A Definitions' Framework for Personal/Egocentric Online Social Networks

Sarra DJEMILI [1], **Claudia MARINICA** [1], **Maria MALEK** [2], **Dimitris KOTZINOS** [1]

1. *ETIS Lab - ENSEA / UCP / CNRS UMR 8051, Cergy-Pontoise, France*
   {sarra.djemili,claudia.marinica}@ensea.fr,dimitrios.kotzinos@u-cergy.fr
2. *QUARTZ Lab - EISTI, Cergy-Pontoise, France*
   mma@eisti.eu

ABSTRACT. *People participate in multiple online social networks, where they create connections by interacting with other users directly or indirectly. These connections constitute what we call a personal or egocentric network. These online personal networks (OPNs) are quite different than the usual social personal networks, since they do not require a personal acquaintance in order to be established and they might serve multiple purposes (e.g. allow or not unidirectional communication). Thus understanding and defining who participates in a user's personal network is a work that needs to be revisited since the underlying conditions have changed. Thus, here, we provide a set of formal definitions for OPNs that capture the realities of online social life and we implement those in an extensible software framework that can be used to extract personal networks from any online social network and can be extended to facilitate the researchers studying the OPNs to easily grasp the evolution and the flow of information within OPNs.*

RÉSUMÉ. *Les utilisateurs participent à de multiples réseaux sociaux en ligne, où des connexions sont crées par l'interaction avec d'autres utilisateurs de manière directe ou indirecte. Ces connexions constituent ce qu'on appelle un réseau personnel ou égocentrique. Ces réseaux personnels en ligne (OPN) sont assez différents des réseaux sociaux personnels habituels, pusiqu'ils ne nécessitent pas un contact direct afin d'être mis en place et peuvent servir à des fins multiples (par exemple permettre ou pas une communication unidirectionnelle). Ainsi, comprendre et définir qui participe au réseau personnel d'un utilisateur est un travail qui doit être revu car les conditions sous-jacentes ont changées. Ainsi, nous proposons un ensemble de nouvelles définitions formelles pour les réseaux personnels en ligne qui englobent les réalités de la vie sociale en ligne et une implémentation de ces définitions dans le cadre d'un logiciel extensible afin d'être utilisé pour extraire des réseaux personnels à partir de n'importe quel réseau social en ligne. Ce logiciel peut être étendu pour faciliter à l'avenir les recherches sur les OPNs afin de comprendre facilement leur évolution et la diffusion de l'information au sein des OPNs.*

KEYWORDS: *online personal networks, online social networks, personal network extraction*

MOTS-CLÉS : *réseaux personnels en ligne, réseaux sociaux, extraction des réseaux personnels*

1. Introduction

Nowadays, people participate in many and diverse Online Social Networks (OSNs). Participation in multiple OSNs becomes a necessity, since each one serves a different objective: e.g. Facebook is used to share private information with friends, LinkedIn is used to build a professional network and Twitter is used to have access to latest news/information and to share short messages. Inside each of these networks, the user has his/her own personal connections which we call a personal (or ego) network, i.e. a network that is composed of the user as its focal point (or ego) and of these actors that the ego is interacting with directly or indirectly. Actually it is the explosion of online social networks which made easy the communication with people who are beyond our immediate social circles, rendering at the same time obsolete the traditional definition of social networks for the online world.

In fact, studying the activities in one or more online personal networks (OPNs) mainly means studying the exchange of information between the ego and other users he/she is connected to and also studying the evolution of the network itself since OSNs (and subsequently OPNs) change over time both in terms of structure and information flow (severity/weight of exchanges). But since interactions are now possible between the ego and users beyond his/her immediate social circle, we need to accordingly expand the notion of the egocentric/personal online social networks. Moreover, the fact that an ego has different ego-networks with different characteristics (e.g. in Facebook the people/nodes that can be contacted by the user are the same as the people/nodes that can contact the user, while on Twitter there is a clear distinction between "followers" and "followees") also imposes an extension to the definition of OPNs. Thus, the roles of creator of information and recipient of information can be either independent or mixed. So, we need to model the incoming information to a node from different sources (e.g. different OPNs) and thus study the effect that the possibly diverse and conflicting information will have on the user. In that respect, we need to expand the available definitions to capture cases and create models where the user will be the recipient of information maybe as part of different personal networks.

