Member Behavior in Dynamic Online Communities: Role Affiliation Frequency Model

Alon Bartal and Gilad Ravid

Abstract—People's social life has become more embedded in dynamic online communities. Each online community can be viewed as a temporal online social network (OSN). The interaction level among OSN members leads to the emergence of dynamic social roles, which change and evolve over time, creating a sequence of temporal roles. These role sequences show diversity in the role-affiliation frequency of members. That diversity enables modeling the dynamic behaviors of individuals. This paper proposes a temporal role-affiliation frequency model (RAFM) which detects the time evolving roles of each member and analyzes her/his role-affiliation frequency to infer her/his latent behavior. Applying the RAFM to real interaction data, collected in four online communities, revealed the identity of influential members. In addition, members with similar temporal behavioral patterns were found to have similar latent behavior patterns. These patterns are manifested via similar role transitions in different OSNs whose temporal interaction rhythms were compatible. These two research findings contribute to OSN research and knowledge via improved understanding of member behavior online based on role-affiliation frequency and role transitions. Thus, member latent behavior can be inferred, and influential members can be identified.

Index Terms—Dynamic online community, influential members, online behavior, role frequency, social roles

1 Introduction

Social life is becoming more embedded in online applications. People spend increasing amounts of time online and perform many tasks in online environments. For example, in a dynamic online social network (OSN), which changes continuously, individuals may exchange information, opinions [1], or find dating partners [2].

Since some OSNs are public, it is possible to observe and explore the dynamic social interactions of members [3] which define their roles [4], [5] and enable analysis of their behavior online [6]. Role discovery is important for many applications such as predicting network links [5].

Although there is no consensus regarding the definition of a social role, some define the notion of a role within a social structure in the sense that members who affiliate with the same role have similar structural patterns [7].

Some studies [e.g., 8] identify and learn the structural positions of members with predefined explicit roles. Other studies [9], [10] analyze structural interaction patterns with unsupervised techniques to uncover implicit roles.

Most role-discovery methods, developed for static networks [6], [11], [12], [13], [14] are inapplicable to dynamic OSNs [9]. Moreover, only a few researchers [e.g., 15] analyzed the changing roles in dynamic OSNs. Modeling dynamic roles is important since member behavior in dynamic OSNs is time-related [5], [16], [17].

Based on the Mixed Membership Stochastic Blockmodel (MMSB) [10], the dynamic MMSB (dMMSB) [15] models role evolution for up to medium dynamic networks. Thus, online role-discovery methods that assign and update role affiliations as the network changes are inapplicable to large OSNs and remain an open area of research [18].

Hence, our goal is to develop a role-affiliation frequency model (RAFM). The RAFM focuses on identification of the temporal role affiliation of an OSN member over time which creates a role-affiliation frequency vector. The proposed model infers member latent behavior that can be highly affiliated with a single role out of one's role sequence. Thus, RAFM aims to discover each member's underlying role-affiliation frequency, describing her/his behavior in the network. Moreover, the RAFM infers the latent behavior of members.

The novel contributions of this study are: 1) identifying the behavioral patterns of members, manifested by role transitions according to their role-affiliation frequencies, 2) showing that members with similar role-affiliation frequencies share the same latent role in different OSNs whose dynamics in terms of interaction rhythms are similar. Additional contributions of this study include: 1) analyzing large, dynamic, and time-evolving networks, 2) basing role discovery only on structural network features, and 3) identifying influential members.

Organization. The next section sets the theoretical background for model development in Section 3. Research hypotheses are presented in Section 4. The datasets are presented in Section 5. The results are presented in Section 6. Finally, conclusions, limitations, and future work are discussed in Section 7.

• *The authors are with the Department of Industrial Engineering and Management, Ben-Gurion University, Beer-Sheva 84105, Israel. E-mail: bartala@post.bgu.ac.il, rgilad@bgu.ac.il.*

Manuscript received 8 Feb. 2018; revised 5 Mar. 2019; accepted 9 Apr. 2019. Date of publication 15 Apr. 2019; date of current version 5 Aug. 2020. (Corresponding author: Alon Bartal.) Recommended for acceptance by G. Chen. Digital Object Identifier no. 10.1109/TKDE.2019.2911067

2 Related Work

Social roles were first discovered in sociology research [19] aiming to explain, for example, member behavior in a community [20]. The broad concept of a community has several definitions: 1) people who interact with one another, 2) sharing common values, beliefs, or behaviors, and 3) interaction within a bounded geographic area [21].

Recent studies use the structure of complex networks to define a community as a group of members, with more connections inside the group and fewer connections outside the group [11]. Structural network analysis can detect different groups of communities [3], [5] and roles [5], [11], [16], [17]. Although related, these groups are of different types that call for applying different detection methods [11]. Community detection methods that are based on the structure of the network, use connectivity patterns of link density among members [22]. Examples of such methods include modularity [22], edge removal [23], graph partitioning [24], and similarity [22]. Community detection methods that are based on probabilistic definitions allow members to affiliate with multiple communities [10]. Research on these methods is still in their infancy [5] and include, for example, community detection by dynamic roles [3] which expend community detection by member roles that capture the contribution of a member to the community. Costa et al., [5] proposed two variations of a Bayesian Hierarchical Latent-Factor model (BHLFM) to discover latent factors that define communities and roles [3], [5] by uncovering hidden affiliation of members to multiple communities and roles.

Role detection methods aim to explain member behaviors through interactions since members affiliated with the same role have similar structural patterns [7]. Roles are thus not required to be connected or close in terms of network density or distance [5], [17] and, hence, a member can play different roles and belong to different communities. However, members in the same community must be close to one another. Recently, community discovery and role assignment were jointly modeled [3], [5], allowing members to belong to multiple communities and roles respectively. This study focuses on analyzing the observed behaviors of members through role detection in online communities.

Online communities enable tracking and analyzing the behaviors of members and their relationships [3]. These relationships are maintained in OSNs such as online forums, and microblogging (e.g., Twitter) [25] where people communicate mostly by text messages [26]. Members' temporal relationships in dynamic OSNs change and evolve [27], and can be characterized as roles which describe member behavior [28]. The importance of roles is acknowledged in many applications such as predicting new network links [5], and modeling dynamic member behavior in large graphs [7], [29].

2.1 Member Roles in Online Social Networks

A community can be represented by a graph of interactions within a group similar to a social network or an OSN [3]. It was found that each role in an OSN has a unique signature that can be defined by observing the patterns of member participation [30]. In most online communities, 90 percent of members never contribute, 9 percent contribute little, and 1 percent account for almost all network activity [31], which implies the member roles. Thus, members who actively participate, and create rich content can highly affect information spread in OSNs [32] and influence other members to participate [33]. However, members are not born influencers; they transition between roles [34].

Members with a *Visitor* role are interested in the community and consume information, yet contribute nothing to the discussions [35]. A Visitor can become a *Novice* by participating in discussions [35]. For a Novice to become an *Activist*, s/he must often participate by consuming and producing information [35]. An Activist can become a *Leader* in a community, also known as a gatekeeper [36], by assuming the role of an opinion maker. A member becomes *Passive* if s/he maintains interest in discussions and in other members, consuming rather than producing content. Lastly, a *Troll* disturbs other members by posting offensive content and causing conflicts in a relatively short period prior to vanishing [37].

Roles that emerge from the structure of the community [17] can be identified by considering network centralities (e.g., Indegree, Outdegree, Closeness, Betweenness, Eigenvector [38]) or ranks (e.g., PageRank, HITS), which uncover structural information about member relations within the community [39]. For example, in the 9/11 terrorist attacks [40], centrality measures were used to identify influencers who played a variety of roles (e.g., "brokers" and "gatekeepers"). Moreover, centrality measures were used to identify member importance in the network [41] (e.g., articulation nodes and selecting the top-k nodes [42]). However, other studies found no correlations between the roles of members and their structural positions [e.g., 43]. This might be since member behavior is dynamic, and richer than what can be captured just by whether or not s/he is highly influential. In addition, since members frequently change roles [15], it is important to discover diverse roles by using member dynamic structural relations.

2.2 Role Discovery by Network Structure

Role detection in a network often involves using unsupervised techniques to assign structurally similar members to the same cluster. A hard assignment assigns a member to a single role [44], whereas a soft assignment allows a member to play multiple roles [9], [10], [15].

