Ties and the Structural Holes theories.

The first one is based on the concept of triadic closure, which states that if nodes A and B have a strong connection and A and C also have a strong connection, B and C will have at least a weak connection. On the other hand, the second mentioned theory makes reference to the supposition that the appearance of holes in social structures supposes the existence of weak ties, conforming these a competitive advantage for the nodes next to such holes.

Therefore, based on these ideas, and in terms of a general context, there are many methods that can be used to evaluate tie strength from social media. These different methods differ from one another in the type of information and the social network that is analysed. That is to say, variables such as lists of friends, belonging to same groups, messages exchanged, frequency of interaction, recency of communication, etc. can be taken as object of study. Depending on the social network analysed, one information or another may be available; and depending on the context in which the analysis is framed, it will be interesting to focus on some parameters or others.

RQ3 – Which Twitter data items are related to social ties and tie strength?

There are different Twitter data items that allow their use in the analysis and evaluation of social ties and tie strength. These data items are made up of a set of indicators and predictors that serve as a tool to determine social ties and tie strength.

It should be noted that depending on the type of search performed when collecting the data, there may be variations in terms of the information available. However, basing the information on the study conducted in the present work, some data items as mentions, retweets, common profile features as language or location, common topics and interests (by analysing the text content of tweets) can be highlighted, among others.

Therefore, there are several sources of information that can be taken into account when analysing social ties and tie strength, this choice being conditioned by the context in which the analysis is framed.

***RQ4** – How can that information be used to obtain interesting conclusions related to tie strength and social ties and useful social ties?*

After conducting the object of study of the present work, the usefulness of the information obtained from Twitter can be concluded. In particular, this utility derives from the possibility of constructing implicit networks from such information obtained from the Twitter data.

Implicit networks are a very useful tool when extracting information about users and the connections among them. Implicit networks allow to analyse especially the existence of possible weak ties and latent ties, since these networks are built on the basis of parameters that connect people that have aspects or features in common. That is to say, it is not about networks built on the basis of obvious symptoms of friendship between users, but it is focused on the use of non-evident connecting links between different users. This allows deepening the analysis of social ties, since it allows to identify relevant nodes, which become the protagonists of useful ties within the network, being those that embody a role of connectors or bridges among others.

This is very useful, since useful information about users (nodes), their connections, as well as possible tendencies and behaviours of these can be extracted. In particular, implicit networks (such as the mentions networks built in the present work) allow to observe information of interest such as the identification of the main topics, the identification of the most relevant nodes, trends and connection among nodes, as well as the recognition of different profile models that represent different sets of nodes.

***RQ5** – How can the analysis of implicit networks (mentions networks) in individual and sequential conferences be used in the recognition of social ties and useful social ties?*

The fact of analysing implicit networks in the conference setting helps to carry out the recognition and evaluation of social ties within it, especially, putting the focus on the weak ties, latent ties and potential ties. This is very useful in this context since it allows to help and enhance the recommendation systems used by conference organizers.

That is, for example, the analysis of implicit networks (mention networks in the present work) allows to identify sets of nodes belonging to different clusters. Each of these sets of nodes show similar characteristics in terms of interests and trends within the conference. And this information is of great interest to the organizers when implementing recommendation systems for, for example, future conferences recommended to different users.

Another important example that reflects the usefulness of this implicit networks analysis is the identification of nodes that reach a relevant position within the network. Thus, certain nodes present a greater degree of centrality within the network, becoming the protagonist nodes of the most useful social ties, since they play a role of "bridges". That is to say, the identification of these nodes is very useful for conference organizers, since they assume a fundamental role if it is wanted to reach or influence any node or group of nodes within the network, being able to reach them through these useful social ties.

It can also be observed the structural dispositions presented by the network, being able to extract information related to the identification of small groups of nodes in which a star structure allows to differentiate those nodes that act as information disseminators or connectors between other nodes.

On the other hand, the fact of analysing sequential editions of the same conference allows to deepen and strengthen the conclusions obtained. That is, with this, the tendencies and preferences followed by the nodes can be studied in a more accurate way. For example, one of the first conclusions observed is the tendency of different nodes to maintain their connections and belonging to the same cluster throughout the different annual editions of the conference. This information allows organizers to detect and confirm trends among users, so that they can implement better recommendation systems among users.

7.2 Conclusion

The main objective of this study is to deepen the understanding of the tie strength analysis and evaluation in the context of an event, specifically, a conference. This objective has been materialized through the formulation and response of supporting research questions. And, these five research questions ultimately provide an answer to the primary ontological question of this study: *How can implicit networks be used to recognize social ties in the context of conferences?*

According to what has been commented in the set of the five previous research questions, it can be highlighted that the fact of analysing implicit networks in the context of conferences is a very useful tool when it comes to identifying mainly weak ties, latent ties and potential ties. That is, implicit networks reveal non-evident connections that allow the identification of links between participants, being these connections motivated by some common factor among such users. The identification of strong ties is more linked to the analysis of non-implicit networks, in which obvious connections can be observed between users that allow detecting real friendships between them.

This focus on weak ties and latent ties is completely adequate for the context of study, conferences. The reason for this is that the main objective of this analysis is the possibility of implementing the information obtained in the recommendation systems by the organizers of the event. As it is argued in the theoretical part of the present work, different studies demonstrate the potential of weak and latent ties when it comes to obtaining competitive advantages in a professional, academic or learning environment. On the other hand, also indicated in the theoretical part of the document, two of the main reasons that lead people to attend conferences are the educational benefits they obtain and networking. Therefore, if these two points mentioned in this paragraph are connected, the adequacy of the analysis of implicit networks in the context of the identification of social ties in a conference setting is deduced.

Therefore, the utility of the analysis of implicit networks in the context of conferences can be concluded, since the construction of implicit networks connect people who are related to a common interest, thus showing weak ties and latent ties within that network.

7.2.1 Limitations

In this section some of the most notable limitations that have been found throughout the realization of this study are presented.

Firstly, it should be taken into account that, in the present work, a longitudinal but single-case based approach has been conducted, thus, the results and conclusions obtained cannot be generalized directly to the rest of events. However, these results and methods can be taken as a preliminary study and as a guide for its implementation and comparison with further studies, which can lead to the generalization of the conclusions obtained.

Regarding the realization of the study itself, the main limitation refers to the available data. The Twitter data that is available to conduct the analysis is limited, that is, it is not possible to extract all kinds of information from it. For example, it is not possible to analyse parameters such as the degree of social overlapping circles between two users from the information coming from the available tweets, since in such data the name or identification of the followers or followees does not appear, it is only indicated the total number of them.

In addition, there may be misinterpretations of the data, that is, it may be that one node appears connected to another and that, however, these nodes do not have common characteristics. This would be again due to a certain lack of information in the data that could be included in the analysis.

On the other hand, the methods used in the present study are relevant in a context in which sufficient information is available to carry out the construction of the implicit networks. That is, for contexts in which it is a very small event with little impact on social networks, this method is not appropriate. However, the greater the volume of the event as well as its impact on networks, the better and more consistent conclusions can be obtained.

Furthermore, the present research is appropriate in a context in which, as is the case of the HICSS conference, the conference is focused on a limited range of topics. That is, the utility of identification of weak ties and latent ties is sufficient to establish potentially useful contacts only in contexts in which the participants are very aware and interested in the topics, so that it may be useful for those participants to identify new contacts with common interests. On the other hand, in events that address a wide range of topics, a large number of weak and latent ties could be identified, without being a real potential useful tie.

Therefore, as a result of the main limitations mentioned above, it can be concluded that the results and conclusions obtained in this study are subject to certain restrictions and that, therefore, its generalization cannot be affirmed at all cases. However, the methods, measures and procedures used can be used as a preliminary study or as a guide for future studies that complement the present study, so that a generalization with a solid base can be achieved. That is, it can be concluded that a generalization of the results and conclusions obtained cannot be done, but that, nevertheless, there are methods and procedures that can be generalized and used in other studies.

7.2.2 Future research

In this section it is tried to give a general vision about the path that can be followed to deepen and advance in obtaining conclusions in the field of study of this work.

Firstly, referring again to the information with which to work, it is proposed to conduct more advanced searches. When it comes to obtaining the Twitter data base of the study, it is proposed to carry out searches that not only take into account the tweets that are published under the hashtag with the name of the conference. That is, other searches can be done that take into account the conversations that derive from these main tweets, or through a more detailed analysis through obtaining information from each of the users who participate in such conversations.

On the other hand, the consideration of a greater number of parameters is proposed as future work, constructing a function that constitutes a compendium of the main measurable parameters when building implicit networks. That is, it is proposed the realization of a function that, through the weighting of different indicators obtained through the data, build a more detailed implicit network and with which to obtain deeper conclusions.

