Social Networks 59 (2019) 61–76

Contents lists available at ScienceDirect

Social Networks

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

Modeling influence on posting engagement in online social networks: Beyond neighborhood effects

Alon Bartal*, Nava Pliskin, Gilad Ravid

Dept. Industrial Engineering and Management, Ben Gurion University of the Negev, Beer Sheva, 84105, Israel

ARTICLE INFO

Keywords:
Online social network (OSN)
Participation shift
Exposure
Local influence
Global influence
Posting engagement

ABSTRACT

In many online social networks (OSN), Member *A* can observe the online participation of Member *B* and thus may be exposed to a shift in the online participation of *B*. Such exposure is often modeled by a neighbor-to-neighbor spreading process that can lead to local influence. In other words, exposure to participation shifts of Neighbor *B* can exert local influence on *A*'s posting engagement. Since many posts nowadays are public rather than private, exposure to participation shifts of non-neighbors outside *A*'s neighborhood can exert global influence on *A*'s posting engagement. We model influence spread in an OSN by differentiating between exposure to participation shifts of local and global influence sources. The results of applying the proposed model to different temporal dynamic OSNs show evidence of both local and global influence on posting engagement of members. This consideration of global influence in OSNs, beyond neighborhood effects, extends past research that focused mainly on local influence and provides new insights about influence sources in general and viral marketing in particular.

1. Introduction

OSNs are efficient platforms for spreading information (Zhang et al., 2016) and influence (Bakshy et al., 2012). Influence can result from exposure to information about views, attitudes, and beliefs of others (Watts and Dodds, 2007), as well as to information promoted by algorithms (Pariser, 2011). Influence is also closely related to behavior adoption in OSNs (Ugander et al., 2012). Modeling how influence spreads in OSNs is of great interest to, among others, those seeking control of the spread of rumors or expedition of product purchase via viral marketing (Lai et al., 2016; Leskovec et al., 2007a).

Many studies about information spread (e.g., Guille et al., 2013), two-step diffusion (e.g., Yin et al., 2016), or product adoption (e.g., Valente et al., 2015) assume influence spreads in a *local* neighbor-to-neighbor process. Only a few central global influence sources, like TV stations, could expose people to information in the past. An OSN member nowadays may post information online and thus may expose another member, not necessarily a neighbor, to her/his online participation (Bakshy et al., 2012). Such exposure may occur, for example, if the exposed OSN member seeks information online (Riedl et al., 2018). Thus, s/he is exposed to the online participation of another member (Oltulu et al., 2018), whether a neighbor or not, and can observe a participation shift of that other member.

According to Gibson (2003), the term *participation shift* refers to shifting between talking and listening activities only. In this work, a participation shift refers to a more diverse set of OSN activities. For example, in the linear threshold model (Chen et al., 2010) each edge has a weight and a member shifts participation from a non-active to an active state if the sum of incoming edge weights from her/his active neighbors exceeds a threshold.

Two main types of models infer influence in an OSN: structural and non-structural. Structural models (such as Chen et al., 2010; Watts, 2002), infer *local* influence by members who expose adjacent neighbors to their participation shifts. Yet, structural models do not infer global influence by non-adjacent members. Since posting activities in many OSNs are publicly accessible (Lottridge and Bentley, 2018), *global* influence can be attributed to a non-local effect as imitation (Leskovec et al., 2009). Non-structural models consider influence mechanisms that do not account for exposure from one member to another. For example, Yang and Leskovec (2010) developed the linear influence model (LIM) that misses structural social effects by inferring the influence of all network members while ignoring the network topology. It is thus crucial to develop a model that infers the influence of both neighbors and non-neighbors at the same time, while also considering the dynamic network topology.

Local neighbor-to-neighbor influence occurs upon exposure of an

* Corresponding author.
E-mail addresses: bartala@post.bgu.ac.il (A. Bartal), pliskinn@bgu.ac.il (N. Pliskin), rgilad@bgu.ac.il (G. Ravid).

https://doi.org/10.1016/j.socnet.2019.05.005

Available online 14 June 2019

OSN member to the participation shifts of neighbors due to direct edge neighborhood effects. On the other hand, global influence occurs upon exposure beyond neighborhood effects to participation shifts of non-neighbors.

Most influence research is focused on modeling only local neighbor-to-neighbor influence (ignoring non-neighbors), or influence by all network members (ignoring the topology of the network). This study extends existing influence models by directly considering in a quantitative way both local influence of adjacent neighbors and global influence of non-adjacent network members. The results show the prevailing nature of global influence, the different temporal patterns between global and local influence, and the ways influence varies across different online environments.

This study aims to model and predict the spread of influence on the posting engagement of OSN members by differentiating between local and global influence sources. The proposed model represents a member's online posting engagement as a network edge and considers both local and global influence as a result of exposure to other members' participation shifts. The model defines influence as the probability that the exposed member would be influenced to shift participation and adopt a similar behavior by engaging the discussion and writing a post. Measuring direct and non-direct edge exposures separately, promotes understanding of how variation patterns at the individual level contribute to the overall *local* and *global* influence.

The next section sets the theoretical background for developing, in the third section, a model that accounts for the duality of local and global influence. The fourth section elaborates upon the configurations of the developed model. As outlined in the fifth section, several datasets from temporal OSNs that contain collective and individual interactions of their members are used for model demonstration and evaluation. The results of applying different configurations of the proposed model to the temporal datasets are discussed in the sixth section. Finally, the contributions and innovations of this work, especially the differentiation between local and global influence, are covered in the concluding section.

2. Related work

Exposure to participation shift of a member in a dynamic OSN can influence the exposed member to engage in a similar shift (Rogers, 2003; Valente et al., 2015; Wang et al., 2012). Most studies focus on neighbor-to-neighbor influence and utilize network structure (Kempe et al., 2005; Ma et al., 2008; Rogers, 2003; Valente et al., 2015). Where a neighbor-to-neighbor spreading process cannot fully explain the influence patterns, a few studies seek to understand out-of-network external influence beyond the structure of the network such as TV stations (Goel et al., 2015; Myers et al., 2012). Our work differs significantly in directly considering not only the local influence resulting from exposure to participation shifts of adjacent neighbors but also the global influence resulting from exposure to participation shifts of non-adjacent network members.

2.1. Structural influence models

There is abundant research demonstrating that network structure affect member behavior. Marsden and Friedkin (1993) explore how the structure of social relations affects the behavior of network members. Research about neighbor-to-neighbor exposure that leads to influence propagation is vast. The cohesion model (Coleman et al., 1966) uses social proximity to explain how exposure to a participation shift of one member can influence another member to shift participation in a similar way. Threshold diffusion Models (Granovetter, 1978; Kreindler and Young, 2014) address the spread of new technology adoption among network members who are influenced by neighbors. The same logic underlies studies about the influence of opinion leaders (Valente and Davis, 1999). For example, physicians can be influenced to prescribe a new drug when exposed to drug prescriptions of their colleagues through links in social networks of friendship, discussion, or advice (Coleman et al., 1966). The Korean Family Planning study (Rogers and Kincaid, 1981) is another example where personal exposure to participation shifts in adopting a family planning innovation is positively associated with network density. Two-step diffusion models (Goel et al., 2015; Yin et al., 2016) argue that information first flows to opinion leaders, and then propagates to a broader population.

OSNs also exhibit neighbor-to-neighbor exposure that influences members to engage in activities (Bakshy et al., 2012) which often generate influence patterns in a tree-like structure (Cheng et al., 2014; Goel et al., 2015). For example, the Independent Cascade Model (Kleinberg, 2007), allows each member to affiliate only with active or inactive states. The model allows each member a single chance to activate its neighbors. Thus, influence is propagated by causing a member to shift participation from inactive to an active state. Then, influenced members can influence their neighbors to also shift participation.

In marketing, consumers have developed a "resistance" to traditional advertising, such as TV and newspaper ads (Zhang et al., 2016). Therefore, retailers encourage customers to exploit OSNs and expose their neighbors to participation shifts by recommending a product (Leskovec et al., 2007a). The exposed neighbors, in turn, can be influenced to buy that product. Customers who shifted to Hotmail, for example, have exposed others to their participation shift by placing a link to a free Hotmail account in the footer of their emails. This exposure influenced 12 million people to open a new Hotmail account (Jurvetson, 2000). Other well-known examples of OSN members exposing their neighbors to their own participation shifts are: the ALS Ice Bucket Challenge (Vaidya, 2014), Old Spice's "The Man Your Man Could Smell Like" (Kaplan and Haenlein, 2011), and Coca-cola's "Friendly Twist" campaign (Fournier and Avery, 2011). Finally, Twitter members who shifted participation from non-active to an active state by actively spreading a rumor, have exposed an average of 45.6 million of the total 51.2 million members within only eight rounds of communications (Doerr et al., 2012).

Influence often leads to adoption of a behavior such as online participation, a practice, or an opinion (Ugander et al., 2012; Marsden and Friedkin, 1993). For example, network members who shifted participation by adopting a new health behavior influenced the health behavior of their neighbors (Centola, 2010). Rogers (2003) classified adopters into five categories based on openness to adopt and found that people can adopt a new behavior at different times. Sociologists who studied member behavior in OSNs, such as Golder and Donath (2004), found that a unique "signature" can be defined for each social behavior by observing the patterns of member participation (Rossi and Ahmed, 2015).

Most OSN members never engage (i.e., lurkers), a few engage a little, and a small minority account for almost all online engagements (Rafaeli et al., 2004; Nielsen, 2006). For example, Nielsen (2006) found that 90% of members never engage, 9% engage a little, and 1% account for the majority of online engagements. Members who engage in discussions, interact online, or create rich content might thus be considered as highly influential discussion catalysts (Himelboim et al., 2009). However, members are not born influencers; they transition between social roles that describe online behavior (Rossi et al., 2012; Rossi and Ahmed, 2015).

Members with a *Visitor* role consume information, yet contribute nothing to the discussions (Stigers, 2011). A Visitor can become a *Novice* by participating in discussions (Stigers, 2011). For a Novice to become an *Activist*, s/he must often participate by consuming and producing information. An Activist can become a *Leader*, also known as a gatekeeper, by assuming an essential role as a rule-maker or an opinion-maker (Barzilai-Nahon, 2006). A member becomes *Passive* if s/he maintains interest in discussions and other members, consuming more than producing content (Sonnenbichler, 2010). Lastly, a *Troll* wishes to disturb interactions (Cheng et al., 2017; Coles and West, 2016).

Describing member behavior (Borgatti and Everett, 1992), some roles emerge from the structure of the network (Rossi and Ahmed, 2015). Such roles can be identified by considering network centralities like Indegree, Outdegree, Closeness, Betweenness, Eigenvector (Agarwal et al., 2008; Domingos, 2005; Valente and Davis, 1999), or ranks (PageRank, HITS) (Page et al., 1999; Kleinberg 1999). One example is the usage of centrality measures to identify influential members after the 9/11 attacks (Memon et al., 2008). In this terrorist network, members played a variety of roles with particular structural attributes, similar to "brokers" and "gatekeepers" (Barzilai-Nahon, 2006). Moreover, centrality measures were used to identify member importance in the network (White and Smyth, 2003), such as articulation nodes and top-k nodes (Borgatti, 2006). However, other studies found no correlation between the roles of members and their structural positions (Ganley and Lampe, 2009; Ravid and Rafaeli, 2004). The reason might be that online member behavior is dynamic and its richness cannot be captured just by whether a member is highly influential or not. In addition, members frequently change roles (Bartal and Ravid, 2019; Rossi et al., 2012; Rossi and Ahmed, 2015). Therefore, to better capture member behavior in OSNs, it is important to discover based on dynamic structural relations of members a more diverse set of roles that represent groups of members with similar behavior.

Threshold models also use the notion of categorizing members into groups with different thresholds of adoption (Kreindler and Young, 2014). According to threshold models, members are influenced to adopt a behavior or an innovation depending on the number of neighbors who shifted participation via adoption. Centola (2010), for instance, found the number of signals required for a member to shift participation from a non-adaptor of health behavior to an adaptor after exposure to a similar shift of her/his neighbors. Sun et al. (2009) showed that members had shifted participation from being a non-fan of a Facebook page to being a fan after a number of their Facebook friends have done so.

Threshold models were generalized to multiplex networks where a non-active member shifts participation to active if the fraction of active neighbors in any layer exceeds a threshold (Karimi and Holme, 2013; Wang et al., 2016). The Linear Threshold Model (Chen et al., 2010) focuses on the increase in the motivation of a non-active member to shift participation to an active state following a shift in the participation of her/his network neighbors. A temporal version of the linear threshold model, which considered the order of events along with the time window that can influence a member, revealed that burst activities of members could explain influence spread (Karimi and Holme, 2013). Other studies found that integrating a non-Markovian process in the threshold model can improve the description of influence spread (Wang et al., 2016, 2015).

The structural-equivalence model (Burt, 1987) considers a wider exposure than a neighbor-to-neighbor effect and depends on the extent to which two members are connected to the same others. According to this model, members engage in a behavior of similar other members depending on the extent of structural equivalence (Burt, 1987). Hence, members should be able to observe each others' activities and be aware of such equivalence. Based on network structure alone, however, one cannot assume that members who are not directly connected are exposed to each other and can observe equivalence. Thus, member equivalence can be assessed at a sociometric distance of only up to 3-hops (Leenders, 2002).

