Interest Assortativity in Twitter

Francesco Buccafurri, Gianluca Lax, Serena Nicolazzo and Antonino Nocera
DIIES, University Mediterranea of Reggio Calabria, Via Graziella, Località Feo di Vito, 89122, Reggio Calabria, Italy

Keywords: Assortativity, Social Network Analysis, Twitter.

Abstract: Assortativity is the preference for a person to relate to others who are someway similar. This property has been widely studied in real-life social networks in the past and, more recently, great attention is devoted to study various forms of assortativity also in online social networks, being aware that it does not suffice to apply past scientific results obtained in the domain of real-life social networks. One of the aspects not yet analyzed in online social networks is interest assortativity, that is the preference for people to share the same interest (e.g., sport, music) with their friends. In this paper, we study this form of assortativity on Twitter, one of the most popular online social networks. After the introduction of the background theoretical model, we analyze Twitter, discovering that users clearly show interest assortativity. Beside the theoretical assessment, our result leads to identify a number of interesting possible applications.

1 INTRODUCTION

Assortativity (often called *assortative mixing*) (Newman, 2002) is an empirical measure describing a positive correlation in personal attributes of people socially connected with each other. Traits exhibiting assortative mixing can be various, like age, physical appearance, education, socio-economic status, religion, etc. For example, we say that age is assortative in a given set of people if the probability that two individuals with close age are friends is higher than the probability that two randomly selected individuals are friends. Despite the difficulty of explaining its cause among *homophily* (Lazarsfeld and Merton, 1954; McPherson et al., 2001), *contagion*, *opportunity structures* and *sociality mechanisms* (Ackland, 2013), the empirical observation of assortative mixing in real-life social networks has been considered by sociologists of remarkable importance, as basis to study a community under the point of view of friendship formation and social influence.

From the birth of Facebook, the extraordinary growth of online social networks (Buccafurri et al., 2013; Buccafurri et al., 2014a) has enforced scientists to enlarge their point of view in order to consider online social networks as a specific social phenomenon. Thus, it is important to study the social dynamics of online social networks being aware that they cannot be trivially understood by applying past scientific results obtained in the domain of real-life social networks. Assortativity is one of these phenomenons. Indeed, it is not obvious whether assortativity, especially that of psychological states (Bollen et al., 2011), takes place also in online social networks. In general, the fact that social ties are only mediated by online networking services instead of physical interactions, might influence the behavior of the community. As a matter of fact, a lot of work exist aimed at studying various people traits from the assortativity perspective (Bliss et al., 2012; Bollen et al., 2011; Ciotti et al., 2015; Ahn et al., 2007; Johnson et al., 2010; Benevenuto et al., 2009).

In this paper, we consider a trait which is basilar in social dynamics, that is people's interests. However, to the best of our knowledge, assortativity of this trait has not been studied so far in online social networks.

Online social networks are a good laboratory to study whether (and how much) interests are assortative because in real-life communities interest assortativity is mostly dominated by opportunity structures and sociality mechanisms (for examples, the physical place attended by individuals who share an interest, like a cinema club). Thus, online social networks may help us to better understand whether friends tend to share interests for homophily or contagion reasons.

To study interest assortativity in online social networks, this paper uses an approach based on public figures of Twitter. The choice of Twitter is related to both the goal of trying to have results little affected by physical friendship and the fact that it has been widely used in several heterogeneous application scenarios (Lax et al., 2016; Buccafurri et al., 2014b). Indeed, as

Buccafurri, F., Lax, G., Nicolazzo, S. and Nocera, A.
Interest Assortativity in Twitter.
In *Proceedings of the 12th International Conference on Web Information Systems and Technologies (WEBIST 2016) - Volume 1*, pages 239-246
ISBN: 978-989-758-186-1

shown in (Panek et al., 2013), differently from Twitter, Facebook users tend to add as friend people they know in real life in order to transform latent ties to weak ties (Ellison et al., 2007). On the other hand, (Hargittai and Litt, 2011) highlights that Twitter use is driven primarily by interest for entertainment news, celebrity news, and sports news. This allows us to map the abstract concept of interest (or topic) to the concrete entity of public figure, to the extent that a public figure in a given field, say Gordon Ramsay, acts as a representative of a topic, *haute cuisine* (i.e., high-level cooking), in our example. Thus, we assimilate the followship of a user to the Twitter profile of Gordon Ramsay to the fact that this user is interested in high-level cooking.

