Int. J. Human-Computer Studies 80 (2015) 56–65

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

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

ARTICLE INFO

Article history:
Received 12 March 2014
Received in revised form
4 March 2015
Accepted 27 March 2015
Communicated by Harith Alani
Available online 11 April 2015

Keywords:
Assortativity
Assortative mixing
Online social networks
Membership overlap
Facebook
Twitter

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.

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 similar to their social ties due 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, such as age, education, socio-economic status, physical appearance, and religion. 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

☆This paper has been recommended for acceptance by Harith Alani.

☆☆A preliminary version of this paper appears in the Proceedings 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).

* Corresponding author.
E-mail addresses: bucca@unirc.it (F. Buccafurri), lax@unirc.it (G. Lax), a.nocera@unirc.it (A. Nocera).

http://dx.doi.org/10.1016/j.ijhcs.2015.03.006
1071-5819/

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 and reciprocity) 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`, and `LiveJournal`, 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*) 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) and 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), and Bliss et al. (2012). The analysis of `Twitter` assortativity (Kwak et al., 2010) showed that users with 1000 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