Structural models indicate influence in one of two ways. One way is by the neighbor-to-neighbor narrow exposure mechanisms, which consider only local influence sources at a distance of 1-hop (e.g., the cohesion model (Coleman et al., 1966)). Another way is by the structural-equivalence model, which considers a wider exposure to influence sources at a distance of at most 3-hops. In any case, structural models do not provide a way to infer influence from more remote network members. Burt (1987) indicates that structural equivalence predicts influence better than does cohesion, while other studies argue the opposite (Coleman et al., 1966; Valente et al., 2015). This contradiction, which is further described in Marsden and Friedkin (1993), is related to the definition of influence and creates a theoretical gap that needs to be addressed.

The effect of influence, while considering exposure to all network members simultaneously, is not fully understood. Combining a structural influence model with local effects, and a structural equivalence model seems to be an attractive, intuitive idea. Yet, as explained in the following two paragraphs, assuming that structurally-equivalent members can observe each other leads to redundancy in addressing influence by both directly-connected members and structurally-equivalent members (Burt, 2009).

OSN members are increasingly exposed to varied opinions, which they value differently (Friedkin and Johnsen, 2011), via posts by editors, via other OSN members (Messing and Westwood, 2014), and via algorithms (Pariser, 2011). Research about the influence of such exposure on a member's opinion (Friedkin and Johnsen, 2011; Friedkin, 1990; Marsden and Friedkin, 1993) shows that members adopt opinions that are equal to the average of their view and adopt the opinion of influential network neighbors (Friedkin and Johnsen, 2011).

Exposure to opinions in OSNs is sometimes limited through "filter bubbles" according to a member's past content consumption (Flaxman et al., 2013), or through "echo chambers" of like-minded members (LaCour, 2013). Echo chambers exist in various forms of online media such as blogs, forums (Edwards, 2013), and social media sites (Quattrociocchi et al., 2016). Garimella et al. (2018), representing opinions on Twitter as content items, measured the echo chamber effect by a shift in member online participation from non-sharing to sharing items that contain similar content.

In the proposed model exposure to a participation shift can influence a member to shift her/his participation by sharing content items over the social network (Bakshy et al., 2015; Halberstam and Knight, 2016) by, for example, writing a post. However, measuring exposures by both a structural model and a non-structural model creates a redundant exposure to participation shifts that leads to an echo chamber effect. For example, consider a structural model were directly-connected members can observe each other. Assuming role homogeneity of network members (McPherson et al., 2001), the "chamber", i.e., the social network around member v_3 (Fig. 1), allows a participation shift of v_3 (e.g., a retweet on Twitter or a share on Facebook) to "echo" to v_1 through v_2, and v_4 as they share it. Hence, v_1 can observe the participation shifts of v_2, and v_4 without knowing that they reflect participation shifts of v_3. Therefore, adding exposures to a global influence source (v_3) results in an overlapping exposure, as the same exposure from v_3 propagates to v_1 via v_2 and v_4. Thus, measuring the exposure of v_1 by both types of models leads to redundant exposure measurement and can result in incorrect influence measurement. This triggers the need to develop a method for measuring non-redundant (effective) exposures.

Albert et al. (2011) found that some social networks are prone to redundancy. For influence to be exerted, one must be exposed to novel non-redundant information, contrary to the homogeneous information provided by neighbors (Burt, 2009). For example, weak ties play a dominant role in the spread of novel information (Bakshy et al., 2012; Granovetter, 1973; Nematzadeh et al., 2014). This triggers the need to develop a model that considers novel information to detect influence.

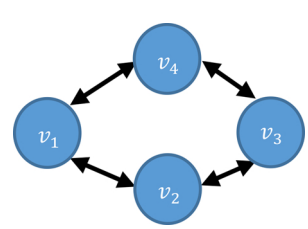

Fig. 1. A network graph illustrating structural equivalence and cohesion.

The structure of OSNs continuously changes, thus creating a temporal network (Rodriguez et al., 2011), which can significantly affect member interaction dynamics, and thus influence patterns. Although numerous studies (e.g. Pastor-Satorras et al., 2015) tried to model the spread of influence using a network structure, simulation results have much space for improvement (Goel et al., 2012). With the growing popularity of OSNs, the mechanisms for influence spread have substantially changed since physical proximity is no longer a constraint for interactions (Kreindler and Young, 2014). In addition, in many OSNs, online participations (e.g., postings) are publicly available, and members can be exposed to participation shifts of non-neighbors beyond neighborhood effect. For example, in Twitter and Reddit members can be exposed to a trending topic that appears on the front/home page. Thus, the need to consider influence beyond network structure arises.

2.2. Non-structural influence models

Non-structural influence models do not rely on the structure of the network. SIR and SIS (Newman, 2003) are two models of epidemiological processes that assume that every member has the same random probability of being infected, i.e., connect to others in a random network. Leskovec et al. (2007b) propose a SIS model where all members have the same probability β to adopt information and members that have adopted the information become susceptible at the next time step. However in social networks, influence is not evenly distributed between all members (Guille et al., 2013) and more likely depends on exposure rates (Myers et al., 2012; Romero et al., 2011). Thus, more complex modeling is proposed in this work, considering an uneven distribution of member exposure to influence sources.

Other studies in OSN, also presented non-structural influence models. For example, Yang and Leskovec (2010) developed the LIM, where the influence functions of members are affected by the overall rate of influenced members in the network. However, LIM assumes the existence of a static network structure and do not identify exposure to influence sources. Wang et al. (2012) developed a logistic model to predict influence by temporal and topological dynamics. However, regarding network structure, they only considered the distance from each member to the influence source. Homophily between network members was also found to affect the influence spreading patterns (Bakshy et al., 2015), and often requires the extraction of personal features of members (Kurka et al., 2015), to apply machine learning algorithms for predicting influence spread patterns (Cheng et al., 2014). This study tests homophily in online engagement as elaborated upon in Section 6.

To summarize, non-structural influence models predict influence spread while ignoring the topology of the network. On the other hand, structural influence models predict which member will influence another neighbor, but do not provide a way to model the influence of non-neighbors. This research identifies and models both local and global influence in a dynamic OSN while considering network structure, thus extending and improving the "classic" network influence models published in the current literature.

2.3. Modeling network dynamics

Exponential Random Graph Models (ERGMs) have been used in recent years to model networks (Hunter et al., 2008). With ERGMs one can model complex processes taking place on static networks, such as the spread of influence (Hu et al., 2014) but, as mentioned before, OSNs are dynamic and change over time. Therefore, the Temporal ERGM (TERGM) was developed (Hanneke et al., 2010) to accommodate intertemporal dependence in time-evolving observed networks (Cranmer et al., 2014; Desmarais and Cranmer, 2012b).

Two approximation methods, whose detailed technical description (Desmarais and Cranmer, 2012b) is beyond the scope of this work, are typically used to estimate the parameters of a TERGM: (1) The Markov chain Monte Carlo maximum likelihood estimation (MCMC-MLE) (Caimo, 2016; Snijders, 2002) method and (2) The Maximum pseudo-likelihood estimation (MPLE) method (Besag, 1974). Contrary to MCMC-MLE, which is based on simulation, the computation of MPLE does not involve simulations. This difference makes MPLE less computationally demanding than MCMC-MLE. MPLE is a consistent estimator since it has been shown to converge in probability to the MLE (Strauss and Ikeda, 1990). Thus, MPLE approaches the MLE in distribution as the size of the analyzed network increases (Leifeld et al., 2015; Strauss and Ikeda, 1990). However, MCMC-MLE has a lower variance than MPLE when implemented with a sufficient number of simulated networks in the MCMC approximation (Van Duijn et al., 2009). A drawback of MCMC-MLE is that it requires a large number of network simulations to guarantee accurate estimations. Since the number of such simulations is unknown in practice, it is difficult to decide whether MCMC-MLE yields more accurate results than MPLE. In addition, MCMC-MLE requires more simulation effort than MPLE as the size of the network increases.

As reviewed so far, research on influence propagation (Fournier and Avery, 2011; Kaplan and Haenlein, 2011; Kreindler and Young, 2014; Vaidya, 2014) in a dynamic OSN (Arnaboldi et al., 2013; Romero et al., 2011) presents explicit relationships between members. Thus, an underlying *activity network* is created, allowing a study of participation shifts (Huberman et al., 2009). Four real datasets of member online interactions are analyzed in this research. Three datasets contain online forums interactions that can be seen as online communities (Gil de Zúñiga and Valenzuela, 2011; Hagel, 1999), and a fourth dataset that describes Twitter micro-blogging interactions.

The four datasets are represented as OSNs where members can be influenced by neighbors with whom they have post-reply relationships but, also, by non-neighbors. Therefore, the relationships between members are implicit and the connections are not well defined as in friendship networks (Gómez et al., 2008). The post-reply relationships between members reflect other reasons for interacting, such as information seeking (Gómez et al., 2008). However, Shi et al. (2009) discovered that the patterns in the influence curves of post-reply relationships are similar to those of stronger relationships in other OSNs.

Next, a local-global influence model is developed. In the model, sources of influence can be local - by exposure to participation shifts of neighbors, and global - by exposure to participation shifts of non-neighbors. An exposure event occurs when a member is exposed to a participation shift of another member. An adoption event occurs when a member influences another member to engage in the discussion by writing a post. The notion of influence resulting from such exposure is defined as the probability of a member to be influenced and engage in posting, similar to De Domenico et al. (2013); Rafaeli et al. (2004).

The model proposed here, which provides multiple pathways in which influence can spread, is the basis for asking the following two Research questions:

a *Does member online posting engagement is influenced by exposure to participation shifts of both local and global influence sources?*
b *Can member online posting engagement be better predicted by exposure to participation shifts of both local and global influence sources rather than only local influence sources?*

3. Model development

Consider the activity network as a graph at time t_k, $G_k = (V_k, E_k, w_k)$. The graph is constructed for each network observation, where nodes (V_k) are *active members* (defined in Section 5), directed edges (E_k) represent interactions between members, and an edge's weight (w_k) indicates the frequency of interactions between time step *k-1* (t_{k-1}) and t_k.

As in similar works (such as Xie et al., 2014; Yang et al., 2013), the online participation of member v_j is measured by a centrality vector $C_{j,k}$, that reflects the participation of v_j at time step t_k. Different indicators can assess the participation of a member in an OSN (Costa et al., 2007)

including, for example, degree, betweenness, and eigenvector centrality (Shi et al., 2009). However, any single indicator is not sufficient to identify multiple and complex characteristics of member participation (Huang et al., 2014). Hence, using a variety of centrality measures can help analyze member online-participation. A *participation shift* of v_j between time steps t_{k-1} and t_k (Δ_{jk}) is measured in this study by the Euclidean (L^2) distance between v_j' s consecutive centrality vectors: $C_{j,k-1}$ and $C_{j,k}$, which reflect her/his network activity at t_{k-1}, t_k. Other distance measures, such as the Cosine distance or the Jaccard distance (Leydesdorff, 2008), are also known to produce similar results in a high-dimensional space (Qian et al., 2004).

In this work, the centrality type in a directed network is prefixed by adding either an "out" or an "in" when calculating the following centralities: Clustering coefficient (CC), Eigenvector centrality, Authority, Hub, PageRank, InKCores, OutKCores, Indegree, Outdegree, Betweenness, InCloseness, OutCloseness. Since some centralities can be highly correlated (Huang et al., 2014), measures with the largest mean absolute correlation above a pre-defined cutoff were removed, as elaborated upon in Section 6. In addition, the centrality vectors are normalized to facilitate working with observations of different sizes (McAuley et al., 2007; Newman, 2003).

Unlike threshold models, which suggest a fixed value of exposure beyond which a member will engage in posting, we follow a different approach guided by exposure curves (Myers et al., 2012). According to this approach, the probability that a member will engage and write a post in the next time step (Rafaeli et al., 2004) depends on the exposure rates. In addition, we argue that a member's posting engagement also depends on personal parameters and preferences (Bakshy et al., 2015; Leenders, 2002) as, for example, affiliation with a role (participation group) which describes her/his participation level in the network, and can also indicate homophily in online posting engagement.

Since the datasets used in this study cover varying time-points, our modeling approach considers their longitudinal character by applying TERGMs. The intensive computational requirement of MCMC-MLE can be overwhelming for a large graph or for a network with multiple observations. Therefore, we use the btergm R package which implements MPLE estimation for TERGMs with bootstrap confidence intervals (Leifeld et al., 2015; Leifeld et al., 2018). The btergm R package uses a similar syntax to that of the ergm package with support functions that are compatible with the statnet suite of packages (Handcock et al., 2016).

3.1. Modeling exposures by local influence sources

When a single member v_j shifts participation, expressed as a structural shift in the network, exposure to this shift by her/his neighbors is propagated over v_j 's outbound edges. Consider a network observation at t_k: if two network neighbors $v_j, v_i \in V_k$ have a directed edge $e_{ji(t_k)} = (v_j, v_{i,i\neq j}) \in E_k$, then member v_j is defined as a local source of influence, exposing member v_i. An edge $e_{ji(t_k)}$ means that v_j posted a message that is referred to v_i at t_k. Hence, v_i is aware of the online participation of v_j and is thus exposed to a participation shift of v_j at t_k by, for example, reading old posts before deciding if or what content to post.

The proposed model follows the notion that a member shares exposures proportional to the edges' strength (Rodriguez et al., 2012). Exposure at t_k to a participation shift of a local influence source, v_j, which is propagated to the targets of its outbound edges, is divided proportionally to v_j' s outbound edges' weights.