The way to study interest assortativity is to observe, for a number of public figures, if the measured probability that two Twitter friends share the followship to the same public figures is higher than the random case (i.e., with no assortativity). The result obtained is that interests are significantly assortative in Twitter. Interestingly, the quantified degree of interest assortativity is much more higher than other forms of assortativity measured in social networks in the past. The knowledge of this form of assortativity may help to understand several aspects of social networks, such as information propagation, identifying influential and susceptible members, network resilience. The contributions of our paper are summarized in the following:

1. we define a new form of assortativity in a social network, the *interest assortativity*, and we study how to measure it;
2. we characterize the *random* graph of the network, commonly denoted as *null model* (Newman, 2002), necessary to quantify the assortativity;
3. we compare the measured value of interest assortativity with other forms of assortativity discovered in the past.
4. we sketch how our result impacts on a number of possible application contexts.

The plan of this paper is as follows. Section 2 presents our assortativity measure. Section 3 describes the experimental campaign carried out on Twitter both to validate the new assortativity measure and to verify whether Twitter shows an assortative/disassortative behavior about interests. In Section 4, we show how the result about interest assortativity reported in this paper may support several applications in the context of social networks. Section 5 deals with literature about assortativity. Finally, in Section 6, we draw our conclusions and discuss possible future work.

2 INTEREST ASSORTATIVITY

In this section, we define the *interest assortativity* of a social network and how this can be measured. First, we model a social network and introduce the notation used in the following.

- An online social network is modelled as a directed graph $G = \langle N, E \rangle$, where N is the set of *nodes* (accounts), and E is the set of edges, i.e., ordered pairs of nodes (representing a relationship between two accounts).
- Given an interest I and a positive integer t, we denote by $N_{t,I}$ the set of the top t indegree nodes followed from other nodes interested in I[1];
- Given a node $n \in N$, we denote by $\Gamma_{out}(n)$ the set $\{n' \in N \ s.t. \ (n,n') \in E\}$, and by $\Gamma_{in}(n)$ the set $\{n' \in N \ s.t. \ (n',n) \in E\}$. $|\Gamma_{out}(n)|$ and $|\Gamma_{in}(n)|$ represent outdegree and indegree of n, respectively.
- The set $\Gamma_{out}(n)$ is said the set of friends (or followings) of n, whereas the set $\Gamma_{in}(n)$ is said the set of followers of n.
- Given $d \geq 0$, we denote by $N_d \subseteq N$ the set of nodes with outdegree d.

We are now ready to introduce interest assortativity. Interest assortativity occurs when the probability that two friends share the same interest is higher than that observed in a network in which friendship edges are set in a random way. As common in this context (Holme and Zhao, 2007; Bliss et al., 2012), assortativity is quantified as the difference between the measure of the studied trait in the observed network and that computed in the corresponding *random* graph of the network (Newman, 2002). Therefore, we need to measure these two quantities. Let us start from the observed network.

Definition 2.1. *Given a network G, an interest I and a positive integer t, we define the Interest Friend Fraction towards the top t nodes of I in G as*

$$IFF_{t,I} = \frac{|L|}{\left|\bigcup_{n \in N_{t,I}} \Gamma_{in}(n)\right|}$$

where $L = \Big|\{a \in N \ s.t. \ \exists\, b \in \Gamma_{out}(a), \ n_1, n_2 \in N_{t,I} \ \wedge \ a \in \Gamma(n_1) \ \wedge \ b \in \Gamma(n_2)\}\Big|$.

In this equation, the numerator is the cardinality of the set L composed of the nodes a such that (1) have at least another node b in their neighborhood (i.e., $b \in \Gamma_{out}(a)$), (2) are followers of a node n_1 in

[1] In Section 3, we will clarify how to find this set.