In addition, in order to broaden the generalization of the methods used in this study to conferences or events that address a wider range of topics, a more in-depth analysis should first be conducted to reduce the list of possible weak or latent ties. For this, more information could be taken into account through bibliographic data of the participants or other social media data, so that this information can be combined to build more appropriate clusters in the context in which the analysis is framed and, thus, allowing the construction of more appropriate recommendation systems.

Finally, thus, it should be noted that new sources of information could be used. That is, the use of data from other social networking platforms can also be considered in addition to Twitter, or the application of other types of information sources such as bibliographic data or publications, in order to obtain a more complete database that supposes a more consistent analysis and closer to reality.

Therefore, it can be concluded that there is a large space for future studies in relation to the field addressed in the present work. And this space is mainly motivated by the construction of a better and more consistent database that allows an in-depth analysis to obtain conclusions that reflect reality in a more faithful and approximate way.

Appendix 1: PYTHON code

The PYTHON code to create the mentions network from Twitter data is presented in this appendix. In addition, the code implemented to obtain the wordclouds is shown also here. Specifically, in both cases the code used for the case of the year 2010 is shown, being the code for the rest of the years analogously constructed.

*Construction of **Mentions Networks**:*

```
#Upload data
#pip install simplejson
import simplejson as json
with open('tweets_hicss2010.json','r') as f:
  tweets2010 = json.load(f)


#Organize useful mentions data
list2010=[]
n=0
for tweet in tweets2010:

list2010.append({'user_from':tweet['user']['screen_name'],'user_
to':[]})
    for i in tweet['entities']['user_mentions']:
        list2010[n]['user_to'].append(i['screen_name'])
    n=n+1


#Create list with origin and destiny nodes
links2010=[]
for j in list2010:
        for k in j['user_to']:
```

```
            links2010.append({'origin':j['user_from'],'destiny':
k})
```

#Give a unique identifier number to each node through the creation of a dictionary that contains such correspondence: user´s screen name - identifier

```
newdictionary={}
newvalue=1

for i in links2010:
    for j in i:
        if str(i[j]) not in newdictionary:
            newdictionary[str(i[j])]=newvalue
            newvalue = newvalue + 1



link2010=links2010
for i in link2010:
    i['origin']=newdictionary[str(i['origin'])]
    i['destiny']=newdictionary[str(i['destiny'])]
```

#Create the network

```
import networkx as nx
network = nx.DiGraph()

for m in link2010:
    userfrom=m['origin']
    userto=m['destiny']
    if not network.has_edge(userfrom,userto):
      network.add_edge(userfrom,userto,weight=0)
    network[userfrom][userto]['weight'] += 1
```

#Save network as gephi file

```
nx.readwrite.gexf.write_gexf(network,'network2010.gexf',
  encoding='utf-8', version='1.2draft')
```

*Construction of **Wordcloud**:*

#Organize useful data with the content of tweets

```
textlist=[]
for tweet in tweets2010:
    textlist.append({'user':tweet['user']['screen_name'],
'text':tweet['text']})
```

#Give the corresponding identifier to each node

```
for i in textlist:
    if str(i['user']) in newdictionary:
        i['user']=newdictionary[str(i['user'])]
```

#Create a list with the information corresponding to those nodes belonging to the cluster or community from which it is wanted to get the wordcloud

```
clusterlist=[]
for i in textlist:
    if  i['user']==3  or  i['user']==18  or  i['user']==26  or
i['user']==27 or i['user']==33 or i['user']==34:
        clusterlist.append(i['text'])
```

#Import packages to create the wordcloud

```
import numpy as np
import matplotlib.pyplot as plt
import re
from PIL import Image
#pip install wordcloud
from wordcloud import WordCloud, STOPWORDS
```

#Prepare tweets content data for the subsequent construction of the wordcloud: remove links, mentions, special characters, words without interest for the wordcloud, duplicated whitespaces, etc. And creation of a string from the list of texts

```
no_links=[]
for j in clusterlist:
    no_links.append(re.sub(r'http\S+', '', j))
no_mentions=[]
for k in no_links:
    no_mentions.append(re.sub(r"@\S+", "", k))
no_special_characters=[]
for i in no_mentions:
    no_special_characters.append(re.sub('[^A-Za-z ]+', '', i))

raw_string = ' '.join(no_special_characters)
no_unicode = re.sub(r"\\[a-z][a-z]?[0-9]+", ' ', raw_string)
lower_characters=no_unicode.lower()
no_hicss_word=re.sub(' hicss ', ' ', lower_characters)
no_tweet_word=re.sub(' tweet', ' ', no_hicss_word)
no_retweet_word=re.sub(' retweet', '', no_tweet_word)
shortword = re.compile(r'\W*\b\w{1,3}\b')
no_short_words=shortword.sub('', no_retweet_word)

final_string = re.sub(r'\s+', ' ', no_short_words)
words = final_string.split(" ")
words = [w for w in words if w not in STOPWORDS]
```

#Create wordcloud

```
clean_string = ','.join(words)
wordcloud = WordCloud().generate(clean_string)

mask = np.array(Image.open('/Users/Desktop/wordcloudimage.png'))
```

```
wordcloud = WordCloud(background_color="white", max_words=100,
mask=mask).generate(clean_string)


f = plt.figure(figsize=(50,50))

plt.imshow(mask, cmap=plt.cm.gray, interpolation='bilinear')

plt.axis("off")

plt.imshow(wordcloud, interpolation='bilinear')

plt.show()


#Save wordcloud

wordcloud.to_file("/Users/Desktop/cluster1_2010.png")
```

Appendix 2: Timelines of tweets (Tableau)

In this appendix the different timelines obtained for each of the years of study thanks to the Tableau tool is presented. In these figures, the values highlighted with a red dot correspond to those central days during which the conference takes place in each of the editions held.

Timeline 2010:

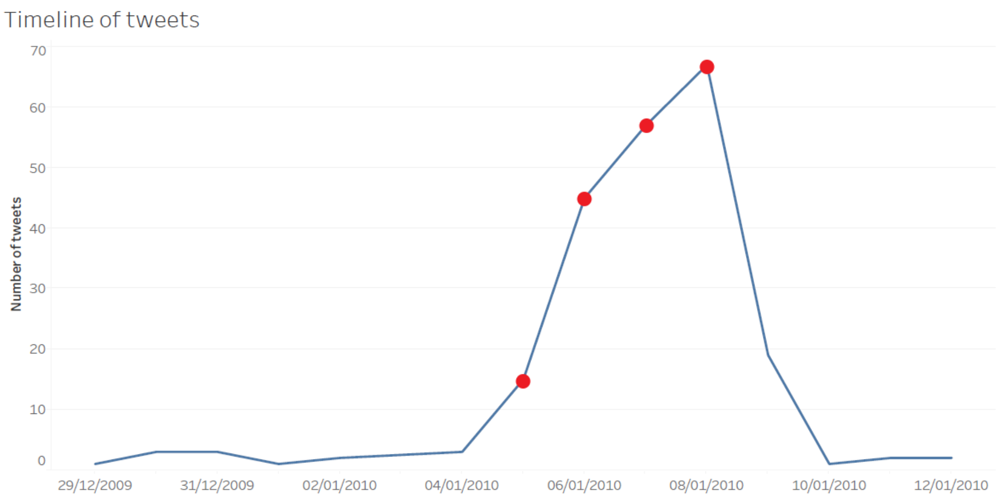

Illustration 34. Timeline 2010

Timeline 2011:

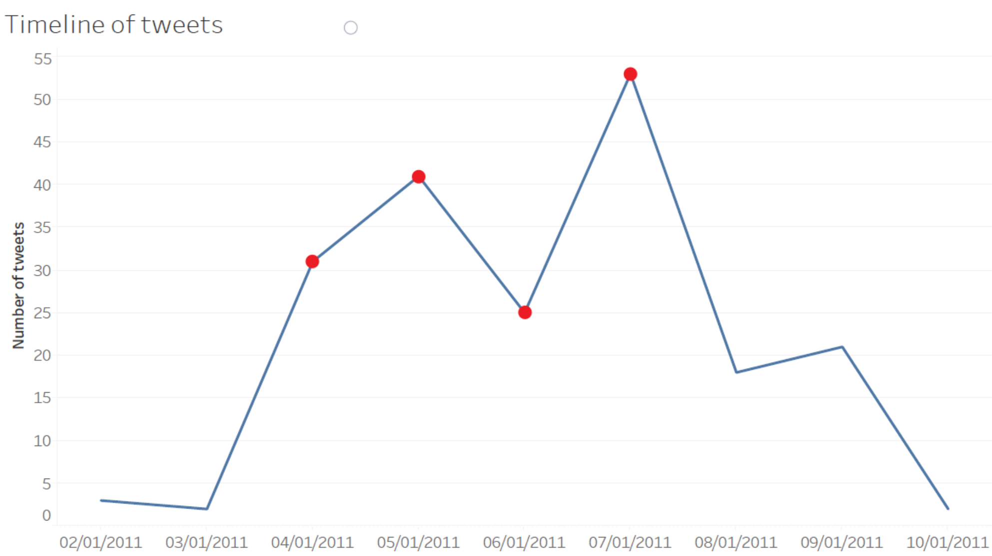

Illustration 35. Timeline 2011

Timeline 2012:

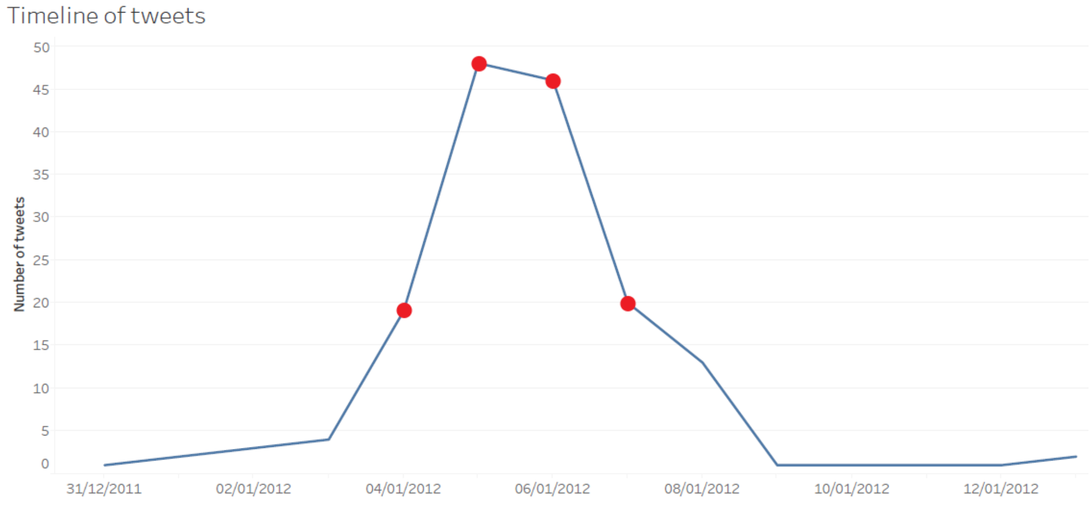

Illustration 36. Timeline 2012

Timeline 2013:

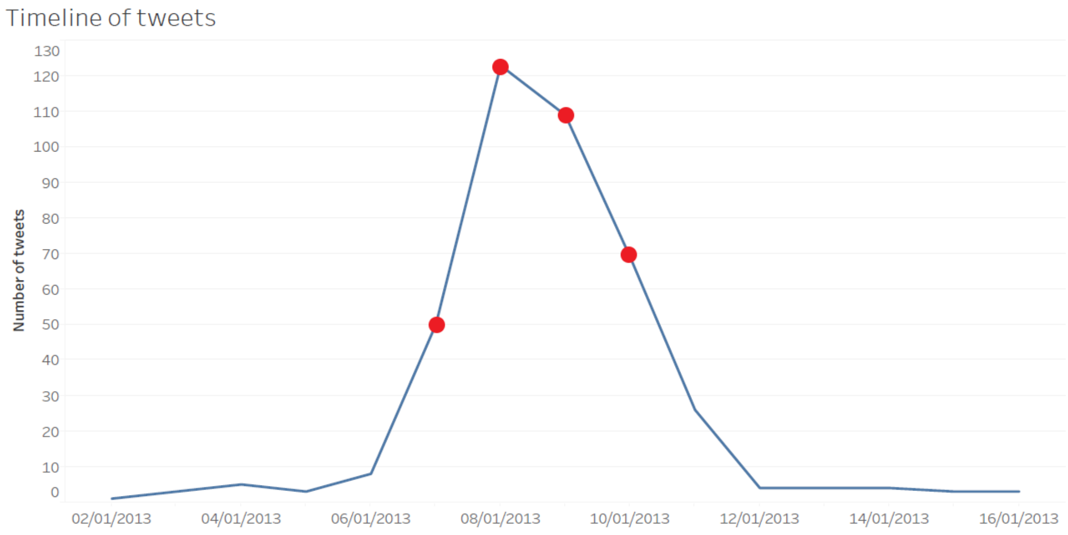

Illustration 37. Timeline 2013

Timeline 2014:

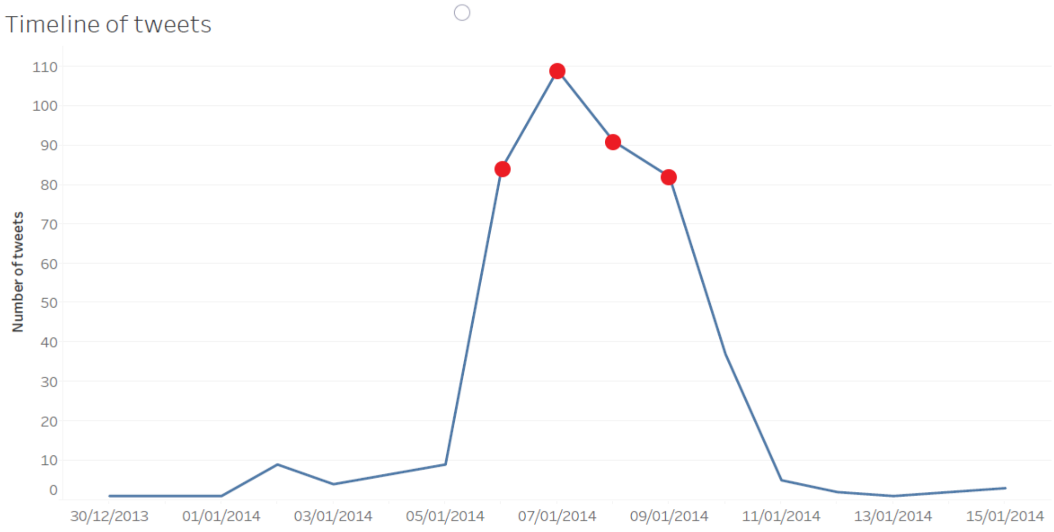

Illustration 38. Timeline 2014

Timeline 2015:

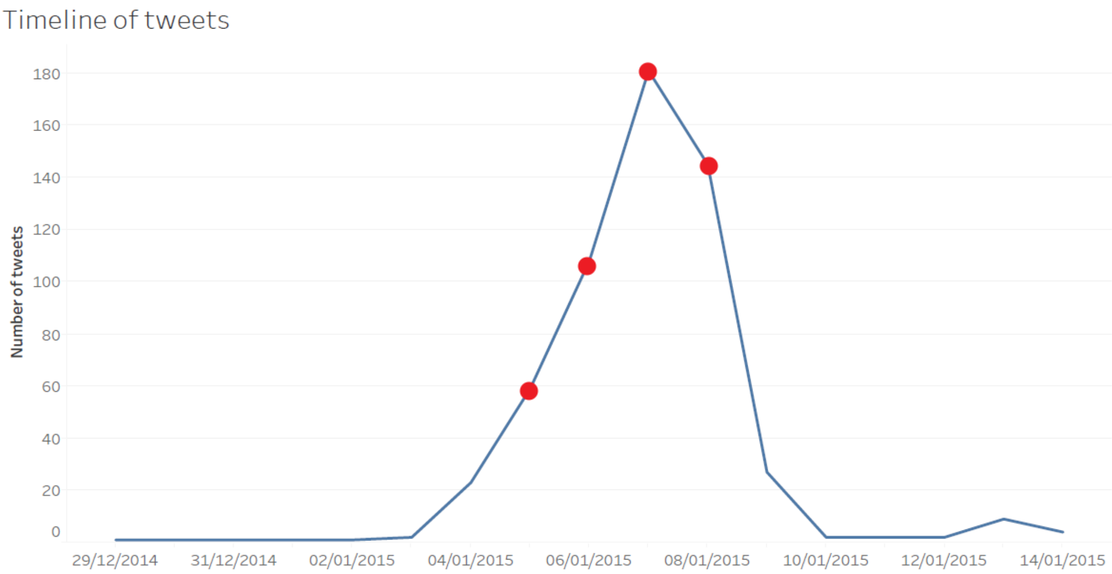

Illustration 39. Timeline 2015

Timeline 2016:

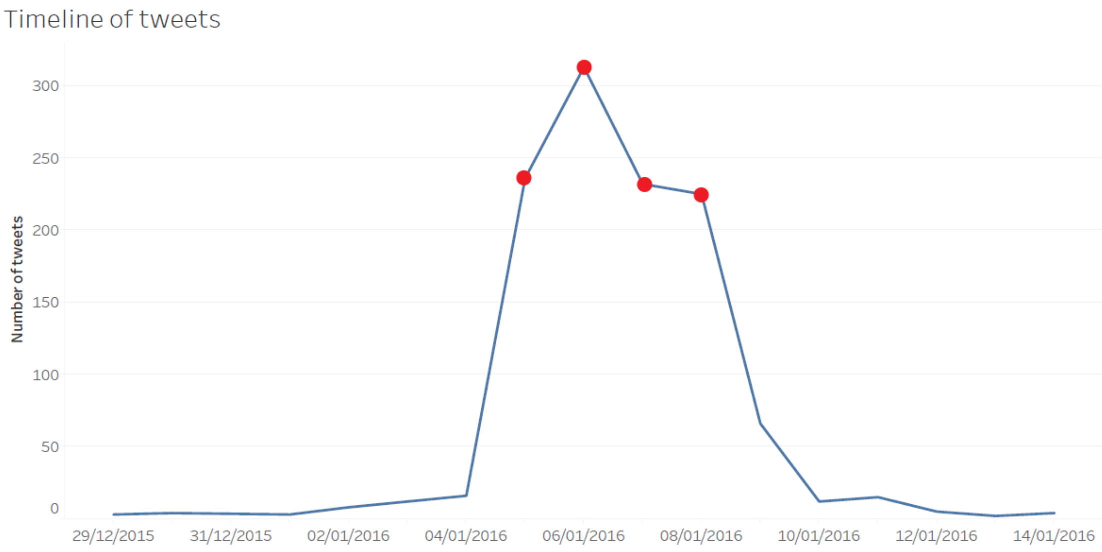

Illustration 40. Timeline 2016

Timeline 2017:

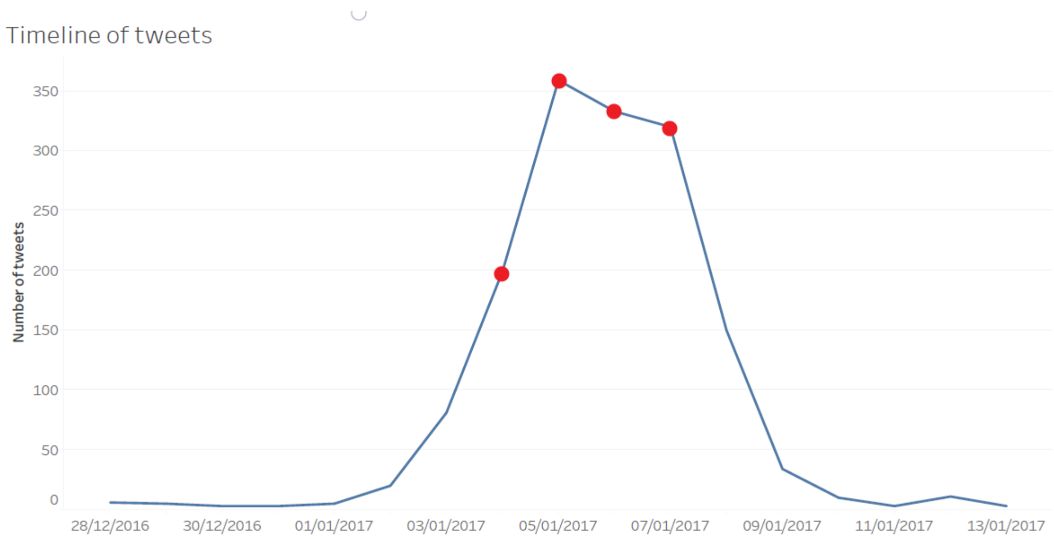

Illustration 41. Timeline 2017

Timeline 2018:

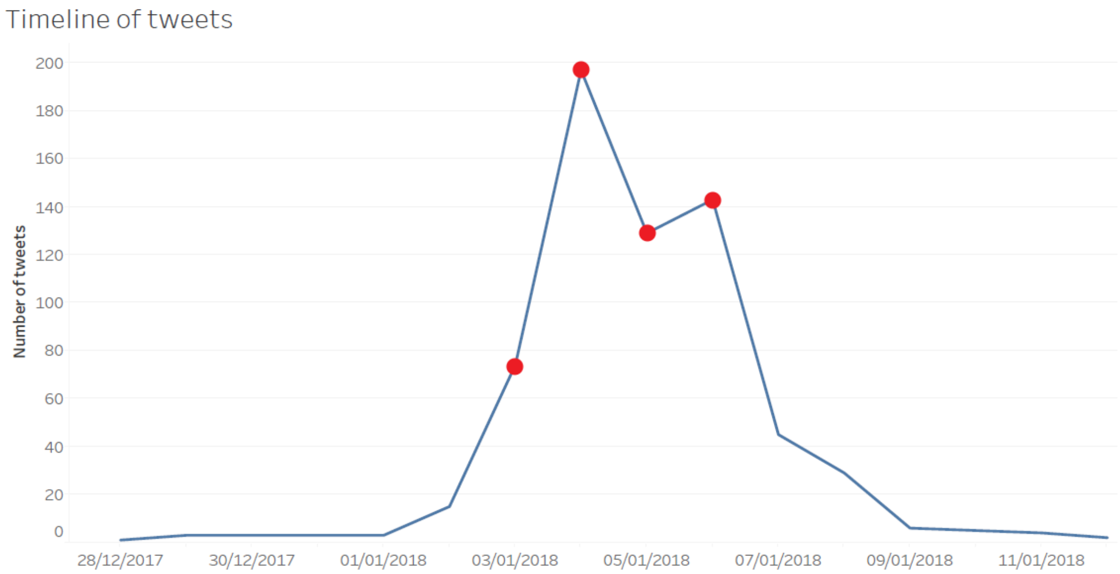

Illustration 42. Timeline 2018

Appendix 3: Rest of the networks (Gephi)

In this appendix the mentions networks obtained with the Gephi tool for the years that have not been shown in the results chapter are presented. As indicated in this chapter, in these networks the identifiers of those nodes that have already appeared in the networks of previous years are indicated (with the exception of the network obtained for 2010, in which all the identifiers are shown).