The following three steps describe how exposure of v_i to a local influence source v_j at t_k via each outbound edge $e_{ji(t_k)}$ is calculated. (i) Create an adjacency matrix, where each entry represents the weights $w_{ji(t_k)}$ of $e_{ji(t_k)}$; (ii) Normalize edges' weights by dividing each outbound edge weight by the total sum of v_j' s outbound edges' weights $\sum_{v_j}(w_{out(t_k)})$; (iii) Multiply (ii) by the participation shift value of v_j at t_k (Δ_{jk}).

We designate the exposure of v_i to the participation shift of a local influence source v_j at t_k by L_{ij}, as shown in (1).

$$L_{ij}(t_k, \Delta_{jk}) = \frac{w_{ji(t_k)}}{\sum_{v_j}(w_{out(t_k)})}\Delta_{jk} \tag{1}$$

Next, we aggregate v_i' s exposures by all local sources $v_j \in H_k^i$, where H_k^i is the group of v_i' s local influence sources at t_k (2).

$$L_i(t_k) = \sum_{v_j \in H_k^i} L_{ij} \tag{2}$$

Member v_i is exposed to participation shifts of local influence sources as the exposures are directly propagated to her/him. But v_i can also be exposed to participation shifts of global influence sources as discussed next.

3.2. Modeling exposures by global influence sources

Considering member v_i, a global influence source (global source) is defined as a network member: $v_{g,\, g\neq i} \in V_k \backslash H_k^i$. As many online participations are publicly accessible, v_i has a probability of being exposed (3) to a participation shift that a global source v_g shares by its outbound edges at t_k. Note that the outbound edges of v_g are not connected to v_i by definition. Hence, v_i is exposed to the participation shifts of v_g by other mechanisms. P_f is a probability that depends on a function $f(N_k)$ of the number of members, N_k, in the observed activity network at t_k. If v_i has in-links from all members, then $P_f = 0$ since all members are considered as local influence sources.

$$P_f := P(f, t_k) = \begin{cases} f(N_k), & indegree(v_i) < N_k - 1 \\ 0, & indegree(v_i) = N_k - 1 \end{cases} \tag{3}$$

The expression 'exposure to a global/local source' is used hereinafter to abbreviate the expression 'exposure to participation shifts of a global/local influence source'. A member v_i can be exposed to a global source at any point in time and can be exposed to an unlimited number of sources. For example, member v_3 in Fig. 2 is exposed not only to local sources (v_2, v_4) at t_k but also to global sources (v_1, v_5) which can lead to redundant exposure to a participation shift. In Fig. 2, assuming role homogeneity, v_3 has already received the exposures that v_1 sent to v_2 and v_4 through exposures to local sources, thus creating a redundant exposure measurement. However, v_3 lacks the exposure that v_1 sent to v_5. Thus, a correct measuring of the exposure of member v_3 at t_k needs to consider non-redundant (effective) exposures to global sources in addition to exposures to local sources (from v_2, v_4).

As discussed, there is a theoretical gap between structural and non-structural models when trying to predict influence. We suggest a solution that measures effective exposures for each member v_i at t_k. The goal is to consider only effective exposures (informative outbound edges of global sources) by following five steps. First, extract the Boolean adjacency matrix (M_k) of G_k. Each entry in a row vector $\vec{v_g} \in M_k$, represents outbound edges by which a source shares exposure to a participation shift with other members. Second, calculate the transpose matrix, M_k^t. Each entry in a row vector $\vec{v_i} \in M_k^t$ indicate the inbound

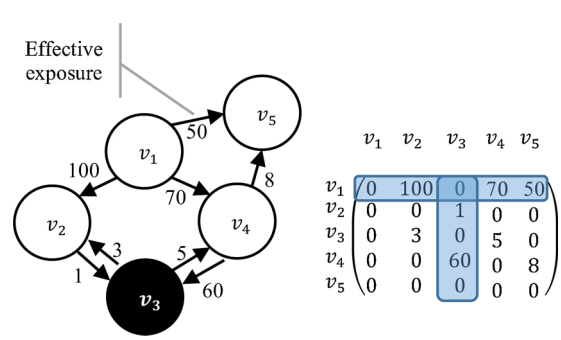

Fig. 2. Illustration of a redundant global exposure of member v_3.

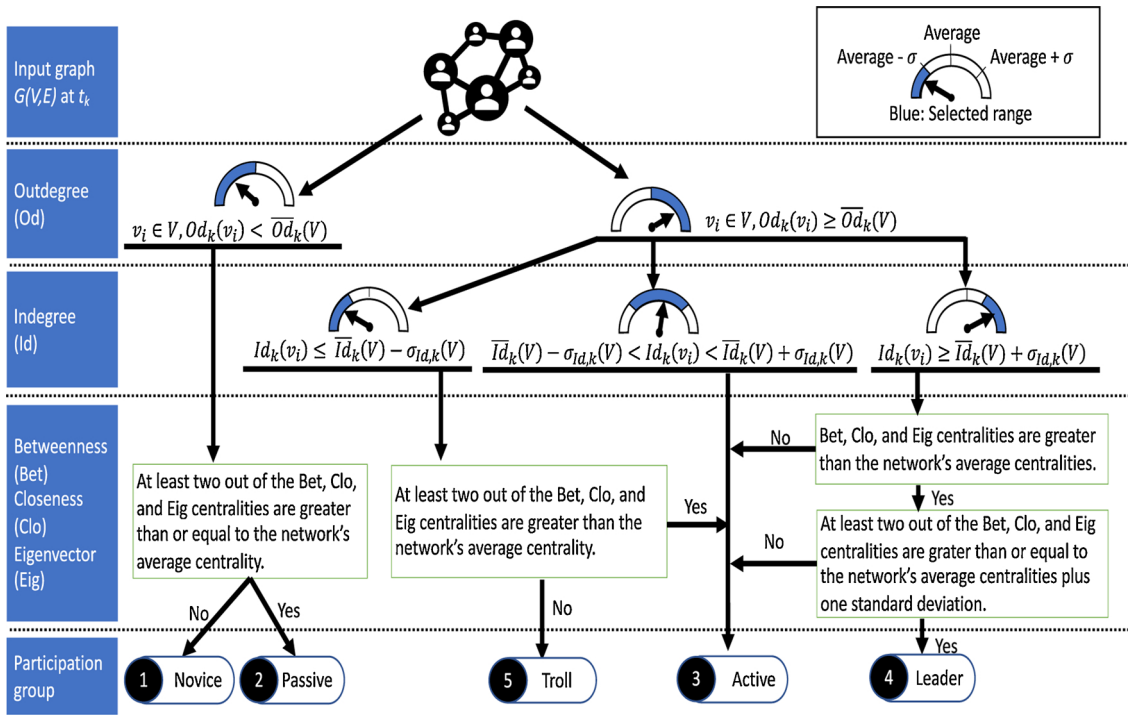

Fig. 3. Group detection sub-model.

edges of $\vec{v_i}$, representing incoming exposures to local sources, as an intermediate step towards finding effective exposures to global sources. Third, compute the intersection between $\vec{v_i}$ and $\vec{v_g}$. Next, subtract the intersection of both vectors from $\vec{v_g}$ which results in a vector $\vec{V}^g_{eff\,(t_k)}$ that contains outbound edges of v_g that represent at t_k only effective exposures from v_g to v_i (4).

$$\vec{V}^g_{eff\,(t_k)} = \vec{v_g} - (\vec{v_i} \bigcap \vec{v_g}) \tag{4}$$

In Fig. 2, v_i is chosen to be v_3, and v_g is chosen to be v_1. The Boolean vector $\vec{V}^g_{eff\,(t_k)}$ contains effective exposures to v_1 (edge e_{15}), which are indicated by 1′s in the corresponding coordinates of $\vec{V}^g_{eff\,(t_k)}$. This step is repeated for each row vector $\vec{v_g}$ from M_k and a row vector $\vec{v_i}$ from M^t_k.

The exposure to participation shifts, which is shared next by outbound edges of v_g at t_k, is divided proportionally to the outbound edge's strength (Rodriguez et al., 2012). In the fourth step, only effective exposures from v_g are summed by the following two sub-steps. In the first sub-step, the outbound edges' weight vector of a global source v_g at time t_k, $\vec{w}_{g\,(t_k)}$ is normalized by the sum of v_g' s outbound edges' weights (5). In the second sub-step, the result of the scalar product of $\vec{\eta}_{g\,(t_k)}$ and $\vec{V}^g_{eff\,(t_k)}$ is multiplied by v_g' s participation shift (Δ_{gk}), resulting in the sum of the proportional effective exposure of v_i to a global source v_g at t_k. Lastly, we multiply the result by the exposure probability to a global source, P_f in (3), yielding the exposure of v_i to the participation shift of v_g at t_k (6).

$$\vec{\eta}_{g\,(t_k)} = \begin{cases} \dfrac{\vec{w}_{g\,(t_k)}}{\Sigma_{v_g}(w_{out\,(t_k)})}, & outdegree(v_g) > 0 \\ 0, & outdegree(v_g) = 0 \end{cases} \tag{5}$$

$$G_{ig}(t_k, P_f, \Delta_{gk}) = (\vec{V}^g_{eff\,(t_k)} \cdot \vec{\eta}_{g\,(t_k)}) \Delta_{gk} P_f \tag{6}$$

Returning to the example in Fig. 2, the exposures of v_3 to participation shifts of global sources (v_1, v_5) are as follows. The global source v_1 shares exposures proportionally by using three outbound edges (e_{12}, e_{14}, e_{15}), but we only consider effective exposures - the outbound edge $e_{15(t_k)}$: $\left(\frac{50}{100+70+50}\right)\Delta_{1k}$. Considering member v_3, the global source v_5 has no outbound edges and, thus, shares zero exposures.

Lastly, in the fifth step, we aggregate the exposures of v_i to all global sources v_g at t_k (7).

$$G_i(t_k) = \sum_{\substack{v_g \in V \setminus H^i_k \\ v_g \neq v_i}} G_{ig} \tag{7}$$

Individuals vary in their willingness to engage and shift participation (Granovetter, 1978; Rogers, 2003; Valente et al., 2015). Thus, members are differentiated here by grouping members with similar network participation level into the same *participation group*, such that at each time step a member can be in any of the groups which are mutually exclusive, as explained next.

3.3. Modeling personal online participation preferences

The group affiliation of each member in each observed network represents common behavior patterns in G_k. These patterns are estimated by a set of features, leveraging Bartal and Ravid (2019). The following centralities derive the features of member participation: Oudegree, Indegree, Betweenness, Closeness, and Eigenvector. The design philosophy for using these centralities is explained in the following paragraphs.

To account for temporal network dynamics, the normalized centrality measures of each member in every OSN observation at t_k are calculated first. Then, the average and standard deviation of the centralities are calculated for each network observation.

To facilitate the definition and detection of a member's *participation group* (Group) at t_k, the group affiliation of each member in every network observation is estimated (Fig. 3). Recall that this study defines influence as the probability that an exposed OSN member will shift participation by engaging the discussion and writing a post. Since the social role of a member can reflect her/his tendency to engage online (Forestier et al., 2012), the sub-model in Fig. 3 facilitates member classification into five participation groups that are represented by roles (Rossi and Ahmed, 2015). Rogers (Rogers, 2003) found five behavioral

Table 1
Member Exposure Probability To a Global Source.

$\#f$	$f(N_k)$	Description
f_1	$\begin{cases} \frac{1}{N_k - \lvert H_k^i \rvert - 1}, & N_k - \lvert H_k^i \rvert > 1 \\ 1, & N_k - \lvert H_k^i \rvert = 1 \\ 0, & N_k - \lvert H_k^i \rvert = 0 \end{cases}$	A member has a uniformly at random exposure probability to a global source at t_k, considering the set of global sources only. N_k: the number of members in the observed network. H_k^i: v_i's local sources at t_k.
f_2	$\frac{1}{N_k - 1}$	A member has a uniformly at random probability of being exposed to a global source at t_k, considering all network members (N_k).
f_3	PageRank	Member exposure probability is proportional to the PageRank value of a global source in the network.
f_4	Gamma correlation	The exposure probability of a member to a global source depends on the extent to which they are structural-equivalent.

groups, as well as Fu et al. (2009) who analyzed the Enron email communication network.

Leaders (Group #4 in Fig. 3) have the highest activity levels, can be considered as influencers (Valente and Davis, 1999; Watts and Dodds, 2007; Will, 2016), and are highly relevant to understanding the diffusion of topics (Romero et al., 2011; Valente and Davis, 1999). They create rich content and promote discussions, which can be measured by high centrality measures (Aghdam and Navimipour, 2016). Therefore, members of Group #4 are likely to have Outdegree values that are greater than or equal to the network's average Outdegree value (Sonnenbichler, 2010). Leaders have many close relationships with other network members and strong ego-networks due to their high activity levels, as can be measured by high Indegree value (Aghdam and Navimipour, 2016). Therefore, Leaders are expected to have Indegree values that are greater than or equal to the network's average Indegree value plus one standard deviation. Since Leaders have high centrality measures, they are expected to have Betweenness, Closeness, and Eigenvector degree centralities that are greater than the network's average centralities (Aghdam and Navimipour, 2016). But Leaders do not always possess the highest centrality measures (Borgatti, 2006). Thus, this expectation is relaxed by allowing at least two out of the three latter values to be greater than or equal to the network's average centralities plus one standard deviation.