Since roles can be inferred from member interactions, as reflected in the network structure, several algorithms classify members into roles using the structure of the graph [15], [45]. Blockmodel algorithms are used to detect roles in social networks [10], [15], [44] by identifying members with equivalent network structures [28] or measure the extent to which they are structurally equivalent [46], [47].

Another approach for detecting roles is the latent position cluster model [45]. This model estimates unknown role affiliations by a set of probabilities of edge formation between members where each member is associated with exactly one role. This single role affiliation assumption was extended in the MMSB by allowing a member to affiliate with several roles [10]. However, the MMSB assumes a fixed network structure, whereas OSNs are dynamic. Thus, dMMSB [9] and a few other models [3], [5] consider network evolution, where members play various roles that evolve over time. These definitions imply that members with similar roles,

share common features and relation patterns, even without direct relationships [16], [47].

An important aspect that has not been thoroughly addressed is role-evolution modeling in dynamic OSNs which considers member time-dependent role transitions.

The RAFM considers dynamic networks where each member is affiliated with a role at each time step. Thus, a member can play multiple roles similar to a fuzzy logic approach [48]. Such approach in the RAFM assumes that the latent behavior of a member can take on a different blend of social roles (for example, s/he can have some Leader qualities, but mainly be an Active member) that can change over time. However, measuring a network at a given time step creates a static snapshot of the network.

Hence, at each snapshot only the role that a member plays during that time step, out of the blend of social roles is observed. Considering the temporal roles played, each member can be highly affiliated with a single most descriptive role (i.e., latent role) out of the blend of social roles, based on the role degree affiliation inferred by analyzing the member's role-affiliation frequency, as will be further explained in Section 3.

As discussed, dMMSB [9] analyzes dynamic networks. However, it has several drawbacks: 1) it assumes a specific parametric form, where the groups are defined based on linkages to specific members rather than on the more general behaviors of members, 2) it assumes a fixed number of members in the network, and 3) it cannot handle large networks in a feasible time [15]. Few studies identified roles in large dynamic networks. For example, Gupte and Ravindran [29] used MapReduce to identify roles but, their algorithm considered large sparse graphs with a small number of time steps (three). Rossi, et al. [49] generated a role transition model per member. However, the way in which they defined roles is not intuitive, making the descriptive behavior of each role hard to understand. Thus, identifying and comparing the same roles (e.g., influencers) in different OSNs is impossible. Hence, they found role meaning by network centralities after identifying roles.

Addressing these limitations, five main innovations are made by developing the RAFM. *First*, role detection by RAFM is based only on network relationships and, therefore, the underlying dynamic process is very intuitive to understand based on the interpretation of network centralities. *Second*, the RAFM is applicable to dynamic and large time-evolving networks without restricting the number of members or relations by handling the network in small time steps. *Third*, unlike other models (e.g., blockmodels), the RAFM captures general member behavior without restriction by a member's structural position. It considers the general behavior of members by grouping members with similar temporal behavior according to network activity (centralities), as well as by the role evolution of each member who may affiliate with several roles over time. *Fourth*, the RAFM infers a member's latent role, which is the role with which s/he is most affiliated with, based on analysis of her/his role-affiliation frequency. This inference reveals that, members with the same latent role in different OSNs have similar role paths. *Lastly*, the RAFM is applicable to detection of influential members by identifying role transitions and predicting structural changes in the network.

The temporal dynamics of an OSN can be analyzed to better understand member behavior online. Lack of sufficient knowledge about the relationship between online roles and behaviors raises the need for new models that, via role identification, will enable better analysis of interactions in OSNs. Thus, our goal is to model time-dependent changes in members' behavioral patterns in dynamic OSNs, to better understand member behavior.

3 Model Development

Given a dynamic OSN, for each network observation between time steps $k-1$ and k (denoted t_k), a graph $G_k = (V_k, E_k, w_k)$ is constructed, where nodes (V_k) are active members who posted at least once at t_k, directed edges (E_k) represent interactions between members, such that $e_{ij} \in E_k$ indicates that member i wrote a post directed to j at t_k, and weights (w_k) of edges indicate interactions frequency.

The two goals of the RAFM are to discover a set of underlying roles in the network, representing online behavior of members at t_k by their role affiliations, and to observe the role evolution of members over time, inferring their latent roles. These goals are achieved by four steps.

Step #1 estimates the role affiliations, which represent behavior patterns, for each member in each observation $G_1, \ldots, G_k$, by defining the features of each role and representing each member in G_k by a set of features. These features are derived from member activity and include the following centralities: $D = \{\text{Indegree}, \text{Outdegree}, \text{Betweenness}, \text{Closeness}, \text{Eigenvector}\}$. The design philosophy for using these centralities is explained in the following paragraphs. To account for temporal network dynamics, the normalized centrality measures at t_k of each member: $C_{d,\,k}(v_i),\ v_i \in V_k,\ d \in D$, in every OSN observation (G_k) are calculated first. Then, the average (1) and standard deviation (2) of the centralities are calculated for each G_k.

$$Avg\big(C_{d,k}\big) \ := \bar{C}_{d,k} = \frac{\sum_{i=1}^{|V_k|} C_{d,k}(v_i)}{|V_k|}\ ,\ d \in D \tag{1}$$

$$std\big(C_{d,k}\big) := \ \sigma_{d,k} = \sqrt{\frac{\sum_{i=1}^{|V_k|} \left(C_{d,k}(v_i) - \bar{C}_{d,k}\right)^2}{|V_k|}}\ ,\ d \in D. \tag{2}$$

Next, member role affiliation in each G_k is estimated by comparing her/his centralities to (1) and (2). This facilitates the definition and detection of a member's role at t_k, and the creation of the role-detection sub-model (Fig. 1) of the RAFM. Fig. 1 facilitates the classification of members into five potential roles, which is also the number of roles used by Fu, et al. [9], and Choobdar, et al. [32] who analyzed member roles in spreading information in OSNs.

In Fig. 1, Leaders (role 4), who are very active by our definition, can be considered either opinion leaders or influencers [1], [50]. They are highly relevant to understanding the diffusion of topics [51], as they promote discussions, and actively create content, which can be measured by high Outdegree values [52]. Due to their high activity levels, Leaders have strong ego-networks and close relationships with others, as can be measured by high Indegree value [52]. Opinion leaders in the network will most likely have high

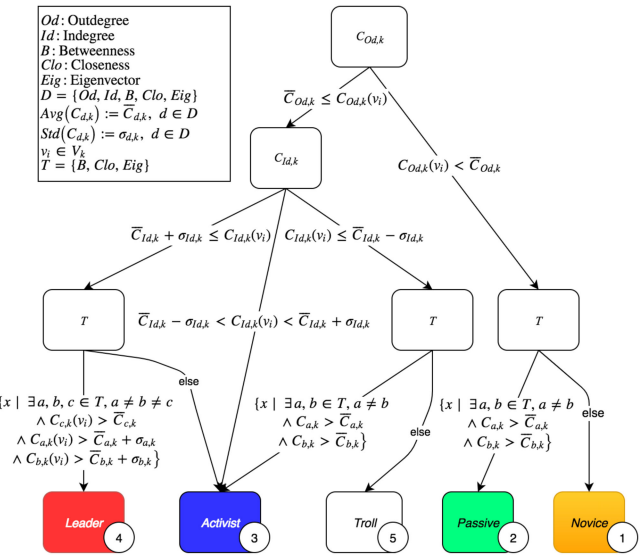


Fig. 1. Role-detection sub-model.

centrality measures. Therefore, Leaders are expected to have Outdegree ($C_{Od,\,k}$) values that are greater than or equal to the network's average Outdegree ($\bar{C}_{Od,\,k}$) value [34], and to have high Indegree ($C_{Id,\,k}$) values [52] that are greater than or equal to the network's average Indegree value plus one standard deviation ($\bar{C}_{Id,\,k} + \sigma_{Id,\,k}$). Leaders are also expected to have Betweenness, Closeness, and Eigenvector degree centralities that are greater than the network's average centralities [52]. However, Leaders do not always possess the highest centrality measures [42]. Thus, this expectation is relaxed by allowing at least two out of the three later values to be greater than or equal to the network's average centralities plus one standard deviation.

Active members (role 3) consume and produce information [34], and have high Outdegree and Indegree. They are less active than Leaders, as they foster existing relationships and gain new contacts who are mostly within their local network, rather than the whole network [34]. Therefore, they are expected to have high Betweenness, Closeness, and Eigenvector values, but less than Leaders.