*Mentions network **2010**:*

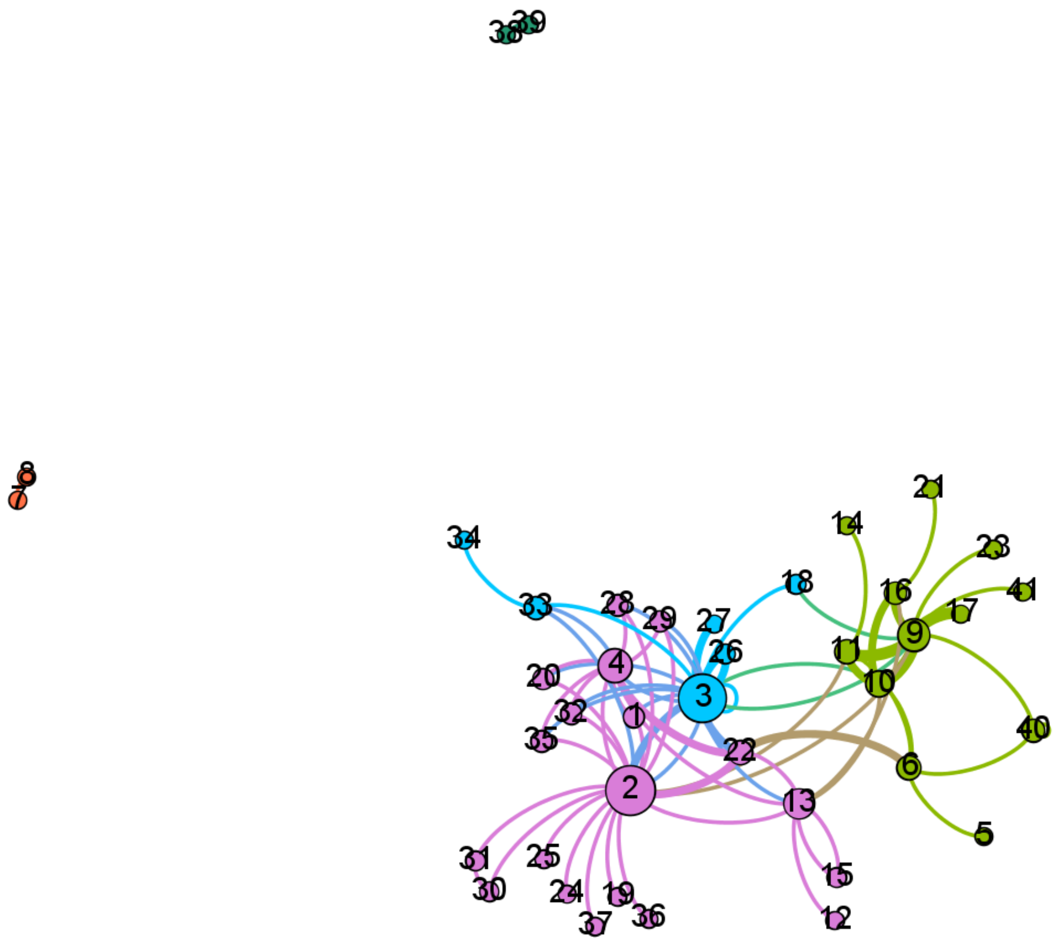

Illustration 43. Network 2010

*Mentions network **2011**:*

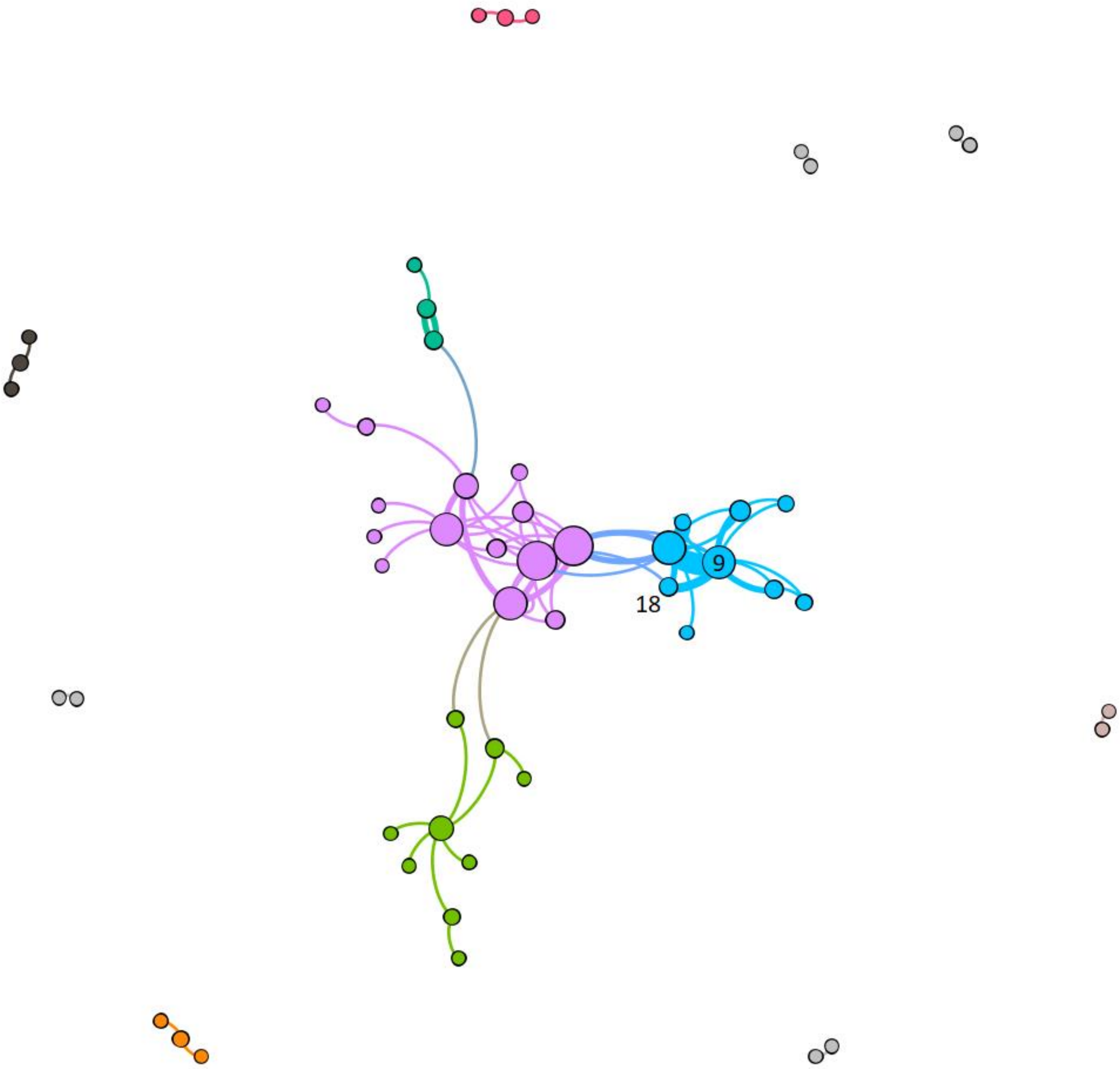

Illustration 44. Network 2011

*Mentions network **2012**:*

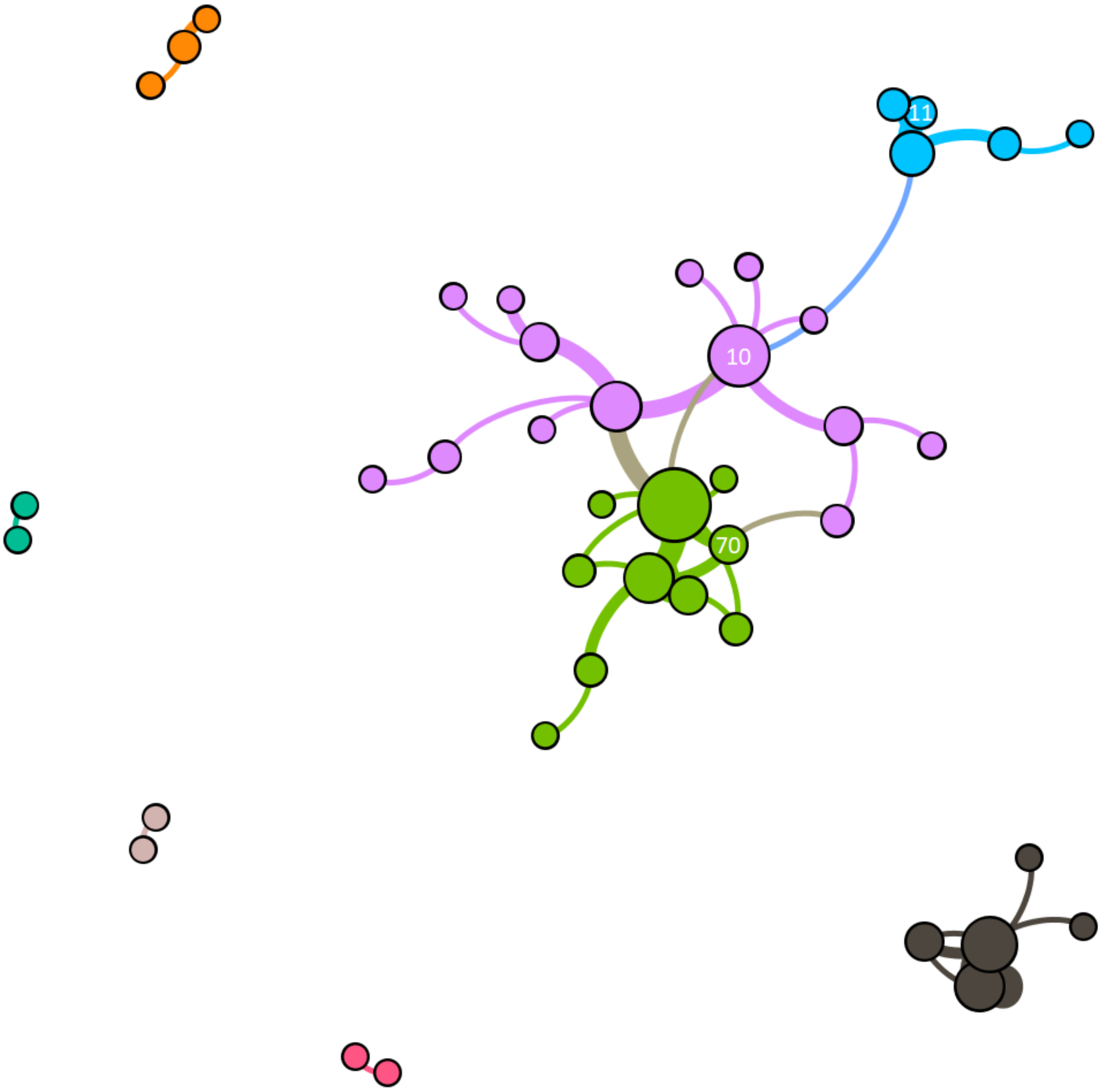

Illustration 45. Network 2012

*Mentions network **2013**:*

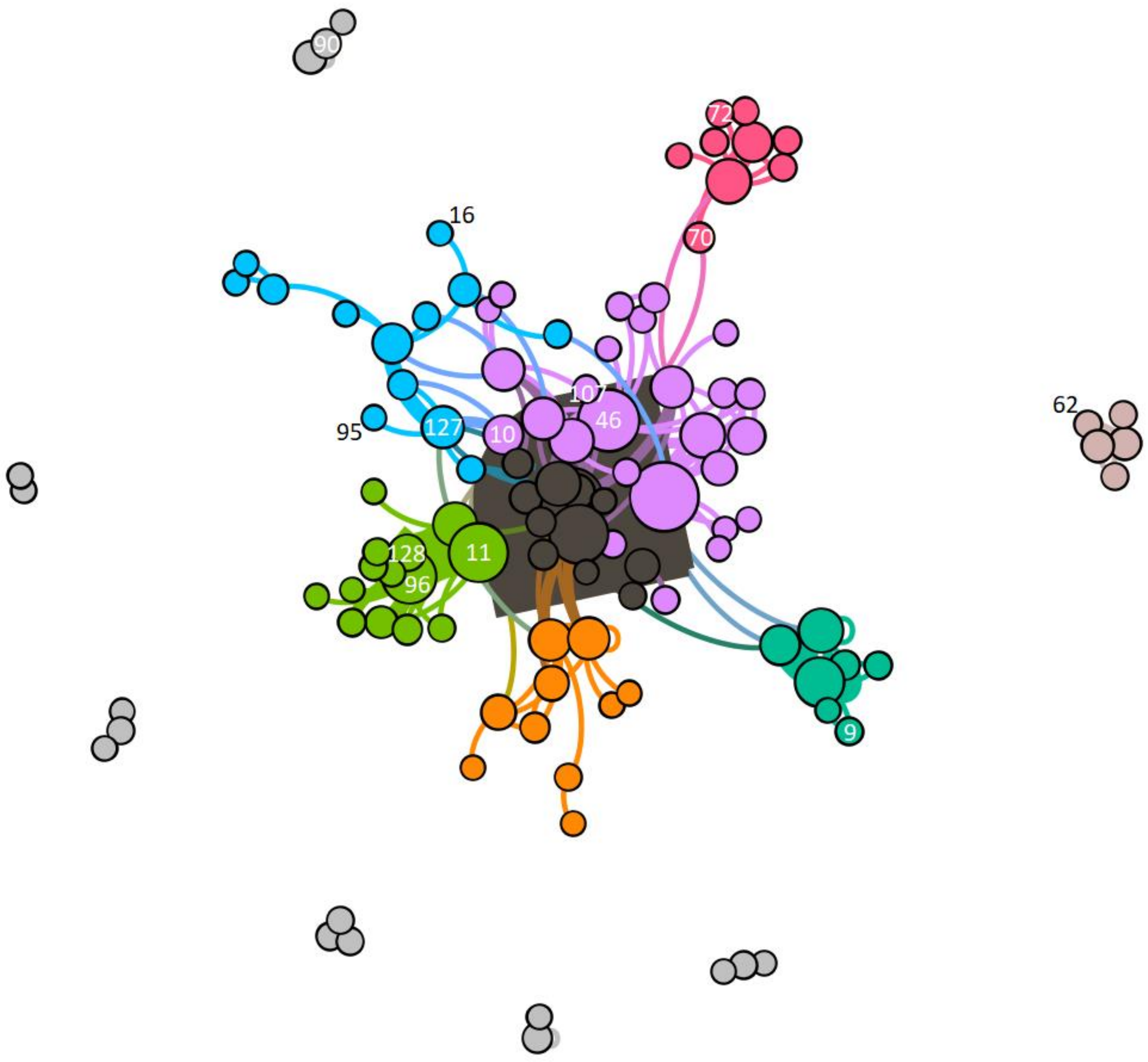

Illustration 46. Network 2013

*Mentions network **2017**:*

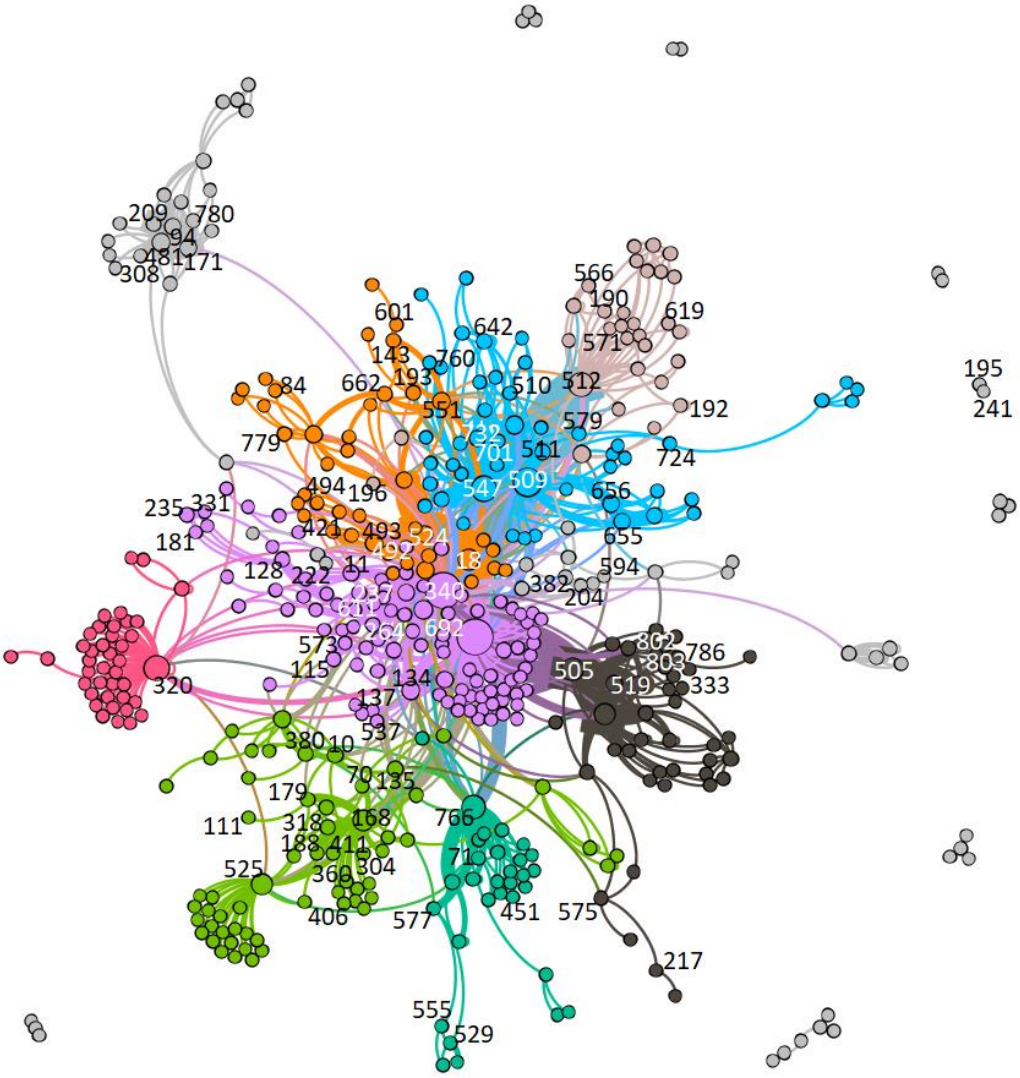

Illustration 47. Network 2017

*Mentions network **2018**:*

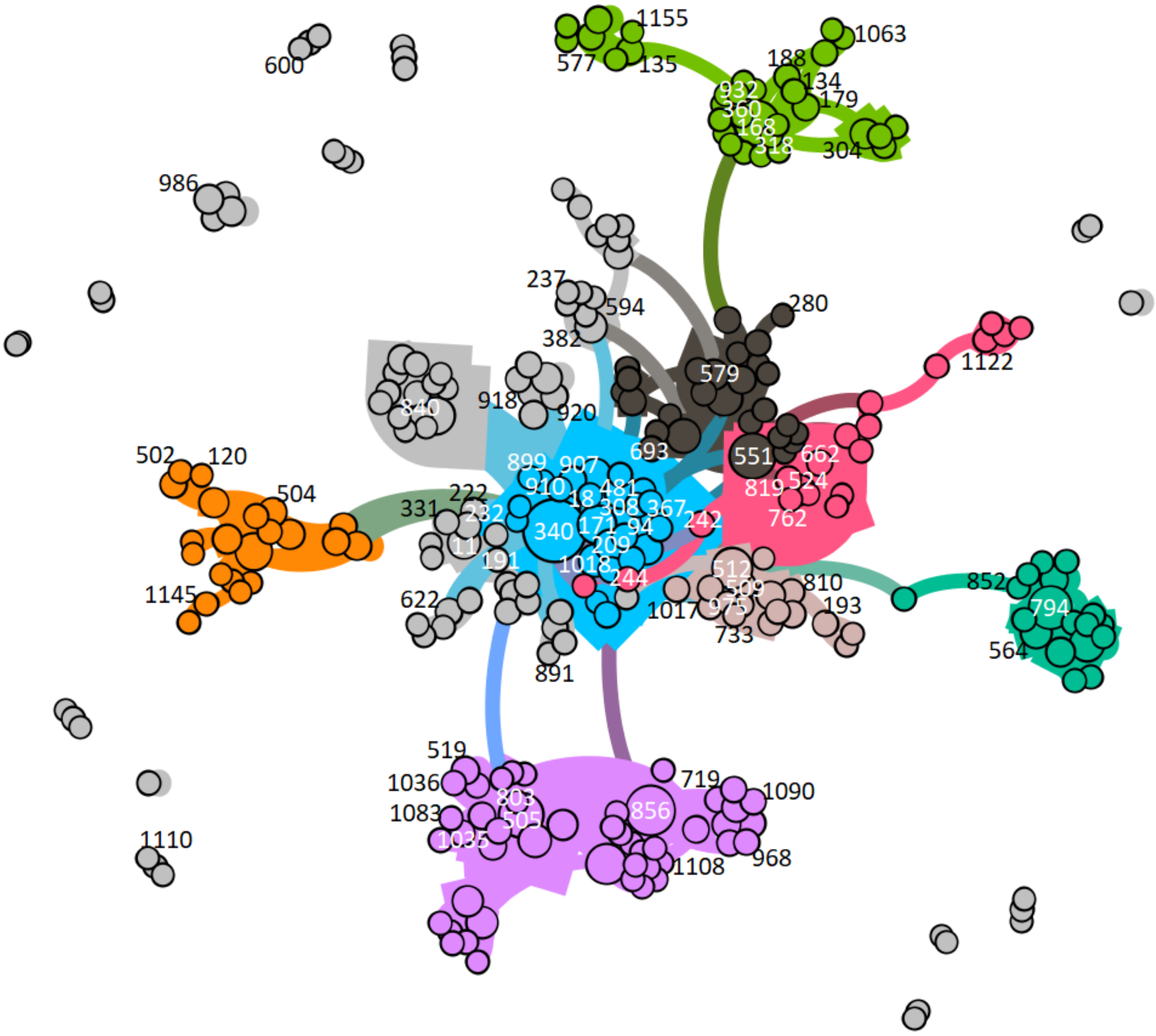

Illustration 48. Network 2018

Bibliography

Alba, R. D. and Kadushin, C. (1976) ‘The intersection of social circles: A New Measure of Social Proximity in Networks’, Sociological Methods & Research, 5(1), pp. 77–102. doi: 10.1177/004912417600500103.

Anaconda Distribution — Anaconda 2.0 documentation. Available at: https://docs.anaconda.com/anaconda/.

Aramo-Immonen, H. *et al.* (2016) ‘Visualizing informal learning behavior from conference participants’ Twitter data with the Ostinato Model’, *Computers in Human Behavior*, 55, pp. 584–595. doi: 10.1016/j.chb.2015.09.043.

Aramo-Immonen, H., Jussila, J. and Huhtamäki, J. (2014) ‘Visualizing informal learning behavior from conference participants Twitter data’, (October), pp. 603–610. doi: 10.1145/2669711.2669962.

Aramo-Immonen, H., Jussila, J. and Huhtamäki, J. (2015) ‘Exploring co-learning behavior of conference participants with visual network analysis of Twitter data’, *Computers in Human Behavior*. Elsevier Ltd, 51, pp. 1154–1162. doi: 10.1016/j.chb.2015.02.033.

Arnaboldi, V., Guazzini, A. and Passarella, A. (2013) ‘Egocentric online social networks: Analysis of key features and prediction of tie strength in Facebook’, *Computer Communications*, 36(10–11), pp. 1130–1144. doi: 10.1016/j.comcom.2013.03.003.

Bastian, M., Heymann, S. and Jacomy, M. (2009) ‘<Gephi_Bastian-Jacomy_AAAI_2009.pdf>’, *Icwsm*, pp. 361–362. doi: 10.1136/qshc.2004.010033.

Beggs, J. J. and Hurlbert, J. S. (1997) ‘The Social Context of Men’s and Women’s Job Search Ties: Membership in Voluntary Organizations, Social Resources, and Job Search Outcomes’, *Sociological Perspectives*. SAGE Publications Inc, 40(4), pp. 601–622. doi: 10.2307/1389465.