Active members (Group #3 in Fig. 3) consume and produce information (Sonnenbichler, 2010), resulting in high Indegree and Outdegree values. They maintain existing relationships and gain new contacts who are mostly within their local network, rather than the whole network (Sonnenbichler, 2010). Therefore, members of Group #3 are less active than members of Group #4, and are expected to have high values of Betweenness, Closeness, and Eigenvector centralities, but lower than those of Leaders.

Trolls (Group #5 in Fig. 3) post offensive content and cause conflict in a relatively short period before vanishing (Coles and West, 2016). Therefore, they have a high activity level at the beginning of their interactions which rapidly decreases to a very low level toward their disappearance for extended periods (Cheng et al., 2017). Thus, compared to the average network values, their Outdegree values are high yet, unaccepted by other members, their Indegree, Betweenness, Closeness, and Eigenvector centralities are expected to be low.

Passive and *Novice* members (respectively Groups #2 and #1 in Fig. 3) have lower activity levels than other members. Passive members are likely to have low Outdegree values since they stay in loose contact with other members. Passive members produce more content than Novices, who mostly observe from the outside. This low activity level by Novices results in a sparse personal network, characterized by low centrality measures. Novice and Passive members have lower Outdegree values than the average Outdegree value for the network. However, since Passive members are more active than Novices, Passive members have at least two out of the Betweenness, Closeness, and Eigenvector centralities that are greater than or equal to the average centrality values for the network, with no restrictions on the value of the third centrality. Novices have the lowest centrality values.

4. Model configurations

This section elaborates upon the different model configurations constructed and analyzed toward the understanding of both local and global influence effects. Above, the developed model measures member exposure to participation shifts of local sources (2) and of global sources (7) at t_k, with a certain probability (3). Focusing on global influence effects and developing a model that measures effective exposures for each member v_i at t_k, four model configurations (*G1* to *G4*) are created here to test v_i' s online engagement resulting from exposures to global sources.

The first model configuration - *G1*, assigns in (7) the participation shift of each global source v_g at t_k (Δ_{gk}) and a probability of exposure (3) to v_g : f_1 (Table 1). Thus, a member has a uniformly at random probability P_{f_1} of being exposed to a global source, considering only the set of global sources at t_k. A possible flaw of P_{f_1}, is in the case of a single global source in the network. Then, the unrealistic exposure probability to that source is $P_{f_1} = 1$. Thus, we defined P_{f_2}: a uniformly at random probability of being exposed to a global source at t_k, while considering all network members (Table 1). Therefore, the second model configuration - *G2*, follows the same steps as *G1* but uses P_{f_2}, instead of P_{f_1}. The third model configuration - *G3*, considers a different scenario where a member can be exposed to a global source proportionally to its PageRank score in the graph (f_3 in Table 1). The fourth model configuration - *G4*, considers the extent of structural-equivalence (Burt, 1987) between an exposed member v_i and a global source v_g as designated by f_4 in Table 1. In the structural-equivalence model, (Burt, 1987) a member v_i is influenced by a global source v_g, depending on the extent to which they are structurally equivalent. Thus, the exposure probability of v_i to v_g is set according to the gamma correlation (Goodman and Kruskal, 1972) between v_i and v_g.

Next, the proposed model is compared to two baseline models. According to the first baseline model, the cohesion model, a member v_i is influenced only by local sources. Therefore, the *Co* baseline model (Table 2) aggregates the participation shifts of local sources, disregarding the weights of the edges. According to the second baseline model, the structural-equivalence model, a member is influenced by the extent of equivalence to other members. Therefore, the *SE* baseline model (Table 2) aggregates the participation shift of a member multiplied by the similarity index, $S_{ij} \in [-1,1]$, which indicates the gamma correlation value between an exposed member v_i and an influence source v_j, disregarding the weights of the edges and redundant information.

All the above model configurations consider the un-signed

Table 2
Exposure By Existing Models.

Model	Configuration	Exposures of member v_i
Cohesion	*Co*	$C_i(t_k) = \sum_{v_j \in H_k^i} (\Delta_{jk})$
Structural-equivalence	*SE*	$SE_i(t_k) = \sum_{v_j \neq i} (S_{ij} \Delta_{jk})$

Table 3
Datasets Description.

#	Dataset	|N|	|E|	|T|	OBS
1	Political online forum	638	62,024	2-weeks	44
2	Apple's support forum	11,118	37,545	2-weeks	98
3	FIFA Sub-Reddit	11,508	101,789	12-hours	56
4	(a) Twitter interactions	456,626	985,590	1-hour	168
	(b) Twitter relationships	456,626	14,855,842	–	–

|N|: number of nodes, |E|: number of edges, |T|: time interval, OBS: number of network observations.

participation shift of each influence source, regardless of whether an exposed member is influenced by either an increase or a decrease in the participation of influence sources. Other tested model configurations, which considered the signed participation shift of each influence source, produced less accurate results and, hence, are not discussed further.

5. Data description

To demonstrate and validate the proposed model, four datasets (Table 3) are analyzed in this study: three online forum-hosting sites, and posted communications from Twitter. By analyzing these four empirical datasets of human communication, temporal processes that indicate influence spread are observed. Since the sequence of discrete temporal events can be interpreted as a time-dependent point process (Karsai et al., 2012), member interactions were split into different time intervals to create network observations. However, due to diversity in the dynamic nature of different OSNs, the higher the network activity, the smaller must the interval size be if one wishes to capture the natural rhythm of human activities. For all networks, only active members were included. A member is considered active at time step t_k if s/he has posted a message at least once within time steps t_{k-1} and t_k.

The interactions in the three analyzed online forums (#1 to 3) originate from a root post that initializes a topic for discussion. After initialization, the discussions follow when members read existing posted messages and reply to those that interest them. Each posted message contains the unique identification (ID) number of the posting member, the date and time of its posting, and its parent message.

The first dataset contains written communications posted in a **political discussion online forum** between Mar. 30, 2015 and Dec. 19, 2016. The second dataset contains written communications posted in **Apple's technical support forum** between Sep. 19, 2003 to Oct. 19, 2007, as described by (Wang et al., 2011). Members of this support forum worldwide, either seek help or provide help to other members for technical problems encountered with Apple hardware and software by writing posts in English. Two preprocessing steps helped refine the data of both datasets: 1) member interactions were split into intervals of two weeks (similar to Galster and Tofan, 2013), creating at each time step a directed bi-weekly network observation, represented as a graph G_k, and 2) Early bi-weekly network observations with less than 20 nodes were removed. This stage yielded 44 bi-weekly directed network observations in the first dataset with a total of 638 unique registered members and 62,024 edges, and 98 bi-weekly directed network observations, with a total of 11,118 unique members and 37,545 edges, in the second dataset.

The third dataset contains member interactions on **Reddit.com**, initiated by Baumgartner (2015). Reddit allows members to comment and vote on posts of other members. Reddit presented a remarkable growth in recent years with 3.4 billion page views over 42.9 million unique visitors during August 2012 (Weninger et al., 2013). Towards the 2018 FIFA World Cup in Russia, the analyzed dataset captured the February's member interactions in a FIFA subreddit. Member interactions were split during preprocessing into intervals of 12 h since the third dataset presents more frequent interactions between members compared to the other two online forum datasets. Such setting of the interval size follows (Karsai et al., 2012), who found that member reaction time lasts up to 12 h in dynamic social spaces. This preprocessing stage yielded 56 directed network observations, with a total of 11,508 unique registered members and 101,789 edges.

The fourth dataset contains interactions on **Twitter** between Jul. 1, 2012, and Jul. 7, 2012 regarding the Higgs boson particle, described by De Domenico et al. (2013). The dataset contains 985,590 edges generated by 456,626 nodes with at least one of the following keywords or hashtags: lhc, cern, boson, higgs. This activity network describes four types of interactions: tweet (TW), retweet (RT), mention (MT), or reply (RE). The activity network was split during preprocessing by one-hour intervals, resulting in 168 time-steps. The time interval was set to one hour as sufficient to capture the spreading patterns in the network because De Domenico et al. (2013) found that the decay time scale of member activity was approximately 1.13 h. Contrary to the three previous datasets, Twitter enables members to form relationships, also allowing an analysis of the relationship network where nodes are tweet authors, and edges represent the followee/follower relationships. This preprocessing stage yielded 456,626 nodes and 14,855,842 directed relationship edges.

6. Analysis and results

The following sub-sections describe the construction of the different configuration models and their analysis over the four datasets. Sub-section 6.1, following the recommendations of Leifeld et al. (2015), estimates the parameters and uses the within-sample–goodness of fit (GOF) to evaluate the model for the first dataset. Then, to gain more confidence in the robustness of the results, the out-of-sample–GOF is applied to evaluate the model fit (Leifeld et al., 2015). The same analysis procedures are pursued in the three other datasets in Sub-section 6.2, revealing the underlying dynamics of influence sources in different OSNs and contributing to the generality of the results. Lastly, the comparison to the two baseline model configurations is reported in Sub-section 6.3.

6.1. Analysis results for the first dataset – a political forum

The model postulate that exposure to participation shifts, measured by the distance between two centrality vectors, influences the online posting engagement of a member. Since some centrality measures can be highly correlated, the Caret R package was used with a cutoff of 0.8 to remove the centrality measure with the highest mean absolute correlation, leaving for consideration the following centrality measures: CC, Hub, InKCores, Betweenness, InCloseness, OutCloseness.

TERGMs with and without temporal effects were set up to address the two questions of this study. The parameters of a TERGM without temporal effects can be interpreted as an average across the networks cross-section, whereas the parameters of a TERGM with temporal effects reflect cross-sectional dependencies over time. The following explanatory parameters (terms) were estimated with the btergm package for R (Leifeld et al., 2015) using terms coined in this software package.

Structural terms:

1 *Edges*- a baseline parameter that expresses the balance between creating and deleting edges.
2 *Mutual* - expresses a tendency to reciprocate edges.
3 *Ttriple* - a set of edges *(i→j), (j→k), (i→k)* that expresses the number of transitive triples in the network.

Terms that account for personal posting preferences:

4 *Pop* - computes the popularity ratio of a member's posts in each network observation, similarly to Chatzopoulou et al. (2010), by dividing the sum of replies to a member's posts by the sum of views (exposures) that her/his posts received. Exposures are discovered by

the number of members clicking on a post to view its content.

5 *ParticipationGroup* - a five-level categorical term (one per each role in Fig. 3) that accounts for the personal posting preferences of members.

6 *Nodematch-Participation-Group* – expresses the homophily of two members who share the same participation group while engaging online posting.

Terms that account for different exposure rates:

7 *Local* - sums the exposures to participation shifts of local sources (2).

8 *Global* - sums the effective exposures to participation shifts of global sources (7).

Temporal dependency terms (Leifeld et al., 2015):

9 *Memory* - a class of intertemporal dependencies designed to capture temporal processes, where each memory term can separately test for one of four temporal effects:
   a *Innovation* - whether previously unconnected nodes tend to tie in the current network.
   b *Loss* - captures edge loss by counting the number of edges that existed in t_{k-1} and do not exist in t_k.
   c *Autoregression* - whether previous ties are carried over to the current network.
   d *Stability* - whether both edges and non-edges are stable between the previous and the current network.

10 Single period delayed reciprocity (*delrcp*) - captures whether edges are reciprocated over time, to test if an edge from member *A* to *B* at t_{k-1}, increases the odds of edge formation from *B* to *A* at t_k.

The model was built in stages by first setting up a TERGM without cross-temporal dependencies based on 43 network observations out of 44, and then iteratively adding terms to create different model configurations. The coefficients of the terms in the model and their significance were estimated by using 1,000 replications. Then, each model configuration fit was examined by a within-sample–GOF that incorporated 300 networks simulated from each time step. Next, the distributions of the simulated and observed network matrices were compared. Lastly, each model configuration was applied to the observed network toward predicting a real network observation at a future time that the model has not seen yet (t_{44}). The model fit was then examined by an out-of-sample–GOF, comparing the matrices of the simulated networks to the matrices of the observed last network observation (that was omitted). The best model configuration was selected by using the ROC and PR curves to compare different models (Leifeld et al., 2015). Consolidating the performance of the model into a single measure, the area under the ROC curve (AUC) was calculated after each analysis. Since an AUC of 1.0 means perfect classification, while randomly- guessed classification scores 0.5, a larger AUC implies better model performance.

Model *F0* included Terms #1 and #2 as a null model, since at least two model terms must be provided to estimate a TERGM. *F0* produced significant parameters' effects, with AUC 0.54. **Model *F1*** included Terms #1 to #3 and produced significant parameters' effects, with AUC 0.61. To account for exposure to *local* sources, **Model *F2*** was constructed by adding Term #7 to *F1* as a node term (*nodecov.L*). *F2* produced significant parameters' effects, with AUC 0.71, and outperformed *F1* in influence prediction. To account for exposures to *global* sources, Term #8 was separately tested with different configurations *G1 to G4* and was added to *F2* as a node term to define **Model *F3***. Out of the four configurations, *F3* yielded the best results in influence prediction for the *G2* (*nodecov.G2)* configuration. *F3* with *G2* (referred hereafter as *F3*) produced significant parameters' effects, with AUC 0.77, and outperformed *F2*, in influence prediction.

The significant parameters' effects of *F3* suggest that member posting engagement is influenced by exposures to local *and* global sources, which provides an answer to the first question in this study. Moreover, the results provide an answer for the second question as well since influence prediction of *F3* outperforms *F2*.