In the existing literature we cannot find adequate definitions (section 2) that capture the diversity of today's OPNs. So we provide formal definition for OPNs by studying their characteristics (section 3) and then we implement these definitions in a software framework (section 4) that allows the extraction of any (type of) OPN within a given OSN and provides a set of metrics for the OPN. We also describe part of the experiments we run using the software framework that show that we can successfully extract different types of OPNs from a large data set and present it to the user in versatile ways and we conclude (section 5) offering a summary of the work described in the paper and some pointers to extend the current work in the near future.

2. Related Work

In the literature, the problem of describing OSNs has been widely addressed (Ahn *et al*., 2007), (Ugander *et al*., 2011), but the study of personal/ego-networks has re-

ceived less attention so far. On the contrary, there is a significant literature coming from the social sciences studying offline egocentric social networks, e.g. (Dunbar, Spoors, 1995). Unfortunately, the definitions of offline (or societal) egocentric social networks are not suitable in online settings. In the rest of the section, we present the available ego-networks' definitions and evaluate their suitability for OPNs' needs.

2.1. Online Social Networks

Before going into OPNs' definitions, let us provide the definition of an OSN.

Definition 1. *An Online Social Network is a graph $G(V, E)$ where V is the set of nodes representing the social actors and E is the set of edges representing the links between them.*

Example 1. *Figure 1-a represents an OSN (our running example in the paper) described by the graph $G(V, E)$ where: $V = \{e, a1, a2 \ldots, a15\}$ and E is composed of all the links between nodes in V.*

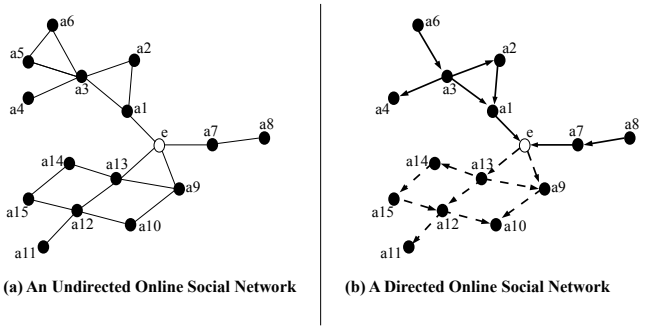

Figure 1. An Online Social Network.

2.2. Egocentric/Personal Online Social Networks

In OSNs, an egocentric or personal (we will use these terms interchangeably throughout the paper) network EN^e, organized around an ego node e is a sub-network of an OSN $G(V, E)$ and it is described by a graph $G'(V', E') \subset G(V, E)$ where V' is the set of nodes (containing the ego e and the $alters$, i.e. all the other nodes) and E' is the set of links of the ego-network. In the literature, two cases of online personal networks are identified depending on the distance of the alters from the ego: 1-level and k-level.

2.2.1. 1-level Ego-Networks

Studies in social sciences consider that an offline ego-network is composed only of the ego and the alters directly connected to it (Dunbar, Spoors, 1995), (Hill, Dunbar, 2003), (Roberts *et al.*, 2009); thus we have 1-level connections and 1-level ego-networks. This definition inspired many researchers in OSNs (Conti *et al.*, 2011),

(Arnaboldi *et al*., 2011), (Arnaboldi *et al*., 2012), (Arnaboldi *et al*., 2013) who provided a series of definitions of 1-level online ego-networks, differentiated by the links that they are considering. The set of nodes remains the same across these definitions and can be formulated for a 1-level ego-network EN^e of an ego node e as: $V' = \{x \in V \mid (e,x) \in E\} \cup \{e\}$ composed of the focal individual e and the individuals that are directly connected to it. In the OSN in Figure 1-a, the node set of the 1-level ego-network of e is $V' = \{e, a1, a7, a9, a13\}$. The set E' of edges, will be defined hereafter for each different case.

