Universidade do Minho

Escola de Engenharia

Departamento de Informática

Jorge Caldas

**Analysis and Visualisation of
Dynamic Social Networks**

October 2016

Universidade do Minho
Escola de Engenharia
Departamento de Informática

Jorge Caldas

Analysis and Visualisation of Dynamic Social Networks

Master dissertation
Master Degree in Computer Science

Dissertation supervised by
Pedro Rangel Henriques
Alda Lopes Ganarski

October 2016

ACKNOWLEDGEMENTS

Write acknowledgements here

ABSTRACT

Write abstract here (en) or import corresponding file

RESUMO

Escrever aqui resumo (pt) ou importar respectivo ficheiro

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

O

OSN Online Social Network.

S

SN Social Network.

SNA Social Network Analysis.

1

INTRODUCTION

1.1 CONTEXT AND PROBLEM

1.2 MOTIVATION

1.3 GOALS

SOCIAL NETWORKS IN SOCIOLOGY

Nowadays is hard to find something that is not organized as a network, if one tries to understand something about the world around us, then definitely one needs to know something about networks.

Curiously if you look up the term *Social Network (SN)* in the Dictionary (2002), we may face the following:

> *"a website or computer program that allows people to communicate and share information on the Internet using a computer or mobile phone"*

But, even if today we automatically think in SNs as websites (or web applications), deep down we know when talking about SNs, we refer to a much more broader term, that said, we may consider a SNs as the following:

> *"A social structure made of nodes that are generally individuals or organizations. A social network represents relationships and flows between people, groups, organizations, animals, computers or other information/knowledge processing entities. The term itself was coined in 1954 by J. A. Barnes."* Beal (2016)

One may say that networks work like pipes, and trough them things flow, from individual to individual inside the network. It's trough networks that big institutions can organize themselves, and actually add value to society despite the large number of individuals.

2.1 ORIGINS OF SOCIAL NETWORKS

> *"The network concept is one of the defining paradigms of the modern era."* Kilduff and Tsai (2003)

Before talking of network from the sociology perspective, one needs to review the network concept, which is broadly used across multiple fields of study, this include, physics, biology, linguistic, anthropology, mathematics, computer science and more recently computer networks.

But why is the network approach so adopted in such diversification fields? According to Kilduff and Tsai (2003), the answer is, because networks allows us to capture the interactions of any individual unit within the larger field of activity to which the unit belongs.

2.1.1 Sociology Perspective

> "(...) many people attribute the first use of the term "social network" to Barnes (1954). The notion of a network of relations linking social entities, or of webs or ties among social units emanating through society, has found wide expression throughout the social sciences. (...)" Wasserman and Faust (1994)

The SN concept has been around for many years now, maybe not in the exact format that nowadays, we are familiarized with ("*web way*", in a manner of speaking), but in a more abstract sense, applied in real life within real connections. Wasserman and Faust (1994), refer that this term has first came into discussion in 1954, introduced by Barnes, J.A.

> "Social relations in Bremnes, Norway, fall into three categories: relatively stable formal organizations serving many different purposes, unstable associations engaged in fishing, and interpersonal links that combine to form a social network and on which perceptions of class are based. In fishing situations, orders are given and obeyed; in the other social settings, consensus decisions are reached obliquely and tentatively." Barnes (1954)

In the above citation, John Arundel Barnes, does a very well succeed reflection about the relationships of the people from Bremnes (Norway). The author points out that relations can form organizations for serving a specific purpose, and today we clearly see that the chosen path of SNs and also OSNs, was narrow down SNs to very specific purposes, such as professional networks. So one may say that John Arundel Barnes not only coined the term SN, but also was one of the first who described **interest-based social networks**.

2.2 RELEVANT SN RELATED TERMS

In this section talk about some inherent concepts of SN, only if they are found relevant. (Review this theories. Why are they important in sociology? What is their placement (fitting) in the thesis?)

- Homophily and Heterophily

- Structuralism

- Structural functionalism

- Conflict theories

- Social constructionism

3

ONLINE SOCIAL NETWORKS

People need to connect other people, and the urge for connection, bring to us what today are known as OSNs. This web sites allows to define a profile as an individual, and to share and visualize content with other individuals in the network, therefore connecting.