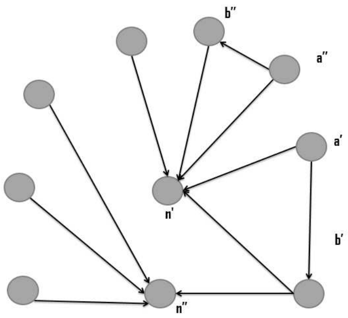

Figure 1: *Interest Friend Fraction* computation.

the top t indegree nodes in I (i.e., $n_1 \in N_{t,I}$) and (3) b is follower of a node n_2 in the top t indegree nodes in I; the denominator is the cardinality of the set of the followers of any node n in the top t indegree nodes in I. Observe that, we use the set $N_{t,I}$ to assimilate the abstract concept of interest (or topic) to the concrete entity of public figures, seen as representatives of interests. Broadly speaking, the *Interest Friend Fraction* measures the fraction of the nodes interested in I having at least one friend interested in I too. In Figure 1, it is represented a network with 10 nodes in which the arrows show the direction of the following relationships among nodes. Assuming that n' and n'' are the top 2 nodes in the interest *music* (thus, they belong to $N_{2,music}$), then the Interest Friend Fraction towards the top 2 nodes of *music* $IFF_{2,music} = \frac{2}{10} = 0,2$, because two nodes (i.e., a' and a'') have as friends nodes following nodes in $N_{2,music}$ (i.e., b' and b'').

After measuring the phenomenon in the observed network by IFF, we need to characterize the *random* graph of the network, commonly denoted as *null model* (Newman, 2002). This graph models the case in which no assortativity occurs.

Definition 2.2. *Given a network G, the* null model *of G is the random graph $\hat{G} = \langle N, \hat{E} \rangle$ such that for each $v \in N$, it holds that:*

1) $\left|\{(x,w) \in E \ s.t. \ x = v\}\right| = \left|\{(x,w) \in \hat{E} \ s.t. \ x = v\}\right|$

and

2) $\left|\{(w,x) \in E \ s.t. \ x = v\}\right| = \left|\{(w,x) \in \hat{E} \ s.t. \ x = v\}\right|$.

In words, it is obtained by maintaining the nodes of G and by replacing the deterministic occurrence of edges by a random variable in such a way that indegree (condition 1) and outdegree (condition 2) of nodes are maintained. Now, we define how to measure IFF in the null model.

Definition 2.3. *Given a network G, an interest I, and a positive integer t, we define the* Interest Friend Fraction *towards the top t nodes of I in $\hat{G}$ as:*

$$\widehat{IFF_{t,I}} = \mathcal{P}\Big(y \in F \ s.t. \ y \in \Gamma_{out}(x) \wedge \ x \in F\Big)$$

where $\mathcal{P}(X)$ stands for probability of X and the set $F = \bigcup_{n \in N_{t,I}} \Gamma_{in}(n)$.

In words, $\widehat{IFF_{t,I}}$ is the probability that a node interested in I has at least one friend interested in I too. Clearly, nodes interested in I are modelled by the set F, which is composed of all nodes following any of the top t nodes for the interest I. The following theorem allows us to compute this probability.

Theorem 2.1. Given a network G, an interest I, and a positive integer t, then:

$$\widehat{IFF_{t,I}} = \sum_{d=2}^{u} \frac{|N_d|}{|N|} \cdot \left(1 - \gamma\Big(|N| - 2, d-1, |F| - 1\Big)\right)$$

where:

$$\gamma(a,b,c) = \prod_{k=0}^{c} \frac{a-b-k}{a-k} \quad \text{and} \quad F = \bigcup_{n \in N_{t,I}} \Gamma_{in}(n)$$

and N is the number of nodes, N_d is the number of nodes with outdegree d, and $\Gamma_{in}(n)$ is the set of followers of n.

Proof. $\widehat{IFF_{t,I}}$ is obtained as the sum for any degree $2 \leq d \leq u$ of the product of two factors: the probability of having a node with outdegree d, and the probability that this node and at least one of its friend follow a node in F. ∎

In the theorem above, it is necessary to compute $|F|$, which is the number of followers in the null model of any node in $N_{t,I}$. The next theorem allows us to do this.

Theorem 2.2. Given a null model $\widehat{G}$, an interest I, and a positive integer t, then:

$$\left| \bigcup_{n \in N_{t,I}} \Gamma_{in}(n) \right| = |N| \cdot \left(1 - \prod_{n \in N_{t,I}} \left(1 - \frac{|\Gamma_{in}(n)|}{|N|}\right)\right)$$

Proof. $\frac{|\Gamma_{in}(n)|}{|N|}$ is the probability to be follower of the node n, which has indegree $|\Gamma_{in}(n)|$, whereas 1 minus this value is the probability of not following it. In the null model, the probability to follow a node n' is independent of that to follow a node n'', so that the products $\prod_{n \in N_{t,I}} \left(1 - \frac{|\Gamma_{in}(n)|}{|N|}\right)$ is the probability to not follow any node in $N_{t,i}$. Thus, 1 minus this value is the probability to follow any node in $N_{t,i}$. Finally, this probability multiplied by the overall number of nodes in the network is the searched value. ∎

Now we are ready to give the formal definition of Interest Assortativity.