The work in (Arnaboldi *et al*., 2012) focuses on 1-level ego-networks that do not take into account the alter-alter links. If we consider the ego-network EN^e of the ego node e, the edges set is composed of the connections between e and all the alters: $E' = \{(e,x) \in E \mid x \in V' \setminus \{e\}\}$. From the OSN in Figure 1-a, the link set of the 1-level ego-network of e is $E' = \{(e,a1),(e,a7),(e,a9),(e,a13)\}$. Investigating the similarity of structures of online and offline social networks, different studies have used the above definition, including one concerning Facebook (Arnaboldi *et al*., 2011). (Tchuente, 2013) proposes to take into account for 1-level ego-networks the alter-alter links and not the ego-alters links. Thus, the set of edges of an ego-network EN^e of the ego node e is composed of the connections only between its alters: $E' = \{(x,y) \in E \mid x \in V' \setminus \{e\} \wedge y \in V' \setminus \{e\}\}$. From the OSN in Figure 1-a, the link set the 1-level ego-network of e is $E' = \{(a9,a13)\}$. They justify this under the hypothesis that the relations between the ego and its alters were not useful as the authors were interested in detecting communities inside the ego-network of a user. The previous definitions were describing an ego-network where the set of edges was composed *only* of ego-alter links *or* alter-alter links. Other studies, e.g. (Yen, 2014) and (Quercia *et al*., 2012), consider *both* types of links as part of the ego-network. Then the set of edges of an ego-network EN^e of the ego node e is: $E' = \{(x,y) \in E \mid x \in V' \wedge y \in V'\}$. From the OSN in Figure 1-a, the link set of the 1-level ego-network of e is $E' = \{(e,a1),(e,a7),(e,a9),(e,a13),(a9,a13)\}$, the ego-alter links $(e,a1),(e,a7),(e,9),(e,a13)$ are added to the alter-alter link $(a9,a13)$. This is justified because they wanted to do quantitative analysis on OSN data to measure network metrics such as $network\ constraint$.

In this section we presented three definitions of 1-level ego-networks that differ by the considered links, each one used for different aims. But these definitions are inadequate for OSNs, which go beyond the 1st level and where users might interact with users not directly related to them (e.g. on Twitter a retweet might reach someone not directly connected to the user). Thus revisiting these definitions is necessary.

2.2.2. k-level Ego-Networks

In addition to 1-level ego-networks, recent works, such as (Tchuente, 2013) and (Gatti *et al*., 2013) propose to define ego-networks by including potential alters which are at a maximum distance k from the ego (what is called a k-ego network). The k-ego-network of an ego e is the network of the individual e consisting of (1) e and the individuals situated at a maximum distance k of e (namely V'), and (2) all the edges

between the two individuals (alters) in the k-ego-network except those with e (namely E') (Tchuente, 2013): $V' = \{x \in V \mid d_G(e,x) \leq k\} \cup \{e\}$,
$E' = \{(x,y) \in E \mid x \in V' \setminus \{e\} \wedge y \in V' \setminus \{e\}\}$

If we consider the running example in Figure 1-a, k is equal to 3 (maximum distance between e and the other nodes is 3). The node set will contain all nodes ($V' = V$) and the link set will cover all the links except those connecting e ($E' = E \setminus \{(e,a1),(e,a7),(e,a9)(e,a13)\}$). (Gatti *et al.*, 2013) also considered k-ego-networks but for Twitter's following graph. The structure adopted could be seen as a particular case of a previous definition since it includes the links with the ego and omits those between alters situated at the ego-network last level. Based on the running example in Figure 1-a, the link $(a5,a6)$ will be omitted. Despite extending the definitions from 1- to k-level ego-networks, omitting connections or not specifying their direction renders them incomplete and limits their applicability only in specific cases.