> *"We define Online Social Networks as web-based services that allow individuals to construct a public or semi-public profile within a bounded system, articulate a list of other users with whom they share a connection, and view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site.* Ellison et al. (2007) [1]

---

1 A table is presented on the next page. The blank space is due to the size of the table.

Name	Year of launch	Registered Users	Active Users[2]	Description/Purpose
Facebook	2004	-	1 712 000 000	**General**. Photos, videos, blogs, apps.
Google+	2011	1 600 000 000	-	**General**. Google+ is an interest-based social network that is owned and operated by Google.
Youtube	2005	1 000 000 000	-	Allows billions of people to discover, watch and share originally-created videos. Provides a forum for people to connect, inform, and inspire others.
Qzone	2005	-	652 000 000	**General**. It allows users to write blogs, keep diaries, send photos, listen to music, and watch videos. It's only available in Chinese.
Twitter	2006	645 750 000	313 000 000	**General**. Micro-blogging, RSS, updates.
Tumblr	2007	-	555 000 000	Microblogging platform and social networking website.
Instagram	2010	-	500 000 000	A photo and video sharing site.
Sina Weibo	2009	300 000 000	282 000 000	Social microblogging site in mainland China.
VK	2006	249 409 900	100 000 000	**General**, including music upload, listening and search. Popular in Russia and former Soviet republics.
LinkedIn	2003	200 000 000	106 000 000	Business and professional networking.
Vine	**2013**	200 000 000	-	Short-form video sharing service where users can share six-second-long looping video clips.
Pinterest	2010	176 000 000	100 000 000	The worlds catalog of ideas. Find and save recipes, parenting hacks, style inspiration and other ideas to try.
Reddit	2005	35 000 000	-	Social media, social news aggregation, web content rating, and discussion website.
Flickr	2007	32 000 000	-	Helping people make their photos available to the people who matter to them. Enable new ways of organising photos and video.
Meetup	**2002**	27 590 000	-	World's largest network of local groups. Meetup makes it easy for anyone to organize a local group or find one of the thousands already meeting up face-to-face. meetup.com (2016)
Couchsurfing	2004	12 000 000	-	Couchsurfing connects travellers with a global network of people willing to share in profound and meaningful ways, making travel a truly social experience. Is commonly used by travellers to find free hosts across the globe. couchsurfing.com (2016)
ResearchGate	2008	10 000 000	-	Built by scientists, for scientists. Connect the world of science and make research open to all. researchgate.net (2016)

Table 1: Table of OSNs (statista.com (2016), expandedramblings.com (2016))

The Table 1 lists the most used and popular OSNs, **ordered by the estimated number of registered users**.

The first obvious comment on the listed OSNs is that general purpose OSNs have more users (social networks with the word *General* in bold), being Youtube an exception, since it is not a general purpose OSNs, neither is focused on individuals, it is build around **social objects**, the videos.

The grey scale in the first column of Table 1 divides OSNs in three groups: the first and smallest, the 1 billion or more users OSNs; the second the OSNs with less than 1 billion users and more then 100 million; finally, the third group, OSNs with less then 100 million users. At this point, we begin to observe that **the narrower purpose OSNs** such as ResearchGate (mainly for researchers) or Couchsurfing (mainly for open minded travellers), **have a smaller number of registered users**, which is expected since the target audience is also smaller.

Other OSNs not listed in the Table 1, but still worth mentioning include **Classmates** (helps users finding classmates form kindergarten, primary school, high school etc.) known for being one of the first OSNs, since it was launched in 1995, and **Ask.fm** (allows users to interact with other users asking and answering questions (revealing identity is optional)).

An important note on the listed OSNs in Table 1 is that, only Qzone, Vine, Couchsurfing and ResearchGate don't provide any web APIs to fetch data or publish content, while all the others, offer a wide variety of web services for developers to consume and use as they please within the terms and policies of use of course of each OSN.

