Socii: A Tool to Analyze and Visualize Dynamic Social Networks*

Jorge Daniel Caldas[1], Alda Lopes Gançarski[2], and Pedro Rangel Henriques[3]

1 **Centro Algoritmi/Dpt.Informática, Universidade do Minho, Braga, Portugal**
a67691@alunos.uminho.pt
2 **Institut Télécom, Télécom SudParis, CNRS Samovar, Evry, France**
alda.gancarski@telecom-sudparis.eu
3 **Centro Algoritmi/Dpt.Informática, Universidade do Minho, Braga, Portugal**
prh@di.uminho.pt

Abstract

Social media network analysis represents a major challenge for data scientists in every aspect, since the extraction all the way to the visualization. Despite representing a major technological challenge, social media data analysis has an additional motivation, that is the daily usage in every country across the planet, making Online Social Network (OSN) a universal tool for communication, such as radio or TV, but with the technological flavor of the 21st century. In the present article, we propose a system, called Socii, for social networks analysis and visualization, as part of an ongoing work under a master's dissertation. This system overlaps two main scientific fields, sociology (more concisely social networks) and computer science. Socii aims at helping OSNs users to know and understand social structures through a user friendly interface. The system relies in four main principles, namely simplicity, accessibility, OSNs integration and contextual analysis.

1998 ACM Subject Classification D.3.4 Processors

Keywords and phrases Social Networks, Online Social Networks, Social Network Analysis, Data Analytics

Digital Object Identifier 10.4230/OASIcs.SLATE.2017.15

1 Introduction

In the mid 1950s sociologists introduced the term Social Network (SN), that despite being a familiar term for today general public because of the Online Social Networks (OSNs) platforms such as Facebook, Instagram or Twitter, it is a deeper and more mature concept. It was in the 2000s that much of the OSNs we know today start emerging, so it took at least ten years to people to adopt the concept and the new way of living, so today billions of people use these online platforms as channels for socializing, connect with each other and share their daily lives.

From the user's point of view we may consider that all the platforms offer a microscopic perspective from within the network, people have a public profile, and they can visualize their friend's profile (this is a typical scenario that we observe today in the majority of the OSNs), and normally have access to a timeline that displays friends activity. The point is

* This work has been supported by COMPETE: POCI-01-0145-FEDER-007043 and FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/ 00319/2013.

that, to the users of these online platforms, it is not provided a mean to visualize and analyze their network structure in a more abstract and generalized sense, where users are given the opportunity to observe their social network from a macroscopic perspective, and, with that, all the metrics for measuring nodes and relationships within the network.

This work is being developed under the master's dissertation which aims to prove that, a software tool may be designed and implemented, in order to actually improve the analysis of social phenomena, allowing not only sociologists, but also the public in general, to explore with greater detail the connections of individuals within a network, being OSNs the base of analysis for such a tool.

2 Origins of Social Networks (a Sociology Perspective)

Event if today people automatically think in Social Networks (SNs) as websites (or web applications), they are not aware that when talking about SNs, we refer to a much more broader term, that said, we may consider a SN 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.* [2]

The network concept is broadly used across multiple fields of study, including, physics, biology, linguistic, anthropology, mathematics, computer science and more recently computer networks, we may certainly say that *"the network concept is one of the defining paradigms of the modern era."* [8].

3 Online Social Networks

People need to connect other people, and the urge for connection brings to us what today are known as OSNs. These Web sites allow us 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 (...)* [6]

OSNs have been around for more than a decade now, but these systems have gain world wide popularity since the global adoption of platforms such as Facebook, Youtube or Twitter, which are platforms that are today massively used across all cultures and age groups, and represent a paradigm shift on social interaction that we not yet fully understand.

4 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 creating metrics to characterize the network. Then one tries to figure what is the structure of the

network and why does it have that structure. Social Network Analysis (SNA) is also about looking at the individuals inside the network and where are those individuals located.

We try to explore social network analysis in order to understand the OSNs reality, for that purposes we explored concepts such as clustering or centrality measures.

4.1 Social Network Analysis Software

In this section we talk about software tools that we analyzed with greater detail in the master's pre-dissertation [4], but that we found relevant to refer in the present article.

