A New Form of Assortativity in Online Social Networks[☆]

Francesco Buccafurri[§], Gianluca Lax, Antonino Nocera

DIIES, University Mediterranea of Reggio Calabria,
Via Graziella, Località Feo di Vito, 89122 Reggio Calabria, Italy
{*bucca, lax, a.nocera*}*@unirc.it*
[§] *Corresponding Author*

Abstract

The term assortativity indicates the tendency, for a network node, to be directly connected to other nodes that are someway similar. In more technical terms, a given feature is assortative in a network if the probability that an arc exists between two nodes having this feature is greater than the probability that an arc exists between two generic nodes. The role of assortativity in real-world and online social networks has been largely investigated in the literature, in which, starting from degree assortativity, several forms of assortativity have been analyzed. When moving from a single-social-network to a multiple-social-network perspective, new specific traits can be studied, also under the assortativity magnifying glass. This is the case of *membership overlap* among networks (i.e., the fact that people belong to more online social networks) as expression of different traits of users' personality. In this paper, we deal with the above issue, by defining two different measures of membership overlap assortativity, called *Loose* and *Constrained Inter-social-network Assortativity*, respectively and by observing that in two of the most representative online social networks, namely `Facebook` and `Twitter`, membership overlap is assortative.

Keywords: assortativity, assortative mixing, online social networks, membership overlap, `Facebook`, `Twitter`

1. Introduction

In real-world social interactions, individuals tend to associate with similar ones, having common (social or demographic) characteristics, thus favoring *homophilic* relationships (Lazarsfeld and Merton, 1954; McPherson et al., 2001). Moreover, it may happen that individuals act similarly to their social ties due

[☆] A preliminary version of this paper appears in the Proc. of the International Conference on Social Computing and Its Applications (SCA 2013), Karlsruhe, Germany, 2013, IEEE, under the title "Internetworking Assortativity in Facebook" (Buccafurri et al., 2013b).

to some form of mutual influence, often referred as *contagion*. Homophily and contagion, together with *opportunity structures* influencing social tie formation (e.g., spatial proximity, working in the same organization) and *sociality* mechanisms (unlike homophily, independent of the attributes of actors in the dyad) are the main reasons why a real-world social network exhibits *assortative mixing* (Ackland, 2013). Assortative mixing (often called *assortativity*) (Newman, 2002) is an empirical measure describing a positive correlation in the traits and personal attributes of people socially connected with each other, as age, education, socio-economic status, physical appearance, religion, etc. In other words, considering for example socio-economic status, we say that it is assortative in a community if the probability that two people with similar socio-economic status belonging to this community are friends is higher than the probability that randomly selecting two people, they are friends.

While assortativity can be in general empirically observed and there are a number of reasonable ways to measure its level in social networks, it is more difficult, sometimes impossible by means of pure observational studies, to understand why people in a social network are assortatively mixing w.r.t. a given dimension (Shalizi and Thomas, 2011). Indeed, both opportunity structures and sociality mechanisms can mask the real level of homophily. Moreover, when assortativity is detected with respect to a changeable attribute or cultural preference, it becomes very hard to understand whether this characteristic is influencing friendship formation (following the homophilic rule encoded into the old adage "birds of a feather flock together") or, vice versa, it is friendship that influences attitudes and preferences (as effect of social contagion, possibly restricted to the case of imitation).

Despite the difficulty of explaining the exact underlying process, the empirical observation of assortative mixing of a social network has been considered of remarkable importance since many years, with strong interest by sociologists, as it represents the fundamental initial step to understand the phenomenon of friendship formation and social influence in a community. In recent years, the rapid growth of online social networks has reinforced interest in assortativity, moving the center of gravity towards computer science, still keeping the role of sociological aspects always crucial. Moreover, online social networks, with the abundance of embedded information about people, even related to their sentimental state and physical health (Shirazi et al., 2013), are huge living laboratories for studying assortativity. On the other hand, it is not obvious whether assortative mixing, especially that of psychological states (Bollen et al., 2011), takes place also in situations where social ties are not mediated by physical contacts but only by online networking services. Finally, online social networks introduce new specific characteristics (e.g., Likes, reciprocity, etc.) which can be analyzed under the assortativity magnifying glass, to improve our knowledge about how people interpret and metabolize social network tools and the psycho-sociological implications.

For all these reasons, studying for which properties online social networks exhibit assortative mixing is an important issue in social network analysis. As a matter of fact, *degree-degree* Newman (2002), *BC-BC* (where *BC* stands for

betweenness centrality) (Goh et al., 2003), and *happiness* assortativity (Bliss et al., 2012; Bollen et al., 2011) are types of assortativity already studied in the context of online social networks. Data extracted from an online social network, such as `Facebook`, `Twitter`, `LiveJournal`, etc., are typically used to characterize it in terms of degree of assortativity (even negative, talking in this case about *disassortativity*) with respect to a given trait, but also to infer general rules concerning social influence in online social networks.

However, to the best of our knowledge, no observation aimed at studying assortative mixing with respect to multi-social-network traits has been provided so far. Indeed, a single user can join multiple social networks, leading to have membership overlap among different social networks. Thus, membership overlap occurs whenever a user belongs to different online social networks. This feature plays an important role in online communities, as it allows the expression of different traits of users' personality (sometimes almost different identities), also enabling, as side effect, the passage of information from one social network to another. Moreover, a recent study has shown that higher levels of membership overlap are positively associated with higher survival rates of online communities (Zhu et al., 2014).

From all the above observations, it clearly follows that studying whether online social networks exhibit assortative mixing with respect to membership overlap is a new, challenging, and important problem. In more technical words, the problem to address is to understand whether two users of a given online social network S are friends in S with higher probability than the generic case if they both belong to other online social networks.

In the present work, we study this issue, concerning explicit membership overlap. Explicit membership overlap occurs when a user shows in the home page of his account in a social network the link to his account in another social network. We introduce two different definitions of assortativity (called *Loose* and *Constrained Inter-social-network Assortativity*, respectively) and measure their value in `Facebook` and `Twitter`, two of the most representative online social networks (Gjoka et al., 2010; Patriquin, 2007; Vasalou et al., 2010). The results obtained in this paper show that both real-life social networks exhibit assortativity according to the Loose and Constrained notions.

A relationship between explicit membership overlap assortativity and implicit membership overlap (i.e., when membership overlap is not declared by the user) is also studied, showing that our assortativity can be related to a form of social behavior which, as side effect, may reduce privacy consisting in keeping separated two accounts in case of implicit membership overlap.

The plan of this paper is as follows: Section 2 presents related literature about assortativity. The reference scenario is illustrated in Section 3. Section 4 presents our assortativity measures. Section 5 describes the experimental campaign carried out on real social networks both to validate the new assortativity measures and to compute the assortativity/disassortativity degree of social networks. Moreover, the interpretation of the results is also discussed. Section 6 illustrates an important implication of membership overlap assortativity in the context of privacy. Finally, in Section 7, we draw our conclusions.

2. Related Work

The concepts of assortativity and degree assortativity have been introduced in the renowned paper of Newman (Newman, 2002). Here, the author defines a measure of connection assortativity for networks and shows that real social networks are often assortative. A further important study concerning social network assortativity has been proposed in (Newman and Park, 2003), in which the relation between clustering and assortativity in the communities composing a social network is investigated. In the wake of (Newman, 2002), Catanzaro et al. (2004b) showed that, while the majority of technological and biological networks appear to be disassortative with respect to the degree, social networks are generally assortative.

A study about the relationship between assortativity and centrality can be found in (Goh et al., 2003). Degree assortativity for co-author networks is studied in (Catanzaro et al., 2004a). Xulvi-Brunet and Sokolov (2005) present two algorithms to change the correlation degree among nodes in a network by keeping unchanged the degree distribution. They show that, although the degree distribution remains unchanged, the variations on assortativity level cause significant changes on several other parameters, such as clustering coefficient, shell structure and percolation. Kossinets (2006) performs some sensitivity analyses, showing that, as for other structural parameters of social networks, assortativity can be dramatically altered by missing data. Ahn et al. (2007) analyze assortativity on Cyworld, MySpace and Orkut. They compute the degree assortativity of these networks and find that online social networks, encouraging activities that cannot be copied in real life, do not show a similar degree correlation pattern to real-life social networks. An opposite behavior is observed for those online social networks handling activities similar to real-life ones. Hu and Wang (2009) study the structural evolution of large online social networks and argue that, with the huge increase of the size of these networks, many network properties, such as density, clustering, heterogeneity, and modularity, show a non-monotone behavior. In (Wilson et al., 2009), the authors found that interaction graphs present a higher assortativity than social graphs and proved their conjectures on `Facebook`. An interesting application of degree assortativity is proposed by Benevenuto et al. (2009) to classify `YouTube` users in spammers, promoters, and legitimates. Johnson et al. (2010) study the relationship between Shannon entropy and degree assortativity, finding that the maximum entropy does not typically correspond to neutral networks but to either assortative or disassortative ones.

