The ML-model for multi-layer social networks

Matteo Magnani
Department of Computer Science
University of Bologna
Via M. Anteo Zamboni 7, 40127 Bologna
Tel: +39 051 2094848
Email: matteo.magnani@cs.unibo.it

Luca Rossi
Department of Communication Studies
University of Urbino "Carlo Bo"
via Saffi 15, 61029 Urbino
Tel: +39 0722305727
Email: luca.rossi@uniurb.it

***Abstract*—In this paper we introduce a new model to represent an interconnected network of networks. This model is fundamental to reason about the real organization of on-line social networks, where users belong to and interact on different networks at the same time. In addition we extend traditional SNA measures to deal with this multiplicity of networks and we apply the model to a real dataset extracted from two microblogging sites.**

I. Introduction

Our contemporary social experience is largely based on our ability to master a growing number of social contexts. Job interviews, chatting among friends or casual conversation at the bus stop, all these social interactions require a specific perception of our audience and the ability of choosing the right topic and tone among a list of many possibilities. We need to perform our social role starting from a clear definition of the social frame [1] we are performing in.

When we move into the on-line context this scenario is, if possible, even harder to manage. On-line communication offers undoubtedly many opportunities to experience this multiplicity of identities and this has been one of the classical topics in Computer Mediated Communication studies [2]. Within this perspective on-line identity has often been described as a self conscious performance of identity practices [3] played in different on-line contexts. The collapse of social contexts followed by the growing success of Social Network Sites [4] shift this phenomenon even forward. Within social network sites e.g. Facebook or even more prominently within microblogging sites like Twitter or Friendfeed, users hardly have a clear perception of their final audience and the control of the social context where their communicative interactions are taking place is often a challenge [3].

As a partial solution to this problem we saw over the last few years how, despite a general movement toward a reduction of the number of Social Network Sites (largely due to the incredible success of Facebook), dedicated social networks seem to be still very popular. Services like LinkedIn - for professional networking - or match.com - for on-line dating - show how, despite a huge social pressure pushing us toward a merging of our on-line presence, there is an interest in maintaining a higher control over specific areas of our on-line presence. Therefore, at the moment, users' on-line presence is no longer scattered through hundreds of difference services but, at the same time, it is not yet unified within a single service (and it seems this will not happen anytime soon).

On a metaphorical level contemporary on-line presence is similar to a pillar, representing a single user connected to other users on several autonomous layers (or floors of this architecture made of multiple networks). Two users might be connected by many layers at the same time - e.g. two friends may be *friends* in Facebook, in Flickr and YouTube and followers in Twitter - while other users might be connected on just one layer - e.g. like co-workers connected only through LinkedIn. Pillars are therefore the linkage between several networks each one representing only a layer of that user's on-line presence. The result is a highly complex system made of several layers, one for each on-line service where users can be present, and where users (nodes) can be connected to each other through several edges (connections) with different properties. The way in which different users exploit the different kinds of connections that are available to them is related to their own personal strategy of on-line identity management. The choice about what to share and using which layer/system is obviously part of this strategy. At the same time users have the ability to move conversations (or topics) from a layer to another.

The impact of studying the whole system instead of each single network should be clear if we think of real and typical phenomena like on-line information diffusion. YouTube users are interconnected through following and friendship links allowing them to know when other users have posted new videos. However, news, *memes* and almost any kind of on-line information usually spread through the network bouncing from a service to another one: e.g. people post videos on YouTube but this videos often reach a high visibility when users start posting about these videos on their blogs or on Twitter or Facebook. This diffusion on Facebook may then have consequences on YouTube, where for instance new connections can be created between users who where exposed to their videos on Facebook. Studying this kind of phenomena in depth is impossible if we do not consider all the networks involved, and in this paper we also show that even considering all networks but separately can lead us to inaccurate results.

For this reason we define what we call the ML Model (Multi Layer Model). A ML model is able to deal with the high complexity of our contemporary on-line experience especially when we want to observe how social phenomena, as well as information, propagate through several networks at the same time. As we can see in our everyday experience and as we have exemplified in the previous paragraph the information propagation process rarely involves only one service or one network. This kind of phenomena implies at least two different social actions: on one side users act in different on-line social contexts dealing with different perceptions of their audience and - at the same time - their activity contributes to build their own on-line identity.

The contributions of this work are the following:

- We define a model for the representation of multi layer networks.
- We extend classical SNA centrality measures to apply them to multi layer networks.
- We report on the extraction of a real large multi layer network from two microblogging sites, which has been made publicly available.
- We present the results of the application of the extended measures to this network.

In summary, we will see that studying a single network often limits our knowledge of on-line organization and dynamics, and even studying separately the organization of multiple networks may not be sufficient to understand the overall role and position of some users. In addition, by comparing SNA metrics computed on the single networks and on the ML network it is possible to understand how much the single networks are complementary to each other or have a similar social function.