3.1 HISTORY OF ONLINE SOCIAL NETWORKS

Although the first platform possessing some of the main characteristics that define OSNs, according to Ellison et al. (2007), the first recognizable OSN launched in 1997 as we can observe in the Figure 1. *SixDegrees.com* allowed users to create personal profiles, connect with friends and consult friends of friends lists. The profile feature came from the online dating sites and online communities, while the surfing trough register users in the network and consult friends was an existing feature in Classmates.com. *SixDegrees.com* was the first to combine these features.

SixDegrees promoted itself as a tool to help people to connect, but in 2000, it became an unsustainable business and the service closed. At the time the creators conclude that *SixDegrees* was a service that was very ahead of its time.

Until 2002 many OSNs have emerged, but still incapable of projecting themselves at a global scale. As we can observe in the timeline of Figure 1 from 2002 and 2005 the *big players* came to existence, in these period, OSN such as Friendster, LinkedIn, MySpace, Hi5, Facebook and Youtube were born, shaping the business, cultural and research landscape.
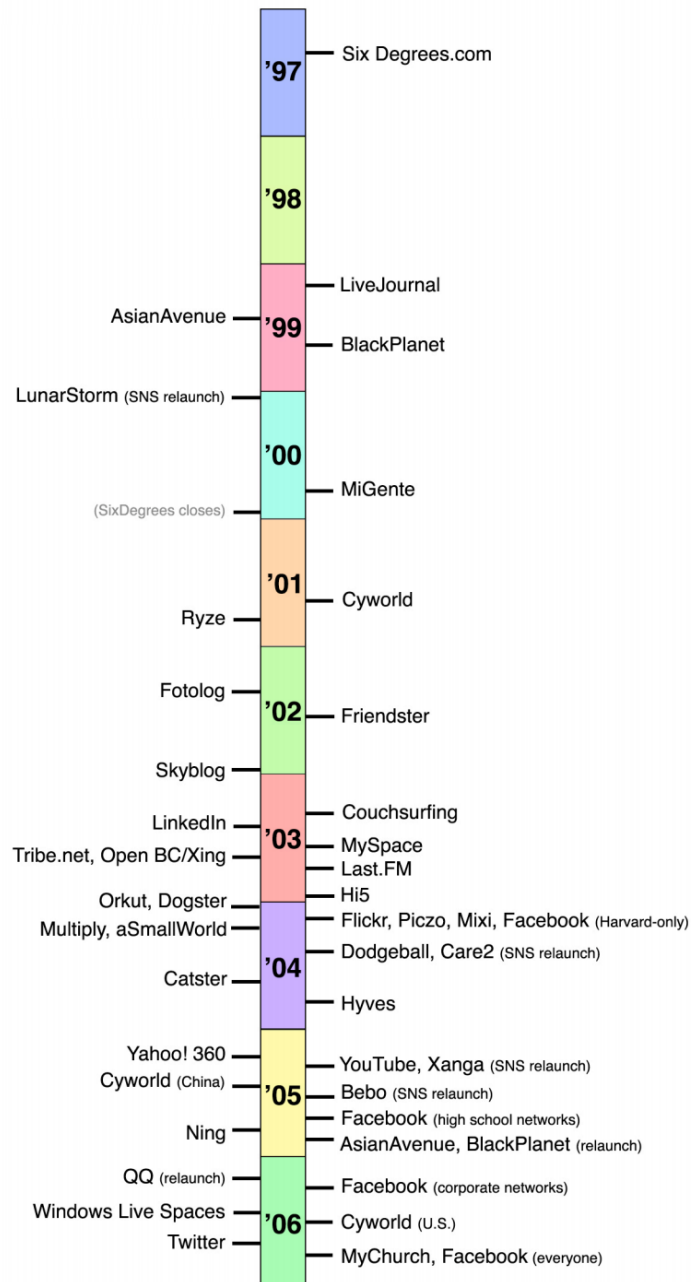
Figure 1: Lauch dates of major OSNs. (Ellison et al. (2007))

3.2   PORTUGUESE AND ONLINE SOCIAL NETWORKS

From Table 1, we get a good overview on OSNs usage among modern society. In this section we do a deep exploration of the most adopted OSNs by portuguese citizens, and get to compare then with the more global scenario presented in Table 1, also, other interesting

facts will be revealed where appropriate.