The most relevant and recent studies on `Twitter` assortativity have been carried out by Kwak et al. (2010); Bollen et al. (2011); Bliss et al. (2012). The analysis of `Twitter` assortativity (Kwak et al., 2010) showed that users with 1,000 followers or less are likely to be geographically close to their reciprocal-friends and also have similar popularity with them. Bollen et al. (2011) investigate the assortativity of psychological states in `Twitter` and show that assortativity takes place at the level of happiness or subjective well-being. A study on the assortativity of happiness in `Twitter` has been performed by Bliss et al. (2012).

The main result is that average happiness scores of users are correlated with those of their neighbors.

Our paper lies in the wake of the literature about assortativity mentioned above. However, to the best of our knowledge, it represents the first attempt to define assortativity on multiple social networks instead of on single social networks. This paper extends the preliminary study on assortativity appeared in (Buccafurri et al., 2013b). Herein, the Loose and the Strict Internetworking Assortativity have been initially defined and measured for `Facebook`. The additional contributions of this paper can be summarized as follows. First, we measure and study the Loose Inter-social-network Assortativity for `Twitter`. Second, having investigated the limits of the Strict Internetworking Assortativity, we define the Constrained Inter-social-network Assortativity, which is measured and studied for both `Facebook` and `Twitter`. Third, we provide the interpretation of the experimental results that regard behavioral and sociological aspects of social network people. Fourth, we introduce here a theoretical framework more precise than (Buccafurri et al., 2013b), as the null model preserves degree distribution and assortativity of the studied network whereas, in (Buccafurri et al., 2013b), the null model assumes degree uniform distribution and absence of degree assortativity. Finally, we identify an interesting relationship between explicit membership overlap assortative mixing and implicit membership overlap, which discovers the (surprising) result that assortativity may be source of private information leakage, as it can improve the chance of disclosing implicit membership overlap.

3. Reference Scenario

We refer to a (real-life) scenario in which users operate in a multi-social-network environment (Okada et al., 2005; Buccafurri et al., 2014a; Zhu et al., 2014), thus joining multiple social networks. As usual in social network analysis, we model social networks as graphs. To capture the interaction among nodes belonging to different social networks, a special type of edge, namely `me` edge, which interconnects different social networks, is introduced. A `me` edge from a to b indicates that a and b are two accounts (in two different social networks) of the same user.

The resulting graph is called *Multi-Social-Network System* and is defined as follows.

Definition 3.1. A *Multi-Social-Network System (MSNS)* Ω is a directed graph $\langle \mathcal{N}^{set}, \mathcal{E}^{set} \rangle$, where $\mathcal{N}^{set}$ is the set of *nodes*, $\mathcal{E}^{set}$ is the set of *edges* (i.e., ordered pairs of nodes), and $\mathcal{N}^{set}$ is partitioned into subsets each corresponding to a social network. Given a social network S belonging to Ω, we denote by $\mathcal{N}^{set}(S)$ the partition of $\mathcal{N}^{set}$ including the nodes of S and by $N(S)$ its cardinality.

Given $d \geq 0$, we denote by $\mathcal{N}_d^{set}(S) \subseteq \mathcal{N}^{set}(S)$ the set of nodes of S with degree d and by $N_d(S)$ its cardinality. Each social network is said *to belong* to Ω. Given a node $a \in N$, we denote by $S(a)$ the social network which a belongs to. $\mathcal{E}^{set}$ is partitioned into two subsets $\mathcal{E}_f^{set}$ and $\mathcal{E}_m^{set}$.

$\mathcal{E}_f^{set}$ is said the set of *friendship edges* and $\mathcal{E}_m^{set}$ is the set of `me` *edges*. $\mathcal{E}_f^{set}$ is such that for each $(a,b) \in \mathcal{E}_f^{set}$, $S(a) = S(b)$, whereas $\mathcal{E}_m^{set}$ is such that for each $(a,b) \in \mathcal{E}_m^{set}$, $S(a) \neq S(b)$. Given a social network S belonging to Ω, an *i-bridge* b (of S) is a node of S such that there exists a `me` edge (b,x) such that $S(b) = S$.

We say that b is an i-bridge (from $S(b)$) *towards* $S(x)$. Given a node $a \in \mathcal{N}^{set}$, we denote by $\Gamma(a)$ the set of nodes in $S(a)$ such that $(a,b) \in \mathcal{E}_f^{set}$ for each $b \in \Gamma(a)$. $\Gamma(a)$ is said the set of *neighbors* of a. Given two social networks S and T belonging to Ω, we denote by $\mathcal{B}^{set}(S)$ the set of the i-bridges of S and by $\mathcal{B}^{set}(S,T)$ the set of the i-bridges of S towards T. $B(S)$ and $B(S,T)$ denote, respectively, the cardinalities of the sets $\mathcal{B}^{set}(S)$ and $\mathcal{B}^{set}(S,T)$. Given $d \geq 0$, we denote by $\mathcal{B}_d^{set}(S) \subseteq \mathcal{B}^{set}(S)$ the set of i-bridges of S with degree d and by $\mathcal{B}_d^{set}(S,T) \subseteq \mathcal{B}^{set}(S,T)$ the set of i-bridges of S towards T with degree d. Finally, we denote by $B_d(S)$ and $B_d(S,T)$ the cardinalities of the sets $\mathcal{B}_d^{set}(S)$ and $\mathcal{B}_d^{set}(S,T)$.

Each node a of an MSNS represents a user account in $S(a)$. The occurrence of an edge $(a,b) \in \mathcal{E}_f^{set}$ means that b is a friend of a (observe that both a and b belong to the same social network). Moreover, for some social networks, such as `Facebook`, the friendship relation is symmetric, so that the corresponding subgraph has a symmetric edge relation too. An edge $(c,c') \in \mathcal{E}_m^{set}$ means that c and c' are accounts of the same user in two different social networks. As a consequence, c is an i-bridge[1] and a `me` edge exists from $S(c)$ towards $S(c')$.

Example 3.1. Figure 1 shows an example of an MSNS, according to Definition 3.1, involving three social networks, namely `Twitter`, `Facebook`, and `Google+`. For the sake of presentation, sometimes, we shorten "`Facebook`" into "`Fb`", "`Twitter`" into "`Tw`", and "`Google+`" into "`G+`". In this example of MSNS, we have that $\mathcal{N}^{set} = \{n_1, n_2, \ldots, n_{16}\}$, $\mathcal{N}^{set}$(`Fb`)= $\{n_1, n_2, \ldots, n_4\}$, $\mathcal{N}^{set}$(`Tw`)= $\{n_5, n_6, \ldots, n_{12}\}$, $\mathcal{N}^{set}$ (`G+`)= $\{n_{13}, n_{14}, n_{15}, n_{16}\}$, N(`Tw`)= 8, N(`Fb`)= 4, N(`G+`)= 4. An example of `me` edge is the edge (n_7, n_1). This models the fact that n_7 and n_1 are two accounts of the same user in `Twitter` and `Facebook`, respectively. The others `me` edges in $\mathcal{E}_m^{set}$ are (n_6, n_4), (n_8, n_{13}), and (n_{11}, n_{14}). Similarly to (n_7, n_1), the two nodes of each `me` edge denote two accounts of the same user in two different social networks. Friendship edges $\mathcal{E}_f^{set}$ are the remaining edges. In the picture, `me` edges are represented by dashed lines. Gray nodes (i.e., nodes $n_1 - n_4$) represent accounts of `Facebook`, black nodes ($n_5 - n_{12}$) are `Twitter` accounts, whereas the remaining white nodes ($n_{13} - n_{15}$) are `Google+` accounts. Thus, for instance, $S(n_2)$ = `Fb`, $S(n_7)$ = `Tw`, $S(n_{15})$ = `G+`. Observe that `Facebook`, differently from `Twitter` and `Google+`, is a social network with symmetric friendship relation and, therefore, a friendship edge (n_i, n_j) exists

[1] The prefix "i-" stands for "inter-social-networks" and is used to avoid ambiguity with the classic notion of "bridge" (Easley and Kleinberg, 2010). Observe that an i-bridge is a node, whereas a (classic) bridge is an edge.