A Troll (Role 5) expresses antisocial behavior [53], disturbing other members with offensive posts and causing conflicts over a relatively short time before vanishing [37]. Therefore, their activity level is high at the beginning but decreases to a very low level toward disappearance. Compared to the average network values, Trolls are expected to have high Outdegree values, but, being unaccepted by others, they have low Indegree, Betweenness, Closeness, and Eigenvector values.

Passive (role 2) and Novice (role 1) members have lower activity levels than others [34]. Passive members enjoy the community, stay in loose contact with other members and, hence, have low Outdegree values. However, they produce more content than Novices, who observe the community from the outside [43]. As such, Novices have a very sparse network, which is characterized by low centralities. In other words, Passive and Novice members have lower Outdegree values than the average Outdegree value for the network. However, since Passive members are more active than Novices, at least two out of the Betweenness, Closeness, and Eigenvector centralities for Passive members are equal to or above the average centrality values for the whole network, with no restrictions on the value of the third centrality. Novices have the lowest network centralities.

Step #2 is designed to analyze the affiliation frequency of a member with each role at every time-step. Given a sequence of dynamic behaviors $G_1, \ldots, G_k$, how behavior in the network changes over time is learned by forming a role frequency matrix $M \in \mathbb{R}^{mXr}$, where m indicates the number of members and r indicates the number of roles. Each entry in M represents the frequency with which a member plays a role along all time-intervals. Step #2 is intended to automatically discover the number of groups with similar role affiliation patterns (common patterns of behavior) based on member characteristics, on which the roles are defined, and on the role frequency vector of each member in M. The number of groups (clusters) in M are inferred using the NbClust R package [54]. Each cluster represents a latent role that best describes the behavior of its affiliated members and can be highly affiliated with one role out of the roles from Fig. 1. The latent role explains the temporal observed behavior of a member. Note, M is not limited to the sub-model (Fig. 1) and can use any role model.

Step #3 is designed to assign members to groups that were discovered in Step #2. The role frequency vector of each member, which is represented by a row vector in M, is clustered using k-means [8], [49]. Clustering is accomplished by the partition of members into disjoint groups (i.e., clusters), where members affiliated with the same cluster exhibit similar behavior in terms of the network activity levels (i.e., centrality values) and temporal behavior (i.e., role-affiliation frequency). For role frequency vectors $\overrightarrow{V_i}, \overrightarrow{V_j} \in M$ of members i and j, the k-means classifies members with similar vectors to the same cluster.

We innovate by considering two similarity dimensions as indicators of behavioral similarity of members in playing a role: structural centrality similarity (expressed by member role) and frequency similarity (expressed by M), which account for temporal effects. To uncover the unique latent behaviors of members in the same cluster, the role transitions within each cluster are analyzed next.

Step #4 is designed to analyze the role transitions in each of Step #3's groups as graphs (G_r), whose nodes represent roles, directed edges represent transitions between roles, and weights represent role transition frequencies. G_r presents the next role of a member given the role attained in the current network observation.

A member's role frequency, captured by the matrix M, helps infer her/his latent role. As depicted in Fig. 1, a member's role is affected by member activity which is more likely to occur due to current events than due to historical events [55]. Moreover, future member behavior is better predicted by recent member activity than by activities in the distant past [17]. The recent behavior of a member, expressed by a role in G_k can, therefore, capture the temporal member behavior and infer a future role in G_{k+1}.

If a member is active at t_k, but inactive at t_{k+1}, then s/he is given at t_{k+1} a role of zero. If a member is not active at t_k but is active at t_{k+1}, then s/he is given at t_k a role of zero. Role 0 indicates zero activity, with the aim of uncovering the unique structural transition patterns between roles in

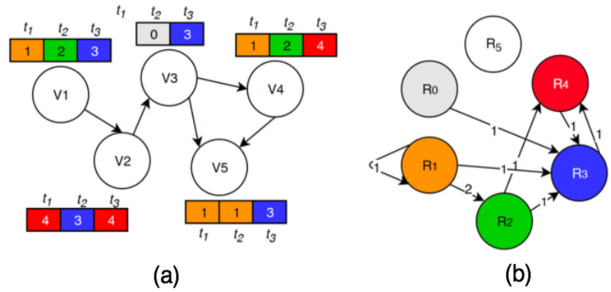


Fig. 2. The dynamic role assignment of members. (a) A network with members role sequence in three time-steps. (b) Role transition graph of Fig. 2a, R_i denotes role i based on Fig. 1.

the same cluster and, thus, identify the latent behaviors of members. Note that different members may have role sequence vectors of different sizes, since members are active during some, but usually not all, time steps.

Fig. 2a illustrates an OSN, where nodes represent active members, edges represent interactions between members, and the member role affiliation in each time step $[t_1, t_3]$ is indicated by a role number inside a box array. For example, node v_3 was inactive at t_1, t_2 but was active at t_3. Therefore, v_3 has no role assignment at t_1, and a zero role at t_2, since it was active at t_3 (role 3). Fig. 2b summarizes the member role transitions of Fig. 2a by illustrating the G_r graph structure. In Fig. 2a, v_1 and v_4 are the only nodes that transitioned from role 1 to role 2. Hence, G_r in Fig. 2b has a directed edge with a weight of 2 from node R_1 to R_2.

Since a member can play any role, s/he can transition between any two roles. To capture the representative patterns of role transitions [27] that correspond to the latent behavior in each cluster, only key role transitions are accounted for in Step #4. Thus, in G_r, the only edges (i.e., role transitions) considered are those that have a formation probability greater than would be expected at random. In a random process, each edge occurs independently with a uniform creation probability [56]. Comparing a graph with a random graph not only facilitates the discovery of functional dependencies that correspond to dominant processes in the formation of the graph but also characterizes relationships among roles. Fig. 3 summarizes the steps of the RAFM.

In the context of the background and model development, three hypotheses are addressed in this work.

4 Research Hypotheses

The RAFM seeks to model behavior and member role affiliation in dynamic OSNs. Although important, role discovery is not the only goal of this research. Examining role evolution (defined by members' dynamic interactions that create a sequence of played roles) which affects member role-affiliation-frequency, will allow us to deepen our understanding of individuals behavior in temporal and dynamic OSNs. The detailed data contained in these OSNs about member interactions, enable studying collective and individual interactions of members. Hence, three hypotheses are proposed:

H1. The social behavior of members can be modeled via the role affiliations exhibited in online discussions.

Moreover, analysis of time-evolving OSNs might reveal similar role-affiliation frequencies of different members along time, leading us to hypothesize the following:

H2. Members with similar temporal role-affiliation frequencies have similar behavioral patterns.

The latent behavior of a member explains her/his temporal observed behavior. Thus, members in different OSNs who share the same latent behavior might show similar behavioral patterns.

H3. Members with the same latent roles, particularly influential members, exhibit similar social behaviors, expressed through similar role transitions.

To answer the three research hypotheses the RAFM was applied and validated on real datasets, as described next.

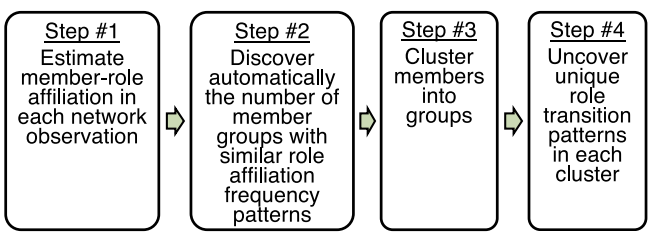


Fig. 3. Summary of the four steps in the RAFM.

TABLE 1
Dataset Description

Parameter/ Dataset	Apple	Tapuz	Twitter(a)	Twitter(b)
# Networks	1	320	1	1
Interval size	2-weeks	2-weeks	1-hour	1-hour
# Observations	98	11,199	168	360
# Nodes	11,118	564,877	456,626	267,043
# Weighted edges	37,545	1,511,833	985,590	1,438,918
Processing time [min]/network	9.41	4.47	525.41	308.56

5 Datasets Description

Datasets from four OSNs are analyzed in this study. Two online forum-hosting sites and two Twitter networks.

Members of these dynamic OSNs infrequently interact, with addition and deletion of nodes and edges occurring over time. By analyzing these human communication datasets, temporal processes that indicate role evolution are observed. Since the sequence of discrete temporal events can be interpreted as a time-dependent point process [57], member interactions were split by time intervals. 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. The datasets are described next, and further details are provided in Table 1.