To account for personal preferences, Terms #4 to #6 were added to *F3*, creating **Model *F4***. *F4* produced significant parameters' effects, with AUC 0.83, and outperformed *F3* in predicting influence. Considering the structural terms of *F4*, the *edges* term captures the density of the networks, suggesting that most members interact with a few others. A positive coefficient of the *mutual* term indicates that the probability of forming mutual edges in the network is higher than forming non-mutual edges. The *ttriple* term shows a positive triangulation coefficient, suggesting closed structures with a hierarchical ordering that correspond to hierarchical threads in the forum. The model also shows a significant positive local effect (*nodecov.L*), which indicates that members with high exposure to local influence sources seem more likely to engage in online postings. In addition, a significant *global* effect (*nodecov.G2*) indicates that a non-edge mechanism can also influence a member. A negative coefficient indicates that members seem less likely to engage in posting when exposed to participation shift of a global influence source.

Recall that the analyzed dataset in this sub-section is a political discussion forum. Therefore, it is likely that the local and global effects in the model uncover the following growing concern. Although members were exposed to a significant amount of cross-cutting content, they seem to have become more polarized about political issues by preferring to engage with local content consistent with their views (An et al., 2014; Bakshy et al., 2015) rather than global cross-cutting content. This behavior can lead to political echo chambers. In addition, this behavior indicates a psychological mechanism known as the selective exposure theory (Frey, 1986). Selective exposure means that members tends to favor information that aligns with their pre-existing local views while avoiding global contradictory information. Moreover, recent research on political forums (Garimella et al., 2018) found that members who try to bridge the local echo chambers pay a "price of bipartisanship" in terms of their network centrality and content appreciation, which reduces one's willingness to bridge.

Regarding personal preferences (Term #5), members were categorized into five participation groups (Fig. 3). However, the number of members in group #5 (Trolls) was small. This finding is in line with previous studies indicating that the prevalence of antisocial behavior online is relatively uncommon, and Trolls are a small number of individuals (Binns, 2012; Cheng et al., 2017). Trolls were detected in only few network observations at time steps: t_2, t_{17}, t_{24}, t_{39} and then vanished (Coles and West, 2016). In addition, the participation level of members who affiliate with group #5 compared to members who affiliate with group #1, were not significant (results are not presented for the sake of brevity). Therefore, members of group #5 were discarded from the analysis. Without Trolls, the results show that members of participation group #2 are more likely to engage than are members of group #1 and, generally, the higher the member participation (expressed by affiliation with a higher participation-level group), the more likely that members who affiliate with that group will engage.

Surprisingly, a significant negative coefficient of Term #6 indicates that members who affiliate with the same participation group are less likely to engage and exchange information with each other, indicating a lack of homophily. Intuitively, one would expect that two members who affiliate with the same participation group will more likely engage in a discussion (Barone and Coscia, 2018; Halberstam and Knight, 2016). Homophily may also explain engagement (Aral et al., 2009). Thus, a lack of homophily strengthens our findings that the model predicts engagement due to influence (and not due to homophily).

Finally, there is a significant positive popularity effect (*nodecov.Pop*). The higher the popularity of a member's posts, the more likely that the member who created those posts will engage. A member who creates a popular post receives benefits in the form of popularity and is encouraged to post (Kankanhalli et al., 2005; Lin, 2007).

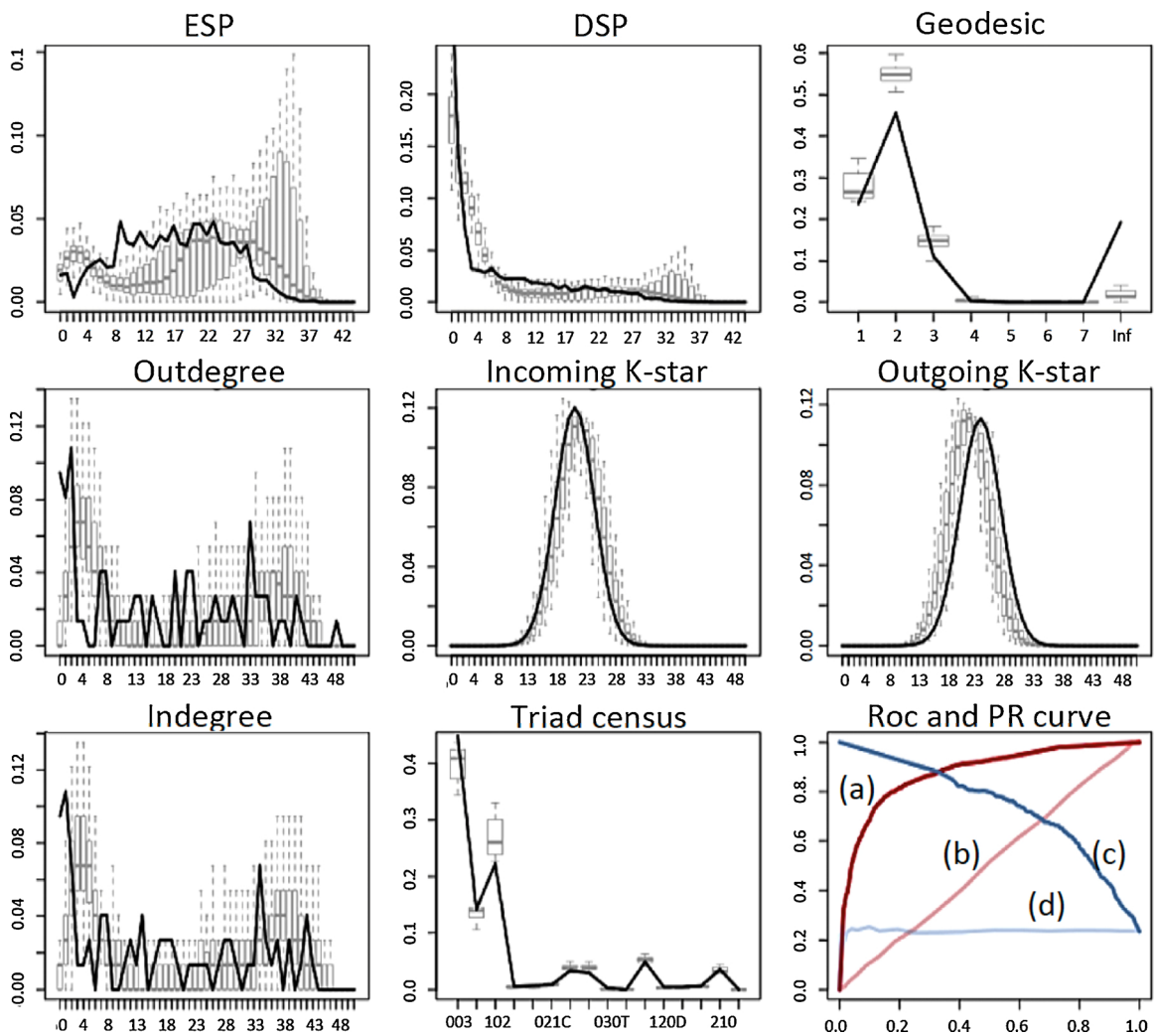

Fig. 4. Out-of-sample GOF for TERGM without temporal effects (*F4*).

The **out-of-sample-GOF** evaluation of *F4* (Fig. 4) presents the distributions according to eight auxiliary network statistics of 300 networks simulated (boxplots) according to *F4* versus the observed network (black line). The horizontal lines in the boxplots denote the median of a statistic across simulated networks. The boxes denote upper and lower quartile. The upper and lower whiskers denote maximum and minimum values excluding outliers. The closer the solid black line gets to the median of the boxplots the better. On the lower right box in Fig. 4, line *(a)* is the ROC curve, line *(b)* is the ROC curve of a random graph of the same size with the same density, line *(c)* is the PR curve, and line *(d)* is the PR curve of the null model. *F4* shows a good fit to the distributions of dyad-wise shared partners (DSP), geodesic distance, Indegree, Outdegree, triad-census and acceptable, but a less good fit to the incoming k-star, outgoing k-star, and edge-wise shared partners (ESP), which led us to consider the use of temporal effects.

Next, the TERGM was tested with temporal effects, reflecting intertemporal dependencies and designed to capture temporal processes (Leifeld et al., 2015). In other words, a previous network observation was treated as a covariate for the current network observation. **Model *F5***, was constructed by adding temporal terms to *F4*, produced significant parameters' effects, with AUC 0.86. Two temporal terms performed best. The first top performer was *autoregression*, which presented a positive coefficient. This finding reflects the tendency of members to keep interacting over time and to bond, indicating the existence of an online community (Sonnenbichler, 2010). The second top performing term was *delrcp*, which presented a negative coefficient. This finding indicates that members do not tend to reciprocate with a single time-step delay but rather reciprocate immediately, as indicated by the *mutual* term. Members seem to be interested in discussing political issues, as indicated by quick responses (Velasquez, 2012).

The inclusion of temporal terms changed some of the effects sizes in *F5* compared to *F4*, but not their qualitative interpretation. The GOF result for *F5* (Fig. 5) is particularly good, especially for the incoming and outgoing k-star distributions. A less good, but still acceptable fit, can be seen for ESP. In all, *F5* shows a good and better fit than that of *F4*. Out of the tested set of models, Table 4 (for the sake of brevity) only presents *F5*, which is the model with the best configuration.

To summarize, two questions were asked in this study. To answer the two questions, we tested and evaluated the proposed model on an online political forum. The results suggest that influence modeling becomes more accurate by distinguishing local and global effects. Aiming to further generalize these results, the proposed model is applied next to several diverse datasets: the Apple, the Twitter, and the Reddit datasets spanning thousands of nodes and edges.

6.2. Analysis results for the second to the fourth datasets

Each model configuration was tested on each of the three datasets, following the same stages as described in Sub-section 6.1. For brevity sake, only the best models are reported hereafter. Since no data regarding popularity (Term #4) was available in the datasets, the in-2-star term (*istar2*) was used to test for member popularity effect, as done

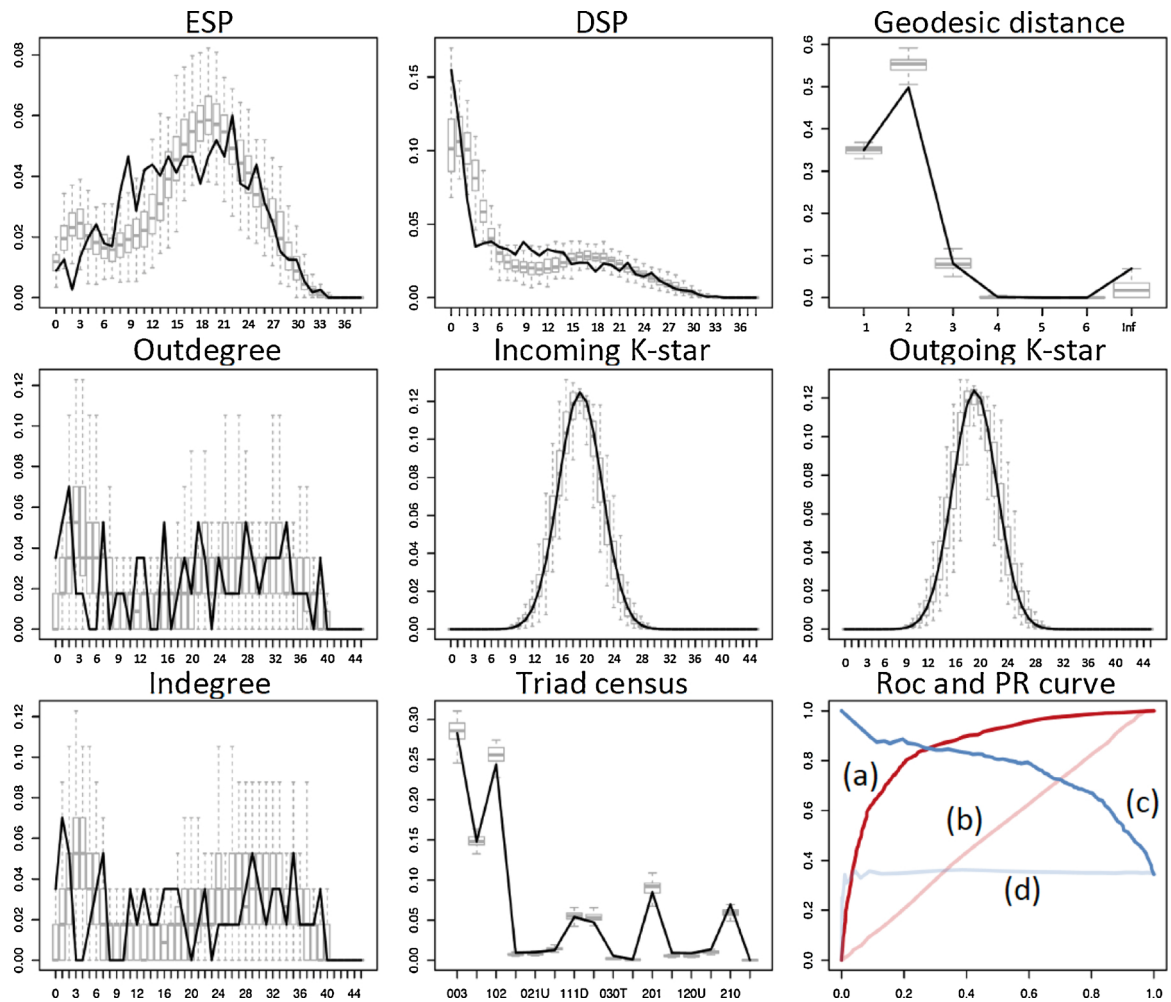

Fig. 5. Out-of-sample GOF for TERGM with temporal effects (*F5*).

in Desmarais and Cranmer (2012a), by measuring how often node *i* connects to nodes *j* and *k*. The term *istar2* can be used to represent popularity since the more edges on a node, the more an additional edge will create a two stars structure. In addition, four *participation groups* (Term #5) were analyzed in both the Apple and Reddit forums since they presented a similar data-behavior to that of the political forum (elaborated upon in Sub-section 6.1).