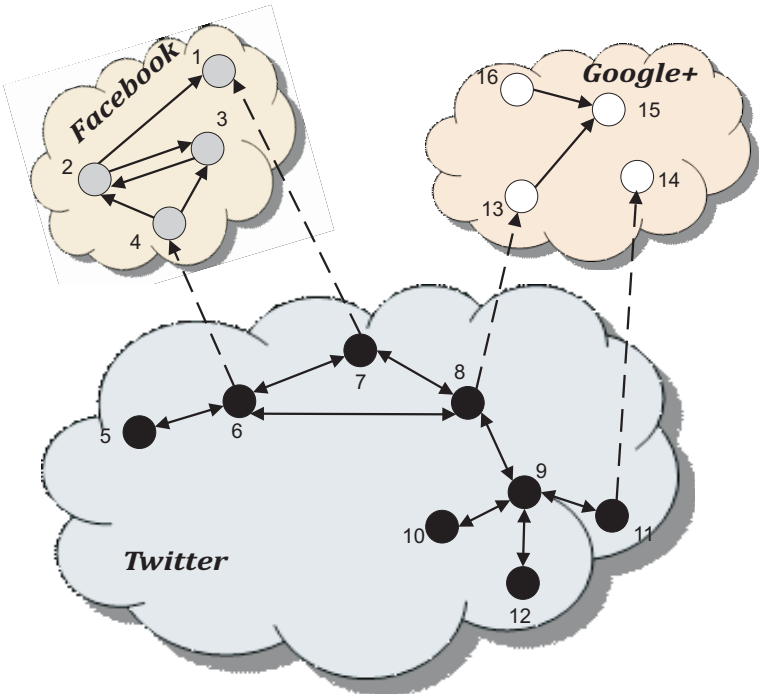

Figure 1: A visual representation of the MSNS of Example 3.1 composed of three social networks.

in `Facebook` if and only if the edge (n_j, n_i) exists too. n_6 is an i-bridge towards `Facebook`, because there exists the `me` edge (n_6, n_4) and n_4 belongs to `Facebook`. The set of i-bridges of `Twitter` is $\mathcal{B}^{set}(\texttt{Fb}) = \{n_6, n_7, n_8, n_{11}\}$, having cardinality $B(\texttt{Tw}) = 4$. The set of the i-bridges of `Twitter` towards `Facebook` is $\mathcal{B}^{set}(\texttt{Tw},\texttt{Fb}) = \{n_6, n_7\}$, with cardinality $B(\texttt{Tw},\texttt{Fb}) = 2$. The set of *neighbors* of n_4 (i.e., $\Gamma(n_4)$) consists of the nodes n_2 and n_3.

From now on, consider given an MSNS Ω.

4. Membership Overlap Assortativity

In this section, we define how to measure explicit membership overlap assortativity. Recall that explicit membership overlap occurs when a user belonging to a social network S_1 declares to have an account also on a second social network S_2. To do this, the user first creates an HTML link to the URL of the account on S_2, and publishes this link in the home page of his account in S_1. According to the model introduced in the previous section, a user with explicit membership overlap is an i-bridge and is the source of a `me` edge.

We measure two forms of membership overlap assortativity, depending on the characteristics analyzed. The first characteristic is just to be an i-bridge

(towards *any* social network). The second characteristic is to be an i-bridge towards the *same* target social network. In particular, we define:

- *Loose Inter-social-network Assortativity* (LIA), as the positive correlation of nodes of an online social network in the characteristic of being an i-bridge.
- *Constrained Inter-social-network Assortativity* (CIA), the positive correlation of nodes of an online social network S in the characteristic of being an i-bridge from S to a given social network T (different from S).

To formally define the above measures, we need some preliminary definitions.

Definition 4.1. Let S and T be two social networks of Ω. We define:

1. The *Loose i-Bridge Friend Fraction* of S as

$$LBF_S = \begin{cases} \frac{\left|\{b \in \mathcal{B}^{set}(S) \mid \Gamma(b) \cap \mathcal{B}^{set}(S) \neq \emptyset\}\right|}{B(S)} & \text{if } B(S) > 0 \\ 0 & \text{otherwise} \end{cases}$$

2. The *Constrained i-Bridge Friend Fraction* of S towards T as

$$CBF_{S,T} = \begin{cases} \frac{\left|\{b \in \mathcal{B}^{set}(S,T) \mid \Gamma(b) \cap \mathcal{B}^{set}(S,T) \neq \emptyset\}\right|}{B(S,T)} & \text{if } B(S,T) > 0 \\ 0 & \text{otherwise} \end{cases}$$

In words, the Loose i-Bridge Friend Fraction of a social network S measures the fraction of the i-bridges of S having at least one friend that is an i-bridge. The Constrained i-Bridge Friend Fraction of a social network S represents the fraction of the i-bridges of S towards T having at least one friend that is an i-bridge towards T too.

Example 4.1. Consider the MSNS represented in Figure 1. In this case, $\mathcal{B}^{set}(\texttt{Tw})$ is the set $\{n_6, n_7, n_8, n_{11}\}$, $\mathcal{B}^{set}(\texttt{Tw}, \texttt{Fb})$ is the set $\{n_6, n_7\}$, whereas $\mathcal{B}^{set}(\texttt{Tw}, \texttt{G+})$ is the set $\{n_8, n_{11}\}$. As a consequence, $LBF_{\texttt{Tw}} = \frac{3}{4}$ because, differently from n_{11}, the nodes n_6, n_7, and n_8 have at least one i-bridge among their friends. $CBF_{\texttt{Tw,Fb}} = 1$ because both n_6 and n_7 have a `me` edge towards `Twitter` and are friends of each other. Vice versa, $CBF_{\texttt{Tw,G+}} = 0$ because neither n_8 nor n_{11} have an i-bridge towards `Google+` among their friends.

Now, we introduce the concept of *null model* of a graph (Newman, 2002; Bayati et al., 2010) adapted to our scenario. This notion provides the theoretical reference necessary to measure the bias of real-life social networks w.r.t. the case in which no membership overlap assortativity exists. This approach is commonly adopted in literature in this context (Holme and Zhao, 2007; Bliss et al., 2012).

Definition 4.2. The *null model* of $\Omega = \langle \mathcal{N}^{set}, \mathcal{E}_f^{set} \cup \mathcal{E}_m^{set} \rangle$ is the random MSNS $\hat{\Omega} = \langle \mathcal{N}^{set}, \mathcal{E}_f^{set} \cup \hat{\mathcal{E}}_m^{set} \rangle$ such that:

$$\mathcal{P}\Big((a,b) \in \hat{\mathcal{E}}_m^{set}\Big) = \frac{\Big|\{(p,q) \in \mathcal{E}_m^{set} \mid p \in S(a), q \in S(b), p \sim a, q \sim b\}\Big|}{\Big|\{(p,q) \mid p \in S(a), q \in S(b), p \sim a, q \sim b\}\Big|}$$

where $\mathcal{P}(X)$ stands for probability of X and $x \sim y$ denotes that $|\Gamma(x)| = |\Gamma(y)|$. Given a social network S in Ω, we denote by $\hat{S}$ the corresponding random social network in $\hat{\Omega}$.

Therefore, the null model of Ω is obtained by keeping the nodes of Ω and its friendship edges and by randomly replacing the source node a and the target node b of any `me` edge with, respectively, a node $a' \in S(a)$ with the same degree as a and a node $b' \in S(b)$ with the same degree as b. This way, we preserve node degree distribution and node degree assortativity, so that the measure of membership overlap assortativity is not affected by the node degree distribution and assortativity of the networks analyzed.

We recall that, social networks show degree assortativity meaning that there is positive correlation between the degree distribution of two nodes at the end of an edge randomly chosen Newman (2002). Degree assortativity is the most common form of assortativity used in network analysis, whereby similarity between nodes is defined in terms of the number of connections the nodes have Piraveenan et al. (2012).

The next step is to compute Loose i-Bridge Friend Fraction (for a given social network S in Ω) and Constrained i-Bridge Friend Fraction (for a given pair of social networks S and T in Ω) in the null model. This allows us to define our measures of assortativity as a difference between the value of Loose i-Bridge Friend Fraction and Constrained i-Bridge Friend Fraction (as defined in Definitions 4.1) observed in the real-life social networks and the theoretical values computed in the null model.

Given a social network S in Ω, its Loose i-Bridge Friend Fraction in the null model $\hat{\Omega}$ is the probability of an i-bridge of $\hat{S}$ to have another i-bridge as a friend. Similarly, given two social networks S and T in Ω, the Constrained i-Bridge Friend Fraction of S toward T in the null model $\hat{\Omega}$ is the probability of an i-bridge of $\hat{S}$ towards $\hat{T}$ to have an i-bridge towards the same social network $\hat{T}$ as a friend.

This is encoded in the following definition.

