

ONLINE SOCIAL NETWORK INTERNETWORKING ANALYSIS

Bassant E. Youssef ^{1,2}

- 1- Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, Virginia, USA.
- 2- Department of Engineering Mathematics and Physics, Faculty of Engineering, Alexandria University.

ABSTRACT

Online social networks (OSNs) contain data about users, their relations, interests and daily activities and the great value of this data results in ever growing popularity of OSNs. There are two types of OSNs data, semantic and topological. Both can be used to support decision making processes in many applications such as in information diffusion, viral marketing and epidemiology. Online Social network analysis (OSNA) research is used to maximize the benefits gained from OSNs' data. This paper provides a comprehensive study of OSNs and OSNA to provide analysts with the knowledge needed to analyse OSNs. OSNs' internetworking was found to increase the wealth of the analysed data by depending on more than one OSN as the source of the analysed data.

Paper proposes a generic model of OSNs' internetworking system that an analyst can rely on. Two different data sources in OSNs were identified in our efforts to provide a thorough study of OSNs, which are the OSN User data and the OSN platform data. Additionally, we propose a classification of the OSN User data according to its analysis models for different data types to shed some light into the current used OSNA methodologies. We also highlight the different metrics and parameters that analysts can use to evaluate semantic or topologic OSN user data. Further, we present a classification of the other data types and OSN platform data that can be used to compare the capabilities of different OSNs whether separate or in a OSNs' internetworking system. To increase analysts' awareness about the available tools they can use, we overview some of the currently publically available OSNs' datasets and simulation tools and identify whether they are capable of being used in semantic, topological OSNA, or both. The overview identifies that only few datasets includes both data types (semantic and topological) and there are few analysis tools that can perform analysis on both data types. Finally paper present a scenario that shows that an integration of semantic and topologic data (hybrid data) in the OSNA is beneficial.

KEYWORDS

Network Protocols, networks architecture, Social network analysis, Data visualization

1. INTRODUCTION

Online social network (OSN) is defined as a digital representation of the relations between registered entities, individuals or institutions [1] used to maintain, strengthen, and support offline social relations.

OSNs contain within them a lot of topological and semantic information. OSN analysis, OSNA, analyses semantic or topological data separately [16,17,18,19,20,21,22,23,24,25]or both data

types as in [24]. Till now no thorough study of OSNs is available to provide the basic knowledge that OSN analysts can use in future research.

Online social networks' internetworking gathers all the data from interconnected OSNs in one place." Social Networks Internetworking" allows a user with multiple accounts in different OSNs to reuse his own data, messages and photos with friends among different OSNs and allows users to interact, even if they belong to different online social networks. OSNs Internetworking removes barriers between OSNs and creates a universal social graph with all ties between users in all OSN, as seen in [79] and [30].

In this paper we propose using the concept of data virtualization where we access data via a virtualization layer. We identify OSNs' internetwork system's main components, OSNs and the virtualization layer and propose a generic architecture of each.

OSN user data and OSN platform data are the two data sources in our anatomy of OSNs. Our focus is on how OSN user data is represented and what metrics and parameters are used in its evaluation. To compare different OSNs in the internetwork, platform data are classified as resources, protocols and other non-functional aspects [31]. We present sufficient knowledge about OSN's data and new proposed OSNA methods that integrate both semantic and topologic data into the analyser. Finally, we identify which type of data (semantic, topologic or both) the publically available datasets contain and which type of data can be handled with the available simulation tools to elaborate what is available for analysts to perform SN's data analysis.

The paper is organized as follows; first it gives a brief introduction to OSNs and OSNA in section II. In section III, we discuss the different OSNA models presenting architecture of OSN, virtualization layer architecture and an OSN user experience scenario. Section IV presents online social network analysis different parameters and metrics. In section V, we illustrate the OSNs user and platform data sets analysis models with different tools used in OSN. Section VI presents proposed OSNs' user data analysis methods classification indicating the need for integrating semantic and topologic data (hybrid data) in OSN analysis methodologies. Finally section VII is the conclusion.

2. ONLINE SOCIAL NETWORKS

2.1. Definition

OSN is a platform, or a website whose main focus is building and maintaining relation or links between registered users.