6.2.1. Analysis results for the apple dataset

After removing correlated centrality measures, the following centralities were considered in computing the exposure to participation shifts in (2) and (7) above: Hub, Betweenness, CC, Eigenvector Centrality, Authority InKCores, OutKCores, Indegree, and OutCloseness. *Model A* was constructed similar to *F5* and produced significant effects (Table 4) with AUC 0.82, except for Term #6. Although Model *A* presented a change in the effect sizes compared to *F5*, the interpretation of Terms #1, #2, #3, #5, and #9c has not changed. However, the *delrcp* term showed a significant positive effect, indicating that the response of one member to an interaction initiated by another member is not necessarily immediate. This is reasonable to expect in a support forum where members exchange their knowledge by asking and answering questions (Tomasello et al., 2012). Members also post answers to enhance their reputation (Anderson et al., 2013; Cavusoglu et al., 2015), or to receive benefits such as satisfaction from helping others (Kankanhalli et al., 2005; Lin, 2007). In other words, this finding suggests that one member tends to delay a response to a question of another member because the high-quality answer expected takes time to produce.

The interpretations of the local effect, the global effect, and the *istar2* effect are explained in the following paragraphs, under the FIFA analysis, since both forums share similar behavior.

6.2.2. Analysis Results for the FIFA subreddit dataset

Model R was constructed similar to *F5* and used the following centralities to compute exposure to participation shifts in (2) and (7) above: Hub, Betweenness, CC, Eigenvector Centrality, Authority InKCores, OutKCores, Indegree, and InCloseness. Model *R* produced significant effects (Table 4) with AUC 0.79 and, except for the temporal effects, the interpretation of the terms has not changed compared to *F5*.

A significant positive coefficient of the *delrcp* term indicates that members tend to reciprocate within a time step. On Reddit, registered members can develop a discussion by commenting on a submission, and vote on content items such as submissions and comments. Reddit allows its members to evaluate other members using two main measures. The first is the amount of time a Reddit account has been active. The second measure is the "karma" number (the votes count of a member's posted comments), indicating the extent to which a member is active and respected. Therefore, it is likely that Reddit members invest time and effort in preparing high-quality answers before online posting, possibly resulting in a delayed response.

Term #9a's significant positive effect indicates that members tend to create interactions with other members with whom they did not interact within the previous time step. An acceptable strategy to maximize one's karma is to interact with other members with whom s/he did not interact earlier (Richterich, 2014) aiming to receive additional karma.

A significant negative effect of the *istar2* term in both the Reddit and

Table 4
TERGM Estimation Results of The Best Models.

	Term	Model F5 Political forum	Model A Apple	Model R Reddit	Model T Twitter (a)
1	edges	−2.40 *	−4.35 *	−7.39 *	−4.62 *
		[-2.58; -2.13]	[-4.95; -2.91]	[-8.16; -2.90]	[-5.17; -4.12]
2	mutual	2.38 *	3.48 *	6.97 *	2.82 *
		[2.29; 2.45]	[3.22; 3.78]	[6.75; 7.20]	[2.75; 2.89]
3	ttriple	0.02 *	0.25 *	0.50 *	1.73 *
		[0.02; 0.02]	[0.15; 0.33]	[0.40; 0.56]	[1.33; 2.38]
4	nodecov.Pop	0.27 *	–	–	–
		[0.15; 0.37]			
5	nodefactor. Participation	0.24 *	0.64 *	1.26 *	–
	Group.2	[0.20; 0.28]	[0.46; 0.69]	[1.21; 1.35]	
	nodefactor. Participation	0.38 *	0.82 *	1.48 *	1.51 *
	Group.3	[0.32; 0.43]	[0.72; 0.94]	[1.42; 1.54]	[1.40; 1.69]
	nodefactor. Participation	0.44 *	0.94 *	1.95 *	–
	Group.4	[0.38; 0.51]	[0.82; 1.09]	[1.82; 2.06]	
6	nodematch. Participation	−0.15 *	−0.04	−0.13 *	−0.27 *
	Group	[-0.19; -0.12]	[-0.22; -0.09]	[-0.22; -0.06]	[-0.32; -0.20]
7	nodecov.L	0.02 *	0.02 *	0.36 *	0.01 *
		[0.02; 0.02]	[0.01; 0.02]	[0.34; 0.40]	[0.00; 0.02]
8	nodecov.G	−0.40 * (*G2*)	−0.28 *(*G2*)	−0.94 *(*G2*)	0.42 *(*G3*)
		[-0.47; -0.35]	[-0.56; -0.18]	[-3.97; -0.60]	[0.32; 0.56]
9	memory	0.81 *(c)	1.04 *(c)	1.32 * (a)	3.56 * (c)
		[0.76; 0.86]	[0.79; 1.25]	[0.80; 1.64]	[3.17; 4.12]
10	delrcp	−0.21 *	1.05 *	4.41 *	2.89 *
		[-0.26; -0.15]	[0.71; 1.05]	[3.85; 4.65]	[2.44; 3.36]
11	istar2	–	−0.12 *	−0.22 *	0.33 *
			[-0.19; -0.04]	[-0.24; -0.21]	[0.29; 0.40]

* Significant at the 0.95 level when 0 is outside the confidence interval.
Memory: (a) Innovation; (b) Loss; (c) Autoregression; (d) Stability.
nodecov.G: *Gi* designates the ith global configuration in Section 6.

the Apple forums indicates that a member is less likely to interact with members who have many edges. This can be attributed to one's shyness to engage with people whom s/he does not know (Fisher et al., 2011), or to one's feeling that s/he neither needs nor has much to contribute (Fisher et al., 2011).

Significant negative exposure to global sources in both the Reddit and the Apple forums indicates that a member is less likely to engage in posting when exposed to a participation shift of a global influence source. This behavior can reflect, for example, the bystander effect (Plötner et al., 2015) according to which one is less likely to engage in helping others, by contributing to the discussion, in the presence of other potential helpers. From a game-theoretical view helping others is costly, hence a member's likelihood of engaging decreases as the number of other potential helpers increases (Diekmann, 1985; Fisher et al., 2011). One's shyness to engage with people whom s/he does not know (Fisher et al., 2011) or one's feeling that s/he neither needs nor has much to contribute (Fisher et al., 2011) are two other examples.

The positive and significant exposure to local sources in both the Reddit and the Apple forums indicates that a member with high exposure to local sources is more likely to engage in online postings. Thus, one's behavior is influenced by the behavior of her/his neighbors. Members who are exposed to a neighbor's behavior tend to converge towards that same behavior and follow the norm of posting (Velasquez, 2012).

6.2.3. Analysis results for the Twitter dataset

Model T was constructed similar to *F5* and used the following centralities to compute exposure to participation shifts in (2) and (7) above: Hub, Betweenness, CC, Eigenvector Centrality, Authority InKCores, OutKCores, Indegree, Outdegree, and OutCloseness. The interpretations of Terms #1, #2, #3, #7, and #9c have not changed compared to *F5*. However, the effects of the other terms were not significant compared to *F5*.

Considering Term #5, using the group detection sub-model (Fig. 3), the dataset contained 47.47% *Novices*; 0.04% *Passive* members; 52.45% *Active* members; 0.04% *Leaders*; but no *Trolls*. The *Novices*, and *Activists* participation groups indicate reduced and high (bursts) posting participations respectively. This is indicative of the bursty nature of member interactions, where members tend to send several tweets in short periods, separated by long periods of no (or reduced) activity (Barabasi, 2005; Karsai et al., 2012). The effects of the four participation groups (Term #5) and Term #6 were not significant. Therefore, to better represent member behavior in the dataset, members of Group #4 were merged into Group #3 due to high activity, and members of Group #2 were merged into Group #1 due to low activity. With two participation groups that represent member behavior. After this merging the effects of the participation groups (Term #5) and Term #6 were significant (Table 4).

Contrary to Apple and the Reddit forums, a significant positive effect of the *istar2* term in the Twitter dataset indicates that members are more likely to interact with members who have many edges. Twitter is different from online forums by, for example, limiting the length of posts and by a higher frequency of content updates (Java et al., 2007). Moreover, Twitter members behave differently than other OSN members by maintaining weak social relationships with a high turnover of contacts in their networks (On-at et al., 2017).

In Model *T*, the *local* term has presented a similar effect as in *F5*, and the *global* term (Table 1) was not significant with the *G2* configuration but was significant with the *G3* configuration. This finding regarding the *global* term indicates that the probability of member exposure to global sources is proportional to the PageRank score of a global source in the network. The significance of Model *T* with the *G3* configuration, rather than with the *G2* configuration, could be explained by the fact that the PageRank considers structural properties of the graph as opposed to individual measures of a member. This finding is in line with the finding regarding the *istar2* term as elaborated upon above. It is also in line with the On-at et al. (2017) finding that unlike members behavior in online forums, Twitter members maintain weak social relationships with the global environment beside their local network. Model *T*'s positive *global* effect indicates that members exposed to participation shift of a global influence source are more likely

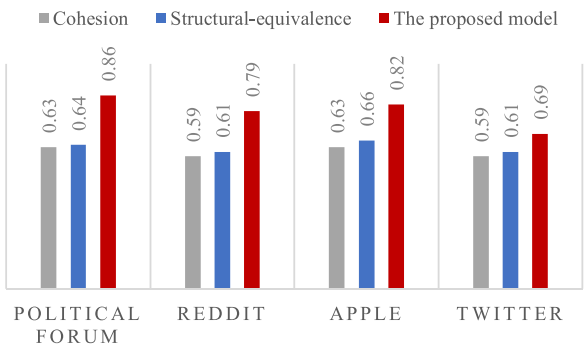

Fig. 6. AUC of the best models Vs. baseline models.

to engage in posting, and that likelihood is proportional to the source's PageRank score.

The coefficient of the *delrcip* term was positive for Model *T*, whereas it was negative for *F5*, indicating that Twitter members tend to reciprocate with delay of a single period. Indeed, De Domenico et al. (2013) found that the times between subsequent member actions is characterized by bursts of rapidly occurring events separated by periods of inactivity (Karsai et al., 2012). In other words, members tend to send several tweets in short periods, separated by long periods of no (or reduced) activity. Barabasi (2005) found this behavior to be related to a decision-based queuing process, where individuals tend to act in response to some perceived priority.

To further test for a local effect for the Twitter dataset, the activity network and the followee/follower relationships network (discussed in Section 5) were considered. For each activity network observation, the relationships edges between members were considered to calculate the value of exposure to local sources (1). Note that, in this case, exposure to local influence sources is measured by the effects of relationship edges instead of the interaction activity edges. To account for the sum of exposers to local influence sources, the *edgecov* term which is a square matrix of edge-covariates was used. This term adds one statistic to the model, equals to the sum of the covariate values for each edge appearing in the relationship network.

Upon incorporating both Term #7 and the *edgecov* term in Model *T*, Term #7 was significant, but the *edgecov* term was not. However, the *edgecov* term was significant when Term #7 was removed from the model. According to this finding, the influence resulting from local neighbors is better captured by the activity network rather than by the relationship network, which considers the followee/follower list of a member. This finding can be explained by the Dunbar's number (Gonçalves et al., 2011), which represents the theoretical cognitive limit of the number of members with whom one can maintain stable social relationships. Thus, members are influenced by strong ties (members with whom they interact) rather than by weak ties (members whom they follow). Overall, Model *T* (Table 4) achieved an accuracy of AUC 0.69.

6.3. Results of comparison to baseline configuration models

Performance of the *F5, A, T,* and *R* models, each of which introduces local effects by considering network structure, global effects, temporal, and personal preferences, were compared to the cohesion, and structural-equivalence baseline models. The cohesion model or the structural-equivalence model is constructed for each dataset by using respectively the *Co* or the *SE* configuration (Table 2), with Terms #1 to #3 accounting for structural effects. In terms of predicting influence, both baseline models achieved less accurate results (Fig. 6) than did each of *F5, A, T,* and *R* models. The results also show that, for all datasets, the proposed influence model is more accurate at catching influence spread than the baseline models by distinguishing between exposures to participation shifts of local and global influence sources.

Overall, the results suggest that influence spread is better explained by exposure to participation shifts of both local and global influence sources rather than by exposure to participation shifts of only local sources. For all four datasets, the *local* term presents a significant positive effect on member influence defined as the probability of creating an edge. This finding indicates that influence spread over the networks is driven by social reinforcement at the neighborhood level. In addition, influence is also driven by a non-edge mechanism beyond neighborhood effects as evident by a significant *global* effect.

Although the interpretation of the coefficient effects is similar across all models for the most part, in some cases some differences are notable. For instance, the prediction results of Models *A*, *T,* and *R* are less accurate than those of Model *F5*. Yet, the focus of this research was on showing the presence of both local and global influence on member posting engagement. Specifically, this research demonstrated that the behavior of non-neighbors, expressed by forming an edge in the network, can help predict influence spread in OSNs. The results confirm that not only a local influence effect exists, but that a global influence effect, transmitted beyond neighbor-to-neighbor interactions exists as well.