Definition 2.4. *Given a network G, an interest I, and a positive integer t, we define the* Interest Assortativity *towards the top t nodes of I in G as:*

$$IA_{t,I} = IFF_{t,I} - \widehat{IFF_{t,I}}$$

This measure gives us an index of how much a social network is biased w.r.t. the null model in terms of probability of finding friends sharing the same interest. The higher the value of assortativity, the higher the correlation in being friends and sharing the same interest.

We conclude this section by showing how to compute the interest assortativity towards the top 2 nodes of *music* for the network depicted in Figure 1. We have already computed $IFF_{2,music} = 0.2$. Concerning the null model of this network, by using Theorem 2.2 we obtain that $|F| = 6$ in the null model (i.e., there are 6 nodes following n' or n'' therein). Concerning the degree, we have 2 nodes, n' and n'', with outdegree 0 (thus, $N_0 = 2$), 3 nodes a', a'' and b'', with outdegree 2 (thus, $N_2 = 3$), and the remaining nodes with outdegree 1 (thus, $N_1 = 5$). Consequently,

$$\widehat{IFF_{2,music}} = \frac{|N_2|}{|N|} \cdot \left(1 - \gamma\left(8,1,5\right)\right) = \frac{3}{10}\left(1 - \frac{3}{8}\right) = 0.1875.$$

Finally, we have that:

$$IA_{2,music} = IFF_{2,music} - \widehat{IFF_{2,music}} = 0.2 - 0.1875 = 0.0125.$$

This result indicates that the network in Figure 1 does not show interest assortativity, as the measured value is very close to zero. From a qualitative point of view, this means that the particular configuration of the network, with only two nodes (i.e., a' and a'') whose friends (i.e., b' and b'') are interested in music, is a result that we could obtain with a high probability also by randomly setting the edges of the network. Therefore, no relation between node friendship and interest in music exists in this network.

3 EXPERIMENTS

This section aims at describing the tests carried out to measure interest assortativity on real-life data. For this purpose, we choose Twitter as target social network both because it is one of the most popular and studied social sites (Gjoka et al., 2010; Patriquin, 2007; Kwak et al., 2010) and for some of its features which fit well with our definitions. Indeed, one of the major advantage of Twitter is that all accounts are publicly available and accessible through a set of APIs[2]. Moreover, because it is focused on the diffusion of information embedded inside short messages (Tweets), the notion of public figures (i.e., celebrities) is very prominent as people often use Twitter to get news from their favorite celebrities. As mentioned above, we use the "following" relationships (see Definition 2.1) between users and public figures as an explicit declaration of interest to the field to which the celebrities belong to.

Table 1: Dataset characteristics.

	# Vip	indegree (Min-Max)	# Checks
Music	30	18.038.914-65.389.090	6.435.561
Sport	30	6.850.604-33.686.429	8.207.264
Cinema	30	7.262.979-26.732.512	10.581.658

In our experiments, we considered the three categories *Music*, *Cinema* and *Sport* as interests and we identified the t most followed celebrities on Twitter for each category. The value t is chosen by applying as a criterium that the number of followers of the top t chosen public figures is of the same magnitude order as the number of Twitter users. We found out that $t = 30$ satisfies the above criterium for the chosen interests. Consequently, we built three sets each including 30 celebrities, which corresponds to the sets $N_{t,I}$ introduced in Section 2). For each celebrity, we extracted its first- and second-level neighbors to obtain information necessary for the computation of our assortativity measure.

The characteristics of the analyzed real-life dataset are presented in Table 1. The second column reports the number of followers of the 30^{th} and 1^{st} celebrity (min and max, resp.), whereas the last column reports how many of such followers have been randomly selected and analyzed in the computation of the interest assortativity. It is worth noting that the number of visited nodes (seemingly small compared with the actual size of the Twitter network) does not limit the validity of the results, because (1) such users are randomly selected and (2) the value of assortativity measured for each of them is extremely high (as we will show in the following). Moreover, the cardinality of our dataset is consistent with that of other researches addressing the same topic (see, for example, the datasets used in (Bliss et al., 2012)).