Table 1 is intended to remove confusion between OSNs and social media (SM) by presenting a detailed comparison:

Table 1- Differentiation between OSNs and SM

	Social Media	Social Networking
Definition	A tool for media broadcasting, or sharing information with a broad audience.	A tool for connecting two entities for reasons as common interests, or like-minds.
Communication Style	Akin to a communication channel. A format that delivers messages and disseminates information 'to' others.	Communication is a two-way as conversations are the core of social networking through which relationships are developed.
Examples before internet	Radio broadcasting, television, newspapers, magazines, etc	Referred to the physical connections between people
Examples after internet	Encompasses several different types of media broadcasting, such as videos, blogs, etc.	Mostly refers to OSNs , OSN, or local OSN websites.
Connections	Connections between users(if any exist) are short lived	As connections are the building block of OSN, they are long lived (unless terminated by entities)
Purpose	Social media let users share content that each of them can then share with other people, for the purpose of information diffusion.	OSNs main concern are the connection between entities for several purposes such as interactions or revenue.

2.2. Perspectives for Online Social Networks

The OSNs system necessitates participation of three parties, the user, the service provider and the analyst. Each party regards the OSN biased by its specific need, usage, and benefits from the OSNs as shown in Table 2. Some aspects are common between the parties. Figure 1 illustrates the inter-relation between the three aspects.

Table 2 OSNs users- service providers and analysts aspects

OSN user aspects	OSN service provider aspects	OSN's analysts aspects
Choosing a suitable OSN type that satisfies his requirements	Increasing number of permanent users by maintaining user satisfaction	Modeling the OSN websites accurately
Maintaining his privacy	Adding new features and applications to keep user interested	Tracing the flow of information (and other resources) through OSNs
Having an easy to use attractive interface	Adding infrastructure updates for scalability and reliability guarantee	Need to define parameters that evaluate the performance of OSNs to compare them
Choosing a popular OSN	Traffic control and congestion prevention	Representing data to be processed in a formal format
Easily communicating with others	Facilitating users' communication	Understanding communication patterns
Choosing a OSN with a reliable server	Maintaining user privacy	Privacy modeling to understand it and enhance it
Being secure while using the OSN	Tracing flow of information in the network	Evaluation of the OSN's QoS
Receiving his desired QoS	Maintaining the promised QoS	
	Establish a user friendly interface	

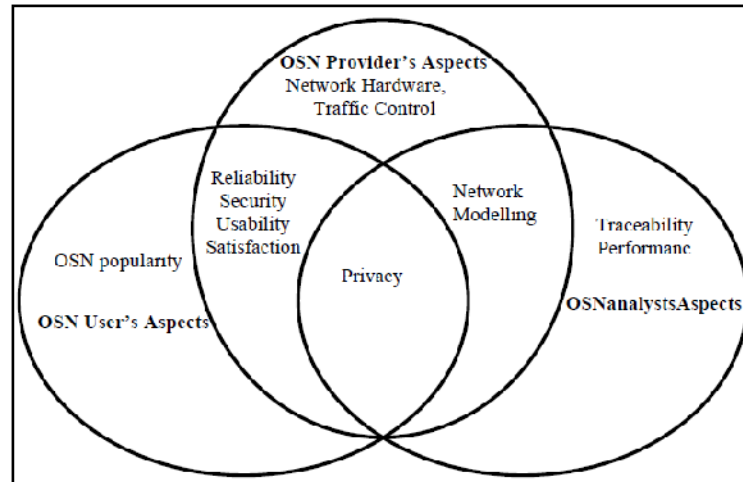


Figure 1. Venn diagram of OSNs users, service providers and analyst's aspects intersections

3.ONLINE SOCIAL NETWORK INTERNETWORKING SYSTEM ARCHITECTURE

Online Social Network Internetwork can provide the analyst with data from various OSNs. The proposed virtualization layer interconnects these OSNs. Next section provides generic architectures for the OSN internetworking system components.

3.1. Online Social Network Architecture

In this paper OSN is regarded from an analyst perspective as an architecture, OSNs' main components and their interactions with each other and with actors. OSN contains components needed to eliminate the barriers and allows the actor to mash his accounts in different OSNs via a virtualization layer so all accounts are accessible from one place. Figure 2 illustrates the proposed OSNs' architecture. The OSN platforms have four interacting different modules :

a) Management Module: This monitors other modules and sends control and timing signals to them. It consists of:

- Regulations and policies unit: Regulating actor's actions either allowed or prohibited by OSN providers. Policies refer to the agreed attitudes that actor is expected to follow. Actor is punished if a policy is broken. Universal regulations and policies are mandatory for a OSN to participate in OSNs internetworking.
- Authentication and Authorization unit: concerned with authenticating actor logging and responsible for access control so only actors specified by the user can access to his stored data.
- Timing Unit: contains the clock that assigns time for each and every action performed in the OSN.