The *first* dataset contains written communications posted on *Apple's technical support forum* between September 19, 2003 and October 19, 2007 [58]. Members of this forum seek help or provide help to others about technical problems.

The *second* dataset is a large online community named *Tapuz*, which operates sub-forums under the same domain, with an average of 3.9 million monthly visitors (60 percent female). Written communications posted in 521 sub-forums between January 1, 2013, and November 10, 2014, were collected. As explained next, sub-forums with low activity were discarded, resulting in 320 online sub-forums with ground truth role information (i.e., explicit roles).

TABLE 2
Tapuz's K-means Centroids

Cluster #	R_1-Novice	R_2- Passive	R_3-Active	R_4-Leader	R_5- Troll	Cluster size	Latent role
1	4.1	9.4	3.4	0.6	0	11,005	Passive
2	2.5	7.3	19.3	4.6	0	2,939	Active
3	0.4	2.1	10.3	28.3	0	693	Leader
4	0.5	1.0	0.2	0	0	211,941	Novice

R_i Denotes the role number, according to Fig. 1.

Three steps helped refine the datasets #1, and #2: 1) member interactions were split into intervals of two weeks similar to [59], creating at each time step a directed network observation, represented as a graph (G_k), 2) for the Tapuz dataset, non-active forums with less than 40 consecutive observations were discarded. Altogether, 321 forums were analyzed, one forum from the first dataset and 320 forums from the second dataset, 3) remove initial time steps form each forum that contained network observations with less than 20 nodes.

The interactions in the above online forums, originate from a root post that initializes a topic for the discussion. Members read posted messages and reply to those that interest them. Each posted message contains the unique identification number of the posting member, the date and time of its posting, and its parent message.

The *third* dataset: The *Twitter(a)* dataset, contains interactions between July 1, and July 7, 2012, regarding the Higgs boson particle [60]. It contains four types of interactions: a tweet (TW), a retweet (RT), a mention tweet (MT), or a reply tweet (RE), with at least one of the following keywords or hashtags: lhc, cern, boson, higgs. The activity was split by 1-hour intervals as sufficient to capture temporal patterns, since De Domenico, et al. [60] found that the decay time scale of member activity in this dataset was $\sim$1.13 hours.

The *fourth* dataset: The *Twitter(b)* dataset, contains TW, RT, MT, RE, collected before, during and after President Trump's announcement that the US will withdraw from the Iran nuclear deal (May 1 to May 15, 2018). The locations of the tweets were limited to fall within a defined radius of a 12km from Tehran's center[1] in the geo-coordinates N35°42′ 55.073″ E51°24′15.634″. Like in Twitter(a) dataset, interactions were split by using a short interval of 1-hour.

6 Data Analysis and Results

By applying the four steps of the RAFM to the four datasets, in Step #1 a role was assigned to each member in each network observation. The member role frequencies, which are represented by matrix M, were analyzed in Step #2. Since the Tapuz dataset included 320 different forums under the same domain, each role frequency matrix $(M_1, \ldots, M_{320})$ for Tapuz was normalized by the number of observations in each forum to enable the aggregation of the 320 matrices into a single representative one. However, there was no need to normalize each single role frequency matrix in each of the other three datasets.

Next, using the NbClust R package with 26 indexes, the number of groups was inferred. NbClust searched for the best number of clusters over the M matrix of each dataset, ranging between 2 and 15 clusters, as the number of roles is usually quite small for a variety of network types [49]. For the first and second datasets a 4-cluster solution was recommended since few members affiliated with the Troll role. This finding is similar to previous studies which found that online antisocial behavior (Trolls) is uncommon [e.g., 53]. Thus, the NbClust package did not recommend assigning Trolls to a fifth cluster.

For each of the Twitter datasets, a 2-clusters solution was recommended by RAFM based on each role frequency matrix, which presented the following fraction of members in each group. Twitter(a): Novice 57.31 percent; Passive 0.02 percent; Active 42.58 percent; Leader 0.08 percent; Troll 0.009 percent; and Twitter(b): Novice 47.47 percent; Passive 0.04 percent; Active 52.44 percent; Leader 0.04 percent; Troll 0.01 percent.

Twitter members interact in a bursty nature, where members send several tweets in short periods, separated by long periods of reduced activity [57]. Thus, the Novice, and the Active groups capture a reduced and high (bursts) posting activity respectively.

Using the k-means algorithm, the row vectors of each matrix M in every dataset, were clustered in Step #3 into groups. The centroids of each cluster serve as the prototype of the cluster [49], which represents the behavior of its affiliated members, as the centroid's coordinates depend on the member role-affiliation frequency. Thus, the member's latent role can be inferred from the weights of the coordinates in each centroid. For brevity sake, Table 2 only presents centroids of the second dataset. In Table 2, the latent role of members who are classified into: 1) Cluster 1 is Passive, 2) Cluster 2 is Active, and 3) Cluster 3 is Leader. Members who are classified into Cluster 4 present similar behavioral patterns to members of Cluster 1, by mostly affiliating with the Passive and Novice roles. However, on average, the frequency at which they affiliate with each role is very low compared to members of Cluster 1. Thus, Cluster 4 corresponds to the Novice latent role. In addition, the sizes of the clusters which describe the numbers of members who were classified into each cluster, support the concept of participation inequality in OSNs [31]. Similarly, for both Twitter datasets members are clustered into two groups, corresponding to the Novice, and the Active latent roles.

A different method for classifying members to groups was tested, according to which member group affiliation was determined based on the maximum frequency with which a member affiliates to a role (instead of the k-means method) but achieved less accurate results in detecting roles, as demonstrated in the following paragraphs.

1. https://www.latlong.net/place/tehran-iran-4703.html

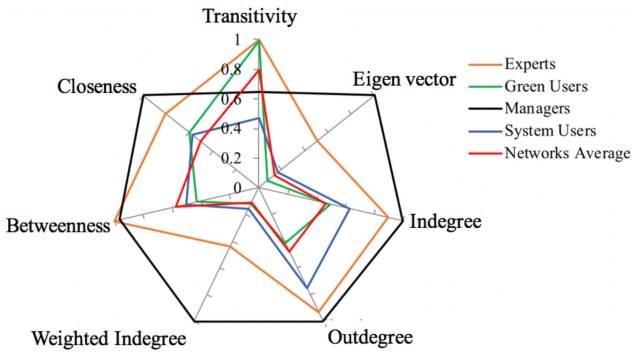


Fig. 4. Average normalized centrality measures for members with explicit roles in the second dataset.

To gain more confidence in the role-detection sub-model (Fig. 1), additional information was obtained[2] about the explicit roles that the Tapuz forum administrative staff assigned to members. Fig. 4 and Table 3 present the centrality measures, and descriptive statistics for the Tapuz members with explicit roles. Weighted Indegree (Fig. 4) was used similarly to [61] to reflect frequent interactions.

Managers and Experts were expected to be classified to the most active group (Cluster 3). However, according to Table 4, which presents a cross-tabulated view of explicit roles, and according to the cluster affiliations in the second dataset, most members with an explicit Manager role were classified by k-means not only to the most active group (Cluster 3) but also to the least active group (Cluster 1). This revealed a contradiction vis-a-vis the k-means classification. As further analysis revealed, however, there is actually no contradiction since each forum can have several Managers (Table 3), some of whom presented high-level activity while others presented low-level of activity.

In addition, although Experts were mostly classified into the Active or Leader cluster, some were classified to less active clusters. This demonstrates the OSN dynamics, as members dynamically change their ties which affects their temporal role. Thus, the k-means accurately uncovered the actual network structure in terms of the role variable, exhibited by online discussions. Therefore, the roles capture the behaviors of members according to their explicit roles, thereby supporting hypothesis *H1*.

According to the findings, each member can play any role, but a member's behavior can be highly affiliated with a single latent role, based on her/his role-affiliation frequency. For example, members who are classified into Cluster 3, as depicted in Table 2, highly affiliate with the Leader role but will play other roles as well. Since members with similar roles share common features and relation patterns [16], it is reasonable to hypothesize that members with the same role-affiliation frequency, manifested by affiliation with the same cluster, will have similar behavioral patterns in terms of role transitions $(H2)$. To test $H2$, the role transitions of members who were classified to the same cluster, implying a non-random behavior, are furthere analyzed.