Benassi, M., Greve, A. and Harkola, J. (1999) ‘Looking For a Network Organization: The Case of GESTO’, *Journal of Market-Focused Management*, 4(3), pp. 205–229. doi: 10.1023/A:1009875019259.

Betweenness Centrality (Centrality Measure) - GeeksforGeeks. Available at: https://www.geeksforgeeks.org/betweenness-centrality-centrality-measure/.

Blumstein, P. and Kollock, P. (1988) ‘Personal relationships.’, *Annual Review of Sociology*. US: Annual Reviews, pp. 467–490. doi: 10.1146/annurev.so.14.080188.002343.

Borgatti, S. and Halgin, D. (2011) ‘On Network Theory’, *Organization Science*, 22(5), pp. 1168–1181. doi: 10.1287/orsc.1110.0641.

Boyd, D. M. and Ellison, N. B. (2007) ‘Social network sites: Definition, history, and scholarship’, *Journal of Computer-Mediated Communication*, 13(1), pp. 210–230. doi: 10.1111/j.1083-6101.2007.00393.x.

Burt, R. S. (1992) ‘Structural Holes versus Network Closure as Social Capital’, *Advances in Modern Environmental Toxicology*, pp. 19–54.

Burt, R. S. (1995) ‘Structural Holes: The Social Structure of Competition’, *Harvard University Press*. doi: 10.4135/9781412994170.n343.

Closeness Centrality (Centrality Measure) - GeeksforGeeks. Available at: https://www.geeksforgeeks.org/closeness-centrality-centrality-measure/.

Coşkun, M. and Ozturan, M. (2018) ‘#europehappinessmap: A framework for multi-lingual sentiment analysis via social media big data (a Twitter case study)’, *Information (Switzerland)*, 9(5). doi: 10.3390/info9050102.

Easley, D. and Kleinberg, J. (2010) ‘Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge University Press.’ doi: 10.1017/CBO9780511761942.

Ebner, M., Rohs, M. S. and Schön, S. (2010) ‘Getting granular on twitter: tweets from a conference and their limited usefulness for nonparticipants’, *Key Competencies in the Knowledge Society*. doi: 10.1007/978-3-642-15378-5.

Feld, S. L. (1982) ‘Social Structural Determinants of Similarity among Associates’, *American Sociological Review*. [American Sociological Association, Sage Publications, Inc.], 47(6), pp. 797–801. doi: 10.2307/2095216.

Friedkin, N. (1980) ‘A test of structural features of granovetter’s strength of weak ties theory’, *Social Networks*. North-Holland, 2(4), pp. 411–422. doi: 10.1016/0378-8733(80)90006-4.

Gilbert, E. (2012) ‘Predicting tie strength in a new medium’, p. 1047. doi: 10.1145/2145204.2145360.