Definition 4.3. Given two social networks S and T in Ω, we define:

1. The *Loose i-Bridge Friend Fraction* of S in the null model $\hat{\Omega}$ as:

$$\widehat{LBF}_S = \mathcal{P}\left(\Gamma(b) \cap \mathcal{B}^{set}(\hat{S}) \neq \emptyset \mid b \in \mathcal{B}^{set}(\hat{S})\right)$$

2. The *Constrained i-Bridge Friend Fraction* of S towards T in the null model $\hat{\Omega}$ as:

$$\widehat{CBF}_{S,T} = \mathcal{P}\left(\Gamma(b) \cap \mathcal{B}^{set}(\hat{S},\hat{T}) \neq \emptyset \mid b \in \mathcal{B}^{set}(\hat{S},\hat{T})\right)$$

where $\mathcal{P}(X)$ stands for probability of X and, we recall, $\mathcal{B}^{set}(\hat{S})$ is the set of i-bridges of $\hat{S}$ and $\mathcal{B}^{set}(\hat{S},\hat{T})$ is the set of the i-bridges of $\hat{S}$ towards $\hat{T}$ – see Definition 3.1.

The above probabilities can be computed by the following theorem, where we model degree assortativity by increasing by a coefficient e the probability that the ends of an edge chosen at random are nodes with same degree, w.r.t. the case of absence of assortativity.

Theorem 4.1. *Let S and T be two social networks of Ω, u be the maximum degree of i-bridges of S, and e the degree assortativity parameter, defined as the increment of probability of having as neighbor an i-bridge with the same degree w.r.t. the no-degree-assortative case (i.e., $\frac{N_g(S)}{N(S)}$), then:*

$$\widehat{LBF}_S = \sum_{d=1}^{u} \frac{B_d(S)}{B(S)} \cdot \left(1 - \prod_{g=1}^{u} \mathcal{P}\big(N_g(S) - \delta_{gd}, \gamma_{gd}, B_g(S) - \delta_{gd}\big)\right) \quad (1)$$

$$\widehat{CBF}_{S,T} = \sum_{d=1}^{u} \frac{B_d(S,T)}{B(S,T)} \cdot \left(1 - \prod_{g=1}^{u} \mathcal{P}\big(N_g(S) - \delta_{gd}, \gamma_{gd}, B_g(S,T) - \delta_{gd}\big)\right) \quad (2)$$

where:

$$\mathcal{P}(n,a,b) = \prod_{k=0}^{b} \frac{n-a-k}{n-k}$$

$$\gamma_{gd} = \begin{cases} \left(\frac{N_g(S)}{N(S)} + e\right) \cdot d & if \quad g = d \\ \left(\frac{N_g(S)}{N(S)} - \frac{e}{u-1}\right) \cdot d & otherwise \end{cases}$$

$$\delta_{gd} = \begin{cases} 1 & if \quad g = d \\ 0 & otherwise \end{cases}$$

and, we recall from Definition 3.1, $N(S)$ is the number of nodes of S, $B(S)$ is the number of i-bridges of S, $B(S,T)$ is the number of i-bridges of S towards T, $N_d(S)$ is the number of nodes of S with degree d, $B_d(S)$ is the number of i-bridges of S with degree d and, finally, $B_d(S,T)$ is the number of i-bridges of S towards T with degree d.

Proof. First, we prove (1). $\widehat{LBF}_S$ is computed as the sum for any degree $1 \leq d \leq u$ of the product of two factors: (i) the probability of having an i-bridge with degree d (i.e., $\frac{B_d(S)}{B(S)}$) and (ii) the probability that this i-bridge has another i-bridge among its neighborhood. To compute the second term, the d neighbor nodes of the considered i-bridge are partitioned on the basis of

their degree into u sets, one for each possible degree and $\gamma_{gd} \cdot d$ is the size of the g-th set. The probability of not finding an i-bridge inside the g-th set is $\mathcal{P}\big(N_g(S) - \delta_{gd}, \gamma_{gd}, B_g(S) - \delta_{gd}\big)$ where $N_g(S)$ and $B_g(S)$ are the number of nodes with degree g and the number of i-bridges with degree g, respectively. Observe that when $g = d$ the total number of nodes and i-bridges have to be reduced by 1 because we must ignore the starting i-bridge (δ_{gd} plays this role). The function γ_{gd} takes into account the degree assortativity, modeled by e. Specifically, the probability of having as neighbor an i-bridge with the same degree ($\frac{N_g(S)}{N(S)}$) is increased by e and this additional probability is uniformly subtracted from the remaining possible degrees. $\mathcal{P}(n, a, b)$ can be computed by observing that it follows a hypergeometric distribution and, thus:

$$\mathcal{P}(n,a,b) = \frac{\left(\begin{array}{c} b \\ 0 \end{array}\right) \cdot \left(\begin{array}{c} n-b \\ a \end{array}\right)}{\left(\begin{array}{c} n \\ a \end{array}\right)} = \frac{(n-b)!}{a! \cdot (n-b-a)!} \cdot \frac{a! \cdot (n-a)!}{n!} =$$

$$= \frac{(n-b)!}{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-b)!} \cdot \frac{(n-a)!}{(n-a-b)!} = \frac{1}{n \cdot (n-1) \dots (n-b+1)} \cdot$$

$$\cdot \frac{(n-a) \cdot (n-a-1) \dots (n-a-b+1) \cdot (n-a-b)!}{(n-a-b)!} = \prod_{k=0..b} \frac{n-a-k}{n-k}$$

The proof of (1) is thus concluded.

The proof of (2) is obtained as for (1), by considering the set of i-bridges from S towards T instead of the set of all i-bridges. ∎

Now we are ready to give the formal definition of Loose Inter-social-network Assortativity and Constrained Inter-social-network Assortativity.

Definition 4.4. Let S be a social network of Ω. We define the *Loose Inter-social-network Assortativity* of S as:

$$LIA_S = LBF_S - \widehat{LBF}_S$$

LIA measures how much a social network is biased w.r.t. the null model in terms of probability of finding i-bridges among the friends of an i-bridge. CIA can be defined as follows:

Definition 4.5. Let S and T be two social networks of Ω. We define the *Constrained Inter-social-network Assortativity* of S w.r.t. T as:

$$CIA_{S,T} = CBF_{S,T} - \widehat{CBF}_{S,T}$$

Also in this case, this measure gives us an index of how much the behavior of i-bridges is far from the random case. In particular, the higher the value of $CIA_{S,T}$, the higher the correlation among i-bridges from S to T. To measure this form of bias when considering any target social network T, we introduce the following definition.

Definition 4.6. Let S be a social network of Ω. We define the *Constrained Inter-social-network Assortativity* of S as:

$$CIA_S = \frac{\sum_{T \in (\Omega \setminus \{S\})} B(S,T) \cdot CIA_{S,T}}{\sum_{T \in (\Omega \setminus \{S\})} B(S,T)}$$

Intuitively, CIA of a social network measures how much it is biased w.r.t. the null model in terms of probability of finding i-bridges among the friends of an i-bridge, which are coherent in terms of target social network. This is obtained by computing the weighted mean of the Constrained Inter-social-network Assortativity values of S w.r.t. the other social networks, where the weight of $CIA_{S,T}$ is the number of i-bridges from S towards T. Indeed, we expect that the higher the number of i-bridges the more relevant a social network in the computation of CIA. As a matter of fact, our notion of Constrained Inter-social-network Assortativity is conceived to improve Strict Internetworking Assortativity presented in (Buccafurri et al., 2013b), which associates the same weight with each social network. The drawback of Strict Internetworking Assortativity is that this measure is too much susceptible to the measured CIA w.r.t. very marginal and little used social networks.

Example 4.2. Consider again the MSNS represented in Figure 1. To compute $\widehat{LBF}_{\texttt{Tw}}$, we need $B(\texttt{Tw})= 4$, $B_1(\texttt{Tw})= B_2(\texttt{Tw})= 1$, $B_3(\texttt{Tw})= 2$, $N(\texttt{Tw})= 8$, $N_1(\texttt{Tw})= 4$, $N_2(\texttt{Tw})= N_4(\texttt{Tw})= 1$, $N_3(\texttt{Tw})= 2$, $u = 3$. For the sake of simplicity, we assume that this network has no degree assortativity (i.e., $e = 0$). As a consequence, $\widehat{LBF}_{\texttt{Tw}} = \frac{1}{4} \cdot \big(1 - \mathcal{P}(3, 1/2, 0) \cdot \mathcal{P}(1, 1/8, 1) \cdot \mathcal{P}(2, 1/4, 2)\big) + \frac{1}{4} \cdot \big(1 - \mathcal{P}(4, 1, 1) \cdot \mathcal{P}(0, 1/4, 0) \cdot \mathcal{P}(2, 1/2, 2)\big) + \frac{1}{2} \cdot \big(1 - \mathcal{P}(4, 3/2, 1) \cdot \mathcal{P}(1, 3/8, 1) \cdot \mathcal{P}(1, 3/4, 1)\big) \approx \frac{1}{4} \cdot \big(1 - 1 \cdot 0.88 \cdot 0.66\big) + \frac{1}{4} \cdot \big(1 - 0.75 \cdot 1 \cdot 0.38\big) + \frac{1}{2} \cdot \big(1 - 0.63 \cdot 0.63 \cdot 0.25\big) = 0.74$

In words, `Tw` shows no Loose Inter-social-network Assortativity. Indeed, three of four of its i-bridges have another i-bridge among their neighbors, thus $LBF_{\texttt{Tw}} = 0.75$ and $LIA_{\texttt{Fb}} = LBF_{\texttt{Fb}} - \widehat{LBF}_{\texttt{Fb}} = 0.01$. Obviously, as we are dealing with a toy syntectic example, the results about assortativity are meaningless.