We reference tools such as UCINET[1] that are very mature and well known among SNA community, then we looked into more modern tools with a set of advanced features and more modern user interfaces, Gephi [1] and Social Network Visualizer[2] (SocNetV). We look also into more specific purpose tools such as Structure[3] which focus on trying to understand population flow. Investigating more deeply academic work on this field, we found Vizster [7], a visualization system for playful end-user exploration and navigation of large-scale online social networks.

5 Socii: our proposal

Given the complexity of SNA for a common user, we propose a software tool, called Socii, that allows the user easily understand and interpret his network structure. For that, not only useful metrics are computed, but also they are visualized in a convenient way, together with the network structure.

Before diving into the architectural details of the system, we will present Socii positioning. This is result of the state of the art summary, that concisely describes what is the positioning of this project in light of the existent explored SNAs tools. This comparative study can be consulted in the master's pre-dissertation document [4].

Simplicity

Aside of **Vizster** [7], the majority of the previously presented tools such as Gephi or Social Network Visualizer, are very complex tools with very heavy interfaces, that have a big learning and are meant for users that have particular advanced knowledge in SNs and SNAs. The system we are developing could also serve for less expert users, providing a set of core basic functionalities (e.g only allow users to load and visualize their networks), and then, allow the user to build complexity from there enabling and disabling other features.

Accessibility

All the software that we presented above exist in the form of desktop applications. This applications need to be downloaded, and installed in a compatible machines (sometimes with dependencies on other software that is not installed by default). Nowadays almost every application is Web based, this allows users to access them every where trough a browser, making Web apps a solution that is Operating System and device agnostic. This said,

[1] Available from `https://sites.google.com/site/ucinetsoftware/home`.
[2] Available from `http://socnetv.org/`.
[3] Available from `http://web.stanford.edu/group/pritchardlab/structure.html`.

building a Web based social networks analysis tool could be a way of tackle the accessibility of such tools.

A Web based application, it is good for sake of accessibility but in another hand it is a performance culprit when it comes to performance. This is a decision to take into account, **but always having in mind that tackling performance is not the main goal this master's thesis**, also, the mentioned tools are mature projects that are highly performant and are capable of rendering huge networks.

Online Social Network (OSN) integration

Social Network Visualizer, allows to *scrap* Web sites to build networks, but for this feature relays only on links to build the network (it blindly scraps recursively some url to build the network). By allowing the user the power to the user to analyze networks that are directly reporting their social network status would be a differentiation factor from the other tools, and would certainly be a more meaningful and valuable analysis for the end user. This feature of OSNs integration may be simulated by data generators that are specially deliver fictitious data sets within the domain of a given OSN.

Drawing Accurate Conclusions

As we state before when talking about simplicity, the mentioned SNAs tools provide generic metrics on networks such as network density or actor centrality. The values outputted from these tools are the result of running generic formulas and algorithms against some networks, so it is very common for current SNAs researchers to be worried about the size of the network, being their focus on **quantitive analysis**.

In a hypothetical analysis scenario where some researcher has a network with a few thousand nodes, **what is the meaning of his assumptions when analyzing the network**? Since this is a pure quantitive analysis the numbers will seem reasonable for the given network, but this will not allow him to extract contextual conclusions, because in this case analyzing data form Facebook or analyzing data from LinkedIn will sound just like the same, because deep down it all comes down to the network. A better approach for drawing conclusions would be to have a mixture between **quantitive analysis** and **quality analysis**, the tool could do some content and context analysis to help the end user to get a more meaningful conclusion rather than just simple metrics.

6 System Architecture Proposal and Implementation Details

In this section we present a more concrete image of the overall system. In Figure 1 we may observe an abstract diagram that represents the system architecture.

As the interaction of the software components may be understood from simply observing the diagram in Figure 1, a description of the specific role per component is needed in order to explain the system overall functionality.

Database

We use a **MongoDB**[4] (a document oriented database) to store data, this gives us more flexibility in manipulating complex JSON structures that are persisted in documents. These

[4] Available from `http://networkx.github.io/`.