According to Step #4, role transitions in each cluster are represented by a directed graph (G_r), where nodes represent the six roles (0 to 5), and weighted edges indicate the number of member transitions among roles. As explained in Section 3, an edge in G_r represents the role transitioning of a member and, in each cluster, only key transitions (edges in G_r) were considered. Five different methods to find key role transitions were tested. Insignificant results are omitted hereinafter. The first three methods tested keep only edges with weights above a corresponding weight of a "knee" located on G_r's weighted edges distribution curve [62]. The location of the point is found via analysis of: 1) the first derivative of the curve where the highest jump in value occurs, 2) the data point with the largest second derivative, or 3) the point on the curve that is furthest from a line fitted to the entire curve. Of the remaining two methods tested, which find key role transitions either by 4) analyzing G_r without removing edges, or by 5) keeping edges that have a higher frequency than would be expected at random, the latter performed best, as reported next.

For the Apple, and the Tapuz datasets, Fig. 5a illustrates the key role transitions in G_r by four heat maps, one per cluster. Fig. 5b illustrates the same for both Twitter datasets by two heat maps. The probability of transitioning or taking on the structural behavior of role j given a member's current role i and latent role (cluster) is represented by G_r. Thus, making it possible to identify future role transition of a member. For example, in Fig. 5a $\{g, h\}$, a member who plays role 1 is more likely to play role 0 at the next time step, if s/he affiliates with the Novice latent role. A member who affiliates with the Leader latent role (Fig. 5a $\{a, b\}$) and is playing role 3 at t_k is more likely to play role 3 at t_{k+1} (indicated by X) but can also transition to the leader role. The results show that members in the same cluster (have similar role-affiliation frequencies) present similar behavioral patterns (make similar role transitions) supporting $H2$ and implying that clusters represent the same latent behavior, as evaluated next.

TABLE 3
Descriptive Statistics for Members with Explicit Roles in The Second Dataset

Explicit Role	AVG	MIN	MAX	Description
Green	13.06	13	15	Members that pay a monthly fee to Tapuz company
Experts	4.68	1	110	Members with vast knowledge that help other members
System	14.99	14	15	Forum's administration staff
Managers	4.81	4	7	Members who expressed a desire to lead discussions

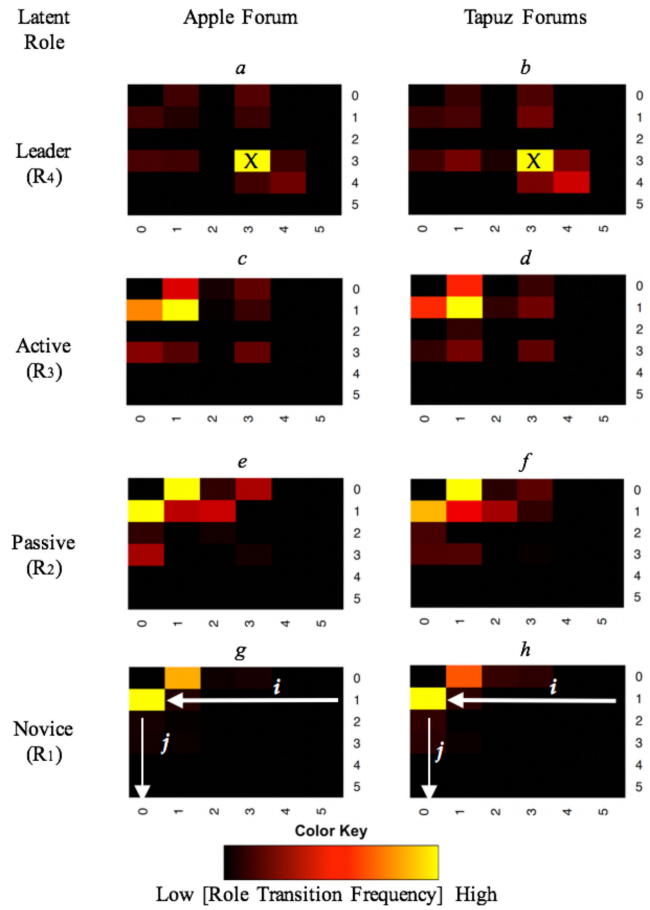
6.1 Uncover Latent Behavior

To test hypothesis $H3$, key role transitions (Fig. 5) were compared among clusters that represent the same latent role but

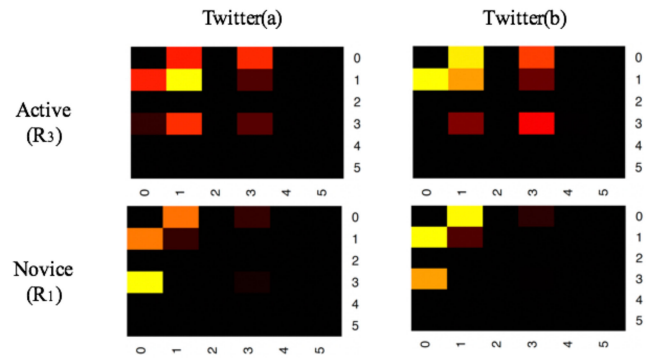
TABLE 4
Cross-Tabulation of Explicit Roles and Cluster Affiliation in The Second Dataset

	Cluster			
Explicit Role	Passive (1)	Active (2)	Leaders (3)	Novice (4)
Green	23.76%	0%	0%	76.24%
Experts	16.12%	42.51%	29.15%	12.23%
System	27.30%	1.46%	1.46%	69.79%
Managers	49.91%	15.96%	24.70%	9.43%

2. Sigal Tweig, Marketing Coordinator at Tapuz, December. 11, 2014.



(a) Apple, and Tapuz role transitions for each latent role.



(b) Twitter role transitions for each latent role.

Fig. 5. Role transitions for each latent role (cluster).

are based on different datasets. First, datasets #1 to #2 were compared in terms of the role transitions among clusters. In Fig. 5a, $\{a, b\}$; $\{c, d\}$; $\{e, f\}$; and $\{g, h\}$ were compared, thereby enabling comparison of key behavior transitions among clusters of the Apple and Tapuz communities. Similarly, for the Twitter datasets, role paths in the Active, and Novice clusters (Fig. 5b) were compared.

Although the Active (and the Novice) clusters in Fig. 5a, compared to Fig. 5b present similar patterns, the dynamic nature of Twitter is different from online forums by, for example, higher frequency of content update [63]. Moreover, Twitter members behave differently than other OSN members by maintaining weak social relationships with a high turnover of contacts in their networks [64]. Therefore, role

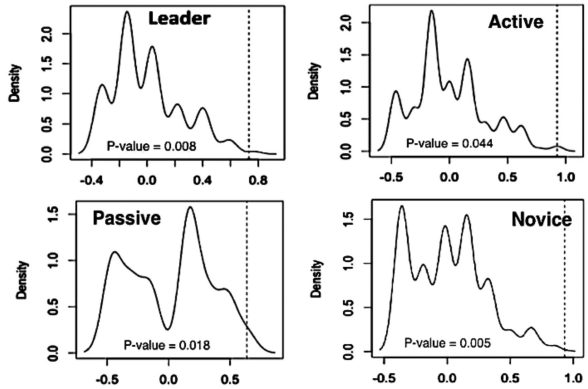


Fig. 6. Estimated QAP replications densities of Apple and Tapuz.

transitions within the online forums, and role transitions within the Twitter datasets were separately analyzed.

The relations (role transitions) in G_r are not independent and can bias significance tests of relationships between graphs [65]. Hence, a quadratic assignment procedure (QAP) test [65] was used to determine which role transitions in different datasets were related. QAP is a restricted permutation test, where each permutation preserves the auto-correlational structure of the data.

Four QAP-tests (Fig. 6), one for each pair of clusters that represent the same latent role from the Apple and Tapuz datasets were analyzed, producing significant correlations among clusters: 1) Leader 0.74, 2) Active 0.96, 3) Passive 0.68, and 4) Novice 0.91. Similarly, for the Twitter datasets, two QAP-tests were analyzed. The following significant correlations (and p-values) among clusters were found: 1) Active 0.76 (0.032), and 2) Novice 0.94 (0.004).

As mentioned before, clusters represent latent roles, and thus, members with the same latent role from different OSNs that have similar dynamic nature, make similar role transitions, supporting $H3$. The results also confirm that leaders who might be influential members have similar social behavior, expressed by similar role transitions.

Next, the utility of the RAFM is presented by predicting structural changes in a time-evolving network.