5. Experiments

The aim of this section is to test the the proposed assortativity measures on real-life data sets. We studied membership overlap assortativity on `Facebook` and `Twitter`, which are the online social networks with the highest number of users and have attracted the attention of many researchers (see, for instance,

(Gjoka et al., 2010; Patriquin, 2007; Kwak et al., 2010)). According to our theoretical framework, we consider `Facebook` and `Twitter` as part of a Multi-Social-Network System (MSNS) composed of 9 real-life online social networks. In the following sections, we describe the datasets used for our analysis, the null model instance, the results obtained, and, finally, we discuss about the main issues arisen from our study.

5.1. Collected Data and Sample Significance

In our experiments, we consider an MSNS consisting of the following social networks: `Facebook`, `Twitter`, `YouTube`, `LiveJournal`, `Flickr`, `MySpace`, `LinkedIn`, `Google+`, and `VK`, selected among the most popular. To extract relationships between social network accounts (both *friendship edges* and `me` *edges*), we use XFN and FOAF standards. XFN (XHTML Friends Network) (XFN, 2013) uses an attribute, called `rel`, to specify the kind of relationship between two accounts. Possible values of `rel` are `me`, `friend`, `contact`, `co-worker`, and `parent`. FOAF (Friend-Of-A-Friend) (Brickley and Miller, 2013) is a human readable ontology serialized into an XML document encoding human relationships.

We cannot use existing datasets as they do not contain information about `me` edges. For this reason, we extract samples by ourselves. To do this, we cannot rely on a specific crawling technique because, in this case, the way of proceeding of the crawler introduces some biases in the parameter estimation (for instance, a crawler specific for i-bridges, such as BDS (Buccafurri et al., 2012a, 2014b) and SNAKE (Buccafurri et al., 2014c), produces a sample with a fraction of i-bridges higher than the average fraction of i-bridges in the network, thus biasing the estimation of LIA and CIA). Thus, to avoid biases, we uniformly sample the networks of our interest (i.e., `Facebook` and `Twitter`). The uniform sampling of a social network is generally not trivial. However, for `Facebook` and `Twitter`, this activity is facilitated by how user identifiers are organized. Both social networks adopt 64-bit identifiers for user accounts. In particular, the `URL` address of the profile page of a `Facebook` (resp. `Twitter`) user is `http://www.facebook.com/YYY` (resp., `http://twitter.com/account/redirect_by_id?id=YYY`), where `YYY` is a 64-bit numeric identifier. Thanks to this mechanism, to obtain a uniform sample, it suffices to generate numbers uniformly at random in a suitable interval and, for each number, to verify whether it corresponds to an existing account (because an account could have been deleted). If this is the case, we compute the number of its friends, to estimate the average degree of the investigated network (which is one of the parameters of the network null model). Moreover, if the account is an i-bridge, then its first- and second-level neighbors are visited, to obtain the information necessary for the computation of Loose and Constrained Inter-social-network Assortativity. These MSNS samples are clearly centered on the social network to investigate, because they mainly represent its users, their friendship and `me` edges towards the other networks of the MSNS. Observe that, to avoid bias, nodes of the first- and second-level neighbors of the

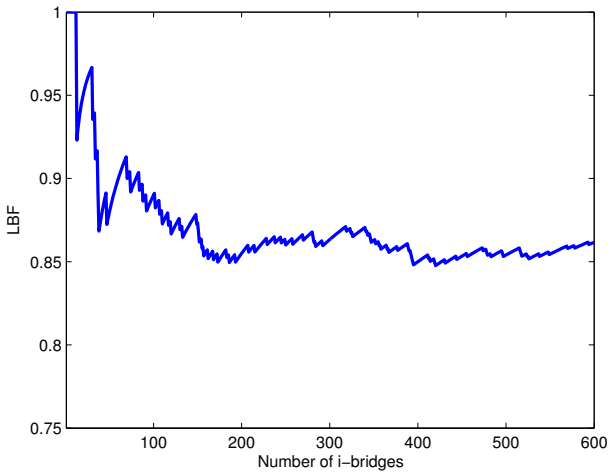

Figure 2: Convergence of LBF versus number of i-bridges.

Number of seen nodes	9,452,867
Number of visited nodes	136,103
Number of i-bridges	1142
Number of edges	9,311,127

Table 1: The characteristics of our dataset.

starting i-bridge at each iteration are not considered in the computation of our assortativity measures and in the estimation of social network parameters.

As it generally happens in sample-based analysis, one of the major problems of the extraction of data for social networks analysis is to obtain a sample with size sufficient to correctly represent the original social network w.r.t. a given parameter. In our case, because we focus on the LIA parameter, we use a convergence-based approach, by iterating the i-bridge population described above until a stable value of LBF is obtained. Figure 2 shows the trend of LBF measured during the data collection. We observe that, after collecting about 100 i-bridges, the value of LBF measured keeps stable in the interval $[0.85, 0.90]$ and converges to about 0.86. As a consequence, we conclude that also a limited number of i-bridges is sufficient to estimate LBF with good precision. However, our experiments have been carried out on a number of i-bridges much higher than the minimum one obtained by the convergence analysis reported above.

The characteristics of the real-life dataset so obtained are presented in Table 1. The dataset can be download at `http://www.infolab.unirc.it/moa.html`.

S	$N(S)$	$B(S)$	d
`Twitter`	$554,750,000$	$5,502,461$	62.9
`Facebook`	$1.06 \cdot 10^9$	$1,253,120$	78.7

Table 2: Number of nodes $N(S)$, number of bridges $B(S)$, and average degree d of the null model build for `Twitter` and `Facebook`.

5.2. Building the Null Model

To compute our measures of assortativity on `Facebook` and `Twitter`, we need to build an instance of the null model defined in 4.2. To do this, we have to estimate the parameters of these social networks as required by Theorem 4.1. Specifically, the number of users of `Facebook` is taken from the annual report of December 2012 (Facebook, 2012), whereas that of `Twitter` is taken from Brain (2013), which are the reports closest to the period of our sampling activity. The number of i-bridges is obtained by multiplying the fraction of i-bridges in our sample by the number of users. As for degree distribution of nodes and i-bridges it is well known that degree distribution in social networks follows a power law (Lu and Wang, 2014; Buccafurri et al., 2013a). Thus, in our model, we used a power law distribution approximating the social network characteristics measured in our sample. In particular, concerning the average degree (d), we compute both that measured for only publicly accessible accounts (as done in (Gjoka et al., 2010)) and the global average degree obtained as the ratio between the number of edges and the number of visited nodes. We set d to the last one because we consider all nodes of our dataset. Also the number of i-bridges towards the diverse social networks has been inferred from that measured in our dataset. Finally, for degree assortativity, we used the value 0.2 as estimated by Ugander et al. (2011). This value is consistent with earlier studies in which it ranges from 0.120 to 0.363 (Newman, 2002, 2003). The values of all these parameters are summarized in Table 2, where $N(S)$ and $B(S)$ denotes the number of nodes and the number of bridges of the network S (Definition 3.1). After the estimation of the null model parameters for `Facebook` and `Twitter`, we are ready to start our experimental campaign.

5.3. Measuring the Membership Overlap Assortativity

In this section, we compute our assortativity measures on `Twitter` and `Facebook`. First, we measure the Loose i-Bridge Friend Fraction LBF as the fraction of i-bridges having an i-bridge in its neighborhood in our sample. Then, we compute $\widehat{LBF}$ by applying Theorem 4.1. Finally, LIA is computed on the basis of Definition 4.4. The values of all these measures are reported in Table 3.