Gilbert, E. and Karahalios, K. (2009) ‘Predicting Tie Strength with Social Media, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ´09. ACM, New York, NY, USA, pp. 211-220. doi:10.1145/1518701.1518736’, pp. 211–220. doi: 10.1145/1518701.1518736.

Golbeck, J. (2006) 'Generating Predictive Movie Recommendations from Trust in Social Networks BT - Trust Management', in Stølen, K. et al. (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 93–104.

Granovetter, M. S. (1973) 'The Strength of Weak Ties Author (s): Mark S . Granovetter Source : American Journal of Sociology , Vol . 78 , No . 6 (May , 1973), pp . 1360-1380 Published by : The University of Chicago Press Stable URL : http://www.jstor.org/stable/2776392 The Stre', 78(6), pp. 1360–1380.

Gupta, J. P. (2016) 'Evaluating Tie Strength From Twitter Data in Con- Ference Setting : Case Cmad', (April).

Haythornthwaite, C. (2002) 'Strong, weak, and latent ties and the impact of new media', *Information Society*, 18(5), pp. 385–401. doi: 10.1080/01972240290108195.

Haythornthwaite, C. (2005) 'Caroline Haythornthwaite SOCIAL NETWORKS AND INTERNET CONNECTIVITY EFFECTS', (May 2012), pp. 37–41. doi: 10.1080/13691180500146185.

Hornick, M. and Tamayo, P. (2012) 'Extending Recommender Systems for Disjoint User/Item Sets: The Conference Recommendation Problem', *IEEE Transactions on Knowledge and Data Engineering*, 24(8), pp. 1478–1490. doi: 10.1109/TKDE.2011.90.

Hossmann, T. *et al.* (2012) 'Collection and analysis of multi-dimensional network data for opportunistic networking research', *Computer Communications*. Elsevier B.V., 35(13), pp. 1613–1625. doi: 10.1016/j.comcom.2012.05.003.

Jussila, J. *et al.* (2013) 'Information visualization of Twitter data for co-organizing conferences', pp. 139–145. doi: 10.1145/2523429.2523482.

Jussila, J. *et al.* (2014) 'Visual network analysis of twitter data for co-organizing conferences: Case CMAD 2013', *Proceedings of the Annual Hawaii International Conference on System Sciences*. IEEE, pp. 1474–1483. doi: 10.1109/HICSS.2014.190.

Jussila, J. J., Kärkkäinen, H. and Aramo-Immonen, H. (2014) 'Social media utilization in business-to-business relationships of technology industry firms', *Computers in Human Behavior*. Elsevier Ltd, 30, pp. 606–613. doi: 10.1016/j.chb.2013.07.047.