This paper is organized as follows: in the next section we review existing work in this area. While we cannot be exhaustive for space reasons and for the number of studies potentially related to the topic of this paper, we have tried to indicate the main references to draw the sociological context in which our work can be inserted and to indicate related and complementary researches. Then we formally define the ML model and extend centrality and closeness measures to apply to it. In the following section we describe the dataset used to test our model and metrics and how we built it from real social data. We conclude with a discussion of the experimental evidences.

II. Related work

This paper deals with at least three fields of related researches. On one side the ML model deals with the sociological researches on on-line presentation of the self by providing a formal model of the several *layers* composing our everyday on-line presence. On another side it deals with complex social network theories by offering a new approach for social network analysis and, at the same time, the final goal of the ML model is to provide a wider perspective on on-line propagation phenomena. Due to this highly interdisciplinary approach a large base of previous researches have to be taken into consideration.

On-line self presentation has to be understood as a technological mediated aspect of the broader issue of identity construction which can be described from a situationist perspective [5] as an understanding of the context where the communicative interaction takes place leading to a specific choice about what kind of role is more suitable to perform [1].

As clearly observed by Meyrowitz's *No Sense of Place* [6] media change in a radical way the places where our communication happens. Electronic media - from radio to television - delete walls and boundaries that have often kept specific social contexts isolated. This phenomenon is not only affecting the way in which we define the distinction between what is public or what is private but affects also our possibility to have a clear perception of our *audience* and therefore to have a clear understanding of the social context. The perception of the audience is therefore always more the perception of an *imagined audience* which is used in order to perform our Self-conscious identity construction processes [3]. This consciousness of an audience may change depending on what kind of on-line service we are studying [7] (in goal oriented spaces people are more conscious of that) nevertheless all on-line users have to be understood as multiple selections of possibilities operated according to the goals and the *imaginated* audience of the on-line places they populate.

Dynamic Network Analysis is an emerging research area partially related to the goal of this paper. Dynamic Network Analysis (DNA) brings together social network analysis (SNA), link analysis (LA) and multi-agent systems (MAS) within network science. This interesting and emerging approach can describe networks with multiple kinds of nodes and multiple types of links connecting them [8], at the same time, similarly to the model we are proposing, nodes can belong to several networks at the same time. This approach and the related meta-matrix model [8] have been widely adopted in organization studies [9] and there has also been a great interest on the study of Dynamic Social Networks [10] that we expect to increase with respect to on-line networks in the future.

This multiplicity has to be taken into consideration when we start investigating the problem of on-line propagation. As we have highlighted in previous work [11], [12] the choice between propagating or not a specific information item and through what kind of service doing that is mainly a choice related to our perception of our audience and related to what kind of identity we are constructing within that specific on-line place. On a more general level the propagation of items through networks is a very abstract and general problem which has been studied in several fields. In general, according to our review of existing formalizations, it appears to be a very complex problem for which a simple mathematical solution does not exist. As a result all approaches have carefully exploited the specificities of their application fields according to the specific items traversing the network, e.g., viruses or Internet surfers, and developing specific solutions that worked well under those assumptions but are not meant to be general answers to this problem. Our goal here is not to explain in detail every approach that has been used but to highlight how previous researches and uses in different fields can provide insights into the topic. It is important to highlight that we are not going to move seamless ideas and concepts (*such as viral or propagation*) from a scientific field to another. Every discipline has its own specificity and moving concepts (and research methods) around will only generate greater confusion rather than real knowledge. Within this perspective being able to stress differences appears to be as important as stressing similarities.

Epidemiology has always tried to understand how *viruses* and other *pathogens* spread over the population at different times and with different modalities. This is aimed at being able to foresee how many people are susceptible to be infected with a specific disease and to be able to undertake proper preventive actions. The introduction of Social Network Analysis in epidemiological research has been based both on Social Sciences and Graph Theory achievements [13] and rose new and unexplored methodological problems. Observing how a disease spreads in the real word assuming the existence of a network of social relations underneath pushes researches toward the identification of *meaningful social connections*. From an epidemiological point of view relevant connections are those able to spread the *pathogens*:

> If networks are to be used for epidemiological purposes, then connections should only be included if they describe relationships capable of permitting the transfer of infection. [13]

This leads to the methodological challenge of being able to observe and trace only what is relevant from a specific point of view, and highlights the role of social networks in *pathogens* propagation. Nevertheless there are many crucial differences between the propagation of an epidemic and the spreading of information in an on-line context. Those differences can be grouped in two major areas: differences related to the nature of the *connection identification* and differences in the *node behaviour*.

Epidemiologists struggle with the challenge of identifying what constitutes a *meaningful connection* between two nodes. This changes dramatically if the traced disease is a *sexual transmitted disease* or an *airborne pathogen*. The identification of what constitutes a link is, after all, up to the researcher and to his/her evaluation. In the on-line context, on the opposite, the definition of the explicit connection between users, the *meaningful link*, is part of the SNS itself and its establishment appears to be an explicit choice of the user. Nevertheless the existence of any kind of on-line connection among two users does not imply the actual use of that connection. This leads us to the second difference.