To compute our assortativity measure for each cat-

[2] The API documentation is available at https://dev.twitter.com/overview/documentation.

egory, we need to obtain $IFF_{t,I}$ and $\widehat{IFF_{t,I}}$ (see Definition 2.4). These computations proceed as follows. Given a public figure n in the top t nodes of a given category, we consider the nodes who are followers of n (i.e., the first-level neighbors of n). Then, we analyze the neighbors of these nodes (i.e., the second-level neighbors of n) and verify whether they are followers of any of the top t nodes.

To compute $\widehat{IFF_{t,I}}$ for each category defined above, we need to build three instances of the null model according to Definition 2.2. To do this, we have to compute the parameters required by Theorem 2.1 and Theorem 2.2. Specifically, the number of users of Twitter is set to 645,750,000 according to the annual report of 2015[3]. Concerning the degree distribution of nodes, it follows a power law as proved in (Lu and Wang, 2014). For this reason, we used this kind of distribution to approximate the social network characteristics measured in our sample and we found that the average degree d of Twitter is about 63.

In our experiments, we randomly selected a large number of Twitter accounts following a given celebrity and checked if such an account shows interest assortativity, measured according to Definition 2.4.

In Table 2, we report the values of interest assortativity on Twitter measured for each category at the end of the experiment.

Table 2: The computed values of IA for the three categories.

Category	IFF	$\widehat{IFF}$	IA
Music	1.000	0.263	0.737
Cinema	0.988	0.177	0.811
Sport	0.973	0.197	0.776

To fully understand the results of our experiments, it is worth recalling that values of assortativity greater than 0.3 means that the phenomenon observed in the social network is very prominent. Indeed, in the first paper on assortativity (Newman, 2002), which will be discussed in Section 5, the most assortative network was the physics coauthorship one, with an assortativity value equal to 0.363 (see Table 3). Under this reasoning and by considering our results, we can state that Twitter is highly assortative w.r.t. the considered trait (i.e., interests). Moreover, the comparison between the three categories (i.e., the three different topics) shows very little differences among them. This means that, the assortative behavior of Twitter users is not bound to a single and well-defined interest, but the general trend is uniform w.r.t. different topics.

[3] Available at http://twittercounter.com/pages/100.

Table 3: Degree assortativity measured on a number of different real-world networks.

Network	Assortativity
physics coauthorship (Newman, 2002)	0.363
biology coauthorship (Newman, 2002)	0.127
mathematics coauthorship (Newman, 2002)	0.120
film actor collaborations (Newman, 2002)	0.208
company directors (Newman, 2002)	0.276
SCN (Zhang et al., 2012)	0.161
CA-HepTh (Zhang et al., 2012)	0.268
CA-GrOc (Zhang et al., 2012)	0.659

4 APPLICATIONS

In this section, we show how the result about interest assortativity reported in this paper can be qualitatively related to some applications. A quantitative study is left as future work. In our study, we used public figures to characterize interests, and showed that the probability that two friends follow the same interest (i.e., the same public figure) is (much) higher than the probability that a pair of users randomly chosen do this. This means that interests are assortative in the social network.

Preliminarily, observe that the applications of interest assortativity may rely also on the fact that our notion (as it usually happens for assortativity measures) is not simply on/off, but gives us a measure of the degree of correlation between sharing of interests and being friends of a given network/subnetwork. Moreover, observe that even though the aim of this paper is to compute interest assortativity of Twitter, the same method can be used to compute interest assortativity of another social network or that of a meaningful subnetwork of a given social network – for example the ego-network (Leskovec and Mcauley, 2012) of a give user. The applications sketched in this section are thought considering that the above extensions are easily feasible.

Consider this first application. Suppose we would like to propagate some information in a social network (say Twitter) with the target of reaching the largest set of people sharing a given interest I. In principle, we could use as starters the highest possible number of public figures representing this interest. A reasonable way to do this is to take the top-t public figures (in terms of followers), where t is such that the magnitude order of the involved followers is that of Twitter. Observe that this is just the criterium we have followed to measure the interest assortativity of Twitter w.r.t. I. Denote by IA_I this value. We have to consider that, typically, reaching an agreement with a public figure for this job could be very hard and expensive. As seen earlier, a realistic value for t could be 30 (for the interests we have considered in this paper), which is definitely high for the application we are considering. We should drastically reduce the number of starters. But how to do this? Is a blind way acceptable? Plausibly, our interest assortativity measure could drive the above minimization. Indeed, we could start by considering the top-1 public figure and by measuring the gap existing between IA_I and the value computed by considering only this public figure. We expect that the measured assortativity is lower than IA_I. Then, we could consider the top-2 public figures and iterate the measure, and so on, until an acceptable closeness between the measured assortativity and IA_I is reached. This allows us to reach a number of followers whose friends share the given interest with high probability, thus heuristically increasing the probability of reaching a large community sharing this interest.