b) Service Module: Concerned with the services performed in an OSN, It consists of two units:

- Network services: This unit handles all the services provided by the OSN service providers performable upon actor registration to the OSN without need to participate in any additional application. These include
 - Signing up in the OSN and creation of new nodes

- Linking which is the core of building an OSN as it connects actors to each other and to objects
- Creating or joining a group or an event
- Sharing (messaging , data posting(text or media))
- Tagging actors in shared data
- 3rd Party APIs: includes external developers' applications which an actor can share his data with and use.
- Virtual layer services Unit: contains new services only needed for OSNs' internetworking.

c) Storage Module: Contains all the data stored within the OSN and consists of :

- User data unit: contains all the data supplied by the actor about himself (his profile), published data (content), data about his connections (links), applications he is using
- Network data: contains data from different actors and objects interactions that constitutes OSN's social graph.

d) Interface Module: Acts as gateway facilitating seamless user access to OSN platform or all other modules. It contains:

- Mobile and wireless access unit: used by actors having smart phones and tablets to check their OSNs status.
- Compiler unit: used to translate programming languages between user actions and the OSN.
- Web browser unit: unit ensures that the user can use the browser of his preference seamlessly.
- Virtual layer interface unit: It is the gateway between the internetworked OSNs and the virtualization layer.

Each module sends interaction signals to the other modules. The management module sends control and timing signals to all the other modules and monitors their operations via receiving monitoring signals. The interface module sends access signals to the service module which represents the actor usage of the services in the service module and any output of the service module communicates with the user via the interface module. Finally the service module actions' create data that is stored in the storage module and the services use the stored data during their operation.

An actor using OSNs' internetworking should pass through the virtualization layer and then have access to different OSNs whose architecture is the same as in figure 2. The actor would experience the same generic scenario we described in figure 3. To investigate how the OSNs' internetworking operates we must define the components of the virtualization layer.

3.2. Virtualization layer Architecture

The architecture of the virtualization layer is also composed of four modules with interaction signals between the modules as in the original OSN architecture.

- a) Interface module: Responsible for data transfer between actors and different OSNs. It contains a virtual switch unit, web browser unit, compiler unit, wireless and mobile access unit.
- b) Service Module: Consists of legacy services and virtual layer services units including caching service.
- c) Management Module: Manages all the security tasks such as actor authentication, authorizing his access only to the resources which he has the right to access and encrypting his data.

- d) Data Module: Heart of the virtualization layer operation consisting of Data extraction, Data cleansing unit, Data Abstraction / Transformation unit, Data federation unit, and Cache memory.