2.3. Discussion

In this section, definitions for ego-networks found in the literature were presented so as to assess their suitability to represent OPNs. Several limitations of these definitions were identified:

– 1-level ego-networks' models are inadequate (Quercia *et al.*, 2012) since the existing definitions are focusing on undirected networks while current OSNs are usually better represented with directed graphs (allowing for unilateral communications);

– In OSNs users can easily interact beyond their immediate social circle (beyond level 1). This changes extensively the notion of the personal network (considering it as a network of interacting actors) and this is not covered in the presented definitions;

– The intensity of a relationship/link between persons/nodes is important since it allows us to differentiate the connections; this is what we usually capture using the tie strength (of the link). The tie strength is prominent in studying social influence between individuals (Aral, Walker, 2014), (Bakshy *et al.*, 2012), but also it affects the information diffusion process as demonstrated in (Ferrara *et al.*, 2012). The inability to express tie strength, limits our ability to properly describe the OPNs.

In the next section, we propose formal definitions for different types of OPNs, which carry characteristics that have been insufficiently or not at all addressed above. Our contribution is to provide inclusive and usable definitions that suit different needs and to the best of our knowledge this is the first work, which is devoted to systematically define a formal framework for OPNs.

3. Definitions of Online Personal/Egocentric Social Networks

Generic OSNs are represented by graphs as given in Definition 1. In this section, we are providing a set of definitions that are flexible enough to cover current OPNs and be extensible for the future. We are presenting three definitions: the first one

focuses on undirected OPNs, the second definition is for directed OPNs and the third one extends the first and the second and defines weighted OPNs where the links are enriched by a value representing a given property.

3.1. Undirected Ego-Networks

Definition 2. *We define an* undirected ego-network EN^e *of an ego individual* e *as being composed of (1) the ego and the individuals who are connected to it directly or indirectly (that we call alters), and (2) of all the connections between* e *and his alters, but also between the alters.*

$$EN^e = G'(V', E') \subset G(V, E)$$

$$V' = \{x \in V \mid d_G(e, x) \leq k\} \cup \{e\}$$

$$E' = \{\{x, y\} \in E \mid x \in V' \wedge y \in V'\}$$

where V' *represents the set of nodes including the ego node* e *and all nodes (1 or more) that are connected to* e *via a shortest path of maximum length* k *given by the condition* $d_G(e, x) \leq k$*, where* d_G *is the number of edges contained in the shortest path from* e *to a node* x*, and* E' *holds the set of* ***all possible*** *edges linking* V'*'s nodes.*

It is important to note that we use here the function d_G in order to restrict the selection of set of nodes of the personal network of an ego e; nevertheless, once the set of nodes V' selected, we will consider in the personal network all the existing edges between the selected nodes. This means that the resulting network might contain nodes that are connected to the ego e via a path longer than k (but via a shortest path shorter or equal to k). The function d_G will have the same role in the following definitions.

Definition 2 is close to the one given in (Gatti *et al*., 2013) for Twitter without the restriction of excluding links between last-level alters. Compared to the definition given by (Tchuente, 2013) presented in section 2.2.2, our new definition considers in addition to alter-alter links, the links between the ego node and its direct alters as part of the ego-network. This is important as we aim to study information flow between users in the OPN, then we need to have the ego-alter connections. Thus, our definition expands the notion of a personal network allowing to better represent diverse current OPNs unlike (Gatti *et al*., 2013), where the definition was solely for Twitter.

Example 2. *Based on our new definition, the OSN in Figure 1-a is actually an undirected ego-network. We note that here* $k = 3$ *because the ego* e *is situated at maximum a distance of* 3 *from the farthest alter. This ego-network consists of: the node set* $V' = V$*, and the link set* $E' = E$*.*

3.2. Directed Ego-Networks

The previous definition is dedicated to undirected ego-networks allowing to describe their nodes and connections. In this case, the connections are symmetric (i.e.

in Facebook if you have a friend in your list, this person will have you in his/her list), but if we take Twitter as example, the connections are not following anymore a symmetric model as you can follow someone without him/her following you back, so it is a directed relationship and thus, we provide a definition based on directed graphs. To this end, we need to define two distinct concepts: incoming and outgoing OPNs (and of course their union that gives the full directed ego-network).

