Review

Role-Aware Information Spread in Online Social Networks

Alon Bartal *[ID] and Kathleen M. Jagodnik [ID]

The School of Business Administration, Bar-Ilan University, Ramat Gan 5290002, Israel;
kathleen.jagodnik@biu.ac.il
* Correspondence: alon.bartal@biu.ac.il

Abstract: Understanding the complex process of information spread in online social networks (OSNs) enables the efficient maximization/minimization of the spread of useful/harmful information. Users assume various roles based on their behaviors while engaging with information in these OSNs. Recent reviews on information spread in OSNs have focused on algorithms and challenges for modeling the local node-to-node cascading paths of viral information. However, they neglected to analyze non-viral information with low reach size that can also spread globally beyond OSN edges (links) via non-neighbors through, for example, pushed information via content recommendation algorithms. Previous reviews have also not fully considered user roles in the spread of information. To address these gaps, we: (i) provide a comprehensive survey of the latest studies on role-aware information spread in OSNs, also addressing the different temporal spreading patterns of viral and non-viral information; (ii) survey modeling approaches that consider structural, non-structural, and hybrid features, and provide a taxonomy of these approaches; (iii) review software platforms for the analysis and visualization of role-aware information spread in OSNs; and (iv) describe how information spread models enable useful applications in OSNs such as detecting influential users. We conclude by highlighting future research directions for studying information spread in OSNs, accounting for dynamic user roles.

1. Introduction

Information spread is the process by which a piece of information reaches different individuals [1,2]. The ways in which people consume and share information have dramatically changed due to the rapid development of online social networks (OSNs) that enable their users to interact worldwide by, for example, publishing messages, sharing experiences, and exchanging opinions [3,4]. Some OSNs support two types of networks: (i) a *social network* is typically represented as a graph in which nodes represent users, and edges (links) represent social relationships (e.g., friendships on Facebook); and (ii) an *interaction network* describing how the information spreads, in which nodes represent users in the social network who adopted the information and edges represent interactions among users (e.g., posting a personal message, re-posting information, or liking a post). Laying the interaction network over the social network (Figure 1) creates a multi-layer network with multiple types of edges, defined as a tuple $< V, E, L >$ where: V represents nodes; E is a set of tuples $< v_i, v_j, l >, v_i, v_j \in V, l \in L, v_i \neq v_j$ and for any two tuples $< v_i, v_j, l >, < v_i', v_j', l' > \in E$ if $< v_i, v_j, l > \neq < v_i', v_j', l' >, v_i = v_i'$ and $v_j = v_j'$ then $l \neq l'$; and L is a set of distinct layers (types of relationships).

The spread of information in OSNs mainly relies on interactions among users [5], with local user-to-user diffusion over the social network as a key mechanism found in the spreading of information in networks describing technological innovations [6], product purchasing in viral marketing [5,7–10], and adoption of opinions [1]. Thus, understanding

information spread in OSNs is essential for a variety of applications, including controlling the spread of rumors [11] and monitoring public opinions [12].
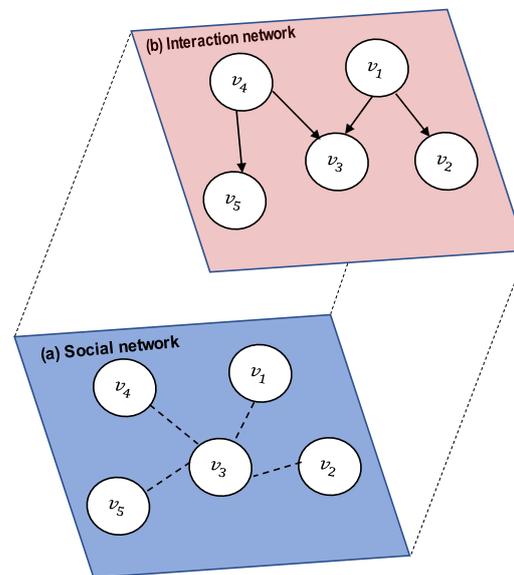

Figure 1. Illustration of two types of networks with the same set of nodes and different edges (links): (**a**) a social network in which dashed edges represent social relationships among users (e.g., Facebook friendships), and (**b**) a directed interaction network laid over the social network, in which solid edges represent user interactions (e.g., user v_4 retweeted a message originated by v_3 and another message originated by v_5).

A piece of information can spread within minutes to millions of people via online social interactions [13]. The spreading speed and reach size of the information in OSNs is affected by factors including information sharing of highly connected users [14] and their motivation in sharing the information based on their interests [15], which implies a variety of user social roles. An OSN user who decides to share a piece of information posted by another user (e.g., a retweet on Twitter) is considered influenced [16–19]. The influence of one user on another is a complex process that depends on many factors, such as user personal interests, reflected by the content of a message [16,20]; and users' attributes, for example, resulting from network position and age [21–27]; these attributes are often referred to as user roles. To better understand information spread processes such as the spread of misinformation [28,29], researchers have tried to model information spreading paths resulting from influence [16,19,30–32].

The online behaviors of users determine their social roles and affect how information spreads in OSNs [33]. Highly active users have a potential to spread information quickly to multiple people [31] and influence them to spread the information, as well. For example, it was found that half of the URLs posted on Twitter were tweeted by a highly active minority of users (fewer than 1%) who act as opinion leaders [34]. Those influence events that result in the spread of information produce dynamic changes in users' online behaviors that are reflected in the structure of the network [31,35,36] and can be modeled by learning the interaction characteristics among OSN users [31]. For example, a User *A* who was exposed to interesting information originated by another User *B*, might decide to follow User *B*, thus forming a new edge in the social network. In addition, User *A* might be influenced to share the information originated by User *B*, thus creating a link in the interaction network. To better explain the observed information spreading patterns in both types of networks, it is essential to identify the categories of social roles of individuals who affect the information spreading process.

Studies on modeling information spread in OSNs can be categorized into three main approaches of (i) structure-based models; (ii) non-structural models; and (iii) hybrid models.

Structure-based models utilize the structure of the social network to infer information spread. These models include, for example, the Linear Threshold Model [37] in which network edges are weighted, and a user can transition between non-active and active roles if the sum of incoming edge weights from active neighbors exceeds a threshold. Using structure-based models, the probability that a user who was exposed to a piece of information will share it depends also on the social roles of the neighbors [33]. In contrast to earlier decades when the mainstream media (e.g., TV and radio stations) were the main sources of information, currently, people use OSNs to consume information [3,4] posted by other users, not necessarily neighbors in the social network, as well as information promoted by content recommendation algorithms and retail publishers [20,27,38]. For example, the Twitter Home timeline exposes a user to content posted by his/her followers, but also to promoted content and personalized recommended content originated by non-neighbors that Twitter thinks will interest the user. This mechanism of sharing global information, not limited to node-to-node spread, is addressed in non-structural based models [39–41] that infer the spread of information by non-structured mechanisms (e.g., homophily [42,43]) other than node-to-node propagation. This means that the role of a user in spreading information via OSNs is not restricted to detecting similar local connections. Thus, users can be affiliated with the same role even if they are remote and unconnected. Extending the idea of detecting users with the same role who do not share structural network similarity (e.g., using homophily), topic-aware information spread models have been utilized as non-structural models to identify users who interact with a similar topic distribution, under the assumption that similar interests imply similar behavioral patterns (i.e., roles) [15]. Hybrid models simultaneously consider structural and non-structural mechanisms to explain information spread. For example, Myers et al. [27] developed a model for explaining how information spreads in OSNs by considering both local node-to-node propagation and external information sources.

A systematic review of the literature about information spread in OSNs reveals the following limitations of existing models that call for future research. First, structure-based models utilize the structure of the network and incorrectly assume that information can diffuse only via the edges of a social network. In addition, these models ignore structural differences between users that imply user roles, neglect personal preferences or interests of users (e.g., Zhang [44]), and omit global information spread mechanisms by non-neighbors (e.g., exposure to information by mainstream media sources). Second, non-structural models consider spreading mechanisms beyond local user-to-user propagation but ignore the topology of the network and only consider the global rate of information spreading patterns [45]. Several studies addressed the challenge of modeling both approaches (structural and non-structural) in a single model by creating hybrid models that combine both local and global features. For example, [46] generates structural features by analyzing the social network or the interaction network, and non-structural features by analyzing topics resulting from user exposure to broadcast news, posts from friends, and users' interests [15]. However (third), hybrid models to date have failed to integrate rich features that affect the complex process of information spread such as local spread mechanisms, global spread mechanisms, the context of a message, social roles reflecting personal user behavior, and personal user interests, leaving much room for improvement. Although several techniques [33,46,47] have been developed for modeling information spread by extracting latent topics of posts and hidden user social roles, accurately characterizing the dynamic role-aware, topic-dependent process of information spread is still not fully understood. Lastly, the fourth limitation of most recent information spread studies in networks (e.g., [22–25,48–52]) is related to focusing on the spread of viral information that leads to influence, but largely ignoring the spread of non-viral information. For example, Nahon et al. [53] studied the viral spread of the "Yes We Can" slogan in the 2008 U.S. presidential election in social and mainstream media first by opinion leaders and then by other users. In addition, there is a lack of consensus among viral information spread studies about the number of influence events that define an information as viral.

The number of users who are influenced by a piece of information follows a long-tailed distribution [22,54]. Thus, most information nuggets are shared only a scant number of times [17,20,22,55], and many information nuggets are not shared at all. This finding implies that influence resulting from viral versus non-viral information spread follows different infection mechanisms [19] and necessitates future research about the spread of non-viral information.

Some of the presented gaps have been addressed in recent review papers on information spread in OSNs [12,32,56,57], covering algorithms for modeling the local node-to-node spread of viral information. However, these reviews present three main gaps. First, previous reviews have not fully considered the roles of users in the spread of information. Second, neglecting to address global information spread beyond OSN edges (links) via non-neighbors through, for example, information pushed via content recommendation algorithms; Third, neglecting to address the spread of non-viral information with low reach size. To address these gaps from previous review papers on information spread in OSNs, this review paper contributes by: (i) presenting a comprehensive survey of the latest studies on role-aware information spread in OSNs, including the different temporal spreading patterns of viral and non-viral information; (ii) surveying modeling approaches that consider structural, non-structural, and hybrid features, and providing a taxonomy of these approaches; (iii) surveying software platforms for the analysis and visualization of role-aware information spread in OSNs; and (iv) describing how information spread models facilitate useful applications in OSNs such as detecting influential users. We conclude by presenting future research directions for studying information spread in OSNs, accounting for dynamic user roles.

Organization. In Section 2, we cover literature about social roles, followed by a survey of computational models aimed at discovering social roles in Section 3. Section 4 addresses the *first* gap in existing review papers of not fully considering user roles in the spread of information, by discussing how user social roles affect information spread. We also address the *second* gap in existing reviews involving their coverage of only node-to-node information spread mechanisms, by reviewing structural, non-structural, and hybrid information spread models in OSNs. More specifically, we provide a brief overview on information spread including its mechanisms, the motivation for modeling information spread in OSNs, and how exposure to information can lead to influence that affects users' behavior and network structure. In Section 5, we address the *third* gap in existing review papers of their covering only viral information spread by describing and comparing viral vs. non-viral information spread models. Section 6 reviews software platforms for exploring user roles in OSN information spread. Finally, we conclude this review and share our perspectives regarding future research in Section 7.

2. Social Roles

Social roles were first identified by sociology research [58] and have been used to categorize the behaviors of individuals. Early definitions of the term *role* include, for example, "a cluster of related and goal-directed behaviors characteristic of a person within a specific situation" [59], and "behavioral expectations of a person" [60]; for example, a person holding the role of a teacher is expected to be an educational figure. Moreover, it is expected that different individuals with the same role behave in the same way. In the physical world, outside OSNs, role definitions sometimes include a formal job description with a list of responsibilities [61]. Biddle [62] defined five major types of social roles, each described by a developed theory: (i) Functional Role Theory—focuses on social roles with an emphasis on the characteristic behaviors of the people who serve in these roles; (ii) Symbolic Interactionist Role Theory—assesses how individuals interpret observed behavior differently and how they affect each other's behavior via interactions; (iii) Structural Role Theory—utilizes mathematical models to examine the influence of the society as a whole on role selection by individuals; (iv) Organizational Role Theory—examines role development

in organizations; and (v) Cognitive Role Theory—analyzes an individual's expectations of a person holding a role compared with the observed behaviors of that person.

In recent years, especially during the restrictive regulations of the COVID-19 pandemic, people have increasingly used OSNs to maintain social relationships and interact globally [63]. Some of these OSNs allow the formation of virtual groups of people (online communities) who interact with each other and share common values, beliefs, behaviors, interests, personal preferences, and other characteristics [64,65]. In most online communities, the majority of users never contribute (90% in [66]), a small group of users account for almost all network activity (1% in [66]), and the rest of the users (9% in [66]) contribute minimally, which implies the existence of user roles. For example, *influentials* are a small minority group of very active users affecting information spread in OSNs [67] who influence multiple users to be active, as well [68].

Studying user roles in OSNs can improve our understanding of how individuals contribute to their community and spread information [59]. The importance of roles is acknowledged in many applications such as predicting new network links [69] and modeling dynamic user behavior in large graphs [70,71]. The growing usage of OSNs establishes the need to better understand user behavior online to recognize emergent behaviors characterized as social roles that define user contributions in generating and spreading information.

Several studies have identified the distinct roles of OSN users via their activity levels in OSNs [31,72,73]. A *visitor* is a user who consumes information, but who does not participate in discussions or interact with other users [31]. A user who participates in discussions is designated as a *novice* [31]. With frequent participation, as well as the activities of viewing and generating content in the community, the novice is upgraded to the *activist* role [31]. An activist earns the designation of a *leader* in a community, also referred to as a *gatekeeper* [73], by engaging in the role of an opinion maker who influences many other community members to engage in similar activities. In the opposite direction, a user becomes *passive* if his/her activity is limited to maintaining interest in discussions and in other users, only consuming content rather than generating it. Finally, a *troll* causes conflicts with other users by posting offensive content, typically engaging in such antisocial behavior over a relatively short period, prior to disappearing from the community [31,72].

While OSN users can be assigned formal roles, such as administrators [74], more often, roles in OSNs are not formally defined, thus challenging other OSN users to identify and understand the roles and behaviors of their community members. Many studies [68,70,75–77] have reported roles that emerge from the structure of the network and that can be identified via network centrality metrics (e.g., Indegree, Outdegree, Closeness, Betweenness, Eigenvector [75]) or ranks (e.g., PageRank, HITS), which reveal structural information about user relationships within the community [78]. For example, when analyzing the 9/11 terrorist attacks [79], influencers who served in a variety of roles (e.g., "gatekeepers" and "brokers") were identified using centrality measures. Additionally, centrality measures have been employed to assess the importance of users in networks [80] (e.g., via articulation nodes and selection of the top-k nodes [81]). In contrast, other studies [82] have identified no correlations between user roles and their structural positions. Such findings might result from the analysis of dynamic user behavior online as in [31], which identifies more nuanced roles than simple analyses of a single, static snapshot of the network, when considering whether a user is highly influential or not. Additionally, since OSN users often change roles [31,83], it is important to use the dynamic structural relations of users in order to identify diverse, shifting roles, as reviewed in Section 3.

Each role in an OSN has a unique signature that can be defined by observing the patterns of user participation [84]. These patterns typically include structural network centralities such as Closeness [85], Degree [86], and Betweenness [87], or changes in node embeddings (a method of representing nodes as vectors) [88] over time, but can also include temporal patterns such as the time interval between posts [20]. Users' interactions change and evolve over time [31,89], and can be characterized as roles that describe

user behavior [31,90]. For example, a user who interacts with users who affiliate with two isolated communities is considered a mediator. The underlying assumption of role detection models is that users who affiliate with the same role have similar network structural patterns [70]. This assumption does not require that users with the same role will be neighbors or close in the network [69,77]. Moreover, a user can play different roles and affiliate with different communities [69]. Users who affiliate with the same community must be close in the network, whereas users who affiliate with the same role can be far from each other. Recent studies utilize the structure of a network to define a community as a group of users who are highly connected among themselves and have fewer connections outside the group [58]. By analyzing the structure of the network, we can identify different communities and their affiliated members [69,91] and roles [58,92].

The association of a single affiliation of a member to a community for a specific snapshot of the network was expanded to allow community detection algorithms to simultaneously assign individual users to several communities [93]. For example, the Bayesian Hierarchical Latent-Factor model (BHLFM) was designed to uncover hidden affiliations of members to multiple communities and roles [69]. Users who simultaneously affiliate with several communities in the analyzed network are considered to have an important role due to their structural location in the network that enables them to potentially control the spread of information among those communities [94]. These users who affiliate with several communities have served central roles in increasing the spread of information reach about presidential campaigns as well as propaganda [95].

Identifying important users in networks is essential for understanding and managing the flow of information in OSNs. This strategy has been employed to contain or extinguish outbreaks of epidemic infections [96]. For example, during the COVID-19 crisis, governments have needed to identify superspreaders of the SARS-CoV-2 virus and isolate them to contain the disease spread [96]. Similarly, identifying important users in OSNs is necessary for delaying or preventing the spread of misinformation in OSNs [97,98]. For example, a small minority of 0.1% of Twitter users (superspreaders) were found to be responsible for spreading 80% of fake news [99]. Aiming to locate users with a significant potential to control the spread of information, Chang et al. [100] developed a model to identify a small minority of nodes describing unique physical locations that act as superspreaders, accounting for the majority of infections. Since these superspreader and influence nodes have a massive influence on a large population [31], numerous studies have tried to identify them. Classic examples include detecting influential researchers in a citation network [101]; understanding user communication patterns in email communication networks [102]; identifying important junctions in road networks that might create bottlenecks [103]; and analyzing critical stop locations of flights between airports [104]. These studies analyzed nodes' behavior to infer their informal (latent) social role, defined by the set of activities that users perform [105,106].

Sociological studies reviewed in the current section aim to describe and define social roles. As modern life becomes increasingly integrated with online systems, the concept of social roles becomes valuable as a tool for categorizing patterns of action, recognizing distinct user types, and cultivating and managing online communities. Utilizing sociological definitions of roles, computational models for role discovery in OSNs combine psychological, behavioral, and structural attributes for modeling user behavior online. Moreover, computational role discovery models often focus on the detection of latent roles using unsupervised methods, as elaborated next.

3. Role Discovery Models

Networks have been widely studied to learn complex interactions among users and discover their roles by analyzing the network structures [107]. Early network-based methods for role discovery used network centrality measures to identify user importance by, for example, selecting the top-k nodes with the highest measure [108]. For example, [109] proposed a weighted Degree decomposition method to rank and identify cheaters who can

lead to the collapse of cooperation process in a dynamic evolving cooperation network. To improve node ranking, [110,111] computed the entropy of nodes using node Degree and the total Degree of a network. However, Degree centrality is limited to 1-hop neighbors only. Other network centralities such as Betweenness and Closeness centralities, when used for role discovery in networks, can become computationally extensive for medium and large-scale networks. In contrast, the k-core centrality [112,113] is much more efficient, in terms of consuming computational power, for role discovery in networks. k-core is an iterative graph pruning algorithm that groups nodes into layers, where the innermost core nodes have the highest k-core value and the outer nodes have the smallest k-core value. Hybrid k-core methods for node ranking for role detection use the k-core algorithm with a combination of other network centrality metrics such as Degree, Closeness, and Shortest Distance [85–87].