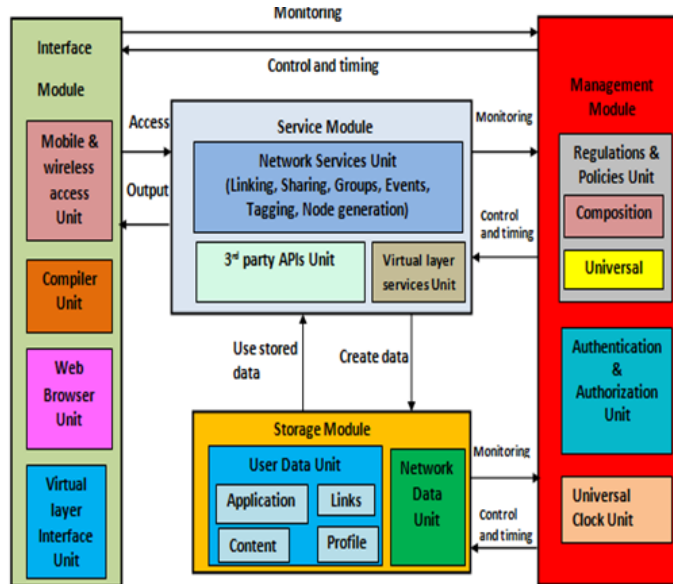


Fig 2- Architecture of OSN

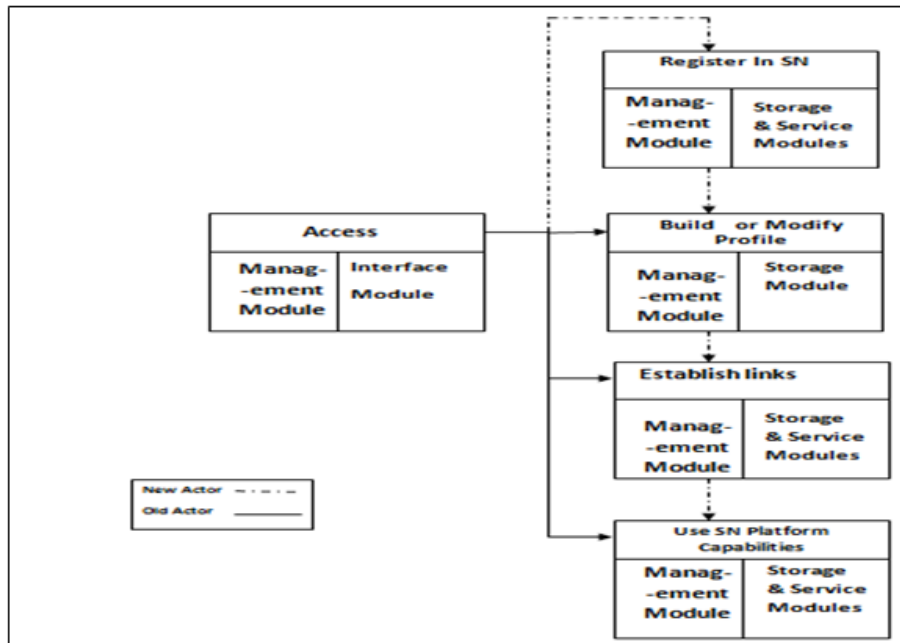


Fig 3- OSN User experience Scenario

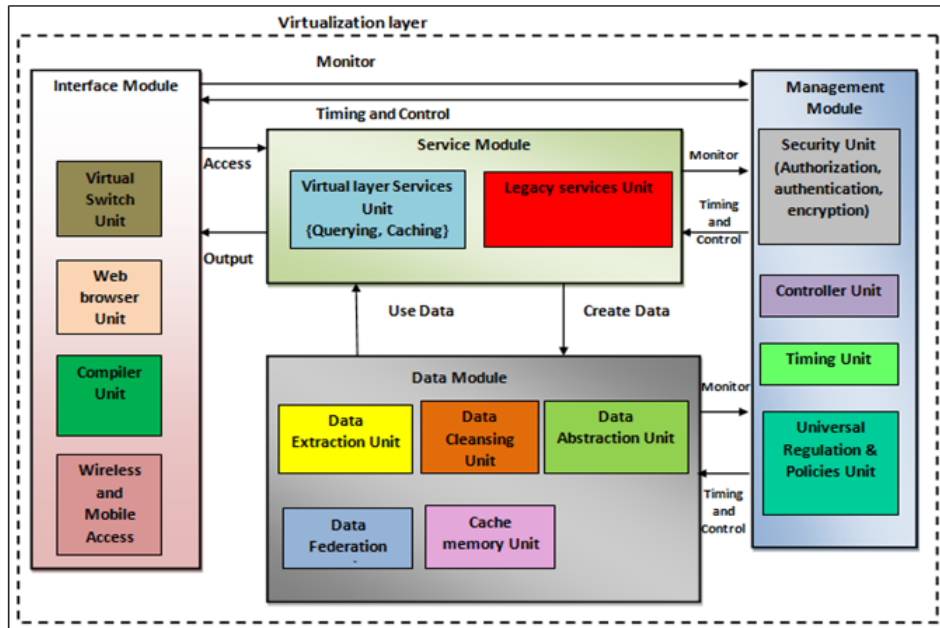


Fig.4Virtualization layer architecture

4.ONLINE SOCIAL NETWORKS ANALYSIS

OSNA represents an effective method to study rumour spreading, diffusion of innovations as well as threat detection or surveillance. OSNA deals with OSN links, OSN entities, and the platform provided by the OSN itself and is defined as the “mapping and measuring of relationships (or flows) between different entities whose relations are represented as a network”.

4.1. OSNA process

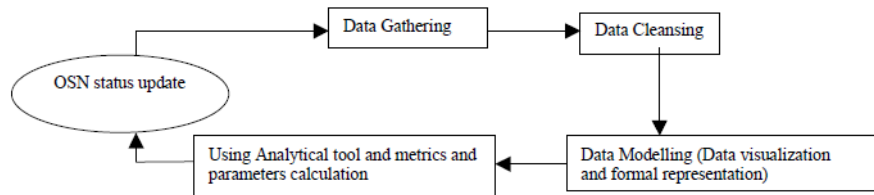


Figure 5 OSNA process

As seen in Figure 5 OSNA process is comprised of:

- 1) **Data gathering:** Gathered from the analyzed OSN using surveys, actual data aggregation, or using publically available dataset.
- 2) **Data visualization and formal representation:** It represents the data mapping and OSN visualization as a graph with nodes and edges that can be manipulated by mathematical methods with data re-represented in some formal formats, resulting in the following benefits
 - o Represent the data compactly and systematically.