A second application could concern the identification of one (or more) expert(s) in a given field, represented by a given interest I. Suppose we have a number of social network (say Twitter) profiles representing possible candidates to play the role of experts on I. We can argue that the degree of expertise of this user about the topic I is positively correlated to the degree of interest assortativity of her/his egonetwork. Indeed, the higher the assortativity, the higher the probability that people in her/his egonetwork are directly connected by a friendship relation. Thus, the higher the exchange of information regarding the topic I involving (directly or indirectly) U.

Finally, another possible application of our results is related to the resilience of the social network w.r.t. information diffusion. Indeed, public figures are typically source of information: news, opinions, comments, links, etc. Suppose now that the public figure P tweets a piece of information I. Obviously, the full success of the diffusion occurs whenever all the followers re-tweets I (and, in turn, the same happens for their followers). Now, it is well known that the social network universe can be thought as a set of communities, with inner strong correlations (strong ties) weakly connected by means of weak ties. What is important is that the information I reaches at least one element (or a few elements) of the community, as the high communication activity among members of the community will allow the whole community to be reached by I (eventually). Now, the results we have here demonstrated about interest assortativity tell us that in case a given follower U of P receives the information I but fails in re-tweetting it, there is a considerable probability that some users at distance 1 from U are also followers of P, thus recipients of the information I. Therefore, the probability that one of these users re-tweets I (thus, propagating it) is higher than the case of absence of interest assortativity. In other words, interest assortativity increases the resilience of the network w.r.t. its capability of propagating information, in case of failure of some nodes.

5 RELATED WORK

Newman was the first author to introduce a formal definition and a metric for the concept of assortativity. In particular, in (Newman, 2002) he demonstrates that social networks are often assortatively mixed, in the sense that the nodes in the network having many relationships tend to be connected to other nodes highly connected themselves. Starting from (Newman, 2002) further studies concerning social network assortativity have been proposed, such as (Newman and Park, 2003; Catanzaro et al., 2004; Goh et al., 2003) Specifically, the authors of (Newman and Park, 2003) present a deep analysis about the relation between clustering and assortativity in the communities composing a social network. As result they obtain that these communities are characterized by both high levels of clustering and assortative mixing. By contrast, Catanzaro et al. (Catanzaro et al., 2004) compare technological and biological networks and social networks, showing that, while the former appear, in general, to be disassortative with respect to the degree, social networks are typically assortative. Moreover, in (Goh et al., 2003) a study on the relationship between assortativity and betweenness centrality correlation for scale-free networks is presented. The work proposed in (Hu and Wang, 2009) analyzes the structural evolution of large online social networks and concludes that, with the huge increase of their size, many network properties show a non-monotone behavior. This is the case of density, clustering, heterogeneity, and modularity. Some results focusing on degree assortativity are presented in (Ciotti et al., 2015; Ahn et al., 2007; Johnson et al., 2010; Benevenuto et al., 2009) Ciotti et al. (Ciotti et al., 2015) investigate degree correlations in two online social networks. The major result of their paper is that, while subnetworks characterized by assortative mixing by degree have in general links expressing a positive connotation (i.e., endorsement or trust), networks in which links have a negative connotation, (i.e., disapproval and distrust) are described by disassortative patterns. The authors of (Ahn et al., 2007) compute the degree assortativity of Cyworld, MySpace and Orkut and find that these online social networks do not show a degree correlation pattern similar to that of real-life social networks. This can be due to the fact that they encourage activities that cannot be copied in real life. Indeed, an opposite behavior is observed for those online social networks handling activities similar to real-life 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) proved that users with 1,000 followers or less are likely to have the same number of followers (that is, the same popularity) of their reciprocal-friends and also to be geographically close to them. An attempt to define assortativity on multiple social networks instead of on single social networks is presented in (Buccafurri et al., 2015). In particular, the authors measure the tendency of users of associating their Facebook and Twitter account with others in different social sites. Moreover, they provide the behavioral and sociological interpretation of the experimental results. Finally, the authors identify an interesting relationship between explicit membership overlap assortative mixing and implicit membership overlap. This led to the discovery that assortativity may be source of private information leakage, as it can improve the chance of disclosing implicit membership overlap. Our perspective is quite different from that of the works already present in literature, because it deals with assortativity in people's interests, that is a basilar trait in social dynamics. We carry out our analysis by relying on users' interests in Twitter and, to the best of our knowledge, this form of assortativity has not yet been studied so far in online social networks.