Utilizing network structure, recent studies use clustering [114] for detecting user roles in networks by affiliating users (nodes) to distinct or overlapping groups. The task of roles detection is complementary to the notion of community detection, which involves partitioning the network into cohesive subgroups. Network clustering has been extensively used for community detection [64,115–117] and role discovery [2,31,77] tasks. Community detection algorithms aim to find groups of users (nodes) that frequently interact with each other and have sparse interactions with other users not affiliated with the group [118]. Structural-based role discovery algorithms, on the other hand, group nodes based on their structural similarity [119]. For example, bridge and hub nodes [120] share structural similarity by connecting previously unconnected groups or users. The difference between the tasks of community and role detection is that users who affiliate with the same community are likely to be connected to each other, whereas users who affiliate with the same role can be distant in a network, and unconnected. Recent methods aimed to detect user roles in OSNs typically use unsupervised techniques to cluster structurally similar users together. The latent position cluster model [121] estimates unknown role affiliations by a set of probabilities of edge formation between users, where each user is associated with exactly one role.

Many role detection methods assume that user roles can be inferred by analyzing social interactions, and that users with similar network structural patterns affiliate with the same role [64,68,77]. Examples of models that utilize the structure of the graph to discover user roles include semi-supervised role inference models [122] and Blockmodels [123]. In another study [67], users' social roles are identified by using the structure of the social network; it analyzed how the interactions between users with different roles impact information diffusion paths. The authors found five user roles in two categories of users' ability to influence others in spreading the information, and in blocking (delaying) information spread. The single role affiliation assumption of the latter models was extended in the Mixed-Membership Stochastic Blockmodels (MMSB) [93] by allowing a user to affiliate with several roles. MMSB assumes a fixed network structure; however, OSNs are dynamic. Thus, dynamic MMSB (dMMSB) [124] and a few other models [31,69,91] consider network evolution, where users serve in various roles that evolve over time. The definition of roles in dMMSB implies that users with similar roles share common features and relation patterns, even without direct relationships [92]. More recently, network embedding methods [29,121,125–127] have gained popularity in studying graph structure, achieving state-of-the-art performance in downstream tasks such as node classification. These methods represent network data as vectors in a latent space to preserve the topological structure and properties of the original network. The most widely used network embedding methods can be categorized into (i) matrix factorization; (ii) random walk based methods; and (iii) deep learning models.

Matrix factorization methods [128], such as singular value decomposition (SVD) and non-negative matrix factorization (NMF), aim to learn node embedding via low-rank approximation and analysis of the adjacency matrix (denoted as A): a binary matrix

describing a network, the elements of which a_{ij} are 1 if there is an edge between these nodes, else 0, as defined in Equation (1).

$$a_{ij} = \begin{cases} 1, & \text{if there is an edge connecting nodes } i \text{ and } j \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The REACT [129] algorithm generates embeddings via matrix factorization to identify both communities and roles. RolX [130] also utilizes matrix factorization. First, it computes primary node features (Degree, weighted Degree, and average Clustering Coefficients) using ReFeX [131] and recursively aggregates the features of a node's neighbors. Second, NMF is used to generate node embeddings. GLRD [120] extends RolX by modeling role detection as a constrained NMF problem, where the guidance is provided as convex constraints and specified per role. RIDϵRs [92] first partitions the graph using the connectivity patterns among roles by applying the concept of ϵ-Equitable partitions, and then performs role discovery using NMF. GraphWave [132] is an unsupervised method for learning node embeddings based on structural similarity in networks. It treats graph diffusion kernels as probability distributions over networks and calculates embeddings by using characteristic functions of the distributions. HONE [133] analyzes motifs of a weighted graph where an edge's weight is the count of the co-occurrences of the two endpoints in a specific motif. The main limitation of these matrix factorization methods is low computational efficiency resulting from calculating pair-wise node similarity.

Methods based on random walks capture node proximity to generate embeddings for capturing structural network similarity of nodes that are more likely to appear in the same sequence, and these methods map nodes with similar structural features to the same role. Regarding role detection, network nodes that affiliate with a role have the same embeddings. For example, struc2vec generates node features by performing random walks on a graph in which edges are weighted based on structural distances to capture structural information, assuming that nodes with similar network structure have similar roles [134]. Role2Vec [76] maps nodes into several disjoint roles using clustering methods that analyze node features and attributes such as K-means. Then, random walks are performed, and the generated sequence IDs are used as role indicators. struc2gauss [135] models both structural similarity by using RoleSim [136], and uncertainty using a Gaussian distribution. Each node is mapped to a Gaussian distribution, and the variance captures the uncertainties. SPINE [137] incorporates structural features of both local and global node proximity to learn embeddings. RiWalk [127] assumes that nodes with different functionalities have different roles in a network, and the structure of the network can be learned using random-walk-based node embedding. These methods reconstruct the edges between nodes based on the structural similarities so that the context nodes obtained by random walks are structurally similar to the central nodes. However, they do not jointly consider structure, content, and label information. struc2vec better represents role information than SPINE [138], which leads to better embeddings but results in higher complexities of computational time and space.

Deep learning methods for graphs were developed in recent years in an effort to optimize node embeddings representations. Therefore, to date, only a few studies [125,139,140] have leveraged deep learning methods for role-oriented network representation learning. For example, DRNE [125] develops a deep learning method with a normalized long short-term memory (LSTM) layer to learn regular equivalence by recursively aggregating neighbors' representations for each node. GAS [141] utilizes graph neural networks (GNN) to capture node structure by applying sum-pooling propagation instead of the graph convolutional networks (GCN) [142] to better capture local node structures. For training, GAS extracts features similarly to ReFeX and aggregates them only once. Similarly, RESD [140] also relies on ReFeX [131] for generating features that are input to an autoencoder to learn node embeddings while reducing data noise during the learning stage. GraLSP [139] uses a GNN to learn node low-ranked representations by considering the node's local structural patterns as well as the its neighboring nodes via random walks.

Deep learning methods have achieved state-of-the-art results in several graph-based downstream tasks such as node classification [143–145] and link prediction [145,146] that were not used for user role identification. Most deep learning methods mainly use autoencoders and GCNs that generally consist of an encoder, a similarity function, and a decoder. An autoencoder is a neural network with an encoder and a decoder architecture. The encoder takes a data point as input and converts it to a lower-dimensional representation. Next, the decoder takes this lower-dimensional representation and aims to reconstruct the original input as accurately as possible. A GCN is a convolutional network operating directly on graphs. For example, NetSMF [126] uses an autoencoder and graph convolutional networks to learn node representation vectors.

Among the three network embedding methods of matrix factorization, random walks, and deep learning, random walks perform best for small networks, whereas matrix factorization methods perform best for large networks with up to tens of thousands of nodes [147]. These methods analyze feature matrices that are characterized by having lower dimension than the original graph matrices analyzed in the random walk-based methods. Deep learning methods scale well for huge networks having millions of nodes, outperforming other methods for role discovery [138]. Huge networks contain multiple examples that permit sufficient training of deep learning models, thus allowing them to achieve excellent performance. In addition, neural network graph-based models work well when nodes are labeled by their structural roles (structure aware). However, they fail when nodes are labeled by their position in the graph (e.g., community affiliation) since they have the same embeddings.

A few studies have identified roles in large dynamic networks. For example, the Role Affiliation Frequency Model (RAFM) [31] considers dynamic networks in which each user is affiliated with a role at each time step. Thus, a user can play multiple roles. RAFM assumes that users' latent behavior has a different blend of social roles (e.g., a user can be a novice but also have the qualities of a leader) that can change over time. Gupte and Ravindran [71] used MapReduce to identify roles in large sparse graphs with only three time steps. Rossi et al. [148] generated a role transition model per user where role meaning is defined by network centralities. Other studies for role detection in OSNs analyzed text inserts of millions of users, where user role is determined using different word selections, as discussed next.

Textual role discovery models based on Natural Language Processing (NLP) can be used to automatically identify social roles in OSNs [149] via analysis of the latest user textual interactions that capture the most up-to-date behavioral cues in users' language [150]. For example, [151] learned roles in movies and novels from the language of plot summaries and dialogue. Welser et al. [152] identified four roles in Wikipedia users: substantive experts, technical editors, vandalism fighters, and social networkers, by analyzing users' textual inserts. Fazeen et al. [153] classified Twitter users into laders, lurkers, associates, and spammers by analyzing user tweets. Yang et al. [106] analyzed the editing behaviors of Wikipedia users and applied a Latent Dirichlet Allocation (LDA)-based model to detect editor roles. Ferschke et al. [154] found four roles: workers, critiquers, encouragers, and managers, by applying a similar approach to [106]. Maki et al. [155] proposed a supervised graphical model with an outcome measure to define editor roles based on interactions of users who contributed content on Wikipedia. Other role studies that analyze texts identified experts of a topic [156,157], opinion leaders [158], and influential bloggers who are not necessarily the most active users, by measuring the lengths of posts [75]. Recently, analysis of health discussions online revealed that the presence of a moderator user led to increased user engagement [159]. Users who continuously contribute rich content can strongly affect information spread in OSNs and influence other users to contribute.

To summarize, most role detection studies focus on (i) estimating users' importance by ranking them based on their capability to spread information; and (ii) finding a minimal set of users to maximize information spread in terms of reach size, often referred to as the Influence Maximization problem [160–163]. The spread of information in OSNs that leads

to influence is discussed in detail in the following section with an emphasis on user roles, where we review approaches and models of information spread in OSNs, discuss their shortcomings, indicate open questions, and summarize them in a taxonomy (Figure 2).
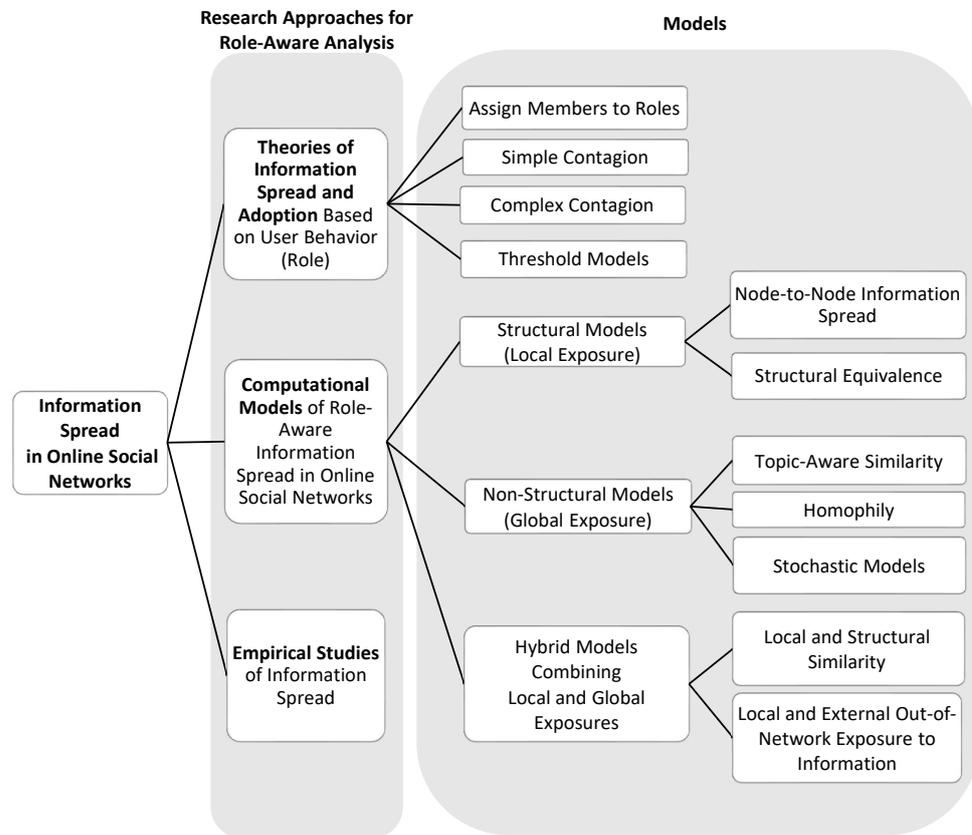

Figure 2. High-level taxonomy of analysis strategies for information spread in online social networks. Grey panels group research approaches for role-aware analysis (left panel), and the associated models (right panel). Refer to Section 4 for a list of studies exemplifying structural models, non-structural models, hybrid models, and models employing external information.

4. Role-Aware Information Spread in Online Social Networks (OSNs)

The widespread usage of OSN platforms such as Facebook, LinkedIn, Sina Weibo, Tumblr, and Twitter enables information to spread quickly. Exposure to information might influence the exposed user to engage by, for example, re-sharing the information [16,19,20]. Relying on the structure of the network, some studies on the spread of information assume that ideas and behaviors have similar spreading mechanisms as diseases [35]. However, these models cannot fully explain the observed spreading patterns of information in OSNs that leads to influence (also referred to as contagion, or adoption) [27,164,165], such as complex contagion. Infectious diseases typically follow *simple contagion*—a contagion probability that is independent of the number of exposures between a source who was infected and a non-infected entity [166]. However, in OSNs *complex contagion* is also at play—a contagion probability that requires multiple exposures to one or more infected entities [166]. Complex contagion is more successful at describing the spread of ideas, technologies, or behaviors than simple contagion [166,167]. Figure 3 illustrates simple and complex contagion.
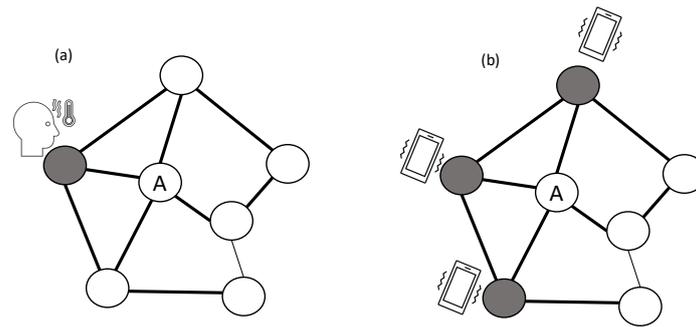
Figure 3. An illustration of simple and complex contagion in a social network. (**a**) Simple contagion: node *A* was infected by a disease after exposure to a single infected node (colored in gray). (**b**) Complex contagion: node *A* adopted a product (a smartphone) after being exposed by three nodes (colored in gray) who adopted the product.

OSN users who have a strong influence on others typically hold the role of opinion leaders who can quickly and efficiently spread information. Many researchers have pursued the identification of opinion leaders in social networks [68,70,75–77,95,168,169] by calculating centrality measures (Table 1). Identifying opinion leaders is important for understanding how information spreads in events including political campaigns [170,171], the adoption of new technologies [172], and planning online marketing and advertising [173]. A research area related to the detection of opinion leaders is the Influence Maximization problem [160–163], introduced in Section 3, that involves selecting a minimal set of unique nodes that will maximize the spread of information and has an NP-complete complexity [163].

Table 1. Centrality Measure Algorithms for Opinion Leader Identification.

Centrality Metric	Algorithm Description	Citation
Betweenness$(v_i) = \sum_{v_k \neq v_j \neq v_i} \frac{\sigma_{v_k,v_j}(v_i)}{\sigma_{v_k,v_j}}$	Given a graph $G = (V, E)$, σ_{v_k,v_j} is the number of shortest paths from $v_k \in V$ to $v_j \in V$, and $\sigma_{v_k,v_j}(v)$ is the number of shortest paths from v_k to v_j that pass through a node $v_i \in V$.	[174]
Closeness$(v_i) = \sum_{v_i \neq v_j; v_i, v_j \in V} \frac{1}{\sigma_{v_i,v_j}}$	Measure how close a node $v_i \in V$ is to all other nodes. Ranges from 0 (far) to $1/(|V|-1)$ (very central). σ is defined as in Betweenness.	[85]
Degree$(v_i) = \sum_j a_{ij}$, $a_{ij} \in A$ defined in Equation (1)	Number of edges connected to a node $v \in V$, or the number of neighbors of a node. Indegree - number of edges connecting into a node in a directed graph. Outdegree - number of edges going out of a node in a directed graph.	[175]
HITS $h(v_i) = \sum_{v_j \in V_{from}} a(v_j)$ $a(v_i) = \sum_{v_j \in V_{to}} h(v_j)$	In the HITS algorithm (hubs and authorities), each node v_i has both a hub score $h(v_i)$ and an authority score $a(v_i)$. We initialize $a(v_i) = h(v_i) = 1$. V_{from} - all nodes that v_i links to. V_{to} - all nodes linking to v_i.	[176]
H-index$(v_i) = H(k_j), j \in \Gamma_i$	Computes the interrelationships between publication quantities and numbers of citations, and defines a researcher's academic influence in a particular domain. H is an operator on a set of real-valued variables $[y_1, \ldots, y_n]$ and returns the maximum integer h such that there are at least h members with a value (degree k) no less than h where Γ_i is the set of neighbors of node v_i.	[177]
Extended H-index$(v_i) = \sum_{v_j \in \Gamma_i} CMC(v_i)$	Utilize structural information from a node's neighbors to compensate for the H-index algorithm that ignores network structure. The extended H-index considers the degrees of the neighbors of nodes (Γ_i) using: (i) $S(v_i)$ is the cumulative degree of the neighbors of node v_i; (ii) $S_k(v_i)$ is the value of the kth index of vector $S(v_i)$; and (iii) $CMC(v_i) = \sum_{k=1}^{h} p^{1+k*\frac{p}{r}} \times s_k(v_i)$, where h is defined in H-index; and $p, r \in [0,1]$ r.	[178]

Table 1. *Cont.*

Centrality Metric	Algorithm Description	Citation
Weighted H-index	A weighted H-index is calculated by constructing an operator H on weighted edges. The accumulation of weighted H-index in the node's neighborhood defines the spreading, then utilizes the SIR model to investigate a spreading process, and define the most influential spreaders.	[179]
$\text{HybridRank}(v_i) = \sum_{v_j \in \Gamma_i} C(v_j) \times EC(v_i)$ were Γ_i is the set of v_i's neighbors	Computes node's importance using two centralities: (i) Eigenvector centrality (*EC*); and (ii) the Coreness (*C*) (see k-core in Table 1) sum of its neighbors. *EC* is a proxy of user influence in terms of connections with high-scored (central) nodes.	[180]
Improved $\text{HybridRank}(v_i) = \sum_{v_j \in \Gamma_i} C(v_j) \times \text{H-index}(v_i)$	Combines two centralities: the Extended Neighborhood Coreness centrality and the H-index centrality. Then, it uses SIR.	[181]
k-core $H \subseteq G, \delta(G) \leq k$	The k-core of graph (*G*) is a maximal subgraph *H* in which each node has at least degree k (In or Out degree). The coreness of a node is k if it belongs to the k-core but not to the $(k+1)$-core.	[112]
Weighted k-core	Applies the same pruning routine as k-core, but measures both the degree of a node and the weights of its links.	[182]
LeaderRank $s_i(t+1) = \sum_{j=1}^{N+1} \frac{a_{ij}}{k_j^{out}} s_j(t)$	The ranking process assigns 1-unit prestige to all (*N*) nodes in a directed network except the ground node (a node connected with every node). The unit prestige of the nodes is evenly distributed to neighboring nodes via links until a steady state is reached. Using random walks, the score of node i at time step t is $s_i(t)$; a_{ij} is an element of an adjacency matrix.	[183]
Weighted LeaderRank $s_i(t+1) = \sum_{j=1}^{N+1} \frac{w_{ij}}{\sum_{k=1}^{N+1} w_{jk}} s_j(t)$	Node ranking is calculated using two models: The first model measures the users' relative influence based on quality of tweet, ratio of retweets, and topic similarity among users; and the second model calculates the user network global influence. This is an expansion of LeaderRank where the score from node i to node j is proportional to the weight w_{ij} as defined in [184].	[184]
$\text{PageRank}(v_i) = \sum_{i \to j} \beta \frac{v_i}{d_i} + (1-\beta) \frac{1}{|V|}$	Measures the importance of a node in a graph $G(V,E)$ with $|V|$ nodes, by counting the number of edges to a node to determine its importance. Important nodes are likely to receive more links from others. β: dumping factor; d_i: out degree of v_i.	[185]
VoteRank $S_i = \sum_{j \in \Gamma_i} V_i$ were Γ_i is the set of v_i's neighbors	Each node v_i is represented by (S_i, V_i) where V_i is v_i's voting ability and S_i is the score of v_i, i.e., the sum S_j where v_j is v_i's neighbor. Initially, V_i is set to 1. At each time step, v_i with the largest score is selected into the target set, and then (i) the voting ability V_i is set to zero; (ii) for each of v_i's neighbors, its voting ability decreases by a factor $\frac{1}{k}$, where k is the average degree, and if $V_j < 0$, we reset it as $V_j = 0$.	[186]