- Allow the application of computers to data analysis which makes the analysis more scalable and accurate.
 - Can lead analysts to discover and see patterns in the data that might not be clear with word description.
- 3) **Analysis methodologies** choose and apply the analysis methodology suitable to chosen formal data representation.
 - 4) **Analytic tool** choose the suitable analytic or simulation tools to solve the problem using chosen analysis methodology.
 - 5) **Evaluation Parameters** choose parameters that should be evaluated.

4.2. Online Social networks User Data and Platform Data

There are two main sources of data in the OSN, user and platform. The OSN user data includes user’s semantic data, his profile or characteristics, behaviour, and content and his topological data, describing his graph interconnections. The other source of data is the OSN’s platform. An actor performance on OSNs is affected by what the OSNs themselves offer. The OSN platform data is represented by its resources, the applied protocols and its non- functional aspects. Hybrid data in OSNA is the data resulting from the coupling or integration of semantic and topological data. Hybrid data analysis can be: Atomistic Actor Analysis, Holistic Actor Analysis, Atomistic Content Analysis, Holistic Content Analysis, Atomistic Actor with Atomistic Content Analysis or Holistic Actor with Holistic Content Analysis are chosen according to the chosen type of hybrid data [6,7]. New parameters may arise due to the semantic and topologic data integration.

5.ONLINE SOCIAL NETWORKS USER DATA ANALYSIS MODELS

5.1. OSNA types

Hybrid data in different OSNs is the coupling or integration of semantic and topological data as shown in following Table 3.

Table 3 Hybrid data sets

OSNs	Topology	Semantics
Bernard & killworth fraternity [8]	Yes	No
kapferer tailor shop [8]	Yes	No
twitter word usage [9]	No	Yes
Facebook Social Graph [10]	Yes	Yes
Facebook Applications [10]	yes	Some(User’s used applications)
Network 1: Facebook-like Social Network [11]	Yes	No
Network 2: Facebook-like Forum Network [11]	Yes	Yes
Freeman network analysis. Evolving social relations. [11]	Yes	No
Newman’s scientific co authorship [11]	Yes	No
Collaboration network in science of networks[12]	Yes	No
Wordnet [12]	No	Yes
BLOGS06 [13]	Yes	Yes
YouTube Entertainment Category [14]	No	Yes
soc-Epinions1 [15]	Yes	No
soc-LiveJournal1 [15]	Yes	No
soc-Slashdot0811 [15]	Yes	No
wiki-Vote [15]	Yes	Yes
email-Enron [15]	Yes	Yes
Political blogs [16]	No	Yes

OSNA can use stochastic or deterministic models in the analysis of topological or semantic data. Deterministic models' outcomes are precisely determined a priori through the mathematical relationships among systems 'states and events. It produces the same exact output for the given set of initial conditions. In deterministic analysis the observed values are regarded as the population of interest.

Conversely, in stochastic models, randomness is present, and variables are defined using ranges of values or probability distributions [19,20,22]. Analysts seek the probability distributions for the evaluated quantities and observed values are samples of the population of possible observations. Time can be one of the observed quantities in OSNA if networks' interactions change over time at high rate, which is called (dynamic online social networks) [17].

5.2. Data Models

After identifying the type of User OSN data to be analysed and OSNA types, the next step is data modelling. As mentioned, 'to analyse data, it must be modelled in a proper analysable form '. OSNs data can be modelled as: ontologies, graphs, or sets.