Definition 3. *An incoming ego-network EN^e_{in} of an individual ego e is a directed sub-graph $G'(V', E')$ of the directed graph G composed of the set of individuals x that are connected to e via a shortest path of maximum length k, where x is the start node and e the end node, and the set of links composed of the links being part of the incoming paths to e:*

$$EN^e_{in} = G'(V', E') \subset G(V, E)$$
$$V' = \{x \in V \mid d_G(x, e) \leq k\} \cup \{e\}$$
$$E' = \{(x, y) \in E \mid x \in V' \setminus \{e\} \wedge y \in V'\}$$

where V' is composed of e and all the nodes (1 or more) with the condition that for each node, there exists an incoming (i.e. a path for which the end node is the ego) shortest path of maximum length k from the given node to e, and E' is the set of directed edges linking nodes in V'.

Through the definition of an incoming ego-network, we can model the flow of the information *explicitly received* by a particular user. For example, we would be interested in investigating whether this particular user is getting information from conflicting sources or if the obtained information affects its participation in various communities or even we could determine the path followed by the information. It is important to note that this type of analysis was not possible with the definitions presented in Section 2 due to the lack of definition of directed ego-networks.

Example 3. *Based on the graph in Figure 1-b, the directed sub-graph with nodes connected by continuous edges corresponds to an incoming ego-network using the definition of EN^e_{in} with k=3. The node set is $V' = \{e, a1, a2, a3, a6, a7, a8\}$; it is important to note that the node $a4$ is not part of V' because there is no path going from $a4$ to e. Furthermore, all the nodes in V′ are at a shortest path with maximum length 3 from the ego. The set of directed edges is $E' = \{(a1, e), (a2, a1), (a3, a1), (a3, a2), (a6, a3), (a7, e), (a8, a7)\}$.*

Definition 4. *An outgoing ego-network EN^e_{out} of an individual ego e is a directed sub-graph $G'(V', E')$ of the directed graph G composed of the set of individuals x that are connected to e via an outgoing (i.e. a path that the ego is the start node) shortest path of length of maximum k going from e to x, and the set of links composed of the links being part of the outgoing paths from e:*

$$EN^e_{out} = G'(V', E') \subset G(V, E)$$
$$V' = \{x \in V \mid d_G(e, x) \leq k\} \cup \{e\}$$

$$E' = \{(x, y) \in E \mid x \in V' \wedge y \in V' \setminus \{e\}\}$$

where V' is composed of the ego node e and all the nodes at a maximum shortest path length k from it such that there exists an outgoing shortest path from e to each one of these nodes, and E' is the set of directed edges linking nodes in V'.

With the definition of outgoing ego-networks, we provide a way to consider only the outflow of information from a single ego node. This model can be used for understanding the spread of information starting from a specific source.

Example 4. *Given the example in Figure 1-b, the directed sub-graph with nodes connected by dotted edges corresponds to outgoing ego-network EN^e_{out} with k=3. The node set is $V' = \{e, a9, a10, a11, a12, a13, a14, a15\}$; since all the nodes in V′ are at a maximum shortest path length 3 from the ego. Also, the set of directed edges is $E' = \{(e, a9), (e, a13), (a9, a13), (a9, a10), (a13, a14), (a13, a12), (a14, a15), (a15, a12), (a12, a10), (a12, a11)\}$.*

Thus, the directed ego-network EN_d of the ego node e is given by the following property.

PROPERTY 1. — *A directed ego-network EN^e_d of an individual e is the union of the incoming and outgoing ego-networks: $EN^e_d = EN^e_{in} \cup EN^e_{out}$.*

Thereby, we proposed a definition for directed ego-networks, while distinguishing between incoming and outgoing models of information diffusion which was not possible with previous definitions and thus capture the complex reality of OSNs.