The second difference when we shift our focus from viral diffusion to information/cultural spreading is about the nature of the *virus* itself and the freedom of the host/node. The metaphor of media virus [14] had (and still has) great success among the large audience. Those *viruses*, often referred as memes, are described as media-based viruses not conceptually dissimilar to what we know from biological studies. Memes are described [14] as units of information capable of:

1) retaining their informational content;
2) inducing people to reproduce the meme itself;
3) staying alive as long as they are able to be reproduced.

According to Jenkins [15], who is investigating how cultural contents spread through our society, there are many crucial differences in the way viruses and cultural content spread. The epidemiological metaphor, even if it is very attractive, should not be used. Jenkins' point stresses the role of the end users in the propagation process. While in virus spreading people are almost passive carriers of viruses (they cannot choose if they want to be infected or not and, if infected, they have no choice between spreading the virus as it is or changing it) memes need some kind of collaboration to their propagation. If it is obviously possible that someone is unintentionally exposed to any kind of *unit of information* the choice between spreading it or not and the way in which it has to be done is definitely up to the single person. This means that the spreading of specific information can be done also to pursue specific personal interests, to enforce personal relationships between users or according to a personal definition of relevance [16]. Information spreading in socio-technical context is not only matter of what has the major chance of being replicated but also of how this replication is used by the members of a specific cultural context.

Within this perspective the *nodes* of a social network involved in the spreading of information, as well as those involved in the spreading of any kind of cultural object, are substantially different from those involved in the spreading of a viral agent. We have to take into consideration nodes with a set of characteristics able to explain why in cultural contexts: exposition $\neq$ contagion $\neq$ spreading. The active role of media audiences has been part of any media spreading theory since long time [17]. A theoretically founded research on propagation in SNSs should not try to simply show how information propagates through a SNS but also understand what is the role of SNS structures and connections in the larger process of propagation of cultural information. Within this process the belonging of every single node to several networks (that we have defined as layers) and the active role of every single node in acting through these layers according to his/her strategies or his/her perception of the audience have to be part of any model aimed at understanding on-line social phenomena.

III. The ML model for multi-layer networks

In this section we provide some general formal definitions, with the aim of covering a large number of potential applications, then describe a possible interpretation of these definitions in the context of multiple on-line social networks. These definitions are necessary to extend traditional social network analysis metrics to multi-layer networks. In particular, later in this section we extend two classical centrality measures (degree and closeness) and discuss why these extensions allow us to extract more information than what we can obtain from the analysis of the single networks.

A. Data structures

The definition of a multi-layer model is based on simple weighted networks.

Definition 1: A Network Layer is a weighted graph (V, w) where V is a set of vertexes and $w : (V \times V) \rightarrow [0, 1]$.

In general social networks are dynamic environments, where the number of nodes and weights change over time. However in this paper we are more interested in investigating the aspects related to the coexistence of multiple networks, therefore we will focus on static data structures to avoid the introduction of unnecessary formalisms.

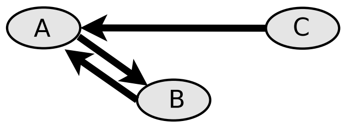

Fig. 1. A single-layer network.

Example 1: In Figure 1 we have represented a single network with three users and three edges (without specifying the weights). This may correspond to a portion of a Twitter network, with following relationships. Weights can be used to represent the strength of the relationships from the point of view of information flows: for instance they can be used to represent the probability that a user will reply to another.

When we start considering multiple networks, we need to know which nodes in one network correspond to nodes in the other. This is done using a Node Mapping.

Definition 2: A Node Mapping from a Network Layer $L_1 = (V_1, w_1)$ to a Network Layer $L_2 = (V_2, w_2)$ is a function $m : V_1 \times V_2 \rightarrow [0, 1]$. For each $u \in V_1$, the set $\mathcal{C}(u) = \{v \in V_2 \mid m(u, v) > 0\}$ is the set of V_2 nodes corresponding to u.

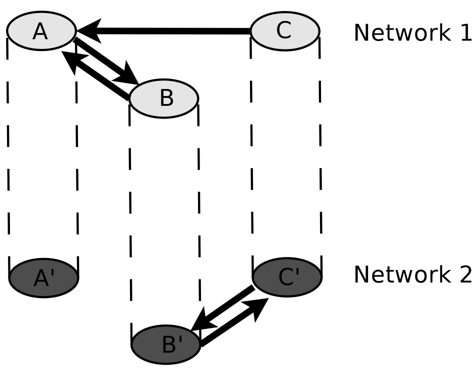

Fig. 2. A Pillar Multi-Network.