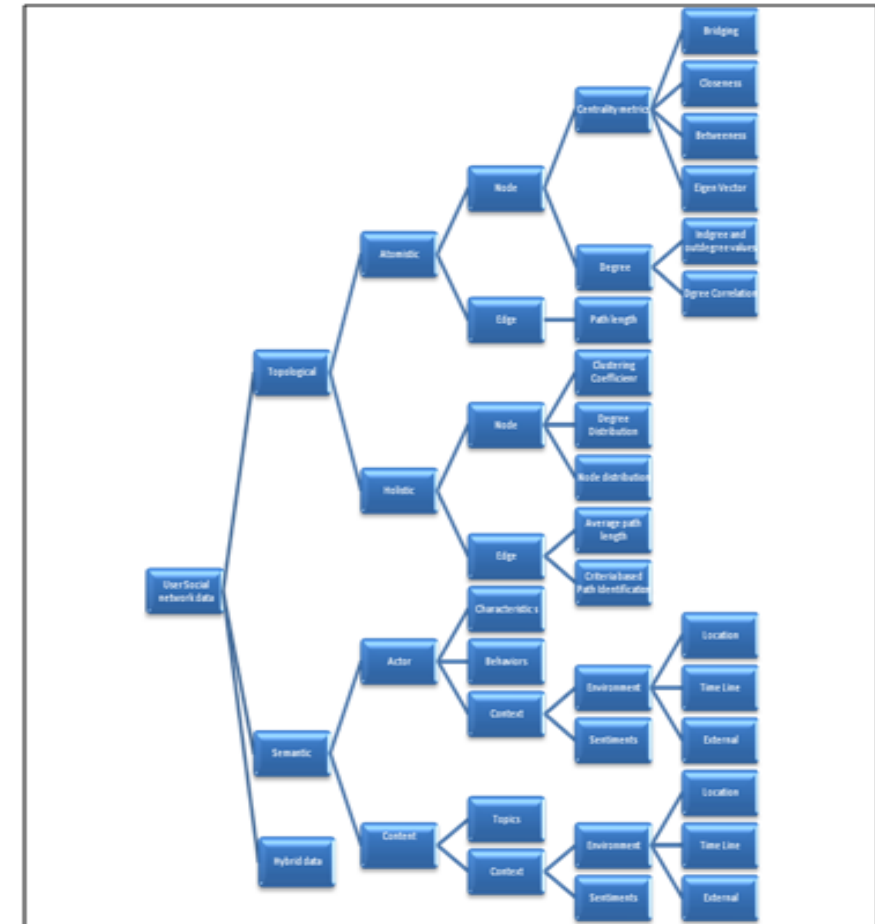


Fig 6- OSN User data



Fig 7- OSNs' user data Analysis methods classification

6. PROPOSED OSNs' USER DATA ANALYSIS METHODS CLASSIFICATION

As seen in figure 6, we start with classifying the user OSN data according to its type whether topological, semantic or hybrid. Each of them can be then classified according to the analysis model. We first choose the analysis type (stochastic or deterministic). We see from Figure 7 that stochastic data can be modelled as graphs or sets. Deterministic data can be modelled as graphs [2,17,27], sets [5,23] or ontologies [2,3,21,26,28]. Graph analysis methodology can be simple graph analysis or matrix manipulation. While sets use algebraic analysis methods, ontologies use ontology analysis. Choosing the right form of data representation to be used in OSNA, depends on the benefits and limitations that each representation provides and also on the available data that can be used to perform the analysis.

6.1. OSN Platform Data

The Social network platform data is classified according to its resources, the applied protocols and its non-functional aspects as shown in figure 8. Resources of an OSN include the services provided which differ from one OSN to another and the network's infrastructure. Resources also

include the network's infrastructure. This includes the physical description of the network including its capacity, or bandwidth allocated for users. Applied protocols include access control, regulations and policies and privacy settings. Protocols also include the different privacy settings provided for the users. Non-functional aspects affect the platform or capabilities provided to the users such as Usability, Security, and Performance measures of the OSN.