7. Conclusions and limitations

This research provides an innovative view of influence spread in OSNs by developing, demonstrating and evaluating a novel local-global influence model that can be used to predict whether a member will be influenced to engage in online postings. Exposure to participation shifts, often and mostly modeled as a flow within network edges like a flow of viruses, ideas, or information, influences a member of a dynamic OSN to shift participation. Here, however, both local influence and global influence on member online posting engagement are identified according to the presented model which was tested using four real datasets.

Model testing revealed evidence of not only a positive local influence effect on member posting engagement but also found evidence of a significant global influence effect on member posting engagement. Identifying different effects on the different datasets, the conclusion emerging based on all four datasets is that influence can reach a member either via the edges of the social network or through global effects. Moreover, prediction accuracy is improved by considering both local and global influence. Lastly, a comparison of the proposed temporal model to a couple of baseline models shows that by incorporating personal preferences, as well as exposures to local and global influence sources, can help predict influence better than a cohesion model or a structural-equivalence model.

The main research objective of the study, to demonstrate that the behavior of non-neighbors can help predict influence spread in OSNs, was accomplished. Yet, while presenting novel modeling and interesting results, this study also has three main limitations that call for future work. First, the prediction of posting preferences was achieved by grouping members into participation levels only. However, since human behavior is complex, richer parameters as age and interests could reveal the larger-scale dynamics of online behavior. Moreover, external exposure to media sources like TV can also influence member behavior. Another limitation is that only public messages were analyzed, while influence can also be due to private interactions between OSN members. Finally, the networks were split by time intervals, but the interval size might bias the findings. Thus, this study is just one step towards understanding the role of global influence in online spaces, demonstrating that different OSNs may present different local and global effects. It is left for future work to consider more diverse datasets such as biology, or neuroscience networks. Another interesting future research direction involves the detection of influencers based on the trails of influence spread.

The proposed model can be used to infer influence propagation in OSNs and thus predict the behavior of an OSN ecosystem. For example, this modeling can help predict which post will go viral by detecting members that are likely to engage. In sum, the research provides new

insights of relevance to social network research about the spread of online influence, in general, and about viral marketing in OSNs, in particular.

References

Agarwal, N., Liu, H., Tang, L., Yu, P.S., 2008. Identifying the influential bloggers in a community. Proceedings of the 2008 International Conference on Web Search and Data Mining 207–218.

Aghdam, S.M., Navimipour, N.J., 2016. Opinion leaders selection in the social networks based on trust relationships propagation. Karbala Int. J. Mod. Sci. 2 (2), 88–97.

Albert, R., DasGupta, B., Hegde, R., Sivanathan, G.S., Gitter, A., Gürsoy, G., Paul, P., Sontag, E., 2011. A new computationally efficient measure of topological redundancy of biological and social networks. Phys. Rev. E 84, 036117.

An, J., Quercia, D., Cha, M., Gummadi, K., Crowcroft, J., 2014. Sharing political news: the balancing act of intimacy and socialization in selective exposure. EPJ Data Sci. 3 (1), 12.

Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J., 2013. Steering user behavior with badges. Proceedings of the 22nd International Conference on World Wide Web 95–106.

Aral, S., Muchnik, L., Sundararajan, A., 2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proc. Natl. Acad. Sci. 106 (51), 21544–21549.

Arnaboldi, V., Conti, M., Passarella, A., Dunbar, R., 2013. Dynamics of personal social relationships in online social networks: a study on Twitter. Proceedings of the First ACM Conference on Online Social Networks 15–26.

Bakshy, E., Messing, S., Adamic, L.A., 2015. Exposure to ideologically diverse news and opinion on Facebook. Science 348 (6239), 1130–1132.

Bakshy, E., Rosenn, I., Marlow, C., Adamic, L., 2012. The role of social networks in information diffusion. Proceedings of the 21st International Conference on World Wide Web 519–528.

Barabasi, A.-L., 2005. The origin of bursts and heavy tails in human dynamics. Nature 435 (7039), 207–211.

Bartal, A., Ravid, G., 2019. Member behavior in dynamic online communities: role affiliation frequency model. IEEE Trans. Knowl. Data Eng. https://doi.org/10.1109/TKDE.2019.2911067.

Barone, M., Coscia, M., 2018. Birds of a feather scam together: trustworthiness homophily in a business network. Soc. Networks 54, 228–237.

Barzilai-Nahon, K., 2006. Gatekeepers, virtual communities and the gated: multi-dimensional tensions in cyberspace. Int. J. Commun. Law Policy 11 9-9.

Baumgartner, J., 2015. Complete Public Reddit Comments Corpus. Retrieved March 4, 2018, from Google BigQuery:. https://bigquery.cloud.google.com/table/fh-bigquery:reddit_comments.2018_02.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc. Ser. B (Methodological) 192–236.

Binns, A., 2012. DON'T FEED THE TROLLS! Managing troublemakers in magazines' online communities. J. Pract. 6 (4), 547–562.

Borgatti, S.P., 2006. Identifying sets of key players in a social network. Comput. Math. Organ. Theory 12 (1), 21–34.

Borgatti, S.P., Everett, M.G., 1992. Notions of position in social network analysis. Sociol. Methodol. 1–35.

Burt, R.S., 1987. Social contagion and innovation: cohesion versus structural equivalence. Am. J. Sociol. 92 (6), 1287–1335.

Burt, R.S., 2009. Structural Holes: the Social Structure of Competition. Harvard University Press.

Caimo, A., 2016. Exponential random graph modelling of static and dynamic social networks. Dynamic Networks and Cyber-Security 157–170.

Cavusoglu, H., Li, Z., Huang, K.-W., 2015. Can gamification motivate voluntary contributions?: The case of StackOverflow Q&A community. Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work and Social Computing 171–174.

Centola, D., 2010. The spread of behavior in an online social network experiment. Science 329 (5996), 1194–1197.

Chatzopoulou, G., Sheng, C., Faloutsos, M., 2010. A first step towards understanding popularity in YouTube. Proceedings of the INFOCOM Conference on Computer Communications Workshops 1–6.

Chen, W., Yuan, Y., Zhang, L., 2010. Scalable influence maximization in social networks under the linear threshold model. Proceedings of the 10th International Conference on Data Mining (IDCM) 88–97.

Cheng, J., Adamic, L., Dow, P.A., Kleinberg, J.M., Leskovec, J., 2014. Can cascades be predicted? Proceedings of the 23rd International Conference on World Wide Web 925–936.

Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., Leskovec, J., 2017. Anyone can become a troll: causes of trolling behavior in online discussions. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing 1217–1230.

Coleman, J.S., Katz, E., Menzel, H., 1966. Medical Innovation: a Diffusion Study. Free Press, New York.

Coles, B.A., West, M., 2016. Trolling the trolls: online forum users constructions of the nature and properties of trolling. Comput. Human Behav. 60, 233–244.

Costa, Ld.F., Rodrigues, F.A., Travieso, G., Villas Boas, P.R., 2007. Characterization of complex networks: a survey of measurements. Adv. Phys. 56 (1), 167–242.

Cranmer, S.J., Heinrich, T., Desmarais, B.A., 2014. Reciprocity and the structural determinants of the international sanctions network. Soc. Networks 36, 5–22.

De Domenico, M., Lima, A., Mougel, P., Musolesi, M., 2013. The anatomy of a scientific rumor. Sci. Rep. 3, 2980.

Desmarais, B.A., Cranmer, S.J., 2012a. Micro-level interpretation of exponential random graph models with application to estuary networks. Policy Stud. J. 40 (3), 402–434.

Desmarais, B.A., Cranmer, S.J., 2012b. Statistical mechanics of networks: estimation and uncertainty. Phys. A Stat. Mech. Its Appl. 391 (4), 1865–1876.

Diekmann, A., 1985. Volunteer's dilemma. J. Conflict Resolut. 29 (4), 605–610.

Doerr, B., Fouz, M., Friedrich, T., 2012. Why rumors spread so quickly in social networks. Commun. ACM 55 (6), 70–75.

Domingos, P., 2005. Mining social networks for viral marketing. IEEE Intell. Syst. 20 (1), 80–82.

Edwards, A., 2013. (How) do participants in online discussion forums create 'echo chambers'?: the inclusion and exclusion of dissenting voices in an online forum about climate change. J. Argum. Context. 2 (1), 127–150.

Fisher, P., Kruegar, J., Greitemeyer, T., Vogrincic, C., Kastenmuller, A., Dieter, F., Heene, M., Wicher, M., Kambacher, M., 2011. The bystander-effect: a meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. Psychol. Bull. 137 (4), 517–537.

Flaxman, S., Goel, S., Rao, J.M., 2013.). Ideological segregation and the effects of social media on news consumption. SSRN Scholarly. Paper ID 2363701. Social Science Research Network, Rochester, NY.

Forestier, M., Stavrianou, A., Velcin, J., Zighed, D.A., 2012. Roles in social networks: methodologies and research issues. Web Intell. Agent Syst.: Int. J. 10 (1), 117–133.

Fournier, S., Avery, J., 2011. The uninvited brand. Bus. Horiz. 54 (3), 193–207.

Frey, D., 1986. Recent research on selective exposure to information. Advances in Experimental Social Psychology 19. Academic Press, New York, pp. 41–80.

Friedkin, N.E., 1990. Social networks in structural equation models. Soc. Psychol. Q. 316–328.

Friedkin, N.E., Johnsen, E.C., 2011. Social Influence Network Theory: A Sociological Examination of Small Group Dynamics Vol. 33 Cambridge University Press.

Fu, W., Song, L., Xing, E.P., 2009. Dynamic mixed membership blockmodel for evolving networks. Proceedings of the 26th Annual International Conference on Machine Learning 329–336.

Galster, M., Tofan, D., 2013. Exploring possibilities to analyse microblogs for dependability information in variability-intensive open source software systems. Proceedings of the International Symposium on Software Reliability Engineering Workshops (ISSREW) 321–325.

Ganley, D., Lampe, C., 2009. The ties that bind: social network principles in online communities. Decis. Support Syst. 47 (3), 266–274.

Garimella, K., De Francisci Morales, G., Gionis, A., Mathioudakis, M., 2018. Political discourse on social media: echo chambers, gatekeepers, and the price of bipartisanship. Proceedings of the 2018 World Wide Web Conference on World Wide Web 913–922.

Gibson, D.R., 2003. Participation shifts: order and differentiation in group conversation. Soc. Forces 81 (4), 1335–1380.

Gil de Zúñiga, H., Valenzuela, S., 2011. The mediating path to a stronger citizenship: online and offline networks, weak ties, and civic engagement. Commun. Res. 38 (3), 397–421.

Goel, S., Anderson, A., Hofman, J., Watts, D.J., 2015. The structural virality of online diffusion. Manage. Sci. 62 (1), 180–196.

Goel, S., Watts, D.J., Goldstein, D.G., 2012. The structure of online diffusion networks. Proceedings of the 13th ACM Conference on Electronic Commerce 623–638.

Golder, S.A., Donath, J., 2004. Social roles in electronic communities. Internet Res. 5, 19–22.

Gómez, V., Kaltenbrunner, A., López, V., 2008. Statistical analysis of the social network and discussion threads in slashdot. Proceedings of the 17th Inter National Conference on World Wide Web 645–654.

Gonçalves, B., Perra, N., Vespignani, A., 2011. Modeling users' activity on Twitter networks: validation of Dunbar's number. PLoS One 6 (8), e22656.

Goodman, L.A., Kruskal, W.H., 1972. Measures of association for cross classifications, IV: simplification of asymptotic variances. J. Am. Stat. Assoc. 67 (338), 415–421.

Granovetter, M., 1978. Threshold models of collective behavior. Am. J. Sociol. 83 (6), 1420–1443.

Granovetter, M.S., 1973. The strength of weak ties. Am. J. Sociol. 1360–1380.

Guille, A., Hacid, H., Favre, C., Zighed, D.A., 2013. Information diffusion in online social networks: a survey. Acm Sigmod Rec. 42 (2), 17–28.

Hagel, J., 1999. Net gain: expanding markets through virtual communities. J. Interact. Mark. 13 (1), 55–65.

Halberstam, Y., Knight, B., 2016. Homophily, group size, and the diffusion of political information in social networks: evidence from Twitter. J. Public Econ. 143, 73–88.

Handcock, M.S., Hunter, D.R., Butts, C.T., Goodreau, S.M., Krivitsky, P.N., Bender-deMoll, S., Morris, M., 2016. statnet: Software Tools for the Statistical Analysis of Network Data. The Statnet Project. URL https: //CRAN.R-project.org/package=statnet. http://www.statnet.org/.

Hanneke, S., Fu, W., Xing, E.P., 2010. Discrete temporal models of social networks. Electron. J. Stat. 4, 585–605.

Himelboim, I., Gleave, E., Smith, M., 2009. Discussion catalysts in online political discussions: content importers and conversation starters. J. Comput. Commun. 14 (4), 771–789.

Hu, Y., Havlin, S., Makse, H.A., 2014. Conditions for viral influence spreading through multiplex correlated social networks. Phys. Rev. 4 (2), 021031.

Huang, S., Lv, T., Zhang, X., Yang, Y., Zheng, W., Wen, C., 2014. Identifying node role in social network based on multiple indicators. PLoS One 9 (8), e103733.

Huberman, B.A., Romero, D.M., Wu, F., 2009. Social networks that matter: twitter under the microscope. First Monday 14 (1), 2009.