3.3. Weighted Ego-Networks

Definition 5. *A weighted ego-network is the ego-network EN^e of an ego node e given in Definition 2 having as an extra element a function f which assigns to each undirected edge in G' a value representing the tie strength of the edge.*

$$EN^e_w = G'(V', E') \subset G(V, E) \mid V', E' \text{as defined in Definition } 2$$
$$\exists f,\ f : E' \to \mathbb{R}$$
$$f(x, y) = a, \text{ where } (x, y) \in E' \text{ and } a \in \mathbb{R}.$$

It is important to note that the definition of d_G remains the same as in Definition 2. This means that we do not use the weights of the edges in order to compute the shortest paths when building the set V', but we only count the number of edges which are part of the shortest path.

Example 5. *The ego-network graph in Figure 1-a is a weighted graph that can be seen as corresponding to a personal undirected Facebook graph of the user e and that on each edge between two users in the social graph we have an integer value representing the number of private messages exchanged between them during a certain period.*

Thus, this definition provides a tool to study topics around the influence among online social network users or the strength of their connection. The same process can respectively be applied to define a weighted and directed ego-network.

4. A Software Framework for the Analysis of Personal Online Social Networks

4.1. The software framework PERSONA

The formal definitions for OPNs presented in Section 3 have been implemented in a software framework (named PERSONA: PERSonal Online social Networks' Analytics) that allows us to extract OPNs based on any of the definitions presented.

This allows firstly to validate the definitions and secondly to offer a versatile tool to anyone who wants to perform research around OPNs. The framework was implemented using the Java programming language. We offer two versions, one as a standalone desktop tool and another as a set of web services that allows us to manipulate OPNs over the web and build the functionality into other applications and tools.

The different definitions of OPNs were implemented as base classes in the framework as it can be seen in Figure 2. We started by the generic notion of a network using a same named class and then we derived classes corresponding to undirected OPN (according to Definition 2) and to directed OPNs depending on whether incoming (Definition 3) or outgoing links (Definition 4) are of interest. We can also derive the complete directed network by combining the incoming and the outgoing. Weighted networks, according to the Definition 5, can also be derived in combination with the previous. In all cases, one needs to specify only the ego she/he is interested in exploring. Finally we can express additional constraints on top of these exported OPNs either by providing the OPN's maximum path length from the ego (previously denoted as k) or by adding a constraint on top of the other properties that might exist in the database and concern the nodes.

The software framework can connect to any kind of relational database and as long as the database respects the documented format, it can retrieve the necessary data, independently of the kind of the network or the rest of its characteristics. It can also support an unlimited number of additional properties for the nodes and, as already described, weights for the edges of the network.

After the extraction of the OPN we offer the ability to export it in a set of different formats (GraphML, GML, DOT, CSV, etc.) and provide some basic visualization capabilities. Furthermore the user can analyze both the extracted OPN and the whole social network in order to compute various network metrics that might be useful in order to better understand the OPN and further analyze it. A non exhaustive list of these available metrics is presented in Table 1. We are currently expanding the available metrics to include additional ones; this also demonstrates the extensibility of the framework, which was one of the main goals of its design.

Figure 2. PERSONA's UML class diagram.

Metric	Description
Number of nodes and edges	Counts the number of nodes and edges that composes the OPN
Number of nodes per level	Counts the number of nodes existing on each level of the OPN
Degree centrality	Counts the number of connections a given node has in the OPN
Betweenness centrality	Measures the extent to which the ego lies on paths between the other vertices in the OPN (Freeman, 1977)
Density	Counts the number of possible edges between the nodes composing the OPN
Effective size	Measures the ego impact inside its 1-level OPN, given by the number of ego's immediate alters minus the average number of links between the ego's 1st level alters (Burt, 1995)
Global clustering coefficient	Computes the number of closed triplets over the total number of triplets in the OPN (Newman, 2003)

Table 1. Main metrics computed by PERSONA.