A recent study, Marktest (2016), revels portuguese relationship with OSNs. This study, has been made by *Marktest Consulting* since 2011, with the goal of know the notoriety, utilization, opinion and habits of portuguese concerning social networks. The study information was collected trough online interviews. The sample was built from 819 interviews from individuals with age between 15 and 64 years, living in Portugal and using OSNs in a daily basis.

Some of the most interesting facts revealed in this study, relative to the participants are:

- 94% has a Facebook account and 43% a Youtube account;

- 21% has abandoned a social network in the past year;

- 27% considers that their dedicated time to social media has increased;

- 67% follows celebrities and 62% follows brands;

- 87% is used to watch videos in social networks.

These are indeed interesting conclusions, but what about the top used OSNs, when it comes to that, **the most used are the following (by order): Facebook, Youtube, Google+, Instagram, LinkedIn, Twitter and Pinterest**.

Relatively to Marktest (2016) past studies, Facebook has maintain its top position, maintaining a grow tendency that has been standing out in the past years.

Going back to Table 1, we may now comment the portuguese OSNs comparing it to the global scenario. As one may notice Facebook still rules users preferences within portuguese. The other noticeable point is that the OSNs preferred among portuguese are general propose ones, but with a slight tendency to content sharing networks (mainly photos).

Concerning to global time related usage statistics, according to Marktest (2016), **portuguese spend 91 minutes a day with social networks**, 68% considers that this is the ideal time to spent with social media, despite 1 in each 4 saying that in the past year has dedicated even more time to them. Even if people spent more than one hour and an half in this platforms, the study, concluded that **67% of the users that visit OSNs several times a day only 41% does daily publications**.

The prime time for using OSNs is between 8pm and 10pm, being the smartphone the most used device in this time. Also in this short period the featured OSN is Facebook, the majority of the interviewed say that is the most credible site, the one that provides better and useful information, the most interesting and addictive.

3.3 EXPLORING SPECIFIC ONLINE SOCIAL NETWORKS

In this section we are going to explore in greater detail some of the OSNs presented in the Table 1. The selection of the social networks was not aleatory, we are going to study deeply the OSNs that gather some important characteristics, that will be of use in the future when we design the system for analyzing and visualizing social networks. First, the OSN must be accessible, this said, one must be capable of extracting information from the platform in order to analyze it. Second, in order to obtain large networks, and since this project is being developed in Portugal, OSNs that are known to be massively adopted by portuguese (this topic was addressed in the previous section, where we did an overview to Marktest (2016)). Third and at last, the OSNs must be the most diversified as possible in terms of their porposes, so that we can draw different types of conclusions derived from different kind of analysis, for then give proof of the adaptability of the system to different kinds of OSNs.

This said, these are the following OSNs that will explored deeply in the proceeding sections:

- Facebook

- Twitter

- LinkedIn

- ...

3.3.1 *Facebook*

Facebook is an OSN, created by Mark Zuckerberg in 2004, which started out by being an exclusive social network for Harvard students, but came later to spread across the country and the globe, having today more than one billion users.

Before diving into details of Facebook's domain, one must first point out some of its general aspects. Facebook basically allows anyone with a valid email address to create a public and personalized profile, we say personalized in terms of displayed content or information such as profile photo, name, work, homeland, education etc. . The next fundamental step is connect with other users, by sending friendship requests to other Facebook users (this are bidirectional relations). The base entity of the network is the user, but entities such as brands, companies can also be part of the platform, appearing normally in the form of page, being a page a public place inside the network with marketing or business related purposes (celebrities, public institutions also use pages as form of appearing in Facebook).

The next parts of this section will clarify the roles of this entities and their way of interact with each other, also other important concepts will be presented.