Now, we consider the Constrained Inter-social-network Assortativity. Specifically, towards each social network of the MSNS, we compute the Constrained i-bridge fraction CBF (according to Definition 4.1), and the Constrained i-bridge fraction $\widehat{CBF}$ in the null model (by means of Theorem 4.1). Then, we compute

	LBF	$\widehat{LBF}$	LIA
Twitter	0.861	0.110	0.751
Facebook	0.464	0.042	0.422

Table 3: Values of Loose i-Bridge Friend Fraction (LBF), Loose i-Bridge Friend Fraction of the null model ($\widehat{LBF}$), and Loose Inter-social-network Assortativity (LIA) measured in Twitter and Facebook.

SN	$CBF_{\texttt{Tw,SN}}$	$\widehat{CBF}_{\texttt{Tw,SN}}$	$CIA_{\texttt{Tw,SN}}$	weight
Facebook	0.701	0.024	0.677	0.380
YouTube	0.603	0	0.603	0.157
LiveJournal	0.923	0	0.923	0.015
Flickr	0.683	0	0.683	0.073
MySpace	0.574	0.002	0.572	0.163
LinkedIn	0.882	0	0.882	0.118
Google+	0.280	0	0.280	0.029
VK	0.839	0	0.839	0.065

Table 4: Constrained i-bridge fraction CBF, Constrained i-bridge fraction $\widehat{CBF}$ in the null model, and Constrained Inter-social-network Assortativity (CIA) of Twitter towards the other social networks.

the Constrained Inter-social-network Assortativity CIA w.r.t. the other social networks, as described in Definition 4.5. The results are reported in Table 4 for Twitter and Table 5 for Facebook.

We recall that the Constrained Inter-social-network Assortativity defined in Definition 4.6 is obtained by computing the weighted mean of the *Constrained Inter-social-network Assortativities* of S w.r.t. the other social networks. To this aim, the last column of Tables 4 and 5 reports also the weight of such terms.

This way, we compute $CIA_{\texttt{Tw}} = 0.675$ and $CIA_{\texttt{Fb}} = 0.894$ according to Definition 4.6.

To study the dependency of our results on degree assortativity and node degree distribution included in the null model, we repeated our analysis for different values of degree assortativity and skew of the power law distribution. We observed that the more the skew, the more the membership overlap assortativity. The same happens for degree assortativity, although in a much weaker way. In particular, the lowest values of LIA and CIA (i.e., $LIA_{\texttt{Tw}} = 0.397$, $LIA_{\texttt{Fb}} = 0.407$, $CIA_{\texttt{Tw}} = 0.514$ and $CIA_{\texttt{Fb}} = 0.833$) were obtained in the case of degree assortativity and skew equal to 0. Thus, moving from the simplified null model (adopted in Buccafurri et al. (2013b)), where no degree assortativity is considered and degrees are distributed uniformly (skew equal to 0), towards a realistic null model, allows us to better highlight the strong tendency of real-life social networks to exhibit assortative mixing in membership overlap. An intuitive explanation of this phenomenon is that assuming the uniform de-

SN	$CBF_{\texttt{Fb,SN}}$	$\widehat{CBF}_{\texttt{Fb,SN}}$	$CIA_{\texttt{Fb,SN}}$	weight
Twitter	1	0	1	0.135
YouTube	0.870	0.001	0.869	0.173
LiveJournal	1	0	1	0.023
Flickr	1	0	1	0.090
MySpace	0.859	0.015	0.844	0.534
LinkedIn	1	0	1	0.045
Google+	0	0	0	0
VK	0	0	0	0

Table 5: Constrained i-bridge fraction CBF, Constrained i-bridge fraction $\widehat{CBF}$ in the null model, and Constrained Inter-social-network Assortativity (CIA) of Facebook towards the other social networks.

gree distribution in the null model gives i-bridges more chances to have other i-bridges in the neighborhood w.r.t. the case of skewed distribution, where the most i-bridges have a so low degree that such a chance is negligible. Similar consideration can be done for degree assortativity, as the tendency for high degree i-bridges to mix only with other high degree nodes, reduces the number of potential i-bridge friends. Thus, including in the null model degree assortativity and skewed node degree distribution decreases the (ground) membership overlap assortativity w.r.t. the simplified null model.

5.4. Analysis of the Results and Discussion

Now, we analyze the main results obtained in our experiments, which are summarized in Table 6. To fully understand them, it is worth recalling that even the most assortative networks have an assortativity degree less than 0.4. For instance, in the paper of Newman (Newman, 2002) discussed in Section 2, the most assortative network was the physics coauthorship one, which had an assortativity value equal to 0.363. On the basis of this reasoning, we can conclude that both Twitter and Facebook are highly assortative, as far as membership overlap assortativity is concerned.

In practice, we can state that, given a user of Facebook or Twitter with an account also in another social network (say T), it is very likely that at least one of his friends has an account in another social network (say T') – this is expressed by the high value of LIA – and that T and T' coincide – this is implied by the high value of CIA.

Keeping in mind that this paper does not attempt to separate homophily and contagion (or possible further causes of assortative mixing), as usually done in papers dealing with assortativity (e.g., (Bliss et al., 2012; Bollen et al., 2011)), we try in some cases to give an intuitive explanation of our empirical observations. Anyway, future work could be done to investigate these aspects.

Intuitively, the high LIA could be related to the propensity of people to imitate their acquaintances in the declaration of me edges or, in general to be influenced. As a matter of fact, the declaration of a me edge typically results in an insertion of the logo/url of the target social network in the home page

	LIA	CIA
`Twitter`	0.751	0.675
`Facebook`	0.422	0.894

Table 6: Loose Inter-social-network Assortativity (LIA) and Constrained Inter-social-network Assortativity (CIA) of `Twitter` and `Facebook`.

of the user. Thus, the friends of this user could be enticed to declare their secondary accounts in other social networks too. Conversely, it appears little plausible that the characteristic of having a declared `me` edge in the user's home page can cause a homophilic friendship formation. However, this cannot be completely excluded, because membership overlap could be seen as a trait with social value, which could be related to the perceived level of expertise in the artificial-technological dimension (with respect to which a homophilic behavior could occur).

Consider now the results about CIA. By analyzing the results reported in Table 4 and 5, it is evident that $\widehat{CBF}$ is always very low. Indeed, in this case we focus on a single target social network, so that very few i-bridges are considered in the null model. Therefore, this implies a very low probability of finding i-bridges having as neighbors i-bridges towards the same social network.

Note that, the low value of weight associated with some social networks is due to the low number of their i-bridges. Indeed, these social networks have a low number of users w.r.t. the others and the density of i-bridges is intrinsically low. Clearly, the low number of i-bridges of these social network could lead to a few accurate measure of the actual CIA w.r.t. these specific social networks. However, recall that the final value of the CIA of a social network is computed by weighting (on the basis of the number of i-bridges) the contribution coming from the CIA w.r.t. each social network of the MSNS. As a consequence, the final value measured for CIA is few susceptible to the measure error of the contributions associated with these marginal and little used social networks.

The value of CIA of `Twitter` is lower than that of `Facebook`. This is motivated by considering that `Facebook` allows its users to declare more `me` edges, whereas `Twitter` allows for just one `me` edge to be declared. As a consequence, in `Facebook` the same i-bridge may contribute to the increment in the CIA w.r.t. more than one social network (by contrast, in `Twitter` an i-bridge contributes only one time).

The decrement of the `Twitter` CIA may have several justifications. As for `Flickr`, which is used to share and embed personal pictures, consider that `Facebook` is a "personal" social network (i.e., it is centered on the person), and, thus, it is natural that a user completes his personal profile by a `me` edge to his personal photo collection. By contrast, because `Twitter` is centered on topics, it is less presumable that its members use their unique `me` edge at disposal to link their photo albums. Concerning `MySpace`, the number of its active users has steadily declined since 2008 (Torkjazi et al., 2009). As a consequence, whereas

Facebook users (who have the possibility to declare more me edges) have no problem to keep their existing references to MySpace, Twitter users prefer to point their unique me edge to a more popular social network.

An interesting observation can be drawn by examining the last row of Tables 4 and 5. In fact, there are no i-bridges from Facebook to VK (VKontakte) (its weight is 0), whereas there are i-bridges from Twitter to VK (the weight is higher than 0). This fact could be explained by considering that VK is the "Russian Facebook". As a consequence, it may happen that a user joining Facebook is not motivated to join VK, and vice versa.

Observe that our assortativity measures are not influenced, as might appear from a first analysis, by the popularity of the target social network, because the null model used to compute the bias already takes into account the numerosity of each social network, as the involved random variables have constrained cardinality. For instance, consider the results concerning me edges from Twitter to Facebook. We observe that most of Twitter users (about 70%) declaring a me edge towards Facebook have a friend behaving in the same way. Obviously, a relevant portion of this percentage is not related to the membership overlap correlation, as Facebook is highly attractive in the world of online social networks, and the measure of assortativity must be able to isolate the correlation component. This is what happens in our case, as the value of $\widehat{CBF}_{\texttt{Tw,Fb}}$ (which is measured on the null model) demonstrates that if a Twitter user would choose the target of his me edge in a random way (i.e., with no correlation), then the probability of choosing Facebook is by far the highest one. In this sense, Facebook "dazzles" social network users in their me edge declaration, but the measured CIA is not influenced by this phenomenon.