Numerous studies [68,70,75–77] have aimed to identify central users who can best spread information quickly and efficiently. Other studies have focused on modeling the spreading process of information instead of finding top spreaders. These models typically aim to predict the next node in a network that will share a piece of information (i.e., become infected) such as the Susceptible-Infected (SI) [187], Susceptible-Infected-Susceptible (SIS) [187], Susceptible–Infected–Recovered (SIR) [188], Susceptible–Infected–Recovered–Susceptible (SIRS) [187], Susceptible–Exposed–Infected–Recovered (SEIR) [189], the Independent Cascade model (IC) [190], and the Linear Threshold model (LTM) for influence maximization [37] that are summarized in Figure 4. The mechanisms behind those models will be covered in the following subsections.
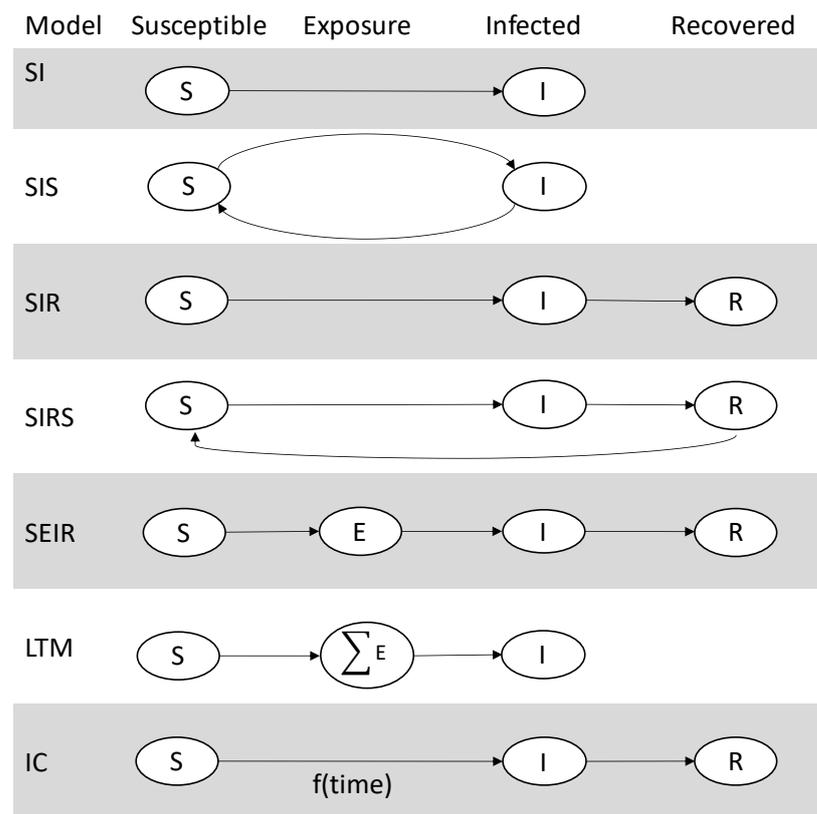
Figure 4. An illustration of the models Susceptible-Infected (SI) [187], Susceptible-Infected-Susceptible (SIS) [187], Susceptible-Infected-Recovered (SIR) [188], Susceptible-Infected-Recovered-for-Susceptible (SIRS) where immunity lasts for only a short period of time [187], Susceptible-Exposed-Infected-Recovered (SEIR) [189], the Linear Threshold model (LTM) for influence maximization [37], and the Independent Cascade model (IC) [190]. In the LTM, a node is exposed to its neighbors, and if the number of infected neighbors exceeds a threshold, the exposed node is infected. In the IC model, each infected node stays active during one time step only and tries to infect its susceptible neighbors with a certain probability. The attempts are independent random events. A susceptible node that was infected will attempt to infect its neighbors at the next time step.

Users' tendency to stay active by creating and sharing information (e.g., a retweet on Twitter or a share on Facebook) can be employed to predict user roles [31] that affect the spread of information [68]. OSN users' motivation to act is affected by, for example, their will to interact with other users or to express interest in the discussed content [191]. Online interactions affect the reach size (often termed *cascade*) and the speed by which information spreads [192]. For example, online interactions by followers of celebrities on Twitter led to multiple retweets of content published by those celebrities [193]. As illustrated in the taxonomy provided in Figure 2, information spread models can be classified into three main groups of (i) Structural, (ii) Non-structural, and (iii) Hybrid information spread models that combine characteristics of both of those model classes, as elaborated next.

4.1. Structural Information Spread Models

Users have different motivations for spreading information [27,194,195], and these motivations inform the study of social behavior as expressed as OSN user roles. The study of information spread in OSNs typically assumes that a User *A* who follows another User *B* is more likely to spread information created or shared by User *B* than users who do not follow User *B*. In other words, structural-based models try to model the spread of information by assuming the existence of underlying node-to-node spreading mechanisms [19,20]. Based on the assumption that information spreads from node to node over network edges, recent studies using structure-based models [19,196] analyzed the social network of a user to infer

the spread of information. Several models have been proposed and studied, demonstrating that the structure of the network affects user behavior and activities [197–200]. For example, the cohesion model [199] is an empirical model measuring the time of adoption to explain the phenomenon of physicians innovating to prescribe the drug tetracycline to patients after directly communicating (cohesion) with other physicians. The authors of [199] detected the time the information was shared by considering the time a prescription was written and found that highly connected physicians in the social network wrote a prescription more rapidly than those who have sparse connections with other physicians, implying different user roles. Valente [197] also presented an empirical model with which he analyzed the time of adopting an innovation.

Similarly to real-world social networks, in OSNs, network structure can be utilized to detect user-to-user exposure to information that can lead to the spread of this information. For example, Kramer et al. [201] found that the emotions of Facebook users that were seen by their friends via the Facebook Wall feature led the exposed friends to express similar emotions, thus influencing those users to spread emotional information online. Bakshy et al. [21] found that exposure of a user to information posted in a URL on the Facebook feed (that presents his/her Facebook friends' activities) increases the user's likelihood to spread that information by sharing the URL address. Sun et al. [202] studied how exposure to information about online activities of a neighbor can influence the online participation of users in fan pages.

The spreading patterns of information via a user-to-user mechanism were found to generate a tree-like structure [22,164]. For example, Kleinberg [190] suggested a structural transition model with user-to-user information spreading mechanisms that creates a tree-like structure, and analyzed the local spread of information by allowing a user to activate inactive nodes at a distance of 1-hop (neighbors) in the network. Two-step diffusion models [164,203] use a slightly different modeling approach in which information first spreads from the mainstream media to opinion leaders. Then, it spreads using a local user-to-user mechanism from opinion leaders to a broader population and creates a tree-like structure.

In retail marketing, exposing people to information via traditional advertising, such as radio, TV, and newspaper commercials, is becoming less effective in convincing people to buy a product [4]. Thus, retailers have been engaging in numerous efforts to convince customers who bought a product to expose their neighbors in OSNs to their experience by recommending other users to buy that product [8]. For example, Leskovec et al. [204] studied a product recommendation network, measuring the extent to which exposure to information about a user's activity in recommending a product can influence their network neighbors to purchase that product. Many other examples of OSN users who exposed their neighbors to information about product usage exist, such as the Old Spice "The Man Your Man Could Smell Like" campaign [9], the ALS Ice Bucket Challenge [26], and Coca-Cola's "Friendly Twist" campaign [10]. According to Rogers' Diffusion of Innovations theory [205], people can be assigned to roles (see taxonomy in Figure 2) of Innovators (2.5%), Early Adopters (12.5%), Early Majority (34%), Late Majority (34%), and Laggards (17%) based on how quickly they adopt an innovation.

Following the notion of categorizing users into roles, threshold diffusion models in OSNs define groups (roles) of users so that users who are affiliated with different groups have unique behavior with varying thresholds of adoption [206]. In other words, users are influenced to adopt a behavior or an innovation depending on the number of their OSN neighbors who were influenced. For example, Centola [207] analyzed the number of exposures to information about health behaviors of neighbors in an OSN that are required for a user to adopt a health behavior. The LTM [208] also utilizes user-to-user information spread with a threshold criterion by focusing on a user's tendency to be influenced resulting from exposure to information by neighbors. The LTM was expanded [209] to consider the temporal order of influence events by considering the time window in which a user can

be influenced, revealing that user bursty activities are able to explain information spread in OSNs.

The structural information spread models reviewed so far in the current section are limited by their inability to consider the spread of information beyond 1-hop neighbors such as user exposure to promoted content. A few studies have addressed the limitation of local user-to-user information spread modeling. For example, the structural-equivalence model [210] considers exposure to information by analyzing the extent to which two users are connected to the same other users in a network. The structural-equivalence model argues that the larger the overlap between user *A*'s set of neighbors and user *B*'s set of neighbors, the more likely user *A* and user *B* will be exposed to similar information [210]. Following the structural-equivalence model, network users should be able to discern the overlap of their neighbor sets and also be able to access each other's activities. However, it is unlikely that users who are not neighbors in a network can observe the overlap of their neighbor sets based on network structure alone. Addressing this problem, Leenders [211] states that the overlap between the neighbors of two network users can be assessed at a sociometric distance of maximum 3-hops.

Structural models account for information by user-to-user exposure mechanisms, which consider only local information spread sources at a distance of 1-hop (e.g., the cohesion model [199]), and the structural-equivalence model considers a wider exposure to information sources at a distance of at most 3-hops. Taken together, those structural information spread models cannot infer the spread of information from users who are located at more remote distances. While in the past, human interactions were limited by physical proximity, currently, with the wide use of the Internet and the increased usage of OSNs, the lack of physical proximity is no longer a constraint for interacting [206].

Recently, numerous studies have modeled information spread by analyzing the diffusion paths of information and using the structure of a social network to predict the next node that will spread the information [212]. Many of these analyses have used neural networks [88,213,214]. For example, [215] applied a recurrent neural network (RNN) to jointly predict the next infected user and estimate the reach size of infected users.

Models that aim to explain information spreading paths that account for user roles by learning user representation typically analyze information spreading patterns as sequences of data. For example, the TopoLSTM model [88] built dynamic directed acyclic graphs (DAGs) consisting of information spread paths and extended Long Short Term Memory (LSTM) network to learn the DAG structure in order to automatically learn user representation. CYAN-RNN [214] applied an attention mechanism to the spreading patterns of information represented using a tree-like structure to capture cross-dependencies of information spreading sequences. Similarly, the DeepDiffuse model [216] used an attention mechanism together with embeddings to predict the time of the next infection and the user who will be infected next.

Most of these models are limited by ignoring the impact of the changing preferences of users over time. To address this limitation, Cao et al. [217] developed a dynamic structural-temporal graph neural network (DySTGNN) that considers both the structure of the social network and temporal features in the information spread graph. Applying a graph attention network (GAT) to the information spread graph, the authors learn user embeddings representing short-term preferences of users. Then, they employ a graph convolutional network (GCN) to the social network and learn structural embedding for each user.

With the phenomenon of global exposure beyond local exposure to information, the mechanisms for online information spread have substantially changed during the last two decades and, thus, information spread has been considered beyond the structure of a network. The local user-to-user exposure to information that is the basis of structural information spread models misses the global mechanisms by which OSN users are exposed to varied content by non-neighbors, beyond network structure, that can lead to global information spread [27,68]. On Twitter, for example, users can be exposed to hashtags

that are posted by non-neighbors [218] as well as to promoted information on a user's Home timeline [219] that contains tweets of followees as well as information tweeted by non-neighbors either by purchased advertisements or tweets ranked as having a large engagement potential [220]. Reddit and Facebook are other examples of network services that allow global exposure via trending topics that appear on a user's front page [221]. Therefore, beyond the local exposure to information by network neighbors, it is important to consider global exposure to information by non-neighbors when modeling information spread in OSNs, as discussed in the next subsection.

4.2. Non-Structural Information Spread Models

In contrast to structural information spread models, non-structural models largely ignore the structure of the network to infer the spread of information that can lead to influence, adoption, or contagion [41,222,223]. Two commonly studied contagion spread models in epidemiology are named SIR and SIS [224] (Figure 4), where nodes with the *S* role are Susceptible, the *I* role are Infectious, and the *R* role are Recovered. In the SIS model, individuals can transition from the role of *S* to *I* and back, and in the SIR model, individuals can transition across the roles *S* to *I* to *R*. SIR and SIS assume that every user has the same random probability to become infected, i.e., users have the same contact rate that is indicated by an edge formation in a network.

Although the SIR and SIS models fail to accurately explain biological contagion processes, particularly at large scales [225], those paradigms have been used to model the adoption of ideas [226], diffusion of innovations [227], and the spread of rumors [11]. Extending the SIS model, Leskovec et al. [204] propose an SIS model in which all users share the same probability β to adopt a piece of information. Users who adopted the information hold the susceptible role at the following time step. Since the assumption that influence is evenly distributed among OSN users is not valid [32], more complex modeling strategies such as exposure rates [16,27] have been developed. For example, Myers et al. [27] measured the level of exposure to information using hazard functions that facilitate exposure curves, such that each new exposure increases the probability of influence.

Other OSN studies that used non-structural models to explain information spread that leads to influence include, for example, the Linear Influence Model (LIM) [39], in which the influence functions of users are affected by the overall rate of influenced users in the network. LIM assumes a static network structure, and integrates exposure effects from a single source at each time step. The Heat Diffusion model [228] assumes a similar logic between heat diffusion in physical systems and information propagation in a network. In this model, the source of the information is analogous to the role of a heat source, and information flows from a node with higher temperature (who was influenced) to a node with lower temperature (non-influenced). The authors [228] developed three diffusion models for selecting the best marketing candidates (users) who will maximize information spread in a network, leading to influence. The three diffusion models developed in [228] involve undirected social networks, directed social networks, and directed social networks with prior knowledge of their diffusion probabilities.

The models reviewed above employ stochastic analyses [229,230], which estimate the probability that a user will spread a piece of information. Stochastic modeling has been established as a useful strategy to analyze OSNs [231–234].

Some studies have combined a topic model and information diffusion, for example, [33]. In contrast to structural information spread models, topic-aware models for explaining information spread in OSNs analyze textural information and treat the latent topics identified in the texts as representing users' interests [235,236]. In these models, topics represent the collection of a user's interests that imply the user's intentions to interact with other users [15,44]. For example, the authors in [235] proposed a mixed latent topic model to predict users' re-posting behaviors. They assumed that users' posting behavior is influenced by three factors: breaking news, posts from OSN friends, and users' personal

interests. Some topic-aware information spread models [15,46] use the Latent Dirichlet Allocation (LDA) topic model [237] to designate that users with the same topic distribution share the same behavioral pattern (role). Most topic-aware information diffusion models consider the topics that a user engages with, but neglect the user's structural attributes, for example, neglecting the effects of role-topic pairs on the information diffusion process.

Overall, the non-structural information spread models in OSNs reviewed so far focus on modeling the spread by all network users (not only neighbors) but largely ignore the structure of the network and, thus, do not differentiate between local and global exposures to information that can lead to local and global influence, respectively. As described in Section 4.1, structural information spread models focus on modeling only local information spread as a result of local user-to-user exposure while ignoring exposure effects by non-adjacent neighbors in the network. Detecting both local and global information spread while considering network structure in a hybrid model, as described in the next subsection, is crucial to better understand human behavior online that determines users' roles [68].

4.3. Hybrid Information Spread Models

Bartal et al. [68] identified user exposure to local information spread in a dynamic OSN by considering network structure, as well as exposure to global information spread by non-neighbors. Thus, the authors extended and improved the existing network information spread models that can explain influence spread in OSNs. In another study, Bartal et al. [19] empirically detected local and global exposure to information that can lead to local/global influence in the form of retweeting a message on Twitter, considered a directed OSN $G = (V, E)$ in which nodes are users and edges represent following relationships among users. Then, the authors tracked the temporal retweet sequence of an original tweet by users in G who were represented as nodes in a temporal interaction network $G_{Tw} = (V_{Tw}, E_{Tw})$ in which edges represent retweeting activities. Overlaid together, the social network G and the interaction network G_{Tw} allow the detection of local and global information spread that resulted in influence events. *Local influence* was detected if a user retweeted an original tweet after one of the users whom s/he follows had retweeted or posted an original tweet. *Global influence* was detected if a user retweeted an original tweet before any of the users s/he follows had retweeted or posted it. Figure 5 illustrates an example of how local and global influence were detected in [19]. Given a social network G and an interaction network G_{Tw}, user v_0 posted an original tweet at time t_0, exposing user v_2 who follows v_0. Then, at time t_1, v_2 retweeted the original tweet that was posted by v_0, demonstrating local influence since v_2 follows v_0 in G. At time t_1, user v_1 also retweeted the original tweet that was posted by v_0, thus demonstrating global influence since v_1 does not follow v_0 in G. User v_1 might have been exposed to the tweet posted by v_0 via, for example, Twitter's content recommender algorithm, or actively browsing for information.

Cha et al. [238] also studied the spread of information that is not limited to user-to-user spreading mechanisms. The authors analyzed the different roles (influential users, and ordinary users who became influentials) that users play in OSNs by analyzing (i) user popularity via Indegree network centrality; (ii) information spread of a specific topic measured using retweeting activities; and (iii) value of a user in an OSN via mentioning activities by other users.

Additional studies have modeled the spread of information beyond internal network sources of information. For example, Myers et al. [27] focused on exposure to viral information that spreads to multiple network users in a short period of time via information sources that are located outside of the network such as the mainstream media. More specifically, as mentioned above, they measured the level of exposure to information using hazard functions that facilitate exposure curves, such that each new exposure increases the probability of influence. Leskovec et al. [239] analyzed a collection of 90 million articles and tracked how information in the form of phrases or memes spreads through the 1.6 million mainstream media sites and online blogs assessed. Their study presents a quantitative

analysis of how global news broadcasts spread between mainstream and social media. The authors found that there is a lag of 2.5 h between the peaks of attention to a phrase in the news media and in blogs. Regarding topics, the Topical Role Model (TRM) presents a hybrid approach [240] that analyzes how topical interests affect the information spread process and also considers user roles to explain information spreading patterns in OSNs. The TRM model assesses the role-aware topic-level diffusion analysis, which emphasizes the interplays between user role-topic pairs and their influence on information diffusion. Another non-structural information spread model in OSNs applied a logistic model [40] that predicts influence by mainly focusing on temporal and topological dynamics. As opposed to other non-structural models that ignore the structure of the network, this model takes into account one topological feature that is the distance from a user who holds an infecting role (the influence source) to a user who holds an infected role, for detecting information spread that resulted in influence.
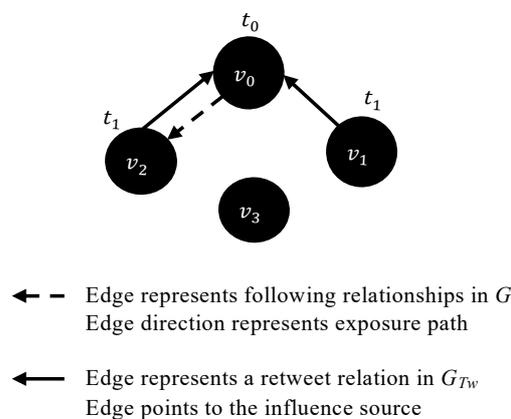

Figure 5. A directed social network $G(V, E)$ (e.g., Twitter Following–Followee relationships) with a directed interaction network $G_{Tw}(V_{Tw}, E_{Tw})$ (e.g., retweets). The interaction network at time t_1 contains the set of nodes $V_{Tw} = \{v_0, v_1, v_2\}$, and the social network contains the set of nodes $V = \{v_0, v_1, v_2, v_3\}$. In G, node v_2 follows node v_0, indicated by a dashed edge (link). Thus, v_0 exposes v_2 to information. v_2 and v_1 retweeted v_0's original tweet at time $t1$, indicated by two solid edges.