Domain Model
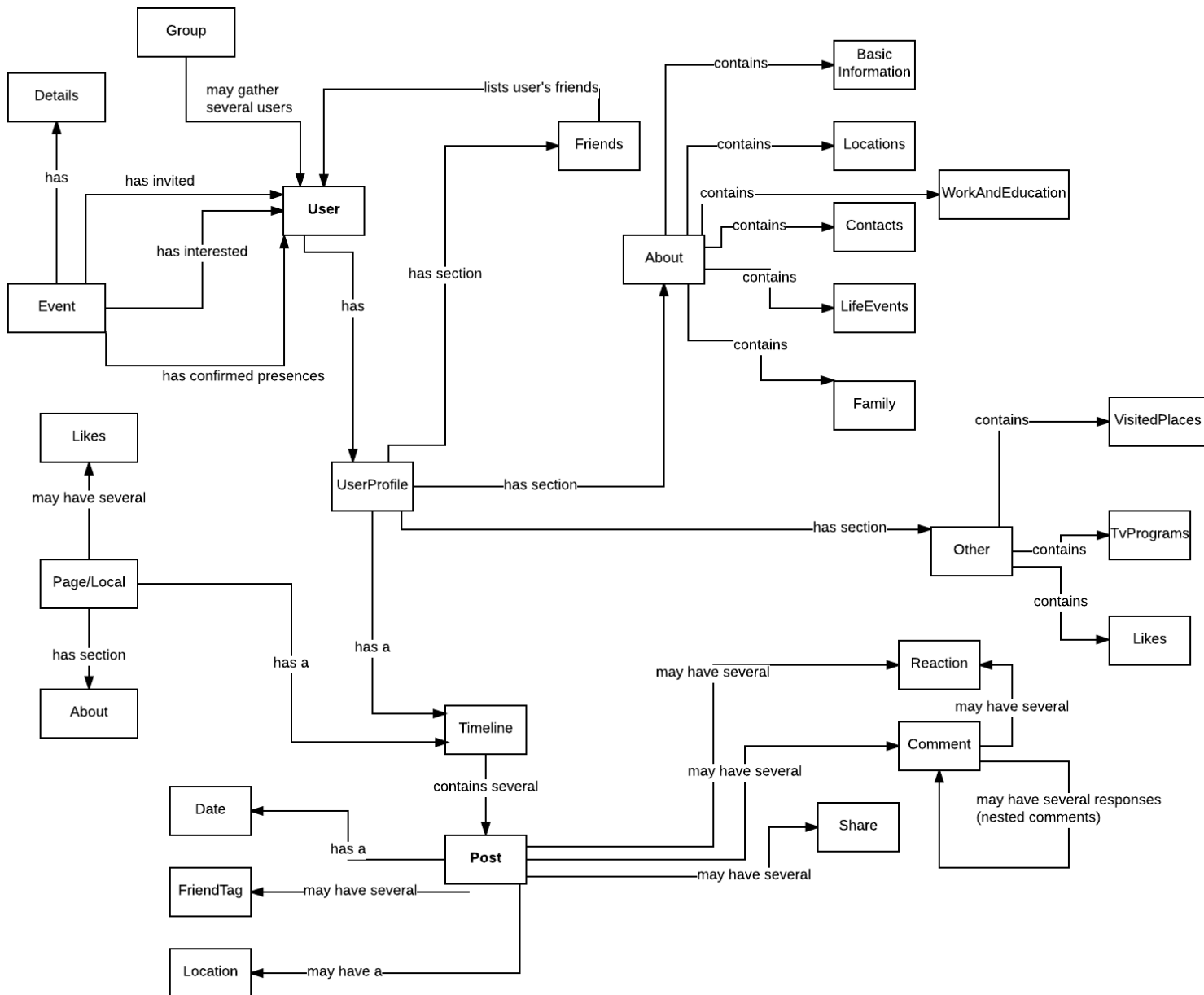
Figure 2: Facebook domain model schema.

In this section we explore the domain of Facebook represented in Figure 2 in detail, what are the pieces that conceptually build this platform, and how they relate. The schema in Figure 2 represents a macroscopic perspective among Facebook components and their organization.