Finally, note that although the fraction of i-bridges w.r.t. the total number of social network users is low (about 1 account in 250 is an i-bridge), the results obtained in this paper are still significant because they concern a population of almost 7 million of users (because Facebook and Twitter together involved more than 1,6 billions of accounts in 2013).

6. A Privacy Threat related to Membership Overlap Assortativity

In this section, we show that membership overlap assortativity can be used as a form of correlation to improve, as done in statistical attacks, the effectiveness of those techniques able to discover that two accounts belonging to different social networks are associated with the same user. Indeed, for disparate reasons (often related to privacy concerns (Lee et al., 2013)), users do not always make their role of *i-bridge* explicit by specifying their me edges. In this case, we talk about *implicit* membership overlap. In the underlying graph, implicit membership overlap results in a big number of missed me edges. Discovering these edges, also in case of anonymized profiles, may represent an issue with potential (business) benefits for third parties but, obviously, also a serious threat to users' privacy. On the other hand, solving this missing-link-detection problem gives us information about the dynamic evolution of the MSNS, because we may expect that a portion of missing me edges will be inserted in the graph later.

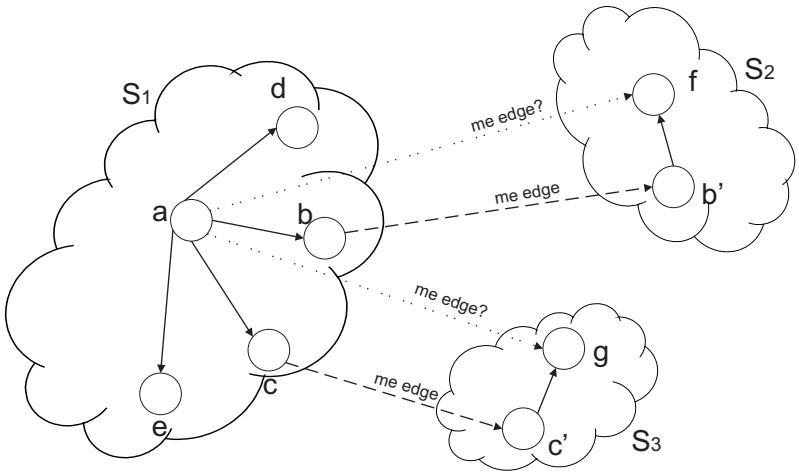

Figure 3: A fragment of an MSNS in which me edges and possible me edges are shown.

This issue has been recently investigated in the literature (Buccafurri et al., 2012b; Narayanan and Shmatikov, 2009). All these solutions, besides lexical similarity between account names, use information coming from the neighborhood to compute the likelihood that two accounts belonging to different social networks are associated with the same user.

In this section, we highlight that a relationship between explicit membership overlap assortativity (studied in this paper) and implicit membership overlap exists, in the sense that assortativity can be related to a form of social behavior which, as side effect, may be source of private information leakage, as it can improve the chance of disclosing implicit membership overlap. In other words, the knowledge acquired in this paper about explicit membership overlap assortativity allows us to increase the effectiveness of those approaches (such as (Buccafurri et al., 2012b; Narayanan and Shmatikov, 2009)) that, on the basis of neighborhood information, discover implicit membership overlap.

Consider the example of Figure 3, showing a fragment of an MSNS composed of three social networks (S_1, S_2, and S_3). The node a is the account in S_1 of the user u, for whom we are looking for other accounts in other social networks of the MSNS.

Observe that, if LIA of $S(a)$ is high, then it is presumable that if a is an i-bridge, then at least one of its neighbors is an i-bridge too. As a consequence, if no friend of a is an i-bridge, then it is expectable that a is not an i-bridge. Otherwise (i.e., a friend of a is an i-bridge), we expect that u is an i-bridge.

Therefore, we start a search to find other accounts of u.

The first contribution of our study in increasing the effectiveness of the above approaches regards the selection of the accounts to be analyzed to discover that/those belonging to a given user. Indeed, it would be time-consuming to test a large number of accounts, but thanks to the results obtained about assortativity, this test can be performed only on a limited set of nodes. In particular, with reference to Figure 3, we can consider first the neighbors of a (i.e., nodes $b \dots e$) and then those nodes having a `me` edge (nodes b and c). The neighbors of the nodes target of the above `me` edges are promising for being other accounts of u.

The second contribution of our study regards the decision about whether a selected account belongs to a given user. In particular, given the account a of the user u, the probability that an account x, belonging to the set of candidates, is associated with u can be biased also on the basis of $CIA_{S(a),S(x)}$. Indeed, a high value of $CIA_{S(a),S(x)}$ increases the above probability due to the fact that $S(x)$ is a "preferential" social network for the i-bridges of $S(a)$.

To be more concrete, consider again the example shown in Figure 3 and suppose that the node a is the account of Mr. John Smith on `Facebook`, which has *johnsmith* as screen name (i.e., the URL associated with the node a is `https://www.facebook.com/johnsmith`), the node g is an account in `MySpace` with screen name *j_smith* (`https://myspace.com/j_smith`) and the node f is an account in `Twitter` with again screen name *j_smith* (`https://twitter.com/j_smith`). In this example, as already said, the nodes f and g are selected as possible accounts of Mr. John Smith. Every discovering techniques operating on string similarity and/or neighborhood information (such as (Buccafurri et al., 2012b; Narayanan and Shmatikov, 2009)) would return the same probability of being account of the same user for the pair (a, f) and the pair (a, g), because f and g have the same screen name and equal neighbors (i.e., no friend). So, on the basis of the above techniques, the probability that f is an account of Mr. Smith is the same as the probability that g is an account of Mr. Smith.

By contrast, thanks to the knowledge about the explicit membership overlap assortativity of the involved social networks, and, in particular, that CIA of `Facebook` towards `Twitter` is much higher than that of `Facebook` towards `MySpace`, we guess that `https://twitter.com/j_smith` is more likely an account of Mr. Smith than `https://myspace.com/j_smith`. Thus, assortativity allowed us to prefer one possibility w.r.t. another one, which the previous approaches considered equivalent.

7. Conclusion

In this paper, we have observed that online social networks exhibit assortativity with respect to explicit membership overlap. To do this, we have provided two measures of assortativity, which captures two different traits with respect to which assortative mixing is studied. According to the approach commonly used in network theory, based on the usage of the null model as a term of comparison, we have verified that both `Facebook` and `Twitter` are assortative w.r.t.

these two measures. This result is very interesting, insofar the past literature has shown that studying assortativity in social networks is per se important but also because the present study does not check whether a given assortative behavior still exists when moving from real-life social networks to online social networks, but it deals with assortativity with regard to a characteristic specific of the online-social-network world. In particular, membership overlap is a crucial trait in this world seen in its globality, where hundreds of online social networks offer different (even opposite) characteristics in which one can recognize and where the passage of information from one social network to another can be conveyed only through membership overlap. Therefore, our result about assortativity w.r.t. membership overlap affects the knowledge on how the information flow crossing two social networks is structured (issue that we plan to analyze in the future). The significance of the present study about assortativity is thus related to the role that the concept of assortativity has in the comprehension of complex networks. For example, it has been realized that assortative networks w.r.t. node degree manifest resilience to node deletion (Newman, 2002). Our form of assortativity leads to a similar result regarding the interconnection between social networks, by showing that points of interconnections are more resilient than the ground-truth case. As an another example, it was recently shown in Ciglan et al. (2013) that degree assortativity has an important role in community detection of real-world networks. It could be thus interesting to investigate whether the new form of assortativity studied in this paper can affect community detection in online social networks when the target is to find communities overlapping between multiple social networks. As further proof of significance of our research we have concluded our study by identifying an interesting relationship between explicit membership overlap assortative mixing and implicit membership overlap, which discovers the surprising result that assortativity may be source of private information leakage, as it can improve the chance of disclosing implicit membership overlap. Keeping in mind that assortativity is only an empirical observation, we tried to explain, through homophily and contagion, why our form of assortativity holds in online social networks. However, this issue merits further analysis involving also sociological aspects and different methodologies, which we plan to deal with in our future research.

Acknowledgment

This work has been partially supported by the Program "Programma Operativo Nazionale Ricerca e Competitività" 2007-2013, Distretto Tecnologico CyberSecurity and project BA2Know (Business Analytics to Know) PON03PE_00001_1, in "Laboratorio in Rete di Service Innovation", both funded by the Italian Ministry of Education, University and Research.