Example 2: The three users in the previous example may also have an account on Facebook, and here we can use another network to represent these three accounts and their relationships. This scenario has been represented in Figure 2, where we exemplified how the connections between Facebook friends may not correspond to Twitter connections — for instance a user does not necessarily follow the Twitter status of his friends, and often we follow public figures that are not

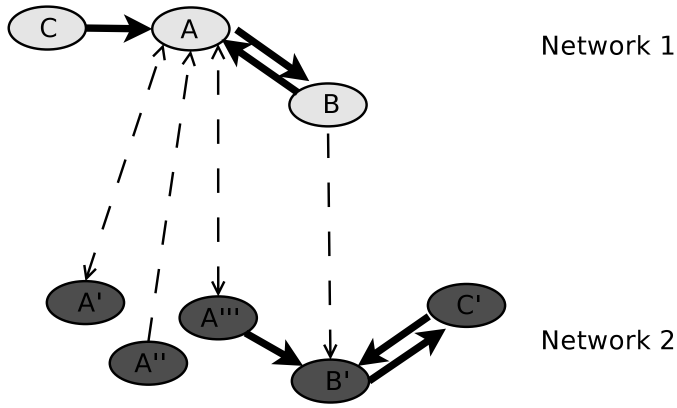

Fig. 3. A general Multi Layer Network

our Facebook friends. In this example every user has exactly one account on each layer. For this reason we call this a *Pillar multi-network*, where every user can be seen as a pillar traversing several ”floors” and posed on the lower layer, not represented here, that is the level of physical reality, with a specific geographic location. A Pillar model is characterized by $|\mathcal{C}(u)| \in \{0, 1\}$.

Here weights can be used to indicate for each account in V_1 the probability that a message will be exported to V_2, and vice-versa. In fact, having two or more accounts does not necessarily mean that the same user will produce the same content on the two layers — which by the way can be made difficult by the different technological settings of the SNSs. In practice we can use the value of the Node Mapping function to represent the probability that an information present on a node u on one network will be posted to nodes in $\mathcal{C}(u)$. Thinking of real systems, this probability can be 1 in case one network is a social media aggregator where the user has registered the other account — like Twitter messages imported automatically into Friendfeed, it can be close to 1 — like when we check in somewhere with Foursquare and posting it on Facebook involves just a confirmation if we registered the account — or it can be close to 0, when the audiences of the two networks are too different — like when we exchange on Facebook links that would not be appropriate for a professional network like LinkedIn.

Finally we can define a generic Multi Layer Network, which consists in a set of networks and a matrix of Node Mappings.

Definition 3: A Multi Layer Network is a tuple MLN $= (L_1, \ldots, L_n, \text{IM})$ where $L_i = (V_i, w_i), i \in 1, \ldots, n$ are Network Layers and IM (Identity Mapping) is an $n \times n$ matrix of Node Mappings, with $\text{IM}_{i,j} : V_i \times V_j \rightarrow [0, 1]$. If we define $v = \{v\}$, $\text{IM}_{i,i}$ is the identity mapping for the i^{th} layer: $\text{IM}_{i,i}(v, v) = 1$.

Example 3: In Figure 3 we have represented a multi layer network more complex than a simple Pillar model. Here the same node in one network may correspond to multiple nodes in another. This is a typical case in social media aggregators, where for example we can have a BBC_journalists account

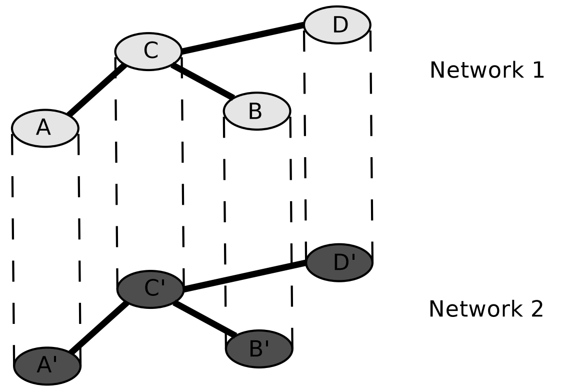

Fig. 4. Multi-Layer degree centrality: a user not increasing his/her audience

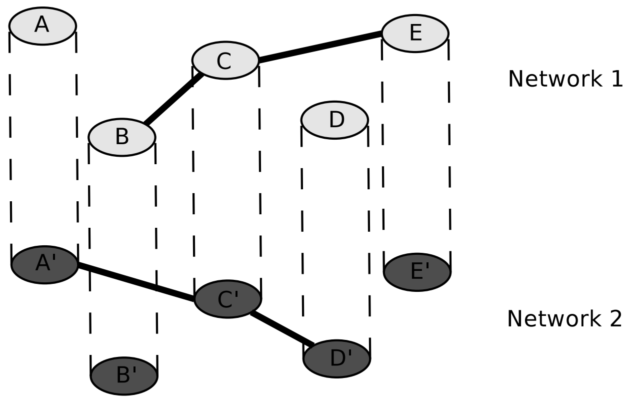

Fig. 5. Multi-Layer degree centrality: a user increasing his/her audience

following all registered BBC journalists and providing a single access point to all their updates. As a consequence, in this case nodes do not represent only users but more in general accounts.

B. Analysis metrics

In this section we extend two fundamental SNA metrics to the context of Multi-Layer networks. It will be clear how this extension enhances our analysis power with respect to considering each network separately.

In Figure 4 we have represented a user with three connections in each network. If we look at the correspondences, however, we can see that the six connections include the same three users, therefore the Multi-Layer audience will be composed by six nodes but limited to three people. On the contrary, in Figure 5 we have represented a case where looking at the two single networks the central user would look less connected than the one in the previous example. In fact he/she has two connections in the first network and two in the other, against the three and three of the previous example.