There are two entities with bold labels in the schema, this are, **User** and **Post**, being *User* the base entity in the network (the node in the network graph basically), and *Post* the most

basic unit of content sharing in Facebook.

Facebook is interesting in terms of data gathering, because despite offering users' basic information and to whom that users are related (*Friends* box), it has a collection of other interesting data such as the family relationships (*Family* box), geographical locations where the user lives, or visited locations (*Locations* and *VisitedPlaces* boxes respectively), and among other things, user information may contain the personal interests that were explicitly inputed by the user (*Likes* box).

In what concerns to user activity in the platform, the *Timeline*, provides all the user Posts chronologically ordered, this is where Facebook dynamism takes place, users are constantly adding content to their timeline, it may be life related events or simply sharing other users posts linking content.

Facebook has, with time, become, more then a user profile centralized network, it has invested in expand its horizons, becoming the place where pages of brands, companies, organizations (media, political, non-profitable etc.), or places (cities, monuments, bars etc.) live (*Page/Local* box). This entities that are now cohabiting with users in the Facebook ecosystem, take advantage of the platform and its range to get their updates to most people as possible. The profile for this pages are in many ways different form the user's profile, it also haves a timeline, but the about information and other details represent a smaller part of this profiles. The most important metric for pages, is its number of *likes* (*Likes* box), it represents the number of users in the network that follow the page, it might be users that simply have a certain relation with the entity or simply want to keep in touch by regularly receiving this entities updates in their Facebook walls [3].

Other Facebook entities not yet mentioned, are events (*Event* box). This are events inputed in the platform that allow users to keep updated about relevant events happening mainly in their area. Users can tag the event as *interested in*, showing their friends the will of participating in some event, or they can simply reject the event. Users also can confirm participation on events showing their network that they will be present. Events keep three separated counters for users, they count the number of invited users, number of interested users and number of confirmed users (this relations are expressed as links between the *Event* box and the *User* box).

In Facebook is also possible to join groups of users, this groups may be public or private, and they generally are focused on a specific matter, or gather users from one same institution or organizations (e.g. Facebook group of students of the University of Minho). Having this feature of groups, clustering users by they interests one may say that groups, some way, transform Facebook in, a "multi interest-based OSN".

---

3 Facebook wall an area where users can see the posts of their friends and/or liked pages, in a chronological order

Facebook Graph API

Facebook has today several software *kits* for developers to interact with the platform in the most diversified and imaginable ways. Facebook developers offers a range of variated software products, this vary from monetization programs, that focus on how to make users profit from Facebook, Analytics to developers who have their apps embedded in the Facebook platform understand their audience and the performance of their apps, etc. (developers.facebook.com/products/ (2016)).

In this master's thesis context, the relevant software that Facebook has available is the Facebook Graph API. This API basically allows developers to collect information from Facebook such as posts, photos, videos, pages etc. According to developers.facebook.com/docs/graph-api/common scenarios (2016), the common scenarios for using the Graph API are the following: determine whether two people are friends on Facebook; publishing new status and updates, uploading content (photos, video etc.); sharing links. But in this project what we seek is build the most biggest and detailed network as possible, with analysis and visualization purposes in mind.

For building the network fetching users friends information is crucial, this was possible until Facebook Graph API v2.0 (trough the router */me/friends*), one could actually retrive friends information and build a network from there. From v2.0 on, to achieve what was explained before, one must request a special permission called **user_friends** from each user. The permission **user_friends** is no longer included by default in every login. This change breaks down the possibility of gather Facebook information via its Graph API, this said, we need in the future to look up alternative paths to extract data from Facebook.

3.3.2 *Twitter*

3.3.3 *LinkedIn*

3.4 HOW SOCIAL NETWORKS HAVE CHANGED THE WORLD

...

SOCIAL NETWORK ANALYSIS

Social Network Analysis (SNA) is the study of how people are connected to each other, basically it studies a set of relations among a set of entities, these entities may be individuals, organizations, or even countries.

The common analysis procedure consists in mapping the network and then create metrics to characterize the network. Then one tries to figure what is the structure of the network and why does it have that structure. SNA is also about look at the individuals inside the network and where are those individuals located.