References

Ackland, R., 2013. Web social science: Concepts, data and tools for social scientists in the digital age. Sage.

Ahn, Y., Han, S., Kwak, H., Moon, S., Jeong, H., 2007. Analysis of topological characteristics of huge online social networking services. In: Proc. of the International Conference on World Wide Web (WWW'07). ACM, Banff, Alberta, Canada, pp. 835–844.

Bayati, M., Kim, J. H., Saberi, A., 2010. A sequential algorithm for generating random graphs. Algorithmica 58 (4), 860–910.

Benevenuto, F., Rodrigues, T., Almeida, V., Almeida, J., Gonçalves, M., 2009. Detecting spammers and content promoters in online video social networks. In: Proc. of the International Conference on Research and Development in Information Retrieval (SIGIR '09). ACM, Boston, MA, USA, pp. 620–627.

Bliss, C., Kloumann, I., Harris, K., Danforth, C., Dodds, P., 2012. Twitter reciprocal reply networks exhibit assortativity with respect to happiness. Journal of Computational Science 3 (5), 388–397.

Bollen, J., Gonçalves, B., Ruan, G., Mao, H., 2011. Happiness is assortative in online social networks. Artificial life 17 (3), 237–251.

Brain, S., 2013. Twitter statistics. `http://www.statisticbrain.com/twitter-statistics/`, [Online; accessed December 2013].

Brickley, D., Miller, L., 2013. The Friend of a Friend (FOAF) project. `http://www.foaf-project.org/`.

Buccafurri, F., Foti, V., Lax, G., Nocera, A., Ursino, D., 2013a. Bridge Analysis in a Social Internetworking Scenario. Information Sciences 224, 1–18, elsevier.

Buccafurri, F., Lax, G., Nicolazzo, S., Nocera, A., 2014a. A Model to Support Multi-Social-Network Applications. In: Proc. of the International Conference Ontologies, DataBases, and Applications of Semantics (ODBASE 2014). Springer, Amantea, Italy, pp. 639–656.

Buccafurri, F., Lax, G., Nocera, A., Ursino, D., 2012a. Crawling Social Internetworking Systems. In: Proc. of the International Conference on Advances in Social Analysis and Mining (ASONAM 2012). IEEE Computer Society, Istanbul, Turkey, pp. 505–509.

Buccafurri, F., Lax, G., Nocera, A., Ursino, D., 2012b. Discovering Links among Social Networks. In: Proc. of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2012). Lecture Notes in Computer Science. Springer, Bristol, United Kingdom, pp. 467–482.

Buccafurri, F., Lax, G., Nocera, A., Ursino, D., 2013b. Internetworking assortativity in Facebook. In: Proc. of the International Conference on Social Computing and its Applications (SCA 2013). IEEE Computer Society, Karlsruhe, Germany, pp. 335–341.

Buccafurri, F., Lax, G., Nocera, A., Ursino, D., 2014b. Moving from social networks to social internetworking scenarios: The crawling perspective. Information Sciences 256, 126–137, elsevier.

Buccafurri, F., Lax, G., Nocera, A., Ursino, D., 2014c. A system for extracting structural information from social network accounts. Software: Practice and Experience DOI: 10.1002/spe.2280.

Catanzaro, M., Caldarelli, G., Pietronero, L., 2004a. Assortative model for social networks. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics 70(3), 037101–037104.

Catanzaro, M., Caldarelli, G., Pietronero, L., 2004b. Social network growth with assortative mixing. Physica A: Statistical Mechanics and its Applications 338 (1), 119–124.

Ciglan, M., Laclavík, M., Nørvåg, K., 2013. On community detection in real-world networks and the importance of degree assortativity. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 1007–1015.

Easley, D., Kleinberg, J., 2010. Networks, crowds, and markets. Vol. 8. Cambridge University Press, Cambridge, UK.

Facebook, 2012. 2012 annual report. `https://materials.proxyvote.com/Approved/30303M/20130409/AR_166822/`, [Online; accessed December 2013].

Gjoka, M., Kurant, M., Butts, C., Markopoulou, A., 2010. Walking in Facebook: A case study of unbiased sampling of OSNs. In: Proc. of the International Conference on Computer Communications (INFOCOM'10). IEEE, San Diego, CA, USA, pp. 1–9.

Goh, K., Oh, E., Kahng, B., Kim, D., 2003. Betweenness centrality correlation in social networks. Physical Review E 67 (1), 017101.

Holme, P., Zhao, J., 2007. Exploring the assortativity-clustering space of a network's degree sequence. Physical Review E 75 (4), 046111.

Hu, H. B., Wang, X. F., 2009. Evolution of a large online social network. Physics Letters A 373 (12), 1105–1110.

Johnson, S., Torres, J., Marro, J., Munoz, M., 2010. Entropic origin of disassortativity in complex networks. Physical review letters 104 (10), 108702.

Kossinets, G., 2006. Effects of missing data in social networks. Social networks 28 (3), 247–268.

Kwak, H., Lee, C., Park, H., Moon, S., 2010. What is Twitter, a social network or a news media? In: Proc. of the International Conference on World Wide Web (WWW'10). ACM, Raleigh, NC, USA, pp. 591–600.

Lazarsfeld, P., Merton, R., 1954. Friendship as a social process: A substantive and methodological analysis. Freedom and control in modern society 18 (1), 18–66.

Lee, H., Park, H., Kim, J., 2013. Why do people share their context information on Social Network Services? A qualitative study and an experimental study on users' behavior of balancing perceived benefit and risk. International Journal of Human-Computer Studies 71 (9), 862–877.

Lu, J., Wang, H., 2014. Variance reduction in large graph sampling. Information Processing & Management 50 (3), 476–491.

McPherson, M., Smith-Lovin, L., Cook, J., 2001. Birds of a feather: Homophily in social networks. Annual review of sociology 27, 415–444.

Narayanan, A., Shmatikov, V., 2009. De-anonymizing social networks. In: Proc. of the International IEEE Symposium on Security and Privacy. IEEE Computer Society, Oakland, California, USA, pp. 173–187.

Newman, M., 2002. Assortative mixing in networks. Physical Review Letters 89 (20), 208701.

Newman, M., 2003. Mixing patterns in networks. Physical Review Letters 67 (2), 026126.

Newman, M., Park, J., 2003. Why social networks are different from other types of networks. Physical Review E 68 (3), 036122.

Okada, Y., Masui, K., Kadobayashi, Y., 2005. Proposal of Social Internetworking. In: Proc. of the International Human.Society@Internet Conference (HSI 2005). Lecture Notes in Computer Science, Springer, Asakusa, Tokyo, Japan, pp. 114–124.

Patriquin, A., 2007. Connecting the Social Graph: Member Overlap at OpenSocial and Facebook. Compete. com blog.

Piraveenan, M., Chung, K. S. K., Uddin, S., 2012. Assortativity of links in directed networks. Proceedings of Fundamentals of Computer Science.

Shalizi, C. R., Thomas, A., 2011. Homophily and contagion are generically confounded in observational social network studies. Sociological Methods & Research 40 (2), 211–239.

Shirazi, A., Clawson, J., Hassanpour, Y., Tourian, M., Schmidt, A., Chi, E., Borazio, M., Laerhoven, K. V., 2013. Already up? using mobile phones to track & share sleep behavior. International Journal of Human-Computer Studies 71 (9), 878–888.

Torkjazi, M., Rejaie, R., Willinger, W., 2009. Hot today, gone tomorrow: on the migration of MySpace users. In: Proc. of the ACM Workshop on Online Social Networks. ACM, Barcelona, Spain, pp. 43–48.

Ugander, J., Karrer, B., Backstrom, L., Marlow, C., 2011. The anatomy of the facebook social graph. arXiv preprint arXiv:1111.4503.

Vasalou, A., Joinson, A., Courvoisier, D., 2010. Cultural differences, experience with social networks and the nature of "true commitment" in Facebook. International Journal of Human-Computer Studies 68 (10), 719–728.

Wilson, C., Boe, B., Sala, A., Puttaswamy, K., Zhao, B., 2009. User interactions in social networks and their implications. In: Proc. of the ACM European Conference on Computer systems (EuroSys'09). ACM, Nuremberg, Germany, pp. 205–218.

XFN, 2013. XHTML Friends Network. `http://gmpg.org/xfn`.

Xulvi-Brunet, R., Sokolov, I., 2005. Changing correlations in networks: assortativity and dissortativity. Acta Physica Polonica B 36 (5), 1431–1455.

Zhu, H., Kraut, R. E., Kittur, A., 2014. The impact of membership overlap on the survival of online communities. In: Proceedings of the 32nd annual ACM conference on Human factors in computing systems. ACM, pp. 281–290.