However, this second user is exploiting the two networks in different ways, managing distinct audiences in one and the other. As a consequence, the overall degree centrality will be 4, more than in the previous example.

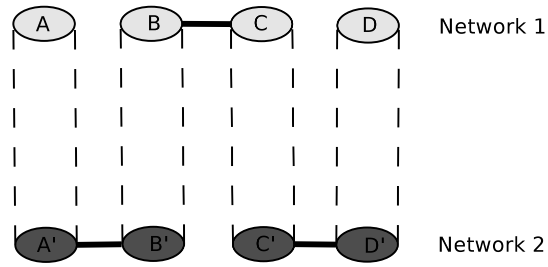

Fig. 6. Node reachability in a Multi-Layer Network

The definition of the extended degree centrality just described by example is not immediate, and we can proceed in three steps. First, given a Multi-Layer Network MLN $= (L_1, \ldots, L_n, \mathrm{IM})$ and a node $v \in V_i$ we can get all the nodes in all networks connected to it. If we focus on the case of directed unweighted networks this is expressed by $\bigcup_{i \in [1,n], (u,v) \in E_i} u$. In Figure 4 this set computed on C corresponds to $\{$A, B, D, A', B', D'$\}$ while in Figure 5 it is $\{$B, E, A', D'$\}$. At this point we want to remove nodes that already contribute in another network to the audience of C. To do this we define an equivalence class $v \; eq_{IM_i} u$ iff $IM_i(u,v) > 0$. As an example, in Figure 4 nodes A and A' are equivalent, while there is no node equivalent to another in Figure 5. This equivalence class defines a partition $\mathcal{P}_{eq_{IM_i}}$ of our set of users and the number of sets in this partition corresponds to the real audience. In our example represented in Figure 4 the resulting partitioned set would be $\{\{$A,A'$\}$,$\{$B,B'$\}$,$\{$D,D'$\}\}$. Each partition indicates a set of corresponding nodes, e.g., the account of the same person on different networks, therefore they contribute only once to the computation of the centrality measure. In this example the resulting degree would be $|\{\{$A,A'$\}$,$\{$B,B'$\}$,$\{$D,D'$\}\}| = 3$. On the contrary the Degree of the multi layer network in Figure 5 would be $|\{\{$A'$\}$,$\{$B$\}$,$\{$C'$\}$,$\{$D$\}\}| = 4$.

Definition 4: Let MLN $= (L_1, \ldots, L_n, \mathrm{IM})$ be a Multi-Layer Network with weights 1. The In-Degree Centrality of a node v is defined as:

$$\delta(v) = |\mathcal{P}_{eq_{IM_i}}(\bigcup_{i \in [1,n], (u,v) \in E_i} u)|$$

The definitions for Out-Degree and Degree are simple modifications of this formula.

Now we can consider closeness, i.e., a measure of how nodes are close to each other. In this definition we will consider also the weights of edges, but we introduce it through simple unweighted examples.

Figure 6 shows that two users A and D whose accounts are not connected to each other can be in fact connected if we consider the ML network. In fact, an information item could reach D from A through the path A $\rightarrow$ A' $\rightarrow$ B' $\rightarrow$ B $\rightarrow$ C $\rightarrow$ C' $\rightarrow$ D' $\rightarrow$ D. This information flow process involves some normal in-network propagations and the *choices* of some users that the information is worth propagating also in the other network.

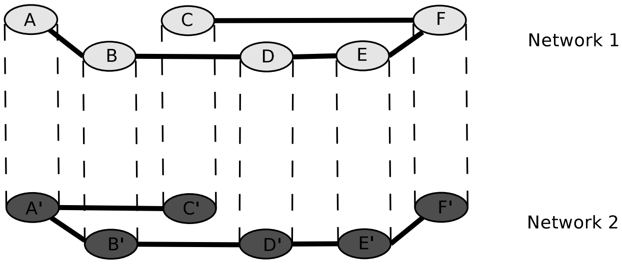

Fig. 7. Decreasing node distances in a ML Network

Figure 7 shows that even for users that are already connected to each other in one or both networks their distance may decrease. Also in this case this depends on the transfer of information from one network to the other.

From the point of view of defining closeness centrality, i.e., the average inverse distance of one node from the others, these examples highlight how we can compute the extended distances by considering a single network obtained starting from the MLN.

Definition 5: Let MLN $= (L_1, \ldots, L_n, \mathrm{IM})$ be a Multi-Layer Network and flat(MLN) = (V, w) where:

- $V = \bigcup_i V_i$
- $w(u, v) = w_i(u, v)$ if $u, v \in V_i$
- $w(u, v) = IM_{i,j}(u, v)$ if $u \in V_i, v \in V_j, i \neq j$

The distance of two nodes u, v in the MLN is defined as their distance in flat(MLN).

IV. Experimental results

Our experimental phase has been designed to investigate how the use of the ML model impacts on the ability to describe a complex real world situation made of users with multiple accounts on several social network sites. Therefore we have selected a group of users each one with an active account on two microblogging sites: FriendFeed and Twitter. The two services appear to be very similar regarding their general goal: both are microblogging services (even if Friendfeed is more an aggregator of on-line content) and both allow the sharing of various kinds of information toward a list of followers. In both cases there is no technical requirement of reciprocity in the following mechanism. Given this situation we are claiming that the best way to describe this scenario is to use the proposed ML model and not separate models for each network.