6.2 Predicting Structural Changes in a Network

The RAFM is used here to demonstrate an application for predicting structural changes in a network. The Apple dataset is analyzed here since it has more time steps than for each of the 320 Tapuz forums, and more latent roles than for each of the Twitter datasets. Thereby, enabling identification of richer behavioral patterns than each of the other datasets. The interaction level among members defines the roles that emerge from specific behaviors [4]. Thus, one can predict future member behavior by predicting the network edges, which represent member interactions. Our goal is to show that the role of a member and the latent role of a member at t_k affect her/his behavior in terms of creating new edges which, in turn, affect her/his role (Fig. 1). Given G_k, the goal is thus to predict G_{k+1} by the role and latent role of a member at t_k.

Since the dataset covers varying time-points, the modelling approach here considers the dataset's longitudinal character by applying temporal exponential random graph models (TERGMs) [66], [67]. To cope with the large 98 network

TABLE 5
Coefficients of the TERGM

Term	Coefficient	Term	Coefficient
edges	-6.22 *	mix.LatRole.1.1	0.94
mutual	4.49 *	**mix.LatRole.2.1**	**3.22** *
ttriple	0.61 *	**mix.LatRole.3.1**	**2.12** *
nodefactor. Role.2	0.33 *	**mix.LatRole.4.1**	**0.89** *
		mix.LatRole.1.2	0.49
nodefactor. Role.3	0.48 *	mix.LatRole.2.2	0.49
		mix.LatRole.3.2	0.88 *
nodefactor. Role.4	0.76*	mix.LatRole.4.2	0.68 *
		mix.LatRole.1.3	**1.03** *
nodefactor. Role.5	-0.01	mix.LatRole.2.3	1.90 *
		mix.LatRole.3.3	1.42 *
nodematch. Role	-0.25 *	mix.LatRole.4.3	0.88 *
		mix.LatRole.1.4	-0.32
autoregression	0.14 *	mix.LatRole.2.4	1.29 *
		mix.LatRole.3.4	0.53 *
		mix.LatRole.4.4	-0.37

* *Significant at the 0.05 level*

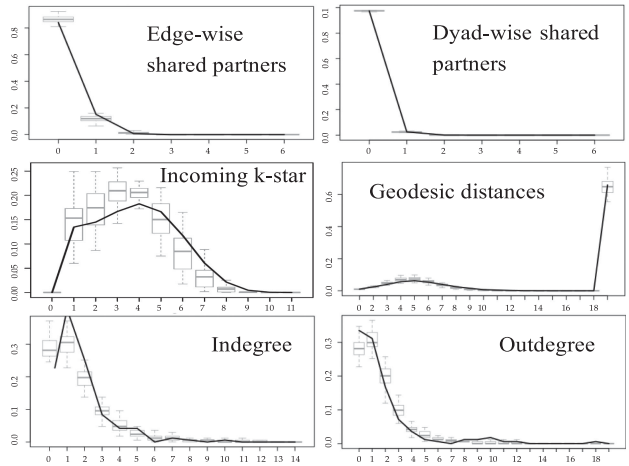


Fig. 7. Goodness of fit for TERGM using Apple's online forum.

observations (Table 1), using MCMC-MLE [67], which is an effective procedure when the networks are modestly sized with few observations, was ruled out. Instead, this study uses the btergm R package [66] which implements bootstrap methods for TERGMs, estimated by maximum pseudolikelihood [68]. For a detailed TERGM technical review, see [67], [68].

The TERGM included the following parameters (i.e., terms): *edges* – expresses a balance between creating and deleting edges; *mutual* - expresses a tendency to reciprocate edges; *ttriple* – a set of edges $(i \rightarrow j), (j \rightarrow k), (i \rightarrow k)$ that expresses the number of transitive triples in the network; *nodefactor.Role.X* – a categorical variable with five levels ($X \in [1, 5]$ in Fig. 1), which indicates the number of times that members with role X in each observed network appears within the edge set; *nodematch.Role* - captures the number of edges whose incident members match in terms of role; $mix.LatRole.X.Y, X, Y \in$ {latent role 1 to 4 in Fig. 5a} captures a tendency to form an edge between two members who affiliate with a latent role X, and a latent role Y respectively; *Autoregression* - a memory term, captures if previous edges are carried over to the current network.

The terms' coefficients (Table 5) were estimated by 1,000 replications, based on G_1 to G_{97}. Then, the model fit was examined by an out-of-sample–goodness of fit (GOF) [66], comparing simulated networks to G_{98}. The contribution of each term was tested by its significance and AUC.

According to the results, *edges* correspond to the density of the observed networks. The *mutual*, *ttriple*, *nodefactor.Role*, and the *autoregression* terms show a positive coefficient. Thus, the *mutual* term indicates that members interact with one another regarding technical problems. The *ttriple* term suggests closed structures with hierarchical ordering, which correspond to hierarchical threads regarding technical problems in Apples' forum. The *nodefactor.Role* term indicates that the more active a member, the more likely s/he to form an edge, except for role 5 (Troll), which was insignificant. Surprisingly, a negative coefficient of the *nodematch.Role* term indicates a lack of homophily. Members who affiliate with the same role are less likely to form an edge. This finding was further analyzed by separately testing member interactions by their broader representative temporal behavior (*mix.LatRole*). Indeed, Table 5 shows insignificant results for members who affiliate with the same latent role, except for latent role 3. Although some coefficients of the *mix.LatRole* term are insignificant, we can learn the behavior of each latent role from those that are significant. For example, in Table 5, the highest positive coefficient for *mix.LatRole.4.Y* is achieved when $Y = 1$. Hence, a member who affiliates with latent role 4 is more likely to form an edge with a member who affiliates with latent role 1 than with any other member who affiliates with a different latent role. Similar results (bolded in Table 5) were found for all latent roles, which indicates that all members help Novices. However, Novices (role 1) tend to interact with Active members who are considered as experts. Finally, the *autoregression* term indicates that members continue to interact by posting questions and answers over time.

The GOF test shows a good fit of the boxplots (Fig. 7) by 300 simulations of each network observation based on the TERGM. Overall, the TERGM presents good prediction results $(\mathrm{AUC} = 0.74)$. Thus, member roles and latent roles affect her/his behavior, expressed as an edge formation.

6.3 Model Performance

Let $|V|$ be the number of nodes, t be the number of timesteps, and $|E|$ be the number of edges. The most demanding algorithms in Step #1 are Betweenness and Closeness with $O(t(|V| * |E|))$ complexity. However, to better capture temporal changes [57], the network is split to observations such that the higher the network's dynamics the smaller the time interval, resulting in several small networks.

The most demanding algorithm in Step #2, which was used for performance evaluation of the NbClust package, is of complexity $O(|V| * |E|)$ [54]. In Step #3, in which members are clustered into groups using the k-means algorithm, the complexity is $O(|V| * k * i * d)$ where k is the number of clusters (a small number), i is the number of iterations until convergence (often small), and d is the number of attributes (five according to Fig. 1). Finally, Step #4 constructs the role transitions graphs with a complexity of $O(k(|V| + |E|))$. The RAFM is bounded by $O(|V| * |E|)$ since t, and k are small compared to $|V|$, and $|E|$.

In practice, the RAFM handled large networks (Table 1) by using small time-steps, where V and E tend to remain

relatively small in each network observation. Contrary to previous studies, the RAFM handled hundreds of time steps in a feasible time (Table 1). Having been applied to small or medium networks with few time steps, similar models [9], [11], [13], [15] are unable to handle large networks. The dMMSB model, for instance, handled a network with ~1,000 nodes in a day [15]. The RAFM can handle much larger dynamic OSNs, with hundreds of time steps, hundreds of thousands of nodes, and millions of edges, within 525.41 minutes on a 4 x Intel Xeon Processor E5-4610v2 2.3 GHz with 752 Gb of memory machine (without parallel processing). Moreover, network centralities can be calculated in parallel in step #1.

7 Conclusions and Limitations

Many studies on role discovery address static networks, or dynamic networks of up to medium size, or ignore the meaning of role transitions in different networks. Furthermore, many studies lack a quality evaluation of the proposed models, due to the lack of ground truth data on roles. The new behavioral model proposed here for large and dynamic networks infers member behavior based on her/his temporal role-affiliation frequency, enabling comparison of the latent behavior of members in different OSNs with similar dynamic nature. The number of latent roles (clusters) was inferred in the present study based on temporal role frequency of each member. Then, the latent role of a member was inferred by the k-means algorithm that classifies members to latent roles based on the distance between their role frequency vectors in a Euclidean 'role-frequency space'. Thus, latent roles describe a mixture of distributions of role frequency vectors, each corresponding to a cluster. Finally, the unique behavior of members in each cluster was uncovered by analyzing the role transitions of members who belong to the same cluster.