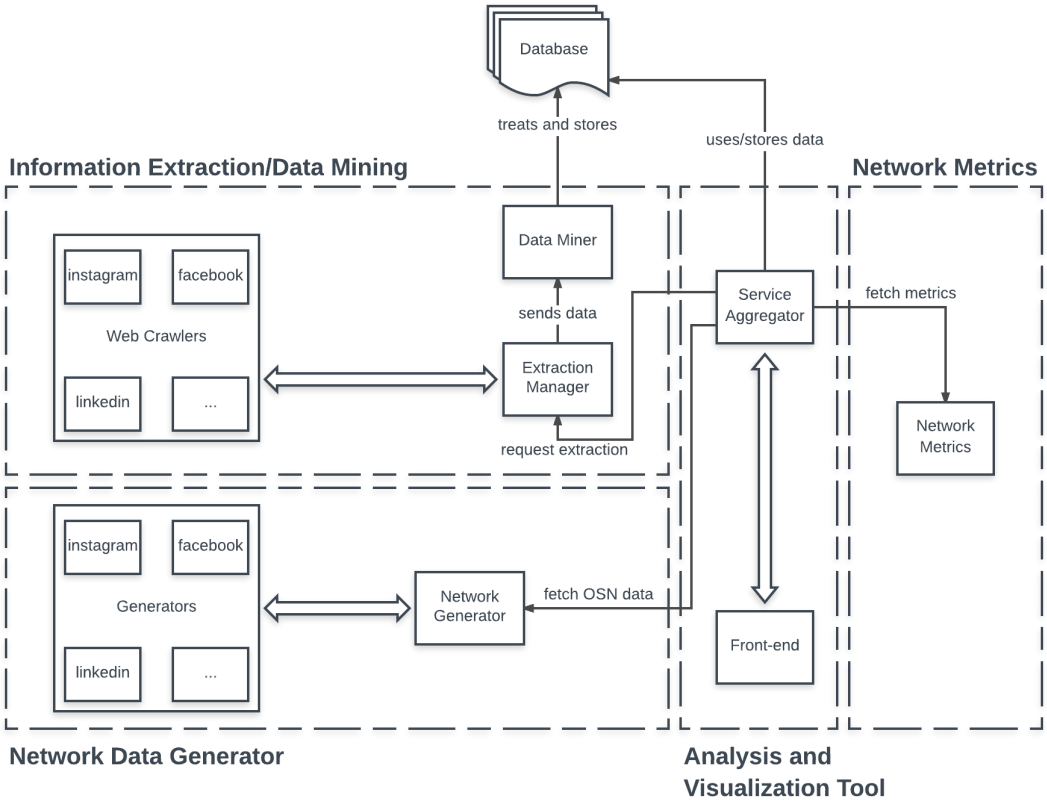

Figure 1 System architecture proposal.

flexibility and interoperability would be considerably more complex to achieve using relational databases.

Information Extraction and Data Mining

The purpose of this component is to actually extract some user network from a given OSNs. This is achieved by calling the web crawler modules (each one dedicated to a specific OSNs), that return the extracted information to the extraction manager which then sends the data through a simple data mining process in order to get data normalized before it is stored into the database. The web crawlers are implemented in Python that plays nicely along side with Selenium[5] and PhantomJS[6] to perform crawling operations using **XPath** [5] selectors to extract the information that we need to build the network.

We are aware that this kind of work is not easily achieved in a reasonable amount of time, but still it is the only way to fully access the data since majority of today OSNs have limited Application Program Interfaces (APIs) that do not allows us to extract the data to build the network. We use a **Flask**[7] APIs that executes background processes to perform parallel extractions (run web crawlers in parallel) to mitigate the slowness of web crawling, still it is a slow and not scalable solution, this is the motivation for building the network

[5] Available from `http://www.seleniumhq.org/`.
[6] Available from `http://phantomjs.org`.
[7] Avaialble from `http://flask.pocoo.org/`.

generator module that we present in a following section, where we retrieve generated mocks that are within a domain of some OSNs.