Table 2 presents a summary of the main studies about information spread in networks that are covered in the current survey. The studies are grouped into three information spread research approaches: local, global, and external information spread. In this review, 'external' information spread sources are located outside of the network, whereas 'global' information spread sources are network nodes that expose non-neighboring nodes. Hybrid models span both local and global information spread models.

Table 2. A summary of studies about information spread in networks that lead to influence, grouped by influence type.

Information Spread in the Study	Structural Models	Non-Structural Models	External Information
Resharing of online content (e.g., a message or a photo)	[21,22,49,50,202,241] [23–26,51]	[11,16]	[27,238,239]
Hybrid models of re-sharing online content	[19,20,40,68,165,240]		
Information first spreads to opinion leaders, and then from node to node	[30,164,203]		
Information on who adopted an innovation or behavior (e.g., prescribing a drug, imitation, or emotional contagion)	[199,201,210]	[18]	

Information Spread in the Study	Structural Models	Non-Structural Models	External Information
Information on who adopted an innovation or behavior depending on the user's number of adopting OSN neighbors	[22,52,206,207,209,242]	[41]	
Purchasing a product	[5,7–10,204]	[54,228]	

Whereas local influence occurs following exposure to information by network neighbors, global exposure results from exposure to information by non-neighbors and can lead to global influence. One way by which global influence can occur is by homophily between network users [18] and often requires extracting features of users [243], to apply machine learning algorithms for predicting information spread [22].

4.4. Homophily-Related Role-Aware Models

Homophily is defined as "the degree to which people who interact are similar in beliefs, education, social status", and other characteristics [244] and has been found to increase the effectiveness of communication among individuals [245]. In other words, individuals prefer to associate with others whom they perceive to be similar to themselves in terms of both values and status characteristics that reflect users' roles. Zhao et al. found that users with high homophily tend to have similar roles [122]. McCroskey et al. [245] viewed homophily as a four-dimensional construct, consisting of attitude, background, morality, and appearance. These dimensions of homophily include: (i) attitude homophily reflecting the extent to which a person perceives that another person shares his/her attitudes; (ii) background homophily indicating the extent to which a person perceives that another person shares his/her social background; (iii) value homophily denoting the extent to which a person perceives that another person shares his/her values and morals; and (iv) appearance homophily measuring the extent to which a person perceives that another person looks similar to him or her. Opinion leaders were found to have higher homophily than others on the dimensions of values, attitude, and background [245].

Similarity measures are typically used to quantify homophily among users and were able to explain the formation of relationships, and strength of interactions [246]. For example, Hanks et al. [247] found that perceived similarity to other customers in the services industries can influence both self-image congruence and self-brand congruence. Moreover, user similarity increases their commitment to the community and user perception of information quality [246,248]. Wang et al. [248] report how homophily plays an important role in determining credibility perceptions and influencing the persuasive process on both websites and online discussion groups.

Homophily affects people's intentions of seeking opinions and consumers' insights on social network sites. It may explain consumers' reactions to online content generated by another consumer, opinion leader, or celebrity, such as a vlogger [249,250]. For example, consumers are influenced by recommendations from reviewers who are similar to them [249]. People feel similar to other people who reflect their own self-image [247]. Indeed, the more similar a person feels to another person, the more likely s/he will interact frequently with that person [251].

Previous studies show that homophily is an essential aspect in the study of OSN users [251,252]. YouTube viewers are more likely to recommend a vlogger and purchase the products featured in the vlog when they perceive that vlogger to be more similar to themselves [253]. Kim et al. [254] found that higher homophily between the consumer and the website lead to a positive attitude toward the website and toward the associated information posted on that website. Sakib et al. [255] found that homophily between a vlogger and his/her audience increased social interactions (in a context of weight loss).

The majority of recent influence spread studies in networks (e.g., [22–25,48–52]) focus on influence spread of viral information, but largely ignore influence spread of non-viral information, as discussed next.

5. Viral vs. Non-Viral Information Spread in Online Social Networks (OSNs)

In their study of viral information spread that produces influence via complex contagion, Romero et al. [16] evaluated users' exposure to the 500 most frequent hashtags in a Twitter dataset consisting of over 3 billion messages. As mentioned above, the "Yes We Can" slogan [53] was employed in the 2008 U.S. presidential election in both mainstream and social media, initially by users who played the role of opinion leaders, and subsequently by standard users who served other roles. This slogan reached millions of voters and contributed to mobilizing millions of volunteers to actively support the ultimately successful presidential campaign [256]. The above are examples of viral information spread serving as the primary source of influence spread.

To date, most studies have focused on the spread of viral information in networks; however, no consensus exists among those studies regarding the number of influence events that define the virality of an information nugget, a unit of information such as a message or an image. For instance, Dow et al. [17] assessed a Facebook dataset consisting of approximately 1 million images, in order to study how influence diffuses via users' activities to share images on that platform. The authors identified users who hold an influential role, using two linear regression models fitted with least squares that predict the importance of a node, defined by: (i) number of nodes who shared the information posted by the node in focus, and (ii) the size of the cascade generated by the node in focus. The linear regression results (maximal $R^2 = 0.49$) show that the audience size variable yielded the greatest explanatory power; other variables explained less than 1% of the variance. That study designated images that were shared a minimum of 100 times as being viral; this definition represents a fairly small number of influence events, relative to other studies. Providing an even broader definition of virality, Myers et al. [27] specified that posts shared at least 50 times are viral; this is one of the lowest thresholds of event counts among studies to detect influence. The authors presented a model in which information can reach a node via: (i) the links of the social network, modeled using a hazard function: $\lambda_{int}(t)dt \equiv P(i$ exposes $j \in [t, t + dt]|i$ has not exposed j yet), where nodes i and j are neighbors, and t is the time passed since node i was infected. λ_{int} represents the time it takes a node to observe an infected neighbor; or (ii) through the influence of external sources: $\lambda_{ext}(t)dt \equiv P(i$ receives exposure $\in [t, t + dt))$; here t is the time passed since any contagion occurred. The authors found that approximately 71% of the information can be attributed to network diffusion, and the remaining 29% is due to exposure to information outside the network. Liben-Nowell and Kleinberg [49] quantified the depth reach of influence events involving the re-sharing of information by modeling the waiting time of a user before infection. The waiting time was modeled according to the density function $f(x) = x^{-\alpha}$. The authors found that events with broad reach can influence millions of users. In that study, the depth reach of influence events was measured starting from the originating user to the farthest user connected by a path of network edges to the originating user.

Numerous studies have reported that the number of users influenced by information nuggets follows a long-tailed distribution [22,54], that is, the majority of information instances are shared minimally [17,20,22,55], and most information nuggets are never shared. This observation suggests that the influence resulting from viral versus non-viral information spread occurs via different mechanisms that are detectable by observing the spreading patterns characteristic of these phenomena [19]. Furthermore, detecting these spreading patterns of influence can inform predictions during early stages of information transmission about whether an information nugget will go viral. For example, it was found [19] that the spread type of a contagion can be predicted in its early stages using two logistic regression models, one for viral ($F1 = 0.84$), and the other for non-viral ($F1 = 0.87$) information with Δt, the time difference from the posting of the original

tweet variable, presenting the greatest predictive power. Other studies [23,48] developed a strategy to detect whether an information instance will become viral by assessing the reach size (or cascade) of information using data including hashtags [50], behavioral dynamics characteristics [24], and YouTube views [257].

Network features have also informed predictions of influence cascades' reach size [22,52]. Gleeson and Durrett [258] reported that network structure and temporal dynamics are able to explain the observed spreading patterns of non-viral cascades. Cui et al. [51] developed a logistic model that employs the importance-rank of each network node, provided a list of users who were previously influenced. While these strategies are effective for predicting the reach size of information diffusion, the majority are unable to predict virality during the early stages of information spread.

This challenge of predicting whether an information nugget will become viral during the early phases of its transmission, before it has had the opportunity to become viral, was successfully addressed in [19]. In that study, the authors innovated a Back-In-Time (BIT) strategy to assess the spreading patterns of viral tweets during the time period before those tweets became viral. They subsequently compared the spreading patterns of viral tweets before they became viral with the spreading patterns associated with non-viral tweets. In that study, an information nugget was defined as viral if it had been shared a minimum of 100 times, and as non-viral if it had been shared 10 to 99 times [19]. Information nuggets that had been shared fewer than 10 times were excluded from that analysis because the influence spread signature of such instances was too low to permit the effective assessment of influence spread.

In more detail, the BIT strategy initially assesses a viral tweet back in time at a point when it had been non-viral, that is, it had been retweeted 10 to 99 initial times. These tweets that were rolled back in time are termed BIT-tweets. To test whether there are significant differences in the early spreading patterns of BIT vs. non-viral tweets, the authors plotted the empirical cumulative distribution functions (ECDFs) of the time differences between each two consecutive retweets (inter-retweet times) of BIT and non-viral tweets. A Kolmogorov–Smirnov test [259] revealed significant differences ($p_{value} \leq 2.2 \times 10^{-16}$) in the distance (D-statistic) between the ECDFs of BIT vs. non-viral tweets. Thus, BIT-tweets and non-viral tweets were characterized by different temporal spreading patterns that were identified by comparison of inter-retweet times.

These authors [20] analyzed the spread of non-viral information on Twitter that can cause the spread of either local or global influence. They considered the topic of a tweet via analysis of users' engagement in retweeting non-viral content, and they reported that influence spread both locally and globally. The percentages of events associated with global influence were discovered to vary by topic, ranging from 8% for topics related to Crime, to 28% for topics associated with Technology. Bartal et al. [20] reported that global influence is more likely to be a result of users' exposure to information promoted by external sources that can include mass media, information searches via the Twitter search box, or content recommender algorithms.

Recommender systems aim to predict items or information that a user will enjoy consuming by providing users with a personalized experience [260]. However, these systems often narrow user exposure to several types of information termed 'filter bubbles' [38] and might lead users to ideological polarization due to repeated exposure to similar content [261]. While many recommender systems ignore unpopular or newly introduced content, some systems encourage the recommendation of unpopular content [262]. These unpopular content items affiliate with the long tail of information distribution where most information is unpopular, and the minority of information is highly popular [22,54].

Local and global influence spread following exposure to viral and non-viral topics on Twitter was examined in greater depth in [19]. In their analysis, a topic's influence spread was detected when an individual user tweeted several messages about the same subject, and each tweet was retweeted by various other network members. In this study [19], a topic's influence was defined somewhat differently than the influence of an individual

information nugget: given a social network G, an interaction network G_{Tw}, and a set of tweets W, posted on the same topic by user v_j: (i) *Local topic contagion* of user v_i occurs if $e_{ij} \in E_{Tw}$ and $e_{ji} \in E$. In other words, v_i retweeted $w \in W$ after one of the users s/he follows had retweeted or written any $w \in W$; and (ii) *Global topic contagion* of user v_i occurs if $e_{ji} \notin E$ and $e_{ij} \in E_{TW}$. That is, v_i shared $w \in W$ prior to when any of the users s/he follows posted or shared any $w \in W$. Figure 6 illustrates local and global topic influence processes by which a user can be influenced by different tweets on the same topic.
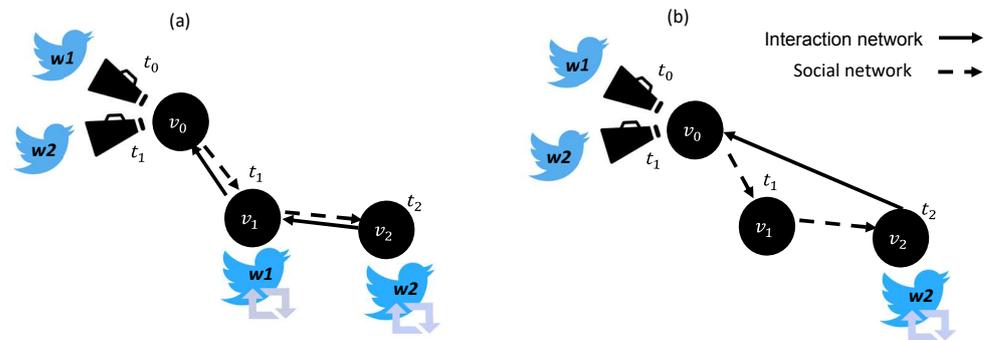

Figure 6. An illustration of local and global topic influence. Local and global topic influence of a set of tweets on the same topic $W = \{w_1, w_2\}$ that were posted by user v_0 at times t_0 and t_1 respectively. (**a**) Local topic influence: user v_1 who follows user v_0 retweets w_1 at time t_1. Another example of local topic influence occurs after user v_2 was exposed to w_1 at t_1 by user v_1 whom s/he follows. Then, user v_2 retweets tweet w_2 on the same topic as w_1. (**b**) Global topic influence: user v_2 retweets w_2 before any of the users who are followed by v_2 retweeted/posted tweet from W.

In their study, the virality of a topic was predicted by summing the number of retweets of each tweet addressing the same topic that were posted by an individual user. While this study reported evidence of global influence, the authors reported that local influence was the most frequent infection mechanism for the spread of both a single information nugget and a topic [19]. In addition, viral information and viral topics were characterized by similar influence spreading patterns. Similarly, non-viral information and non-viral topics had similar spreading patterns of influence [19]. In contrast, viral and non-viral information and topics exhibited substantially different spreading patterns from each other, and local and global influence spread patterns were also seen to differ significantly [19].

The analysis of role-aware information spread in OSNs typically involves the development of software, and many platforms have been developed to explore user roles in information spread in OSNs, as discussed next.

6. Software Platforms for Role-Aware Analysis of Online Social Networks (OSNs)

Reviewing the literature, we differentiate between two types of software platforms for the role-aware analysis of OSNs, focused on an egocentric perspective of a single user and a global view of the network as a whole. The following section presents recent platforms exemplifying both analytical approaches (Table 3).

6.1. Platforms for Egocentric Role-Aware Social Network Analyses

In contrast to whole-network analyses, egocentric analysis involves personal communities related to single individuals, permitting detailed insights into "network neighborhoods" [263]. The focus of egocentric analysis is on a single central node (the ego) and its associations with its neighbors (alters). As an early example of role-aware information spread in OSNs, Xiong and Donath [264] created PeopleGarden, a platform that provides data portraits of users. Each user is represented by a flower-like icon, the properties of which are defined by the quantity and characteristics (initial posts vs. replies) of their messages. The roles of community members, and their relationships, are easily visible using these flower representations. Communities of users are represented as "gardens" of

these icons, and can reveal high-level characteristics of OSN groups. PeopleGarden permits analysts to identify experts in the community, determine whether the group welcomes newcomers, understand the involvement of individual group members, and assess the overall activity patterns of the group.

A more recent example includes the D-map+ [265] algorithm that provides an integrated egocentric and event-centric model to interactively explore user behaviors and information spreading patterns. The platform's egocentric analysis involves collecting data for the users who re-posted a message shared by the central user (ego) of interest. In contrast, its event-centric analysis focuses on one particular event, and examines all of the users associated with sharing and re-posting messages related to that event. D-map+ visualizes OSN users as hex nodes, with their behaviors and roles encoded by color and size. This platform permits the visual identification of important users and events.

Dynamic Analyses. Some of these platforms permit dynamic analyses of ego-networks. The Episogram software of Cao et al. [266] represents the agents and objects involved in interaction processes as a dynamic tripartite network, permitting the visualization of behavior patterns. This platform permits the identification of initiators and responders participating in interaction processes. Adopting a user-centric strategy, VASABI [267] provides dynamic hierarchical user profiles using data collected during OSN usage sessions, in addition to tasks extracted via topic modeling. This platform permits analysts to stratify user behavior and view high-level summaries of trends, and analysis can be performed at the levels of individual users and groups.

To permit the visualization of evolutionary influence graphs, Eiffel [268] provides visual summaries of influence graphs based on individual nodes, node relations, and temporal analyses. The relationships between an influencer of interest and the other categories of users whom they influence are presented in a novel flow map visualization that supports both flip-book and movie-based evolutionary visualizations to explore graph dynamics.

Anomaly Detection. Some egocentric visualization platforms are specifically designed to recognize anomalous user behaviors, which can be used to identify bots and other malicious entities in OSNs. The TargetVue platform of Cao et al. [269] is intended to detect anomalous users using unsupervised learning. The platform provides three novel glyph visualizations that present the ego node's features, communication activities, and social interactions. These glyphs are arranged on a triangle grid that visualizes similarities across users and permits comparing the behaviors of users. The authors demonstrated the power of this platform to detect bots on the Twitter platform.

The egoStellar platform [270] provides the ability to analyze the communication behaviors of mobile users via an ego-network perspective; a specific objective of this platform is to identify anomalous behaviors and associated fraudulent and solicitor users. Three views are provided in its graph model, including a high-level statistical view to display the distribution of mobile users; a group view that permits classifying users and extracting features for anomalous behavior detection; and an egocentric view that details the interactions of each ego and its alters. More recently, the egoDetect platform [271] provides a novel visualization system for anomaly detection. This platform uses unsupervised machine learning that offers efficient anomaly detection without the need for training. The novel glyph visualization provided by this network permits the exploration of each ego's topology and its relationships with its alters. Specifically, this platform permits the identification of different categories of alters, including local alters and alien alters, with the latter category indicating agents exhibiting anomalous behaviors.

6.2. Additional Platforms for Role-Aware Social Network Analyses

Additional platforms provide role-aware analysis of OSNs from other, more holistic perspectives than egocentricity (Table 3). Developed for use by service providers who wish to understand the characteristics of their users, iVIS [272] provides interactive visualizations of behavioral patterns and clustered views of behaviors. It permits the identification of light vs. heavy users, as well as more descriptive user categories including testers (those

who used the service only for testing purposes) and frustrated users (those who examined a site's data, but did not ultimately use it).

Table 3. Software platforms for role-aware analyses of online social networks. Rows are sorted primarily by year of publication, and secondarily by platform name.

Platform	Description	Publication Year	Reference
egoDetect	Detect and explore anomalies (alien alters) via unsupervised learning; novel glyph for ego topology	2020	[271]
Eiffel	View nodal, relational and temporal dimensions of evolutionary influence graphs to see influencer effects on other users	2020	[268]
iVIS	Identify light/heavy users and user categories via clustering	2020	[272]
VASABI	Analyze dynamic hierarchies at individual and group levels to identify user roles	2019	[267]
D-map+	View egocentric and event-centric information diffusion patterns; identify behaviors and roles	2018	[265]
egoStellar	Visualize anomalous users and behaviors via egocentric perspective	2018	[270]
MessageLens	Analyze learner attitudes, interactions among students, and discussion topics	2018	[273]
VisForum	Explore user groups in forums; identify high-impact forum members	2018	[274]
iForum	Analyze users, posts, and threads on three different scales; identify new, active and inactive users	2016	[275]
Episogram	Display dynamic egocentric social interactions to identify initiators and responders	2015	[266]
TargetVue	Identify anomalous users and behaviors via glyph visualizations	2015	[269]

VisForum [274] enables analysts to assess users and their roles in online forums, by presenting three novel glyph visualizations, at the group, user, and set levels, with different granularities. Their novel sorting algorithm reduces noise in the data, and their forum-index concept is used to identify high-impact forum members. For analyses focused on education, numerous platforms have been developed to understand the behavior and roles of course discussion forums [276]. Analytical objectives can include predicting students' performance in courses, understanding sentiments, and predicting patterns of social behavior [276]. Examples of software for the prediction of behavior include the iForum platform [275], which permits the analysis of users, posts, and threads on three different scales, ranging from the complete forum down to the level of an individual thread or user. This platform allows the identification of various categories of users, including new users, active users, and inactive users. Providing a different perspective, the MessageLens visual analytics platform [273] allows instructors to understand their course forums based on discussion topic, interactions among students, and learner attitudes, permitting students to be classified by their behaviors.