The findings from applying the RAFM to real-world OSNs, out of which 320 sub-forums in the second dataset had ground truth role information, lend support to the RAFM by validating the three research hypotheses. First, explicit roles in the second dataset were compared with the inferred roles. Then, the k-means accurately uncovered the actual network structure (exhibited by online discussions) in terms of the role variable. Thus, $H1$ was supported by showing that the roles capture members' behaviors.

Based on the four datasets, the heatmaps along with the QAP-tests results showed that member cluster affiliation depends on a member's role-affiliation frequency, supporting $H2$. Moreover, members with the same latent role from different OSNs with similar dynamic nature, make similar role transitions, regardless of the discussed topics. Hence, the next role of a member depends on the role attained in the current network and on the latent role of a member. Particularly, leaders who might be influential members have similar social behavior, i.e., similar role transitions, supporting $H3$.

RAFM's effectiveness is evident by its demonstrated ability to predict structural changes in the network via identification of member roles and latent roles, and to enable the simulation of realistic networks by considering a member's temporal role and latent role. Thus, increasing our understanding of network evolution.

Five main contributions are suggested in this study. First, analyzing large, dynamic, and time-evolving networks with a large unfixed number of nodes and edges. Second, flexible role discovery based only on structural network features makes the comprehension of the underlying dynamic process very intuitive due to the interpretation of network centralities. Third, identifying influential members. Fourth, identifying the behavioral patterns of members, as manifested by role transitions, according to their role-affiliation frequencies. Lastly, showing that members with similar role-affiliation frequencies share the same latent roles in different OSNs with similar dynamics in terms of interaction rhythms.

This study also has three main limitations. First, since members' behaviors in OSNs are influenced by the specific mechanisms that the OSN site provides, the findings regarding member activities may change as new features are added to OSNs. Second, since our analysis relies on thresholds, different findings could result upon changing the threshold of role classification. Finally, the role frequencies of a "rising star" (a member moving to the center of the network) might be similar to those of a "falling star" (doing the opposite), and thus, both will be classified to the same cluster. Overcoming these limitations is left for future research which may also investigate techniques to learn the appropriate time interval between observations automatically and to evaluate the parameter sensitivity.

The results of this study can help obtain a better understanding of member behaviors online, improve member participation in OSNs, and detect members with similar behaviors who can potentially be opinion leaders. Hence, this study contributes to the study of role detection in large, dynamic, and time-evolving OSNs.

Acknowledgments

The contribution of Professor Emeritus Nava Pliskin to the manuscript is acknowledged with gratitude. This research was partially funded by the ISF 924/16 equipment grant.

References

[1] D. J. Watts and P. S. Dodds, "Influentials, networks, and public opinion formation," *J. Consum. Res.*, vol. 34, no. 4, pp. 441–458, 2007.
[2] S. McWilliams and A. E. Barrett, "Online dating in middle and later life: Gendered expectations and experiences," *J. Family Issues*, vol. 35, no. 3, pp. 411–436, 2014.
[3] G. Costa and R. Ortale, "Overlapping communities meet roles and respective behavioral patterns in networks with node attributes," in *Proc. Int. Conf. Web Inf. Syst. Eng.*, 2017, pp. 215–230.
[4] E. Goffman, *The Presentation of Self in Everyday Life*. Garden City, NY, USA: Doubleday, 1959, 2002.
[5] G. Costa and R. Ortale, "Mining overlapping communities and inner role assignments through Bayesian mixed-membership models of networks with context-dependent interactions," *Trans. Knowl. Discovery Data*, vol. 12, no. 2, 2018, Art. no. 18.
[6] F. Du, Y. Liu, X. Liu, J. Sun, and Y. Jiang, "User role analysis in online social networks based on Dirichlet process mixture models," in *Proc. Int. Conf. Adv. Cloud Big Data*, 2016, pp. 172–177.
[7] R. A. Rossi, B. Gallagher, J. Neville, and K. Henderson, "Modeling dynamic behavior in large evolving graphs," in *Proc. 6th ACM Int. Conf. Web Search Data Mining*, 2013, pp. 667–676.
[8] Y. Ruan and S. Parthasarathy, "Simultaneous detection of communities and roles from large networks," in *Proc. 2nd ACM Conf. Online Soc. Net.*, 2014, pp. 203–214.
[9] W. Fu, L. Song, and E. P. Xing, "Dynamic mixed membership blockmodel for evolving networks," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 329–336.