Fig 8- Online Social network platform data

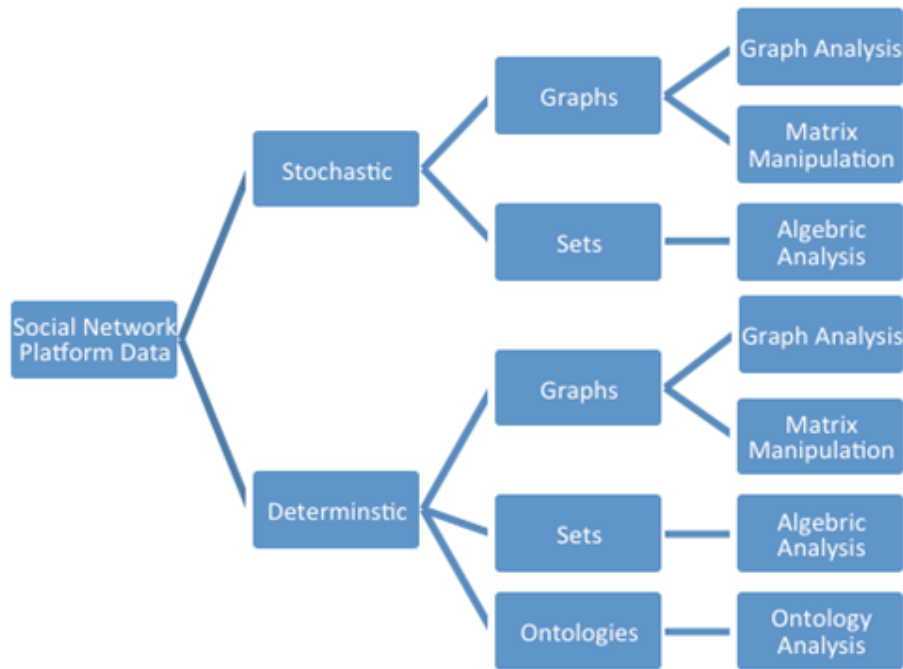


Fig 9- Online Social network platform analysis models

6.2. OSN Platform Data Analysis Models

As Figure 9 illustrates, the analysis of OSN platform data that follows the same analysis model as OSN user data's analysis. We start by determining the type of analysis to be performed which can be stochastic analysis or deterministic analysis. The next step is formal representation of the analysed data or data modelling. Data models can be graphs and sets for stochastic analysis type and they can be graphs, sets or ontologies for deterministic analysis type. The analysis methodology is dependent on the data mode.

7. CONCLUSIONS

OSNs are a valuable sources of information due to their wide spread and publicity. Many decision makers make their decisions according to OSNs' data. This paper started with a comprehensive anatomy of OSNs and OSNA which is expected to aid analysts' future efforts. We used the concept of internetworking of OSNs to further increase the data gained from OSNs where we had to define architectures for OSNs and the virtualization layer. We were concerned with the analyst perspective of OSNs to guide him for whom we differentiated between the sources of OSNs' data, the user or the OSN platform. We proposed a classification of OSN user data according to analysis models it uses for both data types and this classification will be beneficial for analysts as it specifies current OSNA methodologies. We also introduced a classification of User OSN data according to its evaluation parameters and metrics. We classified OSN platform data according to its data source (resources, protocols and non-functional aspects) and highlighted the fact that it can be used to differentiate different OSNs in the internetworking system. These classifications are beneficial to analysts and help them to know the limitations and capabilities of current OSNA methods to aid them in proposing new OSNA methodologies. An overview of currently publically available datasets and OSNA simulation tools was given.

From this overview, we notice that only few datasets includes both data types (semantic and topological) and there are few analysis tools that can perform analysis on both data types. We finally presented a scenario that showed that analysis depending on only one data type (either topological or semantic) may lead to inaccurate decisions and that integrating both data types or hybrid data analysis can create new information that won't exist with individual data type analysis.

We will be working on building a distributed analysis system that provides a systematic methodology for co-analysing and co-modelling topological and semantic data gained from different OSNs. Our system would be able to benefit from all identified OSNs' data sources.