The software platforms presented in this section, and listed in Table 3, permit the efficient role-aware analysis of information spread in OSNs. Establishing a standard to make this software more freely available would help to advance research on the behaviors and roles associated with complex human networks.

7. Conclusions and Future Directions

This review has surveyed the state of research on the role-aware analysis of information spread in online social networks (OSNs). Previous reviews of information spread in OSNs have mainly addressed local user-to-user propagation of viral information; however, recent studies have reported the importance of global spread mechanisms and their associated influence, and the impact of non-viral information spread. This review has addressed those gaps by: (i) providing a comprehensive survey of the latest studies on role-

aware information spread in OSNs, describing the different temporal spreading patterns of viral and non-viral information and how user roles affect the spread of information; (ii) surveying modeling approaches for information spread in OSNs that consider structural and non-structural features as well as recent hybrid models that integrate both strategies, and presenting an associated taxonomy; (iii) providing an overview of software platforms used for the analysis and visualization of role-aware information spread in OSNs; and (iv) describing how role-aware information spread models enable useful applications in OSNs such as detecting influential users.

The current state of research on the role-aware analysis of information spread in OSNs suggests many promising future directions that call for future research, as described next.

The Internet of People (IoP). The Internet of Things (IoT) paradigm [277] provides a framework connecting the devices, sensors, actuators, protocols, and cloud services that have become important elements in managing the daily modern life. However, this paradigm is infrastructure-centric while regarding the human users as peripheral, thus neglecting an essential component of the system. To address this oversight, Conti et al. [278] have proposed an Internet of People (IoP) paradigm in which humans and their devices are active elements; the IoP paradigm is intended to enhance and extend, rather than replace, the existing IoT infrastructure [279]. In this framework, Dunbar's ego-network model [280,281] can be used to map each individual's social network to a concentric circle specifying different levels of membership, including the inner support clique, an intermediate sympathy group, a larger affinity group that is composed of an extended set of more distant friends, and an active network composed of 150 alters with which the individual maintains annual contact [278]. This model, originally developed to analyze real-world relationships, has subsequently been reported to usefully represent online social relationships, as well [282]. By accounting for these relationship structures, information diffusion in OSNs has been effectively modeled [283].

A central feature of the IoP is that human behavior models are embedded in its algorithms, requiring a broad integration among disciplines including psychology, sociology, anthropology, economics, and the science of complex networks [279]. Efforts to accurately represent the complexity of human behavior in OSNs via the integration of these disciplines are in their early stages. Continuing research on the behaviors and roles of OSN users will certainly inform the future developments of the Internet of People.

Software Availability. Unfortunately, the majority of the software platforms listed in Table 3 and described in Section 6 are not freely available; this prevents the widespread use of these platforms to inform research. Looking forward, it will be ideal if funding agencies require the open sharing of publicly funded software platforms, to ensure that these resources will be widely available to advance our understanding of the behavior and roles of individuals in OSNs.

Visualization. Novel immersive, interactive virtual reality (VR) and augmented reality (AR) visualization strategies have recently been revolutionizing many disciplines and fields of research [284,285]. While the full potential of the nascent VR and AR technologies has yet to be realized [286], the use of these technologies for network analysis offers a number of benefits. Three-dimensional (3D) VR representations of networks are reported to provide numerous benefits when compared with classic 2D networks [287], including facilitating accurate distance assessments between nodes [288], path finding among large numbers of nodes [289,290], community detection in complex graphs [291,292], and facilitating the accuracy of users' mental models [293]. VR and AR platforms also lend themselves to collaborative work [294–297], including the creation of a 3D layout of a large Twitter network that provides collaborative interactions [298].

In recent years, these visualization strategies have started to be applied to the analysis of OSNs. The VRige platform [299] permits social network interactions to be explored in an immersive virtual environment. The VR-based framework of Sorger et al. [300] provides two navigation modes, overview exploration and immersive detail analysis, to analyze large dynamic networks. Looking ahead, the new field of cross-virtuality [301] has the

goal of seamlessly integrating and transitioning among conventional 2D visualizations, augmented reality, and virtual reality; such integrative implementations have not yet been explored for the analysis of OSNs, and these represent an interesting future direction. Ens et al. [294] suggest that the use of immersive analyses might change the analytical process itself, promising novel strategies for the analysis of OSNs. Among these possibilities, integrating machine intelligence with human intelligence in immersive environments [294] promises to reveal new pathways for exploration of and discovery in OSNs. Finally, and importantly, to date, efforts to use VR and AR for network analysis have only recently started to include a role-aware focus [287], and this remains a promising and potentially very helpful direction for future work.

Understanding the Complexity of Human Behavior Online. As discussed in Section 2, OSN users' roles are dynamic and varied. Humans are complex agents whose personality traits [302–305], habits [306], preferences [307], and interests [304,308] can evolve over time, and may vary based on a variety of factors. To date, role-aware analyses of OSNs have captured only a fraction of the complexity of dynamic human personality and behavior, and future research will benefit from considering additional facets of these phenomena to achieve more accurate predictions.

Many studies have developed strategies to identify anomalous behaviors for the purpose of eliminating malicious users and improving the social network experience for community members [309–314]. Rather than online behavior that is labeled as antisocial being limited to a small, vocal group of provocateurs, recent research has reported that standard OSN users can be provoked into becoming trolls or engaging in other antisocial behavior, influenced by the user's mood and the context of the discussion [315]. Similarly, communications labeled as hate speech can be triggered in users by their OSN community members' behaviors [316,317] as well as being more likely when specific topics are discussed [318,319]. Such research suggests that at least a subset of OSN users perceived as malevolent might become more positive contributors to OSNs by the thoughtful shaping of online environments.

Differentiating among various forms of behaviors labeled as antisocial in OSNs [315] will permit further granularity and informed control when managing these networks. In this vein, analysis to understand users' intentions can inform the effective management of OSNs. For example, a user who posts messages that are assessed to be trolling content could have malicious intent, or might simply be expressing a sincere opinion that differs from those of the other forum members [315,320]. Additionally, users banned from OSN communities have been reported to form two distinct groups: those whose posts were regularly deleted by moderators prior to being banned, and those whose posts were only recently deleted before they were banned [321]. Extending research to differentiate habitual trolls from users who may have engaged in a specific heated debate [321] would be helpful to moderate online communities effectively, to avoid deplatforming users who do not habitually cause problems. Because the demographics, experiences, beliefs, and customs of OSN users can vary widely across OSN platforms, the definitions of social acceptability and toxic content differ by platform [322–324]. Accordingly, toxic content classifiers that use a single set of rules regardless of the OSN community will not capture the nuance of individual communities, and work is ongoing to tailor such standards to different groups of users [322]. Accurately characterizing users' behaviors [325,326] is necessary to define roles and manage online communities effectively. Because ignoring the perspectives and needs of stakeholders in online communities can result in significant problems [323], it is important for role-aware OSN analyses to consider the perspectives of individual users, to ensure that each community member has an opportunity to participate and to feel that s/he is being heard. Ongoing work to understand and define social contexts [327], social acceptability [327,328], and the roles of OSN users will guide efforts to effectively manage OSN communities.

As discussed in the Internet of People subsection above, the coordination of network science with diverse fields including psychology, sociology, anthropology, and philosophy

will allow us to achieve a deeper understanding of individuals' behavior in interacting with information and other users in OSNs. Insightful integration of research across these diverse disciplines will facilitate network science, with a role-aware focus, to develop and moderate OSNs that will maximally benefit each individual community member.

Author Contributions: Author contributions: Conceptualization, A.B. and K.M.J.; resources, A.B. and K.M.J.; writing—original draft preparation, A.B. and K.M.J.; writing—review and editing, A.B. and K.M.J.; writing—revision based on peer reviewers' feedback, A.B. and K.M.J.; table preparation, A.B. and K.M.J.; visualization, A.B. and K.M.J.; taxonomy development, A.B. and K.M.J.; Project administration, A.B. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

2D	2-dimensional
3D	3-dimensional
AR	Augmented reality
BIT	Back-In-Time
BHLFM	Bayesian Hierarchical Latent-Factor Model
DAGs	Dynamic directed acyclic graphs
DMMSB	Dynamic Mixed-Membership Stochastic Blockmodels
DySTGNN	Dynamic structural-temporal graph neural network
GAT	Graph attention network
GCN	Graph convolutional network
GNN	Graph neural networks
IC	Independent Cascade model
IoP	Internet of People
IoT	Internet of Things
LDA	Latent Dirichlet Allocation
LIM	Linear Influence Model
LTM	Linear Threshold Model
LSTM	Long Short-Term Memory
MMSB	Mixed-Membership Stochastic Blockmodels
NLP	Natural Language Processing
NMF	Non-negative Matrix Factorization
OSN	Online social network
RNN	Recurrent neural network
RAFM	Role Affiliation Frequency Model
SVD	Singular Value Decomposition
SEIR	Susceptible–Exposed–Infected–Recovered
SIR	Susceptible–Infected–Recovered
SIRS	Susceptible–Infected–Recovered-for Susceptible
SIS	Susceptible-Infected-Susceptible
TRM	Topical Role Model
VR	Virtual reality

References

1. Adali, S.; Golbeck, J. Predicting personality with social behavior. In Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Istanbul, Turkey, 26–29 August 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 302–309.
2. Lin, K.C.; Wu, S.H.; Chen, L.P.; Yang, P.C. Finding the key users in Facebook fan pages via a clustering approach. In Proceedings of the IEEE International Conference on Information Reuse and Integration, San Francisco, CA, USA, 13–15 August 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 556–561.
3. Watts, D.J.; Dodds, P.S. Influentials, Networks, and Public Opinion Formation. *J. Consum. Res.* **2007**, *34*, 441–458. [CrossRef]
4. Zhang, Z.K.; Liu, C.; Zhan, X.X.; Lu, X.; Zhang, C.X.; Zhang, Y.C. Dynamics of Information Diffusion and Its Applications on Complex Networks. *Phys. Rep.* **2016**, *651*, 1–34. [CrossRef]
5. Turner, J.C. *Social Influence*; Thomson Brooks/Cole: Pacific Grove, CA, USA, 1991.
6. Tu, C.; Liu, Z.; Sun, M. PRISM: Profession identification in social media with personal information and community structure. In Chinese National Conference on Social Media Processing; Springer: Berlin/Heidelberg, Germany, 2015; pp. 15–27.
7. Jurvetson, S. What Exactly Is Viral Marketing. *Red Herring* **2000**, *78*, 110–112.
8. Richardson, M.; Domingos, P. Mining Knowledge-sharing Sites for Viral Marketing. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, AB, Canada, 23 July 2002; pp. 61–70.
9. Kaplan, A.M.; Haenlein, M. Two Hearts in Three-quarter Time: How to Waltz the Social Media/viral Marketing Dance. *Bus. Horizons* **2011**, *54*, 253–263. [CrossRef]
10. Fournier, S.; Avery, J. The Uninvited Brand. *Bus. Horizons* **2011**, *54*, 193–207. [CrossRef]
11. Daley, D.J.; Kendall, D.G. Stochastic Rumours. *Ima J. Appl. Math.* **1965**, *1*, 42–55. [CrossRef]
12. Li, M.; Wang, X.; Gao, K.; Zhang, S. A survey on information diffusion in online social networks: Models and methods. *Information* **2017**, *8*, 118. [CrossRef]
13. Hsu, L.C.; Wang, K.Y.; Chih, W.H.; Lin, K.Y. Investigating the ripple effect in virtual communities: An example of Facebook Fan Pages. *Comput. Hum. Behav.* **2015**, *51*, 483–494. [CrossRef]
14. Borge-Holthoefer, J.; Banos, R.A.; González-Bailón, S.; Moreno, Y. Cascading Behaviour in Complex Socio-technical Networks. *J. Complex Netw.* **2013**, *1*, 3–24. [CrossRef]
15. Xu, Z.; Zhang, Y.; Wu, Y.; Yang, Q. Modeling user posting behavior on social media. In Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, Portland, OR, USA, 12–16 August 2012; pp. 545–554.
16. Romero, D.M.; Meeder, B.; Kleinberg, J. Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter. In Proceedings of the 20th International Conference on World Wide Web, Hyderabad, India, 28 March–1 April 2011; ACM: New York, NY, USA, 2011; pp. 695–704.
17. Dow, P.A.; Adamic, L.A.; Friggeri, A. The Anatomy of Large Facebook Cascades. In Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, Cambridge, MA, USA, 28 June 2013; pp. 145–154.
18. Bakshy, E.; Messing, S.; Adamic, L.A. Exposure to Ideologically Diverse News and Opinion on Facebook. *Science* **2015**, *348*, 1130–1132. [CrossRef]
19. Bartal, A.; Pliskin, N.; Tsur, O. Local/Global Contagion of Viral/non-viral Information: Analysis of Contagion Spread in Online Social Networks. *PLoS ONE* **2020**, *15*, e0230811. [CrossRef]
20. Bartal, A.; Ravid, G.; Tsur, O. Global Contagion of Non-Viral Information. In Proceedings of the 53rd Hawaii International Conference on System Sciences, Maui, HI, USA, 7–10 January 2020; pp. 2803–2812.
21. Bakshy, E.; Rosenn, I.; Marlow, C.; Adamic, L. The Role of Social Networks in Information Diffusion. In Proceedings of the 21st International Conference on World Wide Web, Lyon, France, 16 April 2012; ACM: New York, NY, USA, 2012; pp. 519–528.
22. Cheng, J.; Adamic, L.; Dow, P.A.; Kleinberg, J.M.; Leskovec, J. Can Cascades Be Predicted? In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Korea, 27 April 2014; ACM: New York, NY, USA, 2014; pp. 925–936.
23. Subbian, K.; Aggarwal, C.; Srivastava, J. Content-centric Flow Mining for Influence Analysis in Social Streams. In Proceedings of the 22nd ACM International Conference on Information and Knowledge Management, San Francisco, CA, USA, 27 October 2013; ACM: New York, NY, USA, 2013; pp. 841–846.
24. Yu, L.; Cui, P.; Wang, F.; Song, C.; Yang, S. From Micro to Macro: Uncovering and Predicting Information Cascading Process With Behavioral Dynamics. In Proceedings of the IEEE International Conference on Data Mining, Atlantic City, NJ, USA, 14–17 November 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 559–568.
25. Subbian, K.; Prakash, B.A.; Adamic, L. Detecting Large Reshare Cascades in Social Networks. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017; International World Wide Web Conferences Steering Committee: Geneva, Switzerland, 2017; pp. 597–605.
26. Vaidya, M. Ice Bucket Challenge Cash May Help Derisk ALS Drug Research. *Nature* **2014**, *201*, 4. [CrossRef]
27. Myers, S.A.; Zhu, C.; Leskovec, J. Information Diffusion and External Influence in Networks. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; ACM: New York, NY, USA, 2012; pp. 33–41.
28. Tambuscio, M.; Ruffo, G.; Flammini, A.; Menczer, F. Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks. In Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, 18–22 May 2015; pp. 977–982.

29. Yuan, C.; Ma, Q.; Zhou, W.; Han, J.; Hu, S. Jointly embedding the local and global relations of heterogeneous graph for rumor detection. In Proceedings of the IEEE International Conference on Data Mining (ICDM), Beijing, China, 8–11 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 796–805.

30. Yin, M.; Gray, M.L.; Suri, S.; Vaughan, J.W. The Communication Network Within the Crowd. In Proceedings of the 25th International Conference on World Wide Web, Montreal, QC, Canada, 11–15 April 2016; pp. 1293–1303.

31. Bartal, A.; Ravid, G. Member behavior in dynamic online communities: Role affiliation frequency model. *IEEE Trans. Knowl. Data Eng.* **2019**, *32*, 1773–1784. [CrossRef]

32. Guille, A.; Hacid, H.; Favre, C.; Zighed, D.A. Information diffusion in online social networks: A survey. *ACM Sigmod Rec.* **2013**, *42*, 17–28. [CrossRef]

33. Yang, Y.; Tang, J.; Leung, C.W.k.; Sun, Y.; Chen, Q.; Li, J.; Yang, Q. Rain: Social role-aware information diffusion. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 9 February 2015.

34. Wu, S.; Hofman, J.M.; Mason, W.A.; Watts, D.J. Who says what to whom on twitter. In Proceedings of the 20th International Conference on World Wide Web, Hyderabad, India, 28 March–1 April 2011; pp. 705–714.

35. Peng, S.; Yu, S.; Yang, A. Smartphone Malware and Its Propagation Modeling: A Survey. *IEEE Commun. Surv. Tutor.* **2013**, *16*, 925–941. [CrossRef]

36. Zhou, X.; Jin, Q. A heuristic approach to discovering user correlations from organized social stream data. *Multimed. Tools Appl.* **2017**, *76*, 11487–11507. [CrossRef]

37. Chen, W.; Yuan, Y.; Zhang, L. Scalable influence maximization in social networks under the linear threshold model. In Proceedings of the IEEE International Conference on Data Mining, Sydney, NSW, Australia, 13–17 December 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 88–97.

38. Pariser, E. *The Filter Bubble: What the Internet Is Hiding from You*; Penguin: London, UK, 2011.

39. Yang, J.; Leskovec, J. Modeling Information Diffusion in Implicit Networks. In Proceedings of the IEEE International Conference on Data Mining, Sydney, NSW, Australia, 13–17 December 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 599–608.

40. Wang, W.; Zhou, H.; He, K.; Hopcroft, J.E. Learning Latent Topics From the Word Co-occurrence Network. In *National Conference of Theoretical Computer Science*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 18–30.

41. Wang, F.; Wang, H.; Xu, K. Diffusive Logistic Model Towards Predicting Information Diffusion in Online Social Networks. In Proceedings of the 32nd International Conference on Distributed Computing Systems Workshops, Macau, China, 18–21 June 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 133–139.

42. Mele, A. A structural model of homophily and clustering in social networks. *J. Bus. Econ. Stat.* **2021**. (Just Accepted). [CrossRef]

43. McPherson, M.; Smith-Lovin, L.; Cook, J.M. Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* **2001**, *27*, 415–444. [CrossRef]

44. Zhang, J.; Liu, B.; Tang, J.; Chen, T.; Li, J. Social influence locality for modeling retweeting behaviors. In Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, Beijing, China, 30 June 2013.

45. Aslay, C.; Barbieri, N.; Bonchi, F.; Baeza-Yates, R. Online Topic-aware Influence Maximization Queries. In Proceedings of the 17th International Conference on Extending Database Technology (EDBT), Athens, Greece, 24–28 March 2014.

46. Zhang, J.; Tang, J.; Zhuang, H.; Leung, C.; Li, J. Role-aware conformity modeling and analysis in social networks. In Proceedings of the AAAI Conference on Artificial Intelligence, Québec, QC, Canada, 21 June 2014; Volume 28.

47. Burt, R.S. *Structural Holes*; Harvard University Press: Cambridge, MA, USA, 1992.

48. Leskovec, J.; McGlohon, M.; Faloutsos, C.; Glance, N.; Hurst, M. Patterns of Cascading Behavior in Large Blog Graphs. In Proceedings of the 2007 SIAM International Conference on Data Mining, Minneapolis, MN, USA, 26 April 2007; SIAM: Philadelphia, PA, USA, 2007; pp. 551–556.

49. Liben-Nowell, D.; Kleinberg, J. Tracing Information Flow on a Global Scale Using Internet Chain-letter Data. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4633–4638. [CrossRef]

50. Yang, J.; Leskovec, J. Patterns of Temporal Variation in Online Media. In Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, Hong Kong, China, 9–12 February 2011; ACM: New York, NY, USA, 2011, pp. 177–186.