Network Data Generator

This component allows Socii aggregator to be fed with mock objects that follow pre-defined data schema specially suited for each OSNs. This is a replacement for the network extraction component, in order to validate Socii instead of having extracted data we can have several fake networks and test several different scenarios. This is also a proof of the architecture interoperability since if in the future some OSNs decide to make their APIs public we can just implement a different extraction component that instead of crawling the OSNs uses data from its APIs. This said we consider the extraction component a pluggable part of the system this architecture is designed to consume any other data retrieval component that returns the same data schema. This generators are developed with NodeJS,[8] and are available as a microservice, one just need to ask for a specific number of users for a specific OSNs.

Network Metrics

Network metrics is a microservice available trough a web API that performs calculations and runs algorithms against a given network. To consume this microservive, a client (let's say the service aggregator) only sends one HTTP request with the graph and the metrics to calculate upon the given graph. The microservice returns the computed metrics to the client. This microservice is implemented Python (again with Flask), this allows us to use the powerful library NetworkX[9] for performing graph calculations that we refer in Section 4.

Analysis and Visualization Tool

The tool that directly interacts with the end user is composed by a **Service Aggregator** that communicates with all the available services (extraction, generator and metrics api), and the front end application that provides the visualization and interaction features and it only communicates with the aggregator that it plays a facade/interface role for communication with the other services. Among the usual suspects of front end development (HTML, CSS and Javascript through the React framework,[10] in the context of Socii front end it made sense to develop a flexible and configurable graph visualization abstraction, this work originated a open source React component called **react-d3-graph**[11], which focuses on providing easy graph rendering, interaction and visualization using **D3.js** [3] for core functionalities such as rendering and node positioning calculation, it also allows the consumer to register to graph interaction events via callback functions that allowing custom application reactions upon simple graph interactions such as node and link click.

The Socii tool has a simple interface, where a user is capable of choosing between generating Facebook or LinkedIn networks, and then observe the social structure with react-d3-graph and with the metrics and other relevant data that Socii provides in a convenient way. Socii has two main areas. The configuration area where the user chooses which social network wants to generate, what metrics he wants to analyze. From here the user enters in the main area, the network vizualizer, where the user has in the center the network rendered

[8] Available from `https://nodejs.org/en/`.
[9] Available from `https://networkx.github.io/`.
[10] Available from `https://facebook.github.io/react/`
[11] Available from `https://danielcaldas.github.io/react-d3-graph/docs/`.

with react-d3-graph, on the bottom the user has some main interactions and global network metrics are displayed. When the user interacts with some node in the network, a panel is open where some metrics are shown as well the OSNs data so that the user is able to cross that data in order to derive conclusions. Some particular features were implemented for each OSNs, in Facebook we have a simple sentimenta analysis where user's posts reactions (reactions such as "*likes*") are used to determine users' sentiments in the network, for that the graph assumes a color schema that maps nodes to reactions. In LinkedIn we implemented human resources discovery where we try to find nodes with particular professional skillset and what are the best ways to approach a certain individual through interpersonal relationships within the network.

7 Conclusion

In this paper, we propose a new system, Socii, for OSNs analysis and visualization, which allows for contextual network analysis, with a relatively low complexity. Socci is available through the Web and it intends to integrate data from different OSNs. As Socii is based in OSNs platforms that are prune to change, Socii architecture was designed with interoperability in mind in order to allow different data sources to be consumed, being this operations transparent to the end user, not regarding the data that one observers but regarding the behavior and functionality of the application.

References

1 Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: an open source software for exploring and manipulating networks. In *Third International Conference on Weblogs and Social Media*, pages 361–362, 2009.

2 Vangie Beal. Definition for 'social network', 2016. Webopedia.

3 Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3 Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011.

4 Jorge Caldas. Analysis and visualization of dynamic social networks, 2017. Pre-Dissertation Report, Master Degree in Computer Science at University of Minho.

5 James Clark and Steve DeRose. *XML Path Language (XPath) Version 1.0*. World Wide Web COnsortium, 1999. URL: `https://www.w3.org/TR/xpath/`.

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

7 Jeffrey Heer and danah m. boyd. Vizster: Visualizing Online Social Networks. In *IEEE Symposium on Information Visualization*, page 5, 2005.

8 Martin Kilduff and Wenpin Tsai. *Social networks and organizations*. SAGE Publications, 2003.