[10] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed membership stochastic blockmodels," *J. Mach. Learn. Res.*, vol. 9, no. Sep, pp. 1981–2014, 2008.
[11] D. Vega, M. Magnani, D. Montesi, R. Meseguer, and F. Freitag, "A new approach to role and position detection in networks," *Soc. Netw. Anal. Mining*, vol. 6, no. 1, 2016, Art. no. 39.
[12] E. Medina, D. Vega, R. Meseguer, H. Medina, S. F. Ochoa, and M. Magnani, "Using indirect blockmodeling for monitoring students roles in collaborative learning networks," in *Proc. 20th Int. Conf. Comput. Supported Cooperative Work Des.*, 2016, pp. 164–169.
[13] T. Hecking, I.-A. Chounta, and H. U. Hoppe, "Investigating social and semantic user roles in MOOC discussion forums," in *Proc. 6th Int. Conf. Learn. Anal. Knowl.*, 2016, pp. 198–207.
[14] S. Gilpin, T. Eliassi-Rad, and I. Davidson, "Guided learning for role discovery (glrd): framework, algorithms, and applications," *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 113–121.
[15] E. P. Xing, W. Fu, and L. Song, "A state-space mixed membership blockmodel for dynamic network tomography," *Ann. Appl. Statist.*, vol. 4, no. 2, pp. 535–566, 2010.
[16] P. V. Gupte, B. Ravindran, and S. Parthasarathy, "Role discovery in graphs using global features: Algorithms, applications and a novel evaluation strategy," in *Proc. 33rd Int. Conf. Data Eng.*, 2017, pp. 771–782.
[17] R. A. Rossi and N. K. Ahmed, "Role discovery in networks," *IEEE Tran. Knowl. Data Eng.*, vol. 27, no. 4, pp. 1112–1131, 2015.
[18] S. Vijayakumar, A. D'souza, and S. Schaal, "Incremental online learning in high dimensions," *Neural Comput.*, vol. 17, no. 12, pp. 2602–2634, 2005.
[19] R. K. Merton, *Social Theory and Social Structure*. New York, NY, USA: Simon and Schuster, 1968.
[20] S. Harkness, C. P. Edwards, and C. M. Super, "Social roles and moral reasoning: A case study in a rural African community," *Develop. Psychology*, vol. 17, no. 5, 1981, Art. no. 595.
[21] Z. P. Neal, *The Connected City: How Networks are Shaping the Modern Metropolis*. Evanston, IL, USA: Routledge, 2012.
[22] M. E. Newman, "Detecting community structure in networks," *Eur. Phys. J. B*, vol. 38, no. 2, pp. 321–330, 2004.
[23] M. E. Newman, "Detecting community structure in networks," *Eur. Phys. J. B*vol. 38, no. 2, pp. 321–330, 2004.
[24] M. E. Newman, "Spectral methods for community detection and graph partitioning," *Phys. Rev. E*, vol. 88, no. 4, 2013, Art. no. 042822.
[25] C.-T. Li, Y. Lin, and M.-Y. Yeh, "The roles of network communities in social information diffusion," in *Proc. IEEE Int. Conf. Big Data*, 2015, pp. 391–400.
[26] P. Schultz, C. Stava, M. Beck, and R. Vassilopoulou, "Internet message board use by patients with cancer and their families," *Clin. J. Oncology Nursing*, vol. 7, pp. 663–667, 2003.
[27] M. Revelle, C. Domeniconi, and A. Johri, "Persistent roles in online social networks," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2016, pp. 47–62.
[28] S. P. Borgatti and M. G. Everett, "Notions of position in social network analysis," *Soc. Methodology*, vol. 22, pp. 1–35, 1992.
[29] P. V. Gupte and B. Ravindran, "Scalable positional analysis for studying evolution of nodes in networks," *arXiv preprintarXiv:1402.3797*, 2014.
[30] H. T. Welser, E. Gleave, D. Fisher, and M. Smith, "Visualizing the signatures of social roles in online discussion groups," *J. Soc. Struct.*, vol. 8, no. 2, pp. 1–32, 2007.
[31] J. Nielsen, "Participation inequality: Encouraging more users to contribute," *Jakob Nielsen's Alertbox*, vol. 9, 2006, Art. no. 2006.
[32] S. Choobdar, P. Ribeiro, S. Parthasarathy, and F. Silva, "Dynamic inference of social roles in information cascades," *Data Mining Knowl. Discovery*, vol. 29, no. 5, pp. 1152–1177, 2015.
[33] A. Bartal, "Modeling influence on posting engagement: The Gaza great return march analyzed on Twitter," in *Proc. Int. Conf. Adv. Soc. Netw. Anal. Mining*, 2018, pp. 577–578.
[34] A. C. Sonnenbichler, "A community membership life cycle model," *Technical Report*, Karlsruhe Institute of Technology, 2010.
[35] R. L. Stigers, "Online social network behaviors as predictors of personality," M. S. thesis. Chico, CA: California State University, 2011.
[36] K. Barzilai-Nahon, "Gatekeepers, virtual communities and the gated: Multidimensional tensions in cyberspace," *Int. J. Commun. Law Policy*, vol. 11, 2006, Art. no. 9.
[37] B. A. Coles and M. West, "Trolling the trolls: Online forum users constructions of the nature and properties of trolling," *Comput. Hum. Behav.*, vol. 60, pp. 233–244, 2016.
[38] N. Agarwal, H. Liu, L. Tang, and P. S. Yu, "Identifying the influential bloggers in a community," in *Proc. ACM Int. Conf. Web Search Data Mining*, 2008, pp. 207–218.
[39] T. van der Valk, M. M. Chappin, and G. W. Gijsbers, "Evaluating innovation networks in emerging technologies," *Technological Forecasting Soc. Change*, vol. 78, no. 1, pp. 25–39, 2011.
[40] N. Memon, H. L. Larsen, D. L. Hicks, and N. Harkiolakis, "Retracted: Detecting hidden hierarchy in terrorist networks: some case studies," in *Proc. Int. Conf. Intel. Security Inf.*, 2008, pp. 477–489.
[41] S. White and P. Smyth, "Algorithms for estimating relative importance in networks," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2003, pp. 266–275.
[42] S. P. Borgatti, "Identifying sets of key players in a social network," *Comput. Math. Org. Theory*, vol. 12, no. 1, pp. 21–34, 2006.
[43] G. Ravid and S. Rafaeli, "Asynchronous discussion groups as small world and scale free networks," *First Monday*, vol. 9, 2004, Art. no. 9.
[44] K. Nowicki and T. A. B. Snijders, "Estimation and prediction for stochastic blockstructures," *J. Amer. Statist. Assoc.*, vol. 96, no. 455, pp. 1077–1087, 2001.
[45] M. S. Handcock, A. E. Raftery, and J. M. Tantrum, "Model-based clustering for social networks," *J. Royal Statist. Society: Series A*, vol. 170, no. 2, pp. 301–354, 2007.
[46] A. Arockiasamy, A. Gionis, and N. Tatti, "A combinatorial approach to role discovery," in *Proc. 16th Int. Conf. Data Mining*, 2016, pp. 787–792.
[47] L. Li, L. Qian, V. E. Lee, M. Leng, M. Chen, and X. Chen, "Fast and accurate computation of role similarity via vertex centrality," in *Proc. Int. Conf. Web-Age Inf. Manage.*, 2015, pp. 123–134.
[48] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*. Englewood Cliffs, NJ, USA: Prentice hall, 1995.
[49] R. Rossi, B. Gallagher, J. Neville, and K. Henderson, "Modeling temporal behavior in large networks: A dynamic mixed-membership model," Lawrence Livermore Nat. Lab., Livermore, CA, USA, Tech. Rep. LLNL-TR-514271, 2011.
[50] T. E. Will, "Flock Leadership: Understanding and influencing emergent collective behavior," *Leadership Quart.*, vol. 27, pp. 261–279, 2016.
[51] D. Romero, B. Meeder, and J. Kleinberg, "Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter," in *Proc. 20th Int. Conf. WWW*, 2011, pp. 695–704.
[52] S. Aghdam and N. Navimipour, "Opinion leaders selection in the social networks based on trust relationships propagation," *Int. J. Modern Sci.*, vol. 2, no. 2, pp. 88–97, 2016.
[53] J. Cheng, M. Bernstein, C. Danescu-Niculescu-Mizil, and J. Leskovec, "Anyone can become a troll: Causes of trolling behavior in online discussions," *in Proc. ACM conf. comput. supported cooperative work and social computing*, 2017, pp. 1217–1230.
[54] M. Charrad, N. Ghazzali, V. Boiteau, and A. Niknafs, "Package 'NbClust'," *J. Statist. Softw.*, vol. 61, pp. 1–36, 2014.
[55] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts, "Who says what to whom on twitter," in *Proc. 20th Int. Conf. World Wide Web*, 2011, pp. 705–714.
[56] P. Erdos and A. Rényi, "On the evolution of random graphs," *Publication Math. Inst. Hungarian Acad. Sci.*, vol. 5, no. 1, pp. 17–60, 1960.
[57] M. Karsai, K. Kaski, A.-L. Barabási, and J. Kertész, "Universal features of correlated bursty behaviour," *Sci. Rep.*, vol. 2, 2012, Art. no. 397.
[58] H. Wang, C. Wang, C. Zhai, and J. Han, "Learning online discussion structures by conditional random fields," in *Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2011, pp. 435–444.
[59] M. Galster and D. Tofan, "Exploring possibilities to analyse microblogs for dependability information in variability-intensive open source software systems," in *Proc. Int. Symp. Softw. Rel. Eng. Workshops*, 2013, pp. 321–325.
[60] M. De Domenico, A. Lima, P. Mougel, and M. Musolesi, "The anatomy of a scientific rumor," *Scientific reports*, vol. 3, pp. 2980, 2013.
[61] M. E. Newman, "Analysis of weighted networks," *Phys. Rev. E*, vol. 70, no. 5, 2004, Art. no. 056131.
[62] S. Salvador and P. Chan, "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," in *Proc. 16th Int. Conf. Tools Artif. Intel.*, 2004, pp. 576–584.
[63] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: Understanding microblogging usage and communities," in *Proc. 9th WebKDD 1st SNA-KDD Workshop Web Mining Soc. Netw. Anal.*, 2007, pp. 56–65.

[64] S. On-at, A. Quirin, A. Péninou, N. Baptiste-Jessel, M.-F. Canut, and F. Sèdes, "A parametric study to construct time-aware social profiles," in *Proc. Trends Soc. Netw. Anal.*, 2017, pp. 21–50.
[65] D. Krackardt, "QAP partialling as a test of spuriousness," *Soc. Netw.*, vol. 9, no. 2, pp. 171–186, 1987.
[66] P. Leifeld, S. J. Cranmer, and B. A. Desmarais, "Temporal exponential random graph models with btergm: Estimation and bootstrap confidence intervals," *J. Statist. Softw.*, vol. 83, no. 6, pp. 1–36, 2018.
[67] S. Hanneke, W. Fu, and E. P. Xing, "Discrete temporal models of social networks," *Electron. J. Statist.*, vol. 4, pp. 585–605, 2010.
[68] B. A. Desmarais and S. J. Cranmer, "Statistical mechanics of networks: Estimation and uncertainty," *Physica A: Statist. Mech. Appl.*, vol. 391, no. 4, pp. 1865–1876, 2012.

Alon Bartal received the BSc and MSc degrees in information systems (IS) from the IEM Department, Ben-Gurion University (BGU), and the PhD degree in IS from the Department of Industrial Engineering & Management (IEM), BGU. He is a postdoctoral researcher in the Department of Software & IS Engineering. BGU. His research focuses on modeling and mining online social network dynamics using networks analysis, machine learning, and natural language processing.

Gilad Ravid received the BSc degree from the Technion, the MBA degree from Hebrew University, and the PhD degree from Hafia University. He is a senior faculty member in information systems (IS) in the Department of Industrial Engineering & Management, Ben-Gurion University. He was a postdoctoral at the Annenberg Center for Communication, University of Southern California. His main interests are focused on the relationship between social structure and human behavior and computer-mediated communication. His work includes research as social structure in web-based spaces, civil information, and information sharing in groups. He has been published in top journals including *IS Research*, *First Monday*, and the *IS Journal*.