51. Cui, P.; Jin, S.; Yu, L.; Wang, F.; Zhu, W.; Yang, S. Cascading Outbreak Prediction in Networks: A Data-driven Approach. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11 August 2013; ACM: New York, NY, USA, 2013; pp. 901–909.

52. Wang, S.; Yan, Z.; Hu, X.; Philip, S.Y.; Li, Z. Burst Time Prediction in Cascades. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015; pp. 325–331.

53. Nahon, K.; Hemsley, J.; Walker, S.; Hussain, M. Fifteen Minutes of Fame: The Power of Blogs in the Lifecycle of Viral Political Information. *Policy Internet D* **2011**, *3*, 1–28. [CrossRef]

54. Bild, D.R.; Liu, Y.; Dick, R.P.; Mao, Z.M.; Wallach, D.S. Aggregate Characterization of User Behavior in Twitter and Analysis of the Retweet Graph. *ACM Trans. Internet Technol. (TOIT)* **2015**, *15*, 4. [CrossRef]

55. Weng, L.; Flammini, A.; Vespignani, A.; Menczer, F. Competition among memes in a world with limited attention. *Sci. Rep.* **2012**, *2*, 1–9. [CrossRef] [PubMed]

56. Yujie, Y. A Survey on Information Diffusion in Online Social Networks. In Proceedings of the 2020 European Symposium on Software Engineering, Rome, Italy, 6–8 November 2020; pp. 181–186.

57. Chang, B.; Xu, T.; Liu, Q.; Chen, E.H. Study on information diffusion analysis in social networks and its applications. *Int. J. Autom. Comput.* **2018**, *15*, 377–401. [CrossRef]

58. Vega, D.; Magnani, M.; Montesi, D.; Meseguer, R.; Freitag, F. A new approach to role and position detection in networks. *Soc. Netw. Anal. Min.* **2016**, *6*, 1–23. [CrossRef]

59. Mumford, T.V.; Van Iddekinge, C.H.; Morgeson, F.P.; Campion, M.A. The Team Role Test: Development and validation of a team role knowledge situational judgment test. *J. Appl. Psychol.* **2008**, *93*, 250. [CrossRef]

60. Heiss, J. Social roles. In *Social Psychology*; Routledge: Oxford, UK, 2017; pp. 94–130.

61. Ebaugh, H.R.; Ebaugh, H.R.F. *Becoming an ex: The Process of Role Exit*; University of Chicago Press: Chicago, IL, USA, 1988.

62. Biddle, B.J. *Role Theory: Expectations, Identities, and Behaviors*; Academic Press: Cambridge, MA, USA, 2013.

63. Drouin, M.; McDaniel, B.T.; Pater, J.; Toscos, T. How parents and their children used social media and technology at the beginning of the COVID-19 pandemic and associations with anxiety. *Cyberpsychol. Behav. Soc. Netw.* **2020**, *23*, 727–736. [CrossRef]

64. Girvan, M.; Newman, M.E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826. [CrossRef] [PubMed]

65. Neal, Z.P. *The Connected City: How Networks Are Shaping the Modern Metropolis*; Routledge: Oxford, UK, 2012.

66. Nielsen, J. Participation Inequality: Encouraging More Users to Contribute. 2006. Available online: http://www.useit.com/alertbox/participation_inequality.html (accessed on 21 September 2021).

67. Choobdar, S.; Ribeiro, P.; Parthasarathy, S.; Silva, F. Dynamic inference of social roles in information cascades. *Data Min. Knowl. Discov.* **2015**, *29*, 1152–1177. [CrossRef]

68. Bartal, A.; Pliskin, N.; Ravid, G. Modeling Influence on Posting Engagement in Online Social Networks: Beyond Neighborhood Effects. *Soc. Netw.* **2019**, *59*, 61–76. [CrossRef]

69. Costa, G.; Ortale, R. Mining overlapping communities and inner role assignments through Bayesian mixed-membership models of networks with context-dependent interactions. *ACM Trans. Knowl. Discov. Data (TKDD)* **2018**, *12*, 1–32. [CrossRef]

70. Rossi, R.A.; Gallagher, B.; Neville, J.; Henderson, K. Modeling dynamic behavior in large evolving graphs. In Proceedings of the Sixth ACM International Conference on Web Search And Data Mining, Rome, Italy, 4 February 2013; pp. 667–676.

71. Gupte, P.V.; Ravindran, B. Scalable positional analysis for studying evolution of nodes in networks. *arXiv* **2014**, arXiv:1402.3797.

72. Coles, B.A.; West, M. Trolling the trolls: Online forum users constructions of the nature and properties of trolling. *Comput. Hum. Behav.* **2016**, *60*, 233–244. [CrossRef]

73. Barzilai-Nahon, K. Gatekeepers, virtual communities and the gated: Multidimensional tensions in cyberspace. *Int. J. Commun. Law Policy* **2006**, *11*, 1–28.

74. Akerlof, G.A.; Kranton, R.E. Economics and identity. *Q. J. Econ.* **2000**, *115*, 715–753. [CrossRef]

75. Agarwal, N.; Liu, H.; Tang, L.; Yu, P.S. Identifying the influential bloggers in a community. In Proceedings of the 2008 International Conference on Web Search and Data Mining, Tempe, AZ, USA, 21–25 February 2008; pp. 207–218.

76. Ahmed, N.; Rossi, R.A.; Lee, J.; Willke, T.; Zhou, R.; Kong, X.; Eldardiry, H. Role-based graph embeddings. *IEEE Trans. Knowl. Data Eng.* **2020**, 1. (Just Accepted). [CrossRef]

77. Rossi, R.A.; Ahmed, N.K. Role discovery in networks. *IEEE Trans. Knowl. Data Eng.* **2014**, *27*, 1112–1131. [CrossRef]

78. van der Valk, T.; Chappin, M.M.; Gijsbers, G.W. Evaluating innovation networks in emerging technologies. *Technol. Forecast. Soc. Chang.* **2011**, *78*, 25–39. [CrossRef]

79. Memon, N.; Larsen, H.L.; Hicks, D.L.; Harkiolakis, N. Retracted: Detecting hidden hierarchy in terrorist networks: Some case studies. In *International Conference on Intelligence and Security Informatics, 1–3 July 2018*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 477–489.

80. White, S.; Smyth, P. Algorithms for estimating relative importance in networks. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 24–27 August 2003; pp. 266–275.

81. Borgatti, S.P. Identifying sets of key players in a social network. *Comput. Math. Organ. Theory* **2006**, *12*, 21–34. [CrossRef]

82. Ravid, G.; Rafaeli, S. Asynchronous discussion groups as small world and scale free networks. *Firstmonday* **2004**. (Just Accepted). [CrossRef]

83. Xing, E.P.; Fu, W.; Song, L. A state-space mixed membership blockmodel for dynamic network tomography. *Ann. Appl. Stat.* **2010**, *4*, 535–566. [CrossRef]

84. Welser, H.T.; Gleave, E.; Fisher, D.; Smith, M. Visualizing the signatures of social roles in online discussion groups. *J. Soc. Struct.* **2007**, *8*, 1–32.

85. Sabidussi, G. The centrality index of a graph. *Psychometrika* **1966**, *31*, 581–603. [CrossRef] [PubMed]

86. Freeman, L.C. Centrality in social networks conceptual clarification. *Soc. Netw.* **1978**, *1*, 215–239. [CrossRef]

87. Freeman, L.C. A set of measures of centrality based on betweenness. *Sociometry* **1977**, *40*, 35–41. [CrossRef]

88. Wang, J.; Zheng, V.W.; Liu, Z.; Chang, K.C.C. Topological recurrent neural network for diffusion prediction. In Proceedings of the 2017 IEEE International Conference on Data Mining (ICDM), New Orleans, LA, USA, 18–21 November 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 475–484.

89. Revelle, M.; Domeniconi, C.; Johri, A. Persistent roles in online social networks. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases; Springer: Berlin/Heidelberg, Germany, 2016; pp. 47–62.

90. Borgatti, S.P.; Everett, M.G. Notions of position in social network analysis. *Sociol. Methodol.* **1992**, *22*, 1–35. [CrossRef]

91.   Costa, G.; Ortale, R. Overlapping communities meet roles and respective behavioral patterns in networks with node attributes. In Proceedings of the International Conference on Web Information Systems Engineering, Puschino, Russia, 7–11 October 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 215–230.

92.   Gupte, P.V.; Ravindran, B.; Parthasarathy, S. Role discovery in graphs using global features: Algorithms, applications and a novel evaluation strategy. In Proceedings of the IEEE 33rd International Conference on Data Engineering (ICDE), San Diego, CA, USA, 19–22 April 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 771–782.

93.   Airoldi, E.M.; Blei, D.M.; Fienberg, S.E.; Xing, E.P. Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **2008**, *9*, 1981–2014.

94.   Tang, F.; Zhang, B.; Zheng, J.; Gu, Y. Friend recommendation based on the similarity of micro-blog user model. In Proceedings of the IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, Beijing, China, 20–23 August 2013; pp. 2200–2204.

95.   Rehman, A.U.; Jiang, A.; Rehman, A.; Paul, A.; Sadiq, M.T. Identification and role of opinion leaders in information diffusion for online discussion network. *J. Ambient. Intell. Humaniz. Comput.* **2020**, 1–13. (Just Accepted). [CrossRef]

96.   Cave, E. COVID-19 super-spreaders: Definitional quandaries and implications. *Asian Bioethics Rev.* **2020**, *12*, 235–242. [CrossRef] [PubMed]

97.   Kumar, S.; Cheng, J.; Leskovec, J.; Subrahmanian, V. An army of me: Sockpuppets in online discussion communities. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017; pp. 857–866.

98.   Ghoshal, A.K.; Das, N.; Das, S. Misinformation containment in osns leveraging community structure. In Proceedings of the 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST), Morioka, Japan, 23–25 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.

99.   Grinberg, N.; Joseph, K.; Friedland, L.; Swire-Thompson, B.; Lazer, D. Fake news on Twitter during the 2016 US presidential election. *Science* **2019**, *363*, 374–378. [CrossRef]

100. Chang, S.; Pierson, E.; Koh, P.W.; Gerardin, J.; Redbird, B.; Grusky, D.; Leskovec, J. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **2021**, *589*, 82–87. [CrossRef]

101. Da Silva, L.N.; Malacarne, A.; e Silva, J.W.S.; Kirst, F.V.; De-Bortoli, R. The Scientific Collaboration Networks in University Management in Brazil. *Creat. Educ.* **2018**, *9*, 1469. [CrossRef]

102. Guimera, R.; Danon, L.; Diaz-Guilera, A.; Giralt, F.; Arenas, A. Self-similar community structure in a network of human interactions. *Phys. Rev. E* **2003**, *68*, 065103. [CrossRef] [PubMed]

103. Šubelj, L.; Bajec, M. Robust network community detection using balanced propagation. *Eur. Phys. J. B* **2011**, *81*, 353–362. [CrossRef]

104. Batagelj, V.; Mrvar, A. Pajek-program for large network analysis. *Connections* **1998**, *21*, 47–57.

105. Arazy, O.; Daxenberger, J.; Lifshitz-Assaf, H.; Nov, O.; Gurevych, I. Turbulent stability of emergent roles: The dualistic nature of self-organizing knowledge coproduction. *Inf. Syst. Res.* **2016**, *27*, 792–812. [CrossRef]

106. Yang, D.; Halfaker, A.; Kraut, R.; Hovy, E. Who did what: Editor role identification in Wikipedia. In Proceedings of the International AAAI Conference on Web and Social Media, Cologne, Germany, 17–20 May 2016; Volume 10.

107. McCallum, A.; Wang, X.; Corrada-Emmanuel, A. Topic and role discovery in social networks with experiments on enron and academic email. *J. Artif. Intell. Res.* **2007**, *30*, 249–272. [CrossRef]

108. He, J.L.; Fu, Y.; Chen, D.B. A novel top-k strategy for influence maximization in complex networks with community structure. *PLoS ONE* **2015**, *10*, e0145283.

109. Yang, G.; Benko, T.P.; Cavaliere, M.; Huang, J.; Perc, M. Identification of influential invaders in evolutionary populations. *Sci. Rep.* **2019**, *9*, 1–12. [CrossRef]

110. Wang, M.; Li, W.; Guo, Y.; Peng, X.; Li, Y. Identifying influential spreaders in complex networks based on improved k-shell method. *Phys. Stat. Mech. Its Appl.* **2020**, *554*, 124229. [CrossRef]

111. Guo, C.; Yang, L.; Chen, X.; Chen, D.; Gao, H.; Ma, J. Influential nodes identification in complex networks via information entropy. *Entropy* **2020**, *22*, 242. [CrossRef]

112. Carmi, S.; Havlin, S.; Kirkpatrick, S.; Shavitt, Y.; Shir, E. A model of Internet topology using k-shell decomposition. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11150–11154. [CrossRef]

113. Kitsak, M.; Gallos, L.K.; Havlin, S.; Liljeros, F.; Muchnik, L.; Stanley, H.E.; Makse, H.A. Identification of influential spreaders in complex networks. *Nat. Phys.* **2010**, *6*, 888–893. [CrossRef]

114. Serrano, M.Á.; Boguna, M. Clustering in complex networks. I. General formalism. *Phys. Rev. E* **2006**, *74*, 056114. [CrossRef] [PubMed]

115. Rosvall, M.; Bergstrom, C.T. An information-theoretic framework for resolving community structure in complex networks. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7327–7331. [CrossRef]

116. Cheng, J.; Chen, M.; Zhou, M.; Gao, S.; Liu, C.; Liu, C. Overlapping community change-point detection in an evolving network. *IEEE Trans. Big Data* **2018**, *6*, 189–200. [CrossRef]

117. Jin, D.; Yu, Z.; Jiao, P.; Pan, S.; Yu, P.S.; Zhang, W. A survey of community detection approaches: From statistical modeling to deep learning. *arXiv* **2021**, arXiv:2101.01669.

118. Fortunato, S. Community detection in graphs. *Phys. Rep.* **2010**, *486*, 75–174. [CrossRef]

119. Liu, G.; Wang, Y.; Orgun, M.A. Social context-aware trust network discovery in complex contextual social networks. In Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, ON, Canada, 22–26 July 2012.
120. Gilpin, S.; Eliassi-Rad, T.; Davidson, I. Guided learning for role discovery (GLRD) framework, algorithms, and applications. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; pp. 113–121.
121. Goyal, P.; Ferrara, E. Graph embedding techniques, applications, and performance: A survey. *Knowl.-Based Syst.* **2018**, *151*, 78–94. [CrossRef]
122. Zhao, Y.; Wang, G.; Yu, P.S.; Liu, S.; Zhang, S. Inferring social roles and statuses in social networks. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; pp. 695–703.
123. Faust, K.; Wasserman, S. Blockmodels: Interpretation and evaluation. *Soc. Netw.* **1992**, *14*, 5–61. [CrossRef]
124. Fu, W.; Song, L.; Xing, E.P. Dynamic mixed membership blockmodel for evolving networks. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; pp. 329–336.
125. Tu, K.; Cui, P.; Wang, X.; Yu, P.S.; Zhu, W. Deep recursive network embedding with regular equivalence. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 2357–2366.
126. Qiu, J.; Dong, Y.; Ma, H.; Li, J.; Wang, C.; Wang, K.; Tang, J. Netsmf: Large-scale network embedding as sparse matrix factorization. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 1509–1520.
127. Ma, X.; Qin, G.; Qiu, Z.; Zheng, M.; Wang, Z. RiWalk: Fast structural node embedding via role identification. In Proceedings of the 2019 IEEE International Conference on Data Mining (ICDM), Beijing, China, 8–11 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 478–487.
128. Cui, P.; Wang, X.; Pei, J.; Zhu, W. A survey on network embedding. *IEEE Trans. Knowl. Data Eng.* **2018**, *31*, 833–852. [CrossRef]
129. Pei, Y.; Fletcher, G.; Pechenizkiy, M. Joint role and community detection in networks via l 2, 1 norm regularized nonnegative matrix tri-factorization. In Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Vancouver, BC, Canada, 27–30 August 2019; pp. 168–175.
130. Henderson, K.; Gallagher, B.; Eliassi-Rad, T.; Tong, H.; Basu, S.; Akoglu, L.; Koutra, D.; Faloutsos, C.; Li, L. Rolx: Structural role extraction & mining in large graphs. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; pp. 1231–1239.
131. Henderson, K.; Gallagher, B.; Li, L.; Akoglu, L.; Eliassi-Rad, T.; Tong, H.; Faloutsos, C. It's who you know: Graph mining using recursive structural features. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 663–671.
132. Donnat, C.; Zitnik, M.; Hallac, D.; Leskovec, J. Learning structural node embeddings via diffusion wavelets. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 1320–1329.
133. Rossi, R.A.; Ahmed, N.K.; Koh, E.; Kim, S.; Rao, A.; Abbasi-Yadkori, Y. A structural graph representation learning framework. In Proceedings of the 13th International Conference on Web Search and Data Mining, Houston, TX, USA, 3–7 February 2020; pp. 483–491.
134. Ribeiro, L.F.; Saverese, P.H.; Figueiredo, D.R. struc2vec: Learning node representations from structural identity. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 385–394.
135. Pei, Y.; Du, X.; Zhang, J.; Fletcher, G.; Pechenizkiy, M. struc2gauss: Structural role preserving network embedding via Gaussian embedding. *Data Min. Knowl. Discov.* **2020**, *34*, 1072–1103. [CrossRef]
136. Jin, R.; Lee, V.E.; Hong, H. Axiomatic ranking of network role similarity. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 922–930.
137. Guo, J.; Xu, L.; Liu, J. Spine: Structural identity preserved inductive network embedding. *arXiv* **2018**, arXiv:1802.03984.
138. Jiao, P.; Guo, X.; Pan, T.; Zhang, W.; Pei, Y. A Survey on Role-Oriented Network Embedding. *arXiv* **2021**, arXiv:2107.08379.
139. Jin, Y.; Song, G.; Shi, C. GraLSP: Graph neural networks with local structural patterns. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 4361–4368.
140. Zhang, W.; Guo, X.; Wang, W.; Tian, Q.; Pan, L.; Jiao, P. Role-based network embedding via structural features reconstruction with degree-regularized constraint. *Knowl.-Based Syst.* **2021**, *218*, 106872. [CrossRef]
141. Guo, X.; Zhang, W.; Wang, W.; Yu, Y.; Wang, Y.; Jiao, P. Role-Oriented Graph Auto-encoder Guided by Structural Information. In Proceedings of the International Conference on Database Systems for Advanced Applications, Jeju, Korea, 13 December 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 466–481.
142. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.
143. Li, B.; Pi, D. Learning deep neural networks for node classification. *Expert Syst. Appl.* **2019**, *137*, 324–334. [CrossRef]
144. Hamilton, W.L.; Ying, R.; Leskovec, J. Inductive representation learning on large graphs. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 1025–1035.
145. Grover, A.; Leskovec, J. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 855–864.