Hunter, D.R., Handcock, M.S., Butts, C.T., Goodreau, S.M., Morris, M., 2008. Ergm: a

package to fit, simulate and diagnose exponential-family models for networks. J. Stat. Softw. 24 (3), 54–60.
Java, A., Song, X., Finin, T., Tseng, B., 2007. Why we twitter: understanding microblogging usage and communities. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis 56–65.
Jurvetson, S., 2000. What exactly is viral marketing. Red Herring 78, 110–112.
Kankanhalli, A., Tan, B.C., Wei, K.-K., 2005. Contributing knowledge to electronic knowledge repositories: an empirical investigation. Mis Q. 113–143.
Kaplan, A.M., Haenlein, M., 2011. Two hearts in three-quarter time: how to waltz the social media/viral marketing dance. Bus. Horiz. 54 (3), 253–263.
Karimi, F., Holme, P., 2013. Threshold model of cascades in empirical temporal networks. Phys. A Stat. Mech. Its Appl. 392 (16), 3476–3483.
Karsai, M., Kaski, K., Barabási, A.-L., Kertész, J., 2012. Universal features of correlated bursty behaviour. Sci. Rep. 2, 397.
Kempe, D., Kleinberg, J.M., Tardos, É., 2005. Influential nodes in a diffusion model for social networks. Proceedings of the International Colloquium on Automata, Languages, and Programming 1127–1138.
Kleinberg, J.M., 1999. Authoritative sources in a hyperlinked environment. J. ACM 46 (5), 604–632.
Kleinberg, J., 2007. Cascading behavior in networks: algorithmic and economic issues. Algorithmic Game Theory 24, 613–632.
Kreindler, G.E., Young, H.P., 2014. Rapid innovation diffusion in social networks. Proc. Natl. Acad. Sci. 111 (Supplement 3), 10881–10888.
Kurka, D.B., Godoy, A., Von Zuben, F.J., 2015. Online Social Network Analysis: a Survey of Research Applications in Computer Science. arXiv preprint arXiv:1504.05655.
LaCour, M.J., 2013. A balanced information diet, not echo chambers: evidence from a direct measure of media exposure. SSRN Scholarly Paper ID 2303138. Social Science Research Network, Rochester, NY, pp. 2013.
Lai, Z., Wong, W.K., Xu, Y., Yang, J., Zhang, D., 2016. Approximate orthogonal sparse embedding for dimensionality reduction. IEEE Trans. Neural Netw. Learn. Syst. 27 (4), 723–735.
Leenders, R.T.A., 2002. Modeling social influence through network autocorrelation: constructing the weight matrix. Soc. Netw. 24 (1), 21–47.
Leifeld, P., Cranmer, S.J., Desmarais, B.A., 2015. Estimating temporal exponential random graph models by bootstrapped pseudolikelihood. R Package Vignette for Xergm 1.4. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.432.4592&rep=rep1&type=pdf.
Leifeld, P., Cranmer, S.J., Desmarais, B.A., 2018. Temporal exponential random graph models with btergm: estimation and bootstrap confidence intervals. J. Stat. Softw. 83 (6).
Leskovec, J., Adamic, L.A., Huberman, B.A., 2007a. The dynamics of viral marketing. ACM Trans. Web (TWEB) 1 (1), 5.
Leskovec, J., Backstrom, L., Kleinberg, J., 2009. Meme-tracking and the dynamics of the news cycle. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 497–506.
Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N., Hurst, M., 2007b. Patterns of cascading behavior in large blog graphs. Proceedings of the 2007 International Conference on Data Mining 551–556.
Leydesdorff, L., 2008. On the normalization and visualization of author co-citation data: salton's Cosine versus the Jaccard index. J. Am. Soc. Inf. Sci. Technol. 59 (1), 77–85.
Lin, H.-F., 2007. Effects of extrinsic and intrinsic motivation on employee knowledge sharing intentions. J. Inf. Sci. 33 (2), 135–149.
Lottridge, D., Bentley, F.R., 2018. Let's hate together: how people share News in messaging, social, and public networks. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems 60–73.
Ma, H., Yang, H., Lyu, M.R., King, I., 2008. Sorec: social recommendation using probabilistic matrix factorization. Proceedings of the 17th ACM Conference on Information and Knowledge Management 931–940.
Marsden, P.V., Friedkin, N.E., 1993. Network studies of social influence. Sociol. Methods Res. 22 (1), 127–151.
McAuley, J.J., da Fontoura Costa, L., Caetano, T.S., 2007. Rich-club phenomenon across complex network hierarchies. Appl. Phys. Lett. 91 (8), 084103.
McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a feather: homophily in social networks. Annu. Rev. Sociol. 27 (1), 415–444.
Memon, N., Larsen, H.L., Hicks, D.L., Harkiolakis, N., 2008. Retracted: detecting hidden hierarchy in terrorist networks: some case studies. Proceedings of the International Conference on Intelligence and Security Informatics 477–489.
Messing, S., Westwood, S.J., 2014. Selective exposure in the age of social media: endorsements trump partisan source affiliation when selecting news online. Communic. Res. 41 (8), 1042–1063.
Myers, S.A., Zhu, C., Leskovec, J., 2012. Information diffusion and external influence in networks. Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 33–41.
Nematzadeh, A., Ferrara, E., Flammini, A., Ahn, Y.-Y., 2014. Optimal network modularity for information diffusion. Phys. Rev. Lett. 113 (8), 088701.
Newman, M.E., 2003. The structure and function of complex networks. Siam Rev. 45 (2), 167–256.
Nielsen, J., 2006. Participation inequality: encouraging more users to contribute. Jakob Nielsen's Alertbox 9, 2006.
Oltulu, P., Mannan, A.A.S.R., Gardner, J.M., 2018. Effective use of Twitter and Facebook in pathology practice. Hum. Pathol. 73, 128–143.
On-at, S., Quirin, A., Péninou, A., Baptiste-Jessel, N., Canut, M.-F., Sèdes, F., 2017. A parametric study to construct time-aware social profiles. Proceedings of Trends in Social Network Analysis 21–50.
Page, L., Brin, S., Motwani, R., Winograd, T., 1999. The PageRank citation ranking: Bringing order to the web. Stanford InfoLab.
Pariser, E., 2011. The Filter Bubble: What the Internet Is Hiding From You. Penguin, UK.
Pastor-Satorras, R., Castellano, C., Van Mieghem, P., Vespignani, A., 2015. Epidemic processes in complex networks. Rev. Mod. Phys. 87 (3), 925.
Plötner, M., Over, H., Carpenter, M., Tomasello, M., 2015. Young children show the bystander effect in helping situations. Psychol. Sci. 26 (4), 499–506.
Qian, G., Sural, S., Gu, Y., Pramanik, S., 2004. Similarity between euclidean and cosine angle distance for nearest neighbor queries. Proceedings of the 2004 ACM Symposium on Applied Computing 1232–1237.
Quattrociocchi, W., Scala, A., Sunstein, C.R., 2016. Echo Chambers on Facebook. Available at SSRN 2795110. .
Rafaeli, S., Ravid, G., Soroka, V., 2004. De-lurking in virtual communities: a social communication network approach to measuring the effects of social and cultural capital. Proceedings of the 37th Annual Hawaii International Conference on System Sciences 10.
Ravid, G., Rafaeli, S., 2004. Asynchronous discussion groups as small world and scale free networks. First Monday 9 (9).
Richterich, A., 2014. 'Karma, precious karma!'Karmawhoring on Reddit and the Front Page's econometrisation. J. Peer Prod. 4 (1).
Riedl, C., Bjelland, J., Canright, G., Iqbal, A., Engø-Monsen, K., Qureshi, T., Sundsøy, P., Lazer, D., 2018. Product diffusion through on-demand information-seeking behaviour. J. R. Soc. Interface 15 (139), 20170751.
Rodriguez, M.G., Balduzzi, D., Schölkopf, B., 2011. Uncovering the temporal dynamics of diffusion networks. Proceedings of the 28th International Conference on Machine Learning (ICML'11) 2011, 561–568.
Rodriguez, M., Leskovec, J., Krause, A., 2012. Inferring networks of diffusion and influence. ACM Trans. Knowl. Discov. Data (TKDD) 5 (4), 21.
Rogers, E.M., 2003. Diffusion of Innovations, 5th ed. Free Press.
Rogers, E.M., Kincaid, D.L., 1981. Communication Networks: Toward a New Paradigm for Research. Free Press, New York.
Romero, D.M., Meeder, B., Kleinberg, J., 2011. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. Proceedings of the 20th International Conference on World Wide Web 695–704.
Rossi, R., Gallagher, B., Neville, J., Henderson, K., 2012. Role-dynamics: fast mining of large dynamic networks. Proceedings of the 21st International Conference on World Wide Web 997–1006.
Rossi, R.A., Ahmed, N.K., 2015. Role discovery in networks. IEEE Trans. Knowl. Data Eng. 27 (4), 1112–1131.
Shi, X., Zhu, J., Cai, R., Zhang, L., 2009. User grouping behavior in online forums. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 777–786.
Sonnenbichler, A.C., 2010. A Community Membership Life Cycle Model. arXiv preprint arXiv:1006.4271.
Stigers, R.L., 2011. Online Social Network Behaviors As Predictors of Personality. Citeseer. M. S. thesis. California State University, Chico, CA.
Strauss, D., Ikeda, M., 1990. Pseudolikelihood estimation for social networks. J. Am. Stat. Assoc. 85 (409), 204–212.
Sun, E., Rosenn, I., Marlow, C., Lento, T.M., 2009. Gesundheit! Modeling contagion through facebook news feed. Proceedings of the 3rd International Conference on Weblogs and Social Media 146–153.
Snijders, T.A., 2002. Markov chain Monte Carlo estimation of exponential random graph models. J. Soc. Struct. 3 (2), 1–40.
Tomasello, M., Melis, A.P., Tennie, C., Wyman, E., Herrmann, E., 2012. Two key steps in the evolution of human cooperation. Curr. Anthropol. 53, 673–692.
Ugander, J., Backstrom, L., Marlow, C., Kleinberg, J., 2012. Structural diversity in social contagion. Proc. Natl. Acad. Sci. 109 (16), 5962–5966.
Vaidya, M., 2014. Ice bucket challenge cash may help derisk ALS drug research. Nat. Methods 20 (10), 1080–1181.
Valente, T.W., Davis, R.L., 1999. Accelerating the diffusion of innovations using opinion leaders. Ann. Am. Acad. Pol. Soc. Sci. 566 (1), 55–67.
Valente, T.W., Dyal, S.R., Chu, K.-H., Wipfli, H., Fujimoto, K., 2015. Diffusion of innovations theory applied to global tobacco control treaty ratification. Soc. Sci. Med. 145.
Van Duijn, M.A., Gile, K.J., Handcock, M.S., 2009. A framework for the comparison of maximum pseudolikelihood and maximum likelihood estimation of exponential family random graph models. Soc. Netw. 31 (1), 52–62.
Velasquez, A., 2012. Social media and online political discussion: the effect of cues and informational cascades on participation in online political communities. New Media Soc. 14 (8), 1286–1303.
Wang, F., Wang, H., Xu, K., 2012. Diffusive logistic model towards predicting information diffusion in online social networks. Proceedings of the 32nd International Conference on Distributed Computing Systems Workshops (ICDCSW) 133–139.
Wang, H., Wang, C., Zhai, C., Han, J., 2011. Learning online discussion structures by conditional random fields. Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval 435–444.
Wang, W., Tang, M., Shu, P., Wang, Z., 2016. Dynamics of social contagions with heterogeneous adoption thresholds: crossover phenomena in phase transition. New J. Phys. 18 (1), 013029.
Wang, W., Tang, M., Zhang, H.-F., Lai, Y.-C., 2015. Dynamics of social contagions with memory of nonredundant information. Phys. Rev. E 92 (1), 012820.
Watts, D.J., 2002. A simple model of global cascades on random networks. Proc. Natl. Acad. Sci. 99 (9), 5766–5771.
Watts, D.J., Dodds, P.S., 2007. Influentials, networks, and public opinion formation. J. Consum. Res. 34 (4), 441–458.
Weninger, T., Zhu, X.A., Han, J., 2013. An exploration of discussion threads in social news sites: a case study of the reddit community. Proceedings of the International

Conference on Advances in Social Networks Analysis and Mining (ASONAM) 579–583.
White, S., Smyth, P., 2003. Algorithms for estimating relative importance in networks. Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 266–275.
Will, T.E., 2016. Flock leadership: understanding and influencing emergent collective behavior. Leadersh. Q. 27 (2), 261–279.
Xie, K., Yu, C., Bradshaw, A.C., 2014. Impacts of role assignment and participation in asynchronous discussions in college-level online classes. Internet High. Educ. 20, 10–19.
Yang, D., Sinha, T., Adamson, D., Rosé, C.P., 2013. Turn on, Tune in, drop Out: anticipating student dropouts in massive Open online courses. Proceedings of the 2013 NIPS Data-Driven Education Workshop 11, 14.
Yang, J., Leskovec, J., 2010. Modeling information diffusion in implicit networks. In: Proceedings of the 2010 International Conference on Data Mining. IEEE. pp. 599–608.
Yin, M., Gray, M.L., Suri, S., Vaughan, J.W., 2016. The communication network within the crowd. Proceedings of the 25th International Conference on World Wide Web 1293–1303.
Zhang, Z.-K., Liu, C., Zhan, X.-X., Lu, X., Zhang, C.-X., Zhang, Y.-C., 2016. Dynamics of information diffusion and its applications on complex networks. Phys. Rep. 651, 1–34.