Kaplan, A. M. and Haenlein, M. (2010) 'Users of the world, unite! The challenges and opportunities of Social Media', *Business Horizons*, 53(1), pp. 59–68. doi: 10.1016/j.bushor.2009.09.003.

Krackhardt, D. (1992) 'The strength of strong ties: the importance of philos in organizations', *Organizations and networks: structure, form and action*, (January 1992), pp. 216–239.

doi: 10.1177/1043463103154001.

Kumar, S., Morstatter, F. and Liu, H. (2014) *Twitter Data Analytics*. doi: 10.1007/978-1-4614-9372-3.

Levin, D. Z. and Cross, R. (2004) ‘The Strength of Weak Ties You Can Trust: The Mediating Role of Trust in Effective Knowledge Transfer’, *Management Science*. INFORMS, 50(11), pp. 1477–1490.

Levin, D. Z., Walter, J. and Murnighan, J. K. (2011) ‘Dormant Ties: the Value of Reconnecting.’, *Academy of Management Proceedings*, 2008(1), pp. 1–6. doi: 10.5465/ambpp.2008.33641687.

Liberatore, F. and Quijano-Sanchez, L. (2017) ‘What do we really need to compute the Tie Strength? An empirical study applied to Social Networks’, *Computer Communications*. Elsevier B.V., 110, pp. 59–74. doi: 10.1016/j.comcom.2017.06.001.

Lin, N., Dayton, P. W. and Greenwald, P. (1978) ‘Analyzing the Instrumental Use of Relations in the Context of Social Structure’, *Sociological Methods & Research*. SAGE Publications Inc, 7(2), pp. 149–166. doi: 10.1177/004912417800700203.

Lin, N., Ensel, W. M. and Vaughn, J. C. (1981) ‘Social Resources and Strength of Ties : Structural Factors in Occupational Status Attainment, Am. Sociol. Rev. 46 (4) (1981) 393-405’, *American Sociological Review*, 46(4), pp. 393–405.

Lin, N., Woelfel, M. W. and Light, S. C. (1985) ‘The buffering effect of social support subsequent to an important life event.’, *Journal of Health and Social Behavior*. US: American Sociological Assn, pp. 247–263. doi: 10.2307/2136756.

Malek, A., Mohamed, B. and Ekiz, E. H. (2011) ‘An analysis of conference attendees’ motivations: Case of international conference attendees in Singapore’.

Marsden, P. V. and Campbell, E. K. (1984) ‘Measuring Tie Strength’, 63(2), pp. 482–501.

Marsden, P. V. and Campbell, K. E. (2012) ‘Reflections on conceptualizing and measuring tie strength’, *Social Forces*, 91(1), pp. 17–23. doi: 10.1093/sf/sos112.

Mathews, K. M. *et al.* (1998) ‘Association of Indicators and Predictors of TIE Strength’, *Psychological Reports*. SAGE Publications Inc, 83(3_suppl), pp. 1459–1469. doi: 10.2466/pr0.1998.83.3f.1459.

Nisar, T. M., Prabhakar, G. and Strakova, L. (2019) ‘Social media information benefits, knowledge management and smart organizations’, *Journal of Business Research*. Elsevier, 94(May 2018), pp. 264–272. doi: 10.1016/j.jbusres.2018.05.005.

Number of monthly active Twitter users worldwide from 1st quarter 2010 to 4th quarter 2018 (in *millions) | Statista*. Available at: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/.

Oppermann, M. and Chon, K.-S. (1997) ‘Convention participation decision-making process’, *Annals of Tourism Research*, 24(1), pp. 178–191. doi: 10.1016/s0160-7383(96)00049-7.

Otte, E. and Rousseau, R. (2002) ‘Social network analysis: a powerful strategy, also for the information sciences’, *Journal of Information Science*, 28(6), pp. 441–454. doi: 10.1177/016555102762202123.

Page Rank Algorithm and Implementation - GeeksforGeeks. Available at: https://www.geeksforgeeks.org/page-rank-algorithm-implementation/.

Perlman, D. and Fehr, B. (1987) ‘The development of intimate relationships.’, in *Intimate relationships: Development, dynamics, and deterioration.* Thousand Oaks, CA, US: Sage Publications, Inc, pp. 13–42.

Petroczi, A., Nepusz, T. and Bazsó, F. (2007) ‘Measuring tie-strength in virtual social networks’, *Connections*, 27(2), pp. 39–52.

Pfeffer, J., Mayer, K. and Morstatter, F. (2018) ‘Tampering with Twitter’s Sample API’, *EPJ Data Science*. The Author(s), 7(1). doi: 10.1140/epjds/s13688-018-0178-0.

Quijano-Sánchez, L., Díaz-Agudo, B. and Recio-García, J. A. (2014) ‘Development of a group recommender application in a Social Network’, *Knowledge-Based Systems*, 71, pp. 72–85. doi: 10.1016/j.knosys.2014.05.013.

Reinhardt, W. *et al.* (2009) ‘How people are using Twitter during conferences’, *Creativity and Innovation Competencies on the Web, Proceeding of 5 EduMedia conference*, (March), pp. 145–156.

Retzer, S., Yoong, P. and Hooper, V. (2010) *Inter-organisational knowledge transfer in social networks: A definition of intermediate ties*, *Information Systems Frontiers - ISF*. doi: 10.1007/s10796-010-9250-4.

Rittichainuwat, B. N., Beck, J. A. and Lalopa, J. (2001) ‘Understanding Motivations, Inhibitors, and Facilitators of Association Members in Attending International Conferences’, *Journal of Convention & Exhibition Management*. Routledge, 3(3), pp. 45–62. doi: 10.1300/J143v03n03_04.

Ross, C. *et al.* (2011) ‘Enabled backchannel: conference Twitter use by digital humanists’, *Journal of Documentation*, 67(2), pp. 214–237. doi: 10.1108/00220411111109449.

Sarah Crowe, Kathrin Cresswell, Ann Robertson, Guro Huby, A. A. and A. S. (1993) ‘The Case

Study Approach', *Business Communication Quarterly*. doi: 10.1177/108056999305600409.

ScholarSpace at University of Hawaii at Manoa: Hawaii International Conference on System Sciences (HICSS). Available at: https://scholarspace.manoa.hawaii.edu/handle/10125/39610.

ScholarSpace at University of Hawaii at Manoa: Hawaii International Conference on System Sciences 2019. Available at: https://scholarspace.manoa.hawaii.edu/handle/10125/59440.

Servia-Rodríguez, S. *et al.* (2014) 'A tie strength based model to socially-enhance applications and its enabling implementation: MySocialSphere', *Expert Systems with Applications*. Elsevier Ltd, 41(5), pp. 2582–2594. doi: 10.1016/j.eswa.2013.10.006.

Severt, D. *et al.* (2007) 'Examining the motivation, perceived performance, and behavioral intentions of convention attendees: Evidence from a regional conference', *Tourism Management*, 28(2), pp. 399–408. doi: 10.1016/j.tourman.2006.04.003.

Shi, X., Adamic, L. A. and Strauss, M. J. (2007) 'Networks of strong ties', *Physica A: Statistical Mechanics and its Applications*, 378(1), pp. 33–47. doi: 10.1016/j.physa.2006.11.072.

Sousa, D., Sarmento, L. and Rodrigues, E. M. (2010) 'Characterization of the twitter @replies network: Are user ties social or topical?', *2nd International Workshop on Search and Mining User-Generated Contents, SMUC'10, Co-located with 19th International Conference on Information and Knowledge Management, CIKM'10, October 26, 2010 - October 30, 2010*, pp. 63–70.

Spyder — Anaconda 2.0 documentation. Available at: https://docs.anaconda.com/anaconda/user-guide/tasks/integration/spyder/.

Weller, K. *et al.* (2014) *Twitter and Society*.

Wellman, B. (1982) 'Studying Personal Communities'.

Wellman, B. and Wortley, S. (1990) 'Different Strokes from Different Folks: Community Ties and Social Support', *American Journal of Sociology*, 96(3), pp. 558–588. doi: 10.1086/229572.

Wollan, R., Smith, N. and Zhou, C. (2010) 'The Social Media Management Handbook: Everything You Need To Know To Get Social Media Working In Your Business'.

Xiang, R., Neville, J. and Rogati, M. (2010) 'Modeling relationship strength in online social networks', p. 981. doi: 10.1145/1772690.1772790.

Zhang, A., Bhardwaj, A. and Karger, D. (2016) 'Confer: A Conference Recommendation and

Meetup Tool', pp. 118–121. doi: 10.1145/2818052.2874340.

Zhang, B. *et al.* (2017) 'The role of weak ties in enhancing knowledge work', pp. 210–215. doi: 10.1145/3131085.3131119.

Zhong, Y., Yang, J. and Nugroho, R. (2015) 'Incorporating Tie Strength in Robust Social Recommendation', in *2015 IEEE International Congress on Big Data*, pp. 63–70. doi: 10.1109/BigDataCongress.2015.19.