146. Keikha, M.M.; Rahgozar, M.; Asadpour, M. DeepLink: A novel link prediction framework based on deep learning. *J. Inf. Sci.*
2019, *47*, 0165551519891345. [CrossRef]
147. Agibetov, A. Graph embeddings via matrix factorization for link prediction: Smoothing or truncating negatives? *arXiv* **2020**,
arXiv:2011.09907.
148. Rossi, R.; Gallagher, B.; Neville, J.; Henderson, K. *Modeling Temporal Behavior in Large Networks: A Dynamic Mixed-Membership
Model*; Technical Report; Lawrence Livermore National Lab. (LLNL): Livermore, CA, USA, 2011.
149. Yang, D. Computational Social Roles. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 2019.
150. Kittur, A.; Kraut, R.E. Harnessing the wisdom of crowds in wikipedia: Quality through coordination. In Proceedings of the 2008
ACM Conference on Computer Supported Cooperative Work, San Diego, CA, USA, 8–12 November 2008; pp. 37–46.
151. Bamman, D.; O'Connor, B.; Smith, N.A. Learning latent personas of film characters. In Proceedings of the 51st Annual Meeting
of the Association for Computational Linguistics (Volume 1: Long Papers), Sofia, Bulgaria, 4–9 August 2013; pp. 352–361.
152. Welser, H.T.; Cosley, D.; Kossinets, G.; Lin, A.; Dokshin, F.; Gay, G.; Smith, M. Finding social roles in Wikipedia. In Proceedings of
the 2011 Conference, Seattle, WA, USA, 8–11 February 2011; pp. 122–129.
153. Fazeen, M.; Dantu, R.; Guturu, P. Identification of leaders, lurkers, associates and spammers in a social network: Context-
dependent and context-independent approaches. *Soc. Netw. Anal. Min.* **2011**, *1*, 241–254. [CrossRef]
154. Ferschke, O.; Yang, D.; Rosé, C.P. A lightly supervised approach to role identification in wikipedia talk page discussions. In
Proceedings of the Ninth international AAAI Conference on Web and Social Media, Oxford, UK, 26–29 May 2015.
155. Maki, K.; Yoder, M.; Jo, Y.; Rosé, C. Roles and success in wikipedia talk pages: Identifying latent patterns of behavior. In
Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Taipei,
Taiwan, 27 November–1 December 2017; pp. 1026–1035.
156. Lin, S.; Hong, W.; Wang, D.; Li, T. A survey on expert finding techniques. *J. Intell. Inf. Syst.* **2017**, *49*, 255–279. [CrossRef]
157. Zhang, J.; Tang, J.; Li, J. Expert finding in a social network. In Proceedings of the International Conference on Database Systems
For Advanced Applications; Springer: Berlin/Heidelberg, Germany, 2007; pp. 1066–1069.
158. Chen, F.; Gao, X.; Peng, Y.; He, Y.; Xue, C. Identifying Weibo Opinion Leaders with Text Sentiment Analysis. *Data Anal. Knowl.
Discov.* **2019**, *3*, 120–128.
159. Wadden, D.; August, T.; Li, Q.; Althoff, T. The Effect of Moderation on Online Mental Health Conversations. *arXiv* **2020**,
arXiv:2005.09225.
160. Ju, W.; Chen, L.; Li, B.; Liu, W.; Sheng, J.; Wang, Y. A new algorithm for positive influence maximization in signed networks. *Inf.
Sci.* **2020**, *512*, 1571–1591. [CrossRef]
161. Li, Y.; Fan, J.; Wang, Y.; Tan, K.L. Influence maximization on social graphs: A survey. *IEEE Trans. Knowl. Data Eng.* **2018**,
30, 1852–1872. [CrossRef]
162. Qing, Y.; Peng, Z. A review of the influence maximization problem in social networks. *Comput. Eng. Sci.* **2015**, 2. (Just Accepted).
163. Kempe, D.; Kleinberg, J.; Tardos, É. Maximizing the spread of influence through a social network. In Proceedings of the Ninth
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 24–27 August 2003;
pp. 137–146.
164. Goel, S.; Anderson, A.; Hofman, J.; Watts, D.J. The Structural Virality of Online Diffusion. *Manag. Sci.* **2015**, *62*, 180–196.
[CrossRef]
165. Bartal, A. Modeling Influence on Posting Engagement: The Gaza Great Return March Analyzed on Twitter. In Proceedings of
the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Barcelona, Spain, 28–31
August 2018; pp. 577–578.
166. Min, B.; San Miguel, M. Competing Contagion Processes: Complex Contagion Triggered by Simple Contagion. *Sci. Rep.* **2018**,
8, 10422. [CrossRef]
167. Mønsted, B.; Sapieżyński, P.; Ferrara, E.; Lehmann, S. Evidence of Complex Contagion of Information in Social Media: An
Experiment Using Twitter Bots. *PLoS ONE* **2017**, *12*, e0184148. [CrossRef] [PubMed]
168. Altinel, A.B.; Hakkoz, M.A.; Bozdag, E.B.; Ganiz, M.C. Identifying Topic-based Opinion Leaders in Social Networks by Content
and User Information. *Int. J. Intell. Syst. Appl. Eng.* **2020**, *8*, 214–220. [CrossRef]
169. Aleahmad, A.; Karisani, P.; Rahgozar, M.; Oroumchian, F. OLFinder: Finding opinion leaders in online social networks. *J. Inf. Sci.*
2016, *42*, 659–674. [CrossRef]
170. Cherepnalkoski, D.; Karpf, A.; Mozetič, I.; Grčar, M. Cohesion and coalition formation in the European Parliament: Roll-call
votes and Twitter activities. *PLoS ONE* **2016**, *11*, e0166586. [CrossRef]
171. Adalat, M.; Niazi, M.A.; Vasilakos, A.V. Variations in power of opinion leaders in online communication networks. *R. Soc. Open
Sci.* **2018**, *5*, 180642. [CrossRef]
172. Orr, G. Diffusion of innovations, by Everett Rogers (1995). *Retrieved J.* **2003**, *21*, 2005.
173. Trusov, M.; Bodapati, A.V.; Bucklin, R.E. Determining influential users in internet social networks. *J. Mark. Res.* **2010**, *47*, 643–658.
[CrossRef]
174. Wasserman, S.; Faust, K. *Social Network Analysis: Methods and Applications*; Cambridge University Press: Cambridge, UK, 1994.
175. Newman, M.E.J. The Structure and Function of Complex Networks. *SIAM Rev.* **2003**, *45*, 167–256. [CrossRef]

176. Kleinberg, J.M.; Kumar, R.; Raghavan, P.; Rajagopalan, S.; Tomkins, A.S. The web as a graph: Measurements, models, and methods. In Proceedings of the International Computing and Combinatorics Conference; Springer: Berlin/Heidelberg, Germany, 1999; pp. 1–17.
177. Lü, L.; Zhou, T.; Zhang, Q.M.; Stanley, H.E. The H-index of a network node and its relation to degree and coreness. *Nat. Commun.* **2016**, *7*, 1–7. [CrossRef]
178. Zareie, A.; Sheikhahmadi, A. EHC: Extended H-index centrality measure for identification of users' spreading influence in complex networks. *Phys. A Stat. Mech. Its Appl.* **2019**, *514*, 141–155. [CrossRef]
179. Gao, L.; Yu, S.; Li, M.; Shen, Z.; Gao, Z. Weighted h-index for Identifying Influential Spreaders. *Symmetry* **2019**, *11*, 1263. [CrossRef]
180. Ahajjam, S.; Badir, H. Identification of influential spreaders in complex networks using HybridRank algorithm. *Sci. Rep.* **2018**, *8*, 1–10. [CrossRef] [PubMed]
181. Bhat, N.; Aggarwal, N.; Kumar, S. Identification of Influential Spreaders in Social Networks using Improved Hybrid Rank Method. *Procedia Comput. Sci.* **2020**, *171*, 662–671. [CrossRef]
182. Al-garadi, M.A.; Varathan, K.D.; Ravana, S.D. Identification of influential spreaders in online social networks using interaction weighted K-core decomposition method. *Phys. Stat. Mech. Its Appl.* **2017**, *468*, 278–288. [CrossRef]
183. Lü, L.; Zhang, Y.C.; Yeung, C.H.; Zhou, T. Leaders in social networks, the delicious case. *PLoS ONE* **2011**, *6*, e21202. [CrossRef] [PubMed]
184. Li, Q.; Zhou, T.; Lü, L.; Chen, D. Identifying influential spreaders by weighted LeaderRank. *Phys. Stat. Mech. Its Appl.* **2014**, *404*, 47–55. [CrossRef]
185. Page, L.; Brin, S.; Motwani, R.; Winograd, T. *The PageRank Citation Ranking: Bringing Order to the Web*; Technical Report; Stanford InfoLab: Stanford, CA, USA, 1999.
186. Lü, L.; Chen, D.; Ren, X.L.; Zhang, Q.M.; Zhang, Y.C.; Zhou, T. Vital nodes identification in complex networks. *Phys. Rep.* **2016**, *650*, 1–63. [CrossRef]
187. Pastor-Satorras, R.; Castellano, C.; Van Mieghem, P.; Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **2015**, *87*, 925. [CrossRef]
188. Lloyd, A.L.; May, R.M. How viruses spread among computers and people. *Science* **2001**, *292*, 1316–1317. [CrossRef]
189. He, S.; Peng, Y.; Sun, K. SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dyn.* **2020**, *101*, 1667–1680. [CrossRef]
190. Kleinberg, J. Cascading behavior in networks: Algorithmic and economic issues. *Algorithmic Game Theory* **2007**, *24*, 613–632.
191. Daugherty, T.; Eastin, M.S.; Bright, L. Exploring consumer motivations for creating user-generated content. *J. Interact. Advert.* **2008**, *8*, 16–25. [CrossRef]
192. Saito, K.; Nakano, R.; Kimura, M. Prediction of information diffusion probabilities for independent cascade model. In Proceedings of the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems; Springer: Berlin/Heidelberg, Germany, 2008; pp. 67–75.
193. Kamiński, M.; Szymańska, C.; Nowak, J.K. Whose tweets on COVID-19 gain the Most attention: Celebrities, political, or scientific authorities? *Cyberpsychol. Behav. Soc. Netw.* **2021**, *24*, 123–128. [CrossRef] [PubMed]
194. Chen, H. Relationship between Motivation and Behavior of SNS User. *J. Softw.* **2012**, *7*, 1265–1272. [CrossRef]
195. Easley, D.; Kleinberg, J. Information cascades. In *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*; Cambridge University Press: Cambridge, UK, 2010.
196. Hodas, N.O.; Lerman, K. The simple rules of social contagion. *Sci. Rep.* **2014**, *4*, 1–7. [CrossRef] [PubMed]
197. Valente, T.W. *Network Models of the Diffusion of Innovations*; Hampton Press: New York, NY, USA, 1995; Volume 152, p. 157.
198. Valente, T.W.; Davis, R.L. Accelerating the Diffusion of Innovations Using Opinion Leaders. *Ann. Am. Acad. Political Soc. Sci.* **1999**, *566*, 55–67. [CrossRef]
199. Coleman, J.S.; Katz, E.; Menzel, H. *Medical Innovation: A Diffusion Study*; Bobbs-Merrill Company: Indianapolis, IN, USA, 1966.
200. Rogers, E.M.; Kincaid, D.L. *Communication Networks: Toward a New Paradigm for Research*; Free Press: New York, NY, USA, 1981.
201. Kramer, A.D.; Guillory, J.E.; Hancock, J.T. Experimental Evidence of Massive-scale Emotional Contagion Through Social Networks. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 8788–8790. [CrossRef] [PubMed]
202. Sun, E.; Rosenn, I.; Marlow, C.A.; Lento, T.M. Gesundheit! Modeling Contagion Through Facebook News Feed. In Proceedings of the Third International AAAI Conference on Weblogs and Social Media, San Jose, CA, USA, 17–20 May 2009; Volume 3, pp. 146–153.
203. Katz, E. The Two-step Flow of Communication: An Up-to-date Report on an Hypothesis. *Public Opin. Q.* **1957**, *21*, 61–78. [CrossRef]
204. Leskovec, J.; Adamic, L.A.; Huberman, B.A. The Dynamics of Viral Marketing. *ACM Trans. Web (TWEB)* **2007**, *1*, 5. [CrossRef]
205. Rogers, E.M. *Diffusion of Innovations*; Simon and Schuster: New York, NY, USA, 2010.
206. Kreindler, G.E.; Young, H.P. Rapid Innovation Diffusion in Social Networks. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 10881–10888. [CrossRef] [PubMed]
207. Centola, D. The Spread of Behavior in an Online Social Network Experiment. *Science* **2010**, *329*, 1194–1197. [CrossRef]
208. Granovetter, M. Threshold Models of Collective Behavior. *Am. J. Sociol.* **1978**, *83*, 1420–1443. [CrossRef]
209. Karimi, F.; Holme, P. Threshold Model of Cascades in Empirical Temporal Networks. *Phys. Stat. Mech. Its Appl.* **2013**, *392*, 3476–3483. [CrossRef]

210. Burt, R.S. Social Contagion and Innovation: Cohesion Versus Structural Equivalence. *Am. J. Sociol.* **1987**, *92*, 1287–1335. [CrossRef]

211. Leenders, R.T.A. Modeling Social Influence Through Network Autocorrelation: Constructing the Weight Matrix. *Soc. Netw.* **2002**, *24*, 21–47. [CrossRef]

212. Singh, S.S.; Singh, K.; Kumar, A.; Shakya, H.K.; Biswas, B. A survey on information diffusion models in social networks. In Proceedings of the International Conference on Advanced Informatics for Computing Research, Gurugram, India, 26–27 December 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 426–439.

213. Yuan, C.; Li, J.; Zhou, W.; Lu, Y.; Zhang, X.; Hu, S. DyHGCN: A Dynamic Heterogeneous Graph Convolutional Network to Learn Users' Dynamic Preferences for Information Diffusion Prediction. *arXiv* **2020**, arXiv:2006.05169.

214. Wang, Y.; Shen, H.; Liu, S.; Gao, J.; Cheng, X. Cascade Dynamics Modeling with Attention-based Recurrent Neural Network. IJCAI **2017**, 2985–2991. (Just Accepted).

215. Yang, C.; Wang, H.; Tang, J.; Shi, C.; Sun, M.; Cui, G.; Liu, Z. Full-Scale Information Diffusion Prediction With Reinforced Recurrent Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, 1–13. (Just Accepted). [CrossRef]

216. Islam, M.R.; Muthiah, S.; Adhikari, B.; Prakash, B.A.; Ramakrishnan, N. Deepdiffuse: Predicting the 'who' and 'when' in cascades. In Proceedings of the IEEE International Conference on Data Mining (ICDM), Sorrento, Italy, 17–20 November 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1055–1060.

217. Cao, Z.; Han, K.; Zhu, J. Information Diffusion Prediction via Dynamic Graph Neural Networks. In Proceedings of the IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Dalian, China, 5–7 May 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1099–1104.

218. Chang, H. A New Perspective on Twitter Hashtag Use: Diffusion of Innovation Theory. *Proc. Assoc. Inf. Sci. Technol.* **2010**, *47*, 1–4. [CrossRef]

219. Wakamiya, S.; Kawai, Y.; Aramaki, E. Twitter-based Influenza Detection After Flu Peak Via Tweets With Indirect Information: Text Mining Study. *JMIR Public Health Surveill.* **2018**, *4*, e65. [CrossRef] [PubMed]

220. Nicolas, K.; Anton, A. Using Deep Learning at Scale in Twitter's Timelines. 2017. Available online: https://blog.Twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-Twitters-timelines.html (accessed on 17 May 2019).

221. Richterich, A. 'Karma, Precious Karma!'Karmawhoring on Reddit and the Front Page's Econometrisation. *J. Peer Prod.* **2014**, 4, 1–12.

222. Çelikkanat, A.; Malliaros, F.D. Topic-aware latent models for representation learning on networks. *Pattern Recognit. Lett.* **2021**, *144*, 89–96. [CrossRef]

223. Barbieri, N.; Bonchi, F.; Manco, G. Topic-aware social influence propagation models. *Knowl. Inf. Syst.* **2013**, *37*, 555–584. [CrossRef]

224. Bailey, N.T. *The Mathematical Theory of Infectious Diseases and Its Applications*; Charles Griffin and Company Ltd.: Glasgow, UK, 1975.

225. Alanazi, S.A.; Kamruzzaman, M.; Alruwaili, M.; Alshammari, N.; Alqahtani, S.A.; Karime, A. Measuring and Preventing COVID-19 Using the SIR Model and Machine Learning in Smart Health Care. *J. Healthc. Eng.* **2020**, *2020*. [CrossRef]

226. Goffman, W.; Newill, V. Generalization of Epidemic Theory. *Nature* **1964**, *204*, 225–228. [CrossRef]

227. Bass, F.M. A New Product Growth for Model Consumer Durables. *Manag. Sci.* **1969**, *15*, 215–227. [CrossRef]

228. Ma, H.; Yang, H.; Lyu, M.R.; King, I. Mining Social Networks Using Heat Diffusion Processes for Marketing Candidates Selection. In Proceedings of the 17th ACM Conference on Information and Knowledge Management, Napa Valley, CA, USA, 26–30 October 2008; pp. 233–242.

229. Rezvanian, A.; Meybodi, M.R. Stochastic graph as a model for social networks. *Comput. Hum. Behav.* **2016**, *64*, 621–640. [CrossRef]

230. Rossetti, G.; Pappalardo, L.; Pedreschi, D.; Giannotti, F. Tiles: An online algorithm for community discovery in dynamic social networks. *Mach. Learn.* **2017**, *106*, 1213–1241. [CrossRef]

231. Amati, V.; Lomi, A.; Mira, A. Social network modeling. *Annu. Rev. Stat. Its Appl.* **2018**, *5*, 343–369. [CrossRef]

232. Snijders, T.A. Stochastic actor-oriented models for network dynamics. *Annu. Rev. Stat. Its Appl.* **2017**, *4*, 343–363. [CrossRef]

233. Snijders, T.A. Statistical models for social networks. *Annu. Rev. Sociol.* **2011**, *37*, 131–153. [CrossRef]

234. Wasserman, S. Analyzing social networks as stochastic processes. *J. Am. Stat. Assoc.* **1980**, *75*, 280–294. [CrossRef]

235. Xiong, F.; Liu, Y.; Zhang, Z.j.; Zhu, J.; Zhang, Y. An information diffusion model based on retweeting mechanism for online social media. *Phys. Lett. A* **2012**, *376*, 2103–2108. [CrossRef]

236. Tang, J.; Sun, J.; Wang, C.; Yang, Z. Social influence analysis in large-scale networks. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France, 28 June–1 July 2009; pp. 807–816.

237. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.

238. Cha, M.; Haddadi, H.; Benevenuto, F.; Gummadi, P.K. Measuring User Influence in Twitter: The Million Follower Fallacy. *Icwsm* **2010**, *10*, 30.

239. Leskovec, J.; Backstrom, L.; Kleinberg, J. Meme-tracking and the Dynamics of the News Cycle. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 28 June–1 July 2009; ACM: New York, NY, USA, 2009; pp. 497–506.

240. Xu, H.; Wei, J.; Yang, Z.; Ruan, J.; Wang, J. Probabilistic topic and role model for information diffusion in social network. In Pacific-Asia Conference on Knowledge Discovery and Data Mining; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–15.