A. Data extraction

Friendfeed is a social media aggregator. In this system while users can directly post messages and comment on other messages much like in Facebook and other similar SNSs, they can also register their accounts on other systems. In this way, using the Friendfeed API we could retrieve the multiple accounts of the same users for several social services.

In our dataset, which is available for download on the project Web site (http://larica.uniurb.it/sigsna) we collected 322,967 users who registered at least one service outside Friendfeed, with a total number of 1,587,273 services. In

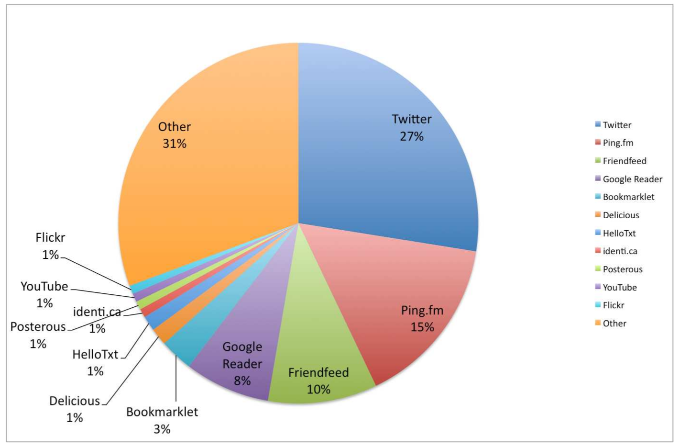

Fig. 8. Top 10 Sources of posts

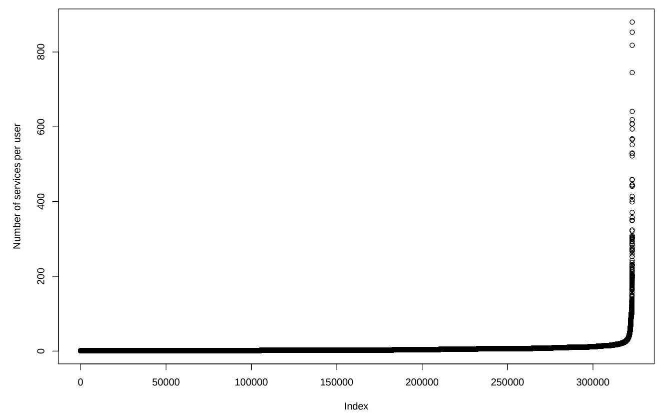

Fig. 9. Number of services registered by each user with at least one service

Figure 8 we show how these sources contributed to the User Generated Content on the site in August-September 2010, and in Figure 9 we show the distribution of the number of registered services.

At this point we focused on two networks: the same Friendfeed and Twitter. Therefore we extracted all users who registered exactly one Twitter account and whose Twitter account was associated to exactly one Friendfeed user, this to remove collective accounts not corresponding to single persons and to build a Pillar ML Model. As a result we selected 155,804 users.

The final step of the data extraction phase consisted in the retrieval of the mutual connections between these users, which was done respectively using the Friendfeed and Twitter APIs and resulted in a Friendfeed network with 5,939,687 arcs and a Twitter network with 13,142,341 arcs.

The first analysis that we carried on was the comparison of the degree centrality among the two networks (Friendfeed and Twitter) and the ML network [18] [19]. Our data showed some interesting results. Figures 10 and 11 show how there is only a limited linear correlation between degree centrality in the two microblogging services and in the ML network. This