6 CONCLUSION

In this paper, we have defined a new form of assortativity, called Interest Assortativity and studied it in Twitter. Our analysis has been carried out by measuring the value of interest assortativity for real-life accounts. The approach used to identify interests is based on Twitter public figures: in particular, we have considered the "following" relationships between users and public figures as an explicit declaration of interest towards the field to which the celebrities belong to. The results of our study allow us to state that Twitter is highly assortative in users interests and that there are not significant differences for the three categories of interests we have considered. This means that users behave uniformly w.r.t. different topics. We showed also possible applications of our results related to information propagation and social network resilience.

Future work could extend the analysis by considering other online social networks and by studying from a quantitative point of view the relationship between interest assortativity and social network resilience.

ACKNOWLEDGEMENTS

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., and 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)*, pages 835–844, Banff, Alberta, Canada. ACM.

Benevenuto, F., Rodrigues, T., Almeida, V., Almeida, J., and 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)*, pages 620–627, Boston, MA, USA. ACM.

Bliss, C., Kloumann, I., Harris, K., Danforth, C., and 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., and Mao, H. (2011). Happiness is assortative in online social networks. *Artificial life*, 17(3):237–251.

Buccafurri, F., Lax, G., Nicolazzo, S., and 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)*, pages 639–656, Amantea, Italy. Springer.

Buccafurri, F., Lax, G., Nicolazzo, S., Nocera, A., and Ursino, D. (2013). Measuring Betweennes Centrality in Social Internetworking Scenarios. In *Proc. of International Workshop on Social and Mobile Computing for collaborative environments (SOMOCO'13)*, pages 666–673, Gratz, Austria. Springer Verlag.

Buccafurri, F., Lax, G., Nicolazzo, S., Nocera, A., and Ursino, D. (2014b). Driving Global Team Formation in Social Networks to Obtain Diversity. In *Proc. of the International Conference on Web Engineering (ICWE 2014)*, pages 410–419, Toulouse, France. Springer.

Buccafurri, F., Lax, G., and Nocera, A. (2015). A new form of assortativity in online social networks. *International Journal of Human-Computer Studies*, 80:56–65.

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

Ciotti, V., Bianconi, G., Capocci, A., Colaiori, F., and Panzarasa, P. (2015). Degree correlations in signed social networks. *Physica A: Statistical Mechanics and its Applications*, pages 25–39.

Ellison, N. B., Steinfield, C., and Lampe, C. (2007). The benefits of facebook friends: social capital and college students use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4):1143–1168.

Gjoka, M., Kurant, M., Butts, C., and 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)*, pages 1–9, San Diego, CA, USA. IEEE.

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

Hargittai, E. and Litt, E. (2011). The tweet smell of celebrity success: Explaining variation in twitter adoption among a diverse group of young adults. *New Media & Society*, 13(5):824–842.

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

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

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

Kwak, H., Lee, C., Park, H., and 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)*, pages 591–600, Raleigh, NC, USA. ACM.

Lax, G., Buccafurri, F., Nicolazzo, S., Nocera, A., and Fotia, L. (2016). A new approach for electronic signature. In *Proc. of the International Conference on Information Systems Security and Privacy (ICISSP 16))*, Rome, Italy.

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

Leskovec, J. and Mcauley, J. J. (2012). Learning to discover social circles in ego networks. In *Advances in neural information processing systems*, pages 539–547.

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

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

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

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

Panek, E. T., Nardis, Y., and Konrath, S. (2013). Mirror or megaphone?: How relationships between narcissism and social networking site use differ on facebook and twitter. *Computers in Human Behavior*, 29(5):2004–2012.

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

Zhang, G.-Q., Cheng, S.-Q., and Zhang, G.-Q. (2012). A universal assortativity measure for network analysis. *arXiv preprint arXiv:1212.6456*.