4.2. Usage Examples

We used the software framework in order to extract examples of OPNs from the DBLP (Digital Bibliography & Library Project) dataset[1], which consists of all the available published papers that are recorded in DBLP until 2013. The DBLP data set offers the possibility to retrieve personal networks according to the three cases of the proposed definitions i.e: undirected, directed and weighted. We were actually able to correctly (we verified this by hand) extract all these different kinds of personal networks. We mainly focused on publications in the Computer Networks field that holds 38134 papers from 1588 conferences and journals published between 1975 and 2013. From which we identified a set of 23010 authors. Based on this data, we built two social networks. The first one is the collaboration network which can be defined as an undirected graph where a node represents an author and an edge represents a collaboration (co-authorship) between two authors. Furthermore this network can be enriched with an additional information that characterizes the edges which consists in the number of collaborations between each pair of authors (weight). Thereby we get an undirected and weighted network. The second network we constructed is the citation network, which is a directed graph where nodes reflect papers and edges reflect a citation relationship between two papers.

Once we built these two social networks, we can deploy the developed framework to extract personal networks that fit the different definitions described in section 3. In

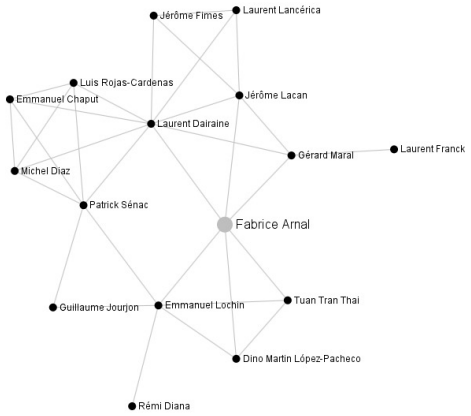

Figure 3. Personal network of author "Fabrice Arnal".

the collaboration network, each author has its own personal network. For any given focal author, we can extract the corresponding personal network and we can further restrict it based on both the k (the maximum distance an author will have from the focal author) and the year of the first collaboration (a property in the database). In

1. https://aminer.org/citation

Figure 3, we give the 2-level personal network of the author "Fabrice Arnal" (ego), retrieved according to Definition 2 for k=2 and until 2013. For the citations' network, one can express for a given paper (ego), the incoming personal network for instance i.e the personal network where we represent the papers that cite that focal paper (k=1) and extend it by including the papers that cite each paper that cite the ego paper (k=2), etc. Similarly we can define the outgoing personal network of a paper (ego) where the papers at distance k=1 from it represent the papers cited by that ego paper, then the papers cited by these papers at k=2, etc.

5. Conclusions and Future Work

In this paper, we propose a set of formal definitions for online personal social networks (OPNs). The definitions are independent of any application or online social network and capture all diverse existing cases, including weighted, directed (incoming and outgoing) and undirected OPNs. To the best of our knowledge there is no other systematic and formal recording of these concepts in a way that provides a universal framework for studying personal online social networks in general. We validated these definitions by providing an extensible software framework for OPNs' extraction, visualization and analysis. This framework is currently being extended to provide additional metrics and to allow for building on top of it evolutionary predictive models that will be used to express the dynamics of OPNs in the future, something valuable for all online social networks.

References

Ahn Y.-Y., Han S., Kwak H., Moon S., Jeong H. (2007). Analysis of topological characteristics of huge online social networking services. In *Proceedings of the 16th international conference on world wide web*, pp. 835–844. New York, NY, USA, ACM. Retrieved from http://doi.acm.org/10.1145/1242572.1242685

Aral S., Walker D. (2014). Tie strength, embeddedness, and social influence: A large-scale networked experiment. *Management Science*, Vol. 60, No. 6, pp. 1352–1370. Retrieved from http://dx.doi.org/10.1287/mnsc.2014.1936

Arnaboldi V., Conti M., Passarella A., Pezzoni F. (2012). Analysis of ego network structure in online social networks. In *Socialcom/passat*, p. 31-40. IEEE. Retrieved from http://dblp.uni-trier.de/db/conf/socialcom/socialcom2012.html#aboldiCPP12