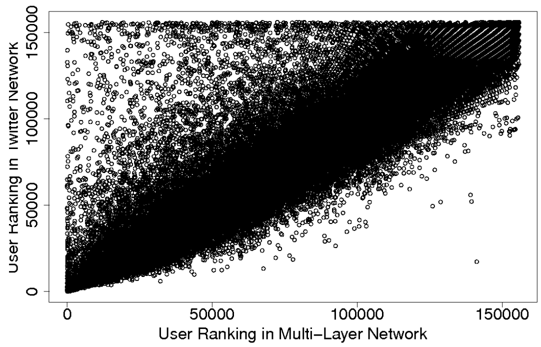

Fig. 10. Degree Centrality Ranking ML Network - Twitter Network

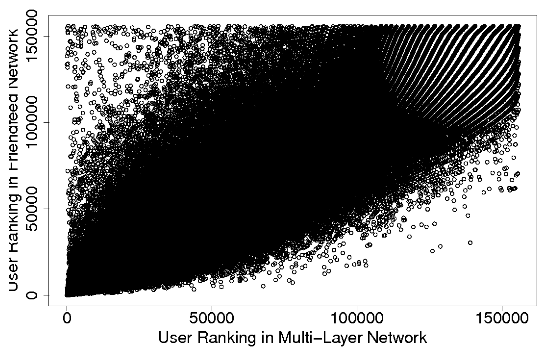

Fig. 11. Degree Centrality Ranking ML Network - FriendFeed Network

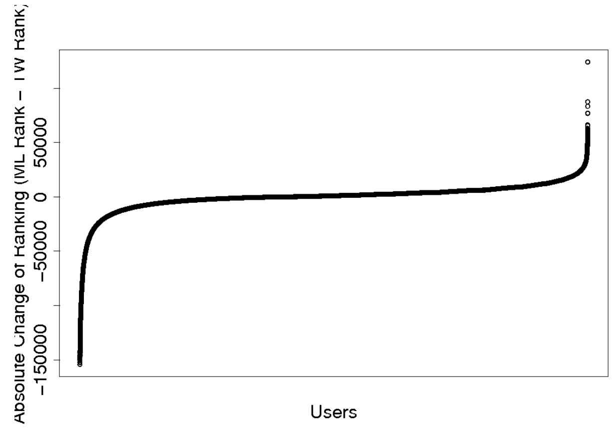

Fig. 12. Increasing in centrality degree ranking per user: Twitter Network

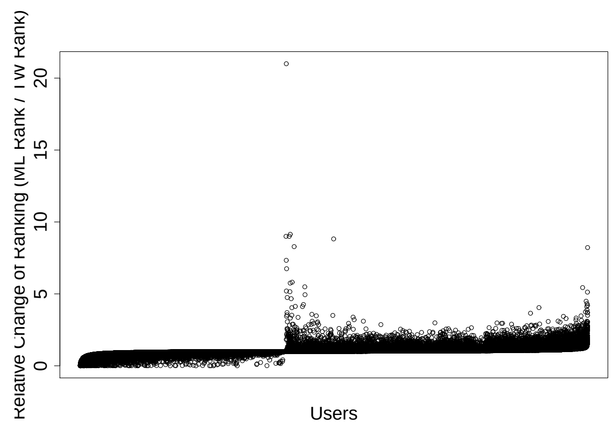

Fig. 13. Increasing in centrality degree value per user: Twitter Network

means that users may have a very different degree centrality in the two services: a single users might have a huge number of connections on one system (e.g. Friendfeed) and a small number of connections on the other (e.g. Twitter). Despite the close similarity in services' goal it appears to be clear that social behaviors on every single system change the users' centrality within the network, therefore given two systems – even though very similar from a technical point of view - a proper measurement of the degree centrality in our scenario can be obtained only through the analysis of the ML network.

The evaluation of the degree centrality on the ML network rises a couple of more specific questions: how much the degree centrality ranking changes by shifting from a single network approach to a ML network approach? As a consequence of that how much does this affect the single user? Figures 12 and 13 highlight possible answers to these questions. Figure 12 represents the gain obtained within the degree centrality ranking by using the ML network instead of the single Twitter network. It appears that for the majority of cases there is a very small increment of ranking positions but at the same time there are nodes losing or gaining many positions.

If the ranking of a whole network is a useful piece of information to describe it, switching the perspective to how many times the degree centrality of a single user changes when using the ML network model might provide many insights into the single persons's use of the networks where he/she has an active account. Figure 13 shows the increasing factor of the degree centrality of every single user when it is calculated on the ML network (the comparison is made with the Twitter network). While the average is quite small it appears evident that there are users that have up to twenty times more connections when these are calculated on the ML network. This suggests a very different kind of usage of the two services. Even if we are not able so far to make a qualitative evaluation about how users exploit the two networks we can surely claim that a proper evaluation of their on-line presence cannot de observed only through a single service and that the ML model we are suggesting offers a better description of this scenario.

The second analysis is related to shortest paths and to connected components. Introducing the ML model may in fact have two direct consequences on these aspects, as we have exemplified when we introduced the extended distance function: on one side the distance that separates two nodes might be shorter when we take into consideration the opportunity to switch to another network. In our case two users might not be directly connected on Friendfeed (even if they might belong to the same connected component of the network) and be directly connected on Twitter. On the other side, as we illustrated in previous examples, two users might not belong to the same

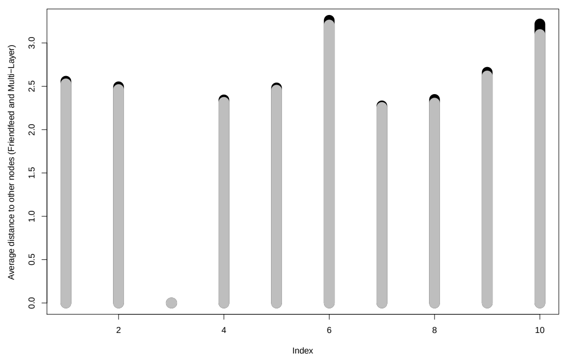

Fig. 14. Closeness for ten nodes of the Friendfeed network (black) and the Multi-Layer Network

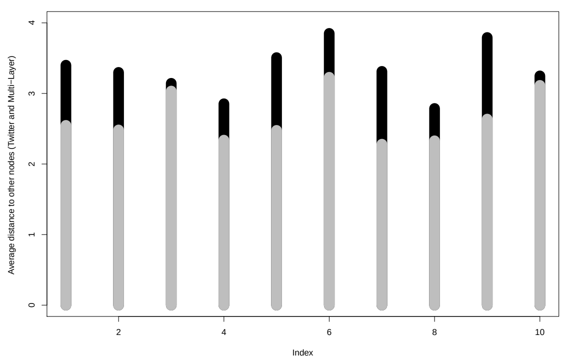

Fig. 15. Closeness for ten nodes of the Twitter network (black) and the Multi-Layer Network

connected component on one network while they might be very close on the other.