241. Bakshy, E.; Karrer, B.; Adamic, L.A. Social Influence and the Diffusion of User-created Content. In Proceedings of the 10th ACM Conference on Electronic Commerce, Stanford, CA, USA, 6–10 July 2009; ACM: New York, NY, USA, 2009; pp. 325–334.
242. Wang, W.; Tang, M.; Shu, P.; Wang, Z. Dynamics of Social Contagions With Heterogeneous Adoption Thresholds: Crossover Phenomena in Phase Transition. *New J. Phys.* **2016**, *18*, 013029. [CrossRef]
243. Kurka, D.B.; Godoy, A.; Von Zuben, F.J. Online Social Network Analysis: A Survey of Research Applications in Computer Science. *arXiv* **2015**, arXiv:1504.05655.
244. Eyal, K.; Rubin, A.M. Viewer aggression and homophily, identification, and parasocial relationships with television characters. *J. Broadcast. Electron. Media* **2003**, *47*, 77–98. [CrossRef]
245. McCroskey, J.C.; Richmond, V.P.; Daly, J.A. The development of a measure of perceived homophily in interpersonal communication. *Hum. Commun. Res.* **1975**, *1*, 323–332. [CrossRef]
246. Zhang, K.Z.; Barnes, S.J.; Zhao, S.J.; Zhang, H. Can consumers be persuaded on brand microblogs? An empirical study. *Inf. Manag.* **2018**, *55*, 1–15. [CrossRef]
247. Hanks, L.; Line, N.; Yang, W. Status seeking and perceived similarity: A consideration of homophily in the social servicescape. *Int. J. Hosp. Manag.* **2017**, *60*, 123–132. [CrossRef]
248. Wang, Z.; Walther, J.B.; Pingree, S.; Hawkins, R.P. Health information, credibility, homophily, and influence via the Internet: Web sites versus discussion groups. *Health Commun.* **2008**, *23*, 358–368. [CrossRef] [PubMed]
249. Ismagilova, E.; Slade, E.; Rana, N.P.; Dwivedi, Y.K. The effect of characteristics of source credibility on consumer behaviour: A meta-analysis. *J. Retail. Consum. Serv.* **2020**, *53*, 101736. [CrossRef]
250. Shang, S.S.; Wu, Y.L.; Sie, Y.J. Generating consumer resonance for purchase intention on social network sites. *Comput. Hum. Behav.* **2017**, *69*, 18–28. [CrossRef]
251. Lee, J.E.; Watkins, B. YouTube vloggers' influence on consumer luxury brand perceptions and intentions. *J. Bus. Res.* **2016**, *69*, 5753–5760. [CrossRef]
252. Li, F.; Du, T.C. Who is talking? An ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs. *Decis. Support Syst.* **2011**, *51*, 190–197. [CrossRef]
253. Ladhari, R.; Massa, E.; Skandrani, H. YouTube vloggers' popularity and influence: The roles of homophily, emotional attachment, and expertise. *J. Retail. Consum. Serv.* **2020**, *54*, 102027. [CrossRef]
254. Kim, S.; Kandampully, J.; Bilgihan, A. The influence of eWOM communications: An application of online social network framework. *Comput. Hum. Behav.* **2018**, *80*, 243–254. [CrossRef]
255. Sakib, M.N.; Zolfagharian, M.; Yazdanparast, A. Does parasocial interaction with weight loss vloggers affect compliance? The role of vlogger characteristics, consumer readiness, and health consciousness. *J. Retail. Consum. Serv.* **2020**, *52*, 101733. [CrossRef]
256. Bang, H.P. 'Yes We Can': Identity Politics and Project Politics for a Late-modern World. *Urban Res. Pract.* **2009**, *2*, 117–137. [CrossRef]
257. Crane, R.; Sornette, D. Robust Dynamic Classes Revealed by Measuring the Response Function of a Social System. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 15649–15653. [CrossRef]
258. Gleeson, J.P.; Durrett, R. Temporal Profiles of Avalanches on Networks. *Nat. Commun.* **2017**, *8*, 1227. [CrossRef]
259. Chakravarti, I.M.; Laha, R.G.; Roy, J. Handbook of methods of applied statistics. In *Wiley Series in Probability and Mathematical Statistics (USA) eng*; Wiley: New York, NY, USA, 1967.
260. Nguyen, T.T.; Hui, P.M.; Harper, F.M.; Terveen, L.; Konstan, J.A. Exploring the filter bubble: The effect of using recommender systems on content diversity. In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Korea, 7–11 April 2014; pp. 677–686.
261. Spohr, D. Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Bus. Inf. Rev.* **2017**, *34*, 150–160. [CrossRef]
262. Zanardi, V.; Capra, L. Social ranking: Uncovering relevant content using tag-based recommender systems. In Proceedings of the 2008 ACM Conference on Recommender Systems, Lausanne, Switzerland, 1 January 2008; pp. 51–58.
263. Perry, B.L.; Pescosolido, B.A.; Borgatti, S.P. *Egocentric Network Analysis: Foundations, Methods, and Models*; Cambridge University Press: Cambridge, UK, 2018; Volume 44.
264. Xiong, R.; Donath, J. PeopleGarden: Creating data portraits for users. In Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology, Asheville, NC, USA, 7–10 November 1999; pp. 37–44.
265. Chen, S.; Chen, S.; Wang, Z.; Liang, J.; Wu, Y.; Yuan, X. D-map+ interactive visual analysis and exploration of ego-centric and event-centric information diffusion patterns in social media. *ACM Trans. Intell. Syst. Technol. (TIST)* **2018**, *10*, 1–26.
266. Cao, N.; Lin, Y.R.; Du, F.; Wang, D. Episogram: Visual summarization of egocentric social interactions. *IEEE Comput. Graph. Appl.* **2015**, *36*, 72–81. [CrossRef]
267. Nguyen, P.H.; Henkin, R.; Chen, S.; Andrienko, N.; Andrienko, G.; Thonnard, O.; Turkay, C. Vasabi: Hierarchical user profiles for interactive visual user behaviour analytics. *IEEE Trans. Vis. Comput. Graph.* **2019**, *26*, 77–86. [CrossRef] [PubMed]
268. Huang, Y.; Shi, L.; Su, Y.; Hu, Y.; Tong, H.; Wang, C.; Yang, T.; Wang, D.; Liang, S. Eiffel: Evolutionary flow map for influence graph visualization. *IEEE Trans. Vis. Comput. Graph.* **2019**, *26*, 2944–2960. [CrossRef]
269. Cao, N.; Shi, C.; Lin, S.; Lu, J.; Lin, Y.R.; Lin, C.Y. TargetVue: Visual analysis of anomalous user behaviors in online communication systems. *IEEE Trans. Vis. Comput. Graph.* **2015**, *22*, 280–289. [CrossRef]

270. Han, M.; Wang, Q.; Wei, L.; Zhang, Y.; Cao, Y.; Pu, J. egoStellar: Visual Analysis of Anomalous Communication Behaviors from Egocentric Perspective. In International Computer Symposium; Springer: Berlin/Heidelberg, Germany, 2018; pp. 280–290.

271. Pu, J.; Zhang, J.; Shao, H.; Zhang, T.; Rao, Y. egoDetect: Visual Detection and Exploration of Anomaly in Social Communication Network. *Sensors* **2020**, *20*, 5895. [CrossRef]

272. Kim, Y.; Bae, Y.; Kim, J.; Nam, Y. iVIS: Interpretable Interactive Visualization for User Behavior Clusters. In International Conference on Human-Computer Interaction; Springer: Berlin/Heidelberg, Germany, 2020; pp. 47–52.

273. Wong, J.S. MessageLens: A visual analytics system to support multifaceted exploration of MOOC forum discussions. *Vis. Inform.* **2018**, *2*, 37–49. [CrossRef]

274. Fu, S.; Wang, Y.; Yang, Y.; Bi, Q.; Guo, F.; Qu, H. VisForum: A visual analysis system for exploring user groups in online forums. *ACM Trans. Interact. Intell. Syst. (TiiS)* **2018**, *8*, 1–21. [CrossRef]

275. Fu, S.; Zhao, J.; Cui, W.; Qu, H. Visual analysis of MOOC forums with iForum. *IEEE Trans. Vis. Comput. Graph.* **2016**, *23*, 201–210. [CrossRef]

276. Wu, M.; Dewan, M.A.A.; Lin, F.; Murshed, M. Visualization of course discussion forums: A short review from online learning perspective. In Proceedings of the IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), Edmonton, AB, Canada, 5–8 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–4.

277. Laghari, A.A.; Wu, K.; Laghari, R.A.; Ali, M.; Khan, A.A. A Review and State of Art of Internet of Things (IoT). *Arch. Comput. Methods Eng.* **2021**, 1–19. (Just Accepted). [CrossRef]

278. Conti, M.; Passarella, A.; Das, S.K. The Internet of People (IoP): A new wave in pervasive mobile computing. *Pervasive Mob. Comput.* **2017**, *41*, 1–27. [CrossRef]

279. Conti, M.; Passarella, A. The Internet of People: A human and data-centric paradigm for the Next Generation Internet. *Comput. Commun.* **2018**, *131*, 51–65. [CrossRef]

280. Zhou, W.X.; Sornette, D.; Hill, R.A.; Dunbar, R.I. Discrete hierarchical organization of social group sizes. *Proc. R. Soc. Biol. Sci.* **2005**, *272*, 439–444. [CrossRef] [PubMed]

281. Hill, R.A.; Dunbar, R.I. Social network size in humans. *Hum. Nat.* **2003**, *14*, 53–72. [CrossRef] [PubMed]

282. Dunbar, R.I.; Arnaboldi, V.; Conti, M.; Passarella, A. The structure of online social networks mirrors those in the offline world. *Soc. Netw.* **2015**, *43*, 39–47. [CrossRef]

283. Arnaboldi, V.; La Gala, M.; Passarella, A.; Conti, M. Information diffusion in distributed OSN: The impact of trusted relationships. *Peer -Peer Netw. Appl.* **2016**, *9*, 1195–1208. [CrossRef]

284. Liberatore, M.J.; Wagner, W.P. Virtual, mixed, and augmented reality: A systematic review for immersive systems research. *Virtual Real.* **2021**, *25*, 1–27. [CrossRef]

285. Sommer, B.; Baaden, M.; Krone, M.; Woods, A. From virtual reality to immersive analytics in bioinformatics. *J. Integr. Bioinform.* **2018**, *15*. (Just Accepted). [CrossRef]

286. Fonnet, A.; Prie, Y. Survey of immersive analytics. *IEEE Trans. Vis. Comput. Graph.* **2019**, *3*, 2101–2122. [CrossRef]

287. Sorger, J.; Arleo, A.; Kán, P.; Knecht, W.; Waldner, M. Egocentric Network Exploration for Immersive Analytics. *arXiv* **2021**, arXiv:2109.09547.

288. Wagner Filho, J.A.; Rey, M.F.; Freitas, C.M.; Nedel, L. Immersive visualization of abstract information: An evaluation on dimensionally-reduced data scatterplots. In Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Lisbon, Portugal, 27 March–1 April 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 483–490.

289. Büschel, W.; Vogt, S.; Dachselt, R. Augmented reality graph visualizations. *IEEE Comput. Graph. Appl.* **2019**, *39*, 29–40. [CrossRef]

290. Belcher, D.; Billinghurst, M.; Hayes, S.; Stiles, R. Using augmented reality for visualizing complex graphs in three dimensions. In Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality, Tokyo, Japan, 10 October 2003; IEEE: Piscataway, NJ, USA, 2003; pp. 84–93.

291. Kraus, M.; Weiler, N.; Oelke, D.; Kehrer, J.; Keim, D.A.; Fuchs, J. The impact of immersion on cluster identification tasks. *IEEE Trans. Vis. Comput. Graph.* **2019**, *26*, 525–535. [CrossRef]

292. Greffard, N.; Picarougne, F.; Kuntz, P. Visual community detection: An evaluation of 2d, 3d perspective and 3d stereoscopic displays. In International Symposium on Graph Drawing; Springer: Berlin/Heidelberg, Germany, 2011; pp. 215–225.

293. Kotlarek, J.; Kwon, O.H.; Ma, K.L.; Eades, P.; Kerren, A.; Klein, K.; Schreiber, F. A Study of Mental Maps in Immersive Network Visualization. In Proceedings of the IEEE Pacific Visualization Symposium (PacificVis), Tianjin, China, 3–5 June 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–10.

294. Ens, B.; Bach, B.; Cordeil, M.; Engelke, U.; Serrano, M.; Willett, W.; Prouzeau, A.; Anthes, C.; Büschel, W.; Dunne, C.; et al. Grand challenges in immersive analytics. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2021; pp. 1–17.

295. Heidrich, D.; Meinecke, A.; Schreiber, A. Towards a Collaborative Experimental Environment for Graph Visualization Research in Virtual Reality. Available online: https://diglib.eg.org/bitstream/handle/10.2312/evp20211068/009-011.pdf?sequence=1&isAllowed=y (accessed on 19 September 2021).

296. Kraus, M.; Klein, K.; Fuchs, J.; Keim, D.A.; Schreiber, F.; Sedlmair, M. The Value of Immersive Visualization. *IEEE Comput. Graph. Appl.* **2021**, *41*, 125–132. [CrossRef] [PubMed]

297. Sereno, M.; Wang, X.; Besançon, L.; McGuffin, M.J.; Isenberg, T. Collaborative work in augmented reality: A survey. *IEEE Trans. Vis. Comput. Graph.* **2020**, 1. (Just Accepted). [CrossRef] [PubMed]

298. Royston, S.; DeFanti, C.; Perlin, K. A collaborative untethered virtual reality environment for interactive social network visualization. *arXiv* **2016**, arXiv:1604.08239.

299. Drogemuller, A.; Cunningham, A.; Walsh, J.; Ross, W.; Thomas, B.H. VRige: Exploring social network interactions in immersive virtual environments. In Proceedings of the International Symposium on Big Data Visual Analytics (BDVA), Konstanz, Germany, 17–19 October 2017; IEEE: Piscataway, NJ, USA, 2017.

300. Sorger, J.; Waldner, M.; Knecht, W.; Arleo, A. Immersive analytics of large dynamic networks via overview and detail navigation. In Proceedings of the IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), Utrecht, The Netherlands, 14–18 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 144–1447.

301. Riegler, A.; Anthes, C.; Jetter, H.C.; Heinzl, C.; Holzmann, C.; Jodlbauer, H.; Brunner, M.; Auer, S.; Friedl, J.; Fröhler, B.; et al. Cross-Virtuality Visualization, Interaction and Collaboration. Available online: https://www.researchgate.net/profile/Andreas-Riegler-2/publication/346082457_Cross-Virtuality_Visualization_Interaction_and_Collaboration/links/5fba8af7299bf104cf6cda09/Cross-Virtuality-Visualization-Interaction-and-Collaboration.pdf (accessed on 17 September 2021).

302. De Vries, J.H.; Spengler, M.; Frintrup, A.; Mussel, P. Personality Development in Emerging Adulthood—How the Perception of Life Events and Mindset Affect Personality Trait Change. *Front. Psychol.* **2021**, *12*, 671421. [CrossRef]

303. Graham, E.K.; Weston, S.J.; Gerstorf, D.; Yoneda, T.B.; Booth, T.; Beam, C.R.; Petkus, A.J.; Drewelies, J.; Hall, A.N.; Bastarache, E.D.; et al. Trajectories of big five personality traits: A coordinated analysis of 16 longitudinal samples. *Eur. J. Personal.* **2020**, *34*, 301–321. [CrossRef]

304. Hoff, K.A.; Song, Q.; Einarsdóttir, S.; Briley, D.A.; Rounds, J. Developmental structure of personality and interests: A four-wave, 8-year longitudinal study. *J. Personal. Soc. Psychol.* **2020**, *118*, 1044. [CrossRef]

305. Roberts, B.W.; Mroczek, D. Personality trait change in adulthood. *Curr. Dir. Psychol. Sci.* **2008**, *17*, 31–35. [CrossRef]

306. Carden, L.; Wood, W. Habit formation and change. *Curr. Opin. Behav. Sci.* **2018**, *20*, 117–122. [CrossRef]

307. Cadilhac, A.; Asher, N.; Lascarides, A.; Benamara, F. Preference change. *J. Logic. Lang. Inf.* **2015**, *24*, 267–288. [CrossRef]

308. Hoff, K.A.; Briley, D.A.; Wee, C.J.; Rounds, J. Normative changes in interests from adolescence to adulthood: A meta-analysis of longitudinal studies. *Psychol. Bull.* **2018**, *144*, 426. [CrossRef] [PubMed]

309. Chaabene, N.E.H.B.; Bouzeghoub, A.; Guetari, R.; Ghezala, H.H.B. Deep learning methods for anomalies detection in social networks using multidimensional networks and multimodal data: A survey. *Multimed. Syst.* **2021**, 1–11. [CrossRef]

310. Mahmood, B.; Alanezi, M. Structural-Spectral-Based Approach for Anomaly Detection in Social Networks. *Int. J. Comput. Digit. Syst.* **2021**, *10*, 343–351. [CrossRef]

311. Rengarajan, R.; Babu, S. Anomaly Detection using User Entity Behavior Analytics and Data Visualization. In Proceedings of the 8th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 17–19 March 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 842–847.

312. Wanda, P.; Jie, H.J. DeepFriend: Finding abnormal nodes in online social networks using dynamic deep learning. *Soc. Netw. Anal. Min.* **2021**, *11*, 1–12. [CrossRef]

313. Noorossana, R.; Hosseini, S.S.; Heydarzade, A. An overview of dynamic anomaly detection in social networks via control charts. *Qual. Reliab. Eng. Int.* **2018**, *34*, 641–648. [CrossRef]

314. Savage, D.; Zhang, X.; Yu, X.; Chou, P.; Wang, Q. Anomaly detection in online social networks. *Soc. Netw.* **2014**, *39*, 62–70. [CrossRef]

315. Cheng, J.; Bernstein, M.; Danescu-Niculescu-Mizil, C.; Leskovec, J. Anyone can become a troll: Causes of trolling behavior in online discussions. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, Portland, OR, USA, 25 February–1 March 2017; pp. 1217–1230.

316. Masud, S.; Dutta, S.; Makkar, S.; Jain, C.; Goyal, V.; Das, A.; Chakraborty, T. Hate is the New Infodemic: A Topic-aware Modeling of Hate Speech Diffusion on Twitter. In Proceedings of the IEEE 37th International Conference on Data Engineering (ICDE), Chania, Greece, 14–22 April 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 504–515.

317. Makkar, S.; Chakraborty, T. Hate Speech Diffusion in Twitter Social Media. Ph.D. Thesis, IIIT-Delhi, New Delhi, India, 2020.

318. Castano-Pulgarín, S.A.; Suárez-Betancur, N.; Vega, L.M.T.; López, H.M.H. Internet, social media and online hate speech. Systematic review. *Aggress. Violent Behav.* **2021**, *25*, 101608. [CrossRef]

319. Salminen, J.; Sengün, S.; Corporan, J.; Jung, S.g.; Jansen, B.J. Topic-driven toxicity: Exploring the relationship between online toxicity and news topics. *PLoS ONE* **2020**, *15*, e0228723. [CrossRef]

320. Lee, S.H.; Kim, H.W. Why people post benevolent and malicious comments online. *Commun. ACM* **2015**, *58*, 74–79. [CrossRef]

321. Cheng, J.; Danescu-Niculescu-Mizil, C.; Leskovec, J. Antisocial behavior in online discussion communities. In Proceedings of the Ninth International AAAI Conference on Web and Social Media, Oxford, UK, 26–29 May 2015.

322. Kumar, D.; Kelley, P.G.; Consolvo, S.; Mason, J.; Bursztein, E.; Durumeric, Z.; Thomas, K.; Bailey, M. Designing Toxic Content Classification for a Diversity of Perspectives. *arXiv* **2021**, arXiv:2106.04511.

323. Fiesler, C. Toward a Multi-Stakeholder Perspective for Improving Online Content Moderation. Ph.D. Thesis, University of Michigan, Ann Arbor, MI, USA, 2019.

324. Kooti, F.; Yang, H.; Cha, M.; Gummadi, K.P.; Mason, W.A. The emergence of conventions in online social networks. In Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, Dublin, Ireland, 4–7 June 2012.

325. Chen, S.; Andrienko, N.; Andrienko, G.; Adilova, L.; Barlet, J.; Kindermann, J.; Nguyen, P.H.; Thonnard, O.; Turkay, C. LDA ensembles for interactive exploration and categorization of behaviors. *IEEE Trans. Vis. Comput. Graph.* **2019**, *26*, 2775–2792. [CrossRef] [PubMed]

326. Benevenuto, F.; Rodrigues, T.; Cha, M.; Almeida, V. Characterizing user behavior in online social networks. In Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, Chicago, IL, USA, 4–9 November 2009; pp. 49–62.

327. Uhde, A.; Tretter, S.; von Terzi, P.; Koelle, M.; Diefenbach, S.; Hassenzahl, M. Interaction in the Public: Aesthetics, Social Acceptability, and Social Context. *Mensch Und Comput.-2021-Work.* **2021**. (Just Accepted).

328. Uhde, A.; Hassenzahl, M. Towards a Better Understanding of Social Acceptability. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021; pp. 1–6.