Arnaboldi V., Conti M., Passarella A., Pezzoni F. (2013). Ego networks in twitter: An experimental analysis. In *Infocom*, p. 3459-3464. IEEE. Retrieved from http://dblp.uni-trier.de/db/conf/infocom/infocom2013.html#aboldiCPP13

Arnaboldi V., Passarella A., Tesconi M., Gazzè D. (2011). Towards a characterization of egocentric networks in online social networks. In *Otm workshops*, Vol. 7046, p. 524-533. Springer. Retrieved from http://dblp.uni-trier.de/db/conf/otm/otm2011-w.html#ArnaboldiPTG11

Bakshy E., Rosenn I., Marlow C., Adamic L. A. (2012). The role of social networks in information diffusion. *CoRR*, Vol. abs/1201.4145. Retrieved from http://dblp.uni-trier.de/db/journals/corr/corr1201.html#abs-1201-4145

Burt R. S. (1995). Le capital social, les trous structuraux et l'entrepreneur. *Revue française de sociologie*, pp. 599–628.

Conti M., Passarella A., Pezzoni F. (2011). A model for the generation of social network graphs. In *Wowmom*, p. 1-6. IEEE. Retrieved from http://dblp.uni-trier.de/db/conf/wowmom/wowmom2011.html#ContiPP11

Dunbar R., Spoors M. (1995). Social networks, support cliques, and kinship. *Human Nature*, Vol. 6, No. 3, pp. 273-290.

Ferrara E., Meo P. D., Fiumara G., Provetti A. (2012). The role of strong and weak ties in facebook: a community structure perspective. *CoRR*, Vol. abs/1203.0535. Retrieved from http://arxiv.org/abs/1203.0535

Freeman L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, pp. 35–41.

Gatti M. A. d. C., Appel A. P., Santos C. N. dos, Pinhanez C. S., Cavalin P. R., Neto S. B. (2013). A simulation-based approach to analyze the information diffusion in microblogging online social network. In *Proceedings of the 2013 winter simulation conference: Simulation: Making decisions in a complex world*, pp. 1685–1696. Piscataway, NJ, USA, IEEE Press. Retrieved from http://dl.acm.org/citation.cfm?id=2676128.2676193

Hill R., Dunbar R. (2003). Social network size in humans. *Human Nature*, Vol. 14, No. 1, pp. 53-72. Retrieved from http://dx.doi.org/10.1007/s12110-003-1016-y

Newman M. E. (2003). The structure and function of complex networks. *SIAM review*, Vol. 45, No. 2, pp. 167–256.

Quercia D., Capra L., Crowcroft J. (2012). The social world of twitter: Topics, geography, and emotions. In J. G. Breslin, N. B. Ellison, J. G. Shanahan, Z. Tufekci (Eds.), *Icwsm*. The AAAI Press. Retrieved from http://dblp.uni-trier.de/db/conf/icwsm/icwsm2012.html#QuerciaCC12

Roberts S. G. B., Dunbar R. I. M., Pollet T. V., Kuppens T. (2009). Exploring variation in active network size: Constraints and ego characteristics. *Social Networks*, Vol. 31, No. 2, pp. 138-146. Retrieved from http://dblp.uni-trier.de/db/journals/socnet/socnet31.html#RobertsDPK09

Tchuente D. (2013). *Modélisation et dérivation de profils utilisateurs à partir de réseaux sociaux : approche à partir de communautés de réseaux k-égocentriques*. Thèse de doctorat, Université Paul Sabatier, Toulouse, France.

Ugander J., Karrer B., Backstrom L., Marlow C. (2011). The anatomy of the facebook social graph. *CoRR*, Vol. abs/1111.4503. Retrieved from http://dblp.uni-trier.de/db/journals/corr/corr1111.html#abs-1111-4503

Yen T.-J. (2014, August). *Analysis of egocentric network data.* (Lecture available at http://www3.stat.sinica.edu.tw/summerschool2014/document/0801-2.pdf Retrieved on 23/02/2015)