As a result of building the ML Network, about 5,000 nodes on Friendfeed and about 200 in Twitter became connected to their respective giant components, becoming reachable from the majority of nodes. These numbers are not very relevant if we consider the size of the analyzed networks. More interestingly, Figures 14 and 15 show for a sample of ten nodes their inverse closeness centrality (i.e., the average distance between them and all nodes reachable on the single networks), both on the two single networks (black) and the ML network (gray, where we computed the distance function defined in this paper). These results show that on the Friendfeed network the distance between nodes is already low in general: considering another network does not introduce many better paths. On the contrary, although more *dense* in average given the larger number of edges, the Twitter network is more scattered and there seem to be less connected regions slowing down the flow of information (i.e., increasing distances). As a consequence the impact of adding another layer is much more important in this network, showing again that extending our perspective we can find a different scenario with respect to the one visible from the point of view of a single layer.

V. Conclusion

In this paper we have proposed a model for the representation of multi layer networks, together with extensions of classical SNA centrality measures. Based on this proposal, other centrality and SNA metrics will be extended in the future. The application of these metrics to a real large multi layer network has confirmed that considering a complete network-of-networks model allows us to extract results from our analyses that do not correspond completely to the ones that can be obtained from each network separately.

We think that introducing multi-layer models in different kinds of studies of on-line networks could boost open research directions and potentially open new ones, and in particular the study of the connections between different layers, i.e., the dynamics of information propagation from one network to the others, and also the topic of data integration for social networks that appears to be a fundamental activity to build unified user profiles from distributed on-line accounts.

References

[1] E. Goffman, *Frame analysis: an essay on the organization of experience*, ser. Harper colophon books ; CN 372. New York: Harper & Row, 1974.

[2] S. Turkle, *Life on the screen: identity in the age of the Internet*. New York: Simon & Schuster, 1995.

[3] d. boyd, *Privacy and Publicity in the Context of Big Data.* WWW, Raleigh, North Carolina, 2010.

[4] d. boyd, *Taken Out of Context: American Teen Sociality in Networked Publics*, Ph.D. dissertation, University of California-Berkeley, School of Information, 2008.

[5] H. Garfinkel, *Studies in ethnomethodology*. Englewood Cliffs, N.J.: Prentice-Hall, 1967.

[6] J. Meyrowitz, *No sense of place: the impact of electronic media on social behavior*. New York: Oxford University Press, 1985.

[7] N. Ellison, R. Heino, and J. Gibbs, *Managing Impressions Online: Self-Presentation Processes in the Online Dating Environment*, *Journal of Computer-Mediated Communication*, vol. 11, no. 2, Jan. 2006.

[8] R. L. Breiger, K. M. Carley, and P. Pattison, *Dynamic Social Network Modeling and Analysis: workshop summary and papers,* Washington, D.C.: National Academies Press, 2003.

[9] J. Merrill, M. Rockoff, S. Bakken, and K. Carley, *Organizational network analysis: a method to model information in public health work*, *AMIA Annu Symp Proc*, 2006.

[10] G. Kossinets and D. J. Watts, *Empirical analysis of an evolving social network*, *Science*, vol. 311, no. 5757, Jan 2006.

[11] M. Magnani, D. Montesi, and L. Rossi, *Information propagation analysis in a social network site*, in *ASONAM*. IEEE Computer Society, Los Alamitos, 2010.

[12] ——, *Friendfeed breaking news: Death of a public figure*, in *IEEE SocialCom - International Symposium on Social Computing Application*. IEEE Computer Society, Los Alamitos, 2010.

[13] M. J. Keeling and K. T. D. Eames, *Networks and epidemic models*, *Journal of the Royal Society, Interface / the Royal Society*, vol. 2, no. 4, 2005.

[14] D. Rushkoff, *Media Virus!: Hidden Agendas in Popular Culture*, Ballantine Books (T), 1994.

[15] H. Jenkins, X. Li, A. Krauskopf, and J. Green, *'If It Doesn't Spread, It's Dead : Media Viruses and Memes*, 2010.

[16] d. boyd, G. Scott, and L. Gilad, *Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter*, in *HICSS-43. IEEE: Kauai, HI, January 6.*, 2010.

[17] D. Morley, *The Nationwide audience: structure and decoding,* British Film Institute, 1980.

[18] L. C. Freeman, *Visualizing social networks*, *Journal of Social Structure*, vol. 1, no. 1, 2000.

[19] ——, *Centrality in Social Networks Conceptual Clarification*, *Social Networks*, vol. 1, no. 1, 1979.