4.1 FUNDAMENTAL CONCEPTS FOR NETWORK ANALYSIS

The concepts listed below are of key importance to understand SNA.Wasserman and Faust (1994)

- *Actor* - SNA is concerned with understanding the linkages among social entities and the implications of these linkages, these social entities are described as actors. Actors are are discrete individual, corporate, or collective social units.

- *Relational Tie* - Actors are linked to one another trough *social ties*. The type of ties may be extensive, and it describes the nature of the connection. Some example of ties:

  - **Evaluation** of one person by another;

  - **Transference** of resources (bussiness transactions);

  - **Association** (to social event or cause);

  - **Behavioural** interactions (communicating);

  - **Moving** between places or statuses (migration, social or physical mobility);

  - Others may be: physical connection (roads, rivers), formal relations (authority), biological relationship;

- *Dyad* - The most basic relationship that can be established is a dyad, a connection between two actors.

- *Triad* - A relation established between three actors. Many studies included breaking SNs down to small groups (triads), this allowed a more clear conclusion about the transitivity of the connections.

- *Subgroup* - It defines any subset of actors in a SN (conceptually, subgroups come after dyads and triads).

- *Group* - A finite set of actors who for conceptual, theoretical or empirical reasons are treated as a finite set of individuals in which network measurements are made.

- *Relation* - A collection of ties of a specific kind among members of a group is called a **relation** (e.g. a connection in *LinkedIn* is a relation while evaluating our connections of sending them messages are ties).

- *SN* - With the definitions of actor, group and relation, a SN consists of a finite set or sets of actors and and the relation or relations defined on them. The presence of relation information is critical and defining feature of a SN.

4.2   NETWORK ANALYSIS

4.2.1   *Scientific Background*

Graphs

Statistics

...

4.2.2   *Power Law*

4.2.3   *Centrality Measures*

4.2.4   *Link Analysis*

4.2.5   *...*

4.3   SIX DEGREES OF SEPARATION

4.4   NETWORK VISUALISATION

4.5   REAL WORLD APPLICATIONS

5

THE PROBLEM AND ITS CHALLENGES

6

PROPOSED SOLUTION

6.1 SOLUTION REQUIREMENTS

6.1.1 *Requirements Analysis*

6.1.2 *Requirements Specification*

6.1.3 *Requirements Prioritisation*

6.2 SYSTEM MODELING

6.3 SYSTEM ARCHITECTURE

6.4 TECHNOLOGY SELECTION

6.4.1 *Technology A*

6.4.2 *Technology B*

6.4.3 *Technology C*

6.4.4 *Technology Comparison*

6.4.5 *Decision*

7

IMPLEMENTATION

7.1 DATA EXTRACTION

7.1.1 *Data Sources*

7.2 DATA MINING

7.3 BACK END

7.4 FRONT END

7.5 OUTCOMES

<div style="text-align: right; font-size: 3em;">8</div>

CASE STUDIES

Application of main result (examples and case studies)

8.1 RESULTS

8.2 DISCUSSION

8.3 SUMMARY

9

CONCLUSION

Conclusions and future work.

9.1 CONCLUSIONS

9.2 PROSPECT FOR FUTURE WORK

BIBLIOGRAPHY

John Arundel Barnes. *Class and committees in a Norwegian island parish*. Plenum New York, 1954.

Vangie Beal. Webopedia definition for social network, 2016.

couchsurfing.com. Couchsurfing about page, 2016.

developers.facebook.com/docs/graph-api/common scenarios. Facebook developer graph api, 2016.

developers.facebook.com/products/. Facebook developer products, 2016.

Cambridge Dictionary. Cambridge dictionaries online, 2002.

Nicole B Ellison et al. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 2007.

expandedramblings.com. Social media statistics, 2016.

Martin Kilduff and Wenpin Tsai. *Social networks and organizations*. Sage, 2003.

Marktest. Os portugueses e as redes sociais, 2016.

meetup.com. Meetup about page, 2016.

researchgate.net. Researchgate about page, 2016.

statista.com. statista, global social media ranking, 2016.

Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994.