REFERENCES

- [1] L.A. Cuttillo, R. Molva, T. Strufe, "Privacy Preserving Social Networking Through Decentralization", Proceedings of the Sixth international conference on Wireless On-Demand Network Systems and Services, 2009.
- [2] Malgorzata Marciniak and Agnieszka Mykowiecka. "Aspects of Natural Language Processing"
- [3] Ereto, G., et al., Semantic Social Network Analysis, a concrete case, in Handbook of Research on Methods and Techniques for Studying Virtual Communities: Paradigms and Phenomena. 2011, IGI Global: Hershey, PA.
- [4] Mohsen Jamali and Hassan Abolhassani. 2006. Different Aspects of Social Network Analysis. In Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI '06). IEEE Computer Society, Washington, DC, USA, 66-72.
- [5] http://www.cs.umbc.edu/help/theory/group_def.shtml
- [6] Nanda, S.; Kotz, D.; , "Localized Bridging Centrality for Distributed Network Analysis," Computer Communications and Networks, 2008. ICCCN '08. Proceedings of 17th International Conference on , vol., no., pp.1-6, 3-7 Aug. 2008
- [7] Yong-Yeol Ahn, Seungyeop Han, Haewoon Kwak, Sue Moon, and Hawoong Jeong. 2007. Analysis of topological characteristics of huge online social networking services. In Proceedings of the 16th international conference on World Wide Web (WWW '07). ACM, New York, NY, USA, 835-844.
- [8] <http://vlado.fmf.uni-lj.si/pub/networks/data/ucinet/ucidata.htm#bkfrat>
- [9] http://www.infochimps.com/datasets/twitter-word-usage#overview_tab
- [10] http://odysseas.calit2.uci.edu/doku.php/public:online_social_networks#available_datasets
- [11] http://toreopsahl.com/datasets/#online_social_network
- [12] <http://vlado.fmf.uni-lj.si/pub/networks/data/collab/netscience.htm>
- [13] http://ir.dcs.gla.ac.uk/test_collections/blog06info.html
- [14] <http://an.kaist.ac.kr/traces/IMC2007.html>
- [15] <http://OSNap.stanford.edu/data/>
- [16] L. A. Adamic and N. Glance, "The political blogosphere and the 2004 US Election", in Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem (2005).
- [17] John Tang, Mirco Musolesi, Cecilia Mascolo, Vito Latora, and Vincenzo Nicosia. 2010. Analysing information flows and key mediators through temporal centrality metrics. In Proceedings of the 3rd Workshop on Social Network Systems (OSNS '10). ACM, New York, NY, USA, , Article 3 , 6 pages.
- [18] J. Pfeiffer III and J. Neville. Probabilistic Paths and Centrality in Time. In Proceedings of the 4th OSNA-KDD Workshop, KDD, 2010
- [19] Ding Zhou, Eren Manavoglu, Jia Li, C. Lee Giles, and Hongyuan Zha. 2006. Probabilistic models for discovering e-communities. In Proceedings of the 15th international conference on World Wide Web (WWW '06). ACM, New York, NY, USA, 173-182.
- [20] Menezes, T.; Roth, C.; Cointet, J.-P.; , "Precursors and Laggards: An Analysis of Semantic Temporal Relationships on a Blog Network," Social Computing (SocialCom), 2010 IEEE Second International Conference on , vol., no., pp.120-127, 20-22 Aug. 2010
- [21] P. Bhattacharyya, A. Garg, and S. Wu. Analysis of user keyword similarity in online social networks. Social Network Analysis and Mining, pages 1-16, 2010. ISOSN 1869-5450. URL <http://dx.doi.org/10.1007/s13278-010-0006-4>.
- [22] Ville H. Tuulos and Henry Tirri. 2004. Combining Topic Models and Social Networks for Chat Data Mining. In Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web

- Intelligence(WI '04). IEEE Computer Society, Washington, DC, USA, 206-213. DOI=10.1109/WI.2004.44
- [23] Mohsen Jamali, GholamrezaHaffari, and Martin Ester. 2011. Modeling the temporal dynamics of social rating networks using bidirectional effects of social relations and rating patterns. InProceedings of the 20th international conference on World wide web (WWW '11). ACM, New York, NY, USA, 527-536
- [24] Cointet, J.-P.; Roth, C.; , "Socio-semantic Dynamics in a Blog Network," Computational Science and Engineering, 2009. CSE '09. International Conference on , vol.4, no., pp.114-121, 29-31 Aug. 2009
- [25] FabrícioBenevenuto, Tiago Rodrigues, Meeyoung Cha, and Virgílio Almeida. 2009. Characterizing user behavior in online social networks. In Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference (IMC '09). ACM, New York, NY, USA, 49-62.
- [26] <http://www.fmsasg.com/>
- [27] Zhanikeev, M.; Tanaka, Y.; , "Application of graph theory to clustering in delay space," Information and Telecommunication Technologies (APSITT), 2010 8th Asia-Pacific Symposium on , vol., no., pp.1-6, 15-18 June 2010
- [28] Chunying Zhou.;Huajun Chen.; Tong Yu.; , "Social network mashup: Ontology-based social network integration for statistic learning," Information Reuse and Integration, 2008. IRI 2008. IEEE International Conference on , vol., no., pp.143-146, 13-15 July 2008
- [29] Meo, Pasquale De; Quattrone, Giovanni; Rosaci, Domenico; Ursino, Domenico; , "Dependable Recommendations in Social Internetworking," Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT '09. IEEE/WIC/ACM International Joint Conferences on , vol.2, no., pp.519-522, 15-18 Sept. 2009 doi: 10.1109/WI-IAT.2009.205
- [30] Barbara Carminati, Elena Ferrari, and Andrea Perego. 2009. Enforcing access control in Web-based social networks. ACM Trans. Inf. Syst. Secur. 13, 1, Article 6 (November 2009), 38 pages. DOI=10.1145/1609956.1609962 <http://doi.acm.org/10.1145/1609956.1609962>
- [31] Chelmiss, Charalampos; Prasanna, Viktor K.; , "Social Networking Analysis: A State of the Art and the Effect of Semantics," Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on and 2011 IEEE Third International Confernece on Social Computing (SocialCom) , vol., no., pp.531-536, 9-11 Oct. 2011.